The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-2/W9-2025
ISPRS Intl. Workshop “Photogrammetric and computer vision techniques for environmental and infraStructure monitoring, Biometrics and Biomedicine”
PSBB25 , 9–11 June 2025, Moscow, Russia

Adaptive Hybrid Transformer Network with Frequency Attention for All-in-One Image Restoration

Cong Thang Pham[1,*], An Hung Nguyen[1], Quoc Cuong Nguyen[1], Minh Nhat Phan[1], Thanh Than Nguyen[1]

[1] The University of Danang - University of Science and Technology, Danang, 550000, Vietnam
{pcthang}@dut.udn.vn, {102210009,102200249, 102210095}@sv1.dut.udn.vn, thanhthan2k1@gmail.com

Commission II, WG II/8

Keywords: Image Restoration, Transformer, Frequency Attention, Hybrid Convolution and Attention, All-in-One model.

Abstract

Image restoration is a critical task in computer vision, aiming to reconstruct high-quality images from degraded inputs caused by environmental factors or sensor limitations. Traditional restoration methods are often designed around prior knowledge of specific degradation types, such as Gaussian noise, rain streaks, or haze. This specificity constrains their flexibility and limits their effectiveness in real-world scenarios where degradations are diverse and unpredictable. To address this limitation, this study proposes a unified image restoration framework capable of handling multiple degradation types without requiring explicit prior knowledge. Specifically, the proposed approach targets three common and challenging degradation scenarios: Gaussian noise, rain, and haze, which are known to exhibit distinct patterns in the frequency domain. The core of the framework is a Hybrid Convolution and Attention (HCA) mechanism. This mechanism integrates the localized feature extraction capability of convolutional neural networks with the global context modeling strength of attention mechanisms, allowing the network to adaptively capture both spatial details and long-range dependencies. Additionally, a Frequency Attention (FA) module is introduced to enhance the model's sensitivity to frequency-domain features. This enables more effective discrimination of degraded image structures and improves restoration accuracy across tasks. To further improve convergence and perceptual quality, the training process is guided by a composite loss function combining Multi-Scale Structural Similarity (MS-SSIM) and l_1 loss. Experimental evaluations conducted on benchmark datasets demonstrate that the proposed method consistently outperforms existing approaches, achieving a PSNR of 34.97 dB and an SSIM of 0.950 when trained jointly across all degradation types. Remarkably, the model attains a PSNR of 39.19 dB and an SSIM of 0.990 on the Dehazing (SOTS) dataset, highlighting its strong generalization and robustness in diverse restoration scenarios.

1. Introduction

In recent years, image restoration and enhancement have emerged as a crucial computer vision research domain, driven by applications in surveillance, medical imaging, and media processing. Advances in deep learning have spurred the development of solutions that overcome the limitations of traditional methods (Zhang et al., 2017, Gao et al., 2016). Real-world images frequently suffer from degradations like noise, rain, haze, or poor lighting, which compromise visual information and can undermine downstream vision systems (Xue el al., 2022). Conventional restoration approaches typically target specific degradation types in isolation, relying on prior knowledge of distortion characteristics. Models like the Image Processing Transformer (IPT) use multi-input, multi-output frameworks for specific scenarios (Chen, 2021), while encoder-decoder architectures often specialize in individual degradation types (Singh et al., 2024). Despite the recent integration of degradation-aware modules, these methods still struggle with complex or mixed degradations, such as rain combined with Gaussian noise.

Traditionally, researchers developed task-specific approaches for distinct degradation categories, such as dehazing (Ren et al., 2018, Pham et al., 2025), deraining (Jiang et al., 2020, Ren et al., 2019), and denoising (Zhang et al., 2017, Zhang et al., 2018). While effective in targeted scenarios, these methods often struggled with complex real-world degradations. Complementing spatial-domain solutions, frequency-domain-based strategies (Cui et al. 2024) emerged to enhance robustness through frequency-specific representations. Recognizing the limitations of specialized models, recent research has shifted toward unified or all-in-one frameworks capable of handling multiple restoration tasks within a single model. Notable contributions include Restormer (Zamir et al., 2022), which leverages a transformer-based architecture with self-attention mechanisms to capture long-range dependencies across various tasks; PromptIR (Potlapalli et al., 2023), which introduced a prompt-based framework for adaptive restoration; and more recent architectures like AdaIR (Cui et al., 2025) and HAIR (Cao et al., 2024) that combine frequency-domain learning with conditional adaptation. Despite these advances, achieving a balanced model that performs well across various degradation types while remaining lightweight and efficient remains challenging.

