ADVANCED PHOTONICS RESEARCH Open Access

www.adpr-journal.com

Metasurface-Encoded Single-Pixel Hyperspectral Imaging

*Haitao Nie, Yaping Zhao, Yifei Zhang, Yanmin Zhu, Jingyan Chen, Yunfei Tian, and Edmund Y. Lam**

Hyperspectral imaging in the visible spectrum offers significant potential for diverse applications, but is often constrained by bulky hardware and limited robustness in low-light conditions. To overcome these challenges, a simulation-based proof-of-concept for a metasurface-encoded single-pixel hyperspectral imaging system (MESH) is presented, in which structured spatial modulation is combined with a compact set of 50 broadband metasurface filters designed using a binary pattern generation strategy to ensure low interfilter correlation. Hyperspectral datacubes comprising 301 channels from 400 to 700 nm are reconstructed via a sparsity-constrained optimization algorithm, while a physics-enhanced deep learning model is further introduced to enable fast and accurate recovery. Simulation results demonstrate that MESH achieves a spectral resolution of 1.17 nm. Even at a total compression ratio of 2.1%, the deep learning model maintains high reconstruction quality, with a peak signal-to-noise ratio of 30.96 dB, structural similarity of 0.8526, and spectral angle mapping of 0.0742 rad, indicating accurate intensity recovery, structural preservation, and spectral integrity. The present study provides a simulation-based verification of feasibility and design guidelines, laying the groundwork for future experimental validation of the MESH system, which is expected to further demonstrate its practical applicability and performance for deployment in low-light and resource-constrained environments.

1. Introduction

Hyperspectral imaging (HSI) is a powerful optical imaging technique that simultaneously acquires spatial and detailed spectral information.[1,2] In contrast to conventional RGB imaging, which provides only limited spectral channels, HSI captures densely sampled spectral profiles at every spatial position. This capability enables sensitive detection and identification of subtle chemical or structural differences, making HSI an invaluable tool in various critical fields such as remote sensing,[3] biomedical diagnostics,[4] environment monitoring,[5,6] and material identification.[7,8]

Despite these significant advantages, conventional HSI faces three major practical limitations. First, traditional HSI approaches typically rely on sequential scanning in spatial or spectral dimension,[9] resulting in slow acquisition speed and bulky optical systems. Second, directly capturing hyperspectral data incurs high storage demands and large transmission bandwidth,[10] which, in turn, increase power consumption and processing overhead, imposing severe constraints on system performance—particularly in portable or embedded imaging applications.[11] Third, under low-light or photon-starved conditions, conventional HSI systems often suffer from poor signal-to-noise ratio (SNR),[12] leading to degraded image quality and unreliable spectral reconstructions, which further hinders their use in real-time or field applications.

To address these issues, single-pixel HSI based on compressive sensing (CS) theory has attracted growing interest,[13–15] as it enables hyperspectral datacube reconstruction from a sequence of encoded measurements acquired by a single-pixel detector while reducing data dimensionality and improving SNR under low-light conditions. To spectrally resolve incident light in such systems, researchers have employed various strategies, including dispersive optics,[13,16,17] interferometric architectures,[14] and broadband optical filters.[15,18] While each approach offers specific benefits, they also involve inherent trade-offs. Grating-based methods typically require bulky optical setups, which limit their portability, whereas interferometric approaches demand precise optical alignment and exceptional system stability, making them highly sensitive to vibrations and environmental disturbances. Broadband optical filters offer advantages such as compact structure, alignment-free operation, and ease of integration, making them widely used in portable imaging systems. However, the limited design flexibility of such filters hinders the implementation of complex and versatile spectral encoding schemes, which are crucial for achieving high spectral resolution in compressive reconstruction, thereby constraining the overall performance of the imaging system.

Meanwhile, metasurfaces have emerged as powerful optical components that enable versatile manipulation of the amplitude,

H. Nie, Y. Zhao, Y. Zhu, J. Chen, Y. Tian, E. Y. Lam
Department of Electrical and Electronic Engineering
The University of Hong Kong
Hong Kong, China
E-mail: elam@eee.hku.hk

Y. Zhang
Tsinghua Shenzhen International Graduate School
Tsinghua University
Shenzhen 518055, China

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adpr.202500181.

DOI: 10.1002/adpr.202500181

phase, and polarization of the light field.[19–21] In particular, dielectric metasurfaces offer high optical efficiency and broad spectral modulation capability, making them well suited for spectral encoding in HSI systems.[22–25] Prior work has demonstrated the use of phase-change metasurfaces as tunable filters for spectral modulation, enabling single-pixel HSI in the short-wave infrared range.[26] In this work, we propose the metasurface-encoded single-pixel hyperspectral imaging system (MESH), which in the visible range integrates broadband spectral encoding using metasurface-based optical filters, single-pixel detection, and computational reconstruction algorithms into a unified framework. Furthermore, we develop a physics-enhanced deep learning reconstruction approach to accelerate and improve hyperspectral recovery, which incorporates prior knowledge of the optical encoding process to improve reconstruction fidelity. This integration of broadband metasurface encoding, efficient single-pixel measurements, and data-driven reconstruction is demonstrated through simulations as a compact and high-performance framework suitable for low-light and resource-constrained HSI applications.

2. Experimental Section

2.1. Joint Spatial–Spectral Acquisition and Reconstruction

The architecture of the proposed MESH system is illustrated in **Figure 1**a. As shown in Figure 1a, a broadband LED illuminates the object, which is imaged onto the digital micromirror device (DMD) by lens L1. The DMD applies spatial encoding via binary patterns, directing the modulated light toward a mirror. The beam then passes through lens L2 and enters a rotating filter wheel that holds a set of metasurface-based broadband spectral filters. Each filter exhibits a unique wavelength-dependent transmission profile over the 400–700 nm range. After spectral encoding, the light is focused by lens L3 onto a single-pixel detector. By cycling through spatial patterns and spectral filters, the system records a series of intensity measurements that compress joint spatial–spectral information from the scene into a low-dimensional set of signals.

Let the hyperspectral datacube have spatial resolution $M \times M$ and C spectral bands. It is vectorized as $\mathbf{X} \in \mathbb{R}^{M^2 \times C}$, where each row corresponds to a spatial pixel and each column to a spectral band. In our case, $M = 64$ and $C = 301$, with spectral bands uniformly sampled from 400 to 700 nm at 1 nm intervals. Spatial modulation is performed using P patterns drawn from a Hadamard basis, reordered via a cake-cutting strategy[27,28] to prioritize low-frequency structures, forming the spatial modulation matrix $\mathbf{H} \in \mathbb{R}^{P \times M^2}$. Spectral encoding is implemented using Q metasurface filters with distinct transmission profiles, which constitute the spectral modulation matrix $\mathbf{F} \in \mathbb{R}^{C \times Q}$. The resulting intensity measurements form a matrix $\mathbf{Y} \in \mathbb{R}^{P \times Q}$, where each element corresponds to one spatial–spectral projection. The forward measurement model, illustrated in Figure 1b, is given by

$$\mathbf{Y} = \mathbf{HXF} + \mathbf{N} \tag{1}$$

where $\mathbf{N}$ denotes additive measurement noise. The compression ratio for the spatial dimension is defined as P/M^2 and the compression ratio for the spectral dimension as Q/C. The total compression ratio, denoted as β, is given by the product of the two

