Spectro-polarimetric-depth Imaging by Inverse-Designed Single-Cell Metasurface

Dasen Zhang\textsuperscript{1}, Ting Ma\textsuperscript{1}, Yuqi Peng\textsuperscript{1}, Xianjin Liu\textsuperscript{1}, Qiwen Bao\textsuperscript{1}, and Jun-Jun Xiao\textsuperscript{1}

\textsuperscript{1}Affiliation not available

March 30, 2024
Spectro-polarimetric-depth Imaging by Inverse-Designed Single-Cell Metasurface

Dasen Zhang, Ting Ma, Yuqi Peng, Xianjin Liu, Qiwen Bao, and Jun-Jun Xiao

Abstract—Metasurface-assisted computational imaging has emerged as a promising route for complex optical field reconstruction. Multi-channel lensless imaging has been proved by leveraging the diverse ingredients of the optical field that can be deliberately manipulated by flat-optics. However, achieving single-shot multi-channel images by a single-cell metasurface is challenging due to the apparent capacity limitation of a single kind of meta-atom. In this work, we introduce an end-to-end inverse design method based on gradient-descent optimization algorithm for multi-channel lensless computational imaging, realizable by a single-cell metasurface. The proposed method directly provides single-cell metasurfaces without the need of complex meta-atom design strategy, which offers high efficiency and eases the fabrication. Particularly, multi-spectral, multi-depth (3D) and polarimetric imaging is demonstrated by a phase-only metasurface across the visible and near-infrared band. Our results prove a viable scheme towards practical metasurface generated lensless image reconstruction, by demonstrating the high-density information coding of single-cell metasurface. The scheme is expected to enable optical storage, optical encryption, holographic display, and full-color and three-dimensional imaging applications.

Index Terms—Metasurface, inverse design, computational imaging.

I. INTRODUCTION

METASURFACES that exhibit unique optical properties have ushered in a new era of meta-optics with extraordinary ultra-compact optical devices [1]. Metasurface enables spatially varying phase, amplitude and polarizations control on an incoming optical beam, for example, as for high-numerical-aperture focusing or imaging [2], beam steering over large angles [3], and computer-generated holography [4], [5], [6]. In particular, metasurface-based multi-channel computational imaging has attracted great attention due to the compact size and strong light manipulation ability of metasurfaces [7], [8], [9], [10]. So far, such metasurface-assisted multi-channel imaging has been actively reported for obtaining multi-polarization,[11] multi-color (spectral) [12], wide-field-of-view [13], and three-dimensional (3D) image [14]. Basically, these studies can be recognized in two distinct categories: lens and lensless imaging [15], [16]. In lens imaging system, metasurface has been designed as a color filter sitting right above an image sensor, more like the ordinary Bayer filter. For example, Rao et. al. combined a CMOS image sensor with a designed metasurface to achieve a snapshot hyperspectral image sensor for a practical anti-spoofing face recognition [17]. Xiong et. al. designed a real-time spectral imaging chip based on a reconfigurable metasurface supercell on a CMOS image sensor for brain hemodynamics imaging [18]. In sharp contrast, in the lensless imaging systems, a metasurface-based multi-channel imaging system that can recover different scenes from different channels by jointly optimizing the metasurface and the recovery algorithm [19]. Arya et. al. have presented a method for the end-to-end optimization of multi-wavelength and depth imaging system that reconstructs targets from different channels by the concept of compressed sensing [20]. These methods, however, need to build a surrogate model that maps the geometry of meta-atom to its complex transmission coefficient in a broadband, which is differentiable but requires complicated design process and suffers from variety of meta-atom geometry.

In this work, we employ an inverse design approach to produce a single-cell metasurface by combining a multi-channel imaging optimization method with the Pancharatnam-Berry (PB) phase (i.e., geometric phase) offered by rectangular meta-atoms. This represents a joint solution by combining optimized optical frontend by artificial nanophotonic structures and electronic backend algorithm. Given a certain geometry of rectangular meta-atom, the phase jump of the outgoing circularly polarized light is twice the rotation angle of the meta-atom, for incident orthogonal circularly polarized light. This simple phase manipulation capability holds for right-hand circular polarization (RCP) to left-hand circular polarization (LCP) channel, and vice versa. Such an approach has been exploited in meta-grating, metalens and metasurface-based light router design for years due to its outstanding ability for light manipulation, easy of the structure fabrication, and high efficiency [21], [22], [23]. Recently, So et al. applied the approach to encode the metasurface for multi-color and 3D images by exploiting the spatial dispersion of the transmitted light [24]. However, the multi-color, multi-depth and multi-polarization...
imaging based on the geometric-phase type metasurface is not reported yet. In this context, we proposed an inverse design scheme based on the geometric phase concept and designed single-cell metasurfaces for multi-channel computational imaging. Specifically, we applied the end-to-end optimization scheme (see Fig. 1) based on gradient optimization method to find a phase mask for multi-channel computational imaging, and cast the phase profile into a single-cell metasurface based on the PB phase method. We thus built a multi-wavelength, multi-depth and multi-polarization imaging system over the visible and near-infrared band. The final results are verified by finite-difference time-domain (FDTD) full-wave simulations. To find optimal cross-polarization conversion efficiency of the PB metasurface, we combined Bayesian optimization and the FDTD solver to design the meta-atom, rather than searching it in a database via brute force. Note that the Bayesian optimization approach yields meta-atom with high conversion efficiency of circularly cross-polarized light over all the interested optical channels for the computational imaging.

II. METHOD

A. Multi-channel imaging

In the scenario of multispectral polarimetric 3D imaging, a ground truth object can be regarded as a high-dimension tensor spanned by the three-dimensional space, the wavelength dimension and the polarization state. For convenience, we can consider the object $I$ as a set of 2D $(x, y)$ intensity image $I_{\lambda,z,\sigma}(x,y)$ indexed by different channels of wavelength ($\lambda$), depth ($z$) and polarization ($\sigma$). For simplicity and conciseness, we label it as $I(z,\lambda,\sigma)$ and omit the spatial coordinates $(x,y)$ for each depth slice. Figure 1 shows the flowchart of the whole optimization process. Notice that in the optical frontend the light fields coming from these multi-channel scenes $I(z,\lambda,\sigma)$ are modulated by the phase mask and then diffracted onto the sensor array, as shown in Fig. 1. Under the paraxial approximate regime, the response $v$ of the sensor can be written as the summation of the convolutions of $I(\lambda,z,\sigma)$ with the corresponding point spreading functions (PSFs) offered by the metasurface modulation and the free-space propagation

$$v = \sum_{\lambda,z,\sigma} PSF_{\lambda,z,\sigma} \otimes I(\lambda,z,\sigma) + \kappa,$$

where

$$\lambda \in [\lambda_1, \lambda_2, \cdots, \lambda_n], \quad z \in [z_1, z_2, \cdots, z_m], \quad \sigma \in [\sigma_1, \sigma_2, \cdots, \sigma_p]$$