This study proposes a novel framework, called the Adaptive Hybrid Transformer Network (AdaHybridNet), that eliminates dependency on degradation-type priors, focusing on three representative degradations (Gaussian noise, rain, and haze) by leveraging their distinctive frequency-domain signatures. Inspired by AirNet's contrastive learning approach (Li et al., 2022) and Restormer's multi-level Transformer architecture (Zamir et al., 2022, Liang et al., 2021), AdaHybridNet integrates both convolutional and attention-based modules while incorporating frequency-domain learning for robust feature extraction under diverse image corruption scenarios.

The contributions of this study are threefold, centered around

* Corresponding author: pcthang@dut.udn.vn

Publisher's note: Copernicus Publications has not received any payments from Russian or Belarusian institutions for this paper.
This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-2-W9-2025-213-2025

three key innovations:
–**Hybrid Convolution and Attention (HCA) mechanism**, which combines the local feature extraction strength of convolutional neural networks (CNNs) with the global dependency modeling capabilities of attention mechanisms to optimize the restoration process. –**Frequency Attention (FA) module**, designed to emphasize frequency-domain information, enabling the model to distinguish more effectively between degraded and clean features, thus improving restoration accuracy.
–**Composite loss function**, which integrates the Multi-Scale Structural Similarity (MS-SSIM) loss with the l_1 loss to ensure a balance between perceptual quality and pixel-wise fidelity during the training process.
The proposed method is extensively evaluated on widely used benchmark datasets, including BSD400 (Ai et al., 2024) and BSD68for Gaussian noise removal (Martin et al., 2001), Rain100L for rain streak removal (Yang et al., 2019), and RESIDE-6K for haze removal (Lu et al., 2024). Experimental results demonstrate not only significant improvements in restoration quality but also establish a promising direction for practical deployment in surveillance, healthcare imaging, and other real-world computer vision systems.

2. Proposed Method

This section presents an all-in-one model architecture designed for multi-task image restoration, as illustrated in Figure (1). The proposed model is based on the Restormer (Zamir et al., 2022) architecture, which follows a layered design to progressively learn features at deeper levels and smaller image scales. In this approach, the traditional decoder layers are replaced with Hybrid Convolutional and Attention (HCA) modules and Frequency Attention Modules. These modifications enable the model to effectively learn and apply filters, which are then used to identify and restore the specific pixels that require recovery within the image. The integration of these advanced attention mechanisms allows for improved feature extraction and pixel-level restoration, making the model highly effective for high-resolution image restoration tasks.

2.1 Transformer Encoder

The Transformer Encoder is a fundamental component of the Transformer architecture, designed to process input data and extract meaningful representations through the combined use of the Self-Attention mechanism and Feed-Forward Networks (FFN). This architecture enables the model to capture dependencies and contextual relationships among input elements without being constrained by spatial or temporal distances, making it particularly effective for tasks in natural language processing and computer vision domains (Vaswani et al., 2017, Devlin et al., 2019). The Self-Attention mechanism allows the model to dynamically weight the importance of each input element relative to others, thereby facilitating the learning of both local and global dependencies. Mathematically, the output of the attention mechanism is defined as:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{1}$$

where Q, K, and V represent the query, key, and value matrices, respectively, and d_k is the dimensionality of the key vectors, used for scaling. Following the attention layer, the output is passed through a position-wise Feed-Forward Network (FFN), which further transforms the feature representations through two fully connected layers with a non-linear activation function:

$$\text{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2. \tag{2}$$

This combination of Self-Attention and Feed-Forward computation allows the Transformer Encoder to effectively learn hierarchical and context-aware features, which are essential for handling complex structured data.

