



# Self-attention module in a multi-scale improved U-net (SAM-MIU-net) motivating high-performance polarization scattering imaging

BING LIN, XUEQIANG FAN, AND ZHONGYI GUO\* 

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

\*[guozhongyi@hfut.edu.cn](mailto:guozhongyi@hfut.edu.cn)

**Abstract:** Polarization imaging has outstanding advantages in the field of scattering imaging, which still encounters great challenges in heavy scattering media systems even though there are helps from deep learning technology. In this paper, we propose a self-attention module (SAM) in multi-scale improved U-net (SAM-MIU-net) for the polarization scattering imaging, which can extract a new combination of multidimensional information from targets effectively. The proposed SAM-MIU-net can focus on the stable feature carried by polarization characteristics of the target, so as to enhance the expression of the available features, and make it easier to extract polarization features which help to recover the detail of targets for the polarization scattering imaging. Meanwhile, the SAM's effectiveness has been verified in a series of experiments. Based on proposed SAM-MIU-net, we have investigated the generalization abilities for the targets' structures and materials, and the imaging distances between the targets and the ground glass. Experimental results demonstrate that our proposed SAM-MIU-net can achieve high-precision reconstruction of target information under incoherent light conditions for the polarization scattering imaging.

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

## 1. Introduction

High-quality imaging through scattering media is of great significance for atmospheric remote sensing [1–3], underwater imaging [4–6], biological tissues imaging [7–8], and other applications [9–11]. So far, various methods have been proposed to improve imaging quality [12–19], in which polarization-based methods, as one of the most effective techniques, have received many achievements. J. S. Tyo et al. have proposed the Polarization Difference (PD) method for improving the imaging quality through scattering media [20]. Y.Y. Schechner has added polarization effects into the model of atmospheric defogging to improve the defogging effect [21]. Liu et al. have proposed an active polarization imaging method based on wavelength selection [22], which takes advantage of the dependence of scattered light at different wavelengths in the turbid underwater environment. Liang et al. have proposed the estimation parameter of the angle of polarization (AoP) can be used in dense fog environments [23], which significantly improves the clarity of blurred images. Li et al. have presented a method based on Stokes vector images to recover objects in turbid water [24]. Guo et al. have obtained the Muller matrix (MM) of a scattering medium based on the Monte Carlo (MC) algorithm [25] and proposed a method of polarization inversion to recovery targets in layered dispersion systems [26–27]. In recent years, deep learning-based methods have developed rapidly and are considered to be a method that surpasses traditional methods and improves the performance of polarization scattering imaging. Li et al. have established a dataset consisting of the Stokes vector images and proposed a polarized image denoising network (PDRDN) based on the residual dense network [28]. Hu et al. have presented a method of polarized underwater image defogging based on deep learning [29], and on this basis, they mathematically converted the two polarization-related

parameters into a separate parameter, enabling the network to learn the polarization modulation parameters and obtain a clear de-scattering image [30]. Li et al. have considered changes in polarization information when light interacts with targets and transmits in the turbid system. They have combined polarization theory and deep learning and designed an end-to-end network for target reconstruction in a scattering environment [31]. Polarization is a superior characteristic of light, but it is not presenting in a direct way. We must use some optical elements to be able to observe it, and the detection methods are also indirectly detected by the intensity. What is more, the energy loss caused by the optics will also lead to a serious decrease in the signal-to-noise ratio of the picture.

Therefore, when polarization pictures captured by the detector are directly entered into the model, the network may not extract enough useful information sufficiently and accurately, resulting in the inability to recover the target efficiently. Therefore, here, we try to provide enough information for the reconstruction network by entering multidimensional information about the target. On this basis, by introducing a self-attention module (SAM) into the multi-scale improved U-net (MIU-net) to form an SAM-MIU-net, the network focuses on the target information carried by the polarization characteristics and enhances the expression of features by giving weights to the feature matrix itself, reducing redundant output and improving the robustness of the network. The experimental results prove that our proposed SAM-MIU-net has significantly improved the reconstruction result, and the test results of complex structures, different materials, and different imaging distances also show that our proposed SAM-MIU-net has superior effectiveness and generalization.

## 2. Methodology

### 2.1. Polarization information

The Stokes vector is a common expression of the polarization information, and it can represent polarized and unpolarized light with four components  $S = (I, Q, U, V)^T$ , all of which are scalars and express the light intensity information without phase [32]. The Stokes vector can be calculated as:

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

where  $I$  is the total light intensity, which provides global information of targets;  $Q$  is the difference between horizontal and vertical components, which is the difference between the components in two orthogonal directions, so  $Q$  images have a certain inhibition effect on backscatter;  $U$  is the difference between  $45^\circ$  and  $135^\circ$  components, and  $V$  is the difference between right-handed and left-handed components. In addition, further polarization information such as the degree of polarization (DoP) and AoP can be obtained by the Stokes vectors. The degree of linear polarization (DoLP) represents the ratio of the linear-polarization component to the total light intensity:

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

We can get more detailed information about the target through DoLP images. Previous research [33] has fused intensity pictures with DoLP pictures to provide each other with complementary information about the target to improve the resolution of images. Here, we consider that a single polarization picture cannot provide sufficient target characteristics for the network, so we regard the light intensity,  $Q$ , and DoLP images as three-dimensional data and input them into

the network. By taking advantages of the powerful data mining and learning capabilities of the neural networks, three-dimensional information can be fully extracted and fused for the target reconstruction.

