Limitations of Data-Driven Spectral Reconstruction: An Optics-Aware Analysis

Qiang Fu*, Matheus Souza*, Eunsue Choi, Suhyun Shin, Seung-Hwan Baek, Wolfgang Heidrich *Fellow, IEEE*

***Abstract*—Hyperspectral imaging empowers machine vision systems with the distinct capability of identifying materials through recording their spectral signatures. Recent efforts in data-driven spectral reconstruction aim at extracting spectral information from RGB images captured by cost-effective RGB cameras, instead of dedicated hardware. Published work reports exceedingly high numerical scores for this reconstruction task, yet real-world performance lags substantially behind.**

In this paper we *systematically analyze* the performance of such methods with three groups of dedicated experiments. First, we evaluate the practical overfitting limitations with respect to current datasets by training the networks with less data, validating the trained models with unseen yet slightly modified data, and cross-dataset validation. Second, we reveal *fundamental limitations* in the ability of RGB to spectral methods to deal with metameric or near-metameric conditions, which have so far gone largely unnoticed due to the insufficiencies of existing datasets. We achieve this by validating the trained models with metamer data generated by metameric black theory and re-training the networks with various forms of metamers. This methodology can also be used for data augmentation as a partial mitigation of the dataset issues, *although the RGB to spectral inverse problem remains fundamentally ill-posed*.

Finally, we analyze the potential for modifying the problem setting to achieve better performance by exploiting some form of optical encoding provided by either incidental optical aberrations or some form of deliberate optical design. Our experiments show that such approaches do indeed provide improved results under certain circumstances, however their overall performance is limited by the same dataset issues as in the plain RGB to spectral scenario. We therefore conclude that *future progress on snapshot spectral imaging will heavily depend on the generation of improved datasets which can then be used to design effective optical encoding strategies*. Code can be found at https://github.com/vccimaging/OpticsAwareHSI-Analysis.

***Index Terms*—Hyperspectral imaging, Spectral reconstruction from RGB, Metamerism, Overfitting, Aberration.**

I. Introduction

Hyperspectral imaging is a method that involves recording the light in a scene in the form of many, relatively narrow, spectral bands, rather than projected into three broadband RGB color channels. Where RGB imaging utilizes the trichromaticity theory of human color vision, spectral imaging provides additional information that can help discriminate between different materials and lighting conditions that are hard to tell apart in RGB images. For example, red stains in a crime scene could be blood, or paint, or a dyed cloth, which cannot be distinguished from their RGB colors. Skin tumors could not be diagnosed from surrounding tissues of the same color. It is difficult to spot and sort out plastic leaves from living plants by their greenish colors. Therefore, spectral imaging has been applied in many fields, including computer graphics [49], machine vision [34], [65], healthcare [57], agriculture [28], and environment [11], to name just a few.

However, conventional hyperspectral cameras require scanning mechanisms [43], [56] to acquire the 3D hyperspectral datacube with 2D sensors. To simplify the demanding hardware, extensive efforts have been made in the development of various snapshot hyperspectral cameras [47], [67], [79].

On the extreme end of these hardware simplification efforts, deep learning methods have emerged in recent years that attempt to solve the problem entirely in software by reconstructing spectral data from RGB images (RGB2HS). This has resulted in three CVPR-hosted NTIRE challenges [5]–[7] and various network architectures [15], [52], [68], [102], [105]. Yet it remains unclear how these methods generalize to unseen data, how they deal with the difficult but important problem of resolving metamerism [40], [61], and how they depend on the optical system of both the RGB source and the spectral cameras used to capture the datasets.

At its core, estimating spectral information from RGB colors is an under-determined one-to-many mapping problem. As stated By Pharr et al. [63] (Ch. 4), "any such conversion is inherently ambiguous due to the existence of metamers". Intuition would therefore suggest that the achievable spectral fidelity of RGB to spectral methods is limited. However, this intuition flies in the face of very high numerical test results reported in recent NTIRE challenges [5]–[7].

One possible explanation for the experimental success of spectral reconstruction from an RGB image is that the networks learn to exploit spatial structures, or scene semantics, to estimate spectral information. However, spectral images are usually employed when RGB images do not provide sufficient information for downstream tasks. Therefore, it is questionable to use scene semantics to resolve spectral ambiguities. Instead, in computational imaging, the idea is to *measure* the spectral information to help better understand the scene semantics, particularly in difficult scenarios. Clearly, these two methodologies feature reversed information flow. We argue that the computational imaging approach is more compatible with spectral imaging itself.

In practice, metameric or near-metameric colors often occur in situations where the spatial structure is also similar, *e.g.*, vein finding (the global geometry is a segment of a forearm, but vein structure is unknown), or biometrics (*e.g.*, distin-

Q. Fu, M. Souza, and W. Heidrich are with King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia. E. Choi, S. Shin, and S. Baek are with Pohang University of Science and Technology (POSTECH), Pohang, Korea. Corresponding author: Q. Fu (qiang.fu@kaust.edu.sa). This paper has supplementary downloadable material provided by the author. The material includes additional experimental results to provide further evidence for the major findings in this work. This material is 14.2 MB in size.

* Joint first authors.

guishing real faces from masks or images). To illustrate this effect, we show an example from the smaller and older CAVE dataset [96] where several fake and real objects are presented in two groups, as shown in Fig. 1. Two sample points from the two red peppers (one real and the other fake, but it's unclear which is which) have the RGB values (86, 21, 10) and (86, 21, 8), respectively. As can be seen, the real spectra of the real and fake red peppers differ substantially in the red part of the spectrum. A pre-trained model (arbitrarily chosen from previous NTIRE challenges trained on the much larger and newer ARAD1K dataset [7]) predicts spectra that differ from the ground truth, but more importantly, the predicted spectra for the real and fake pepper are almost identical, illustrating the failure to adequately deal with metamers.

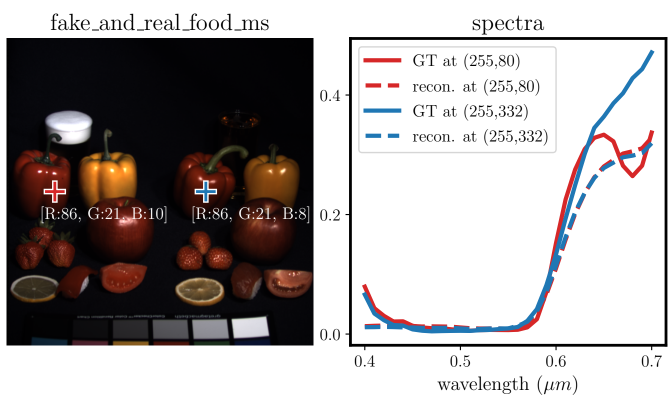

Fig. 1. An example scene fake_and_real_food_ms from the CAVE dataset [96] consists of objects with visually similar colors, but actually different spectra. Left: Color image with highlighted points on the red peppers. Their RGB values are nearly the same. Right: Ground-truth and reconstructed spectra at the corresponding points show their spectral differences. The reconstructed spectra are predicted by the pre-trained MST++ model [15] on the ARAD1K dataset [7]. The neural network struggles to distinguish either the two spectra from each other, or from their true spectra.

In this paper, we take a systematic and outside-the-box look at all the above aspects. To the best of our knowledge, we are the first to analyze, document, and discuss the inherent shortcomings of this research theme. We highlight realistic conditions under which recent efforts fall short, aiming to constructively instigate, debate, provide insights, and forge a new path regarding the physical phenomena that have been overlooked. By conducting a series of adversarial attacks and thorough analysis, we reveal a number of shortcomings in both current datasets and reconstruction methods. Specifically, we find that:

- Existing hyperspectral image datasets severely lack in diversity especially with respect to metameric colors but also other factors including nuisance parameters such as noise and compression ratios.
- State-of-the-art methods suffer from *atypical* overfitting problems that arise from various factors in the image simulation pipeline, such as noise, RGB data format, and lack of optical aberrations.
- Optical aberrations in RGB images, while currently ignored by all methods, are actually *beneficial* rather than harmful to spectral reconstruction *if modeled accurately*.
- Crucially, the limitations of the datasets that we document not only affect the RGB to spectral work, but also any other spectral reconstruction and processing that uses the same training data [8], [45], [90]. We show how *metameric augmentation* can be used to at least partially overcome the dataset issues.

A seemingly apparent observation from the results we show in this paper reinforces that it is impossible to distinguish metamers solely from RGB colors. Remarkably, this fundamental limitation has been largely overlooked within the research in this field. Through the evidence in this work, we contribute to a deepened understanding of the limitations of current datasets as well as of underlying sources that result in the limitations of spectral reconstruction accuracy. The results of the interplay between metameric spectra and optical aberrations open the door for new approaches for spectral recovery down the road.

II. Related Work

A. Hyperspectral cameras

Conventional hyperspectral imaging systems require filter wheels, liquid-crystal tunable filters, or mechanical motion (*e.g.*, pushbroom) [43], [56] to scan the 3D hyperspectral datacube. To enable snapshot acquisition, coded-aperture snapshot spectral imager (CASSI) [25], [79] has been proposed to achieve high spectral accuracy using spectrum-dependent coded patterns. Based on this hardware architecture, supervised learning (such as TSA-Net [59], BiSRNet [16], VmambaSCI [103], SpeCAT [95]) and unsupervised learning (such as LRSDN [23], SAH-SCI [100], CEINR [58]) algorithms have been proposed in recent years to address the inverse problem on hyperspectral datasets. Other variants, such as dual-camera CASSI [82] and reconstruction algorithms have also been proposed. Another category of emerging methods also exploit spectrally encoded point spread functions (PSFs) to computationally reconstruct a hyperspectral image [10], [20], [47]. Various DOE designs for optimal spectral PSFs, such as equalization DOE [92], non-serial quantization-aware deep optics [81], tunable phase encoding [101], Double-DOE [73], have been proposed along with corresponding reconstruction algorithms over the past few years. In general, great efforts have been made to simplify hyperspectral camera hardware by software reconstruction.

B. Spectral reconstruction from RGB images

A recent trend to solve the snapshot hyperspectral imaging problem is to exploit hyperspectral data with deep neural networks to reconstruct spectral information from RGB images [4]. Owing to the wide availability of RGB cameras, this approach seems to be a promising candidate for hyperspectral imaging if successful. A large number of neural network architectures have been proposed in the past three NTIRE spectral recovery challenges [5]–[7] and other venues afterwards. Our analysis in this paper focuses on this class of methods to gain insights on their strengths and limitations. In particular, we comprehensively evaluate 17 open-sourced neural networks to date. HSCNN+ [68] is one of the first networks that employs CNN as the backbone for spectral reconstruction which won the first challenge in 2018 [5]. It features dense residual blocks (HCNN-R) and densely-connected structures (HCNN-D). EDSR [55] introduces an enhanced deep super-resolution network by removing unnecessary modules in conventional residual networks. It was originally designed for single-image super-resolution, and later extended for spectral reconstruction. HRNet [105] employs a Hierarchical Regression Network that consists of 4 levels followed by PixelShuffle layers for inter-level interaction, followed by a residual dense block and a residual global block to reconstruct the hyperspectral images. AWAN [52] utilizes a backbone stacked with multiple dual residual attention blocks for dual residual learning. The sensor response function is used as a finer constraint to improve the reconstruction quality. MIRNet [98] adopts a multi-scale residual block that learns contextual information from multiple scales to enhance spatial resolution for image restoration tasks, and later extended for spectral reconstruction. HINet [22] proposes a Half Instance Normalization Block that was originally designed to boost image restoration networks, and later extended for spectral reconstruction. MPRNet [99] is a multi-stage network to learn the projection from degraded measurements to the high-quality images, with a couple of manageable steps. It can also be extended to hyperspectral reconstruction with proper modifications. HDNet [42] proposes a dual domain learning network with a spatial-spectral attention module for pixel-level features, and a frequency domain learning to narrow the frequency domain discrepancy. Restormer [97] was also designed for image restoration tasks initially and then extended for spectral reconstruction, with a focus on designing the building block with the Transformer architecture to capture long-range pixel interactions. MST [14] and MST++ [15] are Transformer-based networks that employ spectral-wise multi-head self-attention to fully make use of spatial sparsity and spectral self-similarity for efficient spectral reconstruction in a coarse-to-fine manner. HySAT [80] employs an exhaustive correlation Transformer to simultaneously model spectral-wise similarity with a token-independent mapping mechanism and particularity with a spectral-wise re-calibration mechanism. HRPN [86] integrates comprehensive multisource priors, in particular the semantic prior of RGB inputs, to regularize and optimize the solution space with a Transformer-based holistic prior-embedded relation network. SSRNet [31] employs a model-guided network based on cross fusion that uses the image formation model and the sensor spectral response function to guide the training of a CNN-backed network. SSTHyper [91] introduces a sparse spectral transformer model to learn shallow and deep spatial-spectral priors and allows adaptive masking of non-significant details. Computational cost is reduced by a cross-level fusion network architecture. MSFN [87] is a multi-stage UNet structure that captures both spatial and spectral features in a multiscale manner. A feature alignment scheme is proposed to preserve spatial correlations and spectral self-similarities. GMSR [83] builds on a more recently developed architecture Mamba [39] as the backbone and develops a lightweight model for global feature representation. Spatial gradient attention and spectral gradient attention are proposed to improve the spectral reconstruction.

