All-In-One Image Restoration

Feipeng Rong 30920241154572[1], Ying Ye 30920241154557[1], Xian Wu 36920241153259[2*], Peijie Xu 20420241152053[3]

[1]School of Informatics, Xiamen University
[2]Artificial Intelligence Research Institute, Xiamen University
[3]College of Chemistry and Chemical Engineering, Xiamen University
30920241154572@stu.xmu.edu.cn,3118406833@qq.com,1035082699@qq.com,xpj886d@gmail.com

Abstract

In real-world applications, images are often affected by various unknown degradations, such as noise, fog, and rain, which can severely impact safety, especially in fields like autonomous driving. To address this issue, we propose an All-In-One, end-to-end image restoration model designed to recover images from various types and levels of unknown degradation. Our approach involves a prompt-guided multi-scale, end-to-end image restoration algorithm. We introduce a multi-head dual self-attention mechanism to capture dependencies between spatial and channel dimensions in the image, enhancing the network's ability to learn image features. Moreover, we propose a prompt block that leverages the concept of prompting to implicitly encode degradation information, guiding the network to distinguish between different degradation types at each layer. We demonstrate the algorithm's superiority through extensive experiments on established benchmark datasets.

Introduction

As an efficient and intuitive information carrier, digital images play a vital role in the development of digital information technology. In recent years, the demand for digital images has grown explosively, with high-quality image applications penetrating various aspects of everyday life. However, obtaining high-quality images is not easy, as many complex and uncertain factors can affect image quality, leading to varying degrees of degradation. For instance, inherent hardware defects in imaging devices can result in various artifacts and deterioration; lighting conditions and photoelectric conversion devices may introduce Gaussian noise, thermal noise, or speckle noise; and harsh weather conditions often cause images to be affected by rain, fog, snow, or dust.

These degraded images not only significantly affect human visual perception quality but also increase the difficulty of subsequent image content analysis and understanding.

It is worth noting that, in the real world, images are often affected by more than just one known type of degradation. For example, during image acquisition in autonomous driving, a vehicle may be subjected to multiple unknown degradations, such as noise, fog, and rain, either consecutively or simultaneously, which can severely impact driving safety. To address this issue, developing an All-In-One end-to-end image restoration model capable of recovering images from various unknown types and levels of degradation holds significant research value for many scenarios that rely on high-quality digital images.

The main contributions can be summarized as follows:

- We propose a novel multi-scale, prompt-guided All-In-One image restoration algorithm that leverages a comprehensive multi-scale framework to effectively address diverse and unknown corruption types and levels.
- We design the Dsa Block and Prompt Block to augment the restoration capabilities of our algorithm, enabling selective attention to damaged regions and incorporating external guidance to enhance recovery, thereby improving both structural integrity and visual quality of the restored images.
- Extensive experiments and analysis are conducted, including both quantitative and visual comparisons with baseline methods. Ablation studies further validate the effectiveness of each proposed module, while multiple degradation types are combined to investigate their respective impacts on the algorithm's performance.

Related work

Current image restoration algorithms are classified into two categories: Image Restoration for Single Degradation (IRSD) and Image Restoration for Multiple Degradations (IRMD). The All-In-One image restoration algorithm builds upon the framework of IRMD.

IRSD serves as the foundation for IRMD, focusing on restoring clean images from those affected by a single type of degradation. This algorithm effectively addresses single-type degradation issues with a relatively simple and targeted design, yielding good results for specific degradations; however, its performance is limited in IRMD tasks.

Zhang et al.(Zhang et al. 2017a) introduced DnCNN, an early denoising model leveraging deep learning to achieve favorable outcomes. DnCNN expands TNRD through deep residual learning. A common method for solving restoration problems is semi-iterative splitting, utilized by IRCNN (Zhang et al. 2017b), which employs neural networks to

*Corresponding author.

