# PromptCIR: Blind Compressed Image Restoration with Prompt Learning

Bingchen Li[1], Xin Li[1], Yiting Lu[1], Ruoyu Feng[1], Mengxi Guo[2], Shijie Zhao[2], Li Zhang[2], Zhibo Chen[1†]

[1]University of Science and Technology of China    [2]Bytedance Inc.

{lbc31415926, lixin666, luyt31415, ustcfry}@mail.ustc.edu.cn,
{guomengxi.qoelab, zhaoshijie.0526, lizhang.idm}@bytedance.com, chenzhibo@ustc.edu.cn

## Abstract

*Blind Compressed Image Restoration (CIR) has garnered significant attention due to its practical applications. It aims to mitigate compression artifacts caused by unknown quality factors, particularly with JPEG codecs. Existing works on blind CIR often seek assistance from a quality factor prediction network to facilitate their network to restore compressed images. However, the predicted numerical quality factor lacks spatial information, preventing network adaptability toward image contents. Recent studies in prompt-learning-based image restoration have showcased the potential of prompts to generalize across varied degradation types and degrees. This motivated us to design a prompt-learning-based compressed image restoration network, dubbed PromptCIR, which can effectively restore images from various compress levels. Specifically, PromptCIR exploits prompts to encode compression information implicitly, where prompts directly interact with soft weights generated from image features, thus providing dynamic content-aware and distortion-aware guidance for the restoration process. The light-weight prompts enable our method to adapt to different compression levels, while introducing minimal parameter overhead. Overall, PromptCIR leverages the powerful transformer-based backbone with the dynamic prompt module to proficiently handle blind CIR tasks, winning first place in the NTIRE 2024 challenge of blind compressed image enhancement track. Extensive experiments have validated the effectiveness of our proposed PromptCIR. The code is available at* https://github.com/lbc12345/PromptCIR-NTIRE24.

## 1. Introduction

The rapid growth of digital images and videos has necessitated the utilization of lossy compression techniques [3, 24, 48, 52, 62] to optimize storage and bandwidth utilization. Among popular compression codecs, JPEG is favored for its computational efficiency and straightforward implemen-

tation. It compresses an image by segmenting images into $8 \times 8$ blocks, applying discrete cosine transform (DCT), and quantizing the coefficients. During quantizing, the information is inevitably lost, resulting in unfavorable compression artifacts. The degree of compression artifacts is affected by the quality factor ranging from 0 to 100, with lower values indicating worse image qualities.

With the vast development of deep neural networks (DNNs), researchers have explored the application of DNNs in image restoration tasks [6, 22, 25, 26, 29, 33, 42, 60]. Pioneering works [69] utilizes convolutional neural networks (CNNs) to tackle compressed image restoration (CIR) problems. In recent years, Transformers [7, 34, 35] have showcased powerful representative capabilities with self-attention modules. Models incorporating transformer blocks [6, 17, 29, 60, 71] have surpassed previous CNN-based methods, restoring more rich texture details. However, most methods focus on restoring compressed images with predefined JPEG quality factors, significantly restricting their practicality in real-world scenarios where the quality factors are often unknown. To mitigate this problem, a series of works [14–16, 54] attempt to predict the quality factor as an auxiliary guidance, avoiding predefining the quality factor before network training.

Although straightforward, the prediction network still faces two challenges: 1) The predicted numerical quality factor loses all the spatial-wise information, lacking content-aware adaptability which is essential for image restoration (IR) tasks. 2) The prediction network usually causes significant parameter overhead, which constrains its deployment on edge devices like mobile phones [58]. To tackle the above challenges, we propose PromptCIR in this paper, a prompt-learning-based blind CIR approach.

Recent advances in prompt-learning-based IR [28, 37, 39, 43] have demonstrated the great generalization ability of prompts for universal IR tasks. Particularly, PromptIR [43] leverages lightweight prompts to encode degradation information as a guidance for restoration networks. Inspired by this, we explore an efficient method of restoring blind compressed images with prompts, dubbed PromptCIR. Differ-

ent from previous prediction-based methods which directly estimate numerical quality factors for guidance, PromptCIR learns to encode compression information implicitly through the interaction between prompt components, and prompt weights generated from distorted image features, thus providing more content-aware adaptability for blind restorations with a few additional parameters. However, it is noteworthy that most existing prompt-learning-based IR methods [39, 43] set the prompt size to be the same as the size of image feature, constraining the generalization ability for various input sizes. To mitigate this problem, UCIP [21] employs a content-aware dynamic prompt. Specifically, they set the prompt size to $1 \times 1$ and generate spatial-wise prompt weights with the same resolution of image features to avoid the potential problem of overfitting to training image size. Additionally, to improve the content-aware and distortion-aware representation ability, they use several prompt bases to dynamically encode task-adaptive information. Inspired by this, we adopt the same prompts design from UCIP [21] to adaptively encode compression quality information in PromptCIR, while preserving necessary spatial-wise knowledge.

Concretely, the U-shape structure has demonstrated strong representation ability for IR tasks [9, 38, 43, 47, 60, 68, 71]. Building on the latest powerful restoration model Restormer [71], we design our PromptCIR with a U-shape structure to hierarchically restore image features. Moreover, recent studies [27, 65, 67] have showcased that larger datasets can further boost the performance of restoration networks in terms of both objective and subjective qualities, as they encompass wider-range of texture details and more fine-grained semantics. Following their success, we further enhance the performance of our method by utilizing the open-source high-quality dataset LSDIR [27]. With dynamic prompts, representative network structure, and a powerful dataset, our PromptCIR has achieved first place in the NTIRE 2024 challenge of blind compressed image enhancement track [66].

## 2. Related Works

### 2.1. Compressed Image Restoration

Compressed Image Restoration aims to reduce compression artifacts caused by various codecs, especially for JPEG artifact reduction. Early works mainly focus on designing filters in the DCT domain. ARCNN [69], as a pioneer work, first introduced a deep-learning-based convolutional neural network (CNN) to restore compressed images. To further improve the representation ability of the network, some works explore non-local attention [76] and residual-learning [75], which achieve great restored results. Apart from the above studies in the pixel domain, DDCN [11], DMCNN [74] and D$^3$ [59] utilize DCT do-

main priors to further facilitate the restoration of compressed images. Inspired by these, Fu *et al.* [10] leverages dual domain knowledge with sparse coding to build a more compact and explainable restoration network. Similarly, [16, 77] make the most of dual domain information to achieve favorable restoration results. Recently, transformer-based methods have demonstrated great potential in IR tasks [5, 29, 60, 71]. Particularly, shifted window-based attention further boost the performance of CIR [29] and super-resolution [6, 17, 19] by long-range pixel information interaction.