(2)

denotes the considered spectral channel and the object distance, respectively, and $\kappa$ is a noise term modeled by zero-mean Gaussian white noise with standard deviation $\delta$. In the above imaging model, $\otimes$ denotes convolution operation. We assume pure phase modulation by the phase mask which offers pixel-wised phase modulation $\phi(x,y)$, and the distance between the mask and the sensor is
The computational images are then obtained via the following optimization problem

$$\min \left\{ v - \sum_{\lambda, z, \sigma} \text{PSF}_{\lambda, z, \sigma} \otimes \hat{I}(\lambda, z, \sigma) \right\} + \Gamma(\hat{I}(\lambda, z, \sigma)),$$

where the reconstructed object $\hat{I}(\lambda, z, \sigma)$ is the optimal solution. Here, a regularization term $\Gamma(\cdot)$ is added to make the inverse problem well-posed and well-conditioned as well as to impose any prior information such as sparsity or smoothness. A simplest choice (with minimal prior information) is the Tikhonov regularization where $\Gamma(\cdot) = \alpha \| \cdot \|^2$, leading to

$$\hat{I} = (\hat{P}^T \hat{P} + \alpha \hat{\mathbf{A}})^{-1} \hat{P}^T v,$$

where $\hat{I} = [\hat{I}(\lambda, z, \sigma)]$ collects all the reconstructed images at the multiple channels. Note that in Eq. (4) $\hat{P} = \sum_{\lambda, z, \sigma} \text{PSF}_{\lambda, z, \sigma} \otimes \hat{I}(\lambda, z, \sigma)$ represents the system’s Green function subjected to point source illumination, $\alpha$ is a constant that would be selected empirically, $\hat{P}^T$ denotes the transpose of $\hat{P}$, and $\hat{\mathbf{A}}$ is the identity matrix. Note further that in this model, all the PSFs are assumed to be depend on the wavelength $\lambda$, the object distance $z$ and the polarization $\sigma$, and the objects considered are under incoherent illumination. The PSFs are computed by the angular spectrum (AS) algorithm, given a phase mask $\phi(x, y)$ enabled by the metasurface. It is important to remark that the PSFs depend on the detailed structure of the metasurface and shall be optimized in the whole close loop. Once it is settled, the imaging would simply involve the computational decoder part. Equation (4) can be iteratively solved by the conjugate-gradient (CG) method to avoid computing the matrix inversion directly [19].

The loss function is defined as

$$L = \frac{1}{N} \sum_{\lambda, z, \sigma} \| I(\lambda, z, \sigma) - \hat{I}(\lambda, z, \sigma) \|^2,$$

where $N$ is the total number of the considered channel. Here we randomly generate the inputs $I(\lambda, z, \sigma)$ and initialize $\phi(x, y)$ to start the iteration cycle [see Fig. 1].

Considering the fact that the design of metasurface is based on the local potential approximation (LPA), mapping a rapidly changed $\phi(x, y)$ into metasurface structure would reduce its performance. To solve that problem, the designed $\phi(x, y)$ generally is quantized into $l$ discrete values with identical interval $2\pi/l$ before it is mapped to metasurface. However, this degrades performance from phase-based simulation to metasurface-based simulation, to some extent. Here, we introduce soft-quantization function $\Theta(\cdot)$ to gradually quantize the $\phi(x, y)$ into quantized $\phi_q(x, y)$ during the optimization process [25]:

$$\phi_q(x, y) = \Theta(\phi(x, y))$$

$$\phi_q(x, y) = 2\pi \sum_{i=0}^{l-1} \text{sig}(\tau(\phi(x, y) - \frac{\pi}{l} - 2\pi k)), \quad \text{sig}(\phi) = \frac{1}{1 + e^{-\phi}}, \quad \tau = 2 + 2\eta \eta,$$

where $\eta$ is the temperature of quantization, and $\eta$ increases with the iterations $b$. $\phi_q(x, y)$ would gradually map to $l$ discrete values with identical interval $2\pi/l$ [25]. Note that since we replace $\phi(x, y)$ with $\phi_q(x, y)$ in optimization process, the corresponding PSF in Fig. 1 is related to the factor $e^{i\phi(x, y)}$ instead of $e^{i\phi(x, y)}$. Without loss of generality, we choose $l = 8$ in this paper. By Eq. (6), we built the analytical relation between $\phi_q(x, y)$ and $\phi(x, y)$. Therefore, in our end-to-end framework, the gradient of Eq. (5) with respect to the variable $\phi(x, y)$ is back-propagated through the entire pipeline, and is efficiently handled by the autograd package [26].

**B. Inverse design of single-cell metasurfaces**

With the above process (Sec. II A), a unique phase mask $\phi_q(x, y)$ can be designed which is assumed to work for all the relevant working channels ($\lambda, z, \sigma$). We then seek to map it into a single-cell metasurface by the PB phase scheme, as shown in the dash box of Fig. 1. Firstly, we need to find an optimal geometry of the met-atom with appropriate length $L_x$, width $L_y$, height $H$ and period $\lambda$, as shown in the inset of Fig. 2(a). In order to maximize the conversion efficiency for the circular polarization. Here we have combined the Bayesian optimization and finite-different time-domain (FDTD) solver to design the meta-atom. The details are shown by Fig. 1. Note that Bayesian optimization is a powerful framework based on Gaussian process, which is able to find the globally optimal solution with as few iterations as possible [27, 28, 29]. Figure 2(a) shows that the corresponding conversion efficiency $\eta$ reaches 67%, 87%, 95%, and 72% for wavelength $\lambda = 470$ nm, 545 nm, 648 nm, and 800 nm (marked by black dots), respectively. The designed TiO$_2$
meta-atom (nanoblock) with $L_g = 200 \text{ nm}$, $L_g = 100 \text{ nm}$, $\Lambda = 235 \text{ nm}$ and $H = 940 \text{ nm}$ sitting on SiO$_2$ substrate with a fixed height of $1 \text{ \mu m}$. Secondly, since the outgoing phase of the nanoblock is about twice its rotation angle with respect to the principle axes [30], as shown in Fig. 2(b), the phase profile can be mapped into a metasurface via rotating each meta-atom by an appropriate angle with respect to the coordinate axes. Figure 2(b) shows that employing realistic, finite-sized nanoblock metaatom results in negligible variations of the PB phase with respect to the theoretical relationship $\varphi = 2\theta$ (see Fig. 1) [24], [31].

Thirdly, we validated the performance of the multi-channel imaging system by replacing the phase mask with the designed metasurface, and we calculate the transmitted field of the metasurface by full-wave FDTD solver, as well as the PSF by the AS method directly from the optimal phase $\hat{\phi}(x, y)$.

### III. RESULTS

For simplicity, we denote the dimensions of a ground truth object that is going to be imaged by the lensless metasurface as $m_0 = N \times n \times n$, where $N$ represents the number of channel (i.e., characterized by wavelength $\lambda$, depth $z$ and polarization $\sigma$) each of which consists of an object with $n \times n$ pixels, and the total number of sensor pixel is set as $m_t$. Since the Tikhonov regularization we used in Eq. (3) is suitable for over-determined inverse problems, in the following examples we impose $m_t > m_0$ to guarantee the performance of the proposed multichannel imaging.

