PRELIMINARY PREPRINT VERSION – DO NOT CITE
This is not the version of record. The final version will appear in the Proceedings of IJCAI-ECAI 2026.

Diffract and Conquer: Hyperspectral Imaging from Any RGB Camera via Optical Encoding and Learning

Alexey Pronin[1], **Daniil Vladimirov**[1,2,3], **Andrei Korepanov**[1,4], **Artem Muzyka**[1], **Andrey Makarov**[1], **Sofya Podtikhova**[1], **Egor Ershov**[2,3], **Andrey Rastorguev**[1], **Roman Skidanov**[1] and **Artem Nikonorov**[1]
[1]Samara National Research University, Samara, Russia
[2]MIRAI, Moscow, Russia
[3]AXXX, Moscow, Russia
[4]iMotions, Copenhagen, Denmark
artniko@gmail.com

Abstract

Hyperspectral imaging (HSI) provides detailed spectral information but is often impractical due to high cost, size, and hardware complexity. While learning-based methods attempt to recover hyperspectral images from RGB sensors, they are constrained by limited spectral measurements and noise. We propose a unified optical–computational approach that converts any standard RGB camera into a snapshot hyperspectral imaging system. Our method introduces a diffractive optical adapter that replaces the conventional lens with a diffractive lens array optimized for spectral encoding. To reconstruct hyperspectral images from the resulting measurements, we design a neural network specialized for diffractively encoded RGB data, capable of compensating for optical distortions and recovering high-quality spectra. The proposed system achieves up to 10 dB improvement in PSNR over RGB-based hyperspectral reconstruction and enhances the performance of existing state-of-the-art models by up to 6 dB when used with the proposed adapter. Our results demonstrate that diffractive optical encoding combined with learned reconstruction enables practical and scalable hyperspectral imaging using commodity RGB cameras.

1 Introduction

Computer vision technologies are widespread in various monitoring and inspection tasks, particularly in application domains such as precise agriculture [Arad *et al.*, 2020; Kakani *et al.*, 2020] or automated waste sorting [Kiyokawa *et al.*, 2022]. In practice, these systems rely on RGB or RGB-D technical vision cameras due to their affordability and ease of integration, while hyperspectral imaging technologies have gained attention only in recent years [Kar and Dhal, 2025; Benelli *et al.*, 2020; Tamin *et al.*, 2023]. By capturing dense spectral features at each spatial pixel of an image, hyperspectral imaging allows precise material properties and color analysis. Hyperspectral data are conventionally captured using

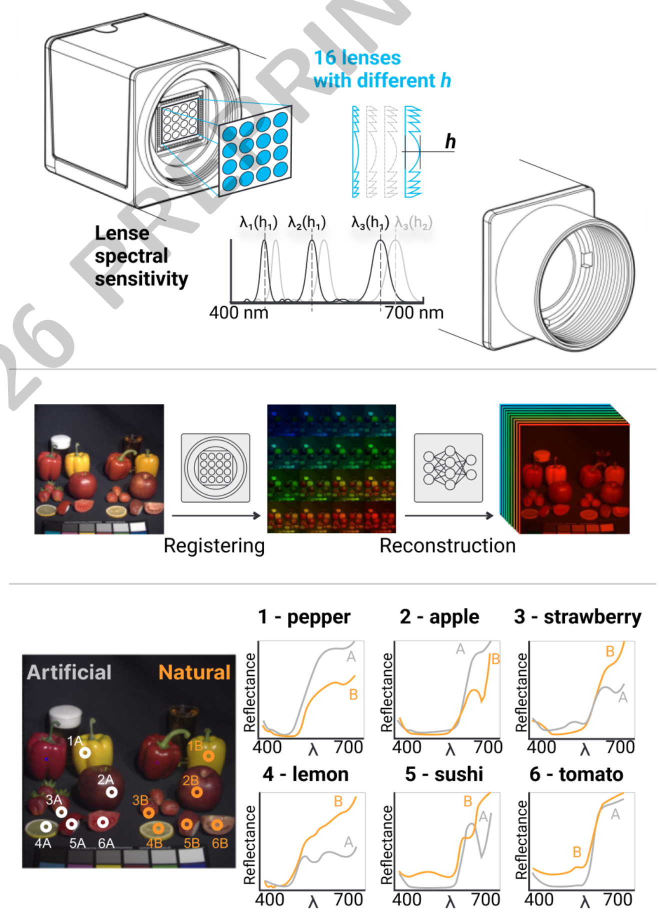

Figure 1: Overview of the proposed snapshot imaging system. **The 1st row**: Array of 16 harmonic diffractive lenses allows to encode the spectral information into 2D sensor. **The 2nd row**: The proposed reconstruction framework estimates the HSI from encoded information. **The 3rd row**: Raw signal captured with proposed optical system and reconstructed spectra allows to determine the natural and artificial objects.

industrial scanning hyperspectral cameras based on whiskbroom or pushbroom imaging techniques. Such systems are often expensive and require a time-consuming acquisition procedure, which limits their applications in dynamic or real-time scenarios [Mukhtar *et al.*, 2025].

Snapshot hyperspectral imaging is an alternative that aims to capture the hyperspectral data in a single exposure, significantly reducing acquisition time and motion artifacts. A common strategy for snapshot systems relies on spatial-spectral encoding via optical elements as prisms [Arce *et al.*, 2013] or diffractive optical elements (DOEs) [Shi *et al.*, 2024]. Such optical encoding produces hyperspectrally mixed measurements, making software reconstruction a necessary component of the HSI registration pipeline. Recovering dense spectral information from these measurements is an ill-posed inverse problem [Mukhtar *et al.*, 2025]. A simple example of this paradigm is HSI reconstruction from RGB images [Arad and Timofte, 2022] or Coded-aperture snapshot spectral system (CASSI) images [Meng *et al.*, 2020], where a neural network reconstructs the high-dimensional spectra from compressed observations.

The spectral resolution of RGB sensor equipped with Bayer color filter array (CFA) can be enhanced with DOE-based optical elements. Initially, the a Bayer CFA usage constrains optical efficiency [Shi *et al.*, 2024] and spectral fidelity. However, these limitations can be mitigated by moving the spectral sensitivity enhancement to the DOEs configuration process. Extending this idea, we propose a snapshot hyperspectral imaging system based on a 4×4 matrix of harmonic diffractive lenses (HDLs).

