NTIRE 2025 Image Denoising ($\sigma = 50$) Challenge Factsheet -Enhanced Blind Image Restoration with Channel Attention Transformers and Multi-Scale Attention Prompt Learning-

Jianhua Hu
Macao Polytechnic University
Macao, 999078, China
p2412994@mpu.edu.mo

K. L. Eddie Law
Macao Polytechnic University
Macao, 999078, China
eddielaw@mpu.edu.mo

1. Introduction

Image restoration technology aims to reconstruct high-quality (HQ) clear images from low-quality (LQ) degraded images. Its applications are extremely extensive, including image denoising, deblurring, deraining, and removal of compression artifacts. However, due to the complexity and irreversibility of the image degradation process, accurately restoring lost details and textures from LQ images is highly challenging. In particular, during the degradation process, some image information may be completely lost, making the restoration task even more difficult.

In recent years, with the development of deep learning technologies, convolutional neural networks (CNNs) have made significant progress in the field of image restoration, thanks to their powerful feature extraction capabilities. CNNs can learn the nonlinear mapping relationship between low-quality and high-quality images, providing an effective solution for image restoration. However, CNNs have limitations in handling long-range dependencies, which restricts their performance in image restoration tasks.

In recent years, the Transformer architecture[3] has achieved great success in the field of natural language processing with its multi-head self-attention (Self-Attention, SA) mechanism and has gradually been introduced into the field of image processing [18] [17] [6][10]. The Transformer can model long-range dependencies, offering new ideas for image restoration. Many researchers have begun to explore using the Transformer architecture for image restoration and have achieved good results. For example, in the fields of image denoising and removal of compression artifacts, the Transformer architecture has demonstrated strong capabilities. However, existing Transformer-based image restoration methods mostly focus on the extraction of high-frequency information while neglecting the importance of low-frequency information. This leads to models being unable to fully capture the overall structure and semantic information of images when dealing with image restoration tasks.

Moreover, existing blind image restoration methods have certain limitations when dealing with different types of image degradations. For example, different noise coefficients (such as the value of the noise level σ) and image compression quality factors (such as JPEG compression quality factors) can lead to different degrees of image degradation. Therefore, training a separate model for each specific type of degradation is impractical, as it not only increases the storage burden of models but also significantly raises training time and complexity. To overcome these challenges, many researchers have begun to explore blind image restoration methods, that is, restoring images without knowing the specific degradation parameters. For example, the PromptCIR [7] method in 2024 achieved blind compressed image restoration through Prompt learning, and the PromptIR [12] method in 2023 achieved full-function blind image restoration through Prompt technology. These methods, by introducing the Prompt mechanism, enable models to better adapt to different types of image degradations and improve the generalization ability of models.

Nevertheless, existing Prompt-based image restoration methods still have significant limitations in feature extraction and multi-dimensional attention mechanisms. On the one hand, the Transformer architecture, with its powerful self-attention mechanism, can efficiently capture high-frequency information in image restoration tasks, such as details and textures. However, this mechanism often focuses excessively on high-frequency information while neglecting the extraction of low-frequency information, such as the overall structure and semantic information of images. This imbalance can lead to models performing inadequately when restoring the overall structure of images, thereby affecting the integrity and accuracy of the restoration results. On the other hand, although existing Prompt learning methods can adapt to different types of image degradations through the Prompt mechanism, they still have deficiencies in multi-scale feature extraction and multi-dimensional attention allocation. Specifically, these methods find it difficult to take into account both the details and the overall structure of images simultaneously, leading to limitations in the generalization ability and restoration accuracy of models when dealing with complex blind image restoration tasks.

To overcome these limitations, this paper proposes a blind image restoration method based on Channel Attention Transformers and Efficient Multi-Scale Attention Prompt learning. Our method is named CTMP. Specifically, we have improved the Transformer module by designing a Transformer architecture that integrates Channel Attention with the self-attention mechanism, combining the strengths of both Transformer and Channel Attention. The Transformer focuses on extracting high-frequency information, capturing the details and textures of images; while Channel Attention excels at capturing low-frequency information, extracting the overall structure and semantic information of images. Through this integrated design, the model can efficiently extract both high-frequency and low-frequency information from images, thereby overcoming the shortcomings of existing Transformer architectures in low-frequency information extraction.