C. Multispectral and hyperspectral image fusion

Another class of related spectral reconstruction methods is to fuse images with low spatial resolution but high spectral resolution with images that have high spatial resolution but low spectral resolution [30], [78]. Different from spectral reconstruction from RGB images, it requires two inputs and a final image with high spatial and spectral resolution is obtained. In recent years, neural networks have also been extensively employed in solving the image fusion problem. Similar datasets as well as remote sensing images are usually used to evaluate the performance. CNN and Transformer are popular backbones in the design of such networks [17], [19], [54], [89], [93]. In this paper, however, we focus on data-driven methods for hyperspectral recovery from single RGB images, where the limitations arising from existing datasets are more prominent.

D. Dataset bias and data augmentation

Deep neural networks are prone to suffer from data bias [32], [72] and overfitting problems [12]. Overfitting can lead to the inability of trained models to generalize in real-world applications [50]. Although overfitting can sometimes be detected by inspecting the training and validation performance over the course of training, it can often be imperceivable in challenging problems. A useful technique to detect overfitting is to use adversarial examples [85] generated from the original dataset. On the other hand, it is important to address overfitting when the amount of data is limited. Data augmentation [66], [69] techniques are usually employed to improve the robustness of deep neural networks.

E. Metamerism

Metamerism is a physical phenomenon where distinct spectra produce the same color [2] as the high-dimensional spectral space is projected down to three dimensions of a trichromatic vision system (either the human eye or an RGB camera). This phenomenon has been studied in color science [37], [71], spectral rendering [46], [74], [84], and hyperspectral imaging [36], [40]. In hyperspectral imaging, it is crucial in many applications to distinguish between metamers or near-metamers (*i.e.*, different spectra that project to *similar* RGB values) [27], [61]. Indeed spectral imaging is usually employed when the RGB color differences between two materials or features are too small to reliably distinguish between them. Therefore, hyperspectral imaging systems require special attention in the system design to acquire accurate spectral signatures [41], [48].

However, modeling metamerism is challenging in data-driven spectral reconstruction as metamerism is hard to capture since they are relatively rare in everyday environments [37], although they are vital for many applications of spectral imaging. Previous works make use of illumination and camera spectral response as means of providing additional information to help improve the spectral reconstruction. For example,

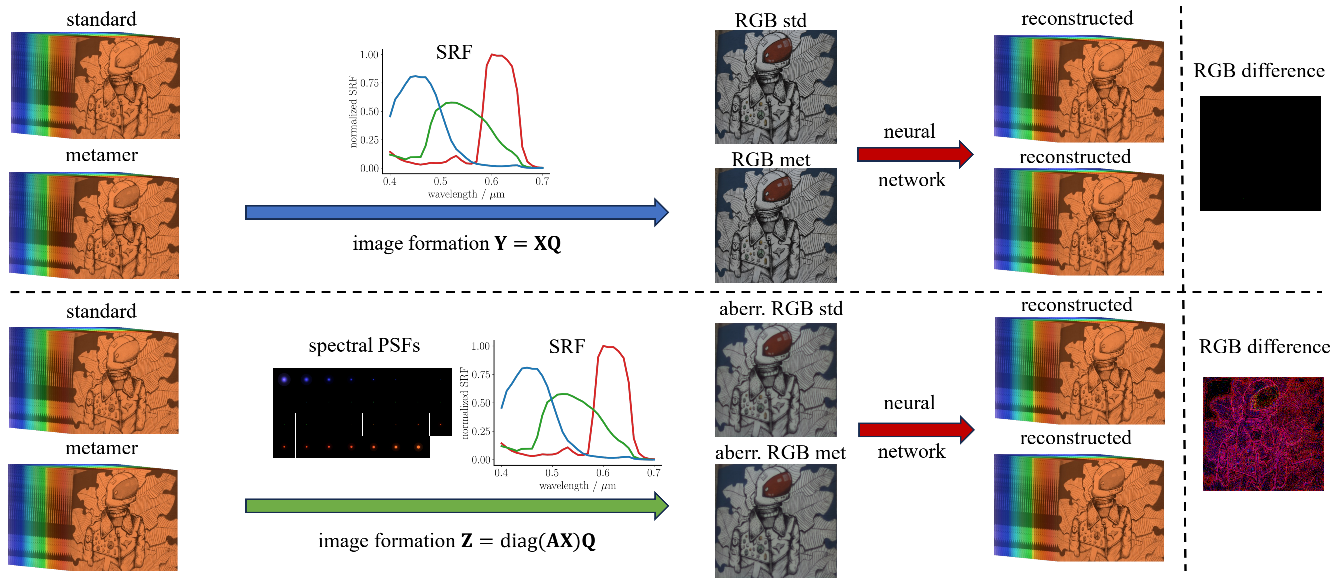

Fig. 2. Spectral image formation models used in the analysis in this work. Top: In the NTIRE spectral recovery challenges, an RGB image is considered as a linear projection from a high-dimensional hyperspectral datacube to a 3D color image. The existence of metamerism results in identical RGB images for different spectra. The neural network trained in this way cannot distinguish their corresponding spectra. Bottom: A possible mitigation to this problem is to include the optical aberrations of the lens in the image formation model. Spectral information is encoded into the aberrated RGB images, enabling the neural network to tell the difference between metamers. In both cases, the RGB image differences are shown on the right (intensity enhanced for better visualization).

Fu et al. [38] propose to select optimal camera response functions from a dataset and recover hyperspectral images with a CNN-based network. This is equivalent to observer metamerism [33] where different cameras observe the spectra differently. Another technique is to explore the structure of illuminant space [2]. For example, Baek et al. [9] employ 29 CIE standard illuminants to augment the hyperspectral dataset for joint hyperspectral and depth reconstruction. Cao et al. [18] develop an unsupervised network to recover spectral information under two lighting conditions. Although such methods enrich the spectral content of hyperspectral datasets, they model metamersim indirectly. Our work differs from these techniques since we directly generate metamers using the metameric black theory [35], [76], [104]. This guarantees the underrepresented metamerism phenomena (the same color from different spectra) in existing datasets can be modeled more effectively.

III. Fundamentals

A. Spectral image formation

We denote the hyperspectral image as a matrix $\mathbf{X} \in \mathbb{R}^{MN\times K}$, where M, N are the number of pixels, and K is the number of spectral bands. Note that we model the spectral radiance here, not spectral reflectance. Illumination spectrum is included. We have stacked the 2D spatial dimensions in rows of $\mathbf{X}$. The spectral response function (SRF) of the camera can be expressed as a matrix $\mathbf{Q} \in \mathbb{R}^{K\times 3}$. Therefore, the spectrum-to-color projection results in a color image

$$\mathbf{Y} = \mathbf{X}\mathbf{Q}, \tag{1}$$

where $\mathbf{Y} \in \mathbb{R}^{MN\times 3}$. This is the color formation model in the NTIRE 2022 challenge [7]. The inverse problem is to recover $\mathbf{X}$ from $\mathbf{Y}$.

In the past NTIRE challenges [5]–[7], optical aberrations have not been included in the image simulation pipeline. However, the optical system of the RGB camera inevitably introduces spectrally-varying blurs to the spectral images, which is modeled as PSFs. This optical process can be described by a linear matrix-vector product in each spectral band followed by a sum over the spectral dimension. The spectral images through the optical system are $\mathbf{W} = \mathrm{diag}\,(\mathbf{A}\mathbf{X})$, where $\mathbf{A} \in \mathbb{R}^{KMN\times MN}$ is a block matrix that stacks the spectral PSF matrices vertically, and $\mathrm{diag}(\cdot)$ extracts the diagonal blocks. The final RGB image is then

$$\mathbf{Z} = \mathrm{diag}\,(\mathbf{A}\mathbf{X})\,\mathbf{Q}, \tag{2}$$

where $\mathbf{Z} \in \mathbb{R}^{MN\times 3}$. With the optical image formation model accounted for, the inverse problem is to recover $\mathbf{X}$ from $\mathbf{Z}$. It is evident that the optical property in $\mathbf{A}$ spreads the spectral information to the RGB channels, offering side-channel information to help spectral reconstruction. See the Appendix for the full derivation.

The two spectral image formation models are illustrated in Fig. 2. Without considering optical aberrations (Eq. (1)), the neural network struggles to reconstruct the real hyperspectral images in the presence of metamers. The spectrally-varying PSFs (Eq. (2)) are helpful to mitigate this issue since the aberrated RGB images from metamers are different. In the following sections, we will discuss the limitations of existing data-driven spectral reconstruction based on these two image formation models in detail.

B. Hyperspectral datasets and data diversity

Compared to very large color (RGB) image datasets (*e.g.*, ImageNet [29], DIV2K [1]), hyperspectral datasets are far

TABLE I
BASIC INFORMATION OF FOUR EXISTING HYPERSPECTRAL DATASETS.

Dataset	Spectra (nm)	Resolution (x, y, λ)	Amount	Device	Scene
CAVE [96]	400:10:700	$512 \times 512 \times 31$	32	monochrome sensor + tunable filters	lab setup
ICVL [4]	400:10:700	$1392 \times 1300 \times 31$	201	HS camera (Specim PS Kappa DX4)	outdoor
KAUST [53]	400:10:700	$512 \times 512 \times 31$	409	HS camera (Specim IQ)	outdoor
ARAD1K [7]	400:10:700	$482 \times 512 \times 31$	1000	HS camera (Specim IQ)	outdoor

smaller in size, primarily limited by the unavailability of high-quality hyperspectral cameras and the difficulty in acquiring outside the lab with moving target scenes. The largest dataset so far is ARAD1K used in the NTIRE 2022 challenge [7]. In addition, we also include the CAVE [96], ICVL [4], and KAUST [53] datasets that share the same spectral range (400 nm to 700 nm) to extend our experiments. The datasets are summarized in Table I. Although other datasets, such as Harvard [21], KAIST [24], and TokyoTech [60] exist, they cover slightly different spectral bands (420 nm to 720 nm), making it difficult to directly compare and cross-validate results among different datasets. We therefore restrict our analysis to the datasets listed in the table.

The difficulties in the data capture not only affect the size but also the *diversity* of the datasets. In particular, effects like metamerism, which are comparatively rare in everyday environments [37], [64], yet crucial for many actual applications of spectral imaging, are under-represented in the datasets. While the CAVE dataset [96] contains some fake-and-real pairs of objects (*e.g.*, Fig. 1) to account for metamerism, the total amount of such data is still very low. We analyze the general data diversity issue in Section IV and the specific case of metamerism in Section V.

C. Modeling metamerism

Since there are not enough examples of metamerism in existing datasets, their effects in spectral reconstruction went unnoticed in prior works. On the one hand, we need adversarial examples to reveal the unexplored problems. On the other hand, we want to investigate how they can complement the current datasets. Therefore, we propose a new form of data augmentation in our experiments. *Metameric augmentation* starts with existing spectral images and creates a new, *different* spectral image that however maps to the same RGB image (given a specific set of RGB spectral response functions). In Section V, we first use metameric augmentation as an adversarial example to reveal the previously omitted effects of metamerism on the performance discrepancy. In Section VI, we show metameric augmentation is beneficial to mitigate the raised problems along with an aberration-aware training strategy.

