Bracketing Image Restoration and Enhancement with High-Low Frequency Decomposition

Genggeng Chen[1†]   Kexin Dai[2†]   Kangzhen Yang[2]   Tao Hu[1,2]   Xiangyu Chen[3]

Yongqing Yang[4]   Wei Dong[1]   Peng Wu[2]   Yanning Zhang[2]   Qingsen Yan[2*]

[1]Xi'an University of Architecture and Technology [2]Northwestern Polytechnical University

[3]University of Macau [4]Xi'an Institute of Optics and Precision Mechanics of CAS

Abstract

In real-world scenarios, due to a series of image degradations, obtaining high-quality, clear content photos is challenging. While significant progress has been made in synthesizing high-quality images, previous methods for image restoration and enhancement often overlooked the characteristics of different degradations. They applied the same structure to address various types of degradation, resulting in less-than-ideal restoration outcomes. Inspired by the notion that high/low frequency information is applicable to different degradations, we introduce HLNet, a Bracketing Image Restoration and Enhancement method based on high-low frequency decomposition. Specifically, we employ two modules for feature extraction: shared weight modules and non-shared weight modules. In the shared weight modules, we use SCConv to extract common features from different degradations. In the non-shared weight modules, we introduce the High-Low Frequency Decomposition Block (HLFDB), which employs different methods to handle high-low frequency information, enabling the model to address different degradations more effectively. Compared to other networks, our method takes into account the characteristics of different degradations, thus achieving higher-quality image restoration.

1. Introduction

In real-world scenarios, the presence of various image degradations makes it challenging to capture high-quality, clear content photos. Low exposure can lead to increased noise, especially in dark areas, potentially causing loss of detail. Similarly, bright areas in high-exposure images may lose detail due to overexposure. Despite nu-
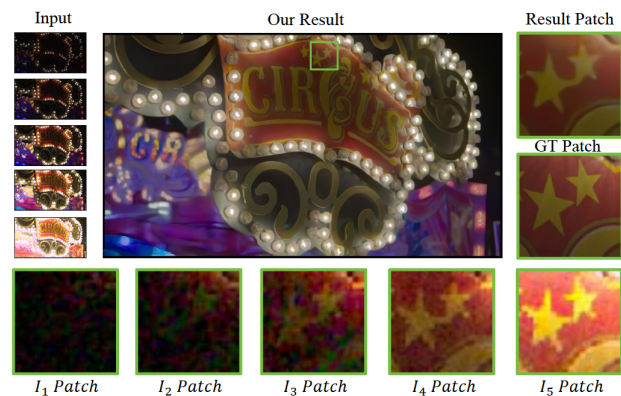
Figure 1. Our HLNet excels in restoring image details and texture, effectively enhancing the edge details of the stars in the figure.

merous single-image restoration methods proposed, such as denoising[1, 5, 17, 28, 51, 53], deblurring[7, 32, 36, 40, 51], super-resolution[10, 26, 29] and high dynamic range image reconstruction[15, 25], their performance is constrained by the insufficient information present in single images.

Due to the limitations of single image restoration and enhancement, such as insufficient information and exposure time impact, an increasing number of methods are utilizing multiple frames for restoration. Burst image restoration methods[13, 14] use multiple consecutive frames for super-resolution and denoising, while multi-exposure HDR imaging[22, 30, 37, 38, 45–50] reconstructs HDR images from LDR images with different exposures. However, these methods only consider a single degradation scenario, overlooking other degradation situations. In recent times, TMRNet[55] has proposed a viable solution in the framework design for unified image restoration and enhancement tasks. It takes multi-exposure images as input and progressively blends non-reference frames with reference frames. Although they have considered the issues of commonality and specificity of different degradations, using shared weight modules and non-shared weight modules, they used the same structure when dealing with the characteristics of

different degradations, ignoring some characteristics of different degradations.

To address the aforementioned issues, we propose HLNet, which takes into account the characteristics of various degradations. Similar to TMRNet, our model utilizes both shared-weight modules and non-shared-weight modules to extract features. In the shared-weight module, we introduce the Spatial Channel Enhancement Block (SCEB), which utilizes SCConv to simultaneously consider spatial and channel information, effectively extracting common features of different degradations. In the non-shared weight module, different degradations are suited to be processed with different frequency information. For instance, denoising and deblurring usually require enhancing high-frequency information to restore image details and textures, super-resolution requires adding high-frequency details on top of restoring low-frequency information, while HDR reconstruction requires capturing low-frequency information from different frames. Hence, we propose the High-Low-Frequency Decomposition Block (HLFDB).

Specifically, in the HLFDB, high-frequency features can capture detailed local information, making them more suitable for denoising and deblurring enhancement. Therefore, we extract local feature maps through multiple convolutional blocks and enhance high-frequency details among multiple frames via dense connection mechanisms. Low-frequency features can capture most structural information and global features in the image, making them more suitable for super-resolution and multi-exposure HDR reconstruction tasks. Hence, we employ multi-level channel self-attention to learn long-range dependencies and utilize a scale-wise feature fusion method based on wavelet transform to avoid the loss of structural information caused by downsampling. Therefore, our model fully considers the characteristics of different degradations when it comes to image restoration and enhancement, enabling better performance in unified image restoration and enhancement tasks.