## 2.2. Measurement system

We construct the experimental setup as shown in Fig. 1 to obtain the polarization dataset. A linear polarizer is placed in front of the LED light source, which allows the captured picture to contain more pronounced polarization information, aiding in subsequent image recovery [34]. The light of  $S = (1, 1, 0, 0)^T$  can transmit from the ground glass, which will be irradiated to the target and reflected from it. Then the reflective light that carries the targets' information of transmits through the ground glass and will be captured by the commercial DoFP (division of focal plane) polarization camera (LUCID, PHX055S-PC) with pixel counts of  $2048 \times 2448$ . The pixel array surface of DoFP is covered with a polarization array consisting of four micro-polarizers with four different polarization orientations of  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$  respectively. Then the needed images for the polarization dataset can be calculated by them easily. In our experiments, the target is put at a certain distance behind the ground glass of 4 mm, and the distance between the target and the ground glass is defined as  $d$ .

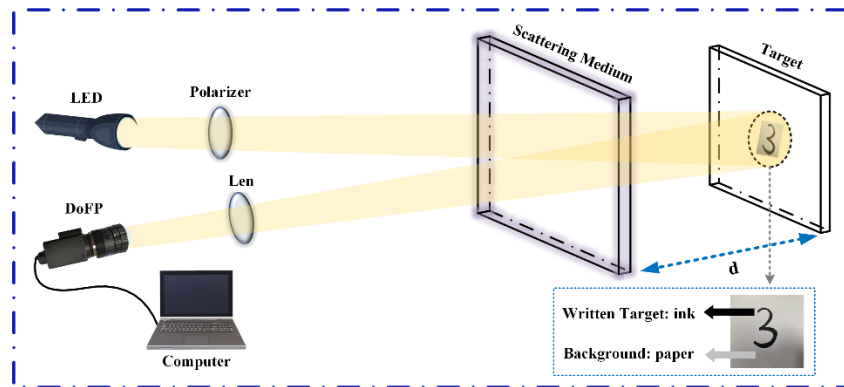


Fig. 1. Schematic of the experimental setup.

## 2.3. Network design

Inspired by the network architecture proposed by Zhen et al. [35], we have improved the network structure used in the previous work [31], in which we use different-size down-sampling strategies in the middle part of the U-Net to extract target features at different levels. Here, the newly proposed network consists of two stages, in which the former part includes a multi-scale feature extraction network, and the latter part is a fusion reconstruction stage network of different levels of features. We utilize the SAM to optimize the feature exchange between two stages, to ensure that the polarization features that are most conducive to the target reconstruction can be well transmitted to the next stage, improving the reconstruction quality and generalization of the network.

As shown in Fig. 2, the backbone network is modified based on the improved U-Net of the previous work [31]. The former part is a feature extraction part composed of three sets of dense blocks, and at the end of each block, the max-pooling of the size of  $2 \times 2$  reduces the feature to half of the previous level. If we input the target's multi-dimensional information into the network directly, it will not be conducive for information to play its role when the features are modulated inside the network, but also will cause information redundancy, which is not conducive

to improving the efficiency and robustness of the network. Therefore, when the feature size is reduced to  $64 \times 64$ , we divide the network into four branches and use  $1 \times 1$ ,  $2 \times 2$ ,  $4 \times 4$ , and  $8 \times 8$  max pooling respectively to extract the existing features in different levels and degrees. Through this method, the different roles of multi-dimensional information in the reconstruction of the target can be refined, and features are divided into different channels to the next stage, thereby reducing the superposition of information and improving the efficiency of the network. Most networks that require multi-branch fusion connect features directly and then pass them to the next stage. However, we want more than just fusion, but also filtering redundant information and enhancing polarization characteristics that facilitate target reconstruction. Wang et al. [36] proposed that the SAM can aggregate global information from the feature map. Therefore, here, we introduce the SAM to reduce the redundant output of the first-level network, aggregate effective polarization features by establishing the interaction of feature information between different channels, and enhance the input features to the next stage for feature fusion.

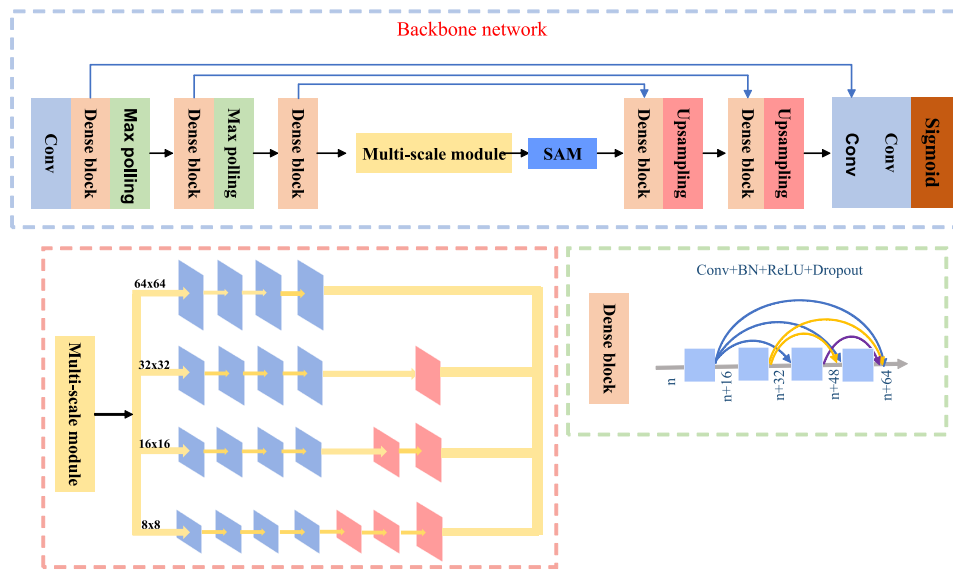


Fig. 2. The overall structure of the used U-net.