achieve similar goals as DnCNN. Given that semi-iterative splitting is an iterative approach, IRCNN trains a series of denoising networks applicable not only to denoising but also to various restoration tasks. FFDNet(Zhang, Zuo, and Zhang 2018) innovatively incorporates noise levels as inputs alongside noisy images, effectively addressing diverse noise levels and spatial variations. Subsequent denoising studies(Wang et al. 2022)have concentrated on developing more efficient network structures. Zamir et al.(Zamir et al. 2022)adapted concepts from the Transformer(Vaswani 2017)model used in natural language processing to image restoration, enhancing the computational efficiency of visual Transformers (VIT)(Dosovitskiy 2020)through modifications to the self-attention mechanism. Image deraining is also a prominent research area with extensive applications, with future trends likely to integrate both data-driven and model-driven methodologies. Zhang et al.(Zhang et al. 2019)proposed a coarse-to-fine model training strategy focused on datasets, yielding a simple yet high-performing network structure.

The image restoration task involving multiple degradations has emerged only in recent years. Prior research typically addressed the IRMD task through multi-input and multi-output network architectures. Li et al.(Li, Tan, and Cheong 2020)developed a model that mitigates weather effects to tackle various adverse weather degradations, such as rain, fog, and snow, with each degradation handled by dedicated encoders. Chen et al. (Chen et al. 2021)proposed a transformer-based image restoration approach that employs a multi-head and multi-tail architecture to address multiple degradations; additionally, this method necessitates pretraining on large-scale datasets. In summary, while the aforementioned IRMD methods represent advancements toward All-In-One image restoration, they still depend on prior degradation information to appropriately route inputs to the respective correction heads.

In 2022, Li et al.(Li et al. 2022)introduced an All-In-One model called AirNet for denoising, deraining, and dehazing. This model is the first of its kind to recover images from various degradation types and levels. It employs contrastive learning to train an image encoder, effectively modeling critical information regarding the degradations. These representations are then used to predict deformable convolution offsets in a separate network for the restoration process. The method necessitates two stages of training, where the successful selection of positive and negative pairs, as well as the amount of available data, significantly impacts the effectiveness of contrastive learning.

Although AirNet achieves state-of-the-art performance, it still faces challenges in modeling representations for different types of degradation. Additionally, the two-stage training process, which relies on an extra encoder for contrastive learning, leads to increased training complexity. Potlapalli et al.(Potlapalli et al. 2023)proposed a prompt learning-based approach called PromptIR for All-In-One image restoration, achieving state-of-the-art results in denoising, deraining, and dehazing. With the advent of the pretrained language-vision model CLIP(Radford et al. 2021), Jiang et al.(Jiang et al. 2023)developed AutoDIR, an All-In-One image restoration model that utilizes text prompts. Its blind image quality assessment module identifies the primary degradation in degraded images and guides the restoration process via text prompts in the latent diffusion model. Additionally, a structural correction module enhances the details of the restored images. AutoDIR offers flexible user control and editing during runtime, allowing users to alternate between text prompts generated by the blind image quality assessment module and additional user-provided text prompts.

Proposed Solution

The algorithm we propose utilizes a multi-scale architecture, with the baseline model being the Unet(Ronneberger, Fischer, and Brox 2015) network. By employing a multi-scale feature fusion strategy, the model can simultaneously process low-level detail and high-level semantic information. This strategy strengthens the model's capability to learn features at various scales by integrating feature maps of different resolutions at different stages of the network. The network architecture is illustrated in Figure 1.

Dual Self Attention Block

In 2017, Vaswani et al.(Vaswani 2017)introduced the self-attention (SA) mechanism in the Transformer model. Prior to this, most attention-based models for sequence tasks focused on capturing the relationship between source and target sequences. These models processed input as vectors of varying dimensions, with inherent correlations between them. However, these correlations were not effectively utilized during training, leading to suboptimal model performance. The self-attention mechanism addresses this issue by enhancing the network's capacity to capture the relationships between different parts of the input data. Through self-attention, the network can identify and emphasize dependencies between various regions of the input, allowing for more precise feature extraction and better integration of information, particularly when handling complex data.

We use Multiple Head Dual Self Attention (MHDSA) to replace the conventional SA, which has quadratic complexity that increases with the input image resolution. MHDSA consists of two parallel attention modules, Multiple Head Position Self Attention (MHPSA) and Multiple Head Channel Self Attention (MHCSA), which apply SA along the spatial and channel dimensions, respectively. This approach enables the network to capture richer contextual information, understand both the global structure and local details of the image, model complex image structures and long-range dependencies, and enhance overall performance. The structure of MHDSA is illustrated in Figure 2.