2.2 Frequency Attention Module (FAM)

The Frequency Attention Module (FAM) is a deep neural network module that combines both spatial and frequency domain processing to enhance feature extraction. The effectiveness of feature representation in the frequency domain has been previously demonstrated in another study on the Efficient Complex Valued Neural Network with Fourier Transform for Image Denoising (Pham et al., 2022). The input is first processed through a 3×3 convolutional layer to extract local spatial features. Subsequently, the data is transformed into the frequency domain using the Fast Fourier Transform (FFT) (Pham et al., 2022). In the frequency domain, a learnable global filter (Rao et al., 2021) adjusts the information before it is returned to the spatial domain using the Inverse FFT (IFFT). An additional branch of the module uses Global Average Pooling (GAP) to compute the global mean across each channel, followed by two 1×1 convolutional layers that learn weights. The Sigmoid activation function is then applied to generate an attention mask, which helps adjust the channel weights. The GAP calculation is given by:

$$GAP_C = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_{c,i,j}, \tag{3}$$

where $X_{c,i,j}$ is the value at position (i, j) of channel c. The learned attention weight for each channel is computed as:

$$S_c = \sigma\left(W_2 \cdot \text{ReLU}\left(W_1 \cdot GAP_c\right)\right), \tag{4}$$

where W_1 and W_2 are the weights of the two 1×1 convolutional layers, and σ is the Sigmoid activation function. The attention weight is then applied to the features X_c, resulting in the modified feature representation:

$$X'_c = X_c \cdot S_c. \tag{5}$$

These two branches are combined using Channel-Wise Cross Attention (CWCA) (Wang et al., 2022), enabling the module to focus on important channels. Finally, a 1×1 convolutional layer is applied to adjust the output before the data undergoes CWCA processing again. By integrating both spatial and frequency domains, FAM effectively learns important features, thereby improving performance in image restoration tasks at various levels.

2.3 Hybrid Convolutional and Attention (HCA)

Hybrid Convolutional and Attention (HCA) is a model that combines Convolutional Neural Networks (CNN) and the Attention mechanism to simultaneously leverage local feature information and long-range dependencies within the feature space. Initially, the input is processed through a convolutional layer (Conv) to extract local features, and is simultaneously passed through a fully connected (FC) layer to map the features to a higher-dimensional space. The results from these two processes are

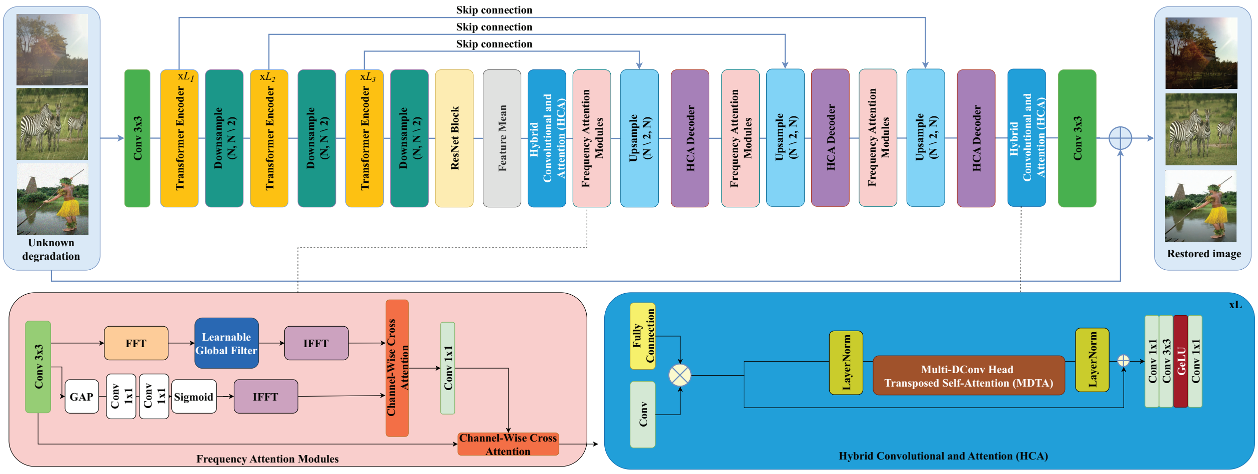

Figure 1. The overall architecture of the proposed AdaHybridNet model.

