

Generating Content for HDR Deghosting from Frequency View

Tao Hu^{1†} Qingsen Yan^{1†*} Yuankai Qi² Yanning Zhang¹
¹Northwestern Polytechnical University ²Macquarie University

hutaoxuat@gmail.com, ¹{qingsenyan, ynzhang}@nwpu.edu.cn, ²yuankai.qi@mq.edu.au

Abstract

Recovering ghost-free High Dynamic Range (HDR) images from multiple Low Dynamic Range (LDR) images becomes challenging when the LDR images exhibit saturation and significant motion. Recent Diffusion Models (DMs) have been introduced in HDR imaging field, demonstrating promising performance, particularly in achieving visually perceptible results compared to previous DNN-based methods. However, DMs require extensive iterations with large models to estimate entire images, resulting in inefficiency that hinders their practical application. To address this challenge, we propose the Low-Frequency aware Diffusion (LF-Diff) model for ghost-free HDR imaging. The key idea of LF-Diff is implementing the DMs in a highly compacted latent space and integrating it into a regression-based model to enhance the details of reconstructed images. Specifically, as low-frequency information is closely related to human visual perception we propose to utilize DMs to create compact low-frequency priors for the reconstruction process. In addition, to take full advantage of the above low-frequency priors, the Dynamic HDR Reconstruction Network (DHRNet) is carried out in a regression-based manner to obtain final HDR images. Extensive experiments conducted on synthetic and real-world benchmark datasets demonstrate that our LF-Diff performs favorably against several state-of-the-art methods and is 10× faster than previous DM-based methods.

1. Introduction

Multiple exposure High Dynamic Range (HDR) imaging aims to restore missing details from exposure-varied Low Dynamic Range (LDR) images. However, in dynamic scenes, it often causes ghosting artifacts due to object or camera movement, limiting practical applications. Researchers are actively exploring ghost-free image recon-

*Corresponding author. † The first two authors contributed equally to this work. This work was partially supported by NSFC (62301432, 62306240), NSBRPS (2023-JC-QN-0685, QCYRCXM-2023-057). Y. Qi is not supported by the above mentioned fundings.

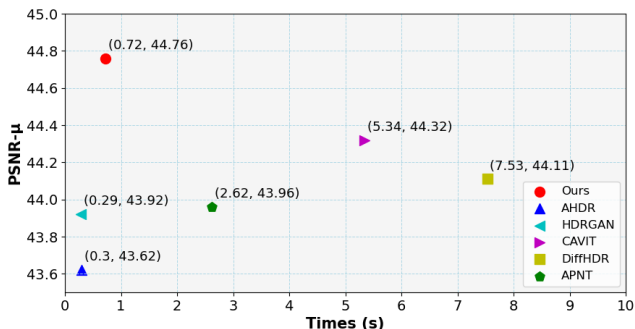


Figure 1. Average PSNR vs. inference time on the Kalantari’s dataset [11]. Our method performs favorably and is 10× faster than the previous diffusion-model-based method DiffHDR [44].

struction methods to enable seamless capture of dynamic scenes with high dynamic ranges.

In recent years, the rise of Deep Neural Networks (DNNs) has brought significant advancements to HDR imaging field. Numerous DNN-based methods have emerged, employing Convolutional Neural Networks (CNNs) [11, 41, 47] or Vision Transformers (ViTs) [16, 28] for HDR image reconstruction. Despite their advancements, DNN-based approaches face challenges when essential information (e.g., content, details) for overexposed areas is missing due to object or camera movement. More recently, Diffusion Models (DMs) [8, 26] have exhibited impressive performance in image synthesis [4, 29] and image reconstruction tasks (including HDR imaging) [24, 44, 46]. DMs achieve this by iteratively denoising Gaussian noise to generate high-fidelity images. Yan *et al.* [44] reconstruct complete HDR images from pure Gaussian noise, yielding impressive results. Note that, compared to previous Deep generative models [14, 21], DMs generate a more accurate target distribution without facing issues like optimization instability or mode collapse, making them a promising choice for enhancing HDR image quality. Nonetheless, DMs demand a substantial number of iteration steps on denoising models to capture intricate data details, which is time-consuming even with a high-end GPU card. For example, as shown in Fig. 1, DiffHDR [44] takes about 7.5 seconds to generate an HDR image on a single A100 GPU.

In this paper, our goal is to devise a DM-based method that effectively harnesses the robust distribution mapping capabilities of DMs for HDR image reconstruction. We note that it is unnecessary to reconstruct a complete HDR image from pure noise, because the LDR reference images already provide most of the necessary content information for HDR images. As low-frequency information is closely related to human visual perception [34], we propose to utilize DMs to create compact low-frequency priors in the latent space. These priors are then integrated into the regression-based model to predict the low-frequency content of the reconstructed HDR images.

To achieve our goal, we propose the Low-Frequency Aware Diffusion Model (LF-Diff), comprising a Low-Frequency Prior Extraction Network (LPENet), a denoising network, and a Dynamic HDR Reconstruction Network (DHRNet). Following established methods [3, 23, 39], we employ a two-stage training strategy for pre-train LF-Diff and DM training. In the first stage, LPENet learns to extract a compact Low-frequency Prior Representation (LPR) from ground-truth images, guiding the DHRNet. DHRNet consists of two modules: a Prior Integration Module (PIM) that fuses the LPR with intermediate features of DHRNet, and a Feature Refinement Module (FRM) that further processes the fused features to HDR image. Notably, LPENet and DHRNet are optimized together in this phase. In the second stage, we train DM to learn the compact LPR directly from LDR images. Since the LPR is lightweight and only used to low-frequency content for HDR imaging, our DM can estimate the LPR with extremely low computational cost, ensuring stable visual results after multiple iterations.

The main contributions can be summarized as follows:

- We introduce LF-Diff, a straightforward and efficient DM-based method for HDR imaging. LF-Diff leverages the capabilities of diffusion models to generate informative low-frequency priors, which are then integrated into the imaging process to enhance results.
- We propose PIM and FRM in the DHRNet to fully exploit the LPR. While PIM efficiently fuses the LPR with intermediate features of DHRNet, FRM further processes the fused features to reconstruct high-quality HDR images.
- Extensive experiments demonstrate that the proposed LF-Diff method achieves SOTA performance in HDR imaging tasks, and produces visually appealing results that align with human visual perception. But consuming significantly fewer computational resources compared to other DM-based methods.

2. Related Work

Ghost-free HDR Reconstruction. Traditional methods often employ motion rejection [7, 40], motion registration [31, 36], and patch matching [9, 25] approaches to align LDR images and reconstruct high-quality results. However,

these methods heavily rely on the performance of preprocessing techniques and often face challenges when dealing with motions across large spatial extents. Subsequently, DNN-based methods have emerged as the mainstream approach, thanks to their superior nonlinear expression capability. Researchers have explored various applications of DNNs in HDR imaging, devising sophisticated network architectures and model optimization schemes, such as attention [1, 41], transformer [16, 28, 30], optical flow [11, 22], GAN [6, 14, 21], and others [5, 15, 42, 43, 47]. Nevertheless, when LDR images lack sufficient information due to motion or saturation, they often manifest artifacts not aligned with human visual perception, commonly referred to as ghosting.

