



TSMPN-PSI: high-performance polarization scattering imaging based on three-stage multi-pipeline networks

XUEQIANG FAN, BING LIN, KAI GUO, BINGYI LIU, AND ZHONGYI GUO^{*}

School of Computer and Information, Hefei University of Technology, Hefei 230009, China
^{*}guozhongyi@hfut.edu.cn

Abstract: Polarization imaging, which provides multidimensional information beyond traditional intensity imaging, has prominent advantages for complex imaging tasks, particularly in scattering environments. By introducing deep learning (DL) into computational imaging and sensing, polarization scattering imaging (PSI) has obtained impressive progresses, however, it remains a challenging but long-standing puzzle due to the fact that scattering medium can result in significant degradation of the object information. Herein, we explore the relationship between multiple polarization feature learning strategy and the PSI performances, and propose a new multi-polarization driven multi-pipeline (MPDMP) framework to extract rich hierarchical representations from multiple independent polarization feature maps. Based on the MPDMP framework, we introduce a well-designed three-stage multi-pipeline networks (TSMPN) architecture to achieve the PSI, named TSMPN-PSI. The proposed TSMPN-PSI comprises three stages: pre-processing polarization image for de-speckling, multiple polarization feature learning, and target information reconstruction. Furthermore, we establish a real-world polarization scattering imaging system under active light illumination to acquire a dataset of real-life scenarios for training the model. Both qualitative and quantitative experimental results show that the proposed TSMPN-PSI achieves higher generalization performance than other methods on three testing data sets refer to imaging distances, target structures, and target materials and their background materials. We believe that our work presents a new framework for the PSI and paves the way to its pragmatic applications.

© 2023 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

Accomplishing high-quality imaging in scattering environments is a long-standing optical imaging problem of great importance. It plays a vital role in a wide range of applications across various domains including but not limited to remote sensing observation, medical imaging, and autonomous driving [1–4]. However, optical scattering imaging is susceptible to interference from various types of scattering media, such as fog [5], haze [6–8], biological tissues [9], and turbid water [10,11]. Specifically, when light carrying target information passes through a scattering medium, the target information will be seriously degraded by the medium, leading to poor imaging quality [12]. Tremendous efforts in optical theories and experiments, such as wavefront shaping [13], transmission matrices [14], optical coherence tomography [15], and correlated imaging [16–19], have been made to reconstruct target information in scattering media. Nevertheless, the optical methods possess inherent limitations, such as being time-consuming, labor-intensive, and expensive. Facing the significance of optical scattering imaging and the difficulty in experimentally reconstructing target signal, it is highly desired to develop an automated and cost-effective computational approach for achieving high-performance imaging in scattering environments.

In recent years, the significance of polarization imaging has been increasingly recognized by many researchers [20–24]. Unlike traditional intensity imaging methods, polarization imaging not only provides the distribution of light intensity in a scene, but also the distribution of polarization feature, such as the angle of polarization (AoP) and degree of polarization (DoP). By leveraging the multi-dimensional characteristics of the polarization information, several computational imaging methods aided by polarization-information have emerged for solving target reconstruction in scattering media. The current polarization scattering imaging (PSI) methods can be roughly classified into two categories: (i) physical model-based methods and (ii) deep learning (DL)-based methods. Physical model-based methods lead the trend in the field of PSI at the early stage. These methods including dark channel prior (DCP) [7], maximum intensity prior (MIP) [25], inverse imaging with the underwater image formation model (IFM) [26], and polarization difference imaging (PDI) [10], mainly rely on hand-crafted statistics and knowledge of the physical characteristics of the scene to obtain a suitable solution for target information recovery. For example, Liu *et al.* [27] use the DCP algorithm and logarithmic transformation to restore polarization-based underwater image; Zhao *et al.* [28] employ genetic algorithm to scan the degree of linear polarization (DoLP) image of the target light and the backscattered light to obtain the target light images with the highest contrast. It is undeniable that the physical model-based methods enable a well-informed interpretation of the intrinsic physical mechanisms of imaging through scattering medium. Nevertheless, physical model-based PSI methods have a common drawback: they generally are based on idealized conditions, which seriously restricts their applicability.

To overcome the defects of physical model-based PSI methods, a range of DL-based methods have been proposed for the PSI, e.g., Polarimetric-Net [29], AOD-Net [30], IPLNet [31], PDRDN [32], PFNet [33], U²R-pGAN [34], MU-DLU [22], and attention-based residual neural network [35]. Generally, these methods for polarization image reconstruction involve three primary steps: dataset construction, feature representation, and model design and optimization. For instance, Polarimetric-Net employs a commercial division of focal plane (DoFP) polarization camera to take the polarization images of specific scenes. Subsequently, these images are split into four polarized images with polarization directions of 0° ($I_{0^\circ}(x, y)$), 45° ($I_{45^\circ}(x, y)$), 90° ($I_{90^\circ}(x, y)$), and 135° ($I_{135^\circ}(x, y)$) including the linear polarization information. Finally, three polarized images, i.e., $I_{0^\circ}(x, y)$, $I_{45^\circ}(x, y)$, and $I_{90^\circ}(x, y)$, are serially merged as one feature source to feed into a polarization dense network for underwater image restoration. Similar to Polarimetric-Net, PDRDN also uses the DoFP polarization camera to generate four polarized images and enters them into a residual dense network for denoising. In contrast to the aforementioned DL-based methods, MU-DLU utilizes a Monte Carlo algorithm to establish an optical simulation platform that models the actual atmospheric scattering environment and collects a synthetic polarization dataset. The Q-component of the Stokes vector is then coupled with a modified U-net to retrieve the target information influenced by the scattering media.

These methods have demonstrated significant advancements in addressing the problem of polarization image reconstruction. Nevertheless, we notice there is still room for further improvement in the performance of existing methods by addressing their potential defects. To be specific, most of existing PSI methods use either one single-view polarization feature map or blindly concatenate multiple single-view polarization feature maps directly as inputs to the DL algorithms, which fails to mine sufficient discriminative information effectively. Furthermore, although the usage of fused multi-view features can represent the polarization information, in most of the cases it contains overlapping information that will seriously reduce the efficiency of PSI model. On the other hand, facing the complicated natural scattering environment, designing an effective DL framework is also a major challenge for researchers.

To tackle the aforementioned critical issues, we herein explore the optimization of effective multiple polarization feature learning for maximizing the usage of the network capability, and

propose a well-designed three-stage multi-pipeline networks (TSMPN) architecture for the PSI, named as TSMPN-PSI. The novelties of our work compared with previous PSI frameworks can be briefly summarized as follows:

- 1) We present a new DL-based method, called TSMPN-PSI, established in a systematic TSMPN architecture, to further improve the reconstruction performance of polarization images, which is comprising three parts, namely pre-processing of de-speckling polarization images, multiple polarization feature learning, and a down-stream structure with cascaded network branches for image reconstruction.
- 2) A new multi-polarization-driven multi-pipeline (MPDMP) framework is designed, which effectively acquiring targets' multi-polarization information. The MPDMP framework contains five sub-pipelines that undertake different tasks. Furthermore, we design a novel multi-tiered recurrent de-speckling module (MTRDS), to suppress the potential noise in polarized scattering images.
- 3) We construct one benchmark scattering image dataset under active light illumination conditions, including a wider range of real-world scenarios, such as varying imaging distances through scattering media, diverse target structures and materials, and their corresponding background materials. This dataset will facilitate the further PSI development.
- 4) We validate the efficacy of TSMPN-PSI through extensive experiments on challenging real-world image datasets. The benchmarking results demonstrate that our proposed TSMPN-PSI significantly outperforms several existing state-of-the-art approaches, delivering more robust performance.

