Monitoring Coastal Environments using UAS Imagery and Deep Learning

Brandon Hobley

A thesis presented for the degree of
Doctor of Philosophy

School of Computer Science
University of East Anglia
United Kingdom
January 2023

Abstract

Coastal monitoring is a complex mapping problem for environments that exhibit distinct physical variations through the energy expended from water and sediment movement. In recent years, sensor platforms that capture imagery from these environments have reached centimeter level pixel resolution, which allowed object-based image processing methods to become a standard mapping tool. However, this tool still adheres to shallow machine learning methods, whereby the construction of a learning system is broken into two steps: feature extraction and machine learning model optimisation.

In the last decade, deep learning and convolutional neural networks have established state-of-the-art performance on a myriad of computer vision applications. However, deep learning models perform best with large, labelled, training datasets. For coastal monitoring, ground-truth observations can be acquired either in-situ or through post-processed imagery, but both avenues require manual process in producing the ground-truth annotations. In turn, this requires laborious and expensive efforts with domain expertise of coastal processes, posing a bottleneck and challenge for accurate coastal monitoring.

In this thesis, practical applications of coastal monitoring using deep learning and convolutional neural networks are discussed. These methods attempt to improve the performance and generalisation of convolutional neural networks with limited amounts of labelled data, which could ease costs of producing ground-truth annotations. A number of approaches are described that reduce the effort required to produce them, or analyse the feasibility of non-domain expert labels.

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate ‘take down’ action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

Monitoring Coastal Environments using UAS Imagery and Deep Learning

Brandon Hobley

© This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Acknowledgements

I would like to thank my supervisory team, Dr. Michal Mackiewicz and Prof. Graham Finlayson for their support, advice, and help throughout this project.

I would like to thank the supervisory team from the Centre for Environment, Fisheries and Aquaculture Science for guiding and assisting me throughout the research project. During my research, I had the opportunity to collaborate with Julie Bremner, Riccardo Arosio and Tony Dolphin whom helped me join the site survey for the methods developed in Chapter 4 and understand the ecological importance for each study site. We would also like to thank the Environment Agency and Cefas for providing very high resolution and annotated records for the work conducted in Chapters 3 and 4

I would also like to thank participants from Cefas and the University of East Anglia for annotating the imagery necessary to conduct an inter-observer experiment in Chapter 3.

I would like to thank my loving family and partner for supporting and motivating me throughout the research. I would also like to thank my colleagues, Chloe Game, Artjom Gorpincenko and Jake McVey for making office days entertaining and social throughout. Finally, I'd like to thank my good friend Tyson Lloyds for pushing to embark on a research project.

This research was funded by Cefas, EA and the Natural Environmental Research Council through Industrial CASE, grant number NE/R007888/1, titled "Blue eyes: New tools for monitoring coastal environments using remotely piloted aircraft and machine learning"

Contents

Abstract **i**

Acknowledgements **iv**

Contents **iv**

List of Figures **ix**

List of Tables **xxiii**

1 Introduction **1**

1.1 Coastal remote sensing 3

1.2 Deep learning for remote sensing 4

1.3 Thesis outline 9

1.4 Publications 12

2 Literature Review **14**

2.1 Sensor platforms 14

2.1.1 Satellite imagery 15

2.1.2 Commercial satellite imagery 18

2.1.3 Uncrewed aircraft system imagery 21

2.1.4 Critical analysis 24

2.2 Thematic mapping 25

2.2.1 Pixel-based mapping 26

2.2.2 Object-based mapping 28

2.2.3 Critical analysis 32

2.3 Deep neural networks . . . 34
2.3.1 Image classification . . . 34
2.3.2 Semantic segmentation . . . 37
2.3.3 Transfer learning . . . 43
2.3.4 Data-augmentation . . . 44
2.3.5 Semi-supervision . . . 46
2.4 Hyperspectral reconstruction . . . 55
2.5 Multi-task learning . . . 58
2.6 Conclusions . . . 60

3 Budle Bay - semi-supervised and crowd sourced learning for intertidal seagrass mapping 65

3.1 Introduction . . . 65
3.2 Data collection and in-situ survey . . . 66
3.3 Semi-supervised intertidal seagrass mapping . . . 71
3.3.1 Methodology . . . 73
3.3.2 Accuracy assessment . . . 88
3.3.3 Experiments and Results . . . 89
3.3.4 Discussion . . . 94
3.3.5 Summary . . . 100
3.4 Crowdsourcing experiment for intertidal seagrass mapping . . . 101
3.4.1 Background . . . 102
3.4.2 Inter-observer variability experiment with Cochran's Q . . . 104
3.4.3 Results and interpretation . . . 117
3.4.4 Summary . . . 124
3.5 Conclusions . . . 125

4 Sizewell - hyperspectral reconstruction and multi-task learning 129

4.1 Introduction . . . 129

4.2 Sizewell study site: background and rationale 130
4.2.1 Data collection and in-situ survey 134
4.2.2 Outline 139
4.3 Hyperspectral reconstruction on ICVL 142
4.3.1 Sparse-dictionary representation for HSI reconstruction 143
4.3.2 HSCNN-R and HS-UNet-R for HSI reconstruction 145
4.3.3 Quality assessment 150
4.3.4 Discussion 153
4.3.5 Summary 157
4.4 Hyperspectral reconstruction for strandline and sand-dune communities . . 157
4.4.1 Background 158
4.4.2 Methodology 160
4.4.3 Quality assessment 162
4.4.4 Discussion 163
4.4.5 Summary 173
4.5 Multi-task learning: segmentation with hyperspectral reconstruction 174
4.5.1 Background 175
4.5.2 Methodology 176
4.5.3 Quality assessment 181
4.5.4 Discussion - species level mapping with FCNs 209
4.5.5 Discussion - merged classes and comparison with OBIA 216
4.5.6 Summary 221
4.6 Conclusions 223

5 Conclusions and future work **226**

5.1 Semi-supervised intertidal seagrass mapping 226
5.2 Crowdsourcing experiment for intertidal seagrass mapping 227
5.3 Hyperspectral reconstruction on multispectral Sizewell imagery 229

5.4 Multi-task learning for species at Sizewell study site 230
5.5 Thoughts on coastal remote sensing and Future work 231

6 Bibliography **240**

A Colour transfer for improved image registration **303**
A.1 Background 305
A.2 Linear Monge-Kantorovich transform for colour transfer 307
A.3 Experiments and results 307
A.4 Summary 310

List of Figures

2.1 Output of multi-resolution segmentation over a high-resolution orthomosaic of Budle Bay. The segmentation algorithm begins clustering image pixels together to form image-objects. The pixel clustering method maximises homogeneity in image-objects and heterogeneity between image-objects by leveraging statistical moments in each candidate cluster. 31

2.2 The progression of convolutional blocks for CNN feature extraction. The VGG architecture standardised network design with with 3×3 kernels, batch normalisation and a non-linear activation function [Simonyan and Zisserman, 2014]. The addition of residual connections allows training signals to propagate the identity function, and thus learn the residual function [He et al., 2016]. The ConvNeXt explores a different strategy by approximating a MLP with 1×1 discrete convolutions (as per network in network [Lin et al., 2013]). 37

2.3 Different segmentation architectures commonly found in literature which can be broadly described from (left to right) in the following categories: dilated convolutions, image pyramid, encoder-decoder, and spatial pyramid pooling. Each architecture has a unique method for performing pixel-wise classification but the strongest pixel-classifiers revolve around multi-scale feature extraction (dilated convolutions and spatial pyramid pooling) and encoder-decoder architectures with skip connections for accurate classification. . . . 42

2.4 Different semi-supervised network architectures used for generating consistency-based loss functions. The Π-Model (top architecture) generates pairs of predictions using a single model and two different stochastic image augmentation processes, generating two different sets of predictions, $\hat{y}$ and $\bar{y}$. The latter drive the unsupervised consistency loss function and the label y and $\hat{y}$ drive the supervised loss. Temporal ensembling (middle architecture) generates a single set of predictions $\bar{y}$ which drives the standard supervised loss and the unsupervised consistency loss through an EMA of predictions from the previous iteration $\hat{y}$. Mean-teacher (bottom architecture) uses two networks to generate pairs of predictions, the student network which drives the supervised loss and a teacher network which is an EMA of the weights in the student network. The unsupervised consistency loss is driven by the predictions generated from the student and teacher network [Tarvainen and Valpola, 2017] 63

2.5 Different MTL paradigms and subsequent effects on network design. On the left, the hard parameter sharing uses a single network to output to learn multiple image tasks that are dictated by the amount of output heads. On the right, the soft parameter sharing uses multiple networks for each image task but employs a multi-task loss to jointly optimise each network. 64

3.1 Orthomosaics of Budle Bay using a SONY ILCE-6000 (left) and a MicaSense RedEdge3 multispectral camera (right). Driving the display uses the Red, Green and Blue image bands. A close-up of each orthomosaics is shown in Figure 3.2. 68

3.2 A close-up of each orthomosaics with water channels and bordering intertidal vegetation such as *Enteromorpha sp.* and Seagrass (*Zostera noltii* and *Angustifolia*) among background sediment. 69

3.3 Example of quadrat sampling for monitoring *Zostera marina* at Porth Dinllaen. Figure from [Davies et al., 2017] 71

3.4 Distribution of recorded tags during the in-situ survey performed by a team of expert ecologists from Cefas and the EA. Each point records the percentage cover of each target class using 300×300mm quadrats. Then, each sample is reduced to a single label by selecting the highest percentage cover. 72

3.5 The process of generating training images for FCNs. First, the orthomosaics were split into 6000×6000 non-overlapping tiles. Second, the classified in-situ points, as shown in Figure 3.4, provide the basis to draw polygons using expert photo-interpretation. Lastly, the list of tiles and the rasterised polygons were combined by searching through each image tile for a rasterised polygon and then cropping the tile to a 256×256 image which is used to drive optimisation of FCNs. 75

3.6 Gallery of images for each class with accompanying in-situ points that were used to annotate polygons. On the right are rasterised polygons which were generated through expert photo-interpretation and the knowledge conveyed by the classified in-situ point which can be seen in the middle column of images. The left column of images provides the same orthomosaic crop with no overlay of in-situ classified points and rasterised polygons. 76

3.7 Gallery of images with annotated polygons. Each polygon contains a unique semantic value that drives the optimisation of fully convolutional neural networks. OM - Other Macroalgae inc. *Fucus*; MB - *Microphytobentos*; EM - *Enteromorpha*; SM - Saltmarsh; SG - Seagrass; DS - Dry Sand; OB - Other Bareground. Images with white polygons are examples of polygons used for modelling. 77

3.8 U-Net architecture and loss calculation. The input channels are stacked and passed through the network. The encoder network applies repeated convolution and max pooling operations to extract feature maps, while in the decoder network upsamples these and stacks features from the corresponding layer in the encoder path. The output is a segmented map that is compared with the mask using cross-entropy loss. The computed loss is used to train the network, through gradient descent optimisation 79

3.9 Distribution of labelled pixels within the polygons for each class and non-labelled pixels on a log-scale. Given each training image was a 256×256 crop of the orthomosaic centered around each polygon, the number of non-labelled pixels refers to image pixels outside the polygon that reside within the 256×256 crop. 80

3.10 Mean-Teacher architecture for semi-supervised segmentation using pseudo-labels. The student network produces predictions that are used to compute the supervised loss with known polygons. The teacher network produces predictions that are then converted to hard pseudo-labels which are then used to drive the unsupervised loss. The student is updated using the combined loss and gradient descent, and the teacher network is updated with an EMA of the student network weights. 83

3.11 Segmented orthomosaics using the multiresolution segmentation algorithm. On the top-left (**A**) is a crop of the orthomosaic captured with the SONY-ILCE 6000. On the top-right (**B**) is is a crop of the orthomosaic captured with the MicaSense RedEdge3 multispectral camera. The corresponding segmented orthomosaics crop are shown on the bottom-left and bottom-right (**C** and **D**, respectively for the SONY-ILCE 6000 and the MicaSense RedEdge3). Both segmented orthomosaics were extracted with the same hyper-parameters. The scale, shape and compactness were respectively 200, 0.1 and 0.5. Given the spatial resolution from both orthomosaics is different, the resulting segmented orthomosiacs are also different since both use the same scale parameter 86

3.12 Confusion matrices for both methods using the SONY camera. 90

3.13 Confusion matrices for both methods using the RedEdge3 multispectral camera. 92

3.14 Segmented habitat maps for both cameras and methods. The top row - RedEdge3 and SONY cameras orthomosaics. The second row - habitat maps using the OBIA approach. The third row - FCNN maps in a supervised setting. The bottom row - FCNN maps in a semi-supervised setting. The left column - RedEdge3 images and segmented maps, the right column - the SONY images and maps. Legend: OM - Other Macroalgae inc. *Fucus*; MB - *Microphytobentos*; EM - *Enteromorpha*; SM - Saltmarsh; SG - Seagrass; DS - Dry Sand; OB - Other Bareground. 93

3.15 Sample images representative of vegetation classes used during the analyses. The ground photographs were taken during the in-situ survey by expert ecologists. 108

3.16 The user interface that was provided for participant annotations during the experiment. The user interface uses ArcMap 10.6.1 to visualise and annotate samples. 109

3.17 Confusion matrices for the majority vote annotations for each control group. 111
3.18 U-Net architecture and loss calculation. The encoder network, now a VGG-13, extracts feature maps. The decoder network upsamples features from the corresponding layer in the encoder path. 116
3.19 Confusion matrices for FCNN models trained using different versions of labelled data. Results for models trained on in-situ labels (top-left) and majority-vote annotations for group A (top-right), group B (bottom-left) and group C (bottom-right). The normalised percentage accuracy is shown along the diagonal of the confusion matrix 118
3.20 Confusion matrices for FCNN models trained using set of in-situ labels (left), and using the same in-situ set supplemented with majority-vote annotations for groups A, B and C (top-right, bottom-left, bottom-right). The normalised percentage accuracy is shown along the diagonal of the confusion matrix . . 119

4.1 Study site within the U.K. (top-left). Ortho-registered images of Sizewell using the Sentera multispectral sensor. On the right there is a close-up using Red, Green and Blue channels as well as a false-colour RGB using Red, Near-Infrared and Green channels. 131
4.2 A close-up of the orthomosaic shown in Figure 4.1. The close-up shows that the Eastern part closest to the sea is dominated by shingle and sand sediment. Then, transitioning from East to West along the orthomosaic, shows the pioneering species on shingle or sand sediment mainly belong to SD1 and SD2 NVCs and the tall grassland communities after the pioneering species that mainly belong to SD6 and SD7 NVCs. Red dots on the orthomosaic correspond to in-situ geo-referenced tags with known classifications of shingle vegetation species. 132

4.3 Plots of reflectance measurements using the hand-gun. The left plot shows the entire bandwidth of the FieldSpec4 (400-2500nm) and the right plot has the same bandwidth as the Sentera multispectral camera (400-900nm). Each species plot is an average of all collected hand-gun measurements during the survey. AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SU- Silene uniflora 136
4.4 Sentera camera responses for RGB colour space (400-700nm) and RGB + NIR colour space (400-900nm). 138
4.5 Shingle vegetated species that belong to SD1, SD2 and SD6 NVCs. These species were sampled and classified by an expert taxonomist during the in-situ survey at Sizewell. For each sample, a hyperspectral measurement, along with a photograph and a RTK GPS measurement were used for further processing. Each image was captured in either the first or second ground-survey. 140
4.6 A gallery of sample images and various outdoor scenes in the ICVL dataset. The images represent a particular hyperspectral channel in the visible spectrum between 400-700nm. The images projected to the Sentera multispectral colour space integrate the continuous hyperspectral signature captured with Specim PS Kappa DX4 with the filters shown in Figure 4.4 (left-plot). . . . 143
4.7 Flowchart of imagery and methods used in [Arad and Ben-Shahar, 2016]. . 145
4.8 Encoder network for HSCNN-R [Shi et al., 2018]. C - 3×3 represents a convolutional block with ReLU non-linearity. 146
4.9 The HS-UNet-R architecture using the HSCNN-R [Shi et al., 2018] as an encoder network. C - 3×3 represents a convolutional block with ReLU non-linearity. TC - 2×2 represents a transposed convolutional block for learnt feature map upsample. Before each residual convolutional block in the encoder is a pooling operation that downsamples the image resolution by a factor of 2 147

4.10 Plots for predicted reconstructed hyperspectral response from RGB. Top-left shows a comparison of supervised models with the shallow method in [Arad and Ben-Shahar, 2016]. Top-right shows plots for semi-supervised models. Bottom-left show plots for self-supervised models. And, the bottom-right plots shows the effects of varying the smoothing loss gain, L_{smo}, for different training procedures. The plots show continuous spectral measurements for Red, Green, Blue samples from a colour checker present in some images of the dataset. The error units are the same units as the quantity being estimated using the Specim PS Kappa DX4 hyperspectral camera which in this case are 12-bit pixel values. 151

4.11 Gallery of predicted images for the methods stated in Sections 4.3.1 and 4.3.2. The predicted images for HSCNN-R and HS-UNet-R are trained in a supervised setting. 152

4.12 Gallery of images with corresponding masks used for training and testing. AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SU- Silene uniflora 162

4.13 Hand-gun measurement for a sample of Crambe maritima. Twenty five hyperspectral measurements were recorded at different locations of this plant of Crambe maritima. The measurements were averaged to a single spectral signature, as shown with the red line (left-plot), and then copied to every pixel in the polygon corresponding to this particular plant of Crambe maritima (arrows to right-figure). 163

4.14 Plots for predicted reconstructed hyperspectral reflectance from RGB. Plots show supervised models using equation 4.10 to reconstruct reflectance. . . 165

4.15 Visual results for reconstructed hyperspectral reflectance from multispectral Sentera imagery using HSCNN-R. Visual results show supervised models using equation 4.10 to reconstruct reflectance. 166

4.16 Visual results for reconstructed hyperspectral reflectance from multispectral Sentera imagery using HS-UNet-R. Visual results show supervised models using equation 4.10 to reconstruct reflectance. 167
4.17 Plots for predicted reconstructed hyperspectral radiance from RGB. Plots show self-supervised models using equation 4.12 to reconstruct radiance. . 170
4.18 Visual results for reconstructed hyperspectral radiance from multispectral Sentera imagery using HSCNN-R. Visual results show self-supervised models using equation 4.12 to reconstruct radiance. 171
4.19 Visual results for reconstructed hyperspectral radiance from multispectral Sentera imagery using HS-UNet-R. Visual results show self-supervised models using equation 4.12 to reconstruct radiance. 172
4.20 The MTL architecture used for the results shown in Section 4.5.3. The shared model is the HS-UNet-R and the proposed MTL architecture alternates between each task to be learnt by forward passing the segmentation dataset, followed by one of the versions of the hyperspectral reconstruction dataset. 180
4.21 A close-up of cropped orthomosaics of Sizewell beach segmented using the multiresolution segmentation algorithm for Red, Green and Blue channels. The scale, shape and compactness were respectively 50, 0.2 and 0.5. 182
4.22 The generated orthomosaic from the aerial survey projected to RGB colour space. The multispectral orthomosaic is narrow which prevents the visual analysis of generated thematic maps from each method described in Section 4.5.2. 184
4.23 Cropped close-ups of the orthomosaic used for the visual analysis of FCNs optimised using various strategies as shown in Sections 3.3 and 4.5. 185
4.24 Orthomosaic of Sizewell beach with geographical coordinates. The figures **A**, **B** and **C** show close-ups of the shingle beach along with assemblages of vegetated shingle communities used for the comparison of FCNs with the OBIA. For the display, the Red, Green and Blue image bands were used. . 186

4.25 Confusion matrices for both supervised, semi-supervised and MTL using the HS-UNet-R. MTL - supervised refers to models trained for supervised segmentation with supervised reflectance recovery. Each confusion matrix for FCNs shows the average pixel accuracy over 5 independent train and test run. MTL - self-supervised refers to models trained for supervised segmentation with self-supervised radiance recovery. 188

4.26 Confusion matrices showing pixel accuracy scores for OBIA and FCNs optimised in a supervised, semi-supervised with consistency regularisation and teacher/student networks and semi-supervised multi-task learning. Each confusion matrix for FCNs shows the average pixel accuracy over 5 independent train and test run. Legend: Mature G - Mature Grassland; Pioneering G. - Pioneering grassland; Young P. - Young pioneering; Crambe M. - Crambe Maritima 189

4.27 Thematic maps for supervised HS-UNet-R without MTL. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora. 190

4.28 Thematic maps for supervised HS-UNet-R without MTL. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora. 191

4.29 Thematic maps for supervised HS-UNet-R without MTL. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora. 192

4.30 Thematic maps for supervised HS-UNet-R without MTL. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora. 193
4.31 Thematic maps for semi-supervised HS-UNet-R without MTL. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora. 194
4.32 Thematic maps for semi-supervised HS-UNet-R without MTL. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora. 195
4.33 Thematic maps for semi-supervised HS-UNet-R without MTL. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora. 196
4.34 Thematic maps for semi-supervised HS-UNet-R without MTL. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora. 197
4.35 Thematic maps for supervised HS-UNet-R with MTL using supervised reflectance as auxiliary. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora. 198

4.36 Thematic maps for supervised HS-UNet-R with MTL using supervised reflectance as auxiliary. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora. 199
4.37 Thematic maps for supervised HS-UNet-R with MTL using supervised reflectance as auxiliary. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora. 200
4.38 Thematic maps for supervised HS-UNet-R with MTL using supervised reflectance as auxiliary. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora. 201
4.39 Thematic maps for supervised HS-UNet-R with MTL using self-supervised radiance as auxiliary. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora. 202
4.40 Thematic maps for supervised HS-UNet-R with MTL using self-supervised radiance as auxiliary. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora. 203

4.41 Thematic maps for supervised HS-UNet-R with MTL using self-supervised radiance as auxiliary. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora. 204
4.42 Thematic maps for supervised HS-UNet-R with MTL using self-supervised radiance as auxiliary. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora. 205
4.43 Segmented habitat maps and close-up of **A** in Figure 4.24 for each of the mentioned methods in Section 4.5. The visual results were generated with the best performing network from each optimisation strategy. Legend: Mature G - Mature Grassland; Pioneering G. - Pioneering grassland; Young P. - Young pioneering; Crambe M. - Crambe Maritima 206
4.44 Segmented habitat maps and close-up of **B** in Figure 4.24 for each of the mentioned methods in Section 4.5. Legend: Mature G - Mature Grassland; Pioneering G. - Pioneering grassland; Young P. - Young pioneering; Crambe M. - Crambe Maritima 207
4.45 Segmented habitat maps and close-up of **C** in Figure 4.24 for each of the mentioned methods in Section 4.5. Legend: Mature G - Mature Grassland; Pioneering G. - Pioneering grassland; Young P. - Young pioneering; Crambe M. - Crambe Maritima 208

A.1 A is the high-resolution reference image with corresponding target image (B) and the application of the linear MK transform (C) mapping the colours in the reference image (A) to match those within the target image (B). 305
A.2 An example illustrating the mapping of multivariate Gaussian distributions for colour distributions u and v. 306

A.3 Gallery of images to be registered. Top-row are high-resolution reference images from the SONY camera and the bottom-row are multispectral target images from the RedEdge3 camera. 308
A.4 Registered images from both cameras after conversion to grey scale. A - pre-processed with the MK colour transfer and B – registered using the original reference. 310

List of Tables

1.1 Various types of coastal environments with associating communities. 1

3.1 Precision, recall and F1 scores for both algorithms on both cameras. The results for FCNs reflect models trained using semi-supervision. DS - Dry Sand; OB - Other bareground; EM - Enteromorpha; MB - Microphytobentos; OM - Other macroalgae; SG - Seagrass; SM - Saltmarsh 94

3.2 Participant annotation accuracy and Cochran Q test statistic results. Participants were grouped into three different groups. Group A were expert ecologists or geomorphologists, present at the in-situ survey and/or had previous experience with annotating marine biology for the study site based. Group B were expert ecologists or geomorphologists that were not present at the in-situ survey and/or did not have experience with annotating marine biology for the study site. Group C were non an experts without experience in annotating marine biology from aerial imagery. The significance level, defined by the parameter α, give critical values according to a χ^2 distribution which in turn may or may not reject the statistical test. 110

3.3 Precision, recall and F1 scores for models trained with in-situ labels and for models trained with majority-vote annotations from group A. DS—Dry Sand; OB—Other bareground; EM—Enteromorpha; MB—Microphytobentos; OM—Other macroalgae; SG—Seagrass; SM—Saltmarsh 117

4.1 Various NVCs present at the Sizewell study site. 133

4.2 Recorded species from SD1B, SD2, SD6A, SD7C and SD19 NVCs along with the number of recorded plants per species, number of HS samples and corresponding DSLR captures 137

4.3 Root mean squared error and relative RMSE for predicted HSI and projected to RGB colour space using the sensor response function in Figure 4.4 (left-plot). The listed methods are sparse-dict as per Arad and Ben-Shahar [2016] and both HSCNN-R [Shi et al., 2018] and HS-UNet-R trained in supervised, semi-supervised or self-supervised settings 153

4.4 Root mean squared error and relative RMSE for predicted HSI and projected to RGB colour space using the sensor response function in Figure 4.4 (righ-plot) 164

4.5 Recall, precision and F1-scores for supervised, semi-supervised and MTL using the HS-UNet-R. The supervised column shows models optimised with equation 3.3, the semi-supervised column shows models optimised using both equations 3.3 and 3.4. MTL - supervised refers to models trained for supervised segmentation with supervised reflectance recovery. MTL - self-supervised refers to models trained for supervised segmentation with self-supervised radiance recovery. 187

4.6 Recall, precision and F1-scores for OBIA and FCNs in supervised, semi-supervised, and MTL semi-supervised with the architecture shown in Figure 4.20. Legend: Mature G - Mature Grassland; Pioneering G. - Pioneering grassland; Young P. - Young pioneering; Crambe M. - Crambe Maritima . . 209

A.1 Mean (± standard deviation) of the number of SIFT matches, percentage of inlier SIFT matches and Euclidean distance of pixel locations between control points in pairs of registered images 309

1 Introduction

Presently around 40% of the Earth's population live within 100 kilometers of the coast but far more benefit from resources and ecosystem services derived from coastal environments. As population density and economic activity grow, coastal environments will be subject to increasing pressure to meet human needs [Seas and Plans, 2011; Millennium ecosystem assessment, 2005]. Burke et al. [2001] describe coastal environments as either near-shore, intertidal, benthic and pelagic, with these habitats often coexisting to represent dynamic systems that directly or indirectly provide a vast range of ecosystem services for humans. Some of these services include: sequestering carbon [Fourqurean et al., 2012], cycling nutrients and elements [Nixon, 1981] and providing nurseries and fishing grounds for commercial fisheries [Sheaves et al., 2015]. Table 1.1 lists different coastal zones along with the common marine communities within them.

From these ecosystems, the intertidal coastal zone comprises some of the world's most productive and ecologically significant ecosystems. These environments represent physically varying environments through the energy expended with water and sediment movement [Alongi, 2020]. In these environments, intertidal seagrass and algae play an important role due to their contribution to tidal and energy management from currents and waves [Bouma et al., 2005], sediment quality and stability [Koch, 1999; Fonseca et al., 1983], with

Near-shore	Dunes, cliffs, rocky and sandy shores, urban, industrial and agricultural landscapes
Intertidal	Estuaries, deltas, lagoons, mangrove forests, mudflats, salt marshes, salt pans and aquaculture beds
Benthic	Kelp forests, seagrass beds and coral reefs
Pelagic	Open waters and freestanding fish farms:

Table 1.1: Various types of coastal environments with associating communities.

studies indicating sediment erosion following seagrass loss [Ramage and Schiel, 1999]. Furthermore, intertidal seagrass meadows contribute to the development of coastal ecosystem health by providing safe and rich fish nurseries [Whitfield and Pattrick, 2015]. The cyclical and delicate balance of nutrients, trophic pathways and ecological energetics that maintain ecosystem health are subtle and complex, with much of the temporal and spatial variation in intertidal marine organism stocks explained by intrinsic factors, such as genetics and reproductive strategies [Alongi, 2020]. Climate change poses an extrinsic challenge for intertidal ecosystems as changing atmospheric and ocean temperatures, sea levels, ocean chemistry and weather patterns disturb the delicate nutrient cycle and cause intertidal seagrass extents to regress [Waycott et al., 2011; Mieszkowska et al., 2013]. The impacts result in increased sediment erosion [Amos et al., 2004; Adriano et al., 2005] that in turn degrade coastal ecosystem health and reduce estuarine fish stocks due to underlying impacts to the nurturing grounds for fish nurseries [Moussa et al., 2020]. The evidence for extrinsic human pressure affecting global marine communities in intertidal coastal zones has been well documented over the last six decades. These include: excessive coastal development and sediment deposits, overfishing, mechanical damage by boats and fishing gear, logging and impacts from invasive species [Duarte et al., 2008; Adam, 2002; Duarte, 2002; Bellwood et al., 2004; Lotze et al., 2006]. Furthermore, sea-level rise consequent of climate change has long-term consequences for coastal ecosystems such as wetlands and coral reefs [Morris et al., 2002].

Open shore environments also represent a dynamic and variable environment in coastal zones [Alongi, 2020]. In particular, shoreline change analysis is a common monitoring application for evaluating health dynamics and vulnerability of communities and habitats to erosion hazards [Dewi et al., 2016]. However, extrinsic human pressure and climate change allow for erosion hazards, such as sea level rise and increased storminess [Dolan et al., 1991; Pendleton, 2010]. Furthermore, shorelines provide various vital regional and local services, including tourism, recreation, fisheries, trade, and aesthetic and cultural value [Astsatryan et al., 2022]

These pressing concerns to marine coastal ecology emphasise the need to create and act on strategies that maintain a sustainable balance of coastal ecosystem health, while also effectively managing the use of resources that are derived from these ecosystems [McCarthy et al., 2017; Pereira et al., 2010]. Coastal marine ecology requires the investigation of organisms and their environmental setting which can be provided with spatially explicit data, given the basic need for knowledge about the location and distribution of species [Aplin, 2005]. Consequently, aerial imagery through remote sensing has become a common data acquisition approach for ecological investigations providing ecologists with tools to monitor biophysical properties and controlling processes at high spatial and temporal resolutions, while also integrating in-situ and field techniques for mapping methodologies [Kerr and Ostrovsky, 2003; Klemas, 2009].

1.1 Coastal remote sensing

Richards and Richards [1999] define remote sensing as the process of measuring reflected energy from the Earth's surface using a sensor mounted on an aircraft platform. The generated image data from these sensors provide a platform for ecologists to assess and monitor sites on a wide variety of applications [Gens, 2010]. But also pose a challenge from an image processing perspective due to high-volumes of generated data [Chi et al., 2016].

Traditionally, satellite remote sensing is an excellent tool for monitoring coastal waters. The periodic sample period allows routine collection of a variety of observations over large and often inaccessible expanses of the coast and adjacent waters [Miller et al., 2005, Ch. 1, p. 21] [McCarthy et al., 2017].

However, due to different satellite sensors, imagery is captured at various temporal, spatial and spectral scales. Therefore, in order to efficiently use satellite imagery the user must consider three key parameters: spatial resolution, Field of View (FoV) and sampling frequency [Pease, 1991]. For coastal monitoring, spatial resolution is the key parameter of consider-

ation because of its obvious and apparent effects on imagery. If the spatial resolution is low (approx. 10m), then objects of interest will exhibit coarse texture. Contrarily, if the spatial resolution is high (approx. 10cm), then objects of interest may exhibit fine-grained texture. The choice of sensor should take into account the objective needs of the coastal monitoring application and the trade-off between high resolution imagery and generated data volume.

Fixed-wing Remotely Piloted Aircraft (RPA) along with commercially available cameras are a prominent avenue for coastal remote sensing as the altitude during data capture is low enough to produce high-resolution imagery. Adding to this, collecting overlapping Very-high resolution (VHR) images allows for Structure from Motion (SfM) techniques to be leveraged in order to create high-resolution orthomosaics that span past the FoV of a single image whilst maintaining high-resolution (commonly 0.1 - 2m per pixel) [Duffy et al., 2018; Turner et al., 2012]. Latest advancements in remotely sensed data acquisition make use of rotor-based drones for stable capture of optical imagery. The ability to pilot drones at lower altitudes is the driving factor for increased spatial resolution (usually less than 0.1m) [Bansod et al., 2017; Tang and Shao, 2015; Gray et al., 2018]. Miniaturization and integration of multispectral cameras into rotor-based drones also increase spectral resolution and prevent common problems in multispectral data acquisition such as registration and subsequent fusion [Wang et al., 2018; Ghassemian, 2016]. However, image registration is still a challenge in remotely sensed imagery, if images are captured at different sample times and illuminant [Zitova and Flusser, 2003].

1.2 Deep learning for remote sensing

Generally, the main goal for coastal monitoring is to accurately map various species and organisms within a coastal environment in order to capture the current health and intrinsic, or extrinsic, dynamics that contribute to the coastal ecosystem [Klemas, 2015]. The use of aerial imagery to depict spatial data is used to create a habitat map by segmenting images

into sets of meaningful classes such that the spatial distribution of ecological features can be assessed [Foody, 2002]. The need for rapid, cost-effective methods that capture the highly varying nature of intertidal and open-shore environments not only necessitates the use of remote sensing but also accurate mapping methodology [Hardisky et al., 1986; Cracknell, 1999].

Object Based Image Analysis (OBIA) [Blaschke, 2010] has been the main approach for supervised and unsupervised habitat mapping of coastal environments [Sreekesh et al., 2020; Ventura et al., 2018; Dronova, 2015; Heumann, 2011b]. The latter method is defined in two-stages: first, an initial unsupervised segmentation clusters pixels into image-objects which provides the grounds for extracting textural, spatial and spectral features [Su et al., 2008; Flanders et al., 2003]. Then, in a supervised learning setting, in-situ data are transcribed and superimposed on generated image-objects in order to learn the underlying relationship between known outcomes (in-situ records) and extracted features [Husson et al., 2016; Rasuly et al., 2010; Duffy et al., 2018; Innangi et al., 2019; Janowski et al., 2020] using conventional machine learning models such as: Decision Trees [Quinlan, 1990], Random Forests [Breiman, 2001], Naive Bayes [Rish et al., 2001] and Support Vector Machines (SVMs) [Boser et al., 1992]. The stated method for semantic segmentation fits under conventional supervised machine learning techniques, where the construction of a learning system can be broken down into two components: feature extraction and model tuning.

The issues for the current approach are two-fold:

1. Feature extraction is known as the process of transforming raw natural data into suitable internal representations of descriptive features that require careful engineering and considerable domain expertise. Therefore, ecologists in coastal monitoring are required to understand the underlying methodology of OBIA and correlate the image-object outputs with spatial features found in aerial imagery.

2. Recently deep learning has shown to be a robust alternative to conventional machine learning techniques and also produce state-of-the-art performance on various computer vision applications, including semantic segmentation Long et al. [2015].

Therefore, in parallel to the advancements in remote sensing and data acquisition, the field of Computer Vision (CV) has also improved in the last decade with Deep Learning (DL) and the introduction of Convolutional Neural Networks (CNNs) [Krizhevsky et al., 2012].

Deep learning methods are a form of representation learning that use raw natural data as an input to extract multiple levels of representation obtained by composing hierarchical non-linear modules in an end-to-end fashion [Bengio et al., 2013]. The impetus to each non-linear module is two-fold:

1. A discrete convolution filters the image. This operation has a plethora of applications in image processing but the desirable property of filtering images with respect to weights in a convolutional kernel [Goodfellow et al., 2016, Chapter 9.1].

2. A pooling operation semantically merges local patches in an image to a single pixel which reduces the spatial dimensions of internal representations and introduces translation invariance [Goodfellow et al., 2016, Chapter 9.2].

Each convolution kernel in a non-linear module is connected to subsequent modules through a shared weight mechanism [Bengio et al., 2013]. Therefore, each non-linear module layer filters the input image with respect convolutional weights and pools filtered images in order to create hierarchical representations, also known as feature maps. These operations exploit the hierarchical description of natural signals where higher-level features are composed of lower-level features. For instance in an image, local combinations of edges form motifs, motifs assemble into parts, and parts form objects [Zeiler and Fergus, 2014]. The network topology of CNNs can also be adapted for a wide variety of CV applications, such as image classification [Krizhevsky et al., 2012], semantic segmentation [Long et al., 2015], object detection [He et al., 2017].

Irrespective of network topology and intended application, the optimisation of CNNs in a supervised setting is based on an objective error metric computed between CNN outcomes and known outcomes, equation 1.2.

$$f(x) = f^M(...(f^2(f^1(x))) \tag{1.1}$$

$$E = L(y, f(x)) \tag{1.2}$$

Where, x is the input image, f^1 is the first non-linear layer, f^2 is the second layer, and so on until the last layer M. E is the computed error between known outcomes y and CNN outcomes $f(x)$ using an objective error metric L. The choice of L is fundamental in order to achieve state-of-the-art CNN performance as the error derivative of E is used to optimise the weights of convolutional kernels in the network. Since each layer is connected through a shared weight mechanism, the initial error derivative between y and $f(x)$ can be propagated from the output layer to the initial layer using the chain rule of derivatives [Rojas, 1996; LeCun et al., 2015a]. Then, each weight is adjusted using a gradient-descent solver, e.g. SGD [Sutskever et al., 2013], Adam [Kingma and Ba, 2014]. The ability of CNNs to learn hierarchical abstract representations of input imagery in a self-learning fashion using gradient descent in effect combines feature learning and supervised classifier training in one optimisation [LeCun et al., 2015a].

The known outcomes y in equation 1.2 are also referred to as labels. Labels are a pivotal concern in many real-world scenarios as CNNs are optimised based on an objective error metric between model outcomes and labels. For coastal environments, labels can be obtained through in-situ surveys which involves high logistic efforts, potential inaccuracies due to geo-location errors as well as sampling and observation bias [Congalton, 1991; Leitão et al., 2018], or through visual identification and delineation of polygons directly from orthomosaics [Kattenborn et al., 2019b; Wagner et al., 2019; Lopatin et al., 2019]. And so, the quality of labelled records in an image dataset is just as important as the choice of error metric [Zlateski et al., 2018; Alonso, 2015].

In spite of CNNs' success, these models perform best with large labelled training datasets [Tarvainen and Valpola, 2017] that are often unavailable in coastal monitoring. Semi-supervised deep learning are a branch of methods that attempt to achieve state-of-the-art performance with a few labelled training examples in tandem with a significant amount of unlabeled samples. In such a setting, semi-supervised methods are more applicable to real-world applications where the unlabeled data are readily available and easy to acquire, while labeled instances are often hard, expensive, and time-consuming to collect [Ouali et al., 2020] such as the case in coastal monitoring.

Overall, with improvements to remote sensing data acquisition and the introduction of CNNs, an opportunity surfaces to leverage the parallel improvements in both fields with application for accurate and efficient mapping of coastal marine features. Deep learning methods have been applied successfully to remotely sensed imagery in a variety of applications [Bowler et al., 2020; Xu et al., 2018; Hamdi et al., 2019; Li et al., 2017]. However, the general contributions of the thesis focus on the application of semi-supervised deep learning for coastal marine features with small labelled datasets in an attempt to bridge the gap between laborious labelling tasks, e.g., in-situ surveys and photo-interpretation, and efficient thematic habitat mapping. The following aims are listed:

The first aim and contribution was to develop a consistency-based regularisation loss function to aid the optimisation of fully convolutional neural networks in the presence of unlabelled training samples in the image dataset. This work was inspired by the use of mean-teacher networks and dual loss functions that achieve state-of-the-art performance with a subset of labelled samples for image classification and semantic segmentation datasets [Tarvainen and Valpola, 2017; French et al., 2020a].

As mentioned, labels for coastal monitoring can be acquired either through in-situ surveys, or through visual identification and delineation of polygons directly from orthomosaics [Kattenborn et al., 2019b; Wagner et al., 2019; Lopatin et al., 2019]. The second aim was to investigate the feasibility of crowdsourcing labels directly from very high resolution

orthomosaics, and compare the performance of fully convolutional neural networks trained with image samples that use crowdsourced labels versus labels derived from the in-situ survey. This aims to reduce laborious labelling efforts by a single domain expert and to investigate the feasibility of supplementing image datasets with crowdsourced labels in coastal monitoring applications.

The third and last aim was to develop a novel semi-supervised approach using fully convolutional neural networks that leverages multi-task learning. Multi-task deep learning aims to enhance the performance of a main image task by leveraging internal representation from auxiliary image tasks [Ruder, 2017]. In this scenario, the main image task is semantic segmentation given the objective in coastal mapping is to identify and accurately map multiple species in a particular coastal environment. The auxiliary image task is unsupervised spectral reconstruction which also provides another semi-supervised approach since the optimisation of fully convolutional neural networks is also performed with a significant amount of unlabeled image samples.

1.3 Thesis outline

During the research period I collaborated with the Centre for Environmental Fisheries and Aquaculture Sciences (Cefas) and the Environmental Agency (EA) whom provided two datasets with VHR imagery and in-situ records for developing supervised and semi-supervised deep learning models. Given each dataset represent different mapping objectives, the following chapters examine each study site individually. The following subsection lists the thesis structure as contributions to literature.

Chapter 2 reviews relevant literature regarding data acquisition in remote sensing and methods for habitat mapping in a variety of environments. The reviewed methods cover both pixel-based and object-based methods, both of which are heavily used for coastal habitat mapping. The literature review also introduces various CNN architectures for image

classification and semantic segmentation, and the final subsections cover semi-supervised deep learning methods for datasets with limited amounts of labelled data.

Chapter 3 shows the first contribution to research using consistency based dual loss with a mean-teacher framework to train Fully Convolutional Neural Networks (FCNs) on multiple marine species [Hobley et al., 2021a], section 3.3. The first dataset was an intertidal estuary located in Budle Bay, Northumberland, England (55.625°N, 1.745°W), captured using two miniaturized sensors with complementing properties. The intent with this dataset was to map species of intertidal seagrass among other vegetation species and unvegetated sediment. First, a description of the study site shows captured imagery and describes the target class domain for the mapping objective. The discussion includes mapping results and conclusions by providing a comparison with the object-based method known as OBIA and supervised and semi-supervised FCNs trained for semantic segmentation. Previous attempts to semi-supervised seagrass mapping perform image classification using methods such as domain adaption and consistency based regularisation with mean-teacher [Islam et al., 2020; Noman et al., 2021].

The second contribution in Chapter 3 focuses on analysing the feasibility of incorporating crowdsourced labels with an inter-observer variability experiment. Participants were invited to label a set of points spread across imagery of Budle Bay. The experiment population included experts in geomorphology and marine ecology as well as other fields in computing science and chemistry. The goal was to determine whether given crowdsourced annotations can supplement or even replace in-situ or single expert photo-interpreted polygons from aerial imagery in an effort to reduce logistical costs, Section 3.4.

The last contribution for the Budle Bay dataset (shown in the Appendix A) uses the linear Monge-Kantorovich Transform (MK-T) for accurate image registration using multiple cameras with different properties. Aerial surveys of Budle Bay were performed with two complementing cameras. One camera had high spatial resolution and low spectral resolution, while the other had low spatial but high spectral resolution. Common image registration

methods for remote sensing, e.g., SIFT + RANSAC for homography estimation, can be improved by applying the MK-Transform before image-registration in order to reduce the covariance shift between colour distributions in pairs of images to be registered [Hobley et al., 2021b].

Chapter 4 shows the final contribution in the thesis using a semi-supervised approach with Multi-task learning (MTL). The second dataset was an open-shore beach in Sizewell, Suffolk, England (55.2°N, 1.633°W). The target species to map were substantially different to Budle Bay and belong to strandline, supra-tidal and sand-dune communities. In Britain, the plant communities from natural, semi-natural and common artificial habitats are classified into distinct categories known as National Vegetation Classes (NVCs) [Rodwell and nature conservation committee , GB]. For Sizewell, the in-situ survey recorded samples for SD1, SD2, SD6 and SD7 NVCs that represent a rare and declining habitat worldwide that is found around the UK coastline [Randall, 2004]. Classifying these particular NVCs poses a challenge due to the variable and short-lived nature of these species. Shingle foreshores are unvegetated or sparsely vegetated, and the specialist plants belonging to these NVCs adapt to survive in harsh coastal conditions [Fuller and Randall, 1988; Scott, 1963; Fuller, 1987]. In turn, accurate mapping of coastal vegetated shingle can provide an indicator for coastal erosion and shoreline analysis.

The in-situ survey also presented an opportunity to incorporate hyperspectral measurements into the optimisation of deep learning models and allowed for the evaluation of deep hyperspectral reconstruction methods. A high-resolution spectroradiometer measured intrinsic reflectance properties of various species in the stated NVCs between 350-2500nms.

First, two methods for hyperspectral reconstruction were evaluated on the ICVL dataset in Section 4.3. These methods include a shallow method described in [Arad and Ben-Shahar, 2016] and deep learning models adapted for hyperspectral reconstruction [Shi et al., 2018]. The deep learning models developed for the ICVL dataset were applied to imagery of Sizewell in Section 4.4. Finally, Section 4.5 shows an MTL framework with a shared model

to learn semantic segmentation and hyperspectral reconstruction. The discussion of results compares fully convolutional neural networks trained in supervised and semi-supervised settings as shown in Section 3.3 with fully convolutional neural networks trained in a MTL framework, as well as a further comparison with OBIA.

Chapter 5 presents the final conclusions of the thesis and discusses future work.

1.4 Publications

This thesis covers three publications.

Semi-Supervised segmentation for coastal monitoring seagrass using RPA imagery - This work uses consistency-based regularisation with a teacher-student network topology to map multiple species of seagrass and algae among sediment. Published as "Semi-supervised segmentation for coastal monitoring seagrass using RPA imagery" Hobley, Brandon and Arosio, Riccardo and French, Geoffrey and Bremner, Julie and Dolphin, Tony and Mackiewicz, Michal. Remote Sensing, 13(9), 1741 [Hobley et al., 2021a].

Improving image registration using colour transfer methods in remote sensing applications - Leverages the properties of each camera used to survey Budle Bay and applies the MK-Transform to transfer the colour statistics of high spectral resolution images onto high spatial resolution images in an effort to improve the subsequent image registration process. Published as "Improving image registration using colour transfer methods in remote sensing applications" by Hobley, Brandon and Finlayson, G. D. and Arosio, Riccardo and Bremner, Julie and Dolphin, Tony and Mackiewicz, Michal. In The Congress of the International Color Association (No. 14, pp. 299-304) [Hobley et al., 2021b].

Crowdsourcing experiment and deep learning techniques for coastal remote sensing of seagrass and macro-algae - Explores the feasibility of crowdsourced labels for mapping multiple species of seagrass and algae at Budle Bay. An inter-observer experiment is conducted with multiple participants grouped into different levels of expertise. Then, the

discussion compares FCNs trained with crowdsourced labels versus FCNs trained solely with transcribed in-situ labels. Published as "Crowdsourcing experiment and fully convolutional neural networks for coastal remote sensing of seagrass and macro-algae" Hobley, Brandon and Bremner, Julie and Dolphin, Tony and Arosio, Riccardo and Mackiewicz, Michal. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing [Hobley et al., 2023].

2 Literature Review

The following sections review relevant literature in coastal remote sensing and deep learning methods. As such, the review is split into two main sections: first, different sensor platforms in remote sensing are compared in order to contrast different spatial and temporal resolutions for data acquisition. Then, two main methodologies for coastal mapping are reviewed that also contrast increasing spatial resolution in data acquisition. The second part introduces deep learning principles and various state-of-the-art network architectures for image classification and semantic segmentation. Then, methods for training networks with limited amounts of labelled data in an image dataset are reviewed. Finally, Sections 2.4 and 2.5 review methods for hyperspectral reconstruction and multi-task learning.

2.1 Sensor platforms

As mentioned in Section 1.1, coastal remote sensing platforms allow ecologists to assess the intrinsic, or extrinsic, factors of a coastal ecosystem by processing spatially explicit aerial imagery across multiple temporal perspectives [Anderson and Gaston, 2013]. A variety of coastal monitoring applications require broad captures of spatial extents to efficiently observe large areas of interest [Klemas, 2015]. Therefore, remotely sensed imagery and mapping techniques address application needs such as identifying and detailing the biophysical characteristics of species habitats, predicting the distribution of species and spatial variability, and detecting natural or human-caused changes at various scales [Kerr and Ostrovsky, 2003].

Sensor platforms for coastal mapping are available in a myriad of complementing properties regarding spatial, spectral and temporal scales. To distinguish spectral resolutions the cen-

ter wavelengths corresponding to each filter in a camera are registered on an electromagnetic spectrum. Standard Red Green and Blue (RGB) imagery captured from commercial wide-band cameras filter reflect light at wavelengths ranging from 400-700nm but for remote sensing platforms, and also in particular for coastal remote sensing, narrow-band multi-spectral cameras are used to filter reflected light at wavelengths ranging from 400-1200nm [Khorram et al., 2012]. A further extension is the use of hyperspectral sensor that comprise of large number of narrow-band filters relative to multispectral cameras and cover a higher range of wavelengths, commonly between 400-2500nm. These sensor platforms are expensive and produce a hyperspectral data cube for further Geographic Information System (GIS) processing. The distinction between standard wide-band RGB, narrow-band multi-spectral and narrow-band hyperspectral is key for remote sensing monitoring applications as the resulting data volume and or available imagery will directly impact the mapping outcome [Ramanath et al., 2005].

2.1.1 Satellite imagery

The use of imagery obtained from satellite instruments provide regional to global scale observations at repeated sampling intervals and are appropriate for continued monitoring of large geographic areas [Gould, 2000]. Furthermore, satellite data are in digital format and can be integrated with GIS that are useful for processing imagery and deriving habitat maps [Dahdouh-Guebas, 2002]. Given orbit cycles of satellites, this branch of sensor platforms are also useful for change detection studies on large geographical areas [Willis, 2015; Asokan and Anitha, 2019].

Current sensor platforms for satellite imagery include: Landsat 8-9 and Sentinel-2. The Landsat program is a NASA driven project that is the longest running project for satellite imagery acquisition of the Earth's surface. The seminal satellite known as Landsat-1 was launched in July 23rd, 1972 with a repeat coverage of 18 days, a spectral range of 0.5-1.1 μm and a spatial resolution of 60m per pixel. The latest launch was Landsat-8, which launched in February 11th, 2013 with a repeat coverage period of 16 days, a spectral range of 0.43-1.38

μm and a spatial resolution that varies between 15m for panchromatic image channels, 30m for multispectral channels and 100m for thermal imaging channels [Acharya and Yang, 2015]. The Sentinel-2 is another Earth observation mission lead by the Coppernicus Program that acquires optical imagery at high spatial resolution over land and coastal waters. The current revisit period for Sentinel-2A is ten days with a multispectral resolution that covers 0.44-2.2 μm wavelengths and a spatial resolution that varies between 10m for RGB and Near-infrared bands, 20m for multispectral tailored for vegetation mapping and 60m for aerosol and water vapour detection bands.

These image sources and associated projects are free and open-source and have been extensively used for a variety of coastal monitoring activities. The following studies show the use of satellite sensors for various coastal mapping applications and change detection.

Chopra et al. [2001] mapped the coastal wetland ecosystems of Harike in Punjab, India where large areas of land use surrounding a coastal wetland were classified into the following categories: built-up land, agricultural land, forest, wasteland and wetland. The scale of captured images was 1:50000 which was complementary for the mapping objective as the features mapped cover large geographical areas.

Fuller et al. [1998] used Landsat-TM imagery with 30m spatial resolution and integrates in-situ information to map the Sango Bay that comprises of swamps, grasslands, cultivated land and forests, bordering the shore of Lake Victoria in Uganda. The imagery was processed and classified into 14 land-cover classes that cover a total area of 2250km^2. The scale of features in tandem with low resolution imagery from the Landsat-TM used a maximum likelihood classifier for coarse pixel classification.

Doren et al. [1999] also used Landsat-TM imagery to map the Florida Everglades in the USA that comprises of freshwater marshes and coastal mangrove estuaries that support a rich diversity of plants and animals. The imagery was classified into broad land cover classes. However, the mapping results have salt-and-pepper appearance which presents a confusing map compared to the manual mapping efforts described in Doren et al. [1999].

Ramsey III and Laine [1997] used Landsat-TM imagery with 30m spatial resolution to map the Louisiana marsh in the USA over a period of 3 years to examine loss of land. The change detection study examines a binary classification problem defined as wetland loss and gain. However, Ramsey III and Laine [1997] states that even with a simple binomial classification scheme, complex marsh systems exhibiting extremely convoluted and heterogeneous landscapes cause problems in the classification process. These scenarios are a product of the low spatial resolution provided by the Landsat-TM that hamper the generation of a binomial land and water mask.

Macleod and Congalton [1998] also used two image captures from the Landsat-TM to monitor eelgrass change detection in Great Bay estuary of New Hampshire. Eelgrass is subtidal plant specie that live on very low shores and present long and thin leaves, however the study site was a large geographical area with 20.7 km^2 which was suitable for the low resolution imagery provided by the Landsat-TM. The orbital cycle of the Landsat-TM was also suitable for the change detection study in order to analyse the net gain or loss of eelgrass extents.

Hossain et al. [2015] used Landsat-5, Landsat-7 and Landsat-8 with a spatial resolution of 30m to photograph aerial extents of multi-species seagrass meadows on the shores of Lawas, Pengkalan and Paka in Malaysia. Respectively, the study areas cover large geographical areas of 185 km^2, 100 km^2 and 20 km^2 that was suitable for the low resolution imagery provided by the satellite sensors. The objective was to analyse the spatial distribution change of subtidal and intertidal seagrass meadows that in turn was suitable given the orbital cycles of the sensor platform. The seagrass meadows vary in size from 120m to km in length and 50m to 120m in width which again was suitable for the spatial resolution provided by the satellite sensors.

Ha et al. [2021] used Landsat-4 TM, Landsat-5 TM, Landsat-7 ETM+ and Landsat-8 OLI to with a spatial resolution of 30m to map the change of intertidal seagrass over a three decades of the Tauranga harbour in New Zealand. The study site is a large intertidal coastal

environment (201km^2) with widely distributed but patchy seagrass cover and the availability of historic ground-truth information. Images were acquired in 1990, 2001, 2011, 2014, and 2019 at low-tide to maximise the visibility of intertidal seagrass. The binary classification models were used to calculate net gain a loss masks over the mentioned sample years.

The mentioned studies cover large geographical areas which are suitable for the imagery collected with Landsat and Sentinel satellites. The meter resolution from these sensor platforms can allow for large land-cover classes to be mapped, or allow for large geographical areas to be surveyed. The orbital cycles of these sensor platforms are also pivotal for change detection studies but struggle with complex and convoluted landscapes as product of provided low spatial resolution [Ramsey III and Laine, 1997]. Therefore, large land cover classes are suitable but mapping objectives with multiple individual species may struggle due to spatial resolution relative to the size of the feature to be mapped [Fuller et al., 1998; Ramsey III and Laine, 1997; Doren et al., 1999].

2.1.2 Commercial satellite imagery

Another avenue for remotely sensed imagery acquisition is to purchase imagery from multispectral or hyperspectral sensors on board satellite platforms. Commercial satellites provide higher spatial resolution relative to open-source imagery provided by platforms such as Landsat and Sentinel and are often part of constellation satellites that communicate together as a system that provide finer temporal resolution [Patino and Duque, 2013]. The higher spatial resolution relative to open-source platforms is because these image sources produce a panchromatic band. This capture is a very high-resolution image band with a spatial resolution that ranges between 30cm to 1m. Panchromatic images exhibit high spatial resolution which is complemented by the corresponding multispectral or hyperspectral capture that exhibits lower spatial resolution but higher spectral resolution. Combining the panchromatic image with the multispectral images is known as pan-sharpening that is achieved through image registration and fusion. A survey of image fusion techniques for remote sensing applications can be found in [Ghassemian, 2016]. However, multispectral

image bands also present higher spatial resolution relative to imagery from open-source platforms such as Landsat and Sentinel.

Current commercial sensor platforms include WorldView-3 and Planet satellite systems. WorldView-3 is a commercial Earth observation satellite owned by DigitalGlobe. It was launched on 13 August, 2014 with a repeat cycle of three days, a spectral range of 0.45-0.92 μm and a spatial resolution of 30cm per pixel for panchromatic channels and 120cm per pixel for multispectral channels. The Planet Labs satellite system is also a commercial Earth observation satellite that consists of a constellation of satellite platforms. The company currently operates a global constellation of over 200 active satellites [Marshall and Boshuizen, 2013]. Therefore, the temporal resolution of this particular satellite system is high relative to open-source platforms as the constellation of satellites can be coordinated and tailored for application needs and objectives.

Again, imagery captured from commercial satellite platforms have been extensively used for a variety of coastal monitoring activities. The following studies show the use of commercial satellite platforms for various coastal mapping applications and change detection.

Valderrama-Landeros et al. [2018] used multiple satellite platforms to map mangrove forests in Teacapán-Agua Brava-Las Haciendas, Pacific coast Mexico. The study site is large geographical area with an estimated mangrove forest cover of 80000 ha which is suitable for the low resolution imagery provided by Sentinel-2 and Landsat-8. However, the results stated in Valderrama-Landeros et al. [2018] show that images captured with the WorldView-2 sensor platform had the best objective performance and also exhibited clear inter-class boundary separation due to the higher spatial resolution relative to Landsat and Sentinel platforms.

Zhu et al. [2015] used imagery photographed from WorldView-2 to map mangrove forests at Lingding Bay, Guangdong Province, China. Again, the study site is a large geographical area with 700 ha with a dynamic landscape characterised by uneven-aged trees and high spatial variability. The target species were *S. Apetala* and *K. Candel* which are both tree

species that flourish at different tidal zones. The spatial resolution was 2m which similarly to the study conducted by Valderrama-Landeros et al. [2018] provided precise inter-class boundary definition due to the high resolution imagery and the high spectral resolution provided the means to use several vegetation indices for accurate classification. The high spatial resolution allowed for object-based image segmentation methods to be used instead of pixel classification methods as shown in Fuller et al. [1998].

Wang et al. [2004b] used Quickbird and IKONOS satellite platforms to map the mangrove forests at Punta Galeta on the Caribbean coast of Panama. The study site is a large geographical area with three tree species that comprise the canopy of the mangrove forest. These species are *Avicennia germinans*, *Laguncularia racemosa* and *Rhizophora mangle*. The average crown cover of these tree species are respectively 164 to 231 m^2, 90 to 141 m^2 and 127 to 241m^2 which is suitable for satellite platforms used to photograph aerial imagery. The spatial resolution of the Quickbird was 1m for panchromatic channels and 4m for multispectral channels whereas the IKONOS had 0.7m for panchromatic and 2.8m for multispectral channels. The mapping results showed that images captured with the IKONOS satellite platform were objectively better by a fine margin. Wang et al. [2004b] recommends the use of commercial satellite platforms that can achieve less than 10m resolution for accurate delineation of canopy crown cover for mangrove forests if the study site is a large geographical area.

Collin et al. [2018a] used imagery captured from the WorldView-3 satellite platform to map subtidal saltmarsh in Emerald Coast, Brittany, France. The study site has an area of 2.11km^2 with most the cover populated with saltmarsh ecosystems, however small dendritic tidal channels (approx. 2m wide) are present and bordering the channels are areas of saltmarsh. The study used a combination of elevation data, very-high resolution imagery captured from a rotor-based drone and hyperspectral image bands from the WorldView-3 satellite platform. In this scenario, spatial resolution was not the key factor for accu-

rate mapping but elevation combined hyperspectral bands yielded the best objective results.

Campbell and Wang [2019] also used imagery from WorldView-2 and WolrdView-3 platforms to map intertidal saltmarsh ecosystems in Fire island, New York, USA. The study site is a large geographical area with 7000 ha and images had 1m spatial resolution that was optimal to map small dendritic tidal channels often present in saltmarsh ecosystems. Similarly to Zhu et al. [2015], the mapping results were achieved with object-based image segmentation methods to be used instead of pixel classification because of a higher spatial resolution relative to open-source satellite platforms.

Comparing the mentioned studies with the work conducted using open-source satellite platforms shows that commercial satellite platforms are also suitable for surveying large geographical areas. The increased spatial resolution from commercial satellite platforms relative to open-source platforms has the potential of providing better inter-class boundary separation, while also maintaining a large field of view to survey large geographical areas. The latter is because current commercial satellite platforms orbit at the same altitude as Landsat-7 and Sentinel-2. Studies such as Valderrama-Landeros et al. [2018] also show that increasing spatial resolution can resolve ambiguity in mapping results.

2.1.3 Uncrewed aircraft system imagery

A recent prospect for very-high resolution remote sensing addresses operational issues related to commercial satellite missions with the use of Uncrewed Aircraft System (UAS). These platforms for data capture are lightweight, low-cost and are operated from the ground. UASs offer ecologists a cost-effective solution with suitable temporal and spatial resolutions for monitoring environmental phenomena with a smaller field of view related to satellite platforms appropriate to the scales of regional to local objectives [Anderson and Gaston, 2013]. These platforms can be classified into the following categories [Watts et al., 2012]: large to medium UASs, and small to nano UASs.

Large to medium UASs are adapted from military-grade platforms and incur high operational costs relative to smaller platforms. This is due to complex ground operations that often require a runway for take-off and landing, and expert ground support staff for every mission. Even with these constraints, this class of UASs have been used in low-altitude missions for Earth science investigations such as: tropospheric chemistry and Arctic ice reconnaissance [Fladeland et al., 2011], and broader applications including ecological surveys [Fladeland et al., 2008].

Small to nano UASs are suitable aircraft platforms for most ecological surveys, including coastal monitoring [Duffy et al., 2018; Collin et al., 2018b; Rossiter et al., 2020]. The successful use of these platforms has been facilitated by miniaturization and cost reductions among inertial sensors, global positioning system (GPS) devices, and embedded computers [Berni et al., 2009]. Presently, there are now numerous miniaturized sensors suited to UAS deployment.

Two main types of small UASs exist: fixed-wing and rotor-based systems. Fixed-wing systems can travel at faster speeds and are larger than rotor-based systems. During flight, navigation is available through GPS-based autopilot guidance tools that navigate the aircraft along a predetermined flight path [Hardin and Jensen, 2011]. Rotor-based systems differ in capability and are able to hover over fixed targets that is suitable for vertical profiling experiments and spatial surveys [Anderson and Gaston, 2013]. The ability for stable capture of imagery allows for overlapping Very-high resolution (VHR) imagery and Structure from Motion (SfM) techniques to create high-resolution orthomosaics [Duffy et al., 2018; Turner et al., 2012] (commonly less than 0.1m pixel resolution). Rotor-based drones are widely used in fields such as hydrology [DeBell et al., 2015], forestry science [Inoue et al., 2014], polar studies [Ryan et al., 2015] and wildlife monitoring [Chabot et al., 2015; Hodgson et al., 2013]. Furthermore, rotor-based drones have also been used for coastal monitoring activities.

Gonçalves and Henriques [2015] used a lightweight Swinglet UAV to monitor the topographic dune change in Aguda on the bank of river Douro, Portugal. The study site was relatively smaller to mentioned studies in previous sub-sections with 2.1km^2 and the spatial resolution of the stitched orthomosaic from overlapping imagery was 4.5cm which allowed for sharp and accurate delineation of dune topography. However, the key-factor for the study described in Gonçalves and Henriques [2015] was the revisit period because of ease with deployment and lightweight nature of the UAS used to photograph the site. Rotor-based drones are not constrained by operational requirements such as orbital cycles of satellite platforms.

Ventura et al. [2016] used a quadcopter with a ArduPilot guidance system to map four species of *Diplodus spp.* which are seagrass species that provide nursing grounds for fisheries. The study site was Giglio Island and again was relatively small to previous studies with a coastline extent of 2.6km. The four species of *Diplodus spp.* were *D. puntazzo*, *D. sargus*, *D. annularis* and *D. vulgaris* all of which represent dynamic ecosystems that present variable cover. The spatial resolution was 1cm which allowed for accurate delineation of seagrass cover in relation to background sediment that often presented a diameter of 1-3m. Ventura et al. [2016] also provided a comparison between maximum-likelihood pixel classification and object-based segmentation methods, with the latter providing stable and better objective scores.

Duffy et al. [2018] used a multi-rotor drone to map seagrass species of *Zostera noltii* in two study sites at Pembrokeshire in Wales: Angle Bay and Garron Pill with both sites covering an area of 2km^2. *Z. noltii* requires specific environmental conditions to successfully grow and survive and therefore can present a dynamic and variable cover from an aerial point of view. The spatial resolution of generated orthomosaics was 4mm which allowed for very fine delineation of dendritic channels of water that could impact *Z. noltii* growth. Duffy et al. [2018] also shows the impacts of increasing spatial resolution for mapping species such as lugworms and cockels with an apparent distortion of features that represent each specie as the spatial resolution is lower.

Rossiter et al. [2020] used a DJI Matrice rotor drone to map intertidal macroalgae at Kilkieran Bay, Ireland. The species were *Ascophyllum nodosum*, Fucus spp., *Himanthalia elongata*, *Laminaria digitata* and *Pelvetia canaliculata* that individually represent small species of macro-algae that can grow up to 2m. Rossiter et al. [2020] states that for fine-scale intertidal macro-algae mapping is limited by the coarse spatial resolution and restricted operational flexibility of satellite platforms, whereas the sub-decimeter resolution allows for accurate results. Furthermore, high-resolution imagery can be used to collect ground-truth data, or labels, only possible because of the low flight altitude and consequent high spatial resolution afforded by UAVs.

2.1.4 Critical analysis

From a chronological view, the open-source satellite platforms available with Landsat and Sentinel have shown to be useful for mapping objectives where the study site is a large geographical area in tandem with target classes that represent agglomerated landcover semantics that span from 20m up to 50m [Fuller et al., 1998; Chopra et al., 2001], e.g. distinguishing wetland, built-up land, among others. The latter is because of the altitude to which Landsat and Sentinel satellites operate as well as constraints to the sensor used to photograph imagery. The spatial resolution for sensors on board open-source satellite platforms is approximately 30m therefore coastal features that represent a physical size inferior to the spatial resolution of the sensor would appear blurred and cause cascading issues for the mapping objective, as shown in Ramsey III and Laine [1997].

The commercial satellite platforms with WorldView and Planet Lab constellation systems have also proven to be useful for mapping objectives where the study site is a large geographical area. However, unlike the open-source satellites, commercial platforms produce the same field of view but have higher spatial and spectral resolution. This has allowed coastal monitoring studies for large areas but also fine scaled mapping of individual species, as opposed to agglomerated target classes. The spatial resolution for sensors on board commercial satellite platforms is approximately 1m which allows finer scale mapping for coastal features such

as corral reefs, saltmarsh ecosystems and mangrove forests [Valderrama-Landeros et al., 2018; Collin et al., 2018a]. Commercial satellites also operate as a constellation offering higher temporal resolution than open-source satellites that is suitable for change detection studies.

The switch to fixed-wing and rotor-based UASs for coastal monitoring further pushed the capabilities for fine scale mapping of individual target species. This is because of the operational altitude of these sensor platforms that in turn capture very-high imagery with a limited field of view. SfM allows for production of VHR orthomosaics that span past the field of view of a single photograph, however data and memory constraints can become an issue for large geographical areas surveyed at sub-decimeter spatial resolution. The spatial resolution for sensors on board UASs is often sub-decimeter that allows for very small coastal features such as lugworms and intertidal macro-algae to be identified [Duffy et al., 2018; Rossiter et al., 2020]. Lastly, the production of very-high resolution orthomosaics allow user's to collect ground-truth data, or labels, directly from processed imagery instead of laborious in-situ that is only possible because of the low flight altitude and consequent high spatial resolution afforded by UASs [Hobley et al., 2023].

Given the properties of different sensor platforms, the use of small rotor-based UASs provides the appropriate spatial and temporal resolutions necessary to capture the complex and highly varying nature of coastal environments that reside in the intertidal coastal zone. In this thesis, both study sites described in Sections 3.2 and 4.2.1 represent local mapping objectives with complex and varying target class domains. For this reason, the thesis focuses on imagery derived from UAS platforms.

2.2 Thematic mapping

As mentioned in Section 1.2, the main goal of coastal monitoring with remote sensing is to produce a thematic map such that the spatial distribution of multiple biophysical properties and controlling processes that govern a particular coastal site can be analysed over time

[Phinn et al., 1999]. In image processing this is also known as semantic segmentation, the process of assigning pixels to a semantic class such that clusters of classified pixels delineate objects of interest within imagery [Thoma, 2016]. The following sections will review established methods in coastal remote sensing for semantic segmentation.

2.2.1 Pixel-based mapping

A wide range of classification methods have been developed to derive semantic information from remotely sensed images [Qian et al., 2007]. The Maximum Likelihood Classification (MLC) is an approach to remote sensing mapping [Foody et al., 1992]. The latter method leverages multispectral radiometric pixel properties to assign a class according to the spectral similarities between a test pixel and a reference train dataset [Jensen, 1986; Gong et al., 1992; Casals-Carrasco et al., 2000; Vatsavai et al., 2011]. MLC assumes that the data for each ecological class to be mapped are normally distributed in each image band of data. Selected training samples, either through visual identification from post-processed imagery or in-situ surveying, are used to build normal distribution models and classification is carried in maximum likelihood fashion [Gao, 1999; Richards and Richards, 1999]. Formally, each pixel is represented as an n-dimensional feature vector and compared to a prototype feature vector constructed from prior spectral information. Spectral features include mean grey levels for individual bands, vegetation indices, e.g, normalised difference vegetation index (NDVI), principal components and band ratios [Wang et al., 2004a; Green et al., 1998; Zerrouki and Bouchaffra, 2014; Held et al., 2003]. The main drawback of MLC is the failure to incorporate contextual texture and spatial features in the per-pixel classification process [Zhou and Robson, 2001].

An alternative approach to MLC is Spectral Angle Mapper (SAM). Again, this approach considers pixels as n-dimensional feature vectors, whereby the number of dimensions is equal to the number of image bands. The magnitude of the vector represents pixel brightness, while the angle represents the spectral feature of the pixel [Kruse et al., 1993]. Then, SAM classifies pixels according to the angular distance between two vectors, with small angles

reflecting similar spectral feature vectors. During classification, a minimum spectral angle threshold is defined for class-boundary separation [Demuro and Chisholm, 2003; Kamal and Phinn, 2011].

Another approach is Linear Spectral Unmixing (LSU) derived from Linear Mixture Models (LMM). The LSU assumes that the reflectance of a pixel is a linear combination of the reflectance of all sub-pixel components also known as endmembers [Horwitz et al., 1971], and the linear combination weighted by the respective endmember abundance [Adams et al., 1995]. Thus, each pixel contains information about the proportion and spectral response of each component, and each pixel spectrum of a multispectral image can be modeled as a linear combination of a finite set of components [Sohn and McCoy, 1997]. Formally, given a linear system of m equations (image bands) and n unknowns (endmembers), a system of equations can be formulated:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 ... + a_{1m}x_n + \epsilon_i &= b_1 \\ a_{21}x_1 + a_{22}x_2 ... + a_{2m}x_n + \epsilon_i &= b_2 \\ &\cdot \\ &\cdot \\ &\cdot \\ a_{m1}x_1 + a_{m2}x_2 ... + a_{mn}x_n + \epsilon_i &= b_m \end{aligned} \tag{2.1}$$

The system of equations $Ax = b$ has a unique solution $x = A^{-1}b$. The linear system of equations is generally solved in a least-squares approach [Drake et al., 1999; Gong and Zhang, 1999; Van der Meer and Jia, 2012]. The integrity of the model is dependent on the assumption that photons only have a single interaction with a surface represented by a pixel [Adams, 1993].

The endmember selection is a crucial step in LSU since the residual error is based on the number of selected endmembers. Semi-automated endmember extraction algorithms include the Pixel Purity Index (PPI) [Chang, 2013], that are aimed at isolating and selecting purest

spectral pixels from a dataset. However, this method is prone to errors through hyper-parameter selection [Chang and Plaza, 2006]. Adding to this, a standard assumption in spectral mixture is that a single endmember signature can be an exact and an invariable representation of its reciprocal component. However, in practice, image components are capable of experiencing multiple degrees of intra-class variability, and subsequent inter-class separability [Wang and Jia, 2009]. In an ecological context, this could be due to plant phenology [Song, 2005] and biochemical properties [Smith et al., 1994; Serrano et al., 2002]. Numerous LSU techniques have been developed to account for spectral variance within endmembers. Spectral averaging has traditionally been a popular technique [Small, 2012], however this approach is based on the assumption that spectral variations of an endmember are normally distributed [Somers et al., 2011]. An alternative method known as Multiple Endmember Spectral Mixture Analysis (MESMA) does not assume the underlying distribution of candidate endmember signatures [Roberts et al., 1998].

2.2.2 Object-based mapping

Images captured by early remote sensing sensors such as Landsat-MSS and TM, SPOT and AVHRR had pixels large enough to cover the same ground feature, often requiring sub-pixel or per-pixel feature mapping. The launch of commercial satellites and UASs with miniaturized multispectral sensors significantly increased spatial resolution of remotely sensed imagery which in turn nullified pixel-based methods used for moderate and low-resolution [Blaschke, 2010; Hossain et al., 2015]. OBIA is an alternative to a pixel-based method where the basic unit is an image-object instead of individual pixels [Castilla and Hay, 2008]. By grouping a number of pixels into shapes with a meaningful representation of objects, the aim of OBIA is to address more complex classes that are defined by spatial and hierarchical relationships during the classification process [Lang, 2008].

OBIA is defined in three main phases: an initial image segmentation to delineate image-objects, feature extraction in image-objects and classification using machine learning models. The most important step is image segmentation as the accuracy of the following feature

extraction and classification mainly depends on the quality of image segmentation [Blaschke et al., 2008; Cheng et al., 2001; Mountrakis et al., 2011; Su and Zhang, 2017]. Image segmentation is defined as a method of dividing an image into homogeneous regions [Pal and Pal, 1993]. In OBIA and remote sensing, these regions could represent land covers such as buildings, trees, water bodies and grasslands [Costa et al., 2018; Heumann, 2011a].

Many methods for high-resolution remote sensing image segmentation exist but many algorithms are not applicable in OBIA [Zhang, 2006; Davis and Wang, 2003]. Literature for OBIA image segmentation can be categorised into the following: edge-based, region-based and hybrid methods.

Edge-based methods first identify edges and then close each boundary using a countour algorithm [Zhou et al., 1989; Cao et al., 2016]. The assumption is that between edges the pixel properties change abruptly, and therefore edges are regarded as boundaries between objects [Shih and Cheng, 2004; Martin et al., 2004]. Edge detection is split into three steps [Jain et al., 1995]: filtering, enhancement, and detection. After edge detection, the next step is to connect edges to form closed boundaries. Multiple edge-linking and Hough transform are suitable linkage algorithms, while also attempting to exclude noisy edges from the filtering process [Lu and Chen, 2008; Ballard, 1981].

Region based segmentation methods are the counterpart to edge based methods. Region based methods start from inside an object and then expand outward until meeting the object boundaries [Zhang, 2006]. Theoretically, edge-based and region-based are different representations of the same object. However, region-based approaches may generate radically different results than edge-based approaches [Kavzoglu and Tonbul, 2017]. Region-based methods assume that neighboring pixels within the same region have similar values and have two basic operations: merging and splitting [Tremeau and Borel, 1997; Fan et al., 2001]. The basic approach is to either obtain an initial (over or under) segmentation of the image and merge or split those adjacent segments according to a criterion of similarity or dissimilarity and repeat until no segments should be merged or split [Bins et al., 1996].

One method to region-based segmentation is the watershed transformation based on Mathematical morphology [Vincent and Soille, 1991; Hossain and Chen, 2019]. Watershed simulates a flooding approach and transforms the image into a gradient that identifies objects with a topographical surface [Mezaris et al., 2004; Muñoz et al., 2003]. Performance of watershed segmentation depends on the method used to compute gradients. Typical gradient operators produce an over-segmented result in watershed segmentation due to noise or texture patterns in remotely sensed imagery [Zuva et al., 2011]. Another common region-based segmentation for OBIA is Multi-resolution segmentation (MRS) [Blaschke, 2010; Nussbaum and Menz, 2008]. The segmentation starts with individual pixels and clusters of pixels to image-objects using one or more criteria of homogeneity. The subsequent clustering of two adjacent image-objects or image-objects that are a subset is based on the criterion that evaluates the change in homogeneity during fusion of image-objects. If this change exceeds a certain threshold value, then the fusion is not performed. In contrast, if the change in image-objects is below the threshold, then both candidates are clustered to form a larger region. The segmentation procedure stops when no further fusions are possible without exceeding the threshold value. Multi-resolution segmentation has been extensively implemented in remote sensing research and OBIA for inter-disciplinary applications [Rossiter et al., 2020; Husson et al., 2016; Gao et al., 2017; Mallinis et al., 2008; Li et al., 2016; Johnson, 2013; Schmidt et al., 2004]. Figure 2.1 shows the output of multi-resolution segmentation.