MHCSA and MHPSA first perform layer normalization (LN) independently on each sample. Then, by applying a 1×1 convolution followed by a 3×3 depthwise convolution, they obtain the feature map $U \in R^{H \times W \times C}$, which is subsequently divided into smaller sections. Based on the differing dimensions involved in the two self-attention mechanisms, the feature map is then reshaped into a channel dimension $M_c \in R^{C \times N}$ and a spatial dimension $M_p \in R^{N \times C}$, where $N = H \times W$. From these sections, the Query (Q), Key (K),

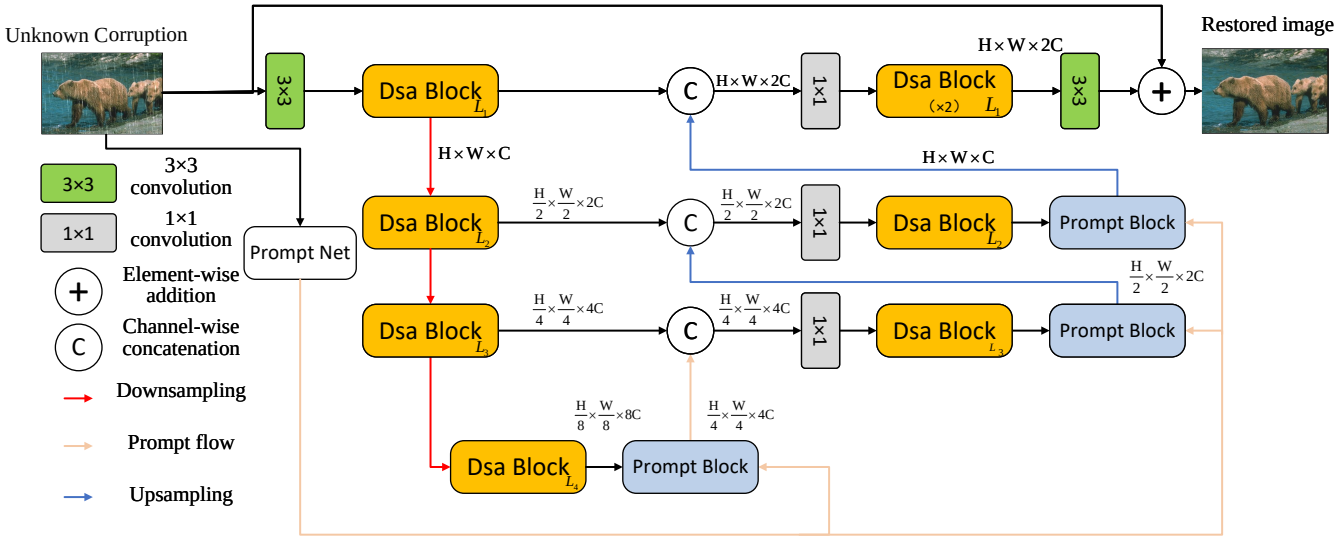

Figure 1: Architecture of the proposed model. The network architecture diagram illustrates the components of the system, which include the **Dual Self Attention (Dsa) Block, Prompt Block, and Prompt Net**. High-resolution degraded images are first passed through the Prompt Net to acquire information about the type of degradation. In the encoder, multiple Dsa Blocks progressively increase the number of channels while reducing spatial resolution. The low-resolution latent features are then gradually restored to high-resolution clean outputs through the decoder. At each step of the decoding process, the Prompt Block, which has been processed using the prompt information flow, is embedded to guide the restoration.

