



# End-to-end joint optimization of metasurface and image processing for compact snapshot hyperspectral imaging

Qiangbo Zhang<sup>a</sup>, Zeqing Yu<sup>a</sup>, Xinyu Liu<sup>a</sup>, Chang Wang<sup>a,b</sup>, Zhenrong Zheng<sup>a,b,\*</sup>

<sup>a</sup> State Key Laboratory of Modern Optical Instrumentation, College of Optical Science and Engineering, Zhejiang University, Hangzhou 310027, China

<sup>b</sup> Intelligent Optics & Photonics Research Center, Jiaxing Research Institute Zhejiang University, Jiaxing 314000, China

## ARTICLE INFO

### Keywords:

Hyperspectral imaging  
End-to-end joint optimization  
Metasurface  
Neural network  
Computational optics

## ABSTRACT

Traditional snapshot hyperspectral imaging systems generally require multiple refractive-optics-based elements to modulate light, resulting in bulky framework. In pursuit of a more compact form factor, a metasurface-based snapshot hyperspectral imaging system, which achieves joint optimization of metasurface and image processing, is proposed in this paper. The unprecedented light manipulation capabilities of metasurfaces are used in conjunction with neural networks to encode and decode light fields for better hyperspectral imaging. Specifically, the extremely strong dispersion of metasurfaces is exploited to distinguish spectral information, and a neural network based on spectral priors is applied for hyperspectral image reconstruction. By constructing a fully differentiable model of metasurface-based hyperspectral imaging, the front-end metasurface phase distribution and the back-end recovery network parameters can be jointly optimized. This method achieves high-quality hyperspectral reconstruction results numerically, outperforming separation optimization methods. The proposed system holds great potential for miniaturization and portability of hyperspectral imaging systems.

## 1. Introduction

Hyperspectral imaging captures the spectral distribution of a scene as a data cube that describes the spectral intensity of each wavelength at each pixel location. At present, hyperspectral imaging technology has been widely applied in many fields, such as agriculture surveillance [1], food safety [2], face recognition [3], military reconnaissance [4], etc.

For acquiring the huge spectral data cube, traditional spectral scanning systems require long exposure time which limits their use in real-time applications [5,6], thus a number of snapshot spectral imaging systems consisting of cascaded refractive-optics-based devices have been proposed for faster imaging. However, the improved functionality comes at the expense of bulky volume, impeding their portability [7,8]. To solve this problem, diffractive optical elements (DOEs) are introduced to snapshot hyperspectral imaging systems, facilitating the system compactness and portability [9–11]. Especially, Jeon et al. achieved high-quality spectral image reconstruction by applying a single DOE in front of the bare sensor [9]. Yet, the defect of only modulating the phase of light limits the further development of DOE-based hyperspectral imaging systems.

Metasurfaces, composed of two-dimensional arrays of subwavelength optical scatterers, are regarded as powerful substitutes to conventional diffractive and refractive optics [12–18]. Compared with DOEs, metasurfaces inherently has a higher space-bandwidth product

due to its subwavelength pitch [19]. In addition, metasurfaces with more powerful wavefront manipulation capabilities can steer the phase, amplitude, and polarization of light, while DOEs merely control the phase of light [17,18]. Recently, the combination of metasurfaces and computational imaging has yielded remarkable results [19–25], such as full-color achromatic imaging [19], three-dimensional imaging [23], etc. However, the end-to-end snapshot hyperspectral imaging system based on joint optimization of the metasurface and image processing has not been well investigated.

In this paper, a compact snapshot hyperspectral imaging system implemented by joint optimization of metasurface and image processing is proposed. It takes advantage of the extremely strong dispersion of metasurfaces to construct the point spread function (PSF) that varies greatly with wavelengths and provides extra freedom for hyperspectral imaging. Simultaneously, an end-to-end neural network based on joint optimization of metasurface phase distribution and image decoding is developed for hyperspectral imaging reconstruction. Fig. 1 shows the overview of this hyperspectral image reconstruction system. The metasurface-based image formation is modeled fully differentiable by consisting of the following differentiable steps: metasurface phase design, PSF calculation, convolution operation, sensor capture, and noise addition. Hyperspectral images can be reconstructed by the prior-based unfolding recovery network and then compared with the ground true

\* Corresponding author at: State Key Laboratory of Modern Optical Instrumentation, College of Optical Science and Engineering, Zhejiang University, Hangzhou 310027, China.

E-mail addresses: [changwang\\_optics@zju.edu.cn](mailto:changwang_optics@zju.edu.cn) (C. Wang), [zzr@zju.edu.cn](mailto:zzr@zju.edu.cn) (Z. Zheng).