Given the basic unit for OBIA is a segmented partition of an image because of an image segmentation algorithm, each image-object provides the basis to extract several features for machine learning models to learn the problem domain [Bengoufa et al., 2021; Clark et al., 2022]. A common approach is to calculate vegetation and non-vegetation indices that are band ratios that leverage reflected signals captured in the red edge and near infrared spectral range. In addition, these features are combined with expert site knowledge, either through in-situ surveying or photo-interpretation, in order to associate features with a semantic class during the classification process [Andrés et al., 2017; Belgiu et al., 2014; Gu et al., 2017]. Texture features can also be extracted with a popular and classic method in

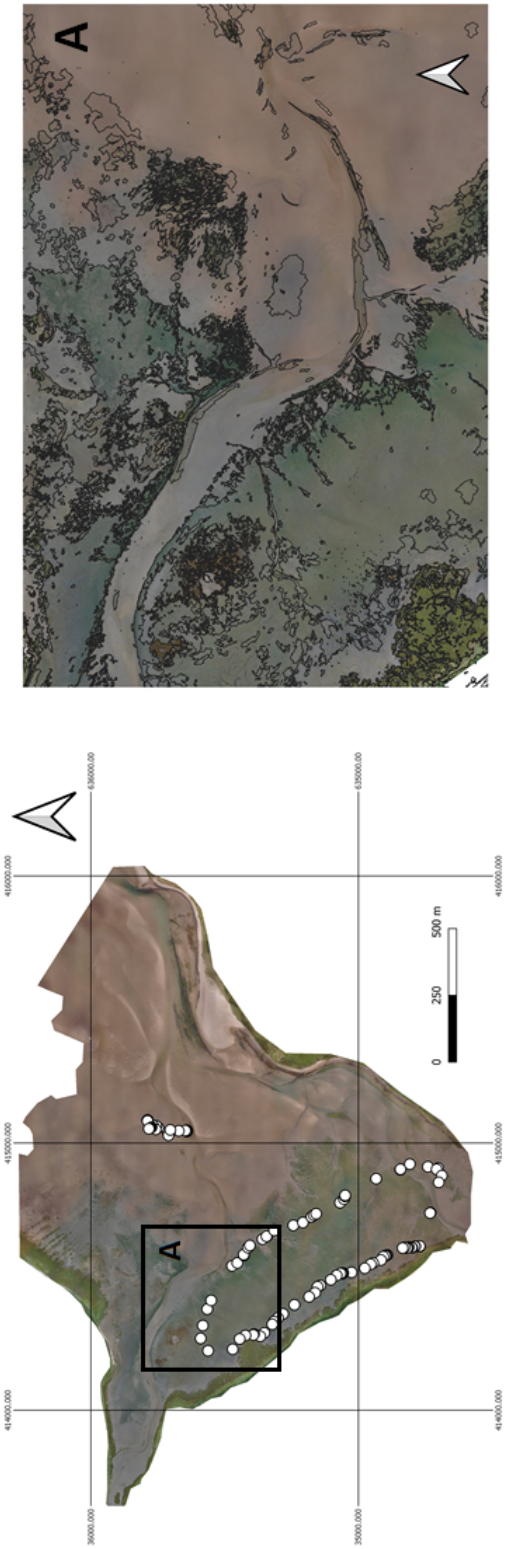

Figure 2.1: Output of multi-resolution segmentation over a high-resolution orthomosaic of Budle Bay. The segmentation algorithm begins clustering image pixels together to form image-objects. The pixel clustering method maximises homogeneity in image-objects and heterogeneity between image-objects by leveraging statistical moments in each candidate cluster.

CV - the Grey Level Co-occurance Matrix (GLCM). The latter is a tabulation of different combinations of pixel values in a greyscale image with respect to their abundance [Haralick et al., 1973]. From a GLCM matrix different features can be calculated with homogeneity proving to be suitable for remote sensing applications [Zhang et al., 2010]. Principal Component Analysis (PCA) can also be used in an OBIA framework to reduce the set of extracted features by transforming the data to a new coordinate system such that the underlying variance is captured in the first couple of principal components [Li et al., 2021b; Kavzoglu and Tonbul, 2017].

2.2.3 Critical analysis

Section 2.1.4 provided an analysis of different sensor platforms with respect to spatial resolution. The analysis correlated the size, or scale, of features to be mapped with the associated sensor platform used to survey the study site. The analysis showed a chronological pattern of increasing spatial resolution due to sensor improvements, or lower flight altitude of the sensor platform, used to photograph a study site. In turn, this allowed for smaller features to be detected which in turn increased the complexity of mapping objectives.

The choice of classifier for image data that are derived from remote sensing platforms is in direct relation to the physical size of each pixel that followed the aforementioned chronological pattern of increasing spatial resolution from sensor platforms. Pixel-based methods were originally used for satellite imagery as the information conveyed in each pixel in relation to the features to be mapped were appropriate. For instance, mapping land cover classes that represent agglomerated landcover semantics that span from 20m up to 50m [Fuller et al., 1998; Chopra et al., 2001]. However, with increasing spatial resolution, the volume of data associated to large geographical study sites in tandem with 1m spatial resolution nullified the use of pixel-based method due to computational costs and lack of accurate classification results. The paradigm shift to object-based methods are largely due to higher resolution imagery and failure to integrate contextual spatial information during the classification process for pixel-based methods.

Multiple studies using UAS and satellite imagery with high resolution (below 5m range) and very-high resolution (below 1dm range) imagery of coastal environments have shown that the use of object-based methods outperform pixel-based methods for supervised semantic segmentation [Hantson et al., 2012; Kamal and Phinn, 2011; Zerrouki and Bouchaffra, 2014; De Giglio et al., 2019]. The common factor of discussion among these studies was the use of a prior image segmentation algorithm in order to cluster homogeneous image-pixels that in turn allowed contextual spatial information to be incorporated in the classification process. Furthermore, the clustering of pixels to image-objects also allowed separability between derived features, e.g., statistical moments and image-band ratios, from each image-object and target classes.

Object-based methods have been shown to perform well in supervised semantic segmentation applications of intertidal and subtidal species in coastal environments using conventional machine learning models, such as Random Forests and SVMs. [Butler et al., 2020; Husson et al., 2016; Purkis et al., 2019; Rasuly et al., 2010; Schmidt et al., 2004; Rossiter et al., 2020; Duffy et al., 2018; Fakiris et al., 2019; Innangi et al., 2019; Janowski et al., 2020]. However, the performance of machine learning models is dependant on the choice of hyper-parameters used during the feature extraction which in turn requires domain expertise [LeCun et al., 2015a]. For instance, with object-based methods the resulting image segmentation/clustering of pixels directly impacts the segmentation accuracy as over-segmentation, or under-segmentation, affect the distribution of features used for conventional machine learning models, as well as the underlying spatial structure of predicted image-objects during inference. This process of feature extraction and subsequent modelling with sophisticated machine learning models can also be referred to as shallow learning or classical supervised machine learning.

Deep learning deviates from the latter method by combining the feature extraction and model tuning processes into a single joint optimisation. The use of convolutional neural networks and fully convolutional neural networks for computer vision applications, such as

classification, object detection and semantic segmentation, have pushed the boundaries of objective results and state-of-the-art which provides a tantalising opportunity to leverage these highly parameterised models for remotely sensed optical imagery. Indeed, for coastal environments, deep learning has already found successful applications whilst outperforming well-established methods in remote sensing literature, e.g, OBIA with SVMs or Random Forests, [Tsiakos and Chalkias, 2023].

Given the advances in computer vision machine learning with the advent of convolutional neural networks, the use of deep learning methods will be reviewed in the following Section. Then, for both study sites described in Sections 3.3.1 and 4.2.1, practical applications of fully convolutional neural networks will be used to map complex target class domains using very-high resolution optical imagery derived from UAS platforms.

2.3 Deep neural networks

The following sub-sections review literature in deep neural networks. The first sub-sections introduce convolutional neural networks and fully convolutional networks for image classification and semantic segmentation, respectively Sections 2.3.1 and 2.3.2. Then, Section 2.3.5 will review methods to improve CNN and FCN generalisation with limited labelled datasets. Lastly, Sections 2.4 and 2.5 review methods for hyperspectral reconstruction and multi-task learning.

2.3.1 Image classification

Deep neural networks have been state-of-the-art models for image classification for nearly a decade. The ImageNet challenge is an image classification benchmark dataset used to evaluate image classifiers [Russakovsky et al., 2015]. The introduction of CNNs and effective use of GPUs allowed these models to reduce the top-5 test set error on the ImageNet challenge from from 28.2% to 17.0% [Krizhevsky et al., 2012]. The top-5 test set error rate shows the percentage of times the image classifier failed to include the correct class among

the top-5 highest probable classes. Previous methods revolved around separating two key components to machine learning: feature extraction and model tuning. Generally, feature extraction methods for image classification relied on carefully engineered scale invariant image features (SIFT [Lowe, 1999]) with a powerful classifier to discern different features to corresponding semantic classes in an image [Sánchez et al., 2013]. In contrast, CNNs replace the separation of machine learning with a joint optimisation of both procedures that is an enabling factor for their success. The feature extraction process consists of repeated convolution and pooling operations that transform the input image into hierarchical abstract representations of data.

The joint optimisation is achieved by adjusting convolutional kernel weights and biases that minimises the error between network outputs and annotated labels [LeCun et al., 2015a]. The choice of error metric is fundamental in order to achieve good CNN performance as the error derivative drives the optimisation of convolutional weights in the network. Since each layer is connected through a shared weight mechanism the initial error derivative is propagated from the output layer to the initial layer using the chain rule of derivatives [Rojas, 1996; LeCun et al., 2015a]. Then, each weight is adjusted using a gradient-descent solver, e.g. SGD [Sutskever et al., 2013], Adam [Kingma and Ba, 2014]

While the introduction of CNNs can be traced all the way to an architecture called LeNet [LeCun et al., 1989, 2015b], the rise of CNNs starts with an architecture commonly referred to as AlexNet [Krizhevsky et al., 2012]. Since then, research has tailored more complex and sophisticated network topologies that further improve image classification accuracy.

The introduction of VGG architectures reduced the top-5 test set error to 8.0% [Simonyan and Zisserman, 2014]. Prior work commonly used convolutional layers with large kernels, whilst the VGG architecture use smaller 3×3 kernels in convolutional layers. These offer the same receptive field while requiring fewer parameters and also improving accuracy. Convolutional layers comprised of 3×3 kernels have become common practice for architectural design and engineering because of their effectiveness and simplicity.

The first very deep networks such as VGG required the training procedure to be done in two stages; the first convolutional blocks were pre-trained before attaching the final layers. The introduction of batch normalization permits deeper networks to be trained in end-to-end fashion by reducing the internal covariance shift that occurs after multiple convolutional layers [Ioffe and Szegedy, 2015]. CNNs are highly parameterised machine learning models comprised of several non-linear modules or layers. Prior to batch normalization, the depth of CNNs would be limited and thus plateau performance of image classifiers.

The introduction of residual architectures further reduced the top-5 test set error rate to 4.49%. Figure 2.2 shows different convolutional blocks that comprise a layer. For residual blocks, instead of stacking multiple convolutional layers in order to fit the underlying non-linear relationship between input image to classification, a skip connection in a residual layer acts as an identity function, thus allowing convolutional layers to learn the residual mapping [He et al., 2016].

ResNet architectures present a low top-5 test set error rate but still present a top-1 test set error rate of 21.2%. The next generation CNN architecture at the start of the 2020s does not explicitly change the overall residual nature for each convolutional block in a ResNet but alters the contents of each residual convolutional block. Again, Figure 2.2 shows the progression of convolutional blocks. The introduction of VGG networks standardised network design with 3×3 kernels, Batch Normalisation (BN) to reduce the covariance shift and a non-linear activation function, such as Rectified Linear Unit (ReLU) [Nair and Hinton, 2010; Dahl et al., 2013]. A recent architecture known as ConvNeXt explores using convolutional blocks with 7×7 kernels, layer normalisation instead of batch normalisation, a multi-layer perceptron approximated with 1×1 discrete convolutions (as per network in network [Lin et al., 2013]) followed by GeLU and a final multi-layer perceptron with a residual skip [Liu et al., 2022b]. This architecture reduced the top-1 test set error to 13.4%.

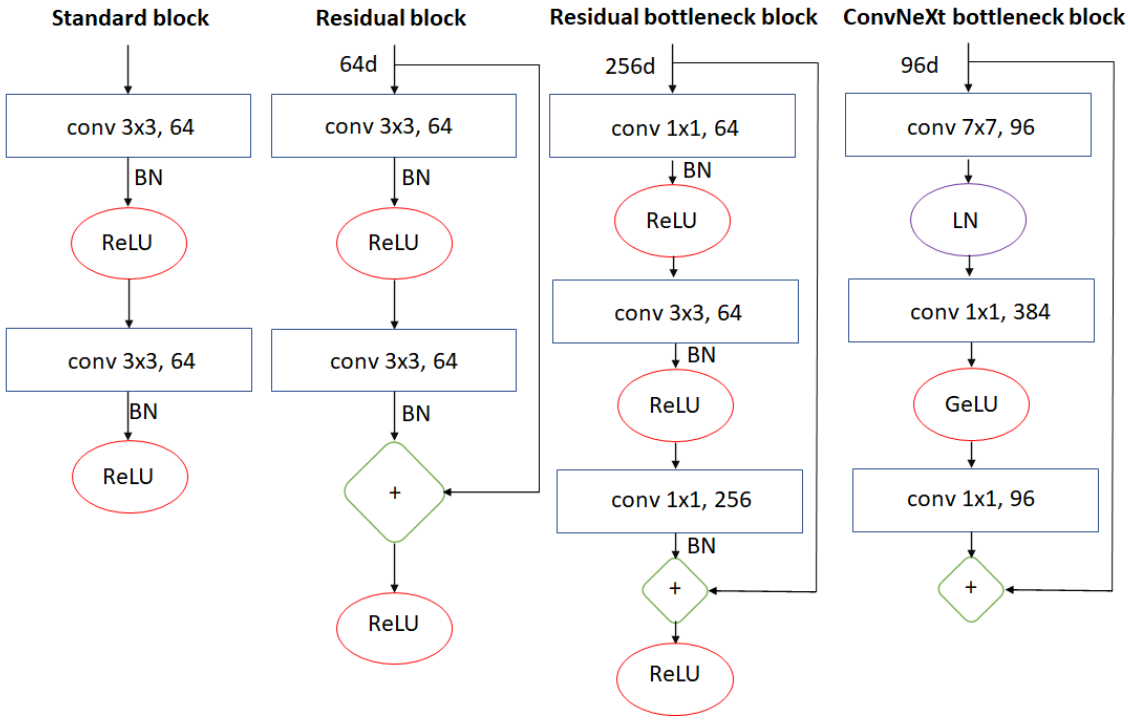

Figure 2.2: The progression of convolutional blocks for CNN feature extraction. The VGG architecture standardised network design with with 3×3 kernels, batch normalisation and a non-linear activation function [Simonyan and Zisserman, 2014]. The addition of residual connections allows training signals to propagate the identity function, and thus learn the residual function [He et al., 2016]. The ConvNeXt explores a different strategy by approximating a MLP with 1×1 discrete convolutions (as per network in network [Lin et al., 2013]).

2.3.2 Semantic segmentation

Semantic segmentation systems can be regarded as pixel classifiers, predicting the class of the object or material type that covers each pixel in an image.

Many semantic segmentation algorithms reported in literature are derived from Fully Convolutional Neural Networks (FCNs) [Long et al., 2015]. This particular deep neural network architecture adapted for segmentation by discarding fully-connected layers that lead to predict image class probabilities. These layers are replaced with new layers that generate pixel-wise class predictions across the image, commonly known as dense predictions. This is possible using of 1×1 convolutional layers, where the number of output channels can

be modified without decimating spatial information. Furthermore, it has been shown that fully connected layers can be approximated with 1×1 convolutional layers [Lin et al., 2013] which in effect allow 1×1 convolutional layers to act as pixel-wise classifiers.

Pooling operations in CNNs reduce the spatial resolution of extracted feature maps to merge local patches to a single pixel and introduce translation invariance [LeCun et al., 2015a]. Fully convolutional neural networks need to restore lower resolution feature maps to the original spatial resolution of the input image. Bilinear interpolation is a common image upsample method used for restoring feature map resolution, and transposed convolutions, also known as deconvolution, can also upsample feature maps while also providing learnable parameters. The upsample method is a trade-off between classifier accuracy and accurate inter-class delineation as transposed convolutions offer learnable parameters but introduce checkerboard artifacts [Gao et al., 2019; Sanjar et al., 2020].

Fully convolutional networks demonstrate the effectiveness of deep neural networks for segmenting the PASCAL VOC 2012 dataset [Everingham and Winn, 2012]. The pioneering implementation of FCNs used dense prediction classifiers to the third and fourth convolutional block and replaced the fully connected layer with a convolutional equivalent to a VGG-16 encoder network [Long et al., 2015; Lin et al., 2013]. This results in class predictions at $1/8^{th}$, $1/16^{th}$ and $1/32^{nd}$ resolution respectively. The final layer upsamples dense predictions by a factor of two using transposed convolution, and combines these with dense predictions from the fourth convolutional block. In turn, the combined feature maps are also upsampled and combined with predictions from the third block. Finally, the combined feature map is upsampled by a factor of eight, resulting in a full-resolution pixel-wise class prediction for the image. This architecture is also known as FCN-8 [Long et al., 2015].

Further work on semantic segmentation explores modifying network topology. Figure 2.3 shows different semantic segmentation architectures which can be broadly described in the following categories: image pyramid, encoder-decoder, context models, spatial pyramid pooling and dilated convolutions.

Image pyramid architectures leverage the same model, typically with shared weights, to multiple resized copies of the input image at various resolutions. Feature responses from low resolution copies encode long-range context, while high resolution copies preserve and encode detail in smaller objects. Some examples transform the input image through a Laplacian pyramid, feed each scale input to a CNN and merge the feature maps from all the scales [Farabet et al., 2012]. Other methods apply multi scale copies sequentially from coarse-to-fine resolution [Eigen and Fergus, 2015] or resize the input image at several scales and then fuse the features from all scales [Lin et al., 2016]. The main drawback to these models are failure to scale well with deeper CNNs due to limited GPU memory. Therefore, most of the multi scale copies are usually applied during the inference stage [Dai et al., 2015].

Encoder-decoder networks consist of two parts: the encoder network where the spatial dimension of feature maps are gradually reduced, allowing low resolution information to be captured in later convolutional blocks. A decoder network that in effect mirrors the encoder network with equivalent decoder layers in opposite order. Encoder-decoder architectures draw data from intermediate layers in the encoder to assist in the decoding process. This adds fine detail as later layers in the encoder tend to represent high level features rather than precise detail [Long et al., 2015]. Encoder networks can be standard network architectures described in Section 2.3.1 but the decoder network can explore various strategies or architectures. Pooling indices from max-pooling layers of the encoder drive equivalent unpooling layers in the decoder in an architecture known as SegNet [Badrinarayanan et al., 2017]. Other strategies use a transposed convolution to learn the upsample process in the decoder network [Long et al., 2015; Noh et al., 2015]. Skip connections from the encoding stages of the network to the corresponding decoding stages were found to be powerful for fine scale semantic segmentation resulting in a network architecture known as U-Net [Ronneberger et al., 2015]. The latter has widespread use in many semantic segmentation applications due to its interchangeability of encoders, learnt upsample process with transposed convolutions and skip connections for accurate pixel prediction.

Context models contain extra modules laid out in cascade manner to encode low resolution detail. An effective method is to incorporate dense conditional random fields to CNNs [Krähenbühl and Koltun, 2011; Chen et al., 2017a]. The use of Conditional Randon Fields (CRFs) for segmentation has a rich previous work body prior to the wide adoption of deep learning [Krähenbühl and Koltun, 2011]. When used for semantic labelling, CRFs perform probabilistic inference and incorporate assumptions such as class agreement between similar or neighbouring pixels [Zheng et al., 2015]. CRFs can be seen to smoothen the classifier predictions within similarly coloured regions, while encouraging semantic segmentation boundaries to align to edges in the RGB image. Therefore, the use of CRFs can refine the segmentation output of the network and encourage segmentation to align with boundaries in the RGB [Chen et al., 2017a].

Spatial pyramids have been an effective multi-scale feature extraction method for many computer vision applications [Grauman and Darrell, 2005; Lazebnik et al., 2006]. CNNs for segmentation can incorporate pyramid pooling layers to extract large scale contextual features to aid in the pixel classification. The contextual features provide the segmentation head with additional information on the surroundings of a particular region of the image, improving the accuracy of the segmentation of objects against common backdrops [Zhao et al., 2017].

Dilated convolutions can also incorporate multi-scale information in CNNs [Chen et al., 2017a,b, 2018]. Dilated convolutions insert zeros between two consecutive filter values along each spatial dimension. A standard discrete convolution is a case where there is no inserted zeros, and dilated convolutions allow the network to filter at various fields of view by changing the number of inserted zeros between consecutive filter values. These convolutions are explored in a spatial pyramid sense in a module called Atrous Spatial Pyramid Pooling (ASPP) [Chen et al., 2017a]. This module is inspired by the success of spatial pyramid pooling that is an effective method to resample features at different scales for accurate pixel classification of regions at various scales [Zhao et al., 2017]. ASPP applies

four parallel dilated convolutions with different dilation rates on extracted feature maps and also applies a global average pooling layer [Chen et al., 2017b]. Dilated convolutions can also be explored in an encoder-decoder architecture [Chen et al., 2018]. These architectures are also known as DeepLabV2 [Chen et al., 2017a], DeepLabV3 [Chen et al., 2017b] and DeepLabV3Plus [Chen et al., 2018].

Critical analysis

Section 2.1.4 showed a clear progression in spatial resolution from imagery derived using satellite or UAS sensor platforms. Consequently, in Section 2.2.3 the shift to object-based mapping methods for coastal remote sensing is mainly because of the increase in spatial resolution for remotely sensed imagery that rendered the capabilities of pixel-based methods since these were not capable of using contextual spatial data in the classification process.

The use of deep learning methods, and in particular, pixel-wise classifiers such as fully convolutional neural networks were not motivated by data and increasing spatial resolution. Instead, the use of fully convolutional neural networks in coastal remote sensing is fueled by the advances in computer vision machine learning, and because of the fact that this branch of algorithms have shown state-of-the-art performance on various benchmark datasets from different computer vision applications, e.g., semantic segmentation, image classification, object detection [Ronneberger et al., 2015; He et al., 2016, 2017]. Thus, fully convolutional neural networks for coastal remote sensing semantic segmentation has found extensive successful use for various coastal study sites at different spatial resolutions.

For instance, Lin et al. [2017] used multi-scale fully convolutional neural networks to delineate shorelines and classify ships using Google Earth satellite imagery, and Li et al. [2018a] used a U-Net to perform sea/land segmentation also with Google Earth imagery.

Liu et al. [2018] compared fully convolutional neural networks with object-based methods using several different conventional classifiers, such as: random forests and support vector

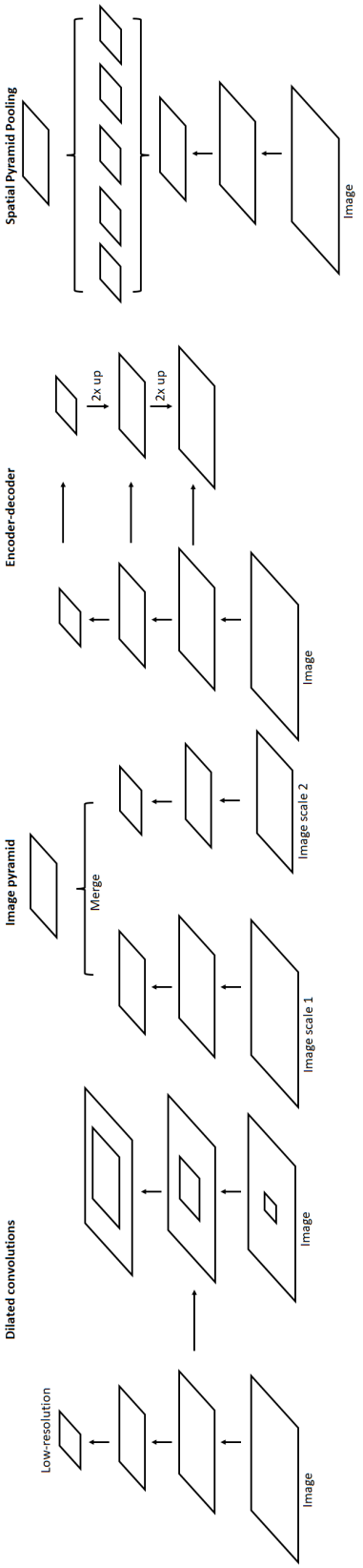

Figure 2.3: Different segmentation architectures commonly found in literature which can be broadly described from (left to right) in the following categories: dilated convolutions, image pyramid, encoder-decoder, and spatial pyramid pooling. Each architecture has a unique method for performing pixel-wise classification but the strongest pixel-classifiers revolve around multi-scale feature extraction (dilated convolutions and spatial pyramid pooling) and encoder-decoder architectures with skip connections for accurate classification.

machines. The comparison was conducted on very-high resolution UAS imagery (approx. 6cm per pixel) and classified large interconnecting marsh of native grass wetlands with fully convolutional neural networks outperforming object-based with conventional classifiers.

La Rosa et al. [2021] used a multi-task learning framework with an encoder-decoder architecture to map tree species dense forests for a coastal site in Santa Catarina state, southern region of Brazil. The study site was captured at very-high resolution (approx. 11cm) and also provided a comparison with object-based methods using random forests and SVMs. The multi-task U-Net was found to outperform OBIA with conventional classifiers.

Hobley et al. [2021a] used U-Nets to map intertidal seagrass at Budle Bay, England using very-high resolution orthomosaics (approx. 3cm). The study provided a comparison of mapping results with OBIA and random forest and also showed that fully convolutional neural networks provide more accurate objective scores.

These use-cases along with extensive review publications [Osco et al., 2021; Yuan et al., 2021] showcase the practicality of deep methods for pixel-wise classification of remotely sensed coastal imagery. Given this, both study sites described in Sections 3.3.1 and 4.2.1, show practical applications of fully convolutional neural networks to map complex target class domains using very-high resolution optical imagery derived from UAS platforms.

2.3.3 Transfer learning

Transfer learning is the process of using an existing model that has already learnt a task from a different domain and using it to transfer knowledge for a related task in order to improve the learning/training procedure [Torrey and Shavlik, 2010]. The most common approach involves adapting a network pre-trained on an ImageNet dataset to a different purpose. The VGG and ResNet network architectures are also commonly used for this purpose. High level convolutional features extracted by AlexNet could be used in place of features extracted using classical computer vision feature extraction methods, e.g. Histograms of

Gradients (HoG) [Dalal and Triggs, 2005], for the purpose of training a classifier, e.g. SVM, for a different classification [Donahue et al., 2014]. Transfer learning can also be applied to different computer vision applications such as semantic segmentation. In this scenario, a pre-trained network on the ImageNet dataset is tuned for image classification but by removing the fully connected layers and replacing the final layer with a new layer suitable for semantic segmentation in the form of pixel-wise classification. The results can be improved through careful fine-tuning where the pre-trained layers are further optimised at a lower learning rate [Long et al., 2015]. Transfer learning has been successfully used in many computer vision tasks, including object detection and image segmentation [He et al., 2016; Long et al., 2015; Yosinski et al., 2014].

2.3.4 Data-augmentation

Data augmentation is a simple method used to increase the variability of samples during training procedures. This process artificially expands the training set by applying linear affine transforms, e.g. rotation, translation, flips and shear, to existing image samples while also preserving ground truth quality. Many state of the art image classifiers incorporate data augmentation during training regimes [Krizhevsky et al., 2012; Szegedy et al., 2015; He et al., 2016].

For small image datasets such as CIFAR-10, random crops are a useful method to augment training samples by padding each 32×32 image with four pixels on each edge to a resolution of 40×40. Then, a random crop of 32×32 resolution is selected which effectively is equivalent to random translation [Krizhevsky et al., 2009]. For larger datasets such as ImageNet more elaborate augmentation schemes have been reported. The Inception architecture for image classification uses an augmentation method known as Inception crop. A random crop is chosen from the image such that it covers between 8% and 100% of the image area with the aspect ratio also varying between 3/4 and 4/3. This crop is extracted from the image and resized to the network input size [Szegedy et al., 2015].

Significant effort in literature has also been devoted to elaborate augmentation schemes in order to push networks to reach state of the art results. The CutOut method augments an image by masking a randomly chosen rectangular region to zero. The rectangles have a fixed size but are randomly positioned [DeVries and Taylor, 2017]. In effect, this is similar to a geometric DropOut, encouraging the network to utilise a wider variety of image features by randomly choosing regions of the image to mask out [Srivastava et al., 2014]. CutOut yielded significant improvement on supervised image classification accuracy on various benchmark datasets. Similarly, another method known as RandErase randomly chooses a rectangle to be replaced with noise also improves performance in classifier accuracy [Zhong et al., 2020].

The MixUp method uses interpolated samples during training. Pairs of input images and target labels are randomly chosen, along with corresponding per-pair blending factors, p. The images, x_a and x_b, and labels, y_a and y_b, are blended using the blending factors: $x_m = (1-p) * x_a + p * x_b$ and $y_m = (1-p) * y_a + p * y_b$, with x_m and y_m used for training [Zhang et al., 2017]. CutMix combines aspects of MixUp and CutOut. Instead of mixing samples using a constant per-pair blending factor, the blending procedure uses a mask and blends target labels with respect to the blending factor. In effect, a rectangular region from one image is cut and pasted over the other. In contrast to CutOut, CutMix uses a rectangle whose size is randomly selected from a normal distribution, such that: $p \sim U(0,1)$. For supervised classification problems, CutMix was found to outperform CutOut and MixUp [Yun et al., 2019].

Recently and inspired on the performance of CutMix, CowMix is also an augmentation procedure where a region of one image is cut and pasted over another. However, the main difference is that CowMix uses a Gaussian filter at a certain scale σ to normally distributed noise. The filtering process produces Friesian cow-like masks that are applied on image pairs before blending them together [French et al., 2020b]. This method also achieved state of the art results in semi-supervised classification - see Section 2.3.5.

Mixing and masking methods are one avenue of data augmentation, another path is to explore the use of rich augmentation schemes during training. AutoAugment and the more recent RandAugment use a repository of 14 image transformations and learned augmentation policies [Cubuk et al., 2019, 2020]. An augmentation policy comprises five sub-policies, each of which combines two image augmentation operations that are applied with a given probability and strength. The probability, strength and choice of operations are optimised to maximise classification performance using reinforcement learning that in turn requires a large amount of computation. RandAugment reduces computational demand with two hyper-parameters: the number of image operations to use to augment each sample and a global strength parameter that determines the strength of every operation used. The hyper-parameters are optimised using grid search [Cubuk et al., 2020].

2.3.5 Semi-supervision

Deep neural networks have set state-of-the-art results in many computer vision problems [Krizhevsky et al., 2012; He et al., 2017; Ronneberger et al., 2015]. However, benchmark datasets such as ImageNet, PASCAL VOC and COCO contain thousands of images with corresponding high-quality labels in order to ensure algorithmic fairness ([Russakovsky et al., 2015; Everingham and Winn, 2012; Lin et al., 2014]). Commercial and widely available cameras provide the means for large quantities of image data to be acquired at very low cost. But, producing ground-truth labels for imagery and desired application is often a laborious bottleneck that is time consuming and expensive, if expert knowledge is required. As mentioned in Section 1.2, coastal remote sensing is also prone to bottlenecks related to labelling issues. In these environments, ground-truth labels can be obtained through in-situ surveys, or through visual identification and delineation of polygons directly from orthomosaics [Congalton, 1991; Leitão et al., 2018; Kattenborn et al., 2019b; Wagner et al., 2019; Lopatin et al., 2019] that often results in a small ratio between area covered via in-situ surveying and the total area covered in imagery [Bowler et al., 2020; Hobley et al., 2021a].

Semi-supervised training methods offer a potential solution to practical applications where only a subset of ground truth labels from training dataset are used, and the remaining unlabelled image samples are incorporated in an unsupervised fashion; while still maintaining classifier performance.

However, semi-supervision is effective under certain assumptions of the underlying problem structure in the training dataset. If these assumptions are not met, this could hinder classification performance [Zhu, 2005]. There are three main assumptions in semi-supervised learning:

- Smoothness assumption - If two data points x_1, x_2 are similar and reside in a high-density region, then the corresponding outputs $\hat{y}_1$, $\hat{y}_2$ should also be similar. Meaning that if two inputs are of the same class and belong to the same cluster that is a high-density region of the input space, then their corresponding outputs should be similar.
- Cluster assumption - If two data points x_1, x_2 are in the same cluster, then the corresponding semantic classes y_1, y_2 should be the same.
- Manifold assumption - high-dimensional data lie on a low-dimensional manifold. In high dimensional spaces, where the volume grows exponentially with the number of dimensions, it is hard to estimate the true data distribution. If the input data lies on a lower-dimensional manifold, then a low dimensional representation can be found using unlabeled data. Then, the labelled task can be solved in a lower-dimensional manifold.

Further to the assumptions on data distribution, semi-supervised approaches assume two learning paradigms: transductive or inductive learning. Inductive semi-supervised learning attempts to generalise a classifier to unobserved instances at test time from both labelled and unlabelled data in the image dataset [Zhu, 2005; Ouali et al., 2020]. Transductive semi-supervised learning attempts to make predictions on a specific set of test instances, given

a combination of both labeled and unlabeled data during the training phase. The model is trained to leverage the relationships and structures within the unlabeled data to improve its predictions for the particular instances of interest. Therefore for most applications of semi-supervised deep learning, inductive learning is more popular because of the practical implications of learning a classifier capable of generalising to unobserved instances at test time instead of a predefined set of instances.

Ouali et al. [2020] describes five distinct methods to semi-supervision, these being: consistency regularisation, pseudo-labels, generative models and graph-based models. Graph-based models will not be included in this review as it refers to the transductive approach Grover and Leskovec [2016] which is out of scope to this thesis.

Consistency regularisation

Consistency regularization describes a class of techniques in that the network is encouraged to give consistent predictions for unlabelled samples under a perturbation, e.g., Gaussian noise and/or standard linear transform augmentations. These methods enforce models to be in line with the cluster assumption that states that decision boundaries must lie in low-density regions. Therefore, if a realistic perturbation is applied to an unlabeled example, then the prediction should not change significantly [Zhu, 2005].

More formally, given a neural network model f with weight parameters θ. A consistency loss favours functions f_θ that return consistent predictions for similar data points, x and $\hat{x}$. Therefore, given an unlabeled data point $x \in D_u$ and the perturbed version $\hat{x}$, the objective is to minimise the distance between the two outputs $d(f_\theta(x), f_\theta(\hat{x}))$. Popular distance metrics for d are mean squared error, Kullback-Leibler divergence and Jensen-Shannon divergence.

The majority of literature in consistency regularisation revolves around complex architecture engineering such that the inputs, x and $\hat{x}$, can be processed to produce outputs, $f_\theta(x)$ and $f_\theta(\hat{x})$, that enforce the cluster assumption. One architecture makes use of ladder networks to

train a CNN in semi-supervision for image classification [Rasmus et al., 2015]. This network consists of two encoders, one for clean and the other for perturbed inputs, and a decoder to remove the perturbation from noisy predictions. The unsupervised training loss is then the mean squared error between the activations of the clean encoder and the reconstructed activations of the noisy encoder. A variant of this network known as **Γ-Model** removes the decoder network and computes the mean squared error between the outputs $f_\theta(x)$ and $f_\theta(\hat{x})$ [Rasmus et al., 2015].

The **Π-Model** is a simplification of the Γ-Model, where the noisy encoder is removed and the same network is used to get the predictions for both clean and perturbed inputs [Laine and Aila, 2016]. This model takes advantage of common regularization techniques, such as data augmentation and dropout, that do not alter predictions. Formally, for a given input x, two augmentation schemes will produce inputs $\hat{x}$ and $\bar{x}$ producing outputs $f_\theta(\hat{x})$ and $f_\theta(\bar{x})$. Then, an unsupervised loss is computed using mean squared error between $f_\theta(\hat{x})$ and $f_\theta(\bar{x})$. An extension to the Π-Model uses **temporal ensembling** to aid with the unsupervised task. Instead of perturbing an input with two separate stochastic augmentations, an exponential moving average (EMA) of predictions is used to compute the mean squared error between the current output and the EMA output [Laine and Aila, 2016].

Inspired with Π-Models and temporal ensembling, the **mean-teacher** approach is another extension where two networks are used: a student network f_θ and a teacher network $f_{\hat{\theta}}$. Both networks process the same input x and produce two sets of predictions $f_\theta(x)$ and $f_{\hat{\theta}}(x)$, however both networks are updated differently. The supervised loss is cross-entropy between $f_\theta(x)$ and known labels y and the unsupervised loss is the mean-square error between $f_\theta(x)$ and $f_{\hat{\theta}}(x)$. The combined loss updates the student network using standard gradient descent, while the teacher network is updated using an EMA of student weights [Tarvainen and Valpola, 2017]. Figure 2.4 lists different semi-supervised network architectures.

Self-training methods

Self-training methods are a class of semi-supervision algorithms that produce labels from unlabeled data using predictions from a model without any supervision. These generated labels are then incorporated with known labeled data to provide more training samples for the model to learn from [Ouali et al., 2020].

More formally, given a model f with parameters θ, the labeled dataset D_l is initially used to train a prediction function f_θ. The model is then used to assign labels to unlabeled data points $x \in D_u$. Given an output $f_\theta(x)$, for an unlabeled data point x in the form of a probability distribution. The new data entry $(x, argmax(f_\theta(x)))$ is added to the labeled set if the probability exceeds a threshold t [Riloff, 1996; Riloff and Wiebe, 2003].

The impact of self-training is similar to that of entropy minimization such that the network learns more confident predictions. However, this could lead to the model amplifying erroneous labels on unlabeled data points. Some approaches to self-training use mean teacher models to generate labels by allowing the teacher model to learn the problem domain using labeled examples, while also generating soft labels on unlabeled data. The student is updated using the labeled set and the generated self-trained labels from the teacher model [Xie et al., 2020]. In addition to image classification, self-training can also be applied to semantic segmentation [Babakhin et al., 2019].

Pseudo-labeling is similar to self-training but the objective of pseudo-labeling is to generate labels that enhance the learning process [Lee et al., 2013; Iscen et al., 2019]. A first attempt at adapting pseudo-labeling for deep learning constrained the usage of the pseudo-labels during the fine-tuning stage after pre-training the network for ImageNet classification [Lee et al., 2013]. Label-propagation with pseudo-labeling alternates between training the network on labeled examples and pseudo-labels and then leveraging the learned representations to build a nearest neighbor graph where label propagation is applied to refine pseudo-labels [Iscen et al., 2019]. However, naive pseudo-labeling overfit to incorrect pseudo-labels due to the confirmation bias. In this context, confirmation bias refers to the tendency of

the model to reinforce its own incorrect predictions, i.e, if the initial pseudo-labels are inaccurate, then models may develop a bias for incorrect labels, leading to poor performance. Some methods use MixUp and set a minimum number of labeled samples per mini-batch in order to reduce confirmation bias [Arazo et al., 2020].

Meta pseudo-labels show that adding pseudo-labels to the training set is a key feature to ensure optimal classifier performance. Previous work show that using a mean teacher model, where the teacher model produces pseudo-labels based on an efficient meta-learning algorithm called Meta Pseudo Labels (MPL) [Pham et al., 2021]. This algorithm encourages the teacher to adjust the target distributions of training examples in a manner that improves the learning of the student model. The teacher is updated through gradient descent by evaluating the student model on a validation set.

Generative Adversarial Networks

Generative Adversarial Networks (GANs) are a recent trend of unsupervised generative models that attempt to match the distribution of a dataset [Goodfellow et al., 2020]. A GAN is composed of two networks: a discriminator and a generator. The discriminator learns to discern training samples that are either part of the true data set or from the generator. The generator learns to produce samples that can fool the discriminator. Thus, the loss gradient is propagated from discriminator to the generator through its ability to judge real from fake samples. Intermediate feature maps in the discriminator can be used to aid the classification process [Radford et al., 2015].

Due to the robust and symbiotic nature of discriminators and generators, GANs can be trained in semi-supervised or unsupervised fashion by allowing samples from the generator to be included in the training process. This way the discriminator operates as a classifier and is trained to maximise the entropy predictions for real samples and maximise entropy for generated ones. The generator learns to generate samples that will maximise discriminator predictions for a given class [Springenberg, 2015].

Other approaches use a discriminator with $N + 1$ classes, with the extra class representing classified fake samples [Salimans et al., 2016]. The discriminator learns to minimise the error between network predictions and labelled samples, and also learns to classify generated samples from the generator. In Salimans et al. [2016], two techniques for improved GANs generalisation are also introduced, these being: mini-batch discrimination and feature matching. **Mini-batch discrimination** allows the discriminator network to operate on multiple samples rather than individual samples. This allows to detect lack of diversity among generated samples which is a good metric for generator performance by checking whether the generator has a constant output. **Feature matching** uses a generator trained to produce samples that induce latent features in the discriminator network, and attempts to match the latent features induced by real samples.

GANs are trained such that the discriminator is used to guide the generator towards producing samples whose distribution closely approximates that of the target dataset. Semi-supervised classification performance can be improved by training a complement generator to approximate a target distribution that assigns high densities for data points with low densities in the true distribution [Dai et al., 2017].

Semi-supervised segmentation

A standard approach for semi-supervised semantic segmentation is to use additional data. For instance, two datasets from different domains can be used to learn the similarity between per-class embeddings from each dataset [Kalluri et al., 2019].

However, there are few approaches for semi-supervised segmentation that use the techniques mentioned thus far. GANs can be used to generate dense predictions and allow the discriminator network to distinguish between predicted segmentation maps and per-pixel labels. For unsupervised samples the segmentation model learns to fool the discriminator network by producing fine-grained accurate segmentation maps [Hung et al., 2018; Mittal et al.,

2019]. The work in Mittal et al. [2019] also reports using a mean teacher model, further improving performance.

Consistency regularisation for segmentation tasks is a challenging problem [French et al., 2019, 2020a]. Standard linear transformations for augmentation procedures drive the consistency metric and thus enforce the cluster assumption. In a segmentation task, low-density regions do not correspond to class boundaries [French et al., 2019], and therefore, semi-supervised segmentation through consistency regularisation has to be achieved without the cluster assumption. Applications of consistency regularisation methods originate from the medical imaging community [Perone and Cohen-Adad, 2018; Li et al., 2018b]. These methods use a MRI volume dataset to detect skin lesions with standard augmentation techniques for perturbed samples. The mean teacher model has also been documented with methods exploring augmenting unlabelled images with noise [Cui et al., 2019].

Given the advances in fully convolutional neural networks and semi-supervised optimisation strategies [Long et al., 2015; Ronneberger et al., 2015; Tang and Shao, 2015], an opportunity surfaces to not only demonstrate a practical application of fully convolutional neural networks using very-high resolution orthomosaics derived from UASs but also to explore the use of semi-supervised optimisation methods given that very-high resolution orthomosaics of coastal environments may cover a substantial spatially-continuous area with respect to the real-world, yet the ratio between the area covered via in-situ surveying and the total area covered in imagery is often relatively small Bowler et al. [2020]; Hobley et al. [2021a]. This said, the use of semi-supervised segmentation methods have been used for remote sensing:

Wang et al. [2022] used consistency-based regularisation and student/teacher networks with pseudo-labels to drive a semi-supervised loss function with fully convolutional neural networks. The method was evaluated on the Inria aerial dataset with a spatial resolution of 0.3m where images cover densely populated cities to alpine towns. The objective was a binary class problem of building and no building. The method was also evaluated on the

15-class ISAID dataset. In both scenarios, the use of a semi-supervised loss function was found to improve objective performance.

Wang et al. [2020b] also used consistency-based regularisation and pseudo-labels to drive a semi-supervised loss function using UNets and DeeplabV3. The method was evaluated on the 15-class ISAID dataset, and again, the use of a semi-supervised loss function was found to improve objective performance.

Sun et al. [2020] used a complex semantic segmentation architecture with channel-weighted multi-scale feature modules, boundary attention modules to alleviate boundary blur in predicted segmented maps. Furthermore, the semi-supervised approach used an auxiliary discriminator network designed to generate high-confidence pseudo-labels for unlabeled images. The method was evaluated on a ten class problem aerial dataset known as ISPRS Vaihingen. The dataset has tiled orthomosaics with a spatial resolution of 9cm.

Patel et al. [2021] evaluated a self-supervised network with and semi-supervised method, respectively known as SimCLR and FixMatch [Chen et al., 2020; Sohn et al., 2020], using a DeeplabV3+ neural network. These methods were evaluated on three datasets: a riverbed segmentation derived from Google Earth of several riverbeds in India with a spatial resolution of 4m, the publicly available Chesapeake Land Cover dataset sourced from the Chesapeake conservancy region in the eastern United States with a spatial resolution of 1m, and Sen1Floods11 that pairs raw satellite imagery with classified permanent and flood water at a 10m spatial resolution. The discussion of results showed that the SimCLR architecture using FixMatch outperformed the baseline DeeplabV3+ (trained with only the supervised loss).

These use-cases show the practicality of semi-supervised methods for semantic segmentation remotely sensed imagery. In general, the use-cases leverage consistency-based regularisation with student/teacher network architectures to produce pseudo-labels with the main discussion of results noting the increase in objective performance metrics when semi-supervision is enabled. Both study sites described in Sections 3.3.1 and 4.2.1, show practical applications

of semi-supervised fully convolutional neural networks to map complex target class domains using very-high resolution optical imagery derived from UAS platforms. The method described in Section 3.3 used a similar method to the aforementioned described use-cases, and the method described in Section 4.5 shows a novel semi-supervised optimisation strategy that leverages multi-task learning and an unsupervised auxiliary image task to promote accurate and correct delineation of target classes.

2.4 Hyperspectral reconstruction

Hyperspectral Imagery (HSI) correspond to captures of scenes or objects where each pixel contains an approximation of a continuous spectral curve to identify the substance of the corresponding objects. In coastal remote sensing, acquiring HSI can aid distinguish different species of vegetation [Liu et al., 2020a, 2022a] and better understand plant phenology [Song, 2005]. However, the devices for acquiring HSIs are complex and incur high costs over common cameras [Descour and Dereniak, 1995; Cao et al., 2011]. Common methods for data acquisition rely on precise scanning to generate hyperspectral data cubes which limits sensor portability unlike standard commercial RGB cameras that require a single snapshot with a centre perspective. Instead, these sensors rely on capturing detailed information about the electromagnetic spectrum across numerous narrow and contiguous wavelength bands with a general scan of the scene. Some literature attempts to solve portability issues by recording a hyperspectral snapshot without scanning at the compromise of degrading spatial resolution [Wagadarikar et al., 2009].

Another avenue is to extract hyperspectral information from a standard RGB image in a process known as hyperspectral reconstruction. The latter is an ill-posed problem as there are many physically plausible hyperspectral metamers that could correspond to the same RGB capture [Cohen and Kappauf, 1982; Morovic and Finlayson, 2006]. Equations 2.2 and

2.3 state the fundamental image formation equation for RGB images.

$$I_c(x, y) = \int_\omega R(x, y, \omega) L(\omega) S_c(\omega)\, d\omega \tag{2.2}$$

$$I_c(x, y) = \int_\omega H(x, y, \omega) S_c(\omega)\, d\omega \tag{2.3}$$

Where, I_c is an RGB image with channels $c = (r, g, b)$, R is the spectral reflectance property at a pixel location (x, y) and wavelength ω, S_c is the spectral sensitivity function of the camera at a channel c and wavelength ω and L is the light spectrum for the scene at a wavelength ω. A hyperspectral image is defined as the multiplication of the reflectance matrix R with the light spectrum L. Therefore, hyperspectral reconstruction attempts to invert the image formation model to find the data-cube H from an RGB image I_c.

To promote solutions, several benchmark datasets have been released to objectively evaluate methods for hyperspectral reconstruction. Open-sourced datasets include: CAVE, ICVL and both NITRE challenges, respectively [Yasuma et al., 2010; Arad and Ben-Shahar, 2016; Arad et al., 2018, 2020] that contain high-quality hyperspectral images of scenes and objects along with RGB images projected into sRGB colour space. From these datasets, two main types of hyperspectral reconstruction methods have been formulated in literature [Zhang et al., 2022]: prior-base and data-driven using deep learning.

Prior-based methods attempt to represent an HSI data cube as a linear combination of basis spectra, known as endmembers, weighted proportionally to the abundance of each endmember in a image pixel. Formally, the HSI can be defined as the multiplication of the basis spectra matrix (E) with its proportional abundance (A) for each pixel i.

$$H_i = EA_i \tag{2.4}$$

Then, the HSI cube can be projected to RGB camera space (X) using the sensor S.

$$X_i = ESA_i \tag{2.5}$$

The latter equation has many non unique solutions which can be constrained with prior knowledge of the hyperspectral dataset.

HSI exhibit sparse encoding and spectral information which can be expressed as a sparse combination of basis spectra [Chakrabarti and Zickler, 2011]. One approach is to create a dictionary representation to store the basis functions of E and abundance coefficients A. Once a dictionary is built from hyperspectral imagery, the sensor response S can project to RGB space and the projected dictionary can then be used to recover the hyperspectral information of an input RGB image. Arad and Ben-Shahar [2016] creates a hyperspectral dictionary by randomly selecting pixels in HSI that in turn correspond to continuous spectral measurements. For selected samples, K-SVD [Aharon et al., 2006] creates a sparse dictionary that can be projected to RGB colour space using the sensor responses. During inference, the Orthogonal Match Pursuit (OMP) algorithm [Pati et al., 1993] maps the input RGB image to an intermediate representation using the projected dictionary and then the intermediate representation is mapped to hyperspectral dictionary, thus recovering spectral information. Another method establishes the RGB to HSI mapping with a local dictionary instead of a global sparse dictionary. Then, it solves the mapping abundance coefficients from neighboring anchor points in a least-squares optimisation [Aeschbacher et al., 2017].

Prior-based methods rely on selected samples and known spectral responses. Deep learning approaches leverage the highly parameterised nature of CNNs and high-quality labels from benchmark datasets to learn accurate RGB to HSI mappings. One of the first approaches to accurate hyperspectral reconstruction was achieved with a CNN architecture known as HSCNN. This deep model restores hyperspectral information from RGB using principles from the spatial super-resolution algorithm VDSR [Kim et al., 2016]. Networks are trained using mean squared error and achieve good reconstruction fidelity but error rates increase as network depth also increases [Xiong et al., 2017]. An extension to HSCNN tackles model depth by introducing residual connections in each convolutional block (HSCNN-

R) or dense blocks with path-widening fusion scheme (HSCNN-D), and achieves spectral upsample through 1×1 convolutional layers [Shi et al., 2018]. Networks are trained with mean relative absolute error on 50×50 image patches.

The current state of the art for hyperspectral reconstruction leverages attention modules for accurate hyperspectral reconstruction. In reconstruction tasks, an adaptive weighted attention network known as SRWAN uses the sensor responses as a prior and integrates various strategies, such as: attention and residual learning [Li et al., 2020]. The attention module is inspired from Squeeze-and-Excitation Hu et al. [2018]. Networks are trained using a combined loss from the sensor response prior and hyperspectral reconstructed images both evaluated using mean relative absolute error.

2.5 Multi-task learning

Generally in deep learning, networks are optimised to learn an objective metric in order to achieve a specific image task, such as image classification [Krizhevsky et al., 2012], semantic segmentation [Long et al., 2015], object detection [He et al., 2017], among other image tasks. While optimising networks for a single specific task achieves state-of-the-art performance on established benchmark challenges, auxiliary image tasks can aid the optimisation of a single image task. By sharing internal representations and allowing information from auxiliary training signals to flow through the network, models can learn different internal representations and prevent overfit on the original image task. This approach is called Multi-task learning (MTL) [Ruder, 2017].

The goal of MTL is to improve model generalisation by leveraging the domain-specific information contained in the training signals of related tasks [Caruana, 1997]. The architecture design of deep neural networks trained to perform MTL follow two paradigms: **hard parameter sharing** and **soft parameter sharing**. Figure 2.5 shows the general network architecture for each MTL paradigm.

The standard approach for network architectures with **hard parameter sharing** is to use the same hidden layers, or convolutional layers in a CNN, and have multiple output layers for each specific image task. Hard parameter sharing reduces the risk of overfit by allowing networks to consider different solutions from auxiliary image tasks. In general, as the number of tasks increase, neural networks will attempt to find different internal representations that capture all of the image tasks to be learnt jointly [Baxter, 1997]. Thus, reducing the likelihood of overfit on each specific image task. Deep relationship networks are a hard parameter that incorporates matrix priors on fully connected layers that allow the shared model to learn the relationship between auxiliary image tasks [Long et al., 2017]. Adaptive feature sharing is another hard parameter sharing method that proposes a bottom-up approach. An initial thin network is dynamically widened during training using a criterion that promotes grouping of similar image tasks [Lu et al., 2017]. Dynamic network widening entails that the topology of the network is adjusted during training to accommodate more complex or diverse tasks or data representations. This widening can involve adding more layers, units, or other architectural elements to the network as needed.

The alternative method for MTL is **soft parameter sharing**. In this scenario, multiple networks are trained, each for a specific image task, and a joint multi-task loss drives the optimisation of all the networks. Networks are then encouraged to learn auxiliary image tasks by allowing networks to optimise based on representations from auxiliary networks. One form of soft parameter sharing is to calculate the Euclidean distance and trace norm between auxiliary networks [Duong et al., 2015; Yang and Hospedales, 2016]. Another method for soft parameter sharing is to use individual networks for each image task. But, the input to each layer is a linear combination of the outputs of the previous layer from every auxiliary network [Misra et al., 2016]. This neural network architecture is also known as Cross-stitch networks.

2.6 Conclusions

The review in chapter 2 covered topics related to sensor platforms, methods for thematic mapping in coastal remote sensing, state-of-the-art deep learning architectures, different optimisation procedures with limited amounts of labelled records and deep learning applications for hyperspectral reconstruction and MTL.

Section 2.1.4 provided a critical review of different sensor platforms in coastal remote sensing. The choice of a particular platform depends on the application and scale of the coastal study site. For large coastal extents with target classes that may span several meters, features of interest can be captured using satellite imagery at repeated sampling intervals and are appropriate for continued monitoring of large geographic areas [Gould, 2000]. For regional to local coastal extents where the scale of features may present in sub-decimeter range, the use of UAS imagery provides suitable temporal and spatial resolutions for monitoring environmental phenomena, such as coastal vegetation for intertidal and open-shore beaches [Anderson and Gaston, 2013]. In particular rotor-based drones allow for stable capture of optical imagery where detailing features for target classes resolve to centimeters [Duffy et al., 2018]. Features and mapping objectives vary from one coastal environment to another but the choice of sensor must consider the scale of features to be mapped.

Section 2.2.3 showed a chronological progression of methods used to map features from remotely sensed coastal imagery. The first set of methods, known as pixel-based methods, whereby the analysis and subsequent pixel-classification do not include contextual spatial detail. With increasing spatial resolution due to advances to sensor platforms, these methods were super-seeded by object-based methods whereby the use of contextual spatial information improved objective performance in various coastal mapping surveys. Sections 2.3.2 and 2.3.5 showed modern deep neural networks for pixel-wise classification, or semantic segmentation that provides an equivalent output to methods described in 2.2.3. Deep learning and fully convolutional neural networks have shown to outperform shallow supervised

machine learning, and therefore an opportunity surfaces to apply this branch of computer vision to very-high resolution optical imagery using rotor-based UASs.

During the research period, Cefas provided two datasets with VHR imagery captured with UAS instruments and miniaturised multispectral sensors that provided the basis to extract fine scale orthomosaics of Budle Bay, Northumberland, England (55.625°N, 1.745°W) and Sizewell, Suffolk, England (55.207°N, 1.602°W) with multispectral resolution using SfM [Turner et al., 2012]. The Environmental Agency (EA) provided in-situ data for Budle Bay, while the in-situ survey for Sizewell was a combined effort of Cefas and the UEA. Given these datasets, the following chapters examine each study site individually. Each dataset had a different mapping objective, and each chapter attempts to map either problem domain using different methods.

The main goal for Budle Bay is to map intertidal seagrass extents due to its contribution to intertidal coastal ecosystem health. The scale of intertidal seagrass, and other macro-algae species, present at the study site warrant the use of fixed wing UASs to allow for stable capture at very-high resolutions (approx. 3cm per pixel resolution). Furthermore, Cefas was able to provide thematic maps of Budle Bay that were generated using object-based methods with supervised machine learning, as described in Section 2.2.3, which provides the opportunity to compare these methods with fully convolutional neural networks and sophisticated optimisation procedures, as shown in Section 2.3.2 and 2.3.5. Section 3.3 shows a practical application of consistency-based regularisation methods for intertidal seagrass mapping and the discussion addresses the challenges and problems associated with mapping these species of intertidal seagrass among species of algae. Section 3.4 explores these problems by conducting an inter-observer experiment to investigate the feasibility of crowdsourcing labels. The discussion was two-fold: the variability among participants in the experiment was analysed with respect to discipline expertise, and then the use of participant annotations to train deep learning models is also analysed to discuss the feasibility of crowdsourcing labels from aerial imagery.

The main goal for the Sizewell study site is to map the strandline and sand-dune communities belonging to SD1, SD2, SD6 and SD7 National Vegetation Classes (NVCs). Again, the scale of shingle vegetation species present at Sizewell necessitate the use of rotor-based drones to allow for stable capture at very-high resolutions (approx. 1cm per pixel resolution). In particular, the detailing features that separate each of species found at Sizewell are often less than 15cm that further stressed the need to use UAS. Furthermore, the in-situ survey for Sizewell also provided the opportunity to evaluate hyperspectral reconstruction methods described in Section 2.4. Therefore, an opportunity also surfaces to evaluate the use of CNNs for hyperspectral reconstruction with a comparison of previous methods. Then, in Section 4.5, the use of an MTL framework to jointly learn hyperspectral reconstruction and semantic segmentation is investigated with the discussion comparing results of supervised and semi-supervised methods, as shown in Section 3.3, and fully convolutional networks trained using MTL.

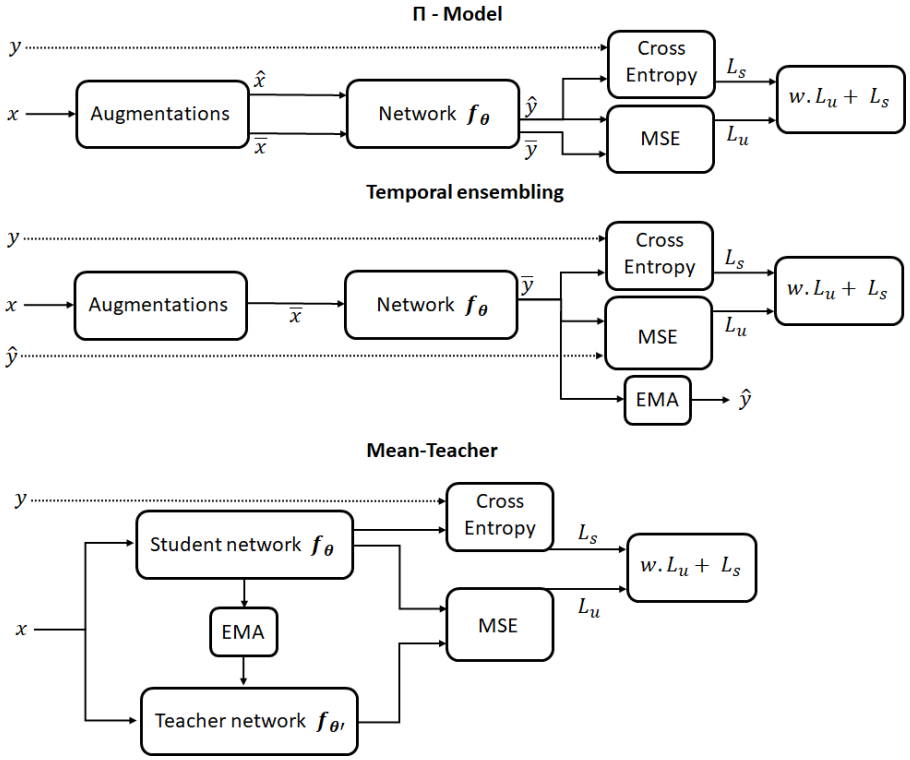

Figure 2.4: Different semi-supervised network architectures used for generating consistency-based loss functions. The Π-Model (top architecture) generates pairs of predictions using a single model and two different stochastic image augmentation processes, generating two different sets of predictions, $\hat{y}$ and $\bar{y}$. The latter drive the unsupervised consistency loss function and the label y and $\hat{y}$ drive the supervised loss. Temporal ensembling (middle architecture) generates a single set of predictions $\bar{y}$ which drives the standard supervised loss and the unsupervised consistency loss through an EMA of predictions from the previous iteration $\hat{y}$. Mean-teacher (bottom architecture) uses two networks to generate pairs of predictions, the student network which drives the supervised loss and a teacher network which is an EMA of the weights in the student network. The unsupervised consistency loss is driven by the predictions generated from the student and teacher network [Tarvainen and Valpola, 2017]

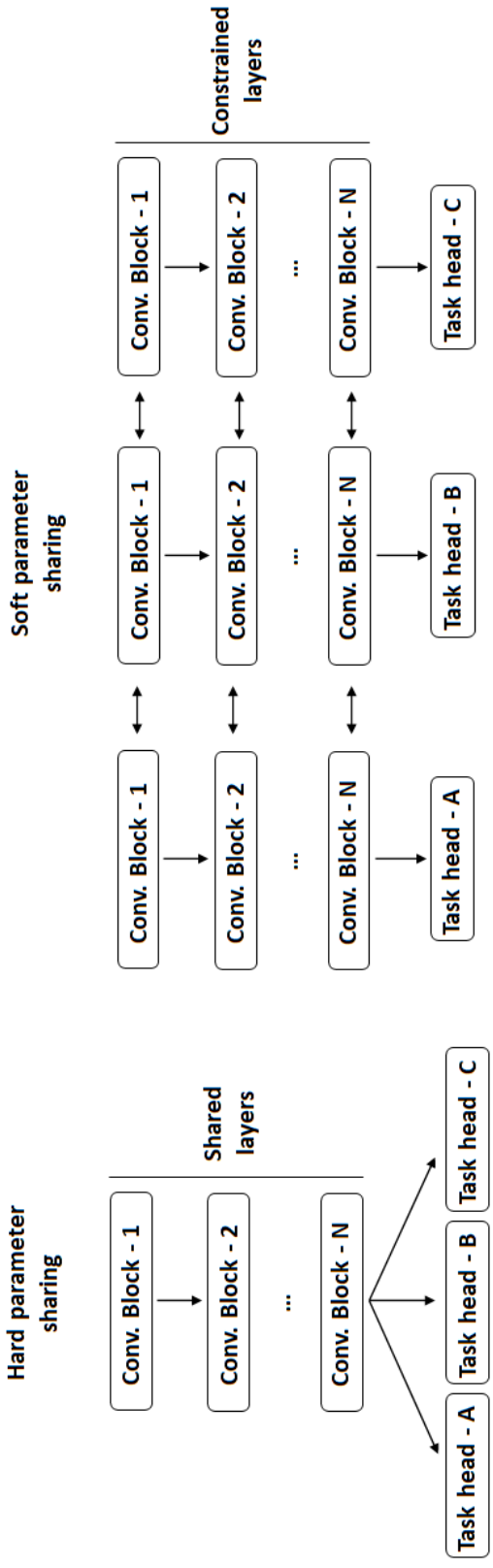

Figure 2.5: Different MTL paradigms and subsequent effects on network design. On the left, the hard parameter sharing uses a single network to output to learn multiple image tasks that are dictated by the amount of output heads. On the right, the soft parameter sharing uses multiple networks for each image task but employs a multi-task loss to jointly optimise each network.

3 Budle Bay - semi-supervised and crowd sourced learning for intertidal seagrass mapping

3.1 Introduction

The research for this chapter was focused on Budle Bay, Northumberland, England (55.625°N, 1.745°W). The coastal site has one tidal inlet, with previous maps also detailing the same inlet [Ladle, 1975; Meyer, 1973; Olive, 1993]. Sinuous and dendritic tidal channels are present within the estuary, and bordering the channels are areas of seagrass and various species of macroalgae. The tidal range varies between 1-4m for the majority of the year and the estuary is fully drained on low spring tides.

The research for this chapter will focus on developing methods to map intertidal seagrass in Budle Bay using limited amounts of labelled data. As mentioned in Section 1, these environments are physically varying through the energy expended with water and sediment movement [Alongi, 2020]. And in particular, intertidal seagrass and algae play an important role to tidal and energy management from currents and waves [Bouma et al., 2005], sediment quality and stability [Koch, 1999; Fonseca et al., 1983].

First, Section 3.2 shows the collected data and imagery for Budle Bay. The target class domain for the mapping exercise is detailed as well as the cameras used to survey the study site.

Then, Section 3.3 shows a semi-supervised approach for semantic segmentation using deep learning models for intertidal seagrass mapping. The discussion includes a comparison with standard mapping techniques for intertidal seagrass, and compares objective scores and visual habitat maps derived from models trained in supervised and semi-supervised settings, with the results obtained using OBIA.

The second method in Section 3.4 continues to find alternatives for mapping intertidal seagrass with limited amounts of labelled data. Instead of devising semi-supervised methods for semantic segmentation, an annotation experiment was conducted to illustrate the feasibility of supplementing training labels derived from the in-situ survey with labels obtained directly from aerial imagery. First, the goal is to examine inter-observer variability subject to annotator expertise, and then deep learning models are trained with crowdsourced annotations.

3.2 Data collection and in-situ survey

Ground and aerial surveys of Budle Bay were conducted in September 2017 by the Centre for Environmental Fisheries and Aquaculture Sciences (Cefas) and the Environmental Agency (EA). The aerial survey performed two flights using a fixed-wing UAS with each flight using one of two attached available sensors: a SONY ILCE-6000 camera with filters for Red, Green and Blue channels and a ground sampling distance of approximately 3 cm (Figure 3.1, bottom right). And a MicaSense RedEdge3 camera with five narrow banded filters for Red (655-680 nms), Green (540-580 nms), Blue (459-490 nms), Red Edge (705-730 nms) and Near Infra-red (800-880 nms) channels and a ground sampling distance of approximately 8 cm (Figure 3.1, top right).

Very high resolution orthomosaics of Budle Bay were created with Agisoft's MetaShape [Agisoft, 2018] and SfM. SfM techniques rely on estimating intrinsic and extrinsic camera parameters from overlapping imagery [Cunliffe et al., 2016]. A combination of appropriate flight planning in terms of altitude and aircraft speed, overlap between successive pho-

tographs, weather conditions, and the camera's field of view were important for producing good quality orthomosaics.

The resulting VHR orthomosaic was further processed with GPS logs of camera positions and ground control markers spread out across the site to ensure that the mosaic was well aligned with real-world coordinates and ecological features present within the coastal site. The multispectral orthomosaic from the MicaSense RedEdge3 sensor had 32,647×26,534 pixels in five image bands, while the SONY ILCE-6000 was 87,730×72,328 pixels in three image bands. For ease of processing, each orthomosaic was split into 6,000×6,000 non-overlapping tiles along with geographic information to be used for further processing. The SONY orthomosaic was split into 140 tiles and the MicaSense RedEdge3 into 24. Figure 3.1 shows very-high resolution orthomosaics of the study site with the SONY ILCE-6000 camera and the MicaSense RedEdge3 multispectral camera and Figure 3.2 shows a close up of each orthomosaics with intertidal vegetation amongst background sediment.

During the in-situ survey, expert ecologists from Cefas and the EA surveyed the Western, Central and Southern parts of Budle Bay estuary. The survey found 13 ecological features of interest that can be grouped into background sediment, algae, seagrass and saltmarsh. Classes defining background sediment were rock, gravel, mud and sand. These measurements of unvegetated sediment were predominately in the presence of water and moisture. However, as parts of the orthomosaic included dry sand, an extra sediment class was added through photo-interpretation from VHR orthomosaics (16 polygons). For the purpose of this work, two heuristics for delineating dry sand polygons were defined: first, the spectral reflectance of sand varies with presence of surface moisture and presents higher reflectance intensity for patches of dry sand [Nolet et al., 2014]. Therefore, polygons were delineated by examining bright unvegetated areas in Figure 3.1. Second, each generated polygon was cross-checked with the topographic Digital Surface Model (DSM) to ensure that patches of dry sand only occur if the surface level was raised.

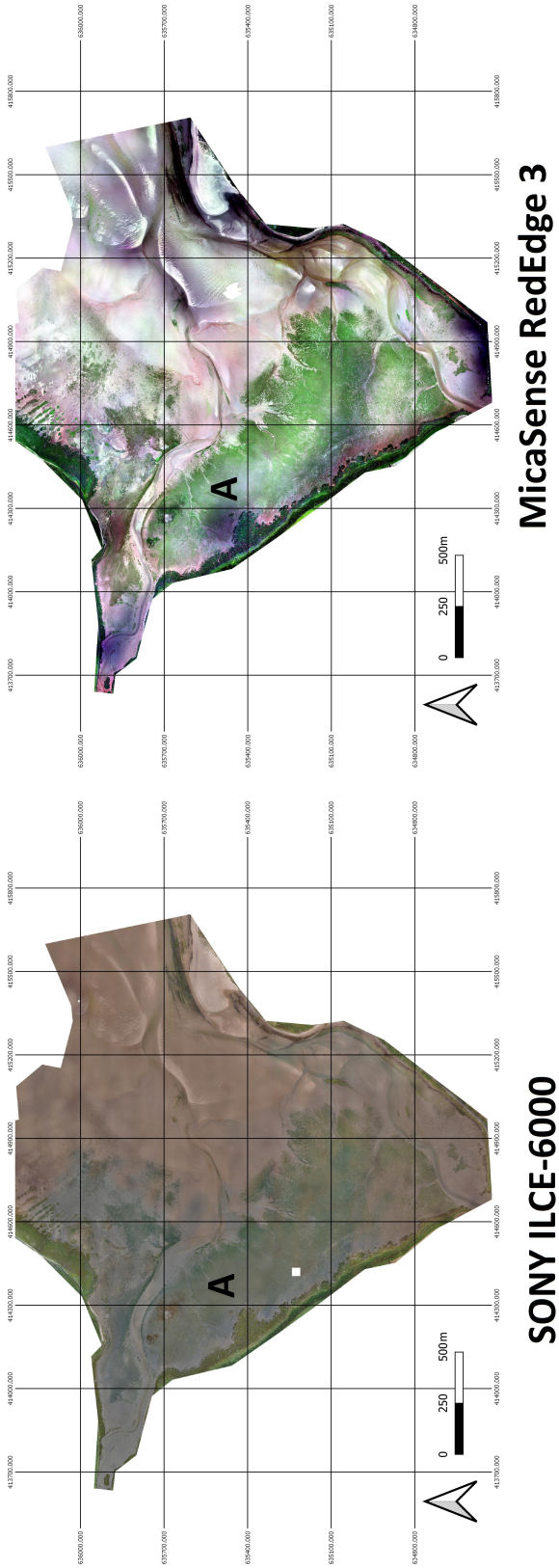

Figure 3.1: Orthomosaics of Budle Bay using a SONY ILCE-6000 (left) and a MicaSense RedEdge3 multispectral camera (right). Driving the display uses the Red, Green and Blue image bands. A close-up of each orthomosaics is shown in Figure 3.2.

Figure 3.2: A close-up of each orthomosaics with water channels and bordering intertidal vegetation such as *Enteromorpha sp.* and Seagrass (*Zostera noltii* and *Angustifolia*) among background sediment.

For vegetation, the following species for algae were found: *Microphytobenthos*, *Enteromorpha spp.* and a generic other macroalgae which included *Fucus*. The remaining vegetation classes were seagrass and saltmarsh. Given the aim of the mapping objective for Budle Bay was to examine areas of intertidal seagrass, both species of seagrass that were found: *Zostera noltii* and *Angustifolia* were merged to a single class.

Therefore, a total of seven target classes can be listed that includes background sediment features, and vegetation species of seagrass, algae and saltmarsh.

- Background sediment: Dry sand
- Background sediment: Other bareground
- Algae: *Microphytobenthos*,
- Algae: *Enteromorpha*
- Algae: Other macroalgae (including *Fucus*)
- Seagrass: *Zostera noltii* and *Angustifolia* merged to a single class
- Other plants: Saltmarsh

The team of expert ecologists from Cefas and the EA sampled 108 geographically referenced tags. For each tag, quadrat sampling was used to estimate the percentage cover of classes of interest, with each quadrat defining a 300×300mm square [Shuman and Ambrose, 2003; Mumby et al., 1997]. Quadrat sampling is a common method for ecologists to sample ecological features within a study site and accurately calculate the percentage cover of each target class (as shown in the list of bullet points). The process entails placing a see through square over the stated area and then sub-divisions within the square allow ecologists to quantify the percentage cover of each target class by adding up the number of sub-divisions that contain a particular ecological feature. This process can be seen in Figure 3.3. In order to apply fully convolutional neural networks for semantic segmentation, each sample point that contains the percentage cover of seven target classes was reduced to a single label by

selecting the class value with maximum percentage. Figure 3.4 shows the distribution of recorded tags across the site, and as mentioned, these were dispersed mainly on the Western, Central and Southern portions of the site.



Figure 3.3: Example of quadrat sampling for monitoring *Zostera marina* at Porth Dinllaen. Figure from [Davies et al., 2017]

3.3 Semi-supervised intertidal seagrass mapping

Intertidal seagrass plays a vital role in estimating the overall health and dynamics of coastal environments due to its interaction with tidal changes. However, most seagrass habitats around the globe have been in steady decline due to human impacts, disturbing the already delicate balance in environmental conditions that sustain seagrass. Miniaturization of multi-

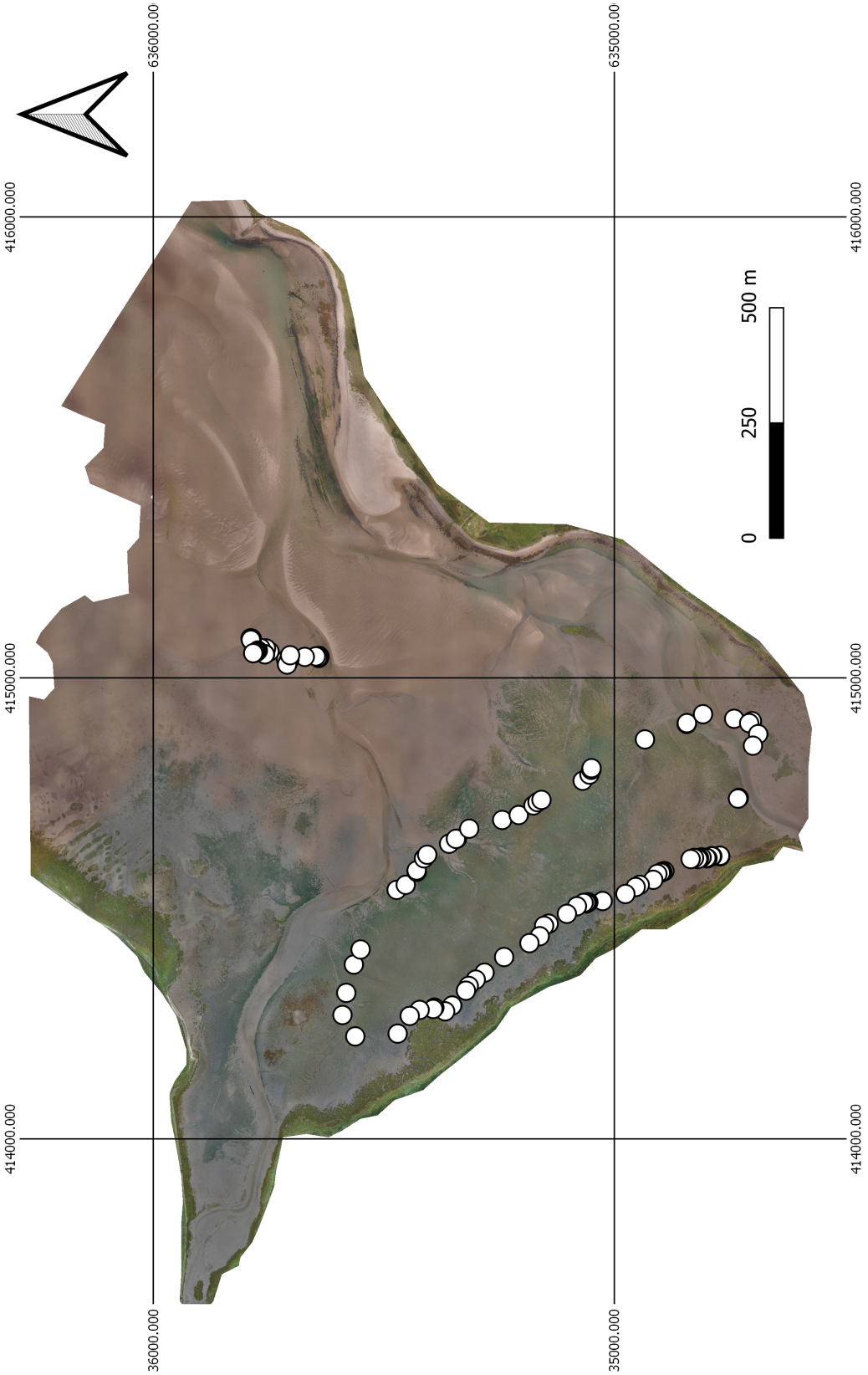

Figure 3.4: Distribution of recorded tags during the in-situ survey performed by a team of expert ecologists from Cefas and the EA. Each point records the percentage cover of each target class using 300×300mm quadrats. Then, each sample is reduced to a single label by selecting the highest percentage cover.

spectral sensors has facilitated very high resolution mapping of seagrass meadows that significantly improve the potential for ecologists to monitor changes.

Accurate and efficient mapping of intertidal seagrass ecosystems play a key role for estimating and assessing the health and dynamics of coastal ecosystems due to their sensitive response to tidal processes [Fonseca et al., 1983; Fonseca and Bell, 1998; Gera et al., 2013; Pu et al., 2014]; or human-made artificial interference [Short and Wyllie-Echeverria, 1996; Marbà and Duarte, 2010; Duarte, 2002]. Furthermore, seagrass plays a vital part in various coastal processes such as: sediment stabilization [McGlathery et al., 2012], pathogen reduction [Lamb et al., 2017], carbon sequestration [Fourqurean et al., 2012; Macreadie et al., 2014] and as an indicator for water quality [Dennison et al., 1993]. However, there is evidence seagrass areas have been in steady decline due to human disturbance for decades [Waycott et al., 2009].

In the following Section, two analytical approaches are used for classifying intertidal seagrass habitats; reviewed in chapter 2 are investigated: Object-based Image Analysis (OBIA) and Fully Convolutional Networks (FCNs). Both methods produce equivalent outputs in the form of semantically segmented maps. OBIA has been a prominent solution for coastal remote sensing, with many studies leveraging in-situ data and multiresolution segmentation to create habitat maps [Rasuly et al., 2010; Ventura et al., 2018; Butler et al., 2020; Husson et al., 2016; Purkis et al., 2019; Janowski et al., 2020]. This work demonstrates the utility of FCNs in two settings: a standard supervised approach and semi-supervised setting to map seagrass and other coastal features.

