NTIRE 2024 Challenge on Blind Enhancement of Compressed Image: Methods and Results

Ren Yang* Radu Timofte* Bingchen Li Xin Li Mengxi Guo Shijie Zhao Li Zhang Zhibo Chen Dongyang Zhang Yash Arora Aditya Arora Yuanbin Chen Hui Tang Tao Wang Longxuan Zhao Bin Chen Tong Tong Qiao Mo Jingwei Bao Jinhua Hao Yukang Ding Hantang Li Ming Sun Chao Zhou Shuyuan Zhu Zhi Jin Wei Wang Dandan Zhan Jiawei Wu Jiahao Wu Luwei Tu Hongyu An Xinfeng Zhang Woon-Ha Yeo Wang-Taek Oh Young-Il Kim Han-Cheol Ryu Long Sun Mingjun Zhen Jinshan Pan Jiangxin Dong and Jinhui Tang Yapeng Du Ao Li Ziyang He Lei Luo Ce Zhu Xin Yao Sunder Ali Khowaja IK Hyun Lee Jaeho Lee Seongwan Kim Sharif S M A Nodirkhuja Khujaev Roman Tsoy

Abstract

This paper reviews the Challenge on Blind Enhancement of Compressed Image at NTIRE 2024, which aims at enhancing the quality of JPEG images which are compressed with unknown quality factor. The challenge requires that the total size of codes and pre-trained model(s) cannot exceed 300 MB, since we encourage solutions for blind enhancement with generalized models, instead of separately training several models for each quality factor. In this report, we summarize the detailed settings of the challenge, the final results, and the solutions proposed by the participants. The challenge has 129 registered participants and received 13 valid submissions. ***Several teams (including all TOP 3 teams) have publicly released the codes (see Sec. 4)****. They gauge the state-of-the-art of blind quality enhancement of compressed image.*

1. Introduction

Compression plays an important role in the efficient transmission of images through the limited-bandwidth Internet. However, image compression unavoidably leads to compression artifacts, which can severely degrade the visual quality. Therefore, quality enhancement of compressed image has become a popular research topic. This challenge focuses on the scenario of the most commonly used image compression standard JPEG, and aims at enhancing JPEG images with unknown quality factor.

In the past decade, a great number of works were proposed for the reduction of JPEG artifacts [11,12,15,35,48]. Among them, the FBCNN method [15] was proposed for blind JPEG enhancement, *i.e.*, improving the quality of JPEG images at various unknown quality factors with one model. Therefore, we use FBCNN as a baseline method. In this challenge, we totally received 13 valid submissions, in which 11 submissions achieve better performance compared to FBCNN, gauging the state-of-the-art of blind compressed image enhancement. Especially, the top 3 teams have publicly released their codes and models, which makes their solutions convincing and we believe they are beneficial for academic research in the future. Please refer to the links provided in Section 3.

This challenge is one of the NTIRE 2024 Workshop [1] associated challenges on: dense and non-homogeneous dehazing [2], night photography rendering [4], blind compressed image enhancement [44], shadow removal [38], efficient super resolution [33], image super resolution (×4) [8], light field image super-resolution [41], stereo image super-resolution [39], HR depth from images of specular and transparent surfaces [45], bracketing image restoration and enhancement [49], portrait quality assessment [6], quality assessment for AI-generated content [26], restore any image model (RAIM) in the wild [24], RAW image super-resolution [10], short-form UGC video quality assessment [20], low light enhancement [27], and RAW burst alignment and ISP challenge.

*Ren Yang and Radu Timofte are the organizers of this challenge. Others are participants to the challenge. This challenge is part of the NTIRE 2024 workshop: https://cvlai.net/ntire/2024.

[1] https://cvlai.net/ntire/2024/

Table 1. The final results of the challenge. We use FBCNN [15] as a baseline method.

Team	PSNR (dB)	Model (MB)	Speed (s)	Hardware	Ensemble	Extra training data
IMCL-BVQE	34.4749	270	51.89	Tesla V100	2×rotation, 2×models	Flickr2K [36] and LSDIR [22]
PixelArtAI	34.3507	269	10	RTX 3090	4×flip	See Sec. 4.2
Titans	34.3205	110	240	A100	2×flip, tiling	Flickr2K [36] and LSDIR [22]
BinYCn	34.2403	39	162.94	RTX 4090	4×flip/rotation, tiling	-
OldFe666	34.2316	63	520	A800	8×flip/rotation, tiling	BSDS500 [3], Flickr2K [36], WED [30], HQ-50K [43]
FVL-T1	34.2181	104	25	A800	8×flip/rotation	-
UCAS_SCST	34.0993	269	72	A100	8×flip/rotation, tiling	Flickr2K [36] and LSDIR [22]
SYU-HnVLab	33.9784	135	3.73	Tesla V100	8×flip/rotation	-
VPEG	33.9645	185	2.3	RTX 3090	8×flip/rotation	LSDIR [22]
Tempest	33.9388	122	4.5	RTX 3080	-	LSDIR [22]
FlyingBunny	33.9241	275	14.2	RTX 4090	8×flip/rotation	Flickr2K [36]
FBCNN [15]	33.6676	275				
AIVerse	33.5772	274	2.5	RTX 3090Ti	-	-
Unicorns	33.1474	102	0.3	A100	-	-
JPEG	31.2827					

2. NTIRE 2024 Challenge

In this section, we introduce the detailed settings of the challenge, including the datasets, the JPEG settings and the regulations of the challenge.

2.1. Dataset

The DIV2K [1] dataset consists of 1,000 high-resolution images with diverse contents. In this challenge, we use validation (100 images) and test (100 images) sets of DIV2K for validation and test, respectively. We provide the training set (800 images) of DIV2K as an example training set, and the participants are allowed to employ more datasets for training their models. The commonly used training sets include Flickr2K [36] and LSDIR [22]. In addition, other datasets are also used by a few teams.

During the development phase, the participants are not allowed to use the LIVE1 dataset for training, since it is used as a cross-validation set. We observed that the rank of the top 4 teams is exactly the same when testing their codes on LIVE1, validating the effectiveness, reproducibility, and generalizability of the prize winners.

