

# A Lightweight Network for High Dynamic Range Imaging

Qingsen Yan<sup>1†</sup>, Song Zhang<sup>2†</sup>, Weiye Chen<sup>3†</sup>, Yuhang Liu<sup>1</sup>, Zhen Zhang<sup>1</sup>,  
Yanning Zhang<sup>4</sup>, Javen Qinfeng Shi<sup>1</sup>, Dong Gong<sup>5\*</sup>

<sup>1</sup>The Australian Institute for Machine Learning, The University of Adelaide, Australia

<sup>2</sup>School of Artificial Intelligence, Xidian University, China

<sup>3</sup>Guangzhou Institute of Technology, Xidian University, China

<sup>4</sup>School of Computer Science, Northwestern Polytechnical University, China

<sup>5</sup>School of Computer Science and Engineering, The University of New South Wales, Australia

## Abstract

Multi-frame high dynamic range (HDR) reconstruction methods try to expand the range of illuminance with differently exposed images. They suffer from ghost artifacts when camera jittering or object moving. Several methods can generate high-quality HDR images with high computational complexity, but the inference process is too slow. However, the network with small parameters will produce unsatisfactory results. To balance the quality and computational complexity, we propose a lightweight network for HDR imaging that has small parameters and fast speed. Specifically, following AHDRNet, we employ a spatial attention module to detect the misaligned regions to avoid ghost artifacts. Considering the missing details in over-/under- exposure regions, we propose a dual attention module for selectively retaining information to force the fusion network to learn more details for degenerated regions. Furthermore, we employ an encoder-decoder structure with a lightweight block to achieve the fusion process. As a result, the high-quality content and features can be reconstructed after the attention module. Finally, we fuse high-resolution features and the encoder-decoder features into the HDR imaging results. Experimental results demonstrate that the proposed method performs favorably against the state-of-the-art methods, achieving a PSNR of 39.05 and a PSNR- $\mu$  of 37.27 with 156.12 GMACs in NTIRE 2022 HDR Challenge (Track 2 Fidelity).

## 1. Introduction

Natural scenes have a wide range of illumination, but a low dynamic range (LDR) image captured by standard digital camera sensors has a limited dynamic range. LDR images often have over-exposure or under-exposure regions

\*† indicates equal contribution.

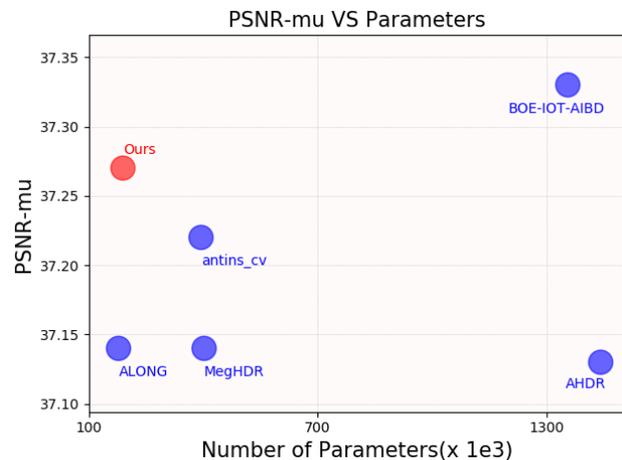


Figure 1. Trade-off of PSNR- $\mu$  and Parameters between the proposed method and several compared methods in the NTIRE 2022 HDR Challenge. The proposed method achieves satisfactory results with fewer parameters.

which severely affect the visual result. On the other hand, High dynamic range (HDR) images can display affluent appearances (e.g., brightness, contrast, and scene details). To obtain a HDR image, the common methods use multiple LDR images captured with different exposure times to combine the well exposed regions. This approach only can generate high-quality HDR image on static scenes. However, ghosting artifacts occur in final HDR image on dynamic scenes or hand-held cameras.

Currently, several solutions have been proposed to solve this problem. The rejection-based methods [3, 4, 6, 7, 10, 16, 18, 21, 26, 37] can fast to detect misaligned regions and remove these regions during the fusion process. Although they have better performance for mostly static scenes, these methods still suffer from ghosting artifacts

when dynamic objects cannot be effectively detected. The alignment-based methods leverage the explicit approach to align the non-reference image to a selected reference image. There are two types of alignment method: rigid registration approaches [24, 25] which cannot cope with complex object motion, and non-rigid registration approaches [5, 11, 14, 35, 38] which are error prone for motions and occlusions. The patch-based methods [8, 12, 23] attempt to generate pure static LDR images from the dynamic input images. This method often can obtain high-quality results than the above methods, but it has high computational complexity and spends more time to infer a scene.

