*Article*

# Physically Plausible Spectral Reconstruction †

## Yi-Tun Lin * and Graham D. Finlayson

School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, UK; G.Finlayson@uea.ac.uk

*   Correspondence: Yi-Tun.Lin@uea.ac.uk
†   This paper is an extension version of the conference paper: Lin, Y.T.; Finlayson, G.D. Physically Plausible Spectral Reconstruction from RGB Images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020.

check for updates

**Abstract:** Spectral reconstruction algorithms recover spectra from RGB sensor responses. Recent methods—with the very best algorithms using deep learning—can already solve this problem with good spectral accuracy. However, the recovered spectra are physically incorrect in that they do not induce the RGBs from which they are recovered. Moreover, if the exposure of the RGB image changes then the recovery performance often degrades significantly—i.e., most contemporary methods only work for a fixed exposure. In this paper, we develop a physically accurate recovery method: the spectra we recover provably induce the same RGBs. Key to our approach is the idea that the set of spectra that integrate to the same RGB can be expressed as the sum of a unique fundamental metamer (spanned by the camera's spectral sensitivities and linearly related to the RGB) and a linear combination of a vector space of metameric blacks (orthogonal to the spectral sensitivities). Physically plausible spectral recovery resorts to finding a spectrum that adheres to the fundamental metamer plus metameric black decomposition. To further ensure spectral recovery that is robust to changes in exposure, we incorporate exposure changes in the training stage of the developed method. In experiments we evaluate how well the methods recover spectra and predict the actual RGBs and RGBs under different viewing conditions (changing illuminations and/or cameras). The results show that our method generally improves the state-of-the-art spectral recovery (with more stabilized performance when exposure varies) and provides zero colorimetric error. Moreover, our method significantly improves the color fidelity under different viewing conditions, with up to a 60% reduction in some cases.

**Keywords:** spectral reconstruction; hyperspectral imaging; multispectral imaging

## 1. Introduction

Hyperspectral imaging devices are developed to capture scene radiance spectra at high spectral resolution. In the context of machine vision, hyperspectral imaging distinguishes different material properties at pixel level, which is commonly used in remote sensing [1–5], anomaly detection [6] and medical imaging [7,8]. Furthermore, the devices (sensors or displays), light sources and object surfaces are commonly characterized by spectral measurements [9–11]. Practical applications include scene relighting [12] and digital art archiving [13].

However, existing technologies by which high-resolution spectra are directly measured [14–16] often suffer from physical bulkiness, restricted mobility, poor light sensitivity and/or long capturing time. For fast and less costly alternatives, using compressed sensing, the spatial and spectral information is jointly encoded in the captured 2D images and decoded by specialized algorithms [17–23]. Most of these approaches use learning algorithms to solve for the complex and ill-posed decompression.

As one of the learning approaches, spectral reconstruction (SR) seeks to reconstruct hyperspectral information from spectral images of fewer spectral channels. While many works in the literature propose ways to increase the number of captured spectral channels—including using a multispectral

color filter array [24–26], a color filter wheel [27], multiple RGB cameras [28], multiple LED light sources [29,30], a stereo camera [31] and faced reflectors [32]—there are many works that focused on recovering hyperspectral information from the RGB images of a single camera [33–51]. Indeed, research has shown that in a natural scene a significant portion of spectral variation is captured by its color appearance [52], which makes it possible for learning approaches to infer much spectral information from the RGB data. Moreover, spectral recovery might be further improved when colors are part of the spatial context (the patterning of RGB colors), e.g., [47,48].

In this paper, we concern ourselves with the physical plausibility of SR from RGB images. Clearly, spectra and RGBs are physically related: an RGB camera integrates spectra with the spectral sensitivities of three different color sensors, resulting in the 3-value RGB colors, yet this physical fact is generally not employed by the best SR algorithms. Indeed, it is shown in [48] that the top deep learning approaches [50,53] recover spectral estimates that do not physically induce the same RGBs. This color fidelity issue is of practical importance. For instance, in some applications where color accuracy is concerned (e.g., art archiving), we clearly do not wish to use an algorithm that cannot regenerate the original colors.

Figure 1 illustrates our physical plausibility test for SR. While the ground-truth RGBs can be generated from the hyperspectral data (red curve), we test the color fidelity of the reconstructed spectra (blue dotted curve—estimated by each tested SR algorithm) when reintegrated with the same set of spectral sensitivities. In Figure 2, we give an example of the color errors introduced by polynomial regression SR [34] and one of the leading deep-learning models, HSCNN-R [46]. We can clearly see that HSCNN-R—despite claiming the state-of-the-art spectral accuracy [47]—performs much worse in color than the regression-based polynomial regression model. However, the very existence of the non-zero color errors indicates that "both methods are physically implausible".
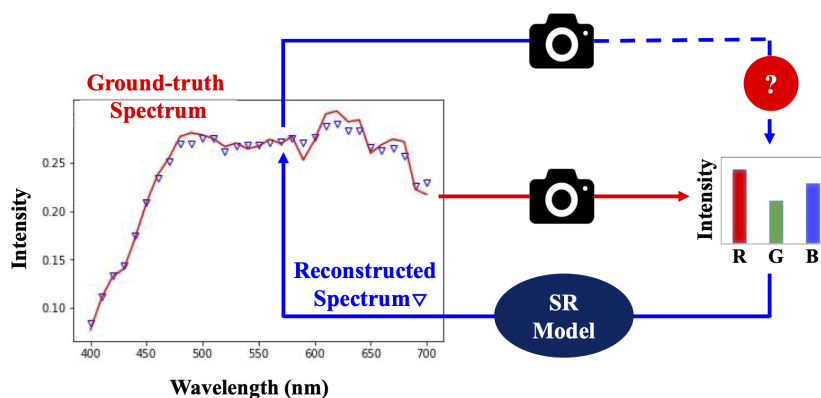


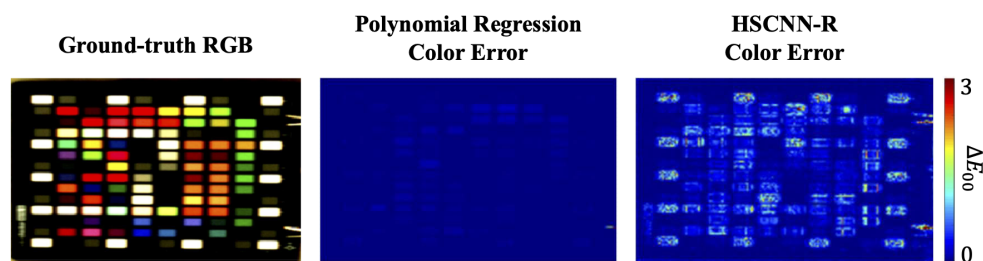**Figure 1.** Our physical plausibility (color fidelity) test for SR.



**Figure 2.** The color errors introduced by polynomial regression SR [34] (**left**) and HSCNN-R [46] (**right**). The color errors are measured in CIE $\Delta E$ 2000 ($\Delta E_{00}$) [54].

The second problem inherent in the current state of the art is exposure invariance. There are several factors that can result in exposure change, including there being more or less light; the same object being viewed in different parts of the scene (and being recorded as brighter or darker); or the device itself might globally or locally change its exposure setting (e.g., the EV value). In this case,

physically, the corresponding spectra are a scalar scaling apart [35]. That is, the magnitude of the physical spectrum changes but its shape remains the same. This said, we might expect that, for the two RGBs from the same but differently exposed physical object, the SR algorithms should recover two spectra with a scaling apart. Unfortunately, this is normally not the case. It is shown in [35] that the leading deep learning methods only work for a fixed exposure; i.e., the shape of the recovered spectra also changes with exposure. Moreover, the change in shape can be surprisingly large. The example in Figure 3 exhibits the extent of deterioration as we seek to reconstruct spectra from a 50% dimmer RGB image using the HSCNN-R model, in comparison to the primitive but exposure-invariant linear regression SR [33].
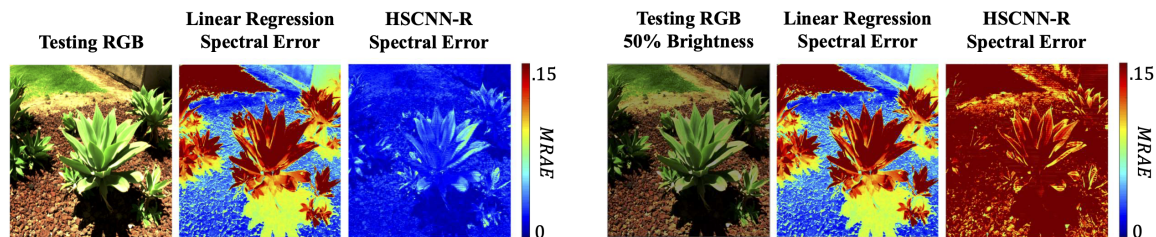


**Figure 3.** Spectral reconstruction under varying exposure by linear regression [33] and HSCNN-R [46]. The spectral errors are calculated in mean relative absolute error (MRAE) [47,48].

In this paper, we extend the existing SR algorithms to ensure that they return physically plausible spectra and that they continue to work well when the exposure changes. To solve the physical plausibility issue (to make the recovered spectra reintegrate to the same initial RGBs), our insight is to represent the output spectral space of the SR algorithm by the sum of the unique fundamental metamer of the given RGB and a non-unique metameric black. The fundamental metamer is in the space spanned by the spectral sensitivities of the camera and the metameric black is orthogonal to the spectral sensitivities. We reformulate SR estimation so that the reconstruction adheres to the fundamental metamer plus metameric black decomposition. In effect, we change the estimation problem from one of recovering the "most likely spectrum" to recovering the "most likely metameric black". Importantly, our method can be directly implemented in all methodologies, including deep learning.

Our solution to stabilizing the models under varying exposure is more pragmatic. We randomly modulate the exposure of the data (i.e., spectra and corresponding RGBs) when training the models. This simple data augmentation approach can make a dramatic difference in the accuracy of recovered spectra when the exposure changes.

We tested our methods on both the regression-based models (which includes the leading sparse coding and a shallow network solution) [33–37] and an exemplar leading deep neural network (DNN) model [46]. Experiments show that we can ensure the physical plausibility of the recovered spectra without negatively affecting recovery performance. Additionally, incorporating exposure variation in training leads to a significant uplift in recovery performance when exposure changes.

Finally, since we are recovering spectra which can be physically projected to the desired RGBs, this means we can change the illumination spectra and/or the camera's spectral sensitivities and get new RGBs for another viewing condition. We present experiments which demonstrate that a physically plausible spectral recovery results in better cross-viewing-condition color prediction (Figure 4 shows an example of the cross-illumination color fidelity result when using our physically plausible approach).

| Original RGB | RGB under CIE Illuminant A | RBFN Color Error | Physically Plausible RBFN Color Error |
| --- | --- | --- | --- |
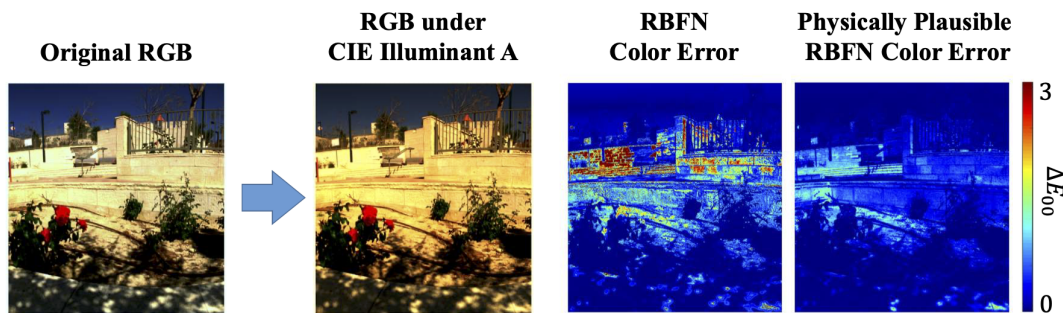
**Figure 4.** The scene relighting color fidelity of one example hyperspectral image recovered by the RBFN algorithm [36] and by our physically plausible modification of RBFN. The results are shown as the error maps of CIE $\Delta E$ 2000 color differences ($\Delta E_{00}$) [54].

