Ultra-High-Definition Image HDR Reconstruction via Collaborative Bilateral Learning

Zhuoran Zheng[1,2,3], Wenqi Ren[2,3*], Xiaochun Cao[3], Tao Wang[4], and Xiuyi Jia[1,2*]
[1]School of Computer Science and Engineering, Nanjing University of Science and Technology
[2]Jiangsu Key Laboratory of Image and Video Understanding for Social Safety, Nanjing University of Science and Technology [3]SKLOIS, IIE, CAS [4]Huawei Noah's Ark Lab

Abstract

Existing single image high dynamic range (HDR) reconstruction methods attempt to expand the range of illuminance. They are not effective in generating plausible textures and colors in the reconstructed results, especially for high-density pixels in ultra-high-definition (UHD) images. To address these problems, we propose a new HDR reconstruction network for UHD images by collaboratively learning color and texture details. First, we propose a dual-path network to extract the content and chromatic features at a reduced resolution of the low dynamic range (LDR) input. These two types of features are used to fit bilateral-space affine models for real-time HDR reconstruction. To extract the main data structure of the LDR input, we propose to use 3D Tucker decomposition and reconstruction to prevent pseudo edges and noise amplification in the learned bilateral grid. As a result, the high-quality content and chromatic features can be reconstructed capitalized on guided bilateral upsampling. Finally, we fuse these two full-resolution feature maps into the HDR reconstructed results. Our proposed method can achieve real-time processing for UHD images (about 160 fps). Experimental results demonstrate that the proposed algorithm performs favorably against the state-of-the-art HDR reconstruction approaches on public benchmarks and real-world UHD images.

1. Introduction

High dynamic images can display rich appearances, such as brightness, contrast, and texture details. However, most mobile devices can only capture images within a limited dynamic range due to the physical limitations of the hardware device. Existing methods fuse LDR images of different exposures into a single HDR image [9, 14]. However, this technique only works well on static scenes, while ghosting artifacts often occur in dynamic scenes or hand-held cameras. Furthermore, it is difficult to get multiple LDR images with different exposure levels in the same scene.

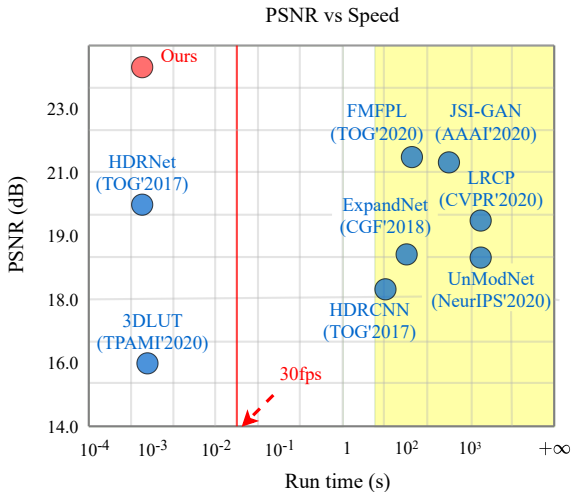


Figure 1. Trade-off of speed and accuracy between our proposed enhancement method and state-of-the-art methods on the FiveK dataset [8]. The red line indicates the real-time method for UHD image reconstruction. The right **yellow region** represents the methods that cannot handle UHD images directly and need to use the downsampling-enhancement-upsampling (DEU) strategy. For example, the maximum resolution can be handled by ExpandNet [26], HDRCNN [15], FMFPL [29] and JSI-GAN [21] is around 2K, while LRCP [25] and UnModNet [42] can only run on images around 512 $\times$ 512 resolution. The proposed algorithm generates enhanced images efficiently and accurately at UHD resolution (4K or more).

Recently, several methods [10, 16, 21, 25, 29, 36, 39, 42] have been developed to reconstruct an HDR image from an LDR input using translational invariance models (CNNs). However, these methods have the following natural limitations [41]. First, since the parameters of existing deep models are fixed, these networks need to enhance saturation issues and texture loss with the same weights. Second, existing models usually enhance an LDR image with the help of a learnable model, which inevitably consumes a large

*Corresponding authors.

number of computational resources. For example, the recent single image HDR reconstruction methods of UnModNet [42] and LRCP [25] cannot directly enhance ultra high resolution images (4K) on a GPU with 24G RAM. Although the early light-weight deep models HDRCNN [15] and ExpandNet [26] can run on 2K images, the performances of the evaluation metrics are below FMFPL [29] and ours as shown in Figure 1. Therefore, recovering lost edges and colors from LDR images is still a tricky problem.

In summary, designing a deep network with both high accuracy and high efficiency for reconstructing the edges and colors of UHD images is still a challenge. To achieve this, we propose a collaborative learning framework to fuse various information with an efficient and interpretable filtering module in bilateral space [11]. We design a dual-path network with edge-aware affine modules for collaborative learning color and texture details. Specially, our algorithm extracts low-resolution content and chromatic features for bilateral grids learning and restores two high-quality feature maps (one is mainly focused on edge and texture, the other is for color). However, we note that the learned bilateral grid by the dual-path network tends to result in new edges, a halos, or noises in the restored image. Therefore, we present a 3D Tucker reconstruction scheme to prevent pseudo edges and noises amplification based on the low-rank characteristic. Finally, we fuse the two high-resolution features to yield the reconstructed UHD HDR image.

Since the change of the bilateral grid occurs according to the content and color of the local area, our proposed algorithm enables the recovery of spatial changes. The proposed dual-path network can also help refine color and texture details in the learned bilateral affine model. In addition, our method processes UHD images in less than 6 ms on a single Titan RTX GPU.

