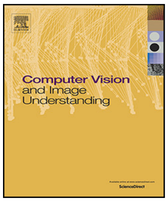
Computer Vision and Image Understanding 238 (2024) 103881

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu



Wavelet-based network for high dynamic range imaging

Tianhong Dai [a,1], Wei Li [c,1], Xilei Cao [d], Jianzhuang Liu [c], Xu Jia [c], Ales Leonardis [c], Youliang Yan [c], Shanxin Yuan [b,*]

[a] *University of Aberdeen, Aberdeen, AB24 3FX, United Kingdom*
[b] *Queen Mary University of London, London, E1 4NS, United Kingdom*
[c] *Huawei Noah's Ark Lab, London, SE10 0ER, United Kingdom*
[d] *Huawei Technologies Co., Ltd., Beijing, 100085, China*

ARTICLE INFO

Communicated by Yung-Yu Chuang

Dataset link: https://github.com/TianhongDai/wavelet-hdr

MSC:
68U10
68T45

Keywords:
High dynamic range image
Ghosting artifacts
Discrete Wavelet Transform
RAW HDR imaging dataset

ABSTRACT

High dynamic range (HDR) imaging from multiple low dynamic range (LDR) images has been suffering from ghosting artifacts caused by scene and objects motion. Existing methods, such as optical flow based and end-to-end deep learning based solutions, are error-prone either in detail restoration or ghosting artifacts removal. Comprehensive empirical evidence shows that ghosting artifacts caused by large foreground motion are mainly low-frequency signals and the details are mainly high-frequency signals. In this work, we propose a novel frequency-guided end-to-end deep neural network (FHDRNet) to conduct HDR fusion in the frequency domain, and Discrete Wavelet Transform (DWT) is used to decompose inputs into different frequency bands. The low-frequency signals are used to avoid specific ghosting artifacts, while the high-frequency signals are used for preserving details. Using a U-Net as the backbone, we propose two novel modules: merging module and frequency-guided upsampling module. The merging module applies the attention mechanism to the low-frequency components to deal with the ghost caused by large foreground motion. The frequency-guided upsampling module reconstructs details from multiple frequency-specific components with rich details. In addition, a new RAW dataset is created for training and evaluating multi-frame HDR imaging algorithms in the RAW domain. Extensive experiments are conducted on public datasets and our RAW dataset, showing that the proposed FHDRNet achieves state-of-the-art performance.

1. Introduction

High dynamic range (HDR) imaging using multiple low dynamic range (LDR) images as inputs is a technique used in computational photography to generate high-quality HDR images. This technique achieves a large range of luminosity by utilizing the information from multiple LDR images. A digital camera usually captures an LDR image with only a limited range of luminosity at a time, where there might appear some over-exposed and/or under-exposed regions, degrading the image quality. Cameras embedded in wearable devices usually have small optical sensors and small apertures, which limit the number of electrons to reach each pixel, making them difficult to capture HDR images at a time. A practical solution for wearable devices is to capture several LDR images with different exposure times and fuse them into a single HDR image. To generate an HDR image, the method should be able to restore the missing information (over-exposed and under-exposed regions) from multiple LDR images, and more importantly, be ghost-free.

Existing methods (Kalantari and Ramamoorthi, 2017; Wu et al., 2018; Yan et al., 2020; Prabhakar et al., 2020; Yan et al., 2019a,b) suffer from different kinds of artifacts, including ghosting, missing details, color degradation, etc. The traditional method by Debevec and Malik (1997) can generate a decent high quality HDR image by merging several static LDR images with different exposures, but it might introduce ghosting artifacts when there is motion. Other early works (Khan et al., 2006; Pece and Kautz, 2010; Li et al., 2014; Bogoni, 2000; Gallo et al., 2015) try to deal with motion through detecting and rejecting moving pixels (Khan et al., 2006; Pece and Kautz, 2010; Li et al., 2014), or through aligning and merging LDR images (Bogoni, 2000; Gallo et al., 2015). They can address a small range of motion but they cannot handle moving content effectively.

Recently, deep learning-based methods (Kalantari and Ramamoorthi, 2017; Wu et al., 2018; Yan et al., 2019a) have been proposed and made great improvements over traditional methods, benefiting from CNN's good representation ability and large amount of training data. These methods either use optical flow to align the inputs, followed by a merging module (Kalantari and Ramamoorthi, 2017), or formulate

[*] Corresponding author.
E-mail address: shanxin.yuan@qmul.ac.uk (S. Yuan).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.cviu.2023.103881
Received 29 August 2022; Received in revised form 28 October 2023; Accepted 3 November 2023
Available online 7 November 2023
1077-3142/© 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

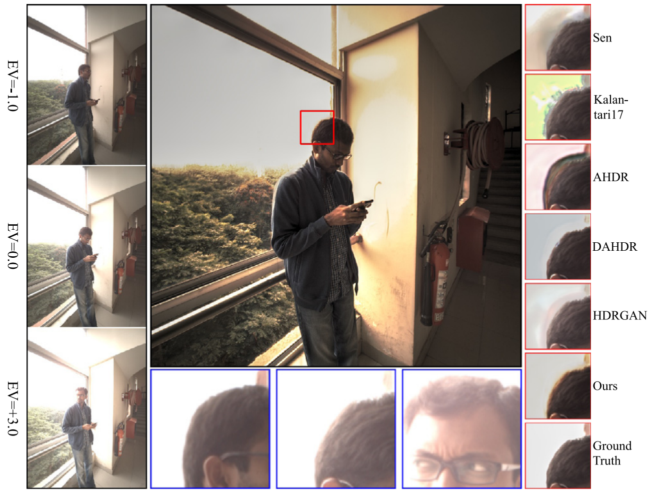
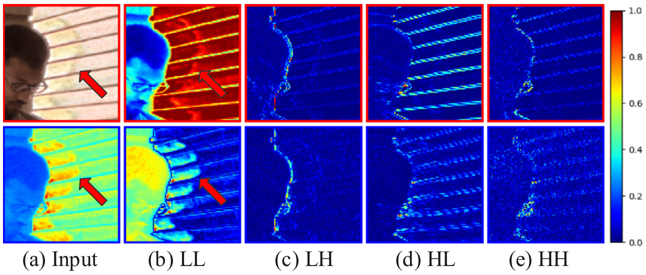


Fig. 1. Comparison between our method and other baselines on the Prabhakar dataset (Prabhakar et al., 2019). Left: Three LDR images with different exposures (low, medium, and high). Center: Our generated HDR image after tone mapping and cropped LDR patches. Right: Results of six methods and the ground truth.



(a) Input (b) LL (c) LH (d) HL (e) HH

Fig. 2. The visualization of frequency sub-bands in the wavelet domain. The first and second row are the results of DWT decomposition of an RGB image and a feature map as the input. (a) Inputs. (b)–(e) Visualization of different frequency sub-bands.

the HDR imaging task as an image-to-image translation problem (Wu et al., 2018; Yan et al., 2019a). Although these methods have made great progress in this area, they still suffer from the ghosting problem (see Fig. 1). We notice that none of the existing methods tries to exploit the fact that the ghosting artifacts caused by large foreground motion are mainly of low-frequency, while the details are of high-frequency. We argue that it is beneficial to separate these low-frequency and high-frequency signals and deal with them separately. Frequency operation has also been used in a few existing HDR imaging methods (Pouli et al., 2014; Hasinoff et al., 2016), *e.g.*, Pouli et al. (2014) decomposes HDR frames into different frequency bands, where the most suitable band is selected adaptively to prevent ghosts, Hasinoff et al. (2016) uses pairwise frequency-domain temporal filter operation for a robust and fast alignment.

In this paper, we choose Discrete Wavelet Transform (DWT) to decompose signals into different frequency bands. Compared with other methods, such as Discrete Fourier Transform (DFT) and Discrete Cosine Transform (DCT), DWT can capture both frequency and spatial information of the images (or feature maps), which helps to preserve detailed texture. In order to verify our hypothesis, we select the output of AHDR (Yan et al., 2019a) in Prabhakar dataset (Prabhakar et al., 2019), where a distinct ghosting artifact is presented because of the object motion, as an example to visualize the decomposed signals in each frequency sub-band. After decomposing, the corresponding frequency sub-bands are given in the first row of Fig. 2. It clearly shows that the ghosting artifacts are mainly in the low-frequency sub-band (LL), while the high-frequency sub-bands (LH, HL, HH) include textures in different directions. In order to extend this verification to the feature space, we also investigate the feature map from the last but one layer of AHDR. In the second row of Fig. 2, it presents a similar trend in the feature space where the ghosting artifacts caused by large foreground motion is mainly in the low-frequency sub-band. Thus, it is highly worth exploring frequency-specific processing in both the RGB and deep feature domains for the HDR imaging task.

In this work, we propose a frequency-guided network (FHDRNet) to explicitly deal with signals of different frequency sub-bands for HDR imaging (see Fig. 3). FHDRNet also performs well in the RAW domain. For RAW domain evaluation, we propose a new RAW dataset.s[2] Processing HDR fusion in the RAW domain has the following advantages, especially for wearable devices: (1) From the Image Signal Processing (ISP) pipeline's perspective, it can bring the HDR fusion module to the early stage (*e.g.*, earlier than demosaicing) of the whole ISP pipeline. It can save computations for other modules (*e.g.*, demosaicing) that otherwise have to be done three times, each for one LDR RAW image; (2) RAW data usually have higher bit width (*e.g.*, 16 bit) and contain more metadata. HDR fusion in the RAW domain will recover more original useful information than that in the RGB domain (8 bit).

The paper's contributions can be summarized as:

- The proposed method — FHDRNet, working in the wavelet domain, is the first to explicitly deal with frequency-specific problems in the HDR imaging task, *e.g.*, ghosting caused by large foreground motion, where the attention mechanism is used on the low-frequency sub-band for fusion to remove such artifacts. The high-frequency sub-bands are used to preserve details (*e.g.*, texture) in the generated HDR image.
- A novel frequency-guided upsampling module is proposed to fuse multiple components with different frequency sub-bands from different images into a single set of low and high-frequency sub-bands to upscale the output using Inverse Discrete Wavelet Transform (IDWT).
- A new dataset is built for training and evaluating HDR algorithms in the RAW domain, which includes 85 and 15 sets of training and testing samples. This is the first RAW dataset for HDR imaging.
- Our method achieves state-of-the-art performance on several public datasets and the new RAW dataset. It also has a good balance between quality and computational efficiency.