The SAM's network structure is shown in Fig. 3. The inputs are first fed into three different convolutional layers (i.e., Conv1, Conv2, Conv3) with kernel size of  $1 \times 1$ , which do not change the spatial size of feature, to generate the query ( $Q$ ), key ( $K$ ), and value ( $V$ ) matrix. Then, according to the equations (3) and (4), the  $Q$  vector interacts with  $K$  vector using the dot-product operators to produce a scalar weight (i.e., the attention map  $Att$ ) for the corresponding  $V$  vector. Subsequently, the attention map  $Att$  vector is applied on the  $V$  vector to generate the  $Y$  vector.

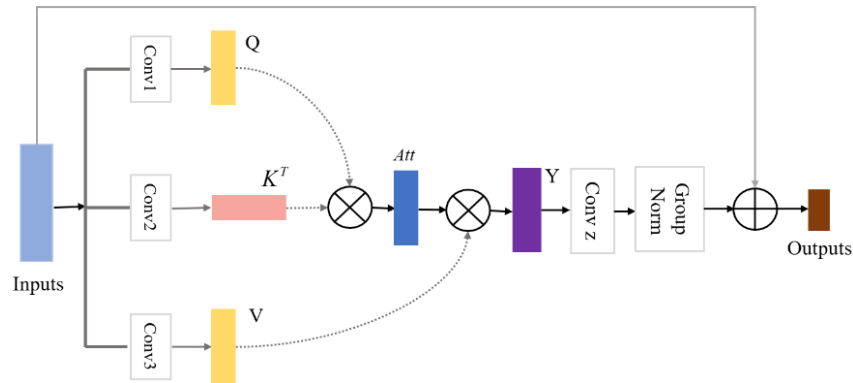
$$Att = Q \cdot K^T, \quad (3)$$

$$Y = V \cdot Att, \quad (4)$$

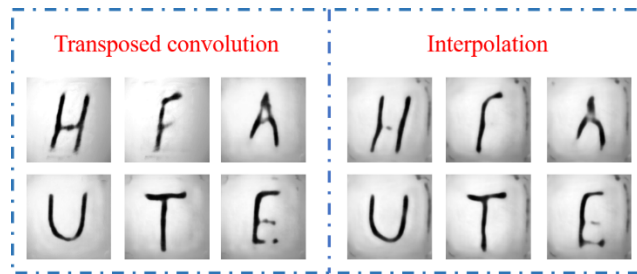
$$Outputs = GroupNorm(Convz(Y, W_\mu)) + Inputs, \quad (5)$$

Next, the computational complexity is reduced by convolution. The final output of the SAM is shown by Eq. (5), in which a residual structure is formed with the initial input to ensure that the feature information is not lost. Finally, the features enhanced and focused by the SAM are input into the dense block, and then recomposed into the target image of  $256 \times 256$  by up-sampling and convolution layers. Furthermore, to better reconstruct the target's information, two up-sampling

schemes, i.e., transposed convolution and bilinear interpolation, are considered in the decoder and multiscale module. Fig. 4 demonstrates the performances' comparison between using transposed convolution and utilizing bilinear interpolation in the decoder and multiscale module. It can be seen that the transposed convolution is more conducive to the final reconstructions of targets. Hence, in subsequent experiments, we have used transposed convolution as the up-sampling operation.



**Fig. 3.** Schematic of the SAM.



**Fig. 4.** The effects of different up-sampling operations.

During training, the MAE acts as a loss function to drive the interaction of polarization features within the network.

$$MAE = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N ||X(i,j) - Y(i,j)||, \quad (6)$$

We trained the model in an image processing unit (NVIDIA RTX 3080) using a Pytorch framework with Python 3.6, training 150 epochs. The optimizer is the Adam (Add Momentum Stochastic Gradient Descent) with a learning rate of 0.001.

#### 2.4. Imaging quality

Mean Squared Error (MSE) and Structural Similarity Index (SSIM) [37] are two common data metrics used to evaluate imaging quality. The MSE between the original image and the predicted

image ( $O, K$ ) with the size of  $m \times n$  can be expressed by:

$$MSE = \frac{1}{MN} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [O(i,j) - K(i,j)]^2, \quad (7)$$

The smaller its value means the better the recovery results.

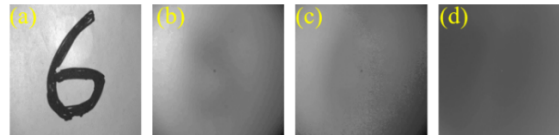
In addition to the evaluation of the noise situation, the quality of the reconstruction image can be evaluated by introducing structural similarity based on the degradation of structural information. The SSIM compares the brightness, contrast, and structure of the two images by:

$$SSIM(X, Y) = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)}, \quad (8)$$

where  $\mu_x$  is the mean of  $X$ ,  $\mu_Y$  is the mean of  $Y$ ,  $\sigma_x$  is the variance of  $X$ ,  $\sigma_Y$  is the variance of  $Y$ ,  $\sigma_{XY}$  is the covariance of  $X$  and  $Y$ , and  $C_1$  and  $C_2$  are small normal numbers used to avoid the zero denominator. The SSIM value range is 0 to 1. The higher value of the SSIM, the more similar the two images, which means a better network reconstruction.