Furthermore, to more efficiently train a model to achieve blind image restoration for various strengths of image noise and different qualities of compressed images, we have improved the existing Prompt module by introducing an Efficient Multi-Scale Attention module (EMA). The EMA module, by combining Channel Attention and spatial attention mechanisms, further enhances the model's ability to represent image features. The EMA module reshapes some channels to the batch dimension and divides the channel dimension into multiple sub-features, which can better distribute spatial semantic features and thus improve the model's ability to capture features at different scales. This enables the model to more flexibly adapt to different types of degradations when dealing with complex image restoration tasks, while also making up for the deficiencies of existing Prompt learning methods in multi-scale feature extraction.

The main contributions of this paper include:

1. **Improving the Transformer module by integrating Channel Attention mechanism**
   This study improves the Transformer module by introducing Channel Attention and integrating it with the self-attention mechanism of Transformer. The Transformer, with its multi-head self-attention mechanism, excels at capturing high-frequency information in images (such as details and textures), while Channel Attention is adept at extracting low-frequency information (such as overall structure and semantic information). Through this integrated mechanism, the model can efficiently extract both high-frequency and low-frequency information, thereby achieving a more comprehensive feature representation in image restoration tasks.
2. **Designing a Prompt module based on Efficient Multi-Scale Attention (EMA)**
   To achieve blind image restoration for various strengths of image noise and different qualities of compressed images, this paper designs a Prompt module based on Efficient Multi-Scale Attention (EMA). The EMA module combines Channel Attention and spatial attention mechanisms, effectively extracting multi-scale features from images. Moreover, by reshaping some channels to the batch dimension and dividing the channel dimension into multiple sub-features, the EMA module can better distribute spatial semantic features, thereby enhancing the model's adaptability to different types of degradations. This design enables the model to more flexibly cope with different types of image degradations when dealing with blind image restoration tasks.
3. **Unified experimental design for image denoising and compression artifact removal**
   This paper designs a unified experimental scheme for the fields of image denoising and compression artifact removal, enabling a comparison of different blind image restoration methods. Experimental results show that the proposed CTMP method demonstrates excellent performance in various image restoration tasks, especially in blind image restoration, where it can better adapt to different types of image degradations. This provides a new solution for the field of image restoration.

2. Team details

- mpu_ai
- Jianhua
- Macao,China, 15875602058,p2412994@mpu.edu.mo
- K. L. Eddie Law
- Macao Polytechnic University
- bvngh3247 Enhanced Blind Image Restoration with Channel Attention Transformers and Multi-Scale Attention Prompt Learning **CTMP**
- We tested the DVIK validation results, and the PSNR value was 29.06.
- https://github.com/gdit-ai/CTMP

3. Method

Existing deep learning-based image restoration methods exhibit inadequate generalization capabilities when faced with a variety of noise types and intensities, thereby significantly impeding their broad application in real-world scenarios. To tackle this challenge, this paper proposes a novel prompt-based learning approach, namely Blind Image Restoration Using Dual-Channel Transformers and Multi-Scale Attention Prompt Learning (CTMP), as depicted in Figure 1. The CTMP model features a U-shaped architecture grounded in the Transformer framework, constructed from the enhanced Channel Attention Transformer Block (CATB). During the image restoration process, CTMP adopts a blind image restoration strategy to address diverse noise types and intensities. It integrates an Efficient Multi-Scale Attention Prompt Module (EMAPM) that is based on prompts. Within the EMAPM, an Enhanced Multi-scale Attention (EMA) module is specifically designed. This module extracts global information across different directions and employs dynamic weight calculations to adaptively modulate the importance of features at various scales. The EMA module subsequently fuses the enhanced multi-scale features with the input feature maps, yielding a more enriched feature representation. This fusion mechanism empowers the model to more effectively capture and leverage features at different scales, thereby markedly bolstering its capacity to restore image degradations and showcasing superior generalization capabilities.

3.1. Transformer Block Incorporating Channel Attention and Residual Connections

The Transformer Block serves as the cornerstone of our entire model, harnessing the Transformer architecture to extract image features through the self-attention mechanism. In pursuit of enhanced performance, we have refined the Transformer module by devising a novel architecture that integrates Channel Attention with the self-attention mechanism, thereby combining the strengths of both Transformer and Channel Attention. Specifically, the Transformer focuses on extracting high-frequency information to capture the fine details and textures of images, while Channel Attention excels at capturing low-frequency information to extract the overall structure and semantic information of images. This integration further boosts the image denoising effect.As depicted in Figure 2, the improved Transformer architecture, named the Channel Attention Transformer Block (CATB), primarily consists of the following three modules: Multi-DConv Head Transposed Self-Attention (MDTA), Channel Attention (CA), and Gated-Dconv Feed-Forward Network (GDFN).