In recent years, with the rise of Convolutional Neural Networks (CNNs), the HDR deghosting [13, 27, 29, 30, 34] have generated visually delightful results, especially for complex dynamic scenes. Unluckily, these networks have the following natural limitations. First, since these methods employ full-resolution convolution networks to hold on the details of input or encoder-decoder network to capture different scales information, these networks need to employ more layers to learn semantics or particulars. Second, existing models usually remove ghosting artifacts with more parameters, which inevitably consumes a large number of computational resources. For example, the recent AHDRNet [29] requires GPU with more memory to process high-resolution images. Similarly, ADNet [15] utilizes a branch to align the input with deformable convolution, which spends a large amount of time predicting result. These methods are not suitable for portable devices with limited computational resources. Although several light-weight deep models (*e.g.*, ALONG Team’s, Antinsecv Team’s) have few parameters, the performances of the evaluation metrics are below ours as shown in Figure 1.

As discussed above, designing a CNN with both high accuracy and high efficiency for reconstructing the ghosting-free HDR image is still a challenge. To alleviate this problem, we propose a hybrid framework to fuse high-resolution features and multi-scale features with a lightweight block. Unlike previous CNN-based methods, we integrate the high-resolution and encoder-decoder structure into a model. Following AHDRNet, we introduce spatial attention to remove the motions and refine the features of non-reference images in the feature extraction stage. In addition, we use a residual operation [2, 29], which highlights useful information (well-exposed) based on dual attention module, to forces the fusion stage to learn more details for degenerated regions (saturation, under-exposure). In the fusion stage, we design a lightweight (LW) module to effectively fuse features with fewer parameters. For the high-resolution branch, we use the depthwise separable convolution to maintain the high-resolution features. On the other hand, we note that the encoder-decoder network tends to rapidly capture a larger receptive field for HDR deghosting.

Thus, we insert LW block into encoder-decoder network to learn different scale features. Finally, the estimated image is generated with a depthwise separable convolution.

To sum up, the main contributions of this paper are three-fold:

- We propose an end-to-end trainable hybrid network called HUNet with two branches (**H**igh-resolution branch and **U**Net branch) for HDR image deghosting and performs better effectively and efficiently than prior work.
- We propose a dual attention module to extract the features of well-exposed regions, and force the fusion network to learn more details for degenerated regions with residual operation.
- We utilize a lightweight (LW) module to effectively fuse features with fewer parameters and achieve better performance.

## 2. Related Works

The related work can be divided into four categories, *i.e.*, rejection-based approach, alignment-based approach, patch-based approach, CNN-based approach.

**Rejection-based approach.** Assuming that the images are globally registered, these approaches [3, 16, 21, 26] use different methods to detect motion regions from static regions. They merge the static regions to generate the final HDR image. Grosch *et al.* [7] estimated a motion map that uses the color constancy criteria of inputs to obtain the ghost-free HDR image. Gallo *et al.* [4] employed patch-wise comparison in the logarithmic domain to detect moving areas. Sparse representation [31] has also been used to detect motions in logarithmic domain. Zhang and Cham [36] used quality measures on image gradients to generate a weighting map over the inputs. Raman and Chaudhuri [22] conducted a comparison in the super-pixels level to improve motion segmentation accuracy along edges. Although they have better performance for mostly static scenes, these methods still suffer from ghosting artifacts when dynamic objects cannot be effectively detected.

**Alignment-based approach.** These approaches align all LDR image to a reference image using rigid or non-rigid algorithms. Bogoni [1] estimated motion vectors using optical flow and used parameters to warp pixels in the exposures. Tomaszewska and Mantuik [24] performed RANSAC after SIFT to refine the matches. Kang *et al.* [14] transformed intensities of LDR images to the luminance domain using exposure time information and computed the optical flow to find corresponding pixels among the LDR images. Pece and Kautz [18] calculated the median threshold bitmap (MTB) for input images to identify motion regions. Zimmer *et al.* [38] used a modern energy-based op-

tic flow approach that takes into account the varying exposure conditions to perform the alignment step. Since the rigid and non-rigid alignment methods are not robust against huge motions, occlusions and brightness changes, they are error-prone for complicated regions.

**Patch-based approach.** The patch-based method can cope with both camera and object motion, which uses dense correspondences to perform patch-based alignment between the exposure images. Sen *et al.* [23] proposed a novel patch-based energy-minimization formulation that integrates alignment and reconstruction in a joint optimization. Hu *et al.* [12] aligned images in an HDR image stack to produce a new exposure stack where all the images are aligned and appear as if they were taken simultaneously, even in the case of highly dynamic scenes. Hafner *et al.* [8] proposed an energy minimization approach that simultaneously calculates HDR irradiance and displacement fields. These approaches reconstruct each HDR region by searching for the best matching patch in LDR images. Although these methods often can obtain high-quality results than the above methods, but they have high computational complexity and spends more time to infer a scene.

