NTIRE 2024 Challenge on Bracketing Image Restoration and Enhancement: Datasets, Methods and Results

Zhilu Zhang* Shuohao Zhang* Renlong Wu* Wangmeng Zuo* Radu Timofte*
Xiaoxia Xing Hyunhee Park Sejun Song Changho Kim Xiangyu Kong
Jinlong Wu Jianxing Zhang Jingfan Tan Zikun Liu Wenhan Luo Wenjie Lin
Chengzhi Jiang Mingyan Han Zhen Liu Ting Jiang Jinting Luo Shen Cheng
Linze Li Xinhan Niu Shuaicheng Liu Kexin Dai Kangzhen Yang Tao Hu
Xiangyu Chen Yu Cao Qingsen Yan Yanning Zhang Genggeng Chen
Yongqing Yang Wei Dong Xinwei Dai Yuanbo Zhou Xintao Qiu Hui Tang
Wei Deng Qingquan Gao Tong Tong Peng Zhang Yifei Chen Wenbo Xiong
Zhijun Song Pu Cheng Taolue Feng Yunqing He Daiguo Zhou Ying Huang
Xiaowen Ma Peng Wu

Abstract

Low-light photography presents significant challenges. Multi-image processing methods have made numerous attempts to obtain high-quality photos, yet remain unsatisfactory. Recently, bracketing image restoration and enhancement has received increased attention. By leveraging the full potential of multi-exposure images, several tasks (including denoising, deblurring, high dynamic range enhancement, and super-resolution) can be jointly addressed. This paper reviews the NTIRE 2024 challenge on bracketing image restoration and enhancement. In the challenge, participants are required to process multi-exposure RAW images to generate noise-free, blur-free, high dynamic range, and even higher-resolution RAW images. The challenge comprises two tracks. Track 1 does not incorporate the super-resolution task, whereas Track 2 does. Each track featured five teams participating in the final testing phase. The proposed methods establish new state-of-the-art performance benchmarks.

1. Introduction

Low-light photography is a widely desired yet challenging problem. Recent years have witnessed significant advancements in enhancing low-light image quality through learning-based methods. In comparison with single-image restoration (*e.g.*, denoising [1, 9, 30, 45, 49, 94, 98, 99], deblurring [18, 62, 66, 74, 95, 97], and super-resolution (SR) [41, 46, 48, 51, 100, 102, 103]) and enhancement (*e.g.*, high dynamic range (HDR) reconstruction [15, 26, 42, 55, 69, 110]), multi-image processing methods offer more advantages in mitigating the ill-posed nature of this problem and can generate results with higher fidelity.

Several multi-image processing works are summarized in Tab. 1. Within these works, burst and dual-exposure images have limited dynamic range, making it difficult to restore underexposed or overexposed details. Multi-exposure images contain richer information, while most methods based on them require ideal prerequisites (*e.g.*, no noise or no blur). More recently, Zhang *et al.* [106] consider a more realistic situation of multi-exposure images, encompassing noisy, blurry, underexposed, overexposed content, as well as inter-frame misalignment. They utilize the complementarity of exposure bracketing images to integrate image restoration and enhancement tasks, thereby generating a noise-free, blur-free, HDR, and high-resolution image. This new objective, termed bracketing image restoration and enhancement, is proposed.

This paper reviews the NTIRE 2024 Challenge on Bracketing Image Restoration and Enhancement, which aims to foster further research and establish state-of-the-art benchmarks. Participants are tasked with generating noise-free, blur-free, HDR, and even higher-resolution RAW images from five multi-exposure RAW images. The challenge consists of two tracks: Track 1 and Track 2. Track 1 does not incorporate the super-resolution task, whereas Track 2 does. In the final testing phase, each track featured the participation of five teams. This report briefly describes their

*Zhilu Zhang, Shuohao Zhang, Renlong Wu, Wangmeng Zuo, and Radu Timofte are the NTIRE 2024 challenge organizers, as shown in Appendix A. The other authors are participants in this challenge, as shown in Appendices B and C.

solutions and reports their performance. Their proposed methods have set new state-of-the-art benchmarks in bracketing image restoration and enhancement.

2. NTIRE 2024 Challenge

Bracketing Image Restoration and Enhancement is one of the NTIRE 2024 associated challenges (https://cvlai.net/ntire/2024/), which include dense and non-homogeneous dehazing [3], night photography rendering [4], blind compressed image enhancement [91], shadow removal [77], efficient super resolution [70], image super resolution (×4) [16], light field image super-resolution [80], stereo image super-resolution [78], HR depth from images of specular and transparent surfaces [93], bracketing image restoration and enhancement [107], portrait quality assessment [10], quality assessment for AI-generated content [53], restore any image model (RAIM) in the wild [47], RAW image super-resolution [19], short-form UGC video quality assessment [44], low light enhancement [54], and RAW burst alignment and ISP challenge.

2.1. Overview

The objectives of NTIRE 2024 Challenge on Bracketing Image Restoration and Enhancement are as follows: (1) highlighting the problems and promoting the research in bracketing image restoration and enhancement, (2) establishing high-quality benchmarks for bracketing image restoration and enhancement and facilitating comparisons between various methods, (3) providing a platform for academic and industrial participants to engage, discuss, and potentially establish collaborations.

The challenge has two tracks, *i.e.*, Track 1 (https://codalab.lisn.upsaclay.fr/competitions/17573) and Track 2 (https://codalab.lisn.upsaclay.fr/competitions/17574). Track 1 utilizes bracketing photography to unify basic restoration (*i.e.*, denoising and deblurring) and enhancement (*i.e.*, HDR reconstruction), named BracketIRE. Track 2 appends the ×4 super-resolution task, dubbed BracketIRE+. The inputs are five multi-exposure RAW images. The output should be a noise-free, blur-free, HDR, and even higher-resolution (only for Track 2) RAW image. Participants can develop novel network architectures and other techniques to achieve the goal. Besides, we provide a simple image signal processing (ISP) toolkit (https://github.com/cszhilu1998/BracketIRE/tree/master/NTIRE2024) for converting the output RAW images into 16-bit RGB images for evaluation and visualization. We also provide codes and pre-trained models of a baseline (*i.e.*, TMRNet [106]) at this URL.

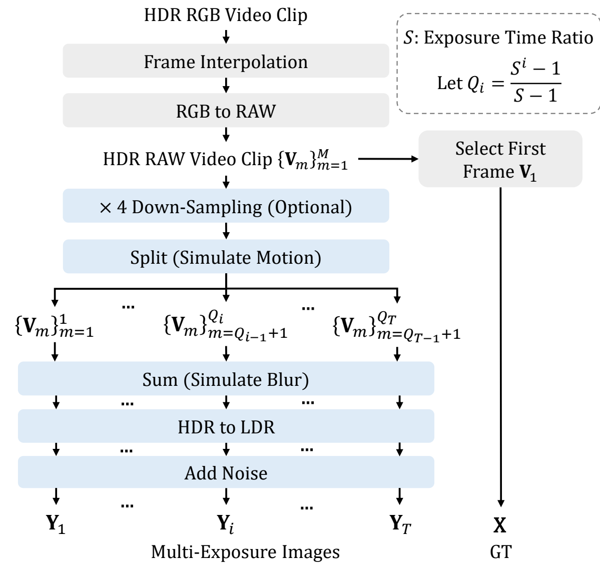


Figure 1. Overview of data simulation pipeline. HDR video is utilized to synthesize multi-exposure images $\{\mathbf{Y}_i\}_{i=1}^{T}$ and the corresponding GT image $\mathbf{X}$. S denotes the exposure time ratio between $\mathbf{Y}_i$ and $\mathbf{Y}_{i-1}$. $\mathbf{Y}_i$ is obtained by summing and processing S^{i-1} ($\{i–1\}$-th power of S) images from HDR raw video $\mathbf{V}$. Q_i denotes the total number of images from $\mathbf{V}$ that participate in constructing $\{\mathbf{Y}_k\}_{k=1}^{i}$.