### 2.5. Dataset preparation

The scattering images captured by the camera as shown in Fig. 5 are cloudy under incoherent light conditions. When the distance between the ground glass and the target is moved, the sharpness of the scattering images will also change, that is, the greater the distance, the blurrier the picture. The training set is composed of three kinds of scattering images,  $S_0$ ,  $S_1$ , and DoP, at the distance of  $d = 4$  cm between the ground glass and the target as shown in Fig. 5, and the following inputs for different test experiments are also composed of three-dimensional data. Our dataset includes 200 groups of polarized images, and each has four sets of images corresponding to different polarization orientations ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$ ). On this basis, the images needed for training can be calculated by Eqs. (1,2). Also, the targets are all made up of digits in different fonts. In addition, 200 scattering images are expanded to 2000 images as training sets. All of our data are grayscale images, and the final outputs are also grayscale images. After collecting and classifying the data, the proposed methods can be used.



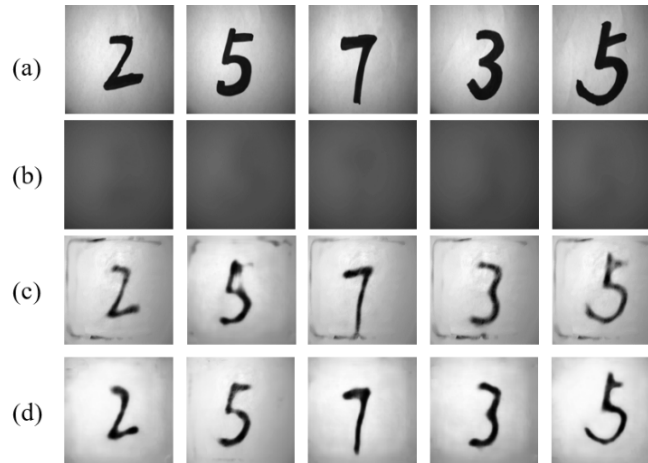
**Fig. 5.** (a) Original target; (b) Scattering imaging by  $S_0$ ; (c) Scattering imaging by  $S_1$ ; (d) Scattering imaging by DoP.

## 3. Results and discussions

### 3.1. Enhanced performances from the SAM

Polarization characteristics can distinguish the ballistic and scattering photons to some extents, so the network trained with polarization information will be more stable. In this section, in order to demonstrate that the SAM can help the network to focus and enhance the stable target characteristics carried by polarizations, we conduct the comparative experiments with or without the SAM in the network. We name WSAM-MIU-net for the circumstance of MIU-net without SAM. We use the same training set to train the SAM-MIU-net and WSAM-MIU-net, respectively, and obtain the corresponding optimal model for comparative testing as shown in Fig. 6, in which

Fig. 6(b) is the scattering images of the light intensity, but the test sets also consisted of scattering images of  $S_0$ ,  $S_1$  and DoP (for the sake of image simplicity, we only show the scattering images of the light intensity in the following result display).



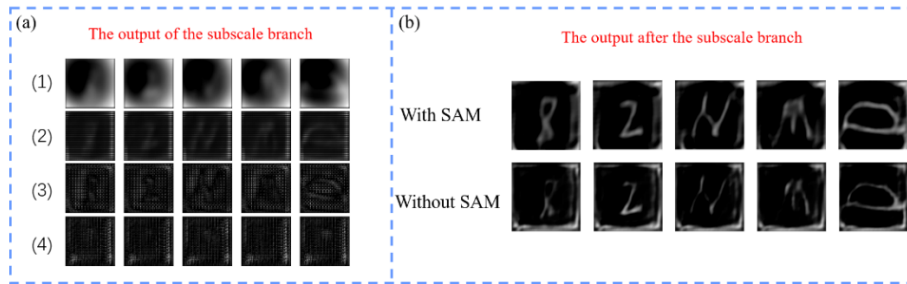
**Fig. 6.** The reconstruction result of different models. (a)Untrained digital images;(b) Scattering images;(c) The reconstruction results without SAM; (d) The reconstruction results with SAM.

It can be clearly seen that the results of the SAM-MIU-net all have higher contrast and clarity, and the background has no interference noise. Nevertheless, the results of the WSAM-MIU-net cannot incompletely rebuild more complex targets, such as the second target (“5”), and all the test results contain background noise. The results prove that the SAM can guide the network to focus on and enhance the target characteristics carried by the polarization information, filter out the redundant information, and improve the reconstruction performance. At the same time, we also calculated the average SSIM and MSE of the reconstructed result that be shown in Table 1. Ultimately, the reconstructing performance of the model with SAM has been greatly improved, where the SSIM has increased by 7.4% and the MSE has decreased by 11%.

**Table 1. The average SSIM and MSE of the different models**

Model	SSIM ↑	MSE ↓
SAM-MIU-net	<b>0.8175</b>	<b>0.00304</b>
WSAM-MIU-net	0.7609	0.00345

In addition, we output the features in the middle of the network to explore the physical process of the proposed network. Firstly, we output the features of the four branches as shown in Fig. 7(a). It can be seen that different branches extract the different aspect features, and refine the contribution of multi-dimensional information to the target reconstruction, which will avoid redundancy caused by information superposition and affect the performance of the model. After that, we export the feature maps from the networks with SAM and without SAM respectively, which has been shown in Fig. 7(b). By visualizing the features, we can know the aggregation and enhancement effect of the SAM on the polarization features. Through the above operations, the features are exported to the next modules for information fusion, which will improve the performance of reconstruction under incoherent conditions.

