Check for updates

Learning degradation-aware visual prompt for maritime image restoration under adverse weather conditions

Xin He[1]*, Tong Jia[2] and Junjie Li[3]

[1]School of Basic Sciences for Aviation, Naval Aviation University, Yantai, China, [2]School of Art and Design, Yantai Institute of Science and Technology, Yantai, China, [3]School of Electromechanical and Automotive Engineering, Yantai University, Yantai, China

Adverse weather conditions such as rain and haze often lead to a degradation in the quality of maritime images, which is crucial for activities like navigation, fishing, and search and rescue. Therefore, it is of great interest to develop an effective algorithm to recover high-quality maritime images under adverse weather conditions. This paper proposes a prompt-based learning method with degradation perception for maritime image restoration, which contains two key components: a restoration module and a prompting module. The former is employed for image restoration, whereas the latter encodes weather-related degradation-specific information to modulate the restoration module, enhancing the recovery process for improved results. Inspired by the recent trend of prompt learning in artificial intelligence, this paper adopts soft-prompt technology to generate learnable visual prompt parameters for better perceiving the degradation-conditioned cues. Extensive experimental results on several benchmarks show that our approach achieves superior restoration performance in maritime image dehazing and deraining tasks.

KEYWORDS

maritime image, image restoration, image deraining, image dehazing, prompt learning, deep learning, visual transformer

1 Introduction

Adverse weather conditions, including rain and haze, frequently occur in our everyday environment. These conditions result in diminished visual quality in captured images and significantly affect the effectiveness of numerous maritime vision systems, such as autonomous ships for ocean observation (Zheng et al., 2024). In maritime navigation and transportation, correctly identifying and interpreting environmental information from images is vital for safety. Figure 1 shows the physical imaging process of different adverse weather conditions. Thus, image processing under adverse weather conditions contributes

to enhancing the safety of maritime traffic and navigation by reducing accidents and collisions (Lu et al., 2021).

To solve image restoration under adverse weather conditions, early algorithms are predominantly based on traditional prior models. In the context of image dehazing, one common approach was the atmospheric scattering model (He et al., 2010), which assumed that haze in an image could be represented as a result of light scattering due to atmospheric particles (Li et al., 2018a). These algorithms typically aim to estimate and remove the haze from images, enhancing visibility. On the other hand, for image deraining, a prevalent technique was the linear superposition model. This model assumed that the observed image under rainy conditions could be expressed as a linear combination of the clean background scene and the rain streaks (Chen et al., 2023b). These early deraining algorithms focus on separating the rain streaks from the desired scene, thus improving the clarity of the image. However, these early prior-based algorithms struggled to adapt to complex and rapidly changing scenes, as they relied heavily on predefined models that could not effectively account for the wide range of scenarios encountered in real-world environments.

With the rise of big data and artificial intelligence, a plethora of image restoration methods based on deep learning have emerged. These techniques aim to learn the mapping relationship between degraded images and their corresponding clear counterparts. Convolutional Neural Networks (CNNs) have emerged as a powerful tool for image restoration due to their inherent ability to capture and learn complex hierarchical features from data. We have witnessed the rapid advancement of CNNs in image dehazing and deraining (Li et al., 2020; Zhou et al., 2021; Chen et al., 2022). However, due to the inherent characteristics of convolution operations, specifically the use of local receptive fields and the independence of input content, CNNs struggle to effectively model spatially-long feature dependencies of images (Chen et al., 2023c).

Later, Transformer-based models (Vaswani et al., 2017) originally bring significant breakthroughs to the natural language processing (NLP) field. The vision Transformer (ViT), as a new network backbone, has been widely applied to various tasks. It has also been utilized in image restoration tasks Zamir et al. (2022) and has achieved better performance compared to CNNs due to its ability to model non-local features effectively. Albeit these approaches have achieved commendable restoration performance in the given weather situation, they often exhibit suboptimal results when applied to maritime images.

The reason behind this can be summarized as follows: (1) Maritime images are often captured over large bodies of water, which introduces additional complexities due to the presence of reflective surfaces, varying water conditions, and dynamic backgrounds. These factors can exacerbate the impact of weather-related degradation. (2) Weather conditions at sea can change rapidly, with haze and rain appearing and dissipating quickly. This dynamic nature poses challenges for image restoration, as algorithms must adapt to evolving weather conditions. Thus, effective image restoration techniques tailored to the unique characteristics of maritime environments are essential for ensuring safe and efficient maritime activities (Zheng et al., 2020).

This raises a question: how to better help image restoration in adapting to the complex and ever-changing maritime scenes under adverse weather conditions? Recent trends of prompt learning (Wang et al., 2022) in artificial intelligence, may offer a potential solution. Prompt learning empowers deep models to adapt swiftly to complex and dynamically changing environments. It allows for the creation of tailored prompts that can capture the intricacies of specific situations, ensuring the model's responsiveness to various challenges. Therefore, this motivates us to introduce prompt learning to better encode degradation features of different weather conditions. This paper proposes a prompt-based learning method with degradation perception for maritime image restoration. The proposed method comprises two essential components: a restoration module and a prompting module. The restoration module is employed for image restoration, while the prompting module encodes weather-related degradation-specific information to modulate the restoration module. Specifically, the main contributions of this paper are as follows:

- This paper presents a new solution for image restoration in adverse weather conditions for maritime images. By incorporating prompt learning into the Transformer-based restoration network, it enhance the adaptability of deep networks to various weather degradation characteristics, enabling our model to adaptively learn more useful features to facilitate better restoration.

- This paper employs a prompt creation block to generate a set of learnable parameters by implicitly predicting degradation-conditioned soft prompts. In addition, this paper further introduces a prompt fusion block to guide the restoration process by interacting with the network backbone.

- Quantitative and qualitative experiments demonstrate that our proposed method achieves favorable performance on multiple benchmark datasets, and can better reconstruct clear images and restore image details compared to previous methods.

2 Related work

In this section, this paper presents a review of recent work related to maritime image restoration and prompt learning.

2.1 Maritime image restoration

To deal with the uncertain task of maritime image restoration, considerable efforts have been made. Existing approaches can be categorized into strategies based on priors and learning-based strategies. Hu et al (Hu et al., 2019). proposed a haze removal method based on illumination decomposition. This method decomposes the hazy image into a haze layer by separating the glow layer. It estimates the transmission rate using haze-line prior, thereby restoring the haze-free image. Luo et al (Lu et al., 2021).

proposed a novel CNN-based visibility dehazing framework aimed at enhancing the visual quality of images captured by maritime cameras under hazy conditions. This framework comprises two subnetworks: the coarse feature extraction module and the fine feature fusion module. Hu et al. (Hu et al., 2021) proposed a deep learning-based variational optimization method for reconstructing haze-free images from observed hazy images. This method fully leverages a unified denoising framework and strong deep learning representation capabilities. Guo et al. (Guo et al., 2021) designed a heterogeneous twin birth haze removal network, HTDNet, to enhance maritime surveillance capabilities in haze environments. The network consists of a twin feature extraction module for learning coarse haze features and a feature fusion module for integration and enhancement.