#### A. Multi-spectral imaging over the visible and infrared bands

Without loss of generality, we firstly use four sets of random binary objects as input to train computational imaging system with the working wavelengths of $\lambda = 470$ nm, 545 nm, 648 nm, and 800 nm. We set the distance between the phase mask and the object planes as $z = 2 \text{ cm}$, and that between the phase mask and the sensor as $d = 62 \text{ \mu m}$ (hereafter referred them as object-image distance). It takes around two hundred iterations for the loss function $L$ to converge, using the moving method of asymptotes (MMA) to update the gradient [4].

Figure 3(a) shows the designed phase profile (normalized by $2\pi$). Figure 3(b) shows the sensor-received pattern with some obvious copies of the objects (ground truth) located in different positions. Figures 3(c) and 3(d) show the corresponding PSF generated by the optimized ideal phase mask and the actual metasurface, respectively, and the panels in the column from left to right represent the spectral channel of wavelength $\lambda = 470$ nm, 545 nm, 648 nm, and 800 nm, respectively (PSFs for wavelengths in 5 nm increments ranging from 460 nm to 800 nm are shown in Fig. S2). Such results show strong agreement despite that there is a slight spatial shift of the main spot for the metasurface-based results. Since we use the AS method to calculate the PSFs, they have the same size of pixel with the transmitted field of metasurface (i.e., $\Lambda \times \Lambda$), which is too small to match the pixel size of commercial sensor. Therefore, we take each block with $5 \times 5$ pixels of PSF as one super-pixel of size $\sim 1.2 \times 1.2 \text{ \mu m}^2$ by replacing the intensity value of the super-pixel with the summation of the intensities contributed by the super-pixel containing $5 \times 5$ blocks.

Figures 3(e)-3(g) show the reconstruction results with images of $10 \times 10$ pixels under the noise of standard deviation $\delta = 5\%$. Figures 3(e)-3(g) represent the ground truth, reconstruction image by the designed phase mask, and reconstruction image by the actual metasurface, respectively. (It is emphasized that the ground truth in Fig. 3(e) is randomly generated for performance test of the designed imaging system, which is completely different from that in the optimization process.) The results in the column from the left to the right are for the wavelength channels of $\lambda = 470$ nm, 545 nm, 648 nm, and 800 nm, respectively.
respectively, and all of them are normalized between 0 and 1. It is seen that the reconstruction image based on the phase mask looks nearly identical to the ground truth with the PSNR over 33 dB for each channel, as shown in Table S1. By contrast, the PSNR of the metasurface-based reconstruction image is slightly larger with the PSNR between 29 dB and 31 dB. For comparison, the reconstruction result under the noise of standard deviation $\delta=1\%$ are plotted in Fig. S3 and the corresponding PSNR and SSIM metric are shown in Table S2.

The reconstruction by the designed phase mask and the ground truth match very well with the corresponding PSNRs over 28.9 dB, whereas that by the designed metasurface shows similar patterns with the corresponding PSNR ranging from 22.0 dB to 33.6 dB, as shown in Table S5. Table S6 shows that all the MSEs by the designed phase mask (metasurface) for each channel is below 0.12% (0.63%), while Table S7 shows the SSIM for each channel is over 0.995, which demonstrates the efficiency of the end-to-end inverse design framework for multi-wavelength and multi-depth imaging. For comparison, the PSNR of reconstructed image under the noise of standard deviation $\delta=1\%$ are shown in Table S8, which shows that the corresponding PSNR of phase-based metasurface-based reconstruction in each channel is over 34.6 dB and 27.9 dB, respectively.

By far, we have demonstrated a general inverse design scheme for four-channel computational imaging based on a phase mask and a single-cell metasurface, respectively. To achieve high resolution of imaging, we increase the pixel number of phase mask up to 1920×1920 and re-optimize the computational system with random binary input. Figure 4(a) shows the designed phase mask. To verify the performance of imaging system, natural scene is taken as the channel imaging system in Fig. 1, working at the input (i.e. ground truth) with 128×128 pixels and the corresponding sensor image under the noise of standard deviation $\delta=1\%$ is shown in Fig. 4(b). Figure 4(c) and 4(d) shows the corresponding ground truth and reconstruction image by phase mask, which show the strong agreement with each other. To further demonstrate the power of such an imaging system, we take another natural case as input and complete the reconstruction. Figure 4(e) and 4(f) shows the corresponding ground truth and reconstruction imaging. There is strong agreement between them. The results in the column from the left to the right are for the wavelength channels of $\lambda=470$ nm, 545 nm, 648 nm, and 800 nm, respectively. Table S4 shows their corresponding PSNR and SSIM.

Generally, separating scenes encoded at various wavelengths on sensor image facilitates recovery algorithm to accurately reconstruct images. In order to further improve the quality of image reconstruction, we have incorporated some design constraint in the end-to-end process to mitigate this overlapping, which increases the accuracy of image reconstruction, as shown in Fig. S4 and Table S3.

**B. Multi-spectral and multi-depth imaging**

In many multi-channel imaging scenarios, both the depth and spectral information of object are vital, such as in applications of target detection, infrared and visible image fusion, and three-dimensional imaging [32], [33], [34]. Therefore, we applied the inverse design scheme to optimize a lensless imaging system for image reconstruction task of both depth and spectrum detection. Figure 5 shows an example of eight-channel image reconstruction with four wavelengths and two depths. Figures 5(a) and (b) show the optimized phase profile (again normalized by $2\pi$) with 320×320 pixels and sensor image with $54\times54$ pixels under the noise standard deviation $\delta=2\%$, respectively. The corresponding PSFs are shown in Fig. S6. Figures 5(c) and 5(d) show the reconstruction results for the object distance $z=1$ mm and $z=2$ mm, respectively, wherein the panels in the upper, middle and lower rows represent the ground truth, reconstruction images by a phase mask, and the reconstruction images by the designed metasurface. The panels in the column from left to right in Figs. 5(c) and 5(d) represent the images at the spectral channels of wavelength $\lambda=470$ nm, 545 nm, 648 nm, and 800 nm, respectively.

The reconstruction by the designed phase mask and the ground truth match very well with the corresponding PSNRs over 28.9 dB, whereas that by the designed metasurface shows similar patterns with the corresponding PSNR ranging from 22.0 dB to 33.6 dB, as shown in Table S5. Table S6 shows that all the MSEs by the designed phase mask (metasurface) for each channel is below 0.12% (0.63%), while Table S7 shows the SSIM for each channel is over 0.995, which demonstrates the efficiency of the end-to-end inverse design framework for multi-wavelength and multi-depth imaging. For comparison, the PSNR of reconstructed image under the noise of standard deviation $\delta=1\%$ are shown in Table S8, which shows that the corresponding PSNR of phase-based and metasurface-based reconstruction in each channel is over 34.6 dB and 27.9 dB, respectively.