and Value (V) are derived for the self-attention mechanism. Next, Q and K are reshaped, and the attention matrices of various dimensions are obtained by computing $Q^T \times K$. Applying the softmax operation to these matrices yields the attention weights affecting each dimension. The resulting similarity matrix after softmax is then multiplied by V, and the outcome is passed through a 1×1 convolution and added to the original feature map via a residual connection. Mathematically,

$$\begin{aligned} Y &= w_1 \hat{X}_C + w_2 \hat{X}_P \\ \hat{X}_C &= W_{1\times1} Attention_C\left(\hat{Q},\hat{K},\hat{V}\right) + X \\ \hat{X}_P &= W_{1\times1} Attention_P\left(\hat{Q},\hat{K},\hat{V}\right) + X \\ Attention\ &\left(\hat{Q},\hat{K},\hat{V}\right) = \hat{V}\cdot softmax\left(\hat{K}\cdot\hat{Q}/\alpha\right) \end{aligned} \tag{1}$$

Where w_1 and w_2 are two learnable parameters used to fuse feature maps, balancing the outputs of the channel and spatial self-attention mechanisms. $Attention_C$ and $Attention_P$ represent the channel-wise and spatial self-attention mechanisms of different dimensionalities, respectively. $W_{1\times1}$ denotes a 1×1 convolution. The parameter α is a learnable scaling factor used to control the magnitude of the dot product between K and Q before applying the softmax function.

Prompt Block

We propose a prompt-guided approach to image restoration, where prompts serve as an effective strategy for All-In-One restoration tasks. This method not only restores clean images but also leverages knowledge of various degradation types to enhance the model's capability to handle different forms of degradation.As shown in Figure 3.

Prompt Block is composed of a Prompt Generation Module (PGM) and a Prompt Interaction Module (PIM). First, the Prompt Net is used to produce prompts that contain degraded category information, and these prompts are then utilized during the restoration process to guide the model. The Prompt Net itself is an image classification network; in this study, ResNet50 (He et al. 2016)was selected as the Prompt Net. The computation of the Prompt Block as defined by:

$$\hat{F} = PIM\left(PGM\left(P_c, P_i, F\right), F\right) \tag{2}$$

whereP_c denotes the learnable Prompt components, P_i represents the prompt information generated by the Prompt Net, and F is the output feature.

The prompt component P_c refers to a set of learnable parameters that interact with input features and prompt information to embed regression information. The PGM dynamically predicts attention-based weights from input features and applies them to generate the prompt component P_c. To derive prompt weights from input features F, the PGM first applies global average pooling across the spatial dimensions to generate feature vectors. Then, a 1×1 convolutional layer reduces the number of channels to obtain compact prompt features, followed by a softmax function to generate prompt weights. Finally, the prompt weights and prompt information are used to adjust the prompt component dynamically, with a 3×3 convolutional layer further refining the adjust-

