Computational Spectral Imaging with Unified Encoding Model and Beyond

Xinyuan Liu Lingen Li Lin Zhu Lizhi Wang*
Beijing Institute of Technology
{xinyuanliu, lingenli, linzhu, wanglizhi}@bit.edu.cn

Abstract

Computational spectral imaging is drawing increasing attention owing to the snapshot advantage, and amplitude, phase, and wavelength encoding systems are three types of representative implementations. Fairly comparing and understanding the performance of these systems is essential, but challenging due to the heterogeneity in encoding design. To overcome this limitation, we propose the unified encoding model (UEM) that covers all physical systems using the three encoding types. Specifically, the UEM comprises physical amplitude, physical phase, and physical wavelength encoding models that can be combined with a digital decoding model in a joint encoder-decoder optimization framework to compare the three systems under a unified experimental setup fairly. Furthermore, we extend the UEMs to ideal versions, namely, ideal amplitude, ideal phase, and ideal wavelength encoding models, which are free from physical constraints, to explore the full potential of the three types of computational spectral imaging systems. Finally, we conduct a holistic comparison of the three types of computational spectral imaging systems and provide valuable insights for designing and exploiting these systems in the future.

1. Introduction

Spectral imaging is a valuable technique that captures both spatial and spectral information of a scene along a specific range of wavelengths in the electromagnetic spectrum. The analysis of the information facilitates material discrimination and object classification. Consequently, spectral imaging has found applications in various fields, such as biology [36], medicine [70], and agriculture [43]. Conventional spectral imaging systems are bulky, require time-consuming scanning to acquire spatial and spectral information, and are difficult to use in dynamic scenes. In contrast, computational spectral imaging systems capture spectral scenes into a single snapshot measurement using well-designed optical encoders and then reconstruct them using corresponding digital decoders. Therefore, computational spectral imaging is the most promising solution to capture spectral information of dynamic scenes without scanning efficiently.

Currently, physically implementable computational spectral imaging systems can be divided into three representative categories based on the encoding method: Amplitude encoding systems based on compressive sensing [5, 24, 40], phase encoding systems that rely on diffractive optics [26, 28, 31], and wavelength encoding systems that employ filters [48, 52]. Research on the three kinds of computational spectral imaging systems has been limited to specific categories, resulting in the absence of a comprehensive understanding of the technology. Moreover, due to inherent differences among these systems, artificial optics specifically designed for one cannot be transferred to another, thus presenting challenges for fair comparisons. Therefore, conducting comprehensive, impartial, and inclusive comparative investigations within a unified context is urgently needed to gain a broader perspective and thorough understanding of these systems.

In this paper, we propose the Unified Encoding Model (UEM) to comprehensively compare the performance of the three types of spectral imaging systems under a unified framework. Representative spectral imaging systems can be modeled using physical coding techniques such as amplitude (AEM-P), phase (PEM-P), or wavelength (WEM-P). We choose three typical physical systems: SD-CASSI as AEM-P, Rotationally Symmetric DOE as PEM-P, and Selected Filter as WEM-P.

Moreover, to fully explore the potential performance of the three types of computational spectral imaging systems in our comparative study, we propose ideal models with greater degrees of freedom that eliminate the constraints of physical models. These ideal models, including the ideal amplitude encoding model (AEM-I), ideal phase encoding model (PEM-I), and ideal wavelength encoding model (WEM-I), offer an opportunity to push the performance limits of each type of spectral imaging system. The AEM-I model enables the free modulation of amplitude to any possible value (i.e., 32-bit floating point number) in both spatial

*Corresponding author: Lizhi Wang.

and spectral dimensions, the PEM-I model enables any form of the point spread function (PSF), and the WEM-I model allows any shape of the spectral response curve. The encoding properties obtained through the high-degree-of-freedom optimization process provide valuable guidance for computational spectral imaging.

Furthermore, We construct an end-to-end computational spectral imaging framework consisting of the UEMs and three types of decoding models. To ensure a fair comparison between the three types of computational spectral systems, we employ a joint encoder-decoder optimization approach [56] to optimize the three casted UEMs and corresponding decoding models for the same spectral imaging objective.

In summary, our contributions are as follows:

- We propose the UEM for the three physically implementable computational spectral imaging systems, enabling a fair comparison under the unified framework.
- We extend the UEM to the ideal version with increased degrees of freedom, eliminating the physical constraints of the three models. The ideal UEM allows for a thorough exploration of the potential performance of the three types of computational spectral imaging systems.
- We construct a computational spectral imaging framework with the UEMs and decoding models, using a joint encoder-decoder optimization approach to compare three types of computational spectral systems.

2. Related Work

Amplitude Encoding System. The amplitude encoding system is based on the theory of compressive sensing [9, 17, 66], which combines optics, mathematics, and optimization theory. This system employs an encoding mask to block or filter the input light and then passes through the dispersion element to acquire an image with compressed information [2, 10, 18, 19, 44, 50]. The image generation is expressed as a point multiplication process between a mask that is consistent with the image size and a monochromatic object image for each wavelength. Finally, a specific reconstruction algorithm decodes and obtains the underlying spectral image.

HyperReconNet [67] jointly learns the coded mask and the corresponding CNN for reconstruction. The BinaryConnect [15] adds the encoded mask to the network as a layer, using a floating-point mapping to the binary-encoded mask entity. E. Salazar et al. [55] proposed an optimal coded mask algorithm for the Spatial Spectral Compressive Spectral Imager (SSCSI) based on a discrete measurement model of his design. D2UF [30] encoded the mask and multispectral color filter array of the coded aperture snapshot spectral imager (CASSI) and performed spectral reconstruction using an unrolling-based network.

Phase Encoding System. The phase encoding system is based on diffractive imaging [28], employing diffractive optical elements(DOEs) [62] or metasurfaces [47]. These elements are specifically designed to control the phase modulation in terms of the PSF to act as encoders that collect desired spectral information.

Several research efforts have been made to improve the phase encoding system. Peng et al. [51] proposed a shape-invariant PSF design approach for high-quality color imaging. In contrast, Jeon et al. [32] designed a spectrally varying PSF that rotates regularly with wavelength and encodes spectral information. Xiong et al. [20] decomposed the PSF formula into Bessel functions, simplified these functions according to rotational symmetry, and reconstructed the network using a W-Net composed of Res-U-Net units. Arguello et al. [4] introduced a color-coded aperture and designed the spatially variable PSF.

To address the quantization issue of DOE height maps, Li et al. [35] integrated the quantization operation into the DOE optimization process, using an adaptive mechanism to optimize the physical height of each volume layer. Tseng et al. [63], and Makarenko et al. [42] used metasurface instead of DOE as the encoding element, achieving excellent results.