**CNN-based approach.** These approaches [13, 28, 29, 32–34] adopt different structures to fuse LDR images to HDR image. Kalantari *et al.* [13] constructed the first HDR imaging dataset and proposed a simple CNN network to fuse inputs that are aligned with optical flow. Yan *et al.* [30] employ multi-scale structure to refine the results. Wu *et al.* [27] proposed an image-wide homography to perform background alignment, and fuse the aligned images to HDR image by CNN. Despite their performance is improved, these approaches still suffer from misalignment and ghosting, especially for fast-moving objects. Yan *et al.* [29] employed an attention mechanism to suppress undesired information before the merging stage. Prabhakar *et al.* [20] employed several networks on a lower resolution and upscaled back to full resolution. Niu *et al.* [17] proposed a GAN-based approach for HDR imaging which is able to synthesize missing details. These methods are not suitable for portable devices with limited computational resources.

### 3. Method

Given a sequence of LDR images of a dynamic scene with different exposure values, our target is to recover an HDR image  $H$  aligned to a reference image  $I_r$ . In this paper, we assign the medium LDR image as  $I_r$ . With the dataset of NTIRE 2022 HDR challenge, we utilize three LDR images  $(I_1, I_2, I_3)$ , we use  $I_2$  as the reference image  $I_r$ .

Following previous works [13, 29], we form a linearized image  $L_i$  for each  $I_i$  as follows:

$$L_i = I_i^\gamma / t_i, \quad (1)$$

where  $t_i$  denotes the exposure time of LDR image  $I_i$ ,  $\gamma$  represents the gamma correction parameter, we set  $\gamma$  to 2.24,  $i = 1, 2, 3$ . As we can see dividing by the exposure time can rectify all the images to have consistent brightness, these  $L_i$  is helpful to detect the motion regions. We concatenate  $I_i$  and  $L_i$  along the channel dimension to form a 6 channel input  $X_i = [I_i; L_i]$ . Given a sequence of inputs  $X_1, X_2, X_3$ , the proposed method produces the HDR image  $\hat{H}$  by:

$$\hat{H} = f(X_1, X_2, X_3; \theta), \quad (2)$$

where  $f(\cdot)$  denotes the network,  $\theta$  is the parameter of the network.

### 3.1. Overview

The proposed method is designed to alleviate the ghosting artifacts for HDR imaging with the lightweight network. The proposed network, called HUNet, consists of two sub-networks, *i.e.*, the attention network and the fusion network (See Figure 2). The first part (attention network) is to detect motions regions, and the second part (fusion network) is to generate the details of degenerated regions. The attention network not only removes the motions of non-reference images, but also highlights the useful regions of the reference image as the guidance for fusion stage with residual learning. Considering the computation complexity and the number of parameters, the fusion network employs UNet architecture with a lightweight module for obtaining a larger receptive field. To decrease the parameters of the network further, we employ a depthwise separable convolution layer and ghost block as the main blocks in UNet.

### 3.2. Attention Network

In the attention network, following AHDRNet, we employ spatial attention to remove misaligned regions between no-reference images and reference images. As discussed before, we obtain  $X_i, i = 1, 2, 3$  from three LDR images, then the attention network extracts features  $F_i$  of  $X_i$  using encoding layer  $e(\cdot)$ .

$$F_i = e(X_i), i = 1, 2, 3 \quad (3)$$

Note that the extracted features  $F_r$  from  $X_r$  (*i.e.*,  $X_2$ ). As shown in Figure 2, to obtain the misalignments between the reference and the non-reference images, we put  $F_i, i = 1, 3$  of the non-reference images into the spatial attention module  $a_i(\cdot), i = 1, 3$  with  $F_r$ . Then, the attention map  $A_i, i = 1, 3$  can be calculated by:

$$A_i = a_i(F_i, F_r). \quad (4)$$

The values of attention map are in the range 0-1. Then, we compute the element-wise multiplication of the  $F_i, i =$

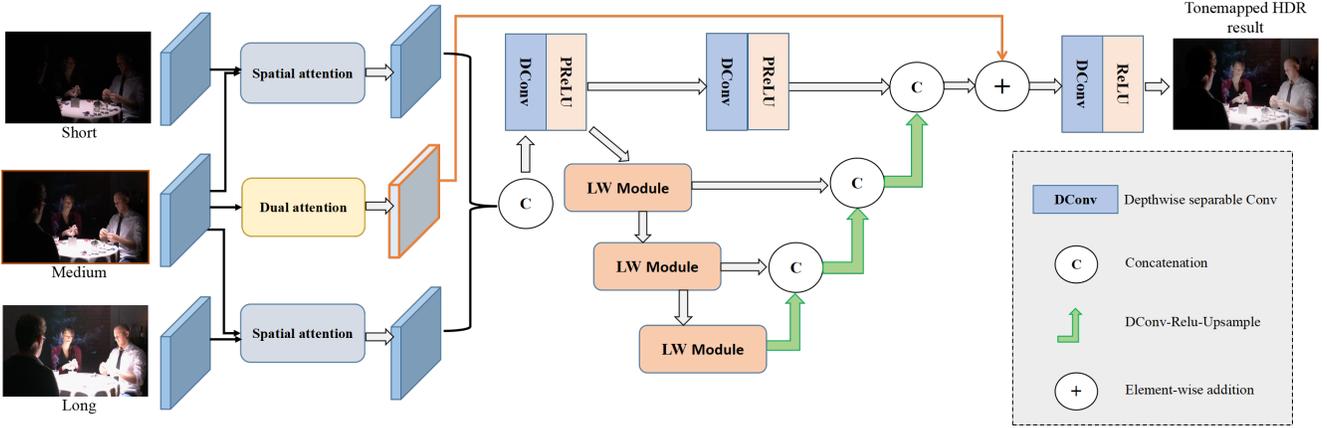


Figure 2. The framework of the proposed method.