3.3.1 Methodology

Data pre-processing for FCNs

The recorded percentage covers were used to classify each point in Figure 3.4 to a single ecological class listed in Section 3.2 based on the highest estimated cover during the in-situ survey. The classification for each point provides the basis to create geographically

referenced polygon files through photo interpretation. This process generated a total of 72 polygons (56 generated from the site survey + 16 polygons from dry sand) that were split into train and test sets. The train set had 50 polygons and the test set 22. The use of photo interpretation instead of selecting segmented image-objects from OBIA segmentation is to avoid introducing bias from the OBIA method during FCN training.

Polygons to segmentation masks for FCNs

Each polygon contains a unique semantic value depending on the recorded class. FCNs were trained with segmentation maps that contain a one-to-one mapping of pixels encoded with a semantic value, with the goal to optimise this mapping [Long et al., 2015]. Segmentation maps used for training FCNs were created using the geographic coordinates stored in each polygon and converting real-world coordinates for each vertex to image-coordinates. As mentioned, each orthomosaic was split into 6000×6000 non-overlapping tiles. If a polygon fits within an tile, then the tile was cropped to an image size of 256×256 centered on the polygon. By cropping images centered on polygons the edges of each image have a number of pixels that were not labelled, Figure 3.7. The difference in spatial resolution for each camera results in a difference in labelled pixels, since each polygon covers the same area within the real-world. This process generated 534 images with the MicaSense RedEdge3 multispectral camera that were split into 363 images for training, 69 images for validation and 102 images for testing. The SONY camera produced 1108 images that were split into 770 images for training, 125 images for validation and 213 images for testing. Figure 3.5 shows the process of generating training images for FCNs using generated rasterised polygons and orthomosaics tiles. Figure 3.6 shows the examples of converting each in-situ point, shown in Figure 3.4, to a rasterised polygon through photo interpretation for each target class listed in Section 3.2. Figure 3.7 displays a gallery of images for each class with some example polygons.

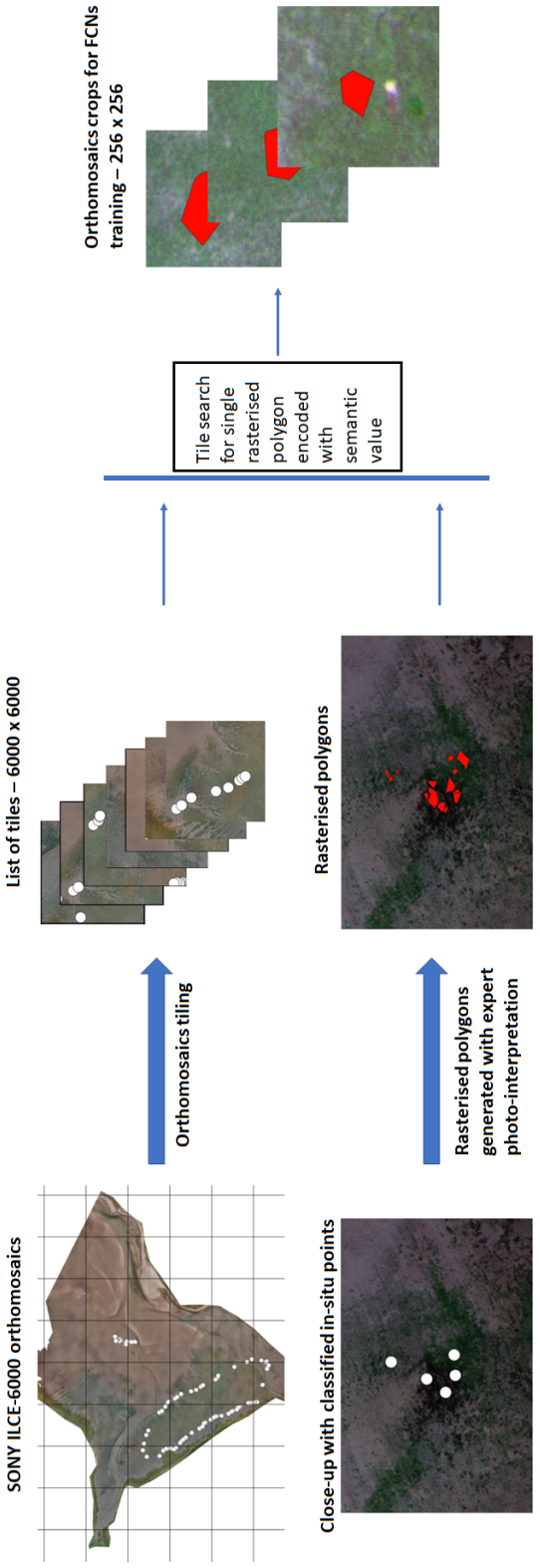

Figure 3.5: The process of generating training images for FCNs. First, the orthomosaics were split into 6000×6000 non-overlapping tiles. Second, the classified in-situ points, as shown in Figure 3.4, provide the basis to draw polygons using expert photo-interpretation. Lastly, the list of tiles and the rasterised polygons were combined by searching through each image tile for a rasterised polygon and then cropping the tile to a 256×256 image which is used to drive optimisation of FCNs.

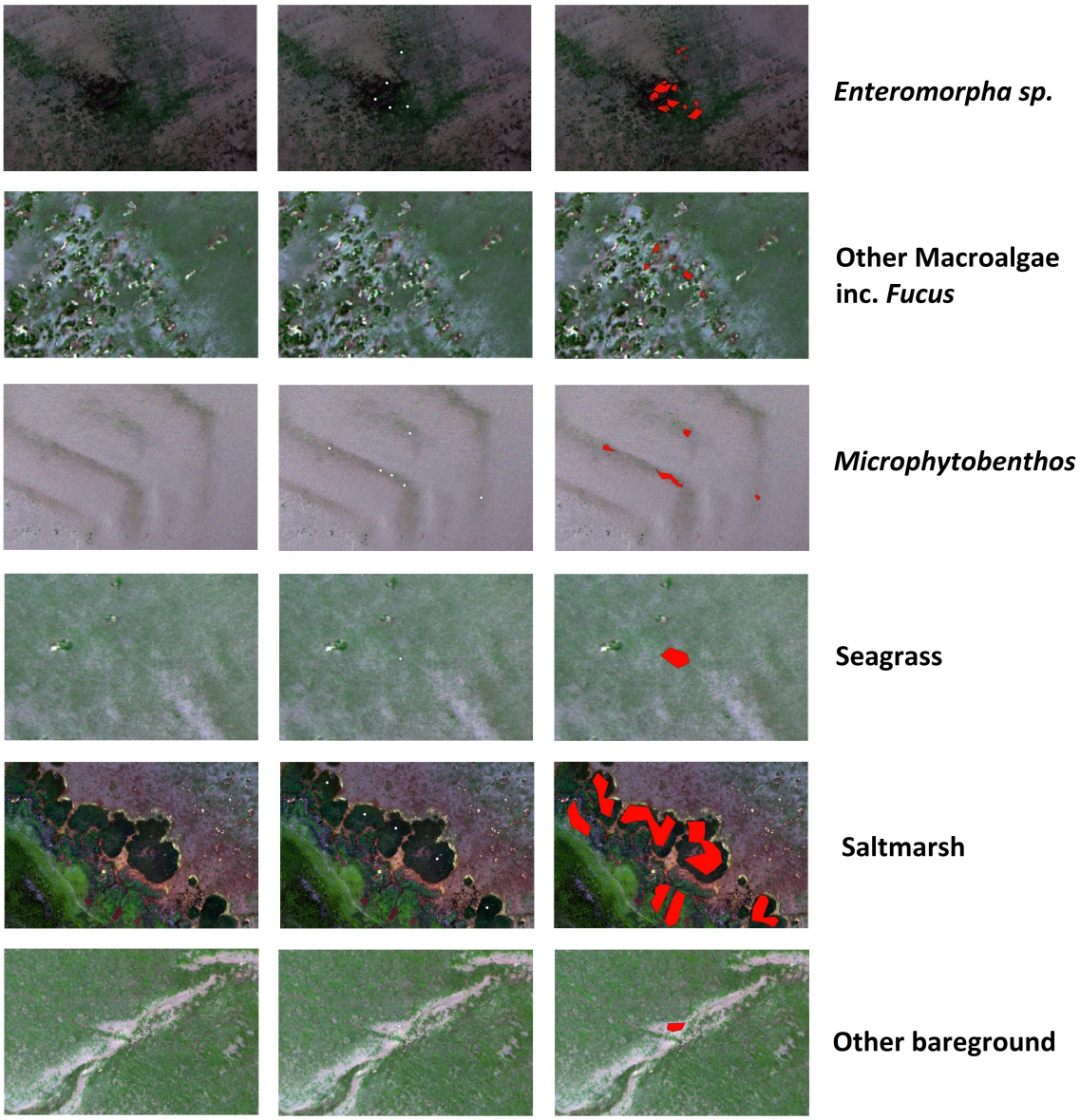

Figure 3.6: Gallery of images for each class with accompanying in-situ points that were used to annotate polygons. On the right are rasterised polygons which were generated through expert photo-interpretation and the knowledge conveyed by the classified in-situ point which can be seen in the middle column of images. The left column of images provides the same orthomosaic crop with no overlay of in-situ classified points and rasterised polygons.

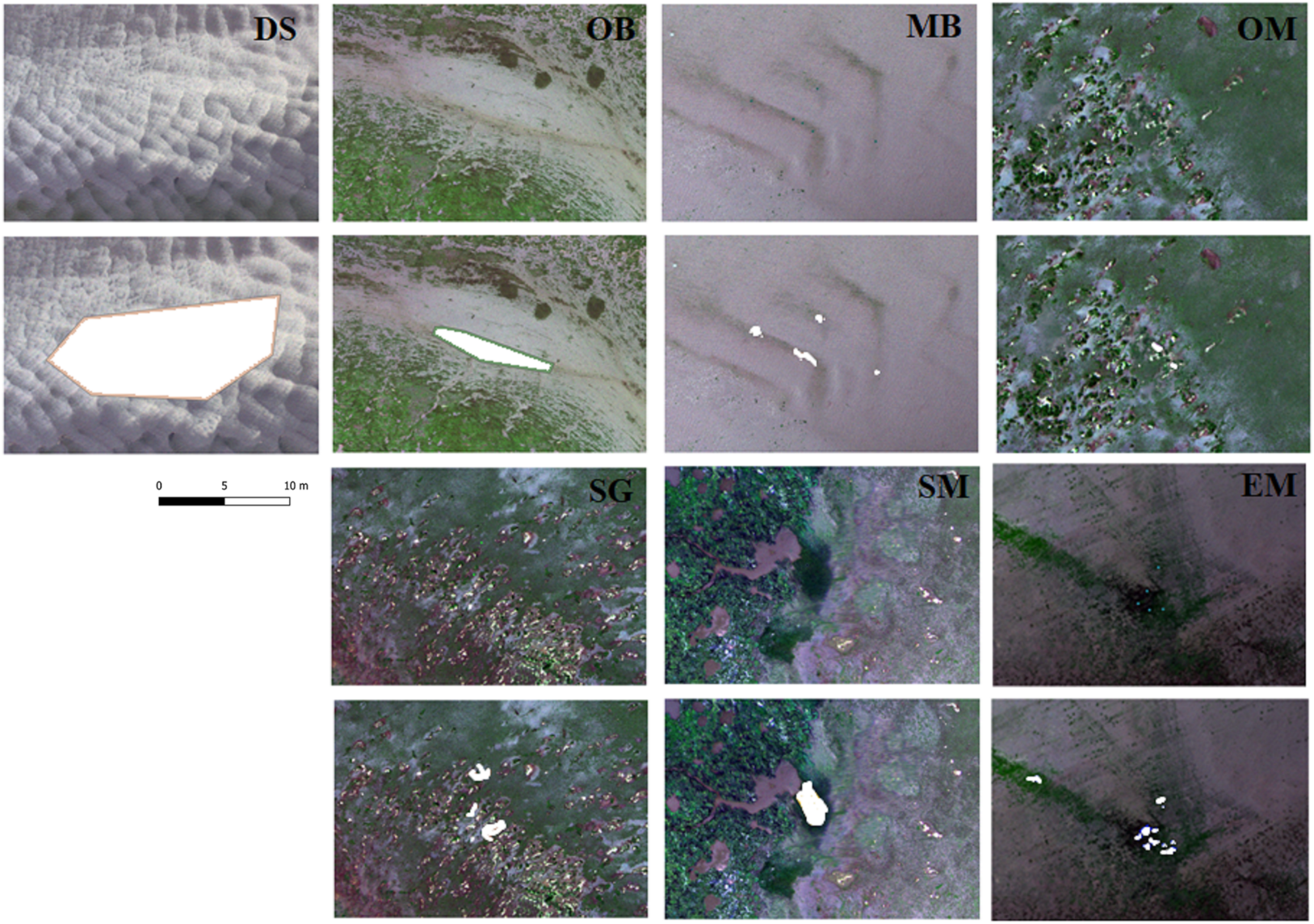

Figure 3.7: Gallery of images with annotated polygons. Each polygon contains a unique semantic value that drives the optimisation of fully convolutional neural networks. OM - Other Macroalgae inc. *Fucus*; MB - *Microphytobentos*; EM - *Enteromorpha*; SM - Saltmarsh; SG - Seagrass; DS - Dry Sand; OB - Other Bareground. Images with white polygons are examples of polygons used for modelling.

Vegetation, soil and atmospheric indices for FCNs

Vegetation, soil and atmospheric indices are derivations from standard Red, Green and Blue and/or Near-infrared image bands that can aid discerning multiple vegetation classes [Xue and Su, 2017]. Near-infrared, Red, Green and Blue bands from the MicaSense RedEdge3 were used to compute a variety of indices, adding five bands of data to each input image. These extra bands were: Normalised Difference Vegetation Index (NDVI) [Rouse et al., 1974], AAtmospheric Resistant Vegetation Index (IAVI) [Ren-hua et al., 1996], Modified Soil Adjusted Vegetation Index (MSAVI) [Qi et al., 1994], Modified Chlorophyll Absorption Ratio Index (MCARI) [Daughtry et al., 2000] and Green Normalised Difference Vegetation Index (GNDVI) [Louhaichi et al., 2001]. The Red, Green and Blue channels for both cameras were used to compute additional four indices, namely: Visible Atmospherically Resistant Index (VARI) [Gitelson et al., 2002], Visible-band Difference Vegetation Index (VDVI) [Xiaoqin et al., 2015], Normalised Green-Blue Difference Vegetation Index (NGBDI) [Verrelst et al., 2008] and Normalised Green-Red Difference Vegetation Index (NGRDI) [Tucker, 1979]. The choice of these indices was mostly due to the importance of the Green channel for measuring reflected vegetation spectra, while also providing more data for FCNs to model complex one-to-one mappings for each pixel.

The mentioned index images were stacked onto the channel dimension which resulted in images for the MicaSense RedEdge3 and the Sony camera having 14 and seven bands, respectively. Furthermore, each individual image band was scaled to a value between zero and one.

Fully Convolutional Networks

Fully Convolutional Neural Networks [Ronneberger et al., 2015; Long et al., 2015; Chen et al., 2018] are an extension of traditional CNN architectures for image classification [LeCun et al., 2015b; Krizhevsky et al., 2012] adapted for semantic segmentation. Figure 3.8 displays the architecture used for this chapter. The overall architecture is a U-Net [Ron-

neberger et al., 2015] and the encoder network is a ResNet101 [He et al., 2016] pre-trained on ImageNet. Residual learning has proven to surpass very deep neural networks [He et al., 2016] and is a suitable encoder network for the overall U-Net architecture. The decoder network applies a transposed 2×2 convolution for upsampling while also concatenating feature maps from each encoding stage at appropriate resolutions followed by a 3×3 convolution. The final layer uses 1×1 convolution and condenses feature maps without spatial decimation to have the same number of channels as the total number of classes before a softmax transfer function classifies each pixel.

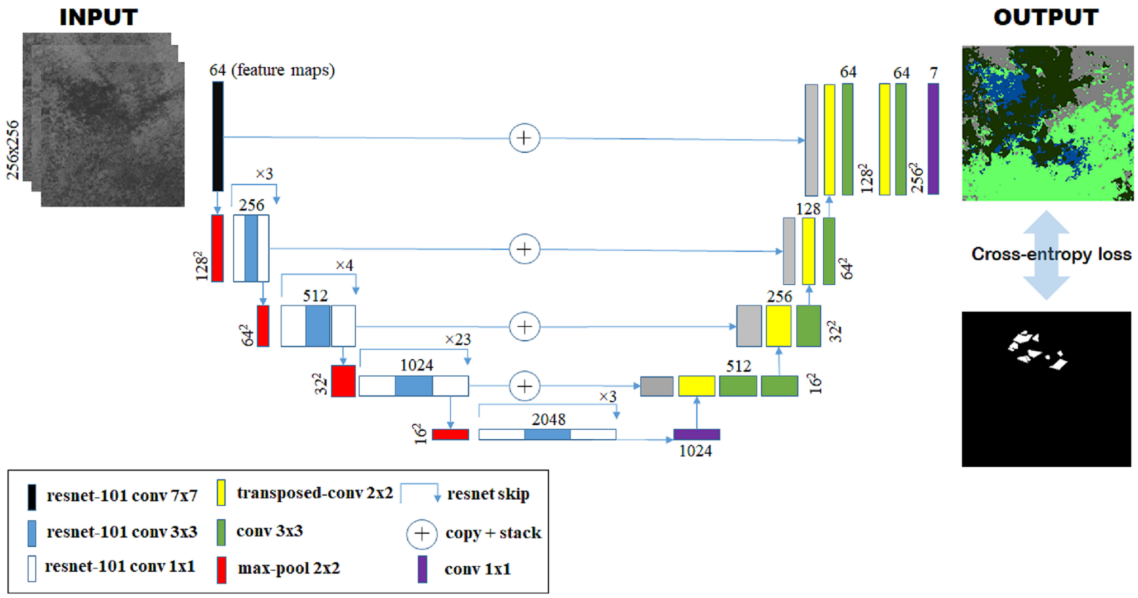

Figure 3.8: U-Net architecture and loss calculation. The input channels are stacked and passed through the network. The encoder network applies repeated convolution and max pooling operations to extract feature maps, while in the decoder network upsamples these and stacks features from the corresponding layer in the encoder path. The output is a segmented map that is compared with the mask using cross-entropy loss. The computed loss is used to train the network, through gradient descent optimisation

For semi-supervised training the Teacher-Student method was used [Tarvainen and Valpola, 2017]. This approach requires two networks: a teacher and a student, both having the same architecture as shown in Figure 3.8. The student network is updated through gradient descent minimising the sum of two loss terms: a supervised loss calculated on labelled

pixels of each segmentation map, and conversely, an unsupervised loss calculated using non-labelled pixels. The teacher network is updated using an Exponential Moving Average (EMA) of weights from the student network. The EMA is a popular statistical calculation used to analyse data points over time, with a particular emphasis on the most recent data. Therefore, the moving average provides more weight to recent data points which in turn allows responsive to changes in the underlying data. In this scenario, the teacher network provides the EMA which passes knowledge about the underlying trends to guide the learning process given the student learns from both labeled and unlabeled data.

Weighted training for FCNs

Section 3.3.1 detailed the process of creating segmentation maps from polygons. Both sets of images from each camera had an imbalanced target class distribution. Figure 3.9 shows the number of labelled pixels per class and also the number of non-labelled pixels for each camera. The recorded distribution poses a challenge for classes such as other macroalgae

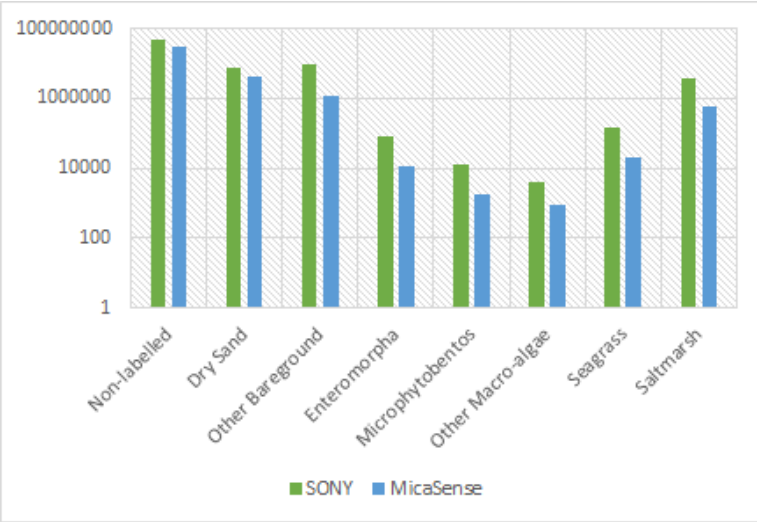

Figure 3.9: Distribution of labelled pixels within the polygons for each class and non-labelled pixels on a log-scale. Given each training image was a 256×256 crop of the orthomosaic centered around each polygon, the number of non-labelled pixels refers to image pixels outside the polygon that reside within the 256×256 crop.

and *Microphytobentos* due to the relatively low number of labelled pixels in comparison

with the remaining classes. The pixel counts shown in Figure 3.9 were used to calculate the probability of each class in the training set. For each class, a weight was computed using the inverse of each probability. During training the supervised loss was scaled with respect to calculated weights based on class abundance.

$$w_i = (p_i K)^{-1} \tag{3.1}$$

Where, w_i is i^{th} weight for a given class probability p_i and K is the total number of classes.

Supervised loss

For the supervised loss term, consider $X \in \mathbb{R}^{B\times C\times H\times W}$ and $Y \in \mathbb{Z}^{B\times H\times W}$ to be respectively, a mini-batch of images and corresponding segmentation maps; where B, C, H and W are respectively, batch size, number of input channels, height and width. Processing a mini-batch with the student network outputs per-pixel scores $\hat{Y} \in \mathbb{R}^{B\times K\times H\times W}$; where K is the number of target classes. The softmax transfer function converts network scores into probabilities by normalising all K scores for each pixel to sum to one.

$$P_k(\mathbf{x}) = \frac{\exp \hat{Y}_k(\mathbf{x})}{\sum_{k'=1}^{K} \exp \hat{Y}_{k'}(\mathbf{x})} \tag{3.2}$$

Where, $\mathbf{x} \in \Omega$; $\Omega \subseteq \mathbb{Z}^2$ is a pixel location and $P_k(\mathbf{x})$ is the probability for the kth channel at pixel location $\mathbf{x}$, with $\sum_{k'=1}^{K} P_{k'}(\mathbf{x}) = 1$. Then, the negative log-likelihood loss is calculated between segmentation maps and network probabilities.

$$L_s(P, Y) = \begin{cases} 0, & \text{if } Y(\mathbf{x}) = -1 \\ -\sum_{k=1}^{K} Y_k(\mathbf{x}) \log(P_k(\mathbf{x})), & \\ & \text{if } Y(\mathbf{x}) \neq -1 \end{cases} \tag{3.3}$$

For each image, the supervised loss is the sum of all losses for each pixel using eq. 3.3 and scaled according to the number of labelled pixels in Y.

Unsupervised loss

Previous work in semi-supervised segmentation details using a Teacher-Student model and advanced data augmentation methods in order to create two images for each network to process [French et al., 2020a; Olsson et al., 2021]. While this work did not use data augmentation methods, pairs of images were created by leveraging labelled and non-labelled pixels in Y.

As with the supervised loss term, a mini-batch of images is passed through both the student and the teacher networks, respectively producing per-pixel scores $\hat{Y}$ and $\bar{Y}$. Again, pixel scores are converted to probabilities with softmax, equation 3.2, respectively producing $\hat{P}$ and $\bar{P}$. The maximum-likelihood of teacher predictions was used to create pseudo-labels to compute the loss for non-labelled pixels in Y. Thus, the unsupervised loss is also calculated similarly to equation 3.3 but the negative log-likelihood is computed between predictions from the student model ($\hat{P}$) and a pseudo-label map (Y^p) generated by the teacher model on pixels that are initially non-labelled.

$$L_u(\hat{P}, Y^p) = \begin{cases} 0, & \text{if } Y(\mathbf{x}) \neq -1 \\ -\sum_{k=1}^{K} Y_k^p(\mathbf{x}) \log(\hat{P}_k(\mathbf{x})), & \\ & \text{if } Y(\mathbf{x}) = -1 \end{cases} \tag{3.4}$$

For each image, the unsupervised loss was the sum of all losses for each pixel using equation 3.4 scaled according to the number of non-labelled pixels within Y. Confidence thresholds were also used so that only confident or high probability predictions from the teacher network would aid the unsupervised loss, and also so that initial optimisation steps focus more on the supervised loss term. Classes with a relatively low number of labelled pixel would benefit from the unsupervised loss term, as confident teacher predictions can guide

the decision boundaries of student models by adding pseudo-label maps to consider. Figure 3.10 shows the overall architecture of the method that was used to generate the results in Section 3.3.3.

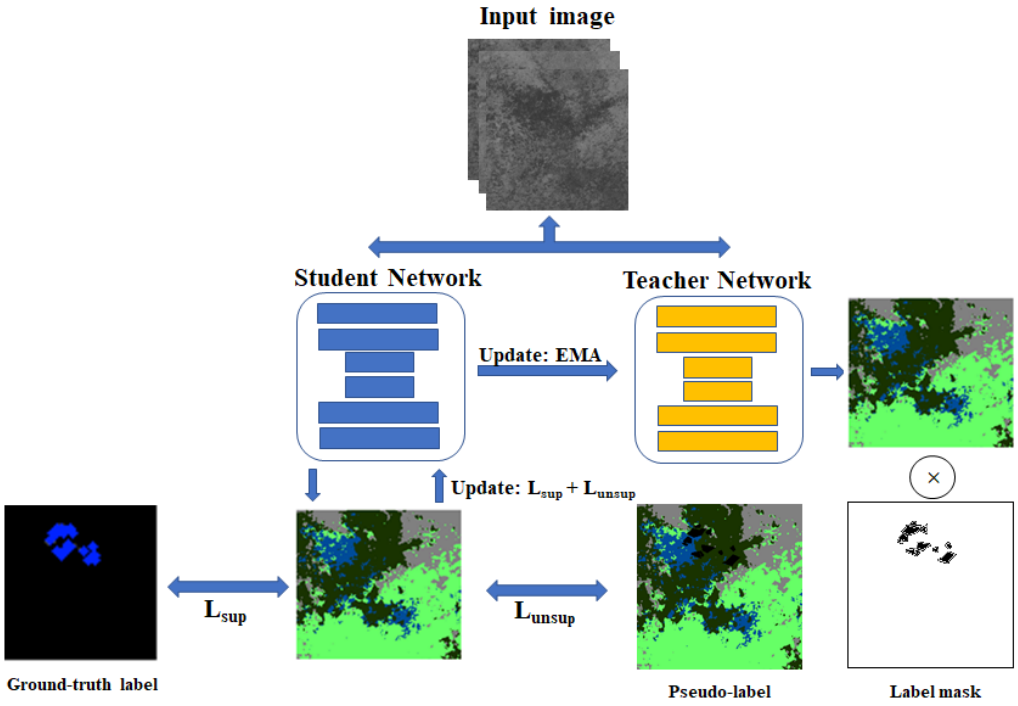

Figure 3.10: Mean-Teacher architecture for semi-supervised segmentation using pseudo-labels. The student network produces predictions that are used to compute the supervised loss with known polygons. The teacher network produces predictions that are then converted to hard pseudo-labels which are then used to drive the unsupervised loss. The student is updated using the combined loss and gradient descent, and the teacher network is updated with an EMA of the student network weights.

Training parameters

Combining both loss terms yields the objective cost used for optimising FCNs in a semi-supervised setting.

$$L = wL_s + \gamma L_u \tag{3.5}$$

Where, L_s and L_u are respectively the supervised and unsupervised loss term. The supervised loss was scaled according to the weights computed in Eq. 3.1 and the unsupervised loss to γ that was set to 0.1 for all experiments.

All networks were pre-trained on ImageNet. Networks for each camera were trained for 150 epochs with a batch-size of 16 using Adam optimiser [Kingma and Ba, 2014]. The learning rate was initially set to 0.001 and reduced by a factor of ten every 70 epochs of training. The confidence threshold for teacher predictions was set to 0.97. The hyper-parameters were chosen through an exhaustive search of various settings and monitoring the convergence with loss plots and accuracy metrics using the validation set without cross-validation. All FCNNs were implemented and trained using Pytorch version 10.2.

OBIA

The OBIA method for modelling multiple coastal features was performed using eCognition v9.3 [Benz et al., 2004]. This software possesses the tools to process high resolution orthomosaics and shape file exports from GISs to create supervised machine learning models. Section 3.3.1 detailed a number of methods used to pre-process the orthomosaics and shape polygons, however the OBIA does not require this.

The first step in OBIA is to process each orthomosaic using a multi-resolution segmentation algorithm to partition the image into segments [Benz et al., 2004]. The bottom-up segmentation process starts with individual pixels and clusters pixels to image-objects using one or more criteria of homogeneity. The subsequent clustering of two adjacent image-objects or image-objects that are a subset of each other is based on the following criterion:

$$h = \sum_{c} N(o_c^m - o_c^1) + M(o_c^m - o_c^2), \tag{3.6}$$

Where o^1, o^2 and o^m respectively represent the pixel values for objects one, two and a candidate virtual merge m. N and M are the number of total pixels, respectively for objects one and two. This criterion evaluates the change in homogeneity during fusion of

image-objects. If this change exceeds a hyper-parameter threshold value, then the fusion is not performed. In contrast, if the change in image-objects is below the threshold, then both candidates are clustered to form a larger region. The segmentation procedure stops when no further fusions are possible without exceeding the threshold value. In eCognition, this hyper-parameter is also known as the scale parameter. The geometry of each shape is defined by two other hyper-parameters: shape and compactness. For this work, the scale parameter was set to 200, the shape to 0.1 and the compactness to 0.5. Figure 3.11 shows image objects overlaid on top of both orthomosaics. These values were determined by the expert geomorphologist at Cefas responsible for generating the results with OBIA. Again, the choice of these hyper-parameter was through an exhaustive search and correlating the segmented objects, using multi-resolution segmentation, with the physical size of ecological features that were present within the orthomosaics. This in turn requires expert domain knowledge to accurately correlate the outputs of the segmentation with ecological features.

In Section 3.3.1, the split of polygons for training and testing was detailed. Each polygon, as shown in Figure 3.7, from the training set was superimposed on top of image-objects to select the candidate segments for extracting spectral features. Selected image-objects create a database for the in-built Random Forest [Breiman, 2001] in eCognition. The spectral features for the MicaSense RedEdge3 camera were: channel mean and standard deviation, vegetation and soil indices (NDVI, RVI, GNDVI, SAVI), ratios between Red/Blue, Red/Green and Blue/Green image layers and the intensity and saturation components of the HSI colour space. The features for the SONY were the same but the vegetation and soil indices were not added. Once the features and image-objects were selected, the Random Forest modeller produced a number of Decision Trees [Quinlan, 1986] with each tree being optimised using the GINI Index [Lerman and Yitzhaki, 1984].

Overall efforts - time and hyper-parameter adjustments

The overall effort in terms of time spent between the OBIA and the proposed FCN methods was similar.

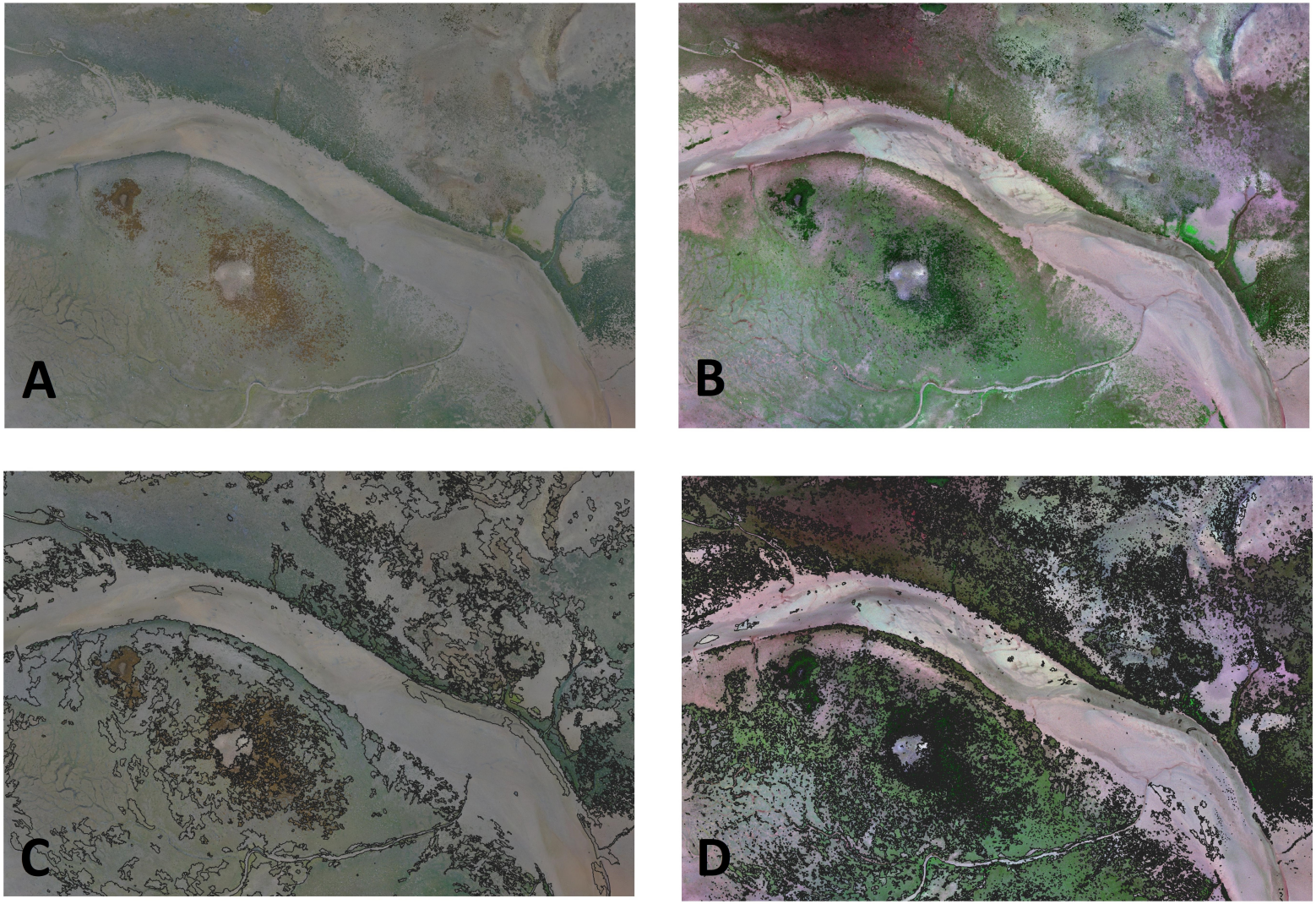

Figure 3.11: Segmented orthomosaics using the multiresolution segmentation algorithm. On the top-left (**A**) is a crop of the orthomosaic captured with the SONY-ILCE 6000. On the top-right (**B**) is is a crop of the orthomosaic captured with the MicaSense RedEdge3 multispectral camera. The corresponding segmented orthomosaics crop are shown on the bottom-left and bottom-right (**C** and **D**, respectively for the SONY-ILCE 6000 and the MicaSense RedEdge3). Both segmented orthomosaics were extracted with the same hyper-parameters. The scale, shape and compactness were respectively 200, 0.1 and 0.5. Given the spatial resolution from both orthomosaics is different, the resulting segmented orthomosiacs are also different since both use the same scale parameter

To start with, both methods require the same labels, or polygons, in order to drive the optimisation of machine learning models. The use of spatially explicit labels was the key factor for optimisation, and both methods require this in order to effectively learn complex relationships from features derived using orthomosaics to target classes. Therefore, the time spent creating the labelled dataset via rasterised polygons exported from GIS was the same for both OBIA and FCNs.

The use of OBIA relies on the use of eCognition v9.3 [Benz et al., 2004] which streamlines the process of training and testing and thus circumvents the pre-processing steps required for FCNs. During supervised training procedures, the OBIA method requires two steps of hyper-parameter adjustment. The first step was to adjust hyper-parameters for the MRS, such as scale, shape and compactness, with the scale parameter being critical for image-object creation. In turn, this requires the user using eCognition v9.3 [Benz et al., 2004] to understand the target class domain for a particular mapping objective in order to correlate segmented image-objects with known aerial extents of the class domain which often requires multiple runs of the MRS in a trial and error fashion. The latter step can vary on time depending on size of the orthomosaics, which impacts the run-time of MRS, and the satisfaction of the end user when correlating image-objects with the underlying features in the orthomosaics, which requires domain expertise. The second step was to to adjust the hyper-parameters used for optimising Machine Learning models, such as Random Forests and Support Vector Machines (SVMs). During inference procedures, OBIA allows for the segmented orthomosaics (as a product of MRS) to be loaded and then individual image-objects were classified using optimised models in order to create thematic maps. Again, this process can vary on time depending and the satisfaction of the end user with regards to objective metrics as well as subjective visual analysis of generated thematic maps.

As mentioned, the proposed FCN method requires several pre-processing steps such that the dataset format is suitable for FCN optimisation. This requires tiling the orthomosaics and center cropping images using the rasterised polygons with additional geographic in-

formation. The time spent pre-processing the orthomosaics and rasterised polygons was static and depends on the number of polygons and the size of the orthomosaics. During training, the main objective was to ensure that the convergence of FCNs was appropriate by adjusting the batch size, learning rate and total number of iterations parameters. The latter was achieved by monitoring loss plots and validation accuracies that required several runs in a trial and error fashion. During inference, optimised FCNs were used to classify the non-overlapping tiles which were then stitched together using QGIS v3.10. The time taken for stitching individual tiles was also static and depends on the total number of tiles generated during the pre-processing stages.

Overall, both methods require similar time in order to produce accurate thematic maps with appropriate training and inference. The OBIA has the advantage with the use of eCognition v9.3 that allows each step to be streamlined. However, the main drawback with OBIA was to adjust two sets of hyper-parameters, one for feature extraction with MRS and the other during model tuning. The use of FCNs requires additional pre-processing steps for training and then additional post-processing steps after inference but the time taken for optimising FCNs was relatively lower.

3.3.2 Accuracy assessment

The measurements used to objectively quantify results were pixel accuracy, precision, recall and F1-score. Pixel accuracy is the ratio between pixels that were classified correctly and the total number of labelled pixels in the test set for a given class. Precision and recall are metrics that can show how a classifier performs for each specific class. F1-score is the harmonic mean of recall and precision and is therefore a suitable metric to quantify classifier performance when a single figure of merit is needed. Equation 3.7 details each of these metrics where TP, TN, FP and FN were respectively, True Positive, True Negative, False Positive and False Negative pixel classifications.

$$pixel\,accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3.7}$$

$$precision = \frac{TP}{TP + FP} \tag{3.8}$$

$$recall = \frac{TP}{TP + FN} \tag{3.9}$$

$$F1 = 2 \times \frac{recall \times precision}{recall + precision} \tag{3.10}$$

3.3.3 Experiments and Results

The outputs for both the FCNs and OBIA were compared with the test set polygons. Figures 3.12 and 3.13 display confusion matrices scoring outputs from each method and camera as pixel accuracy. The confusion matrices also show pixel accuracies for FCNs that were optimised using equation 3.3 and models that were optimised using both equations 3.3 and 3.4. The confusion matrices are average results over three sequential train-test runs with the set of hyper-parameters described in Section 3.3.1. Overall results for OBIA and FCNs in a semi-supervised setting for each camera can be viewed in Table 3.1, where precision, recall and F1-score are reported. Figure 3.14 displays habitat maps for each method and camera.

SONY ILCE-6000 results

Predictions with the OBIA method had an average pixel accuracy of 90.6%. Classes related to sediment had scores of 100% and 98.38%, respectively for dry sand and other bareground. Algal classes scored 97.6%, 88.09% and 83.18%, respectively for *Enteromorpha*, *Microphytobentos* and other macroalgae (inc. *Fucus*). Seagrass predictions were found to score 93.67% and saltmarsh was the worst performing class for the OBIA with 73.32%.

FCNs yielded an average class accuracy of 76.79% and 83.3%, respectively for supervised and semi-supervised settings. Both approaches scored close to 100% for dry sand and

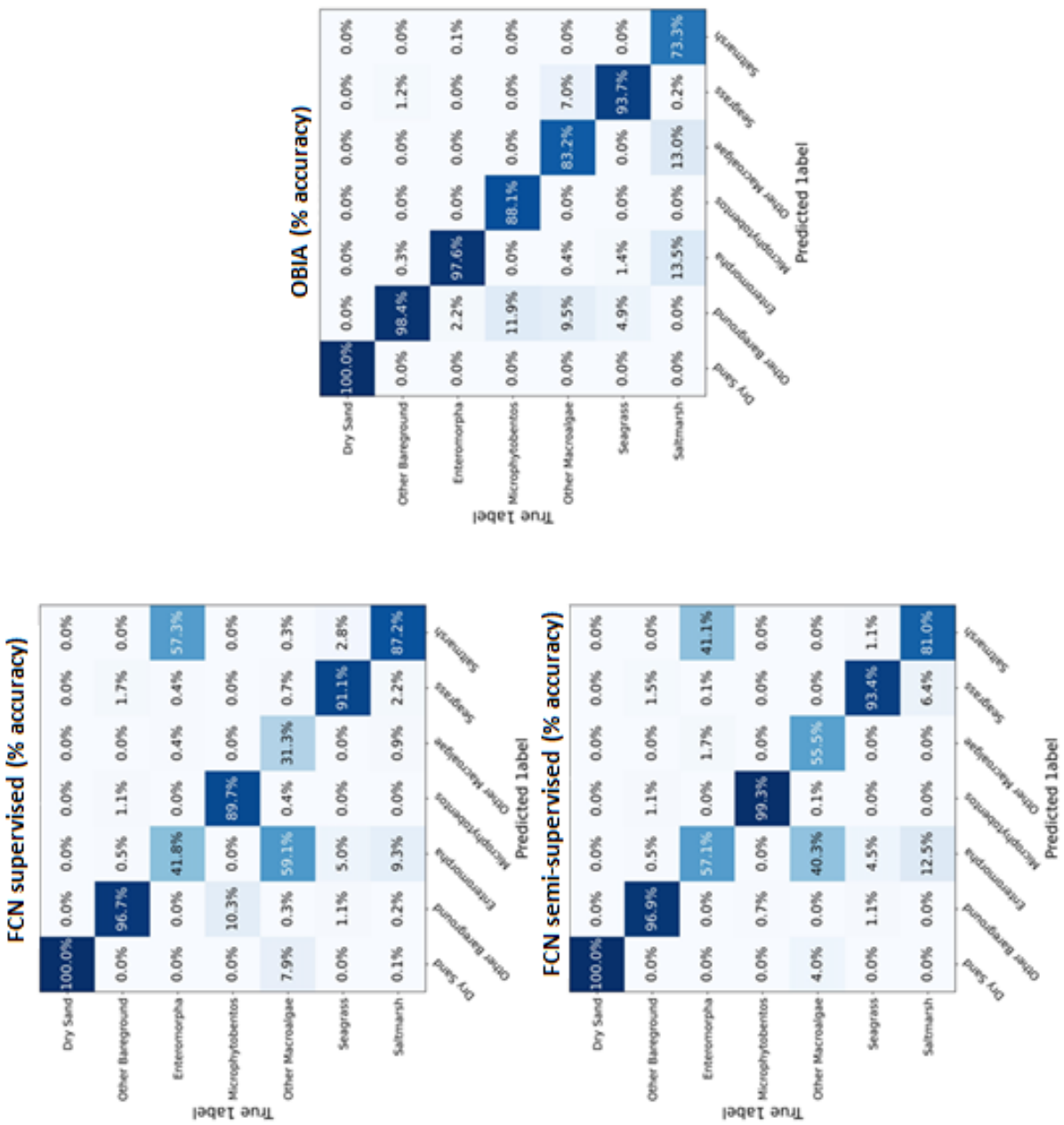

Figure 3.12: Confusion matrices for both methods using the SONY camera.

other bareground performed better in a semi-supervised setting scoring 96.88%. Scores for *Enteromorpha* and other macroalgae (inc. *Fucus*) were respectively 38.72% and 32.29% for supervised training and 57.05% and 55.90% for semi-supervised training. Seagrass scored similarly in both training settings with approximately 90% and saltmarsh scored better in a supervised setting with 87.78%, while the semi-supervised setting scored 81%.

MicaSense RedEdge3 results

The OBIA method had an average pixel accuracy of 73.44%. Sediment classes such as dry sand and other bareground scored 63.18% and 42.80%. Algal classes scored 93.42%, 72.54% and 49.31%, respectively for *Enteromorpha*, *Microphytobentos* and other macroalgae (inc. *Fucus*). The remaining vegetation classes of seagrass and saltmarsh both presented high scores of 95.48% and 96.38.

FCNs yielded an average class accuracy of 83.68% and 88.7%, respectively for supervised and semi-supervised settings. Both models had good scores for sediment classes scoring above 95% in pixel accuracy. Algal classes of *Enteromorpha*, *Microphytobentos* and other macroalgae (inc. *Fucus*) respectively scored for supervised and semi-supervised training (91.5%, 91.3%), (87.3%, 93.6%) and (45.5%, 63.6%). Seagrass predictions scored 69.6% and 76.2, while saltmarsh was found to score 97.4% and 98.9%, respectively for supervised and semi-supervised training.

Overall results

Table 3.1 shows the scores for precision, recall and F1-score for OBIA and FCNs trained in semi-supervision.

Habitat maps

Figure 3.14 shows the habitat maps of Budle Bay for each camera and method previously described.

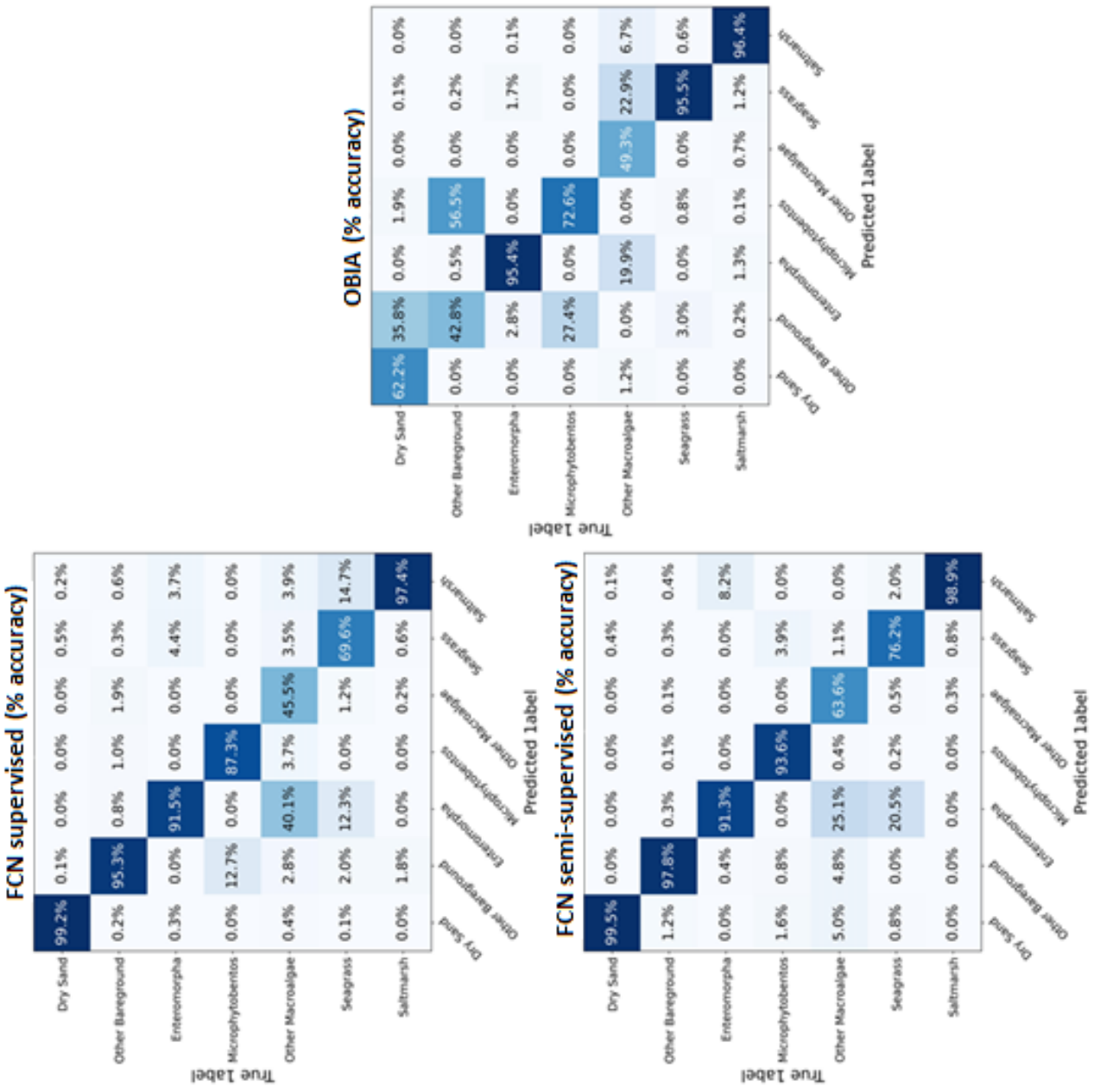

Figure 3.13: Confusion matrices for both methods using the RedEdge3 multispectral camera.

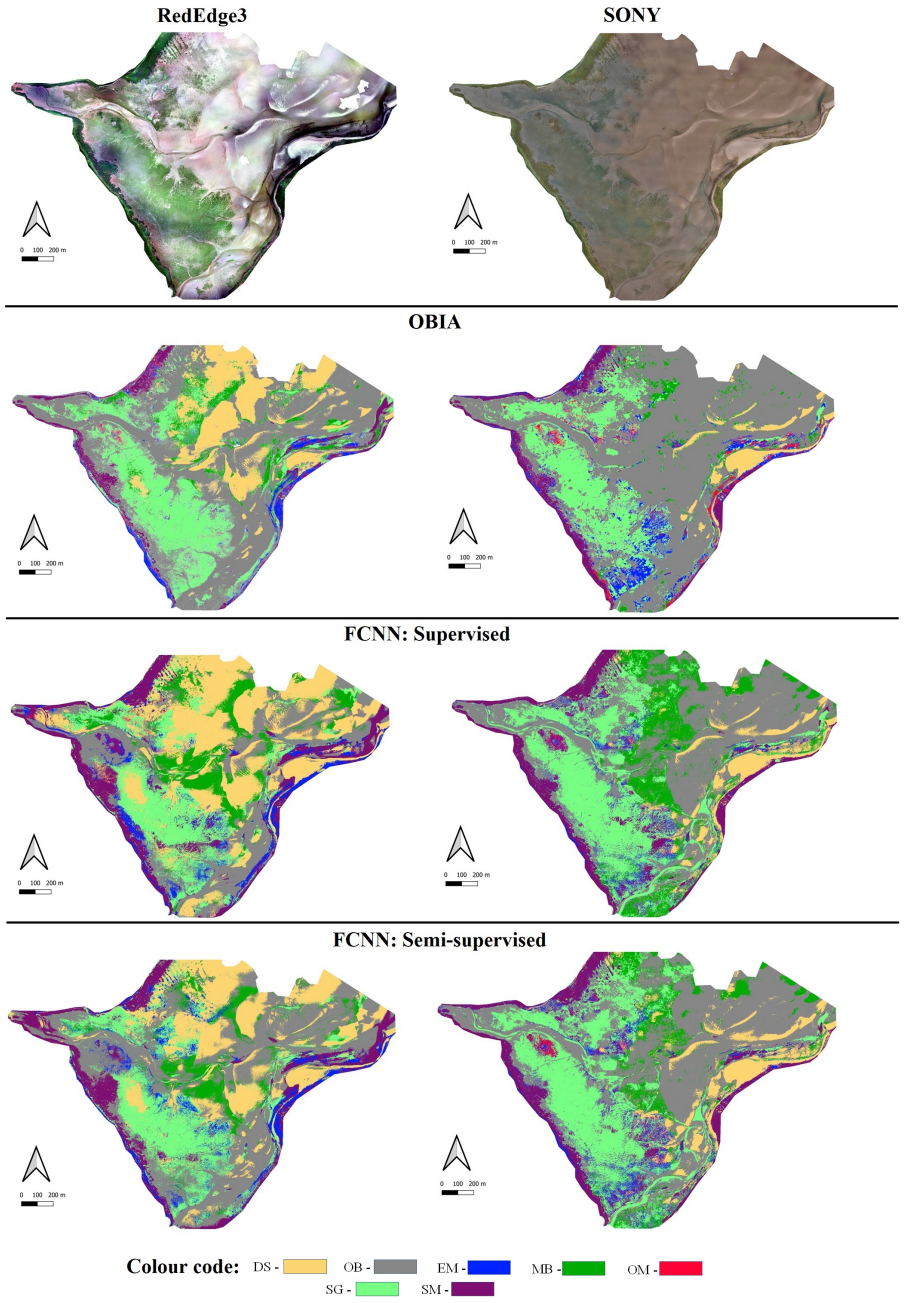

Figure 3.14: Segmented habitat maps for both cameras and methods. The top row - RedEdge3 and SONY cameras orthomosaics. The second row - habitat maps using the OBIA approach. The third row - FCNN maps in a supervised setting. The bottom row - FCNN maps in a semi-supervised setting. The left column - RedEdge3 images and segmented maps, the right column - the SONY images and maps. Legend: OM - Other Macroalgae inc. *Fucus*; MB - *Microphytobentos*; EM - *Enteromorpha*; SM - Saltmarsh; SG - Seagrass; DS - Dry Sand; OB - Other Bareground.

	P	R	F1	P	R	F1	P	R	F1	P	R	F1
DS	0.99	0.62	0.76	0.98	0.99	**0.99**	1.0	1.0	**1.0**	0.99	1.0	0.99
OB	0.56	0.42	0.48	0.99	0.97	**0.98**	0.99	0.98	**0.99**	0.99	0.97	0.98
EM	0.73	0.95	**0.83**	0.77	0.91	**0.83**	0.25	0.97	**0.40**	0.18	0.57	0.27
MB	0.008	0.72	0.01	0.84	0.93	**0.88**	1.0	0.88	**0.93**	0.30	0.99	0.46
OM	0.25	0.49	0.33	0.47	0.63	**0.54**	0.02	0.83	0.05	0.66	0.55	**0.60**
SG	0.67	0.95	**0.78**	0.67	0.76	0.71	0.64	0.93	**0.76**	0.27	0.93	0.42
SM	0.99	0.96	**0.98**	0.98	0.98	**0.98**	0.99	0.73	0.84	0.97	0.81	**0.88**
MicaSense: **OBIA**				MicaSense: **FCNN**			SONY: **OBIA**			SONY: **FCNN**		

Table 3.1: Precision, recall and F1 scores for both algorithms on both cameras. The results for FCNs reflect models trained using semi-supervision. DS - Dry Sand; OB - Other bareground; EM - Enteromorpha; MB - Microphytobentos; OM - Other macroalgae; SG - Seagrass; SM - Saltmarsh

3.3.4 Discussion

Figures 3.12, 3.13 and 3.14 as well as Table 3.1 indicate that FCNs provide comparable performance to OBIA. Figures 3.12 and 3.13 also show an increase in performance for the semi-supervised FCN models in comparison to the supervised setting.

FCNs convergence

The convergence of FCNs was analysed by testing multiple settings and hyper-parameters. The optimal set of hyper-parameters was determined by assessing computed confusion matrices and validation losses over three sequential train/test runs with a given set of hyper-parameters. Figures 3.12 and 3.13 show average pixel accuracy scores over three sequential runs with the same hyper-parameters described in Section 3.3.1.

SONY ILCE-6000 analysis

Habitat maps from the SONY camera were found to perform better with the OBIA than FCNs in terms of average pixel accuracy and F1-score. Respectively, the OBIA had an average accuracy and F1-score of 90.6% and 0.71, while FCNs in a semi-supervised setting had 83.3% and 0.65.

Sediment class predictions for both methods scored well, with both metrics either scoring above 90% or above 0.9, respectively for pixel accuracy and F1-score. This suggests that the OBIA and FCNs methods successfully predicted test polygons for sediment classes while also avoiding false positive and false negative pixel classifications.

Algal classes were found to have mixed performance depending on the method used. Scores in Figure 3.12 with OBIA noted that classes of *Enteromorpha* and other macroalgae (inc. *Fucus*) scored better, while *Microphytobentos* were more accurate with FCNs. However, scores in Table 3.1 for the same classes suggest that OBIA performed better for classes of *Enteromorpha* and *Microphytobentos*, while FCNs scored better for other macroalgae. Analysing areas in Figure 3.14 that were predicted as *Enteromoprha* with OBIA and comparing these areas with FCN habitat maps show that the latter method interchangeably predicts *Enteromorpha* and saltmarsh. This observation can be supported by Figure 3.12 where 60.43% and 41.14% of test labels for *Enteromorpha* were predicted as saltmarsh, respectively for supervised and semi-supervised settings. These points suggest that habitat maps detailing areas for *Enteromorpha* with OBIA were more likely to be correct. Pixel classifications in Figure 3.12 for *Microphytobentos* indicate that FCNs performed well and accurately mapped test polygons of *Microphytobentos*, however figures for precision and F1 in Table 3.1 also indicate that FCNs have high false positive rate for this class. Conversely, OBIA produced a perfect figure for precision which indicates that no pixel classifications for test polygons were false positive. This high false positive rate for *Microphytobentos* can be noticed by comparing the areas mapped as other bareground using OBIA that were mapped as *Microphytobentos* for FCNs. Therefore, habitat maps with OBIA were more likely to be correct for predictions of *Microphytobentos*. Other macroalgae (inc. *Fucus*) was found to be a problematic class for FCNs due to the low number of labelled pixels relative to the rest of the dataset (Figure 3.9). Confusion matrices in Figure 3.12 show that other macroalgae were often classified as *Enteromorpha* which is another algae present in Budle Bay. However, they also show that the semi-supervised results were much better than the results in the supervised setting that supports the premise in Section 3.3.1 that an unsupervised

loss term on pseudo-label segmentation maps can help datasets with a relative low number of labelled pixels. While scores show that OBIA performs better on classification of other macroalgae, Table 3.1 shows that the F1-score was lower with OBIA than FCNs that was mainly due to the OBIA low precision score. Habitat maps in Figure 3.14 show that most areas classified as other macroalgae are similar for both approaches.

The confusion matrix also shows that scores for seagrass are high for both methods. However, Table 3.1 also shows that precision figures were 0.64 and 0.27, respectively for OBIA and FCNs. This again suggests a high false positive rate for FCNs, with habitat maps in Figure 3.14 also detailing more areas mapped as seagrass with FCNs than with OBIA. Therefore, areas mapped as seagrass with OBIA were more likely to be correct than FCNs. The results for saltmarsh were in general very similar for both methods. Scores in the confusion matrix show that saltmarsh polygons was 73.32% for OBIA, and 87.78% and 81.0% for FCNs, respectively for supervised and semi-supervised settings. The F1-score was 0.84 and 0.88, respectively for OBIA and FCNNs. This suggests that OBIA was more likely to classify pixels within saltmarsh polygons incorrectly, although overall both maps present similar areas mapped as saltmarsh.

MicaSense RedEdge3 analysis

Habitat maps from the MicaSense RedEdge3 multispectral camera were found to be more correct with the FCNs than OBIA in terms of both average pixel accuracy and F1-score. The OBIA had an average accuracy and F1-score of 73.4% and 0.60, while semi-supervised FCNN had 88.7% and 0.84.

In terms of both pixel accuracy and F1-score for sediment classes, FCNs were found to perform better than OBIA. The confusion matrix for the latter method in Figure 3.13 shows that 35.82% of pixels in dry sand polygons were classified as other bareground, while Table 1 shows figures of 0.99 for precision and 0.62 for recall. This would suggest that false negative classifications for dry sand were mostly other bareground. Figure 3.13 shows FCNs

in both settings achieved scores of 98% and the semi-supervised setting had an F1-score of 0.98 that suggests that FCNs accurately mapped dry sand test polygons. However, the habitat maps in Figure 3.14 note some differences in areas mapped as dry sand for each method. In particular, supervised FCNs were found to classify larger areas as dry sand, whereas semi-supervised FCNs produced similar results to OBIA. OBIA classified 56.49% of other bareground polygon pixels as *Microphytobentos*. In Section 3.2, other bareground was noted to include wet sand, while *Microphytobentos* is a unicellular eukaryotic algae and cyanobacteria that grow within the upper millimeters of illuminated sediments, typically appearing only as a subtle greenish shading [MacIntyre et al., 1996]. This could provide some reasoning for other bareground and *Microphytobentos* being interchangeably classified with one another with OBIA. Similarly to dry sand, FCNs performed well in terms of both pixel accuracy and F1-score which suggest that other bareground polygons were classified correctly without producing many false positives.

Figure 3.13 and Table 3.1 show the scores for algal classes were higher with FCNs than with OBIA. However, both methods were in fact similar in terms of these figures, with the exception of F1-score for other macroalgae with OBIA. The confusion matrix in Figure 3.13 shows that both OBIA and FCN classifications for *Microphytobentos* exhibited poor precision. Similarly to the SONY camera, this can be noticed by large areas in Figure 3.14 being predicted as *Microphytobentos* instead of other bareground, especially for FCNs in a supervised setting. Both methods mapped *Enteromorpha* in similar areas but FCNs included classifications for *Enteromorpha* in the center and the south eastern boundary of the site, while OBIA predicted mostly seagrass and other bareground for the same stated areas. Other macroalgae class was found to have better results with FCNs over OBIA. Moreover, comparing supervised and semi-supervised models notes an increase in performance when the unsupervised loss term was added to the training algorithm which again supports the initial hypothesis that the unsupervised loss term aids FCNs with target classes that have a low number of labelled pixels relative to the remaining classes.

The remaining vegetation classes of seagrass and saltmarsh were found to have good performance with both methods, however the OBIA was found to perform better with respect to seagrass classifications. Both Figure 3.13 and Table 3.1 supported this with recall scores being lower with FCNs than OBIA. As mentioned, low recall indicates high false negative rate and interestingly all FCNs did not predict seagrass along the north western part of the site (area covered in Figure 3.11). While it is not possible to quantify which method was correct without surveying the site again, the confidence in seagrass predictions for OBIA along with FCNs predicting bareground sediment instead of vegetation can lead to users being more confident with OBIA for seagrass mapping. Both methods performed the same for saltmarsh predictions and habitat maps in Figure 3.14 show that most predicted areas were similar. However, FCNs were found to be more likely to interchangeably classify saltmarsh and seagrass that is also supported by Figure 3.13, where each confusion matrix for FCNs predicted a number of seagrass test polygon pixels as saltmarsh.

Overall analysis

In the discussion of the results for both cameras two key results were established.

The first result is that OBIA continues to be a suitable method for intertidal seagrass mapping while assessing multiple coastal features of algae and sediment within a site. Figures 3.12 and 3.13 as well as Table 3.1 reported pixel accuracy and F1-score that would suggest some degree of confidence for areas classified as seagrass with OBIA in the maps shown in Figure 3.14. Many other studies have mapped intertidal seagrass using OBIA with encouraging results [Martin et al., 2020; Ventura et al., 2018; Duffy et al., 2018; Chand and Bollard, 2021]. However, this work also attempted to make a direct comparison between FCNs and OBIA and showed that the latter outperformed the proposed method with respects to intertidal seagrass mapping. Furthermore, the provided analysis recorded accuracies for supervised classifications at a pixel-level. Some work on intertidal seagrass mapping give confusion matrices for supervised classification where accuracies reflect the percentage of segmented image-objects through multi-resolution segmentation that were

classified correctly [Ventura et al., 2018] and geographically referenced shape points [Chand and Bollard, 2021]. The work in [Martin et al., 2020] also performed an analysis of OBIA for intertidal seagrass mapping at a pixel-level, however this work also considered mapping intertidal seagrass at various density levels which adds complexity to the mapping task. In fact, seagrass mapping can also be considered as a regression problem instead of classification [Duffy et al., 2018; Perez et al., 2020]. Other work using FCNs for seagrass mapping was found in [Reus et al., 2018; Weidmann et al., 2019; Yamakita et al., 2019]. However, these studies were mainly concerned with subtidal seagrass meadows instead of intertidal seagrass. FCNs have been used for mapping intertidal macroalgae [Balado et al., 2021] with reported average accuracies for a five class problem to be 91.19%. Yet, this work considered mapping intertidal macroalgae, seagrass and sediment features at a coarser resolution.

The second key result is that although FCNs performed less well for seagrass mapping, overall results shown in Section 3.3.3 noted that FCNs had a comparable performance with OBIA in terms of average pixel accuracy and F1-score. Moreover, Figures 3.12 and 3.13 as well as habitat maps in Figure 3.14 showed that a semi-supervised setting could increase the overall performance of FCNs, reducing the need for more labelled data. This was particularly true for other macroalgae (inc. *Fucus*) that benefited the most from a semi-supervised training mode. Recent applications for semi-supervised segmentation have shown to produce state of the art results with subsets of labelled data [French et al., 2020a; Olsson et al., 2021; Kervadec et al., 2019; Perone and Cohen-Adad, 2018] that can provide alternate modelling approaches for FCNs in practical applications where labelled data is limited. Studies within remote sensing often have very limited amounts of labelled data while the recent trends show the use of weakly-supervised and semi-supervised training regimes may be utilised to overcome this problem [Wang et al., 2020c; Kang et al., 2019; Islam et al., 2020]. In particular, [Islam et al., 2020] applies adversarial training for seagrass mapping to overcome the domain shift from mapping in different coastal environments, whereas Section 3.3.1 leverages non-labelled parts of each image to produce pseudo-labels in a Teacher-Student framework.

3.3.5 Summary

Section 3.3 showed that FCNs trained from a small set of polygons can be used for segmentation of intertidal habitat maps in high resolution aerial imagery. Each FCN was evaluated in two training modes, supervised and semi-supervised, with results indicating that semi-supervision helps with segmentation of target classes that have a small number of labelled pixels. This prospect may be of benefit in studies where in-situ surveying is an expensive effort to conduct.

This section also showed that OBIA continues to be a robust approach for monitoring multiple coastal features in high resolution imagery. In particular, OBIA was found to be more accurate than FCNs in predicting seagrass for both cameras. However, as noted in Section 3.3.3, OBIA results were highly dependant on the initial parameters used for MRS, with the scale parameter being critical for image-object creation. Therefore, OBIA requires the user to understand the target class domain for a particular mapping objective in order to correlate segmented image-objects with known aerial extents of the class domain. Figure 3.11 shows the disparity of using the same scale parameter for imagery with different spatial resolutions, even though the underlying geographical area is the same. Without knowledge or prior experience of the study site, the choice of segmented map that best describes the underlying vegetation features can be a limiting factor for accurate thematic mapping.

The study site and problem formation exhibits a complex mapping exercise given that the classes listed in Section 3.2 pertinent to intertidal vegetation show similar colour and texture from an aerial point of view. This in turn can make confidence in seagrass predictions decrease as ambiguity over multiple vegetation classes increases, in particular with *Enteromorpha* sp.. OBIA was found to overcome this for both cameras accurately predicting seagrass polygons while maintaining relatively high precision when compared to FCNs. On the other hand, FCNs were found to be more accurate in classifying algae classes, in particular other macroalgae which had the least number of labelled pixels. Therefore, while this work shows that OBIA is a suitable method for intertidal seagrass mapping, other

applications in remote sensing for coastal monitoring with restricted access to in-situ data can utilise semi-supervised FCNs.

3.4 Crowdsourcing experiment for intertidal seagrass mapping

Section 3.3 describes and analyses the use of semi-supervision for FCNs. However, crowdsourcing can also bridge the gap between laborious labelling efforts by a single individual that in turn limits training data. Crowdsourcing labels can increase the volume of training data but may compromise label quality and consistency. The following section assesses the reliability of crowdsourced labels in order to provide a cost effective alternative for acquiring labels in remotely sensed environmental mapping. A crowdsourcing experiment is conducted in order to assess the statistical differences in human annotations for estuarine coastal plant species and (unvegetated) sediment. The statistical differences were evaluated using the Cochran's Q-test and the annotation accuracy of each group was examined for observation biases. Subsequently, several FCNs were trained with majority-vote annotations from each group to check whether observation biases were propagated into the FCN performance. The analysis covers two tests: first, a comparison between FCNs trained on crowdsourced annotations are compared with FCNs trained solely from in-situ data, and second, a crowdsourcing scenario is mimicked whereby in-situ reference data was supplemented with crowdsourced annotations. Discipline experts (ecologists and geomorphologists) familiar with the survey site performed better than experts with no prior knowledge of the site and non-experts. The combined dataset of in-situ annotations and crowdsourced labels from the best performing group yields a normalised average accuracy of 89.6% , while FCNs trained on the same imagery with in-situ labels achieved 87.8%.

3.4.1 Background

Section 3.3 noted the importance of intertidal seagrass for monitoring coastal ecosystem health or human made interference [Fonseca et al., 1983; Fonseca and Bell, 1998; Gera et al., 2013; Pu et al., 2014; Short and Wyllie-Echeverria, 1996; Marbà and Duarte, 2010; Duarte, 2002].

This said, the quantity and quality of data labels is a pivotal concern in many real-world scenarios because deep learning models perform best with large, labelled, training datasets [LeCun et al., 2015a; Eickhoff and de Vries, 2013]. In remote sensing, reference observations involve high logistic efforts, potential inaccuracies due to geo-location errors as well as sampling and observation bias [Congalton, 1991; Leitão et al., 2018]. Moreover, the volume of data generated with UAS imagery may cover a substantial spatially-continuous area with respect to the real-world, yet the ratio between the area covered via in-situ surveying and the total area covered in imagery is often relatively small [Bowler et al., 2020; Hobley et al., 2021a]. Methods such as transfer learning [Tan et al., 2018], data-augmentation [Shorten and Khoshgoftaar, 2019] and semi-supervision [Tarvainen and Valpola, 2017; French et al., 2019] can provide tools for FCNs to self-learn if there are limited amounts of labelled data. However, an alternative for efficient in-situ data collection is visual identification and delineation of training data directly from orthomosaics [Kattenborn et al., 2019b; Wagner et al., 2019; Lopatin et al., 2019] - possible in UAS imagery because the resolution is sufficiently high that even features as small as 10×10 cm can often be accurately identified and labelled. Further to this, the use of crowdsourced labels can provide an even more cost-effective alternative to laborious labelling procedures from aerial imagery involving individual domain specific experts. Indeed, previous work in Information Retrieval (IR) applications describes comparable performance in aggregated crowdsourced labels to expert labels [Alonso et al., 2008; Kazai and Milic-Frayling, 2009]. However, instead of delineating and assigning a meaningful value to polygons from high-resolution imagery, participants are queried whether a particular document is relevant to a topic of interest. Still, the same

concerns regarding label quality from crowdsourcing efforts in IR are echoed in supervised FCN training. Furthermore, the notion that aggregated labels can provide better quality generalisation in machine learning modelling also draws parallels with field of expert frameworks and ensemble learning [Hinton, 2002; Polikar, 2012].

Remote sensing applications have also leveraged the use of crowdsourced labels to supplement aerial imagery datasets in a variety of manners [Saralioglu and Gungor, 2020]. Commonly, web-based applications prompt participants to classify binary tasks with known GPS information for accurate geo-location. This has led to successful workflows that combine deep learning and crowdsourcing for several study sites: Guatemala, Laos and Malawi using MapSwipe [Herfort et al., 2019]; the Missing Maps humanitarian project using OpenStreetMap [Albuquerque et al., 2016]; settlements in Nigeria, Somalia, Pakistan and Afghanistan using Tomnod platform [Gueguen et al., 2016]; and for crop mapping in South East India using Plantix [Wang et al., 2020d]. Furthermore, coastal surveying has also leveraged crowdsourced annotations for deep learning applications of litter mapping in the shores of Xabelia beach in Lesvos, Greece [Papakonstantinou et al., 2021] and shoreline change mapping in two open-coast sandy beaches located within the Sydney metropolitan area [Harley et al., 2019].

However, the aforementioned studies focus on combining crowdsourced data with deep learning models on binary problem domains to avoid ambiguity for participants and erroneous labelling [Saralioglu and Gungor, 2020]. In contrast, coastal mapping requires the identification of multiple feature classes, some of that are superficially similar depending on the situation (e.g. sand and mud, seagrass and filamentous algae). The problem of training data for these types of ecosystem is tackled with a complex multi-class classification target domain for estuarine vegetation (including seagrass, saltmarsh, macro-algae) and unvegetated sediment.

3.4.2 Inter-observer variability experiment with Cochran's Q

Class distribution

The class distribution of in-situ measurements was not balanced, see Figure 3.9 which may add cognitive bias and consequently skew results in human annotations for the experiment [Eickhoff, 2018]. Recognising biases during crowdsourced data collection efforts is an important step to countering the effect these may impose on model training and is an enabling factor for algorithmic fairness [Hajian et al., 2016]. As mentioned in Section 3.2, the team of expert ecologists from Cefas and the EA sampled 108 geographically referenced tags which were reduced to a single label by selecting the class value with maximum percentage cover. These points, as shown in Figure 3.4, show the distribution of recorded tags across the study site. However, the target class distribution of recorded in-situ points was not balanced. Therefore, a set of points from the in-situ survey were combined with a set of extra points added through expert photo-interpretation in order to balance the class distribution for the experimental setup. From the original set of 108 geographically referenced tags, a balanced set of 53 points was chosen and the remaining 55 in-situ points were selected to evaluate FCN performance. For added points, the photo-interpretation was based on class dependant heuristics.

First, no extra points for dry sand were added as the set of photo-interpreted polygons covered a substantial area to generate enough points for both the experiment and FCN testing. Other bareground was a sediment class that comprised wet sediment features such as wet sand and mud. Selected points presented dark brown or gray color, rugged texture and low elevation values relative to the rest of the site. Generally, added points were sampled within a close vicinity of known in-situ records. But this was not considered as an important factor for other bareground points as long as color, texture and elevation within a 30×30cm square which corresponds to a 6×6 image crop in the RedEdge3 multispectral orthomosaics.

Similarly to the previous analysis on semi-supervised methods, vegetation classes were split into three sets: algae, seagrass and saltmarsh. The geo-location of extra points for vegetation classes was always in the vicinity of known in-situ points to establish a baseline for comparing colour and texture.

Saltmarsh points were found to be easily identifiable due to slight elevation changes in the DSM but also because coastal saltmarsh occupy the interface between land and sea [Adam, 1993]. Therefore, saltmarsh points were most present on estuary borders. Identifying points for both species of intertidal seagrass was dependant on the following texture and colour features: both species occur in mixed beds of waterlogged depressions between free-draining hummocks dominated by *Zostera noltii* and presented sparse leaves with light yellow green or green colour [Hootsmans et al., 1987; Jiménez et al., 1987; Hodges and Howe, 1997].

Microphytobenthos are microscopic organism that inhabit the upper millimetres of illuminated wet sediments, typically appearing only as a subtle greenish shading [MacIntyre et al., 1996]. Identifying extra points for microphytobenthos was only possible within very close vicinity of known in-situ points, with colour (greenish shading) used as the identifier. Extra points for *Enteromorpha* sp. had to present bright green colour while other macroalgae (inc. *Fucus*), with similar texture to *Enteromorpha* sp., was presented in a dark brownish color [Tillin and Budd, 2016; Catarino et al., 2018]. *Enteromorpha* sp. and other macroalgae were spatially continuous compared to seagrass that were more likely to be sparse. This further aided distinguishing and picking extra points for these classes. While the vegetation species may be found in other circumstances (e.g. saltmarsh hummocks can grow amongst seagrass slightly away from estuary borders), the intent was to maximise confidence that selected points were classified correctly rather than to select across the range of possible appearances for each species. Overall, an extra 54 points were added through expert photo-interpretation to maintain the class distribution balance. Therefore, the set of points to be annotated for each participant comprised 119 points where 53 points were drawn from the

in-situ survey and an extra 54 were created through photo-interpretation with the remaining 12 points being randomly selected from dry sand polygons.

Experiment setup and data analysis

The experimental population consisted of 12 participants split into three groups based on their discipline and level of expertise in habitat mapping. The experiment was analogous to crowdsourcing labelled data in remote sensing applications as participants were prompted to classify predetermined points. The experimental setup comprised two sets of points: a set where the true semantic value of each human annotation was known according to the in-situ survey described in Section 3.2, and an extra set of points created through expert photo-interpretation to balance class distribution, Section 3.4.2.

The experiment proceeded as follows: first, an inter-observer variability analysis was performed by assessing the annotations in each group using Cochran's Q test, while also reporting accuracy metrics. Second, an analysis of crowdsourced annotations of target classes for the study site was performed to assess any potential biases for each group.

Each participant was presented with a unique and random order of points to be annotated and a small set of labelled sample images representative of the vegetation classes, to assist with identification. Figures 3.15 and 3.16 respectively display the set of labelled sample images presented to each participant and the user interface available to participants during the experiment. Participants used ArcMap 10.6.1 to visualise and annotate samples. Each participant generated 119 annotations with each cell containing a semantic value corresponding to the class domain in Section 3.4.2.

The participant population was split into three groups based on their level of expertise to explore whether prior knowledge of the study site, research background and/or previous experience with marine annotation could influence experimental results. The criteria separating each group were as follows:

- Group A: expert ecologist or geomorphologist, present at the in-situ survey and/or had previous experience with annotating marine biology for the study site.
- Group B: expert ecologist or geomorphologist, but was not present at the in-situ survey and/or did not have experience with annotating marine biology for the study site.
- Group C: not an expert ecologist or geomorphologist, nor had experience with annotating marine biology from aerial imagery.

Therefore, annotations were grouped into three sets based on the stated groupings.

