Optics and Lasers in Engineering 178 (2024) 108194

Contents lists available at ScienceDirect

Optics and Lasers in Engineering

journal homepage: www.elsevier.com/locate/optlaseng

Improved polarization scattering imaging using local-global context polarization feature learning framework

Xueqiang Fan, Weiyun Chen, Bing Lin, Peng Peng, Zhongyi Guo *

School of Computer and Information, Hefei University of Technology, Hefei 230009, China

ARTICLE INFO

Keywords:
Polarization imaging
Scattering medium
Swin transformer
Convolutional neural networks
Local-global context polarization feature

ABSTRACT

Polarization scattering imaging (PSI) is an extremely challenging problem due to the fact that scattering media can lead to severe degradation of the object information. In this paper, we propose a novel high-performance computational method, the PSI based on a well-designed local-global context polarization feature learning (LGCPFL) framework, named PSI-LGCPFL, to efficiently retrieve the target's information. In the LGCPFL framework, the dilated convolutional neural network and Swin Transformer are concurrently introduced to capture local polarization feature information and long-range polarization dependencies in the polarization scattering images, respectively. We in-depth investigate the relationship between the geometry of target, the scattering imaging distance, and the constituent materials and the PSI quality. The performance of our proposed PSI-LGCPFL is benchmarked and evaluated on three testing datasets established in real-life scenarios, *i.e.*, STR24, DIS50 and MAT18, which cover untrained target's geometry, various untrained targets lying in untrained distances between scattering medium and targets, and diverse untrained target's materials with different background materials, respectively. The experimental results demonstrate that the proposed PSI-LGCPFL has a superior performance on retrieving the target's information with high-generalization abilities and reasonably inferring speed, and outperforms several existing state-of-the-art methods. To our knowledge, PSI-LGCPFL is the first approach to achieve the PSI by polarization characteristics and a deep learning model with Swin Transformer. It also highlights the prospect of accurately reconstruction of remote sensing target's information in scattering medium via using polarization information and deep learning

1. Introduction

Imaging through scattering medium plays an important role in a wide variety of applications, such as ocean exploration [1], remote sensing observation [2], biomedical imaging [3], autonomous driving [4], and other fields [5–8]. Hence, it is invaluable and of great interest to achieve clear vision in scattering media. However, when light travels through the scattering media, *e.g.*, haze [9], turbid water [10,11], biological tissues [3], and gross glass [12], it will result in severe degradation of the target information, which will subsequently lead to target detection failure [13]. Traditionally, target information of interest can be reconstructed by the classical optical experimental methods, *i.e.*, optical coherence tomography [14], wavefront shaping [15], correlated imaging [16], transmission matrices [17], and point spread function [18]. Nevertheless, these techniques are cost-intensive, laborious, and time-consuming. In view of this situation, it is highly desirable and important to develop cost-effective computational methods for high-performance and accurate imaging through scattering media.

Over the past decades, tremendous efforts have been made to uncover the intrinsic physical mechanism of imaging through scattering medium. In particular, polarization imaging has been proven to be a remarkably effective approach for target reconstruction in scattering media, depending on the multidimensional characteristics of the polarization information [13,19-22]. Up until now, a series of polarization information-assisted imaging methods have been developed. Generally speaking, these existing methods can be broadly classified into two categories, which are physical model-based and deep-learning-based methods, respectively. In the early stage, physical model-based methods, such as polarization difference imaging [10,23,24], dominated in the field of imaging in scattering media. These physical model-based methods generally concentrate on the accurate estimation of medium transmission. For instance, in Schechner *et al.*'s method [10], two orthogonal polarization images, which are taken in the same scene, are employed to realize the underwater image reconstruction. In Zhao *et*

* Corresponding author.
E-mail address: guozhongyi@hfut.edu.cn (Z. Guo).

https://doi.org/10.1016/j.optlaseng.2024.108194
Received 18 December 2023; Received in revised form 18 February 2024; Accepted 14 March 2024

al.'s method [25], the genetic algorithm is applied to scan the degree of linear polarization (DoLP) of the target light and the backscattered light to obtain the target light images with the highest contrast. In Hu *et al.*'s method [26], Stokes parameters, which are calculated from several images photographed by a polarizer in specific orientations, are used to realize the polarization image recovery. Whereas, applying physical model widely to reconstruct target in scattering media inevitably suffers from certain disadvantages. For example, the performance of physical model-based polarization imaging methods heavily relies on the knowledge related to the physical characteristics of the scene. Furthermore, physical model-based polarization imaging methods are routinely grounded in idealized conditions, making them difficult to scale and limiting potential applications.

To overcome the shortcomings of physical model-based methods, a few deep learning (DL) techniques that utilize multi-layer artificial neural networks to learn tasks, have been successfully applied to solve computational imaging problems, including polarization image recovery, such as AOD-Net [27], DualCNN [28], IPLNet [29], PFNet [30], PDRDN [31], MU-DLN [32], U2R-pGAN [33], SAM-MIU-net [34]. These methods often employ only one polarization image or multiple polarization images to achieve the reconstruction of polarized images with one or more DL algorithms, such as convolutional neural networks (CNN) [35,36], fully-connected hidden layers [37], and support vector machine [38]. For example, in PFNet [30], an unsupervised deep neural network combined intensity and DoLP images together to address the polarization image fusion problem. In PDRDN [31], to capture sufficient polarization information, four angles of polarization pictures, *i.e.*, 0°, 45°, 90°, and 135°, are fed into PDRDN, which is mainly constructed by residual dense network [39], to eliminate noise for polarization underwater images. In SAM-MIU-net, a self-attention module in multi-scale improved U-net, which can extract a new combination of multi-dimensional information from targets, is proposed to further enhance PSI.

Nevertheless, despite achieving the efficiency and accuracy, the existing methods more-or-less have several critical deficiencies. It can be concluded that all the above methods only consider local polarization feature information using convolution operation, whereas ignoring long-range polarization dependencies. Also, from the perspective of polarization properties, the relationship between the scattering imaging distance, and the constituent materials and the polarization scattering imaging quality needs to be explored in-depth, rather than just considering the geometry of the target. Motivated by these discoveries, in this paper, we propose a novel PSI method, based on a local-global context polarization feature learning (LGCPFL) framework, named as the PSI-LGCPFL, which can further enhance PSI's recovery performances efficiently. Compared with previous PSI methods, the novelties of our work can be briefly summarized as follows:

i. We model polarization imaging though scattering media as an inverse imaging problem computationally and proposed a novel high-performance PSI method called PSI-LGCPFL based on LGCPFL framework trained from a self-built dataset.
ii. Unlike existing PSI methods, a newly customized LGCPFL framework, which is equipped with local context polarization feature learning branches (LCPFLB) and global context polarization feature learning branches (GCPFLB), is designed to capture the relationship between the object intensity image and the polarization information well from the sole DoLP map.
iii. At present, due to the extreme difficulty of obtaining polarization images in a scattering environment, there are no or a paucity of publicly available datasets with richer real-world scenarios. Hence, we establish one scattering system under active light illumination and construct one polarization scattering image dataset, covering more abundant real scenarios, which facilitates further development of PSI techniques.
iv. We validate the efficacy of PSI-LGCPFL through extensive experiments on the challenging real image datasets. Benchmarking results demonstrate that the PSI-LGCPFL achieves competitive performance on retrieving the target's information with high-generalization abilities and outperforms several existing state-of-the-art approaches, highlighting its promising potential in solving the PSI problem.

2. Method

The proposed PSI-LGCPFL consists of three steps of establishing one benchmark dataset, extracting polarization information encoding by DoLP, and LGCPFL model training, where the flowchart is depicted in Fig. 1.