## 2. Background

Spectral reconstruction (SR) has been intensively studied in both the color science and computer vision communities. Maloney and Wandell [38] represented reflectances using a 3-dimensional linear model. With respect to this model the spectra are related to RGBs by a simple $3 \times 3$ matrix transform. Additionally, RGBs for the same surfaces viewed under a pair of different lights must be a $3 \times 3$ linear transform apart. However, several subsequent studies showed that to adequately represent spectra, a higher than 3-dimensional linear model is required [55–59]. For higher-dimensional models the spectral reconstruction problem is ill-posed. Indeed, so long as the model has four or more degrees of freedom, we can always find (e.g., using "singlar value decomposition", referring to pp. 382–391 in [60]), one or more axes in the spectral space that are orthogonal to the spectrum-to-RGB projection. Since the values along these axes do not influence the resulting RGB values (i.e., the same RGB can be derived from different spectra with differences only in these axes), there must be a set of infinite spectra—called the metamers [61]—corresponding to one given RGB. In this paper we say the metamers belong to the plausible set of a given RGB.

Spectral recovery in the ill-posed case seeks to find the most likely spectrum for a given RGB. Recovery methods range from simple statistical approaches, including least-squares regression [33–35], Bayesian approaches [41,42] and iterative methods [43,44], to data clustering-based algorithms, such as the radial basis function network [36] and sparse coding [37,45], to the newest deep neural networks (DNN) [46–51].

A key seductive argument made about DNN approaches is that—perhaps at an object description level—a pixel is viewed in the context of an image, which helps determine the object and hence the shape of the spectrum. This idea clearly has some merit. After all, almost all cameras now automatically find faces in images, and the reflectance of skin has a characteristic spectral shape [62,63]. However, in experiments— as per [37] and the results presented in this paper—DNNs deliver only a modest performance increment compared to simpler methods.

Providing some motivation for the approach we develop in this paper, there were already studies that used the physics of image formation to improve spectral reconstruction. Agahian et al. [39] proposed to characterize each 3-dimensional reflectance dynamically with emphasis on the reflectance data of close-by colors. Zhao et al. [40] developed a matrix-R approach to colorimetrically post-facto correct the linear regression-based SR. Morovic and Finlayson [42] used metamer sets [61] as the physical constraints of Bayesian inference (and recovered spectra that are physically plausible). However, the performance of that method—developed over 10 years ago—is not competitive with today's leading methods. Bianco [43] proposed an iterative algorithm which includes color difference in the optimization function. Most recently in the NTIRE 2020 Spectral Recovery Challenge [48], the first-place winner Li et al. [50] included color difference in their learning cost function, and Joslyn Fubara et al. [51] designed an unsupervised learning approach based on the physics prior. However, even these last two methods still recover physically implausible spectra (spectra of wrong colors) [48].

*2.1. Image Formation*

The radiance spectrum is an intensity distribution across wavelengths, denoted as a spectral function $r(\lambda)$. Correspondingly, the R, G and B sensors are characterized by spectrally-varying sensitivities, denoted as $s_k(\lambda)$ with $k = R, G, B$. Based on this nomenclature, the RGB image formation is written as [64]:

$$\int_{\Omega} s_k(\lambda) r(\lambda) d\lambda = \rho_k \,, \tag{1}$$

where $\Omega$ refers to the visible range, which is set to $[400, 700]$ nanometers in this paper, and $\rho_k$ is the color value in the $k$ channel.

In reality, the ground-truth spectra are measured discretely—at $n$ evenly spaced wavelengths—by hyperspectral cameras. Hence, one can vectorize Equation (1):

$$\mathbf{S}^{\mathsf{T}} \underline{r} = \underline{\rho} \,, \tag{2}$$

where $\underline{r} \in \mathbb{R}^n$ is the discrete representation of spectra, $\mathbf{S} = (\underline{s}_R, \underline{s}_G, \underline{s}_B)$ is the $n \times 3$ spectral sensitivity matrix and $\underline{\rho} = (R, G, B)^{\mathsf{T}}$ represents the 3-value RGB color. This $\underline{\rho}$ vector refers to the linear color or raw camera response, which is commonly used as ground-truth RGBs for training the SR algorithms, e.g., in [36,37,45] and the "clean track" in the yearly NTIRE Spectral Recovery Challenge: [47,48]. Essentially, this simple model depicts the physical relationship between the RGBs and spectra.

*2.2. Spectral Reconstruction*

Spectral reconstruction algorithms map RGB colors to the spectral estimates. If we denote an SR algorithm as a mapping function $\Psi : \mathbb{R}^3 \mapsto \mathbb{R}^n$, SR can be simply expressed as:

$$\Psi(\underline{\rho}) \approx \underline{r} \,. \tag{3}$$

For DNN algorithms, a spectrum is recovered given the image context. Let us denote the set of proximal pixels to $\underline{\rho}$ as $Prox(\underline{\rho})$. A more general form of spectral reconstruction is then written as:

$$\Psi(\underline{\rho}; Prox(\underline{\rho})) \approx \underline{r} \,. \tag{4}$$

Equation (4) makes the dependence on context explicit; we will henceforth—to simplify the notation–denote SR algorithms using the notation $\Psi(\underline{\rho})$. In all cases the more general form of SR can be substituted without changing any argument made on our part.

2.2.1. Spectral Reconstruction by Regression

Many algorithms for spectral recovery can be formulated as regressions (linear or non-linear). The standard formulation of the regression-based SR is written as [33]:

$$\Psi(\underline{\rho}) = \mathbf{M}\varphi(\underline{\rho}) \approx \underline{r} \,, \tag{5}$$

where $\varphi : \mathbb{R}^3 \mapsto \mathbb{R}^p$ is a bespoke feature mapping for each algorithm, and $\mathbf{M}$ is called the regression matrix, which linearly maps the $p$-dimensional features to spectra. If spectra are represented by $n$ numbers, then $\mathbf{M}$ is an $n \times p$ matrix. Recasting Equation (5) as an optimization, the least-squares regression-based SR seeks the $\mathbf{M}$ that minimizes:

$$\min_{\mathbf{M}} \sum_i ||\mathbf{M}\varphi(\underline{\rho}_i) - \underline{r}_i||^2 \,, \tag{6}$$

where here—and throughout this paper—$|| \cdot ||^2$ denotes the sum-of-squares or the squared Frobenius norm. Here, $i$ indexes over the training set of paired RGBs and corresponding ground-truth spectra.

Let us consider the meaning of $\varphi(\underline{\rho})$. For linear regression [33], $\varphi(\underline{\rho}) = \underline{\rho}$, i.e., it is the identity transform. For polynomial regression [34] and root-polynomial regression [35], the $\varphi$ functions are respectively the polynomial and root-polynomial expansions of $\underline{\rho}$ up to a given order. Another non-linear model is the radial basis function network (RBFN) [36], where $\varphi$ corresponds to the set of outputs from the radial basis functions centered at a given number of representative RGBs. This model is often seen as a shallow neural network solution (consisting only one hidden layer, compared to significantly more for DNNs).

The leading sparse coding algorithm, A+, is also regression based (which is shown to deliver performance close to the DNN solutions) [37]. In sparse coding, we assume that all spectra can be represented as a convex combination of neighboring spectra, and the same combination coefficients will also derive their projected RGBs. In A+, a fixed set of anchor spectra and RGBs are determined by K-SVD clustering [65] (from which the clusters' centers are selected). Then, around each of the anchor spectra, a given (fixed) number of nearest neighbors are used to solve a linear map (i.e., $\varphi(\underline{\rho}) = \underline{\rho}$), which is the same as the linear regression model but with only the neighboring data. In reconstruction, the nearest anchor RGB of the input RGB is found, and the trained map of that specific anchor is applied to the input RGB to recover spectra.

All regression algorithms are tuned with regularization [33,66], which is a tool for tackling the overfitting problem [67] (the details of regularization theory fall outside the scope of this paper, but the interested reader is pointed to [33,66]).

### 2.2.2. An Exemplar DNN Algorithm

In Figure 5, we illustrate the HSCNN-R architecture. HSCNN-R [46] was the second-place winner in the 2018 NTIRE Spectral Recovery Challenge [47], and is based on a deep residual learning framework [68]. 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 $3 \times 3$. In the original setting, the network maps $50 \times 50 \times 3$ (height $\times$ width $\times$ spectral dimension) RGB image patches to the corresponding $50 \times 50 \times 31$ hyperspectral image patches (i.e., the ground-truth hyperspectral images used for training have 31 spectral channels). The reader who is interested in how the network is trained is pointed to [46].
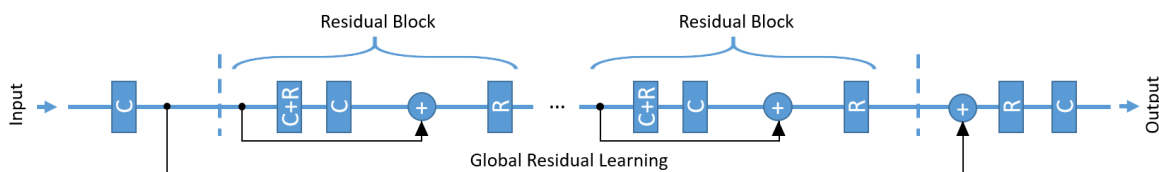


**Figure 5.** The HSCNN-R architecture [46]. "C" means $3 \times 3$ convolution and "R" refers to the ReLU activation.

## 3. Physically Plausible Spectral Reconstruction

Figure 6 contrasts physically plausible and implausible spectral recovery. On the left we show implausible spectral reconstruction which represents how many current algorithms work. An image RGB is mapped to a spectrum and this spectrum is almost always outside the plausible set. In this scenario, when the recovered spectrum is integrated with the camera sensors, the resultant RGB is different from the one we started with. On the right of Figure 6, we show physically plausible spectral reconstruction. Here the recovered spectrum is inside the plausible set and so integrates to the same RGB that we started with.

A spectral reconstruction algorithm is said to be physically plausible if and only if for all RGBs (viewed in all contexts), the recovered spectrum integrates to the same RGB:

$$\mathbf{S}^{\mathsf{T}}\Psi(\underline{\rho}) = \underline{\rho} \, . \tag{7}$$

Here we adopt the notation introduced in the background section: $\underline{\rho}$, $\Psi(\underline{\rho})$ and $\mathbf{S}$, respectively, denote an RGB, the recovered spectrum (an $n \times 1$ vector) and the spectral sensitivities of the camera (an $n \times 3$ matrix). We call Equation (7) the color fidelity constraint.
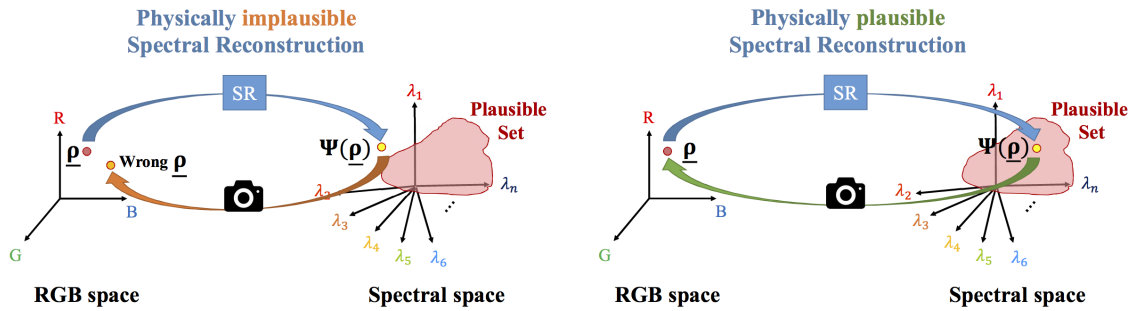


**Figure 6.** Physically implausible (**left**) and physically plausible spectral reconstruction (**right**).

### 3.1. The Plausible Set

Based on the color fidelity constraint, we define the plausible set as all spectra that integrate to the same RGB, which depends on a given RGB and the spectral sensitivities of the camera:

$$\mathcal{P}(\underline{\rho}; \mathbf{S}) = \left\{ \underline{r} \,\middle|\, \mathbf{S}^{\mathsf{T}} \underline{r} = \underline{\rho} \right\}. \tag{8}$$

Let us consider the plausible set in more detail. First we assume that all three sensors—the columns of $\mathbf{S}$—are linearly independent of one another (none can be written as a sum of the other two). In the language of vector spaces, $\mathbf{S}$, is thus a basis defining a 3-dimensional subspace of the $n$-dimensional spectral space. There is a complimentary $n \times (n-3)$ basis $\mathbf{B}$ whose columns are linearly independent and together span an $(n-3)$-dimensional subspace of $\mathbb{R}^n$, and such that $\mathbf{B}^{\mathsf{T}}\mathbf{S} = \mathbf{0}$, where $\mathbf{0}$ is an $(n-3) \times 3$ matrix of zeros signifying that $\mathbf{B}$ is orthogonal to $\mathbf{S}$. Combined, the $n \times n$ matrix $[\mathbf{S}\ \mathbf{B}]$ is a basis for the $n$-dimensional space of spectra.

