

Physically Plausible Spectral Reconstruction from RGB Images

Yi-Tun Lin and Graham D. Finlayson

Colour & Imaging Lab, School of Computing Sciences
University of East Anglia, UK

Yi-Tun.Lin@uea.ac.uk, G.Finlayson@uea.ac.uk

Abstract

Recently Convolutional Neural Networks (CNN) have been used to reconstruct hyperspectral information from RGB images, and this **spectral reconstruction** problem (SR) can often be solved with good (low) error. However, little attention has been paid on whether these models' behavior can adhere to physics. We show that the leading CNN method introduces unexpected 'colorimetric errors', which means the recovered spectra do not reproduce ground-truth RGBs, and sometimes this discrepancy can be large. The problem is further compounded by exposure change. Indeed, most CNN models over-fit to fixed exposure and we demonstrate that this can result in poor performance when exposure varies.

In this paper we show how CNN learning can be extended so that the physical plausibility of SR is enforced. Remarkably, our physically plausible CNN solutions advance both spectral and colorimetric performance of the original network, while the application of data augmentation trades off the network performance for model stability against varying exposure.

1. Introduction

Hyperspectral imaging devices are developed to capture high resolution radiance spectra at every pixel in an image, namely the *hyperspectral images*. These images often record additional scene information that are 'invisible' to human eyes and consumer RGB cameras (where the spectral information is recorded with only 3 intensity values per pixel), which has been found useful in numerous computer vision applications including remote sensing [40, 12, 20, 39, 11], anomaly detection [24] and medical imaging [44, 45], as well as computer graphics applications such as scene relighting [26] and digital art archiving [43].

Recent development in hyperspectral technology seeks faster image capturing speed compared to the conventional scanning-based techniques [19, 21]. Several attempts have

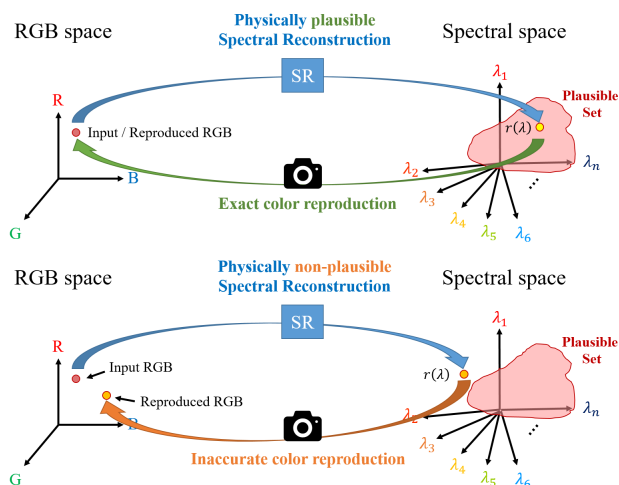


Figure 1: Physically plausible spectral reconstruction (top) and non-plausible spectral reconstruction (bottom).

been made for real-time multi-channel capturing [9, 42, 32, 38]. However, these devices are complicated and/or bulky that limits their usefulness. Other designs deploy novel optical components with specialized post-processing algorithms [13, 18, 6, 17, 28, 34, 47]. But, these devices trade off spatial resolution and/or light sensitivity.

Spectral reconstruction (SR) is an alternative approach to recording hyperspectral information, where hyperspectral images are recovered from RGB images [31, 3, 27, 25, 36, 16, 8, 27, 23, 4, 1, 16, 35, 5]. The idea is not as naïve as it might first appear. Indeed, we are expecting an RGB, which has just 3 numbers, to recover much more than 3 degrees of freedom in spectra. Fortunately, in natural scenes significant portion of spectral variation is covered by its color appearance (*i.e.* the RGBs) [10], which makes it possible for learning approaches to give rather accurate spectral approximations. Recent approaches, leading by Convolutional Neural Networks (CNN), utilize the images' high-level contents to further enhance the recovery accuracy.

A key concern of this paper is the *physical plausibility* of the SR algorithms. In reality, the RGB colors are physi-

cally related to spectra: using the spectral sensitivities of an RGB camera, the RGBs can be accurately reproduced from the spectra. We argue that a physically plausible SR **must** recover spectra that can also reproduce (following the underlying physical conversion and explicit image processing pipeline) the exact ground-truth RGBs. Unfortunately, we do not see any deep-network-based SR explicitly ensures the compliance to this reality. Indeed, as we shall show in this paper, the reproduced RGBs can be quite far from the actual ones. This implies that the modern SR solutions ignore the known physical relation between actual spectra and RGBs, and yet giving some predictions that clearly contradict the physics - put bluntly, these algorithms provide the estimations of spectra which **must be the wrong answers**. Therefore, we re-consider the usage of this physical prior as in some early works [2, 46, 48, 7, 30] to further bound the spectral accuracy of a CNN model.

In Figure 1 we illustrate our problem definition. In the top diagram, a physically plausible SR recovers a radiance spectra (*i.e.* a spectral power distribution $r(\lambda)$, see the right side of the image) from an RGB (the red point on the left in RGB space). Now we reintegrate the spectra with the camera’s spectral sensitivities - just like taking a picture of this spectrum - which gives a reproduced RGB. In this case the input RGB and the reproduced counterpart are the same.

The diagram in the bottom half of Figure 1 shows a spectral recovery which produces incorrect color when reintegrated with the camera sensitivities. This ‘physically non-plausible SR’ is the norm (and is a feature exhibited by all deep-network solutions we are aware of).

In Figure 2 we show a pictorial example. Two reconstructed spectra in the bottom-right panel - red and purple dotted curves - have the same spectral difference from the ground-truth (blue solid curve). Integrating with the color matching functions of sRGB displays (the top-right panel), the purple curve reproduces the background color exactly as the original painting on the left. In contrary, the red curve reproduces the image in the middle, which shows severe background color shift. This teaches that the level of spectral error does *not* determine colorimetric error, and the uncertainty can be significant.

Yet another physical reality addressed in this paper is *exposure invariance*. Clearly, if the light intensity in the scene changes, both the physically measured spectra (hyperspectral images) and the linear RGB images (commonly considered as the ground-truth RGBs) will be linearly scaled. However, Lin and Finlayson [29] showed that the CNN-based ‘state-of-the-art’ models were trained for a confined exposure condition and they perform poorly when a different exposure setting is tested.