then combined through element-wise multiplication, enabling the model to effectively capture the correlation between the two feature sets derived from different transformations.

$$F_c = \text{Conv}(X), \quad F_{fc} = \text{FC}(X), \quad F = F_c \odot F_{fc}. \tag{6}$$

The output from this stage is normalized through Layer Normalization (LayerNorm) before being passed into the Multi-DConv Head Transposed Self-Attention (MDTA) (So et al., 2022). MDTA is a variant of Self-Attention where weight-shared convolutions are used to reduce the number of parameters and optimize computational cost. This mechanism allows the model to better capture relationships between pixels in the feature space without significantly increasing the computational complexity.

$$F_{att} = \text{MDTA}(\text{LayerNorm}(F)). \tag{7}$$

After passing through MDTA, the features are once again normalized via LayerNorm before being processed through a Convolutional Feed-Forward Network (CFFN). CFFN includes a series of transformations, such as a 1×1 convolution (to adjust the number of feature channels), a 3×3 convolution (to extract deeper local features), a non-linear activation function (GELU), and a final 1×1 convolution to re-normalize the feature space.

$$F_{ffn} = \text{GELU}(\text{Conv}_{3\times3}(\text{Conv}_{1\times1}(\text{LayerNorm}(F_{att})))), \tag{8}$$

$$F_{cffn} = \text{Conv}_{1\times1}(F_{ffn}). \tag{9}$$

The output of the CFFN is combined with the output of MDTA through a residual connection, helping to mitigate information loss during the backward pass and optimize the training process.

$$F_{HCA} = F_{att} + F_{cffn}. \tag{10}$$

The HCA model leverages the advantages of CNNs for local feature extraction and the attention mechanism for capturing global relationships between pixels. This architecture improves performance in image processing tasks, especially those requiring the model to combine information from multiple spatial ranges simultaneously.

3. Experiment Results

3.1 Experiment Setup

The proposed model was systematically evaluated across three distinct image degradation tasks: Gaussian noise removal, rain streak removal, and haze removal. For the Gaussian denoising task, synthetic noise levels of $\phi = \{15, 25, 50\}$ were applied to the BSD400 dataset (Ai et al., 2024) for model training, while the BSD68 dataset (Martin et al., 2001) was used for testing and performance evaluation. For the deraining task, the Rain100L dataset (Yang et al., 2019), consisting of 200 paired images, was employed for training, and a separate set of 100 images was used for testing. Regarding dehazing, a compact subset named RESIDE-6K (Lu et al., 2024), extracted from the larger RESIDE benchmark, was adopted to reduce computational cost and training time. The model's generalization capability was further assessed on the SOTS benchmark dataset (Li et al., 2018).

All experiments were conducted over 100 training epochs, with input images resized to a fixed resolution of 256×256 pixels. The optimization process was carried out using the AdamW optimizer, configured with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and an initial learning rate set to 1×10^{-4}. To ensure alignment with human visual perception, the training objective combined Multi-Scale SSIM (MS-SSIM) and l_1 loss as the loss function, formalized as:

$$L_{Composite}(y, \hat{y}) = \alpha \cdot \text{MS-SSIM}(y, \hat{y})) + \beta \cdot \|y - \hat{y}\|_1.$$

where y and $\hat{y}$ denote the ground-truth and predicted images, respectively. The parameters $\alpha = 0.84$ and $\beta = 0.16$ were empirically selected based on prior studies investigating the influence of loss function weighting on image restoration performance (Zhao et al., 2017).

The quantitative performance of the model was evaluated using two widely adopted metrics: Structural Similarity Index (SSIM) (Wang et al., 2004) and Peak Signal-to-Noise Ratio (PSNR) (Gonzalez et al., 2022). Model training was conducted on an NVIDIA RTX 4090 GPU equipped with 24 GB of VRAM. All experiments were implemented using PyTorch version 2.4.1 and executed on the CUDA 12.3 platform. Inference and validation were performed on a workstation featuring an Intel Xeon E5-2676 processor, 32 GB RAM, and an NVIDIA RTX 2060 Super GPU with 8 GB of VRAM, operating under Ubuntu Desktop 22.04.