Any given radiance spectrum $\underline{r}$ can be uniquely split into two components: one is the projected component on the basis $\mathbf{S}$, and the other part lies in $\mathbf{B}$:

$$\underline{r} = \mathbf{P^S}\underline{r} + \mathbf{P^B}\underline{r}, \tag{9}$$

where

$$\begin{cases} \mathbf{P^S} = \mathbf{S}(\mathbf{S}^{\mathsf{T}}\mathbf{S})^{-1}\mathbf{S}^{\mathsf{T}} \\ \mathbf{P^B} = \mathbf{I} - \mathbf{P^S} \end{cases} \tag{10}$$

are the $n \times n$ projector matrices of $\mathbf{S}$ and $\mathbf{B}$, respectively ($\mathbf{I}$ is the $n \times n$ identity matrix). The significance of "projection" is that $\mathbf{P^S}\underline{r}$ and $\mathbf{P^B}\underline{r}$ are respectively, over all other vectors in the span of $\mathbf{S}$ and $\mathbf{B}$, closest to the original radiance $\underline{r}$ in a least-squares sense (pp. 219–232; [60]).

Projector matrices have the natural property that their rank is equal to the dimension of the subspace on which they project. Thus, from this projector $\mathbf{P^B}$ it follows that we can solve for basis $\mathbf{B}$. From elementary linear algebra, we know that $\mathbf{P^S}$ has rank 3 (since $\mathbf{S}$ is 3-dimensional) and $\mathbf{P^B}$ has the complementary rank $n-3$ (pp. 135–149; [60]). The basis $\mathbf{B}$ is then the $n-3$ linearly independent columns of $\mathbf{P^B}$, which can be found using, e.g., the Gram–Schmidt orthogonalization procedure [69].

In Equation (9), the spectral components $\mathbf{P^S}\underline{r}$ and $\mathbf{P^B}\underline{r}$ are respectively called the "fundamental metamer" and "metameric black" [70]; henceforth, we denote them as $\underline{r}^f$ and $\underline{r}^b$, respectively. Returning to the definition of a plausible set, Equation (8), the color fidelity constraint $\mathbf{S}^{\mathsf{T}}\underline{r} = \underline{\rho}$ ensures that all spectra $\underline{r}$ in $\mathcal{P}(\underline{\rho}; \mathbf{S})$ have the same fundamental metamer $\underline{r}^f$. Indeed, since

$$\underline{r}^f = \mathbf{P^S}\underline{r} = \mathbf{S}(\mathbf{S^T S})^{-1}(\mathbf{S^T}\underline{r}) \,. \tag{11}$$

It follows:

$$\underline{r}^f = \mathbf{S}(\mathbf{S^T S})^{-1}\underline{\rho} \,. \tag{12}$$

In other words, $\underline{r}^f$ can be derived directly from the RGB vector $\underline{\rho}$; therefore, **no** estimation is needed. What is also indicated in Equation (12) is that an RGB has a unique fundamental metamer and vice versa.

Now let us consider the other part of the spectra, the metameric black component $\underline{r}^b$. $\underline{r}^b$ lies in the basis **B** which is orthogonal to **S**, and when integrated with the spectral sensitivities, induces a zero color response, i.e., $\mathbf{S^T}\underline{r}^b = \underline{\mathbf{0}}$ (here, $\underline{\mathbf{0}}$ is a 3-vector of zeros). Given only the input RGB, it follows that all metameric blacks which lies in **B** are possible solutions (since it is not constrained by the color fidelity constraint). We can represent the set of all metameric blacks as:

$$\underline{r}^b = \mathbf{B}\underline{b} \,, \tag{13}$$

where $\underline{b}$ is an $(n-3) \times 1$ coefficient vector.

Based on the derivations above, we write $\mathcal{P}(\underline{\rho}; \mathbf{S})$ in the form of $[\underline{r}^f + \underline{r}^b]$:

$$\mathcal{P}(\underline{\rho}; \mathbf{S}) = \left\{ \underline{r}^f + \mathbf{B}\underline{b} \,\middle|\, \underline{b} \in \mathbb{R}^{n-3} \right\} \,. \tag{14}$$

*3.2. Estimating Physically Plausible Spectra from RGBs*

The aim of spectral reconstruction is to recover a radiance spectrum $\Psi(\underline{\rho})$ from an RGB $\underline{\rho}$ that is as close to the correct answer $\underline{r}$ (the ground-truth) as possible. All algorithms $\Psi$ have tunable parameters that seek to minimize the recovery error: the distance between the recovered spectrum and the ground-truth radiance.

The error between one spectral estimate $\Psi(\underline{\rho})$ and the correct ground-truth $\underline{r}$ is written as:

$$recovery\ error = ||\ \underline{r} - \Psi(\underline{\rho})\ || \,. \tag{15}$$

Remember that we are representing a spectrum as a sum of the spectrum's fundamental metamer and a metameric black: $\underline{r} = \underline{r}^f + \underline{r}^b$. At the core of our physically plausible SR approach is to derive (instead of estimate) the exact $\underline{r}^f$ from the RGB i.e., using Equation (12). It follows that the recovery error only depends on how well the $\underline{r}^b$ part of the spectrum is recovered.

Let us denote an algorithm which recovers the metameric black part of the spectrum as $\Psi^b$. Given a set of training spectra and RGBs, $\underline{r}_i$ and $\underline{\rho}_i$ ($i$ indexes an individual data pair), we seek to minimize:

$$\min_{\Psi^b} \sum_i ||\ \underline{r}_i^b - \Psi^b(\underline{\rho}_i)\ || \,, \tag{16}$$

where the ground-truth $\underline{r}_i^b$ can be calculated by the projector matrix $\underline{r}_i^b = \mathbf{P^B}\underline{r}_i$. To ensure that $\Psi^b$ must recover estimates that lie in basis **B**, we restrict the estimated metameric black to comply with the linear combination form $\Psi^b(\underline{\rho}_i) = \mathbf{B}\underline{b}_i$ (Equation (13)). Equation (16) can then be, equivalently, written as:

$$\min_{\underline{b}_i} \sum_i ||\ \underline{r}_i^b - \mathbf{B}\underline{b}_i\ || \,. \tag{17}$$

Counterintuitively, Equations (16) and (17) teach that the physically plausible spectral recovery involves estimating the part of radiance that a camera cannot see.

In Figure 7, we compare our physically plausible method with the conventional approach (which does not recover physically plausible spectra). In the standard approach (top of the figure) the training/estimation scheme directly maps the RGBs to spectra. Here, $\underline{r}$ may not integrate to $\underline{\rho}$ (the RGB from

which it was recovered). In the physically plausible approach, the reconstruction is split into two streams. In the first stream the fundamental metamer—which is the only part that contributes to the RGB formation—is calculated directly from the input RGB. Then, the second stream seeks to find the best estimate for the metameric black. By construction the recovered spectrum (the sum of the fundamental metamer and the metameric black) must integrate to the same RGB.
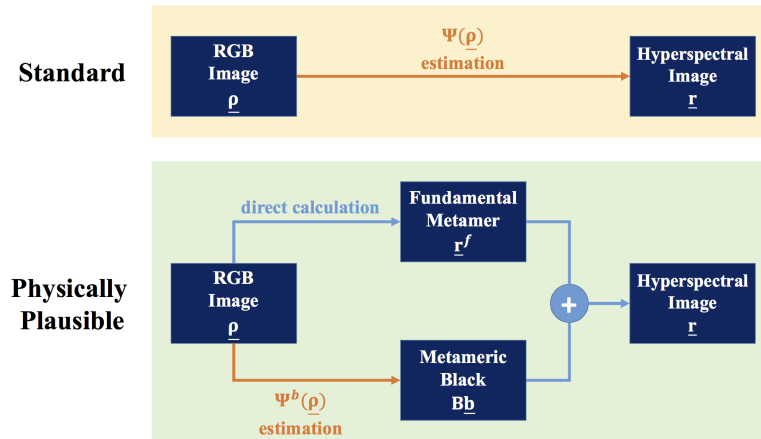


**Figure 7.** The standard SR scheme (**top**) versus our physically plausible SR scheme (**bottom**).

### 3.2.1. Physically Plausible Regression-Based Models

In the case of regression, we return to the formulation of regression-based SR in Equations (5) and (6). We now in turn solve for the map from the RGB—or more generally from its feature expansion $\varphi(\underline{\rho})$—to the metameric black. With $\Psi^b(\underline{\rho}) = \mathbf{M}^b \varphi(\underline{\rho})$, we minimize:

$$\min_{\mathbf{M}^b} \ ||\mathbf{M}^b \varphi(\underline{\rho}) - \underline{r}^b|| \, . \tag{18}$$

Further, according to Equation (17) we have to constrain $\Psi^b(\underline{\rho})$ such that it only recovers metameric black. It follows that we can decompose $\mathbf{M}^b$ into:

$$\mathbf{M}^b = \mathbf{B}\mathfrak{M} \, , \tag{19}$$

where $\mathfrak{M}$ is an $(n-3) \times p$ matrix (remember $\mathbf{B}$ is the $n \times (n-3)$ orthogonal basis spanning the set of metameric blacks). Then, we can rewrite Equation (18) as:

$$\min_{\mathfrak{M}} \ ||\mathbf{B}\mathfrak{M}\varphi(\underline{\rho}) - \underline{r}^b|| \, . \tag{20}$$

Since $\mathbf{B}$ is an orthogonal matrix, we know that $\mathbf{B}^\mathsf{T}\mathbf{B} = \mathbf{I}$ and $||\mathbf{A}|| = ||\mathbf{B}^\mathsf{T}\mathbf{A}||$ for any arbitrary matrix $\mathbf{A}$. Hence,

$$||\mathbf{B}\mathfrak{M}\varphi(\underline{\rho}) - \underline{r}^b|| = ||\mathbf{B}^\mathsf{T}\left(\mathbf{B}\mathfrak{M}\varphi(\underline{\rho}) - \underline{r}^b\right)|| = ||\mathfrak{M}\varphi(\underline{\rho}) - \mathbf{B}^\mathsf{T}\underline{r}^b|| \, . \tag{21}$$

Finally, the physically plausible spectral recovery as a regression problem sets out to find the $\mathfrak{M}$ that minimizes this norm.

### 3.2.2. Physically Plausible Deep Neural Networks

Likewise for the DNNs, we can replace the regression mapping $\mathfrak{M}\varphi(\underline{\rho})$ in the above discussion by a DNN model such that

$$DNN(\underline{\rho}) \approx \mathbf{B}^\mathsf{T}\underline{r}^b \, , \tag{22}$$

that is, to modify the original DNN algorithm to estimate $\mathbf{B}^\mathsf{T}\underline{r}^b$ instead of spectra. Following the same logic in Equations (18) and (19) we have

$$\Psi^b(\underline{\rho}) = \mathbf{B}DNN(\underline{\rho}) \approx \underline{r}^b \,, \tag{23}$$

which recovers the metameric black.

However, for many DNN models (including the one considered in this paper), the output layer is constricted to return positive values (since its original usage is to recover all positive spectra), and yet $\mathbf{B}^\mathsf{T}\underline{r}^b$ **must** have some values that are negative. For this reason we investigated the range of $\mathbf{B}^\mathsf{T}\underline{r}^b$ in our testing dataset. Assume that the maximum value in the original hyperspectral images is $v_{max}$ (e.g., in our case the images are 12-bit, so $v_{max} = 2^{12} - 1 = 4095$), empirically, we found that $\mathbf{B}^\mathsf{T}\underline{r}^b$ are bounded by $[-v_{max}, v_{max}]$. Without changing the original model, we set the DNN algorithm to recover instead the offset values $\frac{1}{2v_{max}}(\mathbf{B}^\mathsf{T}\underline{r}^b + v_{max})$, which is then corrected back to $\mathbf{B}^\mathsf{T}\underline{r}^b$ after reconstruction.