In contrast to single-wavelength DOEs, our HDL focuses three wavelengths, featuring spectral profiles tailored to Bayer filter transmittance. To reconstruct the high-fidelity spectral signals from the captured snapshots, we introduce a compact network inspired by Kolmogorov-Arnold Networks, enabling efficient and accurate spectral reconstruction. The overall system is illustrated in the Figure 1. In summary, the main contributions of our work are as follows:

- We introduce a snapshot hyperspectral imaging system recovering 31 spectral bands (400–700 nm) via a 4×4 HDLs matrix. The architecture exploits multi-order diffraction of HDLs to acquire 48 raw channels on a standard Bayer RGB sensor, followed by a neural network that mitigates optical aberrations for high-fidelity spectral reconstruction.
- We propose a learned reconstruction approach inspired by Kolmogorov-Arnold Networks (KANs). Distinct from mentioned, our approach incorporates a learnable basis of Gaussian primitives. The parameters of these primitives are generated by a hypernetwork for each HSI pixel. Extensive experiments demonstrate that our design outperforms state-of-the-art methods by up to 10 dB in PSNR, while maintaining a compact size of 70K parameters.
- We develop and validate a fully functional prototype of the proposed snapshot hyperspectral camera. The prototype is experimentally evaluated under real-world imaging conditions and benchmarked against Specim IQ, an industrial-grade scanning hyperspectral camera. The proposed system demonstrates comparable spectral reconstruction accuracy and spatial registration quality.

2 Related Works

Optical Solutions for Snapshot Hyperspectral Imaging. Snapshot HSI, essential for dynamic scene analysis, is generally categorized into two paradigms following the taxonomy in [Mukhtar *et al.*, 2025]. The first paradigm focuses on encoding spatial-spectral information via compressive sensing. While originating from the CASSI concept, this field remains active [Zhao *et al.*, 2023; Ma *et al.*, 2024], with recent architectures integrating periodic [Arguello *et al.*, 2021] or random Fabry-Perot masks [Yako *et al.*, 2023]. Alternative encoding mechanisms exploit dispersion via prisms [Baek *et al.*, 2017], engineer spectrally varying PSFs using diffractive elements [Li *et al.*, 2022; Xu *et al.*, 2023], or employ multi-channel diffractive designs [Shi *et al.*, 2024]. Although these approaches facilitate compact optics, they require solving ill-posed inverse problems heavily reliant on scene priors such as spectral smoothness [Baek *et al.*, 2017; Shi *et al.*, 2024; Mukhtar *et al.*, 2025]. Conversely, the second paradigm utilizes spatial multiplexing to map the 3D datacube directly onto a 2D sensor. Prominent designs employ microlens arrays (MLA) with discrete filters or interferometers [Mu *et al.*, 2019; Hubold *et al.*, 2021; Picone *et al.*, 2023; Wang *et al.*, 2024], or utilize fiber-optic remapping [Lu *et al.*, 2023]. While the pixel-to-voxel isomorphism of this approach eliminates the computational overhead of iterative reconstruction [Mukhtar *et al.*, 2025], existing counterparts are hindered by significant hardware complexity. Specifically, [Mu *et al.*, 2019] is rendered economically impractical by stringent filter array requirements, whereas [Hubold *et al.*, 2021], utilizing a linear filter and tilted optics, is constrained by manufacturing intricacies and spectral broadening artifacts.

We propose a novel architecture synergizing a harmonic lens system with standard Bayer color filters and a neural reconstruction backbone. By integrating the CMV4000 sensor, our system achieves a spatial resolution of 450 × 450 pixels, effectively circumventing the resolution bottlenecks of existing spatial multiplexing hyperspectrometers and broadening their application scope.

Learned Hyperspectral Reconstruction techniques. Modern hyperspectral reconstruction (HSR) from RGB is software-only alternative to costly hyperspectral-imaging hardware, enabling spectral recovery using standard RGB sensors [Arad and Timofte, 2022]. The problem is inherently ill-posed due to spectral information loss during RGB formation. The data-driven methods have been widely adopted to learn the mapping from RGB to spectral space. Modern state-of-the-art (SoTA) methods employ transformer-based deep neural networks to achieve high-quality reconstruction [Cai *et al.*, 2022c; Wu *et al.*, 2024; Yang *et al.*, 2024; Yang *et al.*, 2026].

Beyond RGB inputs, similar learning-based reconstruction approaches have been developed for CASSI measurements [Meng *et al.*, 2020; Cai *et al.*, 2022a; Zhao *et al.*, 2023;

Ma *et al.*, 2024; Meng *et al.*, 2025]. Although these systems provide richer measurements than RGB images, the captured signals remain strongly compressed, and their reconstruction pipelines commonly adopt transformer-based network architectures [Yao *et al.*, 2024; Cai *et al.*, 2022b; Li *et al.*, 2024]. In contrast, snapshot hyperspectral systems based on DOE arrays capture more discriminative spectral information. These optical advancements enable SoTA results using simpler dense CNN-based architectures, eliminating the need for computationally demanding self-attention mechanisms while achieving superior reconstruction accuracy [Li *et al.*, 2022; Arguello *et al.*, 2021; Shi *et al.*, 2024].

3 Snapshot Hyperspectral Camera with Harmonic Diffractive Optics and Gaussian Primitives

Figure 2 illustrates the optical architecture of the imaging spectrometer utilizing a harmonic lens matrix. The corresponding hardware implementation, featuring a 4×4 lens array, is depicted in Figure 3.

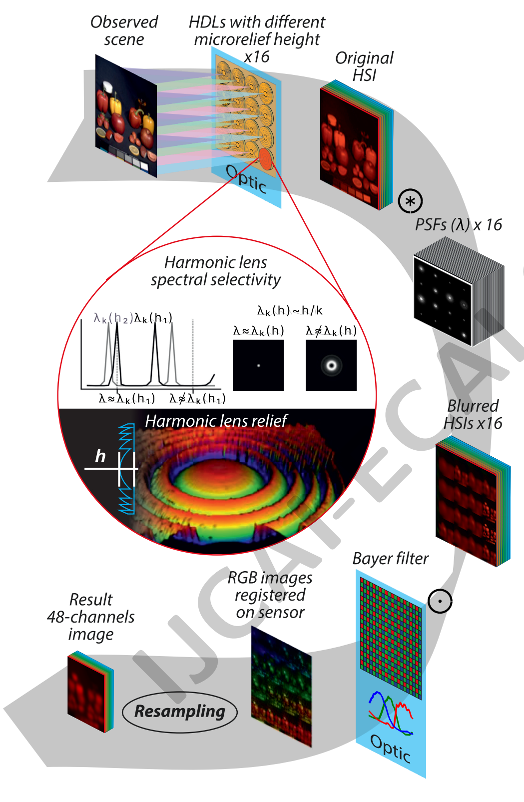

Figure 2: Optical scheme configuration. Each HDL has a distinct microrelief, resulting in a unique spectral PSF. The spectral selectivity of each lens is carefully optimized to match Bayer CFA transmittance. The system configuration enables the acquisition of 48 spectral channels.