The contributions of this paper are as summarized as:

- We propose a new dual-path network by collaboratively learning textural and chromatic features in the bilateral space, which enables the proposed network to process a UHD HDR image in real-time.
- We enforce a smoothness term in the bilateral grid learning process by a 3D Tucker reconstruction block, which prevents pseudo edges and noises amplification in the reconstructed results.
- We propose a LeakAdaIN and a self-evolving loss function for training acceleration and visual perception enhancement. Experimental results on synthetic and real-world images demonstrate the proposed algorithm performs favorably against the state-of-the-art HDR reconstruction methods on arbitrary spatial sizes.

2. Related Work

Multi-image HDR reconstruction. The most traditional HDR reconstruction techniques rely on multiple exposure LDR images. [14, 32]. Image alignment and post-procession are required to eliminate artifacts for dynamic scenes. Recent approaches [20, 30, 34, 37] apply CNNs or combine some other methods to fuse multiple LDR images. The difference is that we focus on constructing an HDR image from a single over- or under-exposure LDR image.

Single image HDR reconstruction. Single LDR image reconstructed into HDR image is more challenging than multi-image fusion reconstruction because of the lack of color and texture information. Convention approaches estimate the density of illumination and saturation to improve the dynamic range [1, 2, 3, 4, 5]. However, with the advances of deep learning [10, 21, 25, 29, 42], some methods have been developed to learn a mapping function from LDR input to HDR output. For instance, the HDRCNN method focuses on recovering missing details by falling back upon encoder-decoder architecture in the over-exposure or under-exposure regions [15]. Marnerides et al. [26] build a multi-branch network to obtain the high and the low frequency information of the image by dilated convolution. In addition, a fixed inverse camera response function (CRF) is applied to reconstruct missing information that was lost from the original signal. However, the fixed CRF may not be applicable to images captured from multi-exposure images.

Some recent methods [25, 42] are constrained by strict prior knowledge and physical rules. Due to a large number of rule constraints, these models are divided into multiple processes to enhance a single LDR image. For example, LRCP [25] employs dequantization, nonlinear mapping and clipping the dynamic range to contruct the final result. However, these methods consume a large amount of computational resources. In contrast, our method directly reconstructs an HDR image by learning color and texture information to enhance the LDR input in an end-to-end manner and saves a lot of computing resources.

UHD image enhancement. Some approaches [17, 18, 33, 35, 38, 40] have been proposed to reconstruct high-resolution images in real-time. Most of the existing methods achieve real-time processing by using a few convolutional layers [40]. In particular, bilateral filters/grids have attracted long term attention in acceleration [7, 12, 13], which is an edge-aware manipulation of images in the bilateral space [6, 33, 35]. For example, HDRNet [17] casts the enhancement task in the bilateral space via affine transformation of pixel-level perception. However, there are two drawbacks to this locally-affine model. First, a single network that directly extracts and transforms the input tends to lose the color and texture details due to limited cell storage capacity in a single bilateral grid. Second, directly transforming the bilateral grid to the high-resolution image ignores the noise amplification of the bilateral grid. In contrast, our proposed method recovers colors and edges in a

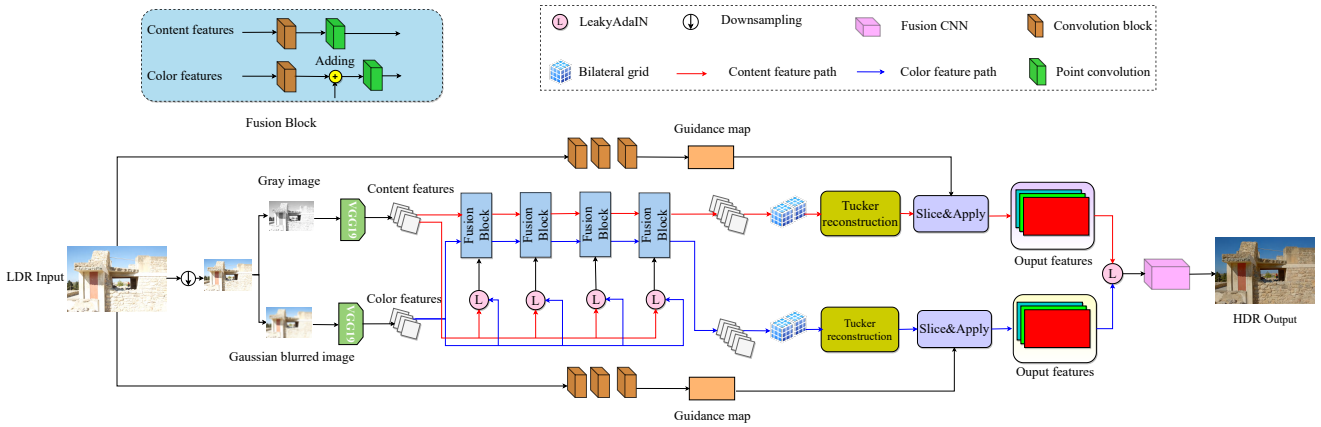


Figure 2. Architecture of the proposed single image LDR-to-HDR network, which consists of three parts. The first step starts with a low-resolution coefficient prediction stream (the dual-path network) that jointly learns the content and chromatic features to fit two affine bilateral grids. Then we use a 3D Tucker reconstruction scheme to remove pseudo edges and noises of the bilateral grid, which is used to restore the high-resolution texture and color features using slicing and application operators. Finally, a feature fusion block combines these two high-quality features to yield an enhanced result. Our proposed algorithm supports UHD images HDR reconstruction at 6 ms on a single Titan RTX GPU shader.

