Enhanced Blind Image Restoration with Channel Attention Transformers and Multi-Scale Attention Prompt-based Learning

Jianhua Hu P2412994@MPU.EDU.MO
K. L. Eddie Law* EDDIELAW@MPU.EDU.MO
Faculty of Applied Sciences, Macao Polytechnic University, Macao, China

Editors: Hung-yi Lee and Tongliang Liu

Abstract

Deep learning models today are indispensable tools for image compression and restoration. However, despite recent progress, many existing models often lack generalization upon facing with different types and coding strength designs of image restoration, thus limiting their practical application. In this paper, a novel approach called *dual-Channel Transformers and Multi-scale attention Prompt learning (CTMP)* is introduced to bridge the gap on blind image restoration. The prompt-based learning approach is employed in the model to address two key image restoration tasks: 1) compressed image artifact removal, and 2) image denoising. By utilizing adaptive prompts to accommodate varying quantization parameter (QP) values and noise conditions, and enhancing adaptability through the integration of multi-scale attention mechanisms, an advanced Transformer architecture in our model can tackle diverse image degradations in blind image restoration. That is, our Transformer module is improved through merging and harnessing the strengths of both channel attention and self-attention. The design is adept at extracting both high-frequency details and low-frequency structures, thereby significantly enhancing overall restoration performance. Using the Kodak dataset in experiments, our model outperforms conventional deep learning techniques with a 2.44% reduction of BD-rate in blind mode. It shows a 29.21% improvement over traditional JPEG compression and a 0.14 dB improvement in blind denoising. The experiments demonstrate that our approach is capable of training a single model effectively for both compressed image artifact removal and image denoising. The code is publicly available on GitHub at https://github.com/gdit-ai/CTMP.

Keywords: Deep learning; image restoration; Transformer; multi-scale attention.

1. Introduction

Image restoration technology is pivotal for converting low-quality (LQ) images into high-quality (HQ) counterparts, with applications including denoising, deblurring, deraining, and removing compression artifacts, etc. Despite its broad utility, the process is intricate due to the irreversible loss of image details during degradation, which poses a significant challenge for accurate restoration (Chitty-Venkata et al., 2022).

Deep learning models, particularly the convolutional neural networks (CNNs), make strides by learning the nonlinear mappings between LQ and HQ images. However, CNNs are generally limited in managing long-range dependencies which hinder their effectiveness in certain restoration tasks (Zamir et al., 2022). The Transformer and its variants, well

* Corresponding author. The MPU internal document identity number is fca.5cc2.bf19.a

known for the success in natural language processing through multi-head self-attention mechanisms, are increasingly being adopted in image processing (Tu et al., 2024a; Jiang et al., 2021b; Liu et al., 2020). The Transformers can model long-range dependencies, but often overlook low-frequency information. It is thus critical for capturing the overall structure and semantics of images (Potlapalli et al., 2023).

Blind image restoration methods face limitations with varying noise types and intensities. Traditional approaches require separate models for different degradations, which is impractical due to the increased storage and training complexities. Recent advancements, e.g., PromptIR (Potlapalli et al., 2023) and PromptCIR (Li et al., 2024a), introduce the prompt mechanism to adapt to diverse degradations, enhancing model generalization. However, these prompt-based methods struggle with feature extraction and multi-dimensional attention. Moreover, prompt learning methods lack in multi-scale feature extraction and attention allocation, limiting their ability to handle complex restoration tasks (Li et al., 2024a).

To address these issues, we introduce the Channel Attention Transformers and Efficient Multi-scale Attention Prompt learning (CTMP) model for blind image restoration in this paper. By integrating channel attention mechanism into the Transformer module, the CTMP leverages comprehensive feature extraction. That is, by capturing both high-frequency details and low-frequency structures, to overcome the limitations of earlier Transformer designs. Additionally, we introduce an Enhanced Multi-scale Attention (EMA) module to extract feature representation by combining both channel attention and spatial attention. This module improves the model's adaptability to various degradations and compensates the shortcomings of existing prompt learning methods in multi-scale feature extraction.

Our proposed CTMP method represents a substantial leap forward in blind image restoration. It provides a unified approach to address varying compression qualities and diverse noise levels, thereby significantly enhancing the model's generalization capability and restoration accuracy. In this paper, our contributions are:

1. **Improving the Transformer module by integrating Channel Attention mechanism**: This study enhances the Transformer module by integrating Channel Attention with its self-attention mechanism, allowing efficient extraction of both high- and low-frequency image features for comprehensive restoration.

2. **Designing a Prompt module based on Enhanced Multi-scale Attention (EMA)**: This paper introduces an Efficient Multi-scale Attention-based Prompt module for blind image restoration across diverse noise levels and compression qualities. The EMA module synergizes channel and spatial attention to adeptly handle multi-scale features, enhancing model adaptability to various degradations.

