Optics & Laser Technology 181 (2025) 111664

Contents lists available at ScienceDirect

Optics and Laser Technology

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

Full length article

Transformer-based improved U-net for high-performance underwater polarization imaging

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

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

ARTICLE INFO

Keywords:
Polarization
Underwater polarization imaging
Transformer
U-Net

ABSTRACT

Polarization imaging plays an important role in underwater imaging technology, as it allows for the suppression of backscattering light and thus enhances the quality of underwater images. Furthermore, with advancements in computer hardware technology, deep learning (DL) has experienced rapid development in recent years. However, existing methods based on convolutional neural networks (CNN) have the limitations of local feature extraction and fail to full use of the salient target's polarization features. In this paper, we propose a transformer-based improved U-Net (TIU-Net) to further improve the performance of underwater polarization imaging. The proposed TIU-Net leverages CNN and transformer module to extract short-range and long-range features respectively for underwater target reconstruction. Meanwhile, by entering multi-dimensional information, each of which has a different emphasis on the expression of target features, to enhance the performance and stability of the reconstruction model. Experimental results on our established polarization underwater dataset show the superiority of our proposed method for underwater imaging, achieving the efficient imaging and high-quality generalization imaging in highly turbid underwater environments.

1. Introduction

Underwater imaging technology plays a crucial role in comprehending, developing, utilizing and preserving the ocean [1,2]. However, underwater imaging usually has poor performance, due to the scattering and absorption of particles in the water. These particles induce scattering phenomena, thereby deteriorating the final imaging quality. Tremendous efforts in optical theories and experiments, such as wavefront shaping [3], transmission matrices [4], optical coherence tomography [5], and correlated imaging [6–8], have been made to reconstruct target information in scattering media. Notably, polarization, as inherence property of light, performs well against the scattering effects resulting from the particle in the water, due to its unique physical characteristics [9–12]. During the propagation process of light, its state of polarization (SoP) is less susceptible to the scattering media, which makes it possible to effectively suppress the affections from scattering. Polarization imaging [13–15] has been considered an effective method for imaging in scattering environments, including underwater environment. At present, a variety of underwater imaging restoration methods based on polarization imaging have been proposed [16–21]. Y.Y. Schechner has proposed an inversion algorithm based on polarimetric pictures to improve the quality of underwater imaging [18]. However, it only solved the problem in the relatively simple underwater environments. Besides, the considered conditions by the corresponding parameter estimation were always not comprehensive enough to achieve the higher-quality imaging. As a result, it has become one of the schemes to improve the quality of polarized underwater imaging by analyzing the characteristics of polarization information propagation underwater to achieve an accurate estimation of imaging model parameters [19–22] in order to obtain clearer underwater pictures and adapt to more underwater environments. However, there are also some limitations: 1) need for professionals to choose the proper parameters; 2) the inability to obtain outstanding performance in surroundings with more complicated scattering; 3) the rigorous requirements for imaging systems.

In recent years, deep learning (DL) technology has performed very well in the field of imaging through scattering media [23–27]. Convolutional neural networks (CNN), such as GAN [28], ResNet [29], has become the backbone of image processing due to the widespread application of DL technology. Similarly, in the field of polarization imaging, CNN-based underwater polarization dehazing algorithms have also performed well [30]. Furthermore, in view of the problem that the

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

https://doi.org/10.1016/j.optlastec.2024.111664
Received 9 June 2024; Received in revised form 1 August 2024; Accepted 20 August 2024
Available online 26 August 2024

degree of polarization (DoP) pictures and the angle of polarization (AoP) pictures will amplify noise during the acquisition process, the CNN-based processing method also has a good drying effect [31]. Moreover, it achieves noteworthy results to employ CNN-based target detection with DoP and AoP as the features [32–35]. In addition, considering the advantages of polarization information during the transmission process, utilizing the powerful feature extraction ability of CNN and special network structures, such as the attention mechanism, can enhance the expression of target polarization characteristics in the training process to improve the efficacy of target reconstruction in more intricate environments and enhance the generalization capability of the network model [36,37]. There is no denying that the performance of convolution operations has improved significantly compared to traditional methods, and it also overcomes some of constraints of traditional methods. But most CNN-based image reconstruction methods focus more on feature extraction of local details, which is not conducive to feature extraction over long distances. Particularly as the scattering environment grows more complex, the target information in the image requires more layers of convolution operation to adequately assess the target features. The network depth can be extended because the presence of the ResNet structure without encountering issues like gradient disappearing [29], but deeper network structures also mean requiring larger datasets and facing the challenge of how to guarantee generalization. At the same time, the gathering of datasets for target reconstruction is also a very involved process, and to train a huge dataset effectively is also a challenge for hardware devices. As a result, it's critical to fully leverage all available data within an appropriate network architecture.

Zhang et al. have found that the output field's component of ballistic light is directly related to the model's adaptability, and by boosting the detected component of ballistic light, the network's generalization ability can be improved [38]. Importantly, compared to the intensity of light, polarization of light can better suppress the backscattering light while maintaining more of ballistic light component. Therefore, maximizing the benefits of polarization information is one of the effective ways to significantly increase the network's capacity for generalization. So, a crucial question is how to fully exploit the benefits of polarization characteristics while working with limited datasets. Recently, the transformer [39,40] has demonstrated good performance in natural language processing [41,42] and computer vision tasks. Compared to CNNs, transformer-based models can obtain better long-range information. Moreover, the transformer, compared with other structures that can extract information over long distances, can process data in parallel, making it more suitable for processing images with relative smaller consuming resources. Hence, in this paper, by introducing the transformer module to an improved U-Net, we have proposed a flexible transformer-based improved U-Net (TIU-Net) for reconstructing hidden objects with complex underwater scattering states. Combined the CNN and the transformer together for effectively perceiving polarization characteristics for reconstruction goals in underwater with high turbidity. In the case of using the prior from polarization physics of the actual scattering scenes, giving full play to the advantages of polarization information, the training set with one visible scattering environment and one fixed scattering imaging distance (SID) can effectively improve the learning process of the DL.

The structure of this paper is as follows. Chapter 2 introduces the physical basis and network structure of the proposed method, while Chapter 3 presents the parameter settings, experimental results, and analysis. Finally, in Chapter 4, we summarize our work.

