

# Domain-inspired image processing and computer vision to support deep-sea benthic ecology

by

**Chloe Amanda Game**

A thesis submitted for the degree of  
Doctor of Philosophy

School of Computing Sciences  
University of East Anglia

December 2022

© 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 therefrom must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

## Abstract

Optical imagery is a necessary methodological tool for ecological research within marine environments, particularly in deeper waters. For benthic (seafloor) surveys, interpretation of image data is crucial to creating high-resolution maps of seabed habitats. This is fundamental to marine spatial planning and mitigating long-term damage of anthropogenic stressors such as growing resource demand, climate change and pollution. However there are numerous, and significant, issues in extracting a reliable ground-truth from imagery to support this process.

Analysis of benthic images is difficult, due in part to the extreme variation and inconsistency in image quality - caused by complex interactions between light and water. It is also time-consuming. This thesis is dedicated to providing solutions to manage these challenges, from a strong perspective of the end-user. Specifically, we aim to improve the annotation of benthic habitats from imagery in terms of quality, consistency and efficiency. Throughout, we consider the purpose the imagery serves and work closely with end-users to best optimize our solutions.

First, and for the majority of this thesis, we investigate image processing techniques to improve the appearance of image features important for habitat classification. We find that tone mapping is an effective and simple (and thus accessible) method through which to improve image quality for interpretation. We describe beneficial (expert-informed) properties for brightness distributions in underwater images and introduce a novel tone-mapping algorithm, Weibull Tone Mapping (WTM), to enhance benthic images. WTM theory operates within general constraints that model image requirements (properties) specified by image analysts, yet possesses a suitable degree of flexibility and customisation. As a tool, WTM provides analysts with a fast and ‘user-friendly’ method to improve benthic habitat classification.

---

Second, we consider computer vision methods that could automatically identify benthic habitats in imagery, relieving the analysis bottleneck. We find that baseline transfer learning of machine learning models, with limited optimization, will better facilitate adoption by novice users, yet still provides a powerful means to swiftly extract and assess benthic data.

## **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.

## Acknowledgements

I would like to thank my supervisor Graham Finlayson for giving me the chance to explore a new discipline. It's been both challenging and fascinating to investigate the nexus of Marine Ecology and Computer Science during this PhD and something i will be eager to continue in the future.

Thank you to my supervisor Michael Thompson for his support and reminder of the field 'from whence i came', helping me to materialize the contributions in these thesis.

I would also like to thank my previous supervisors Kerry Howell and Emma Sheehan for the opportunities and support that got me here, as well as other members (and friends) of MBERC at Plymouth University that motivated and inspired me.

I am grateful to the Gardline Environmental Reporting Team, past and present, for their involvement in the experiments in this work, words of encouragement and cake.

Thank you to the 'bros' (and pigs) for keeping me laughing throughout. I truly valued your company on this journey. Likewise, to my other Norwich friends, its been a blast. Thank you for the necessary distractions. To the friends from afar - your support never went unnoticed!

My heartfelt thanks to Nils Piechaud. Your patience, ear, humour and sass have helped me no end.

To my parents, i cannot thank you enough for encouraging me from a young age to dream big and work hard to achieve my goals. I would not be here without you. To my peas, your constant backing, belief and banter has meant the world.

*“On ne découvre pas de terre nouvelle sans consentir à perdre de vue, d’abord et  
longtemps, tout rivage.”*

---

*“One doesn’t discover new lands without first, and for a long time, accepting to completely  
lose sight of the shore”*

–André Gide, *Les faux-monnayeurs* (1925)

# Contents

<b>Abstract</b>	<b>i</b>
<b>Acknowledgements</b>	<b>iii</b>
<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xiv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Problem . . . . .	1
1.2 Approach . . . . .	2
1.3 Thesis Outline . . . . .	5
<b>2 Background</b>	<b>7</b>
2.1 Underwater Imagery . . . . .	7
2.1.1 Supporting deep-sea benthic ecology . . . . .	7
2.1.2 Ecological data extraction and processing . . . . .	10
2.1.3 Challenges for data extraction and processing . . . . .	12
2.2 Image processing for enhancing image quality . . . . .	14
2.2.1 Digital images and their histograms . . . . .	15
2.2.2 Tone Mapping for image enhancement . . . . .	18
2.3 Computer vision for automatic classification . . . . .	26
2.3.1 Image features . . . . .	26
2.3.2 Classification approaches: simplicity vs. complexity . . . . .	27
2.4 Summary . . . . .	36

<b>3</b>	<b>Image Datasets</b>	<b>37</b>
<b>4</b>	<b>Bespoke tone-mapping of underwater imagery</b>	<b>42</b>
4.1	Introduction . . . . .	42
4.2	Background . . . . .	46
4.2.1	Desirable brightness distributions . . . . .	46
4.2.2	Control-point tonal manipulations . . . . .	49
4.3	Experiments . . . . .	51
4.3.1	Domain-expert tone mapping . . . . .	51
4.3.2	Modelling image histograms . . . . .	57
4.4	Conclusion . . . . .	63
<b>5</b>	<b>Weibull Tone Mapping (WTM): for enhancing underwater imagery</b>	<b>64</b>
5.1	Introduction . . . . .	64
5.2	Background . . . . .	65
5.2.1	Evaluating image similarity . . . . .	66
5.2.2	Psychophysical Evaluation . . . . .	68
5.3	Weibull Tone Mapping (WTM) . . . . .	69
5.3.1	Creating a proxy brightness distribution using the Weibull distribution . . . . .	69
5.3.2	Calculating the Weibull Tone Map . . . . .	73
5.3.3	Applying the Weibull Tone Map . . . . .	74
5.4	Experiments . . . . .	76
5.4.1	Approximating bespoke tone maps . . . . .	76
5.4.2	Preference experiments . . . . .	85
5.5	Conclusion . . . . .	91
<b>6</b>	<b>Weibull Tone Mapping (WTM): as an interactive enhancement tool</b>	<b>93</b>
6.1	Introduction . . . . .	93

6.2	Background . . . . .	95
6.3	WTM as a parameterized enhancement tool . . . . .	96
6.4	Experiments . . . . .	97
6.4.1	WTM suitability for underwater imagery . . . . .	97
6.4.2	Simplification of control-point tonal manipulations . . . . .	102
6.4.3	Behaviour of WTM enhancements . . . . .	106
6.4.4	Psychophysical Evaluation . . . . .	110
6.5	Conclusion . . . . .	114
<b>7</b>	<b>Simplifying automatic classification of benthic habitats</b>	<b>116</b>
7.1	Introduction . . . . .	116
7.2	Background . . . . .	118
7.2.1	Machine Learning Approaches . . . . .	120
7.2.2	Performance Evaluation . . . . .	123
7.3	Experiments . . . . .	125
7.3.1	Extracting and interpreting a deep feature space . . . . .	125
7.3.2	Classifying deep benthic features . . . . .	128
7.3.3	Time Considerations . . . . .	137
7.4	Simplicity vs. Complexity . . . . .	139
7.4.1	Improving performance . . . . .	141
7.5	Conclusion . . . . .	143
<b>8</b>	<b>Conclusion</b>	<b>145</b>
8.1	Contributions . . . . .	145
8.2	Future Work . . . . .	147
8.2.1	Image Processing . . . . .	147
8.2.2	Computer Vision . . . . .	148
8.3	Final thoughts . . . . .	150
8.4	Publications . . . . .	150

<b>Abbreviations</b>	<b>152</b>
<b>9 Bibliography</b>	<b>153</b>
<b>A Gardline Seabed Classification Guide</b>	<b>184</b>

# List of Figures

2.1	Selection of image platforms: (A) AUV (133), (B) ROV, (C) Towed array (166), (D) BRUVS (168), (E) Drop-frame mounted camera (94) and (F) Diver-operated stereo-video camera (62) . . . . .	8
2.2	CountEM software annotations: polygon, point and marked extent. Green lasers for scaling are also present (38) . . . . .	11
2.3	Image histogram of RGB colour channels & brightness channel . . . . .	17
2.4	An example of tone mapping: (A) an input image, (B) an input & enhanced output brightness distribution, (C) the tone map and (D) an output image . . . . .	18
2.5	A diagram of tone mapping . . . . .	19
2.6	Brightness histograms of images in Figure 2.7(A), including an unenhanced input and a HE & CLHE enhanced output . . . . .	21
2.7	Enhancement example: (A) Unenhanced image, (B) HE image and (C) CLHE image with a minimum slope of $0.5/N$ & a maximum slope of $2/N$ , where $N = 256$ bins . . . . .	22
2.8	HE & CLHE tone mapping: (A) An Input PDF (HE proxy histogram) and its slope-limited (CLHE) proxy, (B) their respective CDF's or tone maps. Grey lines in (A) depict upper and lower slope bounds of $2/N$ and $0.5/N$ respectively, where $N = 256$ bins. Grey shading in (B) highlights area of each tone map that fall within the slope thresholds. . . . .	23
2.9	Diagram of a simple CNN for image classification, with 3 output classes. . . . .	29
2.10	A diagram of Support Vector Machines . . . . .	34
2.11	Multi-class Support Vector Machines: a diagram of the One vs. One strategy . . . . .	35
3.1	Example of habitat classes encountered across image dataset . . . . .	40

3.2	Proportion of each habitat class across datasets . . . . .	41
4.1	Example of brightness tone mapping. In (A) we show an input brightness histogram and three modified outputs. The corresponding tone maps are shown in (B). Adjustment 1 demonstrates a brightness increase, 2 a contrast increase and 3 a combined brightness and contrast increase . . . . .	43
4.2	Visual representation of tonal manipulations in Figure 4.1 From left to right we show: the unenhanced input image, a brighter output, a more contrasted output and a brighter output with more contrast. . . . .	43
4.3	An selection of unenhanced benthic images . . . . .	44
4.4	Corresponding brightness distributions of unenhanced images shown in Figure 4.3 . . . . .	45
4.5	Variations of the Rayleigh & Weibull distribution . . . . .	48
4.6	A diagram of tone map creation and interactive manipulation . . . . .	49
4.7	GIMP tone-mapping tool: original ‘identity mapping’ (left) and adjusted tone map (right) . . . . .	50
4.8	Image selection to represent optical challenges present in the experiment dataset	54
4.9	Brightness histograms of images shown in Figure 4.8 . . . . .	55
4.10	Selection of input images (1 per habitat class), bespoke tonal manipulations and their visual effect on output images . . . . .	56
4.11	Mean KL-Divergence between the dataset brightness distributions and their best Rayleigh & Weibull approximation. In (A), results are displayed according to enhancement type. Note that Output = Outputs 1, 2 & 3. In (B) mean KL-divergence (across output images) is viewed according to habitat class. Error bars represent 95% confidence intervals. . . . .	60
4.12	Some example brightness distributions and their most similar Rayleigh and Weibull Distribution for (A) input images and (B) analyst-adjusted output images. KL-Divergence scores, denoted here as $KLd$ , are presented in each case. . . . .	62

5.1	Weibull Tone Mapping flow chart. *Step 3 can either approximate an existing tonal enhancement (using Equation 5.5) or specify an original WTM enhancement, see Chapter 6 . . . . .	70
5.2	Parameter control over WD properties: <b>(A)</b> $\lambda$ modification and <b>(B)</b> $k$ modification . . . . .	71
5.3	WTM method showing (A) WTM proxies of an Input and Output (analyst-adjusted) brightness PDF and (B) the corresponding tone maps that transform PDFs from Input to Output. Dashed and smooth lines depict User and WTM respectively . . . . .	72
5.4	Comparison of an unenhanced input image (A), an analyst enhanced image (B), Chromatic WTM (C) and Chromaticity-preserving WTM (D) . . . . .	74
5.5	An example of two Weibull tone maps which differ from the identity map to varying degrees . . . . .	75
5.6	Image comparison of colour mapping methods Chromatic and Chromaticity-preserving with two WTM tone maps (A & B) presented in Figure 5.5 . . . . .	76
5.7	Average intersection, or <i>match</i> , of analyst tone maps and their Weibull approximation. A 45 degree line (slope= $1/N$ ) indicates a perfect match. Standard deviation (SD) and 95% confidence intervals (CI) are also presented. . . . .	79
5.8	Similarity metrics between the analyst tone maps and their best Weibull approximation. Results are displayed according to enhancement type. Error bars represent 95% confidence intervals. . . . .	80
5.9	Similarity metrics between the analyst tone maps and their best Weibull approximation. Results are displayed according to habitat class. Error bars represent 95% confidence intervals. . . . .	81

5.10	Example of analyst tone maps and their best Weibull approximation. In (A) Weibull tone maps resulted in good approximation of analyst images, with Chromatic WTM, scoring a mean $\Delta E_{ab}^* < 5$ compared to (B) which scored a mean $\Delta E_{ab}^* > 5$ , approximating adjustments less well. Tone maps in each subplot are numbered and their mean $\Delta E_{ab}^*$ listed. . . . .	82
5.11	Example images corresponding to tone maps presented in Figure 5.10 (A). .	83
5.12	Example images corresponding to tone maps presented in Figure 5.10 (B). .	84
5.13	Mean $\Delta E_{ab}^*$ between WTM & analyst-enhanced images across (A) analysts and (B) habitat classes. . . . .	85
5.14	Preference Experiment Scores for Chromatic and Chromaticity-preserving WTM, denoted by subscripts $C$ & $CP$ , respectively. Error bars represent 95% confidence intervals. . . . .	88
5.15	Preference Experiment Scores for Chromaticity-preserving WTM with the addition of CLHE pairwise comparisons. Error bars represent 95% confidence intervals. . . . .	90
6.1	WTM example: (A) Input image, (B) Analyst-adjusted Output, (C) WTM approximation of Output image (B) and (D-F) parameterized WTM adjustments of (A) . . . . .	96
6.2	Experiment GUI . . . . .	98
6.3	Example of WTM approximation of control-point control-point adjustments. Colour difference between the custom and WTM image is summarized by mean $\Delta E_{ab}^*$ across pixel pairs. . . . .	104
6.4	Example of observer tone-maps using WTM and the custom (control-point) enhancement. Custom tone-maps here can not be approximated by WTM i.e. they did not meet mean $\Delta E_{ab}^* (< 5)$ threshold for approximation. . . .	105

6.5	WTM parameter variation with image class (habitat): (A) mean difference in WTM parameters between input and output WD and (B) mean WD parameters across input images. All values are presented as standardized Z-Scores for fair comparison. $\lambda$ and $k$ differences in (A) refer only to cases in which they were decreased and increased, respectively. Error bars represent 95% confidence intervals. . . . .	107
6.6	Mean standardized (Z-Score) difference in WTM parameters between all input and output WDs. Error bars represent 95% confidence intervals. . . .	108
6.7	Preference Experiment Scores. Error bars represent 95% confidence intervals.	112
7.1	Graphical representation of ML approaches ( <b>CNN</b> & <b>CNN+SVM</b> ) used in this chapter . . . . .	119
7.2	VGG16 architecture . . . . .	121
7.3	2-Dimensional visualisation of the 3-Dimensional feature spaces for each dataset, created using PaCMAP . . . . .	127
7.4	PaCMAP 2-Dimensional visualisation of Dataset 1 (Table 3.1) feature space: a) 2D point cloud and b) zoomed region in which points are replaced by image thumbnails for better conceptualization of feature variance. . . . .	128
7.5	Mean performance across Datasets for each habitat. Error bars demonstrate 95% confidence intervals . . . . .	135
7.6	Overall, class-averaged and individual class performance across test datasets - in order of decreasing training set size . . . . .	136
7.7	Time required for training and testing phase, in order of decreasing dataset size. Duration is presented in $\log(\text{minutes})$ for phases run both with a GPU and without (CPU) . . . . .	138

# List of Tables

2.1	Hyperparameter glossary for CNN & SVM training . . . . .	32
3.1	Benthic Image Datasets used in this study . . . . .	37
3.2	Habitat classes encountered within the image dataset . . . . .	38
3.3	Broad habitat classes and their association with EUNIS habitats . . . . .	39
4.1	Summary statistics for KL-Divergence between the dataset brightness distributions (input and outputs) and their best Rayleigh & Weibull approximation	59
4.2	Summary statistics for KL-Divergence, according to enhancement type, between the dataset brightness distributions and their best Rayleigh & Weibull approximation . . . . .	61
4.3	Summary statistics for KL-Divergence, according to habitat class, between the dataset brightness distributions and their best Rayleigh & Weibull approximation. . . . .	61
5.1	Summary statistics for three metrics measuring similarity between each of the analyst tone maps and their closest Weibull tone map. . . . .	78
5.2	Summary statistics for similarity metrics between analyst images and their respective Weibull approximations. Chromatic and chromaticity-preserving Weibull tone map application are denoted by subscripts $C$ & $CP$ , respectively. . . . .	80
5.3	Mean image quality metrics for input images and enhanced outputs . . . . .	90

6.1	Mean image quality metrics for input images and enhanced outputs. Mean output metrics are also displayed according to enhancement type: WTM and custom control-point adjustment. . . . .	100
6.2	Intra-observer variability across image pairs. . . . .	101
6.3	Inter-observer variability across image pairs. . . . .	102
6.4	Summary statistics for mean $\Delta E_{ab}^*$ (across pixel pairs) between control-point tone-mapped images and their WTM approximation . . . . .	104
6.5	Mean image quality metrics for custom tone-mapped images and their successful WTM approximations (mean $\Delta E_{ab}^* < 5$ ) . . . . .	105
6.6	Mean image quality metrics for input images and enhanced outputs . . . . .	114
7.1	Required steps in our ML workflows . . . . .	130
7.2	Classifier training performance: with tuned hyperparameters, mean cross-validation (CV) and final training accuracy . . . . .	133
7.3	Classifier training performance: with overall test accuracy and class-averaged metrics (Recall, Precision and F <sub>1</sub> Score) . . . . .	134

# 1 Introduction

## 1.1 Problem

Around 70% of the world's surface is covered by ocean. The majority of which is considered 'Deep Sea'; the largest ecosystem on earth, a region extending from 2km to around 11km below sea level (54; 184). Oceans both provided the necessary foundation for the evolution of life and ensures its continuity. As a natural resource, its value with respect to ecological, economic and social importance is immense (29; 10; 180); occupying a crucial role in climate regulation, supporting primary and secondary food production and biochemical and pharmaceutical developments, to name a few. The extensive natural capital associated with these assets strongly motivates its study and protection.

Anthropogenic impacts on the marine environment are both ubiquitous, cumulative and increasing (70). They stem from growing resource demand within the fisheries, transport, energy, mineral and recreation sectors and have thus driven the development of marine spatial planning and conservation goals to mitigate long-term damage (47; 186; 152; 185; 82). A fundamental contribution to the success of such endeavours, is the creation of extensive and accurate maps of benthic (seafloor) habitats (27; 73; 13). These establish baselines and support monitoring of impacts and recovery.

Benthic habitats may consist of multiple components: substrate, species and/or communities, their environmental tolerances and preferences (34; 39). They often act as simplified but powerful proxies of biodiversity, by allowing inference of occurring organisms through known ecological associations. Their physical aspects can be well described using acoustic data, collected by instruments such as a Multibeam Echosounder (MBES) and Side-Scan

Sonar (SSS) (22). This can be efficiently employed over large geographic and bathymetric ranges and at high resolutions (tens of m). However, minus the reef-forming species or large shellfish aggregations, this methodology is unable to measure biological components of benthic habitats (22). The density and size of many benthic taxa (taxonomic groups of organisms), in comparison to their surrounding environment, often results in a neutral acoustic signature. Thus, dependent on the survey aims, these features must be processed via alternative means such as grab sampling, trawling or photography. In this thesis, we focus on the latter methodology, a preferred tool for environmental surveys within marine environments, particularly in deeper waters.

Environmental monitoring of benthic habitats from imagery requires annotation by biologists to identify and quantify seafloor objects. Broadly, there are two main problems that hinder this process: (1) that underwater image quality is often poor, primarily due to the physical behaviour of light in water and (2) that annotation is incredibly labour-intensive. In this thesis, we work closely with a multidisciplinary marine survey company, Gardline Ltd., to manage these challenges and improve their analysis pipeline of benthic images. We provide solutions with respect to image processing (enhancement) and computer vision (automatic classification). Each of these are highly-domain inspired, aiming to have practical, real-world outputs with a high capacity for implementation. Importantly, the solutions proposed are focused on improving the user-experience for benthic analysts and enhancing data quality. We note that these contributions are applicable beyond our industry partner, optimizing benthic image analysis more widely.

## 1.2 Approach

Considering the challenges to benthic analysis previously introduced, we engaged in discussions with Gardline to identify specific hindrances in the image analysis pipeline that could be improved.

The illumination challenges of photographing in an aquatic medium present difficulty for image analysts in visualising and identifying features of interest from imagery. Common visual effects in underwater imagery include reduced colour, a bright illumination cone or halo surrounded by dark pixels (vignetting) and low contrast. It is therefore necessary to process images to improve visibility of image features and textures, and standardize their appearance. At the onset of this thesis, Gardline had some image enhancement functionality within their ‘in-house’ annotation software. These included a gamma correction and colour filtering tool for (scaling) lasers (38). However, most popular of these was a tone mapping tool, in which a function is used to transform image pixel values to a more desirable output. Specifically, they use the ‘venerable’ Histogram Equalization (HE) (88; 64). This is used to tonally adjust the contrast of images, an important enhancement to improve conspicuity of image features through improved edge details and textural information. Explanation on tone mapping and the HE algorithm is provided in Chapter 2. The HE tool is a fast and simple enhancement option, requiring just the click of a button. However, the general adjustments it enforces are not tailored or suited to the task at hand. Images often appear over-enhanced, with dark and bright intensities strengthened. This can obscure image features and is thus counter-intuitive; causing the very opposite of the desired outcome. Therefore, a first aim of this thesis was to ascertain if we could provide improvements in this area without much compromising on efficiency or ease of use for image analysts.

Our approach was to further investigate tone-mapping functions and determine what properties they should have to support analysts image interpretation. We achieve this through bespoke tonal adjustment experiments and psychophysical assessments with end-users. This found automatic tone mapping operations to be insufficient for underwater image analysis and conventional methods inconsistent in their benefit. Instead, analysts indicated the importance of custom tonal enhancements. However, these are time-consuming to generate on a per-image-basis and thus unsuitable in analytical scenarios, which are already laborious. This led us to develop a tool that optimizes the generation of custom tonal manipulations for an end-user in underwater imagery.

We call this Weibull Tone Mapping (WTM). This name follows from an analysis of how benthic image analysts make tonal adjustments in relation to the brightness distributions in input (unenhanced) and output (enhanced) images. In WTM, both the input and output distributions are represented by a Weibull distribution. Where the WTM tone curve is the brightness mapping taking the former to the latter.

On completion of this thesis, WTM is not *fully* automated, however the tool allows analysts to tonally adjust imagery by modifying two parameters in a Graphical User Interface (GUI). Importantly, analysts can view ‘live’ adjustments as they modify the parameter ‘sliders’ and simultaneously annotate the image. This provides a significant improvement to their image analysis pipeline, offering a range of enhancements that can aid various aspects of the annotation task. We offer them the ability to construct a bespoke and extremely tailored tonal enhancement but much faster and simpler. WTM allows analysts to improve the quality and efficiency of benthic image annotations. This improved data extraction phase supports the next priority in improving their analysis workflow: automating annotation tasks.

There are increasingly more applications of machine learning (ML) to benthic image annotation, many of which are focused on single taxa (1; 40; 95; 136). An important, yet under-represented, application in the literature is on classifying benthic habitats, particularly in the Deep Sea. Additionally, Gardline have a high requirement for annotation of the image habitat, for habitat mapping, and thus felt their image analysis could highly benefit from the automation of habitat classification. Although highly motivated to integrate automatic capabilities into their analysis pipeline, these approaches are inaccessible due to a technical skills-gap and limited time to focus on this task. This is mirrored amongst the majority of benthic ecologists/biologists external to Gardline. With this in mind, we investigate machine learning approaches with an appropriate trade-off in complexity, implementation ease and performance. This will better facilitate uptake of these methods within Gardline and more widely. We take a hierarchical approach, aiming to first classify

broad-level habitats in images. These images can then be passed to an existing classifier for detecting individual taxa or further features of interest in appropriate habitats. Equally, it could be used to efficiently prepare suitable training data for the same task. Classifiers that detect individual taxa, such as sea pens, could also provide a higher-resolution habitat classification through detection of sub-habitats, see Table 3.2.

We find deep features from a pre-trained ('off-the-shelf') Convolutional Neural Network (CNN) are suitable to train a Support Vector Machine (SVM) to classify broad benthic habitats. This requires little theoretical understanding and is more straightforward programmatically. It also provides a 'primer' for biologists who wish to use deep learning methods, which are popular and sought after within the community due to their reputation for good performance.

### 1.3 Thesis Outline

Chapter 2 reviews the background to the topics covered in this thesis. We discuss in more detail the significant challenges associated with using underwater imagery to support deep-sea benthic ecology. We then focus on how we can improve the interpretation of imagery and therefore the extraction of ecological data. Potential solutions are discussed encompassing (1) enhancing (visual) quality of imagery and (2) automatic analysis i.e. classification.

Chapter 3 introduces the benthic image datasets used in this work. We provide details on image acquisition, such as the camera platforms used and geographic location. We also discuss the procedure for data extraction, including the classification schemes utilised by Gardline to identify benthic habitats, and some associated challenges.

In Chapter 4 our main focus is on tone mapping as a tool to improve image quality. We propose desirable properties a tone mapping operation (TMO) should have, moving beyond merely the aesthetic drivers to consider the purpose the imagery serves and its audience). Expert-designed TMOs, to better identify benthic habitats, are discussed. We show that they have specific properties that can be captured by the 2-parameter Weibull probability

distribution (WD) i.e. analysts enhance imagery according to this distribution. This provides a foundation upon which to optimize current tone mapping approaches, from generic tools into enhancements specialized for underwater image analysis.

Chapters 5 and 6 introduce our novel tone-mapping algorithm, Weibull Tone Mapping (WTM). In Chapter 5, we explore the intuitive parameters of the WD, showing them to correspond to common and effective tonal manipulations, namely brightness and contrast. We present the WTM theory in which we derive simple and smooth approximations of bespoke tonal adjustments created by analysts using the WD. Importantly, analysts find this suitable to enhance visual interpretation of marine habitats. In Chapter 6 we develop the WTM theory into an interactive enhancement tool that provides analysts with a quicker and simpler means of deriving bespoke TMOs. Notably, analysts prefer enhancing image quality with our WTM tool over automatic tonal manipulations. We also show, in these chapters, that WTM is competitive in terms of preference (and its consistency) with another traditional TMO that requires parameter selection, Contrast Limited Histogram Equalization (CLHE). Yet WTM provides end-users with more customisation without compromising ease-of-use and is modelled on the brightness statistics of underwater imagery sought by analysts.

In Chapter 7 we consider the challenges in quality, consistency and efficiency of annotating benthic habitats in imagery. We present methods that offer a suitable trade-off between simplicity of operation and comprehension for non-specialist users and good, consistent performance. We found that classifying 'off-the-shelf' Convolutional Neural Network features with a shallow learning algorithm, a Support Vector Machine is a powerful method by which benthic data can be quickly extracted and evaluated.

Finally, in Chapter 8 we present our conclusions, thesis contributions and discuss areas for future of work.

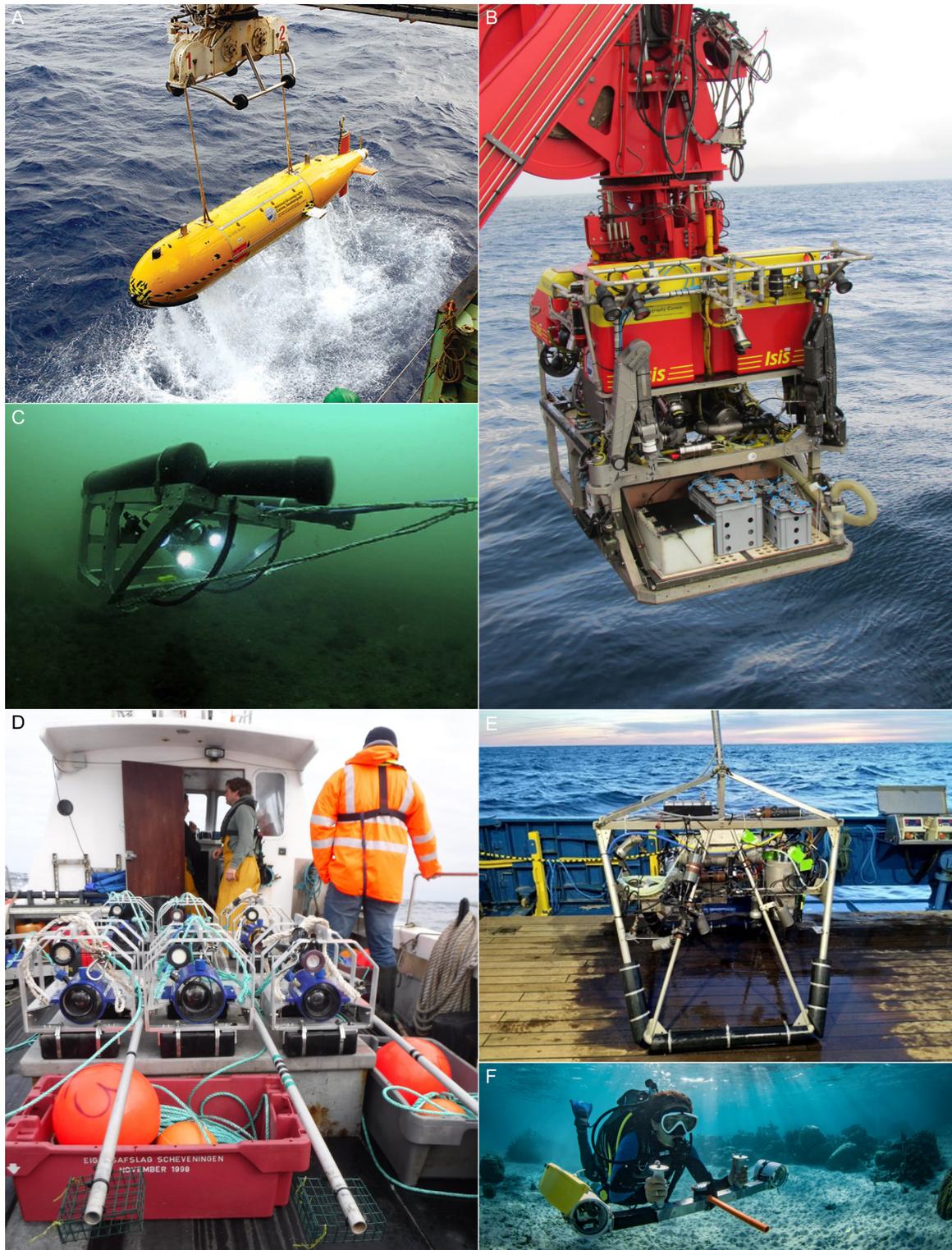
# 2 Background

In this chapter, we provide background for the thesis. In Section 2.1 we provide an introduction to underwater imagery, more specifically its acquisition and role in supporting deep-sea benthic ecology; introducing some of the challenges in this endeavour. We focus on two main challenges: (1) extracting a reliable ground-truth from imagery and (2) coping with the analysis bottleneck, framing each in the perspective of the end-user. In Section 2.2 we discuss the technical components of imagery and how we might process (enhance) them to address the first challenge whereas in Section 2.3 we focus on the latter, considering how computer vision, specifically automatic classification, can optimize analysis. We finish the chapter with a brief summary that will lead into the remainder of the thesis.

## 2.1 Underwater Imagery

### 2.1.1 Supporting deep-sea benthic ecology

Optical imaging has a long history within marine ecology and has been utilized via a number of platforms (123; 174), such as SCUBA divers, baited remote underwater video (BRUVS), remotely operated vehicles (ROVs), autonomous underwater vehicles (AUVs), ocean floor observation and towed video systems, see Figure 2.1. The advantage of imaging as a tool to survey the marine environment is largely attributed to its non-destructive, rapid and repeatable nature over broad temporal and spatial scales. In addition, marine use of optical imaging has been supported by the increasing availability of high quality cameras that are affordable, compact and efficient (99).



**Figure 2.1:** Selection of image platforms: (A) AUV (133), (B) ROV, (C) Towed array (166), (D) BRUVS (168), (E) Drop-frame mounted camera (94) and (F) Diver-operated stereo-video camera (62)

Imagery can provide both qualitative and quantitative data, to support a range of ecological research, describing biodiversity and community composition (155; 20; 44). This research can comprise evaluation of anthropogenic and/or naturally-driven environmental disturbance (167) and temporal and spatial variation (163; 21; 178), as well as supporting habitat mapping and the designation of Marine Protected Areas (80; 149). Unlike other traditional sampling techniques, such as benthic grabs, trawling and diver surveys, the acquisition of imagery is not spatially restrictive and well suited to a multitude of habitats and locations (20). Given the variety and success of its application, imagery is a fundamental and methodological standard to countless research efforts in this field.

A particularly significant application of underwater imagery has been employed in surveying the deep sea, the largest ecosystem on the planet. Of the permanently available living space for animals globally, >98% of the volume is provided by deep sea waters and 63% of the area by the deep seafloor (180). The sheer size, isolation and challenging environmental conditions of this ecosystem, such as extreme pressures, low temperature, low light and high salinity, has historically created numerous difficulties in data acquisition. However, since the pioneer of remotely operated deep-sea cameras in the 1940s (51; 50) the volume of photographic data has been growing, enhanced by the development and use of robotic platforms such as ROVs, but most notably by AUVs (116; 196; 199; 89).

Such data is imperative not only to support research, but conservative efforts of the deep-sea environment. Pressures to deep-sea ecosystems are increasing (61) with over-exploitation currently the greatest threat to this ecosystem (141). Increasing demand for offshore resources, such as fish and minerals, is challenging the persistence of deep sea fauna, due to altered biomass and community structure, decreased diversity and a reduction in habitat building species such as cold-water corals (26; 141). This is especially concerning given the importance of this system to the planet, in terms of matter exchange, energy, biodiversity and climate regulation (10). Maximising the amount of data available for ecological research will better support our understanding of these ecosystems and their response and/or

potential resilience to anthropogenic pressures. It will also help to inform marine policy and resource management in the drive for blue growth and sustainability and support efforts in restoration (82; 81).

### **2.1.2 Ecological data extraction and processing**

Extracting ecological data from underwater image datasets, to support research and conservation, requires time-consuming annotation by image analysts. Image annotation in this context refers to interpretation, or translation, of image information to a semantic level (63). Analysts must localise and quantify marine objects, in terms of abundance or size i.e. number of instances or percentage cover within an image. For example, annotations may be polygons, to estimate coverage of encrusting species and habitat such as rocky reef, or point locations of species of interest. Marked extent can also be highlighted by a box drawn around individual fauna. In Figure 2.2 we show an example of benthic image annotation software used by our collaborator Gardline. This demonstrates a range of annotation formats by image analysts. Notice also the presence of green lasers, which are often mounted on camera platforms and directed downwards to the seabed. These are a popular method by which analysts can determine object size (laser scaling) due to the known distance between the two lasers (38).

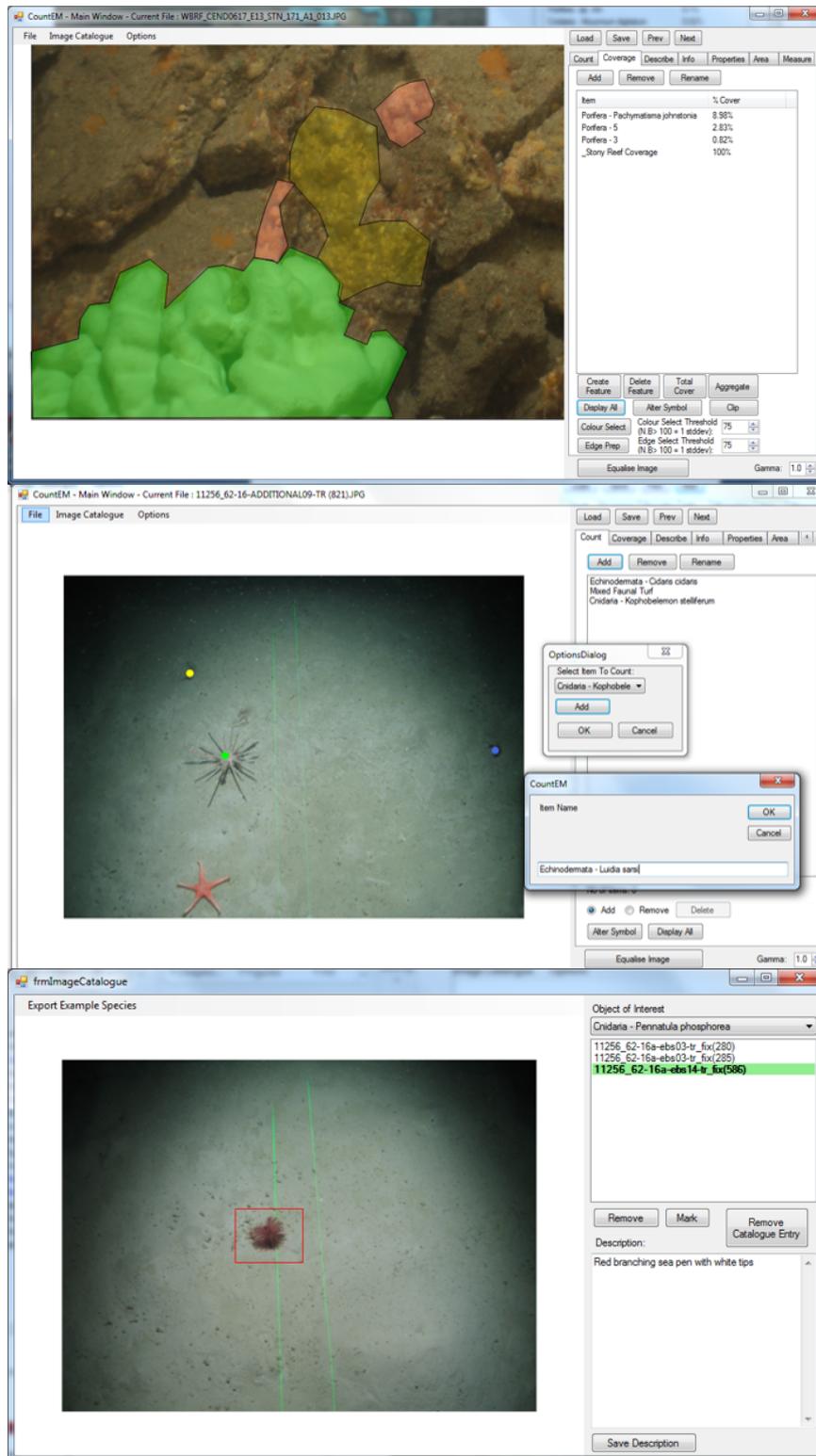


Figure 2.2: CountEM software annotations: polygon, point and marked extent. Green lasers for scaling are also present (38)

### 2.1.3 Challenges for data extraction and processing

Generating annotations in underwater optical images, as shown in Figure 2.2, is challenging for image analysts. One aspect of this is linked to the quality of images and their suitability for image annotation. Imaging in an aquatic medium, and near the seafloor, has a number of difficulties; many of which even the most prepared and sophisticated acquisition methodologies cannot avoid. For example seabed sediments may be disturbed by interaction of the imaging platform and the seabed, as well as animal behaviour, perhaps in response to the image platform. This can obscure the field-of-view and make image annotation extremely difficult, if not impossible. Image quality can not only control whether annotations are possible, but to what extent, for example the resolution of marine taxa identifications. Additionally, species can be cryptic to avoid predation. Small changes in visual acuity can therefore have pronounced implications on what is achievable in analysis.

A more prevalent and significant issue is linked to illumination. Complex interactions between light and water result in extremely varied and inconsistent image quality and can cause image features to look quite different both within an image and across image datasets. Underwater light penetration is one factor attributed to these issues and may be influenced by several variables including water clarity, turbidity, depth and distance (55). Due to the limited penetration of light within water, particularly within the deep sea, strong artificial lighting is employed on camera platforms. This can create non-uniform illumination within images as well as shadows, which when imaging from moving platforms such as underwater vehicles can shift position, resulting in reduced correspondence between images (77). This is further exaggerated by varying power requirements for different vehicles (i.e. ROVs versus AUVs), expending varied amounts of energy in lighting the scenery and altered lighting patterns (48). Light from vehicles such as ROVs has also been seen to attract large shoals of demersal fish, which can hinder the investigation of benthic communities, as a result the illumination level may be intentionally limited. Furthermore, light attenuates with depth, due to increasing wavelength absorption and scattering, caused both by the water medium

itself and suspended particles (15). Both these behaviours can hinder image quality due to colour reduction, low contrast and blurring effects. Further difficulties are encountered when imaging deep sea environments, due to extreme bathymetric pressures, requiring high-pressure camera housings and depth-rated lenses. Imperfections in this, can result in non-linear distortion of images and a risk of light refraction between water/glass and glass/air interfaces, which could result in image deformation (77). Datasets can also suffer from sporadic technical issues such as the failing and over-exposure of camera flashes.

Aside from the quality of underwater imagery, there are challenges associated with the annotation process itself that can hinder the extraction and processing of image data to support benthic ecology. Regardless of the image platform, annotation of resulting imagery is often inconsistent and error prone due to observer bias, fatigue, distraction and short-term memory limitations (31). It is also costly (in the absence of volunteers) and incredibly labour-intensive. This reality is particularly realised with AUV usage, in which one survey (~50 hours) can produce over 170,000 images (200). The time required to process such data (generating annotations) causes a significant bottleneck in deriving ecological information from imagery (17; 158).

The importance of, and high-demand for, benthic image annotations require these challenges to be addressed. Considering our industrial collaborator (Gardline), improving the efficiency and quality of data processing, using readily available methods, will result in improved regulatory decisions on impact assessments and monitoring. In the context of an analysis pipeline, pre-processing, or enhancing, underwater images could improve image quality and support better extraction of a reliable ground-truth by analysts. Coping with the analysis bottleneck requires computer vision solutions to automate elements of the analysis pipeline. However, since many of the optical issues that impede manual analysis may also present difficulties for computer vision research (128), it is first important to consult this part of the analysis pipeline. Thus, in the following section we discuss image processing for enhancing

quality of underwater imagery before moving on the role of computer vision in support imaging analysis.

## 2.2 Image processing for enhancing image quality

Image pre-processing can be an effective method to suppress unwanted lighting effects and enhance image features, supporting further processing such as feature extraction and image analysis(100; 175). Operations may be point, line, area, algorithmic or data conversion based and include a variety of techniques such as illumination, blur and focus corrections, filtering and noise removal, edge enhancements, colour space conversions, region processing and filters, and segmentation to name a few (100).

Pre-processing methods can be separated into broad categories; namely image restoration, correction and enhancement. Restoration methods aim to recover degraded or lost information from an image due to effects mentioned prior whereas enhancement aims to improve image appearance, typically aesthetically. Restoration algorithms include polarization and dark-channel prior methods which provide information on light scattering which can then be used to recover a more realistic scene. There are also deep learning approaches which focus on denoising, deblurring and resolution manipulation. Whereas enhancement algorithms include retinex, fusion and/or histogram-based methods. These can aid colour constancy, reduce blurring and noise as well as manipulate contrast and brightness to improve image clarity. As with restoration, there are (increasingly more) applications of deep learning in image enhancement. Comprehensive reviews of implemented algorithms for image restoration, correction and enhancement can be found in (154; 112; 85). Evident from these reviews is that existing restoration methods are rigorous and beneficial to handle uneven illumination, colour infidelity and distortion as well as haze or blurring. However, they are complex, often computationally demanding and time-consuming. Typically they require a large number of parameters, including knowledge of the environment such as depth, object/camera distance or water quality (15). Their performance may also be limited due to high de-

pendence on degradation model assumptions which may not be accurate, as well as highly variable parameters such as attenuation and diffusion coefficients which represent water turbidity. Although enhancement methods operate within a subjective framework and thus can produce inconsistent results, which may lack fidelity with the original imagery, they provide a flexible, faster and often simpler tool, requiring no *a priori* knowledge (154; 112; 85). This combined with their presence in photo-editing software, at least for histogram-based methods, increases their accessibility for users lacking technical expertise.

Given that the focus of this thesis is to aid biologists interpretation of underwater imagery whilst improving accessibility and the user experience, we thus focus on histogram-based enhancement methods, which manipulate image brightness/illumination characteristics. We explore methods that vary in complexity, both in user interaction and methodologically. Before introducing how histogram enhancements are achieved, it is first necessary to introduce and describe the digital image and its histogram.

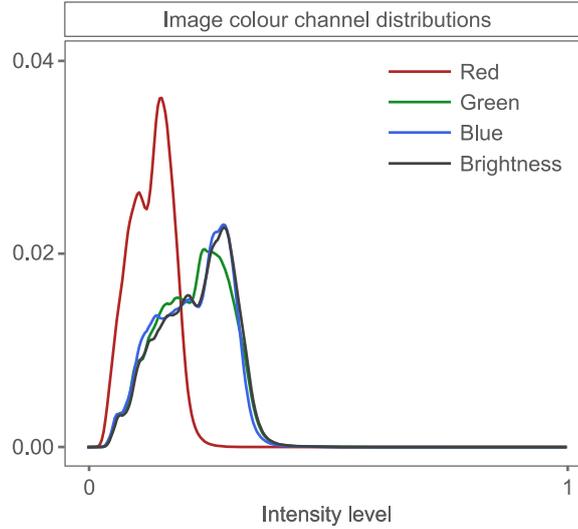
### 2.2.1 Digital images and their histograms

A digital image is a matrix, denoted as  $\underline{I}(x, y)$ , where  $x$  &  $y$  represent the spatial dimensions of the image scene and  $\underline{I}$  the spectral dimensions. The spectral dimensions can be represented in a number of colour spaces. For benthic images, they are typically encoded as RGB (device-dependant), or Red-Green-Blue, (64) and thus have 3 spectral channels i.e.  $\underline{I}(x, y)=[R,G,B]$ . These colour channels mimic the trichromatic visual system of the human eye which possesses three cones. These are sensitive to long, middle and short wavelengths of light, representing the blue, green and red portions of the electromagnetic spectrum, respectively (145). RGB pixel values in each of these channels range from  $[0, 255]$  for 8-bit images. The combination of these values across the channels is what encodes the colour at each pixel in an image. For example, a black pixel is created by values of 0 across all colour channels whereas a value of 255 across all channels will create a white pixel. Note that digital images are not limited to three colour channels, they can possess more, sometimes hundreds, in the case of images captured by hyperspectral cameras (113). Image size

$x \times y \times z$ ) varies with camera models and they are typically stored in JPEG, PNG or TIFF file formats (64).

During image processing, images are often converted from RGB to an alternative colour space. This may include colour information or represent brightness only (where  $z = 1$ ). A popular colour space, defined by the Commission Internationale d'Eclairage (CIE), is CIELAB (CIE 1976  $L^*a^*b^*$ ) (201). CIELAB is a uniform colour space with axes that approximately correlate to perceptual attributes. This can be useful when enhancing imagery. It encodes an image in terms of its brightness or *luminance* ( $L^*$ ) and chromatic information ( $a^*$  &  $b^*$ ), where at each pixel they have a  $L^*$ ,  $a^*$  and  $b^*$  triplet (145). The  $a^*$  channel summarizes the red/green component of images whereas the  $b^*$  channel summarises the blue/yellow component. Pixel values fall between  $[-128, 127]$  in the colour channels while brightness values lie between  $[0, 100]$ . However, pixel values in CIELAB are often normalized, to  $[0, 1]$ , particularly before attempting alterations.

It is far simpler and, typically, preferential to alter the brightness channel of an image rather than the colour channels. Performing colour adjustments can change the colour balance and result in unnatural colour casts in the image. Additionally, brightness based on luminosity, is a closer representation of how the human eye perceives brightness, more closely mimicking the discrimination of colours by a human observer. The CIELAB colour space is therefore helpful in this regard, allowing modification of the brightness component  $L^*$  independently of the colour channels  $a^*$  &  $b^*$ . HSV (172), or Hue-Saturation-Value, is another colour space that separates brightness and chromatic information. Hue refers to the perception of colour (i.e. blue, green, red, yellow), saturation the colourfulness relative to its own brightness and value (or brightness) the brightest colour component, calculated as  $\max(R, G, B)$  (64; 145). HSV images do not require normalization since values are already scaled between  $[0, 1]$ .