The main contributions are summarized as follows:

- We propose the SCEB, which utilizes SCConv to simultaneously consider spatial and channel information, effectively extracting common features of different degradations.
- We have proposed HLFDB, which fully considers that different degradations require different high and low-frequency information, effectively addressing the challenge of simultaneously restoring these degradations.
- HLNet outperformed previous state-of-the-art (SOTA) models in both metrics and visual quality, achieved fourth place in track 2 of the Bracketing Image Restoration and Enhancement Challenge.

2. Related Work

Burst Image Restoration Burst image refers to a series of images captured in rapid succession over a short period of time. In this sequence of images, there may be slight variations, such as camera movement, object motion, or changes in lighting conditions. Burst image restoration typically involves several main categories: denoising, deblurring, super-resolution. Many methods focused on denosing have been widely studied in literature [16, 18, 19, 33, 35, 43, 44]. In certain methodologies, recurrent fully convolutional deep neural network [16] are employed, while some others opt for spatially varying kernel estimation [33, 35, 43, 44]. Notably, in [18], offset estimation is additionally leveraged to address challenges arising from substantial object motion. In [19] proposed a two-stage training scheme, sequentially aligning at the patch level and pixel level, to achieve robust alignment between image frames.

Some methods explore the potential of burst image super-resolution [2, 12, 34, 42], Both [42] and [34] employ a transformer architecture; however, [42] diverges by omitting pixel-wise alignment, opting instead for a straightforward homography alignment based on structural geometry. [12]'s central concept revolves around generating a collection of pseudo-burst features, seamlessly amalgamating complementary information from all input burst frames for effective information exchange.

Multi-frame HDR Restoration Multi-frame HDR restoration involves creating High Dynamic Range (HDR) images from multiple low dynamic range (LDR) frames. In classic HDR restoration methods, Debevec *et al.* [9] first proposed the idea of merging multiple LDR images into a single HDR image. Subsequently, many methods have employed ways to align other frames to the reference frame, including optical flow, energy optimization, rank minimization, and others. Zhang and Cham [54] recalibrated motion region weights using image gradients. Bogoni [4] computed flow vectors for alignment purposes. Sen *et al.* [39] employed a patch-based energy minimization method to optimize subsequent alignment and reconstruction. However, these methods perform poorly when faced with significant movement of foreground objects or excessive pixel loss in over-exposed or under-exposed areas.

With the development of deep learning, deep methods have also been applied to the field of multi-frame HDR fusion. an *et al.* [45]proposed a method of attention-guided image fusion, which reduces the presence of ghosting artifacts. SCTNet [41] utilizes spatial attention and channel attention modules, aiming to simultaneously leverage dynamic and static contextual information for better image generation. However, these methods can only handle individual degraded images.
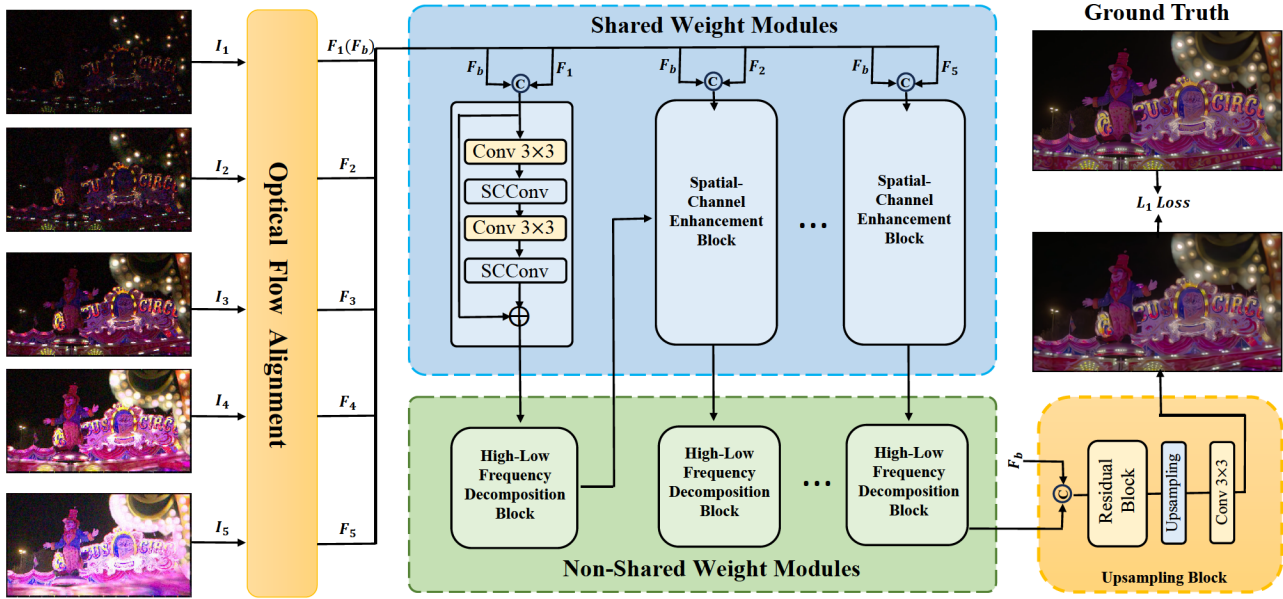
Figure 2. Overview of HLNet. In HLNet, feature alignment is performed first, followed by the gradual feeding of each frame into the network. The feature extraction stage consists of shared weight modules and non-shared weight modules. Each frame needs to pass through both modules, first through the shared weight module and then through the non-shared weight module.