However, most of the aforementioned methods rely on a predefined compression quality factor, limiting their usage in practical scenarios [14]. Ehrlich *et al.* [8] utilizes the quantization table as prior knowledge to facilitate the restoration of blind compressed images. Similarly, some works [15, 16, 54] propose to estimate the compression quality factor to handle blind CIR problems. Particularly, FBCNN [14] proposes a quality factor predictor to estimate the quality factor from image features and a controller to guide the subsequent restoration network, achieving state-of-the-art performance on blind CIR. Nevertheless, the predicted numerical quality factor lacks spatial information on image features, restricting their content-aware adaptability. Moreover, such a method requires a carefully designed predictor and controller network with non-negligible overhead in terms of additional parameters. Different from explicitly predicting quality factor, we adopt lightweight prompts to implicitly encode compression information while preserving spatial-wise flexibility.

### 2.2. Prompt-Learning-Based Image Restoration

With the exponential growth of language and vision model [4, 32, 41, 50, 78], prompt learning has begun to play an increasingly important role [13, 46]. Inspired by the success of prompt learning in language and vision tasks, several studies have started to employ prompts in IR tasks [28, 37, 39, 43]. Early work PromptIR [43] enhances powerful Restormer [71] backbone, applying a set of prompts to encode distortion information and identify different restoration tasks such as image denoising, dehazing, and deraining. The prompts will interact with weights generated from image features to perform implicit guidance for the restoration network. Simultaneously, ProRes [39] leverages an image-like prompt during fine-tuning to conduct efficient prompt-tuning for restoration. Moreover, PIP [28] employs a dual prompt approach with one prompt responsible for distortion types and another prompt for texture enhancement. Prompt-learning-based IR also attracts interests from diffusion-based models. DA-CLIP [37] uses prompts during diffusion steps to facilitate universal IR, while PromptRR [55] utilizes diffusion to refine prompts for better reflection removal. In this work, we explore prompts
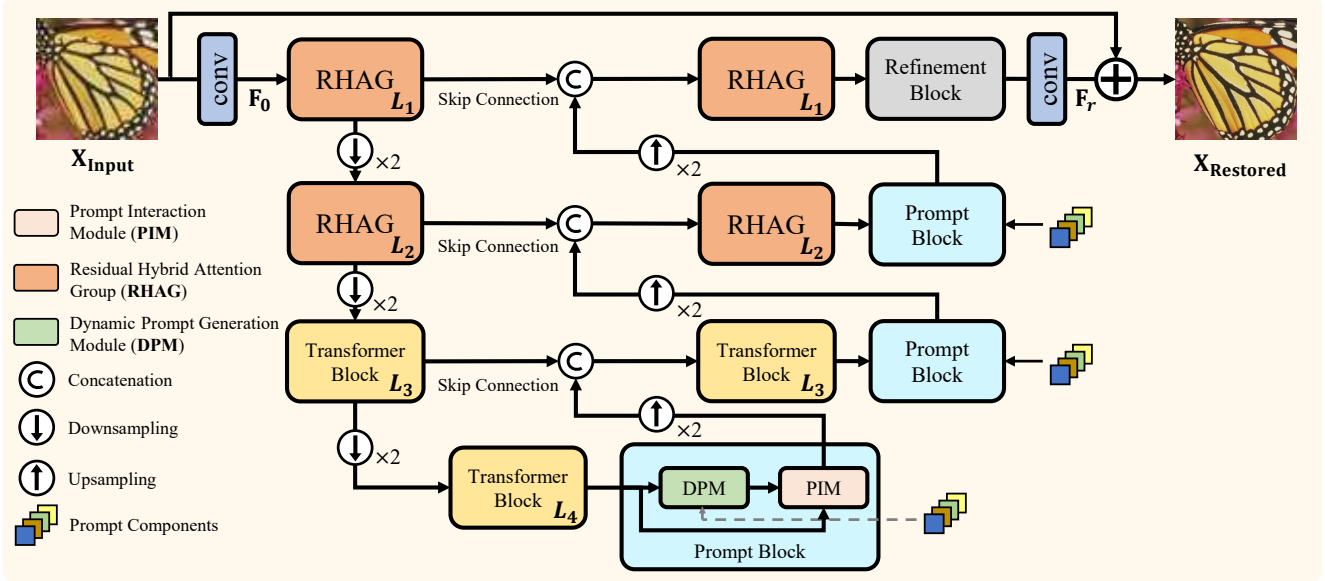
Figure 1. The overall framework of our proposed PromptCIR. We introduce prompt learning for blind compressed image restoration. To preserve content-aware knowledge while efficiently encoding distortion-aware information, we utilize DPM from UCIP [21] to provide implicit guidance for the restoration process. To enhance the representation ability of the network, we replace the transformer blocks in the first two stages with RHAG inherited from HAT [5].

with blind CIR task. We use a set of dynamic prompts [21] to perform comprehension of distortion levels while maintaining the capability of learning content-aware information for blind restoration.

## 3. Method

In this section, we first demonstrate the adaptability of our prompts towards content-aware and distortion-aware modeling. Then, we introduce the advantage of hybrid attention block [5] against the original transformer block used in Restormer [71]. Finally, we present the overall framework of our proposed PromptCIR.

### 3.1. Prompt Block

**Preliminary.** With the development of large models [32, 41, 50], prompt learning has played essential role for efficient tuning. In such a training scheme, the original large model is usually fixed [18, 37], while the lightweight prompts are optimized for new tasks. However, in IR fields, the well-trained large foundation model for various degradation restoration is missing. To restore images with unknown degradations, researchers use prompts as an efficient component [2] to adaptively encode distortion-aware information to guide the restoration process with a single all-in-one restoration model, while the prompts and the model are jointly *trained from-scratch* [21, 28, 43]. Blind CIR is very similar to all-in-one restoration. In blind CIR, the distortion type is consistent, but its degree remains unknown to

the model. Therefore, using prompts in blind CIR would be more effective in modeling distortion-aware information.

**Structure.** The prompt block is composed of a dynamic prompt generation module (DPM) and a prompt interaction module (PIM). Compared to other prompt generation module, DPM [21] features more adaptability in content-aware representation, as it uses several prompt bases to model spatial information. Taking prompt components with $N$ prompt bases $P_b \in \mathbb{R}^{1 \times 1 \times \hat{C}}$ and image features $F_I \in \mathbb{R}^{H \times W \times C}$ as inputs, DPM first generates weights $w_i \in \mathbb{R}^{H \times W \times N}$ directly from $F_I$, thus preserving distortion-aware adaptability. Next, the weights interact with the prompt bases in a manner similar to the operation of attention mechanisms to form the final prompts $P \in \mathbb{R}^{H \times W \times \hat{C}}$. Each spatial position of $P$ is dynamically composed by prompt bases through weights, thus realizing content-aware flexibility, which is crucial for blind CIR. This process can be formulated as:

$$P = \mathrm{Conv}\left(\sum_{b=1}^{N} w_i \odot P_b\right), \quad w_i = \mathrm{Softmax}(\mathrm{MLP}(F_I)) \tag{1}$$