A harmonic lens differs from a conventional diffraction lens by a significantly higher microrelief h [Sweeney and

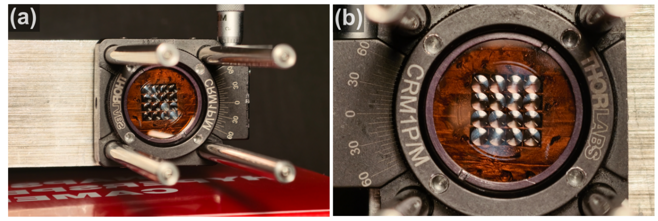

Figure 3: Prototype of the snapshot hyperspectral camera. (a) General view. (b) Close-up of the 4×4 HDL matrix.

Sommargren, 1995]:

$$m = \frac{h\,(n-1)}{\lambda_0}, \tag{1}$$

where m is an order of harmonic lens, showing how many times the microrelief height of a harmonic lens is greater than the microrelief height of a diffraction lens at wavelength λ_0. In this case, several wavelengths arise for a harmonic lens λ_κ [Sweeney and Sommargren, 1995]:

$$\lambda_k = \frac{m\lambda_0}{k} = \frac{h\,(n-1)}{k}, \tag{2}$$

where k is an integer.

We demonstrate an imaging spectrometer where specific microrelief heights align three diffraction harmonics with standard RGB transmission windows.

Consequently, a 4×4 harmonic lens array yields 48 independent image channels. The optical performance of this HDL system is evaluated via sequential simulation of individual components. We formulate the image formation process for a specific HDL indexed by d ($d = 1, ..., 16$). The optical system focuses the incident wavefront from the scene onto a local region of the sensor plane equipped with a Bayer filter array. Integrating the spectral sensitivity of the photodetector $T_c\,(\lambda)$ (where $c = 1, 2, 3$ corresponds to the R, G, B channels), the captured image $I_{d,c}\,(x, y)$ is expressed as follows:

$$\begin{aligned} I_{d,c}\,(x,y) = &\int\limits_{\Delta\lambda_c} \left[PSF_d\,(x,y,\lambda) * I\,(x,y,\lambda)\right] \times \\ &\times T_c\,(\lambda)\,d\lambda, \end{aligned} \tag{3}$$

where $PSF_d\,(x, y, \lambda)$ denotes the spatially and spectrally variant point spread function corresponding to the d-th HDL, and $I\,(x, y, \lambda)$ represents the spectral radiance of the scene. The integration domain $\Delta\lambda_c$ is determined by the spectral bandwidth of the respective Bayer color filters c.

To compute the point spread function PSF_d of the d-th HDL, we consider an incident spherical wave $U_{in}\,(x_1, y_1, \lambda)$ originating from a point source located at a distance $z = 2f$, where f denotes the lens focal length. Consequently, by incorporating the complex transmission function, the output complex amplitude is obtained as follows [Doskolovich *et al.*, 2020]:

$$\begin{aligned} &U_{out}\,(x_1,y_1,\lambda) = U_{in}\,(x_1,y_1,\lambda) \times \\ &\times exp\left(i\frac{2\pi}{\lambda}\,(n-1)\,h_d\,(x_1,y_1)\right), \end{aligned} \tag{4}$$

where n is the wavelength-dependent refractive index, and $h_d\left(x_2, y_2\right)$ defines the geometry of the microrelief.

Leveraging (4), the PSF at distance z is defined as:

$$PSF_d\left(x_2, y_2, \lambda\right) = \left| \iint U_{out}\left(x_1, y_1, \lambda\right) \times \right. \\ \left. \times \exp\left(\frac{i\pi}{\lambda z}\left[\left(x_2 - x_1\right)^2 + \left(y_2 - y_1\right)^2\right]\right) dx_1 dy_1 \right|^2 . \tag{5}$$

Considering the stochastic nature of the signal due to photon and electronic noise, and based on [Shi *et al.*, 2024], the sensor readout is modeled as:

$$N_{d,c} = N_p\left(I_{d,c}, \sigma_p\right) + N_g\left(I_{d,c}, \sigma_g\right), \tag{6}$$

where N_p constitutes the Poisson noise component, and N_g represents the Gaussian noise term. Eqs. (3) – (6) describe the formation of the $d \times c$ spectral bands $I_{d,c}$.

3.1 Harmonic Lens Wavelength Optimization

In our 4×4 snapshot configuration, the aggregate spectral sampling is determined by the union of harmonics produced by all HDL elements. We address the visible spectral range of 400–700 nm, discretized into $|\Lambda|$ = 31 bands with a uniform step of 10 nm.

A target wavelength $\lambda_t \in \Lambda$ is considered covered by an HDL with height h and maximum order m, as defined in Eq. (1), if there exists an integer $k \leq m$ such that $|\lambda_t - \lambda_k(h)| < \delta\lambda$, where $\lambda_k(h) = (n-1)h/k$.

Candidate microrelief heights are generated via an inverse design approach: λ_t and admissible k, we compute

$$h = \frac{\lambda_t k}{n-1}, \tag{7}$$

then restrict h to $[h_{\min}, h_{\max}]$ and quantize by Δh.