3. **Unified experimental design for image denoising and compression artifact removal**: This paper presents a unified experimental framework for compression artifact removal and image denoising, facilitating a comparison of blind restoration methods. The CTMP method excels, particularly in blind restoration, offering a novel solution for image restoration challenges.

2. Related Work

In this Section, we discuss related work across key areas: deep learning models, more specifically, the Vision Transformers (ViTs), and the prompt learning for image restoration.

2.1. Deep Learning in Image Processing

Non-blind image restoration refers to the recovery of images given known degradation types and parameters. It primarily focuses on two aspects: 1) image denoising, and 2) image deblocking. In the field of image denoising, CNNs become the mainstream approach due to their powerful feature extraction capabilities. For instance, DnCNN (Zhang et al., 2017) effectively removes Gaussian noise by incorporating deep residual learning. RNAN (Zhang et al., 2019) further enhances the model's adaptability to complex noise by introducing a non-local attention mechanism. These methods achieve efficient noise removal by learning the mapping between noisy and clean images.

In the realm of image deblocking, we focus on eliminating compression artifacts from images. Early methods, for example, ARCNN (Dong et al., 2015) employed convolutional neural networks to restore compressed images. In recent years, Transformer-based variants, such as, SwinIR (Liang et al., 2021) and Restormer (Zamir et al., 2022), significantly improved deblocking performance by integrating multi-head self-attention mechanisms. By capturing long-range dependencies, these designs can better restore image details and textures.

2.2. Vision Transformers in Image Processing

Blind image restoration, on the other hand, refers to the recovery of images without knowledge of any degradation types and associated parameters. Mostly, we study blind image denoising and blind image deblocking. In recent years, research on blind image denoising (Luo et al., 2021, 2023) has predominantly been moving to self-supervised learning areas. For examples, the Noise2Noise (Lehtinen et al., 2018) and Noise2Self (Batson and Royer, 2019) models typically leverage internal priors of noisy images or self-supervised learning strategies to eliminate the reliance on clean images. These methods posit image noise as white, enabling the creation of training pairs from either separate images or disjoint regions of a single image. However, these models still exhibit performance limitations upon dealing with unseen noise types. To address this issue, e.g., the LAN (Kim et al., 2024), a novel framework that adapts to the specific noisy images, reduces the mismatch and enhances the adaptability of the model to unseen noise.

In the realm of blind image deblocking, PromptCIR (Li et al., 2024b) employs a method based on prompt learning. It implicitly encodes compression information through a dynamic prompt module, and enables image recovery across varying compression levels. A lightweight prompt module is incorporated to not only boost the adaptability to compress artifacts, but also reduce amount of parameter overhead. In fact, the effectiveness of PromptCIR was validated by its first-place achievement in the blind compressed image enhancement track of the NTIRE 2024 Challenge.

2.3. Prompt Learning for Image Restoration

Vision Transformers have recently redrawn the landscape of low-level vision, routinely outperforming their convolutional predecessors by reasoning about pixels as high-dimensional tokens rather than local grids. IPT (Chen et al., 2021) inaugurated this paradigm, showing that a plain ViT pretrained on ImageNet can be repurposed as a powerful denoiser/desnower without architectural surgery. Uformer (Wang et al., 2022) subsequently packaged self-attention inside a U-net skeleton, while Restormer (Zamir et al., 2022) trimmed quadratic complexity by relegating attention to the channel axis, enabling full-resolution processing on modest GPUs.

Later tweaks (Zhang et al., 2022) experiment with criss-cross or axial windows to let tokens peek beyond their panes; nevertheless, the partition itself remains a bottleneck for distance-agnostic context. By stacking windows hierarchically and re-parameterizing branches into a single inference path, IPT-V2 (Tu et al., 2024b) presently tops both denoising and deraining benchmarks while keeping the parameter count in check.

3. CTMP

Unfortunately, most deep learning image restoration methods exhibit inadequate generalization capabilities upon facing a variety of noise types and intensities. This significantly impedes the expansion of their usages to broad applications in real-world scenarios. To tackle this challenge, we propose to add a novel prompt-based learning approach to the design.