The Multi-DConv Head Transposed Self-Attention (MDTA) module enhances the self-attention mechanism's perception of local image features by incorporating multi-scale depthwise convolution operations, effectively capturing detailed image information. The Channel Attention (CA) module, dedicated to information processing along the channel dimension, computes the importance weights of each channel to perform weighted fusion of channel features, thereby strengthening the model's perception of the overall image structure. The Gated-Dconv Feed-Forward Network (GDFN) module combines the gating mechanism with depthwise convolution operations, aiming to further optimize the nonlinear transformation of features. By introducing the gating mechanism, the model can adaptively adjust the transmission and updating of features based on the dynamic characteristics of the input features, thereby enhancing the flexibility and adaptability of feature representation. Through the synergistic action of these three modules, the improved Transformer architecture can more effectively handle both high-frequency and low-frequency information in images, thereby significantly enhancing the performance of image denoising and restoration. In image restoration tasks, feature extraction and representation are crucial steps. Traditional convolutional neural networks (CNNs) and Transformer architectures primarily focus on feature extraction in the spatial domain, while paying less attention to the weighting of features in the channel dimension. To address this limitation, we introduce a Channel Attention module in the Transformer Block, creating a Transformer Block that incorporates Channel Attention and Residual Connections. This module weights the channel dimension through global average pooling and fully connected layers, enhancing important channel features while suppressing less important ones. This weighting mechanism enables the model to focus more effectively on key information, thereby improving the quality of restored images. Additionally, the introduction of residual connections further enhances the model's robustness and performance. Residual connections ensure that the information of the input features is fully retained after processing by the Channel Attention module by adding the input features directly to the output features. This design not only aids gradient propagation but also retains the original information of the input features when the weighting effect of the Channel Attention module is suboptimal, further boosting the model's robustness.

The Channel Attention Module aims to enhance important channel features and suppress less important ones by leveraging global average pooling and fully connected layers. The specific implementation steps are as follows:

1. **Global Average Pooling (GAP)**:

$$\begin{gathered}\text{avg\_pool}(X) \in \mathbb{R}^{B\times C\times 1\times 1}\\ \text{avg\_pool}(X) \rightarrow \mathbb{R}^{B\times C}\end{gathered}$$

2. **Fully Connected Layers (FC)**:

$$\begin{aligned}\text{fc1}(\text{avg\_pool}(X)) &\in \mathbb{R}^{B\times(C//\text{reduction})}\\ \text{ReLU}(\text{fc1}(\text{avg\_pool}(X))) &\in \mathbb{R}^{B\times(C//\text{reduction})}\\ \text{fc2}(\text{ReLU}(\text{fc1}(\text{avg\_pool}(X)))) &\in \mathbb{R}^{B\times C}\\ \text{Sigmoid}(\text{fc2}(\text{ReLU}(\text{fc1}(\text{avg\_pool}(X))))) &\in \mathbb{R}^{B\times C}\end{aligned}$$