Diffusion Models. Diffusion models have recently demonstrated promising results in various low-level vision tasks, including super-resolution [13, 24], HDR reconstruction [44, 46], colorization [12, 35], and deblurring [37]. DiffHDR [44] applied the conventional diffusion model paradigm to reconstruct HDR images, yielding remarkable outcomes. However, due to the high computational cost of diffusion models along with the requirement for numerous iterative steps, inference speed is limited. Despite efforts to mitigate this constraint [2, 17, 27], the overall complexity remains high, especially for high-resolution images commonly encountered in HDR scenarios. Recently, several methods [3, 23, 39] have addressed this issue by conducting diffusion modeling in the latent space. Rombach *et al.* [23] employed an autoencoder model to compress images into features equivalent to the image space. HI-Diff [3] leverages diffusion-based vector features for assisting image deblurring. However, given the unique characteristics of HDR imaging, further exploration is needed to select an appropriate latent space for HDR reconstruction tasks.

3. Preliminaries

The diffusion model introduced by Sohl-Dickstein *et al.* [26] is inspired by nonequilibrium thermodynamics. Here, we provide a brief overview of the "variance-preserving" diffusion model from [8], which encompasses diffusion processes and reverse processes.

Diffusion Process. Given a clean image distribution, the diffusion process gradually injects isotropic Gaussian noise to generate x_t according to a variance schedule $\beta_1, \dots, \beta_T \in (0, 1)$. Let $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$:

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1 - \bar{\alpha}_t)\mathbf{I}). \quad (1)$$

The closed-form for this process can be represented as:

$$x_t(x_0, \epsilon_t) = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon_t, \epsilon \in (0, \mathbf{I}), \quad (2)$$

where x_t represents the output at time t , x_0 is the initial clean image, and ϵ_t denotes isotropic Gaussian noise.

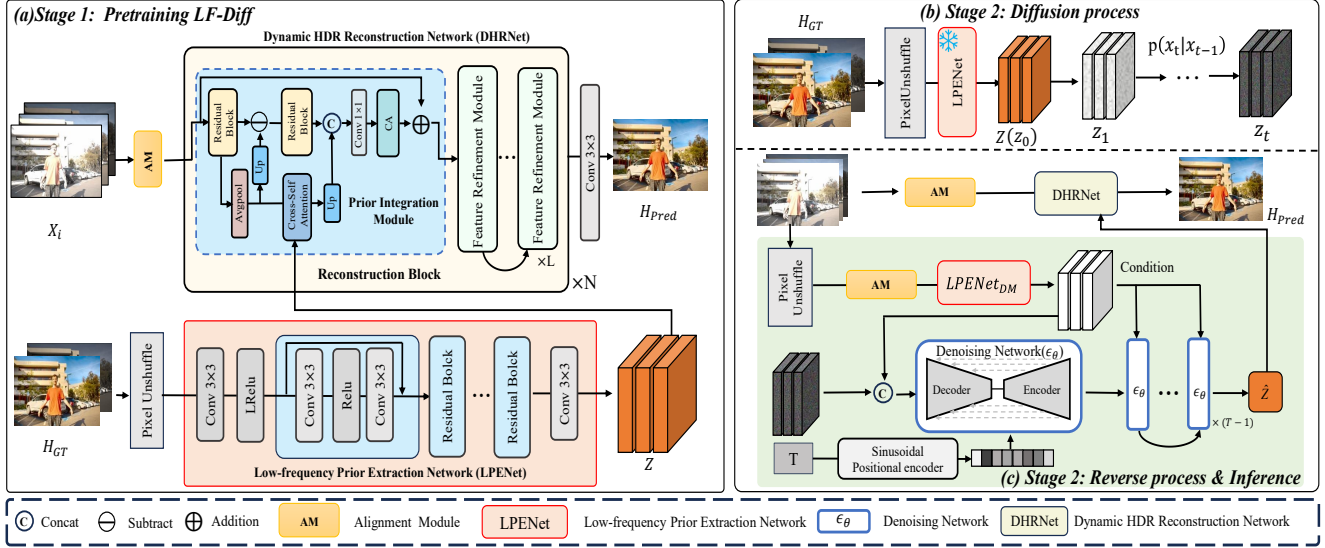


Figure 2. The proposed LF-Diff comprises DHRNet, LPRNet, and a denoising network. LF-Diff undergoes two training stages: Pretrain LF-Diff (Sec. 4.1) and DM training (Sec. 4.2). Notably, during the inference stage, we do not input the ground-truth image into LPENet_{DM} and the denoising networks. Instead, we solely utilize the reverse process of DMs.

Reverse Process. The reverse process attempts to approximate the data distribution $q(x_0)$ by a Markovian process, starting with random Gaussian noise $x_T = \mathcal{N}(x_T; \mathbf{0}, \mathbf{I})$:

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \tilde{\mu}_t(x_t, x_0), \tilde{\beta}_t \mathbf{I}), \quad (3)$$

where $\tilde{\mu}_t$ and $\tilde{\beta}_t$ are distribution parameters. Since the true reverse process Eq.(3) relies on $q(x_0)$ and is intractable, the neural network f_θ serves as the denoiser to estimate $p_\theta(x_{t-1}|x_t, t)$ instead of $q(x_{t-1}|x_t, x_0)$:

$$\begin{aligned} p_\theta(x_{t-1}|x_t, t) &= \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \\ &= q(x_{t-1}|f_\theta(x_t, t)), \end{aligned} \quad (4)$$

where μ_θ and Σ_θ are the parameters to be estimated in the reverse process. Let fixed variances ($\Sigma_\theta(x_t, t) = \tilde{\beta}_t \mathbf{I}$), the optimization objective can be defined as:

$$\mathbb{E}_{t, x_0, \epsilon_t} [\|\epsilon_t - f_\theta(\sqrt{\alpha_t}x_0 + \sqrt{1 - \alpha_t}\epsilon_t, t)\|^2]. \quad (5)$$

4. Methodology

As depicted in Fig. 2, we introduce the proposed LF-Diff framework, which consists of two training stages. During the first stage, we pre-train LF-Diff to extract accurate low-frequency prior representations from ground truths using LPENet, while DHRNet learns how to utilize the above priors to reconstruct HDR images. During the second stage, we train the DM to accurately predict the low-frequency prior representations directly from LDR images, and jointly optimize it with DHRNet to generate final high-quality HDR images.

