CONSTRUCTION THEN RECONSTRUCTION: CONTENT-ADAPTIVE ONE-STEP FEATURE RECONSTRUCTION FOR HDR DEGHOSTING

Anonymous authors
Paper under double-blind review

ABSTRACT

HDR deghosting reconstructs a high dynamic range image from multiple LDR images captured at different exposures. However, scene motion, severe exposure variations, and local saturation often cause misalignment and missing content, making HDR reconstruction challenging. To address this, we propose ***C**onstruction then **R**econstruction for **HDR** Deghosting (**CRHDR**)*, which first constructs a target representation for high-quality HDR reconstruction and then reconstructs it from incomplete multi-exposure observations. Specifically, we use optical flow and valid region masks to selectively align and fuse the input exposures, producing a coarse HDR image, which is then paired with a high-quality HDR image to construct target features for reconstruction. Next, we introduce a content-adaptive refinement (CAR) module that selects different processing paths according to local content, allowing high-quality target features to be incorporated only where the coarse representation is insufficient. Finally, we perform one-step diffusion in feature space, directly mapping coarse features to high-quality features for HDR reconstruction. Extensive experiments on three public real-world benchmarks show that our method significantly outperforms existing state-of-the-art methods in both quantitative metrics and visual quality, while making inference up to 40$\times$ faster. Our code and models will be made publicly available.

1 INTRODUCTION

Natural scenes contain a wide range of illumination perceived by human eyes, while digital cameras are limited by the dynamic range of imaging sensors (Fossum & Hondongwa, 2014). For example, in a sunset scene, the sky can be much brighter than the surrounding landscape, causing either overexposed highlights or underexposed regions in a single shot. A common solution is exposure bracketing, which captures a sequence of low dynamic range (LDR) images at different exposures and merges them into a high dynamic range (HDR) image (Debevec & Malik, 1997; 2008; Mertens et al., 2009). When the input images are well aligned, existing fusion methods can already produce high-quality HDR results (Debevec & Malik, 1997; Zhang & Cham, 2012; Ma et al., 2017; Ram Prabhakar et al., 2017; Ma et al., 2020). However, camera shake and object motion introduce misalignment among the LDRs, leading to severe ghosting artifacts in the reconstructed HDR image.

Traditional HDR deghosting methods mainly handle such misalignment by aligning the input images before fusion (Bogoni, 2000; Hu et al., 2013; Kang et al., 2003) or rejecting inconsistent pixels in misaligned regions (Gallo et al., 2009; Pece & Kautz, 2010; Jacobs et al., 2008; Khan et al., 2006). However, accurate alignment becomes difficult in the presence of large motion, occlusion, and severe saturation, especially under exposure differences. Meanwhile, rejecting unreliable pixels may lead to insufficient information in dynamic regions, making it difficult to recover complete scene content.

As for deep learning based multi-exposure HDR methods, main solutions can be summarized into two paradigms. The first follows an alignment-and-fusion pipeline (Kalantari & Ramamoorthi, 2017; Prabhakar et al., 2019; Wu et al., 2018; Prabhakar et al., 2020), where cross-exposure motion is first estimated and the aligned inputs are then fused to reconstruct the HDR image. Several methods further design specialized motion estimation networks to improve cross-exposure alignment (Catley-Chandar et al., 2022; Yan et al., 2023a; Chung & Cho, 2023), but accurate motion

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

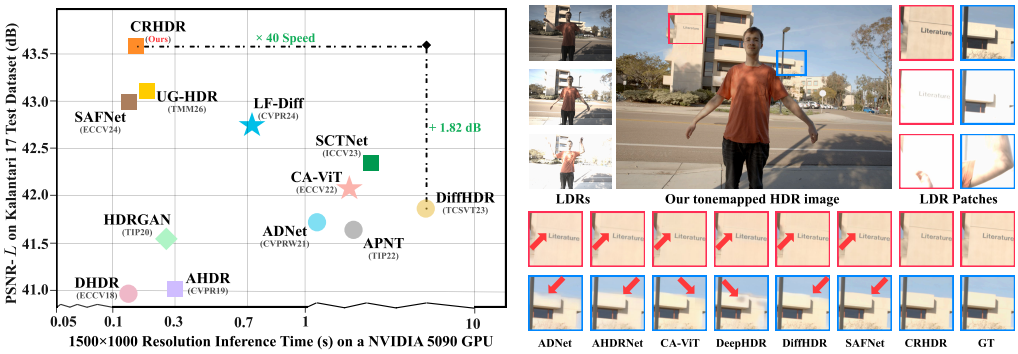


Figure 1: Comparison on Kalantari 17 test dataset (Kalantari & Ramamoorthi, 2017). Our method achieves state-of-the-art HDR imaging accuracy while with fast inference speed.

estimation under severe saturation and occlusion remains challenging. The second paradigm avoids explicit alignment and instead adopts end-to-end networks with attention-based feature interaction to fuse information across exposures (Yan et al., 2019; Chen et al., 2022; Yan et al., 2020; Song et al., 2022; Liu et al., 2022; Tel et al., 2023). Although effective, complex attention interactions often introduce considerable computational overhead. More recently, diffusion models (Hu et al., 2024; Yan et al., 2023b; Zhang et al., 2024a) have shown promising performance in HDR deghosting, benefiting from their strong capability to recover missing content in challenging scenes. Despite this progress, existing methods often require multi-step reconstruction of the entire HDR image or rely heavily on high-quality target information. Such designs lead to long reconstruction processes and make the final results more sensitive to prediction errors. These limitations raise a natural question: *can we avoid repeated full-image reconstruction and reduce the dependence on such target information, while maintaining high reconstruction quality?*

Motivated by these observations, we formulate HDR deghosting as a Construction then Reconstruction process (CRHDR). In the construction stage, optical flow and valid region masks are used to selectively align and fuse input exposures, while high-quality HDR images are employed to construct the target features. To avoid uniformly relying on these high-quality features, we design a content-adaptive refinement (CAR) that selects processing paths for local patches according to their content, so that reconstructed information is used only when necessary. In the reconstruction stage, instead of repeatedly denoising the entire HDR image, we use one-step diffusion reconstruction in feature space to transform coarse HDR representations toward the target features. With these designs, CRHDR achieves a favorable balance between quality and inference efficiency. As shown in the left of Fig. 1, it surpasses state-of-the-art approaches in PSNR-L on the Kalantari 17 test dataset while remaining highly efficient. Compared with DiffHDR (Yan et al., 2023b), which also employs diffusion models for HDR reconstruction, our method improves PSNR-L by about 1.82 dB while achieving approximately 40$\times$ faster inference. Visual results in the right of Fig. 1 further show that our method better suppresses ghosting artifacts and preserves fine structures in challenging regions.