2.1. Physical mechanism

Generally, Stokes parameters [40] $S = (\mathrm{I},\ \mathrm{Q},\ \mathrm{U}, \mathrm{V})^T$ is used to represented the polarization information. Each component in the Stokes parameters can be calculated from four polarization images (*i.e.*, $I_{0^\circ}(x,\ y)$, $I_{45^\circ}(x,\ y)$, $I_{90^\circ}(x,\ y)$, and $I_{135^\circ}(x,\ y)$) taken by orienting linear polarizer at 0°, 45°, 90°, and 135° in the same scene. The Stokes parameters can be expressed as:

$$I(x,\ y) = I_{0^\circ}(x,\ y) + I_{90^\circ}(x,\ y) \tag{1}$$

$$Q(x,\ y) = I_{0^\circ}(x,\ y) - I_{90^\circ}(x,\ y) \tag{2}$$

$$U(x,\ y) = I_{45^\circ}(x,\ y) - I_{135^\circ}(x,\ y) \tag{3}$$

$$V(x,\ y) = I_{R^\circ}(x,\ y) - I_{L^\circ}(x,\ y) \tag{4}$$

where $I(x,\ y)$ denotes the total intensity received by the camera, $Q(x,\ y)$ refers to the intensity difference between the horizontal and vertical components, $U(x,\ y)$ is the intensity difference between the 45° and 90° components, and $V(x,\ y)$ means the intensity difference between right-handed and left-handed components. It should be noted that we use linearly polarized light throughout the experiment, thus $V(x,\ y)$=0. The DoLP is chosen as the sole feature source in this paper, which can be simply calculated as:

$$DoLP = \frac{\sqrt{Q(x,\ y)^2 + U(x,y)^2}}{I(x,\ y)} \tag{5}$$

When light travels through the scattering media, the polarization characteristics of the reflected or transmitted light will generally change, and the process can be formulated as

$$\widehat{S} = M \cdot S_{obj} \tag{6}$$

where M is the Muller matrix [41] of the scattering medium, S_{obj} means the Stokes parameters of the incident light, and $\widehat{S}$ represents the Stokes parameters of the outgoing light after interaction with the scattering media. In the case that the concentration of the scattering medium becomes larger or the imaging distance becomes longer, the detector will receive information containing a large amount of speckle, which will affect the imaging quality. The detector is also unable to effectively distinguish between the target and its background, when the target and its background have similar polarization characteristics. Furthermore, the polarization information is very sensitive to the composition of the target material and its structural properties. The reconstruction of the object image can be regard as an inverse problem of imaging in scattering medium, and the process is formulated as the following objective function

$$S_{obj} = M^{-1} \cdot \widehat{S} \tag{7}$$

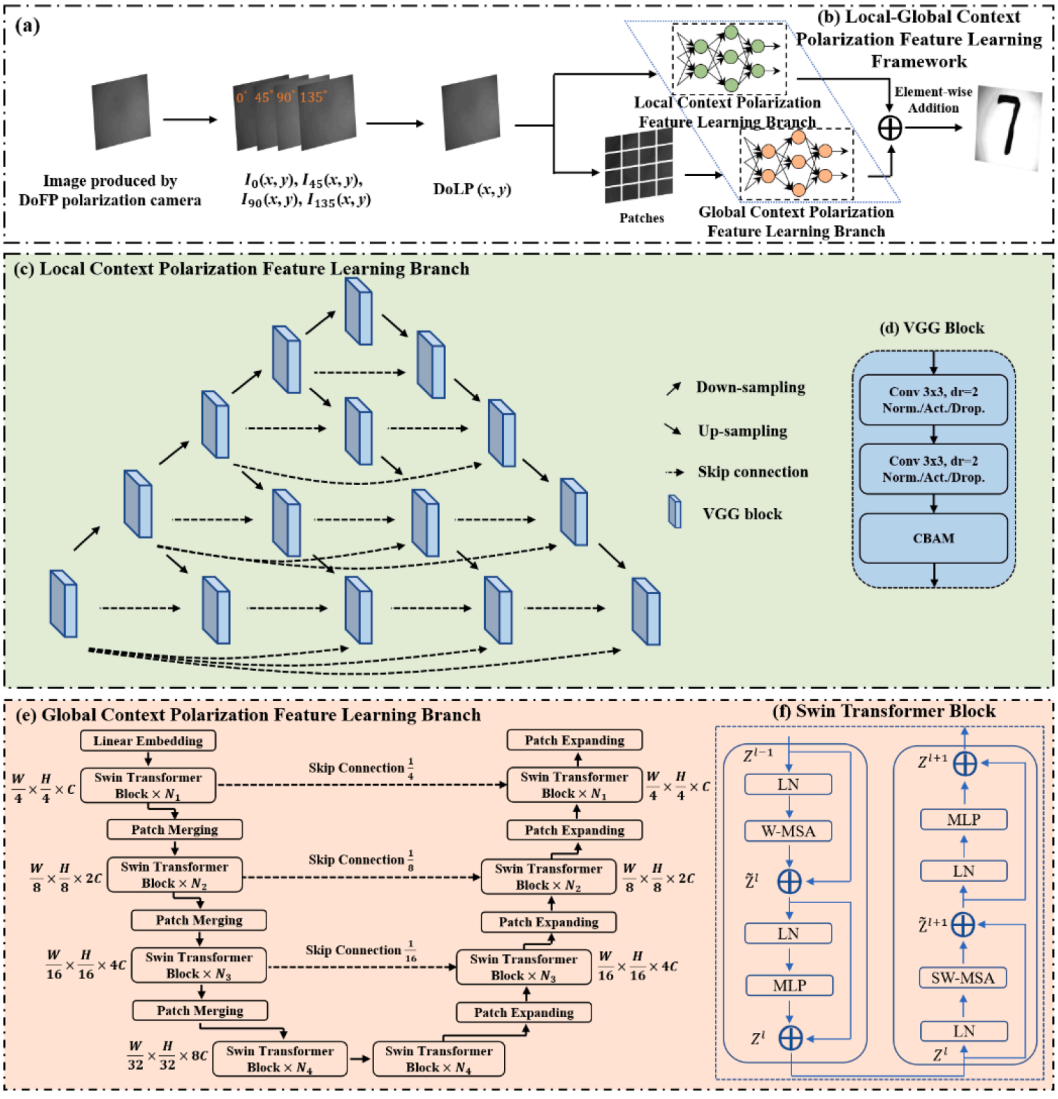

Fig. 1. (a) Pipeline of the PSI-LGCPFL for polarization image reconstruction. (b) The LCPFLB and GCPFLB in the LGCPFL framework. (c) Illustration of the LCPFLB's architecture. (d) Illustration of modified VGG block. "Norm." denotes the batch normalization layer, "Act." denotes the linear rectification activation function (relu), and "Drop." denotes the dropout ratio. (e) Illustration of the GCPFLB's architecture. (f) Illustration of Swin Transformer block.