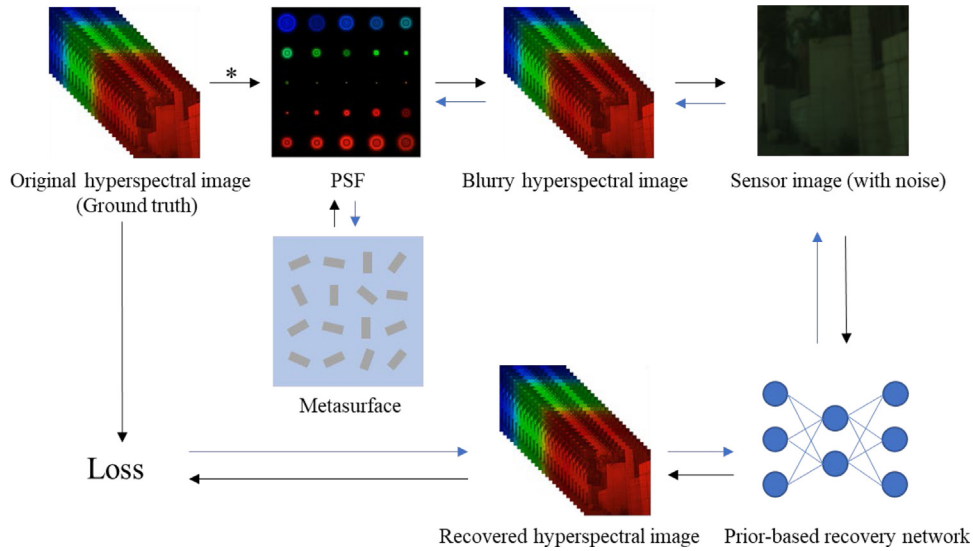


Fig. 1. Overview of the proposed hyperspectral image reconstruction system. Black arrows and blue arrows indicate forward propagation and back propagation, respectively.

to calculate the loss function for backward propagation. During backward propagation, the recovery network parameters and metasurface phase distribution as optimizable variables are updated. Compared with manually setting the phase distribution, our proposed hyperspectral imaging system can achieve better reconstruction results through the co-optimization of front-end optics and back-end recovery networks. In summary, this proposed hyperspectral imaging system realizes high image recovery performance with a compact form factor, further promoting the miniaturization and portability of hyperspectral imaging systems.

Specifically, the contributions of the proposed system are summarized as follow:

1. We introduce the geometric metasurface into the compact snapshot hyperspectral imaging system, exploiting its strong dispersion to distinguish spectral information.
2. We achieve joint optimization of hardware and recovery algorithm by differentiating the phase distribution of the geometric metasurface, enabling it to be co-optimized with the hyperspectral reconstruction network for optimal performance.
3. A back-end prior-based unfolding recovery network is employed as the hyperspectral reconstruction network. It is developed by joint optimization with the front-end metasurface, achieving better hyperspectral reconstruction results than those of separation optimization methods.

## 2. Method

### 2.1. Metasurface design

Phase profile is a crucial part of metasurfaces design. For geometric metasurfaces, phase modulation can be ideally defined as  $J(x, y)|L\rangle = e^{-i2\theta}|R\rangle$ , where  $J(x, y)$  and  $\theta$  are respectively the Jones matrix and rotation angle of each unit cell,  $|L\rangle$  and  $|R\rangle$  denote left-circularly polarized (LCP) and right-circularly polarized (RCP) light. When the geometry of unit cell is constant and LCP light incidents on a geometric metasurface, the phase shift of the unit cell is only linearly dependent on the rotation angle of the nanofin and independent of the incident wavelengths [16,26]. Therefore, geometric metasurfaces can simplify phase profile design by maintaining phase distribution for broad bandwidth. In comparison, polarization-independent metasurfaces should be optimized via complicated differentiable models [19], which are

unrealistic to predict phase responses corresponding to tens of wavelengths for various structure parameters. Hence, geometric metasurface is selected in this work as it modulates the phase more conveniently for broad bandwidth.

The schematic diagram of the unit cell is depicted in Fig. 2(a), where silicon dioxide ( $\text{SiO}_2$ ) and silicon nitride ( $\text{Si}_3\text{N}_4$ ) are chosen as the substrate and the nanofin, respectively. Full-wave finite-difference time-domain (FDTD) simulations are employed to optimize silicon nitride nanofins of different structural parameters for obtaining peak polarization conversion efficiency (PCE) from LCP to RCP in the 460–700 nm band, which is shown in Fig. 2(b). In the FDTD simulations, periodic boundary condition is applied at x-direction as well as y-direction, and perfect matching layer (PML) boundary condition is applied at z-direction. The optimal unit cell is selected due to its high PCE with smooth fluctuations for the entire spectrum. Finally, the chosen unit cell parameters are  $L = 315$  nm,  $W = 115$  nm,  $H = 750$  nm and  $C = 400$  nm. Geometric phase modulation makes phase shift of the output circularly polarized light only linearly dependent on the rotation angle of the nanofin as shown in Fig. 2(c), and phase modulation is identical and basically twice of the rotational angle of the nanofin for different wavelengths. The metasurface phase profile  $\phi$  with respect to current polar coordinate  $r$  is designed in the form of

$$\phi(r) = \frac{2\pi}{\lambda} \left( f - \sqrt{r^2 + f^2} \right) + \sum_{i=0}^n a_i \left( \frac{r}{R} \right)^{2i}, \quad (1)$$

where the first term is defined as the hyperboloidal phase  $\phi_1(r)$ , other terms denote the polynomial phase  $\phi_2(r)$ ,  $\lambda$  means the default wavelength,  $f$  is the focal length,  $R$  is the radius of the metasurface,  $\{a_0, \dots, a_n\}$  indicate optimizable coefficients and  $n$  is the number of polynomial terms. Traditional metalens using hyperboloidal phase indeed generates perfect spherical wavefront [27], but its coefficients are fixed and cannot be optimized. As for polynomial phase, its polynomial coefficients can be jointly optimized with the back-end neural network [19], realizing an overall consideration of both the hyperspectral imaging part as well as the image recovery part. This work proposes a strategy of combining the advantages of both phases, as the optimizable polynomial phase is added to the fixed hyperboloidal phase, enabling joint optimization of polynomial coefficients  $\{a_0, \dots, a_n\}$  and parameters of the back-end neural network for better recovery of spectral images than results of only using hyperboloidal phase.