To evaluate the inter-observer variability within each group the Cochran's Q test was used to investigate the statistical significance of differences between K observations on the same n elements with binomial distribution [Patil, 1975; Kanji, 2006]. For this work, K series of observations corresponded to participants in a group and elements for each observation were individual annotations of participants. Therefore, the null hypothesis was that annotations for participants in a group were drawn from one common dichotomous distribution which would imply low variability in annotations. However, the Cochran's Q test states that each annotation must be dichotomous and represented as zero or one. Since the experimental annotation setup was a complex multi-class problem, each annotation was compared with the assigned label (either in-situ or photo-interpreted) and represented as one if correct, otherwise the annotation was represented as zero.

The Cochran's Q test statistic with $K-1$ degrees of freedom follows a χ^2 distribution and is given in equation 3.11.

$$Q = \frac{K(K-1)\sum_j (C_j - \bar{C})^2}{KS - \sum_i R_i^2} \tag{3.11}$$

Where, C_j is a column total, R_i is a row total, $\bar{C}$ is the average column total and S is the total score, i.e. $S = \sum_i R_i = \sum_j C_j$. In this context, a column total is the sum of correct annotations for a single participant, and a row total is the sum of correct annotations for a single point across all participants.

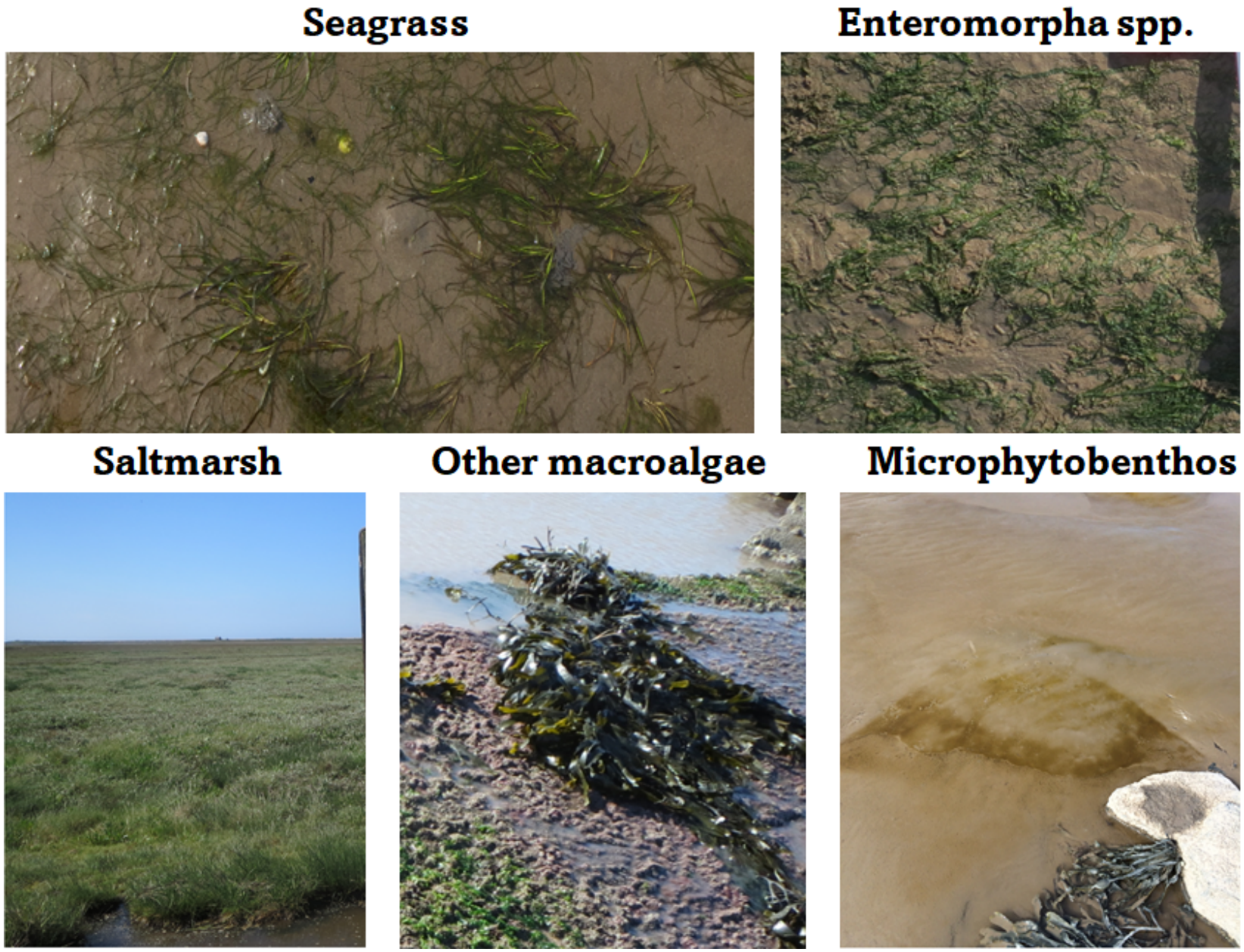

Figure 3.15: Sample images representative of vegetation classes used during the analyses. The ground photographs were taken during the in-situ survey by expert ecologists.

Figure 3.16: The user interface that was provided for participant annotations during the experiment. The user interface uses ArcMap 10.6.1 to visualise and annotate samples.

Experiment results and interpretation

Participants	Accuracy (%)	Group	Cochran Q	DoF /α	Outcome
1	70.09%	A	2.0842	4 / 0.05	Not reject
2	76.64%				
3	70.09%				
4	72.90%				
5	59.81%	B	78.8	6 / 0.05	Reject
6	27.10%				
7	63.55%				
8	61.68%				
9	75.70%				
10	63.55%	C	14.39	3 / 0.05	Reject
11	42.06%				
12	55.14%				

Table 3.2: Participant annotation accuracy and Cochran Q test statistic results. Participants were grouped into three different groups. Group A were expert ecologists or geomorphologists, present at the in-situ survey and/or had previous experience with annotating marine biology for the study site based. Group B were expert ecologists or geomorphologists that were not present at the in-situ survey and/or did not have experience with annotating marine biology for the study site. Group C were non an experts without experience in annotating marine biology from aerial imagery. The significance level, defined by the parameter α, give critical values according to a χ^2 distribution which in turn may or may not reject the statistical test.

Table 3.2 and Figure 3.17 give the results of the inter-observer experiment. The significance level for each control group was set to 5% and the degrees of freedom were set according to the number of participants in a particular group. Therefore, the critical values according to a χ^2 distribution were 9.49, 12.59 and 7.81, for participant groups A, B and C respectively.

By comparing each annotation with the known in-situ label and representing correct annotations as one and incorrect as zero, the Cochran's Q test evaluates whether annotations, which can be correct or incorrect, were drawn from the same binomial distribution. Therefore, the test statistic for a group may not allow us to reject the null hypothesis which would imply low inter-observer variability but participants in that group could collectively

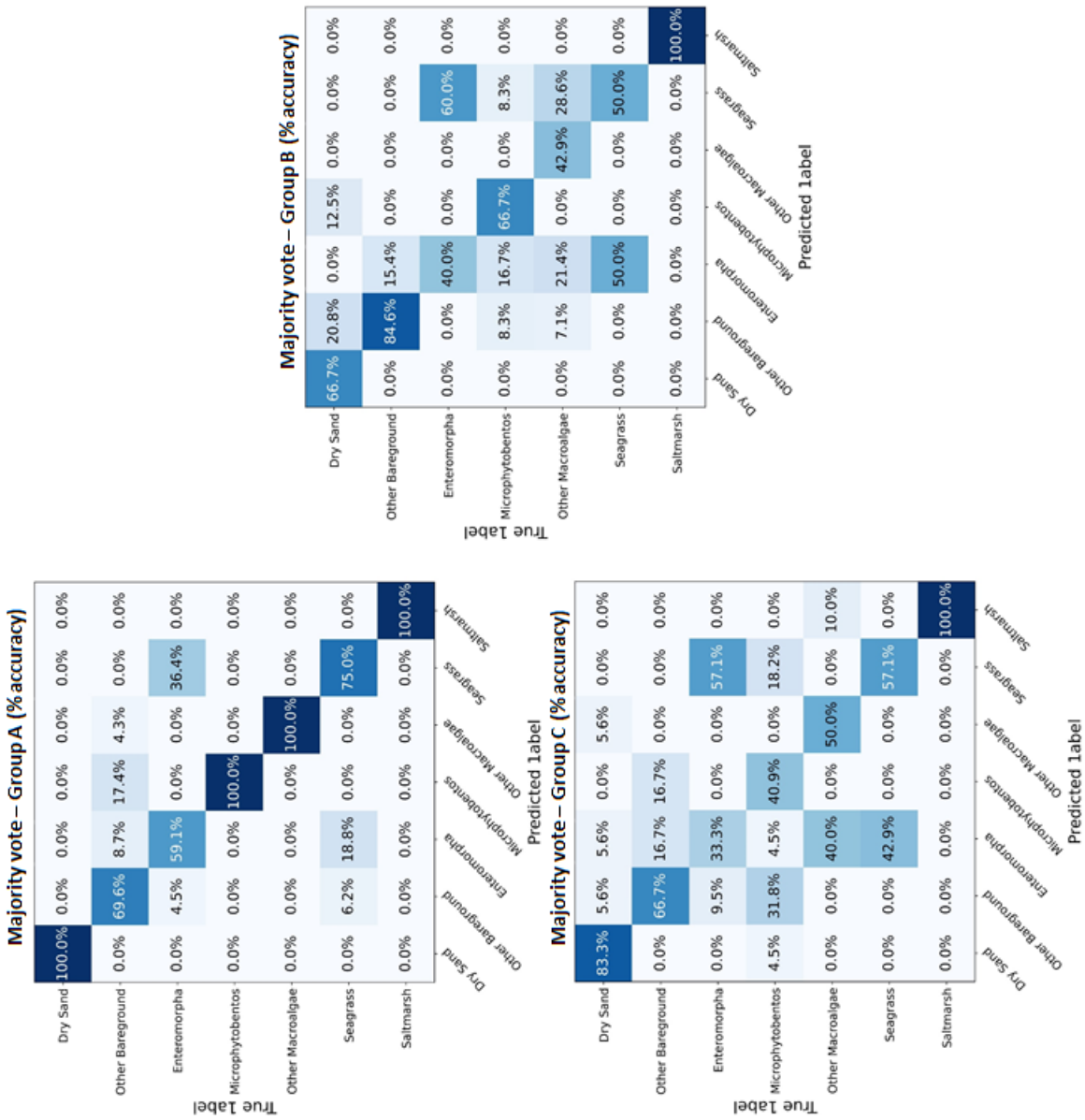

Figure 3.17: Confusion matrices for the majority vote annotations for each control group.

annotate test points incorrectly. Indeed, participants were more likely to be collectively incorrect than correct due to different incorrect annotations being represented as zero. For example, if the class label for a given point was dry sand but participants annotate the said point as other bareground and microphytobenthos, then both annotations were represented as zero that would contribute to a smaller test statistic value. Hence, the test statistic was analysed along with the annotation accuracy metrics so that emphasis was placed on groups that were collectively correct and also yielded a test statistic that does not reject the null hypothesis.

From the results in Table 3.2, the null hypothesis that participant annotations were drawn from the same distribution was not rejected only in group A. Moreover, group A also exhibited the highest mean and lowest variance in accuracy for annotations with $72.43^{+}_{-}3.10\%$ that showed that participants in group A were more likely to be correct than the other two groups. The pre-exposure of participants in group A to the target classes at the study site justified the lowest test statistic for participant annotations in this particular group. Furthermore, the latter statement can be also supported by examining the majority vote confusion matrix for group A (top-left matrix in in Figure 3.17), where the accuracy of the majority vote annotations was 81.31% for group A - higher than the highest accuracy of any participant in the experiment. This illustrates that annotations for participants in group A were better if performed collectively and as a whole group A were good candidates for crowdsourcing labels for this particular study site. Given the low variability in annotations for group A, examining Figure 3.17 also informed us about the problematic classes to annotate from aerial imagery. As mentioned in Section 3.3.4, other bareground was a sediment class composed of rock, mud and wet sand, and microphytobenthos typically appearing only as a subtle greenish shading on wet sediment [MacIntyre et al., 1996], justifies why both classes were mutually misannotated. The same reasoning can be applied to annotations for *Enteromorpha* sp. and seagrass, since both classes exhibit similar colour and texture from an aerial point of view.

The null hypothesis for participants in group B was rejected by a significant margin. This could be due to: (1) participants in this group were not familiar with annotating aerial imagery for this study site. In Information Retrieval (IR) crowdsourcing, this is also known as the ambiguity effect where missing information makes annotations appear more difficult and consequently less attractive [Ellsberg, 1961]. Alternatively, (2) the participant population contained experts from different disciplines who may have conflicting biases during annotation. If participants do not agree with each other, then the test statistic yields a high value based on whether annotations were correct or not. Specifically, the second highest overall annotation accuracy was from participant 9 while the lowest accuracy was from participant 6, both of whom belong to group B. In fact, participant 9 is a benthic ecologist with specific knowledge at identifying intertidal algae, while participant 6 is an expert in sedimentology. This contrast in discipline is reflected in annotations and subsequently in the test statistic due to correct or incorrect annotation on the same test points. The average accuracy was lower than in group A - $57.50^{+}_{-}18.16\%$ and the majority vote confusion matrix paints a similar picture - high variability and feature ambiguity lead to erroneous labelling, with an overall normalised majority-vote accuracy of 64.41% (middle-right matrix in Figure 3.17).

For participants in the final group C, the null hypothesis was also rejected, however by a smaller margin than group B. Again, this implies that participants in this group exhibit high inter-observer variability. Both the average accuracy and majority-vote accuracy were the lowest out of all groups, with $53.5^{+}_{-}10.82$ and 60.75% (bottom-left matrix in Figure 3.17) which also reflected low confidence in participant annotations. However, even with lower accuracy, participants within group C showed less variability in correct/incorrect annotations than the group B participants. This could be due to participants in group C not having any prior knowledge of the study site or with annotating aerial imagery and associating similar colour and texture based on the sample images in Figure 3.15 to the same class. The confusion matrix for group C provides insights into problematic target classes to annotate for subjects with the least experience. Algae classes, e.g. *Enteromorpha* sp.

and other macroalgae, were often mutually mislabelled, while seagrass was often annotated as *Enteromorpha* sp.. This implies that vegetation classes were hard to discern from an aerial point of view with no prior knowledge. Furthermore and similarly to group A, other bareground was also incorrectly annotated as microphytobenthos that again implies that these two classes are hard to discern from each other.

Summary

To sum up, this analysis covers three groups and assessed the inter-observer variability in participants with different backgrounds and expertise, while also assessing the accuracy of each participant, average group accuracy and majority vote accuracy. Participants in group A showed to have low inter-observer variability while also correctly annotating 81.31% of the points collectively. Participants in group B and C exhibited high inter-observer variability. Examining the criteria separating each group, having discipline expertise, prior knowledge of the site and/or previous experience annotating marine biology play an important role in minimising inter-observer variability and ensuring accurate annotation. Conversely, the lack of exposure to these criteria leads to high variability and low confidence. While the results suggest that an expert ecologist or geomorphologist without in-situ exposure produced similar overall accuracy annotations as non-experts, this was influenced by the individual accuracy result of participant 6 since the majority of participants in group B yielded a higher accuracy in annotations than two of three participants in group C. Lastly, aggregating labels based on majority-vote annotations also draw parallels with field of expert frameworks in low-level image processing and ensemble learning [Hinton, 2002; Roth and Black, 2009, 2005; Polikar, 2012]. These frameworks model high-dimensional probability distributions by taking the product of several expert distributions, where each expert works on a low-dimensional subspace that is relatively easy to model. This is similar and accurate for annotations in all groups. In general, aggregating labels showed an increase in accuracy scores of 8.88%, 6.91% and 7.25%, respectively for groups A, B and C. This alludes to the

specific and complementing nature of different research backgrounds aiding the accurate annotation.

Data pre-processing

As in Section 3.3.1, FCNs were trained with segmentation maps that contain a one-to-one mapping of pixels encoded with a semantic value. As with Section 3.3.1, segmentation maps were generated using the geographic coordinates stored in each point and converting real-world coordinates to image-coordinates. If a geographically referenced tag, corresponding to a point in Figure 3.4, resided in an non-overlapping 6000×6000 image tile, then the tile was cropped to an image size of 256×256 centered on the point. For each point, a bounding box corresponding to a square area of 30×30cm was overlaid. This was to be consistent with the area covered with quadrat sampling during the in-situ survey, whereas Section 3.3.1 detailed the use of polygons for semi-supervised methods. Images for training are similar to figure 3.7 but the polygon is replaced with a bounding box. For this work, RedEdge3 multispectral orthomosaic is used.

Fully Convolutional Networks and training parameters

Two tests with trained FCNs were performed: an initial test with several trained FCNs on different versions of labelled data based on majority-vote annotations for each group. These models are compared with FCNs trained solely with in-situ labels to evaluate whether biases in crowdsourced annotations were propagated in FCN performance. The second test mimicked a crowdsourcing scenario where in-situ reference data were supplemented with crowdsourced annotations.

Again, the U-Net architecture was found to be a suitable network. However, the encoder network is a VGG-13 [Simonyan and Zisserman, 2014] pre-trained on ImageNet. Figure 3.18 shows the network architecture used for this experiment.

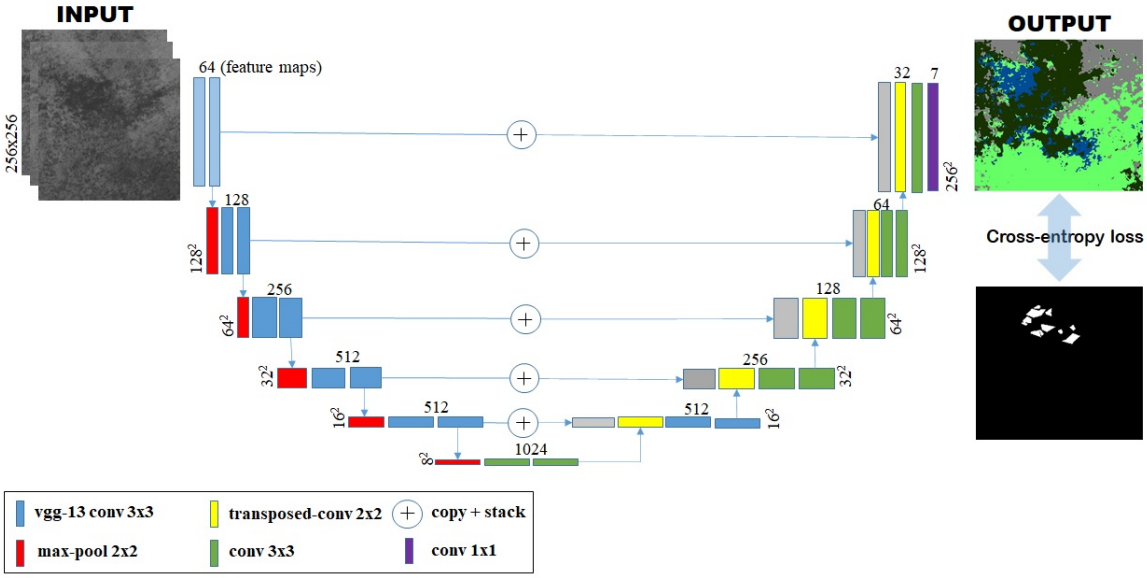

Figure 3.18: U-Net architecture and loss calculation. The encoder network, now a VGG-13, extracts feature maps. The decoder network upsamples features from the corresponding layer in the encoder path.

The loss was computed in the same manner to the steps described in Section 3.3.1. For each image, the loss was the sum of all individual pixel losses using equation 3.3 and averaged according to the number of labelled pixels in Y. The unsupervised loss described in Section 3.3.1 improves the generalisation and performance of FCNs, as shown in Section 3.3.3. However, the use of an unsupervised loss term would influence the analysis by allowing networks to adjust weights based on non-labelled parts of the image, whereas the goal is to determine the effects of aggregated crowdsourced labels. Therefore, the analysis of results does not include the unsupervised loss during model training.

During training, each image was augmented with stochastic transformations that consist of rotations up to 25° and horizontal or vertical flips. Each network was trained for 200 epochs with a batch-size of 12 with Adam optimiser. The optimiser learning rate was constant and set to 0.001. All FCNs were implemented and trained using Pytorch version 10.2.

	In-situ labels			Majority-vote group A		
	P	**R**	**F1**	**P**	**R**	**F1**
DS	0.982	0.956	0.968	0.982	0.997	**0.989**
OB	0.721	0.668	0.693	0.921	0.647	**0.76**
EM	0.433	0.769	0.554	0.517	0.738	**0.608**
MB	0.972	0.814	0.885	1.0	0.921	**0.959**
OM	0.99	1.0	**0.995**	0.982	0.809	0.887
SG	0.579	0.995	**0.73**	0.672	0.711	0.691
SM	0.928	0.944	**0.936**	0.918	0.915	0.917

Table 3.3: Precision, recall and F1 scores for models trained with in-situ labels and for models trained with majority-vote annotations from group A. DS—Dry Sand; OB—Other bareground; EM—Enteromorpha; MB—Microphytobentos; OM—Other macroalgae; SG—Seagrass; SM—Saltmarsh

3.4.3 Results and interpretation

The metrics to quantify FCNs are the same as those described in Section 3.3.2. The results in Table 3.3 and Figures 3.19 and 3.20 show better results than described in Section 3.3.2. This was mainly due to the nature of polygons used to drive the optimisation and subsequent model testing described in Section 3.4.2. The pre-processing of labels for FCN training used bounding boxes that were equivalent to the size of the quadrats used to sample the study site. In turn, this results in an easier test case given the model does not have to predict complex spatial relationships that would occur when segmenting background sediment with foreground vegetation and in essence, the results obtained with FCNs during the crowdsourcing experiment had a much simpler test set relative to the results obtained with FCNs and semi-supervised optimisation.

The evaluation consisted of two different tests: the first test shows the effects of training FCNs on different versions of labelled data based on majority-vote annotations from each group. This test evaluated whether errors in the annotation experiment were propagated to the FCN performance. For training the FCNs, the same points as in the inter-observer variability experiment used - this includes a set of 53 randomly selected points from the in-situ survey, an additional 54 points chosen through expert photo-interpretation and 12 points selected from dry sand polygons. The remaining 55 points recorded in-situ and a

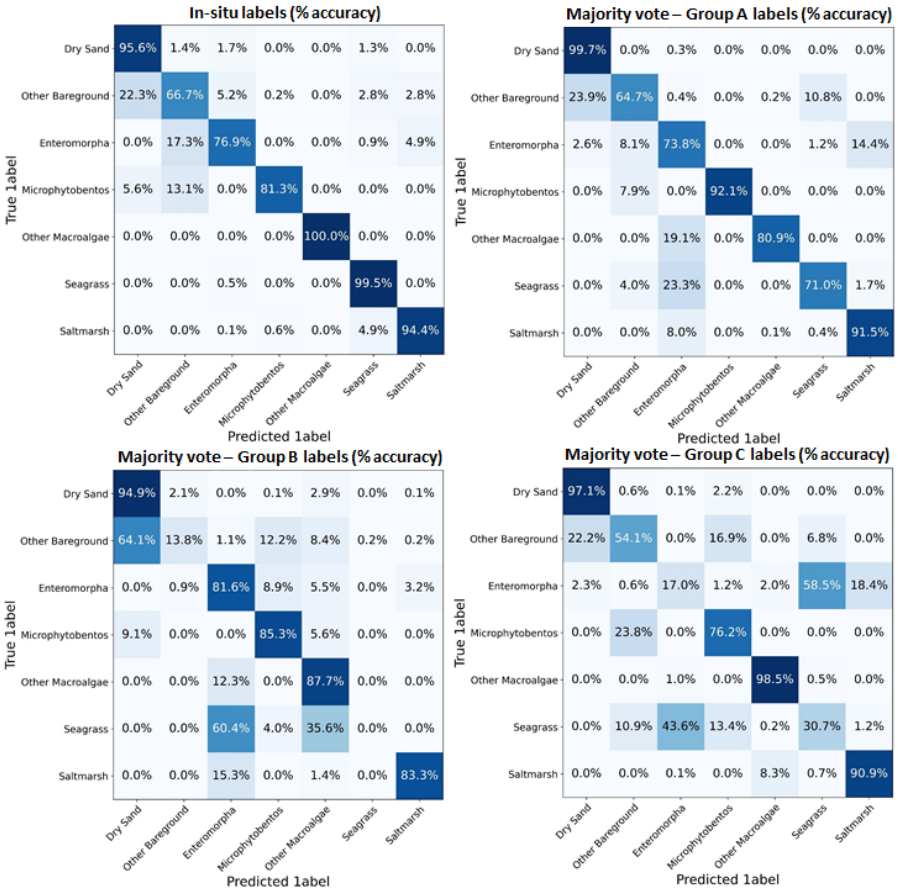

Figure 3.19: Confusion matrices for FCNN models trained using different versions of labelled data. Results for models trained on in-situ labels (top-left) and majority-vote annotations for group A (top-right), group B (bottom-left) and group C (bottom-right). The normalised percentage accuracy is shown along the diagonal of the confusion matrix

further 12 points from dry sand polygons were used for model testing . Therefore, FCNs were trained on the combined set of 119 points and the remaining 67 points comprised the test set.

For the second test, a crowdsourcing scenario was mimicked by reducing the combined training set to the same initial set of 53 randomly selected in-situ points and replacing labels for the remaining 66 points (54 from photo-interpretation plus 12 points from dry sand polygons) with majority-vote annotations from each group. The goal of the second exper-

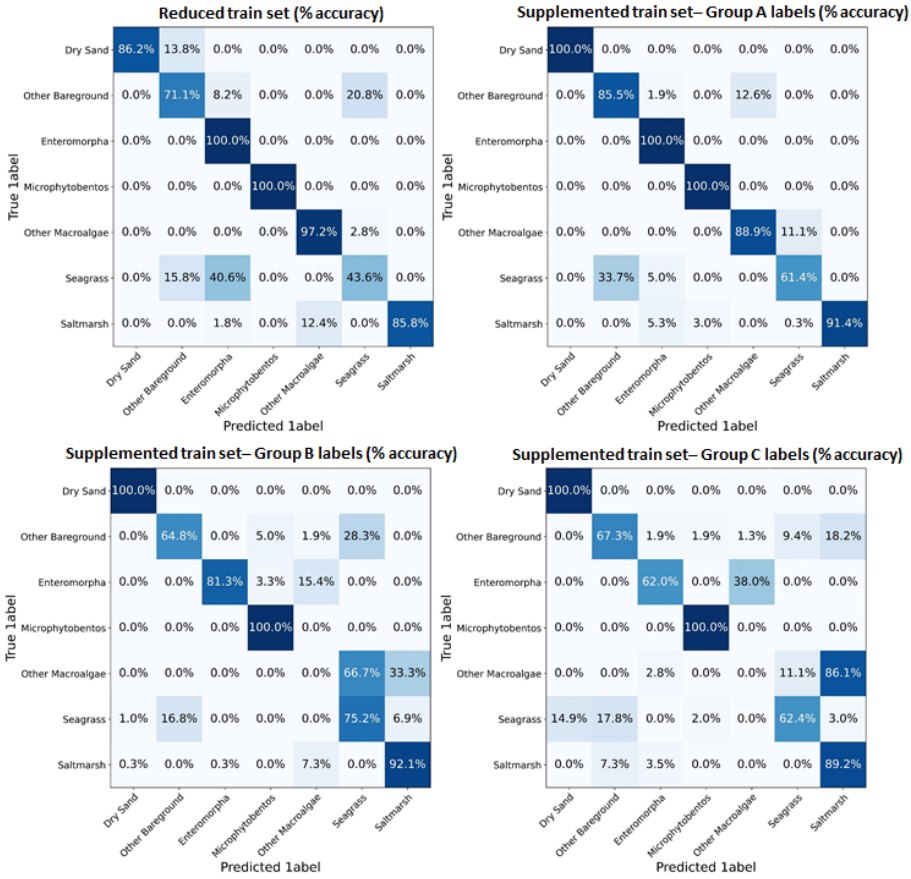

Figure 3.20: Confusion matrices for FCNN models trained using set of in-situ labels (left), and using the same in-situ set supplemented with majority-vote annotations for groups A, B and C (top-right, bottom-left, bottom-right). The normalised percentage accuracy is shown along the diagonal of the confusion matrix

iment was to determine whether supplementing a reduced training set with majority-vote annotations still achieves comparable results to models trained with in-situ labels.

Figure 3.19 shows the results of the first experiment and Table 3.3 provides further insight into class specific performance on FCNs trained with in-situ data versus FCNs trained with majority-vote annotations from group A. Figure 3.20 shows the results of training FCNs on a reduced dataset of in-situ labels versus FCNs trained on a combined train set of in-situ labels and majority-vote annotations. The confusion matrices and tabled metrics contain the average results of 5 sequential train and test runs.

Different versions of labelled data

The first test in the evaluation considered several FCNs trained with different versions of the labelled data.

First, FCNs trained with in-situ labels (top-left matrix in Figure 3.19) were viewed as the baseline for the remaining FCNs trained on majority-vote annotations from each group. The normalised accuracy with in-situ labels was 87.79% and models exhibited high confidence and accurate predictions for dry sand, other macroalgae, seagrass and saltmarsh. Other bareground proved to be a problematic class to model with a majority of predictions confused with microphytobenthos and *Enteromorpha* sp.. This paints a similar picture to majority-vote annotations for participants in group A (top-left matrix in Figure 3.17) whereby microphytobenthos was mislabelled as other bareground. However, FCNs do not mutually mislabel seagrass with *Enteromorpha sp.* that implies that FCNs were better at discerning these two specific vegetation classes than participants from group A.

The normalised accuracy for FCNs trained with majority-vote annotations from participants in group A was 81.99% (top-right matrix in Figure 3.19). As mentioned in Section 3.4.2, this particular group exhibited low inter-observer variability and accurate annotations with the exception of microphytobenthos and other bareground; which may be due to both classes being present in wet sand. Furthermore, *Enteromorpha* sp. was mutually mislabelled with seagrass because both classes showed similar colour and texture from an aerial point of view. The latter bias in annotations from participants in group A was propagated to CNN performance - where 23.3% of seagrass labels were predicted as *Enteromorpha* sp. (top-right in Figure 3.19). However, examining *Enteromorpha* sp. predictions showed that this particular class was over represented due to erroneous predictions and confusion with other vegetation classes such as saltmarsh, seagrass and other macroalgae. Therefore, erroneous labels from participants in group A caused FCNs not only to mutually mislabel *Enteromorpha* sp. with seagrass but also resulted in cascading errors for other vegetation classes due to overfitting for *Enteromorpha* sp.. Similarly to previous work using aerial

imagery for annotation, this test also showed that empirical models can compensate certain degrees of erroneous human annotations [Kattenborn et al., 2019b,a].

FCNs trained with majority-vote annotations from participants in group B yielded a normalised accuracy of 63.72% (bottom-left matrix in Figure 3.19). The analysis in Section 3.4.2 showed that annotations from subjects in group B exhibited high inter-observer variability, resulting in low confidence in majority-vote annotations. This was due to conflicting biases between experts, i.e., ecologists, geomorphologists and sedimentologists, and the ambiguity effect through lack of exposure to the in-situ survey or aerial annotation of marine vegetation species from the study site. The main trends in human annotations from this group were other bareground mislabelled as dry sand, and a general confusion of vegetation classes between *Enteromorpha* sp., other macroalgae and seagrass. These errors were also propagated into CNN performance as 64.1% of other bareground predictions were mislabelled as dry sand and seagrass was severely misclassified and predicted as *Enteromorpha* sp. and other macroalgae, respectively 60.4% and 35.6% (bottom-left matrix Figure 3.19).

The final set of majority-vote labels from group C yielded a normalised accuracy of 66.36% (bottom-right matrix in Figure 3.19). Even though the average and majority-vote accuracy for annotations provided by group C were lower than results yielded by group B - FCNs trained with majority-vote annotations from subjects in group C yielded a higher test set accuracy than majority-vote annotations from group B. The experiment also showed that participants in group C presented high inter-observer variability but by less of a margin than group B (Table 3.2 in Section 3.4.2). The analysis also showed that non-expert participants in group C exhibited low confidence predictions for other bareground with 31.8% of points labelled as microphytobenthos (bottom-left matrix in Figure 3.17). Similarly to participants in group B, they exhibited a general confusion in annotations for vegetation classes - in particular, seagrass and *Enteromorpha* sp. were often mutually misannotated. Again, these

errors in human annotations were propagated to CNN errors, e.g. mutual misclassifications for seagrass and *Enteromorpha* sp. classes.

The analysis supports the hypothesis that errors in crowdsourced human annotation were propagated into the FCN performance. All groups had a similar trend whereby annotations for microphytobenthos were mislabelled with wet sediment classes. This bias was propagated into all models trained with majority-vote annotations where other bareground was either under represented (bottom-left matrix in Figure 3.19), over represented (bottom-right matrix in Figure 3.19) or confused with dry sand (top-right matrix in Figure 3.19). The mutual mislabelling of *Enteromorpha* sp. and seagrass points for participants in group A caused the FCN to misclassify all vegetation classes as *Enteromorpha* sp.. This showed that poor annotations not only propagated errors into the CNN performance but could also cause cascading errors with classes that exhibit similar colour and texture from an aerial point of view. This stresses the need for good quality labels as FCNs optimise their weights and biases based on a non-linear one-to-one mappings between image pixels and labelled maps [Long et al., 2015]. However, results also showed that FCNs trained with low inter-observer variability and high confidence annotations, as shown with subjects in group A, can demonstrate comparable performance to the FCNs trained with in-situ labels. Conversely, training with annotations from groups B or C that manifested high inter-observer variability and higher rates of erroneous labelling, severely degraded CNN performance.

Balanced in-situ only versus crowdsourced supplemented labelled data

The second and final experiment in the evaluation considered several FCNs trained with only the balanced in-situ labels supplemented with the majority-vote annotations from each group. Therefore, the training set was the initial balanced set of 53 in-situ labels, refer to Section 3.4.2, and the labels for the remaining 66 photo-interpreted points were replaced with the semantic value of majority-vote annotations.

For comparison, a FCN was trained with just the balanced set of 53 in-situ label. This model yielded a normalised test set accuracy of 82.9% (top-left matrix in Figure 3.20). The accuracy was lower than FCNs trained with the combined full training set of 53 in-situ labels and 66 photo-interpreted labels (top-left Figure 3.19). This was expected as FCNs learn hierarchical representations of data through gradient descent [LeCun et al., 2015a], and if FCN kernel weight and bias adjustments were based on fewer image examples, then model performance and generalisation also degrades. The main affected and under represented class was seagrass where the accuracy dropped from 99.5% (top-left matrix in Figure 3.19) to 43.6% (top-left matrix in Figure 3.20).

The normalised accuracy for FCNs trained with the in-situ set supplemented with the labels from the participants in group A was 89.6% (top-right matrix in Figure 3.20) that was also the highest accuracy of all FCNs in the analysis. This setting improved the test set accuracy compared to the model trained with just in-situ labels. This was due to two reasons: first, supplementing the dataset allows for more unique samples to be incorporated into the training set, and second, the supplemented crowdsourced portion of the training set from group A exhibited low inter-observer variability and accurate annotations. Furthermore, this particular result provided an interesting comparison with the CNN trained on in-situ plus photo-interpreted labels (the top-left matrix in Figure 3.19). Both CFNs yielded satisfactory results which confirms that aggregated labels from multiple annotators within group A were as good as the efforts of a single expert annotator (lead author). This comparison also showed that in-situ efforts can be combined successfully with aerial imagery annotation that could reduce costs and labour from in-situ surveys.

The accuracy for FCNs trained using in-situ labels supplemented with the labels from participants in groups B and C were respectively 73.34% and 68.7% (bottom-left and bottom-right matrices in Figure 3.20). The analysis of both datasets was performed jointly as FCNs trained in both settings paint a similar picture. Both sets of models failed to achieve better results than models trained with just the balanced set of in-situ labels (top-left in Figure

3.20) that again stresses the need for good quality crowdsourced labels. FCNs trained with majority-vote annotations from participants in group B over represented seagrass and also misclassified all other macroalgae pixels, mostly as seagrass (bottom-left matrix in Figure 3.20). A similar outcome happened for models supplemented with the labels provided by group C - again all other macroalgae class instances are misclassified, this time mostly as saltmarsh (bottom-left matrix in Figure 3.20). In both settings this would be due to poor annotation performance from these two groups, Figure 3.17.

3.4.4 Summary

This section analysed a crowdsourcing experiment with a population of 12 participants split into three sets of groups based on discipline expertise and previous experience with either annotating aerial imagery for this study site or marine biology in general. The aims were to assess for statistical differences and biases for each group and to study the subsequent effects on CNN model performance.

The results confirmed that discipline expertise, prior knowledge of the site and/or previous experience annotating marine biology play an important role in minimising inter-observer variability and ensuring accurate annotation, and that lack of exposure to either these criteria leads to high variability and low confidence. Furthermore, the results also point to a small performance gain between annotators with expert discipline knowledge versus annotators with no previous experience in marine biology annotation or domain expertise. Participant 6 can be viewed as an outlier to the experiment given the poor annotation accuracy. However, erroneous annotations from participant 6 should not influence the confusion matrices shown in Figure 3.17 given the annotations were merged to form majority-vote annotations. Therefore, by using majority-vote annotations, individual miss annotations were suppressed and the general trends shown in the confusion matrix paint general miss classifications between target classes that exhibit similar colour and texture from an aerial point of view, i.e., separating species of algae and even separating algae from seagrass.

This work also stressed the difficulty of labelling a complex multi-class marine biology problem. Pre-exposure to the study site is important for intertidal classification,if good quality labels are to be guaranteed, and that in-situ groundtruthing may be unavoidable to prevent confusion by site experts. For instance, the general confusion between microphytobenthos with other bareground and *Enteromorpha* sp. with seagrass, Sections 3.4.2 and 3.4.3. Therefore, site surveying is necessary but may result in sparse data points with respect to the size of the coastal site. Domain experts can enhance training datasets in coastal remote sensing but domain experts present during the site survey yield the best quality labels.

Lastly, multiple FCNs were trained on different versions of labelled data based on the inter-observer experiment. The results also showed that FCNs trained with low inter-observer variability and high confidence annotations, as shown with subjects in group A, demonstrate comparable performance to the CNNs trained with in-situ labels. For the mimicked crowdsourcing scenario, whereby the balanced set of in-situ points was suplemented with crowdsourced labels, the normalised accuracy for models trained in-situ labels plus majority-vote labels from participants in group A was 89.6% (top-right matrix in Figure 3.20) that was also the highest accuracy of all FCNs in the analysis. This showed that in-situ efforts can be combined successfully with crowdsourced aerial imagery annotation which could reduce costs and labour from in-situ surveys, given that crowdsourced labels are consistent and accurate.

However, this work does not fully exclude in-situ surveying but merely affirms that good quality labels can be found in-situ but a healthy quantity of labels can also be supplemented from aerial imagery which would reduce in-situ efforts and costs.

3.5 Conclusions

Here the findings for chapter 3 are summarised.

The literature reviewed in Section 2.3.2 identified fully convolutional neural networks for semantic segmentation that provides an equivalent output to object-based methods for coastal remote sensing applications. Section 3.3 showed the utility of FCNs for a coastal remote sensing application in order to map intertidal seagrass and algae.

Furthermore, Section 3.3 also showed a comparison with OBIA. The latter continues to be a robust approach for monitoring multiple coastal features in high resolution imagery. In particular, OBIA was found to be more accurate than FCNs in predicting seagrass for both cameras. However, as noted in Section 3.3.3, OBIA results were highly dependant on the initial parameters used for MRS, with the scale parameter being critical for image-object creation. In turn, this requires the user to understand the target class domain for a particular mapping objective in order to correlate segmented image-objects with known aerial extents of the class domain. Therefore, while this work shows that OBIA is a suitable method for intertidal seagrass mapping, other applications in remote sensing for coastal monitoring with restricted access to in-situ data can utilise semi-supervised FCNs.

In essence, both methods produce equivalent outputs, but also require the same labels, or polygons, in order to drive the optimisation of machine learning models. The use of spatially explicit labels is the key factor for optimisation, and both methods require this in order to effectively learn complex relationships from features derived using orthomosaics to target classes such as intertidal seagrass and algae. Therefore, the use of FCNs can be adapted for other applications of coastal remote sensing given, as shown in Section 2.3.2, the requirements to drive the optimisation of FCNs is the same as object-based methods in a supervised setting.

However, as mentioned in Section 2.3.5, deep neural networks require substantial amounts of labelled imagery in order to effectively learn such relationships from orthomosaics to target classes, irrespective of the class domain. Hence, the use of FCNs in two training modes: supervised and semi-supervised. The results in Section 3.3.3 indicate that consistency-based semi-supervised methods improve the pixel accuracy of target classes with a low number

of labelled pixel counts. This prospect may benefit studies where in-situ surveying is an expensive effort to conduct. Another avenue to explore was to supplement the number of labels with crowdsourcing efforts in order to tackle the main drawback of FCNs. But, the feasibility of crowdsourcing labels for coastal remote sensing applications may depend on the complexity of the target class domain.

Section 3.4 analysed a crowdsourcing experiment with a population of 12 participants split into three sets of groups based on discipline expertise and previous experience with aerial imagery annotation. The results of the experiment confirmed that discipline expertise, prior knowledge of the site and/or previous experience annotating marine biology play an important role in minimising inter-observer variability and ensuring accurate annotation, and that lack of exposure to either of these criteria leads to high variability and low confidence. This experiment also stressed the difficulty of labelling a complex multi-class marine biology problem. Therefore, pre-exposure to the study site is important for intertidal classification, if good quality labels are to be guaranteed and that in-situ groundtruthing may be unavoidable to prevent confusion by site experts. For instance, the general confusion between microphytobenthos with other bareground and *Enteromorpha* sp. with seagrass, Sections 3.4.2 and 3.4.3.

The results from Section 3.4 also showed that FCNs trained with low inter-observer variability and high confidence annotations, as shown with subjects in group A, demonstrate comparable performance to FCNs trained with in-situ labels. This is further confirmed during the crowdsourcing scenario, where a combination of the balanced set of in-situ points supplemented with crowdsourced labels were used to train FCNs, resulted in the highest accuracy of all FCNs in the analysis. This showed that in-situ efforts can be combined successfully with crowdsourced aerial imagery annotation which could reduce costs and labour from in-situ surveys, given that crowdsourced labels are consistent and accurate. However, this work does not fully exclude in-situ surveying but merely confirms that good quality

labels can be found in-situ but a healthy quantity of labels can also be supplemented from aerial imagery which would reduce in-situ efforts and costs.

To sum up, the use of FCNs can provide an alternative tool for coastal remote sensing applications given the requirements for optimisation are the same as object-based methods. However, FCNs require substantial amounts of labelled imagery that was tackled in this Chapter with the use of semi-supervision and crowdsourcing.

4 Sizewell - hyperspectral reconstruction and multi-task learning

4.1 Introduction

The research for this chapter was focused on Sizewell, Suffolk, England (55.207°N, 1.602°W). The coastal site is a narrow shingle beach with a mixture of shingle, strandline and sand-dune communities.

Coastal vegetated shingle is a rare and declining habitat worldwide that is found around the UK coastline [Randall, 2004]. Understanding germination characteristics of shingle beach species can aid ecologists understand and restore vegetation in shingle environments [Walmsley and Davy, 1997a,b]. An alternative route is the use of remote sensing and accurate thematic mapping in order to understand the underlying phenology of species belonging to shingle, strandline and sand-dune communities. Therefore, the main goal in this study was to map the pioneering marine species from these communities on shingle and sand sediment with limited amounts of labelled data. The in-situ survey to the Sizewell study site also collected hyperspectral measurements of vegetation and sediment features from shingle, strandline and sand-dune communities.

The following chapter continues to leverage consistency-based semi-supervised segmentation, an also provides two alternative methods to optimise deep learning models using multi-task learning (MTL) to incorporate the hyperspectral measurements.

Section 4.2 details the ecological importance of shingle, strandline and sand-dune communities present at the Sizewell study site. This section also shows collected data and imagery

from the in-situ survey and details the mapping objective for the results described in Section 4.5.

4.2 Sizewell study site: background and rationale

As mentioned in Chapter 1, beaches and open shore coastal environments provide essential ecosystem services, such as natural buffering of inland areas from the damaging impacts of waves and elevated water levels during storm events [Splinter and Coco, 2021]. The coastal zone for the Sizewell study site, see Figure 4.1, includes species communities that belong to the SD national vegetation class (NVC). In Britain, the NVC describes plant communities from natural, semi-natural and common artificial habitats and classifies them into distinct categories [Rodwell and nature conservation committee , GB] - where the SD NVC identifier refers to shingle, strandline and sand-dune communities. In particular, the study site had species and assemblages of SD1, SD2, SD6, SD7, SD10, SD11 and SD19 NVCs. The guideline described in Rodwell and nature conservation committee [GB] also states that each NVC identifier can be further expanded to include sub-communities of other species. For instance, SD1 defines a shingle community with Rumex crispus and Glaucium flavum species, and an extension to SD1 is SD1B which also defines a shingle community for the same species with sub-communities such as Lathyrus japonicus and Crambe maritima. Table 4.1 lists the mentioned NVCs with a general description of species belonging to each SD assemblage. Figure 4.2 shows a close-up of the orthomosaic shown in Figure 4.1. In this close-up, the Eastern part closest to the sea is dominated by shingle and sand sediment. Then, transitioning from East to West along the orthomosaic, the pioneering species on shingle or sand sediment mainly belong to SD1 and SD2 NVCs and the tall grassland communities after the pioneering species mainly belong to SD6 and SD7 NVCs.

The species present in the Sizewell study site belonging to the NVCs listed in Table 4.1 pose a challenge due to the variable and short-lived nature of these species on shingle sediment [Fuller and Randall, 1988; Scott, 1963; Fuller, 1987; Walmsley and Davy, 1997a]. Shingle

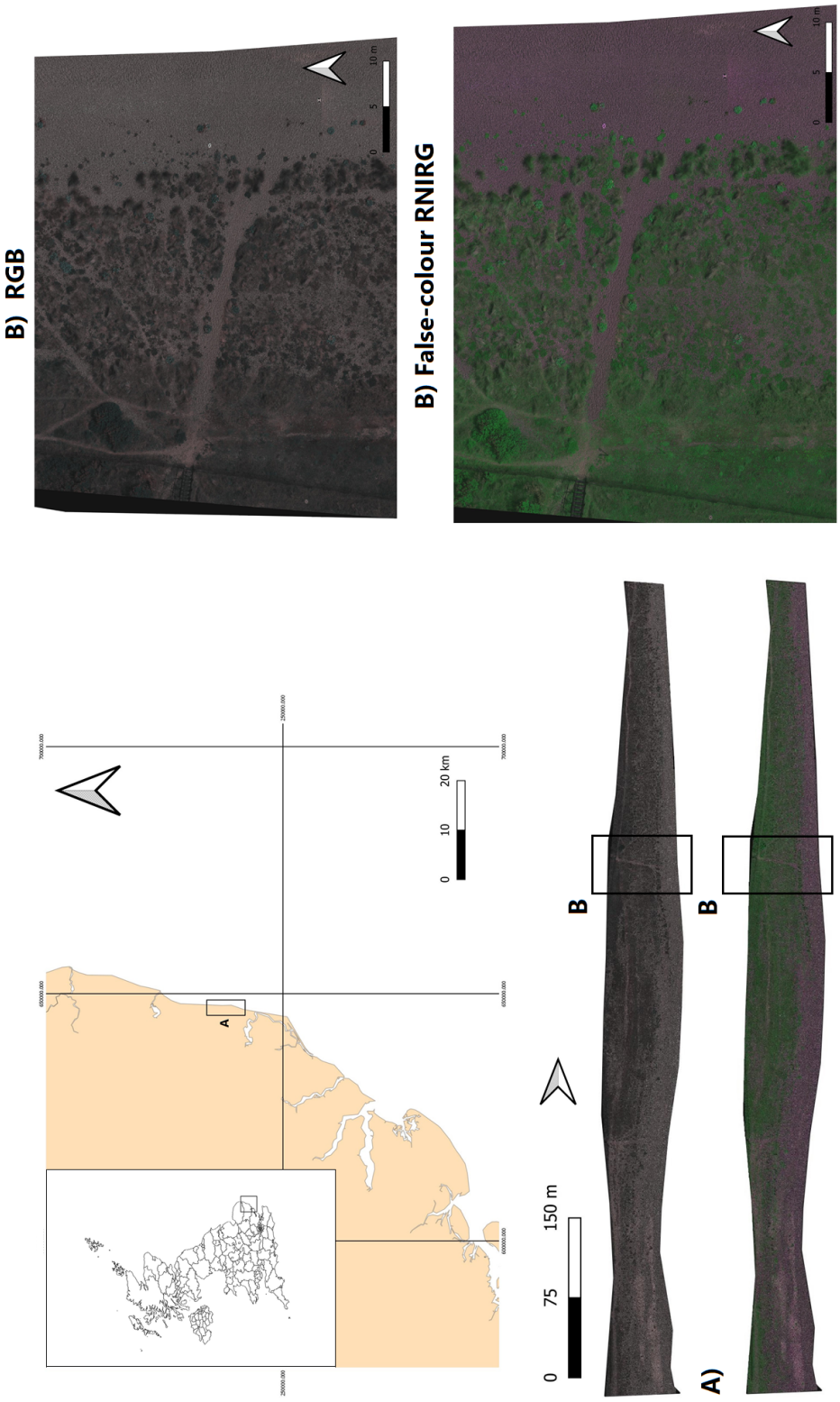

Figure 4.1: Study site within the U.K. (top-left). Ortho-registered images of Sizewell using the Sentera multispectral sensor. On the right there is a close-up using Red, Green and Blue channels as well as a false-colour RGB using Red, Near-Infrared and Green channels.



Figure 4.2: A close-up of the orthomosaic shown in Figure 4.1. The close-up shows that the Eastern part closest to the sea is dominated by shingle and sand sediment. Then, transitioning from East to West along the orthomosaic, shows the pioneering species on shingle or sand sediment mainly belong to SD1 and SD2 NVCs and the tall grassland communities after the pioneering species that mainly belong to SD6 and SD7 NVCs. Red dots on the orthomosaic correspond to in-situ geo-referenced tags with known classifications of shingle vegetation species.

NVC	Species present in community
SD1	Rumex crispus - Glaucium flavum shingle community
SD2	Honkenya peploides - Cakile maritima strandline community
SD6	Ammophila arenaria mobile dune community
SD7	Ammophila arenaria - Festuca rubra semi-fixed dune community
SD10	Carex arenaria dune community
SD11	Carex arenaria - Cornicularia aculeata dune community
SD19	Phleum arenarium - Arenaria serpyllifolia dune annual community

Table 4.1: Various NVCs present at the Sizewell study site.

beaches represent continuously varying environments due to beach composition and shingle mobility, as well as interaction with other factors, such as topography, climate and water supply. The variability in sediment features dictates the variability in vegetation habitats from these particular NVCs [Scott, 1963]. Furthermore, the work related to the study site described in Walmsley and Davy [1997a,b] requires laborious in-situ surveys and domain expertise for species to be accurately identified which is time-consuming and an expensive task. In order to create and act on conservation plans, accurate and efficient thematic mapping is necessary in order to detect changes to the shoreline communities present at Sizewell from external factors, such as climate change.

As such, the main goal was to circumvent laborious in-situ surveys and map the species that compose assemblages belonging to SD1, SD2 and SD6 NVCs using processed VHR orthomosaics and deep learning models. The focus to map these particular assemblages was two-fold: during the site survey, the relative abundance of species that form these assemblages stated in Table 4.1 were greater than other NVCs, and the delicate existence of these species due to shingle and tidal processes can provide an indicator to extrinsic factors, human or natural, affecting the overall ecosystem health [Walmsley and Davy, 1997a,b].

Section 4.2.1 details the in-situ survey and data collected from the study site, and defines the mapping objective for the results shown in Section 4.5.

4.2.1 Data collection and in-situ survey

The following sections describe the data collected during the site survey. One ground and aerial survey were conducted in August 2020. The author was present at the ground survey. The main objectives of the ground-survey were to identify species at the Sizewell study site that assemble to SD1, SD2 and SD6 NVCs and collect a database of hyperspectral measurements with a portable spectroradiometer. The aerial survey had the objective of collecting overlapping VHR imagery which as discussed in Section 3.2, can create VHR orthomosaics using Structure from Motion (SfM) techniques [Cunliffe et al., 2016].

Ground survey

The ground-survey was conducted by the Cefas and University of East Anglia on the 10th of August 2020. The intent of this survey was to collect three sets of data:

1. a set of hand-gun hyperspectral measurements with a portable spectroradiometer
2. corresponding captures with a DSLR (NIKON D5100) and a X-Rite colour checker [Pascale, 2006]
3. a RTK GPS logs for accurate transcription of in-situ data onto generated VHR orthomosaics

The vegetation classes recorded for this ground survey belong to SD1B, SD2, SD6A, SD7C and SD19 NVCs. As mentioned, the SD1B NVC is an extension to SD1 that includes Lathyrus japonicus and Crambe maritima sub-communities, SD6A extends SD6 to include Elytrigia juncea and Honckenya peploides sub-communities and the SD7C NVC details Ammophila arenaria-Festuca rubra semi-fixed dune communities with Ononis repens sub-communities. Lastly, SD19 represents seasonal species from Phleum arenarium-Arenaria serpyllifolia dune annual communities [Rodwell and nature conservation committee , GB]. The following lists the communities identified during the ground-survey for each NVC:

- SD1B - Lathyrus japonicus, Glaucium flavum, Silene uniflora, Crambe maritima, Senecio squadilus, Leontodon saxatilis, and Rumex crispus
- SD2 - Honckenya peploides, Atriplex prostastes, Eryngium maritimum, Silene uniflora and Crambe maritima
- SD6A – Ammophila arenaria, Elytria juncea, Senecio jacobea, Senecio viscosus and Honckenya peploides
- SD7C – Ononis repens, Phleum arenarium, Ammophila arenaria, Carex arenaria, Leontodon saxatilis, Eryngium maritimum and Taraxacum officialis
- SD19 – Moss and Lichen

For each data entry during the in-situ survey, ten, fifteen or twenty five hyperspectral measurements were recorded at different locations of the marine species or unvegetated sediment. The choice of fifteen or twenty five recordings was based on the area covered by the recorded data entry. Figure 4.3 shows plots of collected HS measurements using the hand-gun instrument. Each species plot was an average of all collected HS measurements during the survey. Certain bandwidths (1300-1400nm, 1800-1950nm and 2450-2500nm) have radiometric distortions that could be caused by temperature and humidity [Hueni and Bialek, 2017]. However, these bandwidths were not part of the analysis in Section 4.4. Table 4.2 gives details of recorded samples during the site survey. Along with the hyperspectral measurements was a corresponding photograph of the identified species captured with a DSLR (NIKON D5100) and a X-Rite colour checker [Pascale, 2006] and a RTK GPS log which similarly to Section 3.3.1 was used to create geographically referenced polygons and subsequent masks.

The recorded HS samples with the hand-gun instrument allow natural day-light to be included in the measurement. To calibrate the sensor before each sample, a white-point sample was measured in order to record a day-light illuminant spectrum. Then, the illu-

minant spectrum was removed from each measurement so that HS measurements capture intrinsic reflectance properties.

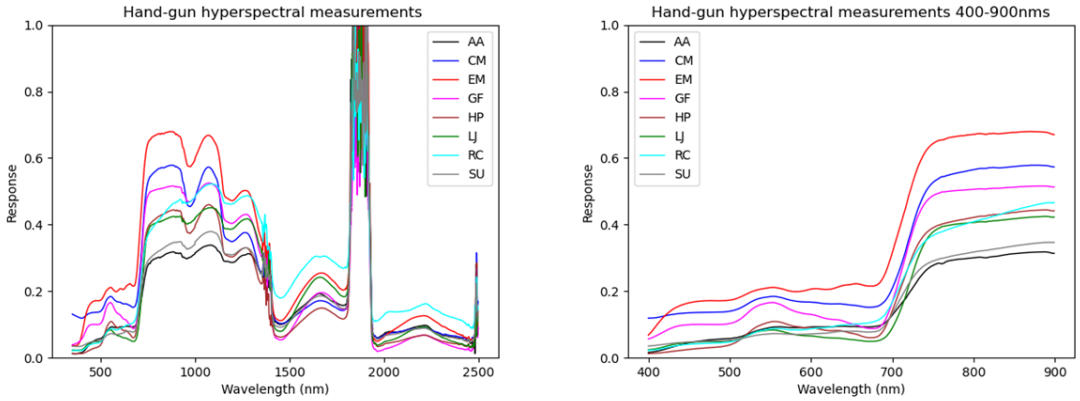

Figure 4.3: Plots of reflectance measurements using the hand-gun. The left plot shows the entire bandwidth of the FieldSpec4 (400-2500nm) and the right plot has the same bandwidth as the Sentera multispectral camera (400-900nm). Each species plot is an average of all collected hand-gun measurements during the survey. AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SU- Silene uniflora

Aerial survey

The aerial survey was performed with a rotor-based UAS using a Sentera multispectral camera with five narrow band filters for for Red (620-690 nm), Green (510-590 nm), Blue (450-500 nm), Red Edge (735-750 nm) and Near Infra-red (840-870 nm) channels with a ground sampling distance of approximately 1cm (Figure 4.1, top right and bottom right). Figure 4.4 (right-plot) shows the sensor response function for wavelengths between 400-900nm.

Very high resolution orthomosaics of Sizewell were created using Pix4D [Cubero-Castan et al., 2018] and SfM techniques. Similarly to Section 3.2, the resulting VHR was orthorectified using GPS logs of camera positions and ground control markers spread out across the site ensuring the mosaic was well aligned with the GPS logs of the ecological features sampled during the in-situ survey. The multispectral orthomosaic had 101,618×6,822 pixels

Table 4.2: Recorded species from SD1B, SD2, SD6A, SD7C and SD19 NVCs along with the number of recorded plants per species, number of HS samples and corresponding DSLR captures

Class	# recorded plants	# of HS samples	# of DSLR captures
Ammophila arenaria	5	75	15
Arctium minus	1	15	3
Beta vulgaris spp. Maritima	4	100	12
Crambe maritima	5	125	15
Elytrigia juncea	1	10	3
Eryngium maritimum	4	60	12
Glacium flavum	4	100	12
Honckenya peploides	3	45	9
Lathyrus japonicus	4	100	12
Lichen	1	15	3
Leontodon saxatalis	1	25	3
Moss	1	10	3
Ononis repens	3	50	9
Phleum arenarium	1	10	3
Rumex crispsus	6	100	18
Sand	2	20	3
Shingle	2	20	3
Senecio squadilus	2	30	6
Silene uniflora	5	125	15
Senecio sylvaticus	2	15	6

in five image bands with 1cm spatial resolution. For ease of processing, each orthomosaic was split into 3,000×3,000 non-overlapping images along with geographic information to be used for further processing. The orthomosaic was split into 102 tiles. Figure 4.1 shows the study site and its geographical position in England and Figure 4.2 shows a close-up of the study site with brightness adjusted for viewing purposes.

Mapping objectives

The mapping objective for the methods described in Section 4.5 focus on a reduced target class domain for species that belong to the SD1B, SD2 and SD6A NVCs. The reduced class domain was based on three heuristics: the number of recorded plants for a particular species during the in-situ survey had to be greater than three (see Table 4.2), the recorded vegetation species had to cover an area greater than 20×20cm and each class had to be paired with hyperspectral measurements from the spectroradiometer positioned with RTK

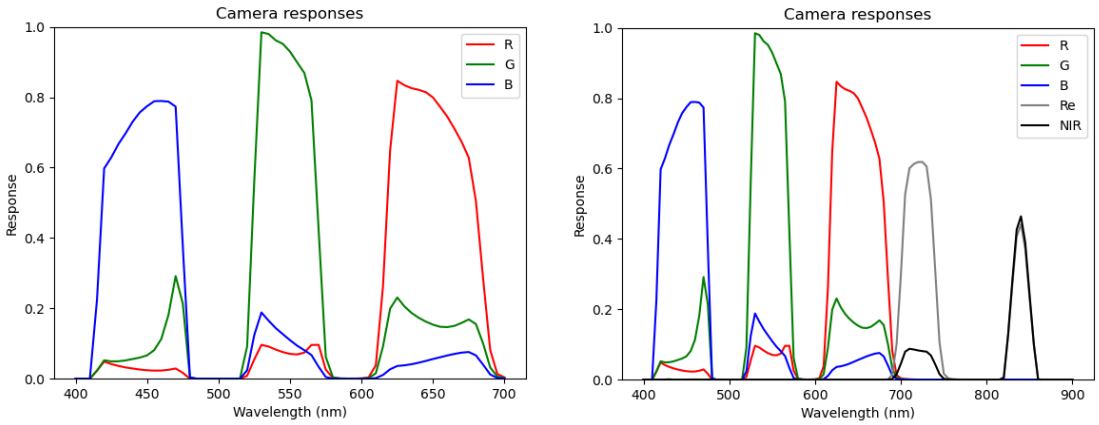

Figure 4.4: Sentera camera responses for RGB colour space (400-700nm) and RGB + NIR colour space (400-900nm).

GPS logs. Therefore, the class domain for the segmentation task in Section 4.5 using deep learning models is as follows:

- SD1B - Lathyrus japonicus
- SD1B - Glaucium flavum
- SD1B - Rumex Crispus
- SD1B and SD2 - Silene uniflora
- SD1B and SD2 - Crambe maritima
- SD2 - Honckenya peploides
- SD2 - Eryngium maritimum
- SD6A – Ammophila arenaria

The class domain had eight vegetation classes and also included two unvegetated sediment classes: shingle and sand. Figure 4.5 shows the reduced class domain for vegetation species with ground captures using the DSLR (NIKON D5100) and a X-Rite colour checker [Pascale, 2006].

Cefas provided the results for the OBIA method described in Section 4.5.3. The results with OBIA were obtained with a further reduced target class domain [Private communication, Arosio, 2021]. These classes were Ammophila arenaria, Crambe maritima, pioneering grassland, young pioneering species, sand and shingle. Pioneering grassland and young pioneering species were target classes that join multiple species from SD1 and SD2 NVCs.

- Pioneering grassland: SD1 species, such as Rumex crispus and Glaucium flavum
- Young pioneering species: SD1 and SD2 species, such as Silene uniflora, Lathyrus japonicus and Honckenya peploides.

The deep learning models and OBIA classified the orthomosaic of the study site with different target class domains which creates a discrepancy between the objective and visual results described in Section 4.5 and the listed species in Table 4.2. Therefore, a further analysis was shown to provide an equal comparison of methods by merging predictions of deep learning models in the same manner as Cefas did with OBIA.

Furthermore, the target class domain for both methods shown in Section 4.5.3 do not encompass the entire bio-diversity of vegetated shingle species found at the study site. In particular, Ammophila arenaria, as shown in Figure 4.5, forms stiff and hardy stems that can grow up to 1.2 metres, and is mostly present on Western parts of the orthomosaic, as shown in Figures 4.1 and 4.2. However, these parts of the orthomosaic are predominately represented by SD6 and SD7 NVCs that also represent other grass species such as Festuca rubra.

4.2.2 Outline

Given the collected data and target class domains defined in Section 4.2.1, the following sections focus on developing methods that can leverage collected hyperspectral measurements and incorporate these into the optimisation of deep learning models for semantic segmentation tasks.



Figure 4.5: Shingle vegetated species that belong to SD1, SD2 and SD6 NVCs. These species were sampled and classified by an expert taxonomist during the in-situ survey at Sizewell. For each sample, a hyperspectral measurement, along with a photograph and a RTK GPS measurement were used for further processing. Each image was captured in either the first or second ground-survey.

First, Section 4.3 shows methods for hyperspectral reconstruction from RGB images. The shallow method described in Arad and Ben-Shahar [2016] was compared with deep learning models tailored for hyperspectral reconstruction [Shi et al., 2018]. Furthermore, deep learning models were optimised in different training modes, and in particular the discussion in Section 4.3.3 shows that self-supervised deep learning models can be developed under the assumption that there exists many hyperspectral physically plausible metamers for the same input image [Morovic and Finlayson, 2006]. The experiments also investigated extending the proposed architecture in Shi et al. [2018] to a U-Net architecture. These methods were evaluated on the ICVL challenge [Arad and Ben-Shahar, 2016].

Then, given the deep learning models developed in Section 4.3 and the hyperspectral measurements described in Section 4.2; Section 4.4 shows two methods for hyperspectral reconstruction from multispectral aerial imagery of the study site. Section 4.2.1 detailed collected hyperspectral measurements data from the site survey that show intrinsic reflectance measurements for SD1, SD2, SD6 and SD7 NVCs. This poses a challenge for evaluating the methods developed in Section 4.3, since collected hyperspectral measurements discount the illuminant spectrum, yet physically plausible hyperspectral radiance reconstruction needs to account for the illuminant in the scene, as shown in equation 2.2. Therefore, one method trains deep learning models in a supervised setting to learn hyperspectral reflectance reconstruction, and the self-supervised method in Section 4.3 instead reconstructs radiance in accordance to equation 2.2.

Sections 4.3 and 4.4 investigate extending the architecture described in Shi et al. [2018] to a U-Net architecture. Given the results in Section 4.4 with the U-Net architecture were satisfactory, Section 4.5 uses a multi-task learning framework with a shared model to learn both semantic segmentation and hyperspectral reconstruction. As mentioned in Section 2.5, the premise for adding an auxiliary image task is to prevent overfit on the main image task which in this case was mapping the pioneering marine species from the stated NVCs in Section 4.2.1. The auxiliary image task (hyperspectral reconstruction) causes shared

models to prefer new hypothesis in higher-dimensional space. This in turn can prevent models trained for semantic segmentation to overfit and seek better solutions [Ruder, 2017]. An analysis of models trained in supervised, semi-supervised, as described in Section 3.3, MTL settings and OBIA was provided, both objectively and through visual subjective analysis.

4.3 Hyperspectral reconstruction on ICVL

As mentioned in Section 2.4, the ICVL challenge is a hyperspectral imaging dataset that contains various scenes and objects with the objective of developing hyperspectral reconstruction solutions from RGB imagery [Arad and Ben-Shahar, 2016]. The database contains 200 images that were acquired using a Specim PS Kappa DX4 hyperspectral camera and a rotary stage for spatial scanning. Images are captured at 1392×1300 resolution over 519 spectral bands for a bandwidth of 400-1,000nm at 1.25nm increments. The number of spectral bands causes data volume constraints for image loading and subsequent model training, therefore available images are downsampled to 31 spectral bands from 400-700nm at 10nm increments. Each image pixel represents a continuous spectral measurement of the underlying object across 31 spectral channels. Figure 4.6 displays a gallery of images from ICVL along with a corresponding projection using the Sentera multispectral camera responses (Figure 4.4 left-plot) to RGB colour space. For the next sub-sections the following hyperspectral reconstruction methods were evaluated:

- sparse-dictionary representation of HSI [Arad and Ben-Shahar, 2016]
- a deep learning implementation with the the HSCNN-R [Shi et al., 2018]

The HSCNN-R was also extended to a U-Net architecture with added pooling operations and a subsequent upsample track in the network design. The images were randomly split to a 70/30 ratio, respectively for training and testing. Images from the train set that include a colour checker were removed and added to the test set for evaluation, Section 4.3.3.

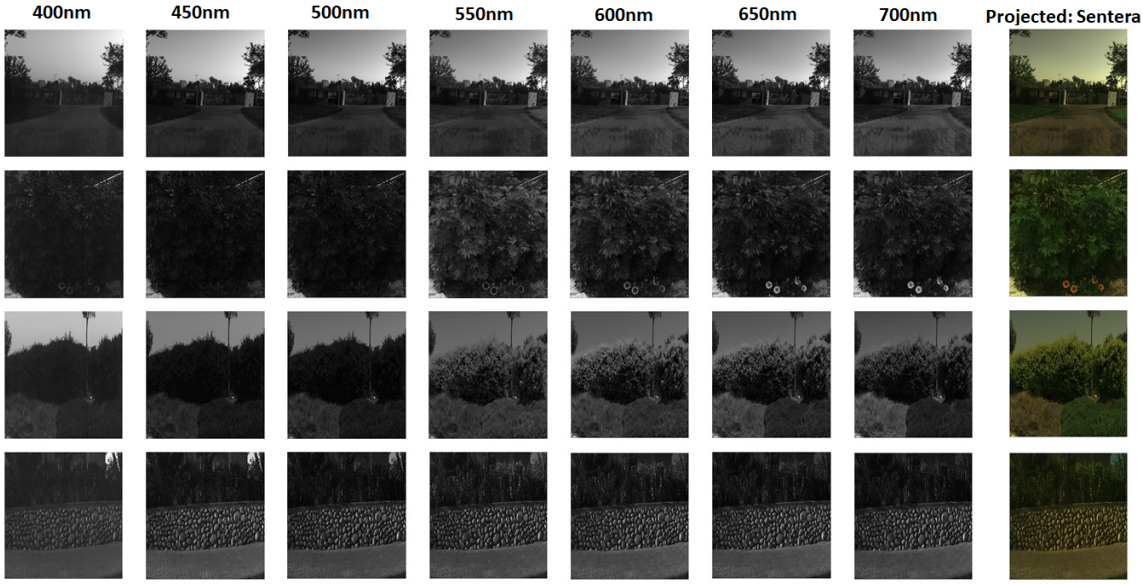

Figure 4.6: A gallery of sample images and various outdoor scenes in the ICVL dataset. The images represent a particular hyperspectral channel in the visible spectrum between 400-700nm. The images projected to the Sentera multispectral colour space integrate the continuous hyperspectral signature captured with Specim PS Kappa DX4 with the filters shown in Figure 4.4 (left-plot).

4.3.1 Sparse-dictionary representation for HSI reconstruction

The method presented in Arad and Ben-Shahar [2016] can be broken down into two steps.

The first part was to create a prior sparse-dictionary representation from HSI. As mentioned in Section 2.4, HSI exhibits sparse encoding and spectral information that can be expressed as a sparse combination of basis spectrum [Chakrabarti and Zickler, 2011], as shown in equation 2.4. Therefore, the goal was to create a dictionary representation such that each key represents a basis-spectrum and the value is the relative abundance of each basis spectrum. The first step was to collect a rich hyperspectral prior by randomly selecting pixels from the HSI dataset. Then, the prior was reduced computationally to a dictionary of hyperspectral signatures (D_h) using K-SVD [Aharon et al., 2006].

$$D_h = \{\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_n\} \tag{4.1}$$

The hyperspectral dictionary can be projected to RGB colour space (D_{rgb}) using the Sentera sensor responses (S) from Figure 4.4 (left-plot). These projections were expressed as the inner product of the sensor response matrix S with D_h.

$$D_{rgb} = \{\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_n\} = D_h.S \tag{4.2}$$

Where, $\boldsymbol{c}_i = (\boldsymbol{c}_r, \boldsymbol{c}_g, \boldsymbol{c}_b)^T$ such that:

$$\boldsymbol{c}_i = \boldsymbol{h}_i.S \,\forall\, \boldsymbol{c}_i \in D_{rgb} \tag{4.3}$$

Given the prior hyperspectral dictionary, the second step was to reconstruct a hyperspectral image given an input image in RGB colour space. For each test pixel, $\boldsymbol{c}_i = (\boldsymbol{c}_r, \boldsymbol{c}_g, \boldsymbol{c}_b)^T$ in the RGB image, an intermediate weight vector $\boldsymbol{w}$ was computed using orthogonal match pursuit (OMP) [Pati et al., 1993]. The sparsity imposed on the dictionary D_h during K-SVD needs to match the sparsity imposed on the OMP.

$$\boldsymbol{c}_i = \boldsymbol{h}_i.S = D_{rgb}.\boldsymbol{w} \implies \boldsymbol{h}_i = D_{rgb}.\boldsymbol{w}.S^{-1} \tag{4.4}$$

The underlying hyperspectral structure $\boldsymbol{h}_i$ projected to $\boldsymbol{c}_i$ was estimated using the linear combination of basis functions found in D_h. The accuracy of $\boldsymbol{h}_i$ generated the pixel $\boldsymbol{c}_i$ depends on the representational power of the dictionary.

$$\boldsymbol{h}_i = D_h.\boldsymbol{w} \tag{4.5}$$

Implementation details

The method was implemented in MATLAB using standard toolboxes for K-SVD [Aharon et al., 2006] and OMP [Pati et al., 1993]. Images projected to RGB colour space use the sensor response functions from a miniaturised Sentera multispectral camera (Figure

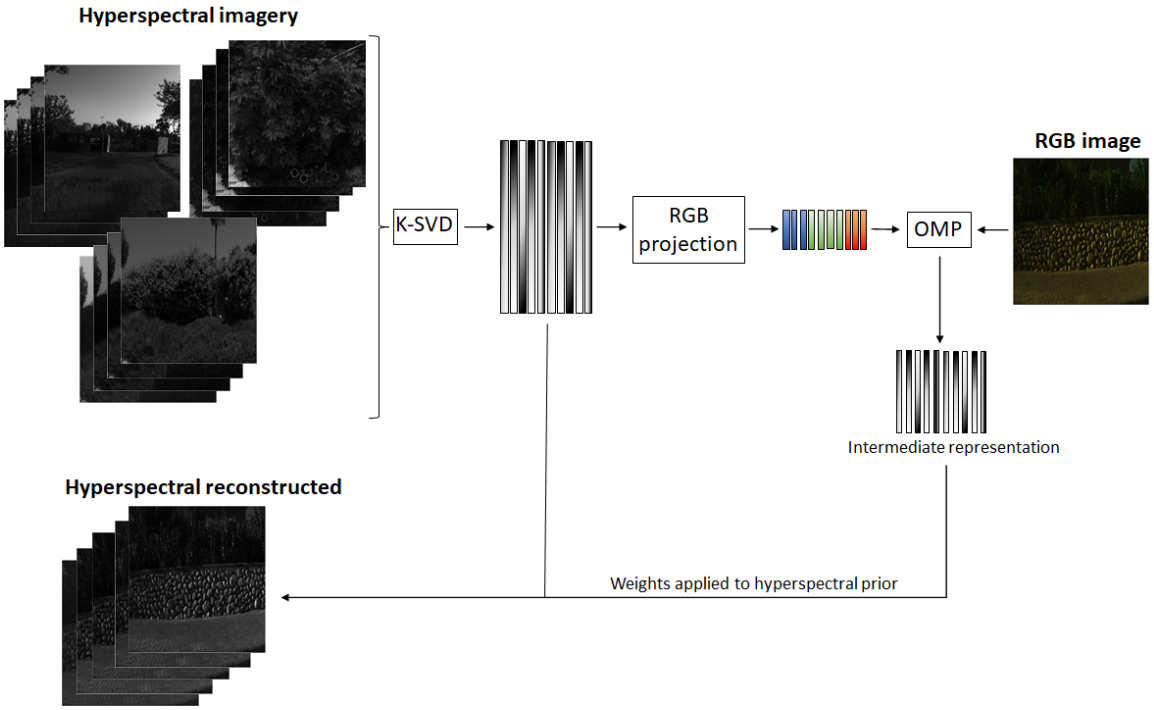

Figure 4.7: Flowchart of imagery and methods used in [Arad and Ben-Shahar, 2016].

4.4 left-plot). For each image, 100 random samples were selected and combined to create the over complete hyperspectral dictionary D_h using the K-SVD toolbox. The dictionary size was limited to 500 atoms with a sparsity constraint of 28 non-zero weights per atom. The resulting dictionary was then projected to RGB colour space to form D_{rgb}. Once all these components have been obtained, the dictionary representation of each RGB pixel was computed with the OMP using the same sparsity constraint.

4.3.2 HSCNN-R and HS-UNet-R for HSI reconstruction

HSCNN is one of the first CNN-based methods for hyperspectral recovery from a single RGB image [Xiong et al., 2017], inspired by the VDSR network for single image super-resolution [Kim et al., 2016]. However, this network architecture has two limitations that hinder performance: the use of a prior bicubic interpolation for spectral upsample [Smith et al., 1994] and failure to solve RGB to HSI mapping when model depth increases. Furthermore,

the prior spectral upsample requires knowledge of the sensor spectral response function and is found to be sub-optimal [Xiong et al., 2017].

The HSCNN-R introduces two methods to solve the mentioned issues.

1. The first is to remove the prior spectral upsample and replace the input layer with a single 1×1 convolutional layer for learnt spectral upsample
2. The second is the use of residual skips in each convolutional block to promote the network to learn the residual mapping [He et al., 2016]

The proposed method also disables pooling and batch normalisation during training and inference as decimation of spatial resolution and/or the shift and centering of statistical moments from each convolutional layer degrades performance [Shi et al., 2018].

The following sections also investigated the utility of U-Nets for hyperspectral recovery with added pooling operations before each convolutional block after the input layer. Therefore, the network topology of the encoder was the same as HSCNN-R but includes a pooling operation before each residual block. The decoder network used transposed 2×2 convolution for learnt feature map upsample and concatenates feature maps from each encoding stage at appropriate resolutions followed by a 3×3 convolution. The final layer used 1×1 convolution to output the same number of hyperspectral channels found in the ICVL dataset. In the experiments, this network was named HS-UNet-R. Figures 4.8 and 4.9 show the network architectures used for the experiments and results in Section 4.3.3.

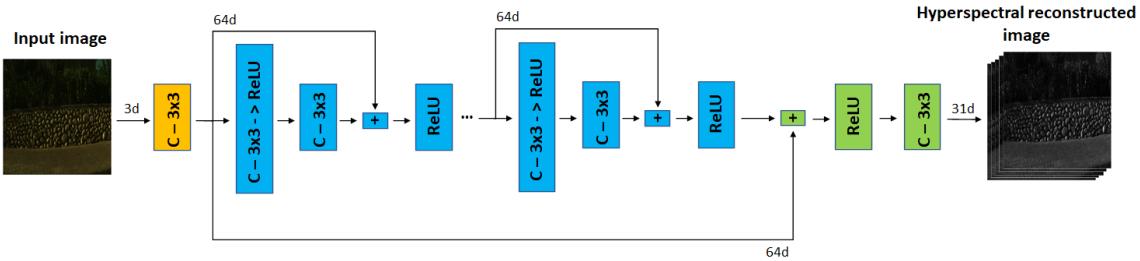

Figure 4.8: Encoder network for HSCNN-R [Shi et al., 2018]. C - 3×3 represents a convolutional block with ReLU non-linearity.

