Gamma-enhanced Spatial Attention Network for Efficient High Dynamic Range Imaging

Fangya Li[1*] Ruipeng Gang[2†*] Chenghua Li[3] Jinjing Li[1] Sai Ma[2]
Chenming Liu[2] Yizhen Cao[1†]
[1]State Key Laboratory of Media Convergence and Communication(CUC), Beijing 100024, China
[2]Academy of Broadcasting Sciencience, NRTA, Beijing 100866, China
[3]Institute of Automation, Chinese Academy of Science, Beijing 100190, China
fly72109@163.com gangruipeng@abs.ac.cn caoyizh@cuc.edu.cn

Abstract

High dynamic range(HDR) imaging is the task of recovering HDR image from one or multiple input Low Dynamic Range (LDR) images. In this paper, we present Gamma-enhanced Spatial Attention Network(GSANet), a novel framework for reconstructing HDR images. This problem comprises two intractable challenges of how to tackle overexposed and underexposed regions and how to overcome the paradox of performance and complexity trade-off. To address the former, after applying gamma correction on the LDR images, we adopt a spatial attention module to adaptively select the most appropriate regions of various exposure low dynamic range images for fusion. For the latter one, we propose an efficient channel attention module, which only involves a handful of parameters while bringing clear performance gain. Experimental results show that the proposed method achieves better visual quality on the HDR dataset. The code will be available at: https://github.com/fancyicookie/GSANet

Figure 1. The first three rows show are the LDR inputs with various exposures. The final row are our reconstructed HDR.We show differences in the zoomed-in patches.

1. Introduction

Dynamic range is the contrast between the brightest and darkest parts of an image. Most digital photography sensors can only measure a limited fraction of this range. The resulting low dynamic range (LDR) images thus often have over or underexposed regions and don't reflect the human ability to see details in both bright and dark areas of a scene. The high dynamic range imaging technique aims at recovering an HDR image from one or several LDR images. Compared with single-frame HDR imaging, multi-frame HDR imaging is more practical and promising due to its informative bracket LDR inputs. Therefore, during HDR reconstruction, we need to fill in the missing details of various exposure LDR images first.

Some specialized hardware devices [11, 18] have been proposed to produce HDR images directly, but they are usually too expensive to be widely adopted. An alternative is to create HDR content from several LDR images in virtual environments using existing software. However, this approach is mostly explored in the entertainment industry [19].

Recently, several learning-based methods have been explored. Most multi-exposure HDR image reconstruction methods adopt two steps: learning to align the LDR images and merging them to get an HDR image [19]. Kalantari *et al*. proposed the first deep convolutional neural network (CNN) for HDR imaging of dynamic scenes. They first

†corresponding author
∗equal contribution

aligned the LDR images with optical flow and then fused the aligned images with a CNN [10]. However, some researchers argue that classic optical flow algorithms could lead to considerable misalignment errors [13, 15]. ADNet [9] is the first application of deformable alignment module for multi-frame HDR imaging. Their results achieve the best but the method has a high computational cost. Generally, aligning LDR images is one of the steps of the traditional approach. However, this alignment method can lead to artifacts or some other error-prone. By contrast, the correlation-guide feature is more flexible and effective [19], such as using the attention mechanism to exclude misaligned features.

In this paper, we propose Gamma-enhanced Spatial Attention Network(GSANet), a new pipeline to tackle such problems. This network hierarchically utilizes exposure information and can make full use of complementary information from gamma-corrected images to recover missing details for LDR images. Instead of processing LDR images and gamma-corrected images separately, we divide LDRs and gamma-corrected images into three groups and hierarchically conduct information integration. In other words, the method is a two-stage framework. We first integrate information in each group and then fuse information across groups. Specifically, before grouping, we process the first gamma-corrected by a small UNet [16] to remove noise and ghost artifacts. A spatial attention module [22] is used for extracting the first and third groups' attention features for better fusion. Such design is motivated by the intuition that the brightness of the gamma-corrected image is biased towards the medium image. Layers of progressive contrast information can further complement the poor exposure areas of the image.

Our main contribution can be summarized as follows:

- We propose a two-stage pipeline for multi-frame HDR imaging of dynamic scenes. Unlike previous methods, we treat LDR and gamma-corrected images uniformly and divide them into three groups to obtain more details about over and underexposed regions. Then we process each group with spatial attention to extract feature.

- We introduce an efficient channel attention to fuse the concatenated features of LDR images and gamma-corrected images and overcome high complexity cost.

- Experimental results show that the proposed method achieves good performance under the constraint of operations and also has a significant improvement in visual quality.

2. Related Work

We classify HDR reconstruction from multi-frame LDR images into two categories according to whether there are alignment steps or not.

