

LEARNING A PHYSICAL-AWARE DIFFUSION MODEL BASED ON TRANSFORMER FOR UNDERWATER IMAGE ENHANCEMENT

Chen Zhao, Chenyu Dong, Weiling Cai[†]

School of Artificial Intelligence, Nanjing Normal University

ABSTRACT

Underwater visuals undergo various complex degradations, inevitably influencing the efficiency of underwater vision tasks. Recently, diffusion models were employed to underwater image enhancement (UIE) tasks, and gained SOTA performance. However, these methods fail to consider the physical properties and underwater imaging mechanisms in the diffusion process, limiting information completion capacity of diffusion models. In this paper, we introduce a novel UIE framework, named PA-Diff, designed to exploiting the knowledge of physics to guide the diffusion process. PA-Diff consists of Physics Prior Generation (PPG) Branch and Physics-aware Diffusion Transformer (PDT) Branch. Our designed PPG branch is a plug-and-play network to produce the physics prior, which can be integrated into any deep framework. With utilizing the physics prior knowledge to guide the diffusion process, PDT branch can obtain underwater-aware ability and model the complex distribution in real-world underwater scenes. Extensive experiments prove that our method achieves best performance on UIE tasks.

Index Terms— Underwater image enhancement, physics model, diffusion model, transformer

1. INTRODUCTION

The restoration of images beneath the water’s surface is a pragmatic yet intricate technology within the realm of underwater vision, extensively applied in endeavors such as underwater robotics[1] and tracking underwater objects[2]. Owing to the phenomena of light refraction, absorption, and scattering in underwater surroundings, images captured underwater are often subject to substantial distortion, characterized by diminished contrast and inherent blurriness [3]. Consequently, the acquisition of clear images in underwater scenarios assumes paramount importance, especially in disciplines necessitating interactions with the submerged environment. The primary objective of underwater image enhancement (UIE) lies in the attainment of superior-quality images, achieved through the elimination of scattering effects and rectification of color distortions prevalent in degraded images. UIE stands as an indispensable component for tasks related to vision in the underwater domain.

To address this problem, conventional UIE approaches, which hinge on the intrinsic characteristics of underwater images, have been presented [4, 5, 6]. These techniques delve into the physical aspects contributing to degradation, such as color cast or scattering, seeking to rectify and enhance the underwater images. Nevertheless, these models grounded in physics exhibit restricted representational capacities, rendering them insufficient for comprehensively addressing the intricate physical and optical elements inherent in underwater scenes. Consequently, their efficacy diminishes, yielding sub-optimal enhancement outcomes, particularly in the face of highly intricate and diverse underwater scenarios. In recent times, many learning-based methodologies [7, 8, 9, 10] have been introduced to yield superior results. Leveraging the potent feature representation and nonlinear mapping capabilities of neural networks, these methods excel. They can effectively learn the transformation from a degraded image to a clear one through extensive paired training data.

Recently, there has been a surge of interest in image synthesis [11, 12] and restoration tasks [13, 14, 15], with a notable focus on diffusion-based techniques like DDPM [16] and DDIM [17]. This heightened attention can be attributed to the remarkable generative capabilities of diffusion models. These methods can gradually recover images from noisy images generated in the forward diffusion process and achieve high-quality mapping from randomly sampled Gaussian noise to target images in the reverse diffusion process [16], which poses a new perspective for underwater image enhancement task. Tang et al. [7] present an image enhancement approach with diffusion model in underwater scenes, and WF-Diff [18] proposed a underwater image enhancement framework based on frequency domain information and diffusion models. However, these methods fail to consider the physical properties and underwater imaging mechanisms in the diffusion process, limiting their information completion capacity. Moreover, the diffusion models lack awareness regarding the difficulty of enhancing different image regions, a critical consideration for modeling the complex distribution in real-world underwater scenes.