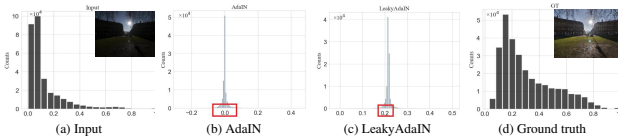


Figure 3. Effectiveness of LeakyAdaIN for fitting the data distribution. (a) Histogram of the LDR input images. (b) The fitted histogram by AdaIN. (c) Our proposed LeakyAdaIN models the data distribution close to the ground truth in (d).

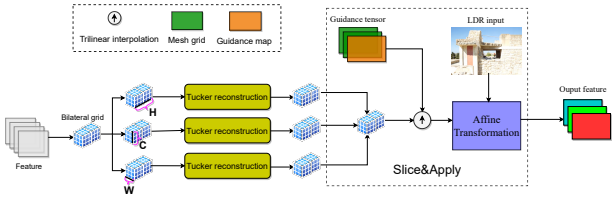


Figure 4. The pipeline of high-quality features reconstruction using 3D Tucker reconstruction.

collaborative dual-path network, while the information of bilateral grid is denoised by Tucker reconstruction.

3. Proposed Method

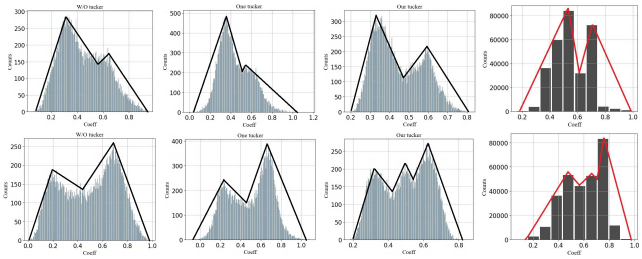
Given a UHD LDR image, our model first reconstructs two bilateral coefficient grids via the proposed dual-path network at a reduced resolution. We propose a new leaky adaptive instance normalization (LeakyAdaIN) to adaptively fuse color and texture features in the dual-path network. Meanwhile, to eliminate pseudo edges and remove noises in the learned bilateral grids, we introduce 3D Tucker decomposition and reconstruction to constrain these affine matrices to vary smoothly. Capitalizing on these two regressed affine bilateral grids, we can generate high-quality edge/texture and color feature maps via the guidance tensor of full-resolution. Figure 2 shows the architecture of the proposed UHD image HDR reconstruction network.

3.1. Collaborative Bilateral Grid Learning

Although two pixels in an LDR input across a weak edge are close in the spatial and frequency domains because of the degradation of contrast and visibility, these two pixels are distant from bilateral filter perspective. Therefore, we consider predicting affine models to restore sharp contents (structures, edges, and textures) and colors in the bilateral space by a dual-path network. Specifically, the upstream path uses the gray image of the LDR input to extracts content information, while the downstream path takes the Gaussian blurred color image as the input to focus on the chromatic information.

Collaborative learning via LeakyAdaIN. Given an LDR input, we first reduce the UHD image to a fixed resolution of 256×256. Then we split the network into a dual-path by feeding the corresponding gray image and a Gaussian blurred color image, and extract content F_c and chromatic features F_c using the pre-trained VGG19 [31], respectively.

During the learning process, we propose a new LeakyAdaIN to help regularize the transforms of each path, so that the network yields a vivid color output while respecting edges. Our LeakyAdaIN is an extension of adaptive instance normalization (AdaIN) [19] that contains a content input x and a style input y and aligns the channel-wise mean



(a) w/o Tucker (b) 1D Tucker (c) Ours (d) GT

Figure 5. 3D Tucker reconstruction allows our model to learn better data distribution in the bilateral space. (a) and (b) show the data distribution of the learned bilateral grid without Tucker reconstruction and with 1D Tucker reconstruction, respectively. (c) is the data distribution learned by our model, which is close to the target distribution of (d).

μ and variance σ of x,

$$\text{AdaIN(x,y)} = \sigma(y)\left(\frac{x-\mu(x)}{\sigma(x)}\right) + \mu(y). \quad (1)$$

Generally, AdaIN is the nearest method of distributing data toward the target, which is cost-effective and often used in style transfer. However, (1) has strong physical constraints and offsets for data distribution fusion of two inputs. In this paper, we consider the content features that lack color information need to replenish the style of color features. Inspired by [19], we propose a LeakyAdaIN to yield fused information with a higher degree of freedom,

$$\text{LeakyAdaIN(x,y)} = s(p(y))\sigma(y)\left(\frac{x-\mu(x)}{\sigma(x)}\right) + s(p(y))\mu(y), \quad (2)$$

where p denotes the global pooling and linear transformation of style data, s is the sigmoid function. To evaluate the effectiveness of the proposed LeakyAdaIN, we compare our LeakyAdaIN with AdaIN, and show the histogram of a learned feature map after the last LeakyAdaIN layer in Figure 3. As shown, using LeakyAdaIN can better model the content distribution (c) than the AdaIN in (b).

Bilateral grid via 3D Tucker reconstruction. We then reshape the content feature F_c and color feature F_g into two bilateral grids, where each coordinate is in three-dimension. The reshaped F_l and F_g can be viewed as a $16 \times 8 \times 8$ bilateral grid, where each grid cell contains 12 numbers. Then, we use the *grid_sample*[1] function to restore the high-resolution content and color feature maps. As shown in Figure 4, two guidance maps with coordinate guidance are constructed with the same dimension of the LDR input. We need the grid cell of the bilateral grid and use trilinear interpolation to fill the guidance tensor (a tensor with the 12 channels). Finally, we employ affine transformation by using 12-channel tensor cutting and three R, G, B channels of the LDR input as a linear transformation, and then generates 3-channel high-quality full-resolution feature maps.