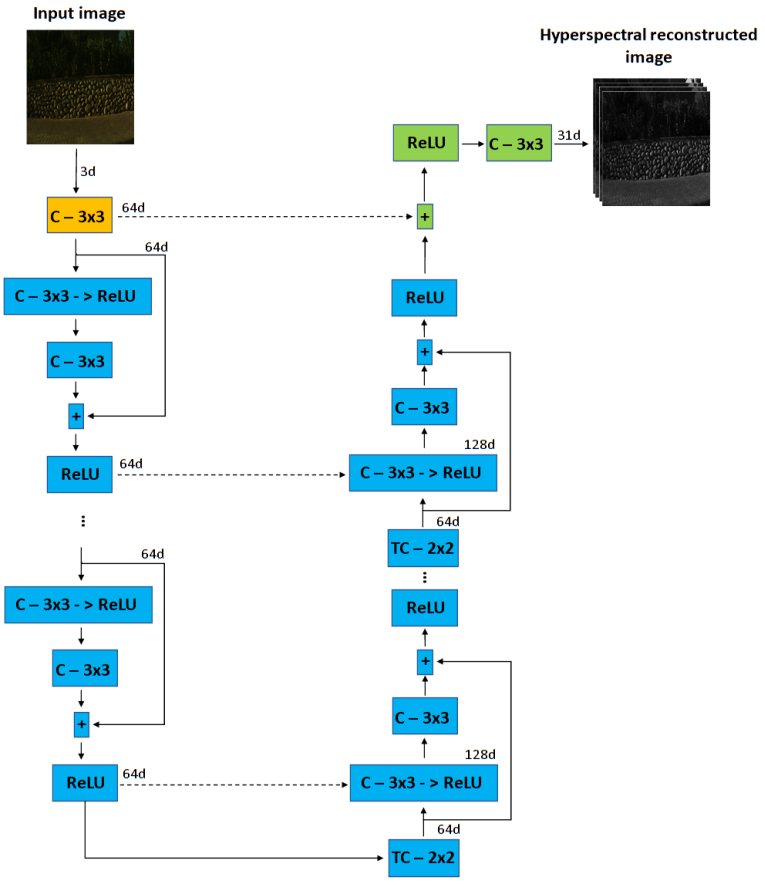

Figure 4.9: The HS-UNet-R architecture using the HSCNN-R [Shi et al., 2018] as an encoder network. C - 3×3 represents a convolutional block with ReLU non-linearity. TC - 2×2 represents a transposed convolutional block for learnt feature map upsample. Before each residual convolutional block in the encoder is a pooling operation that downsamples the image resolution by a factor of 2

Loss functions

The experiments show networks trained in supervised, semi-supervised and self-supervised settings in order to provide a complete analysis.

For the supervised loss term, consider $X \in \mathbb{R}^{B\times C\times H\times W}$ and $Y \in \mathbb{R}^{B\times N\times H\times W}$ to be respectively, a mini-batch of RGB images and corresponding HSI; where B, C, N, H and W are respectively, batch size, number of RGB channels, number of HSI channels, height and width. Processing a mini-batch outputs per-pixel continuous spectral measurements

$\hat{Y} \in \mathbb{R}^{B\times N\times H\times W}$. Then, the mean relative absolute error (MRAE) was computed between Y and $\hat{Y}$. The choice for MRAE over MSE loss is due to the luminance variation among different bands in a given pixel, where a MSE loss favours image bands with high luminance levels [Shi et al., 2018].

$$L_s(Y,\hat{Y}) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|Y_i - \hat{Y}_i|}{Y_i}\right) \tag{4.6}$$

Where, $\mathbf{i} \in \Omega$; $\Omega \subseteq \mathbb{Z}^2 \in (H \times W)$ is a pixel location.

For the unsupervised loss term, consider the same input image and the spectral response function of the miniaturised Sentera camera, Figure 4.4 left-plot, respectively $X \in \mathbb{R}^{B\times C\times H\times W}$ and $S \in \mathbb{R}^{N\times C}$; where B, C, N, H and W are respectively, batch size, number of RGB channels, number of HSI channels, height and width. The predicted HSI cube $\hat{Y} \in \mathbb{R}^{B\times N\times H\times W}$ was flattened to $\hat{Y} \in \mathbb{R}^{(B\times H\times W)\times N}$. Then, the spectral response function was used to project the predicted tensor to RGB colour space.

$$\hat{X} = \hat{Y}S \tag{4.7}$$

The reconstructed RGB image $\hat{X}$ was permuted to match the original input image $\hat{X} \in \mathbb{R}^{B\times C\times H\times W}$. Similarly to the supervised loss, the MRAE was computed between X and $\hat{X}$.

$$L_u(X,\hat{X}) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|X_i - \hat{X}_i|}{X_i}\right) \tag{4.8}$$

Where, $\mathbf{i} \in \Omega$; $\Omega \subseteq \mathbb{Z}^2 \in (H \times W)$ is a pixel location.

Furthermore, for self-supervised networks, a smoothing loss was added to constraint neighbouring channels in $\hat{Y}$. Again, the predicted tensor was flattened to $\hat{Y} \in \mathbb{R}^{(B\times H\times W)\times N}$. Then, the MSE of neighbouring channels was computed.

$$L_{smo} = \frac{1}{n}\sum_{i=1}^{n-1}(\hat{Y}_i - \hat{Y}_{i+1})^2 \tag{4.9}$$

Where, $\mathbf{i} \in \Omega$; $\Omega \subseteq \mathbb{Z} \in N$ is a predicted hyperspectral channel in $\hat{Y}$.

Implementation details

The objective loss function used for optimising networks depends on the mode used for training, as shown in equations 4.10, 4.11 and 4.12. If the mode is supervised, then networks were trained with only L_s. If the mode is semi-supervised, then networks were trained with L_s and L_u and the smoothing loss was omitted. If the mode is self-supervised, then networks were trained with L_u and L_{smo}. The reason for omitting the smoothing loss in supervised and semi-supervised settings was due to L_s already imposing a smoothing constraint on predicted spectra due to smooth high-quality labels from the ICVL challenge.

$$L = L_s \tag{4.10}$$

$$L = L_s + \alpha L_u \tag{4.11}$$

$$L = L_u + \beta L_{smo} \tag{4.12}$$

Where α was set to 0.1 and β to 0.0001 for all experiments in Section 4.3.3.

The HSCNN-R and HS-UNet-R variant were trained on 50×50 RGB patches and the corresponding hyperspectral data cubes. Each network architecture was trained for 800 epochs with a batch-size of 32 using Adam optimiser [Kingma and Ba, 2014]. The HSCNN-R had a depth of 12 residual layers while the HS-UNet-R had a depth of five. The learning rate was initially set to 0.0002 and reduced using polynomial decay of learning rate with the power set to 1.5. The hyper-parameters were chosen through an exhaustive search of various settings and monitoring the convergence with loss plots and accuracy metrics using the test set without cross-validation. The optimal set of parameters also matched the set of parameters described in [Shi et al., 2018]. During testing, the full size RGB image was processed in order to obtain a hyperspectral image. All HSCNN-R and extension to U-Net were implemented and trained using Pytorch version 10.2.

4.3.3 Quality assessment

The following objective metrics were used to evaluate the stated methods: Root Mean Square Error (RMSE) and Relative Root Mean Square Error (RRMSE).

$$RMSE(Y, \hat{Y}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{(Y_i - \hat{Y}_i)^2}{n}\right)} \tag{4.13}$$

$$RRMSE(Y, \hat{Y}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}\hat{Y}_i)^2}\right)} \tag{4.14}$$

Where, $Y, \hat{Y} \in \mathbb{R}^{N\times H\times W}$ are respectively the ground-truth and predicted HSI and $\mathbf{i} \in \Omega$; $\Omega \subseteq \mathbb{Z}^2 \in (H \times W)$ is a pixel location. Table 4.3 lists the results using the stated metrics for each method mentioned in the previous sub-sections. Figure 4.10 shows plots of predicted spectra for each method and different training modes. The plots show continuous spectral measurements for Red, Green, Blue samples from a colour checker present in some images of the dataset. The number of pixels for each colour was 100 which is a significantly lower number than the total number of pixels used to compute the values in Table 4.3. Figure 4.11 shows the visual results for each method at different wavelengths between 400-700nm in a supervised setting with a corresponding image displaying absolute error. The error units for equations 4.13 and 4.14 shown in Table 4.3 and Figure 4.10 are the same units as the quantity being estimated using the Specim PS Kappa DX4 hyperspectral camera which in this case are 12-bit pixel values.

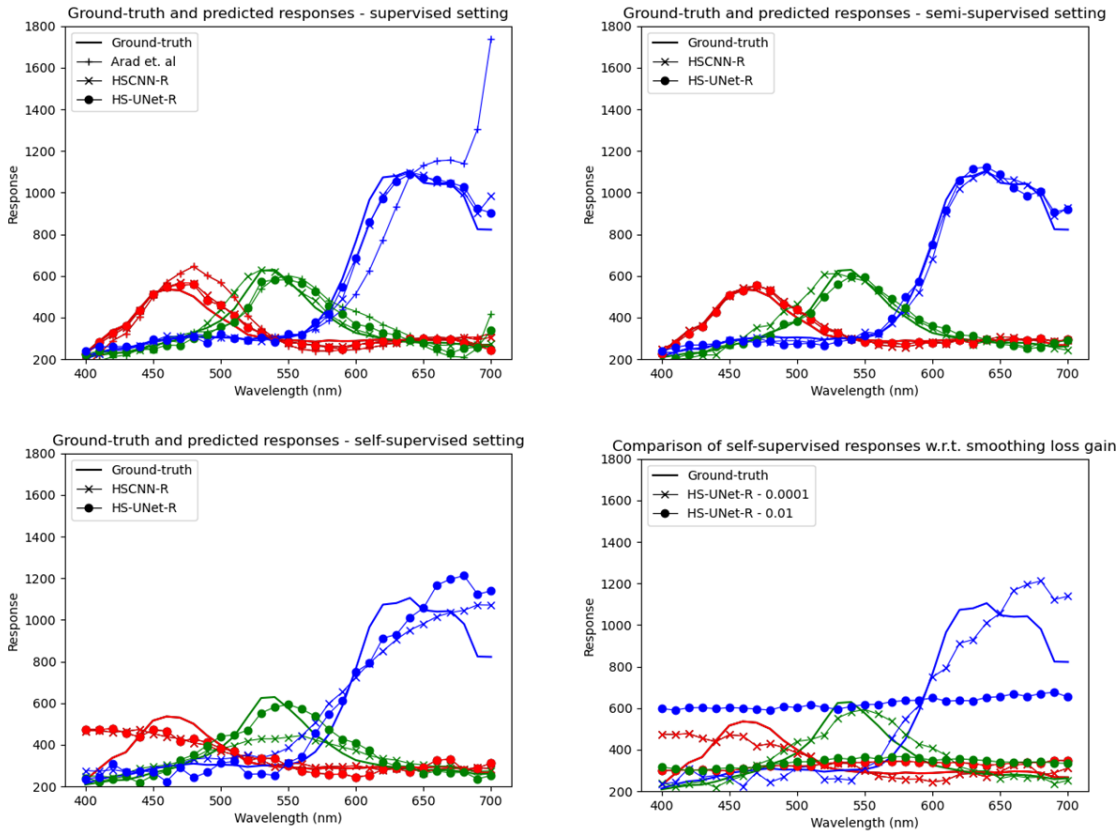

Figure 4.10: Plots for predicted reconstructed hyperspectral response from RGB. Top-left shows a comparison of supervised models with the shallow method in [Arad and Ben-Shahar, 2016]. Top-right shows plots for semi-supervised models. Bottom-left show plots for self-supervised models. And, the bottom-right plots shows the effects of varying the smoothing loss gain, L_{smo}, for different training procedures. The plots show continuous spectral measurements for Red, Green, Blue samples from a colour checker present in some images of the dataset. The error units are the same units as the quantity being estimated using the Specim PS Kappa DX4 hyperspectral camera which in this case are 12-bit pixel values.

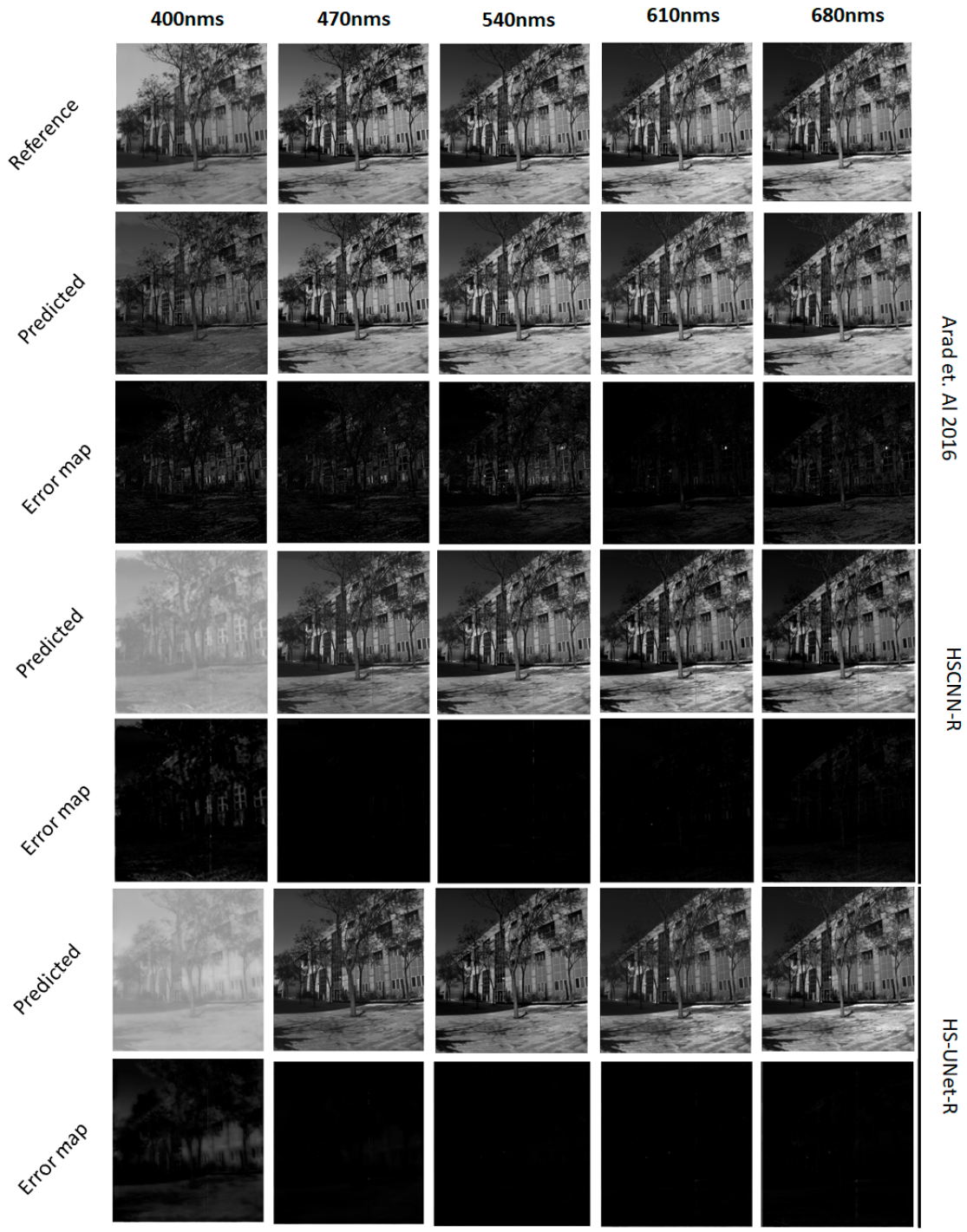

Figure 4.11: Gallery of predicted images for the methods stated in Sections 4.3.1 and 4.3.2. The predicted images for HSCNN-R and HS-UNet-R are trained in a supervised setting.

	Losses	RMSE	RMSE (RGB)	RRMSE	RRMSE (RGB)
Sparse-dict	-	3.663	31.784	0.0618	0.323
HSCNN-R	L_s	2.134	20.573	0.0232	0.238
	$L_s + \alpha L_u$	2.168	20.707	0.0221	0.239
	$L_u + \beta L_{smo}$	7.889	20.339	0.0783	0.239
HS-UNet-R	L_s	2.742	20.618	0.0272	0.238
	$L_s + \alpha L_u$	2.778	20.657	0.0281	0.236
	$L_u + \beta L_{smo}$	7.956	20.796	0.0808	0.241

Table 4.3: Root mean squared error and relative RMSE for predicted HSI and projected to RGB colour space using the sensor response function in Figure 4.4 (left-plot). The listed methods are sparse-dict as per Arad and Ben-Shahar [2016] and both HSCNN-R [Shi et al., 2018] and HS-UNet-R trained in supervised, semi-supervised or self-supervised settings

.

4.3.4 Discussion

The following sections discuss the results for Figures 4.10 and 4.11 and Table 4.3. First, the convergence of HSCNN-R and HS-UNet-R is described. Then, the shallow method in [Arad and Ben-Shahar, 2016] is compared with the deep learning models HSCNN-R and HS-UNet-R in a supervised setting. Finally, different training modes are compared for each network architecture.

HSCNN-R and HS-UNet-R convergence

The convergence of HSCNN-R and HS-UNet-R was analysed by testing multiple settings and hyper-parameters. The optimal set of hyper-parameters was determined by assessing computed errors for RMSE and RRMSE over five sequential runs; for each network and hyper-parameter setting. Table 4.3 and Figures 4.10 and 4.11 show results over five sequential runs with the same hyper-parameters described in Section 4.3.2. The results in Table 4.3 were an average of five sequential runs and the plots in Figure 4.10 were the best performing models from these sequential runs.

HSCNN-R and HS-UNet-R versus [Arad and Ben-Shahar, 2016]

From the results in Table 4.3, the method described in [Arad and Ben-Shahar, 2016] yielded 3.663 and 0.0618, respectively for RMSE and RRMSE, which was the worst performing method for hyperspectral reconstruction, if considering supervised and semi-supervised settings. This confirms the findings stated in [Xiong et al., 2017; Shi et al., 2018]. Furthermore, for images projected back to RGB colour space using the sensor responses in Figure 4.4 (left-plot), the same method had 31.784 and 0.323, respectively for RMSE and RRMSE, which was again the worst performing method. Both error metrics for HSI reconstruction can be confirmed with the plots shown in Figure 4.10 (top-left) - where the method in Arad and Ben-Shahar [2016] deviates more often from the HSI label. The poor performance with respects to both HSI reconstruction and RGB projection can be seen in Figure 4.11 - where the error map shows stronger intensity, especially around the edges of windows and walls. This was due to artifacts around edges or image patches, where the intensity changes abruptly. The latter issue was also stated in Aeschbacher et al. [2017]; Xiong et al. [2017].

The results for supervised HSCNN-R yielded the best results for HSI reconstruction in the experiment with 2.134 and 0.0232, respectively for RMSE and RRMSE. This can be confirmed with the plots shown in Figure 4.10 (top-left) - where predicted spectra were closer to the label compared to the method in [Arad and Ben-Shahar, 2016] and the HS-UNet-R. The error maps in Figure 4.11 paint a similar picture with less intensity for wavelengths beyond 470nm. However, predictions for 400nm were often blurred. The projection of reconstructed HSI to RGB colour space yielded 20.573 and 0.238, respectively for RMSE and RRMSE, which scored lower than the method in [Arad and Ben-Shahar, 2016]. As mentioned, this was mainly due to artifacts around image edges.

The results for supervised HS-UNet-R yielded 2.742 and 0.0238, respectively for RMSE and RRMSE, for HSI reconstruction. This particular architecture performed worse than the HSCNN-R in [Shi et al., 2018] but better than the method in [Arad and Ben-Shahar,

2016]. Again, this can be confirmed with the plots shown in Figure 4.10 (top-left) - where predicted spectra were closer to the label compared to the method in [Arad and Ben-Shahar, 2016] but not as precise as HSCNN-R. The error maps in Figure 4.11 show a similar problem as the HSCNN-R - where predictions for 400nm were often blurry. The projection of reconstructed HSI to RGB colour space yielded 20.618 and 0.238, respectively for RMSE and RRMSE, which scored lower than the method in [Arad and Ben-Shahar, 2016] but higher than HSCNN-R.

Overall this experiment showed that deep learning methods outperformed the method presented in Arad and Ben-Shahar [2016] with respects to both HSI reconstruction and RGB projection. Although the HS-UNet-R performed worse than the HSCNN-R, the degradation in HSI reconstruction still yields better results than the method in [Arad and Ben-Shahar, 2016]. Furthermore, as the U-Net architecture is a robust model for semantic segmentation, an opportunity surfaces to leverage an multi-task learning.

Different training modes for HSCNN-R and HS-UNet-R

From the results in Table 4.3, models trained in supervised and semi-supervised settings yielded similar scores, while the self-supervised setting performed worse. Since both the HSCNN-R and HS-UNet-R performed similarly with respect to different training modes, the analysis of both architectures was conducted jointly.

As mentioned, in a supervised setting, the HSCNN-R yielded 2.134 and 0.0232, while the HS-UNet-R yielded 2.742 and 0.238, respectively for RMSE and RRMSE. The same models in a semi-supervised setting yielded slightly higher errors with 2.168 and 0.0221 for the HSCNN-R and 2.778 and 0.238 for the HS-UNet-R, respectively for RMSE and RRMSE. Comparing the plots in Figure 4.10 (top-left with top-right) confirms that both training settings produce precise RGB to HSI mappings with both network architectures, and that while the unsupervised loss term degrades performance, the underlying structure of predicted spectra was maintained. The projection to RGB colour space paints a similar picture - where the

HSCNN-R in both settings had 20.579 and 20.707 RMSE, and the HS-UNet-R had 20.618 and 20.657 RMSE, respectively for supervised and semi-supervised settings.

The results for HSCNN-R trained with the self-supervised loss function yielded 7.889 and 0.0783, and the HS-UNet-R 7.956 and 0.808, respectively RMSE and RRMSE, which were the worst results for HSI reconstruction from RGB. This was expected due to the ill-posed and unconstrained nature of the problem, since the task was to predict a higher-dimensional space from a lower-dimensional manifold without any known prior. This can be confirmed by analysing the plots in Figure 4.10 (bottom-left) - where predictions deviate substantially from the HSI label. However, the results for HSI to RGB projection confirm the hypothesis that many physically plausible hyperspectral metamers could correspond to the same RGB capture [Cohen and Kappauf, 1982; Morovic and Finlayson, 2006]. For self-supervised models, the projection to RGB for HSCNN-R had 20.339 and 0.239, and for HS-UNet-R 20.796 and 0.241, respectively for RMSE and RRMSE, which were similar scores to models trained in supervised and semi-supervised settings. For self-supervised methods, the effects of changing the gain value for the smoothing loss in equation 4.9 was analysed. The gain value was important to produce physically plausible continuous spectral predictions. If the gain was too high, then the smoothing loss takes over the generalisation and produces continuous predictions for spectra with no underlying structure. If the gain was too low, the lack of constraint for neighbouring channels causes ragged predictions, Figure 4.10.

This analysis showed that models trained in a supervised setting produced the lowest error for HSI reconstruction. Furthermore, the addition of the unsupervised loss term causeed an increase in RMSE and RRMSE. For self-supervised models, the hypothesis that many physically plausible hyperspectral metamers could correspond to the same RGB capture was confirmed [Cohen and Kappauf, 1982; Morovic and Finlayson, 2006]. However, while the HSI reconstruction fidelity was not as precise as models trained in supervised and semi-supervised settings (Figure 4.10 - bottom left), the projection to RGB from HSI was as precise for all training modes.

4.3.5 Summary

This section shows two methods for HSI reconstruction. The first conclusion confirms the findings in [Xiong et al., 2017] - where the shallow method in [Arad and Ben-Shahar, 2016] was outperformed by deep implementations, such as HSCNN-R and HS-UNet-R. Furthermore, the addition of pooling operations and a subsequent upsample track in the network topology results in objective score degradation that confirms the findings in [Zhang et al., 2022]. However, while the pooling operation was not ideal for network design, the HS-UNet-R still outperformed the shallow method in [Arad and Ben-Shahar, 2016] and presents the opportunity to leverage multi-task learning (MTL); since the U-Net is a powerful model for semantic segmentation, refer to Section 2.3.2.

The analysis of different training modes also confirmed that many physically plausible hyperspectral metamers correspond to the same RGB capture [Cohen and Kappauf, 1982; Morovic and Finlayson, 2006] with the projection to RGB from HSI scoring the same in all training settings, Table 4.3.

4.4 Hyperspectral reconstruction for strandline and sand-dune communities

Section 4.3.2 showed the use of deep learning methods trained in supervised, semi-supervised and self-supervised settings for hyperspectral reconstruction.

As mentioned in Section 4.2.1, prior to each hyperspectral measurement, a white-point sample was measured in order to record a day-light illuminant spectrum. Then, the illuminant spectrum was removed from each measurement so that recorded hyperspectral measurements detail the intrinsic reflectance of plants that belong to SD1B, SD2, SD6A, SD7C and SD19 NVCs, displayed in Figure 4.3. The HSCNN-R and HS-UNet-R, described in Section 4.3.2, learn RGB to HSI mappings in accordance to the image formation fundamentals given in equation 2.2 - where the hyperspectral data cube, H, is the product of a illuminant

spectrum, L, with the intrinsic reflectance of the object, R. Therefore, given equation 2.2, the use of in-situ hyperspectral measurements for supervised or semi-supervised settings was not physically plausible without estimating the illuminant spectrum of the scene. In this section, a supervised method for hyperspectral reflectance recovery was developed, and the self-supervised method described in Section 4.3 instead reconstructs radiance, as given in equation 2.2.

Section 4.4.1 provides a background for hyperspectral imagery in remote sensing and coastal monitoring applications. Section 4.4.2 details the data pre-process for transcribing in-situ records onto VHR orthomosaics of the study site and the network architecture used for the experiments and results detailed in Section 4.4.3. For the results, the recorded hyperspectral measurements were used to evaluate the supervised and self-supervised methods.

4.4.1 Background

In remote sensing, hyperspectral imagery is costly and rare avenue for data acquisition since even with the introduction and widespread use of commercial satellites [Goetz, 2009]. HSI sensors capture an image where each pixel has continuous measurements of the underlying spectral structure that provide the opportunity to analyse land-cover materials or enhance intended applications [Plaza et al., 2009], such as endmember extraction [Zhang et al., 2011], spectral unmixing [Bioucas-Dias et al., 2013], target detection [Zhang et al., 2013] and image classification [Kang et al., 2013]. For coastal remote sensing, HSI capture has found many applications that aid in identification and discrimination of coastal and inland features [Banerjee and Shanmugam, 2021]. These applications include: classifying algal blooms [Ghatkar et al., 2019], red tides for large extents of the Korean South Sea, Yellow Sea, East China Sea and Bohai Sea [Ahn and Shanmugam, 2006], coral reef mapping from two sites on the Great Barrier Reef, Australia [Hedley et al., 2018], water quality assessment at Muttukadu Lake and Chilika Lake on the east coast of India [Kulshreshtha and Shanmugam, 2018], seagrass mapping on the shores of the Ionian Sea and Greek part of the Aegean Sea [Traganos et al., 2018]. Furthermore, many other studies have correlated vegetation health

with leaf reflectance properties that in turn can provide the basis to investigate the overall health and dynamics of a particular coastal ecosystem [Zinnert et al., 2012; Kearney et al., 2009; Naumann et al., 2008; Peñuelas and Filella, 1998].

The main limitations to mentioned work is two-fold. First, the category of sensor platform relates to commercial satellites which as mentioned in Section 2.1, incur high logistical costs per scene, oblique views that cause geometric and radiometric distortions to pixel [Loarie et al., 2007], and the spatial resolution of generated imagery can be too coarse for regional to local mapping objectives [Gould, 2000]. Second, the manufacturing cost of hyperspectral cameras is high and often results in a trade-off between high spectral but low spatial resolution [Gutiérrez et al., 2019; Behmann et al., 2018]. Pan-sharpening is a method to overcome coarse resolution in HSI by fusing a very-high resolution greyscale image with a corresponding HSI capture [Zhou et al., 2016]. However, the generated HSI cube is dependant on the quality of image fusion [Ghassemian, 2016]. Miniaturization of hyperspectral cameras [Basedow et al., 1995; Gonzalez et al., 2016] with rotor-based UAS could also solve issues related to spatial resolution. But, current implementations of miniaturized sensors are physically limited [Deng et al., 2021].

As mentioned in Section 2.4, another avenue is to extract hyperspectral information from a standard RGB image in a process known as hyperspectral reconstruction. In recent years, deep learning methods have pushed the fidelity in hyperspectral recovery from RGB images [Xiong et al., 2017; Shi et al., 2018; Lin and Finlayson, 2020] by using CNNs to learn accurate RGB to HSI mappings from HSI datasets [Yasuma et al., 2010; Arad and Ben-Shahar, 2016; Arad et al., 2018, 2020].

Given the hyperspectral measurements described in Section 4.2.1, this section also investigates recovering hyperspectral reflectance from a multispectral image. Deep implementations of hyperspectral reflectance reconstruction have been proposed in [Deeb et al., 2019; Zhang et al., 2020; Gong et al., 2022] that produce high fidelity reflectance recovery from RGB images. In particular, Gong et al. [2022] consider correlating predicted leaf reflectance

with chlorophyll concentration which is a marker for assessing plant health [Zinnert et al., 2012; Naumann et al., 2008].

4.4.2 Methodology

Data pre-processing for HSCNN-R and HS-UNet-R

The recorded data points during the in-situ survey for species belonging to SD1B, SD2, SD6A, SD7C and SD19 NVCs listed in Section 4.2.1 provide the basis to create geographically referenced polygon files through photo interpretation, in the same manner as described in Section 3.3.1. This process generated 57 polygons, with one polygon for every recorded sample in Table 4.2. These polygons contain recorded hyperspectral reflectance measurements that can be accurately transcribed onto VHR orthomosaics because of RTK GPS logs. For the experiments in Section 4.4.3, two datasets were created for each mode used to train the HSCNN-R and HS-UNet-R.

The self-supervised dataset used all the polygons for model evaluation, since this method does not require labelled data. The dataset used for supervised hyperspectral reflectance recovery had polygons split to an 80/20 ratio, respectively for train and test set. Therefore, the train set used 47 polygons and the test set used the remaining 10 polygons. The test set polygons belong to the reduced class domain described in Section 4.2.1, with one polygon for each class.

Polygons to masks for HSCNN-R and HS-UNet-R

HSCNN-R and HS-UNet-R models were trained and evaluated with masks that contain one-to-one mappings of continuous hyperspectral measurements. These masks were created using the geographic coordinates stored in each polygon and converting real-world coordinates for each vertex to image-coordinates. If a polygon fits in an image, then the candidate image was sampled into 256×256 image tiles centered on polygons. Figure 4.12 shows ex-

amples of cropped aerial imagery with the corresponding mask. This process generated one image per polygon.

For the self-supervised method, training images were collected using random 256×256 image crops of the orthomosaic. A selection of 200 random images spread across the site were used for model training. These models were evaluated on all 57 images with known hyperspectral measurements. The supervised method for intrinsic reflectance recovery had a different dataset with 47 images for training and ten images for model testing.

As mentioned in Section 4.2.1, fifteen or twenty five hyperspectral measurements were recorded at different locations of each marine species. For each data entry recorded in Table 4.2, the associated hyperspectral measurements were averaged so that each polygon contains a single spectral signature. Then, the averaged spectral signature was copied to every pixel in the polygon. Figure 4.13 shows this process for a sample of Crambe maritima, where each pixel represents the averaged spectral signature along the channel dimension. Furthermore, for input images used to train supervised reflectance recovery, the pixel values for each channel in the polygon were averaged. Then, the average value was copied to every pixel.

HSCNN-R and HS-UNet-R training parameters

The details for HSCNN-R and HS-UNet-R were described in Section 4.3.2. Figures 4.8 and 4.9 show the network architectures used for this experiment. This section evaluated the HSCNN-R and the HS-UNet-R optimised with equations 4.10 and 4.12.

Self-supervised networks were optimised using equation 4.12 with the parameter β set to 1.0 for all results. Supervised networks were optimised with equation 4.10.

Each network architecture was trained for 600 epochs with a batch-size of four using Adam optimiser [Kingma and Ba, 2014]. The HSCNN-R had a depth of 12 residual layers while the HS-UNet-R had a depth of five. The learning rate was initially set to 0.0005 and reduced using polynomial decay of learning rate with the power set to 1.5. The hyper-

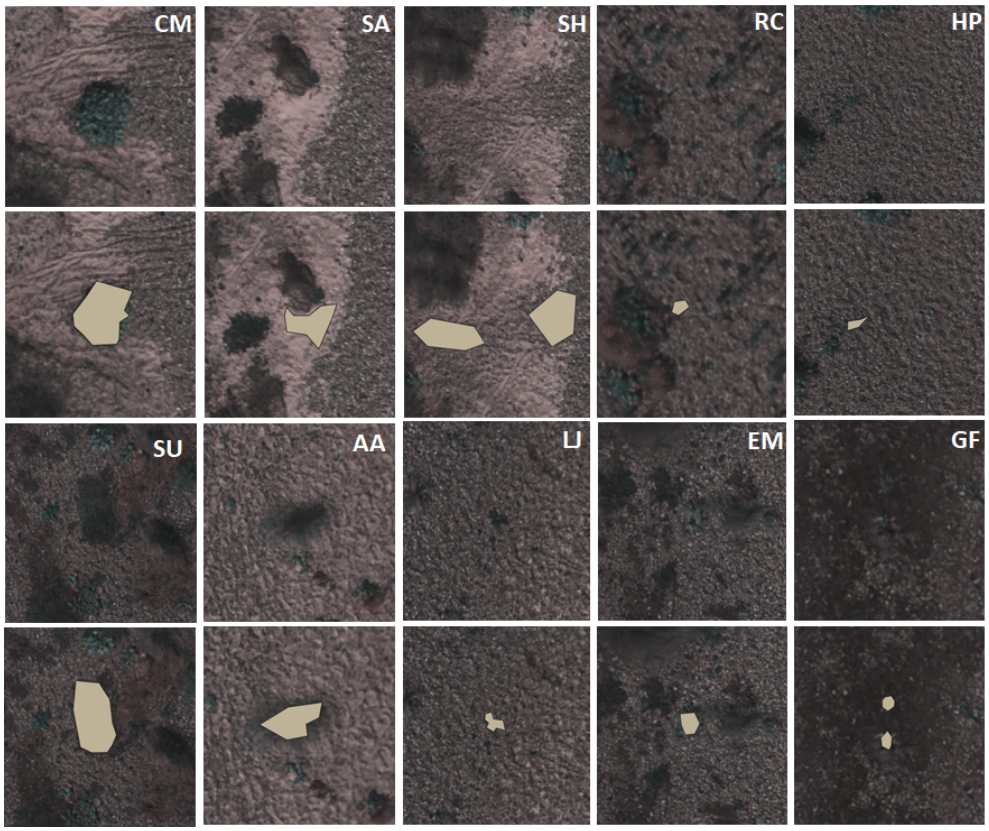

Figure 4.12: Gallery of images with corresponding masks used for training and testing. AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SU- Silene uniflora

parameters were chosen through an exhaustive search of various settings and monitoring the convergence with loss plots and accuracy metrics. All HSCNN-R and extension to U-Net were implemented and trained using Pytorch version 10.2.

4.4.3 Quality assessment

The objective metrics used to evaluate the stated methods are given in equations 4.13 and 4.14. Table 4.4 lists the results for each method using the stated metrics. Figure 4.14 shows plots of predicted reflectance using the supervised loss metric defined in Eq. 4.10 and Figure 4.17 show plots of predicted radiance using the self-supervised loss metric defined in Eq. 4.12. The plots show continuous spectral measurements for each species belonging to

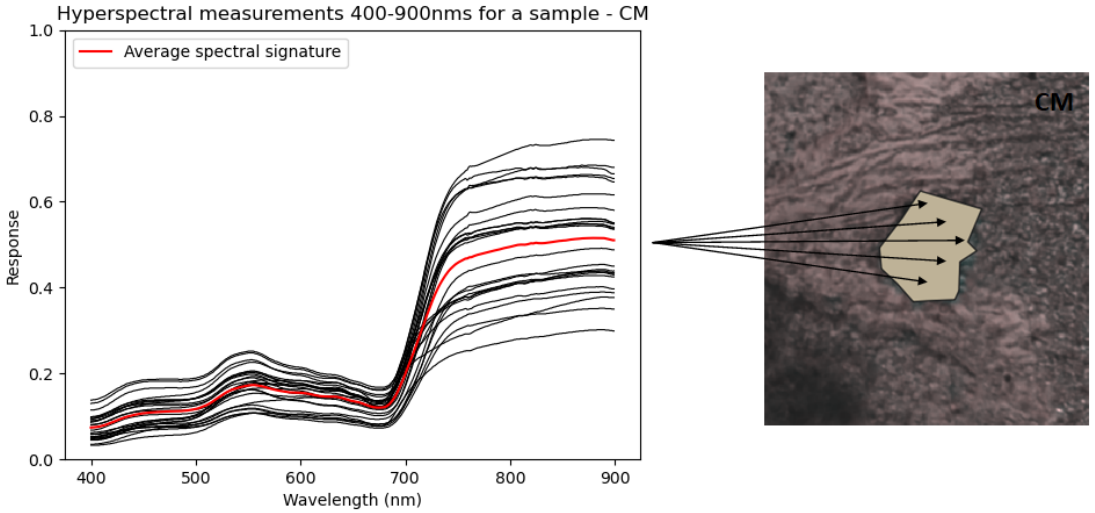

Figure 4.13: Hand-gun measurement for a sample of Crambe maritima. Twenty five hyperspectral measurements were recorded at different locations of this plant of Crambe maritima. The measurements were averaged to a single spectral signature, as shown with the red line (left-plot), and then copied to every pixel in the polygon corresponding to this particular plant of Crambe maritima (arrows to right-figure).

the reduced class domain described in Section 4.2.1. Figures 4.15 -4.19 show visual results at different wavelengths between 400-900nm for each network architecture optimised with either Eq. 4.10 in a supervised setting or Eq. 4.12 in a self-supervised setting.

4.4.4 Discussion

The following sections discuss the results in Figures 4.14 - 4.19 and Table 4.4. First, the convergence of HSCNN-R and HS-UNet-R is discussed. Then, both network architectures are compared in a supervised setting (reflectance reconstruction) and self-supervised setting (radiance reconstruction). The results in Table 4.4 show errors for each dataset described in Section 4.4.2. The self-supervised method was evaluated on all 57 images, whereas the supervised method was evaluated on ten images. Figures 4.14 - 4.19 were generated using the 10 images from supervised method test set to provide a comparison of both methods.

	Losses	RMSE	RMSE (RGB)	RRMSE	RRMSE (RGB)
HSCNN-R	L_s	0.2404	222.63	0.0049	3.461
	$L_u + \beta L_{smo}$	1.794	0.656	0.0171	0.049
HS-UNet-R	L_s	0.3023	240.10	0.0066	3.564
	$L_u + \beta L_{smo}$	1.677	0.667	0.0166	0.063

Table 4.4: Root mean squared error and relative RMSE for predicted HSI and projected to RGB colour space using the sensor response function in Figure 4.4 (righ-plot)

.

HSCNN-R and HS-UNet-R convergence

The convergence of HSCNN-R and HS-UNet-R was analysed by testing multiple settings and hyper-parameters. The optimal set of hyper-parameters was determined by assessing computed RMSE and RRMSE over five sequential runs for each network and hyper-parameter setting. Table 4.4 and Figures 4.14 - 4.19 show the results over five sequential runs with the same hyper-parameters described in Section 4.4.2. The results in Table 4.4 were the average from these sequential runs, and the plots and visual results in Figures 4.14 - 4.19 show the best performing model.

HSCNN-R and HS-UNet-R for supervised hyperspectral reflectance reconstruction

From the results in Table 4.4, the HSCNN-R trained for hyperspectral reflectance reconstruction yielded 0.2402 and 0.0049, respectively for RMSE and RRMSE, which suggests that these models produced accurate hyperspectral reflectance recovery for the dataset stated in Section 4.4.2. These errors can be confirmed with the plots shown in Figure 4.14 - where continuous reflectance predictions were close to the hyperspectral measurement, with the test image for Rumex crispus deviating the most. However, the projection to multispectral colour space using the Sentera camera responses (Figure 4.4 - right-plot) had 222.63 and 3.461, respectively RMSE and RRMSE. This was expected given that the projection was not in accordance to equation 2.2, since the illuminant spectrum of the scene was not known. The recorded illuminant spectrum during the calibration process described

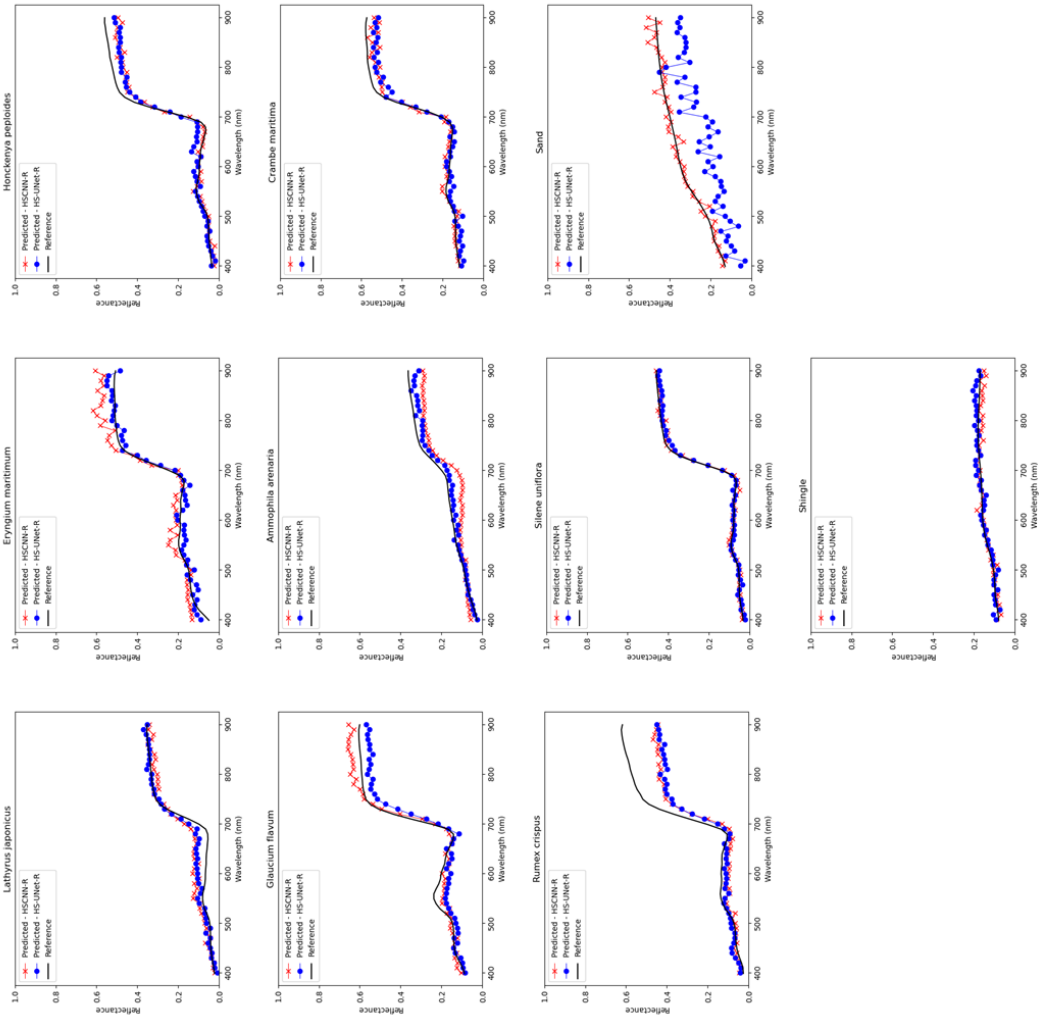

Figure 4.14: Plots for predicted reconstructed hyperspectral reflectance from RGB. Plots show supervised models using equation 4.10 to reconstruct reflectance.

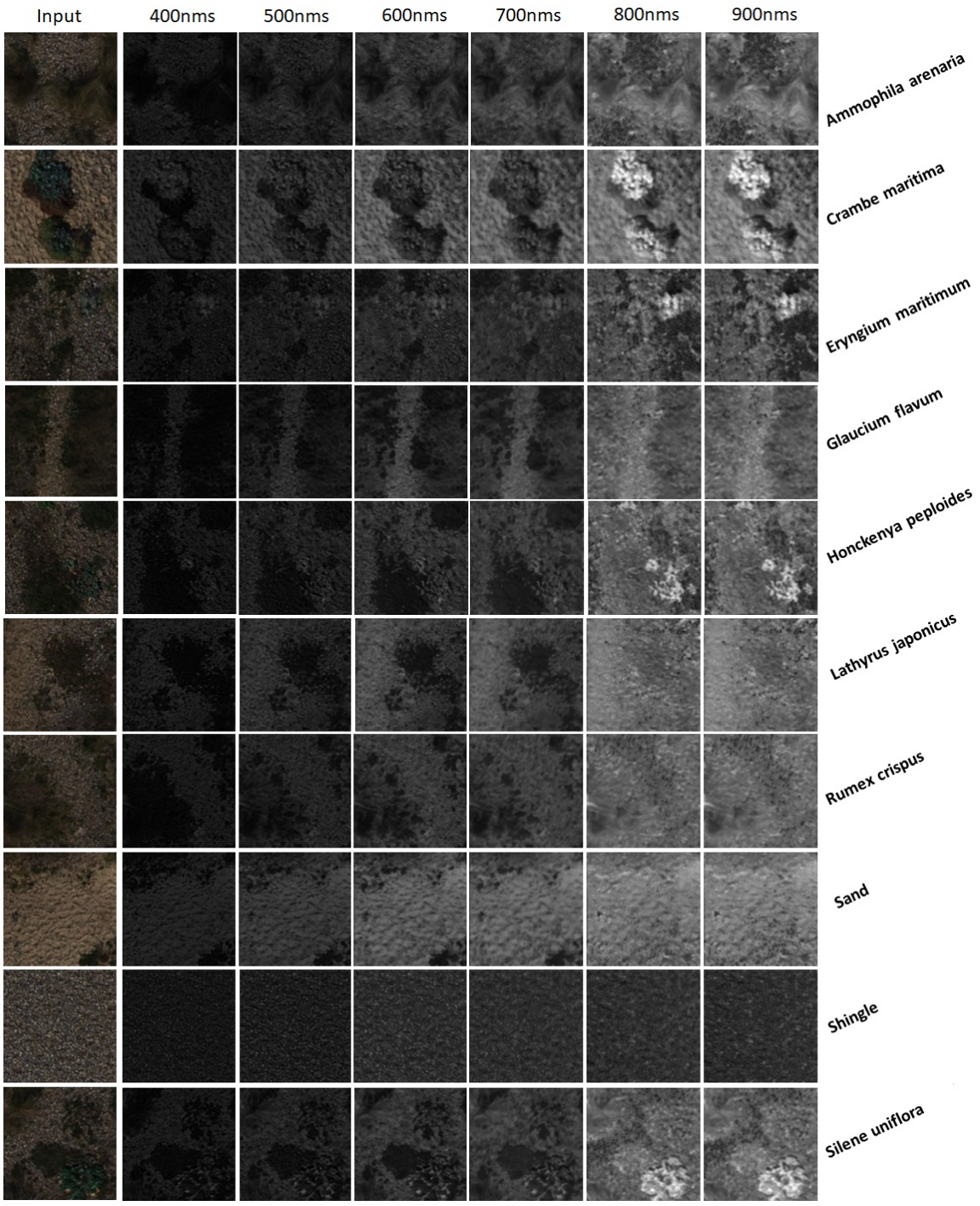

Figure 4.15: Visual results for reconstructed hyperspectral reflectance from multispectral Sentera imagery using HSCNN-R. Visual results show supervised models using equation 4.10 to reconstruct reflectance.

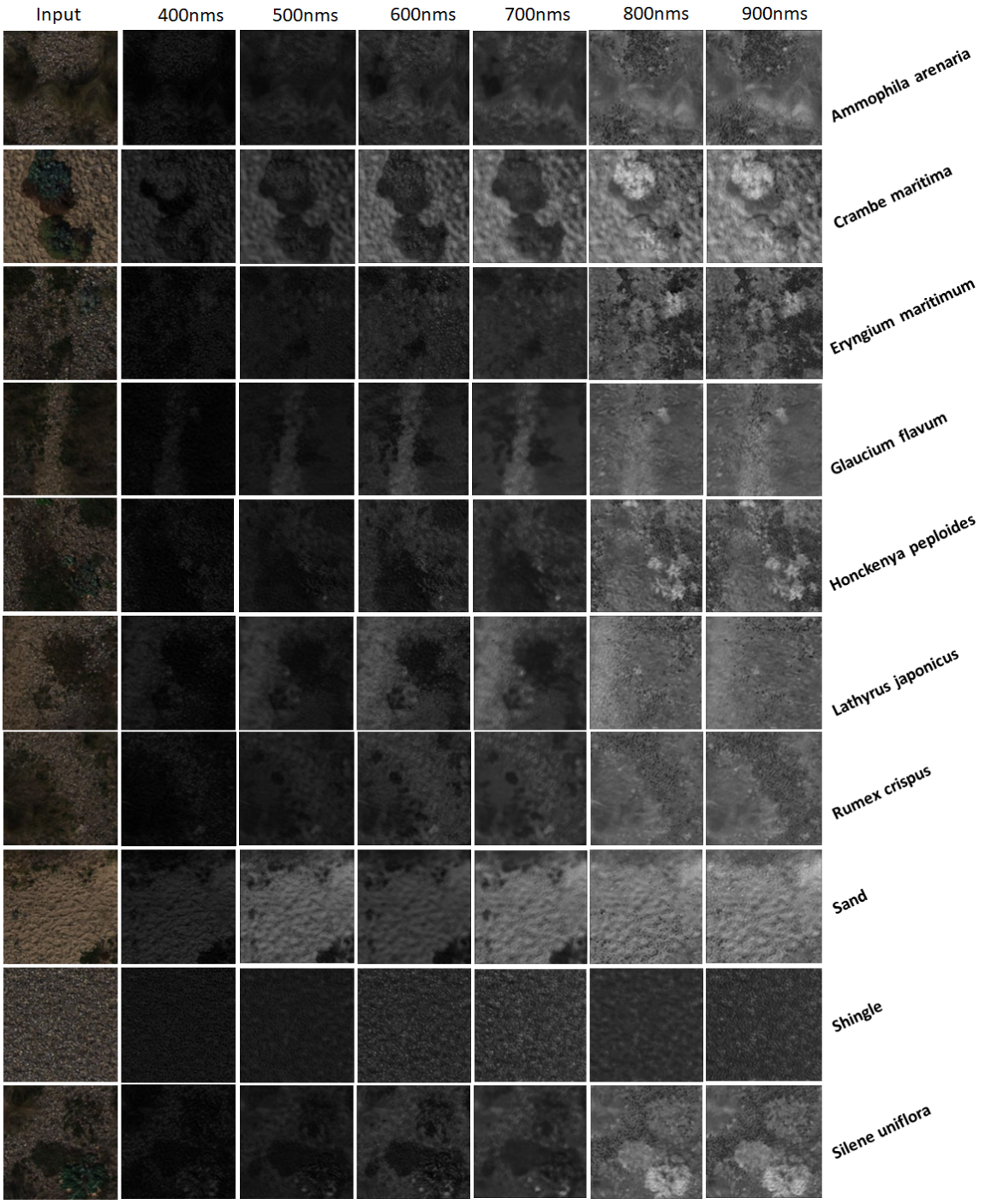

Figure 4.16: Visual results for reconstructed hyperspectral reflectance from multispectral Sentera imagery using HS-UNet-R. Visual results show supervised models using equation 4.10 to reconstruct reflectance.

in Section 4.2.1 was not suitable for model training. The goal for recording a white-sample was to calibrate and scale the recorded reflectance in Figure 4.3 to a range between zero to one. However, the recorded scale of the illuminant spectrum was at a scale defined by the spectroradiometer instrument. Therefore, subsequent scaling of recorded illuminant spectrum to the same range was possible but would result in ad-hoc tuning during model training.

The HS-UNet-R yielded 0.3023 and 0.0066, respectively RMSE and RRMSE. This was the same result as discussed in Section 4.3.4, where the HS-UNet-R was worse than the HSCNN-R. But, these scores still suggest that the HS-UNet-R can produce accurate hyperspectral reflectance recovery for the dataset stated in Section 4.4.2. This was confirmed with the plots shown in Figure 4.14 - where the HS-UNet-R deviates more often from the HSI measurements than the HSCNN-R, and in particular for the test image with sand. The projection to multispectral colour space had 240.10 and 3.564, respectively RMSE and RRMSE, which was worse than the HSCNN-R. Again, this was expected given that the projection was not in accordance to equation 2.2.

Overall this experiment shows that the HSCNN-R and HS-UNet-R can learn hyperspectral reflectance reconstruction using the dataset described in Section 4.4.2. Indeed, both models learn reflectance but also learn to discount the illuminant from input images which entails that this method cannot be used on another site without re-training the model with the same set of hyperspectral measurements. This is a common issue in remote sensing that requires normalised inputs from one study site to another which in this case would be the reflectance of different shingle vegetation species.

Figures 4.15 and 4.16 also showed some of the limitations of the dataset described in Section 4.4.2. First, the number of images used for training and testing the HSCNN-R and HS-UNet-R was limited in order to conclude whether this method for hyperspectral reflectance was comparable to the work done in [Deeb et al., 2019; Zhang et al., 2020; Gong et al., 2022]. Furthermore, the pre-process of hyperspectral measurements and input images for generated

polygons described in Section 4.4.2 results in visual artifacts and coarse boundaries for reconstructed vegetation features. This was noticeable for rows showing showing Glaucium flavum and Silene uniflora in Figures 4.15 and 4.16, where the vegetation features lack fine-grain detail. Another visual artifact can be found for rows showing visual predictions of sand, where salt and pepper artifacts can be found among casting shadows of sand pockets, and the nature of shingle that may present small pebbles with distinct colours thus returning different hyperspectral signatures. These artifacts could also be caused by the small number of labelled samples in the dataset.

HSCNN-R and HS-UNet-R for self-supervised hyperspectral reconstruction

The results in Table 4.4 show that self-supervised models yielded similar scores for both the HSCNN-R and HS-UNet-R. Therefore, the analyses for both network architectures was conducted jointly.

The HSCNN-R had 1.794 and 0.0171, and the HS-UNet-R 1.677 and 0.0166, respectively for RMSE and RRMSE, which performed worse than supervised models. As mentioned in Section 4.3.4, this was expected due to the ill-posed and unconstrained nature of the problem, since the task was to predict a higher-dimensional space from a lower-dimensional manifold without any known prior. Again, this can be confirmed with the plots in Figure 4.17 - where hyperspectral predictions deviate substantially from the hyperspectral measurement. In fact, irrespective of the target vegetated or non-vegetated class in the reduced class domain, each predicted response details similar structure. The responses between 400-700nm vary from species to species but then step up similarly from 700-900nm. Furthermore, it was important to note that the hyperspectral measurements in Figure 4.17 show reflectance measurements of the reduced class domain but the goal with self-supervised models was to reconstruct radiance, as given in equation 2.2.

For the projection to multispectral colour space using the responses in Figure 4.4 (right-plot), the HSCNN-R had 0.656 and 0.0049, and the HS-UNet-R 0.667 and 0.063, respectively

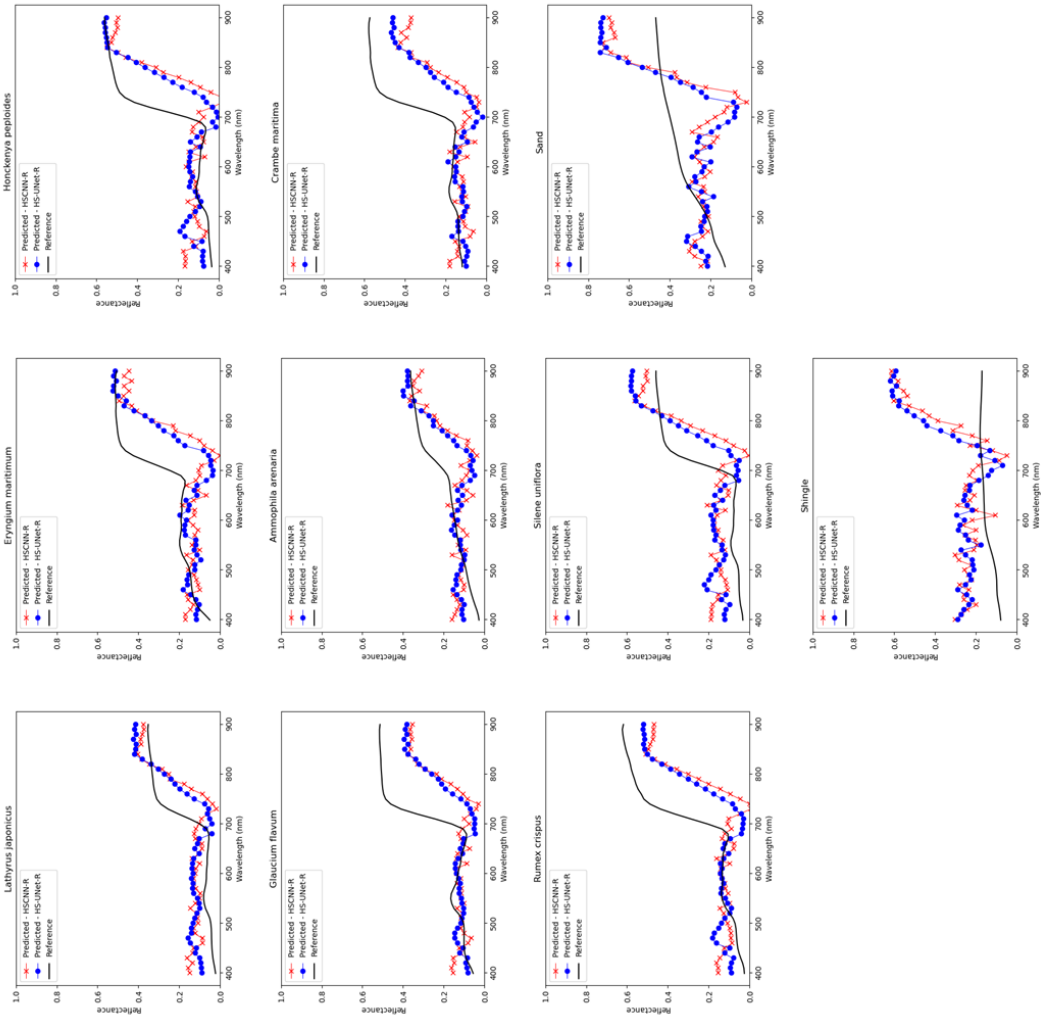

Figure 4.17: Plots for predicted reconstructed hyperspectral radiance from RGB. Plots show self-supervised models using equation 4.12 to reconstruct radiance.

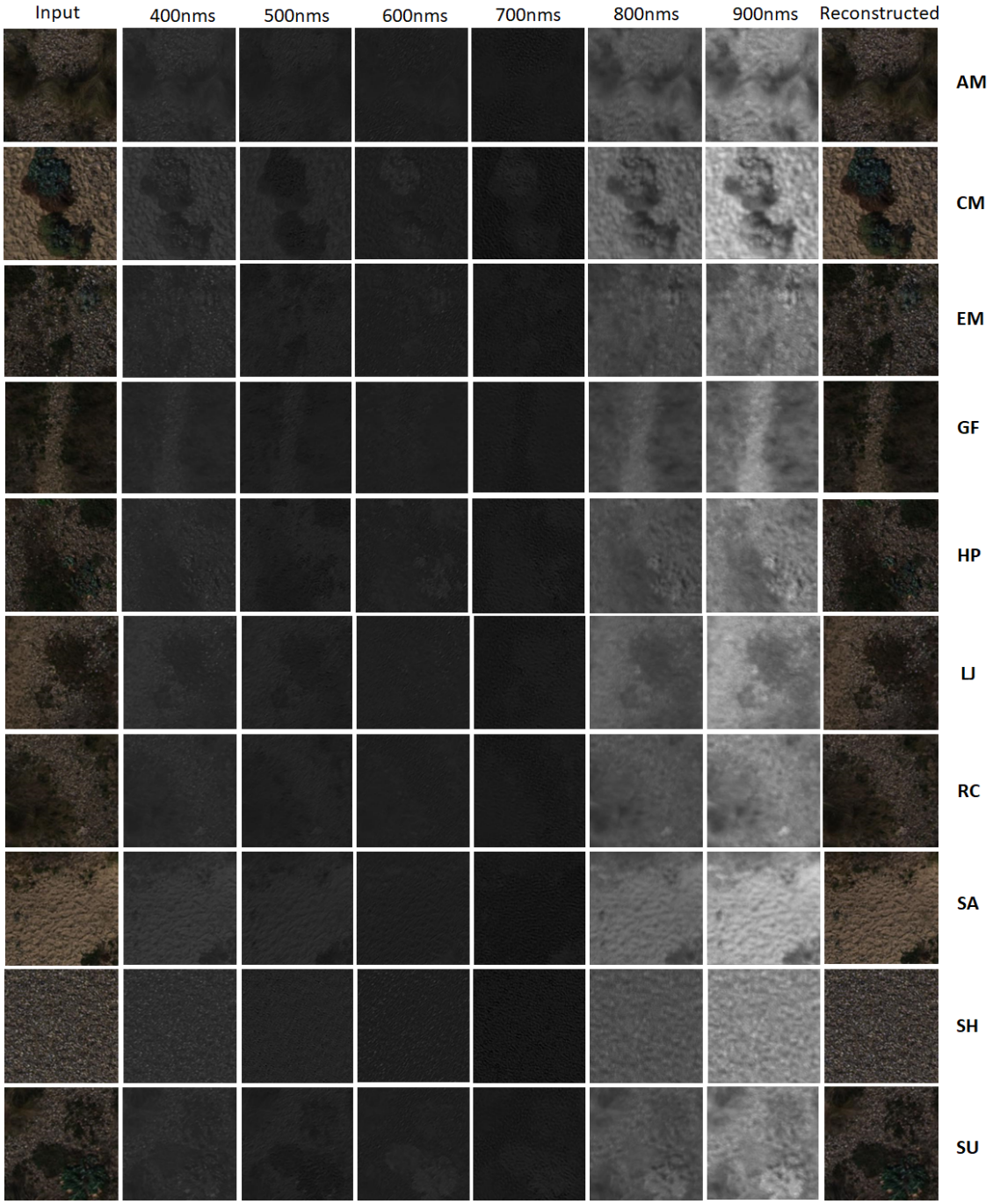

Figure 4.18: Visual results for reconstructed hyperspectral radiance from multispectral Sentera imagery using HSCNN-R. Visual results show self-supervised models using equation 4.12 to reconstruct radiance.
AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SU- Silene uniflora

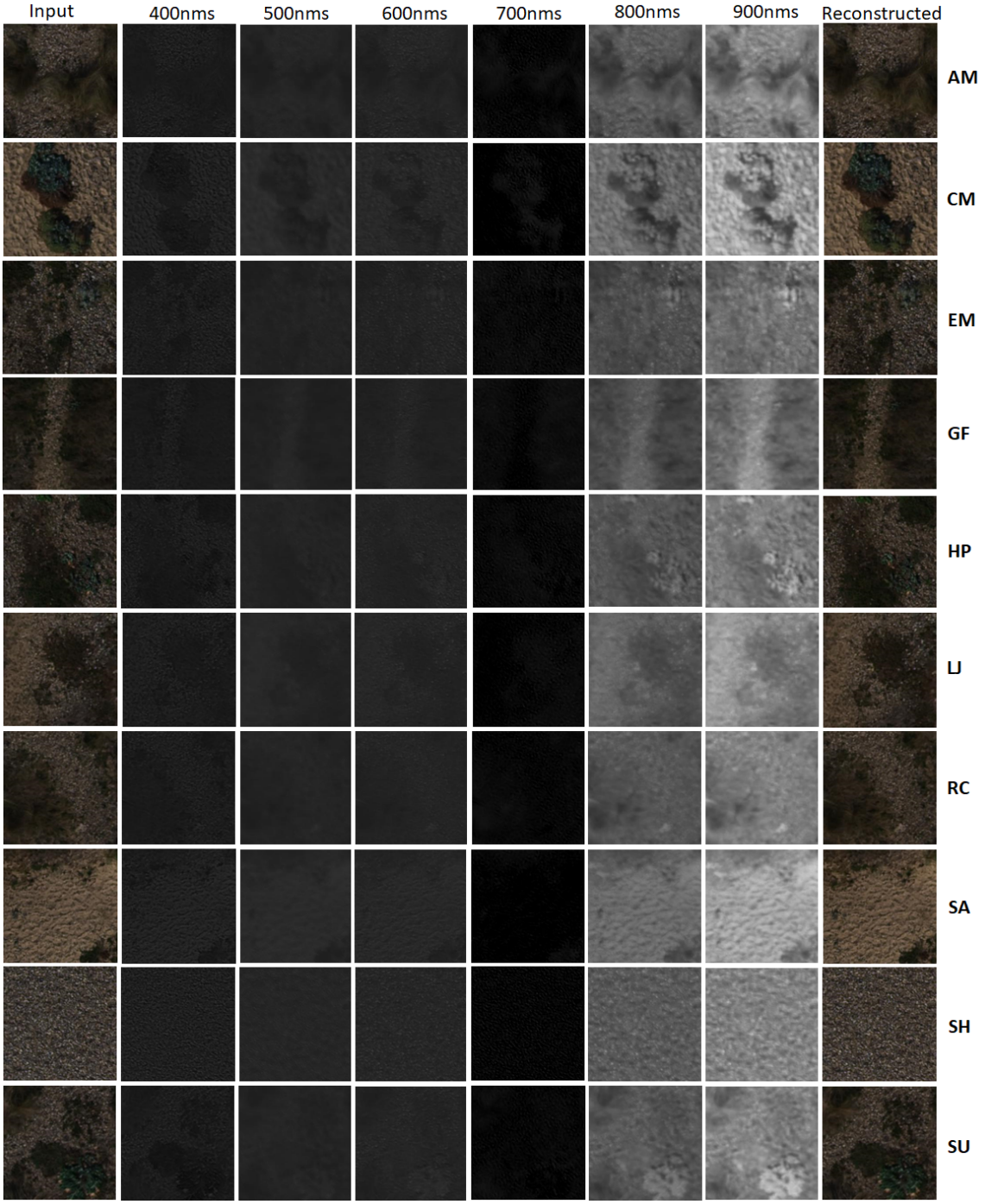

Figure 4.19: Visual results for reconstructed hyperspectral radiance from multispectral Sentera imagery using HS-UNet-R. Visual results show self-supervised models using equation 4.12 to reconstruct radiance.
AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SU- Silene uniflora

for RMSE and RRMSE. This suggests that both network architectures produce accurate reconstructed multispectral images when optimised with equation 4.12. This can be confirmed with the visual results in Figures 4.18 and 4.19, where the reconstructed multispectral image was identical to the input image. Again, this confirms that many physically plausible hyperspectral metamers can correspond to the same multispectral capture [Cohen and Kappauf, 1982; Morovic and Finlayson, 2006].

This experiment confirms the findings in Section 4.3.4, where the HSI reconstruction fidelity was not as precise for self-supervised, as shown in Figure 4.17 but the projection to multispectral colour space from HSI was precise. Furthermore, comparing Figures 4.15 and 4.16 with Figures 4.18 and 4.19 shows that the visual artifacts, as a result of the pre-processing steps described in Section 4.4.2, were not present for self-supervised models, and detailing features for vegetation target classes from the input image were maintained for each reconstructed hyperspectral channel.

While it is hard to confirm whether HSCNN-R and HS-UNet-R have overfit on the train data given the test set comprises of ten images and the lack of a further dataset split to include a validation set. The plots shown in Figure 4.14 show different spectral signatures for each shingle vegetation specie with both models accurately predicting each hyperspectral signature corresponding to a particular specie correctly that may confirm that the model has learnt hyperspectral reflectance recovery on unseen data.

4.4.5 Summary

Section 4.4 continued to develop the methods in Section 4.3.

The hyperspectral measurements described in Section 4.2.1 were used to optimise the HSCNN-R and HS-UNet-R in a supervised setting to learn reflectance recovery. The results with the supervised setting confirm the findings detailed in Section 4.3.4, where the HSCNN-R outperformed the HS-UNet-R. However, the dataset and pre-processing stages of hyperspectral measurements described in Section 4.4.2 resulted in visual artifacts for

reconstructed hyperspectral channels, as shown in Figures 4.15 and 4.16. Furthermore, the test set was limited to ten images that in turn does not provide the basis to compare with methods presented in [Deeb et al., 2019; Zhang et al., 2020; Gong et al., 2022].

The self-supervised hyperspectral radiance reconstruction described in Section 4.3 was applied to multispectral imagery of the study site. The results in Section 4.4.3 also confirm the findings described in Section 4.3.4, where the HSI reconstruction fidelity was not precise due to the ill-posed and unconstrained nature of reconstructing a higher-dimensional space from a lower-dimensional manifold. However, the projection to multispectral colour space from HSI was precise.

4.5 Multi-task learning: segmentation with hyperspectral reconstruction

Sections 4.3 and 4.4 showed the use of deep learning methods for hyperspectral reflectance and radiance reconstruction. The utility of a U-Net architecture for hyperspectral reconstruction was investigated to note for performance degradation, as stated in Zhang et al. [2022]. However, given the performance degradation, the U-Net architecture still outperformed the shallow method described in Section 4.3.1.

In Section 3.3, deep learning models were trained using supervised and semi-supervised settings. In the next section, an alternative semi-supervised optimisation method that leverages multi-task learning (MTL) was used for semantic segmentation, instead of consistency-based regularisation. The goal was to use MTL in order to learn both semantic segmentation and hyperspectral reconstruction with a shared model. As mentioned in Section 2.5, by sharing internal representations from auxiliary training signals, models can learn different internal representations and prevent overfit on the original main image task [Ruder, 2017]. In this case, the main image task was semantic segmentation, given the main objective was to map the pioneering species belonging to SD1B, SD2 and SD6A NVCs, and the auxiliary image task was hyperspectral reconstruction.

The HS-UNet-R architecture described in Section 4.3.2 was adapted for the purpose of learning both tasks. The tasks for MTL were either supervised semantic segmentation with supervised reflectance recovery, or supervised semantic segmentation with self-supervised radiance recovery. Therefore, the mapping results compared supervised and semi-supervised methods, as described in Section 3.3, with supervised models trained in MTL. The mapping results also include an analyses of the OBIA method with the objective metrics used in Section 3.3.2.

4.5.1 Background

In remote sensing, MTL frameworks that attempt to solve semantic segmentation as a main image task often attempt to improve inter-class boundary separation as an auxiliary task. For instance, Volpi and Tuia [2018] learns semantic segmentation and semantic boundary detection jointly. Bischke et al. [2019] uses a shared model to learn semantic segmentation and also incorporates a self-supervised distance metric for accurate inter-class boundary separation. Another method presented Li et al. [2021a] uses a boundary attention module for refined segmentation boundaries. These methods leverage an auxiliary image task closely related to the main image task. In Lu et al. [2022], the MTL framework learns two different problem domains which in this case are semantic segmentation and depth estimation. Another method following this trend presented in Wang et al. [2020a] learns semantic segmentation in tandem with object detection. For coastal remote sensing, the trend of MTL framework with semantic segmentation as the main image task is to also use auxiliary boundary refinement tasks [Jing et al., 2021; Liu et al., 2020b] or semantic segmentation with depth estimation [Carvalho et al., 2019].

The following section shows an MTL framework to jointly learn self-supervised radiance reconstruction or supervised reflectance recovery with supervised semantic segmentation. The premise was that the auxiliary training signals, unrelated to the main image task, can refine semantic segmentation boundaries or improve inter-class separation. Section 4.5.2 shows the method for MTL with a shared model. Section 4.5.4 provides a discussion of

visual mapping results and objective scores of supervised and semi-supervised methods, as described in Section 3.3, supervised models trained with MTL and OBIA.

4.5.2 Methodology

Data pre-processing for MTL

Again, as in Section 4.4.2, the recorded data points during the in-situ survey were used to create geographically referenced polygon files through photo interpretation. As described in Section 4.2.1, the segmentation task had a reduced class domain. Therefore, from the 57 polygons described in Section 4.4.2, 40 polygons belonging to the reduced class domain were selected.

From these classes, extra polygons were added in the vicinity of known points. The target classes with extra polygons were Ammophila arenaria (14 polygons), Crambe maritima (11 polygons), Silene uniflora (five polygons), sand (eight polygons) and shingle (five polygons). This brings the total number of polygons for the segmentation dataset to 83 polygons. The use of photo-interpretation also draws from the conclusions stated in Section 3.4, where good quality labels can be found in-situ, but a healthy quantity of labels can also be supplemented from aerial imagery.

Therefore, each image task to be learnt used different sets of polygons. The supervised reflectance reconstruction task used the 57 polygons described in Section 4.4.2, the semantic segmentation task used 83 polygons and the self-supervised radiance reconstruction does not require polygons for training.

The polygons for the segmentation task were split to an 80/20 ratio, with 68 polygons for model training and the remaining 15 polygons for testing.

Polygons to masks for MTL

Depending on the task, each polygon either contains a semantic value for segmentation, or hyperspectral signature for reconstruction. The HS-UNet-R were trained with segmentation

maps that contain a one-to-one mapping of pixels encoded with a semantic value, and also continuous hyperspectral measurements, if considering the supervised reflectance recovery task.

Masks used for training HS-UNet-R were created using the geographic coordinates stored in each polygon and converting real-world coordinates for each vertex to image-coordinates. If a polygon fits within an image, then the candidate image was sampled into 256×256 image tiles centered on labelled sections of the image. By cropping images centered on polygons the edges of each image had a number of pixels that were not labelled. In turn, this allowed the use of semi-supervised methods described in Section 3.3. This process generated one image per polygon.

Therefore, the MTL framework had three datasets. The semantic segmentation task had 75 images, with 60 for model training and the remaining 15 images for testing. The hyperspectral reflectance recovery had 57 images and the self-supervised dataset had 200, as described in Section 4.4.2.

Loss functions

For each dataset and task, different loss functions were used to optimise the network.

The segmentation task considers two training modes: a supervised mode using equation 3.3 and semi-supervised mode using a teacher-student architecture, see Figure 3.10, with equations 3.3 and 3.4. As mentioned, the mapping results in networks trained using MTL only consider the supervised loss with equation 3.3 to avoid bias from the teacher network. The results for the consistency-based semi-supervised setting in Section 4.5.3 combined both losses in the same way as equation 3.5. The parameter γ was set to 0.1 and the weights, w, were computed using equation 3.1 and the class distribution of labelled pixels. Furthermore, for teacher predictions the confidence threshold was set to 0.97.

The hyperspectral reconstruction task also considered two training modes depending on the reconstruction objective. The supervised setting for hyperspectral reflectance reconstruction

used equation 4.6 and the self-supervised setting for radiance reconstruction used equations 4.8 and 4.9. Given the main objective was to map the pioneering species on shingle and sand sediment, the hyperspectral reconstruction task was viewed as auxiliary. Therefore, for networks optimised with MTL, the following loss functions were used depending on the hyperspectral reconstruction task.

If the goal was to apply MTL for supervised segmentation with supervised reflectance recovery, then the loss function combines Eqs. 3.3 and 4.6 as follows:

$$L = wL_{seg} + \alpha L_{recon} \tag{4.15}$$

Where, w was computed using equation 3.1 and the class distribution of labelled pixels, L_{seg} was the supervised segmentation loss (Eq. 3.3), L_{recon} was the supervised hyperspectral reflectance loss (Eq. 4.6) and α was set to 0.1.

If the goal was to apply MTL for supervised segmentation with self-supervised radiance recovery, then the loss function combines Eqs. 3.3, 4.8 and 4.9 as follows: were respectively scaled down by a factor of 100 and ten.

$$L = wL_{seg} + \alpha L_{recon} + \beta L_{smo} \tag{4.16}$$

Where, w was computed using equation 3.1 and the class distribution of labelled pixels, L_{seg} was the supervised segmentation loss (Eq. 3.3), L_{recon} was the self-supervised hyperspectral reflectance loss (Eq. 4.8) with α set to 0.1 and L_{smo} was the smoothing loss (Eq. 4.9) with β set to 0.01.

The hyper-parameters α and β were chosen through an exhaustive search of various settings and monitoring the convergence with loss plots and accuracy metrics.

HS-UNet-R implementation and training parameters for MTL

The details for the HS-UNet-R were described in Section 4.4.2. Figure 4.8 shows the encoder network for the U-Net version in Figure 4.9. As described in Shi et al. [2018], the number of feature maps generated from each convolutional layer was consistent, with 64 feature maps per convolutional layer. The experiments in Section 4.4 with the HS-UNet-R also had a consistent number of feature maps generated per convolutional layer in the encoding network, see Figure 4.9. Generally, in computer vision the number of feature maps expands by a factor of two from each encoding stage [Simonyan and Zisserman, 2014; He et al., 2016, 2017; Long et al., 2015; Ronneberger et al., 2015]. Therefore, for the MTL framework, the implemented HS-UNet-R increased the number of feature maps for each encoding stage, then decreased with each decoding stage.

The MTL architecture shared the weights of the HS-UNet-R for each task to be learnt but each image task had separate input and output layers. A full epoch consisted of forward passing the segmentation dataset, followed by one of the versions of the hyperspectral reconstruction dataset depending on the MTL framework, see Figure 4.20. Each network architecture was trained for 200 epochs with a batch-size of six using Adam optimiser [Kingma and Ba, 2014]. The HS-UNet-R had a depth of five with the following feature map expansion after each encoding stage: 64, 128, 256, 512 and 512. The learning rate was initially set to 0.001. All HS-UNet-Rs were implemented and trained using Pytorch version 10.2.