### 3.3. Intensity-Scaling Data Augmentation

The same object viewed in different parts of the same scene or viewed under different intensities of light and/or different camera exposure settings can appear brighter or darker. The brightness change due to there being more or less light is called a change of exposure. Let us model exposure change by a scaling factor $k$ multiplying the radiance spectrum: $\underline{r} \to k\underline{r}$. Clearly, according to Equation (2) the corresponding RGB is scaled by the same factor:

$$\mathbf{S}^\mathsf{T}(k\underline{r}) = k(\mathbf{S}^\mathsf{T}\underline{r}) = k\underline{\rho} \,. \tag{24}$$

Unfortunately, as shown in [35,48], in most existing algorithms—and all of the leading DNN-based spectral reconstruction approaches,

$$\Psi(k\underline{\rho}) \neq k\Psi(\underline{\rho}) \,. \tag{25}$$

That is, the shape of the recovered spectrum changes as the exposure changes (not just its magnitude as prescribed by the physics).

Our solution to this problem is pragmatic. Given a pair of RGB and spectrum for training, $(\underline{\rho}_i, \underline{r}_i)$, we multiply them with a random scaling factor $k$, such that $(k\underline{\rho}_i, k\underline{r}_i)$ is used as a replacement of the original pair in training. We, of course, must use many different scaling factors (for different training pairs). We argue that the random distribution of $k$ should follow a uniform distribution on a log scale:

$$\log_\beta k \sim Uniform(-1, 1) \,, \tag{26}$$

where $\beta$ controls the range of the distribution, e.g., for $\beta = 10$, the distribution is bounded by $[\frac{1}{10}, 10]$.

The justification of using this random distribution is demonstrated in Figure 8. Let us compare the proposed distribution ($\beta = 10$; right panel) with the straightforward uniform distribution between $[0, 10]$ (left panel). From both distributions we drew 5000 random numbers and show the histogram with 100 bins on the log scale (linear to the "geometric progression" of the exposure modes of a usual imaging device). Evidently, the straightforward uniform distribution generates exponentially more bright scaling factors than the dim ones, while our proposed distribution provide equal chances for bright and dim factors to be chosen.
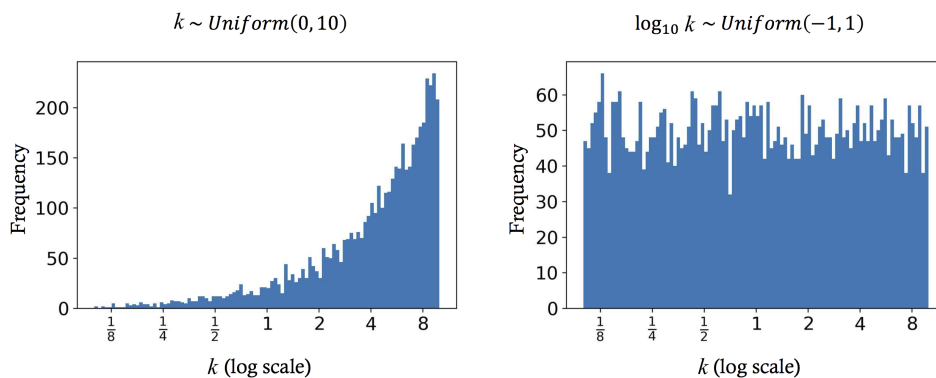
**Figure 8.** The comparison between drawing the scaling factor $k$ from the straightforward uniform distribution (**left**) and from our proposed distribution (**right**).

For the regression-based models, we simply apply the random scaling factors to all individual pairs of spectrum and RGB prior to the training. Now for the DNN model we implement data augmentation slightly differently. By virtue of the iterative training process of DNN, instead of generating all augmented data before training, we apply the random scaling factors in real-time—different image patches and the same patches in different training epochs are applied with different scaling factors. This setting in effect provides far more chances for the model to see the introduced exposure variation.

Another implementation detail is that, once we allow different exposure scaling factors, we essentially stretch the range of the output space of the physically plausible DNN to $[-\beta v_{\max}, \beta v_{\max}]$ (which was originally $[-v_{\max}, v_{\max}]$; see the discussion in Section 3.2.2). Hence, in our case that the considered DNN model only allow positive output values, we need to apply an offset following: $\frac{1}{2\beta v_{\max}}(\mathbf{B}^{\mathsf{T}}\underline{r}^b + \beta v_{\max})$.

## 4. Experiments

In Table 1, we list six exemplar algorithms we tested (see table for algorithm names and abbreviations), which comprise five regression-based algorithms and one exemplar DNN approach (these algorithms are reviewed in Section 2). According to [35], LR, RPR and A+ are exposure-invariant, which means they perform equally well for a varying exposure as they do for fixed exposure conditions. This means we did not need an additional data augmentation process (detailed in Section 3.3) to ensure their generalizability to different exposure conditions.

**Table 1.** Exemplar spectral recovery algorithms.

| Exposure-Invariant Models | Non-Exposure-Invariant Models |
|---|---|
| Linear Regression (LR) [33] | Radial Basis Function Network (RBFN) [36] |
| Root-Polynomial Regression (RPR) [35] | Polynomial Regression (PR) [34] |
| A+ Sparse Coding (A+) [37] | HSCNN-R Deep Neural Network (HSCNN-R) [46] |

We will compare spectral recovery for all considered algorithms where a standard training methodology is used (color fidelity is not enforced) and with our new physically plausible SR formulation (that guarantees color fidelity). All implemented codes are provided as the Supplementary Materials.

### 4.1. Image Dataset

In this paper, we used the ICVL database [45], which consists of 201 hyperspectral images of both indoor and outdoor scenes. The spatial dimensions of the scenes are $1300 \times 1392$, and the spectra were measured from 400 to 700 nanometers (nm) with 10-nm intervals, resulting in 31 spectral channels. All values of the images are encoded in 12 bits. Some example scenes from the database are shown in Figure 9.

**Figure 9.** Example scenes from the ICVL hyperspectral image database [45].

Then, we simulated the ground-truth RGB images following the linear RGB image formation (Equation (2)), with the CIE 1964 color matching functions (CMF) [71] as the camera's spectral sensitivities. This choice of using CMF is so we follow the standard methodology of the yearly NTIRE competition on spectral recovery [47,48]. We also remark that the CIE 1964 CMF is a revised version of the CIE 1931 CMF [72], which addressed the influence within the $10°$ viewing angle of the standard observer, as opposed to the $2°$ viewing angle considered in CIE 1931 CMF.

*4.2. Cross Validation*

In this paper we use a 4-trial cross validation setting. We randomly allocate all images into four groups—conceptually, group $A$, $B$, $C$ and $D$. We designed a compact 4-trial setting:

- Trial 1—Train set: $A + B$, Validation set: $C$, Test set: $D$,
- Trial 2—Train set: $A + B$, Validation set: $D$, Test set: $C$,
- Trial 3—Train set: $C + D$, Validation set: $A$, Test set: $B$,
- Trial 4—Train set: $C + D$, Validation set: $B$, Test set: $A$.

In each trial, two groups of images were used for training, one group for validation and one group for testing. Note that for regression-based methods the model validation refers to selecting proper regularization parameters to fit the validation set images (we point the interested readers to [34,35,37] for the implementation details), whereas for the deep learning model we used the validation set data to determine the terminating epoch in the iterative training process. The cross-validated error statistics are then the averaged testing performance over the four trials.

*4.3. Evaluation Metrics*

4.3.1. Spectral Difference

In this paper, we use the following metrics to measure the spectral error. Given a pair of ground-truth spectrum $\underline{r}$ and reconstructed spectrum $\underline{\hat{r}}$:

- Mean relative absolute error:

$$\text{MRAE } (\%) = 100 \times \frac{1}{n} \left\| \frac{\underline{r} - \underline{\hat{r}}}{\underline{r}} \right\|_1, \tag{27}$$

where $n$ is the number of spectral channels (in our case $n = 31$), the division is element-wise and the L1 norm is calculated. Essentially, this MRAE metric measures the averaged percentage absolute deviation over all spectral channels. This metric is regarded as the standard metric to rank and evaluate SR algorithms in the recent benchmark [47,48].

- Goodness of fit coefficient:

$$\text{GFC} = \frac{\underline{r}}{||\underline{r}||} \cdot \frac{\underline{\hat{r}}}{||\underline{\hat{r}}||}, \tag{28}$$

where the inner product of the normalized spectra is calculated. According to [56], acceptable reconstruction performance refers to $\text{GFC} \geq 0.99$, $\text{GFC} \geq 0.999$ is regarded as very good performance and $\text{GFC} \geq 0.9999$ means nearly exact reconstruction.

- Root mean square error:

$$\text{RMSE} = \sqrt{\frac{1}{n} ||\underline{r} - \hat{\underline{r}}||_2^2}, \tag{29}$$

where $n$ is the number of spectral channels. Note that RMSE is scale dependent, that is, the overall brightness level in which the compared spectra reside will reflect on the scale of RMSE. Thus, bear in mind that the images in the ICVL database [45] use 12-bit encoding (i.e., all values are bounded by [0, 4095]) when interpreting the presented results.

- Peak signal-to-noise ratio:

$$\text{PSNR} = 20 \times \log_{10}\left(\frac{v_{\max}}{\text{RMSE}}\right), \tag{30}$$

where $v_{\max} = 2^{12} - 1 = 4095$ is the maximum possible value for 12-bit images. Similarly to RMSE, PSNR is scale dependent.

### 4.3.2. Color Difference

In addition to the spectral error measures, we pay special attention to the models' colorimetric performances. We used the CIE $\Delta E$ 2000 color difference formula ($\Delta E_{00}$) [54] to measure the difference between the ground-truth and reconstructed colors. The implementation of $\Delta E_{00}$ is rather complex: we refer the readers to [54] for details. Practically, a $\Delta E_{00}$ equaling 1 between two color stimuli correlates with a color difference that is just noticeable to a human observer.

Note that the $\Delta E_{00}$ is defined upon the CIELAB [73] color coordinates—one of the standard (device independent) color spaces [74]. From our ground-truth color space, CIEXYZ, there exists direct transformation to CIELAB given a ground-truth white point color (i.e., the illumination color) [74]. In our experiments, we obtained this white point information by hand-crafting the "brightest near-achromatic spectrum" from each ground-truth hyperspectral image and then integrating this white-surface radiance spectrum with the CIE 1964 XYZ color matching functions.

## 5. Results

### *5.1. Effectiveness of Data Augmentation*

We can only create the augmented data within a given "range" of exposure variation (it is not feasible to include all possible exposure changes, since the physical brightness level is unbounded). Returning to the random distribution used in our data augmentation approach (Section 3.3; Equation (26)), the range of the random scaling is bounded by $[\frac{1}{\beta}, \beta]$. Clearly, if we choose a larger $\beta$, the trained models will have a wider range of generalizability in terms of exposure change. Note that we simulated the scaled images in floating point numbers (no darkened pixels were digitized to 0), and we allowed values exceeding the camera's original dynamic range (brightened pixels were not clipped at $v_{\max}$). Under this setting, assuming that there is no under- or over-exposed image in the database, there will also not be any of such images among our brightened/darkened images.

In Table 2 and Figure 10, we show how the value of $\beta$ influences the models' performance and generalizability. We trained the models with $\beta = 1$ (i.e., the original training regime), $2.5, 5, 7.5$ and 10 (for the deep learning based HSCNN-R we only trained for $\beta = 1, 5$ and 10). Under this training arrangement, we tested the models with all testing images scaled by factors of 1 (the original images), 0.5 (half exposure) and 2 (double exposure), denoted as "**1x**", "**0.5x**" and "**2x**", respectively. The performances of the exposure-invariant models are also given on top of each result table as baselines for comparison, and plotted as dotted lines in the figures.