Wavelength Encoding System. The wavelength encoding system uses RGB narrowband filters or other band filters directly to encode the scene in the wavelength dimension and perform spectral reconstruction of the captured image. [25, 57, 60] Wavelength encoding systems produce better spatial quality in reconstructed results due to the absence of modulation on spatial dimensions.

The prevalent wavelength encoding system directly regards the imaging pipeline of an arbitrary or well-selected RGB camera as the encoding process, constructing reconstruction algorithms from RGB images. Xiong et al. [68] proposed HSCNN, which applied CNN to upsampled RGB and amplitude-encoded measurements for spectral reconstruction. Galliani et al. [23] proposed a learned spectral super-resolution using CNN for end-to-end RGB to spectral image reconstruction. The current popular attention mechanism [64] has also been applied in spectral imaging. Cai et al. [7] proposed Transformer-based MST++, which uses cells based on HSI spatial sparsity and spectral self-similarity properties as the base module.

Admittedly, existing RGB narrowband filters are designed to satisfy human vision and are intuitively suboptimal for spectral imaging. Therefore, some researchers chose filters to achieve a suitable spectral response for spectral imaging [22], and some optimized the response to obtain better spectral imaging performance. Nie et al. [49] encoded the RGB filter and combined this function with the reconstruction network for end-to-end optimization on the spectral reconstruction task.

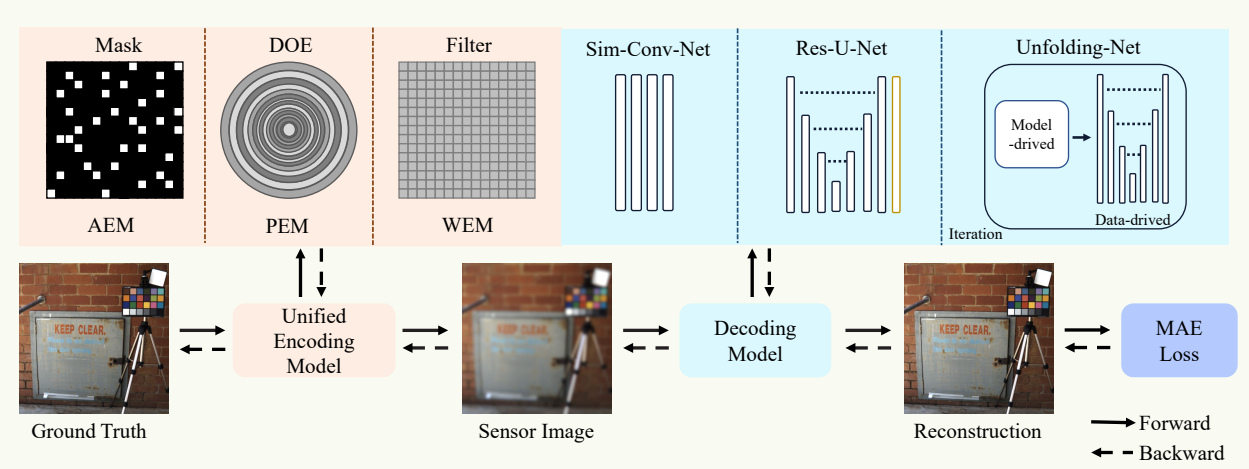


Figure 1. Joint encoder-decoder optimization framework optimizes the UEM and the decoding model toward the same objective.

Table 1. Physical systems corresponding to each physical UEM.

Encoding Model	Physical Systems
AEM-P	CASSI [3, 13, 38, 65], SSCSI [55]
PEM-P	Rotationally Symmetric DOE [20], Free DOE [6]
WEM-P	Selected Filter [21], Learned Filter [49]

Joint Encoder-decoder Optimization. With the advancements in computational photography, joint encoder-decoder optimization methods have become the mainstream technique to enhance imaging performance. Instead of designing the encoder and the decoder separately like traditional approaches, the joint encoder-decoder optimization algorithm optimizes both the encoder and decoder simultaneously for a given task. The superior performance of the joint encoder-decoder optimization method has been demonstrated in various computational imaging tasks such as HDR imaging [45, 59], depth estimation [6, 29, 39] and computational spectral imaging [20, 35]. However, joint encoder-decoder optimization in computational spectral imaging is still limited to specific categories, lacking comprehensive understanding.

3. Unified Encoding Model

3.1. Overview

We propose the UEM that covers all the possible types of optical modulations for computational spectral imaging, which can be cast into AEM, PEM, and WEM to represent amplitude, phase, and wavelength encoding systems, respectively. Specifically, for the amplitude modulation, the scene light is element-wise multiplied with the matrix $A(x, y, \lambda)$ of the encoding mask; For phase modulation, the light is convolved with the PSF $P(x, y, \lambda)$ of the DOE; For wavelength modulation, the light is multiplied by the spectral response $W(\lambda)$ of the filter, integrated along the wavelength dimension λ, and converted into a measurement I_{rgb} on the RGB sensor. We give the overall formula for the UEM as follows, with some terms becoming unit terms under specific settings:

$$I_{rgb} = \int W(\lambda) \cdot P(x, y, \lambda) * (A(x, y, \lambda) \cdot I(x, y, \lambda)) d\lambda, \tag{1}$$

where $*$ is the convolution operator, $\cdot$ is dot product operation, $I(x, y, \lambda)$ denotes the natural spectral scene, and I_{rgb} is the RGB encoded image captured by the sensor.

In the following, we introduce two versions of UEMs: the physical and the ideal versions. The physical UEM is modeled with physical constraints in the encoding process, corresponding to a series of physically implementable encoding systems for spectral imaging. Based on the physical UEM, we can put the three types of physically implementable spectral imaging systems into one unified setting for comparative study. Beyond, the ideal UEMs are extended from the physical UEMs by eliminating the physical constraints of the encoding process, which means the ideal UEMs have higher degrees of encoding freedom. We employ the ideal UEMs to investigate the maximum potential of each type of spectral imaging system. For convenience, we use the suffix '-P' to denote the physical UEMs and the suffix '-I' to indicate ideal UEMs.

3.2. Physical Model

We first introduce the physical UEMs and cast them to AEM-P, PEM-P, and WEM-P for physically implementable amplitude, phase, and wavelength encoding systems, respectively. The representative physical systems that correspond to each physical UEM are presented in Table 1.

Physical Amplitude Encoding Model. The AEM-P only employs the amplitude modulation part of the UEM, with

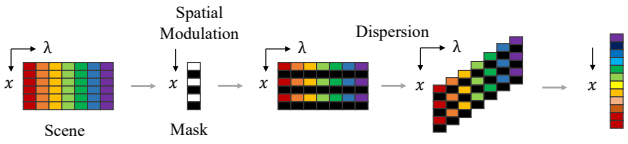


Figure 2. Optical encoding propagation of AEM-P.

