

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

Feng Han, Tingkui Mu*, Haoyang Li, and Abudusalamu Tuniyazi

MOE Key Laboratory for Nonequilibrium Synthesis and Modulation of Condensed Matter, School of Physics, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, P.R. China

*tkmu@mail.xjtu.edu.cn

Supplementary Note 1:

The SR-PM consists of a multiple-order retarder (MOR) followed by a horizontally linear polarizer (HLP). The Mueller matrices of HLP and MOR are expressed as:

$$M_p(0^\circ) = \frac{1}{2} \begin{bmatrix} q+r & (q-r) & 0 & 0 \\ (q-r) & (q+r) & 0 & 0 \\ 0 & 0 & 2\sqrt{qr} & 0 \\ 0 & 0 & 0 & 2\sqrt{qr} \end{bmatrix}, \quad (S1)$$

$$M_R^{(k)}(\theta, \delta^{(k)}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos^2 2\theta + \sin^2 2\theta \cos \delta^{(k)} & [1 - \cos \delta^{(k)}] \sin 2\theta \cos 2\theta & -\sin 2\theta \sin \delta^{(k)} \\ 0 & [1 - \cos \delta^{(k)}] \sin 2\theta \cos 2\theta & \sin^2 2\theta + \cos^2 2\theta \cos \delta^{(k)} & \cos 2\theta \sin \delta^{(k)} \\ 0 & \sin 2\theta \sin \delta^{(k)} & -\cos 2\theta \sin \delta^{(k)} & \cos \delta^{(k)} \end{bmatrix}, \quad (S2)$$

where q and r are the maximum and minimum transmittance of the HLP, $\delta^{(k)} = 2\pi d[n_e^{(k)} - n_o^{(k)}] / \lambda^{(k)}$, d is the thicknesses, n_e and n_o are the extraordinary and ordinary refractive indices of the MOR, respectively. According to the Muller-Stokes calculus, the output Stokes vector $S_{\text{out}}^{(k)}$ of the SR-PM is

$$S_{\text{out}}^{(k)} = M_p(0^\circ) M_R^{(k)}(\theta, \delta^{(k)}) S_{\text{in}}^{(k)}. \quad (S3)$$

We only need carry out the multiplication of the first row of the Mueller matrix with the input Stokes vector because the imaging spectrometer only measures the output intensity signal $I^{(k)} = S_{\text{out},0}^{(k)}$. Thus, the measured intensity is

$$I^{(k)} = \frac{1}{2} \left[M_0^{(k)} S_{\text{in},0}^{(k)} + M_1^{(k)} S_{\text{in},1}^{(k)} + M_2^{(k)} S_{\text{in},2}^{(k)} + M_3^{(k)} S_{\text{in},3}^{(k)} \right], \quad (S4)$$

where

$$M_0^{(k)} = (q+r), \quad (S5a)$$

$$M_1^{(k)} = (q-r) [\cos^2 2\theta + \sin^2 2\theta \cos \delta^{(k)}], \quad (S5b)$$

$$M_2^{(k)} = (q-r) \sin^2 \frac{\delta^{(k)}}{2} \sin 4\theta, \quad (S5c)$$

$$M_3^{(k)} = -(q-r) \sin 2\theta \sin \delta^{(k)}. \quad (\text{S5d})$$

are the arguments of measurement matrix $\mathbf{M}^{(k)}$. where q and r are the maximum and minimum transmittance that account for imperfection of the polarizer, $\delta^{(k)} = 2\pi d[n_e^{(k)} - n_o^{(k)}] / \lambda^{(k)}$, d is the thicknesses, $\lambda^{(k)}$ is k -th wavelength, $n_e^{(k)}$ and $n_o^{(k)}$ are the extraordinary and ordinary refractive indices of the MOR, respectively.

Supplementary Note 2:

Legendre polynomials basis is expressed as:

$$B_{n,L}^{(k)} = \begin{bmatrix} P_1(x_1) & \cdots & P_L(x_1) \\ \vdots & \ddots & \vdots \\ P_1(x_{N_x N_y}) & \cdots & P_L(x_{N_x N_y}) \end{bmatrix}, \quad (\text{S6})$$

where L is the number of Legendre polynomials which is signal-dependent; $x_1, \dots, x_{N_x N_y}$ uniformly sample the interval $[-1,1]$. Correspondingly, the Legendre polynomial $P_m(x)$ is

$$P_m(x) = 2^m \sum_{l=0}^m x^l \binom{m}{l} \binom{m+l-1}{m}. \quad (\text{S7})$$