We select at most $D = 16$ HDLs by solving a budgeted maximum-coverage problem:

$$\max_{|S|\leq D} \left| \bigcup_{(h,M)\in S} \mathrm{cov}(h, M) \right|, \\ \mathrm{cov}(h, M) = \{\, \lambda_t \in \Lambda \mid \exists\, k \leq M, \\ |\lambda_t - \lambda_k(h)| < \delta\lambda \,\}. \tag{8}$$

We use a greedy maximum-coverage procedure (ties: smaller M, then smaller h);

3.2 Generated Gaussian Primitives Image Reconstruction Framework

Spectral Projector and Aggregator Blocks

Unlike prior methods that reconstruct full hyperspectral images (HSI) from a small number of channels [Shi *et al.*, 2024] or from RGB inputs [Cai *et al.*, 2022c], our proposed optical design acquires the full spectral range directly, eliminating the need for spectral super-resolution. Consequently, the reconstruction objective is formulated purely as an image enhancement task. The cmKAN framework [Nikonorov *et al.*, 2025], which employs pixel-wise parameter generation for Kolmogorov–Arnold Networks (KAN), has demonstrated superior performance in RGB image enhancement. It leverages the spline capabilities of Kolmogorov-Arnold Networks (KANs) to model the color matching between source and target distributions. Specifically, it uses a hypernetwork that generates spatially varying weight maps to control the nonlinear splines of a KAN, enabling accurate color matching. The number of parameters generated per-pixel for M input channels is defined as:

$$n_{cmKAN} = M(G + k + 2), \tag{9}$$

where k, and G denote spline order and group size for the KAN spline basis, respectively. According to (9), processing RGB data requires 30 parameters per channel, totaling 90 parameters. However, extending (9) to HSI with 31 channels results in 310 per output channel, and 9,610 parameters in total, making cmKAN impractical for HSI applications.

To address this problem, we introduce a Spectral Projector block (Fig. 4), which functions as a sliding window along the spectral dimension to capture local dependencies within a three-channel receptive field. Furthermore, it incorporates a mechanism to compensate for the chromatic aberration inherent to HDLs. By modeling aberration as energy dispersion controlled by a PSF, we approximate the energy redistribution using a non-linear modulation term. Specifically, the feature refinement is defined as $x + x \cdot \mathrm{conv2D}(x)^2$, where the squared convolution term acts as an attention-like gate to correct PSF-induced artifacts.

The Spectral Aggregator block then projects the latent features back to the original HSI spectral space to restore energy balance. As shown in Fig. 4, a Feed-Forward Network (FFN) block [Cai *et al.*, 2022c] serves as a bottleneck, reducing dimensionality from N to 3 prior to rearrangement. By integrating these projection blocks into the cmKANlight backbone [Nikonorov *et al.*, 2025], we establish our baseline architecture, cmKAN++, specifically adapted for HSI reconstruction.

Gaussian Primitives Layer

Recent studies indicate that reducing the spline parameter count in KAN layers and employing Gaussian approximations for the spline bases [Li, 2024; Ta, 2024] improve HSI analysis performance [Firsov *et al.*, 2024]. Motivated by these results and appealing to the Kolmogorov-Arnold Theorem in the form of Sprecher's generalization [Braun and Griebel, 2009], we propose a Gaussian Primitives Layer, which models the output channels c_j, $0 \leq j < N$ as a linear combination of parameterized 1D Gaussian functions:

$$c_j = d_j \sum_{i}^{M} exp\left(-0.5\left(a_{ij} s_i + b_{ij}\right)^2\right), \tag{10}$$

where d_j, a_{ij}, b_{ij} are layer parameters and s_i, $0 \leq i < M$ are input channels. Unlike fastKAN, the Gaussian primitives in (10) are parameterized to adapt to the local data context. The layer processes M input channels generated by the Spectral Projector. The number of output channels N is a hyperparameter that can be tuned to optimize performance; in our work, we evaluated $N \in 3, 5, 7, 9, 11$. For each channel, the parameter count is given by:

$$n_{GGP} = N(2M + 1). \tag{11}$$

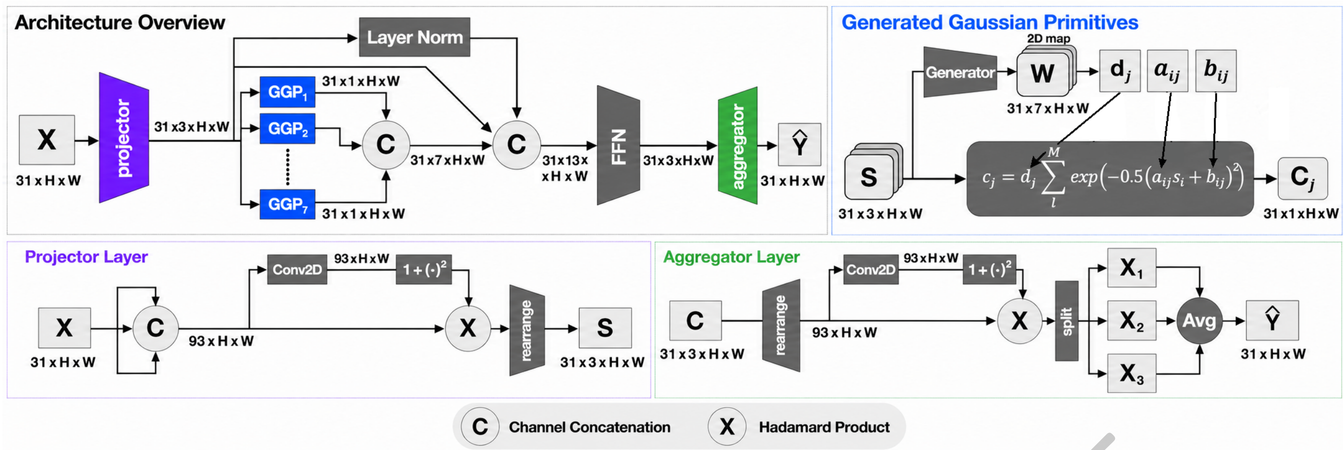

Figure 4: Generated Gaussian Primitives Image Restoration framework. Projector layer reduces the data dimensionality from 31 channels to 3, using the grouped convolution with a size 3×3 kernel. The aggregator layer employs grouped convolution with a 3×3 kernel to expand data dimensionality back to 31 channels. Generated Gaussian Primitives Layer obtains Gaussian parameters from Generator and applies the resulting Gaussian function to the input tensor, where Generator is Light cmKAN Encoder.

For our range of N, this configuration results in at most 63 parameters, which is significantly fewer than the 310 parameters required by (9). Consequently, replacing traditional regular B-spline bases with learnable, non-regular Gaussian-based basis allows the Gaussian Primitives Layer to adapt to local image structure using fewer basis functions, thereby achieving substantial parameter efficiency.

Generated Gaussian Primitives Framework

We implement the Generated Gaussian Primitives (GGP) layer using a hypernetwork to generate spatially varying parameter maps that control the Gaussian bases in (10). The generator hypernetwork incorporates Illumination Estimator and Color Transformer blocks adopted from cmKANlight and redesigned specifically for Gaussian basis. To reduce memory consumption, we employ independent generators for each output channel.