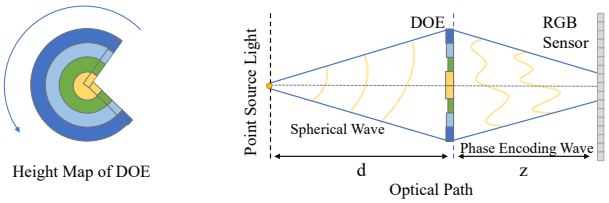


Figure 3. Encoding parameterization (height map) and the optical path of PEM-P.

physical constraints on the modulation that should be implementable using a mask.

CASSI [5, 14, 37, 40, 54, 61, 65] stands as a widely employed physical amplitude encoding system. Employing a binary encoding mask to modulate the spatial amplitude of incident light, the system utilizes a dispersion element introducing distinct offsets for various wavelengths, effectively compressing spectral information along the spatial dimension.

AEM-P, an extension of the classical SD-CASSI, employs a binary spatial encoding mask and dispersion elements in the encoding process, illustrated in Figure 2. Here, we introduce a learnable parameter representing the spatial dimension of the image as a mask, adopting values of 0 or 1 to signify the blocking or passing of light through the mask. The dispersion operation then shifts spatial information upward by one unit in each spectral band of the spectral dimension. Specifically, the $A(x, y, \lambda)$ of AEM-P encodes the unit shift of the mask upwards layer by layer in the spectral dimension.

$$A(x, y, \lambda) = Mask_P(x, y, \theta_{mask_p}) \cdot Dispersive(\lambda), \quad (2)$$

where, the $Mask_P$ denotes the spatial dimensional encoding mask, the θ_{mask_p} is a learnable parameter, and the $Dispersive(\lambda)$ represents the dispersion operation.

The imaging process of AEM-P can be written based on UEM as follows:

$$I_{rgb} = \int W(\lambda) \cdot (A(x, y, \lambda) \cdot I(x, y, \lambda))d\lambda. \quad (3)$$

Physical Phase Encoding Model. The PEM-P controls the height map or metasurface structure of the DOE to produce different PSFs, which describe the image-blurring effect by point light source. The resulting phase delay is used as phase modulation. Following the standard settings, we model PEM-P based on a rotationally symmetric DOE, which uses a one-dimensional learnable encoding parameter to represent the entire DOE height map. The one-dimensional parameter is rotated and filled to obtain the final DOE height map, as shown in Figure 3. The parameterization reduces both the number of learnable parameters and manufacturing complexity. The height map $H(x, y)$ of a rotationally symmetric DOE can be represented as follows:

$$H(x, y) = H_P(x, y, \theta_{heightmap}), \quad (4)$$

where the H_P is the height map, and the $\theta_{heightmap}$ is a one-dimensional learnable parameter.

The PSF $P(x, y, \lambda)$ can be derivated from the height map $H(x, y)$ according to diffraction law. For the detailed derivation, please refer to the supplementary materials for additional information. Then, the imaging process of PEM-P can be written based on UEM:

$$I_{rgb} = \int W(\lambda) \cdot (P(x, y, \lambda) * I(x, y, \lambda))d\lambda. \quad (5)$$

Physical Wavelength Encoding Model. The WEM-P algorithm selects the best spectral response curve from a database while keeping other system components fixed, ensuring optimal performance for non-learnable spectral response curves [21].

$$W(\lambda) = \{w_i | w_i \in \mathbf{W}, w_i \; is \; the \; most \; suited\}, \quad (6)$$

where $w_1, w_2, w_3, ... \in \mathbf{W}$ represents a fixed response function dataset.

The imaging process of WEM-P can be written from UEM by eliminating unused terms:

$$I_{rgb} = \int W(\lambda) \cdot I(x, y, \lambda)d\lambda. \quad (7)$$

3.3. Ideal Model

We then introduce the ideal UEMs, AEM-I, PEM-I, and WEM-I, free from any physical constraint in the physical UEMs.

Ideal Amplitude Encoding Model. Compared to AEM-P, which uses binary spatial dimensional encoding, we design the AEM-I using floating point encoding to encode spatial and spectral dimensional encoding mask $A_I(x, y, \lambda)$. We construct the mask with a learnable parameter consistent with the image size as a spatial-spectral co-encoding.

The imaging model of AEM-I is obtained by replacing $A(x, y, \lambda)$ in Eq. (3) with $A_I(x, y, \lambda)$.

$$A_I(x, y, \lambda) = Mask_I(x, y, \lambda, \theta_{mask}), \quad (8)$$

where the $Mask_I$ is the spatial and spectral dimensional encoding mask, the θ_{mask} is a learnable parameter.

Ideal Phase Encoding Model. We encode the PSF $P_I(x, y, \lambda)$ directly for the PEM-I without considering the structure of the DOE of PEM-P. We construct a learnable parameter consistent with the image channel dimension as a freely encoding PSF with a spatial size less than or equal to the spatial size of the image. The imaging model of PEM-I is obtained by replacing $P(x, y, \lambda)$ in Eq. (5) with $P_I(x, y, \lambda)$.

$$P_I(x, y, \lambda) = PSF_I(x, y, \lambda, \theta_{psf}), \tag{9}$$

where the PSF_I is the encoding PSF, the θ_{psf} is a learnable parameter.

Ideal Wavelength Encoding Model. When designing filter surface coatings, customizing the spectral curve must consider filter material properties. Unlike WEM-P, our WEM-I encodes the spectral curves directly with a learnable parameter but does not consider the material limitations. The imaging model of PEM-I is obtained by replacing $W(\lambda)$ in Eq. (7) with $W_I(\lambda)$.

$$W_I(\lambda) = Response_I(\lambda, \theta_{response}), \tag{10}$$

where the $Response_I$ is the encoding response function, and the $\theta_{response}$ is a learnable parameter.

4. Joint Optimization

This section introduces the decoding models we use and the optimization objective in the joint optimization framework for the comparative study.

Decoding Model. Since we mainly explore the performance of several encoding models rather than the decoding models, we design three typical reconstruction networks as decoding models. These include a Sim-Conv-Net which is a simple CNN [33, 34, 41], Res-U-Net [27, 74] which is a variant of the popular U-Net [46, 53] and Unfolding-Net which uses deep unfolding methods [8, 16, 58, 71–73, 75] as the reconstruction network. For the detailed network structure, please refer to the supplementary materials.