2.2. Datasets

We follow the data simulation pipeline [106] to provide synthetic datasets for BracketIRE and BracketIRE+ tasks, as shown in Fig. 1. First, we perform ×32 frame interpolation with RIFE [34] on HDR videos from Froehlich *et al.* [28]. Then, we convert the RGB videos to RAW space, getting HDR RAW sequences$\{\mathbf{V}_m\}_{m=1}^{M}$.

Next, we introduce degradations to construct multi-exposure images, mainly including the following 5 steps. (1) Bicubic ×4 down-sampling is applied to obtain low-resolution images, which only serves for BracketIRE+ task. (2) Each video is split into T non-overlapped groups, where i-th group should be used to synthesize $\mathbf{Y}_i$. (3) Denote the exposure time ratio between $\mathbf{Y}_i$ and $\mathbf{Y}_{i-1}$ by S. We sequentially move S^{i-1} ($\{i-1\}$-th power of S) consecutive images into the above i-th group, summing them up to simulate blurry images. (4) We transform the HDR blurry images into low dynamic range (LDR) ones by cropping values outside the specified range and mapping the cropped values to 10-bit unsigned integers. (5) We add the heteroscedastic Gaussian noise [9, 32, 79] to LDR images to generate the final multi-exposure images (*i.e.*, $\{\mathbf{Y}_i\}_{i=1}^{T}$). The noise variance is a function of pixel intensity, whose parameters are estimated from the real-world images captured by Xiaomi 10S smartphone when ISO is set to 1,600.

The first frame $\mathbf{V}_1$ in HDR video clip is taken as a ground truth (GT) image, whose resolution is 1,920×1,080.

Table 1. Comparison between various multi-image processing manners.

Setting	Methods	Input Images	Supported Tasks Denoising	Deblurring	HDR	SR
Burst Denoising	[29, 31, 64, 71, 86]	Burst	✓			
Burst Deblurring	[2, 21, 68, 83]	Burst		✓		
Burst SR	[22, 82, 84]	Burst				✓
Burst Denoising and SR	[5–8, 24, 25, 39, 61, 63, 85]	Burst	✓			✓
Burst Denoising and HDR	[27, 33]	Burst	✓		✓	
Dual-Exposure Image Restoration	[12, 38, 65, 72, 92, 104, 108]	Dual-Exposure	✓	✓		
Basic HDR Imaging	[36, 57, 67, 75, 88, 89, 105]	Multi-Exposure			✓	
HDR Imaging with Denoising	[17, 52, 69]	Multi-Exposure	✓		✓	
HDR Imaging with SR	[73]	Multi-Exposure			✓	✓
HDR Imaging with Denoising and SR	[40]	Multi-Exposure	✓		✓	✓
BracketIRE	[106]	Multi-Exposure	✓	✓	✓	
BracketIRE+	[106]	Multi-Exposure	✓	✓	✓	✓

We set the exposure time ratio S to 4 and the frame number T to 5. Finally, we obtain 1,335 data pairs from 35 scenes. 1,045 pairs from 31 scenes are used for training, and the remaining 290 pairs from the other 4 scenes are used for testing. The validation set is selected from the testing set, including 29 pairs from 4 scenes.

2.3. Challenge Rules

To ensure fairness and equitable comparisons to the greatest extent possible, the following competition rules are established: (1) Reproducibility of methods is a must. (2) The method should input multi-exposure RAW images and output an HDR RAW image. (3) The participants cannot change the provided ISP. (4) During inference, the number of network parameters should be less than 100M, GPU memory should be controlled within 24G and the use of a self-ensemble strategy is prohibited. (5) Each participant can only join one team. Each team can only submit one algorithm for final ranking. (6) Any testing adaptive strategies are not permitted. In other words, the model cannot be fine-tuned using data related to the test images. For example, using the testing results of Track 1 to fine-tune the model of Track 2 is prohibited.

2.4. Challenge Phases

There are two phases in the challenge: (1) development and validation phase, (2) testing phase.

Development and Validation Phase. Participants have access to both the training data and validation data (refer to Sec. 2.2 for dataset details). Note that the training data includes multi-exposure images and the corresponding GT images, while the validation data only includes the input multi-exposure images. Participants can upload their results to the validation server to calculate PSNR metrics and receive feedback.

Testing Phase. Participants have access to the testing multi-exposure images to generate final results. Note that the GT images are not available for them. They can submit their results to the server and email their code and fact sheet to the organizers. The organizers executed the provided code to verify the results, which were then shared with participants at the end of the challenge.

3. Challenge Results

There are 100 and 92 participants registered for Track 1 and Track 2, respectively. Each track has five teams submitting the final testing results (including model outputs, codes, and fact sheets). Their solutions are described in Sec. 4.

3.1. Evaluation Metrics

The restored RAW images are post-processed to 16-bit RGB ones by the provided ISP. We evaluate the results on the 16-bit RGB domain. The widely used PSNR [35], SSIM [81] and LPIPS [101] are employed as the quantitative evaluation metrics. The ranking is based on the PSNR metric of the full images. Besides, given that several pixels around the image are invalid, we also calculate the metrics on the cropped images that exclude the invalid pixels. To compare the inference cost, #FLOPs, testing time, and GPU memory are measured when generating a 1920×1080 RAW image on a single NVIDIA RTX A6000 GPU. The number of model parameters is also measured. In addition, the performance of a baseline model (*i.e.*, TMRNet [106]) is reported for reference.

3.2. Track 1: BracketIRE

Tab. 2 shows that all teams outperform the baseline (*i.e.*, TMRNet[106]) with a large margin, improving PSNR results ranging from 0.27dB to 2.35dB. In particular, Samsung team achieves the best performance, obtaining 2.35dB

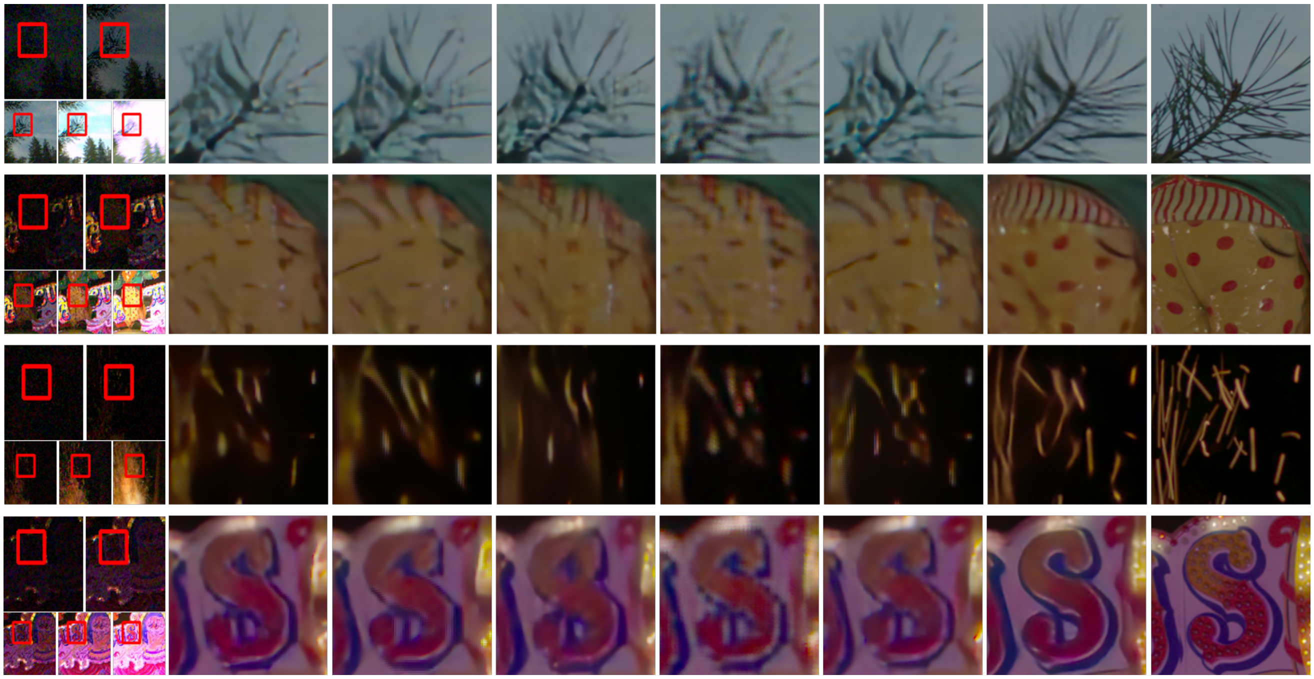
Figure 2. Qualitative results on Track 1 of Bracketing Image Restoration and Enhancement Challenge.