4.1. Stage One: Pretrain LF-Diff

In the first stage, our objective is to train the LPENet how to extract accurate low-frequency prior features from ground truth and embed this representation into the DHRNet to guide the reconstruction process. As illustrated in Fig. 2 (a), we utilize the Prior Integration Module (PIM) to incorporate these prior features into the DHRNet via the cross-attention mechanism. Within DHRNet, multiple Feature Refinement Modules (FRM) are stacked to directly learn the mapping relationship from input features to the target image distribution, while PIM leverages prior features to assist in recovering the low-frequency content of the reconstructed images. Next, we elaborate on above descriptions.

Low-frequency Prior Extraction Network. The LPENet has a straightforward structure, as shown in Fig. 2 (a), comprising several residual blocks designed to extract the Low-frequency Prior Representation (LPR). Starting with the ground truth image $I_{gt} \in \mathbb{R}^{H \times W \times 3}$, we initially perform a tonemapping operation $\mathcal{T}(\cdot)$ to obtain the LDR domain ground truth image $\mathcal{T}(I_{gt})$. We concatenate these images along the channel dimension and employ the PixelUnshuffle operation to downsample, generating the input for LPENet. Subsequently, LPENet extracts the LPR $z \in \mathbb{R}^{\frac{H}{k} \times \frac{W}{k} \times 3}$ as follows:

$$z = \text{LPENet}(\text{PixelUnshuffle}(\text{Concat}(I_{gt}, \mathcal{T}(I_{gt})))), \quad (6)$$

where $\mathcal{T}(x) = \frac{\log(1+\mu x)}{\log(1+\mu)}$ denotes the tonemapping operator, $\mu = 5000$ is a predefined constant parameter that controls the degree of compression applied to the input signal, k denotes the sampling multiple of PixelUnshuffle.

Dynamic HDR Reconstruction Network. As illustrated in Fig. 2 (a), DHRNet consists of multiple stacked Reconstruction Blocks, with each block comprising one Prior Integration Module (PIM) and N Feature Refinement Module (FRM). PIM fuses the LPR with intermediate features of DHRNet, we will give a detailed description in the following part. FRM is designed based on the understanding that natural images encompass diverse frequencies, since high and low frequencies serve distinct roles in image encoding.

In Fig. 3 (a), the structure of FRM involves several blocks. Specifically, we begin by processing the features through a residual block to obtain F_{n-1} . Subsequently, we decouple these features into two components: high-frequency features F_{n-1}^{high} , which remain resolution-invariant, and low-frequency features F_{n-1}^{low} with reduced resolution, following the approach described in [18]:

$$\begin{aligned} F_{n-1}^{\text{low}} &= \text{Avgpool}(F_{n-1}, k), \\ F_{n-1}^{\text{high}} &= F_{n-1} - \text{Upsample}(F_{n-1}^{\text{low}}), \end{aligned} \quad (7)$$

where k denotes the kernel size of the avgpooling layer, *Upsample* denotes the upsampling operator. In essence, low-frequency features capture global dependencies within the input (image/features) and do not necessitate a high-resolution feature map, but require global attention. On the other hand, high-frequency features capture detailed local information, requiring a high-resolution feature map, and only can be achieved through local operators. Therefore, we employ the residual block (RB) to process F_{n-1}^{high} and utilize the self-attention (SA) from [45] to handle F_{n-1}^{low} . Subsequently, we fuse the low- and high-frequency features to preserve the initial details, resulting in the feature F'_{n-1} . This operation can be represented as follows:

$$F'_{n-1} = \text{Concat} \left[\text{RB} \left(F_{n-1}^{\text{high}} \right), \text{Upsample} \left(\text{SA} \left(F_{n-1}^{\text{low}} \right) \right) \right]. \quad (8)$$

Considering that F'_{n-1} is a concatenation of two features, a $1 \times 1 \text{ Conv}$ is employed to effectuate a reduction in the number of channels. This downsizing step is followed by a channel attention module, designed to accentuate those channels exhibiting heightened activation values.

In order to effectively merge the prior feature derived from the LPENet with the intermediate feature produced by DHRNet, we replaced the self-attention mechanism in the low-frequency branch of FRM with a carefully designed cross-attention mechanism, resulting in the Prior Integration Module (PIM). As depicted in Fig. 2 (a), a PIM is positioned before each reconstruction block. Within each PIM, cross-attention computations are performed between the prior and intermediate features to facilitate feature fusion. This module enables the aggregation of information from the prior feature into the features of DHRNet.

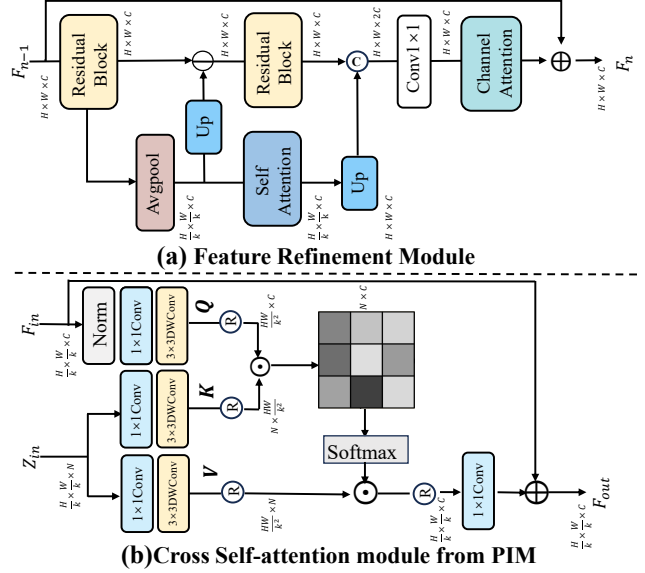


Figure 3. The architecture of (a) Feature Refinement Module (FRM) and (b) Cross-attention mechanism.

Specifically, as illustrated in Fig. 3 (b), our cross-attention module receives two types of input: the intermediate feature $\mathbf{F} \in \mathbb{R}^{\hat{H} \times \hat{W} \times \hat{C}}$ and the prior feature $z \in \mathbb{R}^{\hat{H} \times \hat{W} \times \hat{N}}$. We project \mathbf{F} into a query vector $\mathbf{Q} = W_d^Q W_c^Q \mathbf{F}$ using point-wise 1×1 convolution and depth-wise 3×3 convolution with weights W_d^Q and W_c^Q . Similarly, z is transformed into key \mathbf{K} and value \mathbf{V} through analogous operations. Next, we reshape these projections into matrices suitable for attention computation: $\hat{\mathbf{Q}} \in \mathbb{R}^{\hat{H}\hat{W} \times \hat{C}}$, $\hat{\mathbf{K}} \in \mathbb{R}^{\hat{N} \times \hat{H}\hat{W}}$, and $\hat{\mathbf{V}} \in \mathbb{R}^{\hat{H}\hat{W} \times \hat{N}}$. We then calculate a more computationally efficient attention map $A \in \mathbb{R}^{\hat{N} \times \hat{C}}$ by performing dot product between $\hat{\mathbf{Q}}$ and $\hat{\mathbf{K}}$. The process can be described as follows:

$$\hat{\mathbf{F}} = W_c \hat{\mathbf{V}} \cdot \text{Softmax}(\hat{\mathbf{K}} \cdot \hat{\mathbf{Q}}/\gamma) + \mathbf{F}, \quad (9)$$

where γ is a trainable parameter. Notably, we do not implement multi-head attention in PIM.