2. Methodology

The proposed TSMPN-PSI is composed of three stages: de-speckling polarization image, multi-polarization feature learning, and image reconstruction. The flowchart is illustrated in Fig. 1.

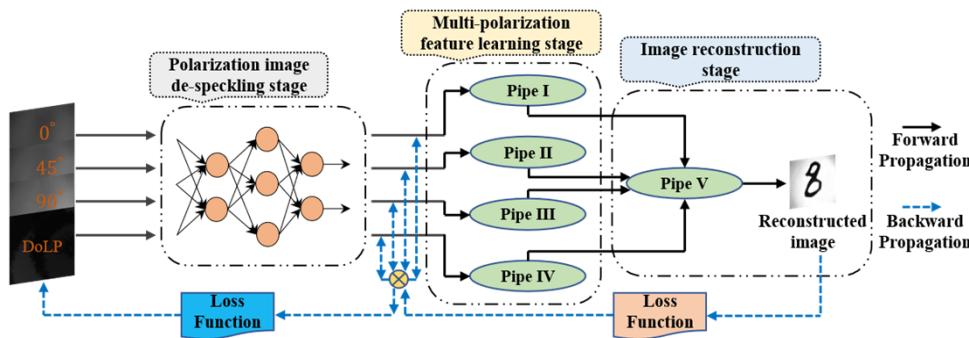


Fig. 1. Flowchart of the proposed TSMPN-PSI.

2.1. Theoretical background

As a higher-dimension information, the Stokes vector is more suitable for characterizing the polarization properties of natural light [12]. Polarized light or non-polarized light can be represented by the Stokes vector $S = [I(x, y), Q(x, y), U(x, y), V(x, y)]^T$, where $I(x, y)$ denotes the total light intensity, $Q(x, y)$, $U(x, y)$, and $V(x, y)$ are the light intensity difference between different orthogonal polarization states. For example, $Q(x, y)$ refers to the light intensity

difference between 0° and 90° polarization directions [36]. In general, circularly polarized light is rarely available in the natural environment in visible band, so $V(x, y)$ component is ignored in this work. In practice, the first three components of the Stokes parameters can be obtained by measuring the intensities of light along four different directions. More specifically, researchers use the DoFP to measure the intensities of light along four directions of 0° , 45° , 90° , and 135° in the same scene, which are denoted as $I_{0^\circ}(x, y)$, $I_{45^\circ}(x, y)$, $I_{90^\circ}(x, y)$, $I_{135^\circ}(x, y)$ respectively. The first three components in the Stokes parameters and the degree of linear polarization (DoLP) can be respectively calculated as:

$$\begin{aligned} I(x, y) &= I_{0^\circ}(x, y) + I_{90^\circ}(x, y) \\ Q(x, y) &= I_{0^\circ}(x, y) - I_{90^\circ}(x, y) \\ U(x, y) &= I_{45^\circ}(x, y) - I_{135^\circ}(x, y) \end{aligned} \quad (1)$$

$$DoLP = \sqrt{Q(x, y)^2 + U(x, y)^2} / I(x, y) \quad (2)$$

In a scattering environment, natural light traveling through a scattering medium and interacting with it ordinarily brings about an alteration in the polarization characteristics of the reflected or transmitted light. Now, we assume that the Stokes vector of the incident light is $J(x, y)$, the Stokes vector of the outgoing light after interacting with the medium is $J^*(x, y)$, and the Muller matrix (MM) [37] can formulate their relationship as follows:

$$J^*(x, y) = MM \cdot J(x, y) \quad (3)$$

This process makes the detector receive object information containing a large number of speckles, which will affect the imaging quality. Meanwhile, the quality of scattering imaging deteriorates as the concentration of the scattering medium becomes larger or the imaging distance becomes longer. It is more challenging that the detector can't effectively distinguish the target in background, when the target and its background have similar polarization characteristics. In this work, we regard the reconstruction of the object image as an inverse imaging problem computationally, and the process can be formulated as the following objective function:

$$J(x, y) = MM^{-1} \cdot J^*(x, y) \quad (4)$$

where MM^{-1} is the inverse of MM . DL has proven to be an effective resolution to the inverse imaging problem. Therefore, DL is incorporated with the polarization physical priors to learn the statistical distribution of polarization information in an effort to achieve the mapping of low-quality primary scattered images to high-quality restored image in this work.

2.2. Measurement system

Some existing DL-based methods have achieved considerable progresses in reconstructing polarization images on synthetic dataset, but there is dissimilarity in polarization characteristics between the synthetic and the real-word polarization image, which seriously restricts the development of polarization scattering imaging. Hence, we establish a real-world polarization scattering imaging system under active light illumination to collect a real-life scenarios dataset to directly verify the effectiveness of polarization image recovery methods. The schematic diagram of the experimental setup is illustrated in Fig. 2(a). ‘Dis’ denotes the distance between the scattering medium and the objects. A linear polarizer is employed as the polarization state generator to provide polarized illumination $S = (1, 1, 0, 0)^T$ in front of an LED light source. The ground glass is utilized as the scattering medium device. The light is focused by means of a convex lens. We use DoFP polarization camera (LUCID, PHX055S-PC) to take simultaneously

four polarized images with polarization directions of 0° , 45° , 90° , and 135° , called $I_{0^\circ}(x, y)$, $I_{45^\circ}(x, y)$, $I_{90^\circ}(x, y)$, and $I_{135^\circ}(x, y)$, respectively. Taking hand-written digits written in ink on the white paper as an example, Fig. 2(b) shows separately the $I_{0^\circ}(x, y)$, $I_{45^\circ}(x, y)$, $I_{90^\circ}(x, y)$, $I_{135^\circ}(x, y)$, DoLP, and the ground truth respectively.

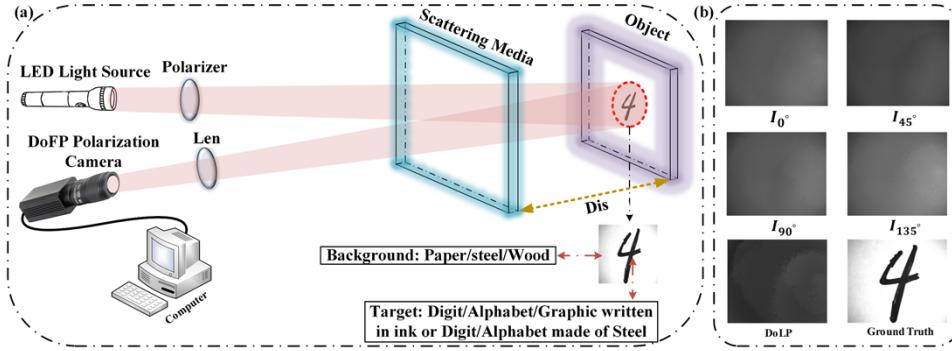


Fig. 2. (a) Experimental setup for the PSI through scattering medium, (b) the $I_{0^\circ}(x, y)$, $I_{45^\circ}(x, y)$, $I_{90^\circ}(x, y)$, $I_{135^\circ}(x, y)$, DoLP, and the ground truth.

2.3. Three-stage multi-pipeline networks (TSMPN) architecture

With increasing popularity of DL techniques in computational imaging domains, DL techniques have shown superior results over traditional methods across a wide variety of tasks [38,39], particularly in imaging through scattering media. In this work, leveraging the power of the Convolutional Long Short-Term Memory (ConvLSTM) [40] module, the multi-scale residual structure (Res2Net) [41], and the Atrous Spatial Pyramid Pooling (ASPP) [42] module, we design and implement our proposed TSMPN-PSI pipeline based on one newly designed MTRDS module and one novel designed MPDMP framework.