Joint Optimization Framework. We combine the UEM with the decoding model to compose a joint encoder-decoder optimization framework. In UEM, the scene light passes through the mask to get the amplitude modulation $f_A(I, \theta_A)$, passes through the DOE to get the phase modulation $f_P(I, \theta_P)$, passes through the filter and sensor to get the wavelength modulation $f_W(I, \theta_W)$. We can describe the overall imaging process as follows:

$$I_{rgb} = f_W(f_P(f_A(I, \theta_A), \theta_P), \theta_W), \tag{11}$$

where θ_A,θ_P, and θ_W are learnable parameters for the AEM, the PEM, and the WEM. For each system only optimizes one parameter, e.g., for AEM systems, only θ_A is optimized. After that, the decoding model reconstructs the spectral images from the RGB measurement I_{rgb}, and this process can be written as:

$$\tilde{I} = f_D(I_{rgb}, \theta_D), \tag{12}$$

where the θ_D is the learnable parameter of the decoding model.

Then, the optimization objective of the UEM is to find a set of $\theta_A, \theta_P, \theta_W, \theta_D$ that minimizes the mean absolute error (MAE) loss between the reconstructed spectral images and the ground truth:

$$\theta_A, \theta_P, \theta_W, \theta_D = \underset{\theta_A, \theta_P, \theta_W, \theta_D}{\arg\min} \sum \left\| \tilde{I} - I \right\|_1. \tag{13}$$

Since each stage of the UEM is differentiable, we can solve this problem using the first-order gradient-based optimization methods. The irrelevant parameter will be eliminated when we cast the UEM to a specific encoding model. That is, we only optimize θ_A, θ_D for AEMs, θ_P, θ_D for PEMs, and θ_W, θ_D for WEMs.

Finally, we combine the UEM, decoding model, and optimization objective to build a joint encoder-decoder optimization framework, as shown in Figure 1.

5. Experiments and Analysis

5.1. Experimental Setup

In our experiments, the Sim-Conv-Net serves as the primary decoding model, and we establish a baseline by employing nearest neighbor interpolation from RGB images to the spectral image. The camera response curve of the FLIR BFS_U3_04S2C_C is utilized in experiments where the response is fixed (AEM-P/I, PEM-P/I, and WEM-P). Response selection of WEM-P experiments are detailed in the supplementary materials.

The initialization of encoding is crucial in the ideal model. A comprehensive description of the initialization setup, along with experimental demonstration, is provided in the supplementary materials.

Our joint encoder-decoder optimization framework undergoes evaluation on four datasets: ICVL[1], Harvard [11], CAVE [69], and KAIST [12], with cross-validation. The dataset sizes are as follows: ICVL has 167 training, 17 validation, and 17 testing images. Harvard has 40 training, 4 validation, and 6 testing images. CAVE has 28 training, 1 validation, and 3 testing images. Lastly, KAIST has 30 testing images exclusively. All images are cropped to 512 × 512 patches.

We use MAE Loss as the main result presentation in our experiments, and the experimental results for mean square error (MSE) loss and ERGAS loss are the supplementary materials.

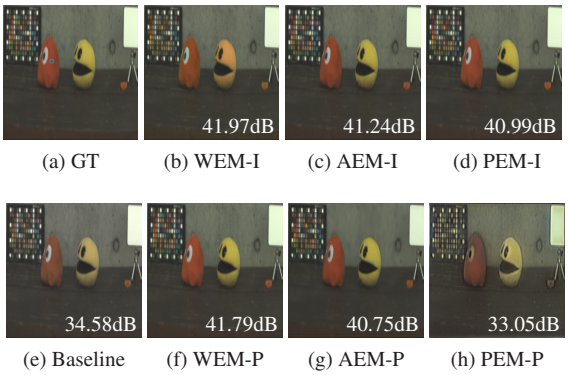
We consider peak signal-to-noise ratio (PSNR), PSNR-SI, spectral angle mapping (SAM), and eror relative global

Table 2. Spectral imaging performance of systems using UEMs on different datasets, with the best results in bold.

	Physical Model					Ideal Model				
Datasets	**Encoding Model**	**PSNR↑**	**PSNR-SI↑**	**SAM↓**	**ERGAS↓**	**Encoding Model**	**PSNR↑**	**PSNR-SI↑**	**SAM↓**	**ERGAS↓**
ICVL (training & testing)	**WEM-P**	44.96	38.81	0.0399	6.23	**WEM-I**	**45.25**	38.51	**0.0328**	**5.23**
	AEM-P	40.21	32.66	0.0491	8.96	**AEM-I**	45.08	**38.95**	0.0405	6.24
	PEM-P	34.07	26.35	0.0669	16.16	**PEM-I**	44.9	38.68	0.0398	6.28
	Baseline	32.21	28.98	0.1342	33.3					
Harvard (training & testing)	**WEM-P**	38.72	36.91	0.07538	8.01	**WEM-I**	**39.9**	36.31	**0.05152**	8.09
	AEM-P	32.7	30.01	0.0852	14.04	**AEM-I**	38.84	**36.88**	0.07513	**7.97**
	PEM-P	27.24	24.42	0.10196	25.17	**PEM-I**	38.71	**36.88**	0.07561	8.12
	Baseline	26.28	24.87	0.21845	143.05					
CAVE (training & testing)	**WEM-P**	36.02	37.57	0.1961	18.38	**WEM-I**	**37.65**	37.98	0.1682	17.48
	AEM-P	33.13	33.13	0.2193	26.11	**AEM-I**	36.96	**38.00**	**0.1590**	**16.85**
	PEM-P	28.99	28.76	0.2974	48.23	**PEM-I**	35.77	37.55	0.1894	18.77
	Baseline	29.50	30.36	0.3906	41.74					
CAVE (training), **KAIST** (testing)	**WEM-P**	32.02	33.58	0.4819	36.82	**WEM-I**	33.04	31.93	0.4363	43.45
	AEM-P	32.21	33.69	0.4707	36.22	**AEM-I**	**33.30**	**34.42**	**0.3679**	**31.82**
	PEM-P	28.39	26.20	0.4909	70.45	**PEM-I**	30.68	29.60	0.4541	48.82
	Baseline	30.78	31.06	0.6339	46.47					

Table 3. Spectral imaging performance of systems using UEMs on different decoding models, with the best results in bold.

	Physical Model					Ideal Model				
Decoding Model	**Encoding Model**	**PSNR↑**	**PSNR-SI↑**	**SAM↓**	**ERGAS↓**	**Encoding Model**	**PSNR↑**	**PSNR-SI↑**	**SAM↓**	**ERGAS↓**
Res-U-Net	**WEM-P**	38.62	32.00	8.07	0.0373	**WEM-I**	**40.54**	**33.96**	**6.73**	**0.0360**
	AEM-P	36.72	29.53	10.39	0.0422	**AEM-I**	39.28	32.80	7.51	0.0396
	PEM-P	34.87	28.10	13.08	0.0654	**PEM-I**	38.57	31.83	8.15	0.0385
Unfolding-Net	**WEM-P**	45.09	38.93	4.22	0.0289	**WEM-I**	**46.43**	**40.86**	**3.69**	**0.0257**
	AEM-P	44.63	38.55	4.43	0.0295	**AEM-I**	46.33	40.30	3.71	0.0280
	PEM-P	38.70	31.80	8.69	0.0514	**PEM-I**	45.94	40.53	3.84	0.0286



(a) GT (b) WEM-I (c) AEM-I (d) PEM-I

(e) Baseline (f) WEM-P (g) AEM-P (h) PEM-P

Figure 4. Visual comparison of reconstruction images, the PSNR metrics are labeled in the lower left corner.

