Accredited SINTA 2 Ranking
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online at: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 9 No. 2 (2025) 404 - 415 e-ISSN: 2580-0760

Efficient Hybrid Network with Prompt Learning for Multi-Degradation Image Restoration

Muhammad Yusuf Kardawi[1]*, Laksmita Rahadianti[2]
[1,2]Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia
[1]my.kardawi@gmail.com, [2]laksmita@cs.ui.ac.id

Abstract

Image restoration aims to repair degraded images. Traditional image restoration methods have limited generalization capabilities due to the difficulty in dealing with different types and levels of degradation. On the other hand, contemporary research has focused on multi-degradation image restoration by developing unified networks capable of handling various types of degradation. One promising approach is using prompts to provide additional information on the type of input images and the extent of degradation. Nonetheless, all-in-one image restoration requires a high computational cost, making it challenging to implement on resource-constrained devices. This research proposes a multi-degradation image restoration model based on PromptIR with lower computational cost and complexity. The proposed model is trained and tested on various datasets yet it is still practical for deraining, dehazing, and denoising tasks. By unification convolution, transformer, and dynamic prompt operations, the proposed model successfully reduces FLOPs by 32.07% and the number of parameters by 27.87%, with a comparable restoration result and an SSIM of 34.15 compared to 34.33 achieved by the original architecture for the denoising task.

Keywords: *all-in-one; degradation; prompt; restoration*

How to Cite: Muhammad Yusuf Kardawi and Laksmita Rahadianti, "Efficient Hybrid Network with Prompt Learning for Multi-Degradation Image Restoration ", *J. RESTI (Rekayasa Sist. Teknol. Inf.)* , vol. 9, no. 2, pp. 404 - 415, Apr. 2025.
DOI: https://doi.org/10.29207/resti.v9i2.6381

1. Introduction

Image restoration attempts to reconstruct high-quality images from images degraded by weather conditions, such as rain, snow, and fog, and degradations caused by real-world computer vision systems, such as noise, resolution loss, defocus, color imbalance, etc. Many previous studies have focused on developing methods that often handle image degradation separately or are specific to only one type of degradation (single degradation), such as deblurring [1], denoising [2], deraining [3], dehazing [4], and low-light enhancement [5].

Image restoration methods focusing on a single degradation can be divided into two main categories. First, task-specific image restoration methods are designed and trained to handle one type of degradation, such as denoising, dehazing, or deraining. Second, methods with task-aligned image restoration, where one model is trained separately to address multiple types of degradation. While both approaches have yielded very effective results, they have difficulties in real-world scenarios due to their limited generalization ability [6]. Moreover, separately training models from the same network for each degradation, level, or even across different datasets is computationally expensive and highly impractical.

There has been an increasing research interest in developing all-in-one image restoration methods capable of handling different types of degradation with a single model [7]. The approaches used are also diverse, such as the use of domain translation [8] with multi-attentive feature learning and progressive multi-domain deformable alignment (PMDA), and IDR [9] which carries the concept of ingredient-oriented to improve scalability. GAN-based models such as MACGAN [10] have also been developed with an architecture consisting of four encoder-decoder blocks and an attention block to improve restoration in various domains (land, air, and sea).

The various models mentioned have shown significant progress in multi-degradation image restoration. However, prompt learning is emerging as a promising

Received: 08-02-2025 | Accepted: 12-04-2025 | Published Online: 20-04-2025

new paradigm in deep learning, especially in image restoration. This method allows models to adapt to various tasks by utilizing prompts that encode task-specific information. Mperceiver [11], DA-CLIP [12], CAPTNet [13], and PromptIR [14] have been examples of all-in-one image restoration models based on prompt learning and have shown awe-inspiring capabilities in handling different types of degradation with one unified model.

This research focuses on developing an all-in-one image restoration model that utilizes the capabilities of prompt learning to address various types of image degradation simultaneously. PromptIR has introduced an all-in-one image restoration network by utilizing prompts that provide additional information about the type and level of degradation of the image. Based on this information, it can direct the model in performing image restoration with appropriate degradation. PromptIR [14] can handle various types and levels of degradation without requiring separate training for each condition, making it more efficient than conventional methods that require specialized models for each type of degradation. In addition, PromptIR works using a blind restoration approach that allows the model to restore images based solely on the input without knowing the type of degradation in advance. The flexibility offered by prompt learning in directing the model's focus on specific degradations reinforces the reason for choosing PromptIR as the baseline of this research. The proposed model is a modification to the PromptIR [14] model with the intention to achieve a lower computational cost with equal or better restoration performance.

Therefore, this research aims to create an all-in-one model with lower computational cost and complexity. By improving and modifying the PromptIR [14], this research achieves an optimal balance between performance, computational cost, and model complexity. The lighter model will be highly relevant for real-world applications that require reliable and fast image restoration, especially on devices with limited computational resources.

The main contributions can be summarized as follows: Developed a more efficient all-in-one image restoration model with prompt learning based on hybrid convolutional networks and transformers; The proposed method significantly reduces the computational cost and complexity of the model without drastically compromising the restoration performance; A comprehensive evaluation with full-reference (PSNR, SSIM) and no-reference (PIQE, NIQE) metrics provides a more holistic understanding of the model's performance from various perspectives.

2. Research Methods

2.1 Image Restoration on Single Degradation Images

Single degradation image restoration focuses on improving the quality of images that suffer from one type of degradation or predominantly. The degradation can be noise, blur, haze, or others. However, single-degradation image restoration can only repair one type of degradation that significantly affects the overall image quality. This is because, in this type of restoration, the focus is on addressing that specific degradation by applying algorithms explicitly developed to study one particular type.

The single degradation approach makes it easier to select the correct algorithm and optimize parameters more efficiently to achieve optimal restoration results, since the focus is only on one type of degradation. To understand the different types of degradation in images, Park et al. [15] examines physical models for single degradation, such as rain, snow, and haze, with Equation 1.

$$x_{rain,snow,haze} = T \odot (x^{gt} + S) + (1 - T) \odot A \qquad (1)$$

$x_{rain,snow,haze}$ denotes rain, snow or haze degraded image. $\odot, 1, A, T, S$ is the element multiplication, matrix one, atmospheric light, medium transmission map, rain or snow. It can be noticed that the physical corruption models of rain, snow and fog degradation have similar structures and characteristics. As for the image with noise, it can be represented as Equation 2.

$$x_{noise} = x^{gt} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2 I) \qquad (2)$$

$x_{noise}, x^{gt}, \epsilon$ and σ_ϵ^2 denotes noise image, clean image, noise vector and variance, respectively. Usually, the denoising function shows the characteristics of a low-pass filter, which reduces the noise's high-frequency information.