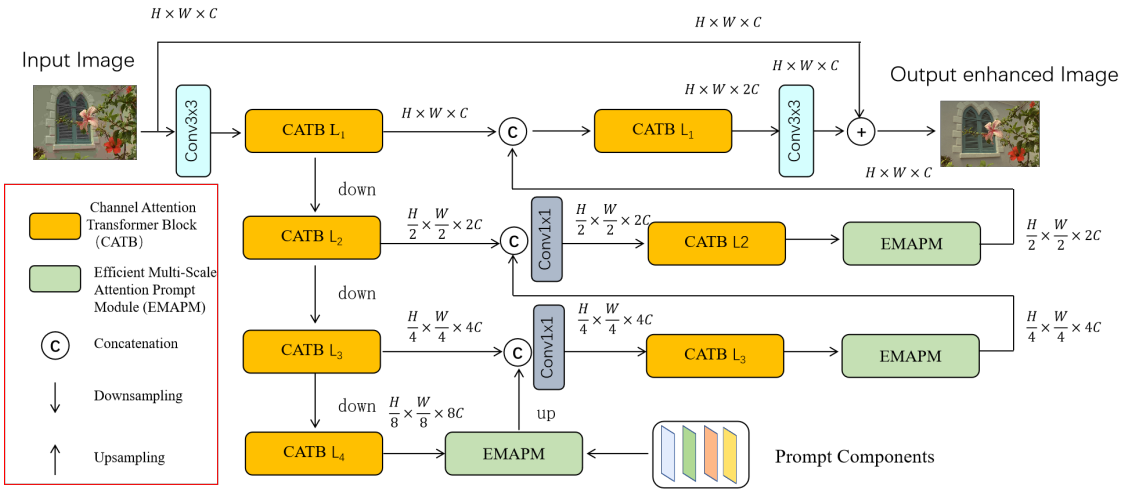

Figure 1: The CTMP Model.

As depicted in Fig. 1, we propose a model named dual-Channel Transformers and Multi-scale attention Prompt learning (CTMP) for blind image restoration. The CTMP features a U-shaped architecture grounded in the Transformer framework, constructed based on the Channel Attention Transformer Block (CATB), as shown in Fig. 2. During 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),

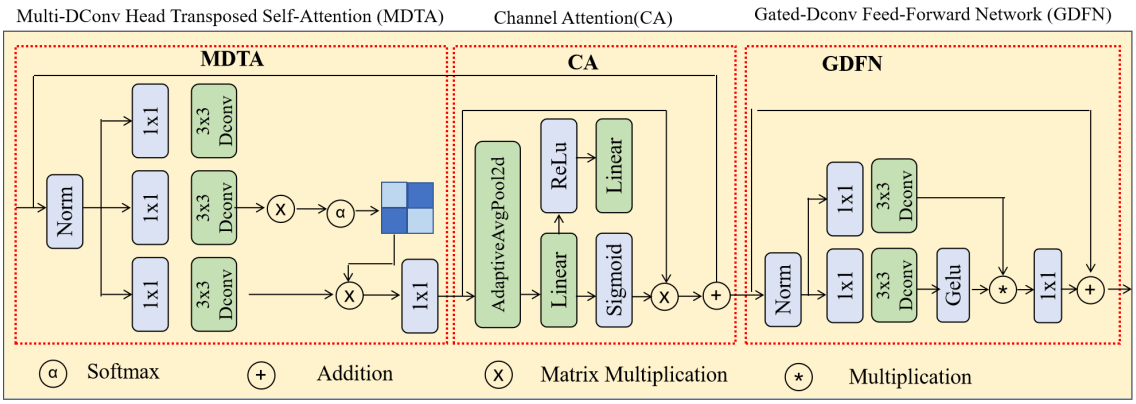

Figure 2: Structure of Channel Attention Transformer Block (CATB).

as shown in Fig. 3, that is based on prompt design. Inside the EMAPM, an Enhanced Multi-scale Attention (EMA) module is specifically for extracting global information across different directions and employing 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. It yields an enriched feature representation. This fusion mechanism empowers the model to more effectively capture and leverage features at different scales, thereby remarkably bolstering its capacity to restore image degradations and showcasing superior generalization capabilities.

3.1. Transformer Block: Channel Attention and Residual Connections

Transformer block is a central component in our design. Typically, it utilizes self-attention for feature extraction. Through the integration of the channel attention's low-frequency structure and semantic extraction, and the Transformer's high-frequency detail capture, the dual mechanism significantly improves image denoising capabilities. In Fig. 2, the improved Transformer architecture is called Channel Attention Transformer Block (CATB), which is primarily consisted of three submodules: 1) Multi-DConv head Transposed self-Attention (MDTA), 2) Channel Attention (CA), and 3) Gated-DConv Feed-forward Network (GDFN).

The MDTA module boosts local feature detection via multi-scale depth-wise convolutions, thus capturing enriched image details. Meanwhile, the CA module evaluates channel importance with a weighted feature fusion, and improves structural perception. Then the GDFN integrates gating mechanism to refine feature transformations adaptively. Collectively, these modules empower the Transformer architecture to manage a broad frequency spectrum in images, substantially improving denoising and restoration performance.