---

**Fig. 4. Results of four-wavelength high resolution imaging reconstruction.**

(a) The designed phase profile normalized by $2\pi$. (b) Sensor raw image with noise. (c) and (d) The corresponding ground truth and reconstruction images by the designed phase mask, respectively. (e) and (f) are the same plots as (c) and (d) by the same phase mask but with different input scene (i.e. ground truth). Note that the panels in the column from left to right in (c)-(f) represent the spectral channel of wavelength $\lambda=470$ nm, 545 nm, 648 nm, and 800 nm, respectively.
C. Spectro-polarimetric-depth imaging

Spectro-polarimetric-depth imaging in the visible and infrared region is pivotal across various applications, spanning from night vision and machine perception to gas detection [11], [35]. In this regard, we applied the inverse design scheme to optimize a spectro-polarimetric-depth imaging system based on a single-cell metasurface. Here we ingeniously applied a single-phase profile to achieve multi-polarization imaging, and then extend it to a spectro-polarimetric-depth imaging. Specifically, according the principle of geometric phase, the delay phase between the incident and transmitted waves is linearly proportional to the rotation angle $\theta$ of meta-atom. Consider that the incident and transmitted beam is with left (right) and right (left) circular polarization, respectively, the relation between incident electric field $E_{i\text{LCP/RCP}}$ and transmitted electric field $E_{t\text{RCP/LCP}}$ can be described by [21]:

$$E_{t\text{RCP/LCP}} = e^{i\varphi}E_{i\text{LCP/RCP}},$$

(10)

where the delay phase is $\varphi = 2\theta$. Under the incident light with different polarizations, the phase delay is opposite with $\pm\varphi$. We use the relation to re-build a surrogate model for phase mask with the response of $e^{i\varphi}$ for cross circular polarization design, and then embed the model in the inverse design scheme in Fig. 1 for multi-polarization imaging design. To demonstrate this method, we first inverse designed a two-polarization imaging system to obtain image reconstruction, as shown in Fig. S7, and then apply it to optimize a spectro-polarimetric-depth imaging system.

Figure 6 shows an example of sixteen-channel image reconstruction with four wavelengths, two depths and two polarizations. Figures 6(a) and 6(b) show the optimized phase profile (again normalized by $2\pi$) with $470 \times 470$ pixels and sensor image with $84 \times 84$ pixels under the noise standard deviation $\delta = 1\%$, respectively. The corresponding PSFs are hot spot-like pattern, as shown in Figs. S8 and S9, respectively. This leads to the overlap of different scene (i.e., capital letters) in the sensor image. Figure 6(c) and 6(d) show the reconstruction results for the object distance $z = 1\text{ mm}$ and $z = 2\text{ mm}$, respectively, wherein the panels in the upper, middle and lower rows represent the ground truth, reconstruction images by a phase mask, and the reconstruction images by the designed metasurface. The panels in the column from left to right in Figs. 6(c) and 6(d) represent the images at the spectral channels of wavelength $\lambda = 470\text{ nm}$, $\lambda = 545\text{ nm}$, $\lambda = 648\text{ nm}$, and $\lambda = 800\text{ nm}$, respectively. The odd and even column represent the images in the LCP and RCP, respectively. The reconstruction by the designed phase mask and the ground truth match very well with the PSNR for each channel over 37.4 dB, whereas the PSNR of the metasurface-based reconstructed image in each channel is above 32.6 dB, as shown in Table S9, which demonstrates the efficiency of the end-to-end inverse design framework for spectro-polarimetric-depth imaging.

IV. DISCUSSION AND CONCLUSION

We have demonstrated a general inverse design scheme for multi-channel lensless computational imaging based on a single-cell metasurface. Since the FDTD simulation for
the large-scale metasurface costs large computational memory and resources, the designed metasurface is set to include 470×470 meta-atoms maximally, which limits the total number of pixels in the reconstruction image in Figs. 3-5. In fact, in our inverse design approach the size of phase mask and the number of pixels in image can be easily extended to \(10^3 \lambda \times 10^3 \lambda\) and \(2 \times 10^3\), respectively, and the phase mask could be realized as PB-phase type metasurface and verified experimentally. More examples of multispectral computational imaging based on a large-scale phase mask can be found in Figures S10 and S11. Figure S10 shows a four-wavelength channel imaging example which realizes the reconstruction of gray image (with 128×128 pixels) for the letter “B”, “G”, “R”, and combined “IR”. In addition, Fig. S11 shows Fashion-MNIST image reconstruction for four wavelengths and two depths.

Since the performance of the designed metasurface highly depends on the assumption of LPA [36], it is important to make the designed phase profile as smooth as possible and in turn prevent the geometry of the resulting adjacent meta-atoms varying rapidly. Therefore, we took the metasurface with identical meta-atom (i.e. a phase mask with same identical phase value in each pixel) as the initial structure in the training process. On top of that, we digitalize the phase profile into eight levels by soft-quantization function and then map it into metasurface by the PB phase method. We finally use FDTD method to calculate the metasurface-based PSFs and apply them to reconstruct the object from different channels.

As a matter of fact, to achieve multi-channel imaging, one needs to accurately solve the ill-posed problem [i.e., Eq. (4)]. Physically, the accurate solution of Eq. (4) prefers the well-designed PSFs with low degree of correlation [32]. Therefore, the correlation degree of PSFs in different channel becomes weaker and weaker during training process, and they finally turn into sharp peak-like pattern automatically. To quantify that, we defined the correlation degree of PSFs as

\[
C = \ln \left(1/ \sum_{i,j} \frac{\|\text{PSF}_i - \text{PSF}_j\|^2}{N} \right),
\]

where \(i\) and \(j\) represent different channel of \((\lambda, z)\). Figures 7(a) and 7(b) explicitly show how the correlation degree between PSFs decreases with iterations, for the case of four-wavelength channel imaging in Fig. 3 and the case of four-wavelength-two-depth channel imaging Fig. 5, respectively. In this work, the end-to-end design scheme manages to reduce the correlation degree \(C\) by optimizing a phase mask so as to improve the performance of multi-channel imaging. However, once the phase profile is mapped into a metasurface, the coupling of meta-atoms

![Image](image-url)
would inevitably introduce background noise to the resulting metasurface-based PSF and in turn lead to increased $C$. Therefore, the PSNRs of the phase-based reconstruction image are larger than that of the metasurface-based reconstruction, as shown in Tables S1 and S5.

![Fig. 7. The correlation degree between PSFs for (a) four-wavelength channel and (b) four-wavelength-two-depth channel imaging change with iterations.](image)

In summary, we have proposed an end-to-end inverse design scheme for multi-channel imaging which is based on a physically interpretable optimization method. The approach connects both the optical domain and the electronic domain ingredients. Different from the multi-channel imaging design by theoretically specifying the optical element with spectral and spatial-varying PSF [14], [33], we apply the inverse design scheme to optimize the optical frontend which finally and simultaneously generates PSFs that are highly dependent on the working wavelength, the object distance and the helicity of circular polarization. Compared to the design of multi-channel imaging by machine learning algorithm [13], our scheme does not require a large number of training dataset (only one object for each channel), which saves a lot of computation time and resources in the training process. The reason of that is that our computational image model is fundamentally rooted in an imaging model that employs the paraxial approximation, which makes the PSF as being shift-invariant. This means that in each channel $(\lambda, z, \sigma)$, all the inputs (or objects) $I(\lambda, z, \sigma)$ are corresponding to the same PSF $\text{PSF}_{\lambda, z, \sigma}$, as shown in Eq. (1), no matter what the intensity distribution of $I(\lambda, z, \sigma)$ is. Consequently, if a computational imaging system is capable of reconstructing one object within a specific channel $(\lambda, z, \sigma)$, it can inherently reconstruct a variety of different objects within that same channel.