1, 3 and the estimated attention map  $A_i$  to obtain the refined feature each non-reference image  $F'_i$ .

$$F'_i = F_i \odot A_i, \quad (5)$$

where  $\odot$  denotes the point-wise multiplication.

Although spatial attention can effectively alleviate misalignment between reference and non-reference images, directly learning clear features is still hard for network [2]. To handle this problem, following [2], we propose a dual attention module to highlight the useful regions of the reference image, which forces the fusion network to learn the details of degenerated regions with a residual. In another word, the highlighted regions of reference image will give a clue for the fusion network, and the fusion network can pay more attention to learn the residual of degenerated regions. As shown in Figure 2, the dual attention module is only used in the medium (reference) frame. The dual attention module consists of spatial attention  $sa(\cdot)$  and channel attention  $ca(\cdot)$  which are helpful to refine the features of the reference image. The spatial attention  $sa(\cdot)$  has several depthwise separable convolutional layers and one sigmoid activation to generate the spatial attention weights. The channel attention  $ca(\cdot)$  includes one global pooling, one fully connected layer and one sigmoid activation to generate the channel attention weights. The refined features of the reference image can be written as:

$$A_2 = ca(sa(F_2)), \quad (6)$$

$$F'_2 = A_2 \odot F_2. \quad (7)$$

After attention network, we obtain  $F'_1, F'_2, F'_3$  which are excluded the harmful information of inputs. Then we stack the extracted features for merging.

$$F = Concat(F'_1, F'_2, F'_3), \quad (8)$$

where  $Concat(\cdot)$  represents the concatenation operation. We will use  $F$  in the fusion network.

### 3.3. Fusion Network

To design a CNN with both high accuracy and high efficiency for feature fusion, we propose a hybrid framework to fuse high-resolution features and multi-scale features with a lightweight block. Unlike previous CNN based methods, we integrate the high-resolution and encoder-decoder structure into a model.

For the high-resolution branch, we use the depthwise separable convolution to maintain the high-resolution features.

$$F_h = PReLU(DC(PReLU(DC(F)))), \quad (9)$$

where  $DC()$  is the depthwise separable convolution layer, and  $PReLU()$  denotes the PReLU activation. We use depthwise separable convolution to decrease the GMACs of the network. On the other hand, we note that the encoder-decoder network tends to rapidly capture a larger receptive field for HDR dehazing. To effectively fuse features with fewer parameters, we design a lightweight (LW) module which can be inserted into encoder-decoder network to learn different scale features. As shown in Figure 2, the encoder-decoder network consists of 3 LW Modules, 3 depthwise separable convolutional layers and upsamplings.

Since the complexity of the encoder-decoder network is dependent on the block in the network, we design a lightweight (LW) module. As shown in Figure 3, the LW module includes a ghost block and depth-wise convolutional layer with stride=2. Considering the redundancy in feature maps, we use the ghost block [9] to generate more feature maps from cheap operations. It is unnecessary to generate these redundant feature maps one by one with a large number of GMACs and parameters [9]. Thus, it decreases the GMACs of the network and improves the performance of the proposed method. Given the input feature  $F_{in} \in \mathcal{R}^{c \times h \times w}$ , where  $c$  is the number of input channels

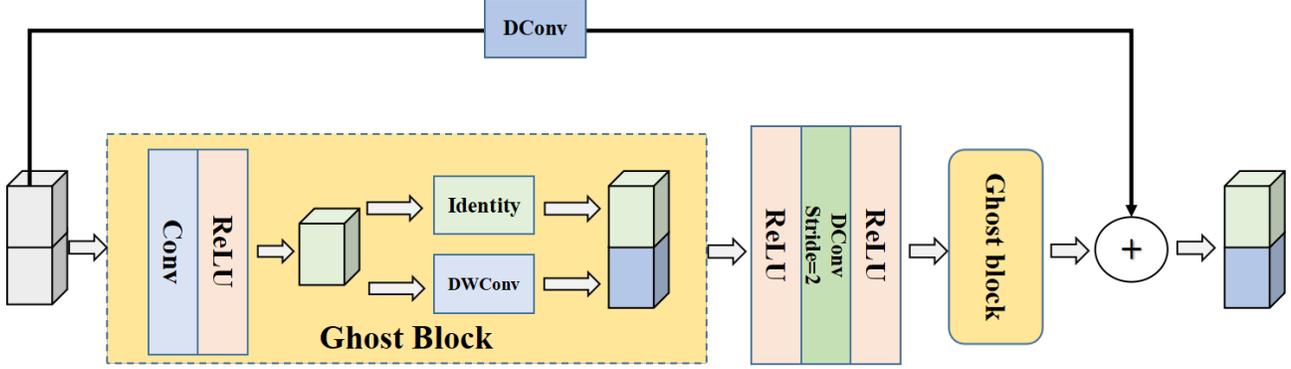


Figure 3. The architecture of the proposed LW module.