Van et al. (Van Nguyen et al., 2021) proposed a haze removal algorithm for maritime environment images based on texture and structure priors in illumination decomposition. This method utilizes a haze removal algorithm to eliminate the haze component from the glow-free layer and employs illumination compensation to restore natural illumination in the glow layer. Yang et al. (Yang et al., 2022) proposed a multi-head pyramid large kernel encoder-decoder network (LKEDN-MHP) for denoising tasks in maritime images. This method utilizes the transmission map extracted from the guidance image as an additional input to improve the network performance. Liu et al. (Liu et al., 2022) proposed a CNN-based dual-channel two-stage image dehazing network, which utilizes an attention mechanism to achieve adaptive fusion of multi-channel features. Hu et al. (Hu et al., 2022) proposed a maritime video dehazing algorithm based on spatiotemporal information fusion and improved dark channel prior. This method utilizes an enhanced dark channel prior model to restore each frame image, thereby achieving video dehazing.

Recently, Huang et al. (Huang et al., 2023) proposed an improved convex optimization model based on an atmospheric scattering model to achieve image dehazing. This method integrates simplified atmospheric light value estimation and the V channel in the HSV color space to obtain more local information. He et al. (He and Ji, 2023) improved MID-GAN is capable of training with non-paired adversarial learning. It consists of a CycleGAN cycle framework with two constraint branches. And it introduced an effective attentionrecursive feature extraction module to gradually extract haze components in an unsupervised manner. Chen et al. (Chen et al., 2022) introduced a contrastive learning mechanism based on the CycleGAN framework to improve dehazing performance. However, due to the limited performance of the aforementioned methods in the task of maritime image restoration and the relative saturation of model capabilities, there is a need to explore an effective approach to address these issues.

2.2 Prompt learning

Prompt learning was initially introduced in the field of natural language processing (NLP) and has proven to be highly effective, it

has been applied to various vision-related tasks. Prompt learning is divided into two different methods: hard prompts and soft prompts. Hard prompts refer to explicit and predefined instructions given to the model during training. These prompts provide specific information and guide the model to produce the desired output. Soft prompts are more flexible and adaptive. They are not explicitly defined but rather generate prompt information based on input data or learn from the training process. Soft prompts allow the model to dynamically adjust its behavior based on input and context, enabling it to capture more specific and nuanced contextual information.

Recently, Zhou et al. (Zhou et al., 2022) demonstrated that a simple design based on conditional prompt learning performs exceptionally well in various problem scenarios, including generalization from base classes to novel classes, cross-dataset prompt transfer, and domain generalization. Potlapalli et al. (Potlapalli et al., 2023) demonstrated the effectiveness of their designed prompt block in integrated image restoration by integrating it into state-of-the-art restoration models. The prompt block can interact with input features, dynamically adjust representations, and adapt the restoration process to the relevant degradation. Li et al. (Li et al., 2023b) proposed a novel prompt-in-prompt learning for universal image restoration. The method involves simultaneous learning of high-level degradation-aware prompts and low-level basic restoration prompts to generate effective universal restoration prompts. By utilizing a selective promptfeature interaction module to modulate features most relevant to the degradation. Ai et al. (Ai et al., 2023) proposed a multi-modal prompt learning method called MPerceiver, which includes cross-modal adapters and image restoration adapters to learn holistic and multiscale detail representations. The adaptability of text and visual prompts is dynamically adjusted based on degradation prediction, enabling effective adaptation to various unknown degradations. Kong et al. (Kong et al., 2024) proposed sequential learning strategy and prompt learning strategy, respectively. These two strategies are effective for both CNN and Transformer backbones, and they can complement each other to learn effective image representations.

Inspired by these methods, this paper proposes a prompt-based learning approach to guide the maritime image restoration process, facilitating the integration and communication of information.

3 Proposed method

In this section, this paper first describes the overall pipeline of the model. Then, this paper provides details of the restoration module and prompting module, which serve as the fundamental building blocks of the approach. The restoration module mainly comprises two key elements: multi-head self-attention (MHSA) and dual gated feed-forward network (DGFN). The prompting module mainly consists of two key elements: prompt creation block (PCB) and prompt fusion block (PFB).

3.1 Overall pipeline

The overall pipeline of the proposed model, as illustrated in the Figure 2, is based on a hierarchical encoder-decoder framework (Chen et al., 2023a). Given a maritime degraded image $I_{rain} \in \mathbb{R}^{H \times W \times 3}$, where $H \times W$ denotes the spatial resolution of the feature maps, and C represents the channels, this paper performs feature projection embedding using a 3×3 convolution. On the network backbone, this paper stacks 4 levels of hierarchical encoder-decoders, where the encoder-decoder serves as the restoration module of the model, extracting rich spatially variant degradation distribution features. To extract multiscale representations from degradation information, each level of the restoration module covers its specific spatial resolution and channel dimensions. Beginning with high-resolution input, the restoration module aims to progressively decrease spatial resolution while enhancing channel capacity, resulting in a low-resolution latent representation $F \in \mathbb{R}^{H/8 \times W/8 \times 8C}$. During the stage of high-resolution image restoration, this paper incorporates a prompting module into the framework to generate prompts and enrich input features for dynamically guiding the restoration process of the restoration module. This paper also introduces skip connections (Li et al., 2023a) to bridge consecutive intermediate features, ensuring stable training. Next, this paper provides a detailed description of the proposed restoration module, prompting module, and their core building blocks.

3.2 Restoration module

This paper develops a restoration module as a feature extraction unit, which can be used to encode degradation information to recover output clean restored images. Formally, given the input features of the $(l-1)$-th block X_{l-1}, the encoding of the restoration module process can be represented as Equations (1, 2):

$$X_l' = X_{l-1} + MHSA(LN(X_{l-1})), \tag{1}$$

$$X_l = X_l' + DGFN(LN(X_l')), \tag{2}$$

where LN denotes the layer normalization, X_l' and X_l represent the outputs of MHSA and DGFN, which are described below.