As PIM is not our contribution, we directly adopt the well-designed module from PromptIR [43]. PIM (shown in Fig. 3) leverages a transposed self-attention transformer block to efficiently fuse information of prompt guidance with image features, as shown in Fig. 2.
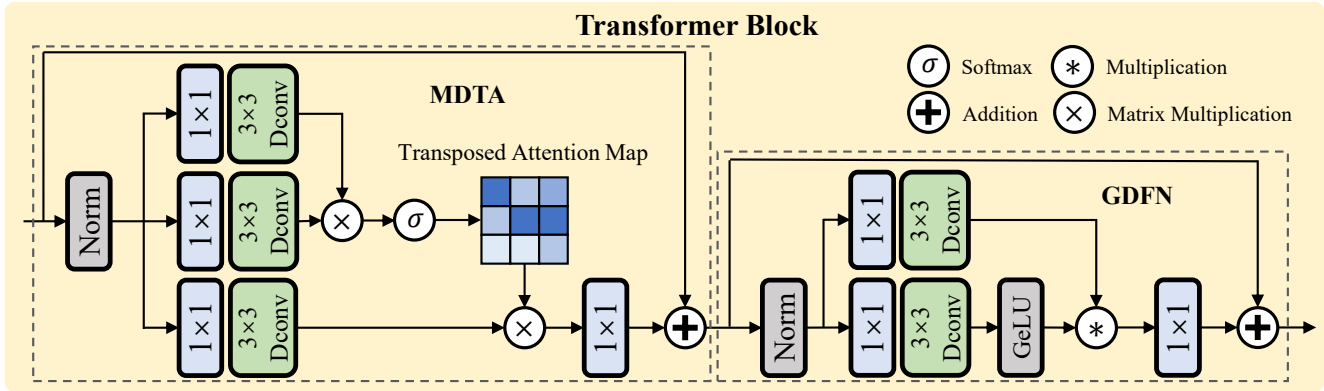
Figure 2. The structure of transformer block [43, 71]. It is composed of two modules, including the Multi Dconv head transposed attention module (MDTA) and the Gated Dconv feed-forward network (GDFN). Compared to traditional self-attention block, transposed attention mechanism provides more efficient information extraction with less computational complexity.
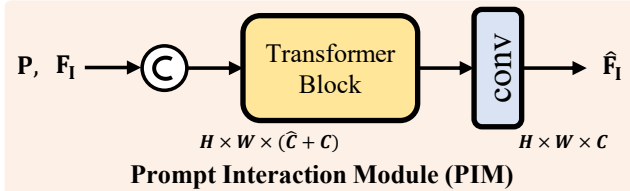


**Prompt Interaction Module (PIM)**

Figure 3. The structure of PIM. Here a $1 \times 1$ convolution is used to reduce channel numbers to the same as input features.

## 3.2. Hybrid Attention Block

Blind CIR task requires that the network possess strong representation capability, encompassing both content-aware and distortion-aware comprehension. Specifically, the JPEG codec compresses images by dividing them into $8 \times 8$ blocks, necessitating networks to possess a robust capability for aggregating local information in order to effectively restore compression artifacts. Although the transformer block of PromptIR [43] has computational advantages, it fails to meet the local information extraction needs of blind CIR due to its reliance on global-wise token attention. Recently, the shifted window-based attention mechanism has demonstrated great potential in IR tasks [5, 6, 19, 29], primarily because it integrates the ability to extract local features with the capacity for long-range pixel dependency modeling enabled by window-shifting. Notably, HAT [5] integrates a channel attention block in parallel with the original shifted window-based attention block, as convolution facilitates the transformer to enhance representation abilities [20, 61, 64, 70]. Furthermore, HAT introduces the overlapped window-based attention to directly establish cross-window connections, boosting the representative capabilities. Concretely, HAT establishes a residual hybrid attention group (RHAG) by combining these two designs, *possessing strong local and global information extraction abilities that*
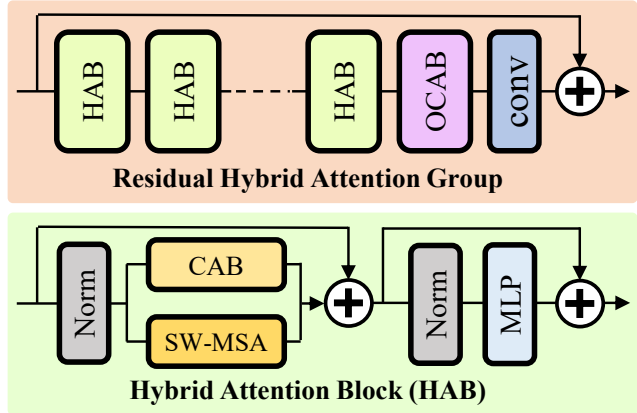


Figure 4. (Upper) illustrates the structure of RHAG [5], it combines several hybrid attention blocks (Bottom) with an overlapped cross-attention block (OCAB) followed by one convolution. Such design enhances the representation capabilities of the network in both local and global information extraction, and highly meets our needs for blind CIR.

*meet our needs for blind CIR*. Therefore, we replace the first two transformer blocks in both the downsampling and upsampling stages with RHAG. The structure of RHAG is illustrated in Fig. 4

**Discussion.** As image features go deeper, *i.e.*, stage $L_3$ and $L_4$, local feature extraction becomes less important that global information modeling [12, 51]. Besides, the spatial resolution of feature maps is insufficient for window-shifting in deeper stages. Therefore, we only apply RHAG to the first two stages.

## 3.3. Overall Pipeline

To tackle the challenging blind CIR problem, we propose a prompt-learning-based framework, named PromptCIR. Compared with previous blind CIR and prompt-learning-

Table 1. Quantitative comparison for blind CIR. Notice that, for a fair comparison on the scale of the training dataset, we train PromptIR†and PromptCIR with only DF2K, marking with *. Results are tested in terms of PSNR↑/SSIM↑/PSNRB↑. The best performances are **bolded**.

| | JPEG | FBCNN [14] | Wang *et al*. [58] | PromptIR†* [43] | PromptCIR* |
|---|---|---|---|---|---|
| Training data | - | DF2K | DIV2K+BSDS500(train) | DF2K | DF2K |
| Params(M) | - | 71.92 | 11.11 | 36.54 | 34.75 |
| LIVE1 | 30.02/0.872/28.18 | 32.08/0.904/31.44 | 32.26/0.907/31.70 | 32.34/0.907/31.75 | **32.45/0.908/31.87** |
| BSDS500 | 30.58/0.876/28.29 | 32.41/0.903/31.45 | **32.75/0.908/31.78** | 32.67/0.907/31.67 | 32.74/0.907/31.76 |
| ICB | 33.87/0.842/33.05 | 35.90/0.865/36.33 | 35.93/0.865/36.35 | 36.35/0.868/36.75 | **36.41/0.870/36.79** |
| DIV2K | 31.71/0.885/30.02 | 33.96/0.918/33.41 | 34.13/0.919/33.62 | 34.38/**0.922**/33.81 | **34.45/0.922/33.88** |

