

Deep image prior plus sparsity prior: toward single-shot full-Stokes spectropolarimetric imaging with a multiple-order retarder

Feng Han, Tingkui Mu^{✉,*}, Haoyang Li, and Abudusalamu Tuniyazi

Xi'an Jiaotong University, School of Physics, MOE Key Laboratory for Non-equilibrium Synthesis and Modulation of Condensed Matter, Xi'an, China

Abstract. Compressive full-Stokes spectropolarimetric imaging (SPI), integrating passive polarization modulator (PM) into general imaging spectrometer, is powerful enough to capture high-dimensional information via incomplete measurement; a reconstruction algorithm is needed to recover 3D data cube (x , y , and λ) for each Stokes parameter. However, existing PMs usually consist of complex elements and ensnare to accurate polarization calibration, current algorithms suffer from poor imaging quality and are subject to noise perturbation. In this work, we present a single multiple-order retarder followed a polarizer to implement passive spectropolarimetric modulation. After building a unified forward imaging model for SPI, we propose a deep image prior plus sparsity prior algorithm for high-quality reconstruction. The method based on untrained network does not need training data or accurate polarization calibration and can simultaneously reconstruct the 3D data cube and achieve self-calibration. Furthermore, we integrate the simplest PM into our miniature snapshot imaging spectrometer to form a single-shot SPI prototype. Both simulations and experiments verify the feasibility and outperformance of our SPI scheme. It provides a paradigm that allows general spectral imaging systems to become passive full-Stokes SPI systems by integrating the simplest PM without changing their intrinsic mechanism.

Keywords: spectropolarimetry; Stokes parameters; deep image prior; compressive sensing.

Received Mar. 4, 2023; revised manuscript received Apr. 7, 2023; accepted for publication Apr. 13, 2023; published online May 11, 2023.

© The Authors. Published by SPIE and CLP under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.

[DOI: [10.1117/1.APN.2.3.036009](https://doi.org/10.1117/1.APN.2.3.036009)]

1 Introduction

High-dimensional optical information, such as irradiance, spectrum, space, polarization, and phase, are vital for comprehensively noninvasive characterization of targets over diverse scenes.^{1,2} The acquisition of maximal information using a single integrating system is highly desirable in consideration of volume, weight, integration, portability, and cost.³⁻⁸ As shown in Fig. 1, spectropolarimetric imaging (SPI), which integrates a polarization modulator (PM) into an imaging spectrometer, is a kind of such a versatile integrating system,^{5,9-15} which can obtain a spatio-spectrally resolved 3D data cube (x , y , and λ) for each of interested Stokes parameters (S_0 , S_1 , S_2 , and S_3). It has aroused wide applications in the fields of aerosol detection,¹⁶ planetary exploration,¹⁷ remote sensing,¹⁸ biomedical diagnosis,^{19,20} etc.

The design chain of SPI system involves three aspects: the PM, the imaging spectrometer, and a reconstruction algorithm. It is extremely meaningful to design a kind of advanced PM and algorithm that can directly adapt to general imaging spectrometers (multispectral/hyperspectral systems), including scanning (whisk broom, push broom, windowing, and framing)²¹ and snapshot modes.^{6,7}

The PM is a key component that determines the number of measurements in the polarization dimensionality. According to the sampling mechanism, the PMs can be classified into two types^{5,7}: a well-sampled PM and an undersampled PM. For the former, the number of polarization measurements is equal to the number of interested Stokes parameters. In this case, time-sequence active scanning hardware in a single optical path or snapshot passive hardware using multiple parallel optical paths are usually employed.⁵ As a result, the acquisition, storage, and processing of these measurements may result in significant time

*Address all correspondence to Tingkui Mu, tkmu@mail.xjtu.edu.cn

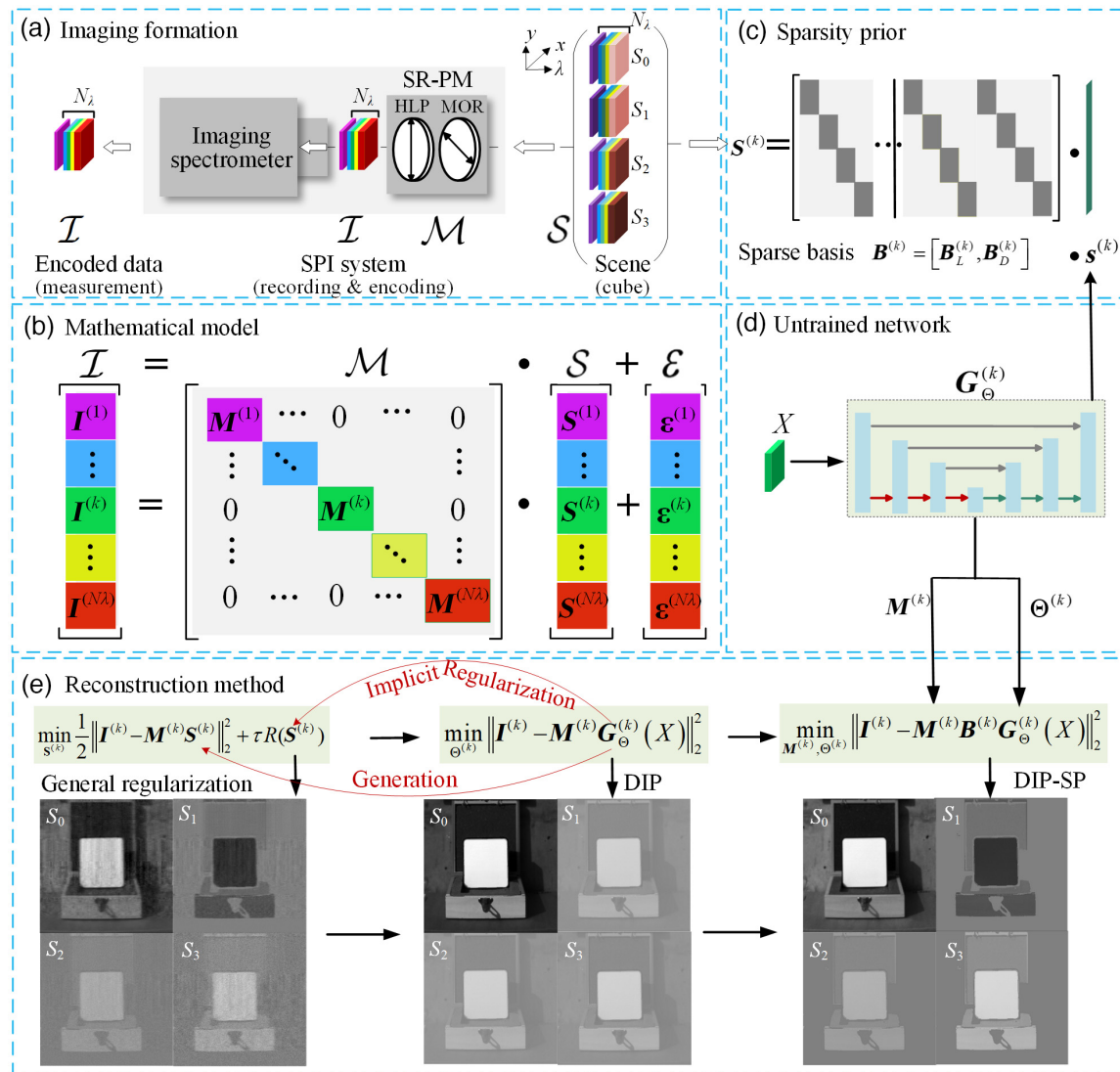


Fig. 1 Schematic of the SPI framework with the passive SR-PM scheme. (a) Imaging formation by integrating the SR-PM with general spectrometer. (b) Forward mathematical model with different colors indicating different spectral bands. (c) Combined sparse representations in transform domain to achieve faster convergence. (d) Untrained network acts as implicit regularization and generator. (e) The reconstruction method is transferred from the CS method with apparent regularization and manual fine-tuning to the unconstrained DIP that uses an untrained network without manual tuning regularization, then to the DIP-SP with the sparse representation constraint and self-calibration ability, which achieves the best performance.

costs, power consumption, space/memory footprint, and possible human resource costs. Meanwhile, designing and building such hardware is expensive and complex.⁷ For the latter, the number of measurements is less than the number of interested Stokes parameters. In this case, part of the burden from hardware is shifted to a postprocessing algorithm by designing the active or passive undersampled PMs and jointly developing corresponding reconstruction algorithms. As a result, the SPI based on the undersampled PM becomes a kind of computational imaging mechanism that includes a hardware encoder plus a software decoder.

As a typical passive undersampled PM scheme, double multiple-order retarders (MORs) followed a polarizer (termed DR-PM) along a single optical path have the advantages of

passive modulation, compact structure, and freedom design space, which have led to many SPI system designs based on existing general imaging spectrometers.^{9–15} Initially, the Fourier transform method (FTM) based on an analytically physical model was developed and then frequently used to extract polarization channels in the Fourier domain of modulated spectra.¹⁰ Although the FTM is a straightforward method, it has some drawbacks, such as channel cross talk, band-limited high-frequency loss, sensitivity to noise, and needing accurate polarization calibration.^{22–26} Later, optimization-based algorithms, such as compressive sensing (CS) with prior knowledge, have been proposed to overcome the above issues.^{27,28} While the high-frequency loss and noise sensitivity have been alleviated, accurate polarization calibration with extra precise elements/setup

is still needed. The rigorous design of an SPI system is thus transferred to the polarization calibration system, which is a change in form but not in content. Meanwhile, we have developed a continuous sliding iterative method (CSIM) without the need for polarization calibration,^{29,30} which can converge well by just letting the initialization be of nominal value. The spectropolarimetric resolution recovered from the CS and CSIM retains the intrinsic spectral resolution of the imaging spectrometer used. This point is very important for the imaging spectrometers that only have modest spectral resolution.^{12,19,31,32} However, the FTM, CS, and CSIM methods are still subject to noise perturbation, since the SPI systems usually have lower flux due to a narrow spectral band and the use of a polarizer. In addition, the FTM and CS methods are enslaved to systematic errors originating from double MORs. Therefore, much labor is spent on the compensation and calibration of the systematic errors such as phase shift, azimuthal misalignment, and temperature effects.^{22–26} Recently, data-driven deep-learning-based techniques have been proposed to enhance reconstruction quality.^{33–35} However, the collection of a large amount of labeled data for kinds of application scenes is challenging. Further, the training data and trained networks are usually system-specific, which is difficult to be extended to other SPI systems. Conversely, model-driven reconstruction methods are still expected to label the data for feeding such data-driven deep-learning-based approaches.

To avoid the above dilemma, it is important to determine whether it is feasible to construct the simplest undersampled full-Stokes PM just using a single MOR followed a polarizer (termed SR-PM) along a single optical path and build novel model-driven method to reconstruct a spatio-spectrally resolved 3D cube for each of four Stokes parameters (S_0 , S_1 , S_2 , and S_3) from incomplete measurement. Although such an SR-PM scheme definitely has the advantages of volume, alignment, assembly, and cost compared to the DR-PM scheme, the reconstruction algorithm crisis would be aggravated in the absence of clarity. Recently, state-of-the-art deep-image prior (DIP) methods using untrained networks have been proposed for image denoise and restoration with provable convergence.^{36–47} Therein, the structure of an untrained network with randomly initialized weights can function as a prior on image statistics without any training, mainly because deep neural networks are good at representing and generating realistic images.^{48–50} Specifically, an untrained network is paired with a physically differentiable forward-imaging model in which the network weights are updated through a loss function comparing the experimental measurement and the generated measurement from the network output passed through the forward imaging model. Therefore, the DIP method is expected to be promising to solve the reconstruction issues of the SPI. However, preliminary applications of the DIP methods on other computational imaging problems^{38–47}, would not account for practical issues in the SPI systems. It was found that if we directly use the DIP method to solve the inverse problem of the SPI, the inversion results tend to be trapped in a pseudo-solution, where all generated images show the similar still detail, but modulated with constants (see Secs. 4 and 5), mainly because the problem in the SPI is more challenging owing to its high dimensionality. This kind of challenge is our motivation to further constrain the DIP by incorporating proper prior knowledge about optical physics in systems or scenes. That is, the physical imaging models of the SPI should be built reasonably, and further regularization

should be introduced for the inverse problem. Hereby, we devise a DIP plus sparsity prior (DIP-SP) method to make the encoding and decoding of the SR-PM scheme become feasible and more efficient. Meanwhile, self-calibration is achieved with the algorithm itself; then accurate polarization calibration is not necessary.

Toward the above-mentioned ends, the novelty of this paper through the encoding hardware to decoding software is three-fold: (1) The principle and forward-imaging model of SPI based on the SR-PM is presented. (2) By incorporating SP into the SPI forward model, a physically constrained DIP-SP method based on an untrained network is developed. In particular, the method does not need any training data and accurate polarization calibration and can simultaneously reconstruct the 3D data cube of each Stokes parameter and achieve the self-calibration of measurement matrix. (3) As an instance, we integrate the SR-PM in front of our miniature snapshot imaging spectrometer to form a single-shot SPI prototype. Both simulation and experimental results show that the proposed DIP-SP algorithms make the SR-PM feasible and more robust to noisy perturbation. The rest of paper is organized as follows. Section 2 presents the SPI forward model. The proposed DIP-SP method is presented in Sec. 3. Extensive simulations and real-world experiments using our miniature SPI prototype are provided in Sec. 4. Section 5 provides a discussion and Sec. 6 provides the conclusion.