3.2.1 Multi-head self-attention

In reviewing the standard self-attention mechanism in Transformers (Zamir et al., 2022), given queries Q, keys K, and values V, the output of dot-product attention is typically represented as Equation (3):
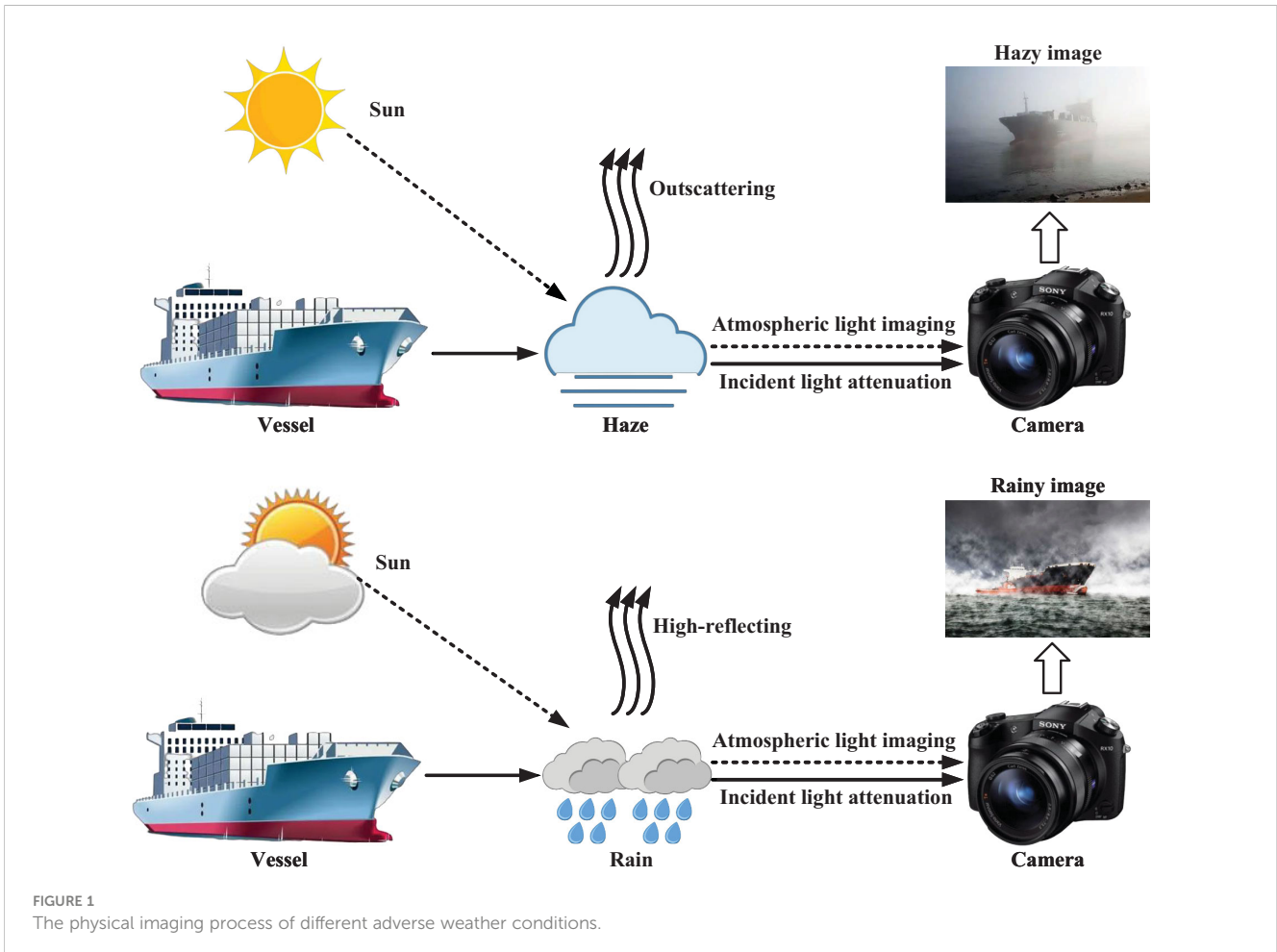

FIGURE 1
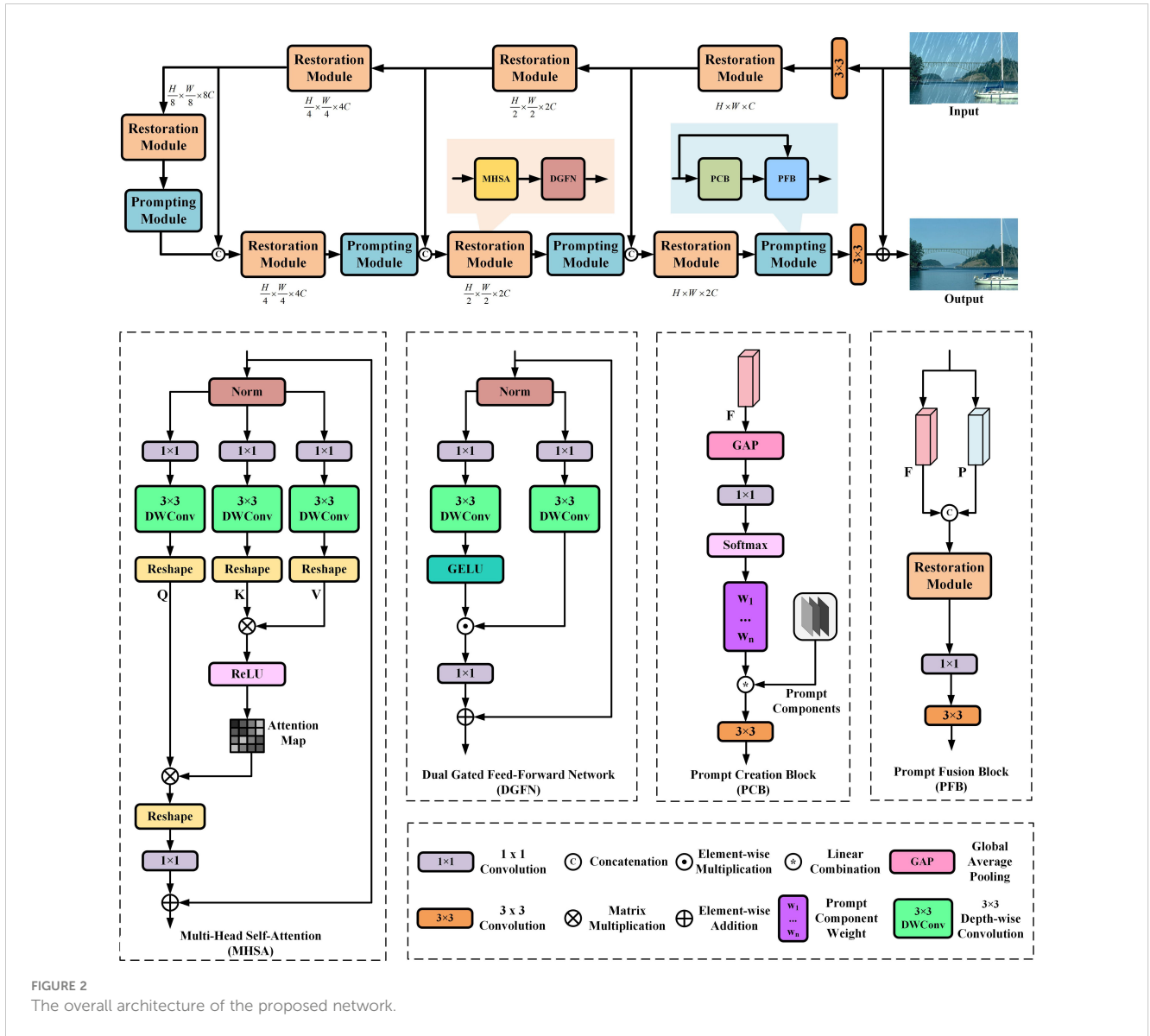The physical imaging process of different adverse weather conditions.

FIGURE 2
The overall architecture of the proposed network.

$$Attention' = Softmax\left(\frac{Q^{\top}K}{\alpha}\right) \cdot V, \qquad (3)$$

where α is a learnable parameter; Q, K, and V represent the matrix forms of Q, K, and V, respectively. It is noted that computing self-attention using the Softmax function may lead to unstable gradients due to the presence of exponential functions, which could also limit the network's ability for nonlinear fitting. This work replaces it with the ReLU activation function, which can alleviate the issues of gradient vanishing or exploding, and aid in learning better feature representations. Specifically, this work starts by aggregating pixel-level cross-channel context through the application of a 1×1 convolution. Subsequently, a 3×3 depthwise convolution is applied to encode channel-wise context. This work employs bias-free convolution layers in the network. Next, it reshapes the projections of queries and keys to allow their dot product interaction to generate a transposed attention map of size $\mathbb{R}^{\hat{C} \times \hat{C}}$, rather than a massive regular attention map of size

$\mathbb{R}^{\hat{H}\hat{W} \times \hat{H}\hat{W}}$. Then, the attention map is further interacted with the reshaped projections of values to complete the self-attention computation. Overall, the MHSA process is defined as Equations (4, 5):

$$Attention = ReLU\left(\frac{Q^{\top}K}{\alpha}\right) \cdot V, \qquad (4)$$

$$\hat{X} = \underset{1\times1}{\mathrm{Conv}}(Attention(Q, K, V)) + X, \qquad (5)$$

where X and $\hat{X}$ are the input and output feature maps, $Conv_{1\times1}$ $(\cdot)$ denotes 1×1 convolution.