Under the condition of the finite dimensions (or pixel number) of the camera sensor, the PSNR of multi-channel computational imaging would be inversely proportional to the total pixel number of the reconstruction image $m_i = N\Phi$, where $N$ represents the number of designed channel and $\Phi$ represents the total pixel number of reconstruction image in each channel. In essence, the theoretical and mathematical frameworks of lensless computational imaging suggest that multichannel image recovery is feasible for sensors with either sufficiently high pixel resolutions or, conversely, for those with relatively low pixel resolutions. The quality of the reconstructed images depends not only on the interplay between the number of channels, the comprehensive sampling points of the object, and the sensor's pixel resolution but is also influenced by the presence of noise, such as photon shot noise, thermal noise, and dark current noise. Additionally, the backend algorithm plays a pivotal role in this process. Given the above factors, it is challenging to identify a fundamental principle governing the trade-offs involved.

REFERENCES


Supporting Information

Spectro-polarimetric-depth Imaging by Inverse-Designed Single-Cell Metasurface

DASEN ZHANG,1,2 TING MA,1,2, YUQI PENG,1,2 XIANJIN LIU,1,2 QIWEN BAO,1,2 AND JUN-JUN XIAO*1,2

1College of Electronic and Information Engineering, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China
2Shenzhen Engineering Laboratory of Aerospace Detection and Imaging, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China

*Corresponding author: eiexiao@hit.edu.cn

Section 1: Meta-atom designed by Bayesian optimization
Section 2: Phase-based point spread function and evaluation of imaging quality for four-wavelength channel imaging system in Fig. 3.
Section 3: Optimization of imaging quality for four-wavelength channel imaging system in Fig. 4 by mitigating overlapping on sensor image.
Section 4: MSE, PSNR and SSIM metric for four-wavelength, two-depth and two-polarization imaging system in Fig. 5.
Section 5: Computational imaging system for two polarizations
Section 6: Point spread functions and the evaluation of image quality for sixteen channel imaging in Fig. 6
Section 7: Multi-channel imaging based on a large-scale phase mask
Section 1: Meta-atom designed by Bayesian optimization

To design meta-atoms with high conversion efficiency, we apply the Bayesian optimization approach to optimize meta-atom. To achieve that, we need to maximize the objective function defined as:

$$ Y = \sum_{k=1}^{M} Y_{\lambda_k}, $$

(1)

where $M = 4$, and $Y_{\lambda_k}$ is the cross circular polarization conversion efficiency at the wavelength of $\lambda_k$, $k = 1, 2, 3, 4$, where $\lambda_1 = 470 \text{ nm}$, $\lambda_2 = 545 \text{ nm}$, $\lambda_3 = 648 \text{ nm}$, $\lambda_4 = 800 \text{ nm}$.

Figure S1 depicts the Bayesian optimization framework. First, we create a database $X$ for the geometry of meta-atom, such as length $L_x$, width $L_y$, height $H$, period $\Lambda$. Second, we randomly choose some structures from the database (i.e. initial structures) and then use an electromagnetic solver (based on FDTD method) to calculate their corresponding cross circular polarization conversion efficiency $Y$. Third, a probabilistic regression model is learned from the training set $(X, Y)$. Fourth, this model is used to estimate the next candidates $X_{\text{new}}$, and then compute the corresponding $Y_{\text{new}}$ by the electromagnetic solver. Fifth, we retrain the regression model by updating the training set $X = X \cup X_{\text{new}}$ and $Y = Y \cup Y_{\text{new}}$. The fourth and fifth steps are repeated until an optimum or pre-set criterion is achieved.

![Fig. S1. Bayesian optimization flowchart for the design of meta-atom](image)

#### Section 2: Phase-based point spread function and evaluation of imaging quality for four-wavelength channel imaging system in Fig. 3.

Point Spread Function (PSF) is a fundamental characteristic of optical systems, detailing the system’s response to a point light source and influencing image resolution and quality. Figure S2 shows the phase-based PSF customized for wavelengths spanning 460 nm to 800 nm in 5 nm increments, effectively encompassing the full visible spectrum. Figures S2(a1)-S2(a69) illustrate the PSFs corresponding to each wavelength, with increments of 5 nm, from $\lambda_1 = 460 \text{ nm}$ to $\lambda_4 = 800 \text{ nm}$. Normalization of all PSFs is performed against the maximum PSF value at $\lambda_1 = 470 \text{ nm}$ (see Figure S2(a3)), establishing a baseline for comparison.

It is evident that the PSF’s intensity profile shifts with variations in wavelength. Notably, as the wavelength approaches $\lambda_4 = 800 \text{ nm}$, the PSF intensity deviates from the designed benchmarks ($\lambda = 470 \text{ nm}$, $545 \text{ nm}$, $648 \text{ nm}$, and $800 \text{ nm}$), there is a corresponding increase or decrease in the maximum intensity value of the PSF. This fluctuation in intensity is crucial; and a higher PSF intensity is conducive to enhancing the quality of the reconstructed images. These observations underscore the importance of inverse design in the computational imaging system, ensuring optimal performance across the specified wavelength range.

To assess the imaging quality of the four-wavelength channel system in Fig. 3, we compared the Mean Square Error (MSE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity (SSIM) indices for each wavelength channel in Table S1. The analysis reveals that the MSE for phase-based reconstructions is consistently lower than that for metasurface-based reconstructions when compared to the ground truth, indicating superior accuracy. Concurrently, the PSNR values for phase-based images exceed those of metasurface-based images for all channels, further substantiating the higher quality of phase-based reconstructions. Notably, all SSIM values for both phase-based and metasurface-based reconstructions are unity, signifying perfect structural preservation. The PSNR values surpass 29.2 dB, and the MSE values are all below 0.14%, which is under the specified noise standard deviation $\delta = 5\%$, underscoring the robust performance of the computational system across the four-wavelength channels.

To further validate the system’s performance under varying noise conditions, we adjusted the noise standard deviation to match the levels used in Figs. 6 and 4 ($\delta = 1\%$), and performed image reconstruction anew. The resulting images and their quantitative metrics are presented in Fig. S3 and Table S2, respectively. Figures S3(a)-S3(c) display the ground truth, phase-based reconstruction, and metasurface-based reconstruction, respectively. The columns, from left to right, correspond to the wavelength channels of 470 nm, 545 nm, 648 nm, and 800 nm. The visual alignment is strikingly precise, with the quantitative PSNR
exceeding 42.3 dB and the SSIM maintaining a perfect score of 1, as detailed in Table S2. These findings confirm the exceptional imaging fidelity and resilience of the system against noise disturbances.

Fig. S2. (a1)-(a69) represent the intensity distribution of PSF for wavelengths ranging from 460 nm to 800 with 5 nm increments.