Methods with alignment. These methods argue that the quality of the alignment is crucial in the reconstructed HDR images. A common method of alignment for LDR images applies optical flow algorithms or networks. Kalantari *et al*. [6] proposed the first deep multi-frame HDR imaging method for dynamic scenes. The LDR images are first aligned with optical flow and then blended by a fusion subnet [7]. Chen *et al*. [1] first perform image alignment and HDR fusion in the image space and then in the feature space. Prabhakar *et al*. [15] and Q. Yan *et al*. [23] use PWC-Net [17], a lightweight pyramidal optical flow estimation network for alignment. Another approach is applying deformable convolution for alignment. Liu *et al*. [9] first proposed the alignment of gamma-corrected images with a PCD alignment module instead of optical flow. However, these methods often lead to artifacts due to inaccurate alignment information and also have high complexity.

Methods without alignment. These methods involve direct feature concatenation and attention mechanisms in deep learning methods primarily. Omrani *et al*. [12] proposed to merge LDR images directly and the image wavelet coefficients are used to reconstruct more details and make data reduction. Wu *et al*. [21] formulate simple translation network that can automatically hallucinate plausible HDR details in the presence of total occlusion. Unlike previous methods stacking the LDR images or features for merging directly, Chen *et al*. [2] use only two LDR images to warp the underexposed images to the overexposed images and an attention module is applied to reduce artifacts before being fed to merging network. AHDRNet [22] applies attention modules to guide the merging according to the reference image.

These methods perform gamma correction based on whether the data set has a file about the exposure information. Most of the previous methods do not take advantage of the information available from the exposed image. Liu *et al*. [9] only used gamma-corrected images for alignment. Inspired by the above review methods and gamma-corrected images, we adopt spatial attention mechanisms to process gamma-corrected images with more information. Especially, the work most related to ours is [5], which also reorganized the input frames into several groups. However, in [5], the method is mainly used for super-resolution and the groups are composed of the reference frame. In addition, they pay attention to the temporal information of different video frames. While in LDR to HDR task, we pay more attention to the information supplement of images with different brightness for the same frame. Our method divides input LDR images into three groups and effectively

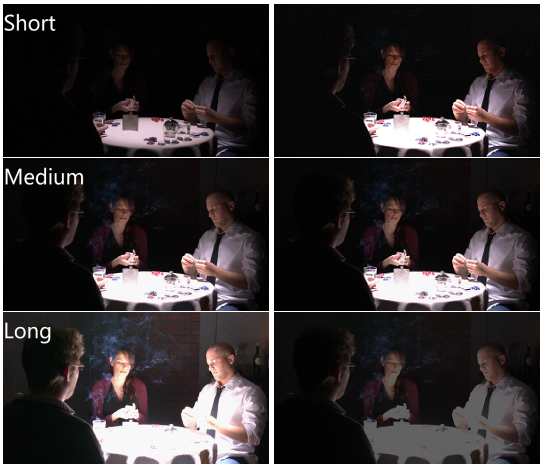


Figure 2. The comparison of LDRs and their gamma-corrected images. The left images are LDR images: short, medium and long. The right images are the corresponding images that mapped from the left images into the HDR domain.

hierarchically integrates gamma-corrected information.

3. Proposed Method

3.1. Overview

In this paper, like other existing learning-based methods, we first map the input LDR images to the HDR domain with gamma correction and then concatenate them directly as the network input. Given 3 LDR images, i.e.,I_i, $i = s, m, l$, that is, I_s, I_m, I_l as input, the gamma transformed outputs are I_s^γ and I_l^γ of the short and long exposure images. Due to the use of exposure information of the LDRs, there are noise and artifacts in the gamma images obtained from the underexposure images. Their image comparison is shown in Fig. 2. The mapping formula is defined as:

$$I_i^\gamma = f(I_i, e_i) \tag{1}$$

where e_i is the exposure information of I_i, γ is the gamma correction parameter, and $f(\cdot)$ denotes the mapping relation of LDR images and their corresponding gamma-corrected images. And note that $I_m = I_m^\gamma$.

As shown in Fig. 3, before grouping, we use one 'convolution+relu' module to extract features of the inputs, denoted as f_s, f_m, f_l, f_l^γ. Each feature has 64 channels. To remove noise and ghost effects, I_s^γ is processed by a small UNet and then the 'convolution+relu' module, and get the cleaned feature $f_s^{\gamma*}$. Then, these features are grouped into three subsets: $\mathcal{G}_1 = \{f_s, f_s^{\gamma*}, f_m\}$, $\mathcal{G}_2 = \{f_s^{\gamma*}, f_m, f_l^\gamma\}$, $\mathcal{G}_3 = \{f_m, f_l^\gamma, f_l\}$. Then the three groups are further processed by an attention network and a fusion network.

$$I^H = \mathcal{F}(\mathcal{A}(\mathcal{G}_1), \mathcal{G}_2, \mathcal{A}(\mathcal{G}_3)) \tag{2}$$

where I^H denotes the reconstructed HDR image. $\mathcal{A}$ denotes a spatial attention module and $\mathcal{G}_1$ and $\mathcal{G}_1$ are processed to get $\mathcal{A}_1^{atten}$ and $\mathcal{A}_3^{atten}$. $\mathcal{F}$ denotes the final fusion net. We concatenate $\mathcal{G}_2$ with $\mathcal{A}_1^{atten}$ and $\mathcal{A}_3^{atten}$ and get the final feature $Fea = [\mathcal{A}_1^{atten}, \mathcal{G}_2, \mathcal{A}_3^{atten}]$. The following fusion and HDR reconstruction steps are made up of efficient channel attention and dilated residual dense block, which takes Fea as input. Finally, the fusion net would output an image with high dynamic range.