2. Related work

In this section, we review the most relevant works, including HDR imaging (Khan et al., 2006; Kalantari and Ramamoorthi, 2017; Wu et al., 2018; Yan et al., 2020, 2019a,b, 2021; Prabhakar et al., 2019; Niu et al., 2021) and learning in the wavelet domain (Li et al., 2020; Abdelhamed et al., 2020; Liu et al., 2018).

2.1. High dynamic range imaging

When the scene and camera are completely static, the traditional method (Debevec and Malik, 1997) can generate a high quality HDR image by merging them together. But it generates ghosting artifacts when there is motion among the LDR images. Early works (Khan et al., 2006; Pece and Kautz, 2010; Li et al., 2014) that try to detect and reject the moving pixels fail to handle moving content effectively. To make use of the moving content, Bogoni (2000) and Gallo et al. (2015) first align the input images and then merge the aligned images into one HDR image. These methods (Bogoni, 2000; Gallo et al., 2015) simply merge the aligned LDR images, and are unable to avoid alignment artifacts.

Recently, many deep learning-based methods (Kalantari and Ramamoorthi, 2017; Wu et al., 2018; Yan et al., 2020) have been developed. Kalantari and Ramamoorthi (2017) propose a deep learning-based model that first aligns the LDR image using optical flow, and

[2] More details about the RAW dataset are included in Section 4.1.

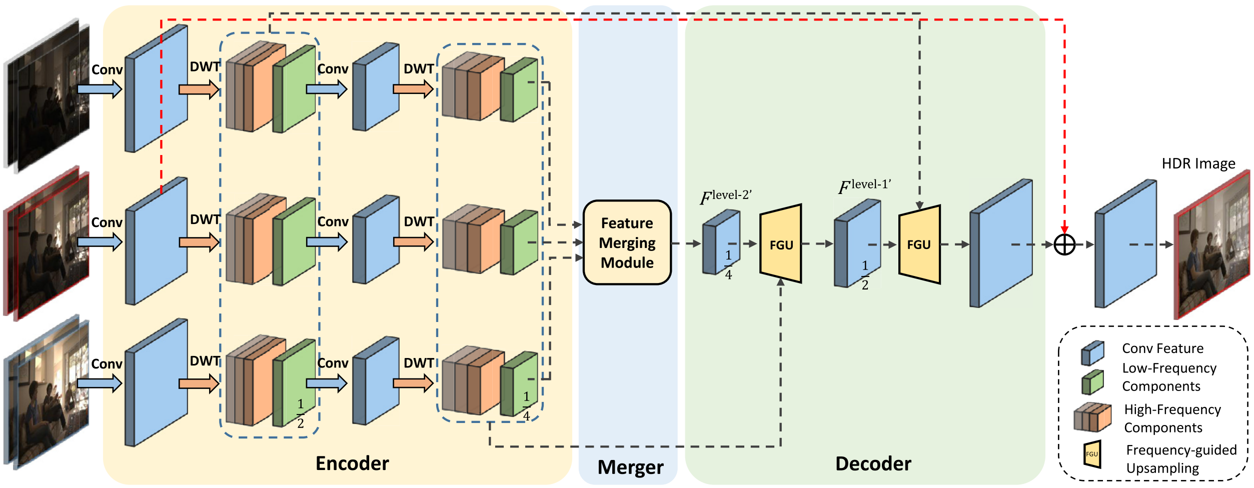


Fig. 3. Overall architecture of the proposed FHDRNet. We consider three pairs of LDR and its HDR as inputs to our FHDRNet and the final reconstructed HDR output is viewed after tone mapping. The proposed network structure contains three parts: an encoder, a merger and a decoder. In the encoder, the input feature maps are decomposed into different frequency sub-bands for further fusion and reconstruction by using DWT. In the merger, the low-frequency sub-bands from the last layer of the encoder are used to generate a single fused feature map. In the decoder, the pre-saved frequency sub-bands are used along with the fused feature map to reconstruct into an upper scale feature map through a frequency-guided upsampling module (FGU). Finally, a global residual connection is used to enhance the feature representation ability of the network. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

then uses a convolutional neural network to generate an HDR image. However, it is difficult to correct the misalignment errors of optical flow, *e.g.*, in the moving area, particularly when there also exists occlusion. Wu et al. (2018) treat the HDR imaging as an image translation problem and use a U-Net to cope with large foreground motion. Though it can reduce the ghosting artifacts, it also blurs image details and hallucinates fine details in the over/under-exposed regions. Yan et al. (2019b) adopt three sub-networks with different scales to reconstruct the HDR image gradually. NHDRRNet (Yan et al., 2020) uses a U-Net to extract features in a low dimension, and then the features are sent into a global non-local network which can fuse the features from inputs according to their correspondence. This method can remove the ghosting artifacts from the final output efficiently. AHDR (Yan et al., 2019a) employs an attention mechanism to solve misalignment and avoids the ghosting artifacts. On the basis of AHDR, DAHDR (Yan et al., 2021) designs a recurrent spatial and channel attention module to improve the performance. HDRGAN (Niu et al., 2021) proposes a novel adversarial training paradigm to restore missing content in the predicted HDR outputs, utilizing an extra reference-based residual merging block to remove artefacts caused by misalignment. HDRGAN also achieves the state-of-the-art results on the public dataset (Kalantari and Ramamoorthi, 2017). SCHDR (Prabhakar et al., 2019) uses a lightweight optical flow PWC-Net (Sun et al., 2018) followed with refinement to align the LDR images first, and then conducts feature aggregation and feature merging to generate an HDR image. These methods fail to explicitly remove the ghosting artifacts and fully exploit the useful information in the inputs.

2.2. Learning in the wavelet domain

Learning in the wavelet domain has the advantage of explicitly dealing with signals in different frequency sub-bands, and it has been applied to some high-level vision and low-level vision problems, such as classification (Li et al., 2020; Williams and Li, 2018; Ji et al., 2012; Cid et al., 2017), style transfer (Yoo et al., 2019), video watermarker (Amini et al., 2018; Huan et al., 2021), image denoising (Abdelhamed et al., 2020; Remenyi et al., 2014; Ho and Hwang, 2012), image demoireing (Liu et al., 2020), image deblurring (Yue et al., 2017), image/video compression (Suzuki, 2019; Haghighat et al., 2019; Mishra et al., 2021), network compression (Gueguen et al., 2018), and super-resolution (Huang et al., 2017; Liu et al., 2018), etc. One of the classical image denoising approach is through image shrinkage (Donoho, 1995), where the noisy image is decomposed into low and high-frequency components and then thresholding is applied to the high-frequency coefficients to remove high-frequency noise. For image super-resolution (Robinson et al., 2010), the classical approaches are to estimate or interpolate the coefficients of wavelet sub-bands for refining image details. Recently, DWT has also been applied in deep learning-based image denoising. The winner of the NTIRE 2020 Denoising Challenge (Abdelhamed et al., 2020) proposes a multi-level wavelet ResNet for image denoising, where DWT and IDWT are used for downsampling and upsampling. Guo et al. (2017) propose a deep wavelet super-resolution model to recover the residuals of wavelet coefficients of the low resolution image. Bae et al. (2017) present a wavelet residual network for image denoising and image super-resolution. Both Guo et al. (2017) and Bae et al. (2017) only use one level wavelet transformation. Liu et al. (2020) develop WDNet for image demoireing working directly in the wavelet domain. Liu et al. (2018) propose a multi-level wavelet-CNN that shows good performance on several image restoration tasks.

In recent years, discrete wavelet transform (DWT) has also been applied in HDR imaging. Omrani et al. (2020) propose a wavelet-based method that aims to use the high-frequency sub-bands obtained from the wavelet decomposition of the input images to recover the details. However, it does not fully utilize the low-frequency sub-band. Kaftan et al. (2009) introduce a wavelet-based method to remove noise from the input images with a correlation analysis among them. Ramakrishnan and Pete (2022) use Haar wavelet to decompose input images into four different frequency bands. Then, different frequency sub-bands are fused using different predefined fusion rules. Finally, IDWT is applied to achieve the final HDR image using the fused frequency sub-bands. Zheng et al. (2022) introduce the cross-transform domain neural network for HDR imaging, which consists of a merging module and a restoration module. In the restoration module, DWT is used to build a cross-transform domain learning block to effectively remove the ghosting artifacts. In the experiments, we observe that in HDR imaging, ghosting artifacts caused by large foreground motion are of low-frequency, while the details are of high-frequency. Since DWT can decompose the input image into different frequency sub-bands, we can take advantage of this property to remove ghosting artifacts or recover details using different frequency sub-bands. Therefore, different from above approaches, our FHDRNet combines wavelet transform

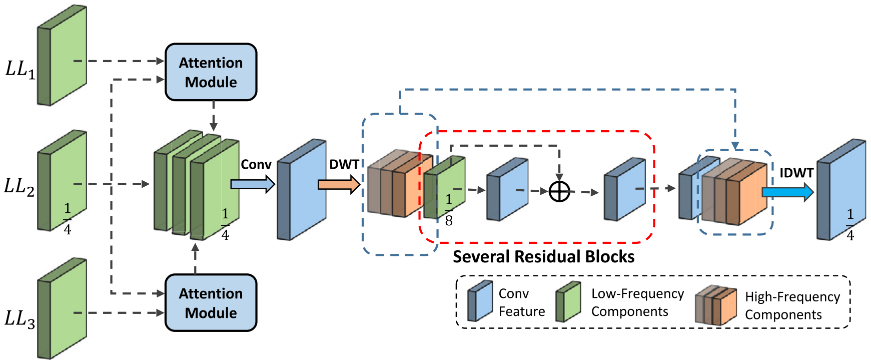


Fig. 4. Structure of the merging module. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with deep learning based method to treat different frequency sub-bands separately. The attention module is used to remove the ghosting artifacts on the low-frequency sub-band and high-frequency sub-bands are used to restore details.