**Figure 2.3:** Image histogram of RGB colour channels & brightness channel

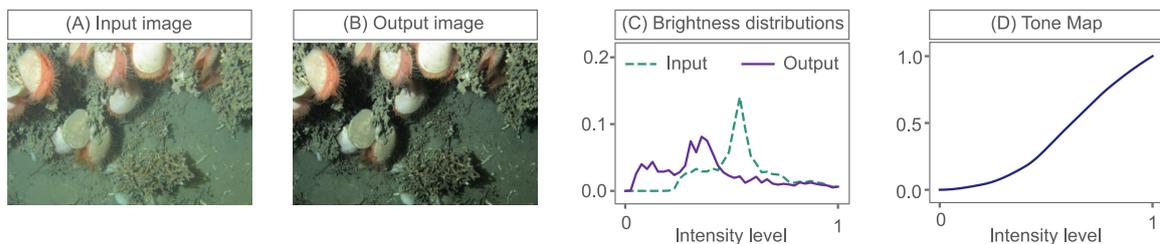
Before attempting to enhance an input RGB image, it is helpful to translate or define undesirable perceptual qualities in a technical sense. One such way is to represent pixel information in a histogram. Here, pixel values within an image, can be organised into a vector demonstrating the frequency of pixels assigned to any given pixel value. This is called a histogram,  $h$ , and can be displayed graphically. Histograms can be used to display the distribution of colour values in an image i.e.  $h(c) = \text{hist}(C(x, y))$  where,  $C$  represents a colour channel  $\in [R, G, B]$  for example with pixel values  $c$ . For a typical 8-bit encoded image, there are 256 possible values; integers in the interval  $[0, 255]$ . However for simpler explanation, we map the image values to the interval  $[0, 1]$  i.e. we divide each pixel value by 255. It follows that if  $\underline{h}$  is a vector of 256 values, then  $h_j$ , where  $j$  is in  $[1, 256]$ , encodes the frequency that the  $j$ th value - which is calculated as  $((j - 1)/255)$  - appears in the image. Given a brightness image  $L(x, y)$ ,  $h$  can therefore also be used to describe the distribution of pixel (brightness) intensities i.e.  $h(b) = \text{hist}(L(x, y))$ , where  $b$  is the brightness values. In Figure 2.3 we present image histograms. On the left we show histograms for each of the colour channels in an RGB image and on the right a brightness histogram. It is often useful to normalise  $h$ , by dividing the raw frequencies of the histogram by its sum. A

histogram whose total bin count sums to one is often referred to a probability density function (PDF).

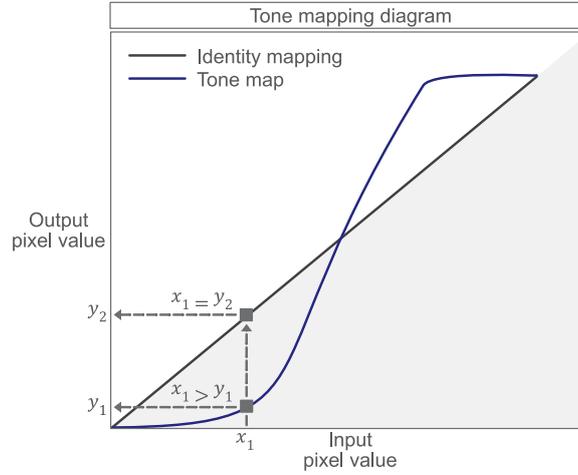
Histograms can summarize a great deal of information about an image. They can describe the statistical distribution of colours as well as the tonality of an image; detailing to some extent contrast and brightness. Interrogation of an image histogram can thus provide a technical foundation upon which to design a suitable image enhancement. The skewness of a histogram, a measure of its asymmetry, can identify whether an image is particularly bright or dark. The brightness histogram in Figure 2.3, for example, is highly skewed towards the darkest pixel intensities and may therefore benefit from a brightness enhancement. A histogram that is Gaussian, or Normal, indicates perfect symmetry and that most brightness values are within the mid-dynamic range. The kurtosis, the spread of the central peak, can indicate the contrast of an image, with a narrow peak (small spread) indicating the need for contrast enhancement. The modality of the histogram, the number of peaks, can describe the prevalence of pixel value ranges and complexity of an image. Image histograms can be unimodal (single peak), bimodal (two peak) and even multimodal ( $>2$  peaks).

### 2.2.2 Tone Mapping for image enhancement

Often in image processing pipelines (and workflows), an input image is modified to make an *improved* output image. Tone mapping, which manipulates image histograms, can be an effective method to enhance the appearance and/or visibility of image features and suppress non-desirable lighting effects. Tone mapping can be framed as a problem of mapping



**Figure 2.4:** An example of tone mapping: (A) an input image, (B) an input & enhanced output brightness distribution, (C) the tone map and (D) an output image



**Figure 2.5:** A diagram of tone mapping

an input distribution  $h(b)$  to a desirable target (output) distribution  $h_{targ}(b)$ , such as in Figure 2.4. Here we have an input image (A) and its enhanced counterpart (B), with their corresponding brightness histograms in (C). Note that for this example, and the remainder of the thesis, a brightness image is defined as the maximum of R,G and B (as in HSV), unless specified otherwise. It is calculated as follows:  $L(x, y) = \max(\underline{I}(x, y))$ . To modify the input brightness image  $L_{in}(x, y)$  to become the output  $L_{out}(x, y)$ , we use a tone curve  $t()$  shown in (D).

A tone curve is simply an increasing function, or transform curve, that defines how to change image pixel values. In Figure 2.5 we show a diagram of tone mapping and how it functions as a look-up table. Here we see an identity mapping (a line at 45 degrees), which maps an input pixel value  $x_1$  to the same output value  $y_2$ . The tone curve however maps  $x_1$  to a lower output pixel value  $y_1$ . Tone curves can be applied individually, to each colour channel. However, more commonly, they are applied to a brightness image.

To modify image pixel values in a brightness image ( $L_{in}(x, y)$ ), with a tone curve we use the following:

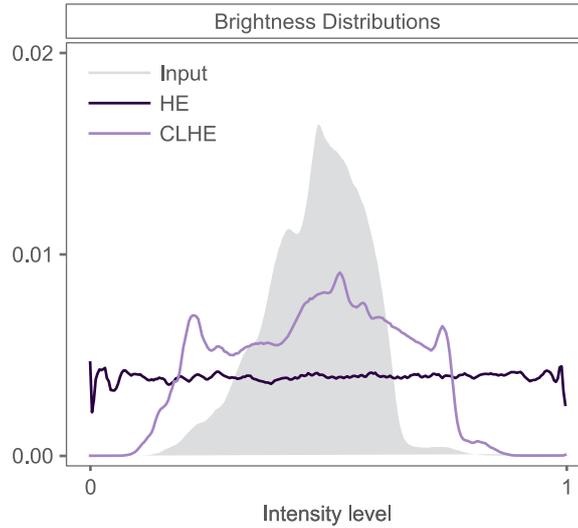
$$L_{out}(x, y) = t(L_{in}(x, y)) \quad (2.1)$$

Typically, the next step is to convert the output brightness image back to the original colour space, in this case RGB. Using the maximum as a definition of brightness, rather than the mean (for example), allows for simpler calculation of output colour images. Given an input color image  $\underline{I}_{in}(x, y)$  the corresponding max brightness input image is denoted  $L_{in}(x, y)$ . Given an output brightness image  $L_{out}(x, y)$  (created by tone mapping, for example) the output colour image,  $\underline{I}_{out}(x, y)$  is computed as:

$$\underline{I}_{out}(x, y) = \underline{I}_{in}(x, y) \frac{L_{out}(x, y)}{L_{in}(x, y)} \quad (2.2)$$

Since our brightness is the per-pixel maximum of R, G and B,  $L_{out}(x, y) \in (0, 1]$ . That is, the max definition ensures that  $\underline{I}_{out}(x, y)$  is in the display range, see (146). In contradistinction, if we had defined brightness to be the mean of R, G and B,  $L(x, y) = mean(\underline{I}(x, y))$ , then the output colour image (Equation 4.3) can have values larger than 1 and thus not be displayed directly. By using the maximum of R, G and B as our brightness, we avoid the question of what to do if the manipulated brightness values fall outside of the display range.

Tone mapping is a common tool in photo-editing software; employed either practically, through interactive manipulation of a tone map, or algorithmically, which can require the user to supply parameters or function automatically. Tone mapping can be used to enforce a variety of desirable tonal output qualities. In Figure 2.4, the output histogram is left-skewed, describing a darker image than the input. The histogram also appears flatter, an adjustment enforced by the tone mapping algorithm Histogram Equalization (HE) (88; 64). In HE, the target distribution,  $h_{targ}(b)$ , is a uniform or equalised histogram. The predominant motivation for the usage of HE is to improve image contrast. In poor contrast images, pixel values may be confined to a specific range of the histogram. A dark image for example, may have pixel intensity values aggregated towards zero, the darker portion of the dynamic range (the range of possible pixel values). In order to increase contrast, the histogram can be stretched and flattened across the full dynamic range such that its mean brightness is shifted to the middle of the dynamic range. According to information theory, this increased



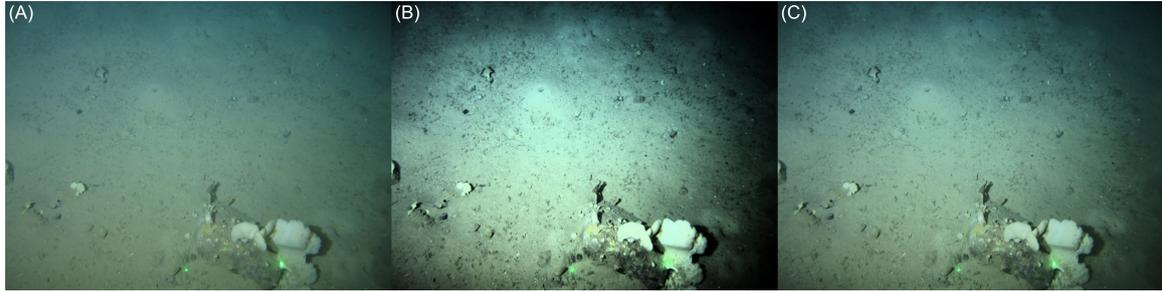
**Figure 2.6:** Brightness histograms of images in Figure 2.7(A), including an unenhanced input and a HE & CLHE enhanced output

uniformity in the histogram will yield an increase in entropy, and therefore an increase in information or details within the image (164; 210).

In HE, the tone mapping function  $t()$  is the cumulative distribution function (CDF), or integral, of the input brightness distribution; a histogram that is normalised to sum to one.

$$t(b) = \int_0^b h(b) db \quad (2.3)$$

We show a worked example of HE in Figures 2.6, 2.7 and 2.8. In Figure 2.6 we show an input brightness distribution and its *flattened* HE output. Observing the corresponding images in Figures 2.7 (A) & (B), we see that this contrast enhancement has improved the conspicuity of edge details. The HE method can not only improve the visibility and therefore segmentation and identification of features in an image, such as species or habitat, it can be used to standardize the appearance of features under different illumination (162). This is particularly important for automated image analysis, since recognition systems are known to be sensitive to varying brightness conditions (4). In addition, HE is not computationally

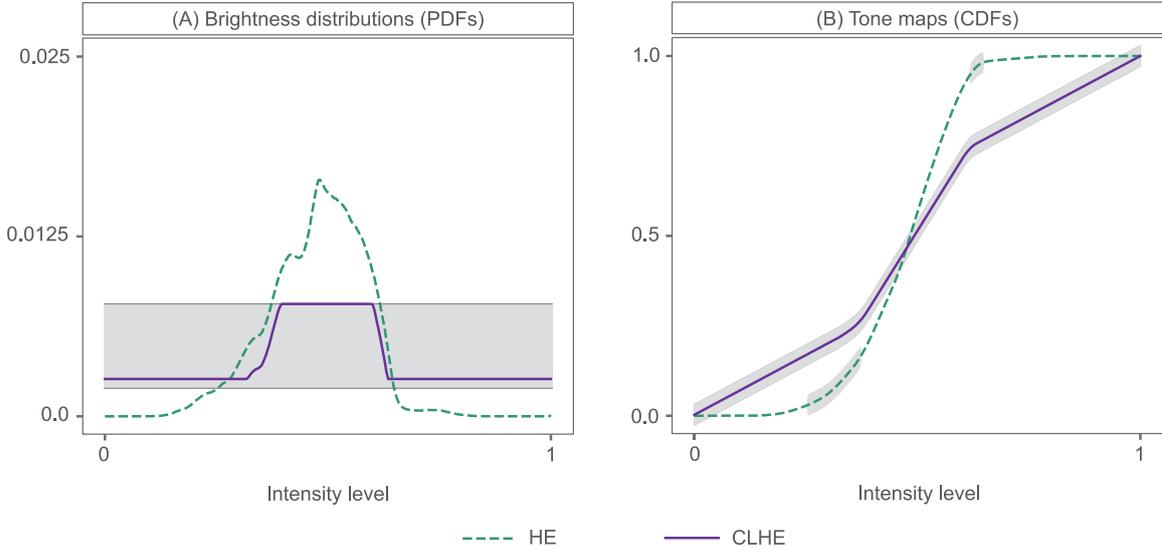


**Figure 2.7:** Enhancement example: (A) Unenhanced image, (B) HE image and (C) CLHE image with a minimum slope of  $0.5/N$  & a maximum slope of  $2/N$ , where  $N = 256$  bins

intensive, is easy to use (given it is automatic) and an operation that can be easily reverted if the transformation function is known. These factors make HE a popular choice in image processing.

However, rather than improving an image, HE can often give a worse result. In Figure 2.8(A) we show the corresponding brightness distribution of the input image from Figure 2.7(A), labelled as ‘HE’. In Figure 2.8(B) we show the HE tone curve - the cumulative distribution of the input (or HE) histogram (A) - that delivers the output image in Figure 2.7(B). Note that tone curve has low slopes for small and large brightness values. In the equalised image (Figure 2.7 (B)), this translates to a loss of details in darker regions (i.e. bottom right) and bright regions (i.e. sponges on the cobble). In the mid-brightness range of the tone curve, the steepness of the slope, driven by sharp differences in bin frequencies, cause adjacent pixel values to map to very different values. This causes the equalised output to appear overly enhanced, with too much contrast. Histogram equalized images often appear over-enhanced, with significant and abnormal brightness changes.

To address the issue of over-enhancement, the HE method can be constrained. In Figure 2.7 we show a second enhanced image (C), possessing more detail than the input image (A), yet without the artefacts present in (B). This has been enhanced using the Contrast Limited Histogram Equalisation (CLHE) algorithm (138; 214). In Figure 2.6 it is evident that the brightness histogram following CLHE is more flat, indicating greater contrast, yet not fully



**Figure 2.8:** HE & CLHE tone mapping: (A) An Input PDF (HE proxy histogram) and its slope-limited (CLHE) proxy, (B) their respective CDF's or tone maps. Grey lines in (A) depict upper and lower slope bounds of  $2/N$  and  $0.5/N$  respectively, where  $N = 256$  bins. Grey shading in (B) highlights area of each tone map that fall within the slope thresholds.

equalized. The CLHE tone curve to enforce such a distribution is shown in Figure 2.8(B). We can see from this that the tone curve never has a slope less than  $0.5/N$  or greater than  $2/N$ , shown by the shaded areas. This slope-constrained tone curve is the cumulative distribution of the CLHE-derived histogram shown in Figure 2.8(A).

CLHE strives to find a histogram  $h'(b) \approx h(b)$ , such that the minimum slope ( $m$ ) and maximum slope ( $M$ ) of its cumulative histogram is bounded. Where  $h(b)$  represents our input (HE) histogram and, in this case,  $m = 0.5$  &  $M = 2$ , as depicted by the grey horizontal lines in Figure 2.8(A). To understand how this is achieved, let us move to the discrete domain. We map the continuous histogram  $h(b)$  to a discrete  $N$ -vector  $\underline{h}$ , which has  $N$  bins and sums to 1. Here, the input domain  $(0, 1]$  is uniformly sampled into  $N$  regions. As such the  $j$ th bin is in the interval  $(\frac{j-1}{N}, \frac{j}{N}]$  and  $h_j$  is the proportion of brightness values in the image that fall in this interval; under the assumption that histograms are normalised to sum to one.

In the continuous domain, the tone curve for HE is the integral of  $h()$ , see Equation 2.3. In the discrete domain, we have an  $N$ -element tone curve,  $\underline{s}$ , which is analogously defined. Here the integral is replaced by a summation:

$$s_j = \sum_{i=1}^j h_i \quad (2.4)$$

As the size of each histogram bin is  $1/N$ , the slope of the tone curve at the  $j$ th brightness level  $s_j$  is defined as:

$$s_j = \frac{s_j - s_{j-1}}{1/N} \quad , \text{ if } j > 1 \quad (2.5)$$

$$s_1 = \frac{s_1}{1/N}$$

Clearly, we can also write the slope as:

$$s_j = \frac{h_j}{1/N} = Nh_j \quad (2.6)$$

If  $h_j = 1/N$ , then the slope of the corresponding tone curve (cumulative histogram) is always equal to one. That is the tone-curve is a line at 45 degrees. And, as we would expect, if we try and equalise an image that already has a flat histogram then the tone curve is a *null* operation i.e. each input brightness maps to the same output brightness.

In the case of CLHE, we would like to flatten an input histogram, yet limit the extent of this imposed uniformity, using a tone curve with a bounded slope that is neither too large nor too small. The CLHE algorithm achieves this by finding a proxy histogram  $\underline{h}'$  that is similar to the original but in which the slope conditions are adhered to. Mathematically, we would like the distance,  $||\underline{h}' - \underline{h}||$ , to be small and the slope constraint  $m \leq N\underline{h}' \leq M$  to be met.

However, what, exactly, is meant by ‘small’ is not well defined. As discussed in (129), CLHE appears to empirically minimize a least-squares error; indeed it generally returns the same proxy as an algorithm that minimizes the least-squares error. So, to a first approximation we can consider ‘small’ to mean a minimum least-squares error. The integral of this *proxy* brightness histogram (Figure 2.8(A)) creates the CLHE tone map shown in Figure 2.8(B).

Both CLHE & HE are global tone mapping operations (TMOs) and thus may smooth over small local variability in pixel values (214). This could be problematic for species or seabed features that are either small and/or not very distinctive. CLHE is therefore often deployed in its adaptive form which is called Contrast Limited *Adaptive* Histogram Equalization (CLAHE). In CLAHE, an image is tiled into several non-overlapping regions, say a  $4 \times 4$  grid. We calculate the histogram of each tile and use CLHE to derive a slope-limited tone curve. This tone curve is associated with the central pixel in each tile. Individual pixels are then mapped to output values by bilinearly interpolating the output values found for the 4 tone curves (from the 4 neighbouring centres). This better accommodates local variability and enhances contrast within an image, whilst constraining the introduction of noise, particularly within homogenous regions i.e. shadowed or dark regions. HE also has a locally-adaptive form, Adaptive Histogram Equalization (AHE) (138), but is less used. Although these tone mapping algorithms enforce a uniform target distribution on the histogram of the output image, this need not be the case. Following initial recommendation by (48), many CLAHE applications to underwater images seek a target histogram modelled by a Rayleigh distribution (RD) (143). Further details on the RD are provided in Chapter 4.

For analysts, employing CLHE in its global or adaptive form is more complex than its HE equivalent, given that it requires more parameters to be supplied. However, these are relatively intuitive, given that they correspond to the level of contrast and number of regions in the image. If integrated into image manipulation software, they thus have the potential

to be easily toggled by analysts. Importantly, the additional parameters in CLHE and CLAHE allow analysts to generate more personalized enhancements; supporting individual preferences and those related to image content. However, studies rarely consider potential benefits in such a task-focused context, typically concentrating solely on improved aesthetics ???. This is an important issue given the usage of imagery extends beyond aesthetics - such as the application in this thesis.

## 2.3 Computer vision for automatic classification

Incorporating computer vision into benthic image analysis workflows, specifically automatic classification or detection, is widely believed to be essential. Not only to cope with accumulating data from autonomous imaging platforms but to optimize current analysis methods, improving the efficiency, consistency and quality of ecological data extraction from images.

### 2.3.1 Image features

Image classification is simply the assignment of a label to an image based on visual patterns. Labels are simply categories that describe the image contents or objects that can be represented textually or numerically (one-hot encoded). Whether undertaken by a human or a computer, image classification is *broadly* the same. For example, analysts identify a species or habitat from an image by scrutinizing the physical and/or biological characteristics based on domain knowledge and experience (or training). Additionally, this decision is often supported by contextual information about the environment in which the image was captured, such as geographic region and bathymetric depth (distance below sea level).

A computer vision or machine learning algorithm can be trained to achieve the same task. Simply put, given an input image, some information or features must be extracted which can then be passed to a learning algorithm; this *learns* to assign these features to an appropriate classification. Features could merely be pixel values in an image, however this would

perform poorly in most classification tasks. Alternatively, feature extraction algorithms, or feature descriptors, exist to derive more complex patterns from an image based on colour, tonal or textural attributes. Some low-level features descriptors include Scale Invariant Feature Transform (SIFT) (115), Histogram of Orientated Gradients (HOG) (33), Speeded Up Robust Features (SURF) (14), Linear Binary Patterns (LBP) (6), Grey Level Co-occurrence Matrices (GLCM) (71), Gradient Location and Orientation Histogram (GLOH) (131), Maximally Stable Extremal Regions (MSER) (125). However, in complex classification tasks, these are not always suitable.

Instead, deep features extracted by Convolutional Neural Networks (108; 66) have proven to be extremely useful in image classification (102; 170; 176; 74; 86). These are artificial neural networks that learn bespoke feature descriptors, using convolutional operations to aggregate simpler features into more complex features. We provide further details on CNNs (as both a feature extractor and classification algorithm) in the following section, broadly discussing their architecture and function.

### **2.3.2 Classification approaches: simplicity vs. complexity**

A number of machine learning algorithms have been used to classification image features, such as Decision Trees (140; 76), K-Nearest Neighbour (KNN) (67), Linear & Logistic Regression, Bayesian Network Classifiers (53), Adaptive Boosting (AdaBoost) (187), Convolutional Neural Networks and Support Vector Machines (28), to name a few. In this work we focus on the latter two, which differ in their theoretical and practical complexity for a non-specialist such as a marine ecologist.

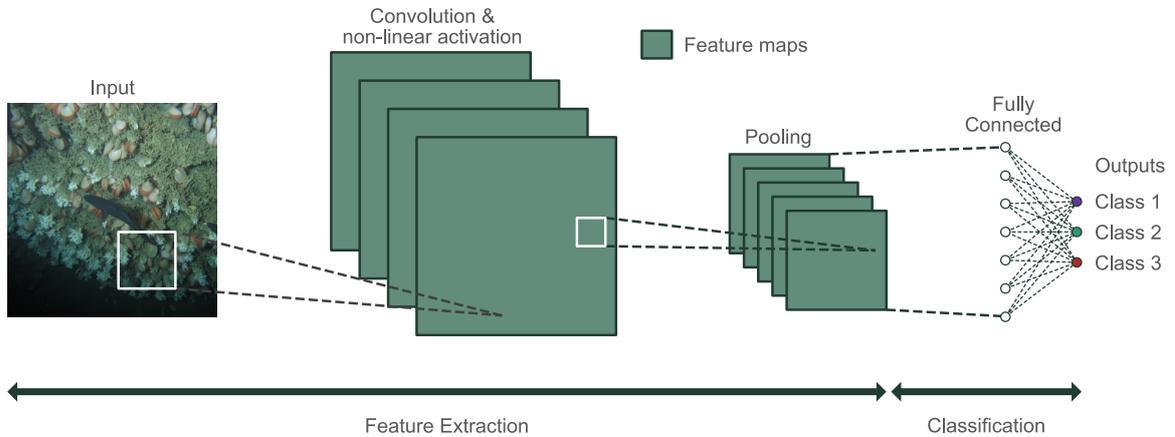
#### **Convolutional Neural Networks**

CNNs are biologically-inspired deep learning (DL) models (a subset of ML models) that use convolution to transform image data and extract features (108; 66; 9). Convolution, or a convolutional operation, refers to a kernel or filter (a matrix of discrete values), that moves across an image (a *sliding window*) altering pixel values through multiplication and

summation. This happens many times with different kernels, building a representation of the image in which patterns, or features, are evident. They may highlight edges, textures or even objects within the image. Features (the transformed pixel values) are then flattened into a 1-Dimensional vector and act as input variables to a neural network (NN) within the CNN for classification. Note that though they can be used independently, when housed within a CNN, NNs are referred to as a fully connected (FC) layer(s).

FC layers are a convenient way of mapping all feature information into a specified number of outputs. CNNs typically house multiple ( $>2$ ) FC layers in succession. Although each of them perform the same function in a technical sense, it is helpful to think of the first FC layers as reducing the large number of features extracted following convolutions or further clarifying patterns within these features. Whereas the last FC layer can be thought of as the true classification or output layer. Here the number of outputs is equivalent to the number of classes in the dataset. FC layers map feature data to the specified outputs by fitting complex curves or hyper-surfaces to the data. They do this by (1) linear transformations, multiplying features by weights and adding biases and (2) non-linear transformations, adjusting features with an activation function. Similar to regression, the FC layers are therefore merely approximating relationships (or functions) between the extracted image features and each of the outputs. The output values are interpreted *like* a probability (logits), showing which class the image features best correspond to. This is decided by taking the maximum output value, known as the *argmax*. Note that these are not true probabilities; their sum may be greater than 1. To ascertain the normalized probability that each image represents one of the possible classes, the output values can be converted using a *softmax* function.

We simplify and recap this process in Figure 2.9. For each image that passes through a CNN, it will first be modified by numerous convolutional operations, creating multiple image versions. These are referred to as feature maps and are then flattened and combined together in a 1D feature vector. The features are then classified by passing through the FC



**Figure 2.9:** Diagram of a simple CNN for image classification, with 3 output classes.

layer and compressed into the specified number of outputs, in this example 3. The image is classified as the class with the maximum output value.

Without training, passing an image through a CNN will likely result in an inaccurate class prediction. However, both the convolutional operations and FC layers can be optimized such that the predicted outputs better match the real classes. Suffice to say that for each image, or group of images (batch), that is passed through the CNN, the parameters (convolutional filters, weights and biases) that transform pixel values and the subsequent extracted features, are updated through a process known as backpropagation (150). In short, this monitors the error between predictions output from the final FC layer and the observed classes. For multi-class classification, cross-entropy loss is typically used as the error metric, or loss function. The CNN then uses gradient descent (an optimization algorithm) to update the CNN parameters according to the size of this error multiplied by the learning rate (step-size). In effect, large errors will result in a large change to parameters, scaled according to the learning rate. Backpropagation will gradually optimize CNN parameters and improve performance as more images, and thus examples of each class, are passed through. When all training image data has been passed through the CNN, this is termed an epoch. Backpropagation is repeated such that the CNN is trained for a

specified number of epochs, or until prediction accuracy either stabilises or reaches a set threshold.

CNNs are highly effective at extracting and classifying visual features (102; 170; 176; 74; 86). They start by recognising low-level features such as edges and textures, gradually building to more complex, high-level feature representations such as parts of objects or entire objects. For example this could be the fins, markings and body contour of a fish which contribute to its full appearance and thus enable the CNN to predict its class (153; 52). To do this, they requiring enormous volumes of imagery (1000s) for training and computer resources such as high-priced Graphical Processing Units (GPUs) which may not be available. As a result, many CNN applications rely on a process known as *Transfer Learning* in which a CNN, trained for a different classification task (or dataset), is adapted by training further on their own classification problem. Using pre-trained networks as a head-start enables extraction of generic high-level features that can support classification. These can then be refined, or optimised, with further training on the new dataset. Common CNNs used for this purpose include VGG-16 (170), ResNet-50 & ResNet-101 (74).

Many studies have demonstrated the possibility of using CNNs to classify benthic taxa in optical imagery (124; 104; 137; 106; 43; 1; 40; 95; 136) and this continues to grow each year. Shallow-water corals are a particularly common benthic target for CNN classification (18; 120; 68; 69; 121), largely thanks to the availability of open-source datasets such as MLC (16), EILAT and RSMAS. In some cases, reef-forming corals can be considered to occur as both a species and a habitat (83). However, the examples in these datasets are typically highly-zoomed images of coral texture and thus do not represent well the broader structure and contextual appearance of the habitat at the captured image resolution.

Currently, there is less literature employing CNNs at the habitat level in benthic applications (142; 122; 119; 132; 204; 202). The majority of these utilise shallow-water datasets, including open-source datasets such as Benthos15 (19) and Tasmania (197), as well as those on corals mentioned previously (MLC, EILAT, RSMAS). Such datasets are important for bench-

marking of methods however they do not represent the application in this thesis which varies with respect to the region covered and their associated habitats, as well as camera and image properties (platform, lighting), see Chapter 3.

Common amongst these studies is the use of pre-trained CNN's (transfer learning) as a feature extractor to support classification with Support Vector Machines (SVM), an alternative machine learning algorithm, see Section 2.3.2 for more details. Mahmood et al., (122) showed that 'off-the-shelf' ResNet50 features outperform traditional hand-crafted features in the classification of kelp habitats using Linear SVM's. Similarly, Mahmood et al., (119) used the same to classify benthic images from multiple datasets, including MLC, EILAT and RSMAS. However, they found that combination of features from different ResNet layers resulted in even more powerful feature descriptors for classification. Mohamed et al., (132) instead utilised non-linear (polynomial) SVM classification of shallow benthic habitats including corals, algae, seagrass and sediments. They found that features extracted from a pre-trained VGG16 could be improved with additional features provided by a pre-trained 'bagging of features' (BOF) algorithm. Overall performance could also be improved, to a lesser extent, if the VGG16 features were combined with ResNet50 features.

One key benefit of these approaches, aside from good performance, is the reduced requirement for training imagery by employing transfer learning and the use of ML classifiers which are more appropriate with smaller datasets (SVMs - see Section 2.3.2). Other habitat applications use a different approach to reduce labelling effort; unsupervised learning. Yamada et al., (204; 202) use image location information to guide autoencoder training to identify subsets of unlabelled benthic images that include kelp, reef, bacterial mats and varied substrates. These can then guide manual annotation efforts and/or pseudo-labelling based on relative locations, in latent representations, to annotated imagery. Subsequent performances yielded competitive performance conventional CNN transfer learning (AlexNet, ResNet18 and ResNet152) as well as more traditional classifiers like SVMs. In a similar sense, Rao et al., (142) explored autoencoders guided by acoustic bathymetry (ocean depth) to improve

**Table 2.1:** Hyperparameter glossary for CNN & SVM training

---

<b>CNN:</b>	
Batch size	The number of images you send to the model in each iteration. <i>Model parameters are updated after each batch during training.</i>
Epochs	How many times you pass the full image dataset through the model.
Loss function	The error metric that you wish to minimize. <i>e.g. Cross entropy loss for multi-class classification.</i>
Learning rate	A small number (0, 1] that determines the amount to alter parameters during training with respect to the loss. <i>Also known as the step size.</i>
Optimizer	An algorithm that modifies CNN parameters according to a particular strategy to minimize the loss. <i>e.g. the Adam optimizer sets the learning rate adaptively for faster and more efficient training.</i>

---

<b>SVM:</b>	
$C$	A regularization parameter that offers a trade-off between the maximum-margin and misclassification rate. <i>e.g. A large <math>C</math> enforces a small margin hyper-plane maximizing classification accuracy.</i>
$\gamma$	A value to determine the distance over which support vectors influence the hyperplane. <i>e.g. A high <math>\gamma</math> considers only points that are close to each other and causes the decision-boundary to be highly curved.</i>

---

*N.B.* This list of hyperparameters is not exhaustive.

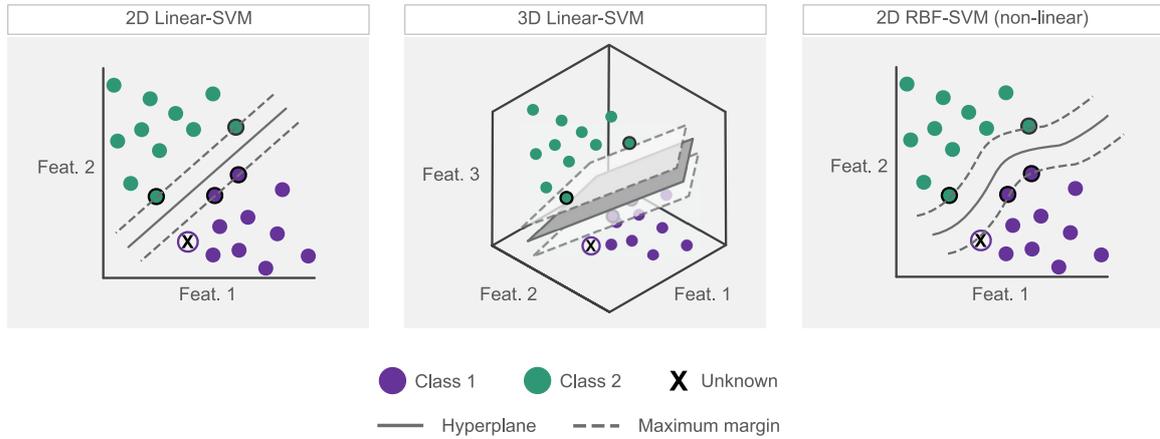
CNN performance in benthic habitats. They found that visual imagery of benthic habitats is suitable for automatic classification and does not benefit from the addition of bathymetric data i.e. encoding the relationship of visual features with bathymetric data. However, the inverse was significantly improved.

The methodologies of the approaches presented are not always straightforward from a user perspective, for example the training of a CNN is highly intimidating, and often inaccessible, to a non-specialist such as marine ecologists, without collaboration with experts (183). Understanding their mechanics and theory is complex and requires large investments in time for comprehension and implementation. They require several hyper-parameters, which must

be set manually by the user, see Table 2.1. These also require a good deal of theoretical study to understand, train and use them effectively. Additionally, for individuals without more in-depth programming knowledge, particularly in Python or Matlab, they might be challenging to train. A better option for novice users, as already employed by some benthic applications, could be to use a pre-trained CNN (feature extractor) and pair with a traditional machine learning algorithm, such as a SVM, that is simpler to train. However the simplicity of this method does vary with respect to the CNN employed (122). ResNet features, for example, require processing of the convolution block(s), into a 1-dimensional feature vector, to be extracted. This is due to the single FC layer being unsuitable since it maps feature outputs to the ImageNet classes. By comparison, VGG16 (which contains 3 FC layers) is simpler as you can simply extract features from the penultimate FC layer.

### **Support Vector Machines**

SVMs are a shallow learning technique that classify images well using deep features extracted by a CNN (144; 11; 153; 121; 122; 119; 132). Created for binary classification problems, they function by identifying the best boundary, or hyperplane, between data points that enables the distinction of two classes. A hyperplane is a flat affine subspace of dimensions  $n-1$  in an  $n$ -dimensional space. In its simplest case, consider a 2D ( $n = 2$ ) scatter plot of data represented by two features (or variables); one on the x-axis and the other on the y-axis. Data points, in this 2D space, are classified as class 1 or 2. An SVM will find a hyperplane, in this case a 1D ( $n - 1$ ) linear boundary (or line) that maximises the distance between the data points of each class to this boundary. SVMs classify data points simply by observing where they lie with respect to this hyperplane separated 2D space, see Figure 2.10. Unlike a CNN, the output is a predicted class rather than a probability that it is either of the classes. It is important to note that the data points in an SVM classification are not restricted to  $n=2$  dimensions (2 features). In Figure 2.10 we also show an SVM example when  $n=3$ . However, in reality, data points can be the intersection of a significant

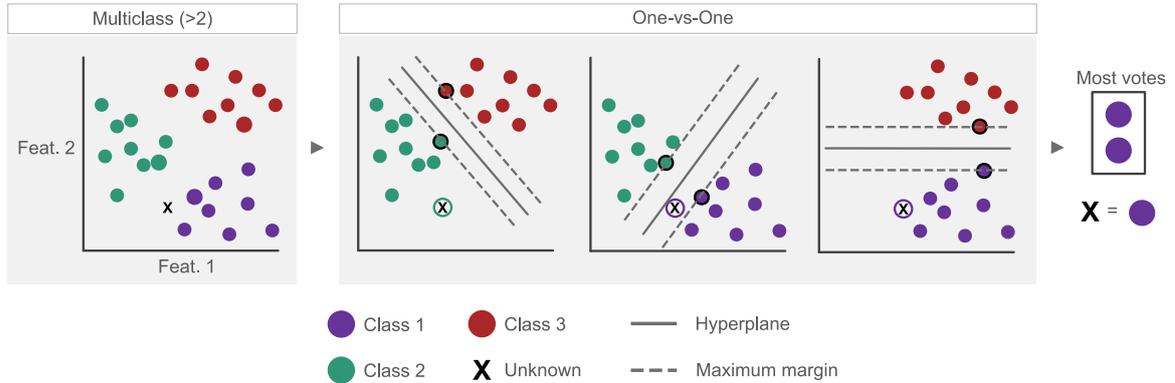


**Figure 2.10:** A diagram of Support Vector Machines

number of features. The VGG-16 CNN feature extractor (170) for example, extracts 4096 features from each image, creating a feature space of 4096 dimensions.

The SVM uses the points closest to the hyperplane, known as support vectors, to guide hyperplane placement. The support vectors are the hardest points to classify, given the potentially close proximity of support vectors of each class. Optimal placement is therefore found by maximising the distance between the support vectors of each class and the hyperplane (known as the margin) such that the misclassification rate is minimized. This is why SVMs are referred to as maximum-margin classifier; they find the hyperplane that is equidistant between the two classes. Using only the support vectors, and thus a subset of the data, to place the hyperplane in this way is very memory efficient compared to a CNN which uses all data in training.

In the event that data points are not linearly separable in the original  $n$ -dimensional space, an SVM will transform the data points to a higher number of dimensions; such that a linear hyperplane that separates the two classes can be found. Reducing the dimension of this optimal hyperplane back to  $n-1$  dimensions will then result in a non-linear separation of the data points. For simpler conceptualization, say that the SVM moves a 2D feature space into 3D, such that a 2D hyperplane is found. This hyperplane is then converted to 1D, creating a non-linear line.



**Figure 2.11:** Multi-class Support Vector Machines: a diagram of the One vs. One strategy

The movement of data points into a higher-dimensional space is achieved by kernels. Note that kernels in this case are mathematical functions. In truth, kernels do not physically transform the data as this is computationally expensive. Instead, they compute and optimize distances between support vectors and the hyperplane, assuming they were projected in a higher number of dimension. This is referred to as the *kernel-trick*. (84) recommend the use of a Radial Basis Function (RBF) kernel as a good default choice, as it can find both a linear and non-linear hyperplane at high dimensions, see Figure 2.10.

The RBF kernel requires only 2 parameters:  $C$  &  $\gamma$  (gamma), see Table 2.1.  $C$  is a regularization parameter that controls the compromise between the margin maximization and the number of misclassifications. Whilst  $\gamma$  enforces the region (distance from the margin) in which support vectors have influence, controlling the curvature and complexity of the decision boundary.

In the event of a multi-class classification ( $>2$  classes), as is the case in this thesis, there are a few strategies that enable use of an SVM. We use a typical approach called One-Vs-One (OVO). This splits the dataset into multiple binary classification problems that are assessed per each pair of classes. Compiling the classifications of all binary SVMs allows a final classification to be made for each data point, based on the class that received the most votes, see Figure 2.11.

SVMs guarantee an optimal solution to be found when classifying data. They always find the best placement for the hyperplane. Neural networks, or FC layers, however may never converge to an optimum during gradient descent. The number of SVM hyper-parameters required is also far less; only 1 or 2 depending on the kernel used (Table 2.1). SVM models are computationally efficient and can therefore easily be re-trained with different parameter combinations to find an optimal set. Training a CNN however requires a great deal more time and effort. Since the FC layers use all of the data and require a large number of intricately connected parameters, this leaves them prone to over-fitting and poor generalization (102; 9). For this reason, it is necessary to use large and diverse datasets when training CNNs. For small datasets, an SVM classifier may be more appropriate. SVM classifiers are less restrictive with regards to resources (time, computing & imagery) and experience.

## 2.4 Summary

This chapter has provided a background on some of the challenges associated with using imagery data for marine ecological research. We focused on histogram manipulation with tone mapping as a simple, yet powerful image enhancement tool to improve interpretation of underwater imagery. We also discussed how the efficiency and quality of benthic image analysis can be bettered through automatic classification using computer vision. In the following chapters of this thesis, we aim to improve the image analysis pipeline for ecologists, exploring solutions in these areas of image processing and computer vision. We design a tone mapping tool that is tailored to underwater image analysis, being both simple and quick to operate by end-users and possessing a more appropriate degree of customization. We also present machine learning approaches to classify benthic habitats from imagery, that offer a good trade-off between performance and accessibility for non-specialists.

### 3 Image Datasets

Three underwater image datasets were provided by Gardline Ltd. for this thesis, see Table 3.1. They cover multiple geographic regions; 3 unique locations within the Norwegian Sea. Further geographic information is unavailable due to commercial sensitivity. The datasets span two types of imaging platform commonly used in seabed surveys; ROVs and Drop Cameras, the specifications of which are provided (Table 3.1). Examples of these platforms, as well as others, can be found in Figure 2.1. Note that all images presented in this thesis belong to either Gardline Ltd or the thesis author, unless otherwise stated.

**Table 3.1:** Benthic Image Datasets used in this study

<b>ID</b>	<b>Platform</b>	<b>Components</b>	<b>Image Specifications</b>	<b>Location</b>
1	ROV	Imenco Tiger Shark 14mpx with external flash (Lantern Shark), 10 × ROS (MV-4000) and 4 × Innova Gas lights.	8353 RGB images (4320×3240, <i>.jpeg</i> )	Norwegian Sea
2	ROV	Kongsberg/Simrad (OE14-208) 5.0mpx, 1 × forward-facing strobe, 2 × fixed & 2 × mobile LED lamps.	1240 RGB images (2592×1944, <i>.jpeg</i> )	Norwegian Sea
3	Drop camera	Kongsberg/Simrad (OE14-208) 5.0mpx, 1 × forward-facing strobe, 2 × fixed & 2 × mobile LED lamps.	574 RGB images (2592×1944, <i>.jpeg</i> )	Norwegian Sea

The surveys to collect each of these datasets involved lowering the imaging platform to the seafloor, with height controlled using a USBL beacon and the real-time video feed - transmitted via an umbilical/sonar cable. The ROV is manually controlled by pilots on the accompanying vessel whereas the drop-camera system consists of a stainless steel frame that moves progressively along the seabed via thrusters or drift. For each platform, seabed

**Table 3.2:** Habitat classes encountered within the image dataset

<b>ID</b>	<b>Habitat</b>	<b>Included sub-habitats</b>
SS1	Soft Substrate (SS)	Heavily bioturbated SS, Single sea pen & Sea pen community
SS2	SS Sponge Community	-
HS1	Hard Substrate (HS)	Gravel area, Scattered Cobbles, Cobble and boulder area, Boulder area
HS2	HS Sponge Community	-
Cor1	Reef Framework	Coral rubble zone, Dead & Live <i>Desmophyllum pertusum</i> reef framework
Cor2	Soft Corals	Lone soft coral, Multiple soft coral colonies, HS soft coral community

images were taken using a mounted digital stills camera system with dedicated strobe and video lamp(s). Note that each platform has a unique camera and lighting configuration (Table 3.1). Images are captured consistently throughout a transect, every 30m. Manual shot-selection is also possible.

Each of these datasets are visually inspected by analysts to localise, identify and enumerate features of interest. For habitat classification, the primary habitat of each image was categorized according to an in-house seabed classification guide, which can be found in Appendix A. However, a simplified version is presented in Table 3.2. Analysts assign the appropriate class to each image by looking for both the presence and absence of specific features. For example Soft Substrate habitats should contain no hard substrates and the presence of Sponges or Sea Pens can alter the classification. To perform this task, analysts regard the full (contextual) view of the image, as well as focusing and zooming in on finer details.

As mentioned in Section 2.1.3, this manual procedure suffers multiple limitations; it is impeded by a number of optical challenges related to the physical behaviour of light in water as well as environmental and human factors. Operation of the imaging platforms in these

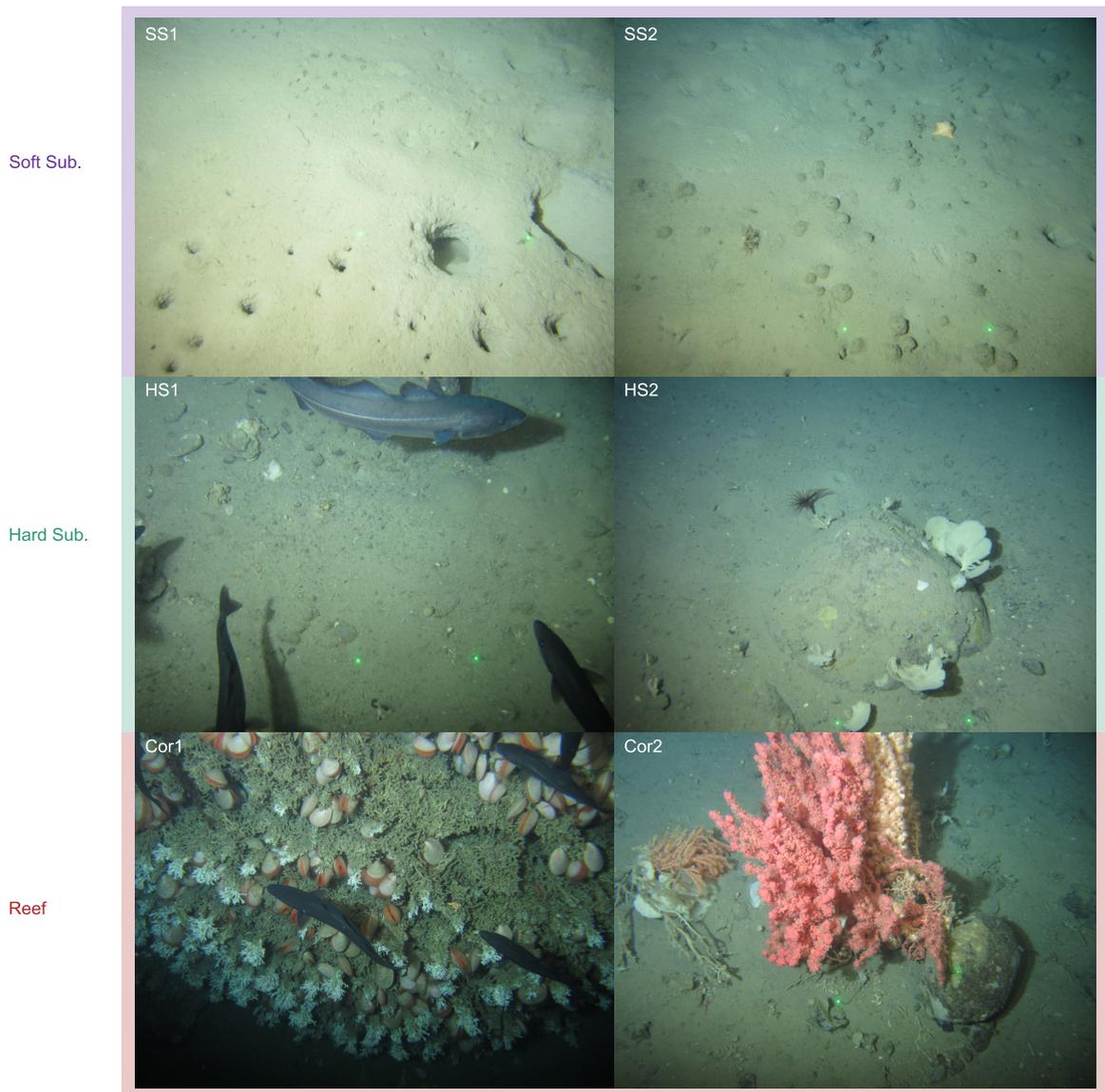
**Table 3.3:** Broad habitat classes and their association with EUNIS habitats

Broad habitat	Abbreviation	IDs	EUNIS (Level 2)
Soft Substrate	Soft Sub.	SS1, SS2	ME5: Upper bathyal sand & ME6: Upper bathyal mud
Hard Substrate	Hard Sub.	HS1, HS2	ME1: Upper bathyal rock
Reef	Reef	Cor1, Cor2	ME2: Upper bathyal biogenic habitat

datasets also introduces further challenges, for example those related to the fluctuation in the height of the camera above the seabed and its angle. This alters the size and perspective of habitat components. It also contributes to the inconsistent lighting patterns that occur in underwater image surveys - in particular those that use artificial lighting. Variable height and angle of the platform can adjust the illumination direction and angle, size of the illumination halo, vignetting patterns, as well as the brightness and contrast levels. Each of these factors can cause seabed features to appear more dissimilar between images and become more obscure. This enhances the difficulty of the image analysis task. Where possible, analysts employ simple image enhancements to aid annotation. However although quick, these enhancements are mostly insufficient, fixed (lacking customisation) and only used in worst-case scenarios in which their benefit is minimal. Analysts require tools that are better optimised, providing benefits to multiple images without compromising on efficiency - i.e. they should be simple and quick to use for an end-user - avoiding further delays in an already time-consuming procedure.

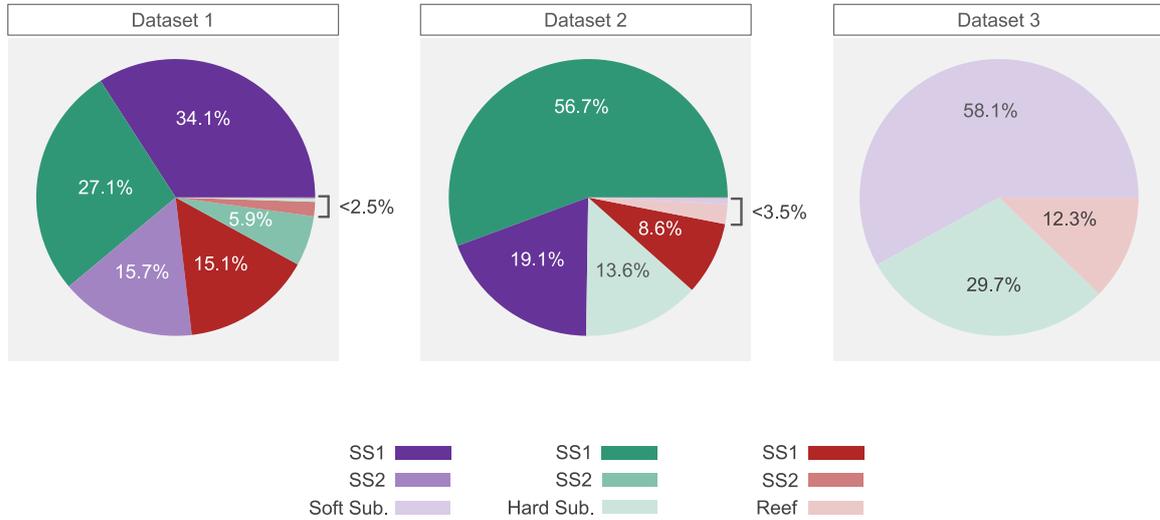
Benthic images in Datasets 1 & 2 were manually annotated by Gardline analysts using CountEM, a bespoke image analysis program developed at Gardline, see Section 2.1.2 and Figure 2.2. All habitat labels were quality controlled for accuracy. In the event of misclassification, labels were corrected with their broad-scale habitat classification for time efficiency. Broad-scale habitat classes are presented in Table 3.3 alongside their European Nature Information System (EUNIS) 2022 classification (49; 46) equivalent for wider context. For brevity, we often refer to habitats by their abbreviation or ID in this thesis, see

Table 3.3. Gardline annotations were unavailable for Dataset 3. Thus to ensure its inclusion in this body of work and in order to provide an accurate and rapid classification, each image was categorised at the broadest level described in Table 3.3. Image examples of habitats encountered across datasets are presented in Figure 3.1.