3.2. Network Structure

Gamma-corrected images

Like [21], we argue that HDR imaging is an image translation problem where optical flow is not the main problem. Therefore, the crucial problem with HDR reconstruction lies on information fusion on overexposed and underexposed areas for the same frame. Gamma correction is a nonlinear operation on the gray value of the input image, which makes the gray value of the output image and the gray value of the input image show an exponential relationship. So before feeding the LDR images to the network, we first map the input LDR images to the HDR domain relying on gamma correction. As we describe in Sec. 3.1, after using exposures given by official for gamma correction, the details of the overexposed area is shown on I_l^γ. However, the I_s^γ, which comes from underexposed image, becomes significantly more similar to medium images. Therefore it shows some artifacts and noise. (shown in Fig. 2) To address this issue, we propose a small UNet (shown in Fig. 4) for removing noise and ghost effects. The effectiveness of this module is proved by the ablation experiment in Sec. 4.2.2. Later, the LDR images and gamma-corrected images are processed by the 'convolution+relu' module and get all the cleaned feature.

Attention Network

Grouping. In contrast to the previous work, all features f_s, f_m, f_l, f_l^γ and $f_s^{\gamma*}$ are split to three groups: $\mathcal{G}_1 = \{f_s, f_s^{\gamma*}, f_m\}$, $\mathcal{G}_2 = \{f_s^{\gamma*}, f_m, f_l^\gamma\}$, $\mathcal{G}_3 = \{f_m, f_l^\gamma, f_l\}$. Note that the gamma-corrected feature appears in each group. The grouping allows explicit and efficient integration of information about the same frame with different exposure regions for some reasons: 1) Comparing with short and medium images, gamma-corrected image I_s^γ increases the contrast of the dark part of an image. Similarly, I_l^γ reduces the overexposure area and adds the details of the bright part. That is, information of different groups complements each other. 2)The gamma-corrected features in each group guide the model to extract beneficial information from LDR images, allowing efficient information extraction and fusion.

Spatial Attention Module for groups. To better integrate features from different groups, spatial attention module is introduced. In the first group $\mathcal{G}_1$ and the third group $\mathcal{G}_3$, we

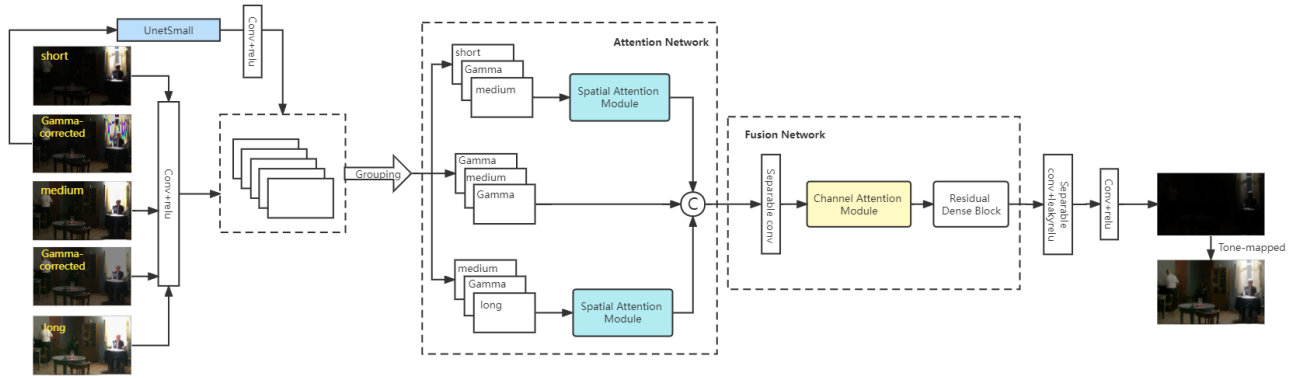


Figure 3. The pipeline of GSANet

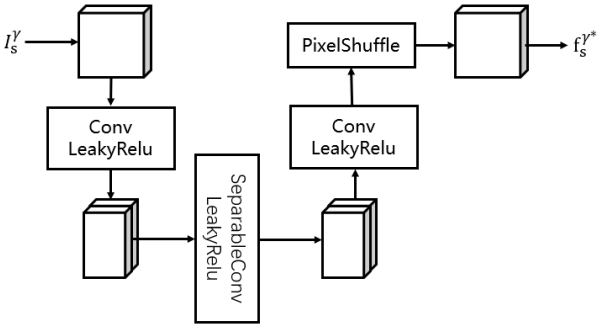


Figure 4. A Small UNet for Denoising. The details can be seen on our code.

