泽彬 温

Hierarchical Adaptive Attention and Multi-Scale for Ghost-Free HDR Imaging

No repository

Document Details

Submission ID
trn:oid:::3618:99936562

Submission Date
Jun 9, 2025, 9:25 AM GMT+8

Download Date
Jun 9, 2025, 9:28 AM GMT+8

File Name
917816300296732672_泽彬 温_Hierarchical Adaptive Attention and Multi-Scale Transformer for ....docx

File Size
652.4 KB

12 Pages

7,590 Words

42,737 Characters

0% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Detection Groups

0 AI-generated only 0%
Likely AI-generated text from a large-language model.

0 AI-generated text that was AI-paraphrased 0%
Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?
The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?
Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.

Hierarchical Adaptive Attention and Multi-Scale Transformer for Ghost-Free HDR Imaging

ZEBIN WEN[1], SHU GONG[2]
[1]Guangdong University of Science and Technology, Dong Guan 523000 CHINA (e-mail: 2079331865@qq.com)
[2]Guangdong University of Science and Technology, Dong Guan 523000 CHINA (e-mail: @qq.com)

Corresponding author: GONG SHU (e-mail: @qq.com).

This work was supported in part by the Ministry of Education of China under Grant 23JDSZ3152.

ABSTRACT High Dynamic Range (HDR) imaging synthesizes vivid images by merging multiple low dynamic range (LDR) images of different exposures. Nevertheless, in dynamic scenes, object motion or camera commonly introduce ghosting artifacts, severely degrading image quality. Although numerous DNN-based methods have been proposed to address this issue, existing solutions remain unsatisfactory. Spatial attention-based approaches often struggle to cope with complex scenarios characterized by random luminance fluctuations and large-scale motion, while conventional HDR deghosting models that rely on CNN during the fusion stage are hampered by limited receptive fields, lack of dynamic weighting and the absence of multi-scale capabilities.To overcome these limitations, we propose two innovative modules. The Luminance Adaptive Channel Attention (LACA) module dynamically and adaptively modulates channel-wise weights across multiple scales. This enables precise information balancing among channels, effectively suppressing ghosting artifacts and alleviating color saturation issues, thereby yielding refined feature representations that enhance the HDR fusion process.The Multi-Scale Residual Swin Transformer Block (MSRSTB), empowered by a multi-scale Transformer architecture, provides an expansive receptive field and dynamic weighting mechanism. It adeptly handles diverse motion patterns, integrating features in a hierarchical, coarse-to-fine manner, and efficiently manages regions with varying exposure levels. As a result, it significantly reduces saturation artifacts and mitigates ghosting, facilitating high-quality HDR image reconstruction in challenging scenarios. Comprehensive qualitative and quantitative evaluations emonstrate that our proposed modules outperform state-of-the-art methods.

INDEX TERMS High Dynamic Range imaging, Ghosting artifacts, Saturation, Luminance Adaptive Channel Attention, Multiple Scales, Multi-Scale Residual Swin Transformer Block, Coarse-to-Fine, Expansive Receptive Field, Dynamic Weighting Mechanism.

I. INTRODUCTION

NATURAL scenes exhibit a vast spectrum of illumination, yet standard digital camera sensors are only capable of capturing a restricted dynamic range. Consequently, camera-captured images frequently contain saturated or underexposed areas, resulting in poor visual quality due to significant loss of detail. High Dynamic Range (HDR) imaging has emerged as a solution to these constraints, enabling the display of more detailed visual content. A typical approach in HDR imaging involves fusing a sequence of Low Dynamic Range (LDR) images with varying exposures. While this method can generate high-quality HDR images in static scenes with stationary cameras, it often produces ghosting artifacts when dealing with moving objects or images captured by handheld cameras.

Currently, numerous solutions have been put forward to tackle this issue. Rejection-based methods [1]–[4], [87]–[89] are capable of rapidly detecting misaligned regions and eliminating them during the image fusion process. While these methods demonstrate relatively good performance in predominantly static scenes, they still struggle with ghosting artifacts when dynamic objects cannot be accurately detected. Alignment-based methods adopt an explicit strategy to align non-reference images with a pre-selected reference image. These methods either fail to handle complex object motions effectively or are highly error-prone when dealing with motion and occlusion. Patch-based methods aim to generate pure static low dynamic range images from dynamic input images. However, it is characterized by high computational complexity and requires a longer time for scene inference.

The advent of Deep Neural Networks (DNNs) has spurred numerous studies that utilize Convolutional Neural Networks (CNNs) to directly model the complex mapping between Low Dynamic Range (LDR) and High Dynamic Range (HDR) images. These models [17], [61] commonly first align LDR images using optical flow or homography. However, such alignment methods are error-prone which leads to ghosting artifacts in the presence of complex foreground motion. Conversely, the attention-based models [75], [78] mitigate motion and saturation issues through spatial attention mechanisms. Building on AHDRNet, subsequent works [66], [77], [82], [83] have aimed to further eliminate ghosting artifacts. The spatial attention modules in these works generate attention maps, which modulate non-reference features by element-wise multiplication. This mechanism selectively suppresses motion and saturation and accentuates informative content which improves HDR image quality.