In this paper, we develop a novel physics-aware diffusion model, fully exploiting physical information to guide the diffusion process, called PA-Diff. It mainly consists of two branches: Physics Prior Generation (PPG) Branch and

Physics-aware Diffusion Transformer (PDT) Branch. The first branch aims to generate the transmission map and the global background light as physics prior information by utilizing the modified Koschmieder light scanning model. The transmission map is exploited as the confidence guidance for the PDT branch which enables our PA-Diff with underwater-aware ability. The PDT branch employs the strong generation ability of diffusion models to restore underwater images, with the guidance of physical information. Specifically, we design a physics-aware diffusion transformer block which contains a physics-aware self-attention (PA-SA) and multi-scale dynamic feed-forward network (MS-FFN) to exploit physics prior information and capture the long-range diffusion dependencies. Our PA-Diff achieves unprecedented perceptual performance in UIE task. Extensive experiments demonstrate that our developed PA-Diff performs the superiority against previous UIE approaches, and ablation study can demonstrate the effectiveness of all contributions.

In summary, the main contributions of our PA-Diff are as follows:

- We propose a novel UIE framework based on physics-aware diffusion model, named PA-Diff, which consists of Physics Prior Generation (PPG) Branch and Physics-aware Diffusion Transformer (PDT) Branch. To the best of our knowledge, it is the first the Diffusion model with the guidance of physical knowledge in underwater image enhancement tasks.
- We design a physics-aware diffusion transformer block, which not only enables PA-Diff with underwater-aware ability to guide the diffusion process, but also captures the long-range diffusion dependencies.
- Extensive experiments compared with SOTAs considerably demonstrate that our PA-Diff performs the superiority against previous UIE approaches, and extensive ablation experiments can demonstrate the effectiveness of all contributions.

2. RELATED WORKS

2.1. Underwater Image Enhancement

Currently, existing UID methods can be briefly categorized into the physical and deep model-based approaches [4, 6, 7, 8, 9]. Most UID methods based on the physical model utilize prior knowledge to establish models, such as water dark channel priors [4], attenuation curve priors [19], fuzzy priors [20]. In addition, Akkaynak and Treibitz [21] proposed a method based on the revised physical imaging model. The manually established priors restrain the model’s robustness and scalability under the complicated and varied circumstances. Recently, deep learning-based methods [7, 8, 9] achieve acceptable performance, and some complex frameworks are pro-

posed and achieve the-state-of-the-art performance [4, 5, 22]. The previous methods neglect the underwater imaging mechanism and rely only on the representation ability of deep networks. Ucolor [23] combined the underwater physical imaging model in the raw space and designed a medium transmission guided model. ATDCnet [24] attempted to use transmission maps to guide deep neural networks. Therefore, how to fully exploit the knowledge of physics in deep neural networks is a very crucial problem.

2.2. Diffusion Model

Diffusion Probabilistic Models (DPMs) [16, 17] have been widely adopted for conditional image generation [13, 14, 25, 26, 27]. Saharia et al. [28] proposes Palette, which has demonstrated the excellent performance of diffusion models in the field of conditional image generation, including colorization, in-painting and JPEG restoration. Currently, diffusion models was employed to a variety of low-level visual tasks, and gained SOTA performance such as image restoration [29] and low light image enhancement [30]. Recently, Tang et al. [7] presented an image enhancement approach with diffusion model in underwater scenes. WF-Diff [18] proposed a underwater image enhancement framework based on frequency domain information and diffusion models. However, these methods fail to consider the physical properties and underwater imaging mechanisms in the diffusion process, limiting information completion capacity of diffusion models.

3. METHODOLOGY

3.1. Overall Framework

The overall framework of PA-Diff is shown in Fig. 1. PA-Diff is designed to leverage the physical properties of underwater imaging mechanisms to guide the diffusion process, which mainly consists of Physics Prior Generation (PPG) Branch and Physics-aware Diffusion Transformer (PDT) Branch. The PPG branch aims to generate the transmission map and the global background light as physics prior information by utilizing the modified Koschmieder light scanning model. The transmission map is exploited as the confidence guidance for the PDT branch which enables our PA-Diff with underwater-aware ability. The PDT branch employs the strong generation ability of diffusion models to restore underwater images, with the guidance of physical information. We adopt the diffusion process proposed in DDPM [16] to construct the PDT branch, which can be described as a forward diffusion process and a reverse diffusion process.