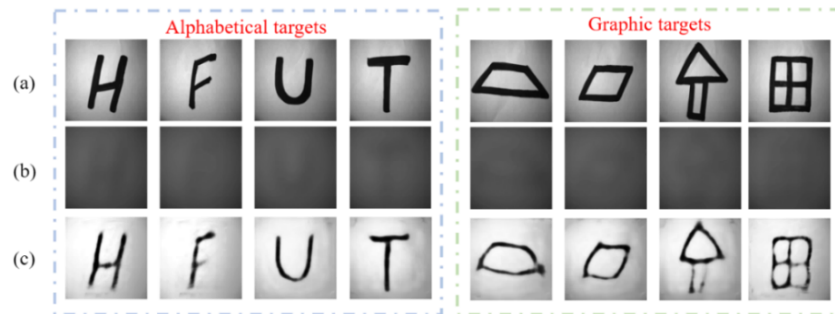


**Fig. 7.** Visualization of features in the middle of the networks. (a) The output of the subscale branch; (b) The output after the subscale branch for the networks with SAM and without SAM.

### 3.2. Performances of SAM-MIU-net on untrained targets

#### 3.2.1. Untrained different-structure targets

In this section, we test the trained SAM-MIU-net with more complexity targets while other conditions are unchanged to further verify the generalization. The alphabetical targets and graphic targets, which are not in the training sets, are entered into the trained network, and the reconstruction results are obtained in Fig. 8. Alphabetical targets and graphic targets, which belong to different structural types from training sets, can be also accurately reconstructed by our method. In case of the limited number of training sets, the target structure is reestablished without excessive noise in the background. The structure of graphic targets is the weakest correlation with the training dataset, but the results are still reconstructed with little distortion. From the Fig. 8, our proposed method can reestablish untrained objectives with high contrast, and there is no excessive noise in the background. It can be proved that the SAM-MIU-net has excellent generalization ability for targets' structures.



**Fig. 8.** The reconstruction result of SAM-MIU-net. (a)Original images; (b) Scattering images; (c) Reconstruction results.

In addition to visual effects, the superior effectiveness and generalization of the SAM-MIU-net can be also seen from the average SSIM and MSE of results in Table 2. Even the less correlated graphical targets also have better performances of SSIM and MSE.

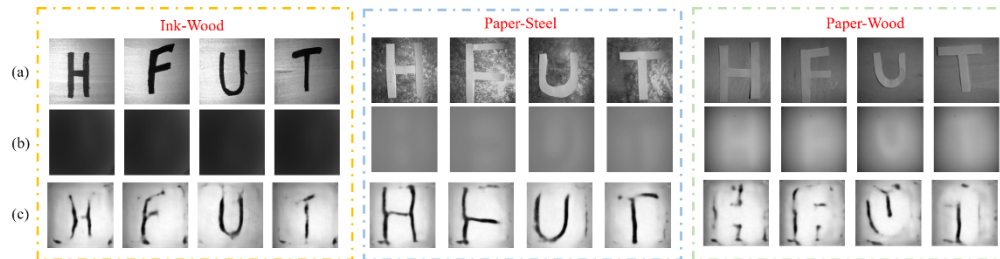
**Table 2.** The average SSIM and MSE of the different targets

Target	SSIM ↑	MSE ↓
Alphabetical targets	0.7882	0.00320
Graphic targets	0.7628	0.00338



### 3.2.2. Untrained different-material targets

Polarization properties are very sensitive to different materials, so we explored the concrete effects of different materials on the trained model. We change the materials of target-background to Ink-Wood, Paper-Steel, and Paper-Wood sequentially, and place them in the same experimental environment to obtain scattering images of different components. The test results can be obtained by entering their scattering images into the SAM-MIU-net trained by the target-background of Ink-Paper, as depicted in Fig. 9.



**Fig. 9.** The reconstruction results of different target-background: Ink-Wood, Paper-Steel, and Paper-Wood. (a) Original images; (b) Scattering images; (c) Their reconstruction results from the SAM-MIU-net trained by the target-background of Ink-Paper.

Table 3 [3,38] shows the corresponding elements in the Mueller matrix (MM) of the materials used in our experiment. When objects are set as “Ink-Wood”, the difference of corresponding MM elements of paper and wood are small, so, the model can reconstruct the ink target relatively completely. Besides, when the target materials are set as “Paper-Steel”, other conditions remain unchanged. From Table 3, the corresponding MM elements of ink and steel are similar, so, their scattering images are similar to the “Ink-Paper”, resulting in targets can be reconstructed like result got by Ink-Paper model. Lastly, when “Paper-Wood” targets are set, due to the similar corresponding MM elements of paper and wood, only the target profile can be distinguished. From the results, it can be known that when the material is not trained by the neural network, the performances of the target reconstruction will be decreased, and the performance of the reconstruction is related to the difference in polarization characteristics between the training material and the test material. More importantly, SAM-MIU-net reconstructs the target with more completeness and higher contrast. The SAM also will enhance features that are similar to the polarization characteristics of training targets, which is of great effect on model scalability imaging. Besides, we have also calculated the average SSIM and MSE of the reconstructed different-material targets shown in Table 4.