Overall, this paper makes three main contributions:

- We evaluate one of the state-of-the-art CNN-based models [5, 35] to gauge the extent that it delivers physically

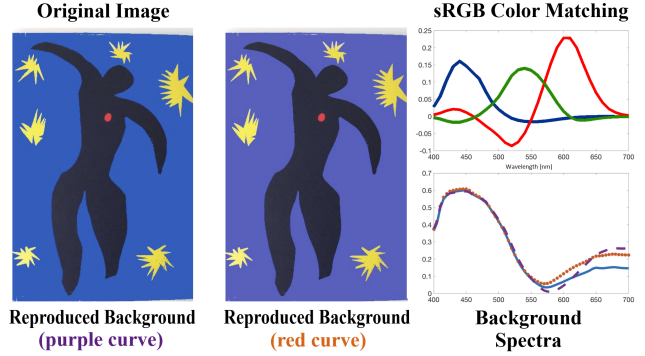


Figure 2: Background color reproduction from two spectra (the purple and red curves in the bottom-right panel).

plausible spectral recovery, either in the sense of reproducing the input RGBs or being resilient to a varying exposure.

- We propose a novel framework which ensures *exact* color reproduction in CNN-based SR that further improves the model’s spectral recovery performance.
- We deploy a data augmentation process that maintains model stability over different exposure settings.

2. Related Work

Hyperspectral imaging. There exist technologies where hyperspectral images can be directly captured, and these include using a prism-mask system [9], multiple cameras [41, 32] and faced reflectors [38]. However, the practical application of these devices is limited by their complex configurations and/or their physical bulkiness. Alternately, in *compressive imaging*, a scene’s spectral information is encoded in alternative forms on the sensed 2-D images. But, there is the overhead of decompressing the signal. Examples include multi-spectral color filter array [13], coded aperture [6, 17, 18], diffractive gratings [28], digital micro-mirror device [34] and most recently random printed mask [47]. Other problems inherent in compressive sensing are the need for specialized optics and the inherent trade-off between the number of sensors and/or light sensitivity and the spatial resolution.

Spectral reconstruction (SR). Rather than building new hardware for capturing hyperspectral images, spectral reconstruction attempts to map RGB images to their spectral counterparts. Shallow-learned methods - of which sparse coding is the best example [4, 1] - have the advantage of model simplicity and quick training. However, these models are effectively implementing a ‘one-to-one’ lookup table, which contradicts the fact that *many* (in fact *infinite*) spectra can reproduce the same RGB.

In the CNN approach the implementation complexity is much higher as so the hardware requirements but the re-

construction is richer. The promise of these methods is that, in an intermediate representation, they might identify scene contents which are associated with the target spectra and then effectively use these information in the recovery process. Indeed, it is well known that faces, chlorophyll (in foliage) and daylights have very characteristic shapes (amongst other scene features). Of the current developments, deep neural networks [3, 25, 36, 16, 35] provide the leading performance in spectral reconstruction.

Physical plausibility. Some of the early models can already provide accurate color reproduction (but are generally believed to perform poorer in spectral recovery comparing to the recent CNN methods). This includes color-difference-weighted PCA [2], linear regression with colorimetric correction [46] and the colorimetrically constrained iterative optimizations [48, 7] and Bayesian inference [30].

In the recent NTIRE Challenge on Spectral Reconstruction from RGB Images in 2018 (hereinafter abbreviated as NTIRE18) [5], all 12 leading entries out of 73 attendants (on the ‘Clean Track’) involve the implementation of deep neural networks. However, none of these methods explicitly ensure the spectra can reproduce the input RGBs.

As for the issue of exposure invariance, we remark that in many learning-based computer vision tasks, the model stability over intensity change are considered; that is, the model are ensured to work well even as the scene exposure changes. However, Lin and Finlayson [29] demonstrated that the leading SR models in NTIRE18 perform poorly when different exposure settings are considered, and this has raised an issue that many modern developments of SR may not work in the wild where exposure can vary.

3. Physically Plausible Spectral Reconstruction

At each pixel of a hyperspectral image, a high-resolution radiance spectrum is recorded. For training the SR model we simulate the ground-truth linear RGBs by calculating the *inner products* between the measured radiance spectra and the spectral sensitivities of a given RGB camera:

$$\rho_k = \sum_{\lambda \in \Omega} s_k(\lambda) r(\lambda), \quad (1a)$$

where $k = 1, 2, 3$ refer to the red, green and blue channels of the RGB image, ρ_k , $s_k(\lambda)$ and $r(\lambda)$ are respectively the k^{th} camera response, the k^{th} camera’s spectral sensitivity function and the radiance’s spectral power distribution, λ denotes the wavelength dimension, and Ω is the *visible range*. Of course for this inner-product model (as oppose to an integral) of image formation to work, we must sample the spectra at a sufficient resolution across the visible spectrum. Let us vectorize Equation (1a):

$$\underline{\rho} = \mathbf{S}^T \underline{\mathbf{r}}, \quad (1b)$$

where $\underline{\rho} = (\rho_1, \rho_2, \rho_3)^T$ is the 3-dimensional RGB vector, $\underline{\mathbf{r}}$ is the n -dimensional radiance spectra with n to be the number of spectral bands, and $\mathbf{S} = (\underline{\mathbf{s}}_1, \underline{\mathbf{s}}_2, \underline{\mathbf{s}}_3)$ is an $n \times 3$ matrix with its columns to be the three camera’s spectral sensitivity functions. In all simulations we report later in this paper, we assume the visible range runs from 400 through 700 nanometers, and the spectra are sampled every 10 nanometers; hence $n = 31$ (this is the common assumption made in most studies, including in the NTIRE18).

Note that in this paper we only consider the linear RGBs derived by Equation (1a) and (1b). This setting aligns with the ‘Clean Track’ of NTIRE18 rather than the ‘Real-world Track’ where the camera’s spectral sensitivities are unknown and the RGBs are non-linearly rendered. Our intention is to, without further complicating the problem, discuss and manipulate the intrinsic properties of an SR model.

In the ordinary SR framework, a radiance spectrum $\underline{\mathbf{r}}$ is recovered from its RGB camera response $\underline{\rho}$: the algorithm searches for the best estimate of $\underline{\mathbf{r}}$ within the entire spectral space (*i.e.* \mathbb{R}^n) that statistically minimizes the distance error between the estimation and the ground-truth. However, this framework does not ensure that the recovered $\underline{\mathbf{r}}$ must reproduce $\underline{\rho}$ - the algorithm may find a solution which is spectrally close to the ground-truth but reproduces distant color (as per the example we showed in Figure 2).

Let us now develop a method to ‘constrain’ the algorithm only to search for the estimated radiance within the set of spectra that reproduce the correct input RGB. For this purpose, we propose a ‘plausible set’ concept, which is defined as *the set of all spectra that reproduce a given RGB*.

The derivation of our plausible set is analogous to, but simpler than, the *metamer set* in [15, 30]: their focus was on the reflectances instead of our case on the radiance spectra.