Table 1. Comparison of model performance on combined rain, noise, hazy datasets

Model	SOTS (Dehazing)	BSD68 (Denoising) $\phi = 15$	BSD68 (Denoising) $\phi = 25$	BSD68 (Denoising) $\phi = 50$	Rain100L (Deraining)	Average
BRDNet (Tian et al., 2020)	23.23/ 0.895	32.26 / 0.898	29.76 / 0.836	26.34 / 0.693	27.42 / 0.895	27.80 / 0.843
LPNET (Gao et al., 2019)	20.84/ 0.828	26.47 / 0.778	24.77 / 0.748	21.26 / 0.552	24.88 / 0.784	23.64 / 0.738
FDGAN (Dong et al., 2020)	24.71/ 0.929	30.25 / 0.910	28.81 / 0.868	26.43 / 0.776	29.89 / 0.933	28.02 / 0.883
MPRNet (Zamir et al., 2021)	25.28/ 0.955	33.54 / 0.927	30.89 / 0.880	27.56 / 0.779	33.57 / 0.954	30.17 / 0.899
DL (Fan et al., 2019)	26.92/0.931	33.05 / 0.914	30.41 / 0.861	26.90/ 0.740	32.62 / 0.931	29.98 / 0.876
AirNet (Li et al., 2022)	27.94/0.962	33.92 / 0.933	31.26 / 0.888	28.00/0.797	34.90 / 0.968	31.20 / 0.910
Restormer (Zamir et al., 2022)	27.78/0.958	33.72 / 0.930	30.67 / 0.865	27.63 / 0.792	33.78 / 0.958	30.75 / 0.901
PromptIR (Potlapalli et al., 2023)	30.58/ 0.974	33.98 /0.933	31.31 / 0.888	28.06 / 0.799	36.37 / 0.972	32.06 / 0.913
AdaIR (Cui et al., 2025)	31.06/ 0.980	34.12 / 0.935	31.45 /0.892	28.19 /0.802	**38.64 /0.983**	32.69 / 0.918
HAIR (Cao et al., 2024)	30.98/ 0.979	34.16 / 0.935	31.51 /0.892	28.24 /0.803	38.59 /0.983	32.70 / 0.919
AdaHybridNet (ours)	**39.19/ 0.990**	**35.46 / 0.959**	**33.26 /0.938**	**30.25 /0.893**	36.67 /0.972	**34.97 / 0.950**

Figure 2. Qualitative comparison of the dehazing performance of the proposed method against existing approaches.

Figure 3. Qualitative comparison of the deraining performance of the proposed method against existing approaches.

3.2 Experiment Results

The performance of the proposed **AdaHybridNet** was comprehensively evaluated against several state-of-the-art image restoration models across three representative degradation tasks: haze removal, Gaussian denoising, and rain streak removal. As reported in Table (1), AdaHybridNet consistently outperformed all baseline methods across both quantitative metrics PSNR and SSIM.

For the *dehazing* task on the SOTS dataset, our model achieved a PSNR of 39.19 and an SSIM of 0.990, substantially surpassing prior methods, including Restormer (Zamir et al., 2022) (27.78 dB/ 0.958) and AdaIR (Cui et al., 2025) (31.06 dB/ 0.980). In the *Gaussian denoising* task evaluated on the BSD68 dataset at noise levels $\phi = \{15, 25, 50\}$, AdaHybridNet achieved PSNR/SSIM scores of 35.46 dB/0.959, 33.26 dB/0.938, and 30.25 dB/0.893, respectively, outperforming recent competitive models such as PromptIR (Potlapalli et al., 2023) and HAIR (Cao et al., 2024).

For the *deraining* task on the Rain100L dataset, the proposed AdaHybridNet achieved a PSNR of 36.67 dB and an SSIM of 0.972. Although the result is competitive and comparable to state-of-the-art methods, it slightly underperforms the highest PSNR score of 38.64 dB reported by AdaIR (Cui et al., 2025). This performance gap can be attributed to the distinct frequency-domain distribution characteristics of rain streak artifacts, which differ considerably from those of Gaussian noise and haze. Since AdaHybridNet leverages frequency-domain learning combined with complex-valued feature representations, the irregular and non-uniform spectral distribution of rain streaks presents a greater challenge during model optimization, thereby limiting its capacity to fully suppress this type of degradation. Nevertheless, the model still demonstrates strong generalization ability across diverse restoration tasks.