2. Methodology

2.1. Polarization information

The modeling and analysis of polarized optical systems are very important for polarization applications. Various polarization characterization methods exhibit distinct sensitivities towards specific features of the target. So, in this section, we will introduce the different polarization features, which is essential for our investigations and analysis. George Gabriel Stokes has proposed that the Stokes vector can be obtained directly by measuring the intensity [43], which can be calculated by receiving the intensity in different directions (i.e., 0°, 45°, 90°, 135° and left-/right-handed circular polarizations) through the detector, and the its mathematical expression can be expressed as follows:

$$S=\begin{bmatrix} I \\ Q \\ U \\ V \end{bmatrix}=\begin{bmatrix} \left\langle E_{0x}E_{0x}^{*}+E_{0y}E_{0y}^{*}\right\rangle \\ \left\langle E_{0x}E_{0x}^{*}-E_{0y}E_{0y}^{*}\right\rangle \\ \left\langle E_{0x}E_{0y}^{*}+E_{0y}E_{0x}^{*}\right\rangle \\ i\left\langle E_{0x}E_{0y}^{*}-E_{0y}E_{0x}^{*}\right\rangle \end{bmatrix}=\begin{bmatrix} I_{0^\circ}+I_{90^\circ} \\ I_{0^\circ}-I_{90^\circ} \\ I_{45^\circ}-I_{135^\circ} \\ I_{R^\circ}-I_{L^\circ} \end{bmatrix} \tag{1}$$

where S stands for the Stokes vector; I is the total intensity; Q is the intensity difference between the horizontal and vertical polarization directions (two orthogonal components), which can suppress backscattering light to some extents [44]; U represents the intensity difference between the 45° and 135° polarization directions; V represents the intensity difference between right-handed and left-handed circular polarization light. Other physical quantities that can represent the polarization information of received light, such as DoP and AoP, can be calculated by the Stokes vector. Among them, for linearly polarized light, the linear degree of polarization (DoLP) can be calculated, which represents the ratio of the components of linear polarizations to the entire light intensity:

$$DoLP=\frac{\sqrt{Q^2+U^2}}{I} \tag{2}$$

The fusion of polarization information and intensity information can provide complementary information (detail information and global information of the image), which will improve the quality and detailed expression of the target image [45]. At the same time, the multi-dimensional information will provide richer feature information for target reconstruction, when the detector captures less target information due to the scattering system being more complex. Therefore, in this paper, we use the intensity (S_0), Q component (S_1), and DoLP information to construct the training dataset and testing dataset.

The complexity of the scattering system can be expressed by the optical thickness (OT), which can be expressed as [46]:

$$OT=\mu\times d \tag{3}$$

where μ is the attenuation coefficient, and d is the path length through the scattering media. In this paper, the scattering media employs a turbid suspension of water-milk system. From Eq (3), it can be known that increasing the SID (increasing d) will increase the OT of the corresponding scattering environment. And then, the scattering system's attenuation coefficient depends on the particle density and radius in the scattering system. In fact, it is still challenging to precise control over particle density and radius, and their variations in the scattering medium can influence the attenuation coefficient at a macroscopic level. So, when we modify the water-milk suspension's concentration to change the turbidity, the scattering system's attenuation coefficient will vary, thereby influencing the OT.

2.2. Measurement system

We obtained the underwater polarization dataset through the following experimental setup as shown in Fig. 1, where (a) is a schematic diagram of the experimental setup and (b) is the actual experimental setup. Our experiments are conducted in a darkroom. In order to simulate the real underwater environment, in this paper, we add different volumes of milk to clean water to simulate the turbid

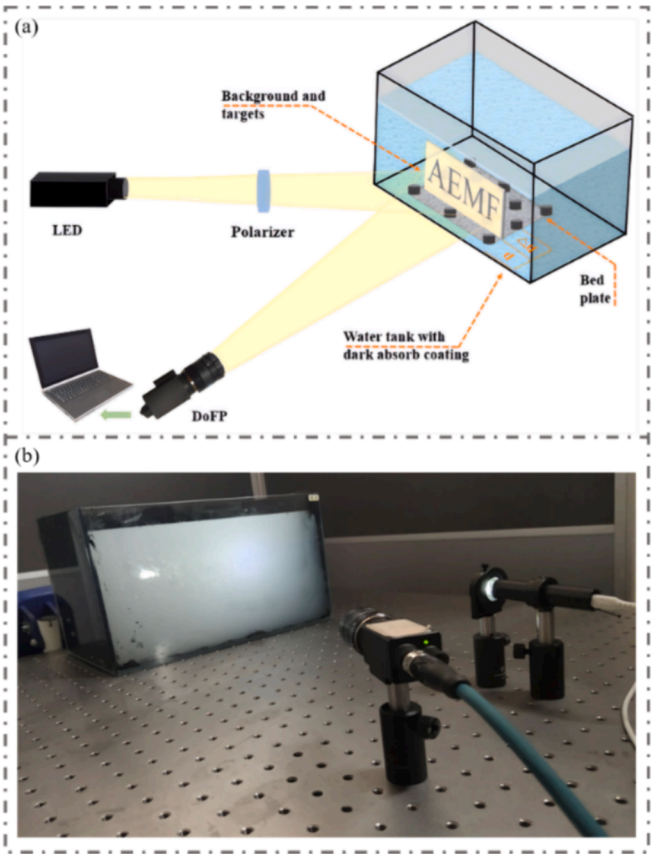

Fig. 1. Experimental setup for underwater polarization imaging. (a) Schematic illustration of the experimental setup; (b) Actual illustration of the experimental setup.

underwater environment to obtain scattering images. We use a clear glass tank as a container, pour clean water with a volume of 340 mm × 190 mm × 140 mm, and paste black shields on three side-walls and the bottom of the glass tank to avoid ambient light interference. The green LED light is selected for active lighting, and a linear polarizer is fixed in front of the LED light to make the horizontally polarized light, that is $S = (1, 1, 0, 0)^T$. In order to obtain the pixel-level corresponding dataset of objects under turbid water, we use building blocks as the bed plate and supporting connectors that can be fixed on the bed plate. The objects adhering on these connectors ensure the concrete position of the targets, which provides a physical basis for the subsequent shooting of pixel-level corresponding datasets. The training targets are made of steel. In addition, we treat the images taken in clear water as the clear image, that is, labels, instead of the image taken in the air as labels. That is because the refractive index of air and water is different, so the field of view of the image will be different when the object is captured in the air and underwater in the same position, which will cause the image not to correspond at the pixel level. To capture the training dataset of turbid underwater, we add 10 ml milk to clean water, stir continuously to make the water evenly cloudy, and fix the target at a distance of 9 cm from the front surface of water, that is, SID of $d = 9$ cm. And in this situation, the turbidity is measured to be 133NTU by the turbidimeter.