### 2.2. PSF calculate

In this section, the PSF of the metasurface is calculated by introducing the model of diffraction imaging and the schematic diagram of the

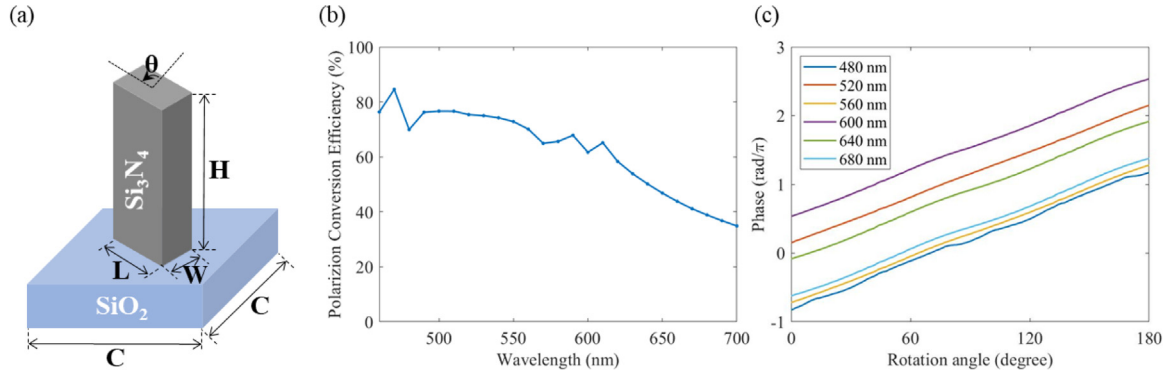


Fig. 2. (a) Schematic of the unit cell containing a silicon nitride nanofin on a silica substrate. (b) Polarization conversion efficiency from LCP to RCP of the unit cell in the 460~700 nm band. (c) The relationship of phase shift and the rotation angle of the chosen nanofin at different wavelengths.

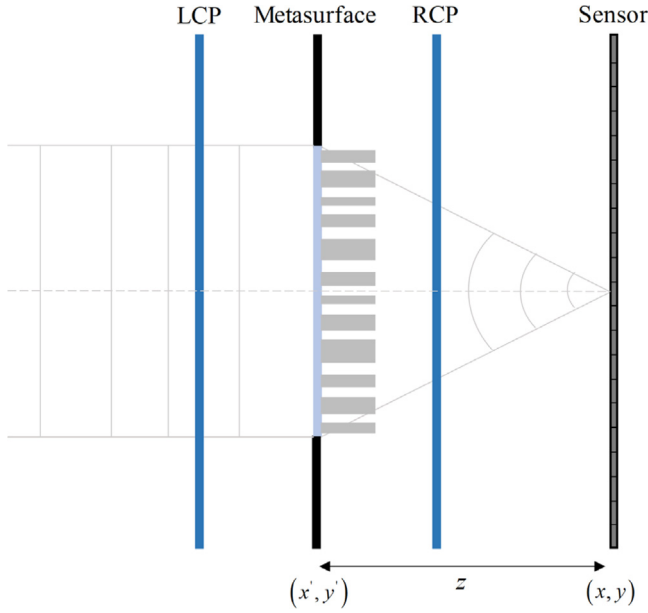


Fig. 3. Schematic diagram of diffraction imaging with the metasurface.

diffraction imaging is demonstrated in Fig. 3. The incident light passes through the left-circularly polarizer, the metasurface, and the right-circularly polarizer successively to be received by the sensor. Suppose the light source is located at infinity, the parallel monochromatic light of wavelength  $\lambda$  with normalized amplitude passes through the left-handed circular polarizer and then vertically hits on the metasurface. The wave field  $u_{\text{meta}}$  at position  $(x', y')$  along the metasurface after passing through the metasurface can be formulated as

$$u_{\text{meta}}(x', y') = \sqrt{\frac{1}{2} \text{PCE}_{\lambda}} \exp[i(\phi_{\text{meta}}(x', y'))], \quad (2)$$

where  $\text{PCE}_{\lambda}$  is polarization conversion efficiency of the metasurface for different wavelengths, and  $1/2$  indicates the energy loss caused by the polarizer. The sensor is located at the distance of  $z$  behind the metasurface. When  $z \gg \lambda$ , the wave field  $u_{\text{sensor}}(x, y)$  can be calculated approximated by the Fresnel diffraction as

$$u_{\text{sensor}}(x, y) = \frac{1}{i\lambda z} \exp(ikz) \iint u_{\text{meta}}(x', y') \times \exp\left\{\frac{ik}{2z} [(x-x')^2 + (y-y')^2]\right\} dx' dy', \quad (3)$$

where  $k = 2\pi/\lambda$  is the wavenumber. Finally, the PSF becomes the distribution of light intensity, that is, the square value of the wave field

$u_{\text{sensor}}$ . The PSF  $P_{\lambda}(x, y)$  can be derived by Fourier transform as:

$$P_{\lambda}(x, y) \propto \left| \mathcal{F} \left\{ \sqrt{\frac{1}{2} \text{PCE}_{\lambda}} \exp\left\{i\left[\phi_{\text{meta}}(x', y') + \frac{\pi}{\lambda z} (x'^2 + y'^2)\right]\right\} \right\} \right|^2. \quad (4)$$

### 2.3. Sensor image formation

Since the hyperspectral images of  $c$  channels are captured by a traditional RGB image sensor, it is essential to apply the spectral response of the sensor to compress the  $c$  spectral channels into three RGB channels. Meanwhile, considering the noise  $n$  during imaging, the final RGB image  $J_{\text{RGB}}$  received from the sensor can be expressed as

$$J_{\text{RGB}} = \int (I_{\lambda} * P_{\lambda}) \cdot S_{\lambda, \text{rgb}} d\lambda + n, \quad (5)$$

where  $I_{\lambda}$  denotes the original hyperspectral image,  $S_{\lambda, \text{rgb}}$  is the spectral sensitivity function of the sensor and  $*$  is defined as the convolution operator. Further, Eq. (5) can be rewritten in the form of vector and matrix:

$$J = \Phi I + N, \quad (6)$$

where  $J \in \mathbb{R}^{HW \times 3}$  and  $I \in \mathbb{R}^{HW \times c}$  are respectively defined as the captured image vector and original hyperspectral image vector,  $H$  and  $W$  are the image size.  $\Phi = SP$ , where  $S \in \mathbb{R}^{HW \times 3 \times HW \times c}$  and  $P \in \mathbb{R}^{HW \times c \times HW \times c}$  denote the sensor spectral sensitivity matrix and PSF matrix, respectively.  $N \in \mathbb{R}^{HW \times 3 \times 1}$  represents the noise vector.