Forward Diffusion Process. The forward diffusion process can be viewed as a Markov chain progressively adding Gaussian noise to the data. Given a clean data x_0 , then introduce Gaussian noise based on the time step, as follows:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t I), \quad (1)$$

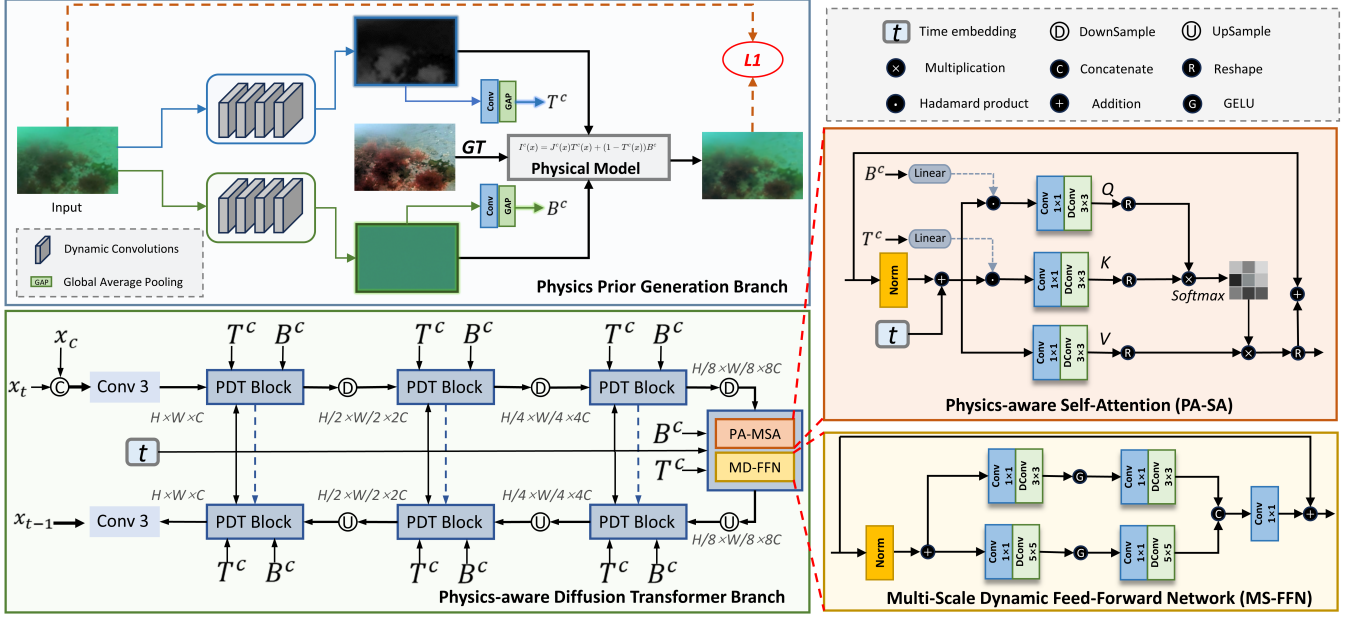


Fig. 1. Overall framework of our proposed PA-Diff. PA-Diff mainly consists of two cooperative branches: Physics Prior Generation (PPG) Branch and Physics-aware Diffusion Transformer (PDT) Branch. Our designed PPG branch is a plug-and-play universal module to produce the prior knowledge of physics, which can be integrated into any deep learning framework. With utilizing the physics prior knowledge to guide the diffusion process, PDT branch can obtain underwater-aware ability and model the complex distribution in real-world underwater scenes.

where β_t is a variable controlling the variance of the noise. Introducing $\alpha_t = 1 - \beta_t$, this process can be described as:

$$x_t = \sqrt{\alpha_t}x_{t-1} + \sqrt{1 - \alpha_t}\epsilon_{t-1}, \quad \epsilon_{t-1} \sim \mathcal{N}(0, \mathcal{Z}). \quad (2)$$

With Gaussian distributions are merged, We can obtain :

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1 - \bar{\alpha}_t)I). \quad (3)$$

Reverse Diffusion Process. The reverse diffusion process aims to restore the clean data from the Gaussian noise. The reverse diffusion can be expressed as:

$$p_\theta(x_{t-1}|x_t, x_c) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, x_c, t), \sigma_t^2 \mathcal{Z}), \quad (4)$$

where x_c refers to the conditional image I (underwater input image). $\mu_\theta(x_t, x_c, t)$ and σ_t^2 are the mean and variance from the estimate of step t , respectively. We follow the setup of [16], they can be expressed as:

$$\mu_\theta(x_t, x_c, t) = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{(1 - \bar{\alpha}_t)}\epsilon_\theta(x_t, x_c, t)), \quad (5)$$

$$\sigma_t^2 = \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t}\beta_t, \quad (6)$$

where $\epsilon_\theta(x_t, x_c, t)$ is the estimated value with a Unet.