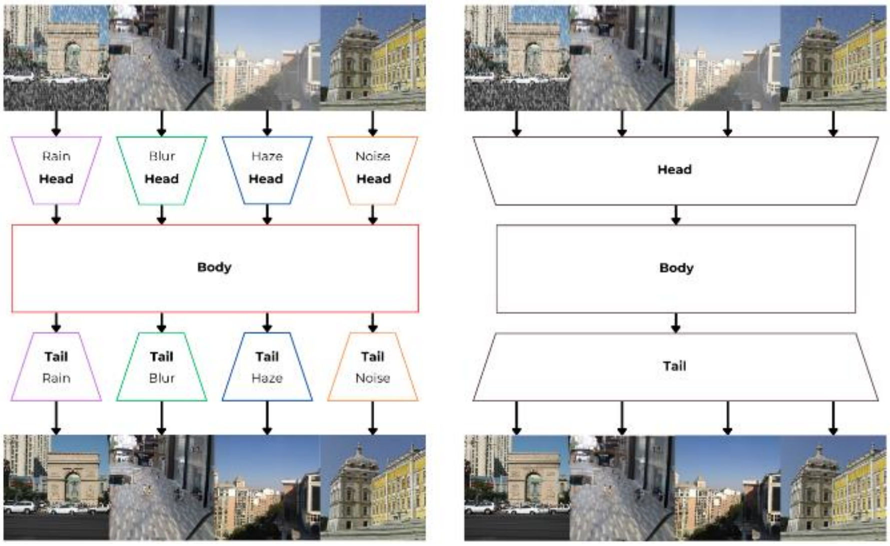
Research focusing on developing task-specific image restoration models has been very popular. In image dehazing, HEDehazeNet [16] produces a diverse transmission map, LID-Net [17] uses a multi-scale lightweight architecture, and DADRNet [18] utilizes domain adaptation and disentangled representation. In denoising, CFNet [19] uses conditional adaptive filters, and IFGLT [20] combines transformers with feature enhancement and information compensation modules. In contrast, MLFAN [21] applies attention to LBP-based multi-level features. Finally, in deraining, DPNet [22] uses transfer learning and frequency domain processing, while UC-former [23] introduces a transformer architecture with channel across attention and multi-scale feature fusion for more efficient and accurate restoration.

In addition, some task-aligned models are SwinIR [24], using residual SWIN transformer blocks for deep feature extraction. Uformer [25] introduces locally-enhanced window transformer blocks and multi-scale restoration modulators. Restormer [26] combines multi-head attention and a feed-forward network to capture remote pixel interactions. Meanwhile, MAXIM [27] uses a multi-axis MLP architecture with multi-axis gated and cross-gating blocks in a UNet hierarchical structure to combine local and global visual information.

The aforementioned image restoration methods have achieved a very good level of performance. This is due to the specialization of models that are developed and trained specifically to handle one particular type of degradation, such as denoising, deblurring or deraining. By focusing on one type of visual artefact, the model can learn the patterns and characteristics of such degradation in greater detail and accuracy. However, such models cannot be applied to images that have multiple types of degradation at once, such as images found in real-world applications.

2.2 All-in-one Image Restoration

Image restoration is transitioning from models with specific degradations to all-in-one methods that are more versatile and can handle more degradations in one network. The all-in-one model aims to build a single network capable of handling different types of degradation, such as denoising, deblurring, dehazing, deraining, and desnowing. This approach simplifies the workflow, increases efficiency, and improves generalization by learning from multiple types of degradation simultaneously. As shown in Figure 1 (a), several types of all-in-one methods have been proposed using different types of inputs to address several adverse weather conditions [28]. Other studies have also developed pre-trained image restoration frameworks with specific outputs for each type of degradation [29]. While other frameworks, as shown in Figure 1 (b), are designed to handle different types of image degradation simultaneously using only one encoder and one decoder. Therefore, the all-in-one method illustrated in Figure 1 (b) is a more promising and practical approach to removing various degradations simultaneously.



(a) All-in-One-Like (b) All-in-One

Figure 1. Illustration of the all-in-one image restoration framework

In single-degradation-based models, the training process is done on a specific type and severity, often making facing real-world images with an unknown mix of degradations and varying severity. In contrast, the all-in-one method uses a single unified network architecture that is capable of performing the restoration of images that have been affected by various degradations, making it a more practical solution for real-world applications.

In addition, task-aligned restoration methods also have the limitation of relying on prior knowledge about the degradation present in the input image, such as the specific degradation type and noise level. This reliance on pre-determined information is impractical in real-world applications where such knowledge is often unavailable or difficult to determine accurately. All-in-one image restoration aims to train a single model capable of restoring a clean image from a degraded input without prior knowledge of the specific type of degradation.

The unified model, as used in the all-in-one framework, attempts to restore an unknown degraded image through noise, blur, rain, snow, or fog by minimizing the following loss function, as shown in Equation 3.

$$\mathcal{L}(\theta_{um}) = \sum_{d=1}^{k}\sum_{n=1}^{N_d} ||G\left(x_{n,d};\theta_{um}\right) - x_{n,d}^{gt}|| \quad (3)$$

d denotes the index for a particular type of degradation, such as noise, rain, and blur, N_d is the number of training samples for a particular task d, G, and θ_{um} denote the network structure and UM network parameters, $||\cdot||$ denotes $l1_{norm}$, and x_{nd} is the degraded image for degradation d with the corresponding ground truth $x_{n,d}^{gt}$.

2.3. All-in-one Image Restoration with CNN

The development of deep learning in recent years has opened up many new opportunities in image restoration research. Convolutional Neural Networks (CNN) have demonstrated outstanding capabilities in image restoration tasks. CNN mimics the workings of the human visual system by utilizing local receptive areas and hierarchical structures similar to biological neural networks. These advantages make CNN very effective in learning the representation of important features of an image, which in turn significantly improves the performance of image restoration methods [30].

Several all-in-one CNN methods have been proposed for image restoration. AirNet [31] uses contrastive learning and deformable convolution but has a high computational cost, so U-WADN [32] was developed with an adaptive backbone width. ADMS [15] uses a degradation-specific filter that is adaptively selected by the degradation classifier. The method proposed by [33] uses two-stage knowledge distillation of task-specific models to handle various weather disturbances. AoSRNet [34] combines CNN with traditional image enhancement techniques for scene recovery under low visibility conditions.

2.4 All-in-one Image Restoration with Transformer

Although initially designed for NLP processing, the transformer architecture, with its self-attention mechanism, has shown great potential in various computer vision tasks, including image restoration. The ability of transformers to model long-range dependencies and capture spatial relationships globally opens up new opportunities for developing more adaptive image restoration methods. Unlike CNNs that tend to focus on local features, transformers can learn a holistic representation of image degradation and guide the restoration process more effectively.

Some transformer-based methods are NDR-Restore [35] using an encoder-decoder architecture with transformers as an attentional mechanism to learn the degradation representation. AIRFormer [6] uses transformers with selective spatial frequency processing through the frequency-guided encoder and frequency-refined decoder to handle weather-induced degradation. TransWeather [36] uses transformers with weather-type embeddings in the decoder to adapt to different types of weather degradation, while the encoder uses intra-patch transformer blocks for detailed feature extraction.