By integrating the GGP Layer with a Spectral Projector and a Spectral Aggregator — which maps the N Gaussian layer outputs back to the original M spectral channels — we arrive at the proposed Generated Gaussian Primitives Image Restoration (GGPIR) architecture (see Fig. 4). The model is then defined as:

$$\hat{Y} = S_{proj}(GPL(\mathcal{G}(S_{proj}(X), \theta), S_{aggr}(X))), \qquad (12)$$

where S_{proj} and S_{aggr} are the spectral projector and aggregator, GPL is the layer of Generated Gaussian Primitives, $\mathcal{G}$ is the generator block with parameters θ, and X, $\hat{Y}$ are the input and output hyperspectral images.

4 Computational Experiments

In this section, we provide a detailed analysis and evaluation of the proposed optical design and our reconstruction framework on four publicly available datasets.

4.1 Evaluation on Synthetic Data

To validate the proposed system, we utilized 4 diverse HSI datasets: NTIRE 2022 [Arad and Timofte, 2022] (950 images), CAVE [Yasuma *et al.*, 2010] (32 images), ICVL [Arad and Ben-Shahar, 2016] (400 images), and CZ_HSDB [Chakrabarti and Zickler, 2011] (50 images). As a recent study [Fu *et al.*, 2025] indicates that reconstruction methods are sensitive to scene content and HS device variations, therefore we evaluate each method within a single dataset without cross-dataset validation. We follow the official split for NTIRE and use a 4:1 training-to-testing ratio for the remaining datasets.

It is important to highlight that we compare various HSI estimation methods, including both software-only and optics-dependent approaches. This allows us to highlight the question "How precisely can we reconstruct spectral information using a similar sensor, but different (including apparatus) methods?". The first group consists of methods that reconstruct HSIs from RGB inputs, thus relying on a simple and accessible imaging system. For comparison, we selected SoTA neural network models from different years: MST++ [Cai *et al.*, 2022c], MSFN [Wu *et al.*, 2024], CESST [Yang *et al.*, 2024]. The methods in the second group jointly utilize optical hardware and computational reconstruction, typically based on DOE-arrays: SCCD [Arguello *et al.*, 2021], QDO [Li *et al.*, 2022], and the method proposed in [Shi *et al.*, 2024].

For the software-only approaches group, we use RGB images, provided by the dataset authors, as inputs. Since CZ_HSDB does not include corresponding RGB images, we rendered RGB responses from the HSIs to sRGB color space using XYZ spectral sensitivities, following a procedure similar to other datasets. To generate the input images for the second group, we utilize the optical system simulators released by the respective authors.

For the evaluation of the proposed approach, which utilizes the combination of hardware equipment and neural reconstruction network, we employed a synthetic data generation pipeline simulating our hyperspectral camera model accordingly (3)–(6). This simulation was used for each dataset, as detailed explicitly in the previous section. The lens configura-

Method type	Estimation method	NTIRE 2022	ICVL	CAVE	HARVARD
		SAM, ° ↓ / PSNR, dB ↑ / SSIM, ↑	SAM, ° ↓ / PSNR, dB ↑ / SSIM, ↑	SAM, ° ↓ / PSNR, dB ↑ / SSIM, ↑	SAM, ° ↓ / PSNR, dB ↑ / SSIM, ↑
Software only	MST++	5.88 / 28.37 / 0.89	2.38 / 35.12 / 0.98	7.58 / 35.14 / **0.98**	5.44 / 34.98 / 0.92
	CESST	5.76 / 28.71 / 0.91	2.43 / 34.68 / 0.98	12.08 / 27.62 / 0.95	12.19 / 30.34 / 0.90
	MSFN	4.94 / 25.99 / 0.85	2.42 / 35.16 / 0.98	7.48 / 34.97 / **0.98**	6.71 / 34.73 / 0.92
Hardware and Software	SCCD	**3.97** / 33.24 / **0.97**	2.29 / 31.92 / 0.97	12.19 / 25.83 / 0.78	8.94 / 33.19 / 0.82
	QDO	14.26 / 28.50 / 0.82	5.87 / 28.85 / 0.91	15.44 / 27.76 / 0.86	- / - / -*
	[Shi *et al.*, 2024]	11.11 / 32.72 / 0.90	3.86 / 36.40 / 0.98	11.04 / 32.65 / 0.94	6.76 / 32.99 / 0.87
	GGPIR (Proposed)	4.80 / **38.47** / 0.968	**1.24 / 46.37 / 0.996**	**6.78 / 35.58** / 0.937	**4.083 / 39.5 / 0.96**

Table 1: Comparison of different HSI reconstruction methods on 4 datasets. The **best** and second best results are highlighted. *The QDO results could not be reproduced on HARVARD.

tion was optimized as described in Section 3.2 to produce 31 spectral channels approximating a uniform spanning of 400—700 nm with a 10 nm stride. The simulation parameters were set as follows: 20 mm focal length, 40 mm object/image distance, and 2.8 mm aperture. Subsequently, layer-wise normalization was applied strictly to the training subset.

To measure different aspects of reconstructed images, we employ 3 evaluation metrics: Spectral Angle Mapper (SAM) [Kruse *et al.*, 1993] to estimate the accuracy of the reconstructed spectral shape, PSNR to measure the fidelity of reconstructed intensity values, and Structural Similarity Index Measure (SSIM) [Wang *et al.*, 2004] to evaluate the spatial consistency of the reconstructed HSI.

The quantitative comparisons are summarized in the Table 1. On large-scale datasets such as NTIRE 2022, ICVL, HARVARD, the GGPIR achieves the highest reconstruction fidelity in terms of PSNR, outperforming the second-best method by at least 5 dB. At the same time, our approach shows competitive results in terms of SAM and SSIM, indicating good spectral accuracy and structural consistency. These results demonstrate the advantage of the proposed approach over both other software-only and hardware-software co-design methods.

On smaller and diverse datasets, such as CAVE, the performance gap is less pronounced. This reduced margin may be attributed to the tendency of software-only methods, which contain at least 1.5 M parameters, to overfit when trained on limited data. In contrast, the proposed HDLs-based optical system allows the use of a compact reconstruction network with less than 100 K trainable parameters, while still maintaining competitive reconstruction accuracy.