and  $h$  and  $w$  are the height and width of the input feature, respectively. First, ghost block learns intrinsic features  $G_1 \in \mathcal{R}^{m \times h \times w}$  by smaller size and produced by standard convolution filters, where  $m = c/2$ .

$$G_1 = \text{ReLU}(\text{Conv}(F_{in})). \quad (10)$$

To further obtain the desired  $c$  channels feature maps, the ghost block applies a cheap linear operations on intrinsic feature  $G_1$  to generate ghost features according to the following function:

$$G_2 = \text{DWConv}(G_1), \quad (11)$$

where  $\text{DWConv}()$  denotes the depth-wise convolution layer. The output is:

$$G = \text{Concat}(G_1, G_2). \quad (12)$$

Thanks to LW modules, the encoder-decoder network has fewer parameters with better performance for HDR imaging. The final HDR image is generated by combing the features of high-resolution and encoder-decoder structure with depth-wise convolutional layer and ReLU activation.

### 3.4. Training Loss

From the proposed network, we can obtain the HDR image  $\hat{H}$ . As the HDR images are not displayed in the linear domain, we use  $\mu$ -law to map the linear domain to the tonemapped domain which is more effective than training directly in the HDR domain.

$$\mathcal{T}(H) = \frac{\log(1 + \mu H)}{\log(1 + H)}, \quad (13)$$

where  $\mathcal{T}(H)$  is the tonemapped image,  $\mu = 5000$ . Based on this tonemapping, we calculate the tonemapped L1 loss between the estimated result and the ground truth:

$$\mathcal{L} = \|\mathcal{T}(H) - \mathcal{T}(\hat{H})\|_1 \quad (14)$$

### 3.5. Implementation Details

In training stage, we crop the  $256 \times 256$  patches with stride 128 for training. We use Adam optimizer and set the batch size and learning rate as 16 and 0.001, respectively, we set  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$  and  $\epsilon = 10^{-8}$ . We implement our model using PyTorch with 2 NVIDIA GeForce 3090 GPUs. We select the best model using the PSNR- $\mu$  score calculated on our validation set.

We use convolution layer to extract 32 features with  $3 \times 3$  kernels. We apply  $3 \times 3$  and  $1 \times 1$  kernels in the Depth-wise separable Conv layers, which are followed by PReLU activations, if not specified otherwise. Depth-wise separable Conv layer is employed in spatial and dual attention. We set the stride size for LW module as 2. We define the output layer to produce 3-channel images.

## 4. Experiments

### 4.1. Experimental Settings

**Dataset.** We use the NTIRE2022 High Dynamic Range Challenge [19] to train and test the proposed method. This dataset includes 1494 different scenes, *we randomly use 149 scenes as the validation set* and keep the remaining as the training set. Each sample has 3 LDR images with different exposures, *i.e.*, short, medium and long exposures, and a related ground-truth HDR image aligned with the central medium frame.

**Evaluation Metrics.** We evaluate the performance with PSNR and PSNR- $\mu$ . PSNR is calculated on the linear domain, PSNR- $\mu$  is PSNR values for images after tonemapping using  $\mu$ -law.

### 4.2. Qualitative Evaluations

We evaluate the performance of the proposed method for the HDR dehazing task and compare it to other state-of-the-art methods. The compared methods contains AHDR-Net [29], DeepHDR [27], ADNet [15], which are all CNN-based methods. Note that ADNet is the champion solu-