Note that data augmentation has proven to be effective in deep learning to mitigate data shortage. Color image augmentation techniques have been focusing mainly on geometric transformations and intensity adjustment. Although these techniques have been employed in prior methods, they only augment the spatial dimensions in hyperspectral images. To the best of our knowledge, metameric augmentation beyond RGB colors, accounting for the physical phenomenon of metamerism, has not been adopted before in spectral reconstruction problems. Metameric augmentation can greatly enrich the **spectral content** in existing datasets to mitigate the lack of diversity.

Interestingly, metamer generation from existing spectra has been studied in color science and spectral rendering to accurately model the scenes using various methods, *e.g.*, metameric black [3], [35], [76] and spectral uplifting [13], [46], [74]. In this work, to support our analysis, we adopt the metameric black approach to generate metamers, whereas we note that other metamer generation methods can also be employed for the same purpose.

A spectrum $\mathbf{S}$ can be projected onto two orthogonal subspaces, one for the fundamental metamer $\mathbf{S}^*$, and the other for metameric black $\mathbf{B}$ [26], [77]. That means the original spectrum can be decomposed as $\mathbf{S} = \mathbf{S}^* + \mathbf{B}$. The fundamental metamer is a particular solution to Eq. (1), and the metameric black always leads to zero tristimulus, *i.e.*, no impact on the color appearance. The set of all possible metamers is called a metamer set [35]. Wyszecki [88] first introduces a decomposition technique to calculate the metameric black. To generate new metamers, it is possible to add a linear combination of metameric blacks to the fundamental metamer [35]. In a linear algebra perspective, metameric blacks lie in the null space (or kernel) of the camera SRF. Any scalar multiplication to a vector in the null space remains in the null space according to the scalar multiplication property [70]. Inspired by the metameric black theory and the mathematical properties, we propose a simple yet effective way to generate metamers by scaling the metameric black component. A new metameric spectrum $\mathbf{S}'$ is then

$$\mathbf{S}' = \mathbf{S}^* + \alpha \mathbf{B}, \tag{3}$$

where $\mathbf{S}^* = \mathbf{Q}\left(\mathbf{Q}^T\mathbf{Q}\right)^{-1}\mathbf{Q}^T\mathbf{S}$ and $\mathbf{B} = \mathbf{S} - \mathbf{S}^*$. Since adding metameric black does not alter the RGB color, we can vary the coefficient α to generate different spectra that are all metamers. To avoid negative spectral radiance, we clip the negative values in the generated spectra and re-calculate the RGB colors for the affected pixels. See the Appendix for the analysis on the effects of clipping to non-negative values while generating metamer data.

D. Performance evaluation metrics

Consider a hyperspectral image $\mathbf{X}_{k,i,j}$ and its estimate $\hat{\mathbf{X}}_{k,i,j}$, where k is the spectral index, and i, j are spatial indices. The reconstruction quality can be evaluated in various ways. The NTIRE 2022 spectral reconstruction challenge [7] adopts two numerical metrics, Mean Relative Absolute Error (MRAE),

$$\mathbf{MRAE} = \frac{1}{KMN}\sum_{k,i,j}\left|\hat{\mathbf{X}}_{k,i,j} - \mathbf{X}_{k,i,j}\right| / \mathbf{X}_{k,i,j}, \quad (4)$$

and Root Mean Square Error (RMSE)

$$\mathbf{RMSE} = \sqrt{\frac{1}{KMN}\sum_{k,i,j}\left(\hat{\mathbf{X}}_{k,i,j} - \mathbf{X}_{k,i,j}\right)^2}. \quad (5)$$

Another metric widely used in hyperspectral imaging is the Spectral Angle Mapper (SAM) [51], [62], [75], although it has not yet found its way into the relevant computer vision literature. SAM emphasizes the spectral accuracy compared to the previous metrics, which are more forgiving of large errors in individual spectral channels:

$$\mathbf{SAM} = \frac{1}{MN}\sum_{i,j}\cos^{-1}\left(\frac{\sum_k \hat{\mathbf{X}}_{k,i,j}\mathbf{X}_{k,i,j}}{\sqrt{\sum_k \hat{\mathbf{X}}_{k,i,j}^2 \sum_k \mathbf{X}_{k,i,j}^2}}\right). \quad (6)$$

Finally, we also inspect the spatial quality in individual spectral channels, and calculate the spectrally averaged Peak Signal-to-Noise Ratio (PSNR),

$$\mathbf{PSNR} = \frac{1}{K}\sum_k 20\log_{10}\left(\frac{\text{MAX}}{\sqrt{\text{MSE}_k}}\right), \quad (7)$$

where MAX is the maximum possible value, and MSE_k is the mean squared error in the k-th spectral band. This metric complements RMSE to account for performance variation in individual spectral bands.

E. Training details

In our study, we conduct all the experiments across various datasets and network architectures. Following the methodologies proposed by the latest champion network MST++ [15], we employ their patch-wise training approach (patches of 128×128 pixels).

1) Data Preparation: Throughout the experiments in this work, we follow the same data format (Matlab-compatible `mat` files) for hyperspectral images in the ARAD1K dataset [7]. To be consistent, we also convert the raw hyperspectral datacubes in the CAVE [96], ICVL [4], and KAUST [53] datasets to this format. The data values are normalized by their respective bit-depths such that the data range is [0.0, 1.0]. The training and validation sets in ARAD1K are kept the same as offered in the NTIRE 2022 spectral recovery challenge [7], *i.e.*, 900 files for training, and 50 for validation. We split the CAVE, ICVL, and KAUST datasets by 90% for training, and 10% for validation. Following the training strategy in MST++ [15], we keep the training and validation lists fixed.

2) Metamer Generation: We adopt the metameric black method [3], [35], [76] to generate metamers from the original hyperspectral data. By varying the coefficient of the metameric black term, we could generate metamers that project to the same RGB color. Note that the original datacube corresponds to $\alpha = 1$, and the fundamental metamer corresponds to $\alpha = 0$. For the experiments with fixed metamers, we use the fundamental metamers to complement the original standard data. This is sufficient to demonstrate our findings. Other arbitrary values would result in the same conclusions. A more aggressive setting is to vary α as a variable to account for the infinite possible metamers in a more realistic situation.

3) Training and Validation Procedure: Following the training strategies in MST++ [15], we sub-sample the hyperspectral datacubes and the corresponding RGB images into overlapping patches of 128×128. Spatial augmentations, such as random rotation, vertical flipping, and horizontal flipping are randomly applied to the training patches. In the validation step, we calculate the evaluation metrics (MRAE, RMSE, PSNR, and SAM) on the full spatial resolution for the ARAD1K (482×512), CAVE (512×512), and KAUST (512×512) datasets. Note that this is different from MST++ [15], where only the central 256×256 regions are evaluated. The ICVL dataset has a very large spatial resolution (1300×1392). To be consistent with other datasets, we evaluate ICVL only in the central 512×512 regions.

Similar as MST++, in each epoch, we train the networks for 1000 iterations, with a total number of 300 epochs. All the reported results are evaluated at the end of the training epochs, *i.e.*, 300k iterations. We find that the training iterations are sufficient to achieve convergence in all our proposed experiments. Hyperparameters, such as learning rate and batch size, are tuned to achieve the best performance for each network on each dataset. All the experiments are conducted on an NVIDIA A100 GPU (80 GB memory).

IV. Finding 1: Atypical Overfitting

Although it is well-known that deep neural networks may suffer from overfitting problems, we find that the overfitting behavior in spectral reconstruction is atypical and difficult to notice with standard evaluations. Here we introduce minimalist changes to the ARAD1K dataset used in NTIRE 2022 challenge [7] in three experiments to demonstrate it. We exhaustively evaluate a total of 17 open-sourced neural network architectures to-date, namely MST++ [15], MST-L [14], MPRNet [99], Restormer [97], MIRNet [98], HINet [22], HDNet [42], AWAN [52], EDSR [55], HRNet [105], HSCNN+ [68], HySAT [80], HPRN [86], SSTHyper [91], MSFN [87], GMSR [83], and SSRnet [31].

A. Training with less data

First, we make a simple change to the training of the participating networks in the NTIRE 2022 challenge [7]. While keeping all the training settings intact, we randomly choose only 50% or 20% of the original training data, respectively, to train the candidate networks and validate the performance on the original validation data. We illustrate the validation curves for MST++ [15] in Fig. 3. See the Supplementary Material for the results of other networks. We summarize the results for 100% and 50% training data in Table II for all networks.

Although the performance with less training data deviates mildly in MRAE, RMSE, and PSNR, the spectral accuracy SAM (highlighted in bold in Table II) is surprisingly less affected. In particular, some networks (*e.g.*, MST++ [15],

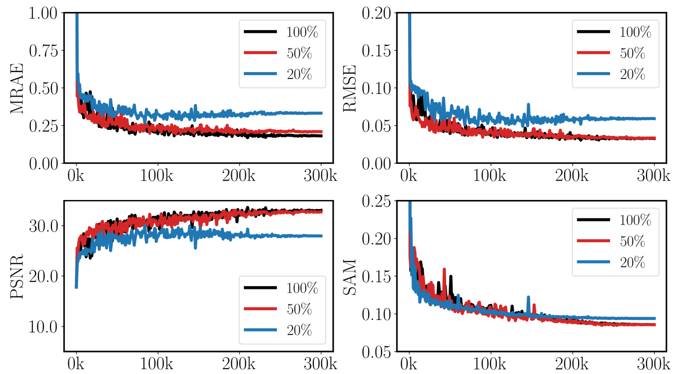

Fig. 3. Validation performance for MST++ [15] with 100%, 50%, and 20% of the original training data on ARAD1K [7].

MIRNet [98]) achieve exactly the same SAM scores. MST-L [14] (50%) even improves SAM slightly, placing itself the best among all.

TABLE II
PERFORMANCE COMPARISON WITH 100% AND 50% OF TRAINING DATA FOR DIFFERENT METHODS ON THE ORIGINAL ARAD1K DATASET.

Network	Data	MRAE↓	RMSE↓	PSNR↑	SAM↓
MST++	100%	0.182	0.033	33.0	**0.086**
	50%	0.209	0.033	32.7	**0.086**
MST-L	100%	0.184	0.031	33.5	**0.084**
	50%	0.253	0.042	30.6	**0.080**
MPRNet	100%	0.212	0.034	32.5	**0.084**
	50%	0.293	0.039	31.3	**0.091**
Restormer	100%	0.204	0.033	33.2	**0.083**
	50%	0.304	0.041	31.4	**0.092**
MIRNet	100%	0.186	0.030	33.7	**0.082**
	50%	0.214	0.033	32.6	**0.082**
HINet	100%	0.234	0.036	32.3	**0.085**
	50%	0.267	0.041	30.7	**0.090**
HDNet	100%	0.223	0.038	31.2	**0.095**
	50%	0.296	0.047	28.9	**0.097**
AWAN	100%	0.213	0.034	32.2	**0.091**
	50%	0.273	0.042	30.5	**0.095**
EDSR	100%	0.358	0.052	27.3	**0.095**
	50%	0.430	0.059	26.1	**0.093**
HRNet	100%	0.388	0.057	26.4	**0.094**
	50%	0.413	0.065	25.5	**0.096**
HSCNN+	100%	0.428	0.066	25.4	**0.098**
	50%	0.462	0.068	25.0	**0.101**
HySAT	100%	0.176	0.028	34.6	**0.085**
	50%	0.254	0.037	31.6	**0.089**
HPRN	100%	0.257	0.044	30.4	**0.098**
	50%	0.261	0.041	30.8	**0.098**
SSTHyper	100%	0.181	0.030	33.6	**0.083**
	50%	0.241	0.039	31.1	**0.083**
MSFN	100%	0.226	0.038	31.9	**0.084**
	50%	0.271	0.044	30.4	**0.096**
GMSR	100%	0.308	0.056	27.5	**0.113**
	50%	0.393	0.077	25.6	**0.127**
SSRnet	100%	0.270	0.048	29.5	**0.097**
	50%	0.307	0.052	29.0	**0.108**

We therefore paradoxically find that despite the small size of hyperspectral datasets, the data already seems to be redundant. This serves as a first indication that the *diversity* of the datasets is severely lacking. We analyze this effect in more depth in the following experiments.

B. Validation with unseen data

To further scrutinize the underlying issue, we validate existing pre-trained models with "unseen" data synthesized from the original dataset used in the NTIRE challenge. The challenge organizers state that "the exact noise parameters and JPEG compression level used to generate RGB images for the challenge was kept confidential" [7]. Only the spectrum-to-color projection was considered, and no aberrations of the optical system were simulated.