4.2 Assessment of the Optical Scheme

We conducted a quantitative evaluation of the proposed optical design on the CAVE and NTIRE datasets (Table 2), comparing the proposed 4×4 system featuring an optimized height profile against a variant with random height distribution (in the 2.0–4.5 µm range) as well as the 5×5 and 6×6 configurations. The 3×3 configuration was omitted due to insufficient spectral sampling.

Experiments demonstrated that a random lens-height distribution consistently degrades the resulting quality. This is attributed to the spectral mismatch between lens harmonics and target wavelengths (400–700 nm with a 10 nm step). Furthermore, transitioning to a 5×5 or 6×6 array with a fixed sensor size reduces the aperture of individual lenses, which significantly degrades image contrast.

Height profile	CAVE 4×4	CAVE 5×5	CAVE 6×6	NTIRE 4×4	NTIRE 5×5	NTIRE 6×6
Optimized	**27.59**	23.09	21.01	**29.29**	25.06	22.70
Random	27.21	22.97	20.92	29.18	24.98	22.65

Table 2: Quantitative comparison (PSNR) of lens profiles on CAVE and NTIRE datasets (4×4, 5×5 and 6×6 configurations). Best results are **bolded**.

4.3 Evaluation of the Reconstruction Framework

To validate the generalization ability of the proposed methods, we utilize three datasets as sources of HSIs: NTIRE 2022, ICVL, CAVE (see Table 3). To simulate corresponding input images we used image formation model of the proposed imaging system (3)–(6). As a baseline, we used cmKAN++ [Nikonorov *et al.*, 2025] with the additional first layer — a projector reducing the data dimensionality from 31 channels to 3, along with the additional last layer - aggregator expanding the data dimensionality back to 31 channels, as described in Section 3.3. The proposed GGPIR architecture has been investigated in various configurations differing in the number of Gaussian layers: N=3 (S3), N=5 (S5), N=7 (S7) and N=9 (S9), and also a single configuration (M7) utilizing the 3D Gaussians instead of the summation of 1D Gaussians. Increasing the number of Gaussian layers, we get a higher capacity, however the number of network parameters increases proportionally: S3 contains 47K parameters, S7 and S11 contain 71K and 95K parameters respectively (i.e. 6K per each Gaussian layer), while the cmKAN++ contains 76K parameters. The optimal number of layers differs for each dataset: we get S7 as an optimal design for ICVL (leading to a 3.86 dB improvement in PSNR over the cmKAN++) and for CAVE (1.15dB improvement in PSNR), while the optimal design for NTIRE 2022 is S9 (2.79dB improvement in PSNR). Using 3D Gaussians (M7) does not bring any benefits; the PSNR and SSIM values are quite close to those obtained for S7. On the CAVE-HSI dataset the proposed architecture is inferior to the MST++ in terms of PSNR (35.58dB for S7 vs 36.36dB), while the S7 contains more than 22 times fewer trainable parameters (71K vs 1.6M). One extra architecture (SC) has been investigated: SC is based on S7 with the only block $1 + x^2$ replaced with the block $1 + x\sigma(x)$ in the projector and aggregator (Figure 4), where $\sigma(\cdot)$ is a sigmoid function.

Notably, our main competitor is the Transformer-based MST++ baseline. However, despite using orders of magnitude fewer parameters, GGPIR achieves better reconstruction quality: as shown in Table 3, GGPIR$_{S9}$ reaches 38.47 dB / 0.968 on NTIRE'22 with only 83K parameters, outperforming MST++ (37.88 dB / 0.96) that uses 1.6M parameters. This highlights the strong parameter-efficiency of the proposed GGPIR framework.

Method	# params	NTIRE'22 $M_1\uparrow$ / $M_2\uparrow$	ICVL $M_1\uparrow$ / $M_2\uparrow$	CAVE $M_1\uparrow$ / $M_2\downarrow$
MST++	1.6M	37.88/0.960	40.11/0.967	**36.36/0.940**
cmKAN++	76K	35.68/0.960	42.51/0.993	34.43/0.927
GGPIR$_{S3}$	47K	36.73/0.953	44.74/0.995	33.74/0.921
GGPIR$_{S5}$	59K	36.46/0.957	45.62/0.995	34.57/0.916
GGPIR$_{S7}$	71K	<u>38.16</u>/0.964	**46.37/0.996**	<u>35.58</u>/0.937
GGPIR$_{S9}$	83K	**38.47/0.968**	44.64/0.995	34.52/0.913
GGPIR$_{S11}$	95K	37.74/<u>0.968</u>	44.51/0.995	34.37/0.915
GGPIR$_{M7}$	71K	38.14/0.964	<u>46.37</u>/<u>0.996</u>	35.56/<u>0.938</u>
GGPIR$_{SC}$	71K	37.11/0.965	44.87/0.995	34.69/0.923

Table 3: Ablation study on various configurations of the proposed hyperspectral image reconstruction framework GGPIR in comparison with two SoTA architectures, M_1 and M_2 denotes PSNR and SSIM metrics. The **best** and <u>second best</u> results are highlighted.

4.4 Qualitative Assessment

The proposed experimental setup was evaluated against ground truth data from a Specim IQ camera. Comparisons of reconstructed RGB imagery and spectral profiles (Figure 5) reveal a close alignment between our acquisition and reconstruction and the reference across all 31 spectral bands. The system achieved exposure times of 15 ms and 200 ms for halogen and RGB-LED light sources, respectively. These exposure times are comparable to those of the Specim IQ camera, confirming the feasibility of our snapshot implementation.

5 Conclusion

We presented an easy-to-implement snapshot hyperspectral imaging system that captures 31 spectral bands using a plug-and-play optical adapter for conventional RGB cameras. The system combines an array of 16 harmonic diffractive lenses with a lightweight neural reconstruction model, enabling hyperspectral imaging without dedicated, less accessible devices. Each diffractive lens exploits multi-order diffraction to encode multiple spectral channels into RGB measurements. For reconstruction, we proposed GGPIR, a Gaussian-primitive-based framework with fewer than 70k parameters, which achieves accuracy comparable to or better than state-of-the-art models. Experimental results demonstrate consistent improvements of at least 5 dB over RGB-only reconstruction and prior optical adapter-based methods. Our results highlight the effectiveness of jointly designing diffractive optical encoding and neural reconstruction for practical hyperspectral imaging. The source code and supplementary is available at https://github.com/Liptee/Diffract-and-conquer-hsi.