- 1) *MTRDS module:* A pre-requisite for developing powerful computational models to reconstruct polarization image is the suppression of noise and irrelevant information in the initial feature source. Hence, our designed MTRDS module attempts to record speckle degradations via convolutional sequence-to-sequence process, which can guide the subsequent network to focus on the potential target regions [43]. The whole structure of our MTRDS module is shown in Fig. 3. To be specific, the proposed MTRDS module includes six iterations, each of which is consisting of an initial convolutional layer followed by a ConvLSTM layer, eight Res2Net blocks, and a convolutional layer in the end.
- 2) *Res2Net:* Different from common ResNet block [44], the recently proposed network architecture Res2Net is designed to obtain multi-scale features more efficiently. As illustrated in Fig. 3(B), in Res2Net, the feature maps denoted by x first are input into a 1×1 convolutional layer. Then the transformed feature maps are split into several subsets (i.e., x_1 , x_2 , x_3 , and x_4) in channel accordingly, followed by different operations. Finally, multi-scale features are fed into another 1×1 convolutional layer after merging, and are summed with x as the output y .
- 3) *ConvLSTM:* Inspired by the ConvLSTM superior ability to learn context-dependent information, we utilize it to capture relevant information relating to target of interest. The ConvLSTM mainly contains three basic gates, i.e., an input gate, a forget gate, and an output gate. The forget gate determines which information should be discarded from the cell state while updating; meanwhile, the input gate decides what new data can be stored in

the cell state, and the output gate decides what information can be output based on the cell state [45].

- 4) *MPDMP framework*: the extraction of effective polarization features is considered the most important step in developing accurate computational methods for retrieving target information. Hereby, a customized MPDMP framework, which is composed of multiple polarization feature learning and image reconstructing stages, is designed to extract the discriminative information from four single-view feature maps to reconstruct polarization image. As shown in Fig. 4, the MPDMP consists of five sub-pipelines named Pipes I, II, III, IV, and V. In the multiple polarization feature learning stage, the sub-pipelines I, II, III, and IV corresponding to input feature maps $I_{0^\circ}(x, y)$, $I_{45^\circ}(x, y)$, $I_{90^\circ}(x, y)$, and DoLP, respectively, are set the same network's hyper-parameters, including a Block a (i.e., an initial convolutional layer), Block c repeated five times, and a Block d . Taking the process of extracting discriminative features of $I_{0^\circ}(x, y)$ from Pipe I as an illustration, the feature map $I_{0^\circ}(x, y)$ is first entered an initial convolutional layer, which is transformed into a spatial vector. Subsequently, the transformed feature maps are fed into five Block c of various depths (i.e., N_1 , N_2 , N_3 , N_4 , and N_5) to make the extraction and fusion of multi-dimensional features throughout the forwarding propagation. Here, each Block c consists of x Block b (i.e., residual network) and a max pooling layer. Finally, a Block d , i.e., ASPP, utilizes multiple parallel filters to generate multi-scale feature from the output of the last Block c and fuses them [42]. At the stage of image reconstructing, Pipe V is designed to integrate the diverse levels of feature information mined in the first stage, which consists of five Block e . The detailed structure of the Block e is illustrated in Fig. 4. To be concrete, in Block e , a calibrator adjusts the number of channels and size of two adjacent features before fusing two feature maps. The fused maps are then processed through a bilinear up-sampling module, and followed by a 3×3 convolutional layer. Finally, they are fed into a 1×1 convolutional layer to adjust the output feature map's channel count to the specified value. It should be noted that after each convolutional layer, batch normalization and Relu activation functions are employed. A dropout strategy with a ratio of $d\%$ is utilized to avoid overfitting.

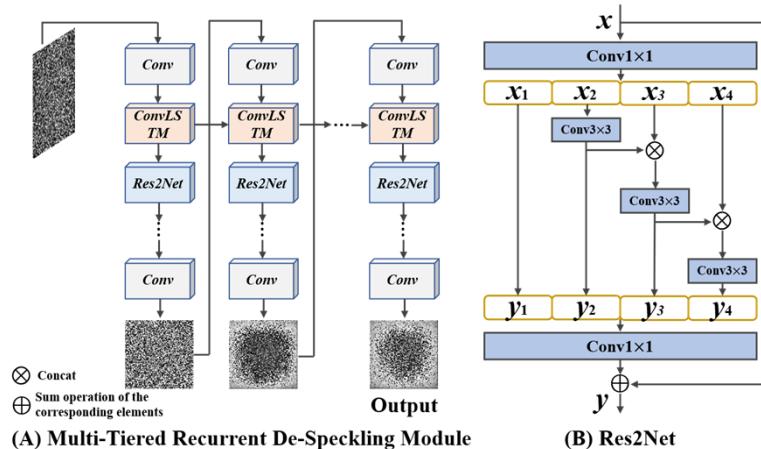


Fig. 3. (A) Block diagram of the network architecture of MTRDS module. (B) The structure of Res2Net module.

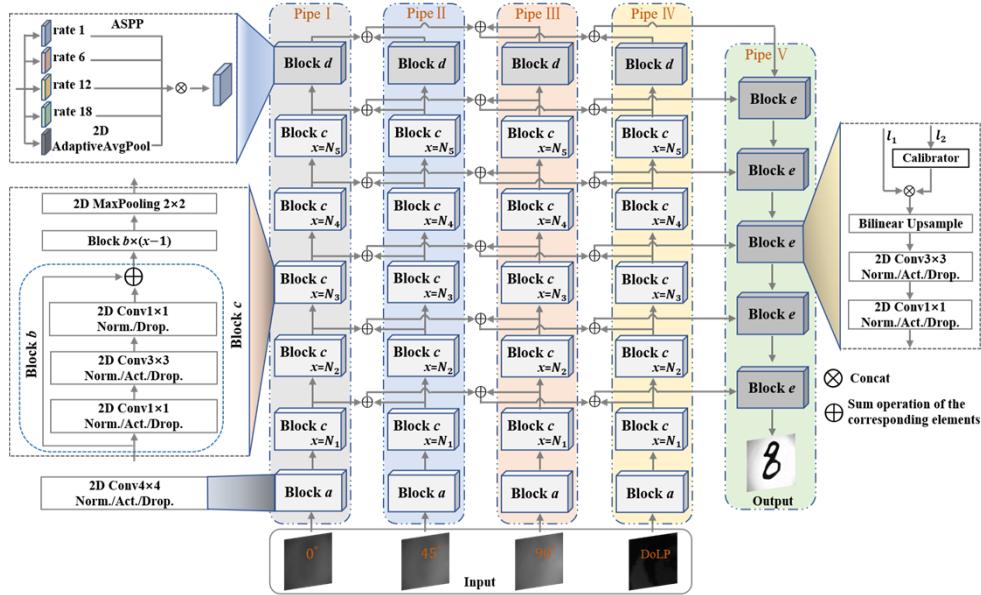


Fig. 4. The network layout of the MPDMP framework. It consists of multiple polarization feature learning stages (i.e., Pipe I, Pipe II, Pipe III, and Pipe IV) and image reconstructing stage (i.e., Pipe V). ‘Norm.’ denotes the batch normalization layer, ‘Act.’ denotes the linear rectification activation function (Relu), and ‘Drop.’ denotes the dropout ratio.