However, existing spatial attention-based methods mainly focus on spatial-level suppression and enhancement, which may fall short in handling complex scenarios with randomly varying luminance and large-scale motion. In such challenging conditions, the ghosting artifacts and saturation issues in HDR imaging persist due to the inability to effectively manage the information flow across different channels in a multi-scale context. To address these limitations, we propose the Luminance Adaptive Channel Attention (LACA) module. Unlike traditional approaches, LACA adaptively modulates the weights of different channels in a multi-scale and dynamic manner. By doing so, it can precisely capture and balance the information from various channels, even under extreme luminance changes and significant motion. This multi-scale dynamic weight adjustment strategy not only suppresses ghosting artifacts more effectively but also mitigates saturation issues. As a result, it provides a more refined feature representation, which is highly beneficial for the fusion stage in HDR reconstruction. This ensures that the final HDR images exhibit higher quality with enhanced details and fewer visual artifacts.

Moreover, existing HDR deghosting models that predominantly rely on CNNs in the fusion stage are constrained by their limited receptive fields and are devoid of dynamic weighting mechanisms and multi-scale capabilities. As a result, they struggle to effectively fuse features from motion-affected and saturated regions in a coarse-to-fine manner, particularly in complex scenarios with large luminance variations and large-scale motions.To overcome these limitations, we introduce the Multi-Scale Residual Swin Transformer Block (MSRSTB). Leveraging the power of a multi-scale Transformer architecture, MSRSTB offers an extensive receptive field which enables the model to capture global context while also attending to fine-grained details. The dynamic weighting scheme within the MSRSTB adaptively adjusts the importance of different features across multiple scales, allowing it to handle a wide range of motions, from subtle local displacements to large-scale global movements. This multi-scale and dynamic feature integration mechanism enables the MSRSTB to more effectively manage regions with varying exposure levels, reducing saturation artifacts and mitigating ghosting issues. By fusing features in a hierarchical, coarse-to-fine fashion, theMSRSTB provides a more comprehensive and accurate representation of the scene which is crucial for the high-quality reconstruction of HDR images even in the most challenging scenarios.

In summary, we introduce the LACA and MSRDTB modules which are designed to dynamically process motion-affected and saturated regions with multi-scale capabilities and a large receptive field. These modules effectively handle ghosting artifacts and enable superior fusion of HDR images, as shown in Figure 1. In conclusion, the principal contributions of this paper can be summarized as follows:

- We propose the Luminance Adaptive Channel Attention (LACA) module and the Multi-Scale Residual Swin Transformer Block (MSRSTB). The LACA module dynamically and adaptively modulates channel-wise weights across multiple scales which effectively suppressing ghosting artifacts and alleviating color saturation under extreme luminance variations and significant motions. This results in refined feature representations which are highly conducive to the fusion stage.
- The proposed MSRSTB is powered by a multi-scale Transformer architecture which offers an expansive receptive field. Its dynamic weighting mechanism adaptively integrates features across various scales enabling it to handle diverse motion patterns ranging from fine-grained local displacements to large-scale global movements. By fusing features in a hierarchical and coarse-to-fine manner, the MSRSTB effectively manages regions with different exposure levels, reducing saturation and ghosting issues, and thus facilitating high-quality HDR image reconstruction even in complex scenarios.
- Extensive qualitative and quantitative experiments conducted on two datasets with ground truth and two without ground truth demonstrate the superior performance of our proposed modules, validating their effectiveness and robustness.

II. RELATED WORK

Existing HDR deghosting algorithms are primarily classified into three categories: image registration methods, motion rejection methods and CNN-based methods.

Image registration methods. Bogoni employed optical flow to estimate motion vectors and utilized specific parameters to warp pixels within the exposure images. Kang et al. first transformed LDR images intensities into the luminance domain by leveraging exposure time information. Then, they estimated optical flow to merge these LDR images. Zimmer et al. leveraged the energy-based optic flow method to align the LDR images in order to achieve more accurate alignment results Sen et al. presented a new patch-based energy-minimization method, which combines alignment and reconstruction through joint optimization. Hu et al. carried out alignment of images within an HDR image stack. This

FIGURE 1. The proposed approach is capable of efficiently eliminating ghosting artifacts and delivers excellent visual quality.

process propagates intensity and gradient information on the transformed domain. Hafner et al. put forward an energy minimization approach that computes HDR irradiance and displacement fields simultaneously. Although image registration methods generate dense matching, they still suffer from large motion and occlusion. In addition, they exhibit high computational complexity and require more time for scene inference.

Motion rejection methods. Motion rejection methods first perform global registration on the LDR images and detect motion regions, then they reject the misalignment pixels and merge static regions to reconstruct HDR images. Grosch et al. [4] identified motion regions and estimated an error which was based on alignment color differences to generate ghost-free HDR images. Pece et al. [27] used the median threshold bitmap of the LDR images to detect and reject motion areas. Jacobs et al. marked misaligned locations through the variance of weighted intensity. Zhang et al. [59] calculated a motion weighting map using quality measures based on image gradients. Moreover, several methods [32], [58] utilized rank minimization approach to identify motion regions to generate an HDR image without ghosted regions. Unfortunately, these approaches often exhibit significant limitations, resulting in the lack of contents. This is attributable to the fact that valuable information is irretrievably lost during the pixel rejection phase.

CNN-based method. Kalantari et al. [17] first performed optical flow alignment then employed a convolutional neural network to merge LDR images. Wu et al. [61] utilized an image homography transformer to account for camera motion by framing the HDR imaging as an image translation task. Yan et al. [75] developed an attention module to suppress motion and saturation in order to reconstruct ghost-free HDR images. Yan et al. [6] used multi-scale network to refine the generated results. Yan et al. [76] designed a nonlocal block to capture the constraints of the local receptive field for global HDR merging. Niu et al. [62] introduced HDR-GAN which is leveraging Generative Adversarial Networks to synthesize missing content.