2 SPI Forward Model

As shown in Fig. 1(a), a full-Stokes SPI system is built by integrating the SR-PM into a general imaging spectrometer.^{5–7,21} Herein, the SR-PM consists of a fixed MOR following a horizontally linear polarizer (HLP), which is the simplest undersampled full-Stokes PM. The sampling mechanism of the SPI mainly consists of two steps. First, the spatio-spectrally distributed 3D data cube (x , y , and λ) for each of four Stokes parameters (S_0 , S_1 , S_2 , and S_3) are compressively encoded by the SR-PM. Second, the encoded spectropolarimetric data cube is then recorded by the imaging spectrometer used. The working mode of the imaging spectrometer can be whisk broom, push broom, windowing, framing, or snapshot. Either way, the encoded spectropolarimetric data cube should be output for subsequent reconstruction.

Mathematically, a 2D scene can be characterized by a spatio-spectrally distributed 3D data cube (x , y , and λ) for each of the Stokes parameters (S_0 , S_1 , S_2 , and S_3), as shown in the right part of Fig. 1(a). The data cube slices of the k 'th spectral band are expressed as $\{S^{(k)}\}_{k=1}^{N_s} \in \mathbb{R}^{N_x \times N_y \times N_s}$, where N_x , N_y , N_s , and N_s denote the scene height, scene width, the number of spectral bands, and the number of interested Stokes parameters (herein $N_s = 4$), respectively. It is first compressively modulated by the SR-PM with measurement matrix $\{M^{(k)}\}_{k=1}^{N_s} \in \mathbb{R}^{N_x \times N_y \times N_s}$; then the encoded spectropolarimetric slice $\{I^{(k)}\}_{k=1}^{N_s} \in \mathbb{R}^{N_x \times N_y}$ is recorded by the imaging spectrometer as

$$I^{(k)} = \sum_{n=0}^{N_s-1} M_n^{(k)} \odot S_n^{(k)} + e^{(k)}, \quad (1)$$

where $\{e^{(k)}\}_{k=1}^{N_s} \in \mathbb{R}^{N_x \times N_y}$ indicates noise; $M_n^{(k)} = M^{(k)}(:, :, n)$ and $S_n^{(k)} = S^{(k)}(:, :, n)$ denote the n 'th measurement matrix ($n = 0, 1, \dots, N_s - 1$) and corresponding 2D distribution of

Stokes parameter, respectively; and \odot represents the Hadamard (element-wise) product.

As shown in Fig. 1(b), the forward model of SPI at the k 'th spectral band can be linearly expressed as

$$I^{(k)} = M^{(k)}S^{(k)} + \varepsilon^{(k)}, \quad (2)$$

where $\{I^{(k)}\}_{k=1}^{N_\lambda} = \{\text{Vec}(I^{(k)})\}_{k=1}^{N_\lambda} \in \mathbb{R}^{N_x N_y}$ and $\{\varepsilon^{(k)}\}_{k=1}^{N_\lambda} = \{\text{Vec}(\varepsilon^{(k)})\}_{k=1}^{N_\lambda} \in \mathbb{R}^{N_x N_y}$ with $\text{Vec}(\cdot)$ vectorizing the 2D matrix by lexicographically stacking columns. Correspondingly, the vectorization $\{S^{(k)}\}_{k=1}^{N_\lambda} \in \mathbb{R}^{N_x N_y N_s}$ of the whole Stokes parameters $\{S^{(k)}\}_{k=1}^{N_\lambda} \in \mathbb{R}^{N_x \times N_y \times N_s}$ is

$$\begin{aligned} S^{(k)} &= \text{Vec}(S^{(k)}) \\ &= [\text{Vec}(S_0^{(k)})^T, \text{Vec}(S_1^{(k)})^T, \text{Vec}(S_2^{(k)})^T, \text{Vec}(S_3^{(k)})^T]^T, \end{aligned} \quad (3)$$

where T indicates transpose operation. The measurement matrix $M^{(k)} \in \mathbb{R}^{N_x N_y \times N_x N_y N_s}$ is a concatenation of diagonal matrices,

$$\begin{aligned} M^{(k)} &= \frac{1}{2} \left\{ \text{diag}[\text{Vec}(M_0^{(k)})], \text{diag}[\text{Vec}(M_1^{(k)})], \right. \\ &\quad \left. \text{diag}[\text{Vec}(M_2^{(k)})], \text{diag}[\text{Vec}(M_3^{(k)})] \right\}, \end{aligned} \quad (4)$$

where the constant coefficient 1/2 means the flux is halved due to the use of a polarizer. The slice of each spectral band is characterized by similar expressions. Correspondingly, the measurement of all spectral bands is obtained by integrating all slices,

$$\mathcal{I} = \mathcal{M}\mathcal{S} + \mathcal{E}, \quad (5)$$

where

$$\mathcal{I} = [(I^{(1)})^T, \dots, (I^{(k)})^T, \dots, (I^{(N_\lambda)})^T]^T, \quad (6a)$$

$$\mathcal{M} = \text{diag}[M^{(1)}, \dots, M^{(k)}, \dots, M^{(N_\lambda)}], \quad (6b)$$

$$\mathcal{S} = [(S^{(1)})^T, \dots, (S^{(k)})^T, \dots, (S^{(N_\lambda)})^T]^T, \quad (6c)$$

$$\mathcal{E} = [(\varepsilon^{(1)})^T, \dots, (\varepsilon^{(k)})^T, \dots, (\varepsilon^{(N_\lambda)})^T]^T. \quad (6d)$$

As can be seen, the modulation mechanism of SR-PM is characterized by the measurement matrix $M^{(k)}$, which can be derived using Muller–Stokes calculus as (Note 1 in the [Supplementary Material](#))

$$\begin{aligned} M_0^{(k)} &= (q + r), \\ M_1^{(k)} &= (q - r)[\cos^2 2\theta + \sin^2 2\theta \cos \delta^{(k)}], \\ M_2^{(k)} &= (q - r)\sin^2 \frac{\delta^{(k)}}{2} \sin 4\theta, \\ M_3^{(k)} &= -(q - r)\sin 2\theta \sin \delta^{(k)}, \end{aligned} \quad (7)$$

where q and r are the maximum and minimum transmittance that account for the imperfection of the polarizer;

$\delta^{(k)} = 2\pi d[n_e^{(k)} - n_o^{(k)}]/\lambda^{(k)}$; d is the thickness, $\lambda^{(k)}$ is the k 'th wavelength; and $n_e^{(k)}$ and $n_o^{(k)}$ are the extraordinary and ordinary refractive indices of the MOR, respectively. It is observed that the number and complexity of effective arguments depend on the two free parameters (θ, d) . If we set $\theta = 0^\circ/45^\circ/90^\circ/135^\circ$ to simplify the arguments, only partial Stokes parameters could be modulated. In addition, if we let the MOR be an achromatic waveplate for simplification, the system spectral range would be narrowed and unnecessary cost would be induced. To simultaneously modulate all $N_s = 4$ Stokes parameters ($S_0, S_1, S_2,$ and S_3) with generality, we should let $0^\circ < \theta < 180^\circ$ and $\theta \neq 45^\circ/90^\circ/135^\circ$ regardless of the phase retardance $\delta^{(k)}$, although the arguments would be not so clear and simple. The effects of free parameters (θ, d) on the reconstruction quality will be explored in Sec. 5 for the optimal measurement matrix $M^{(k)}$ and its tolerance. For the DR-PM scheme,^{9–15} we also can derive the corresponding measurement matrix, and the following methods also adapt to its modulation.

3 Reconstruction Methods

For the inverse problem of SPI, the measurement does not resemble the reconstructed image; instead, the scene is related to the measurement through the forward-imaging model that describes the physics of the image formation problem. To reconstruct Stokes parameters at different wavelengths from Eq. (5), one may use the pseudo-inverse of measurement matrix \mathcal{M} as a fast method. However, it should be noted that the forward-imaging model in Eqs. (2) or (5) is ill-posed, since the number of unknowns is $N_s = 4$ times that of knowns in the measurement. This means that the measurement is incomplete, and computational reconstruction is needed. As a matter of convenience, it is recommended to independently reconstruct the Stokes parameters at each wavelength. We formulate the inverse problem by creating the mathematical model with an inspiration from the CS. The inverse problem aims to recover the image slices of the four Stokes parameters from the encoded spectropolarimetric slice via proper regularization,

$$\min_{S^{(k)}} \frac{1}{2} \|I^{(k)} - M^{(k)}S^{(k)}\|_2^2 + \tau R(S^{(k)}), \quad (8)$$

where $\|\cdot\|_2$ is the data-fidelity term and $R(S^{(k)})$ is the regularization term with a tuning parameter τ . To obtain the 3D data cube of each Stokes parameter with high quality, generally the measurement matrix $M^{(k)}$ should be calibrated accurately. However, it is hard to solve Eq. (8) directly, even using state-of-the-art solvers [e.g., the gradient projection for sparse reconstruction (GPSR)⁵¹ or the two-step iterative shrinkage/thresholding algorithm (TwIST)]⁵² with elaborately handcrafted regularization such as the sparsity or total variation (TV). Its noise robustness and reconstruction quality are subject to handcrafted regularizations or scene types, as shown in Fig. 1(e). In addition, it is necessary that the measurement matrix $M^{(k)}$ be definitely known using accurate polarization calibration. Thus, it is desirable to develop an advanced algorithm with powerful regularization but without manual tuning of parameters, meanwhile incorporating self-calibration of the measurement matrix $M^{(k)}$ into the algorithm.

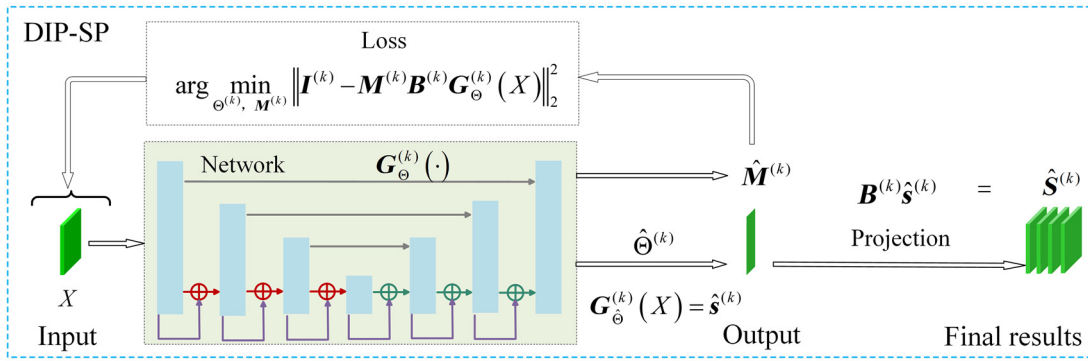


Fig. 2 Processing pipeline of the proposed DIP-SP method.

3.1 DIP Plus SP Method

The state-of-the-art DIP via untrained network has been proposed for the inverse problem of computational imaging with provable convergence^{36,37} and shown to be particularly effective for many applications.^{38–47} In the following sections, we use DIP to solve the SPI inverse problem. As shown in Fig. 1(d), assuming the mapping function $G_{\hat{\Theta}}^{(k)}(\cdot)$ of the untrained network is specified by a set of parameters $\hat{\Theta}^{(k)}$, the inverse problem in Eq. (8) is then replaced as

$$\hat{\Theta}^{(k)} = \arg \min_{\Theta^{(k)}} \|I^{(k)} - M^{(k)} G_{\Theta}^{(k)}(X)\|_2^2, \quad (9)$$

where X indicates the input of the untrained network. It is interesting to note that the function of regularization term $R(S^{(k)})$ in Eq. (8) is replaced by the implicit DIP captured by the untrained network $G_{\hat{\Theta}}^{(k)}(\cdot)$ in the data-fidelity term. With the known measurement matrix $M^{(k)}$, the iterative optimization of Eq. (9) is to find a set of optimal parameters $\hat{\Theta}^{(k)}$ of the untrained network that map the input X to the desired signal as

$$\hat{S}^{(k)} = G_{\hat{\Theta}}^{(k)}(X). \quad (10)$$

As can be seen, the main advantage of the DIP method is that the input is just the X without any need of the ground truth (GT) $S^{(k)}$. However, we found that it is also hard to optimize such preliminary DIP method in Eq. (9) with satisfactory results. The inversion results tend to be trapped in a pseudo-solution, where all generated images show similar detail, but modulated with constants (see Sec. 4.2 and Sec. 5), mainly because the SPI is higher-dimensional and more ill-posed compared to previous applications.^{38–47} In addition, the measurement matrix $M^{(k)}$ in Eq. (9) also needs to be definitely determined before the reconstruction process. These kinds of challenges are our motivation to further constrain the DIP by introducing reasonable regularization.