2.2. JPEG settings

In this challenge, we use the Pillow library in Python to produce JPEG images, which are with random quality factor ranging from 10 to 70. The following shoes the Python codes for compressing one image:

```
import PIL
from PIL import Image
from PIL.features import check_feature
import random

assert(PIL.__version__=="10.0.1")
assert(Image.core.jpeglib_version=="9.0")
assert not check_feature("libjpeg_turbo")

img = Image.open('img.png')
qf = random.randint(10, 71)
img.save('img.jpg', "JPEG", quality=qf)
```

2.3. Challenge regulations

This challenge encourages participants to propose overall solutions to the enhancement of blind compressed images, instead of training separate models to each quality factor. Therefore, the challenge requires that the total size of submitted codes and models cannot exceed 300 MB. Besides, to ensure the fairness, the challenges requires the participants to submit their codes and models before the test phase begins, and the models are prohibited to be fine-tuned or altered during the test phase.

3. Challenge results

The challenge results are shown in Table 1. As we can see from Table 1, all methods proposed in this challenge achieves $>$ 2 dB PSNR enhancement of the JPEG inputs, and 11 out of the 13 teams achieve superior performance than the baseline method FBCNN [15]. Besides, the model sizes of all solutions are comparable or smaller than FBCNN. Table 1 also shows that the self-ensemble strategy [37] is widely used in the top teams, and 8 teams employed extra training data to boost the performance.

The IMCL-BVQE Team, PixelArtAI Team and Titans Team rank first, second, and third, respectively. The model sizes of IMCL-BVQE and PixelArtAI are comparable, while PixelArtAI achieves a much faster inference speed. Titans uses a smaller model, but their method works with lower speed.

Note that the model size indicated the total size of codes and model(s) submitted by each team before the test phase. The running time is reported in the factsheet of each team.

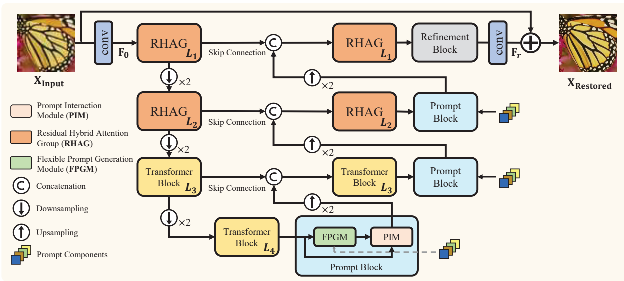

Figure 1. The PromptCIR approach [18] of IMCL-BVQE.

4. Teams and Methods

The codes of the TOP 3 teams are publicly released:
IMCL-BVQE: https://github.com/lbc12345/PromptCIR-NTIRE24
PixelArtAI: https://github.com/zdyshine/BasicSR-NTIRE2024-Compressed-Image
Titans: https://github.com/yasharora102/UnifyFormer

Publicly released codes of other teams:
VPEG: https://github.com/sunny2109/SAFMN
AIVerse: https://github.com/sander-ali/JPEG_Compression_Enhancement_RDAB

4.1. IMCL-BVQE Team

The IMCL-BVQE team proposed the PromptCIR approach [18], which implicitly perceives the quality factor while restoring image details through prompt learning.

Following PromptIR [32], they design a U-shape like network architecture for blind CIR, as illustrated in Fig. 1. To better capture the local information for restoration, they leverage overlapping cross-attention module and hybrid attention mechanism [7] in the first two down-sampling and up-sampling stages of PromptCIR. Transformer Block [32,46] is utilized to capture global pixel relations in deeper stages. Prompt is introduced in the up-sampling stages to facilitate the restoration of images. Specifically, prompt will first interact with image features to generate compression-adaptive prompt features. Then, prompt features are concatenate with image features to further provide implicit but flexible guidance for restoring images. To mitigate producing suboptimal enhancement results due to the interpolation operation in the original prompt generation process [32], they adopt a Flexible Prompt Generation Module (FPGM) [19]. Compared with original design, FPGM has fewer parameters while achieving higher performance.

The IMCL-BVQE team adopts two-stage training strategy to optimize our model. In the pre-training stage, apart from provided training images, they generate compressed images with 7 quality factors [10,20,30,40,50,60,70] for Flickr2K [36] and LSDIR [22] to enrich the training datasets. The training images are paired-cropped into 128×128 patches, and augmented by random horizontal flips, vertical flips and rotations. The AdamW optimizer is used to optimize the model parameters with an initial learning rate of 2×10^{-4}. A Cosine Annealing [28] scheduler is employed to decay the learning rate, with the total number of iterations set to 800k.

In the fine-tuning stage, they optimize our model using compressed image pairs generated online with randomly selected quality factors in range [10, 70]. The initial learning rate is set to 1e-4, and the total number of iterations is set to 600k. To further boost the performance, they increase the patch size to 192 to tune the model parameters for another 200k iterations with a fixed learning rate of 1×10^{-5}.

4.2. PixelArtAI Team

The PixelArtAI team proposes a blind JPEG artifacts removal method via enhanced Swin-Conv-UNet, which is developed based on SCUNet [47] with several modifications to enhance the model's performance while maintaining a maximum model size of 300 MB.

Firstly, they increased the number of channels in the model from [64, 128, 256, 512] to [96, 192, 384, 768]. Secondly, they increased the number of downsampling and upsampling modules in the model. This changes in SCUNet allows the model to capture more features and details in the input data. By increasing the model parameters, the model has gained stronger learning capabilities. Compared to the baseline, the performance metrics have been significantly improved.

The PixelArtAI team uses DIV2K [1] and Flickr2K [1], DIV8K [13], LSDIR [22], Unsplash2K, nomos8k and nomosuni as training data. The training data is degraded online (i.e. random JPEG compression). The input image size is $128 \times 128 \times 3$, the batch size is 16. Adam optimizer with the initial learning rate set to 0.001. The training is divided into two stages: First, the learning rate is 1×10^{-4} and the loss is L1. This stage is trained for 60k iterations. Second, only the PSNR Loss is calculated, and the input image size is $256 \times 256 \times 3$, the initial learning rate set to 0.0002,and is halved by 20k iterations.

