Spectra without Spectra: Learning Spectral Reconstruction from Real-World RGB Data with Sparse Supervision*

Shira Werman
Samsung Israel R&D Center
Tel-Aviv, Israel
shira.werman@samsung.com

Genady Paikin
Samsung Israel R&D Center
Tel-Aviv, Israel
genady.p@samsung.com

Roy Shaul
Samsung Israel R&D Center
Tel-Aviv, Israel
roy.s@samsung.com

Evgeny Soloveichik
Samsung Israel R&D Center
Tel-Aviv, Israel
evgeny.s@samsung.com

Abstract

Our work addresses hyperspectral reconstruction from RGB images by identifying and mitigating a significant gap between synthetic and real-world data. Methods mostly rely on synthetic RGB data generated from hyperspectral datasets, which we show generalizes poorly to real-world camera inputs. To bridge this gap, we propose a novel training framework that leverages accessible real-world datasets containing RGB images with color charts and illumination measurements. Our mixed-supervision strategy, which is adaptable to existing state-of-the-art models, combines explicit supervision in regions with known spectra and physics-based self-regularization techniques across entire images. In particular, sparse supervision is provided on color chart patches with measured spectral ground truth, serving as reliable anchors to guide learning across the full image. We further strengthen reconstruction quality through two complementary regularizations to propagate spectral constraints across the entire image: a self-supervised RGB to RGB loss that enforces physically plausible reconstructions, and a relighting consistency loss that provides per-pixel target in the hyperspectral domain. Experiments using the MST++ backbone on the BeyondRGB dataset demonstrate a substantial improvement in angular error (from 18.91° to 5.52°) compared to models trained on synthetic data. This work offers a practical, accessible path toward robust real-world spectral reconstruction, moving beyond the reliance on synthetic data.

*The supplemental material can be found at https://github.com/shirawerman/Spectra-without-Spectra

1. Introduction

Hyperspectral (HS) imaging captures detailed spectral data for precise material identification, classification, and anomaly detection, benefiting applications like environmental monitoring, agriculture, and medical diagnostics. However, HS sensors are costly, require calibration, and have slow acquisition rates, limiting dynamic scene capture. In contrast, fast and accessible RGB imaging has driven research in spectral reconstruction to recover HS data from RGB images. These techniques derive a pixel-to-pixel mapping that maps each pixel value from the RGB domain to the HS domain. Given the ill-posed nature of the problem, early methods used prior knowledge, while recent deep learning approaches improve accuracy through data-driven techniques.

The NTIRE HS Reconstruction Challenges datasets [4–6] have become benchmarks for spectral reconstruction methods, providing synthetically generated, pixel-aligned RGB-HS image pairs. The published datasets and the NTIRE challenge have facilitated and inspired the advancement of numerous spectral reconstruction algorithms, both supervised and unsupervised, as it provides a valuable resource for training and evaluating these methods. However, recent works identify a gap between synthetic and real-world data. Models trained on spatially low-resolution synthesized RGB images struggle with high-resolution real-world imagery [8], and limited HS dataset variety hinders generalization to complex scenes [11]. Our experimental results strongly support this, showing performance degradation when models trained on synthetic data are applied to mobile camera RGB images.

In this work we aim to challenge the current practice and answer the question: Can real RGB camera data be used for