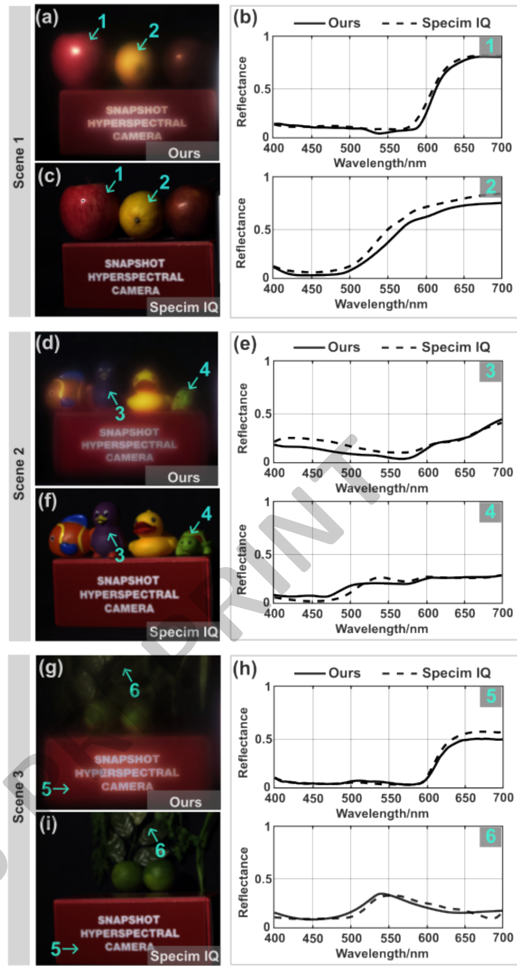

Figure 5: Experimental Assessment. We evaluate the proposed experimental snapshot hyperspectral camera with reconstruction in an indoor environment under halogen lamp illumination (scenes 1, 3) and RGB-LED illumination (scene 2), comparing the results against the commercial Specim IQ hyperspectral camera. Reconstructed RGB images compared with the reference Specim IQ: (a), (c), (d), (f), (g), (i). Spectral validation plots for six control points: (b), (e), (h).

Acknowledgments

This work was supported by the Ministry of Economic Development of the Russian Federation (agreement identifier 000000C313925P3U0002, grant No 139-15-2025-003 dated 16.04.2025).

References

[Arad and Ben-Shahar, 2016] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *Proceedings of the IEEE/CVF the European Conference on Computer Vision*, 2016.

[Arad and Timofte, 2022] Boaz Arad and Radu Timofte. Ntire 2022 spectral recovery challenge and data set. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[Arad *et al.*, 2020] Boaz Arad, Jos Balendonck, Ruud Barth, Ohad Ben-Shahar, Yael Edan, Thomas Hellström, Jochen Hemming, Polina Kurtser, Ola Ringdahl, Toon Tielen, et al. Development of a sweet pepper harvesting robot. *Journal of field robotics*, 37(6):1027–1039, 2020.

[Arce *et al.*, 2013] 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*, 31(1):105–115, 2013.

[Arguello *et al.*, 2021] 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.

[Baek *et al.*, 2017] Seung-Hwan Baek, Incheol Kim, Diego Gutierrez, and Min H Kim. Compact single-shot hyperspectral imaging using a prism. *ACM Transactions on Graphics (TOG)*, 36(6):1–12, 2017.

[Benelli *et al.*, 2020] Alessandro Benelli, Chiara Cevoli, Angelo Fabbri, et al. In-field hyperspectral imaging: An overview on the ground-based applications in agriculture. *Journal of Agricultural Engineering*, 51(3):129–139, 2020.

[Braun and Griebel, 2009] Jürgen Braun and Michael Griebel. On a constructive proof of kolmogorov's superposition theorem. *Constructive approximation*, 30(3):653–675, 2009.

[Cai *et al.*, 2022a] Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Coarse-to-fine sparse transformer for hyperspectral image reconstruction. In *European conference on computer vision*, pages 686–704. Springer, 2022.

[Cai *et al.*, 2022b] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17502–17511, 2022.

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

[Chakrabarti and Zickler, 2011] Ayan Chakrabarti and Todd Zickler. Statistics of real-world hyperspectral images. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2011.

[Doskolovich *et al.*, 2020] Leonid L Doskolovich, Roman V Skidanov, Evgeni A Bezus, Sofiya V Ganchevskaya, Dmitry A Bykov, and Nikolay L Kazanskiy. Design of diffractive lenses operating at several wavelengths. *Optics Express*, 28(8):11705–11720, 2020.

[Firsov *et al.*, 2024] Nikita Firsov, Evgeny Myasnikov, Valeriy Lobanov, Roman Khabibullin, Nikolay Kazanskiy, Svetlana Khonina, Muhammad A Butt, and Artem Nikonorov. Hyperkan: Kolmogorov–arnold networks make hyperspectral image classifiers smarter. *Sensors*, 24(23):7683, 2024.

[Fu *et al.*, 2025] Qiang Fu, Matheus Souza, Eunsue Choi, Suhyun Shin, Seung-Hwan Baek, and Wolfgang Heidrich. Limitations of data-driven spectral reconstruction: an optics-aware analysis. *IEEE Transactions on Computational Imaging*, 2025.

[Hubold *et al.*, 2021] Martin Hubold, Elisabeth Montag, René Berlich, Robert Brunner, and Robert Brüning. Multi-aperture system approach for snapshot multispectral imaging applications. *Optics Express*, 29(5):7361–7378, 2021.

[Kakani *et al.*, 2020] Vijay Kakani, Van Huan Nguyen, Basivi Praveen Kumar, Hakil Kim, and Visweswara Rao Pasupuleti. A critical review on computer vision and artificial intelligence in food industry. *Journal of Agriculture and Food Research*, 2:100033, 2020.

[Kar and Dhal, 2025] Debashish Kar and Sambandh Bhusan Dhal. Advancing food security through drone-based hyperspectral imaging: applications in precision agriculture and post-harvest management. *Environmental monitoring and assessment*, 197(3):283, 2025.

[Kiyokawa *et al.*, 2022] Takuya Kiyokawa, Jun Takamatsu, and Shigeki Koyanaka. Challenges for future robotic sorters of mixed industrial waste: a survey. *IEEE Transactions on Automation Science and Engineering*, 21(1):1023–1040, 2022.