Table 2. Quantitative comparison for non-blind CIR evaluation. Notice that, for a fair comparison on the scale of the training dataset, we train PromptIR†and PromptCIR with only DF2K, marking with *. Results are tested in terms of PSNR↑/SSIM↑/PSNRB↑. The best and second best performances are **bolded** and underlined, respectively. We also include PromptCIR trained with DF2K and LSDIR for a comprehensive comparison.

| Dataset | QF | JPEG | FBCNN [14] | Wang *et al*. [58] | PromptIR†* [43] | PromptCIR* | PromptCIR |
|---|---|---|---|---|---|---|---|
| LIVE1 | 10 | 25.69/0.743/24.20 | 27.77/0.803/27.51 | 27.80/0.805/27.57 | 27.98/0.806/27.66 | <u>28.00/0.806/27.73</u> | **28.28/0.813/27.94** |
| | 20 | 28.06/0.826/26.49 | 30.11/0.868/29.70 | 30.23/0.872/29.85 | 30.31/0.872/29.82 | <u>30.32/0.872/29.86</u> | **30.61/0.876/30.12** |
| | 30 | 29.37/0.861/27.84 | 31.43/0.897/30.92 | 31.58/0.900/<u>31.13</u> | 31.66/0.900/31.07 | <u>31.69/0.901</u>/31.10 | **31.96/0.904/31.38** |
| | 40 | 30.28/0.882/28.84 | 32.34/0.913/31.80 | 32.53/0.916/32.04 | 32.58/0.915/31.96 | <u>32.64/0.917/32.12</u> | **32.88/0.919/32.26** |
| ICB | 10 | 29.44/0.757/28.53 | 32.18/0.815/32.15 | 32.05/0.813/32.04 | 32.70/0.821/33.01 | <u>32.73/0.822/33.12</u> | **33.15/0.825/33.45** |
| | 20 | 32.01/0.806/31.11 | 34.38/0.844/34.34 | 34.32/0.842/34.31 | 34.84/0.847/35.21 | <u>34.89/0.848/35.24</u> | **35.21/0.850/35.57** |
| | 30 | 33.20/0.831/32.35 | 35.41/0.857/35.35 | 35.37/0.856/35.35 | 35.85/0.860/36.24 | <u>35.89/0.861/36.26</u> | **36.19/0.863/36.58** |
| | 40 | 33.95/0.840/33.14 | 36.02/0.866/35.95 | 35.99/0.860/35.97 | 36.45/0.869/36.85 | <u>36.52/0.871/36.92</u> | **36.78/0.872/37.18** |

based methods, PromptCIR has several advantages: 1) We use lightweight dynamic prompts [21] to implicitly encode content-aware and distortion-aware information as flexible guidance for restoration networks. Prompts have more spatial adaptability than numerical predicted quality factors. 2) We use RHAG [5] in the first two stages to enhance the local and global modeling capabilities of network, which is more helpful for removing compression artifacts.

The framework of PromptCIR is demonstrated in Fig. 1. Given a compressed input image $X_{\text{Input}} \in \mathbb{R}^{H \times W \times 3}$, where H and W are the spatial resolutions, the network first uses a convolution layer to convert image into feature $F_0 \in \mathbb{R}^{H \times W \times C}$, where C is the embedded channel number. $F_0$ further passes through a 4-stage U-shape encoder-decoder structure. In the encoder stage, the image feature is iteratively downsampled to $F_l \in \mathbb{R}^{\frac{H}{8} \times \frac{W}{8} \times 8C}$. In the decoder stage, the image feature will gradually upsampled from $F_l$ to the original spatial resolution. We utilize prompts to implicitly learn the content-aware and distortion-aware information as a guidance for each decoder stage. Additionally, following [43, 71], a refinement block with 4 transformer blocks is used to further enhance image features. A convolution layer is leveraged to project the channel number of restored feature $F_r$ from $C$ back to 3. Finally, we get restored clean image $X_{\text{Restored}} \in \mathbb{R}^{H \times W \times 3}$ by adding restored feature $F_r$ with $X_{\text{Input}}$.

## 4. Experiments

### 4.1. Implementation Details

**Training Datasets.** We use DF2K (800 images from DIV2K [1] and 2650 images from Flickr2K [49]) and LS-DIR [27] to form our training datasets. It is noteworthy that LSDIR is a recently proposed large-scale high-quality training datasets for IR tasks. Following the suggestion of NTIRE 2024 official code[1], we apply PIL toolkit to generate compressed training images at 7 quality factors ranging from 10 to 70 with a step size of 10.

**Optimization Details.** We train our PromptCIR by 8 NVIDIA Tesla V100 16GB GPUs with a total batchsize of 24. To explore the upper limits of our model, we adopt a two-stage training strategy to optimize our PromptCIR. Specifically, in the first stage, the model is pre-trained on 7 aforementioned quality factors, fully developing its content-aware information extraction abilities. In the second stage, the model is fine-tuned on online compressed images with quality factors randomly selected from [10, 70]. At this stage, the model focuses more on distortion-aware information encoding, while still retaining the ability to extract content-aware information through prompt bases [21].

During the whole training process, the training images are paired-cropped into $128 \times 128$ patches following [14, 43], while applying random flips and rotations. In
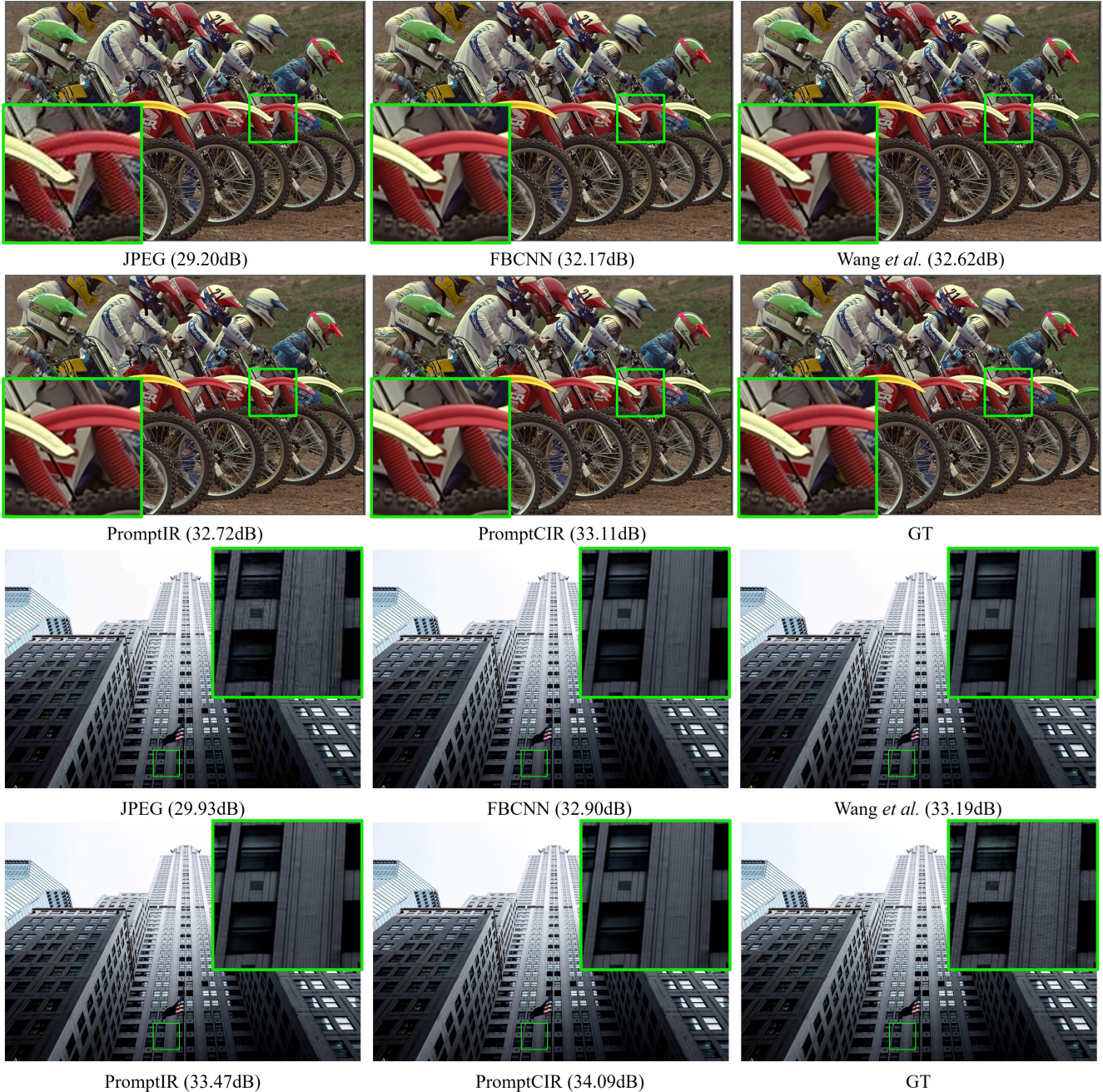
---