**Table 3. Mueller matrix elements of different materials [3,38]**

Material	m22	m33
Paper	0.265	0.247
Wood	0.215	0.16
Ink	0.892	0.921
Steel	0.980	0.977

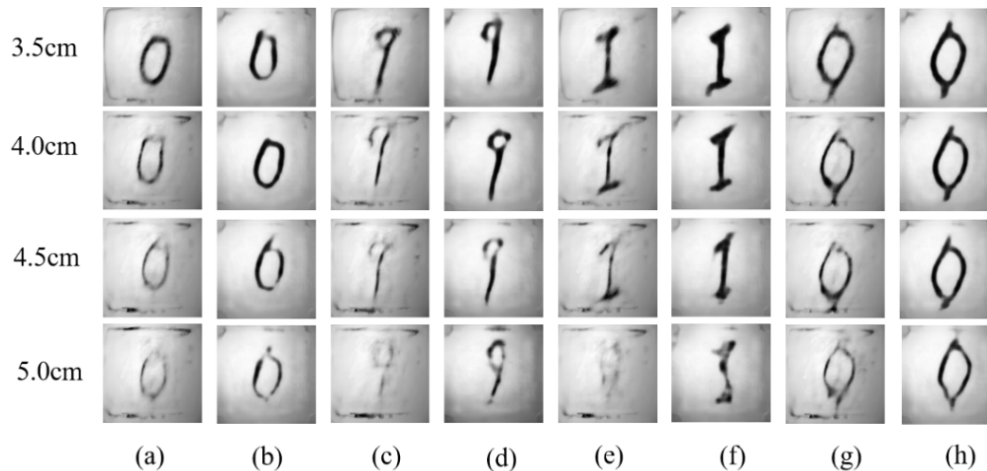
In summary, our proposed method has a certain ability and generalization due to that the SAM-MIU-net will enhance the expression of polarization features in the process of reconstructing the targets.

**Table 4. The average SSIM and MSE of the different-material targets**

Target	Ink-Wood	Paper-Steel	Paper-Wood
SSIM $\uparrow$	0.4969	0.4147	0.3372
MSE $\downarrow$	0.00341	0.00357	0.00378

### 3.2.3. Untrained different-distance imaging

The values of MM for different materials will be different, so their polarization properties will be also different. In addition, when targets and scattering media are determined in a system, the MM of those will not change. As a result, the network trained by polarization information is more robust. So, our proposed SAM-MIU-net can reconstruct the targets with different distances (the targets move within a certain range). In this section, we will demonstrate that the proposed method has strong stability by reconstructing the scattering images obtained at different distances of  $d$ , as shown in Fig. 10.



**Fig. 10.** (a)(c)(e)(g) The result without SAM; (b)(d)(f)(h) The result with SAM.

When  $d = 3.5$  cm, there is enough target information for target reconstruction, and the stable polarization information improves the quality of results. Besides, we also show the test results from the WSAM-MIU-net, in which the reconstruction results from the SAM-MIU-net are more stable than those from the WSAM-MIU-net when the distance is longer than 4 cm. Particularly, when  $d = 5$ , the SAM-MIU-net can still distinguish the targets, but the WSAM-MIU-net can't. Comparison results prove that the aggregation effect of SAM on the polarization characteristics enhancing the expression of stable target features carried by the polarization and improving the flexibility of network. Ultimately, the SAM-MIU-net is capable of extended imaging beyond the distance of the training set by 25%, and achieve efficient elastic imaging. At the same time, we calculated the average SSIM and MSE of the reconstruction result at different  $d$  for both cases. From Table 5, the data for the model with SAM is generally better, and the values at different  $d$  are more stable.

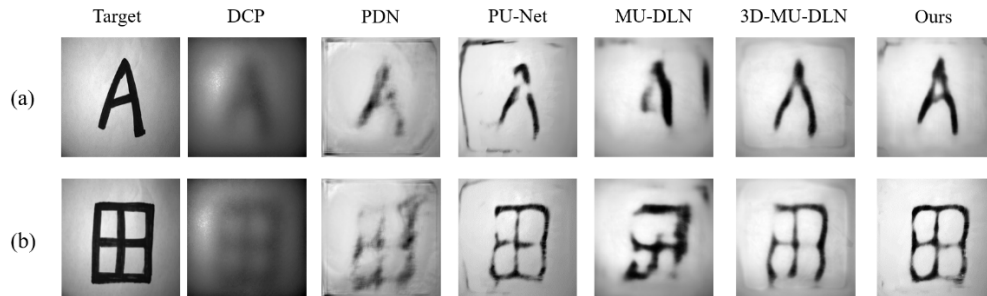
### 3.3. Performance comparison with other existing methods

In this section, we compared our proposed SAM-MIU-net with several existing methods, including Dark Channel Prior (DCP) [39], PDN [29] (Hu proposed the method based on RDB to dehaze using  $0^\circ$ ,  $45^\circ$  and  $90^\circ$ ), PU-Net [40] (Zhang proposed the method based on U-Net to dehaze

**Table 5. The average SSIM and MSE of the different networks at different distances of d**

SAM-MIU-net	d = 3.5cm	d = 4.0cm	d = 4.5cm	d = 5.0cm
SSIM ↑	0.8029	0.7928	0.7758	0.7729
MSE ↓	0.00351	0.00312	0.00356	0.00358
WSAM-MIU-net	d = 3.5cm	d = 4.0cm	d = 4.5cm	d = 5.0cm
SSIM ↑	0.7565	0.7374	0.7352	0.7253
MSE ↓	0.00362	0.00334	0.00367	0.00369

using  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$  and  $S_0$ ), and MU-DLN [31] (Li proposed the method based on modified U-Net using Q-component). The corresponding results are shown in Fig. 11 respectively. To make a fair performance comparison, except for the DCP method, all the methods first used the same training set to learn the model of polarization scattering imaging, and then employed the same testing set to verify their performances accordingly.