Training Strategy. As proposed in [11], we initially transform the provided set of input LDR images $\{L_1, L_2, \dots, L_N\}$ into their corresponding HDR versions H_i using gamma correction. Subsequently, we concatenate each L_i with its corresponding H_i along the channel axis to generate six-channel input tensors $X_i = [L_i, H_i]$. Similar to [16, 41, 44], we employ an alignment module [41] to process the input LDR images, align the features implicitly, and feed them into DHRNet. Then, the LPR feature z extracted from LPENet is injected into DHRNet via PIM. The reconstructed HDR image \hat{H} is generated by:

$$\hat{H} = \text{Conv}_{3 \times 3}(\text{DHRNet}(X_i, z)). \quad (10)$$

Following previous approaches [16, 41], we utilize both

tonemapped per-pixel loss and perceptual loss as the image reconstruction loss function \mathcal{L}_r . This dual-loss strategy optimizes both pixel-level accuracy and high-level feature representations in the generated HDR results:

$$\mathcal{L}_r = \|\mathcal{T}(H) - \mathcal{T}(\hat{H})\|_1 + \lambda \|\phi_{i,j}(\mathcal{T}(H)) - \phi_{i,j}(\mathcal{T}(\hat{H}))\|_1, \quad (11)$$

where $\phi_{i,j}(\cdot)$ signifies the j -th convolutional feature extracted from the VGG19 network after the i -th maxpooling operation, and $\lambda = 1e-2$ is a hyperparameter that balances the contribution of the each component.

4.2. Stage Two: Diffusion Models for HDR Imaging

After the above-mentioned learning procedure, we now possess LDR images and their corresponding LPR z . In the second stage (Fig. 2 (b, c)), our objective is to efficiently harness the powerful distribution estimation capability of the DMs. Specifically, we utilize the LPENet from stage one to capture the LPR as the denoising target of the DM. The DM will learn how to extract accurate LPR from the LDR inputs and perform joint optimization with DHRNet.

Diffusion Model for LRP Learning. After capturing the LPR $z \in \mathbb{R}^{\frac{H}{k} \times \frac{W}{k} \times N}$ from the pretrained LPENet, we transform the clean LPR feature z into a noisy version z_T using Eq.(2), which can be described as:

$$q(z_T | z) = \mathcal{N}(z_T; \sqrt{\bar{\alpha}_T} z, (1 - \bar{\alpha}_T) \mathbf{I}), \quad (12)$$

where T represents the total number of iterations, while $\bar{\alpha}_t$ and α denote the pre-defined variance schedule.

In the reverse process, we iteratively generate the LPR from a pure Gaussian distribution, starting from z_T and moving backward to z_0 , which utilizes the posterior distribution described in Eq. (3). Following previous work [24, 44], we utilize a neural network to estimate $p_\theta(x_{t-1}|x_t, t)$ instead of $q(x_{t-1}|x_t, x_0)$ for each step. Specifically, we first use $LPENet_{DM}$ to obtain a conditional feature $D \in \mathbb{R}^{\frac{H}{k} \times \frac{W}{k} \times N}$ from aligned LDR feature:

$$\mathbf{D} = LPENet_{DM}(AM(\text{PixelUnshuffle}(X_i))), \quad (13)$$

where $LPENet_{DM}$ maintains a similar architecture to LPENet but with a modified input dimensionality for its first convolution layer, $AM(\cdot)$ denotes alignment module from [41]. The denoising neural network predicts the noise based on both z_t and the derived conditional feature D , *i.e.*, $\epsilon_\theta(\text{Concat}(z_t, \mathbf{D}), t)$. Upon substituting this estimated noise term into the equation governing the reverse process Eq. (4), we arrive at the following sampling formula:

$$z_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(z_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(z_t, \mathbf{D}, t) \right) + \sqrt{1 - \alpha_t} \epsilon_t, \quad (14)$$

where $\epsilon_t \sim \mathcal{N}(0, \mathbf{I})$. By iteratively sampling z_t using Eq. (14) T times, we progressively reconstruct the predicted LPR representation z_0 (*i.e.* \hat{z}) $\in \mathbb{R}^{\frac{H}{k} \times \frac{W}{k} \times N}$.

Training Strategy. Traditional DMs solely learn the probability distribution by optimizing the weighted variational bound (Eq. (5)), resulting in a slight deviation between the predicted prior and the actual prior. Integrating the DM directly with DHRNet might lead to a misalignment issue, thereby hampering the overall image processing performance. To address this, we joint training the DM and DHRNet. During each training iteration, we first sample the noise sample z_t through the diffusion process (Eq. (2)). Given that our denoising neural network is lightweight, we use a reverse process for S iterations based on the DDIM strategy [27] to generate the predicted prior feature \hat{z} . This \hat{z} guides DHRNet via PIM. The loss function in the second stage consists of the reconstruction loss \mathcal{L}_r (Eq. (11)) and the diffusion loss \mathcal{L}_{diff} :

$$\mathcal{L}_{diff} = E_{z_0, t, \epsilon_t \sim \mathcal{N}(0, \mathbf{I})} \left[\|\epsilon_t - \epsilon_\theta(z_t, D, t)\|^2 \right] + \|\hat{z} - z\|_1. \quad (15)$$

4.3. Inference

In the inference stage, $LPENet_{DM}$ extracts a conditional feature \mathbf{D} from aligned LDR feature (Eq. (13)), and we randomly sample a Gaussian noise \hat{z}_T . Then, denoising neural network utilizes the \hat{z}_T and \mathbf{D} to estimate LPR after S iterations (Eq. (14)) based on DDIM [27]. After that, DHRNet exploits the LPR to restore HDR images as Eq. (10). More details about the DM training and the model inference can be found in the supplementary material.

5. Experiments

5.1. Experimental Settings

Datasets. All methods are trained using two publicly available datasets, employing identical training settings: Kalantari’s dataset [11] and Hu’s dataset [10]. Kalantari’s dataset consists of 74 samples for training and 15 for testing, all captured under authentic environmental conditions. Each sample comprises three LDR images with exposure variations of $\{-2, 0, +2\}$ or $\{-3, 0, +3\}$. In contrast, Hu’s dataset is a synthetic dataset designed to emulate sensor realism, generated through a game engine. This dataset contains images captured at three distinct exposure levels $\{-2, 0, +2\}$, with our primary focus on dynamic scene images within this collection. Following the settings outlined in [10], we allocate the initial 85 samples for training, reserving the remaining 15 for testing. Moreover, to further validate the model’s generalizability, we incorporate Sen *et al.*’s dataset [25] and Tursun *et al.*’s dataset [32] exclusively for qualitative assessment, as these datasets lack ground truths.