3.2.2 Dual gated feed-forward network

To enhance the enrichment of contextual information, this paper introduces a dual gated feed-forward network that operates on each pixel. It incorporates two branches based on a gating mechanism. They initially undergo feature Y transformation by using 1×1 convolutions, followed by 3×3 depth-wise convolutions

to encode information from spatially adjacent pixel positions, facilitating the learning of local image details for effective restoration. One branch extends the feature channels, while the other branch, activated with GELU non-linearity, reduces the channels back to the original input dimensionality, enabling the discovery of non-linear contextual information in hidden layers. The DGFN is formulated as Equations (6, 7):

$$Gated = GELU(\text{Conv}_{3\times3}(\text{Conv}_{1\times1}(X))) \odot \text{Conv}_{3\times3}(\text{Conv}_{1\times1}(Y)), \tag{6}$$

$$\hat{Y} = \text{Conv}_{1\times1}(Gated(Y)) + Y, \tag{7}$$

where $\hat{Y}$, $Conv_{3\times3}(\cdot)$, $\odot$ denote outputs, 3×3 depthwise convolution and element-wise multiplication, respectively. Overall, compared to MHSA, DGFN plays a distinctly different role by governing the flow of information across various levels in our pipeline, thereby enabling each level to focus on fine details complementary to other levels.

3.3 Prompting module

Different from (Zhou et al., 2023) that models global features, this paper further proposes a prompting module designed to perceive features of interfering information in degraded images and dynamically generate valuable prompts to guide high-quality maritime image restoration. Given the input features F, the prompting module first employs PCB to generate prompts for the distribution of degraded features. Subsequently, PFB collaboratively fuses the input features F with the generated prompt information to obtain the output features F_b. The overall procedure of prompting module is defined as Equation (8):

$$\hat{F} = PFB(PCB(F, F_{pc}), F), \tag{8}$$

where F_{pc} represents the learnable prompt components, the prompt creation block and prompt fusion block are described below.

3.3.1 Prompt creation block

The PCB dynamically captures the distribution characteristics in degraded maritime images. This capability enhances the provision of useful prompt information for the restoration process. Here, this paper employs a soft prompt (Potlapalli et al., 2023) to generate a set of learnable parameters, which encode distinctive deteriorative features related to various weather conditions. For input features F, the network first applies global average pooling, followed by a 1×1 convolution to obtain a compact feature vector. Subsequently, a Softmax function is applied to derive the prompt weights $W \in \mathbb{R}_N$, where the value of N is determined by the prompt length. In soft prompt learning, the length of the prompt determines the amount of information and guidance provided to the model. In fact, it's essential to strike a balance between the impact of prompt length, choosing an appropriate length that balances information, guidance, and diversity of generation to achieve satisfactory results. This paper

will analyzes its impact in Section 4.4.2. Next, the weights are adjusted within the prompting components using a linear combination. Finally, a 3×3 convolution is applied to obtain the conditional prompt P. The overall procedure of PCB is defined as Equations (9, 10):

$$W_i = Softmax(\text{Conv}_{1\times1}(GAP(F))), \tag{9}$$

$$P = \text{Conv}_{3\times3}(Linear(W_i \cdot F_{pc})), \tag{10}$$

where $GAP(\cdot)$, $Linear(\cdot)$ denote the global average pooling and linear combination, respectively.

3.3.2 Prompt fusion block

To facilitate more effective interaction between the prompt information and input features for better guidance in the restoration process, this paper designs a PFB. In this module, the input features F are concatenated with the degradation distribution prompt P along the channel dimension. Subsequently, the concatenated information is further processed by the restoration module to generate the transformation of the input features. Finally, through operations of 1×1 convolution and 3×3 convolution, the features are smoothed and mapped to the output features $\hat{F}$. The overall procedure of PFB is defined as Equation (11):

$$\hat{F} = \text{Conv}_{3\times3}(\text{Conv}_{1\times1}(\mathcal{RM}(\mathcal{C}(F \cdot P)))), \tag{11}$$

where $\mathcal{RM}(\cdot)$ and $\mathcal{C}(\cdot)$ denote the process of involving restoration module and concatenation, respectively.

3.4 Loss function

To supervise the training progress of our network, this paper employs the L_1 pixel loss function. The final output is the restored image I_{rec}. It is obtained by adding the residual image I_{res} to the input degraded image I_{deg}, where $I_{res} \in \mathbb{R}^{H \times W \times 3}$. During training, the network minimizes the loss function ,which is defined as Equation (12):

$$L_{pixel} = \left\| I_{rec} - I_{gt} \right\|_1, \tag{12}$$

where I_{gt} represents the ground truth image, and $\| \cdot \|_1$ denotes the L_1-norm.

4 Experiments

In this section, this paper evaluates our method on benchmarks for image dehazing and image deraining tasks. The main experiments are conducted using PyTorch and trained on 4 TESLA V100 GPUs.

4.1 Experimental settings

Following (Zamir et al., 2022), our method employs a 4-level encoder-decoder framework. From level-1 to level-4, the number of

restoration module is [4,6,6,8], attention heads are [1,2,4,8], and number of channels is [48,96,192,384]. The prompt length in the prompting module is 5. The batch size and patch size are configured as 16 and 128, respectively. Our model is trained using the AdamW optimizer for a total of 300,000 iterations with cosine annealing scheme (Loshchilov and Hutter, 2016) to gradually decrease the initial learning rate from 3×10^{-4} to 1×10^{-6}. Specifically, the learning rate is initially set at 3×10^{-4} for the 92,000 iterations and subsequently reduced to 1×10^{-6} over the next 208,000 iterations.

4.2 Experimental results on image dehazing

To evaluate the dehazing performance of the method, this paper trains on the RESIDE SOTS-Outdoor (Li et al., 2018a), but since this dataset lacks maritime scenes, this paper conducts testing on real maritime hazy images. It should be noted that there are no ground truth images for real maritime hazy images. Here, this paper compares our method with 6 popular image dehazing algorithms, including DCP (He et al., 2010), DehazeNet (Cai et al., 2016), AODNet (Li et al., 2017), GridDehazeNet Liu et al. (2019), MSBDN (Dong et al., 2020), and DeHamer (Guo et al., 2022). For fair comparison, all comparison algorithms use consistent pre-trained weights trained on the training set. In the case of hazy maritime images without ground truth data, this paper employs non-reference metrics such as NIQE (Naturalness Image Quality Evaluator) (Mittal et al., 2012b) and BRISQUE (Blind/Referenceless Image Spatial Quality Evaluator) (Mittal et al., 2012a). When the NIQE or BRISQUE scores are lower, it means that the image is considered to have higher quality in terms of naturalness (for NIQE) or overall spatial quality (for BRISQUE).