2.5 All-in-one Image Restoration with Prompt Learning

A technique called prompting is used to direct the model in accomplishing a particular task. A prompt is an instruction or context given to the model to guide its output. Although LLM models such as GPT-4, Gemini, and Claude have revolutionized the field of NLP with their ability to understand and generate text, similar concepts are also beginning to show great potential in the visual domain. In healthcare, Zhang et al. [37] proposed Semantic-oriented Visual Prompt Learning to develop a grading model for diabetic retinopathy. Zhu et al. [38] proposes Prompt-based Learning for Unpaired Image Captioning to generate image captions or descriptions without using paired image-text.datasets. In Image Quality Assessment, Fu et al. [39] proposes a model to assess the quality of images generated by AI-generated images by utilizing the Contrastive Language-Image Pretraining model and multi-modal prompt learning. In the field of remote sensing, Gao et al. [40] applies source-free domain adaptation segmentation for remote sensing images that focuses on the use of vision foundation models and prompt learning without the need for direct source data.

This is similar to the use of prompts in the field of NLP. In image restoration, the prompt also serves as a guide for the model. However, the difference is that in image restoration, the prompt is not in text but rather additional information used to direct the model in performing the restoration process according to the type of image degradation present. Prompts will help the model to understand the specific conditions of the degradation that need to be repaired so that the model can restore image quality more precisely and effectively. Prompting is an efficient [41] and an appropriate method to equip the model with relevant knowledge about the type of degradation to produce better image restoration [14].

Several all-in-one method that utilizes prompt learning for image restoration is Mperceiver [11], which uses multimodal prompt learning (textual and visual) with Stable Diffusion to improve adaptation and restoration quality, then DA-CLIP [12] utilizes the vision-language model (CLIP) and prompt learning to guide restoration based on degradation information. Also, PromptIR [14] introduces a prompt block consisting of a prompt generation module (PGM) and prompt interaction module (PIM) to integrate degradation information into the decoder features of the restoration network, thus guiding the restoration process more effectively.

This research focuses on modifying the architecture of the PromptIR [14] model for image restoration with the primary objective of reducing computational cost and model complexity without sacrificing restoration performance. The modifications that have been done in this research start from the hypothesis that the convolution-based PromptIR will show comparable or even superior restoration performance compared to the transformer-based PromptIR and have lower computational cost and complexity. To test these hypotheses, this study will evaluate the performance of the PromptIR and the proposed model on three types of image degradation: haze, rain, and noise.

The choice to modify PromptIR is based on three main reasons, namely: The state-of-the-art performance achieved by PromptIR and its innovative architecture based on prompt learning; PromptIR has demonstrated highly competitive image restoration results and great potential for further development; The prompt mechanism in PromptIR offers great flexibility and adaptability.

Another advantage of the prompt architecture in PromptIR is its plug-and-play nature. The prompt module can be easily integrated into various other architectures. This flexibility opens up vast opportunities for modification and development of PromptIR. One is by combining prompt learning with the convolution block [13], which is expected to produce an image restoration model that is lower in computational cost and complexity without sacrificing performance.

The proposed method uses convolution blocks on each encoder and decoder. The convolution block was adopted from CAPTNet [13], namely the Nonlinear Activation Free Block (NAFBlock) developed by Chen et al. [42]. The selection of CAPTNet as the source of convolutional blocks is based on several factors. First, CAPTNet also carries a prompt learning-based architecture, so the principles and mechanisms of prompts are already integrated into the convolution block. This facilitates integration with the PromptIR framework, which is also prompt-based. Second, CAPTNet has demonstrated competitive and comparable performance with other state-of-the-art research in image restoration and demonstrated the effectiveness and potential of the NAFBlock.

2.6 Utilizing Convolution Blocks

This study offers modifications and developments to the PromptIR architecture [14], namely offering a hybrid approach that combines the advantages of the convolution block (Conv Block) of CAPTNet [13] based on Nonlinear Activation Free Block (NAFBlock) [42] with the framework of PromptIR [14]. This hybrid model can reduce the computational cost and complexity of the model while maintaining or even improving the image restoration capability of various types of degradation.

This study compares the performance of the proposed model with PromptIR [14] as a baseline. Performance evaluation will use four main metrics as described in Section 4.1.2. In addition to image quality, this study will also analyze the computational cost and complexity of the model. The number of FLOPs (Floating Point Operations) and parameters of each model will be calculated and compared. FLOPs will measure the number of floating-point operations required by the model, while the number of parameters indicates the size and complexity of the model. FLOPs and parameter analysis will compare the proposed model's computational cost with the baseline model. Thus, this study not only focuses on the quality of image restoration but also the cost and complexity of the model. This analysis and comparison can provide a deep understanding of the effect of replacing the transformer block with a convolution block. The proposed method is shown in Figure 2.

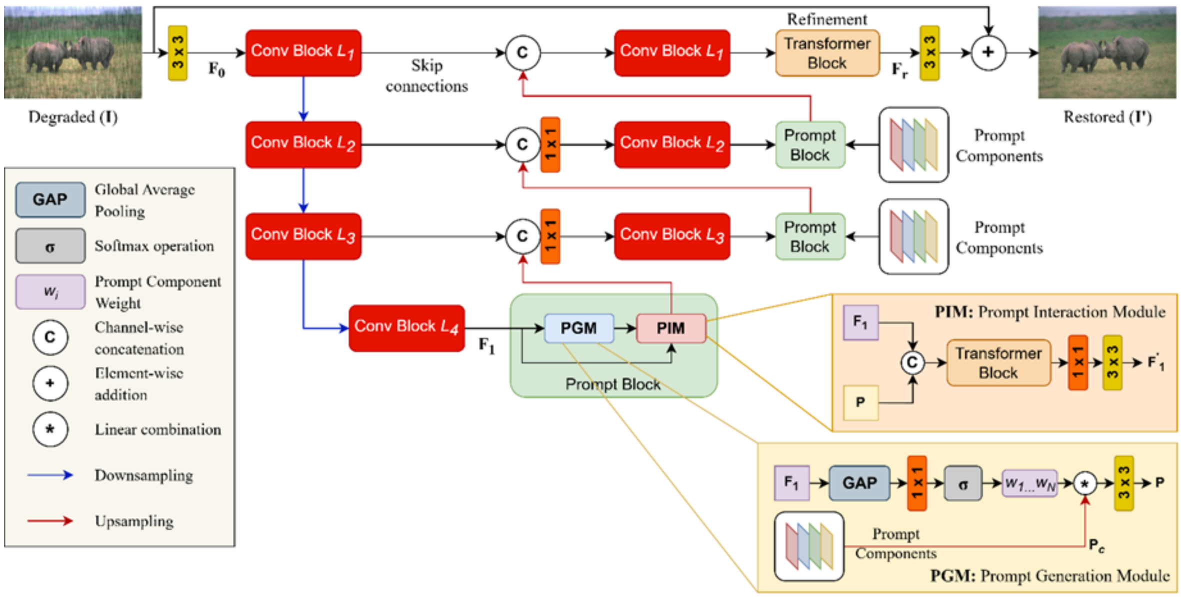


Figure 2. The proposed method with the red block indicates the main modification

The workflow of the proposed method begins with a degraded image (I), which is processed by four consecutive convolution blocks (Conv Block L1 to L4). These convolution blocks are responsible for extracting important features from the image. Figure 3 shows the convolution block used in this architecture. The feature extracted by Conv Block L4, F1, will be used by Prompt Block. Prompt Block consists of Prompt Generation Module (PGM) and Prompt Interaction Module (PIM). PGM generates prompt (P) based on features F1 and Prompt Components. In PGM, a set of learnable prompt components (P_c) is considered latent representations of various degradation characteristics. PGM takes features from the input image to generate attention weights.