$$\beta = \left(\frac{P}{M^2}\right)\left(\frac{Q}{C}\right) \tag{2}$$

Our goal is to recover the high-dimensional datacube $\mathbf{X}$ from the compressed measurements $\mathbf{Y}$. To make this ill-posed inverse problem tractable, we exploit the fact that hyperspectral data can be sparsely represented in an appropriate basis. For narrowband spectra, sparsity arises naturally due to energy concentration within limited spectral ranges. For broadband spectra typical of natural scenes, sparse representations can still be obtained using data-driven dictionaries learned from hyperspectral datasets; see Note S1, Supporting Information, for implementation details. In this formulation, the datacube $\mathbf{X} \in \mathbb{R}^{M^2 \times C}$ is expressed as $\mathbf{X} = \mathbf{Z}\mathbf{\Psi}^T$, where $\mathbf{\Psi} \in \mathbb{R}^{C \times S}$ is a spectral basis, $\mathbf{Z} \in \mathbb{R}^{M^2 \times S}$ contains sparse coefficients, and $\mathbf{\Psi}^T$ denotes the matrix transpose of $\mathbf{\Psi}$.

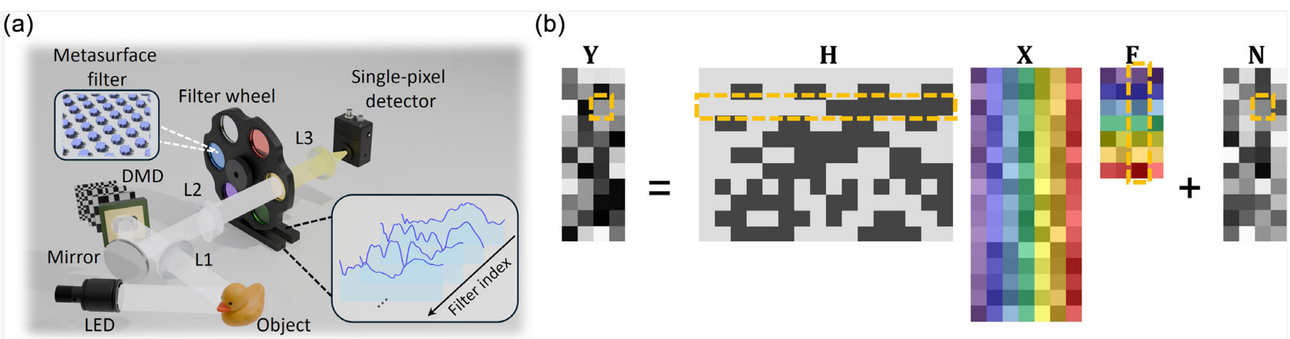

Figure 1. a) Schematic diagram of MESH. The object is illuminated by a broadband LED and imaged onto the DMD by lens L1. Reflected light is directed by a mirror and passes through lens L2, which guides the modulated beam toward the rotating filter wheel equipped with metasurface filters. The spectrally encoded light is then focused by lens L3 onto a single-pixel detector. b) The measurements of a hyperspectral datacube represented as a series of matrix multiplications in the spatial and spectral domains. The orange dashed lines highlight a single spatial pattern and a single spectral filter, whose combined response contributes to one single-pixel measurement.

Substituting this representation into the forward model yields

$$\mathbf{Y} = \mathbf{H}\mathbf{Z}\mathbf{\Psi}^T\mathbf{F} + \mathbf{N} \tag{3}$$

To facilitate optimization, we vectorize the measurement equation using the identity $\text{vec}(\mathbf{AZB}) = (\mathbf{B}^T \otimes \mathbf{A})\text{vec}(\mathbf{Z})$, where $\otimes$ denotes the Kronecker product and $\text{vec}(\cdot)$ denotes the columnwise vectorization of a matrix. Let $\mathbf{y} = \text{vec}(\mathbf{Y}) \in \mathbb{R}^{PQ\times 1}$ and $\mathbf{z} = \text{vec}(\mathbf{Z}) \in \mathbb{R}^{M^2S\times 1}$. The reconstruction problem can then be formulated as the following Lasso optimization, where $\gamma > 0$ is a regularization parameter that controls the sparsity level of the solution

$$\min_{\mathbf{z}} \|\mathbf{y} - ((\mathbf{F}^T\mathbf{\Psi}) \otimes \mathbf{H})\mathbf{z}\|_2^2 + \gamma\|\mathbf{z}\|_1 \tag{4}$$

After solving this problem, the datacube is recovered by reshaping $\mathbf{z}$ into $\mathbf{Z} \in \mathbb{R}^{M^2\times S}$ and computing $\hat{\mathbf{X}} = \mathbf{Z}\mathbf{\Psi}^T$.

In the special case where the spatial modulation matrix $\mathbf{H} \in \mathbb{R}^{M^2\times M^2}$ consists of a complete Hadamard basis, the measurement model simplifies significantly. Leveraging the orthogonality property $\mathbf{H}^T\mathbf{H} = M^2\mathbf{I}$, we can left-multiply the measurement equation by $\mathbf{H}^T/M^2$, yielding

$$\frac{1}{M^2}\mathbf{H}^T\mathbf{Y} = \mathbf{X}\mathbf{F} + \frac{1}{M^2}\mathbf{H}^T\mathbf{N} \tag{5}$$

Let $\widetilde{\mathbf{Y}} = \frac{1}{M^2}\mathbf{H}^T\mathbf{Y} \in \mathbb{R}^{M^2\times Q}$. Each row $\widetilde{\mathbf{y}}_i^T \in \mathbb{R}^{1\times Q}$ of $\widetilde{\mathbf{Y}}$ corresponds to a linear projection of the spectral signature at a specific spatial location through the metasurface filters. Let $\widetilde{\mathbf{y}}_i \in \mathbb{R}^{Q\times 1}$ denote its column vector form used in the reconstruction. Each spectral signature is recovered by solving the following optimization problem

$$\min_{\mathbf{z}_i} \|\widetilde{\mathbf{y}}_i - \mathbf{F}^T\mathbf{\Psi}\mathbf{z}_i\|_2^2 + \gamma\|\mathbf{z}_i\|_1 \tag{6}$$

where $\mathbf{z}_i \in \mathbb{R}^{S\times 1}$ is the sparse coefficient vector, and the reconstructed spectrum is given by $\hat{\mathbf{x}}_i = \mathbf{\Psi}\mathbf{z}_i$. Here, $\hat{\mathbf{x}}_i^T \in \mathbb{R}^{1\times C}$ corresponds to the ith row of the reconstructed datacube $\hat{\mathbf{X}} \in \mathbb{R}^{M^2\times C}$. Since each spatial location is processed independently, the reconstruction can be implemented in a fully parallel manner, leading to high computational efficiency. Further implementation details of the optimization solver are provided in Note S1, Supporting Information.

2.2. Design of Metasurface-Based Optical Filter

According to the principles of CS, the spectral encoding functions associated with different filters should exhibit minimal mutual correlation to ensure a well-conditioned inverse problem and enhance reconstruction fidelity. To meet this requirement, we aim to design a set of metasurface filters that produce diverse, low-redundancy spectral transmission profiles across the visible spectrum. Crystalline silicon is selected for its high refractive index contrast, which enables broadband spectral modulation across the visible range.