As presented in Table 1, our proposed method achieves notably lower NIQE and BRISQUE scores, indicating that it produces high-quality outputs characterized by clearer content and superior perceptual quality when compared to other models in maritime scenarios. This underscores the effectiveness of our model in enhancing image quality within maritime contexts. To provide compelling evidence, this paper illustrates a visual quality comparison between two samples generated by recent methods in Figure 3. It is observed that the performance of DCP (He et al., 2010) is suboptimal, particularly in the sky region, where undesirable halo effects occur. DehazeNet (Cai et al., 2016) and AODNet (Li et al., 2017), are found to exhibit insufficient learning capabilities, resulting in their inability to effectively remove haze from images. The limitations in their learning capabilities become particularly evident when dealing with challenging and complex hazy scenes, such as those with dense fog or severe haze. The

dehazing results obtained from GridDehazeNet Liu et al. (2019), MSBDN (Dong et al., 2020), and DeHamer (Guo et al., 2022) still exhibit residual haze, indicating limitations in the generalization of their algorithms to maritime images. These residual haziness issues suggest that their models may struggle to effectively adapt to the unique challenges posed by maritime scenarios. In contrast, our method demonstrates the capability to recover significantly clearer images, particularly in the sailboat regions. This suggests that the introduction of prompt learning cues can substantially improve the adaptation of dehazing algorithms to more challenging maritime images. The enhanced performance underscores the effectiveness of incorporating prompt learning, which enables better haze removal and results in visually superior outcomes in maritime scenes.

4.3 Experimental results on image deraining

To evaluate the deraining performance of the method, this paper carries out comprehensive experiments using the Rain13K dataset (Jiang et al., 2020), comprising 13,700 pairs of clean and rainy images. To evaluate our approach, this paper employs 4 benchmarks [Test100 (Zhang et al., 2019), Rain100H (Yang et al., 2017), Rain100L (Yang et al., 2017), and Test2800 (Fu et al., 2017b)] for testing purposes. Here, this paper compares our method with 9 popular image deraining algorithms, including DerainNet (Fu et al., 2017a), SEMI (Wei et al., 2019), DIDMDN (Zhang and Patel, 2018), UMRL (Yasarla and Patel, 2019), RESCAN (Li et al., 2018b), PReNet (Ren et al., 2019), MSPFN (Jiang et al., 2020), MPRNet (Zamir et al., 2021), and IDT (Xiao et al., 2022). Here, this paper uses full-reference image evaluation metrics, as there are ground truth images available. This paper quantitatively assesses our results by computing both PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index) scores (Wang et al., 2004), specifically using the Y channel within the YCbCr color space. This allows us to make objective comparisons, focusing on luminance information, which is crucial for evaluating image quality accurately. These metrics provide valuable insights into the fidelity and structural similarity between the deraining images and their corresponding ground truth references.

Table 2 presents the quantitative results of various algorithms for image deraining. In comparison to the recent IDT method (Xiao et al., 2022), our approach demonstrates an average improvement of 1dB across the four datasets. This significant enhancement highlights our method's ability to adapt more effectively to diverse rainy conditions. Our results suggest that our approach outperforms existing methods by providing superior deraining performance across a range of challenging rain scenarios.

TABLE 1 Quantitative comparisons of different methods on hazy maritime images.

Method	DCP	DehazeNet	AODNet	GridDehazeNet	FFA Net	DeHamer	Ours
NIQE	3.464	3.466	3.433	3.504	3.530	3.483	3.015
BRISQUE	22.149	25.884	25.006	24.113	24.847	25.054	22.028

FIGURE 3
Image dehazing comparisons for different methods on hazy maritime images.

Figures 4 and 5 show the visual results on the Rain100H and Test100 datasets. Observing these visual results, it becomes evident that DerainNet (Fu et al., 2017a), SEMI (Wei et al., 2019), and DIDMDN (Zhang and Patel, 2018) struggle to effectively remove heavy rain artifacts. UMRL (Yasarla and Patel, 2019), RESCAN (Li et al., 2018b), and PReNet (Ren et al., 2019) still leave residual rain streaks in their recovery results. MSPFN (Jiang et al., 2020), MPRNet (Zamir et al., 2021), and IDT (Xiao et al., 2022) exhibit

shortcomings in preserving local image details, especially in regions such as the ship's hull. In contrast, our approach stands out due to its incorporation of prompt learning, enabling adaptive feature extraction. As a result, it excels in eliminating rain streaks while effectively retaining intricate image structures. This demonstrates the robustness and effectiveness of our method in addressing the challenges posed by rainy conditions and preserving fine-grained image details.

TABLE 2 Quantitative comparisons of different methods on the Rain13K dataset.

| Datasets Method	Test100		Rain100H		Rain100L		Test2800		Average	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DerainNet	22.77	0.810	14.92	0.592	27.03	0.884	24.31	0.861	22.26	0.787
SEMI	22.35	0.788	16.56	0.486	25.03	0.842	24.43	0.782	22.09	0.725
DIDMDN	22.56	0.818	17.35	0.524	25.23	0.741	28.13	0.867	23.32	0.738
UMRL	24.41	0.829	26.01	0.832	29.18	0.923	29.97	0.905	27.39	0.872
RESCAN	25.00	0.835	26.36	0.786	29.80	0.881	31.29	0.904	28.11	0.852
PReNet	24.81	0.851	26.77	0.858	32.44	0.950	31.75	0.916	28.94	0.894
MSPFN	27.50	0.876	28.66	0.860	32.40	0.933	32.82	0.930	30.35	0.900
MPRNet	30.27	0.897	30.41	0.890	36.40	0.965	33.64	0.938	32.68	0.923
IDT	29.69	0.905	29.95	0.898	37.01	0.971	33.38	0.937	32.51	0.928
Ours	31.08	0.907	30.85	0.900	38.30	0.974	34.02	0.941	**33.56**	**0.930**