Table S1. MSE between the phase-based and metasurface-based reconstruction image and the ground truth, and PSNR and SSIM for four-wavelength channel imaging system in Fig. 3 under the noise of standard deviation $\delta = 5\%$.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Scheme</th>
<th>$\lambda = 470$ nm</th>
<th>$\lambda = 545$ nm</th>
<th>$\lambda = 648$ nm</th>
<th>$\lambda = 800$ nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>Phase mask</td>
<td>0.0004</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0004</td>
</tr>
<tr>
<td></td>
<td>Metasurface</td>
<td>0.0012</td>
<td>0.0014</td>
<td>0.0008</td>
<td>0.0010</td>
</tr>
<tr>
<td>PSNR</td>
<td>Phase mask</td>
<td>33.7</td>
<td>35.9</td>
<td>35.6</td>
<td>33.9</td>
</tr>
<tr>
<td></td>
<td>Metasurface</td>
<td>29.2</td>
<td>28.6</td>
<td>30.8</td>
<td>30.0</td>
</tr>
<tr>
<td>SSIM</td>
<td>Phase mask</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Metasurface</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Fig. S3 (a) Ground truth. (b) and (c) Reconstruction images by the designed phase mask and the metasurface, respectively. Note that the panels in the column from left to right in (a)-(c) represent the spectral channel of wavelength $\lambda = 470$ nm, 545 nm, 648 nm, and 800 nm, respectively.

Table S2. PSNR and SSIM for four-wavelength channel imaging system in Fig. 3 under the noise of standard deviation $\delta = 1\%$.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Scheme</th>
<th>$\lambda = 470$ nm</th>
<th>$\lambda = 545$ nm</th>
<th>$\lambda = 648$ nm</th>
<th>$\lambda = 800$ nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td>Phase mask</td>
<td>47.5</td>
<td>49.1</td>
<td>49.0</td>
<td>47.5</td>
</tr>
<tr>
<td></td>
<td>Metasurface</td>
<td>43.1</td>
<td>42.3</td>
<td>44.6</td>
<td>43.7</td>
</tr>
<tr>
<td>SSIM</td>
<td>Phase mask</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Metasurface</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Section 3: Optimization of imaging quality for four-wavelength channel imaging system in Fig. 4 by mitigating overlapping on sensor image.

Generally, separating scenes encoded at various wavelengths on sensor image facilitates recovery algorithm to accurately reconstruct images. Here we have incorporated the design constraint in the end-to-end process to mitigate this overlapping. For that purpose, we can change the original objective to

$$L = \frac{1}{N} \sum_i \| i(\lambda_i) - \hat{i}(\lambda_i) \|^2 + \frac{w}{\sum_{\lambda_j} \| v(\lambda_j) - \hat{v}(\lambda_j) \|^2},$$

(S2)

$$v(\lambda_j) = \text{PSF}_\lambda \otimes I(\lambda_j)$$

(S3)

where $N$ is the total number of the considered channel, $w = 4 \times 10^9$ is a weight factor chosen empirically, and we can explicitly set $\lambda_1 = 470$ nm, $\lambda_2 = 545$ nm, $\lambda_3 = 648$ nm, and $\lambda_4 = 800$ nm. By incorporate the constraint of the second term in Eq. (S2) under which we expect to minimize cross talk of different channel scenes on the sensor, we have successfully reduced the overlapping in the sensor image, which improves the performance of computational imaging system as shown in Fig. S4. Table S3 shows the corresponding PSNR and SSIM for each wavelength channel. For comparison, Table S4 shows the same metric of the image reconstruction in Fig. 4 whose imaging system is designed by the objective function in Eq. (1). It can be seen that when we replace the objective function Eq. (1) by Eq. (S2), the average PSNR of Fig. S4(d) and Fig. S4(f) across different channel increases by 4.6% and 4.7%, respectively, compared to Fig. 4(d) and Fig. 4(f). Whereas the average SSIM for Fig. S4(d) and Fig. 4(d) (as well as for Fig. S4(f) and Fig. S4(f)) is almost equal, as shown in Table S3 and Table S4. Although we also notice that the PSNR of Fig. S4(d) and Fig. S4(f) in the channel $\lambda = 545$ nm and $\lambda = 800$ nm is slightly lower than that of Fig. 4(d) and Fig. 4(f), this problem can be solved by fine-tuning the weight factor in Eq. (S2) or developing a more reasonable objective function.
Table S3. PSNR and SSIM of the phase-based reconstruction image for four-wavelength high-resolution imaging system in Fig. 4 by using the Eq. (S2) as objective function.

<table>
<thead>
<tr>
<th>image</th>
<th>Metric</th>
<th>$\lambda = 470$ nm</th>
<th>$\lambda = 545$ nm</th>
<th>$\lambda = 648$ nm</th>
<th>$\lambda = 800$ nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. S4(d)</td>
<td>PSNR</td>
<td>32.0</td>
<td>25.8</td>
<td>31.8</td>
<td>26.9</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.992</td>
<td>0.972</td>
<td>0.994</td>
<td>0.939</td>
</tr>
<tr>
<td>Fig. S4(f)</td>
<td>PSNR</td>
<td>31.6</td>
<td>25.5</td>
<td>31.5</td>
<td>26.3</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.992</td>
<td>0.971</td>
<td>0.994</td>
<td>0.973</td>
</tr>
</tbody>
</table>

Fig. S4 Results of four-wavelength high resolution imaging reconstruction. (a) The designed phase profile normalized by $2\pi$. (b) Sensor raw image with noise. (c) and (d) The corresponding ground truth and reconstruction images by the designed phase mask, respectively. (e) and (f) are the same plots as (c) and (d) by the same phase mask but with different input scene (i.e. ground truth). Note that the panels in the column from left to right in (c)-(f) represent the spectral channel of wavelength $\lambda = 470$ nm, 545 nm, 648 nm, and 800 nm, respectively.

Table S4. Same plot as Table S3 but with the Eq. (1) as objective function.

<table>
<thead>
<tr>
<th>image</th>
<th>Metric</th>
<th>$\lambda = 470$ nm</th>
<th>$\lambda = 545$ nm</th>
<th>$\lambda = 648$ nm</th>
<th>$\lambda = 800$ nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 4(d)</td>
<td>PSNR</td>
<td>27.4</td>
<td>27.7</td>
<td>28.5</td>
<td>27.8</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.984</td>
<td>0.985</td>
<td>0.986</td>
<td>0.955</td>
</tr>
<tr>
<td>Fig. 4(f)</td>
<td>PSNR</td>
<td>26.9</td>
<td>27.4</td>
<td>28.2</td>
<td>27.2</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.984</td>
<td>0.984</td>
<td>0.987</td>
<td>0.978</td>
</tr>
</tbody>
</table>
Section 4: MSE, PSNR and SSIM metric for four-wavelength, two-depth and two-polarization imaging system in Fig. 5.

To assess the imaging quality of the four-wavelength and two-depth channel system in Fig. 5, we compared PSNR, MSE, and SSIM indices for each wavelength channel in Tables S5-S7, respectively. As shown in Table S5, the PSNR value for phase-based image in each channel exceeds 28.9 dB, while that of metasurface-based image varies from 22.0 dB to 33.6 dB. Table S6 shows the MSE value for phase-based image in each channel below 0.12% and that for metasurface-based image ranging from 0.04% to 0.63%. Table S7 shows that all SSIM values for both phase-based and metasurface-based reconstructions exceed 99.5%. Overall, the quality of phase-based image is better than that of metasurface-based image and all of them shows a very high SSIM with the ground truth.