Table 2. Quantitative results on Track 1 of Bracketing Image Restoration and Enhancement Challenge. Quantitative metrics are calculated on the full images and the cropped images that exclude the surrounding 10 invalid pixels. #FLOPs, inference time, and GPU memory are measured when generating a 1920×1080 RAW image. We use NVIDIA RTX A6000 GPU to calculate the inference time and adopt THOP [109] toolkit to calculate #FLOPs. The ranking is based on the PSNR metric of the full images.

Rank	Team	Full Images PSNR ↑ / SSIM↑ / LPIPS↓	Cropped Images PSNR ↑ / SSIM↑ / LPIPS↓	#Params (M)	#FLOPs (T)	Time (s)	Memory (GB)
1	Samsung	40.54 / 0.9637 / 0.077	41.77 / 0.9633 / 0.076	94.34	48.238	3.102	20
2	MegIRE	39.78 / 0.9556 / 0.102	39.82 / 0.9550 / 0.105	19.75	30.751	2.383	16
3	UPN1	39.03 / 0.9500 / 0.117	39.02 / 0.9493 / 0.120	13.32	10.409	1.090	6
4	CVG	38.78 / 0.9543 / 0.102	39.89 / 0.9557 / 0.104	33.40	21.340	7.518	11
5	FZU_DXW	38.46 / 0.9527 / 0.105	39.61 / 0.9540 / 0.107	14.04	22.283	1.829	16
-	TMRNet [106]	38.19 / 0.9488 / 0.112	39.06 / 0.9498 / 0.115	13.29	21.340	1.874	15

Table 3. Quantitative results on Track 2 of Bracketing Image Restoration and Enhancement Challenge. Quantitative metrics are calculated on the full images and the cropped images that exclude the surrounding 16 invalid pixels. #FLOPs, inference time, and GPU memory are measured when generating a 1920×1080 RAW image. We use NVIDIA RTX A6000 GPU to calculate the inference time and adopt THOP [109] toolkit to calculate #FLOPs. The ranking is based on the PSNR metric of the full images.

Rank	Team	Full Images PSNR ↑ / SSIM↑ / LPIPS↓	Cropped Images PSNR ↑ / SSIM↑ / LPIPS↓	#Params (M)	#FLOPs (T)	Time (s)	Memory (GB)
1	Samsung	34.26 / 0.8913 / 0.206	34.80 / 0.8913 / 0.208	95.00	5.285	0.813	5
2	NWPU	30.59 / 0.8728 / 0.268	29.93 / 0.8633 / 0.274	13.37	1.426	0.887	3
3	FZU_DXW	29.82 / 0.8537 / 0.282	31.35 / 0.8660 / 0.277	14.34	1.500	0.493	3
4	CYD	29.66 / 0.8598 / 0.284	30.27 / 0.8632 / 0.285	17.60	1.560	0.751	4
5	CVG	29.25 / 0.8521 / 0.278	30.63 / 0.8645 / 0.275	71.82	6.898	0.679	4
-	TMRNet [106]	28.91 / 0.8572 / 0.273	30.65 / 0.8725 / 0.270	13.58	1.441	0.489	4



Input Frames TMRNet [106] CVG CYD FZU_DXW NWPU Samsung GT

Figure 3. Qualitative results on Track 2 of Bracketing Image Restoration and Enhancement Challenge.

and 0.76dB PSNR improvements than the baseline and second-place team (*i.e*., MegIRE), respectively. Besides, #Params and GPU Memory of all teams meet the challenge rules. The visual comparisons are shown in Fig. 2. It can be seen that the results from Samsung team have more photo-realistic details, and are more consistent with the GTs.

3.3. Track 2: BracketIRE+

This track appends $\times 4$ super-resolution task based on Track 1, making it more challenging. The results are shown in Tab. 3. It can be seen that all teams perform better than the baseline (*i.e*., TMRNet[106]). Among all teams, Samsung team performs best and outperforms other methods by a large margin. Compared with the baseline, it obtains 5.35dB PSNR gains. Compared with the second-place team (*i.e*., NWPU), it achieves 3.67dB improvements. Besides, the inference cost of all methods follows the challenge requirements. The visual comparisons in Fig. 3 show that Samsung team produces more fine-scale details and fewer artifacts.

4. Challenge Methods

4.1. Track 1: BracketIRE

4.1.1 Samsung Team †

The team proposes a high quality Reference feature for Two stage bracketing Image Restoration and Enhancement Network, called RT-IRENet [87]. Specifically, the first module is based on TMRNet [106] which fuses 5 RAW images into the coarse restored result. The second module is based on NAFNet [14] which refines the output of the first module into the final result with much more details.

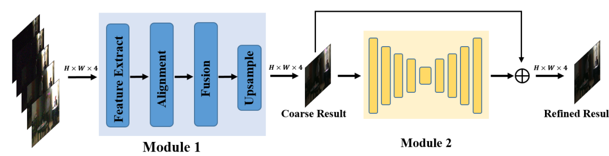


Figure 4. Structure of RT-IRENet proposed by Samsung team.

Network Architecture. The network architecture is shown in Fig. 4. For the first module, the team increases the number of channels from 64 to 96 in TMRNet [106]. Furthermore, the team chooses the second frame instead of the first frame of input as a reference frame. The reason for this is that the second frame has less noise and acceptable blur, which is beneficial in avoiding noise information from the first frame feature on the fused feature. It's worth noting that the second frame has only a tiny position misalignment with ground truth, thus such an operation has less impact on calculating loss terms. For the second module, the team adopts NAFNet [14], and changes the number of input and output channels to 4.

Inference Strategy. During inference, the team crops surrounding 5 pixels for input RAW images and feeds the cropped images to the model. Then they pad 5 pixels for

† Full name: Samsung MX (Mobile eXperience) Business & Samsung Research China - Beijing (SRC-B).

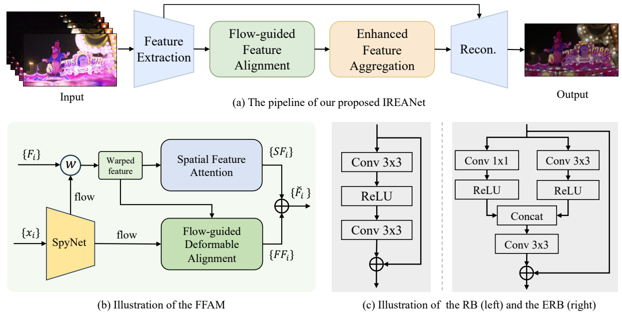


Figure 5. Illustration of IREANet proposed by MegIRE team.

the output RAW image. This ensures a high-quality output around the image.