We optimize an objective function for the noise estimated by the network and the noise ϵ actually added. Therefore, the diffusion loss is:

$$L_{dm}(\theta) = \|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, x_c, t)\|. \quad (7)$$

3.2. Physics Prior Generation

Mathematically, the underwater images can be expressed by the modified Koschmieder light scanning model:

$$I^c(x) = J^c(x)T^c(x) + (1 - T^c(x))B^c, \quad (8)$$

where $c \in \{R, G, B\}$ is the color channel, I^c is the observed underwater image, J^c is the restored clean image, $T^c(x)$ is the medium transmission map, and B^c is the global background scattered light. In Physics Prior Generation (PPG) branch, our goal is to produce $T^c(x)$ and B^c from the underwater image I^c , exploiting the physics model.

Network architecture of PPG branch is shown in Fig. 1, and consists of two sub-networks: transmission map generation sub-network $\mathcal{T}(x)$ and background light generation sub-network $\mathcal{B}(x)$. The two sub-networks consist of duplicated Dynamic Convolutions (DC) [31]. DC dynamically combines numerous parallel convolutional kernels according to their attention, thereby enhancing the model's complexity and performance without augmenting the network's depth or breadth.

Given a underwater image as input I and its corresponding ground truth (GT), we can obtain via Eq. 8:

$$\hat{I} = GT \cdot \mathcal{T}(I) + (1 - \mathcal{T}(I))\mathcal{B}(I), \quad (9)$$

We constrain the PPG branch by supervised manner. The physics reconstruction loss can be expressed as:

$$\mathcal{L}_p = \|\hat{I} - I\|_1, \quad (10)$$

Finally, we use a Conv and GAP to extract physics prior information \mathbf{T}^c and \mathbf{B}^c to guide the diffusion process.

3.3. Physics-aware Diffusion Transformer

Physics-aware diffusion transformer (PDT) branch aims to exploit the extracted physics prior information from PPG branch for better guiding underwater image restoration. Fig. 1 shows the overall of PDT branch, which is a U-shaped structure with physics-aware diffusion transformer blocks (PDTB). We also utilize skip connections to connect the features at the same level.

Given a degraded underwater image as condition image $x_c \in R^{H \times W \times 3}$, we first obtain the noise sample $x_t \in R^{H \times W \times 3}$ at time step t according to the forward process, and concatenate x_t and x_c at channel dimension to obtain the input of PDT branch. Then, we obtain the embedding features $\mathbf{F} \in R^{H \times W \times C}$ through convolution, where C means the number of channel. \mathbf{F} is encoded and decoded by PDTB, which consists of physics-aware self-attention (PA-SA) and multi-scale dynamic feed-forward network (MS-FFN). In our PA-MSA, We first embed the time embedding \mathbf{T} into the input features \mathbf{F} :

$$\tilde{\mathbf{F}} = \text{Norm}(\mathbf{F}) + \mathbf{T}, \quad (11)$$

where Norm means layer normalization. Then, we integrated \mathbf{T}^c and \mathbf{B}^c as the corresponding dynamic modulation parameters to generate transmission-aware feature \mathbf{F}_t and light-aware feature \mathbf{F}_b :

$$\begin{aligned} \mathbf{F}_t &= Li(\mathbf{T}^c) \odot \tilde{\mathbf{F}}, \\ \mathbf{F}_b &= Li(\mathbf{B}^c) \odot \tilde{\mathbf{F}}, \end{aligned} \quad (12)$$

where $Li()$ means linear layer. Afterward, we aggregate the obtained embeddings by projecting \mathbf{F}_b into query $\mathbf{Q} = \mathbf{W}_d \mathbf{W}_p \mathbf{F}_b$, \mathbf{F}_t into key $\mathbf{K} = \mathbf{W}_d \mathbf{W}_p \mathbf{F}_t$ and transforming $\tilde{\mathbf{F}}$ into value $\mathbf{V} = \mathbf{W}_d \mathbf{W}_p \tilde{\mathbf{F}}$, where \mathbf{W}_p and \mathbf{W}_d respectively denote 1×1 point-wise convolution and 3×3 depth-wise convolution. We employ self-attention and get the output of PA-SA $\hat{\mathbf{F}}$:

$$\hat{\mathbf{F}} = \mathbf{F} + \text{softmax}(\mathbf{Q}\mathbf{K}^T/\alpha) \cdot \mathbf{V}, \quad (13)$$

where α is a learnable parameter. Consequently, PA-SA introduces physical guidance to fully exploit physics knowledge at the feature level and use self-attention mechanism to implicitly model the features of transmission map and background scattered light, which can help the diffusion model restore missing details and correct color distortion.

Finally, we design a multi-scale dynamic feed-forward network (MS-FFN) for local feature aggregation. In order to expand the receptive field, we employ multi-scale kernel depth-

wise convolutions. MS-FFN adopts GELU to ensure the flexibility of feature aggregation. Thus, the multi-scale feature \mathbf{F}^3 and \mathbf{F}^5 of MS-FFN can be expressed as:

$$\begin{aligned} \mathbf{F}^3 &= \mathbf{W}_d^3 \mathbf{W}_p(GELU(\mathbf{W}_d^3 \mathbf{W}_p(\text{Norm}(\hat{\mathbf{F}})))), \\ \mathbf{F}^5 &= \mathbf{W}_d^5 \mathbf{W}_p(GELU(\mathbf{W}_d^5 \mathbf{W}_p(\text{Norm}(\hat{\mathbf{F}})))), \end{aligned} \quad (14)$$

The output feature F_{out} of MS-FFN can be expressed as:

$$\mathbf{F}_{out} = \text{Conv}(\text{Concat}(\mathbf{F}^3, \mathbf{F}^5)) + \hat{\mathbf{F}}. \quad (15)$$

4. EXPERIMENTS

4.1. Experimental Settings

Implementation details. Our network, implemented using PyTorch 1.7, underwent training and testing on an NVIDIA GeForce RTX 3090 GPU. We employed the Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate was established at 0.0001. During the training stage, to balance the batch size and image size, the batch and image size are set to 8 and 128×128 , respectively. The number of training iterations reached one million. The pixel values of the image are normalized to $[-1, 1]$. The time step of the diffusion model is set to 2000. In the testing stage, following the setting in [7], the input size of the image is 256×256 . By using the skip sampling strategy, and the sampling times are set to 10 times to balance the performance and runtime.

Datasets. We utilize the real-world UIEBD dataset [8] and the LSUI dataset [9] for training and evaluating our model. The UIEBD dataset comprises 890 underwater images with corresponding labels. Out of these, 700 images are allocated for training, and the remaining 190 are designated for testing. LSUI dataset contains 5004 underwater images and their corresponding high-quality images. Compared with UIEBD, LSUI contains diverse underwater scenes, object categories and deep-sea and cave images. In the paper, The LSUI dataset is randomly partitioned into 4500 images for training and 504 images for testing.

Evaluation Metrics. We employ standard full-reference image quality assessment metrics, including PSNR and SSIM [32], for quantitative comparisons at both pixel and structural levels. Higher PSNR and SSIM values indicate superior image quality. Additionally, LPIPS [33] and FID [34] are utilized to evaluate perceptual performance. Lower LPIPS and FID scores signify a more effective UIE approach.

4.2. Results and Comparisons

Table 1 shows the quantitative results compared with different SOTA methods on the UIEBD and LSUI datasets, including with UIEC²-Net [35], Water-Net [8], UIEWD [36], UWCNN [37], U-color [23], U-shape [9], and DM-water [7]. We mainly use PSNR, SSIM, LPIPS and FID as our quantitative indices for UIEBD and LSUI datasets. The results

Table 1. Quantitative comparison on the UIEBD and LSUI datasets. The best results are highlighted in bold and the second best results are underlined.