**Figure 3.1:** Example of habitat classes encountered across image dataset

Although the exact proportions of each habitat varies across datasets, all are dominated by soft and hard substrate habitats, followed by reef as the least dominant habitat type, see



**Figure 3.2:** Proportion of each habitat class across datasets

Figure 3.2. This reflects well the typical constituents of the deep sea benthos. As is normal with deep Soft Sub. habitats (the most common benthic habitat), images categorized as such will often contain signs of burrows, trails and epifauna. This habitat supports a huge diversity of organisms, many not perceptible from benthic imagery, that influence a number of ecological processes (173; 160). Hard or mixed substrates, such as Hard Sub. or Reef however, are often dominated by sessile epifauna, encrusting organisms, and shelter motile epifauna. The increased complexity offered by hard substrates has been linked with greater diversity (103), as they provide stability for sessile organisms, micro-habitats and access to food (via elevation up into current flows) (23). Reef habitat in particular plays an important role in the functioning of deep sea communities and drives biodiversity (147; 23). This is also true for deep-sea sponge grounds (23). We note the presence of both soft and hard substrate sponge communities in these datasets. However, the sponge species within each are different. The variety of habitats, multitude of imaging platforms and geographic separation of these datasets ensure that associated analyses and findings will have a great analytical relevance for both the wider ecology and computer science community.

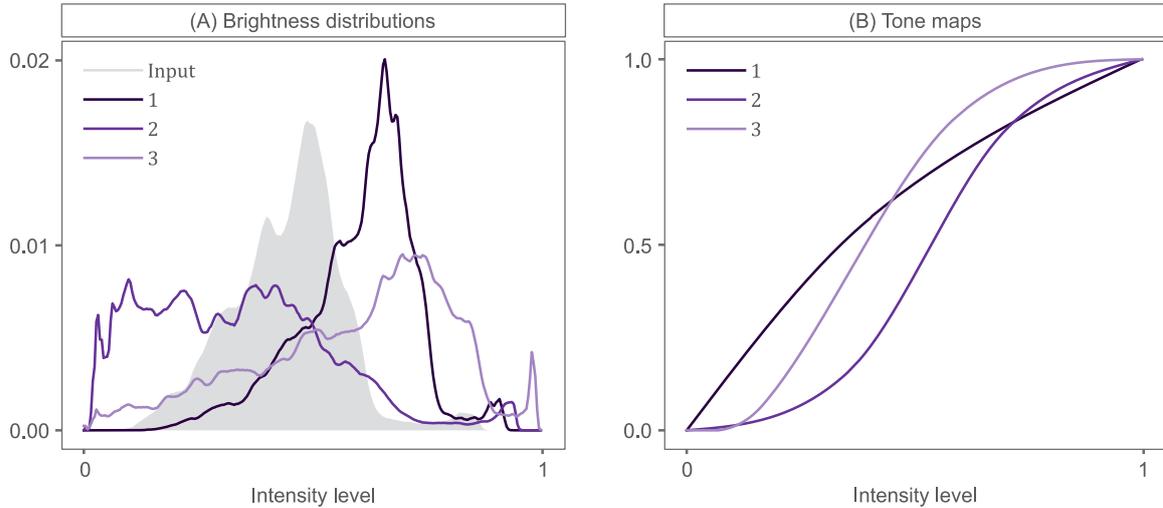
# 4 Bespoke tone-mapping of underwater imagery

Heavy reliance on underwater imaging to support environmental monitoring and conservation in combination with the optical challenges presented whilst capturing said imagery (44), strongly motivate the need to develop image processing pipelines that will better support the interpretation of image data.

In this chapter, we focus on image enhancement, specifically tone-mapping operations (TMOs); a common tool to improve image quality within the industry (as well as our collaborator Gardline Ltd.). Understanding of how a TMO should behave, in order to support image analysis, is still in its infancy within underwater image applications. We therefore provide expert-informed guidance on suitable characteristics for TMOs to support the extraction of ecological information from underwater imagery and explore pathways for their future automation.

## 4.1 Introduction

Underwater images typically lack colour (or are strongly colour-casted), have low contrast, are poorly saturated, blurry and intrinsically dark. They are also subject to non-uniform illumination and shadows within images (77). As discussed in Chapter 2, this is a major hindrance to manual annotation and can also reduce the success of automatic image annotation (156; 44). Since many of these undesirable qualities are explicable with reference to image histograms (or brightness distributions), a simple approach to visually improve images is



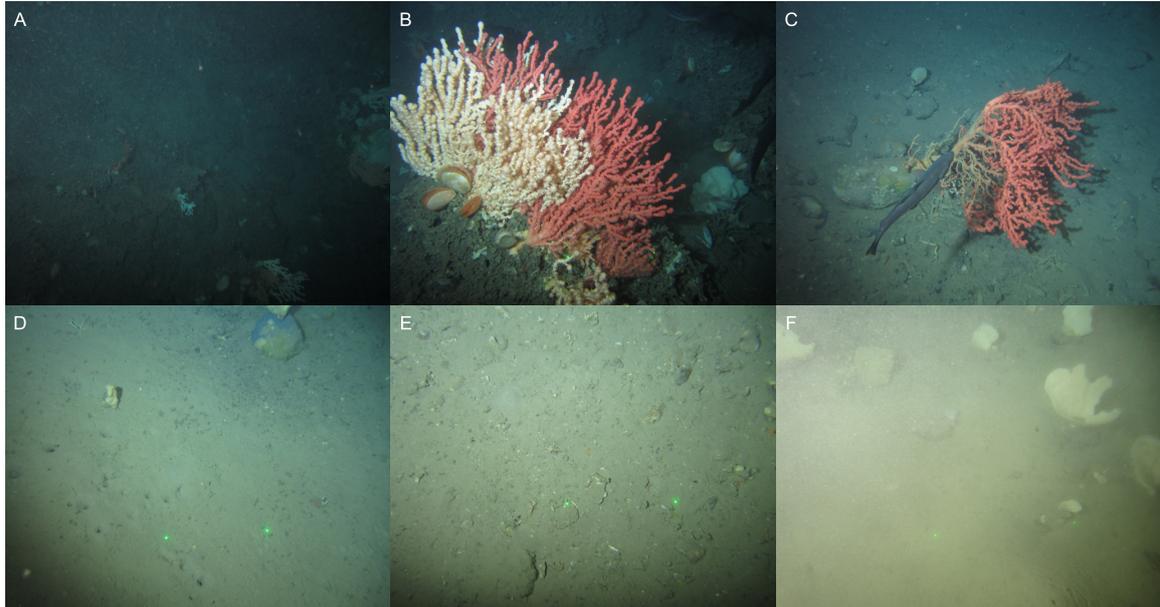
**Figure 4.1:** Example of brightness tone mapping. In (A) we show an input brightness histogram and three modified outputs. The corresponding tone maps are shown in (B). Adjustment 1 demonstrates a brightness increase, 2 a contrast increase and 3 a combined brightness and contrast increase

to alter their pixel values, and by extension their corresponding brightness distribution, see Chapter 2.2.2.

In Figure 4.1 (A), we show an input brightness distribution and three modified output histograms, labelled 1 to 3. Each of these were created by applying the corresponding tone maps in Figure 4.1 (B) to the input brightness image. We see that applying tone map 1, in (B), makes the input image brighter as the peak of the output brightness histogram in (A) is shifted towards brighter values (or higher pixel intensities). Comparatively, tone map 2 increases the contrast (and darkens) the output image, demonstrated by a flattening of its



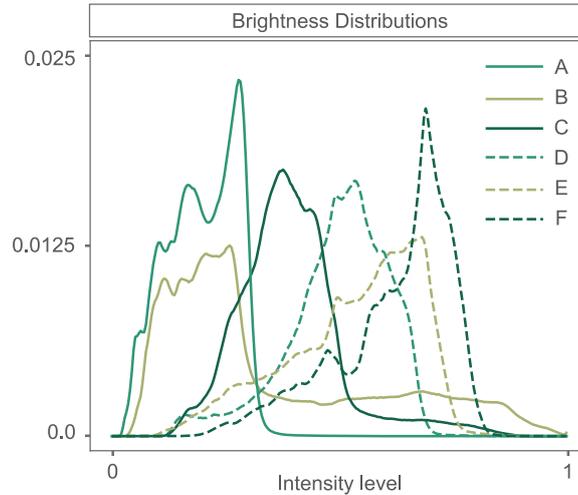
**Figure 4.2:** Visual representation of tonal manipulations in Figure 4.1 From left to right we show: the unenhanced input image, a brighter output, a more contrasted output and a brighter output with more contrast.



**Figure 4.3:** An selection of unenhanced benthic images

brightness histogram. Thus the combination of these effects, driven by tone map 3, is a brighter output image with greater contrast. This is indicated by a histogram that is both flatter than the input and skewed towards a higher pixel intensity. We show the results of applying these three tone curves in terms of a real input image in Figure 4.2. From left to right we show the input image and the modified output following application of tone curves 1-3. The intuitive results seen in the output histograms (1 is brighter, 2 is darker with more contrast and 3 is brighter with more contrast) are evident in these tone-mapped images.

Although the three TMOs are fairly standard and simple enhancements that can improve the visual quality of imagery, their usefulness is dependant on the underlying data. Clearly, if an image is already bright, we probably wouldn't want to make the image even brighter, for example through the application of tone curve 1. In Figure 4.3 we show six unenhanced underwater images. Viewing their six brightness histograms in Figure 4.4 we see that there is significant variation between their distributions. There is variation in the degree of skewness and flatness, as well as variation in modality - though most appear to be unimodal, gaussian-



**Figure 4.4:** Corresponding brightness distributions of unenhanced images shown in Figure 4.3

like distributions. The different histograms, indirectly, point both to how the content and illumination of underwater scenes vary across images.

It is difficult to therefore suggest a *fixed* tone-mapping operator that will provide an optimal, or at least highly beneficial, enhancement in all cases. Additionally, the data collected from underwater imagery, and its usage, is highly diverse and thus this may influence the kind of tonal enhancement that is required.

Of course, the obvious ‘solution’ to this problem is to use a tone-map that is tied to the brightness distribution itself. For example., Histogram Equalization (HE), its various adaptive and contrast-limited versions (138; 214; 75) and others (91; 59) have been widely used to automatically map an input image to an output that has more visible detail (or, equivalently, a flatter histogram). Often these approaches work well but not always. Part of the reason for failure, is that the tone-map objective function is expressed in the language of signal processing (here ‘maximizing information’), which may or may not fit with the adjustments sought for a particular task. General algorithms do not consider the purpose the imagery serves (and its audience).

If a fixed or auto-generated tone curve is not fit for task, a third alternative is ask a user to make the tone-curve. Making a bespoke tonal enhancement may be more useful to end-users, though at the expense of time. It gives more flexibility which is highly suitable for the diverse appearance of underwater images and better supports the subjective preferences of end-users to aid their analyses. Advantageously, investigating these tonal enhancement may also serve to guide future development and automation of underwater tone-mapping and/or image enhancement algorithms. That is, given a corpus of bespoke adjustments made by users in enhancing underwater images it is, perhaps, possible to learn what adjustments are made for a given image. This learned behaviour effectively implements a domain-specific (here, underwater images) tone-adjustment algorithm.

In, this chapter, we focus on the tone mapping problem in the context of underwater imagery, with the long-term aim of developing automated image enhancement tools. We began by investigating images processed by end-users in the field *i.e. biologists*, who manually manipulated input images so that details sought for analytical purposes were more conspicuous and did not contain artefacts. We describe characteristics of unenhanced input histograms and their user-adjusted outputs to guide development of more tailored image tonal manipulations. Finally, we explore their characteristics and possibility of automation by comparing them to known probability distributions.

## 4.2 Background

### 4.2.1 Desirable brightness distributions

As described in Chapter 2, tone mapping can be framed as as simply mapping an input image with a brightness histogram  $h(b)$  to an output image with a desired target brightness distribution  $h_{targ}(b)$ . For underwater images, particularly those in deeper waters, it is important to preserve the fact that it is dark, as well as image shadows (which can aid object identification). When designing a bespoke tone map  $t$ , if the target distribution  $h_{targ}(b)$  - to which this tone map adjusts an input image- does not tail toward zero, then the processed

image will be over-enhanced and dark noisy pixels may become apparent. Additionally, as images are often dominated by a strong ‘spot-light’ in the center and floating particulates, there are often many extremely bright pixels which should be reduced in intensity.

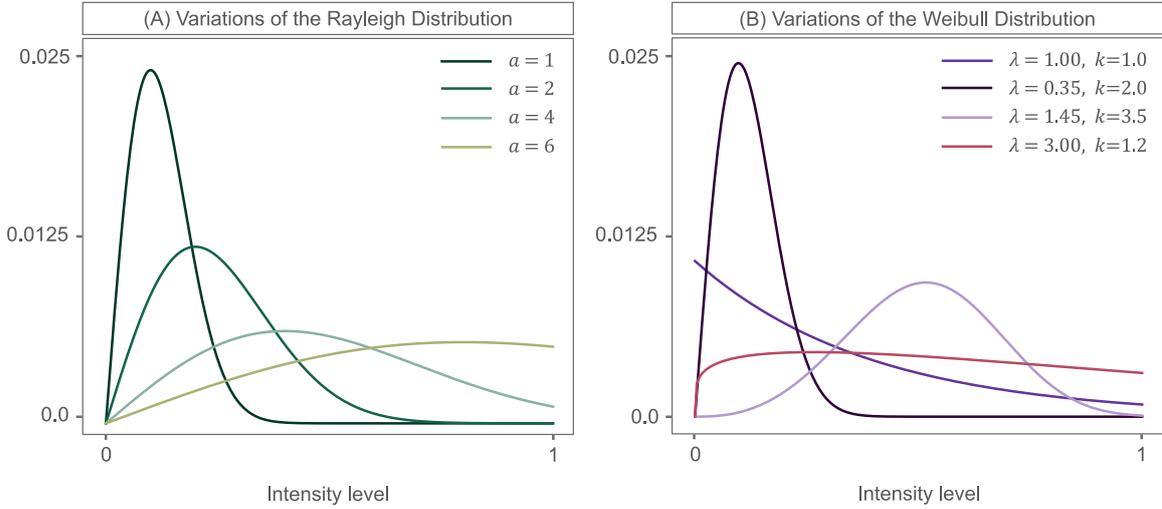
For these reasons, it has been suggested that a Rayleigh distribution (143; 93) is a favourable target distribution for images (48). This is a continuous probability distribution for positive-valued random variables (127), often resembling a bell-shape, and has been frequently enforced within underwater image enhancement, particularly as an adaptation of the CLAHE algorithm (48; 75; 59; 60). The RD  $h_R$  is defined as:

$$h_R(b; a) = \frac{b}{a^2} e^{-b^2/(2a^2)}, \quad b \geq 0, \quad a > 0, \quad (4.1)$$

where  $b$  is the brightness value and  $a$  the scale parameter; broadly corresponding to the peak of the histogram.

In Figure 4.5 (A) we show variations of the Rayleigh distribution. Here we see that the Rayleigh distribution, in this case when  $a = 4$ , can preserve the darkness of pixels that should not be enhanced as well as bringing back details that are compressed within the spot illumination. That is to say that this ‘target’ Rayleigh distribution has brightness values mildly concentrated in the middle of the brightness range, but there is still coverage in the bright and dark areas. In the adaptation of CLAHE (48), the input image brightness values are modified so that the processed image histogram is moved toward the Rayleigh distribution.

The Weibull probability distribution (WD) (194; 93) exhibits similarities to the RD and as one of the contributions of this chapter, we propose it has desirable properties that make it a good target distribution for underwater images. Although not yet used for this purpose, to our knowledge, it has been demonstrated that the WD can explain the contrast statistics



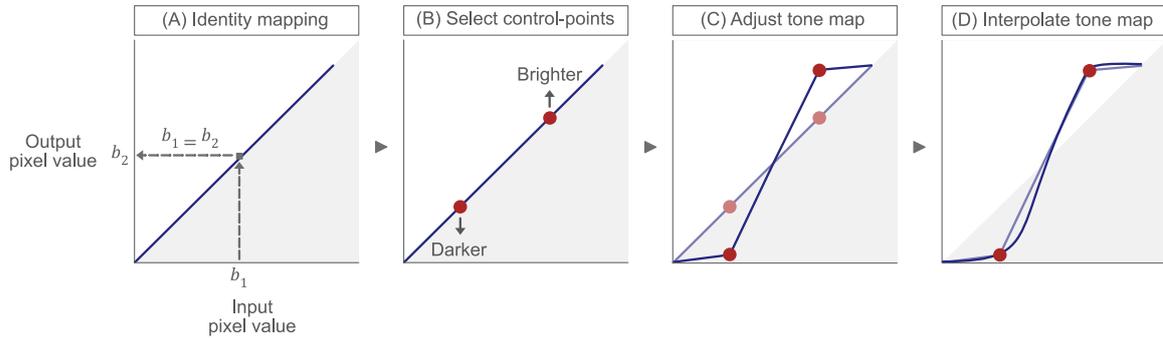
**Figure 4.5:** Variations of the Rayleigh & Weibull distribution

of natural images (57; 208) and is correlated with our own perception of natural images (159). In its two-parameter form, the Weibull Distribution  $h_W$  is denoted as:

$$h_W(b; \lambda, k) = \frac{k}{\lambda} \left(\frac{b}{\lambda}\right)^{k-1} e^{-(b/\lambda)^k}, \quad b \geq 0, \quad \lambda > 0, \quad k > 0, \quad (4.2)$$

where  $b$  is the brightness value,  $\lambda$  is the scale parameter and  $k$  is the shape parameter. The Weibull parameters  $\lambda$  and  $k$ , broadly account for the peak position and the slope or spread of the distribution, respectively. The WD has the benefit of yielding a RD, when  $k = 2$ , yet is far more flexible as shown by Figure 4.5 (B). From this we see that the 2-parameter function can resolve an extensive range of distributions, including those equivalent to the exponential distribution when  $k = 1$ , approaching the normal distribution when  $k \sim 3.5$  and informally appearing *like* the uniform distribution when  $\lambda = 3$  &  $k = 1.2$ .

Our experiments in this chapter will describe the characteristics of domain-experts' desirable target distributions and the extent to which these align with previous assumptions of ideal, or desirable, target distributions in underwater images i.e. the Rayleigh distribution. Given the relatively limited variation in properties the Rayleigh Distribution can manifest, and the



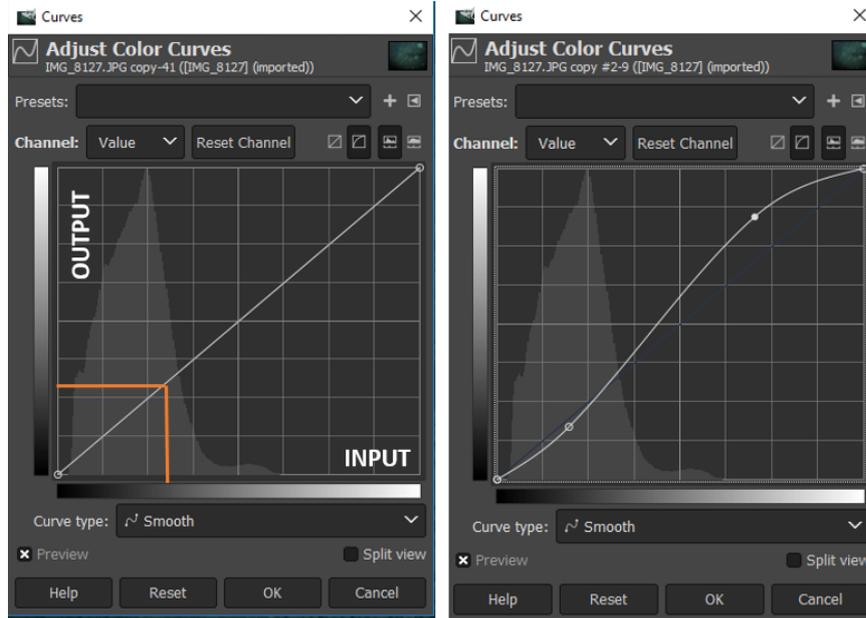
**Figure 4.6:** A diagram of tone map creation and interactive manipulation

high degree of freedom domain-experts had in their tonal manipulations, we also explore the potential of the more general Weibull Distribution as a suitable target distribution.

#### 4.2.2 Control-point tonal manipulations

Creating a bespoke tone mapping operation can be achieved using photo-editing tools such as Photoshop or GIMP. In this example, a user can, effectively, make their own input to output (brightness) tone curve, see Figure 4.6. At the start, a user is presented with an input to output 'identity' mapping; simply a graph with a line at 45 degrees, depicted in (A). This linear tone curve means that an input brightness, of say  $b_1$ , maps to an equivalent output brightness,  $b_2$  i.e.  $b_1 = b_2$ . This is true for all brightness values in this case. A user can then select a control-point, as shown in (B), and move it up or down. Movement upwards causes brightness values to increase, or become brighter, whereas movement downwards will cause values to reduce, becoming darker. Multiple control-points might be chosen by a user to alter the curve shape (C). Finally, a smooth interpolation may be made between the adjusted control points to define the overall tone adjustment, shown here in (D). Importantly, as the user moves the control-point(s) they can simultaneously view its effect on the output image. The user stops making adjustments when they are happy with the output result.

In this chapter, GIMP was the chosen platform for creating bespoke tone-maps. It exhibits similar features to photoshop, but with the benefit of being an open-source image manipulation programme. GIMP contains a tone-mapping tool, *Curves*, in which the image tone



**Figure 4.7:** GIMP tone-mapping tool: original ‘identity mapping’ (left) and adjusted tone map (right)

curve can be manipulated to map input values to adjusted output values, see Figure 4.7. This process is identical to that described in Figure 4.6 and is both simple and intuitive for the end-user. Analysts simply modify the input to output ‘identity’ mapping from a line at 45 degrees to a tone map of their choosing, whilst simultaneously viewing its effect on the now tonally-adjusted output image.

The GIMP curve tool allows users to create a brightness tone-mapping  $t()$ , where brightness is defined as *value* in the HSV colour space (64) or the  $\max(R, G, B)$  of an RGB input image  $\underline{I}_{in}(x, y)$ , see Chapter 2. GIMP then calculates the enhanced colour output image  $\underline{I}_{out}(x, y)$  as

$$\underline{I}_{out}(x, y) = t(\underline{I}_{in}(x, y)) \quad (4.3)$$

Applying the same tone map to each colour channel greatly simplifies the task of tonal adjustment. The analyst constructs only one curve, avoiding the need to carefully manipu-

late each colour channel. This approach has the benefit of preserving the hue, or dominant colours in an image, whilst allowing the analysts to enhance the saturation or chroma, improving the vibrancy of colours. This can be particularly helpful as underwater images often suffer from colour loss due to wavelength attenuation.

## 4.3 Experiments

### 4.3.1 Domain-expert tone mapping

The following experiment involved the development of custom-per-image tone maps by analysts, to determine the type of enhancements that will aid their annotation efforts. Analysts were asked to, manually, tonally adjust images, for the specific purpose of maximising the conspicuousness of details required to annotate the content of the image (i.e. the habitat).

It is important to acknowledge that our assumptions of desirable output images in this chapter are based on 3 analysts who were available to undertake the experiment. Given the subjective nature of image enhancement, desirable tonal characteristics described may thus not be optimised to other analysts. However, each of the 3 analysts are highly experienced in benthic image interpretation and have therefore been exposed to a range of image datasets that vary in content and quality. Their responses will thus provide a useful foundation for explorations of image enhancement in a broad sense, since they have a good grasp of the image characteristics that will support accurate identification more widely. In addition, in the remaining thesis experiments we explore desirable tonal enhancements more extensively, with additional analysts, to better capture the subjective variation amongst analysts and distinguish patterns.

#### Setup

For the experiment, all analysts enhanced the same dataset - a random selection of 60 images from Dataset 1, see Chapter 3. This image sample covered each of the 6 broad habitat classes

(10 of each habitat), representing the breadth of biological and physical features expected in the dataset, see Table 3.2 & Figure 3.1. This data subset was intentionally small as we use it for pairwise preference evaluations in Chapter 5, a time-consuming process for experiment participants.

To accurately track alterations to pixel values and extract analyst tone curves, a strip of pixels in the bottom left corner of each image was replaced with known values. This strip contained  $10 \times 10$  pixel blocks according to a RGB colour space gradient  $[0, 255]$ . This created a pixel strip of  $10 \times 2560$ , as demonstrated by Equation 4.4:

$$S = \left[ P_0 \mid P_1 \mid \cdots \mid P_{255} \right] \quad (4.4)$$

where  $P$  is a  $10 \times 10$  matrix of the same value (depicted by the subscript) and  $S$ , the augmented matrix, represents the pixel strip. This known  $10 \times 2560$  pixel gradient was applied to each colour channel in the RGB images. Due to potential compression-induced pixel alterations that may arise after exporting adjusted images from GIMP, a  $4 \times 4$  sub-block was extracted from the centre of each  $10 \times 10$  pixel block. These blocks were then collapsed into a  $16 \times 256$  array and the mean extracted to produce the recreated  $1 \times 256$  tone curve.

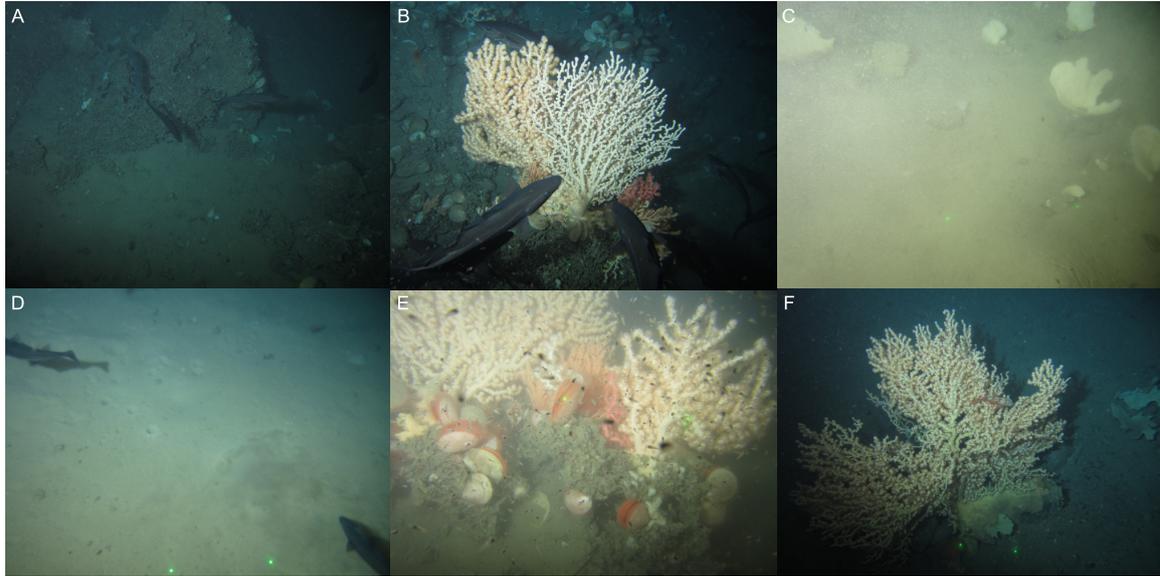
Prior to the experiment, annotators were provided with training. This explained the aims of the project, tone curves and their manipulation, image histograms and how to use the tone-mapping tool. Training involved a small presentation demonstrating the features of the curves tool and then a demonstration of its usage and effects of various kinds of tonal adjustments. This information was also shared with annotators via a background document for reference purposes. Analysts had complete freedom in the design of their tone maps, the only stipulation was that the tone maps must be monotonically-increasing functions to prevent introduction of image artefacts and undesirable effects. To gain confidence with the usage of the software and tone curve manipulation before the experiment took place,

annotators were provided with practice images. These images were not present with the experimental dataset.

All annotators performed adjustments under ISO standard 3664:2009 viewing conditions (92); sitting approximately 70 cm from the display in a neutrally painted and darkened room. A duration of 1-2 minutes, per image adjustment was suggested (though not enforced) to participants. This was considered sufficient to find a suitable enhancement and would prevent annotators from over-analyzing their tonal enhancements. Additionally, the short time-frame of the experiment would help to reduce inconsistent performance linked to fatigue. The image order was randomised for each analyst to randomize any effects of fatigue and variable concentration and prevent bias in adjustments linked to the order of the images.

## **Results**

Underwater images used in the experiment suffered from the typical optical challenges present in many underwater datasets. As shown in Figure 4.8, they were typically dark with a strong spot-light in the center of each image - particularly when the camera platform was at an increased height above the seafloor, to avoid contact with reef for example. Halo and vignetting effects were associated with the strong artificial lighting employed and many images were poorly contrasted and lacking distinct colouration. Identification of the habitat from these images was further complicated by high turbidity, due to the presence of particulate matter in the water column, and high abundance of small pelagic organisms such as Krill.

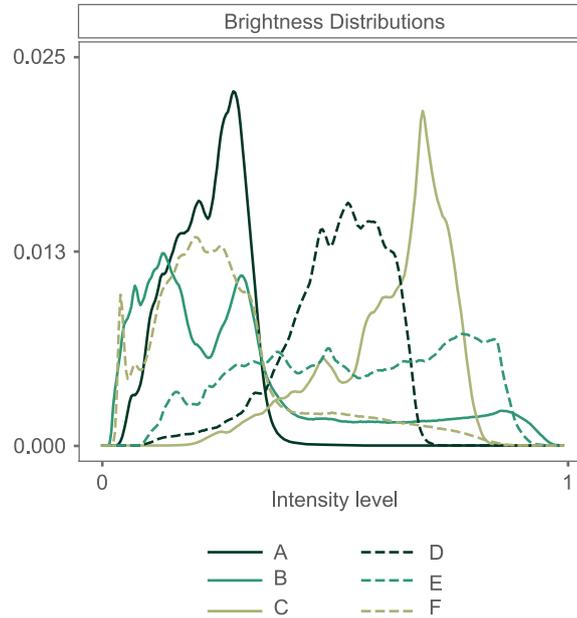


**Figure 4.8:** Image selection to represent optical challenges present in the experiment dataset

These illumination effects were evident within the underlying brightness distributions, as shown in Figure's 4.9 & 4.10. These were typically *gaussian-like* (bell-shaped), broadly unimodal, skewed and poorly contrasted. Thus enhancements to benefit such distributions could simply include increasing contrast and adjusting brightness.

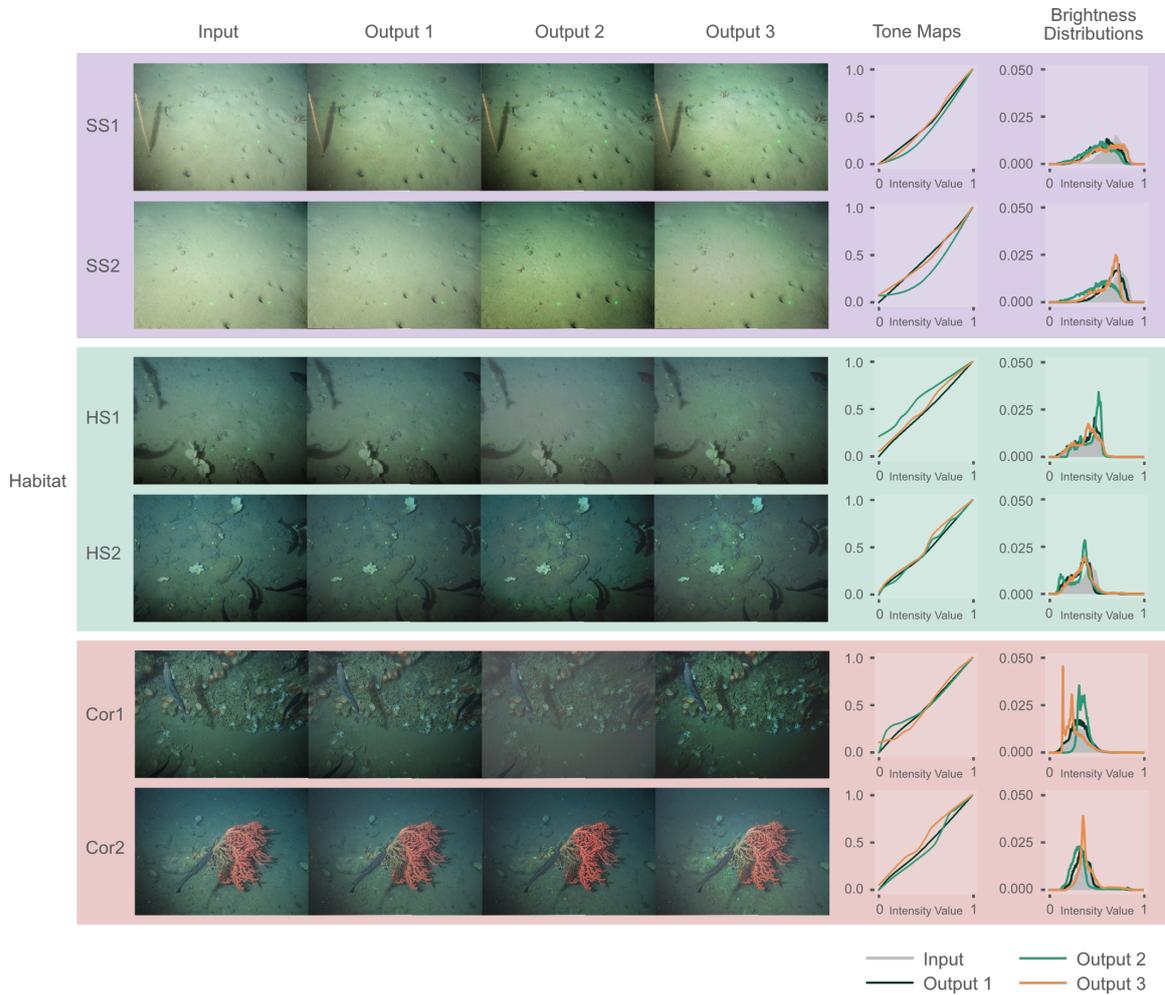
The experiment resulted in the production of 180 tonal enhancements (60 images  $\times$  3 analysts). A selection of these are shown in Figure 4.10. Here we include the input and tonally-adjusted output images as well as the corresponding tone maps. Outputs are numbered according to the analyst who produced the tonal manipulation. We found that generally, the shape of tone curves was variable, however a large proportion were mildly *wavy* and linear in design. Others appeared as stretched S-shaped curves, indicating more gentle contrast enhancement and some more exponential-like in shape, causing some compression of the mid-tones.

Contrast and brightness of a brightness image ( $L(x, y)$ ) can be measured simply by calculating the standard deviation and mean, respectively (193). Comparison of the input



**Figure 4.9:** Brightness histograms of images shown in Figure 4.8

and output brightness images revealed that analysts increased contrast in 51% of the output images. This proportion was also relatively consistent with individual analysts, who enhanced contrast in between 53-63% of images. Increasing contrast would have allowed analysts to improve seafloor texture visualization, highlighting appearance of burrows, bioturbation, small seabed depressions and sand ripples as well as acknowledging presence of hard substrata such as gravels and pebbles. Interestingly, for 79% of the dataset, analysts chose to decrease brightness. This may have helped to lessen the intensity of the light cone and halo-effect, as well as reduce the saturation of bright seafloor organisms such as corals. However, increasing brightness would have helped analysts fully visualize the content of the image by eliminating the dominance of the darkest pixel intensities, particularly present around the image exterior. Instead analysts often opted to shrink the dynamic range in some cases, indicated by a clipped tone map that no longer stretches between 0 & 1, for example Output 2 for habitat class Cor1 in Figure 4.10. Shrinking the dynamic range will reduce contrast, however it can also be used to reduce the intensity of stark bright and dark regions. This is particular useful when images are dominated by bright features such as



**Figure 4.10:** Selection of input images (1 per habitat class), bespoke tonal manipulations and their visual effect on output images

corals in the image foreground and surrounded by a very dark background. Reducing these intensities would better reveal obscure features.

It is interesting to note that tonal manipulations did not always adhere to what would be considered a 'natural' appearance. For such an application, enhancements need not be aesthetically pleasing - nor conform to general standards of tonal enhancement such as improved contrast - only that they allow the user to better identify cryptic features that enable identification of the habitat. This may explain the fact that a large number of tonal manipulations reduced contrast and brightness. Enhanced images also appeared

unusual due to the boosting of colours through GIMP’s application of the brightness tone map to each colour channel. In the event that colour is a distinguishing feature, this chromatic method could support discrimination of marine features. However given that colour constancy and fidelity in underwater imagery is unreliable (7), it is important to apply caution when using colour as a discriminating feature, both before and after colour enhancement.

### 4.3.2 Modelling image histograms

The domain-expert tone mapping experiment revealed a variety of characteristics suitable to improve the identification of marine habitats from imagery. As expected, it also demonstrated the subjective nature of user tonal enhancements. In this experiment, we further explore the behaviour of the tonal manipulations and their influence on the underlying brightness distributions. We show that these properties can be described mathematically, and as such provide a foundation for development of a tailored image enhancement tool for end-users.

#### Setup

Given the previous recommendations of the Rayleigh distribution for underwater imagery and the importance of the Weibull distribution for summarizing natural images, we compare them both to the input and output (analyst-enhanced) brightness distributions. In this experiment we determine the degree of conformity between the compared distributions to help to determine the extent to which enhanced brightness distributions can be parameterised. This will also help to build on previous work, summarising natural image statistics, with a specific focus on underwater images.

The brightness distribution  $h(b)$  of each image  $\underline{I}(x, y)$  was extracted from their corresponding  $\max(R, G, B)$  colour space using  $\text{hist}(L(x, y))$ , where  $L(x, y) = \max(\underline{I}(x, y))$ . To determine their similarity to the Rayleigh & Weibull distributions, we simply needed to find a proxy histogram  $h'(b)$  of each brightness histogram such that  $h'(b) \approx h(b)$ . Here

the proxy histogram refers to the *closest* Rayleigh  $h' = h_R(b; a)$  or Weibull distribution  $h' = h_W(b; \lambda, k)$ . This was found by finding the parameters of each distribution,  $a$  and  $\lambda$  and  $k$  respectively, that minimizes the Kullback-Leibler (KL) divergence - as a measure of closeness. The KL-divergence is a probabilistic measure of the difference between two distributions, in this case  $h(b)$  and its proxy  $h'(b)$ . It is calculated as follows:

$$\min \int_0^1 h(b) \log\left(\frac{h(b)}{h'(b)}\right) db \quad (4.5)$$

Also known as information divergence or relative entropy, if the two distributions are highly similar then the KL-divergence will be low, with 0 reached only when  $h(b)=h'(b)$ . Remembering that the proxy histogram  $h'(b)$  refers to either a Rayleigh or Weibull distribution in this case, we thus seek the parameters for each probability distribution that return the lowest KL-divergence. We approach this problem discretely, searching the parameter  $a \in \{0, 0.1, 0.2, \dots, 8\}$  for the RD and parameter pairs of  $\lambda \in \{0, 0.1, 0.2, \dots, 3\}$  and  $k \in \{0, 0.1, 0.2, \dots, 15\}$  for the WD. The KL-divergence was therefore also used for comparative purposes to quantify which of the probability distributions, Rayleigh or Weibull, best described the observed data.

## Results

Despite the prior recommendations for underwater brightness distributions to follow a Rayleigh distribution, we found that the RD poorly captures the brightness characteristics of underwater images, with an average KL-divergence of 0.36 ( $\pm 0.02$ ) between the dataset brightness distributions and their *closest* Rayleigh distribution, see Table 4.1. Instead we find that brightness properties are significantly better described by the Weibull Distribution, with an average KL-divergence of 0.11 ( $\pm 0.01$ ). In Figure 4.11 we show that this also holds true for images regardless of (1) whether they are an input or enhanced output and (2) the habitat.

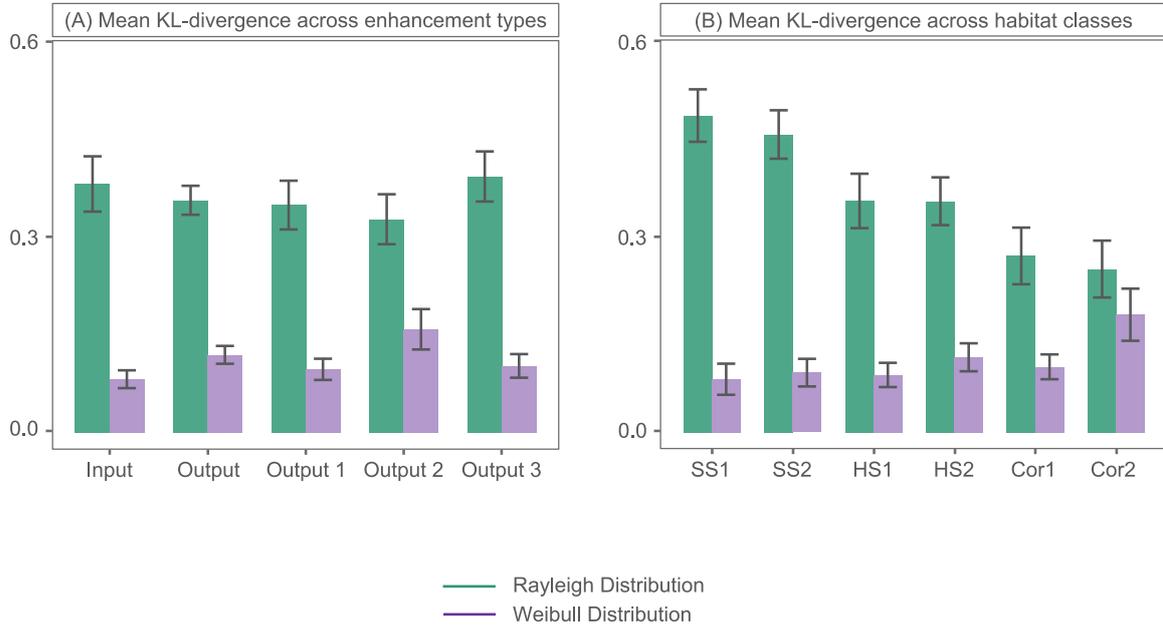
**Table 4.1:** Summary statistics for KL-Divergence between the dataset brightness distributions (input and outputs) and their best Rayleigh & Weibull approximation

Type	Mean	SD	Min.	Median	99 <sup>th</sup> p.	Mean difference
<b>Rayleigh</b>	0.36 ( $\pm 0.02$ )	0.16	0.04	0.35	0.72	0.25 ( $\pm 0.02$ )
<b>Weibull</b>	0.11 ( $\pm 0.01$ )	0.09	0.01	0.09	0.49	

<sup>1</sup>95% confidence intervals are shown in parentheses <sup>2</sup>SD = Standard Deviation <sup>3</sup>p. = Percentile

In Figure 4.11 (A), KL-divergence is grouped according to enhancement type i.e. input images in which no-enhancement is applied, output images for each of the 3 analysts and the combined analyst outputs (1, 2 3). The lowest mean KL-divergence was only 0.08 ( $\pm 0.01$ ) for the Weibull fit of input distributions, and increased slightly when approximated their adjusted outputs, scoring 0.12 ( $\pm 0.01$ ). Considering the individual analysts, we see that this reduced fit was largely driven by analyst 2, scoring 0.16 ( $\pm 0.03$ ) on average. However, this was still an improvement on the Rayleigh fit, achieving a divergence  $\sim 50\%$  better. In Table 4.2 we further summarize the KL-divergence scores across the dataset. Calculating the mean difference in KL-divergence of the two probability distributions, we see that the RD was consistently worse than the WD, across input and output distributions, scoring between 0.17-0.3 ( $\mu = \sim 0.23$ ,  $\mu = \text{mean}$ ) higher. This inability to adequately describe brightness distributions is further demonstrated by the absence of Rayleigh properties within the best Weibull approximations i.e. when  $k = 2$ . Only 1/240 Weibull models simulated a Rayleigh distribution with  $k = 2$ .

Not only does the WD adequately mimic the intrinsic and adjusted properties of the underwater images, we see that it is also generally not influenced by the content (habitat) of the image, see Figure 4.11 (B). KL-divergence in this case was grouped based on the expert-assigned habitat. Note that we show this for output images only. No relationship was observed between KL-divergence in input images and their associated habitat class, for both the Rayleigh and Weibull fits. In Figure 4.11 (B), mean KL-divergence is shown to be relatively consistent across habitat classes, always scoring  $< 0.12$  for Weibull, excluding Cor2. Cor2 habitat's occurrence across either boulder areas with predominance of soft



**Figure 4.11:** Mean KL-Divergence between the dataset brightness distributions and their best Rayleigh & Weibull approximation. In (A), results are displayed according to enhancement type. Note that Output = Outputs 1, 2 & 3. In (B) mean KL-divergence (across output images) is viewed according to habitat class. Error bars represent 95% confidence intervals.

substrate, as well as predominant hard substrate or reef substrate regions, means there is likely greater variation within the brightness distributions themselves. The subsequent fit of brightness statistics was significantly poorer (though still generally good) at 0.18. This directly contradicts the behaviour of the RD in which clear differences were noted between broad habitat levels. Rayleigh approximations performed worst for Soft Substrate (SS) habitats followed by Hard Substrate (HS) and Reef (Cor). Thus the suitability of the RD to summarize enhanced output distributions appeared to be highly influenced by image content and the subsequent properties analysts enforce to better expose image details. In Table 4.3 we summarize performance of the approximations according to the image content. As with comparisons based on enhancement type (Table 4.2), large differences in KL-divergence, between Rayleigh and Weibull fits, were consistently present across groups. The largest mean difference, between the quality of two distribution fits was observed within SS habitats ( $\mu = 0.39$ ) and the smallest within reef habitats ( $\mu = 0.12$ ).

**Table 4.2:** Summary statistics for KL-Divergence, according to enhancement type, between the dataset brightness distributions and their best Rayleigh & Weibull approximation

Type	Mean	SD	Min.	Median	99 <sup>th</sup> p.	Mean difference
<b>Input<sub>R</sub></b>	0.38 (± 0.04)	0.17	0.05	0.38	0.65	0.3 (± 0.05)
<b>Input<sub>W</sub></b>	0.08 (± 0.01)	0.05	0.01	0.07	0.29	
<b>Output<sub>R</sub></b>	0.36 (± 0.02)	0.15	0.04	0.33	0.73	0.24 (± 0.02)
<b>Output<sub>W</sub></b>	0.12 (± 0.01)	0.09	0.01	0.1	0.51	
<b>Output1<sub>R</sub></b>	0.35 (± 0.04)	0.15	0.08	0.35	0.72	0.25 (± 0.04)
<b>Output1<sub>W</sub></b>	0.1 (± 0.02)	0.06	0.03	0.09	0.37	
<b>Output2<sub>R</sub></b>	0.33 (± 0.04)	0.15	0.04	0.28	0.8	0.17 (± 0.03)
<b>Output2<sub>W</sub></b>	0.16 (± 0.03)	0.12	0.01	0.12	0.56	
<b>Output3<sub>R</sub></b>	0.39 (± 0.04)	0.15	0.11	0.38	0.71	0.17 (± 0.03)
<b>Output3<sub>W</sub></b>	0.1 (± 0.02)	0.07	0.01	0.08	0.43	

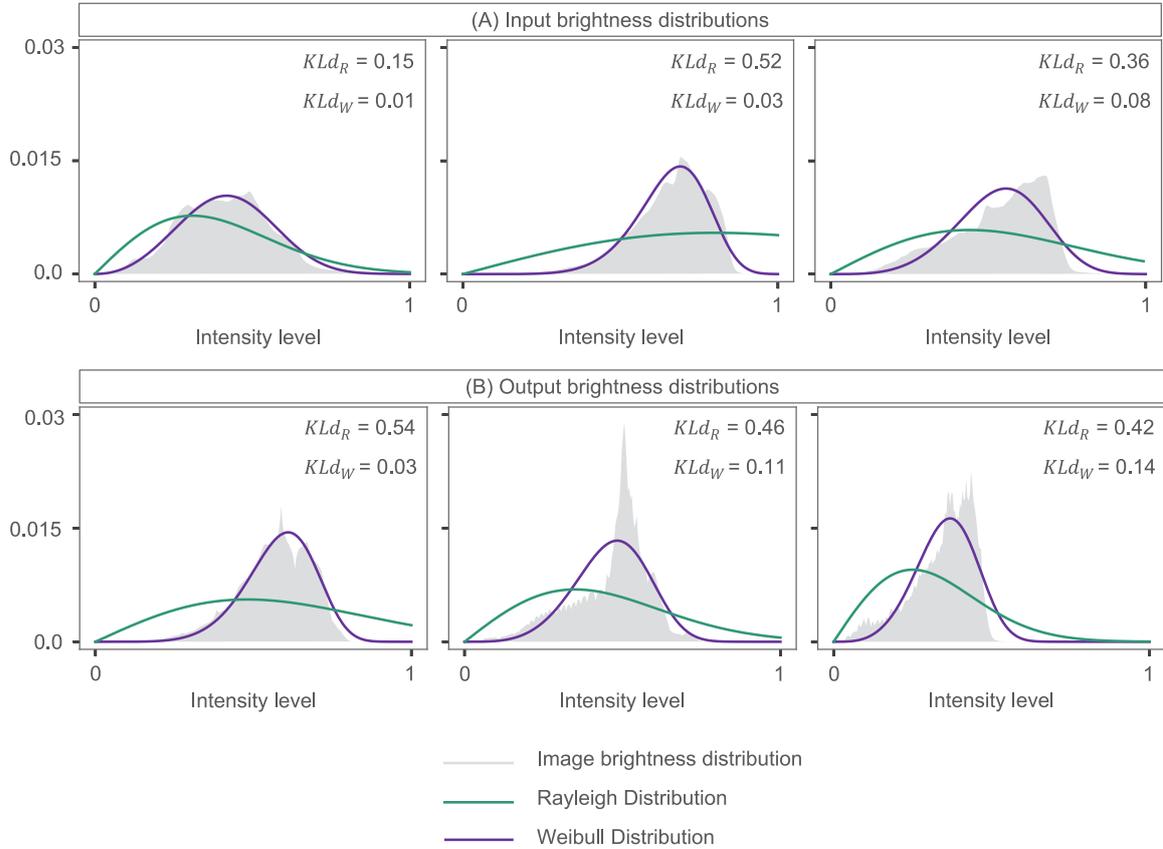
<sup>1</sup>Numeric subscript in Type refers to the probability distribution <sup>2</sup>95% confidence intervals are shown in parentheses <sup>3</sup>SD = Standard Deviation <sup>4</sup>Output = Outputs 1,2 & 3 <sup>5</sup>p. = Percentile

**Table 4.3:** Summary statistics for KL-Divergence, according to habitat class, between the dataset brightness distributions and their best Rayleigh & Weibull approximation.