On average across all tasks, AdaHybridNet achieved a PSNR of 34.97 and an SSIM of 0.950, marking a significant improvement over existing approaches. These results validate the effectiveness of combining convolutional and attention mechanisms along with adaptive feature learning in enhancing the model's generalization across multiple image restoration scenarios.

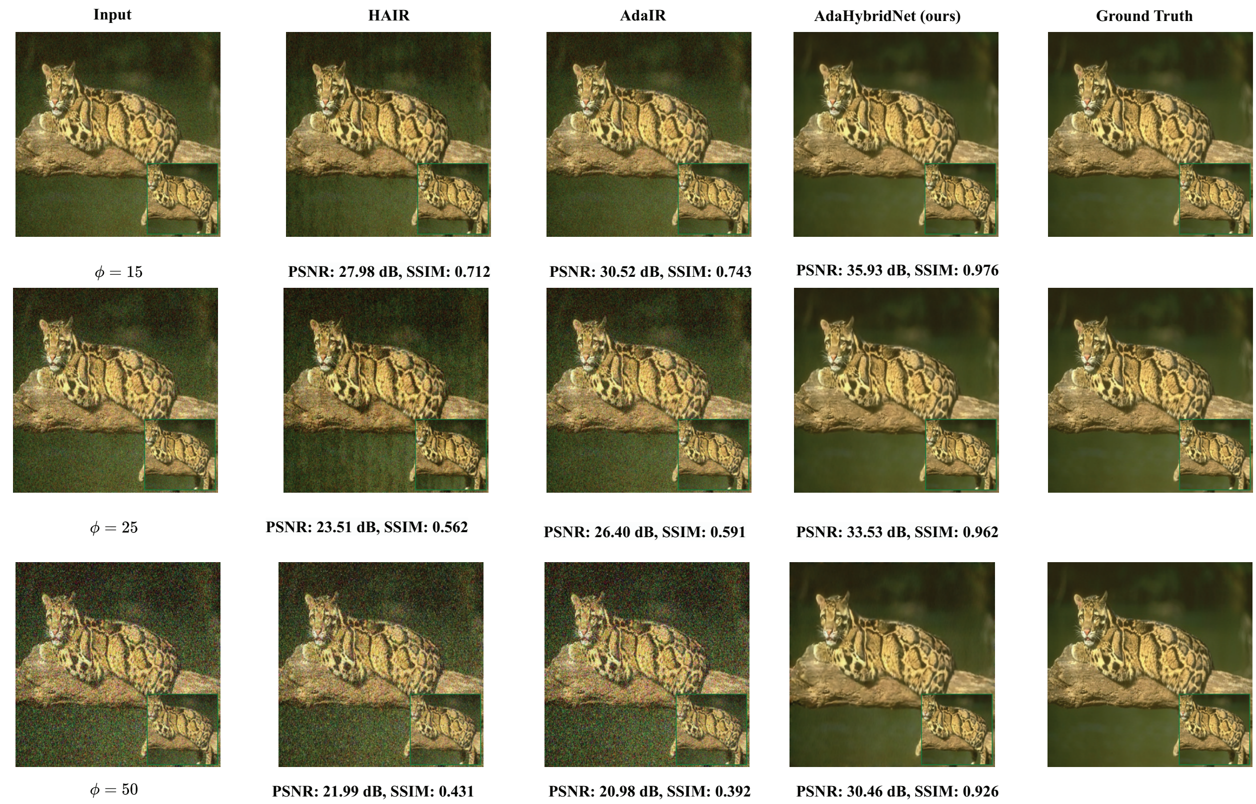

Figure 4. Qualitative comparison of denoising performance on Gaussian noise with three levels ($\phi = 15, 25, 50$) using the proposed method and existing approaches.