3. Methodology

Given a set of LDR images $\{L_1, L_2, \dots, L_n\}$ with different exposure times, the task of HDR imaging aims to reconstruct an HDR image H that is aligned with the reference frame L_{ref} (*e.g.*, the medium exposure LDR image) In this paper, we follow Kalantari and Ramamoorthi (2017), Wu et al. (2018), Yan et al. (2019a) and use three pairs of LDR and HDR images as input. The corresponding HDR images are obtained from the LDR inputs using a gamma correction function:

$$H_i = \frac{L_i^\gamma}{t_i}, i = 1, 2, 3, \tag{1}$$

where γ is set to 2.2 as the default gamma parameter, and t_i is the exposure time of L_i. The final input of the network is the concatenation of the LDR and the corresponding HDR images, forming a 3-pair 6-channel input:

$$\{I_1, I_2, I_3\} = \{\{L_1, H_1\}, \{L_2, H_2\}, \{L_3, H_3\}\}. \tag{2}$$

3.1. Overview of our network structure

The proposed network has a U-Net like structure, as shown in Fig. 3, containing an encoder, a merger and a decoder with skip connections. In the encoder, the inputs $\{I_1, I_2, I_3\}$ are sent into three independent sub-networks. In each sub-network, DWT is used for decomposing the feature maps into different frequency sub-bands $\{LL_i, LH_i, HL_i, HH_i\}$ ($i = 1, 2, 3$), among which only the low-frequency sub-band LL_i is used for the next stage (scale) processing. All frequency sub-bands are also sent to the corresponding frequency-guided upsampling modules through skip connections. The merger fuses the three inputs (in the low-frequency sub-band) into a ghost-free one, which is then sent to the decoder. The network also includes two significant modules: merging module (Section 3.3) and frequency-guided upsampling module (Section 3.4). The merging module takes only low-frequency components of the previous stage as input and generates a merged result, focusing on structural information. In the decoder, the frequency-guided upsampling module is used to process features in the low-frequency and high-frequency sub-bands independently and then reconstruct the feature maps to a finer scale using IDWT. A global residual connection is also used to enhance the feature representation ability of the network. The output of the network passes through a tone mapping function (using μ-law) to generate the final tone-mapped HDR image:

$$\mathcal{T}(H) = \frac{\log(1 + \mu H)}{\log(1 + \mu)}, \tag{3}$$

where H is the generated HDR output and μ is set to 5000 as default to adjust the compression level.

3.2. Encoder using DWT

The original inputs $\{I_1, I_2, I_3\}$ are firstly sent into three independent sub-networks to extract features individually. The features after the first convolution layer (*conv1*) are transformed into different frequency sub-bands through DWT, including one low-frequency component $LL_i^{\text{level-1}}$ and three high-frequency components, $\{LH_i^{\text{level-1}}, HL_i^{\text{level-1}}, HH_i^{\text{level-1}}\}$, where i denotes the ith input. According to Liu et al. (2020), the low-frequency sub-band contains more structure information and the high-frequency sub-bands contain more detailed information. In order to effectively leverage the decomposed data, the low-frequency component $LL_i^{\text{level-1}}$ is used for further decomposition. In the corresponding frequency-guided upsampling module, the high-frequency components can provide details. So we keep them $\{LH_i^{\text{level-1}}, HL_i^{\text{level-1}}, HH_i^{\text{level-1}}\}$ for reconstruction. Then, $LL_i^{\text{level-1}}$ goes through the feature extraction (*conv2*) and DWT again. The resulting high-frequency components $\{LH_i^{\text{level-2}}, HL_i^{\text{level-2}}, HH_i^{\text{level-2}}\}$ are kept for later reconstruction, while the low-frequency component $LL_i^{\text{level-2}}$ is sent to the feature merging module (Section 3.3) to conduct feature fusion.

3.3. Merging module

The merging module aims at reducing the low-frequency artifacts (*e.g.*, ghosting) by fusing only the low-frequency components (see Fig. 4). Inspired by AHDR (Yan et al., 2019a), attention mechanism is applied to deal with the misalignment and saturated regions. The support frames $\{LL_1^{\text{level-2}}, LL_3^{\text{level-2}}\}$ are firstly sent into the attention modules along with the reference frame $LL_2^{\text{level-2}}$ to generate corresponding weighted masks M_1 and M_3. The attention module includes two convolution layers (3 × 3 kernel size), with stride and zero padding equal to 1. A sigmoid function is used to normalize the values of the masks to [0, 1]. Next, the feature maps of the support frames are masked and weighted with the masks using element-wise multiplication to get the filtered feature maps $\{LL_1^{\text{level-2}'}, LL_3^{\text{level-2}'}\}$:

$$LL_i^{\text{level-2}'} = M_i \odot LL_i^{\text{level-2}}, i = 1, 3, \tag{4}$$

where $\odot$ denotes element-wise multiplication. These filtered feature maps and the reference frame's feature maps are concatenated and go through a convolution layer. DWT is applied again to decompose the previous feature map into frequency components with a lower scale for efficient fusion, where the low-frequency component $LL_i^{\text{level-3}}$ goes through 9 residual blocks to conduct feature fusion. Finally, the pre-saved high-frequency components $LH_i^{\text{level-3}}, HL_i^{\text{level-3}}, HH_i^{\text{level-3}}$, and the merged feature are used as the input of IDWT to recover fused feature maps $F^{\text{level-2}'}$ for the frequency-guided upsampling module.

3.4. Frequency-guided upsampling module

Different from the previous works (Liu et al., 2018; Guo et al., 2017) that use IDWT to reconstruct feature maps from the filtered frequency

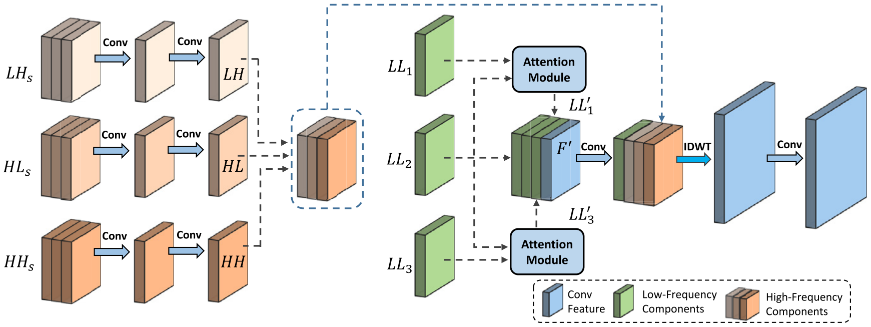


Fig. 5. Structure of the frequency-guided upsampling module. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

sub-bands that all go through the same process, our method leverages the decomposed components that go through different processes with the aim of further fusing lower frequency components. As shown in Fig. 5, three sets (each for one input) of decomposed components are used for restoration. Firstly, the high-frequency components are re-grouped into three groups according to their frequency sub-bands: LH_s, HL_s, HH_s, where $LH_s = \{LH_1, LH_2, LH_3\}$, etc. Then, each group is fused by two convolution layers to generate a single set of high-frequency components. The low-frequency components are fused in a similar way to the merging module by going through the attention modules, and $\{LL'_1, LL_2, LL'_3\}$ along with fused feature maps F' (from the previous stage) are concatenated and go though a convolution layer for fusion. Finally, IDWT is applied on the fused low and high-frequency components to reconstruct the feature maps. An extra convolution layer is used to squeeze the output's size.

3.5. Training loss

Two types of loss are used to train our network: reconstruction loss and Sobel loss. The reconstruction loss is ℓ_1 loss which is the sum of the pixel-wise errors between the generated HDR image and the ground truth. We adopt ℓ_1 loss because it is proved effective for image restoration tasks (Yan et al., 2019a). For the HDR imaging problem, it has been shown that ℓ_1 loss of the tone-mapped images is better than the ℓ_1 loss in the linear space. The tone mapping function $\mathcal{T}(\cdot)$ is applied to the output to generate the HDR image using μ-law. A basic reconstruction loss is defined as below:

$$\mathcal{L}_{\mathcal{R}} = \left\| \mathcal{T}(\hat{H}) - \mathcal{T}(H) \right\|_1, \tag{5}$$

where $\hat{H}$ is the predicted HDR linear RGB image and H is the ground truth.

In order to keep the structure information in the generated HDR image, we also use the Sobel loss, which is:

$$\mathcal{L}_{sobel} = \|\nabla_x \mathcal{T}(\hat{H}) - \nabla_x \mathcal{T}(H)\|_1 \tag{6}$$

$$+ \|\nabla_y \mathcal{T}(\hat{H}) - \nabla_y \mathcal{T}(H)\|_1, \tag{7}$$

where ∇_x and ∇_y are the Sobel edge operator in the x direction and y direction respectively. Our final loss is defined as:

$$\mathcal{L}_{total} = \mathcal{L}_{\mathcal{R}} + \lambda \cdot \mathcal{L}_{sobel}, \tag{8}$$

and λ is a balancing parameter.

4. Experiments and results

4.1. Datasets

In Section 1, we have introduced the advantages that processing HDR fusion in the RAW domain. In order to satisfy the requirements that developing HDR imaging algorithms in the wearable devices (*e.g.*, smart phone). We create a new dataset for training and evaluating HDR imaging algorithms in the RAW domain. The data capturing and ground truth merging is according to the method in Kalantari (Kalantari and Ramamoorthi, 2017) and the device is SONY ILCE-7RM2. We capture two sets of images for the same scene: the static set and the dynamic set. Each set contains three images captured with different exposure bias and with high resolution (5120 × 3456) using RAW format. In the static set, the object is kept static during the capturing, and these images are mainly used to generate the ground truth HDR images. In the dynamic set, the object will do some different movements, and these images are used as inputs for the network. In our dataset, we capture both classic HDR imaging scenes and objective scenes. In order to have an objective evaluation of the generated HDR images, some professional standards are introduced to our dataset, such as Film calibration plate (*e.g.*, details) and SpyderCheckr (*e.g.*, color). The examples of our dataset are in Fig. 6.

Furthermore, we also provide the corresponding RGB images and metadata (*e.g.*, ISO, F-number, exposure time, exposure bias, and white balance coefficients) for each set of samples, and these extra data can be used for future works (*e.g.*, training deep learning based end-to-end ISP pipelines). The provided metadata can also be used to calculate the precise exposure ratio (ER) between images, instead of using exposure bias to get an approximate value.

Finally, We capture 253 sets of samples in total and keep 100 sets, where there is no scene motion or object motion in the static sets (*e.g.*, pixel shift is smaller than 5 pixels) for producing better ground truth HDR images. In the experiment, 85 sets of samples are used for training, and 15 sets of samples are used for evaluating.