Then, these weights are applied to the prompt components by concatenating them to produce P values.

This prompt is forwarded to the PIM, which interacts with the F1 feature through the transformer block. The transformer block plays a role in integrating the prompt information into the decoder features and guiding the restoration network to focus on aspects relevant to the specific degradation of the input image. The transformer block performs self-attention and feed-forward operations on the combined input values. The decoder features and the prompt interact within this transformer block. Then, self-attention connects information from various parts of the image while the feed-forward network processes the information. The presence of a prompt in the input affects the self-attention and feed-forward calculations so that specific degradation information can be integrated into the feature.

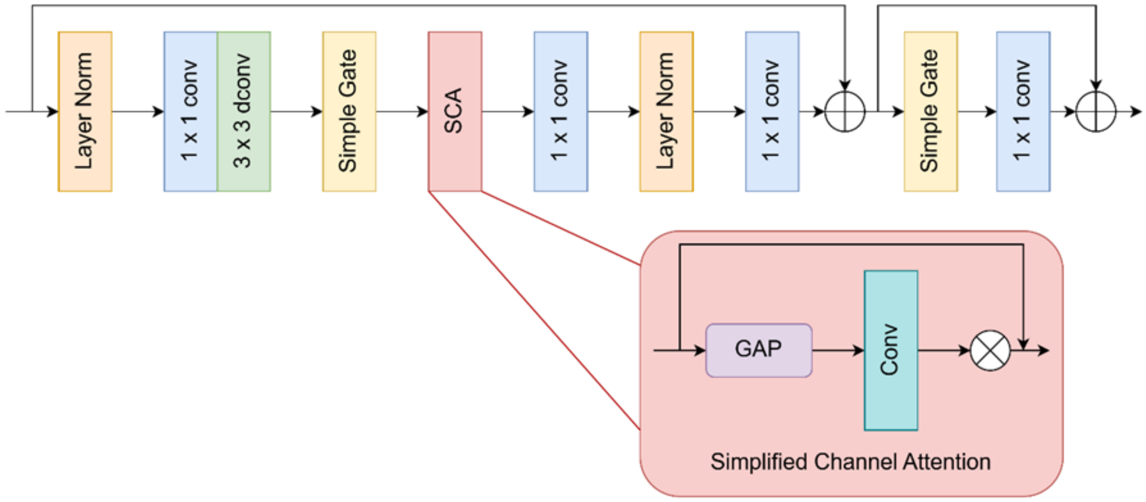


Figure 2. Architecture of the convolution block used

This interaction allows the prompt to guide the restoration process by providing specific degradation information. Next, three Conv Blocks in the decoder use the features modified by the Prompt Block (Conv Block L3 to L1). Skip connection combines features from the encoder to the decoder and retains detailed information. Finally, a 3x3 convolution is applied to the last feature (Fr) and summed with the input image (I) to produce the restored image (I').

The convolution block used is the Nonlinear Activation Free Block (NAFBlock) [42], which consists of convolution, Simplified Gate (SG), and Simplified Channel Attention (SCA). Convolution acts as an essential feature extraction operation to capture local spatial patterns in the image. Simplified Gate (SG) acts as a gating mechanism that controls the flow of information, allowing the model to forward relevant features and limit less important features selectively. This can help to improve model efficiency and prevent overfitting. Simplified Channel Attention (SCA) assigns weights to different feature channels [42], that the model can focus on more informative channels for the restoration task.

The convolutional block, as demonstrated in NAFNet [42] and CAPTNet [13], offers proven effectiveness and efficiency for image processing tasks. Convolutional layers are renowned for their computational efficiency compared to transformers, especially in capturing local spatial dependencies. This efficiency stems from the localized receptive fields of convolutional filters, which require fewer computations than the global attention mechanism employed by transformers. These advantages make convolutional blocks a compelling choice for building efficient and performant image restoration models.

2.7 Hybrid Model with Prompt Learning

This study combines the dynamic prompts of PromptIR with the convolution block of CAPTNet to form a new hybrid model network. In addition, this study also extends the performance evaluation of the hybrid model by adding no-reference evaluation metrics such as PIQE and NIQE. The proposed method of replacing the transformer block with the convolution block of CAPTNet is a rational and promising step. Using the NAFBlock, which has proven effective in CAPTNet, this study can significantly reduce the computational cost and complexity of the PromptIR. Integrating the prompt mechanism of PromptIR into the convolution block of CAPTNet allows the hybrid model to utilize the prompt learning capability to adapt to various types of degradation. This combination produces an image restoration model that is lower in computational cost and complexity and easier to implement on devices with limited resources without sacrificing high restoration performance. In addition, the fundamental differences in the prompt mechanisms between PromptIR and CAPTNet also open up opportunities for further exploration and optimization in designing more effective and efficient prompt interactions.

Several key differences exist in the prompt learning mechanisms used by PromptIR and CAPTNet. These differences lie in how prompts are generated (prompt generation), how prompts interact with the model (prompt interaction), and the main purpose of using prompts.

The explanation of each difference is as follows: In the aspect of prompt generation, PromptIR uses the Prompt Generation Module (PGM) to dynamically generate prompts based on the content of the input image. PGM learns the context of the image and generates prompts specific to that image. In contrast, CAPTNet learns prompts directly from the data without a separate prompt generation module like PGM in PromptIR. Prompts in CAPTNet are learned implicitly during the training process and are located inside the transformer block; In the aspect of prompt interaction, PromptIR uses prompts to guide the transformer block inside the Prompt Interaction Module (PIM). Prompts interact with the features generated by the transformer block and direct the restoration process. On the other hand, CAPTNet integrates prompts directly into the attention mechanism, namely Multi-head Rearranged Attention with Prompts (MRAP). Prompts become an integral part of the attention process and directly affect the attention weights; In terms of objectives, PromptIR prioritizes blind restoration, which is image restoration without prior information about the type of degradation, and easy integration of the prompt module into various architectures. In contrast, CAPTNet focuses more on efficient design and scalability to handle various types of degradation with low computational cost.

This study also visualizes the Prompt Block of the trained model to understand how prompts affect the restoration process. Visualization of each prompt used in the training stage can be seen in Figure 4. The first row is a visualization of the prompt from the haze image. The second row is a visualization of the noise image (σ = 50), and the last row is a visualization of the prompt from the rain image. These three prompts are used due to the architecture of the PromptIR [14], which places the Prompt Block at three different levels in the decoder.