3. Proposed Method

In this study, inspired by the collaborative potential of multi-exposure images and the applicability of high-low frequency information to different degradations, we propose a method for synthesizing and enhancing images through high-low frequency decomposition. The objective is to achieve clear, high dynamic range, and high-resolution images. Specifically, our input consists of five different-exposure raw images, denoted as $\{R_1, R_2, R_3, R_4, R_5\}$, which are then processed.

Initially, the number of multiple exposure images we input is 5, and we label the raw image captured with exposure time Δt_i as Y_i, where $i \in \{1, 2, ..., 5\}$ and $\Delta t_i < \Delta t_{i+1}$. Subsequently, adhering to the guidelines established by multi-exposure HDR reconstruction methodologies[20, 45, 47, 49], we merge the gamma-transformed image Y_r with the original image Y_i through concatenation, resulting in F_i, where $i \in \{1, 2, ..., 5\}$. This process can be formulated as follows:

$$Y_i = \left(\frac{R_i}{\Delta t_i/\Delta t_1}\right), Y_r = \left(\frac{R_i}{\Delta t_i/\Delta t_1}\right)^\gamma, \quad (1)$$

$$F_i = \text{Concat}(Y_i, Y_r), \quad (2)$$

where γ represents the gamma correction parameter and is generally set to $\frac{1}{2.2}$. Finally, we feed these concatenated images into the HLNet model. The resulting image is denoted as $\hat{H}$, and the process can be represented as:

$$\hat{H} = f(F_1, F_2, F_3, F_4, F_5; \theta), \quad (3)$$

where $f(\cdot)$ represents the imaging function, and θ refers to the network's parameters.

3.1. Overview of the HLNet

The main framework of our model is illustrated in Fig. 2. Feature alignment is conducted first, inspired by TMRNet, where both shared-weight modules and non-shared-weight modules are utilized for feature extraction. Finally, upsampling is performed to obtain larger-sized images.

During the feature alignment stage, we select the first frame of the five input images as the reference frame and use optical flow[3] to warp the features of other frames to align with the reference frame. Then, we further align the features using Deformable Convolutional Networks (DCN)[8].

In the feature extraction stage, we utilize both shared-weight modules and non-shared-weight modules. This is because in burst and video restoration tasks, some degradation types among multiple input frames are typically similar. Therefore, we employ shared-weight modules here. Shared weights not only enhance the model's generalization ability, enabling it to learn more universal feature representations, but also reduce the model's parameter count. By contrast, in TMRNet, only regular convolutions were used, without fully considering image details and contextual information. Hence, we introduce the Spatial-Channel Enhancement Block (SCEB), employing SCConv to simultaneously consider spatial and channel information, thus effectively capturing image details and contextual information. Additionally, relying solely on shared-weight modules

is insufficient because other degradations are varying. For instance, differences in exposure time and image blur exist. Therefore, we introduce the High-Low Frequency Decomposition Block (HLFDB) to learn the specificity of different degradation types. This block explicitly separates the high and low-frequency information of the feature and processes them differently based on their characteristics.

At this stage, each aligned frame image F_i, where $i \in \{1, 2, ..., 5\}$, is sequentially inputted into the network. Each frame image first passes through a shared-weight module, followed by a non-shared-weight module. Simultaneously, along with inputting each frame image, we also input the reference frame $F_b(F_1)$ and the output of the previous frame through the non-shared-weight module. This process can be expressed using the following formula:

$$F_{i+1_{out}} = \text{HLFDB}(\text{SCEB}(\text{Concat}(F_{i+1}, F_b, F_{i_{out}}))),$$
(4)

$F_{i+1_{out}}$ denotes the output of the $(i + 1)$-th non-shared-weight module. $\text{SCEB}(\cdot)$ represents the feature map after passing through the Spatial-Channel Enhancement Block, and $\text{HLFDB}(\cdot)$ represents the feature map after passing through the High-Low Frequency Decomposition Block.

In the final stage, the obtained feature maps are upsampled to map low-resolution images or feature maps to high resolution. This process is achieved through skip connections, feature fusion, and multiple upsampling operations.

3.2. Spatial-Channel Enhancement Block

In the Spatial-Channel Enhancement Block, we utilize SCConv[27], a module that integrates spatial and channel dimensions in convolutional operations, enabling more effective capture of image details and contextual information, thus facilitating the extraction of common features across different degradations. In our module, we alternate between regular convolutions and SCConv, augmented with residual connections. This is because regular convolutions focus on capturing local spatial features, effectively extracting texture and shape information from the image. Meanwhile, SCConv considers the relationship between spatial and channel information, enabling a more comprehensive understanding of contextual information and inter-channel dependencies. By alternating between the two, we ensure that the model not only extracts powerful local features but also comprehends their importance in the global context, thus achieving a richer feature representation.The process can be expressed by the following formula:

$$F_{out} = F + \text{SCConv}(\text{Conv}(\text{SCConv}(\text{Conv}(F)))), \quad (5)$$

where F_{out} represents the output of the Spatial-Channel Enhancement Block, $\text{Conv}(\cdot)$ denotes 3×3 convolution, and $\text{SCConv}(\cdot)$ denotes Spatial-Channel Convolution.
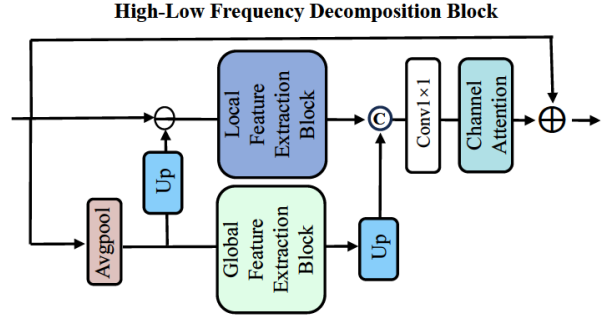
High-Low Frequency Decomposition Block

Figure 3. The architecture of the proposed High-Low Frequency Decomposition Block starts by using Average pooling to extract the low-frequency information from the features. Then, subtracting this low-frequency information from the overall features yields the high-frequency information.