Discrete cosine transform matrix with (n_x, n_y) th entry is expressed as:

$$B_{n,D}^{(k)}(n_x, n_y) = \begin{cases} \sqrt{\frac{2}{D}} \cos\left(\frac{\pi}{2D}(2n_y - 1)(n_x - 1)\right), & \text{for } n_x = 2, \dots, D \\ \sqrt{\frac{1}{D}}, & \text{for } n_x = 1 \end{cases}. \quad (\text{S8})$$

We combine Legendre polynomials with discrete cosine transform bases into sparse matrices, and the total bases coefficients N_c is the summation of the Legendre polynomials coefficients L and the discrete cosine transform coefficients $D = N_x N_y$,

$$B_n^{(k)} = \begin{bmatrix} B_{n,L}^{(k)} & B_{n,D}^{(k)} \end{bmatrix}. \quad (\text{S9})$$

Supplementary Note 3:

The modified Res-Unet is a neat end-to-end convolutional neural network (CNN) architecture for multiscale feature extraction from input images [56]. As shown in Fig. S1, it consists of two parts, a downsampling encoder that takes encoded spectropolarimetric images as input, and an upsampling decoder that outputs the decoded Stokes vector for each spectral band. In the encoder part there are five encoder blocks, and each block consists of two convolutional operations (3×3 convolutional + batch normalization + ReLU) followed by a 2×2 max pooling operation for downsampling. A residual block connects the two convolutions. After each encoder block, we double the number of feature channels. Other two similar convolutional operations in the valley connect the encoder part and the decoder part. Symmetrically, in the decoder part has five decoder blocks, and each block consists of a 2×2 up-convolution operation for upsampling followed by the two similar convolutional operations. After each decoder block the number of feature channels are halved. The feature channels in the encoder blocks are concatenated to the corresponding decoder blocks using skip connections in the middle. Finally, the last additional 1×1 convolutional layer followed a sigmoid activation function outputs the reconstructed result.

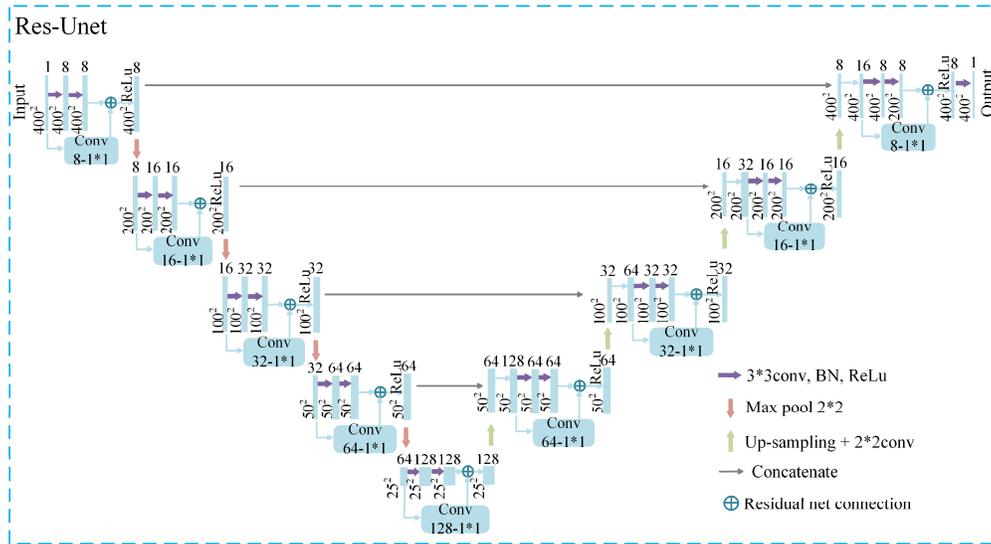


Fig. S1. The Res-UNet framework.