OBIA

The results from the OBIA method were provided by Cefas. As in Section 3.3, the OBIA method for modelling multiple coastal features was achieved using eCognition v9.3 [Benz et al., 2004].

The first step in OBIA was to process the orthomosaic using a multi-resolution segmentation algorithm to partition the image into segments [Benz et al., 2004]. The scale parameter

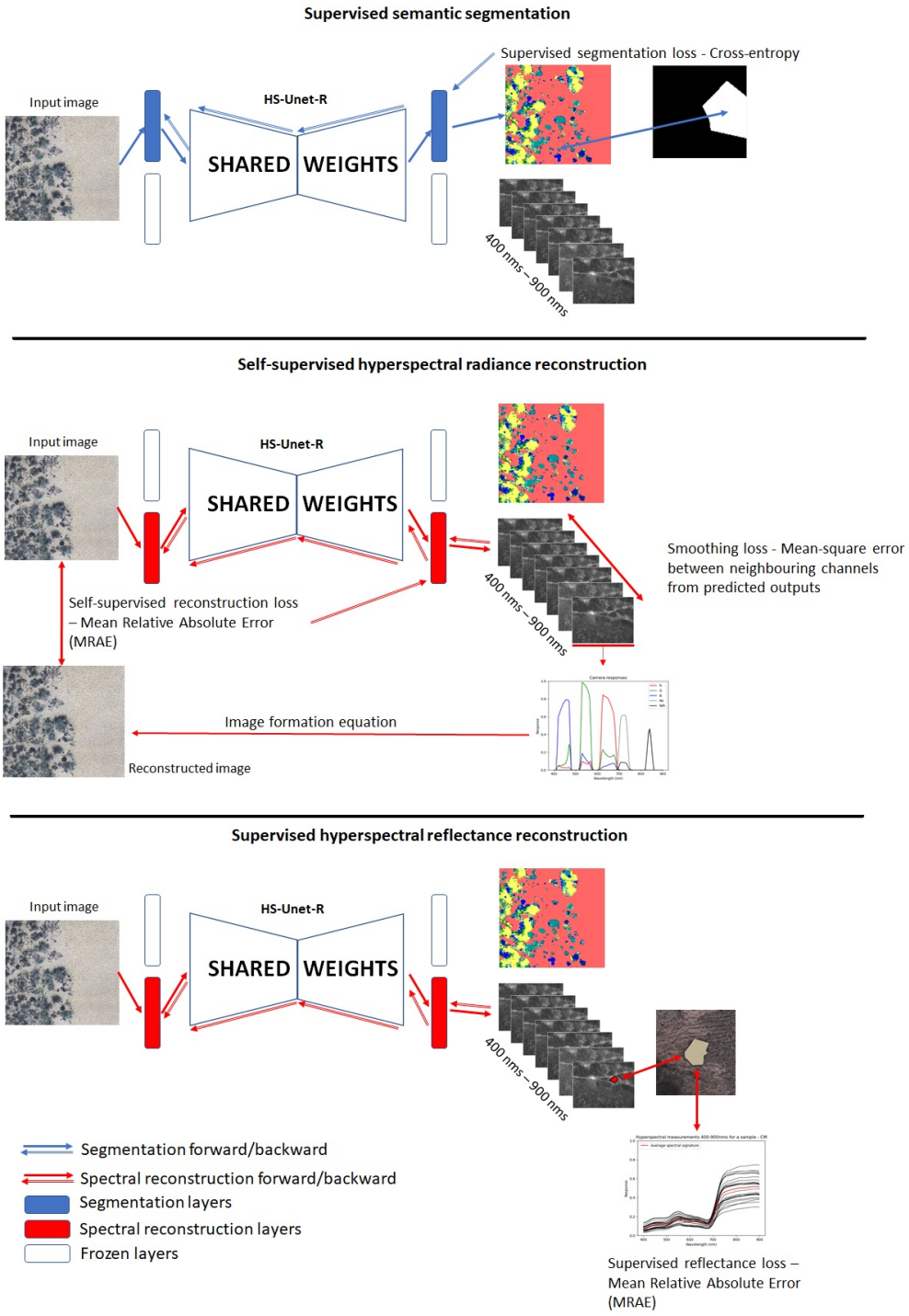

Figure 4.20: The MTL architecture used for the results shown in Section 4.5.3. The shared model is the HS-UNet-R and the proposed MTL architecture alternates between each task to be learnt by forward passing the segmentation dataset, followed by one of the versions of the hyperspectral reconstruction dataset.

dictated whether two adjacent image-objects were fused, as given in equation 3.3.1. If the criterion exceeds the scale parameter value, then the fusion was not performed. In contrast, if the criterion was below the scale parameter value, then both candidates were clustered to form a larger region. The segmentation procedure stops when no further fusions were possible without exceeding the scale parameter. The geometry of each image-object was defined by two other hyper-parameters known as shape and compactness. For the results in Section 4.5.3, the scale parameter was set to 50, the shape to 0.2 and the compactness to 0.5. The input layers for the segmentation process were Red, Green and Blue bands, the Red Edge band and Near-Infrared, normalised difference vegetation index (NDVI) and the digital surface model (DSM) [Private communication, Arosio, 2021]. Figure 4.21 shows image objects overlaid on top of a crop of the study site.

In Section 3.3, the use of polygons derived from the in-situ survey were superimposed on top of image-objects to select the candidate segments for extracting features to train shallow machine learning models. However, for the results provided by Cefas in Section 4.5.3, segmented image-objects were manually selected to create a dataset for the in-built Random Forest in eCognition [Breiman, 2001]. The manual selection was a combination of expert photo-interpretation and accurate transcription of data points from the site survey. As mentioned in Section 4.2.1, the target class domain for the OBIA method was further reduced by joining Rumex crispus and Glaucium flavum labels to a single class called pioneering grassland, and merging Silene uniflora, Lathyrus japonicus and Honckenya peploides labels to another single class named young pioneering species [Private communication, Arosio, 2021]. For each selected image-object, statistical moments, such as channel mean and standard deviation, NDVI, ratios between Red/Blue, Red/Green and Blue/Green image layers and the DSM were used to train and validate the Random Forest modeller [Breiman, 2001].

4.5.3 Quality assessment

The objective metrics discussed in Section 4.5.4 will be pixel accuracy, precision, recall and F1-score. As mentioned in Section 3.3.2, pixel accuracy measures the ratio between pixels

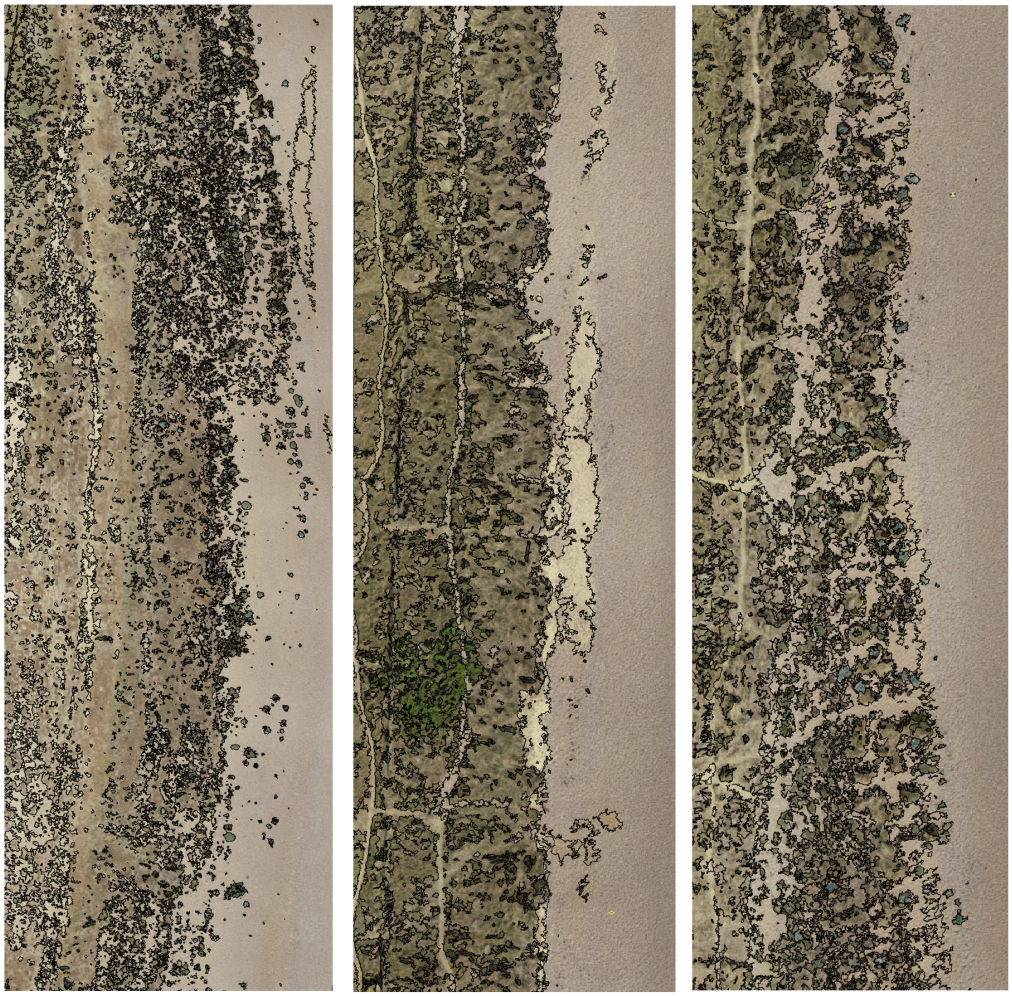

Figure 4.21: A close-up of cropped orthomosaics of Sizewell beach segmented using the multiresolution segmentation algorithm for Red, Green and Blue channels. The scale, shape and compactness were respectively 50, 0.2 and 0.5.

that were classified correctly and the total number of labelled pixels in the test set for a given class. Precision and recall are metrics that can show how a classifier performs for each specific class. F1-score is the harmonic mean of recall and precision and is a suitable metric to quantify classifier performance when a single figure of merit is needed. Equation 3.7 shows each objective metric. The analysis of results will be split into two threads of discussion.

First, an analysis of results for FCNs trained with different optimisation strategies in order to achieve fine-grained specie level mapping, as noted with the target class domain stated in Section 4.2.1. Figure 4.22 shows the VHR orthomosaic generated from the aerial survey described in Section 4.2.1 with an overlay of each part of interest used for this particular discussion. Figure 4.23 show cropped close-ups of the orthomosaic used in the visual analysis.

As mentioned, the results from the OBIA with a further reduced target class domain which creates a discrepancy between the objective and visual results with the listed species in Table 4.2 and the discussion in Section 4.5.4. Therefore, while the discussion in Section 4.5.4 points to trends and patterns found in objective scores and generated thematic maps for each optimisation strategy, Section 4.5.5 provides a one-to-one comparison by merging predictions obtained with FCNs using the heuristics described in Section 4.2.1 for defining the mapping schema used for OBIA. Figure 4.24 shows the VHR orthomosaic along with the cropped regions of interest used for the visual analysis of results.

Experiments and Results

The mapping outputs of the HS-UNet-R trained in multiple settings were compared with the 15 polygons that were not used for training. Figure 4.25 show confusion matrices scoring outputs from each MTL setting as pixel accuracy. The confusion matrices also show pixel accuracies for models without MTL that were optimised using equation 3.3 and models that



Figure 4.22: The generated orthomosaic from the aerial survey projected to RGB colour space. The multispectral orthomosaic is narrow which prevents the visual analysis of generated thematic maps from each method described in Section 4.5.2.

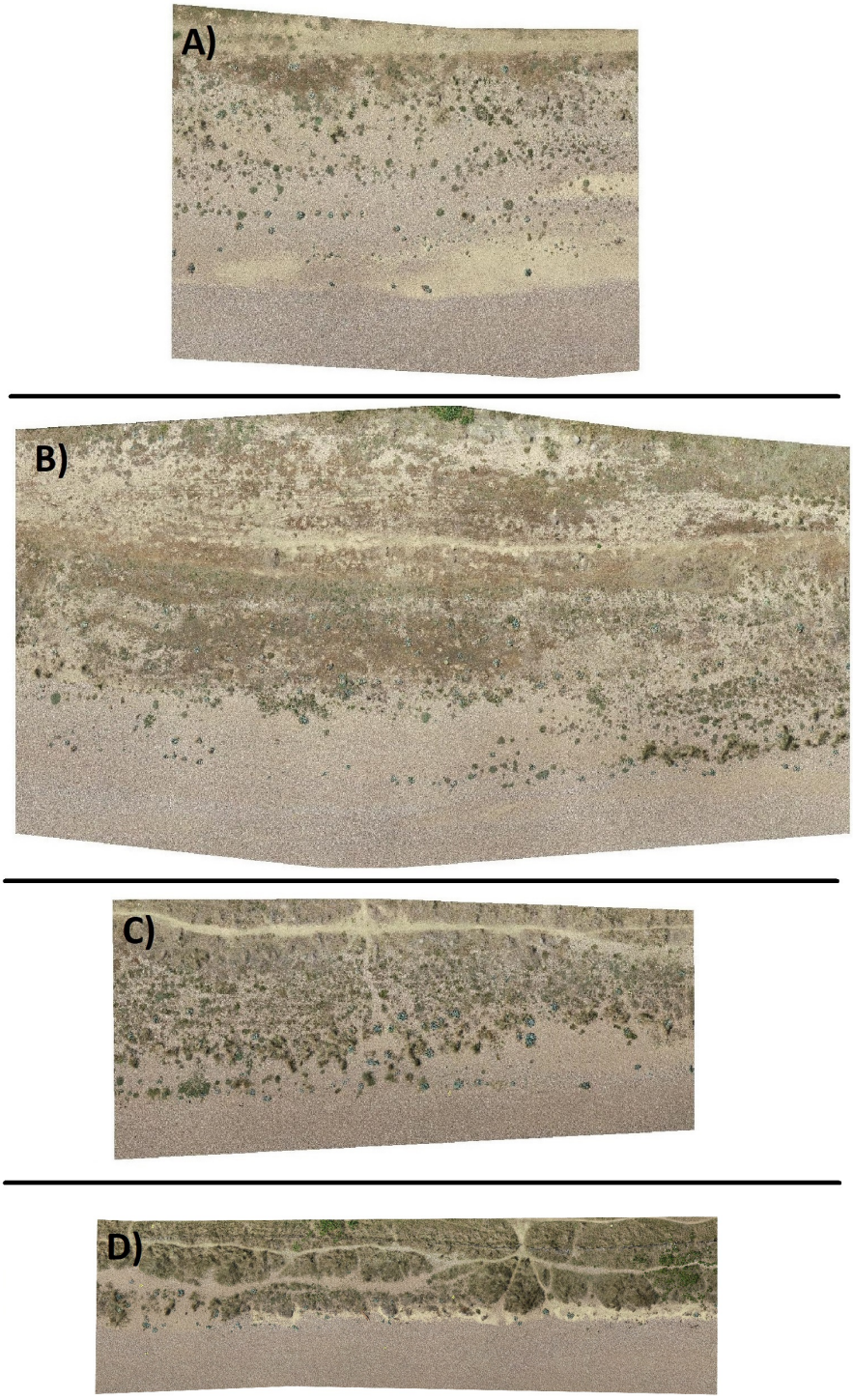

Figure 4.23: Cropped close-ups of the orthomosaic used for the visual analysis of FCNs optimised using various strategies as shown in Sections 3.3 and 4.5.

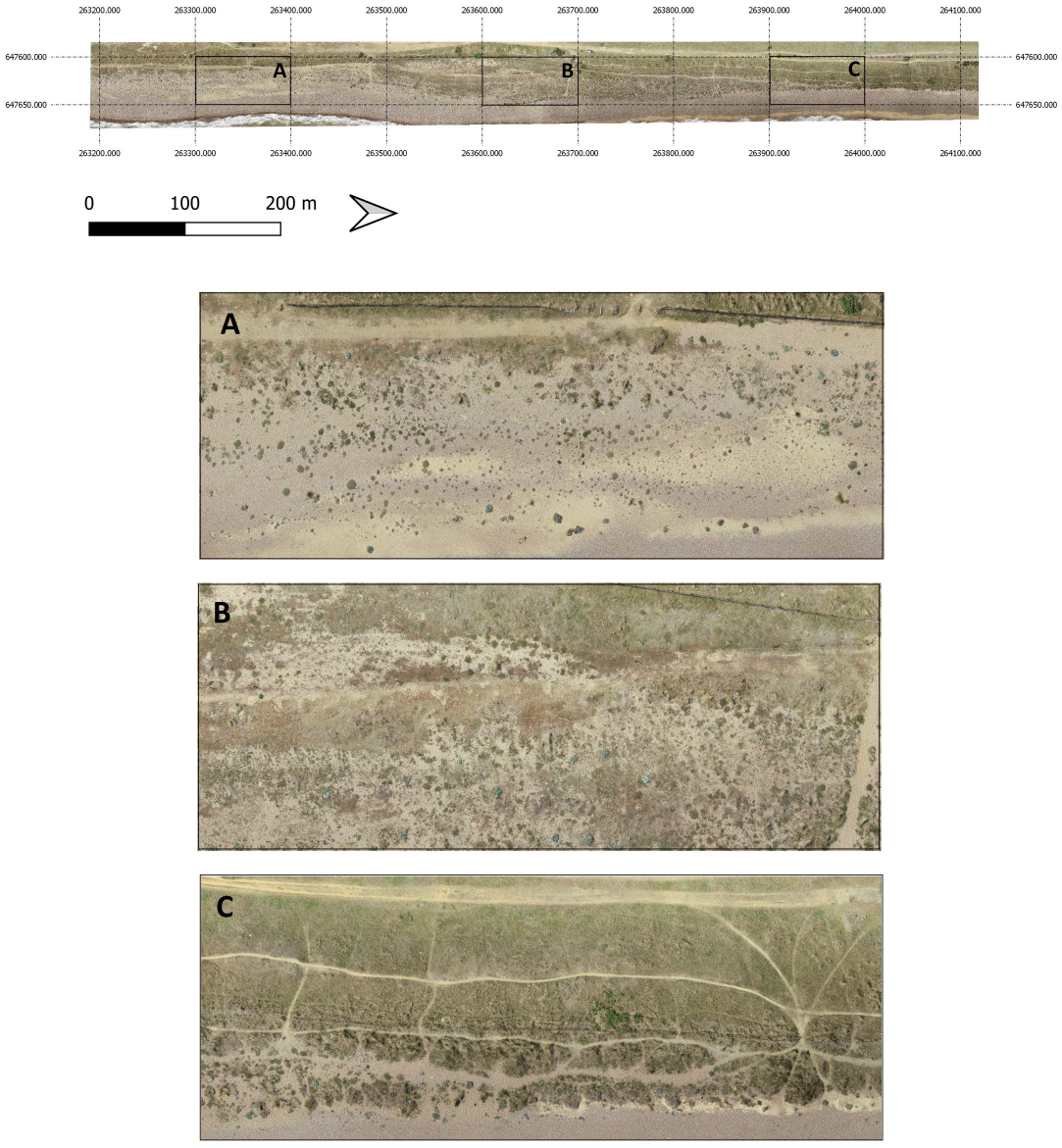

Figure 4.24: Orthomosaic of Sizewell beach with geographical coordinates. The figures **A**, **B** and **C** show close-ups of the shingle beach along with assemblages of vegetated shingle communities used for the comparison of FCNs with the OBIA. For the display, the Red, Green and Blue image bands were used.

	Supervised			Semi-supervised			MTL: supervised			MTL: self-supervised		
	R	**P**	**F1**	**R**	**P**	**F1**	**R**	**P**	**F1**	**R**	**P**	**F1**
Shingle	0.933	0.994	**0.963**	0.907	0.991	0.947	0.891	0.984	0.935	0.775	0.974	0.863
Sand	0.997	0.929	**0.962**	0.997	0.897	0.945	1	0.857	0.923	1	0.767	0.868
Ammophila A	0.583	0.927	0.716	0.950	0.936	**0.943**	0.988	0.708	0.825	0.987	0.706	0.823
Hockenya P	0.958	0.294	0.450	0.703	0.700	0.702	0.764	0.747	**0.752**	0.437	0.248	0.317
Lathyrus J	0.006	0.208	0.0135	0.055	0.220	**0.089**	0.036	0.371	0.066	0.002	0.153	0.005
Glaucium F	0.906	0.174	0.292	0.891	0.490	**0.632**	0.702	0.480	0.570	0.631	0.482	0.546
Silene U	0.968	0.694	0.809	0.951	0.834	**0.889**	0.709	0.888	0.789	0.727	0.830	0.775
Crambe M	0.672	0.983	0.798	0.868	0.965	0.914	0.933	0.948	**0.940**	0.889	0.960	0.923
Eryngium M	0.871	0.121	0.211	0.842	0.163	0.273	0	0	0	0.457	0.263	**0.334**
Rumex C	0.069	0.489	0.122	0.158	0.500	0.240	0.313	0.565	**0.403**	0.215	0.788	0.338
Avg. score	0.702	0.532	0.533	0.732	0.664	0.563	0.633	0.654	0.62	0.612	0.617	0.579

Table 4.5: Recall, precision and F1-scores for supervised, semi-supervised and MTL using the HS-UNet-R. The supervised column shows models optimised with equation 3.3, the semi-supervised column shows models optimised using both equations 3.3 and 3.4. MTL - supervised refers to models trained for supervised segmentation with supervised reflectance recovery. MTL - self-supervised refers to models trained for supervised segmentation with self-supervised radiance recovery.

were optimised using both equations 3.3 and 3.4. The results in Table 4.5 reflect the specie level mapping objective.

For the comparison with OBIA, the predictions for FCNs optimised with 3.3, or both equations 3.3 and 3.4, and FCNs trained in semi-supervision using MTL were merged in the same manner as the heuristics described for the further reduced target class domain used with OBIA. The same subset of 15 rasterised polygons that were not used for training compare both methods jointly. Figure 4.26 display confusion matrices scoring outputs from each method as pixel accuracy. Overall results for OBIA and FCNs in supervised, semi-supervised with consistency regularisation, and semi-supervised with MTL were reported in Table 4.6.

Thematic maps

Figures 4.27 - 4.34 show thematic maps for HS-UNet-R trained in a supervised or semi-supervised setting as described in Section 3.3. Figures 4.35 - 4.42 show thematic maps for HS-UNet-R trained with MTL where the auxiliary image task was supervised reflectance recovery or self-supervised radiance recovery as described in Section 4.5.2.

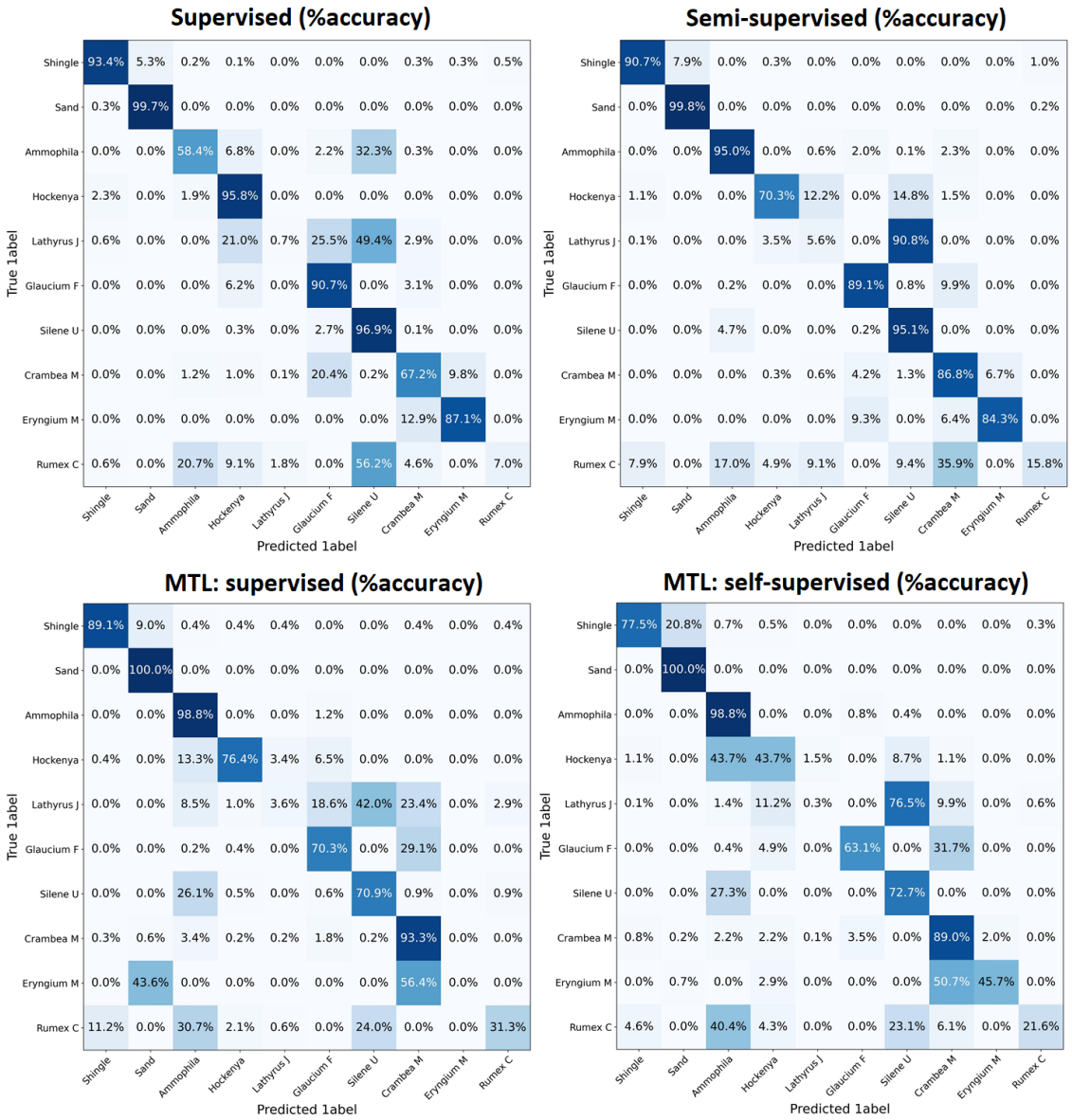

Figure 4.25: Confusion matrices for both supervised, semi-supervised and MTL using the HS-UNet-R. MTL - supervised refers to models trained for supervised segmentation with supervised reflectance recovery. Each confusion matrix for FCNs shows the average pixel accuracy over 5 independent train and test run. MTL - self-supervised refers to models trained for supervised segmentation with self-supervised radiance recovery.

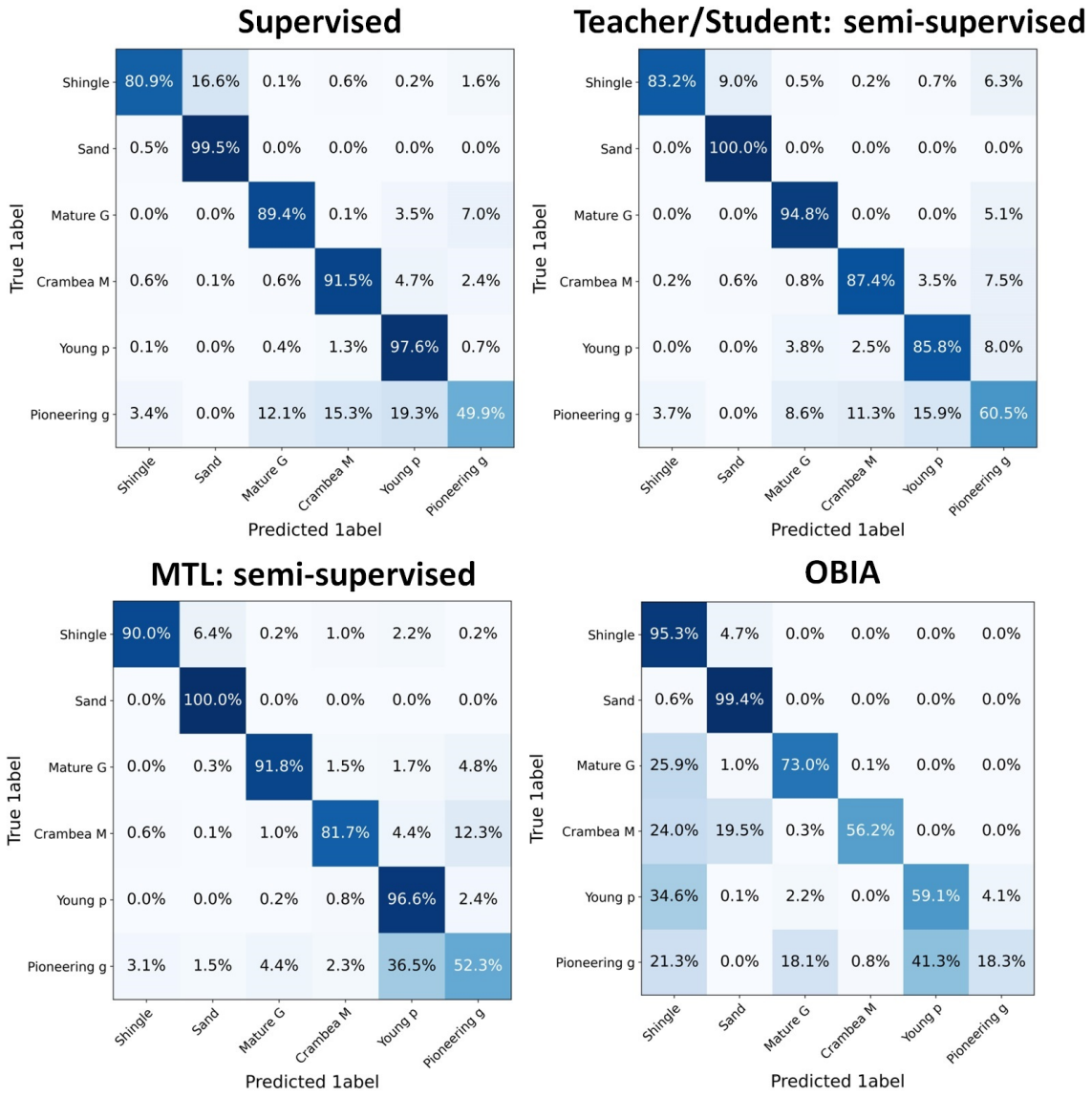

Figure 4.26: Confusion matrices showing pixel accuracy scores for OBIA and FCNs optimised in a supervised, semi-supervised with consistency regularisation and teacher/student networks and semi-supervised multi-task learning. Each confusion matrix for FCNs shows the average pixel accuracy over 5 independent train and test run. Legend: Mature G - Mature Grassland; Pioneering G. - Pioneering grassland; Young P. - Young pioneering; Crambe M. - Crambe Maritima

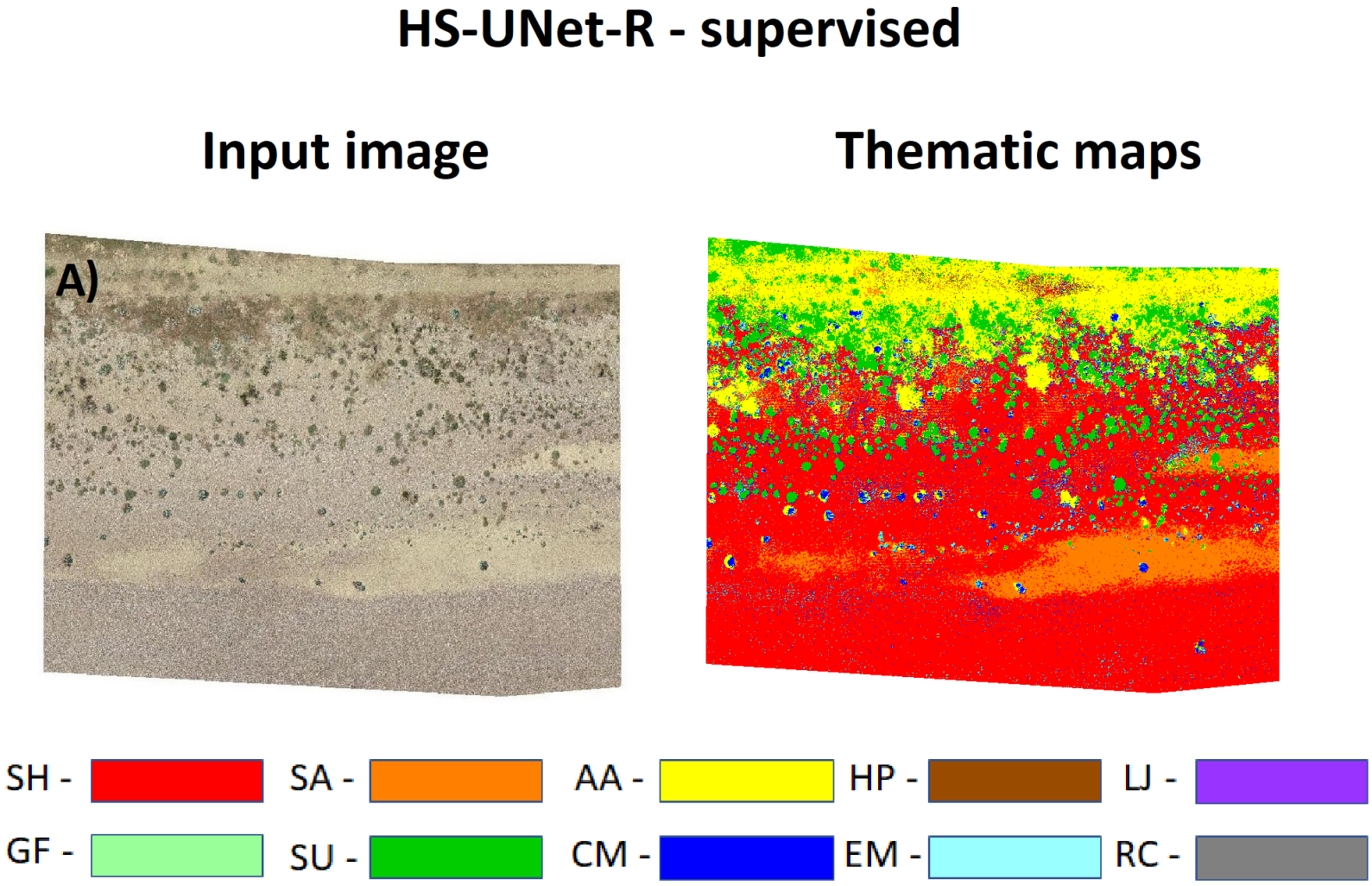

Figure 4.27: Thematic maps for supervised HS-UNet-R without MTL. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora.

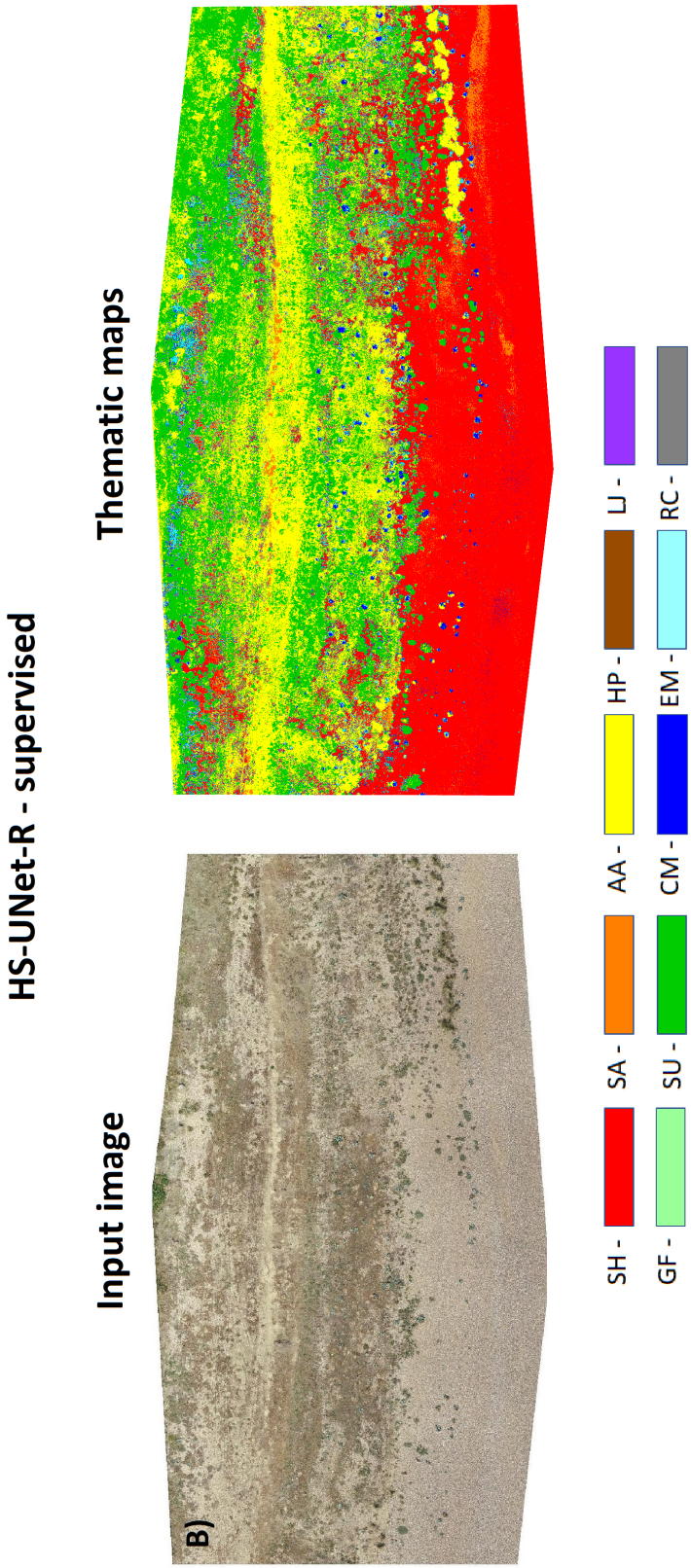

Figure 4.28: Thematic maps for supervised HS-UNet-R without MTL. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora.

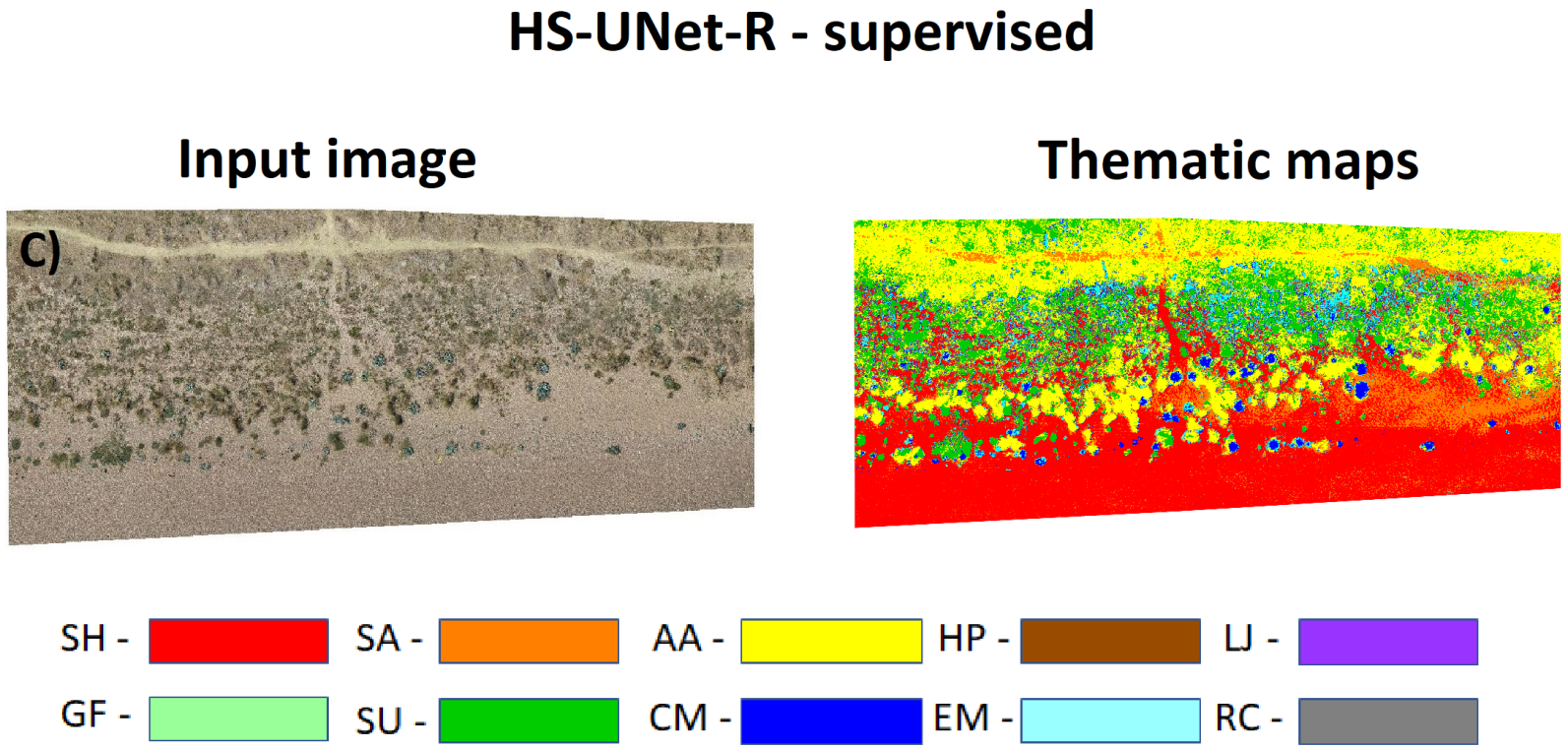

Figure 4.29: Thematic maps for supervised HS-UNet-R without MTL. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora.

Figure 4.30: Thematic maps for supervised HS-UNet-R without MTL. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora.

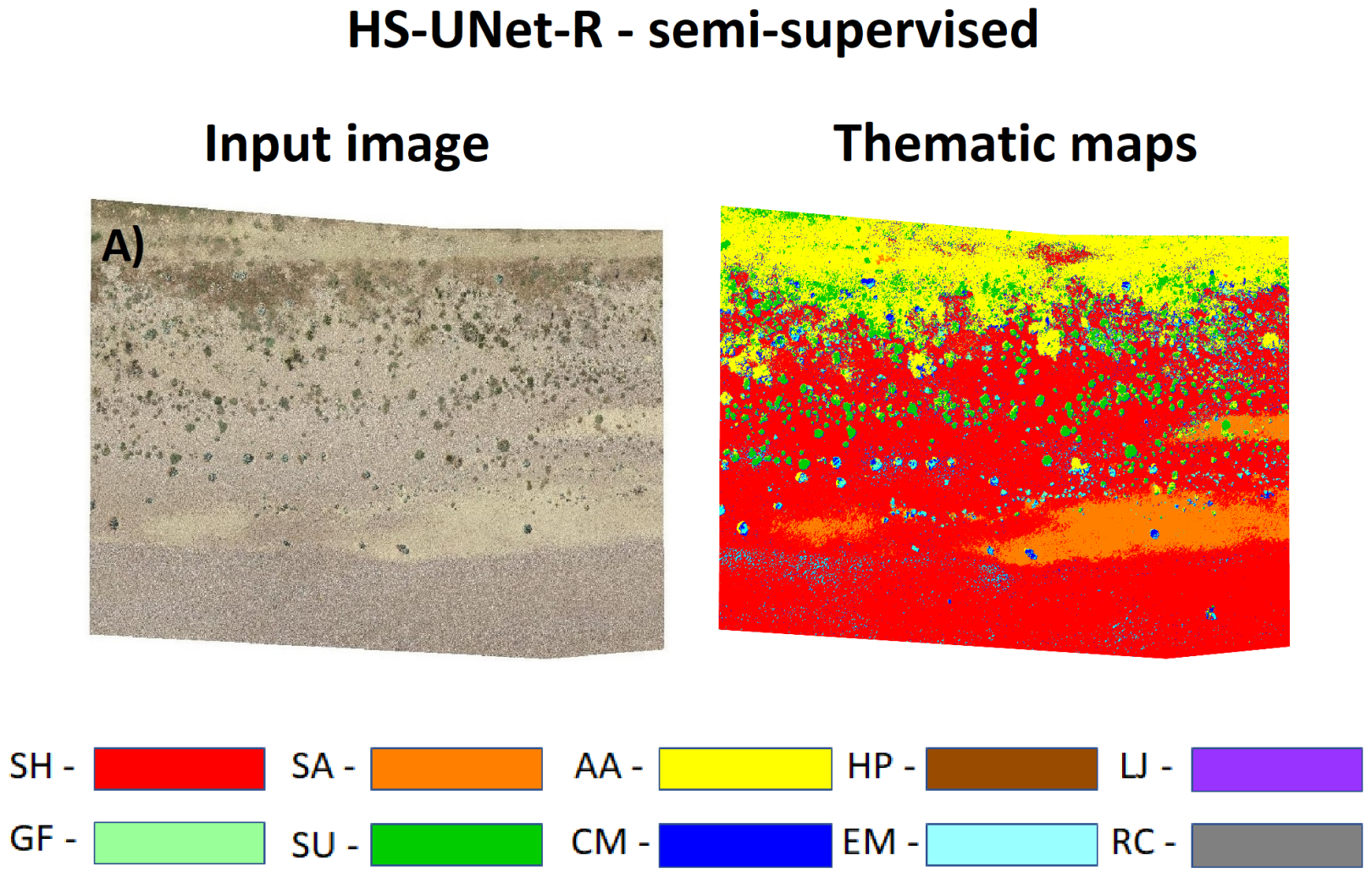

Figure 4.31: Thematic maps for semi-supervised HS-UNet-R without MTL. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora.

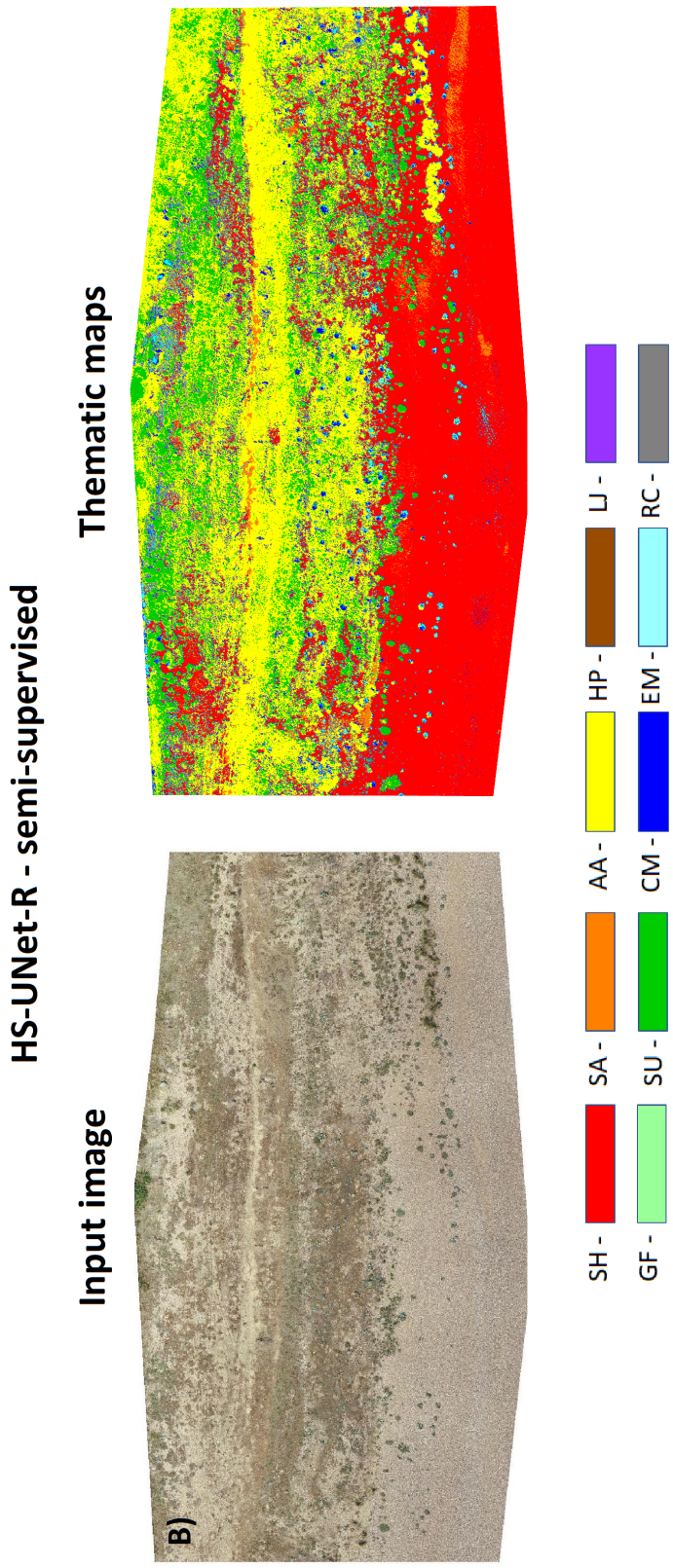

Figure 4.32: Thematic maps for semi-supervised HS-UNet-R without MTL. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora.

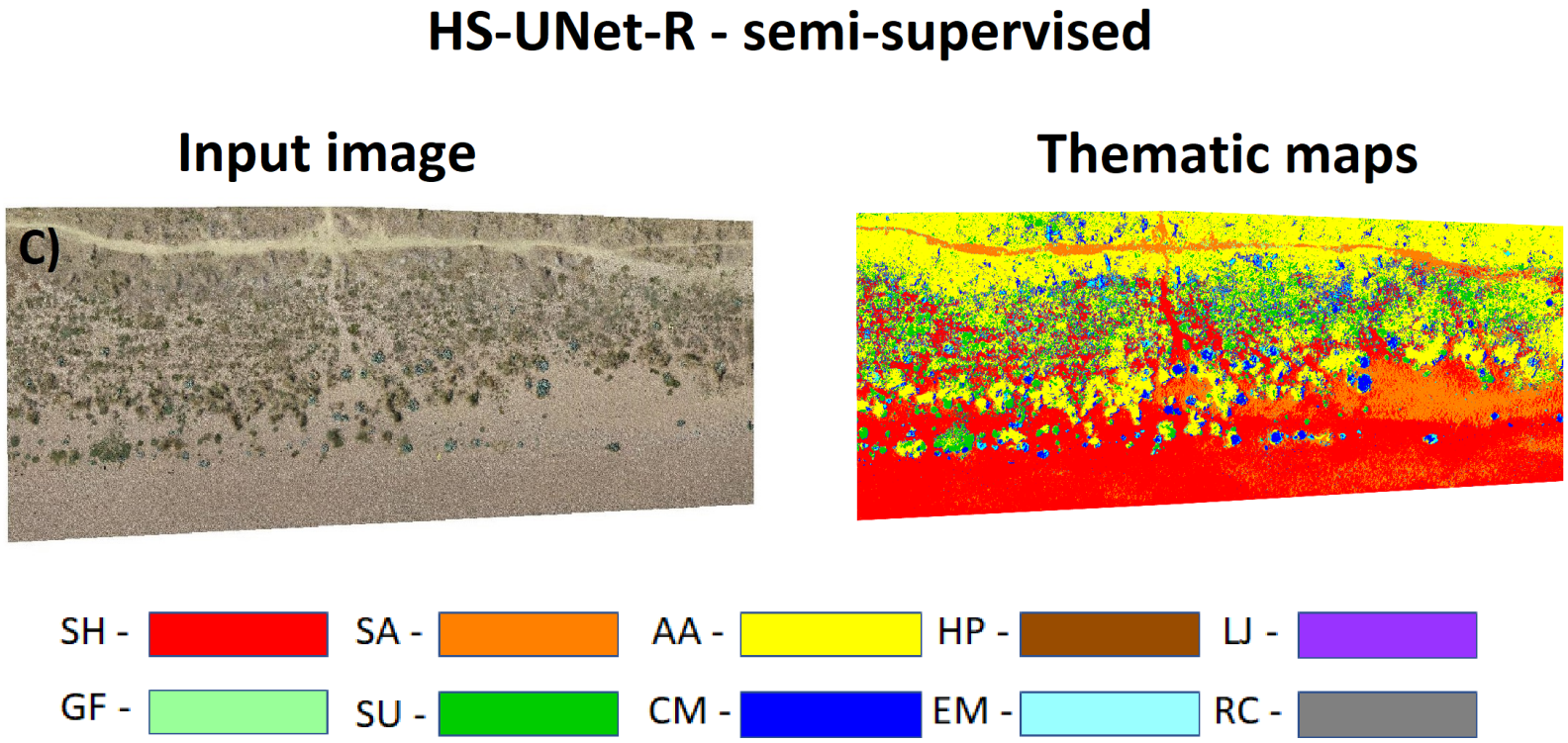

Figure 4.33: Thematic maps for semi-supervised HS-UNet-R without MTL. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora.

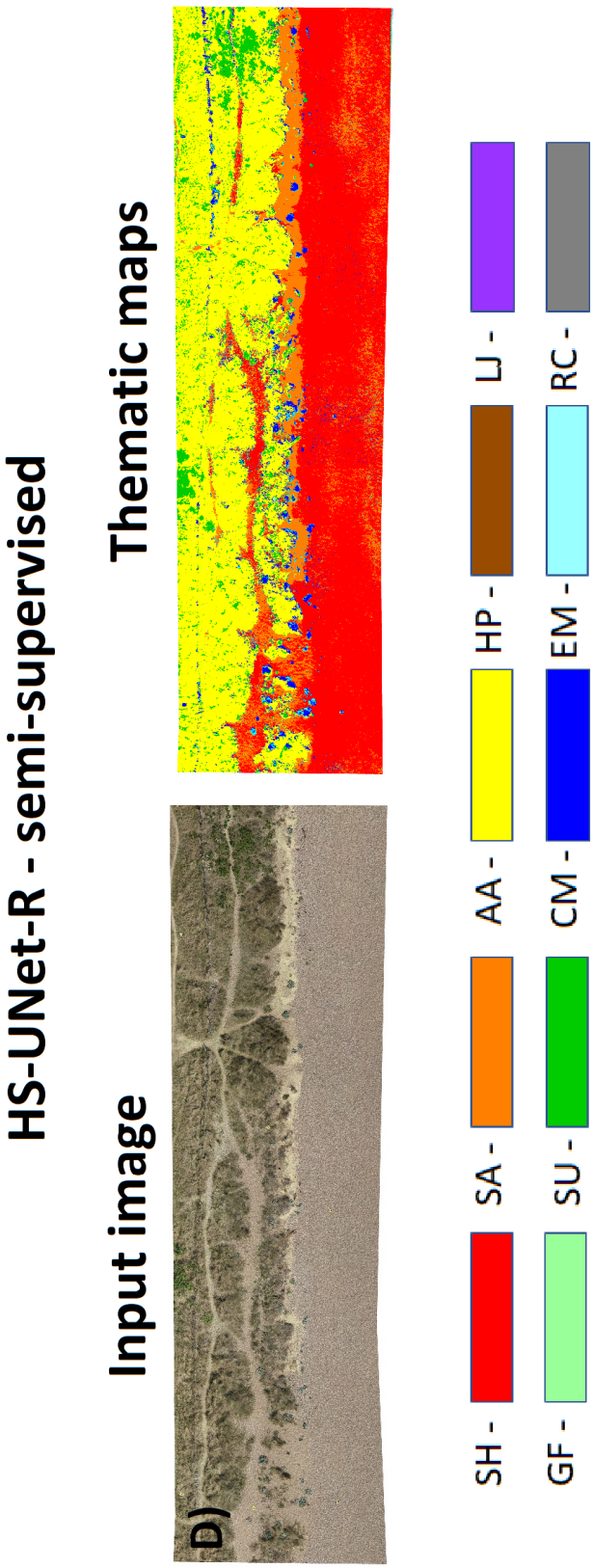

Figure 4.34: Thematic maps for semi-supervised HS-UNet-R without MTL. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora.

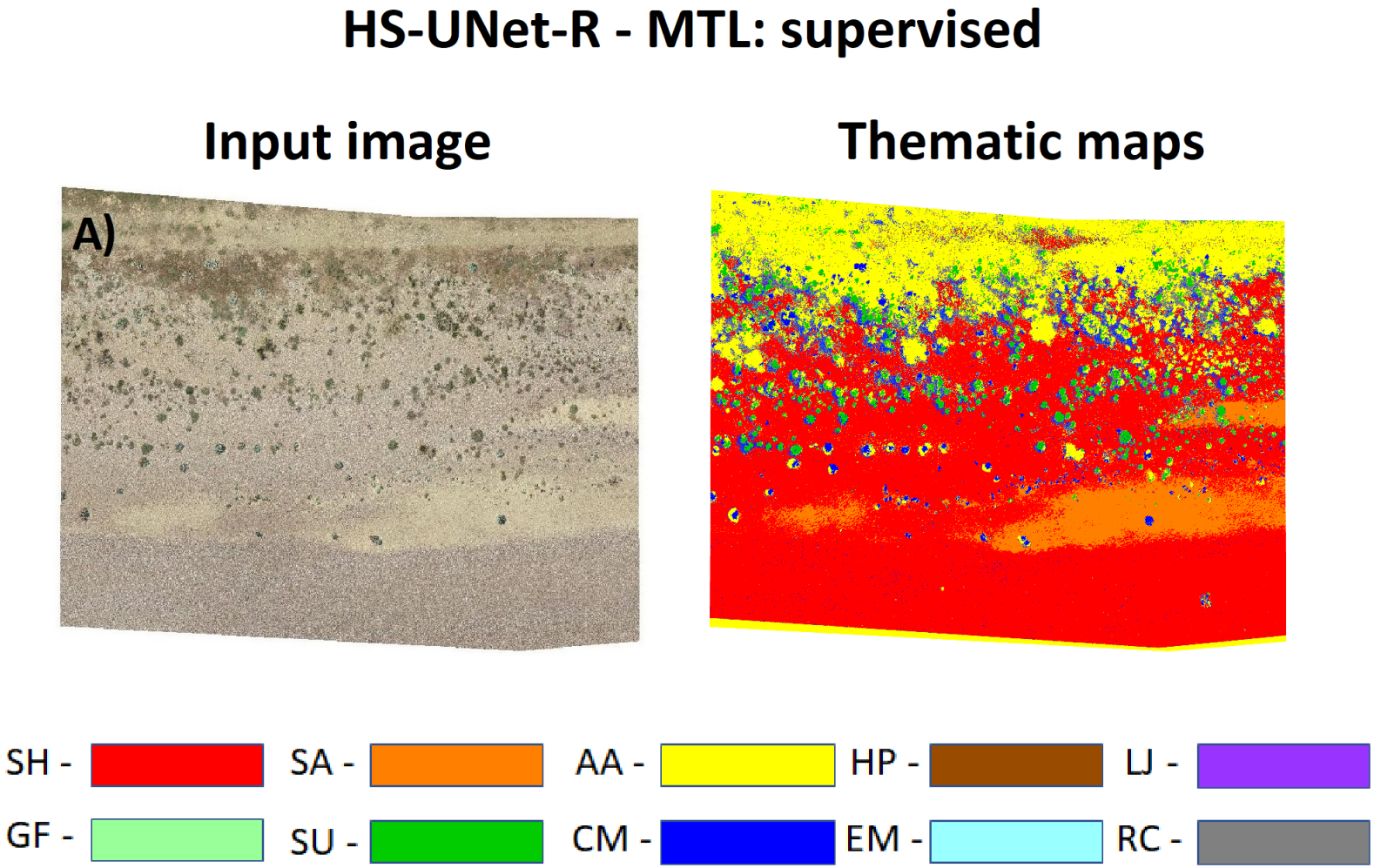

Figure 4.35: Thematic maps for supervised HS-UNet-R with MTL using supervised reflectance as auxiliary. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora.

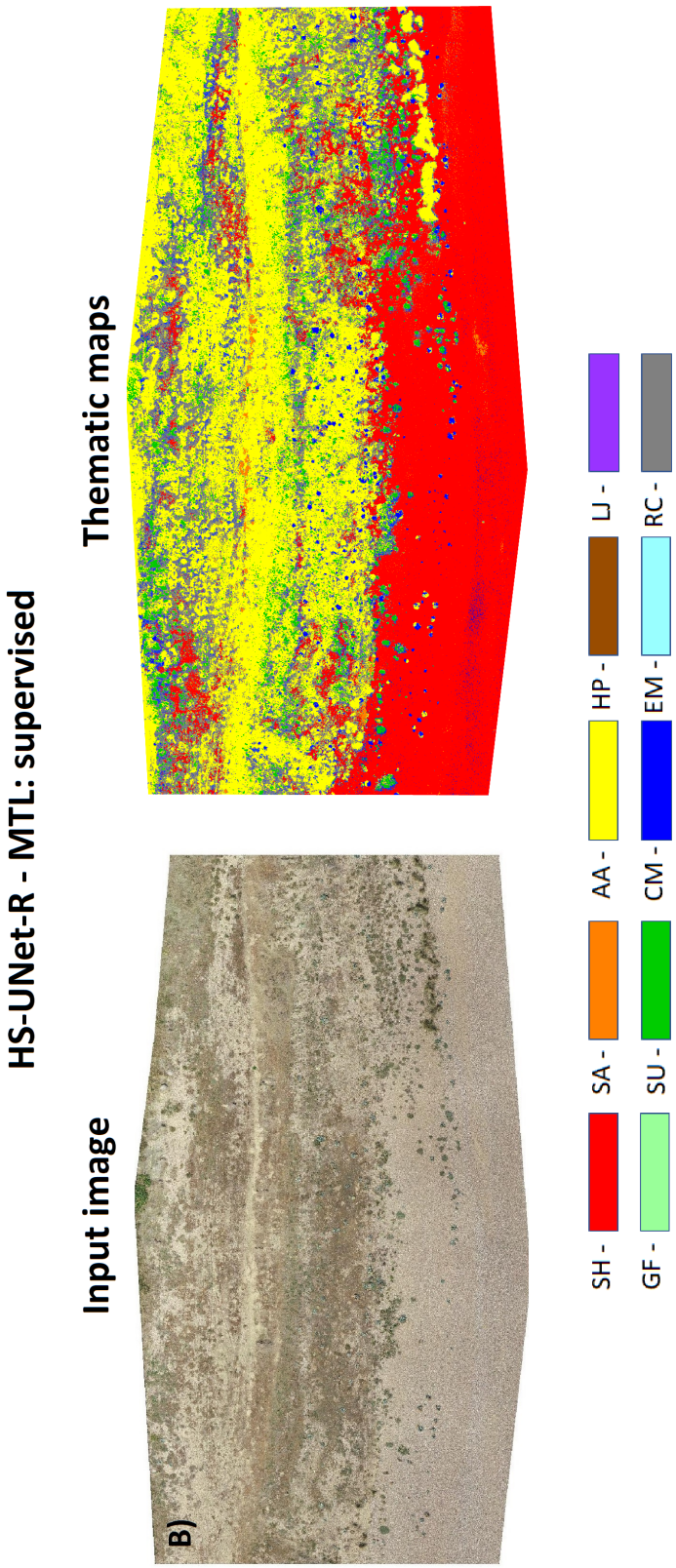

Figure 4.36: Thematic maps for supervised HS-UNet-R with MTL using supervised reflectance as auxiliary. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora.

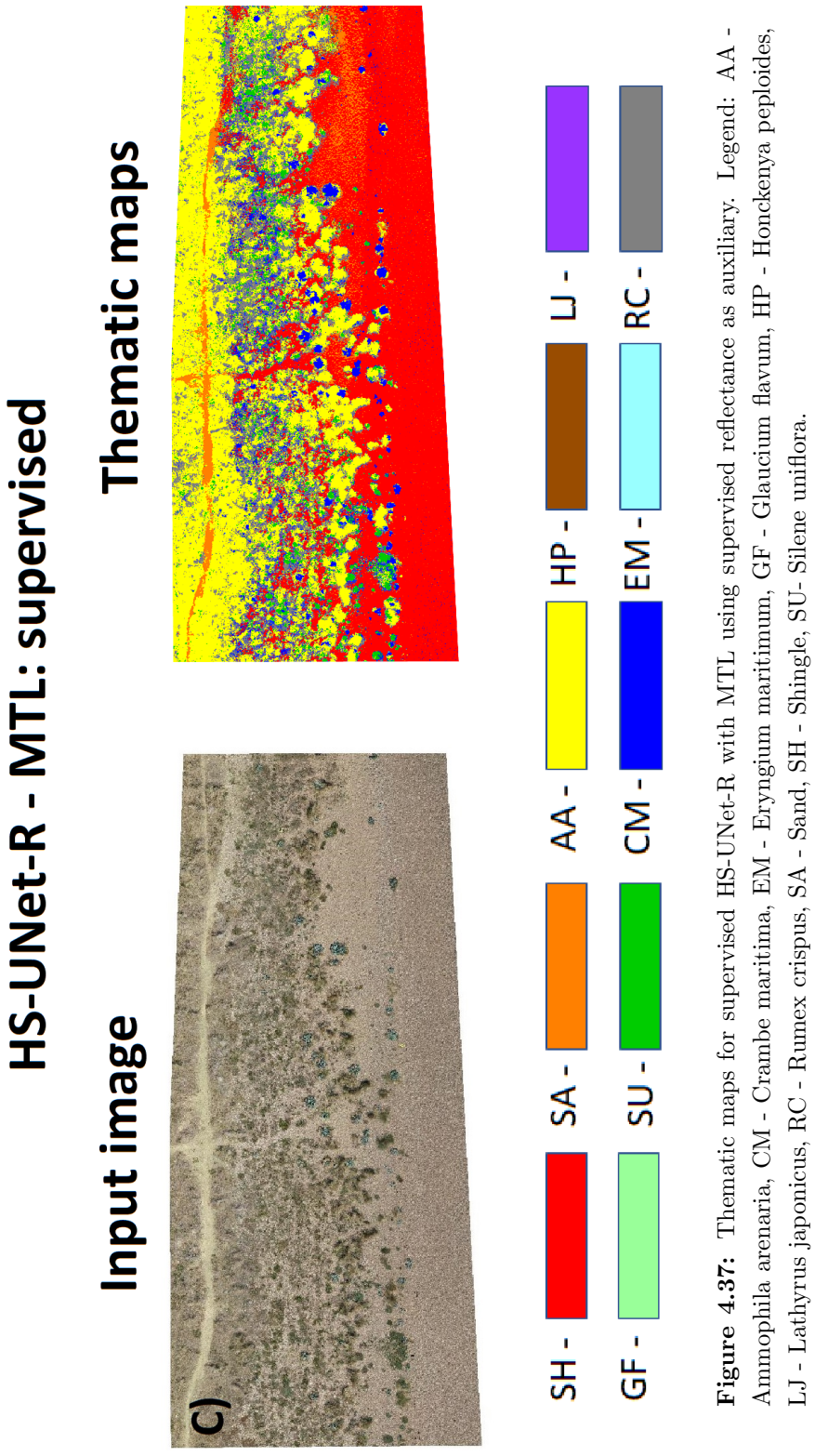

Figure 4.37: Thematic maps for supervised HS-UNet-R with MTL using supervised reflectance as auxiliary. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora.

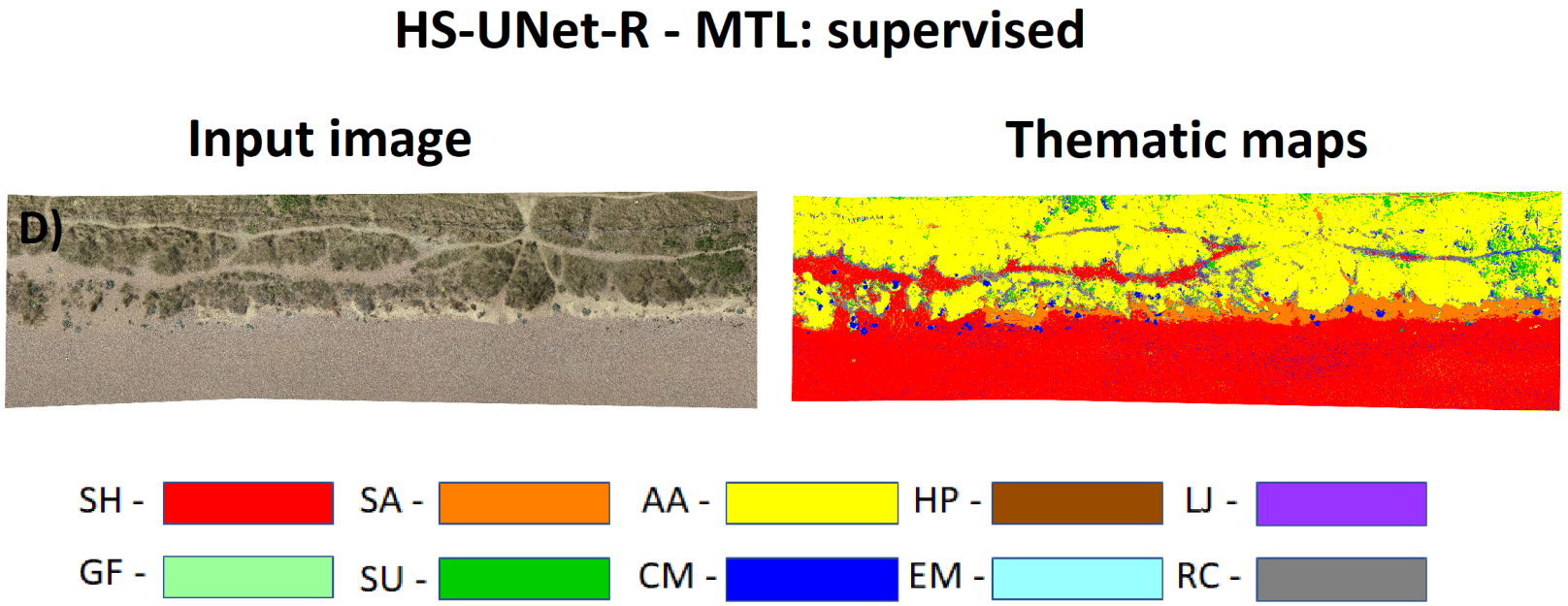

Figure 4.38: Thematic maps for supervised HS-UNet-R with MTL using supervised reflectance as auxiliary. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora.

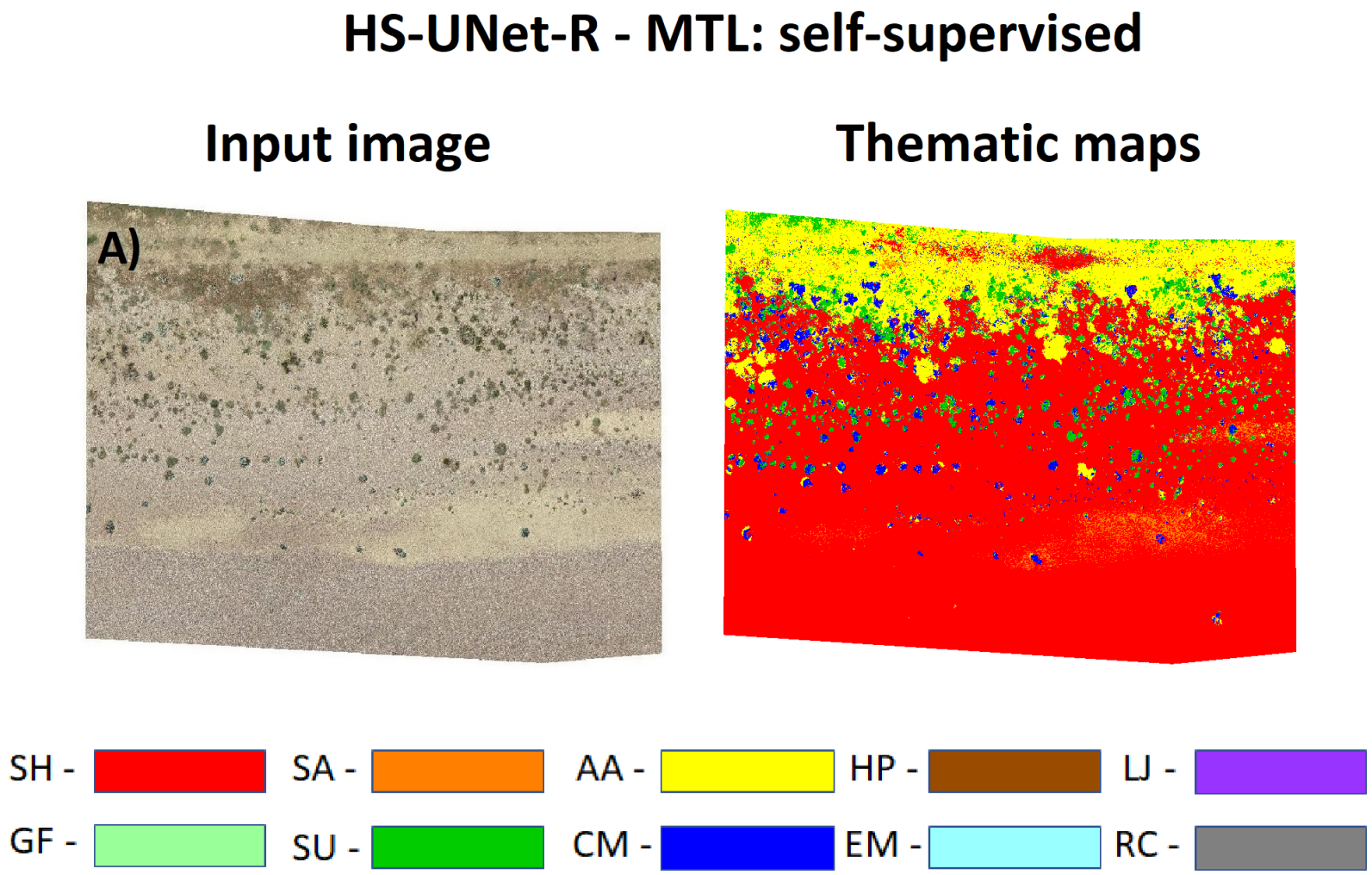

Figure 4.39: Thematic maps for supervised HS-UNet-R with MTL using self-supervised radiance as auxiliary. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora.

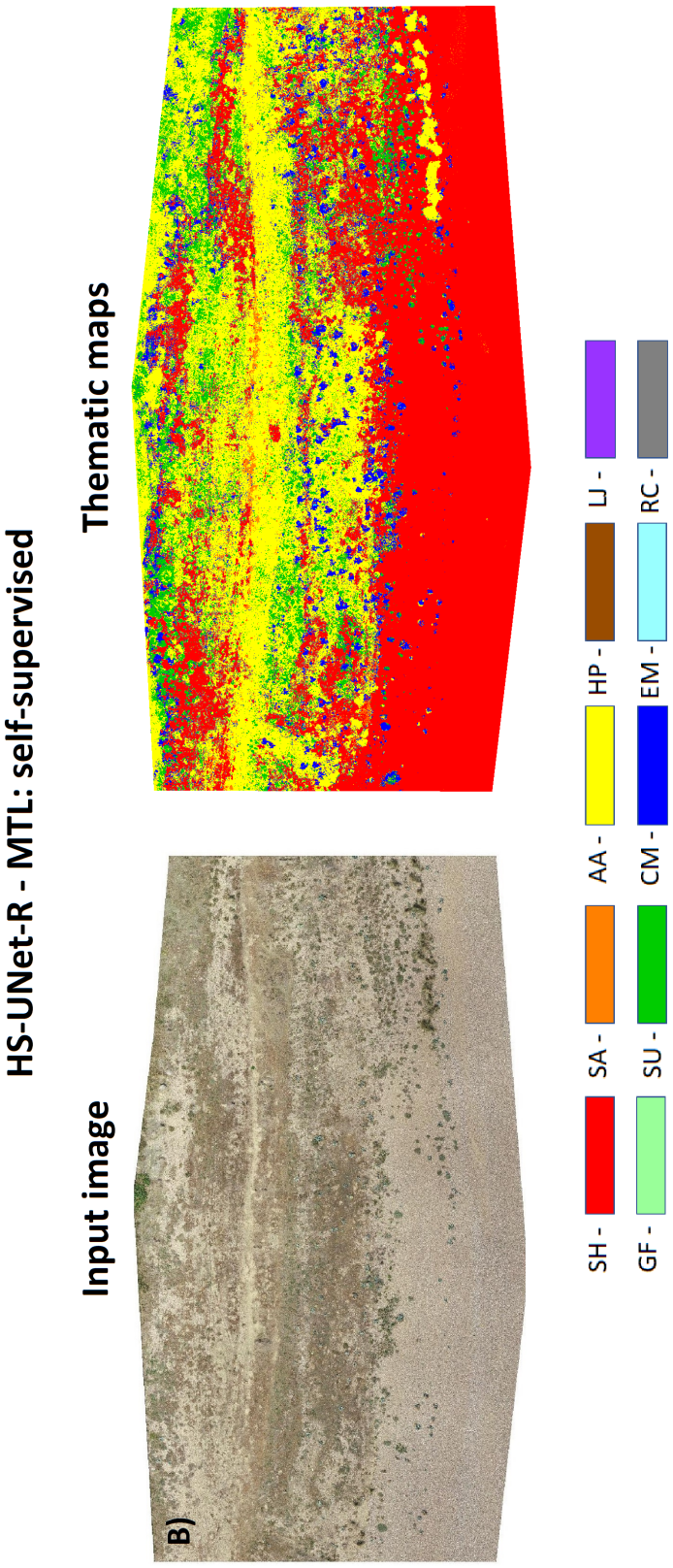

Figure 4.40: Thematic maps for supervised HS-UNet-R with MTL using self-supervised radiance as auxiliary. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora.