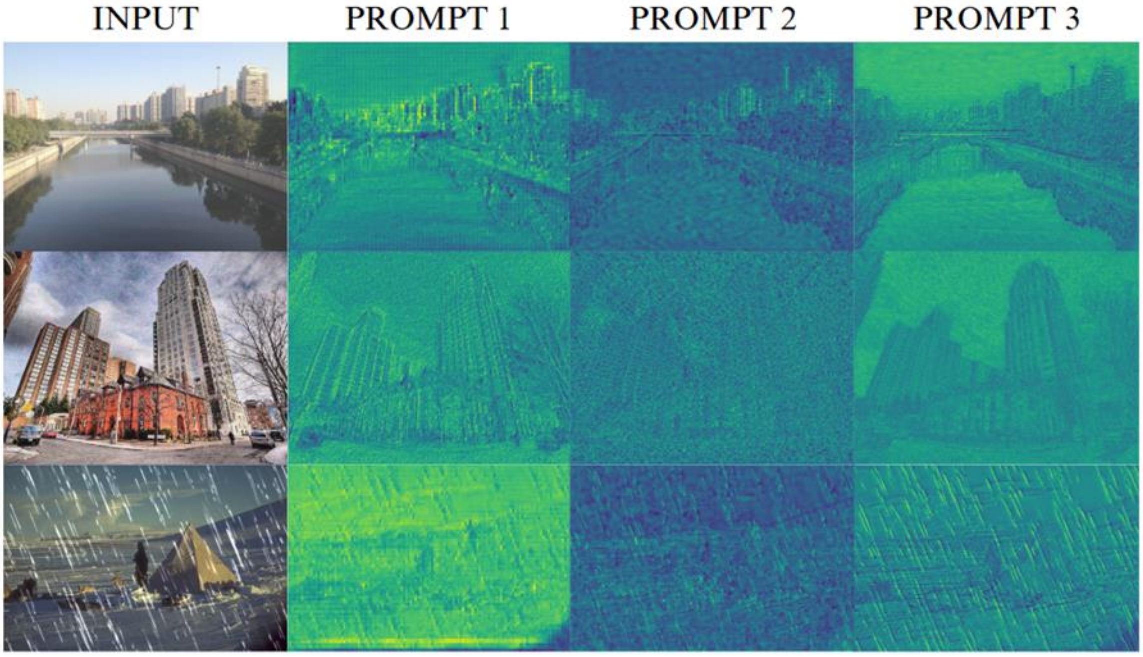


Figure 3. Visualization of the Prompt Block

The main difference between the three prompts lies in the feature level they interact with. Prompt one interacts with features at the lowest resolution level in the decoder, prompt two interacts with features at the middle resolution level, and prompt three interacts with features at the highest resolution level. Since each level of the decoder processes features of different sizes and details, the prompts must also be adjusted. The prompt at the lowest resolution level (prompt 1) focuses on more general information, while the prompt at the highest (prompt 3) focuses on more detailed information, such as texture. Thus, these three different prompts work together to guide the restoration process at different sizes and levels of detail, allowing the model to adapt to different types and severities of degradation.

3. Results and Discussions

3.1 Experimental Setup

This study uses six datasets based on three image degradation types: deraining, denoising, and dehazing. For deraining, the Rain13K [43] dataset is used as training data to train the model, while Rain100L [44] is used as test data to evaluate model performance. For the denoising task, the WED [45] and BSD400 [46] datasets are combined and used as training data, while Urban100 is used as test data. Urban100 [46] is chosen as test data because it has different and more complex image characteristics than the training data, so that it can test the generalization ability of the model. Finally, for dehazing, the RESIDE [47] dataset is used as training and test data. This dataset division ensures a comprehensive and measurable evaluation of each image restoration task. Details of the number of datasets used in this study can be seen in Table 1.

Table 1. Details of the number of datasets used

Degradation	Train	Total	Test	Total
Derain	Rain13K	13.711	Rain100L	100
Denoise	WED BSD400	4.744 400	Urban100	100
Dehaze	RESIDE	13.990	SOTS	500

The performance of the image restoration model in this study will be evaluated using four metrics commonly used in image processing. These metrics include full-reference (FR) metrics such as Peak-Signal to Noise-Ratio (PSNR) and Structural Similarity Index (SSIM). Peak Signal-to-Noise Ratio (PSNR) is a commonly used metric in image restoration to measure the quality of the reconstructed or restored image compared to the original image. It calculates the ratio between the signal's maximum power and the image-damaging noise power, indicating that the noise power is relatively lower than the original signal power. PSNR is a full-reference metric requiring comparison access to a clean original image [48]. PSNR measures how close the restored image is to the original image in terms of pixel difference. Structural Similarity Index Measure (SSIM) is used in image restoration to measure the structural similarity between the restored and original images. Unlike PSNR, which focuses on the difference in pixel values, SSIM is based on the idea that structural changes in the image significantly impact the human visual system. SSIM integrates the three main components of perceptual similarity, namely luminance (brightness), contrast (brightness difference between adjacent pixels), and structure (spatial arrangement of pixels).

Two no-reference (NR) metrics are also used, namely Perception Image Quality Evaluator (PIQE) and Natural Image Quality Evaluator (NIQE). PIQE [49] and NIQE [50] are no-reference metrics for assessing image quality without requiring the original image. PIQE extracts local features and predicts quality based on a generalized distortion statistical model, with lower values indicating better quality. NIQE compares the Multivariate Gaussian model (MVG) of the tested image with the MVG learned from the natural image, with lower values indicating statistical characteristics more similar to the natural image. Using NR metrics such as PIQE and NIQE helps obtain a more comprehensive image quality evaluation. NR metrics do not require target or ground truth images; they rely on statistical values obtained from natural image data.

This allows image quality assessment in real-world scenarios because these scenarios usually do not have ground truth images. By combining FR and NR metrics, this study can calculate the similarity of restored images to ground truth images and study the characteristics of image quality in general.

In the training stage, this study performs parameter tuning with 200 epochs, a learning rate of 0.0002, a batch size of 8, a patch size 128x128, and 16 num workers. These parameters regulate the training process, such as the length of training, the speed of model weight adjustment, the number of data samples in each iteration, the size of the processed image pieces, and the number of processes used to load the data. In this experiment, training data augmentation is still carried out, such as random horizontal and vertical flips on the input image. The process of training and testing the model was carried out on the NVIDIA A100-SXM4-40GB environment.

3.2 Result Evaluation

This study evaluates the effectiveness of the convolution block [13] in improving the computational efficiency of the model. The measurement of the number of FLOPs and parameters will show how much computational cost and model complexity reduction is achieved by replacing the transformer block with the convolution block [13]. This analysis will provide empirical evidence of the effectiveness of the proposed method in achieving the research goal, which is to improve the efficiency of computational cost without sacrificing restoration performance. The evaluation results of the model can be seen in Table 2.

Table 2. Performance Evaluation Results

Degradation	PromptIR				Proposed			
	PSNR	SSIM	PIQE	NIQE	PSNR	SSIM	PIQE	NIQE
Deraining	33.92	0.98	12.52	15.85	32.26	0.97	12.94	16.36
Dehazing	16.36	0.80	11.26	11.73	21.57	0.89	7.04	11.48
Denoising (σ = 15)	34.33	0.96	19.68	8.96	34.15	0.95	18.34	9.24
Denoising (σ = 25)	32.00	0.94	21.03	9.15	31.76	0.93	20.11	9.58
Denoising (σ = 50)	28.75	0.89	22.66	9.83	28.44	0.89	21.51	10.46

Overall, the proposed model performs very close to the PromptIR baseline model, especially on denoising and deraining tasks. On denoising, the proposed model achieves competitive PSNR and SSIM values with the PromptIR baseline at various noise levels (σ = 15, 25, 50). These results indicate that replacing the transformer block with the convolution block [13] does not significantly degrade the model's performance in denoising. Likewise, on the deraining task, the proposed model achieves a PSNR value of 32.26 and an SSIM of 0.9726, indicating that the performance of the proposed method is comparable to the PromptIR baseline. However, on the NR metric, the proposed model successfully provides better PIQE values than the PromptIR baseline model, namely 18.34, 20.11, and

21.51 at each noise level. While on the NIQE metric, the results obtained are similar to the PromptIR.