	Methods	UIEWD	UWCNN	UIEC ² -Net	Water-Net	U-color	U-shape	DM-water	Ours
UIEBD	FID↓	85.12	94.44	35.06	37.48	38.25	46.11	<u>31.07</u>	28.76
	LPIPS↓	0.3956	0.3525	0.2033	0.2116	0.2337	0.2264	<u>0.1436</u>	0.1324
	PSNR↑	14.65	15.40	20.14	19.35	20.71	<u>21.25</u>	21.88	21.14
	SSIM↑	0.7265	0.7749	0.8215	0.8321	0.8411	<u>0.8453</u>	0.8194	0.8620
LSUI	FID↓	98.49	100.5	34.51	38.90	45.06	28.56	<u>27.91</u>	22.15
	LPIPS↓	0.3962	0.3450	0.1432	0.1678	0.123	<u>0.1028</u>	0.1138	0.0923
	PSNR↑	15.43	18.24	20.86	19.73	22.91	24.16	27.65	<u>25.89</u>
	SSIM↑	0.7802	0.8465	0.8867	0.8226	0.8902	<u>0.9322</u>	0.8867	0.9354

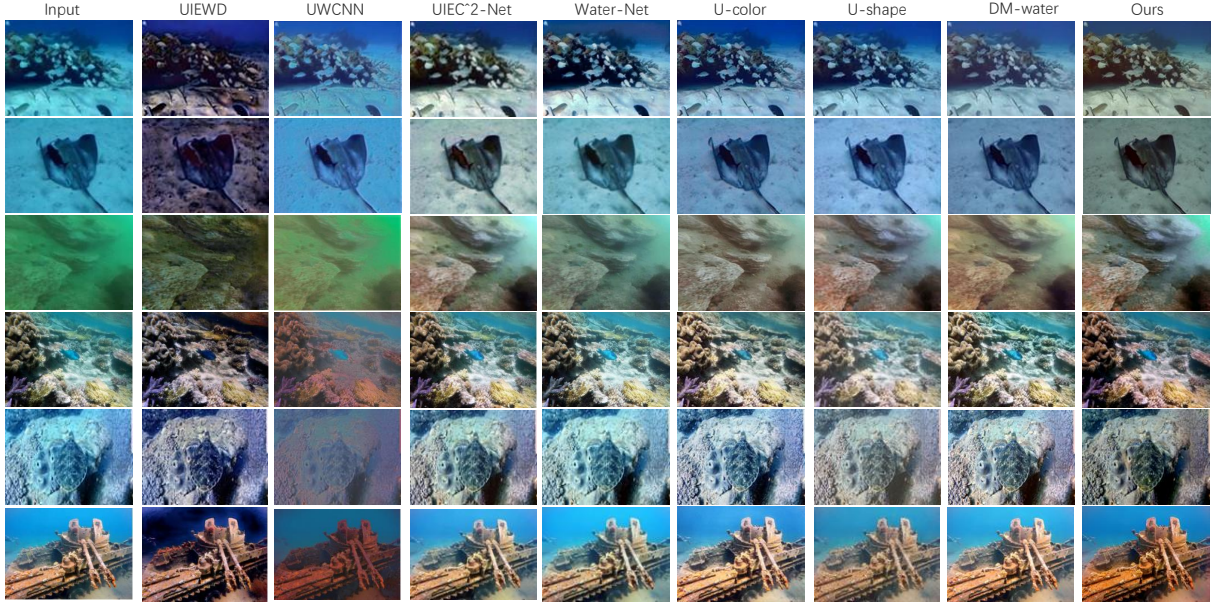


Fig. 2. Qualitative comparison with other SOTA methods on the UIEBD and LSUI datasets.

in Table 1 show that our algorithm outperforms state-of-the-art methods obviously, especially in terms of perceptual metrics. To better validate the superiority of our methods, Fig. 2 shows the visual comparison results with other methods. The six examples are randomly selected on the UIEBD and LSUI datasets. Our methods consistently generate natural and better visual results, strongly proving that PA-Diff has good generalization performance for real-world underwater images.

4.3. Ablation Study

To evaluate the impact of each strategy on the diffusion model, we conduct an ablation study. Table 2 shows the ablation results on the LSUI dataset. Model A means that we just use the normal transformer structure. Compared with model A, model B, C and D achieve better results, suggesting that the knowledge of physics can help diffusion model to better restore image. Model B outperforms C, suggesting that

the prior information of the transmission map is more important than the background light. Model E achieves the best performance, proving that our all contribution is effective for UIE tasks. Model E underperforms model F. The possible reason for this is that there is still a gap between the generated physics information (the transmission map and the background light) and the ideal true value. Therefore, how to generate more accurate a prior knowledge of physics remains a problem we need to research in the future.