Our contributions: **(1)** We propose CRHDR, which formulates HDR deghosting into a target construction and reconstruction process. **(2)** The content-adaptive refinement performs patch-wise routing according to local content, allowing high-quality features to be integrated only in necessary regions. **(3)** One-step diffusion is designed in feature space, directly transforming coarse features toward high-quality target features. **(4)** Extensive experiments on real-world benchmarks demonstrate that CRHDR achieves superior reconstruction quality while providing up to 40$\times$ faster inference.

2 Related Work

Traditional HDR Deghosting. Traditional multi-exposure HDR methods handle motion through pixel rejection or image registration. Pixel rejection methods identify inconsistent regions across exposures and exclude unreliable pixels before fusion (Pece & Kautz, 2010; Jacobs et al., 2008; Khan et al., 2006; Lee et al., 2014; Oh et al., 2015). Although effective for moderate motion, discarding misaligned pixels may remove useful textures and lead to missing content in dynamic regions. Another line of work aligns non-reference exposures to the reference frame using optical flow or patch-based registration (Bogoni, 2000; Kang et al., 2003; Sen et al., 2012; Hu et al., 2013). However, accurate registration remains difficult under large motion, occlusion, and severe saturation.

Deep Learning based HDR Deghosting. Early learning based HDR deghosting methods follow the alignment and fusion pipeline. Kalantari (Kalantari & Ramamoorthi, 2017) align multi-exposure inputs using optical flow and reconstruct HDR images with a convolutional network, while DeepHDR (Wu et al., 2018) adopt image-level alignment before CNN-based fusion. Subsequent methods improve cross-exposure alignment using specialized motion estimation modules (Catley-Chandar et al., 2022; Chung & Cho, 2023; Kong et al., 2024) or feature-level alignment mechanisms (Liu et al., 2021; Yan et al., 2023a). Despite improved performance, accurate alignment under severe saturation and large motion remains challenging. To avoid explicit alignment, another group of methods performs feature interaction and fusion across exposures. NHDRRNet (Yan et al., 2020) introduces a non-local attention for HDR fusion. PSFNet (Ye et al., 2021) progressively selects useful regions across LDR exposures. STHDR (Song et al., 2022) selectively employs convolution and transformer branches for different regions. CA-ViT (Liu et al., 2022) integrates local context modeling with global transformer attention. SCTNet (Tel et al., 2023) combines spatial and channel attention to improve semantic consistency. These methods achieve strong reconstruction quality, but complex attention interactions introduce considerable computational overhead.

Diffusion based HDR Deghosting. Diffusion models have recently shown strong reconstruction ability in various low-level vision tasks (Li et al., 2022; Saharia et al., 2022; Wang et al., 2022; Kawar et al., 2022; Whang et al., 2022; Zhang et al., 2024a; Xia et al., 2023; Rombach et al., 2022). Their application to HDR imaging has also received increasing attention. DiffHDR (Yan et al., 2023b) directly applies the conventional diffusion paradigm to reconstruct HDR images. Although effective, reconstructing the entire HDR image through iterative denoising requires considerable computation and leads to slow inference, especially for high-resolution inputs. LF-Diff (Hu et al., 2024) alleviates this problem by performing diffusion in a compact latent space, where low-frequency priors are generated and then incorporated into a regression-based HDR reconstruction network. However, the method still relies on iterative diffusion to estimate the target, and the final reconstruction is closely related to the quality of the generated prior. In contrast, our method selectively controls the use of high-quality features according to local content and performs one-step reconstruction from coarse HDR features to target features, thus providing a more efficient and flexible reconstruction process.

3 METHOD

As shown in Fig. 2, CRHDR follows the Construction then Reconstruction process. In the first stage (Sec. 3.2), we selectively align and fuse the LDRs to obtain a coarse HDR image, while constructing coarse and target feature priors for content-adaptive refinement (CAR, Sec. 3.3). In the second stage (Sec. 3.4), a one-step diffusion model is trained in feature space to reconstruct the target prior from the coarse prior. During inference, the reconstructed features are used for final HDR reconstruction.

3.1 PRELIMINARIES

Diffusion models (Ho et al., 2020) define a forward process that gradually perturbs clean data with Gaussian noise and a reverse process that progressively recovers the data distribution.

Diffusion Process. Given a clean sample x_0, the forward process adds Gaussian noise according to a variance schedule $\beta_1, \ldots, \beta_T \in (0, 1)$. Let $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$. The distribution of the noisy sample x_t at timestep t is given by

$$q(x_t \mid x_0) = \mathcal{N}\left(x_t; \sqrt{\bar{\alpha}_t} x_0, (1 - \bar{\alpha}_t)\mathbf{I}\right). \tag{1}$$

Using the reparameterization form, x_t can be directly sampled as

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t}\boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{2}$$

where $\boldsymbol{\epsilon}_t$ denotes the Gaussian noise injected at timestep t.

Reverse Process. Starting from $x_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, the reverse process progressively removes noise to recover a clean sample. The posterior conditioned on x_0 is expressed as

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

where $\tilde{\mu}_t$ and $\tilde{\beta}_t$ denote the posterior mean and variance, respectively.