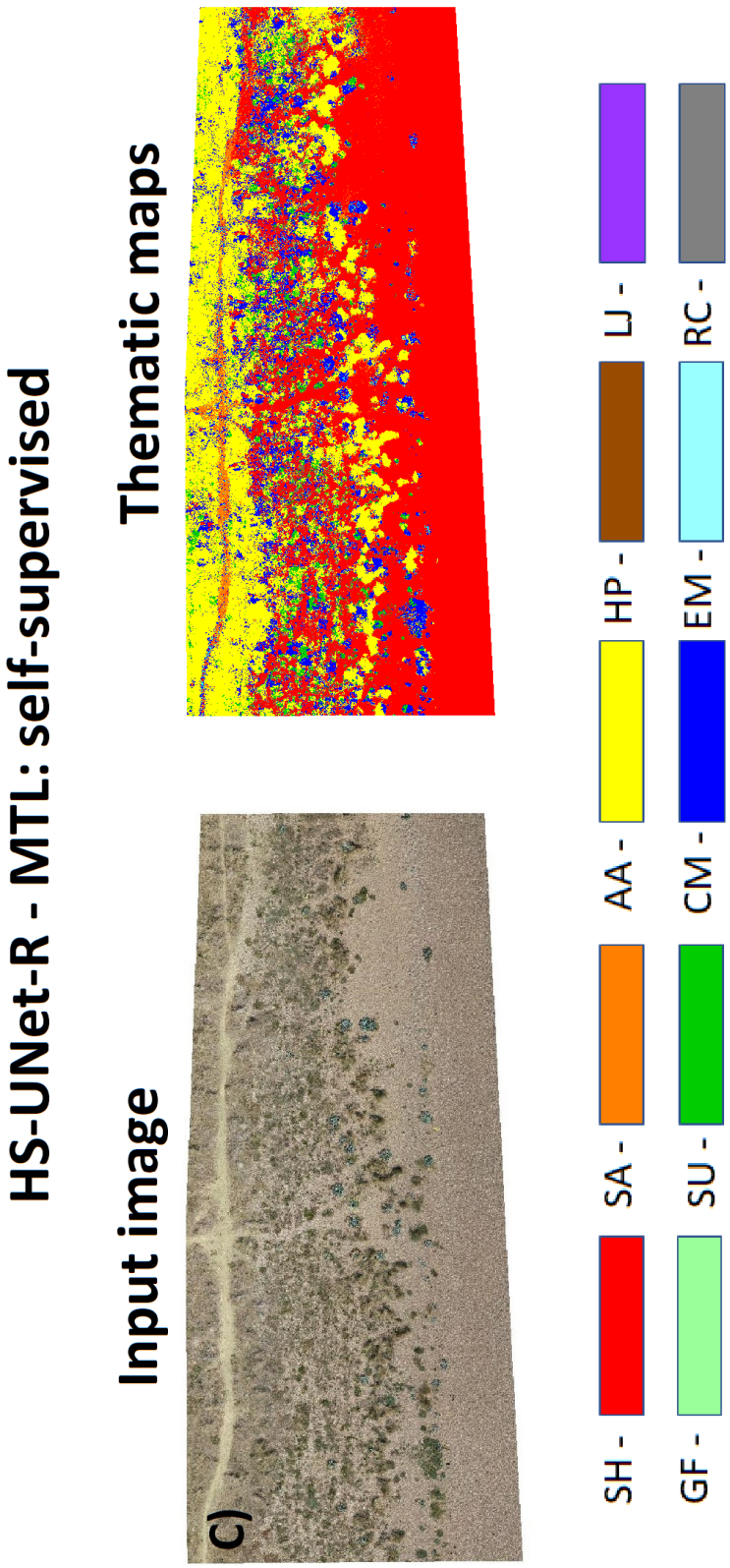

Figure 4.41: Thematic maps for supervised HS-UNet-R with MTL using self-supervised radiance as auxiliary. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora.

Figure 4.42: Thematic maps for supervised HS-UNet-R with MTL using self-supervised radiance as auxiliary. Legend: AA - Ammophila arenaria, CM - Crambe maritima, EM - Eryngium maritimum, GF - Glaucium flavum, HP - Honckenya peploides, LJ - Lathyrus japonicus, RC - Rumex crispus, SA - Sand, SH - Shingle, SU- Silene uniflora.

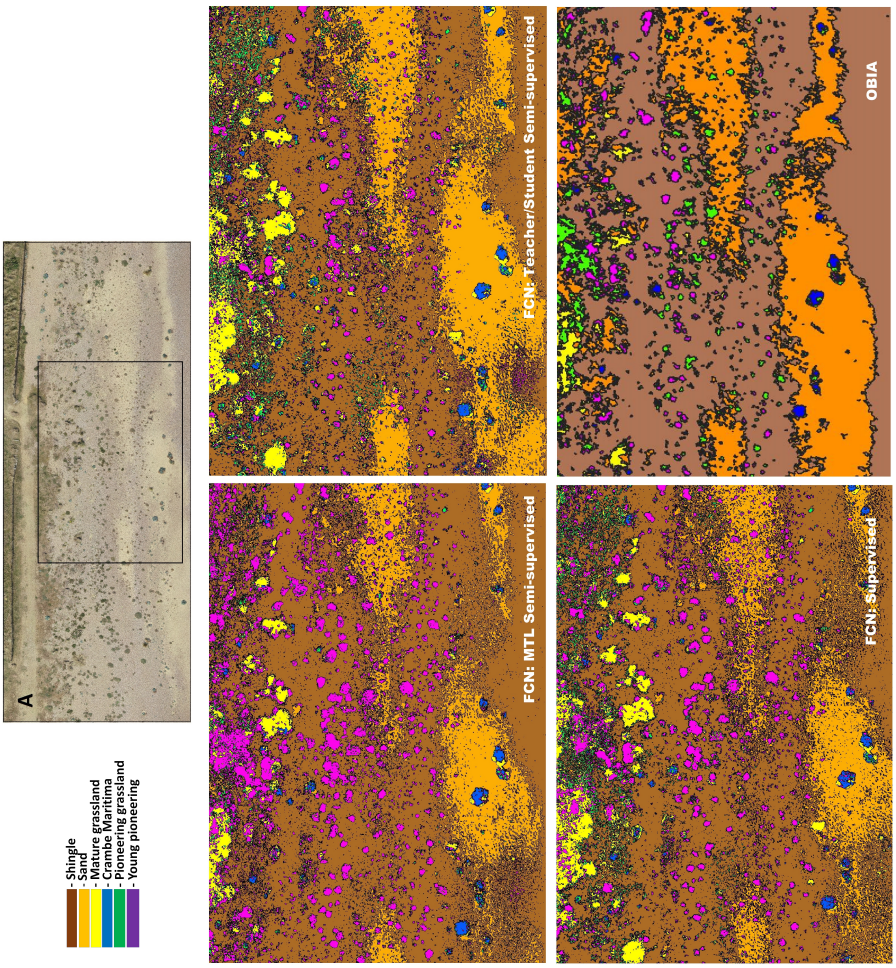

Figure 4.43: Segmented habitat maps and close-up of **A** in Figure 4.24 for each of the mentioned methods in Section 4.5. The visual results were generated with the best performing network from each optimisation strategy. Legend: Mature G - Mature Grassland; Pioneering G. - Pioneering grassland; Young P. - Young pioneering; Crambe M. - Crambe Maritima

Figure 4.44: Segmented habitat maps and close-up of **B** in Figure 4.24 for each of the mentioned methods in Section 4.5. Legend: Mature G - Mature Grassland; Pioneering G. - Pioneering grassland; Young P. - Young pioneering; Crambe M. - Crambe Maritima

Figure 4.45: Segmented habitat maps and close-up of **C** in Figure 4.24 for each of the mentioned methods in Section 4.5. Legend: Mature G - Mature Grassland; Pioneering G. - Pioneering grassland; Young P. - Young pioneering; Crambe M. - Crambe Maritima

	OBIA			Supervised			Semi-supervised			MTL: semi-supervised		
	R	**P**	**F1**	**R**	**P**	**F1**	**R**	**P**	**F1**	**R**	**P**	**F1**
Shingle	0.953	0.387	0.550	0.808	0.971	0.882	0.832	0.986	0.902	0.899	0.981	**0.938**
Sand	0.993	0.589	0.739	0.994	0.805	0.89	1	0.873	0.932	1	0.906	**0.951**
Mature G.	0.729	0.983	0.838	0.894	0.935	0.923	0.948	0.912	0.929	0.917	0.963	**0.941**
Pioneering G.	0.183	0.311	0.230	0.498	0.233	0.317	0.605	0.24	**0.344**	0.522	0.223	0.323
Young P.	0.590	0.934	0.726	0.975	0.881	**0.926**	0.857	0.909	0.883	0.965	0.861	0.911
Crambe M.	0.562	0.997	0.719	0.915	0.972	**0.943**	0.874	0.969	0.919	0.816	0.977	0.889
Avg. score	0.668	0.701	0.673	0.847	0.799	0.813	0.853	0.815	0.818	0.854	0.821	**0.825**

Table 4.6: Recall, precision and F1-scores for OBIA and FCNs in supervised, semi-supervised, and MTL semi-supervised with the architecture shown in Figure 4.20. Legend: Mature G - Mature Grassland; Pioneering G. - Pioneering grassland; Young P. - Young pioneering; Crambe M. - Crambe Maritima

4.5.4 Discussion - species level mapping with FCNs

The results in Table 4.5 show precision, recall and F1-score for HS-UNet-Rs trained supervised, semi-supervised and MTL settings. In Table 4.5 and Figure 4.25, MTL supervised refers to models trained for supervised segmentation with supervised reflectance recovery and MTL self-supervised refers to models trained for supervised segmentation with self-supervised radiance recovery. Figures 4.27 - 4.42 show thematic maps for each method and for each cropped area shown in Figure 4.23.

HS-UNet-R convergence

The convergence of HS-UNet-Rs was analysed by testing multiple settings and hyper-parameters. The optimal set of hyper-parameters was determined by assessing computed confusion matrices and F1-scores over five sequential train/test runs with a given set of hyper-parameters. Figure 4.25 show the average pixel accuracy score over five sequential runs with the same hyper-parameters described in Section 4.5.2. Table 4.5 also shows average scores over five sequential runs. The generated thematic maps display the best performing model for a given train/test run.

HS-UNet-R - supervised

The average normalised accuracy and average F1-scores for the HS-UNet-R trained in a supervised setting were respectively, 70.23% and 0.533. The results for this setting provide a baseline for the remaining deep learning models.

Sediment pixel classifications for shingle and sand yielded the highest pixel accuracies and F1-scores, respectively scoring 93.3% and 99.7%, and 0.963 and 0.962. The thematic maps in Figures 4.27 - 4.30 show accurate predictions for sand and shingle in all input images. In particular, the separation between sand and shingle was clearly shown for input image A). However, input images C) and D) also showed failure to accurately delineate sediment channels among species belonging to SD6 NVC, e.g. Ammophila arenaria.

Classifications for Ammophila arenaria yielded the lowest pixel accuracy and F1-score, respectively 58.3% and 0.716. The top-left confusion matrix in Figure 4.25 notes 32.3% of labels for Ammophila arenaria were miss-classified as Silene uniflora. The latter error can be seen in Figures 4.28 - 4.29 for input images B) and C), where a general confusion can be noted between extents predicted as Ammophila arenaria and Silene uniflora.

The predictions for target classes belonging to SD1B NVC had a mixed performance. From this NVC, Lathryrus japonicus and Rumex crispus yielded unsatisfactory results, respectively scoring 0.7% and 7.0%, and 0.0135 and 0.122, in terms of pixel accuracy and F1-score. The top-left confusion matrix in Figure 4.25 showed that 49.4% of Lathyrus japonicus labels were miss-classified as Silene uniflora. Figures 4.5 and 4.12 showed ground and aerial captures of each species for the mapping objective. From an aerial point of view, Silene uniflora and Lathyrus japonicus can exhibit similar spectral and texture features, as shown in Figure 4.12, and cover similar areas on the ground, as shown in Figure 4.5. This could justify the over-representation of Silene uniflora and the under-representation of Lathryrus japonicus for the thematic maps displayed in Figures 4.27 - 4.30. Figures 4.5 and 4.12 should also allude to the complexity of mapping Rumex crispus from aerial imagery. In particular, Figure 4.12 shows similar texture and spectral features between Rumex crispus and shingle

sediment. Target classes such as Glaucium flavum and Silene Uniflora both yielded high pixel accuracies, respectively scoring 90.6% and 96.8%. However, the results in Table 4.5 also show poor precision scores that suggests that performance for both target classes were often false-positive. The thematic maps in Figures 4.28 and 4.29 for input images B) and C) showed this issue, where Silene Uniflora was generally over-represented. The last target class belonging to SD1B was Crambe maritima that yielded a pixel accuracy of 67.2% and a F1-score of 0.798. These results were the lowest pixel accuracy and F1-score for Crambe maritima. However, Table 4.5 showed that this class exhibited high precision that suggests a low probability of false-positive predictions.

Pixel predictions for Honckenya peploides yielded high pixel accuracy with 95.8% but also low precision. Again, this suggests a higher probability to generate false-positives. In fact, 21.0% of Lathyrus japonicus labels were miss-classified as Honckenya peploides. Eryngium maritimum paints a similar picture with high pixel accuracy but low precision score. In this case, 9.8% of Crambe maritima labels were miss-classified as Eryngium maritimum.

HS-UNet-R - semi-supervised

The average normalised accuracy and average F1-score for the HS-UNet-R trained in a semi-supervised setting were respectively, 73.22% and 0.563. HS-UNet-Rs trained in a semi-supervised setting yielded the highest average normalised accuracy but not the highest F1-score.

Sediment predictions were found to have a similar performance to models trained in a supervised setting. Respectively, shingle and sand pixel were 90.3% and 99.8% correct and the F1-score also yields high scores with 0.947 and 0.945. Again, the thematic maps in Figures 4.31 - 4.34 showed accurate delineation for sand and shingle in all input images, with input image A) clearly detailing accurate classification and delineation of shingle and sand. However, as with supervised models, input image D) showed failure to accurately delineate

sediment channels among species belonging to SD6 NVC, e.g. Ammophila arenaria, but this time input image C) showed accurate classification of the same sediment channels.

The scores in Table 4.5 showed that Ammophila arenaria yielded the highest F1-score, with 0.943 and high pixel accuracy with 95.8%. In general, for semi-supervised HS-UNet-Rs, predictions for this target class exhibit high confidence and accurate separation between vegetation and sediment among sediment channels with transitioning SD6 NVC. The latter can be seen in Figure 4.33 with input image C), where the sediment track was clearly delineated among Ammophila arenaria.

Again, the performance for Lathryrus japonicus and Rumex crispus yielded unsatisfactory results, respectively scoring 5.6% and 15.8% in terms of pixel accuracy, and 0.089 and 0.240 in F1-score. The top-right confusion matrix in Figure 4.25 noted that 90.8% of Lathyrus japonicus labels were miss-classified as Silene uniflora. Again, this could be due to both species possessing similar spectral and texture features from an aerial point of view. The poor performance of Rumex crispus also stressed the complexity of mapping this particular species from aerial imagery. Pixel classifications for target classes such as Glaucium flavum and Silene Uniflora both yielded high pixel accuracies, respectively scoring 89.1% and 95.1%. The results in Table 4.5 also showed that adding the unsupervised loss term with a teacher-student architecture described in Section 3.3 increased the precision scores of these particular target classes that in turn suggests that both target classes were less likely to be false-positive. Examining input images B) and C) in Figures 4.28 - 4.29 and Figures 4.32 and 4.33 confirms this result, where Silene Uniflora with the semi-supervised HS-UNet-R were not over-represented among sediment gaps and Ammophila arenaria. The last target class belonging to SD1B was Crambe maritima that yielded good results. The pixel accuracy for this class was 86.8% and the F1-score 0.914. The semi-supervised method has high recall with high precision that is apparent for input images C) and D) in Figures 4.33 and 4.34, where accurate predictions for pioneering Crambe maritima can be found on shingle and sand sediment.

Honckenya peploides had 70.2% pixel accuracy which is lower than supervised models. However, the latter exhibited low precision, while semi-supervised models show higher precision. Figure 4.25 show that 12.2% and 14.8% of Honckenya peploides labels were respectively miss-classified as Silene uniflora and Lathyrus japonicus. Again, suggesting that these species were complex target classes to discern from an aerial point of view. Eryngium maritimum paints a similar picture to models trained in a supervised setting, where Figure 4.25 show high pixel accuracy but Table 4.5 show low precision score. In this case, 9.3% and 6.4% of Eryngium maritimum labels were respectively miss-classified as Glaucium flavum and Crambe maritima.

HS-UNet-R - MTL: supervised

The average normalised accuracy and average F1-score for the HS-UNet-R trained in MTL with supervised segmentation and supervised reflectance recovery were respectively, 63.36% and 0.620. HS-UNet-Rs trained in this particular MTL setting yielded the highest average F1-score.

Sediment classes when using supervised reflectance recovery had a pixel accuracy of 89.1% and 100%, and F1-score of 0.935 and 0.923, respectively for shingle and sand. However, 9.0% of shingle labels were miss-classified as sand, and 43.6% of Eryngium maritimum labels were also miss-classified as sand. Suggesting that this particular sediment class was over-represented. However, Figures 4.27 - 4.30 and 4.35 - 4.38 show similar patterns with respect to shingle and sand predictions, where the separation between both classes was clearly shown for input image A) but fail to accurately delineate sediment channels among species belonging to SD6 for input images C) and D).

Ammophila arenaria yielded the highest pixel accuracy with 98.8%. However, the middle-left confusion matrix in Figure 4.25 also shows that labels for target classes belonging to SD1B NVC were often classified as Ammophila arenaria. Furthermore, Table 4.5 shows lower precision with this particular MTL setting that suggests that this particular class was

more likely to be false-positive. This error can be seen in Figures 4.36 and 4.37 for input images B) and C), where Ammophila arenaria was clearly mapped more often compared than the same images in Figures 4.28 - 4.29 and 4.32 - 4.33.

The mapping results for Lathyrus japonicus yielded unsatisfactory results, following the same trend as supervised and semi-supervised settings. The pixel accuracy and F1-score were respectively 3.6% and 0.066, with 42% of Lathyrus japonicus labels miss-classified as Silene uniflora. As mentioned, this could be due to both species possessing similar spectral and texture features from an aerial point of view. Rumex crispus pixels were 31.3% correct and the F1-score was 0.403 which was the highest pixel accuracy and F1-score. However, Figures 4.36 - 4.37 clearly shows that Rumex crispus was over-represented for input images B) and C). Therefore, while this particular training setting yields the highest accuracy and F1-score - low scores for recall and precision clearly affect the visual results shown in Figures 4.35 - 4.42. Crambe maritima was 93.3% correct with a F1-score of 0.94 which was also the highest pixel accuracy and F1-score. Figures 4.35 - 4.42 shows accurate predictions for pioneering Crambe maritima in all input images on shingle and sand sediment. Target classes such as Glaucium flavum and Silene Uniflora both yielded satisfactory pixel accuracies, respectively with 70.3% and 70.9%. However, the pixel accuracy was lower than both training settings without MTL, with 29.1% of Glaucium flavum labels miss-classified as Crambe maritima and 26.1% of Silene uniflora labels miss classified as Ammophila arenaria. However, both classes exhibited higher precision than models trained in a supervised setting using equation 3.3 that suggests that these models were less likely to generate false positives for Glaucium flavum and Silene uniflora. Both species exhibit similar spectral and texture features from an aerial point of view and adding the spectral reflectance signature of these species into the optimisation process may aid discern specific shingle plants and prevent false positives.

The performance for target classes belonging to SD2 NVCs was also mixed. Honckenya peploides had 76.4% pixel accuracy, with 13.3% of the pixel labels miss-classified as Am-

mophila arenaria. As mentioned, this may be due to the latter being over-represented, as shown in Figures 4.35 - 4.42 and Table 4.5, where the precision score was relatively lower. The F1-score for Honckenya peploides prediction was 0.752 that was also the highest score. The pixel predictions for Eryngium maritimum were unsatisfactory. For all train/test runs this particular class was not predicted a single time. Instead, labels were predicted as Ammophila arenaria and Crambe maritima.

HS-UNet-R - MTL: self-supervised

The average normalised accuracy and average F1-score for the HS-UNet-R trained in MTL with supervised segmentation and self-supervised radiance recovery were respectively, 61.2% and 0.579. HS-UNet-Rs trained in this particular MTL setting yielded the lowest average normalised accuracy but the second highest average F1-score.

Sediment classes had a pixel accuracy of 77.5% and 100%, and an F1-score of 0.863 and 0.868, respectively for shingle and sand which were the lowest scores for sediment. The pixel accuracy for shingle was also the lowest out of all training settings with 20.8% of test labels predicted as sand. The thematic maps in Figures 4.39 - 4.42 show accurate pixel classifications for shingle in all input images but fail to predict sand extents in input image A). However, for input image B), the shingle sediment was accurately separated from vegetation that was not possible for Figures 4.28 and 4.36. Furthermore, and similarly to thematic maps generated from semi-supervised models, this training setting fails to accurately delineate sediment channels among species belonging to SD6 in image D) but accurate classification for sediment channels was shown in image C).

Ammophila arenaria also yielded the highest pixel accuracy with 98.8%. However, the middle-right confusion matrix in Figure 4.25 also showed that labels for target classes, such as Honckenya peploides, Silene uniflora and Rumex crispus were often classified as Ammophila arenaria. Again, Table 4.5 shows that the precision of Ammophila arenaria was also lower for this particular MTL setting, compared to supervised and semi-supervised HS-

UNet-R without MTL. However, this error was not as prominent in thematic maps shown in Figures 4.39 - 4.42, where Ammophila arenaria was mapped similarly to thematic maps in Figures 4.31 - 4.34.

Following the trend, the mapping results for Lathryrus japonicus yielded unsatisfactory results with 0.3% pixel accuracy and 0.005 F1-score. Again, this particular class had 76.5% of test labels predicted as Silene uniflora. Rumex crispus were 21.6% correct with an F1-score of 0.338. Glaucium flavum and Silene uniflora yielded satisfactory pixel accuracies, respectively 63.1% and 72.7%. The F1-score for the same classes was 0.546 and 0.775. Silene uniflora exhibit high precision but lower recall due to 27.3% of test labels being predicted as Honckenya peploides. Comparing Figures 4.39 - 4.42 with the remaining thematic maps showed that Silene uniflora was not predicted as often in all input images. The last target class belonging to SD1B was Crambe maritima that had 88.9% test pixels classified correctly. Crambe maritima also exhibited high recall and high precision that also suggests high confidence for Crambe maritima predictions in Figures 4.39 - 4.42.

Honckenya peploides had 43.7% pixel accuracy, with 43.7% of the pixel labels miss-classified as Ammophila arenaria. Again, this suggests that Ammophila arenaria was over-represented but this was not noticed in Figures 4.39 - 4.42. The F1-score for Honckenya peploides prediction was 0.317 that was also the lowest score. The performance for Eryngium maritimum had 45.7% pixel accuracy that was lower than both training settings without MTL. However, the F1-score for Eryngium maritimum was the highest out of all training settings with 0.334. This was mainly due to low precision scores for supervised and semi-supervised HS-UNet-Rs without MTL.

4.5.5 Discussion - merged classes and comparison with OBIA

The results in Table 4.5 show precision, recall and F1-score for HS-UNet-Rs trained supervised, semi-supervised and MTL with a self-supervised auxiliary image task. Table 4.6 and Figure 4.26 show that the average normalised accuracy and average F1-score for the OBIA

were respectively, 66.8% and 0.633 with a reduced class domain, where Rumex crispus and Glaucium flavum were merged to form a new class called pioneering grassland and Silene uniflora, Lathyrus japonicus and Honckenya peploides were merged to form the young pioneering species class. The predictions with FCNs were merged in the same way in order to provide a one-to-one comparison with OBIA. Figures 4.43 - 4.45 show thematic maps for each method and for each cropped area shown in Figure 4.24.

Comparison with OBIA

The results in the Section 4.5.3 show two key findings.

The first was that the confusion matrices in Figure 4.26 and the scores in Table 4.6 showed that FCNs, and in particular the U-Net architecture in Figure 4.9, performed better than the standard OBIA methodology with eCognition [Nussbaum and Menz, 2008] used for coastal habitat mapping. The F1-scores in Table 4.6 also reflected the same performance with each target-class scoring lower F1 with the OBIA method than FCNs with any optimisation strategy.

This trend of results has become standard with other studies also showing that CNN-based applications for semantic segmentation and object detection can produce comparable or better performance than OBIA mapping applications Guirado et al. [2017]; Huang et al. [2020]; Hobley et al. [2021a]; Zheng et al. [2022]. Further research has attempted to combine OBIA with FCNs by allowing the multiresolution segmentation to generate candidate image-objects for FCN training. These studies have also shown to outperform standard OBIA with shallow machine learning models, such as random forests and support vector machines Zaabar et al. [2022]; Detka et al. [2023]. However, deep learning methods excel at learning internal representations of input signals in end-to-end fashion, and therefore constraining the input of deep learning networks to the outputs of multiresolution segmentation may limit performance LeCun et al. [2015a].

The second key finding was the increase in performance using semi-supervised optimisation strategies instead of standard supervised training. The confusion matrices in Figure 4.26 and the scores in Table 4.6 reflected this finding for every target-class, except Crambe Maritima and young pioneering species from SD1B and SD2.

The semi-supervised method using consistency regularisation described in Section 3.3.1 showed the same pattern of results for the dataset described in Section 4.2.1. As mentioned, the dataset pre-process strategy resulted in a number of pixels on the edge of each training sample to remain unlabelled. The loss function in Equation 3.3 discards these pixels for supervised optimisation by use of a binary mask but the method in Section 3.3.1 leverages the unlabelled pixels to gain a small performance boost without the addition of more labelled training samples (rasterised polygons). This also confirms the validity of the methods described in Tarvainen and Valpola [2017]; French et al. [2020a] for practical applications of coastal remote sensing.

The method described in Section 4.5 yielded the best average recall (or pixel accuracy), precision and F1 which also shows the practicality of semi-supervised MTL optimisation for datasets where the quantity and distribution of labelled data within a coastal environment may be limited due to associated costs.

Other applications of MTL in remote sensing tend to follow two trends. One trend was to leverage two, or more, supervised image tasks to improve segmentation accuracy or inter-class boundary separation [Lu et al., 2019; Jing et al., 2021; Ruiwen et al., 2022; Lu et al., 2022]. However, the proposed method deviates from these studies by only leveraging one labelled dataset whilst improving segmentation accuracy. The other trend in MTL applications in remote sensing was to incorporate temporal information by allowing a change detection loss from multiple surveys to be added into the optimisation strategy [Hong et al., 2023; Cui and Jiang, 2023]. This trend has shown to produce fine-grained accurate segmentation for the datasets in Hong et al. [2023]; Cui and Jiang [2023] and could be incorporated into future work for study site revisits.

The results in Table 4.6 also follow a similar trend reported for other applications of semi-supervised MTL. In particular, a study for lung cancer diagnosis found the use of semi-supervised multi-task learning to produce better results than dual supervised image tasks with a single shared model [Khosravan and Bagci, 2018].

The research in Wang et al. [2022] uses two partially labelled image tasks: semantic segmentation and depth estimation. However, instead of allowing a single epoch to be the forward propagation of both image datasets as shown in Figure 4.20, their method continuously switches labelled samples for each image task, e.g., one epoch has supervised segmentation and unsupervised depth estimation and the next epoch had unsupervised segmentation with supervised depth estimation.

The study in Castillo-Navarro et al. [2021] also used semi-supervised MTL for remotely sensed imagery. This particular study also showcased the use of reconstruction image loss functions, such as: mean relative absolute difference, mean squared difference, and other unsupervised loss functions, such as: relaxed k-means and Mumford-Shah [Kim and Ye, 2019]. For shared network architectures, the unsupervised reconstruction loss was found to produce the best objective results which was also achieved with our proposed method. Incorporating the self-supervised reconstruction loss encodes the network with internal representations that approximate the semantic segmentation performance to be as photo realistic and fine-grained as the input image [Xia and Kulis, 2017]. However, the main difference between our proposed method and the method described in Castillo-Navarro et al. [2021] was the intermediate spectral upsample step to the original multispectral colour space with known sensor sensitivities before reconstructing the images. As mentioned, this unsupervised intermediate step does not achieve high-fidelity hyperspectral reconstruction but many hyperspectral metamers can integrate to the same multispectral image capture. Furthermore, the use of hyperspectral reconstruction can bridge the domain gap when different sensors are used for consecutive site surveys. On the whole, the proposed MTL optimisation strategy should

allow shared networks to achieve similar photo realism and improved segmentation as shown in Castillo-Navarro et al. [2021].

The latter can be seen with Figures 4.43 to 4.45 where the separation between background sediment and foreground vegetation was more accurate with MTL semi-supervised than OBIA and semi-supervision with consistency regularisation but was similar to supervised FCNs. The main difference between supervised and semi-supervised MTL networks was that networks trained with the former optimisation strategy tend to over-represent mature grassland, as shown in Figure 4.44 and the precision scores in Table 4.6.

Visual analysis

This section provides a qualitative visual analysis of the segmented maps shown in Figures 4.43 to 4.45 using information and knowledge gained from the in-situ survey.

The thematic maps in Figures 4.43 to 4.45 produced fine-grained and similar results for FCNs with any optimisation strategy that provides additional evidence regarding the accuracy of the results. However, the visual results for OBIA were found to be coarse in relation to FCNs that can be due to the partitioning of high-resolution orthomosaics into homogeneous regions using multiresolution segmentation. Furthermore, OBIA was found to produce large areas of erroneous pioneering grassland predictions that can be reflected in Figure 4.44.

The qualitative analysis for FCNs in Figure 4.43 shows that the method in Section 4.5.2 produced the most accurate maps for both shingle vegetation and sediment classes. Supervised FCNs optimised with Equation 3.3 yielded similar results but was found to have erroneous classifications of pioneering grassland instead of young pioneering. Lastly, the semi-supervised method with consistency regularisation had the same issue with erroneous predictions of pioneering grassland and sand.

Figure 4.44 showed that the proposed method yielded the most accurate separation between foreground vegetation and background sediment that shows that incorporating the self-

supervised reconstruction loss encodes the network approximate segmentation performance to be as photo realistic and fine-grained as the input image. However, the same method was found to over-represent vegetation as young pioneering, whereas the area shown in Figure 4.44 was most found to have a mixture of young pioneering species and pioneering grassland. Supervised FCNs optimised with Equation 3.3 reflected this mixture of vegetation classes but indeed struggled to separate vegetation from sediment. The semi-supervised method with consistency regularisation was found to be the worst among the FCNs due to erroneous over-representation of mature grassland.

Figure 4.45 paints a very similar picture to the results in Figure 4.44. The separation of sediment channels among mature grassland was found to be the best with the proposed semi-supervised method. The supervised FCNs performed similarly to the proposed method but the delineation of sediment channels was not as refined and the semi-supervised method with consistency regularisation was found to over-represent mature grassland and failed to find the same sediment channels.

4.5.6 Summary

Section 4.5 showed the utility of HS-UNet-Rs trained with a small set of polygons to segment shingle, strandline and sand-dune communities belonging to SD1, SD2 and SD6 NVCs. Each HS-UNet-R was evaluated in four training modes: supervised and semi-supervised without MTL, and supervised with MTL, where the auxiliary task was either supervised reflectance recovery or self-supervised radiance recovery. The results show that models trained without MTL yield high accuracies but also lower precision which indicates a higher likelihood to produce false-positives. The semi-supervised setting confirms the findings in Section 3.3, where the unsupervised loss term with a teacher-student architecture helps with segmentation performance. Models trained with MTL using self-supervised radiance recovery also show an alternative method for segmentation that improves objective scores without adding labels to the training dataset. Supervised models with MTL using supervised reflectance recovery as an auxiliary task had the highest average F1-score which suggests that hyper-

spectral signatures recorded in-situ can help with discerning vegetation species. However, the latter results required supplementing the dataset with spectroradiometer measurements which is not often available in coastal monitoring.

This section also showed a comparison with OBIA that yielded two key findings. The first was that FCNs were a good alternative and the proposed architecture in Figure 4.9 was shown to produce better objective scores and visually pleasing segmented maps. The second key finding was the use of semi-supervision in two different strategies was found to perform better than networks trained in standard supervised methodology.

The proposed network architecture and optimisation strategy in Figures 3.8 and 4.20 was a simple extension to classic encoder–decoder U-Net architectures and was shown to be effective in a semi-supervised scenario. With this architecture, an unsupervised auxiliary loss based on hyperspectral reconstruction was used alongside with semantic segmentation. The unsupervised hyperspectral reconstruction was based on the assumption that many hyperspectral metamers correspond to the same multispectral capture Cohen and Kappauf [1982]; Morovic and Finlayson [2006], and therefore while the method does not achieve high-fidelity hyperspectral reconstruction, it does indeed produce accurate image reconstruction. The experiments in Section 4.5.3 have shown that adding an unsupervised auxiliary image task for the purpose of image reconstruction from higher-dimensional spectral cubes improved semantic segmentation maps and allowed to generate finer and more homogeneous thematic maps. Furthermore, the use of hyperspectral reconstruction can bridge the domain gap when different sensors are used in consecutive surveys of the same study site.

However, many other semi-supervised strategies exist that could be developed in future work. Further work could explore the use of change detection to add a temporal aspect to the mapping process and unsupervised domain adaption could provide an alternative route to generating labels from synthesised data Li et al. [2022]; Gao et al. [2023].

4.6 Conclusions

Chapter 4 continues to show the utility of fully convolutional neural networks for coastal remote sensing and explores alternative semi-supervised optimisation strategies for partially labelled datasets. In Chapter 3 FCNs were trained in two modes: supervised and semi-supervised. The results in Section 3.3.3 show that consistency-based semi-supervised methods improve the average pixel accuracy for intertidal seagrass mapping.

In Chapter 4, FCNs were trained using multi-task learning in order to map the pioneering shingle vegetation species of SD1, SD2 and SD6 NVCs that can be found in the open-shore beach of Sizewell in Suffolk, England. The in-situ survey described in Section 4.2 shows collected data from the study site and provided an ecological context for the mapping objectives detailed in Section 4.5. The in-situ survey also collected hyperspectral reflectance samples of several species from the stated NVCs that provided the basis of the multi-task learning framework. In this framework, FCNs were trained for semantic segmentation, given this provides an equivalent output with OBIA, and hyperspectral reconstruction as an auxiliary image task.

Section 4.3 showed the use of HSCNN-R [Xiong et al., 2017] and the HS-UNet-R for HSI reconstruction on the ICVL dataset. The utility of the U-Net was investigated in order to check the feasibility of employing this architecture with multi-task learning. The results show a small degradation in objective score that confirms the findings in [Zhang et al., 2022].

Section 4.4 uses the methods derived in Section 4.3 for supervised hyperspectral reflectance reconstruction based on the measurements described in Section 4.2.1. Section 4.4.3 confirms the results in Section 4.3.3, where the HSCNN-R outperforms the HS-UNet-R in a supervised setting. However, the dataset and pre-processing of hyperspectral measurements described in Section 4.4.2 also results in visual artifacts for reconstructed hyperspectral channels, as shown in Figures 4.15 and 4.16. Furthermore, the test set was limited to

10 images that in turn does not provide the basis to compare with methods presented in [Deeb et al., 2019; Zhang et al., 2020; Gong et al., 2022]. The second method described a self-supervised hyperspectral radiance reconstruction method. The results in Section 4.4.3 confirm the findings described in Section 4.3.4, where the HSI reconstruction fidelity was not precise due to the ill-posed and unconstrained nature of reconstructing a higher-dimensional space from a lower-dimensional manifold. However, the projection to multispectral colour space from HSI was precise.

The results in Section 4.4.3 show that the HS-UNet-R was also suitable for hyperspectral reconstruction which provides the opportunity to leverage MTL.

Section 4.5 showed the use of MTL to jointly learn self-supervised hyperspectral reconstruction or supervised hyperspectral reflectance recovery with semantic segmentation. The main objectives were to incorporate hyperspectral measurements in the optimisation process, and note whether the auxiliary image task could improve the boundary delineation of per pixel predictions from the HS-UNet-R.

Section 4.5 also showed the utility of HS-UNet-Rs trained with a small set of polygons to segment shingle vegetation communities belonging to SD1, SD2 and SD6 NVCs. Each HS-UNet-R was evaluated in four training modes: supervised and semi-supervised without MTL, and supervised with MTL, where the auxiliary task was either supervised reflectance recovery or self-supervised radiance recovery. The results show that models supervised models with MTL using supervised reflectance recovery as an auxiliary task had the highest average F1-score which suggests that hyperspectral signatures recorded in-situ can help with discerning vegetation species. However, the latter results required supplementing the dataset with spectroradiometer measurements that is not often available in coastal monitoring.

A comparison with OBIA also yielded two key findings. The first was that FCNs were a good alternative and the proposed architecture in Figure 4.9 was shown to produce better objective scores and visually pleasing segmented maps. The second key finding was the use

of semi-supervision in two different strategies was found to perform better than networks trained in standard supervised methodology.

The proposed network architecture and optimisation strategy was shown to be effective in a semi-supervised scenario. The experiments in Section 4.5.3 have shown that adding an unsupervised auxiliary image task for the purpose of image reconstruction from higher-dimensional spectral cubes improved semantic segmentation maps and allowed to generate finer and more homogeneous predictions. Furthermore, the use of hyperspectral reconstruction can bridge the domain gap when different sensors are used in consecutive surveys of the same study site.

To sum up, this Chapter continues to show the use of FCNs as an alternative tool for coastal remote sensing applications given the requirements for optimisation are the same as object-based methods. Furthermore, an alternative method for semi-supervised semantic segmentation using multi-task learning provided better objective scores than models trained with standard supervised techniques. This builds on the conclusions set in Chapter 3 where semi-supervised optimisation strategies can help bridge the gap between laborious in-situ labelling efforts and accurate, yet effecient mapping methodologies with FCNs.

5 Conclusions and future work

In this chapter, the findings and contributions are summarised along with a discussion of potential future work.

During the research period, Cefas provided two datasets with VHR imagery captured with UAS instruments and miniaturised multispectral sensors that provided the basis to extract fine scale orthomosaics with multispectral resolution using SfM [Turner et al., 2012]:

1. Budle Bay, Northumberland, England (55.625°N, 1.745°W) with in-situ data provided by the Environmental Agency.
2. Sizewell, Suffolk, England (55.207°N, 1.602°W) with in-situ data provided by Cefas and the UEA.

Given these datasets, each chapter concerns different mapping objectives, and also considers different methods to map either problem domain that aim to reduce the quantity of ground-truth labels required.

5.1 Semi-supervised intertidal seagrass mapping

Chapter 3 shows the work conducted using the imagery and annotated samples from the Budle Bay study site. The main goal for this particular site was to map intertidal seagrass extents due to its contribution to intertidal coastal ecosystem health.

The literature reviewed in Section 2.3.2 identified fully convolutional neural networks for semantic segmentation that provides an equivalent output to object-based methods. Given both methods produce equivalent outputs and require the same labels, Section 3.3 showed a successful application of FCNs for intertidal seagrass mapping. Section 3.3 also showed

a semi-supervised approach for semantic segmentation and the discussion addresses the challenges and problems associated with mapping intertidal seagrass among species of algae and compares FCNs in two training modes: supervised and semi-supervised. The results indicate that semi-supervision helps with segmentation of target classes that have a small number of labelled pixels. Furthermore, Section 3.3 also showed that OBIA continues to be a robust approach for monitoring multiple coastal features in high resolution imagery. In particular, OBIA was found to be more accurate than FCNs in predicting seagrass for both cameras. However, these results were highly dependant of the initial parameters used for MRS, with the scale parameter being critical for image-object creation. However, OBIA requires the user to understand the target class domain for a particular mapping objective in order to correlate segmented image-objects with known aerial extents of the class domain.

In essence, fully convolutional neural networks were identified and used efficiently with semi-supervised optimisation strategies and the results show that FCNs can be used to create effective tools or methodologies for robust analysis across different sites, given the common use of spatially explicit labels, or polygons, between FCNs and OBIA.

5.2 Crowdsourcing experiment for intertidal seagrass mapping

Given Section 3.3 identified FCNs as an alternative mapping tool to OBIA. Section 3.4 continues to explore the use of FCNs for the same and also investigates alternative methodologies for efficient intertidal seagrass mapping.

An alternative for in-situ data collection could be visual identification and delineation of training data directly from orthomosaics [Kattenborn et al., 2019b; Wagner et al., 2019; Lopatin et al., 2019]. Section 3.4 showed the feasibility of crowdsourcing labels from aerial imagery in order to provide a cost-effective alternative to laborious labelling procedures from single domain specific experts. The second contribution explores the problems associ-

ated with crowdsourced labels by conducting an inter-observer variability experiment and training FCNs with crowdsourced labels.

The results in Section 3.4.2 confirmed that discipline expertise, prior knowledge of the site and/or previous experience annotating marine biology play an important role in minimising inter-observer variability and ensuring accurate annotation, and that lack of exposure to the above leads to high variability and low confidence. Furthermore, the results also point to a small performance gain between annotators with expert discipline knowledge versus annotators with no previous experience in marine biology annotation or domain expertise. However, this may be skewed due to the annotations from one of the participants, who is an expert geomorphologist with no prior knowledge of the study site. Participant 6 can be viewed as an outlier to the experiment given the poor annotation accuracy. However, erroneous annotations from participant 6 should not influence the confusion matrices shown in Figure 3.17 given the annotations were merged to form majority-vote annotations. Therefore, by using majority-vote annotations, individual miss annotations were suppressed and the general trends shown in the confusion matrix paint general miss classifications between target classes that exhibit similar colour and texture from an aerial point of view, i.e., separating species of algae and even separating algae from seagrass.

This Section stressed the difficulty of labelling a complex multi-class marine biology problem and confirms that pre-exposure to the study site was important for intertidal classification, if good quality labels were to be guaranteed, and that in-situ ground-truthing may be unavoidable to prevent confusion by site experts. Therefore, site surveying was necessary but may result in sparse data points with respect to the size of the coastal site. Domain experts can enhance training datasets in coastal remote sensing but domain experts present during the site survey yielded the best quality labels.

The results also showed that FCNs trained with low inter-observer variability and high confidence annotations demonstrate comparable performance to the FCNs trained with in-situ labels. The crowdsourcing scenario also showed that in-situ efforts can be combined

successfully with crowdsourced aerial imagery annotation. Having said that, this work does not fully exclude in-situ surveying but merely affirms that a good quality labels can be found in-situ and a healthy quantity of labels can also be supplemented from aerial imagery which would reduce in-situ efforts and costs.

5.3 Hyperspectral reconstruction on multispectral Sizewell imagery

Chapter 4 shows the work conducted using the imagery and annotated samples from the Sizewell study site. The main goal for this particular site was to map shoreline species that belong to strandline, sand and shingle communities. This Chapter also continues to use fully convolutional neural networks and investigates alternative semi-supervised optimisation strategies using multi-task learning with a shared model. Given the data collected, as shown in Section 4.2.1, the MTL framework attempts to jointly learn semantic segmentation and hyperspectral reconstruction.

Section 4.3 showed two methods for HSI reconstruction: a shallow method described in Arad and Ben-Shahar [2016], and two deep implementations, the HSCNN-R [Shi et al., 2018] and an extension U-Net architecture, known as HS-UNet-R. The results confirm the findings in [Xiong et al., 2017] - where the shallow method in [Arad and Ben-Shahar, 2016] was outperformed by deep implementations. Furthermore, the addition of pooling operations and a subsequent upsample track in the network topology resulted in objective score degradation that confirms the findings in [Zhang et al., 2022]. However, while the pooling operation was not ideal for network design, the HS-UNet-R still outperformed the shallow method in [Arad and Ben-Shahar, 2016]. These network designs were optimised in three settings: supervised, semi-supervised and self-supervised. The unsupervised loss used the spectral response functions of the miniaturized Sentera multispectral camera to project higher-dimensional hyperspectral data cubes to a lower-dimensional manifold in RGB colour space. The predictions for self-supervised networks were optimised on the basis that many

physically plausible hyperspectral metamers correspond to the same RGB capture [Cohen and Kappauf, 1982; Morovic and Finlayson, 2006].

Given the deep learning models described in Section 4.3; the following methods showed a supervised method to reconstruct hyperspectral reflectance from a multispectral image using the hyperspectral measurements described in Section 4.2.1. And, the self-supervised method was used to reconstruct hyperspectral radiance. The results confirm the findings shown with the ICVL dataset. But, given the dataset and pre-processing stages of hyperspectral measurements described in Section 4.4.2, the results showed visual artifacts for reconstructed hyperspectral channels. Furthermore, the analysis was constrained to ten test images that in turn does not provide the basis to compare with methods presented in [Deeb et al., 2019; Zhang et al., 2020; Gong et al., 2022]. The self-supervised hyperspectral radiance reconstruction described in Section 4.3 was also applied to multispectral imagery of the study site.

5.4 Multi-task learning for species at Sizewell study site

As discussed in Section 5.3 the HS-UNet-R showed robust results which provided the opportunity to leverage a multi-task learning.

Section 4.5 showed the utility of HS-UNet-Rs trained with a small set of polygons to segment shingle vegetation belonging to SD1, SD2 and SD6 NVCs. Each HS-UNet-R was evaluated in four training modes: supervised and semi-supervised without MTL, and supervised with MTL where the auxiliary task was either supervised reflectance recovery or self-supervised radiance recovery. The semi-supervised setting confirmed our earlier findings from Section 3.3, where the unsupervised loss term with a teacher-student architecture helps with segmentation performance. Moreover, models trained with MTL using self-supervised radiance recovery also showed an alternative method for segmentation that improves objective scores without adding labels to the training dataset.

Most of MTL frameworks aimed at semantic segmentation attempt to improve inter-class boundaries and delineation but also leverage auxiliary image tasks closely related to segmentation task [Volpi and Tuia, 2018; Bischke et al., 2019; Li et al., 2021a; Jing et al., 2021; Liu et al., 2020b], or depth estimation [Lu et al., 2022; Wang et al., 2020a; Carvalho et al., 2019]. In this case, the auxiliary task was hyperspectral reconstruction that also achieved the objective set out in prior works - predictions of our method were sharper between inter-class boundaries, and that leveraging an auxiliary image task that optimises image reconstruction forces the MTL network to convey realistic and sharp features in imagery to the later network stages.

Another comparison with OBIA that yielded two key findings. The first was that the HS-UNet-R shown in Figure 4.9 was shown to produce better objective scores and visually pleasing segmented maps. The second key finding was the use of semi-supervision in two different strategies was found to perform better than networks trained in standard supervised methodology.

Section 4.5 continues to show the use of FCNs as an alternative tool for coastal remote sensing applications given the requirements for optimisation are the same as object-based methods. Furthermore, an alternative method for semi-supervised semantic segmentation using multi-task learning provided better objective scores than models trained with standard supervised techniques. This builds on the conclusions set in Chapter 3 where semi-supervised optimisation strategies can help bridge the gap between laborious in-situ labelling efforts and accurate, yet efficient mapping methodologies with FCNs.

5.5 Thoughts on coastal remote sensing and Future work

Coastal remote sensing is an avenue of acquiring data that has seen improvements to sensor platforms [Anderson and Gaston, 2013] that in turn have increased spatial resolution of captured imagery. This provides an opportunity for accurate and fine scale mapping of coastal features but also presents a gap in labelled data. The methods described in Chapters

3 and 4 attempt to bridge the gap between large and fine scale orthomosaics and sampled in-situ records.

The thesis reviewed deep learning literature and identified fully convolutional neural networks as an alternative tool for mapping remotely sensed imagery with Chapters 3 and 4 using the U-Net architecture for mapping multiple coastal features in order to achieve comparable, or better, performance to OBIA. Both methods produce equivalent outputs but also require the same labels, or polygons, in order to drive the optimisation of machine learning models. The use of spatially explicit labels is the key factor for optimisation, and both methods require this in order to effectively learn complex relationships from features derived using orthomosaics to target classes such as intertidal seagrass and algae. Therefore, the use of FCNs can be adapted for other applications of coastal remote sensing given the requirements to drive the optimisation of FCNs is the same as object-based methods in a supervised setting, and ecologists may consider the use of deep learning models as an alternative tool for robust analysis across different study sites.

However, as mentioned, deep learning models perform the best with large datasets with high-quality labels [Everingham and Winn, 2012]. Practical applications to coastal remote sensing where the amount of labels are limited may prefer simpler architectures in order to achieve better model generalisation. This thesis also focused on deriving efficient optimisation strategies with semi-supervised semantic segmentation in two scenarios:

- consistency-based regularisation
- multi-task learning with self-supervised auxiliary image tasks.

And also focused on alternative label procurement through crowdsourcing. However, one avenue that was not explored during the research project was domain adaption and adversarial training for semantic segmentation purposes [Souly et al., 2017; Luc et al., 2016].

Domain adaptation aims to transfer knowledge in the presence of the domain gap [You et al., 2019]. More formally, real-world applications, including coastal remote sensing, can

capture datasets that result in different distributions due to many factors, such as collection of data from different sources or time [Farahani et al., 2021]. Domain adaptation is a sub-field within machine learning that aims to align the disparity between domains. Coastal remote sensing can experience domain shifts, where the same study site is analysed but in a different point in time and imaging sensor. These methods have found use in remote sensing and in particular, Islam et al. [2020] applies adversarial training for seagrass mapping to overcome the domain shift from mapping in different coastal environments.

Also, Section 4.5 showed the utility of using hyperspectral reconstruction as an auxiliary image task to improve inter-class predictions. While the results indicate that hyperspectral reconstruction indeed improves inter-class delineation, future work could leverage an even simpler auxiliary task that optimises image reconstruction in order to convey realistic and sharp features in imagery to the later stages of the MTL network.

This work also stressed the importance of site surveys for mapping objectives that pertain to local study sites. A range of studies in coastal remote sensing derive training labels directly from VHR orthomosaics. However, while this may be cost effective, pre-exposure to the study site can provide important biases that yield good quality labels, and that in-situ survey may be unavoidable to prevent confusion by annotators from an aerial point of view. Furthermore, coastal monitoring through remote sensing is an inter-disciplinary problem that emphasises the need for ecologists to work in tandem with computer scientists. Therefore, while site surveys concern ecologists due to their site knowledge of species that contribute to coastal ecosystem health, the presence of computer scientists can also be beneficial which would allow them to better understand the mapping objectives.

In 2021, the Sizewell study site was revisited with intent to identify the same species listed in Section 4.2.1. This time, the aerial survey was performed with a ten band MicaSense Blue/RedEdge-MX dual camera system that captured the study site with very fine multispectral resolution between 400-900nms. The second dataset to the Sizewell study site has two applications that could be investigated. The first is efficient and accurate image

registration between generated orthomosaics of both aerial surveys. Therefore, the goal would be to register a five band multispectral orthomosaic from the Sentera multispectral camera with the ten band multispectral orthomosaic from the MicaSense Blue/RedEdge-MX dual camera system. This work can leverage the method developed in the Appendix A, and combine it with spatial transformer networks [Jaderberg et al., 2015]. The latter networks have shown state-of-the-art performance on image registration [Hernandez-Matas et al., 2017; Hering et al., 2022] but the author's hypothesis is that the linear MK-Transform prior to the network forward pass can improve the subsequent registration process.

The second application would be to conduct a change detection study for the species and NVCs stated in Section 4.5. Efficient thematic mapping is an important task in coastal monitoring but accurate cataloging on a temporal scale in tandem with change detection can provide ecologists with different perspectives of contributing factors to coastal ecosystem health [Cook, 2017; Morgan and Hodgson, 2021]. The ten band MicaSense Blue/RedEdge-MX dual camera system which as mentioned results in a domain shift of the underlying image data distributions. This fits the key requirement for applying domain adaption [You et al., 2019; Farahani et al., 2021]. The use of finer spectral resolution could allow for improved hyperspectral radiance reconstruction which in tandem with MTL could improve the mapping of plant species and communities present at the study site. Therefore, domain adaption methods can be applied to the hyperspectral reconstruction task that in turn would connect both surveys to the Sizewell study site over different periods of time.

Abbreviations

ASPP Atrous Spatial Pyramid Pooling.

BN Batch Normalisation.

Cefas Centre for Environmental Fisheries and Aquaculture Sciences.

CNNs Convolutional Neural Networks.

CRFs Conditional Randon Fields.

CV Computer Vision.

DL Deep Learning.

DSM Digital Surface Model.

EA Environmental Agency.

EMA Exponential Moving Average.

FCNs Fully Convolutional Neural Networks.

FoV Field of View.

GANs Generative Adversarial Networks.

GIS Geographic Information System.

GLCM Grey Level Co-occurance Matrix.

GNDVI Green Normalised Difference Vegetation Index.

GPS Global Positioning System.

HS Hyperspectral.

HSI Hyperspectral Imagery.

IAVI Atmospheric Resistant Vegetation Index.

LMM Linear Mixture Models.

LSU Linear Spectral Unmixing.

MCARI Modified Chlorophyll Absorption Ratio Index.

MESMA Multiple Endmember Spectral Mixture Analysis.

MK-T Monge-Kantorovich Transform.

MLC Maximum Likelihood Classification.

MRS Multi-resolution segmentation.

MSAVI Modified Soil Adjusted Vegetation Index.

MTL Multi-task learning.

NDVI Normalised Difference Vegetation Index.

NGBDI Normalised Green-Blue Difference Vegetation Index.

NGRDI Normalised Green-Red Difference Vegetation Index.

NVCs National Vegetation Classes.

OBIA Object Based Image Analysis.

OMP Orthogonal Match Pursuit.

PCA Principal Component Analysis.

PPI Pixel Purity Index.

RANSAC Random Sample Consensus.

ReLU Rectified Linear Unit.

RGB Red Green and Blue.

RMSE Root Mean Square Error.

RPA Remotely Piloted Aircraft.

RRMSE Relative Root Mean Square Error.

RTK Real-Time Kinetics.

SAM Spectral Angle Mapper.

SfM Structure from Motion.

SGD Stochastic Gradient Descent.

SIFT Scale Invariant Feature Transform.

SVMs Support Vector Machines.

UAS Uncrewed Aircraft System.

VARI Visible Atmospherically Resistant Index.

VDVI Visible-band Difference Vegetation Index.

VHR Very-high resolution.

Glossary

crowdsourced The process of acquiring labels for Machine Learning algorithm training using members of the general public or non-experts.

hyperspectral sensor A type of camera that uses many narrow-band filters (typically in the hundreds) to capture wavelengths between 400-2500nm. The range covers a wide range of wavelengths in order to capture the spectral signature of an object in a pixel. These sensors have a variety of scanning mechanisms dedicated for producing a hyperspectral data cube at a comprise of reduced spatial resolution.

image segmentation The process of clustering image pixels, without supervision, in order to delineate objects of interest.

narrow-band multispectral A type of camera that uses narrow-band filters (typically ranging from five to ten) to capture wavelengths between 400-1200nm. The range covers the visible spectrum (red, green and blue channels) as well near infra-red. In comparison a commercial camera uses wide-band filters to capture wavelengths between 400-700nm.

point A point is an in-situ RTK GPS measurement from the study site that can be used to locate features in the orthomosaic using a GIS. For Budle Bay, each in-situ RTK GPS had the percentage cover of features of interest estimates using quadrat sampling. For Sizewell, each RTK GPS had an identified specie classified using an expert taxonomist.

polygon A polygon is a rasterised shape file from a Geographic Information System (GIS). Each polygon was drawn using the semantic information from an in-situ point and

leveraging the aerial point-of-view from the orthomosaic. The particular shape of each polygon uses photo-interpretation of colour and texture and depends on the type ecological feature to be mapped. Rasterised polygons were used to training image samples for Fully Convolutional Neural Networks (FCNs).

quadrat sampling The process of used by ecologists to sample ecological features in a small area (typically 50×50cm) in order to accurately calculate the percentage cover of each ecological feature of interest..

semantic segmentation The process of assigning each pixel in an image to a semantic value in order to delineate objects of interest.

semi-supervised The process of training Machine Learning algorithms with known outcomes and without known outcomes in order to learn the relationship between said outcomes and input features whilst also leveraging non-labelled samples in the dataset.

spatial resolution The physical distance measured in a single pixel.

supervised The process of training Machine Learning algorithms with known outcomes in order to learn the relationship between said outcomes and input features.

tiles A tile is a rectangular orthomosaic crop that corresponds to a portion of the study site. The image size for tiles were either 3000×3000 or 6000×6000.

unsupervised The process of training Machine Learning algorithms without known outcomes in order to find clusters of data.

6 Bibliography

Acharya, T. D. and Yang, I. (2015). Exploring lLandsat 8. *International Journal of IT, Engineering and Applied Sciences Research (IJIEASR)*, 4(4):4–10.

Adam, P. (1993). *Saltmarsh Ecology*. Cambridge University Press.

Adam, P. (2002). Saltmarshes in a time of change. *Environmental conservation*, 29(1):39–61.

Adams, J. B. (1993). Imaging spectroscopy: Interpretation based on spectral mixture analysis. *Remote geochemical analysis: Elemental and mineralogical composition*, pages 145–166.

Adams, J. B., Sabol, D. E., Kapos, V., Almeida Filho, R., Roberts, D. A., Smith, M. O., and Gillespie, A. R. (1995). Classification of multispectral images based on fractions of endmembers: Application to land-cover change in the brazilian amazon. *Remote sensing of Environment*, 52(2):137–154.

Adriano, S., Chiara, F., and Antonio, M. (2005). Sedimentation rates and erosion processes in the lagoon of venice. *Environment International*, 31(7):983–992.

Aeschbacher, J., Wu, J., and Timofte, R. (2017). In defense of shallow learned spectral reconstruction from RGB images. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 471–479.

Agisoft, L. (2018). Agisoft metashape user manual, Professional edition, Version 1.5. *Agisoft LLC, St. Petersburg, Russia, from https://www. agisoft. com/pdf/metashape-pro_1_5_en. pdf, accessed June*, 2:2019.

Aharon, M., Elad, M., and Bruckstein, A. (2006). K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11):4311–4322.

Ahn, Y.-H. and Shanmugam, P. (2006). Detecting the red tide algal blooms from satellite ocean color observations in optically complex Northeast-Asia Coastal waters. *Remote Sensing of Environment*, 103(4):419–437.

Albuquerque, J. P. d., Herfort, B., and Eckle, M. (2016). The tasks of the crowd: A typology of tasks in geographic information crowdsourcing and a case study in humanitarian mapping. *Remote Sensing*, 8(10):859.

Alongi, D. M. (2020). *Coastal ecosystem processes*. CRC press.

Alonso, O. (2015). Challenges with label quality for supervised learning. *Journal of Data and Information Quality (JDIQ)*, 6(1):1–3.

Alonso, O., Rose, D. E., and Stewart, B. (2008). Crowdsourcing for relevance evaluation. In *ACM SigIR Forum*, volume 42, pages 9–15. ACM New York, NY, USA.

Amos, C., Bergamasco, A., Umgiesser, G., Cappucci, S., Cloutier, D., DeNat, L., Flindt, M., Bonardi, M., and Cristante, S. (2004). The stability of tidal flats in Venice Lagoon—the results of in-situ measurements using two benthic, annular flumes. *Journal of Marine Systems*, 51(1-4):211–241.

Anderson, K. and Gaston, K. J. (2013). Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Frontiers in Ecology and the Environment*, 11(3):138–146.

Andrés, S., Arvor, D., Mougenot, I., Libourel, T., and Durieux, L. (2017). Ontology-based classification of remote sensing images using spectral rules. *Computers & Geosciences*, 102:158–166.

Aplin, P. (2005). Remote sensing: ecology. *Progress in Physical Geography*, 29(1):104–113.

Arad, B. and Ben-Shahar, O. (2016). Sparse recovery of hyperspectral signal from natural RGB images. In *European Conference on Computer Vision*, pages 19–34. Springer.

Arad, B., Ben-Shahar, O., Timofte, R. N., Van Gool, L., Zhang, L., and Yang, M. N. (2018). challenge on spectral reconstruction from RGB images. In *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA*, pages 18–22.

Arad, B., Timofte, R., Ben-Shahar, O., Lin, Y.-T., and Finlayson, G. D. (2020). Ntire 2020 challenge on spectral reconstruction from an rgb image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 446–447.

Arazo, E., Ortego, D., Albert, P., O'Connor, N. E., and McGuinness, K. (2020). Pseudo-labeling and confirmation bias in deep semi-supervised learning. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Asokan, A. and Anitha, J. (2019). Change detection techniques for remote sensing applications: a survey. *Earth Science Informatics*, 12(2):143–160.

Astsatryan, H., Grigoryan, H., Abrahamyan, R., Asmaryan, S., Muradyan, V., Tepanosyan, G., Guigoz, Y., and Giuliani, G. (2022). Shoreline delineation service: using an earth observation data cube and sentinel 2 images for coastal monitoring. *Earth Science Informatics*, pages 1–10.

Babakhin, Y., Sanakoyeu, A., and Kitamura, H. (2019). Semi-supervised segmentation of salt bodies in seismic images using an ensemble of convolutional neural networks. In *German Conference on Pattern Recognition*, pages 218–231. Springer.

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495.

Balado, J., Olabarria, C., Martínez-Sánchez, J., Rodríguez-Pérez, J. R., and Pedro, A. (2021). Semantic segmentation of major macroalgae in coastal environments using high-resolution ground imagery and deep learning. *International Journal of Remote Sensing*, 42(5):1785–1800.

Ballard, D. H. (1981). Generalizing the Hough transform to detect arbitrary shapes. *Pattern recognition*, 13(2):111–122.

Banerjee, S. and Shanmugam, P. (2021). Novel method for reconstruction of hyperspectral resolution images from multispectral data for complex coastal and inland waters. *Advances in Space Research*, 67(1):266–289.

Bansod, B., Singh, R., Thakur, R., and Singhal, G. (2017). A comparision between satellite based and drone based remote sensing technology to achieve sustainable develop-

ment: A review. *Journal of Agriculture and Environment for International Development (JAEID)*, 111(2):383–407.

Basedow, R. W., Carmer, D. C., and Anderson, M. E. (1995). HYDICE system: Implementation and performance. In *Imaging Spectrometry*, volume 2480, pages 258–267. SPIE.

Baxter, J. (1997). A Bayesian/information theoretic model of learning to learn via multiple task sampling. *Machine learning*, 28(1):7–39.

Behmann, J., Acebron, K., Emin, D., Bennertz, S., Matsubara, S., Thomas, S., Bohnenkamp, D., Kuska, M. T., Jussila, J., Salo, H., et al. (2018). Specim IQ: evaluation of a new, miniaturized handheld hyperspectral camera and its application for plant phenotyping and disease detection. *Sensors*, 18(2):441.

Belgiu, M., Hofer, B., and Hofmann, P. (2014). Coupling formalized knowledge bases with object-based image analysis. *Remote Sensing Letters*, 5(6):530–538.

Bellwood, D. R., Hughes, T. P., Folke, C., and Nyström, M. (2004). Confronting the coral reef crisis. *Nature*, 429(6994):827–833.

Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828.

Bengoufa, S., Niculescu, S., Mihoubi, M. K., Belkessa, R., Rami, A., Rabehi, W., and Abbad, K. (2021). Machine learning and shoreline monitoring using optical satellite

images: case study of the Mostaganem shoreline, Algeria. *Journal of applied remote sensing*, 15(2):026509.

Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I., and Heynen, M. (2004). Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of photogrammetry and remote sensing*, 58(3-4):239–258.

Berni, J. A., Zarco-Tejada, P. J., Suárez, L., and Fereres, E. (2009). Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle. *IEEE Transactions on geoscience and Remote Sensing*, 47(3):722–738.

Bins, L. S., Fonseca, L. G., Erthal, G. J., and Ii, F. M. (1996). Satellite imagery segmentation: a region growing approach. *Simpósio Brasileiro de Sensoriamento Remoto*, 8(1996):677–680.

Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., and Chanussot, J. (2013). Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and remote sensing magazine*, 1(2):6–36.

Bischke, B., Helber, P., Folz, J., Borth, D., and Dengel, A. (2019). Multi-task learning for segmentation of building footprints with deep neural networks. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 1480–1484. IEEE.

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS journal of photogrammetry and remote sensing*, 65(1):2–16.

Blaschke, T., Lang, S., and Hay, G. (2008). *Object-based image analysis: spatial concepts for knowledge-driven remote sensing applications*. Springer Science & Business Media.

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152.

Bouma, T., De Vries, M., Low, E., Peralta, G., Tánczos, I. v., van de Koppel, J., and Herman, P. M. J. (2005). Trade-offs related to ecosystem engineering: A case study on stiffness of emerging macrophytes. *Ecology*, 86(8):2187–2199.

Bowler, E., Fretwell, P. T., French, G., and Mackiewicz, M. (2020). Using deep learning to count albatrosses from space: Assessing results in light of ground truth uncertainty. *Remote Sensing*, 12(12):2026.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Brown, M., Lowe, D. G., et al. (2003). Recognising panoramas. In *ICCV*, volume 3, page 1218.

Burke, L., Kura, Y., Kassem, K., Revenga, C., Spalding, M., McAllister, D., and Caddy, J. (2001). *Coastal ecosystems*. World Resources Institute Washington, DC.

Butler, J. D., Purkis, S. J., Yousif, R., Al-Shaikh, I., and Warren, C. (2020). A high-resolution remotely sensed benthic habitat map of the qatari coastal zone. *Marine Pollution Bulletin*, 160:111634.

Campbell, A. and Wang, Y. (2019). High spatial resolution remote sensing for salt marsh mapping and change analysis at Fire Island National Seashore. *Remote Sensing*, 11(9):1107.

Cao, W., Li, J., Liu, J., and Zhang, P. (2016). Two improved segmentation algorithms for whole cardiac CT sequence images. In *2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 346–351. IEEE.

Cao, X., Du, H., Tong, X., Dai, Q., and Lin, S. (2011). A prism-mask system for multispectral video acquisition. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2423–2435.

Caruana, R. (1997). Multitask learning. *Machine learning*, 28(1):41–75.

Carvalho, M., Le Saux, B., Trouvé-Peloux, P., Champagnat, F., and Almansa, A. (2019). Multitask learning of height and semantics from aerial images. *IEEE Geoscience and Remote Sensing Letters*, 17(8):1391–1395.

Casals-Carrasco, P., Kubo, S., and Madhavan, B. B. (2000). Application of spectral mixture analysis for terrain evaluation studies. *International Journal of Remote Sensing*, 21(16):3039–3055.

Castilla, G. and Hay, G. (2008). Image objects and geographic objects. In *Object-based image analysis*, pages 91–110. Springer.

Castillo-Navarro, J., Le Saux, B., Boulch, A., Audebert, N., and Lefèvre, S. (2021). Semi-supervised semantic segmentation in Earth observation: The minifrance suite, dataset analysis and multi-task network study. *Machine Learning*, pages 1–36.

Catarino, M. D., Silva, A., and Cardoso, S. M. (2018). Phycochemical constituents and biological activities of Fucus spp. *Marine drugs*, 16(8):249.

Chabot, D., Craik, S. R., and Bird, D. M. (2015). Population census of a large common tern colony with a small unmanned aircraft. *PloS one*, 10(4):e0122588.

Chakrabarti, A. and Zickler, T. (2011). Statistics of real-world hyperspectral images. In *CVPR 2011*, pages 193–200. IEEE.

Chand, S. and Bollard, B. (2021). Low altitude spatial assessment and monitoring of intertidal seagrass meadows beyond the visible spectrum using a remotely piloted aircraft system. *Estuarine, Coastal and Shelf Science*, page 107299.

Chang, C.-I. (2013). *Hyperspectral data processing: algorithm design and analysis*. John Wiley & Sons.

Chang, C.-I. and Plaza, A. (2006). A fast iterative algorithm for implementation of pixel purity index. *IEEE Geoscience and Remote Sensing Letters*, 3(1):63–67.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2017a). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848.

Chen, L.-C., Papandreou, G., Schroff, F., and Adam, H. (2017b). Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818.

Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR.

Cheng, H.-D., Jiang, X. H., Sun, Y., and Wang, J. (2001). Color image segmentation: advances and prospects. *Pattern recognition*, 34(12):2259–2281.

Chi, M., Plaza, A., Benediktsson, J. A., Sun, Z., Shen, J., and Zhu, Y. (2016). Big data for remote sensing: Challenges and opportunities. *Proceedings of the IEEE*, 104(11):2207–2219.

Chopra, R., Verma, V., and Sharma, P. (2001). Mapping, monitoring and conservation of Harike wetland ecosystem, Punjab, India, through remote sensing. *International Journal of Remote Sensing*, 22(1):89–98.

Clark, A., Moorman, B., Whalen, D., and Vieira, G. (2022). Multiscale object-based classification and feature extraction along Arctic coasts. *Remote Sensing*, 14(13):2982.

Cohen, J. B. and Kappauf, W. E. (1982). Metameric color stimuli, fundamental metamers, and Wyszecki's metameric blacks. *The American journal of psychology*, pages 537–564.

Collin, A., Lambert, N., and Etienne, S. (2018a). Satellite-based salt marsh elevation, vegetation height, and species composition mapping using the superspectral WorldView-3 imagery. *International Journal of Remote Sensing*, 39(17):5619–5637.

Collin, A., Ramambason, C., Pastol, Y., Casella, E., Rovere, A., Thiault, L., Espiau, B., Siu, G., Lerouvreur, F., Nakamura, N., et al. (2018b). Very high resolution mapping of

coral reef state using airborne bathymetric LiDar surface-intensity and drone imagery. *International journal of remote sensing*, 39(17):5676–5688.

Congalton, R. G. (1991). Remote sensing and geographic information system data integration: error sources and. *Photogrammetric Engineering & Remote Sensing*, 57(6):677–687.

Cook, K. L. (2017). An evaluation of the effectiveness of low-cost UAVs and structure from motion for geomorphic change detection. *Geomorphology*, 278:195–208.

Costa, H., Foody, G. M., and Boyd, D. S. (2018). Supervised methods of image segmentation accuracy assessment in land cover mapping. *Remote Sensing of Environment*, 205:338–351.

Cracknell, A. (1999). Remote sensing techniques in estuaries and coastal zones an update. *International Journal of Remote Sensing*, 20(3):485–496.

Cubero-Castan, M., Schneider-Zapp, K., Bellomo, M., Shi, D., Rehak, M., and Strecha, C. (2018). Assessment of the radiometric accuracy in a target less work flow using Pix4D software. In *2018 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–4. IEEE.

Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., and Le, Q. V. (2019). Autoaugment: Learning augmentation strategies from data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 113–123.

Cubuk, E. D., Zoph, B., Shlens, J., and Le, Q. V. (2020). Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the*

IEEE/CVF conference on computer vision and pattern recognition workshops, pages 702–703.

Cui, F. and Jiang, J. (2023). MTSCD-Net: A network based on multi-task learning for semantic change detection of bitemporal remote sensing images. *International Journal of Applied Earth Observation and Geoinformation*, 118:103294.

Cui, W., Liu, Y., Li, Y., Guo, M., Li, Y., Li, X., Wang, T., Zeng, X., and Ye, C. (2019). Semi-supervised brain lesion segmentation with an adapted mean teacher model. In *International Conference on Information Processing in Medical Imaging*, pages 554–565. Springer.

Cunliffe, A. M., Brazier, R. E., and Anderson, K. (2016). Ultra-fine grain landscape-scale quantification of dryland vegetation structure with drone-acquired structure-from-motion photogrammetry. *Remote Sensing of Environment*, 183:129–143.

Dahdouh-Guebas, F. (2002). The use of remote sensing and GIS in the sustainable management of tropical coastal ecosystems. *Environment, development and sustainability*, 4(2):93–112.

Dahl, G. E., Sainath, T. N., and Hinton, G. E. (2013). Improving deep neural networks for LVCSR using rectified linear units and dropout. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 8609–8613. IEEE.

Dai, J., He, K., and Sun, J. (2015). Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision*, pages 1635–1643.

Dai, Z., Yang, Z., Yang, F., Cohen, W. W., and Salakhutdinov, R. R. (2017). Good semi-supervised learning that requires a bad gan. *Advances in neural information processing systems*, 30.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. Ieee.

Daughtry, C. S., Walthall, C., Kim, M., De Colstoun, E. B., and McMurtrey Iii, J. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote sensing of Environment*, 74(2):229–239.

Davies, J., Wray, B., and Brazier, P. (2017). Intertidal SAC monitoring of Zostera marina at Porth Dinllaen, Pen.

Davis, C. and Wang, X. (2003). Planimetric accuracy of Ikonos 1 m panchromatic orthoimage products and their utility for local government GIS basemap applications. *International Journal of Remote Sensing*, 24(22):4267–4288.

De Giglio, M., Greggio, N., Goffo, F., Merloni, N., Dubbini, M., and Barbarella, M. (2019). Comparison of pixel-and object-based classification methods of unmanned aerial vehicle data applied to coastal dune vegetation communities: Casal borsetti case study. *Remote sensing*, 11(12):1416.

DeBell, L., Anderson, K., Brazier, R. E., King, N., and Jones, L. (2015). Water resource management at catchment scales using lightweight UAVs: Current capabilities and future perspectives. *Journal of Unmanned Vehicle Systems*, 4(1):7–30.

Deeb, R., Van de Weijer, J., Muselet, D., Hebert, M., and Tremeau, A. (2019). Deep spectral reflectance and illuminant estimation from self-interreflections. *JOSA A*, 36(1):105–114.

Demuro, M. and Chisholm, L. (2003). Assessment of Hyperion for characterizing mangrove communities. In *Proceedings of the 12th JPL AVIRIS airborne earth science workshop, Pasadena, CA, USA*, volume 31.

Deng, L., Sun, J., Chen, Y., Lu, H., Duan, F., Zhu, L., and Fan, T. (2021). M2H-Net: A reconstruction method for hyperspectral remotely sensed imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173:323–348.

Dennison, W. C., Orth, R. J., Moore, K. A., Stevenson, J. C., Carter, V., Kollar, S., Bergstrom, P. W., and Batiuk, R. A. (1993). Assessing water quality with submersed aquatic vegetation: habitat requirements as barometers of Chesapeake Bay health. *BioScience*, 43(2):86–94.

Derpanis, K. G. (2010). Overview of the RANSAC algorithm. *Image Rochester NY*, 4(1):2–3.

Descour, M. and Dereniak, E. (1995). Computed-tomography imaging spectrometer: experimental calibration and reconstruction results. *Applied optics*, 34(22):4817–4826.

Detka, J., Coyle, H., Gomez, M., and Gilbert, G. S. (2023). A drone-powered deep learning methodology for high precision remote sensing in California's coastal shrubs. *Drones*, 7(7):421.

DeVries, T. and Taylor, G. W. (2017). Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552*.

Dewi, R. S., Bijker, W., Stein, A., and Marfai, M. A. (2016). Fuzzy classification for shoreline change monitoring in a part of the northern coastal area of Java, Indonesia. *Remote sensing*, 8(3):190.

Dolan, R., Fenster, M. S., and Holme, S. J. (1991). Temporal analysis of shoreline recession and accretion. *Journal of coastal research*, pages 723–744.

Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., and Darrell, T. (2014). Decaf: A deep convolutional activation feature for generic visual recognition. In *International conference on machine learning*, pages 647–655. PMLR.

Doren, R. F., Rutchey, K., and Welch, R. (1999). The Everglades: A perspective on the requirements and applications for vegetation map and database products. *Photogrammetric Engineering and Remote Sensing*, 65(2):155–161.

Drake, N. A., Mackin, S., and Settle, J. J. (1999). Mapping vegetation, soils, and geology in semiarid shrublands using spectral matching and mixture modeling of SWIR AVIRIS imagery. *Remote Sensing of Environment*, 68(1):12–25.

Dronova, I. (2015). Object-based image analysis in wetland research: A review. *Remote Sensing*, 7(5):6380–6413.

Duarte, C. M. (2002). The future of seagrass meadows. *Environmental conservation*, 29(2):192–206.

Duarte, C. M., Dennison, W. C., Orth, R. J., and Carruthers, T. J. (2008). The charisma of coastal ecosystems: addressing the imbalance. *Estuaries and coasts*, 31(2):233–238.

Duffy, J. P., Pratt, L., Anderson, K., Land, P. E., and Shutler, J. D. (2018). Spatial assessment of intertidal seagrass meadows using optical imaging systems and a lightweight drone. *Estuarine, Coastal and Shelf Science*, 200:169–180.

Duong, L., Cohn, T., Bird, S., and Cook, P. (2015). Low resource dependency parsing: Cross-lingual parameter sharing in a neural network parser. In *Proceedings of the 53rd annual meeting of the Association for Computational Linguistics and the 7th international joint conference on natural language processing (volume 2: short papers)*, pages 845–850.

Eickhoff, C. (2018). Cognitive biases in crowdsourcing. In *Proceedings of the eleventh ACM international conference on web search and data mining*, pages 162–170.

Eickhoff, C. and de Vries, A. P. (2013). Increasing cheat robustness of crowdsourcing tasks. *Information retrieval*, 16(2):121–137.

Eigen, D. and Fergus, R. (2015). Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In *Proceedings of the IEEE international conference on computer vision*, pages 2650–2658.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The quarterly journal of economics*, pages 643–669.

Evans, L. C. (1997). Partial differential equations and Monge-Kantorovich mass transfer. *Current developments in mathematics*, 1997(1):65–126.

Everingham, M. and Winn, J. (2012). The PASCAL visual object classes challenge 2012 (VOC2012) development kit. *Pattern Anal. Stat. Model. Comput. Learn., Tech. Rep*, 2007:1–45.

Fakiris, E., Blondel, P., Papatheodorou, G., Christodoulou, D., Dimas, X., Georgiou, N., Kordella, S., Dimitriadis, C., Rzhanov, Y., Geraga, M., et al. (2019). Multi-frequency, multi-sonar mapping of shallow habitats—efficacy and management implications in the national marine park of Zakynthos, Greece. *Remote Sensing*, 11(4):461.

Fan, J., Yau, D. K., Elmagarmid, A. K., and Aref, W. G. (2001). Automatic image segmentation by integrating color-edge extraction and seeded region growing. *IEEE transactions on image processing*, 10(10):1454–1466.

Farabet, C., Couprie, C., Najman, L., and LeCun, Y. (2012). Learning hierarchical features for scene labeling. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1915–1929.

Farahani, A., Voghoei, S., Rasheed, K., and Arabnia, H. R. (2021). A brief review of domain adaptation. *Advances in Data Science and Information Engineering: Proceedings from ICDATA 2020 and IKE 2020*, pages 877–894.