3.3. High-Low Frequency Decomposition Block

To address the issue of the varying characteristics of different degradations during image restoration, we were inspired by ESRT[31] and recognized the applicability of high-low frequency information to different degradation types. High-frequency information typically reflects local details in images, aiding in the restoration of fine image details, which is beneficial for tasks like deblurring and denoising. Meanwhile, low-frequency information generally represents the overall structure of the image, assisting in the restoration of image backgrounds and outlines, which is helpful for tasks like super-resolution reconstruction and multi-frame HDR reconstruction. Therefore, we separate the high and low-frequency information of the image and select different methods for feature extraction based on their characteristics.

As shown in Fig. 3, we choose to use Avgpooling to obtain the low-frequency information of the features. Then, by subtracting the low-frequency information from the overall features, we obtain the high-frequency information of the features. Subsequently, we employ different methods to extract features from these two types of information. This process can be represented by the following formula:

$$F_{\text{low}} = \text{Avgpool}(F), \tag{6}$$

$$F' = \text{Upsampling}(F_{\text{low}}), F_{\text{high}} = F - F', \tag{7}$$

where $\text{Avgpool}(\cdot)$ represents average pooling, and $\text{Upsampling}(\cdot)$ denotes bilinear interpolation upsampling. F_{low} represents the separated low-frequency information, while F_{high} represents the separated high-frequency information.

As high-frequency information represents the details of the image, we need to adopt smaller receptive fields to better focus on local image information for finer detail restoration. To address high-frequency information, we propose the Local Feature Extraction Block for feature extraction,
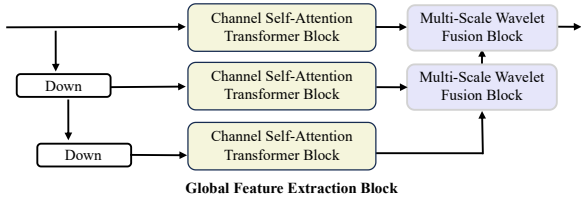
Figure 4. In the architecture of the proposed Global Feature Extraction Block, three successive downsampling operations are required. To prevent the loss of structural information during downsampling, the module employs cross-scale feature fusion based on wavelet transform.

comprising convolutions with multiple small convolution kernels and dense connections. Small convolution kernels allow for better focus on detail areas, while residual connections excel at exploring high-frequency information[11, 24]. Thus, this combination is highly effective for extracting high-frequency information.

For the low-frequency information of images, it contributes to the restoration of image backgrounds and outlines. Given that background and outline information occupies a considerable proportion of the image, having long-distance dependencies is beneficial for their restoration. Therefore, as shown in Fig. 4, we introduce the Global Feature Extraction Block. In this module, we employ multi-scale feature fusion to consider long-distance interactions, and utilize Transformer during feature learning to establish global contextual relationships. Taking inspiration from [52], the Transformer architecture implemented here eschews spatial self-attention in favor of channel-wise self-attention. This is because spatial self-attention would incur an unacceptable computational burden.

In addition, although multi-scale feature extraction can achieve long-range dependencies, structural information may be lost during the downsampling process. Inspired by the ability of wavelet transformation to model scale information in images[21], we propose the Multi-Scale Wavelet Fusion Block for multi-scale information fusion. The Discrete Wavelet Transform separates large-scale feature information into {HH, HL, LH, LL}. We first fuse the original small-scale information with LL, and then use Inverse Discrete Wavelet Transform to fuse the merged information with {HH, HL, LH}. This approach helps avoid the loss of structural information when directly upsampling small-scale feature maps and merging them with large-scale feature maps. Overall, the Global Feature Extraction Block downsamples the input features three times, applies channel-wise self-attention on feature maps of different sizes, and finally utilizes the Multi-Scale Wavelet Fusion Block to merge features of different scales.

3.4. Training Loss

Training the network on tonemapped images is more effective than training directly in the HDR domain due to the common practice of displaying HDR images after tonemapping. Upon receiving an HDR image H in the HDR domain, we compress the image's range through the μ-law transformation.

$$T(H) = \frac{\log(1 + \mu H)}{\log(1 + \mu)}, \tag{8}$$

where μ denotes a parameter that specifies the extent of compression, while $T(H)$ signifies the tonemapped image. Throughout our study, we constrain the values of H to fall within the interval [0, 1], and we fix μ at 5000.

$$L_1 = \|T(H) - T(\hat{H})\|_1, \tag{9}$$

where $\hat{H}$ represents the predicted outcome derived from our HLNet model, while H denotes the ground truth. In this method, we utilize the L_1 loss function to calculate the loss.

4. Experiments

4.1. Experiments Settings

Datasets. The dataset we utilized is the training set from the Bracketing Image Restoration and Enhancement Challenge - Track 2 BracketIRE+ Task. The data was obtained through a simulation process proposed by Zhang [55]. The dataset comprises a total of 1,335 data pairs from 35 scenes. Each data pair includes two inputs of different sizes, ×2 and ×4, with our task utilizing the ×4 size. Each input consists of five frames of raw images with different exposures, sized (4,135,240). In this dataset, 1,045 data pairs from 31 scenes were used for training, while the remaining 290 data pairs from four other scenes were reserved for testing.