2.4. Training details and evaluation metrics

We perform all experiments on Linux Server (Ubuntu 20.04) Intel Core i7-7700 CPU @ 3.6 Hz 32.0GB of RAM, and Python 3.7 programming. TSMPN architecture, which is implemented using *PyTorch* software (Version 1.7.1), is trained on one graphics processing unit (Nvidia GeForce RTX 3090) to speed up training. In the model training process, we use the mean squared error (MSE) function to calculate the loss and optimize the model by the Adam algorithm with a learning rate of lr and a batch size of bs . We use the strategy of grid search and adjust the network’s hyper-parameters, i.e., $N_1, N_2, N_3, N_4, N_5, d, lr$, and bs , by observing the model performance on the training dataset (see “Data Acquisition” section for details) over five-fold cross-validation tests. Lastly, according to the best/sub-best performance of TSMPN model, we use the following values for the above hyper-parameters: $N_1=3, N_2 = 6, N_3 = 8, N_4 = 10, N_5 = 5, d = 25\%, lr = 0.0001$, and $bs = 15$.

We employ four evaluation metrics, i.e., Structural Similarity Index Measure (SSIM), Pearson Correlation Coefficient (PCC), MSE, and Peak Signal-to-Noise Ratio (PSNR), to evaluate the quality of image recovery. SSIM is utilized to evaluate the structure and texture between the reconstructed and real images, and a higher SSIM value reflects a more similar structure and texture. PCC is employed to quantify the relationship between the reconstructed and real pixel-level values of each image, and its value is between -1 and 1, whereas MSE is applied to quantitatively measure the average deviation between the reconstructed and ground truth pixel-level values of each image. Furthermore, we also use PSNR to quantify the content between the reconstructed and ground truth images, and a higher PSNR value represents closer image content.

3. Results and discussions

3.1. Data acquisition

Training data set. In this work, based on our established PSI system under active light illumination (refer to the section ‘Experimental Setup’), we collect 200 samples of polarization scattering images as the training data set to infer the parameters of our proposed TSMPN-PSI. Specifically, firstly, we employed a linear polarizer as the polarization state generator to provide polarized illumination $S = (1, 1, 0, 0)^T$ in front of an LED light source. Meanwhile, we put the object of interest away from the DoFP polarization camera to generate ground truth images. Then, we place the scattering medium device (i.e., ground glass) at a distance of 40 mm from the object and use the DoFP polarization camera to generate scattered images containing polarization information as preliminary feature source. Finally, based on scattered images of different directions, we use Eq. (1) and 2 to calculate the Stokes vector and DoLP accordingly. It should be noted that each polarization scattered image in the training data set is imaged at a distance of 40 mm, and its target is simply hand-written digit written in ink on white paper. In addition, in the training phase, the Scikit-image [46] python library is employed to enhance the training data set by rotating and flipping the existing images, and all images are adjusted to a fixed size, i.e., 256×256.

Testing data set. We evaluate the performance of the proposed method on three independent testing data sets. In contrast to the previous method, the targets in our testing data sets are completely invisible to observation, and include more abundant real-life scenarios. More specifically, the testing data sets are divided into three groups, i.e., Groups I, II, and III. Group I, for characterizing the generalization capability of the proposed method concerning the untrained targets with different structures, we collect 24 pairs of images having the same imaging distance (i.e., 40 mm) as the training data set, which contain untrained hand-written digits, hand-written alphabets, and hand-written graphic patterns, written in ink on the white paper. Group II, to further characterize the generalization capability of our proposed method in term of imaging distance, we generate 50 pairs of images at distances of 35 mm, 40 mm, 42.5 mm, 45 mm, and 50 mm between ground glass and the targets, whose target structures and background materials same as Group I. Group III, to characterize the generalization capability of the proposed method concerning the untrained object and background materials, we produce 18 pairs of images having the same imaging distance as Group I, which consist of three classifications, i.e., Paper-Steel, Wood-Ink, and Wood-Steel, where Paper-Steel, Wood-Ink, and Wood-Steel denote digits (or alphabets) made of steel against paper background, digits (or alphabets) written in ink against wood, and digits (or alphabets) made of steel against wood background, respectively.

3.2. Performance of multi-pipeline learning strategy

To demonstrate that our proposed multi-pipeline learning framework (MPLF) (introduced in the “Three-Stage Multi-Pipeline Networks Architecture” section) can extract more discriminative information from the four stand-alone single-view feature maps, i.e., $I_{0^\circ}(x, y)$, $I_{45^\circ}(x, y)$, $I_{90^\circ}(x, y)$, and DoLP, we compare the performances between the proposed MPLF and single-pipeline learning framework (abbreviated as SPLF). Unlike MPLF framework, we integrate MTRDS module, Pipe I, and Pipe V to build SPLF, and employ the fusion feature maps of the above-mentioned four single-view feature maps as feature source. Specifically, for each framework, we use it to train a model on the training data set with the optimal or sub-optimal parameters selected in cross-validation and validate the performance of the trained model on the untrained targets with different structures, i.e., Group I testing data set. Some cases in Group I testing data set are randomly selected for performance analysis, and the results are displayed in Fig. 5.

As can be seen from Fig. 5, in terms of both the detail and structural integrity of the recovered images, proposed MPLF strategy consistently outperforms the SPLF. Further, Table 1 also lists



Fig. 5. Visual performance comparison between MPLF and SPLF frameworks on the untrained targets with different structures. (a) Ground truth. (b) Scattering images with polarization directions of 0° . (c) Reconstructed images by SPLF framework. (d) Reconstructed images by MPLF framework.

the average SSIM, PCC, MSE, and PSNR values of reconstructed images by the MPLF and SPLF in Fig. 5, in which it is easy to see that the MPLF outperforms SPLF concerning the four evaluation indexes. Concretely, the SSIM, PCC, MSE, and PSNR values of MPLF are 0.75, 0.81, 0.043, and 13.97, which are 4.17%, 6.58%, 23.21%, and 8.38% higher than those of SPLF respectively. The above comparison results demonstrate that the MPLF could extract more discriminative information from multi-view feature maps to enhance the polarization image recovery performance. That is to say, MPLF is able to get useful complementary information from multiple independent polarization feature maps, which helps in reconstruction of target information.

Table 1. Performance comparison between MPLF and SPLF on the untrained targets with different structures.

Model	SSIM	PCC	MSE	PSNR
SPLF	0.72	0.76	0.056	12.89
MPLF	0.75	0.81	0.043	13.97

3.3. Performance comparison with existing methods

In this section, to evaluate the performance of our proposed TSMPN-PSI, we will experimentally compare it with several state-of-the-art methods, including DCP [7], Polarimetric-Net [29], MU-DLU [22], and our implementation version of UNet-DoLP [20] that trains the U-Net with solely DoLP feature map. In four methods, the former one, i.e., DCP, is physical model-based methods, while the other three, i.e., Polarimetric-Net, MU-DLU, and UNet-DoLP, are DL-based methods. For an objective and fair comparison, except for DCP method, the remaining methods are respectively trained on the same training data set and evaluated on the same independent testing data sets.

1) *Performance Comparison on Targets with Different Geometries:* The purpose of the sub-section is to experimentally demonstrate the efficacy of the proposed TSMPN-PSI by comparing it with four state-of-the-art methods mentioned above on the untrained targets with different structures i.e., the Group I testing data set. The results of two untrained handwritten digit testing targets, two untrained handwritten alphabet testing targets, and two untrained handwritten graphic pattern testing targets are illustrated in Fig. 6 and Table 2 to intuitively compare the performance by TSMPN-PSI and other methods. By observing Fig. 6 and Table 2, it is easy to find that TSMPN-PSI is consistently superior to four methods with regard to SSIM, PCC, PSNR, and MSE evaluation indexes. Compared

with UNet-DoLP, the second-best method among all methods, the values of SSIM, PCC, PSNR, and MSE of TSMPN-PSI are enhanced by 3.75%, 13.16%, 37.50% and 12.27% on Fig. 6 (a) sample, respectively. As expecting, the DCP method, which is developed based on physical model, gains the lowest reconstruction performance in terms of four evaluation indexes.