Although the bilateral grid is an effective feature storage container that maintains detailed edges and textures in the LDR input, it is easy to introduce some pseudo edges and noises since the lack of physical constraints on the bilateral grid, especially for UHD images. To address this problem, we enforce a smoothness term in the bilateral grid by introducing Tucker reconstruction [22]. Different from [22], we propose a 3D Tucker reconstruction block to remove artifacts in the learned affine models. As shown in Figure 4, we first make three copies of the learned bilateral grid B. We can regard the bilateral grid as a 3D tensor ($H \times W \times C$) by ignoring the grid cell, and use the Tucker decomposition and reconstruction method to operate each bilateral grid under three perspectives of H, W, and C, respectively,

$$F_z[l,m,n] \leftarrow TR^r_{1\to z}(TR^d_{z\to 1}(B_u[l',m',n'])), \quad (3)$$

where $[l'=16, m'=8, n'=8]$ denote the coordinates of the bilateral grid cell, d and z are tensor decomposition and reconstruction process, respectively. We use the $tucker\_decomposition$[2] approach to compress the dimension $l^{'}$ to 1, and then we use the *tucker_reconstruction*[2] to expand the dimension 1 to $l^{'}$. We set the $tucker\_rank = [l^{'}//2, m^{'}//2, n^{'}//2]$, $init =' random'$, and $tol = 10e{-}5$ for the $tucker\_reconstruction$[2]. This operation first compresses a certain dimension z to a low-rank representation, which indicates that the space has only one layer of data, and finally uses this layer of data to restore the original space dimension z. The reconstructed three grids are averaged to generate the final smoothed bilateral grid. 3D Tucker reconstruction enforces the learned bilateral grid to remove pseudo edges and noises, and acts as a low-rank prior to regularize the data distribution of clear images in the bilateral space. Without 3D Tucker reconstruction, the network erroneously estimate the data distribution as shown in Figure 5.

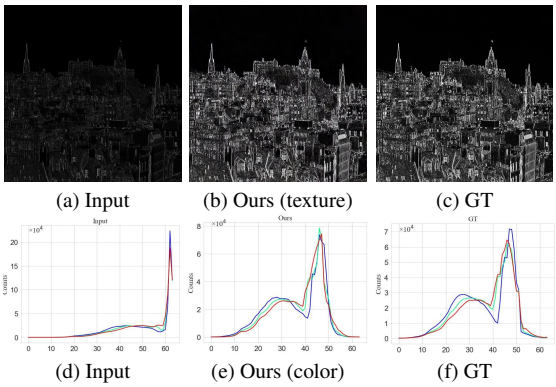
As a result, high-quality content and color features at the UHD resolution can be reconstructed after the slice and application operations. Figure 6 shows a comparison between the LDR input and our reconstructed features. As shown, both the content and color features reconstructed by the collaborative bilateral learning are close to the distribution of the ground truth. Furthermore, grayscale images can express the spatial and structural information of an image, and blurred images can express color characteristics.

3.2. High-Quality Features Fusion

To effectively blend the reconstructed high-quality content and color features, we concatenate three multi-layer convolution blocks with skip connections in each block to

[1] https://pytorch.org/docs/master/nn.functional.html#torch.nn.functional.grid_sample

[2] http://tensorly.org/stable/auto_examples.html



(a) Input (b) Ours (texture) (c) GT

(d) Input (e) Ours (color) (f) GT

Figure 6. Our collaborative bilateral learning method can effectively extract content (edge/texture) details and color distribution. (a) shows the edge information of the LDR input; The recovered high-quality edges in (b) is close to the ground truth in (c). (d) shows the color histogram of the LDR input. The recovered high-quality color histogram in (e) is close to the ground truth in (f).

filter the important features. Each convolution block is brought into inverted residuals [28], which reduces computational effort while maintaining performance.

3.3. Loss Function

We optimize the proposed network by minimizing the L_1 loss and a new self-evolving loss $\mathcal{L}_{se}$ on the training set,

$$\mathcal{L} = \frac{1}{D}\sum_{i=1}^{D} \|I_i - J_i\| + \lambda \mathcal{L}_{se}, \tag{4}$$

where D is the number of training images, I is the result of our network output, and J is the corresponding ground truth of the network output. We introduce an adaptive self-evolving loss, which allows the network to automatically learn an appropriate gamma correction I^g of the network output I during each iteration,

$$\mathcal{L}_{se} = \frac{1}{D}\sum_{i=1}^{D} \left\| I_i - \left[I_i(a,b) \right]^{\gamma[a,b,N(a,b)]} \right\|, \tag{5}$$

where

$$\gamma[a,b,N(a,b)] = 2^{[128-\text{mask}(a,b)]}, \tag{6}$$

where the mask is generated by performing inverse color processing on the original image, and then convolved by a 3×3 Gaussian kernel with stride 2. In addition, a and b denote the pixel of the image network training process. We also tried the adversarial losses, but we notice that the L_1 loss and $\mathcal{L}_{se}$ can obtain vivid colors and clear texture in the enhanced LDR. As shown in Figure 7, we validate the effect of adaptive gamma correction on the training process.

4. Experimental Results

Our proposed model is evaluated on synthetic datasets and real-world images. All the results are compared

(a) I (40 epo.) (b) I^g (40 epo.) (c) I (60 epo.) (d) I^g (60 epo.) (e) I (80 epo.) (f) I^g (80 epo.)

Figure 7. Outputs (I) of the network and the corresponding gamma corrected image (I^g) at different epochs in the training phase. As shown, our proposed adaptive gamma correction darken the output if it is too bright (e.g., 40 and 60 epochs), while tend to brighten the output is it is too dark (e.g., 80 epoch).