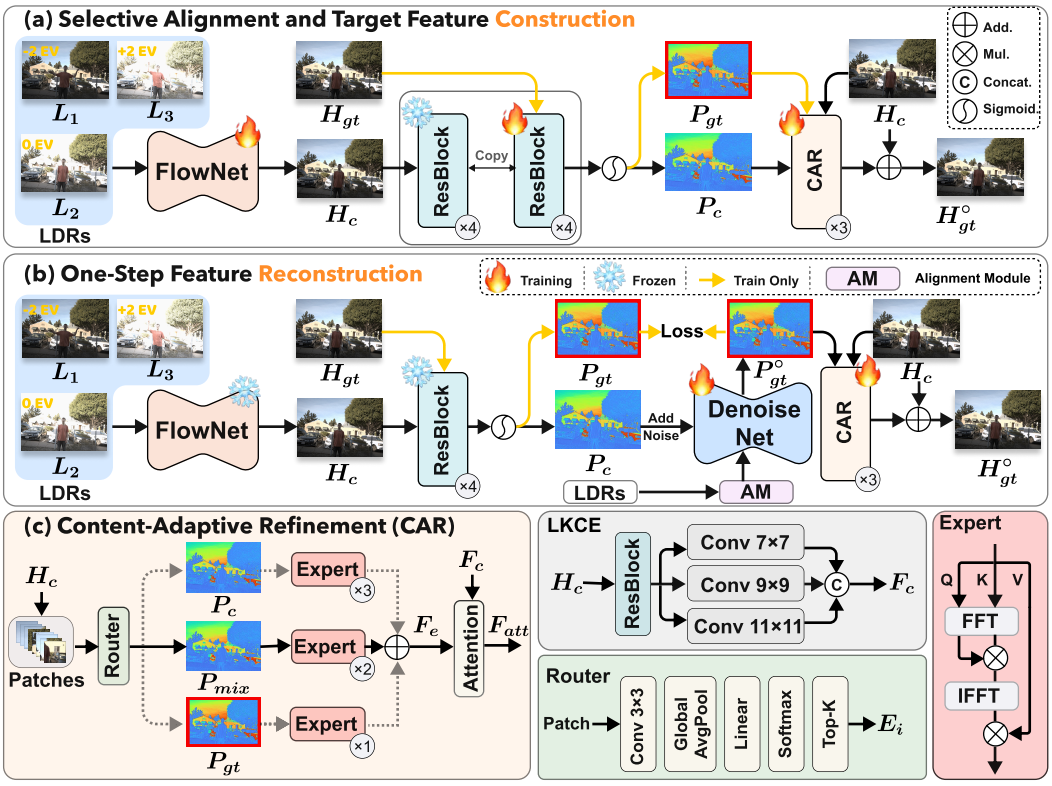


Figure 2: Overview of CRHDR. **(a)** Selective alignment and target feature construction. The LDR inputs are aligned by FlowNet (Kong et al., 2024) and merged into a coarse HDR image H_c, while H_c and H_{gt} are used to construct the coarse and target priors, respectively. **(b)** One-step feature reconstruction. DenoiseNet reconstructs the target prior P_{gt} from P_c in a single diffusion step, conditioned on the multi-exposure information from AM (Yan et al., 2019). **(c)** Content-Adaptive Refinement (CAR), which selectively exploits priors of different quality for HDR refinement.

Since x_0 is unavailable during inference, a denoising network $\boldsymbol{\epsilon}_\theta$ is trained to predict the noise from x_t. Following DDPM (Ho et al., 2020), the noise prediction objective is

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{t,x_0,\boldsymbol{\epsilon}_t}\left[\|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(x_t, t)\|_2^2\right]. \tag{4}$$

3.2 Selective Alignment and Target Feature Construction

Given three input images L_1, L_2, and L_3 at different exposures, we take the medium exposure image L_2 as the reference. Following SAFNet (Kong et al., 2024), FlowNet estimates cross exposure optical flow and valid region masks to align the non-reference frames, which are then selectively merged with the reference frame to produce a coarse HDR image H_c.

Although H_c integrates reliable information from LDRs, it remains incomplete in challenging regions. We therefore construct a high quality target representation from ground truth H_{gt} for accurate HDR reconstruction. For both H_c and H_{gt}, the linear HDR image is concatenated with its tonemapped counterpart. They are processed by four shared ResBlocks to construct the coarse prior P_c and target prior P_{gt}. As shown in Fig. 2 (a), gradients are stopped along the H_c branch, while the H_{gt} branch optimizes the shared ResBlocks. We denote this prior construction process as $\Phi(\cdot)$:

$$\boldsymbol{\Phi}(H) = \sigma(\text{ResBlocks}(\text{Concat}(H, \mathcal{T}(H)))). \tag{5}$$

The target and coarse priors are then obtained as:

$$P_{gt} = \boldsymbol{\Phi}(H_{gt}), \qquad P_c = \text{stopgrad}(\boldsymbol{\Phi}(H_c)). \tag{6}$$

where $\mathcal{T}(H) = \frac{\log(1+\mu H)}{\log(1+\mu)}$ denotes the tonemapping function with $\mu = 5000$, and $\sigma(\cdot)$ denotes the sigmoid function. P_c represents the coarse multi-exposure information, while P_{gt} provides the high quality target representation with richer structural details. These two priors define a compact feature space supporting subsequent reconstruction and Content-Adaptive Refinement.

3.3 CONTENT-ADAPTIVE REFINEMENT

Given coarse prior P_c and target prior P_{gt}, we introduce Content-Adaptive Refinement (Fig. 2(c)) to exploit priors of different quality. Since P_c and P_{gt} represent two distinct quality levels, an intermediate prior P_{mix} is constructed to form a continuous hierarchy of prior representations:

$$P_{mix} = \boldsymbol{\alpha} \odot P_{gt} + \boldsymbol{\beta} \odot P_c. \tag{7}$$

The weights $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are predicted from the concatenated priors $[P_{gt}, P_c]$ through a convolution. In this way, P_c, P_{mix}, and P_{gt} provide three priors with different quality levels for expert processing.

To select processing paths according to content, we divide H_c into local patches and route each patch independently. As shown in Fig. 2, each patch is processed by a convolution, global average pooling, and a linear layer to obtain the routing scores. Softmax is then applied to normalize the scores, followed by Top-K selection to determine the experts E_i. The priors P_c, P_{mix}, and P_{gt} are refined using the same Expert module but with different stacking depths. Specifically, P_c, P_{mix}, and P_{gt} are assigned three, two, and one Expert blocks, respectively. This hierarchical design applies deeper refinement to lower quality priors, providing greater processing capacity where more refinement is required. Each Expert employs a feature interaction module based on FFT (Kong et al., 2023). The selected expert features are fused into F_e, and we introduce a Large Kernel Context Extractor (LKCE, Fig. 2) to refine H_c. Specifically, LKCE employs a ResBlock followed by parallel 7×7, 9×9, and 11×11 convolutions to capture contextual information across different receptive fields. Guided by F_c extracted from H_c, relevant information from F_e is aggregated through an attention module following (Zamir et al., 2022), producing the refined feature F_{att} for HDR reconstruction.

3.4 ONE-STEP FEATURE RECONSTRUCTION

Existing diffusion based HDR methods rely on iterative denoising. DiffHDR (Yan et al., 2023b) reconstructs a HDR image through multiple reverse steps, while LF-Diff (Hu et al., 2024) performs diffusion in a compact prior space but still requires iterative sampling to estimate the target prior. To avoid this, we directly reconstruct the target prior P_{gt} from the coarse prior P_c in a single step.