In order to make targets for being photographed, we engrave different patterns with a laser on the steel plate. We fixed the target in turbid water in the same order as that in clear water and used the commercial DoFP (division of focal plane) polarization camera (LUCID, PHX055S-PC) with pixel counts of 2048 × 2448 to capture images. Then the images are divided into four different polarization orientations of 0°, 45°, 90°, and 135° respectively, in which the size of each polarization image is 1224 × 1024. The intensity image (S_0), Q-component image (S_1), and DoLP image required for the training set can be calculated by Eq. (1) and Eq. (2) from original polarization images.

2.3. Dataset preparation

The scattering underwater images captured by the camera under incoherent light are shown in Fig. 2. When the OT increases, the captured picture will be more blurred due to increasing scattering effect. The training set is composed of scattering images of S_0, S_1 and DoLP in turbid water, as shown in Fig. 2, and the following different testing data sets are also composed of these data (scattering images of S_0, S_1 and DoLP). We take 110 groups of polarization images, each group with four polarization directions (0°, 45°, 90°, 135°). In addition, we expand them to 2000 to get the training set. Limited by the hardware memory and computing power of training neural networks, the input image size of the neural network in this article is 256 × 256 pixels.

2.4. Network design

As shown in Fig. 3, the backbone of the network is the improved U-Net, which is consisted of dense blocks [47]. And then, the high-resolution spatial information of the target is extracted from the input through the three-layer dense blocks in the part of the front-end encoder, and the first two-layer of dense blocks are followed by down-sampling. After that, the features are entered into the transformer module to integrate the global context features. These features are input to the part of the back-end decoder and combined with high-resolution convolutional features, which are from the front-end encoder, through skip connection.

The transformer module consists of four layers of transformer layer, which is based on the Multi-head Self-Attention (MSA) [39]. Features size H×W×C, which are output by the encoder, are reshaped to a size of $HW/M^{\hat{}}2 \times M^{\hat{}}2 \times C$, and then these are divided into non-overlapping windows of size M×M, before being manipulated by the transformer layer. The structural design of the self-attention is based on our previous work [31]. We perform three attention functions in parallel and concatenate the results as a window multi-head self-attention (WMSA) [48]. Finally, feature transformation is operated by a multilayer perceptron (MLP) with a Gaussian error linear unit (GELU). And LayerNorm (LN) is added before the WMSA and MLP. Then, the global feature and the local feature both are input into the decoder, which consists of dense block and up-sampling. Subsequently, features output from decoder are input into the two layers of convolution, and then the clear target image of 256 × 256 can be obtained from the output of the decoder.

We use MAE as a loss function to drive the interaction of polarization features in the network, which can be expressed as:

$$MAE = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} \|X(i,j) - Y(i,j)\| \tag{4}$$

where X(i, j) represents the pixel of the original target, Y(i, j) represents the pixel of the reconstructed image, and M and N represent the size of the image. We trained the model in an image processing unit (NVIDIA RTX 3090) using a Pytorch framework with Python 3.6. To get the best optimal model, we trained 150 epochs. The optimizer is the Adam (Add Momentum Stochastic Gradient Descent) with a learning rate of 0.0001.

2.5. Imaging quality

In this paper, to assess the quality of the output of the network, we adopt several evaluation metrics, i.e., Pearson Correlation Coefficient (PCC), MAE, Mean Squared Error (MSE), Root Mean Square Error (RMSE) and Peak Signal-to-Noise (PSNR) [49,50]. The PCC with the value between 0–1, is a way to measure the similarity of images. It can be expressed as [51]:

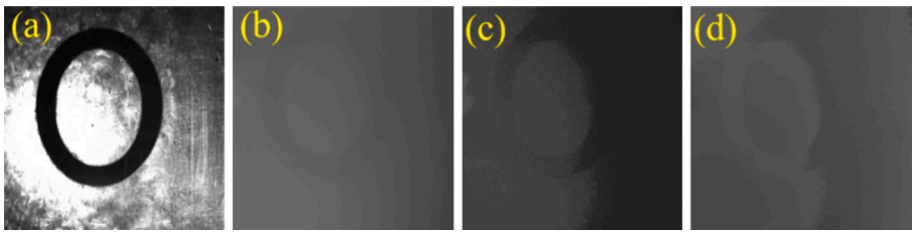

Fig. 2. (a) Original target; (b) Scatting imaging by I (S_0); (c) Scatting imaging by Q (S_1); (d) Scatting imaging by DoLP.

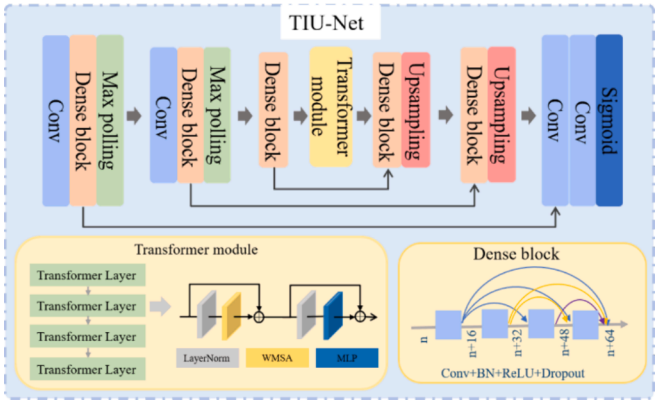

Fig. 3. The overall structure of the proposed TIU-Net.

$$PCC = \frac{\sum_{i=1}^{w}\sum_{j=1}^{h}(Y(i,j)-Y_1)(G(i,j)-G_1)}{\sqrt{\sum_{i=1}^{w}\sum_{j=1}^{h}(Y(i,j)-Y_1)^2}\sqrt{\sum_{i=1}^{w}\sum_{j=1}^{h}(G(i,j)-G_1)^2}} \tag{5}$$

where Y(i, j) represents the pixel of the reconstructed image, G(i, j) represents the pixel of the original target, G_1 and Y_1 represent the mean of the original target and the network reconstruction image respectively, and M and N represent the size of the image. The MSE between the original image and the predicted image (G, Y) of size M×N can be represented as:

$$MSE = \frac{1}{MN}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[G(i,j)-Y(i,j)]^2 \tag{6}$$

The RMSE measures the deviation between the predicted image and the original image and is sensitive to outliers.

$$RMSE = \sqrt{\frac{1}{MN}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[G(i,j)-Y(i,j)]^2} \tag{7}$$

The PSNR is defined as:

$$PSNR = 10 \times \log_{10}\frac{MAX^2}{MSE} \tag{8}$$

where MAX is the maximal value in the image.