The experiments are conducted on four public datasets, including three real datasets: Kalantari dataset (Kalantari and Ramamoorthi, 2017), Prabhakar dataset (Prabhakar et al., 2019) and Tursun dataset (Tursun et al., 2016), and one synthetic dataset: Samsung dataset (Hu et al., 2020). Among them, Kalantari, Prabhakar, and Samsung datasets are used for quantitative and qualitative evaluation, while Tursun dataset is used for qualitative comparison only as it does not provide ground truth. In addition, our RAW dataset is also used for training and evaluating the proposed method in the RAW domain. The Kalantari dataset (Kalantari and Ramamoorthi, 2017) includes 74 training samples and 15 testing samples. Each sample contains three LDR images which are captured with different exposure biases: $\{-2, 0, 2\}$ or $\{-3, 0, 3\}$, and the size of each image is 1500 × 1000. Prabhakar dataset (Prabhakar et al., 2019) has 116 testing samples and it is used only for evaluation. The Samsung dataset (Hu et al., 2020) is a synthetic one, containing 100 samples. The dataset is created in a similar way to the Kalantari dataset, except that all the data are synthesized through a game engine. We choose the first 85 samples for training, and the last 15 for testing. Our RAW dataset is also created in a similar way with higher resolution, and it includes 85 training samples and 15 testing samples.

In the experiments, the training and evaluation are divided into three parts: (1) For real images, the model is trained on the Kalantari

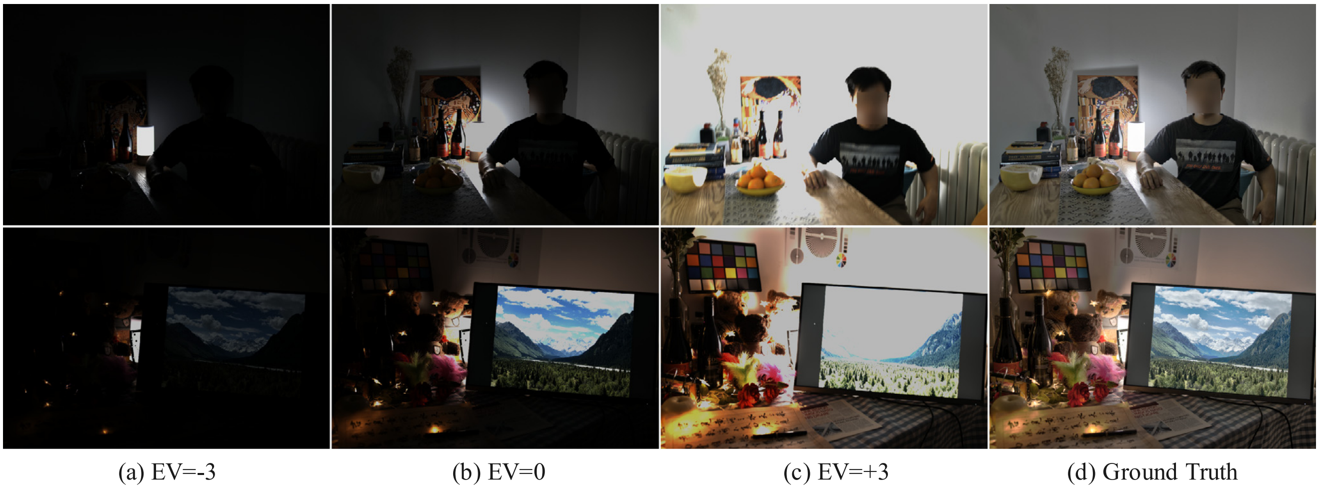


Fig. 6. Examples of LDR inputs and corresponding ground truth HDR images in the RAW dataset. (a)–(c) LDR inputs with different exposure bias. (d) Ground truth HDR images. Face has been processed by using Gaussian filter for privacy protection.

and Ramamoorthi's (2017) training samples and evaluated on the Kalantari and Ramamoorthi (2017) and Prabhakar et al. (2019) testing samples; (2) For synthetic images, the training and evaluation are on the Samsung dataset (Hu et al., 2020); (3) For RAW images, training and evaluation are on the RAW dataset. For those training samples with ground truth, during training, the images are randomly cropped into 256 × 256 small patches and then data augmentation (*e.g.*, flip and rotate) is applied for effective training. During evaluation, the entire test images are fed into the network to predict the HDR images.

4.2. Experimental settings

(1) Implementation Details: During training, Adam (Kingma and Ba, 2015) is selected as the optimizer. The initial learning rate is 2×10^{-4}. After 20,000 epochs, it is reduced to 2×10^{-5}, and after 20,000 epochs, it is further reduced to 2×10^{-6}. We train the network for 60,000 epochs. The batch size is 16. Haar wavelet is used for frequency decomposition. The balancing parameter λ is set to 0.25. The code and the dataset are available at: https://github.com/TianhongDai/wavelet-hdr.

(2) Compared Methods and Evaluation Metrics: Our model is compared with 9 state-of-the-art methods, including Sen et al. (2012), Hu et al. (2013), Kalantari and Ramamoorthi (2017), Wu18 (Wu et al., 2018), NHDRRNet (Yan et al., 2020), SCHDR (Prabhakar et al., 2019), AHDR (Yan et al., 2019a), DAHDR (Yan et al., 2021) and HDRGAN (Niu et al., 2021). Among them, there are two patch-based (Sen et al., 2012; Hu et al., 2013), two optical flow based with CNNs (Kalantari and Ramamoorthi, 2017; Prabhakar et al., 2019) and five CNN based without using optical flow (Wu et al., 2018; Yan et al., 2019a, 2020, 2021; Niu et al., 2021). For quantitative evaluation, we follow Kalantari and Ramamoorthi (2017) to compare PSNR and SSIM results for the linear RGB images and for the tone-mapped images. PSNR-μ and SSIM-μ are for the HDR images after tone mapping using μ-law. PSNR-L and SSIM-L are for the HDR images in the linear space. PSNR-PU and SSIM-PU are for the HDR images using the perceptual uniform encoding (Azimi et al., 2021), and the value of peak luminance is 4000. PSNR-M is for the tone-mapped HDR images using the MATLAB built-in function. In addition, we also use HDR-VDP-2 (Mantiuk et al., 2011), which is an evaluation metric specially designed to evaluate the visual quality of HDR images, and following parameters are used: display diagonal is 21, viewing distance is 1, and color encoding is "rgb-native". Furthermore, in order to provide a fair comparison, we re-train HDRGAN (Niu et al., 2021) by using the official code with a 256 × 256 patch size (the original model uses a patch size of 512 × 512), and other hyperparameters are the same as the default setting.

4.3. Comparison with state-of-the-arts

(1) Quantitative Results: Table 1 shows the quantitative comparison between the state-of-the-art models and ours on three datasets: Kalantari and Ramamoorthi (2017), Prabhakar et al. (2019), and Samsung (Hu et al., 2020). Our model outperforms most of them, especially on the Kalantari and the Samsung datasets, where it achieves six best and two second best results in eight evaluation metrics on the Kalantari dataset, and achieves five best and two second best results in eight evaluation metrics on the Samsung dataset. For the Prabhakar dataset, our model has four best and three second best results in eight evaluation metrics. HDRGAN (Niu et al., 2021) achieves the second best results in most of metrics on the Kalantari dataset, because it uses adversarial learning to restore the missing information in the generated HDR. However, HDRGAN performs poorly on the Samsung dataset, probably due to the sensitivity of the adversarial training to hyperparameters and network structure. Different from other optical flow free methods using the U-Net structure, AHDR (Yan et al., 2019a) and DAHDR (Yan et al., 2021) adopt a network structure in a fixed scale. Therefore, AHDR and DAHDR can preserve more information during encoding and merging. With the assistance of the attention mechanism which can detect the misalignment and saturated regions, AHDR and DAHDR achieve the top three scores in most cases. Compared with AHDR and DAHDR, our model consistently outperforms them across all three datasets.

(2) Qualitative Results: From Figs. 7 to 11, we show the qualitative comparisons on three public datasets (Kalantari and Ramamoorthi, 2017; Prabhakar et al., 2019; Hu et al., 2020). Sen et al. (2012) and Hu et al. (2013) generate strong ghosting artifacts in the images with large foreground motion. These traditional methods have worse performances compared with deep learning-based methods. Optical flow based methods Kalantari17 (Kalantari and Ramamoorthi, 2017) and SCHDR (Prabhakar et al., 2019), in which the input frames are aligned using optical flow before the further merging operation, benefiting a lot from the explicit alignment. But inaccurate optical flow estimation leads to ghosting artifacts, especially in the areas of large motion (see Figs. 7 and 8). Wu18 (Wu et al., 2018) and NHDRRNet (Yan et al., 2020) produce gridding artifacts (see Figs. 8 and 11), because of deconvolution for upsampling. AHDR (Yan et al., 2019a) and DAHDR (Yan et al., 2021) also produce ghosts in Fig. 8 and Fig. 7, respectively. From these results, our method shows better details than other baselines, because details are preserved in high-frequency sub-bands. Through merging features using low-frequency components with the attention mechanism, the ghosting artifacts are also relieved compared with other methods.

Table 1
Quantitative comparison between the baselines and our proposed network on three public testing datasets: Kalantari (Kalantari and Ramamoorthi, 2017), Prabhakar (Prabhakar et al., 2019) and Samsung (Hu et al., 2020).

