# DomainPlus: Cross-Transform Domain Learning towards High Dynamic Range Imaging

Bolun Zheng*
Hangzhou Dianzi University
HangZhou, Zhejiang, CHINA
blzheng@hdu.edu.cn

Xiaokai Pan*
Hangzhou Dianzi University
HangZhou, Zhejiang, CHINA
panxiaokai@hdu.edu.cn

Hua Zhang†
Hangzhou Dianzi University
HangZhou, Zhejiang, CHINA
zhangh@hdu.edu.cn

Xiaofei Zhou†
Hangzhou Dianzi University
HangZhou, Zhejiang, CHINA
xfzhou@hdu.edu.cn

Gregory Slabaugh
Queen Mary University of London
London, UK
g.slabaugh@qmul.ac.uk

Chenggang Yan
Hangzhou Dianzi University
HangZhou, Zhejiang, CHINA
cgyan@hdu.edu.cn

Shanxin Yuan
Huawei Noah's Ark Lab
London, UK
shanxin.yuan@huawei.com

## ABSTRACT

High dynamic range (HDR) imaging by combining multiple low dynamic range (LDR) images of different exposures provides a promising way to produce high quality photographs. However, the misalignment between the input images leads to ghosting artifacts in the reconstructed HDR image. In this paper, we propose a cross-transform domain neural network for efficient HDR imaging. Our approach consists of two modules: a merging module and a restoration module. For the merging module, we propose a Multiscale Attention with Fronted Fusion (MAFF) mechanism to achieve coarse-to-fine spatial fusion. For the restoration module, we propose fronted Discrete Wavelet Transform (DWT) and Discrete Cosine Transform (DCT)-based learnable bandpass filters to formulate a cross-transform domain learning block, dubbed DomainPlus Block (DPB) for effective ghosting removal. Our ablation study and comprehensive experiments show that DomainPlus outperforms the existing state-of-the-art on several datasets.

## CCS CONCEPTS

• **Computing methodologies** → **Image processing**.

## KEYWORDS

dynamic HDR, fronted fusion, learnable bandpass filter, discrete wavelet transform, and domain plus

*Bolun Zheng and Xiaokai Pan contributed equally to this research.
†Corresponding authors: Hua Zhang (zhangh@hdu.edu.cn) and Xiaofei Zhou (zxforchid@outlook.com).


*MM '22, October 10–14, 2022, Lisboa, Portugal*

ACM ISBN 978-1-4503-9203-7/22/10…$15.00
https://doi.org/10.1145/3503161.3547823

**ACM Reference Format:**
Bolun Zheng, Xiaokai Pan, Hua Zhang, Xiaofei Zhou, Gregory Slabaugh, Chenggang Yan, and Shanxin Yuan. 2022. DomainPlus: Cross-Transform Domain Learning towards High Dynamic Range Imaging. In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22), October 10–14, 2022, Lisboa, Portugal.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3503161.3547823

## 1 INTRODUCTION

Reconstructing high dynamic range (HDR) scenes using low dynamic range sensors is a fundamental problem in signal processing and computer vision. A general approach is to capture multiple low dynamic range (LDR) images with different exposure times and then merge them into a composite HDR image[9]. Previously, several multi-exposure based methods[5, 16, 42] were used to reconstruct HDR images. These methods usually assign one of the LDR images as the reference, and compensate it with the information from the other LDR images. However these methods require the camera and the scene to be totally static, otherwise ghosting artifacts emerge. This requirement is strong, since many scenes are dynamic, and in handheld photography the camera moves between exposures. When there is motion between the LDR images, alignment becomes a major challenge for accurate merging of multiple exposure images. Subsequent studies pay more attention to fusion-based methods [24, 30] and motion compensation [2, 43]. Fusion-based methods[24, 30] adopt a rejection strategy to reject the misaligned patches, and reconstruct the HDR image with the remaining candidate patches and intensity-rendered patches. Because the rejection is only determined by a local patch, and ignores the consistency of neighbor patches, errors occur around over- and under-exposed areas of the reference image. Motion compensation-based methods perform a pixel-wise[2] or patch-wise[43] alignment between the reference image and the other exposure images before merging. However, the alignment suffers from luminance gaps and the different types of noise resulting from different exposures. Further, the large image resolution makes pixel-wise alignment very expensive computationally.

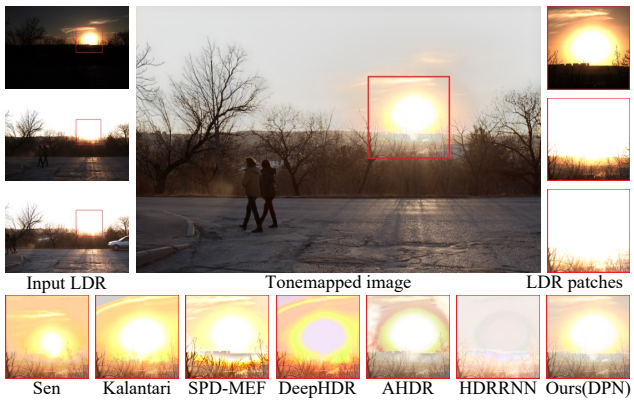


**Figure 1: Compared with other methods, our proposed DPN method generates results that better match the ground truth tone-mapped image particularly in challenging dynamic range regions.**

Recently, several deep neural networks (DNNs) have been proposed for HDR imaging. Kalantari *et al.* [17] propose a convolutional neural network (CNN) to adaptively estimate the aligned pixels and the merging of the different exposure images. However the CNN has difficulty in handling complex motions and suffers from inaccuracy in optical flow estimation and merging. Lee *et al.*[23] use CNNs to perform pre-registration of the input images, which alleviates motion compensation errors. However, this pre-registration requires additional training datasets and is inefficient to run.

Generally, existing work concentrates on the alignment-merge-enhance formulation and tries to convert the dynamic scenes HDR imaging into static scenes HDR imaging. In this work, we focus on the consistencies of local structural representation along the merge-restore direction. We propose a novel cross-transform domain learning block named *DomainPlus Block* (DPB), and construct a multi-scale architecture dubbed *DomainPlus Network* (DPN) for efficient HDR imaging of dynamic scenes. Our DPN is a multi-scale CNN consisting of two parts: a merging module, and a restoration module. The DPB is primarily constructed by two transforms, discrete wavelet transform (DWT) and discrete cosine transform (DCT), where the DWT is first used to decompose the features into different frequency components, then the DCT-based learnable bandpass filters (LBFs)[60] are introduced to generate consistent local features with the decomposed components. Moreover, the multi-scale fused attention, the multi-scale reconstruction and the structural loss function are also proposed to construct the DPN.

## 2 RELATED WORK

In this section, we briefly review existing approaches for dynamic scene HDR imaging and also frequency domain learning methods related to this study.

### 2.1 Dynamic scene HDR imaging

Early HDR imaging methods that address camera and object motion can be grouped into three categories: region alignment-based methods, rejection-based methods, and deep learning-based methods.