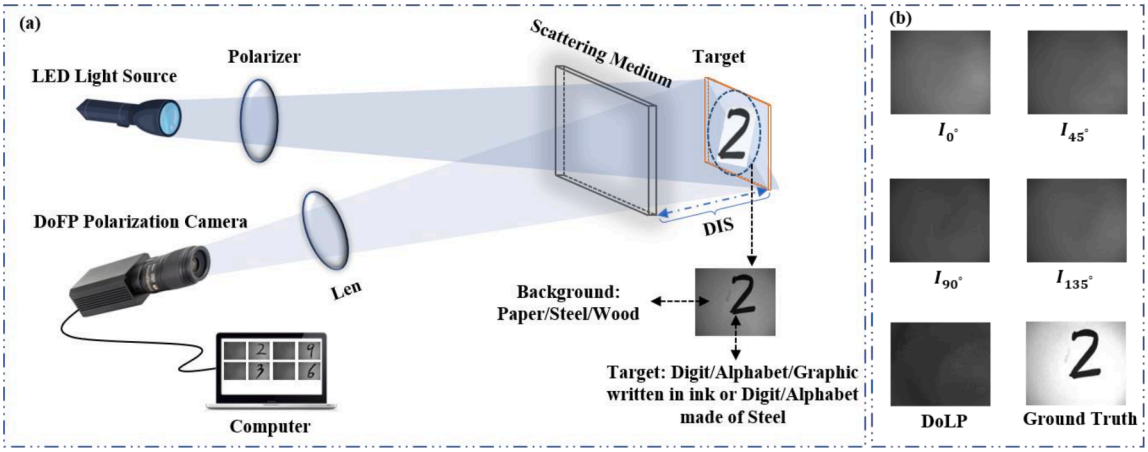

Fig. 2. (a) Experimental setup for the PSI through scattering medium, (b) The obtained images ($I_{0^\circ}(x, y)$, $I_{45^\circ}(x, y)$, $I_{90^\circ}(x, y)$, and $I_{135^\circ}(x, y)$) from the DoFP polarization Camera, the calculated DoLP image, and the ground truth.

where M^{-1} is the inverse of M. Hence, to efficiently address the inverse problem, we attempt to couple the polarization physical priors with DL techniques to map the low-quality original speckle images to high-quality retrieved images from the statistical distribution of polarization information in the training dataset.

2.2. Experimental setup

High-quality benchmark datasets are important to establish reliable polarization image recovery models. By revisiting existing DL-based methods, it can be found that DL models of some existing methods, *e.g.*, Modified U-Net based DL network [32], are trained on one synthetic dataset. Although these methods have achieved considerable progresses in reconstructing polarization images, their applicability is still difficult in real world. Hence, building a real-life scenarios dataset to directly verify the effectiveness of polarization image recovery methods is highly desired and can be well complementary to optical experimental methods.

Here, we establish a real-world polarization scattering imaging system under active light illumination, and the schematic diagram of the experimental setup is illustrated in Fig. 2(a). A linear polarizer is employed as the polarization state generator to provide polarized illumination in front of the LED light source. The ground glass is utilized as the scattering medium device. The light is focused by means of a convex lens. DIS means the distance between the scattering medium and the objects. One real-world polarization dataset is established using a commercial division of focal plane (DoFP) polarization camera (Model type: LUCID, PHX055S-PC), whose spatial resolution is 2448 × 2048. For each single shot, DoFP polarization camera can photograph 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. Using the digit written in ink 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 ground truth.

2.3. The LGCPFL framework

DL has recently been proven to be a fascinating algorithm and successfully applied in a wide variety of computational imaging domains, including polarization imaging in scattering media. Here, a novel custom-made LGCPFL framework is established to achieve polarization image recovery. The overall architecture of our proposed LGCPFL framework is presented in Fig. 1. LGCPFL framework consists of two main modules, LCPFLB (shown in Fig. 1(c)) and GCPFLB (shown in Fig. 1(e)).

In Fig. 1(c), the original DoLP image signal is entered into the LCPFLB to distill local contextual feature of polarization information. The architecture of LCPFLB, whose overall design follows the UNet++ [42] model, consists of a series of down-sampling layers to derive increasing complex context feature representation of the input image, followed by a sequence of up-sampling layers. This structure fuses the high-resolution feature maps from the encoder network with the corresponding semantic-rich feature maps from the decoder network to capture the structural and detailed polarization features of the target more efficiently. Multiple lateral connections between down-sampling and up-sampling layers are employed to fill in contextual information [42]. Each down-sampling layer or up-sampling layer is implemented by using modified VGG block [43] (shown in Fig. 1(d)). In the modified VGG block, it first integrates two dilated convolutional layers with kernel size of 3 × 3 and the dilation rate of 2 to mine discriminative information. After each dilated convolutional layer, batch normalization [44] and ReLU activation function [45] are employed. A dropout [46] strategy with a ratio of 25 % is utilized to avoid overfitting. Furthermore, one convolutional block attention module [47] is embedded the modified VGG block to capture key position information.

Convolutional kernels used in LCPFLB naturally integrate key information from nearby context areas. Since long-range dependencies existing in a whole image should influence the polarization properties, we also design the GCPFLB module to integrate distant information using the Swin Transformer-based expansion architecture inspired by Swin-unet model [48]. As shown in Fig. 1(e), the input of GCPFLB is a series of patches of the original DoLP image. The architecture of GCPFLB mainly includes three modules, *i.e.*, down-sampling layers based on Swin Transformer blocks [49] (shown in Fig. 1(f)) and one patch merging layer, up-sampling layers based on Swin Transformer blocks [49] and one patch expanding layer, and single skip connection between down-sampling and up-sampling layers [48]. In GCPFLB, the feature dimension of each patch is first projected into custom dimension (*i.e.*, C) using a linear embedding layer. Second, the transformed patches are entered into several down-sampling layers to produce the hierarchical features, followed by a series of up-sampling layers. Next, the hierarchical features are merged with multi-scale features from up-sampling layers via single skip connection. Finally, the last patch expanding layer is used to generate the output of GCPFLB. The output of GCPFLB incorporates that of LCPFLB to reconstruct the polarization image.

2.4. Model implementation

All experiments are performed on Linux Server (Version 5.15.0–41) Intel Core i7–7700 CPU @3.6Hz 32.0GB of RAM, and Python 3.7 programming. One training dataset TR200 and three independent validation datasets, *i.e.*, STR24, DIS50 and MAT18 (see details in the 'Benchmark Dataset' section), are employed to train and test the model of PSI-LGCPFL, respectively. In the training process, we use Scikit-image Python library to enhance the training dataset by rotating and flipping the existing images. Furthermore, all images are adjusted to a fixed size, *i.e.*, 256 × 256, when fed into the networks, and the pixel values will undergo normalized to [0, 1].

The model of PSI-LGCPFL is implemented in PyTorch library (Version 1.7.1) [50] and trained by using the Adam optimization algorithm. The initial learning rate is set as 0.0005 and 0.0001 for the first 240 epochs and the last 60 epochs respectively. To speed up training, the models are trained on one graphics processing unit (GPU: Nvidia GeForce RTX 3090) with a batch size of 10. In the model training process, the mean squared error function (MSE) is used to calculate the loss, which is calculated by:

$$MSE = \frac{\sum_{i=1}^{H}\sum_{j=1}^{W}(R(i,\ j) - G(i,j))^2}{H \times W} \tag{8}$$

where H and W are the length and width of the picture; $R(i,\ j)$ and $G(i,j)$ mean the pixel value of ith row and jth column in the reconstructed images and ground truths, respectively. We use the strategy of a grid search and adjust the network's hyper-parameters, *i.e.*, N_1, N_2, N_3, and N_4, by observing the model performance on the training dataset over five-fold cross-validation tests. Finally, according to the best performance of PSI-LGCPFL model, we use the following values for the above hyper-parameters: N_1=2, N_2=6, N_3=2, and N_4=1.