Evaluation Metrics. To conduct a quantitative comparison, we utilize five objective metrics: PSNR- μ , SSIM- μ , PSNR-L, SSIM-L, and HDR-VDP-2 [20]. Here, the sub-

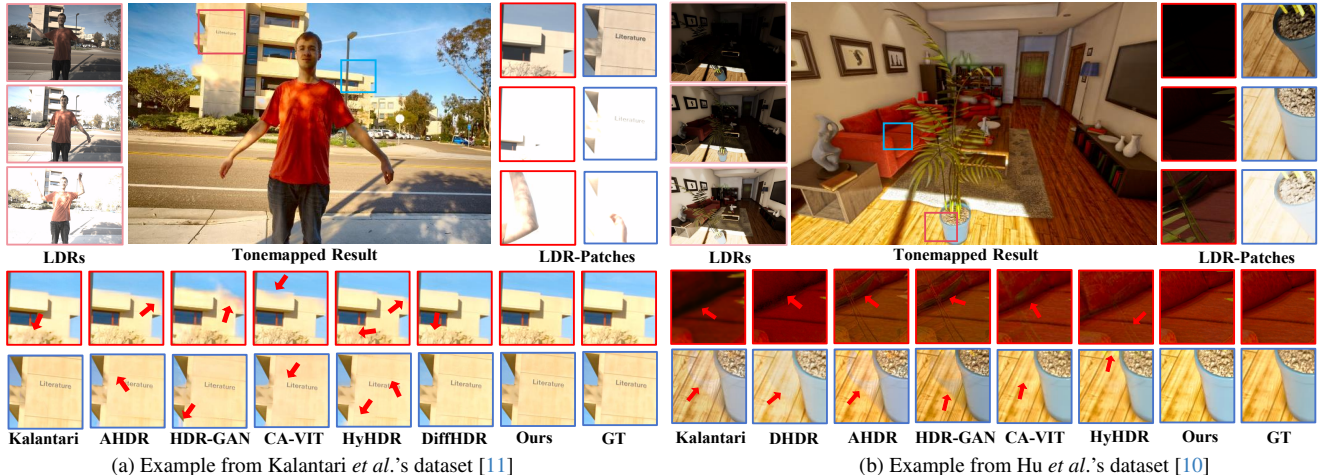


Figure 4. Visual comparisons are conducted on testing data from various datasets, focusing on zoomed-in local areas of the HDR images estimated by our method and the compared techniques. Our model demonstrates the ability to generate HDR images of superior quality.

scripts μ and L denote that these metrics are computed in the tonemapped and linear domains, respectively.

Implementation Details. Our implementation is conducted using PyTorch, and each training stage converges after 300 epochs on four NVIDIA A100 GPUs. We employ the Adam optimizer with an initial learning rate of $1e-4$, which decays by a factor of 0.1 after every 50 epochs. The training dataset is processed by cropping 128×128 patches with a stride of 64, and the batch size is set to 64. For LF-Diff, we adopt a variant of the commonly used U-Net architecture from [19] as the denoising network, with 3 levels of blocks $\{2, 2, 2\}$. During training, the time step T is set to 200, and the implicit sampling step S is set to 10 for both training and inference phases to achieve efficient restoration. In DHR-Net, the parameters $N_i \in \{L_1, L_2, L_3\}$ are set to $\{3, 3, 3\}$, and the channel C is set to 60. The number of attention heads for the self-attention branch at the same level is set to $\{6, 6, 6\}$, and the channel expansion factor in FFN is 2.66. The kernels for avgpool in PIM and FRM are set to 4 and 2, respectively. For LPENet, it comprises 4 residual blocks, with a Pixelunshuffle downsampling factor of 4 and the output channel of 3.

5.2. Comparison with the State-of-the-art Methods

In this section, we evaluate the performance of the proposed LF-Diff method and present experimental results to validate its reconstruction performance compared to state-of-the-art techniques. Specifically, we evaluate LF-Diff against two patch-based methods [9, 25], a optical flow-based method [11], five CNN-based methods [1, 15, 21, 38, 41], two ViT-based methods [16, 43], along with a diffusion-based method [44]. It's worth noting that all deep learning methods employ the same training dataset and settings for consistency.

Datasets w/ Ground Truth. The quantitative outcomes for LF-Diff on two datasets are presented in Tab. 1. Our method is compared with various state-of-the-art approaches using the testing data from [11] and [10], which includes challenging samples characterized by saturated backgrounds and foreground motions. All quantitative results are averaged across testing images. Notably, LF-Diff exhibits a significant improvement over other leading methods, exceeding the performance of the DM-based method DiffHDR [44], by 0.65 dB and 0.86 dB in PSNR- μ and PSNR- L metrics, respectively, based on Kalantari's dataset [11]. Moreover, LF-Diff showcases superior performance compared to the runner-up method, HyHDR [43], with improvements of 0.28 dB and 0.19 dB in PSNR- μ and PSNR- L metrics, respectively, on Hu's dataset [10].

As shown in Fig. 4 (a) and (b), the datasets present some challenging samples due to information loss in the LDR images. Most existing approaches struggle in these areas, producing ghosting artifacts due to large motions and occlusions. Kalantari's method [11] and DHDR [38] struggle with background motion due to error-prone alignments (*e.g.*, optical flow), resulting in undesirable ghosting. While HDR-GAN [21] exhibits noticeable ghosting artifacts around the balcony area and introduces erroneous color information, AHDR [41] aligns LDR images using convolutional spatial attention. However, it unintentionally suppresses valuable contextual information and encounters difficulties with significant motions in over/underexposed scenes. CA-ViT [16], relying on patch-based sampling, produces noticeable blocky ghosting. In contrast, aided by the DM, both Diff-HDR [44] and our LF-Diff generate HDR images aligned with human perception. Notably, our approach outperforms Diff-HDR with faster inference speeds and a smaller computational overhead.