Following the forward diffusion process in Eq. 2, we perturb P_c at timestep t to obtain

$$Z_t = \sqrt{\bar{\alpha}_t} P_c + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{8}$$

Given Z_t, t, and the conditioning feature $\mathbf{c}$ generated by AM (Yan et al., 2019), DenoiseNet predicts the noise $\boldsymbol{\epsilon}_\theta(Z_t, \mathbf{c}, t)$. The target prior is then reconstructed in a single reverse step:

$$P^o_{gt} = \frac{Z_t - \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon}_\theta(Z_t, \mathbf{c}, t)}{\sqrt{\bar{\alpha}_t}}. \tag{9}$$

Substituting Z_t into the reconstruction gives

$$P^o_{gt} = P_c + w_t\left(\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(Z_t, \mathbf{c}, t)\right), \tag{10}$$

where $w_t = \sqrt{\frac{1-\bar{\alpha}_t}{\bar{\alpha}_t}}$ is a time-dependent coefficient controlling the update magnitude. The noise bias $\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(Z_t, \mathbf{c}, t)$ drives P_c toward P_{gt}. This enables one step prior reconstruction without iterative sampling, avoiding repeated denoising in feature space. During inference, the reconstructed P^o_{gt} replaces P_{gt} in Content-Adaptive Refinement and is combined with P_c for the final reconstruction.

3.5 LOSS FUNCTIONS

CRHDR is optimized in two stages. In the construction stage, the reconstructed H^o_{gt} is supervised by an L_1 loss in the tonemapped domain, $\mathcal{L}_\mu = \|\mathcal{T}(H^o_{gt}) - \mathcal{T}(H_{gt})\|_1$, together with a VGG perceptual loss $\mathcal{L}_{\text{per}}$. An auxiliary load balancing loss (Riquelme et al., 2021) $\mathcal{L}_{\text{route}} = \frac{1}{2}\,\text{CV}(\text{Load}(x))^2$ is also applied to balanced expert utilization. The overall construction objective is:

$$\mathcal{L}_{\text{con}} = \mathcal{L}_\mu + \lambda_{\text{per}}\mathcal{L}_{\text{per}} + \lambda_{\text{route}}\mathcal{L}_{\text{route}}. \tag{11}$$

In the reconstruction stage, the one-step feature reconstruction is supervised by noise prediction and feature reconstruction losses. Given the noisy Z_t and prior P_{gt}, the target noise is defined as:

$$\boldsymbol{\epsilon}^{gt}_t = \frac{Z_t - \sqrt{\bar{\alpha}_t}P_{gt}}{\sqrt{1-\bar{\alpha}_t}}. \tag{12}$$

Corresponding losses are defined as:

$$\mathcal{L}_{\text{noise}} = \left\| \boldsymbol{\epsilon}_\theta(Z_t, \mathbf{c}, t) - \boldsymbol{\epsilon}_t^{gt} \right\|_1, \qquad \mathcal{L}_{\text{feat}} = \left\| P_{gt}^o - P_{gt} \right\|_1. \tag{13}$$

The overall reconstruction objective is

$$\mathcal{L}_{\text{rec}} = \mathcal{L}_\mu + \lambda_{\text{per}}\mathcal{L}_{\text{per}} + \lambda_{\text{noise}}\mathcal{L}_{\text{noise}} + \lambda_{\text{feat}}\mathcal{L}_{\text{feat}} + \lambda_{\text{route}}\mathcal{L}_{\text{route}}. \tag{14}$$

The loss weights λ_{per}, λ_{noise}, λ_{feat}, and λ_{route} are set to 0.01, 0.05, 1.0, and 0.01, respectively.

4 Experiments

4.1 Implementation Details

Datasets. We conduct experiments on three publicly available real-world multi-exposure HDR datasets, including Kalantari (Kalantari & Ramamoorthi, 2017) , Tel (Tel et al., 2023), and Challenge123 dataset (Kong et al., 2024). Kalantari dataset contains 74 training scenes and 15 test scenes. Each scene consists of three LDR images captured at different exposures together with the corresponding HDR ground truth, with exposure intervals of {-2,0,+2} or {-3,0,+3}. Tel dataset contains 108 training scenes and 36 test scenes, covering diverse foreground objects, environmental conditions, and motion patterns. Challenge123 contains 96 training scenes and 27 test scenes, with larger motion between input exposures and more severe saturation in the reference image, providing a more challenging benchmark for HDR reconstruction. In addition, we employ Sen dataset (Sen et al., 2012) and Tursun dataset (Tursun et al., 2016) for qualitative evaluation. Since these datasets do not provide HDR ground truths, they are used only for visual comparisons.

Evaluation Metrics. For quantitative evaluation, we adopt five metrics, including PSNR-μ, SSIM-μ, PSNR-L, SSIM-L, and HDR-VDP2 (Mantiuk et al., 2011). The subscripts μ and L indicate that the metrics are computed in the μ-law tonemapped and linear domains, respectively. HDR-VDP2 is employed to evaluate the perceptual quality of the reconstructed HDR images.

Implementation Details. Our method is implemented in PyTorch and trained in two stages. In the first stage, we construct the target prior and train Content-Adaptive Refinement (CAR). The Adam optimizer is used with an initial learning rate of 2×10^{-4}, which is gradually reduced to 1×10^{-5} using a cosine learning rate schedule. This stage is trained for 18K iterations with 512×512 patches. In the second stage, the model trained in the first stage is loaded, while FlowNet (Kong et al., 2024) and the shared ResBlocks are frozen. CAR is further optimized together with DenoiseNet and AM. The learning rate is initialized to 5×10^{-5} and reduced to 1×10^{-7} using the same cosine schedule. This stage is also trained for 18K iterations with 128×128 image patches. DenoiseNet follows the U-Net architecture of DiffHDR. The total number of diffusion timesteps is set to T=200, with a linear noise schedule ranging from β_1=10^{-4} to β_T=2×10^{-2}. During training stage, timestep t is randomly sampled, while a fixed timestep t =120 is used during inference for the one-step reconstruction.