Type	Mean	SD	Min.	Median	99 <sup>th</sup> p.	Mean difference
<b>SS1<sub>R</sub></b>	0.49 (± 0.04)	0.13	0.22	0.52	0.72	0.4 (± 0.04)
<b>SS1<sub>W</sub></b>	0.08 (± 0.02)	0.08	0.01	0.07	0.49	
<b>SS2<sub>R</sub></b>	0.46 (± 0.04)	0.12	0.26	0.46	0.67	0.37 (± 0.05)
<b>SS2<sub>W</sub></b>	0.09 (± 0.02)	0.07	0.03	0.06	0.36	
<b>HS1<sub>R</sub></b>	0.35 (± 0.04)	0.13	0.16	0.32	0.73	0.27 (± 0.05)
<b>HS1<sub>W</sub></b>	0.09 (± 0.02)	0.06	0.01	0.08	0.32	
<b>HS2<sub>R</sub></b>	0.35 (± 0.04)	0.12	0.14	0.35	0.71	0.24 (± 0.05)
<b>HS2<sub>W</sub></b>	0.12 (± 0.02)	0.07	0.03	0.11	0.31	
<b>Cor1<sub>R</sub></b>	0.27 (± 0.04)	0.14	0.1	0.26	0.8	0.17 (± 0.04)
<b>Cor1<sub>W</sub></b>	0.1 (± 0.02)	0.06	0.01	0.09	0.24	
<b>Cor2<sub>R</sub></b>	0.25 (± 0.04)	0.14	0.04	0.22	0.61	0.07 (± 0.02)
<b>Cor2<sub>W</sub></b>	0.18 (± 0.04)	0.13	0.04	0.14	0.56	

<sup>1</sup>Numeric subscript in Type refers to the probability distribution <sup>2</sup>95% confidence intervals are shown in parentheses <sup>3</sup>SD = Standard Deviation <sup>4</sup>p. = Percentile

Despite analysts having complete freedom in their tonal manipulations, the alterations are captured sufficiently well by the 2-parameter WD. In Figure 4.12 we show example distri-



**Figure 4.12:** Some example brightness distributions and their most similar Rayleigh and Weibull Distribution for (A) input images and (B) analyst-adjusted output images. KL-Divergence scores, denoted here as  $KLd$ , are presented in each case.

butions and their associated WD. This demonstrates the close approximations of Weibull across some input and output distributions and their associated KL-divergence. Both the skewness (or peak location) and thus the brightness of the images is accurately described as well as the spread of the distribution which encapsulates the contrast statistics. For comparison, we also include the corresponding Rayleigh fits. This demonstrates that the 1-parameter RD, does not possess the flexibility to adequately capture the skewness and kurtosis of the brightness distributions. In these cases the peak location of the RD is often offset to the left of the true peak location, translating to a darker image. The RD also simulates images with amplified contrast, shown by the spread and flatness of the distributions; a contrast far greater than that enforced by the analysts for example.

The WD is effective at representing the brightness and contrast statistics of underwater images with an even smoother and simpler function. Given these elements are the exact properties that are modified in tone mapping, there is thus scope to utilise the WD as enhancement tool.

## 4.4 Conclusion

In this chapter, we have described the custom tonal manipulations by domain experts, demonstrating that although varied, these are simple in design and enforce brightness properties that follow a Weibull Distribution. Contrary to previous knowledge, we show that the Rayleigh distribution is not suitable to model the underlying brightness characteristics of underwater images, nor does it possess properties that align with the needs of image analysts for interpreting imagery. The suitability of the WD is generally dependable despite variation in image content. It thus offers a mechanism through which to explore targeted image enhancement, with the aim of providing a tool to support underwater image annotation.

In the next chapter we will develop this work further, with this aim in mind. We will show how the Weibull distribution can be used enhance underwater imagery and improve upon bespoke tonal manipulations, exploring how its usage can best support the analyses of end-users.

# 5 Weibull Tone Mapping (WTM): for enhancing underwater imagery

In Chapter 4 we described the bespoke tonal manipulations by analysts to support the interpretation of underwater imagery, finding their brightness statistics to exhibit behaviour that can be summarized by the Weibull Distribution. In this chapter we build on this knowledge to introduce a tone mapping algorithm, Weibull Tone Mapping (WTM). This algorithm is designed, by construction, to approximate user adjustments but result in smoother and simpler tonal manipulations that are preferred by image analysts. We also show how the method by which we map the colour aspect of images (given a brightness only tonal adjustment) has a significant impact on users' subjective preferences.

## 5.1 Introduction

Adjusting images with control-point tonal manipulations allows domain-experts to more precisely highlight features of interest to aid interpretation of marine imagery. As demonstrated in Chapter 4, these enhancements often conform to previous broad assumptions of what are considered beneficial tonal enhancements such as increasing contrast in many cases. However they also demonstrated that analysts are willing to enforce tonal manipulations at the expense of aesthetics, should this aid them in their visual assessments. This for example could include reducing contrast, dynamic range and most notably, brightness.

Significantly, Chapter 4 found that the majority of these modified brightness qualities, as well as their original intrinsic properties, can be well captured by the Weibull distribution.

The 2-parameter WD is a highly general function and can recreate the shape of many histograms but with a far smoother representation.

Since the input and output brightness distributions of underwater images are generally like the WD, it is logical to presume that the mapping that converts the input WD to an output WD will be similar to the tone map defined by analysts. Application of this tone map to an unenhanced image should therefore mimic the adjustments made by analysts. In this chapter, we confirm this very hypothesis, validated by objective image quality assessments and subjective evaluations (with domain experts). This demonstrates that the process to generate tone maps can be optimized as WTM offers a faster and easy method to achieve the same enhancement. In addition, since the WD is a continuous smooth function (though handled discretely in this work) this mapping should have the additional benefit of being smoother than the tone map that the analysts themselves applied. This would mimic the adjustments made by analysts whilst limiting the introduction of image artefacts. For example, abrupt shifts between brightness tones could lead to posterization and banding effects as well as a loss of fine details and texture, causing the image to appear unnatural and flat. These benefits of WTM are crucial in image analysis pipelines to improve efficiency and offset labour demands whilst enhancing the quality of data acquired.

## 5.2 Background

In this chapter we present a method, Weibull Tone Mapping, which can approximate the bespoke tonal manipulations of end users. To demonstrate the success of these approximations we use both objective and subjective evaluation methods. Objective methods involve using metrics with explicit numerical comparisons that aim to mimic the quality assessments of a human observer (148; 114; 193; 189; 111) whereas subjective methods are based on human quality assessment (90). For reviews of underwater image quality assessments, see (112; 110). In this work, we use a mixture of objective and subjective evaluations; these are introduced in the following sections.

### 5.2.1 Evaluating image similarity

When approximating, or evaluating the distortion between, images a variety of objective metrics have been employed to assess the success or performance of these operations. CIELAB Delta E (or  $\Delta E_{ab}^*$ ) (148), is a commonly employed metric and is used as a measure of colour difference. Given two RGB images,  $A$  &  $B$ , with which we wish to measure similarity,  $\Delta E_{ab}^*$  involves first converting each to their corresponding CIELAB (CIE 1976  $L^*a^*b^*$ ) image, where at each pixel they have a  $L^*$ ,  $a^*$  and  $b^*$  triplet. As discussed in 2.2.1, in the CIELAB colourspace,  $L^*$  represents brightness whilst  $a^*$  and  $b^*$  encode the chromatic aspects of an image (145). Denoting dependence on images  $A$  &  $B$ , using the subscripts  $A$  and  $B$ , respectively, the  $\Delta E_{ab}^*$  difference for one pair of corresponding pixels is calculated as:

$$\Delta E_{ab}^* = \sqrt{(L_A^* - L_B^*)^2 + (a_A^* - a_B^*)^2 + (b_A^* - b_B^*)^2} \quad (5.1)$$

Significantly, CIELAB was designed to be a perceptually uniform space that correlates with perceived color difference (145). In that regard a  $\Delta E_{ab}^*$  of approximately 1 coincides with a *just noticeable difference*. In terms of images, if the average  $\Delta E_{ab}^*$  calculated between images (across all pixel pairs) is up to 5, then images will appear the same, or similar (114).

Aside from colour difference, it is important to consider the deviation in luminance, contrast and correlation between images as these contribute to perceived difference by an observer. The structural similarity index metric (SSIM) (193) considers each of these components making it a commonly used evaluation tool in image quality assessments, with a maximum SSIM of 1 achieved when  $A = B$ . SSIM is calculated as follows:

$$SSIM = l(A, B)^\alpha \cdot c(A, B)^\beta \cdot s(A, B)^\gamma \quad (5.2)$$

where  $l$  is the luminance, or brightness comparison,  $c$  the contrast comparison,  $s$  the structural comparison which considers the local luminance patterns and  $\alpha$ ,  $\beta$  &  $\gamma$  are the positive constants. For further detail, please refer to (193).

Another distortion metric is the mean squared error (MSE), a measure of absolute error. It is calculated as:

$$MSE = \sum_{x,y} \frac{(A - B)^2}{x \times y} \quad (5.3)$$

However taking square root of MSE, or root mean squared error (RMSE), is preferable to MSE as the units become the same as the dependant variable, in this case the difference in pixel value. RMSE is calculated as  $\sqrt{MSE}$ .

Another metric derived from MSE is the peak signal-to-noise ratio (PSNR), in which MSE is scaled by the maximum pixel value (255 in the case of an 8-bit image) and reported in decibels.

$$PSNR = 10 \log_{10} \left( \frac{255^2}{MSE} \right) \quad (5.4)$$

Although considered objective, MSE, RMSE and PSNR must be used with caution. Neither possess defined criteria with which to establish an appropriate value (or threshold) at which two images are perceptually considered the same. MSE & RMSE are always positive valued, decreasing towards zero as error reduces, therefore it is important to minimize these metrics as much as possible. Whereas for PSNR, typically a value between 30-35 is reported as good. These metrics also correlate poorly with perceived difference (165). We include them in our work for descriptive purposes, and to abide with classical objective comparison protocols.

## 5.2.2 Psychophysical Evaluation

For subjective assessment of image quality and/or similarity, psychophysical evaluations are undertaken that aim to derive a preference signal from human participants. A number of protocols exist (90; 112), yet one of the simplest and most precise is the method of paired comparison. This is an experiment in which participants are shown two stimuli at a time, say two images  $A$  &  $B$ , and choose one based on some pre-defined criteria, for example the image that they prefer for the purposes of their analyses. For each pair in this forced-choice experiment, the observer can choose either image  $A$  or image  $B$ , with the selected image automatically assigned a score of 1 and the other 0. This can be also be extended such that the observer can judge the images to be equal - providing a third choice of *no preference*. In this case, both images are awarded a score of 0.5.

In its simplest case, observers are presented with a selection of images categorized into 2 variants, as an example: those that either unenhanced and enhanced. Although in practice, such experiments can include many more variants though the methodology remains the same (72). Importantly, the observer is unaware of the image variants presented when indicating their preference, to remove bias. Pair comparison experiments are very straightforward, not requiring any discrete, continuous or categorical scoring by the observer, and are thus highly suitable for inexperienced participants (135).

They are derived from Thurstone's law of comparative judgements (181), a law based on the idea that given two stimuli, for example images  $A$  &  $B$ , the degree to which  $A$  and  $B$  differ determines the proportion of times  $A$  will be rated greater than  $B$  by observers (56). Thurstone proposed that this selective process between two stimuli, results in normally distributed observer responses. In Thurstone's Case V, the simplest and most widely used of the cases, it is assumed that each variant will have equal variance and equal (or zero) correlations. Thus observer responses are simply normalized by linear transformation into standard z-scores to ensure their scales are comparable. This scaling maps the preference

data to numerical quality scores (on a continuous interval scale), conveying both the ranking of each variant and the magnitude of the difference.

### 5.3 Weibull Tone Mapping (WTM)

In this section we describe our Weibull Tone Mapping algorithm; in which we approximate and smooth a pre-existing tonal adjustment, in this case created by an analyst. Note that in Chapter 6 we demonstrate that the WTM method can extend beyond this as a stand-alone enhancement algorithm, with little alteration to the method.

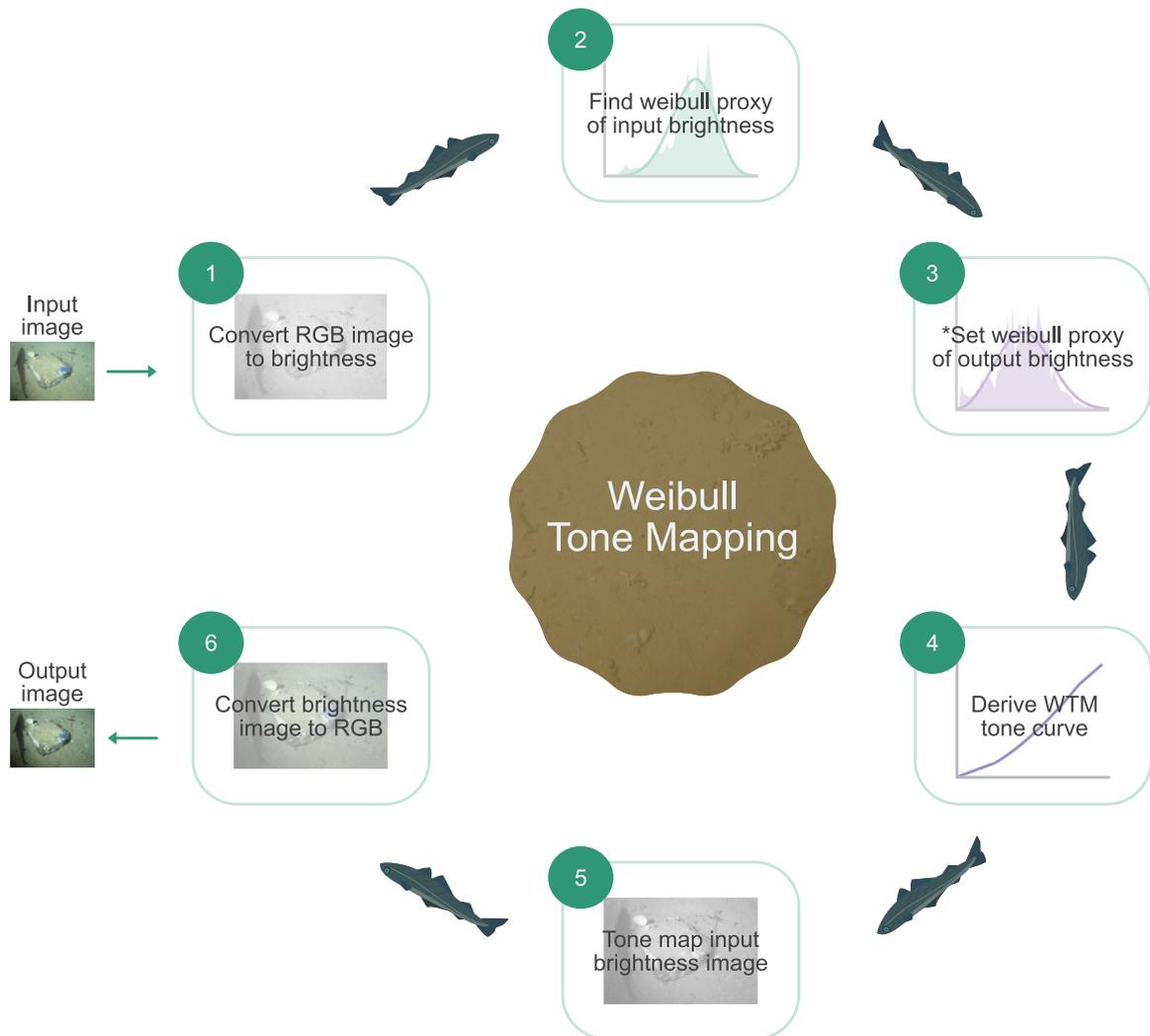
WTM, broadly involves 5 steps. Steps 1 & 2 involve finding the weibull approximation of an input and tonally-adjusted output brightness distribution, as introduced in the previous chapter. In step 3, we derive the tone curve that maps the input WD to the output WD. Finally, in steps 4 & 5 we perform the Weibull tonal mapping operation, applying the weibull tone map to an input image. For clarity, these steps are represented graphically in Figure 5.1. The method by which we map the chromatic aspects of the output image, during tone mapping, (i.e. steps 4 & 5) will be compared in this chapter. For simpler representation, and due to our future development of WTM in Chapter 6, steps 4 & 5 in Figure 5.1 refer to WTM in its chromaticity-preserving capacity only.

In the following sections, we describe each of these steps in further detail.

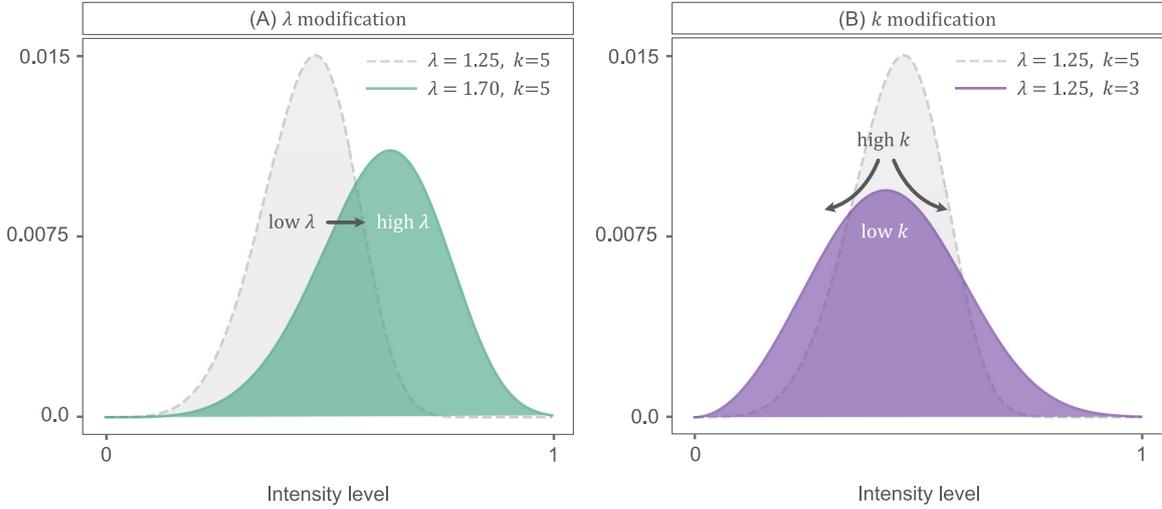
#### 5.3.1 Creating a proxy brightness distribution using the Weibull distribution

Here we wish to use Weibull distributions to drive tone-mapping, by finding the closest approximation of an input and output brightness distribution. Although previously introduced in Chapter 4, we recap and further clarify the method here.

Analogously to CLHE (138; 214), we wish to approximate the histogram  $h(b)$  of a brightness image  $L$ , by a proxy histogram  $h'(b)$ . In CLHE, the proxy brightness distribution is related



**Figure 5.1:** Weibull Tone Mapping flow chart. \*Step 3 can either approximate an existing tonal enhancement (using Equation 5.5) or specify an original WTM enhancement, see Chapter 6



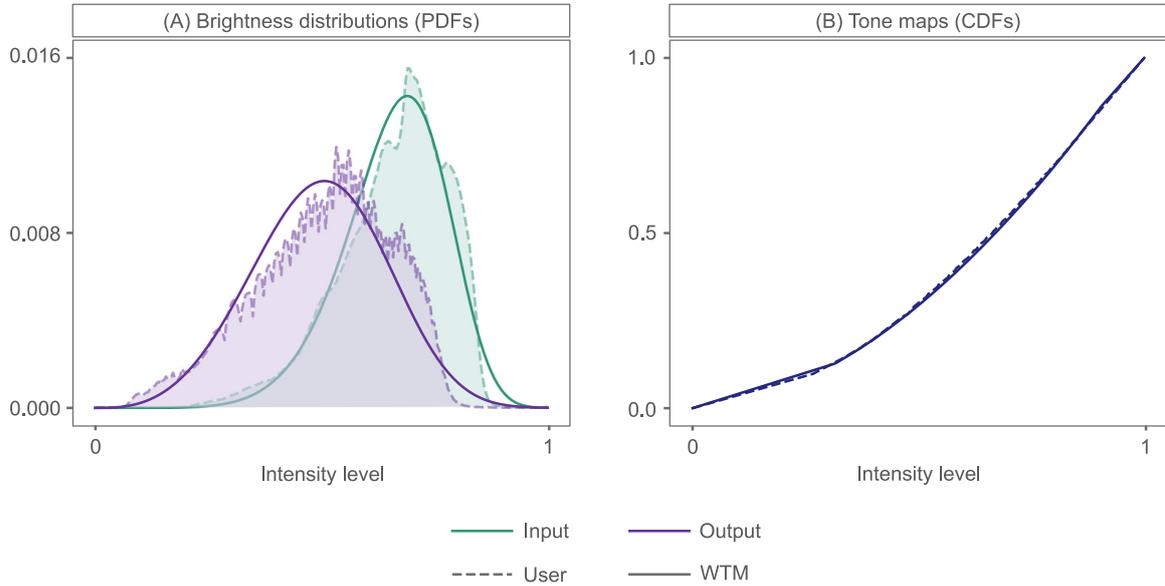
**Figure 5.2:** Parameter control over WD properties: **(A)**  $\lambda$  modification and **(B)**  $k$  modification

to the actual brightness distribution; where the proxy cumulative histogram (the CLHE tone curve), has a bounded slope. This implies that the proxy histogram itself adheres to the slope constraints. Here we wish to further constrain the shape of the proxy histogram, beyond just slope-limitations. Specifically, we will represent brightness histograms by proxies defined by the Weibull Distribution, see Equation 4.2. To ease the exposition, we will present the approach in the continuous domain. However, in practice, all computations are carried out in the discrete domain.

The Weibull proxy,  $h' = h_W(b; \lambda, k)$ , is found by finding the Weibull parameters,  $\lambda$  and  $k$ , that minimizes the Kullback-Leibler (KL) divergence:

$$\min_{\lambda, k} \int_0^1 h(b) \log\left(\frac{h(b)}{h_W(b; \lambda, k)}\right) db \quad (5.5)$$

where  $h(b) = h_W(b; l, k)$  will obtain a KL-divergence of 0. As the difference of two probability distributions increases, so does the KL-divergence. We discretely search parameter pairs of  $\lambda \in \{0, 0.1, 0.2, \dots, 3\}$  and  $k \in \{0, 0.1, 0.2, \dots, 15\}$  creating a large diversity of histogram shapes. In Figure 5.2 we demonstrate how these two parameters control the shape of the



**Figure 5.3:** WTM method showing (A) WTM proxies of an Input and Output (analyst-adjusted) brightness PDF and (B) the corresponding tone maps that transform PDFs from Input to Output. Dashed and smooth lines depict User and WTM respectively

WD. In (A) we see how increasing  $\lambda$  results in a WD in which the peak location is right-skewed, simulating a histogram of a 'brighter image'. Whereas in (B), we see how decreasing  $k$  creates a flatter and wider histogram. If attributed to an image, this would infer higher contrast.

Following the tone mapping experiment in Chapter 4, we have access to an input brightness image  $L_{in}(x, y)$  and an analyst-adjusted output  $L_{out}(x, y)$ , enhanced using a bespoke tone curve, created using a control-point tool (see Figures 4.6 & 4.7). Thus for their respective brightness distributions,  $h_{in}(b)$  and  $h_{out}(b)$ , we find the Weibull proxies,  $h'_{in}(b)$  and  $h'_{out}(b)$ , by minimizing KL-divergence. This results in proxy WDs that closely *match* the input and output brightness distributions, as demonstrated in Figure 5.3 (A), as solid and dashed lines, respectively.

### 5.3.2 Calculating the Weibull Tone Map

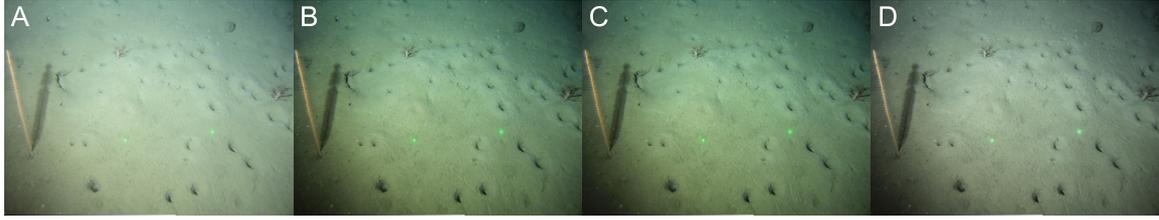
Having determined the closest Weibull approximations of an input and output brightness image, it is necessary to derive the tone curve  $t()$  that will match the input Weibull proxy to the target output Weibull proxy. It is well known that this *histogram matching* can be implemented as a forward and inverse histogram equalisation step. As described in Chapter 2, the HE (88; 64) tone map is simply the cumulative histogram, or cumulative distribution function (CDF), of an input brightness image  $L_{in}(x, y)$ , see Equation 2.3. In a forward HE step, the HE tone map, is applied to  $L_{in}(x, y)$ , transforming pixel values such that its histogram  $h(b)$  becomes uniform  $h_{eq}(b)$ . It therefore follows that an inverse HE step, will alter pixel values in an equalized brightness image  $L_{eq}(x, y)$  such that its histogram  $h_{eq}(b)$  is returned to a distribution  $h'(b) = h(b)$ . The tone map to perform such a step, is simply the inverse HE tone map, or inverse CDF  $t^{-1}()$ .

Using this logic, the tone curve that maps a WD  $h_W(b)$  to a uniform brightness histogram is simply the corresponding CDF of  $h_W(b)$ , denoted as  $t_W(b)$ . It is defined as:

$$t_W(b; \lambda, k) = 1 - e^{-(b/\lambda)^k}, \quad b \in (0, \infty), \quad (5.6)$$

In Weibull Tone Mapping, the tone curves that map the input and output Weibull proxy distributions to a uniform brightness histogram are  $t_{in}(b)$  and  $t_{out}(b)$ , respectively. The inverse of these tone curves, denoted as  $t_{in}^{-1}(b)$  and  $t_{out}^{-1}(b)$ , map the uniform distribution to the input and output proxies. In WTM, we map the original input brightness image  $L_{in}(x, y)$  using the tone curve  $t()$ , so that the input proxy brightness histogram matches the output proxy. We apply:

$$t(b) = t_{out}^{-1}(t_{in}(b)) \quad (5.7)$$



**Figure 5.4:** Comparison of an unenhanced input image (A), an analyst enhanced image (B), Chromatic WTM (C) and Chromaticity-preserving WTM (D)

In Figure 5.3 (B) we see how that this results in a tone map that closely approximates that created by an analyst.

### 5.3.3 Applying the Weibull Tone Map

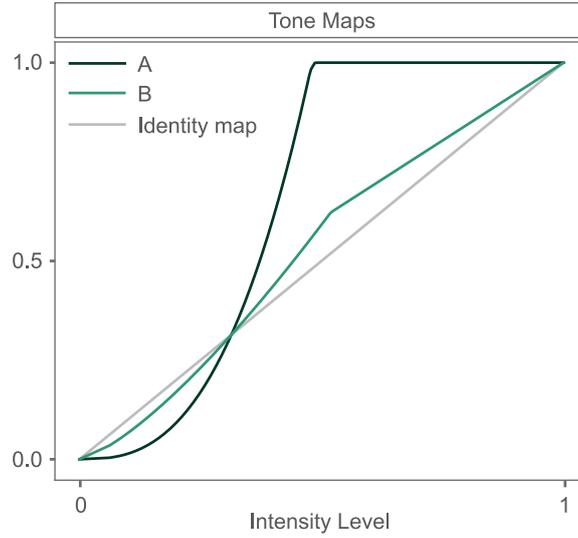
The tone mapping experiment presented in Chapter 4, allowed domain experts to create a tone map based on the brightness component (defined as  $\max(R, G, B)$ ). The photo-editing software GIMP, then applies this to each of the colour channels in an RGB input image to produce an enhanced output image  $\underline{I}_{out}(x, y)$ , see Equation 4.3.

Therefore, to approximate this process with WTM, we merely apply the tone map to R, G and B:

$$\hat{\underline{I}}_{out}(x, y) = t(\underline{I}_{in}(x, y)) \quad (5.8)$$

In Figure 5.4 (A) we show an input image and its analyst enhanced counterpart in (B). Using chromatic WTM (Equation 5.8) we see that this results in an output image  $\hat{\underline{I}}_{out}(x, y)$  in (C) that is indistinguishable from (B) ( $\underline{I}_{out}(x, y)$ ).

Tone mapping an image in this way can produce an colour-boosted output image, which may be beneficial for underwater imagery given that colour loss is common due to light attenuation. Colourfulness is thus a sign of better image quality in some metrics (207; 134; 192)



**Figure 5.5:** An example of two Weibull tone maps which differ from the identity map to varying degrees

However this comes at the expense of distorting the *natural*, or intrinsic, colour properties of the image. A simpler and often preferred alternative to this is to modify the brightness component only (138; 214), retaining the respective ratios of each of the colour channels and thus avoiding unnatural colour casts. To ensure that WTM is chromaticity-preserving, we first adjust the  $\max(R, G, B)$  brightness image:

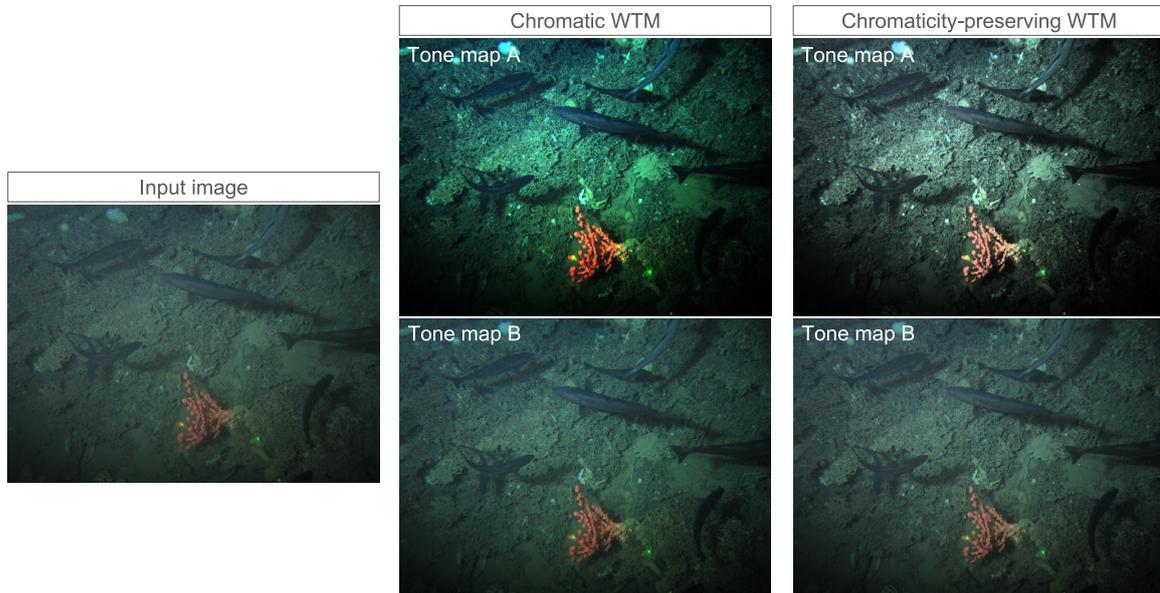
$$\hat{L}_{out}(x, y) = t(L_{in}(x, y)) \quad (5.9)$$

and then scale each of the colour channels by this adjustment:

$$\hat{I}_{out}(x, y) = R \frac{\hat{L}_{out}}{L_{in}}, B \frac{\hat{L}_{out}}{L_{in}}, G \frac{\hat{L}_{out}}{L_{in}} \quad (5.10)$$

This will result in an output image  $\hat{I}_{out}(x, y)$ , see Figure 5.4 (D).

The degree of similarity between an analyst's output and a chromaticity-preserving WTM approximation will vary according to the divergence of the analyst's tone map from the 45 degree identity mapping (in which no tonal adjustment is applied). For example, in



**Figure 5.6:** Image comparison of colour mapping methods Chromatic and Chromaticity-preserving with two WTM tone maps (A & B) presented in Figure 5.5

Figure 5.5 we present two Weibull tone maps: (A) which is most dissimilar to the identity map, providing a significant tonal adjustment and (B) which is, comparatively, gentle and highly similar to the identity map. In Figure 5.6 we show the corresponding output images following application of these tone maps using the colour mapping methods previously described. Application of tone map A using the chromatic method (Equation 5.8) greatly alters each of the colour channels and thus it appears quite different to the chromaticity-preserving output (Equation's 5.9 & 5.10). Alternatively, the tonal adjustments enforced by tone map B are smaller and thus produce similar outputs whether the colour information is preserved or not.

## 5.4 Experiments

### 5.4.1 Approximating bespoke tone maps

This experiment investigates the ability of Weibull Tone Mapping to approximate the analyst's tonal adjustments from Chapter 4. The best approximation of these adjustments

will be achieved by employing WTM in its chromatic form, see Equation 5.8 as it mimics the GIMP tone mapping application. However, for comparison, we also include WTM in its chromaticity-preserving capacity (Equation's 5.9 & 5.10), as it encapsulates the analyst brightness enhancements without distorting the colours. Preserving the *natural* qualities in this way, could better suit analysts' interpretation of underwater imagery by retaining appearance of features in a familiar, but improved sense. We explore this idea further in a psychophysical evaluation in Experiment 5.4.2.

### Setup

The tone mapping experiment in Chapter 4 provided 180 bespoke tonal adjustments, 3 unique tone maps for each of the 60 images presented to analysts. Tone maps were then approximated using WTM, using the algorithm outlined in Section 5.3. We evaluated the similarity of the analyst tone maps and their Weibull approximation using RMSE & PSNR, as described in Section 5.2.1, and a Pearson Correlation Coefficient to assess the strength of a linear relationship between the tone maps. Ranging from -1 to 1, a correlation of 1 or -1 indicates a perfect linear relationship, that is positive or negative respectively.

Each of the Weibull tone maps was applied to their corresponding input image. Given the two colour mapping methods, this resulted in a total of  $180 \times 2 = 360$  WTM images. For brevity, Chromatic WTM and Chromaticity-preserving WTM are referred to as  $WTM_C$  &  $WTM_{CP}$  respectively. We assessed the similarity of WTM images and the output images they approximated using a variety of metrics. We used  $\Delta E_{ab}^*$ , SSIM, RMSE and PSNR, to consider changes in chromaticity, lightness and structural information as well the general quality and/or difference. Details of these metrics can be found in Section 5.2.1.

### Results

Our Weibull Tone Mapping algorithm was found in general to closely mimic the bespoke tone maps designed by domain experts. In Table 5.1 we present a summary of metrics indicating the similarity of analyst tone maps and their Weibull approximation. Note that

tone maps have been normalized to 8 bits, where values are  $\in [0, 2^8 - 1]$  or  $[0, 255]$ , for ease of interpretation. This is specifically useful for RMSE, as error represents the discrete (integer) pixel values, rather than a small normalized value. From this we see that the Weibull tone maps diverge from the analyst maps by an intensity of only  $\sim 9$  on average. PSNR was also reasonable on average, falling within the expected range of 30-50db for 8-bit images.

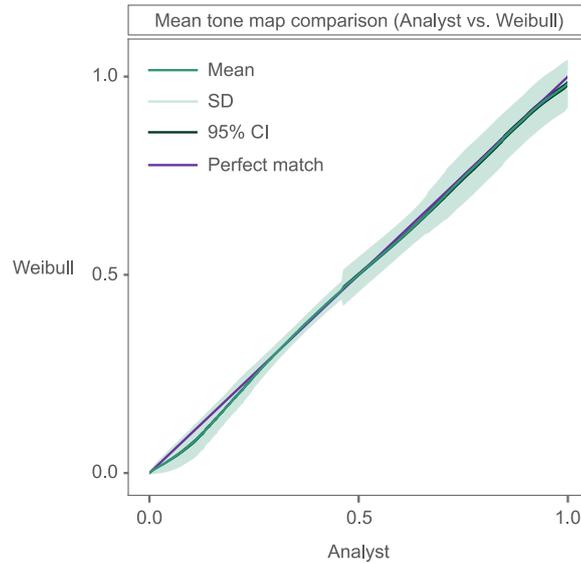
The Weibull tone maps were found to be, consistently, highly positively correlated to analyst tone maps with an average Pearson correlation coefficient of 0.995, see Table 5.1. This is further demonstrated in Figure 5.7. Each tone curve has a natural order from 1 to 255 and records the value which each input brightness is mapped to e.g.  $t(10) = 20$  ‘means’ that an input brightness of value 10 is mapped to 20. Assuming the WTM tone map is similar to that made by the analyst then we would expect  $t_{WTM}(b)$  approx equal  $t_{analyst}(b)$ . Thus, if we plot the tone curves against each other, we expect a line at 45 degrees. To a large extent this is found to be true, see Figure 5.7. Broadly, plotting one tone curve against another is a line at 45 degrees with narrow confidence limits around this line.

Exploring the similarity metrics in more detail, we see that the high degree of correlation between analyst and Weibull tone maps was fairly consistent across analysts; exhibiting only slightly better correlation to Analyst 3’s tone maps, see Output 3 in Figure 5.8. In contrast, Weibull tone maps were considered to best approximate Analyst 1’s tone maps in terms of PSNR and RMSE, with the fits of Analyst 2 & 3’s tone maps equivalent. In

**Table 5.1:** Summary statistics for three metrics measuring similarity between each of the analyst tone maps and their closest Weibull tone map.

Metric	Mean	SD	Min.	Median	99 <sup>th</sup> p.
<b>Corr.</b>	0.995 ( $\pm 0.001$ )	0.01	0.894	0.998	1
<b>RMSE</b>	8.998 ( $\pm 1.045$ )	7.155	0.612	7.791	33.889
<b>PSNR</b>	31.285 ( $\pm 0.971$ )	6.643	11.829	30.396	52.17

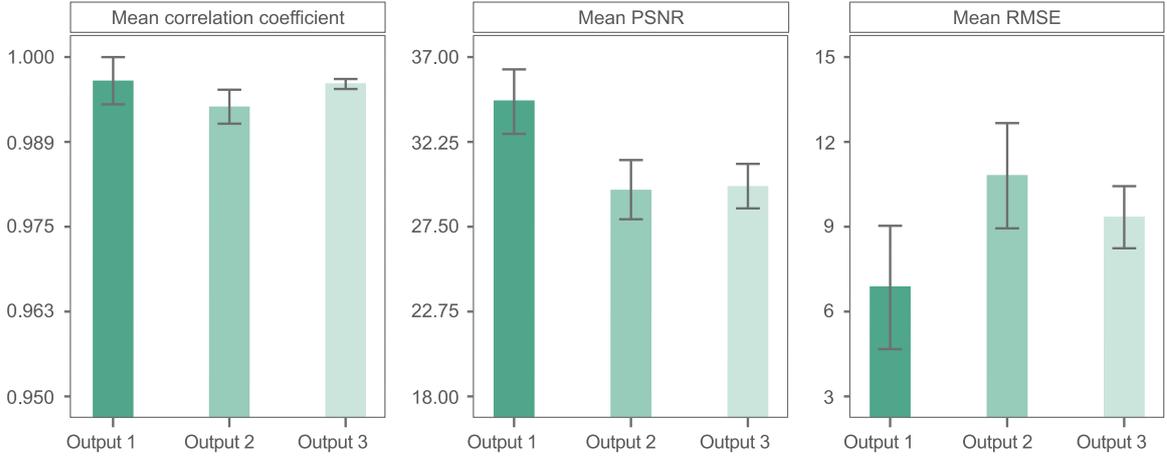
<sup>1</sup>95% confidence intervals are shown in parentheses <sup>2</sup>SD = Standard Deviation <sup>3</sup>Corr. = Pearson correlation coefficient <sup>4</sup>p. = Percentile



**Figure 5.7:** Average intersection, or *match*, of analyst tone maps and their Weibull approximation. A 45 degree line (slope=1/N) indicates a perfect match. Standard deviation (SD) and 95% confidence intervals (CI) are also presented.

general these results indicate that analysts adjust underwater imagery such that the target brightness distributions possess Weibull-like characteristics.

We also see that this is true regardless of the image content to varying degrees, see Figure 5.9. In general the fit of the Weibull tone maps was found to decrease with increasing seabed complexity, with tone maps applied to soft substrate images (SS1 & SS2) approximated best, followed by hard substrate (HS1 & HS2) and reef (Cor1 & Cor2). In Chapter 4, it was found that the WD approximated the output brightness distributions in reef images, specifically soft corals (Cor2), less well. In Figure 5.9 we see the result of this, in that the associated Weibull tone maps exhibited a significantly lower correlation and PSNR than most other habitats.



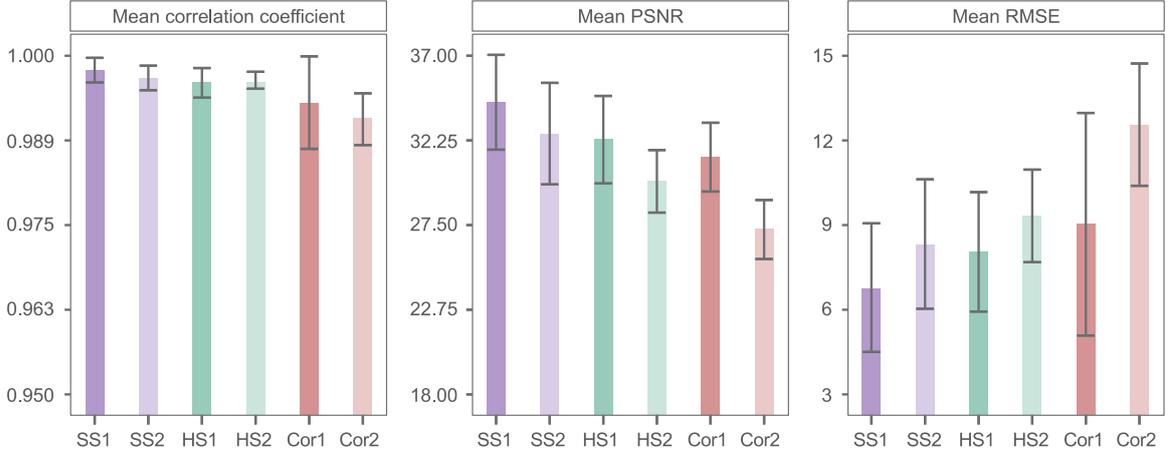
**Figure 5.8:** Similarity metrics between the analyst tone maps and their best Weibull approximation. Results are displayed according to enhancement type. Error bars represent 95% confidence intervals.

**Table 5.2:** Summary statistics for similarity metrics between analyst images and their respective Weibull approximations. Chromatic and chromaticity-preserving Weibull tone map application are denoted by subscripts  $C$  &  $CP$ , respectively.

Metric	Mean	SD	Min.	Median	99 <sup>th</sup> $p.$
Mean $\Delta E_{ab}^*C$	3.568 ( $\pm$ 0.235)	1.613	0.949	3.402	8.404
Mean $\Delta E_{ab}^*CP$	4.99 ( $\pm$ 0.357)	2.453	0.956	4.504	12.117
<b>SSIM<sub>C</sub></b>	0.936 ( $\pm$ 0.002)	0.015	0.847	0.94	0.954
<b>SSIM<sub>CP</sub></b>	0.933 ( $\pm$ 0.002)	0.016	0.837	0.938	0.953
<b>RMSE<sub>C</sub></b>	6.315 ( $\pm$ 0.461)	3.159	1.888	5.994	20.497
<b>RMSE<sub>CP</sub></b>	7.882 ( $\pm$ 0.553)	3.785	1.9	7.341	21.071
<b>PSNR<sub>C</sub></b>	33.673 ( $\pm$ 0.476)	3.269	27.73	33.017	41.754
<b>PSNR<sub>CP</sub></b>	32.301 ( $\pm$ 0.457)	3.135	27.619	31.419	41.749

<sup>1</sup>95% confidence intervals are shown in parentheses <sup>2</sup>SD = Standard Deviation <sup>3</sup> $p.$  = Percentile

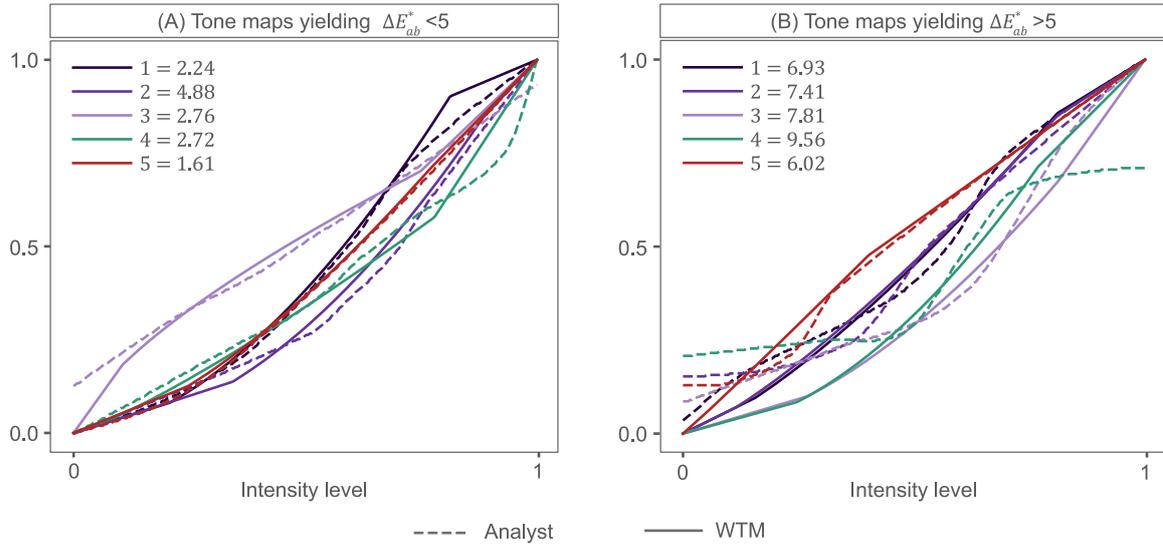
It is important to contextualize the adequacy of Weibull approximations in a visual sense. In Table 5.2 we therefore present summary statistics for the similarity between the analyst tone-mapped images and their associated WTM image, either Chromatic (WTM<sub>C</sub>) or Chromaticity-preserving (WTM<sub>CP</sub>). WTM<sub>C</sub> was significantly better at approximating the analyst tonal adjustments than WTM<sub>CP</sub>, according to all image metrics. This is expected given that the tone map is applied to images in the same method as the user tone maps



**Figure 5.9:** Similarity metrics between the analyst tone maps and their best Weibull approximation. Results are displayed according to habitat class. Error bars represent 95% confidence intervals.

(Equation 5.8). When applied chromatically, the majority (87%) of Weibull tone maps produced images that were highly visually similar to the analysts images, indicated by a mean  $\Delta E_{ab}^* < 5$ . Thus the average colour difference between these images pairs was also small, with a mean  $\Delta E_{ab}^*$  of 3.57 ( $\pm 0.24$ ), see Table 5.2. This is further demonstrated by only an average RMSE of 6.32 ( $\pm 0.46$ ) pixels to the analyst images and a mean PSNR of 33.67 ( $\pm 0.48$ ). Interestingly, the  $WTM_{CP}$  method could still adequately capture a large number of the analyst’s adjustments, with 56% scoring a mean  $\Delta E_{ab}^* < 5$ , resulting in an average colour difference of 4.99 ( $\pm 0.36$ ). SSIM was comparable to the chromatic method, lower by only 0.03 on average. Mean RMSE was higher & PSNR lower, indicating a poorer approximation, but still reasonable, with 7.88 ( $\pm 0.55$ ) and 32.3 ( $\pm 0.46$ ), respectively.