5. CONCLUSION

In this paper, we develop a novel UIE framework, namely PA-Diff. With utilizing physics prior information to guide the diffusion process, PA-Diff can obtain underwater-aware ability and model the complex distribution in real-world underwater scenes. To the best of our knowledge, it is the first the Diffu-

Table 2. Ablation study on LSUI dataset. T and B refer to transmission map and background light prior information, respectively. MS means multi-scale operation in our designed MS-FFN. GAP denotes that we use global average pooling for the extraction of physical prior information from both the generated transmission map and background light.

Method	T	B	MS	GAP	FID↓	SSIM↑
A	×	×	×	×	28.35	0.8943
B	✓	×	×	✓	24.71	0.9193
C	×	✓	×	✓	25.84	0.9077
D	✓	✓	×	✓	23.06	0.9298
E	✓	✓	✓	×	25.44	0.9177
F	✓	✓	✓	✓	22.15	0.9354

sion model based on physical perception in image enhancement tasks. Our designed physics prior generation branch is a plug-and-play universal module to produce physics prior information, which can be integrated into any deep learning framework. PA-Diff shows SOTA performance on UIE task, and extensive ablation experiments can prove that each of our contributions is effective.

6. REFERENCES

- [1] James McMahon and Erion Plaku, "Autonomous data collection with timed communication constraints for unmanned underwater vehicles," *IEEE Robotics Autom. Lett.*, vol. 6, no. 2, pp. 1832–1839, 2021.
- [2] Karin de Langis and Junaed Sattar, "Realtime multi-diver tracking and identification for underwater human-robot collaboration," in *2020 IEEE International Conference on Robotics and Automation, ICRA 2020, Paris, France, May 31 - August 31, 2020*, 2020, pp. 11140–11146.
- [3] Derya Akkaynak, Tali Treibitz, Tom Shlesinger, Yossi Loya, Raz Tamir, and David Iluz, "What is the space of attenuation coefficients in underwater computer vision?," in *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pp. 568–577.
- [4] Yan-Tsung Peng and Pamela C. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1579–1594, 2017.
- [5] Paulo Drews Jr., Erickson Rangel do Nascimento, F. Moraes, Silvia S. C. Botelho, and Mario F. M. Campos, "Transmission estimation in underwater single images," in *2013 IEEE International Conference on Computer Vision Workshops, ICCV Workshops 2013, Sydney, Australia, December 1-8, 2013*, pp. 825–830.
- [6] Yan-Tsung Peng, Keming Cao, and Pamela C. Cosman, "Generalization of the dark channel prior for single image restoration," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2856–2868, 2018.
- [7] Yi Tang, Hiroshi Kawasaki, and Takafumi Iwaguchi, "Underwater image enhancement by transformer-based diffusion model with non-uniform sampling for skip strategy," in *Proceedings of the 31st ACM International Conference on Multimedia, MM 2023, Ottawa, ON, Canada, 29 October 2023- 3 November 2023*, 2023, pp. 5419–5427, ACM.
- [8] Lintao Peng, Chunli Zhu, and Liheng Bian, "U-shape transformer for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 32, pp. 3066–3079, 2023.
- [9] Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2020.
- [10] Cameron Fabbri, Md Jahidul Islam, and Junaed Sattar, "Enhancing underwater imagery using generative adversarial networks," in *2018 IEEE International Conference on Robotics and Automation, ICRA 2018, Brisbane, Australia, May 21-25, 2018*, pp. 7159–7165, IEEE.
- [11] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer, "High-resolution image synthesis with latent diffusion models," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, pp. 10674–10685, IEEE.
- [12] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L. Denton, Seyed Kamyar Seyed Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, Jonathan Ho, David J. Fleet, and Mohammad Norouzi, "Photorealistic text-to-image diffusion models with deep language understanding," in *NeurIPS*, 2022.
- [13] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon, "ILVR: conditioning method for denoising diffusion probabilistic models," in *2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021*, pp. 14347–14356, IEEE.
- [14] Yinhuai Wang, Jiwen Yu, and Jian Zhang, "Zero-shot image restoration using denoising diffusion null-space model," in *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*, 2023.