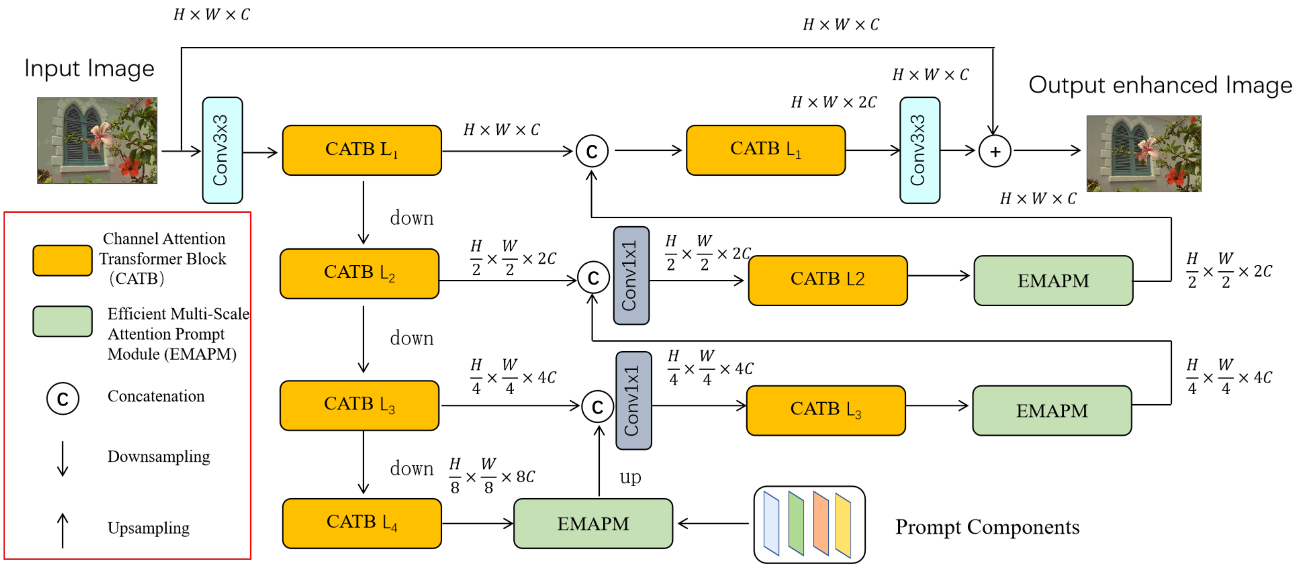

Figure 1. Structure of the CTMP Model

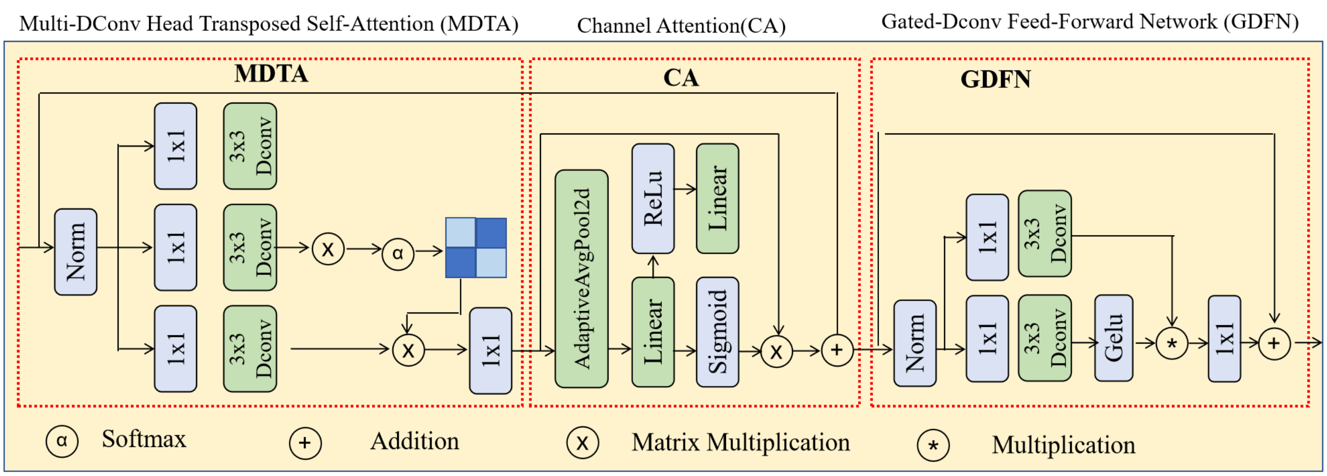

Figure 2. Structure of Channel Attention Transformer Block (CATB)

3. **Channel Weighting**:

$$\text{avg\_out} \in \mathbb{R}^{B\times C\times H\times W}$$
$$X_{\text{enhanced}} = X \odot \text{avg\_out}$$

The Transformer Block integrates the Channel Attention Module and residual connections to enhance feature extraction and representation. The specific implementation is as follows:

1. **Normalization and Self-Attention**:

$$x_{\text{norm1}} = \text{LayerNorm}(x)$$
$$\text{attn\_out} = \text{Attention}(x_{\text{norm1}})$$

2. **Channel Attention**:

$$\text{attn\_out} = \text{ChannelAttention}(\text{attn\_out})$$

3. **Residual Connection**:

$$x = x + \text{attn\_out}$$

4. **Feed-Forward Network (FFN)**:

$$x_{\text{norm2}} = \text{LayerNorm}(x)$$
$$\text{ffn\_out} = \text{FeedForward}(x_{\text{norm2}})$$
$$\text{ffn\_out} = \text{LeakyReLU}(\text{ffn\_out}, \text{negative\_slope} = 0.2)$$

5. **Residual Connection**:

$$x = x + \text{ffn\_out}$$

The proposed model incorporates several key enhancements to improve image restoration quality. Firstly, the Channel Attention Module leverages global average pooling and fully connected layers to selectively enhance important channel features while suppressing less relevant ones. This mechanism enables the model to focus more effectively on critical information, thereby improving the quality of the restored image. Secondly, residual connections are employed to ensure that the original input features are fully retained and added directly to the output features after processing by the Channel Attention Module. This not only aids gradient propagation but also preserves the original information when the weighting effect is suboptimal, thus boosting the model's robustness. Lastly, the LeakyReLU activation function is utilized in the Feed-Forward Network to introduce non-linearity while avoiding the "dying neurons" issue associated with ReLU, further enhancing the model's expressive power. Together, these improvements contribute to a more effective and robust image restoration model.