**Fig. 11.** The obtained results with different methods

From the results, the DCP method achieves a slight dehazing effect and does not fully visualize the targets, and then the PDN method cannot clearly recover the target, which may not be suitable for more blurry pictures in complex environments. Although the PU-Net can relatively recover part of the target structure, there is more noise around it. The MU-DLN method input the polarization information of  $S_1$ , only one of the components of the Stokes vector, and the recovery result is incomplete and not in high contrast. On the basis of MU-DLN, we retrain that by using the training sets (i.e. I, Q, DoP) used in this article to get 3D-MU-DLN. And the recovery results are better than those of MU-DLN, but there are also cases where the target recovery is incomplete. In contrast, our proposed SAM-MIU-net is able to completely reconstruct the target structure, enhance the contrast, and make the background have less excessive noise in complex environments. In Table 6, the SSIM and MSE of different methods are calculated, and our proposed SAM-MIU-net obtained better reconstruction performance than other methods. In addition, we calculate the parameters and Floating Point Operations (FLOPs) for different methods to assess the complexity of the network. It can be seen that our method has high quality while having fewer parameters and calculations.

**Table 6. The evaluation indicators of the different methods**

Method	DCP	PDN	MU-DLN	PU-Net	3D-MU-DLN	Ours
SSIM ↑	0.4969	0.6618	0.6736	0.6972	0.7363	<b>0.7755</b>
MSE ↓	0.1330	0.003484	0.00335	0.00323	0.00352	<b>0.00315</b>
Params/M	\	0.08	53.86	53.87	55.18	<b>1.13</b>
FLOPs/M	\	5322.57	68136.58	68400.82	68604.44	<b>5970.42</b>

#### 4. Conclusion

In this manuscript, given the not obvious polarization characteristics and the limitations of the detection system, we use multidimensional polarization information to characterize the target and use multi-scale extraction to refine the contribution of multi-dimensional information for the reconstruction target, which also will avoid information redundancy caused by information superposition. The SAM is introduced to aggregate global information, and enhance the polarization characteristics which will input subsequent reconstruction modules. Experiments have verified that the SAM-MIU-net has greatly improved generalization and stability. And through the intermediate feature output, we can visualize the influence of our proposed module on the final reconstruction result. It notes that the application of multidimensional information to target features is of great significance for target reconstruction in complex scenarios, such as in the optical remote sensing. In future works, to further improve the performance of polarization scattering imaging, we will focus on the following points: (i) extracting the available information from multi-material target information using the supervised learning algorithms for more scenes reconstruction; (ii) due to the relatively complex acquisition of polarization data sets, it is significant to consider using a small number of training samples to achieve high-quality reconstruction.

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

**Disclosures.** The authors declare no conflicts of interest.

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

#### References

1. Q. Xu, Z. Y. Guo, Q. Q. Tao, W. Y. Jiao, S. L. Qu, and J. Gao, "Multi-spectral characteristics of polarization retrieve in various atmospheric conditions," *Opt. Commun.* **339**, 167–170 (2015).
2. T. W. Hu, F. Shen, K. Wang, K. Guo, X. Liu, F. Wang, Z. Peng, Y. Cui, R. Sun, Z. Ding, J. Gao, and Z. Guo, "Broad-band transmission characteristics of Polarizations in foggy environments," *Atmosphere* **10**(6), 342 (2019).
3. X. Y. Wang, T. W. Hu, D. K. Li, K. Guo, J. Gao, and Z. Y. Guo, "Performances of polarization-retrieve imaging in stratified dispersion media," *Remote Sensing* **12**(18), 2895 (2020).
4. K. Purohit, S. Mandal, and A.N. Rajagoplan, "Multilevel weighted enhancement for underwater image dehazing," *J. Opt. Soc. Am. A* **36**(6), 1098–1108 (2019).
5. Q. Xu, Z. Guo, Q. Tao, W. Jiao, X. Wang, S. Qu, and J. Gao, "Transmitting characteristics of the polarization information under seawater," *Appl. Opt.* **54**(21), 6584–6588 (2015).
6. B. Huang, T. Liu, H. Hu, J. Han, and M. Yu, "Underwater image recovery considering polarization effects of objects," *Opt. Express* **24**(9), 9826–9838 (2016).
7. D. Li, C. Xu, M. Zhang, X. Wang, K. Guo, Y. Sun, J. Gao, and Z. Guo, "Measuring glucose concentration in a solution based on the indices of polarimetric purity," *Biomed. Opt. Express* **12**(4), 2447–2459 (2021).
8. R. Horstmeyer, H. W. Ruan, and C. H. Yang, "Guidestar-assisted wavefront-shaping methods for focusing light into biological tissue," *Nat. Photonics* **9**(9), 563–571 (2015).
9. F. Shen, M. Zhang, K. Guo, H. Zhou, Z. Peng, Y. Cui, F. Wang, J. Gao, and Z. Guo, "The Depolarization Performances of Scattering Systems Based on Indices of Polarimetric Purity," *Opt. Express* **27**(20), 28337–28349 (2019).
10. F. Shen, B. Zhang, K. Guo, Z. Yin, and Z. Guo, "The depolarization performances of the polarized light in different scattering media systems," *IEEE Photonics J.* **10**(2), 1–12 (2018).
11. J. Wu, J. C. Wang, Y. G. Nie, and L. F. Hu, "Multiple-image optical encryption based on phase retrieval algorithm and fractional Talbot effect," *Opt. Express* **27**(24), 35096–35107 (2019).
12. J. Burch, J. Ma, R. I. Hunter, S. A. Schulz, D. A. Robertson, G. M. Smith, J. Wang, and A. Di Falco, "Flexible patches for mm-wave holography," *Appl. Phys. Lett.* **115**(2), 021104 (2019).
13. K. Cheng, Z. Li, J. Wu, Z. Hu, and J. Wang, "Super-resolution imaging based on radially polarized beam induced superoscillation using an all-dielectric metasurface," *Opt. Express* **30**(2), 2780–2791 (2022).
14. K. Guo, X. Li, H. Ai, X. Ding, L. Wang, W. Wang, and Z. Guo, "Tunable oriented mid-infrared wave based on metasurface with phase change material of GST," *Results Phys.* **34**, 105269 (2022).
15. M. Lyu, H. Wang, and G. Li, "Learning-based lensless imaging through optically thick scattering media," *Adv. Photon.* **1**(03), 1 (2019).
16. L. Sun, J. Shi, X. Wu, Y. Sun, and G. Zeng, "Photon-limited imaging through scattering medium based on deep learning," *Opt. Express* **27**(23), 33120–33134 (2019).
17. J. Wu, L. Hu, and J. Wang, "Fast tracking and imaging of a moving object with single-pixel imaging," *Opt. Express* **29**(26), 42589–42598 (2021).