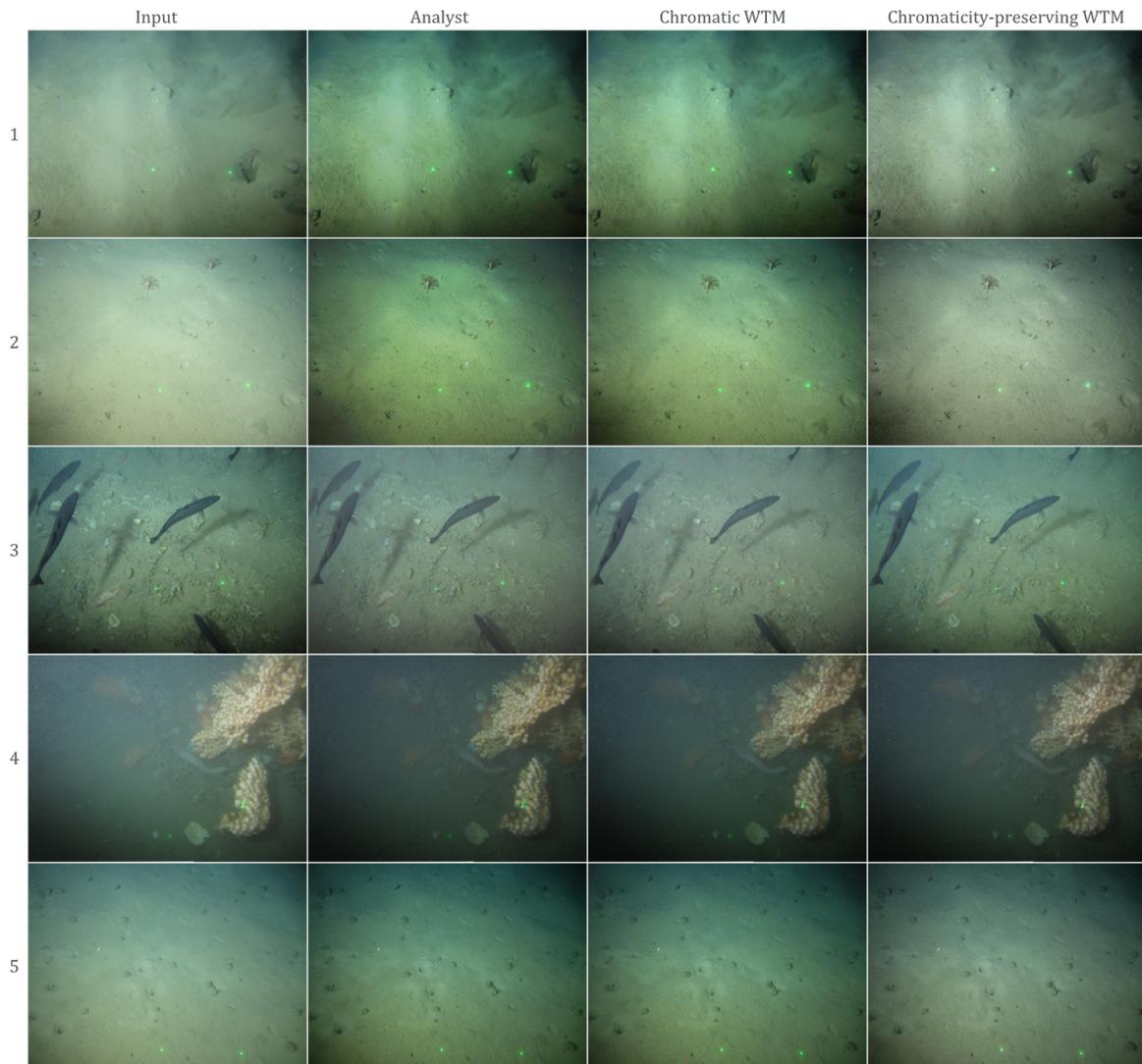
In Figure 5.10 we present a selection of analyst tone maps and their best Weibull approximations. In (A) we show Weibull tone maps that resulted in a mean  $\Delta E_{ab}^* < 5$  between the analyst reference image and the Chromatic WTM image. From this we see that WTM provides general approximations that are smoother. Although the Weibull tone maps are not a perfect match in each case, when we view the corresponding images in Figure’s 5.11 we see that, perceptually, the Chromatic WTM images appear almost the same. Comparatively, in Figure 5.10 (B) presented tone maps resulted in a mean  $\Delta E_{ab}^* > 5$ . The



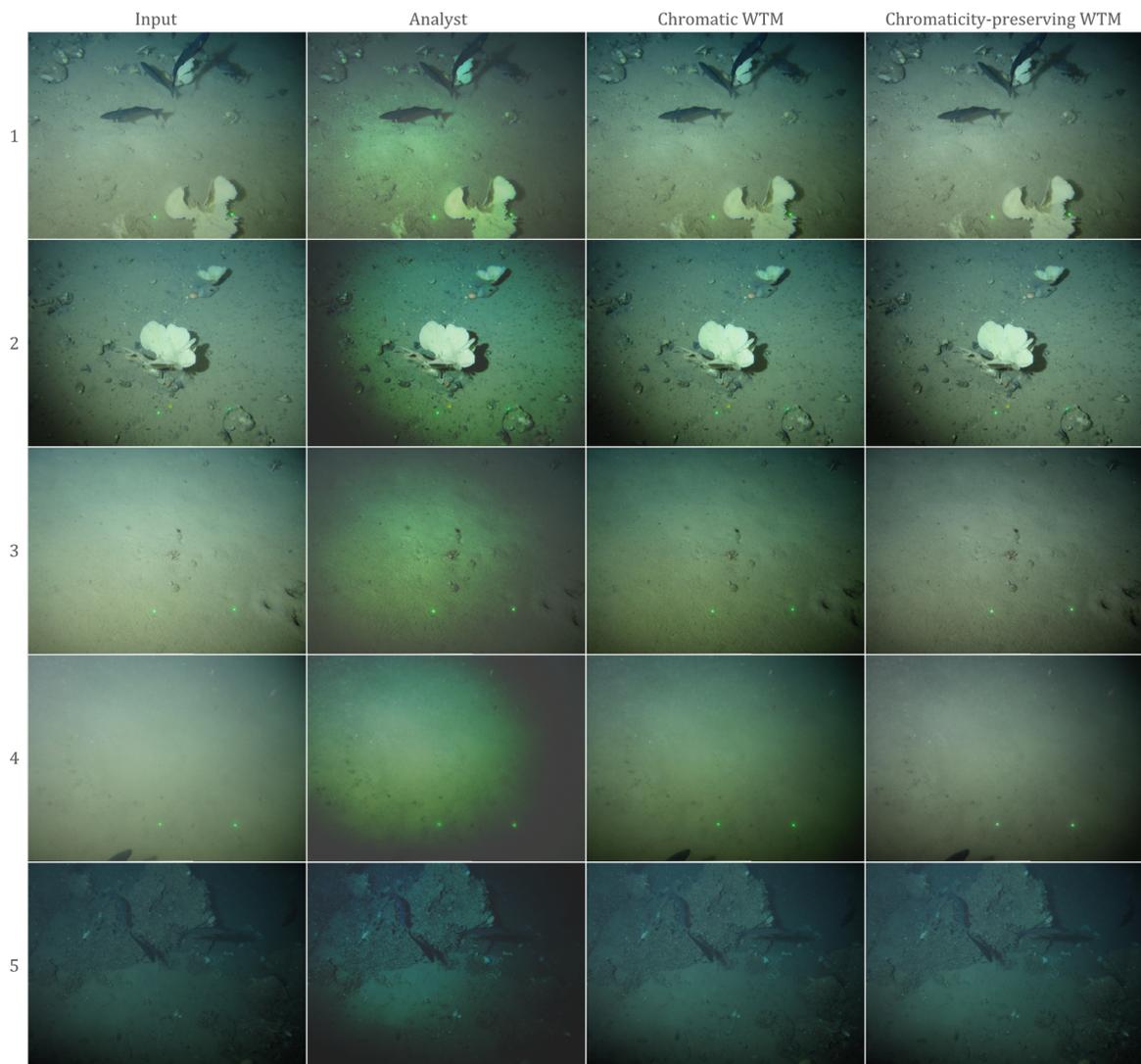
**Figure 5.10:** Example of analyst tone maps and their best Weibull approximation. In (A) Weibull tone maps resulted in good approximation of analyst images, with Chromatic WTM, scoring a mean  $\Delta E_{ab}^* < 5$  compared to (B) which scored a mean  $\Delta E_{ab}^* > 5$ , approximating adjustments less well. Tone maps in each subplot are numbered and their mean  $\Delta E_{ab}^*$  listed.

analyst tone maps in these cases were quite complex. Tone maps fluctuated between very low slopes and high slopes and also often compressed the dynamic range. These properties are not common to the smoother Weibull tone maps, which more typically stretch between 0 & 1. This results in clear differences between analyst and Weibull image pairs as shown in Figure 5.12. The analyst images in Figure 5.12 look significantly modified from their input appearance. The enhancements generally enhance contrast, however at the expense of exacerbating the light cone and vignetting effects. Evidence of dynamic compression can be seen, in that the darkest intensities have been brightened. The colour cast of the images has also been strongly enhanced; images appear extremely and unnaturally green. The fact that the analyst enhancements do not appear aesthetically-pleasing, by typical standards, is irrelevant. They are focused on enhancing visibility of cryptic features, such as fine textural details on sponge species. The analyst will make these adjustments, at the expense of enhancing other undesirable characteristics, such as enhanced vignetting around the edges of the image, that can obscure features in these areas. The Chromatic WTM

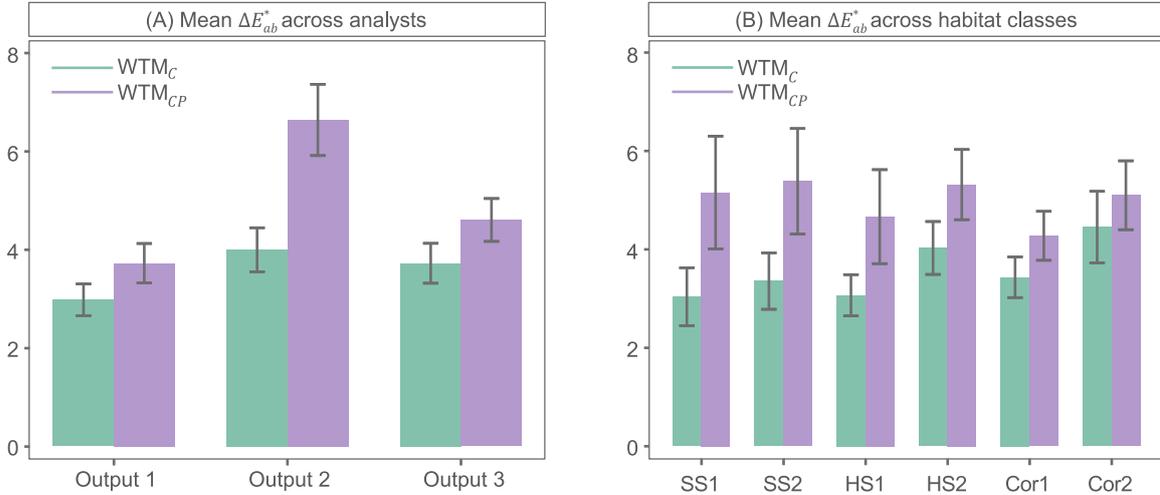
enhancements appear less intense by comparison. They are better contrasted than the input and colour-boosted, however, the vignetting and colour casts affects are less severe. When Weibull tone maps are applied such that they preserve chromaticity these effects are diminished even further.



**Figure 5.11:** Example images corresponding to tone maps presented in Figure 5.10 (A).



**Figure 5.12:** Example images corresponding to tone maps presented in Figure 5.10 (B).



**Figure 5.13:** Mean  $\Delta E_{ab}^*$  between WTM & analyst-enhanced images across (A) analysts and (B) habitat classes.

Considering the mean  $\Delta E_{ab}^*$  scored according to analysts, WTM approximations were found to be significantly more similar for output images made by Analyst 1, see Figure 5.13. When applying  $WTM_C$ , output images for the remaining analysts were significantly similar. Unlike  $WTM_{CP}$  for which mean  $\Delta E_{ab}^*$  was worst for Analyst 2’s output images. As with the tone map similarities presented in Figure 5.9, resulting WTM images did not, in general, significantly differ across habitat classes. Although tone map approximation metrics in Figure 5.9 demonstrated a mild trend of decreasing fit with increasing seabed complexity, these differences did not appear to strongly influence the resulting images. Significant differences were only noted between the fit of HS1 images and, HS2 and Cor2 when using  $WTM_C$ .

#### 5.4.2 Preference experiments

Given the good performance of WTM in the previous experiment, we undertook pair-wise comparison experiments to determine to what extent analysts prefer - for the purpose of image interpretation - our WTM adjustments compared with 1) their own bespoke enhancements (from Chapter 4) and 2) the original images that are unenhanced. We explored their preference signal for both chromatic and chromaticity-preserving WTM. In the latter case,

we also compared this preference against images enhanced with the CLHE algorithm; a widely-used tone mapping tool, see Section 2.2.2.

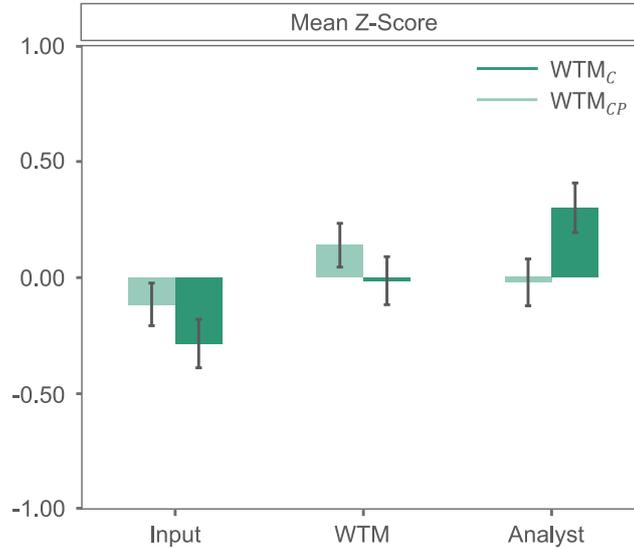
### Setup

We ran two psychophysical evaluations to assess the preference of analysts with respect to (1) Chromatic WTM or (2) Chromaticity-preserving WTM. In each evaluation, analysts at Gardline, were presented with uniquely randomized pairs of images. For every image pair they were asked to ‘Choose the image that best allows identification of the habitat (class) therein, or no preference if the images are equivalent’. Analysts could thus choose one of three options for each image pair; Image 1, Image 2 or No Preference. This was conducted under ISO standard 3664:2009 viewing conditions (92); with analysts sitting approximately 70 cm from the display in a neutrally painted and darkened room. Subject to Gardline’s availability, a total of 6 analysts completed the evaluation featuring Chromatic WTM images, 3 of which created the bespoke tone maps presented in Chapter 4. The same 6 analysts, and a further 4 (total=10), completed the evaluation with Chromaticity-preserving WTM images. Note that in each evaluation, all analysts completed the preference experiment twice.

For each evaluation analysts were presented with identical image datasets except for WTM images, in which the method was varied (either  $WTM_C$  or  $WTM_{CP}$ ). Note that we down-sampled the analyst adjusted images, randomly selecting one of the three possible outputs for each of the 60 original images. For each of the original images we therefore have  $n = 3$  variants in each evaluation, namely the unenhanced input, the analyst adjusted output and the WTM approximation. This results in  $\frac{n(n-1)}{2} = 3$  pair-wise comparisons per image, which is then viewed twice (left-right order switched). Thus, the total number of pairwise comparisons is  $60 \times 2 \times 3 = 360$  pairs; too many to compare in a reasonable time frame. To minimize the duration of the experiment, we thus chose 2 random subsets (one for each experiment repeat) of 18 images, 3 per habitat class (total=6). The number of pairwise comparisons, per experimental repeat, was therefore  $18 \times 2 \times 3 = 108$ .

Preference votes for each image pair were processed by awarding a score of 1 to the chosen image and 0 to the other. If no preference was selected, each image was given a score of 0.5. Votes for each analyst, in each experimental sitting, were converted to  $3 \times 3$  frequency matrices, of which the score at  $[i, j]$  represents the frequency of votes in which image variant  $i$  was preferred over variant  $j$ , across the 18 images. Thurstone's Law of Comparative Judgments, or Thurstone's Case V (181), was then used to convert frequency matrices to z-score (standard score) matrices. This involved first converting each frequency matrix to proportions before deriving the z-scores. Self comparisons on the diagonal of the proportion matrix, when variant  $i = \text{variant } j$ , are assigned a proportion of 0.5 assuming that when judged against itself any variant would get a random amount of votes (41). It is anticipated that each analyst conducting the preference experiment would choose either variant 50% of the time. In reality, these pair comparisons are never presented to the analyst. However, conversion to z-scores scales these proportions to zero and thus they are effectively cancelled out.

In the psychophysical evaluation with  $\text{WTM}_{CP}$  analysts were also shown images enhanced with CLHE. CLHE parameters, that is the minimum and maximum slope of the tone map, were set to be the average upper and lower slope bounds of all WTM tone maps, for fair comparison. This resulted in a minimum and maximum slope of 0.003 and 0.011 respectively, corroborating well with traditional usage of CLHE in which studies often use a default value of 0.01 for the maximum slope (75; 59; 60; 190; 37; 188; 126). Using these parameters, the CLHE tone map was derived for each of the input images in the 2 random subsets. Images were subsequently tone mapped used the same methodology as  $\text{WTM}_{CP}$ , see Equations 5.9 & 5.10. This resulted in an additional  $18 \times 2 = 36$  pairwise comparisons, per experimental repeat. For clarity, and since we do not consider CLHE in the  $\text{WTM}_C$  psychophysical evaluation, we analyse this data separately. Preference votes are treated as previously described, but instead converted into  $4 \times 4$  frequency matrices given the additional algorithm.



**Figure 5.14:** Preference Experiment Scores for Chromatic and Chromaticity-preserving WTM, denoted by subscripts  $C$  &  $CP$ , respectively. Error bars represent 95% confidence intervals.

## Results

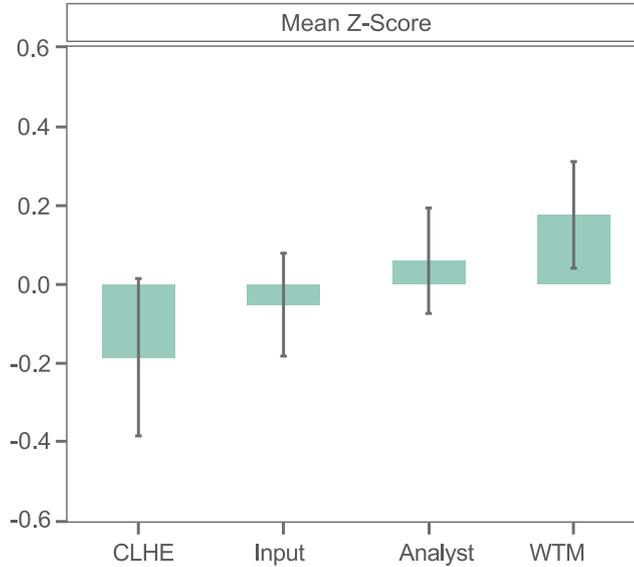
The preference experiment results were interesting. In Figure 5.14 we present the average z-scores across analysts in each evaluation. Note that scores for input and analyst-enhanced images are coloured according to experiment (chromatic or chromaticity-preserving), however the images are the same. Significantly, Figure 5.14 shows that given an unenhanced input image, analysts find a WTM enhancement useful for the extraction of biological data from imagery, regardless of the colour rendering method. Analysts across the experiments also preferred the bespoke (analyst) tonal enhancements over input images, though to variable degrees. These tonal enhancements were created by a subset of the analysts with subjective preferences. The enhancements created may thus not be optimised for all analysts. However, the results in this chapter show that they have general properties which are encapsulated and improved by Weibull Tone Mapping.

Comparing the two methods of colour rendering for WTM, we find for Chromatic WTM - in which we more accurately match the analyst tonal adjustments - domain experts prefer their

own manipulations to the WTM simplification. However when preserving the chromaticity of images, analysts instead find a WTM enhancement (slightly) more preferable to their own bespoke enhancements. Significantly, the 95% confidence intervals for Chromatic WTM overlap with those of the analysts in the chromaticity-preserving experiment. Thus, at this confidence level the 2 methods are, statistically, similarly preferred, although not directly compared. In general, these results demonstrate that how the image colour component is mapped, when based on a brightness tonal adjustment, can significantly influence the preference assessments of end-users. In the original experiment, the analysts were forced to apply their tone maps chromatically. This work suggests that if given the choice, perhaps analysts would prefer the alternative.

Although these preference experiments did not directly compare the two WTM methods, the preference signal demonstrates that Chromaticity-preserving WTM may be better suited for enhancing visualization of marine habitats in imagery. This is despite images being less similar to the analyst tonal adjustments than the Chromatic WTM. Chromaticity-preserving WTM captures the brightness distribution qualities that are desired by analysts but with the benefit of maintaining the existing hue and saturation. The result is an image that is more reminiscent of the original image but with enhanced tonal properties.

Additionally, we found that these WTM qualities are significantly preferred to the traditional and automated tone mapping approach CLHE, see Figure 5.15. Given that custom-per-image enhancements are time-consuming and thus less practical, a good automated tone-mapper should offer a suitable trade-off - at the very least performing equivalently. Clear from these results is that, although popular, CLHE is not up to the task. Of the image variants presented, end-users preferred CLHE least for image analysis, followed by the unenhanced inputs and the custom enhancements. Although differences are noted between these variants, statistically speaking, they are not significantly preferred to the CLHE enhancements. With these additional comparisons,  $WTM_{CP}$  is still preferred by analysts, however the strength of the preference signal is lessened.



**Figure 5.15:** Preference Experiment Scores for Chromaticity-preserving WTM with the addition of CLHE pairwise comparisons. Error bars represent 95% confidence intervals.

**Table 5.3:** Mean image quality metrics for input images and enhanced outputs

Image	UIQM	UCIQE	CCF
Input	0.6 ( $\pm 0.07$ )	0.7 ( $\pm 0.14$ )	7.39 ( $\pm 1.05$ )
Analyst	0.48 ( $\pm 0.11$ )	0.64 ( $\pm 0.07$ )	7.04 ( $\pm 0.97$ )
Weibull	0.56 ( $\pm 0.1$ )	0.7 ( $\pm 0.15$ )	7.64 ( $\pm 1.19$ )
CLHE	<b>1.1</b> ( $\pm 0.04$ )	<b>1.45</b> ( $\pm 0.09$ )	<b>12.52</b> ( $\pm 1.37$ )

<sup>1</sup> 95% confidence intervals are shown in parentheses. <sup>2</sup> Best results are highlighted in bold.

Interestingly, these results are in contradiction to objective metrics of image quality which rank CLHE best of the pairwise comparisons shown and analyst enhancements worst, see Table 5.3. Here we present 3 popular no-reference quality metrics for underwater image enhancement: (1) Underwater Image Quality Metric (UIQM) (134), (2) Underwater Color Image Quality Evaluation (UCIQE) index (207) and (3) a colourfulness, contrast and fog density (CCF) metric (192). UIQM evaluates image colourfulness, sharpness and contrast whereas UCIQE measures the degradation of colour in the CIELAB colour space. CCF is inspired by underwater imaging absorption and scattering characteristics to predict colour loss, blurring and fog/haze. For each, better image quality is associated with a higher value.

Since each of these metrics consider higher contrast to be advantageous in their assessment, our results in Table 5.3 are unsurprising given that CLHE is a contrast-enhancement algorithm and, as discussed in Chapter 4, a large proportion (49%) of the analyst enhancements reduced contrast. This demonstrates that objective assessments of image quality do not always correspond well with subjective perception, as discussed in previous work (112; 78). These metrics should therefore be employed cautiously.

## 5.5 Conclusion

In this chapter, we developed a method for tonally adjusting underwater images called Weibull Tone Mapping (WTM). Given an input and an analyst tone-mapped image, WTM provides an output image derived from the input-output pair. The method works by approximating the input and output distribution, with a Weibull probability density function, and calculating the tone curve that maps between these distributions. This map, the WTM map, is then applied to the input image to generate a new output. The WTM method is designed, by construction, to approximate analyst adjustments but result in smoother and simpler tonal adjustments. Psychophysical evaluations showed that such adjustments improve appearance of underwater imagery for better classification of marine habitats.

We find that application of the WTM tone map to all colour channels (R, G and B), provides a more colour-rich image and a better approximation of the analyst tonal adjustments in general. However, interestingly, this method was not preferred by analysts. Instead analysts favour application of the WTM tone map to a brightness image, preserving chromaticity. Given this altered preference, it therefore seems prudent to consider how colours are rendered when developing and assessing tone-mapping algorithms, as they can have a significant influence on quality perception by end-users.

At this stage, Weibull Tone Mapping is ‘existential’ in nature. Given an analyst adjustment it provides an even better tonal enhancement when preserving chromaticity. In addition, analysts prefer these enhancements to those provided by CLHE, a conventional automated

tone mapper. In the following chapter we will explore how this case of WTM can be employed as an interactive enhancement tool, in which analysts can directly adjust images according to WTM theory.

# 6 Weibull Tone Mapping (WTM): as an interactive enhancement tool

In Chapter 2 we introduced our Weibull Tone Mapping (WTM) algorithm that creates smoother and simple approximations of bespoke tonal adjustments by image analysts. Importantly, we showed that application of these WTM tone maps, in a way that preserves chromaticity, is preferred by end users, over their own adjustments, to interpret underwater imagery. In this chapter, we extend WTM beyond an approximation tool to a interactive enhancement tool; allowing users to directly enhance imagery according to WTM theory. We show that, in general, analysts prefer this tool to more advanced control-point tonal manipulations, significantly shortening the time required to create a bespoke and targeted tonal enhancement to support their analyses.

## 6.1 Introduction

Creation of bespoke tonal enhancements are a relatively simple way to enhance imagery. Moreover, a tonal adjustment tool is easy to create and typically involves adjusting the positions of control points on a curve (through which a final tone curve is interpolated). Their inherent power and flexibility results in improved appearance in output images, that are easier to interpret by analysts. As we have demonstrated previously, since the data collected from underwater imagery, and its usage, is highly diverse, it is crucial to consider the purpose the imagery serves (and its audience) when designing or evaluating an enhancement. The high variability in image appearance within datasets also motivates the need for highly flexible tone mapping approaches.

However, since bespoke tonal adjustments are time consuming and benthic datasets are large, the cost of finding a per-image bespoke tone curve is high. One solution to increasing the speed of tonally adjusting images is to use an automatic adjustment tool/algorithm. Perhaps the best known automatic tone adjustment is the histogram equalization (HE) method. As mentioned in Chapter 2, HE enhances the conspicuousness of image details by enforcing a flat, or uniform, brightness histogram, where all brightness values are in use, and in an equal amount; they are equalized (88; 64).

However, HE is too simplistic in formulation and often produces images with contrasts that are both unnaturally low and high (in different image regions). An alternative is to more subtly enforce uniformity, as is the case in Contrast Limited Histogram Equalization (CLHE) (138). The tone curve that maps an input brightness distribution to a uniform (flat) output is constrained so that the slope is neither too steep or shallow, lessening the degree of flattening in the mapped output brightness distribution. CLHE has the advantage that it is a simple method and often produces improved results. However, neither HE nor CLHE are suited to underwater imagery, which rarely exhibit, nor benefit from, uniform brightness histograms. Tone mapping is, of course, a large field. In the context of underwater imaging, a range of algorithms have been applied including (138; 214; 91; 75; 59; 60; 58; 12; 212; 213). See (154; 85) for a review of underwater image processing. However, to our knowledge, existing automated algorithms are rarely designed from the perspective of an end-user nor reliably produce outputs which are always preferred by the users. As has been a common theme throughout this thesis, when creating or assessing an image enhancement, it is vital to consider the function of the image and its end-user.

We might say that automated algorithms are *constructive* in the sense that they provide an algorithm that makes an output image without user involvement. In this chapter we will consider a *descriptive* approach: we will propose properties that an automated algorithm should have. We then adopt these properties - in a simplified user tone-manipulation scenario - without specifying how an automated algorithm should work. Rather, we are

manipulating the brightness values, in the same way that the data suggests an automated algorithm should alter them. In this context there is still the need for user involvement. However, their selection of a suitable tone curve is guided (and simplified) by incorporating knowledge of how analysts typically tonally adjust underwater images.

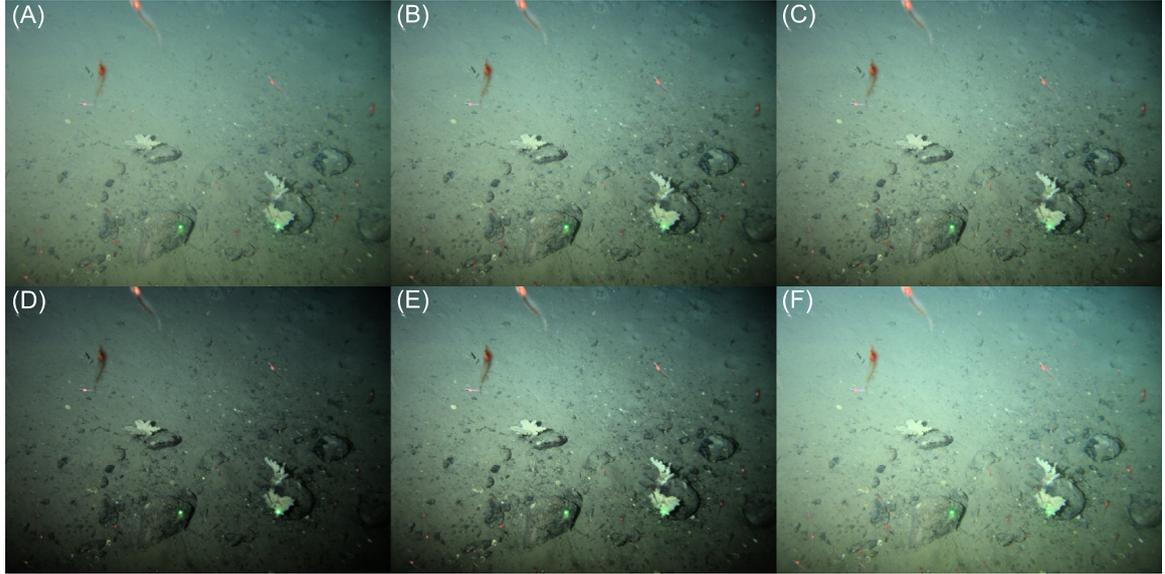
A key contribution of this chapter is to allow analysts to directly enhance images according to WTM theory. We provide analysts with a tool that simplifies the image enhancement task, making users more productive in image annotation and further demonstrate the applicability of the WD to underwater imagery - focusing on a larger image dataset.

## 6.2 Background

Before introducing our interactive WTM tool, we briefly recap here on the prior art of Weibull Tone Mapping (WTM), as presented in Chapter 5. A Weibull distribution (WD) is a smooth & unimodal function that is parameterised by two numbers that control the peak location and spread of the distribution (194). In WTM the brightness distributions of an input and a user-adjusted output image are first represented by WDs, or Weibull proxies,  $h'_{in}(b)$  and  $h'_{out}(b)$ . These are Weibull approximations of the input and output histograms,  $h_{in}(b)$  and  $h_{out}(b)$

Then, the tone map,  $t()$ , that alters the input image to the output, is defined as the function that maps the underlying input WD to the target output WD, see Equation 5.7. Application of  $t()$  to an input brightness image  $L_{in}(x, y)$  results in  $\hat{L}_{out}(x, y)$ , a WTM approximation of a analyst modified output  $L_{out}(x, y)$ , see Equation 5.9.

In Figure 6.1 (A) we present an input image that lacks contrast. In (B), an analyst has enhanced the contrast and, more mildly, the brightness of the image. In (C) we see that the WTM approximation closely mimics these adjustments by an analyst. Here the tone curve is applied to preserve chromaticity, see Section 5.3.3. All results presented in this chapter preserve the input chromaticity in the tone-adjusted output.



**Figure 6.1:** WTM example: (A) Input image, (B) Analyst-adjusted Output, (C) WTM approximation of Output image (B) and (D-F) parameterized WTM adjustments of (A)

### 6.3 WTM as a parameterized enhancement tool

In our WTM method we approximate input and output brightness histograms ( $h_{in}(b)$  and  $h_{out}(b)$ ) by proxies that follow the Weibull distribution ( $h'_{in}(b)$  and  $h'_{out}(b)$ ), see Section 5.3.1. Then the tone mapping,  $t()$ , that is applied to the input brightness image, is the curve that maps the input to output proxy (Section 5.3.2). The output brightness image is mapped from the input according to  $\hat{L}_{out}(x, y) = t(L_{in}(x, y))$ .

In the context of the parameterized tool we develop here, we continue to use the Weibull proxy for the input brightness distribution. However, the behaviour of the output distribution is determined by the user, by adjusting the two parameters that drive the Weibull distribution. In effect, they define the output proxy distribution. An analyst can simultaneously look at the tone-mapped output image, continuing to adjust the Weibull parameters until they find a preferred result.

Under the hood, the users are adjusting the Weibull parameters  $\lambda$  and  $k$ , see Equation 4.2. A high  $\lambda$  is associated with a histogram peak towards the brightest region in the dynamic range of the image. Whereas a high  $k$  decreases the slope, narrowing the peak and its spread across the same dynamic range. Thus, increasing  $\lambda$  in  $\hat{h}_{out}(b)$  will result in a WTM output image,  $\hat{L}_{out}(x, y)$ , that is brighter than its input,  $L_{in}(x, y)$ . Here we use the  $\hat{\cdot}$  notation to indicate a user-defined proxy). If we select a lower  $k$  than the input WD for the target WD, this will cause  $\hat{L}_{out}(x, y)$  to appear to have more contrast than  $L_{in}(x, y)$ , with increased visibility of edge pixels and textures.

## 6.4 Experiments

### 6.4.1 WTM suitability for underwater imagery

In Chapter 5 we established that WTM - matching input distributions to user-defined control-point adjusted images - actually delivered slightly preferred tonal enhancements for analysts, in the sense that the enhanced images help them to identify marine benthic habitats from imagery. Here we explore the usability and performance of WTM implemented as a parameterized tone-mapping tool. We investigate whether users find this approach sufficient compared to classic and manual control-point tone-mapping approaches, that are more complex.

#### Setup

We asked 10 image analysts at our collaborator Gardline to tonally adjust underwater images, using a bespoke GUI (Figure 6.2), so that details required to annotate the content of the image (i.e. the habitat) were made as conspicuous as possible. The same question put to analysts in Chapter 4. Participants were instructed to find a suitable WTM enhancement by manipulating two sliders, that modified the parameters  $\lambda$  &  $k$ , of the target output WD. For user clarity,  $\lambda$  &  $k$  were respectively named as *Brightness* and *Contrast* in the GUI. These terms have two advantages. First, they are intuitively understood by the users;

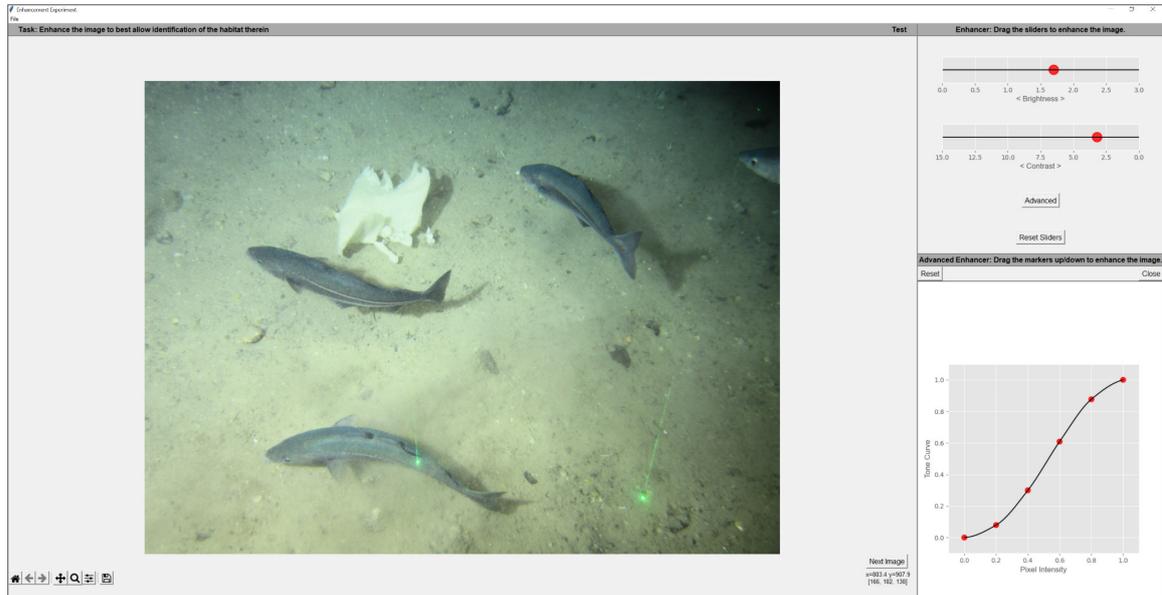


Figure 6.2: Experiment GUI

appearing in almost all image adjustment tools. Second, the effect of WTM on an image often resembles a sort of brightness and contrast change. This is entirely to be expected as the WD parameters,  $\lambda$  and  $k$ , broadly map to the terms brightness (peak of a distribution) and contrast (distribution width or spread). Additionally, as explained in WTM (Section 5.3), the WD of an input image is mapped to the WD of a target and these distributions are described in terms of their peak/brightness and width/contrast. In effect the tone map that modifies one WD to another is effectively making a brightness and contrast adjustment (by definition).

When a suitable WTM (contrast and brightness) adjustment could not be found, an analyst could construct a control-point tone mapping by pressing *Advanced*. This action would plot the current WTM tone map, that they could then modify using 6 control points fixed along the x-axis at  $c \in [0, 0.2, \dots, 1]$ . Only movements that maintained a monotonically increasing tone map were possible. Movements to each control point created a new tone map using a Piecewise Cubic Hermite Interpolating Polynomial, or PCHIP interpolation, through each point. All tone maps in this study were vectors of 256 values in the  $[0,1]$  interval.

Each analyst received GUI training on how to make WTM and control-point tone mapping adjustments, with separate test data, before conducting the experiment. Following this training, they were each asked to adjust (enhance) a unique dataset of 42 images, randomly selected from Dataset 1 (Table 3.1). The random selection contained images of 6 broad habitat classes (7 images per habitat), representing the breadth of biological and physical features expected in the Gardline dataset. These classes are summarised in Table 3.2. As we have 10 analysts, their adjustments resulted in  $42 \times 10 = 420$  unique image enhancements. Augmenting this set, a further *common* sample set of 18 images (3 per habitat class) was shared with each analyst. Each analyst viewed the common sample set twice, creating a further  $(18 \times 2) \times 10 = 360$  enhancements.

This allowed us to investigate the intra- and inter-person variability of the tonal adjustments. We consider the average *relative* extent to which observer's tool parameters varied between image pairs, when they selected the same tool. Parameters here refer to  $\lambda$  &  $k$  for WTM-enhanced images or each of the control-points in the interactively tone-mapped images. We compare variance between observers and parameters using coefficient of variation (*CV*), or normalized standard deviation. *CV* is a standardized measure of dispersion around the mean, calculated as  $CV = \frac{\sigma}{\mu}$ , where  $\sigma$  is the standard deviation and  $\mu$  the mean. A *CV* of zero indicates zero dispersion from the mean and thus equal values.

In each image pair, we determine the *CV* for each parameter. We then summarise by the mean across image pairs (*MCV*), either for each observer (intra) or for all image pairs (inter). For simplicity and brevity, we do not report the *MCV* for each control-point (or tone-map parameter). Instead, we first derive the average control-point *CV* in each image pair (creating one summary measure) and then as before derive the mean of this across image pairs.

This study was conducted under ISO standard 3664:2009 conditions (92); with participants sitting approximately 70 cm from the display in a darkened room. On average, analysts took  $\sim 30$  minutes to complete the experiment.

**Table 6.1:** Mean image quality metrics for input images and enhanced outputs. Mean output metrics are also displayed according to enhancement type: WTM and custom control-point adjustment.

Image	UIQM	UCIQE	CCF
<b>Input</b>	0.68 ( $\pm 0.02$ )	0.66 ( $\pm 0.02$ )	7.84 ( $\pm 0.2$ )
<b>Output</b>	<b>0.74 (<math>\pm 0.02</math>)</b>	<b>0.68 (<math>\pm 0.02</math>)</b>	<b>8.47 (<math>\pm 0.18</math>)</b>
<b>Output<sub>WTM</sub></b>	0.72 ( $\pm 0.02$ )	0.68 ( $\pm 0.02$ )	8.29 ( $\pm 0.19$ )
<b>Output<sub>custom</sub></b>	<b>0.85 (<math>\pm 0.04</math>)</b>	<b>0.72 (<math>\pm 0.06</math>)</b>	<b>9.18 (<math>\pm 0.47</math>)</b>

<sup>1</sup> 95% confidence intervals are shown in parentheses. <sup>2</sup> Best results are highlighted in bold.

## Results

Domain experts in this study overwhelmingly used the 2-slider WTM to adjust the underwater imagery, with 81% of the total image dataset satisfactorily enhanced without recourse to the control-point tone curve adjustment. Moreover, on average individual observers selected WTM for 81% ( $\pm 19\%$ ) of their images. In fact, 60% chose WTM almost ( $\geq 90\%$ ) exclusively in their images. This would suggest that tonal adjustments offered by WTM are highly suitable for enhancing underwater imagery, in accordance with Chapters 4 & 5.

Observers in this study designed their bespoke enhancements to improve image quality to aid image annotation. These enhancements are subjective and tailored and may thus not adhere to aesthetic improvements in the conventional sense, such as those that score highly with objective reference metrics. Indeed objective evaluations of image quality do not always agree well with subjective assessments (112; 78), see Chapter 5. That being said, we find the enhancements in this study are also beneficial objectively. In Table 6.1 we present 3 popular no-reference quality metrics for underwater image enhancement: (1) Underwater Image Quality Metric (UIQM) (134), (2) Underwater Color Image Quality Evaluation (UCIQE) index (207) and (3) a colourfulness, contrast and fog density (CCF) metric (192). For each, a greater value corresponds to a better image quality. Image quality was thus improved in enhanced output images, though not significantly in terms of UCIQE (colourfulness). However this is unsurprising given that the higher values of this metric are

**Table 6.2:** Intra-observer variability across image pairs.

Observer	1	2	3	4	5	6	7	8	9	10	Mean
<b>Agreement</b>	1	1	1	1	0.61	0.94	0.94	1	0.94	0.83	0.93 ( $\pm 0.08$ )
$MCV_\lambda$	0.11	-	0.06	0.07	0.05	0.08	0.08	0.08	0.09	0.03	0.07 ( $\pm 0.01$ )
$MCV_k$	0.11	-	0.13	0.12	0.08	0.13	0.1	0.12	0.12	0.07	0.11 ( $\pm 0.01$ )
$MCV_c$	-	0.19	-	-	0.28	-	-	-	0.14	0.06	0.17 ( $\pm 0.09$ )

<sup>1</sup>**Agreement** = proportion of matching decisions across image pairs <sup>2</sup> $MCV_\lambda, MCV_k$  = mean *CV* for  $\lambda$  &  $k$  in WTM enhanced pairs <sup>3</sup> $MCV_c$  = mean *CV* for control-points in control-point enhanced pairs (*CV* here refers to mean *CV* across 6 control-points) <sup>4</sup>Hyphen denotes when enhancement tool was not selected <sup>5</sup>95% confidence intervals are shown in parentheses

associated with a more colourful image yet the tonal adjustments in this study preserve chromaticity. Assessing the output metrics in more detail, we see that the limited cases of custom control-point adjustments improved image quality more on average than the WTM-enhancements. However, in the following experiment we show that this is independent of the tone mapping method and linked to the enhancements themselves.

Intra-observer variability in this study was low, for image pairs in the common set, see Table 6.2. In row *Agreement*, we detail the proportion of times analysts used the same tool (WTM or control-point) for each of the 18 images in the control set, which they view twice. This showed that they typically selected the same enhancement tool, with 80% of analysts choosing the same tool in >90% of their common images. Mean intra-observer agreement was therefore also high, at 93% ( $\pm 8$ ).

In the remaining rows, we consider the variation in tool parameters between each common-set image pair (for cases where the same tool was selected). Mean coefficient of variation (*MCV*) is used to represent variance; we denote this as  $MCV_\lambda$  and  $MCV_k$  in WTM-enhanced pairs and  $MCV_c$  in control-point pairs. Lastly, if an enhancement tool was not used by an observer, we denote the absent *MCV* values by ‘-’ in Table 6.2.

When using WTM, analysts typically introduced more variation between image pairs, in terms of  $k$ , a proxy for *contrast*, than  $\lambda$ , a proxy for brightness. Comparatively, there was more variation in control-point adjustments between images - with the exception of observer 10 whose control-points varied by  $MCV = 0.06$  on average. Increased *MCV* is

**Table 6.3:** Inter-observer variability across image pairs.

<b>Agreement</b>	$MCV_\lambda$	$MCV_k$	$MCV_c$
0.81 ( $\pm 0.04$ )	0.13	0.21	0.26

<sup>1</sup>**Agreement** = proportion of matching decisions across image pairs <sup>2</sup> $MCV_\lambda, MCV_k$  = mean *CV* for  $\lambda$  &  $k$  in WTM enhanced pairs <sup>3</sup> $MCV_c$  = mean *CV* for control-points in control-point enhanced pairs (*CV* here refers to mean *CV* across 6 control-points) <sup>4</sup>95% confidence intervals are shown in parentheses

unsurprising with the control-point tool, as the potential for difference increases with more parameters (or control-points), here requiring 6 compared to 2 in WTM. These results demonstrate that WTM parameters were more similar and thus analysts adjustments are more consistent.

Variability between observers (inter-personal) is particularly likely in studies such as this, as observers manipulate images according to their own aesthetic preferences. In terms of tool preference, we found, on average, that 81% ( $\pm 4$ ) selected the same enhancement type across images, specifically the favoured tool in the dataset, WTM, see Table 6.3. A comparable proportion to the full dataset.

Using *MCV*, we assessed similarity of selected parameters across images for each tool. As with intra-personal variance, we found that observer adjustments using WTM were more similar across images than those enhanced with the control-point tool, see Table 6.3. For images enhanced with WTM, observers were again more varied in their selection of  $k$ , than  $\lambda$ . For the remaining images, control-point adjustments were, comparatively, more variable on average. Assessing the *MCV* in all of these cases provides only relative comparison, it is not possible to say what is a *good* value. However, a value closer to zero suggests higher similarity of observer adjustments when using WTM. This general behaviour of end-user WTM preferences is promising for future development of an automatic WTM.

### 6.4.2 Simplification of control-point tonal manipulations

In Chapters 4 & 5, it was found that, often, an analyst-defined control-point tone curve could be well approximated by WTM. That is, the control-point tone curve could be interpreted as

the tone adjustment that maps a Weibull approximation of the input brightness distribution to the Weibull approximation of the adjusted brightness distribution. Thus, we wondered if the same result would be found here in the case of control-point adjusted images. In this work, when an analyst cannot obtain a suitable output using the WTM slider adjustment they resort to a custom control-point manipulation. Here we ask whether - like the prior experiments - these control-point adjustments might also be interpreted as a WTM. If this is found to be the case this would imply that the WTM sufficed in general but that finding the best WTM could not always be easily found using the two slider adjustment.

### Setup

To assess whether control-point tonal manipulations could be interpreted as a WTM, we selected all output images from the previous experiment that were enhanced using this *more advanced* tone-mapping tool. Using WTM, we derived the closest Weibull approximation of all images i.e. we computed the WD that best modelled the brightness histogram of the user-optimised output image. Using this Weibull proxy histogram we could derive a new tone curve (the Weibull tone map), which we applied to the input image. Intuitively, if this new image looked similar to the analyst's image, that was control-point adjusted, then we might conclude that the WTM method is working. Instead, it could suggest that the analyst could not find the desired approximation using our tool.

To determine the similarity between an analyst-enhanced control-point image and its WTM approximation, we calculated the colour difference between the image pair using the CIELAB Delta E (or  $\Delta E_{ab}^*$ ) (148) colour difference (Equation 5.1. For complex images, an average  $\Delta E_{ab}^*$  less than 5 indicates that the images are perceptually very similar (114).

