Optics Letters

Large-range polarization scattering imaging with an unsupervised multi-task dynamic-modulated framework

Bing Lin, Longyu Qiao, Xueqiang Fan, and Zhongyi Guo*

School of Computer and Information, Hefei University of Technology, Hefei 230601, China
**guozhongyi@hfut.edu.cn*

Received 12 February 2025; revised 11 April 2025; accepted 11 April 2025; posted 14 April 2025; published 13 May 2025

The adaptability of most polarization scattering imaging, such as supervised learning-related and physics-based algorithms, is still limited, especially in dynamic or variable scattering environments. More importantly, it is a challenge to obtain a sufficient amount of polarization data for practical applications. Hence, we propose a multi-task dynamic-modulated polarization de-scattering framework (MDPF) based on the unsupervised training strategy, which achieves real-time de-scattering for a single image, by using the physical transmission model of light in scattering media as well as gradient descent optimization. Our proposed framework can significantly improve adaptability to large-range scattering conditions, without large-scale datasets for training. In addition, the experiments in real haze scenarios verify the effectiveness and superiority of our proposed method, demonstrating the potential of the proposed method in real-world environments, such as remote sensing or underwater imaging.

https://doi.org/10.1364/OL.559415

With the development of deep learning technology, polarization imaging through scattering media based on deep learning (PIDL) has become one of the important solutions for imaging through scattering media. The PIDL methods based on supervised learning have achieved many excellent results [1–4]. Hu *et al.* have proposed the PDN for polarization underwater imaging, which directly used the images with different directions (0°, 45°,90°, 135°) to train the network and obtained excellent recovery results [1]. To analyze the multidimensional polarization information deeply, Lin *et al.* have proposed the SAM-MIU-net to explore maximizing the utilization of polarization information with limited data [2]. Moreover, Fan *et al.* have proposed a method that combined 3D and 2D convolution to balance the extraction of polarization information between different channels, resulting in a more accurate modeling of the road area [3]. However, on the one hand, the performance of supervised learning-based methods relies on a large amount of data; however, the specificity of polarization data detection makes it relatively difficult to acquire polarization image data, especially in real scenes. On the other hand, the trained model is not flexible enough to adapt to the changing dynamic scattering environment due to the limitation of the type and amount of training data.

So, in this paper, we explore the unsupervised strategies. In particular, based on the theory that incorporating a complete physical model representing the image formation process into the deep neural network [5], we designed a multi-task dynamic-modulated polarization de-scattering framework (MDPF). The MDPF effectively combines the physical model with the feature extraction and optimization capabilities of deep learning, creating a synergistic approach for enhanced performance. For common scattering media, such as atmospheric and underwater, the process of information transmission can be modeled using the Jaffe–McGlamery model [6,7], and most de-scattering or de-hazing models [8,9] obtain a clear image by accurately estimating their parameters, especially the transmission rate map. Because backscattered light has a significant polarization dependence [10], polarization is one of the important factors influencing the quality of imaging through scattering media, and several studies have also proved that there is a close relation between the state of polarization of output light and the characteristics of particles in scattering media [11,12]. Therefore, by using polarization information, the transmission process of target information in the scattering media can be analyzed with more accuracy. This enables effective mitigation of the adverse effects of scattering, thereby enhancing the image quality [13,14].

For this reason, we employ the Jaffe–McGlamery model based on polarization information as the physical drive to guide the unsupervised framework. In particular, we design a network architecture based on the inherent properties of the polarization information, enabling the model to fully capture relevant features and provide an accurate removal of scattering effects. Subsequently, the MDPF is optimized using the gradient descent algorithm to obtain precise parameters of the Jaffe–McGlamery model, thereby getting accurate and high-quality images. Notably, the MDPF is trained without relying on large datasets, allowing for real-time optimization of scattering images. So, it demonstrates strong adaptability to the large-range scattering conditions encountered in real-world environments, such as remote sensing or underwater imaging.

The framework of our proposed MDPF is shown in Fig. 1. Here, our proposed framework is divided into three tasks,

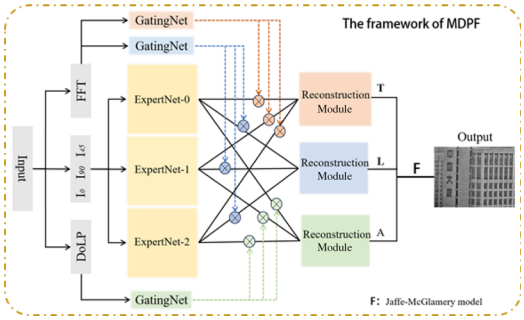

Fig. 1. Framework of MDPF.