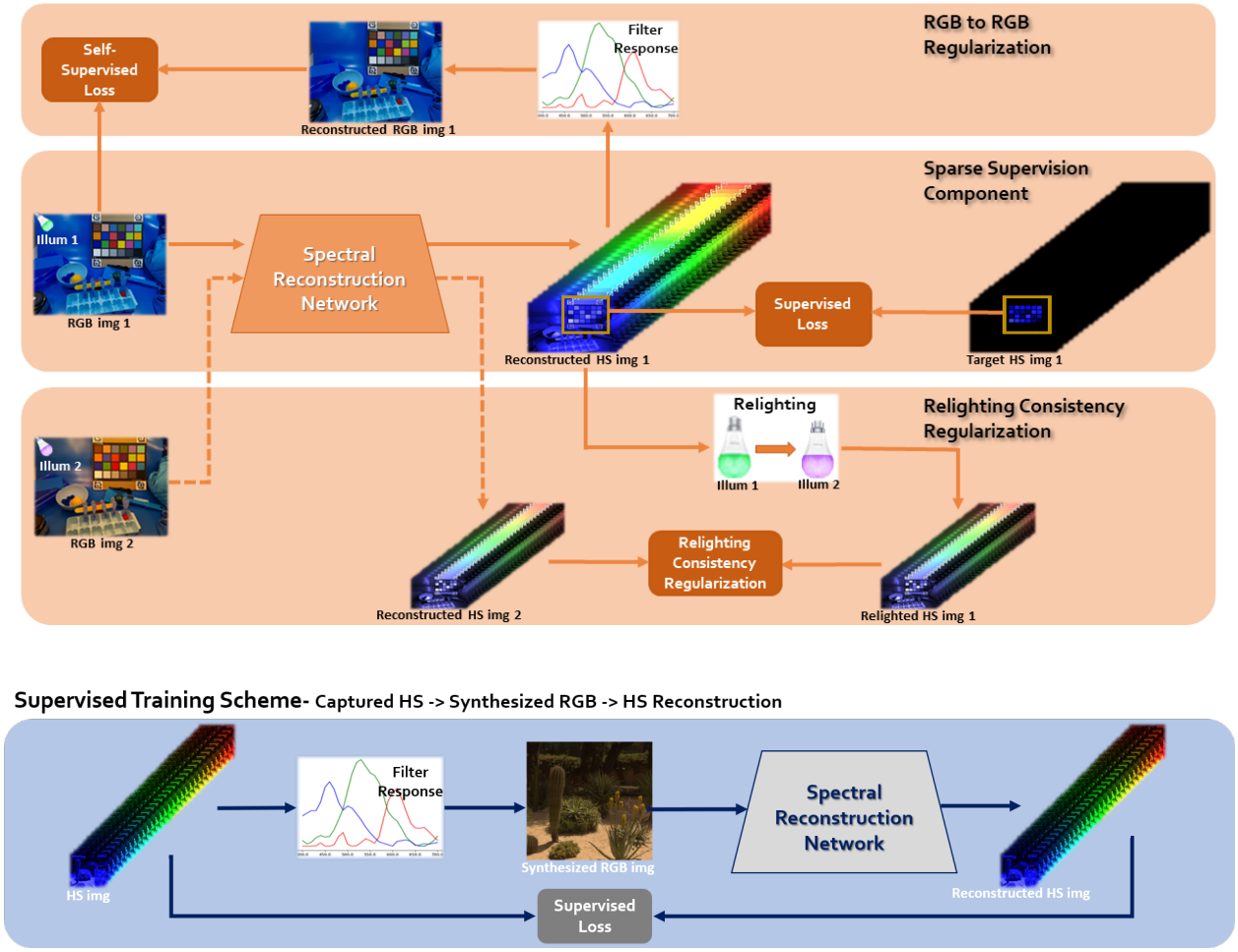

Figure 1. Top- the components of our method utilizing a dataset with color charts as reference GT. The input is a RAW RGB image and the output an HS image. In the middle component we compare the color chart patches to the GT measured values. In the top component we pass the network's output through the camera's filter response function to compare to the original input. In the bottom component we reconstruct two images from the same scene, comparing one with the relighted version of the second. Bottom- supervised training scheme: fully supervised method on RGB images synthesized from HS images.

spectral reconstruction and reduce the domain gap? Diverging from conventional approaches, we propose leveraging known spectral information found in RGB images. With the publication of BeyondRGB [14] this has become possible. In BeyondRGB, originally published for tasks such as white balance and illumination spectrum estimation, each image includes a color chart of type X-Rite ColorChecker Classic with measurable reflectance and an illumination measurement, providing spectral information for specific image regions. This approach is more accessible and cost-effective than traditional HS data collection. While it still requires a spectrum measuring device and color charts, these instruments are significantly more affordable than HS cameras. Moreover, it diminishes the need for simulations and abridges the domain gap.

Using this dataset, we develop a mixed-supervised deep learning framework for spectral reconstruction from RAW RGB images. We adopt the state-of-the-art MST++ architecture [7] as our backbone to demonstrate that our performance gains stem from the training methodology and data utilization rather than architectural changes. Our training approach integrates direct sparse supervision in regions with known spectra and incorporates two regularization techniques applied across all image pixels. Evaluation shows our method achieves a 5.52 angular error compared to 18.91 for fully supervised learning on synthesized RGB data. We also demonstrate the technique's extensibility to multispectral (MS) images available in BeyondRGB.

Our main contributions include:

1. A training framework for leveraging real-world RGB

data in spectral reconstruction, simplifying data collection and mitigating domain shift.
2. Demonstration of the failure of synthetic-data-trained models on real-world inputs.
3. A set of physical regularization losses (RGB consistency and Relighting consistency) that enable effective training with only sparse (chart-based) spectral supervision.

2. Related Work

In recent years, various approaches have been proposed for spectral reconstruction from RGB images. Early spectral reconstruction methods relied on data priors. Arad and Ben-Shahar [3] utilized sparse representation with a global dictionary of hyperspectral signals, while Aeschbacher *et al*. [1] developed local dictionaries. Subsequent improvements included spatial constraints [13] and Gaussian processes [2].

Prior-based methods face limitations in generalization across diverse scenarios and show sensitivity to noise and sensor-specific artifacts. Recent work has shifted toward data-driven approaches. Initial CNN-based models [18, 25, 27, 31, 36] evolved to include physical constraints [21] and attention mechanisms [16, 17, 26, 35]. Notable advances include MST++ [7], which won the NTIRE 2022 spectral reconstruction challenge [6] by using spectral-wise multi-head self-attention to effectively model inter-spectral correlations. The MST++ architecture consists of multiple single-stage spectral-wise transformers arranged in a U-shaped structure to extract multi-resolution contextual information. Alternative approaches explore implicit neural representations [32], different domains for reconstruction [10, 28], and specialized architectures [29, 30]. Most methods train on synthetic RGB data derived from hyperspectral datasets such as ICVL [3], CAVE [33], Harvard [9] and the datasets published in the NTIRE spectral reconstruction challenges [4–6]. Detailed information can be found in the supplementary material.

To address the scarcity of hyperspectral data, unsupervised techniques have emerged [8, 12, 19, 34, 38]. Li *et al*. [19] introduced MFormer with dual spectral-wise multi-head attention and a mask-guided band augmentation module. However, these methods still rely on evaluation on simulated data, so an appropriate evaluation dataset is lacking. Previous studies on spectral reflectance estimation have focused on reconstructing the spectral properties of standard color samples. Among these, some have demonstrated reflectance reconstruction on color chart patches, but primarily under standard illuminations and at limited scale [8, 20, 22, 23, 37]. Unlike these approaches focused on reflectance values, our work aims to recover the recorded light spectrum, which varies with illumination conditions.

3. Method

3.1. Problem Formulation

We rely on the common single illumination image formation model:

$$I_k(x,y) = \int_\Lambda R(x,y;\lambda)L(\lambda)f_k(\lambda)d\lambda \quad k \in \{R,G,B\} \tag{1}$$

Here, (x, y) represent the spatial coordinates of a given pixel and k denotes the channel (R, G or B) in the image I. R is the scene reflectance and L is the scene's illuminant. The system response function for the k^{th} channel is denoted as $f_k(\lambda)$. The parameter λ corresponds to the observed wavelength of light and the integration domain Λ encompasses the wavelengths where the camera exhibits a non-negligible response.