18. C. Xu, D. Li, K. Guo, Z. Yin, and Z. Guo, "Computational ghost imaging with key-patterns for image encryption," *Optics Communications*, In press, (2022).
19. S. Zhu, E. Guo, J. Gu, L. Bai, and J. Han, "Imaging through unknown scattering media based on physics-informed learning," *Photon. Res.* **9**(5), B210–B219 (2021).
20. J. S. Tyo, M. P. Rowe, and E. N. Pugh, "Pugh. Target detection in optically scattering media by polarization-difference imaging," *Appl. Opt.* **35**(11), 1855–1870 (1996).
21. Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Nayar Instant dehazing of images using polarization," *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. IEEE 1, I (2001).
22. F. Liu, P. Han, Y. Wei, K. Yang, S. Huang, X. Li, G. Zhang, L. Bai, and X. Shao, "Deeply seeing through highly turbid water by active polarization imaging," *Opt. Lett.* **43**(20), 4903–4906 (2018).
23. J. Liang, L. Ren, H. Ju, W. Zhang, and E. Qu, "Polarimetric dehazing method for dense haze removal based on distribution analysis of angle of polarization," *Opt. Express* **23**(20), 26146–26157 (2015).
24. X. Li, J. Xu, L. Zhang, H. Hu, and Sh. Chen, "Underwater image restoration via Stokes decomposition," *Opt. Lett.* **47**(11), 2854–2857 (2022).
25. Q. Tao, Y. Sun, F. Shen, Q. Xu, J. Gao, and Z. Guo, "Active imaging with the AIDS of polarization retrieve in turbid media system," *Opt. Commun.* **359**(15), 405–410 (2016).
26. Q. Tao, Z. Guo, Q. xu, W. Jiao, X. Wang, S. Qu, and J. Gao, "Retrieving the polarization information for satellite-to-ground light communication," *J. Opt.* **17**(8), 085701 (2015).
27. D. Li, C. Xu, L. Yan, and Z. Guo, "High-Performance Scanning-mode Polarization based Computational Ghost Imaging (SPCGI)," *Opt. Express* **30**(11), 17909–17921 (2022).
28. X. Li, H. Li, Y. Lin, J. Guo, J. Yang, H. Yue, K. Li, C. Li, Z. Cheng, H. Hu, and T. Liu, "Learning-based denoising for polarimetric images," *Opt. Express* **28**(11), 16309–16321 (2020).
29. H. Hu, Y. Zhang, X. Li, Y. Lin, Z. Z. Cheng, and T. Liu, "Polarimetric underwater image recovery via deep learning," *Optics and Lasers in Engineering* **133**, 106152 (2020).
30. H. Hu, Y. Han, X. Li, L. Jiang, L. Che, T. Liu, and J. Zhai, "Physics-informed neural network for polarimetric underwater imaging," *Opt. Express* **30**(13), 22512–22522 (2022).
31. D. Li, B. Lin, X. Wang, and Z. Guo, "High-Performance Polarization Remote Sensing With the Modified U-Net Based Deep-Learning Network," *IEEE Transactions on Geoscience and Remote Sensing* **60**, 5621110 (2022).
32. S. G. G Stokes, "Mathematical and physical papers [M]," (1901).
33. J. Zhang, J. Shao, J. Chen, D. Yang, B. Liang, and R. Liang, "PFNet: an unsupervised deep network for polarization image fusion," *Opt. Lett.* **45**(6), 1507–1510 (2020).
34. T. Treibitz and Y. Y. Schechner, "Active Polarization Descattering," *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(3), 385–399 (2009).
35. S. Zheng, H. Wang, S. Dong, F. Wang, and G. Situ, "Incoherent imaging through highly nonstatic and optically thick turbid media based on neural network," *Photon. Res.* **9**(5), B220–B228 (2021).
36. Z. Wang, N. Zou, D. Shen, and S. Ji, "Non-local u-nets for biomedical image segmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence* 34(04), 6315–6322 (2020).
37. Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, IEEE 2, 1398–1402 (2003).
38. Y. Zhao, Y. Li, W. He, Y. Liu, and Y. Fu, "Polarization scattering imaging experiment based on Mueller matrix," *Opt. Commun.* **490**(1), 126892 (2021).
39. K. He, J. Sun, and X. Tang, "Guided Image Filtering," *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(6), 1397–1409 (2013).
40. R. Zhang, X. Gui, H. Cheng, and J. Chu, "Underwater image recovery utilizing polarimetric imaging based on neural networks," *Appl. Opt.* **60**(27), 8419–8425 (2021).