In our experiments, we generate new RGB images using the same methodology and calibration data, but different noise and compression settings. Specifically, we use the SRF data for a Basler ace 2 camera (model A2a5320-23ucBAS) known to the networks, and simulate Poisson noise at varying noise levels by controlling the number of photon electrons (npe). We adopt the same rudimentary in-camera image signal processing pipeline. As an illustrative example, we use MST++ [15] in Table III, Row 1 as a reference for comparison; results for other networks can be found in the Supplemental Material.

First, as a baseline, we consider a noiseless (npe = 0) and aberration-free case with moderate JPEG compression quality (Q = 65), shown in Table III, Row 2. The results show significant drops in all the performance metrics. Note that the only differences here compared to the challenge dataset are the noise level and compression quality – the base images are identical! This indicates that the network overfits both the noise and JPEG compression parameters.

Second, in Row 3, we generate noiseless RGB images, but in lossless PNG format, as opposed to the JPEG (Q = 65) in Row 2. Note that JPEG compression is not necessary for the core inverse problem in hyperspectral imaging, since raw data could be readily obtained from the sensors. This results in paradoxical reconstruction performance. MRAE and SAM improve compared to Row 2 (but are still worse than Row 1), while RMSE and PSNR deteriorate further. Considering this only eliminates image compression, and the networks were trained on MRAE [15], we can confirm that the network indeed overfits the specific unknown JPEG compression used in the challenge [7].

Third, we consider a more realistic imaging scenario, in which we eliminate the impact of unnecessary compression by employing the lossless PNG format to save the RGB images (equivalent to using raw camera data). We adopt moderate noise levels (npe = 1000) and realistic optical aberrations from a recent double Gauss lens patent [44] to mimic a real photographic camera. We can observe a further performance drop in Row 4, which provides additional evidence that the network overfits the unknown parameters in the image simulation pipeline [7]. When used under realistic imaging conditions, the performance degrades significantly.

C. Cross-dataset validation

In addition, we inspect the effects of different datasets (*cf.* Table I) on the performance. We train the MST++ network on the four datasets with the same image simulation parameters. To eliminate the impact of other factors, we choose the

TABLE III
EVALUATION OF THE PRE-TRAINED MST++ MODEL ON SYNTHESIZED VALIDATION DATA FROM THE ARAD1K DATASET WITH DIFFERENT NOISE LEVELS, COMPRESSION QUALITY, AND REALISTIC OPTICAL ABERRATIONS.

	Data property				MRAE ↓	RMSE ↓	PSNR ↑	SAM ↓
	Data source	Noise (npe)	RGB format	Aberration				
1	NTIRE 2022	unknown	jpg (Q unknown)	None	0.170	0.029	33.8	0.084
2	Synthesized	0	jpg (Q = 65)	None	0.460	0.049	29.2	0.094
3		0	png (lossless)	None	0.362	0.057	28.7	0.087
4		1000	png (lossless)	CA*	0.312	0.055	28.4	0.118

*CA: chromatic aberration, from a patent double Gauss lens (US20210263286A1).

ideal noiseless and aberration-free condition without compression. In the validation, we use our trained model on ARAD1K dataset to validate on the other three datasets, respectively. In Table IV, we compare the performance with the models both trained and validated on the original datasets. Results for other networks can be found in the Supplementary Material. They all illustrate the same difficulties in generalization.

TABLE IV
CROSS-DATASET VALIDATION USING MST++ [15].

Trained on	Validated on	MRAE↓	RMSE↓	PSNR↑	SAM↓
CAVE	CAVE	0.237	0.034	31.9	0.194
ARAD1K	CAVE	1.626	0.074	24.4	0.376
ICVL	ICVL	0.079	0.019	38.3	0.024
ARAD1K	ICVL	0.627	0.091	22.0	0.110
KAUST	KAUST	0.069	0.013	44.4	0.061
ARAD1K	KAUST	1.042	0.100	22.0	0.370

Even though the imaging conditions are the same and ideal, the network trained on one dataset experiences significant performance drops in all metrics when validated on other datasets. This indicates that the contents of the datasets, as well as the acquisition devices used to capture the datasets, play important roles.

We also point out that the CAVE dataset [96], although smaller and older than the others, is more difficult to train for better performance. This is probably due to the fact that CAVE consists of several challenging scenes of real and fake objects that appear in similar colors, but other datasets comprise less aggressive natural scenes.

D. Discussion

The experiments conducted in this section clearly highlight several shortcomings with respect to the existing datasets. **(1) They lack diversity in nuisance parameters** such as noise and compression ratios. Our experiments show that when the RGB images are slightly modified by changing the nuisance parameters embedded in the image formation (*e.g.*, noise level, compression factor, and optical aberrations), a significant drop in the performance is observed. **(2) They lack scene diversity.** Training modern deep neural networks via supervised learning typically demands large-scale datasets. However, our experiments reveal that reducing the training data volume leads to only a marginal or even no drop in spectral accuracy. This suggests that the dataset lacks diversity in its content. The cross-dataset validation experiments further show that each dataset has its own statistics. Neural networks trained on a single dataset exhibit limited generalization to other datasets, suggesting the lack of sufficient scene diversity in individual datasets. In summary, our experiments provide clear evidence that the performance degradation primarily stems from insufficient scene diversity in existing datasets. While we show here the results for the largest available dataset (ARAD1K), the Supplementary Material shows consistent results for all the datasets. Both of these aspects result in over-fitting and prevent the networks from learning the general spectral image restoration task. Next, we specifically analyze the effect of metamerism; the analysis of the impact of optical aberrations will be deepened in Section VI.

V. FINDING 2: METAMERIC FAILURE

In this section, we inspect the performance of existing methods using metamer as an adversary to validate as well as re-train the neural networks for performance analysis.

A. Validation with metamers

We generate metamer datacubes (metamer data) from the original ARAD1K dataset (standard data) using the metameric black method [35]. For this set of experiments, we fix the coefficient $\alpha = 0$ in Eq. (3). We choose a realistic imaging condition as used in Table III, Row 4, and keep it the same for both cases. The validation results on the ARAD1K dataset for existing pre-trained networks are summarized in Table V. We also visualize the reconstructed spectral images in five arbitrary bands (420 nm, 500 nm, 550 nm, 580 nm, and 660 nm) and spectra of two points in Fig. 4 for Scene ARAD_1K_0944 from the validation set.

From the numerical results in Table V, it is apparent that all the existing methods experience catastrophic performance drop in terms of MRAE and SAM in the presence of metamers, which we call metameric failure. The MRAE (*cf*. Eq. (4)) may yield large values when large errors occur for dark ground-truth pixels (see the exemplary spectra in Fig. 4). The SAM values become large when the spectra are essentially dissimilar with each other. RMSE and PSNR do not capture the spectral differences as well, since they average out differences in the spatial and spectral dimensions.

The visual results in Fig. 4 show that the reconstruction results are very close to each other for both standard and metamer data, because the input RGB images are quite similar. However, distinct differences exist in the scene for certain spectral bands, *e.g.*, the intensities of the yellow and green parts of the slide (blue box) in 500 nm band vary in the

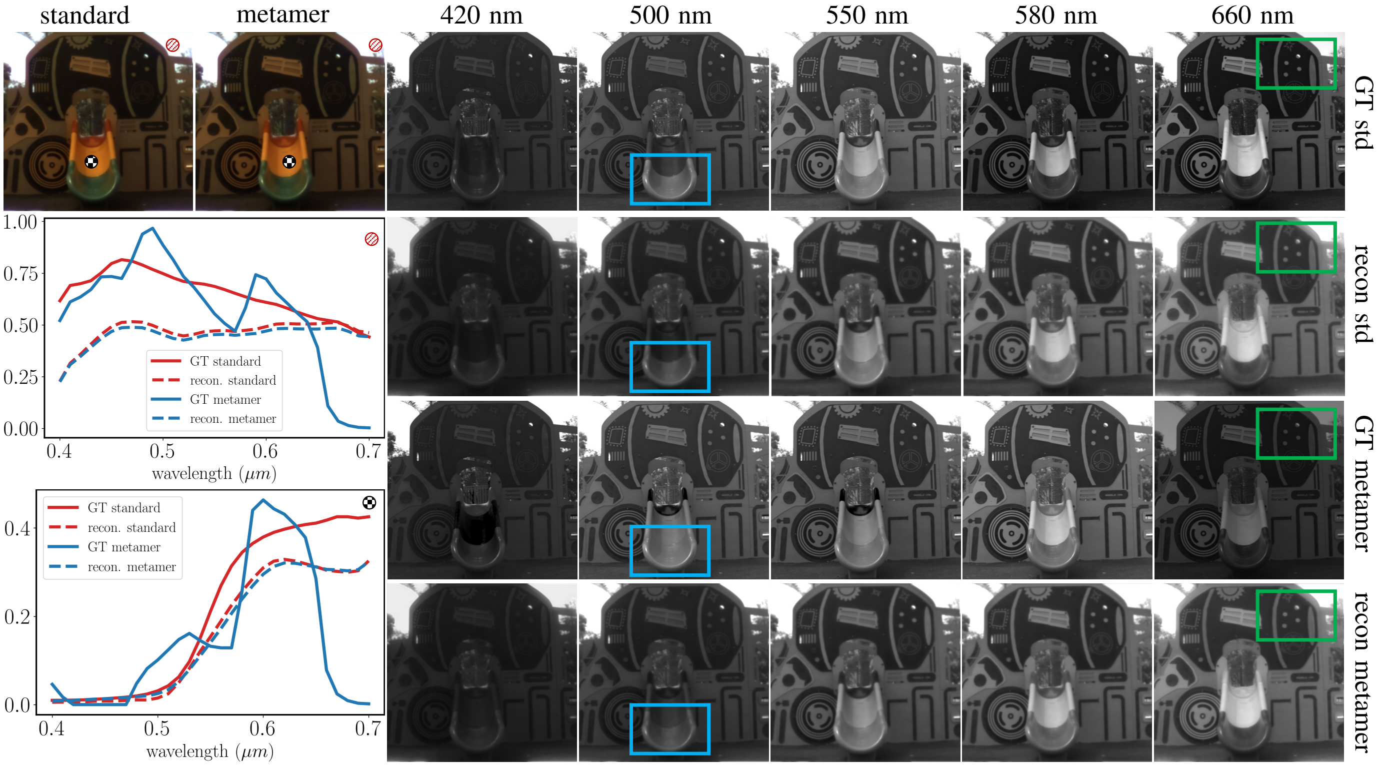

Fig. 4. Validation with metamers for MST++ [15]. An example Scene ARAD_1K_0944 is shown to visualize the standard and metamer datacubes. Top left: the standard and metamer data result in similar color images. Bottom left: ground-truth and reconstructed spectra from two labeled points. Right: ground-truth and reconstructed spectral images in 420 nm, 500 nm, 550 nm, 580 nm, and 660 nm.

standard data, but remain identical in the metamer data. The reconstructions fail to reflect this important difference. All spectral images are displayed on the same global intensity scale, so the brightness differences (green box) in corresponding images reflect the reconstruction artifacts.

B. Training with metamers

The pre-trained models were not explicitly trained to cope with metamers. This raises the question whether it is possible to improve the performance by training the networks with metamer data.

As a first step, we use both the standard and metamer data ($\alpha = 0$) generated from the ARAD1K dataset to train various networks. To eliminate the impact of other factors, we simulate the RGB images in a noiseless, aberration-free condition, and without compression.

However, it is not sufficient to consider only a pair of standard and fixed metamer data. In reality, there are infinite metamers that project to the same color. As a second variant, we train the neural networks with random metamers generated on-the-fly as a spectral augmentation to enhance the spectral content of existing datasets. We vary the coefficient for the metameric black by setting α as a uniformly distributed random number in the range $[-1, 2]$. During validation, we use both the standard validation data and their corresponding metamer data with fixed $\alpha = 0$, which doubles the amount of the original validation data.

As an example, we train MST++ and evaluate its validation performance over the training process. In Fig. 5, we show that it is no longer a good choice to use MRAE as the loss function [15] and evaluation metric [7], because it is completely overwhelmed by metamers. Instead, we find that L1 loss is a more stable loss function.