4.3. Titans Team

The Titans team introduces UnifyFormer, based on recently proposed Restormer [46]. Over the base architecture, they proposed to improve over the self-attention module by capturing global and local context information which furthermore shows significant results on blind image compression task.

Instead of a single convolution, they use multiple convolution and activations as the Patch Embedding block. This early visual processing is a crucial design choice that strikes

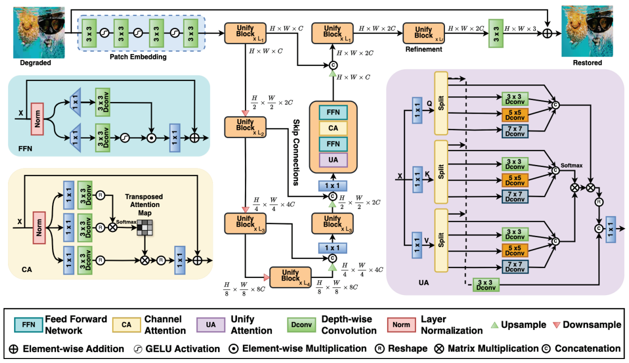

Figure 2. Framework of the Unifyformer method of Titans.

a balance between (hard) inductive biases and the representation learning ability of transformer blocks. Central to our approach is the integration of an innovative component known as `UnifyBlock`, which seamlessly enhances the performance of the overall transformer-based network. This section delves into the details of UnifyFormer's operation, emphasizing the UnifyBlocks' architecture and their role in optimizing the overall performance of the framework.

4.3.1 General method description

Overall pipeline. Fig. 2 presents the overall pipeline of our UnifyFormer architecture. UnifyFormer is designed to operate on low-quality compressed input images, extracting and refining features to reconstruct high-quality counterparts. Initially, the input image $I \in \mathbb{R}^{H\times W\times 3}$ undergoes multiple convolutions to obtain low-level feature embeddings. These embeddings are then processed through an encoder-decoder structure, which consists of our novel UnifyBlocks. The key innovation lies in these blocks, which are strategically placed to enhance the model's attention mechanism, ensuring a superior balance between global and local information processing.

In the proposed UnifyFormer, the core components are: **(a)** Unify Attention (UA), **(b)** Channel Attention (CA) and **(c)** Feed-Forward Network (FFN).

To harness the full potential of Unify Attention, the Titans team strategically sequences its operation with the Channel Attention and Feed-Forward Network components from Restormer. This integration is pivotal, as it combines the strengths of each component to achieve a nuanced amalgamation of spatial and channel-wise analysis. The two attention blocks complement each other as CA focuses on the inter dependencies between channels, while UA enhances local and global context awareness. Simultaneously, FFN further refines the feature representations of CA and UA by leveraging its gating mechanism, emphasizing spatial context and thus, enhancing the restoration quality of intricate image details.

Unify Attention The architecture of Unify Attention is what distinguishes UnifyFormer from traditional approaches. Unify Attention consists of several operations tailored to optimize the compression and subsequent reconstruction of images:

Token Segmentation and Unification: The first step involves segmenting the spatial tokens and then applying a unification operation. This operation transforms individual token features into unified group representations, or "unify proxies," using depth-wise convolutions of varied kernel sizes. This process is designed to capture diverse local patterns within each group, effectively encoding more information into these unified representations.

Modified Attention Mechanism: Unlike standard self-attention that processes queries (Q), keys (K), and values (V) at the individual token level, our approach modifies these components to integrate the unify proxies. This modification facilitates a more nuanced, group-wise attention, enabling the model to understand and preserve complex patterns and textures in the image more effectively.

Global and Local Context Integration: After the attention phase, the model recombines the unify proxies with the original token features. This critical step ensures that both detailed local information and broader contextual insights are maintained and enhanced. This integration is pivotal for recovering intricate details in the image during the decompression phase, leading to higher fidelity in the reconstructed HR images.

The inclusion of Unify Attention within the encoder-decoder framework markedly improves the model's ability to enhance and reconstruct images. By emphasizing both global context and local detail, UnifyFormer preserves essential image content through the compression process. Subsequent stages leverages the enriched feature set, leading to the production of a residual image R, which, when added to the initial degraded input, yields the restored HR image $\hat{I} = I + R$.

Channel Attention Introduced in [46], CA has linear complexity. The key ingredient is to apply SA across channels rather than the spatial dimension, i.e., to compute cross-covariance across channels to generate an attention map encoding the non-local context implicitly. Also the depth-wise convolutions emphasize on the local context before computing feature covariance to produce the global attention map.

Feed-Forward Network Followed by each UA and CA is the FFN component. This further processes the information, utilizing a gating mechanism to enhance information

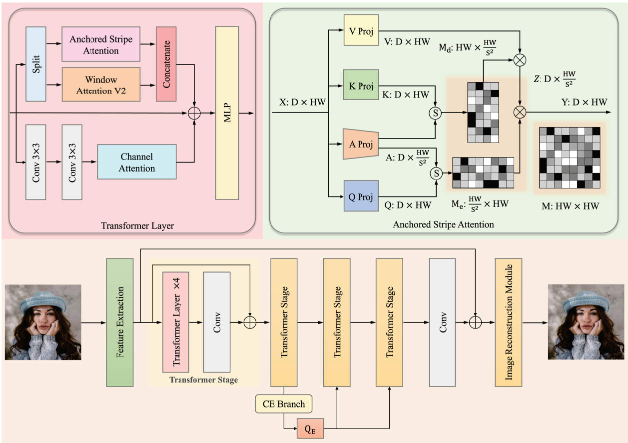

Figure 3. Illustration of the GRL-CIC proposed by BinYCn.

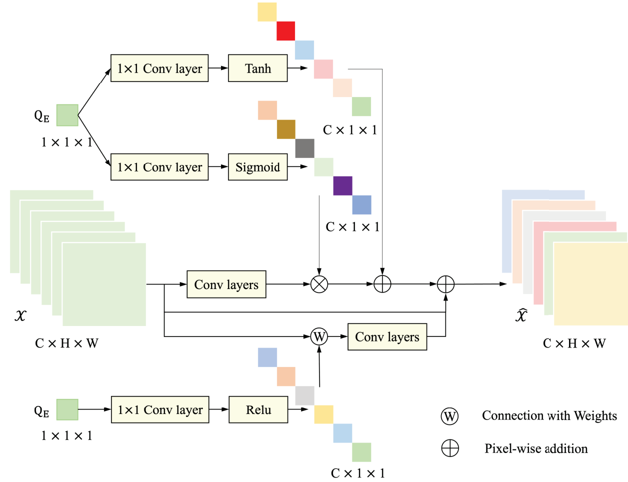

Figure 4. Illustration of the CIC proposed by BinYCn.

flow and focus on spatial context. This is crucial for capturing and refining local image details, setting the stage for effective restoration.