Dataset	Model	Sen	Hu	Kalantari17	Wu18	NHDRRNet	SCHDR	AHDR	DAHDR	HDRGAN	Ours
Kalantari	PSNR-μ	40.9453	32.1872	42.7423	41.6377	42.4769	40.4700	43.6172	43.8400	43.8746	**43.9066**
	SSIM-μ	0.9805	0.9716	0.9877	0.9869	0.9942	0.9931	0.9956	0.9956	**0.9958**	0.9957
	PSNR-L	38.3147	30.8395	41.2518	40.9082	40.1978	39.6800	41.0309	41.3100	41.0931	**41.4736**
	SSIM-L	0.9749	0.9511	0.9845	0.9847	0.9889	0.9899	0.9903	0.9905	**0.9907**	**0.9907**
	PSNR-PU	34.4651	27.8629	36.3597	35.8021	36.0498	36.3154	37.2419	37.0055	37.4420	**37.4677**
	SSIM-PU	0.9783	0.9623	0.9844	0.9811	0.9810	0.9829	0.9848	0.9850	**0.9862**	0.9858
	PSNR-M	30.5507	25.5937	32.0458	31.0998	34.4113	32.3244	33.0429	33.2900	35.2171	**35.4163**
	HDR-VDP-2	60.5425	57.8278	64.6319	58.3739	63.1585	62.6192	64.8465	64.6765	64.7617	**65.3235**
Prabhakar	PSNR-μ	32.7831	30.8200	35.3400	31.3100	33.0926	30.5700	33.7200	35.3408	35.1984	**35.5652**
	SSIM-μ	0.9740	0.9710	0.9782	0.9733	0.9597	0.9715	0.9789	0.9798	**0.9829**	0.9811
	PSNR-L	30.4985	28.8700	32.0800	30.7200	28.8839	31.4400	31.8300	32.1148	30.9183	**33.0187**
	SSIM-L	0.9749	0.9564	**0.9818**	0.9518	0.9389	0.9722	0.9674	0.9784	0.9717	0.9779
	PSNR-PU	26.3589	23.8109	**29.9942**	26.9620	26.3631	26.3823	27.5110	28.8940	28.4855	29.0431
	SSIM-PU	0.9394	0.9346	0.9477	0.9304	0.8984	0.9356	0.9500	0.9478	**0.9560**	0.9520
	PSNR-M	23.5772	27.2642	28.4386	28.2246	27.5843	27.7573	28.8104	29.3938	29.3619	**29.4924**
	HDR-VDP-2	58.4144	59.6765	62.9073	62.4351	59.9271	62.4376	62.3386	61.9452	62.9463	**63.4667**
Samsung	PSNR-μ	22.8929	34.0052	23.0547	41.2544	41.6741	40.1686	45.1167	45.4359	44.0909	**45.6199**
	SSIM-μ	0.8870	0.9896	0.8922	0.9938	0.9942	0.9902	0.9972	**0.9975**	0.9958	0.9974
	PSNR-L	24.0611	30.3692	26.1808	44.0798	43.9048	41.4633	46.4468	47.2869	44.9299	**48.1514**
	SSIM-L	0.9541	0.9872	0.9628	0.9976	0.9976	0.9967	0.9989	0.9990	0.9986	**0.9991**
	PSNR-PU	17.4251	27.5395	17.7897	36.5663	36.0027	34.5844	39.3863	39.7680	38.5820	**39.9739**
	SSIM-PU	0.8545	0.9796	0.8646	0.9883	0.9880	0.9824	0.9943	**0.9947**	0.9926	0.9943
	PSNR-M	21.2455	30.7557	16.0222	30.9670	32.4359	28.6001	28.8571	26.6076	32.5280	**34.1855**
	HDR-VDP-2	55.4090	66.2029	58.2501	71.6399	70.1484	70.7524	74.6180	**75.0042**	74.6636	74.2816

The best and the second best results are bold and underlined, respectively.

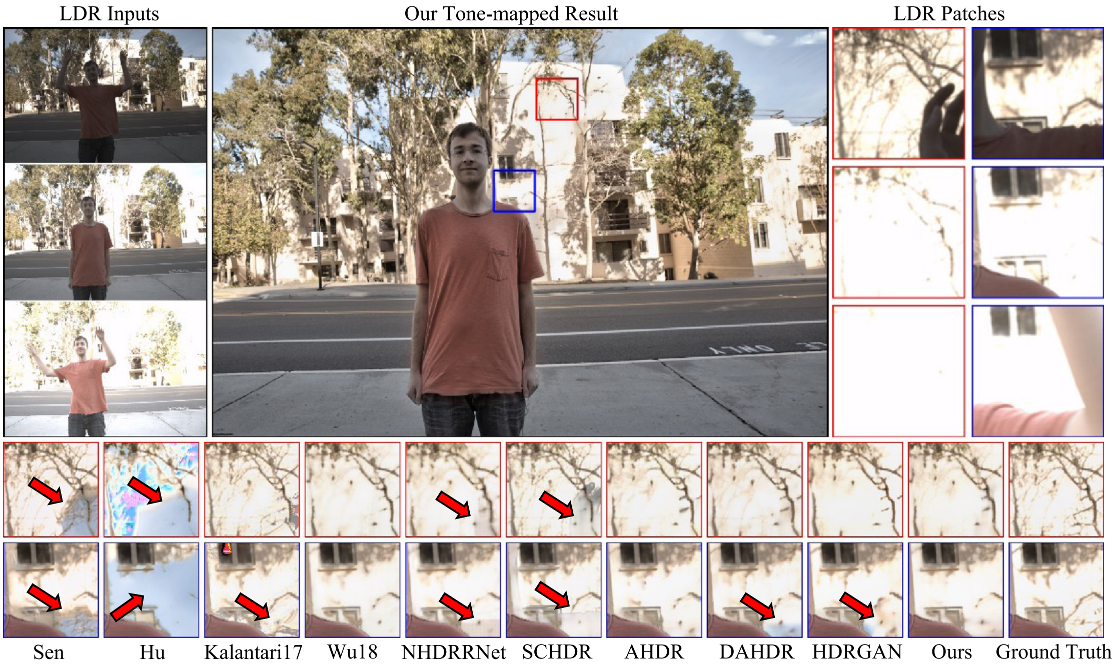


Fig. 7. Qualitative comparison between our method and the baselines on the Kalantari testing dataset (Kalantari and Ramamoorthi, 2017).

4.4. Ablation studies

In this section, we conduct ablation studies to investigate the contribution of each module in our model on the Kalantari dataset. As shown in Table 2, our ablation studies focus on the following parts: (1) only process low-frequency and high-frequency separately, (2) the importance of the attention mechanism, (3) different types of wavelet, (4) different types of methods to fuse high-frequency components in the upsampling module, (5) the Sobel loss function, and (6) the importance of using ONLY the low-frequency component for further processing (next scale) and merging.

(1) Frequency-Specific Processing: We design a "U-Net + DWT" model that is basically a U-Net except that it processes the low and high-frequency sub-bands separately. The naive replacement leads to an improvement of 0.78 dB in terms of PSNR-μ over the U-Net baseline. This model can outperform several state-of-the-art deep learning models (Wu et al., 2018; Yan et al., 2020; Prabhakar et al., 2019) that adopt U-Net as the backbone. This is because the high-frequency components can preserve more details. As shown in Fig. 12, the results of "U-Net + DWT" are smoother and also with better details than U-Net.

(2) Attention Mechanism: Inspired by AHDR, attention modules are also used in both the merging module and the upsampling module of

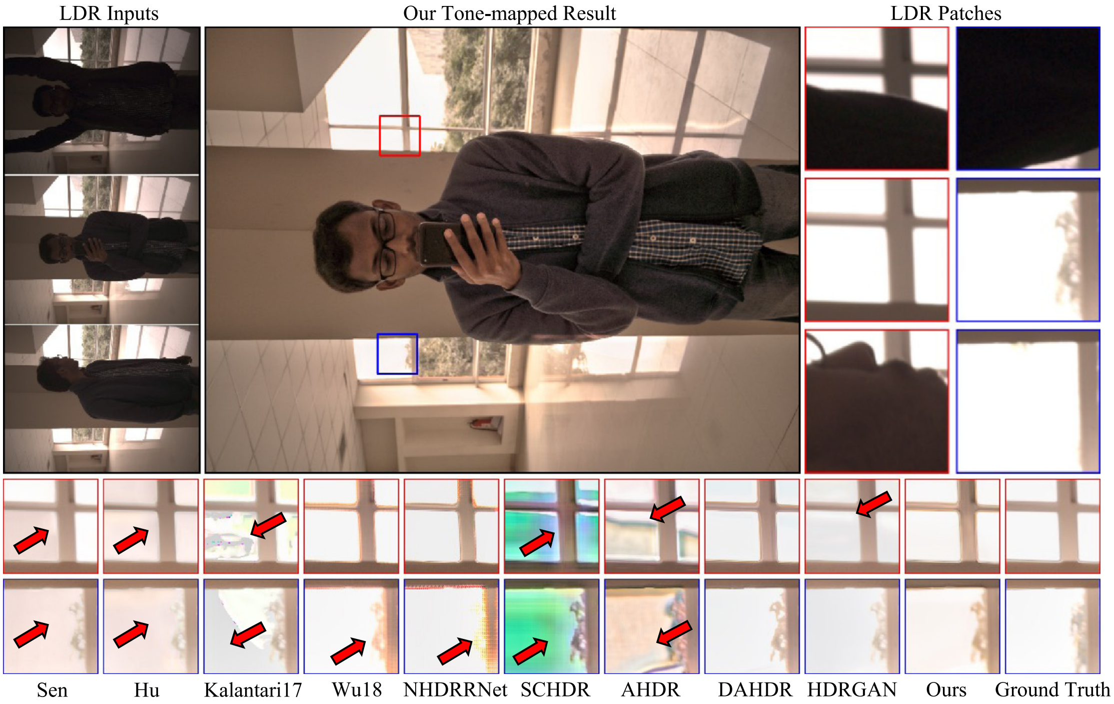


Fig. 8. Qualitative comparison between our method and the baselines on the Prabhakar testing dataset (Prabhakar et al., 2019).