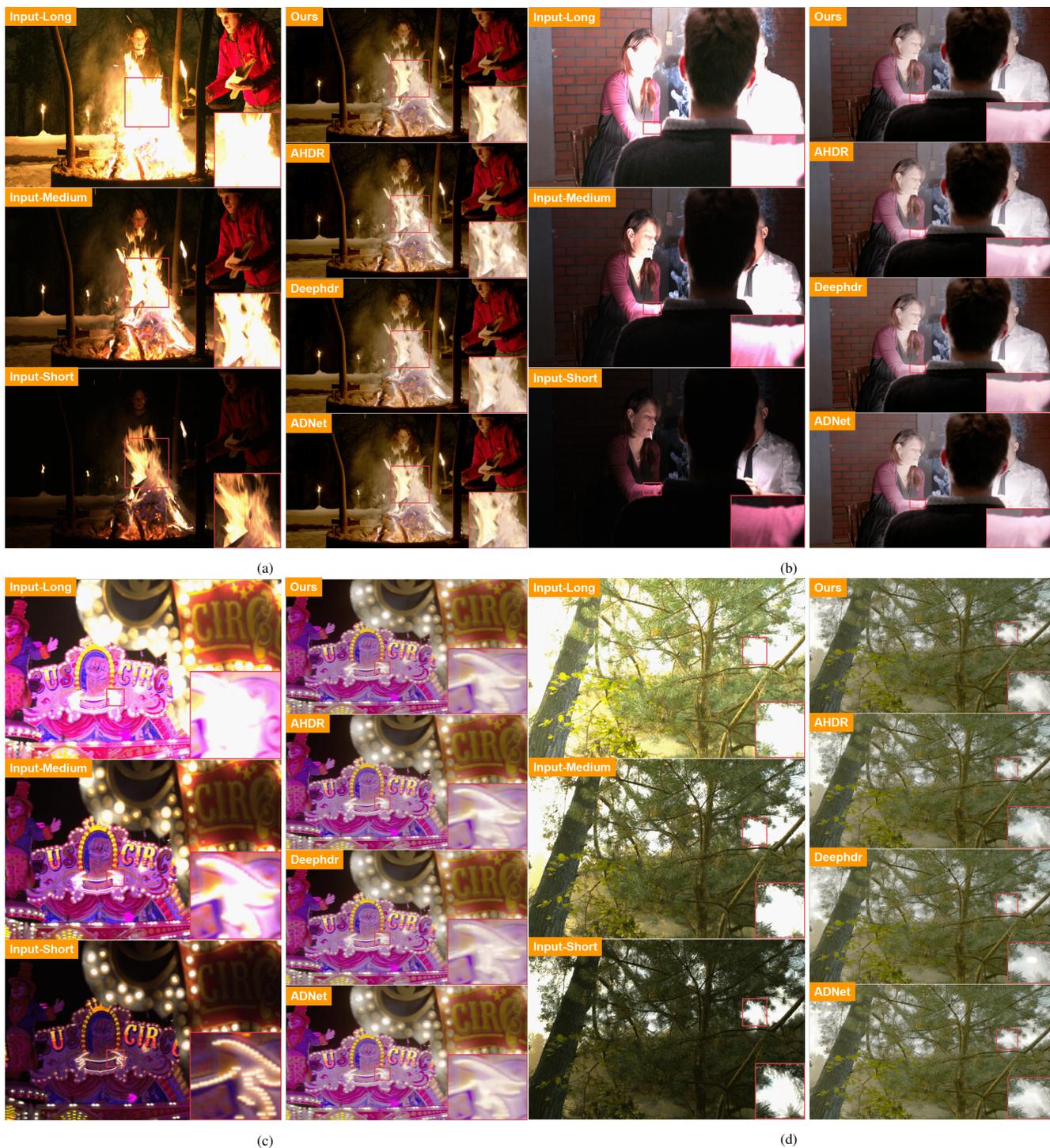


Figure 4. Visual comparisons on the testing data. The LDR images are shown on the left. The propose network can produce a high-quality HDR image.

tion on the NTIRE2021 High Dynamic Range Challenge in Track 2. For fair comparisons, we retrain the compared methods in the challenge data with the same settings.

As shown in Figure 4, we display results in several scenes. The scenes in Figure 4 (a) and (b) are captured with dynamic objects and static camera. For example, the fire in



Figure 5. Attention maps of the spatial attention block. The first row shows 3 LDR images, the second and third rows are attention maps of Long and Short frame.

Figure 4 (a) is changed drastically with severe saturation. And in Figure 4 (b), the lady’s clothes cover over-exposure regions in the medium and long-exposure frame. The proposed method can avoid ghosting artifacts and color distortion, while other compared methods still reconstruct saturated details and blurred boundaries. The scenes in Figure 4 (c) and (d) are captured with dynamic objects and camera. These scenes are more challenging. The proposed method still can recover details of distorted regions. It can be attributed to the proposed hybrid network with two branches for HDR image dehazing. Note that the proposed model has fewer parameters and faster speed.

To verify the visual effect of spatial attention, we display the attention maps of non-reference image in Figure 5. The first row shows 3 LDR images, the second and third rows are attention maps of 1st and 3rd frame. Compared with AHDRNet, the estimated spatial attention map of our method can accurately highlight well-exposure regions, such as the bottle in Figure 5 (a) and the part of fire in Figure 5 (b).

### 4.3. Quantitative Evaluations

To verify the superiority of the proposed HUNet, we compare it with the existing state-of-the-art methods AHDRNet [29], DeepHDR [27], ADNet [15]. To make fair comparison in terms of GMACs, we redesign the small version of DeepHDR and AHDRNet, named DeepHDR\* and AHDRNet\*. From Table 1, we can find that the proposed

method is on-par with the larger models DeepHDR and AHDRNet in terms of PSNR and PSNR- $\mu$ , and 10x smaller than larger models in terms of GMACs. When compared with smaller models, our proposed method has tremendous advantages than DeepHDR\* and AHDRNet\* in terms of PSNR and PSNR- $\mu$ , even with fewer GMACs. We consider that the improved performance can attribute to the proposed dual attention and hybrid fusion stage. Note that since we cannot obtain the ground truth of testing set from NTIRE2022 High Dynamic Range Challenge, we randomly select 149 samples from training set as validation samples (See Sec 4.1), thus the values of PSNR and PSNR- $\mu$  are different with the final official results [19].