3. Result and discussion

In this section, we have designed and explored different experiments to test the trained TIU-Net to demonstrate that our proposed method can take full advantages of polarization information. Then, our method can achieve high-quality and strong-generalization imaging through scattering media in turbid underwater.

3.1. Untrained different-structure targets

One of the indicators for assessing the stability of a reconstructed network model is its ability to accurately predict invisible targets in the training set. To demonstrate the robustness of our model, we selected three sets of same-material targets with varying degrees of correlation to those in the training set and evaluated their reconstruction performance. Firstly, other conditions remain unchanged, we replace the targets with different levels of complexity, which have not appeared in the training set. We set the digital target, alphabetical target, and Chinese-character target separately, and their relevance to the structures of training set decreases in order[52,53]. They are placed in the underwater environment to obtain corresponding scattering polarization images (i.e., S_0, S_1, and DoLP). As can be seen from Fig. 4, the proposed TIU-Net can reconstruct digital targets completely. Even small detail (target "6″) in background can be also reconstructed clearly. Fig. 4(b) is the scattering images of S_0, which can be seen that the target is almost completely masked by the noise. For the sake of conciseness in the image presentation, we only display scattering images of S_0 in the following result display.

Further, in Fig. 5 and Fig. 6, the more complex alphabetical targets and Chinese-character targets can be also reestablished by our proposed TIU-Net effectively. From Fig. 5(c) and Fig. 6(c), it can be seen that the structure of targets is complete and the contrast of images is high. Meanwhile, there is no unnecessary noise in the background of results nearly.

The proposed method is capable of recovering untrained targets of varying complexity, which proves the ability of our network to reconstruct targets got in turbid underwater with high quality. In Table 1, more than 80 % similarity has been achieved between the recovery outcomes for various structures and the matching original pictures. Additionally, from Table 1, the difference among other indicators (MAE, MSE, RMSE and PSNR) of the various objectives varies slightly. It can be seen that the TIU-Net not only achieve high-quality imaging, but also demonstrates well stability. Even the Chinese-characters, whose structures are least relevant to the training set, can be accurately reconstructed with the PCC of more than 82 %, and other relevant data are also satisfactory.

3.2. Untrained different-material targets

In addition, we set targets with different-materials compositions, such as "Steel-Wood" (with Steel as the target and Wood as background)

Fig. 4. The test results of untrained digital targets. (a) Ground truth; (b) The scattering images (S_0); (c) Reconstructed images.

Fig. 5. The test results of untrained alphabetical targets. (a) Ground truth; (b) The scattering images (S_0); (c) Reconstructed images.

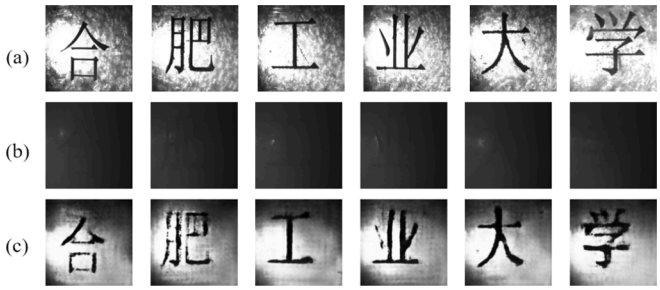

Fig. 6. The test results of untrained Chinese-character targets. (a) Ground truth; (b) The scattering images (S_0); (c) Reconstructed images.

Table 1
The average evaluation indicators of the different targets.

	Digital targets	Alphabetical targets	Chinese-character targets
PCC	0.8488	0.8389	0.8220
PSNR	15.5459	14.6976	14.4135
MAE	0.1136	0.1304	0.1353
MSE	0.0300	0.0358	0.0397
RMSE	0.1704	0.1878	0.1975

and "Paper-Steel" (with Paper as the target and Steel as background), which are different from the materials' compositions of the training set. All else being equal, the camera obtains the corresponding polarized scattering images. Table 2 shows the special elements of the Muller matrix (MM) for the used material [54,55]. The differences in polarization properties between different materials can be expressed through the parameters of MM. The test results can be found in Fig. 7. From Table 2, the steel has high reflectivity and low depolarization characteristics, and wood has high depolarization and low reflectivity. Meanwhile, there is the steel in the targets of the training set. So, when setting targets as "Steel-Wood", the steel part can be reconstructed and relatively bright, while the polarization characteristics of wood are weak, and the background part is not completely reconstructed and there is more noise. When targets are set as "Paper-Steel", the reconstruction of paper part is incomplete, due to its weaker polarization characteristics and large differences from the polarization characteristics of steel, which can be seen from Table 2. While the background part of steel reconstruction is relatively bright and has less noise. In Table 3, evaluation indicators for different materials have been listed and demonstrated, in which it can be seen that the "Steel-Wood" has better performance on data because the steel targets are partially rebuilt. For the "Paper-Steel", although its background is reconstructed, the target is composed of material with a large gap from the training set. So, the PCC of the reconstruction results is only about 46 %, and other indicators are relatively less than those of the "Steel-Wood".

Table 2
The Muller matrix elements of the materials utilized in the experiment [54,55].

Material	m22	m44
Steel	0.975	0.990
Paper	0.265	0.247
Wood	0.215	0.160

From generalization results for different-materials targets, we can know that the TIU-Net's performance of the reconstruction will decrease when the material is not trained by the network, which is related to the difference in polarization characteristics between the training material and the test material. In general, we can improve the adaptability of the model to different materials by enriching the materials of the training set.