Our objective is to recover the discrete spectral value S in each pixel (x, y). Our measurements are between 400nm-700nm with 10nm gaps, so after discretization of the wavelength domain, S is equal to:

$$S(x,y;\lambda) = R(x,y;\lambda)L(\lambda), \quad \lambda = 400 + 10j, \quad j \in \{0,1,2,\ldots,30\}. \tag{2}$$

3.2. Datasets

We use a real RGB camera dataset with known spectral information derived from color charts. Additionally, we present results on a synthetic dataset and demonstrate that supervised methods trained on synthetic data struggle to generalize to real-world scenarios.

3.2.1. Dataset with Color Charts

Recently, a dataset with RGB and multispectral data has been published [14] dubbed BeyondRGB. The images span from lab scenes, and natural scenes including indoors and outdoors scenes, resulting in 1680 images with a wide range of spectra values. See figure 2 a-c for examples. For each image in the dataset there is an illumination measurement and a version with the X-Rite ColorChecker Classic. The lab sub dataset is composed of 13 scenes each recorded under 93 different illuminations. The RGB images were captured with a Samsung Galaxy S21 Plus SM-G996B camera and saved in RAW format without any further processing. For our method we use the RAW images after demosaicing, to minimize the impact on image colors and prevent distortion in spectral reconstruction.

The BeyondRGB dataset published the single illumination value of each image and we obtained the reflectance values of the X-Rite ColorChecker Classic present in the images. To derive the ground truth spectra, we multiply the measured reflectance values with the given single illumination value captured. These patch measurements multiplied

by illuminations are used as the ground truth in our training scheme. Unlike synthetic HS and RGB pairs where the RGB data is derived from the HS image, here the ground truth values were obtained independently of the input image, so the intensity of the color chart patches in the ground truth are unrelated to the intensity of the pixels in the RGB images. This is because the input is significantly influenced by the exposure set during capturing and the camera's sensitivity to light. To address the intensity issue, we normalize the ground truth values by the intensities in the RGB image I, like so:

$$H_{cc}(x, y; \lambda) = \frac{S_{cc}(x, y; \lambda)||I(x, y)||_1}{||f \cdot S_{cc}(x, y)||_1}. \tag{3}$$

Where $S_{cc}(x, y) = R_{cc}(x, y)L_{cc}(x, y)$ is the derived spectrum in a pixel (x, y) in a color chart patch and H_{cc} is the final target value.

If we apply the filter response functions f on H_{cc} and compute the $\mathcal{L}_1$ norm we'll get:

$$\begin{aligned} \|f \cdot H_{cc}(x, y; \lambda)\|_1 &= \left\| f \cdot \frac{S_{cc}(x, y; \lambda)\, \|I(x, y)\|_1}{\|f \cdot S_{cc}(x, y)\|_1} \right\|_1 \\ &= \|I(x, y)\|_1. \end{aligned} \tag{4}$$

ensuring that mapping the GT values to RGB preserves both the pixel's intensity scale and the expected RGB ratios. Our primary metric and training loss is SAM (eq. 9) which is scaleless, making manual scale adjustments relatively insignificant.

We computed the filter response functions f_k for the camera used in the dataset by conducting an optimization process on the BeyondRGB lab images. A more detailed explanation can be found in the supplementary material.

The BeyondRGB dataset comprises multispectral images captured with a low-cost, 16-filter prototype camera designed for mobile use [15, 24]. While the wide, overlapping filters make the images non-hyperspectral, they still offer higher spectral resolution than RGB. We compare spectral reconstruction performance using both RGB and MS data from this dataset. Computing the filters and the GT intensities for the MS data was done in the same ways described above, which extend naturally to 16 channels. We used a simplistic approach to demosaic the MS data by directly rearranging the raw data into a multispectral cube.

3.2.2. Evaluation Dataset

For out of distribution images, we photographed five images of the Digital ColorChecker SG with the Samsung Galaxy S21 Plus (figure 2 d), to show the reconstruction accuracy on unseen color patches. The images were taken under natural lightings indoors and outdoors. The illuminations and patches' reflectance values were measured with an X-Rite I1Pro 3 Plus spectrophotometer.

Figure 2. a-c. Examples from the BeyondRGB dataset. d. Example of an image of the Digital ColorChecker SG captured by us for out-of-distribution evaluation.

3.2.3. Simulative RGB Dataset

In this work, we show the disparity between synthetic data and real data by showing results on the ARAD 1K HS dataset from NTIRE2022 [6]. ARAD 1K, consists of 1000 HS images along with their RGB counterpart. The dataset mainly comprises outdoor scenes under natural illuminations captured with an HS camera. It covers spectral bands between 400–700 nm with a 10 nm spectral resolution, resulting in 31 spectral bands. To generate RGB-HS pairs, an RGB filter response function was applied to the 31-band HS data, producing a three-channel RGB image. To enhance realism, the simulation incorporates a noise model and applies basic image signal processing. We synthesize our own RGB data with the measured RGB camera filter response to match the RGB images we use for the rest of our experiments. In addition, we add shot noise in different scales.

3.3. Architecture and Objective Functions

The backbone of our method is the MST++ network [7], a state-of-the-art spectral reconstruction model that won the NTIRE 2022 challenge. MST++ is a Transformer-based architecture that effectively captures non-local spectral correlations. We adopt this architecture without modification to demonstrate that our contribution lies in the training framework—specifically, the ability to train on real-world RGB images with sparse supervision—rather than in architectural novelty.

Our main objective function is the supervised loss on the ColorChecker patches. In addition, we employ two regularization functions that are applied on all pixels in the image. See figure 1.We reconstruct the image with the network to obtain $\tilde{Y}$ and then re-scale the output similarly to eq. 3, this normalization is differentiable and we apply the loss functions on the rescaled reconstructed image, Y:

$$Y(x, y; \lambda) = \frac{\tilde{Y}(x, y; \lambda)||I(x, y)||_1}{||f \cdot \tilde{Y}(x, y)||_1}. \tag{5}$$

The supervised loss $\mathcal{L}_{\text{sup}}$ is applied to all pixels in the color chart patches, comparing them to the target values specified in eq. 3. For this loss, we use the spectral angle mapper (SAM), as defined in eq. 9.