3.2. Efficient Multi-Scale Attention Prompt Module

Addressing multi-scale image degradations is a crucial challenge in image restoration tasks. Traditional feature extraction methods typically capture features at a single scale, neglecting the fusion and interaction of features across multiple scales. To overcome this limitation, we propose a prompt-based blind image restoration approach, incorporating an Efficient Multi-Scale Attention Prompt Module (EMAPM). The core of the EMAPM is the Enhanced Multi-scale Attention (EMA) module, which extracts global information in different directions and combines dynamic weight calculations to adaptively adjust the significance of features at various scales, thereby generating a richer feature representation. This design not only enhances the model's adaptability to multi-scale image degradations but also strengthens the expressiveness of features, significantly improving the quality of image restoration. The introduction of the EMA module represents a significant innovation in our image restoration approach. Experimental results validate the effectiveness of the EMA module, demonstrating its ability to substantially boost model performance across multiple image restoration tasks. This innovation not only enhances the model's restoration capabilities but also offers new research directions for image restoration tasks.

The Efficient Multi-Scale Attention Prompt Module (EMAPM) is designed to enhance the model's ability to capture multi-scale features in image restoration tasks. By generating adaptive prompts that focus on different scales and characteristics of the input image, EMAPM allows the model to better handle various types of image degradations. The core components and operations of EMAPM are described as follows:

Module Configuration: To configure the EMAPM, several key parameters are defined:

- **Prompt Dimension** (d_p): This determines the dimension of each prompt vector, which represents the feature space for each prompt.
- **Prompt Length** (L_p): This specifies the number of prompt vectors, which controls the diversity of prompts generated.
- **Prompt Size** (S_p): This sets the spatial size of each prompt vector, which affects the resolution of the prompts.
- **Linear Dimension** (d_l): This is the dimension of the input to the linear layer, which processes the embedding of the input feature map.
- **Factor** (f): This defines the number of groups in the EMA module, which influences the grouping mechanism in the attention process.

Mathematical Formulation: Given an input feature map $x \in \mathbb{R}^{B\times C\times H\times W}$, where B is the batch size, C is the number of channels, and $H \times W$ is the spatial dimension, the operations within EMAPM are defined as follows:

1. **Compute Embedding:** The embedding of the input feature map is computed by averaging the spatial dimensions.

$$\text{emb} = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} x_{:,:,i,j} \in \mathbb{R}^{B\times C} \tag{1}$$

2. **Linear Layer and Softmax:** The embedding is passed through a linear layer followed by a softmax function to generate prompt weights.

$$\text{prompt\_weights} = \text{softmax}(\text{linear\_layer}(\text{emb})) \in \mathbb{R}^{B\times L_p} \tag{2}$$

3. **Generate Prompt:** The prompts are generated by weighting the prompt parameters with the prompt weights and then summing them up. The prompts are then interpolated to match the spatial dimensions of the input feature map.

$$\text{prompt} = \sum_{k=1}^{L_p} \text{prompt\_weights}_{:,k} \cdot \text{prompt\_param}_k \in \mathbb{R}^{B\times d_p\times S_p\times S_p} \tag{3}$$

$$\text{prompt} = \text{F.interpolate}(\text{prompt}, (H, W), \text{mode} = \text{"bilinear"}) \tag{4}$$

4. **Enhance Prompt using EMA:** The prompts are enhanced using the Enhanced Multi-scale Attention (EMA) module, which refines the prompts by incorporating multi-scale attention.

$$\text{enhanced\_prompt} = \text{EMA}(\text{prompt}) \in \mathbb{R}^{B\times d_p\times H\times W} \tag{5}$$

5. **Conv3x3:** Finally, the enhanced prompts are processed through a 3x3 convolutional layer to further refine

the feature representation.

$$\text{enhanced\_prompt} = \text{conv3x3}(\text{enhanced\_prompt}) \in \mathbb{R}^{B \times d_p \times H \times W} \tag{6}$$

4. Experiments

In this section, we conducted a series of extensive experiments to comprehensively demonstrate the superior performance of the proposed CTMP model across multiple datasets and benchmarks. The experiments covered a variety of tasks, including denoising and deblocking of compressed images, and were compared with previous state-of-the-art methods. Additionally, we reported the results of ablation studies, which strongly validated the effectiveness of the Channel Attention Transformer Block (CATB) and the Enhanced Multi-scale Attention Prompt Module (EMAPM) within the CTMP architecture.