concatenate the LDR features with the feature of gamma-corrected image as the input of the spatial attention module, generating the attention map with the range of 0-1. We then compute the element-wise multiplication of the LDR feature and its corresponding attention map to generate the spatial attention feature of each LDR image. From the two groups, we obtain four feature maps i.e.M_j, j = 1,2,3,4. Take example for the group $\mathcal{G}_1$ about specific process of spatial attention module. The process can be formulated as

$$M_j = \mathcal{A}(f_i, f_i^{\gamma*}) \tag{3}$$

where $i = s, m, l$ and M_j, $j = 1, 2$ denotes the attention map. In this paper, we adopt the attention module as used in [22]. The structure of the spatial attention module are shown in Fig. 6. The attention module are two small CNNs. The attention module concatenates the input feature maps f_i and $f_i^{\gamma*}$ and obtains the attention map after two separable conv layers. The two conv layers are followed by ReLU activation and a sigmoid activation. As a result, the 32-channel attention map M_j can be obtained with values in range[0,1]. As shown in Fig. 5, there are the attention feature maps. From the visualization of the features, we argue that the model can get the details of overexposed areas from the first group. The dark information in the background is obtained by the third group. Then the concatenated features are benefit for fusion.

As for M_j, $j = 1, 2$, attention weighted feature for the two groups is calculated as:

$$F_i = M_j \odot f_i \tag{4}$$

where M_j represents the weight of the spatial attention map. F_i represents the group-wise features produced by attention maps and features. '$\odot$' denotes element-wise multiplication. The details of the concatenate is shown in Fig. 7. In order to make full use of attention weighted feature over the groups of gamma-corrected images, we first aggregate those features by concatenating them and feed it into the fusion module.

Fusion Network. The feature maps are concatenated together as input of the fusion subnet. The goal of the fusion module is to aggregate information across different groups and produce the HDR image. Different from the channel attention, the spatial attention focuses on 'where' is an informative part, which is complementary to the channel attention. So the fusion network mainly consists of channel attention module and dilated residual dense block. And in order to reduce the number of operations, separable convolution is used to decrease some channels. We produce a channel attention map by exploiting the inter-channel relationship of features. To compute channels attention efficiently, we compress the spatial dimensions of the input features. To aggregate spatial information, average pooling has been commonly adopted so far. That is, we first aggregate spatial information of a feature map by using both average pooling. We then apply two convolution layer to obtain the raw attention map. The final attention map is normalized by the sigmoid function. We use efficient channel attention [24] to supplement to information that may be lost in the channel. And the usage of dilated convolution in-

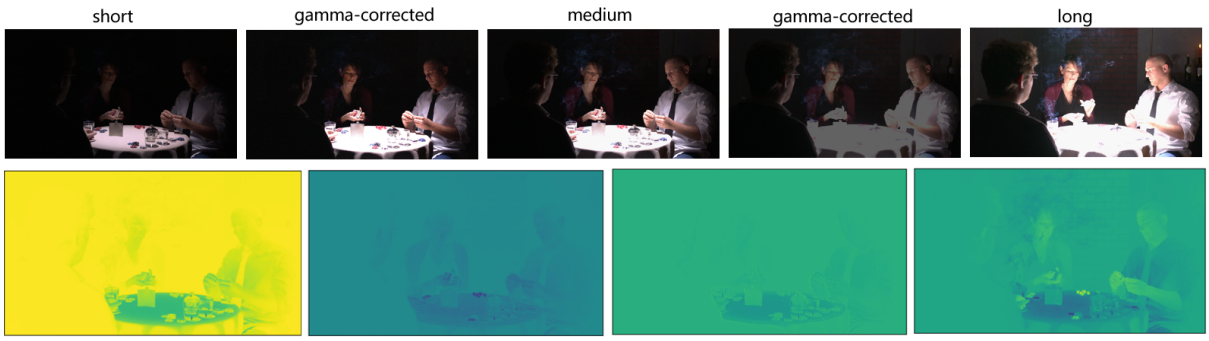


Figure 5. The feature map of the spatial attention. The first three images denotes $\mathcal{G}_1$ and the last three images denotes the $\mathcal{G}_2$. Their feature map comes from the combination of the features of two images above.

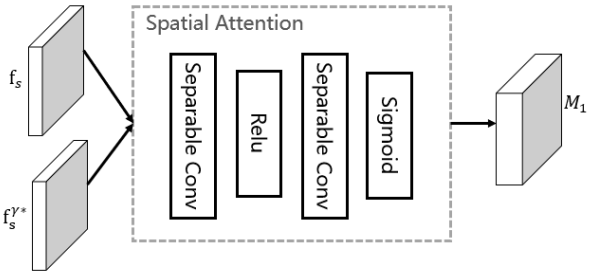


Figure 6. The Spatial Attention Module first concatenates the two inputs and then obtains attention maps.