4.3.2 Training strategy

The Titans team perform progressive learning where the network is trained on smaller image patches in the early epochs and on gradually larger patches in the later training epochs. The model trained on mixed-size patches via progressive learning shows enhanced performance at test time where images can be of different resolutions (a common case in image restoration). The model is trained on the provided 800 training images of the DIV2K dataset [1], 2,650 images of the Flickr dataset [36] and 80,000 images of the LSDIR dataset [22].

In all experiments, the following training parameters are used. From level-1 to level-4, the number of Transformer blocks are [2, 3, 3, 4], attention heads in UA and CA are [1, 2, 4, 8], and number of channels are [48, 96, 192, 384]. The refinement stage contains 4 blocks. The channel expansion factor in FFN is γ=2.66. They train the model with AdamW optimizer (β_1=0.9, β_2=0.999, weight decay $1e^{-4}$) and L_1 loss for 300K iterations with the initial learning rate $3e^{-4}$ gradually reduced to $1e^{-6}$ with the Cosine Annealing scheme [28].

For progressive learning, they start training with patch size 128×128 and batch size 64. The patch size and batch size pairs are updated to [(128^2,48), (160^2,32), (192^2,16), (224^2,16)] at iterations [100K, 170K, 220K, 260K]. For data augmentation, the horizontal and vertical flips are used.

4.4. BinYCn Team

The BinYCn team adopts an image restoration network architecture (GRL-CLC) based on the GRL Transformer [21] and propose a compression information control strategy to estimate the compression quality factor. As illustrated in Fig. 3, the network comprises two main branches: the Transformer branch and the Compression Information Controller (CIC). The Transformer branch consists of a feature extraction layer, GRL Transformer, and an image reconstruction module. The GRL Transformer possesses global, regional, and local range image modeling capabilities, enhancing convolution by parallel computation of anchor stripe attention, window attention, and channel attention. This achieves a balanced modeling of image layers, balancing computational complexity and global dependency modeling capability. The CIC is used to enhance the adaptive ability of compression artifact removal. Only a relatively small prediction branch is added, and the decoder shares parameters for QF estimation and image reconstruction, accelerating the convergence of QF prediction. As shown in Fig. 4, QE is used as additional input to generate gate weights. Then, gate weights are used to rescale feature maps according to different QFs, providing more compression-related information.

During the training phase, the BinYCn team randomly cropped 288 × 288 smaller patches from the processed HR images to serve as the ground truth during the network training process. Subsequently, each of these smaller patches was subjected to JPEG compression to act as the input for the network, with the compression quality factors varying from 10 to 70.

Loss function:

$$L_{total} = L_1(Pred, Y) + \lambda L_1(Pred_q, Q) \quad (1)$$

where L_1 denotes the mean absolute error, with $Pred$ and Y representing the predicted and ground truth images, respectively. Additionally, $Pred_q$ and Q orrespond to the predicted and actual compression quality factors. The parameter λ denotes the loss weight used to balance the influence between the two terms. The model are optimized by the AdamW with β_1 = 0.9 and β_2 = 0.999 with weight decay 1×10^{-8} by default. The initial learning rate was set to 1×10^{-4}, and multi step was chosen as the learning scheme.

4.5. OldFe666 Team

The OldFe666 team adopts ART as an one-stage network for this image restoration task. During the training phase, they use 800 images in training data of DIV2K [1], 200 images in training data of BSDS500 [3], 2650 images of Flickr2K [36] and 4744 images of WED [30] as the training datasets in the first 300k iterations and add other 4000 images from HQ-50K [43] into training data in the rest 100k iterations. In the first 300k iterations, to generally learn from compression degradation with diverse QFs, they use the images in training datasets degraded by JPEG compression with random QFs ranging from 5 to 95 to finetuned the ART model, whose parameters are initialized as the ART official weights on image super-resolution task. And the patchsize and batchsize of input images are set as 126 and 2 respectively, the initial learning rate is 2e-4. While in the rest 100k iterations, they continue to use the compressed images with QFs ranging from 10 to 70 to train the model. The patchsize and batchsize are set as 256 and 1, and the learning rate gets started as 8×10^{-5}. They apply common data augment strategy for improvement, which contains flipping and rotation.

During the testing phase, to reduce the potential loss from improper patchsizes, they use the multi-patchsize and self-ensemble strategies at the same time to improve the effectiveness. The patchsizes are set as 288, 320, and 352.

4.6. FVL-T1 Team

The FVL-T1 team proposes Quality Factor predictor to produce the corresponding QF value of the input, and then inject it into the modified Restormer [46]. The QF Attention Block is added to the original Restormer, to inject the predicted QF values into the extracted features by modulation. During the training phase, they first train the designed Quality Factor Predictor to precisely predict the exact QF value of each compressed image. Then they fix the QF predictor and train Restormer-QF with the predicted QF values.

4.7. UCAS_SCST Team

Transformer-based methods have achieved impressive success in image restoration and enhancement tasks. Inspired by HAT [7], which is an effective Transformer image super-resolution model, the UCAS_SCST team proposes a Transformer image enhancement model, namely High Frequency Transformer (HFT). Based on HAT [7], they removed the super-resolution up-sampling module and incorporate high frequency loss to reduce the compression artifacts. Specifically, they constrained the residuals after Gaussian blurring and images in frequency domain obtained by Discrete Cosine Transform (DCT). Therefore, HFT is more robust to blind JPEG compressed images and reconstructs more high-frequency details.