Training Details. The team uses AdamW [59] optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$) with a weight decay of 0.01. They adopt the cosine annealing strategy where the learning rate gradually decreases from 1×10^{-4} to 1×10^{-6} for 1000 epochs. They utilize PSNR loss to optimize. The batch size is 4, and the input patch size is 256×256. The experiments are conducted on the NVIDIA A100 GPU.

4.1.2 MegIRE Team

The team utilizes flow-guided feature alignment and enhanced feature aggregation to propose IREANet, which is shown in Fig. 5.

Flow-Guided Feature Alignment. It is designed as a dual-branch architecture to perform feature alignment, as shown in Fig. 5(b). The first branch is flow-guided deformable alignment [11], which is also adopted in TMRNet [106]. The second branch performs spatial feature attention on the flow-guided aligned features, obtaining spatial attention features. The spatial attention mechanism has been proven to effectively reduce noise and undesired contents caused by foreground object movement [57, 88]. The features from the two branches can be seen as complementary and are then combined using element-wise addition to acquire the final aligned features.

Enhanced Feature Aggregation. Similar to TMRNet [106], a unidirectional recurrent network is utilized to aggregate temporal features. Furthermore, they introduce an enhanced feature aggregation module, which takes the proposed enhanced residual block as a basic component, as shown in Fig. 5 (c). The proposed module increases the network's nonlinearity and enables better convergence, thus allowing for more effective aggregation of temporal features.

Training Details. The team uses a random combination of Bayer preserving augmentation [50] (see Fig. 6), vertical flip, horizontal flip, and rotation to augment the training data. They adopt AdamW [59] optimizer and the cosine annealing strategy of learning rate. The batch size is 8 and the input patch size is 128×128. Experiments are conducted on 4 NVIDIA TITAN Xp GPUs.

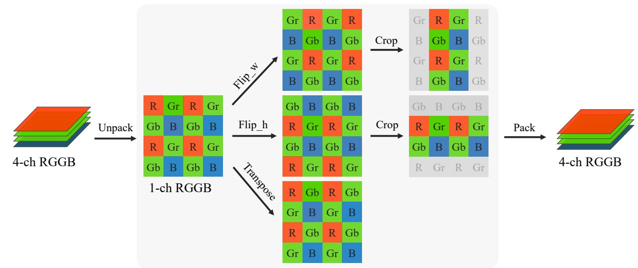


Figure 6. Illustration of Bayer preserving augmentation.

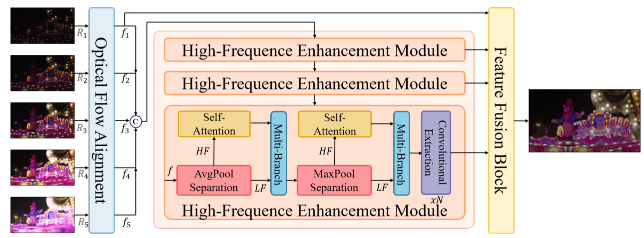


Figure 7. Overview of CRNet proposed by UPN1 team.

4.1.3 UPN1 Team

For the task, image details are of paramount importance. Moreover, to be applicable on edge devices, the inference time and the GPU memory footprint when generating images should also be carefully considered. The team proposes a Convolutional Restoration Network (CRNet) [90] (see Fig. 7), which can produce high-quality image details and has significant advantages in terms of inference time and GPU memory usage.

Network Architecture. TMRNet adopts a frame-by-frame processing manner. However, as the network deepens and the number of input frames increases, the team experimentally observes that this manner may lead to the network gradually forgetting the earlier frames and focusing more on the later frames, resulting in poor image quality. Instead, the team concatenates the five aligned images together for subsequent processing. First, the aligned images go through 3 high-frequency enhancement modules. Each module starts with two different pooling layers to separate high and low-frequency information [60] and enhance them separately, and ends with N convolutional extraction blocks. The frequency separation module utilizes self-attention to enhance the precious high-frequency information and deploys multi-branch [76] blocks to fuse high and low-frequency information. The convolutional extraction block can be seen as a high-frequency filter, and it utilizes large-kernel depth-wise separable convolutions [56] and convolutional FFN [23] for computationally friendly feature enhancement. Then, they

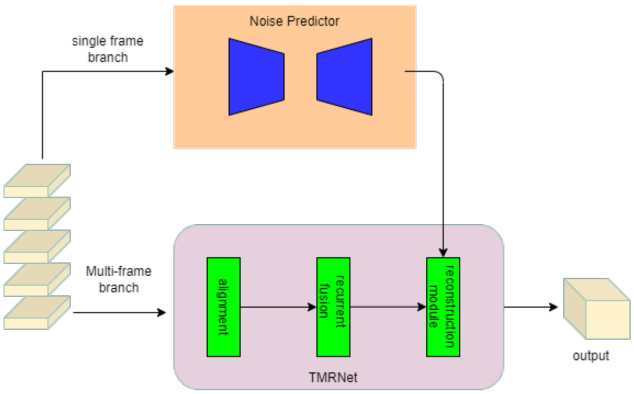


Figure 8. Network architecture proposed by CVG team.

use simple convolutional blocks to fuse the features from the 3 high-frequency enhancement modules and the reference frame to output the final result.

Implemental Details. They adopt the AdamW [59] optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate is initially set to 10^{-4} and decays by half every 80 epochs. The model is trained on 4 NVIDIA A100 GPUs for 500 epochs for 3 days. Additionally, they utilize Hugging Face's Accelerator for parallel training.

4.1.4 CVG Team

In the process of image restoration, denoising is crucial to balance detail and noise. Given the promising capabilities of diffusion models, the team integrates a single-frame diffusion module and a multi-frame processing module, thus proposing a two-branch network (see Fig. 8).

Network Architecture. In the multi-frame branch, the team adopts TMRNet [106]. In the single-frame branch, the team uses a diffusion-based framework to process the reference frame. They utilize both noise-domain and image-domain loss terms to train the diffusion module for conducting stable sampling during inference. Finally, the output of the diffusion-based branch is connected to the reconstruction module in TMRNet.

Training Details. The team first trains the diffusion module for about 300 epochs. Then they optimize the multi-frame branch module for about 1000 epochs. When they train the multi-frame branch, they fix the weight of the diffusion module. The optimizer is Adam [37] with an initial learning rate of 10^{-4}. The batch size is 16 and the patch size is 48×48. Experiments are conducted on 2 NVIDIA Tesla A800 GPUs for about 4 days.

4.1.5 FZU_DXW Team

The team proposes an efficient aggregation restoration network for leveraging inter-frame complementary information effectively, named LGSTANet [20], as shown in Fig. 9.

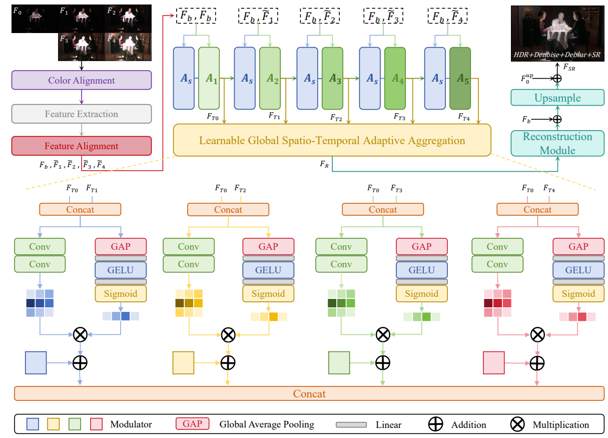


Figure 9. Structure of LGSTANet proposed by FZU_DXW team.