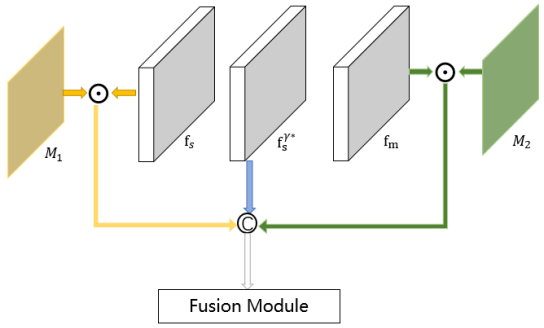


Figure 7. The attention maps is multiplied by its associated feature image and combined with the gamma corrected image.

creases the receptive field. The effectiveness of the module is proved on Sec. 4.2.2.

3.3. Training Loss

Loss functions, such as L1 and L2 loss, are commonly used in previous image restoration work. In HDR reconstruction, HDRUNet [3] argues that it is necessary to consider not only the restoration of the dynamic range, but also the reduction of noise and artifacts. So they propose a specially designed $Tanh\_L_1$ loss for the task. While in ADNet [9], they consider that since HDR images are usually displayed after tone mapped, it is more efficient to train the network on a tonemapping image than directly in the HDR domain. Given an HDR image I^H, they compress the normal range using μ-law:

$$\Gamma(I^H) = \frac{log(1+\mu H)}{log(1+\mu)} \tag{5}$$

where μ ia a parameter that represents the amount of compression and $\Gamma(I^H)$ denotes the tone mapped image. And the loss is defined as:

$$L = \left\|\Gamma(I^H) - \Gamma(I^{GT})\right\| \tag{6}$$

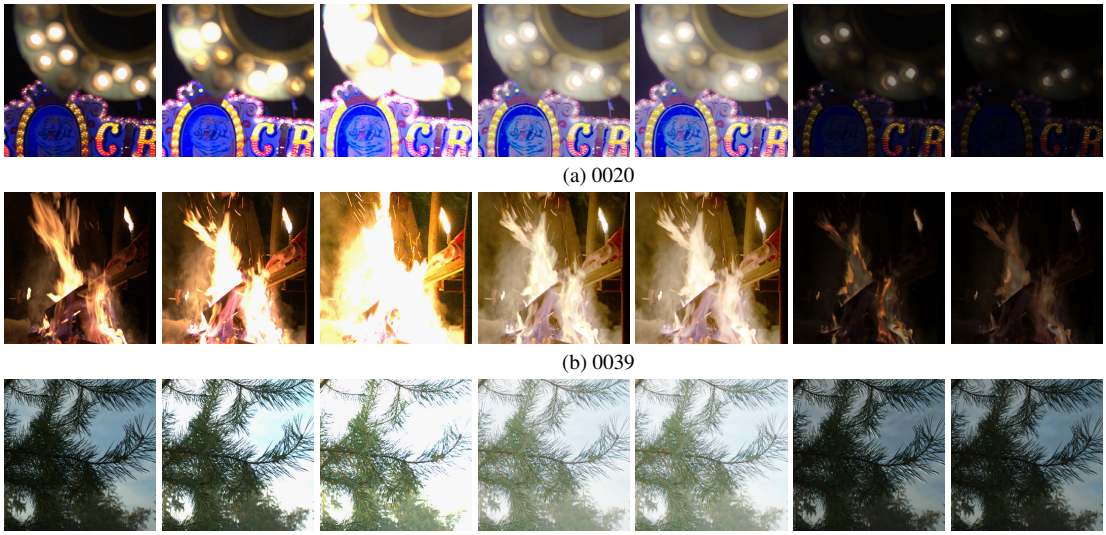
where I^{GT} denotes the tone mapped results of ground truth. Therefore, to obtain better visual quality of the HDR output, we compare a variety of loss and choose the same loss in ADNet, which is called MuLoss. The formula is Eq. (6). The experimental results can be found in Sec. 4.2.

4. Experiment

4.1. Experimental Setup

Dataset. Previous studies [6, 22] train the model on the Kalantari's dataset [7]. In this paper, we use the dataset [4] proposed by NTIRE 2022 HDR Challenge [14]. In the dataset, there are 1494 LDRs/HDR for training, 60 images for validation and 201 images for testing. The 1494 frames consist of 26 long shots. Each scene contains three LDR images, their corresponding exposure and alignment information and HDR ground truth. The size of an image is 1060 × 1900. Since the ground truth of the validation and test sets are not available, we only do experiments on the training set. We select all images as the training set and select 30 images as the validation set. Final scores are based on results from codalab. Although there is a difference between our verified scores and real scores, the trend is the same.