Supplementary Note 4:

In this note, we present more detailed analyses about the influence of the spatial resolution and spectral resolution on the DIP-SP reconstruction. Figure S2 shows the reconstructed images of Stokes parameters (S_0 , S_1 , S_2 , S_3) at the spectral band of 550 nm under different spatial and spectral resolutions, respectively, at the low-noise level ($\sigma = 0.05$). The method of TwIST-SP and DIP-SP perform slightly when reducing spectral resolution or spatial resolution, and the DIP-SP method still achieves high-quality reconstruction. When the spectral resolution is reduced from $400(N_x) \times 400(N_y) \times 100(N_\lambda)$ to $400(N_x) \times 400(N_y) \times 20(N_\lambda)$, for the intensity image S_0 , the PSNRs of TwIST-SP and DIP-SP decrease about 0.37dB and 0.06dB, and the SSIM reduces about 0.001 and 0.001, respectively. For all the polarization images (S_1 , S_2 , S_3), the PSNRs decrease about 0.51dB and 0.02dB, and the SSIM reduces about 0.036 and 0.001, respectively. It shows that the decrease of spectral resolution has low effect on the DIP-SP. When the spatial resolution is reduced from $400(N_x) \times 400(N_y) \times 100(N_\lambda)$ to $100(N_x) \times 100(N_y) \times 100(N_\lambda)$, for the intensity image S_0 , the PSNRs of TwIST-SP and DIP-SP decrease about 0.98dB and 0.74dB, and the SSIM reduces about 0.005 and 0.004, respectively. For all the polarization images (S_1 , S_2 , S_3), the PSNRs decrease about 1.21dB and 0.73dB, and the SSIM reduces about 0.008 and 0.006, respectively. As seen, the DIP-SP is slightly sensitive to the spatial resolution relative to the spectral resolution. This mainly because neural networks are good at representing and generating images, the decrease in spatial resolution expectably affect the fit ability of untrained network. However, the performance of the DIP-SP is still the best and acceptable.

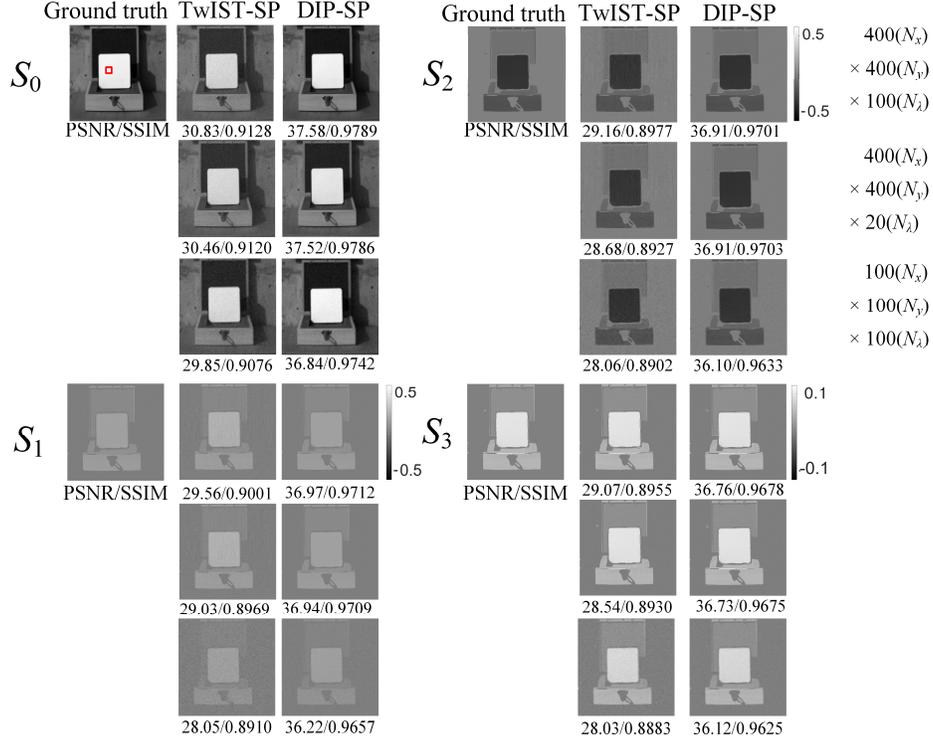


Fig. S2. Simulated results for the reconstructed images of full Stokes parameters (S_0 , S_1 , S_2 , S_3) at the spectral band of 550 nm from different algorithms: TwIST-SP and DIP-SP at the low noise level ($\sigma = 0.05$). For each Stokes parameter, the spatio-spectral resolution of the first row is $400(N_x) \times 400(N_y) \times 100(N_\lambda)$, the second row is $400(N_x) \times 400(N_y) \times 20(N_\lambda)$, and the third row is $100(N_x) \times 100(N_y) \times 100(N_\lambda)$. Average PSNR and SSIM relative to the Ground truth over all spectral bands are presented just below each image.