2.5. Evaluation metrics

Two widely used evaluation indexes, *i.e.*, Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM), are employed to assess the quality for image recovery. PSNR is applied to quantify the content between the reconstructed images and ground truths, and a higher PSNR value represents closer image content. SSIM is used to evaluate the structure and texture between the reconstructed images and ground truths, and a higher SSIM value reflects a more similar structure and texture. The two indexes can be calculated by following equations:

$$PSNR = 10log_{10}\left(MAX^2 \,/\, MSE\right) \tag{9}$$

$$SSIM = \frac{(2\rho_R\rho_G + C_1)(2\beta_{RG} + C_2)}{(\rho_R^2 + \rho_G^2 + C_1)(\beta_R^2 + \beta_G^2 + C_2)} \tag{10}$$

where MAX is the maximal in the image; ρ_R and β_R are the corresponding average and variance values of the reconstructed image R, respectively; ρ_G and β_G are the corresponding average and variance values of the ground true image G, respectively; β_{RG} is the variance value between R and G; C_1 and C_2 are small positive constants utilized to avoid a null denominator.

3. Result and discussion

3.1. Benchmark dataset

To build one benchmark dataset, we perform polarization imaging experiments in scattering system under active light condition, and the schematic diagram of the experimental setup is illustrated in Fig. 2. A linear polarizer is used as the polarization state generator to provide polarized illumination $S = (1,\ 1,\ 0,\ 0)^T$ in front of an LED light source that emits light. Firstly, the object of interest is placed away from the DoFP polarization camera to produce ground truths. Note that at this time no scattering media device is positioned in front of the object. Then, the scattering media device, *i.e.*, ground glass, is placed in front of the object and degraded images are taken as original feature sources. Finally, we experimentally establish one benchmark dataset, which is composed of one training dataset and three independent validation datasets. Concretely, our training dataset (named TR200) consists of 200 pairs of images generated with a ground glass at a distance of 40 mm from the targets, and the targets are only a variety of handwritten digits written in ink on the white paper. For the validation performance of our proposed PSI-LGCPFL, the following three independent validation datasets are included:

(i) Validate the performance of PSI-LGCPFL owing to the change in untrained objects with different structures. It consists of 24 pairs of images produced with a distance of 40 mm between ground glass and the targets. The targets contain three categories, *i.e.*, untrained handwritten digits, handwritten alphabets, and handwritten graphic patterns, written in ink on the white paper. For descriptive purposes, this independent validation dataset is named as STR24.

(ii) Validate the performance of PSI-LGCPFL owing to the untrained objects about untrained distances between the ground glass and the targets. It consists of 50 pairs of images produced with distance of 35 mm, 40 mm, 42.5 mm, 45 mm and 50 mm between ground glass and the targets. The targets include three categories (*i.e.*, untrained handwritten digits, handwritten alphabets, and handwritten graphic patterns) written in ink on the white paper. For descriptive purposes, this independent validation dataset is named as DIS50.

(iii) Validate the performance of PSI-LGCPFL owing to the change in untrained object materials with untrained background materials. It is composed of 18 pairs of images created with a distance of 40 mm between ground glass and the targets. The targets include three classifications, *i.e.*, Paper-Steel, Wood-Steel, and Wood-Ink. The Paper-Steel, Wood-Steel, and Wood-Ink mean digits (or alphabets) made of steel against paper background, digits (or alphabets) made of steel against wood background, and digits (or alphabets) written in ink against wood, respectively. This independent validation dataset is called MAT18, for the sake of description.

3.2. Ablation study

(1) **Comparison of our model LGCPFL with the LCPFLB and the GCPFLB.** To demonstrate the powerful ability of our proposed model, we conduct ablation studies to compare our model LGCPFL with the LCPFLB and the GCPFLB alone on TR200 over five-fold cross-validation tests. Specifically, (i) the training dataset TR200 is randomly divided into five non-overlapping subsets; (ii) in each testing stage, the LGCPFL, LCPFLB, and GCPFLB frameworks train the prediction models on four subsets, meanwhile, the remaining subset is employed to evaluate the performance of the trained models; (iii) the averages of five-fold cross-validation tests are calculated as the final results. Note that for a fair comparison, the inputs of LGCPFL, LCPFLB, and GCPFLB are all DoLP polarization images in this section. The five-fold cross-validation tests results of LGCPFL, LCPFLB, and GCPFLB are summarized in Table 1, respectively. In addition, ten samples are randomly selected from the results of the five-fold cross-validation tests of the training dataset for visual comparison. Fig. 3 displays a visual comparison of recovery results of LGCPFL framework with LCPFLB and GCPFLB frameworks on ten cases.

By observing Table 1, it is easy to find that LGCPFL is consistently superior to LCPFLB and GCPFLB concerning the two evaluation indexes, *i.e.*, SSIM and PSNR. The SSIM and PSNR of LGCPFL are 0.79 and 14.81, which are 8.22 % and 9.38 % higher than those of LCPFLB, and 14.49 % and 59.94 % higher than those of GCPFLB, respectively. The differences in SSIM and PSNR values are both statistically significant, which have p-values$<10^{-7}$ in the Student's t-test, except for the PSNR of LCPFLB. Fig. 3 also exhibits the same phenomenon. Taking ②'s results in Fig. 3 as an example, compared with the LCPFLB framework, LGCPFL achieves the improvement of 9.09 % and 12.36 % on SSIM and PSNR, respectively. Furthermore, the images recovered by the LGCPFL yields better visual effects. The above comparison results can demonstrate that the performance is indeed enhanced after applying our proposed LGCPFL model.

(2) **Comparison of different polarization components**: This section examines to what extents the DoLP polarization component can help to reconstruct the scattering images. Thus, the recovery performance of the LGCPFL framework with six polarization components, *i.e.*, $I_{0^\circ}(x,\ y)$, $I_{45^\circ}(x,\ y)$, $I_{90^\circ}(x,\ y)$, $I(x,\ y)$, and DoLP, are investigated. The performances of our proposed LGCPFL framework with these six polarization components are evaluated by performing five-fold cross-validation tests on the training dataset TR200. Table 2 summarizes the discriminative average performance results of these polarization components. Fig. 4 demonstrates visual comparison of recovery results of the LGCPFL framework with five polarization components on some samples.

From Table 2, we can find that the performance of DoLP is superior to those of other four polarization components in terms of both evaluation indexes. Concretely, compared to the second-best performer I_{0°, the SSIM and PSNR values of DoLP are 0.81 and 15.32, respectively, which are improvements of approximately 3.85 % and 3.10 % over those of I_{0°, respectively. The performance improvement observed for five polarization components is statistically significant as shown by the p-values obtained through paired t-test in Table 2. In Fig. 4, the LGCPFL with DoLP is the sole approach that achieves SSIM$>$0.76 and PSNR$>$14.26 on ten cases. The experimental results demonstrate that the DoLP

Table 1
Performance comparison of recovery results of LGCPFL framework with those of LCPFLB and GCPFLB frameworks on the training dataset TR200 over five-fold cross-validation tests.

Dataset	Method	SSIM *value*	*p*-value	PSNR *value*	*p*-value
TR200	LCPFLB	0.73	6.7×10^{-7}	13.54	2.5×10^{-1}
	GCPFLB	0.69	1.7×10^{-12}	9.26	1.5×10^{-15}
	LGCPFL	0.79		14.81	

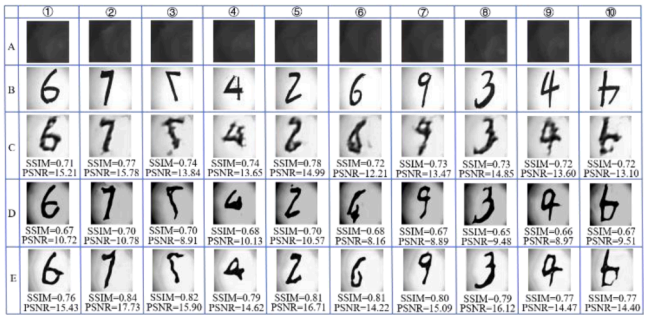

Fig. 3. Visual comparison of the recovery results of LGCPFL framework with those of LCPFLB and GCPFLB frameworks on ten samples selected from the results of the five-fold cross-validation tests of training dataset TR200. (A) the scattering DoLP image of input; (B) ground truths; (C) the images recovered by LCPFLB; (D) the images recovered by GCPFLB; (E) the images recovered by LGCPFL.