Task-L, Task-A, and Task-t, respectively, each of which is used to form parameters in the Jaffe–McGlamery model: the brightness of the scene $L_n(x, y)$, the intensity of ambient light at infinity A_n, and the transmittance $t_n(x, y)$ ((x, y) presents the pixels of the image). The process can be presented following:

$$L_n(x,y) = F_L(f_{in}); A_n = F_A(f_{in}); t_n(x,y) = F_t(f_{in}), \quad (1)$$

where $F_L(.)$, $F_A(.)$, and $F_t(.)$ are the mapping functions of each task and f_{in} is the input feature. Then, the $L_n(x, y)$, A_n, and $t_n(x, y)$ are incorporated into the Jaffe–McGlamery model. Then the degraded image $I_n(x, y)$ calculated by the network can be presented as follows:

$$I_n(x,y) = L_n(x,y)t_n(x,y) + A_n(1 - t_n(x,y)). \quad (2)$$

Then, the mapping function F of MDPF, defined by weights and biases, can be learned from Eq. (3). The weights and bias of the MDPF were optimized by gradient descent using the error between $I_n(x, y)$ calculated by the network and $I(x, y)$ measured in the real environment. As the iterative process progresses, the calculated scattering image $I_n(x, y)$ will converge to the measured scattering image $I(x, y)$. When the error becomes minimum, the network parameters reach optimal, and each task in MDPF will get the effective result. Thereby, the Task-L will get the clear image we need:

$$F_{\theta^*} = \arg\min_{\theta\in\Theta} ||I_n(x,y) - I(x,y)||^2. \quad (3)$$

Importantly, the polarization information has better stability than the light intensity information in the scattering media, and it not only reflects the fine characteristics of the target but also reflects the characteristics of the scattering media during transmission. So, it is also an invaluable tool for imaging in changing environments, and polarization information is more sensitive to the scattering process and scattering distance [15], so the use of polarization information can provide more accurate and precise characteristics for the estimation of the scattering process.

Hence, leveraging the sensitivity and advantages of polarization information in the scattering process is crucial for effectively implementing the unsupervised backscattering scheme. So, we design the corresponding modules in Expert Network (EN) according to the characteristics of polarization information. Considering the multidimensional nature of polarization information, the multi-scale attention mechanism (MAM) is introduced to analyze features at multiple scales. Moreover, we incorporate the Fourier feature extraction (FFE) into the network, enabling it to perform diverse spatial analyses on the extracted features. In addition, we also use gating network (GN) to fuse polarization information according to specific task requirements dynamically. By combining the advantages of polarization information and the characteristics of the MDPF, we can achieve the de-scattering processing of a single image using the unsupervised strategy. This innovative approach allows us to effectively enhance image quality in the presence of scattering media without relying on labeled training data.

The overall structure of MDPF is shown in Fig. 1, which consists of the EN, GN, and reconstruction module (RM). First, we input multidimensional polarization information (I_0, I_{45}, and I_{90}) to the EN-0, EN-1, and EN-2 to obtain the analyzed polarization features. Subsequently, excluding the iterative update driven by the loss function to the algorithm, we also use the GN to provide fusion parameters for adjusting the polarization feature which is obtained by EN. Lastly, the modulated features are input to the RM to reconstruct the corresponding task result. The FLOPs and parameters of MDPF are 38445.3 M and 90.61 M, respectively.

The structure of EN is shown in Fig. 2(a). We innovatively use the Fourier layer (FL) as the FFE module. The FL is based on the spatial Fourier transform module (SFTM), which uses learnable filters and Fourier transform (FT) to process the polarization features. The SFTM is following a similar architecture in [16]. Each FFE consists of four FLs and uses long connections to integrate the features of beginning and end. The SFTM uses a 2D discrete FT to transform the input features into the frequency domain and then transforms them linearly. After the linear transformation, the processed feature is returned to the spatial domain using a 2D discrete inverse FT. Using the SFTM, the initiative and interpretability of the MDPF will be increased through the learnable optimization filter, and in addition, the features can be analyzed from the frequency domain space, thereby the feature information dimension will be increased. What is more, the SFTM can reduce the internal parameters to increase the image inference speed [16]. From Fig. 2(a), after traversing two convolutional layers, the feature is fed into the two layers of FFE, using the channel attention mechanism (CAM) to further enhance the analyzing capacity of this module. Additionally, the output feature from each FFE and CAM will pass through the MAM together. The structure of MAM is shown in Fig. 2(a), the 3×3 convolution with varying expansion rates is used to analyze the polarization features at multiple scales. The model's attention is then focused on key feature expressions, thereby enhancing the role of polarization features throughout the recovery process. Then, the output feature passes through the FFE again. Subsequently, it passes through the self-attention mechanism module (SAM) designed by our previous work [2] together with the output of the CAM.