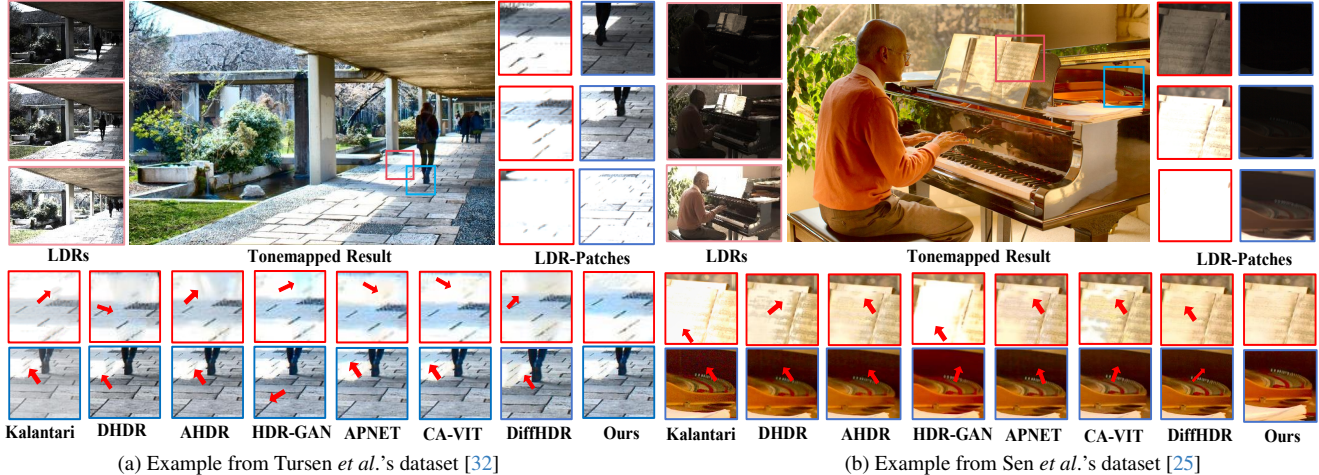


Figure 5. Visual comparisons on the datasets without ground truth.

Datasets	Models	Sen[25]	Hu[9]	Kalantari[11]	DHDR[38]	AHDR[41]	NHDRR[42]	HDRGAN[21]	APNT[1]	CA-ViT[16]	HyHDR[43]	DiffH[44]	Ours
Kalantari	PSNR- μ	40.95	32.19	42.74	41.64	43.62	42.41	43.92	43.94	44.32	44.64	44.11	44.76
	PSNR-L	38.31	30.84	41.22	40.91	41.03	41.08	41.57	41.61	42.18	42.47	41.73	42.59
	SSIM- μ	0.9805	0.9716	0.9877	0.9869	0.9900	0.9887	0.9905	0.9898	0.9916	0.9915	0.9911	0.9919
	SSIM-L	0.9726	0.9506	0.9848	0.9858	0.9862	0.9861	0.9865	0.9879	0.9884	0.9894	0.9885	0.9906
	HDR-VDP-2	55.72	55.25	60.51	60.50	62.30	61.21	65.45	64.05	66.03	66.05	65.52	66.54
Datasets	Models	Sen[25]	Hu[9]	Kalantari[11]	DHDR[38]	AHDR[41]	NHDRR[42]	HDRGAN[21]	APNT[1]	ADNet[15]	CA-ViT[16]	HyHDR[43]	Ours
Hu	PSNR- μ	31.48	36.56	41.60	41.13	45.76	45.15	45.86	46.41	46.79	48.10	48.46	48.74
	PSNR-L	33.58	36.94	43.76	41.20	49.22	48.75	49.14	47.97	50.38	51.17	51.91	52.10
	SSIM- μ	0.9531	0.9824	0.9914	0.9870	0.9956	0.9956	0.9945	0.9953	0.9908	0.9947	0.9959	0.9968
	SSIM-L	0.9634	0.9877	0.9938	0.9941	0.9980	0.9981	0.9989	0.9986	0.9987	0.9989	0.9991	0.9993
	HDR-VDP-2	66.39	67.58	72.94	70.82	75.04	74.86	75.19	73.06	76.21	77.12	77.24	77.35

Table 1. The evaluation results on Kalantari’s [11] and Hu’s [10] datasets. The best results are highlighted in **Bold**.

Datasets w/o Ground Truth. To assess the generalization capability of our proposed HDR imaging method, we evaluated the performance of the model trained on Kalantari *et al.*’s Dataset [11] by testing it on datasets from Sen *et al.* [25] and Tursun *et al.* [33], which lack ground truths. In Fig. 5 (a), numerous current methods encounter difficulties in restoring recover the large saturated region and large motion. Conversely, in Fig. 5 (b), while complete elimination of ghosting remains elusive for any method, our approach notably enhances image sharpness and detail compared to preceding techniques.

Computational Budgets. We also conducted comparisons regarding model parameters and inference times with previous approaches. As illustrated in Table 2, patch-based methods [9, 25] exhibit significantly longer inference times due to their CPU-based computation. Kalantari [11] requires considerable time for initial optical flow alignment. NHDRNet [42], employing a U-shaped network architecture, demonstrates shorter inference times but substantially higher parameter counts compared to other methods. CA-ViT [16] has a large number of standard transformer blocks leading to high computational cost despite fewer param-

eters. DiffHDR [44] has high inference time and parameter count due to reconstructing HDR from pure noise. In comparison, our method effectively leverages the powerful distribution estimation capability of DM with orders of magnitude fewer parameters and computations.

5.3. Ablation Study

In this section, we investigate the impact of various designs within our proposed method. All experiments are conducted using the Kalantari’s dataset [11].

Effects of Diffusion Prior. As shown in Tab. 3, we establish a regression-based baseline model without LPR generation. In this configuration, the PIM in DHRNet is replaced with FRM, and HDR images are reconstructed through a regression-based approach. LF-Diff_{s1} is a pre-trained model in the first stage utilizing ground-truth images to provide LPR. Compared to the baseline, this brings a 8.7dB PSNR-L improvement, demonstrating that an accurate compact prior can greatly enhance results. Variant LF-Diff_{s2} w/o DM does not use the DM, but instead directly learns the LPR using LPENet. This achieves a 0.16dB PSNR gain over the baseline. When learning the LPR based

Methods	Sen [25]	Hu [9]	Kalantari [11]	AHDR [41]	NHDRR [42]	HDRGAN [21]	APNT [1]	CA-ViT [16]	DiffHDR [44]	Ours
Environment	(CPU)	(CPU)	(CPU+GPU)	(GPU)	(GPU)	(GPU)	(GPU)	(GPU)	(GPU)	(GPU)
Times(s)	61.81s	79.77s	29.14s	0.30s	0.31s	0.29s	2.62s	5.34s	7.53s	0.72s
Para.(M)	-	-	0.30M	1.24M	38.10M	2.56M	37.98M	1.22M	74.99M	7.48M

Table 2. Average runtime performance of various methods on the testing set with dimensions of 1000×1500 .

on the DM, LF-Diff_{s2} w/ DM further improves PSNR by 0.24dB compared to variant LF-Diff_{s2} w/o DM, This indicates that the DM demonstrates superior capability in accurately estimating the distribution for LPR prediction. Additionally, compared to DiffHDR, which requires over 70M Params for estimating complete images, LF-Diff only adds 2.39M Params over the baseline. Moreover, we present visual comparisons of the baseline (without the prior feature) and LF-Diff variants in Fig. 6. It can be observed that performing DM in a compact latent space to predict LPR effectively mitigates ghosting issues.