### Results

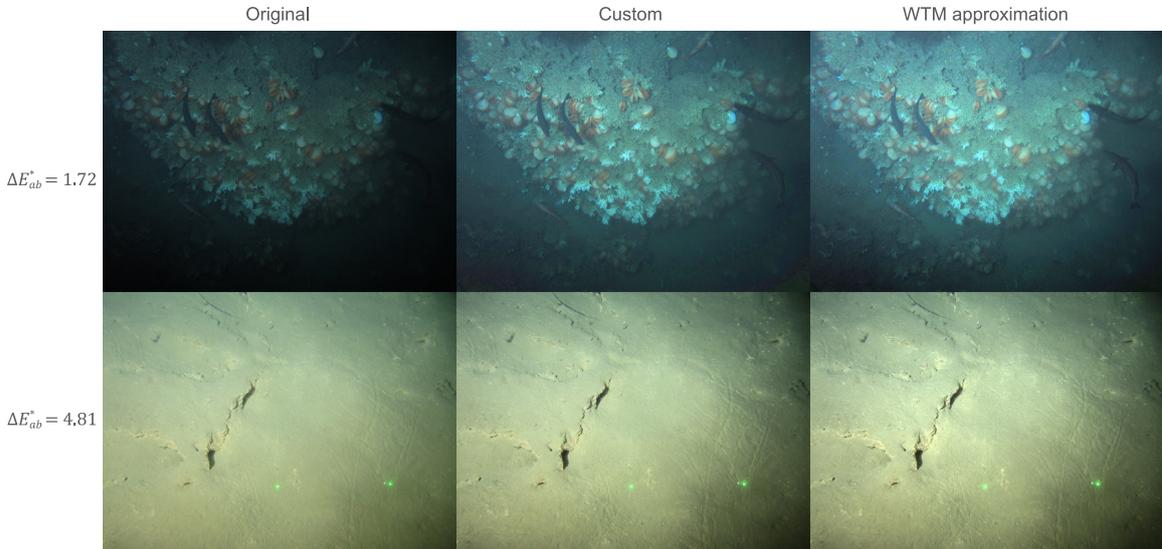
In Table 6.4, we report the colour difference between WTM approximations and their custom-enhanced counterparts. Here, we find that the average colour difference between our WTM- and custom-adjusted image pairs was small, with an average  $\Delta E_{ab}^*$  of 2.98 ( $\pm 0.52$ ).

**Table 6.4:** Summary statistics for mean  $\Delta E_{ab}^*$  (across pixel pairs) between control-point tone-mapped images and their WTM approximation

Mean	% Images <5	% Images <1
2.98 ( $\pm 0.52$ )	91	11

<sup>1</sup> 95% confidence intervals are shown in parentheses

Furthermore, a significant proportion (91%) had a mean  $\Delta E_{ab}^* < 5$ , demonstrating that for the most part control-point tone maps could be convincingly approximated using WTM. In Figure 6.3 we show example images that have been tone-mapped by an expert, and their WTM approximation. The color difference between each pair is  $< 5$  and in each case the differences are near indistinguishable to the observer. These results indicate that when a control-point adjustment is made ( $\sim 20\%$  of adjustments) that these adjusted images can be well approximated by the WTM model (i.e. the closest WTM adjustment results in a similar image visually). Additionally, these successful WTM approximations (mean  $\Delta E_{ab}^* < 5$ ) do not jeopardize image quality in an objective sense, with metrics equivalent to their custom counterpart, see Table 6.5.

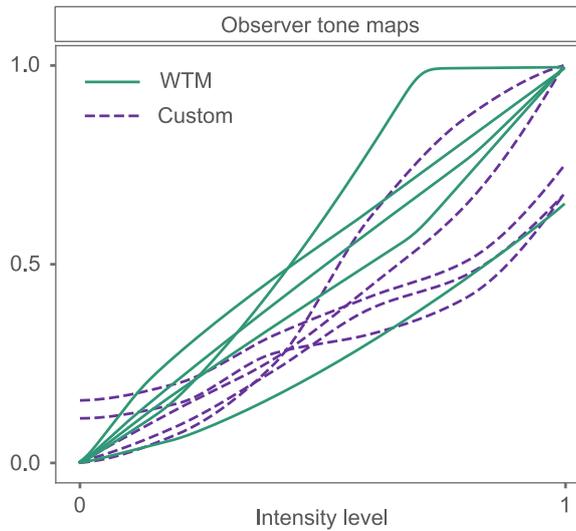


**Figure 6.3:** Example of WTM approximation of control-point control-point adjustments. Colour difference between the custom and WTM image is summarized by mean  $\Delta E_{ab}^*$  across pixel pairs.

**Table 6.5:** Mean image quality metrics for custom tone-mapped images and their successful WTM approximations (mean  $\Delta E_{ab}^* < 5$ )

Image	UIQM	UCIQE	CCF
WTM approx.	<b>0.85</b> ( $\pm 0.04$ )	<b>0.72</b> ( $\pm 0.06$ )	8.94 ( $\pm 0.43$ )
Custom	<b>0.85</b> ( $\pm 0.04$ )	0.71 ( $\pm 0.06$ )	<b>9.14</b> ( $\pm 0.48$ )

<sup>1</sup> 95% confidence intervals are shown in parentheses. <sup>2</sup> Best results are highlighted in bold.

**Figure 6.4:** Example of observer tone-maps using WTM and the custom (control-point) enhancement. Custom tone-maps here can not be approximated by WTM i.e. they did not meet mean  $\Delta E_{ab}^*$  ( $< 5$ ) threshold for approximation.

Control-point tone maps that could not be well matched by WTM were varied in behaviour, see Figure 6.4. In general they appeared to be complex operations, with, sometimes multiple, transitions between very low and steeper gradients; a morphology not consistent with WTM tone maps, which are smoother and simpler. Aesthetically speaking these tone curves can, counter-intuitively, lead to too little and too much contrast in parts of an image and cause it to appear unnatural. However these *outlier* tone adjustments **are** preferred by the analysts as they help them to see details important to identify the image content, in this case the habitat. Thus they need not be aesthetically pleasing. That said there are few outliers and there are only 9% of control-point tone maps not well approximated by the Weibull approximation method. Combining the successful WTM approximations (mean

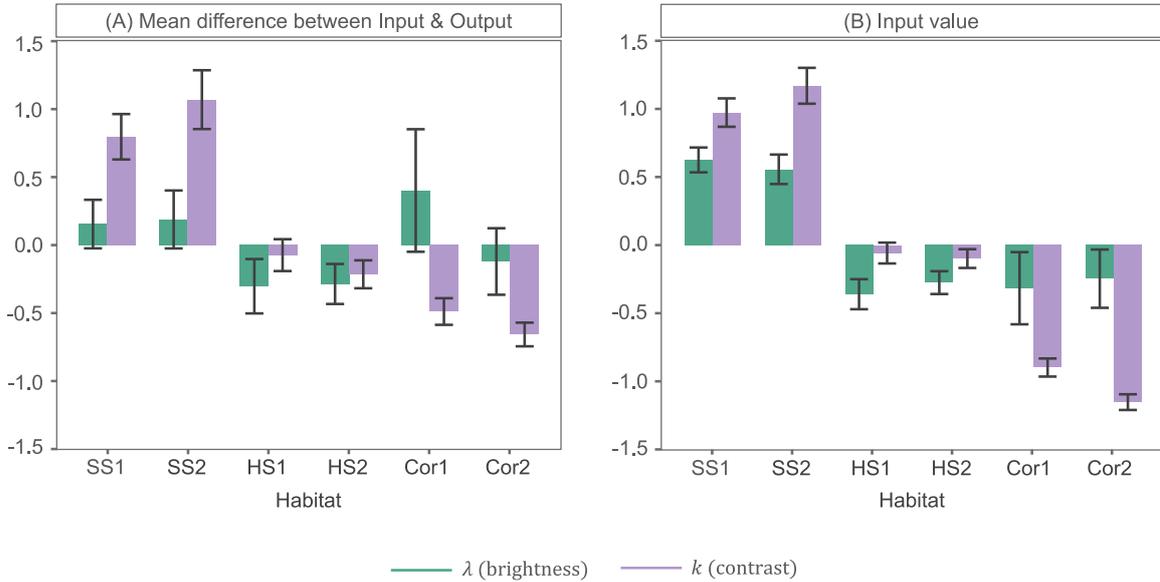
$\Delta E_{ab}^* < 5$ ) with the observer-selected WTM enhancements, we find that for  $>98\%$  of image adjustments there is a suitable WTM to enhance the image, in a way that is useful for the analyst. This offers the promise of supporting analysts to make quicker adjustments of their images, with control-point tone adjustments rarely required.

### 6.4.3 Behaviour of WTM enhancements

The previous experiments demonstrated that the majority of analysts prefer Weibull Tone Mapping to support their benthic image analyses. In this section we discuss these enhancements in more detail, contextualising them by considering the influence of image content. Note that when describing these tonal adjustments, we also include the WTM approximations that had a mean  $\Delta E_{ab}^* < 5$ .

#### Results

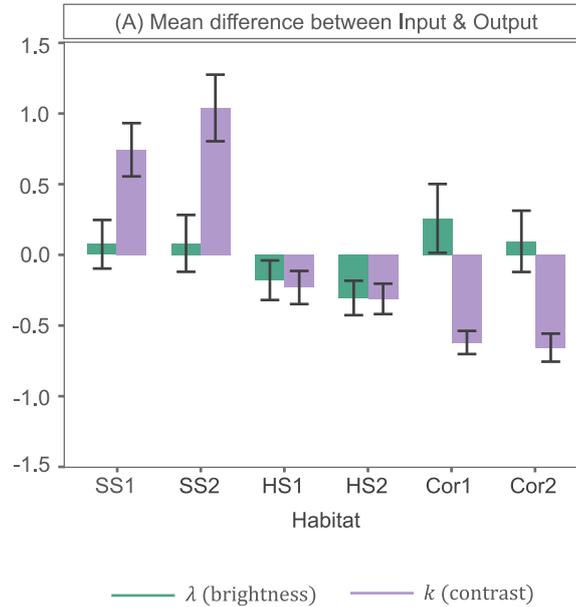
In general, analysts decreased  $\lambda$  between input and output Weibull Distributions in the experiment (61% of images). Decreasing  $\lambda$  (a proxy for brightness) shifts the peak of a brightness histogram towards a lower pixel intensity. This adjustment thus causes a brightness reduction in images following WTM. The prevalence of brightness reductions within the dataset corresponds well with the tonal adjustments made by analysts in Chapter 4. It may have helped to lessen the intensity of the light cone and halo-effect; a common problem in artificially lit underwater imagery. Of these darkened images, 45% were classified as soft substrate (SS) habitats (SS1 & SS2), as described in Table 3.2. In Figure 6.5 (A) we show the mean standardized (Z-Score) difference in WTM parameters, between input and output WDs, for each habitat class in the image dataset. Note that the differences presented for average  $\lambda$ , refer to cases where  $\lambda$  was decreased only. Full results are presented in Figure 6.6. From this we note that not only are SS habitats 1 & 2 representing a large proportion of darkened images, but that the extent to which they are darkened is larger than average.



**Figure 6.5:** WTM parameter variation with image class (habitat): (A) mean difference in WTM parameters between input and output WD and (B) mean WD parameters across input images. All values are presented as standardized Z-Scores for fair comparison.  $\lambda$  and  $k$  differences in (A) refer only to cases in which they were decreased and increased, respectively. Error bars represent 95% confidence intervals.

In Figure 6.5 (B) we detail the average standardized (z-score)  $\lambda$  value across **all** input WDs for each habitat class. From this we see that SS images in this dataset are significantly brighter on average, demonstrated by a high  $\lambda$  value, likely due to the light and homogeneous appearance of sands and muds. Decreasing  $\lambda$ , in output images, may have helped reduce the bright illumination (reflective) effects on the seafloor whilst highlighting the appearance of burrows (a distinguishing habitat feature). Of the remaining images (55%), in which  $\lambda$  was decreased, the other habitats (HS1, HS2, Cor1 & Cor2) were represented roughly equivalently; on average  $\sim 14\%$  each. Yet the extent to which  $\lambda$  was reduced in each was variable, see Figure 6.5 (A). Most notably, the brightness of images classified as hard substrate (HS) (HS1 & HS2, Table 3.2) were, on average, reduced significantly less.

Images of these, more structurally complex, habitats were significantly darker than SS images (Figure 6.5(B)), containing increased shadow presence around topographic features such as boulders and cobbles. Small reductions in  $\lambda$  would help to diminish the unwanted



**Figure 6.6:** Mean standardized (Z-Score) difference in WTM parameters between all input and output WDs. Error bars represent 95% confidence intervals.

brightness effects mentioned prior, yet minimize visibility reduction of distinguishing features within shadowed/darker image regions. Reef framework (Cor1) and soft coral (Cor2) habitats (Table 3.2) can also feature such shadowing due to their pronounced height above the seafloor, in association with topographic highs. Yet these images were, on average, darkened to a greater extent than HS habitats. Corals such as stony coral *Desmophyllum pertusum* in habitat Cor1 and soft coral *Paragorgia arborea* in habitat Cor2, typically appeared very bright (Figure 6.5(B)) and sometimes dominated the field-of-view. A logical enhancement may therefore be to reduce brightness. This is particularly true when increasing contrast, as the intensity of pixels in these regions of interest will be elevated and potentially saturated.

Intensifying contrast is an important tonal adjustment in underwater images (91; 75; 182), to lessen undesirable image effects induced by light absorption in the water medium and scattering due to present particulates (i.e. *marine snow*) (154). It is no surprise therefore that contrast was enhanced in 74% of images, relatively more than analysts in the tone mapping experiment presented in Chapter 4. Analysts achieved this with WTM by de-

creasing  $k$  (a proxy for contrast), which flattens the peak of the brightness histogram. 43% of images in which  $k$  was decreased contained SS habitats (SS1 & SS2). This reduction in  $k$  was also significantly greater for these habitats on average, as shown in Figure 6.5 (A). Images of SS habitats were the most poorly contrasted in this work, indicated by the high  $k$  values in Figure 6.5 (B). Images typically lack edge details and contain more cryptic fauna, such as those that burrow into the sediment. Increasing contrast would therefore allow analysts to improve seafloor texture visualization, in this case concluding absence of gravels and pebbles in order to classify a SS habitat. It would also allow easier searching for sponge presence, to distinguish between habitats SS1 & SS2; a factor equally applicable to discriminate between HS habitats (HS1 & HS2) which represented 37% of  $k$  decreased images. However, as evident from Figure 6.5 (A), contrast was increased significantly less than average in HS images. HS images, as well as the soft coral and reef images (Cor1 Cor2), typically appeared better contrasted (Figure 6.5 (B)) and highly textured. The degree to which  $k$  was reduced in each of these is therefore likely to be less. Few soft coral and reef images received a contrast enhancement by analysts and those that did were adjusted significantly less than average. Enhancing contrast in these images thus appears less important to analysts. In fact, of the 26% of images in which a contrast reduction was enforced, by increasing  $k$ , 66% were classified as soft coral or reef.

A final remark on the behaviour of the WTM adjustments was that 17% compressed the dynamic range. Shrinking the dynamic range will reduce contrast, however it can also be used to lessen the intensity of stark bright and dark regions. It is no surprise therefore that this dynamic compression mostly occurred in images containing soft corals (51%) and Reef (38%). These images are typically dominated by bright corals in the image foreground, surrounded by a very dark background. Reducing these intensities would allow for better visualisation of cryptic features.

As expected, these results demonstrate that the types of tonal adjustments by end-users are logical. They are clearly driven by the underlying brightness distributions of underwater

images, which, in turn, are a product of artificial lighting interactions with the seafloor. For example, a positive trend exists between the brightness of images and the extent to which they are darkened as well as a negative trend between general seabed complexity and the degree of contrast enhancement.

#### 6.4.4 Psychophysical Evaluation

Weibull Tone mapping is designed to aid end-users to make guided adjustments to improve their image annotation of underwater imagery. In this thesis thus far, we have shown that analysts find it extremely suitable and that it can broadly encompass many global tone mapping operations. Here we aim to more deeply evaluate the performance of WTM, and the benefits of bespoke enhancements, by exploring preference over alternative algorithms. In Chapter 5 we considered preference based on enhancements selected from multiple users, finding WTM to be superior over other TMOs. Given the diversity of enhancements deemed desirable by analysts throughout these works, we now consider preference on a more individual basis to investigate whether these preferences are tied to the algorithm or the enhancements themselves. We compare both automatically-generated enhancements and those that are tailored to individuals.

##### Setup

We ran a psychophysical evaluation to assess if analysts prefer their WTM adjustments against other enhancement algorithms. As in previous preference assessments (Section 5.4.2), analysts at Gardline were presented with image pairs and asked to ‘Choose the image that best allows identification of the habitat (class) therein, or no preference if the images are equivalent’. This was conducted under the same viewing conditions, namely, ISO standard 3664:2009 viewing conditions (92); with analysts sitting approximately 70 cm from the display in a neutrally painted and darkened room. Of the 10 analysts that completed the WTM adjustment experiment in Section 6.4.1, only 7 were available to conduct the preference experiment.

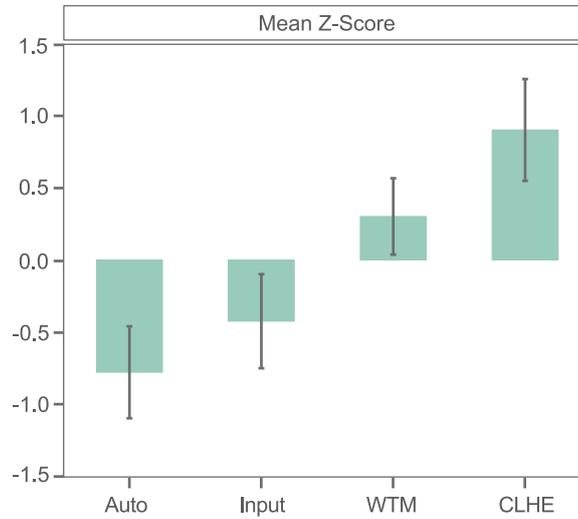
12 images were randomly selected from the common image set in Section 6.4.1, with at least 2 images representing each of the 6 habitat classes. In this experiment we compared  $n = 4$  variants of these images; the unenhanced 'Input' and three enhanced outputs, enhanced either automatically (*magic button*) or with WTM or CLHE. With  $\frac{n(n-1)}{2} = 6$  pair-wise comparisons per image, which is then viewed twice (left-right order switched), the total number of pairwise comparisons for each analyst was therefore  $12 \times 2 \times 6 = 144$  pairs. Note that subject to the availability of image variants in the common set, 1 analyst only viewed  $10 \times 2 \times 6 = 120$  pairs. The order of pairwise comparisons were uniquely randomized for each analyst.

All analysts were presented with identical input and auto-enhanced images, with the latter created using the automatic brightness adjustment in Adobe Photoshop 2023 (5). This produced a tone map which was applied the same as WTM (preserving chromaticity) see Equations 5.9 & 5.10. Although the specific mechanisms of deriving these tone maps are unknown, they always increased the brightness and, to a lesser extent, the contrast.

Alternatively, the WTM and CLHE image variants presented to each analyst were unique. WTM images were those produced by the analyst in the WTM adjustment experiment (Section 6.4.1). CLHE images for each analyst were then created based on the 5<sup>th</sup> and 95<sup>th</sup> slope percentiles of their WTM tone maps (that created the selected WTM images). This resulted in an average upper slope bound of 0.006 ( $\pm 0.001$ ) and lower slope bound of 0.001 ( $\pm 0.001$ ) across analysts. The slope bounds for each analyst were used to derive the CLHE tone map for each input image, which were then applied identically to WTM, using Equations 5.9 & 5.10. Preference vote data was processed using the same methodology as prior experiments, see Section 5.4.2 for details.

## Results

The preference experiment showed that analysts significantly prefer their tailored WTM adjustments to both the unenhanced input images and those enhanced automatically with



**Figure 6.7:** Preference Experiment Scores. Error bars represent 95% confidence intervals.

a *magic button* tonal enhancement, see Figure 6.7. In line with prior experiments in this thesis, this again demonstrates the significance of the Weibull Distribution and WTM in enhancing underwater imagery for annotation.

An interesting distinction in this evaluation however, is the stronger preference signal for CLHE, a contrast-enhancement algorithm. Here CLHE is preferred most of the algorithms, though not significantly more than WTM, whereas in prior work (Section 5.4.2) it was ranked worst. An important difference in this experiment is that CLHE was tailored to the slope statistics of each individual's tonal enhancements, in theory creating an enhancement that they would find more appropriate. Previously we used a more generalised CLHE, with slopes that reflected the tonal enhancements of multiple users. We also previously included assessment by analysts who had not created these enhancements. The fluctuating preference of analysts for CLHE could also reflect the variable contrast manipulations undertaken by analysts throughout this work. In Chapter 4 we saw that analysts' contrast adjustments were split approximately equally between enhancement and reduction. Yet, in this chapter, we saw that analysts often (74%) enhanced contrast to aid image analysis. These results

demonstrate that the suitability of CLHE for improving image annotation is not consistent and heavily dependant upon the input imagery and chosen parameters.

When directly comparing WTM and its *slope-matched* CLHE adjustment, the contrast enhancement offered by WTM will always be more moderate than CLHE, given that the unimodal Weibull Distribution often tails to zero. This likely rendered WTM less suitable to analysts in this particular image sample when higher contrast was prioritized. However, a further reduction in the WTM contrast parameter ( $k$ ) would help to increase contrast further. Although qualities like higher contrast may be broadly desirable in underwater images, which are typically poorly contrasted, there are clearly nuances in analysts preferences and between images. This highlights the difficulty in designing an automatic and general tool, especially in analytical scenarios. It also indicates that enhancement tools should be flexible. This is exactly the use-case for WTM, which extends beyond just a simple contrast enhancement like CLHE. WTM quickly offers diverse tonal adjustments, which adhere to natural underwater image statistics, to better optimise the enhancement process. Unlike CLHE, it can support both contrast reductions and/or modifications to brightness should they be required, as has been the case throughout the experiments in this thesis. Although not optimal in the case of contrast-enhancement when compared to CLHE, WTM is more consistently useful for image analysis and performs well regardless of whether the enhancements belong to the analysts passing judgement. It is also less likely to introduce low-frequency contouring artefacts that can occur with CLHE. These effects are caused by a tone map which contains multiple zero-crossings in its second derivative i.e. the tone map appears stepped or *wiggly*. CLHE tone maps possess this quality, whereas the WTM tone map is extremely smooth.

In general, the psychophysical evaluation in this chapter corroborates well with the general narrative of this thesis which emphasizes the importance of tailored enhancements in underwater image analysis, both with respect to WTM and CLHE. Clear from this work, is that automatic aesthetically-designed enhancements are insufficient (Figure 6.7). As in Sec-

**Table 6.6:** Mean image quality metrics for input images and enhanced outputs

Enhancement	UIQM	UCIQE	CCF
<b>Input</b>	0.69 ( $\pm 0.11$ )	0.69 ( $\pm 0.23$ )	8.11 ( $\pm 1.57$ )
<b>WTM</b>	0.7 ( $\pm 0.05$ )	0.69 ( $\pm 0.06$ )	8.53 ( $\pm 0.56$ )
<b>Auto</b>	0.86 ( $\pm 0.09$ )	0.62 ( $\pm 0.18$ )	9.0 ( $\pm 1.93$ )
<b>CLHE</b>	<b>0.92</b> ( $\pm 0.04$ )	<b>0.77</b> ( $\pm 0.08$ )	<b>10.14</b> ( $\pm 0.71$ )

<sup>1</sup> 95% confidence intervals are shown in parentheses. <sup>2</sup> Best results are highlighted in bold.

tion 5.4.2, this is again in contradiction to objective perceptions of image quality, see Table 6.6. Here WTM is considered a poorer enhancement and the automatic tonal adjustment performs well despite analysts low preference for it. However, CLHE is again rated highest across all image metrics, corresponding well with analyst preferences in this case. These comparisons between objective and subjective ratings further illustrate that caution must be applied as to how we quantify value or quality in an image. These metrics infer quality based on a selection of characteristics such as contrast sharpness as well as loss and degradation of colour. However, in this domain, image adjustments can result in modifications that such metrics would deem to be undesirable and of low quality, although perceptually more useful to analysts. It is therefore prudent to consider both the application and aims and design of the image manipulation when employing such metrics.

## 6.5 Conclusion

Building on our prior work, we have shown here that the Weibull Distribution is highly suitable for both modelling and adjusting the brightness histograms of underwater images. Its properties are driven by two reasonably intuitive parameters which roughly conform to the brightness (peak) and contrast (spread) of a distribution. These preserve well the natural behaviour of pixel intensities in underwater imagery, but can also provide enhancements to support annotation, through their modification using our Weibull Tone Mapping (WTM) algorithm.

As with previous chapters, we demonstrate here that the characteristics of analyst tonal enhancements are tightly linked to the content of underwater imagery and their associated brightness distributions, and the purpose the imagery serves (in this case habitat annotation). Although time-consuming, control-point tonal manipulations offer end-users with more complex and targeted tonal enhancements. That being said, this work showed that given the choice, annotators rarely opt for control-point tonal manipulations. Instead they prefer to utilise WTM, specifying desirable Weibull brightness distributions by simply manipulating its 2 parameter-sliders. Furthermore, in the few cases where a control-point tone map was sought, the majority behaved like WTM.

Since most analysts enhance imagery according to the Weibull distribution, WTM is a useful mechanism to grant analysts the ability to modify an image such that it maintains Weibull properties. It also strikes a good trade-off between the flexibility of a bespoke control-point manipulation and the simplicity and speed of an automated tone-mapper. Thus it can easily be used as a *live* modification tool alongside annotation. WTM also lends itself well to future automation, by highlighting properties that an automated enhancement should have to support underwater image analysis. Analysts found these properties to be significantly better than those produced automatically. In addition, these properties are more consistently desirable across imagery than those offered by CLHE, a popular tone mapping tool.

We note that although the design and current function of WTM is domain-specific (underwater imaging), its usage could extend outside of this scope. For example, future work could investigate its performance on medical imaging or images collected for terrestrial- or aerial-based ecological surveys.

# 7 Simplifying automatic classification of benthic habitats

In the previous chapters we focused on supporting analysts to extract a reliable ground-truth from underwater imagery. We centered on a major obstacle in this task, image quality, investigating the type of tonal enhancements required by analysts and developing a tool on this premise. In this chapter, we move to another challenging component in their analysis pipelines - coping with the processing bottle-neck. We explore automatic methods to interpret underwater imagery, in order to manage the time-consuming process of classification and enumeration in large datasets. We present methods that offer a suitable trade-off between simplicity of operation and comprehension for non-specialist users and good, consistent performance.

## 7.1 Introduction

Optical imaging provides a non-destructive and high resolution tool to monitor the biological and conspicuous physical attributes of the seabed. Regardless of the platform, there are some common truths suffered; (1) annotation of resulting imagery is often inconsistent and error prone due to observer bias, fatigue, distraction and short-term memory limitations (31) (2) it is costly (in the absence of volunteers) and (3) labour-intensive. This reality is particularly realised with AUV usage, in which one survey (~50 hours) can produce over 170,000 images (200).

Machine learning (ML) solutions, to automate the processing of image feature identification and enumeration, are therefore generally believed to be essential to alleviating this

problem. Their use has become increasingly popular in ecology over the last few decades (195). In particular, Convolutional Neural Networks (CNNs), a ML algorithm, have shown promise (109; 108). Although their implementation has traditionally been more limited to the Computer Science community, there are now a number of other scientific disciplines, including Marine Ecology, that are beginning to explore their application. For examples of CNN applications in Marine Ecology, please refer to Section 2.3.2

However, CNNs are highly complex. Without close collaboration with experts, they are often inaccessible for a non-specialist (183). This is mostly as they require large investments in time (for theoretical comprehension and implementation). They are intimidating for those without more extensive programming experience, particularly in languages such as Python and Matlab. Additionally, their demand for high specification machines and 'Big Data' may present barriers to their use. Effort has been made to incorporate automatic classification tools into annotation softwares such as BIIGLE (107), VIAME (35) and CoralNet (25). However these may be unsuitable for those requiring more flexibility in ML approaches or those wishing to integrate ML functionality into their custom annotation programs, as is the case with our collaborator Gardline. They are also tailored for certain automation tasks.

For the reasons stated, we demonstrate here a simpler approach that is more accessible to end-users and focuses on an under-represented application in the literature, deep-sea benthic habitats. Specifically, we use an 'off-the-shelf' CNN (pre-trained and open-source), VGG16 (170), to extract features from contextually-representative benthic images. Image features in this case are pixel patterns that correspond to low-level objects such as edges, circles and lines up to high-level features such as sponge branches for example. These are then used to train a simpler classification algorithm, a Support Vector Machine (SVM) (28; 30). SVMs have been shown to pair well with 'off-the-shelf' CNN features generally (144; 11; 153), as well as in benthic image applications (120; 121; 122; 119; 132). They are a classical method, well documented and offer a good trade-off in terms of complexity,

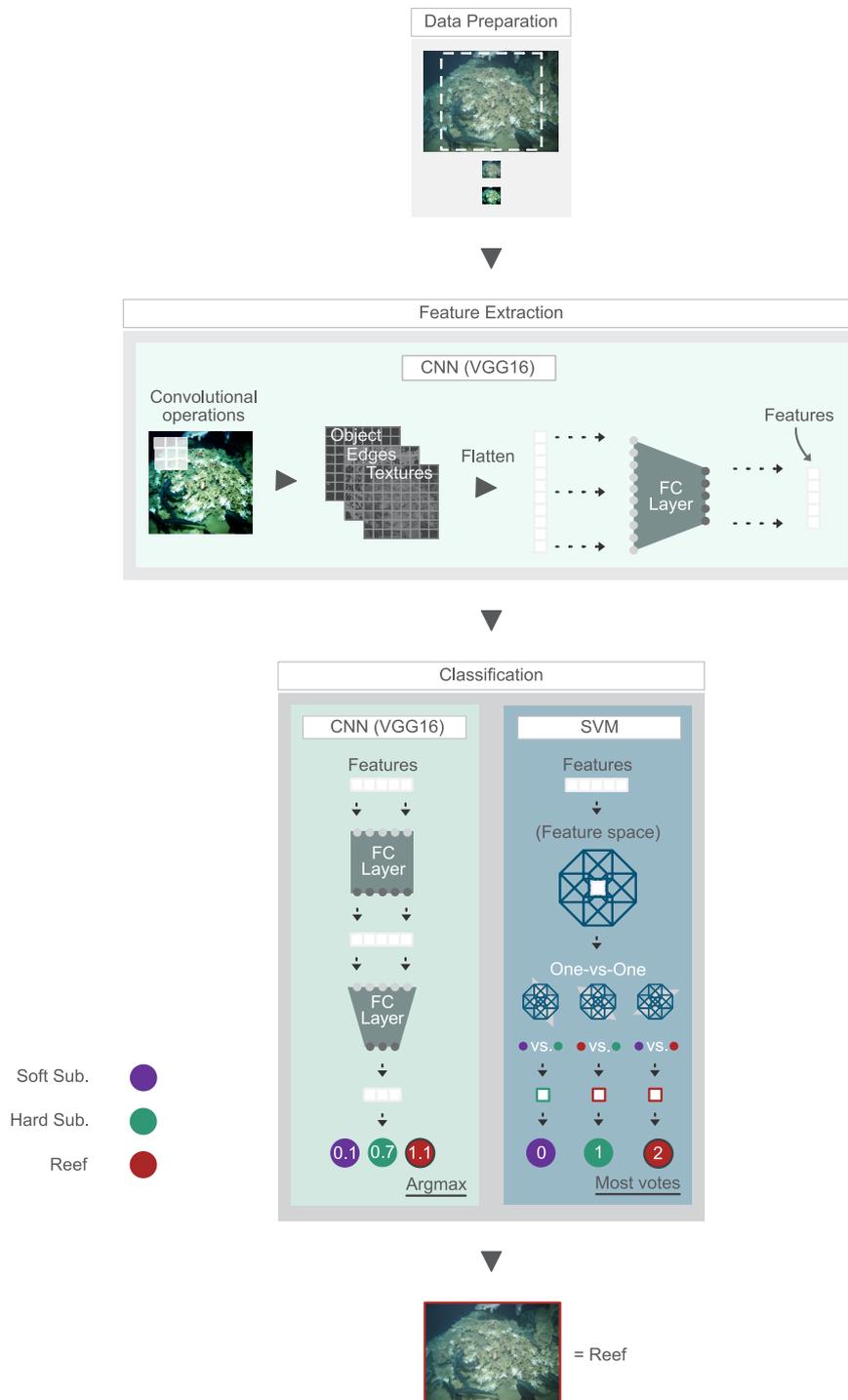
performance, computational demand and time. We compare this approach to training a complex CNN classifier for the same task and demonstrate that, for this application, they have the advantage.

The purpose of this chapter is to act as both a quick primer for novice users, facilitating comprehension and implementation of these approaches, whilst also serving as proof of concept. We demonstrate this extends to multiple datasets that vary in geographic region and imaging platform.

## 7.2 Background

In this chapter, we present two approaches to automate classification of benthic habitats from optical imagery. Each approach consists of two main phases: feature extraction and classification. Simply put, feature extraction learns to find and highlight patterns associated with pixel information (features). The classification phase then learns to link these features to one of the three habitat classes. The first of our methods uses the feature extraction components of a Convolutional Neural Network (109; 108) paired with a Support Vector Machine classifier. The other, a more complex approach, uses a full CNN architecture for both feature extraction and classification. Thus, both approaches in this work differ only with respect to classification. For brevity, these methods are henceforth referred to as **CNN+SVM** and **CNN**, respectively. Please refer to Figure 7.1 for a graphical representation of each.

In the following sections, we provide further details on each ML approach and evaluation metrics. For more detailed background on the the mechanics of each algorithm, please refer to Section 2.3.



**Figure 7.1:** Graphical representation of ML approaches (CNN & CNN+SVM) used in this chapter

### 7.2.1 Machine Learning Approaches

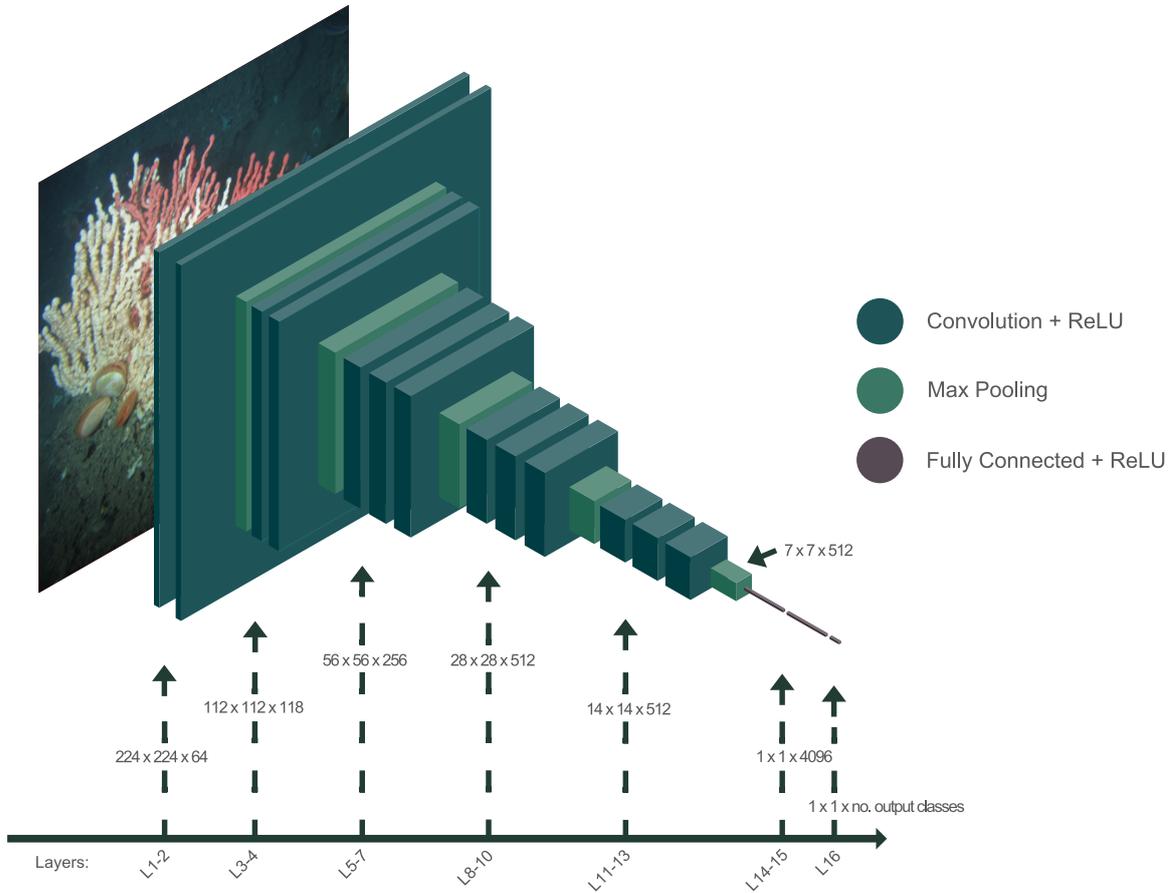
Common to both of our ML approaches is the use of a CNN. Simply put, CNNs transform image data through a process known as convolution to extract features (108; 66; 9). These features are then 'connected' to a final layer of outputs that determine the classes that need to be classified. This last layer is referred to as fully-connected (FC) i.e. all outputs from the penultimate (feature) layer are connected to all classification nodes.

CNNs can be trained to extract and classify image features extremely well (102; 170; 176; 74; 86). However, this training requires vast amounts of imagery (1000s) and computational resources that are not attainable for many scientists such as expensive Graphical Processing Units (GPUs). It also requires experience to monitor training and select appropriate hyperparameters, see Table 2.1.

In our work we will therefore use transfer learning; taking a pre-trained CNN and simply training it a little more. For this we use VGG16 (170), which was trained on the ImageNet dataset (36), see Figure 7.2.

Transfer learning with networks pre-trained on ImageNet has increased performances across a multitude of image classification problems, including benthic imagery (120; 132), and is thus a common component of computer vision pipelines. This is widely speculated to be due to the vast size of ImageNet ( $\approx 1.2$  million images) and the large number (1000) of classes which include both similar and distinct objects (87). However these simple assumptions may not be entirely accurate as empirical investigations are limited (87).

Logically, the performance of transfer learning is in part influenced by the similarity of the training and test data. However, this pre-trained VGG16 is not optimised to recognise features in our own datasets, with the majority of ImageNet classes irrelevant to our underwater imaging application, covering groups such as plants, vehicles, food and animals. However we can still use it as a 'head-start', leveraging the power and success of prior training to extract more general features such as edges, lines, corners and simple textures



**Figure 7.2:** VGG16 architecture

(144; 209; 120; 121). Additionally, there are a number of underwater classes including ‘corals’ and ‘sandbar’ which are broadly more similar to our benthic habitat classes. This suggests that the learnt feature representations from ImageNet may be applicable to our alternative domain beyond a general sense, recognising more complex geometric and textural qualities of seabed features such as corals and sand.

There are several reasons for choosing VGG16 as a basis for transfer learning. VGG is one of the most implemented algorithms for image classification and although several years old remains highly popular with high performance across diverse image applications applications (96; 101; 206; 205; 8; 3). More importantly, it also has a record of good performance across marine classification tasks (117; 211; 65; 98; 121; 122). Preliminary comparisons of

VGG16 to other models (AlexNet (102), ResNet18 & ResNet50 (74) and VGG19 (170)) found VGG16 to produce deep features that were more accurately classified by an SVM. Additionally, for inexperienced users, our target audience in this application, a model architecture such as VGG16 may be preferential over newer state-of-the-art model architectures. This is due to its inclusion in more accessible platforms/frameworks such as PyTorch (139) and TensorFlow (2), its extensive guidance materials for implementation and well-documented high performance.

The VGG16 architecture includes 16 layers with learnable weights, of which 13 are convolutional and 3 are FC. Convolutional layers use a  $3 \times 3$  kernel and are generally followed by a Rectified Linear Unit (ReLU) non-linear activation layer. Down-sampling is also performed after some of the convolutional layers using max-pooling (kernel =  $2 \times 2$ , stride=2). Within the classification component of the network, the first two FC layers have 4096 nodes each whereas the third contains 1000 nodes, corresponding to the number of classes within the training dataset (ImageNet).

Although a powerful ML technique, using a CNN is still intimidating to the novice, both practically and intellectually speaking. Several hyperparameters are still required in transfer learning 2.1, which must be set manually by the user. It also requires a good deal of theoretical study to understand, train and use them effectively. For this specific application, we demonstrate that this trade-off between complexity and its performance is unwarranted.

Shallow-learning SVMs are by comparison far simpler to train. Designed for binary classification problems, they work by finding the best boundary, or hyperplane, between data points that enables distinction of two classes. However, strategies exist to handle multi-class classification ( $>2$  classes). In this chapter, we use One-Vs-One (OVO). This splits the dataset into multiple binary classification problems that are assessed per each pair of classes. Compiling the classifications of all binary SVMs allows a final classification to be made for each data point, based on the class that received the most votes.

There are a number of reasons that encourage the use of SVMs, both widely and in our application. Importantly, SVMs guarantee an optimal solution to be found when classifying data, unlike CNNs. SVMs also require less hyperparameters. In our work we explore both a linear and non-linear (Radial Basis Function) SVM, which require 1 and 2 parameters, respectively, see Table 2.1. Given, the few hyperparameters SVMs can easily be re-trained with different hyperparameter combinations to find an optimal set. They also use only a subset of the data (support vectors) to place the hyperplane and are thus computationally efficient. By comparison, FC layers in the CNN use all of the data and require a large number of intricately connected parameters, thus they are prone to over-fitting and poor generalization (102; 9). Using a SVM classifier, in combination with a CNN feature extractor is thus less restrictive with regards to resources (time, computational, & imagery) and experience.

### 7.2.2 Performance Evaluation

A number of metrics can be used to evaluate the performance of the trained models at classifying classes in test-datasets. Perhaps the simplest of these measures is accuracy, which indicates the number of images that the model classifies correctly, with 1 marking perfect (100%) accuracy. However, this does not reveal information as to its performance within classes. To account for this, classification studies also calculate the recall, precision and F<sub>1</sub> score for each class. Recall, or sensitivity, refers to the proportion of images of each class that were correctly classified. Precision however determines the accuracy of the predictions themselves; it measures the proportion of images that were assigned a correct class when classified. For each habitat class  $c$ , Recall and Precision are calculated as:

$$Recall_c = \frac{TP_c}{TP_c + FN_c} \quad (7.1)$$

$$Precision_c = \frac{TP_c}{TP_c + FP_c} \quad (7.2)$$

where  $TP$  refers to the number of true positives (those correctly classified as class  $c$ ),  $FP$  the false positives (those that are incorrectly classified as class  $c$ ) and  $FN$  the false negatives (those of class  $c$  that are incorrectly classified). Dependant on the classification problem, it may be desirable to have a higher performance for recall or precision. For example you may wish to ensure the model can identify all images of a class or instead maximise its correctness in identifying a class, for recall and precision respectively. For a more balanced view of model performance we calculate the  $F_1$  score, a harmonic mean of the precision and recall. This gives an idea of how well the model recognises images of each class and distinguishes between images of other classes. For each class,  $F_1$  is calculated as:

$$F_{1c} = \frac{2 * (Recall_c * Precision_c)}{(Recall_c + Precision_c)} \quad (7.3)$$

For each of the class metrics used (Recall, Precision and  $F_1$ ), we present an average across classes, also known as the macro-average, alongside their 95% confidence intervals (CI). The macro-average for each metric  $m$  is calculated simply by:

$$Macro - average_m = \frac{\sum_{c=1}^C m_c}{C} \quad (7.4)$$

where  $C$  is the total number of classes. The macro-average weights class importance equally, irrespective of the number of images associated (instances of each class), and therefore represents model performance more reliably. This is particularly useful given that the deep seafloor is largely a soft-sediment habitat with intermittent hard-substrate and reef and thus a class imbalance is typically present in image surveys. Our datasets are no stranger to this phenomenon.

All performance metrics were calculated using the python library *scikit-learn* and are scored between 0 & 1, with optimal performance reached at 1.

## 7.3 Experiments

The following experiments use the pre-trained VGG16 network, to extract deep features from benthic images. We show that feeding the default outputs of VGG16 into a Support Vector Machine (SVM) is a high-performing and consistent ML approach, that focuses on conceptual and practical simplicity to maximise accessibility. For comparison, we also use the VGG16 network directly, in which we re-train the final 2 layers of the net (i.e. we compare against a classical deepnet methodology), though its performance does not warrant the additional complexity.

### 7.3.1 Extracting and interpreting a deep feature space

In this experiment we explain how to employ VGG16 as feature extractor and explore the resulting features. These features are numerous and near-impossible to interpret. This is especially difficult for novice users from a ecological background, who routinely apply field-specific knowledge or experience to aid interpretation of a model's features or input variables. It is also challenging if you wish to assess the underlying suitability of the features for automatic classification. We therefore explore dimensionality reduction of the deep features, as a tool to visualize the feature space and any patterns or clusters within.

#### Setup

The pre-trained VGG16 model was sourced from the *torch* library, the basis of the Pytorch ML framework. The network's parameters were initially frozen to prevent further training. To create a feature extractor tool, the majority of VGG16's classification component is removed i.e all layers following the first FC layer are removed, see Figure's 7.2 & 7.1. Every image that passes through this feature extractor will thus result in a matrix of  $1 \times 4096$  features that will later be passed to a classifier i.e. an SVM.

VGG16 was used to extract deep features from three datasets to explore the generality (and performance) of the pipelines presented. Details of each are listed in Table 3.1. Datasets

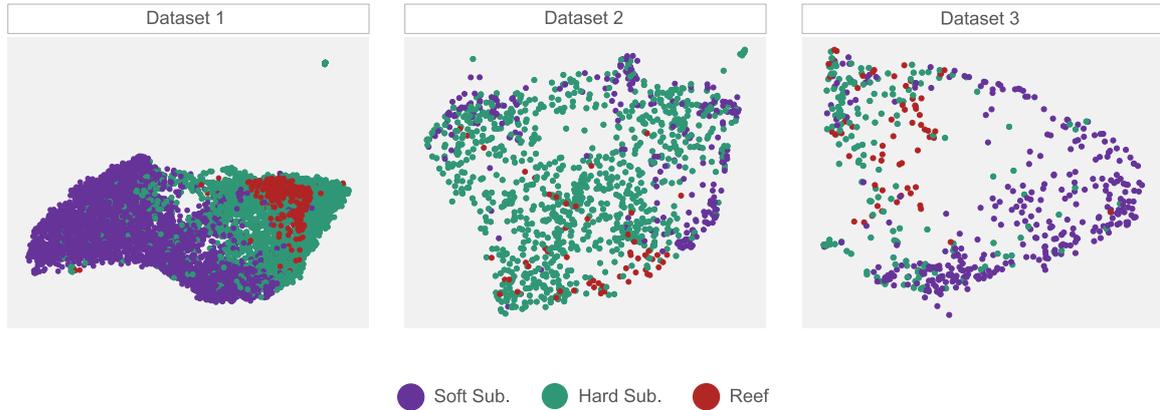
vary according to imaging platform and geographic region. For the purpose of this chapter, we focus on the broad-scale habitat descriptions listed in Table 3.3: 1) Soft Substrate (SS) with SS sponge communities, 2) Hard Substrate (HS) with HS sponge communities and 3) Reef. For brevity, we refer to these as Soft Sub., Hard Sub. and Reef.

Pre-trained CNNs are designed to expect image data in a certain format before feature extraction (or classification). For VGG16 (and other networks trained on ImageNet (151; 74; 79)), RGB images must be 224×224 pixels. We therefore resize images to 224 pixels along the x-axis, preserving their aspect ratio. We then crop the center of images such that they are square. Following standard practice, RGB values were also normalized (centered and scaled) with respect to the ImageNet dataset.

Before undertaking any classification of the extracted high-dimensional feature space, we took steps to better conceptualize it, reducing the number of features (or dimensions) to  $\leq 3$ . Although the features themselves remain cryptic, this allows us to visually assess any structure or clustering in the data that would enable good separation of classes and thus a high performing model. For this task we use a Pairwise Controlled Manifold Approximation (PaCMAP) (191), employed using the python library *pacmap*. This technique quickly finds a low-dimensional representation of the complex feature space (that is structurally most-similar) by calculating Euclidean distances between pairs of observations. By identifying neighbouring pairs, mid-near pairs and further pairs, each with lessening 'attractiveness' or similarity, they optimize a mapping that best maintains these point-based relationships. By maintaining the relative 'attractiveness' of these pair groups, PaCMAP can preserve the local and global structure of the data. The technique is fast and simple to use and has been proven to accurately capture data distributions (191).