It is always important to extract precise feature information in image restoration. Our model integrates a channel attention module in the Transformer block to emphasize significant features via global average pooling and fully connected layers. Residual connections ensure data integrity, aid gradient propagation, and improve robustness. The Leaky-ReLU activation function in the Feed-Forward Network introduces non-linearity, avoiding the lim-

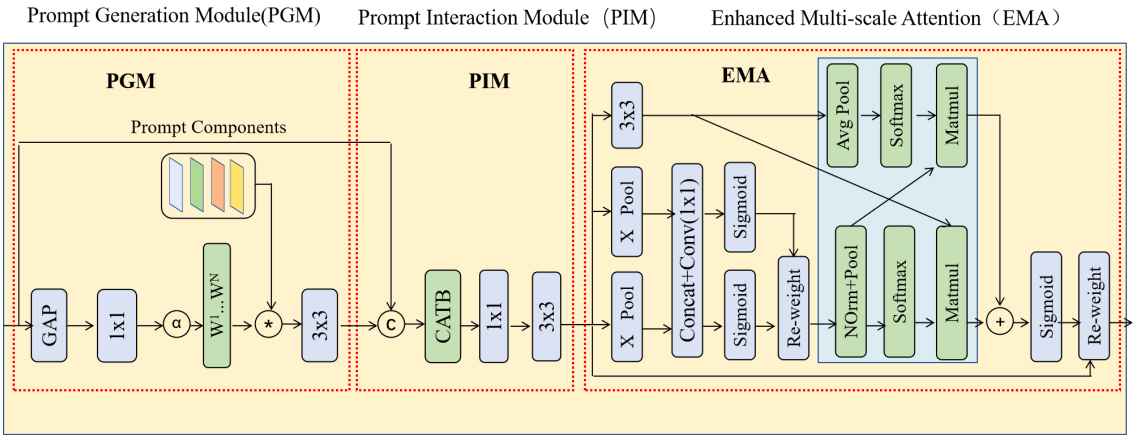

Figure 3: Structure of Efficient Multi-scale Attention Prompt Module (EMAPM).

itation of regular ReLU function. The improvements make our model highly effective for image restoration.

3.2. Efficient Multi-Scale Attention Prompt Module

In our design, the EMAPM is introduced to tackle multi-scale image degradations in restoration, and enhance feature integration across different scales beyond traditional single-scale methods. In Fig. 3, the Enhanced Multi-scale Attention (EMA) module in EMAPM captures global information, dynamically adjusts feature significance across scales, and enriches feature representation. This innovation enhances restoration quality and adaptability in image restoration.

To simplify and streamline the mathematical formulation of the EMAPM, while maintaining clarity. The operations can be merged into five cohesive steps. Given an input feature map $x \in \mathbb{R}^{B\times C\times H\times W}$, then the steps taken by EMAPM are:

1. **Embedding.** Global average pool the spatial dimensions: $\boldsymbol{e} = \frac{1}{HW}\sum_{i,j} \boldsymbol{x}_{:,:,i,j} \in \mathbb{R}^{B\times C}$.
2. **Prompt-weight generation.** A learned linear layer projects $\boldsymbol{e}$ to logit vector $\boldsymbol{z}\in\mathbb{R}^{L_p}$ and applies softmax: $\boldsymbol{z} = \boldsymbol{W}\boldsymbol{e}+\boldsymbol{b}$, $\boldsymbol{w} = \mathrm{softmax}(\boldsymbol{z}) \in \mathbb{R}^{B\times L_p}$, where $\boldsymbol{W}\in\mathbb{R}^{L_p\times C}$ and $\boldsymbol{b}\in\mathbb{R}^{L_p}$ are trainable parameters.
3. **Prompt assembly.** Let $\boldsymbol{\Theta}\in\mathbb{R}^{L_p\times d_p\times S_p\times S_p}$ denote the prompt codebook. The initial prompt map is obtained by weighted summation: $\boldsymbol{p}_1 = \sum_{k=1}^{L_p} w_{:,k}\boldsymbol{\Theta}_k \in \mathbb{R}^{B\times d_p\times S_p\times S_p}$, $\boldsymbol{p} = \texttt{F.interpolate}\big(\boldsymbol{p}_1,(H,W)\big)$.
4. **EMA enhancement and final convolution.** $\boldsymbol{p}_{\mathrm{ema}} = \mathrm{EMA}(\boldsymbol{p})$, $\hat{\boldsymbol{p}} = \mathrm{Conv}_{3\times3}(\boldsymbol{p}_{\mathrm{ema}}) \in \mathbb{R}^{B\times d_p\times H\times W}$.

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 compressed image artifact removal and denoising , and were compared with previous state-of-the-art methods. Additionally, the outcomes of the ablation studies prove the effectiveness and improvements of the design we introduced.

4.1. Implementation Details

CTMP can be tuned from scratch without warm-starting any sub-network. Its backbone stacks four encoder–decoder tiers, each hosting a distinct count of Transformer bricks—4, 6, 6, 8 as depth grows. A lightweight prompt block is sandwiched between every adjacent decoder pair, yielding three such units in total.