They utilize the provided DIV2K dataset, additional Flickr2K and LSDIR datasets as training data. At the same time, they employ the rotating and flipping strategies to enhance the above images. To improve the robustness of the model, they add random level JPEG noise. The Adam optimizer ($\beta_1 = 0.9, \beta_2 = 0.99$) is used for 350 iterations on 8 NVIDIA A100 GPUs. The training batch size is set to 8 and the patch size is 64 $\times$ 64. With the input of 64 $\times$ 64 $\times$ 3, the parameters number of HFT is 33.86 M. During the inference phase, they adopt the tile model due to the limited GPU memory. They also apply the self-ensemble strategy to improve the performance.

4.8. SYU-HnVLab Team

The SYU-HnVLab Team participated in this challenge by utilizing the FFTformer [**?**], a module that inversely exploits the JPEG compression process. Additionally, they proposed a module for estimating the JPEG compression's Quality Factor (QF) to achieve further performance improvements.

The application of FFTformer to this competition, focused on JPEG compressed image enhancement, is inherently justified by the core principles underlying both the JPEG compression method and the novel components introduced in FFTformer. JPEG compression, a widely utilized image compression technique, operates by transforming spatial domain information into the frequency domain, selectively quantizing this frequency domain data to reduce file size while attempting to maintain perceptual quality. This process inherently prioritizes certain frequency components over others, often leading to the loss of high-frequency details which are crucial for image sharpness and clarity.

To address the challenge of the *blindness* to the compression quality factor, the SYU-HnVLab team proposes a quality factor estimation module (QFEM). This module functions as a branch that stems from the first level of features in the encoder of the IEM. It is specifically designed to estimate the quality factor based on these initial feature extractions. By incorporating an additional loss calculation related to the quality factor, the module effectively acts like an auxiliary guide. This innovative approach allows for a more nuanced understanding and handling of the compression quality factor during the image enhancement process, significantly improving the model's ability to restore JPEG compressed images with high fidelity and detail.

Prior to training, all images were cropped into patches of 480 $\times$ 480 with a stride of 240, and these data were used

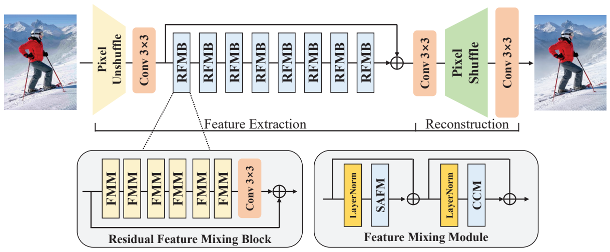

Figure 5. Overview of the model proposed by VPEG.

for training. During training, patches of 480 were randomly cropped again into smaller sizes of 192×192 for the training process. Additionally, images undergo random flipping and rotation with a probability of 0.5 as part of the data augmentation strategy.

The training employs the AdamW optimizer [29], starting with an initial learning rate of $1e^{-3}$. This rate gradually decreases to a minimum of $1e^{-7}$, following a consine annealing scheduler [28]. The goal is to complete a total of 300,000 iterations, utilizing L1 loss and FFT loss as the primary loss functions. FFT loss is given a weight of 0.1. For the estimation of the quality factor, L1 loss is used, with a loss weight set to 0.5.

The dimension of the blocks within the network is configured to 48. The architecture specifies the number of encoding and decoding stages at each level, arranged as [4, 4, 8]. This structured approach to image cropping, coupled with a comprehensive training strategy, aims to optimize the model's performance for image restoration tasks.

4.9. VPEG Team

The VPEG team proposes a CNN-based deep model with spatially-adaptive feature modulation mechanism [34] for blind compressed image enhancement. As shown in Fig. 5, the proposed model consists of the following parts: a stacking of residual feature mixing blocks (RFMBs) and a reconstruction layer. Given a compressed image, the input resolution is first reduced using the PixelUnshuffle operation, and a 3×3 convolutional layer is used to transform the downsampled input into feature space and generate shallow features. The extracted shallow features are then processed by multiple stacked RFMBs, one of which contains 6 feature mixing modules and a 3×3 convolutional layer. Each feature mixing module has a spatially-adaptive feature modulation (SAFM) sub-layer and a convolutional channel mixer (CCM). To recover high-quality image, a global residual connection is further introduced to facilitate the model to learn high-frequency details, and a reconstruction layer is utilized to transform the extracted features to the target image.

In terms of training, LSDIR [22] training set is used, and the size of cropped input patch is set to 192×192. Random horizontal flip, vertical flip, and rotation are introduced into the data augmentation during training. The proposed model consists of 8 RFMBs, and the number of channels is set to 128. Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$ is used. The batch size is set to 16. The initial learning rate is set to 3×10^{-4} and the minimum one to 1×10^{-7}, which is updated by the Cosine Annealing scheme [28]. The total number of iterations is 500 000. In the training process, the loss function is the combination of L1 loss and Fourier based frequency loss [9].

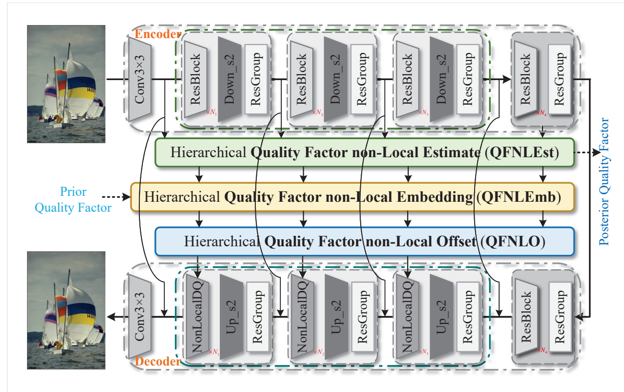

Figure 6. Framework of the model proposed by Tempest.

4.10. Tempest Team

The Tempest team proposes a hierarchical non-local enhance network for blind image compression. Specifically, for JPEG block-based compressed images, they propose a non-local, hierarchical enhancement method. The overall framework, as illustrated in Fig. **??**, involves global and local Quality Factor (QF) estimation and embedding at four different scales during the encoding process. The global QF for all scales is averaged to obtain the final predicted QF (QF posterior). Subsequently, attention is applied between the local QF embedding and the global QF embedding, followed by an update to the local QF embedding. Finally, mapping of both global and local QF embeddings yields global and local QF offsets. During the decoding process, the QF offsets, along with skip connections, are used for hierarchical enhancement at various scales, ultimately producing the enhanced image.