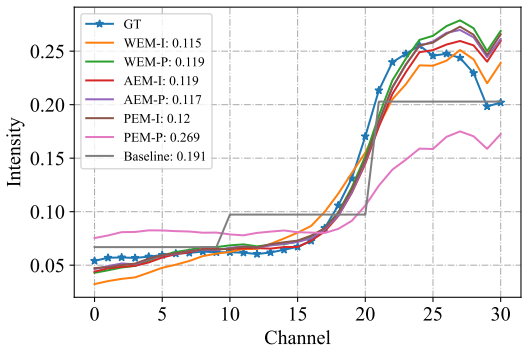


Figure 5. Spectral curve comparison at a single sampling point, the SAM metrics for this point are labeled in the legend.

dimensionless synthesis (ERGAS) for evaluation metrics. PSNR-SI is the average PSNR of each channel of the spectral image, with intensity set to the channel maximum of the corresponding ground truth.

5.2. Results

Comparative Experiments on Different Datasets. Table 2 presents a comprehensive comparison of the spectral imaging performance of the system, incorporating UEMs and Sim-Conv-Net, across different datasets. The imaging effects of various UEM sub-models are carefully examined, revealing consistent regularity. WEM-P outperforms AEM-P among physical models, while PEM-P exhibits relatively inferior performance. Similarly, in the ideal model, WEM-I imaging is superior. Given that WEM-P already has the best performance in the physical model, the WEM-I brings relatively little improvement. In contrast, AEM-P and PEM-P show great potential for improvement when transitioning to the ideal model.

To evaluate the generalization of our systems, we train on the CAVE dataset and test on the KAIST dataset, using only

Table 4. Comparison of the imaging quality of systems using different types of WEM, where the WEM w/ P.C. represents the wavelength encoding model with positive constraints. The best result is in bold, and the next best is underlined.

Encoding Model	PSNR↑	PSNR-SI↑	SAM↓	ERGAS↓
WEM-I	**45.25**	38.51	**0.0328**	**5.23**
WEM-P	44.96	38.81	0.0399	6.23
WEM-I w/ P.C.	44.26	**38.84**	0.0432	5.37

Table 5. Comparison of the imaging quality of systems using different ideal models of UEM and Sim-Conv-Net, with the best results in bold and the next best results underlined.

Encoding Model	PSNR↑	PSNR-SI↑	SAM↓	ERGAS↓
UEM-I	**45.30**	38.48	0.0372	5.63
WEM-I	45.25	38.51	**0.0328**	**5.23**
AEM-I	45.08	**38.95**	0.0405	6.24
PEM-I	44.9	38.68	0.0398	6.28

overlapping bands for both training and testing. Significantly, AEMs demonstrate superior generalization, whereas WEM-P and PEM-I, when trained on CAVE, exhibit noticeable performance dips when tests on KAIST. Additional validation experiments on ICVL are detailed in the supplementary materials.

RGB visualizations of the reconstructed spectral images for all models are displayed in Figure 4. Although WEM-I performs better numerically, we notice slight color inaccuracies in some instances, such as a dominant orange shift (Figure 4b). We attribute this shift to the negative values in the response curve of the free encoding. PEM-P consistently shows inferior reconstruction, as reflected in color and image detail, in line with numerical metrics.

In addition to the numerical and visual evaluations, we select a point in Figure 4 (marked in the blue-circled area in Figure 4a) to obtain the spectral accuracy of each model for comparing the spectral accuracies of the models, as shown in Figure 5. The results indicate that the WEM-I model curve aligns with the ground truth most precisely, while the PEM-P model is less accurate than the baseline. The other physical and idea models display relatively close curves around the ground truth, consistent with the numeric metrics.

Experiments on Different Decoding Model. In this experiment, we replace the Sim-Conv-Net decoding model with two different reconstruction networks: the Res-U-Net and the Unfolding-Net. Table 3 summarizes the results.

Contrary to expectations that increasing network depth and complexity would enhance image reconstruction, Res-U-Net results in lower metrics in our end-to-end optimization experiments. Except for WEM-I, the tables display those models utilizing Sim-Conv-Net have higher metrics sensitive to spatial differences like PSNR. In contrast, models using Res-U-Net as the decoding model tend to have superior metrics sensitive to spectral errors like SAM. Only PEM-P shows enhanced imaging performance when applying Res-U-Net compared to Sim-Conv-Net.

Unfolding-Net incorporates optical encoding information into the inference process and achieves superior decoding performance by optimizing model-driven and data-driven modules iteratively. Compared to Sim-Conv-Net, physical model systems demonstrate significantly improved performance, while ideal model systems exhibit only marginal gains.

Training a large network structure in an end-to-end optimized computational spectral imaging system may be challenging. Thus, we suggest utilizing simpler convolutional neural networks, particularly in situations with limited computing power. Moreover, we advise that simpler optical encoding processes are more compatible with simple networks, while more complex encoding models should prefer deeper networks.

Full Optimization. We combine UEM-I with a neural network as a decoding model to construct a fully optimized ideal computational spectral imaging system. This system can simultaneously optimize mask, PSF, spectral response, and neural networks.

As shown in Table 5, UEM-I outperforms WEM-I in PSNR metrics only, and there is no significant improvement in imaging performance compared to single-mode free encoding. Therefore, we suggest that restricting optimization to a single free encoding is adequate for approaching the upper-performance limit of all free encoding.

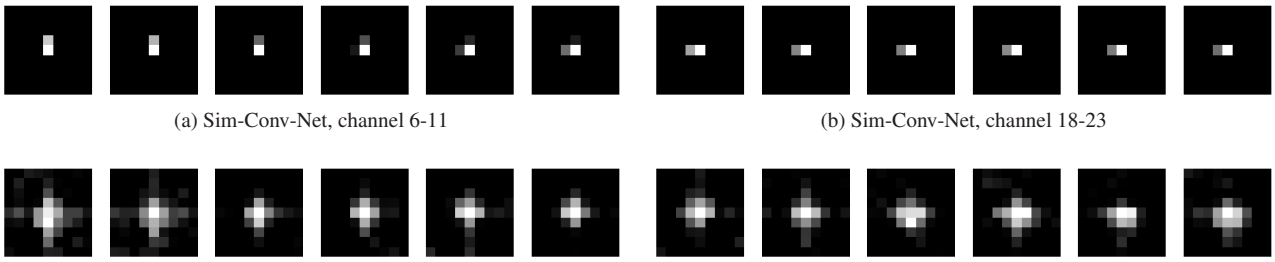
5.3. Breakdown Analysis

In this section, we analyze the characteristics of each submodel of UEM in combination with the experimental results and visualization.