### 4.4. Ablation Study

To investigate the effectiveness of the proposed components in HUNet, we design several different variants. The

Table 1. Compared with SOTA methods. \* denotes the model size in 200 GMACs.

Model	PSNR	PSNR- $\mu$	GMACs	Para.
DeepHDR	37.57	31.93	1983.38	16606339
AHDRNet	38.94	32.60	2916.92	1441283
ADNet	39.73	32.88	6249.43	3132773
DeepHDR*	36.16	31.21	180.79	1096706
AHDRNet*	36.32	31.26	186.17	90679
Ours	39.29	32.73	156.12	188992

Table 2. Ablation study on the network structure.

Model	PSNR	PSNR- $\mu$	GMACs	Para.
Baseline	36.32	31.26	186.17	90679
Model1	37.60	31.49	187.78	91511
Model2	38.31	32.22	150.80	188992
Ours	39.29	32.73	156.12	188992

ablation study is conducted by comparing the following variants of HUNet:

- Baseline (*i.e.*, AHDRNet\*). It is the small version of AHDRNet and its model size is less than 200GMACs.
- Model1. We add a dual attention module to AHDRNet\* model.
- Model2. In AHDRNet\*, we replace the fusion stage with the encoder-decoder structure.
- Ours. The full model of the HUNet.

**Spatial and Dual Attention.** Compared with Baseline and Model1 in Table 2, the Model1 with the dual attention shows better performance, and obtains a 1.28dB gain of PSNR. It is mainly because that our proposed dual attention focuses on the useful regions in reference image and reconstructs more reasonable details.

**Fusion network.** The Model2 in Table 2 shows that the hybrid network with two branches, which integrate the information from high-resolution and encoder-decoder, is more effective for fusing features. Compared with the baseline model, the performance of PSNR is improved about 2dB. The main reason can be concluded that our fusion network captures a larger receptive field and learns different scale features, which obtains a better balance between quality and computational complexity.

## 5. Conclusion

In this paper, we propose a hybrid network with two branches for HDR image deghosting. The proposed model performs better effectively and efficiently than prior work. We introduced a dual attention module to highlight details of useful regions in the reference image. We utilized a lightweight module to effectively fuse features with fewer parameters and achieve better performance. We compared our method to several state-of-the-art approaches obtaining a better balance between performance, speed and parameters.

## References

[1] Luca Bogoni. Extending dynamic range of monochrome and color images through fusion. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, volume 3, pages 7–12. IEEE, 2000.

[2] Xiangyu Chen, Yihao Liu, Zhengwen Zhang, Yu Qiao, and Chao Dong. Hdrunet: Single image hdr reconstruction with denoising and dequantization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 354–363, 2021.

[3] Ashley Eden, Matthew Uyttendaele, and Richard Szeliski. Seamless image stitching of scenes with large motions and exposure differences. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 2498–2505. IEEE, 2006.

[4] Orazio Gallo, Natasha Gelfandz, Wei-Chao Chen, Marius Tico, and Kari Pulli. Artifact-free high dynamic range imaging. In *2009 IEEE International conference on computational photography (ICCP)*, pages 1–7. IEEE, 2009.

[5] Orazio Gallo, Alejandro Troccoli, Jun Hu, Kari Pulli, and Jan Kautz. Locally non-rigid registration for mobile hdr photography. In *Proceedings of the IEEE conference on computer vision and pattern recognition Workshops*, pages 49–56, 2015.

[6] Miguel Granados, Kwang In Kim, James Tompkin, and Christian Theobalt. Automatic noise modeling for ghost-free hdr reconstruction. *ACM Transactions on Graphics (TOG)*, 32(6):1–10, 2013.

[7] Thorsten Grosch. Fast and robust high dynamic range image generation with camera and object movement. *Vision, Modeling and Visualization, RWTH Aachen*, pages 277–284, 2006.

[8] David Hafner, Oliver Demetz, and Joachim Weickert. Simultaneous hdr and optic flow computation. In *2014 22nd International Conference on Pattern Recognition*, pages 2065–2070. IEEE, 2014.

[9] Kai Han, Yunhe Wang, Qi Tian, Jianyuan Guo, Chunjing Xu, and Chang Xu. Ghostnet: More features from cheap operations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1580–1589, 2020.

[10] Yong Seok Heo, Kyoung Mu Lee, Sang Uk Lee, Youngsu Moon, and Joonhyuk Cha. Ghost-free high dynamic range imaging. In *Asian Conference on Computer Vision*, pages 486–500. Springer, 2010.