**Region alignment-based methods.** These methods usually first align the non-reference LDR images to the reference image using optical flow[8, 18, 62] before merging them. Bogoni *et al.*[2] estimated the optical flow field between each non-reference image and the reference image to register local motion. Hu *et al.*[15] estimated the motion of forward warping, and filled the holes resulting from motion estimation errors and incomplete matching. Qin *et al.*[40] carried out more flexible motion estimation, where patches can be rotated or scaled for better matching. An intensity mapping function (IMF) [12] is also used to reduce ghosting artifacts. Li *et al.*[27] adopt bidirectional normalization to detect pixels with inconsistent motion, which is addressed through the IMF. Tomaszewska and Mantiuk[44] use the scale-invariant feature transform (SIFT) to search for keypoints in consecutive images and then eliminate the misalignment between a sequence of LDR images of different exposures. Hu *et al.*[15] and Sen *et al.*[43] propose a patch-based joint optimization for both alignment and reconstruction.

**Rejection based methods.** These methods[6, 7, 10, 11, 32, 37, 57] attempt to detect and reject the moving pixels, as well as over-exposed and under-exposed pixels. The median threshold bitmap[48] for image alignment is used to detect motion and select the best exposure for fusion. Khan *et al.*[20] iteratively compute the fusion weights, then apply them to pixels to determine their contribution to ghost-free images. Li *et al.*[25] introduce a median filter to remove moving objects. Ma *et al.*[30] propose a structural block decomposition method which is combined with IMF to reject inconsistent motion. Heo *et al.*[14] use joint probability density to first roughly detect moving regions, and then employ graph cut-based energy minimization to refine these regions. Lee *et al.*[22] use rank minimization to detect motion. Yan *et al.*[52, 53] handle object motion via a sparse representation.

**Deep learning based methods.** Recently CNNs have significantly improved the performance of HDR imaging [3, 4, 26, 36, 49, 50, 54]. Wu *et al.*[49] and Yan *et al.*[54] employ deep auto-encoder networks to translate multiple LDR images into a ghosting-free HDR image. Choi *et al.*[3] reconstruct HDR videos using interlaced samples with joint sparse coding. Kathirvel *et al.*[19, 36] propose a scalable CNN architecture to efficiently handle the arbitrary LDR inputs. Yang *et al.*[55] propose an end-to-end deep reciprocating HDR transformation model to reconstruct the lost detail in the HDR domain and then transfer the detail to the LDR image for detail correction. Zhang *et al.*[56] use a depth self coding architecture to regress linear and high dynamic range panoramic images from nonlinear, saturated and low dynamic range panoramic images. [1, 34] propose a fast method for selecting pixels from LDR images based on the scene and pixel classification. Ram *et al.*[41] propose a deep learning structure of static multi-exposure images to learn the fusion operation without reference ground truth image for generating artifact-free results. Wang *et al.*[46] use a GAN to create images and video from the adjustable part of a data stream based on an event camera. Xiong *et al.*[50] propose a hierarchical fusion mechanism for practical ghost-free HDR imaging with a CNN.

### 2.2 Frequency domain learning

CNN-based frequency domain learning has stimulated the interest of researchers in recent years. The discrete wavelet transform

(DWT) and discrete cosine transform (DCT) are widely used to construct frequency domain learning architectures. Liu *et al.*[29] improve a U-Net model by introducing the DWT to reduce the size of feature mapping and inverse wavelet transform (IWT) for upsampling. Liu *et al.*[28] propose a DWT based architecture to remove moire patterns in wavelet subbands. Guo *et al.*[13] and Zheng *et al.*[59] introduce DCT domain learning to make full use of the prior knowledge of JPEG compression. Zhao *et al.* propose a multi-scale dual-domain learning architecture with adaptive quantization table estimation for constant bit rate video quality enhancement. Zheng *et al.*[58, 60, 61] propose a block-IDCT based learnable bandpass filter (LBF) for frequency domain learning.

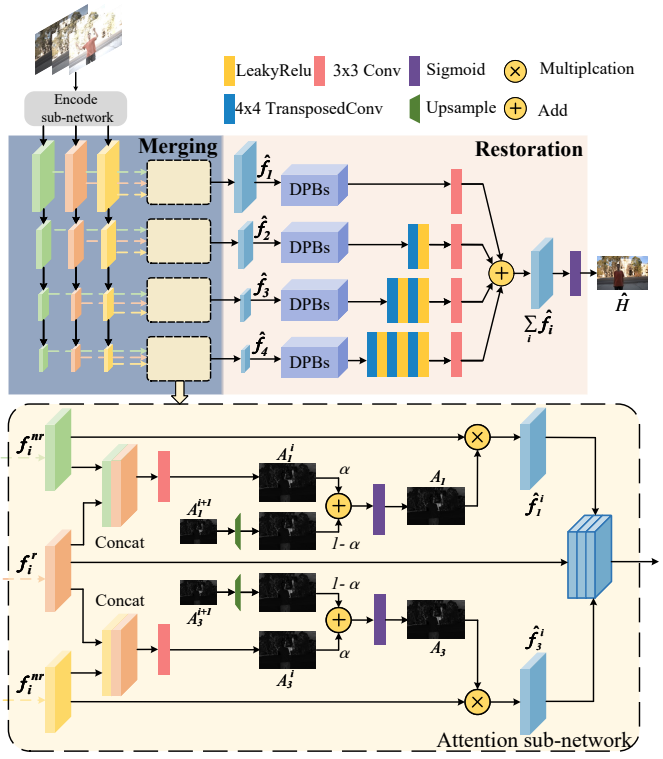


**Figure 2: The architecture of the proposed DPN. The network consists of an *attention module* and *merging module*.**

## 3 PROPOSED METHOD

Given a set of LDR images $\{L_1, L_2, \ldots, L_k\}$ representing different exposures of a dynamic scene, a directly reconstructed high dynamic range image $H$ will suffer from ghosting artifacts when the motion is not compensated. We use a residual model to describe such an image contaminated by ghosting artifacts as:

$$H_{merge} = H_{clean} + N_{ghost} \tag{1}$$

where $H_{clean}$ denotes the HDR image, $H_{merge}$ denotes the image produced by directly merging $\{L_1, L_2, \ldots, L_k\}$ using static scene based methods, and $N_{ghost}$ denotes the ghosting artifacts.

### 3.1 Motivation

Extracting the frequency-domain representation of locally inconsistent patches and performing a correction is an effective way to produce locally consistent results[59, 60]. The recently proposed learnable bandpass filter (LBF)[60, 61] has been shown to be an effective way to make such frequency-domain corrections. The LBF targets learning frequency domain priors of the artifact from various contaminated and clean image pairs and performs bandpass filtering for frequency-domain correction. However, ghosting is a complex artifact and exhibits variable frequency characteristics. To address this challenge, we apply a DWT to decompose the input into different frequency components. We then apply an LBF-based frequency-domain correction to reconstruct a locally consistent output for each frequency component. The DWT first decomposes the unpredictable ghosting effects into different components. Then each LBF works on only one frequency component, which increases the accuracy of the learned frequency passband in LBF.

Following the settings in [17, 51], we use three LDR images ($k = 3$) in the sRGB domain as input, and output an sRGB HDR image aligned to the prescribed reference image $L_2$. As suggested in [17], we first map the LDR images to the HDR domain by gamma correction defined as:

$$H_i = L_i^{\gamma} / t_i \tag{2}$$

where $\gamma > 1$ denotes the gamma correction parameter and $t_i$ denotes the exposure time. Then, we concatenate the $H_i$ and $L_i$ along the channel dimension to formulate a 6-channel input sent into the network. In this work, we set $\gamma = 2.2$. Pre-alignment of the LDR images to the reference image is optional in our solution. As mentioned in several previous studies, pre-alignment does help make a performance improvement. Therefore we no longer introduce and validate the improvement of the pre-alignment in this work.