### 2.4. Optimization problem

Considering the fact that the process of hyperspectral image reconstruction is from blurred to clear, from noisy to denoised, and from RGB three-channel to hyperspectral multi-channel, hyperspectral reconstruction requires a prior item since it is a very underdetermined problem. The unfolding neural network based on image prior has proved to be the very helpful non-blind convolution solver for such inverse image restoration problems [9,28–30], but has not yet been jointly optimized with the front-end optics. Therefore, we also employ the prior-based unfolding neural network as the back-end recovery algorithm jointly optimized with the metasurface. The hyperspectral reconstruction objective function including the spectral prior can be expressed as

$$I = \arg \min_I \frac{1}{2} \|J - \Phi I\|_2^2 + V(I), \quad (7)$$

where the first item is defined as the spectral data item, and the second item indicates the spectral prior item, with  $V(\cdot)$  denoting the prior function of hyperspectral images. Using the half-quadratic splitting method,

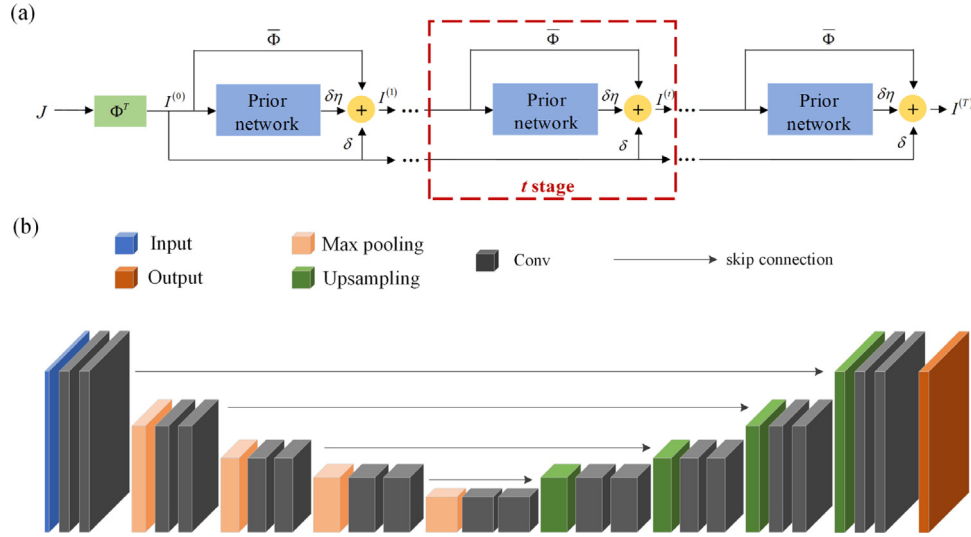


Fig. 4. The network architecture of our method for hyperspectral image restoration. (a) The overall architecture of the proposed deep neural network. (b) The architecture of the hyperspectral image prior network.

the data term and prior term of Eq. (7) can be decoupled and transformed into an unconstrained optimization problem by introducing auxiliary variables  $H \in \mathbb{R}^{HW \times 3 \times 1}$ :

$$(I, H) = \arg \min_{I, H} \frac{1}{2} \|J - \Phi I\|_2^2 + \eta \|I - H\|_2^2 + V(H) \quad s.t. \quad H = I, \quad (8)$$

where  $\eta$  is the penalty parameter. Further, Eq. (8) can be solved by alternately solving the two subproblems:

$$I^{(t+1)} = \arg \min_I \|J - \Phi I\|_2^2 + \eta \|I - H^{(t)}\|_2^2, \quad (9)$$

and

$$H^{(t+1)} = \arg \min_H \|I^{(t+1)} - H\|_2^2 + V(H). \quad (10)$$

An inexact solution of  $I^{(t+1)}$  can be calculated by alternating gradient descent:

$$\begin{aligned} I^{(t+1)} &= I^{(t)} - \delta [\Phi^T (\Phi I^{(t)} - J) + \eta (I^{(t)} - H^{(t)})] \\ &= \bar{\Phi} I^{(t)} + \delta I^{(0)} + \delta \eta H^{(t)}, \end{aligned} \quad (11)$$

where  $\bar{\Phi} = [(1 - \delta\eta)I - \delta\Phi^T\Phi]$ ,  $I^{(0)} = \Phi^T J$ , and  $\delta$  is the parameter controlling the gradient descent step size. Now,  $H^{(t)} = R(I^{(t)})$  can be firstly solved through a neural network of spectral prior, where  $R(\cdot)$  is the nonlinear function from  $I^{(t)}$  to  $H^{(t)}$ . Then  $I^{(t+1)}$  can be solved by Eq. (11). After iteratively solving  $H^{(t)}$  and  $I^{(t)}$ , the optimum spectral image  $I^{(T)}$  can eventually be recovered.