In GN, as shown in Fig. S1(a) in Supplement 1, the transformer structure (TS) based on [17], which is more focused on global information and better at generating sequences, is introduced to provide weights to modulate features for different tasks. The GN is based on the structure in [18] of us, using three layers of TS instead of convolutional layers as an improvement to analyze the input features. Then the features pass through the fully connected (FC) layer to generate the adjusted parameters. To accommodate the diverse tasks, we input task-specific data into the GN to achieve parameter fusion that optimally facilitates the generation of task results. Therefore, as to the Task-L, we aim to obtain clear target information, so the Fourier image, which has the concentrated expression of the target characteristics and is not susceptible to environmental influences, is used to provide dynamic fusion parameters to adjust the polarization feature to

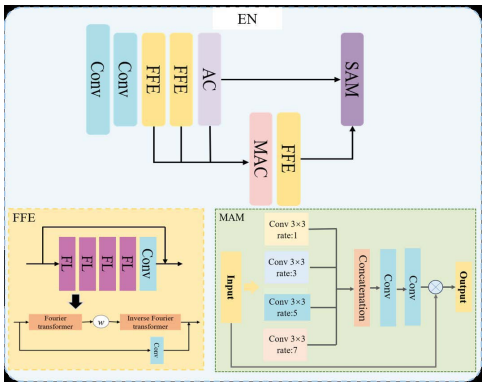

Fig. 2. (a) Structure of EN; (b) structure of GN; (c) structure of RM.

Table 1. Evaluation Metrics for Different Turbidity Levels

	PSNR	Contrast	EN	SD
12.9NTU	13.3778	281.1277	6.0648	55.5993
29.4NTU	12.9747	269.7715	6.3231	31.0542
46.4NTU	11.6337	142.4012	6.0351	25.3822
62.5NTU	10.0440	47.3557	6.7555	22.5029

fusion. For the Task-A and Task-t, we use the DoLP image, which prominently expresses polarization information essential for analyzing scattering conditions, to provide dynamic fusion parameters for estimating the environmental scattering situation and providing a clear representation of the difference between the target and the background.

In Fig. S1(b) in Supplement 1, the RM is composed of the 3 * 3 size of convolution, upsampling, and sigmoid. The features co-modulated by the EN and the GN are input to the RM to get the corresponding task results with the image size of 256×256.

In this paper, we use structural similarity (SSIM) [19] as the loss function to drive the parameter of the MDPF updates. The MDPF works in an image processing unit (NVIDIA RTX 3090) using a Pytorch framework with Python 3.6 in the computational environment that Windows Server 10 (Version 21H1) Inter(R) Core (TM) i7.9750 H CPU at 2.60 Hz, 2.59 GHz, and 16.0 GB of RAM. The optimizer is the Adam (Add Momentum Stochastic Gradient Descent) with a learning rate of 0.0005, $\beta 1 = 0.9$ and $\beta 2 = 0.999$. The MDPF can take 0.2 s for one epoch.

First, we evaluated the performance of our proposed MDPF using the dataset obtained from turbid underwater environments in a controlled laboratory setting. Polarization images are captured in water with turbidity levels ranging from 12.9 NTU to 62.5 NTU (imaging distance fixed at 18 cm). As shown in Figs. 3(b1)–3(b4), the results indicate that our method can effectively adapt to environments with varying scattering levels, successfully recovering the target even at the highest turbidity of 62.5 NTU. These findings also highlight the capability of our method for large-range scattering imaging. In addition, as shown in Table 1, we have calculated the peak signal to noise ratio (PSNR) [20] and contrast to assess the quality of the image, the information entropy (EN) to quantify the richness of image information [21], and the standard deviation (SD) to assess the image clarity [22]. The results demonstrate our method performs well over large-range turbidity environments.

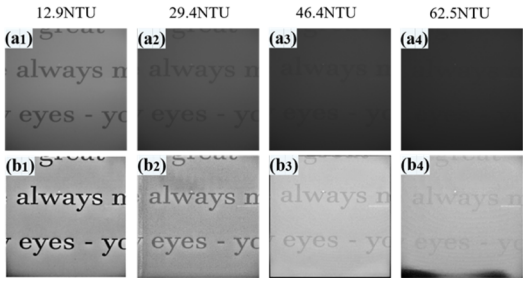

Fig. 3. Test results for different turbidity levels.