### 3.2 Overview of the DPN Architecture

We follow a merge-and-restore approach to formulate the baseline of our DPN. As shown in Figure 2, our DPN mainly consists of two modules: a merging module and restoration module. The merging module extracts multi-scale feature maps from LDR inputs and merges the feature maps using a multi-scale spatial attention mechanism. The restoration module aims at removing the ghosting effects remaining in the merged feature maps via multiple *Domain-Plus blocks* (DPBs) from different scales, and then reconstructs an HDR output.

The merging module first uses an encode sub-network to extract feature maps through different scales. As shown in Figure 3, the encode sub-network first uses multiple 2× pixel unshuffling-based downsampling to generate multi-scale tensors from the 6-channel LDR input, and sequentially uses an $F$-channel convolutional layer to obtain the base feature map at each scale. Compared to strided convolution, the pixel unshuffling can perfectly preserve the image details and ensure none of information is lost during the downsampling. Because the encode sub-network works as a basic feature extractor, its weights are shared for the three inputs. Then the attention sub-network generates a spatial attention map with each non-reference input and the reference input working progressively from the highest scale (lowest resolution) to the original scale. The

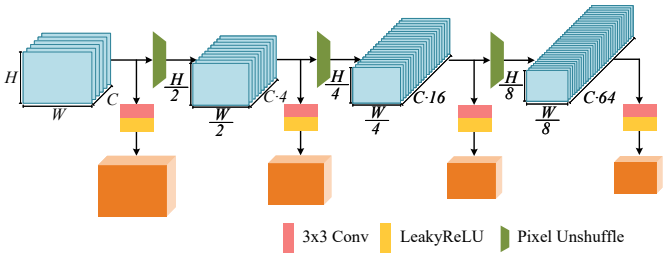


**Figure 3: The structure of encode sub-network.**

two non-reference frames are very different, since one is under-exposed and the other is over-exposed. Therefore, the weights of the two attention sub-networks are not shared with each other.

The restoration module takes the multi-scale feature maps output by the attention module as inputs. The correction sub-network, constructed by stacking DPBs in each scale, is introduced to correct the ghosting artifacts that remain in the feature maps. Then the transposed convolution-based reconstruction sub-network upsamples the corrected feature maps to the original scale and finally outputs a ghost-free HDR image.

## 3.3 Merging with attention

The merging module is developed to combine the multiple LDR inputs based on a spatial attention mechanism to identify the misaligned parts between reference input and non-reference inputs. Classic signal processing methods [30] use color distribution to distinguish the misaligned local patches. CNN-based methods[51] introduce pixel-wise attention to achieve a similar effect. Though these methods may produce a satisfactory result, errors often exist when the over-exposed and under-exposed areas are included in the reference image. The small patch size and poor receptive field are major limitations as there is not enough information to make an accurate estimation.

To overcome this limitation, we propose a coarse-to-fine strategy that estimates multi-scale spatial attention maps progressively from the highest scale to the original scale, resulting in an multi-scale spatial attention. The feature maps from the highest scale get the largest receptive fields, designed to provide a rough but useful spatial attention map at first. Then the feature maps from the lower scales gradually refine the spatial attention map by fusing the spatial attention maps from the upper scale and current estimation. Assuming there are $N$ scales in total, the reference and non-reference feature maps of the $i$-th scale are denoted as $f_i^r$ and $f_i^{nr}$ respectively ($i = 1$ denotes the original scale). The spatial attention map of the $i$-th scale can be obtained as:

$$A_i = \begin{cases} \sigma(C_{att}(< f_i^r, f_i^{nr} >)) & i = N \\ \mathcal{F}(\sigma(C_{att}(< f_i^r, f_i^{nr} >)), U(A_{i+1})) & 1 \le i < N \end{cases} \tag{3}$$

where $<,>$ denotes concatenation, $\sigma$ represents sigmoid activation, $C_{att}$ denotes a $3 \times 3$ convolution outputting a one channel feature map, $U$ denotes the upsampling operation and $\mathcal{F}$ represents the fusion function. Normally, we can use a simple linear function to construct $\mathcal{F}$, which can be expressed as:

$$\mathcal{F}(x_1, x_2) = \alpha \cdot x_1 + (1 - \alpha) \cdot x_2 \tag{4}$$

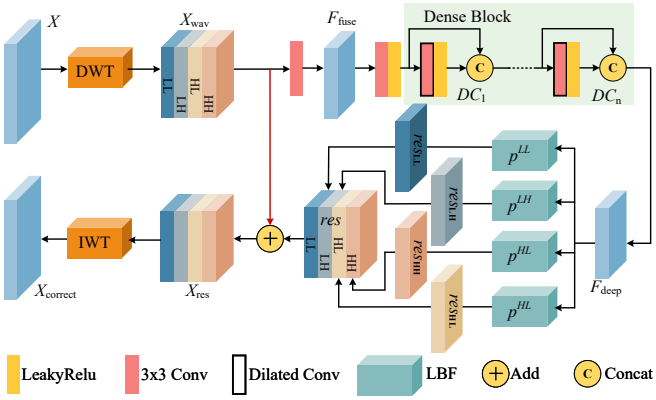


**Figure 4: The struture of DomainPlus Block. The $p^{LL}$, $p^{HL}$, $p^{LH}$ and $p^{HH}$ are four LBFs producing the corresponding residual artifact for $x^{LL}$, $x^{HL}$, $x^{LH}$ and $x^{HH}$ respectively.**

where $\alpha \in [0, 1]$ is a hyper-parameter. However, because the values of each of two inputs lie in the range $[0, 1]$, direct fusion with a linear function will certainly reduce performance of the refinement.

Instead, we introduce a fronted fusion (FF) method to fuse the two spatial attention maps without losing the refinement performance, and propose a multi-scale attention with fronted fusion (MAFF) mechanism for obtaining the coarse-to-fine multi-scale spatial attention map. *Algorithm*1 illustrates the pipeline of the proposed MAFF.

The major difference from the FF is that the spatial attention maps are fused before the sigmoid activation. Specifically, the $f_i^r$ and $f_i^{nr}$ are firstly fused to obtain a temporal attention map $A_i^{temp}$ by:

$$A_i^{temp} = C_{att}(< f_i^r, f_i^{nr} >) \tag{5}$$

Then the exact raw attention map $A_i^{raw}$ to be activated by the sigmoid can be obtained as:

$$A_i^{raw} = \begin{cases} A_i^{temp} & i = N \\ \mathcal{F}(A_i^{temp}, U(A_{i+1}^{raw})) & 1 \le i < N \end{cases} \tag{6}$$

So that the exact spatial attention map $A_i$ is expressed as:

$$A_i = \sigma(A_i^{raw}) \tag{7}$$

Because $A_i$ is supposed to reject the misaligned parts of the non-reference frame, the $A_i$ is applied to the non-reference feature map.