Amplitude Encoding Model. We show the masks in Figure 7b. The AEM-I mask has a narrower value range than the binary AEM-P mask, with clear distinctions between channels, illustrating similar masks amongst neighboring channels. Supplementary material provides a complete view of AEM-I masks.

Retaining more scene information in optical encoding seems advantageous for obtaining superior spectral reconstruction and enhancing robustness in the context of AEMs. Our experimental findings strongly indicate that optimizing AEM masks via a floating-point representation rather than a binary one contributes to boosting the performance of spectral imaging.

Phase Encoding Model. In PEM-I, achieving convergent optimization results becomes challenging when the PSF has the same size as the image. Consequently, we set the PSF size to 9, optimizing only the central 9×9 region and padding the remaining area with zeros. The visualization



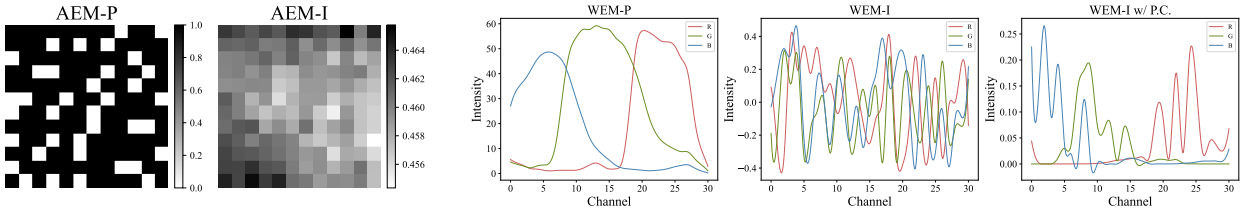
(a) Sim-Conv-Net, channel 6-11

(b) Sim-Conv-Net, channel 18-23

(c) Res-U-Net (depth=7), channel 6-11

(d) Res-U-Net (depth=7), channel 18-23

Figure 6. Visual comparison of the PSFs of PEM-I using different decoding networks. The PSF energy of the Sim-Conv-Net is more convergent at the center, and the PSF pattern is differentiated between channels. The PSFs of the Res-U-Net are more diffuse and have no obvious regularity.



(a) Mask

(b) Filter Response Curve

Figure 7. Visual comparison of masks of AEMs and response curves of WEMs. The mask is zoomed in for better visualization.

of PEM-I PSFs, utilizing two different decoding models, is depicted in Figure 6. Notably, the PSFs of PEM-I with Res-U-Net exhibit significant divergence and inferior imaging performance. Conversely, the Sim-Conv-Net experiment, which attains the best performance, showcases relatively concentrated patterns at the center with slight variations across channels.

Our experiments involve varying PSF sizes and convolutional kernel sizes of the decoding model, revealing that smaller learnable PSF and kernel sizes improve PSF convergence and overall performance. Additional details can be found in the supplementary materials. To achieve clear imaging quality and preserve some spectral information, the design of a concentrated PSF with a slight variance across the channels is necessary for the PEM.

Wavelength Encoding Model.

In WEM-I, a linear layer is employed for encoding instead of a fixed RGB filter response function. We visualize the weights of the linear layer as spectral response curves for comparison with the FLIR BFS-U3-04S2C-C. As shown in Figure 7b, the response curve of the WEM-I displays a densely serrated pattern while the WEM-I with positive constraints (WEM-I w/ P.C.) is smoother but still has more peaks than the ordinary camera response curve. Comparing the performance of the three WEMs in Table 4, the WEM-I still has the best imaging performance, while the WEM-I w/ P.C. and the WEM-P exhibit comparable imaging performance.

The filter and sensor response is crucial in controlling the information bottleneck between the encoder and decoder. When the response is optimized according to the task, the response curve will usually have multiple peaks. According to our results, the ideal shape of the filter response curve tends to have various separated sharp peaks compared to the ordinary response curves of cameras. Therefore, we recommend maximizing peak separation in the design process to improve spectral imaging performance.

6. Conclusion

In this study, we propose the UEM, which allows for a fair comparison of amplitude, phase, and wavelength encoding computational spectral imaging systems through joint encoder-decoder optimization. We utilize the UEM to establish physical models of these systems, enabling a comprehensive performance analysis. Additionally, we extend the UEM to ideal models and explore the potential upper bounds of performance without physical constraints. Finally, by visualizing and analyzing the ideal UEMs, we provide suggestions and insights for the future design of spectral imaging systems.

References

[1] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *European Conference on Computer Vision (ECCV)*, pages 19–34. Springer, 2016. 5

[2] Gonzalo R Arce, David J Brady, Lawrence Carin, Henry Arguello, and David S Kittle. Compressive coded aperture spectral imaging: An introduction. *IEEE Signal Processing Magazine (SPM)*, 31(1):105–115, 2013. 2

[3] Henry Arguello and Gonzalo R Arce. Colored coded aperture design by concentration of measure in compressive spectral imaging. *IEEE Transactions on Image Processing (TIP)*, 23(4):1896–1908, 2014. 3

[4] 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. 2

[5] Yitzhak August, Chaim Vachman, Yair Rivenson, and Adrian Stern. Compressive hyperspectral imaging by random separable projections in both the spatial and the spectral domains. *Applied Optics (AO)*, 52(10):D46–D54, 2013. 1, 4

[6] 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 *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2651–2660, 2021. 3

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

[8] Yuanhao Cai, Jing Lin, Haoqian Wang, Xin Yuan, Henghui Ding, Yulun Zhang, Radu Timofte, and Luc V Gool. Degradation-aware unfolding half-shuffle transformer for spectral compressive imaging. *Advances in Neural Information Processing Systems (NIPS)*, 35:37749–37761, 2022. 5

[9] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine (SPM)*, 25(2):21–30, 2008. 2

[10] Xun Cao, Tao Yue, Xing Lin, Stephen Lin, Xin Yuan, Qionghai Dai, Lawrence Carin, and David J Brady. Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world. *IEEE Signal Processing Magazine (SPM)*, 33(5):95–108, 2016. 2

[11] A. Chakrabarti and T. Zickler. Statistics of Real-World Hyperspectral Images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 193–200, 2011. 5

[12] Inchang Choi, Daniel S. Jeon, Giljoo Nam, Diego Gutierrez, and Min H. Kim. High-quality hyperspectral reconstruction using a spectral prior. *ACM Transactions on Graphics (TOG)*, 36(6):218:1–13, 2017. 5