In addition, the visual results presented in Figures (2), (3), and (4) further validate the quantitative findings, highlighting the effectiveness of the proposed all-in-one **AdaHybridNet** model in addressing three distinct image restoration tasks: dehazing, deraining, and denoising. Across all scenarios, the model consistently demonstrates superior visual quality compared to other baselines. Notably, in the denoising task illustrated in Figure (4), our proposal achieves restored images that closely approximate the groundtruth references, with PSNR values reaching 38.08 dB, 36.17 dB, and 33.39 dB at noise levels $\phi = 15, 25,$ and 50, respectively. These results clearly outperform two of the most recently introduced models, HAIR (Cao et al., 2024) and AdaIR (Cui et al., 2025), confirming both the robustness and generalization capability of the proposed framework under varying degradation conditions.

4. Conclusion

In this study, a novel multi-task learning framework was proposed to address various image degradation problems, including Gaussian noise removal, rain streak removal, and haze removal. Through extensive experiments, the model demonstrated strong generalization capability and robustness across diverse degradation scenarios. The proposed approach leverages the complementary strengths of convolutional feature extraction and attention-based global context modeling, while simultaneously incorporating degradation characteristics in the frequency domain. This combination enables the model to effectively restore fine-grained details and preserve global structural information, thereby improving the visual quality of degraded images.

The promising experimental results highlight the potential of the proposed architecture as a flexible and scalable solution for image restoration tasks in real-world applications. Future research will focus on extending this framework to handle more complex degradations and exploring its applicability in related domains such as video restoration and medical image enhancement.

References

Ai, Y., Huang, H., Zhou, X., Wang, J. & He, R., 2024. Multimodal prompt perceiver: Empower adaptiveness, generalizability and fidelity for all-in-one image restoration. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* , 25432-25444.

Cao, J., Cao, Y., Pang, L., Meng, D. & Cao, X. , 2024. HAIR: Hypernetworks-based All-in-One Image Restoration. *ArXiv Preprint ArXiv:2408.08091*, 1-6.

Chen, H. et al., 2021. Pre-Trained Image Processing Transformer.*IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12299–12310.

Cui, Y., Zamir, S., Khan, S., Knoll, A., Shah, M. & Khan, F., 2025. AdaIR: Adaptive All-in-One Image Restoration via Frequency Mining and Modulation. *The Thirteenth International Conference On Learning Representations*, 1-28.

Devlin, J., Chang, M., Lee, K. & Toutanova, K., 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *The North American Chapter Of The Association For Computational Linguistics*. 4171-4186.

Dong, Y., Liu, Y., Zhang, H., Chen, S. & Qiao, Y., 2020. FD-GAN: Generative adversarial networks with fusion-discriminator for single image dehazing. *The AAAI Conference On Artificial Intelligence*, 10729-10736.

Fan, Q., Chen, D., Yuan, L., Hua, G., Yu, N. & Chen, B., 2019. A general decoupled learning framework for parameterized image operators. *IEEE Transactions On Pattern Analysis And Machine Intelligence*. **43**, 33-47.

Gao, R. & Grauman, K., 2016. On-Demand Learning for Deep Image Restoration. *IEEE International Conference on Computer Vision* , 1095-1104.

Gao, H., Tao, X., Shen, X. & Jia, J., 2019. Dynamic scene deblurring with parameter selective sharing and nested skip connections. *IEEE Conference On Computer Vision And Pattern Recognition (CVPR)*, 3843-3851.

Gonzalez, R. & Woods, R., 2002. *Digital Image Processing*. Prentice Hall,2002), 793 pages.

Jiang, K. et al., 2020. Multi-Scale Progressive Fusion Network for Single Image Deraining. *IEEE/CVF Conference On Computer Vision And Pattern Recognition (CVPR)*, 8346-8355.

Zhang, K., Zuo, W., Chen, Y., Meng, D. & Zhang, L., 2017. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Transactions On Image Processing*. **26**, 3142-3155.

Li, B. et al. Benchmarking single-image dehazing and beyond. *IEEE Transactions On Image Processing*, **28** (1), 492-505

Li, B., Liu, X., Hu, P., Wu, Z., Lv, J. & Peng, X., 2022. All-in-one image restoration for unknown corruption. *IEEE Conference On Computer Vision And Pattern Recognition (CVPR)*, 17431-17441.

Liang, J. et al., 2021. SwinIR: Image Restoration Using Swin Transformer. *International Conference on Computer Vision Workshops* , 1833-1844.