3.1. The Plausible Set

Given known camera’s spectral sensitivity functions \mathbf{S} , the plausible set \mathcal{P} is defined as:

$$\mathcal{P}(\underline{\rho}; \mathbf{S}) = \left\{ \underline{\mathbf{r}} \mid \mathbf{S}^T \underline{\mathbf{r}} = \underline{\rho} \right\}. \quad (2)$$

Geometrically, the outcome of an inner product is only affected by the parts of the two vectors that are ‘parallel’ to each other, whereas the ‘perpendicular’ part do not contribute to the inner product. Given this view, with respect to the constraint $\mathbf{S}^T \underline{\mathbf{r}} = \underline{\rho}$ we can separate $\underline{\mathbf{r}}$ into two parts: the part that is spanned by the column vectors of \mathbf{S} which contributes to $\underline{\rho}$, and the part lies in the null-space of \mathbf{S} which yields *zero projection*.

Define \mathcal{S} as the vector space spanned by the columns of \mathbf{S} , called the *camera subspace*, then an arbitrary spectrum:

$$\underline{\mathbf{r}} \in \mathbb{R}^n = \mathcal{S} \oplus \text{Null}(\mathbf{S}) \quad (3)$$

(recall n is the number of spectral bands). We can further decompose any given \mathbf{r} into two components: $\mathbf{r}^\parallel \in \mathcal{S}$ and $\mathbf{r}^\perp \in \text{Null}(\mathcal{S})$, such that

$$\mathbf{r} = \mathbf{r}^\parallel + \mathbf{r}^\perp \quad (4)$$

subject to

$$\begin{cases} \mathbf{S}^\top \mathbf{r}^\parallel = \underline{\rho} \\ \mathbf{S}^\top \mathbf{r}^\perp = \mathbf{0} \end{cases} \quad (5a)$$

$$(5b)$$

Theorem 1. *All members of a given $\mathcal{P}(\underline{\rho}; \mathcal{S})$ share the same \mathbf{r}^\parallel component.*

Proof. Consider an arbitrary $\mathbf{r} \in \mathcal{P}(\underline{\rho}; \mathcal{S})$, its \mathbf{r}^\parallel component can be derived directly via subspace projection. The projection matrix onto \mathcal{S} is written as:

$$\mathbf{P}^\mathcal{S} = \mathbf{S}(\mathbf{S}^\top \mathbf{S})^{-1} \mathbf{S}^\top, \quad (6)$$

such that

$$\mathbf{r}^\parallel = \mathbf{P}^\mathcal{S} \mathbf{r} = \mathbf{S}(\mathbf{S}^\top \mathbf{S})^{-1} \mathbf{S}^\top \mathbf{r}. \quad (7)$$

Next, the intrinsic constraint of $\mathcal{P}(\underline{\rho}; \mathcal{S})$ is $\mathbf{S}^\top \mathbf{r} = \underline{\rho}$ (see Equation (2)), which gives

$$\mathbf{r}^\parallel = \mathbf{S}(\mathbf{S}^\top \mathbf{S})^{-1} \underline{\rho}. \quad (8)$$

As \mathbf{S} and $\underline{\rho}$ are known factors of $\mathcal{P}(\underline{\rho}; \mathcal{S})$, \mathbf{r}^\parallel is therefore fixed (theorem proved).

Given Theorem 1 and Equation (4), we can further infer that among all members of a given plausible set $\mathcal{P}(\underline{\rho}; \mathcal{S})$, only their \mathbf{r}^\perp components defer them from each other. Since no additional constraint is placed on \mathbf{r}^\perp other than $\mathbf{r}^\perp \in \text{Null}(\mathcal{S})$, we get

$$\mathcal{P}(\underline{\rho}; \mathcal{S}) = \mathbf{S}(\mathbf{S}^\top \mathbf{S})^{-1} \underline{\rho} \oplus \text{Null}(\mathcal{S}). \quad (9)$$

Now $\text{Null}(\mathcal{S})$ can be further represented by the linear combinations of its *orthogonal basis*. Define a basis matrix \mathbf{N} whose columns are the orthogonal basis vectors of $\text{Null}(\mathcal{S})$. This \mathbf{N} matrix (*i.e.* the orthogonal basis) can be obtained via singular value decomposition (SVD) on the null-space projection matrix $\mathbf{P}^\mathcal{N} = \mathbf{I}_{n \times n} - \mathbf{P}^\mathcal{S}$ ($\mathbf{I}_{n \times n}$ is the $n \times n$ identity matrix). Note that there should be $n - 3$ linearly-independent basis vectors that are perpendicular to all 3 column vectors of \mathbf{S} (*i.e.* the camera's spectral sensitivities) in the spectral space \mathbb{R}^n , therefore the rank of $\mathbf{P}^\mathcal{N}$ is $n - 3$ and \mathbf{N} is $n \times (n - 3)$.

Finally, given the calculated null-space basis matrix \mathbf{N} , we reach the following definition of our plausible set:

$$\mathcal{P}(\underline{\rho}; \mathcal{S}) = \left\{ \mathbf{S}(\mathbf{S}^\top \mathbf{S})^{-1} \underline{\rho} + \mathbf{N} \underline{\alpha} \mid \underline{\alpha} \in \mathbb{R}^{n-3} \right\}. \quad (10)$$

We call this $\underline{\alpha}$ vector the *null-space coefficients*, and each one of $\underline{\alpha}$'s corresponds to one \mathbf{r} in \mathcal{P} (*i.e.* \mathbf{r} and $\underline{\alpha}$ is one-to-one).

This way of isolating the constrained constant \mathbf{r}^\parallel while leaving \mathbf{r}^\perp decided by an *unbounded* factor $\underline{\alpha}$ in \mathbb{R}^{n-3} (*i.e.* $\mathbf{r}^\perp = \mathbf{N} \underline{\alpha}$) is the key to implementing our physically plausible SR based on CNN. Indeed, the existing CNN solutions are designed to search for spectral approximations in an unbounded vector space. In the next part of this section, we are going to introduce our physically plausible SR via 'null-space coefficients reconstruction'.

3.2. Reconstructing the Null-space Coefficients

Let us denote \mathbf{r}_{gt} and \mathbf{r}_{rec} as respectively the ground-truth and reconstructed spectra. The ground-truth RGB $\underline{\rho}_{\text{gt}}$ can be calculated by Equation (1b), and naturally \mathbf{r}_{gt} is a member of $\mathcal{P}(\underline{\rho}_{\text{gt}}; \mathcal{S})$. Then, we would like to search for \mathbf{r}_{rec} in $\mathcal{P}(\underline{\rho}_{\text{gt}}; \mathcal{S})$, which implies that \mathbf{r}_{gt} and \mathbf{r}_{rec} belong to the same plausible set so that \mathbf{r}_{rec} is physically plausible.