Evaluation Metrics. We use two objective measures for quantitative comparison: PSNR-μ, SSIM-μ Here, μ indicate that the metrics are calculated in the tonemapped domain, respectively.

Implementation Details. During the training process, both multiple downsampling operations and small input image sizes can affect training effectiveness. We cropped the images to a size of 64×64 with a stride of 32, ensuring that the feature sizes after multiple downsampling steps were sufficient for effective feature extraction. We utilized the PyTorch framework and employed the AdamW optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Training was conducted on a single A100 GPU using the synthetic dataset provided by the BracketIRE+ Task for a total of 200 epochs, requiring 6 days of training.

4.2. Comparison with the State-of-the-art Methods

To validate the superiority of our model, we compared it with three HDR reconstruction models, two

Table 1. The evaluation results on the Bracketing Image Restoration and Enhancement Challenge - Track 2 BracketIRE+ Task's dataset [55]. We use NVIDIA RTX A100 GPU to calculate the inference time. The best and second best results are highlighted in **Bold** and <u>Underline</u>, respectively.

Models	AHDRNet[45]	CA-ViT[30]	XRestormer[6]	TMRNet[55]	Kim[23]	ESRT[31]	Ours
PSNR-$\mu\uparrow$	26.37	27.44	28.79	28.91	29.02	<u>29.11</u>	**29.66**
SSIM-$\mu\uparrow$	0.8479	0.8518	0.8566	0.8524	0.8571	<u>0.8579</u>	**0.8598**
Time(s)	0.691	0.749	0.807	**0.413**	0.684	0.711	<u>0.642</u>
#Params(M)	17.65	17.90	17.88	13.58	18.07	18.12	17.60

super-resolution models, and one unifying image restoration and enhancement model. These models' tasks are closely related to ours. The HDR reconstruction models, AHDRNet[45], CA-ViT[30], and Kim[23], reconstruct HDR images using multiple LDR frames, similar to our task. The super-resolution models, XRestormer[6] and ESRT[31], are currently among the best-performing super-resolution models. The unifying image restoration and enhancement method's model, TMRNet[55], is consistent with our task. To eliminate the influence of parameter differences, we set the model parameters of these six methods to be roughly the same.

For the three HDR reconstruction methods, we only changed the number of input images, and the model used its original alignment and feature extraction methods, and finally added upsampling to increase the image resolution. For the two super-resolution methods, we only used its feature fusion method in the feature fusion stage, and the alignment method and upsampling were consistent with our model.

As shown in Table 1, our method leads the second-best by 0.55 dB in terms of PSNR-μ. Overall, as depicted in Fig. 5(a), the images restored by our method exhibit the best visual effect to the human eye. In terms of detail recovery, our model outperforms others, effectively restoring the original shapes while other models exhibit significant blurriness. In Fig. 5(b), our method still performs the best in detail recovery. We can observe rich details in the ears of our images, while ears restored by other methods appear blurry. Furthermore, ESRT's results are commendable, attributed to its adoption of high-low frequency decomposition methods, although slightly inferior to our model's performance. In Figure Fig. 5(c), only our method does not exhibit blurriness in the edge details of the stick, while all other methods do.

4.3. Ablation Studies

To validate the effectiveness of each component in our model, we conducted ablation experiments on the Track 2 BracketIRE+ Task dataset. We designed four different ablation experiments to assess the importance of different components, including: (1) Removing the Spatial-Channel

Enhancement Block(SCEB) and replacing it with a simple residual block; (2) Removing the High-Low Frequency Decomposition Block(HLFDB) and replacing it with a simple residual block; (3) Modifying the processing method of high-low frequency information in the High-Low Frequency Decomposition Block; (4) Modify the method of decomposing high-low frequency information. Consequently, a total of 6 different models were generated. The specific details of the models will be elaborated below, and the results of the ablation experiments are shown in Table 2.

- **Model1:** In the shared-weight module, we replaced the SCEB module with a simple residual block, and this model was named HLNet-NoSCEB.
- **Model2:** In the non-shared-weight module, we replaced the HLFDB module with a simple residual block, and this model was named HLNet-NoHLFDB.
- **Model3:** In the HLFDB module, when processing high-low frequency information, we used the Local Feature Extraction Block for both, and this model was named HLNet-LL.
- **Model4:** In the HLFDB module, when processing high-low frequency information, we used the Global Feature Extraction Block for both, and this model was named HLNet-GG.
- **Model5:** In the HLFDB module, when processing high-low frequency information, we use the method of processing high-low frequency information in ESRT to process high-frequency features and low-frequency features, and this model was named HLNet-ESRT.
- **Model6:** In the HLFDB module, we replaced the previous method of decomposing high and low-frequency information using Average Pooling with the wavelet transform method, and this model was named HLNet-Wavelet.