(a) Input (PSNR/SSIM) (b) SES (19.26/0.58) (c) DEU (19.54/0.60) (d) PSNR

Figure 8. Comparisons of LDR-to-HDR strategies (SES and DEU). The SES scheme ignores global information, so it is prone to generate artifacts and pseudo edges.

against eight state-of-the-art HDR enhancement methods of LRCP [25], UnModNet [42], 3DLUT [38], ExpandNet [26], HDRCNN [15], FMFPL [29], JSI-GAN [21], and HDRNet [17]. In addition, we performed three ablation studies to demonstrate the effectiveness of each part of our approach.. The implementation code will be made available to the public.

4.1. Training Data

To train and evaluate the proposed network, we use two datasets from [9] and [8] to train the proposed method as well as the comparison methods. The first training data in [9] is multi-exposure fusion data, which is divided into m1-m9 exposure levels, and the normal-exposure can be regarded as the ground truth. We reserve 80 percent of the images for training, and test on the remaining 20 percent of the images. For another dataset [8], we select the image retouched by Expert A as the corresponding enhanced image for the input image, while we use the toolkit (image format converter) to convert the DNP format to the JPG format. We reserve 4500 images for training, and test on the remaining 500 images. For fair comparisons, we have fine-tuned all the compared methods on the same training dataset as our algorithm.

4.2. Implementation Details

The implementation environment is PyTorch 1.7 version and the Adam optimizer is applied to train the model. We use the resolution of 512 $\times$ 512 images with a batch size of 16 to train the network for 3000 epochs in total. Due to the recent HDR enhancement models of HDRCNN [15], ExpandNet [26], FMFPL [29], JSI-GAN [21], LRCP [25], UnModNet [42] cannot directly enhance UHD images on a single Titan TRX GPU. Inspired by [41], we design two strategies for these approaches. First, the downsample-enhancement-upsample (DEU) scheme applies HDR enhancement approaches at a reduced resolution and then upsamples the enhanced images. The other one is splitting-

(a) Input (PSNR/SSIM) (b) LRCP (19.83/0.59) (c) HDRCNN (20.49/0.61) (d) HDRNET (22.17/0.62) (e) LUT3D (21.65/0.73) (f) Our (**24.48/0.0.72**) (g) GT (+∞/1)

Figure 9. Enhanced results on the M-exposure dataset [9]. Our method obtains better visual quality and recovers more image details compared with other state-of-the-art methods.

Table 1. Quantitative evaluations on the FiveK and M-exposure dataset (resolution ranges from 4K to 5K) in terms of PSNR, SSIM, Q-Score [27], and run time, where ⋆ denotes the run time is computed by the DEU scheme.

		HDRCNN [15]	ExpandNet [26]	FMFPL [29]	JSI-GAN [21]	LRCP [25]	UnModNet [42]	HDRNet [17]	3DLUT [38]	Ours
FiveK	PSNR (dB)	17.25	17.80	21.48	23.14	18.94	17.79	19.83	15.75	**23.29**
	SSIM	0.6559	0.6733	0.7581	**0.7711**	0.7195	0.7062	0.7641	0.6152	0.7702
	Q-Score	42.34	44.17	45.88	46.64	43.98	40.15	43.46	39.88	**46.68**
	Time (ms)	18⋆	22⋆	51⋆	55⋆	78⋆	94⋆	5	6	6
M-exposure	PSNR	16.56	16.89	18.51	20.02	18.79	17.97	19.17	14.95	**22.22**
	SSIM	0.5451	0.5981	0.7170	0.7289	0.6993	0.7029	0.7299	0.4731	**0.7593**
	Q-Score	36.29	40.04	43.55	44.78	39.68	41.90	42.44	33.74	**45.76**
	Time (ms)	20⋆	23⋆	78⋆	90⋆	81⋆	95⋆	6	9	8

enhancement-stitching (SES), which splits images into the largest patches that the model can handle, then stitches the enhanced patches to the resolution of the raw image. We compare these two strategies on our UHD test datasets extracting from several public datasets. As shown in Figure 8, SES has pseudo edges since it does not consider the global structure of the image. Figure 8 also demonstrates that DEU obtains higher performance and generates the global structure of the image. So far, we use the DEU strategy for these six deep models [15, 21, 25, 26, 29, 42].