[1]https://codalab.lisn.upsaclay.fr/competitions/17548

Figure 5. Qualitative comparisons between different methods on blind CIR. Zoom in for better views. (Upper: LIVE1_bikes. Bottom: DIV2K_0846)

the first stage, We adopt AdamW [36] optimizer to optimize our model with an initial learning rate of 2e-4. A CosineAnnealing scheduler is employed to decay the learning rate to 1e-6, with the total number of iterations set to 800k. In the second stage, we apply the same strategy with modification of the initial learning rate to 1e-4 and the iterations to 600k. **Network parameters.** We follow the same settings of block numbers ([4,6,6,8]) and head numbers ([1,2,4,8]) of PromptIR [43]. The only difference is that, in the first

two stages of the downsampling and upsampling stages, the block number refers to the number of successive HABs, and the head number refers to the count of heads used in the shifted window-based attention blocks.

### 4.2. Blind Compressed Image Restoration

We evaluate the performance of PromptCIR against the contrastive learning-based method [58] (denoted as Wang *et al*.), FBCNN [14], and the prompt-learning-based method
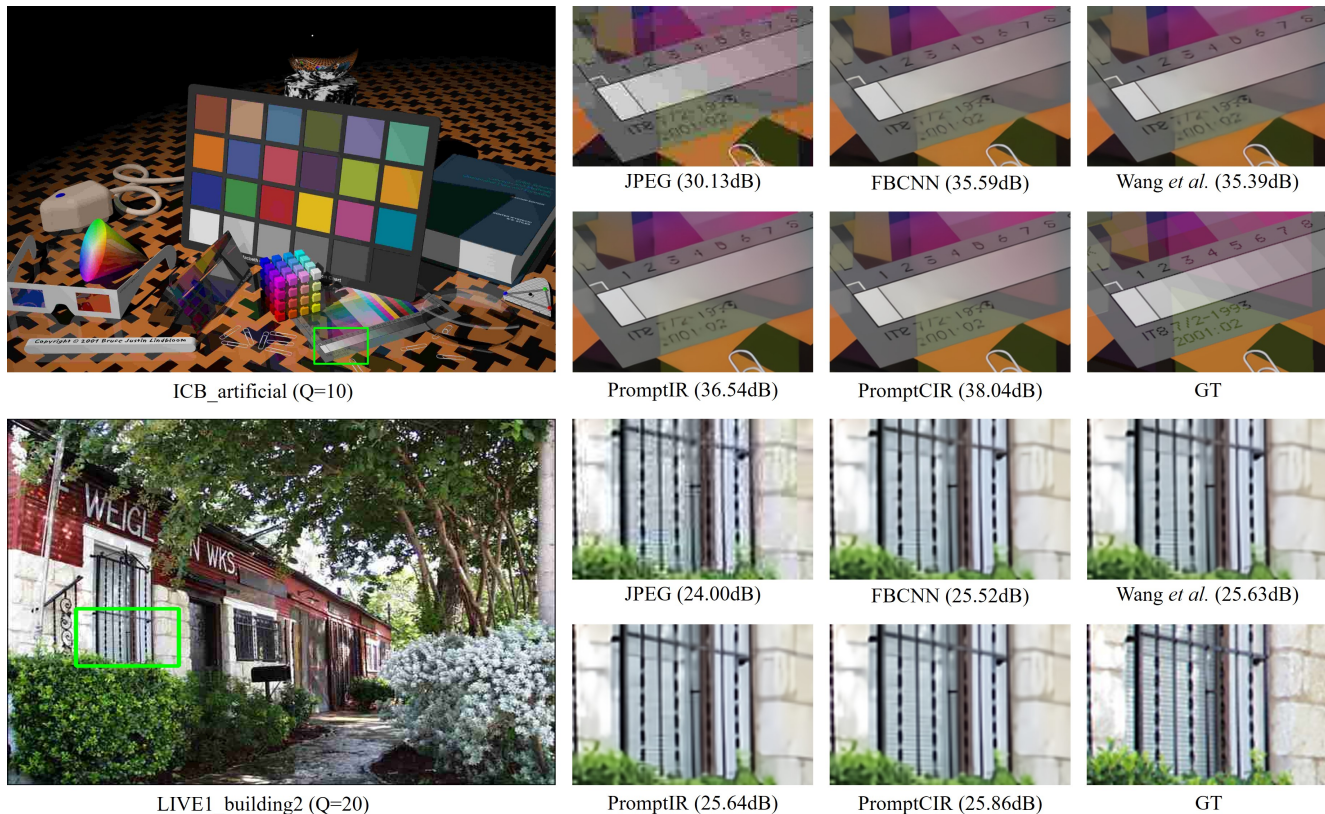
Figure 6. Qualitative comparisons between different methods on ICB with quality factor of 10 and LIVE1 with quality factor of 20. Our PromptCIR achieves visual pleasant restoration results. Zoom in for better views.