On the dehazing task, the proposed model shows significant performance improvement compared to PromptIR, with PSNR values of 21.57, SSIM 0.8966, PIQE 7.04, and NIQE 11.48. This improvement is likely due to the ability of the convolution block [13] to extract local features more effectively in haze images. Haze often exhibits complex local spatial patterns, and the convolution block is better at capturing these patterns than the transformer block, which focuses more on global features. These results indicate that the convolution block not only reduces the cost and complexity of the model but can also improve the restoration performance on certain types of degradation, such as haze. These findings support the initial hypothesis that the convolution-based PromptIR can perform comparable or superior to the transformer-based model on image restoration tasks. The samples of restored images on each degradation task can be seen in Figure 5, Figure 6, and Figure 7.

Figure 4. Image deraining comparison

Figure 6. Image dehazing comparison

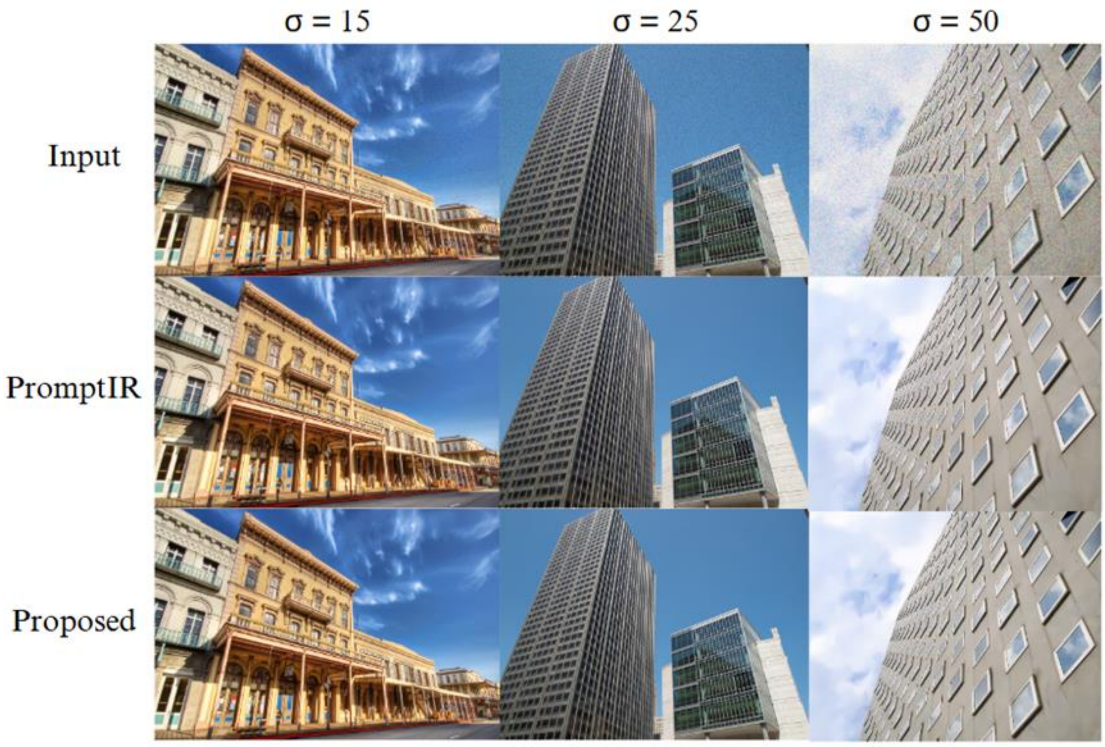


Figure 7. Image denoising comparison

Despite showing good performance in general, the proposed model has limitations in handling rain and haze degradation at different intensity levels. The model successfully removes degradation artefacts on images with low and medium intensity. However, on images with heavy rain and haze, the model is not able to remove the artefacts completely, and some rain and haze artefacts are still visible. This limitation may be due to the unbalanced distribution of data in the training dataset, where images with low and medium-intensity degradation are more dominant than those with heavy degradation. As a result, the model is under-trained for heavy rain and haze cases and has difficulty in generalizing to such scenarios.

3.3 Result Comparison

In addition to providing satisfactory results in PSNR, SSIM, PIQE, and NIQE values, the proposed method also significantly reduced the computational cost. The average number of FLOPs of the proposed method is 664.972 G, which shows a decrease of 32.07% compared to the baseline PromptIR, which has average FLOPs of 978.907 G. This decrease in FLOPs indicates that the proposed method is much more efficient in terms of computation. Table 3 presents a more detailed comparison of the number of FLOPs for each type of degradation task.

Table 3. Comparison of computational costs measured by the number of FLOPs

Method	Denoise	Dehaze	Derain	Average
PromptIR	1620.94 G	945.138 G	370.642 G	978.907 G
Proposed	1101.10 G	642.033 G	251.778 G	664.972 G

This reduction in FLOPs benefits real-time applications or devices with limited computing resources. With lower computational costs, the proposed model can be run faster and requires fewer resources. In addition, Table 4 compares the number of parameters between the proposed and the PromptIR. The proposed model has 23.7798 million parameters, which shows a decrease of 27.87% compared to the PromptIR.

This decrease in the number of parameters indicates that the proposed model is less complex than PromptIR [14]. Model complexity is related to the number of parameters that must be stored and processed during training and inference. More complex models tend to require more memory and processing time. By reducing the number of parameters, the proposed model becomes lighter and easier to implement on devices with limited resources.

Table 4. Comparison of model complexity measured by the number of parameters

Method	Params	Runtime	L1 Loss
PromptIR	32.9644 M	3d 20h 33m 2s	0.01526
Proposed	23.7798 M	3d 12h 43m 28s	0.01630

Finally, Table 4 also compares the loss values and training time between the proposed and PromptIR. The L1 loss value for the proposed model is slightly higher than the PromptIR. This slight difference indicates that despite the architecture modification, the proposed model can still achieve a similar convergence rate during training. Meanwhile, in terms of training time, the proposed model is much shorter than the baseline PromptIR. This reduction in training time shows the efficiency of using convolution blocks. These results make it clear that modifying PromptIR with convolution blocks has increased model efficiency in terms of computational cost (FLOPs) and model complexity (number of parameters).