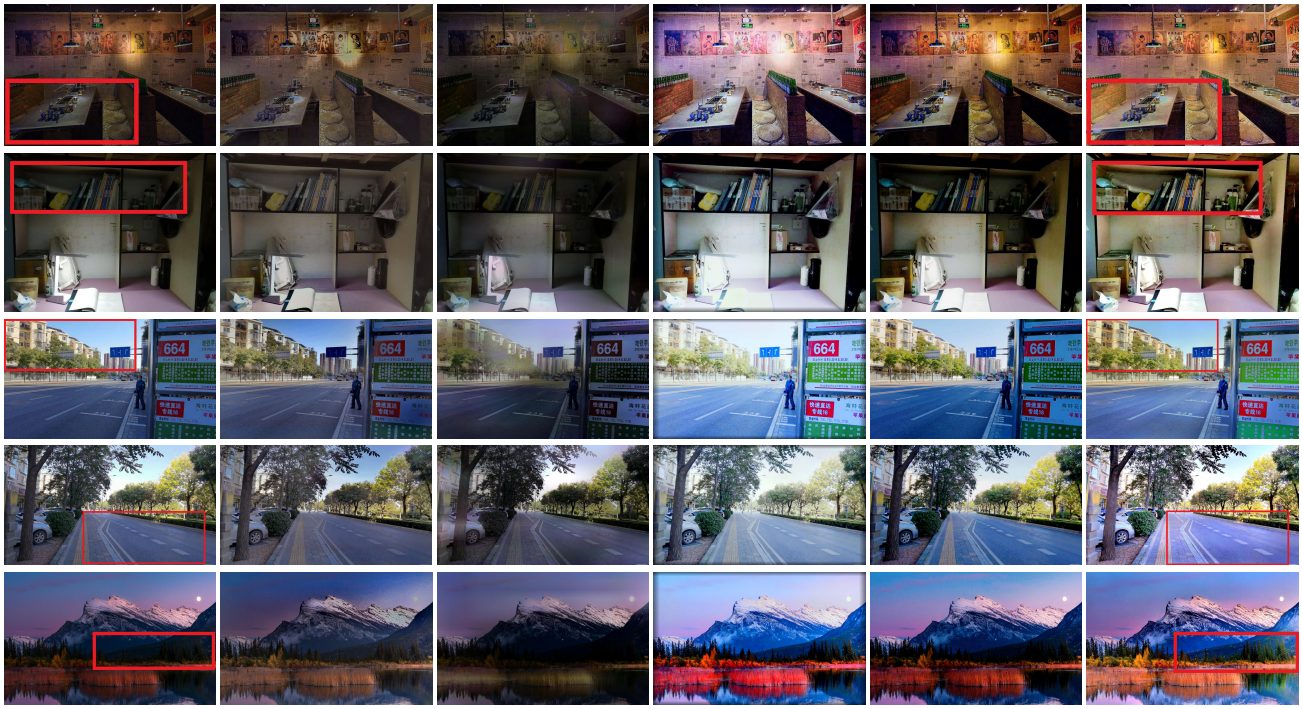
Table 2. Effectiveness of 3D Tucker reconstruction, self-evolving, and the dual-path schemes. Quantitative results demonstrate the effectiveness of each module.

	w/o Tucker	1D Tucker	w/o $\mathcal{L}_{se}$	Single path	Ours
PSNR	20.81	20.35	22.03	19.97	**23.29**
SSIM	0.69	0.68	0.70	0.58	**0.7702**
Q-Score	42.89	43.01	44.79	40.51	**46.68**

4.3. Evaluation

Quantitative evaluation. The proposed method is evaluated on two datasets: test data of M-exposure [9] and FiveK [8]. The comparison results of our proposed method and the other LDR-to-HDR methods are shown in Figure 9. It can be observed that recent deep models [15, 17, 25, 26, 38, 42] still some pseudo edges and noises in the visualized images. However, the enhanced results generated by our algorithm in Figure 9(f) are close to the ground truth images in Figure 9(g). The quantitative results on the M-exposure and FiveK datasets are reported in Table 1. Table 1 demonstrates the effective performance of our method. Note that we fine-tuned the models in the log domain and compute the PSNR, SSIM, and Q-Score in the log domain in the test stage.

Qualitative evaluation. We evaluate the proposed algorithm with different state-of-the-art methods in real-world UHD LDR images. Figure 10 shows the visual comparison of five challenging real-world images. As shown, HDRNet and 3DLUT make the contrast decreased and accompanied by the appearance of artifacts in the enhanced results, LRCP and HDRCNN show color distortion in some regions. In contrast, our algorithm is able to generate vivid colors in Figure 10(f). Because our method considers the global information of the UHD image, the deep model takes into account the contrast and texture information of the whole image.



(a) Input (b) LRCP (c) ExpandNet (d) HDRNet (e) 3DLUT (f) Ours

Figure 10. Enhanced results on real-world LDR images. Our method obtains vivid colors and recovers more image details compared with other state-of-the-art methods. The first two images are 2K resolution, and the last three images are with 4K resolution.

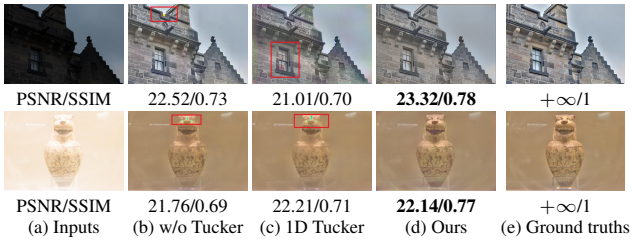
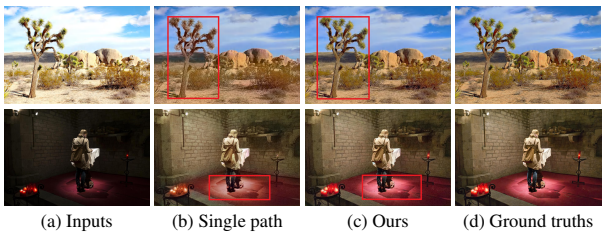


Figure 11. Effectiveness of 3D Tucker reconstruction.



(a) Inputs (b) Single path (c) Ours (d) Ground truths

Figure 12. There is no color information to help restore the image on the dual-path network.

(a) Input (b) Color (c) Ours (d) GT

Figure 13. Effectiveness of gray input in the upper content stream of our proposed network.

4.4. Ablation Study

Effectiveness of 3D Tucker reconstruction. We conduct the following two experiments: 1) removing Tucker reconstruction, and 2) using 1D Tucker reconstruction, on the learned bilateral grids. Table 2 compares our method with these two baselines on the M-exposure [9] dataset. As shown in Figure 11, without Tucker reconstruction or using 1D Tucker reconstruction tends to generate artifacts. Both the quantitative and qualitative results demonstrate that the proposed 3D Tucker reconstruction is able to remove artifacts, pseudo edges, and noises.

Effectiveness of the self-evolving loss. We remove the self-evolving loss function in the training stage and compare it with our proposed algorithm. We note that the training loss with the proposed self-evolving loss function decreases rapidly in the training stage. However, the convergence speed is relatively slow without using the self-evolving loss. Table 2 also demonstrates the effectiveness of the proposed self-evolving loss.