[11] Jun Hu, Orazio Gallo, and Kari Pulli. Exposure stacks of live scenes with hand-held cameras. In *European Conference on Computer Vision*, pages 499–512. Springer, 2012.

[12] Jun Hu, Orazio Gallo, Kari Pulli, and Xiaobai Sun. Hdr deghosting: How to deal with saturation? In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1163–1170, 2013.

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

[14] Sing Bing Kang, Matthew Uyttendaele, Simon Winder, and Richard Szeliski. High dynamic range video. *ACM Transactions on Graphics (TOG)*, 22(3):319–325, 2003.

[15] Zhen Liu, Wenjie Lin, Xinpeng Li, Qing Rao, Ting Jiang, Mingyan Han, Haoqiang Fan, Jian Sun, and Shuaicheng Liu. Adnet: Attention-guided deformable convolutional network for high dynamic range imaging. In *Proceedings of*

- the *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 463–470, 2021.
- [16] Tae-Hong Min, Rae-Hong Park, and SoonKeun Chang. Histogram based ghost removal in high dynamic range images. In *2009 IEEE International Conference on Multimedia and Expo*, pages 530–533. IEEE, 2009.
- [17] 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.
- [18] Fabrizio Pece and Jan Kautz. Bitmap movement detection: Hdr for dynamic scenes. In *2010 Conference on Visual Media Production*, pages 1–8. IEEE, 2010.
- [19] Eduardo Pérez-Pellitero, Sibi Catley-Chandar, Richard Shaw, Ales Leonardi, Radu Timofte, et al. NTIRE 2022 challenge on high dynamic range imaging: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2022.
- [20] 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 *European Conference on Computer Vision*, pages 497–513. Springer, 2020.
- [21] K Ram Prabhakar and R Venkatesh Babu. Ghosting-free multi-exposure image fusion in gradient domain. In *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 1766–1770. IEEE, 2016.
- [22] Shanmuganathan Raman and Subhasis Chaudhuri. Reconstruction of high contrast images for dynamic scenes. *The Visual Computer*, 27(12):1099–1114, 2011.
- [23] 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.
- [24] Anna Tomaszewska and Radoslaw Mantiuk. Image registration for multi-exposure high dynamic range image acquisition. In *International Conference on Computer Graphics, Visualization and Computer Vision*, 2007.
- [25] Greg Ward. Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. *Journal of graphics tools*, 8(2):17–30, 2003.
- [26] Shiqian Wu, Shoulie Xie, Susanto Rahardja, and Zhengguo Li. A robust and fast anti-ghosting algorithm for high dynamic range imaging. In *2010 IEEE International Conference on Image Processing*, pages 397–400. IEEE, 2010.
- [27] Shangzhe Wu, Jiarui Xu, Yu-Wing Tai, and Chi-Keung Tang. Deep high dynamic range imaging with large foreground motions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 117–132, 2018.
- [28] 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*, 130(1):76–94, 2022.
- [29] 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.
- [30] Qingsen Yan, Dong Gong, Pingping Zhang, Qinfeng Shi, Jinqiu Sun, Ian Reid, and Yanning Zhang. Multi-scale dense networks for deep high dynamic range imaging. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 41–50. IEEE, 2019.
- [31] Qingsen Yan, Jinqiu Sun, Haisen Li, Yu Zhu, and Yanning Zhang. High dynamic range imaging by sparse representation. *Neurocomputing*, 269:160–169, 2017.
- [32] Qingsen Yan, Bo Wang, Peipei Li, Xianjun Li, Ao Zhang, Qinfeng Shi, Zheng You, Yu Zhu, Jinqiu Sun, and Yanning Zhang. Ghost removal via channel attention in exposure fusion. *Computer Vision and Image Understanding*, 201:103079, 2020.
- [33] Qingsen Yan, Bo Wang, Lei Zhang, Jingyu Zhang, Zheng You, Qinfeng Shi, and Yanning Zhang. Towards accurate hdr imaging with learning generator constraints. *Neurocomputing*, 428:79–91, 2021.
- [34] 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.
- [35] Qingsen Yan, Yu Zhu, and Yanning Zhang. Robust artifact-free high dynamic range imaging of dynamic scenes. *Multimedia Tools and Applications*, 78(9):11487–11505, 2019.
- [36] Wei Zhang and Wai-Kuen Cham. Gradient-directed multiexposure composition. *IEEE Transactions on Image Processing*, 21(4):2318–2323, 2011.
- [37] Wei Zhang and Wai-Kuen Cham. Reference-guided exposure fusion in dynamic scenes. *Journal of Visual Communication and Image Representation*, 23(3):467–475, 2012.
- [38] Henning Zimmer, Andrés Bruhn, and Joachim Weickert. Freehand hdr imaging of moving scenes with simultaneous resolution enhancement. In *Computer Graphics Forum*, volume 30, pages 405–414. Wiley Online Library, 2011.