Effects of Joint Training Strategy. We conducted an ablation study on the joint training strategy. Under this strategy, only the DM is optimized in the second stage, referred to as Split-Training. Specifically, we first utilize the pre-trained LPENet to generate the prior feature z from the ground truth, and then apply the training objective defined in Eq. (15) to independently train the DM. Subsequently, the trained DM is directly integrated with the DHRNet for evaluation. It’s worth noting that the DHRNet in the second stage utilizes pre-trained weights from the stage one without additional training. It is evident that LF-Diff significantly outperforms Split-Training by 1.51 dB in terms of PSNR value. This comparison underscores the effectiveness of joint training strategy over Split-Training.

Sampling Step. In addition to the lightweight DM structure and compact feature dimensions, our method achieves greater computational efficiency by employing a smaller sampling step, denoted as S , in the reverse process based on DDIM [27]. The performance of our method is demonstrated with different values of S ranging from 5 to 100, as presented in Tab. 4. Variations in sampling steps, when S is set to a value smaller than the training configuration, a noticeable performance degradation occurs. On the other hand, when S is larger than the training setting, the impact on performance is minimal, particularly with the SSIM metric remaining consistently unchanged. While a larger sampling step is known to enhance the visual quality of images in diffusion-based methods [24, 44], in our case, it primarily increases inference time. This observation suggests that larger sampling steps, such as $S = 1000$ in DiffHDR [44] and SR3 [24], may not be essential for diffusion-based HDR reconstruction and other related low-level vision tasks when using our proposed framework.

Method	GT	DM	Joint Training	Para.	PSNR- μ	PSNR-L
Baseline	✗	✗	✗	5.09M	44.23	42.19
LF-Diff _{s1}	✓	✗	✗	5.24M	50.42	50.89
Split training	✗	✓	✗	7.48M	41.93	39.30
LF-Diff _{s2} w/o DM	✗	✗	✓	5.85M	44.45	42.35
LF-Diff _{s2} w/ DM	✗	✓	✓	7.48M	44.76	42.59

Table 3. Ablation studies of Diffusion Prior and Joint Training.

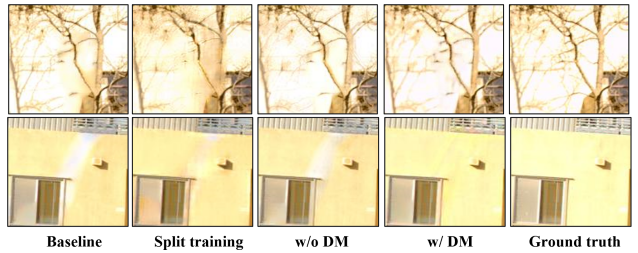


Figure 6. Qualitative results of our ablation study.

Sampling step	PSNR- μ	PSNR-L	SSIM- μ	SSIM-L	Time(s)
S=5	39.76	41.33	0.9819	0.9878	0.049
S=10	44.76	42.59	0.9919	0.9906	0.106
S=20	44.73	42.54	0.9919	0.9906	0.208
S=50	44.66	42.45	0.9919	0.9906	0.479
S=100	44.65	42.44	0.9919	0.9906	0.971

Table 4. Ablation studies of various settings on the sampling step during the reverse process. Time(s) only denotes the time expenditure of the DM for the corresponding setting.

6. Conclusion

The potential of diffusion models in HDR deghosting has shown promising results, particularly in achieving visually perceptible outcomes. Different from image synthesis which generates each pixel from scratch, HDR imaging provides several LDR images as reference. Thus, it is inefficient to reconstruct the full HDR image starting from pure Gaussian noise. In this paper, we propose an efficient diffusion model called LF-Diff, consisting of LPENet, DHRNet, and a denoising network. Specifically, we apply the DM in a compact latent space to predict low-frequency priors of HDR images. These prior features provide explicit guidance for the image reconstruction process, thereby enhancing the details of the reconstructed HDR images. Compared to traditional DM-based methods, LF-Diff achieves accurate estimations and reduces artifacts in reconstructed images with much lower computational cost.

References

- [1] Jie Chen, Zaifeng Yang, Tsz Nam Chan, Hui Li, Junhui Hou, and Lap-Pui Chau. Attention-guided progressive neural texture fusion for high dynamic range image restoration. *IEEE Transactions on Image Processing*, 31:2661–2672, 2022. 2, 6, 7, 8
- [2] Nanxin Chen, Yu Zhang, Heiga Zen, Ron J Weiss, Mohammad Norouzi, and William Chan. Wavegrad: Estimating gradients for waveform generation. In *International Conference on Learning Representations*, 2020. 2
- [3] Zheng Chen, Yulun Zhang, Ding Liu, Jinjin Gu, Linghe Kong, Xin Yuan, et al. Hierarchical integration diffusion model for realistic image deblurring. *Advances in Neural Information Processing Systems*, 36, 2024. 2
- [4] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. In *Advances in Neural Information Processing Systems*, pages 8780–8794. Curran Associates, Inc., 2021. 1
- [5] Li Fang, Qian Wang, and Long Ye. Glnet: light field angular superresolution with arbitrary interpolation rates. *Visual Intelligence*, 2(1):6, 2024. 2
- [6] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020. 2
- [7] YongSeok Heo, KyoungMu Lee, SangUk Lee, Youngsu Moon, and Joonhyuk Cha. Ghost-free high dynamic range imaging. In *IEEE Conference on Asian Conference on Computer Vision (ACCV)*, pages 486–500, 2011. 2
- [8] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, pages 6840–6851. Curran Associates, Inc., 2020. 1, 2
- [9] Jun Hu, O. Gallo, K. Pulli, and Xiaobai Sun. HDR deghosting: How to deal with saturation? In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1163–1170, 2013. 2, 6, 7, 8
- [10] Jinhan Hu, Gyeongmin Choe, Zeeshan Nadir, Osama Nabil, Seok-Jun Lee, Hamid Sheikh, Youngjun Yoo, and Michael Polley. Sensor-realistic synthetic data engine for multi-frame high dynamic range photography. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 516–517, 2020. 5, 6, 7
- [11] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep high dynamic range imaging of dynamic scenes. *ACM Transactions on Graphics*, 36(4):1–12, 2017. 1, 2, 4, 5, 6, 7, 8
- [12] Bahjat Kwar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Advances in Neural Information Processing Systems*, 35:23593–23606, 2022. 2
- [13] Haoying Li, Yifan Yang, Meng Chang, Shiqi Chen, Huajun Feng, Zhihai Xu, Qi Li, and Yueting Chen. Srdiff: Single image super-resolution with diffusion probabilistic models. *Neurocomputing*, 479:47–59, 2022. 2
- [14] Ru Li, Chuan Wang, Jue Wang, Guanghui Liu, Heng-Yu Zhang, Bing Zeng, and Shuaicheng Liu. Uphdr-gan: Generative adversarial network for high dynamic range imaging with unpaired data. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(11):7532–7546, 2022. 1, 2
- [15] Zhen Liu, Wenjie Lin, Xinpeng Li, Qing Rao, Ting Jiang, Mingyan Han, Haoqiang Fan, Jian Sun, and Shuaicheng Liu. Adnet: Attention-guided deformable convolutional network for high dynamic range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 463–470, 2021. 2, 6, 7
- [16] Zhen Liu, Yinglong Wang, Bing Zeng, and Shuaicheng Liu. Ghost-free high dynamic range imaging with context-aware transformer. pages 344–360, 2022. 1, 2, 4, 6, 7, 8
- [17] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. *Advances in Neural Information Processing Systems*, 35:5775–5787, 2022. 2
- [18] Zhisheng Lu, Juncheng Li, Hong Liu, Chaoyan Huang, Lintin Zhang, and Tiejong Zeng. Transformer for single image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 457–466, 2022. 4
- [19] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Refusion: Enabling large-size realistic image restoration with latent-space diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1680–1691, 2023. 6
- [20] Rafat Mantiuk, Kil Joong Kim, Allan G. Rempel, and Wolfgang Heidrich. HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions. In *ACM Siggraph*, pages 1–14, 2011. 5
- [21] Yuzhen Niu, Jianbin Wu, Wenxi Liu, Wenzhong Guo, and Rynson WH Lau. Hdr-gan: Hdr image reconstruction from multi-exposed ldr images with large motions. *IEEE Transactions on Image Processing*, 30:3885–3896, 2021. 1, 2, 6, 7, 8
- [22] Feiyue Peng, Maojun Zhang, Shiming Lai, Hanlin Tan, and Shen Yan. Deep hdr reconstruction of dynamic scenes. In *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, pages 347–351. IEEE, 2018. 2
- [23] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022. 2
- [24] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 1, 2, 5, 8
- [25] Pradeep Sen, Nima Khademi Kalantari, Maziar Yaesoubi, Soheil Darabi, Dan B Goldman, and Eli Shechtman. Robust patch-based hdr reconstruction of dynamic scenes. *ACM Trans. Graph.*, 31(6):203–1, 2012. 2, 5, 6, 7, 8
- [26] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning*, pages 2256–2265, 2015. 1, 2