4. Conclusions

This study successfully developed an image restoration model based on PromptIR with lower computational cost and complexity. The development is done by replacing the transformer block with a convolution block from CAPTNet. The proposed model shows comparable performance to PromptIR on denoising and deraining tasks, even showing significant improvement on the dehazing task. In addition, the proposed model also managed to reduce the computational cost and complexity of the model significantly. This is indicated by a decrease of 32.07% in FLOPs and 27.87% in the number of parameters. These results prove that the convolution block can be an effective alternative to the transformer block in image restoration. Based on the research results obtained, the proposed method still has some limitations on the deraining and dehazing datasets. Although it has applied image augmentation, the limited dataset causes imperfect results in the output image results such as there are remaining rain or haze artifacts that were not successfully restored. The solution that should be considered next is to add new datasets to the degradation or duplicate existing datasets to increase the size of the training data. Besides that, this study needs to be developed by testing on a specially collected real-world dataset as a next step. Testing on real-world data that often contains various types of degradation at once in one image will evaluate the generalization ability of the proposed model in more realistic scenarios. Then, this study also needs to make the CAPTNet model as an additional baseline model. In addition, further research will be carried out by adjusting the model parameters and using more diverse datasets. This is expected to improve further the performance and adaptability of the model in dealing with various types of image degradation in the real world.

Acknowledgements

This research is supported by Lembaga Pengelola Dana Pendidikan (LPDP) under Master Program Scholarship; Tokopedia-UI AI Center; Faculty of Computer Science Universitas Indonesia; and Program Kompetisi Kampus Merdeka under Ministry of Education, Culture, Research, and Technology Republic of Indonesia.

References

[1] S. Cao, H. Fang, L. Chen, W. Zhang, Y. Chang, and L. Yan, "Robust Blind Deblurring Under Stripe Noise for Remote Sensing Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2022, doi: 10.1109/TGRS.2022.3202867.

[2] B. Jiang, Y. Lu, X. Chen, X. Lu, and G. Lu, "Graph Attention in Attention Network for Image Denoising," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 53, no. 11, pp. 7077–7088, 2023, doi: 10.1109/TSMC.2023.3289873.

[3] C. Zhao, W. Cai, C. Hu, and Z. Yuan, "Cycle contrastive adversarial learning with structural consistency for unsupervised high-quality image deraining transformer," *Neural Networks*, vol. 178, no. April, p. 106428, 2024, doi: 10.1016/j.neunet.2024.106428.

[4] H. Sun *et al.*, "Unsupervised Bidirectional Contrastive Reconstruction and Adaptive Fine-Grained Channel Attention Networks for image dehazing," *Neural Networks*, vol. 176, no. April, p. 106314, 2024, doi: 10.1016/j.neunet.2024.106314.

[5] S. Qiao and R. Chen, "Progressive feature fusion for SNR-aware low-light image enhancement," *J. Vis. Commun. Image Represent.*, vol. 100, no. April, p. 104148, 2024, doi: 10.1016/j.jvcir.2024.104148.

[6] T. Gao *et al.*, "Frequency-Oriented Efficient Transformer for All-in-One Weather-Degraded Image Restoration," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 3, pp. 1886–1899, 2024, doi: 10.1109/TCSVT.2023.3299324.

[7] C. Zhang, Y. Zhu, Q. Yan, J. Sun, and Y. Zhang, "All-in-one Multi-degradation Image Restoration Network via Hierarchical Degradation Representation," *MM 2023 - Proc. 31st ACM Int. Conf. Multimed.*, pp. 2285–2293, 2023, doi: 10.1145/3581783.3611825.

[8] P. W. Patil, S. Gupta, S. Rana, S. Venkatesh, and S. Murala, "Multi-weather Image Restoration via Domain Translation," *Proc. IEEE Int. Conf. Comput. Vis.*, no. Iccv, pp. 21639–21648, 2023, doi: 10.1109/ICCV51070.2023.01983.

[9] J. Zhang *et al.*, "Ingredient-oriented Multi-Degradation Learning for Image Restoration," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2023-June, pp. 5825–5835, 2023, doi: 10.1109/CVPR52729.2023.00564.

[10] M. Siddiqua, S. Brahim Belhaouari, N. Akhter, A. Zameer, and J. Khurshid, "MACGAN: An All-in-One Image Restoration under Adverse Conditions Using Multidomain Attention-Based Conditional GAN," *IEEE Access*, vol. 11, no. June, pp. 70482–70502, 2023, doi: 10.1109/ACCESS.2023.3289591.

[11] Y. Ai, H. Huang, X. Zhou, J. Wang, and R. He, "Multimodal Prompt Perceiver: Empower Adaptiveness, Generalizability and Fidelity for All-in-One Image Restoration," *2024 IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 25432–25444, 2023, doi: 10.1109/CVPR52733.2024.02403.

[12] Z. Luo, F. K. Gustafsson, Z. Zhao, J. Sjölund, and T. B. Schön, "Controlling Vision-Language Models for Multi-Task Image Restoration," *12th Int. Conf. Learn. Represent. ICLR 2024*, pp. 1–21, 2024.

[13] H. Gao, J. Yang, Y. Zhang, N. Wang, J. Yang, and D. Dang, "Prompt-based Ingredient-Oriented All-in-One Image Restoration," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 1, pp. 9458–9471, 2024, doi: 10.1109/TCSVT.2024.3398810.

[14] V. Potlapalli, S. W. Zamir, S. Khan, and F. S. Khan, "PromptIR: Prompting for All-in-One Blind Image Restoration," *Adv. Neural Inf. Process. Syst.*, vol. 36, no. NeurIPS, 2023.

[15] D. Park, B. H. Lee, and S. Y. Chun, "All-in-One Image Restoration for Unknown Degradations Using Adaptive Discriminative Filters for Specific Degradations," *2023 IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 5815–5824, 2023, doi: 10.1109/cvpr52729.2023.00563.

[16] W. Li, D. Fan, Q. Zhu, Z. Gao, and H. Sun, "HEDehazeNet: Unpaired image dehazing via enhanced haze generation," *Image Vis. Comput.*, vol. 150, no. June, 2024, doi: 10.1016/j.imavis.2024.105236.

[17] F. Tao, Q. Chen, Z. Fu, L. Zhu, and B. Ji, "LID-Net: A lightweight image dehazing network for automatic driving vision systems," *Digit. Signal Process. A Rev. J.*, vol. 154, no. July, p. 104673, 2024, doi: 10.1016/j.dsp.2024.104673.

[18] X. Li, H. Yu, C. Zhao, C. Fan, and L. Zou, "DADRnet: Cross-domain image dehazing via domain adaptation and disentangled representation," *Neurocomputing*, vol. 544, 2023, doi: 10.1016/j.neucom.2023.126242.

[19] Y. Zuo *et al.*, "CFNet: Conditional filter learning with dynamic noise estimation for real image denoising," *Knowledge-Based Syst.*, vol. 284, no. December 2023, p. 111320, 2024, doi: 10.1016/j.knosys.2023.111320.

[20] F. Liu, Z. Zhou, C. Men, Q. Sun, and K. Huang, "IFGLT: Information fusion guided lightweight Transformer for image denoising," *J. Vis. Commun. Image Represent.*, vol. 97, no. June, p. 103994, 2023, doi: 10.1016/j.jvcir.2023.103994.