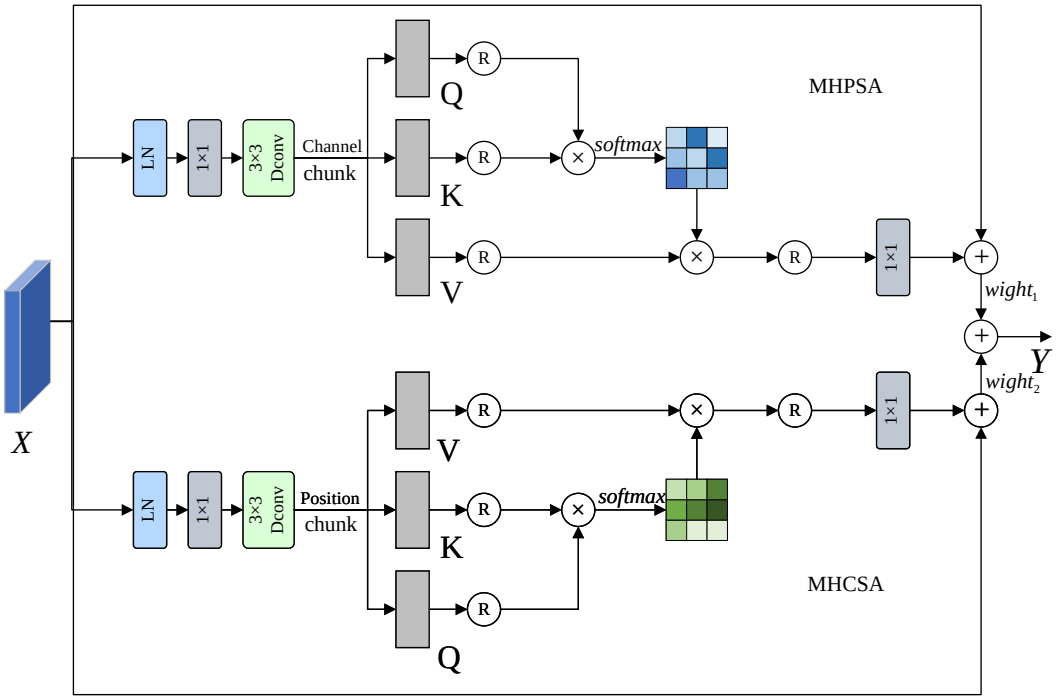

Figure 2: The details of Multiple Head Dual Self Attention.

Method	Denoise			Derain	Dehaze	Average
	BSD68($\sigma = 15$)	BSD68($\sigma = 25$)	BSD68($\sigma = 50$)	Rain100L	SOTS	
BRDNet	31.86/0.895	29.71/0.837	26.25/0.705	27.15/0.843	22.78/0.9	27.55/0.836
LPNet	26.55/0.819	24.65/0.715	21.64/0.502	23.62/0.773	19.89/0.741	23.27/0.71
FDGAN	30.66/0.906	28.52/0.847	26.78/0.726	28.63/0.883	23.94/0.867	27.73/0.846
MPRNet	33.39/0.923	30.73/0.874	27.37/0.765	29.15/0.903	25.38/0.938	29.21/0.881
DL	33.23/0.922	30.61/0.874	27.35/0.771	29.17/0.894	25.13/0.923	29.1/0.877
AirNet	33.59/0.927	30.9/0.878	27.51/0.772	32.89/0.942	25.28/0.937	30.03/0.891
PromptIR	33.79/0.932	31.14/0.885	27.85/0.789	33.99/0.957	27.75/0.959	30.9/0.904
Ours	**33.94/0.932**	**31.29/0.888**	**28.04/0.798**	**35.42/0.968**	**28.99/0.967**	**31.53/0.911**

Table 1: Performance comparisons on three challenging datasets.

ments. The calculation of PGM is shown in:

$$\begin{aligned} P &= Conv_{3\times3}\left(\sum_{c=1}^{N} w_i P_c P_i\right) \\ w_i &= Softmax\left(Conv_{1\times1}(GAP(F))\right) \end{aligned} \quad (3)$$

The main goal of the Prompt Interaction Module (PIM) is to establish the relationship between the input feature F and the prompt P to facilitate guided restoration. In PIM, the generated prompt P is concatenated with the input features along the channel dimension and then passed through the Dsa Block for feature transformation. This transformation utilizes the degradation information encoded in the prompt P to adjust the input features. The calculation of PIM is shown in:

$$\hat{F} = Conv_{3\times3}(Conv_{1\times1}(Dsa([F;P])) \quad (4)$$

Where$[;]$ represents the concatenation operation.

Experiments

To evaluate the proposed method, we carry out comprehensive experiments on BSD dataset(Martin et al. 2001), Rain100L dataset(Yang et al. 2017) and RESIDE dataset(Li et al. 2018). Experimental results demonstrate that method achieves state-of-the-art performance on All-In-One task.

Datasets

BSD: BSD is a dataset proposed in 2001 for image segmentation. In this study, the datasets BSD400 and BSD68, which are subsets derived from BSD, are used. BSD400 consists of 400 images, while BSD68 includes 68 images. These serve as the training and test sets for the image denoising task. The noisy images are generated by adding noise to the clean images.The noise is generated by manually adding Gaussian noise of different levels, specifically $\sigma = 15, 25, 50$, to introduce varying degrees of noise.