Since the SP has been shown to be more effective on computational imaging problems,⁵³ it is reasonable to incorporate the SP into the DIP for solving the SPI problem, forming the DIP-SP method. Therefore, as shown in Fig. 1(c), we make a prior assumption that the Stokes parameters $S^{(k)}$ of the 2D scene can be sparsely represented on suitable bases $\{B_n^{(k)}\}_{k=1}^{N_s} \in \mathbb{R}^{N_x N_y \times N_c}$ as

$$S^{(k)} = B^{(k)} s^{(k)}, \quad (11)$$

$$B^{(k)} = \text{diag}[B_0^{(k)}, B_1^{(k)}, \dots, B_{N_s-1}^{(k)}], \quad (12)$$

where $\{s^{(k)}\}_{k=1}^{N_s} \in \mathbb{R}^{N_c N_s}$ denote the bases coefficients and N_c is the number of bases coefficients. The bases can be obtained from some transform domains (e.g., undecimated wavelet transform, discrete Fourier transform, discrete Hartley transform, discrete cosine transform, Legendre polynomials) or dictionary learning. In this work, we select the bases $B_n^{(k)}$ from the combined domains, including discrete cosine transform and Legendre polynomials, to maximize earnings of the SP constraint (Note 2 in the [Supplementary Material](#)). The Legendre polynomials are an orthogonal basis that are useful to model signals such as linear, quadratic, and cubic polynomials. Further, their scale is small relative to other bases, therefore increasing the speed of calculation. The discrete cosine transform coefficients are helpful in capturing sinusoidal variations. The bases coefficients $s^{(k)} = [(s_L^{(k)})^T, (s_D^{(k)})^T]^T$ are from the combination of two bases $B_n^{(k)} = [B_{n,L}^{(k)}, B_{n,D}^{(k)}]$. The number of total bases coefficients N_c is the summation of the Legendre polynomials coefficients L and the discrete cosine transform coefficients $D = N_x N_y$.

Then, as shown in Fig. 2, we incorporate the SP constraint in Eq. (11) into the unconstrained DIP optimization in Eq. (9) as

$$\{\hat{\Theta}^{(k)}, \hat{M}^{(k)}\} = \arg \min_{\Theta^{(k)}, M^{(k)}} \|I^{(k)} - M^{(k)} B^{(k)} G_{\Theta}^{(k)}(X)\|_2^2. \quad (13)$$

As a result, the prior captured by the untrained network is enhanced due to the additional SP constraint, making convergence easier. In addition, it should be pointed out that we let the network parameters $\Theta^{(k)}$ and measurement matrix $M^{(k)}$ be optimized simultaneously. Therefore, it is not necessary to explicitly know the measurement matrix $M^{(k)}$, depending on the accurate polarization calibration. Only a warm initialization of $M^{(k)}$ is needed for convergence and self-calibration (see discussion in Sec. 5). The optimized mapping function $G_{\hat{\Theta}}^{(k)}(\cdot)$ is now used to estimate the bases coefficients $s^{(k)}$,

$$\hat{s}^{(k)} = G_{\hat{\Theta}}^{(k)}(X). \quad (14)$$

The final Stokes parameters $\hat{S}^{(k)}$ are derived using Eq. (10). The input X can be random input with similar dimension^{36–38} or the practical measurement $I^{(k)}$.^{40,41} If the texture features of the measured spectropolarimetric image $I^{(k)}$ resemble the

reconstructed images, it is helpful to directly use the practical measurement $I^{(k)}$ as the input to speed up convergence.³⁹ By allowing other spectral bands undergo a similar optimization process, the whole Stokes parameters at each spectral band can be inverted one by one. Since the DIP-SP method processes all spectral bands individually, it can be applied to monochromatic, multispectral, and hyperspectral SPI schemes.

3.2 Neural Network Framework and Auto-stopping Criterion

For the DIP method, different kinds of neural networks can be employed.^{36,37} When the DIP method was debuted, Ulyanov et al.³⁶ used an overparameterized autoencoder network (i.e., a network with more weights than the number of unknowns). Other preliminary applications also inherit similar overparameterized networks.³⁹⁻⁴⁷ However, it is observed that the iteration process cannot stop automatically, since loss function does not monotonously decrease with the iterations. The iteration must be manually stopped in time when the loss begins to increase. It is critical to stop the iteration at a reasonable convergence point in a timely manner. To avoid early stopping or any other further regularization, an underparameterized network, deep decoder, was proposed by Heckel and Hand,³⁸ due to its resemblance to the decoder part of the autoencoder. The deep decoder does not involve convolutions, and has a concise structure that makes itself an image model with a lower-dimensional description. Since the underparameterized network provides a barrier to overfitting, it can enhance the regularization of inverse problems. However, its representation ability on high-dimensional images with complex details lacks evidence.

To ensure the feasibility of our methods, we elect to use the overparameterized network for performance verification (see discussion in Sec. 5). Specifically, a combination of Unet⁵⁴ and ResNet,⁵⁵ termed as modified Res-Unet⁵⁶ (Note 3 in the [Supplementary Material](#)), is employed, as shown in Fig. 2(b). While the Unet is a neat end-to-end convolutional neural network for multiscale feature extraction from input images,⁵⁴ the Res-Net with skip connection can avoid overfitting.⁵⁵ Therefore, to prevent the over-parameterized Res-Unet from auto-stopping at an early stage, an autostopping criterion (further regularization) must be built. Since the SPI reconstruction is to obtain the spectral images of Stokes parameters, the structural similarity (SSIM) and peak signal-to-noise ratio (PSNR) are two necessary metrics to evaluate the reconstruction quality. Therefore, we build the autostopping criterion by combining the SSIM and PSNR as follows:

$$|\langle \{\text{SSIM}\}_p^{p+q} \rangle - \langle \{\text{SSIM}\}_{p-q}^p \rangle| \leq \varepsilon_1, \quad (15a)$$

$$|\langle \{\text{PSNR}\}_p^{p+q} \rangle - \langle \{\text{PSNR}\}_{p-q}^p \rangle| \leq \varepsilon_2, \quad (15b)$$

where p is the number of iterations, q is the interval of adjacent iterations for average, and ε_1 and ε_2 are two small values for stopping. We let the iteration process stop when both Eqs. (15a) and (15b) are satisfied simultaneously. As the criterion considers the average SSIM and PSNR through a series of adjacent q iterations, it can avoid the unstable alarm from seldom-rising loss points that cannot easily be distinguished by manual intervention.

4 Implementations and Results

4.1 Competitive Algorithms and Evaluation Metrics

In this section, the performance of the proposed DIP-SP in Eq. (13) is verified by extensive simulations and experiments. In addition, the unconstrained DIP method is also executed for comparison. These DIP methods are implemented with Python version 3.6.13 on the PyTorch 1.9.0 platform. The adaptive moment estimation (Adam) optimizer with a learning rate of 0.01 is used to iteratively update the weights and biases. For comparison, we also implement the CS methods in Eq. (8) with TV regularization and with the same SP, respectively. The TwIST⁴⁹ is selected as the solver. For convenience, we term these two algorithms as the TwIST-TV and TwIST-SP, respectively. It is found that, for achieving the best results, the regularization parameter should be manually tuned to $\tau = 0.1$ and the number of iterations in the minimization step in each TwIST iteration be set at 4. All implementations use a machine equipped with AMD Ryzen 7 3700X 8-Core CPU with 16 GB RAM and NVIDIA GeForce GT 1030 GPU.

Two widely quantitative image quality assessment metrics, the PSNR and SSIM, are used to compare the performance. Both of them are spatial measurements, and the higher values indicate spatial reconstruction is good. In addition, the reconstructed spectropolarimetric curves of selected areas are provided for qualitative evaluation. The root mean square error (RMSE) between the GT and the reconstructed curve is used to globally measure the reconstruction quality. The lower the RMSE value, the less the distortion of the reconstructed curve.

4.2 Simulations

4.2.1 Simulation setting

To determine whether the DIP-SP adapts to a scene that has spectropolarimetric curves with different vibration frequencies, we select hyperspectral images from the public data set ICVL⁵⁷ to manually feed polarization information with different vibration frequencies. The original hyperspectral images in the ICVL have the spatial size of 1392×1300 with 361 spectral bands from 400 to 850 nm at a step of 1.25 nm. To generate the controllable polarized spectra with different vibration frequencies, we simply use cosine and sine functions to modulate the spectra S_0 . To reduce computation, we crop out the middle part with the spatial size of 400×400 that include different objects as the test scene, and just extract 100 spectral bands at a step of 4.5 nm for simulations.

The measurement of the encoded spectropolarimetric data cube is generated following the forward-imaging model in Eq. (2). We let the thickness be $d = 1.1$ mm and the fast-axis orientation be $\theta = 22.5^\circ$ (optimal selections are discussed in Sec. 4.1). Then the practical value of measurement matrix $M^{(k)}$ is derived from Eq. (4). For the TwIST-TV, TwIST-SP, and the unconstrained DIP methods, the practical value is necessary and used for reconstruction. However, for the proposed DIP-SP method, the measurement matrix $M^{(k)}$ is the optimization objective, as described in Eq. (13). To validate the robustness of DIP-SP, we, by design, let the initialization of DIP-SP method have a random deviation of a maximum of 3% from the practical value (its tolerance is discussed in Sec. 5). In addition, to mimic practical measurements as much as possible, the additive Gaussian noise of zero mean $\mu = 0$ with low and

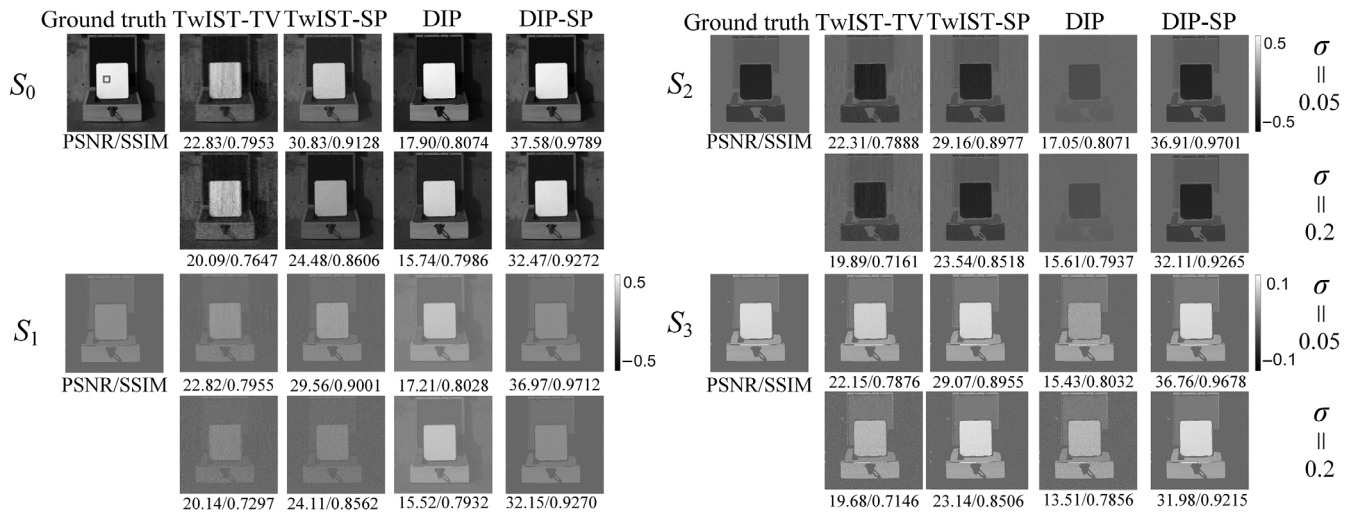


Fig. 3 Simulated results for the reconstructed images of full-Stokes parameters (S_0 , S_1 , S_2 , and S_3) at the spectral band of 550 nm from different algorithms: TwIST-TV, TwIST-SP, DIP, and DIP-SP under the two noise levels ($\sigma = 0.05$ and $\sigma = 0.2$). Average PSNR and SSIM relative to the GT over all 100 spectral bands are presented just below each image.

high standard deviations of $\sigma = 0.05$ and $\sigma = 0.2$, respectively, are added to the normalized measurements for comparison.

4.2.2 Simulation results

Figure 3 shows the reconstructed images of Stokes parameters (S_0 , S_1 , S_2 , and S_3) at the spectral band of 550 nm under the two noise levels ($\sigma = 0.05$ and $\sigma = 0.2$), respectively. It can be seen that for CS methods, while the TwIST-TV performs the worst, the TwIST-SP is sensitive to noise perturbation. For the DIP methods, the unconstrained DIP method is totally invalid. In contrast, the DIP-SP method achieves the best results at any noise level. For the intensity image S_0 , the PSNRs of TwIST-SP and DIP-SP decrease about 6.35 and 5.11 dB, and the SSIM decreases about 0.052 and 0.051, respectively, from the low-noise level ($\sigma = 0.05$) to the high-noise level ($\sigma = 0.2$). For all the polarization images (S_1 , S_2 , and S_3), the PSNRs decrease about 5.5 dB and 4.8 dB, and the SSIM decreases about 0.046 and 0.044, respectively. However, the DIP-SP method still achieves high-quality reconstruction at the high-noise level.