The self-supervised loss guarantees that the reconstruction is physically plausible, ensuring the recovered spectra

accurately reproduce the original RGB values when downsampled using the filter response function f. Let I be the input image, and Y as in eq. 5, then the self-supervised loss is equal to:

$$\mathcal{L}_{\text{self}}(Y) = f_{loss}(I, f \cdot Y). \tag{6}$$

Where f_{loss} can be any loss, in practice we use a weighted combination of $\mathcal{L}_{\text{SAM}}$ and $\mathcal{L}_1$.

To add a per-pixel regularization in the HS domain, we devise the relighting consistency regularization. For two images of the same scene in the lab sub-dataset, the reflectance values are identical, while the illuminations differ. So for two hyperspectral images of the same scene $S_1 = R_{scene}L_1$ and $S_2 = R_{scene}L_2$ we can derive that $S_2 = \frac{S_1 \cdot L_2}{L_1}$. Our relighting consistency relies on this concept. Let Y_1 and Y_2 be the reconstruction of two images of the same scene, then:

$$\mathcal{L}_{\text{relighting}}(Y_1, Y_2) = f_{loss}\left(\frac{Y_1 \cdot L_2}{L_1}, Y_2\right). \tag{7}$$

In each iteration during training, we randomly pick pairs from the lab scenes and apply $\mathcal{L}_{\text{relighting}}$. The other losses are applied on all images in the dataset, including the lab images and natural environment images. In our ablation study (Sec. 4), when we disable the relighting consistency loss (i.e., $\gamma = 0$), the network is still trained on the exact same set of images, including the multiple illumination pairs available in the lab dataset. This ensures that any performance drop is observed solely due to the removal of the physical consistency constraint (Eq. 7) and not due to a reduction in training data diversity or quantity.

The overall loss function is a weighted combination of the above objective functions:

$$\begin{aligned}\mathcal{L}_{\text{total}} &= \alpha \cdot \mathcal{L}_{\text{sup}}(Y, H) + \beta \cdot \mathcal{L}_{\text{self}}(Y) + \gamma \cdot \mathcal{L}_{\text{relighting}}(Y_1, Y_2) \\ &= \alpha \cdot \mathcal{L}_{\text{SAM}}(Y, H) \\ &\quad + \beta_1 \cdot \mathcal{L}_1(I, fY) + \beta_2 \cdot \mathcal{L}_{\text{SAM}}(I, fY) \\ &\quad + \gamma_1 \cdot \mathcal{L}_1\left(\tfrac{Y_1 L_2}{L_1}, Y_2\right) + \gamma_2 \cdot \mathcal{L}_{\text{SAM}}\left(\tfrac{Y_1 L_2}{L_1}, Y_2\right).\end{aligned} \tag{8}$$

3.4. Implementation Details and Experiment Settings

For efficiency, we downsample input RGB images by 4. During training, we randomize patch sizes (128×128, 256×256, 512×512, and full size - 756×1008 for RGB images and 484×646 for MS images) to improve performance across various scales. During inference we input the full image. We use a batch of size 16 with the Adam optimizer (learning-rate $= 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$) for 200 epochs on an NVIDIA RTX H100 GPU. Loss function weights are set empirically to $\alpha = 1$, $\beta_1 = 25$, $\beta_2 = 0.75$, $\gamma_1 = 25$ and $\gamma_2 = 0.3$.

We use an 80/10/10 train/validation/test split on the BeyondRGB dataset. To demonstrate the network's ability to reconstruct colors without overfitting to specific color chart patches or illuminations we devise two split strategies: (1) illumination based split. We split the lab images based on the illuminant, such that each sub-dataset contains unique illuminations. Crucially, while the scenes in the lab are static, splitting by unique illuminations ensures the network cannot simply memorize the spectral response of a specific scene under a specific light, testing its ability to generalize to new lighting conditions. The natural scenes are split randomly, ensuring no overlap between scenes in train and test sets. (2) Color chart patches split. Due to the limited number of color patches (18) in the ColorChecker, a standard 80/10/10 split was not feasible. Instead, we randomly mask out three patches from the training set and evaluate performance on these three patches across all images in the dataset. Hyper-parameters were selected using the validation set from the illumination-based split. All the reported results are on the illumination test set, except when explicitly stated otherwise.

4. Experiments & Results

We evaluate accuracy and consistency using four measures:

X-Rite ColorChecker Classic. For regions with GT signals (color chart areas), we compare the HS reconstruction to the derived HS targets (eq. 3), averaging across all patch pixels in test images. Only one gray patch per chart is included since gray patches under the same illuminant differ only in scale.

Digital ColorChecker SG. We compare the reconstruction with the target values on our evaluation data (Digital ColorChecker SG), including only one gray patch per chart.

Relighting Consistency. For lab images, we measure the relighting consistency in the HS domain (eq. 7) between randomly paired images of the same scene under different illuminations.

RGB Reconstruction. To assess physical truthfulness, we project outputs back to the low-dimensional spectral space and compare with the input (eq. 6).