	Ground Truth	DCP	Polarimetric-Net	MU-DLU	UNet-DoLP	TSMPN-PSI
(a)	7					
(b)	2					
(c)	F					
(d)	A					
(e)	□					
(f)	○					

Fig. 6. Visual performance comparison between TSMPN-PSI and other methods on the untrained targets with different structures. (a), (b), (c), (d), (e), and (f) are untrained handwritten digits, handwritten alphabets, and handwritten graphic patterns, written in ink on the white paper.

Figure 7 illustrates the head-to-head comparisons between TSMPN-PSI and other methods on the Group I testing data set, which contains 8 untrained handwritten digit testing targets, 8 untrained handwritten alphabet testing targets, and 8 untrained handwritten graphic pattern testing targets. From Fig. 7, it is clear that the performance of TSMPN-PSI is superior to that of the other methods in terms of the numbers of higher SSIM, PCC, PSNR, and MSE. For instance, out of the 24 targets, there 24, 24, 24, and 19 cases where TSMPN-PSI has better SSIM values than DCP, Polarimetric-Net, MU-DLU, and UNet-DoLP, respectively.

2) *Performance Comparison on Targets with Different Imaging Distances:* To further examine the effectiveness of proposed TSMPN-PSI, we compare it with DCP, Polarimetric-Net, MU-DLU, and UNet-DoLP on the dependent testing data set Group II, i.e., the untrained targets with different imaging distances. Five samples, which are selected from the Group II, are used for detailed visual comparison between TSMPN-PSI and other methods. Figure 8 and Table 3 show that the TSMPN-PSI's generalization performance about imaging distances is significantly superior to those of the other four methods. Specifically, the SSIM, PCC, PSNR, and MSE of TSMPN-PSI on Fig. 8 (b) sample are 15.49%, 116.67%, 76.00%, and 68.36% higher, respectively, than the corresponding values yielded by Polarimetric-Net that is the recently reported DL-based multi-polarization feature learning methods. Furthermore, by carefully observing Fig. 8 and Table 3, the following three phenomena can be seen: (1) The generalization capability of TSMPN-PSI at $Dis = 40$ mm outperforms those at $Dis = 35$ mm, $Dis = 42.5$ mm, $Dis = 45$ mm, and $Dis = 50$ mm with respect to four evaluation indexes. (2) It is straightforward to find that the image recovered by the TSMPN-PSI have well structural integrity, when $Dis \leq 42.5$ mm. While $Dis > 42.5$ mm, although the performance of the TSMPN-PSI drops slowly in recovering image details, the background and target of the reconstructed image can be effectively distinguished. (3) It has not escaped from our notice that for the case at $Dis = 50$ mm, the imaging distance's generalization reaches to 25%, and the target can be also distinguished, which can be an index for demonstrating excellence of our proposed TSMPN-PSI.

Table 2. Performance comparison between TSMPN-PSI and other methods on the untrained targets with different structures.

Images	Metrics	DCP	Polarimetric-Net	MU-DLU	UNet-DoLP	TSMPN-PSI
(a)	SSIM	0.37	0.74	0.78	0.80	0.83
	PCC	0.35	0.42	0.75	0.76	0.86
	MSE	0.84	0.089	0.034	0.032	0.02
	PSNR	5.45	10.74	15.2	15.48	17.38
(b)	SSIM	0.39	0.74	0.78	0.76	0.83
	PCC	0.28	0.34	0.70	0.71	0.88
	MSE	0.91	0.098	0.045	0.049	0.024
	PSNR	5.11	10.38	13.68	13.71	16.33
(c)	SSIM	0.35	0.70	0.72	0.72	0.79
	PCC	0.34	0.41	0.56	0.70	0.85
	MSE	0.99	0.11	0.066	0.049	0.028
	PSNR	4.73	9.91	12.44	13.71	15.77
(d)	SSIM	0.35	0.71	0.75	0.77	0.77
	PCC	0.23	0.30	0.74	0.77	0.81
	MSE	0.99	0.11	0.045	0.037	0.037
	PSNR	4.79	10.33	13.09	14.79	14.41
(e)	SSIM	0.28	0.64	0.66	0.70	0.73
	PCC	0.29	0.39	0.62	0.72	0.77
	MSE	0.99	0.11	0.0712	0.057	0.051
	PSNR	4.77	9.80	12.17	13.23	13.36
(f)	SSIM	0.31	0.62	0.67	0.69	0.71
	PCC	0.25	0.38	0.74	0.76	0.81
	MSE	0.90	0.12	0.061	0.058	0.045
	PSNR	5.23	9.75	13.11	13.42	13.92

In addition, Fig. 9(a), (b), (c), and (d) display the distribution of SSIM, PCC, PSNR, and MSE values among individual sample in terms of median, 25th, and 75th percentile for all methods on Group II, respectively. The results shown in Fig. 9 clearly demonstrate that TSMPN-PSI outperforms the other methods concerning the median, 25th, and 75th percentile of SSIM, PCC, PSNR, and MSE values. Taking the distribution of PCC as an example, TSMPN-PSI achieves the highest median PCC with the least spread around the median values as compared to other methods. This shows the stable performance of TSMPN-PSI in comparison to other methods.

3) *Performance Comparison on Target with Different Materials:* The structural properties and the constituents of the target's material considerably influence the polarization characteristics of the image signal. In this sub-section, to further highlight the generalization capability of our proposed method, we compare it with DCP, Polarimetric-Net, MU-DLU, and UNet-DoLP by independent validation on Group III testing data set, i.e., the untrained object and background materials. Figures 10 and 11 and Table 4 summarize the compared results.

As demonstrated in Fig. 10 and Table 4, the proposed TSMPN-PSI achieves satisfactory results. Although the existing methods gain a reasonable SSIM of 0.12~0.72 on six cases, the improvement of TSMPN-PSI is also significant. Using the results of Fig. 10(d) (i.e., Wood-Ink) as an example, the SSIM and PSNR, which are two overall measurements of the quality of the

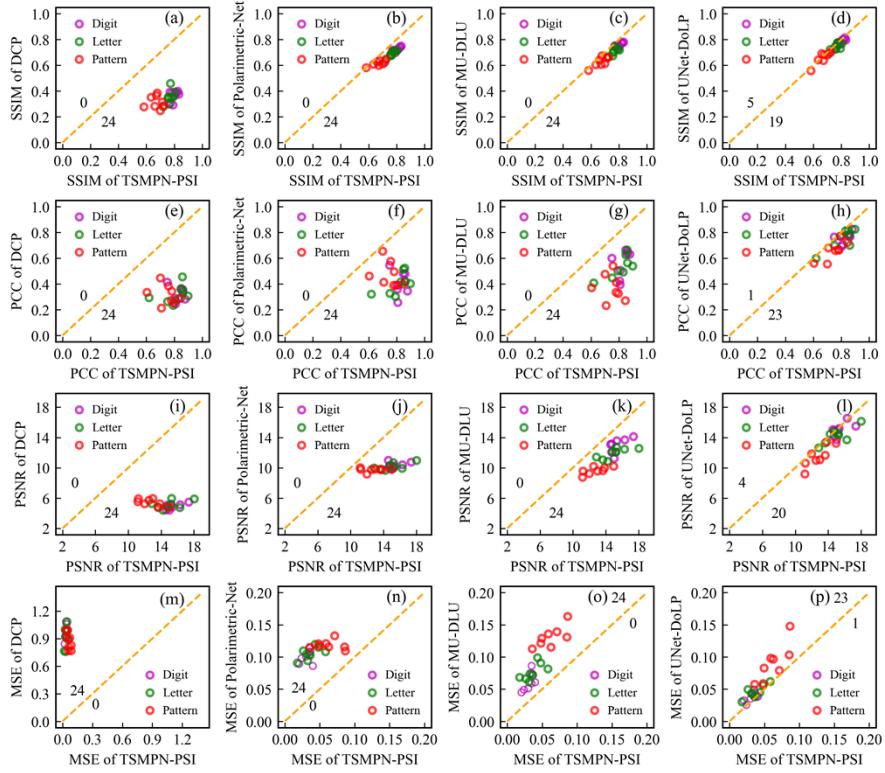


Fig. 7. Head-to-head comparison of SSIM, PCC, PSNR, and MSE between TSMPN-PSI and other methods on the Group I testing data set. Each purple, green, and red circle mean untrained targets of handwritten digits, handwritten alphabets, and handwritten graphic patterns, respectively, written in ink on the white paper. The numbers in each panel represent the number of points in the upper and lower triangles, respectively.