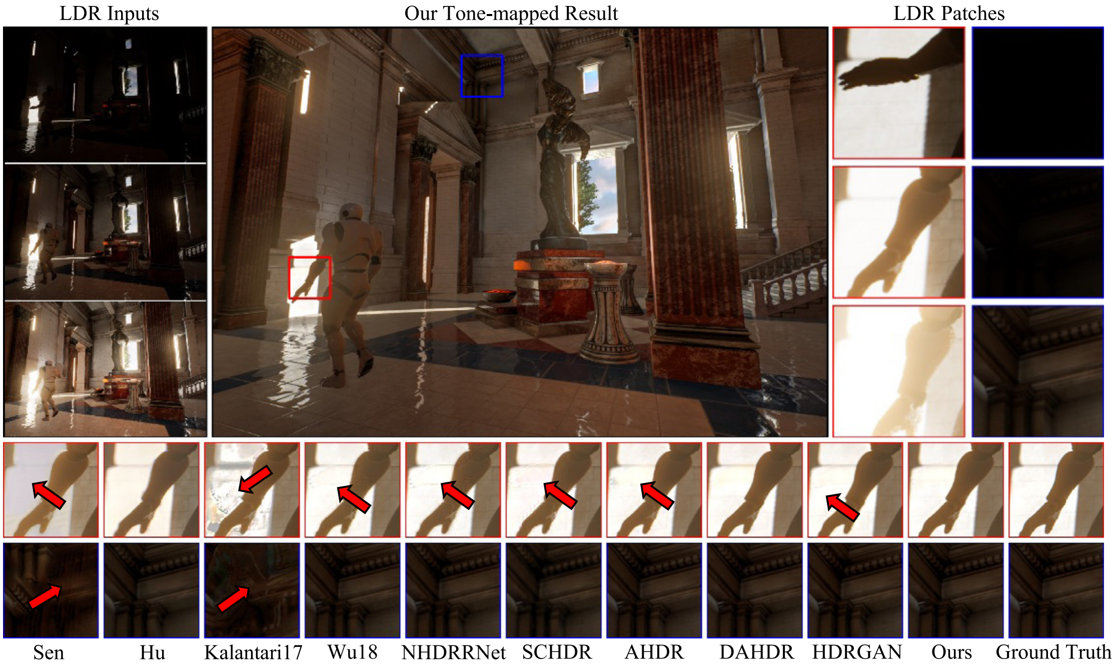


Fig. 9. Qualitative comparison between our method and the baselines on the Samsung testing dataset (Hu et al., 2020).

the proposed model. To verify its contribution, we remove the attention modules (indicated as "w/o Attention" in Table 2). Compared with our final model, it shows that removing the attention modules leads to 0.52dB decrease in terms of PSNR-μ. Different from AHDR and DAHDR which applies the attention to all feature maps of the original scale, we only apply the attention to the low-frequency components of the feature maps on smaller scales (1/8, 1/4, and 1/2). By designing in this way, we specifically align the lower-frequency sub-band to remove ghosting artifacts and also save computation. As shown in Fig. 12, the results of "w/o Attention" have ghosting artifacts.

(3) Types of Wavelet: In addition to the default Haar wavelet, various types of wavelet are also evaluated: Symlet wavelet (indicated as "sym2") and Daubechies wavelets with approximation order 2 and 3 (indicated as "db2" and "db3"). Our model with Haar wavelet outperforms the models with other wavelets. However, using other types of wavelet still gets comparable results, which shows the robustness of our method to the type of wavelet.

(4) Fusion Methods of High-Frequency Components: Another approach to fuse the high-frequency components in the upsampling module is also investigated. Firstly, it groups the components from different

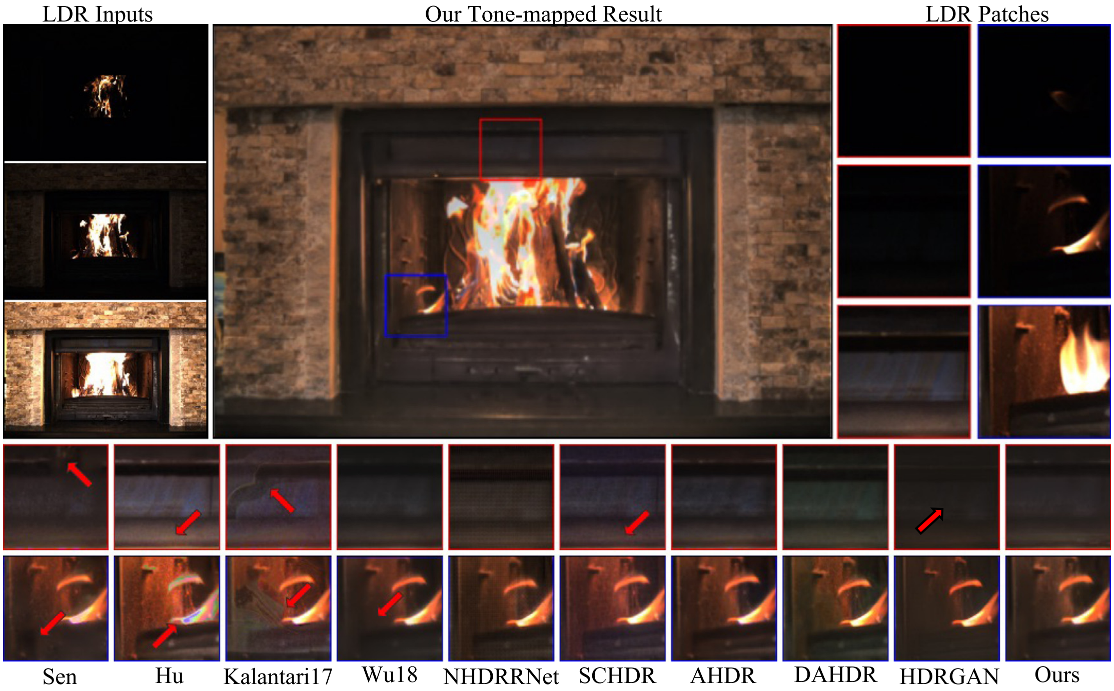


Fig. 10. Qualitative comparison between our method and the baselines on the Tursun dataset (Flames) (Tursun et al., 2016).

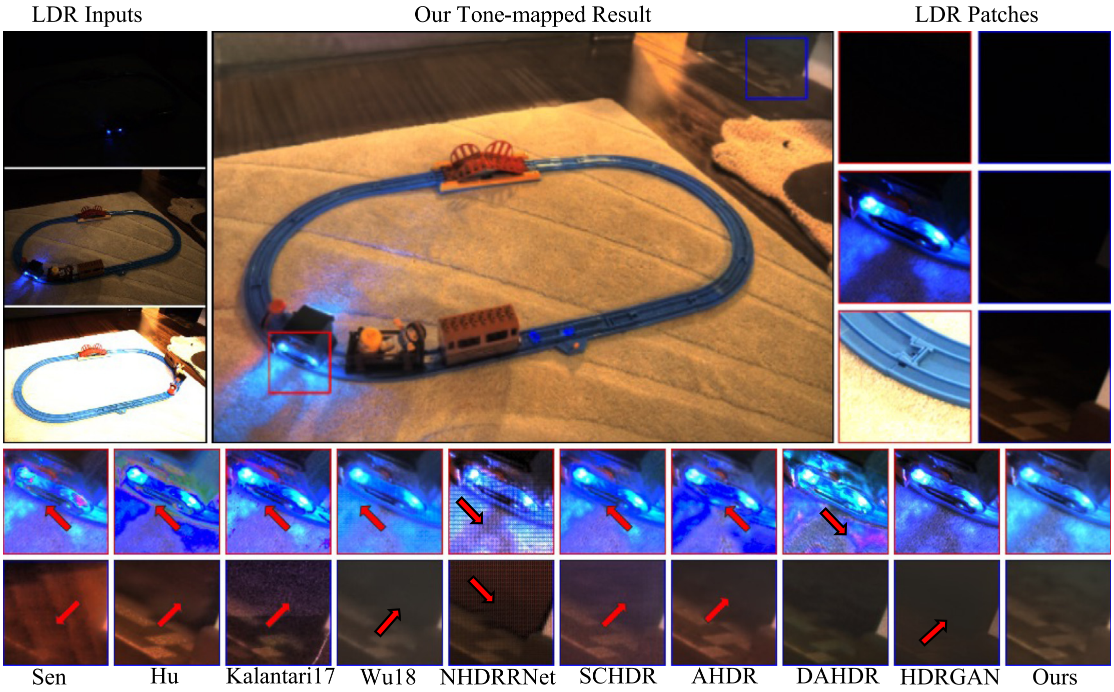


Fig. 11. Qualitative comparison between our method and the baselines on the Tursun dataset (ToyTrain) (Tursun et al., 2016).

inputs with specific frequencies, and then averages the values of these components pixel by pixel to get the fused high-frequency components. The average fusion method has worse performance than the CNN fusion.

(5) Sobel Loss Function: The Sobel loss contributes 0.28 dB improvement for the score of PSNR-μ. It can guide the model to recover better edge information.

(6) Using only Low-Frequency after Decomposition: To verify our design of using only the low-frequency component after decomposition, all frequency sub-bands are used for the next stage's processing (indicated as "Ours^{-}"), and it leads to a decrease of PSNR-μ by 1.01 dB. As shown in Fig. 12, the results (Ours^{-}) contain ghosting artifacts.

4.5. Trade-off between quality and efficiency

High dynamic range (HDR) imaging algorithms are widely used in the real-world devices (*e.g.*, smart phones). Therefore, computational efficiency is also an important factor to evaluate the performance of

Table 2
Quantitative results of ablation studies on the Kalantari testing dataset (Kalantari and Ramamoorthi, 2017).

Model	U-Net	U-Net+DWT	w/o Attention	Wavelet:sym2	Wavelet:db2	Wavelet:db3	Average Fusion	w/o Sobel Loss	Ours^{-}	Ours
PSNR-μ	42.0488	42.8238	43.4657	43.6048	43.5820	43.4675	43.1682	43.6257	42.8969	**43.9066**
SSIM-μ	0.9936	0.9950	0.9953	0.9955	0.9955	0.9954	0.9951	0.9954	0.9947	**0.9957**
PSNR-L	39.0038	40.2282	40.5313	40.8969	40.7055	40.8254	40.7980	40.9293	40.8455	**41.4736**
SSIM-L	0.9852	0.9892	0.9895	0.9899	0.9897	0.9901	0.9903	0.9899	0.9899	**0.9907**
PSNR-M	32.6175	34.4052	34.5538	35.3321	34.5855	**35.7751**	35.0043	35.1833	34.8647	35.4163
HDR-VDP-2	61.3812	63.3845	63.5026	64.8847	64.3280	64.4923	64.6400	64.4789	64.5697	**65.3235**

The best and the second best results are bold and underlined, respectively.

Table 3
Comparison of the running time (second) and GPU memory (GB) with corresponding PSNR-μ (dB) between the baselines and our method for generating a 1500 × 1000 HDR image.

Model	Sen	Hu	Kalantari17	Wu18	NHDRRNet	SCHDR	AHDR	DAHDR	HDRGAN	Ours
Time	51.96	293.61	68.81	0.21	0.34	0.32	0.78	0.92	0.45	0.59
Memory	–	–	–	3.33	3.60	8.60	10.37	14.83	8.48	7.36
PSNR-μ	40.95	32.19	42.74	41.64	42.48	40.47	43.62	43.84	43.87	43.91

The sign "–" denotes the method is evaluated on a CPU.

Table 4
Quantitative results between baselines and our proposed network on our RAW dataset.