[21] A. Ulu, G. Yildiz, and B. Dizdaroglu, "MLFAN: Multilevel

Feature Attention Network With Texture Prior for Image Denoising," *IEEE Access*, vol. 11, no. April, pp. 34260–34273, 2023, doi: 10.1109/ACCESS.2023.3264604.

[22] H. Yang, D. Zhou, J. Cao, and Q. Zhao, "DPNet: Detail-preserving image deraining via learning frequency domain knowledge," *Digit. Signal Process. A Rev. J.*, vol. 130, p. 103740, 2022, doi: 10.1016/j.dsp.2022.103740.

[23] W. Zhou and L. Ye, "UC-former: A multi-scale image deraining network using enhanced transformer," *Comput. Vis. Image Underst.*, vol. 248, no. July, p. 104097, 2024, doi: 10.1016/j.cviu.2024.104097.

[24] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image Restoration Using Swin Transformer," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2021-October, pp. 1833–1844, 2021, doi: 10.1109/ICCVW54120.2021.00210.

[25] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, "Uformer: A General U-Shaped Transformer for Image Restoration," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2022-June, pp. 17662–17672, 2022, doi: 10.1109/CVPR52688.2022.01716.

[26] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M. H. Yang, "Restormer: Efficient Transformer for High-Resolution Image Restoration," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2022-June, pp. 5718–5729, 2022, doi: 10.1109/CVPR52688.2022.00564.

[27] Z. Tu *et al.*, "MAXIM: Multi-Axis MLP for Image Processing," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2022-June, pp. 5759–5770, 2022, doi: 10.1109/CVPR52688.2022.00568.

[28] R. Li, R. T. Tan, and L. F. Cheong, "All in One Bad Weather Removal Using Architectural Search," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 3172–3182, 2020, doi: 10.1109/CVPR42600.2020.00324.

[29] H. Chen *et al.*, "Pre-trained image processing transformer," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 12294–12305, 2021, doi: 10.1109/CVPR46437.2021.01212.

[30] K. Wang, L. Zhuo, J. Li, T. Jia, and J. Zhang, "Learning an Enhancement Convolutional Neural Network for Multi-degraded Images," *Sens. Imaging*, vol. 21, no. 1, pp. 1–15, 2020, doi: 10.1007/s11220-020-00289-0.

[31] B. Li, X. Liu, P. Hu, Z. Wu, J. Lv, and X. Peng, "All-In-One Image Restoration for Unknown Corruption," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2022-June, pp. 17431–17441, 2022, doi: 10.1109/CVPR52688.2022.01693.

[32] Y. Xu, N. Gao, Z. Shan, F. Chao, and R. Ji, "Unified-Width Adaptive Dynamic Network for All-In-One Image Restoration," 2024, [Online]. Available: http://arxiv.org/abs/2401.13221

[33] W. T. Chen, Z. K. Huang, C. C. Tsai, H. H. Yang, J. J. Ding, and S. Y. Kuo, "Learning Multiple Adverse Weather Removal via Two-stage Knowledge Learning and Multi-contrastive Regularization: Toward a Unified Model," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2022-June, no. i, pp. 17632–17641, 2022, doi: 10.1109/CVPR52688.2022.01713.

[34] Y. Lu, D. Yang, Y. Gao, R. W. Liu, J. Liu, and Y. Guo, "AoSRNet: All-in-One Scene Recovery Networks via multi-knowledge integration," *Knowledge-Based Syst.*, vol. 294, 2024, doi: 10.1016/j.knosys.2024.111786.

[35] M. Yao, R. Xu, Y. Guan, J. Huang, and Z. Xiong, "Neural Degradation Representation Learning for All-In-One Image Restoration," vol. 14, no. 8, pp. 1–13, 2023, [Online]. Available: http://arxiv.org/abs/2310.12848

[36] J. M. Jose Valanarasu, R. Yasarla, and V. M. Patel, "TransWeather: Transformer-based Restoration of Images Degraded by Adverse Weather Conditions," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2022-June, pp. 2343–2353, 2022, doi: 10.1109/CVPR52688.2022.00239.

[37] Y. Zhang, X. Ma, K. Huang, M. Li, and P. A. Heng, "Semantic-Oriented Visual Prompt Learning for Diabetic Retinopathy Grading on Fundus Images," *IEEE Trans. Med. Imaging*, vol. 43, no. 8, pp. 2960–2969, 2024, doi: 10.1109/TMI.2024.3383827.

[38] P. Zhu *et al.*, "Prompt-Based Learning for Unpaired Image Captioning," *IEEE Trans. Multimed.*, vol. 26, pp. 379–393, 2024, doi: 10.1109/TMM.2023.3265842.

[39] J. Fu, W. Zhou, Q. Jiang, H. Liu, and G. Zhai, "Vision-Language Consistency Guided Multi-Modal Prompt Learning for Blind AI Generated Image Quality Assessment," *IEEE Signal Process. Lett.*, vol. 31, pp. 1820–1824, 2024, doi: 10.1109/LSP.2024.3420083.

[40] K. Gao, X. You, K. Li, L. Chen, J. Lei, and X. Zuo, "Attention Prompt-Driven Source-Free Adaptation for Remote Sensing Images Semantic Segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, 2024, doi: 10.1109/LGRS.2024.3422805.

[41] M. Jia *et al.*, "Visual Prompt Tuning," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 13693 LNCS, pp. 709–727, 2022, doi: 10.1007/978-3-031-19827-4_41.

[42] L. Chen, X. Chu, X. Zhang, and J. Sun, "Simple Baselines for Image Restoration," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 13667 LNCS, pp. 17–33, 2022, doi: 10.1007/978-3-031-20071-7_2.

[43] S. W. Zamir *et al.*, "Multi-stage progressive image restoration," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 14816–14826, 2021, doi: 10.1109/CVPR46437.2021.01458.

[44] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan, "Deep joint rain detection and removal from a single image," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 1685–1694, 2017, doi: 10.1109/CVPR.2017.183.

[45] K. Ma *et al.*, "Waterloo exploration database: New challenges for image quality assessment models," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 1004–1016, 2017, doi: 10.1109/TIP.2016.2631888.

[46] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2, pp. 416–423, 2001, doi: 10.1109/ICCV.2001.937655.

[47] B. Li *et al.*, "Benchmarking Single-Image Dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, 2019, doi: 10.1109/TIP.2018.2867951.

[48] M. Y. Hossain, M. M. H. Rakib, S. Rajit, I. R. Nijhum, and R. M. Rahman, "Adaptive and automatic aerial image restoration pipeline leveraging pre-trained image restorer with lightweight Fully Convolutional Network," *Expert Syst. Appl.*, vol. 259, no. March 2024, p. 125210, 2025, doi: 10.1016/j.eswa.2024.125210.

[49] N. Venkatanath, D. Praneeth, B. H. Maruthi Chandrasekhar, S. S. Channappayya, and S. S. Medasani, "Blind image quality evaluation using perception based features," *2015 21st Natl. Conf. Commun. NCC 2015*, pp. 1–6, 2015, doi: 10.1109/NCC.2015.7084843.

[50] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, 2013, doi: 10.1109/LSP.2012.2227726.