## 2.5. Network architecture

The network framework shown in Fig. 4(a) is utilized to iteratively calculate  $I^{(T)}$ , with the RGB image  $J$  received by the sensor as the only input, and the output of the first iteration  $I^{(1)}$  is obtained according to the Eq. (11). By repeating the same iteration step continuously, finally the output  $I^{(T)}$  is generated. In the network framework, prior network is the nonlinear function from  $I^{(t)}$  to  $H^{(t)}$ . The U-net structure [31] has been widely employed to hyperspectral image restoration [9], image denoising [32] and depth estimation [23] with excellent performance. Compared with other networks, U-net could achieve good recovery results with a simple structure. Hence, U-net network is utilized to calculate the spectral prior. As shown in Fig. 4(b), the structure of the prior network consists of a contracting path and a symmetric expansive path. The input  $I^{(t)}$  is first extracted to 64 feature channels by two convolutional layers, entering to the contracting path where each block contains a max-pooling layer for downsampling and two convolutional

layers for extracting features. Next in the extensive path, the upsampling block and the corresponding block in the contracting path are concatenated together as the input of two convolutional layers. Finally, a convolutional layer is performed on the output of the extensive path to obtain the output  $H^{(t)}$  of the prior network. In this paper, the number of iterations  $T$  is set to be 6. Meanwhile, in order to increase the flexibility of the network,  $\delta$  and  $\eta$  are set as optimizable variables and are different at each stage [9,28].

The loss function is defined as the sum of the mean squared errors between each intermediate result and the ground truth:

$$Loss = \sum_{t=1}^T \|I^{(t)} - I\|_2^2. \quad (12)$$

This recursive supervised approach is able to accelerate training and the learned optimizer steadily pushes each intermediate result to be closer to the ground truth [30].

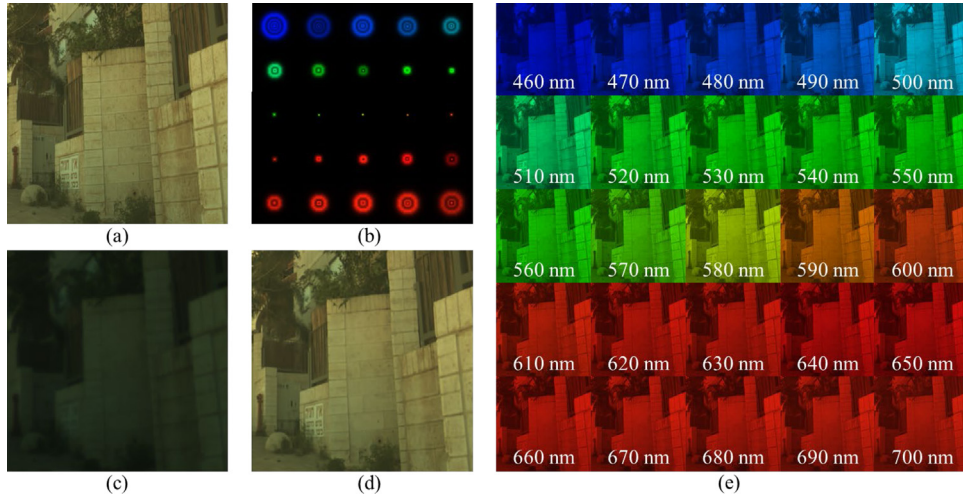
## 3. Results and discussions

### 3.1. Implementation details

In this paper, the joint optimization of front-end metasurface and back-end recovery network is utilized to restore hyperspectral images, with the polynomial phase coefficients and recovery network parameters being the optimizable variables. Specifically, metasurface phase distribution is firstly set to be the hyperboloidal phase  $\phi_1$  and then only the back-end recovery network parameters are optimized. After the loss function converges, the polynomial phase  $\phi_2$  is added and jointly optimized with the parameters of the recovery network. In addition, the alternate update method is employed to optimize the phase coefficients and recovery network parameters for 10 and 30 iterations, respectively [19].

During training, the publicly available ICVL dataset [33] (150 training and 22 testing images) is utilized. The training hyperspectral images include 25 wavelength channels in a range from 460 nm to 700 nm are randomly cropped to  $128 \times 128$  as the input of the entire hyperspectral imaging system. For metasurface design, the focal length is designed as  $f = 15$  mm, the diameter is 1 mm, the default wavelength is set as 580 nm, the number of polynomial terms  $n = 8$ . The spectral response of the Nikon D700 camera is applied to compress hyperspectral channels into RGB channels, and the pixel size of the sensor is set to be 5  $\mu\text{m}$ . The distance between the metasurface and the sensor is set to be 15 mm. In the sensor image formation model, gaussian noise is added with





**Fig. 5.** (a) The original hyperspectral image. (b) Simulated PSFs of 25 channels in the bandwidth from 460 nm to 700 nm. The wavelength channel sequence is the same as (e). (c) The color image received by the sensor. (d) The recovered hyperspectral image. (e) Recovered images for 25 spectral channels from 460 nm to 700 nm.

standard deviation of  $\sigma = 0.005$ . The coefficients of polynomial phase and network parameters are optimized by ADAM optimizer ( $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ). The learning rates are all set to be  $10^{-4}$ , and the batch size is set as 4. The equipment used includes an Intel Core i9-10900K CPU with 64 GB memory and an NVIDIA 3090 GPU with 24 GB memory. It takes 1.28 s to restore a hyperspectral image with  $800 \times 800$  resolution applying the proposed recovery network.