Model	Sen	Hu	Wu18	NHDRRNet	AHDR	DAHDR	Ours
PSNR-μ	32.0580	29.9452	37.532	37.4309	38.7193	38.8811	**39.1243**
SSIM-μ	0.9705	0.9720	0.9869	0.9877	0.9900	0.9903	**0.9911**
PSNR-L	35.7427	28.6619	39.1129	38.6586	39.5321	39.7531	**39.7821**
SSIM-L	0.9386	0.9031	0.9944	0.9943	0.9944	0.9947	**0.9952**
PSNR-PU	25.6776	23.1445	31.5875	31.2615	32.4102	**32.6657**	31.9010
SSIM-PU	0.9363	0.9366	0.9644	0.9639	0.9712	**0.9727**	0.9716
PSNR-M	**29.8978**	27.9381	29.1849	28.4560	29.0868	29.2939	29.1621
HDR-VDP-2	65.2848	59.7245	65.0400	65.6121	65.6350	**65.7841**	65.6577

The best and the second best results are bold and underlined, respectively.

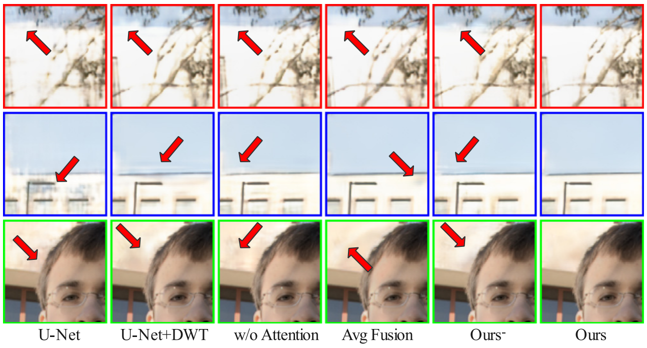


Fig. 12. Qualitative comparison between our method and the baseline variants on the Kalantari testing dataset (Kalantari and Ramamoorthi, 2017).

the algorithms. In this experiment, we test the running time and the memory cost with the corresponding PSNR-μ scores of the baselines and our method in Table 3. The proposed method needs around 0.59 s to generate an HDR image with 1500 × 1000 resolution on a RTX-2080Ti GPU, whereas DAHDR need 0.92 s and has a lower score on PSNR-μ. Besides, DAHDR takes up the most memory among the 8 competitors, because DAHDR merges LDR images in the original scale, and has a "heavy-weighted" network structure. Furthermore, our method, with 50% less memory than DAHDR, can still achieve better performance. Thus, our method has a good balance between quality and efficiency.

4.6. Evaluation on the RAW dataset

We evaluate FHDRNet and compare with state-of-the-art on our new RAW dataset. As shown in Table 4, FHDRNet achieves the best performance in terms of PSNR-μ, SSIM-μ, PSNR-L and SSIM-L, indicating that our model can also be used in the RAW domain. In the qualitative comparison, our method preserves more details in the texture than the baselines (see Fig. 13), because of the efficient utilization of high-frequency sub-bands. For example, in Fig. 13, our method restores better details in the bottle. Thus, the proposed method can also keep its advantages in the RAW domain.

4.7. Limitations and future work

Although in this work our proposed FHDRNet outperforms other baselines on the several datasets, and our new created dataset also provides a platform for training and evaluating HDR algorithms in the RAW domain, these two contributions still have some limitations to be addressed in the future.

(1) Other Challenges: Our proposed FHDRNet and other approaches (Wu et al., 2018; Prabhakar et al., 2019; Yan et al., 2019a, 2021, 2020; Niu et al., 2021) mentioned in this work mainly focus on mitigating the ghosting artifacts caused by large foreground motion in the multi-frame HDR image reconstruction. However, there are some other types of artifacts that need to be addressed in the multi-frame HDR image reconstruction task (Johnson, 2015; Mantiuk et al., 2015):

- Blurry Artifacts: It is usually caused by the global camera motion, such as taking photos with a hand-held camera. The instability of the capturing process will lead misaligned images and make the generated HDR image blurry.
- Noise Artifacts: This is due to some HDR fusin algorithms that are operated on a per-pixel basis. This means the value of each pixel of the generated HDR image is estimated using the value of the pixel at the same location in all input images.
- Glare Artifacts: It is introduced by the defective camera lens or light streaks around the light source created by special filters mounted on the lens.

Thus, in the future, we need to explore more potential solutions so that other types of artifacts can also be addressed at the same time to provide better results.

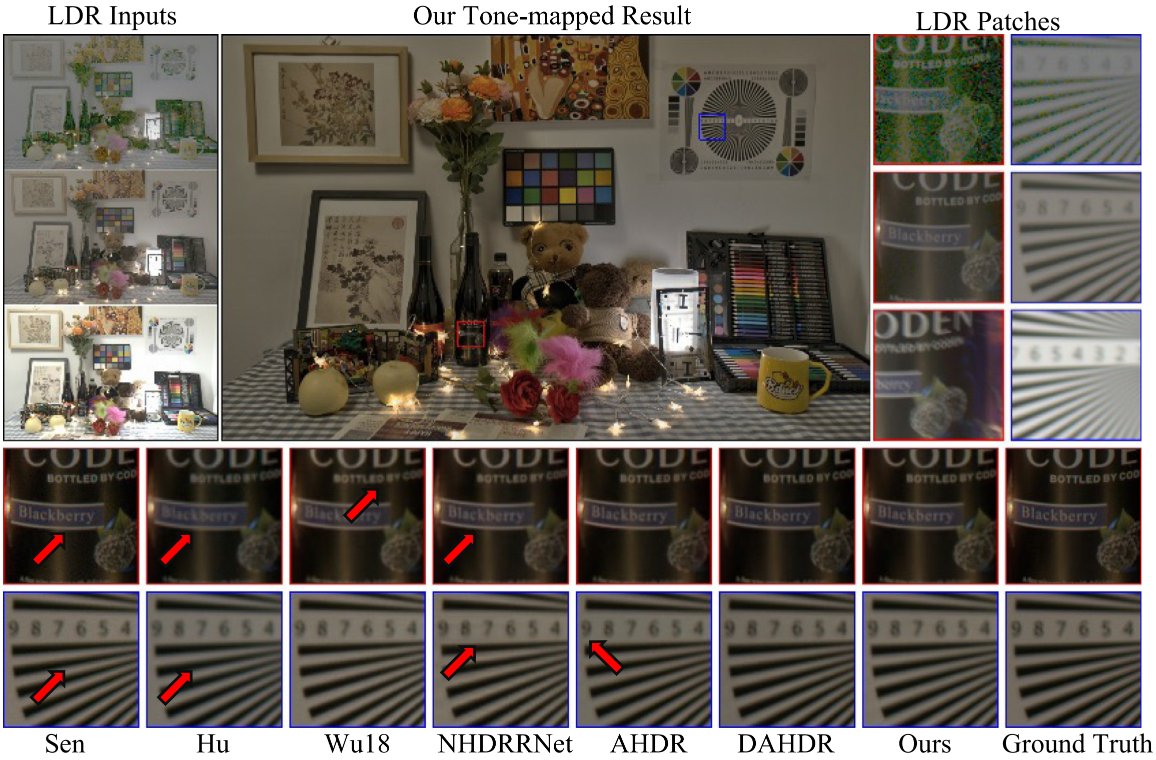


Fig. 13. Qualitative comparison between our method and the baselines on the RAW testing dataset. The RAW images are visualized through the same ISP pipeline.

(2) Raw Dataset: While we capture the images follow a comprehensive pipeline, it still has some aspects to introduce several bias to affect the capability of the models:

- Camera Motion: When we capture the LDR images, a tripod is used to stabilize the camera. However, in a real-world scenario, the tripod is not always available to users, which can occasionally lead to large camera motion. In this case, our dataset does not cover this situation effectively and it can affect the final performance of the model. There are some possible solutions to mitigate this problem: (1) the camera can be held in the hand to capture some new data to expand our current dataset. (2) global motion can be added to the existing data manually (*e.g.*, shift images).
- Limited Scenarios: In our dataset, the images that captured in the outdoor environment are mostly located in the city center, surrounded by buildings and vehicles. Samples with other outdoor scenarios (*e.g.*, natural scenery) are not included in the current dataset, however capturing photos in various environments is also a requirement for the HDR imaging. To further improve the performance of the model in different scenarios, we will collect more samples in various environments to cover more daily use cases in the dataset.
- Limited Size: In this dataset, we keep only 100 high-quality samples for training and evaluation. Although the amount of samples is comparable to the existing datasets, such as Kalantari and Ramamoorthi (2017), Prabhakar et al. (2019) and Samsung (Hu et al., 2020) datasets, it still remains a huge gap compared to the number of samples in datasets from other domains. For example, Flickr2K (Timofte et al., 2017) is a dataset for the single image super-resolution task, which consists of 2650 images with 2K resolution for training and evaluation. The limited training data could affect the generalization ability of the model. Therefore, we will collect more data from different environments to augment our dataset.

5. Conclusion

In this paper, we have proposed a frequency-guided network (FHDRNet) for high dynamic range (HDR) imaging. In the proposed method, the input LDR images are transformed into the wavelet domain using Discrete Wavelet Transform (DWT). The low-frequency sub-bands are mainly used to avoid ghosting artifacts caused by large motion, while the high-frequency sub-bands are used for preserving details. The attention mechanism is adopted to merge low-frequency information to deal with misalignment. The extensive experiments have shown that our method can remove ghosts and preserve more details. It also achieves state-of-the-art results on several public datasets and our RAW dataset with lower computational costs, compared with previous approaches. We believe it has great potential for more extensive applications of HDR imaging.

CRediT authorship contribution statement

Tianhong Dai: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing. **Wei Li:** Methodology, Software, Writing – original draft, Writing – review & editing. **Xilei Cao:** Software, Writing – review & editing. **Jianzhuang Liu:** Supervision, Writing – review & editing. **Xu Jia:** Methodology, Writing – review & editing. **Ales Leonardis:** Supervision, Writing – review & editing. **Youliang Yan:** Writing – review & editing. **Shanxin Yuan:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I have shared the code on GitHub: https://github.com/TianhongDai/wavelet-hdr.

References

Abdelhamed, A., Afifi, M., Timofte, R., Brown, M.S., 2020. Ntire 2020 challenge on real image denoising: Dataset, methods and results. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 496–497.