4.2 Comparison with State-of-the-Arts.

To comprehensively evaluate the performance of our method, we compared it against representative state-of-the-art deep learning-based approaches spanning different architectural paradigms. Specifically, we considered six CNN-based methods, including DHDR (Wu et al., 2018), AHDR (Yan et al., 2019), NHDRR (Yan et al., 2020), SAFNet (Kong et al., 2024), TMRNet (Zhang et al., 2025) and SelfHDR (Zhang et al., 2024b); one GAN-based method, HDR-GAN (Niu et al., 2021); three Transformer-based models, namely CA-ViT (Liu et al., 2022), SCTNet (Tel et al., 2023), and Hy-HDR (Yan et al., 2023a); as well as two diffusion-based methods, DiffHDR (Yan et al., 2023b) and LF-Diff (Hu et al., 2024). Moreover, all results presented in this chapter are obtained either directly from existing publications or reproduced using the publicly available code. Crucially, for methods requiring qualitative evaluation, we rigorously reproduced all results from scratch using the official codes. Across all datasets, we strictly followed the original training protocols.

Datasets w/ Ground Truth. The quantitative outcomes for CRHDR on two datasets are presented in Tab. 1. Our method is compared with SOTA approaches using the testset from Kalantari (Kalantari & Ramamoorthi, 2017) and Tel (Tel et al., 2023), which includes challenging samples characterized by saturated backgrounds and foreground motions. All quantitative results are averaged across

Table 1: Quantitative comparison with methods on Kalantari (Kalantari & Ramamoorthi, 2017) and Tel (Tel et al., 2023) dataset. For each item, the **best** and **second best** performances are highlighted.

Method	*Kalantari (Kalantari & Ramamoorthi, 2017)*					*Tel (Tel et al., 2023)*					Time (s)
	PSNR-μ	SSIM-μ	PSNR-L	SSIM-L	HDR-VDP2	PSNR-μ	SSIM-μ	PSNR-L	SSIM-L	HDR-VDP2	
DHDR (*ECCV'18*)	41.64	0.9869	40.91	0.9858	60.50	40.05	0.9794	43.37	0.9924	67.09	**0.118**
AHDR (*CVPR'19*)	43.62	0.9900	41.03	0.9862	62.30	42.08	0.9837	45.30	0.9943	68.80	0.345
NHDRR (*TIP'20*)	42.41	0.9887	41.08	0.9861	61.21	36.68	0.9590	39.61	0.9853	65.41	**0.045**
HDRGAN (*TIP'21*)	43.92	0.9905	41.57	0.9865	65.45	41.71	0.9832	44.87	0.9949	69.57	0.211
CA-ViT (*ECCV'22*)	44.32	0.9916	42.18	0.9884	66.03	42.39	0.9844	46.35	0.9948	69.23	2.670
HyHDR (*CVPR'23*)	44.64	0.9915	42.47	0.9894	66.05	-	-	-	-	-	-
DiffH (*TCSVT'23*)	44.11	0.9911	41.73	0.9885	65.52	42.18	0.9841	45.63	0.9946	69.88	7.530
SCTNet (*ICCV'23*)	44.49	**0.9924**	42.29	0.9887	**66.65**	42.55	**0.9850**	**47.51**	0.9952	**70.66**	3.460
SelfH (*ICLR'24*)	43.95	0.9907	41.77	0.9889	64.77	-	-	-	-	-	-
SAFNet (*ECCV'24*)	44.66	0.9919	**43.18**	0.9901	66.11	**42.68**	0.9792	47.46	**0.9955**	68.16	0.151
LF-Diff (*CVPR'24*)	**44.76**	0.9919	42.59	**0.9906**	66.54	-	-	-	-	-	0.720
TMRNet (*ICLR'25*)	44.09	0.9905	41.96	0.9891	65.72	-	-	-	-	-	-
CRHDR (*Ours*)	**44.83**	**0.9925**	**43.55**	**0.9908**	**66.98**	**43.47**	**0.9874**	**47.85**	**0.9959**	**70.82**	0.183

Table 2: Quantitative comparison on the Challenge123 (Kong et al., 2024) dataset. Params, Memory, FLOPs and Time are measured on an NVIDIA RTX 5090 GPU under 1500×1000 resolution.

Method	*Challenge123 (Kong et al., 2024)*					Params (M)	Memory (MiB)	FLOPs (T)	Size (Mb)	Time (s)
	PSNR-μ	SSIM-μ	PSNR-L	SSIM-L	HDR-VDP2					
AHDR (*CVPR'19*)	40.44	0.9877	28.13	0.9703	54.58	1.520	9263.44	2.170	**6.10**	0.345
NHDRR (*TIP'20*)	37.82	0.9769	26.75	0.9632	53.38	31.56	-	**0.421**	126.2	**0.045**
CA-ViT (*ECCV'22*)	40.70	0.9881	28.72	0.9731	54.63	1.220	10781.08	1.886	53.4	2.673
SCTNet (*ICCV'23*)	40.65	0.9882	28.73	0.9721	54.35	**0.990**	9957.80	1.547	29.1	3.466
SelfHDR (*ICLR'24*)	40.30	**0.9885**	27.86	**0.9741**	54.53	-	-	-	-	-
SAFNet (*ECCV'24*)	**40.94**	**0.9885**	**28.93**	0.9740	**54.84**	**1.120**	**1028.81**	**0.976**	**4.48**	**0.151**
TMRNet (*ICLR'25*)	40.29	0.9877	28.02	0.9729	54.72	-	-	-	-	-
CRHDR (*Ours*)	**41.28**	**0.9900**	**29.20**	**0.9763**	**54.93**	20.58	**5071.90**	1.520	82.3	0.183

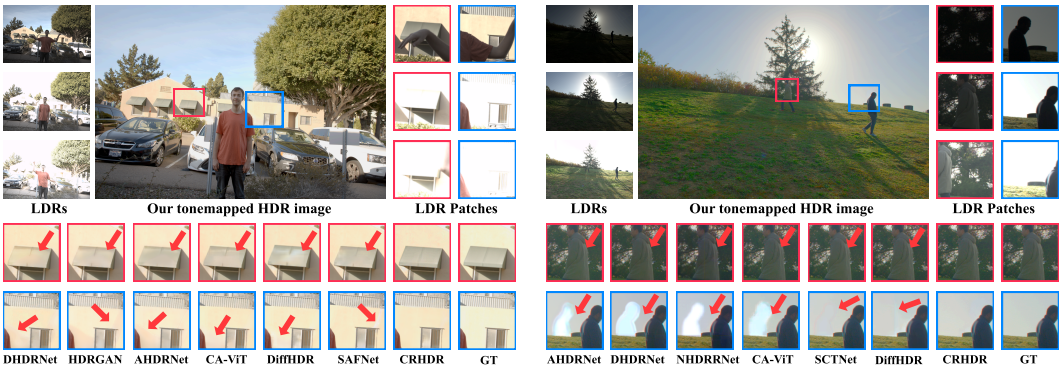
testing images. Notably, CRHDR exhibits a significant improvement over other leading methods, exceeding the DM-based method DiffHDR by 0.72 dB and 1.82 dB in PSNR-μ and PSNR-L, respectively, while achieving nearly 40× faster inference. Moreover, on the Tel dataset, CRHDR outperforms the second best results by 0.79 dB and 0.34 dB in PSNR-μ and PSNR-L, respectively.