[13] Claudia V Correa, Henry Arguello, and Gonzalo R Arce. Compressive spectral imaging with colored-patterned detectors. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7789–7793. IEEE, 2014. 3

[14] Claudia V Correa, Henry Arguello, and Gonzalo R Arce. Snapshot colored compressive spectral imager. *JOSA A*, 32 (10):1754–1763, 2015. 4

[15] Matthieu Courbariaux, Yoshua Bengio, and Jean-Pierre David. Binaryconnect: Training deep neural networks with binary weights during propagations. *Advances in Neural Information Processing Systems (NIPS)*, 28, 2015. 2

[16] Weisheng Dong, Peiyao Wang, Wotao Yin, Guangming Shi, Fangfang Wu, and Xiaotong Lu. Denoising prior driven deep neural network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 41(10): 2305–2318, 2018. 5

[17] David L Donoho. Compressed sensing. *IEEE Transactions on information theory (TIT)*, 52(4):1289–1306, 2006. 2

[18] Marco F Duarte and Richard G Baraniuk. Spectral compressive sensing. *Applied and Computational Harmonic Analysis (ACHA)*, 35(1):111–129, 2013. 2

[19] Marco F Duarte, Mark A Davenport, Dharmpal Takhar, Jason N Laska, Ting Sun, Kevin F Kelly, and Richard G Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine (SPM)*, 25(2):83–91, 2008. 2

[20] Xiong Dun, Hayato Ikoma, Gordon Wetzstein, Zhanshan Wang, Xinbin Cheng, and Yifan Peng. Learned rotationally symmetric diffractive achromat for full-spectrum computational imaging. *Optica*, 7(8):913–922, 2020. 2, 3

[21] Ying Fu, Tao Zhang, Yinqiang Zheng, Debing Zhang, and Hua. Huang. Joint camera spectral response selection and hyperspectral image recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, PP(99):1–1, 2020. 3, 4

[22] Ying Fu, Tao Zhang, Yinqiang Zheng, Debing Zhang, and Hua Huang. Joint camera spectral response selection and hyperspectral image recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 44(1):256–272, 2020. 2

[23] Silvano Galliani, Charis Lanaras, Dimitrios Marmanis, Emmanuel Baltsavias, and Konrad Schindler. Learned spectral super-resolution. *arXiv preprint arXiv:1703.09470*, 2017. 2

[24] Michael E Gehm, Renu John, David J Brady, Rebecca M Willett, and Timothy J Schulz. Single-shot compressive spectral imaging with a dual-disperser architecture. *Optics Express (OE)*, 15(21):14013–14027, 2007. 1

[25] Andrew S Glassner. How to derive a spectrum from an rgb triplet. *IEEE Computer Graphics and Applications (CGA)*, 9(4):95–99, 1989. 2

[26] Michael A Golub, Amir Averbuch, Menachem Nathan, Valery A Zheludev, Jonathan Hauser, Shay Gurevitch, Roman Malinsky, and Asaf Kagan. Compressed sensing snapshot spectral imaging by a regular digital camera with an added optical diffuser. *Applied Optics (AO)*, 55(3):432–443, 2016. 1

[27] Javier Gurrola-Ramos, Oscar Dalmau, and Teresa E Alarcón. A residual dense u-net neural network for image denoising. *IEEE Access*, 9:31742–31754, 2021. 5

[28] Jonathan Hauser, Amit Zeligman, Amir Averbuch, Valery A Zheludev, and Menachem Nathan. Dd-net: spectral imaging from a monochromatic dispersed and diffused snapshot. *Applied Optics (AO)*, 59(36):11196–11208, 2020. 1, 2

[29] Hayato Ikoma, Cindy M Nguyen, Christopher A Metzler, Yifan Peng, and Gordon Wetzstein. Depth from defocus with learned optics for imaging and occlusion-aware depth estimation. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2021. 3

[30] Roman Jacome, Jorge Bacca, and Henry Arguello. D 2 uf: Deep coded aperture design and unrolling algorithm for compressive spectral image fusion. *IEEE Journal of Selected Topics in Signal Processing (J-STSP)*, 2022. 2

[31] 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 Transactions on Graphics (TOG)*, 38(4):1–13, 2019. 1

[32] 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 Transactions on Graphics (TOG)*, 2019. 2

[33] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. 5

[34] Hyungtae Lee and Heesung Kwon. Going deeper with contextual cnn for hyperspectral image classification. *IEEE Transactions on Image Processing (TIP)*, 26(10):4843–4855, 2017. 5

[35] Lingen Li, Lizhi Wang, Weitao Song, Lei Zhang, Zhiwei Xiong, and Hua Huang. Quantization-aware deep optics for diffractive snapshot hyperspectral imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19780–19789, 2022. 2, 3

[36] Qingli Li, Xiaofu He, Yiting Wang, Hongying Liu, Dongrong Xu, and Fangmin Guo. Review of spectral imaging technology in biomedical engineering: achievements and challenges. *Journal of Biomedical Optics*, 18(10):100901–100901, 2013. 1

[37] Xing Lin, Yebin Liu, Jiamin Wu, and Qionghai Dai. Spatial-spectral encoded compressive hyperspectral imaging. *ACM Transactions on Graphics (TOG)*, 33(6):1–11, 2014. 4

[38] Xing Lin, Gordon Wetzstein, Yebin Liu, and Qionghai Dai. Dual-coded compressive hyperspectral imaging. *Optics Letters*, 39(7):2044–2047, 2014. 3

[39] Xin Liu, Linpei Li, Xu Liu, Xiang Hao, and Yifan Peng. Investigating deep optics model representation in affecting resolved all-in-focus image quality and depth estimation fidelity. *Optics Express (OE)*, 30(20):36973–36984, 2022. 3

[40] Patrick Llull, Xuejun Liao, Xin Yuan, Jianbo Yang, David Kittle, Lawrence Carin, Guillermo Sapiro, and David J Brady. Coded aperture compressive temporal imaging. *Optics Express (OE)*, 21(9):10526–10545, 2013. 1, 4

[41] Alessandro Maffei, Juan M Haut, Mercedes Eugenia Paoletti, Javier Plaza, Lorenzo Bruzzone, and Antonio Plaza. A single model cnn for hyperspectral image denoising. *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, 58 (4):2516–2529, 2019. 5

[42] Maksim Makarenko, Arturo Burguete-Lopez, Qizhou Wang, Fedor Getman, Silvio Giancola, Bernard Ghanem, and Andrea Fratalocchi. Real-time hyperspectral imaging in hardware via trained metasurface encoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12692–12702, 2022. 2

[43] Adriano Mancini, Emanuele Frontoni, and Primo Zingaretti. Challenges of multi/hyper spectral images in precision agriculture applications. In *IOP Conference Series: Earth and Environmental Science (EES)*, page 012001, 2019. 1