Without Spatial-Channel Enhancement Block. To validate the effectiveness of the Spatial-Channel Enhancement Block, we replaced it with a regular residual block. By referring to Table 2, we observed a decrease of 0.25dB in the PSNR-μ metric compared to our model. Based on this result, we conclude that SCConv, by considering the relationship between spatial and channel information, can capture more comprehensive contextual information and dependencies between channels. The ability to capture
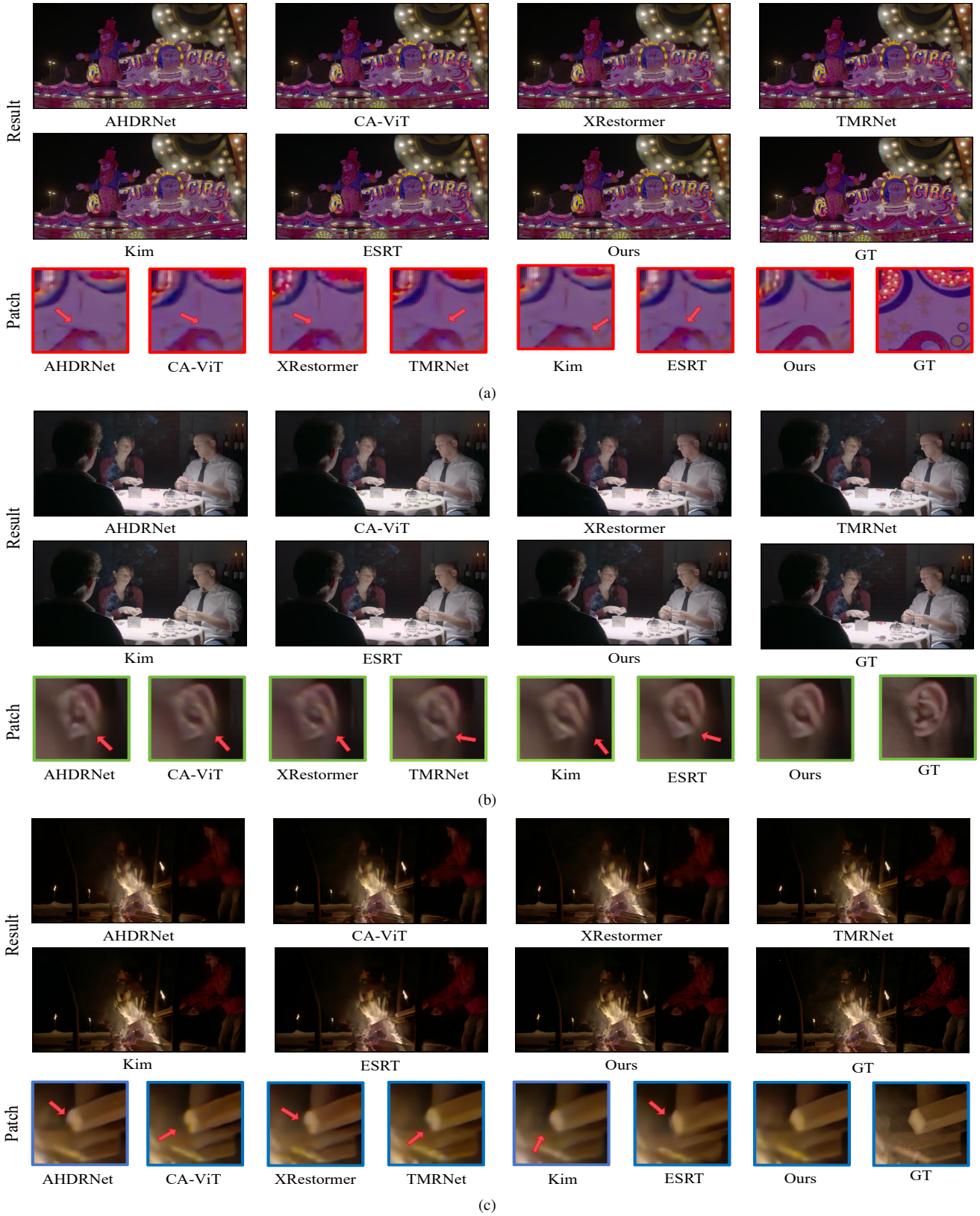
Figure 5. Examples of comparisions on the track 2 of the Bracketing Image Restoration and Enhancement Challenge dataset.

Table 2. The Ablation study of HLNet on the Track 2 BracketIRE+ Task dataset.

Models	PSNR-μ	SSIM-μ
HLNet-NoSCEB	29.41	0.8581
HLNet-NoHLFDB	28.97	0.8560
HLNet-LL	28.69	0.8521
HLNet-GG	28.75	0.8547
HLNet-ESRT	29.27	0.8566
HLNet-Wavelet	29.32	0.8577
Ours	**29.66**	**0.8598**

Figure 6. From the figure, it is evident the importance of the High-Low Frequency Decomposition Block. Models without using HLFDB show poor detail recovery. HLNet-NoHLFDB represents the absence of HLFDB in our model.

contextual information and channel dependencies is highly beneficial for extracting common features across different degradations.

Without High-Low Frequency Decomposition Block. To validate the effectiveness of the High-Low Frequency Decomposition Block, we replaced it with a regular residual block. By referring to Table 2, we observed that this module is crucial, as removing it resulted in a decrease of 0.69dB in the PSNR-μ metric, which is unacceptable. This further confirms the importance of this module. Additionally, we can also perceive the advantages of this module in detail restoration visually. As shown in Fig. 6, removing the High-Low Frequency Decomposition Block significantly reduces the details in the resulting image. There is severe blurring in the hand area. However, with the High-Low Frequency Decomposition Block, the hand area is well restored.

Altering the approach to processing high-low frequency information in the HLFDB module. To validate the rationality and effectiveness of processing high-frequency information with convolution and low-frequency information with Transformer, we designed three sets of experiments: in the first set, both high-frequency and low-

Table 3. Results on track 2 of the Bracketing Image Restoration and Enhancement Challenge[56].