Effectiveness of the collaborative learning. We remove the color stream and the proposed LeakyAdaIN in the network, and directly regress the final output by the content stream. Figure 12 demonstrates that the dual-path network can further enhance the estimated results by ensuring that

(a) Input (b) Ours (c) Ground truths (d) Input (e) Ours (f) Ground truths

Figure 14. Dehazed (first row) and Deblured (second row) results on the public datasets of [24] and [23], respectively.

the fine structural details and vivid colors are captured in the results, such as color details of the green tree and the red carpet shown in Figure 12(c). Table 2 also demonstrates the effectiveness of the proposed collaborative learning in the dual-path network. In addition, we also conduct an experiment by changing the gray input of the content stream into the low-resolution color image. Figure 13 demonstrates that the gray image as input can further enhance texture details of the "art-board".

4.5. Run Time

In terms of run time, the proposed network performs favorably against all the comparison HDR reconstruction approaches [15, 17, 21, 25, 26, 29, 38, 42]. All approaches are executed on the same machine, which was evaluated using an Inter(R) Xeon(R) CPU and an NVIDIA Titan RTX GPU. It should be noted that we only consider the run time of GPU processing.

We counted the average run time for the M-exposure [9] and FiveK [8] datasets are shown in Table 1. These methods are clearly less efficient than our algorithm. The early HDR reconstruction approaches of [15, 26] perform faster than recent approaches of [21, 29], but still have less efficiency than ours. Although 3DLUT and HDRNet have similar run time as ours, these two methods cannot achieve high quantitative results on the Fivek and M-exposure datasets.

4.6. Potentials of Model and Network

Our model can be extended to weather conditions (e.g., hazy days) and motion blurs. We conduct some experiments to show the potential. For image dehazing and deblurring tasks, we randomly select 70% training samples from [24] and [23] to train the proposed model, respectively.

We show some dehazing and deblurring results for hazy and blurry images in Figure 14. The luminance of hazy images tends to have some white drifting due to haze interference, while the luminance range of blurred images is compressed since the pixels are smoothed. Since our model can address the compressed image luminance range problems by the proposed collaborative bilateral learning of texture detail and color range, our algorithm can also remove haze and blur in the degraded inputs. In addition, the construction of the bilateral grid needs to be the edge information of the scene in the image. In the frequency domain analysis, both haze and image blurring are filled with a large number of low-frequency information, and bilateral learning uses filters to extract high-frequency details of the original image. Then, the edge details (high frequency information) of LDR images are used to learn a better bilateral grid acting on the low-frequency regions of the original image.

5. Conclusion

In this paper, we proposed a UHD image HDR reconstruction method via a collaborative learning manner. Our algorithm collaboratively learns the content and color details in the dual-path network capitalized on the proposed LeakyAdaIN layer, and builds corresponding bilateral grids in each patch to maintain detailed content and color in the LDR input. At the same time, we enforce a smoothness term on the learned bilateral grids by 3D Tucker reconstruction to prevent pseudo edges and noise amplification. Quantitative and qualitative results show that our proposed algorithm, compared to other state-of-the-art models, can reach 166fps for a single UHD image and generate satisfactory visual results on real-world UHD images.

Acknowledgment. This work is supported by the National Key R&D Program of China under Grant 2020AAA0109304, National Natural Science Foundation of China (No. 61802403, 62176123, 62025604, U1736219, U1936208), the Natural Science Foundation of Jiangsu Province (BK20191287), the Fundamental Research Funds for the Central Universities (30920021131). Beijing Nova Program (No. Z201100006820074), and Elite Scientist Sponsorship Program by the Beijing Association for Science and Technology.

References

[1] Ahmet Oguz Akyüz, Roland W. Fleming, Bernhard E. Riecke, Erik Reinhard, and Heinrich H. Bülthoff. Do HDR displays support LDR content?: a psychophysical evaluation. *ACM Transactions on Graphics*, 26(3):38, 2007. 2

[2] Francesco Banterle, Kurt Debattista, Alessandro Artusi, Sumanta N. Pattanaik, Karol Myszkowski, Patrick Ledda, and Alan Chalmers. High dynamic range imaging and low dynamic range expansion for generating HDR content. *Computer Graphics Forum*, 28(8):2343–2367, 2009. 2

[3] Francesco Banterle, Patrick Ledda, Kurt Debattista, and Alan Chalmers. Inverse tone mapping. In *CGIT*, 2006. 2

[4] Francesco Banterle, Patrick Ledda, Kurt Debattista, and Alan Chalmers. Expanding low dynamic range videos for high dynamic range applications. In *CG*, 2008. 2

[5] Francesco Banterle, Patrick Ledda, Kurt Debattista, Alan Chalmers, and Marina Bloj. A framework for inverse tone mapping. *The Visual Computer*, 23(7):467–478, 2007. 2

[6] Jonathan T. Barron, Andrew Adams, YiChang Shih, and Carlos Hernández. Fast bilateral-space stereo for synthetic defocus. In *CVPR*, 2015. 2

[7] Jonathan T. Barron and Ben Poole. The fast bilateral solver. In *ECCV*, 2016. 2

[8] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input / output image pairs. In *CVPR*, 2011. 1, 5, 6, 8

[9] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4):2049–2062, 2018. 1, 5, 6, 7, 8

[10] 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. *CoRR*, abs/2103.14943, 2021. 1, 2

[11] Jiawen Chen, Andrew Adams, Neal Wadhwa, and Samuel W. Hasinoff. Bilateral guided upsampling. *ACM Transactions on Graphics*, 35(6):203:1–203:8, 2016. 2

[12] Jiawen Chen, Sylvain Paris, and Frédo Durand. Real-time edge-aware image processing with the bilateral grid. *ACM Transactions on Graphics*, 26(3):103, 2007. 2