To achieve this, we adopt a binary pattern generation strategy that also imposes rotational symmetry to maintain polarization invariance. As illustrated in **Figure 2**a, each pattern is initialized as a 3 × 3 binary matrix with diagonal symmetry. This seed is rotated by 90°, 180°, and 270° to produce four variants,

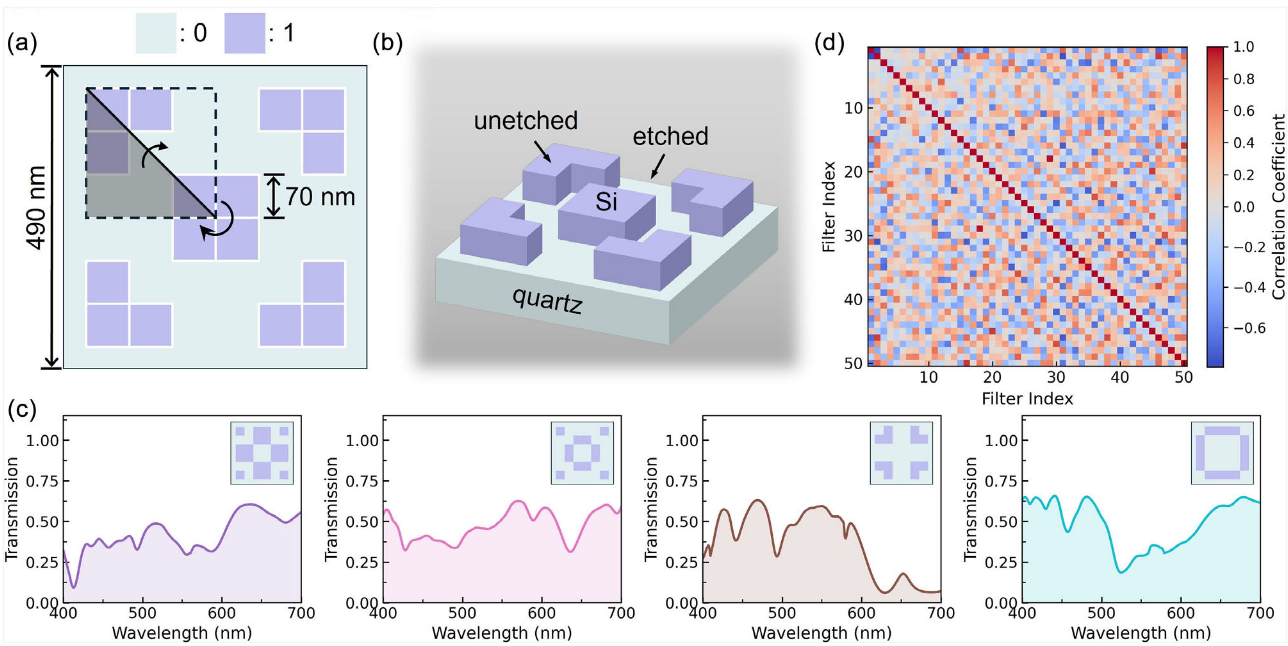

Figure 2. Design and selection of broadband spectral encoders using metasurface filters. a) Schematic of the pattern generation process: a 3 × 3 binary matrix is constructed with diagonal symmetry and then rotated to obtain fourfold rotational (C_4) symmetry. b) Perspective view of a metasurface unit cell, consisting of etched and unetched regions in a silicon-on-quartz structure. c) Representative geometries and their simulated transmission spectra. d) Correlation matrix of the selected 50 spectral encoders.

which are then tiled to form a 6×6 binary mask exhibiting fourfold rotational symmetry (C_4). Each binary mask corresponds to a physical metasurface design, where pixels marked as "1" and "0" represent unetched and etched regions, respectively, on a silicon-on-quartz substrate. A representative unit cell is shown in Figure 2b. The metasurface consists of a 220 nm thick crystalline silicon layer deposited on a 0.5 mm quartz substrate. The active area measures $420 \times 420\,\text{nm}^2$ and is divided into a 6×6 pixel grid, with a 35 nm-wide isolation trench introduced around each unit cell to delineate its boundary from adjacent patterns. The imposed C_4 symmetry ensures polarization-independent spectral responses,[22,24] thereby eliminating polarization-induced variability in the spectral encoding process. The designed binary pattern is periodically repeated to form a large-area metasurface array, and the incident beam can be efficiently coupled into the metasurface area for single-pixel detection. A feasible fabrication route for realizing such metasurface filters is outlined in Note S2, Supporting Information.

We generate 63 unique metasurface configurations through the pattern design process described above. To obtain the spectral response of each design, we perform finite-difference time-domain simulations, with each configuration yielding a distinct broadband transmission spectrum. Figure 2c shows representative examples of the simulated geometries and their corresponding spectral responses. To construct a well-conditioned spectral modulation matrix $\mathbf{F}$, we select a subset of 50 filters with minimal pairwise spectral correlation, as detailed in Note S3, Supporting Information. In practice, we note that fabricated metasurfaces may exhibit deviations from simulated transmission spectra due to fabrication imperfections; therefore, the actual spectral responses will need to be experimentally characterized using a spectrometer, which are used in real hyperspectral reconstruction experiment. The spectral compression ratio of our system is therefore 50/301 ≈16.6%. Figure 2d presents the correlation matrix for the 50 selected filters. The average absolute correlation coefficient across the 50 filters is 0.2672, demonstrating low cross-correlation among encoders. The full spectral responses are provided in Note S3, Supporting Information.

3. Results

In this section, we first evaluate the spectral resolution of the proposed MESH system. We then present reconstruction results on natural scene datasets using an optimization-based method. To enable faster inference and explore spatial compression, we introduce a deep learning-based approach and examine its performance under varying spatial compression ratios. Finally, we investigate the robustness of the deep learning-based model to different levels of measurement noise. Quantitative evaluation is performed using peak SNR (PSNR), structural similarity (SSIM), and spectral angle mapping (SAM), providing a comprehensive assessment of reconstruction fidelity.

To simulate realistic measurement conditions, we first compute the noise-free measurement matrix $\mathbf{B} = \mathbf{HXF}$. Additive Gaussian noise is then introduced to generate the final measurements $\mathbf{Y} = \mathbf{B} + \mathbf{N}$, where $\mathbf{N}$ is drawn from a zero-mean normal distribution with variance σ^2. The variance σ^2 is determined based on a target SNR, which is defined as

$$\mathrm{SNR} = 10\log_{10}\left(\frac{\frac{1}{PQ}\sum_{i=1}^{P}\sum_{j=1}^{Q} B_{ij}^2}{\sigma^2}\right) \tag{7}$$

where $P \times Q$ is the size of $\mathbf{B}$. Unless otherwise specified, the default SNR is set to 40 dB.

3.1. Numerical Evaluation of the Spectral Resolution

We first perform numerical evaluations of the spectral resolving capability at multiple wavelength positions across the visible range (from 410 to 690 nm at 10 nm intervals) using a synthetic hyperspectral scene composed of spatially distinct spectral signatures. As a representative example, we show here the results around 410 nm. As shown in **Figure 3**a (top row), the image contains the characters "HKU" in the upper half and "ISL" in the lower half, each assigned a Gaussian spectral profile centered at 410 and 411 nm, respectively, with a full width at half maximum (FWHM) of 1 nm.