Our primary evaluation metric is the Spectral Angle Mapper (SAM) defined in eq. 9, a widely used measure in spectral analysis. SAM is often preferred over other distance measures for its insensitivity to signal intensity variations and focus on spectral signature shape. In addition, we report the commonly used Mean Relative Absolute Error (MRAE) and Root Mean Square Error (RMSE). We apply these metrics after the rescaling mentioned in eq. 5. Let Y and H be the values we want to compare, and N the number of pixels we average over, then:

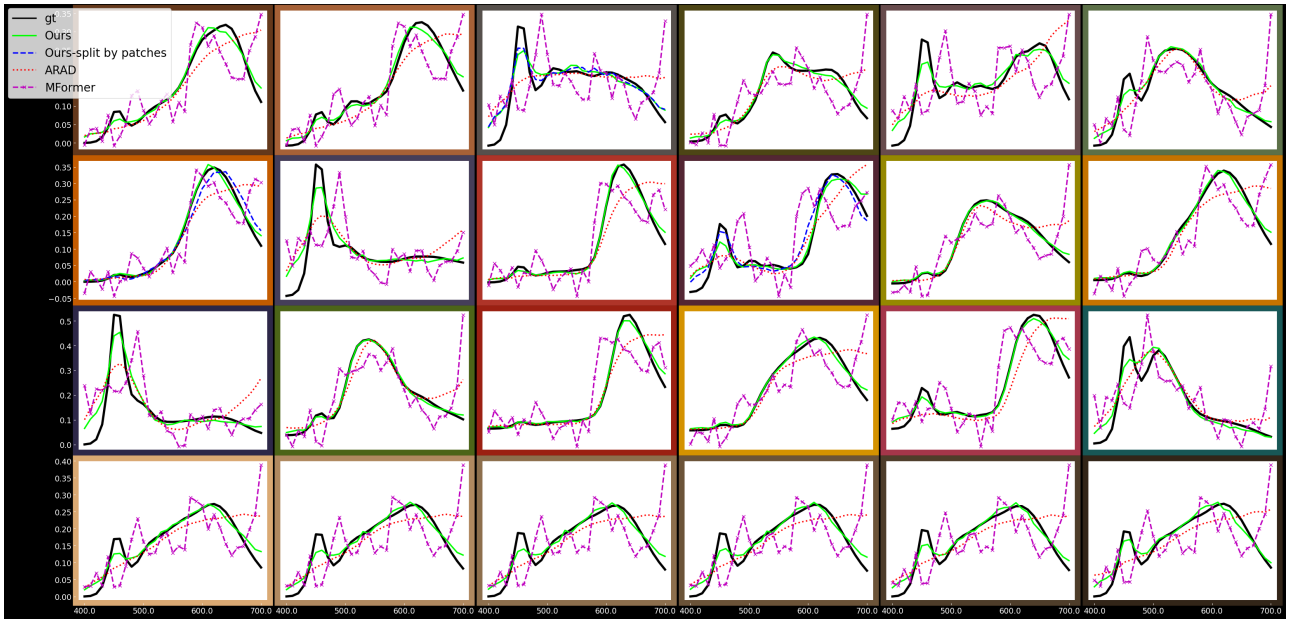

Figure 3. Reconstruction on the X-Rite ColorChecker Classic. The values shown are the mean of the reconstruction on the patch from an indoor image. For the network trained with the patches split we show results on the patches from the test set: (0, 2), (1, 0), (1, 4).

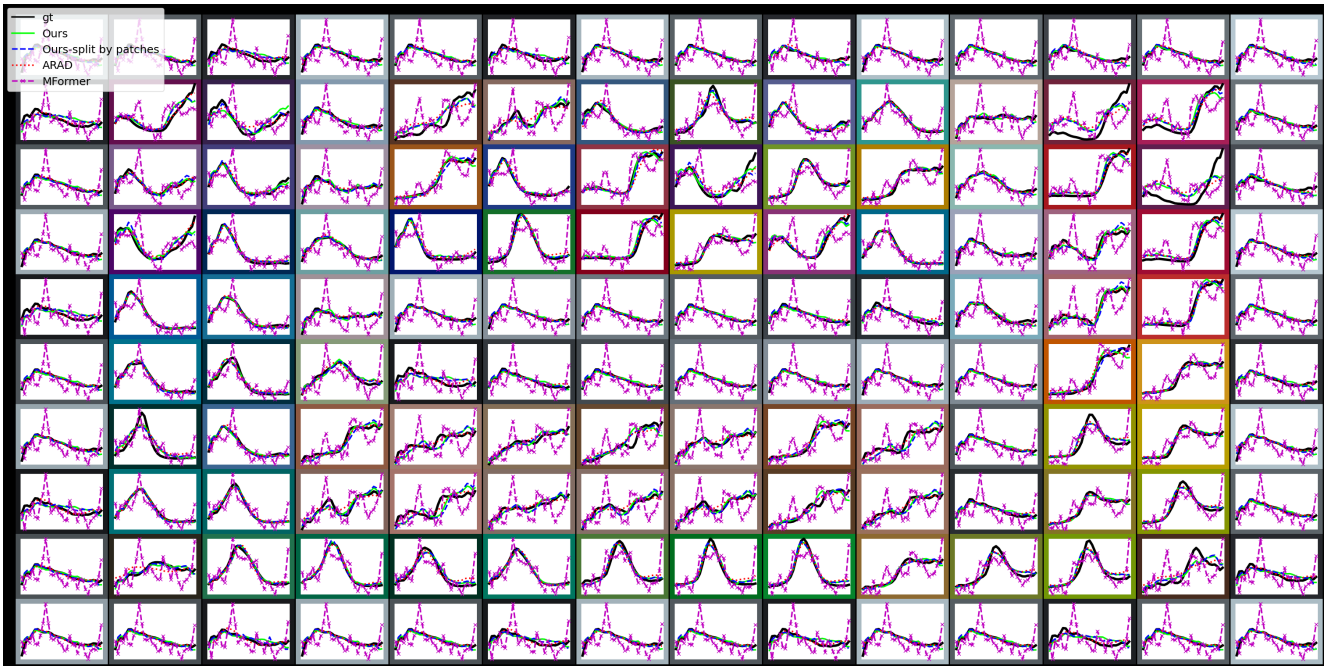

Figure 4. Reconstruction on an example Digital ColorChecker SG image. The values shown are the mean of the reconstruction on the patch.

$$\text{SAM} = \frac{1}{N}\sum_{i=1}^{N}\arccos\left(\frac{Y[i]\cdot H[i]}{\|Y[i]\|\,\|H[i]\|}\right),$$

$$\text{MRAE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|Y[i]-H[i]|}{H[i]},$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\big(Y[i]-H[i]\big)^2}. \qquad (9)$$