Network Architecture. Inspired by video restoration methods, the team adopts an efficient architecture of alignment, aggregation, and reconstruction. Additionally, they introduce a learnable global spatio-temporal adaptive aggregation module to help effectively aggregate inter-frame complementary information. Furthermore, they propose an adaptive restoration modulator to address specific degradation disturbances of various types, thus achieving high-quality restoration outcomes.

Training Details. The team adopts the progressive training strategy [96], which increases patch size and reduces batch size during training. The patch size of the training includes [128, 160, 192, 256, 320, 384]. They adopt AdamW [59] optimizer with $\beta_1 = 0.9$ and $\beta_1 = 0.999$ for training 800 epochs. The initial learning rate was set to 10^{-4}. Cosine annealing strategy [58] is employed to decrease the learning rates to 10^{-6}. All experiments are conducted on a single NVIDIA GeForce RTX 3090 GPU.

4.2. Track 2: BracketIRE+

4.2.1 Samsung Team †

The solution is roughly the same as stated in Sec. 4.1.1. Some differences are as follows: (1) The first module of RT-IRENet further adds ×4 upsampling operation. (2) During inference, they crop 2 pixels for the input raw images and pad 16 pixels for the output. (3) During training, they utilize model weights from Track 1 to initialize the model weights for Track 2. The input patch size is set to 64 × 64.

4.2.2 NWPU Team

The team proposes a CNN-based model named BSCCNet based on TMRNet[106].

Network Architecture. The model includes three components: alignment, reconstruction, and upsampling module.

† Full name: Samsung MX (Mobile eXperience) Business & Samsung Research China - Beijing (SRC-B).

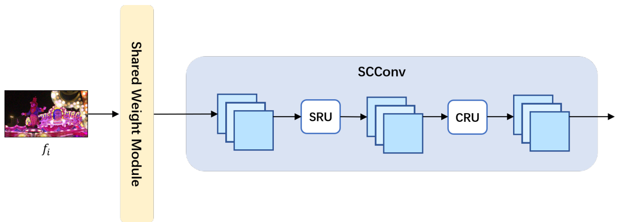


Figure 10. Structure of reconstruction module in BSCCNet proposed by NWPU team. f_i refers to the i^{th} feature after alignment.

In the alignment module, they use flow-guided deformable alignment [11], which is also adopted in TMRNet [106]. In the reconstruction module, they adopt a recurrent mechanism similar to TMRNet, progressively constructing the image through frame-by-frame processing. Fig. 10 shows the reconstruction process. Initially, each frame passes through a convolutional module with shared weights, followed by their respective feature extraction modules. Different from TMRNet, they choose SCConv (Spatial and Channel reconstruction Convolution) [43] as their feature extraction module to emphasize the uniqueness of each frame more. SCConv consists of two units: the Spatial Reconstruction Unit (SRU) and the Channel Reconstruction Unit (CRU), which can help suppress spatial and channel redundancy, aiding in learning more representative features. SCConv not only improves performance but also significantly reduces the model's complexity and computational cost by minimizing redundant features. In the subsequent up-sampling stage, they add the reference frame and use bilinear interpolation to increase the resolution.

Training Details. The team opts for the Adam [37] optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The initial learning rate is set to 10^{-4}. They employ the cosine warm-up method for learning rate reduction with $lr_{min} = 10^{-6}$ and $warmup_t = 5$. The training lasted for 200 epochs and took approximately 120 hours on four NVIDIA A100 GPUs.

4.2.3 FZU_DXW Team

The solution is roughly the same as stated in Sec. 4.1.5. Some differences are as follows: (1) The upsampling operation is added at the end of the proposed model. (2) They first train the ×2 SR model whose weights are initialized with the best model from Track 1. Then they take the pre-trained weights of ×2 SR model to initialize the ×4 SR model. (3) The ×2 and ×4 SR models are trained for 400 and 801 epochs, respectively. The batch size is set to 8 and progressive training strategies are not used.

4.2.4 CYD Team

The proposed model of the team is called HLNet [13], as shown in Fig. 11. The team mainly improves the aggregation module of TMRNet [106].

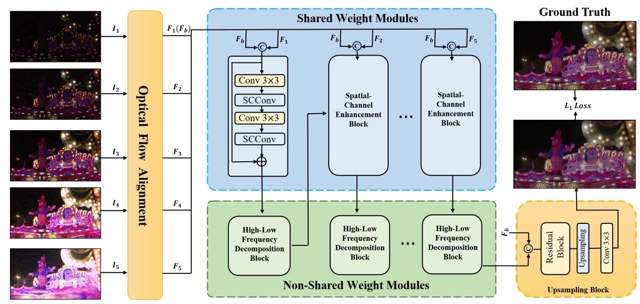


Figure 11. Structure of HLNet proposed by CYD team.

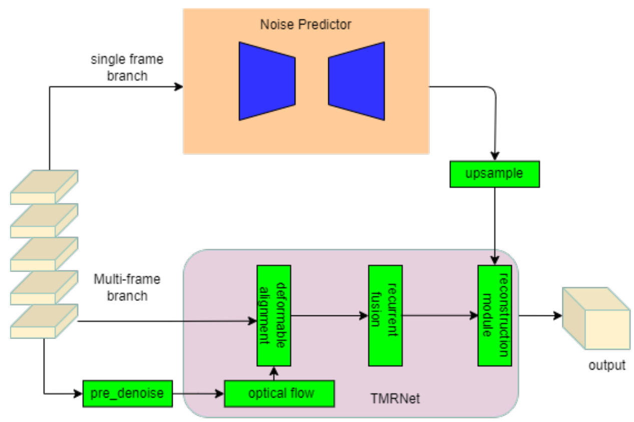


Figure 12. Network architecture proposed by CVG team.

Network Architecture. The aggregation module in TMRNet includes a shared module for all frames and a specific module only for i^{th} frame, which are composed of simple residual blocks. The team replaces the residual blocks of the two modules with Spatial-Channel Enhancement Blocks (SCEB) and High-Low Frequency Separation Blocks (HLFSB), respectively. Specifically, SCEB alternately uses regular convolution and SCConv [43]. SCConv can utilize spatial and channel redundancy and reduce the number of parameters effectively. HLFSB is inspired by ESRT [60], where high- and low- frequency information are processed separately. For high-frequency information, HLFSB adopts small convolution kernels and dense connections, which can better focus on local information, thereby restoring the details of the image. For low-frequency information that requires global information to restore the background and contours of the image, HLFSB adopts multi-scale feature extraction and Transformers to obtain long-distance dependencies. To compensate for the detail loss caused by downsampling, HLFSB fuses features of different scales based on wavelet transform.

Training Details. The team adopts AdamW [59] optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The initial learning rate is set to 10^{-4}, and they use a cosine warmup strategy for learning rate decay. The model is trained for 200 epochs for 6 days on 4 NVIDIA A100 GPUs.

4.2.5 CVG Team

The solution is roughly the same as stated in Sec. 4.1.4. Some differences are as follows: (1) The upsampling operation is added to the end of the proposed two-branch module. (2) For image alignment in the multi-frame branch, they train a noise pre-processing module using extra RAW data collected by the Xiaomi 13 smartphone, as shown in Fig. 12. (3) The multi-frame branch module is trained for about 400 epochs.

Acknowledgements

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).