Figure S3 shows the reconstructed average spectral curves of all polarization parameters over a homogeneous area of 5×5 pixels when the spatio-spectral resolution is $400(N_x) \times 400(N_y) \times 20(N_\lambda)$, and their average absolute errors and RMSEs relative to the ground truth (GT) are listed in Tab. S1. Relative to the result in Fig. 4 with the spatio-spectral resolution is $400(N_x) \times 400(N_y) \times 100(N_\lambda)$, the performances of the TwIST-SP and DIP-SP methods degrade with the decrease of spectral resolution, although the DIP-SP still performs the best. For the DIP-SP method, the average absolute errors of the reconstructed Stokes parameters by the DIP-SP increase from 3×10^{-4} to 4×10^{-4} at the low-noise level ($\sigma = 0.05$) and from 6×10^{-4} to 9×10^{-4} at the high-noise level ($\sigma = 0.2$). This is mainly due to the increase of the spectral interval among adjacent bands. As a result, the initial values from the former band will not be adaptable to the next band, making it susceptible to noise perturbation.

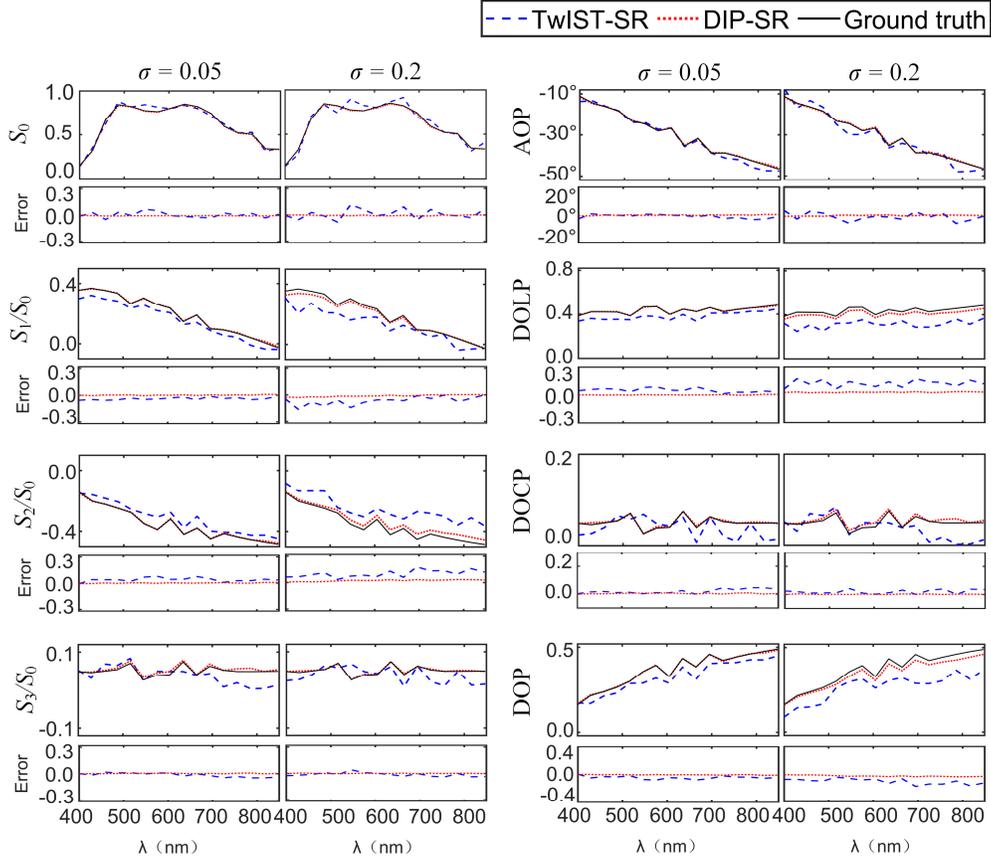


Fig. S3. Simulated results (TwIST-SP: blue dashed line, DIP-SP: red dotted line, Ground truth: black solid line) for the average spectropolarimetric curves and error curves over a homogeneous area of 5×5 pixels at the two noise levels. The spatio-spectral resolution of datacube is $400(N_x) \times 400(N_y) \times 20(N_z)$. The Stokes parameters (S_0 , S_1/S_0 , S_2/S_0 , S_3/S_0) are listed in left column and the derived AOP, DOLP, DOCP, DOP are displayed in right column, respectively.

Table S1. The average absolute errors and RMSEs of spectropolarimetric curves (TwIST-SP and DIP-SP) relative to the Ground truth.