The CTMP framework is end-to-end trainable without the need for pretraining any individual components. Its architecture consists of a 4-level encoder-decoder, with each level equipped with a different number of Transformer modules, specifically [4, 6, 6, 8] from level 1 to level 4. We placed a Prompt module between every two consecutive decoder levels, resulting in a total of 3 Prompt modules across the entire PromptIR network, with a total of 5 Prompt components. During training, the model was trained with a batch size of 2, leveraging the computational power of a Tesla T4 GPU. The network was optimized through L1 loss, using the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$) with a learning rate of 2×10^{-4}. To further enhance the model's generalization ability, we used 128×128 cropped blocks as input during training and augmented the training data by applying random horizontal and vertical flips to the input images.

4.1. Dataset

To comprehensively evaluate the performance of the CTMP algorithm in image restoration tasks, we conducted experiments in two critical areas: image denoising and deblocking of compressed images. For training, we selected the high-quality DIV2K dataset, which comprises 800 high-resolution clean images with rich textures and details, providing ample training samples to enable the model to perform well under various degradation conditions [16]. Additionally, we used 100 clean/noisy image pairs as the validation set to monitor the model's performance during training and adjust the hyperparameters.

During the testing phase, we chose several widely used datasets, including Kodak, LIVE1, and BSDS100, to comprehensively assess the algorithm's performance. The Kodak dataset consists of 24 high-quality images with diverse scenes and textures, commonly used to evaluate the visual effects of image restoration algorithms [1]. The LIVE1 dataset contains a variety of image types and is widely used for image quality assessment tasks, effectively testing the algorithm's performance under different degradation conditions [14]. The BSDS100 dataset includes 100 images with rich textures and edge information, providing a comprehensive evaluation of the algorithm's performance in image restoration tasks [11].

By testing on these representative datasets, we were able to comprehensively evaluate the CTMP algorithm's performance across different degradation types and image conditions, ensuring its effectiveness and reliability in practical applications.

4.2. Image Denoising Results

In this study, we selected four representative state-of-the-art algorithms for evaluation, which cover the fields of deep learning-based image restoration and image denoising. These algorithms include Masked Denoising [2], SwinIR [9], ESWT [15], and PromptIR [13]. These algorithms have been widely applied in image restoration tasks and demonstrate the cutting-edge level of the current image denoising field.

In the non-blind image denoising mode, we trained separate models for different noise levels. To ensure the reliability and reproducibility of the experimental results, we used the high-quality DIV2K dataset for training and selected the Kodak, LIVE1, and BSDS100 datasets as test sets. Specifically, during the training phase, we added Gaussian noise with a fixed standard deviation of $\sigma = 50$ to the images in the DIV2K dataset and trained the model using this data. In the testing phase, we similarly added Gaussian noise with $\sigma = 50$ to the test images and evaluated the model's performance in the non-blind mode. Notably, the trained model was submitted to the NTIRE 2025 Image Denoising Challenge ($\sigma = 50$). As shown in Table 1.

Method	Kodak		LIVE1	
	PSNR (dB)	SSIM	PSNR (dB)	SSIM
MaskedDenoising	28.55	0.7807	27.6	0.7838
SwinIR	29.09	0.7950	28.09	0.7981
ESWT	28.676	0.7800	27.203	0.7726
PromptIR	29.36	0.8063	28.33	0.8092
CTMP (Our)	**29.50**	**0.8089**	**28.47**	**0.8114**

Table 1. Comparison of different denoising methods on Kodak, LIVE1, and BSDS100 datasets.

As can be observed from Table 1, the CTMP algorithm achieves higher Peak Signal-to-Noise Ratio (PSNR) values across all tested datasets compared to other mainstream methods. This indicates that CTMP possesses superior restoration capabilities when the noise levels are known. The enhanced performance can be attributed to the algorithm's ability to effectively capture and integrate multi-scale features, which allows it to better handle various types of image degradations.

4.3. Compressed Image Restoration Results

In this study, we conducted a comprehensive evaluation of eight representative state-of-the-art algorithms that extensively cover the fields of deep learning-based image artifact removal and image restoration. Specifically, these algorithms include not only advanced methods focused on image artifact removal, such as ARCNN [4], FBCNN [5], PromptCIR [8], and IDCN [19], but also those that have demonstrated excellent performance in the field of image restoration, include Masked Denoising [2], SwinIR [9], ESWT [15], and PromptIR [13]. These algorithms have been widely applied in the restoration tasks of JPEG compressed images, showcasing the cutting-edge level of the current image restoration field.