**Table 2.** The dependency of spectral and color accuracy on the $\beta$ factor used for data augmentation. All models were tested under original (**1x**), half (**0.5x**) and double exposure settings (**2x**). The MRAE, GFC and $\Delta E_{00}$ errors are calculated per pixel, and the mean results (over all pixels and images) are shown.

| | **Mean MRAE (%) (Spectral Error)** Baseline Performance: LR = 6.24, RPR = 4.69, A+ = 3.87 | | | | | | | | | | | | | | |
| | **$\beta = 1$** | | | **$\beta = 2.5$** | | | **$\beta = 5$** | | | **$\beta = 7.5$** | | | **$\beta = 10$** | | |
| | 1x | 0.5x | 2x | 1x | 0.5x | 2x | 1x | 0.5x | 2x | 1x | 0.5x | 2x | 1x | 0.5x | 2x |
| RBFN | 2.06 | 18.58 | 8.74 | 4.20 | 5.67 | 4.33 | 6.19 | 6.02 | 5.30 | 6.82 | 7.05 | 6.40 | 7.37 | 7.75 | 6.98 |
| PR | 1.95 | 9.60 | 13.04 | 3.50 | 5.01 | 3.57 | 4.72 | 5.40 | 3.80 | 5.25 | 5.72 | 4.45 | 5.74 | 6.03 | 5.13 |
| HSCNN-R | 1.73 | 16.41 | 6.39 | - | - | - | 2.91 | 2.92 | 2.81 | - | - | - | 2.96 | 2.96 | 2.95 |
| | **Mean GFC (Spectral Error)** Baseline Performance: LR = 0.9966, RPR = 0.9979, A+ = 0.9983 | | | | | | | | | | | | | | |
| | **$\beta = 1$** | | | **$\beta = 2.5$** | | | **$\beta = 5$** | | | **$\beta = 7.5$** | | | **$\beta = 10$** | | |
| | 1x | 0.5x | 2x | 1x | 0.5x | 2x | 1x | 0.5x | 2x | 1x | 0.5x | 2x | 1x | 0.5x | 2x |
| RBFN | 0.9994 | 0.9802 | 0.9959 | 0.9986 | 0.9981 | 0.9983 | 0.9977 | 0.9979 | 0.9981 | 0.9974 | 0.9971 | 0.9977 | 0.9973 | 0.9968 | 0.9976 |
| PR | 0.9994 | 0.9949 | 0.9900 | 0.9989 | 0.9984 | 0.9986 | 0.9984 | 0.9981 | 0.9987 | 0.9981 | 0.9979 | 0.9985 | 0.9979 | 0.9977 | 0.9982 |
| HSCNN-R | 0.9995 | 0.9889 | 0.9972 | - | - | - | 0.9992 | 0.9991 | 0.9992 | - | - | - | 0.9991 | 0.9991 | 0.9991 |
| | **Mean $\Delta E_{00}$ (Color Error)** Baseline Performance: LR = 0.05, RPR = 0.14, A+ = 0.06 | | | | | | | | | | | | | | |
| | **$\beta = 1$** | | | **$\beta = 2.5$** | | | **$\beta = 5$** | | | **$\beta = 7.5$** | | | **$\beta = 10$** | | |
| | 1x | 0.5x | 2x | 1x | 0.5x | 2x | 1x | 0.5x | 2x | 1x | 0.5x | 2x | 1x | 0.5x | 2x |
| RBFN | 0.32 | 0.68 | 1.97 | 0.15 | 0.17 | 0.37 | 0.51 | 0.61 | 0.76 | 0.81 | 1.03 | 1.00 | 0.95 | 1.24 | 1.20 |
| PR | 0.01 | 0.02 | 0.15 | 0.01 | 0.03 | 0.01 | 0.05 | 0.06 | 0.04 | 0.06 | 0.09 | 0.04 | 0.09 | 0.11 | 0.05 |
| HSCNN-R | 0.10 | 0.36 | 0.16 | - | - | - | 0.17 | 0.18 | 0.18 | - | - | - | 0.15 | 0.15 | 0.15 |

Notice that here we only present the results in MRAE, GFC and $\Delta E_{00}$. In [37,45,47], it is argued that RMSE tends to penalize bright pixels more than the dim pixels. We remark that this originates from that RMSE, as we mentioned in Section 4.3.1, is scale dependent. Indeed, if we scale both $\underline{r}$ and $\hat{\underline{r}}$ by 2, the RMSE will also be doubled. It is therefore not suitable to use RMSE for comparing the reconstruction results in different exposure scales. The same argument also applies to PSNR.
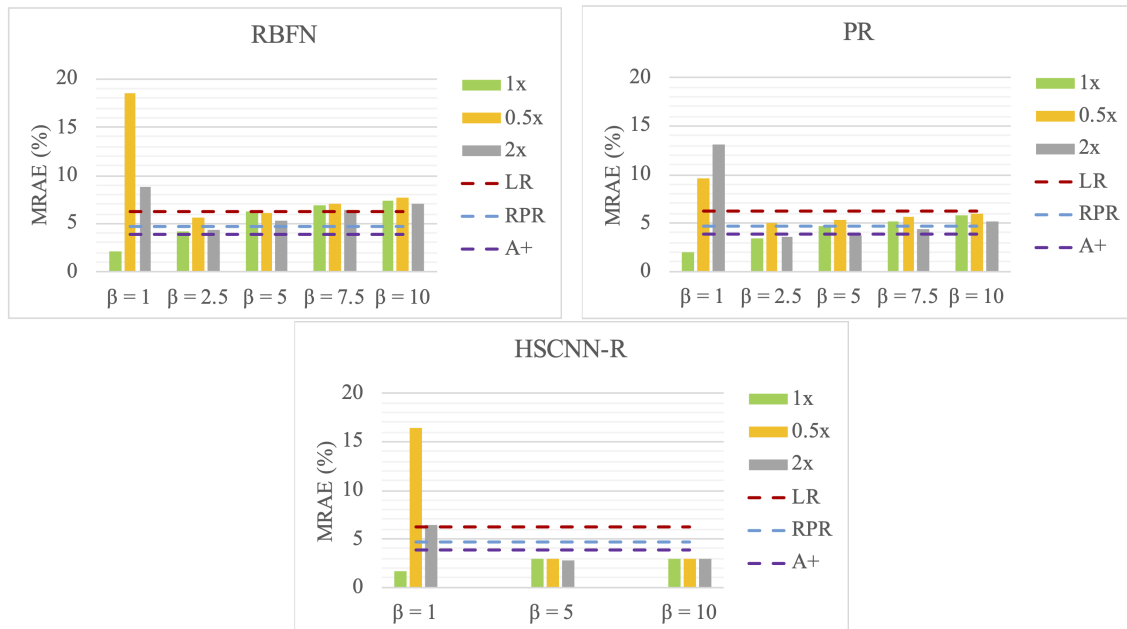


**Figure 10.** Visualizing the performance and generalizability (in mean MRAE) with respect to different $\beta$ factors chosen.

First, we see that RBFN, PR and HSCNN-R trained under the original training regime ($\beta = 1$) deliver superior performance in spectral accuracy compared to those exposure-invariant baseline models when tested with the original testing images (**1x**), but deteriorate under other exposure conditions (much worse than the simplest LR model). This result implies that the images (used for training and testing) in the ICVL database [45] were captured under very similar exposure conditions. Granted, when capturing images we often adjust the exposure settings of the device to fit the dynamic range of the scene, so as to avoid over- and under-exposed images, but in doing so we are in effect training the models only to work on those "nicely captured" scenes—say, if a sudden strong light occurs in the scene (e.g., the cars' headlights) and the rest of the scene darkens for fitting the new dynamic range, the models may not work even for the parts of the image that are not over-exposed.

Through our data augmentation, $\beta$ of higher values stabilizes the models' performances in both spectral and color accuracy—though at the cost of worse overall spectral accuracy. Indeed, the performances of the data-augmented RBFN and PR became worse than the baseline models in some cases, while the data-augmented HSCNN-R still held some advantage over the baseline models.

For HSCNN-R, the selection of $\beta$ does not have much influence on the models' performance. In contrast, for both RBFN and PR, large $\beta$ values lead to performance degradation, to the point that the performance can be much worse than the baseline models. As a result, in the forgoing discussion, we select $\beta = 2.5$ for RBFN and PR, and $\beta = 10$ for HSCNN-R.

Notice that the HSCNN-R with data augmentation clearly delivers good generalizability for the three testing exposures (i.e., small differences between the **1x**, **0.5x** and **2x** results). On the other hand, despite improvement, both RBFN and PR only exhibit limited generalizability. Indeed, for both models the performance for the **0.5x** condition is generally worse than that for the **1x** and **2x** exposure. We note that the powerful HSCNN-R has many more parameters than the polynomial or RBF regressions

(so it is not entirely surprising that the DNN model improves more significantly given the augmented training data).

Another interesting phenomenon can be viewed in the results is that: the spectral accuracy does not imply color accuracy. We see from the mean $\Delta E_{00}$ results, the most primitive LR—albeit much less accurate in spectra—is much accurate in colors than other more complicated models, including RBFN and HSCNN-R (with or without data augmentation). As commented in Figure 2, all presented models are nonetheless physically implausible due to the non-zero color errors.

*5.2. Effectiveness of Physically Plausible Spectral Reconstruction*

5.2.1. Color Fidelity and Spectral Accuracy

In Table 3, we present the effectiveness of physically plausible SR in color and spectral accuracy. Under the "Original" headings, we show the results of the original models (those found in the original citations), and under the "Physically Plausible" headings we present the results of the physically plausible version of the models. For all presented metrics, we calculated the mean and worst-case (99.9 percentile error) of each test image, and then averaged them over all testing images. In continuation of the analysis of the exposure invariance presented in the previous section, we also present the performance of physically plausible SR under varying exposure conditions (i.e., **1x**, **0.5x** and **2x**) in Table 4.

First, let us consider color accuracy. Looking at the error statistics of $\Delta E_{00}$ in Table 3, it is clear that our physically plausible approach forces all models to recover spectra of the exact same colors as the ground-truth—thus, the 0 color error under all circumstances. Then, the spectral accuracy results in all four spectral metrics illustrate that there is no penalty to physically plausible SR (note that for $\Delta E_{00}$, MRAE and RMSE, the lower the numbers are the better, while for GFC and PSNR, the higher the better). Indeed, on average (despite few cases of disagreements among different metrics) enforcing physical plausibility results in a small increase in mean performance. These results indicate that we can, in effect, recover spectra of perfect color fidelity without deteriorating the spectral accuracy. For visualized results, see Figure 13.

Finally, let us look at Table 4. The implementation of physically plausible SR does not influence very much how the models react to exposure change. Indeed, LR, RPR and A+ are still exposure invariant, while $\text{RBFN}^{(\beta = 1)}$, $\text{PR}^{(\beta = 1)}$ and $\text{HSCNN-R}^{(\beta = 1)}$ are not. Additionally, the effectiveness of data augmentation, i.e., $\text{RBFN}^{(\beta = 2.5)}$, $\text{PR}^{(\beta = 2.5)}$ and $\text{HSCNN-R}^{(\beta = 10)}$, still holds for physically plausible SR. Notice that the $\Delta E_{00}$ color error remains zero for all physically plausible models even in the situation that some models' spectral accuracies deteriorate in varying exposure conditions.

Jointly considering the effectiveness of our proposals—intensity-scaling data augmentation and physically plausible SR—we have achieved SR with no color error and stabilized performance under changing exposure.