PromptIR [43]. Notice that, as the released PromptIR are not trained on CIR, we retrain PromptIR with the same strategy as PromptCIR to ensure a fair comparison, denoted as PromptIR†. We include three compression benchmarks LIVE1 [45], BSDS500 [40] and ICB [44]. The high-quality images are randomly compressed by a quality factor ranging from 10 to 70, following the NTIRE 2024 [66] implementation. Furthermore, we adopt the NTIRE 2024 official-released blind DIV2K validation set for evaluation. The results are tested on the RGB channel with PSNR, SSIM, and PSNRB. It is noteworthy that, for a fair comparison on the scale of training datasets, we also train PromptIR†and our PromptCIR with only DF2K, denoted as PromptIR†* and PromptCIR*.

As demonstrated in Tab. 1, we have two observations: 1) prompt-learning-based methods have significant advantages on high-resolution datasets, *i.e.*, ICB [44] and DIV2K [1]. We attribute the reason to the strong representation ability of prompts in modeling content-aware information, which surpasses that of numerical quality factor prediction and contrastive learning. 2) Benefiting from window-based attention, our PromptCIR achieves better performance than PromptIR, which indicates the importance of extracting lo-

cal information. For qualitative results shown in Fig. 5, we can observe that PromptCIR is capable of restoring texture details without blurring or blocking effects.

## 4.3. Non-blind CIR Evaluation

Non-blind CIR targets for evaluating model performance at a specific compression level, while the model is still *trained in a blind way*. It serves as an important validation baseline for blind methods. Following previous works [8, 14, 58], we evaluate our PromptCIR on LIVE1 [45] and ICB [44] with quality factors [10,20,30,40]. The results are listed in Tab. 2. We can observed that, apart from PromptCIR, PromptCIR* achieves the highest performance on almost all quality factors and datasets. As anticipated, owing to the utilization of a large dataset for training, PromptCIR has realized an additional average performance PSNR gain of about 0.3dB compared to PromptCIR*. This performance gain is inseparable from the content-aware and distortion-aware adaptability of the dynamic prompts [21], as well as the robust representation capabilities of the RHAG [5] module. The subjective results are shown in Fig. 6. PromptCIR restores characters and building textures more clearly and vividly, while achieving an average 0.86dB performance

Table 3. Quantitative comparison for different scales of training datasets. Results are tested in terms of PSNR↑/SSIM↑/PSNRB↑ on blind validation sets. We highlight the performance gain of PSNR by ↑value for more intuitive comparisons.

| Dataset | Method | DF2K | DF2K+LSDIR |
|---------|--------|------|-----------|
| LIVE1 | PromptIR | 32.34/0.907/31.75 | $32.42_{\uparrow 0.08}$/0.908/31.83 |
| | Ours | 32.45/0.908/31.87 | $32.65_{\uparrow 0.11}$/0.910/32.06 |
| BSDS500 | PromptIR | 32.67/0.907/31.67 | $32.76_{\uparrow 0.09}$/0.908/31.72 |
| | Ours | 32.74/0.907/31.76 | $32.92_{\uparrow 0.18}$/0.910/31.83 |
| ICB | PromptIR | 36.35/0.868/36.75 | $36.54_{\uparrow 0.19}$/0.870/36.93 |
| | Ours | 36.41/0.870/36.79 | $36.70_{\uparrow 0.29}$/0.781/37.09 |
| DIV2K | PromptIR | 34.38/0.922/33.81 | $34.52_{\uparrow 0.14}$/0.923/33.90 |
| | Ours | 34.45/0.922/33.88 | $34.68_{\uparrow 0.23}$/0.925/34.08 |

gain compared to PromptIR.

## 4.4. Ablation Studies

**The Effectiveness of Large-scale Training Datasets.** In IR tasks, models are usually optimized in supervised ways [23, 29, 30, 71, 72]. Typically, these models learn the pixel-wise or distribution-wise [31, 53, 56, 57, 63, 73] relationships between distorted and clean image pairs through training datasets. However, in real-world scenarios, such relationships are complex and diverse. Merely relying on a few thousand training images is insufficient for the model to learn the correspondence between distorted and clean images applicable to real-world scenarios [27]. Therefore, a larger training dataset has become crucial needs in the field of IR. Luckily, recent work LSDIR [27] has introduced a large-scale training dataset consisting of 84991 high-quality images, which is a hundred times larger than popular DIV2K [1] training dataset, significantly boost the performance of IR networks. Inspired by this, we explore the performance of PromptCIR by optimizing it with LSDIR. The quantitative comparisons are given in Tab. 3. As demonstrated, large-scale training datasets confer two significant advantages: 1) It is noteworthy that with increased resolution and diversity of the dataset (*e.g.*, ICB with $5k \times 3k$ resolution), we observe a larger performance gain. This can be attributed to the fact that larger-scale training datasets offer more rich texture details, thereby enhancing the abilities of models to establish complex mappings between compressed and clean images. 2) As demonstrated in Sec. 3.2, we employ RHAG [5] for its enhanced representation capabilities in extracting both local and global information. It naturally follows that RHAG stands to achieve more performance gain than the transposed transformer block [43, 71] from larger-scale training datasets. We provide qualitative comparisons in Fig. 7. As illustrated, PromptCIR trained with LSDIR achieves better texture restoration qualities.

**The effectiveness of RHAG and DPM.** As demonstrated in Sec. 3.1 and Sec. 3.2, blind CIR requires network to possess powerful extraction abilities for both local and global content-aware and distortion-aware information. To explore
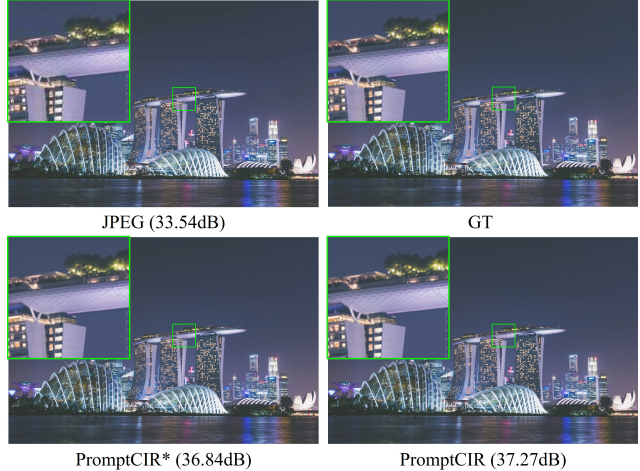


Figure 7. Qualitative comparisons between different training dataset scales. Zoom in for better views. (DIV2K_0814)

Table 4. Quantitative comparison for RHAG and DPM. Results are tested in terms of PSNR↑/SSIM↑ on blind DIV2K set. We adopt ✓ to indicate the usage of specific modules.

| Method | RHAG | DPM | PSNR/SSIM |
|--------|------|-----|-----------|
| PromptIR | | | 34.52/0.923 |
| PromptCIR | | ✓ | 34.56/0.924 |
| | ✓ | | 34.62/0.924 |
| | ✓ | ✓ | **34.68/0.925** |

the effectiveness of RHAG [5] and DPM [21], we conduct ablation studies in Tab. 4. Based on the quantitative results, we have validated that: 1) Compared to the original prompt design of PromptIR [43], DPM [21] achieves better performance with dynamic prompt bases. 2) Compared to the original transformer block [71], RHAG enjoys stronger representation abilities for CIR task.