We feed the system with mini-batches of two 128×128 crops and let a single Tesla T4 handle the workload. Optimization is driven by plain L_1 error, guided by Adam (β_1=0.9, β_2=0.999) at an initial rate of 2×10^{-4}. To guard against overfitting, training patches undergo random left–right and up–down mirroring on the fly.

4.2. Dataset

We evaluated the CTMP algorithm's performance in image restoration, including compressed image restoration and image denoising, using the DIV2K dataset for training and Kodak, LIVE1, and BSDS100 for testing (Timofte et al., 2018; kod; Sheikh et al., 2006; Martin et al., 2001). The LIVE1 dataset, known for its diverse image types, and the BSDS100 dataset, with its rich textures and edges, further validated our algorithm's robustness across different degradation scenarios.

4.3. Compressed Image Restoration Results

In this study, we conducted a comprehensive evaluation of eight representative state-of-the-art algorithms that cover the fields of deep learning-based image artifact removal and image restoration. Specifically, these algorithms include advanced methods for image artifact removal, FBCNN (Jiang et al., 2021a), PromptCIR (Li et al., 2024b), and IDCN (Zheng et al., 2019). Additionally, we also evaluated algorithms that have demonstrated excellent performance in image restoration, including Masked Denoising (Chen et al., 2023), SwinIR (Liang et al., 2021), ESWT (Shi et al., 2023), and PromptIR (Potlapalli et al., 2023). These algorithms have been widely applied in the restoration tasks of JPEG compressed images and represent the cutting-edge level of the current image restoration field.

4.3.1. Objective Comparisons

This study rigorously assesses JPEG artifact reduction algorithms using the DIV2K dataset (Timofte et al., 2018), renowned for its high-quality images. We selected a broad QP range from 10 to 80 to evaluate performance across various compression levels, ensuring a thorough examination of restoration capabilities from low to near-lossless quality. This approach provides a solid benchmark for practical image restoration scenarios. Additionally, we trained the CTMP algorithm on the DIV2K dataset and tested it on Kodak, LIVE1, and BSDS100 datasets under blind conditions with varied QPs, enhancing its generalization. For a fair comparison, we incorporated a diverse QP training set, generating approximately

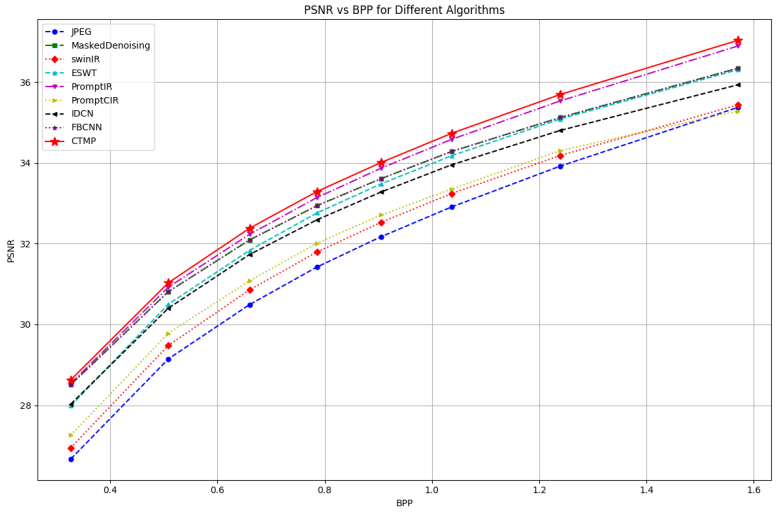

Figure 4: Test results on the Kodak dataset for different QPs.

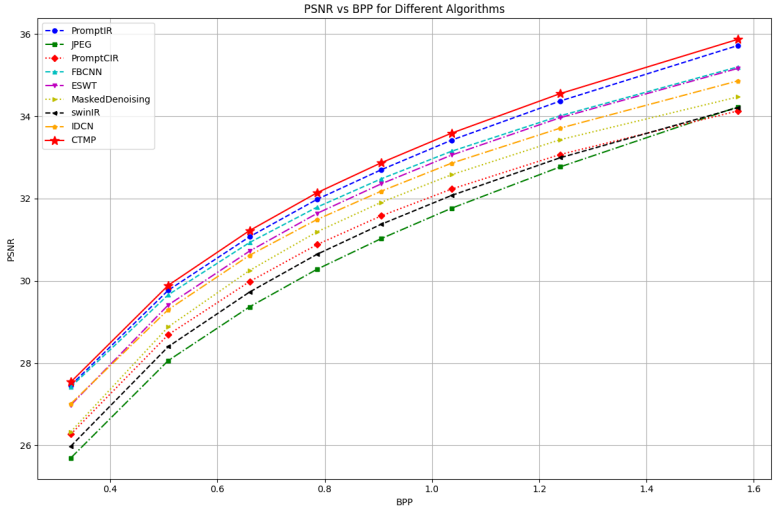

Figure 5: Test results on the LIVE1 dataset for different QPs.