III. METHOD

A. PROBLEM DEFINITION

Given a set of Low-Dynamic-Range (LDR) images captured from a dynamic scene with varying exposure levels, the objective istoreconstruct a High-Dynamic-Range (HDR) image H that is precisely aligned with a designated reference image . In this paper, the intermediate LDR image is selected as the reference. Specifically, we comprise three LDR images ( ) and designate as the reference image .

Following previous work [17], we construct a linearized image for every using gamma correction as follows

$$H = I^{\gamma} / t \tag{1}$$

where indicates the exposure time of LDR image , denotes the gamma correction parameter and we set to 2.2. Subsequently, we combine and along the channel dimension resulting in a 6-channel input denoted as Given the inputs X_1, X_2 and X_3, our model generates an HDR image H through the following process:

$$\hat{H} = f(X; \theta) \tag{2}$$

where denotes the HDR deghosting network, is the parameters of the network.

B. OVERVIEW

The proposed method aimsto remove the ghosted regions and construct a high-quality HDR image. As shown in Figure 2 , our proposed network is composed of two subnetworks, a dynamic alignment subnetwork and a multi-scale Transformer-based fusion subnetwork. The alignment network is to suppress moving regions, and the fusion subnetwork aims to produce details for the degraded regions. The alignment subnetwork incorporates a dynamic weight motion removal mechanism. This mechanism effectively eliminates motion in non-reference images and enhances feature representation simultaneously, thereby suppressing saturated regions. We employ a multi-scale Transformer-based subnetwork to effectively integrate features extracted at different scales. The multi-scale Transformer module adaptively integrates feature representations of moving regions, ranging from fine-grained local motions to large-scale global displacements across multiple scales. Additionally, it dynamically fuses regions with varying exposures to mitigate the saturation problem.

C. ALIGNMENT SUBNETWORK

Feature Extraction. We obtain three 6-channel LDR images , then we extract features of using feature extraction layers .

$$F_{ea} = f_{FE}(X_i) \tag{3}$$

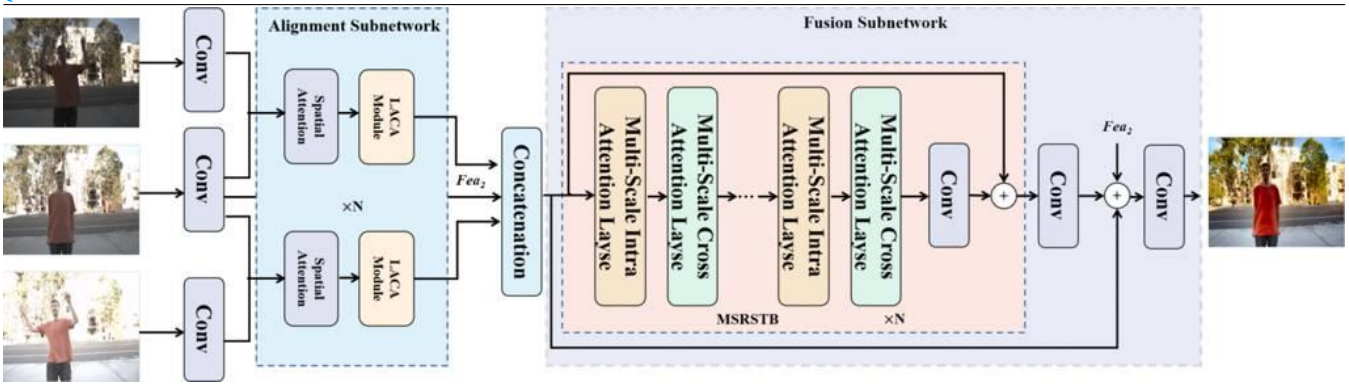

FIGURE 2. The framework of the proposed method.

Note that

Spatial Attention. Inspired by [75], in order to alleviate the ghosting artifacts between the reference and the non-reference images, the non-reference features are processed with the reference feature map Fea_r through a series of specialized spatial attention modules denoted as . It should be noted that is the extracted features from X_2. The attention map is calculated as follows:

(4)

We apply sigmoid function to constrain the values of the attention map within the range of 0 to 1. Subsequently, we perform an element-wise multiplication operation between the feature maps of the non-reference images and their corresponding estimated attetion maps . Finally, we obtain the refined feature maps for each non-reference image.

(5)

where represents the element-wise multiplication. As shown in 1, this mechanism enables the precise identification of spatial inconsistencies, which is fundamental for fusion stage.

LACA Module. Even though spatial attention demonstrates remarkable effectiveness in reducing the misalignment discrepancies between the reference and non-reference images, in scenes with intricate luminance and color variations, spatial attention fails to fully harness channel information, preventing it from effectively removing artifacts in complex motion regions and thus leaving ghosting artifacts. To solve this problem, we design a Luminance Adaptive Channel Attention (LACA) module to adaptively adjust the weight of different channel which can more effectively balance information from different channels, mitigating the motion and saturation problem. The LACA module consists of multi-scale squeeze submodule, adaptive excitation submodule and reweighting submodule.