## 5. Conclusion

In this paper, we present PromptCIR, the prompt-learning-based CIR model, which adopts lightweight prompts as effective content-aware and distortion-aware information guidance for blind CIR task. Different from existing prediction-based networks that estimate numerical quality factors, our PromptCIR has advantages of spatial-wise adaptabilities with dynamic prompt bases. Moreover, we explore the effectiveness of large-scale training dataset, which is capable of further boosting the qualities of restored images. Our method has achieved first place in the NTIRE 2024 challenge of blind compressed image enhancement track. Extensive experiments on blind and non-blind benchmarks have demonstrated the superiority of proposed PromptCIR.

# References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 126–135, 2017. 5, 7, 8

[2] Yuang Ai, Huaibo Huang, Xiaoqiang Zhou, Jiexiang Wang, and Ran He. Multimodal prompt perceiver: Empower adaptiveness, generalizability and fidelity for all-in-one image restoration. *arXiv preprint arXiv:2312.02918*, 2023. 3

[3] Benjamin Bross, Ye-Kui Wang, Yan Ye, Shan Liu, Jianle Chen, Gary J Sullivan, and Jens-Rainer Ohm. Overview of the versatile video coding (vvc) standard and its applications. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(10):3736–3764, 2021. 1

[4] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020. 2

[5] Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22367–22377, 2023. 2, 3, 4, 5, 7, 8

[6] Marcos V Conde, Ui-Jin Choi, Maxime Burchi, and Radu Timofte. Swin2sr: Swinv2 transformer for compressed image super-resolution and restoration. *arXiv preprint arXiv:2209.11345*, 2022. 1, 2, 4

[7] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 1

[8] Max Ehrlich, Larry Davis, Ser-Nam Lim, and Abhinav Shrivastava. Quantization guided jpeg artifact correction. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16*, pages 293–309. Springer, 2020. 2, 7

[9] Zhiyi Feng, Feifei Lee, and Qiu Chen. Srunet: stacked reversed u-shape network for lightweight single image super-resolution. *IEEE Access*, 10:60151–60162, 2022. 2

[10] Xueyang Fu, Zheng-Jun Zha, Feng Wu, Xinghao Ding, and John Paisley. Jpeg artifacts reduction via deep convolutional sparse coding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019. 2

[11] Jun Guo and Hongyang Chao. Building dual-domain representations for compression artifacts reduction. In *European Conference on Computer Vision*, pages 628–644. Springer, 2016. 2

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 4

[13] Menglin Jia, Luming Tang, Bor-Chun Chen, Claire Cardie, Serge Belongie, Bharath Hariharan, and Ser-Nam Lim. Visual prompt tuning. In *European Conference on Computer Vision*, pages 709–727. Springer, 2022. 2

[14] 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, 2, 5, 6, 7

[15] Yoonsik Kim, Jae Woong Soh, Jaewoo Park, Byeongyong Ahn, Hyun-Seung Lee, Young-Su Moon, and Nam Ik Cho. A pseudo-blind convolutional neural network for the reduction of compression artifacts. *IEEE Transactions on circuits and systems for video technology*, 30(4):1121–1135, 2019. 2

[16] Yoonsik Kim, Jae Woong Soh, and Nam Ik Cho. Agarnet: Adaptively gated jpeg compression artifacts removal network for a wide range quality factor. *IEEE Access*, 8: 20160–20170, 2020. 1, 2

[17] Bingchen Li, Xin Li, Yiting Lu, Sen Liu, Ruoyu Feng, and Zhibo Chen. Hst: Hierarchical swin transformer for compressed image super-resolution. *arXiv preprint arXiv:2208.09885*, 2022. 1, 2

[18] Bingchen Li, Xin Li, Hanxin Zhu, Yeying Jin, Ruoyu Feng, Zhizheng Zhang, and Zhibo Chen. Sed: Semantic-aware discriminator for image super-resolution. *arXiv preprint arXiv:2402.19387*, 2024. 3

[19] Honggui Li, Maria Trocan, Mohamad Sawan, and Dimitri Galayko. Cswin2sr: Circular swin2sr for compressed image super-resolution. *arXiv preprint arXiv:2301.08749*, 2023. 2, 4

[20] K Li, Y Wang, J Zhang, P Gao, G Song, Y Liu, H Li, and Y Qiao. Uniformer: Unifying convolution and self-attention for visual recognition. arxiv 2022. *arXiv preprint arXiv:2201.09450*, . 4

[21] Xin Li, Bingchen Li, Yeying Jin, Cuiling Lan, Hanxin Zhu, Yulin Ren, and Zhibo Chen. Ucip: A universal framework for compressed image super-resolution using dynamic prompt. . 2, 3, 5, 7, 8

[22] Xin Li, Xin Jin, Jianxin Lin, Sen Liu, Yaojun Wu, Tao Yu, Wei Zhou, and Zhibo Chen. Learning disentangled feature representation for hybrid-distorted image restoration. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16*, pages 313–329. Springer, 2020. 1

[23] Xin Li, Simeng Sun, Zhizheng Zhang, and Zhibo Chen. Multi-scale grouped dense network for vvc intra coding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 158–159, 2020. 8

[24] Xin Li, Jun Shi, and Zhibo Chen. Task-driven semantic coding via reinforcement learning. *IEEE Transactions on Image Processing*, 30:6307–6320, 2021. 1

[25] Xin Li, Bingchen Li, Xin Jin, Cuiling Lan, and Zhibo Chen. Learning distortion invariant representation for image restoration from a causality perspective. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1714–1724, 2023. 1

[26] Xin Li, Yulin Ren, Xin Jin, Cuiling Lan, Xingrui Wang, Wenjun Zeng, Xinchao Wang, and Zhibo Chen. Diffusion models for image restoration and enhancement–a comprehensive survey. *arXiv preprint arXiv:2308.09388*, 2023. 1

[27] Yawei Li, Kai Zhang, Jingyun Liang, Jiezhang Cao, Ce Liu, Rui Gong, Yulun Zhang, Hao Tang, Yun Liu, Denis Demandolx, et al. Lsdir: A large scale dataset for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1775–1787, 2023. 2, 5, 8

[28] Zilong Li, Yiming Lei, Chenglong Ma, Junping Zhang, and Hongming Shan. Prompt-in-prompt learning for universal image restoration. *arXiv preprint arXiv:2312.05038*, 2023. 1, 2, 3

[29] 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. 1, 2, 4, 8

[30] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 8

[31] Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Ben Fei, Bo Dai, Wanli Ouyang, Yu Qiao, and Chao Dong. Diffbir: Towards blind image restoration with generative diffusion prior. *arXiv preprint arXiv:2308.15070*, 2023. 8