To more precisely verify the system's ability to resolve such narrow spectral features, we increase the spectral sampling density by halving the interval to 0.5 nm while maintaining the 400–700 nm range. This yields a total of 601 spectral channels in the datacube, instead of the standard 301 used elsewhere in the article. For visualization, Figure 3a displays a zoomed-in spectral range from 409.0 to 412.0 nm, where the two peaks are most distinguishable.

We reconstruct the datacube under additive Gaussian noise corresponding to a SNR of 40 dB, using full spatial sampling. As shown in Figure 3a (bottom row), the reconstructed spectral slices accurately highlight each character at its corresponding wavelength, with minimal interference from neighboring bands.

To assess spectral fidelity, we average the pixel intensities within each character region and plot the resulting spectral curves. As shown in Figure 3b, the reconstructed spectra closely match the ground truth in both center wavelength and shape. Each reconstructed spectrum exhibits a peak with a FWHM of 1.17 nm for the "HKU" region and 1.12 nm for the "ISL" region, with negligible signal observed in channels more than 1 nm away from the center wavelength.

The spectral resolution of a single-pixel HSI system is defined as the FWHM of the reconstructed response to a narrowband input, or equivalently as the minimum channel spacing that avoids crosstalk from a spectrally localized signal.[12,14,29] In our case, we input two 1 nm FWHM Gaussian spectra separated by 1 nm, and the recovered spectra exhibit peaks with FWHMs of 1.17 nm for the "HKU" region and 1.12 nm for the "ISL" region, respectively. Taking the average of the two, the spectral resolution of the proposed MESH system at 410 nm is determined to be 1.15 nm. The spectral resolution across the visible range is shown in Figure 3c, with a mean value of 1.17 nm. These results confirm the system's ability to resolve adjacent spectral features with sub-2 nm spacing and minimal crosstalk. A comparative analysis between the proposed MESH system and existing state-of-the-art single-pixel HSI systems is provided in Note S4, Supporting Information.

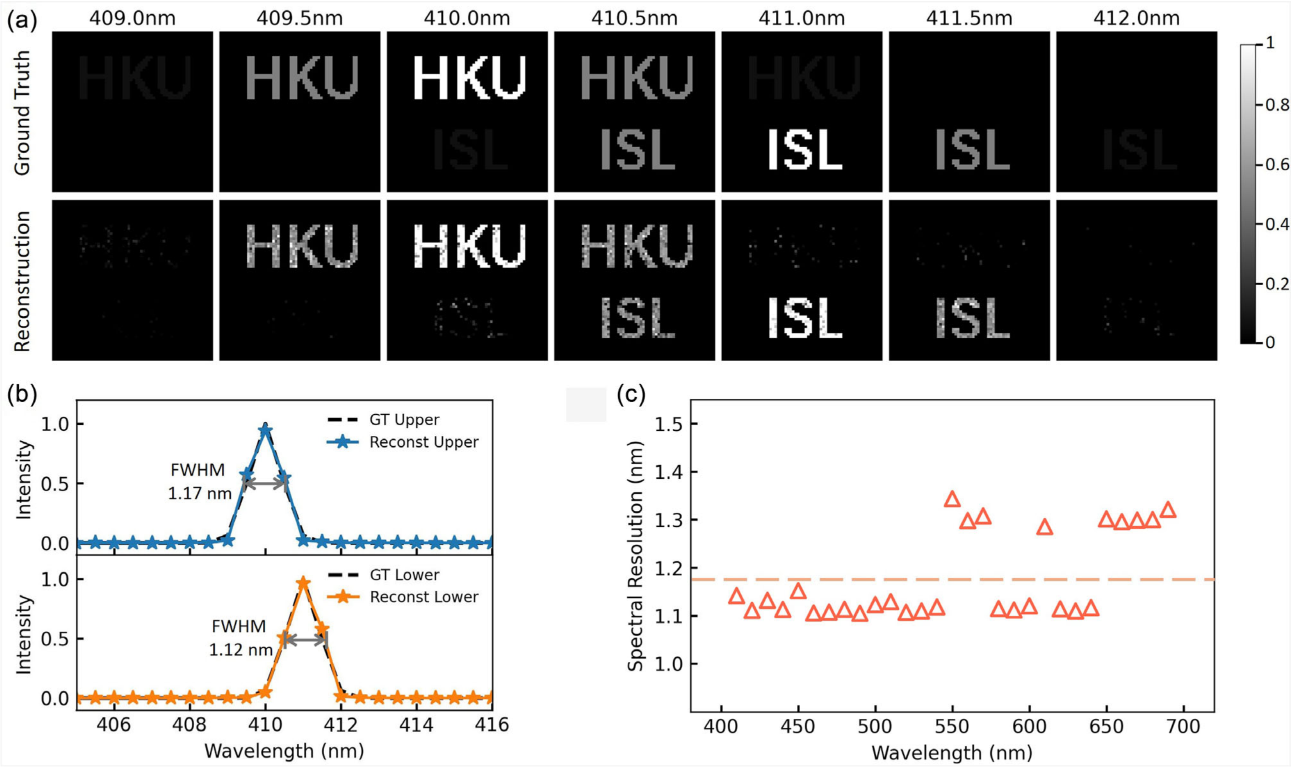

Figure 3. Numerical evaluations of the spectral resolution using synthetic narrowband patterns. a) Visualization of selected spectral slices for a test image composed of characters "HKU" (upper half) and "ISL" (lower half), each assigned a distinct Gaussian spectral signature with 1 nm FWHM. Top row: ground truth slices; bottom row: reconstructed results. b) Averaged reconstructed spectra for each region compared with the corresponding ground truth curves. c) Spectral resolution across the visible band, with a mean value of 1.17 nm (dashed line).

3.2. HSI Results of MESH

To qualitatively demonstrate the spectral reconstruction performance of MESH, we first conduct a numerical experiment on a color checker image selected from the ARAD_1K hyperspectral dataset.[30] To focus exclusively on spectral reconstruction performance, we employ full spatial sampling using Hadamard patterns in this experiment.

Prior to experiment, the original hyperspectral image is center-cropped and resized to a spatial resolution of 192×192. The spectral dimension is then interpolated using cubic splines to obtain 301 bands uniformly sampled from 400 to 700 nm. The resulting datacube is partitioned into nonoverlapping $64 \times 64 \times 301$ patches, each independently reconstructed and later reassembled into the full-resolution output.

The reconstruction is performed on a workstation equipped with an Intel Core i9-14900 K CPU, and parallel spectral recovery is accelerated using the joblib package in Python with 32 workers. Leveraging this parallelization strategy, the full hyperspectral datacube of the color checker scene is recovered in 3 min 39 s. As shown in **Figure** 4a, the recovered color checker image—converted to an RGB rendering using the CIE color matching function—displays high visual fidelity, with all color patches accurately restored in both hue and brightness. This result affirms the system's capability to preserve spectral distinctions across a wide range of chromatic surfaces.