**Table 3.** The colorand spectral accuracy results as the averaged per-image mean and 99.9th percentile (pt). The results are shown respectively in $\Delta E_{00}$, MRAE, GFC, RMSE and PSNR.

| | $\Delta E_{00}$ (Color Error) | | | | MRAE (%) (Spectral Error) | | | | GFC (Spectral Error) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Original | | Physically Plausible | | Original | | Physically Plausible | | Original | | Physically Plausible | |
| | Mean | 99.9 pt | Mean | 99.9 pt | Mean | 99.9 pt | Mean | 99.9 pt | Mean | 99.9 pt | Mean | 99.9 pt |
| LR | 0.05 | 0.79 | 0.00 | 0.00 | 6.24 | 22.45 | 6.23 | 22.53 | 0.9966 | 0.9770 | 0.9966 | 0.9767 |
| RPR | 0.14 | 1.48 | 0.00 | 0.00 | 4.69 | 24.06 | 4.60 | 24.86 | 0.9979 | 0.9712 | 0.9979 | 0.9640 |
| A+ | 0.06 | 2.47 | 0.00 | 0.00 | 3.87 | 21.06 | 3.83 | 20.65 | 0.9983 | 0.9770 | 0.9983 | 0.9770 |
| RBFN$^{(\beta = 1)}$ | 0.32 | 9.24 | 0.00 | 0.00 | 2.06 | 14.44 | 1.96 | 13.09 | 0.9994 | 0.9852 | 0.9994 | 0.9854 |
| RBFN$^{(\beta = 2.5)}$ | 0.15 | 3.36 | 0.00 | 0.00 | 4.20 | 17.25 | 4.15 | 17.00 | 0.9986 | 0.9832 | 0.9986 | 0.9834 |
| PR$^{(\beta = 1)}$ | 0.01 | 0.18 | 0.00 | 0.00 | 1.95 | 12.84 | 1.94 | 12.81 | 0.9994 | 0.9841 | 0.9994 | 0.9843 |
| PR$^{(\beta = 2.5)}$ | 0.01 | 0.07 | 0.00 | 0.00 | 3.50 | 17.95 | 3.46 | 18.38 | 0.9989 | 0.9814 | 0.9989 | 0.9802 |
| HSCNN-R$^{(\beta = 1)}$ | 0.10 | 2.06 | 0.00 | 0.00 | 1.73 | 12.10 | 1.76 | 12.68 | 0.9995 | 0.9864 | 0.9995 | 0.9842 |
| HSCNN-R$^{(\beta = 10)}$ | 0.15 | 2.46 | 0.00 | 0.00 | 2.96 | 16.14 | 2.93 | 21.09 | 0.9991 | 0.9841 | 0.9991 | 0.9686 |

| | RMSE (Spectral Error) | | | | PSNR (dB) (Spectral Error) | | | |
|---|---|---|---|---|---|---|---|---|
| | Original | | Physically Plausible | | Original | | Physically Plausible | |
| | Mean | 99.9 pt | Mean | 99.9 pt | Mean | 99.9 pt | Mean | 99.9 pt |
| LR | 33.26 | 153.49 | 33.23 | 153.35 | 43.34 | 30.24 | 43.36 | 30.33 |
| RPR | 27.80 | 167.17 | 27.49 | 172.33 | 45.49 | 29.93 | 45.71 | 29.84 |
| A+ | 23.97 | 161.69 | 24.36 | 165.61 | 48.23 | 29.79 | 48.21 | 29.65 |
| RBFN$^{(\beta = 1)}$ | 18.30 | 152.57 | 17.50 | 138.23 | 50.63 | 31.04 | 50.98 | 31.62 |
| RBFN$^{(\beta = 2.5)}$ | 27.70 | 142.46 | 27.24 | 139.51 | 45.54 | 30.84 | 45.67 | 31.06 |
| PR$^{(\beta = 1)}$ | 17.05 | 142.31 | 17.06 | 142.55 | 50.86 | 31.72 | 50.86 | 31.71 |
| PR$^{(\beta = 2.5)}$ | 23.88 | 143.93 | 23.75 | 146.78 | 47.03 | 31.07 | 47.10 | 30.96 |
| HSCNN-R$^{(\beta = 1)}$ | 16.33 | 139.58 | 16.34 | 137.24 | 52.34 | 31.58 | 52.08 | 31.70 |
| HSCNN-R$^{(\beta = 10)}$ | 23.56 | 167.82 | 22.67 | 165.65 | 49.07 | 29.47 | 49.38 | 29.55 |

**Table 4.** The spectral and color accuracy of the "physically plausible" SR under original (**1x**), half (**0.5x**) and double exposure settings (**2x**). The results are shown in mean MRAE, mean GFC and mean $\Delta E_{00}$.

|  | Mean MRAE (%) (Spectral Error) Physically Plausible | | | Mean GFC (Spectral Error) Physically Plausible | | | Mean $\Delta E_{00}$ (Color Error) Physically Plausible | | |
|---|---|---|---|---|---|---|---|---|---|
|  | **1x** | **0.5x** | **2x** | **1x** | **0.5x** | **2x** | **1x** | **0.5x** | **2x** |
| LR | 6.23 | 6.23 | 6.23 | 0.9966 | 0.9966 | 0.9966 | 0.00 | 0.00 | 0.00 |
| RPR | 4.60 | 4.60 | 4.60 | 0.9979 | 0.9979 | 0.9979 | 0.00 | 0.00 | 0.00 |
| A+ | 3.83 | 3.83 | 3.83 | 0.9983 | 0.9983 | 0.9983 | 0.00 | 0.00 | 0.00 |
| RBFN$^{(\beta = 1)}$ | 1.96 | 17.6 | 7.63 | 0.9994 | 0.9773 | 0.9958 | 0.00 | 0.00 | 0.00 |
| RBFN$^{(\beta = 2.5)}$ | 4.15 | 5.47 | 4.19 | 0.9986 | 0.9982 | 0.9983 | 0.00 | 0.00 | 0.00 |
| PR$^{(\beta = 1)}$ | 1.94 | 9.72 | 13.07 | 0.9994 | 0.9948 | 0.9899 | 0.00 | 0.00 | 0.00 |
| PR$^{(\beta = 2.5)}$ | 3.46 | 4.93 | 3.55 | 0.9989 | 0.9984 | 0.9986 | 0.00 | 0.00 | 0.00 |
| HSCNN-R$^{(\beta = 1)}$ | 1.76 | 15.33 | 6.39 | 0.9995 | 0.9844 | 0.9972 | 0.00 | 0.00 | 0.00 |
| HSCNN-R$^{(\beta = 10)}$ | 2.93 | 3.00 | 2.88 | 0.9991 | 0.9991 | 0.9991 | 0.00 | 0.00 | 0.00 |

### 5.2.2. Color Fidelity under Different Viewing Conditions

Here, we investigate using the hyperspectral recoveries (delivered by the various algorithms) to predict the colors of the same scene under either a different illumination or a different camera. To change the illumination of the scene, first we divide (component-wise) the whole hyperspectral image by the original white spectra and then multiply the image by a target illuminant's spectrum. Then, this newly derived hyperspectral image can be used to generate the relighted RGB scene using the color formation formula in Equation (2). As for simulating the color responses of a different camera (different from the one that generates the RGBs used to train SR), we need simply to incorporate a different set of spectral sensitivities in Equation (2). The new illumination spectra and camera sensitivities used in the experiments are shown in Figures 11 and 12, respectively.

Note that by relighting the scenes to CIE Illuminant E—which replaces the original illumination spectrum in each image with a "flat" spectral power distribution—effectively, we obtain the new spectra that are (individually) a scaling factor apart from the "reflectance spectra", which are pure object surfaces' properties without the influence of the illumination's spectral property.

We present the color fidelity results of changing the illumination or camera in Table 5. In this experiment we only tested the models under the **1x** (original) exposure condition; i.e., we did not test for exposure variation. Visualized results for CIE Illuminat A relighting can be found in the rightmost column of Figure 13.
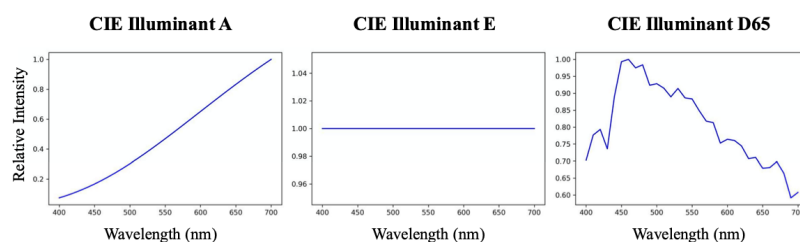


**Figure 11.** Target illuminants for scene relighting: CIE Illuminants A (**left**), E (**middle**) and D65 (**right**).
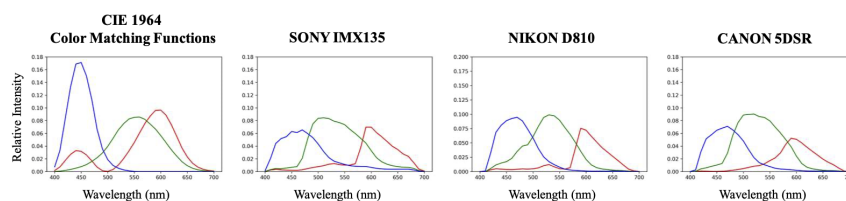


**Figure 12.** The spectral sensitivities of the ground-truth RGBs used for training (CIE 1964 color matching functions) and for testing (SONY IMX135, NIKON D810 and CANON 5DSR).

**Table 5.** The color accuracy when changing the illumination (**top**) or camera (**bottom**). The results are shown in the averaged per-image mean and 99.9th percentile (pt) $\Delta E_{00}$.

| | CIE Illuminant A | | | | CIE Illuminant E | | | | CIE Illuminant D65 | | | |
| | Original | | Physically Plausible | | Original | | Physically Plausible | | Original | | Physically Plausible | |
| | Mean | 99.9 pt | Mean | 99.9 pt | Mean | 99.9 pt | Mean | 99.9 pt | Mean | 99.9 pt | Mean | 99.9 pt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LR | 0.38 | 3.89 | 0.38 | 4.00 | 0.57 | 6.58 | 0.56 | 6.33 | 0.49 | 6.05 | 0.47 | 5.67 |
| RPR | 0.32 | 4.89 | 0.29 | 4.36 | 0.51 | 6.49 | 0.46 | 6.26 | 0.44 | 5.83 | 0.39 | 5.66 |
| A+ | 0.27 | 4.90 | 0.24 | 4.53 | 0.40 | 6.72 | 0.38 | 6.02 | 0.34 | 6.33 | 0.31 | 5.51 |
| RBFN$^{(\beta=1)}$ | 0.37 | 10.22 | 0.16 | 3.66 | 0.39 | 10.67 | 0.14 | 3.24 | 0.38 | 10.74 | 0.13 | 3.18 |
| RBFN$^{(\beta=2.5)}$ | 0.41 | 5.80 | 0.35 | 3.97 | 0.58 | 7.28 | 0.54 | 5.69 | 0.49 | 6.79 | 0.45 | 5.07 |
| PR$^{(\beta=1)}$ | 0.17 | 3.51 | 0.17 | 3.48 | 0.14 | 2.89 | 0.14 | 2.88 | 0.14 | 2.88 | 0.14 | 2.86 |
| PR$^{(\beta=2.5)}$ | 0.26 | 3.77 | 0.25 | 3.74 | 0.46 | 5.36 | 0.45 | 5.30 | 0.38 | 4.79 | 0.37 | 4.73 |
| HSCNN-R$^{(\beta=1)}$ | 0.18 | 4.12 | 0.15 | 3.75 | 0.18 | 3.91 | 0.12 | 2.92 | 0.18 | 4.06 | 0.12 | 2.95 |
| HSCNN-R$^{(\beta=10)}$ | 0.31 | 5.41 | 0.26 | 4.95 | 0.53 | 7.67 | 0.43 | 7.12 | 0.44 | 7.03 | 0.35 | 6.29 |