In the multi-scale squeeze submodule, given an input feature map we first perform global pooling with different kernel sizes to obtain multi-scale global feature. For each channel c in the input feature map we calculate the global feature as follows:

(6)

where is the global feature of the c-th channel under the m-th pooling operation. This results in N sets of global features for . We then concatenate these multi-scale global features to get Z:

(7)

where .

We then use two adaptive fully-connected layers to learn the channel weights. The first fully-connected layer is defined as:

(8)

where , and the Mish function is formulated as:

(9)

The second fully-connected layer is defined as:

(10)

where and is the sigmoid activation function.

Subsequently, we perform a channel-wise multiplication of the learned channel weights with the input feature and get the output feature :

(11)

where and represents the -th non-reference feature.

After the LACA module, we get and feature which contains harmless information. Then we concatenate , and for fusion subnetwork.

(12)

D. FUSION SUBNETWORK

Inspired by SWIN-IR [70], the fusion subnetwork con- sists of several Multi-Scale Residual Transformer Blocks (MSRSTBs) to dynamically integrates features ranging from fine-grained local motions to large-scale global displace- ments across multiple scales. The input feature map $F_{in} \in \mathbb{R}^{H \times W \times C}$ is first embedded into token embeddings, then we adopt several MSRSTBs and a convolution block to recon- struct an HDR image without motion and saturation.

MSRSTB. To further boost multi-scale feature extraction capabilities, we present an innovative approach to reconstruct ghost-free HDR images in a distinct way. We first generate multi-scale Q, K and V matrices by applying convolutional layers with different kernel sizes and strides to the input feature map. For the i-th scale level ($i = 1, 2, \ldots, N$), the query matrix Q_i is obtained as:

$$Q_i = Conv(W_{Q_i}, F_{in}) \quad (13)$$

where $Conv()$ represents the convolutional operation, W_{Q_i} is the weight matrix at the i-th scale, s_i is the stride and k_i is the kernel size. Similarly, the key matrix K_i and value matrix V_i for the i-th scale level is generated by:

$$K_i = Conv(W_{K_i}, F_{in}) \quad (14)$$

$$V_i = Conv(W_{V_i}, F_{in}) \quad (15)$$

Subsequently, we calculate the intra-scale attention for each scale level i:

$$A_i^{ii} = softmax\left(\frac{Q_i^T}{\sqrt{d}} + B\right) \quad (16)$$

$$O_i^{ii} = A_i^{ii} V_i \quad (17)$$

where B is a learnable position encoding, d is the dimension of Q_i. We also adopt shifted windows mechanism. After that, we calculate the cross-scale attention scores for each scale i with respect to all other scales j ($j \neq i$) to capture cross-scale features:

$$A_{ij}^{ij} = softmax\left(\frac{Q_i^T}{\sqrt{d}} + B\right) \quad (18)$$

$$O_{ij}^{ij} = A_{ij}^{ij} V_j \quad (19)$$

Then, we aggregate the intra-scale and cross-scale attention outputs for each scale i:

$$O_i = \gamma_{ii} O_i^{ii} + \sum_{j \neq i} \gamma_{ij} O_i^{ij} \quad (20)$$

where γ_{ii} and γ_{ij} are weighting hyperparameters.

Finally, we use bilinear interpolation to upsample the lower-scale features. Then, we concatenate the upsampled features along the channel dimension:

$$O_{out} = [O_1, O_2, O_3] \quad (21)$$

Lastly, we use convolutional layers to obtain the final output feature:

$$O_{out} = Conv(O_{out}) \quad (22)$$

E. TRAINING LOSS

As HDR images are presented in the tonemapped domain, we adopt the μ-law [17] to transform the images from the linear domain to the tonemapped domain:

$$T(x) = \frac{\log(1+\mu x)}{\log(1+\mu)} \quad (23)$$

where $T(x)$ is the tonemapped function and we set $\mu = 5000$.

Considering the predicted result $\hat{H}$ and the ground truth H, we calculate the tonemapped per-pixel loss and perceptual loss as described:

$$L = \| T(\hat{H}) - T(H) \|_1 + \gamma \| \phi(T(\hat{H})) - \phi(T(H)) \|_1 \quad (24)$$

where ϕ represents the feature extracted from VGG19, $\gamma = 0.01$ is a weighting hyperparameter.

F. IMPLEMENTATION DETAILS

We extract 256 × 256 patches with a stride of 64 for the training. To optimize our model, we employ Adam optimizer. The window size of the fusion subnetwork is set to 8. The batch size is set to 4 and the learning rate is set at 2×10^{-4}. Additionally, we set the hyperparameters of the Adam optimizer such that β_1 is 0.9, β_2 is 0.999 and ϵ is 10^{-8}. Our model is implemented using the PyTorch framework with 2 NVIDIA GeForce 3090 GPUs for 200 epochs. We use the $PSNR\text{-}\mu$ score computed on validation set to save the best checkpoint.

IV. EXPERIMENTS

A. EXPERIMENTAL SETTINGS

Dataset. We train all the methods on two publicly datasets, Kalantari's dataset [17] and Hu's dataset [73]. Kalantari's dataset consists of real-world scenario. It contains 74 samples for training and 15 for testing. In each sample, three distinct LDR images are captured. The exposure biases for these captures are either {—2, 0, +2} or {—3, 0, +3}. On the other hand, Hu's dataset is a synthetic dataset. The images in this dataset are captured at three exposure levels, specifically {—2, 0, +2}. For our experiments, we focus on the dynamic scene images from Hu's dataset and select the 85 samples for training purposes and the remaining 15 samples for testing. To assess the generalization ability of our model, we also conduct evaluations on Sen's dataset [35] and Tursun's dataset [48] which do not have ground truth.