Figure 4 shows the reconstructed average spectral curves of all polarization parameters over a homogeneous area of 5 pixels \times 5 pixels; their average absolute errors and RMSEs relative to the GT are listed in Table 1. It is found that both the TwIST-TV and the unconstrained DIP methods are almost invalid at any noise level. Even though the spectropolarimetric curves of TwIST-SP approximate to the GT just at the low-noise level ($\sigma = 0.05$), they have obvious spectral distortions. Although both the TwIST-SP and DIP-SP methods degrade with the increase of noise, the results of DIP-SP method still conform well to the GT. When the initialization is equal to the practical value of the measurement matrix, the average absolute errors of the reconstructed Stokes parameters from the DIP-SP are 3×10^{-4} at the low-noise level ($\sigma = 0.05$) and 6×10^{-4} at the high-noise level ($\sigma = 0.2$). If the initialization has the random deviation of a maximum of 3%, the average error of reconstruction is less than 0.01, still within the tolerance range, as shown in Table 1. In summary, the DIP-SP method is most robust to the noise perturbation and spectral vibration attribute

to the use of the SP constraint as well as the self-calibration of measurement matrix $M^{(k)}$. More detailed analyses about the influence of the spatial resolution and spectral resolution on the DIP-SP reconstruction are presented in Note 4 in the [Supplementary Material](#).

Figure 5(a) shows the reconstructed images of Stokes parameters (S_0 , S_1 , S_2 , and S_3) just using TwIST-SP and DIP-SP methods over the spectral bands of 450, 550, and 650 nm, respectively, at the low-noise level ($\sigma = 0.05$). As can be seen, the results of DIP-SP preserve more features across the spectral bands, and the results of TwIST-SP have obvious artifacts. The average PSNRs and SSIMs of DIP-SP in each band are higher than those of the TwIST-SP, respectively, by 9.9 dB and 0.078 for the S_0 images, and by 10.52 dB and 0.073 for (S_1 , S_2 , and S_3) images.

Similarly, Fig. 5(b) shows the reconstructed images of Stokes parameters (S_0 , S_1 , S_2 , and S_3) at the high-noise level ($\sigma = 0.2$). The images of TwIST-SP exhibit evident artifacts, especially on the white wall. In contrast, the images of DIP-SP retain clearer details and have the best quality. The PSNR and SSIM of DIP-SP are better than those of the TwIST-SP by 10.95 dB and 0.092 for the S_0 images and by 10.52 dB and 0.091 for (S_1 , S_2 , and S_3) images.

4.3 Experiments

4.3.1 Miniature optically replicating and remapping imaging spectropolarimeter

In this work, we let the imaging spectrometer in Fig. 1(a) be our snapshot ORRIS,⁵⁸ then the SPI system become a snapshot optically replicating and remapping imaging spectropolarimeter (ORRISp).⁵⁹ The passive SR-PM is used to encode the spectropolarimetric images. The ORRIS is used to obtain the spectropolarimetric images encoded by the SR-PM. However, our pioneer prototypes of ORRIS⁵⁸ and ORRISp⁵⁹ are lengthy due to the cascaded imaging objective lens (OL), field stop, and collimating lens. These optical elements are mainly used to prevent adjacent subimages of lenslet array (LA) from overlapping.

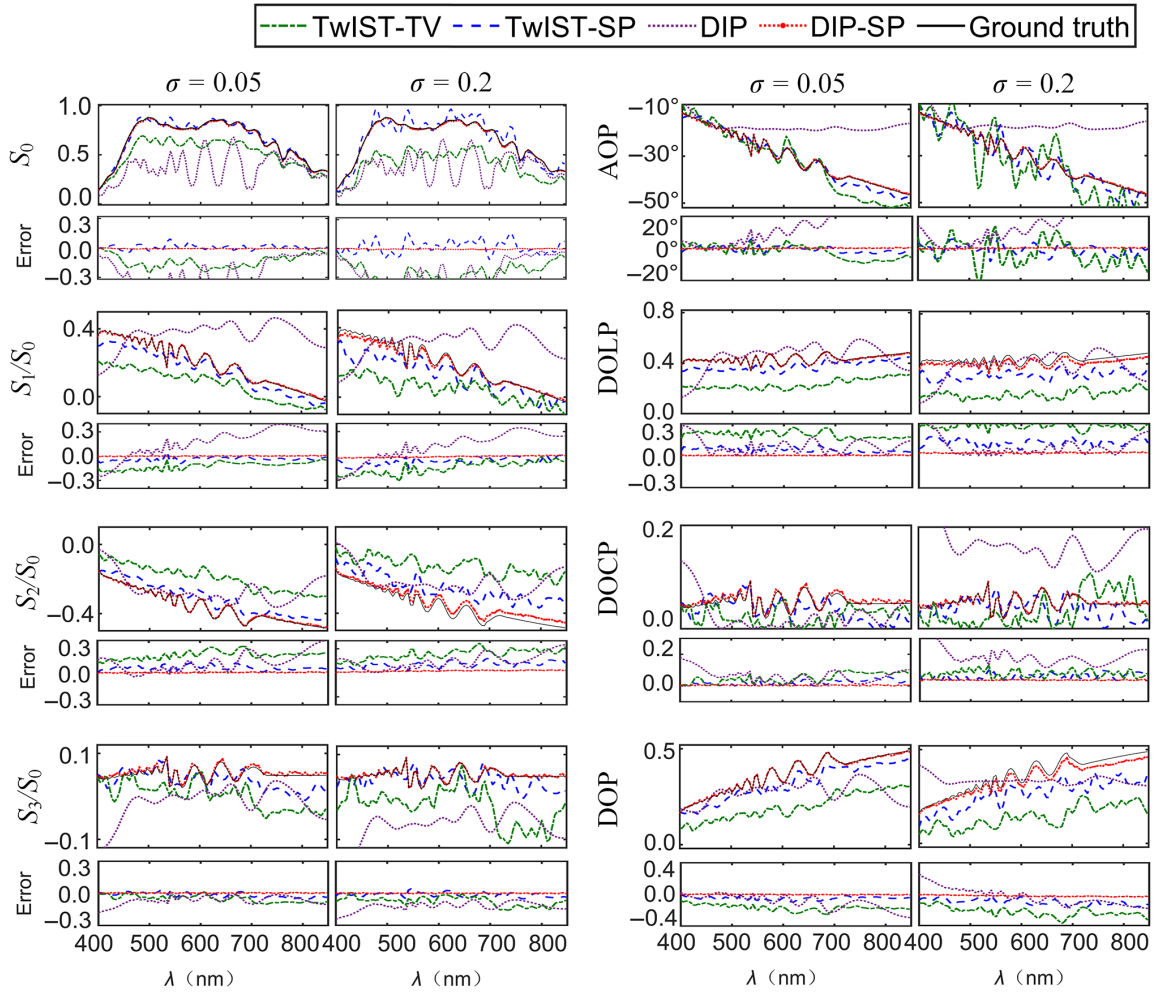


Fig. 4 Simulated results (TwIST-TV, green dashed-dotted line; TwIST-SP, blue dashed line; DIP, purple dotted line; DIP-SP, red star-marked dotted line; GT: black solid line) for the average spectropolarimetric curves and error curves over a homogeneous area of 5 pixels \times 5 pixels. The Stokes parameters (S_0 , S_1/S_0 , S_2/S_0 , and S_3/S_0) are in the left column and derived angle of polarization (AOP), degree of linear polarization (DOLP), degree of circular polarization (DOCP), degree of polarization (DOP) in the right column, respectively.

Table 1 The average absolute errors and RMSEs of spectropolarimetric curves (TwIST-TV, TwIST-SP, DIP, and DIP-SP) relative to the GT.

σ	Average absolute error									RMSE							
	S_0	S_1	S_2	S_3	AOP	DOLP	DOCP	DOP		S_0	S_1	S_2	S_3	AOP	DOLP	DOCP	DOP
TwIST-TV	0.05	0.187	0.203	0.196	0.041	4.5	0.22	0.04	0.22	0.286	0.369	0.363	0.072	6.3	0.32	0.08	0.33
	0.20	0.436	0.245	0.243	0.086	11.2	0.31	0.09	0.32	0.413	0.482	0.480	0.142	17.1	0.46	0.14	0.47
TwIST-SP	0.05	0.034	0.078	0.062	0.024	2.1	0.07	0.02	0.07	0.058	0.134	0.128	0.043	3.1	0.14	0.05	0.15
	0.20	0.081	0.114	0.127	0.059	5.4	0.13	0.04	0.13	0.113	0.252	0.243	0.082	7.4	0.26	0.09	0.27
DIP	0.05	0.302	0.283	0.163	0.162	21.4	0.24	0.17	0.29	0.355	0.422	0.238	0.194	32.3	0.35	0.20	0.40
	0.20	0.486	0.481	0.242	0.263	30.9	0.31	0.27	0.35	0.616	0.580	0.351	0.307	46.2	0.48	0.33	0.58
DIP-SP	0.05	0.003	0.004	0.004	0.002	0.2	0.008	0.002	0.01	0.007	0.014	0.014	0.005	0.8	0.012	0.005	0.01
	0.20	0.004	0.006	0.006	0.003	0.5	0.012	0.003	0.02	0.015	0.026	0.025	0.008	1.3	0.024	0.008	0.03

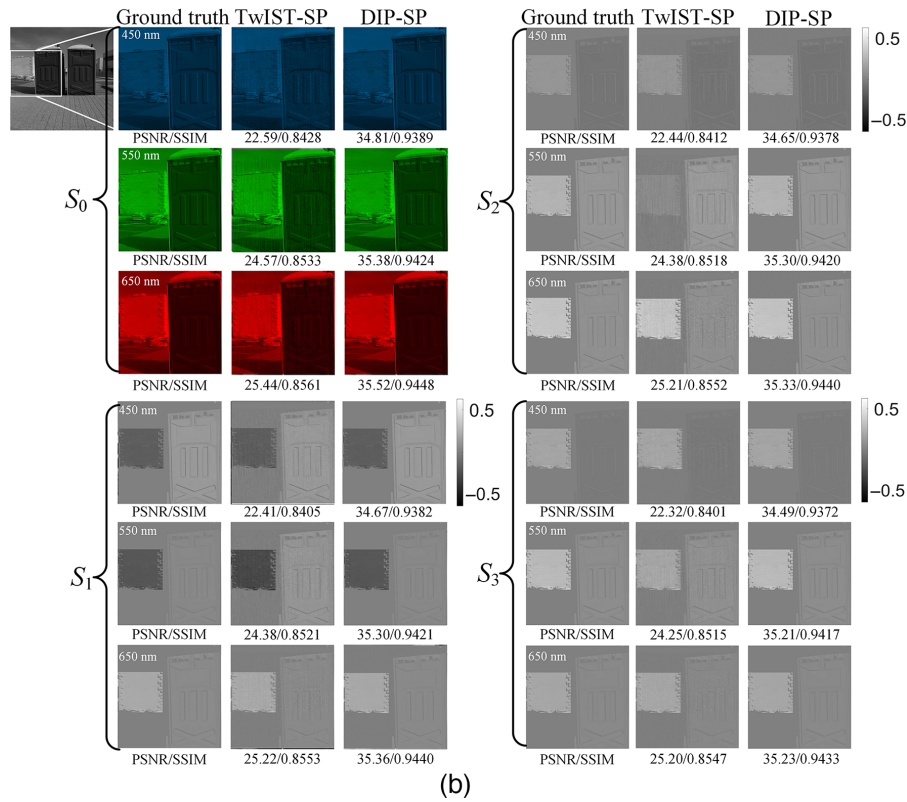
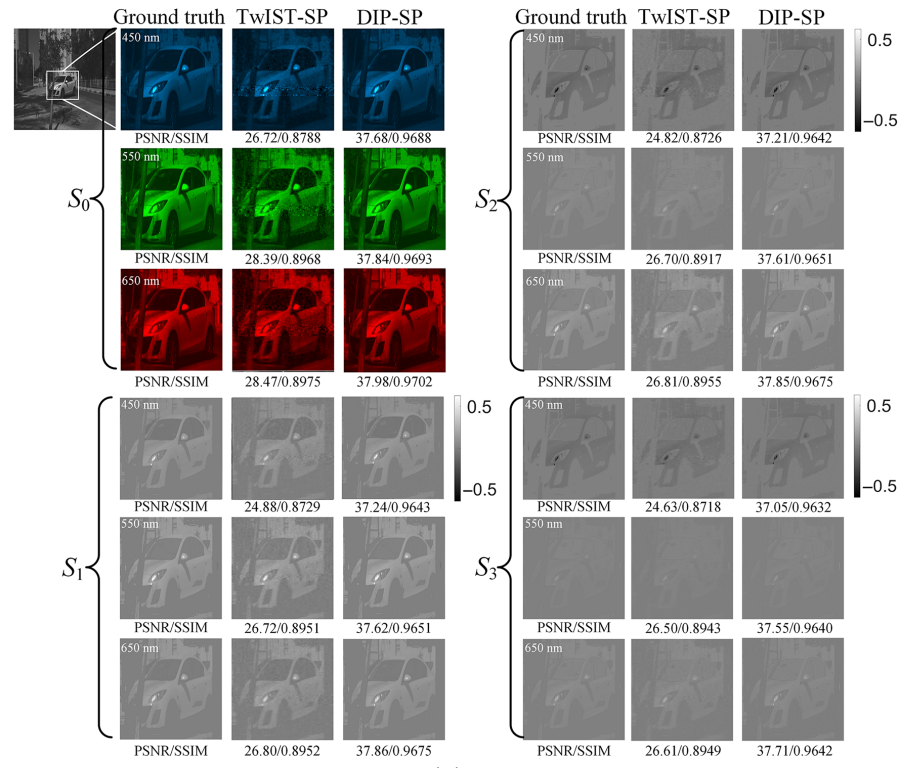


Fig. 5 Simulated results for the reconstructed images of full-Stokes parameters (S_0 , S_1 , S_2 , and S_3) over the spectral bands of 450, 550, and 650 nm from the TwIST-SP to DIP-SP methods, respectively. The PSNRs and SSIMs relative to the GT at each spectral band are provided below each image. (a) Low-noise level ($\sigma = 0.05$) and (b) high-noise level ($\sigma = 0.2$).