Since \mathbf{r} and $\underline{\alpha}$ is one-to-one, given \mathbf{r}_{gt} we can find the corresponding $\underline{\alpha}_{\text{gt}}$ (we shall explain the calculation later). Next, we re-train the SR algorithm; but instead of the original search on \mathbf{r}_{gt} , we now expect the SR model to search for the approximation of the null-space coefficients $\underline{\alpha}_{\text{gt}}$. Denote this modified SR as $\Psi(\cdot) : \mathbb{R}^3 \mapsto \mathbb{R}^{n-3}$, which gives

$$\Psi(\underline{\rho}_{\text{gt}}) = \underline{\alpha}_{\text{rec}} \approx \underline{\alpha}_{\text{gt}}, \quad (11)$$

the reconstructed spectrum \mathbf{r}_{rec} , which is ensured to be a member of $\mathcal{P}(\underline{\rho}_{\text{gt}}; \mathcal{S})$, is then derived by

$$\mathbf{r}_{\text{rec}} = \mathbf{S}(\mathbf{S}^\top \mathbf{S})^{-1} \underline{\rho}_{\text{gt}} + \mathbf{N} \left(\Psi(\underline{\rho}_{\text{gt}}) \right). \quad (12)$$

Note that now the SR model Ψ is trained to map RGBs to the lower-dimensional $\underline{\alpha}$'s; corresponding to the CNN's architecture, we only need to change the spectral dimension in the output layer from n to $n - 3$.

So far, the idea behind our physically plausible SR framework has been established. Still, to train a CNN model the *ground-truth labels* are necessary, which means we are yet to calculate $\underline{\alpha}_{\text{gt}}$ from \mathbf{r}_{gt} in order to train Ψ .

In Equation (6) we calculated the projection matrix $\mathbf{P}^\mathcal{S}$ which projects \mathbf{r} onto the column space of \mathbf{S} that derives \mathbf{r}^\parallel . Likewise, we can also derive \mathbf{r}^\perp by projecting \mathbf{r} onto the column space of the null-space basis \mathbf{N} . The null-space projection matrix $\mathbf{P}^\mathcal{N} = \mathbf{I}_{n \times n} - \mathbf{P}^\mathcal{S}$ can be alternatively written as:

$$\mathbf{P}^\mathcal{N} = \mathbf{N}(\mathbf{N}^\top \mathbf{N})^{-1} \mathbf{N}^\top, \quad (13)$$

such that

$$\mathbf{r}_{\text{gt}}^\perp = \mathbf{P}^\mathcal{N} \mathbf{r}_{\text{gt}} = \mathbf{N}(\mathbf{N}^\top \mathbf{N})^{-1} \mathbf{N}^\top \mathbf{r}_{\text{gt}}. \quad (14)$$

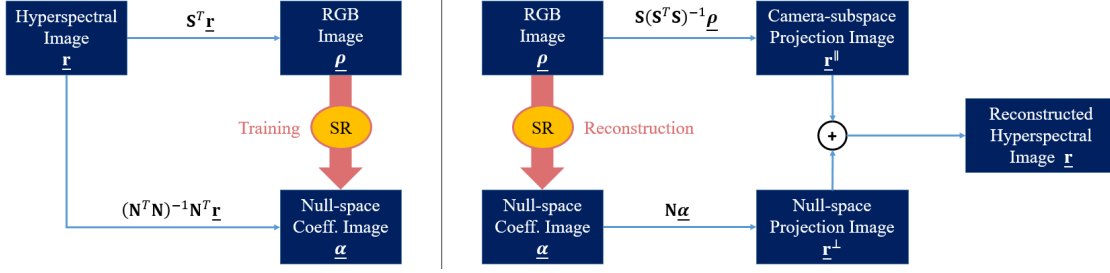


Figure 3: The training (left) and reconstruction scheme (right) of our physically plausible spectral reconstruction.

Since $\underline{\mathbf{r}}^{\perp} = \mathbf{N} \underline{\alpha}$, we get

$$\mathbf{N} \underline{\alpha}_{\text{gt}} = \mathbf{N} \left((\mathbf{N}^T \mathbf{N})^{-1} \mathbf{N}^T \underline{\mathbf{r}}_{\text{gt}} \right). \quad (15)$$

Finally, since the columns of \mathbf{N} are linearly-independent basis vectors, $\underline{\alpha}$ and $\underline{\mathbf{r}}^{\perp}$ is one-to-one, which gives

$$\underline{\alpha}_{\text{gt}} = (\mathbf{N}^T \mathbf{N})^{-1} \mathbf{N}^T \underline{\mathbf{r}}_{\text{gt}}. \quad (16)$$

Figure 3 summarizes our physically plausible SR framework. In the training stage (on the left side of the figure), the SR model is trained to map the RGB image to the null-space coefficients image. In the reconstruction stage (on the right), the camera-subspace projection $\underline{\mathbf{r}}^{\parallel}$ is calculated directly from the input RGB $\underline{\rho}$ while the SR algorithm only concerns the recovery of the null-space coefficients $\underline{\alpha}$, which subsequently decides the null-space projection $\underline{\mathbf{r}}^{\perp}$. As the color reproductions of a spectrum (with the underlying RGB camera’s spectral sensitivities \mathbf{S}) only depend on their $\underline{\mathbf{r}}^{\parallel}$ (see Equation (5a) and (5b)), the reconstructed hyperspectral image is ensured to reproduce exactly the input RGB image. In the next section we are going to integrate this framework with a CNN-based SR model.

4. Implementation

We built our models based on the HSCNN-R architecture, which was the 2nd place entry of the NTIRE18 [5, 35] (whose performance is similar to the 1st place HSCNN-D model; we used the 2nd place architecture simply because it was simpler in our development environment). As illustrated in Figure 4, the HSCNN-R model adopts a deep residual learning framework [22]. Each of the residual blocks is constructed with two convolutional layers and one ReLU layer. The model also adopts a global residual learning structure. All convolutional kernels are set to 3×3 . On training and reconstruction, the network maps 50×50 RGB image patches to the corresponding 31-channel hyperspectral image patches. According to one of the reported settings in [35], we set the filter numbers in each layer to be 256 and the network depth to be 20.

In this paper, we aim for two improvements on HSCNN-R: (1) ensuring exact color reproduction and (2) enhancing the robustness against exposure change. For the former, we integrated our physically plausible framework shown in Figure 3 with HSCNN-R, and for the latter, we adopted a *data augmentation* process. To study the effects of both improvements, 3 new models listed in Table 1 were trained.

Model	Physically Plausible	Data Augmentation
HSCNN-R ^p	V	
HSCNN-R ^d		V
HSCNN-R ^{pd}	V	V

Table 1: List of our new models

4.1. Physically Plausible HSCNN-R

In the original HSCNN-R model, the output layer corresponds to a 50×50 hyperspectral image patch with 31 spectral dimensions. To accommodate our physically plausible framework in HSCNN-R^p and HSCNN-R^{pd}, we reduced the spectral dimension from 31 to 28 in the output layer for recovering the null-space coefficients image ($\underline{\alpha}$ is $(n - 3)$ -dimensional with $n = 31$).