As shown in Fig. 3 (a) and (b), the datasets present challenging samples due to information loss in the LDR images. Most approaches struggle in these areas, producing ghosting artifacts due to large motions and occlusions. DHDR (Wu et al., 2018) struggles with background motion due to error-prone alignments, resulting in ghosting. AHDR (Yan et al., 2019) aligns LDR images using convolutional spatial attention but suppress contextual information in over/underexposed scenes. CA-ViT (Liu et al., 2022), relying on patch-based sampling, produces noticeable blocky ghosting. In contrast, our CRHDR generates HDR images better aligned with human perception.

We further evaluate CRHDR on the more challenging Challenge123 (Kong et al., 2024), which contains larger inter frame motion and severe saturated regions. As shown in Table. 2, CRHDR achieves the best performance across all five evaluation metrics, outperforming the second best results by 0.34 dB and 0.27 dB in PSNR-μ and PSNR-L, respectively. In addition, CRHDR maintains competitive computational efficiency, requiring only 0.181 s to process a 1500×1000 image.

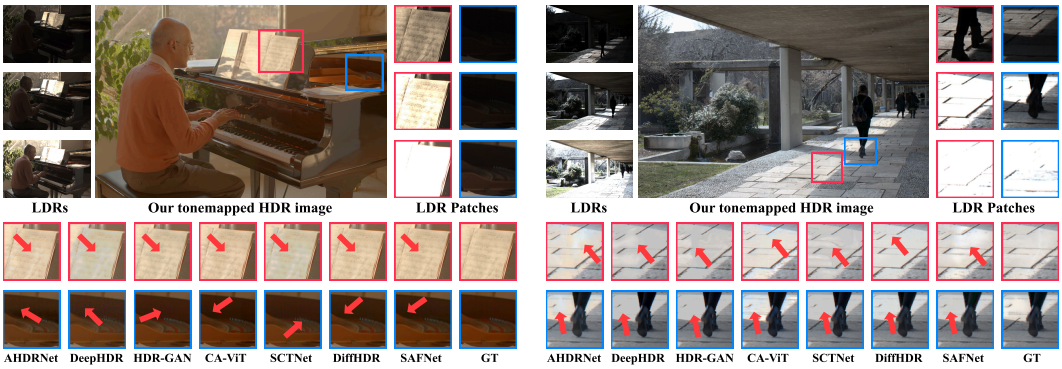
Datasets w/o Ground Truth. To assess the generalization of CRHDR, we evaluated the performance of the model trained on Kalantari (Kalantari & Ramamoorthi, 2017) by testing it on Sen (Sen et al., 2012) and Tursun (Tursun et al., 2016) datasets, which lack ground truths. In Fig. 4 (a), numerous methods encounter difficulties in recovering the large saturated region and large motion. Conversely, in Fig. 4 (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. As shown in Table 1 and 2, NHDRRNet (Yan et al., 2020), which adopts a U-shaped architecture, achieves a shorter inference time but requires substantially more parameters than most competing methods. CA-ViT (Liu et al., 2022) and SCTNet (Tel et al., 2023) contain multiple Transformer blocks, resulting in high computational costs despite their small parameter counts. DiffHDR (Yan et al., 2023b) requires longer inference times due to its iterative denoising process. In contrast, our method obtains a coarse HDR result through optical flow based alignment and then performs one-step denoising in a compact feature space to recover a high quality prior. This design avoids costly iterative sampling while maintaining competitive computational efficiency.



(a) Example from Kalantari (Kalantari & Ramamoorthi, 2017) (b) Example from Tel (Tel et al., 2023)

Figure 3: Visual comparisons on various test datasets show that CRHDR achieves better local reconstruction quality than competing methods. Zoom in for a better view.



(a) Example from Sen (Sen et al., 2012) (b) Example from Tursun (Tursun et al., 2016)

Figure 4: Visual comparisons on the datasets without ground truth

4.3 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 (Kalantari & Ramamoorthi, 2017).

Effect of Prior Assignment Across Experts. To study how prior assignment affects expert refinement, we compare four configurations for the experts. As shown in Table 3, when all experts use P_c, they access only coarse information, limiting reconstruction. P_{mix} provides richer information, but assigning same quality level to all experts reduces diversity and introduces redundancy. Although P_{gt} contains the highest quality information, using it for all experts makes the refinement sensitive to target prior reconstruction . In contrast, assigning P_c, P_{mix}, and P_{gt} to different experts forms a coarse-to-high quality hierarchy, providing diverse information and achieving the best performance.

To further visualize the behavior of the router, as shown in Fig. 5 (a), the coarse HDR image H_c is resized to 1024×1024 and divided into non-overlapping 256×256 patches. Each patch is routed independently, and its selected prior is overlaid on the residual map $H_{gt} - H_c$. For comparison, each patch is passed through all three expert paths, with the the highest local PSNR-L selected as the oracle choice. The corresponding outputs are stitched to form H_{syn}. We visualize $H_{gt} - H_{syn}$ and the local PSNR-L gap between the router-selected and oracle choices. Most regions show a gap close to 0 dB, indicating that the router generally makes near optimal local selections. Notably, P_{gt} is not always preferred; in some regions, selecting P_c yields better local reconstruction, supporting content-adaptive prior routing rather than uniformly relying on the highest-quality prior.

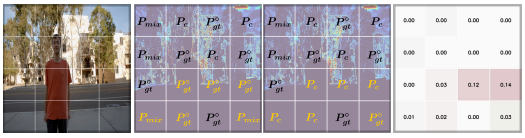
Effect of Different Losses. We evaluate the contribution of different losses by progressively adding $\mathcal{L}_{\text{route}}$, $\mathcal{L}_{\text{noise}}$, and $\mathcal{L}_{\text{feat}}$, while $\mathcal{L}_{\text{per}}$ is used in all settings. As shown in Tab. 4, $\mathcal{L}_{\text{route}}$ mainly regularizes expert utilization and helps maintain balanced routing. Introducing $\mathcal{L}_{\text{noise}}$ further improves PSNR-μ and PSNR-L by 0.12 dB and 0.27 dB, respectively, demonstrating the effectiveness of direct noise supervision for one-step reconstruction. Further incorporating $\mathcal{L}_{\text{feat}}$ provides complementary feature-level supervision, and the full objective achieves the best performance.