3.3. Untrained underwater environments with different turbidities

Polarization characteristics of the target are less susceptible to scattering particles because it can suppress the backscattering light to some extent. Therefore, the network which makes full use of polarization information will be more stable for an unknown scattering environment. In this section, we create the turbid underwater environment with increasing concentration of scattering particles by adding the increasing volume of milk into water, and obtain polarization scattering images to test the stability of our proposed model. We obtain the polarization scattering image in the cases of 10 ml, 11 ml, 12 ml, 13 ml, 14 ml, and 15 ml milks in turns, and calculate the corresponding turbidity by turbidimeter as 133NTU, 148NTU, 162NTU, 174NTU, 187NTU and 197NTU respectively. The test results can be found in Fig. 8.

It can be seen that targets of different structural complexity in the turbid underwater at 133-162NTU can be clearly and completely reconstructed. Relatively simple digital targets in the turbid underwater of 174-187NTU milk can be accurately reconstructed, while alphabetical and Chinese character targets are poorly reconstructed. Until the volume of added milk reaches to 197NTU, the targets cannot be reconstructed totally. When targets in the turbid underwater with 133-174NTU, the polarization information is not affected by the scattering media enough to make the target information be unable to be interpreted and reconstructed by the TIU-Net. From the quantitative indicators in Table 4, in the turbid underwater at 133-174NTU, the TIU-Net is stable and there is not much gap among them. When beyond 174NTU, the scattering images cannot be fully deciphered by the TIU-Net, resulting in incomplete reconstruction of targets and sharp drop in evaluation indicators. Ultimately, the TIU-Net is capable of extended imaging beyond the concentration of the training set by 40 %, when trained with a single concentration. Increased underwater turbidity will result in less target information received by the detector, but our method, combining the advantages of polarization characteristics with the capabilities of TIU-Net, enhances generalization capability of imaging at different concentrations. It notes that the model trained by polarization information of targets is more stable. In addition, while fully extracting detailed features from scattering images obtained in turbid underwater, we integrate long-range information to make full use of the polarization characteristics of the target to improve the target reconstruction.

3.4. Untrained scattering imaging distances (SID: d)

The TIU-Net still produces good results when the SID is extended because the target polarization properties are less impacted by the scattering media during transmission. When the SID increases, the new MMs of the scattering media (M) are calculated by multiplying the MMs of the initial scattering media (M_i) by the additional MMs of thickness (M_a), according to the polarization scattering transport theory, i.e $M=M_i \times M_a$. As a result, when the model generalizes for SIDs, it also generalizes for the variable MMs. Our proposed TIU-Net can obtain the scalable MMs, which belongs to the corresponding scattering media, by the complete study and use of polarization features. Hence, in this section, we change the SID, that is to say moving the target's positions for

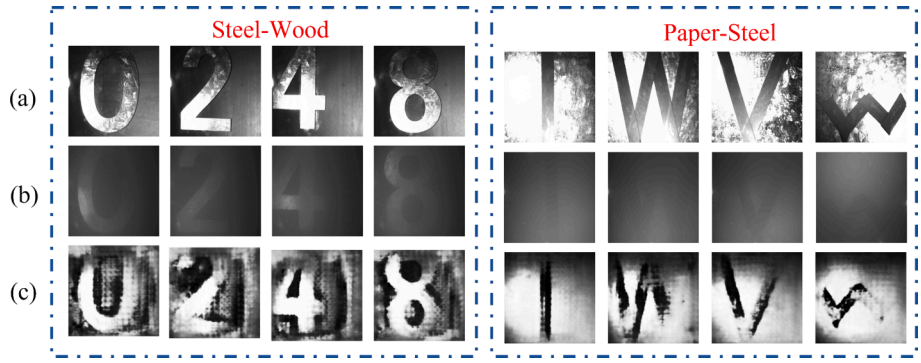

Fig. 7. The test results of targets with untrained materials. (a) Ground truth; (b) The scattering images (S_0); (c) Reconstructed images.

Table 3
The average evaluation indicators of the different materials.

	PCC	PSNR	MAE	MSE	RMSE
Steel-Wood	0.6208	11.6649	0.2054	0.00741	0.2723
Paper-Steel	0.4615	8.9919	0.2706	0.1282	0.3581

obtaining different distances from the front surface in the scattering system, for testing the TIU-Net's generalization on the SIDs. We place the target at the SID of 8 cm, 9 cm, 10 cm, 12 cm, and 13 cm respectively, with other conditions remained unchanged, and the camera captured the corresponding polarization scattering images. The results can be seen in Fig. 9, in which the TIU-Net model can clearly and completely reconstruct the target of all complexity when the SID lies in the range of 8–10 cm. At the SID of 12–13 cm, the model recovery quality is poor, the background noise appears in results, and partial reconstruction of the target is incomplete. In short, the TIU-Net is capable of extending imaging beyond the SID of the training set by 33 %. It notes that when trained with a single SID, our polarization reconstruction network can interpret the target information in the above range of SID. When the SID is less than that of the training set, the captured scattering images have more targets' information features, so our polarization reconstruction network can completely reconstruct the targets. What's more, we calculate relevant metrics to assess the generalization of our method for the SIDs depicted in Table 5. When the SID is within 10 cm, the evaluation indicators still have excellent performance, and the PCC of results are all over 75 %. When the SID increases, the transmission distance of the target information in the scattering media also increases accordingly, resulting in the output field's backward scattering increasing. Then, from the Fig. 9 (12 cm and 13 cm), the background noise becomes noticeable and the matching evaluation indicators decrease distinctly in Table 5.

Based on the analysis of Eq (3), variations in the turbidity of the scattering media and the SID will impact the change of OT. These changes signify fluctuations in the complexity of the scattering media, consequently affecting the discernment of target information. So, from the results of test experiments on the concentrations of scattering media

Table 4
The average evaluation indicators in different turbidities.