Implementation Details. Before training, we preprocess the data by cropping images into 480×480 with a



(a) 0020

(b) 0039

(c) 0046

Figure 8. Qualitative results of a comparison between our method and ADNet. Our model achieves a good effect visually.

step of 240. During training, the patch size is set to 256×256 and the batch size is set to 8. The number of training iterations is set to 1.5×10^6. Adam optimizer and Kaiming-initialization are adopted for training. The initial learning rate is set to 0.0002 and decayed by a factor of 2 after every 2×10^5 iteration. All models are built on the PyTorch framework and trained with NVIDIA RTX3090 GPU. The total training is about 2 days.

4.2. Ablation Study

In this section, we separately conduct ablation studies on training loss, gamma-corrected images and grouping, spatial attention, channel attention and the model size of GSANet. In the following, we demonstrate the effectiveness of the proposed method in detail.

4.2.1 Training loss.

In Sec. 3.3, we find some different loss. In order to verify which loss function is more efficient, we conduct experiments on various loss functions and make quantitative and qualitative comparisons. The quantitative results are shown in Tab. 1. From the table, we can draw the following observations: 1) In our model, compared with L_1, L_2 and $Tanh\_L_1$, $Tanh\_L_1$ get worse quantitative performance with lower PSNR-L and PSNR-μ. 2) The MuLoss in ADNet can be further improved in PSNR. So we adopt MuLoss as the loss function.

Loss	PSNR-L	PSNR-μ
L_1	36.38	35.48
L_2	36.35	35.39
$Tanh\_L_1$	35.91	35.15
$MuLoss$	36.88	35.56

Table 1. Quantitative comparison of different loss functions

4.2.2 Effectiveness of Key Modules

Gamma-corrected images and Grouping. First, we experiment with different ways of organizing the input images. Supposing input images are $\{1, 2, 3, 4, 5\}$, our grouping method is $\{123, 234, 345\}$. And $\{123, 345\}$ groups are processed by spatial attention. Another method is to conduct spatial attention on only the original three LDRs, similar to ADNet. That is, the group is $\{135\}$. To make sure that the number of operations in different models is as similar as possible, we also select the medium LDR image as the reference image. So the grouping method is $\{135, 234\}$.

Model	$\{12345\}$	$\{135\}$	$\{123, 345\}$	$\{135, 234\}$
SA	×	√	√	√
PSNR-L	36.71	36.72	36.80	36.61
PSNR-μ	34.91	35.50	35.82	35.58

Table 2. SA denotes Spatial Attention module. The ablation experiments are on different grouping strategies. The '×' is the module has no spatial attention. We just concatenate all images.

Spatial Attention. In addition, we also evaluate a model which removes the attention module from our whole model. As shown in Tab. 2, $\{12345\}$ denotes we do not use spatial attention. We concatenate all features directly. The model performs worst among these methods. That proves that spa-

tial attention is useful to integrate information across images. $\{135\}$ means that the spatial attention module is used in the features of LDR images as in ADNet. We do not use gamma-corrected images. We only do this from LDR inputs. The result illustrates that integrating gamma-corrected information is a more effective way in HDR imaging. The $\{123, 345\}$ is better than $\{135, 234\}$, which implies the advantage of add the gamma-corrected images in each groups. Adding the gamma-corrected image in the group encourages the model to extract more complementary information which is missing in the LDR images.

Channel Attention. We investigate the architecture of GSANet and validate the importance of channel attention and small UNet components in the whole GASNet. We achieve this ablation study by comparing the proposed GSANet and the following variants of GSANet.

Denoising. We introduce a small UNet for denoising. It is found in the experiment that this module can improve the score to a certain extent.

In order to investigate the architecture of GSANet and validate the importance of different individual components in the whole model. We achieve this ablation study by comparing the proposed GSANet and the following variants of GSANet:

GSANet. The full model of the GSANet.

GSA-NoUNet. We remove the small UNet for denoising in this variant, in which the feature f_s^γ is directly stacked with other features and fed to the attention network.

GSA-DRDBNet. This variant of GSANet does not contain the channel attention and only uses dilated residual dense block in the fusion network.

GSA-CANet. We do not use DRBD module and only use the channel attention module.

Model	PSNR-L	PSNR-μ
GSANet	36.88	35.57
GSA-NoUNet	36.71	35.51
GSA-DRDBNet	36.35	35.40
GSA-CANet	36.39	35.48

Table 3. Quantitative comparisons of different models. All scores are the average across all testing images.

The experiment results are shown in Tab. 3. Note that we set iteration of 1×10^6 for fast training. If we adopt GSA without UNet for denoising, the PSNR-L and PSNR-μ are 36.71dB and 35.51dB, respectively. By adopting the small UNet, the performance is slightly improved. If we only adopt DRDB module or channel attention module, the scores are shown in the table. With the module of the dilated residual dense block and Channel attention, our full model can further achieve higher quantitative results with PSNR-L of 36.88dB and PSNR-μ of 35.57dB. The results validate the effectiveness of the proposed modules. The visualization of the experiments is in Fig. 9. From this figure, we argue that if the model without denoising, the overall color detail is dark and is accompanied by artifacts in the sunset section. A misty shadow might appear in overexposed areas in the GSA-DRDBNet and GSA-CANet. Our result of the full model is better than these variants.