Table 2
Comparison of recovery results of different polarization components with LGCPFL framework on the training dataset TR200 over five-fold cross-validation tests.

Method	Feature	SSIM *value*	*p*-value	PSNR *value*	*p*-value
LGCPFL	I_{0°	0.78	3.4×10^{-3}	14.86	2.6×10^{-2}
	I_{45°	0.73	6.2×10^{-5}	12.39	1.1×10^{-8}
	I_{90°	0.75	1.6×10^{-2}	14.08	3.4×10^{-3}
	I	0.76	5.3×10^{-4}	13.46	7.5×10^{-3}
	DoLP	0.81		15.32	

containing polarization feature information is more helpful for the recovery of scattering images under scattering medium conditions.

3.3. Performance analysis

(1) **Performance on untrained target with untrained geometry structures.** The purpose of this section is to experimentally demonstrate the efficacy of the proposed PSI-LGCPFL by performing independent validation tests on dataset STR24 consisted of 8 untrained handwritten numeric targets, 8 untrained handwritten alphabet targets, and 8 untrained handwritten graphic pattern targets. To evaluate the recovery ability of PSI-LGCPFL more fairly and objectively on untrained targets, the experimental design criteria are set as following: (i) the prediction model of PSI-LGCPFL for all the following independent verification tests are constructed on the training dataset TR200, and the targets in the training dataset are only simply digits written in ink on the white paper (see details in the 'Benchmark Dataset' section); (ii) the target generation in the independent validation dataset STR24 and the training dataset TR200 meet the same experimental setup conditions, *e.g.*, the imaging distance between the ground glass and the objects is set as 40 mm, except for more complexity of target shape in the testing dataset; (iii) the results of PSI-LGCPFL model are reported on STR24, and SSIM and PSNR are used to validate the performance of PSI-LGCPFL model.

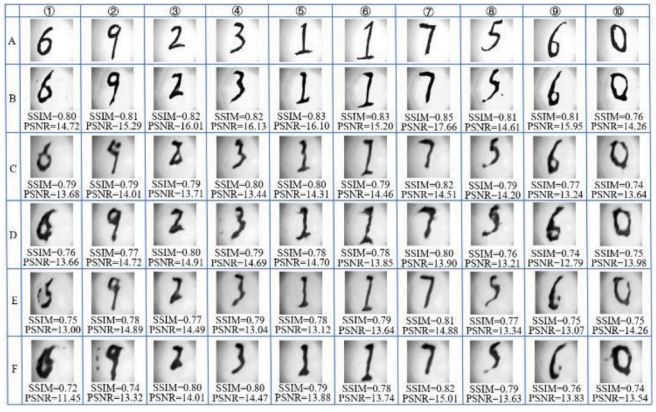

Fig. 4. Visual comparison of the recovery results of different polarization components on the training dataset TR200 over five-fold cross-validation tests using the LGCPFL framework. (A) ground truths; (B) the images recovered by LGCPFL with DoLP; (C) the images recovered by LGCPFL with I_{0°; (D) the images recovered by LGCPFL with I_{45°; (E) the images recovered by LGCPFL with I_{90°; (F) the images recovered by LGCPFL with I.

We first conduct validation experiments on untrained handwritten digits. Fig. 5 shows the visual comparison between recovery results of PSI-LGCPFL and corresponding ground truths for untrained digit targets. Fig. 5A, B, and C present the scattering DoLP images, ground truth, and the reconstructed results, respectively. In Fig. 5, it is clearly demonstrated that PSI-LGCPFL gains relatively superior reconstruction results concerning the two evaluation indexes (SSIM and PSNR) and visual quality. Specifically, out of the 8 targets, there are three cases where PSI-LGCPFL reaches SSIM>0.80 and PSNR>14.65. As expected, well visual details and high-quality structures can be observed between the reconstructed images of PSI-LGCPFL and the corresponding ground truths.

Next, we further empirically examine the generalization and reconstruction performance of the PSI-LGCPFL on 8 untrained handwritten alphabet targets and 8 untrained handwritten graphic pattern targets. The visual comparison between the recovery results of PSI-LGCPFL and ground truths on untrained handwritten alphabets and graphic patterns are demonstrated in Figs. 6 and 7, respectively, in which it is easily found that the PSI-LGCPFL can consistently recover such high-quality images from the scattering images that are difficult to distinguish objects. Using results on handwritten alphabet targets as an example, among the 8 untrained alphabet targets, the PSI-LGCPFL has 5 cases with SSIM>0.75. More specifically, the SSIM and PSNR values, which are important indexes to assess the overall performance of the polarization imaging methods, in ②, ③, ④, ⑤ and ⑥ are 0.79 and 15.01, 0.76 and 14.47, 0.81 and 17.85, 0.76 and 13.55, and 0.75 and 12.80, respectively. Moreover, according to the average SSIM and PSNR of the PSI-LGCPFL for three types of untrained targets on the independent validation dataset STR24 listed in Table 3, we can find that the PSI-LGCPFL has the best performance for the untrained digit targets, followed by untrained alphabet targets and untrained graphic pattern targets.

(2) **Performance on untrained target with untrained imaging distances.** To investigate whether and to what extent in the untrained objects with untrained distances between the ground glass and the objects can influence the performance of the PSI-LGCPFL, we design and conduct a sequence of independent validation tests on the dataset DIS50. The dataset DIS50 consisted of untrained targets with trained/untrained distances is first built. The generalization performance of our proposed PSI-LGCPFL model then is assessed on the dataset DIS50. Table 4 provides a comparison of the predictive performance of the PSI-LGCPFL for various imaging distances. Fig. 8 illustrates a visual comparison of

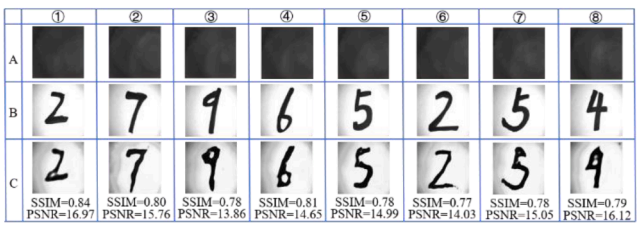

Fig. 5. Visual comparison between the recovery results of PSI-LGCPFL for untrained digit targets and ground truths on the independent validation dataset STR24. (A) the scattering DoLP images of input; (B) ground truths; (C) the images recovered by the PSI-LGCPFL.

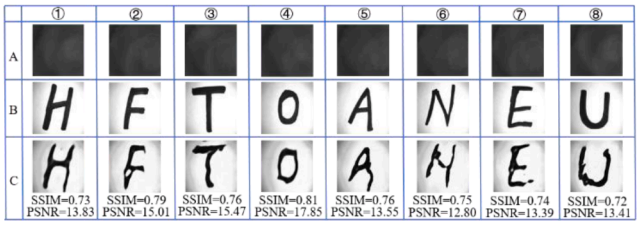

Fig. 6. Visual comparison between the recovery results of PSI-LGCPFL for untrained alphabet targets and ground truths on the independent validation dataset STR24. (A) the scattering DoLP images of input; (B) ground truths; (C) the images recovered by the PSI-LGCPFL.