[44] Gabriel Martín, José M Bioucas-Dias, and Antonio Plaza. Hyca: A new technique for hyperspectral compressive sensing. *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, 53(5):2819–2831, 2014. 2

[45] Christopher A Metzler, Hayato Ikoma, Yifan Peng, and Gordon Wetzstein. Deep optics for single-shot high-dynamic-range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1375–1385, 2020. 3

[46] Xin Miao, Xin Yuan, Yunchen Pu, and Vassilis Athitsos. l-net: Reconstruct hyperspectral images from a snapshot measurement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4059–4069, 2019. 5

[47] Dragomir N Neshev and Andrey E Miroshnichenko. Enabling smart vision with metasurfaces. *Nature Photonics*, 17(1):26–35, 2023. 2

[48] Rang MH Nguyen, Dilip K Prasad, and Michael S Brown. Training-based spectral reconstruction from a single rgb image. In *European Conference on Computer Vision (ECCV)*, pages 186–201. Springer, 2014. 1

[49] Shijie Nie, Lin Gu, Yinqiang Zheng, Antony Lam, Nobutaka Ono, and Imari Sato. Deeply learned filter response functions for hyperspectral reconstruction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4767–4776, 2018. 2, 3

[50] Yaniv Oiknine, Isaac August, Vladimir Farber, Daniel Gedalin, and Adrian Stern. Compressive sensing hyperspectral imaging by spectral multiplexing with liquid crystal. *Journal of imaging*, 5(1):3, 2018. 2

[51] Yifan Peng, Qiang Fu, Felix Heide, and Wolfgang Heidrich. The diffractive achromat full spectrum computational imaging with diffractive optics. In *SIGGRAPH ASIA 2016 Virtual Reality meets Physical Reality: Modelling and Simulating Virtual Humans and Environments*, pages 1–2. 2016. 2

[52] Antonio Robles-Kelly. Single image spectral reconstruction for multimedia applications. In *Proceedings of the ACM International Conference on Multimedia (ICM)*, pages 251–260, 2015. 1

[53] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. 5

[54] Hoover Rueda, Henry Arguello, and Gonzalo R Arce. Compressive spectral imaging based on colored coded apertures.

In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7799–7803. IEEE, 2014. 4

[55] Edgar Salazar and Gonzalo R Arce. Coded aperture optimization in spatial spectral compressive spectral imagers. *IEEE Transactions on Computational Imaging (TCI)*, 6:764–777, 2020. 2, 3

[56] Vincent Sitzmann, Steven Diamond, Yifan Peng, Xiong Dun, Stephen Boyd, Wolfgang Heidrich, Felix Heide, and Gordon Wetzstein. End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Transactions on Graphics (TOG)*, 37(4):1–13, 2018. 2

[57] Brian Smits. An rgb-to-spectrum conversion for reflectances. *Journal of Graphics tools*, 4(4):11–22, 1999. 2

[58] Lu Sun, Weisheng Dong, Xin Li, Jinjian Wu, Leida Li, and Guangming Shi. Deep maximum a posterior estimator for video denoising. *International Journal of Computer Vision (IJCV)*, 129:2827–2845, 2021. 5

[59] Qilin Sun, Ethan Tseng, Qiang Fu, Wolfgang Heidrich, and Felix Heide. Learning rank-1 diffractive optics for single-shot high dynamic range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1386–1396, 2020. 3

[60] Yinlong Sun, F David Fracchia, Thomas W Calvert, and Mark S Drew. Deriving spectra from colors and rendering light interference. *IEEE Computer Graphics and Applications (CGA)*, 19(4):61–67, 1999. 2

[61] Yubao Sun, Ying Yang, Qingshan Liu, and Mohan Kankanhalli. Unsupervised spatial-spectral network learning for hyperspectral compressive snapshot reconstruction. *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, 60:1–14, 2021. 4

[62] Mohammad Reza Taghizadeh, P Blair, B Layet, IM Barton, AJ Waddie, and N Ross. Design and fabrication of diffractive optical elements. *Microelectronic Engineering*, 34(3-4): 219–242, 1997. 2

[63] Ethan Tseng, Shane Colburn, James Whitehead, Luocheng Huang, Seung-Hwan Baek, Arka Majumdar, and Felix Heide. Neural nano-optics for high-quality thin lens imaging. *Nature Communications*, 12(1):6493, 2021. 2

[64] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems (NIPS)*, 30, 2017. 2

[65] Ashwin Wagadarikar, Renu John, Rebecca Willett, and David Brady. Single disperser design for coded aperture snapshot spectral imaging. *Applied Optics (AO)*, 47(10): B44–B51, 2008. 3, 4

[66] Michael B Wakin, Jason N Laska, Marco F Duarte, Dror Baron, Shriram Sarvotham, Dharmpal Takhar, Kevin F Kelly, and Richard G Baraniuk. An architecture for compressive imaging. In *IEEE International Conference on Image Processing (ICIP)*, pages 1273–1276. IEEE, 2006. 2

[67] Lizhi Wang, Tao Zhang, Ying Fu, and Hua Huang. Hyperreconnet: Joint coded aperture optimization and image reconstruction for compressive hyperspectral imaging. *IEEE Transactions on Image Processing (TIP)*, 28(5):2257–2270, 2018. 2

[68] Zhiwei Xiong, Zhan Shi, Huiqun Li, Lizhi Wang, Dong Liu, and Feng Wu. Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections. In *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 518–525, 2017. 2

[69] F. Yasuma, T. Mitsunaga, D. Iso, and S.K. Nayar. Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum. Technical report, 2008. 5

[70] Qi Zeng, Long Wang, Sijun Wu, Guangpu Fang, Mingwei Zhao, Zheng Li, and Wenlong Li. Research progress on the application of spectral imaging technology in pharmaceutical tablet analysis. *International Journal of Pharmaceutics*, page 122100, 2022. 1

[71] Kai Zhang, Luc Van Gool, and Radu Timofte. Deep unfolding network for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3217–3226, 2020. 5

[72] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-play image restoration with deep denoiser prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 44(10):6360–6376, 2021.

[73] Xuanyu Zhang, Yongbing Zhang, Ruiqin Xiong, Qilin Sun, and Jian Zhang. Herosnet: Hyperspectral explicable reconstruction and optimal sampling deep network for snapshot compressive imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17532–17541, 2022. 5

[74] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters(GRSL)*, 15(5):749–753, 2018. 5

[75] Zhonghao Zhang, Yipeng Liu, Jiani Liu, Fei Wen, and Ce Zhu. Amp-net: Denoising-based deep unfolding for compressive image sensing. *IEEE Transactions on Image Processing (TIP)*, 30:1487–1500, 2020. 5