	PCC	PSNR	MAE	MSE	RMSE
133NTU	0.8065	14.4698	0.1616	0.0351	0.2221
148NTU	0.7274	12.6777	0.1644	0.0576	0.2367
162NTU	0.7157	12.5699	0.1735	0.0598	0.2410
174NTU	0.6554	11.7344	0.1917	0.0713	0.2645
187NTU	0.5770	10.8294	0.2136	0.0869	0.2929
197NTU	0.5740	10.7605	0.2138	0.0884	0.2952

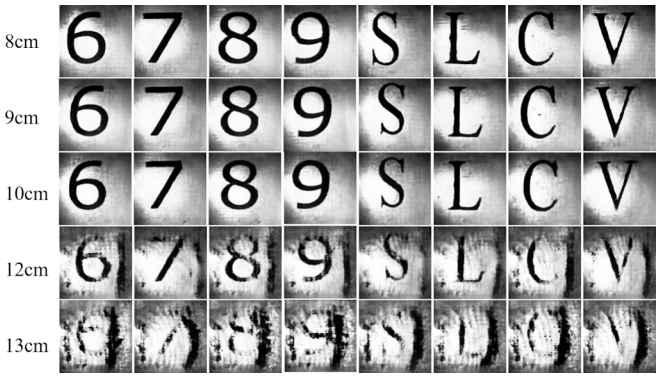

Fig. 9. The test results of untrained SIDs.

Fig. 8. The test results of untrained turbidity of water.

Table 5
The average evaluation indicators of the different SIDs.

	PCC	PSNR	MAE	MSE	RMSE
8 cm	0.8012	14.1890	0.0434	0.1357	0.2000
9 cm	0.7964	14.0736	0.0449	0.1602	0.2522
10 cm	0.7732	13.7789	0.0400	0.1676	0.2614
12 cm	0.5982	11.2101	0.0805	0.2104	0.2820
13 cm	0.4022	9.05657	0.1171	0.2579	0.3411

and SID, we can see that our proposed TIU-Net has superior performance in the underwater scattering environment for varying OTs. That has important implications in real applications of underwater imaging.

3.5. Comparison with other existing methods

To further prove that our method has certain advantages for target reconstruction in turbid underwater, in this section, we will compare our method with other existing underwater imaging techniques, including Dark Channel Prior (DCP) [56], polarization dehazing method (PDM) [57], polarization dehazing net (PDN) [58], polarization based U-Net (PU-Net) [59], modified U-Net (MU-DLN) [30]. The corresponding results are shown in Fig. 10, where all methods use the dataset obtained in the same underwater environment to train and test the model for obtaining a fair performance comparison.

From the results, both the DCP and PDM have a slightly dehazing effect on the scattering pictures obtained in turbid underwater, but the target is not clearly seen. These methods are more suitable for scenarios with less scattering influence and have certain requirements for parameter selection. From Table 6, the PDM introduced polarization information has a better performance. Hence, incorporating polarization features into the restoration algorithm based on the physical model also can enhance the imaging quality.

The PDN method can reconstruct the target, but the sharpness and contrast are not high, so, the result of outputs still looks blurry. This network structure is relatively simple, and available polarization information can't be adequately retrieved for more dense underwater environments. The PU-Net method directly uses (0°, 45°, 90°, 135°) as the training set, and the MU-DLN method only uses the S_1 as the training set. They don't make efficient use of polarization's physical prior, which prevents them from amplifying the effects of polarization properties on the model. As a result, both methods reconstruct the target poorly and incompletely in this environment. The above network structures are based on CNN; however, they didn't consider the global information of datasets with hidden target information. Additionally, we train the improved U-Net without the transformer module (IU-Net) in order to show the effectiveness of the added transformer module, and the test results are shown in the Fig. 10. According to the result in Fig. 10, even if the IU-Net can also reconstruct the target, the extraction of specific features and general features of the model do not strike a balance, leading to distinct background noise and a low contrast of the reconstruction target. In contrast, from Fig. 10, our proposed TIU-Net can reconstruct unknown objects in murky underwater with higher quality. And all of the targets are completely reconstructed without missing details. The resulting pictures have high resolution and contrast. Our approach leverages the further analysis of polarization information, specifically selecting S_0, S_1, and DoP as the training set. By addressing CNN's limitations in long-distance feature extraction through the transformer model, we successfully achieve the efficient reconstruction of hidden targets obscured by dense scattering media. In Table 6, the evaluation indicators of the different methods are also calculated, from which we can see that our method has better performance than other methods. What's more, the parameters and Floating Point Operations (FLOPs) are calculated for different methods to estimate the complexity of the network, from which it can be known that the TIU-Net can improve the quality of images without consuming too many resources. Meanwhile, it should be noted that in our computational environment (Windows Server 10 (Version 21H1) Inter(R) Core (TM) i7-9750H CPU @2.60 Hz 2.59 GHz, and 16.0 GB of RAM), after training, our proposed TIU-Net just requires 0.09628 s for reconstructing an image, which guarantee the higher imaging efficiency.

4. Conclusion

In this article, we propose a framework that integrates the transformer and CNN for extracting global and local polarization features, enabling a comprehensive perception of such features. By incorporating physical priors related to polarization information, such as the S_0, S_1, and DoLP, we effectively integrate them into our network structure for total information mining. Experimental results show that the stability and generalization of our proposed TIU-Net are greatly improved. It can effectively reconstruct the targets with different structures and materials in the turbid underwater environment. And our method can also realize cross-conditional generalization imaging with variational OTs. It is of great significance for target reconstruction in complex underwater scenarios. There are numerous distinctive characteristics of polarization that can be explored, and in the future, these polarization features could be integrated with neural networks in a more scientifically rigorous manner to achieve a more adaptable learning framework.

CRediT authorship contribution statement

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

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.

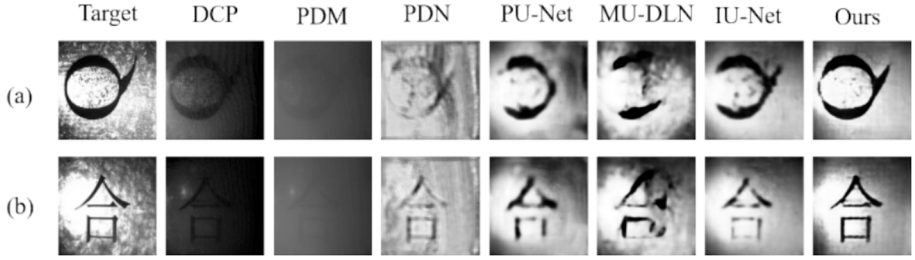

Fig. 10. The test results from different methods.

Table 6
The average evaluation indicators of the different methods.