Lu, L., Xiong, Q., Chu, D. & Xu, B., 2024. MixDehazeNet: Mix Structure Block for Image Dehazing Network. *International Joint Conference on Neural Networks*, 1-10.

Martin, D., Fowlkes, C., Tal, D. & Malik, J., 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. *IEEE International Conference On Computer Vision*, textbf2, 416-423

Pham C.T. et al., 2025. DehazeCLNet: A Contrastive Learning Framework with Advanced Feature Extraction for Image Dehazing. Information and Communication Technology. SOICT 2024. Communications in Computer and Information Science, vol 2350. Springer, 275-286.

Pham M. T. et al., 2022. Efficient Complex Valued Neural Network with Fourier Transform on Image Denoising. *The 5th International Conference On Future Networks And Distributed Systems*, 48-57.

Potlapalli, V., Zamir, S., Khan, S. & Khan, F. PromptIR: PromptIR: prompting for all-in-one blind image restoration. *the 37th International Conference on Neural Information Processing System* , 71275-71293.

Rao Y. & Zhou, J. Global Filter Networks for Image Classification. *the 35th International Conference on Neural Information Processing Systems* , 980-993.

Ren, D., Zuo, W., Hu, Q., Zhu, P. & Meng, D., 2019. Progressive Image Deraining Networks: A Better and Simpler Baseline. *The IEEE/CVF Conference On Computer Vision And Pattern Recognition*, 3932-3941.

Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X. & Yang, M., 2018. Gated Fusion Network for Single Image Dehazing. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3253-3261.

Singh K. et al., 2024. LightClearNet: Lightweight Encoder-Decoder Architecture for Single Image Dehazing. *International Conference on Communication, Computer Sciences and Engineering* , 1824-1828,

So D., Manke, W., Liu, H., Dai, Z., Shazeer, N. & Le, Q. Primer: Searching for Efficient Transformers for Language Modeling. *the 35th International Conference on Neural Information Processing Systems* , 6010 – 6022.

Tian, C., Xu, Y. & Zuo, W., 2020. Image denoising using deep CNN with batch renormalization. *Neural Networks*, **121**, 461-473.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L. & Polosukhin, I., 2017. Attention Is All You Need. *the 31st International Conference on Neural Information Processing Systems*, 6000 – 6010.

Wang, H., Cao, P., Wang, J. & Zaiane, O. UCTransNet: Rethinking the Skip Connections in U-Net from a Channel-wise Perspective with Transformer. *The AAAI Conference on Artificial Intelligence*, **36** (3), 2441-2449.

Wang, Z., Bovik, A., Sheikh, H. & Simoncelli, E., 2004. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions On Image Processing*, **13**, 600-612.

Cui Y., Ren W., Cao X. and Knoll A., 2024. Image Restoration via Frequency Selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **46** (2), 1093-1108.

Xue, Z., Angara, S., Levitz, D., & Antani, S., 2022. Analysis of digital noise reduction methods on classifiers used in automated visual evaluation. *Proceedings of SPIE - The International Society for Optical Engineering*, 1195008.

Yang, W., Tan, R., Feng, J., Guo, Z., Yan, S. & Liu, J., 2019. Joint rain detection and removal from a single image with contextualized deep networks. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. **42**(6),1377-1393.

Zamir, S., Arora, A., Khan, S., Hayat, M., Khan, F., Yang, M. & Shao, L., 2021. Multi-stage progressive image restoration. *The IEEE Conference On Computer Vision And Pattern Recognition*, 14816-14826.

Zamir, S., Arora, A., Khan, S., Hayat, M., Khan, F. & Yang, M. Restormer: Efficient transformer for high-resolution image restoration. *The IEEE Conference On Computer Vision And Pattern Recognition*, 5718-5729.

Zhang, K., Zuo, W. & Zhang, L. 2018. FFDNet: Toward a Fast and Flexible Solution for CNN-based Image Denoising. *arXiv:1710.04026* , 1-15.

Zhao, H., Gallo, O., Frosio, I. & Kautz, J., 2017. Loss Functions for Image Restoration With Neural Networks. *IEEE Transactions On Computational Imaging*, **3**, 47-57.