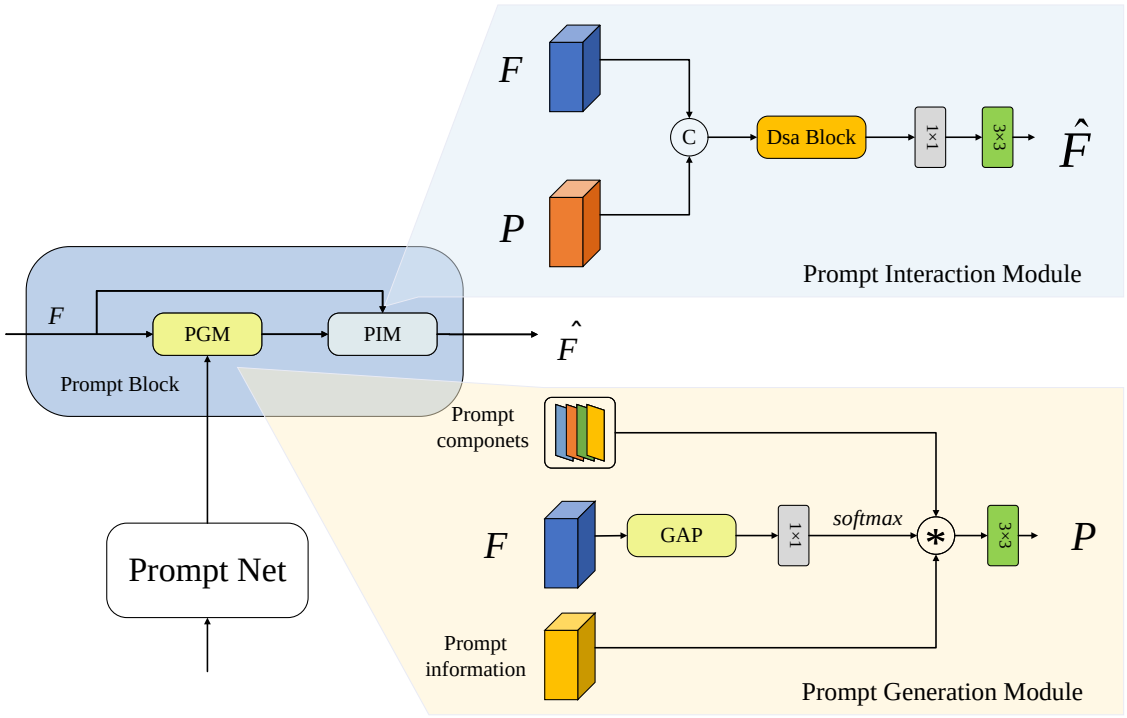

Figure 3: The details of Prompt Block.

Rain100L: Rain100L dataset was proposed in 2017 as a synthetic dataset for rain image removal. This dataset is derived from certain images in the BSD200 dataset, where synthetic raindrop streaks with sharp edges are added by simulating rain streaks in one direction. The angles of the streaks in the images have very little variation. The dataset consists of 300 image pairs, with 200 pairs used as the training set and 100 pairs as the test set.

RESIDE: RESIDE is a large-scale benchmark dataset for image dehazing, proposed in 2018. The RESIDE dataset is divided into five subsets: Indoor Training Set (ITS), Outdoor Training Set (OTS), Synthetic Objective Testing Set (SOTS), Real-world Task-Driven Testing Set (RTTS), and Hybrid Subjective Testing Set (HSTS). Each subset is designed for specific training or evaluation needs to meet different research requirements.In this study, OTS and SOTS are selected as the training and testing sets, respectively. OTS and SOTS consist of 72,135 and 500 image pairs, respectively. We use 800 image pairs from OTS for training.

Baselines and Implementation Details

We conducted a comparison between our proposed method and eight baseline approaches:BRDNet(Tian, Xu, and Zuo 2020), LPNet(Gao et al. 2019), FDGAN(Dong et al. 2020), MPRNet(Zamir et al. 2021), DL(Fan et al. 2019), AirNet(Li, Tan, and Cheong 2020) and PromptIR(Potlapalli et al. 2023).Our proposed method in implemented in PyTorch, utilizing the Adam optimizer with an initial learning rate of 2e-4.The training process includes 200 epochs, a batch size of 4, and random cropping for input data. A learning rate cosine annealing strategy is adopted, with a warm-up phase for the learning rate during the first 15 epochs.