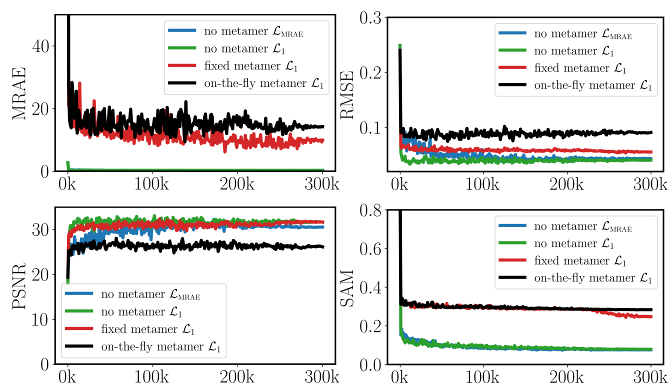

Fig. 5. Training MST++ with metamers. It fails to combat fixed metamers and on-the-fly metamers, in particular on the spectral accuracy SAM.

$$\mathcal{L}_1\left(\hat{\mathbf{X}}, \mathbf{X}\right) = \frac{1}{KMN} \sum_{k,i,j} \left| \hat{\mathbf{X}}_{k,i,j} - \mathbf{X}_{k,i,j} \right|. \tag{8}$$

We then train the network with L1 loss for three cases, no metamer (easy), fixed metamer (medium), and on-the-fly metamer (difficult). Nevertheless, we can see in Fig. 5 that the network fails in particular for the spectral accuracy SAM.

We also train all other candidate networks with fixed and on-the-fly metamers. The results are summarized in Table VI.

TABLE V
VALIDATION PERFORMANCE FOR DIFFERENT PRE-TRAINED MODELS ON STANDARD (STD) DATA AND METAMER (MET) ADVERSARY SYNTHESIZED FROM THE ARAD1K DATASET [7].

Network	Data	MRAE↓	RMSE↓	PSNR↑	SAM↓
MST++	std	0.312	0.055	33.8	0.084
	met	**52.839**	0.091	26.0	**0.580**
MST-L	std	0.327	0.055	28.0	0.118
	met	**51.321**	0.090	25.9	**0.579**
MPRNet	std	0.661	0.066	26.1	0.125
	met	**145.981**	0.122	23.3	**0.547**
Restormer	std	0.510	0.066	25.5	0.126
	met	**79.705**	0.116	23.4	**0.567**
MIRNet	std	0.404	0.077	24.8	0.124
	met	**38.252**	0.089	24.8	**0.570**
HINet	std	0.450	0.063	26.5	0.120
	met	**67.148**	0.096	24.8	**0.552**
HDNet	std	0.450	0.082	23.9	0.126
	met	**34.429**	0.095	23.8	**0.570**
AWAN	std	0.424	0.080	24.6	0.119
	met	**39.854**	0.095	24.4	**0.558**
EDSR	std	0.421	0.066	25.5	0.132
	met	**49.435**	0.100	23.8	**0.564**
HRNet	std	0.514	0.078	23.9	0.128
	met	**43.726**	0.112	22.7	**0.560**
HSCNN+	std	0.508	0.075	24.4	0.148
	met	**42.274**	0.098	23.1	**0.556**
HySAT	std	0.326	0.047	29.4	0.127
	met	**61.516**	0.095	26.2	**0.594**
HPRN	std	0.524	0.104	22.0	0.130
	met	**33.439**	0.102	22.8	**0.574**
SSTHyper	std	0.314	0.058	27.7	0.117
	met	**48.427**	0.089	25.8	**0.575**
MSFN	std	0.328	0.055	28.3	0.119
	met	**51.846**	0.090	26.0	**0.573**
GMSR	std	0.484	0.075	24.9	0.138
	met	**53.063**	0.109	23.6	**0.547**
SSRNet	std	0.419	0.075	25.8	0.130
	met	**38.225**	0.104	24.5	**0.564**

Again, the same performance drop applies to all networks. Finally, we show the results of the top-performing network, MST++ on the CAVE, ICVL, and KAUST datasets in Table VII. (See Supplementary Material for more results). As before, the performance drops similarly in the presence of metamers.

C. Discussion

The experiments conducted in this section clearly highlight the difficulties that the data-driven spectral recovery methods face with metamers: (1) **lack of sufficient metameric data** in current datasets, (2) **training with metamers** alone cannot mitigate the issue when the problem is formulated by Eq. (1), and (3) **spectral estimation from RGB data** is indeed limited in the presence of metamers.

The limitations of spectral estimation from RGB data are ultimately not overly surprising – after all the projection from the high dimensional spectral space to RGB invariably destroys scene information that can be difficult to recover. Spatial context from underrepresented data does not contribute to the spectral estimation, because such information remains the same for metamers. However, our experiments show that this is indeed an issue faced by the state-of-the-art methods, which so far went unnoticed due to the under-representation of metamers in the datasets. This shortcoming will also

TABLE VI
PERFORMANCE COMPARISON FOR TRAINING VARIOUS NETWORKS WITH FIXED AND ON-THE-FLY METAMERS.

Network	Metamer	MRAE↓	RMSE↓	PSNR↑	SAM↓
MST++	no	0.270	0.041	31.6	0.079
	fixed	9.912	0.056	31.6	0.247
	on-the-fly	14.224	0.091	26.1	0.284
MST-L	no	0.269	0.040	32.2	0.081
	fixed	11.398	0.061	30.2	0.289
	on-the-fly	12.155	0.061	27.0	0.258
MPRNet	no	0.346	0.051	29.9	0.076
	fixed	7.359	0.059	30.6	0.224
	on-the-fly	13.492	0.087	26.6	0.264
Restormer	no	0.286	0.041	31.5	0.068
	fixed	8.129	0.059	31.0	0.242
	on-the-fly	10.186	0.089	26.7	0.264
MIRNet	no	0.258	0.040	32.2	0.083
	fixed	9.555	0.061	30.5	0.289
	on-the-fly	13.205	0.088	26.4	0.290
HINet	no	0.315	0.056	28.0	0.081
	fixed	8.322	0.068	29.6	0.296
	on-the-fly	14.238	0.090	25.5	0.288
HDNet	no	0.287	0.045	30.2	0.087
	fixed	8.884	0.064	30.2	0.296
	on-the-fly	18.270	0.087	26.4	0.299
AWAN	no	0.240	0.039	32.0	0.073
	fixed	8.789	0.068	29.6	0.294
	on-the-fly	14.406	0.090	26.0	0.264
EDSR	no	0.415	0.061	26.1	0.084
	fixed	9.717	0.073	26.6	0.297
	on-the-fly	11.723	0.101	23.2	0.290
HRNet	no	0.430	0.065	25.6	0.085
	fixed	8.142	0.076	26.0	0.299
	on-the-fly	13.098	0.103	23.0	0.294
HSCNN+	no	0.516	0.077	24.1	0.082
	fixed	9.362	0.085	24.8	0.297
	on-the-fly	13.311	0.102	22.9	0.286
HySAT	no	0.295	0.036	32.6	0.078
	fixed	7.450	0.063	29.9	0.270
	on-the-fly	12.806	0.087	26.5	0.283
HPRN	no	0.249	0.038	32.7	0.076
	fixed	5.633	0.060	31.6	0.221
	on-the-fly	12.197	0.080	26.7	0.263
SSTHyper	no	0.261	0.036	32.4	0.083
	fixed	10.070	0.060	30.5	0.285
	on-the-fly	12.194	0.083	27.3	0.276
MSFN	no	0.283	0.043	30.4	0.078
	fixed	8.609	0.065	29.1	0.286
	on-the-fly	12.198	0.096	24.9	0.288
GMSR	no	0.333	0.049	29.0	0.114
	fixed	10.612	0.065	29.4	0.302
	on-the-fly	11.674	0.096	24.9	0.300
SSRNet	no	0.354	0.052	29.0	0.098
	fixed	6.678	0.072	28.9	0.286
	on-the-fly	13.359	0.102	25.0	0.291

TABLE VII
TRAINING WITH METAMERS FOR MST++ ON THE CAVE [96], ICVL [4], AND KAUST [53] DATASETS.

Dataset	Metamer	MRAE↓	RMSE↓	PSNR↑	SAM↓
CAVE	no	1.014	0.038	29.9	0.192
	fixed	38.26	0.053	29.6	0.229
	on-the-fly	226.0	0.078	25.2	0.451
ICVL	no	0.067	0.016	40.1	0.027
	fixed	1.454	0.041	34.8	0.229
	on-the-fly	2.615	0.087	24.3	0.268
KAUST	no	0.082	0.016	43.2	0.076
	fixed	2.033	0.022	39.0	0.217
	on-the-fly	1.874	0.032	33.7	0.245

affect other uses of the same datasets, for example in the training of reconstruction methods for spectral computational cameras [10], [20], [47]. A metameric adversary helps to identify this overlooked issue and avoid unrealistically high numerical scores for existing systems. It also underscores that, without side-channel information, no intrinsic property exists in RGB images to distinguish between metamers, even when they are augmented. Note that this does not downplay the effects of metameric augmentation, since the problem formulation of reconstructing spectral information from RGB images in Eq. (1) is fundamentally limited. Once the problem is formulated in Eq. (2), metameric augmentation contributes to improving the network robustness. We will explore it further in the next section.

VI. Finding 3: The Aberration Advantage and Effective Spectral Encoding

A. Aberration-Aware Training with Metameric Augmentation

As shown so far, the existing methods have difficulties distinguishing metamers in the ideal noiseless and aberration-free condition. In this section, we analyze what effect (if any) optical aberrations have on this situation, *i.e.*, aberration-aware training [94]. To this end, we train the networks in a realistic imaging condition with moderate noise level (npe = 1000), lossless PNG format, and aberrations from the same double Gauss lens as before [44]. In short, we simulate, through spectral ray tracing, the effect that an imperfect (*i.e.*, aberrated) optical system has on the RGB image measured when observing a specific spectral scene. The details of this simulation can be found in the Supplementary Material.

In Fig. 6, we show an example with MST++ for the validation on SAM in two situations, one with fixed metamers, and the other with on-the-fly metamers. As a reference, we also show the standard validation without metamers as done in previous works (thin dashed black lines). As we can see, the realistic optical aberrations of the lens actually *improve* the spectral estimation in the presence of metamers as long as metamers are modeled in the training. With chromatic aberrations combined with metameric augmentation, the network can already distinguish fixed metamer pairs, achieving similar accuracy as the standard case. In the more aggressive case of on-the-fly metamers, chromatic aberrations also improve the spectral accuracy, compared with their no-aberration counterparts. Again, this aberration advantage holds for all datasets (Table VIII). See Supplementary Material for details.

TABLE VIII
SAM metrics for MST++ on CAVE [96], ICVL [4], and KAUST [53].

Dataset	Fixed metamers		On-the-fly metamers	
	no aberration	aberration	no aberration	aberration
CAVE	0.251	0.135	0.380	0.167
ICVL	0.028	0.077	0.240	0.085
KAUST	0.212	0.113	0.221	0.113

We carry out an ablation study to examine all possible combinations of aberrations and metameric augmentation, as shown in Fig. 6. This leads to 4 situations: (1) Training without aberrations, and without metameric-augmentation (purple

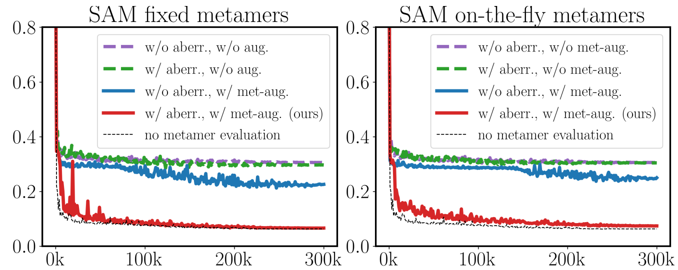

Fig. 6. Chromatic aberrations improve spectral accuracy. Left: fixed metamers. Right: on-the-fly metamers. aberr.: aberrations; met-aug.: metameric augmentation.

lines). It is the existing training method that fails to combat metamers, as we have demonstrated in Section V. (2) Training with aberrations, but without metameric augmentation (green lines). It corresponds to simply a more blurry RGB image without considering metamers, which fails similarly. (3) Training without aberrations, but with metameric augmentation (blue lines). It slightly improves the spectral reconstruction, owing to the fact that noise may introduce slight differences in the RGB images. (4) Training with both aberrations and metameric augmentation. With the combination of aberrations and metameric augmentation, the network can learn the differences in RGB images between metamers, achieving similar performance as existing methods where metamers are actually not evaluated. These results clearly demonstrate again that the spectral reconstruction solely from RGB images fails for metamer data. The problem can be more effectively formulated only when spectral information, such as optical aberrations, is taken into account.