Table 3: Ablation study on prior assignment across experts on Kalantari dataset.

Method	PSNR-μ	SSIM-μ	PSNR-L	SSIM-L
$[P_c, P_c, P_c]$ (*Baseline*)	44.42	0.9914	43.01	0.9886
$[P_{mix}, P_{mix}, P_{mix}]$	44.70	0.9921	43.36	0.9902
$[P_{gt}, P_{gt}, P_{gt}]$	44.60	0.9919	43.39	0.9901
$[P_c, P_{mix}, P_{gt}]$ (*Ours*)	**44.83**	**0.9925**	**43.55**	**0.9908**

Table 4: Effect of different losses: (a), (b), (c), (d) denote $\mathcal{L}_{\mu}$, $\mathcal{L}_{\text{route}}$, $\mathcal{L}_{\text{noise}}$, $\mathcal{L}_{\text{feat}}$, respectively.

Method	PSNR-μ	SSIM-μ	PSNR-L	SSIM-L
(a)	44.57	0.9916	43.18	0.9899
(a)+(b)	44.59	0.9918	43.22	0.9903
(a)+(b)+(c)	44.71	0.9923	43.49	0.9905
(a)+(b)+(c)+(d) (*Ours*)	**44.83**	**0.9925**	**43.55**	**0.9908**



(a) Visualization of expert routing behavior. Zoom in for a better view.



(b) Visual comparisons of different training designs.

Figure 5: Visual results of our ablation study.

Effect of Training Designs. We investigate different designs of CRHDR in Tab. 5. The baseline is identical to the $[P_c, P_c, P_c]$ setting in Tab. 3, where all three experts use only the coarse prior. In split training, DenoiseNet and AM are optimized independently and then directly integrated into the network for inference. This leads to inferior performance, as the reconstructed prior is not optimized under the final HDR reconstruction objective and cannot adapt to CAR refinement. In contrast, joint training optimizes DenoiseNet, AM, and CAR together, enabling the reconstructed prior to better support downstream reconstruction and achieve improved performance. It is worth noting that we further replace the diffusion process with direct regression, where DenoiseNet directly maps paired P_c to P_{gt} without noise perturbation. The inferior performance indicates that direct regression tends to learn a fixed mapping, while diffusion based reconstruction provides a more flexible way to model the transformation from coarse to target priors. Moreover, we present visual comparisons between the baseline and different CRHDR variants in Fig. 5 (b). Performing one-step adaptive reconstruction to predict HDR effectively mitigates ghosting artifacts.

Table 5: Ablation on different training designs.

Method	GT	Diffusion	Joint Training	PSNR-μ	PSNR-L
Baseline				44.42	43.01
CRHDR$_{stage1}$	✓			45.80	46.92
Split Training	✓	✓		44.32	43.19
Regression	✓		✓	44.48	43.12
Ours	✓	✓	✓	**44.83**	**43.55**

Table 6: Ablation Study on timestep settings.

Inference Sampling Timesteps	Training Sampling Timesteps	PSNR-μ	PSNR-L
$t = 10$	[1,200]	44.81	43.52
$t = 60$	[1,200]	44.79	43.51
$t = 120$	[1,200]	**44.83**	**43.55**
$t = 180$	[1,200]	44.80	43.52
$t = 120$	120	44.54	43.07

Sampling Step. In Tab. 6, when the training timestep is randomly sampled from [1,200], different inference timesteps ($t = 10$, 60, 120, 180) produce similar results, indicating that DenoiseNet learns a robust mapping across noise levels. In contrast, fixing the training timestep $t = 120$, even with the same timestep during inference, leads to a performance drop. This suggests that diverse noise levels during training provide effective supervision and improve the generalization of one-step reconstruction, while training at a single noise level limits the learned reconstruction capability.

5 Conclusion

In this paper, we propose CRHDR, a Construction then Reconstruction framework for HDR deghosting. Instead of directly reconstructing HDR images from incomplete multi-exposure observations, our method first constructs high-quality target priors and then learns to recover them from coarse HDR representations. A content-adaptive refinement mechanism further exploits priors of different quality levels through patch-wise expert routing, allowing different regions to follow suitable processing paths. Moreover, one-step diffusion reconstruction is introduced in a compact feature space to recover the target prior without iterative sampling. Extensive experiments on real-world HDR benchmarks demonstrate that CRHDR consistently improves reconstruction quality while maintaining efficient inference, achieving up to 40× acceleration over iterative diffusion-based methods, confirming it as an effective and computationally efficient solution for HDR deghosting.

AI USE STATEMENT

In this work, we used generative AI tools only to improve the readability and language of the manuscript. All AI-assisted edits were reviewed by the authors. We take full responsibility for the final content of this work.

REFERENCES

L. Bogoni. Extending dynamic range of monochrome and color images through fusion. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, volume 3, pp. 7–12 vol.3, 2000. 1, 2

Sibi Catley-Chandar, Thomas Tanay, Lucas Vandroux, Aleš Leonardis, Gregory Slabaugh, and Eduardo Pérez-Pellitero. Flexhdr: Modeling alignment and exposure uncertainties for flexible hdr imaging. *IEEE Transactions on Image Processing*, 31:5923–5935, 2022. 1, 3

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

Haesoo Chung and Nam Ik Cho. Lan-hdr: Luminance-based alignment network for high dynamic range video reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12760–12769, 2023. 1, 3

Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '97, pp. 369–378, 1997. 1

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

Eric R. Fossum and Donald B. Hondongwa. A review of the pinned photodiode for ccd and cmos image sensors. *IEEE Journal of the Electron Devices Society*, 2(3):33–43, 2014. 1

Orazio Gallo, Natasha Gelfandz, Wei-Chao Chen, Marius Tico, and Kari Pulli. Artifact-free high dynamic range imaging. In *2009 IEEE International Conference on Computational Photography (ICCP)*, pp. 1–7, 2009. 1

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf. 3, 4

Jun Hu, Orazio Gallo, Kari Pulli, and Xiaobai Sun. Hdr deghosting: How to deal with saturation? In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1163–1170, 2013. 1, 2