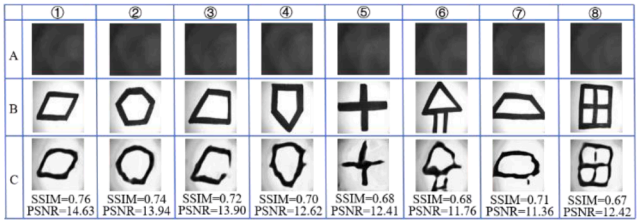

Fig. 7. Visual comparison between the recovery results of PSI-LGCPFL for untrained graphic targets and ground truths on the independent validation dataset STR24. (A) the scattering DoLP images of input; (B) ground truths; (C) the images recovered by the PSI-LGCPFL.

Table 3
Performance of PSI-LGCPFL on the independent validation dataset STR24.

Dataset	Different targets	Digit	Alphabet	Graphic
STR24	SSIM	0.79	0.76	0.71
	PSNR	15.18	14.41	12.76

recovery results of the PSI-LGCPFL for the untrained distances between the ground glass and the objects on the independent validation dataset DIS50.

By carefully observing Table 4, not surprisingly, it is easy to see that the recovery performance of the PSI-LGCPFL at DIS=40 mm outperforms those at DIS=35, DIS=42.5, DIS=45 and DIS=50 mm in terms of SSIM and PSNR. Concretely, compared with DIS=35, the second-best performance's distance from the viewpoint of SSIM, the SSIM and PSNR of PSI-LGCPFL at DIS=40 mm are 0.78 and 15.14, which are 2.63% and 5.80%, respectively, higher than that at DIS=35. Moreover, from Fig. 8, it is easy to find that when DIS=35, 40, and 42.5 mm, the image recovered by the PSI-LGCPFL have well structural integrity. When distance greater than 42.5 mm, although the performance of the PSI-LGCPFL drops slowly in recovering image details, the background and target of the reconstructed image can be effectively distinguished. As expected, the values of two evaluation indexes exhibit the same phenomenon. Taking ③ in Fig. 8 as an example, the SSIM and PSNR values are 0.80 and 15.49, 0.81 and 16.14, 0.77 and 13.87, 0.76 and 14.65, and 0.75 and 13.47, respectively, at DIS=35, 40, 42.5, 45 and 50 mm, which show the superior performance of our proposed PSI-LGCPFL. Meanwhile, it should be noted that for the case at DIS=50 mm, the imaging distance's generalization reaches to 25% that can be an index for demonstrating excellence of our proposed PSI-LGCPFL.

(3) **Performance on untrained target with untrained Materials.**

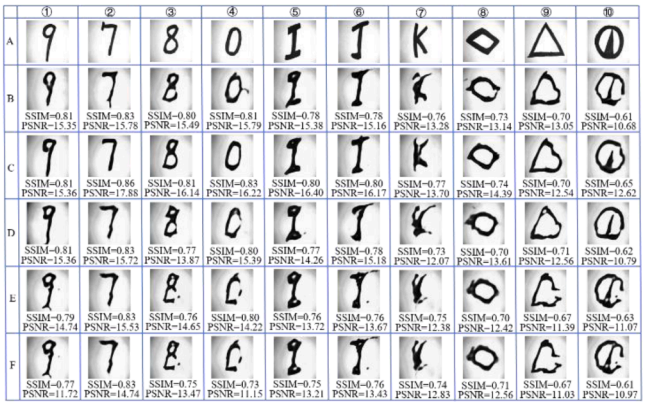

Fig. 8. Visual comparison of the recovery results of PSI-LGCPFL for the untrained distances between the ground glass and the objects on the independent validation dataset DIS50. (A) ground truths; (B) DIS=35 mm; (C) DIS=40 mm; (D) DIS=42.5 mm; (E) DIS=45 mm; (F) DIS=50 mm.

The polarization characteristics of the image signal are heavily influenced by the structural properties and the constituents of the target's material. In this section, to further evaluate the efficacy of proposed PSI-LGCPFL, its generalization performance is also experimentally evaluated on the independent validation dataset MAT18 containing three types untrained targets with diverse background materials, *i.e.*, Paper-Steel, Wood-Ink, and Wood-Steel (see details in the 'Benchmark Dataset' section). We design the following experiments in an orderly manner to fairly and reasonably explore the effects of target materials on the performance of the PSI-LGCPFL. Firstly, we conduct validation experiments on the subset Paper-Steel of MAT18, *i.e.*, untrained digit (or alphabet) targets made of steel against trained paper background, for researching into the polarization characteristic of steel. Secondly, the subset Wood-Ink, *i.e.*, untrained digit (or alphabet) targets made of 'trained' ink against untrained wood background, is used to study the polarization characteristic of wood. Finally, the subset Wood-Steel, *i.e.*, untrained digit (or alphabet) targets made of steel against untrained wood background, is employed to investigate the polarization characteristic of the interaction between steel and wood. Table 5 demonstrates the performance of the PSI-LGCPFL on the independent validation dataset MAT18. Figs. 9, 10, and 11 display visual comparisons of recovery results of PSI-LGCPFL on the subsets Wood-Ink, Wood-Steel, and Paper-Steel, respectively.

Based on the obtained SSIM and PSNR listed in Table 5, we can find that our proposed PSI-LGCPFL show the best performance on the Wood-Ink, followed by Wood-Steel and Paper-Steel. Concretely, among the three subsets, PSI-LGCPFL has the highest values of SSIM and PSNR on the Wood-Ink subset, reaching 0.67 and 13.24, respectively. As described in Figs. 9, 10, and 11, our proposed PSI-LGCPFL can effectively distinguish the target and background. Taking the '8' shape made of steel in Fig. 11(⑥) as an example, despite the SSIM and PSNR of PSI-LGCPFL are 0.44 and 6.65, respectively, the structural outline of the figure '8' can be reconstructed well. It has not escaped from our notice that our proposed PSI-LGCPFL performs poorly on certain targets in the subset Wood-Ink, *e.g.*, the targets of recovery are even more distorted. The reason for this is that the training dataset has only 200 samples. Thus, to improve performance further, constructing a large high-quality dataset will be an effective strategy.

By observing Table 5 and Figs. 9, 10, and 11, we can find an

Table 4
Performance of the PSI-LGCPFL on the independent validation dataset DIS50.

Dataset	Distance (mm)	35	40	42.5	45	50
DIS50	SSIM	0.76	0.78	0.75	0.75	0.73
	PSNR	14.31	15.14	13.88	13.80	12.51

Table 5
Performance of PSI-LGCPFL on the independent validation dataset MAT18.

Dataset	Material	Wood-Ink	Wood-Steel	Paper-Steel
MAT18	SSIM	0.67	0.50	0.47
	PSNR	13.24	5.18	6.34

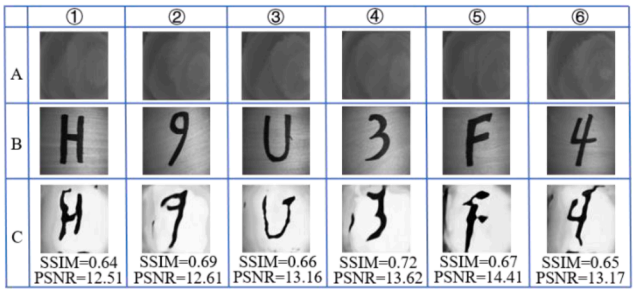

Fig. 9. Visual comparison between the recovery results of the PSI-LGCPFL for the untrained ink targets with untrained wood background material, *i.e.*, Wood-Ink, and ground truths on the independent validation dataset MAT18. (A) the scattering DoLP images of input; (B) ground truths; (C) the images recovered by the PSI-LGCPFL.