A. Organizers of NTIRE 2023 Challenge

Title:
NTIRE 2024 Challenge on Bracketing Image Restoration and Enhancement

Members:
Zhilu Zhang[1] (cszlzhang@outlook.com), Shuohao Zhang[1], Renlong Wu[1], Wangmeng Zuo[1], Radu Timofte[2,3]

Affiliations:
[1]Harbin Institute of Technology, China
[2]University of Würzburg
[3]ETH Zürich

B. Track 1: Teams and Affiliations

(1) Samsung MX (Mobile eXperience) Business & Samsung Research China - Beijing (SRC-B)

Title:
High Quality Reference Feature for Two Stage Bracketing Image Restoration and Enhancement

User Name:
baolinv

Members:
Xiaoxia Xing[1] (xx.xing@samsung.com), Hyunhee Park[2], Sejun Song[2], Changho Kim[2], Xiangyu Kong[1], Jinlong Wu[1], Jianxing Zhang[1], Jingfan Tan[1], Zikun Liu[1], Wenhan Luo[3]

Affiliations:
[1]Samsung Research China - Beijing (SRC-B)
[2]Department of Camera Innovation Group, Samsung Electronics
[3]Sun Yat-sen University

(2) MegIRE Team

Title:
Improving Bracket Image Restoration and Enhancement with Flow-Guided Alignment and Enhanced Feature Aggregation

User Name:
NygmaLin

Members:
Wenjie Lin[1] (linwenjie@megvii.com), Chengzhi Jiang[1], Mingyan Han[1], Zhen Liu[1], Ting Jiang[1], Jinting Luo[1], Shen Cheng[1], Linze Li[1], Xinhan Niu[1], Shuaicheng Liu[1]

Affiliations:
[1]Megvii Technology

(3) UPN1 Team

Title:
CRNet: A Detail-Preserving Network for Unified Image Restoration and Enhancement Task

User Name:
CalvinYang

Members:
Kangzhen Yang[1] (yangkangzhen@mail.nwpu.edu.cn), Tao Hu[1], Kexin Dai[1], Genggeng Chen[1], Yongqing Yang[2], Xiangyu Chen[3], Wei Dong[4], Qingsen Yan[1], Yanning Zhang[1]

Affiliations:
[1]School of Computer Science, Northwestern Polytechnical University, Xi'an, China
[2]Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences
[3]Shenzhen Institute of Advanced Technology
[4]Xi'an University of Architecture and Technology

(4) CVG Team

Title:
Bracketing Image Restoration and Enhancement with diffusion model

User Name:
buxianggaimingzi

Members:
Peng Zhang[1] (13797017129@163.com), Yifei Chen[1], Wenbo Xiong[1], Zhijun Song[1], Pu Cheng[1], Taolue Feng[1], Yunqing He[1], Daiguo Zhou[1], Ying Huang[1]

Affiliations:
[1]Mi Lab

(5) FZU_DXW Team

Title:
Learnable Global Spatio-Temporal Adaptive Aggregation

for Bracketing Image Restoration and Enhancement

User Name:
Xinwei

Members:
Xinwei Dai[1] (xinweidai@163.com), Yuanbo Zhou[1], Xintao Qiu[1], Hui Tang[1], Wei Deng[2], Qingquan Gao[1], Tong Tong[1]

Affiliations:
[1]Fuzhou University
[2]Imperial Vision Technology

C. Track 2: Teams and Affiliations

(1) Samsung MX(Mobile eXperience) Business & Samsung Research China - Beijing (SRC-B)

Title:
High Quality Reference Feature for Two Stage Bracketing Image Restoration and Enhancement

User Name:
baolinv

Members:
Jinlong Wu[1] (jinlong.wu@samsung.com), Hyunhee Park[2], Sejun Song[2], Changho Kim[2], Xiangyu Kong[1], Xiaoxia Xing[1], Jianxing Zhang[1], Jingfan Tan[1], Zikun Liu[1], Wenhan Luo[3]

Affiliations:
[1]Samsung Research China-Beijing
[2]Department of Camera Innovation Group, Samsung Electronics
[3]Sun Yat-sen University

(2) NWPU Team

Title:
BSCCNet

User Name:
hhd04

Members:
Kexin Dai[1] (1595421656@qq.com), Kangzhen Yang[1], Tao Hu[1], Xiangyu Chen[2], Yu Cao[3], Qingsen Yan[1], Yanning Zhang[1]

Affiliations:
[1]School of Computer Science, Northwestern Polytechnical University, Xi'an, China
[2]Shenzhen Institute of Advanced Technology
[3]Xi'an University of Architecture and Technology

(3) FZU_DXW Team

Title:
Learnable Global Spatio-Temporal Adaptive Aggregation for Bracketing Image Restoration and Enhancement

User Name:
Xinwei

Members:
Xinwei Dai[1] (xinweidai@163.com), Yuanbo Zhou[1], Xintao Qiu[1], Hui Tang[1], Wei Deng[2], Qingquan Gao[1], Tong Tong[1]

Affiliations:
[1]Fuzhou University
[2]Imperial Vision Technology

(4) CYD Team

Title:
Bracketing Image Restoration and Enhancement with High-Low Frequency Decomposition

User Name:
Mystic_

Members:
Genggeng Chen[1] (chengeng0613@xauat.edu.cn), Xiaowen Ma[1], Peng Wu[1], Wei Dong[2], Qingsen Yan[1]

Affiliations:
[1]School of Computer Science, Northwestern Polytechnical University, Xi'an, China
[2]Xi'an University of Architecture and Technology

(5) CVG Team

Title:
Bracketing Image Restoration and Enhancement with diffusion model

User Name:
darkbridge

Members:
Peng Zhang[1] (13797017129@163.com), Wenbo Xiong[1], Zhijun Song[1], Yifei Chen[1], Pu Cheng[1], Taolue Feng[1], Yunqing He[1], Daiguo Zhou[1], Ying Huang[1]

Affiliations:
[1]Mi Lab

References

[1] Abdelrahman Abdelhamed, Mahmoud Afifi, Radu Timofte, and Michael S Brown. Ntire 2020 challenge on real image denoising: Dataset, methods and results. In *CVPR Workshops*, 2020. 1

[2] Miika Aittala and Frédo Durand. Burst image deblurring using permutation invariant convolutional neural networks. In *ECCV*, 2018. 3

[3] 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. 2

[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. 2

[5] Goutam Bhat, Martin Danelljan, Luc Van Gool, and Radu Timofte. Deep burst super-resolution. In *CVPR*, 2021. 3

[6] Goutam Bhat, Martin Danelljan, Fisher Yu, Luc Van Gool, and Radu Timofte. Deep reparametrization of multi-frame super-resolution and denoising. In *CVPR*, 2021.

[7] Goutam Bhat, Martin Danelljan, Radu Timofte, Yizhen Cao, Yuntian Cao, Meiya Chen, Xihao Chen, Shen Cheng, Akshay Dudhane, Haoqiang Fan, et al. Ntire 2022 burst super-resolution challenge. In *CVPR Workshops*, 2022.