Tao Hu, Qingsen Yan, Yuankai Qi, and Yanning Zhang. Generating content for hdr deghosting from frequency view. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25732–25741, 2024. doi: 10.1109/CVPR52733.2024.02431. 2, 3, 5, 6

Katrien Jacobs, Celine Loscos, and Greg Ward. Automatic high-dynamic range image generation for dynamic scenes. *IEEE Computer Graphics and Applications*, 28(2):84–93, 2008. 1, 2

Nima Khademi Kalantari and Ravi Ramamoorthi. Deep high dynamic range imaging of dynamic scenes. *ACM Trans. Graph.*, 36(4), 2017. 1, 2, 3, 6, 7, 8

Sing Bing Kang, Matthew Uyttendaele, Simon Winder, and Richard Szeliski. High dynamic range video. *ACM Trans. Graph.*, 22(3):319–325, 2003. 1, 2

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Advances in Neural Information Processing Systems*, 35:23593–23606, 2022. 3

Erum Arif Khan, Ahmet Oguz Akyuz, and Erik Reinhard. Ghost removal in high dynamic range images. In *2006 International Conference on Image Processing*, pp. 2005–2008, 2006. 1, 2

Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5886–5895. IEEE, 2023. 5

Lingtong Kong, Bo Li, Yike Xiong, Hao Zhang, Hong Gu, and Jinwei Chen. Safnet: Selective alignment fusion network for efficient hdr imaging. In *European Conference on Computer Vision*, pp. 256–273. Springer, 2024. 3, 4, 6, 7

Chul Lee, Yuelong Li, and Vishal Monga. Ghost-free high dynamic range imaging via rank minimization. *IEEE Signal Processing Letters*, 21(9):1045–1049, 2014. 2

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. 3

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 *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 463–470, 2021. 3

Zhen Liu, Yinglong Wang, Bing Zeng, and Shuaicheng Liu. Ghost-free high dynamic range imaging with context-aware transformer. In *Computer Vision – ECCV 2022*, pp. 344–360, 2022. 2, 3, 6, 7

Kede Ma, Hui Li, Hongwei Yong, Zhou Wang, Deyu Meng, and Lei Zhang. Robust multi-exposure image fusion: A structural patch decomposition approach. *IEEE Transactions on Image Processing*, 26(5):2519–2532, 2017. 1

Kede Ma, Zhengfang Duanmu, Hanwei Zhu, Yuming Fang, and Zhou Wang. Deep guided learning for fast multi-exposure image fusion. *IEEE Transactions on Image Processing*, 29:2808–2819, 2020. 1

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*, pp. 1–14, 2011. 6

Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion: A simple and practical alternative to high dynamic range photography. In *Computer Graphics Forum*, 2009. 1

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. 6

Tae-Hyun Oh, Joon-Young Lee, Yu-Wing Tai, and In So Kweon. Robust high dynamic range imaging by rank minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37 (6):1219–1232, 2015. 2

Fabrizio Pece and Jan Kautz. Bitmap movement detection: Hdr for dynamic scenes. In *2010 Conference on Visual Media Production*, pp. 1–8, 2010. 1, 2

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

K. Ram Prabhakar, Susmit Agrawal, Durgesh Kumar Singh, Balraj Ashwath, and R. Venkatesh Babu. Towards practical and efficient high-resolution hdr deghosting with cnn. In *Computer Vision – ECCV 2020*, pp. 497–513, 2020. 1

K. Ram Prabhakar, V Sai Srikar, and R. Venkatesh Babu. Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017. 1

Carlos Riquelme, Joan Puigcerver, Basil Mustafa, Maxim Neumann, Rodolphe Jenatton, André Susano Pinto, Daniel Keysers, and Neil Houlsby. Scaling vision with sparse mixture of experts. *Advances in neural information processing systems*, 34:8583–8595, 2021. 5

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*, pp. 10684–10695, 2022. 3

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. 3

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, 6, 7, 8

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*, pp. 288–304, 2022. 2, 3

Steven Tel, Zongwei Wu, Yulun Zhang, Barthélémy 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)*, pp. 12836–12845, 2023. 2, 3, 6, 7, 8

Okan Tarhan Tursun, Ahmet Oğuz Akyüz, Aykut Erdem, and Erkut Erdem. An objective deghosting quality metric for hdr images. In *Computer Graphics Forum*, volume 35, pp. 139–152. Wiley Online Library, 2016. 6, 7, 8

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. 3

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*, pp. 16293–16303, 2022. 3

Shangzhe Wu, Jiarui Xu, Yu-Wing Tai, and Chi-Keung Tang. Deep high dynamic range imaging with large foreground motions. In *Computer Vision – ECCV 2018*, pp. 120–135, 2018. 1, 3, 6, 7

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*, pp. 13095–13105, 2023. 3

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*, pp. 1751–1760, 2019. 2, 4, 5, 6, 7

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, 3, 6, 7

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*, pp. 22211–22220, 2023a. 1, 3, 6

Qingsen Yan, Tao Hu, Yuan Sun, Hao Tang, Yu Zhu, Wei Dong, Luc Van Gool, and Yanning Zhang. Toward high-quality hdr deghosting with conditional diffusion models. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(5):4011–4026, 2023b. 2, 3, 5, 6, 7

Qian Ye, Jun Xiao, Kin-man Lam, and Takayuki Okatani. Progressive and selective fusion network for high dynamic range imaging. In *Proceedings of the 29th ACM International Conference on Multimedia*, pp. 5290–5297, 2021. 3

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 (CVPR)*, pp. 5728–5739, 2022. 5

Wei Zhang and Wai-Kuen Cham. Gradient-directed multiexposure composition. *IEEE Transactions on Image Processing*, 21(4):2318–2323, 2012. 1

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)*, pp. 7660–7664. IEEE, 2024a. 2, 3

Zhilu Zhang, Haoyu Wang, Shuai Liu, Xiaotao Wang, Lei Lei, and Wangmeng Zuo. Self-supervised high dynamic range imaging with multi-exposure images in dynamic scenes. In *International Conference on Learning Representations*, volume 2024, pp. 25867–25884, 2024b. 6

Zhilu Zhang, Shuohao Zhang, Renlong Wu, Zifei Yan, and Wangmeng Zuo. Exposure bracketing is all you need for a high-quality image. In *International Conference on Learning Representations*, volume 2025, pp. 95393–95418, 2025. 6