4.1. Main Results

Numerical results can be seen in table 1. We show results with the two data splits, results on MS data and for comparison, results on the MST++ network trained in a supervised manner on the ARAD 1K dataset. We also report the results of the unsupervised network MFormer trained in an

Method	X-Rite ColorChecker Classic			Relighting Consistency			RGB Reconstruction			Digital ColorChecker SG		
	SAM	MRAE	RMSE	SAM	MRAE	RMSE	SAM	MRAE	RMSE	SAM	MRAE	RMSE
ARAD (sup.)	18.9	0.59	0.04	31.12	0.86	0.1	1.39	0.03	0.002	12.98	0.35	0.39
MFormer (unsup.)	29.7	1.02	0.06	36.36	0.47	0.13	0.11	0.003	0.0002	27.03	0.81	0.08
Ours - illum. split	5.52	0.15	0.01	7.16	1.3	0.03	0.97	0.02	0.002	7.36	0.19	0.02
Ours - MS data	2.33	0.07	0.008	7.27	1.15	0.02	4.95	0.07	0.03	-	-	-
Ours - patch split	8.82	0.18	0.015	5.99	1.26	0.03	0.94	0.02	0.002	9.03	0.23	0.025

Table 1. Results on BeyondRGB and our SG test set. For all metrics, lower is better. We highlight the best RGB to HS method on the test set in the illumination split under the SAM metric, our scaleless and main evaluation method.

Loss Components			X-Rite ColorChecker Classic			Relighting Consistency			RGB Reconstruction			Digital ColorChecker SG		
Supervised	Relighting	Self-Loop	SAM	MRAE	RMSE	SAM	MRAE	RMSE	SAM	MRAE	RMSE	SAM	MRAE	RMSE
✓			6.0	0.17	0.01	15.21	1.82	0.03	2.4	0.06	0.005	8.89	0.23	0.03
✓	✓		6.45	0.17	0.02	1.16	0.38	0.01	23.23	0.55	0.05	8.88	0.23	0.026
✓		✓	6.6	0.18	0.01	15.32	1.88	0.03	0.84	0.02	0.001	7.39	0.19	0.02
✓	✓	✓	5.52	0.15	0.01	7.16	1.3	0.03	0.97	0.02	0.002	7.36	0.19	0.02

Table 2. Results when utilizing the different components in our method.

unsupervised manner on the BeyondRGB dataset.

Comparison Strategy: Our goal is to highlight the effectiveness of training on real-world data versus the standard practice of training on synthetic data. Therefore, we compare our method (Real Data + Sparse Supervision) against two established baselines: (1) ARAD (Supervised): Represents the current state-of-the-art for supervised learning, but trained on synthetic data (as is standard). This comparison highlights the domain gap. (2) MFormer (Unsupervised): Represents the best alternative when no dense spectral labels are available, trained on the same real-world data. This comparison highlights the value of our sparse supervision and consistency losses. While ARAD uses a different training paradigm, it serves as the industry standard baseline we aim to surpass by shifting the data focus.

For synthetic data from ARAD 1K, we introduced varying shot noise levels to better model real-world conditions in BeyondRGB. However, noise intensity adjustments had minimal impact on performance, indicating these simulations don't fully capture real photograph characteristics. Results shown use NPE=100 (number of photo-electrons), which yielded marginally better performance. Complete noise level results are available in the supplementary material. We used the published NTIRE 2022 challenge train-validation split.

The error on the ARAD 1K validation set was $3.83°$. However, as seen in table 1, the results on the X-Rite ColorChecker Classic are fairly high. This may suggest a domain gap arising from two possible causes: first, a mismatch between the colors in the training set and those in the BeyondRGB test set, highlighting a limitation of HS datasets due to their restricted size and content. Second, the synthetic RGB data may not accurately capture the camera's characteristics. From the results on the unsupervised method, it is clear that having some pixels with supervised ground truth provides an advantage over purely unsupervised methods.

As expected, the method trained on MS data performs best on the X-Rite ColorChecker Classic, as mapping from 16 channels to 31 channels is less complex than mapping from 3 to 31 channels.

All modalities perform well on the RGB Reconstruction measure, with the unsupervised method yielding the best results, as it was primarily trained with this loss. The strong performance across all modalities can be attributed to the fact that many spectral values in the high-dimensional space map to the same RGB values in the lower-dimensional space. The MS data, however, shows significantly higher error because its higher-dimensional space (with 16 channels) is more sensitive to small differences, leading to a larger angular error compared to lower-dimensional data like RGB, which has only 3 channels.

We report results on the Digital ColorChecker SG test set using all available RGB networks: the model trained with the illumination split, the model trained with the patch split, the model trained on synthetic data, and the unsupervised model. Our networks achieve substantially better performance than both existing supervised and unsupervised methods. Furthermore, their performance is only slightly below the obtained results on the X-Rite ColorChecker Classic, demonstrating the method's ability to generalize to unseen color patches.

Visual results are presented in Figures 3, 4 and 5. We showcase results from a sample image from the BeyondRGB dataset and an image from the SG color chart dataset. More samples from BeyondRGB are included in the supplementary material. The displayed results correspond to our network with the illumination split, our network with the patches split, MST++ trained in a supervised manner on the ARAD 1K dataset, and the MFormer network trained in an unsupervised manner.

Figure 5. Depiction of the relighting consistency: we input an image to the four networks, relight the prediction, project back to camera RGB and white balance with the new illumination. On the left is the white balanced image under the new illumination. The last column is the RGB value of the projected reconstructions in the marked rectangle in the image.

Figure 3 illustrates the hyperspectral reconstruction for color chart patches in an indoor image. The spectrum in each patch represents the mean of the reconstructed patch, compared to the GT, with all values normalized to have a unit norm. These graphs clearly show that our reconstruction closely resembles the shape of the GT. For our network trained with the patches split we only show results on the patches from the test set, in positions: (0, 2), (1, 0) and (1, 4). We can observe that the reconstruction of completely unseen patches by this method closely matches the GT, highlighting the strong generalization capability of our approach.