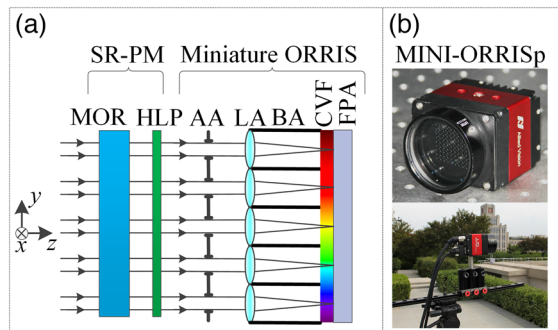


Fig. 6 Scheme of our miniature snapshot ORRISp. (a) Optical scheme and (b) prototype. MOR, multiple-order retarder; HLP, horizontally linear polarizer; AA, aperture array; LA, lenslet array; BA, baffle array; CVF, continuous variable filter; and FPA, focal plane array.

Actually, the size of the subimage can be controlled by inserting a baffle array (BA) between the LA and the focal plane continuous variable filter (CVF), as shown in Fig. 6(a). In this case, the fore optic is removed for miniaturization and then the optical flux is improved, largely due to the reduction of multiple reflection surfaces. In addition, the previous DR-PM with the double MORs is now replaced by the SR-PM with the single MOR for further simplification and easy alignment and assembly. Finally, a miniature ORRISp is developed, which is termed as MINI-ORRISp. Its detailed design, calibration, and testing can be found in reference.⁶⁰ Correspondingly, the working spectral range is extended from 450 to 750 nm for the pioneer ORRISp to 400 to 850 nm for the MINI-ORRISp due to the remove of achromatic quarter-wave plate. More important, the spatio-spectrally resolved 3D data cube (x , y , and λ) of full Stokes parameters (S_0 , S_1 , S_2 , and S_3) can be simultaneously obtained by the MINI-ORRISp prototype, relative to the pioneer ORRISp prototype that just focused on the linear Stokes parameters (S_0 , S_1 , and S_2).

Figure 6(b) shows the final prototype of MINI-ORRISp. The MOR is made from quartz, with a thickness of $d = 1.1$ mm. The HLP is an ultrabroadband wire grid polarizer. They are assembled in the same anodized aluminum lens tube. The LA is 9×9 array with a pitch of 3 mm and a focal length of 9 mm. The row direction of the LA has an inclination angle of 6.34° relative to the CVF's wave band. The size of the CVF is 25 mm \times 30 mm. The center wavelength varies along the 25 mm side in the global y axis. The FPA is a CCD array with the spatial resolution of 3232 pixels \times 4864 pixels and the pixel size of $7.4 \mu\text{m} \times 7.4 \mu\text{m}$. The volume of MINI-ORRISp prototype is largely miniaturized to 80 mm (length) \times 70 mm (width) \times 70 mm (height), and the weight is reduced to 700 g. Within a single exposure time, the MINI-ORRISp can get 81 raw subimages. After registering all subimages to the center one, we can remap out the encoded spectropolarimetric images \mathcal{I} with the spatial resolution of $(N_x = 400) \times (N_y = 400)$ over $N_\lambda = 72$ spectral bands. The spectral resolution at each spectral band is around 3.5% of the central wavelength. For the LA arrangement, two main considerations come into play. The first one is the trade-off between spatial and spectral resolution. The second is how to maximize the utilization of the CVF and effective CCD area while avoiding complex processing. According to the principle of optical replication

and remapping,⁵⁸ the spectral resolution increases with the number of lenses in each row of the LA, but the spatial resolution (N_x and N_y) will decrease due to the limited area of the CCD. Therefore, a reasonable arrangement of the LA must be a compromise between spatial and spectral resolution. Based on the above considerations, the size of the subimage is designed to be 400 (N_x) pixels \times 400 (N_y) pixels, and the LA is designed to be 9×9 . It should be noted that, to maximize the use of the CVF and CCD, a portion of LA on the marginal area is deliberately designed to be outside the effective imaging area. Specifically, the upper half of the first row of the LA and the lower half of the last row of the LA do not cover the CVF, which means these parts do not participate in the replication of subimages and continuous filtering. Therefore, after remapping, the 9×9 LA corresponds to 8×9 spectral bands.

4.3.2 Experimental setting

We first use the MINI-ORRISp to record the encoded spectropolarimetric images \mathcal{I} of a real-world scene in a single shot and then apply both the TwIST-SP and DIP-SP methods to decode a spatio-spectrally resolved 3D data cube of $400(N_x) \times 400(N_y) \times 72(N_\lambda)$ for each of four Stokes parameters (S_0 , S_1 , S_2 , and S_3), respectively. For the two reconstruction methods, the measurement matrix $M^{(k)}$ of system is roughly calibrated using an He-Ne laser and a commercial polarization analyzer (Thorlabs, PAX1000) pointing to several field points of the PM. Other field points and spectral bands are simply interpolated and derived as the initializations. Predictably, the performance of the TwIST-SP method will be degraded further relative to the simulation experiments, in which the practical measurement matrix has been exactly employed for reconstruction. In contrast, the performance of the DIP-SP method will be retained, since the measurement matrix can be self-calibrated during iteration. In addition, a diffuse reflection standard whiteboard is employed as a white reference for each test scene to remove the disturbance of ambient light during acquisition. A monochromatic polarization camera is built and calibrated to obtain the ground truth images of full Stokes parameters. It consists of a high-resolution camera modulated by a rotatable linear polarizer and filtered by a bandpass filter centered at 550 nm with a bandwidth of 10 nm. A standard fiber spectrometer with a fiber probe is used to sample the GT of spectrum S_0 for the selected area. An RGB camera is employed to acquire the GT of color images. We let the polarization camera, RGB camera, and fiber spectrometer work with the best settings to get the data as GTs.

4.3.3 Laboratory scene

For the laboratory scene, the experimental setup is shown in the upper left of Fig. 7. A color-checker covered with linear polarizers and circular polarizers of different polarization directions is used as a test scene in the upper right of Fig. 7. The illumination source is a fluorescent lamp. The MINI-ORRISp, the polarization camera, the RGB camera, and the fiber spectrometer observe the color-checker at the same time. The MINI-ORRISp operates at two exposure time of 150 and 60 ms for noisy comparison. For all measurements, the longer the exposure time, the higher the signal to noise ratio (SNR).

The lower part of Fig. 7 shows the reconstructed Stokes parameter images (S_0 , S_1 , S_2 , and S_3) at 550 nm with the statistic PSNR and SSIM results for each algorithm. Obviously, the reconstruction quality of two methods at the long exposure time is better than that at the short exposure time. While the DIP-SP

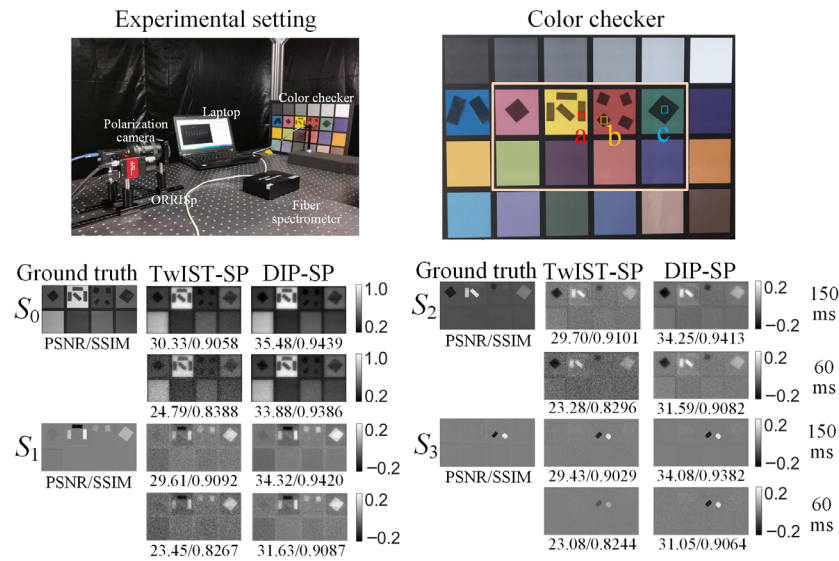


Fig. 7 Lab scene experiment. Experimental setting (top left) and the color-checker covered with different polarizers as test scene (top right). (a) The red-square mark area is the linear polarizer; (b) the yellow-square mark area is the left-circular polarizer; and (c) the azure-square mark area is the linear polarizer. Lower parts are reconstructed Stokes parameters (S_0 , S_1 , S_2 , and S_3) at 550 nm from the TwIST-SP and DIP-SP methods at two exposure time of 60 and 150 ms, respectively.

method achieves better results, the TwIST-SP method is very sensitive to noise perturbation at any exposure time. This phenomenon is consistent with the simulation results in Sec. 4.2.

Figure 8 plots the average spectropolarimetric curves (S_0 , S_1/S_0 , S_2/S_0 , and S_3/S_0) of three selected areas (a, b and c with $5 \text{ pixels} \times 5 \text{ pixels}$) that are shown in the upper right of Fig. 8. Only the S_0 spectra have the GTs that captured using the fiber spectrometer. Clearly, the DIP-SP is champion for the S_0 spectra at each exposure time. In contrast, the TwIST-SP is barely acceptable only after a long exposure time and is futile at the short exposure time. Unsurprisingly, the spectropolarimetric curves (S_1 , S_2 , and S_3) are slowly varying, mainly because the selected areas are almost fully polarized by the covered polarizers. We let the spectropolarimetric curves from the DIP-SP at the long exposure time of 150 ms be references; then the RMSEs of other results are calculated in Table 2. Clearly, the performance of TwIST-SP at any exposure time is worse than that of the DIP-SP at the short exposure time.

4.3.4 Outdoor scene

In the laboratory scene, we have verified the performance of our system and algorithms on a fully polarized scene with different exposure time. However, outdoor natural scenes are totally different due to the lack of a high degree of polarization and the fullness of a complex background. We photographed a car parked on the road on a sunny day, with the exposure time decreased to 50 ms. Figure 9 shows the reconstructed images of Stokes parameters (S_0 , S_1 , S_2 , and S_3) at 480, 550, 600, and 700 nm, respectively. The PSNR and SSIM relative to the GT captured by the polarization camera at 550 nm are calculated in Table 3.

As can be seen, for the TwIST-SP method, the details of images (S_1/S_0 , S_2/S_0 , and S_3/S_0) are blurry in each narrow spectral band caused by noise. The DIP-SP has the better reconstruction quality in all bands. The spectropolarimetric

curves on the vehicle sunroof are plotted. Only the spectra S_0 of the DIP-SP are consistent with the GT that is obtained using the fiber spectrometer. The TwIST-SP result suffers from large vibrations and deviations. Since the sunroof has high reflection, the reflected spectra are partially polarized and slowly varying, as indicated by the spectropolarimetric curves (S_1/S_0 , S_2/S_0 , and S_3/S_0) from the DIP-SP method. The value of each curve at 550 nm complies with the GT from the polarization camera. Since the circular polarization component of nature scene is very weak, both the hardware and software should have good robustness to noise.

To demonstrate the necessity of our MINI-ORRISp as well as the DIP-SP method over previous methods with rotatable polarizers, the application on real-time acquisition of dynamic scenes is presented in Note 5 and Fig. S4 in the [Supplementary Material](#) as well as in [Video 1](#), URL: <https://doi.org/10.1117/1.APN.2.3.036009.s1>.