To systematically evaluate the performance of these algorithms in the task of JPEG artifact reduction, we trained these models on the DIV2K training set, which is widely used in image restoration research for its high-quality image resources [16]. To comprehensively compare the performance of these algorithms across different compression qualities, we carefully selected images with a wide range of Quantization Parameter (QP) values from 10 to 80 as test samples. This extensive range of QP values ensures that we can thoroughly assess the restoration capabilities of these algorithms when dealing with images ranging from very low quality to near-lossless compression, thereby providing a robust reference for image restoration in various compression scenarios in practical applications. Through this systematic evaluation approach, we aim to offer researchers and practitioners in the field of image restoration a comprehensive comparison of algorithm performance to guide their choices when facing different compressed image restoration tasks.

To comprehensively evaluate the performance of the CTMP algorithm in compressed image restoration tasks, we used the DIV2K dataset for training and selected the Kodak, LIVE1, and BSDS100 datasets as test sets. Compressed images were generated using the standard JPEG compression standard. In the blind mode, we trained a unified model using randomly different Quantization Parameters (QP). Under this mode, the model is required to handle compressed images of varying QP qualities to enhance image quality. The advantage of this approach lies in reducing model storage and training complexity while improving the model's generalization ability.

For image restoration models that do not support blind mode, to facilitate comparison, we used training sets with images of different QP qualities. By randomly generating images with QP values ranging from 10 to 80 and ensuring an even distribution of images for each QP quality, approximately 112 images per QP quality were obtained (900/8 = 112.5). This training strategy ensures that the model performs well across a variety of QP qualities.

During the testing phase, we evaluated the models on the Kodak, LIVE1, and BSDS100 datasets for different QPs (10-80). The test results are shown in Figures 3, 4, and 5, respectively, with the x-axis representing bits per pixel (bpp) and the y-axis representing Peak Signal-to-Noise Ratio (PSNR).

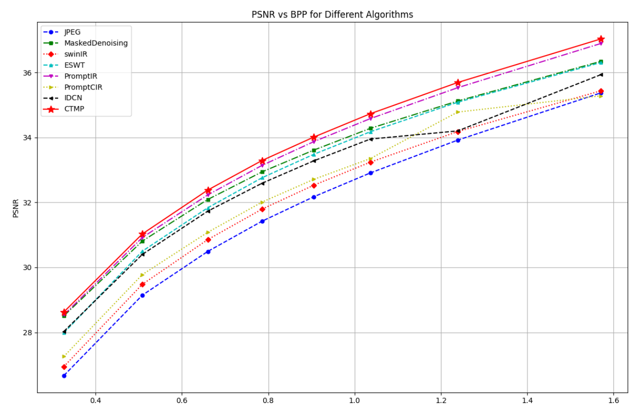

Figure 3. Test results on the Kodak dataset for different QPs.

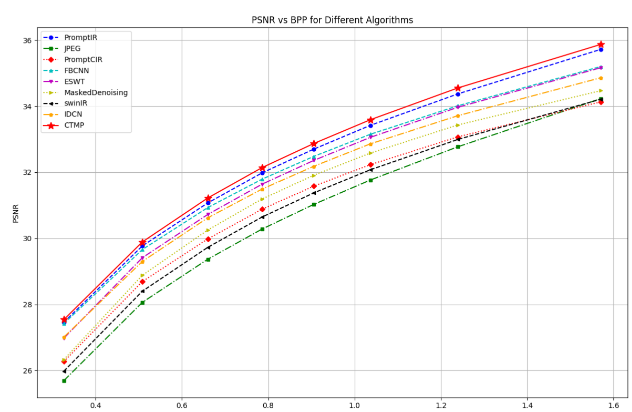

Figure 4. Test results on the LIVE1 dataset for different QPs.

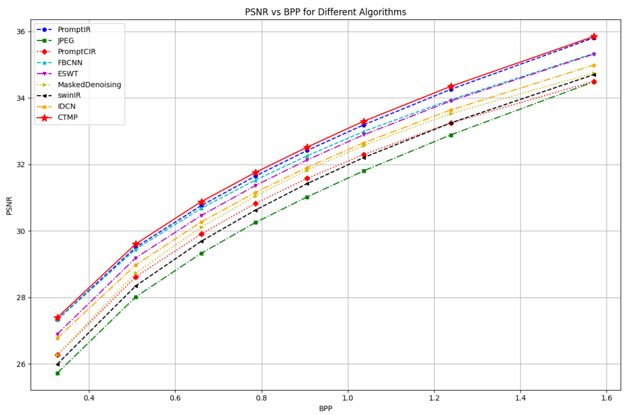

Figure 5. Test results on the BSDS100 dataset for different QPs.