Evaluation Metrics. We use PSNR, PSRN-μ, SSIM, SSIM-μ and HDR-VDP-2 [46] metrics for evaluation, where PSNR denotes linear domain and PSNR-μ represents tonemapped domain and the same goes for SSIM.

B. QUALITATIVE EVALUATIONS

We evaluate the performance of the proposed method and compare it with several state-of-the-art methods, which con-

tains two patch-based methods and five deep learning-based methods: Kalantari [17], AHDRNet [75] , DeepHDR [61], NHDRRNet [76], ADNet [66] and HDR-GAN [62].

As illustrated in Figure 3 and 4, both datasets include challenging samples with extensive foreground motions and over/under-exposed regions. Most competing methods yield ghosting artifacts in these areas because of large motion and occlusion. Kalantari's method and DeepHDR attempt to align the images using optical flow and tomographies. However, due to the error-prone nature of these alignment methods, they fail to effectively handle background motion. This results in undesirable ghosting artifacts which is shown in Figure 3 and 4. Although AHDRNet and ADNet effectively miti- gates ghosting artifacts by exploiting spatial attention mech- anisms, it inevitably dampens some advantageous contextual information. Moreover, it struggles to reconstruct large-scale motion in over/under-exposed regions. These limitations are clearly shown in the highlighted regions of Figure 3 and 4. As for Tursun's [48] and Sen's [35] datasets, which do not have ground truth, visual comparisons are shown in Fig- ure 5. These comparisons reveal that most methods struggle to reconstruct large-scale saturated regions and motions. In Figure 5, as highlighted by the red block, other methods produce a noticeable halo effect around the sun, resulting in saturated ghosting artifacts. Similarly, Figure 5 demonstrates persistent over-exposure issues across these methods. The proposed LACA and MSRSTB modules play important roles in resolving ghosting and saturated problems. The LACA module adaptively modulates the weights assigned to dif- ferent channels, thereby enabling a more sophisticated and effective equilibrium of information across channels. This dynamic weight strategy significantly alleviates motion blur and color saturation which improves the model performance. The MSRSTB module excels at integrating features extracted at various scales. The multi-scale Transformer component within it adaptively combines feature representations of mov- ing regions. It can handle motions from fine-grained local dis- placements to large-scale global movements across multiple scales. Moreover, it dynamically fuses regions with different exposure levels, thereby reducing the saturation problem.