Method	DCP	PDM	PDN	PU-Net	MU-DLN	IU-Net	Ours
PCC	0.1512	0.2498	0.3321	0.7176	0.6342	0.6935	**0.8579**
PSNR	4.3290	5.3052	10.3707	13.2399	11.568	12.6469	**15.8391**
MAE	0.4588	0.3621	0.2680	0.1566	0.1905	0.1742	**0.1085**
MSE	0.3031	0.1865	0.1003	0.0501	0.0720	0.0590	**0.0264**
RMSE	0.5505	0.4319	0.3166	0.2239	0.2683	0.2399	**0.1626**
Params/M	/	/	0.08	53.87	53.86	55.18	**1.13**
FLOPs/M	/	/	5322.57	68400.82	68136.58	68604.44	**5970.42**

Data availability

Data will be made available on request.

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.

Acknowledgment

We acknowledge funding support from National Natural Science Foundation of China (61775050).

References

[1] J. Wang, L. Liang, S. Li, et al., Correction and implement of polarization-difference imaging model for underwater target, Acta Opt. Sin. 39 (2019) 160–166.
[2] N.B. Geoffrey, C. Nicholas, Flemming, Archaeology of the continental shelf: Marine resources, submerged landscapes and underwater archaeology, Quat. Sci. Rev. 27 (2008) 2153–2165.
[3] A. Mosk, A. Lagendijk, G. Lerosey, M. Fink, Controlling waves in space and time for imaging and focusing in complex media, Nat. Photonics 6 (5) (2012) 283–292.
[4] J. Yoon, K. Lee, J. Park, Y. Park, Measuring optical transmission matrices by wavefront shaping, Opt. Express 23 (8) (2015) 10158–10167.
[5] D. Huang, E. Swanson, C. Lin, J. Schuman, W. Stinson, W. Chang, M. Hee, T. Flotte, K. Gregory, C. Puliafito, Optical Coherence Tomography. Science 254 (5035) (1991) 1178–1181.
[6] W. Yu, D. Li, K. Guo, Z. Yin, Z. Guo, Optimized sinusoidal patterns for high-performance computational ghost imaging, Appl Opt 62 (7) (2023) 1738–1744.
[7] C. Xu, D. Li, K. Guo, Z. Yin, Z. Guo, Computational ghost imaging with key-patterns for image encryption, Opt Commun 537 (2023) 129190.
[8] C. Xu, D. Li, X. Fan, B. Lin, K. Guo, Z. Yin, Z. Guo, High-performance deep-learning based polarization computational ghost imaging with random patterns and orthonormalization, Phys Scr 98 (6) (2023) 065011.
[9] Q. Xu, Z. Guo, Q. Tao, W. Jiao, S. Qu, J. Gao, Multi-spectral characteristics of polarization retrieve in various atmospheric conditions, Opt. Commun 339 (15) (2015) 167–170.
[10] P. Wang, et al., Analyzing polarization transmission characteristics in foggy environments based on the indices of polarimetric purity, IEEE Access 8 (2020) 227703–227709.
[11] C. Wang, J. Gao, T. Yao, L. Wang, Y. Sun, Z. Xie, Z. Guo, Acquiring reflective polarization from arbitrary multi-layer surface based on Monte Carlo simulation, Opt. Express 24 (9) (2016) 9397–9411.
[12] F. Shen, M. Zhang, K. Guo, H. Zhou, Z. Peng, Y. Cui, F. Wang, J. Gao, Z. Guo, The depolarization performances of scattering systems based on indices of polarimetric purity, Opt. Express 27 (20) (2019) 28337–28349.
[13] L. Shen, Y. Zhao, Q. Peng, J.C.W. Chan, S.G. Kong, An iterative image dehazing method with polarization, IEEE Trans. Multimedia 21 (5) (2019) 1093–1107.
[14] T.W. Hu, F. Shen, K. Wang, K. Guo, X. Liu, F. Wang, Z. Peng, Y. Cui, R. Sun, Z. Ding, J. Gao, Z. Guo, Broad-band transmission characteristics of Polarizations in foggy environments, Atmos. 10 (6) (2019) 342.
[15] F. Shen, K. Wang, Q. Tao, X. Xu, R. Wu, K. Guo, H. Zhou, Z. Yin, Z. Guo, Polarization imaging performances based on different retrieving Mueller matrixes, Optik 153 (2018) 50–57.
[16] Q. Tao, Y. Sun, F. Shen, Q. Xu, J. Gao, Z. Guo, Active imaging with the aids of polarization retrieve in turbid media system, Opt. Commun. 359 (2016) 405–410.
[17] B. Huang, T. Liu, H. Hu, J. Han, M. Yu, Underwater image recovery considering polarization effects of objects, Opt. Express 24 (9) (2016) 9826–9838.
[18] Y.Y. Schechner, N. Karpel, Recovery of underwater visibility and structure by polarization analysis, IEEE J. Ocean. Eng. 30 (3) (2005) 570–587.
[19] F. Liu, P. Han, Y. Wei, K. Yang, S. Huang, X. Li, G. Zhang, L. Bai, X. Shao, Deeply seeing through highly turbid water by active polarization imaging, Opt. Lett 43 (20) (2018) 903–4906.
[20] Q. Xu, Z. Guo, Q. Tao, W. Jiao, X. Wang, S. Qu, J. Gao, Transmitting characteristics of the polarization information under seawater, Appl. Opt 54 (21) (2015) 6584–6588.
[21] H. Wang, et al., Underwater imaging by suppressing the backscattered light based on mueller matrix, IEEE Photonics J. 13 (4) (2021) 1–6.
[22] X. Li, J. Xu, L.P. Zhang, H. Hu, S.C. Chen, Underwater image restoration via Stokes decomposition Opt, Lett 47 (11) (2022) 2854–2857.
[23] S. Zhu, E. Guo, J. Gu, L. Bai, J. Han, Imaging through unknown scattering media based on physics-informed learning, Photon. Res 9 (5) (2021) B210–B219.
[24] Q. Ma, J. Jiang, X. Liu, J. Ma, Deep Unfolding Network for Spatiospectral Image Super-Resolution, IEEE Trans. Comput. Imaging 8 (2022) 28–40.
[25] Z. Liu, M. Roy, D.K. Prasad, K. Agarwal, Physics-guided loss functions improve deep learning performance in inverse scattering, IEEE Trans. Comput. Imaging 8 (2022) 236–245.
[26] M. Lyu, H. Wang, G.W. Li, S.S. Zheng, G.H. Situ, Learning-based lensless imaging through optically thick scattering media, Adv. Photon 1 (3) (2019) 036002.
[27] J. Gao, G. Wang, Y. Chen, X. Wang, Y. Li, K.-H. Chew, R.-P. Chen, Mueller transform matrix neural network for underwater polarimetric dehazing imaging, Opt. Express 31 (17) (2023) 27213–27222.
[28] J.Y. Ma, H. Xu, J. Jiang, X.G. Mei, X.P. Zhang, DDcGAN: a dual-discriminator conditional generative adversarial network for multi-resolution image fusion, Trans. Img. Proc 29 (2020) 4980–4995.
[29] L. Hui, X. Wu, J. Kittler. Infrared and Visible Image Fusion using a Deep Learning Framework. 2018 24th International Conference on Pattern Recognition (ICPR) 2018; 2705-2710.
[30] H.F. Hu, Y.B. Zhang, X.B. Li, et al., Polarimetric underwater image recovery via deep learning, Opt. Lasers Eng. 133 (2020) 106152.
[31] H. Liu, Y. Zhang, Z. Cheng, J. Zhai, H. Hu, Attention-based neural network for polarimetric image denoising, Opt. Lett 47 (2022) 2726–2729.
[32] G. Wang, J. Gao, Y. Xiang, Y. Li, K.-H. Chew, R.-P. Chen, Deep learning-driven underwater polarimetric target detection based on the dispersion of polarization characteristics, Opt. Laser Technol. 174 (2024) 110549.
[33] Y. Xiang, et al., Underwater polarization imaging recovery based on polarimetric residual dense network, IEEE Photonics J. 14 (6) (2022) 1–6.
[34] G. Wang, et al., Polarization-enhanced underwater detection method for multiple material targets based on deep-learning, IEEE Photonics J. 15 (6) (2023) 1–6.
[35] Y. Xiang, G. Wang, J. Gao, X. Wang, Y. Chen, K.-H. Chew, R. Chen, Fast processing of underwater active polarimetric dehazing imaging without prior knowledge, J. Electron. Imag 32 (3) (2023) 033026.
[36] D. Li, B. Lin, X. Wang, Z. Guo, High-performance polarization remote sensing with the modified U-Net based deep-learning network, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–10.
[37] B. Lin, X. Fan, Z. Guo, Self-attention module in a multi-scale improved U-net (SAM-MIU-net) motivating high-performance polarization scattering imaging, Opt. Express 31 (2023) 3046–3058.
[38] X. Zhang, S. Cheng, J. Gao, Y. Gan, C. Song, D. Zhang, S. Zhuang, S. Han, P. Lai, H. Liu, Physical origin and boundary of scalable imaging through scattering media: a deep learning-based exploration, Photon. Res 11 (2023) 1038–1046.
[39] A. Vaswani, N.M. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin. Attention is All you Need. *ArXiv*, abs/1706.03762 2017.
[40] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A.L. Yuille, Y. Zhou. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. ArXiv, abs/2102.04306 2021.
[41] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison et al. Transformers: State-of-the-Art Natural Language Processing In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations 2020; 38–45.
[42] J. Devlin, M.W. Chang, K. Lee, K. Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies 2019; 1:4171–4186.
[43] G.G. Stokes, Mathematical and physical papers. [M], Cambridge University Press, 1901.
[44] J.S.T. Yo, M.P. Rowe, P.EN. Pugh, Target detection in optically scattering media by polarization-difference imaging, Appl. Opt 35 (11) (1996) 1855–1870.
[45] J. Zhang, J. Shao, J. Chen, D. Yang, B. Liang, R. Liang, PFNet: an unsupervised deep network for polarization image fusion, Opt. Lett 45 (2020) 507–1510.
[46] J.D. Laan, D.A. Scrymgeour, S.A. Kemme, E.L. Dereniak, Detection range enhancement using circularly polarized light in scattering environments for infrared wavelengths, Appl. Opt 54 (2015) 2266–2274.
[47] G. Huang, Z. Liu, L. Maaten, K.Q. Weinberger, Densely Connected Convolutional Networks, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017 (2017) 2261–2269.
[48] Z.e. Liu, et al., Swin Transformer: Hierarchical Vision Transformer using Shifted Windows, in: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 9992–10002.