Results on All-In-One task

As shown in Table 1,the proposed algorithm significantly improves the accuracy of the All-In-One image restoration algorithm. In eight comparative experiments, the proposed algorithm achieves the best average PSNR and SSIM metrics for each image restoration task. Compared to the state-of-the-art PromptIR algorithm, the proposed method improves the average PSNR metric by 0.63dB and the average SSIM metric by 0.007 for the three types of image restoration tasks. The experimental results demonstrate that the multi-scale All-In-One image restoration algorithm based on prompt-guided learning proposed in this study can effectively address All-In-One image restoration challenges.

Conclusion

In this paper, we proposes an end-to-end multi-scale image restoration algorithm guided by prompts.The method provides a comprehensive solution for restoring images affected by various types of corruptions. Extensive experiments demonstrate the exceptional performance of model in both qualitative and quantitative evaluations.

References

Chen, H.; Wang, Y.; Guo, T.; Xu, C.; Deng, Y.; Liu, Z.; Ma, S.; Xu, C.; Xu, C.; and Gao, W. 2021. Pre-Trained Image Processing Transformer. In *Computer Vision and Pattern Recognition*.

Dong, Y.; Liu, Y.; Zhang, H.; Chen, S.; and Qiao, Y. 2020. FD-GAN: Generative adversarial networks with fusion-discriminator for single image dehazing. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 10729–10736.

Dosovitskiy, A. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Fan, Q.; Chen, D.; Yuan, L.; Hua, G.; Yu, N.; and Chen, B. 2019. A general decoupled learning framework for parameterized image operators. *IEEE transactions on pattern analysis and machine intelligence*, 43(1): 33–47.

Gao, H.; Tao, X.; Shen, X.; and Jia, J. 2019. Dynamic scene deblurring with parameter selective sharing and nested skip connections. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 3848–3856.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Jiang, Y.; Zhang, Z.; Xue, T.; and Gu, J. 2023. AutoDIR: Automatic All-in-One Image Restoration with Latent Diffusion. *ArXiv*, abs/2310.10123.

Li, B.; Liu, X.; Hu, P.; Wu, Z.; Lv, J.; and Peng, X. 2022. All-In-One Image Restoration for Unknown Corruption. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 17431–17441.

Li, B.; Ren, W.; Fu, D.; Tao, D.; Feng, D.; Zeng, W.; and Wang, Z. 2018. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1): 492–505.

Li, R.; Tan, R. T.; and Cheong, L. F. 2020. All in One Bad Weather Removal Using Architectural Search. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Martin, D.; Fowlkes, C.; Tal, D.; and Malik, J. 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings eighth IEEE international conference on computer vision. ICCV 2001*, volume 2, 416–423. IEEE.

Potlapalli, V.; Zamir, S. W.; Khan, S. S.; and Khan, F. S. 2023. PromptIR: Prompting for All-in-One Blind Image Restoration. *ArXiv*, abs/2306.13090.

Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; Krueger, G.; and Sutskever, I. 2021. Learning Transferable Visual Models From Natural Language Supervision. In *International Conference on Machine Learning*.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, 234–241. Springer.

Tian, C.; Xu, Y.; and Zuo, W. 2020. Image denoising using deep CNN with batch renormalization. *Neural Networks*, 121: 461–473.

Vaswani, A. 2017. Attention is all you need. *Advances in Neural Information Processing Systems*.

Wang, Z.; Cun, X.; Bao, J.; Zhou, W.; Liu, J.; and Li, H. 2022. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 17683–17693.

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*, 1357–1366.

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*, 5728–5739.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; Yang, M.-H.; and Shao, L. 2021. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 14821–14831.

Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; and Zhang, L. 2017a. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7): 3142–3155.

Zhang, K.; Zuo, W.; Gu, S.; and Zhang, L. 2017b. Learning deep CNN denoiser prior for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3929–3938.

Zhang, K.; Zuo, W.; and Zhang, L. 2018. FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Transactions on Image Processing*, 27(9): 4608–4622.

Zhang, Z.; Xu, Y.; Wang, H.; Ni, B.; and Xu, H. 2019. Single-image rain removal via multi-scale cascading image generation. In *2019 IEEE International Conference on Image Processing (ICIP)*, 2771–2775. IEEE.