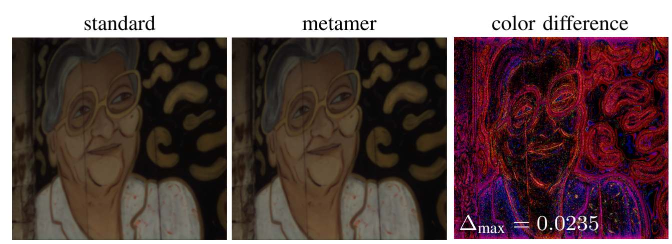

Fig. 7. Chromatic aberration induced informative color differences (right) as spectral cues for metamer pairs (left and middle).

To understand why optical aberrations help improve the reconstruction, consider the simulated images in Fig. 7. The left and middle images are simulations of RGB images for metameric scene pairs, with the difference image on the right. The different spectra of the two scenes are affected *differently* by the optical aberrations, and therefore, although the *scenes* are metamers of each other, the *RGB images* are in fact different. With optical aberrations, spectral information spreads out into adjacent pixels spatially at the cost of slightly making the RGB images blurry. The networks then see these color differences from metamers during training, and are able to correctly learn the mapping from metameric spectra to colors. We visualize the reconstructed spectra for the two example points in Fig. 4 with the MST++ model trained with both aberrations and on-the-fly metameric augmentation in Fig. 8. Now it is clear that when the network is trained with aberrations and metameric augmentation, it can tell the metameric spectra apart, which is not possible otherwise. In effect, the optical aberrations have *encoded* spectral information into the RGB image, which the networks can learn to distinguish, lending credibility to *PSF engineering* methods for hyperspectral encoding [10], [20], [47]. We also note that **incidental aberrations of a lens are only a weak spectral encoding**. Better performance would require optimizing the spectral encodings deliberately.

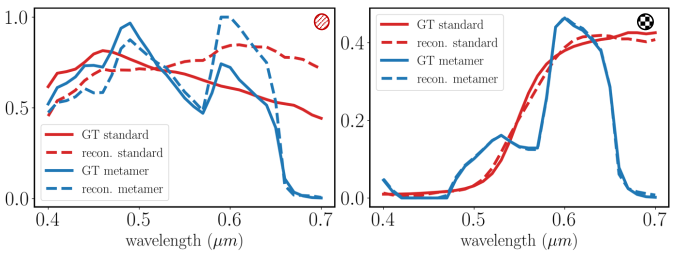

Fig. 8. Reconstructed metameric spectra with MST++ trained with aberrations and metameric augmentation. The two example spectra are exactly the same as in Fig. 4.

B. Effective Spectral Encoding

The above experiments point to a better formulation of the spectral reconstruction problem where spectral encoding plays a key role. Optical aberrations, however, are usually *minimized* in camera lenses, making them imperfect candidates for effective spectral encoding. Existing works have explored some deliberate use of dispersive optical elements for this purpose, including diffraction rotation [47] and grating [10]. However, their performance has historically not been analyzed for metamers either. To this end, we test such spectral encoding schemes on a real challenging scene as shown earlier in Fig. 1. In Fig. 9, we compare four spectral encoding conditions: None (no encoding), Diffraction Rotation (used in [47]), Double Gauss aberrations (used above), and Grating (used in [10]). We train MST++ with metameric augmentation for these spectral encodings, and the SAM results in Fig. 9(b) show that spectral encodings indeed improve the overall spectral accuracy compared with no spectral encoding. The corresponding spectral PSFs are shown in Fig. 9(c). The reconstructed spectra for the same two points in Fig. 1 are shown in Fig. 9(d). Without spectral encoding, the spectral accuracy diverges, while all the spectral encodings improve the spectral quality. Interestingly, different spectral encoding schemes lead to varying spectral accuracy. Diffraction rotation tends to separate the metameric colors more, but the overall SAM is worse than aberrations and grating. We highlight that such challenging metameric spectra have not been properly evaluated in previous works. Although these spectral encoding schemes have been proposed, they are not optimized to deal with metamers yet. Again, the primary culprits are the dataset limitations we have pointed out – metamers are highly underrepresented in existing datasets. While the CAVE dataset includes such examples, they are present in only limited quantities. To achieve better spectral reconstruction in such challenging situations, an effective spectral encoding and a powerful neural network should be trained on a large-scale and diverse dataset in which metamers are well represented. This constitutes a critical yet unresolved challenge in the field of data-driven spectral reconstruction.

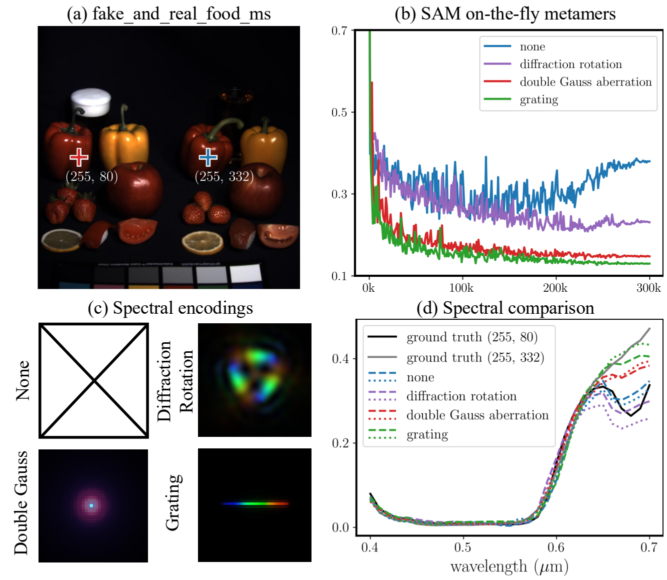

Fig. 9. Comparison of different spectral encoding methods. (a) The fake_and_real_food_ms scene from CAVE. (b) Training performance in SAM for different encoding methods. (c) PSFs for different spectral encodings: none, diffraction rotation [47], double Gauss aberration, and grating [10]. (d) Reconstructed spectra for the two points denoted in (a).

Other spectral encoding methods, such as CASSI, could also benefit from metameric augmentation. In Table IX, we show results on one of the best-performing networks, MST-L [14], trained on the ARAD1K dataset and validated on the other three datasets. All the experiments are carried out with Poisson noise npe = 1000, and aberrations from the double Gauss lens (US20210263286A1). We adopt the CASSI settings in [14].

As indicated by the results, RGB2HS cannot maintain its performance in PSNR and SAM for data from other datasets, even with metameric augmentation to account for metamers during training (see the RGB2HS columns). With the aid of aberrations and metameric augmentation, the reconstruction performance could be boosted for both standard and metamer data. In addition, evident from the CASSI results without metameric augmentation, CASSI offers overall better reconstruction quality, thanks to its spatial-spectral encoding design rooted in the compressive sensing theory. When metameric augmentation is applied to CASSI, its performance has also been boosted, proving the effectiveness of our metameric augmentation scheme. Note that the reconstruction quality varies among datasets, owing to the different spectral content in each dataset.

C. Discussion

We have demonstrated through aberration-aware training with metameric augmentation that only when optical aberrations are considered in the image formation, the problem of spectral reconstruction from RGB images can be better formulated. Our ablation experiments further prove that it is not sufficient to only model metamers without employing spectral

TABLE IX
GENERALIZATION ANALYSIS ON THE EFFECTS OF THE PROPOSED METAMERIC AUGMENTATION FOR VARIOUS SPECTRAL ENCODING SCHEMES.

Trained on	Validated on	No metameric augmentation				With metameric augmentation					
		RGB2HS		CASSI		RGB2HS		RGB2HS + aberrations		CASSI	
		PSNR	SAM	PSNR	SAM	PSNR	SAM	PSNR	SAM	PSNR	SAM
ARAD1K	CAVE std	29.9	0.359	30.6	0.270	22.4	0.692	31.1	0.363	31.7	0.257
	CAVE met	27.3	0.510	24.2	0.382	29.9	0.296	37.9	0.160	27.4	0.295
	ICVL std	24.5	0.090	39.7	0.048	35.5	0.168	34.4	0.164	39.9	0.046
	ICVL met	24.5	0.497	27.8	0.314	33.9	0.290	34.1	0.311	32.8	0.195
	KAUST std	23.5	0.512	41.9	0.099	32.7	0.253	35.5	0.247	44.3	0.078
	KAUST met	25.3	0.775	32.8	0.284	34.0	0.220	39.9	0.073	37.2	0.184

encodings. In addition, we highlight that optical aberrations are incidental but not deliberate spectral encodings, so their effect is still limited. We further examine alternative spectral encoding methods, such as PSF engineering approaches [10], [20], [47], and compressed sensing methods like CASSI [25], [79]. The results of our extensive experiments reiterate the credibility to such computational camera approaches. However, learned reconstruction methods for these approaches also suffer from the same dataset issues as the methods analyzed in this paper, making the collection of large-scale, diverse spectral image data a matter of urgency. These new datasets in turn will enable the design of improved optical encodings in computational spectral imaging systems without overfitting to specific scenarios.

VII. Conclusion

In this work, we have comprehensively analyzed a category of data-driven spectral reconstruction methods from RGB images by reviewing the problem fundamentally from dataset bias to physical image formation, and to reconstruction networks. From an optics-aware perspective, we leverage both metamerism and optical aberrations to reassess existing methodologies.

The major findings of our study reveal important yet previously overlooked limitations in this research direction. **(1)** The limitations of current datasets lead to overfitting to both nuisance parameters (noise, compression), as well as limited scene content. **(2)** Metamerism in particular presents a challenge both in terms of under-representation in the datasets, and in terms of fundamental limitations of spectral reconstruction from RGB input. **(3)** Metameric augmentation along with the targeted use of optical aberrations paves the way to combating the metamer issue, though more effective spectral encodings are demanded to solve the challenge.

Our results systematically demonstrate that it is impossible to accurately reconstruct spectra solely from RGB images. In order to realize the dream of spectral estimation from arbitrary RGB sources, it is necessary to coherently and jointly diversify the spectral contents in hyperspectral image datasets, adopt side-channel information from the optical system, and embrace versatile spectral data augmentation methods to fully enable the power of networks in adaptation to whole families of spectral encodings. We argue that addressing these foundational issues is imperative. Continuing to propose new network designs without rethinking the misdefined problem formulation will fall into the same fundamental shortcomings. Clarifying these limitations will enable the community to focus on solving the real challenges in snapshot spectral imaging.

The dataset limitations we point out in this work may also apply to other spectral reconstruction problems using the referenced datasets, such as CASSI, PSF engineering, and multispectral-hyperspectral fusion. In particular, the same metamerism issue has not yet been extensively evaluated in such domains either. Our findings underscore the broader importance of effective spectral encodings in such snapshot spectral imaging problems. The proposed metameric augmentation technique could inform future directions in optical design, network design, and, more importantly, their joint optimization to cope with metamers.

Acknowledgments

This work was supported by the KAUST Individual Baseline Funding and Center of Excellence for Generative AI. Seung-Hwan Baek was supported by the National Research Foundation of Korean (NRF) grant funded by the Korea Goverment (MSIT) (RS-2024-00438532, RS-2023-00211658) and Basic Science Research Program through the NRF funded by the Ministry of Education (2022R1A6A1A03052954).

Appendix

A. Image Formation Model

Mathematically, the physical image formation of a color image from the spectral radiance can be expressed by

$$g_c(x,y) = \int_{\lambda_1}^{\lambda_2} \left(f(x,y,\lambda) * h(x,y,\lambda) \right) q_c(\lambda)\, d\lambda, \qquad (9)$$

where $f(x,y,\lambda)$ is the spectral image, $h(x,y,\lambda)$ is the spectral point spread function (PSF) of the optical system, $q_c(\lambda)$ is the spectral response function (SRF) of the sensor, and $g_c(x,y)$ is the color image in color channel $c \in [R,G,B]$.