[Kruse *et al.*, 1993] Fred A Kruse, Adam B Lefkoff, Joseph W Boardman, Kathleen B Heidebrecht, AT Shapiro, PJ Barloon, and Alexander FH Goetz. The spectral image processing system (sips)—interactive visualization and analysis of imaging spectrometer data. *Remote sensing of environment*, 44(2-3):145–163, 1993.

[Li *et al.*, 2022] 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*, pages 19780–19789, 2022.

[Li *et al.*, 2024] Chenyu Li, Bing Zhang, Danfeng Hong, Jun Zhou, Gemine Vivone, Shutao Li, and Jocelyn Chanussot. Casformer: Cascaded transformers for fusion-aware computational hyperspectral imaging. *Information Fusion*, 108:102408, 2024.

[Li, 2024] Ziyao Li. Kolmogorov-arnold networks are radial basis function networks. *arXiv preprint arXiv:2405.06721*, 2024.

[Lu *et al.*, 2023] Jiawei Lu, Desheng Zheng, Razvan-Ionut Stoian, Christopher Flynn, David Alexander, and

Tomasz S Tkaczyk. Visible and short-wave infrared fiber-based snapshot imaging spectrometer with a custom high-throughput relay system. *Optics Continuum*, 2(5):1106–1121, 2023.

[Ma *et al.*, 2024] Jingyue Ma, Zhenming Yu, Liming Cheng, Jiayu Di, Ning Zhan, Yue Zhou, Haiying Zhao, and Kun Xu. Ultra-high-speed four-dimensional hyperspectral imaging. *Optics Express*, 32(11):19684–19696, 2024.

[Meng *et al.*, 2020] Ziyi Meng, Jiawei Ma, and Xin Yuan. End-to-end low cost compressive spectral imaging with spatial-spectral self-attention. In *European conference on computer vision*, pages 187–204. Springer, 2020.

[Meng *et al.*, 2025] Ge Meng, Jingyan Tu, Jingjia Huang, Yunlong Lin, Yingying Wang, Xiaotong Tu, Yue Huang, and Xinghao Ding. Sp3ctralmamba: Physics-driven joint state space model for hyperspectral image reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 6108–6116, 2025.

[Mu *et al.*, 2019] Tingkui Mu, Feng Han, Donghao Bao, Chunmin Zhang, and Rongguang Liang. Compact snapshot optically replicating and remapping imaging spectrometer (orris) using a focal plane continuous variable filter. *Optics Letters*, 44(5):1281–1284, 2019.

[Mukhtar *et al.*, 2025] Sani Mukhtar, Amir Arbabi, and Jaime Viegas. Advances in spectral imaging: A review of techniques and technologies. *IEEE Access*, 2025.

[Nikonorov *et al.*, 2025] Artem Nikonorov, Georgy Perevozchikov, Andrei Korepanov, Nancy Mehta, Mahmoud Afifi, Egor Ershov, and Radu Timofte. Color matching using hypernetwork-based kolmogorov-arnold networks. *arXiv preprint arXiv:2503.11781*, 2025.

[Picone *et al.*, 2023] Daniele Picone, Silvère Gousset, Mauro Dalla Mura, Yann Ferrec, and Etienne Le Coarer. Interferometer response characterization algorithm for multi-aperture fabry-perot imaging spectrometers. *Optics Express*, 31(14):23066–23085, 2023.

[Shi *et al.*, 2024] Zheng Shi, Xiong Dun, Haoyu Wei, Siyu Dong, Zhanshan Wang, Xinbin Cheng, Felix Heide, and Yifan Peng. Learned multi-aperture color-coded optics for snapshot hyperspectral imaging. *ACM Transactions on Graphics (TOG)*, 43(6):1–11, 2024.

[Sweeney and Sommargren, 1995] Donald W Sweeney and Gary E Sommargren. Harmonic diffractive lenses. *Applied Optics*, 34(14):2469–2475, 1995.

[Ta, 2024] Hoang-Thang Ta. Bsrbf-kan: a combination of b-splines and radial basis functions in kolmogorov-arnold networks. In *International Symposium on Information and Communication Technology*, pages 3–15. Springer, 2024.

[Tamin *et al.*, 2023] Owen Tamin, Ervin Gubin Moung, Jamal Ahmad Dargham, Farashazillah Yahya, and Sigeru Omatu. A review of hyperspectral imaging-based plastic waste detection state-of-the-arts. *Int. J. Electr. Comput. Eng.(IJECE)*, 13(3):3407, 2023.

[Wang *et al.*, 2004] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[Wang *et al.*, 2024] Pengyu Wang, Guoliang Tang, Shijie Liu, Yang Yang, Shouzheng Zhu, Senyuan Wang, Xinze Liu, Jiawei Lu, Jianyu Wang, Chunlai Li, et al. Study on the lower limit of gas detection based on the snapshot infrared multispectral imaging system. *Optics Express*, 32(16):27919–27930, 2024.

[Wu *et al.*, 2024] Yaohang Wu, Renwei Dian, and Shutao Li. Multistage spatial–spectral fusion network for spectral super-resolution. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–11, 2024.

[Xu *et al.*, 2023] 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. *Optics Express*, 31(12):20489–20504, 2023.

[Yako *et al.*, 2023] Motoki Yako, Yoshikazu Yamaoka, Takayuki Kiyohara, Chikai Hosokawa, Akihiro Noda, Klaas Tack, Nick Spooren, Taku Hirasawa, and Atsushi Ishikawa. Video-rate hyperspectral camera based on a cmos-compatible random array of fabry–pérot filters. *Nature Photonics*, 17(3):218–223, 2023.

[Yang *et al.*, 2024] Xingxing Yang, Jie Chen, and Zaifeng Yang. Hyperspectral image reconstruction via combinatorial embedding of cross-channel spatio-spectral clues. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 6567–6575, 2024.

[Yang *et al.*, 2026] Xingxing Yang, Jie Chen, and Zaifeng Yang. Unraveling metameric dilemma for spectral reconstruction: A high-fidelity approach via semi-supervised learning. *Advances in Neural Information Processing Systems*, 38:134423–134446, 2026.

[Yao *et al.*, 2024] Zhiyang Yao, Shuyang Liu, Xiaoyun Yuan, and Lu Fang. Specat: Spatial-spectral cumulative-attention transformer for high-resolution hyperspectral image reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25368–25377, 2024.

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

[Zhao *et al.*, 2023] Ruixuan Zhao, Qi Cui, Zhaoqiang Wang, and Liang Gao. Coded aperture snapshot hyperspectral light field tomography. *Optics Express*, 31(22):37336–37347, 2023.