In figure 4 visual results of the reconstruction on our evaluation dataset of the Digital ColorChecker SG are shown. Most of the patches in the Digital ColorChecker SG aren't included in the X-Rite ColorChecker Classic where there was a supervised loss in our method. Despite this, the visual results show that the shapes of the reconstructions from our two networks are very similar to the GT signals, and our networks performs better than the supervised and unsupervised methods.

Figure 5 depicts the relighting consistency. We take a pair of images I_1 and I_2 of the same scene, pass I_1 through the spectral reconstruction model to generate Y_1, and relight the predictions to L_2. Then we project the relighted prediction to the camera RGB space and white balance it according to L_2. The first image in each row is I_2 after white balance. On the right column we can see a zoom-in of the RGB value in the marked pixel. The relighting error impacts the projection, resulting in noticeable visual differences between the GT RGB values and those generated by both the supervised and unsupervised methods.

4.2. Ablation Study

We validate the importance of each part of the objective function (Eq. 8) in Table 2. The results show that the reduction in error on the X-Rite ColorChecker Classic is not achieved solely through supervised signals, but rather through the combination of all components, emphasizing their collective contribution to performance. Each regularization term achieves its best performance when used together with the supervised signal. However, disabling one regularization term causes a significant deterioration to that loss. It is important to note that relighting is an additional training signal, and all rows in the table are trained on the same data. Finally, removing the self-loop component has the strongest negative impact on the Digital ColorChecker SG, comprising both indoor and outdoor lighting, suggesting that this component plays a critical role in improving generalization.

5. Conclusion and Limitations

In this paper, we demonstrate that RGB data generated synthetically from HS images for spectral reconstruction does not generalize effectively to real camera RGB data. This finding underscores the limitations of relying solely on synthetic datasets for training and evaluating models intended for real-world spectral reconstruction.

To address this gap, we propose a novel approach utilizing regions in images where reflectance values are known under controlled, measured illuminations, such as color charts. Our deep learning model, equipped with a task-specific loss, improves real-world spectral prediction. We also introduce new metrics that assess accuracy where ground truth is known, while ensuring plausible outputs elsewhere.

Limitations: Our method currently relies on the presence of a color chart and measured illumination during training, which limits data collection to scenes where these can be inserted. However, this is significantly more accessible than capturing full hyperspectral ground truth. Additionally, our evaluation focuses heavily on color chart accuracy due to the lack of dense spectral GT for real scenes. We mitigate this by evaluating on the out-of-distribution Digital

ColorChecker SG, showing generalization beyond the training charts. Future work could explore cross-camera generalization, as our current experiments are limited to a single sensor.

Despite these limitations, our method has the potential to be used as a foundation model in the future. This would enhance downstream tasks by offering robust pretrained representations adaptable to a variety of spectral reconstruction challenges. Expanding datasets to cover broader reflectance and lighting conditions will also be essential for the advancement of real-world spectral reconstruction.

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 Workshops*, pages 471–479, 2017. 3

[2] Naveed Akhtar and Ajmal Mian. Hyperspectral recovery from rgb images using gaussian processes. *IEEE transactions on pattern analysis and machine intelligence*, 42(1): 100–113, 2018. 3

[3] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, the Netherlands, October 11–14, 2016, Proceedings, Part VII 14*, pages 19–34. Springer, 2016. 3

[4] Boaz Arad, Radu Timofte, Ohad Ben-Shahar, Luc Van Gool, Lei Zhang, and et al Yang, Ming-Hsuan. Ntire 2018 challenge on spectral reconstruction from rgb images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2018. 1, 3

[5] Boaz Arad, Radu Timofte, Ohad Ben-Shahar, Yi-Tun Lin, and Graham D Finlayson. Ntire 2020 challenge on spectral reconstruction from an rgb image. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 446–447, 2020.

[6] Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, et al. Ntire 2022 spectral recovery challenge and data set. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 863–881, 2022. 1, 3, 4

[7] Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, Radu Timofte, and Luc Van Gool. Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 745–755, 2022. 2, 3, 4

[8] Xuheng Cao, Yusheng Lian, Zilong Liu, Jin Li, and Kaixuan Wang. Unsupervised spectral reconstruction from rgb images under two lighting conditions. *Optics Letters*, 49(8): 1993–1996, 2024. 1, 3

[9] Ayan Chakrabarti and Todd Zickler. Statistics of real-world hyperspectral images. In *CVPR 2011*, pages 193–200. IEEE, 2011. 3

[10] Puhong Duan, Tianci Shan, Xudong Kang, and Shutao Li. Spectral super-resolution in frequency domain. *IEEE Transactions on Neural Networks and Learning Systems*, 2024. 3

[11] Abdelhamid N Fsian, Jean-Baptiste Thomas, Jon Y Hardeberg, and Pierre Gouton. Spectral reconstruction from rgb imagery: A potential option for infinite spectral data? *Sensors*, 24(11):3666, 2024. 1

[12] Biebele Joslyn Fubara, Mohamed Sedky, and David Dyke. Rgb to spectral reconstruction via learned basis functions and weights. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 480–481, 2020. 3

[13] Yunhao Geng, Shaohui Mei, Jin Tian, Yifan Zhang, and Qian Du. Spatial constrained hyperspectral reconstruction from rgb inputs using dictionary representation. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 3169–3172. IEEE, 2019. 3

[14] Ortal Glatt, Yotam Ater, Woo-Shik Kim, Shira Werman, Oded Berby, Yael Zini, Shay Zelinger, Sangyoon Lee, Heejin Choi, and Evgeny Soloveichik. Beyond rgb: a real world dataset for multispectral imaging in mobile devices. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 4344–4354, 2024. 2, 3