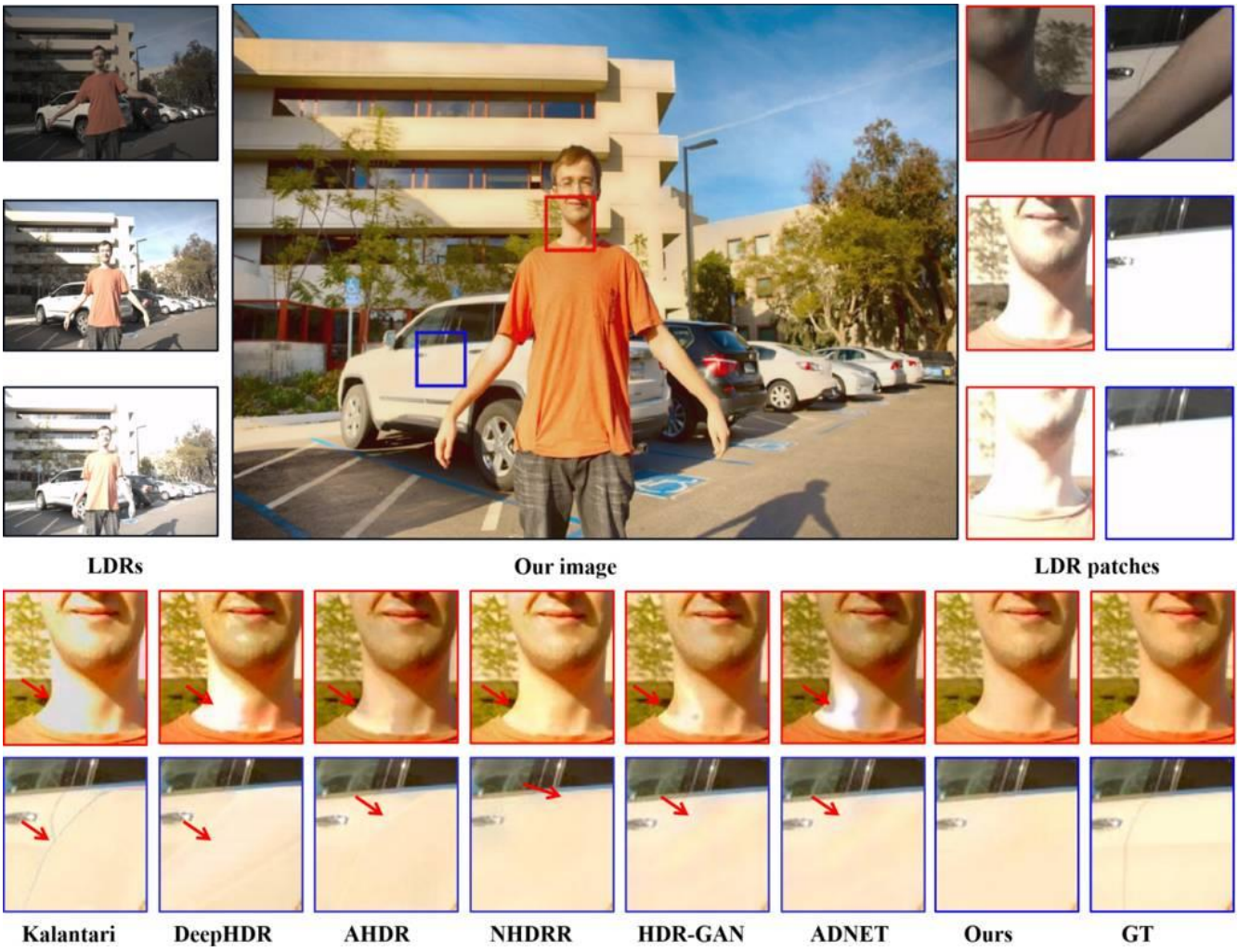

FIGURE 3. **Results of Kalantari's dataset.**

TABLE 1. The evaluation results on Kalantari's [17] and Hu's [73] datasets. The best and the second best results are highlighted in **Bold** and Underline, respectively.

Datasets	Models	Sen	Hu	Kalantari	DeepHDR	AHDRNet	NHDRR	HDR-GAN	ADNet	Ours
Kalantari	PSNR-μ	40.95	35.79	42.74	41.65	43.63	42.41	43.92	43.76	**44.35**
	PSNR-L	38.31	30.76	41.22	40.88	41.14	41.43	41.57	41.27	**42.20**
	SSIM-μ	0.9805	0.9717	0.9877	0.9860	0.9900	0.9887	0.9905	0.9904	**0.9916**
	SSIM-L	0.9726	0.9503	0.9848	0.9858	0.9862	0.9857	0.9865	0.9860	**0.9889**
	HDR-VDP-2	59.38	57.05	63.51	64.90	64.61	61.21	65.45	64.21	**66.03**
Hu	PSNR-μ	31.37	36.52	41.60	41.08	45.69	45.01	45.83	46.68	**47.35**
	PSNR-L	33.52	36.90	43.67	41.16	49.14	48.72	49.10	50.33	**50.92**
	SSIM-μ	0.9529	0.9823	0.9911	0.9868	0.9953	0.9941	0.9956	0.9906	**0.9957**
	SSIM-L	0.9632	0.9874	0.9937	0.9939	0.9976	0.9986	0.9980	0.9986	**0.9990**
	HDR-VDP-2	66.36	67.55	72.91	70.79	75.01	74.83	75.16	76.18	**76.83**