112 images per QP value from 10 to 80. This strategy ensures robust performance across QP qualities and standardizes evaluation for blind image restoration

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 Fig. 4, 5, and 6, respectively, with the x-axis representing bits per pixel (bpp) and the y-axis representing Peak Signal-to-Noise Ratio (PSNR).

As illustrated in Fig. 4, 5, and 6, the proposed CTMP algorithm exhibits outstanding performance in restoring compressed images. To evaluate its effectiveness, we conducted tests using the Kodak dataset and employed the BD-Rate metric, which measures the relative change in average bit rate at a constant Peak Signal-to-Noise Ratio (PSNR) level. Our

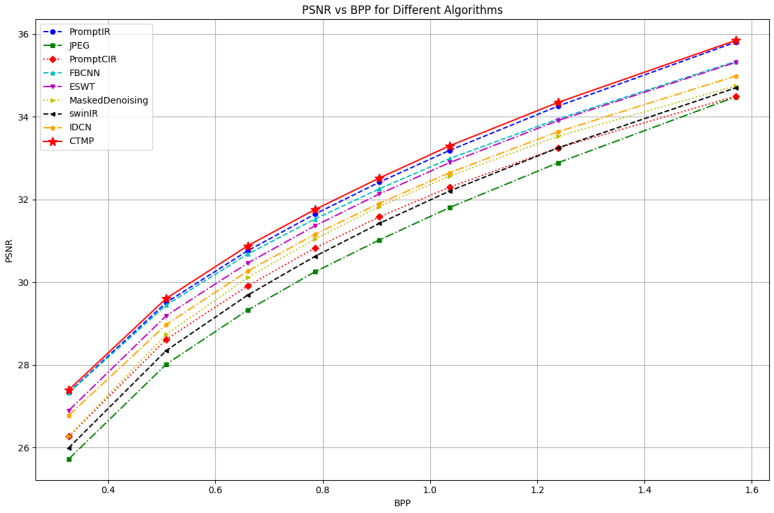

Figure 6: Test results on the BSDS100 dataset for different QPs.

method achieved a significant reduction in BD-Rate by **29.21%** compared to traditional JPEG compression. Furthermore, when compared to the best-performing algorithms in the field of deep learning-based image restoration, our approach demonstrated a **2.44%** improvement in BD-Rate. Across various compression levels and image datasets, our method, CTMP, consistently outperformed existing methods.

4.3.2. Subjective Evaluations

To more intuitively illustrate the performance of different algorithms in the task of compressed image recovery, we conducted a subjective image comparison experiment. As shown in Fig. 7, we selected Kodak images with a quantization parameter (QP) of 40 for comparison. The figure displays the results of recovery processing by various algorithms on the same image.

In the subjective evaluation, the CTMP algorithm effectively removes compression artifacts while preserving image details and structure, even at low quantization parameter (QP) values, such as QP = 40. In contrast, other algorithms may experience detail loss or over-smoothing under low QP conditions. This highlights the CTMP algorithm's superior practicality and reliability in real-world applications.

Through both objective and subjective assessments, we have comprehensively validated the CTMP algorithm's superior performance in compressed image recovery. It outperforms many latest designs in both numerical metrics and visual restoration capabilities. Our comparative evaluation underscores the CTMP algorithm's strengths in visual clarity, artifact reduction, and overall image quality, indicating its strong potential for practical applications in image recovery and enhancement.

Figure 7: Visual comparison of different methods on compress image (QP = 40).

4.4. 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, SwinIR, ESWT, and PromptIR. These algorithms have been widely applied in image restoration tasks and demonstrate the cutting-edge level of the current image denoising field.

4.4.1. Objective Comparisons

In our non-blind image denoising experiments, we developed distinct models for various noise levels, ensuring the reliability and reproducibility of our results by training on the high-quality DIV2K dataset. We assessed the model's performance using the Kodak, LIVE1, and BSDS100 datasets under test conditions with Gaussian noise ($\sigma = 50$). Consistent application of this noise level during both training and testing phases ensured uniformity. Our trained model was subsequently submitted to the NTIRE 2025 Image Denoising Challenge at the specified noise level ($\sigma = 50$).

The results, as detailed in Table 2, demonstrate the CTMP algorithm's superior performance, outperforming other state-of-the-art methods with higher Peak Signal-to-Noise Ratio (PSNR) values across all tested datasets. This enhanced restoration quality in known noise conditions is due to the CTMP algorithm's effective capture and integration of multi-scale features, enabling it to adeptly manage diverse forms of image degradation.

Table 1: Non-blind mode: Comparison of different denoising methods on Kodak, LIVE1, and BSDS100 datasets.