## 3.4 Restoration with DomainPlus Blocks

Human eyes can easily distinguish ghosting artifacts without any training. We incorporate a *learnable bandpass filter* (LBF)[60] to learn subtle frequency domain priors of ghosting artifacts using various pairs of images for training. As introduced in Sec. 3.1, we propose a DWT-based LBF, namely a *DomainPlus* block (DPB). The DPB can be expressed as:

$$DPB(X) = T_{wav}^{-1}(P(T_{wav}(X))) \tag{8}$$

where $T_{wav}$ and $T_{wav}^{-1}$ denotes the DWT and IWT respectively, and $P(\cdot)$ denotes the following LBF-based processing. As suggested by [60], we apply an $8 \times 8$-IDCT-based LBF. Specifically, let the

Algorithm 1: *DomainPlus* Block

Input: feature map $X$
1. Making DWT on $X \rightarrow X_{wav} = T_{wav}(X) = \{x_{LL}, x_{HL}, x_{LH}, x_{HH}\}$
2. Making $P(X_{wav}) \rightarrow X_{correct}$:
2.1 Fusing the $X_{wav} \rightarrow F_{fuse} = C_{fuse}(X_{wav})$
2.2 Producing $F_{deep}$ from $F_{fuse}$ via dense block
2.3 Extracting the residuals for each component of $X_{wav}$:
2.3.1 $res_{LL} = p^{LL}(F_{deep})$
2.3.2 $res_{HL} = p^{HL}(F_{deep})$
2.3.3 $res_{LH} = p^{LH}(F_{deep})$
2.3.4 $res_{HH} = p^{HH}(F_{deep})$
2.4 $res = Concat(res_{LL}, res_{HL}, res_{LH}, res_{HH})$
2.5 Residually connecting to $X_{wav} \rightarrow X_{res} = X_{wav} + res$
3. Making IWT on $X_{res} \rightarrow X_{correct} = T^{-1}_{wav}(X_{res})$
Return: Corrected feature map $X_{correct}$

* $p^{LL}$, $p^{HL}$, $p^{LH}$ and $p^{HH}$ denotes the LBFs for extracting the residuals for the corresponding component of $X_{wav}$.

components obtained after $D_{wav}$ be $X_{wav} = \{x_{LL}, x_{HL}, x_{LH}, x_{HH}\}$. For each $x \in X_{wav}$, a dilated convolution-based dense block first extracts rich features to ensure the size of receptive field is larger than $8 \times 8$.

However this operation introduces too many dense blocks, and therefore requires too much computation and memory space. Noting the redundancy that exists among the four components of $X_{wav}$, we first use a $3 \times 3$ convolution to fuse the $X_{wav}$ into one feature map $F_{fuse}$, then introduce a dense block to extract the deep feature map $F_{deep}$ from the $F_{fuse}$. Then four LBFs are connected in parallel to the dense block, sharing $F_{deep}$ to estimate the residuals for each component of $X_{wav}$. This way, only one dense block is needed for constructing the DPB. Figure 4 presents the structure of the proposed DPB, and $Algorithm1$ illustrates the pipeline of the proposed DPB. In each LBF, a $3 \times 3$ convolutional layer, a IDCT layer with the learnable passband $\theta$ and an another $3 \times 3$ convolutionsl layer are sequentially stacked. This enables the LBF to extract the representation of the ghosting artifacts, which are residually connected to $x$. We denote the operation of the LBF as $p(\cdot)$, which can be expressed as:

$$p(x) = C(T^{-1}_{dct}(C(x) \cdot \theta)) \tag{9}$$

where $C$ denotes the convolution and $T^{-1}_{dct}$ denotes the IDCT.

We construct the multi-scale restoration module with DPBs to produce the final clean HDR image. Using the multi-scale feature maps produced by the merging module as inputs, we first recurrently stack three DPBs at each scale, then use transposed convolution layers to upsample the output of DPBs to the original scale, and finally sum the outputs of all scales and use a $3 \times 3$ convolution layer with sigmoid activation to produce the final output. Specifically, for the situation of multiple DPBs, because $D_{wav}$ and $D^{-1}_{wav}$ are mutual inverse transforms, introducing $D_{wav}$ and $D^{-1}_{wav}$ in each DPB is unnecessary. Therefore, only one $D_{wav}$ is applied at the beginning of the first DPB, and one $D^{-1}_{wav}$ at the end of the last DPB is required for the entire processing at each scale.

## 3.5 Training the network

Similar to previous studies, to produce an HDR domain image, we apply the $\mu$-law function on both the predicted image and ground truth image before they are sent to the loss function. Given an image $H$ as input, the $\mu$-law function can be expressed as:

$$\tau(H) = \frac{\ln(1 + \mu H)}{\ln(1 + \mu)} \tag{10}$$

where $\mu$ is a hyperparameter set to 5000.

We adopt the L1 loss as the basic loss function for pixel-wise supervision. To provide structural information describing the local consistency for improved training of the DPBs, we adopt the dilated advanced Sobel loss (D-ASL) loss function [61] as an additional function, which is defined as:

$$\mathcal{ASL}(\widehat{Z}, Z) = \frac{1}{N} \sum \left| Sobel^{\star}(\widehat{Z}) - Sobel^{\star}(Z) \right| \tag{11}$$

$$\mathcal{D} - \mathcal{ASL}^{\{d_1, d_2, \cdots, d_n\}} = \sum_{i=1}^{n} \mathcal{ASL} \mid_{dilation\_rate=d_i} \tag{12}$$

where $N$ denotes the training batch size, $Z$ and $\widehat{Z}$ denote the groundtruth and the image predicted by network after the $\mu$-law function, $Sobel^{\star}$ denotes the advanced Sobel filtering[60], and $\mathcal{ASL} \mid_{dilation\_rate=d_i}$ denotes the ASL with dilation rate of $d_i$. Following [61], we apply the reference settings that $D = \{1, 2, 3\}$ to formulate the D-ASL. Then the final loss function can be expressed as:

$$Loss(\widehat{Z}, Z) = \mathcal{L}1(\widehat{Z}, Z) + \lambda \cdot \mathcal{D} - \mathcal{ASL}^{D}(\widehat{Z}, Z) \tag{13}$$

where $\lambda = 0.25$ is a hyper-parameter to balance the L1 loss and D-ASL.

# 4 EXPERIMENTS

## 4.1 Implementation details

**Training settings** We use 64 $3 \times 3$ stride one kernels in Conv layers and use zero padding to retain the size of the resulting feature maps. The Adam [21] optimizer is used with initial learning rate set to $10^{-4}$. The LDR images and the corresponding images are cropped into $256 \times 256$ sized patches for training. The batch size is set to 16. Patches are randomly rotated for data augmentation to avoid overfitting during the training stage. During training, we measure the PSNR-$\mu$ of the validation set at each epoch. If the decrease in the validation loss is less than 0.001 dB for four consecutive epochs, the learning rate is halved. The training procedure ends when learning rate is less than $10^{-6}$. We implement our DPN using TensorFlow on an NVIDIA RTX2080Ti GPU.