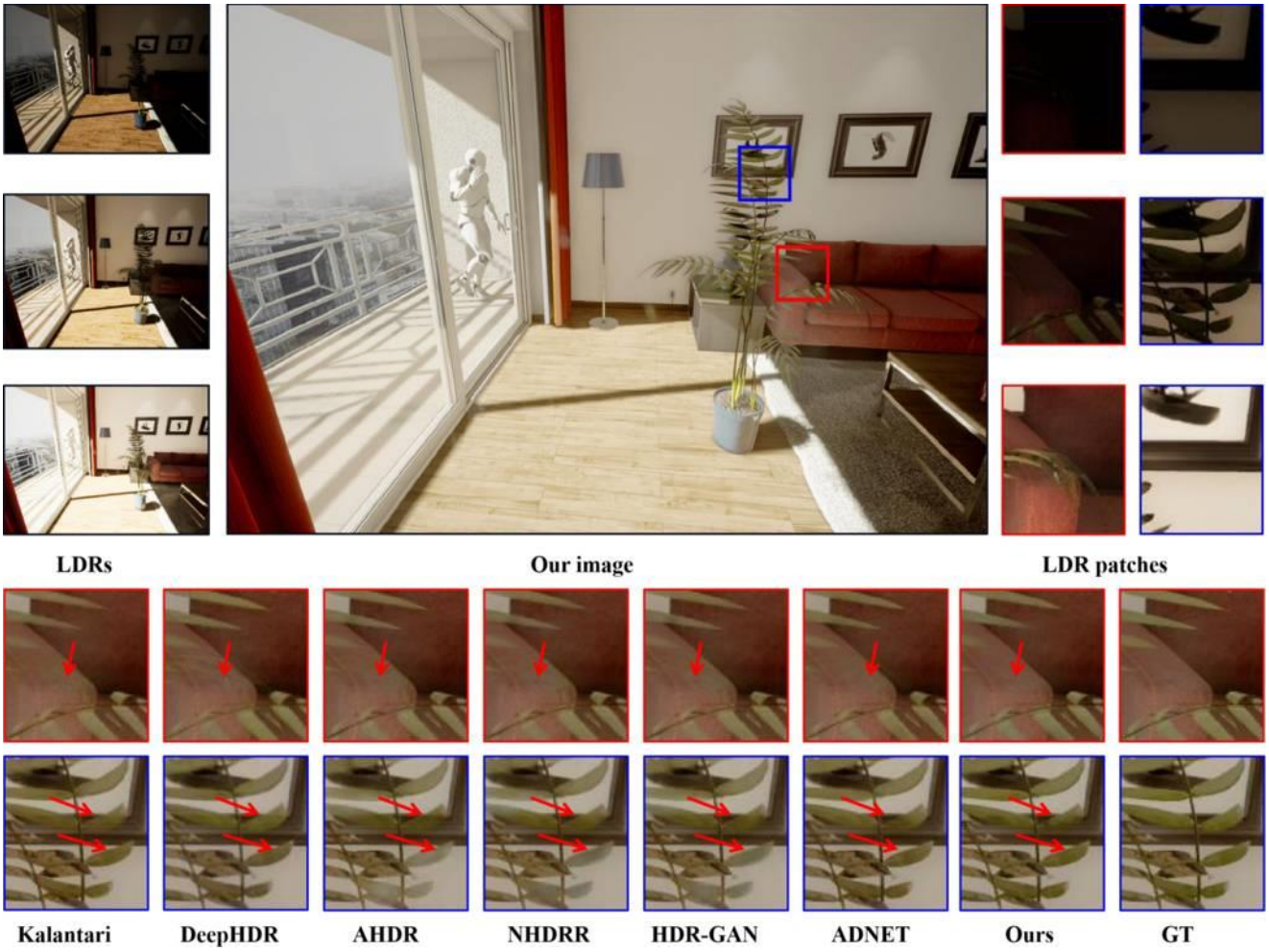

FIGURE 4. **Results of Hu's dataset.**

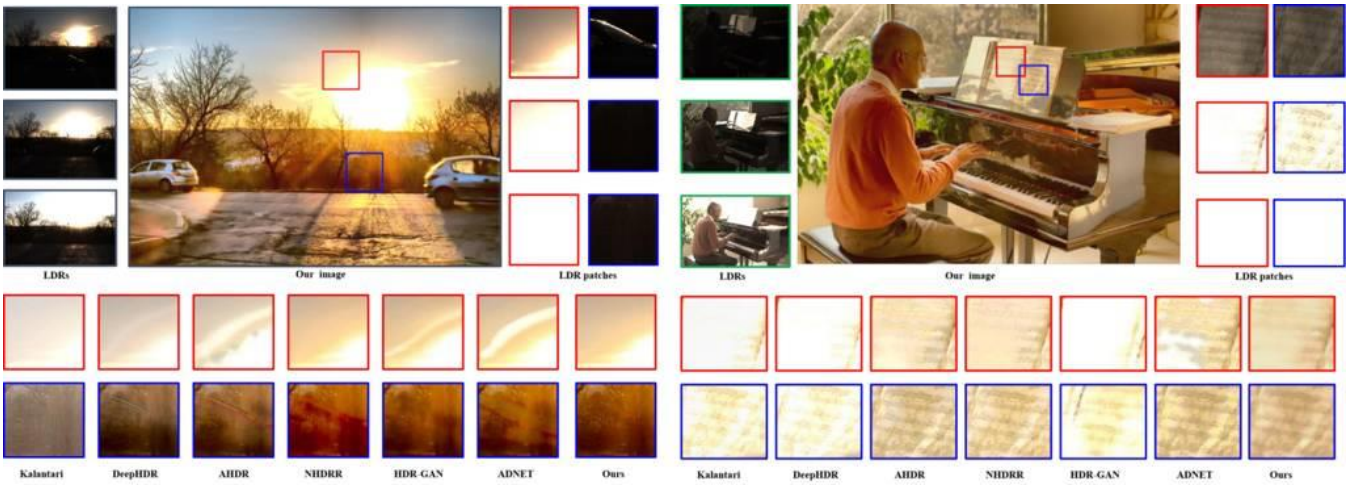

FIGURE 5. **Results of Tursun's and Sen's dataset.**

C. QUANTITATIVE EVALUATIONS

Table 1 displays the quantitative evaluations of the proposed method on two datasets. Remarkably, the proposed approach consistently outperforms existing state-of-the-art (SOTA) methods across all five evaluation metrics for both datasets. Specifically, on Kalantari's dataset [17], our method achieves significant improvements, exceeding the second-best methods by 0.43 dB and 0.63 dB in PSNR-μ and PSNR-L, respectively. Similarly, on Hu's dataset [73], it demonstrates even more pronounced enhancements, outperforming the runner-up by 0.67 dB in PSNR-μ and 0.59 dB in PSNR-L. In terms of SSIM-L and SSIM-μ, our method also significantly surpasses the second-best methods on Kalantari's and Hu's dataset. Our method achieves the best HDR-VDP-2 of all the compared approaches. The proposed LACA and MSRSTB modules contribute significantly to the model's performance. The LACA module adaptively adjusts channel weights, enabling a more refined balance of inter-channel information. This adaptive weighting mechanism effectively mitigates motion blur and color saturation issues, enhancing overall model performance. The MSRSTB module is designed to integrate multi-scale features efficiently. Its multi-scale Transformer component adaptively combines feature representations of moving regions, handling a wide range of motions from fine-grained local displacements to large-scale global movements. Additionally, it dynamically fuses regions with varying exposure levels, reducing saturation problems and further improving the model's effectiveness.

D. ABLATION STUDY

To assess the efficacy of the proposed components within our model, we devise multiple variants. The ablation study proceeds through the comparison of these model variants, aiming to isolate and quantify the individual contributions of each component.

TABLE 2. Ablation study on the network structure.

Model	PSNR-μ	PSNR-L	HDR-VDP-2
Baseline	43.62	41.03	62.30
Model1	43.85	41.43	64.97
Model2	44.02	41.86	65.54
Model3	44.27	42.12	65.96
Ours	44.35	42.20	66.03

- Baseline (AHDRNet). It is the vanilla AHDRNet. [75].
- Model1. We add a LACA module to AHDRNet.
- Model2. Based on Model1, we replace the DRDB fusion module in AHDRNet with RSTB module in SwinIR [70].
- Model3. Based on Model2, we replace the RSTB module with the proposed MSRSTB module.
- Ours. The full model of the proposed method. Based Model3, we add perceptual loss to the model.