	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-SP	0.05	0.035	0.080	0.063	0.025	2.1	0.08	0.02	0.08	0.060	0.134	0.131	0.049	3.2	0.14	0.05	0.14
	0.20	0.083	0.119	0.128	0.060	5.5	0.16	0.04	0.16	0.118	0.254	0.246	0.091	7.5	0.25	0.09	0.26
DIP-SP	0.05	0.004	0.005	0.005	0.003	0.3	0.009	0.003	0.01	0.008	0.015	0.015	0.006	0.8	0.013	0.006	0.02
	0.20	0.009	0.009	0.009	0.006	0.6	0.015	0.005	0.02	0.020	0.029	0.029	0.011	1.3	0.029	0.010	0.03

Supplementary Note 5:

To demonstrate the necessity of our MINI-ORRISp as well as DIP-SP method over previous methods with rotatable polarizers, the application on real time acquisition of dynamic scenes is presented in this note. Herein, we use the MINI-ORRISp with the integration time of 200 ms to record a low-light lab dynamic scene at cloudy afternoon. Since the data volume is large, just the DIP-SP is used for reconstruction with the iterations of 8600 for the first frame according to the auto-stopping criteria. Figure S4 shows a frame of dynamic scene including a man moving through the lab hallway. The CIE color fusion image S_0 from all bands and the gray images ($S_1/S_0, S_2/S_0, S_3/S_0$) at the four selected bands of 480nm, 550nm, 600nm and 700nm are presented, as well as the spectropolarimetric curves from the cellphone screen and the sun glasses, respectively. The derived images AOP, DOLP and DOCP are also provided. It is clearly to found that although the sun glasses have dark image S_0 and spectrum S_0 , their polarization information is high, as a result the sun glasses on the images ($S_1/S_0, S_2/S_0$) become obvious. Similar phenomena also appeared on the watch panel on left forearm. In contrast, the cellphone screen not only has high polarization information (even on the image S_3/S_0) but also has bright spectrum which complies with the ground truth from the fiber spectrometer. The cellphone screen and sun glasses have uniform AOP and DOLP, which show definitely contrast relative to the spectral images S_0 . The average DOLPs of the glasses and screen are close to 0.25 and 0.48 respectively. The images S_3/S_0 and DOCP are weak in all spectral bands, but they still can reveal the cell phone screen from the background. As seen, our system as well as algorithm have the capability of capturing the objects located at low-light level.

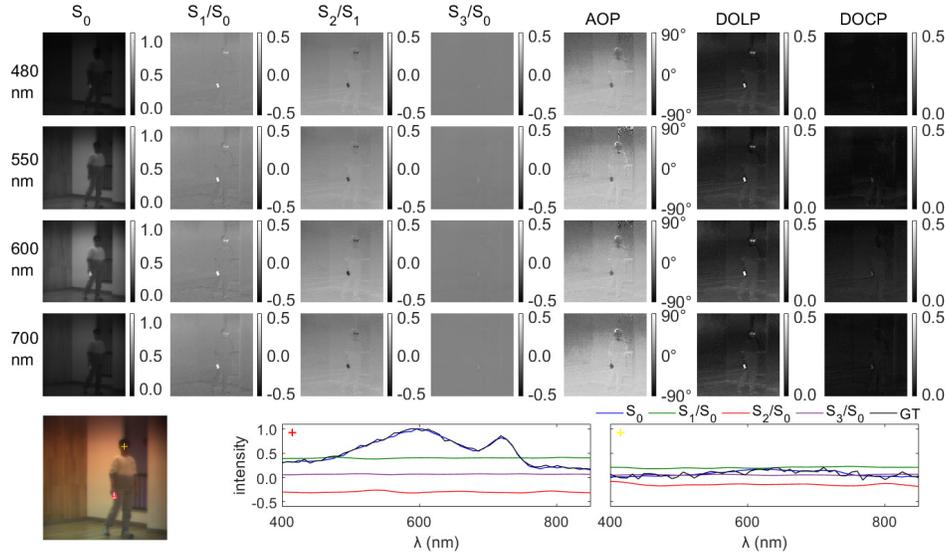


Fig. S4. Lab dynamic scene results with the integration time of 200 ms at the four spectral bands of 480 nm, 550 nm, 600 nm and 700 nm (see [Visualization 1](#)). The first through fourth column are the images of ($S_0, S_1/S_0, S_2/S_0, S_3/S_0$), respectively. The fifth through seventh column are the images of AOP, DOLP, and DOCP, respectively. The bottom left is the CIE color fusion image S_0 , where the cell phone screen is marked with red cross and the sun glasses with yellow cross; the bottom right is the spectropolarimetric curves of two marked areas.