Unlike the original hyperspectral data which only contains positive values, the null-space coefficients $\underline{\alpha}$ allow *negative* entries, and this is not permitted for the ReLU output layer of the HSCNN-R architecture. As a result, it is necessary to apply an *offset* to the ground-truth $\underline{\alpha}$ such that the negative values are prevented.

In our implementation, we found that empirically the entries of the ground-truth $\underline{\alpha}$ range between -1 and 1 . Hence, in the training stage of HSCNN-R^p we adopted a re-centering:

$$\tilde{\underline{\alpha}} = \frac{1}{2}(\underline{\alpha} + 1), \quad (17)$$

while in the reconstruction stage we center $\underline{\alpha}$ back from $\tilde{\underline{\alpha}}$. As we will mention later the HSCNN-R^{pd} model requires a different re-centering function due to the implementation of the additional *data augmentation* process.

The rest of the hyperparameters of HSCNN-R^p were kept the same as the original HSCNN-R model [35]. Our

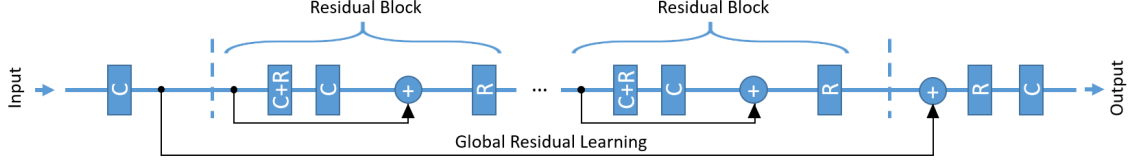


Figure 4: The HSCNN-R architecture [35]. ‘C’ means 3×3 convolution and ‘R’ refers to ReLU activation.

HSCNN- R^p is expected to provide absolute color reproduction. However, as demonstrated in [29], the original HSCNN-R model is not robust against scene exposure change, and this implies HSCNN- R^p may also deliver poor performance when a different testing exposure is applied.

4.2. Intensity-scaling Data Augmentation

We created the augmented data by simulating ‘brighter’ and ‘dimmer’ RGB images from the ground-truth hyperspectral images. Instead of generating all the augmented data before training the model, we decided to draw different scaling constants in real time: all input image patches (and the same patch in different training epochs) were scaled differently, in which way the network is ensured to see adequate intensity variations from the data.

Furthermore, we investigated the proper random distribution from which the scaling constants should be drawn. Presumably, we hope to implement a trained SR algorithm on an RGB camera. We notice that the standard exposure settings of a camera (*i.e.* the aperture size and shutter speed) follow *geometric progressions*; more precisely, the available diameters of the aperture normally follows a sequential scaling change by $\sqrt{2}$, and the shutter speed is adjusted by a factor of 2 between adjacent modes. To ensure that the trained SR model performs equally well in all these settings, we propose to draw the scaling constants ξ from a uniform distribution on a *log scale*:

$$\log_{\beta} \xi \sim \text{Uniform}(-1, 1). \quad (18)$$

In our implementation, we set $\beta = 10$ such that the scaling factor ξ is bounded by $[\frac{1}{10}, 10]$.

We trained HSCNN- R^d which only adopts the intensity-scaling data augmentation but not the physically plausible framework, and all hyperparameters were kept the same as provided in [35]. On the other hand, to train the HSCNN- R^{pd} model (implemented with both the physically plausible framework and data augmentation), similar to the case of HSCNN- R^p , we need to apply an offset on the ground-truth null-space coefficients. Due to the intensity scaling, the range of α is extended to $[-10, 10]$ (since $\beta = 10$), so the required re-centering function becomes:

$$\tilde{\alpha} = \frac{1}{20}(\alpha + 10). \quad (19)$$

Additionally, to make the model converge efficiently, we set

the adaptive learning rate to follow a polynomial decay with the power of 25 (instead of the original 1.5 in [35]).

5. Experiment

5.1. Experimental Setup

The aim of the experiment is to compare the three new models listed in Table 1 with the original HSCNN-R. In addition, we evaluated two shallow-learned models (non-network solutions) as our comparing benchmarks: the generic linear regression (LR) [23] and the leading sparse coding SR (A+) [1].

We trained all models based on the ICVL hyperspectral database (201 images) [4], where we randomly split the database into 100 images for training, 50 for validation and 50 for evaluation. The CIE 1964 color matching functions [14] were selected as the spectral sensitivity functions of the RGB camera (so the simulated ground-truth RGBs are the CIEXYZ color coordinates).

Our experiment concerns the performances of the models in terms of (1) spectral recovery (2) color reproduction (using the original and different RGB cameras) and (3) both performances under different exposure settings. Accordingly we selected the following error metrics:

- Spectral difference: Mean Relative Absolute Error

$$\text{MRAE} = \frac{1}{n} \left\| \frac{\mathbf{r}_{\text{gt}} - \mathbf{r}_{\text{rec}}}{\mathbf{r}_{\text{gt}}} \right\|_1 \quad (20)$$

- Color difference: CIE 1976 color difference

$$\Delta E = \sqrt{(L_{\text{gt}}^* - L_{\text{rec}}^*)^2 + (a_{\text{gt}}^* - a_{\text{rec}}^*)^2 + (b_{\text{gt}}^* - b_{\text{rec}}^*)^2} \quad (21)$$

In Equation (20), respectively \mathbf{r}_{gt} and \mathbf{r}_{rec} refers to the ground-truth and reconstructed radiance spectra, and n is the number of spectral bands. The division is component-wise and the L_1 norm is calculated.

Equation (21) shows the definition of the CIE 1976 color difference formula [33], where $(L_{\text{gt}}^*, a_{\text{gt}}^*, b_{\text{gt}}^*)$ and $(L_{\text{rec}}^*, a_{\text{rec}}^*, b_{\text{rec}}^*)$ are the CIELAB color coordinates of the ground-truth and reconstructed RGB colors, respectively. The transformation between CIEXYZ and CIELAB requires the normalization by the ‘white point’ coordinates (*i.e.* the illumination color) [37], for which we hand-crafted the white points of each images by selecting the RGB of the brightest achromatic pixel.