To further examine the quality of spectral recovery, we focus on the square region marked by the white dashed box in Figure 4a and visualize 30 spectral slices uniformly sampled from 400 to 700 nm within this region, as shown in Figure 4b. As the wavelength increases from left to right and top to bottom, different flowers and foliage gradually emerge in their respective dominant spectral bands. Specifically, violet petals appear most prominently at shorter wavelengths, green leaves dominate the midrange, and red to orange flowers become increasingly visible at longer wavelengths. This progressive transition indicates close correspondence between the reconstructed and ground-truth reflectance spectra, confirming that the reconstructed datacube reliably captures wavelength-specific chromatic variation with high spatial coherence.

For quantitative spectral evaluation, we select six representative color patches from the checker and compare the reconstructed spectra with the ground truth. As plotted in Figure 4c, the reconstructed curves (dashed red) closely match the corresponding ground truth spectra (solid blue) over the full wavelength range. Key spectral features such as peak locations and relative intensities are well preserved, indicating accurate recovery of material-specific reflectance characteristics.

We further compute quantitative metrics over the entire image, obtaining a PSNR of 40.46 dB, SSIM of 0.9729, and SAM of 0.0865 rad. These results confirm that MESH achieves high-fidelity hyperspectral reconstruction under full spatial sampling.

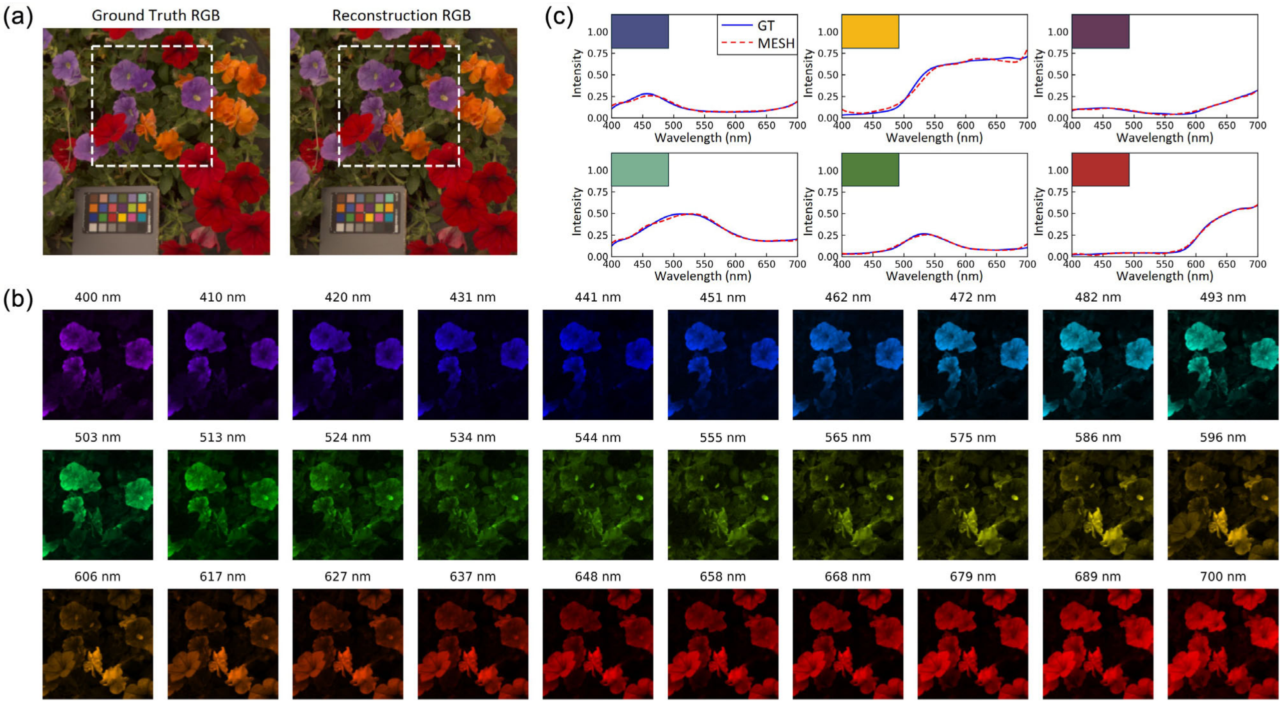

Figure 4. Hyperspectral reconstruction results on a color checker scene. a) RGB rendering of the ground truth (left) and MESH reconstruction (right), with the white dashed box indicating the region of interest. b) Visualization of 30 uniformly sampled spectral slices from 400–700 nm within the selected region. c) Comparison between ground truth and reconstructed spectra for six representative color patches; the color of the square at the upper left of each plot indicates the corresponding patch color.

3.3. Deep Learning-Based Compressive MESH

While the optimization-based reconstruction algorithm introduced earlier yields accurate spectral recovery, its iterative nature often results in considerable computation time. To overcome this limitation, we introduce deep learning-based MESH (DL-MESH), a reconstruction framework that replaces manually designed priors and iterative solvers with a data-driven learning strategy. DL-MESH is trained to directly map compressive single-pixel measurements to hyperspectral datacubes via supervised learning, enabling the network to capture complex spectral–spatial correlations and perform fast inference using GPU acceleration.

3.3.1. Hyperspectral Dataset and Preprocessing

We use the publicly available ARAD_1K dataset to evaluate the proposed deep learning-based MESH reconstruction framework. ARAD_1K contains 950 hyperspectral images of outdoor natural scenes, with spectral bands spanning from 400 to 700 nm at 10 nm intervals. Each image has a spatial resolution of 482×512 pixels and 31 spectral channels.

As the first step, all 950 images are divided into training, validation, and test sets using a 6:2:2 ratio. For each image, we crop the central 482×482 region and resize it to 192×192 pixels. To match the design target of our MESH system, which aims to recover the full visible spectrum from 400 to 700 nm at 1 nm intervals, the original 31-band spectra are interpolated to 301 bands using cubic interpolation. Each interpolated datacube is then divided into nine nonoverlapping patches of size $64 \times 64 \times 301$. This patch-based strategy is adopted due to the memory limitations of GPU during model training. The final dataset contains 5130 training patches, 1710 validation patches, and 1710 test patches.

3.3.2. DL-MESH Network

The overall pipeline of the DL-MESH framework is illustrated in **Figure 5**. DL-MESH is formulated as a physics-enhanced deep learning framework that couples the metasurface-based measurement process with a learnable decoder. The encoder is fixed by the physics of the MESH system: spatial modulation and metasurface filtering are applied to the input hyperspectral datacube to generate single-pixel measurements, and a pseudoinverse layer $\mathbf{H}^{\dagger}$ of the spatial modulation matrix is then applied to produce an initial estimate of the spatial structure under different spectral encodings. This step acts as a physics-enhanced layer that embeds optical priors directly into the network. The learnable part of the framework, referred to as the DL-Decoder, then maps this initialization to the full hyperspectral datacube.