The training dataset was augmented using the LSDIR [22] dataset. They encoded the ground truth images using JPEG compression with ten arbitrary Quality Factors (QFs) within the range of [10, 70], thereby generating low-quality compressed images. The model underwent a total of 800K training iterations on four NVIDIA 3080 GPUs with batch size 16 on each GPU. The initial 300K iterations utilized the DIV2K training set released for the com-

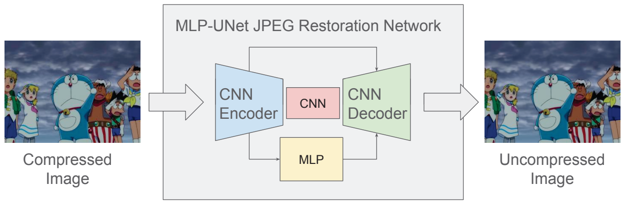

Figure 7. Framework of the model proposed by FlyingBunny.

petition, with a learning rate of $2 \times 10{-}4$. For the subsequent iterations, the LSDIR dataset was employed, with the learning rate being reduced to one-fifth every 100K iterations. An Adam optimizer was used for optimization. Similar to the baseline method [15], L_1 loss was applied to constrain the Quality Factor (QF) posterior and QF prior, while L_2 loss was utilized to constrain the fidelity of the reconstructed images to their ground truth, and the trade off parametr $\lambda = 0.001$.

4.11. FlyingBunny Team

The FlyingBunny team uses a UNet-based network for blind compressed image restoration. They additionally use MLP to extract additional features from CNN-Encoder and feed the features to CNN-Decoder. The architecture is very similar to FBCNN. The difference to FBCNN is that (1) the loss function of quality factor in FBCNN is removed because it was mainly used to adjust the results. (2) batch normalization is added to make the results better. (3) self-ensemble is used in test phase to further improve the performance. The proposed architecture is shown in Fig. 7.

4.12. AIVerse Team

The proposed method by team AIVerse is named as Progressive Residual Dense Attention Block (RDAB) for Compressed Image Enhancement. They build the network around the Retinex theory suggesting that the connection between the compressed image and the enhanced image can be represented by $y = x \otimes z$, where y represents the compressed images, x represents the quality factor component and z represents the desired enhanced image, respectively. The proposed method is inspired by the studies [14, 25, 31] that learn the parameters to enhance the image by introducing mapping with residual term represented by u^t using parameters ϑ at each stage t to improve the task modeling with progressive perspective. They use the residual representation due to its steadiness and better performance as suggested in [25, 40]. The calibration module takes into account the compressed image and extracts the calibrated map s, which presents the difference between the ground truth and the compressed image in each stage. The notation v^t refers to the input that is converted at each stage parameterized by the RDAB blocks using the learnable parameter θ. It is hypothesized that the calibration module learns to enhance the image gradually with each passing stage. The loss functions that helps in enlarging the capacity of the network and maintaining the pixel-wise consistency is used. During the learning phase, each output is designed to produce a result that is close to the desired output. The subsequent may produce the same results as the first block or may be very close results to the first block, thus the inference stage only need one block to provide accelerated inference. They use the spatial attention block and the residual dense block for extracting the representations. The SAB is composed of Convolutional layer and rectified linear units (ReLU) blocks followed by the concatenation. They use eight blocks of the pair followed by global average pooling, ReLU, global max pooling and Sigmoid function. The structure of the residual dense attention block (RDAB) is composed of dilated convolution, ReLU, Concatenation, and channel attention blocks from their previous method that is proposed for image denoising approach in NTIRE 2023 report [23].

The model is trained on the training dataset available from competition website. For the training process, they used ADAM optimizer [16] with the default parameters and the epsilon value set to 10^{-9}. The size of the minibatch was set to 8 and the learning rate was initialized to be $10^{-}3$ with decaying rate of 10^{-7} after every 100 epochs. They trained the network for $1K$ epochs. For the residual connection H_ϑ, they use the setting of three convolutional blocks paired with ReLU layer with the channel size of 3.

4.13. Unicorns Team

The Unicorns team developed a novel transformer-based deep network to learn decompress details from a compressed image, as shown in Fig. **??**. The architecture comprises two separate encoder-decoder blocks (EDB) and a multi-head correlation block (MHCB) to produce plausible images. Illumination mapping is leveraged from the well-known Retinex theory [17] to accelerate the reconstruction performance. They deeply examined the practicability of illumination mapping in generic image restoration techniques such as image decompression. Hence, in the first half of the architecture, the state-of-the-art low-light enhancement method, Retinexformer [5], is incorporated. In shadow removal, Retinexformer can outperform well-known image restoration methods like Uformer [42], Restormer [46], etc. However, like other restoration models, Retinexformer failed to recover the salient details in spatially complex regions. To address this limitation, they proposed to utilize an MHCB, followed by another EDB in our architecture. They leverage the correlated features with intermediate output in the second EDB to perceive better restoration results. In addition to that, they utilized a perceptual loss, including luminance-chrominance guidance, to address the color inconsistency.

The training phase consisted of feeding the JPEG input images and calculating the loss between the JPEG and the corresponding ground-truth image. The training underwent using only the dataset from NTIRE competition. Each dataset image was cropped and the cropped region was used for model training for better spatial results. The model was optimized with an Adam optimizer, whose hyperparameters were tuned as $\beta_1 = 0.9$, $\beta_2 = 0.99$, for 65k steps with a constant learning rate of 1×10^{-4}.

Acknowledgments

This work was partially supported by the Humboldt Foundation. We thank the NTIRE 2024 sponsors: Meta Reality Labs, OPPO, KuaiShou, Huawei and University of Würzburg (Computer Vision Lab).

Appendix: Teams and affiliations

NTIRE 2024 Team

Challenge:
NTIRE 2024 Challenge on Blind Enhancement of Compressed Image
Organizer(s):
Ren Yang[1] (r.yangchn@gmail.com),
Radu Timofte[2] (radu.timofte@uni-wuerzburg.de)
Affiliation(s):
[1] ETH Zürich, Switzerland
[2] Julius Maximilian University of Würzburg, Germany

IMCL-BVQE Team

Member(s):
Bingchen Li (lbc31415926@mail.ustc.edu.cn), Xin Li, Mengxi Guo, Shijie Zhao, Li Zhang, Zhibo Chen
Affiliation(s):
University of Science and Technology of China,
ByteDance Inc.