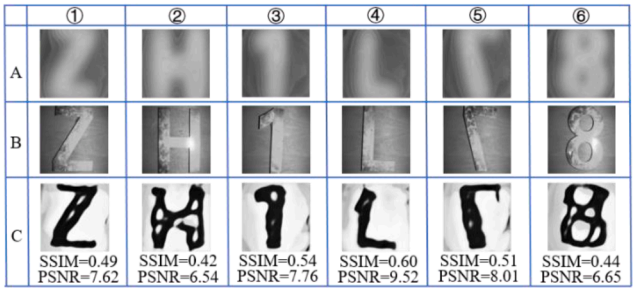

Fig. 10. Visual comparison between the recovery results of PSI-LGCPFL for the untrained steel targets with untrained wood background material, *i.e.*, Wood-Steel, and ground truths on the independent validation dataset MAT18. (A) the scattering DoLP images of input; (B) ground truths; (C) the images recovered by the PSI-LGCPFL.

interesting fact that, although PSI-LGCPFL obtains lower SSIM and PSNR values on both subsets, *i.e.*, Wood-Steel and Paper-Steel, compared with those of Wood-Ink subset, the structural outline and visual quality of the steel targets in the reconstructed images on both subsets Wood-Steel and Paper-Steel by PSI-LGCPFL are superior. This phenomenon can be further speculated that the polarization characteristics of steel is more robustness in scattering medium by reason of its high reflectivity and low deflection characteristics. Meanwhile, the GCPFLB in the PSI-LGCPFL is helpful for reconstructing the targets with wide line width due to its powerful ability of extracting global polarization features.

3.4. Comparison with the state-of-the-art methods

To further evaluate the ability of our PSI-LGCPFL, its performance is also assessed on six representative testing samples, compared to several state-of-the-art methods, including dark channel prior (DCP) [9], Polarimetric-Net [51], MU-DLU [32], and SAM-MIU-net [34]. Of the four methods, the former one, *i.e.*, DCP, is physical model-based methods, while the other three, *i.e.*, Polarimetric-Net, MU-DLU, and SAM-MIU-net are DL-based methods. For an objective and fair comparison, except for non-DL-based methods, the remaining methods are respectively trained on the same training dataset and evaluated on the same testing samples. The comparative results are demonstrated in Table 6 and Fig. 12.

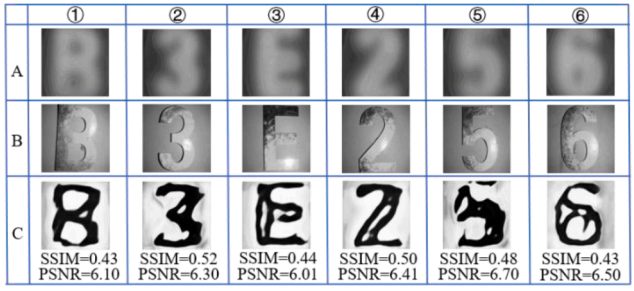

Fig. 11. Visual comparison between the recovery results of the PSI-LGCPFL for the untrained steel targets with trained paper background material, *i.e.*, Paper-Steel, and ground truths on the independent validation dataset MAT18. (A) the scattering DoLP images of input; (B) ground truths; (C) the images recovered by the PSI-LGCPFL.

As described in Table 6, it is easily found that PSI-LGCPFL consistently outperforms other state-of-the-art methods concerning two evaluation indexes. To be more specific, the SSIM and PSNR of PSI-LGCPFL are 0.68 and 11.74, which are 142.86% and 117.01% higher than DCP, 11.48% and 36.35% higher than Polarimetric-Net, 7.94% and 21.28% higher than MU-DLU, and 6.26% and 17.16% higher than SAM-MIU-net, respectively. In addition, the running time of PSI-LGCPFL is measured against other comparative methods. Note that, in order to make the comparison as fair as possible, the running time of PSI-LGCPFL and the control methods are evaluated on the same computational device (Windows Server 10 (Version 21H1) Intel Core i7–9750 CPU @2.60Hz 2.59GHz, and 16.0GB of RAM). Concretely, DCP, Polarimetric-Net, MU-DLU, SAM-MIU-net, and PSI-LGCPFL take 7.73s, 12.16s, 9.12s, 2.69s, and 8.63s, respectively, to reconstruct a polarized image. On the other hand, from the visual-level viewpoint as follows Fig. 12, PSI-LGCPFL yields a higher visual quality compared with other recent methods. It has not escaped from our notice that although SAM-MIU-net make all its efforts to be able to reconstruct some parts of the target geometry, it does not work on various untrained object materials with untrained background materials.

4. Conclusion

We have developed and implemented a novel method, termed PSI-LGCPFL, to reconstruct polarization image from DoLP polarization information. Experimental results have demonstrated the efficacy of the proposed PSI-LGCPFL via validating its performances on three independent validation datasets and comparison to several state-of-the-art PSI methods. The excellent performance of PSI-LGCPFL is mainly attributed to the following reasons: (i) one high-quality benchmark dataset is used; (ii) up to our knowledge, a multi-module network and Swin Transformer block are integrated to achieve optical polarization scattering imaging for the first time; (iii) a well-designed LGCPFL framework can effectively learn a highly expressive optical physical representation between the object radiance and the polarization information.

Despite its good performance, the proposed PSI-LGCPFL still has potential disadvantages. For instance, PSI-LGCPFL may lack the ability to remove noisy information in feature sources during feature preprocessing. Our further research work comprises the following four directions to further enhance the recovery efficacy of polarization image: (i) designing high discriminative feature source; (ii) developing an excellent feature optimization tool to get rid of noise in the feature; (iii) designing a more accurate method by combining PSI-LGCPFL and various macroscopically identical microscopically different scattering medias; (iv) developing useful strategies to extract discriminative feature. Furthermore, 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

Table 6
Performance comparison between PSI-LGCPFL and other methods.

Method	DCP	Polarimetric-Net	MU-DLU	SAM-MIU-net	PSI-LGCPFL
SSIM	0.28	0.61	0.63	0.64	0.68
PSNR	5.41	8.61	9.68	10.02	11.74

Fig. 12. Visual comparison between the recovery results of PSI-LGCPFL and other methods. (*a*) ground truths; (*b*) DCP; (*c*) Polarimetric-Net; (*d*) MU-DLU; (*e*) SAM-MIU-net; (*f*) PSI-LGCPFL.

integration of physics and artificial intelligence could potentially offer a solution to the problem. Finally, although the proposed PSI-LGCPFL still has room for optimization, we believe that it will be exploited as a useful tool to speed up the progress of optical scattering imaging application.

CRediT authorship contribution statement

Xueqiang Fan: Methodology, Validation, Formal analysis, Software, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Weiyun Chen:** Writing – review & editing. **Bing Lin:** Data curation, Writing – review & editing. **Peng Peng:** Writing – review & editing. **Zhongyi Guo:** Conceptualization, Methodology, Formal analysis, Resources, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

This research is funded by the National Natural Science Foundation of China under Grant 61775050. The computation of this research is supported by the HPC Platform of Hefei University of Technology.