Bold indicates the best results.
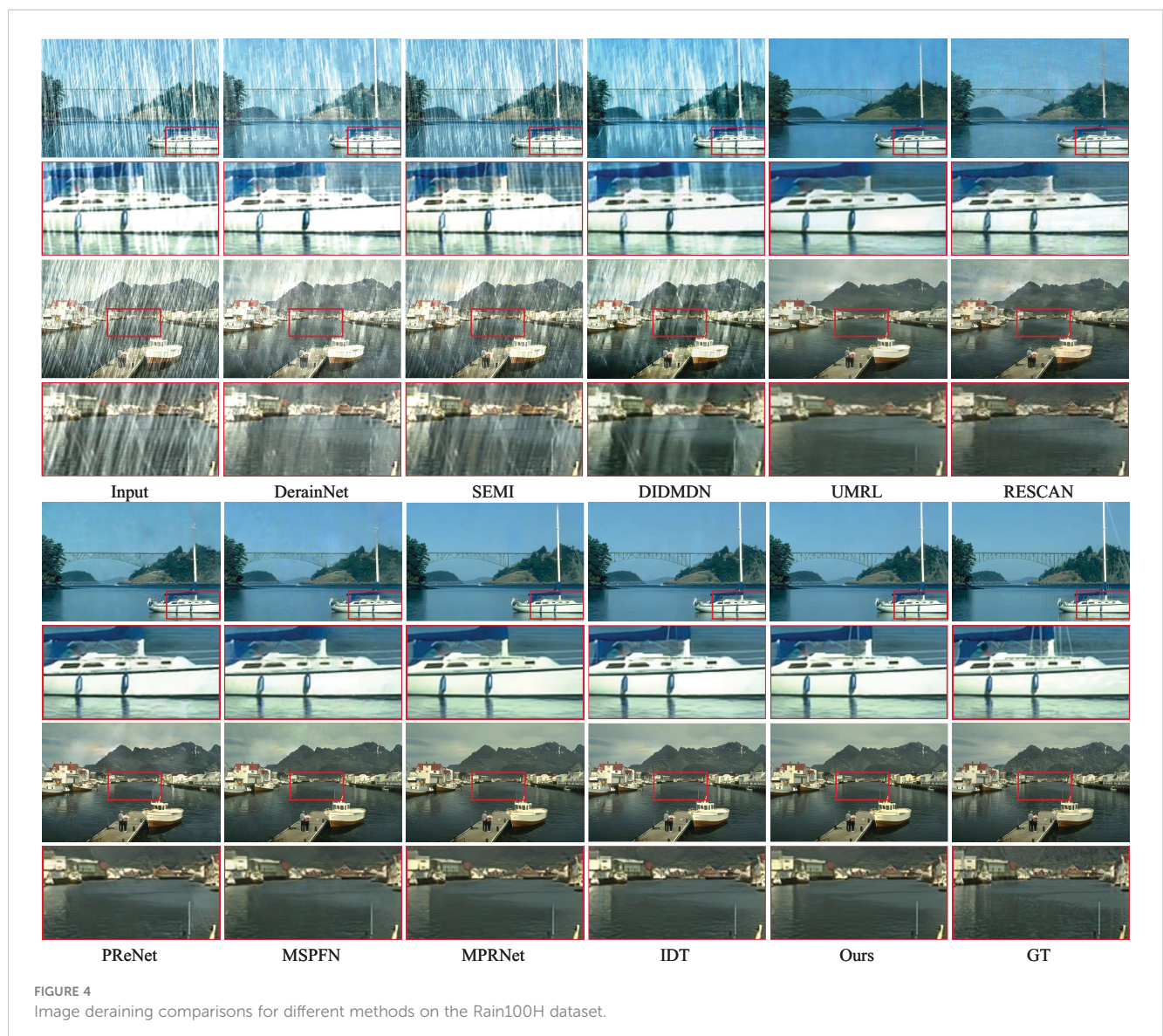
FIGURE 4
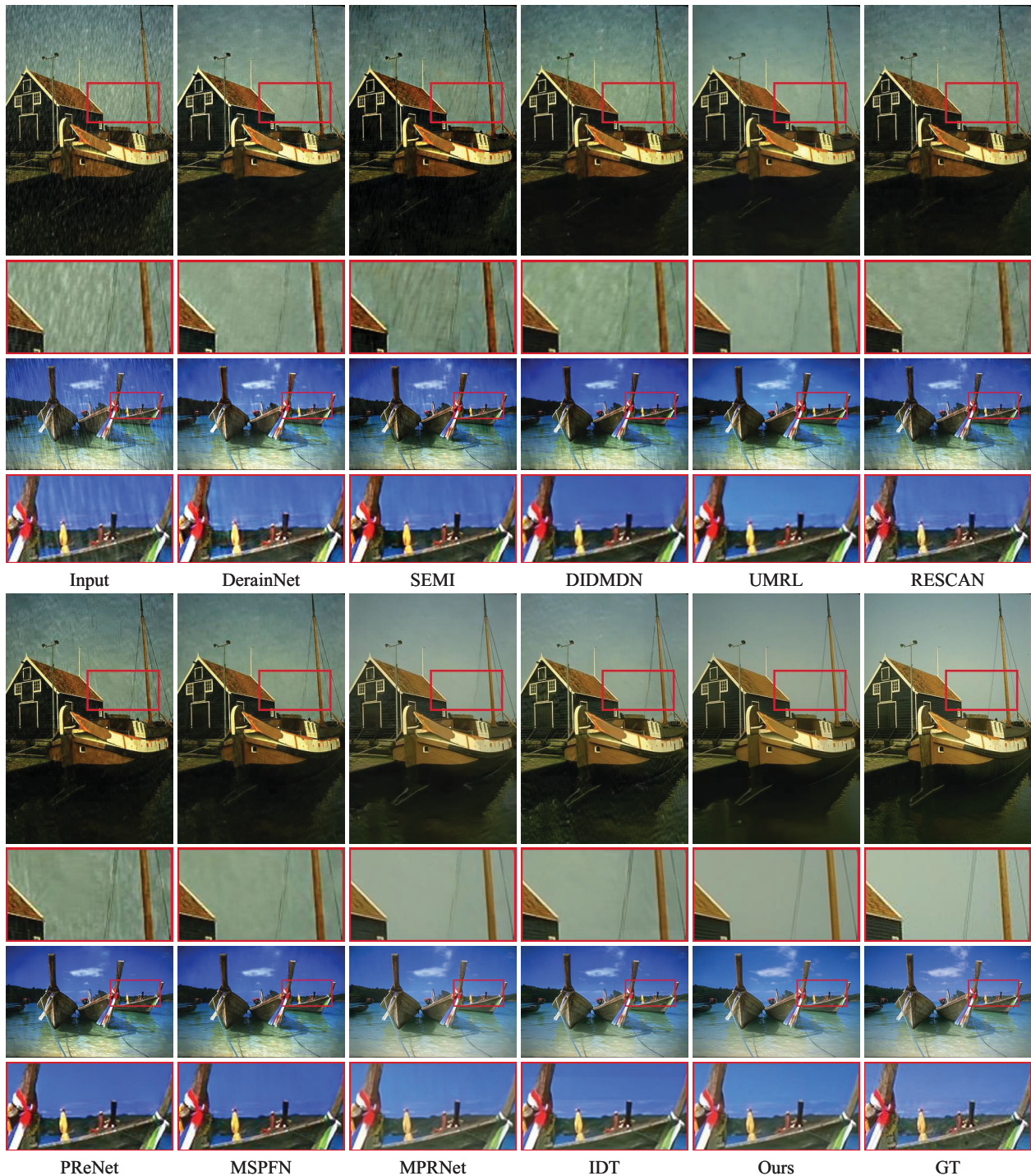Image deraining comparisons for different methods on the Rain100H dataset.

FIGURE 5
Image deraining comparisons for different methods on the Test100 dataset.

4.4 Ablation experiment and analysis

To further analysis the effectiveness of our proposed method, this paper conducts additional ablation experiments. This paper focuses on evaluating the performance of our approach in the image deraining task, specifically examining the effectiveness of prompting modules and the effect of prompt length in prompting modules.

4.4.1 Effectiveness of prompting modules

In this experiment, this paper aims to gauge the contributions of the prompting modules within our methodology. This paper designs three distinct model variants to comprehensively analyze the impact of prompting modules on our approach. These variants encompass models with or without prompting modules, models with or without PFB, and models with prompting modules placed at

TABLE 3 Ablation experimental result on the effectiveness of prompting modules.

| Model | Prompting module | | | Resto-ration module | PSNR/SSIM |
	PCB	PFB	Position		
(a)	×	×	–	✓	33.15/0.924
(b)	✓	×	Enc.	✓	33.39/0.927
(c)	✓	✓	Enc.	✓	33.48/0.929
(d)	✓	✓	Dec.	✓	**33.56/0.930**

Bold indicates the best results.

TABLE 4 Ablation experimental result on the effect of prompt length in prompting modules.

Length	3	4	5	6	7
PSNR/SSIM	33.41/0.928	33.49/0.929	**33.56/0.930**	33.52/0.930	33.44/0.927

Bold indicates the best results.

different positions within the network architecture. The quantitative results for these various model variants are documented in Table 3, shedding light on their respective contributions to the image restoration task.