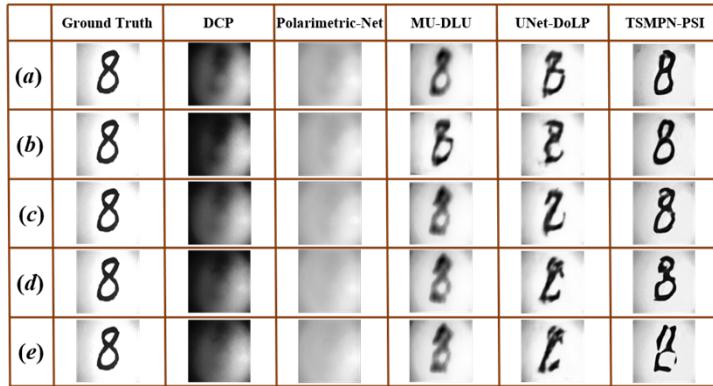
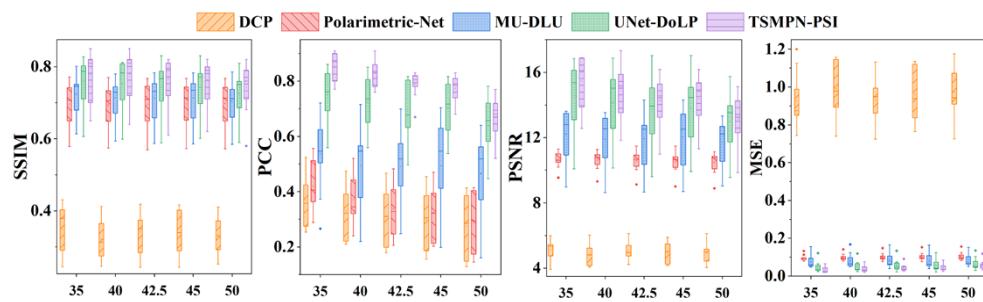


Fig. 8. Visual comparison between TSMPN-PSI and other methods on the untrained targets with different imaging distances. (a) Dis = 35 mm. (b) Dis = 40 mm. (c) Dis = 42.5 mm. (d) Dis = 45 mm. (e) Dis = 50 mm.

Table 3. Performance comparison between TSMPN-PSI and other methods on the untrained targets with different imaging distances.

Images	Metrics	DCP	Polarimetric-Net	MU-DLU	UNet-DoLP	TSMPN-PSI
Dis = 35mm	SSIM	0.37	0.70	0.75	0.78	0.81
	PCC	0.34	0.43	0.74	0.74	0.82
	MSE	0.90	0.10	0.041	0.037	0.034
	PSNR	5.14	10.18	14.45	14.26	14.73
Dis = 40mm	SSIM	0.26	0.71	0.75	0.78	0.82
	PCC	0.32	0.42	0.71	0.72	0.91
	MSE	1.16	0.10	0.044	0.04	0.024
	PSNR	4.06	10.02	14.08	14.66	16.87
Dis = 42.5mm	SSIM	0.35	0.7	0.75	0.75	0.79
	PCC	0.30	0.37	0.69	0.66	0.78
	MSE	0.912	0.10	0.045	0.049	0.04
	PSNR	5.12	10.41	13.36	13.71	14.21
Dis = 45mm	SSIM	0.30	0.71	0.74	0.75	0.79
	PCC	0.29	0.32	0.68	0.71	0.79
	MSE	1.13	0.098	0.045	0.042	0.034
	PSNR	4.15	10.47	13.96	14.24	15.1
Dis = 50mm	SSIM	0.32	0.70	0.73	0.74	0.76
	PCC	0.26	0.28	0.71	0.66	0.60
	MSE	0.99	0.099	0.043	0.05	0.065
	PSNR	4.70	10.43	14.25	13.46	12.29

**Fig. 9.** Boxplot of SSIM, PCC, PSNR, and MSE for TSMPN-PSI and other methods on different imaging distances. On each box, the square and horizontal represent the mean and median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. the outliers are plotted individually using the diamond with filled color.

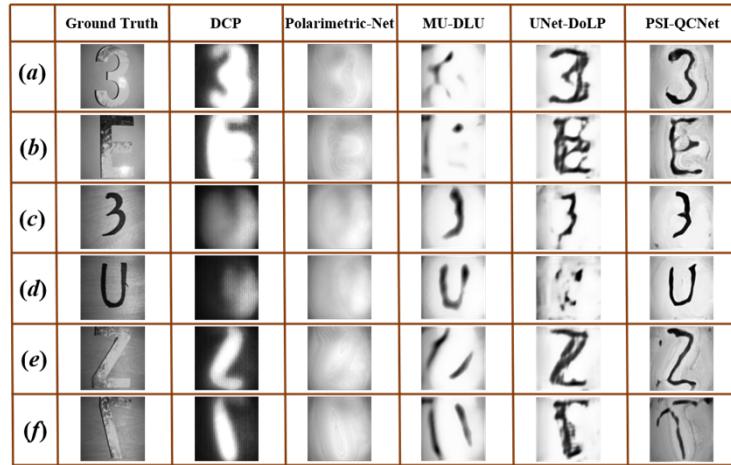


Fig. 10. Visual performance comparison between TSMPN-PSI and other methods on the untrained target and background materials. (a) and (b) are digits (or alphabets) made of steel against paper background, i.e., Paper-Steel. (c) and (d) are digits (or alphabets) written in ink against wood, i.e., Wood-Ink. (e) and (f) are digits (or alphabets) made of steel against wood background, i.e., Wood-Steel.