### 3.2. Results

Fig. 5 shows images and PSFs obtained at different stages during spectral reconstruction with the proposed method. The entire 3D hyperspectral data are synthesized into an RGB image for ease of display in this paper. Fig. 5(a) illustrates the original hyperspectral image. The PSFs of 25 channels after optimization in the bandwidth from 460 nm to 700 nm are demonstrated in Fig. 5(b). The PSF is smallest when the wavelength is around 580 nm and expands with the wavelength varying. Notably, PSFs impart additional spectral information to hyperspectral imaging by virtue of their wavelength-dependent feature. Fig. 5(c) exhibits the image received by the sensor, and it is dim due to the polarization-dependent energy loss. Meanwhile, large PSF kernels and noise lead to the blurry image and low signal-to-noise ratio (SNR). Fig. 5(d and e) show respectively the restored hyperspectral image and the reconstructed images for 25 spectral channels from 460 nm to 700 nm.

To further evaluate the proposed system, we retrain Jeon's rotative PSF method [9] (designing phase to make PSF rotate with wavelength varying, and we apply this to the metasurface) with our recovery network for comparison. Moreover, the phase design methods of only polynomial or hyperboloid are also added to the comparison with same recovery network. Specially, the phase of Jeon's and Hyperboloid methods are pre-configured and only the parameters of the recovery network are optimized, while the phase of Polynomial and Our methods are jointly optimized with the back-end recovery network.

Three quantitative image quality metrics are employed to evaluate the performance of these methods: average peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [34], and spectral angle mapping (SAM) [35]. The performance of different methods is listed in Table 1. It can be seen that our method demonstrates the highest PSNR, and is nearly equally well in SSIM and SAM with the Hyperboloid method.

To show more intuitive results, two representative sets of restored images are visualized as shown in Fig. 6. The results recovered by the proposed method not only achieve the best results on image quality metrics, but also have sharper edges in visual perception, outperforming those of other methods.

**Table 1**

Average PSNR, SSIM, and SAM comparison of four spectral reconstruction methods with the same test dataset. Bold text indicates the highest accuracy.

Method	Jeon's	Polynomial	Hyperboloid	Ours
PSNR(dB)	33.14	34.53	36.06	<b>37.24</b>
SSIM	0.962	0.974	0.980	<b>0.981</b>
SAM	0.048	0.058	<b>0.039</b>	0.040

Additionally, Fig. 7 shows the recovered spectrum by different methods for two points in the selected image. Comprehensively, the reconstructed spectra using the proposed method is closest to the value of ground truth, followed by results of the Hyperboloid method. The basically close performance of these two methods is also consistent with the SAM data given in Table 1.

### 3.3. Discussions

A compact snapshot hyperspectral imaging system jointly optimized by metasurface phase and recovery network is here proposed. Since the optimizable polynomial phase is added to the basis of the general hyperboloidal phase, the phase distribution of the metasurface holds greater flexibility. Furthermore, the phase-adaptive learning to optimize polynomial coefficients is more efficient than the hand-designed phase, and the former can achieve better recovery quality.

However, there are still some limitations in this study. When designing the metasurface, it is defaulted that the rotation has no effect on the spectral response. But in fact, the spectral response of the unit cell changes slightly with the unit cell rotating. However, considering that there will also be errors in the actual manufacturing process, an experimental calibration of PSF is ultimately required. After the PSF experimental calibration, the parameters of the recovery network need to be fine-tuned, after which high-quality spectral reconstruction results can be obtained.

During the PSF calculation, it is assumed that the light source is located at infinity. When the light source is at a finite distance in front of the metasurface, the PSF is not only a function of the spectrum, but also a function of distance and it has been explored by Baek et al. [36]. In this paper, the PSF calculation of limited distance is not considered, making the proposed spectral reconstruction method suitable for shooting distant objects.

Although metasurfaces with an unparalleled ability can control the phase, amplitude, and polarization of light synchronously, this paper only exploits the phase modulation capabilities. Incorporating amplitude and polarization modulation into computational imaging should be put into effect in the future.

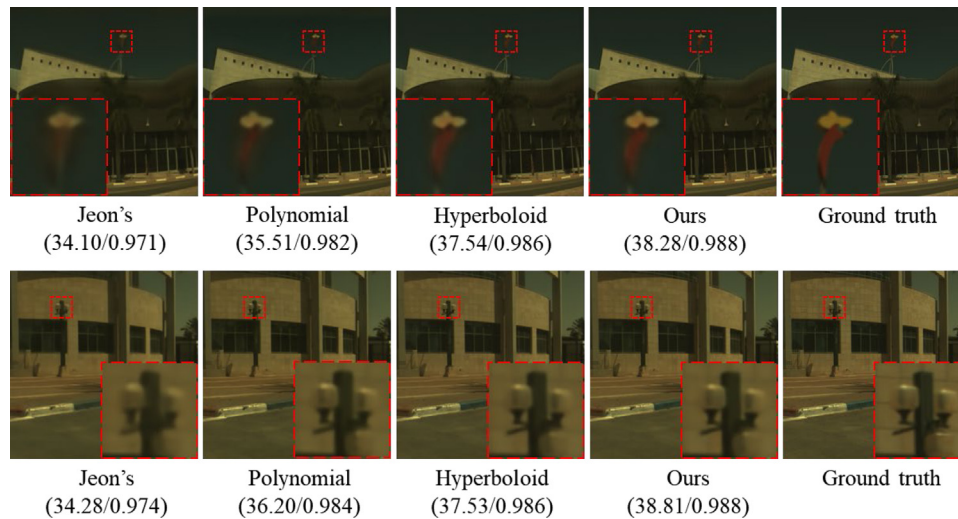


Fig. 6. Visual quality comparison of five methods. The PSNR and SSIM for the recovered results are shown in the parenthesis.