**Datasets & metrics.** We adopt the Kalantari's[17] dataset[1] as the basic training dataset; all models in Sec. 4.2 are trained on this dataset. Furthermore, we also report experiments with Prabhakar's dataset[2] in Sec.4.3. Sen's [43] dataset[3] and the Tursun's [45] dataset[4], which are widely used in the related studies, are also used to evaluate the performance of compared models. We measure the PSNR and SSIM[47] of the predicted HDR image in both the

[1] https://cseweb.ucsd.edu/ viscomp/projects/SIG17HDR/
[2] https://val.cds.iisc.ac.in/HDR/ICCP19/, MIT License
[3] https://web.ece.ucsb.edu/psen/hdrvideo
[4] https://user.ceng.metu.edu.tr/akyuz/files/eg2016/index.html

| Scales | attention fusion | fronted fusion | PSNR-$\mu$ | PSNR-L | SSIM-$\mu$ | SSIM-L |
|---|---|---|---|---|---|---|
| 4 | | | 44.35 | 41.91 | 0.9960 | 0.9920 |
| 4 | ✓ | | 44.07 | 41.89 | 0.9960 | 0.9919 |
| 4 | ✓ | ✓ | 44.45 | 41.94 | 0.9960 | 0.9919 |
| 1 | | | 43.96 | 41.53 | 0.9958 | 0.9911 |
| 2 | ✓ | ✓ | 43.78 | 41.71 | 0.9959 | 0.9912 |
| 3 | ✓ | ✓ | 44.21 | 41.69 | 0.9960 | 0.9914 |
| 5 | ✓ | ✓ | 44.20 | 41.54 | 0.9960 | 0.9913 |

**Table 1: Performance comparison of the multi-scale, attention fusion and fronted fusion options.**

linear domain (-L) and HDR domain (-$\mu$) for quantitative evaluation. We also include HDR-VDP-2[31] as another metric for comparison.

## 4.2 Ablation study

We investigate the value of key components in the proposed DPN, including the merging and restoring modules, multi-scale architectures, *DomainPlus* block and attention mechanism.

*4.2.1 Merging and restoration modules.* We start by investigating the contributions of the merging and restoration modules presented in Sec. 3.2. We first separately show the contribution of the merging module by removing the residual branches of the DPBs in the restoring module. For easy visualization, the reconstruction subnetwork is preserved. As shown in the second column of Figure 5, though the merging module partially rejects the misaligned parts, the ghosting effect is certainly present. The DPBs in restoration module compensate such ghosting artifact and provide improved detail in the over-exposed regions. This observation demonstrates that the two modules work as desired.

Moreover, because both AHDR[51] and our DPN adopt an attention-based architecture, we compare the attention component of the two methods. In DPN, the attention component is named as the "merging module", while the corresponding component in AHDR is named as an "attention network". Figure 5 visualizes the results of attention components in the two methods. Limited by the single scale attention, the attention network in AHDR fails to distinguish the moving pixels. Thanks to the MAFF, our merging module successfully rejects most of the moving pixels and preserves most of details for the following restoration module.

*4.2.2 Multi-scale architecture.* Next, we investigate the contribution of the multi-scale architecture. We construct five networks constructed with scales of 1 to 5 respectively. For the single scale network, there is no attention fusion and fronted fusion options, and the network degrades to an AHDR-like architecture[51]. As shown in Table 1, additional scales contribute to better performance. However the 5-scale model has a 0.25dB PSNR-$\mu$ reduction compared to the 4-scale model, as the DPBs at the over-coarse scale cannot learn reliable priors of the ghosting artifact.

*4.2.3 Attention mechanism.* The multi-scale attention with fronted fusion (MAFF) mechanism is one of the major contributions of this work. As shown in Table 1, using only multi-scale attention fusion leads to a 0.28dB PSNR-$\mu$ reduction in performance. Due to the limited value range of $[0, 1]$, directly fusing the upsampled attention map from the higher scale to the current scale certainly reduces the accuracy of current scale's attention map. The fronted fusion (FF) mechanism greatly overcomes this limitation to let the information from the higher scale efficiently be passed to the lower scale to help the lower scale produce a more accurate attention map. Introducing the FF mechanism significantly improves the multi-scale attention fusion and ultimately leads a 0.1dB PSNR-$\mu$ performance gain from the no fusion network.

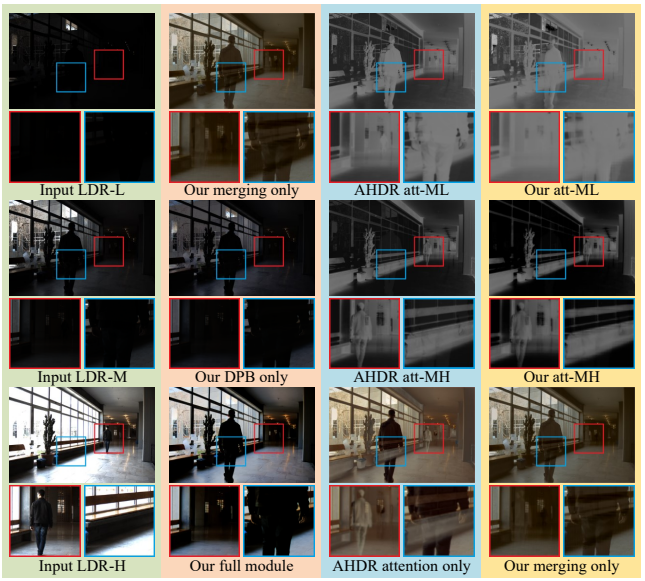


**Figure 5: Visual comparison of the separate effects of the merging module and restoration module of DPN. 'L', 'M' and 'H' refer to low-exposure, medium-exposure and high-exposure respectively. The att-ML and att-MH denote the attention map between the medium-exposure and low-exposure images, and medium-exposure and high-exposure images respectively.**

| DPB | DRDB | DWT | LBF | PSNR-$\mu$ | PSNR-L | SSIM-$\mu$ | SSIM-L |
|---|---|---|---|---|---|---|---|
| ✓ | | ✓ | | 44.35 | 41.72 | 0.9960 | 0.9918 |
| ✓ | | | ✓ | 44.29 | 41.87 | 0.9961 | 0.9918 |
| | ✓ | | | 44.11 | 41.59 | 0.9959 | 0.9914 |
| ✓ | | ✓ | ✓ | 44.45 | 41.94 | 0.9960 | 0.9919 |

**Table 2: Performance comparison for different components and settings of the DPB.**

*4.2.4 DomainPlus block.* We first compare the proposed DPB with the recently proposed dilated residual dense block (DRDB)[51]. Both DRDB and DPB adopt a dilated convolution-based dense block for fronted deep feature extraction. The major difference between them is that DPB introduces cross-transform domain processing in the post-processing. As shown in Table 2, the cross-transform domain processing brings a 0.34dB PSNR-$\mu$ performance gain and clearly removes the ghosting effects as visualized in Figure 10.

Next we investigate the internal contribution of the cross-transform domain process presented in Sec. 3.4. As shown in Table 2, the LBFs bring a performance gain. Including the fronted DWT certainly helps the LBFs achieve better efficiency for learning the frequency prior of ghosting artifacts, and further leads to a 0.16 PNSR-$\mu$ performance gain. We also visualize the intermediate results to demonstrate the contribution of each wavelet component. We choose the LDR images provided in Figure 5 as the input, where the ghosting artifacts exist around the man's legs in the merging only result. Figure 6 shows the results of four different components before IDWT. The ghosting artifacts (red arrows) are mostly removed by the LL component. The LH, HL and HH components further improve the edges and textures details around the ghosting areas. Moreover, we can also observe that the LH, HL and HH components could effectively enhance the details around the over-exposure area (blue arrows).