[32] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *Advances in neural information processing systems*, 36, 2024. 2, 3

[33] Jianzhao Liu, Jianxin Lin, Xin Li, Wei Zhou, Sen Liu, and Zhibo Chen. Lira: Lifelong image restoration from unknown blended distortions. In *European Conference on Computer Vision*, pages 616–632. Springer, 2020. 1

[34] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021. 1

[35] Ze Liu, Han Hu, Yutong Lin, Zhuliang Yao, Zhenda Xie, Yixuan Wei, Jia Ning, Yue Cao, Zheng Zhang, Li Dong, et al. Swin transformer v2: Scaling up capacity and resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12009–12019, 2022. 1

[36] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 6

[37] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Controlling vision-language models for universal image restoration. *arXiv preprint arXiv:2310.01018*, 2023. 1, 2, 3

[38] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Refusion: Enabling large-size realistic image restoration with latent-space diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1680–1691, 2023. 2

[39] Jiaqi Ma, Tianheng Cheng, Guoli Wang, Qian Zhang, Xinggang Wang, and Lefei Zhang. Prores: Exploring degradation-aware visual prompt for universal image restoration. *arXiv preprint arXiv:2306.13653*, 2023. 1, 2

[40] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, pages 416–423. IEEE, 2001. 7

[41] OpenAI. Gpt-4 technical report, 2023. 2, 3

[42] Yingxue Pang, Xin Li, Xin Jin, Yaojun Wu, Jianzhao Liu, Sen Liu, and Zhibo Chen. Fan: Frequency aggregation network for real image super-resolution. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 468–483. Springer, 2020. 1

[43] Vaishnav Potlapalli, Syed Waqas Zamir, Salman Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one blind image restoration. *arXiv preprint arXiv:2306.13090*, 2023. 1, 2, 3, 4, 5, 6, 7, 8

[44] Rawzor. Image compression benchmark. 7

[45] H Sheikh. Live image quality assessment database release 2. *http://live. ece. utexas. edu/research/quality*, 2005. 7

[46] Kihyuk Sohn, Huiwen Chang, José Lezama, Luisa Polania, Han Zhang, Yuan Hao, Irfan Essa, and Lu Jiang. Visual prompt tuning for generative transfer learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19840–19851, 2023. 2

[47] Binbin Song, Xiangyu Chen, Shuning Xu, and Jiantao Zhou. Under-display camera image restoration with scattering effect. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12580–12589, 2023. 2

[48] Gary J Sullivan, Jens-Rainer Ohm, Woo-Jin Han, and Thomas Wiegand. Overview of the high efficiency video coding (hevc) standard. *IEEE Transactions on circuits and systems for video technology*, 22(12):1649–1668, 2012. 1

[49] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, and Lei Zhang. Ntire 2017 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 114–125, 2017. 5

[50] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023. 2, 3

[51] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 4

[52] Gregory K Wallace. The jpeg still picture compression standard. *Communications of the ACM*, 34(4):30–44, 1991. 1

[53] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin CK Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. *arXiv preprint arXiv:2305.07015*, 2023. 8

[54] Menglu Wang, Xueyang Fu, Zepei Sun, and Zheng-Jun Zha. Jpeg artifacts removal via compression quality ranker-guided networks. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 566–572, 2021. 1, 2

[55] Tao Wang, Wanglong Lu, Kaihao Zhang, Wenhan Luo, Tae-Kyun Kim, Tong Lu, Hongdong Li, and Ming-Hsuan Yang.

Promptrr: Diffusion models as prompt generators for single image reflection removal. *arXiv preprint arXiv:2402.02374*, 2024. 2

[56] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018. 8

[57] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1905–1914, 2021. 8

[58] Xi Wang, Xueyang Fu, Yurui Zhu, and Zheng-Jun Zha. Jpeg artifacts removal via contrastive representation learning. In *European Conference on Computer Vision*, pages 615–631. Springer, 2022. 1, 5, 6, 7

[59] Zhangyang Wang, Ding Liu, Shiyu Chang, Qing Ling, Yingzhen Yang, and Thomas S Huang. D3: Deep dual-domain based fast restoration of jpeg-compressed images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2764–2772, 2016. 2

[60] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17683–17693, 2022. 1, 2

[61] Haiping Wu, Bin Xiao, Noel Codella, Mengchen Liu, Xiyang Dai, Lu Yuan, and Lei Zhang. Cvt: Introducing convolutions to vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 22–31, 2021. 4

[62] Yaojun Wu, Xin Li, Zhizheng Zhang, Xin Jin, and Zhibo Chen. Learned block-based hybrid image compression. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(6):3978–3990, 2021. 1

[63] Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc Van Gool. Diffir: Efficient diffusion model for image restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13095–13105, 2023. 8

[64] Tete Xiao, Mannat Singh, Eric Mintun, Trevor Darrell, Piotr Dollár, and Ross Girshick. Early convolutions help transformers see better. *Advances in neural information processing systems*, 34:30392–30400, 2021. 4

[65] Qinhong Yang, Dongdong Chen, Zhentao Tan, Qiankun Liu, Qi Chu, Jianmin Bao, Lu Yuan, Gang Hua, and Nenghai Yu. Hq-50k: A large-scale, high-quality dataset for image restoration. *arXiv preprint arXiv:2306.05390*, 2023. 2

[66] Ren Yang, Radu Timofte, et al. NTIRE 2024 challenge on blind enhancement of compressed image: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2024. 2, 7

[67] Fanghua Yu, Jinjin Gu, Zheyuan Li, Jinfan Hu, Xiangtao Kong, Xintao Wang, Jingwen He, Yu Qiao, and Chao Dong. Scaling up to excellence: Practicing model scaling for photo-realistic image restoration in the wild. *arXiv preprint arXiv:2401.13627*, 2024. 2

[68] Gaocheng Yu, Jin Zhang, Zhe Ma, and Hongbin Wang. Efficient progressive high dynamic range image restoration via attention and alignment network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1124–1131, 2022. 2

[69] Ke Yu, Chao Dong, Chen Change Loy, and Xiaoou Tang. Deep convolution networks for compression artifacts reduction. *arXiv preprint arXiv:1608.02778*, 2016. 1, 2

[70] Kun Yuan, Shaopeng Guo, Ziwei Liu, Aojun Zhou, Fengwei Yu, and Wei Wu. Incorporating convolution designs into visual transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 579–588, 2021. 4

[71] 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, 2, 3, 4, 5, 8

[72] 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. 8

[73] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4791–4800, 2021. 8

[74] Xiaoshuai Zhang, Wenhan Yang, Yueyu Hu, and Jiaying Liu. Dmcnn: Dual-domain multi-scale convolutional neural network for compression artifacts removal. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 390–394. IEEE, 2018. 2

[75] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2472–2481, 2018. 2

[76] 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. 2

[77] 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. 2

[78] Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric. P Xing, Hao Zhang, Joseph E. Gonzalez, and Ion Stoica. Judging llm-as-a-judge with mt-bench and chatbot arena, 2023. 2