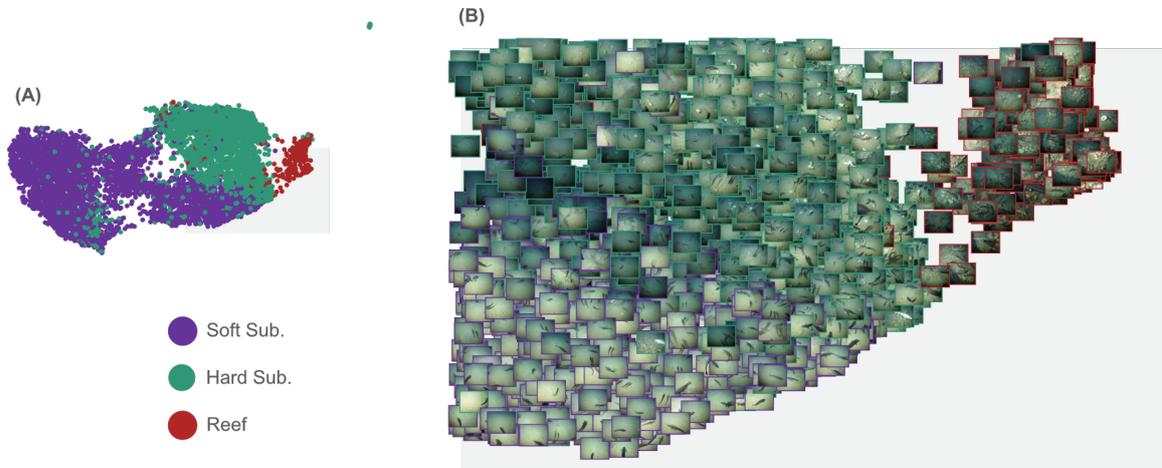
## Results

The VGG16 network activations resulted in a feature space with conspicuous clustering of each habitat, see Fig. 7.3. This suggests the general features provided by the VGG16



**Figure 7.3:** 2-Dimensional visualisation of the 3-Dimensional feature spaces for each dataset, created using PaCMAP

network are suitable for classifying benthic habitats, though the extent to which may be variable across the Datasets used and classes encountered. Across all datasets, the PaCMAP dimensionality reduction showed the strongest partition of Soft Sub. & Hard Sub., aligning with the distinct visual contrast of these habitat classes. Features associated with Reef habitat were also clearly separable from Soft Sub. habitats. However, they do exhibit a degree of overlap with Hard Sub., in 3 dimensions. This may indicate higher similarity with this habitat, with respect to the general features extracted. The overlap of Hard Sub. and Reef feature spaces, particularly within Datasets 2 & 3, is unsurprising given the ecological association of Reef and hard substrates (147; 83). Larvae of coral species such as *Desmophyllum pertusum*, responsible for creating the reef framework in the featured datasets, require hard substrate to settle and build their communities (198; 147). Reef is also more commonly associated with topographic highs (created by boulders and bedrock for example) that enable elevation into currents for enhanced filter feeding (179; 130; 147). Imagery classified as Reef will thus likely contain evidence of hard substrates. Some images would also have featured the sub-class coral rubble, as described in Table 3.2. In a technical sense, the texture and edge distributions of coral rubble in imagery, may exhibit more visual similarity to seabed areas dominated by gravel.



**Figure 7.4:** PaCMAP 2-Dimensional visualisation of Dataset 1 (Table 3.1) feature space: a) 2D point cloud and b) zoomed region in which points are replaced by image thumbnails for better conceptualization of feature variance.

Of the datasets, habitat clusters were particularly notable within Dataset 1, at both higher (3D) (Figure 7.3) and lower (2D) dimensions (7.4). In Fig. 7.4 (B), image thumbnails demonstrate the visual transition between image characteristics of each class for better comprehension of the feature space. A gradient of seabed complexity is clearly evident from the bottom left to the top right of the frame. The proximity of images, within and between classes, can be seen to not only correlate well with seabed characteristics but also shows some clustering based on illumination patterns.

It is important to note that although PaCMAP preserves the structure of high-dimensional feature spaces well (191), this is a simplification. However, we still see reasonably clear class associations across datasets; this could suggest that the class feature spaces may be even more easily separable at higher dimensions.

### 7.3.2 Classifying deep benthic features

The PaCMAP dimensionality reduction demonstrated that the dataset features spaces do exhibit clustering by habitat class and thus should support well-performing automatic classification. In this experiment we evaluate classification performance across the three datasets

showing that combining an SVM classifier with a CNN feature extractor is a high-performing and consistent ML approach and not a significant trade-off with its simplicity. Specifically, we find a non-linear RBF SVM to be most suitable.

### Setup

Interaction with the ML models in this work is conducted using different Python libraries. CNN training and testing is conducted using *torch* whereas *scikit-learn* is used for the SVM. To keep our model training and analysis pipelines equivalent, we therefore use *skorch*, a *scikit-learn* compatible neural network library that wraps PyTorch. *skorch* has a clear and simple interface. Users need only add the prepared datasets and model, and specify the associated hyperparameters (Table 2.1). It then packages the CNN model and allows the same *scikit-learn* training and evaluation procedure to be used for both models. Documentation is available at (171).

To classify benthic habitats from imagery we used the ML workflow summarized in Table 7.1. For each dataset (Table 3.1), images were split into 80% training (including 5-fold cross-validation) and 20% testing subsets. Splits were stratified to preserve the class-ratio. For model preparation, we load a pre-trained VGG16 with ImageNet weights and freeze all layers to prevent further training. For our **CNN+SVM** modelling approach, we duplicate the network and retaining the architecture only up to the first FC layer (FC1). Note that this is the same as the CNN feature extractor developed in Section 7.3.1. We then pair it with a SVM, either a linear SVM or a non-linear RBF SVM; we evaluate both. Whereas for our **CNN** method, we train only the final two FC layers of the downloaded VGG16 network. Thus we unfreeze (or replace) layers FC2 & FC3 to re-initialize the weights for training. We also alter the number output nodes in FC3 to 3, the number of classes in the datasets.

Each of the ML approaches require hyperparameters to classify imagery. Optimizing model hyperparameters during training can increase model performance. For our **CNN+SVM**

**Table 7.1:** Required steps in our ML workflows

<b>Data Preparation</b>
Organise images into class folder structure
Separate image dataset into training and testing
Resize imagery for CNN <i>i.e.</i> 224×224
Normalize images to ImageNet dataset (center & scale)
<b>Model Preparation</b>
Load pre-trained VGG16 (with ImageNet weights)
Freeze all layers
For <b>CNN+SVM</b> :
Duplicate the network
Remove all classification layers, except FC1
Initialize SVM with default parameters
For <b>CNN</b> :
Unfreeze (replace) FC2 & FC3 in classifier
Set FC3 outputs to the number of classes <i>i.e.</i> 3
Set hyperparameters <i>i.e.</i> <i>learning rate, optimizer, loss function</i>
<b>Feature Extraction</b>
Set batch size for feature extractor
Extract features from training and testing set
<b>Classification</b>
For <b>CNN+SVM</b> :
Optimize SVM hyperparameters with grid-search
Train SVM with optimal choice
For <b>CNN</b> :
Optimize learning rate during training
<b>Performance Evaluation</b>
Assess training performance
Predict classes in test data
Evaluate test performance

<sup>1</sup>Note that the order of steps may vary with ML frameworks (i.e. *pytorch, tensorflow*) or may be achievable simultaneously <sup>2</sup>Grey shading denotes model-specific steps

method, we follow recommendations by (84), authors of the *LIBSVM* library (24). We use a non-linear RBF kernel for our SVM and run a  $k$ -fold ( $k = 5$ ) cross-validated fine grid-search (on the training data) of the parameters  $C$  and  $\gamma$  i.e.  $C = 2^3, 2^{3.25}, \dots, 2^7$  and  $\gamma = 2^{-15}, 2^{-13}$  &  $2^{-11}$ . For the linear SVM, we use the same parameter search for  $C$  (its only required parameter). For comparison, we also look at the RBF and linear SVM with default *scikit-learn* values,  $C = 1$  and  $\gamma = 1/(f \times Var(F))$ , where  $f$  is the number of features,  $F$  the features and  $Var$  the variance. Each fold in the training dataset is used once as a validation set, while the  $k - 1$  remaining folds form the new training set.

Hyperparameter combinations that received the highest average accuracy (proportion of correct classifications), across validation folds were selected. This is a simple and relatively fast process, subject to the size of the dataset.

Training a CNN is far more difficult and time-consuming. It requires monitoring of performance (accuracy and loss) across epochs to ensure the model is effectively learning to classify the feature space. It may then be manually interrupted upon issue or when a suitable model is developed. The large number of hyperparameters associated (Table 2.1) and high computational demand, mean that an exhaustive grid-search is inappropriate. Instead our preliminary work showed that recommended default parameters, were suitable for our data. These include a batch-size of 32 images and a cross-entropy loss function. We also used the Adam learning rate optimizer (97). The initial value was set to 1e-03 and then automatically adjusted during training, in a way that improves performance. Adam is computationally efficient and straight-forward to use; you need only specify its use and the rest is handled. In preliminary work, each model was set to train for 100 epochs maximum. However for later time-saving and better automation, we enabled early stopping if the validation error (loss) does not reduce for 10 epochs. This identified a suitable number of epochs for each dataset in the main experiment, 14, 14 and 23 epochs for Datasets 1, 2 and 3, respectively. As with SVM training, we conduct the CNN training protocol using 5-fold cross-validation of the training set; the same k-folds as the SVM training.

Following cross-validation of each modelling approaches, mean validation accuracy across  $k$ -folds was determined. We then undertook final training on the entire training dataset with the best hyperparameter combinations (where relevant). Feature extraction and classification were supported by an NVIDIA GeForce RTX 2080 SUPER GPU (8GB VRAM) and an Intel Core i7-9700 CPU.

## Results

In line with other research (120; 121; 122; 119; 132), we find that general features extracted from imagery by 'off-the-shelf' convolutional neural networks are suitable for classifying benthic imagery with few errors. Of the SVMs compared, we found a non-linear RBF kernel was best suited across datasets, factoring in both validation and test accuracies. Performance was also improved with the optimisation of hyperparameters. In Table 7.2 we list final training accuracy and mean accuracy across the 5-fold cross-validation (following hyperparameter tuning), for each dataset and SVM approach. Training accuracy was always highest for the linear SVMs, regardless of whether the hyperparameters were tuned, however the comparative performance across validation sets was almost always lower, on average by  $2.4x$ , than when employing a non-linear SVM. With the addition of hyperparameter tuning, mean validation accuracy was always highest with an RBF SVM. These results indicate that the Linear SVM may be over-fitting more to the training data.

In Table 7.3 we show the final test accuracy as well as the class-averaged Recall, Precision and  $F_1$  scores. Overall accuracy on the test set was similar to the mean validation accuracy. In each dataset, best performance was found with a hyperparameter-tuned RBF SVM, with an average improvement of 3.3% over Linear SVMs and 1.6% over default RBF parameters. Performance was more variable across the class-averaged metrics. Linear SVM performance was fairly balanced across metrics. RBFs were found to be more precise in their predictions by comparison, sometimes at the expense of recall. For interpreting marine imagery, it is more important that predictions are correct (i.e. high precision) than identifying all images of a class (i.e. high recall). These results therefore suggest that a good SVM approach for our datasets is a non-linear RBF kernel. We note that highest RBF precision across datasets is found using default hyperparameters, however the recall (for dataset 2 & 3) is extremely low by comparison. By tuning RBF hyperparameters we yield higher performance than a Linear SVM and do not compromise the recall as much.

**Table 7.2:** Classifier training performance: with tuned hyperparameters, mean cross-validation (CV) and final training accuracy

Set	Images	Classifier	Hyperparameters	Accuracy	
				Mean CV	Train
1	6682	SVM:Linear	-	0.93 ( $\pm 0.003$ )	<b>1.00</b>
		SVM:Linear <sub>d</sub>	$C = 2^{3.0}$	0.93 ( $\pm 0.003$ )	<b>1.00</b>
		SVM:RBF <sub>d</sub>	-	0.95 ( $\pm 0.003$ )	0.97
		SVM:RBF	$C = 2^{3.0}, \gamma = 2^{-15.0}$	<b>0.96 (<math>\pm 0.003</math>)</b>	0.98
		CNN	$lr = 0.001$	0.92 ( $\pm 0.023$ )	0.98
2	992	SVM:Linear <sub>d</sub>	-	0.87 ( $\pm 0.010$ )	<b>1.00</b>
		SVM:Linear	$C = 2^{3.0}$	0.87 ( $\pm 0.010$ )	<b>1.00</b>
		SVM:RBF <sub>d</sub>	-	0.87 ( $\pm 0.018$ )	0.91
		SVM:RBF	$C = 2^{3.75}, \gamma = 2^{-15.0}$	<b>0.91 (<math>\pm 0.016</math>)</b>	0.97
		CNN	$lr = 0.001$	0.90 ( $\pm 0.015$ )	0.97
3	459	SVM:Linear <sub>d</sub>	-	0.83 ( $\pm 0.021$ )	<b>1.00</b>
		SVM:Linear	$C = 2^{3.0}$	0.83 ( $\pm 0.021$ )	<b>1.00</b>
		SVM:RBF <sub>d</sub>	-	0.78 ( $\pm 0.016$ )	0.84
		SVM:RBF	$C = 2^{5.25}, \gamma = 2^{-15.0}$	<b>0.86 (<math>\pm 0.020</math>)</b>	0.95
		CNN	$lr = 0.001$	0.82 ( $\pm 0.020$ )	0.95

<sup>1</sup>Subscript *d* denotes SVM with default hyperparameters <sup>2</sup>95% confidence intervals are shown in brackets

<sup>3</sup>Bold font dictates best results

For the remainder of this section, we compare the SVM classification performance to the CNN. Note that when referring to the **CNN+SVM** hereforward, we refer explicitly to this *best* SVM case, the hyperparameter-tuned RBF.

As with the SVM, the CNN classifier classified general deep features well. Across both classification methods, training and validation accuracy were high, ranging from 0.95-0.98 ( $\mu = 0.97 \pm 0.01$ ) and 0.82-0.96 ( $\mu = 0.9 \pm 0.04$ ) respectively, see Table 7.2. Here,  $\mu$  indicates the mean and  $\pm$  the 95% CI. Final test accuracy, detailed in Table 7.3, was comparable to mean validation accuracy. Test accuracy even marginally exceeded validation performance (by 2-3%) in Datasets 1 & 2 when using a CNN and by 1% in Dataset 3 when using an SVM classifier. Across all datasets, classifier test accuracy ranged between 0.84-0.95 ( $\mu = 0.9 \pm 0.03$ ). We find that in general the simpler classification approach, **CNN+SVM**,

**Table 7.3:** Classifier training performance: with overall test accuracy and class-averaged metrics (Recall, Precision and F<sub>1</sub> Score)

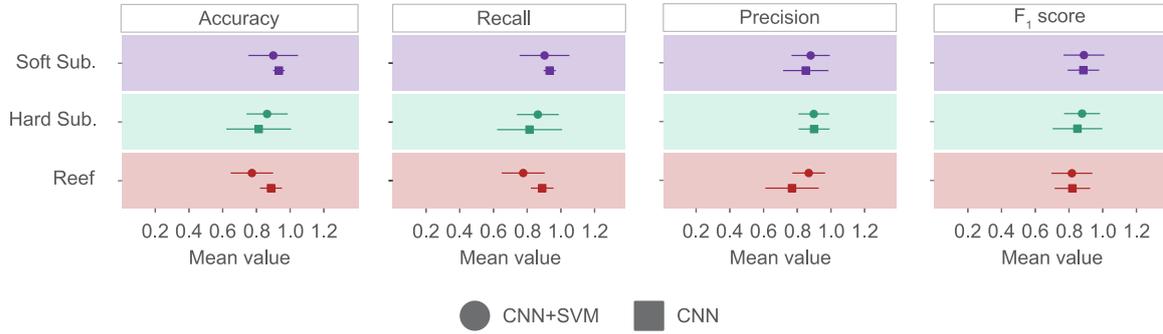
Set	Images	Classifier	Accuracy	Recall	Precision	F <sub>1</sub> Score
1	1671	SVM:Linear <sub>d</sub>	0.93	0.93 (± 0.025)	0.93 (± 0.022)	0.93 (± 0.023)
		SVM:Linear	0.93	0.93 (± 0.025)	0.93 (± 0.022)	0.93 (± 0.023)
		SVM:RBF <sub>d</sub>	<b>0.95</b>	0.92 (± 0.055)	<b>0.97 (± 0.026)</b>	0.94 (± 0.017)
		SVM:RBF	<b>0.95</b>	0.93 (± 0.040)	0.96 (± 0.016)	<b>0.95 (± 0.014)</b>
		CNN	0.95	<b>0.94 (± 0.018)</b>	0.93 (± 0.029)	0.93 (± 0.023)
2	248	SVM:Linear <sub>d</sub>	0.86	0.75 (± 0.140)	0.76 (± 0.123)	0.75 (± 0.130)
		SVM:Linear	0.86	0.75 (± 0.140)	0.76 (± 0.123)	0.75 (± 0.130)
		SVM:RBF <sub>d</sub>	0.88	0.63 (± 0.370)	<b>0.90 (± 0.098)</b>	0.67 (± 0.297)
		SVM:RBF	<b>0.89</b>	0.80 (± 0.106)	0.83 (± 0.070)	<b>0.82 (± 0.088)</b>
		CNN	0.88	<b>0.87 (± 0.039)</b>	0.78 (± 0.145)	<b>0.82 (± 0.085)</b>
3	115	SVM:Linear <sub>d</sub>	0.82	0.82 (± 0.041)	0.82 (± 0.127)	<b>0.82 (± 0.085)</b>
		SVM:Linear	0.82	0.82 (± 0.041)	0.82 (± 0.127)	<b>0.82 (± 0.085)</b>
		SVM:RBF <sub>d</sub>	0.83	0.67 (± 0.315)	<b>0.86 (± 0.126)</b>	0.70 (± 0.207)
		SVM:RBF	<b>0.87</b>	0.80 (± 0.138)	0.84 (± 0.049)	<b>0.82 (± 0.095)</b>
		CNN	0.84	<b>0.83 (± 0.168)</b>	0.80 (± 0.085)	0.80 (± 0.099)

<sup>1</sup>Subscript *d* denotes SVM with default hyperparameters <sup>2</sup>95% confidence intervals are shown in brackets

<sup>3</sup>Bold font dictates best results

competes well with its complex counterpart, **CNN**, increasing accuracy up to 4% across all test and validation sets.

Compared to test accuracy, mean F<sub>1</sub> score was more varied across datasets, ranging between 0.8-0.95 ( $\mu = 0.86 \pm 0.05$ ) across both methods. The lowest values were associated with CNN classification (in Dataset 3). SVM mean F<sub>1</sub> score was always >0.82; either matching or exceeded CNN performance up to 2%. Of the components that contributed to the mean F<sub>1</sub> score, namely recall and precision, precision was always greatest for the SVM classifiers, and higher than their scores for recall. By comparison recall was always highest for CNN classifiers and lower than their scores for precision. This indicates the different priorities of each classification approach. The SVM is more conservative in its predictions whereas the CNN favours over-estimation - ensuring that more of each class is captured. Across both methods, mean precision scored between 0.78-0.96 ( $\mu = 0.86 \pm 0.05$ ). Use of an SVM

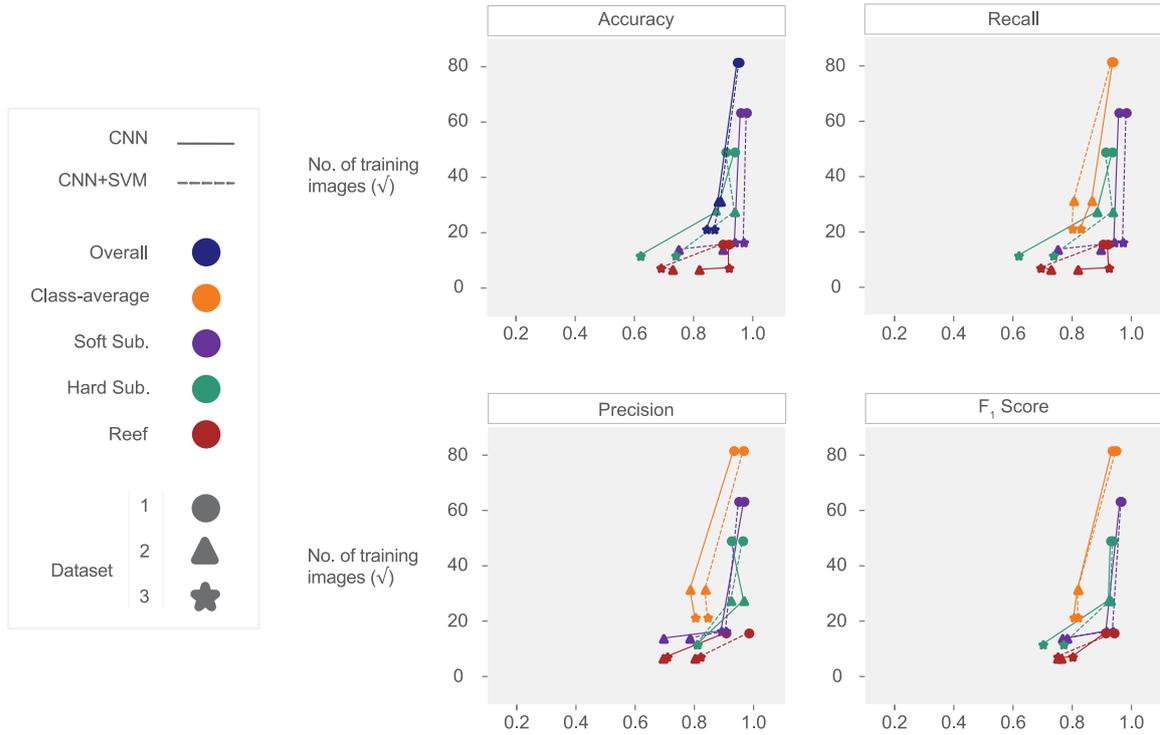


**Figure 7.5:** Mean performance across Datasets for each habitat. Error bars demonstrate 95% confidence intervals

increased mean precision by between 3-5%. Whereas mean recall ranged between 0.8-0.94 ( $\mu = 0.86 \pm 0.05$ ), with SVM decreasing performance by between 1-7%.

Considering the habitats individually, dataset-averaged performance metrics were relatively consistent across habitats and between classifiers (CNN or SVM) for each habitat, see Figure 7.5. Regarding the consistency of performance scores across metrics, for each habitat and classifier, more variation was present in the classification of reef. Average performance metrics for Reef varied between 0.77-0.89 ( $\mu = 0.84 \pm 0.05$ ) when using the CNN classifier and 0.77-0.87 ( $\mu = 0.81 \pm 0.04$ ) with an SVM classifier. Soft Sub. and Hard Sub. performance was more consistent across metrics by comparison. For Soft Sub., average performance metrics ranging from 0.85-0.93 ( $\mu = 0.9 \pm 0.03$ ) for CNN classification compared to 0.88-0.9 ( $\mu = 0.89 \pm 0.01$ ) when using an SVM. Hard Sub. scored between 0.81-0.89 ( $\mu = 0.84 \pm 0.03$ ) with a CNN and 0.86-0.9 ( $\mu = 0.87 \pm 0.01$ ) with an SVM. These figures indicate that regardless of class, dataset-averaged performance metrics are more similar to each other (and thus more stable) when using an SVM. However, the average of these performance metrics do not vary significantly between the two classifiers or between classes.

Comparing datasets, we see that the variability in performance metrics is somewhat reflective of the number of training images. In Figure 7.6 we show the overall and class-averaged metric scores on the test datasets (as detailed in Table 7.3) and the corresponding number of training images. Note that the size of training sets in Figure 7.6 have been square root trans-



**Figure 7.6:** Overall, class-averaged and individual class performance across test datasets - in order of decreasing training set size

formed for easier visual interpretation. Across datasets there is typically a general trend of decreasing performance with shrinking training set size, with Dataset 1 (the largest) scoring best followed by smaller datasets, 2 & 3. However, when focusing on individual classes we see that the number of training images is not always sufficient alone to explain variation in performance. Training sets were typically dominated by Soft Sub. followed by Hard Sub. & Reef, however for each of these, performance often declined significantly between the smaller datasets in which class representation was roughly equivalent. This was particularly pronounced for Soft Sub. and Reef and common to both classifiers.

Overall, performance of our habitat recognition pipelines was found comparable or better than accuracies reported following manual annotation of marine imagery (32; 118; 157; 42). This demonstrates that these approaches meet standards required in the marine science community and can be reliably integrated into image analysis pipelines for this application. We also note that our results compare and compete well with those previously reported in

benthic habitat classification, across various methods (161; 142; 132; 203; 204; 202), despite the comparatively small size of some of our datasets. Interestingly, our structure-based habitat classification also competes well with those that classify benthic imagery from a textural perspective, featuring highly-zoomed imagery (16; 122; 119; 68; 69).

### 7.3.3 Time Considerations

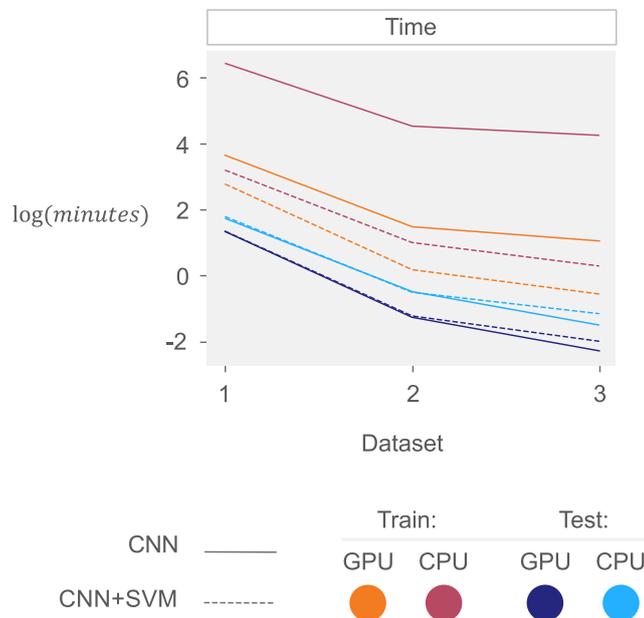
In Section 7.3.2 we presented classification results of two modelling approaches, **CNN+SVM** & **CNN**. In this experiment we further evaluate their performance with respect to training and testing time. We also consider the influence of hardware on computational efficiency.

#### Setup

The success of CNNs in image classification is partially linked to the development and availability of GPUs (108). Standard practice when using CNNs, either for feature extraction and/or classification, therefore involves GPU support. However, the computational efficiency provided varies drastically with the GPU model used and by extension the expense. For some, budgets do not accommodate purchase of such expensive hardware. Cloud services are available which provide access to GPU farms, some free of charge, yet their reliability can be hindered by fluctuating usage limits and hardware availability. By comparison, SVMs are highly memory efficient and do not require the use of a GPU. For this reason, we compare the time required for training and testing for both the **CNN+SVM** & **CNN** approaches, both with and without GPU support, exploring the magnitude of these differences and whether they are significant for an end-user. We use the same GPU & CPU from Section 7.3.2, namely an NVIDIA GeForce RTX 2080 SUPER GPU (8GB VRAM) and an Intel Core i7-9700 CPU. Due to the difference in modelling approaches and the number of hyperparameters required we did not consider time associated with hyperparameter-tuning for fair comparison.

## Results

The two modelling approaches yielded notable differences in training time. In Figure 7.7 we demonstrate final training, and testing time, for each approach. For visual clarity, time is expressed as  $\log(\text{minutes})$ . However, to facilitate better comprehension we discuss time in minutes throughout this section. Training was consistently faster across datasets when using an SVM. With a GPU, SVM training took between 0.58-16.21 ( $\mu = 6 \pm 8.17$ ) minutes,  $\sim 2.4\text{-}5\times$  faster than training a CNN, which took 2.89-38.78 ( $\mu = 15.37 \pm 18.74$ ) minutes. In Figure 7.7, we also demonstrate the decline to training speed when relying solely on a CPU. Training time took, on average, 3.6 ( $\pm 3.92$ ) minutes longer for the **CNN+SVM** when utilising a CPU, ranging between 1.35-24.69 ( $\mu = 9.6 \pm 12.09$ ) minutes. However, CNN training required, on average, 387.09 ( $\pm 472.32$ ) minutes more, ranging between 67.78-1015.3 ( $\mu = 402.46 \pm 491.07$ ) minutes; a significant deterioration in training time. Notably, training each **CNN+SVM** with only a CPU was still faster than training each **CNN** with a GPU.



**Figure 7.7:** Time required for training and testing phase, in order of decreasing dataset size. Duration is presented in  $\log(\text{minutes})$  for phases run both with a GPU and without (CPU)

Testing time was found to be equivalent between the two classification approaches, but vary with hardware. Using a GPU completed the testing phase, on average, 0.87 ( $\pm 0.7$ ) minutes faster across both methods. GPU-enabled testing time ranged between 0.1-3.88 ( $\mu = 1.43 \pm 1.38$ ) minutes compared to 0.28-5.87 ( $\mu = 2.26 \pm 2.9$ ) minutes using just the CPU.

## 7.4 Simplicity vs. Complexity

In this application, we note that SVM classifiers offer a good advantage over the more complex CNN classifier, for the following reasons: 1) performance, 2) time constraint and 3) ease of implementation.

We have demonstrated that SVM classifiers perform well and compete with CNN classifiers - typically matching or improving performance metrics across datasets. SVM classifiers were also more conservative in their predictions, a preferable trait in marine habitat classification. Prediction accuracy is more important than capturing all instances of a class. The use of an SVM classifier as an automated tool will thus be more appropriate for this application. They are also not particularly sensitive to alterations in training set size, handling poorly represented classes such as Reef reasonably well. SVMs always converge to an optimum, thus irrespective of training set size, if the underlying data are suitable they are likely to perform well. CNNs however, train iteratively, gradually improving their ability to fit complex relationships to the high-dimensional feature data to enable class prediction. Naturally, they may therefore require larger quantities of data to achieve higher performance. Interestingly, in this work, the sensitivity of the CNN classifier to training set size appeared similar to the SVM.

In terms of time constraints, we found that SVM's train much faster than CNNs. We also note that they are not as demanding with respect to computational resources. This was evident from our speed comparisons, contrasting time with and without a GPU. Considering that much of the heavy computation in the **CNN+SVM** approach is within the CNN

feature extraction phase, this accounts for the less pronounced increase in training time across datasets when using a CPU. Likely training time could be further improved if the GPU was employed for SVM training. We found no advantage with respect to testing time in utilising an SVM classifier. The classification process itself is not computationally intensive across the two methods, they will therefore require similar time periods in this phase. Since features must be extracted from test images before classification, the delay experienced in absence of a GPU is to be expected. Although these CPU/GPU comparisons are an extreme example, they emphasize well the importance of hardware considerations when selecting modelling approaches in image classification – particularly when using complex deep learning approaches. Note that time demands will likely vary with GPU type and other hardware specifications such as the number of cores and processor model.

We also recommend the use of SVMs due to their ease of implementation. Training an SVM is, programmatically, extremely straightforward, with training and testing possible within only a few simple lines of code. A simple hyperparameter grid-search can also aid optimal placement of the hyperplane, maximising performance. The relative complexity with the **CNN+SVM** approach, and by extension the **CNN** approach, rather lies in data management and feature extraction. The complexity of these steps is not so much related to the actions required, but navigating the literature and knowing "where to start". As we detail in Table 7.1, imagery must be organised within file structures according to class and partitioned into training and test sets. The CNN must also be altered such that it performs feature extraction only and images transformed (resized and normalized), with specialised functions, such that their features can be extracted. There is a wealth of resources available online but this can be overwhelming, most certainly for a novice user. As one of the contributions of this chapter we hope to better guide the user with clear and detailed descriptions of the methodological processes, with principles applicable across ML frameworks. Aside from these preparations, training the **CNN** classifier, is further complicated since no optimal can be found. Performance metrics must therefore be monitored across epochs and decisions made on when to stop training - typically the point at which training and test

accuracy are near equivalent, to minimise over- and under- fitting. Hyperparameters are also more extensive within a CNN and cannot be quickly optimised with a grid-search. Instead experience is required in their selection and interpretation. In this work, we found typical values used in the literature to be suitable - simplifying the approach in this aspect. However, with each application, tuning and exploration of hyperparameters is particularly essential with CNN classifiers to maximise their performance.

### 7.4.1 Improving performance

The approaches presented in this study were designed to minimize complexity and reflect a, realistic, performance baseline. Results that the reader could expect when attempting a similar classification task on their own data. That being said, there are some ways the results could be potentially improved. These improvements can be broadly applied to the following categories 1) the ground-truth (imagery) and 2) the extracted feature space. For further details, see Chapter 2.

It is widely acknowledged that a model is only as good as the data upon which it was trained and the work in this study is no exception. Model performance could thus be improved by enhancing the suitability of the underlying dataset; through quality and/or quantity. Quality-control to minimise error in manually-annotated datasets is important. An interesting observation from our own work was that trained models identified some manual annotation errors. Hence even an imperfect model, with <100% accuracy, could provide a useful screening tool to improve quality of the ground-truth. Model performance could also be improved through handling of class imbalance (105; 43). In the study survey areas, and more widely, much of the seafloor is represented by soft sediment whereas areas of reef framework are more scattered in their distribution. Image datasets therefore typically reflect this fact. Models can develop biases based on these class imbalances, particularly CNNs, such as over-predicting common classes. Down-weighting common classes or up-sampling rare classes could thus improve model performance (105), though this strays from realistic representation of the benthic environment.

Aside from balancing class-proportions, increasing the amount of training data could also increase performance (137; 203; 43). This is echoed in our own work, with increased performance associated with the largest training sets. (43) demonstrated that increased performance due to greater image quantity also extends beyond accuracy-based computer vision metrics and applies to those related to ecology, such as abundance and diversity. Since additional suitable data are not always available, a common method to increase availability is to re-sample existing training data. Known as data augmentation, this creates duplicates of training images that are typically modified through flips, crops and rotations. This aims to create a representation of classes in imagery that better reflects reality. However, great care must be taken to ensure these modifications do reflect reality; many of the common augmentations may not in fact be applicable to marine datasets (177). Note that data augmentation can also be used to solve class imbalance (105).

Increasing availability, and quality, of the ground-truth would also allow for sophistication of feature spaces extracted by the CNN. We used transfer learning to extract general features from our imagery, given the size of training data datasets and our target of minimising complexity across approaches. With more imagery, it would be possible to tune the CNN feature extractor or conduct scratch-based learning on our own datasets. This could in turn create feature representations more suitable for automatic classification of benthic habitats. Further modifications to the feature space that may improve performance link to their reduction, either through importance selection or dimensionality reduction. Large feature spaces suffer from the curse of dimensionality: whereby the space between data points grows with the number of features or dimensions. Finding groups and similar properties amongst data points in high-dimensions is thus difficult as points appear sparse. Data required to fill these 'gaps' increases almost exponentially with dimensionality. It may therefore be prudent to condense their dimensions, using PacMAP (191) for example, if more data are unavailable. Redundant features that have little relevance to decision functions can also introduce noise and so it may be beneficial to remove them. Note that a feature importance assessment cannot be conducted with a CNN (or SVM), instead some *post-*

*hoc* analysis with an alternative ML model can be conducted. Alternatively, employing a network which outputs less features, such as GoogleNet (176) and Inception V3 (176) could be helpful.

## 7.5 Conclusion

In this study, we focused on automating a relatively simple but time-consuming manual task that would benefit marine scientists: classifying benthic habitats in imagery.

We find that general features extracted from imagery by an 'off-the-shelf' convolutional neural network are suitable for classifying benthic imagery. Specifically, we find that they exhibit relatively clear associations with broad deep-sea habitats; contributing to one of a few known cases of this application. We show that SVM classifiers paired with a CNN feature extractor offer a simple, fast and consistent framework for automatically classifying benthic habitats. Using a CNN is, by comparison, less *user-friendly*. It is more conceptually illusive and requires more complex training procedures and significantly higher time and resource demands. Additionally, in its simplest *baseline* form, the performance does not warrant this added complexity. Given the results, we therefore believe it is useful to first employ the **CNN+SVM** approach over a CNN classifier, as it may prove sufficient.

The success of the **CNN+SVM** approach is not only illustrated in terms of high performance, but in the generality of the results across multiple datasets; something that is often missing from ML studies, unless comparing performance across benchmark datasets. As shown in this study, some variation is to be expected across datasets, in part due to their size. To therefore best demonstrate practical implementation, and encourage their use within processing pipelines, research needs to move past 'proof of concept studies' on singular datasets alone.

In order to make the procedures more accessible to non-specialists, our work concentrated on techniques that put simplicity before any potential improvement in habitat resolution. It is particularly suited to offshore use; offering near real-time decision making in the field

and the development of sampling protocols. Data collection can be triaged and quick, albeit crude, insights into habitat presence provided. It can also support automation by grouping images into similar categories (for annotation or model selection) or screen for labelling errors, as well as allow exploration of old datasets.

# 8 Conclusion

## 8.1 Contributions

In this thesis we have, broadly, focused on two problems within the analysis pipeline for benthic images. First that underwater image quality is often too poor and second that it is incredibly labour-intensive to analyse. Each of these problems was approached from a strong perspective of the end-user, aiming to develop solutions that improve the analytical experience for biologists with tools that are efficient and simpler to use. With this in mind, we make eight contributions.

First we evaluate control-point tonal manipulations; describing the general characteristics a TMO should possess to support image analysts in their benthic habitat assessments. This provides a foundation upon which to evaluate and develop tone-mapping tools in supporting benthic image interpretation.

In our second contribution, we explored the potential of the Rayleigh probability distribution as an automatic tool, given its popularity in underwater image enhancement. However, we found that it poorly summarized image brightness distributions, both before and after bespoke tone-mapping by analysts. This suggests it does not possess suitable qualities as previously believed. However, we note that image enhancement studies mostly judge suitability with regards to (subjective) aesthetics, whilst our aim here is to ensure the image is fit for analytical purposes. This does not discount aesthetics in the judgement of quality, but the assessment protocol is different, and potentially more complex.

Instead, for our third contribution, we find brightness statistics are better described by the Weibull probability distribution. This suggests that widely utilised tone mapping tools that

enforce uniform or rayleigh distributions in image brightness histograms, such as HE and CLHE, are unsuited to this application. This work provides a theoretical basis from which to develop an image enhancement tool to better support benthic image annotation.

In our fourth contribution we present a novel algorithm, Weibull Tone Mapping, to model the bespoke TMOs of analysts. In WTM, we fit a WD to brightness histograms of input and analyst-adjusted output images, then solve for the tone map that *matches* the input distribution to the enhanced output. The WTM tone map simplifies the TMOs into a two parameter function which results in smoother tonal enhancements that are less prone to artefacts.

For our fifth contribution we show that analysts find WTM helpful for image annotation but that the method of colour rendering, from a brightness to colour enhancement, influences their preference assessments. Analysts favour WTM images that preserve chromaticity. This maintains the intrinsic colour balance of the input image and thus enhancements appear more *natural*.

In our sixth contribution we provided analysts with an interactive WTM tool that allowed users to enhance imagery according to the two intuitive Weibull parameters, which approximately control brightness and contrast. This significantly shortens the time required to create a bespoke tonal enhancement and was preferred by analysts over tone map construction via interactive control-points. This tool improved upon existing TMO functionality in their annotation software and can easily be integrated into software to support live annotation. Analysts can now quickly create custom tonal manipulations on a per-image-basis quickly with an intuitive tool that is inspired by their own tonal enhancements.

In our seventh contribution, we further demonstrate the importance of personalized enhancements to improve underwater image analysis. We found that analysts prefer their tailored WTM enhancements to automatic TMOs (Adobe Photoshop) that would be even simpler and faster to employ - requiring just the click of a button. We also find that WTM

is more consistently suitable for analysis over CLHE, a popular TMO, for which preference strongly fluctuates.

Finally, for our last contribution, we target improvement of the image analysis bottleneck, providing one of a few known cases of automatic benthic (deep-sea) habitat classification. We provide analysts with an automation pipeline that leverages the power of deep learning (VGG16) and transfer learning but is made simpler and more accessible with the use of a support vector classifier. This is shown to be competitive with a convolutional neural network in terms of performance, time and ease of implementation; multiple datasets validate its practical utility. This work facilitates better comprehension and implementation of these approaches by a non-specialist with finite resources.

## 8.2 Future Work

### 8.2.1 Image Processing

Despite the multiple contributions in this thesis to aid benthic image surveys there are a number of possible areas on which to focus future work.

Although Weibull Tone Mapping simplifies and speeds-up the tone mapping process significantly, it only describes properties an automatic TMO should have. An interesting development to further support end-users in their analysis would be to create a *magic button* so to speak, that automatically provides a suitable enhancement based on the input image. We have yet to resolve this in our investigation of input-output image pairs and Weibull parameters. However, plans to collect more WTM adjustments from analysts on additional datasets will provide opportunity for further exploration. An automatic enhancement need not replace the dynamic use of WTM, merely act as a ‘first pass’ from which analysts can further adjust an image if required. The benefit of WTM is in part linked to its flexibility, creating diverse enhancements which can be quickly applied and re-adjusted

throughout the annotation process, highlighting variable features that aid the analyst to reach a classification.

In this thesis, we have focused only on global TMOs to initialize a strong foundation for WTM theory. However, it is crucial to consider adaptive (local) tonal enhancements in future work, particularly with underwater imaging. Adaptive transformation of pixel values, in which unique tone maps are applied to image sub-regions, better maintains smaller feature variations that could be important for analytical purposes. It also allows stronger brightness modifications to be made without over-enhancement and would help to resolve vignetting effects.

It is important to acknowledge that the Weibull Distribution exists in a 3-parameter form (called the translated, or general, Weibull distribution) (93). The extra parameter allows a given distribution shape to be shifted on the brightness axis. Thus, the ‘meaning’ of the third terms overlaps with the lambda parameter. This said, it may be a more difficult adjustment for a use to make. Currently, we have found that the 2-parameter form suffices in describing and providing suitable brightness enhancements. However, should more complexity and generality of brightness modifications be sought in the future, this provides an option.

Finally we remark that the usage of WTM need not be limited to underwater imagery. Future work could explore its potential on photographic or medical images.

### **8.2.2 Computer Vision**

A natural progression of this thesis and bridge between the two broad themes (Image Processing and Computer Vision) would be to see if image enhancement, with WTM, could improve automatic classification of benthic habitats. We have established that analysts prefer WTM as a supporting tool in benthic image analysis. An intriguing advancement would therefore be to determine if the same is true for a machine. Given that the WD is well suited to explain the brightness statistics of benthic imagery, it might also prove interesting to explore its capacity as a specialized underwater image augmentation tool to support

deep learning. The design and application of data augmentation tools in deep learning is an important area of research to improve the quality of automatic analyses. These tools must reflect reality and be domain-specific (177). The WD, and WTM, therefore presents an intriguing opportunity.

We highlighted the success of classification workflows on three datasets (Table 3.1) in this work. However, an important next step, in integrating this automated solution into image analysis workflows, is to validate the approach on datasets fully distinct to the training data, in a process known as *Domain Adaptation*. This would help to ascertain the transferability of deep generic features. It would also be prudent to determine if there a significant benefit in merging datasets and training together. This may better highlight what constitutes an appropriate dataset for deep learning (task-dependant), but also present a better representation of reality.

Our classification pipeline prioritised simplicity of the modelling approach over any potential gains in resolution of habitat classifications. However, further work is required to focus classification on habitat sub-classes listed in Table 3.2. This will require more sophisticated methodology such as improving the underlying feature space to include more tailored deep features. This could be achieved through training of more layers in a pre-trained network (Transfer Learning) such as VGG16. Scratch-based learning could create even more suitable feature spaces, yet this is unlikely to be applicable to this problem due the the lack of sufficiently large, annotated ‘gold-standard’ datasets. Despite its current resolution, it still forms an important component in developing a hierarchical, or ensemble, analysis tool for benthic imagery. It represent the ‘first-pass’, identifying the next suitable analysis steps and automatic tools to use, such as object detectors for specific taxa found within that habitat.

Future work could also look at better integration of acoustic, optical and environmental data to support contextual-based automation (142; 45; 169; 203). This would mimic analysis protocols by humans and could not only provide improvements to performance - particularly

in the case of visually indistinct taxa - but give biologists more confidence in automatic classifications.

### 8.3 Final thoughts

My experience in undertaking this thesis, has found that navigating the path between marine science and computer science is incredibly challenging but crucial. There is huge drive to optimize ocean observation with computer vision solutions, but this is coupled with a technical illiteracy in these topics that requires large resources in time and funding to overcome. The significant, yet exciting problems to be addressed, present a great opportunity for collaboration with computer scientists. However, in reality these are few in number, and though essential, cannot fully support the needs of the deep-sea community. It is therefore becoming increasingly clear that we need to build multi-disciplinary capacity between these fields. This is an intimidating, yet exhilarating, position to occupy. It is in developing these skill-sets that we can both optimize collaborations and manage priorities between the two disciplines, and bring new perspectives and ideas that could advance either field.

### 8.4 Publications

Game, C. A., Thompson, M. B., and Finlayson, G. D. (2020). Weibull tone mapping for underwater imagery. *Color and Imaging Conference*, 28, 156–161.

Game, C. A., Thompson, M. B., and Finlayson, G. D. (2021). Chromatic weibull tone mapping for underwater image enhancement. *Proceedings of the International Colour Association Congress 2021*, 239–244.

Game, C. A., Thompson, M. B., and Finlayson, G. D. (2023). Weibull Tone Mapping for the Enhancement of Underwater Imagery. *Sensors* 23 (7), 3533.

Game, C. A., Thompson, M. B., and Finlayson, G. D. (2023). Machine learning for non-experts: A more accessible and simpler approach to automatic benthic habitat classification. *In Press: Frontiers of Marine Science*.

# Abbreviations

**CLHE** Contrast Limited (*Adaptive*) Histogram Equalization.

**CNN** Convolutional Neural Network.

**FC** Fully Connected (layer).

**GUI** Graphical User Interface.

**HE** Histogram Equalization.

**ML** Machine Learning.

**RBF** Radial Basis Function.

**RD** Rayleigh Distribution.

**SVM** Support Vector Machine.

**TMO** Tone Mapping Operation.

**WD** Weibull Distribution.

**WTM** Weibull Tone Mapping.

## 9 Bibliography

- [1] Abad-Uribarren, A., Prado, E., Sierra, S., Cobo, A., Rodríguez-Basalo, A., Gómez-Ballesteros, M., and Sánchez, F. (2022). Deep learning-assisted high resolution mapping of vulnerable habitats within the Capbreton Canyon System, Bay of Biscay. *Estuarine, Coastal and Shelf Science*, 275:107957.
- [2] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., and Zheng, X. (2016). TensorFlow: A System for Large-Scale Machine Learning. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, pages 265–283.
- [3] Abosaq, H. A., Ramzan, M., Althobiani, F., Abid, A., Aamir, K. M., Abdushkour, H., Irfan, M., Gommosani, M. E., Ghonaim, S. M., Shamji, V. R., and Rahman, S. (2023). Unusual Driver Behavior Detection in Videos Using Deep Learning Models. *Sensors*, 23(1):311.
- [4] Adini, Y., Moses, Y., and Ullman, S. (1997). Face recognition: the problem of compensating for changes in illumination direction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):721–732.
- [5] Adobe (2023). Adobe Photoshop 2023. Available at: <https://www.adobe.com/uk/products/photoshop/> [Accessed 25 May 2023].

- [6] Ahonen, T., Hadid, A., and Pietikainen, M. (2006). Face Description with Local Binary Patterns: Application to Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041.
- [7] Akkaynak, D. and Treibitz, T. (2018). A Revised Underwater Image Formation Model. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6723–6732.
- [8] Althubiti, S. A., Alenezi, F., Shitharth, S., K., S., and Reddy, C. V. S. (2022). Circuit Manufacturing Defect Detection Using VGG16 Convolutional Neural Networks. *Wireless Communications and Mobile Computing*, 2022:e1070405.
- [9] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., and Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8(1):53.
- [10] Armstrong, C. W., Foley, N. S., Tinch, R., and van den Hove, S. (2012). Services from the deep: Steps towards valuation of deep sea goods and services. *Ecosystem Services*, 2:2–13.
- [11] Azizpour, H., Razavian, A. S., Sullivan, J., Maki, A., and Carlsson, S. (2015). From generic to specific deep representations for visual recognition. *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 36–45.
- [12] Bai, L., Zhang, W., Pan, X., and Zhao, C. (2020). Underwater Image Enhancement Based on Global and Local Equalization of Histogram and Dual-Image Multi-Scale Fusion. *IEEE Access*, 8:128973–128990.

- [13] Baker, E. K. and Harris, P. T. (2020). Chapter 2 - Habitat mapping and marine management. In Harris, P. T. and Baker, E., editors, *Seafloor Geomorphology as Benthic Habitat (Second Edition)*, pages 17–33. Elsevier.
- [14] Bay, H., Tuytelaars, T., and Van Gool, L. (2006). SURF: Speeded Up Robust Features. In Leonardis, A., Bischof, H., and Pinz, A., editors, *Computer Vision – ECCV 2006*, Lecture Notes in Computer Science, pages 404–417, Berlin, Heidelberg. Springer.
- [15] Bazeille, S., Quidu, I., Jaulin, L., and Malkasse, J.-P. (2006). Automatic underwater image pre-processing. In *Characterisation Du Milieu Marin (CMM 06)*.
- [16] Beijbom, O., Edmunds, P. J., Kline, D. I., Mitchell, B. G., and Kriegman, D. (2012). Automated annotation of coral reef survey images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1170–1177.
- [17] Beijbom, O., Edmunds, P. J., Roelfsema, C., Smith, J., Kline, D. I., Neal, B. P., Dunlap, M. J., Moriarty, V., Fan, T.-Y., Tan, C.-J., Chan, S., Treibitz, T., Gamst, A., Mitchell, B. G., and Kriegman, D. (2015). Towards Automated Annotation of Benthic Survey Images: Variability of Human Experts and Operational Modes of Automation. *PLOS ONE*, 10(7):e0130312.
- [18] Beijbom, O., Treibitz, T., Kline, D. I., Eyal, G., Khen, A., Neal, B., Loya, Y., Mitchell, B. G., and Kriegman, D. (2016). Improving Automated Annotation of Benthic Survey Images Using Wide-band Fluorescence. *Scientific Reports*, 6(1):23166.
- [19] Bewley, M., Friedman, A., Ferrari, R., Hill, N., Hovey, R., Barrett, N., Marzinelli, E. M., Pizarro, O., Figueira, W., Meyer, L., Babcock, R., Bellchambers, L., Byrne, M.,

- and Williams, S. B. (2015). Australian sea-floor survey data, with images and expert annotations. *Scientific Data*, 2(1):150057.
- [20] Bicknell, A. W. J., Godley, B. J., Sheehan, E. V., Votier, S. C., and Witt, M. J. (2016). Camera technology for monitoring marine biodiversity and human impact. *Frontiers in Ecology and the Environment*, 14:424–432.
- [21] Billett, D. S. M., Bett, B. J., Reid, W. D. K., Boorman, B., and Priede, I. G. (2010). Long-term change in the abyssal NE Atlantic: The ‘Amperima Event’ revisited. *Deep Sea Research Part II: Topical Studies in Oceanography*, 57(15):1406–1417.
- [22] Brown, C. J., Smith, S. J., Lawton, P., and Anderson, J. T. (2011). Benthic habitat mapping: A review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. *Estuarine, Coastal and Shelf Science*, 92(3):502–520.
- [23] Buhl-Mortensen, L., Vanreusel, A., Gooday, A. J., Levin, L. A., Priede, I. G., Buhl-Mortensen, P., Gheerardyn, H., King, N. J., and Raes, M. (2010). Biological structures as a source of habitat heterogeneity and biodiversity on the deep ocean margins. *Marine Ecology*, 31(1):21–50.
- [24] Chang, C.-C. and Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3):Article 27.
- [25] Chen, Q., Beijbom, O., Chan, S., Bouwmeester, J., and Kriegman, D. (2021). A New Deep Learning Engine for CoralNet. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3693–3702.