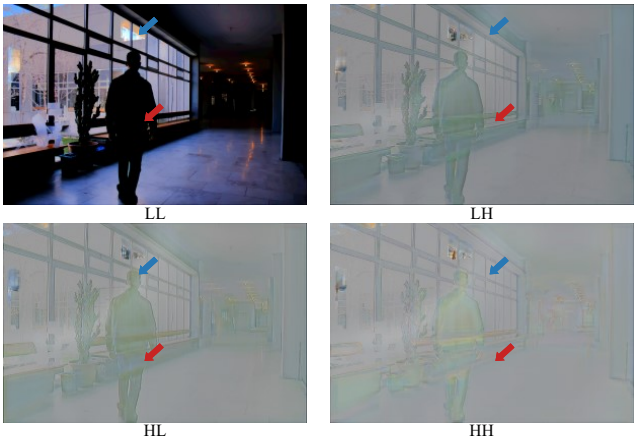


**Figure 6: The contributions of four wavelet components in the *DomainPlus* block.**

## 4.3 Comparison with state-of-the-art methods

We quantitatively and qualitatively compare our DPN method to several state-of-the-art methods including Sen[43], Hu[15], Oh[33], Kalantari[17], SPD-MEF[30], DeppHDR[49], AHDR[51], Robust-HDR[39], NHDRRnet[54], SCHDR[36], Prabhakar[35], HFNet[50] and HDRRNN[19]. Because most of the compared methods are fully supervised, the few-shot learning-based method[38] will not be included in the comparison for fairness.

*4.3.1 Experiments on Kalantari's and Prabhakar's datasets.* We first evaluate all compared methods on Kalantari's dataset. As shown in Table 3, our DPN clearly outperforms all methods on all performance indices. For the HDR-oriented metric HDR-VDP-2, our DPN surpasses the second best method by 0.55. Furthermore, on the more competitive Prabhakar's dataset, our DPN outperforms the second best method by 1.38dB/0.012 in terms of PSNR-$\mu$/SSIM-$\mu$. From the visual results shown in Figure 7, only our DPN properly suppresses the ghosting artifact around the car door, demonstrating the superiority of DPN.


**Figure 7: Visual comparison on Kalantari's dataset.**

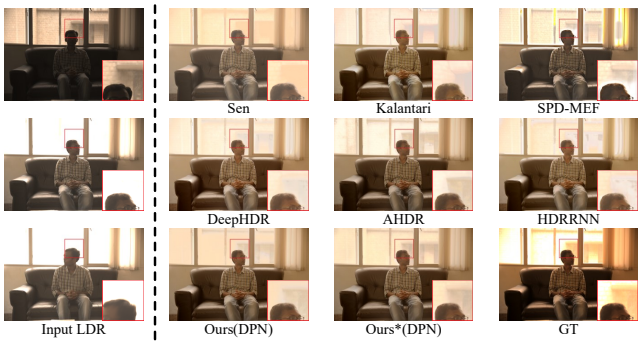


**Figure 8: The visual comparison for cross-dataset evaluation on Prabhakar's dataset. The visual result that we train the model on Prabhakar's[36] is called Our*.**

Existing research only conducts the evaluation individually on these two datasets, it's hard to fairly evaluate the generalization and robustness of the compared methods. Therefore, we conduct a cross-dataset evaluation on Kalantari's and Prabhakar's datasets to demonstrate the generalization and robustness of our DPN. As shown in Table 4, our DPN exhibits better performance than AHDR[51] on both two cross-dataset evaluations using Kalantari's and Prabhakar's datasets. We also provide several visualized comparisons shown in Figure 8. Our DPN trained on Prabhakar's dataset exhibits the best performance among all compared methods. It's worthy to mention that the HDRRNN is also trained on Prabhakar's dataset, but cannot robustly remove the ghosting artifact, unlike our DPN.

*4.3.2 Experiments on additional datasets.* Because neither Sen's and Tursun's datasets provide ground truth, we only conduct qualitative comparisons on these two datasets. As shown in Figure 9 and 10, our DPN exhibits strong performance in over-exposed regions. Benefiting from the MAFF mechanism, our DPN correctly rejects unwanted pixels. The DPB-based restoration module clearly removes the ghosting artifacts and reconstructs natural details.

We also test all compared methods on our self-captured LDR images including both indoor and outdoor HDR scenes with different

| | Sen[43] *TOG'12* | Hu[15] *CVPR'13* | Oh[33] *TPAMI'14* | Kalantari[17] *TOG'17* | SPD-MEF[30] *TIP'17* | DeepHDR [49] *ECCV'18* | AHDR [51] *CVPR'19* | SCHDR[36] *ICCP'19* | RobustHDR[39] *ACCV'20* | NHDRRnet[54] *TIP'20* | Prabhakar[35] *ECCV'20* | HFNet[50] *ACMMM'21* | HDRRNN[19] *TCI'21* | DPN *Ours* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PSNR-$\mu$ | 40.94 | 32.18 | 27.35 | 42.74 | 43.34 | 41.65 | 43.72 | 40.47 | 43.84 | 42.41 | 43.08 | <u>44.28</u> | 42.82 | **44.45** |
| PSNR-L | 38.31 | 31.88 | 26.73 | 41.21 | 40.77 | 40.86 | 40.87 | 39.68 | 41.64 | 41.13 | <u>41.68</u> | 41.48 | 41.68 | **41.94** |
| SSIM-$\mu$ | 0.981 | 0.972 | 0.904 | 0.988 | 0.986 | 0.986 | 0.995 | <u>0.993</u> | 0.991 | 0.988 | - | - | 0.990 | **0.996** |
| SSIM-L | 0.972 | 0.969 | 0.901 | 0.9845 | 0.986 | 0.986 | 0.989 | 0.989 | 0.987 | 0.984 | - | - | <u>0.990</u> | **0.992** |
| HDR-VDP-2 | 55.72 | 55.24 | 46.82 | 60.50 | 61.84 | 61.21 | 62.30 | 61.80 | <u>62.36</u> | 61.21 | 62.21 | 62.33 | - | **62.91** |
| Time(s) | 73.41 | 103.57 | 37.86 | 54.82 | 13.29 | **0.28** | 0.94 | 0.39 | 1.63 | <u>0.31</u> | - | 0.71 | 0.47 | 0.72 |

**Table 3: Comparison results on Kalantari's dataset. The best results are highlighted and the second best results are underlined.**