Rank	Team	PSNR ↑ / SSIM ↑ / LPIPS ↓	#Params(M)
1	SRC-B	34.26 / 0.8913 / 0.206	95.00
2	NWPU	30.59 / 0.8728 / 0.268	13.37
3	FZU_DXW	29.82 / 0.8537 / 0.282	14.34
4	CYD	29.66 / 0.8598 / 0.284	17.60
5	CVG	29.25 / 0.8521 / 0.278	71.82
-	TMRNet [55]	28.91 / 0.8572 / 0.273	13.58

frequency information were processed with convolution; in the second set, both high-frequency and low-frequency information were processed with Transformer; in the third set, we used the method of processing high-low frequency information in ESRT to process features. From Table 2, we observed that all three methods yielded poor results. Therefore, we conclude that it is reasonable and effective to process them with different methods according to the different characteristics of high-low frequency information.

Using wavelet transform to decompose high-low frequency information. To validate the effectiveness of using average pooling for decomposing high and low-frequency information, we conducted an experiment using wavelet transform to decompose high and low-frequency information. From Table 2, we can see that although the method of using wavelet transform to decompose high and low-frequency information does not perform as well as using average pooling for decomposition. This is because wavelet transform, when inappropriate wavelet basis functions are chosen, may lead to the loss of important details in the image, affecting the final image quality.

4.4. Results of NTIRE 2024 Challenge on Bracketing Image Restoration and Enhancement - Track 2 BracketIRE+ Task

We participated in the Bracketing Image Restoration and Enhancement Challenge - Track 2 BracketIRE+ Task at NITRE2024 and achieved the fourth place. The results are shown in Table 3. Our PSNR score is 0.75dB higher than the baseline model TMRNet.

5. Conclusion

In this paper, we propose an approach called HLNet, which is based on high-low frequency decomposition for Bracketing Image Restoration and Enhancement. Typically, image restoration requires low-frequency information, while image enhancement requires high-frequency information. Therefore, to uniformly address different degradation issues, we propose a method based on high-low frequency decomposition, which can simultaneously provide the high-frequency and low-frequency information required for degraded images.

References

[1] Abdelrahman Abdelhamed, Mahmoud Afifi, Radu Timofte, and Michael S Brown. Ntire 2020 challenge on real image denoising: Dataset, methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 496–497, 2020. 1

[2] Goutam Bhat, Martin Danelljan, Luc Van Gool, and Radu Timofte. Deep burst super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9209–9218, 2021. 2

[3] Goutam Bhat, Martin Danelljan, Luc Van Gool, and Radu Timofte. Deep burst super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9209–9218, 2021. 3

[4] Luca Bogoni. Extending dynamic range of monochrome and color images through fusion. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, pages 7–12. IEEE, 2000. 2

[5] Tim Brooks, Ben Mildenhall, Tianfan Xue, Jiawen Chen, Dillon Sharlet, and Jonathan T Barron. Unprocessing images for learned raw denoising. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11036–11045, 2019. 1

[6] Xiangyu Chen, Zheyuan Li, Yuandong Pu, Yihao Liu, Jiantao Zhou, Yu Qiao, and Chao Dong. A comparative study of image restoration networks for general backbone network design. *arXiv preprint arXiv:2310.11881*, 2023. 6

[7] 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*, pages 4641–4650, 2021. 1

[8] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 764–773, 2017. 3

[9] Paul E Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pages 643–652. 2023. 2

[10] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015. 1

[11] Jiangxin Dong, Jinshan Pan, Zhongbao Yang, and Jinhui Tang. Multi-scale residual low-pass filter network for image deblurring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12345–12354, 2023. 5

[12] Akshay Dudhane, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Burst image restoration and enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5759–5768, 2022. 2

[13] Akshay Dudhane, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Burst image restoration and enhancement. In *Proceedings of the ieee/cvf*

[14] Akshay Dudhane, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Burstormer: Burst image restoration and enhancement transformer. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5703–5712. IEEE, 2023. 1

[15] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafał K Mantiuk, and Jonas Unger. Hdr image reconstruction from a single exposure using deep cnns. *ACM transactions on graphics (TOG)*, 36(6):1–15, 2017. 1

[16] Clément Godard, Kevin Matzen, and Matt Uyttendaele. Deep burst denoising, 2017. 2

[17] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1712–1722, 2019. 1

[18] Shi Guo, Zhetong Liang, and Lei Zhang. Joint denoising and demosaicking with green channel prior for real-world burst images. *IEEE Transactions on Image Processing*, 30: 6930–6942, 2021. 2

[19] Shi Guo, Xi Yang, Jianqi Ma, Gaofeng Ren, and Lei Zhang. A differentiable two-stage alignment scheme for burst image reconstruction with large shift, 2022. 2

[20] Tao Hu, Qingsen Yan, Yuankai Qi, and Yanning Zhang. Generating content for hdr deghosting from frequency view. *arXiv preprint arXiv:2404.00849*, 2024. 3

[21] Jun-Jie Huang and Pier Luigi Dragotti. Winnet: Wavelet-inspired invertible network for image denoising. *IEEE Transactions on Image Processing*, 31:4377–4392, 2022. 5

[22] Nima Khademi Kalantari, Ravi Ramamoorthi, et al. Deep high dynamic range imaging of dynamic scenes. *ACM Trans. Graph.*, 36(4):144–1, 2017. 1