[49] Q. Huynh-Thu, M. Ghanbari, Scope of validity of PSNR in image/video quality assessment, Electron. Lett 44 (13) (2008) 800–801.
[50] H. Zhao, O. Gallo, I. Frosio, J. Kautz, Loss functions for image restoration with neural networks, IEEE Trans. Comput. Imaging 3 (1) (2017) 47–57.
[51] Buda, Andrzej. Life time of correlation between stocks prices on established and emerging markets. arXiv: General Finance 2011; n. pag.
[52] X. Fan, B. Lin, K. Guo, B. Liu, Z. Guo, High-performance polarization scattering imaging based on three-stage multi-pipeline networks, Opt. Express 31 (23) (2023) 38097–38113.
[53] B. Lin, X. Fan, P. Peng, Z. Guo, Dynamic polarization fusion network (DPFN) for imaging in different scattering systems, Opt. Express 32 (1) (2024) 511–525.
[54] Y. Zhao, Y. Li, W. He, Y. Liu, Y. Fu, Polarization scattering imaging experiment based on Mueller matrix, Opt. Commun. 490 (1) (2021) 126892.
[55] X.Y. Wang, T.W. Hu, D.K. Li, K. Guo, J. Gao, Z.Y. Guo, Performances of polarization-retrieve imaging in stratified dispersion media, Remote Sens. (Basel) 12 (18) (2020) 2895.
[56] K. He, J. Sun, X. Tang, Guided image filtering, IEEE Trans. Pattern Anal. 35 (6) (2013) 1397–1409.
[57] J. Liang, L. Ren, H. Ju, W. Zhang, E. Qu, Polarimetric dehazing method for dense haze removal based on distribution analysis of angle of polarization, Opt. Express 23 (2015) 26146–26157.
[58] H. Hu, Y. Zhang, X. Li, Y. Lin, Z.Z. Cheng, T. Liu, Polarimetric underwater image recovery via deep learning, Opt. Lasers Eng. 133 (2020) 106152.
[59] R. Zhang, X. Gui, H. Cheng, J. Chu, Underwater image recovery utilizing polarimetric imaging based on neural networks, Appl. Opt. 60 (27) (2021) 8419–8425.