Fladeland, M., Berthold, R., Monforton, L., Kolyer, R., Lobitz, B., and Sumich, M. (2008). The NASA SIERRA UAV: A new unmanned aircraft for earth science investigations. In *AGU Fall Meeting Abstracts*, volume 2008, pages B41A–0365.

Fladeland, M., Sumich, M., Lobitz, B., Kolyer, R., Herlth, D., Berthold, R., McKinnon, D., Monforton, L., Brass, J., and Bland, G. (2011). The NASA SIERRA science

demonstration programme and the role of small–medium unmanned aircraft for earth science investigations. *Geocarto International*, 26(2):157–163.

Flanders, D., Hall-Beyer, M., and Pereverzoff, J. (2003). Preliminary evaluation of eCognition object-based software for cut block delineation and feature extraction. *Canadian Journal of Remote Sensing*, 29(4):441–452.

Fonseca, M. S. and Bell, S. S. (1998). Influence of physical setting on seagrass landscapes near Beaufort, North Carolina, USA. *Marine Ecology Progress Series*, 171:109–121.

Fonseca, M. S., Zieman, J. C., Thayer, G. W., and Fisher, J. S. (1983). The role of current velocity in structuring eelgrass (Zostera marina L.) meadows. *Estuarine, Coastal and Shelf Science*, 17(4):367–380.

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote sensing of environment*, 80(1):185–201.

Foody, G. M., Campbell, N., Trodd, N., and Wood, T. (1992). Derivation and applications of probabilistic measures of class membership from the maximum-likelihood classification. *Photogrammetric engineering and remote sensing*, 58(9):1335–1341.

Fourqurean, J. W., Duarte, C. M., Kennedy, H., Marbà, N., Holmer, M., Mateo, M. A., Apostolaki, E. T., Kendrick, G. A., Krause-Jensen, D., McGlathery, K. J., et al. (2012). Seagrass ecosystems as a globally significant carbon stock. *Nature geoscience*, 5(7):505–509.

French, G., Laine, S., Aila, T., Mackiewicz, M., and Finlayson, G. (2019). Semi-supervised semantic segmentation needs strong, varied perturbations. *arXiv preprint arXiv:1906.01916*.

French, G., Laine, S., Aila, T., Mackiewicz, M., and Finlayson, G. (2020a). Semi-supervised semantic segmentation needs strong, varied perturbations. In *British Machine Vision Conference*, number 31.

French, G., Oliver, A., and Salimans, T. (2020b). Milking cowmask for semi-supervised image classification. *arXiv preprint arXiv:2003.12022*.

Fuller, R. (1987). Vegetation establishment on shingle beaches. *The Journal of Ecology*, pages 1077–1089.

Fuller, R., Groom, G., Mugisha, S., Ipulet, P., Pomeroy, D., Katende, A., Bailey, R., and Ogutu-Ohwayo, R. (1998). The integration of field survey and remote sensing for biodiversity assessment: a case study in the tropical forests and wetlands of Sango Bay, Uganda. *Biological conservation*, 86(3):379–391.

Fuller, R. and Randall, R. (1988). The Orford Shingles, Suffolk, UK—classic conflicts in coastline management. *Biological Conservation*, 46(2):95–114.

Gao, H., Tang, Y., Jing, L., Li, H., and Ding, H. (2017). A novel unsupervised segmentation quality evaluation method for remote sensing images. *Sensors*, 17(10):2427.

Gao, H., Yuan, H., Wang, Z., and Ji, S. (2019). Pixel transposed convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 42(5):1218–1227.

Gao, J. (1999). A comparative study on spatial and spectral resolutions of satellite data in mapping mangrove forests. *International journal of remote sensing*, 20(14):2823–2833.

Gao, K., Yu, A., You, X., Qiu, C., Liu, B., and Zhang, F. (2023). Cross-domain multi-prototypes with contradictory structure learning for semi-supervised domain adaptation segmentation of remote sensing images. *Remote Sensing*, 15(13):3398.

Gens, R. (2010). Remote sensing of coastlines: detection, extraction and monitoring. *International Journal of Remote Sensing*, 31(7):1819–1836.

Gera, A., Pagès, J. F., Romero, J., and Alcoverro, T. (2013). Combined effects of fragmentation and herbivory on Posidonia oceanica seagrass ecosystems. *Journal of ecology*, 101(4):1053–1061.

Ghassemian, H. (2016). A review of remote sensing image fusion methods. *Information Fusion*, 32:75–89.

Ghatkar, J. G., Singh, R. K., and Shanmugam, P. (2019). Classification of algal bloom species from remote sensing data using an extreme gradient boosted decision tree model. *International Journal of Remote Sensing*, 40(24):9412–9438.

Gitelson, A. A., Kaufman, Y. J., Stark, R., and Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote sensing of Environment*, 80(1):76–87.

Goetz, A. F. (2009). Three decades of hyperspectral remote sensing of the Earth: A personal view. *Remote sensing of environment*, 113:S5–S16.

Gonçalves, J. and Henriques, R. (2015). UAV photogrammetry for topographic monitoring of coastal areas. *ISPRS journal of Photogrammetry and Remote Sensing*, 104:101–111.

Gong, L., Zhu, C., Luo, Y., and Fu, X. (2022). Spectral Reflectance Reconstruction from Red-Green-Blue (RGB) Images for Chlorophyll Content Detection. *Applied Spectroscopy*, page 00037028221139871.

Gong, P., Marceau, D. J., and Howarth, P. J. (1992). A comparison of spatial feature extraction algorithms for land-use classification with SPOT HRV data. *Remote sensing of environment*, 40(2):137–151.

Gong, P. and Zhang, A. (1999). Noise effect on linear spectral unmixing. *Geographic Information Sciences*, 5(1):52–57.

Gonzalez, P., Tack, K., Geelen, B., Masschelein, B., Charle, W., Vereecke, B., and Lambrechts, A. (2016). A novel CMOS-compatible, monolithically integrated line-scan hyperspectral imager covering the VIS-NIR range. In *Next-Generation Spectroscopic Technologies IX*, volume 9855, pages 129–137. SPIE.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. `http://www.deeplearningbook.org`.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11):139–144.

Gould, W. (2000). Remote sensing of vegetation, plant species richness, and regional biodiversity hotspots. *Ecological applications*, 10(6):1861–1870.

Grauman, K. and Darrell, T. (2005). The pyramid match kernel: Discriminative classification with sets of image features. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 2, pages 1458–1465. IEEE.

Gray, P. C., Ridge, J. T., Poulin, S. K., Seymour, A. C., Schwantes, A. M., Swenson, J. J., and Johnston, D. W. (2018). Integrating drone imagery into high resolution satellite remote sensing assessments of estuarine environments. *Remote Sensing*, 10(8):1257.

Green, E., Clark, C., Mumby, P., Edwards, A., and Ellis, A. (1998). Remote sensing techniques for mangrove mapping. *International journal of remote sensing*, 19(5):935–956.

Grover, A. and Leskovec, J. (2016). node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864.

Gu, H., Li, H., Yan, L., Liu, Z., Blaschke, T., and Soergel, U. (2017). An object-based semantic classification method for high resolution remote sensing imagery using ontology. *Remote Sensing*, 9(4):329.

Gueguen, L., Koenig, J., Reeder, C., Barksdale, T., Saints, J., Stamatiou, K., Collins, J., and Johnston, C. (2016). Mapping human settlements and population at country scale from VHR images. *IEEE journal of selected topics in applied earth observations and remote sensing*, 10(2):524–538.

Guirado, E., Tabik, S., Alcaraz-Segura, D., Cabello, J., and Herrera, F. (2017). Deep-learning versus OBIA for scattered shrub detection with Google earth imagery: Ziziphus Lotus as case study. *Remote Sensing*, 9(12):1220.

Gutiérrez, S., Wendel, A., and Underwood, J. (2019). Spectral filter design based on in-field hyperspectral imaging and machine learning for mango ripeness estimation. *Computers and Electronics in Agriculture*, 164:104890.

Ha, N.-T., Manley-Harris, M., Pham, T.-D., and Hawes, I. (2021). Detecting multi-decadal changes in seagrass cover in tauranga harbour, new zealand, using landsat imagery and boosting ensemble classification techniques. *ISPRS International Journal of Geo-Information*, 10(6):371.

Hajian, S., Bonchi, F., and Castillo, C. (2016). Algorithmic bias: From discrimination discovery to fairness-aware data mining. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 2125–2126.

Hamdi, Z. M., Brandmeier, M., and Straub, C. (2019). Forest damage assessment using deep learning on high resolution remote sensing data. *Remote Sensing*, 11(17):1976.

Hantson, W., Kooistra, L., and Slim, P. A. (2012). Mapping invasive woody species in coastal dunes in the Netherlands: a remote sensing approach using LIDAR and high-resolution aerial photographs. *Applied vegetation science*, 15(4):536–547.

Haralick, R. M., Shanmugam, K., and Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6):610–621.

Hardin, P. J. and Jensen, R. R. (2011). Small-scale unmanned aerial vehicles in environmental remote sensing: Challenges and opportunities. *GIScience & Remote Sensing*, 48(1):99–111.

Hardisky, M., Gross, M., and Klemas, V. (1986). Remote sensing of coastal wetlands. *Bioscience*, 36(7):453–460.

Harley, M. D., Kinsela, M. A., Sánchez-García, E., and Vos, K. (2019). Shoreline change mapping using crowd-sourced smartphone images. *Coastal Engineering*, 150:175–189.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Hedley, J. D., Roelfsema, C., Brando, V., Giardino, C., Kutser, T., Phinn, S., Mumby, P. J., Barrilero, O., Laporte, J., and Koetz, B. (2018). Coral reef applications of Sentinel-2: Coverage, characteristics, bathymetry and benthic mapping with comparison to Landsat 8. *Remote sensing of environment*, 216:598–614.

Held, A., Ticehurst, C., Lymburner, L., and Williams, N. (2003). High resolution mapping of tropical mangrove ecosystems using hyperspectral and radar remote sensing. *International Journal of Remote Sensing*, 24(13):2739–2759.

Herfort, B., Li, H., Fendrich, S., Lautenbach, S., and Zipf, A. (2019). Mapping human settlements with higher accuracy and less volunteer efforts by combining crowdsourcing and deep learning. *Remote Sensing*, 11(15):1799.

Hering, A., Hansen, L., Mok, T. C., Chung, A. C., Siebert, H., Häger, S., Lange, A., Kuckertz, S., Heldmann, S., Shao, W., et al. (2022). Learn2Reg: comprehensive multi-

task medical image registration challenge, dataset and evaluation in the era of deep learning. *IEEE Transactions on Medical Imaging*.

Hernandez-Matas, C., Zabulis, X., Triantafyllou, A., Anyfanti, P., Douma, S., and Argyros, A. A. (2017). FIRE: fundus image registration dataset. *Modeling and Artificial Intelligence in Ophthalmology*, 1(4):16–28.

Heumann, B. W. (2011a). An object-based classification of mangroves using a hybrid decision tree—support vector machine approach. *Remote Sensing*, 3(11):2440–2460.

Heumann, B. W. (2011b). Satellite remote sensing of mangrove forests: Recent advances and future opportunities. *Progress in Physical Geography*, 35(1):87–108.

Hinton, G. E. (2002). Training products of experts by minimizing contrastive divergence. *Neural computation*, 14(8):1771–1800.

Hobley, B., Arosio, R., French, G., Bremner, J., Dolphin, T., and Mackiewicz, M. (2021a). Semi-supervised segmentation for coastal monitoring seagrass using RPA imagery. *Remote Sensing*, 13(9):1741.

Hobley, B., Finlayson, G. D., Arosio, R., Bremner, J., Dolphin, T., and Mackiewicz, M. (2021b). Improving image registration using colour transfer methods in remote sensing applications. In *The Congress of the International Color Association*, number 14, pages 299–304.

Hobley, B., Mackiewicz, M., Bremner, J., Dolphin, T., and Arosio, R. (2023). Crowdsourcing experiment and fully convolutional neural networks for coastal remote sensing of

seagrass and macro-algae. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Hodges, J. and Howe, M. (1997). Milford Haven waterway monitoring of eelgrass, Zostera angustifolia, following the Sea empress oils spill. *Report to Shoreline & Terrestrial Task Group. Sea Empress Environmental Evaluation Committee*.

Hodgson, A., Kelly, N., and Peel, D. (2013). Unmanned aerial vehicles (UAVs) for surveying marine fauna: a dugong case study. *PloS one*, 8(11):e79556.

Hong, D., Qiu, C., Yu, A., Quan, Y., Liu, B., and Chen, X. (2023). Multi-task learning for building extraction and change detection from remote sensing images. *Applied Sciences*, 13(2):1037.

Hootsmans, M., Vermaat, J., and Van Vierssen, W. (1987). Seed-bank development, germination and early seedling survival of two seagrass species from the Netherlands: Zostera marina l. and Zostera noltii hornem. *Aquatic Botany*, 28(3-4):275–285.

Horwitz, H. M., Nalepka, R. F., Hyde, P. D., and Morgenstern, J. P. (1971). Estimating the proportions of objects within a single resolution element of a multispectral scanner. In *International Symposium on Remote Sensing of Environment, 7th, University of Michigan*.

Hossain, M., Bujang, J., Zakaria, M., and Hashim, M. (2015). Application of Landsat images to seagrass areal cover change analysis for Lawas, Terengganu and Kelantan of Malaysia. *Continental Shelf Research*, 110:124–148.

Hossain, M. D. and Chen, D. (2019). Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150:115–134.

Hu, J., Shen, L., and Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141.

Huang, H., Lan, Y., Yang, A., Zhang, Y., Wen, S., and Deng, J. (2020). Deep learning versus object-based image analysis (obia) in weed mapping of uav imagery. *International Journal of Remote Sensing*, 41(9):3446–3479.

Hueni, A. and Bialek, A. (2017). Cause, effect, and correction of field spectroradiometer interchannel radiometric steps. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(4):1542–1551.

Hung, W.-C., Tsai, Y.-H., Liou, Y.-T., Lin, Y.-Y., and Yang, M.-H. (2018). Adversarial learning for semi-supervised semantic segmentation. *arXiv preprint arXiv:1802.07934*.

Husson, E., Ecke, F., and Reese, H. (2016). Comparison of manual mapping and automated object-based image analysis of non-submerged aquatic vegetation from very-high-resolution uas images. *Remote Sensing*, 8(9):724.

Innangi, S., Di Martino, G., Romagnoli, C., and Tonielli, R. (2019). Seabed classification around Lampione islet, Pelagie Islands Marine Protected area, Sicily Channel, Mediterranean Sea. *Journal of Maps*, 15(2):153–164.

Inoue, T., Nagai, S., Yamashita, S., Fadaei, H., Ishii, R., Okabe, K., Taki, H., Honda, Y., Kajiwara, K., and Suzuki, R. (2014). Unmanned aerial survey of fallen trees in a deciduous broadleaved forest in eastern Japan. *PLoS one*, 9(10):e109881.

Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR.

Iscen, A., Tolias, G., Avrithis, Y., and Chum, O. (2019). Label propagation for deep semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5070–5079.

Islam, K. A., Hill, V., Schaeffer, B., Zimmerman, R., and Li, J. (2020). Semi-supervised adversarial domain adaptation for seagrass detection using multispectral images in coastal areas. *Data Science and Engineering*, 5:111–125.

Jaderberg, M., Simonyan, K., Zisserman, A., et al. (2015). Spatial transformer networks. *Advances in neural information processing systems*, 28.

Jain, R., Kasturi, R., Schunck, B. G., et al. (1995). *Machine vision*, volume 5. McGraw-hill New York.

Janowski, L., Madricardo, F., Fogarin, S., Kruss, A., Molinaroli, E., Kubowicz-Grajewska, A., and Tegowski, J. (2020). Spatial and temporal changes of tidal inlet using object-based image analysis of multibeam echosounder measurements: A case from the Lagoon of Venice, Italy. *Remote Sensing*, 12(13):2117.

Jensen, J. R. (1986). Introductory digital image processing: a remote sensing perspective. Technical report, Univ. of South Carolina, Columbus.

Jiménez, C., Niell, F. X., and Algarra, P. (1987). Photosynthetic adaptation of Zostera noltii Hornem. *Aquatic Botany*, 29(3):217–226.

Jing, W., Cui, B., Lu, Y., and Huang, L. (2021). BS-Net: Using Joint-Learning Boundary and Segmentation Network for Coastline Extraction from Remote Sensing Images. *Remote Sensing Letters*, 12(12):1260–1268.

Johnson, B. A. (2013). High-resolution urban land-cover classification using a competitive multi-scale object-based approach. *Remote Sensing Letters*, 4(2):131–140.

Kalluri, T., Varma, G., Chandraker, M., and Jawahar, C. (2019). Universal semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5259–5270.

Kamal, M. and Phinn, S. (2011). Hyperspectral data for mangrove species mapping: a comparison of pixel-based and object-based approach. remote sens 3: 2222–2242.

Kang, X., Li, S., and Benediktsson, J. A. (2013). Spectral–spatial hyperspectral image classification with edge-preserving filtering. *IEEE transactions on geoscience and remote sensing*, 52(5):2666–2677.

Kang, X., Zhuo, B., and Duan, P. (2019). Semi-supervised deep learning for hyperspectral image classification. *Remote sensing letters*, 10(4):353–362.

Kanji, G. K. (2006). *100 statistical tests*. Sage.

Kattenborn, T., Eichel, J., and Fassnacht, F. E. (2019a). Convolutional Neural Networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution UAV imagery. *Scientific reports*, 9(1):1–9.

Kattenborn, T., Lopatin, J., Förster, M., Braun, A. C., and Fassnacht, F. E. (2019b). UAV data as alternative to field sampling to map woody invasive species based on combined Sentinel-1 and Sentinel-2 data. *Remote sensing of environment*, 227:61–73.

Kavzoglu, T. and Tonbul, H. (2017). A comparative study of segmentation quality for multi-resolution segmentation and watershed transform. In *2017 8th International Conference on Recent Advances in Space Technologies (RAST)*, pages 113–117. IEEE.

Kazai, G. and Milic-Frayling, N. (2009). On the evaluation of the quality of relevance assessments collected through crowdsourcing. In *SIGIR 2009 Workshop on the Future of IR Evaluation*, volume 21.

Kearney, M. S., Stutzer, D., Turpie, K., and Stevenson, J. C. (2009). The effects of tidal inundation on the reflectance characteristics of coastal marsh vegetation. *Journal of Coastal Research*, 25(6):1177–1186.

Kerr, J. T. and Ostrovsky, M. (2003). From space to species: ecological applications for remote sensing. *Trends in ecology & evolution*, 18(6):299–305.

Kervadec, H., Dolz, J., Granger, É., and Ayed, I. B. (2019). Curriculum semi-supervised segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 568–576. Springer.

Khorram, S., Koch, F. H., van der Wiele, C. F., and Nelson, S. A. (2012). *Remote sensing*. Springer Science & Business Media.

Khosravan, N. and Bagci, U. (2018). Semi-supervised multi-task learning for lung cancer diagnosis. In *2018 40th Annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pages 710–713. IEEE.

Kim, B. and Ye, J. C. (2019). Mumford–Shah loss functional for image segmentation with deep learning. *IEEE Transactions on Image Processing*, 29:1856–1866.

Kim, J., Lee, J. K., and Lee, K. M. (2016). Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Klemas, V. V. (2009). Sensors and techniques for observing coastal ecosystems. *Remote sensing and geospatial technologies for coastal ecosystem assessment and management*, pages 17–44.

Klemas, V. V. (2015). Coastal and environmental remote sensing from unmanned aerial vehicles: An overview. *Journal of coastal research*, 31(5):1260–1267.

Koch, E. W. (1999). Sediment resuspension in a shallow Thalassia testudinum banks ex König bed. *Aquatic Botany*, 65(1-4):269–280.

Krähenbühl, P. and Koltun, V. (2011). Efficient inference in fully connected crfs with gaussian edge potentials. *Advances in neural information processing systems*, 24.

Krizhevsky, A., Hinton, G., et al. (2009). Learning multiple layers of features from tiny images.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105.

Kruse, F. A., Lefkoff, A., Boardman, J., Heidebrecht, K., Shapiro, A., Barloon, P., and Goetz, A. (1993). The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote sensing of environment*, 44(2-3):145–163.

Kulshreshtha, A. and Shanmugam, P. (2018). Assessment of trophic state and water quality of coastal-inland lakes based on Fuzzy Inference System. *Journal of Great Lakes Research*, 44(5):1010–1025.

La Rosa, L. E. C., Sothe, C., Feitosa, R. Q., de Almeida, C. M., Schimalski, M. B., and Oliveira, D. A. B. (2021). Multi-task fully convolutional network for tree species mapping in dense forests using small training hyperspectral data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 179:35–49.

Ladle, M. (1975). The Haustoriidae (Amphipoda) of Budle Bay, Northumberland. *Crustaceana*, 28(1):37–47.

Laine, S. and Aila, T. (2016). Temporal ensembling for semi-supervised learning. *arXiv preprint arXiv:1610.02242*.

Lamb, J. B., Van De Water, J. A., Bourne, D. G., Altier, C., Hein, M. Y., Fiorenza, E. A., Abu, N., Jompa, J., and Harvell, C. D. (2017). Seagrass ecosystems reduce exposure to bacterial pathogens of humans, fishes, and invertebrates. *Science*, 355(6326):731–733.

Lang, S. (2008). Object-based image analysis for remote sensing applications: modeling reality–dealing with complexity. In *Object-based image analysis*, pages 3–27. Springer.

Lazebnik, S., Schmid, C., and Ponce, J. (2006). Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, volume 2, pages 2169–2178. IEEE.

LeCun, Y., Bengio, Y., and Hinton, G. (2015a). Deep learning. *nature*, 521(7553):436–444.

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., and Jackel, L. (1989). Handwritten digit recognition with a back-propagation network. *Advances in neural information processing systems*, 2.

LeCun, Y. et al. (2015b). LeNet-5, Convolutional Neural Networks. *URL: http://yann. lecun. com/exdb/lenet*, 20(5):14.

Lee, D.-H. et al. (2013). Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, volume 3, page 896.

Leitão, P. J., Schwieder, M., Pötzschner, F., Pinto, J. R., Teixeira, A. M., Pedroni, F., Sanchez, M., Rogass, C., van der Linden, S., Bustamante, M. M., et al. (2018). From sample to pixel: multi-scale remote sensing data for upscaling aboveground carbon data in heterogeneous landscapes. *Ecosphere*, 9(8):e02298.

Lerman, R. I. and Yitzhaki, S. (1984). A note on the calculation and interpretation of the Gini index. *Economics Letters*, 15(3-4):363–368.

Li, A., Jiao, L., Zhu, H., Li, L., and Liu, F. (2021a). Multitask semantic boundary awareness network for remote sensing image segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14.

Li, J., Wu, C., Song, R., Li, Y., and Liu, F. (2020). Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from RGB images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 462–463.

Li, M., Ma, L., Blaschke, T., Cheng, L., and Tiede, D. (2016). A systematic comparison of different object-based classification techniques using high spatial resolution imagery in agricultural environments. *International Journal of Applied Earth Observation and Geoinformation*, 49:87–98.

Li, R., Liu, W., Yang, L., Sun, S., Hu, W., Zhang, F., and Li, W. (2018a). DeepUNet: A deep fully convolutional network for pixel-level sea-land segmentation. *IEEE journal of selected topics in applied earth observations and remote sensing*, 11(11):3954–3962.

Li, W., Fu, H., Yu, L., and Cracknell, A. (2017). Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. *Remote Sensing*, 9(1):22.

Li, W., Gao, H., Su, Y., and Momanyi, B. M. (2022). Unsupervised domain adaptation for remote sensing semantic segmentation with transformer. *Remote Sensing*, 14(19):4942.

Li, X., Yu, L., Chen, H., Fu, C.-W., and Heng, P.-A. (2018b). Semi-supervised skin lesion segmentation via transformation consistent self-ensembling model. *arXiv preprint arXiv:1808.03887*.

Li, Z., Ding, J., Zhang, H., and Feng, Y. (2021b). Classifying individual shrub species in UAV images—A case study of the Gobi region of Northwest China. *Remote Sensing*, 13(24):4995.

Lin, G., Shen, C., Van Den Hengel, A., and Reid, I. (2016). Efficient piecewise training of deep structured models for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3194–3203.

Lin, H., Shi, Z., and Zou, Z. (2017). Maritime semantic labeling of optical remote sensing images with multi-scale fully convolutional network. *Remote sensing*, 9(5):480.

Lin, M., Chen, Q., and Yan, S. (2013). Network in network. *arXiv preprint arXiv:1312.4400*.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer.

Lin, Y.-T. and Finlayson, G. D. (2020). Physically plausible spectral reconstruction from RGB images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 532–533.

Liu, C., Tao, R., Li, W., Zhang, M., Sun, W., and Du, Q. (2020a). Joint classification of hyperspectral and multispectral images for mapping coastal wetlands. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:982–996.

Liu, K., Sun, W., Shao, Y., Liu, W., Yang, G., Meng, X., Peng, J., Mao, D., and Ren, K. (2022a). Mapping Coastal Wetlands Using Transformer in Transformer Deep Network on China ZY1-02D Hyperspectral Satellite Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:3891–3903.

Liu, T., Abd-Elrahman, A., Morton, J., and Wilhelm, V. L. (2018). Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system. *GIScience & remote sensing*, 55(2):243–264.

Liu, W., Chen, X., Ran, J., Liu, L., Wang, Q., Xin, L., and Li, G. (2020b). LaeNet: a novel lightweight multitask CNN for automatically extracting lake area and shoreline from remote sensing images. *Remote Sensing*, 13(1):56.

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., and Xie, S. (2022b). A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11976–11986.

Loarie, S. R., Joppa, L. N., and Pimm, S. L. (2007). Satellites miss environmental priorities. *Trends in Ecology & Evolution*, 22(12):630–632.

Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440.

Long, M., Cao, Z., Wang, J., and Yu, P. S. (2017). Learning multiple tasks with multilinear relationship networks. *Advances in neural information processing systems*, 30.

Lopatin, J., Dolos, K., Kattenborn, T., and Fassnacht, F. E. (2019). How canopy shadow affects invasive plant species classification in high spatial resolution remote sensing. *Remote Sensing in Ecology and Conservation*, 5(4):302–317.

Lotze, H. K., Lenihan, H. S., Bourque, B. J., Bradbury, R. H., Cooke, R. G., Kay, M. C., Kidwell, S. M., Kirby, M. X., Peterson, C. H., and Jackson, J. B. (2006). Depletion, degradation, and recovery potential of estuaries and coastal seas. *Science*, 312(5781):1806–1809.

Louhaichi, M., Borman, M. M., and Johnson, D. E. (2001). Spatially located platform and aerial photography for documentation of grazing impacts on wheat. *Geocarto International*, 16(1):65–70.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision*, volume 2, pages 1150–1157. Ieee.

Lu, D.-S. and Chen, C.-C. (2008). Edge detection improvement by ant colony optimization. *Pattern Recognition Letters*, 29(4):416–425.

Lu, M., Liu, J., Wang, F., and Xiang, Y. (2022). Multi-Task Learning of Relative Height Estimation and Semantic Segmentation from Single Airborne RGB Images. *Remote Sensing*, 14(14):3450.

Lu, X., Zhong, Y., Zheng, Z., Liu, Y., Zhao, J., Ma, A., and Yang, J. (2019). Multi-scale and multi-task deep learning framework for automatic road extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 57(11):9362–9377.

Lu, Y., Kumar, A., Zhai, S., Cheng, Y., Javidi, T., and Feris, R. (2017). Fully-adaptive feature sharing in multi-task networks with applications in person attribute classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5334–5343.

Luc, P., Couprie, C., Chintala, S., and Verbeek, J. (2016). Semantic segmentation using adversarial networks. *arXiv preprint arXiv:1611.08408*.

MacIntyre, H. L., Geider, R. J., and Miller, D. C. (1996). Microphytobenthos: the ecological role of the "secret garden" of unvegetated, shallow-water marine habitats. I. Distribution, abundance and primary production. *Estuaries*, 19(2):186–201.

Macleod, R. D. and Congalton, R. G. (1998). A quantitative comparison of change-detection algorithms for monitoring eelgrass from remotely sensed data. *Photogrammetric engineering and remote sensing*, 64(3):207–216.

Macreadie, P., Baird, M., Trevathan-Tackett, S., Larkum, A., and Ralph, P. (2014). Quantifying and modelling the carbon sequestration capacity of seagrass meadows–a critical assessment. *Marine pollution bulletin*, 83(2):430–439.

Mallinis, G., Koutsias, N., Tsakiri-Strati, M., and Karteris, M. (2008). Object-based classification using Quickbird imagery for delineating forest vegetation polygons in a Mediterranean test site. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(2):237–250.

Marbà, N. and Duarte, C. M. (2010). Mediterranean warming triggers seagrass (Posidonia oceanica) shoot mortality. *Global Change Biology*, 16(8):2366–2375.

Marshall, W. and Boshuizen, C. (2013). Planet labs' remote sensing satellite system.

Martin, D. R., Fowlkes, C. C., and Malik, J. (2004). Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE transactions on pattern analysis and machine intelligence*, 26(5):530–549.

Martin, R., Ellis, J., Brabyn, L., and Campbell, M. (2020). Change-mapping of estuarine intertidal seagrass (Zostera muelleri) using multispectral imagery flown by remotely piloted aircraft (RPA) at Wharekawa Harbour, New Zealand. *Estuarine, Coastal and Shelf Science*, 246:107046.

McCarthy, M. J., Colna, K. E., El-Mezayen, M. M., Laureano-Rosario, A. E., Méndez-Lázaro, P., Otis, D. B., Toro-Farmer, G., Vega-Rodriguez, M., and Muller-Karger, F. E. (2017). Satellite remote sensing for coastal management: A review of successful applications. *Environmental management*, 60(2):323–339.

McGlathery, K. J., Reynolds, L. K., Cole, L. W., Orth, R. J., Marion, S. R., and Schwarzschild, A. (2012). Recovery trajectories during state change from bare sediment to eelgrass dominance. *Marine Ecology Progress Series*, 448:209–221.

Meyer, A. (1973). *An investigation into certain aspects of the ecology of Fenham flats and Budle Bay, Northumberland*. PhD thesis, Durham University.

Mezaris, V., Kompatsiaris, I., and Strintzis, M. G. (2004). Still image segmentation tools for object-based multimedia applications. *International Journal of pattern recognition and artificial intelligence*, 18(04):701–725.

Mieszkowska, N., Firth, L., and Bentley, M. (2013). Impacts of climate change on intertidal habitats. *MCCIP Science Review*, 2013:180–192.

Millennium ecosystem assessment, M. (2005). *Ecosystems and human well-being*, volume 5. Island press Washington, DC.

Miller, R. L., Del Castillo, C. E., and McKee, B. A. (2005). *Remote sensing of coastal aquatic environments*, volume 511. Springer.

Misra, I., Shrivastava, A., Gupta, A., and Hebert, M. (2016). Cross-stitch networks for multi-task learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3994–4003.

Mittal, S., Tatarchenko, M., and Brox, T. (2019). Semi-supervised semantic segmentation with high-and low-level consistency. *IEEE transactions on pattern analysis and machine intelligence*, 43(4):1369–1379.

Morgan, G. R. and Hodgson, M. E. (2021). A Post-Classification Change Detection Model with Confidences in High Resolution Multi-Date sUAS Imagery in Coastal South Carolina. *International Journal of Remote Sensing*, 42(11):4309–4336.

Morovic, P. and Finlayson, G. D. (2006). Metamer-set-based approach to estimating surface reflectance from camera RGB. *JOSA A*, 23(8):1814–1822.

Morris, J. T., Sundareshwar, P., Nietch, C. T., Kjerfve, B., and Cahoon, D. R. (2002). Responses of coastal wetlands to rising sea level. *Ecology*, 83(10):2869–2877.

Mountrakis, G., Im, J., and Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3):247–259.

Moussa, R. M., Bertucci, F., Jorissen, H., Gache, C., Waqalevu, V. P., Parravicini, V., Lecchini, D., and Galzin, R. (2020). Importance of intertidal seagrass beds as nursery area for coral reef fish juveniles (Mayotte, Indian Ocean). *Regional Studies in Marine Science*, 33:100965.

Mumby, P., Green, E., Edwards, A., and Clark, C. (1997). Measurement of seagrass standing crop using satellite and digital airborne remote sensing. *Marine ecology progress series*, 159:51–60.

Muñoz, X., Freixenet, J., Cufı, X., and Martı, J. (2003). Strategies for image segmentation combining region and boundary information. *Pattern recognition letters*, 24(1-3):375–392.

Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814.

Naumann, J. C., Young, D. R., and Anderson, J. E. (2008). Leaf chlorophyll fluorescence, reflectance, and physiological response to freshwater and saltwater flooding in the evergreen shrub, Myrica cerifera. *Environmental and Experimental Botany*, 63(1-3):402–409.

Nixon, S. W. (1981). Remineralization and nutrient cycling in coastal marine ecosystems. In *Estuaries and nutrients*, pages 111–138. Springer.

Noh, H., Hong, S., and Han, B. (2015). Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision*, pages 1520–1528.

Nolet, C., Poortinga, A., Roosjen, P., Bartholomeus, H., and Ruessink, G. (2014). Measuring and modeling the effect of surface moisture on the spectral reflectance of coastal beach sand. *PLoS One*, 9(11):e112151.

Noman, M. K., Islam, S. M. S., Abu-Khalaf, J., and Lavery, P. (2021). Multi-species Seagrass Detection using Semi-supervised learning. In *2021 36th International Conference on Image and Vision Computing New Zealand (IVCNZ)*, pages 1–6. IEEE.

Nussbaum, S. and Menz, G. (2008). eCognition image analysis software. In *Object-based image analysis and treaty verification*, pages 29–39. Springer.

Olive, P. (1993). Management of the exploitation of the lugworm Arenicola marina and the ragworm Nereis virens (Polychaeta) in conservation areas. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 3(1):1–24.

Olkin, I. and Pukelsheim, F. (1982). The distance between two random vectors with given dispersion matrices. *Linear Algebra and its Applications*, 48:257–263.

Olsson, V., Tranheden, W., Pinto, J., and Svensson, L. (2021). Classmix: Segmentation-based data augmentation for semi-supervised learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1369–1378.

Osco, L. P., Junior, J. M., Ramos, A. P. M., de Castro Jorge, L. A., Fatholahi, S. N., de Andrade Silva, J., Matsubara, E. T., Pistori, H., Gonçalves, W. N., and Li, J. (2021). A review on deep learning in UAV remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 102:102456.

Ouali, Y., Hudelot, C., and Tami, M. (2020). An overview of deep semi-supervised learning. *arXiv preprint arXiv:2006.05278*.

Pal, N. R. and Pal, S. K. (1993). A review on image segmentation techniques. *Pattern recognition*, 26(9):1277–1294.

Papakonstantinou, A., Batsaris, M., Spondylidis, S., and Topouzelis, K. (2021). A citizen Science Unmanned Aerial System Data Acquisition Protocol and Deep Learning Techniques for the Automatic Detection and Mapping of Marine Litter Concentrations in the Coastal Zone. *Drones*, 5(1):6.

Pascale, D. (2006). RGB coordinates of the Macbeth Color Checker. *The BabelColor Company*, 6.

Patel, C., Sharma, S., Pasquarella, V. J., and Gulshan, V. (2021). Evaluating self and semi-supervised methods for remote sensing segmentation tasks. *arXiv preprint arXiv:2111.10079*.

Pati, Y. C., Rezaiifar, R., and Krishnaprasad, P. S. (1993). Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Proceedings of 27th Asilomar conference on signals, systems and computers*, pages 40–44. IEEE.

Patil, K. D. (1975). Cochran's Q test: Exact distribution. *Journal of the American Statistical Association*, 70(349):186–189.

Patino, J. E. and Duque, J. C. (2013). A review of regional science applications of satellite remote sensing in urban settings. *Computers, Environment and Urban Systems*, 37:1–17.

Pease, C. B. (1991). *Satellite imaging instruments: principles, technologies and operational systems*. Prentice Hall.

Pendleton, E. A. (2010). Coastal vulnerability assessment of the Northern Gulf of Mexico to sea-level rise and coastal change.

Peñuelas, J. and Filella, I. (1998). Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends in plant science*, 3(4):151–156.

Pereira, H. M., Leadley, P. W., Proença, V., Alkemade, R., Scharlemann, J. P., Fernandez-Manjarrés, J. F., Araújo, M. B., Balvanera, P., Biggs, R., Cheung, W. W., et al. (2010). Scenarios for global biodiversity in the 21st century. *Science*, 330(6010):1496–1501.

Perez, D., Islam, K., Hill, V., Zimmerman, R., Schaeffer, B., Shen, Y., and Li, J. (2020). Quantifying seagrass distribution in coastal water with deep learning models. *Remote Sensing*, 12(10):1581.

Perone, C. S. and Cohen-Adad, J. (2018). Deep semi-supervised segmentation with weight-averaged consistency targets. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 12–19. Springer.

Pham, H., Dai, Z., Xie, Q., and Le, Q. V. (2021). Meta pseudo labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11557–11568.

Phinn, S., Hess, L., and Finlayson, C. M. (1999). An assessment of the usefulness of remote sensing for wetland inventory and monitoring in australia. *Techniques for enhanced wetland inventory and monitoring*, pages 44–83.

Pitié, F. and Kokaram, A. (2007). The linear monge-kantorovitch linear colour mapping for example-based colour transfer.

Plaza, A., Benediktsson, J. A., Boardman, J. W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A., et al. (2009). Recent advances in techniques for hyperspectral image processing. *Remote sensing of environment*, 113:S110–S122.

Polikar, R. (2012). Ensemble learning. In *Ensemble machine learning*, pages 1–34. Springer.

Pu, R., Bell, S., and Meyer, C. (2014). Mapping and assessing seagrass bed changes in central Florida's west coast using multitemporal Landsat TM imagery. *Estuarine, Coastal and Shelf Science*, 149:68–79.

Purkis, S. J., Gleason, A. C., Purkis, C. R., Dempsey, A. C., Renaud, P. G., Faisal, M., Saul, S., and Kerr, J. M. (2019). High-resolution habitat and bathymetry maps for 65,000 sq. km of Earth's remotest coral reefs. *Coral Reefs*, 38(3):467–488.

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., and Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote sensing of environment*, 48(2):119–126.

Qian, J., Zhou, Q., and Hou, Q. (2007). Comparison of pixel-based and object-oriented classification methods for extracting built-up areas in arid zone. In *ISPRS workshop on updating Geo-spatial databases with imagery & the 5th ISPRS workshop on DMGISs*, volume 8, pages 163–171. Citeseer.

Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1):81–106.

Quinlan, J. R. (1990). Decision trees and decision-making. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(2):339–346.

Radford, A., Metz, L., and Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.

Ramage, D. L. and Schiel, D. R. (1999). Patch dynamics and response to disturbance of the seagrass Zostera novazelandica on intertidal platforms in southern New Zealand. *Marine Ecology Progress Series*, 189:275–288.

Ramanath, R., Snyder, W. E., Yoo, Y., and Drew, M. S. (2005). Color image processing pipeline. *IEEE Signal processing magazine*, 22(1):34–43.

Ramsey III, E. W. and Laine, S. C. (1997). Comparison of Landsat Thematic Mapper and high resolution photography to identify change in complex coastal wetlands. *Journal of coastal research*, pages 281–292.

Randall, R. (2004). Management of coastal vegetated shingle in the United Kingdom. *Journal of Coastal Conservation*, 10(1):159–168.

Rasmus, A., Berglund, M., Honkala, M., Valpola, H., and Raiko, T. (2015). Semi-supervised learning with ladder networks. *Advances in neural information processing systems*, 28.

Rasuly, A., Naghdifar, R., and Rasoli, M. (2010). Monitoring of Caspian Sea coastline changes using object-oriented techniques. *Procedia Environmental Sciences*, 2:416–426.

Ren-hua, Z., Rao, N., and Liao, K. (1996). Approach for a vegetation index resistant to atmospheric effect. *Journal of Integrative Plant Biology*, 38(1).

Reus, G., Möller, T., Jäger, J., Schultz, S. T., Kruschel, C., Hasenauer, J., Wolff, V., and Fricke-Neuderth, K. (2018). Looking for seagrass: Deep learning for visual coverage estimation. In *2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO)*, pages 1–6. IEEE.

Richards, J. A. and Richards, J. (1999). *Remote sensing digital image analysis*, volume 3. Springer.

Riloff, E. (1996). Automatically generating extraction patterns from untagged text. In *Proceedings of the national conference on artificial intelligence*, pages 1044–1049.

Riloff, E. and Wiebe, J. (2003). Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 105–112.

Rish, I. et al. (2001). An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, volume 3, pages 41–46.

Roberts, D. A., Gardner, M., Church, R., Ustin, S., Scheer, G., and Green, R. (1998). Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. *Remote sensing of environment*, 65(3):267–279.

Rodwell, J. S. and nature conservation committee (GB), J. (2006). *National vegetation classification: Users' handbook*. Joint nature conservation committee Peterborough.

Rojas, R. (1996). The backpropagation algorithm. In *Neural networks*, pages 149–182. Springer.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

Rossiter, T., Furey, T., McCarthy, T., and Stengel, D. B. (2020). UAV-mounted hyperspectral mapping of intertidal macroalgae. *Estuarine, Coastal and Shelf Science*, 242:106789.

Roth, S. and Black, M. J. (2005). Fields of experts: A framework for learning image priors. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 860–867. IEEE.

Roth, S. and Black, M. J. (2009). Fields of experts. *International Journal of Computer Vision*, 82(2):205–229.

Rouse, J., Haas, R. H., Schell, J. A., Deering, D. W., et al. (1974). Monitoring vegetation systems in the great plains with ERTS. *NASA special publication*, 351(1974):309.

Ruder, S. (2017). An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*.

Ruiwen, N., Ye, M., Ji, L., Tong, Z., Tianye, L., Ruilong, F., He, G., Tianli, H., Yu, S., Ying, G., et al. (2022). Segmentation of remote sensing images based on U-Net multi-task learning. *Comput., Mater. Continua*, 73(2):3263–3274.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252.

Ryan, J. C., Hubbard, A. L., Box, J. E., Todd, J., Christoffersen, P., Carr, J. R., Holt, T. O., and Snooke, N. (2015). UAV photogrammetry and structure from motion to assess calving dynamics at Store Glacier, a large outlet draining the Greenland ice sheet. *The Cryosphere*, 9(1):1–11.

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., and Chen, X. (2016). Improved techniques for training gans. *Advances in neural information processing systems*, 29.

Sánchez, J., Perronnin, F., Mensink, T., and Verbeek, J. (2013). Image classification with the fisher vector: Theory and practice. *International journal of computer vision*, 105(3):222–245.

Sanjar, K., Bekhzod, O., Kim, J., Kim, J., Paul, A., and Kim, J. (2020). Improved U-net: fully convolutional network model for skin-lesion segmentation. *Applied Sciences*, 10(10):3658.

Saralioglu, E. and Gungor, O. (2020). Crowdsourcing in remote sensing: A review of applications and future directions. *IEEE Geoscience and Remote Sensing Magazine*, 8(4):89–110.

Schmidt, K., Skidmore, A., Kloosterman, E., Van Oosten, H., Kumar, L., and Janssen, J. (2004). Mapping coastal vegetation using an expert system and hyperspectral imagery. *Photogrammetric Engineering & Remote Sensing*, 70(6):703–715.

Scott, G. (1963). The ecology of shingle beach plants. *The Journal of Ecology*, pages 517–527.

Seas, U. R. and Plans, A. (2011). Percentage of total population living in coastal areas.

Serrano, L., Penuelas, J., and Ustin, S. L. (2002). Remote sensing of nitrogen and lignin in mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural signals. *Remote sensing of Environment*, 81(2-3):355–364.

Sheaves, M., Baker, R., Nagelkerken, I., and Connolly, R. M. (2015). True value of estuarine and coastal nurseries for fish: incorporating complexity and dynamics. *Estuaries and Coasts*, 38(2):401–414.

Shi, Z., Chen, C., Xiong, Z., Liu, D., and Wu, F. (2018). Hscnn+: Advanced cnn-based hyperspectral recovery from rgb images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 939–947.

Shih, F. Y. and Cheng, S. (2004). Adaptive mathematical morphology for edge linking. *Information sciences*, 167(1-4):9–21.

Short, F. T. and Wyllie-Echeverria, S. (1996). Natural and human-induced disturbance of seagrasses. *Environmental conservation*, pages 17–27.

Shorten, C. and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):1–48.

Shuman, C. S. and Ambrose, R. F. (2003). A comparison of remote sensing and ground-based methods for monitoring wetland restoration success. *Restoration Ecology*, 11(3):325–333.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Small, C. (2012). Spatiotemporal dimensionality and Time-Space characterization of multitemporal imagery. *Remote Sensing of Environment*, 124:793–809.

Smith, M. O., Adams, J. B., and Sabol, D. E. (1994). Spectral mixture analysis-new strategies for the analysis of multispectral data. In *Imaging spectrometry—a tool for environmental observations*, pages 125–143. Springer.

Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C. A., Cubuk, E. D., Kurakin, A., and Li, C.-L. (2020). Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, 33:596–608.

Sohn, Y. and McCoy, R. M. (1997). Mapping desert shrub rangeland using spectral unmixing and modeling spectral mixtures with TM data. *Photogrammetric Engineering and Remote Sensing*, 63(6):707–716.

Somers, B., Asner, G. P., Tits, L., and Coppin, P. (2011). Endmember variability in spectral mixture analysis: A review. *Remote Sensing of Environment*, 115(7):1603–1616.

Song, C. (2005). Spectral mixture analysis for subpixel vegetation fractions in the urban environment: How to incorporate endmember variability? *Remote sensing of environment*, 95(2):248–263.

Souly, N., Spampinato, C., and Shah, M. (2017). Semi supervised semantic segmentation using generative adversarial network. In *Proceedings of the IEEE international conference on computer vision*, pages 5688–5696.

Splinter, K. D. and Coco, G. (2021). Challenges and opportunities in coastal shoreline prediction. *Frontiers in Marine Science*, 8:1917.

Springenberg, J. T. (2015). Unsupervised and semi-supervised learning with categorical generative adversarial networks. *arXiv preprint arXiv:1511.06390*.

Sreekesh, S., Kaur, N., and Sreerama Naik, S. (2020). An OBIA and rule algorithm for coastline extraction from high-and medium-resolution multispectral remote sensing images. *Remote Sensing in Earth Systems Sciences*, 3(1):24–34.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958.

Su, T. and Zhang, S. (2017). Local and global evaluation for remote sensing image segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130:256–276.

Su, W., Li, J., Chen, Y., Liu, Z., Zhang, J., Low, T. M., Suppiah, I., and Hashim, S. A. M. (2008). Textural and local spatial statistics for the object-oriented classification of urban areas using high resolution imagery. *International journal of remote sensing*, 29(11):3105–3117.

Sun, X., Shi, A., Huang, H., and Mayer, H. (2020). BASNet: Boundary-aware semi-supervised semantic segmentation network for very high resolution remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:5398–5413.

Sutskever, I., Martens, J., Dahl, G., and Hinton, G. (2013). On the importance of initialization and momentum in deep learning. In *International conference on machine learning*, pages 1139–1147. PMLR.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., and Liu, C. (2018). A survey on deep transfer learning. In *International conference on artificial neural networks*, pages 270–279. Springer.

Tang, L. and Shao, G. (2015). Drone remote sensing for forestry research and practices. *Journal of Forestry Research*, 26(4):791–797.

Tarvainen, A. and Valpola, H. (2017). Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30.

Thoma, M. (2016). A survey of semantic segmentation. *arXiv preprint arXiv:1602.06541*.

Tillin, H. and Budd, G. (2016). Green seaweeds (Ulva spp. and Cladophora spp.) in shallow upper shore rockpools.

Torrey, L. and Shavlik, J. (2010). Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 242–264. IGI global.

Traganos, D., Aggarwal, B., Poursanidis, D., Topouzelis, K., Chrysoulakis, N., and Reinartz, P. (2018). Towards global-scale seagrass mapping and monitoring using Sentinel-2 on Google Earth Engine: The case study of the aegean and ionian seas. *Remote Sensing*, 10(8):1227.

Tremeau, A. and Borel, N. (1997). A region growing and merging algorithm to color segmentation. *Pattern recognition*, 30(7):1191–1203.

Tsiakos, C.-A. D. and Chalkias, C. (2023). Use of machine learning and remote sensing techniques for shoreline monitoring: A review of recent literature. *Applied Sciences*, 13(5):3268.

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*, 8(2):127–150.

Turner, D., Lucieer, A., and Watson, C. (2012). An automated technique for generating georectified mosaics from ultra-high resolution unmanned aerial vehicle (UAV) imagery, based on structure from motion (SfM) point clouds. *Remote sensing*, 4(5):1392–1410.

Valderrama-Landeros, L., Flores-de Santiago, F., Kovacs, J., and Flores-Verdugo, F. (2018). An assessment of commonly employed satellite-based remote sensors for mapping mangrove species in Mexico using an NDVI-based classification scheme. *Environmental monitoring and assessment*, 190(1):1–13.

Van der Meer, F. D. and Jia, X. (2012). Collinearity and orthogonality of endmembers in linear spectral unmixing. *International Journal of Applied Earth Observation and Geoinformation*, 18:491–503.

Vatsavai, R. R., Bright, E., Varun, C., Budhendra, B., Cheriyadat, A., and Grasser, J. (2011). Machine learning approaches for high-resolution urban land cover classification: a comparative study. In *Proceedings of the 2nd international conference on computing for geospatial research & applications*, pages 1–10.

Ventura, D., Bonifazi, A., Gravina, M. F., Belluscio, A., and Ardizzone, G. (2018). Mapping and classification of ecologically sensitive marine habitats using unmanned aerial vehicle (UAV) imagery and object-based image analysis (OBIA). *Remote Sensing*, 10(9):1331.

Ventura, D., Bruno, M., Lasinio, G. J., Belluscio, A., and Ardizzone, G. (2016). A low-cost drone based application for identifying and mapping of coastal fish nursery grounds. *Estuarine, Coastal and Shelf Science*, 171:85–98.

Verrelst, J., Schaepman, M. E., Koetz, B., and Kneubühler, M. (2008). Angular sensitivity analysis of vegetation indices derived from CHRIS/PROBA data. *Remote Sensing of Environment*, 112(5):2341–2353.

Vincent, L. and Soille, P. (1991). Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(06):583–598.

Volpi, M. and Tuia, D. (2018). Deep multi-task learning for a geographically-regularized semantic segmentation of aerial images. *ISPRS journal of photogrammetry and remote sensing*, 144:48–60.

Wagadarikar, A. A., Pitsianis, N. P., Sun, X., and Brady, D. J. (2009). Video rate spectral imaging using a coded aperture snapshot spectral imager. *Optics express*, 17(8):6368–6388.

Wagner, F. H., Sanchez, A., Tarabalka, Y., Lotte, R. G., Ferreira, M. P., Aidar, M. P., Gloor, E., Phillips, O. L., and Aragao, L. E. (2019). Using the U-net convolutional network to map forest types and disturbance in the Atlantic rainforest with very high resolution images. *Remote Sensing in Ecology and Conservation*, 5(4):360–375.

Walmsley, C. and Davy, A. (1997a). Germination characteristics of shingle beach species, effects of seed ageing and their implications for vegetation restoration. *Journal of Applied Ecology*, pages 131–142.

Walmsley, C. and Davy, A. (1997b). The restoration of coastal shingle vegetation: effects of substrate composition on the establishment of seedlings. *Journal of Applied ecology*, pages 143–153.

Wang, C., Pei, J., Wang, Z., Huang, Y., Wu, J., Yang, H., and Yang, J. (2020a). When deep learning meets multi-task learning in sar atr: Simultaneous target recognition and segmentation. *Remote Sensing*, 12(23):3863.

Wang, J., HQ Ding, C., Chen, S., He, C., and Luo, B. (2020b). Semi-supervised remote sensing image semantic segmentation via consistency regularization and average update of pseudo-label. *Remote Sensing*, 12(21):3603.

Wang, J.-X., Chen, S.-B., Ding, C. H., Tang, J., and Luo, B. (2022). Semi-supervised semantic segmentation of remote sensing images with iterative contrastive network. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5.

Wang, L. and Jia, X. (2009). Integration of soft and hard classifications using extended support vector machines. *IEEE Geoscience and Remote Sensing Letters*, 6(3):543–547.

Wang, L., Sousa, W., and Gong, P. (2004a). Integration of object-based and pixel-based classification for mapping mangroves with IKONOS imagery. *International journal of remote sensing*, 25(24):5655–5668.

Wang, L., Sousa, W. P., Gong, P., and Biging, G. S. (2004b). Comparison of IKONOS and Quickbird images for mapping mangrove species on the Caribbean coast of Panama. *Remote sensing of environment*, 91(3-4):432–440.

Wang, S., Chen, W., Xie, S. M., Azzari, G., and Lobell, D. B. (2020c). Weakly supervised deep learning for segmentation of remote sensing imagery. *Remote Sensing*, 12(2):207.

Wang, S., Di Tommaso, S., Faulkner, J., Friedel, T., Kennepohl, A., Strey, R., and Lobell, D. B. (2020d). Mapping crop types in southeast India with smartphone crowdsourcing and deep learning. *Remote Sensing*, 12(18):2957.

Wang, S., Quan, D., Liang, X., Ning, M., Guo, Y., and Jiao, L. (2018). A deep learning framework for remote sensing image registration. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145:148–164.

Watts, A. C., Ambrosia, V. G., and Hinkley, E. A. (2012). Unmanned aircraft systems in remote sensing and scientific research: Classification and considerations of use. *Remote Sensing*, 4(6):1671–1692.

Waycott, M., Duarte, C. M., Carruthers, T. J., Orth, R. J., Dennison, W. C., Olyarnik, S., Calladine, A., Fourqurean, J. W., Heck, K. L., Hughes, A. R., et al. (2009). Accelerating loss of seagrasses across the globe threatens coastal ecosystems. *Proceedings of the national academy of sciences*, 106(30):12377–12381.

Waycott, M., McKenzie, L. J., Mellors, J. E., Ellison, J. C., Sheaves, M. T., Collier, C., Schwarz, A.-M., et al. (2011). Vulnerability of mangroves, seagrasses and intertidal flats in the tropical pacific to climate change.

Weidmann, F., Jäger, J., Reus, G., Schultz, S. T., Kruschel, C., Wolff, V., and Fricke-Neuderth, K. (2019). A closer look at seagrass meadows: Semantic segmentation for visual coverage estimation. In *OCEANS 2019-Marseille*, pages 1–6. IEEE.

Whitfield, A. K. and Pattrick, P. (2015). Habitat type and nursery function for coastal marine fish species, with emphasis on the Eastern Cape region, South Africa. *Estuarine, Coastal and Shelf Science*, 160:49–59.

Willis, K. S. (2015). Remote sensing change detection for ecological monitoring in United States protected areas. *Biological Conservation*, 182:233–242.

Xia, X. and Kulis, B. (2017). W-net: A deep model for fully unsupervised image segmentation. *arXiv preprint arXiv:1711.08506*.

Xiaoqin, W., Miaomiao, W., Shaoqiang, W., and Yundong, W. (2015). Extraction of vegetation information from visible unmanned aerial vehicle images. *Transactions of the Chinese Society of Agricultural Engineering*, 31(5).

Xie, Q., Luong, M.-T., Hovy, E., and Le, Q. V. (2020). Self-training with noisy student improves imagenet classification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10687–10698.

Xiong, Z., Shi, Z., Li, H., Wang, L., Liu, D., and Wu, F. (2017). Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 518–525.

Xu, Y., Wu, L., Xie, Z., and Chen, Z. (2018). Building extraction in very high resolution remote sensing imagery using deep learning and guided filters. *Remote Sensing*, 10(1):144.

Xue, J. and Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *Journal of sensors*, 2017.

Yamakita, T., Sodeyama, F., Whanpetch, N., Watanabe, K., and Nakaoka, M. (2019). Application of deep learning techniques for determining the spatial extent and classification of seagrass beds, Trang, Thailand. *Botanica marina*, 62(4):291–307.

Yang, Y. and Hospedales, T. M. (2016). Trace norm regularised deep multi-task learning. *arXiv preprint arXiv:1606.04038*.

Yasuma, F., Mitsunaga, T., Iso, D., and Nayar, S. K. (2010). Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum. *IEEE transactions on image processing*, 19(9):2241–2253.

Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). How transferable are features in deep neural networks? *Advances in neural information processing systems*, 27.

You, K., Long, M., Cao, Z., Wang, J., and Jordan, M. I. (2019). Universal domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2720–2729.

Yuan, X., Shi, J., and Gu, L. (2021). A review of deep learning methods for semantic segmentation of remote sensing imagery. *Expert Systems with Applications*, 169:114417.

Yun, S., Han, D., Oh, S. J., Chun, S., Choe, J., and Yoo, Y. (2019). Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6023–6032.

Zaabar, N., Niculescu, S., and Kamel, M. M. (2022). Application of Convolutional Neural Networks with object-based image analysis for land cover and land use mapping in coastal areas: A case study in Ain Témouchent, Algeria. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:5177–5189.

Zeiler, M. D. and Fergus, R. (2014). Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer.

Zerrouki, N. and Bouchaffra, D. (2014). Pixel-based or object-based: Which approach is more appropriate for remote sensing image classification? In *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 864–869. IEEE.

Zhang, B., Sun, X., Gao, L., and Yang, L. (2011). Endmember extraction of hyperspectral remote sensing images based on the ant colony optimization (ACO) algorithm. *IEEE transactions on geoscience and remote sensing*, 49(7):2635–2646.

Zhang, H., Cisse, M., Dauphin, Y. N., and Lopez-Paz, D. (2017). mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*.

Zhang, J., Su, R., Fu, Q., Ren, W., Heide, F., and Nie, Y. (2022). A survey on computational spectral reconstruction methods from RGB to hyperspectral imaging. *Scientific reports*, 12(1):1–17.

Zhang, L., Wang, X., Zhang, J., Qiao, J., and Li, P. (2020). Reconstruction of Spectral Reflectance Based on Fusion Convolution Neural Network. In *Proceedings of the 2020 9th International Conference on Computing and Pattern Recognition*, pages 250–254.

Zhang, L., Zhang, L., Tao, D., and Huang, X. (2013). Sparse transfer manifold embedding for hyperspectral target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 52(2):1030–1043.

Zhang, T., Li, Q., Yang, X., Zhou, C., and Su, F. (2010). Automatic mapping aquaculture in coastal zone from TM imagery with OBIA approach. In *2010 18th International Conference on Geoinformatics*, pages 1–4. IEEE.

Zhang, Y.-J. (2006). An overview of image and video segmentation in the last 40 years. *Advances in Image and Video Segmentation*, pages 1–16.

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890.

Zheng, J.-Y., Hao, Y.-Y., Wang, Y.-C., Zhou, S.-Q., Wu, W.-B., Yuan, Q., Gao, Y., Guo, H.-Q., Cai, X.-X., and Zhao, B. (2022). Coastal Wetland Vegetation Classification Using Pixel-Based, Object-Based and Deep Learning Methods Based on RGB-UAV. *Land*, 11(11):2039.

Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., and Torr, P. H. (2015). Conditional random fields as recurrent neural networks. In *Proceedings of the IEEE international conference on computer vision*, pages 1529–1537.

Zhong, Z., Zheng, L., Kang, G., Li, S., and Yang, Y. (2020). Random erasing data augmentation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 13001–13008.

Zhou, J., Kwan, C., and Budavari, B. (2016). Hyperspectral image super-resolution: A hybrid color mapping approach. *Journal of Applied Remote Sensing*, 10(3):035024.

Zhou, Q. and Robson, M. (2001). Automated rangeland vegetation cover and density estimation using ground digital images and a spectral-contextual classifier. *International Journal of Remote Sensing*, 22(17):3457–3470.

Zhou, Y.-T., Venkateswar, V., and Chellappa, R. (1989). Edge detection and linear feature extraction using a 2-D random field model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(1):84–95.

Zhu, X. J. (2005). Semi-supervised learning literature survey.

Zhu, Y., Liu, K., Liu, L., Wang, S., and Liu, H. (2015). Retrieval of mangrove aboveground biomass at the individual species level with worldview-2 images. *Remote Sensing*, 7(9):12192–12214.

Zinnert, J. C., Nelson, J. D., and Hoffman, A. M. (2012). Effects of salinity on physiological responses and the photochemical reflectance index in two co-occurring coastal shrubs. *Plant and Soil*, 354(1):45–55.

Zitova, B. and Flusser, J. (2003). Image registration methods: a survey. *Image and vision computing*, 21(11):977–1000.

Zlateski, A., Jaroensri, R., Sharma, P., and Durand, F. (2018). On the importance of label quality for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1479–1487.

Zuva, T., Olugbara, O. O., Ojo, S. O., and Ngwira, S. M. (2011). Image segmentation, available techniques, developments and open issues. *Canadian Journal on Image Processing and Computer Vision*, 2(3):20–29.

A Colour transfer for improved image registration

As noted in Section 3.2 two cameras were used to capture aerial imagery of Budle Bay with complementing characteristics. The SONY ILCE-6000 camera has low spectral resolution with three wide banded filters but high spatial resolution with approximately 3 cm ground sample distance, whereas the MicaSense RedEdge3 camera has high spectral resolution with five narrow banded filters but low spatial resolution with approximately 8 cm sampling.

Image registration is described as aligning several images into a common image coordinate system. For remote sensing, registration is often a key pre-processing step for combining aerial imagery from multiple sensors at different spatial and spectral resolutions. A standard approach to automatic registration is to upsample the multispectral image to match the resolution of the corresponding high-resolution image using interpolation; find correspondences in pairs of images using SIFT features and compute an optimal geometric transformation based on matching correspondences [Brown et al., 2003]. However, errors in registration distort images by blurring edges of objects and affect any subsequent fusion method [Ghassemian, 2016].

Registration errors can be caused either by subpar geometric transformation models or distortions due to interpolation. In this section, the hypothesis is that subpar geometric transformation models fail to incorporate colour information given SIFT functions on grey-scale images [Lowe, 1999]. Adding to this, the use of different cameras results in two different colour response distributions for each camera that needs to be accounted for in order to

incorporate colour information in the registration process. As such, each image is modelled as a continuous probability density function (pdf) with each pixel value as a realization of a colour random variable. Then, a linear colour transform is computed to minimise the underlying covariance shift in colour responses from multiple cameras. Transferring statistical moments between resulting pdfs is possible using the linear Monge-Kantorovitch (MK) solution [Olkin and Pukelsheim, 1982] which has shown to be effective in media production applications [Pitié and Kokaram, 2007]. Therefore, the MK solutions maps the colours from the high-resolution reference image to match those in a multispectral target image.

Given this, the automatic registration is applied, as described in Brown et al. [2003], was applied on the following pairs of images: the high-resolution image from the SONY camera with the multispectral image from the RedEdge3, and the high-resolution image with transferred colour responses from the multispectral image with the same corresponding multispectral image. The goal here is to register the datasets from each sensor used to survey Budle Bay, and that the registration between the high-resolution image from the SONY ILCE-6000 with transferred colour responses improves the automatic registration process.

The method is evaluated using the processed VHR orthomosaics of Budle Bay. In Figure A.1, the reference image from the SONY camera and the multispectral corresponding capture using the RedEdge3 camera are displayed respectively in images (A) and (B) both which cover the same aerial extent but are not in registration. Using the MK transform, the tone and colouring of the target image is mapped onto the reference image, with the MK transformed (C) image looking like the target, even though it is not also registered. By minimizing the shift in colour responses resulting from multiple cameras the evaluation notes that the subsequent registration process is improved.

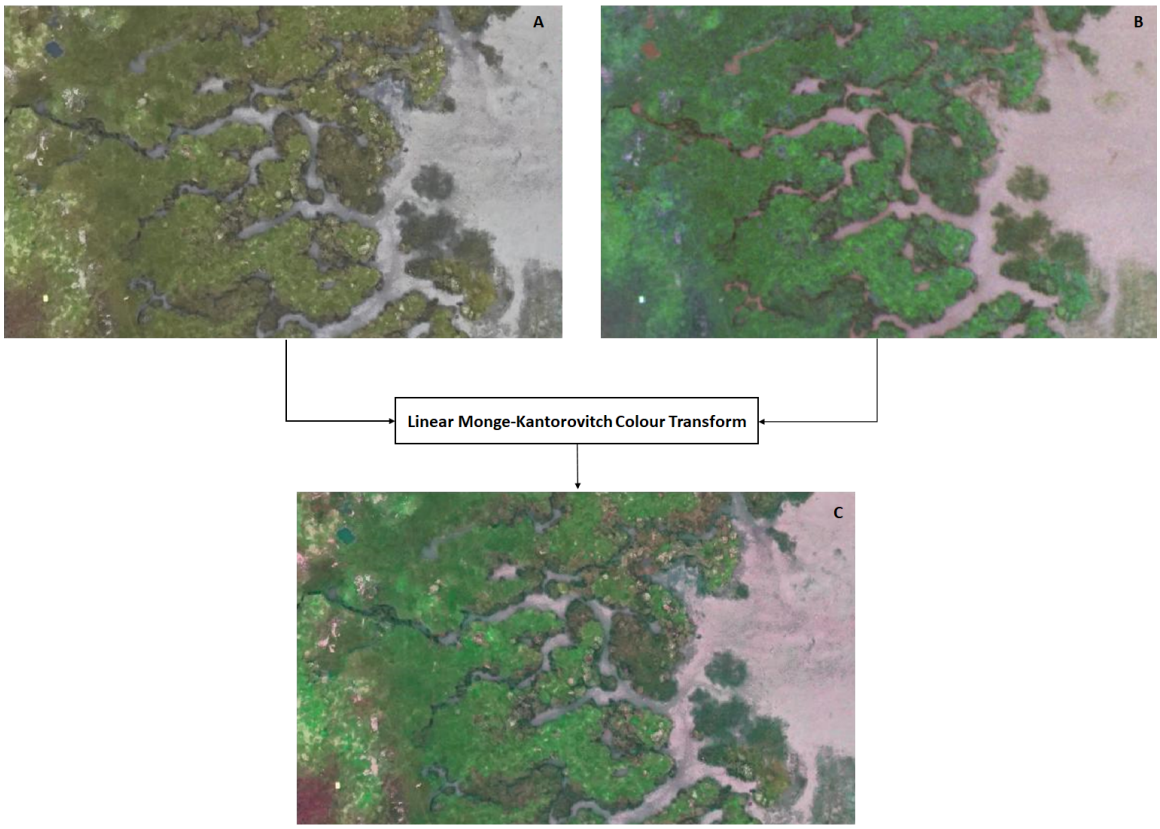

Figure A.1: A is the high-resolution reference image with corresponding target image (B) and the application of the linear MK transform (C) mapping the colours in the reference image (A) to match those within the target image (B).

A.1 Background

Literature describes many different registration methods that can be applied to remote sensing. A review of image registration methods can be found in [Zitova and Flusser, 2003]. For this work, the pipeline based on [Brown et al., 2003] is considered. The first step is to match the resolution in pairs of images using interpolation. Generally, the image with coarser resolution is upsampled to match the high-resolution image. The second step is to extract SIFT features from the image pair [Lowe, 1999]. SIFT features describe local features in an image that are invariant to translation, rotation and scaling, and partially invariant to illumination. As described in Brown et al. [2003], SIFT features are suitable for finding correspondences in image pairs. Given several matching points, the parameters for

an affine transform are found in a linear least squares sense, and finally each transform is scored and accepted/rejected using RANSAC [Derpanis, 2010]. The best scoring transform warps pixels from both images into a common coordinate system.

Figure A.1 shows the affects of applying a linear transform based on the colour statistics of the reference and target images. This linear mapping is achieved by representing each image as a set of RGB colour samples, where in a probabilistic sense, each colour sample is a realization of a three dimensional colour variable. The distributions of colour samples for each image are denoted as u and v, with the assumption that both distributions have a continuous probability density function (pdf) f and g, respectively for reference and target images. The goal then is to find a continuous mapping $u \rightarrow t(u)$, such that the new colour distribution $t(u)$ matches the target distribution g [Pitié and Kokaram, 2007]. Figure A.2 illustrates this problem, also known as the mass preserving transport problem, to which the goal of the MK-transform is to find the minimal displacement mapping.

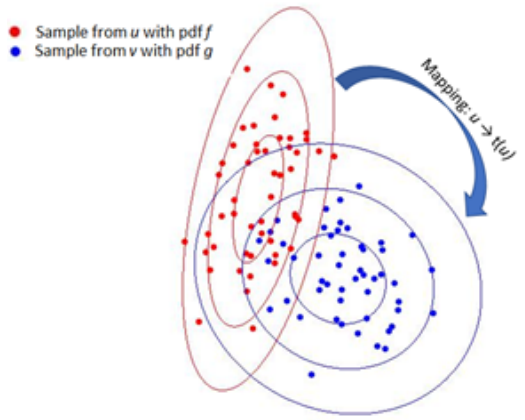

Figure A.2: An example illustrating the mapping of multivariate Gaussian distributions for colour distributions u and v.

A.2 Linear Monge-Kantorovich transform for colour transfer

In Pitié and Kokaram [2007], the goal is to transfer the statistical moments of two images represented as pdfs such that the displacement caused by a continuous mapping function is minimal. In the general case this is known as Monge's optimal transportation problem [Evans, 1997].

Consider two images to be registered, $X \in \mathbb{R}^{H_1 \times W_1 \times 3}$ and $Y \in \mathbb{R}^{H_2 \times W_2 \times 3}$, respectively as a reference and a target image, where $(H_1 \times W_1)$ and $(H_2 \times W_2)$ are the height and width respectively for the reference and target images. Before computing the linear MK-transform the brightness of X is matched with Y by converting both images to CIELAB colour space. Then, the histograms in corresponding lightness channels are matched before converting both images back to the RGB colour space. Each image band is flattened and concatenated column-wise so that each row represents an R, G and B colour sample.

The covariance matrices Σ_X and Σ_Y are computed, respectively from X and Y. Equation A.1 details the linear MK transform. In colour grading, the MK solution is desirable for two reasons: firstly, the solution always exists for continuous pdfs and is unique, meaning that there is no room left for ambiguity; secondly, the solution uses the gradient of a convex function that is the equivalent of monotonicity for one dimensional functions in $\mathbb{R}$. This implies that the brightest areas of a picture remain the brightest areas after mapping.

$$T = \Sigma_X^{-\frac{1}{2}}(\Sigma_X^{\frac{1}{2}}\Sigma_Y\Sigma_X^{\frac{1}{2}})\Sigma_X^{-\frac{1}{2}} \tag{A.1}$$

A.3 Experiments and results

Given the MK-transformed image, two pairs of images are used to compute geometrical linear transforms, as per [Brown et al., 2003]. The first pair is the high-resolution original

reference image with the low-resolution multispectral image (Figure A.1, (A) and (B)), and the second pair is the high resolution MK-transformed with the low-resolution multispectral image (Figure A.1, (C) and (B)). Each image-pair will result in a linear geometrical transform that warp pixel values from the target image to the same image coordinate system as the reference image.

The method was evaluated using cropped imagery from the generated very high resolution orthomosaics of Budle Bay. As mentioned in Section 3.2, each orthomosaic was orthorectified with ground markers that were spread out across the site. For this work, ground control markers will be used to mark control points in images in order to evaluate the pixel location accuracy in registration. The evaluation uses fourteen control points and for each control point an image is sampled to a 512×512 and 193×193 crop, respectively for images captured with the SONY and RedEdge3. Figure A.3 shows a gallery of images to be registered with control points used for evaluation.

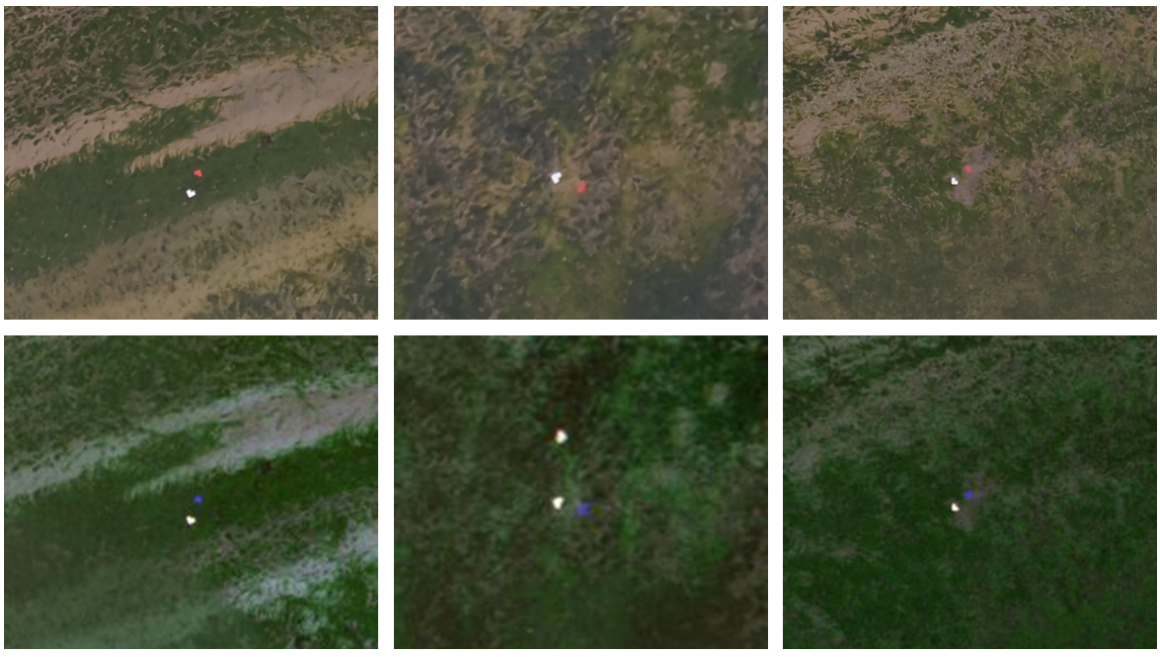

Figure A.3: Gallery of images to be registered. Top-row are high-resolution reference images from the SONY camera and the bottom-row are multispectral target images from the RedEdge3 camera.

	# SIFT matches	% inlier SIFT matches	Euclidean distance
MK-transformed/Target	630±798	77.15±17.4	1.2±0.86
Original/Target	725±940	78.85±14.7	1.86±1.27

Table A.1: Mean (± standard deviation) of the number of SIFT matches, percentage of inlier SIFT matches and Euclidean distance of pixel locations between control points in pairs of registered images

For each registered image, the pixel locations of each edge of the control marker in a target image are recorded and compared with the pixel locations of each edge in the reference image. The list of errors resulting from each edge are averaged so that each image has a single error metric. Table A.1 reports the mean Euclidean distance between control points for all pairs of images after registration in the dataset, as well as the number of SIFT matches and percentage of inlier matches after RANSAC.

The results in Table A.1 show that using MK-transform high-resolution image to compute a geometrical linear transform improves on registration pixel accuracy. The mean Euclidean distance between control points in pairs of registered images is lower for images pre-processed using the MK colour transfer than for images using the original reference image. This tells us that distortions caused by errors in registration, e.g. blur, will be more noticeable for registered images where the reference image is not pre-processed using the MK-transform.

The mean and variance of feature matches in pairs of images in the dataset is greater between the original reference and target images than with reference image pre-processed with MK-transform. This may seem counter intuitive given the euclidean distance in pixel locations is lower for images pre-processed with MK-transform but the number of SIFT matches does not correlate to improved image registration. SIFT generates key-points in images, then correspondences are found in least-squares fashion by estimating linear affine transformations. However, some key-points do not in fact lead to correspondences [Brown et al., 2003], hence the use of RANSAC to eliminate outlier correspondences [Derpanis, 2010]. Therefore, the hypothesis here is that pre-processed images with the MK-transform result

in less but higher quality matches in pairs of images. This is supported by the percentage of inlier SIFT matches where the MK-transformed/Target pair has 77.15% while the Original/Target pair has 78.85%, but the number of matches is much greater for Original/Target image pairs.

Figure A.4 shows the results of registration for pairs of images, where each registered image is converted to grey scale. The left image is the registration result where the linear MK transform is used to map the colours from the target image to the reference image and the right image is the registration result using the original reference image. Figure A.4 subtly confirms the results in Table A.1 - the left image is sharper around the edges of vegetation and soil as opposed to the right image.

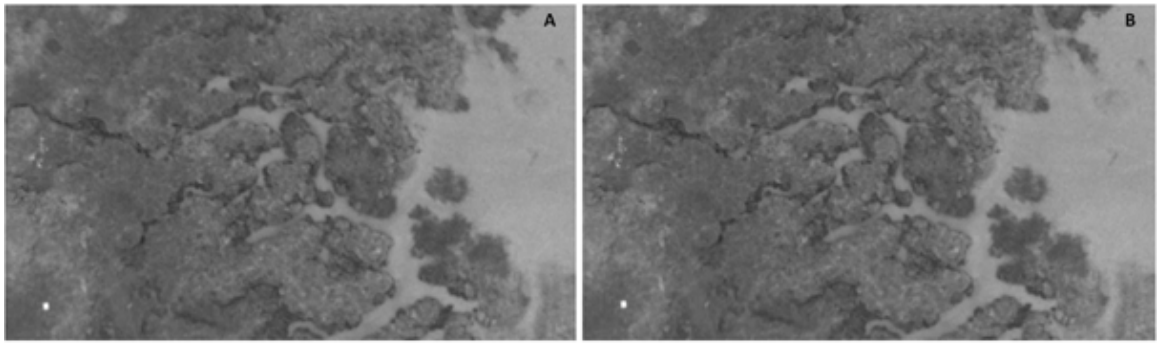

Figure A.4: Registered images from both cameras after conversion to grey scale. A - pre-processed with the MK colour transfer and B – registered using the original reference.

A.4 Summary

Image registration is a key-processing step in remote sensing applications that can be performed in various manners, with more complex methods existing in literature [Zitova and Flusser, 2003]. However, the method used and described in [Brown et al., 2003], where corresponding SIFT features in images are used to compute a linear transform is a common approach for automatic registration. This section shows that using a simple colour

transfer [Pitié and Kokaram, 2007] pre-registration reduces subsequent image registration errors.