5 Discussions

5.1 Optimal Free Parameters (θ , d)

For the SR-PM scheme with the single MOR, the fast-axis orientation θ and the thickness d of the MOR are the two free parameters to encode spectra. The reconstruction quality of the DIP-SP depends on these parameters. Using the simulation data in Sec. 4.2, the PSNR and SSIM as functions of the two parameters, respectively, are plotted in Fig. 10. It is found that the reconstruction quality (PSNR and SSIM) slightly decreases with the increase of d and sensitively vibrate with θ . The optimal values of θ are 22.5° , 67.5° , 112.5° , and 157.5° . So, we have set $\theta = 22.5^\circ$ in simulations and real-world experiments. We have selected $d = 1.1 \text{ mm}$ because the reconstruction quality is relatively high and we have a quartz retarder on hand.

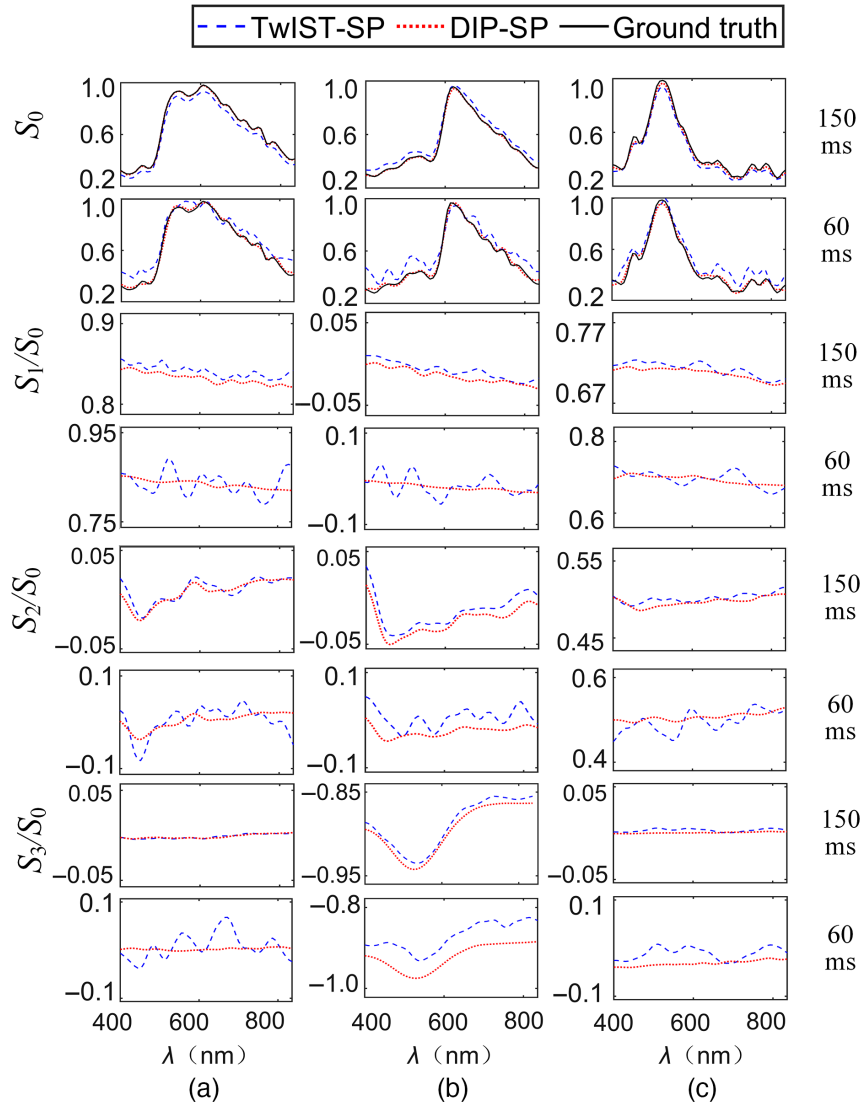


Fig. 8 Lab experimental spectropolarimetric curves (S_0 , S_1/S_0 , S_2/S_0 , and S_3/S_0) of the three selected areas that are shown in the top right of Fig. 7. (a) Red-square mark area; (b) yellow-square mark area; and (c) azure-square mark area.

Table 2 Lab experimental results for the RMSEs of reconstructed spectropolarimetric curves relative to the results from the DIP-SP method at the long exposure time of 150 ms.

	Time	(a)			(b)			(c)		
		S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
TwIST-SP	150 ms	0.015	0.014	0.011	0.015	0.015	0.011	0.013	0.013	0.010
	60 ms	0.055	0.056	0.049	0.056	0.059	0.051	0.051	0.050	0.046
DIP-SP	60 ms	0.008	0.007	0.006	0.009	0.008	0.007	0.008	0.008	0.007

Figure 11 shows the tolerances of the initialization values (d_i, θ_i) from the practical values (d_p, θ_p) , respectively. That is, for the initialization values (d_i, θ_i) within the tolerances, the system can be self-calibrated to the practical value (d_p, θ_p) very well. Obviously, the nominal value $(d_n = 1.1 \text{ mm}, \theta_n = 22.5^\circ)$ is totally embraced in the range of the initialization

values. That means we can directly let the nominal values be the warm initialization values under the corresponding noise levels. However, the tolerances are mainly subject to noise perturbation. The tolerances are maximized to $(0.053 \text{ mm}, 15.19^\circ)$ at $(d_p = 1.12 \text{ mm}, \theta_p = 23^\circ)$ without noise perturbation and they are minimized to $(0.04 \text{ mm}, 5.83^\circ)$ at $(d_p = 1.14 \text{ mm},$

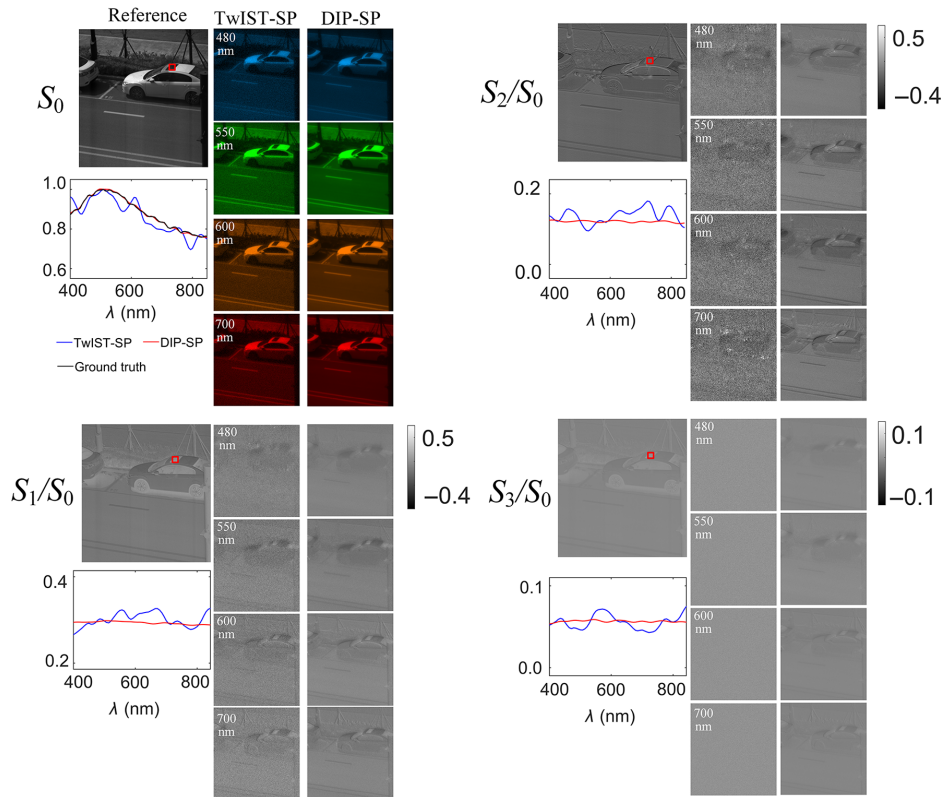


Fig. 9 Outdoor scene experiment. The reconstructed results at the exposure time of 50 ms for the CIE color fusion image S_0 and the gray images (S_1/S_0 , S_2/S_0 , and S_3/S_0) over the four spectral bands of 480, 550, 600, and 700 nm, respectively. The images from the polarization camera at 550 nm are used for the GT. The spectropolarimetric curves (S_0 , S_1/S_0 , S_2/S_0 , and S_3/S_0) of the selected point on the vehicle sunroof are plotted, where only the spectrum S_0 has the GT from the fiber spectrometer.

Table 3 Outdoor experimental results for the average PSNRs and SSIMs at 550 nm of each method.

	PSNR (dB)				SSIM			
	S_0	S_1/S_0	S_2/S_0	S_3/S_0	S_0	S_1/S_0	S_2/S_0	S_3/S_0
TwIST-SP	25.48	25.19	25.16	24.83	0.859	0.852	0.851	0.849
DIP-SP	35.42	34.83	34.60	33.89	0.946	0.942	0.941	0.935

$\theta_p = 25^\circ$) under the highest noise levels. The tolerance on the thickness is tighter and more sensitive to noise than that on the fast-axis orientation, mainly because even a small error on the thickness d will incur considerable error on the retardance δ of the thick retarder. To let the DIP-SP method adapt to general error and noise cases, it is suggested to roughly calibrate the measurement matrix $M^{(k)}$ with elementary setup and operation. In our above experimental results, the measurement matrix $M^{(k)}$ has been roughly calibrated using an He-Ne laser and a commercial polarization analyzer (Thorlabs, PAX1000) pointing to several field points of the PM. Other field points and spectral bands are simply interpolated and derived as the initializations. In the future, it will be necessary to maximize the tolerance by designing advanced regularization or network and then save the rough calibration procedure.

5.2 Overparameterized or Underparameterized Networks

The network frameworks may influence the reconstruction performance. In this section, we compare the performance of an overparameterized network (Res-UNet⁵⁶) and an underparameterized network (deep decoder³⁷) using the simulated scene in Fig. 3. Figure 12 shows the reconstructed PSNR and SSIM of the two networks over the number of iterations, as well as the reconstructed S_0 images at 550 nm every 500 iterations. It is found that, without the SP constraint, the reconstruction results of the overparameterized and underparameterized networks tend to be trapped in a pseudo-solution. In this case, as expected, the convergence of overparameterized network first reaches a higher value, then drops to a lower value. Although the convergence of the underparameterized network does not drop in

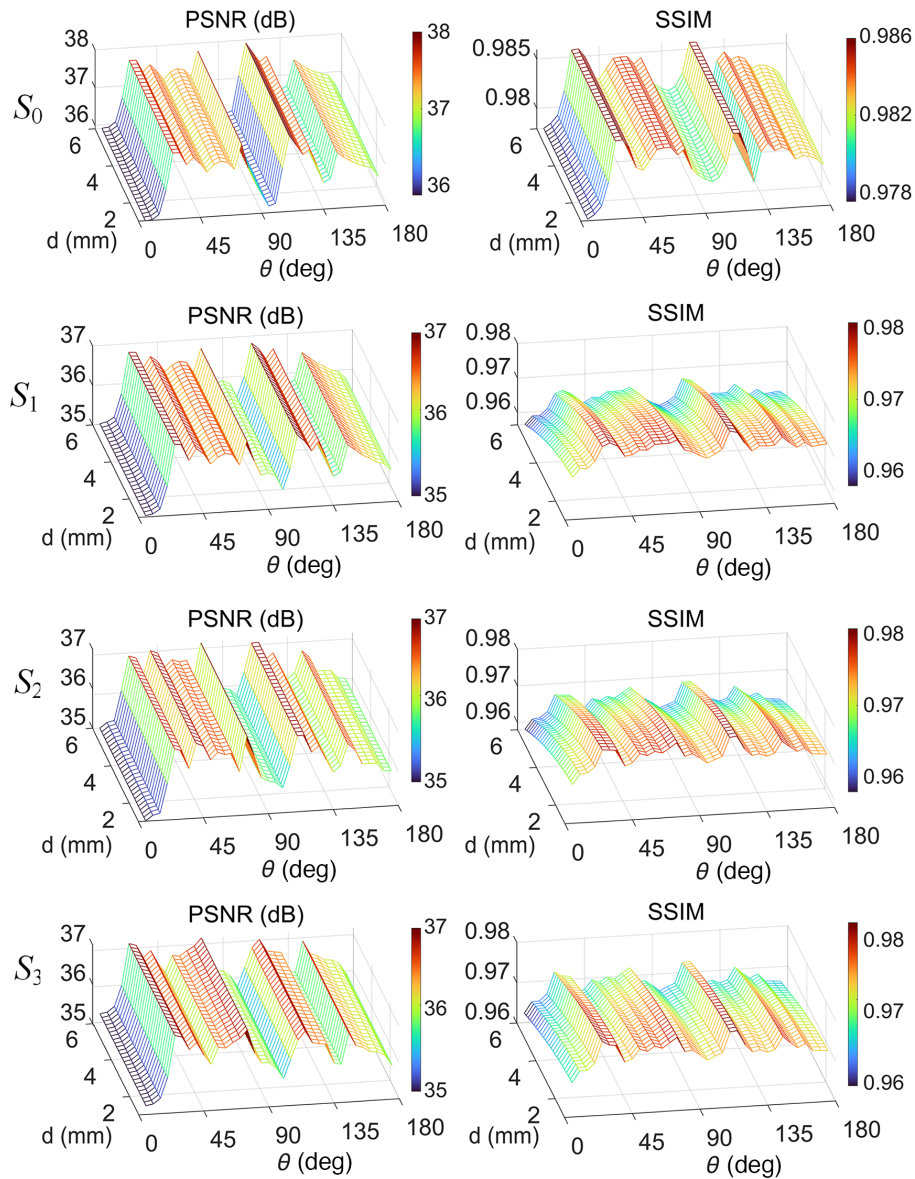


Fig. 10 Dependence of reconstruction quality individually on the fast-axis orientation θ and the thickness d of the MOR in the SR-PM scheme using the DIP-SP method.