[13] Qifeng Chen, Jia Xu, and Vladlen Koltun. Fast image processing with fully-convolutional networks. In *ICCV*, 2017. 2

[14] Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH*, 2008. 1, 2

[15] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafal K. Mantiuk, and Jonas Unger. HDR image reconstruction from a single exposure using deep cnns. *ACM Transactions on Graphics*, 36(6):178:1–178:15, 2017. 1, 2, 5, 6, 8

[16] Yuki Endo, Yoshihiro Kanamori, and Jun Mitani. Deep reverse tone mapping. *ACM Transactions on Graphics*, 36(6):177:1–177:10, 2017. 1

[17] Michaël Gharbi, Jiawen Chen, Jonathan T. Barron, Samuel W. Hasinoff, and Frédo Durand. Deep bilateral learning for real-time image enhancement. *ACM Transactions on Graphics*, 36(4):118:1–118:12, 2017. 2, 5, 6, 8

[18] Bin He, Ce Wang, Boxin Shi, and Ling-Yu Duan. Fhde^{2}net: Full high definition demoireing network. In *ECCV*, 2020. 2

[19] Xun Huang and Serge J. Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *ICCV*, 2017. 3, 4

[20] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep high dynamic range imaging of dynamic scenes. *ACM Transactions on Graphics*, 36(4):144:1–144:12, 2017. 2

[21] Soo Ye Kim, Jihyong Oh, and Munchurl Kim. JSI-GAN: gan-based joint super-resolution and inverse tone-mapping with pixel-wise task-specific filters for UHD HDR video. In *AAAI*, 2020. 1, 2, 5, 6, 8

[22] Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009. 4

[23] Wei-Sheng Lai, Jia-Bin Huang, Zhe Hu, Narendra Ahuja, and Ming-Hsuan Yang. A comparative study for single image blind deblurring. In *CVPR*, 2016. 8

[24] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2019. 8

[25] Yu-Lun Liu, Wei-Sheng Lai, Yu-Sheng Chen, Yi-Lung Kao, Ming-Hsuan Yang, Yung-Yu Chuang, and Jia-Bin Huang. Single-image HDR reconstruction by learning to reverse the camera pipeline. In *CVPR*, 2020. 1, 2, 5, 6, 8

[26] Demetris Marnerides, Thomas Bashford-Rogers, Jonathan Hatchett, and Kurt Debattista. Expandnet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content. *Computer Graphics Forum*, 37(2):37–49, 2018. 1, 2, 5, 6, 8

[27] Manish Narwaria, Rafal K. Mantiuk, Matthieu Perreira Da Silva, and Patrick Le Callet. HDR-VDP-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images. *Journal of Electronic Imaging*, 24(1):010501, 2015. 6

[28] Mark Sandler, Andrew G. Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *CVPR*, 2018. 5

[29] Marcel Santana Santos, Tsang Ing Ren, and Nima Khademi Kalantari. Single image HDR reconstruction using a CNN with masked features and perceptual loss. *ACM Transactions on Graphics*, 39(4):80, 2020. 1, 2, 5, 6, 8

[30] Masahito Shimamoto, Yusuke Kameda, and Takayuki Hamamoto. HDR imaging based on image interpolation and motion blur suppression in multiple-exposure-time image sensor. *IEICE Transactions on Information and Systems*, 103-D(10):2067–2071, 2020. 2

[31] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *ICLR*, 2015. 3

[32] Attila Telegdy. Extending the dynamic range of rf/microwave intermodulation measurements by multiple-carrier cancellation. In *IMRC*, 2000. 2

[33] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *CVPR*, 2019. 2

[34] Shangzhe Wu, Jiarui Xu, Yu-Wing Tai, and Chi-Keung Tang. Deep high dynamic range imaging with large foreground motions. In *ECCV*, 2018. 2

[35] Xide Xia, Meng Zhang, Tianfan Xue, Zheng Sun, Hui Fang, Brian Kulis, and Jiawen Chen. Joint bilateral learning for real-time universal photorealistic style transfer. In *ECCV*, 2020. 2

[36] Xin Yang, Ke Xu, Yibing Song, Qiang Zhang, Xiaopeng Wei, and Rynson W. H. Lau. Image correction via deep reciprocating HDR transformation. In *CVPR*, 2018. 1

[37] Guanghui Yue, Weiqing Yan, and Tianwei Zhou. Reference-less quality evaluation of tone-mapped HDR and multiexposure fused images. *IEEE Transactions on Industrial Informatics*, 16(3):1764–1775, 2020. 2

[38] Hui Zeng, Jianrui Cai, Lida Li, Zisheng Cao, and Lei Zhang. Learning image-adaptive 3d lookup tables for high performance photo enhancement in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, abs/2009.14468, 2020. 2, 5, 6, 8

[39] Jinsong Zhang and Jean-François Lalonde. Learning high dynamic range from outdoor panoramas. In *ICCV*, 2017. 1

[40] Jing Zhang and Dacheng Tao. Famed-net: A fast and accurate multi-scale end-to-end dehazing network. *IEEE Transactions on Image Processing*, 29:72–84, 2020. 2

[41] Zhuoran Zheng, Wenqi Ren, Xiaochun Cao, Xiaobin Hu, Tao Wang, Fenglong Song, and Xiuyi Jia. Ultra-high-definition image dehazing via multi-guided bilateral learning. In *CVPR*, 2021. 1, 5

[42] Chu Zhou, Hang Zhao, Jin Han, Chang Xu, Chao Xu, Tiejun Huang, and Boxin Shi. Unmodnet: Learning to unwrap a modulo image for high dynamic range imaging. In *NeurIPS*, 2020. 1, 2, 5, 6, 8