To further validate the system’s performance under varying noise conditions, we adjusted the noise standard deviation to match the levels used in Figs. 6 and 4 (δ = 1%), and performed image reconstruction anew. The resulting images and their quantitative metrics are presented in Fig. S5 and Table S8, respectively. Figures S5(a) and S5(b) show the reconstruction results with corresponding depth (i.e., object distance) z = 1 mm and z = 2 mm, respectively, and the panels from up do down display the ground truth, phase-based reconstruction, and metasurface-based reconstruction, respectively. In Figure S5(a) or Figure S5(b), the columns, from left to right, correspond to the wavelength channels of 470 nm, 545 nm, 648 nm, and 800 nm. The corresponding phase-based and metasurface-based PSF for each channel is shown in Fig. S6. It can be seen that they agree with each other very well, which demonstrates the power of our inverse design scheme. Table S8 shows that the resulting PSNR of phase-based and metasurface-based image for each channel exceeds 38.1 dB and 27.9 dB, which is larger than the corresponding PSNR under noise level δ = 2% (see Table S5).

Table S5. PSNR of the phase-based and metasurface-based reconstruction image for four-wavelength and two-depth imaging system in Fig. 5.

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Distance</th>
<th>( \lambda = 470 \text{ nm} )</th>
<th>( \lambda = 545 \text{ nm} )</th>
<th>( \lambda = 648 \text{ nm} )</th>
<th>( \lambda = 800 \text{ nm} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase mask</td>
<td>1 mm</td>
<td>28.9</td>
<td>37.2</td>
<td>35.5</td>
<td>33.8</td>
</tr>
<tr>
<td></td>
<td>2 mm</td>
<td>32.7</td>
<td>34.2</td>
<td>35.3</td>
<td>36.0</td>
</tr>
<tr>
<td>Metasurface</td>
<td>1 mm</td>
<td>28.6</td>
<td>32.6</td>
<td>26.9</td>
<td>30.6</td>
</tr>
<tr>
<td></td>
<td>2 mm</td>
<td>31.4</td>
<td>23.8</td>
<td>22.0</td>
<td>33.6</td>
</tr>
</tbody>
</table>

Table S6. MSE between the phase-based and metasurface-based reconstruction image and the ground truth for four-wavelength and two-depth imaging system in Fig. 5.

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Distance</th>
<th>( \lambda = 470 \text{ nm} )</th>
<th>( \lambda = 545 \text{ nm} )</th>
<th>( \lambda = 648 \text{ nm} )</th>
<th>( \lambda = 800 \text{ nm} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase mask</td>
<td>1 mm</td>
<td>0.0012</td>
<td>0.0002</td>
<td>0.0003</td>
<td>0.0004</td>
</tr>
<tr>
<td></td>
<td>2 mm</td>
<td>0.0005</td>
<td>0.0004</td>
<td>0.0003</td>
<td>0.0003</td>
</tr>
<tr>
<td>Metasurface</td>
<td>1 mm</td>
<td>0.0014</td>
<td>0.0005</td>
<td>0.0020</td>
<td>0.0009</td>
</tr>
<tr>
<td></td>
<td>2 mm</td>
<td>0.0007</td>
<td>0.0041</td>
<td>0.0063</td>
<td>0.0004</td>
</tr>
</tbody>
</table>

Table S7. SSIM of the phase-based and metasurface-based reconstruction image for four-wavelength and two-depth imaging system in Fig. 5.

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Distance</th>
<th>( \lambda = 470 \text{ nm} )</th>
<th>( \lambda = 545 \text{ nm} )</th>
<th>( \lambda = 648 \text{ nm} )</th>
<th>( \lambda = 800 \text{ nm} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase mask</td>
<td>1 mm</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>2 mm</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Metasurface</td>
<td>1 mm</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>2 mm</td>
<td>1.0</td>
<td>0.999</td>
<td>0.995</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Fig. S5 (a) and (b) image reconstruction for the object distance in $z = 1$ mm. The first, second and third row represents ground truth, phase-based and metasurface-based reconstruction, respectively, whereas the panels in the column from left to right represent the spectral channel of wavelength $\lambda = 470$ nm, 545 nm, 648 nm, and 800 nm, respectively. (b) same as (a) but for $z = 2$ mm.

Table S8. PSNR of the phase-based and metasurface-based reconstruction image for four-wavelength and two-depth imaging system in Fig. 5 under the noise standard deviation $\delta \sim 1\%$.

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Distance</th>
<th>$\lambda = 470$ nm</th>
<th>$\lambda = 545$ nm</th>
<th>$\lambda = 648$ nm</th>
<th>$\lambda = 800$ nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase mask</td>
<td>1 mm</td>
<td>34.6</td>
<td>43.1</td>
<td>41.5</td>
<td>39.5</td>
</tr>
<tr>
<td></td>
<td>2 mm</td>
<td>38.1</td>
<td>40.0</td>
<td>41.1</td>
<td>41.7</td>
</tr>
<tr>
<td>Metasurface</td>
<td>1 mm</td>
<td>34.5</td>
<td>38.6</td>
<td>32.8</td>
<td>36.6</td>
</tr>
<tr>
<td></td>
<td>2 mm</td>
<td>37.4</td>
<td>29.8</td>
<td>27.9</td>
<td>39.6</td>
</tr>
</tbody>
</table>

Fig. S6. Phase-based PSFs and metasurface-based PSFs for eight-channel imaging shown in Fig. 4. Phase-based PSFs for (a) $z = 1$ mm and (b) $z = 2$ mm. Metasurface-based PSFs for (c) $z = 1$ mm and (d) $z = 2$ mm. The results in from left to right column is for the wavelength $\lambda = 470$ nm, $\lambda = 545$ nm, $\lambda = 648$ nm, and $\lambda = 800$ nm, respectively.
Section 5: Computational imaging system for two polarizations

In order to test the surrogate model, we embedded it into the inverse design scheme in Fig. 1 for two-polarization imaging with $6 \times 6$ pixels, and the corresponding results are shown in Fig. S7. Figure S7(a) shows the optimized phase mask, while Fig. S7(b) shows its corresponding PSF for LCP and RCP under the incident wavelength of $\lambda = 800$ nm. It can be seen that we successfully inverse design the peak-like type PSF. Figure S7(c) shows the corresponding metasurface-based PSF that resembles the phase mask-based PSF. These peak-like PSFs impose the convolution of the input scene to the sensor response. Figs. S7(d) and S7(e) shows the corresponding sensor image by phase mask and metasurface, respectively. Figures S7(f)–S7(h) shows the corresponding reconstruction result, where the upper, middle and bottom panels represent the ground truth, reconstruction image by phase mask, and reconstruction image by metasurface. It can be seen that they look similar to each other, which demonstrates the viability of the inverse design.