To validate the generality of the proposed system, we implement three representative DL-Decoders within the DL-MESH framework. The first is a residual convolutional (RC) Block-based network, inspired by prior work on single-pixel HSI.[13] In our implementation, four RC blocks are stacked, and each block includes an additional residual connection from the network input to enhance feature reuse. The second decoder adopts the AWAN,[31] originally developed for RGB-to-hyperspectral

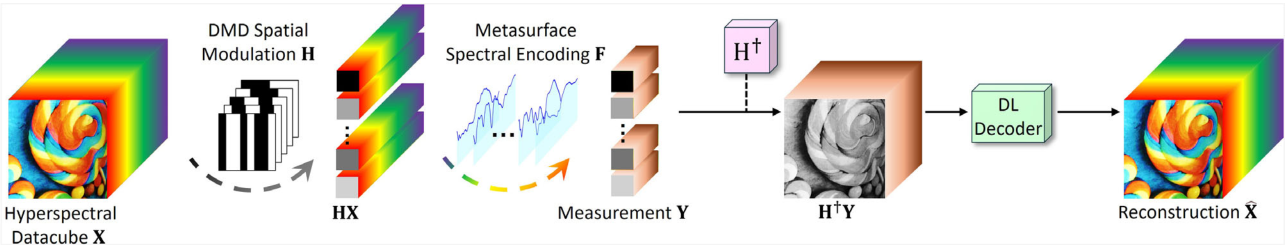

Figure 5. Pipeline of the proposed DL-MESH framework.

reconstruction, with eight dual residual attention blocks. The third decoder employs the MST++,[32] also a leading RGB-to-hyperspectral network, configured with two stages and 301 feature channels. All DL-Decoders are trained end-to-end under identical data preprocessing, optimization settings, and a mean squared error loss that quantifies the difference between input and output datacubes, ensuring a fair comparison.

3.3.3. DL-MESH Performance across Compression Ratios

To evaluate the influence of spatial compression ratio on hyperspectral reconstruction, we train and test DL-MESH under four spatial compression ratios: 1.0, 0.5, 0.25, and 0.125, corresponding to total compression ratios $\beta = 16.6\%$, 8.3%, 4.2%, and 2.1%, respectively. All experiments are conducted under a SNR of 40 dB, and each compression level is associated with a separately trained DL-MESH model.

Quantitative results for the three DL-Decoders are summarized in **Table 1**. All decoders achieve favorable performance across different compression ratios, confirming the robustness of the proposed DL-MESH framework. Among them, MST++ generally attains the highest PSNR and SSIM values, together with competitive SAM scores, indicating overall superior reconstruction quality. Notably, even under the most aggressive

Table 1. Quantitative evaluation of DL-MESH with different DL-Decoders under varying total compression ratios. The best result under each compression ratio is highlighted in bold.

β	RC Block-based			AWAN			MST++		
	PSNR [dB]	SSIM	SAM [rad]	PSNR [dB]	SSIM	SAM [rad]	PSNR [dB]	SSIM	SAM [rad]
16.6%	40.16	0.9740	0.0685	41.05	0.9827	0.0566	**44.26**	**0.9899**	**0.0466**
8.3%	36.04	0.9550	0.0752	37.42	0.9652	0.0588	**38.44**	**0.9717**	**0.0562**
4.2%	32.24	0.9051	0.0880	33.80	0.9208	0.0638	**34.25**	**0.9267**	**0.0623**
2.1%	29.64	0.8397	0.0968	30.90	0.8499	**0.0709**	**30.96**	**0.8526**	0.0742

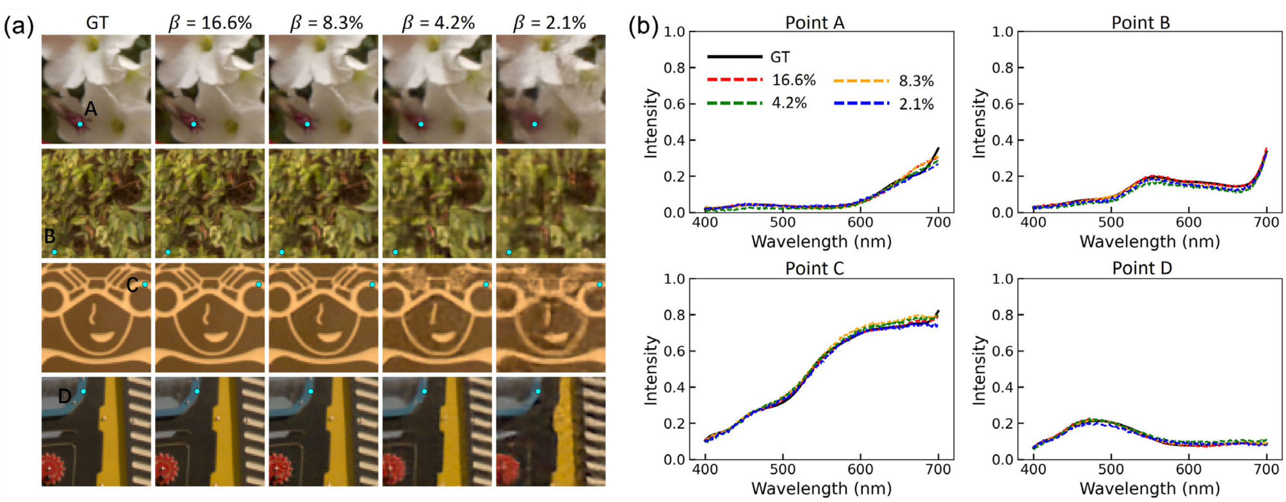

Figure 6. Performance of DL-MESH under varying total compression ratios. a) Reconstructed RGB images of four test scenes under spatial compression ratios of 1.0, 0.5, 0.25, and 0.125, corresponding to total compression ratios $\beta = 16.6\%$, 8.3%, 4.2%, and 2.1%, respectively. Cyan markers denote sampled points for spectral evaluation. b) Spectral curves at points A–D across different β levels.

sampling condition ($\beta = 2.1\%$), MST++ still achieves a PSNR of 30.96 dB, a SSIM of 0.8526, and a SAM of 0.0742 rad, demonstrating that accurate reconstruction remains feasible despite substantial undersampling. Therefore, in the following analysis, we primarily present results obtained using the MST++ decoder as a representative example. The training and inference are performed on a workstation equipped with an NVIDIA RTX 4090 GPU. For example, at a spatial compression ratio of 1.0, the reconstruction of a hyperspectral datacube with a spatial size of 192×192 and 301 spectral channels requires only 0.42 s, which is substantially shorter than the recovery time of the optimization-based algorithm.

Figure 6a shows RGB visualizations of reconstructed scenes across the four spatial compression ratios. While fine textures and high-frequency details are gradually lost at lower sampling rates, global scene structures and dominant color distributions maintain high visual fidelity. This highlights DL-MESH's ability to retain salient spatial features under severe measurement constraints.

To further assess spectral fidelity, Figure 6b plots the reconstructed spectral profiles at four representative spatial locations (A–D). Across all total compression ratios β, the recovered spectra closely follow the ground truth, exhibiting well-aligned peak positions and consistent overall shapes. This stability can be attributed to the physics-enhanced initialization in DL-MESH, which provides strong priors that help preserve spectral fidelity under reduced sampling, together with the smooth and low-dimensional nature of natural spectra that makes them amenable to accurate recovery from compressed measurements. Although minor deviations emerge under the lowest sampling level, material-specific spectral characteristics remain well preserved.