- [26] Clark, M. R., Althaus, F., Schlacher, T. A., Williams, A., Bowden, D. A., and Rowden, A. A. (2015). The impacts of deep-sea fisheries on benthic communities: a review. *ICES Journal of Marine Science*, 73(suppl\textunderscore1):i51–i69.
- [27] Cogan, C. B., Todd, B. J., Lawton, P., and Noji, T. T. (2009). The role of marine habitat mapping in ecosystem-based management. *ICES Journal of Marine Science*, 66(9):2033–2042.
- [28] Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3):273–297.
- [29] Costanza, R. (1999). The ecological, economic, and social importance of the oceans. *Ecological Economics*, 31(2):199–213.
- [30] Cristianini, N. and Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press.
- [31] Culverhouse, P. F., Macleod, N., Williams, R., Benfield, M. C., Lopes, R. M., and Picheral, M. (2014). An empirical assessment of the consistency of taxonomic identifications. *Marine Biology Research*, 10(1):73–84.
- [32] Culverhouse, P. F., Williams, R., Reguera, B., Herry, V., and González-Gil, S. (2003). Do experts make mistakes? A comparison of human and machine identification of dinoflagellates. *Marine Ecology Progress Series*, 247:17–25.
- [33] 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 vol. 1.

- [34] Davies, C. E., Moss, D., and Hill, M. O. (2004). EUNIS habitat classification revised 2004. *Report to: European Environment Agency European Topic Centre on Nature Protection and Biodiversity*.
- [35] Dawkins, M., Sherrill, L., Fieldhouse, K., Hoogs, A., Richards, B., Zhang, D., Prasad, L., Williams, K., Lauffenburger, N., and Wang, G. (2017). An Open-Source Platform for Underwater Image and Video Analytics. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 898–906.
- [36] Deng, J., Dong, W., Socher, R., Li, L. J., Kai, L., and Li, F.-F. (2009). ImageNet: A large-scale hierarchical image database. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255.
- [37] Deperlioglu, O. and Kose, U. (2018). Practical Method for the Underwater Image Enhancement with Adjusted CLAHE. In *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, pages 1–6.
- [38] Dias, F. C., Gomes-Pereira, J., Tojeira, I., Souto, M., Afonso, A., Calado, A., Madureira, P., and Campos, A. (2015). Area Estimation of Deep-Sea Surfaces from Oblique Still Images. *PLOS ONE*, 10(7):e0133290.
- [39] Diaz, R. J., Solan, M., and Valente, R. M. (2004). A review of approaches for classifying benthic habitats and evaluating habitat quality. *Journal of Environmental Management*, 73(3):165–181.
- [40] Downie, A.-L., Noble-James, T., Sperry, J., and White, S. (2022). Automated detection of sea pens in video footage: Applications for time-series monitoring of Vulnerable Marine Ecosystems (VME). Technical report.

- [41] Dunn-Rankin, P., Knezek, G. A., Wallace, S., and Zhang, S. (2014). *Scaling Methods*. Psychology Press, New York, 2 edition.
- [42] Durden, J. M., Bett, B. J., Schoening, T., Morris, K. J., Nattkemper, T. W., and Ruhl, H. A. (2016a). Comparison of image annotation data generated by multiple investigators for benthic ecology. *Marine Ecology Progress Series*, 552:61–70.
- [43] Durden, J. M., Hosking, B., Bett, B. J., Cline, D., and Ruhl, H. A. (2021). Automated classification of fauna in seabed photographs: The impact of training and validation dataset size, with considerations for the class imbalance. *Progress in Oceanography*, 196:102612.
- [44] Durden, J. M., Schoening, T., Althaus, F., Friedman, A., Garcia, R., Glover, A. G., Greinert, J., Stout, N. J., Jones, D. O. B., Jordt, A., Kaeli, J. W., Köser, K., Kuhnz, L. A., Lindsay, D., Morris, K. J., Nattkemper, T. W., Osterloff, J., Ruhl, H. A., Singh, H., and Bett, M. T. . B. J. (2016b). Perspectives in Visual Imaging for Marine Biology and Ecology: From Acquisition to Understanding. In *Oceanography and Marine Biology: An Annual Review*, volume 54, pages 1–72. CRC Press, 1st edition.
- [45] Ellen, J. S., Graff, C. A., and Ohman, M. D. (2019). Improving plankton image classification using context metadata. *Limnology and Oceanography: Methods*, 17(8):439–461.
- [46] European Environment Agency (2022). EUNIS habitat classification. Available at: <https://www.eea.europa.eu/data-and-maps/data/eunis-habitat-classification-1> [Accessed 19 Dec. 2022].

- [47] European Parliament (2008). Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive).
- [48] Eustice, R., Pizarro, O., Singh, H., and Howland, J. (2002). UWIT: underwater image toolbox for optical image processing and mosaicking in MATLAB. In *Proceedings of the 2002 International Symposium on Underwater Technology (Cat. No.02EX556)*, pages 141–145.
- [49] Evans, D., Aish, A., Boon, A., Condé, S., Connor, D., Gelabert, E. M. N., Parry, M., Richard, D., Salvati, E., and Tunesi, L. (2016). Revising the marine section of the EUNIS Habitat classification-Report of a workshop held at the European Topic Centre on Biological Diversity, 12 & 13 May 2016. ETC/BD report to the EEA. Technical report.
- [50] Ewing, M. (1967). Early development of ocean-bottom photography at Woods Hole Oceanographic Institution and Lamont Geological Observatory. In *Deep-sea photography*, volume 3 of *The John Hopkins Oceanographic Studies*, pages 13–41. The John Hopkins Press, Baltimore.
- [51] Ewing, M., Vine, A., and Worzel, J. L. (1946). Photography of the Ocean Bottom. *J. Opt. Soc. Am.*, 36(6):307–321.
- [52] French, G., Mackiewicz, M., Fisher, M., Holah, H., Kilburn, R., Campbell, N., and Needle, C. (2019). Deep neural networks for analysis of fisheries surveillance video and automated monitoring of fish discards. *ICES Journal of Marine Science*, 77(4):1340–1353.
- [53] Friedman, N., Geiger, D., and Goldszmidt, M. (1997). Bayesian Network Classifiers. *Machine Learning*, 29(2):131–163.

- [54] Gage, J. D. and Tyler, P. A. (1991). *Deep-Sea Biology: A Natural History of Organisms at the Deep-Sea Floor*. Cambridge University Press, Cambridge.
- [55] Garrison, T. S. (2012). *Oceanography: an invitation to marine science*. Cengage Learning.
- [56] Gescheider, G. A. (1976). *Psychophysics: Method and theory*. Lawrence Erlbaum, Hillsdale, N.J.
- [57] Geusebroek, J. M. and Smeulders, A. W. M. (2002). A Physical Explanation for Natural Image Statistics. In *2nd International Workshop on Texture Analysis and Synthesis*, pages 47–52. Copenhagen, Denmark: Heriot-Watt University.
- [58] Ghani, A. S. A. (2018). Image contrast enhancement using an integration of recursive-overlapped contrast limited adaptive histogram specification and dual-image wavelet fusion for the high visibility of deep underwater image. *Ocean Engineering*, 162:224–238.
- [59] Ghani, A. S. A. and Isa, N. A. M. (2015a). Enhancement of low quality underwater image through integrated global and local contrast correction. *Applied Soft Computing*, 37:332–344.
- [60] Ghani, A. S. A. and Isa, N. A. M. (2015b). Underwater image quality enhancement through integrated color model with Rayleigh distribution. *Applied Soft Computing*, 27:219–230.
- [61] Glover, A. G. and Smith, C. R. (2003). The deep-sea floor ecosystem: current status and prospects of anthropogenic change by the year 2025. *Environmental Conservation*, 30(3):219–241.

- [62] Goetze, J. S., Bond, T., McLean, D. L., Saunders, B. J., Langlois, T. J., Lindfield, S., Fullwood, L. A. F., Driessen, D., Shedrawi, G., and Harvey, E. S. (2019). A field and video analysis guide for diver operated stereo-video. *Methods in Ecology and Evolution*, 10(7):1083–1090.
- [63] Gomes-Pereira, J. N., Auger, V., Beisiegel, K., Benjamin, R., Bergmann, M., Bowden, D., Buhl-Mortensen, P., De Leo, F. C., Dionísio, G., Durden, J. M., Edwards, L., Friedman, A., Greinert, J., Jacobsen-Stout, N., Lerner, S., Leslie, M., Nattkemper, T. W., Sameoto, J. A., Schoening, T., Schouten, R., Seager, J., Singh, H., Soubigou, O., Tojeira, I., van den Beld, I., Dias, F., Tempera, F., and Santos, R. S. (2016). Current and future trends in marine image annotation software. *Progress in Oceanography*, 149:106–120.
- [64] Gonzales, R. C. and Woods, R. E. (2002). *Digital image processing*. Prentice hall, New Jersey, 3 edition.
- [65] González-Rivero, M., Beijbom, O., Rodriguez-Ramirez, A., Bryant, D. E. P., Ganase, A., Gonzalez-Marrero, Y., Herrera-Reveles, A., Kennedy, E. V., Kim, C. J. S., Lopez-Marcano, S., Markey, K., Neal, B. P., Osborne, K., Reyes-Nivia, C., Sampayo, E. M., Stolberg, K., Taylor, A., Vercelloni, J., Wyatt, M., and Hoegh-Guldberg, O. (2020). Monitoring of Coral Reefs Using Artificial Intelligence: A Feasible and Cost-Effective Approach. *Remote Sensing*, 12(3):489.
- [66] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.
- [67] Guo, G., Wang, H., Bell, D., Bi, Y., and Greer, K. (2003). KNN Model-Based Approach in Classification. In Meersman, R., Tari, Z., and Schmidt, D. C., editors, *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, Lecture Notes in Computer Science, pages 986–996, Berlin, Heidelberg. Springer.

- [68] Gómez-Ríos, A., Tabik, S., Luengo, J., Shihavuddin, A. S. M., and Herrera, F. (2019a). Coral species identification with texture or structure images using a two-level classifier based on Convolutional Neural Networks. *Knowledge-Based Systems*, 184:104891.
- [69] Gómez-Ríos, A., Tabik, S., Luengo, J., Shihavuddin, A. S. M., Krawczyk, B., and Herrera, F. (2019b). Towards highly accurate coral texture images classification using deep convolutional neural networks and data augmentation. *Expert Systems with Applications*, 118:315–328.
- [70] Halpern, B. S., Walbridge, S., Selkoe, K. A., Kappel, C. V., Micheli, F., D’Agrosa, C., Bruno, J. F., Casey, K. S., Ebert, C., Fox, H. E., Fujita, R., Heinemann, D., Lenihan, H. S., Madin, E. M. P., Perry, M. T., Selig, E. R., Spalding, M., Steneck, R., and Watson, R. (2008). A Global Map of Human Impact on Marine Ecosystems. *Science*, 319(5865):948–952.
- [71] Haralick, R. M., Shanmugam, K., and Dinstein, I. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6):610–621.
- [72] Harris, M., Finlayson, G., Tauber, J., and Farnand, S. (2013). Web-based Image Preference. *Journal of Imaging Science and Technology*, 57(2):2050–1 – 20502–12.
- [73] Harris, P. T. and Baker, E. K. (2020). Chapter 1 - Why map benthic habitats? In Harris, P. T. and Baker, E., editors, *Seafloor Geomorphology as Benthic Habitat (Second Edition)*, pages 3–15. Elsevier.

- [74] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.
- [75] Hitam, M. S., Awalludin, E. A., Jawahir Hj Wan Yussof, W. N., and Bachok, Z. (2013). Mixture contrast limited adaptive histogram equalization for underwater image enhancement. In *2013 International Conference on Computer Applications Technology (ICCAT)*, pages 1–5.
- [76] Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, volume 1, pages 278–282 vol.1.
- [77] Horgan, J. and Toal, D. (2009). Computer Vision Applications in the Navigation of Unmanned Underwater Vehicles. In Inzartsev, A. V., editor, *Underwater Vehicles*, pages 195–214. IntechOpen, London.
- [78] Hou, G., Li, Y., Yang, H., Li, K., and Pan, Z. (2023). UID2021: An Underwater Image Dataset for Evaluation of No-Reference Quality Assessment Metrics. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 19(4):151:1–151:24.
- [79] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.
- [80] Howell, K. L., Davies, J. S., and Narayanaswamy, B. E. (2010). Identifying deep-sea megafaunal epibenthic assemblages for use in habitat mapping and marine protected area network design. *Journal of the Marine Biological Association of the United Kingdom*, 90(1):33–68.

- [81] Howell, K. L., Hilário, A., Allcock, A. L., Bailey, D., Baker, M., Clark, M. R., Colaço, A., Copley, J., Cordes, E. E., Danovaro, R., Dissanayake, A., Escobar, E., Esquete, P., Gallagher, A. J., Gates, A. R., Gaudron, S. M., German, C. R., Gjerde, K. M., Higgs, N. D., Le Bris, N., Levin, L. A., Manea, E., McClain, C., Menot, L., Mestre, N. C., Metaxas, A., Milligan, R., Muthumbi, A. W. N., Narayanaswamy, B. E., Ramalho, S. P., Ramirez-Llodra, E., Robson, L. M., Rogers, A. D., Sellanes, J., Sigwart, J. D., Sink, K., Snelgrove, P. V. R., Stefanoudis, P. V., Sumida, P. Y., Taylor, M. L., Thurber, A. R., Vieira, R., Watanabe, H. K., Woodall, L. C., and Xavier, J. R. (2021). A decade to study deep-sea life. *Nature Ecology & Evolution*, 5(3):265–267.
- [82] Howell, K. L., Hilário, A., Allcock, A. L., Bailey, D. M., Baker, M., Clark, M. R., Colaço, A., Copley, J., Cordes, E. E., and Danovaro, R. (2020). A blueprint for an inclusive, global deep-sea ocean decade field program. *Frontiers in Marine Science*, page 999.
- [83] Howell, K. L., Holt, R., Endrino, I. P., and Stewart, H. (2011). When the species is also a habitat: Comparing the predictively modelled distributions of *Lophelia pertusa* and the reef habitat it forms. *Biological Conservation*, 144(11):2656–2665.
- [84] Hsu, C.-W., Chang, C.-C., and Lin, C.-J. (2016). A Practical Guide to Support Vector Classification. Available at: <https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf> [Accessed 22 Dec. 2022].
- [85] Hu, K., Weng, C., Zhang, Y., Jin, J., and Xia, Q. (2022). An Overview of Underwater Vision Enhancement: From Traditional Methods to Recent Deep Learning. *Journal of Marine Science and Engineering*, 10(2):241.

- [86] Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269.
- [87] Huh, M., Agrawal, P., and Efros, A. A. (2016). What makes ImageNet good for transfer learning? Available at: <https://arxiv.org/abs/1608.08614> [Accessed 06 Sep. 2023].
- [88] Hummel, R. (1977). Image enhancement by histogram transformation. *Computer Graphics and Image Processing*, 6(2):184–195.
- [89] Huvenne, V. A., Robert, K., Marsh, L., Lo Iacono, C., Le Bas, T., and Wynn, R. B. (2018). ROVs and AUVs. In Micallef, A., Krastel, S., and Savini, A., editors, *Submarine Geomorphology*, Springer Geology, pages 93–108. Springer International Publishing, Cham.
- [90] International Telecommunication Union (2019). Recommendation 500-14: Methodology for the subjective assessment of the quality of television pictures. Available at: <https://www.itu.int/rec/R-REC-BT.500-14-201910-I/en> [Accessed 22 Dec. 2022].
- [91] Iqbal, K., Salam, R. A., Osman, A., and Talib, A. Z. (2007). Underwater Image Enhancement Using an Integrated Colour Model. *IAENG International Journal of Computer Science*, 34(2):239–244.
- [92] ISO (2009). ISO 3664:2009. Technical report, Geneva, Switzerland: ISO.
- [93] Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). *Continuous univariate distributions, volume 2*, volume 289. John Wiley & Sons, New York.

- [94] Joint Nature Conservation Committee (JNCC) (2022). Data Collection: Survey Methods and Equipment. Available at: <https://jncc.gov.uk/our-work/data-collection-survey-methods-and-equipment/#seabed-imagery-drop-frame-mounted-camera> [Accessed 13 Dec. 2022].
- [95] Kandimalla, V., Richard, M., Smith, F., Quirion, J., Torgo, L., and Whidden, C. (2022). Automated Detection, Classification and Counting of Fish in Fish Passages With Deep Learning. *Frontiers in Marine Science*, 8.
- [96] Kaur, T. and Gandhi, T. K. (2019). Automated Brain Image Classification Based on VGG-16 and Transfer Learning. In *2019 International Conference on Information Technology (ICIT)*, pages 94–98.
- [97] Kingma, D. P. and Ba, J. (2015). Adam: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations*, San Diego, USA.
- [98] Kloster, M., Langenkämper, D., Zurowietz, M., Beszteri, B., and Nattkemper, T. W. (2020). Deep learning-based diatom taxonomy on virtual slides. *Scientific Reports*, 10(1):14416.
- [99] Kocak, D. M., Dalglish, F. R., Caimi, F. M., and Schechner, Y. Y. (2008). A Focus on Recent Developments and Trends in Underwater Imaging. *Marine Technology Society Journal*, 42(1):52–67.
- [100] Krig, S. (2014). *Computer vision metrics: Survey, taxonomy, and analysis*. Apress.
- [101] Krishnaswamy Rangarajan, A. and Purushothaman, R. (2020). Disease Classification in Eggplant Using Pre-trained VGG16 and MSVM. *Scientific Reports*, 10(1):2322.

- [102] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc.
- [103] Lacharité, M. and Metaxas, A. (2017). Hard substrate in the deep ocean: How sediment features influence epibenthic megafauna on the eastern Canadian margin. *Deep Sea Research Part I: Oceanographic Research Papers*, 126:50–61.
- [104] Langenkämper, D., Simon-Lledó, E., Hosking, B., Jones, D. O. B., and Nattkemper, T. W. (2019a). On the impact of Citizen Science-derived data quality on deep learning based classification in marine images. *Plos One*, 14(6):e0218086.
- [105] Langenkämper, D., van Kevelaer, R., and Nattkemper, T. W. (2019b). Strategies for Tackling the Class Imbalance Problem in Marine Image Classification. In Zhang, Z., Suter, D., Tian, Y., Branzan Albu, A., Sidère, N., and Jair Escalante, H., editors, *Pattern Recognition and Information Forensics*, Lecture Notes in Computer Science, pages 26–36, Cham. Springer International Publishing.
- [106] Langenkämper, D., van Kevelaer, R., Purser, A., and Nattkemper, T. W. (2020). Gear-Induced Concept Drift in Marine Images and Its Effect on Deep Learning Classification. *Frontiers in Marine Science*, 7(506).
- [107] Langenkämper, D., Zurowietz, M., Schoening, T., and Nattkemper, T. W. (2017). BIIGLE 2.0 - Browsing and Annotating Large Marine Image Collections. *Frontiers in Marine Science*, 4:83.
- [108] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

- [109] Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.
- [110] Li, C. Y. and Cavallaro, A. (2022). On The Limits of Perceptual Quality Measures for Enhanced Underwater Images. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 4148–4152.
- [111] Li, C.-Y., Guo, J.-C., Cong, R.-M., Pang, Y.-W., and Wang, B. (2016). Underwater Image Enhancement by Dehazing With Minimum Information Loss and Histogram Distribution Prior. *IEEE Transactions on Image Processing*, 25(12):5664–5677.
- [112] Li, C. Y., Mazzon, R., and Cavallaro, A. (2020). Underwater image filtering: methods, datasets and evaluation. Available at: <http://arxiv.org/abs/2012.12258> [Accessed 22 Dec. 2022].
- [113] Li, Q., He, X., Wang, Y., Liu, H., Xu, D., and Guo, F. (2013). Review of spectral imaging technology in biomedical engineering: achievements and challenges. *Journal of Biomedical Optics*, 18(10):100901.
- [114] Lindbloom, B. J. (1989). Accurate color reproduction for computer graphics applications. *ACM SIGGRAPH Computer Graphics*, 23(3):117–126.
- [115] Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110.
- [116] Lucieer, V. L. and Forrest, A. L. (2016). Emerging Mapping Techniques for Autonomous Underwater Vehicles (AUVs). In Finkl, C. W. and Makowski, C., editors,

- Seafloor Mapping along Continental Shelves: Research and Techniques for Visualizing Benthic Environments*, pages 53–67. Springer International Publishing, Cham.
- [117] Lumini, A. and Nanni, L. (2019). Deep learning and transfer learning features for plankton classification. *Ecological Informatics*, 51:33–43.
- [118] MacLeod, N., Benfield, M., and Culverhouse, P. (2010). Time to automate identification. *Nature*, 467(7312):154–155.
- [119] Mahmood, A., Bennamoun, M., An, S., Sohel, F., and Boussaid, F. (2020a). ResFeats: Residual network based features for underwater image classification. *Image and Vision Computing*, 93:103811.
- [120] Mahmood, A., Bennamoun, M., An, S., Sohel, F., Boussaid, F., Hovey, R., Kendrick, G., and Fisher, R. B. (2016). Coral classification with hybrid feature representations. *IEEE International Conference on Image Processing (ICIP)*, pages 519–523.
- [121] Mahmood, A., Bennamoun, M., An, S., Sohel, F. A., Boussaid, F., Hovey, R., Kendrick, G. A., and Fisher, R. B. (2019). Deep Image Representations for Coral Image Classification. *IEEE Journal of Oceanic Engineering*, 44(1):121–131.
- [122] Mahmood, A., Ospina, A. G., Bennamoun, M., An, S., Sohel, F., Boussaid, F., Hovey, R., Fisher, R. B., and Kendrick, G. A. (2020b). Automatic Hierarchical Classification of Kelps Using Deep Residual Features. *Sensors*, 20(2):447.
- [123] Mallet, D. and Pelletier, D. (2014). Underwater video techniques for observing coastal marine biodiversity: A review of sixty years of publications (1952–2012). *Fisheries Research*, 154:44–62.

- [124] Marburg, A. and Bigham, K. (2016). Deep learning for benthic fauna identification. In *OCEANS 2016 MTS/IEEE Monterey*, pages 1–5.
- [125] Matas, J., Chum, O., Urban, M., and Pajdla, T. (2004). Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767.
- [126] MathWorks (2023). adapthisteq : Contrast-limited adaptive histogram equalization (CLAHE). Available at: <https://uk.mathworks.com/help/images/ref/adapthisteq.html> [Accessed 4 Jul. 2023].
- [127] Maymon, G. (2018). Chapter 2 - Some Important Statistical Distributions. In Maymon, G., editor, *Stochastic Crack Propagation*, pages 9–18. Academic Press.
- [128] McGlamery, B. (1980). *A Computer Model For Underwater Camera Systems*, volume 0208 of *Ocean Optics VI*. SPIE.
- [129] McVey, J., Finlayson, G., and Finlayson, G. (2019). Least-Squares Optimal Contrast Limited Histogram Equalisation. *Color and Imaging Conference*, 27:256–261.
- [130] Mienis, F., de Stigter, H. C., White, M., Duineveld, G., de Haas, H., and van Weering, T. C. E. (2007). Hydrodynamic controls on cold water coral growth and carbonate mound development at the SW and SE Rockall Trough Margin, NE Atlantic Ocean. *Deep-Sea Research Part 1. Oceanographic Research Papers*, 54:1655 – 1674.
- [131] Mikolajczyk, K. and Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630.

- [132] Mohamed, H., Nadaoka, K., and Nakamura, T. (2020). Semiautomated Mapping of Benthic Habitats and Seagrass Species Using a Convolutional Neural Network Framework in Shallow Water Environments. *Remote Sensing*, 12(23):4002.
- [133] National Oceanography Center (2022). Autosubs. Available at: <https://noc.ac.uk/facilities/marine-autonomous-robotic-systems/autosubs> [Accessed 13 Dec. 2022].
- [134] Panetta, K., Gao, C., and Agaian, S. (2016). Human-Visual-System-Inspired Underwater Image Quality Measures. *IEEE Journal of Oceanic Engineering*, 41(3):541–551.
- [135] Perez-Ortiz, M. and Mantiuk, R. K. (2017). A practical guide and software for analysing pairwise comparison experiments.
- [136] Piechaud, N. and Howell, K. L. (2022). Fast and accurate mapping of fine scale abundance of a VME in the deep sea with computer vision. *Ecological Informatics*, page 101786.
- [137] Piechaud, N., Hunt, C., Culverhouse, P. F., Foster, N. L., and Howell, K. L. (2019). Automated identification of benthic epifauna with computer vision. *Marine Ecology Progress Series*, 615:15–30.
- [138] Pizer, S. M., Amburn, E. P., Austin, J. D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J. B., and Zuiderveld, K. (1987). Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, 39(3):355–368.

- [139] PyTorch (2023). PyTorch. Available at: <https://pytorch.org/> [Accessed 02 Mar. 2023].
- [140] Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1):81–106.
- [141] Ramirez-Llodra, E., Tyler, P. A., Baker, M. C., Bergstad, O. A., Clark, M. R., Escobar, E., Levin, L. A., Menot, L., Rowden, A. A., Smith, C. R., and Van Dover, C. L. (2011). Man and the Last Great Wilderness: Human Impact on the Deep Sea. *PloS one*, 6(8):e22588.
- [142] Rao, D., De Deuge, M., Nourani–Vatani, N., Williams, S. B., and Pizarro, O. (2017). Multimodal learning and inference from visual and remotely sensed data. *The International Journal of Robotics Research*, 36(1):24–43.
- [143] Rayleigh, L. (1880). XII. On the resultant of a large number of vibrations of the same pitch and of arbitrary phase. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 10(60):73–78.
- [144] Razavian, Ali., S., Azizpour, H., Sullivan, J., and Carlsson, S. (2014). CNN features off-the-shelf: an astounding baseline for recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 806–813.
- [145] Reinhard, E., Khan, E. A., Akyuz, A. O., and Johnson, G. (2011). *Color Imaging: Fundamentals and Applications*. A. K. Peters/CRC Press, New York, 1st edition.
- [146] Robert Sobol (2002). Improving the Retinex algorithm for rendering wide dynamic range photographs. In *Human Vision and Electronic Imaging VII*, volume 4662, pages 341–348.

- [147] Roberts, J. M., Wheeler, A., Freiwald, A., and Cairns, S. (2009). *Cold-Water Corals: The Biology and Geology of Deep-Sea Coral Habitats*. Cambridge University Press, Cambridge.
- [148] Robertson, A. R. (1977). The CIE 1976 Color-Difference Formulae. *Color Research & Application*, 2(1):7–11.
- [149] Ross, R. E. and Howell, K. L. (2013). Use of predictive habitat modelling to assess the distribution and extent of the current protection of ‘listed’ deep-sea habitats. *Diversity and Distributions*, 19(4):433–445.
- [150] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.
- [151] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3):211–252.
- [152] Sala, E., Lubchenco, J., Grorud-Colvert, K., Novelli, C., Roberts, C., and Sumaila, U. R. (2018). Assessing real progress towards effective ocean protection. *Marine Policy*, 91:11–13.
- [153] Salman, A., Jalal, A., Shafait, F., Mian, A., Shortis, M., Seager, J., and Harvey, E. (2016). Fish species classification in unconstrained underwater environments based on deep learning. *Limnology and Oceanography: Methods*, 14(9):570–585.

- [154] Schettini, R. and Corchs, S. (2010). Underwater image processing: state of the art of restoration and image enhancement methods. *EURASIP Journal on Advances in Signal Processing*, 2010:1–7.
- [155] Schlacher, T. A., Williams, A., Althaus, F., and Schlacher-Hoenlinger, M. A. (2010). High-resolution seabed imagery as a tool for biodiversity conservation planning on continental margins. *Marine Ecology*, 31(1):200–221.
- [156] Schoening, T., Bergmann, M., and Nattkemper, T. W. (2012a). Investigation of hidden parameters influencing the automated object detection in images from the deep seafloor of the HAUSGARTEN observatory. In *2012 Oceans*, pages 1–5.
- [157] Schoening, T., Bergmann, M., Ontrup, J., Taylor, J., Dannheim, J., Gutt, J., Purser, A., and Nattkemper, T. W. (2012b). Semi-Automated Image Analysis for the Assessment of Megafaunal Densities at the Arctic Deep-Sea Observatory HAUSGARTEN. *PLOS ONE*, 7(6):e38179.
- [158] Schoening, T., Durden, J., Preuss, I., Branzan Albu, A., Purser, A., De Smet, B., Dominguez-Carrió, C., Yesson, C., de Jonge, D., Lindsay, D., Schulz, J., Möller, K. O., Beisiegel, K., Kuhnz, L., Hoeberechts, M., Piechaud, N., Sharuga, S., and Treibitz, T. (2017). Report on the Marine Imaging Workshop 2017. *Research Ideas and Outcomes*, 3:e13820.
- [159] Scholte, H. S., Ghebreab, S., Waldorp, L., Smeulders, A. W. M., and Lamme, V. A. F. (2009). Brain responses strongly correlate with Weibull image statistics when processing natural images. *Journal of Vision*, 9(4):29.

- [160] Schratzberger, M. and Ingels, J. (2018). Meiofauna matters: The roles of meiofauna in benthic ecosystems. *Journal of Experimental Marine Biology and Ecology*, 502:12–25.
- [161] Seiler, J., Friedman, A., Steinberg, D., Barrett, N., Williams, A., and Holbrook, N. J. (2012). Image-based continental shelf habitat mapping using novel automated data extraction techniques. *Continental Shelf Research*, 45:87–97.
- [162] Shan, S., Gao, W., Cao, B., and Zhao, D. (2003). Illumination normalization for robust face recognition against varying lighting conditions. In *2003 IEEE International SOI Conference. Proceedings (Cat. No.03CH37443)*, pages 157–164.
- [163] Shank, T. M., Fornari, D. J., Von Damm, K. L., Lilley, M. D., Haymon, R. M., and Lutz, R. A. (1998). Temporal and spatial patterns of biological community development at nascent deep-sea hydrothermal vents (9°50N, East Pacific Rise). *Deep Sea Research Part II: Topical Studies in Oceanography*, 45(1):465–515.
- [164] Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423.
- [165] Sharma, G., editor (2003). *Digital Color Imaging Handbook*. CRC Press, USA, 1st edition.
- [166] Sheehan, E. V., Cousens, S. L., Nancollas, S. J., Stauss, C., Royle, J., and Attrill, M. J. (2013). Drawing lines at the sand: Evidence for functional vs. visual reef boundaries in temperate Marine Protected Areas. *Marine Pollution Bulletin*, 76(1):194–202.

- [167] Sheehan, E. V., Stevens, T. F., and Attrill, M. J. (2011). A Quantitative, Non-Destructive Methodology for Habitat Characterisation and Benthic Monitoring at Off-shore Renewable Energy Developments. *PloS one*, 5(12):e14461.
- [168] sheehangroupblog (2019). SEA Wave. Available at: <https://sheehanresearchgroup.com/seawave/> [Accessed 13 Dec. 2022].
- [169] Shields, J., Pizarro, O., and Williams, S. B. (2020). Towards Adaptive Benthic Habitat Mapping. pages 9263–9270.
- [170] Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations*, San Diego, USA.
- [171] skorch (2022). skorch 0.12.1 documentation. Available at: <https://skorch.readthedocs.io/en/stable/> [Accessed 22 Dec. 2022].
- [172] Smith, A. R. (1978). Color gamut transform pairs. *ACM SIGGRAPH Computer Graphics*, 12(3):12–19.
- [173] Snelgrove, P. V. R. (1999). Getting to the Bottom of Marine Biodiversity: Sedimentary Habitats: Ocean bottoms are the most widespread habitat on Earth and support high biodiversity and key ecosystem services. *BioScience*, 49(2):129–138.
- [174] Solan, M., Germano, J. D., Rhoads, D. C., Smith, C., Michaud, E., Parry, D., Wenzhöfer, F., Kennedy, B., Henriques, C., Battle, E., Carey, D., Iocco, L., Valente, R., Watson, J., and Rosenberg, R. (2003). Towards a greater understanding of pat-

- tern, scale and process in marine benthic systems: a picture is worth a thousand worms. *Journal of Experimental Marine Biology and Ecology*, 285-286:313–338.
- [175] Sonka, M., Hlavac, V., and Boyle, R. (2014). *Image processing, analysis, and machine vision*. Cengage Learning.
- [176] Szegedy, C., Wei, L., Yangqing, J., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9.
- [177] Tan, M., Langenkämper, D., and Nattkemper, T. W. (2022). The Impact of Data Augmentations on Deep Learning-Based Marine Object Classification in Benthic Image Transects. *Sensors*, 22(14):5383.
- [178] Taylor, J., Krumpen, T., Soltwedel, T., Gutt, J., and Bergmann, M. (2017). Dynamic benthic megafaunal communities: Assessing temporal variations in structure, composition and diversity at the Arctic deep-sea observatory HAUSGARTEN between 2004 and 2015. *Deep Sea Research Part I: Oceanographic Research Papers*, 122:81–94.
- [179] Thiem, , Ravagnan, E., Fosså, J. H., and Berntsen, J. (2006). Food supply mechanisms for cold-water corals along a continental shelf edge. *Journal of Marine Systems*, 60(3):207–219.
- [180] Thurber, A. R., Sweetman, A. K., Narayanaswamy, B. E., Jones, D. O. B., Ingels, J., and Hansman, R. L. (2014). Ecosystem function and services provided by the deep sea. *Biogeosciences*, 11(14):3941–3963.

- [181] Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34:273–286.
- [182] Town, C., Marshall, A., and Sethasathien, N. (2013). Manta Matcher: automated photographic identification of manta rays using keypoint features. *Ecology and Evolution*, 3(7):1902–1914.
- [183] Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., Mathis, A., Mathis, M. W., van Langevelde, F., Burghardt, T., Kays, R., Klinck, H., Wikelski, M., Couzin, I. D., van Horn, G., Crofoot, M. C., Stewart, C. V., and Berger-Wolf, T. (2022). Perspectives in machine learning for wildlife conservation. *Nature Communications*, 13(1):792.
- [184] Tyler, P. A., Baker, M., and Ramirez-Llodra, E. (2016). Deep-Sea Benthic Habitats. In *Biological Sampling in the Deep Sea*, pages 1–15. John Wiley & Sons, Ltd.
- [185] United Nations (2018). Revised roadmap for the UN decade of ocean science for sustainable development. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000265141> [Accessed 23 Dec. 2022].
- [186] United Nations General Assembly (2015). Transforming our world: the 2030 Agenda for Sustainable Development. Available at: <https://sdgs.un.org/2030agenda> [Accessed 23 Dec. 2022].
- [187] Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I.

- [188] Wang, H., Chen, X., Xu, B., Du, S., and Li, Y. (2021a). An Improved MSCNN Method for Underwater Image Defogging. In *2021 IEEE International Conference on Artificial Intelligence and Industrial Design (AIID)*, pages 296–302.
- [189] Wang, S., Ma, K., Yeganeh, H., Wang, Z., and Lin, W. (2015). A Patch-Structure Representation Method for Quality Assessment of Contrast Changed Images. *IEEE Signal Processing Letters*, 22(12):2387–2390.
- [190] Wang, Y., Ding, X., Wang, R., Zhang, J., and Fu, X. (2017). Fusion-based underwater image enhancement by wavelet decomposition. In *2017 IEEE International Conference on Industrial Technology (ICIT)*, pages 1013–1018.
- [191] Wang, Y., Huang, H., Rudin, C., and Shaposhnik, Y. (2021b). Understanding How Dimension Reduction Tools Work: An Empirical Approach to Deciphering t-SNE, UMAP, TriMap, and PaCMAP for Data Visualization. *Journal of Machine Learning Research*, 22(201):1–73.
- [192] Wang, Y., Li, N., Li, Z., Gu, Z., Zheng, H., Zheng, B., and Sun, M. (2018). An imaging-inspired no-reference underwater color image quality assessment metric. *Computers & Electrical Engineering*, 70:904–913.
- [193] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society*, 13(4):600–612.
- [194] Weibull, W. (1951). A Statistical Distribution Function of Wide Applicability. *Journal of Applied Mechanics*, 18(3):293–297.

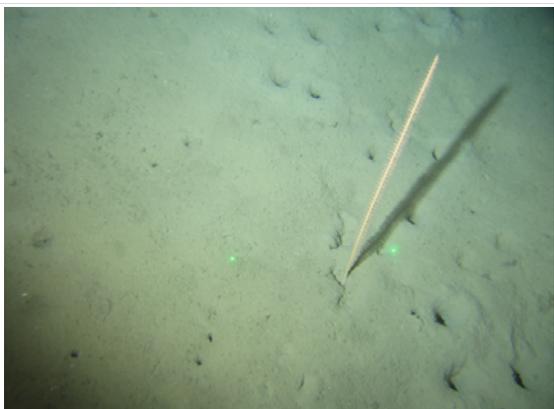
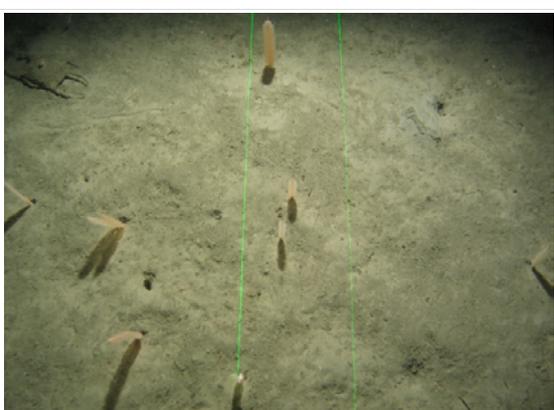
- [195] Weinstein, B. G. (2018). A computer vision for animal ecology. *Journal of Animal Ecology*, 87(3):533–545.
- [196] Williams, S. B., Pizarro, O., Steinberg, D. M., Friedman, A., and Bryson, M. (2016). Reflections on a decade of autonomous underwater vehicles operations for marine survey at the Australian Centre for Field Robotics. *Annual Reviews in Control*, 42:158–165.
- [197] Williams, S. B., Pizarro, O. R., Jakuba, M. V., Johnson, C. R., Barrett, N. S., Babcock, R. C., Kendrick, G. A., Steinberg, P. D., Heyward, A. J., Doherty, P. J., Mahon, I., Johnson-Roberson, M., Steinberg, D., and Friedman, A. (2012). Monitoring of Benthic Reference Sites: Using an Autonomous Underwater Vehicle. *IEEE Robotics & Automation Magazine*, 19(1):73–84.
- [198] Wilson, J. B. (1979). ‘Patch’ development of the deep-water coral *Lophelia Pertusa* (L.) on Rockall Bank. *Journal of the Marine Biological Association of the United Kingdom*, 59(1):165–177.
- [199] Wynn, R., Bett, B., Evans, A., Griffiths, G., Huvenne, V., A.R. Jones, M.R. Palmer, D. Dove, J.A. Howe, T.J. Boyd, and MAREMAP Partners (2013). Investigating the feasibility of utilizing AUV and Glider technology for mapping and monitoring of the UK MPA network. Final report for Defra project MB0118. Technical report, National Oceanography Centre, Southampton, GB.
- [200] Wynn, R. B., Huvenne, V. A. I., Le Bas, T. P., Murton, B. J., Connelly, D. P., Bett, B. J., Ruhl, H. A., Morris, K. J., Peakall, J., Parsons, D. R., Sumner, E. J., Darby, S. E., Dorrell, R. M., and Hunt, J. E. (2014). Autonomous Underwater Vehicles (AUVs): Their past, present and future contributions to the advancement of marine geoscience. *Marine Geology*, 352:451–468.

- [201] Wyszecki, G. and Stiles, W. (1982). *Color Science: Concepts and Methods, Quantitative Data and Formulae*. Wiley, 2nd edition.
- [202] Yamada, T., Massot-Campos, M., Prügel-Bennett, A., Pizarro, O., Williams, S., and Thornton, B. (2022a). Guiding Labelling Effort for Efficient Learning With Georeferenced Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1.
- [203] Yamada, T., Prügel-Bennett, A., and Thornton, B. (2021). Learning features from georeferenced seafloor imagery with location guided autoencoders. *Journal of Field Robotics*, 38(1):52–67.
- [204] Yamada, T., Prügel-Bennett, A., Williams, S. B., Pizarro, O., and Thornton, B. (2022b). GeoCLR: Georeference Contrastive Learning for Efficient Seafloor Image Interpretation. *ArXiv*, abs/2108.06421.
- [205] Yang, D., Martinez, C., Visuña, L., Khandhar, H., Bhatt, C., and Carretero, J. (2021a). Detection and analysis of COVID-19 in medical images using deep learning techniques. *Scientific Reports*, 11(1):19638.
- [206] Yang, H., Ni, J., Gao, J., Han, Z., and Luan, T. (2021b). A novel method for peanut variety identification and classification by Improved VGG16. *Scientific Reports*, 11(1):15756.
- [207] Yang, M. and Sowmya, A. (2015). An Underwater Color Image Quality Evaluation Metric. *IEEE Transactions on Image Processing*, 24(12):6062–6071.
- [208] Yanulevskaya, V., Geusebroek, J.M., Ranchordas, A., Araújo, H., and Intelligent Sensory Information Systems (IVI, FNWI) (2009). Significance of the Weibull distribution

- and its sub-models in natural image statistics. In *Fourth International Conference on Computer Vision Theory and Applications (VISAPP 2009)*, Lisboa, Portugal. INSTICC Press.
- [209] Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). How transferable are features in deep neural networks? In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc.
- [210] Yu, W., Qian, C., and Baomin, Z. (1999). Image enhancement based on equal area dualistic sub-image histogram equalization method. *IEEE Transactions on Consumer Electronics*, 45(1):68–75.
- [211] Zhang, B., Xie, F., and Han, F. (2019). Fish Population Status Detection Based on Deep Learning System. In *2019 IEEE International Conference on Mechatronics and Automation (ICMA)*, pages 81–85.
- [212] Zhang, W., Pan, X., Xie, X., Li, L., Wang, Z., and Han, C. (2021). Color correction and adaptive contrast enhancement for underwater image enhancement. *Computers & Electrical Engineering*, 91:106981.
- [213] Zhang, W., Wang, Y., and Li, C. (2022). Underwater Image Enhancement by Attenuated Color Channel Correction and Detail Preserved Contrast Enhancement. *IEEE Journal of Oceanic Engineering*, 47(3):718–735.
- [214] Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. In *Graphics gems IV*, pages 474–485. Academic Press Professional, Inc., USA.

# A Gardline Seabed Classification Guide

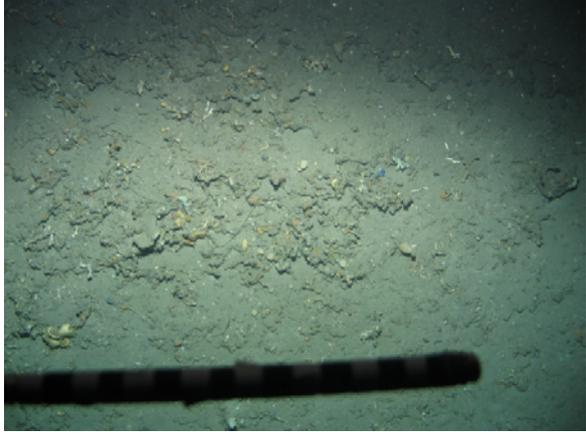
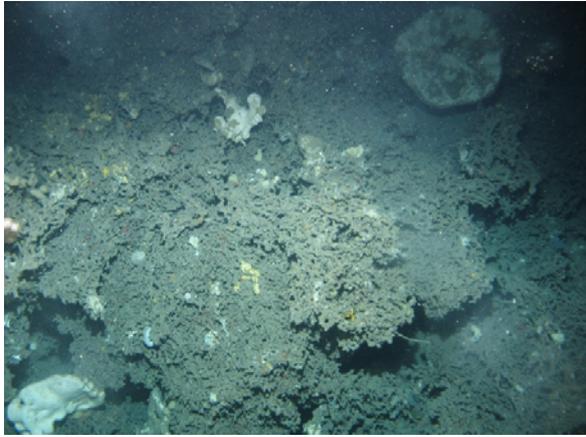
## APPENDIX A SEABED CLASSIFICATION GUIDE

Example Image	Class	Description
	A1	Soft sediments are the typical sediments expected to dominate the low reflectivity areas of the Norwegian Sea
	A2	Heavily bioturbated soft sediments
	A3	Single Sea pen / potential edge of sea pen community
	A4	Multiple Sea pens forming sea pen community.
		<p>Note: Various species of sea pen including, but not limited to: <i>Virgularia mirabilis</i>, <i>Kophobelemnon stelliferum</i>, <i>Funiculina quadrangularis</i>, <i>Umbellula encrinus</i> and <i>Pennatula phosphorea</i></p>

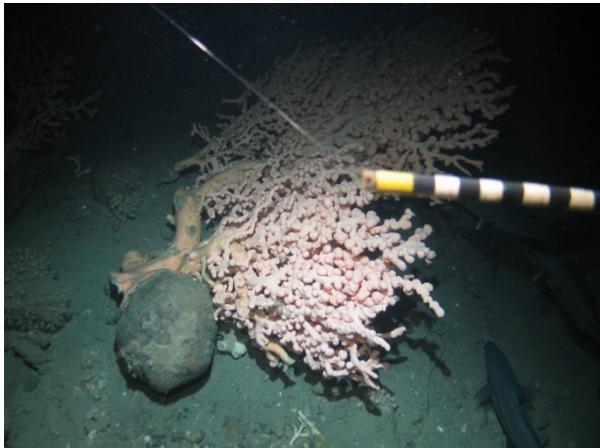
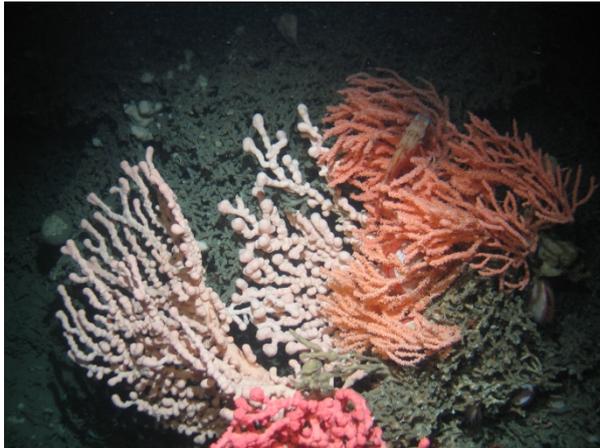
## APPENDIX A SEABED CLASSIFICATION GUIDE

Example Image	Class	Description
	B1	Gravel area
	B2	Scattered cobbles
	B3	Cobble and boulder area
	B4	Boulder area

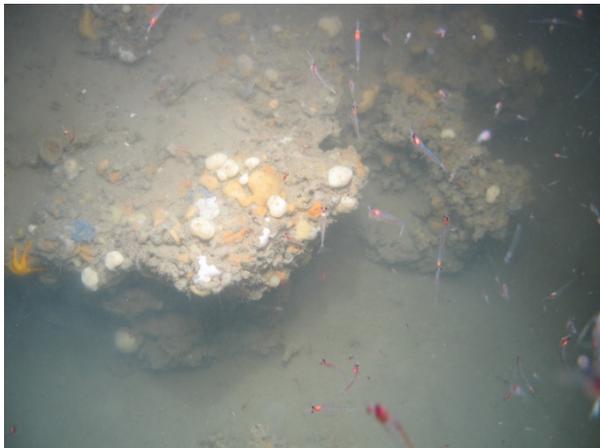
## APPENDIX A SEABED CLASSIFICATION GUIDE

Example Image	Class	Description
	C1	Coral rubble zone
	C2	Dead <i>Lophelia pertusa</i> reef framework
	C3	Live <i>Lophelia pertusa</i> reef,
		<p>Note: The live elements are usually located on the current facing side of the coral mound</p>

## APPENDIX A SEABED CLASSIFICATION GUIDE

Example Image	Class	Description
	D1	Lone soft corals
	D2	<p>Multiple colonies of a single soft coral species</p> <p>Note: Multiple colonies can be identified by their separate attachment points at the base.</p>
	D3	<p>Hard substrate soft coral community, multiple individuals with mixed species present</p> <p>Note: The bubblegum coral <i>Paragorgia arborea</i> is commonly present as two different colour morphs. This image therefore only contains two different species.</p>

## APPENDIX A SEABED CLASSIFICATION GUIDE

Example Image	Class	Description
	E	Soft substrate sponge community. Please refer to Table 1.1 for guide to densities
	F	Hard substrate sponge community. Please also use the density guides in Table 1.1.
	G	Marine derived authigenic carbonate concretions (MDAC)