**Table 4: Cross-dataset evaluation on Kalantari's dataset and Prabhakar's dataset.**

| Module | Training set | Testing set | PSNR-$\mu$ | PSNR-L | SSIM-$\mu$ | SSIM-L |
|---|---|---|---|---|---|---|
| DPN | Kalantari's | Prabhakar's | 33.35 | 31.80 | 0.978 | 0.953 |
| | Prabhakar's | Kalantari's | 41.11 | 37.80 | 0.995 | 0.989 |
| AHDR | Kalantari's | Prabhakar's | 32.67 | 30.65 | 0.978 | 0.947 |
| | Prabhakar's | Kalantari's | 40.87 | 37.11 | 0.994 | 0.987 |

lighting, noise and depth-of-field. These LDR images are captured by Canon 100D camera with the exposure times of $\{\frac{1}{2500}s, \frac{1}{400}s, \frac{1}{100}s\}$, whose corresponding exposure biases are $\{-2.64, 0, 2\}$. As shown in Figure 11, ghosting artifacts and abnormal exposure are still serious challenges for HDR imaging. Compared to other methods, our DPN exhibits the best performance on both ghosting removal and reconstruction of details.

| | Kalantari[17] | DeepHDR [49] | AHDR [51] | SCHDR[36] | Prabhakar[35] | HDRRNN[19] | DPN |
|---|---|---|---|---|---|---|---|
| PSNR-$\mu$ | 35.63 | 38.03 | 38.65 | 36.08 | 38.30 | <u>39.03</u> | **40.41** |
| PSNR-L | 32.50 | 34.40 | 35.28 | 32.74 | 34.98 | <u>36.38</u> | **38.49** |
| SSIM-$\mu$ | 0.961 | 0.971 | 0.973 | 0.959 | 0.970 | <u>0.975</u> | **0.987** |
| SSIM-L | 0.969 | 0.977 | 0.980 | 0.967 | 0.978 | <u>0.983</u> | **0.992** |

**Table 5: Comparison results on Prabhakar's dataset. The best results are highlighted and the second best results are underlined.**

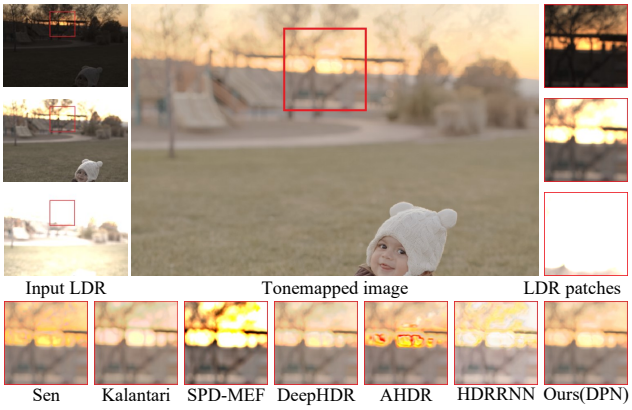


**Figure 9: Visual comparison on Sen's dataset.**


**Figure 10: Visual comparison on Tursun's dataset.**


**Figure 11: Visual comparison on our self-captured image.**

## 5 CONCLUSION

In this paper, we propose a *DomainPlus* network towards efficient HDR imaging in dynamic scenes. Our method clearly surpasses all state-of-the-art methods on all performance indexes. *DomainPlus* provides cross DWT and DCT domain processing to efficiently learn the frequency prior of ghosting artifacts. The multi-scale attention with fronted fusion mechanism is proposed to accurately merge multiple LDR inputs and reject the unwanted regions. Through an ablation study, we demonstrate the importance of the proposed components in the network. Experiments on Kalantari's, Prabhakar's, Sen's and Tursun's datasets and LDR images captured by a Canon camera show that our model outperforms state-of-the-art methods.

## ACKNOWLEDGMENTS

This work is supported by The National Nature Science Foundation of China (62001146, 61901145).

# REFERENCES


[1] Peter van Beek. 2019. Improved image selection for stack-based hdr imaging. *Electronic Imaging* 2019, 4 (2019), 581–1.

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

[3] Inchang Choi, Seung-Hwan Baek, and Min H Kim. 2017. Reconstructing interlaced high-dynamic-range video using joint learning. *IEEE Transactions on Image Processing* 26, 11 (2017), 5353–5366.

[4] Tianhong Dai, Wei Li, Xilei Cao, Jianzhuang Liu, Xu Jia, Ales Leonardis, Youliang Yan, and Shanxin Yuan. 2021. Wavelet-Based Network For High Dynamic Range Imaging. *arXiv preprint arXiv:2108.01434* (2021).

[5] Paul E Debevec and Jitendra Malik. 2008. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 2008 classes*. 1–10.

[6] Ashley Eden, Matthew Uyttendaele, and Richard Szeliski. 2006. 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)*, Vol. 2. IEEE, 2498–2505.

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

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

[9] Miguel Granados, Boris Ajdin, Michael Wand, Christian Theobalt, Hans-Peter Seidel, and Hendrik PA Lensch. 2010. Optimal HDR reconstruction with linear digital cameras. In *2010 IEEE computer society conference on computer vision and pattern recognition*. IEEE, 215–222.

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

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

[12] Michael D Grossberg and Shree K Nayar. 2003. Determining the camera response from images: What is knowable? *IEEE Transactions on pattern analysis and machine intelligence* 25, 11 (2003), 1455–1467.

[13] Jun Guo and Hongyang Chao. 2016. Building dual-domain representations for compression artifacts reduction. In *European Conference on Computer Vision*. Springer, 628–644.

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

[15] Jun Hu, Orazio Gallo, Kari Pulli, and Xiaobai Sun. 2013. HDR deghosting: How to deal with saturation?. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1163–1170.

[16] Katrien Jacobs, Celine Loscos, and Greg Ward. 2008. Automatic high-dynamic range image generation for dynamic scenes. *IEEE Computer Graphics and Applications* 28, 2 (2008), 84–93.

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

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

[19] Ram Prabhakar Kathirvel, Susmit Agrawal, and Venkatesh Babu Radhakrishnan. 2021. Self-Gated Memory Recurrent Network for Efficient Scalable HDR Deghosting. *IEEE Transactions on Computational Imaging* (2021).

[20] Erum Arif Khan, Ahmet Oguz Akyuz, and Erik Reinhard. 2006. Ghost removal in high dynamic range images. In *2006 International Conference on Image Processing*. IEEE, 2005–2008.

[21] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[22] Chul Lee, Yuelong Li, and Vishal Monga. 2014. Ghost-free high dynamic range imaging via rank minimization. *IEEE signal processing letters* 21, 9 (2014), 1045–1049.

[23] Sang-Hoon Lee, Haesoo Chung, and Nam Ik Cho. 2020. Exposure-Structure Blending Network for High Dynamic Range Imaging of Dynamic Scenes. *IEEE Access* 8 (2020), 117428–117438.

[24] Hui Li, Kede Ma, Hongwei Yong, and Lei Zhang. 2020. Fast multi-scale structural patch decomposition for multi-exposure image fusion. *IEEE Transactions on Image Processing* 29 (2020), 5805–5816.

[25] Shutao Li and Xudong Kang. 2012. Fast multi-exposure image fusion with median filter and recursive filter. *IEEE Transactions on Consumer Electronics* 58, 2 (2012), 626–632.

[26] Wei Li, Shuai Xiao, Tianhong Dai, Shanxin Yuan, Tao Wang, Cheng Li, and Fenglong Song. 2022. SJ-HDˆ 2R: Selective Joint High Dynamic Range and Denoising Imaging for Dynamic Scenes. *arXiv preprint arXiv:2206.09611* (2022).