Table 2. Comparison Results of Objective Evaluation Metrics

Method	PSNR	Contrast	EN	SD
Ours	13.2705	130.3689	6.8388	29.0023
DCP	8.4858	122.4173	6.3231	11.5443
AoP-Dehaze	12.4028	76.2565	6.0351	15.7260
TIUNet	11.1808	15.4700	6.7555	30.0892
PID2Net	11.6524	38.2182	6.5293	21.4864

To demonstrate the viability of our proposed approach, we test it by using scattering images taken in real scenes by a polarization camera with 2480×1860 pixels. The real-world dataset consists of scattering images captured under varying degrees of severe haze, with the haze intensity increasing from left to right in Figs. 4(a1)–4(a4). These real-world scenarios involve more complex scattering media and more intricate target structures, reflecting the challenges encountered in practical applications. From Figs. 4(b1)–4(b4), the results of our method are able to recover the target clearly. From light fog to dense fog (left to right), the results are all with clear and high contrast. What is more, we compare our framework with other de-scattering methods in order to demonstrate the superiority of the MDPF. Based on Fig. 4, the results obtained by our method (MDPF) demonstrate more realistic, with satisfactory brightness, and higher contrast, and we have found that the image quality of the DCP [23] method is dark to some extents, and the recovery effect varies considerably depending on the degree of scattering, as shown in Figs. 4(c1)–4(c4). Then, the AoP-based denoising method [24] has limitations in recovering details, as shown in Figs. 4(d1)–4(d4). From Figs. 4(e1)–4(e4) and Figs. 4(f1)–4(f4), the TIUNet [25] and PID2Net [26] exhibit limited generalization performance when applied to outdoor scenes with significantly different scattering environments, primarily due to the constraints of the training dataset. While these methods can capture the general outline of the scene, they struggle to recover fine details, resulting in a substantial loss of information, and we calculate the corresponding estimated indicators in Table 2.

Compared with other methods, the images obtained by MDPF have better image clarity and information richness, consistent with the results of subjective evaluation indicators. What is more, we enlarge the details of the image rebuilt by DCP and the AoP-based method, the image details can be seen in the red box and blue box, respectively. The details of our MDPF have a high resolution and are capable of restoring the scene's intricacies, as shown in the red box of Fig. 5(a1) and the blue box of Fig. 5(a2). The fine-grained features are faithfully preserved, showcasing the effectiveness of our approach in capturing and reconstructing intricate scene details. The performance of the DCP is that

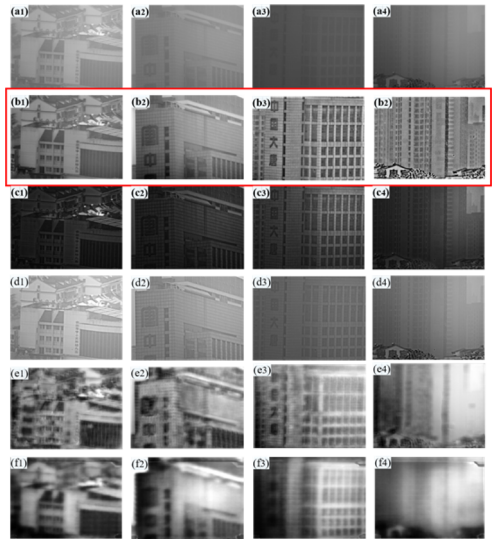

Fig. 4. Results of different models. (a1)–(a4) Scattering image; (b1)–(b4) results of MDPF; (c1)–(c4) results of DCP; (d1)–(d4) results of the AoP-based denoising method; (e1)–(e4) results of the TIUNet; (f1)–(f4) results of the PID2Net.

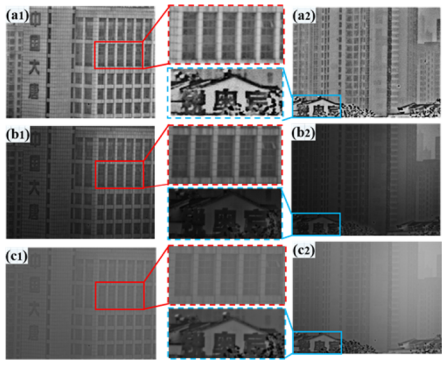

Fig. 5. Enlarged details of images with different methods. (a1) and (a2) MDPF; (b1) and (b2) DCP; (c1) and (c2) AoP-based denoising method.

the brightness of detail is too low, making it difficult to see the Chinese character target in the blue box of Fig. 5(b2). Also, the results of the AoP-based denoising method have poor contrast findings, making it difficult to clearly distinguish between targets in the red box of Fig. 5(c1).