In addition to spatial compression, we also evaluate the performance of DL-MESH under varying spectral compression ratios by using 30, 40, and 50 metasurface filters at spatial

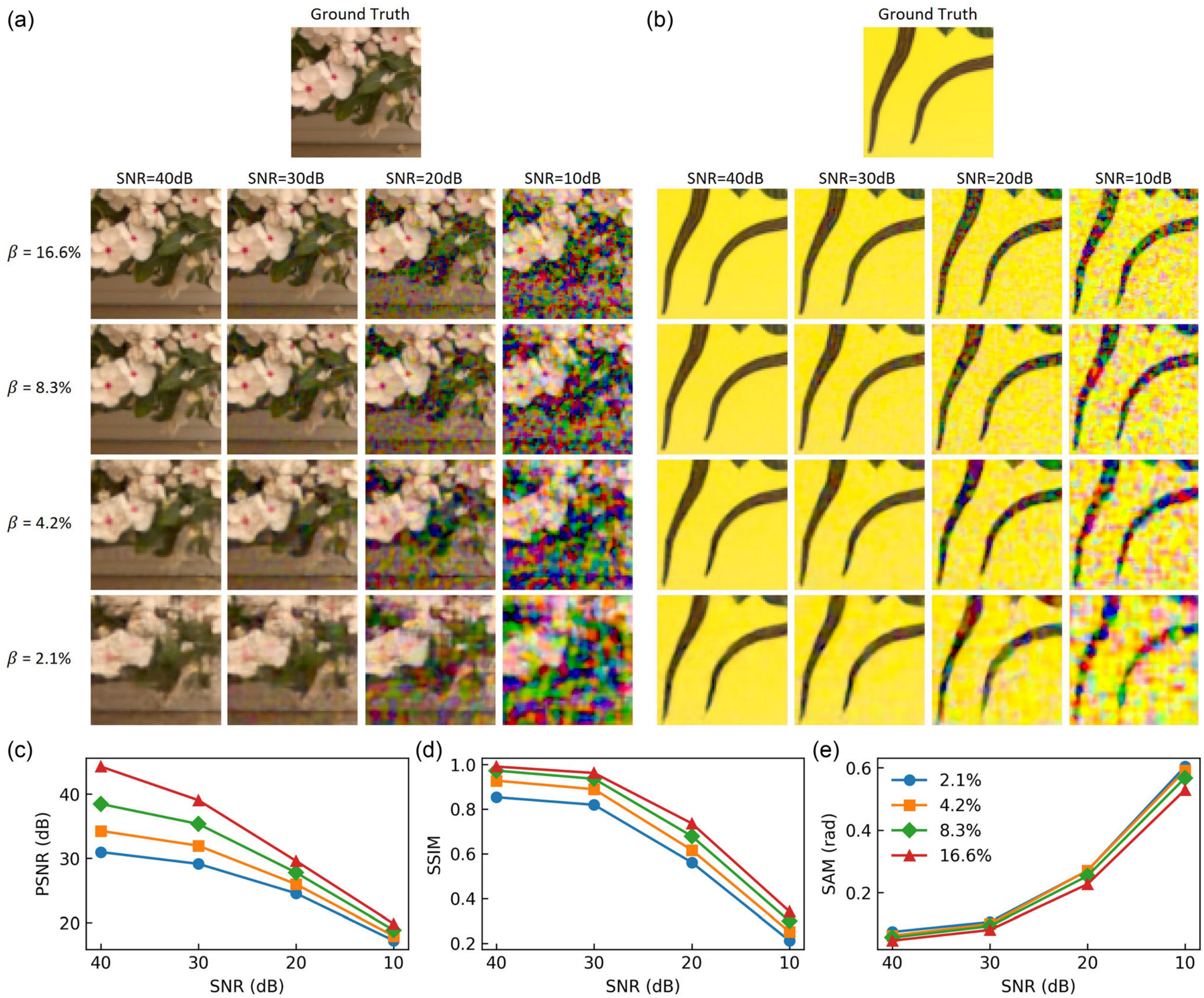

Figure 7. Visual and quantitative evaluation of DL-MESH under varying SNR and total compression ratios. a) Reconstructed RGB renderings of a flower scene and b) a fine-pattern scene. Rows represent different total compression ratios β (top to bottom: 16.6%, 8.3%, 4.2%, 2.1%), and columns represent different SNR levels (left to right: 40, 30, 20, 10 dB). c–e) Corresponding quantitative metrics: (c) PSNR, (d) SSIM, and (e) SAM.

compression ratios of 1.0 and 0.125. The results indicate that more filters improve reconstruction at high spatial sampling, whereas performance is mainly limited by spatial undersampling at low sampling. Detailed analyses are provided in Note S5, Supporting Information. These results demonstrate that DL-MESH remains effective under significant undersampling, with graceful degradation in both spatial and spectral domains.

3.3.4. Performance under Varying Noise Levels

To investigate the robustness of DL-MESH under measurement noise, we evaluate its reconstruction performance across a range of SNRs: 40, 30, 20, and 10 dB. For each spatial compression ratio (corresponding to total compression ratios $\beta = 16.6\%$, 8.3%, 4.2%, and 2.1%), a dedicated model is trained at 40 dB SNR and subsequently tested across all noise levels without further adaptation.

To qualitatively assess visual fidelity, **Figure 7**a,b presents RGB renderings of two representative test scenes reconstructed under varying noise and total compression ratios. The left block shows a natural flower scene rich in fine textures, while the right block depicts a high-contrast pattern composed of well-defined edges. In both cases, reconstruction quality progressively deteriorates with increasing noise and decreasing total compression ratio β, resulting in visible artifacts and color distortion. The flower scene, due to its low contrast and complex structures, is more susceptible to noise. In contrast, the fine-pattern scene—with simpler geometry and higher intrinsic contrast—demonstrates greater resilience to noise at comparable β levels.

Figure 7c–e summarizes the quantitative performance in terms of PSNR, SSIM, and SAM. As the SNR decreases, all metrics degrade consistently. Furthermore, models trained with lower total compression ratios β are more vulnerable to noise. Notably, at SNR $= 20$ dB, all four β settings still achieve PSNR above 24.54 dB, SSIM above 0.5609, and SAM below 0.2699 rad, indicating that the model successfully preserves fundamental spectral signatures and spatial structures under moderate noise. These results confirm that DL-MESH remains robust under moderate noise, consistently delivering stable and perceptually coherent reconstructions across diverse acquisition conditions. It is also noted that, in practical single-pixel hyperspectral measurements, the rotation of the filter wheel may introduce slight alignment errors. In some measurements, the optical path may not pass precisely through the filter center, or part of the beam may be partially blocked by the wheel frame, leading to intensity fluctuations that can be regarded as additional measurement noise. In such cases, the noise robustness of DL-MESH, together with accurate control of the filter wheel rotation, becomes particularly important for maintaining reliable spectral reconstruction.

4. Conclusion

In conclusion, we present a simulation-based proof-of-concept of MESH, a single-pixel HSI system based on metasurface-encoded spectral filtering. The system employs a C_4-symmetric binary pattern generation strategy to generate a set of 50 metasurface filters with low mutual correlation, achieving an average absolute correlation coefficient of 0.2672. This low interfilter correlation is favorable for CS, as it improves the conditioning of the inverse problem and enhances reconstruction stability. By jointly integrating structured spatial modulation, tailored spectral encoding, and sparsity-constrained reconstruction algorithms, MESH enables efficient compressive acquisition and accurate reconstruction of hyperspectral datacubes across 301 spectral channels within the visible range (400–700 nm).