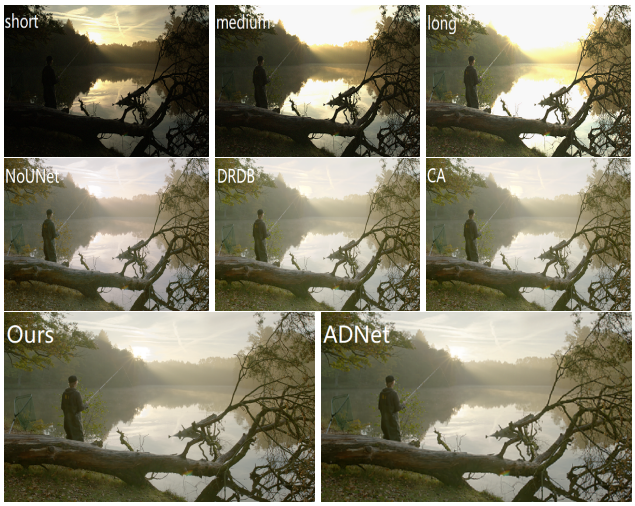


Figure 9. The visual results of variants of GSANet. The first row is LDR inputs and the rests are the results after tonemapping.

Model Size. Due to competition constraints, the number of operations of our model is below 200GMAccs. We found that in our model, generally speaking, the smaller the computation, the smaller the number of parameters in the model. But the number of parameters is not the same trend as the number of operations. That's part of the flaw in our model. In other models with more parameters but less computation, they tend to get better results. The problem we'll look into further.

Settings	GMAcs	Param.(M)	PSNR-L	PSNR-μ
ADNet	6249.43	280	38.71	37.22
+ G(w/G)	–	–	39.01	37.27
- PCD	475.83	0.21	36.97	36.04
- DRDB	243.71	0.16	35.77	35.51
+ SepConv	195.45	0.07	36.35	35.40
+ CA	199.39	0.08	37.57	35.82

Table 4. The details of the changes of model size and the scores based on ADNet.

As listed in Tab. 4, the w/G model denotes that we feed

the gamma-corrected images and grouping in ADNet instead of only LDR images originally. Note that there we have not tested the parameters and GMAccs. But both the parameters and the number of operations are more than ADNet. The results show that ADNet with our gamma-corrected images and grouping produces higher scores, outperforming 0.05dB in terms of PSNR-μ. That proves that the part is useful and can be applied to other models to achieve better results. To reduce the number of operations, our measurement of the original PCD module [9] in ADNet takes a lot of calculation, but the score is decreased by 1.7dB in terms of PSNR-L and 1.2dB in terms of PSNR-μ. And we also test the dilated residual dense block(DRDB). The block takes a few operations but gets higher scores. Then we decide keep dilated residual dense block with SepConv(Separable convolution) to achieve 200 GMAcs. Finally, we found channel attention can get better result. The reason may be that after attention network, the concatenated features get a lot of channels. Although the separable convolution is used to reduce channels and operations, plenty of feature information is missing here. This also is where we're going to improve. Therefore, the proposed method has the potential to get better.

4.3. Comparison with Other Methods

We perform quantitative comparisons of our method with ADNet [9] on the same HDR datasets. To further demonstrate the advantages of the proposed gamma-corrected and grouping in GSANet, we put the part into the ADNet for training and testing. There has never been a limit to complexity in the previous competition. The results shown in Tab. 5.

Model	PSNR-μ	Param(M)	GMAccs
ADNet	37.22	280	6249.43
AD-G(w/G)	37.27	–	–
Ours	35.82	0.08	199.38

Table 5. Quantitative comparisons of application of our the gamma-corrected images and grouping to ADNet.

To demonstrate the superiority and weakness of our proposed GSANet, we compare it with the existing state-of-the-art method ADNet, both quantitatively and qualitatively. As shown in Fig. 8, qualitative results of our method and ADNet are shown in this figure. The (a)–(c) mean the label of the validation images. The first three images are the LDR images: short, medim and long. The fourth and fifth images are results after tone mapping. The last two images are the results from our method and ADNet. Since there is no ground truth of validation set, we compare the final valid result of our model with the visualization result of ADNet model. After tone mapping (the unit16 HDR image is tone mapped to LDR image for visualization), the results look similar to ADNet. So our model achieves a good effect visually. While in unit16 HDR images, the better results look more dark and the overexposed areas appear more blurred. This is also why the difference of PSNR-L value is larger quantitatively.

As listed in Tab. 5, although the score is not best, our model has fewer parameters. The analysis on trading off accuracy for efficiency is shown on Tab. 4. According to the results of the NTIRE2022 [14], we can scale up the parameters to get better scores. First, the separable convolution in fusion network squeezes the channel dimension to 32, leading to information loss. Therefore, we can use channel split [8] to reduce the channel dimension in the bottleneck and add parameters, avoiding too much information loss. In our experiment, the scheme results in a score improvement of 0.07dB with fewer training. Second, our efficient channel attention module also use separable convolution. We find another method that can develop our channel attention module without channel dimensionality reduction [20]. These will be our work in the future.