Accurate analysis of the transmission map of the scene is of great significance in improving the imaging quality. In our analysis, we observed that the transmission map obtained by our method outperforms the one obtained by the DCP method as shown in Fig. S2 in Supplement 1. Our method effectively eliminates most of the scattering effects, corrects inaccurate transmittance estimation, and yields a transmission map with higher contrast from Figs. S2(b1)–S2(b4) in Supplement 1. These improvements contribute to enhancing the final imaging effect, resulting in a higher-quality image compared to existing methods. These results also further validate the efficacy of our proposed method without labeled data in accurately estimating the scene scattering information by leveraging polarization features and a meticulously engineered network architecture.

In conclusion, we have successfully combined the physical model of imaging with the network structure to propose an unsupervised polarization de-scattering scheme. Because the incorporation of polarization prior and imaging models effectively guides and activates the neural networks to extract features, the framework proposed in this paper can constrain the training process without labeled data. Thus, the proposed MDPF can recover targets in real time, providing enhanced adaptability to the large-range scattering environments. In the actual scene, the MDPF is capable of obtaining high-quality and clear images. What is more, it can handle varying levels of haze and can identify the details of distant targets. This work demonstrates the potential for deep learning methods combined with physical models to drive progress in de-scattering research.

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.

Supplemental document. See Supplement 1 for supporting content.

REFERENCES

1. H. Hu, Y. Zhang, X. Li, *et al.*, Opt. Lasers Eng. **133**, 106152 (2020).
2. B. Lin, X. Fan, and Z. Guo, Opt. Express **31**, 3046 (2023).
3. X. Fan, B. Lin, and Z. Guo, IEEE Trans. Intell. Transp. **25**, 12762 (2024).
4. K. Li, M. Qi, S. Zhuang, *et al.*, Opt. Lett. **47**, 4255 (2022).
5. F. Wang, Y. Bian, H. Wang, *et al.*, Light: Sci. Appl. **9**, 77 (2020).
6. B. L. McGlamery, Proc. SPIE **0208**, 221 (1980).
7. S. G. Narasimhan and S. K. Nayar, Int. J. Comput. Vision **48**, 233 (2002).
8. K. He, J. Sun, and X. Tang, IEEE Trans. Pattern Anal. Mach. Intell. **35**, 1397 (2013).
9. L. Shen, Y. Zhao, Q. Peng, *et al.*, IEEE Trans. Multimedia **21**, 1093 (2019).
10. F. C. MacKintosh, J. X. Zhu, D. J. Pine, *et al.*, Phys. Rev. B **40**, 9342 (1989).
11. T. Hu, F. Shen, K. Wang, *et al.*, Atmosphere **10**, 342 (2019).
12. F. Shen, B. Zhang, K. Guo, *et al.*, IEEE Photonics J. **10**, 3900212 (2018).
13. X. Li, J. Xu, L. Zhang, *et al.*, Opt. Lett. **47**, 2854 (2022).
14. D. Li, C. Xu, M. Zhang, *et al.*, Biomed. Opt. Express **12**, 2447 (2021).
15. G. Tremblay and G. Roy, Appl. Opt. **60**, 1217 (2021).
16. Z. Li, N. B. Kovachki, K. Azizzadenesheli, *et al.*, *The 9th International Conference on Learning Representations (ICLR-21)* (2021).
17. A. Vaswani, N. M. Shazeer, N. Parmar, *et al.*, "Attention is All You Need," arXiv (2017).
18. B. Lin, X. Fan, P. Peng, *et al.*, Opt. Express **32**, 511 (2024).
19. Z. Wang and A. C. Bovik, IEEE Signal Process. Lett. **9**, 81 (2002).
20. H. Zhao, O. Gallo, I. Frosio, *et al.*, IEEE Trans. Comput. Imaging **3**, 47 (2017).
21. Y. J. Rao, Meas. Sci. Technol. **8**, 355 (1997).
22. Z. Wang, A. C. Bovik, H. R. Sheikh, *et al.*, IEEE Trans. Image Process. **13**, 600 (2004).
23. K. He, J. Sun, and X. Tang, IEEE Trans. Pattern Anal. Mach. Intell. **33**, 2341 (2011).
24. J. Liang, L. Ren, H. Ju, *et al.*, Opt. Express **23**, 26146 (2015).
25. B. Lin, W. Chen, X. Fan, *et al.*, Opt. Laser Technol. **181**, 111664 (2025).
26. H. Liu, W. Zhang, Y. Han, *et al.*, IEEE Sens. J. **24**, 27803 (2024).