Let us denote the hyperspectral image as a matrix $\mathbf{X} \in \mathbb{R}^{MN\times K}$, where M, N are the number of pixels in spatial dimensions, and K is the number of spectral bands in spectral dimension. Note that we have stacked the 2D spectral images in rows of $\mathbf{X}$. Explicitly, we have

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K], \qquad (10)$$

where each column $\mathbf{x}_k \in \mathbb{R}^{MN \times 1}$ is a vector for the spectral image in spectral channel k. The SRF of the sensor is a matrix $\mathbf{Q} \in \mathbb{R}^{K \times 3}$, *i.e.*,

$$\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3] = \begin{bmatrix} q_{11} & q_{21} & q_{31} \\ q_{21} & q_{22} & q_{32} \\ \vdots & \ddots & \vdots \\ q_{K1} & q_{K2} & q_{K3} \end{bmatrix}, \tag{11}$$

where each column $\mathbf{q}_c \in \mathbb{R}^{K \times 1}$. Therefore, the spectrum-to-color projection results in a color image

$$\mathbf{Y} = \mathbf{X}\mathbf{Q}, \tag{12}$$

where $\mathbf{Y} \in \mathbb{R}^{MN \times 3}$ with three columns

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3], \tag{13}$$

and each column $\mathbf{y}_c \in \mathbb{R}^{MN \times 1}$ is a vector for the image in color channel $c \in [R, G, B]$.

When considering the spectral PSFs in each spectral channel, the optically blurred image can be expressed by

$$\mathbf{w}_k = \mathbf{A}_k \mathbf{x}_k, \quad k \in [1, 2, \ldots, K], \tag{14}$$

where $\mathbf{A}_k \in \mathbb{R}^{MN \times MN}$ is a matrix that represents the spectral PSF in channel k. Similar as $\mathbf{X}$, we concatenate $\mathbf{w}_k$ horizontally to obtain the spectral images through the optical system as

$$\begin{aligned} \mathbf{W} &= [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_K] \\ &= [\mathbf{A}_1\mathbf{x}_1, \mathbf{A}_2\mathbf{x}_2, \ldots, \mathbf{A}_K\mathbf{x}_K]. \end{aligned} \tag{15}$$

We define a block matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \cdots \\ \mathbf{A}_K \end{bmatrix}, \tag{16}$$

which stacks the matrices $\mathbf{A}_k$ vertically, and $\mathbf{A} \in \mathbb{R}^{KMN \times MN}$. Therefore, we have

$$\mathbf{W} = \mathrm{diag}(\mathbf{A}\mathbf{X}), \tag{17}$$

where $\mathrm{diag}(\cdot)$ extracts the K diagonal blocks and concatenate them horizontally,

$$\begin{aligned} \mathbf{A}\mathbf{X} &= \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \cdots \\ \mathbf{A}_K \end{bmatrix} [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K] \\ &= \begin{bmatrix} \mathbf{A}_1\mathbf{x}_1 & \mathbf{A}_1\mathbf{x}_2 & \cdots & \mathbf{A}_1\mathbf{x}_K \\ \mathbf{A}_2\mathbf{x}_1 & \mathbf{A}_2\mathbf{x}_2 & \cdots & \mathbf{A}_2\mathbf{x}_K \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_K\mathbf{x}_1 & \mathbf{A}_K\mathbf{x}_2 & \cdots & \mathbf{A}_K\mathbf{x}_K \end{bmatrix}. \end{aligned} \tag{18}$$

Finally, the color image is

$$\mathbf{Z} = \mathbf{W}\mathbf{Q} = \mathrm{diag}(\mathbf{A}\mathbf{X})\,\mathbf{Q}. \tag{19}$$

where $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3] \in \mathbb{R}^{MN \times 3}$.

B. Effect of clipping to non-negative values.

It is necessary to clip negative values in the generated metamer data to ensure the resulting spectra are physically plausible (*i.e.*, no negative spectral radiance). This may lead to slight deviations in the RGB values, and therefore images that are not *exact* metamers. However, we verify that the resulting difference is actually negligible by comparing the projected RGB images from the metamer pairs. For example, in the experiments of Table 5 in the main paper, 32.9% of the generated metamers produce exactly the same RGB images (exact-metamers). Among the remaining 67.1% that are affected by clipping (*i.e.*, near-metamers), the average PSNR between the RGB pairs is 75.8 dB, with a standard deviation of ±17.7 dB. This indicates that the effect of clipping the negative values in the metamer spectra is negligible.

References

[1] Eirikur Agustsson and Radu Timofte. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, pages 126–135, 2017.

[2] Arash Akbarinia and Karl R Gegenfurtner. Color metamerism and the structure of illuminant space. *J. Opt. Soc. Am. A*, 35(4):B231–B238, 2018.

[3] Ali Alsam and Reiner Lenz. Calibrating color cameras using metameric blacks. *J. Opt. Soc. Am. A*, 24(1):11–17, 2007.

[4] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural RGB images. In *Eur. Conf. Comput. Vis.*, pages 19–34. Springer, 2016.

[5] Boaz Arad, Ohad Ben-Shahar, Radu Timofte, L Van Gool, L Zhang, MH Yang, et al. NTIRE 2018 challenge on spectral reconstruction from RGB images. In *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, pages 1042–1042, 2018.

[6] 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 *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, pages 446–447, 2020.

[7] 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 *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, pages 863–881, 2022.

[8] Henry Arguello, Samuel Pinilla, Yifan Peng, Hayato Ikoma, Jorge Bacca, and Gordon Wetzstein. Shift-variant color-coded diffractive spectral imaging system. *Optica*, 8(11):1424–1434, 2021.

[9] Seung-Hwan Baek, Hayato Ikoma, Daniel S Jeon, Yuqi Li, Wolfgang Heidrich, Gordon Wetzstein, and Min H Kim. Single-shot hyperspectral-depth imaging with learned diffractive optics. In *Int. Conf. Comput. Vis.*, pages 2651–2660, 2021.

[10] Seung-Hwan Baek, Incheol Kim, Diego Gutierrez, and Min H Kim. Compact single-shot hyperspectral imaging using a prism. *ACM Trans. Graph.*, 36(6):1–12, 2017.

[11] Bikram Pratap Banerjee, Simit Raval, and PJ Cullen. UAV-hyperspectral imaging of spectrally complex environments. *Int. J. Remote Sens.*, 41(11):4136–4159, 2020.

[12] Mohammad Mahdi Bejani and Mehdi Ghatee. A systematic review on overfitting control in shallow and deep neural networks. *Artif. Intell. Rev.*, pages 1–48, 2021.

[13] L. Belcour, P. Barla, and G. Guennebaud. One-to-many spectral upsampling of reflectances and transmittances. *Comput. Graph. Forum*, 42(4):e14886, 2023.

[14] Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 17502–17511, 2022.

[15] 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 *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, pages 745–755, 2022.

[16] Yuanhao Cai, Yuxin Zheng, Jing Lin, Xin Yuan, Yulun Zhang, and Haoqian Wang. Binarized spectral compressive imaging. *Adv. Neural Inf. Process.*, 36:38335–38346, 2023.

[17] Xuheng Cao, Yusheng Lian, Jin Li, Kaixuan Wang, and Chao Ma. Unsupervised multi-level spatio-spectral fusion transformer for hyperspectral image super-resolution. *Opt. Laser Technol.*, 176:111032, 2024.

[18] Xuheng Cao, Yusheng Lian, Zilong Liu, Jin Li, and Kaixuan Wang. Unsupervised spectral reconstruction from RGB images under two lighting conditions. *Opt. Lett.*, 49(8):1993–1996, 2024.

[19] Xuheng Cao, Yusheng Lian, Kaixuan Wang, Chao Ma, and Xianqing Xu. Unsupervised hybrid network of transformer and CNN for blind hyperspectral and multispectral image fusion. *IEEE Trans. Geosci. Remote Sens.*, 62:1–15, 2024.

[20] Xun Cao, Hao Du, Xin Tong, Qionghai Dai, and Stephen Lin. A prism-mask system for multispectral video acquisition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(12):2423–2435, 2011.

[21] Ayan Chakrabarti and Todd Zickler. Statistics of real-world hyperspectral images. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 193–200. IEEE, 2011.

[22] Liangyu Chen, Xin Lu, Jie Zhang, Xiaojie Chu, and Chengpeng Chen. HINet: Half instance normalization network for image restoration. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 182–192, 2021.

[23] Yong Chen, Wenzhen Lai, Wei He, Xi-Le Zhao, and Jinshan Zeng. Hyperspectral compressive snapshot reconstruction via coupled low-rank subspace representation and self-supervised deep network. *IEEE Trans. Image Process.*, 33:926–941, 2024.

[24] Inchang Choi, Daniel S. Jeon, Giljoo Nam, Diego Gutierrez, and Min H. Kim. High-quality hyperspectral reconstruction using a spectral prior. *ACM Trans. Graph.*, 36(6):218:1–13, 2017.

[25] Inchang Choi, MH Kim, D Gutierrez, DS Jeon, and G Nam. High-quality hyperspectral reconstruction using a spectral prior. *ACM Trans. Graph.*, 36(6):1–13, 2017.

[26] Jozef B Cohen and William E Kappauf. Metameric color stimuli, fundamental metamers, and Wyszecki's metameric blacks. *Am. J. Psychol.*, pages 537–564, 1982.

[27] William J Cukierski and David J Foran. Metamerism in multispectral imaging of histopathology specimens. In *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 145–148. IEEE, 2010.

[28] Laura M Dale, André Thewis, Christelle Boudry, Ioan Rotar, Pierre Dardenne, Vincent Baeten, and Juan A Fernández Pierna. Hyperspectral imaging applications in agriculture and agro-food product quality and safety control: A review. *Appl. Spectrosc. Rev.*, 48(2):142–159, 2013.

[29] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 248–255. Ieee, 2009.

[30] Renwei Dian, Shutao Li, Bin Sun, and Anjing Guo. Recent advances and new guidelines on hyperspectral and multispectral image fusion. *Inf. Fusion*, 69:40–51, 2021.

[31] Renwei Dian, Tianci Shan, Wei He, and Haibo Liu. Spectral super-resolution via model-guided cross-fusion network. *IEEE Trans. Neural Netw. Learn. Syst.*, 2023.

[32] Simone Fabbrizzi, Symeon Papadopoulos, Eirini Ntoutsi, and Ioannis Kompatsiaris. A survey on bias in visual datasets. *Comput. Vis. Image Underst.*, 223:103552, 2022.

[33] Luca Fascione and Johannes Hanika. A study of observer metamerism for reflectance-induced stimuli. 2024.

[34] Mathieu Fauvel, Yuliya Tarabalka, Jon Atli Benediktsson, Jocelyn Chanussot, and James C Tilton. Advances in spectral-spatial classification of hyperspectral images. *Proc. IEEE*, 101(3):652–675, 2012.

[35] Graham D Finlayson and Peter Morovic. Metamer sets. *J. Opt. Soc. Am. A*, 22(5):810–819, 2005.

[36] David H Foster and Kinjiro Amano. Hyperspectral imaging in color vision research: tutorial. *J. Opt. Soc. Am. A*, 36(4):606–627, 2019.

[37] David H Foster, Kinjiro Amano, Sérgio MC Nascimento, and Michael J Foster. Frequency of metamerism in natural scenes. *J. Opt. Soc. Am. A*, 23(10):2359–2372, 2006.

[38] Ying Fu, Tao Zhang, Yinqiang Zheng, Debing Zhang, and Hua Huang. Joint camera spectral response selection and hyperspectral image recovery. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(1):256–272, 2020.

[39] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.

[40] Bernhard Hill. Color capture, color management, and the problem of metamerism: does multispectral imaging offer the solution? In *Proc. SPIE*, volume 3963, pages 2–14. SPIE, 1999.

[41] Bernhard Hill. Optimization of total multispectral imaging systems: best spectral match versus least observer metamerism. In *Proc. SPIE*, volume 4421, pages 481–486. SPIE, 2002.

[42] Xiaowan Hu, Yuanhao Cai, Jing Lin, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. HDNet: High-resolution dual-domain learning for spectral compressive imaging. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 17542–17551, 2022.

[43] Longqian Huang, Ruichen Luo, Xu Liu, and Xiang Hao. Spectral imaging with deep learning. *Light Sci. Appl.*, 11(1):61, 2022.

[44] Junya Ichimura. Optical system and image pickup apparatus having the same, August 2021. US Patent App. 17/174,832.

[45] Roman Jacome, Pablo Gomez, and Henry Arguello. Middle output regularized end-to-end optimization for computational imaging. *Optica*, 10(11):1421–1431, 2023.

[46] Wenzel Jakob and Johannes Hanika. A low-dimensional function space for efficient spectral upsampling. *Comput. Graph. Forum*, 38(2):147–155, 2019.