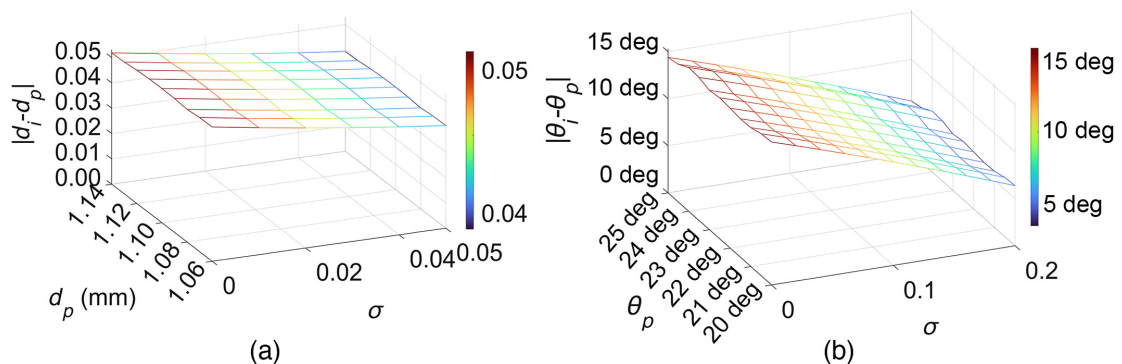


Fig. 11 (a) and (b) The tolerances of the initialization values (d_i, θ_i) from the practical values (d_p, θ_p) of retarder, respectively, under different noise levels.

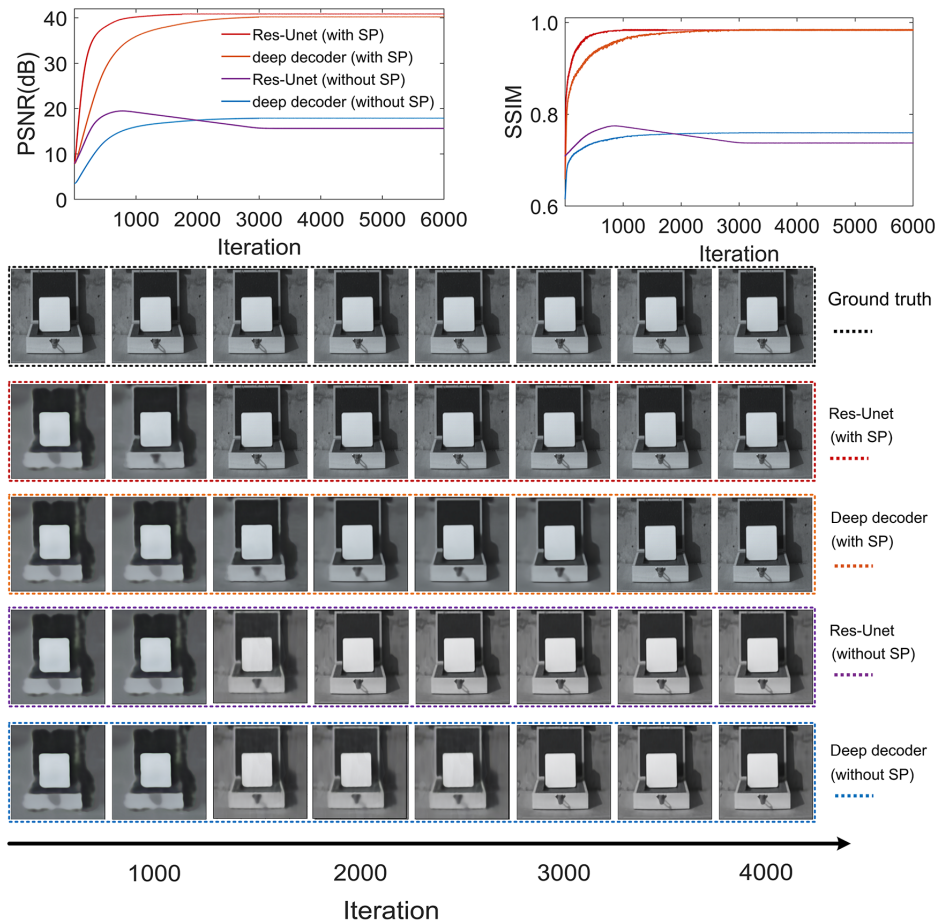


Fig. 12 History of PSNR and SSIM values of the reconstruction results by the overparameterized network (Res-Net) and underparameterized network (deep decoder) with respect to the iterations. The inset represents the S_0 image every 500 iterations (the upper and lower rows are from the underparameterized and overparameterized network, respectively).

expectation, the reconstructed quality is limited to the lower value. That is, both of them cannot fit the image well. In contrast, with the help of the SP constraint and self-calibration operation, the reconstructed PSNR and SSIM of the two networks can approach the best results. In this case, both the overparameterized and underparameterized networks can converge very well, and there are no overfitting phenomena, even with a large number of iterations. It is interesting to find that the underparameterized network needs more than 3000 iterations to converge, and the restored S_0 images become clear at around 3500 iterations. In contrast, the overparameterized network starts to converge just after 1000 iterations, and the restored S_0 images become definite just after 1500 iterations. That is, the overparameterized network (Res-Net) converges faster than the underparameterized network (deep decoder) for our SPI problem, perhaps because the DIP representation ability of the underparameterized network on high-dimensional images is not sufficient.

5.3 Auto-stopping Parameters

For the DIP-SP method, the iterations can be stopped automatically according to the auto-stopping criterion in Eq. (15). As shown in Fig. 12, both the PSNR and SSIM approach to stably

smooth values after 1000 iterations for the over-parameterized network with the SP constraint. To let the DIP-SP automatically stop at the optimal iterations of 1500 nearby, we have set the parameters $q = 5$, $\epsilon_1 = 10^{-4}$, and $\epsilon_2 = 10^{-3}$ in Eq. (15), respectively. These parameters have been used commonly in simulations and real-world experiments. As a result, the optimization at each spectral band experiences about 3000 iterations for the lab data and around 5800 iterations for the outdoor data in Sec. 4.

5.4 Acceleration Strategy

As shown in Sec. 3.1, the forward-imaging model and inversion are mainly illustrated at the k 'th spectral band for simplification. In practice, although we can input the images of all spectral bands simultaneously into the network according to Eq. (5), this would increase the calculation amount and require mass memory. To reduce computation cost, the images can be spatially split into a series of small image patches. But an evident seam would appear by stitching the reconstructed image patches to nonuniform reconstructions.

Hereby, we recommended processing the images of all spectral bands one by one via a similar procedure. As a result, the DIP-SP method reasonably adapts to hyperspectral, multispectral, and monochromatic imaging mechanisms. However, if the

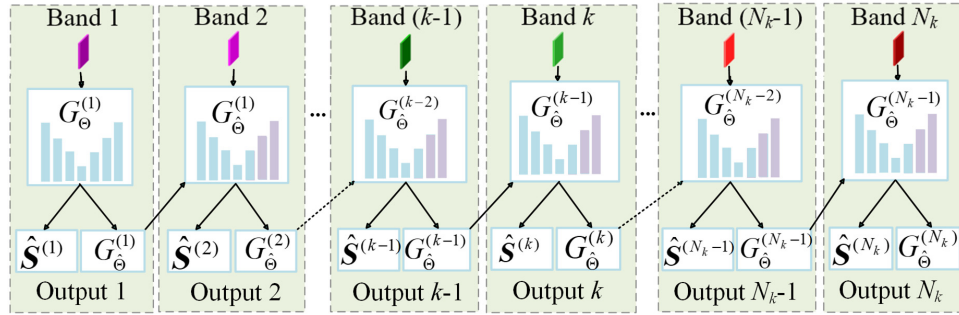


Fig. 13 Acceleration strategy of the DIP-SP processing the images of all spectral bands one by one.

Table 4 Iteration statistics for the simulated and experimental data using the speed-up DIP-SP.

	First Spectral Band		Subsequent Spectral Band		Total time (s)
	Iterations	Iteration time (s)	Iterations	Iteration time (s)	
Simulated data	1500	40	400	0.40	64
Lab data	3000	61	500	0.51	96
Outdoor data	5800	87	650	0.64	133

iteration time of the first spectral band is t , the total time would be $(N_k \times t)$, which is the accumulation of N_k spectral bands. To speed up convergence and reduce iteration time, an acceleration pipeline is proposed in Fig. 13. It is assumed that the images of adjacent bands share similar semantic features. Then, the optimized mapping function $G_{\hat{\Theta}}^{(k-1)}(\cdot)$ for the $(k-1)$ 'th band is considered the prior initial mapping function $G_{\hat{\Theta}}^{(k)}(\cdot)$ for the k 'th band. In addition, it is found that just allowing the last two layers of the optimized network $G_{\hat{\Theta}}^{(k-1)}(\cdot)$ to participate in training for the k 'th band is enough to further reduce iterations. The iteration statistics for the simulated and experimental data are listed in Table 4. As can be seen, the first spectral band needs full iterations; subsequent spectral bands just need fewer iterations. As a result, the total iteration time is reduced around 45 times relative to $(N_k \times t)$.

However, the total reconstruction time of the DIP-SP method remains a stumbling block for real-time reconstruction, although we have devised the acceleration strategy. The iterations would increase linearly to the spatial resolution and spectral bands because each spectral band is processed individually. Interesting topics for future work include developing more efficient acceleration methods. While lightweight networks⁶¹ with higher fitting ability are desirable, optical neural networks may be a possible way for such a fast reconstruction task.^{62,63}

6 Conclusions

The real-time acquisition of a spatio-spectrally resolved 3D data cube (x, y, λ) for each of four Stokes parameters is challenged. In this paper, we have proposed the simplest passive SR-PM with a single MOR for the full-Stokes SPI with compressed sampling for the first time. As a result, the alignment and assembly errors from the DR-PM with double MORs are avoided, and there is a reduction in system bulk and cost. The SR-PM can be easily incorporated into general imaging spectrometers. As an instance, we have successfully integrated

the SR-PM into our snapshot ORRIS to form the MINI-ORRISp prototype. After building the forward-imaging model and incorporating an *a priori* assumption of sparse representation, we have developed a novel advanced reconstruction method, DIP-SP, based on the untrained network. The DIP-SP can achieve the self-calibration of measurement matrix and is very robust to noise perturbation. Extensive results on simulation and real-world data captured by the MINI-ORRISp have verified the outperformance of our full-Stokes SPI scheme.

It is worth pointing out the advantages of the DIP-SP method. First, it makes the encoding and decoding of the simplest passive SR-PM scheme become feasible. Second, the GT does not appear in the inverse problem, and the optimization process is only driven by the measurement at each spectral band. That is, the DIP-SP does not need an additional data set for training and has the potential capability to face diverse applications. Third, the DIP-SP removes the time-consuming and interaction-intensive search for a scene-dependent regularization parameter that is used in the TwIST-based methods. Fourth, since the forward-imaging model is used for optimization, the output of the neural network is physically constrained and is interpretable. Last, we do not need to accurately calibrate the measurement matrix; roughly preliminary polarization calibration is acceptable. The nominal value can serve as a warm initialization if the noise level is not too high. Furthermore, the DIP-SP method also can be applied to the DR-PM schemes by modifying the arguments of measurement matrix $M^{(k)}$ in Eq. (7). These advantages are helpful for promoting the development of a miniature SPI system using free-space optical components^{12,31,32,59} as well as chip-scale integration with silicon photonic circuits⁶⁴ or metasurfaces.⁶⁵⁻⁷²

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 62175196), the Shaanxi Fundamental Science Research Project for Mathematics and

Physics (Grant No. 22JSY020), and the Shaanxi Province Key Research and Development Program of China (Grant No. 2021GXLH-Z-058).