Model	Original exposure ($\xi = 1$)				Half exposure ($\xi = 0.5$)				Double exposure ($\xi = 2$)			
	MRAE ($\times 10^{-2}$)		ΔE		MRAE ($\times 10^{-2}$)		ΔE		MRAE ($\times 10^{-2}$)		ΔE	
	Mean	99.9pt	Mean	99.9pt	Mean	99.9pt	Mean	99.9pt	Mean	99.9pt	Mean	99.9pt
LR [23]	6.39	23.07	0.10	1.75	6.39	23.07	0.10	1.75	6.39	23.07	0.10	1.75
A+ [1]	3.79	23.26	0.02	1.51	3.79	23.26	0.02	1.51	3.79	23.26	0.02	1.51
HSCNN-R [35]	1.76	12.94	0.15	3.29	15.68	47.77	0.47	3.43	6.38	19.56	0.29	6.61
HSCNN-R ^p	1.74	15.12	0.00	0.00	15.84	45.30	0.00	0.00	6.44	20.07	0.00	0.00
HSCNN-R ^d	2.95	17.24	0.29	4.77	2.95	17.32	0.28	4.71	2.94	17.20	0.29	4.83
HSCNN-R ^{pd}	2.73	21.52	0.00	0.00	2.83	20.75	0.00	0.00	2.70	22.69	0.00	0.00

Table 2: The averaged mean and 99.9 percentile hyperspectral image reconstruction errors in MRAE and ΔE under original, half and double exposure settings. Best results are shown in **red**.

ΔE (SONY IMX135), Original exposure					
Physically Non-Plausible			Physically Plausible		
Model	Mean	99.9pt	Model	Mean	99.9pt
HSCNN-R	0.41	14.64	HSCNN-R ^p	0.38	15.12
HSCNN-R ^d	0.60	17.88	HSCNN-R ^{pd}	0.45	16.44

Table 3: Color reproduction error in ΔE under an alternative ‘SONY IMX135’ camera sensitivities.

We tested all models under 3 exposure settings: the original, half and double exposure (respectively $\xi = 1, 0.5$ and 2). For each testing exposure, we uniformly scaled-up all testing images with the same scaling constant, and the recovered hyperspectral images were compared with the ground-truth hyperspectral images scaled by the same constant.

5.2. Result and Discussion

The performance statistics shown in Table 2 are the averaged MRAE and ΔE of the mean and 99.9 percentile errors of individual testing images (the image dimension is around 1300×1392). Hereafter we refer the averaged 99.9 percentile errors to as the ‘worst-case’ performance.

First, compared to the shallow-learned linear regression (LR) and sparse coding (A+) model, the original HSCNN-R shows great advantage in spectral accuracy (MRAE) under original testing exposure (*i.e.* the training and testing exposures are the same). However, HSCNN-R introduces higher worst-case colorimetric error (ΔE), and as per the examples of visual comparisons on rendered images provided in Figure 3-8 of our supplementary material, this color discrepancy can be very perceivable. This tells that HSCNN-R can (and commonly do) recover spectra that are clearly not physically plausible. Also, in the aspect of exposure invariance, LR and A+ retain their performance as the testing exposure is halved and doubled; nevertheless, the HSCNN-R model performs considerably worse when the testing exposure changes. Indeed, under half testing exposure HSCNN-R even performs far worse than the generic LR.

Next, the implementation of our physically plausible framework on HSCNN-R^p and HSCNN-R^{pd} ensures *zero* ΔE and further improves the spectral recovery performance of their non-physically plausible counterparts, respectively

HSCNN-R and HSCNN-R^d. This is shown by the minute decrease in mean MRAE from HSCNN-R to HSCNN-R^p and from HSCNN-R^d to HSCNN-R^{pd}. However, the converse increase in the worst-case MRAE for these physically plausible models indicates possible trade-offs between different parts of the images.

Lastly, the consideration of intensity-scaling data augmentation *seems* to deteriorate spectral recovery. Indeed, the MRAE of HSCNN-R^d and HSCNN-R^{pd} are higher than the cases for HSCNN-R and HSCNN-R^p under the original testing exposure. However, they perform the best overall as testing exposure changes, and this ensures their robustness against varying exposure in real-world applications.

In Table 3 we present the color reproduction performance using a *different camera* (and under original exposure setting). Here, the hyperspectral images are still reconstructed from the CIEXYZ color coordinates, but we re-generated the RGB images by the spectral sensitivities of an alternative camera - SONY IMX135. Remarkably, the physically plausible HSCNN-R^p and HSCNN-R^{pd} models provide superior performance compared to their physically non-plausible counterparts. Especially for HSCNN-R^{pd} whose worst-case MRAE (Table 2) is worse than HSCNN-R^d, and yet it out-performs HSCNN-R^d by 8% in the worst-case ΔE in Table 3.

In Figure 5 and 6 we show respectively the MRAE and ΔE error maps of the spectral recovery on one example image. We can clearly see in Figure 5 that under original exposure (top row), the physically plausible HSCNN-R^p and HSCNN-R^{pd} respectively improve HSCNN-R and HSCNN-R^d in the upper part of the image. It is also evident that HSCNN-R^d and HSCNN-R^{pd}, where intensity-scaling data augmentation is adopted, shows constant performance across different exposures, as opposed to the deterioration introduced by HSCNN-R and HSCNN-R^p. In Figure 6 we can see that non-physically plausible HSCNN-R and HSCNN-R^d are the least colorimetrically accurate models, while our physically plausible HSCNN-R^p and HSCNN-R^{pd} guaranteed absolute color accuracy for all spectra in the image. We remark that the horizontal artifacts in the error maps may come from the patch-wise reconstruction

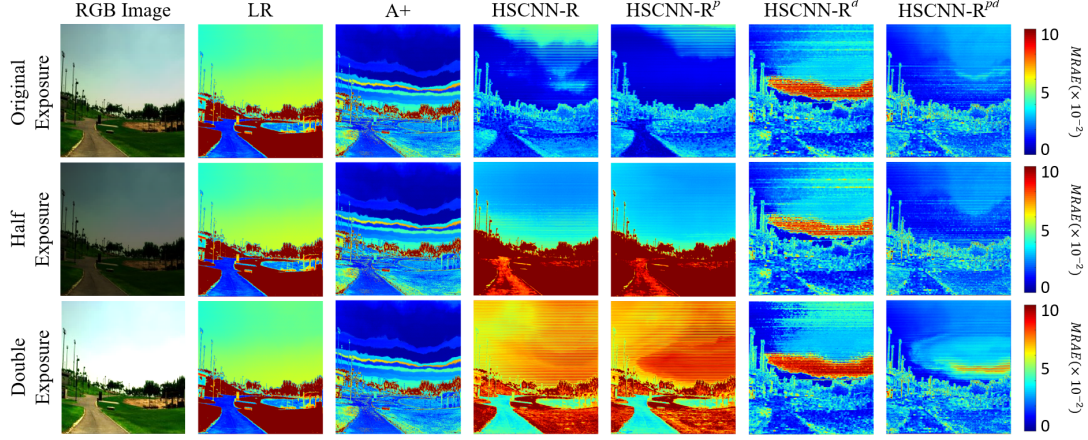


Figure 5: Visualization of spectral recovery errors by MRAE error maps. All models are tested under original exposure (top row), half exposure (middle row) and double exposure (bottom row).