Method	**Kodak**		**LIVE1**		**BSDS100**	
	PSNR (dB)	**SSIM**	**PSNR (dB)**	**SSIM**	**PSNR (dB)**	**SSIM**
MaskedDenoising	28.55	0.7807	27.6	0.7838	27.4	0.7642
SwinIR	29.09	0.7950	28.09	0.7981	27.76	0.7761
ESWT	28.676	0.7800	27.203	0.7726	27.3904	0.7568
PromptIR	29.36	0.8063	28.33	0.8092	27.96	0.7845
CTMP (ours)	**29.50**	**0.8089**	**28.47**	**0.8114**	**28.08**	**0.7876**

Table 2: non-blind mode : Comparison of different denoising methods on Kodak, LIVE1, and BSDS100 datasets.

Method	**Kodak**		**LIVE1**		**BSDS100**	
	PSNR (dB)	**SSIM**	**PSNR (dB)**	**SSIM**	**PSNR (dB)**	**SSIM**
MaskedDenoising	28.55	0.7807	27.6	0.7838	27.4	0.7642
SwinIR	29.09	0.7950	28.09	0.7981	27.76	0.7761
ESWT	28.676	0.7800	27.203	0.7726	27.3904	0.7568
PromptIR	29.36	0.8063	28.33	0.8092	27.96	0.7845
CTMP (ours)	**29.50**	**0.8089**	**28.47**	**0.8114**	**28.08**	**0.7876**

In the blind image denoising experiments, we utilized the DIV2K dataset as our training set and randomly generated three levels of noise ($\sigma = 15$, 25, and 50). For image restoration models that do not support blind mode, we constructed a mixed noise dataset by evenly sampling from these three noise levels. By training a unified model to handle these mixed noise images, we assessed the model's adaptability to unknown noise levels. During the testing phase, we again set up corresponding test sets for different noise levels ($\sigma = 15$, 25, and 50) and conducted tests in blind mode. As shown in Tables 3 and 4, the results demonstrate that the CTMP algorithm can effectively manage various noise levels with a single model, highlighting its excellent generalization capabilities and robust adaptability to diverse noise environments.

Table 3: Blind mode: comparison of different denoising methods on various datasets with different noise levels (PSNR).

Method	**Kodak**			**LIVE1**			**BSDS100**		
	15	**25**	**50**	**15**	**25**	**50**	**15**	**25**	**50**
MaskedDenoising	34.34	31.84	28.62	33.58	31.01	27.68	33.42	30.75	27.46
SwinIR	33.95	31.78	28.83	34.35	31.78	28.48	33.97	31.30	28.04
ESWT	34.13	31.62	28.36	33.39	30.83	27.45	33.17	30.52	27.19
PromptIR	34.86	32.36	29.28	34.06	31.50	28.28	33.79	31.17	27.93
CTMP (ours)	**34.91**	**32.41**	**29.28**	**34.11**	**31.52**	**28.28**	**33.82**	**31.16**	**27.93**

Table 4: Blind mode: Comparison of different denoising methods on various datasets with different noise levels (MSSIM).

Method	Kodak			LIVE1			BSDS100		
	15	**25**	**50**	**15**	**25**	**50**	**15**	**25**	**50**
MaskedDenoising	0.9173	0.8717	0.7820	0.9235	0.8789	0.7856	0.9227	0.8713	0.7659
SwinIR	0.9091	0.8756	0.8128	0.9321	0.8928	0.8113	0.9296	0.8827	0.7873
ESWT	0.9145	0.8664	0.7695	0.9213	0.8742	0.7730	0.9175	0.8626	0.7484
PromptIR	0.9227	0.8794	0.7992	0.9285	0.8862	0.8026	0.9272	0.8782	0.7783
CTMP (ours)	**0.9250**	**0.8839**	**0.8039**	**0.9304**	**0.8897**	**0.8074**	**0.9285**	**0.8809**	**0.7845**

Figure 8: Visual comparison of different methods on image denoising ($\sigma = 50$).

4.4.2. Subjective Evaluations

To more intuitively demonstrate the denoising performance of different algorithms, we selected a Gaussian noise image with a noise intensity of $\sigma = 50$ from the Kodak image dataset for a subjective image comparison experiment. As shown in Fig. 8, We present the outcomes of denoising the identical image using a variety of algorithms. Specifically, using the Kodak dataset as a benchmark, our approach achieved a performance enhancement of 0.14 dB over the best-performing deep learning image restoration methods.

The CTMP algorithm not only excels in numerical metrics but also demonstrates its strengths in visual quality. It is capable of effectively preserving image details and textures while removing noise, resulting in a clearer and more natural restored image.

In the subjective evaluation, the CTMP algorithm proves to be highly effective in noise removal even at high noise levels, while maintaining the image's details and structure. In contrast, other algorithms may suffer from detail loss or over-smoothing under high noise conditions. This highlights the superior practicality and reliability of the CTMP algorithm in real-world applications.