[47] Daniel S Jeon, Seung-Hwan Baek, Shinyoung Yi, Qiang Fu, Xiong Dun, Wolfgang Heidrich, and Min H Kim. Compact snapshot hyperspectral imaging with diffracted rotation. *ACM Trans. Graph.*, 38(4):1–13, 2019.

[48] Yuhyun Ji, Sang Mok Park, Semin Kwon, Jung Woo Leem, Vidhya Vijayakrishnan Nair, Yunjie Tong, and Young L Kim. mHealth hyperspectral learning for instantaneous spatiospectral imaging of hemodynamics. *Proc. Natl. Acad. Sci.*, 2(4):pgad111, 2023.

[49] Min H Kim. 3D graphics techniques for capturing and inspecting hyperspectral appearance. In *2013 International Symposium on Ubiquitous Virtual Reality*, pages 15–18. IEEE, 2013.

[50] Simon Kornblith, Jonathon Shlens, and Quoc V Le. Do better ImageNet models transfer better? In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2661–2671, 2019.

[51] Sarawak Kuching. The performance of maximum likelihood, spectral angle mapper, neural network and decision tree classifiers in hyperspectral image analysis. *J. Comput. Sci.*, 3(6):419–423, 2007.

[52] 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 *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, pages 462–463, 2020.

[53] Yuqi Li, Qiang Fu, and Wolfgang Heidrich. Multispectral illumination estimation using deep unrolling network. In *Int. Conf. Comput. Vis.*, pages 2672–2681, 2021.

[54] YuJie Liang, Zihan Cao, Shangqi Deng, Hong-Xia Dou, and Liang-Jian Deng. Fourier-enhanced implicit neural fusion network for multispectral and hyperspectral image fusion. *Adv. Neural Inf. Process.*, 37:63441–63465, 2024.

[55] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, pages 136–144, 2017.

[56] Vaibhav Lodhi, Debashish Chakravarty, and Pabitra Mitra. Hyperspectral imaging system: Development aspects and recent trends. *Sens. Imaging*, 20:1–24, 2019.

[57] Guolan Lu and Baowei Fei. Medical hyperspectral imaging: a review. *J. Biomed. Opt.*, 19(1):010901–010901, 2014.

[58] Xiaoyin Mei, Yuqi Li, Qiang Fu, and Wolfgang Heidrich. Progressive self-supervised learning for CASSI computational spectral cameras. *IEEE Trans. Comput. Imaging.*, 2024.

[59] Ziyi Meng, Jiawei Ma, and Xin Yuan. End-to-end low cost compressive spectral imaging with spatial-spectral self-attention. In *Eur. Conf. Comput. Vis.*, pages 187–204. Springer, 2020.

[60] Yusukex Monno, Sunao Kikuchi, Masayuki Tanaka, and Masatoshi Okutomi. A practical one-shot multispectral imaging system using a single image sensor. *IEEE Trans. Image Process.*, 24(10):3048–3059, 2015.

[61] Samuel Ortega, Martin Halicek, Himar Fabelo, Gustavo M Callico, and Baowei Fei. Hyperspectral and multispectral imaging in digital and computational pathology: a systematic review. *Biomed. Opt. Express*, 11(6):3195–3233, 2020.

[62] Bosoon Park, WR Windham, KC Lawrence, and DP Smith. Contaminant classification of poultry hyperspectral imagery using a spectral angle mapper algorithm. *Biosyst. Eng.*, 96(3):323–333, 2007.

[63] Matt Pharr, Wenzel Jakob, and Greg Humphreys. *Physically based rendering: From theory to implementation*. MIT Press, 2023.

[64] Dilip K Prasad and Looi Wenhe. Metrics and statistics of frequency of occurrence of metamerism in consumer cameras for natural scenes. *J. Opt. Soc. Am. A*, 32(7):1390–1402, 2015.

[65] Aneesh Rangnekar, Zachary Mulhollan, Anthony Vodacek, Matthew Hoffman, Angel D Sappa, Erik Blasch, Jun Yu, Liwen Zhang, Shenshen Du, Hao Chang, et al. Semi-supervised hyperspectral object detection challenge results-pbvs 2022. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 390–398, 2022.

[66] Sylvestre-Alvise Rebuffi, Sven Gowal, Dan Andrei Calian, Florian Stimberg, Olivia Wiles, and Timothy A Mann. Data augmentation can improve robustness. *Adv. Neural Inf. Process.*, 34:29935–29948, 2021.

[67] Vishwanath Saragadam and Aswin C Sankaranarayanan. KRISM—Krylov subspace-based optical computing of hyperspectral images. *ACM Trans. Graph.*, 38(5):1–14, 2019.

[68] Zhan Shi, Chang Chen, Zhiwei Xiong, Dong Liu, and Feng Wu. HSCNN+: Advanced CNN-based hyperspectral recovery from RGB images. In *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, pages 939–947, 2018.

[69] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *J. Big Data*, 6(1):1–48, 2019.

[70] Gilbert Strang. *Introduction to linear algebra*. SIAM, 2022.

[71] William A Thornton. How strong metamerism disturbs color spaces. *Color Res. Appl.*, 23(6):402–407, 1998.

[72] Antonio Torralba and Alexei A Efros. Unbiased look at dataset bias. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1521–1528. IEEE, 2011.

[73] Sergio Urrea, Roman Jacome, M Salman Asif, Henry Arguello, and Hans Garcia. DoDo: Double doe optical system for multishot spectral imaging. *IEEE Journal of Selected Topics in Signal Processing*, 2024.

[74] Mark Van De Ruit and Elmar Eisemann. Metameric: Spectral uplifting via controllable color constraints. In *ACM SIGGRAPH 2023 Conference Proceedings*, pages 1–10, 2023.

[75] Freek Van der Meer. The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery. *Int. J. Appl. Earth Obs. Geoinf.*, 8(1):3–17, 2006.

[76] Cornelius van Trigt. Metameric blacks and estimating reflectance. *J. Opt. Soc. Am. A*, 11(3):1003–1024, 1994.

[77] Françoise Viénot and Hans Brettel. The verriest lecture: Visual properties of metameric blacks beyond cone vision. *J. Opt. Soc. Am. A*, 31(4):A38–A46, 2014.

[78] Gemine Vivone. Multispectral and hyperspectral image fusion in remote sensing: A survey. *Inf. Fusion*, 89:405–417, 2023.

[79] Ashwin Wagadarikar, Renu John, Rebecca Willett, and David Brady. Single disperser design for coded aperture snapshot spectral imaging. *Appl. Opt.*, 47(10):B44–B51, 2008.

[80] Hongyuan Wang, Lizhi Wang, Chang Chen, Xue Hu, Fenglong Song, and Hua Huang. Learning spectral-wise correlation for spectral super-resolution: Where similarity meets particularity. In *ACM Int. Conf. Multimedia*, pages 7676–7685, 2023.

[81] Lizhi Wang, Lingen Li, Weitao Song, Lei Zhang, Zhiwei Xiong, and Hua Huang. Non-serial quantization-aware deep optics for snapshot hyperspectral imaging. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2024.

[82] Lizhi Wang, Zhiwei Xiong, Guangming Shi, Feng Wu, and Wenjun Zeng. Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(10):2104–2111, 2016.

[83] Xinying Wang, Zhixiong Huang, Sifan Zhang, Jiawen Zhu, Paolo Gamba, and Lin Feng. GMSR: Gradient-integrated mamba for spectral reconstruction from RGB images. *Neural Netw.*, page 108020, 2025.

[84] Andrea Weidlich, Alex Forsythe, Scott Dyer, Thomas Mansencal, Johannes Hanika, Alexander Wilkie, Luke Emrose, and Anders Langlands. Spectral imaging in production: course notes SIGGRAPH 2021. In *ACM SIGGRAPH 2021 Courses*, pages 1–90. Association for Computing Machinery (ACM), 2021.

[85] Roman Werpachowski, András György, and Csaba Szepesvári. Detecting overfitting via adversarial examples. *Adv. Neural Inf. Process.*, 32, 2019.

[86] Chaoxiong Wu, Jiaojiao Li, Rui Song, Yunsong Li, and Qian Du. HPRN: Holistic prior-embedded relation network for spectral super-resolution. *IEEE Trans. Neural Netw. Learn. Syst.*, 2023.

[87] Yaohang Wu, Renwei Dian, and Shutao Li. Multistage spatial–spectral fusion network for spectral super-resolution. *IEEE Trans. Neural Netw. Learn. Syst.*, 2024.

[88] Günter Wyszecki. Evaluation of metameric colors. *J. Opt. Soc. Am. A*, 48(7):451–454, 1958.

[89] Qi Xie, Minghao Zhou, Qian Zhao, Deyu Meng, Wangmeng Zuo, and Zongben Xu. Multispectral and hyperspectral image fusion by ms/hs fusion net. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1585–1594, 2019.

[90] Hao Xu, Shiqi Chen, Haiquan Hu, Peng Luo, Zheyan Jin, Qi Li, Zhihai Xu, Huajun Feng, Yueting Chen, and Tingting Jiang. Wavelength encoding spectral imaging based on the combination of deeply learned filters and an rgb camera. *Opt. Express*, 32(7):10741–10760, 2024.

[91] Meng Xu, Mingying Lin, Qi Ren, and Sen Jia. SSTHyper: Sparse spectral transformer for hyperspectral image reconstruction. In *ACCV*, pages 1918–1935, 2024.

[92] Nan Xu, Hao Xu, Shiqi Chen, Haiquan Hu, Zhihai Xu, Huajun Feng, Qi Li, Tingting Jiang, and Yueting Chen. Snapshot hyperspectral imaging based on equalization designed DOE. *Opt. Express*, 31(12):20489–20504, 2023.

[93] Jun Yan, Kai Zhang, Qinzhu Sun, Chiru Ge, Wenbo Wan, Jiande Sun, and Huaxiang Zhang. Spatial–spectral unfolding network with mutual guidance for multispectral and hyperspectral image fusion. *Pattern Recognition*, 161:111277, 2025.

[94] Xinge Yang, Qiang Fu, Mohamed Elhoseiny, and Wolfgang Heidrich. Aberration-aware depth-from-focus. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2023.

[95] Zhiyang Yao, Shuyang Liu, Xiaoyun Yuan, and Lu Fang. SpeCAT: Spatial-spectral cumulative-attention transformer for high-resolution hyperspectral image reconstruction. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 25368–25377, 2024.

[96] Fumihito Yasuma, Tomoo Mitsunaga, Daisuke Iso, and Shree K Nayar. Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum. *IEEE Trans. Image Process.*, 19(9):2241–2253, 2010.

[97] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5728–5739, 2022.

[98] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *Eur. Conf. Comput. Vis.*, pages 492–511. Springer, 2020.

[99] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 14821–14831, 2021.

[100] Haijin Zeng, Yuxi Liu, Yongyong Chen, Youfa Liu, Chong Peng, and Jingyong Su. SAH-SCI: Self-supervised adapter for efficient hyperspectral snapshot compressive imaging. In *Eur. Conf. Comput. Vis.*, pages 311–328. Springer, 2024.

[101] Chong Zhang, Wenjing Liu, Juntao Li, Siqi Li, Lizhi Wang, Hua Huang, Yuanjin Zheng, Yongtian Wang, Jinli Suo, and Weitao Song. Tunable optimally-coded snapshot hyperspectral imaging for scene adaptation. *Laser & Photonics Reviews*, page 2401921, 2025.

[102] Jingang Zhang, Runmu Su, Qiang Fu, Wenqi Ren, Felix Heide, and Yunfeng Nie. A survey on computational spectral reconstruction methods from RGB to hyperspectral imaging. *Sci. Rep.*, 12(1):11905, 2022.

[103] Mingjin Zhang, Longyi Li, Wenxuan Shi, Jie Guo, Yunsong Li, and Xinbo Gao. VmambaSCI: Dynamic deep unfolding network with mamba for compressive spectral imaging. In *ACM Int. Conf. Multimedia*, pages 6549–6558, 2024.

[104] Xiandou Zhang, Brian Funt, and Hamidreza Mirzaei. Metamer mismatching in practice versus theory. *J. Opt. Soc. Am. A*, 33(3):A238–A247, 2016.

[105] Yuzhi Zhao, Lai-Man Po, Qiong Yan, Wei Liu, and Tingyu Lin. Hierarchical regression network for spectral reconstruction from RGB images. In *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, pages 422–423, 2020.