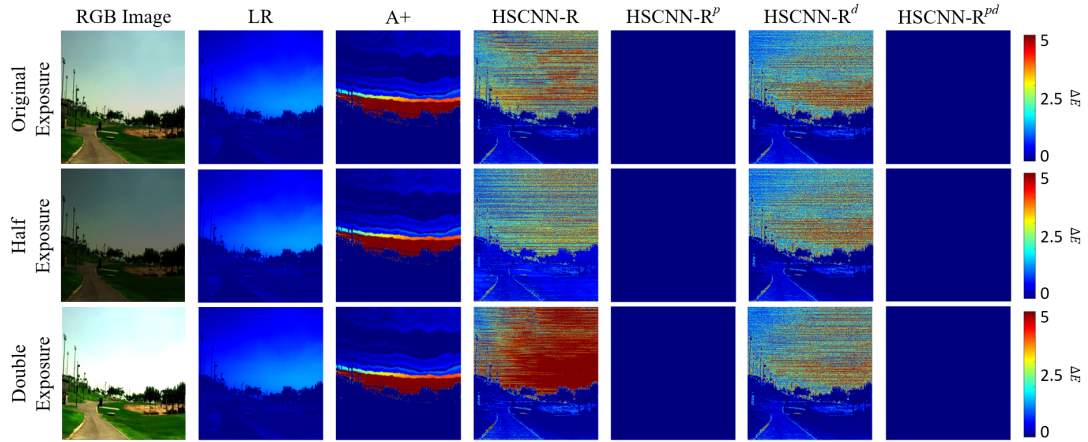


Figure 6: Visualization of color reproduction errors by ΔE error maps. All models are tested under original exposure (top row), half exposure (middle row) and double exposure (bottom row).

of the networks: we split the input image along the vertical spatial dimension for limited GPU memory, and we did find the gaps between the artifacts vary according to this setting.

6. Conclusion

Spectral reconstruction (SR) studies the mapping from RGB to hyperspectral images, which is regarded as a promising solution to low-cost, snapshot and high resolution hyperspectral camera. Recently, the leading models are based on Convolutional Neural Networks (CNN), providing remarkable spectral recovery performance. However, these models only aim to minimize the spectral recovery errors without ensuring the physical plausibility of the output spectra. Physical plausibility is defined as ensuring the recovered spectrum integrates (using the underlying camera sensors) to the same RGB as it is recovered from (the input RGB). Existing method, which do not have this property,

estimate RGBs which are significantly different from those found in the original image.

In this paper we developed a physically plausible framework for SR. Our insight is that all plausible spectra can be represented by a fixed camera-subspace projection spectrum defined by a linear combination of camera spectral sensitivities, and a *null-space* spectrum which do not contribute to the color formation. Relative to this insight, the spectral recovery problem sets out to reconstruct the null-space coefficients from the RGB (instead of the original RGB to radiance mapping), such that the physical plausibility of the predicted radiance is guaranteed. We also addressed the issue of exposure invariance in SR [29] by implementing an intensity-scaling data augmentation to ensure the model robustness against intensity variations. By imposing physical plausibility on CNN-based SR, our models out-perform the original network in spectral recovery as well as cross-device color reproduction.

References

- [1] Jonas Aeschbacher, Jiqing Wu, and Radu Timofte. In defense of shallow learned spectral reconstruction from rgb images. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 471–479, 2017. 1, 2, 6, 7
- [2] Farnaz Agahian, Seyed Ali Amirshahi, and Seyed Hossein Amirshahi. Reconstruction of reflectance spectra using weighted principal component analysis. *Color Research & Application*, 33(5):360–371, 2008. 2, 3
- [3] Aitor Alvarez-Gila, Joost Van De Weijer, and Estibaliz Garrote. Adversarial networks for spatial context-aware spectral image reconstruction from rgb. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 480–490, 2017. 1, 3
- [4] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *European Conference on Computer Vision*, pages 19–34. Springer, 2016. 1, 2, 6
- [5] Boaz Arad, Ohad Ben-Shahar, and Radu Timofte. Ntire 2018 challenge on spectral reconstruction from rgb images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 929–938, 2018. 1, 2, 3, 5
- [6] Henry Arguello and Gonzalo R Arce. Colored coded aperture design by concentration of measure in compressive spectral imaging. *IEEE Transactions on Image Processing*, 23(4):1896–1908, 2014. 1, 2
- [7] Simone Bianco. Reflectance spectra recovery from tristimulus values by adaptive estimation with metameric shape correction. *JOSA A*, 27(8):1868–1877, 2010. 2, 3
- [8] Bin Cao, Ningfang Liao, and Haobo Cheng. Spectral reflectance reconstruction from rgb images based on weighting smaller color difference group. *Color Research & Application*, 42(3):327–332, 2017. 1
- [9] Xun Cao, Hao Du, Xin Tong, Qionghai Dai, and Stephen Lin. A prism-mask system for multispectral video acquisition. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2423–2435, 2011. 1, 2
- [10] Ayan Chakrabarti and Todd Zickler. Statistics of real-world hyperspectral images. In *CVPR 2011*, pages 193–200. IEEE, 2011. 1
- [11] Chen Chen, Wei Li, Hongjun Su, and Kui Liu. Spectral-spatial classification of hyperspectral image based on kernel extreme learning machine. *Remote Sensing*, 6(6):5795–5814, 2014. 1
- [12] Yushi Chen, Xing Zhao, and Xiuping Jia. Spectral-spatial classification of hyperspectral data based on deep belief network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):2381–2392, 2015. 1
- [13] Claudia V Correa, Henry Arguello, and Gonzalo R Arce. Snapshot colored compressive spectral imager. *JOSA A*, 32(10):1754–1763, 2015. 1, 2
- [14] Commission Internationale de l’Eclairage. Cie proceedings (1964) vienna session, committee report e-1.4. 1. 1964. 6
- [15] Graham D Finlayson and Peter Morovic. Metamer sets. *JOSA A*, 22(5):810–819, 2005. 3
- [16] Ying Fu, Tao Zhang, Yinqiang Zheng, Debing Zhang, and Hua Huang. Joint camera spectral sensitivity selection and hyperspectral image recovery. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 788–804, 2018. 1, 3
- [17] Laura Galvis, Daniel Lau, Xu Ma, Henry Arguello, and Gonzalo R Arce. Coded aperture design in compressive spectral imaging based on side information. *Applied optics*, 56(22):6332–6340, 2017. 1, 2
- [18] Hans Garcia, Claudia V Correa, and Henry Arguello. Multi-resolution compressive spectral imaging reconstruction from single pixel measurements. *IEEE Transactions on Image Processing*, 27(12):6174–6184, 2018. 1, 2
- [19] Nahum Gat. Imaging spectroscopy using tunable filters: a review. In *Wavelet Applications VII*, volume 4056, pages 50–64. International Society for Optics and Photonics, 2000. 1
- [20] Pedram Ghamisi, Mauro Dalla Mura, and Jon Atli Benediktsson. A survey on spectral-spatial classification techniques based on attribute profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5):2335–2353, 2014. 1
- [21] Robert O Green, Michael L Eastwood, Charles M Sarture, Thomas G Chrien, Mikael Aronsson, Bruce J Chippendale, Jessica A Faust, Betina E Pavri, Christopher J Chovit, Manuel Solis, et al. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (aviris). *Remote sensing of environment*, 65(3):227–248, 1998. 1
- [22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 5
- [23] Ville Heikkinen, Reiner Lenz, Tuija Jetsu, Jussi Parkkinen, Markku Hauta-Kasari, and Timo Jääskeläinen. Evaluation and unification of some methods for estimating reflectance spectra from rgb images. *JOSA A*, 25(10):2444–2458, 2008. 1, 6, 7
- [24] James A Jablonski, Trevor J Bihl, and Kenneth W Bauer. Principal component reconstruction error for hyperspectral anomaly detection. *IEEE Geoscience and Remote Sensing Letters*, 12(8):1725–1729, 2015. 1
- [25] Sriharsha Koundinya, Himanshu Sharma, Manoj Sharma, Avinash Upadhyay, Raunak Manekar, Rudrabha Mukhopadhyay, Abhijit Karmakar, and Santanu Chaudhury. 2d-3d cnn based architectures for spectral reconstruction from rgb images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 844–851, 2018. 1, 3
- [26] Antony Lam and Imari Sato. Spectral modeling and relighting of reflective-fluorescent scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1452–1459, 2013. 1
- [27] Jinxing Liang and Xiaoxia Wan. Optimized method for spectral reflectance reconstruction from camera responses. *Optics Express*, 25(23):28273–28287, 2017. 1
- [28] Xing Lin, Yebin Liu, Jiamin Wu, and Qionghai Dai. Spatial-spectral encoded compressive hyperspectral imaging. *ACM Transactions on Graphics (TOG)*, 33(6):233, 2014. 1, 2