$\Delta E_{00}$ (Color Error)

| | SONY IMX135 | | | | NIKON D810 | | | | CANON 5DSR | | | |
| | Original | | Physically Plausible | | Original | | Physically Plausible | | Original | | Physically Plausible | |
| | Mean | 99.9 pt | Mean | 99.9 pt | Mean | 99.9 pt | Mean | 99.9 pt | Mean | 99.9 pt | Mean | 99.9 pt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LR | 0.33 | 3.39 | 0.33 | 3.36 | 0.63 | 5.90 | 0.63 | 5.83 | 0.41 | 3.95 | 0.41 | 3.89 |
| RPR | 0.28 | 3.93 | 0.26 | 3.76 | 0.54 | 6.74 | 0.53 | 6.72 | 0.38 | 4.75 | 0.35 | 4.52 |
| A+ | 0.27 | 4.93 | 0.24 | 4.37 | 0.49 | 8.33 | 0.45 | 7.86 | 0.34 | 5.88 | 0.30 | 5.29 |
| RBFN$^{(\beta=1)}$ | 0.43 | 10.75 | 0.23 | 4.75 | 0.56 | 13.01 | 0.39 | 8.12 | 0.47 | 11.55 | 0.26 | 5.46 |
| RBFN$^{(\beta=2.5)}$ | 0.36 | 4.92 | 0.30 | 3.33 | 0.71 | 6.48 | 0.66 | 5.26 | 0.47 | 5.22 | 0.40 | 3.50 |
| PR$^{(\beta=1)}$ | 0.23 | 4.34 | 0.23 | 4.33 | 0.42 | 8.17 | 0.43 | 8.15 | 0.27 | 5.38 | 0.27 | 5.37 |
| PR$^{(\beta=2.5)}$ | 0.24 | 3.52 | 0.23 | 3.53 | 0.51 | 6.11 | 0.50 | 6.16 | 0.33 | 3.94 | 0.32 | 3.97 |
| HSCNN-R$^{(\beta=1)}$ | 0.26 | 5.26 | 0.24 | 4.99 | 0.42 | 8.70 | 0.40 | 8.60 | 0.29 | 5.97 | 0.27 | 5.71 |
| HSCNN-R$^{(\beta=10)}$ | 0.35 | 5.75 | 0.28 | 5.10 | 0.62 | 9.36 | 0.56 | 9.06 | 0.43 | 6.40 | 0.36 | 5.97 |
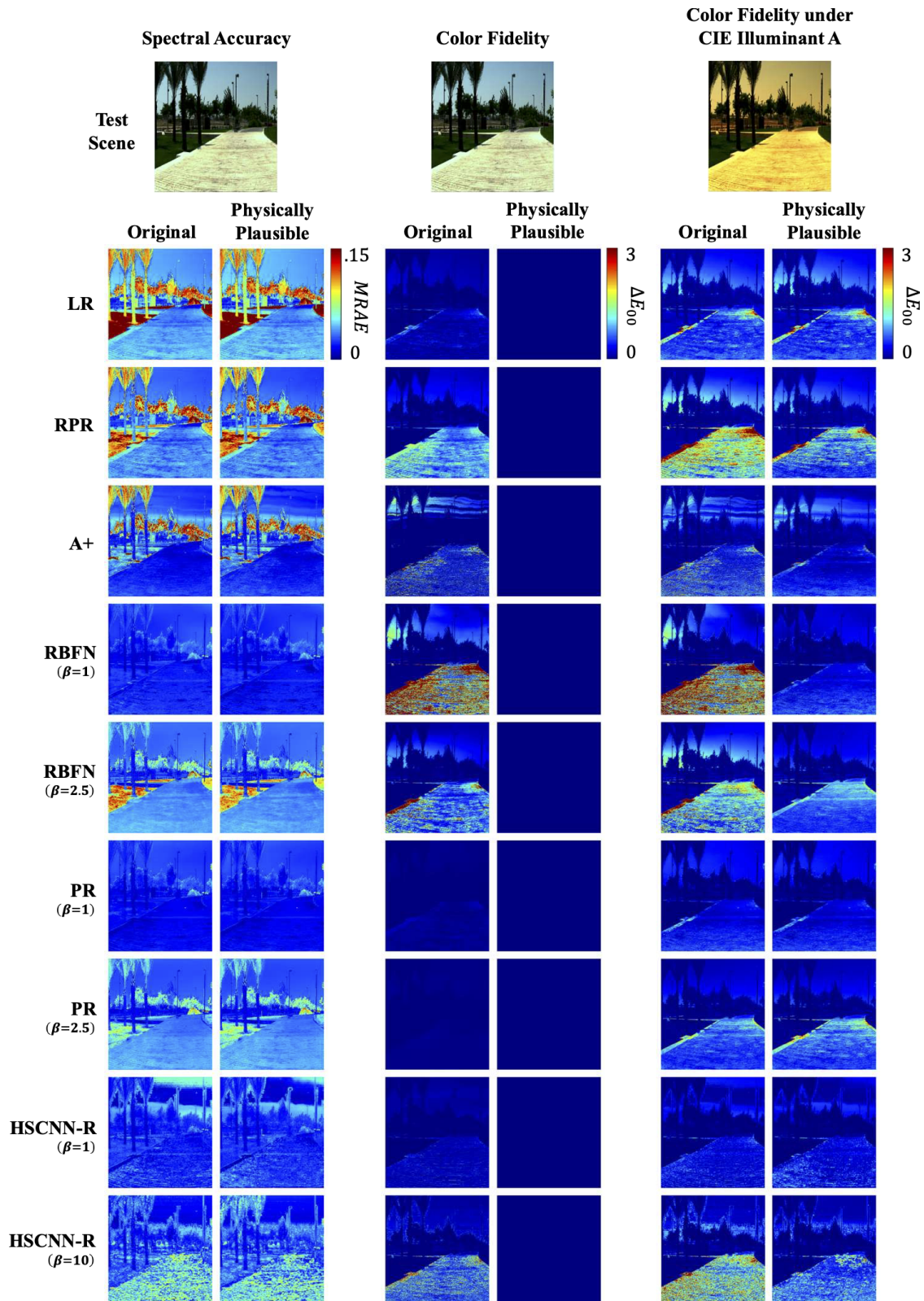
**Figure 13.** The reconstruction error maps of an example scene in terms of spectral accuracy (**left**; in MRAE), color fidelity (**middle**; in $\Delta E_{00}$) and color fidelity under CIE Illuminant A (**right**; in $\Delta E_{00}$).

First, we see that for all models our physically plausible approach in general improves the cross-illumination and cross-camera color reproduction. If we look at the performances of the original models (without the physically plausible implementation and data augmentation), the $PR^{(\beta = 1)}$ model performs

the best in predicting the actual cross-illuminant colors, and is up to 30% better than the DNN-based HSCNN-R$^{(\beta = 1)}$ model. A similar performance increment is also shown when the camera changes. Ironically, compared to the spectral accuracy results (Table 3), we see PR$^{(\beta = 1)}$ recovers spectra that are 13% less accurate in mean MRAE than HSCNN-R$^{(\beta = 1)}$. This result tells that while most SR models primarily aim to minimize spectral errors, that does not ensure better performance in the general context of color fidelity, either under the original or changing viewing conditions. Additionally, we showed that as we implement the "physically plausible" HSCNN-R$^{(\beta = 1)}$—which actually contributes to improving color fidelity—it is then when HSCNN-R$^{(\beta = 1)}$ performs better than PR$^{(\beta = 1)}$.

Next, the physically plausible RBFN$^{(\beta = 1)}$ exhibited the most improvement from the original model compared to others. Indeed, on average a 60% improvement in cross-viewing-condition color fidelity was delivered by making RBFN$^{(\beta = 1)}$ physically plausible. This performance increment also makes it one of the best performing models, on par with PR$^{(\beta = 1)}$ and HSCNN-R$^{(\beta = 1)}$.

Further, if we consider the effect of data augmentation, we see that—similarly to the spectral accuracy results—RBFN$^{(\beta = 2.5)}$, PR$^{(\beta = 2.5)}$ and HSCNN-R$^{(\beta = 10)}$ in general worsen the performance from that of their original counterparts ($\beta = 1$). In various circumstances, those data-augmented models deliver much worse mean and worst-case performances compared to the exposure-invariant LR, RPR and A+. Especially, the physically plausible A+ method performs better "in all conditions" than the data augmented physically plausible HSCNN-R$^{(\beta = 10)}$. We remind the readers that the exposure-invariant models have the benefit of being able to generalize the exact same performance on the whole scale of physical brightness [35] (e.g., A+), as opposed to the finite range of (often suboptimal) generalizability induced purely by data augmentation (e.g., HSCNN-R).

Given all these experimental results the obvious question to ask is "which algorithm should I choose?" Well, consistent with the trend of adopting DNNs, the HSCNN-R solution—where physical plausibility is enforced and data augmentation is implemented to the training regime—is a good choice overall. However, considering its overhead of long training time and required computing resources, in terms of the various aspects we present in this paper, the exemplar DNN model does not appear to be much superior than the rest of the regression-based methods.

## 6. Conclusions

Spectral reconstruction algorithms seek to map RGB images to hyperspectral images. Most models are designed to minimize the spectral error of the reconstruction, but the underlying physical relationship between spectra and colors is not preserved. This physically non-plausible mapping causes the issues of poor color fidelity and inconsistent performance for the same object viewed at different exposures.

In this paper we provide solutions for both issues. First, we show that all plausible spectra can be represented by a fixed fundamental metamer defined by a linear combination of camera spectral sensitivities, and a metameric black which does not contribute to the color formation. Relative to this insight, the spectral recovery sets out to reconstruct only the metameric black's coefficients from the RGBs, while the fundamental metamer is derived directly. This ensures that the predicted spectra are always of the exact same RGBs found in the original images. Secondly, we show that better robustness against exposure change can be achieved by augmenting the training data with randomly-generated intensity scaling factors.

Another contribution of this paper is that we performed extensive studies on the models' colorimetric performances apart from the usual spectral accuracy measure. Our evaluations here included scene relighting and color predictions for different cameras. Our results show that the best performing models— from a color fidelity point of view—do not necessarily correspond to the most spectrally accurate recovery models.

**Supplementary Materials:** The code of the methods introduced in this paper is available at https://github.com/EthanLinYitun/Physically_Plausible_Spectral_Reconstruction.

## References

1. Veganzones, M.; Tochon, G.; Dalla-Mura, M.; Plaza, A.; Chanussot, J. Hyperspectral image segmentation using a new spectral unmixing-based binary partition tree representation. *IEEE Trans. Image Process.* **2014**, *23*, 3574–3589. [CrossRef] [PubMed]

2. Chen, Y.; Zhao, X.; Jia, X. Spectral–spatial classification of hyperspectral data based on deep belief network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2381–2392. [CrossRef]

3. Ghamisi, P.; Dalla Mura, M.; Benediktsson, J. A survey on spectral—spatial classification techniques based on attribute profiles. *IEEE Trans. Geosci. Remote. Sens.* **2014**, *53*, 2335–2353. [CrossRef]

4. Tao, C.; Pan, H.; Li, Y.; Zou, Z. Unsupervised spectral-spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2438–2442.

5. Chen, C.; Li, W.; Su, H.; Liu, K. Spectral-spatial classification of hyperspectral image based on kernel extreme learning machine. *Remote. Sens.* **2014**, *6*, 5795–5814. [CrossRef]

6. Jablonski, J.A.; Bihl, T.J.; Bauer, K.W. Principal component reconstruction error for hyperspectral anomaly detection. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1725–1729. [CrossRef]

7. Zhang, Y.; Mou, X.; Wang, G.; Yu, H. Tensor-based dictionary learning for spectral CT reconstruction. *IEEE Trans. Med Imaging* **2016**, *36*, 142–154. [CrossRef]

8. Zhang, Y.; Xi, Y.; Yang, Q.; Cong, W.; Zhou, J.; Wang, G. Spectral CT reconstruction with image sparsity and spectral mean. *IEEE Trans. Comput. Imaging* **2016**, *2*, 510–523. [CrossRef]

9. Deering, M. Multi-Spectral Color Correction. U.S. Patent 6,950,109, 27 September 2005

10. Abrardo, A.; Alparone, L.; Cappellini, I.; Prosperi, A. Color constancy from multispectral images. In Proceedings of the International Conference on Image Processing, Kobe, Japan, 24–28 October 1999; Volume 3, pp. 570–574.

11. Cheung, V.; Westland, S.; Li, C.; Hardeberg, J.; Connah, D. Characterization of trichromatic color cameras by using a new multispectral imaging technique. *J. Opt. Soc. Am. A* **2005**, *22*, 1231–1240. [CrossRef]

12. Lam, A.; Sato, I. Spectral modeling and relighting of reflective-fluorescent scenes. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 1452–1459.

13. Xu, P.; Xu, H.; Diao, C.; Ye, Z. Self-training-based spectral image reconstruction for art paintings with multispectral imaging. *Appl. Opt.* **2017**, *56*, 8461–8470. [CrossRef]

14. Gat, N. Imaging spectroscopy using tunable filters: A review. In Proceedings of the Wavelet Applications VII, International Society for Optics and Photonics, Orlando, FL, USA, 26 July 2000; Volume 4056, pp. 50–64.