Section 6: Point spread functions and the evaluation of image quality for sixteen channel imaging in Fig. 6

Figures S8 and S9 show the phase-based and metasurface-based PSFs for sixteen-channel imaging system discussed in Fig. 6, corresponding the object distance $z = 1$ mm and $z = 2$ mm, respectively. Figures S8(a) and S8(b) show the phase-based PSF with LCP and RCP, respectively. The corresponding metasurface-based PSF are shown in Figures S8(c) and S8(d). The panels in the column from left to right in Figs. S8(c) and S8(d) represent the images at the spectral channels of wavelength $\lambda = 470$ nm, $\lambda = 545$ nm, $\lambda = 648$ nm, and $\lambda = 800$ nm, respectively. Figures S9(a)-S9(d) show the same plots of Figures S8(a)-S8(d) but with $z = 2$ mm. It can be seen that the phase-based and metasurface-based PSFs agree with each other very well. All the PSFs have the peak-like patterns, which can reduce the correlation between that in different channel, as we discussed in main text.

For evaluating the imaging performance of the system depicted in Fig. 6, which encompasses four wavelengths, two depths, and two polarizations, we have calculated the PSNR and SSIM for each wavelength channel. The results are detailed in Tables S9-S10. Table S9 indicates that the PSNR for phase-based images surpasses 37.4 dB, while metasurface-based images achieve over 32.6 dB across all channels. This signifies that both methods deliver excellent image quality. Moreover, the SSIM values in Table S10 have reached the maximum of 1, denoting perfect similarity with the ground truth. In summary, the phase-based imaging outperforms the metasurface-based approach, yet both exhibit exceptional quality as evidenced by the high PSNR and SSIM scores, indicating a very close match to the original images.

Fig. S7. Optimized results for two-polarization imaging system with the incident wavelength at 800 nm. (a) Phase mask. (b) and (c) Phase-based and metasurface-based PSF. (d) and (e) Sensor image by phase mask and metasurface. (f-h) the ground truth, and reconstruction image based on the phase mask and metasurface.
**Fig. S8** Optimized PSF corresponding to the object distance $z = 1 \text{ mm}$ for sixteen-channel imaging system. (a) and (b) the phase-based PSF with LCP and RCP, respectively. (c) and (d) the corresponding metasurface-based PSF with LCP and RCP. The panels in the column from left to right represent the images at the spectral channels of wavelength $\lambda = 470 \text{ nm}$, $\lambda = 545 \text{ nm}$, $\lambda = 648 \text{ nm}$, and $\lambda = 800 \text{ nm}$, respectively.

**Fig. S9** Same plot as Fig. S8, but with the object distance of $z = 2 \text{ mm}$.
Table S9. PSNR of the phase-based and metasurface-based reconstruction image for four-wavelength, two-depth and two-polarization imaging system in Fig. 6.

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Distance</th>
<th>470 nm</th>
<th>545 nm</th>
<th>648 nm</th>
<th>800 nm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>LCP</td>
<td>RCP</td>
<td>LCP</td>
<td>RCP</td>
</tr>
<tr>
<td>Phase mask</td>
<td>1 mm</td>
<td>38.5</td>
<td>39.1</td>
<td>42.0</td>
<td>42.7</td>
</tr>
<tr>
<td></td>
<td>2 mm</td>
<td>37.5</td>
<td>38.8</td>
<td>39.0</td>
<td>37.7</td>
</tr>
<tr>
<td>Metasurface</td>
<td>1 mm</td>
<td>36.0</td>
<td>36.1</td>
<td>39.1</td>
<td>38.8</td>
</tr>
<tr>
<td></td>
<td>2 mm</td>
<td>45.6</td>
<td>47.1</td>
<td>40.1</td>
<td>39.4</td>
</tr>
</tbody>
</table>

Table S10. SSIM of the phase-based and metasurface-based reconstruction image for four-wavelength, two-depth and two-polarization imaging system in Fig. 6.

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Distance</th>
<th>470 nm</th>
<th>545 nm</th>
<th>648 nm</th>
<th>800 nm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>LCP</td>
<td>RCP</td>
<td>LCP</td>
<td>RCP</td>
</tr>
<tr>
<td>Phase mask</td>
<td>1 mm</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>2 mm</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Metasurface</td>
<td>1 mm</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>2 mm</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Section 7: Multi-channel imaging based on a large-scale phase mask

We optimized a four-wavelength imaging system based on a large-scale phase mask. The corresponding pixel number of the object is set as 128×128 and the object distance is z = 2 mm. Here the pixel size and number of phase mask is Λ = 465 nm and 1920×1920, respectively. Since we use the angular spectrum method to calculate the PSF, phase mask has the same size of pixel with the digital PSF (i.e., Λ×Λ), which is too small to match the pixel size of commercial sensor. Therefore, we take each block with 3×3 pixels of PSF as one super-pixel of size ~1.4×1.4 μm² by replacing the intensity value of the super-pixel with the summation of the intensities contributed by the super-pixel containing 3×3 blocks.

Fig. S10 Image reconstruction for four wavelengths. (a) Phase mask designed. (b) Sensor captured data. (c) The combined input images consisting of the Alphabet letters B, G, R, and combined ones IR. (d)-(g) Ground truth scene for the wavelength of 470 nm (B), 545 nm (G), 648 nm (R), 800 nm (IR), respectively. (h)-(k) Reconstructed images corresponding to (d)-(g), respectively.
Figure S10 shows the results of image reconstruction at four channels (wavelengths). Figure S10(a) shows the optimized phase mask, while Fig. S10(b) shows the corresponding response captured on a raw sensor. Figure S10(c) shows the combined input images consisting of the Alphabet letters B (for $\lambda = 470$ nm), G (for $\lambda = 545$ nm), R (for $\lambda = 648$ nm), and combined ones IR (for $\lambda = 800$ nm), respectively. Figures S10(d)–S10(g) show the ground truth, whereas Figures S10(h)–S10(k) show the reconstructed images under the noise standard deviation $\delta = 0.01$. It can be seen that they match very well with each other, which demonstrates the viability and efficiency of the end-to-end inverse design framework.

Figure S11 shows the results of eight-channel image reconstruction for four wavelengths and two depths. Figures S11(a) and S11(b) show the optimized phase profile and sensor image, respectively. Figures S11(c)–S11(j) [Figs. S11(k)–S11(r)] show reconstruction result for the object distance $z = 2$ cm [$z = 3$ cm]. Figures S11(c)–S11(f) [Figs. S11(k)–S11(n)] show the ground truth selected from Fashion-MNIST with $50 \times 50$ pixels, whereas Figs. S11(g)–S11(j) [Figs. S11(o)–S11(r)] show the corresponding reconstructed image for $\lambda = 470$ nm, $545$ nm, $648$ nm, and $800$ nm.

**Fig. S11** Image reconstruction for four wavelengths and two depths. (a) Phase mask designed. (b) Sensor-captured data. (c)-(f) Ground truth selected from Fashion-MNIST for the wavelength of 470 nm, 545 nm, 648 nm and 800 nm, respectively. (g)-(j) Reconstructed images corresponding to (c)-(f) with object distance of 2 cm, respectively. (k)-(r) Same plots as (c)-(j) but with object distance of 3 cm.