As illustrated in Figures 3, 4, and 5, the proposed CTMP algorithm demonstrates remarkable performance in the restoration of compressed images. Across various compression levels and image datasets, CTMP consistently outperforms existing methods.

5. Conclusion

This paper proposes CTMP, a blind image restoration method using Channel Attention Transformers and Efficient Multi-Scale Attention (EMA) Prompt learning. CTMP enhances image restoration by integrating channel attention with self-attention mechanisms to capture both high-frequency details and low-frequency structures. The EMA module further improves feature representation through multi-scale attention. Experiments demonstrate CTMP's superior performance in denoising and removing compression artifacts, showcasing its strong generalization ability and adaptability to various image degradations.

References

[1] Kodak dataset. http://r0k.us/graphics/kodak/. 6

[2] Haoyu Chen, Jinjin Gu, Yihao Liu, Salma Abdel Magid, Chao Dong, Qiong Wang, Hanspeter Pfister, and Lei Zhu. Masked image training for generalizable deep image denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1692–1703, 2023. 6, 7

[3] Krishna Teja Chitty-Venkata, Murali Emani, Venkatram Vishwanath, and Arun K Somani. Neural architecture search for transformers: A survey. *IEEE Access*, 10:108374–108412, 2022. 1

[4] Chao Dong, Yubin Deng, Chen Change Loy, and Xiaoou Tang. Compression artifacts reduction by a deep convolutional network. In *Proceedings of the IEEE international conference on computer vision*, pages 576–584, 2015. 7

[5] Jiaxi Jiang, Kai Zhang, and Radu Timofte. Towards flexible blind jpeg artifacts removal. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4997–5006, 2021. 7

[6] Jiaxi Jiang, Kai Zhang, and Radu Timofte. Towards flexible blind jpeg artifacts removal. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4997–5006, 2021. 1

[7] Bingchen Li, Xin Li, Yiting Lu, Ruoyu Feng, Mengxi Guo, Shijie Zhao, Li Zhang, and Zhibo Chen. Promptcir: blind compressed image restoration with prompt learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6442–6452, 2024. 1

[8] Bingchen Li, Xin Li, Yiting Lu, Ruoyu Feng, Mengxi Guo, Shijie Zhao, Li Zhang, and Zhibo Chen. Promptcir: blind compressed image restoration with prompt learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6442–6452, 2024. 7

[9] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021. 6, 7

[10] Jiaying Liu, Dong Liu, Wenhan Yang, Sifeng Xia, Xiaoshuai Zhang, and Yuanying Dai. A comprehensive benchmark for single image compression artifact reduction. *IEEE Transactions on image processing*, 29:7845–7860, 2020. 1

[11] 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. In *IEEE International Conference on Computer Vision (ICCV)*, pages 416–423, 2001. 6

[12] Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. *Advances in Neural Information Processing Systems*, 36:71275–71293, 2023. 1

[13] Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. *Advances in Neural Information Processing Systems*, 36:71275–71293, 2023. 6, 7

[14] H. R. Sheikh, M. F. Sabir, and A. C. Bovik. Live image quality assessment database release 2. *http://live.ece.utexas.edu/research/quality/*, 2006. 6

[15] Jinpeng Shi, Hui Li, Tianle Liu, Yulong Liu, Mingjian Zhang, Jinchen Zhu, Ling Zheng, and Shizhuang Weng. Image super-resolution using efficient striped window transformer. *arXiv preprint arXiv:2301.09869*, 2023. 6, 7

[16] Radu Timofte, Veronique D. Smet, and Luc V. Gool. Div2k: A dataset for image super-resolution. *IEEE International Conference on Image Processing (ICIP)*, pages 2496–2500, 2018. 6, 7

[17] Zhijun Tu, Kunpeng Du, Hanting Chen, Hailing Wang, Wei Li, Jie Hu, and Yunhe Wang. Ipt-v2: Efficient image processing transformer using hierarchical attentions. *arXiv preprint arXiv:2404.00633*, 2024. 1

[18] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on*

computer vision and pattern recognition, pages 5728–5739, 2022. 1

[19] Bolun Zheng, Yaowu Chen, Xiang Tian, Fan Zhou, and Xuesong Liu. Implicit dual-domain convolutional network for robust color image compression artifact reduction. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(11):3982–3994, 2019. 7