PixelArtAI Team

Member(s):
Dongyang Zhang (690866816@qq.com)
Affiliation(s):
MGTV, Changsha, China

Titans Team

Member(s):
Yash Arora (yasharora102@gmail.com), Aditya Arora
Affiliation(s):
Amrita Vishwa Vidyapeetham, Kerala, India
York University, Toronto, Canada

BinYCn Team

Member(s):
Yuanbin Chen (binycn904363330@gmail.com), Hui Tang, Tao Wang, Longxuan Zhao, Bin Chen, Tong Tong
Affiliation(s):
College of Physics and Information Engineering, Fuzhou University, Fuzhou, China

OldFe666 Team

Member(s):
Qiao Mo (mqiao568@gmail.com), Jingwei Bao, Jinhua Hao, Yukang Ding, Hantang Li, Ming Sun, Chao Zhou and Shuyuan Zhu
Affiliation(s):
Kuaishou Technology, UESTC

FLV-T1 Team

Member(s):
Zhi Jin (jinzh26@mail.sysu.edu.cn), Wei Wang, Dandan Zhan, Jiawei Wu, Jiahao Wu, Luwei Tu
Affiliation(s):
School of Intelligent Systems Engineering, Shenzhen Campus of Sun Yat-sen University, Shenzhen, China

UCAS_SCST Team

Member(s):
Hongyu An (anhongyu22@mails.ucas.ac.cn), Xinfeng Zhang
Affiliation(s):
School of Computer Science and Technology, University of Chinese Academy of Sciences, China

SYU-HnVLab Team

Member(s):
Woon-Ha Yeo (canal@syuin.ac.kr), Wang-Taek Oh, Young-Il Kim, Han-Cheol Ryu
Affiliation(s):
AI Convergence Research Center, Sahmyook University, Republic of Korea

VPEG Team

Member(s):
Long Sun (cs.longsun@gmail.com), Mingjun Zhen, Jinshan Pan, Jiangxin Dong, and Jinhui Tang
Affiliation(s):
Nanjing University of Science and Technology, Nanjing, China

Tempest Team

Member(s):
Yapeng Du (18200199198@163.com), Ao Li, Ziyang He, Lei Luo, Ce Zhu
Affiliation(s):
University of Electronic Science and Technology of China, China

FlyingBunny Team

Member(s):
Xin Yao (xin.yao.ict@gmail.com)
Affiliation(s):
Politecnico di Torino, Italy

AIVerse Team

Member(s):
Sunder Ali Khowaja[1] (Sandar.ali@usindh.edu.pk), IK Hyun Lee[2,3] (ihlee@tukorea.ac.kr)
Affiliation(s):
[1] Faculty of Computing, Digital and Data, Technological University Dublin (TU Dublin), Ireland
[2] Department of Mechatronics Engineering, Tech University of Korea, Siheung-Si, Republic of Korea
[3] IKLab Inc., Siheung-Si, Republic of Korea

Unicorns Team

Member(s):
Jaeho Lee (jaeho.lee@opt-ai.kr), Seongwan Kim, Sharif S M A, Nodirkhuja Khujaev, Roman Tsoy
Affiliation(s):
Opt-AI

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 (CVPRW)*, pages 126–135, 2017. 2, 3, 5, 6

[2] Cosmin Ancuti, Codruta O Ancuti, Florin-Alexandru Vasluianu, Radu Timofte, et al. NTIRE 2024 dense and non-homogeneous dehazing challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 1

[3] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 33(5):898–916, 2010. 2, 6

[4] Nikola Banić, Egor Ershov, Artyom Panshin, Oleg Karasev, Sergey Korchagin, Shepelev Lev, Alexandr Startsev, Daniil Vladimirov, Ekaterina Zaychenkova, Dmitrii R Iarchuk, Maria Efimova, Radu Timofte, Arseniy Terekhin, et al. NTIRE 2024 challenge on night photography rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 1

[5] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12504–12513, 2023. 8

[6] Nicolas Chahine, Marcos V. Conde, Sira Ferradans, Radu Timofte, et al. Deep portrait quality assessment. a NTIRE 2024 challenge survey. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 1

[7] 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 (CVPR)*, pages 22367–22377, 2023. 3, 6

[8] Zheng Chen, Zongwei WU, Eduard Sebastian Zamfir, Kai Zhang, Yulun Zhang, Radu Timofte, Xiaokang Yang, et al. NTIRE 2024 challenge on image super-resolution (×4): Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 1

[9] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021. 7

[10] Marcos V. Conde, Florin-Alexandru Vasluianu, Radu Timofte, et al. Deep raw image super-resolution. a NTIRE 2024 challenge survey. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 1

[11] 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 (ICCV)*, pages 576–584, 2015. 1

[12] Max Ehrlich, Larry Davis, Ser-Nam Lim, and Abhinav Shrivastava. Quantization guided jpeg artifact correction. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 293–309. Springer, 2020. 1

[13] Shuhang Gu, Andreas Lugmayr, Martin Danelljan, Manuel Fritsche, Julien Lamour, and Radu Timofte. DIV8K: Diverse 8k resolution image dataset. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 3512–3516, 2019. 3

[14] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26(2):982–993, 2016. 8

[15] Jiaxi Jiang, Kai Zhang, and Radu Timofte. Towards flexible blind jpeg artifacts removal. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4997–5006, 2021. 1, 2, 8

[16] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014. arxiv.org/abs/1412.6980. 8

[17] Edwin H Land and John J McCann. Lightness and retinex theory. *Josa*, 61(1):1–11, 1971. 8

[18] 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 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2024. 3

[19] 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. 3

[20] Xin Li, Kun Yuan, Yajing Pei, Yiting Lu, Ming Sun, Chao Zhou, Zhibo Chen, Radu Timofte, et al. NTIRE 2024 challenge on short-form UGC video quality assessment: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 1