Amini, M., Ahmad, M.O., Swamy, M., 2018. A robust multibit multiplicative watermark decoder using a vector-based hidden Markov model in wavelet domain. IEEE Trans. Circuits Syst. Video Technol. 28 (2), 402–413.

Azimi, M., et al., 2021. PU21: A novel perceptually uniform encoding for adapting existing quality metrics for HDR. In: 2021 Picture Coding Symposium. pp. 1–5.

Bae, W., Yoo, J., Chul Ye, J., 2017. Beyond deep residual learning for image restoration: Persistent homology-guided manifold simplification. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 145–153.

Bogoni, L., 2000. Extending dynamic range of monochrome and color images through fusion. In: International Conference on Pattern Recognition. Vol. 3. pp. 7–12.

Cid, Y.D., Müller, H., Platon, A., Poletti, P.-A., Depeursinge, A., 2017. 3D solid texture classification using locally-oriented wavelet transforms. IEEE Trans. Image Process. 26 (4), 1899–1910.

Debevec, P.E., Malik, J., 1997. Recovering high dynamic range radiance maps from photographs. In: ACM SIGGRAPH. pp. 1–10.

Donoho, D.L., 1995. De-noising by soft-thresholding. IEEE Trans. Inform. Theory 41 (3), 613–627.

Gallo, O., Troccoli, A., Hu, J., Pulli, K., Kautz, J., 2015. Locally non-rigid registration for mobile HDR photography. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 49–56.

Gueguen, L., Sergeev, A., Kadlec, B., Liu, R., Yosinski, J., 2018. Faster neural networks straight from jpeg. In: Conference on Neural Information Processing Systems. Vol. 31.

Guo, T., Seyed Mousavi, H., Huu Vu, T., Monga, V., 2017. Deep wavelet prediction for image super-resolution. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 104–113.

Haghighat, M., Mathew, R., Naman, A., Taubman, D., 2019. Illumination estimation and compensation of low frame rate video sequences for wavelet-based video compression. IEEE Trans. Image Process. 28 (9), 4313–4327.

Hasinoff, S.W., Sharlet, D., Geiss, R., Adams, A., Barron, J.T., Kainz, F., Chen, J., Levoy, M., 2016. Burst photography for high dynamic range and low-light imaging on mobile cameras. ACM Trans. Graph. 35 (6), 1–12.

Ho, J., Hwang, W.-L., 2012. Wavelet Bayesian network image denoising. IEEE Trans. Image Process. 22 (4), 1277–1290.

Hu, J., Choe, G., Nadir, Z., Nabil, O., Lee, S.-J., Sheikh, H., Yoo, Y., Polley, M., 2020. Sensor-realistic synthetic data engine for multi-frame high dynamic range photography. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 516–517.

Hu, J., Gallo, O., Pulli, K., Sun, X., 2013. HDR deghosting: How to deal with saturation? In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1163–1170.

Huan, W., Li, S., Qian, Z., Zhang, X., 2021. Exploring stable coefficients on joint subbands for robust video watermarking in DT CWT domain. IEEE Trans. Circuits Syst. Video Technol..

Huang, H., He, R., Sun, Z., Tan, T., 2017. Wavelet-SRNet: A wavelet-based cnn for multi-scale face super resolution. In: International Conference on Computer Vision. pp. 1689–1697.

Ji, H., Yang, X., Ling, H., Xu, Y., 2012. Wavelet domain multifractal analysis for static and dynamic texture classification. IEEE Trans. Image Process. 22 (1), 286–299.

Johnson, A.K., 2015. High dynamic range imaging—A review. Int. J. Image Process. 9, 198.

Kaftan, J.N., Bell, A.A., Seiler, C., Aach, T., 2009. Wavelet based denoising by correlation analysis for high dynamic range imaging. In: IEEE International Conference on Image Processing. pp. 3857–3860.

Kalantari, N.K., Ramamoorthi, R., 2017. Deep high dynamic range imaging of dynamic scenes. ACM Trans. Graph. 36 (4), 1–12.

Khan, E.A., Akyuz, A.O., Reinhard, E., 2006. Ghost removal in high dynamic range images. In: IEEE International Conference on Image Processing. pp. 2005–2008.

Kingma, D.P., Ba, J., 2015. Adam: A method for stochastic optimization. In: International Conference on Learning Representations.

Li, Q., Shen, L., Guo, S., Lai, Z., 2020. Wavelet integrated CNNs for noise-robust image classification. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 7245–7254.

Li, Z., Zheng, J., Zhu, Z., Wu, S., 2014. Selectively detail-enhanced fusion of differently exposed images with moving objects. IEEE Trans. Image Process. 23 (10), 4372–4382.

Liu, L., Liu, J., Yuan, S., Slabaugh, G., Leonardis, A., Zhou, W., Tian, Q., 2020. Wavelet-based dual-branch network for image Demoiréing. In: European Conference on Computer Vision. pp. 86–102.

Liu, P., Zhang, H., Zhang, K., Lin, L., Zuo, W., 2018. Multi-level wavelet-CNN for image restoration. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 773–782.

Mantiuk, R., Kim, K.J., Rempel, A.G., Heidrich, W., 2011. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. ACM Trans. Graph. 30 (4), 1–14.

Mantiuk, R.K., Myszkowski, K., Seidel, H.-P., 2015. High dynamic range imaging. In: Wiley Encyclopedia of Electrical and Electronics Engineering. pp. 1–42.

Mishra, D., Singh, S.K., Singh, R.K., 2021. Wavelet-based deep auto encoder-decoder (WDAED)-based image compression. IEEE Trans. Circuits Syst. Video Technol. 31 (4), 1452–1462.

Niu, Y., Wu, J., Liu, W., Guo, W., Lau, R.W., 2021. HDR-GAN: HDR image reconstruction from multi-exposed ldr images with large motions. IEEE Trans. Image Process. 30, 3885–3896.

Omrani, A., Soheili, M.R., Kelarestaghi, M., 2020. High dynamic range image reconstruction using multi-exposure wavelet hdrcnn. In: International Conference on Machine Vision and Image Processing. pp. 1–4.

Pece, F., Kautz, J., 2010. Bitmap movement detection: HDR for dynamic scenes. In: European Conference on Visual Media Production. pp. 1–8.

Pouli, T., Boitard, R., Chamaret, C., Abebe, M., Serré, C., Touzé, D., François, E., Reinhard, E., 2014. Hdr in the living room. In: ACM SIGGRAPH Studio. p. 1.

Prabhakar, K.R., Agrawal, S., Singh, D.K., Ashwath, B., Babu, R.V., 2020. Towards practical and efficient high-resolution HDR deghosting with CNN. In: European Conference on Computer Vision. pp. 497–513.

Prabhakar, K.R., Arora, R., Swaminathan, A., Singh, K.P., Babu, R.V., 2019. A fast, scalable, and reliable deghosting method for extreme exposure fusion. In: International Conference on Computational Photography. pp. 1–8.

Ramakrishnan, V., Pete, D., 2022. Haar wavelet-based fusion of multiple exposure images for high dynamic range imaging. SN Comput. Sci. 3 (2), 129.

Remenyi, N., Nicolis, O., Nason, G., Vidakovic, B., 2014. Image denoising with 2D scale-mixing complex wavelet transforms. IEEE Trans. Image Process. 23 (12), 5165–5174.

Robinson, M.D., Toth, C.A., Lo, J.Y., Farsiu, S., 2010. Efficient Fourier-wavelet super-resolution. IEEE Trans. Image Process. 19 (10), 2669–2681.

Sen, P., Kalantari, N.K., Yaesoubi, M., Darabi, S., Goldman, D.B., Shechtman, E., 2012. Robust patch-based hdr reconstruction of dynamic scenes. ACM Trans. Graph. 31 (6), 1–11.

Sun, D., Yang, X., Liu, M.-Y., Kautz, J., 2018. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 8934–8943.

Suzuki, T., 2019. Wavelet-based spectral–spatial transforms for CFA-sampled raw camera image compression. IEEE Trans. Image Process. 29, 433–444.

Timofte, R., Agustsson, E., Van Gool, L., Yang, M.-H., Zhang, L., 2017. Ntire 2017 challenge on single image super-resolution: Methods and results. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 114–125.

Tursun, O.T., Akyüz, A.O., Erdem, A., Erdem, E., 2016. An objective deghosting quality metric for HDR images. In: Computer Graphics Forum. Vol. 35. No. 2. pp. 139–152.

Williams, T., Li, R., 2018. Wavelet pooling for convolutional neural networks. In: International Conference on Learning Representations.

Wu, S., Xu, J., Tai, Y.-W., Tang, C.-K., 2018. Deep high dynamic range imaging with large foreground motions. In: European Conference on Computer Vision. pp. 117–132.

Yan, Q., Gong, D., Shi, Q., Hengel, A.v.d., Shen, C., Reid, I., Zhang, Y., 2019a. Attention-guided network for ghost-free high dynamic range imaging. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1751–1760.

Yan, Q., Gong, D., Shi, J.Q., van den Hengel, A., Shen, C., Reid, I., Zhang, Y., 2021. Dual-attention-guided network for ghost-free high dynamic range imaging. Int. J. Comput. Vis. 1–19.

Yan, Q., Gong, D., Zhang, P., Shi, Q., Sun, J., Reid, I., Zhang, Y., 2019b. Multi-scale dense networks for deep high dynamic range imaging. In: Winter Conference on Applications of Computer Vision. pp. 41–50.

Yan, Q., Zhang, L., Liu, Y., Zhu, Y., Sun, J., Shi, Q., Zhang, Y., 2020. Deep HDR imaging via a non-local network. IEEE Trans. Image Process. 29, 4308–4322.

Yoo, J., Uh, Y., Chun, S., Kang, B., Ha, J.-W., 2019. Photorealistic style transfer via wavelet transforms. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 9036–9045.

Yue, T., Suo, J., Cao, X., Dai, Q., 2017. Efficient method for high-quality removal of nonuniform blur in the wavelet domain. IEEE Trans. Circuits Syst. Video Technol. 27 (9), 1869–1881.

Zheng, B., Pan, X., Zhang, H., Zhou, X., Slabaugh, G., Yan, C., Yuan, S., 2022. DomainPlus: Cross transform domain learning towards high dynamic range imaging. In: ACM International Conference on Multimedia. pp. 1954–1963.