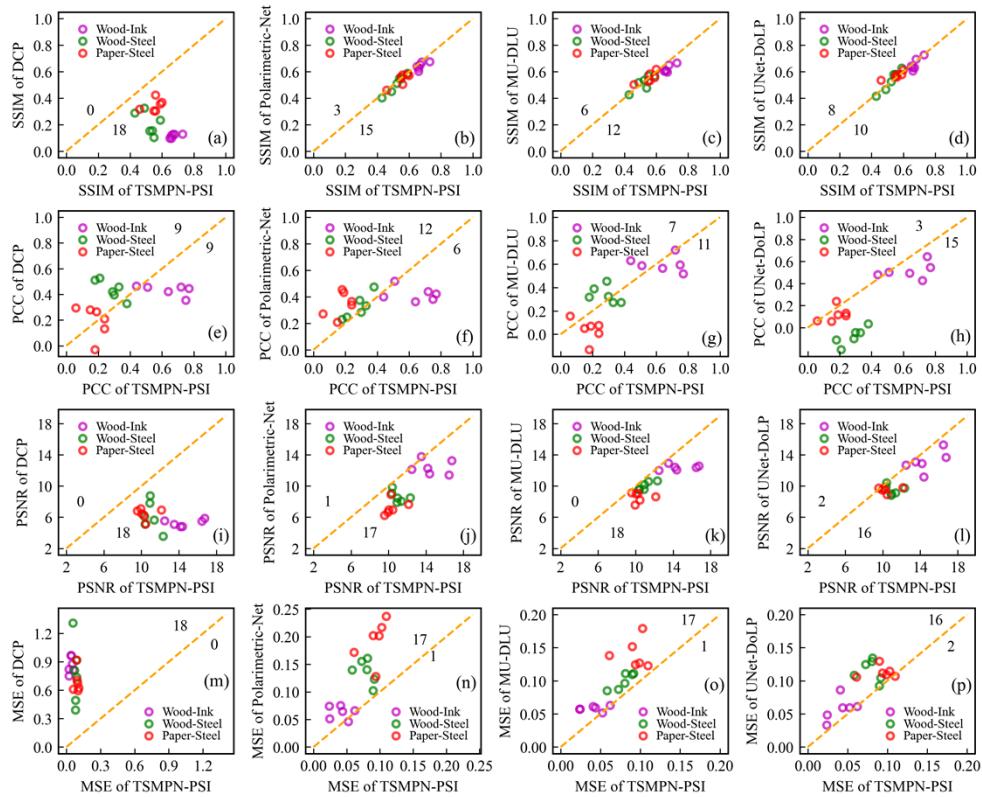


Fig. 11. Head-to-head comparison of SSIM, PCC, PSNR, and MSE between TSMPN-PSI and other methods on the Group III testing data set. Each purple, green, and red circle mean Wood-Ink, Wood-Steel, and Paper-Steel, respectively. The numbers in each panel represent the number of points in the upper and lower triangles, respectively.

Table 4. Performance comparison between TSMPN-PSI and other methods on the untrained target and background materials.

Images	Metrics	DCP	Polarimetric-Net	MU-DLU	UNet-DoLP	TSMPN-PSI
(a)	SSIM	0.36	0.57	0.61	0.61	0.60
	PCC	0.27	0.20	0.049	0.055	0.15
	MSE	0.67	0.20	0.13	0.11	0.098
	PSNR	6.44	6.93	8.98	9.65	10.05
(b)	SSIM	0.42	0.50	0.53	0.55	0.59
	PCC	-0.031	0.45	-0.13	0.24	0.33
	MSE	0.59	0.21	0.17	0.11	0.091
	PSNR	7.08	6.60	7.56	9.48	10.35
(c)	SSIM	0.12	0.67	0.72	0.72	0.73
	PCC	0.35	0.38	0.62	0.64	0.75
	MSE	0.82	0.073	0.032	0.033	0.023
	PSNR	5.48	11.41	15.36	15.26	16.47
(d)	SSIM	0.096	0.62	0.61	0.62	0.66
	PCC	0.42	0.36	0.56	0.49	0.64
	MSE	0.96	0.064	0.059	0.059	0.043
	PSNR	4.78	12.27	12.41	12.88	14.19
(e)	SSIM	0.15	0.51	0.53	0.52	0.53
	PCC	0.52	0.24	-0.20	-0.20	0.21
	MSE	0.81	0.15	0.13	0.12	0.072
	PSNR	5.62	8.04	8.93	9.11	11.38
(f)	SSIM	0.15	0.54	0.53	0.57	0.54
	PCC	0.32	0.47	-0.068	0.033	0.38
	MSE	0.91	0.10	0.13	0.09	0.089
	PSNR	5.09	9.84	9.03	10.40	10.41

image reconstruction, are 0.66 and 14.19 respectively for the TSMPN-PSI, which are 587.50% and 196.86%, 6.45% and 15.65%, 8.20% and 14.34%, and 6.45% and 10.17% higher than those of DCP, Polarimetric-Net, MU-DLU, and UNet-DoLP, respectively. By observing Fig. 10, it is noteworthy that although four compared methods obtain good performance in terms of four evaluation metrics, the images they restored are extremely blurred and the targets are difficult to distinguish. Figure 11 also shows that the performance of TSMPN-PSI is superior to that of the other methods with respect to the numbers of higher SSIM, PCC, PSNR, and MSE, except for when compared to the PCC of Polarimetric-Net.

4. Conclusion

In this work, a new MPDMP framework is designed to effectively learn rich hierarchical representations from multiple polarization feature maps. Using the proposed MPDMP, a robust DL method is developed for polarization image reconstruction, called TSMPN-PSI. By comparison with several state-of-the-art methods on three real-life testing data sets, the efficacy of the proposed method has been demonstrated and verified. The excellent performance of TSMPN-PSI is mainly attributed to the strong capability of MPDMP for effectively dealing with the physical representation between the object intensity and its polarization information.

Although our work achieves some improvements in PSI, there is still room to further enhance its performance due to the following three points. First, it may be a promising approach to further improve the performance of TSMPN-PSI by incorporating more information with heterogeneous features that are complementary to the currently used encoding features. Second, the inner workings of the DL model are a black box to us, and we can obtain more if we can understand the underlying workings. Third, the performance of DL-based computational imaging methods is optimal for high-contrast targets, but these methods are likely to encounter limitations when applied to complex scenes. The integration of physics and artificial intelligence could potentially offer a solution to the problem. On the other hand, the utilization of circularly polarized light may also be an important solution, but its data acquisition is a problem. Lastly, the generalization performance of TSMPN-PSI for underwater PSI is not yet known. Hence, we will investigate the applicability of TSMPN-PSI to underwater imaging in future. In the subsequent works, we will propose targeted strategies to counter the above issues.