15. Green, R.O.; Eastwood, M.L.; Sarture, C.M.; Chrien, T.G.; Aronsson, M.; Chippendale, B.J.; Faust, J.A.; Pavri, B.E.; Chovit, C.J.; Solis, M.; et al. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). *Remote Sens. Environ.* **1998**, *65*, 227–248. [CrossRef]

16. Cao, X.; Du, H.; Tong, X.; Dai, Q.; Lin, S. A prism-mask system for multispectral video acquisition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 2423–2435.

17. Correa, C.V.; Arguello, H.; Arce, G.R. Snapshot colored compressive spectral imager. *J. Opt. Soc. Am. A* **2015**, *32*, 1754–1763. [CrossRef] [PubMed]

18. Garcia, H.; Correa, C.V.; Arguello, H. Multi-resolution compressive spectral imaging reconstruction from single pixel measurements. *IEEE Trans. Image Process.* **2018**, *27*, 6174–6184. [CrossRef] [PubMed]

19. Arguello, H.; Arce, G.R. Colored coded aperture design by concentration of measure in compressive spectral imaging. *IEEE Trans. Image Process.* **2014**, *23*, 1896–1908. [CrossRef]

20. Galvis, L.; Lau, D.; Ma, X.; Arguello, H.; Arce, G.R. Coded aperture design in compressive spectral imaging based on side information. *Appl. Opt.* **2017**, *56*, 6332–6340. [CrossRef]

21. Lin, X.; Liu, Y.; Wu, J.; Dai, Q. Spatial-spectral encoded compressive hyperspectral imaging. *ACM Trans. Graph.* **2014**, *33*, 233. [CrossRef]

22. Rueda, H.; Arguello, H.; Arce, G.R. DMD-based implementation of patterned optical filter arrays for compressive spectral imaging. *J. Opt. Soc. Am. A* **2015**, *32*, 80–89. [CrossRef]

23. Zhao, Y.; Guo, H.; Ma, Z.; Cao, X.; Yue, T.; Hu, X. Hyperspectral Imaging With Random Printed Mask. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 10149–10157.

24. Shrestha, R.; Hardeberg, J.Y.; Khan, R. Spatial arrangement of color filter array for multispectral image acquisition. In Proceedings of the Sensors, Cameras, and Systems for Industrial, Scientific, and Consumer Applications XII, International Society for Optics and Photonics, San Francisco, CA, USA, 25–27 January 2011; p. 787503.

25. Murakami, Y.; Yamaguchi, M.; Ohyama, N. Hybrid-resolution multispectral imaging using color filter array. *Opt. Express* **2012**, *20*, 7173–7183. [CrossRef]

26. Mihoubi, S.; Losson, O.; Mathon, B.; Macaire, L. Multispectral demosaicing using intensity-based spectral correlation. In Proceedings of the International Conference on Image Processing Theory, Tools and Applications, Orleans, France, 10–13 November 2015; pp. 461–466.

27. Brauers, J.; Schulte, N.; Aach, T. Multispectral filter-wheel cameras: Geometric distortion model and compensation algorithms. *IEEE Trans. Image Process.* **2008**, *17*, 2368–2380. [CrossRef] [PubMed]

28. Wang, L.; Xiong, Z.; Gao, D.; Shi, G.; Zeng, W.; Wu, F. High-speed hyperspectral video acquisition with a dual-camera architecture. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 4942–4950.

29. Park, J.I.; Lee, M.H.; Grossberg, M.D.; Nayar, S.K. Multispectral imaging using multiplexed illumination. In Proceedings of the International Conference on Computer Vision, Rio De Janeiro, Brazil, 14–21 October 2007; pp. 1–8.

30. Hirai, K.; Tanimoto, T.; Yamamoto, K.; Horiuchi, T.; Tominaga, S. An LED-based spectral imaging system for surface reflectance and normal estimation. In Proceedings of the International Conference on Signal-Image Technology & Internet-Based Systems, Kyoto, Japan, 2–5 December 2013; pp. 441–447.

31. Shrestha, R.; Hardeberg, J.Y.; Mansouri, A. One-shot multispectral color imaging with a stereo camera. In Proceedings of the Digital Photography VII, International Society for Optics and Photonics, San Francisco, CA, USA, 24–25 January 2011; p. 787609.

32. Takatani, T.; Aoto, T.; Mukaigawa, Y. One-shot hyperspectral imaging using faced reflectors. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4039–4047.

33. Heikkinen, V.; Lenz, R.; Jetsu, T.; Parkkinen, J.; Hauta-Kasari, M.; Jääskeläinen, T. Evaluation and unification of some methods for estimating reflectance spectra from RGB images. *J. Opt. Soc. Am. A* **2008**, *25*, 2444–2458. [CrossRef]

34. Connah, D.; Hardeberg, J. Spectral recovery using polynomial models. In Proceedings of the Color Imaging X: Processing, Hardcopy, and Applications, International Society for Optics and Photonics, San Jose, CA, USA, 17 January 2005; Volume 5667, pp. 65–75.

35. Lin, Y.; Finlayson, G. Exposure Invariance in Spectral Reconstruction from RGB Images. In Proceedings of the Color and Imaging Conference, Society for Imaging Science and Technology, Paris, France, 21–25 October 2019; Volume 2019, pp. 284–289.

36. Nguyen, R.; Prasad, D.; Brown, M. Training-based spectral reconstruction from a single RGB image. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 186–201.

37. Aeschbacher, J.; Wu, J.; Timofte, R. In defense of shallow learned spectral reconstruction from RGB images. In Proceedings of the International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 471–479.

38. Maloney, L.T.; Wandell, B.A. Color constancy: A method for recovering surface spectral reflectance. *J. Opt. Soc. Am. A* **1986**, *3*, 29–33. [CrossRef]

39. Agahian, F.; Amirshahi, S.A.; Amirshahi, S.H. Reconstruction of reflectance spectra using weighted principal component analysis. *Color Res. Appl.* **2008**, *33*, 360–371. [CrossRef]

40. Zhao, Y.; Berns, R.S. Image-based spectral reflectance reconstruction using the matrix R method. *Color Res. Appl.* **2007**, *32*, 343–351. [CrossRef]

41. Brainard, D.H.; Freeman, W.T. Bayesian color constancy. *J. Opt. Soc. Am. A* **1997**, *14*, 1393–1411. [CrossRef]
42. Morovic, P.; Finlayson, G.D. Metamer-set-based approach to estimating surface reflectance from camera RGB. *J. Opt. Soc. Am. A* **2006**, *23*, 1814–1822. [CrossRef]
43. Bianco, S. Reflectance spectra recovery from tristimulus values by adaptive estimation with metameric shape correction. *J. Opt. Soc. Am. A* **2010**, *27*, 1868–1877. [CrossRef]
44. Zuffi, S.; Santini, S.; Schettini, R. From color sensor space to feasible reflectance spectra. *IEEE Trans. Signal Process.* **2008**, *56*, 518–531. [CrossRef]
45. Arad, B.; Ben-Shahar, O. Sparse recovery of hyperspectral signal from natural RGB images. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 19–34.
46. Shi, Z.; Chen, C.; Xiong, Z.; Liu, D.; Wu, F. Hscnn+: Advanced cnn-based hyperspectral recovery from RGB images. In Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops, Perth, Australia, 2–6 December 2018; pp. 939–947.
47. Arad, B.; Ben-Shahar, O.; Timofte, R. NTIRE 2018 challenge on spectral reconstruction from RGB images. In Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 929–938.
48. Arad, B.; Timofte, R.; Ben-Shahar, O.; Lin, Y.; Finlayson, G. NTIRE 2020 challenge on spectral reconstruction from an RGB image. In Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020.
49. Arun, P.; Buddhiraju, K.; Porwal, A.; Chanussot, J. CNN based spectral super-resolution of remote sensing images. *Signal Process.* **2020**, *169*, 107394. [CrossRef]
50. Li, J.; Wu, C.; Song, R.; Li, Y.; Liu, F. Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from RGB images. In Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 462–463.
51. Joslyn Fubara, B.; Sedky, M.; Dyke, D. RGB to Spectral Reconstruction via Learned Basis Functions and Weights. In Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 480–481.
52. Chakrabarti, A.; Zickler, T. Statistics of real-world hyperspectral images. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA 20–25 June 2011; pp. 193–200.
53. Zhao, Y.; Po, L.M.; Yan, Q.; Liu, W.; Lin, T. Hierarchical regression network for spectral reconstruction from RGB images. In Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 422–423.
54. Sharma, G.; Wu, W.; Dalal, E.N. The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Res. Appl.* **2005**, *30*, 21–30. [CrossRef]
55. Hardeberg, J.Y. On the spectral dimensionality of object colours. In Proceedings of the Conference on Colour in Graphics, Imaging, and Vision, Society for Imaging Science and Technology, Poitiers, France, 2–5 April 2002; Volume 2002, pp. 480–485.
56. Romero, J.; García-Beltrán, A.; Hernández-Andrés, J. Linear bases for representation of natural and artificial illuminants. *J. Opt. Soc. Am. A* **1997**, *14*, 1007–1014. [CrossRef]
57. Lee, T.W.; Wachtler, T.; Sejnowski, T.J. The spectral independent components of natural scenes. In Proceedings of the International Workshop on Biologically Motivated Computer Vision, Seoul, Korea, 15–17 May 2000; pp. 527–534.
58. Marimont, D.H.; Wandell, B.A. Linear models of surface and illuminant spectra. *J. Opt. Soc. Am. A* **1992**, *9*, 1905–1913. [CrossRef]
59. Parkkinen, J.P.; Hallikainen, J.; Jaaskelainen, T. Characteristic spectra of Munsell colors. *J. Opt. Soc. Am. A* **1989**, *6*, 318–322. [CrossRef]
60. Strang, G. *Introduction to Linear Algebra*, 5th ed.; Wellesley-Cambridge Press: Wellesley, MA, USA, 2016; pp. 135–149; 219–232; 382–391.
61. Finlayson, G.; Morovic, P. Metamer sets. *J. Opt. Soc. Am. A* **2005**, *22*, 810–819. [CrossRef]
62. Bashkatov, A.; Genina, E.; Kochubey, V.; Tuchin, V. Optical properties of the subcutaneous adipose tissue in the spectral range 400–2500 nm. *Opt. Spectrosc.* **2005**, *99*, 836–842. [CrossRef]
63. Pan, Z.; Healey, G.; Prasad, M.; Tromberg, B. Face recognition in hyperspectral images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 1552–1560.

64. Wandell, B.A. The synthesis and analysis of color images. *IEEE Trans. Pattern Anal. Mach. Intell.* **1987**, *1*, 2–13. [CrossRef] [PubMed]

65. Aharon, M.; Elad, M.; Bruckstein, A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* **2006**, *54*, 4311–4322. [CrossRef]

66. Tikhonov, A.; Goncharsky, A.; Stepanov, V.; Yagola, A. *Numerical Methods for the Solution of Ill-Posed Problems*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013; Volume 328.

67. Webb, G.I. Overfitting. In *Encyclopedia of Machine Learning*; Sammut, C., Webb, G.I., Eds.; Springer: Boston, MA, USA, 2010; p. 744. [CrossRef]

68. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

69. Cheney, W.; Kincaid, D. *Linear Algebra: Theory and Applications*; Jones & Bartlett Learning: Sudbury, MA, USA, 2009; Volume 110, pp. 544–558.

70. Cohen, J.B.; Kappauf, W.E. Metameric color stimuli, fundamental metamers, and Wyszecki's metameric blacks. *Am. J. Psychol.* **1982**, *95*, 537–564. [CrossRef]

71. Commission Internationale de l'Eclairage. *CIE Proceedings (1964) Vienna Session, Committee Report E-1.4*; Commission Internationale de l'Eclairage: Vienna, Austria, 1964.

72. Commission Internationale de l'Eclairage. *Commission Internationale de L'eclairage Proceedings (1931)*; Cambridge University: Cambridge, UK, 1932.

73. Robertson, A.R. The CIE 1976 color-difference formulae. *Color Res. Appl.* **1977**, *2*, 7–11. [CrossRef]

74. Süsstrunk, S.; Buckley, R.; Swen, S. Standard RGB color spaces. In Proceedings of the Color and Imaging Conference, Society for Imaging Science and Technology, Scottsdale, AZ, USA, 16–19 November 1999; pp. 127–134.