5. Conclusion

In this paper, we point out that the gamma-corrected images and grouping are crucial to add more details for the overexposed area and underexposed area. And the spatial attention is a representative approach to suppress ghosting effects in HDR imaging. We have presented GSANet, Gamma-enhanced Spatial Attention Network for efficient high dynamic range imaging. A two-stage pipeline is proposed where we handle the LDR images with gamma-corrected and a spatial attention module, and then tackle features with fusion model, wich consists of an efficient channel attention and dilated residual dense block. Experimental results show that although the number of parameters in the proposed model is small, it can achieve well performance and reconstruct HDR images are of good quality visually. Code used in this work will be publicly available upon publication.

6. Acknowledgement

This paper is supported by the "Video Super-Resolution Algorithm Design and Software Development for Face Blur Problem" (JBKY20220210) and "Research and Simulation Experiment of Lightweight Sports Event Remote Production System" (ZZLX-2020-001) projects of the Academy of Broadcasting Science,National Radio and Television Administration and "National Key Research and development Program" (2019YFB1406201) of Key Laboratory of Convergent Media and Intelligent Technology (Communication University of China).

References

[1] Guanying Chen, Chaofeng Chen, Shi Guo, Zhetong Liang, Kwan-Yee K Wong, and Lei Zhang. Hdr video reconstruction: A coarse-to-fine network and a real-world benchmark dataset. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2502–2511, 2021. 2

[2] Sheng-Yeh Chen and Yung-Yu Chuang. Deep exposure fusion with deghosting via homography estimation and attention learning. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1464–1468. IEEE, 2020. 2

[3] 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. 5

[4] 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*, volume 9023, pages 279–288. SPIE, 2014. 5

[5] Takashi Isobe, Songjiang Li, Xu Jia, Shanxin Yuan, Gregory Slabaugh, Chunjing Xu, Ya-Li Li, Shengjin Wang, and Qi Tian. Video super-resolution with temporal group attention. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8008–8017, 2020. 2

[6] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep hdr video from sequences with alternating exposures. In *Computer graphics forum*, volume 38, pages 193–205. Wiley Online Library, 2019. 2, 5

[7] Nima Khademi Kalantari, Ravi Ramamoorthi, et al. Deep high dynamic range imaging of dynamic scenes. *ACM Trans. Graph.*, 36(4):144–1, 2017. 2, 5

[8] Jie Liu, Jie Tang, and Gangshan Wu. Residual feature distillation network for lightweight image super-resolution. In *European Conference on Computer Vision*, pages 41–55. Springer, 2020. 8

[9] 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. 2, 5, 8

[10] Jiayi Ma, Wei Yu, Pengwei Liang, Chang Li, and Junjun Jiang. Fusiongan: A generative adversarial network for infrared and visible image fusion. *Information Fusion*, 48:11–26, 2019. 2

[11] Shree K Nayar and Tomoo Mitsunaga. High dynamic range imaging: Spatially varying pixel exposures. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*, volume 1, pages 472–479. IEEE, 2000. 1

[12] Alireza Omrani, Mohammad Reza Soheili, and Manoochehr Kelarestaghi. High dynamic range image reconstruction using multi-exposure wavelet hdrcnn. In *2020 International Conference on Machine Vision and Image Processing (MVIP)*, pages 1–4. IEEE, 2020. 2

[13] Feiyue Peng, Maojun Zhang, Shiming Lai, Hanlin Tan, and Shen Yan. Deep hdr reconstruction of dynamic scenes. In *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, pages 347–351. IEEE, 2018. 2

[14] Eduardo Pérez-Pellitero, Sibi Catley-Chandar, Richard Shaw, Ales Leonardis, 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. 5, 8

[15] K Ram Prabhakar, Rajat Arora, Adhitya Swaminathan, Kunal Pratap Singh, and R Venkatesh Babu. A fast, scalable, and reliable deghosting method for extreme exposure fusion. In *2019 IEEE International Conference on Computational Photography (ICCP)*, pages 1–8. IEEE, 2019. 2

[16] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 2

[17] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8934–8943, 2018. 2

[18] Jack Tumblin, Amit Agrawal, and Ramesh Raskar. Why i want a gradient camera. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 103–110. IEEE, 2005. 1

[19] Lin Wang and Kuk-Jin Yoon. Deep learning for hdr imaging: State-of-the-art and future trends. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 1, 2

[20] Q. Wang, B. Wu, P. Zhu, P. Li, and Q. Hu. Eca-net: Efficient channel attention for deep convolutional neural networks. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 8

[21] 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. 2, 3

[22] 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. 2, 4, 5

[23] 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. 2

[24] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018. 4