Numerical evaluations indicate that the proposed MESH system can resolve two regions with 1 nm-FWHM Gaussian spectra, yielding an average spectral resolution of 1.17 nm. The system performs consistently well on natural scene datasets using both optimization-based and deep learning-based reconstruction approaches. Under full spatial sampling, MESH achieves high-fidelity and spectrally consistent recovery. Under compressive sampling, the proposed DL-MESH network maintains robustness: even at a total compression ratio β of 2.1%, it reaches a PSNR of 30.96 dB, SSIM of 0.8526, and SAM of 0.0742 rad under the measurement SNR of 40 dB. Across a range of β values and noise levels, DL-MESH consistently delivers perceptually coherent and spectrally accurate reconstructions.

The proposed metasurface design is compatible with current nanofabrication techniques, and the optical configuration can be readily integrated into standard single-pixel imaging setups. Numerical simulations verify its high spectral resolution and reconstruction performance, while the computational framework demonstrates scalability for real-time implementation. Taken together, these considerations confirm that the proposed MESH system is practically realizable with current fabrication and computational resources. The present study provides a simulation-based proof-of-concept to verify feasibility, and future work will focus on fabrication and experimental validation. The proposed framework also holds strong potential for deployment in portable remote sensing, biomedical diagnostics, and in-field spectral analysis.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The work was supported by the Research Grants Council of Hong Kong (GRF 17201822 and RIF R7003-21).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

compressive sensing, hyperspectral imaging, metasurface filter, single-pixel detection

Received: June 18, 2025
Revised: October 21, 2025
Published online:

[1] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*, Vol. *1*, Springer Science & Business Media, Boston, MA, USA **2003**.
[2] Y. Zhu, Y. Li, J. Huang, Y. Zhang, Y.-W. Ho, J. K.-H. Fang, E. Y. Lam, *Adv. Photonics Res.* **2024**, *5*, 2400038.
[3] T. Adão, J. Hruška, L. Pádua, J. Bessa, E. Peres, R. Morais, J. J. Sousa, *Remote Sens.* **2017**, *9*, 1110.
[4] X. Hadoux, F. Hui, J. K. Lim, C. L. Masters, A. Pébay, S. Chevalier, J. Ha, S. Loi, C. J. Fowler, C. Rowe, V. L. Villemagne, *Nat. Commun.* **2019**, *10*, 4227.
[5] J. Chen, Y. Li, J. Huang, E. Y. Lam, *Imaging Systems and Applications*, Optica Publishing Group, Washington, DC **2024**, p. IM1G–6.
[6] M. B. Stuart, A. J. McGonigle, J. R. Willmott, *Sensors* **2019**, *19*, 3071.
[7] S. Mahesh, D. Jayas, J. Paliwal, N. White, *J. Stored Products Res.* **2015**, *61*, 17.
[8] P. Guan, E. Y. Lam, *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1.
[9] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis, M. R. Olah, *Remote Sens. Environ.* **1998**, *65*, 227.
[10] C.-C. Hsu, C.-H. Lin, C.-H. Kao, Y.-C. Lin, *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 7773.
[11] Y. Li, J. Huang, J. Chen, E. Y. Lam, *Applied Industrial Spectroscopy*, Optica Publishing Group, Washington, DC **2024**, p. AW3A–2.
[12] H. Meng, Y. Gao, X. Wang, X. Li, L. Wang, X. Zhao, B. Sun, *Light: Sci. Appl.* **2024**, *13*, 121.
[13] Y. Xu, L. Lu, V. Saragadam, K. F. Kelly, *Nat. Commun.* **2024**, *15*, 1456.
[14] S. Jin, W. Hui, Y. Wang, K. Huang, Q. Shi, C. Ying, D. Liu, Q. Ye, W. Zhou, J. Tian, *Sci. Rep.* **2017**, *7*, 45209.
[15] C. Tao, H. Zhu, X. Wang, S. Zheng, Q. Xie, C. Wang, R. Wu, Z. Zheng, *Opt. Express* **2021**, *29*, 11207.
[16] Q. Yi, L. Z. Heng, L. Liang, Z. Guangcan, C. F. Siong, Z. Guangya, *Opt. Express* **2020**, *28*, 16126.
[17] M. Song, Z. Yang, P. Li, Z. Zhao, Y. Liu, Y. Yu, L.-A. Wu, *Appl. Opt.* **2023**, *62*, 2610.
[18] Y. Wang, J. Suo, J. Fan, Q. Dai, *IEEE Photonics Technol. Lett.* **2015**, *28*, 288.
[19] M. Liu, P. Lin, P. Huo, H. Qi, R. Jin, H. Zhang, Y. Ren, M. Song, Y.-Q. Lu, T. Xu, *Nat. Commun.* **2025**, *16*, 3994.
[20] H.-H. Hsiao, C. H. Chu, D. P. Tsai, *Small Methods* **2017**, *1*, 1600064.
[21] M. Liu, W. Zhao, Y. Wang, P. Huo, H. Zhang, Y.-Q. Lu, T. Xu, *Nano Lett.* **2024**, *24*, 7609.
[22] H. He, Y. Zhang, Y. Shao, Y. Zhang, G. Geng, J. Li, X. Li, Y. Wang, L. Bian, J. Zhang, L. Huang, *Adv. Mater* **2024**, *36*, 2313357.
[23] L. Zhang, C. Zhou, B. Liu, Y. Ding, H.-J. Ahn, S. Chang, Y. Duan, M. T. Rahman, T. Xia, X. Chen, Z. Liu, *Sci. Adv.* **2024**, *10*, eadp5192.
[24] J. Xiong, X. Cai, K. Cui, Y. Huang, J. Yang, H. Zhu, W. Li, B. Hong, S. Rao, Z. Zheng, S. Xu, *Optica* **2022**, *9*, 461.
[25] M. Makarenko, A. Burguete-Lopez, Q. Wang, F. Getman, S. Giancola, B. Ghanem, A. Fratalocchi, in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, IEEE, Piscataway, **2022**, pp. 12692–12702.
[26] C. Tao, H. Zhu, Y. Zhang, S. Luo, Q. Ling, B. Zhang, Z. Yu, X. Tao, D. Chen, Q. Li, Z. Zheng, *Opt. Express* **2022**, *30*, 33697.
[27] W.-K. Yu, *Sensors* **2019**, *19*, 4122.
[28] Y. Zhang, Y. Li, Z. Liu, F. Wang, G. Situ, M. K. Chen, H. Wang, Z. Geng, *Advanced Photonics Nexus* **2025**, *4*, 036005.
[29] G. Beneti Martins, L. Mahieu-, Williame, T. Baudier, N. Ducros, *Opt. Express* **2023**, *31*, 15599.
[30] B. Arad, R. Timofte, R. Yahel, N. Morag, A. Bernat, Y. Cai, J. Lin, Z. Lin, H. Wang, Y. Zhang, H. Pfister, in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops*, IEEE, Piscataway, **2022**, pp. 863–881.
[31] J. Li, C. Wu, R. Song, Y. Li, F. Liu, in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops,*, IEEE, Piscataway, **2020**, pp. 462–463.
[32] Y. Cai, J. Lin, Z. Lin, H. Wang, Y. Zhang, H. Pfister, R. Timofte, L. Van Gool, in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops*, IEEE, Piscataway, **2022**, pp. 745–755.