References

1. M. T. Eismann, *Hyperspectral Remote Sensing*, SPIE Press (2012).
2. Y. Zhao et al., *Multi-band Polarization Imaging and Applications*, Springer & National Defense Industry Press (2016).
3. J. Wang et al., “Hybrid reconstruction of the physical model with the deep learning that improves structured illumination microscopy,” *Adv. Photonics Nexus* **2**(1), 016012 (2023).
4. Z. Zhang et al., “Ten-mega-pixel snapshot compressive imaging with a hybrid coded aperture,” *Photonics Res.* **9**, 2277–2287 (2021).
5. J. S. Tyo et al., “Review of passive imaging polarimetry for remote sensing applications,” *Appl. Opt.* **45**(22), 5453–5469 (2006).
6. N. Hagen and M. W. Kudenov, “Review of snapshot spectral imaging technologies,” *Opt. Eng.* **52**, 090901 (2013).
7. S. Takizawa and K. Hiramatsu, “High-speed hyperspectral imaging enabled by compressed sensing in time domain,” *Adv. Photonics Nexus* **2**, 026008 (2023).
8. T. Mu et al., “Optimized design, calibration, and validation of an achromatic snapshot full-Stokes imaging polarimeter,” *Opt. Express* **27**, 23009–23028 (2019).
9. K. H. Nordsieck, “A simple polarimetric system for the lick observatory image-tube scanner,” *Pub. Astron. Soc. Pac.* **86**, 324–329 (1974).
10. K. Oka and T. Kato, “Spectroscopic polarimetry with a channeled spectrum,” *Opt. Lett.* **24**, 1475–1477 (1999).
11. F. J. Lannarilli et al., “Polarimetric-spectral intensity modulation (P-SIM): enabling simultaneous hyperspectral and polarimetric imaging,” *Proc. SPIE* **3698**, 474–481 (1999).
12. D. S. Sabatke et al., “Snapshot imaging spectropolarimeter,” *Opt. Eng.* **41**, 1048–1054 (2002).
13. F. Snik, T. Karalidi, and C. U. Keller, “Spectral modulation for full linear polarimetry,” *Appl. Opt.* **48**, 1337–1346 (2009).
14. J. Craven-Jones et al., “Infrared hyperspectral imaging polarimeter using birefringent prisms,” *Appl. Opt.* **50**, 1170–1185 (2011).
15. C. Zhang et al., “High throughput static channeled interference imaging spectropolarimeter based on a Savart polariscope,” *Opt. Express* **24**, 23314–23332 (2016).
16. G. van Harten et al., “Atmospheric aerosol characterization with a ground-based SPEX spectropolarimetric instrument,” *Atmos. Meas. Tech.* **7**, 4341–4351 (2014).
17. T. Karalidi et al., “Observing the Earth as an exoplanet with LOUPE, the lunar observatory for unresolved polarimetry of Earth,” *Planet. Space Sci.* **74**, 202–207 (2012).
18. F. Snik et al., “FlySPEX: a flexible multi-angle spectropolarimetric sensing system,” *Proc. SPIE* **9853**, 985308 (2016).
19. A. Tuniyazi et al., “Snapshot polarized light scattering spectroscopy using spectrally-modulated polarimetry for early gastric cancer detection,” *J. Biophotonics* **14**, e202100140 (2021).
20. A. Tuniyazi et al., “Bayesian neural networks enabled snapshot polarized light scattering spectroscopy with uncertainty quantification,” *Opt. Lasers Eng.* **160**, 107298 (2023).
21. R. G. Sellar and G. D. Boreman, “Classification of imaging spectrometers for remote sensing applications,” *Opt. Eng.* **44**(1), 013602 (2005).
22. A. Taniguchi et al., “Stabilization of a channeled spectropolarimeter by self-calibration,” *Opt. Lett.* **31**, 3279–3281 (2006).
23. M. W. Kudenov et al., “Fourier transform channeled spectropolarimetry in the MWIR,” *Opt. Express* **15**, 12792–12805 (2007).
24. J. Craven-Jones and M. W. Kudenov, “False signature reduction in channeled spectropolarimetry,” *Opt. Eng.* **49**, 053602 (2010).
25. T. Mu et al., “Alignment and retardance errors, and compensation of a channeled spectropolarimeter,” *Opt. Commun.* **294**, 88–95 (2013).
26. X. Ju et al., “Easily implemented approach for the calibration of alignment and retardation errors in a channeled spectropolarimeter,” *Appl. Opt.* **57**, 8600–8613 (2018).
27. D. J. Lee, C. F. LaCasse, and J. M. Craven, “Compressed channeled spectropolarimetry,” *Opt. Express* **25**(25), 32041–32063 (2017).
28. G. Zhou, Y. Li, and K. Liu, “Adaptive linear reconstruction with regularizer for channeled spectropolarimeter,” *Opt. Eng.* **60**, 064111 (2021).
29. F. Han et al., “Iterative reconstruction for general linear imaging polarimetry without polarimetric calibration,” *Opt. Lett.* **45**, 57–60 (2020).
30. F. Han et al., “Iterative reconstruction for snapshot intensity-modulated linear imaging spectropolarimetry without Fourier transform and phase calibration,” *Opt. Lasers Eng.* **134**, 106286 (2020).
31. Y. Qiao et al., “Intensity modulation-based spectral polarization measurement method of coded aperture,” *Opt. Commun.* **437**, 128–132 (2019).
32. X. Lv et al., “Lenslet-based snapshot full-Stokes imaging spectropolarimeter,” *Opt. Laser Tech.* **148**, 107758 (2022).
33. Q. Li et al., “Channeled imaging spectropolarimeter reconstruction by neural networks,” *Opt. Lett.* **46**(17), 4394–4397 (2021).
34. X. B. Lv et al., “Spectral-temporal channeled spectropolarimetry using deep-learning-based adaptive filtering,” *Opt. Express* **29**(22), 35556–35569 (2021).
35. C. Huang et al., “Convolutional neural network-based spectrum reconstruction solver for channeled spectropolarimeter,” *Opt. Express* **30**, 10367–10386 (2022).
36. D. Ulyanov, A. Vedaldi, and V. Lempitsky, “Deep image prior,” in *Conf. Comput. Vision and Pattern Recognit. (CVPR)*, IEEE, pp. 9446–9454 (2018).
37. R. Heckel and P. Hand, “Deep decoder: concise image representations from untrained non-convolutional networks,” in *Int. Conf. Learn. Represent. (ICLR)* (2019).
38. E. Bostan et al., “Deep phase decoder: self-calibrating phase microscopy with an untrained deep neural network,” *Optica* **7**(6), 559–562 (2020).
39. F. Wang et al., “Phase imaging with an untrained neural network,” *Light: Sci. Appl.* **9**, 77 (2020).
40. F. Yang et al., “Robust phase unwrapping via deep image prior for quantitative phase imaging,” *IEEE Trans. Image Process.* **30**, 7025–7037 (2021).
41. X. Zhang, F. Wang, and G. Situ, “BlindNet: an untrained learning approach toward computational imaging with model uncertainty,” *J. Phys. D: Appl. Phys.* **55**, 034001 (2022).
42. K. C. Zhou and R. Horstmeyer, “Diffraction tomography with a deep image prior,” *Opt. Express* **28**(9), 12872–12896 (2020).
43. F. Niknam, H. Qazvini, and H. Latifi, “Holographic optical field recovery using a regularized untrained deep decoder network,” *Sci. Rep.* **11**, 10903 (2021).
44. Y. Zhu et al., “Full scene underwater imaging with polarization and an untrained network,” *Opt. Express* **29**, 41865–41881 (2021).
45. K. Monakhova et al., “Untrained networks for compressive lensless photography,” *Opt. Express* **29**, 20913–20929 (2021).
46. H. Yu et al., “Untrained deep learning-based fringe projection profilometry,” *APL Photonics* **7**, 016102 (2022).
47. F. Wang et al., “Far-field super-resolution ghost imaging with a deep neural network constraint,” *Light: Sci. Appl.* **11**, 1 (2022).
48. Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature* **521**, 436–444 (2015).
49. G. Barbastathis, A. Ozcan, and G. Situ, “On the use of deep learning for computational imaging,” *Optica* **6**, 921–943 (2019).
50. C. Zuo et al., “Deep learning in optical metrology: a review,” *Light: Sci. Appl.* **11**, 39 (2022).
51. M. Figueiredo, R. Nowak, and S. Wright, “Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems,” *IEEE J. Sel. Top. Signal Process.* **1**(4), 586–597 (2007).

52. J. Bioucas-Dias and M. Figueiredo, "A new TwIST: two-step iterative shrinkage/thresholding for image restoration," *IEEE Trans. Image Process.* **16**(12), 2992–3004 (2007).
 53. J. N. Mait, G. W. Euliss, and R. A. Athale, "Computational imaging," *Adv. Opt. Photonics* **10**, 409–483 (2018).
 54. O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci.* **9351**, 234–241 (2015).
 55. K. He et al., "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, IEEE, pp. 770–778 (2016).
 56. K. Wang et al., "Deep learning spatial phase unwrapping: a comparative review," *Adv. Photonics Nexus* **1**(1), 014001 (2022).
 57. B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural RGB images," in *Eur. Conf. Comput. Vision* (2016), pp. 19–34.
 58. T. Mu et al., "Compact snapshot optically replicating and remapping imaging spectrometer (ORRIS) using a focal plane continuous variable filter," *Opt. Lett.* **44**, 1281–1284 (2019).
 59. T. Mu et al., "Snapshot hyperspectral imaging polarimetry with full spectropolarimetric resolution," *Opt. Laser. Eng.* **148**, 106767 (2022).
 60. H. Li et al., "Miniature snapshot optically replicating and remapping imaging spectropolarimeter (MINI-ORRISp): design, calibration and performance," prepared (2023).
 61. H. Chen et al., "Fourier Imager Network (FIN): a deep neural network for hologram reconstruction with superior external generalization," *Light: Sci. Appl.* **11**, 254 (2022).
 62. G. Wetzstein et al., "Inference in artificial intelligence with deep optics and photonics," *Nature* **588**, 39–47 (2020).
 63. F. Hoch et al., "Characterization of multimode linear optical networks," *Adv. Photonics Nexus* **2**(1), 016007 (2023).
 64. Z. Lin et al., "Chip-scale full-Stokes spectropolarimeter in silicon photonic circuits," *Photonics Res.* **8**, 864–874 (2020).
 65. T. Li et al., "Integrating the optical tweezers and spanner onto an individual single-layer metasurface," *Photonics Res.* **9**, 1062–1068 (2021).
 66. Y. Ni et al., "Computational spectropolarimetry with a tunable liquid crystal metasurface," *eLight* **2**, 23 (2022).
 67. X. Hua et al., "Ultra-compact snapshot spectral light-field imaging," *Nat. Commun.* **13**, 2732 (2022).
 68. D. Yang et al., "Multiwavelength high-order optical vortex detection and demultiplexing coding using a metasurface," *Adv. Photonics Nexus* **1**(1), 016005 (2022).
 69. T. Li et al., "Reversible lateral optical force on phase-gradient metasurfaces for full control of metavehicles," *Opt. Lett.* **48**, 255–258 (2023).
 70. K. Qu et al., "Deep-learning-assisted inverse design of dual-spin/frequency metasurface for quad-channel off-axis vortices multiplexing," *Adv. Photonics Nexus* **2**(1), 016010 (2023).
 71. S. Ebel et al., "Optical reflective metasurfaces based on mirror-coupled slot antennas," *Adv. Photonics Nexus* **2**(1), 016005 (2023).
 72. S. Lung et al., "Real-time monitoring of polarization state deviations with dielectric metasurfaces," *Adv. Photonics Nexus* **2**(2), 026003 (2023).
- Feng Han** received his BS degree in physics from Xi'an Jiaotong University, Xi'an, Shaanxi, China, in 2018 and his MS degree in physics from Xi'an Jiaotong University in 2021. Currently, he is pursuing a PhD at the School of Physics of Xi'an Jiaotong University. His research interests include spectral imaging, polarimetric imaging, spectropolarimetric imaging, and computational imaging.
- Tingkui Mu** received his PhD in physics from Xi'an Jiaotong University, Xi'an, Shaanxi, China, in 2012. From 2014 to 2016, he was working as a postdoctoral research fellow at the College of Optical Sciences of The University of Arizona, Tucson, United States. Currently, he is working as a professor at the School of Physics of Xi'an Jiaotong University. His research interests include advanced optical imaging and sensing, spectral and polarimetric imaging techniques and sensors, atmospheric sounding, and computer vision. He has published more than 80 reviewed papers and applied for more than 20 patents.
- Haoyang Li** received his BS degree in applied physics from Qingdao University of Science and Technology, China, in 2016. He is a post-graduate student at Xi'an Jiaotong University, China. His research interests include polarimetric imaging, hyperspectral imaging, and spectropolarimetric imaging.
- Abudusalamu Tuniyazi** received his BS degree in physics from Tongji University, Shanghai, China, in 2012 and his MS degree in optical engineering from Tongji University, Shanghai, China, in 2015. Currently, he is pursuing a PhD in Xi'an Jiaotong University, Xi'an, Shaanxi, China. His research interests include biological and medical hyperspectral imaging, polarization and spectral imaging, and computational imaging.