Funding. National Natural Science Foundation of China (61775050).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. S. Yoon, M. Kim, M. Jang, Y. Choi, W. Choi, S. Kang, and W. Choi, "Deep optical imaging within complex scattering media," *Nat. Rev. Phys.* **2**(3), 141–158 (2020).
2. H.-F. Wang, Y.-H. Shan, T. Hao, X.-M. Zhao, S.-Z. Song, H. Huang, and J.-J. Zhang, "Vehicle-Road Environment Perception Under Low-Visibility Condition Based on Polarization Features via Deep Learning," *IEEE Trans. Intell. Transp. Syst.* (2022).
3. L. V. Wang and H.-I. Wu, *Biomedical optics: principles and imaging* (John Wiley & Sons, 2012).
4. C.-Q. Hu, Z.-Q. Yan, J. Gao, Z.-Q. Jiao, Z.-M. Li, W.-G. Shen, Y. Chen, R.-J. Ren, L.-F. Qiao, and A.-L. Yang, "Transmission of photonic polarization states through 55-m water: towards air-to-sea quantum communication," *Photonics Res.* **7**(8), A40–A44 (2019).
5. T. Hu, F. Shen, K. Wang, K. Guo, X. Liu, F. Wang, Z. Peng, Y. Cui, R. Sun, and Z. Ding, "Broad-band transmission characteristics of polarizations in foggy environments," *Atmosphere* **10**(6), 342 (2019).
6. X. Wang, T. Hu, D. Li, K. Guo, J. Gao, and Z. Guo, "Performances of polarization-retrieve imaging in stratified dispersion media," *Remote Sens.* **12**(18), 2895 (2020).
7. K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(12), 2341–2353 (2010).
8. J. Liang, L. Ren, E. Qu, B. Hu, and Y. Wang, "Method for enhancing visibility of hazy images based on polarimetric imaging," *Photonics Res.* **2**(1), 38–44 (2014).
9. D. Li, C. Xu, M. Zhang, X. Wang, K. Guo, Y. Sun, J. Gao, and Z. Guo, "Measuring glucose concentration in a solution based on the indices of polarimetric purity," *Biomed. Opt. Express* **12**(4), 2447–2459 (2021).
10. Y. Y. Schechner and N. Karpel, "Recovery of underwater visibility and structure by polarization analysis," *IEEE J. Oceanic Eng.* **30**(3), 570–587 (2005).
11. Q. Tao, Y. Sun, F. Shen, Q. Xu, J. Gao, and Z. Guo, "Active imaging with the aids of polarization retrieve in turbid media system," *Opt. Commun.* **359**, 405–410 (2016).
12. X. Li, Y. Han, H. Wang, T. Liu, S.-C. Chen, and H. Hu, "Polarimetric Imaging Through Scattering Media: A Review," *Front. Phys.* **10**, 153 (2022).
13. A. P. Mosk, A. Lagendijk, G. Lerosey, and M. Fink, "Controlling waves in space and time for imaging and focusing in complex media," *Nat. Photonics* **6**(5), 283–292 (2012).
14. J. Yoon, K. Lee, J. Park, and Y. Park, "Measuring optical transmission matrices by wavefront shaping," *Opt. Express* **23**(8), 10158–10167 (2015).
15. D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, and C. A. Puliafito, "Optical coherence tomography," *Science* **254**(5035), 1178–1181 (1991).
16. W. Yu, D. Li, K. Guo, Z. Yin, and Z. Guo, "Optimized sinusoidal patterns for high-performance computational ghost imaging," *Appl. Opt.* **62**(7), 1738–1744 (2023).
17. C. Xu, D. Li, K. Guo, Z. Yin, and Z. Guo, "Computational ghost imaging with key-patterns for image encryption," *Opt. Commun.* **537**, 129190 (2023).
18. C. Xu, D. Li, X. Fan, B. Lin, K. Guo, Z. Yin, and Z. Guo, "High-performance deep-learning based polarization computational ghost imaging with random patterns and orthonormalization," *Phys. Scr.* **98**(6), 065011 (2023).
19. F. Wang, C. Wang, C. Deng, S. Han, and G. Situ, "Single-pixel imaging using physics enhanced deep learning," *Photonics Res.* **10**(1), 104–110 (2022).

20. B. Lin, X. Fan, D. Li, and Z. Guo, "High-Performance Polarization Imaging Reconstruction in Scattering System under Natural Light Conditions with an Improved U-Net," *Photonics* **10**(2), 204 (2023).
21. B. Lin, X. Fan, and Z. Guo, "Self-attention module in a multi-scale improved U-net (SAM-MIU-net) motivating high-performance polarization scattering imaging," *Opt. Express* **31**(2), 3046–3058 (2023).
22. D. Li, B. Lin, X. Wang, and Z. Guo, "High-Performance Polarization Remote Sensing with the Modified U-Net Based Deep-Learning Network," *IEEE Trans. Geosci. Remote Sens.* **60**, 1–10 (2022).
23. C. Zhang, J. Hu, Y. Dong, A. Zeng, H. Huang, and C. Wang, "High efficiency all-dielectric pixelated metasurface for near-infrared full-Stokes polarization detection," *Photonics Res.* **9**(4), 583–589 (2021).
24. C. Zuo, J. Qian, S. Feng, W. Yin, Y. Li, P. Fan, J. Han, K. Qian, and Q. Chen, "Deep learning in optical metrology: a review," *Light Sci. Appl.* **11**(1), 39 (2022).
25. N. Carlevaris-Bianco, A. Mohan, and R. M. Eustice, "Initial results in underwater single image dehazing," in *Oceans 2010 Mts/IEEE Seattle* (IEEE, 2010), pp. 1–8.
26. Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Polarization-based vision through haze," *Appl. Opt.* **42**(3), 511–525 (2003).
27. X.-Y. Liu, J.-L. Zhang, Y.-F. Yin, Y. Yang, C. Zhang, and X.-X. Wu, "Underwater polarization image restoration based on logarithmic transformation and dark channel," *Optoelectronics Letters* **16**(2), 149–153 (2020).
28. Y. Zhao, W. He, H. Ren, Y. Li, and Y. Fu, "Polarization descattering imaging through turbid water without prior knowledge," *Opt. Lasers Eng.* **148**, 106777 (2022).
29. H. Hu, Y. Zhang, X. Li, Y. Lin, Z. Cheng, and T. Liu, "Polarimetric underwater image recovery via deep learning," *Opt. Lasers Eng.* **133**, 106152 (2020).
30. Y. Xiang, Q. Ren, and R.-P. Chen, "A neural network for underwater polarization dehazing imaging," *Proc. SPIE* **11897**, 1189702 (2021).
31. H. Hu, Y. Lin, X. Li, P. Qi, and T. Liu, "IPLNet: a neural network for intensity-polarization imaging in low light," *Opt. Lett.* **45**(22), 6162–6165 (2020).
32. X. Li, H. Li, Y. Lin, J. Guo, J. Yang, H. Yue, K. Li, C. Li, Z. Cheng, and H. Hu, "Learning-based denoising for polarimetric images," *Opt. Express* **28**(11), 16309–16321 (2020).
33. J. Zhang, J. Shao, J. Chen, D. Yang, B. Liang, and R. Liang, "PFNet: an unsupervised deep network for polarization image fusion," *Opt. Lett.* **45**(6), 1507–1510 (2020).
34. P. Qi, X. Li, Y. Han, L. Zhang, J. Xu, Z. Cheng, T. Liu, J. Zhai, and H. Hu, "U2R-pGAN: Unpaired underwater-image recovery with polarimetric generative adversarial network," *Opt. Lasers Eng.* **157**, 107112 (2022).
35. H. Liu, Y. Zhang, Z. Cheng, J. Zhai, and H. Hu, "Attention-based neural network for polarimetric image denoising," *Opt. Lett.* **47**(11), 2726–2729 (2022).
36. Z. Dong, D. Zheng, Y. Huang, Z. Zeng, C. Xu, and T. Liao, "A polarization-based image restoration method for both haze and underwater scattering environment," *Sci. Rep.* **12**(1), 1–11 (2022).
37. J. L. Pezzaniti and R. A. Chipman, "Mueller matrix imaging polarimetry," *Opt. Eng.* **34**(6), 1558–1568 (1995).
38. B. Bai, Y. Li, Y. Luo, X. Li, E. Çetintas, M. Jarrahi, and A. Ozcan, "All-optical image classification through unknown random diffusers using a single-pixel diffractive network," *Light Sci. Appl.* **12**(1), 69 (2023).
39. E. Ashalley, K. Acheampong, L. V. Besteiro, P. Yu, A. Neogi, A. O. Govorov, and Z. M. Wang, "Multitask deep-learning-based design of chiral plasmonic metamaterials," *Photonics Res.* **8**(7), 1213–1225 (2020).
40. X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," *Advances in neural information processing systems* **28** (2015).
41. S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, "Res2net: A new multi-scale backbone architecture," *IEEE Trans. Pattern Anal. Mach. Intell.* **43**(2), 652–662 (2021).
42. L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv*, arXiv:1706.05587 (2017).
43. X. Chen and Y. Huang, "Memory-Oriented Unpaired Learning for Single Remote Sensing Image Dehazing," *IEEE Geosci. Remote Sens. Lett.* **19**, 1–5 (2022).
44. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 770–778.
45. Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Comput.* **31**(7), 1235–1270 (2019).
46. S. Van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu, "scikit-image: image processing in Python," *PeerJ* **2**, e453 (2014).