[15] Jaesoong Lee, Yeonsang Park, Hyochul Kim, Young-Zoon Yoon, Woong Ko, Kideock Bae, Jeong-Yub Lee, Hyuck Choo, and Young-Geun Roh. Compact meta-spectral image sensor for mobile applications. *Nanophotonics*, 11(11): 2563–2569, 2022. 4

[16] Jiaojiao Li, Chaoxiong Wu, Rui Song, Yunsong Li, and Fei Liu. Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from rgb images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 462–463, 2020. 3

[17] Jiaojiao Li, Chaoxiong Wu, Rui Song, Yunsong Li, Weiying Xie, Lihuo He, and Xinbo Gao. Deep hybrid 2-d–3-d cnn based on dual second-order attention with camera spectral sensitivity prior for spectral super-resolution. *IEEE Transactions on Neural Networks and Learning Systems*, 34(2): 623–634, 2021. 3

[18] Jiaojiao Li, Songcheng Du, Chaoxiong Wu, Yihong Leng, Rui Song, and Yunsong Li. Drcr net: Dense residual channel re-calibration network with non-local purification for spectral super resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1259–1268, 2022. 3

[19] Jiaojiao Li, Yihong Leng, Rui Song, Wei Liu, Yunsong Li, and Qian Du. Mformer: Taming masked transformer for unsupervised spectral reconstruction. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–12, 2023. 3

[20] Jinxing Liang, Xin Hu, Zhuan Zuo, Xiao Liu, Yifan Li, Wensen Zhou, Hang Luo, Xinrong Hu, and Kaida Xiao. Exploring multispectral reconstruction based on camera response prediction. In *Proceedings of the 12th Colour and Visual Computing Symposium (CVCS 2024) Gjøvik, Norway, September 5-6, 2024., Gjøvik, Norway, September 5-6, 2024*. CEUR-WS.org, 2024. 3

[21] Yi-Tun Lin and Graham D Finlayson. Physically plausible spectral reconstruction from rgb images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 532–533, 2020. 3

[22] Zhen Liu, Kaida Xiao, Michael R Pointer, Qiang Liu, Changjun Li, Ruili He, and Xuejun Xie. Spectral reconstruction using an iteratively reweighted regulated model from two illumination camera responses. *Sensors*, 21(23):7911, 2021. 3

[23] Rang MH Nguyen, Dilip K Prasad, and Michael S Brown. Training-based spectral reconstruction from a single rgb image. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VII 13*, pages 186–201. Springer, 2014. 3

[24] Young-Geun Roh, Hyochul Kim, Wooshik Kim, Jaesoong Lee, Suyeon Lee, Youngho Jung, Heejin Choi, and Sangyoon Lee. Spectral image sensing in the smartphone and its applications. In *Image Sensing Technologies: Materials, Devices, Systems, and Applications X*, page PC1251408. SPIE, 2023. 4

[25] Zhan Shi, Chang Chen, Zhiwei Xiong, Dong Liu, and Feng Wu. Hscnn+: Advanced cnn-based hyperspectral recovery from rgb images. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1052–10528, 2018. 3

[26] Abhishek Kumar Sinha and Manthira Moorthi S. Lipsspecformer: Non-linear interpolable transformer for spectral reconstruction using adjacent channel coupling. In *34th British Machine Vision Conference 2023, BMVC 2023, Aberdeen, UK, November 20-24, 2023*. BMVA, 2023. 3

[27] 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. 3

[28] Nan Wang, Shaohui Mei, Yi Wang, Yifan Zhang, and Duo Zhan. Whanet: Wavelet-based hybrid asymmetric network for spectral super-resolution from rgb inputs. *IEEE Transactions on Multimedia*, 2024. 3

[29] Xinying Wang, Zhixiong Huang, Sifan Zhang, Jiawen Zhu, Paolo Gamba, and Lin Feng. Gmsr: gradient-guided mamba for spectral reconstruction from rgb images. *arXiv preprint arXiv:2405.07777*, 2024. 3

[30] Chaoxiong Wu, Jiaojiao Li, Rui Song, Yunsong Li, and Qian Du. Repcpsi: Coordinate-preserving proximity spectral interaction network with reparameterization for lightweight spectral super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–13, 2023. 3

[31] Zhiwei Xiong, Zhan Shi, Huiqun Li, Lizhi Wang, Dong Liu, and Feng Wu. Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 518–525, 2017. 3

[32] Ruikang Xu, Mingde Yao, Chang Chen, Lizhi Wang, and Zhiwei Xiong. Continuous spectral reconstruction from rgb images via implicit neural representation. In *European Conference on Computer Vision*, pages 78–94. Springer, 2022. 3

[33] Fumihito Yasuma, Tomoo Mitsunaga, Daisuke Iso, and Shree K Nayar. Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum. *IEEE transactions on image processing*, 19(9):2241–2253, 2010. 3

[34] Junchao Zhang, Dangjun Zhao, Jianlai Chen, Yuanyuan Sun, Degui Yang, and Rongguang Liang. Unsupervised learning for hyperspectral recovery based on a single rgb image. *Optics letters*, 46(16):3977–3980, 2021. 3

[35] Kai Zhang, Tian Jin, Feng Zhang, and Jiande Sun. Long-short attention network for the spectral super-resolution of multispectral images. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2023. 3

[36] Lei Zhang, Zhiqiang Lang, Peng Wang, Wei Wei, Shengcai Liao, Ling Shao, and Yanning Zhang. Pixel-aware deep function-mixture network for spectral super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 12821–12828, 2020. 3

[37] Xiandou Zhang, Qiang Wang, Jincheng Li, Xiaohui Zhou, Yuechuan Yang, and Haisong Xu. Estimating spectral reflectance from camera responses based on cie xyz tristimulus values under multi-illuminants. *Color Research & Application*, 42(1):68–77, 2017. 3

[38] Zhiyu Zhu, Hui Liu, Junhui Hou, Huanqiang Zeng, and Qingfu Zhang. Semantic-embedded unsupervised spectral reconstruction from single rgb images in the wild. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2279–2288, 2021. 3