- [29] Yi-Tun Lin and Graham D Finlayson. Exposure invariance in spectral reconstruction from rgb images. In *Color and Imaging Conference*, volume 2019, pages 284–289. Society for Imaging Science and Technology, 2019. 2, 3, 6, 8
- [30] Peter Morovic and Graham D Finlayson. Metamer-set-based approach to estimating surface reflectance from camera rgb. *JOSA A*, 23(8):1814–1822, 2006. 2, 3
- [31] Rang MH Nguyen, Dilip K Prasad, and Michael S Brown. Training-based spectral reconstruction from a single rgb image. In *European Conference on Computer Vision*, pages 186–201. Springer, 2014. 1
- [32] Wug Seoung Oh, Michael S Brown, Marc Pollefeys, and Seon Joo Kim. Do it yourself hyperspectral imaging with everyday digital cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2461–2469, 2016. 1, 2
- [33] Alan R Robertson. The cie 1976 color-difference formulae. *Color Research & Application*, 2(1):7–11, 1977. 6
- [34] Hoover Rueda, Henry Arguello, and Gonzalo R Arce. Dmd-based implementation of patterned optical filter arrays for compressive spectral imaging. *JOSA A*, 32(1):80–89, 2015. 1, 2
- [35] Zhan Shi, Chang Chen, Zhiwei Xiong, Dong Liu, and Feng Wu. Hscnn+: Advanced cnn-based hyperspectral recovery from rgb images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 939–947, 2018. 1, 2, 3, 5, 6, 7
- [36] Tarek Stiebel, Simon Koppers, Philipp Seltsam, and Dorit Merhof. Reconstructing spectral images from rgb-images using a convolutional neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 948–953, 2018. 1, 3
- [37] Sabine Süsstrunk, Robert Buckley, and Steve Swen. Standard rgb color spaces. In *Color and Imaging Conference*, volume 1999, pages 127–134. Society for Imaging Science and Technology, 1999. 6
- [38] Tsuyoshi Takatani, Takahito Aoto, and Yasuhiro Mukaigawa. One-shot hyperspectral imaging using faced reflectors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4039–4047, 2017. 1, 2
- [39] Chao Tao, Hongbo Pan, Yansheng Li, and Zhengrou Zou. Unsupervised spectral-spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification. *IEEE Geoscience and remote sensing letters*, 12(12):2438–2442, 2015. 1
- [40] Miguel A Veganzones, Guillaume Tochon, Mauro Dalla-Mura, Antonio J Plaza, and Jocelyn Chanussot. Hyperspectral image segmentation using a new spectral unmixing-based binary partition tree representation. *IEEE Transactions on Image Processing*, 23(8):3574–3589, 2014. 1
- [41] Lizhi Wang, Zhiwei Xiong, Dahua Gao, Guangming Shi, and Feng Wu. Dual-camera design for coded aperture snapshot spectral imaging. *Applied optics*, 54(4):848–858, 2015. 2
- [42] Lizhi Wang, Zhiwei Xiong, Dahua Gao, Guangming Shi, Wenjun Zeng, and Feng Wu. High-speed hyperspectral video acquisition with a dual-camera architecture. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4942–4950, 2015. 1
- [43] Peng Xu, Haisong Xu, Changyu Diao, and Zhengnan Ye. Self-training-based spectral image reconstruction for art paintings with multispectral imaging. *Applied optics*, 56(30):8461–8470, 2017. 1
- [44] Yanbo Zhang, Xuanqin Mou, Ge Wang, and Hengyong Yu. Tensor-based dictionary learning for spectral ct reconstruction. *IEEE transactions on medical imaging*, 36(1):142–154, 2016. 1
- [45] Yi Zhang, Yan Xi, Qingsong Yang, Wenxiang Cong, Jiliu Zhou, and Ge Wang. Spectral ct reconstruction with image sparsity and spectral mean. *IEEE transactions on computational imaging*, 2(4):510–523, 2016. 1
- [46] Yonghui Zhao and Roy S Berns. Image-based spectral reflectance reconstruction using the matrix r method. *Color Research & Application*, 32(5):343–351, 2007. 2, 3
- [47] Yuanyuan Zhao, Hui Guo, Zhan Ma, Xun Cao, Tao Yue, and Xuemei Hu. Hyperspectral imaging with random printed mask. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10149–10157, 2019. 1, 2
- [48] Silvia Zuffi, Simone Santini, and Raimondo Schettini. From color sensor space to feasible reflectance spectra. *IEEE Transactions on Signal Processing*, 56(2):518–531, 2008. 2, 3