References

[1] Li C, Anwar S, Hou J, Cong R, Guo C, Ren W. Underwater Image Enhancement via Medium Transmission-Guided Multi-Color Space Embedding. IEEE Trans Image Process 2021;30:4985–5000.
[2] Yoon S, Kim M, Jang M, Choi Y, Choi W, Kang S, Choi W. Deep optical imaging within complex scattering media. Nat Rev Phys 2020;2(3):141–58.
[3] Wang LV, Wu H-i. Biomedical optics: principles and imaging. John Wiley & Sons; 2012.
[4] Wang HF, Shan YH, Hao T, Zhao XM, Song SZ, Huang H, Zhang JJ. Vehicle-Road Environment Perception Under Low-Visibility Condition Based on Polarization Features via Deep Learning. IEEE trans Intell Transp Syst 2022.
[5] Wolff LB. Polarization vision: a new sensory approach to image understanding. Image Vision Comput 1997;15(2):81–93.
[6] Bailey GN, Flemming NC. Archaeology of the continental shelf: Marine resources, submerged landscapes and underwater archaeology. Quat Sci Rev 2008;27(23):2153–65.
[7] Snik F, Craven-Jones J, Escuti M, Fineschi S, Harrington D, De Martino A, Mawet D, Riedi J, Tyo JS. An overview of polarimetric sensing techniques and technology with applications to different research fields. Polarization: measurement, analysis, and remote sensing XI 2014;9099:48–67.
[8] Dong Y, Wan J, Wang X, Xue JH, Zou J, He H, Li P, Hou A, Ma H. A polarization-imaging-based machine learning framework for quantitative pathological diagnosis of cervical precancerous lesions. IEEE Trans Med Imaging 2021;40(12):3728–38.
[9] He K, Sun J, Tang X. Single image haze removal using dark channel prior. IEEE Trans Pattern Anal Mach Intell 2010;33(12):2341–53.
[10] Schechner YY, Karpel N. Recovery of underwater visibility and structure by polarization analysis. IEEE J Oceanic Eng 2005;30(3):570–87.
[11] Roysam B, Cohen AR, Getto PH, Boyce PR. A numerical approach to the computation of light propagation through turbid media: application to the evaluation of lighted exit signs. IEEE Trans Ind Appl 1993;29(3):661–9.
[12] Tominaga S, Kimachi A. Polarization imaging for material classification. Opt Eng 2008;47(12):123201.
[13] Li X, Han Y, Wang H, Liu T, Chen SC, Hu H. Polarimetric Imaging Through Scattering Media: A Review. Front Phys 2022;153.
[14] Huang D, Swanson EA, Lin CP, Schuman JS, Stinson WG, Chang W, Hee MR, Flotte T, Gregory K, Puliafito CA. Optical coherence tomography. Science 1991;254(5035):1178–81.
[15] Mosk AP, Lagendijk A, Lerosey G, Fink M. Controlling waves in space and time for imaging and focusing in complex media. Nat Photonics 2012;6(5):283–92.
[16] Xu C, Li D, Fan X, Lin B, Guo K, Yin Z, Guo Z. High-performance deep-learning based polarization computational ghost imaging with random patterns and orthonormalization. Phys Scr 2023.
[17] Yoon J, Lee K, Park J, Park Y. Measuring optical transmission matrices by wavefront shaping. Opt Express 2015;23(8):10158–67.
[18] Rossmann K. Point spread-function, line spread-function, and modulation transfer function: tools for the study of imaging systems. Radiology 1969;93(2):257–72.
[19] Schechner YY, Karpel N. Clear underwater vision. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2004. pp. I–I.
[20] Demos S, Alfano R. Optical polarization imaging. Appl Opt 1997;36(1):150–5.
[21] Fan X, Lin B, Guo K, Liu B, Guo Z. TSMPN-PSI: high-performance polarization scattering imaging based on three-stage multi-pipeline networks. Opt Express 2023;31(23):38097–113.
[22] Lin B, Fan X, Peng P, Guo Z. Dynamic polarization fusion network (DPFN) for imaging in different scattering systems. Opt Express 2024;32(1):511–25.
[23] Treibitz T, Schechner YY. Active polarization descattering. IEEE Trans Pattern Anal Mach Intell 2008;31(3):385–99.
[24] Guan J, Zhu J. Target detection in turbid medium using polarization-based range-gated technology. Opt Express 2013;21(12):14152–8.
[25] Zhao Y, He W, Ren H, Li Y, Fu Y. Polarization descattering imaging through turbid water without prior knowledge. Opt Lasers Eng 2022;148:106777.
[26] Hu H, Zhao L, Li X, Wang H, Yang J, Li K, Liu T. Polarimetric image recovery in turbid media employing circularly polarized light. Opt Express 2018;26(19):25047–59.
[27] Xiang Y, Ren Q, Chen RP. A neural network for underwater polarization dehazing imaging, optoelectronic imaging and multimedia technology viii. SPIE; 2021, 1189702.
[28] Gao F, Huang T, Wang J, Sun J, Hussain A, Yang E. Dual-branch deep convolution neural network for polarimetric SAR image classification. Applied Sciences 2017;7(5):447.
[29] Hu H, Lin Y, Li X, Qi P, Liu T. IPLNet: a neural network for intensity-polarization imaging in low light. Opt Lett 2020;45(22):6162–5.
[30] Zhang J, Shao J, Chen J, Yang D, Liang B, Liang R. PFNet: an unsupervised deep network for polarization image fusion. Opt Lett 2020;45(6):1507–10.
[31] Li X, Li H, Lin Y, Guo J, Yang J, Yue H, Li K, Li C, Cheng Z, Hu H. Learning-based denoising for polarimetric images. Opt Express 2020;28(11):16309–21.
[32] Li D, Lin B, Wang X, Guo Z. High-Performance Polarization Remote Sensing with the Modified U-Net Based Deep-Learning Network. IEEE Trans Geosci Remote Sens 2022;60:1–10.
[33] Qi P, Li X, Han Y, Zhang L, Xu J, Cheng Z, Liu T, Zhai J, Hu H. U2R-pGAN: Unpaired underwater-image recovery with polarimetric generative adversarial network. Opt Lasers Eng 2022;157:107112.
[34] Lin B, Fan X, Guo Z. Self-attention module in a multi-scale improved U-net (SAM-MIU-net) motivating high-performance polarization scattering imaging. Opt Express 2023;31(2):3046–58.
[35] Li Z, Liu F, Yang W, Peng S, Zhou J. A survey of convolutional neural networks: analysis, applications, and prospects. IEEE Trans Neural Networks Learn Syst 2021.
[36] Si Z, Yu Y, Yang M, Li P. Hybrid solar forecasting method using satellite visible images and modified convolutional neural networks. IEEE Trans Ind Appl 2020;57(1):5–16.
[37] LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks 1995;3361(10):1995.
[38] Noble WS. What is a support vector machine? Nat Biotechnol 2006;24(12):1565–7.

[39] Zhang Y, Tian Y, Kong Y, Zhong B, Fu Y. Residual dense network for image super-resolution. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2018. p. 2472–81.
[40] Berry HG, Gabrielse G, Livingston A. Measurement of the Stokes parameters of light. Appl Opt 1977;16(12):3200–5.
[41] Pezzaniti JL, Chipman RA. Mueller matrix imaging polarimetry. Opt Eng 1995;34 (6):1558–68.
[42] Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. Unet++: a nested u-net architecture for medical image segmentation, deep learning in medical image analysis and multimodal learning for clinical decision support. Springer; 2018. p. 3–11.
[43] Simonyan K, Zisserman A. arXiv preprint. 2014.
[44] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning. PMLR; 2015. p. 448–56.
[45] Agarap AF. arXiv preprint. 2018.
[46] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15 (1):1929–58.
[47] Woo S, Park J, Lee JY, Kweon IS. Cbam: Convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV); 2018. p. 3–19.
[48] Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, Wang M. arXiv preprint. 2021.
[49] Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B. Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2021. p. 10012–22.
[50] Pajankar A, Joshi A, Network Neural, Basics PyTorch. Hands-on machine learning with python. Springer; 2022. p. 215–26.
[51] Hu H, Zhang Y, Li X, Lin Y, Cheng Z, Liu T. Polarimetric underwater image recovery via deep learning. Opt Lasers Eng 2020;133:106152.