[23] Jungwoo Kim and Min H Kim. Joint demosaicing and deghosting of time-varying exposures for single-shot hdr imaging. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12292–12301, 2023. 6

[24] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016. 5

[25] Bruno Lecouat, Thomas Eboli, Jean Ponce, and Julien Mairal. High dynamic range and super-resolution from raw image bursts. *arXiv preprint arXiv:2207.14671*, 2022. 1

[26] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 1

[27] Jiafeng Li, Ying Wen, and Lianghua He. Scconv: spatial and channel reconstruction convolution for feature redundancy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6153–6162, 2023. 4

Conference on Computer Vision and Pattern Recognition, pages 5759–5768, 2022. 1

[28] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1904–1920, 2023. 1

[29] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021. 1

[30] Zhen Liu, Yinglong Wang, Bing Zeng, and Shuaicheng Liu. Ghost-free high dynamic range imaging with context-aware transformer. In *European Conference on Computer Vision*, pages 344–360. Springer, 2022. 1, 6

[31] Zhisheng Lu, Juncheng Li, Hong Liu, Chaoyan Huang, Linlin Zhang, and Tieyong Zeng. Transformer for single image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 457–466, 2022. 4, 6

[32] Xintian Mao, Yiming Liu, Fengze Liu, Qingli Li, Wei Shen, and Yan Wang. Intriguing findings of frequency selection for image deblurring. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1905–1913, 2023. 1

[33] Talmaj Marinc, Vignesh Srinivasan, Serhan Gul, Cornelius Hellge, and Wojciech Samek. Multi-kernel prediction networks for denoising of burst images. In *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019. 2

[34] Nancy Mehta, Akshay Dudhane, Subrahmanyam Murala, Syed Waqas Zamir, Salman Khan, and Fahad Shahbaz Khan. Adaptive feature consolidation network for burst super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 1279–1286, 2022. 2

[35] Ben Mildenhall, Jonathan T. Barron, Jiawen Chen, Dillon Sharlet, Ren Ng, and Robert Carroll. Burst denoising with kernel prediction networks, 2018. 2

[36] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3883–3891, 2017. 1

[37] 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 Transactions on Image Processing*, 30:3885–3896, 2021. 1

[38] K Ram Prabhakar, Susmit Agrawal, Durgesh Kumar Singh, Balraj Ashwath, and R Venkatesh Babu. Towards practical and efficient high-resolution hdr deghosting with cnn. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16*, pages 497–513. Springer, 2020. 1

[39] Pradeep Sen, Nima Khademi Kalantari, Maziar Yaesoubi, Soheil Darabi, Dan B Goldman, and Eli Shechtman. Robust patch-based hdr reconstruction of dynamic scenes. *ACM Trans. Graph.*, 31(6):203–1, 2012. 2

[40] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8174–8182, 2018. 1

[41] Steven Tel, Zongwei Wu, Yulun Zhang, Barthélémy Heyrman, Cédric Demonceaux, Radu Timofte, and Dominique Ginhac. Alignment-free hdr deghosting with semantics consistent transformer. *arXiv preprint arXiv:2305.18135*, 2023. 2

[42] 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 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13233–13242, 2023. 2

[43] Zhihao Xia, Federico Perazzi, Michaël Gharbi, Kalyan Sunkavalli, and Ayan Chakrabarti. Basis prediction networks for effective burst denoising with large kernels, 2020. 2

[44] Xiangyu Xu, Muchen Li, and Wenxiu Sun. Learning deformable kernels for image and video denoising, 2019. 2

[45] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1751–1760, 2019. 1, 2, 3, 6

[46] Qingsen Yan, Lei Zhang, Yu Liu, Yu Zhu, Jinqiu Sun, Qinfeng Shi, and Yanning Zhang. Deep hdr imaging via a non-local network. *IEEE Transactions on Image Processing*, 29: 4308–4322, 2020.

[47] Qingsen Yan, Dong Gong, Javen Qinfeng Shi, Anton van den Hengel, Chunhua Shen, Ian Reid, and Yanning Zhang. Dual-attention-guided network for ghost-free high dynamic range imaging. *International Journal of Computer Vision*, pages 1–19, 2022. 3

[48] Qingsen Yan, Weiye Chen, Song Zhang, Yu Zhu, Jinqiu Sun, and Yanning Zhang. A unified hdr imaging method with pixel and patch level. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22211–22220, 2023.

[49] Qingsen Yan, Tao Hu, Yuan Sun, Hao Tang, Yu Zhu, Wei Dong, Luc Van Gool, and Yanning Zhang. Towards high-quality hdr deghosting with conditional diffusion models. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023. 3

[50] Qingsen Yan, Song Zhang, Weiye Chen, Hao Tang, Yu Zhu, Jinqiu Sun, Luc Van Gool, and Yanning Zhang. Smae: Few-shot learning for hdr deghosting with saturation-aware masked autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5775–5784, 2023. 1

[51] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2696–2705, 2020. 1

[52] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang.

Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022. 5

[53] 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

[54] Wei Zhang and Wai-Kuen Cham. Gradient-directed multiexposure composition. *IEEE Transactions on Image Processing*, 21(4):2318–2323, 2011. 2

[55] Zhilu Zhang, Shuohao Zhang, Renlong Wu, Zifei Yan, and Wangmeng Zuo. Bracketing is all you need: Unifying image restoration and enhancement tasks with multi-exposure images. *arXiv preprint arXiv:2401.00766*, 2024. 1, 5, 6, 8

[56] 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. 8