[27] Zhengguo Li, Jinghong Zheng, Zijian Zhu, and Shiqian Wu. 2014. Selectively detail-enhanced fusion of differently exposed images with moving objects. *IEEE Transactions on Image Processing* 23, 10 (2014), 4372–4382.

[28] Lin Liu, Jianzhuang Liu, Shanxin Yuan, Gregory Slabaugh, Aleš Leonardis, Wengang Zhou, and Qi Tian. 2020. Wavelet-based dual-branch network for image demoiréing. In *European Conference on Computer Vision*. Springer, 86–102.

[29] Pengju Liu, Hongzhi Zhang, Kai Zhang, Liang Lin, and Wangmeng Zuo. 2018. Multi-level wavelet-CNN for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 773–782.

[30] Kede Ma, Hui Li, Hongwei Yong, Zhou Wang, Deyu Meng, and Lei Zhang. 2017. Robust multi-exposure image fusion: a structural patch decomposition approach. *IEEE Transactions on Image Processing* 26, 5 (2017), 2519–2532.

[31] Rafał Mantiuk, Kil Joong Kim, Allan G Rempel, and Wolfgang Heidrich. 2011. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Transactions on graphics (TOG)* 30, 4 (2011), 1–14.

[32] Tae-Hong Min, Rae-Hong Park, and SoonKeun Chang. 2009. Histogram based ghost removal in high dynamic range images. In *2009 IEEE International Conference on Multimedia and Expo*. IEEE, 530–533.

[33] Tae-Hyun Oh, Joon-Young Lee, Yu-Wing Tai, and In So Kweon. 2014. Robust high dynamic range imaging by rank minimization. *IEEE transactions on pattern analysis and machine intelligence* 37, 6 (2014), 1219–1232.

[34] Reza Pourreza-Shahri and Nasser Kehtarnavaz. 2015. Exposure bracketing via automatic exposure selection. In *2015 IEEE international conference on image processing (ICIP)*. IEEE, 320–323.

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

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

[37] K Ram Prabhakar and R Venkatesh Babu. 2016. Ghosting-free multi-exposure image fusion in gradient domain. In *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 1766–1770.

[38] K Ram Prabhakar, Gowtham Senthil, Susmit Agrawal, R Venkatesh Babu, and Rama Krishna Sai S Gorthi. 2021. Labeled From Unlabeled: Exploiting Unlabeled Data for Few-Shot Deep HDR Deghosting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4875–4885.

[39] Zhiyuan Pu, Peiyao Guo, M Salman Asif, and Zhan Ma. 2020. Robust High Dynamic Range (HDR) Imaging with Complex Motion and Parallax. In *Proceedings of the Asian Conference on Computer Vision*.

[40] Xiameng Qin, Jianbing Shen, Xiaoyang Mao, Xuelong Li, and Yunde Jia. 2014. Robust match fusion using optimization. *IEEE transactions on cybernetics* 45, 8 (2014), 1549–1560.

[41] K Ram Prabhakar, V Sai Srikar, and R Venkatesh Babu. 2017. Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs. In *Proceedings of the IEEE international conference on computer vision*. 4714–4722.

[42] Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. 2010. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann.

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

[44] Anna Tomaszewska and Radoslaw Mantiuk. 2007. Image registration for multi-exposure high dynamic range image acquisition. (2007).

[45] Okan Tarhan Tursun, Ahmet Oğuz Akyüz, Aykut Erdem, and Erkut Erdem. 2016. An objective deghosting quality metric for HDR images. In *Computer Graphics Forum*, Vol. 35. Wiley Online Library, 139–152.

[46] Lin Wang, Yo-Sung Ho, Kuk-Jin Yoon, et al. 2019. Event-based high dynamic range image and very high frame rate video generation using conditional generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10081–10090.

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

[48] Greg Ward. 2003. Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. *Journal of graphics tools* 8, 2 (2003), 17–30.

[49] Shangzhe Wu, Jiarui Xu, Yu-Wing Tai, and Chi-Keung Tang. 2018. Deep high dynamic range imaging with large foreground motions. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 117–132.

[50] Pengfei Xiong and Yu Chen. 2021. Hierarchical Fusion for Practical Ghost-free High Dynamic Range Imaging. In *Proceedings of the 29th ACM International Conference on Multimedia*. 4025–4033.

[51] Qingsen Yan, Dong Gong, Qinfeng Shi, Anton van den Hengel, Chunhua Shen, Ian Reid, and Yanning Zhang. 2019. Attention-guided network for ghost-free high dynamic range imaging. In *Proceedings of the IEEE/CVF Conference on Computer*

*Vision and Pattern Recognition*. 1751–1760.

[52] Qingsen Yan, Dong Gong, Pingping Zhang, Qinfeng Shi, Jinqiu Sun, Ian Reid, and Yanning Zhang. 2019. Multi-scale dense networks for deep high dynamic range imaging. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 41–50.

[53] Qingsen Yan, Jinqiu Sun, Haisen Li, Yu Zhu, and Yanning Zhang. 2017. High dynamic range imaging by sparse representation. *Neurocomputing* 269 (2017), 160–169.

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

[55] Xin Yang, Ke Xu, Yibing Song, Qiang Zhang, Xiaopeng Wei, and Rynson WH Lau. 2018. Image correction via deep reciprocating HDR transformation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1798–1807.

[56] Jinsong Zhang and Jean-François Lalonde. 2017. Learning high dynamic range from outdoor panoramas. In *Proceedings of the IEEE International Conference on Computer Vision*. 4519–4528.

[57] Hengrun Zhao, Bolun Zheng, Shanxin Yuan, Hua Zhang, Chenggang Yan, Liang Li, and Gregory Slabaugh. 2022. CBREN: Convolutional Neural Networks for Constant Bit Rate Video Quality Enhancement. *IEEE Transactions on Circuits and Systems for Video Technology* 32, 7 (2022), 4138–4149. https://doi.org/10.1109/TCSVT.2021.3123621

[58] Bolun Zheng, Quan Chen, Shanxin Yuan, Xiaofei Zhou, Hua Zhang, Jiyong Zhang, Chenggang Yan, and Gregory Slabaugh. 2022. Constrained Predictive Filters for Single Image Bokeh Rendering. *IEEE Transactions on Computational Imaging* 8 (2022), 346–357. https://doi.org/10.1109/TCI.2022.3171417

[59] Bolun Zheng, Yaowu Chen, Xiang Tian, Fan Zhou, and Xuesong Liu. 2019. Implicit dual-domain convolutional network for robust color image compression artifact reduction. *IEEE Transactions on Circuits and Systems for Video Technology* 30, 11 (2019), 3982–3994.

[60] Bolun Zheng, Shanxin Yuan, Gregory Slabaugh, and Ales Leonardis. 2020. Image demoireing with learnable bandpass filters. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3636–3645.

[61] Bolun Zheng, Shanxin Yuan, Chenggang Yan, Xiang Tian, Jiyong Zhang, Yaoqi Sun, Lin Liu, Aleš Leonardis, and Gregory Slabaugh. 2021. Learning Frequency Domain Priors for Image Demoireing. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021).

[62] Henning Zimmer, Andrés Bruhn, and Joachim Weickert. 2011. Freehand HDR imaging of moving scenes with simultaneous resolution enhancement. In *Computer Graphics Forum*, Vol. 30. Wiley Online Library, 405–414.