- [15] Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, and Jiayi Ma, "Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model," *CoRR*, vol. abs/2308.13164, 2023.
- [16] Jonathan Ho, Ajay Jain, and Pieter Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.
- [17] Jiaming Song, Chenlin Meng, and Stefano Ermon, "Denoising diffusion implicit models," in *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*, 2021.
- [18] Chen Zhao, Weiling Cai, Chenyu Dong, and Chengwei Hu, "Wavelet-based fourier information interaction with frequency diffusion adjustment for underwater image restoration," *arXiv preprint arXiv:2311.16845*, 2023.
- [19] Yi Wang, Hui Liu, and Lap-Pui Chau, "Single underwater image restoration using adaptive attenuation-curve prior," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 65-I, no. 3, pp. 992–1002, 2018.
- [20] John Yi-Wu Chiang and Ying-Ching Chen, "Underwater image enhancement by wavelength compensation and dehazing," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1756–1769, 2012.
- [21] Derya Akkaynak and Tali Treibitz, "Sea-thru: A method for removing water from underwater images," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pp. 1682–1691, Computer Vision Foundation / IEEE.
- [22] Chen Zhao, Weiling Cai, Chenyu Dong, and Ziqi Zeng, "Toward sufficient spatial-frequency interaction for gradient-aware underwater image enhancement," *arXiv preprint arXiv:2309.04089*, 2023.
- [23] Chongyi Li, Saeed Anwar, Junhui Hou, Runmin Cong, Chunle Guo, and Wenqi Ren, "Underwater image enhancement via medium transmission-guided multi-color space embedding," *IEEE Trans. Image Process.*, vol. 30, pp. 4985–5000, 2021.
- [24] Pan Mu, Jing Fang, Haotian Qian, and Cong Bai, "Transmission and color-guided network for underwater image enhancement," in *2023 IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, 2023, pp. 1337–1342.
- [25] Shilin Lu, Yanzhu Liu, and Adams Wai-Kin Kong, "Tf-icon: Diffusion-based training-free cross-domain image composition," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 2294–2305.
- [26] Dewei Zhou, Zongxin Yang, and Yi Yang, "Pyramid diffusion models for low-light image enhancement," *arXiv preprint arXiv:2305.10028*, 2023.
- [27] Dewei Zhou, You Li, Fan Ma, Zongxin Yang, and Yi Yang, "Migc: Multi-instance generation controller for text-to-image synthesis," *arXiv preprint arXiv:2402.05408*, 2024.
- [28] Chitwan Saharia, William Chan, Huiwen Chang, Chris A. Lee, Jonathan Ho, Tim Salimans, David J. Fleet, and Mohammad Norouzi, "Palette: Image-to-image diffusion models," in *SIGGRAPH '22: Special Interest Group on Computer Graphics and Interactive Techniques Conference, Vancouver, BC, Canada, August 7 - 11, 2022*, 2022, pp. 15:1–15:10, ACM.
- [29] Ozan Özdencizci and Robert Legenstein, "Restoring vision in adverse weather conditions with patch-based denoising diffusion models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.
- [30] Shuzhou Wang, Xuanyu Zhang, Yinhuai Wang, Jiwen Yu, Yuhang Wang, and Jian Zhang, "Diffille: Diffusion-guided domain calibration for unsupervised low-light image enhancement," *arXiv preprint arXiv:2308.09279*, 2023.
- [31] Yinpeng Chen, Xiyang Dai, Mengchen Liu, Dongdong Chen, Lu Yuan, and Zicheng Liu, "Dynamic convolution: Attention over convolution kernels," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11030–11039.
- [32] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.
- [33] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pp. 586–595.
- [34] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, 2017, pp. 6626–6637.
- [35] Yudong Wang, Jichang Guo, Huan Gao, and Huihui Yue, "Uic²-net: Cnn-based underwater image enhancement using two color space," *Signal Process. Image Commun.*, vol. 96, pp. 116250, 2021.

- [36] Ziyin Ma and Changjae Oh, "A wavelet-based dual-stream network for underwater image enhancement," in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2022, Virtual and Singapore, 23-27 May 2022*. 2022, pp. 2769–2773, IEEE.
- [37] Saeed Anwar, Chongyi Li, and Fatih Porikli, "Deep underwater image enhancement," *CoRR*, vol. abs/1807.03528, 2018.