[8] Goutam Bhat, Michaël Gharbi, Jiawen Chen, Luc Van Gool, and Zhihao Xia. Self-supervised burst super-resolution. In *ICCV*, 2023. 3

[9] Tim Brooks, Ben Mildenhall, Tianfan Xue, Jiawen Chen, Dillon Sharlet, and Jonathan T Barron. Unprocessing images for learned raw denoising. In *CVPR*, 2019. 1, 2

[10] 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. 2

[11] Kelvin CK Chan, Shangchen Zhou, Xiangyu Xu, and Chen Change Loy. Basicvsr++: Improving video super-resolution with enhanced propagation and alignment. In *CVPR*, 2022. 6, 8

[12] Meng Chang, Huajun Feng, Zhihai Xu, and Qi Li. Low-light image restoration with short-and long-exposure raw pairs. *IEEE TMM*, 2021. 3

[13] Genggeng Chen, Kexin Dai, Kangzhen Yang, Tao Hu, et al. Bracketing image restoration and enhancement with high-low frequency decomposition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 8

[14] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *ECCV*, 2022. 5

[15] Su-Kai Chen, Hung-Lin Yen, Yu-Lun Liu, Min-Hung Chen, Hou-Ning Hu, Wen-Hsiao Peng, and Yen-Yu Lin. Learning continuous exposure value representations for single-image hdr reconstruction. In *ICCV*, 2023. 1

[16] 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. 2

[17] Yiheng Chi, Xingguang Zhang, and Stanley H Chan. Hdr imaging with spatially varying signal-to-noise ratios. In *CVPR*, 2023. 3

[18] 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 *ICCV*, 2021. 1

[19] 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. 2

[20] Xinwei Dai, Yuanbo Zhou, Xintao Qiu, Hui Tang, Wei Deng, Qinquan Gao, and Tong Tong. Learnable global spatio-temporal adaptive aggregation for bracketing image restoration and enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 7

[21] Mauricio Delbracio and Guillermo Sapiro. Burst deblurring: Removing camera shake through fourier burst accumulation. In *CVPR*, 2015. 3

[22] Michel Deudon, Alfredo Kalaitzis, Israel Goytom, Md Rifat Arefin, Zhichao Lin, Kris Sankaran, Vincent Michalski, Samira E Kahou, Julien Cornebise, and Yoshua Bengio. Highres-net: Recursive fusion for multi-frame super-resolution of satellite imagery. *arXiv preprint arXiv:2002.06460*, 2020. 3

[23] Xiaohan Ding, Xiangyu Zhang, Jungong Han, and Guiguang Ding. Scaling up your kernels to 31x31: Revisiting large kernel design in cnns. In *CVPR*, 2022. 6

[24] Akshay Dudhane, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Burst image restoration and enhancement. In *CVPR*, 2022. 3

[25] Akshay Dudhane, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Burstormer: Burst image restoration and enhancement transformer. *CVPR*, 2023. 3

[26] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafał K Mantiuk, and Jonas Unger. Hdr image reconstruction from a single exposure using deep cnns. *ACM TOG*, 2017. 1

[27] Manfred Ernst and Bartlomiej Wronski. Hdr+ with bracketing on pixel phones, 2021. `https://blog.research.google/2021/04/hdr-with-bracketing-on-pixel-phones.html`. 3

[28] Jan Froehlich, Stefan Grandinetti, Bernd Eberhardt, Simon Walter, Andreas Schilling, and Harald Brendel. Creating cinematic wide gamut hdr-video for the evaluation of tone mapping operators and hdr-displays. In *Digital photography X*, 2014. 2

[29] Clément Godard, Kevin Matzen, and Matt Uyttendaele. Deep burst denoising. In *ECCV*, 2018. 3

[30] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In *CVPR*, 2019. 1

[31] Shi Guo, Xi Yang, Jianqi Ma, Gaofeng Ren, and Lei Zhang. A differentiable two-stage alignment scheme for burst image reconstruction with large shift. In *CVPR*, 2022. 3

[32] Samuel W Hasinoff, Frédo Durand, and William T Freeman. Noise-optimal capture for high dynamic range photography. In *CVPR*, 2010. 2

[33] Samuel W Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM TOG*, 2016. 3

[34] Zhewei Huang, Tianyuan Zhang, Wen Heng, Boxin Shi, and Shuchang Zhou. Real-time intermediate flow estimation for video frame interpolation. In *ECCV*, 2022. 2

[35] Quan Huynh-Thu and Mohammed Ghanbari. Scope of validity of psnr in image/video quality assessment. *Electronics letters*, 2008. 3

[36] Nima Khademi Kalantari, Ravi Ramamoorthi, et al. Deep high dynamic range imaging of dynamic scenes. *ACM TOG*, 2017. 3

[37] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015. 7, 8

[38] Wei-Sheng Lai, Yichang Shih, Lun-Cheng Chu, Xiaotong Wu, Sung-Fang Tsai, Michael Krainin, Deqing Sun, and Chia-Kai Liang. Face deblurring using dual camera fusion on mobile phones. *ACM TOG*, 2022. 3

[39] Bruno Lecouat, Jean Ponce, and Julien Mairal. Lucas-kanade reloaded: End-to-end super-resolution from raw image bursts. In *ICCV*, 2021. 3

[40] Bruno Lecouat, Thomas Eboli, Jean Ponce, and Julien Mairal. High dynamic range and super-resolution from raw image bursts. *ACM TOG*, 2022. 3

[41] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, 2017. 1

[42] Siyeong Lee, Gwon Hwan An, and Suk-Ju Kang. Deep recursive hdri: Inverse tone mapping using generative adversarial networks. In *ECCV*, 2018. 1

[43] Jiafeng Li, Ying Wen, and Lianghua He. Scconv: spatial and channel reconstruction convolution for feature redundancy. In *CVPR*, 2023. 8

[44] 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. 2

[45] Yawei Li, Yulun Zhang, Radu Timofte, Luc Van Gool, Zhijun Tu, Kunpeng Du, Hailing Wang, Hanting Chen, Wei Li, Xiaofei Wang, et al. Ntire 2023 challenge on image denoising: Methods and results. In *CVPR Workshops*, 2023. 1

[46] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *ICCV*, 2021. 1

[47] 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. 2

[48] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *CVPR Workshops*, 2017. 1

[49] Jingbo Lin, Zhilu Zhang, Yuxiang Wei, Dongwei Ren, Dongsheng Jiang, and Wangmeng Zuo. Improving image restoration through removing degradations in textual representations. *arXiv preprint arXiv:2312.17334*, 2023. 1

[50] Jiaming Liu, Chi-Hao Wu, Yuzhi Wang, Qin Xu, Yuqian Zhou, Haibin Huang, Chuan Wang, Shaofan Cai, Yifan Ding, Haoqiang Fan, et al. Learning raw image denoising with bayer pattern unification and bayer preserving augmentation. In *CVPR Workshops*, 2019. 6

[51] Ming Liu, Zhilu Zhang, Liya Hou, Wangmeng Zuo, and Lei Zhang. Deep adaptive inference networks for single image super-resolution. In *ECCV Workshops*, 2020. 1

[52] Shuaizheng Liu, Xindong Zhang, Lingchen Sun, Zhetong Liang, Hui Zeng, and Lei Zhang. Joint hdr denoising and fusion: A real-world mobile hdr image dataset. In *CVPR*, 2023. 3

[53] 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. 2

[54] 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. 2

[55] Yu-Lun Liu, Wei-Sheng Lai, Yu-Sheng Chen, Yi-Lung Kao, Ming-Hsuan Yang, Yung-Yu Chuang, and Jia-Bin Huang. Single-image hdr reconstruction by learning to reverse the camera pipeline. In *CVPR*, 2020. 1

[56] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *CVPR*, 2022. 6

[57] Zhen Liu, Yinglong Wang, Bing Zeng, and Shuaicheng Liu. Ghost-free high dynamic range imaging with context-aware transformer. In *ECCV*, 2022. 3, 6

[58] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. In *ICLR*, 2016. 7

[59] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *ICLR*, 2019. 6, 7, 8

[60] Zhisheng Lu, Juncheng Li, Hong Liu, Chaoyan Huang, Linlin Zhang, and Tieyong Zeng. Transformer for single image super-resolution. In *CVPR Workshops*, 2022. 6, 8

[61] Ziwei Luo, Youwei Li, Shen Cheng, Lei Yu, Qi Wu, Zhihong Wen, Haoqiang Fan, Jian Sun, and Shuaicheng Liu. Bsrt: Improving burst super-resolution with swin transformer and flow-guided deformable alignment. In *CVPR*, 2022. 3

[62] Xintian Mao, Yiming Liu, Fengze Liu, Qingli Li, Wei Shen, and Yan Wang. Intriguing findings of frequency selection for image deblurring. In *AAAI*, 2023. 1