[21] Yawei Li, Yuchen Fan, Xiaoyu Xiang, Denis Demandolx, Rakesh Ranjan, Radu Timofte, and Luc Van Gool. Efficient and Explicit Modelling of Image Hierarchies for Image Restoration . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18278–18289, 2023. 5

[22] 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 (CVPR)*, pages 1775–1787, 2023. 2, 3, 5, 7

[23] Yawei Li, Yulun Zhang, Radu Timofte, Luc Van Gool, Zhijun Tu, Kunpeng Du, Hailing Wang, Hanting Chen, Wei Li, Xiaofei Wang, Jie Hu, Yunhe Wang, Xiangyu Kong, Jinlong Wu, Dafeng Zhang, Jianxing Zhang, Shuai Liu, Furui Bai, Chaoyu Feng, Hao Wang, Yuqian Zhang, Guangqi Shao, Xiaotao Wang, Lei Lei, Rongjian Xu, Zhilu Zhang, Yunjin Chen, Dongwei Ren, Wangmeng Zuo, Qi Wu, Mingyan Han, Shen Cheng, Haipeng Li, Ting Jiang, Chengzhi Jiang, Xinpeng Li, Jinting Luo, Wenjie Lin, Lei Yu, Haoqiang Fan, Shuaicheng Liu, Aditya Arora, Syed Waqas Zamir, Javier Vazquez-Corral, Konstantinos G. Derpanis, Michael S. Brown, Hao Li, Zhihao Zhao, Jinshan Pan, Jiangxin Dong, Jinhui Tang, Bo Yang, Jingxiang Chen, Chenghua Li, Xi Zhang, Zhao Zhang, Jiahuan Ren, Zhicheng Ji, Kang Miao, Suiyi Zhao, Huan Zheng, YanYan Wei, Kangliang Liu, Xiangcheng Du, Sijie Liu, Yingbin Zheng, Xingjiao Wu, Cheng Jin, Rajeev Irny, Sriharsha Koundinya, Vighnesh Kamath, Gaurav Khandelwal, Sunder Ali Khowaja, Jiseok Yoon, Ik Hyun Lee, Shijie Chen, Chengqiang Zhao, Huabin Yang, Zhongjian Zhang, Junjia Huang, and Yanru Zhang. NTIRE 2023 challenge on image denoising: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1905–1921, 2023. 8

[24] Jie Liang, Qiaosi Yi, Shuaizheng Liu, Lingchen Sun, Rongyuan Wu, Xindong Zhang, Hui Zeng, Radu Timofte, Lei Zhang, et al. NTIRE 2024 restore any image model (RAIM) in the wild challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 1

[25] Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *2021 IEEE/CVF International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10561–10570, 2021. 8

[26] Xiaohong Liu, Xiongkuo Min, Guangtao Zhai, Chunyi Li, Tengchuan Kou, Wei Sun, Haoning Wu, Yixuan Gao, Yuqin Cao, Zicheng Zhang, Xiele Wu, Radu Timofte, et al. NTIRE 2024 quality assessment of AI-generated content challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 1

[27] Xiaoning Liu, Zongwei WU, Ao Li, Florin-Alexandru Vasluianu, Yulun Zhang, Shuhang Gu, Le Zhang, Ce Zhu, Radu Timofte, et al. NTIRE 2024 challenge on low light image enhancement: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 1

[28] Ilya Loshchilov and Frank Hutter. SGDR: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016. 3, 5, 7

[29] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 7

[30] Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang. Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing*, 26(2):1004–1016, 2016. 2, 6

[31] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *2022 IEEE/CVF International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5637–5646, 2022. 8

[32] 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 (NeurIPS)*, 36, 2024. 3

[33] Bin Ren, Yawei Li, Nancy Mehta, Radu Timofte, et al. The ninth NTIRE 2024 efficient super-resolution challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 1

[34] Long Sun, Jiangxin Dong, Jinhui Tang, and Jinshan Pan. Spatially-adaptive feature modulation for efficient image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023. 7

[35] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. Memnet: A persistent memory network for image restoration. In

Proceedings of the IEEE International Conference on Computer Vision (ICCV), pages 4539–4547, 2017. 1

[36] 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 (CVPRW)*, pages 114–125, 2017. 2, 3, 5, 6

[37] Radu Timofte, Rasmus Rothe, and Luc Van Gool. Seven ways to improve example-based single image super resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1865–1873, 2016. 2

[38] Florin-Alexandru Vasluianu, Tim Seizinger, Zhuyun Zhou, Zongwei WU, Cailian Chen, Radu Timofte, et al. NTIRE 2024 image shadow removal challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 1

[39] Longguang Wang, Yulan Guo, Juncheng Li, Hongda Liu, Yang Zhao, Yingqian Wang, Zhi Jin, Shuhang Gu, Radu Timofte, et al. NTIRE 2024 challenge on stereo image super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 1

[40] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6849–6857, 2019. 8

[41] Yingqian Wang, Zhengyu Liang, Qianyu Chen, Longguang Wang, Jungang Yang, Radu Timofte, Yulan Guo, et al. NTIRE 2024 challenge on light field image super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 1

[42] 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 (CVPR)*, pages 17683–17693, 2022. 8

[43] 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, 6

[44] Ren Yang, Radu Timofte, et al. NTIRE 2024 challenge on blind enhancement of compressed image: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 1

[45] Pierluigi Zama Ramirez, Fabio Tosi, Luigi Di Stefano, Radu Timofte, Alex Costanzino, Matteo Poggi, et al. NTIRE 2024 challenge on HR depth from images of specular and transparent surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 1

[46] 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 (CVPR)*, pages 5728–5739, 2022. 3, 4, 6, 8

[47] Kai Zhang, Yawei Li, Jingyun Liang, Jiezhang Cao, Yulun Zhang, Hao Tang, Deng-Ping Fan, Radu Timofte, and Luc Van Gool. Practical blind image denoising via swin-conv-unet and data synthesis. *Machine Intelligence Research*, 20(6):822–836, 2023. 3

[48] 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. 1

[49] Zhilu Zhang, Shuohao Zhang, Renlong Wu, Wangmeng Zuo, Radu Timofte, et al. NTIRE 2024 challenge on bracketing image restoration and enhancement: Datasets, methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 1