Starting with Model (a), the performance starkly deteriorates in the absence of prompting modules. This striking decline underscores the pivotal role that prompting modules play in the image restoration process. They serve as critical components in guiding the model's understanding of the task and assisting it in achieving high-quality restoration results. Comparing Model (b) and Model (c), it becomes evident that the absence of PFB in the

prompting modules results in a disconnect between the guidance provided by the prompts and the features processed in the main restoration modules. This lack of synchronization hinders the model's ability to effectively utilize the provided prompts, consequently restricting its overall performance. Further comparison between Model (c) and Model (d) reveals the advantage of introducing prompting modules during the network decoding phase. This strategic placement allows the model to better harness the clear image features, facilitating improved performance and yielding the best network performance among the variants.

4.4.2 Effect of prompt length in prompting modules

To delve deeper into the impact of prompts in our prompting modules, this paper conducts experiments to investigate the effect of prompt length. This paper varies the length of prompts while keeping other parameters constant and examined how it influenced the model's performance. Table 4 presents the quantitative results. Our findings reveal that prompt length plays a crucial role in image restoration process. Specifically, shorter prompts tend to yield faster convergence and better overall results, while longer prompts sometimes lead to overfitting or increased computational complexity. This analysis allows us to optimize the prompt length within our prompting modules to achieve the best balance between performance and efficiency. Finally, this paper determines that a prompt length of 5 is the most suitable configuration for our model.

4.5 User study

This paper conducts a user study to assess the outcomes of various image restoration techniques. These studies are based on restoring real foggy maritime images. A portion of the participants invited to our user study are professionals engaged in maritime
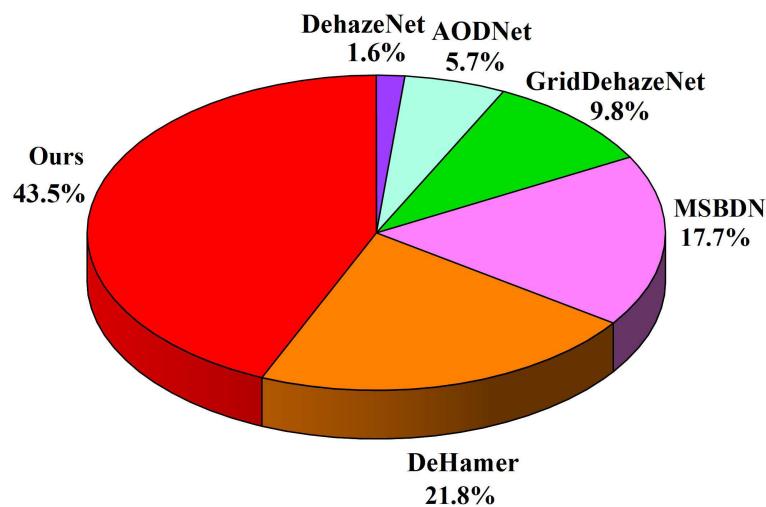

FIGURE 6
Averaged selection percentage of user study.

TABLE 5 Comparison of model efficiency.

	Uformer	Restormer	IDT	DRSformer	Ours
Paramaters (M)	50.8	26.1	16.4	33.7	35.6
Flops (G)	45.9	174.7	61.9	242.9	43.2

activities to ensure the professionalism of our work. Additionally, in order to enhance the universality of our work, this paper has also invited non-maritime professionals to participate in our user study. Participants are presented with a set of images and asked to select the one with the best visual clarity. To maintain fairness, the methods remain anonymous, and the images within each set are randomly ordered. This paper distributes the questionnaire widely to online users and collect responses. Finally, this paper receives responses from 46 human evaluators. Figure 6 depicts the average selection percentage for each method. Based on the majority of human evaluators' feedback, our method consistently outperforms the others.

4.6 Limitations

While our proposed method introduces prompt learning to enhance restoration performance under adverse weather conditions, it comes with relatively high parameter count and computational complexity. The comparison results are presented in the Table 5, compared to other methods, our approach has a higher parameter count. As a result, significant computational resources are also required, which to some extent limits the applicability of the model. In order to enable our method to be more rapidly and conveniently applied in maritime operations, this paper plans to address this issue through strategies such as model compression and pruning, aiming to make the model more suitable for maritime vision applications.

5 Conclusions

This paper has proposed a prompting image restoration approach by learning degradation-aware visual prompt for maritime surveillance. Our proposed approach possesses the capability to interact with input features, allowing for dynamic adjustments in weather-related representations. This adaptability ensures that the restoration process is tailored to the specific degradation being addressed. This paper has validated the effectiveness of our method on extensive experimental datasets, enhancing its restoration performance in various weather conditions, including rain removal and haze removal in maritime images. In future work, this paper plans to explore leveraging text

models such as CLIP as alternative prompts to further guide the image restoration process.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://sites.google.com/view/reside-dehaze-datasets, https://www.deraining.tech/benchmark.html.

Author contributions

XH: Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. TJ: Data curation, Software, Visualization, Writing – original draft. JL: Investigation, Writing – review & editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ai, Y., Huang, H., Zhou, X., Wang, J., and He, R. (2023). Multimodal prompt perceiver: Empower adaptiveness, generalizability and fidelity for all-in-one image restoration. arXiv [preprint] arXiv:2312.02918.

Cai, B., Xu, X., Jia, K., Qing, C., and Tao, D. (2016). Dehazenet: An end-to-end system for single image haze removal. *IEEE Trans. image Process.* 25, 5187–5198. doi: 10.1109/TIP.2016.2598681

Chen, X., Li, H., Li, M., and Pan, J. (2023a). "Learning a sparse transformer network for effective image deraining," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (Canada: IEEE), 5896–5905. doi: 10.1109/CVPR52729.2023.00571

Chen, X., Pan, J., Dong, J., and Tang, J. (2023b). Towards unified deep image deraining: A survey and a new benchmark. arXiv [preprint] arXiv:2310.03535.

Chen, X., Pan, J., Jiang, K., Li, Y., Huang, Y., Kong, C., et al. (2022). "Unpaired deep image deraining using dual contrastive learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 2017–2026.

Chen, X., Pan, J., Lu, J., Fan, Z., and Li, H. (2023c). "Hybrid cnn-transformer feature fusion for single image deraining," in *Proceedings of the AAAI Conference on Artificial Intelligence*. (USA: AAAI Press), Vol. 37. 378–386. doi: 10.1609/aaai.v37i1.25111

Dong, H., Pan, J., Xiang, L., Hu, Z., Zhang, X., Wang, F., et al. (2020). "Multi-scale boosted dehazing network with dense feature fusion," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 2157–2167.

Fu, X., Huang, J., Ding, X., Liao, Y., and Paisley, J. (2017a). Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Trans. Image Process.* 26, 2944–2956. doi: 10.1109/TIP.2017.2691802

Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., and Paisley, J. (2017b). "Removing rain from single images via a deep detail network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (USA: IEEE), 3855–3863. doi: 10.1109/CVPR.2017.186

Guo, C.-L., Yan, Q., Anwar, S., Cong, R., Ren, W., and Li, C. (2022). "Image dehazing transformer with transmission-aware 3d position embedding," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 5812–5820.

Guo, Y., Lu, Y., Liu, R. W., Wang, L., and Zhu, F. (2021). "Heterogeneous twin dehazing network for visibility enhancement in maritime video surveillance," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. (USA: IEEE), 2875–2880.

He, K., Sun, J., and Tang, X. (2010). Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 2341–2353.