[63] Nancy Mehta, Akshay Dudhane, Subrahmanyam Murala, Syed Waqas Zamir, and Khan. Gated multi-resolution transfer network for burst restoration and enhancement. *CVPR*, 2023. 3

[64] Ben Mildenhall, Jonathan T Barron, Jiawen Chen, Dillon Sharlet, Ren Ng, and Robert Carroll. Burst denoising with kernel prediction networks. In *CVPR*, 2018. 3

[65] Janne Mustaniemi, Juho Kannala, Jiri Matas, Simo Särkkä, and Janne Heikkilä. Lsd_2–joint denoising and deblurring of short and long exposure images with cnns. In *BMVC*, 2020. 3

[66] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, 2017. 1

[67] Yuzhen Niu, Jianbin Wu, Wenxi Liu, Wenzhong Guo, and Rynson WH Lau. Hdr-gan: Hdr image reconstruction from multi-exposed ldr images with large motions. *IEEE TIP*, 2021. 3

[68] Fidel Alejandro Guerrero Peña, Pedro Diamel Marrero Fernández, Tsang Ing Ren, Jorge de Jesus Gomes Leandro, and Ricardo Massahiro Nishihara. Burst ranking for blind multi-image deblurring. *IEEE TIP*, 2019. 3

[69] Eduardo Pérez-Pellitero, Sibi Catley-Chandar, Ales Leonardis, and Radu Timofte. Ntire 2021 challenge on high dynamic range imaging: Dataset, methods and results. In *CVPR Workshops*, 2021. 1, 3

[70] 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. 2

[71] Xuejian Rong, Denis Demandolx, Kevin Matzen, Priyam Chatterjee, and Yingli Tian. Burst denoising via temporally shifted wavelet transforms. In *ECCV*, 2020. 3

[72] Shayan Shekarforoush, Amanpreet Walia, Marcus A Brubaker, Konstantinos G Derpanis, and Alex Levinshtein. Dual-camera joint deblurring-denoising. *arXiv preprint arXiv:2309.08826*, 2023. 3

[73] Xiao Tan, Huaian Chen, Kai Xu, Yi Jin, and Changan Zhu. Deep sr-hdr: Joint learning of super-resolution and high dynamic range imaging for dynamic scenes. *IEEE TMM*, 2021. 3

[74] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *CVPR*, 2018. 1

[75] Steven Tel, Zongwei Wu, Yulun Zhang, Barthélémy Heyrman, Cédric Demonceaux, Radu Timofte, and Dominique Ginhac. Alignment-free hdr deghosting with semantics consistent transformer. In *ICCV*, 2023. 3

[76] Huiyuan Tian, Li Zhang, Shijian Li, Min Yao, and Gang Pan. Multi-depth branch network for efficient image super-resolution. *Image and Vision Computing*, 2024. 6

[77] 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. 2

[78] 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. 2

[79] Yuzhi Wang, Haibin Huang, Qin Xu, Jiaming Liu, Yiqun Liu, and Jue Wang. Practical deep raw image denoising on mobile devices. In *ECCV*, 2020. 2

[80] 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. 2

[81] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 3

[82] Pengxu Wei, Yujing Sun, Xingbei Guo, Chang Liu, Guanbin Li, Jie Chen, Xiangyang Ji, and Liang Lin. Towards real-world burst image super-resolution: Benchmark and method. In *ICCV*, 2023. 3

[83] Patrick Wieschollek, Bernhard Schölkopf, Hendrik PA Lensch, and Michael Hirsch. End-to-end learning for image burst deblurring. In *ACCV*, 2017. 3

[84] Bartlomiej Wronski, Ignacio Garcia-Dorado, Manfred Ernst, Damien Kelly, Michael Krainin, Chia-Kai Liang, Marc Levoy, and Peyman Milanfar. Handheld multi-frame super-resolution. *ACM TOG*, 2019. 3

[85] Renlong Wu, Zhilu Zhang, Shuohao Zhang, Hongzhi Zhang, and Wangmeng Zuo. Rbsr: Efficient and flexible recurrent network for burst super-resolution. In *PRCV*, 2023. 3

[86] Zhihao Xia, Federico Perazzi, Michaël Gharbi, Kalyan Sunkavalli, and Ayan Chakrabarti. Basis prediction networks for effective burst denoising with large kernels. In *CVPR*, 2020. 3

[87] Xiaoxia Xing, Hyunhee Park, Sejun Song, Changho Kim, Xiangyu Kong, Jinlong Wu, et al. High quality reference feature for two stage bracketing image restoration and enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 5

[88] Qingsen Yan, Dong Gong, Qinfeng Shi, Anton van den Hengel, Chunhua Shen, Ian Reid, and Yanning Zhang. Attention-guided network for ghost-free high dynamic range imaging. In *CVPR*, 2019. 3, 6

[89] Qingsen Yan, Weiye Chen, Song Zhang, Yu Zhu, Jinqiu Sun, and Yanning Zhang. A unified hdr imaging method with pixel and patch level. In *CVPR*, 2023. 3

[90] Kangzhen Yang, Tao Hu, Kexin Dai, Genggeng Chen, et al. Crnet: A detail-preserving network for unified image restoration and enhancement task. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 6

[91] 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. 2

[92] Lu Yuan, Jian Sun, Long Quan, and Heung-Yeung Shum. Image deblurring with blurred/noisy image pairs. In *SIGGRAPH*, 2007. 3

[93] 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. 2

[94] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Cycleisp: Real image restoration via improved data synthesis. In *CVPR*, 2020. 1

[95] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *CVPR*, 2021. 1

[96] 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 *CVPR*, 2022. 7

[97] Jiawei Zhang, Jinshan Pan, Jimmy Ren, Yibing Song, Linchao Bao, Rynson WH Lau, and Ming-Hsuan Yang. Dynamic scene deblurring using spatially variant recurrent neural networks. In *CVPR*, 2018. 1

[98] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE TIP*, 2017. 1

[99] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE TIP*, 2018. 1

[100] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Learning a single convolutional super-resolution network for multiple degradations. In *CVPR*, 2018. 1

[101] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 3

[102] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *ECCV*, 2018. 1

[103] Zhilu Zhang, Haolin Wang, Ming Liu, Ruohao Wang, Jiawei Zhang, and Wangmeng Zuo. Learning raw-to-srgb mappings with inaccurately aligned supervision. In *ICCV*, 2021. 1

[104] Zhilu Zhang, Rongjian Xu, Ming Liu, Zifei Yan, and Wangmeng Zuo. Self-supervised image restoration with blurry and noisy pairs. *NeurIPS*, 2022. 3

[105] Zhilu Zhang, Haoyu Wang, Shuai Liu, Xiaotao Wang, Lei Lei, and Wangmeng Zuo. Self-supervised high dynamic range imaging with multi-exposure images in dynamic scenes. In *ICLR*, 2024. 3

[106] Zhilu Zhang, Shuohao Zhang, Renlong Wu, Zifei Yan, and Wangmeng Zuo. Exposure bracketing is all you need for unifying image restoration and enhancement tasks. *arXiv preprint arXiv:2401.00766*, 2024. 1, 2, 3, 4, 5, 6, 7, 8

[107] 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. 2

[108] Yuzhi Zhao, Yongzhe Xu, Qiong Yan, Dingdong Yang, Xuehui Wang, and Lai-Man Po. D2hnet: Joint denoising and deblurring with hierarchical network for robust night image restoration. In *ECCV*, 2022. 3

[109] Ligeng Zhu. Thop: Pytorch-opcounter, 2022. https://pypi.org/project/thop. 4

[110] Yunhao Zou, Chenggang Yan, and Ying Fu. Rawhdr: High dynamic range image reconstruction from a single raw image. In *ICCV*, 2023. 1