Through comprehensive assessments from both objective and subjective perspectives, we have thoroughly validated the superior performance of the CTMP algorithm in image denoising. It outperforms mainstream algorithms in numerical metrics and exhibits remarkable visual restoration capabilities.

4.5. Model Parameters and Memory Cost

We benchmark our method against six state-of-the-art competitors: FBCNN (Jiang et al., 2021a), PromptCIR (Li et al., 2024b), Masked Denoising (Chen et al., 2023), SwinIR (Liang et al., 2021), ESWT (Shi et al., 2023), and PromptIR. As reported in Table 5, our CTMP network, with only 35.96 M learnable parameters, consumes 158.4 G FLOPs and 441.6 MB GPU memory, i.e., consistently lower than the most demanding rivals. Despite this frugality, CTMP still embeds 304 convolutional layers, achieving a favorable accuracy-to-cost trade-off that outperforms heavier counterparts, as shown in Tables 5.

Table 5: Computational & memory cost of competing methods and our proposed network. All numbers were measured on the same validation set using a Tesla T4 GPU.

Method	Params. (M)	FLOPs (G)	#Conv2d	GPU memory (MB)
FBCNN	71.9413	182.4223	80	747.439
MaskedDenoising	0.8241	211.8697	17	6963.085
swinIR	11.5042	9.6274	9	1608.220
ESWT	1.7322	390.5868	103	473.457
PromptIR	35.5936	10.4037	324	1053.053
PromptCIR	34.7549	11.6092	334	1040.563
CTMP (Ours)	**35.9200**	**158.400**	**304**	**441.600**

4.6. Ablations Studies

In the ablation studies of this research, we trained a Gaussian color image denoising model solely on image patches of size 128×128 and tested it on the Kodak dataset, with a focus on the challenging noise level of $\sigma = 50$. The results demonstrate that our proposed improvements significantly enhance the model's quality performance, as detailed in Table 6.

We analyze performance gains from our enhanced PromptIR model, featuring Transformer and Prompt Blocks. By replacing the Prompt Block with the EMAPM, we achieved a 0.11 dB boost. Substituting the Transformer Block with the Channel Attention Transformer Block (CATB) added a 0.03 dB improvement, with multi-channel attention enhancing robustness to noise.

Table 6: Performance evaluation of components via ablation studies.

Network	**Component**	**PSNR**	**MSSIM**
Baseline	PromptIR (Potlapalli et al., 2023)	29.36	0.8063
Baseline + EMAPM	Transformer Block + **EMAPM**	29.47	0.8085
Baseline + CATB	**CATB** + EMAPM	29.50	0.8089

5. Conclusion

In this paper, we have proposed a dual-Channel Transformers and Multi-scale attention Prompt learning (CTMP) model. It is designed for blind image restoration using Channel Attention Transformer Blocks (CATBs) and Efficient Multi-scale Attention Prompt Modules (EMAPMs). The CTMP improves image restoration through the integration of channel attention and self-attention mechanisms to capture both high-frequency details and low-frequency information. The EMAPM module further improves feature representation through multi-scale attention. Through thorough experiments, the CTMP demonstrates its superior performance in denoising and removing compression artifacts, thus showcasing its strong generalization ability and adaptability to various image degradations.

References

Kodak dataset. http://r0k.us/graphics/kodak/.

Joshua Batson and Loic Royer. Noise2self: Blind denoising by self-supervision. In *International conference on machine learning*, pages 524–533. PMLR, 2019.

Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition*, pages 12299–12310, 2021.

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.

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.

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.

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, 2021a.

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, 2021b.

Changjin Kim, Tae Hyun Kim, and Sungyong Baik. Lan: Learning to adapt noise for image denoising. In *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition*, pages 25193–25202, 2024.

Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*, 2018.

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, 2024a.

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, 2024b.

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.

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.

Fangzhou Luo, Xiaolin Wu, and Yanhui Guo. Functional neural networks for parametric image restoration problems. *Advances in Neural Information Processing Systems*, 34: 6762–6775, 2021.

Fangzhou Luo, Xiaolin Wu, and Yanhui Guo. And: Adversarial neural degradation for learning blind image super-resolution. *Advances in Neural Information Processing Systems*, 36:21255–21267, 2023.

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.

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.

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.

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.

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.

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*, 2024a.

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*, 2024b.

Z Wang, X Cun, J Bao, W Zhou, J Liu, and H Li. Uformer: A general u-shaped transformer for image restoration. In *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition*, pages 17683–17693, 2022.

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.

Jiale Zhang, Yulun Zhang, Jinjin Gu, Yongbing Zhang, Linghe Kong, and Xin Yuan. Accurate image restoration with attention retractable transformer. *arXiv preprint arXiv:2210.01427*, 2022.

Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017.

Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, and Yun Fu. Residual non-local attention networks for image restoration. *arXiv preprint arXiv:1903.10082*, 2019.

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.