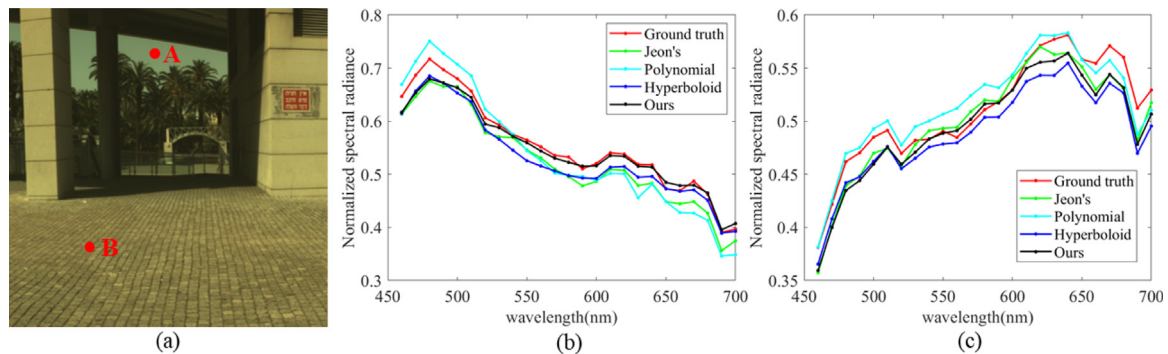


Fig. 7. (a) A hyperspectral image with two randomly selected spatial points A and B. (b and c) Respectively the reconstruction spectra of points A and B using different methods. The reconstruction spectra of this proposed method (the black line) is closest to the ground truth (red line).

#### 4. Conclusion

In this paper, we propose a compact and portable snapshot hyperspectral imaging system, which utilizes a metasurface and achieves joint optimization of hardware and recovery algorithm. Above all, the designed metasurface phase profile, consisting of fixed hyperboloidal phase and optimizable polynomial phase, is jointly optimized with the recovery network to restore the high-quality hyperspectral images, which outperform those of the separation optimization, and there may be opportunities of further improvement by increasing the manipulation dimension of the metasurface. The use of metasurface realized miniaturization of the hyperspectral imaging system configuration, and besides that, our proposed metasurface-based hyperspectral imaging system is easier to assemble and calibrate due to its simple optical components, compared with the complex refractive-optics-based hyperspectral system. Moreover, the fabrication of metasurfaces is compatible with the complementary metal oxide-semiconductor (CMOS) fabrication process [37]. Hence, the proposed system has the potential to be mass-produced with low cost for industrial applications. In addition, hyperspectral imaging systems using metasurfaces can realize multi-functionalities, such as exploiting polarization information from the input hyperspectral light. In summary, our work taking advantage of the thin but powerful light manipulation ability of metasurfaces is expected to promote practical applications of hyperspectral imaging system by further realizing the miniaturization and portability while maintaining high performance.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Data availability

Data will be made available on request.

#### Acknowledgment

This work was supported by the Key Research and Development Plan of Zhejiang Province, China a under Grant No. 2022C01127.

#### References

- [1] R.N. Sahoo, S.S. Ray, K.R. Manjunath, Hyperspectral remote sensing of agriculture, *Current Sci.* 108 (2015) 13.
- [2] J. Qin, M. Kim, K. Chao, D. Chan, S. Delwiche, B.-K. Cho, Line-scan hyperspectral imaging techniques for food safety and quality applications, *Appl. Sci.* 7 (2017) 125.
- [3] V. Sharma, L. Van Gool, Image-level classification in hyperspectral images using feature descriptors, with application to face recognition, 2016, <http://dx.doi.org/10.48550/ARXIV.1605.03428>.
- [4] I. Makki, R. Younes, C. Francis, T. Bianchi, M. Zucchetti, A survey of landmine detection using hyperspectral imaging, *ISPRS J. Photogramm. Remote Sens.* 124 (2017) 40–53.