- [27] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. [2](#), [5](#), [8](#)
- [28] Jou Won Song, Ye-In Park, Kyeongbo Kong, Jaeho Kwak, and Suk-Ju Kang. Selective transhdr: Transformer-based selective hdr imaging using ghost region mask. In *Computer Vision—ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVII*, pages 288–304. Springer, 2022. [1](#), [2](#)
- [29] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020. [1](#)
- [30] Steven Tel, Zongwei Wu, Yulun Zhang, Barthélemy Heyrman, Cédric Demonceaux, Radu Timofte, and Dominique Ginhac. Alignment-free hdr deghosting with semantics consistent transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12836–12845, 2023. [2](#)
- [31] Anna Tomaszewska and Radoslaw Mantiuk. Image registration for multi-exposure high dynamic range image acquisition. In *International Conference in Central Europe on Computer Graphics and Visualization, WSCG’07, 2007*. [2](#)
- [32] Okan Tarhan Tursun, Ahmet Oğuz Akyüz, Aykut Erdem, and Erkut Erdem. An objective deghosting quality metric for HDR images. *Comput. Graph. Forum*, 35(2):139–152, 2016. [5](#), [7](#)
- [33] Okan Tarhan Tursun, Ahmet Oğuz Akyüz, Aykut Erdem, and Erkut Erdem. An objective deghosting quality metric for hdr images. In *Computer Graphics Forum*, pages 139–152. Wiley Online Library, 2016. [7](#)
- [34] Haohan Wang, Xindi Wu, Zeyi Huang, and Eric P Xing. High-frequency component helps explain the generalization of convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8684–8694, 2020. [2](#)
- [35] Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *The Eleventh International Conference on Learning Representations*, 2022. [2](#)
- [36] Greg Ward. Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. *Journal of Graphics Tools*, 8, 2012. [2](#)
- [37] Jay Whang, Mauricio Delbracio, Hossein Talebi, Chitwan Saharia, Alexandros G Dimakis, and Peyman Milanfar. Deblurring via stochastic refinement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16293–16303, 2022. [2](#)
- [38] Shangzhe Wu, Jiarui Xu, Yu-Wing Tai, and Chi-Keung Tang. Deep high dynamic range imaging with large foreground motions. In *European Conference on Computer Vision (ECCV)*, 2018. [6](#), [7](#)
- [39] Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc Van Gool. Diffir: Efficient diffusion model for image restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13095–13105, 2023. [2](#)
- [40] Qingsen Yan, Jinqiu Sun, Haisen Li, Yu Zhu, and Yanning Zhang. High dynamic range imaging by sparse representation. *Neurocomputing*, 269:160–169, 2017. [2](#)
- [41] Qingsen Yan, Dong Gong, Qinfeng Shi, Anton van den Hengel, Chunhua Shen, Ian Reid, and Yanning Zhang. Attention-guided network for ghost-free high dynamic range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1751–1760, 2019. [1](#), [2](#), [4](#), [5](#), [6](#), [7](#), [8](#)
- [42] Qingsen Yan, Lei Zhang, Yu Liu, Yu Zhu, Jinqiu Sun, Qinfeng Shi, and Yanning Zhang. Deep hdr imaging via a non-local network. *IEEE Transactions on Image Processing*, 29: 4308–4322, 2020. [2](#), [7](#), [8](#)
- [43] Qingsen Yan, Weiye Chen, Song Zhang, Yu Zhu, Jinqiu Sun, and Yanning Zhang. A unified hdr imaging method with pixel and patch level. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22211–22220, 2023. [2](#), [6](#), [7](#)
- [44] Qingsen Yan, Tao Hu, Yuan Sun, Hao Tang, Yu Zhu, Wei Dong, Luc Van Gool, and Yanning Zhang. Towards high-quality hdr deghosting with conditional diffusion models. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1, 2023. [1](#), [2](#), [4](#), [5](#), [6](#), [7](#), [8](#)
- [45] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022. [4](#)
- [46] Xiang Zhang, Tao Hu, Jiashuang He, and Qingsen Yan. Efficient content reconstruction for high dynamic range imaging. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7660–7664, 2024. [1](#), [2](#)
- [47] Xiang Zhang, Qiang Zhu, Tao Hu, and Qingsen Yan. Eifffhdr: An efficient network for multi-exposure high dynamic range imaging. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6560–6564, 2024. [1](#), [2](#)