He, X., and Ji, W. (2023). Single maritime image dehazing using unpaired adversarial learning. *Signal Image Video Process.* 17, 593–600. doi: 10.1007/s11760-022-02265-5

Hu, H.-M., Guo, Q., Zheng, J., Wang, H., and Li, B. (2019). Single image defogging based on illumination decomposition for visual maritime surveillance. *IEEE Trans. Image Process.* 28, 2882–2897. doi: 10.1109/TIP.83

Hu, X., Wang, J., Zhang, C., and Tong, Y. (2021). Deep learning-enabled variational optimization method for image dehazing in maritime intelligent transportation systems. *J. Advanced Transportation* 2021, 1–18. doi: 10.1155/2021/6658763

Hu, Q., Zhang, Y., Liu, T., Liu, J., and Luo, H. (2022). Maritime video defogging based on spatiotemporal information fusion and an improved dark channel prior. *Multimedia Tools Appl.* 81, 24777–24798. doi: 10.1007/s11042-022-11921-4

Huang, H., Li, Z., Niu, M., Miah, M. S., Gao, T., and Wang, H. (2023). A sea fog image defogging method based on the improved convex optimization model. *J. Mar. Sci. Eng.* 11, 1775. doi: 10.3390/jmse11091775

Jiang, K., Wang, Z., Yi, P., Chen, C., Huang, B., Luo, Y., et al. (2020). "Multi-scale progressive fusion network for single image deraining," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 8346–8355.

Kong, X., Dong, C., and Zhang, L. (2024). Towards effective multiple-in-one image restoration: A sequential and prompt learning strategy. arXiv [preprint] arXiv:2401.03379.

Li, Z., Lei, Y., Ma, C., Zhang, J., and Shan, H. (2023b). Prompt-in-prompt learning for universal image restoration. arXiv [preprint] arXiv:2312.05038.

Li, Y., Lu, J., Chen, H., Wu, X., and Chen, X. (2023a). "Dilated convolutional transformer for highquality image deraining," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (Canada: IEEE), 4198–4206. doi: 10.1109/CVPRW59228.2023.00442

Li, B., Peng, X., Wang, Z., Xu, J., and Feng, D. (2017). "Aod-net: All-in-one dehazing network," in *Proceedings of the IEEE international conference on computer vision*. (Italy: IEEE), 4770–4778.

Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., et al. (2018a). Benchmarking single-image dehazing and beyond. *IEEE Trans. Image Process.* 28, 492–505. doi: 10.1109/TIP.83

Li, R., Tan, R. T., and Cheong, L.-F. (2020). "All in one bad weather removal using architectural search," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 3175–3185.

Li, X., Wu, J., Lin, Z., Liu, H., and Zha, H. (2018b). "Recurrent squeeze-and-excitation context aggregation net for single image deraining," in *Proceedings of the European conference on computer vision (ECCV)*. (Germany: Springer), 254–269.

Liu, X., Ma, Y., Shi, Z., and Chen, J. (2019). "Griddehazenet: Attention-based multi-scale network for image dehazing," in *Proceedings of the IEEE/CVF international conference on computer vision*. Korea (South): (IEEE), 7314–7323.

Liu, T., and Zhou, B. (2022). Dual-channel and two-stage dehazing network for promoting ship detection in visual perception system. *Math. Problems Eng.* 2022. doi: 10.1155/2022/8998743

Loshchilov, I., and Hutter, F. (2016). Sgdr: Stochastic gradient descent with warm restarts. arXiv [preprint] arXiv:1608.03983.

Lu, Y., Guo, Y., and Liang, W. (2021). Cnn-enabled visibility enhancement framework for vessel detection under haze environment. *J. advanced transportation* 2021, 1–14. doi: 10.1155/2021/7649214

Mittal, A., Moorthy, A. K., and Bovik, A. C. (2012a). No-reference image quality assessment in the spatial domain. *IEEE Trans. image Process.* 21, 4695–4708. doi: 10.1109/TIP.2012.2214050

Mittal, A., Soundararajan, R., and Bovik, A. C. (2012b). Making a "completely blind" image quality analyzer. *IEEE Signal Process. Lett.* 20, 209–212.

Potlapalli, V., Zamir, S. W., Khan, S., and Khan, F. S. (2023). Promptir: Prompting for all-in-one blind image restoration. arXiv [preprint] arXiv:2306.13090.

Ren, D., Zuo, W., Hu, Q., Zhu, P., and Meng, D. (2019). "Progressive image deraining networks: A better and simpler baseline," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 3937–3946.

Van Nguyen, T., Mai, T. T. N., and Lee, C. (2021). Single maritime image defogging based on illumination decomposition using texture and structure priors. *IEEE Access* 9, 34590–34603. doi: 10.1109/ACCESS.2021.3060439

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Adv. Neural Inf. Process. Syst.* 30.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans. image Process.* 13, 600–612. doi: 10.1109/TIP.2003.819861

Wang, Z., Zhang, Z., Lee, C.-Y., Zhang, H., Sun, R., Ren, X., et al. (2022). "Learning to prompt for continual learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 139–149.

Wei, W., Meng, D., Zhao, Q., Xu, Z., and Wu, Y. (2019). "Semi-supervised transfer learning for image rain removal," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 3877–3886.

Xiao, J., Fu, X., Liu, A., Wu, F., and Zha, Z.-J. (2022). Image de-raining transformer. *IEEE Trans. Pattern Anal. Mach. Intell.*

Yang, W., Gao, H., Jiang, Y., and Zhang, X. (2022). A novel approach to maritime image dehazing based on a large kernel encoder–decoder network with multihead pyramids. *Electronics* 11, 3351. doi: 10.3390/electronics11203351

Yang, W., Tan, R. T., Feng, J., Liu, J., Guo, Z., and Yan, S. (2017). "Deep joint rain detection and removal from a single image," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (USA: IEEE), 1357–1366.

Yasarla, R., and Patel, V. M. (2019). "Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 8405–8414.

Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., and Yang, M.-H. (2022). "Restormer: Efficient transformer for high-resolution image restoration," In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 5728–5739.

Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., Yang, M.-H., et al. (2021). "Multi-stage progressive image restoration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, virtual*. (virtual: IEEE) 14821–14831.

Zhang, H., and Patel, V. M. (2018). "Density-aware single image de-raining using a multi-stream dense network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (USA: IEEE), 695–704.

Zhang, H., Sindagi, V., and Patel, V. M. (2019). Image de-raining using a conditional generative adversarial network. *IEEE Trans. circuits Syst. video Technol.* 30, 3943–3956. doi: 10.1109/TCSVT.76

Zheng, S., Sun, J., Liu, Q., Qi, Y., and Zhang, S. (2020). "Overwater image dehazing via cycle-consistent generative adversarial network," in *Proceedings of the Asian Conference on Computer Vision*. (Japan: Springer), doi: 10.3390/electronics9111877

Zheng, K., Zhang, X., Wang, C., Li, Y., Cui, J., and Jiang, L. (2024). Adaptive collision avoidance decisions in autonomous ship encounter scenarios through rule-guided vision supervised learning. *Ocean Eng.* 297, 117096. doi: 10.1016/j.oceaneng.2024.117096

Zhou, M., Huang, J., Guo, C.-L., and Li, C. (2023). "Fourmer: An efficient global modeling paradigm for image restoration," in *International conference on machine learning*. (USA: PMLR), 42589–42601.

Zhou, M., Xiao, J., Chang, Y., Fu, X., Liu, A., Pan, J., et al. (2021). "Image de-raining via continual learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (virtual: IEEE), 4907–4916.

Zhou, K., Yang, J., Loy, C. C., and Liu, Z. (2022). "Conditional prompt learning for vision-language models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (USA: IEEE), 16816–16825.