- [5] 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, O. Williams, Imaging spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), *Remote Sens. Environ.* 65 (1998) 227–248.
- [6] R.G. Sellar, G.D. Boreman, Comparison of relative signal-to-noise ratios of different classes of imaging spectrometer, *Appl. Opt.* 44 (2005) 1614.
- [7] C.V. Correa, H. Arguello, G.R. Arce, Snapshot colored compressive spectral imager, *J. Opt. Soc. Amer. A* 32 (2015) 1754.
- [8] H. Rueda, H. Arguello, G.R. Arce, Compressive spectral testbed imaging system based on thin-film color-patterned filter arrays, *Appl. Opt.* 55 (2016) 9584.
- [9] D.S. Jeon, S.-H. Baek, S. Yi, Q. Fu, X. Dun, W. Heidrich, M.H. Kim, Compact snapshot hyperspectral imaging with diffracted rotation, *ACM Trans. Graph.* 38 (2019) 1–13.
- [10] H. Hu, H. Zhou, Z. Xu, Q. Li, H. Feng, Y. Chen, T. Jiang, W. Xu, Practical snapshot hyperspectral imaging with DOE, *Opt. Lasers Eng.* 156 (2022) 107098.
- [11] H. Arguello, S. Pinilla, Y. Peng, H. Ikoma, J. Bacca, G. Wetzstein, Shift-variant color-coded diffractive spectral imaging system, *Optica* 8 (2021) 1424–1434.
- [12] N. Yu, F. Capasso, Flat optics with designer metasurfaces, *Nature Mater.* 13 (2014) 139–150.
- [13] M. Khorasaninejad, W.T. Chen, R.C. Devlin, J. Oh, A.Y. Zhu, F. Capasso, Metalenses at visible wavelengths: Diffraction-limited focusing and subwavelength resolution imaging, *Science* 352 (2016) 1190–1194.
- [14] M. Khorasaninejad, F. Capasso, Metalenses: Versatile multifunctional photonic components, *Science* 358 (2017) eaam8100.
- [15] M. Khorasaninejad, Z. Shi, A.Y. Zhu, W.T. Chen, V. Sanjeev, A. Zaidi, F. Capasso, Achromatic metalens over 60 nm bandwidth in the visible and metalens with reverse chromatic dispersion, *Nano Lett.* 17 (2017) 1819–1824.
- [16] P. Genevet, F. Capasso, F. Aieta, M. Khorasaninejad, R. Devlin, Recent advances in planar optics: from plasmonic to dielectric metasurfaces, *Optica* 4 (2017) 139.
- [17] J. Engelberg, U. Levy, The advantages of metalenses over diffractive lenses, *Nature Commun.* 11 (2020) 1991.
- [18] W.T. Chen, F. Capasso, Will flat optics appear in everyday life anytime soon? *Appl. Phys. Lett.* 118 (2021).
- [19] E. Tseng, S. Colburn, J. Whitehead, L. Huang, S.-H. Baek, A. Majumdar, F. Heide, Neural nano-optics for high-quality thin lens imaging, *Nature Commun.* 12 (2021) 6493.
- [20] S. Colburn, A. Zhan, A. Majumdar, Metasurface optics for full-color computational imaging, *Sci. Adv.* 4 (2018) eaar2114.
- [21] Q. Guo, Z. Shi, Y.-W. Huang, E. Alexander, C.-W. Qiu, F. Capasso, T. Zickler, Compact single-shot metalens depth sensors inspired by eyes of jumping spiders, *Proc. Natl. Acad. Sci. USA* 116 (2019) 22959–22965.
- [22] J.N. Mait, R.A. Athale, J. van der Gracht, G.W. Euliss, Potential applications of metamaterials to computational imaging, in: *Frontiers in Optics/Laser Science*, OSA, Washington, DC, 2020, p. FTu8B.1.
- [23] S. Tan, F. Yang, V. Boominathan, A. Veeraraghavan, G.V. Naik, 3D imaging using extreme dispersion in optical metasurfaces, *ACS Photonics* 8 (2021) 1421–1429.
- [24] Q. Fan, W. Xu, X. Hu, W. Zhu, T. Yue, C. Zhang, F. Yan, L. Chen, H.J. Lezec, Y. Lu, A. Agrawal, T. Xu, Trilobite-inspired neural nanophotonic light-field camera with extreme depth-of-field, *Nature Commun.* 13 (2022) 2130.
- [25] X. Hua, Y. Wang, S. Wang, X. Zou, Y. Zhou, L. Li, F. Yan, X. Cao, S. Xiao, D.P. Tsai, J. Han, Z. Wang, S. Zhu, Ultra-compact snapshot spectral light-field imaging, *Nature Commun.* 13 (2022) 2732.
- [26] H.-H. Hsiao, C.H. Chu, D.P. Tsai, Fundamentals and applications of metasurfaces, *Small Methods* 1 (2017) 1600064.
- [27] F. Aieta, P. Genevet, M.A. Kats, N. Yu, R. Blanchard, Z. Gaburro, F. Capasso, Aberration-free ultrathin flat lenses and axicons at telecom wavelengths based on plasmonic metasurfaces, *Nano Lett.* 12 (2012) 4932–4936.
- [28] L. Wang, C. Sun, Y. Fu, M.H. Kim, H. Huang, Hyperspectral image reconstruction using a deep spatial-spectral prior, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, Long Beach, CA, USA, 2019, pp. 8024–8033.
- [29] W. Dong, P. Wang, W. Yin, G. Shi, F. Wu, X. Lu, Denoising prior driven deep neural network for image restoration, *IEEE Trans. Pattern Anal. Mach. Intell.* 41 (2019) 2305–2318.
- [30] D. Gong, Z. Zhang, Q. Shi, A. van den Hengel, C. Shen, Y. Zhang, Learning deep gradient descent optimization for image deconvolution, *IEEE Trans. Neural Netw. Learn. Syst.* 31 (2020) 5468–5482.
- [31] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W.M. Wells, A.F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention, MICCAI 2015*, Springer International Publishing, Cham, 2015, pp. 234–241.
- [32] X. Dun, H. Ikoma, G. Wetzstein, Z. Wang, X. Cheng, Y. Peng, Learned rotationally symmetric diffractive achromat for full-spectrum computational imaging, *Optica* 7 (2020) 913.
- [33] B. Arad, O. Ben-Shahar, Sparse recovery of hyperspectral signal from natural RGB images, in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), *Computer Vision – ECCV 2016*, Springer International Publishing, Cham, 2016, pp. 19–34.
- [34] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: From error visibility to structural similarity, *IEEE Trans. Image Process.* 13 (2004) 600–612.
- [35] F.A. Kruse, A.B. Lefkoff, J.W. Boardman, K.B. Heidebrecht, A.T. Shapiro, P.J. Barloon, A.F.H. Goetz, The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data, *Remote Sens. Environ.* 44 (1993) 145–163.
- [36] S.-H. Baek, H. Ikoma, D.S. Jeon, Y. Li, W. Heidrich, G. Wetzstein, M.H. Kim, Single-shot hyperspectral-depth imaging with learned diffractive optics, in: 2021 IEEE/CVF International Conference on Computer Vision, ICCV, IEEE, Montreal, QC, Canada, 2021, pp. 2631–2640.
- [37] M. Pan, Y. Fu, M. Zheng, H. Chen, Y. Zang, H. Duan, Q. Li, M. Qiu, Y. Hu, Dielectric metalens for miniaturized imaging systems: progress and challenges, *Light Sci. Appl.* 11 (2022) 195.