LACA Module. In comparison with the Baseline presented in Table 2, Model 1 integrated with the LACA module demonstrates superior performance. It achieves a notable improvement, with a gain of 0.23 dB in PSNR-μ, 0.4 dB in PSNR-L, and 2.67 in HDR-VDP-2.This enhancement can be primarily attributed to the LACA module we proposed. It adaptively modulates the weights of different channels, which allows for a more sophisticated equilibrium of information across channels. By doing so, this adaptive weighting approach effectively alleviates motion blur and color saturation problems, thereby enhancing the overall performance of the model.

MSRSTB Module. As shown in Table 2, Model 3 which incorporates the MSRSTB module outperforms Model 1 significantly. Model 3 achieves PSNR-L of 44.27, PSNR-μ of 42.12 and HDR-VDP-2 of 65.96 compared to Model 1's 43.85, 41.43, and 64.97, respectively. This results in an increase of 0.69 in HDR-VDP-2, 0.69 in PSNR-μ, and 0.42 in PSNR-L. When we replace the MSRSTB module with the RSTB module, the results decreases 0.25db, 0.26db and 0.45 in terms of PSNR-L, PSNR-μ and HDR-VDP-2. It shows that the proposed MSRSTB module performs better than RSTB module. It is because the MSRSTB module is engineered to efficiently integrate multi-scale features. Its multi-scale Transformer component adeptly combines feature representations of moving regions, effectively managing motions that range from minute local displacements to extensive global movements. Moreover, by dynamically fusing regions with different exposure levels, the MSRSTB module mitigates saturation issues, thereby enhancing the model's overall effectiveness.

Loss function. To validate the efficacy of the perceptual loss, we train the model in two scenarios: with and without the perceptual loss term. The experimental results as presented in the Table 2 clearly demonstrate that incorporating the

perceptual loss significantly enhances the performance of our proposed model.

V. CONCLUSION

In this paper, we introduce two noveld modules. The LACA module adaptively modulates channel-wise weights across multiple scales, precisely balancing information among channels. This effectively suppresses ghosting artifacts and reduces color saturation, enhancing feature representation for HDR fusion. The MSRSTB leverages a multi-scale Transformer architecture to offer a large receptive field and dynamic weighting. It manages diverse motion patterns and integrates features in a coarse-to-fine hierarchical manner, efficiently handling regions with varying exposures. Consequently, it significantly reduces saturation and ghosting, enabling high-quality HDR image reconstruction in challenging scenarios. We also conduct comprehensive qualitative and quantitative evaluations which confirm that our proposed modules outperform existing state-of-the-art methods and improve the quality of HDR reconstruction.

REFERENCES

[1] E. Ashley, U. Matthew, and S. Richard, ''Seamless image stitching of scenes with large motions and exposure differences,'' in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2006, pp. 2498–2505.

[2] O. Gallo, N. Gelfand, W.-C. Chen, M. Tico, and K. Pulli, ''Artifact-free high dynamic range imaging,'' in *IEEE International Conference on Computational Photography (ICCP)*. IEEE,2009, pp. 1–7.

[3] G. Miguel, I. K. Kwang, T. James, and T. Christian, ''Automatic noise modeling for ghost-free hdr reconstruction,'' *ACM Transactions on Graphics*. vol. 36, pp. 1–10, 2013.

[4] T. Grosch, ''Fast and robust high dynamic range image generation with camera and object movement,'' in *IEEE Conference of Vision , Modeling and Visualization*, 2006.

[5] D. Sun, S. Roth, and M. J. Black, ''Secrets of optical flow estimation and their principles,'' in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2010, pp. 2432–2439.

[6] Q. Yan, D. Gong, P. Zhang, Q. Shi, J. Sun, I. Reid, and Y. Zhang, ''Multi-scale dense networks for deep high dynamic range imaging,'' in *IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019, pp. 41–50.

[7] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, ''Automatic differentiation in pytorch,'' 2017.

[8] J. Tumblin, A. Agrawal, and R. Raskar, ''Why I want a gradient camera,'' in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. I, 1 2005, pp. 103–110.

[9] Mann, Picard, S. Mann, and R. W. Picard, ''On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures,'' in *Proceedings of IS&T*, 1995, pp. 442–448.

[10] G. Miguel, A. Boris, W. Michael, T. Christian, S. Hans-Peter, and P. A. L. Hendrik, ''Optimal HDR reconstruction with linear digital cameras,'' in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 215–222.

[11] Debevec, E. Paul, Malik, and Jitendra, ''Recovering high dynamic range radiance maps from photographs,'' *Proc Siggraph*, vol. 97, pp. 369–378, 1997.

[12] E. Reinhard, G. Ward, S. Pattanaik, and P. E. Debevec, *High dynamic range imaging : acquisition, display, and image-based lighting*. Princeton University Press,, 2005.

[13] K. Jacobs, C. Loscos, and G. Ward, ''Automatic high dynamic range image generation of dynamic environments,'' *IEEE Computer Graphics and Applications*, vol. 28, no. 2, pp. 84–93, 2008.

[14] O. T. Tursun, A. O. Akyüz, A. Erdem, and E. Erdem, ''The state of the art in HDR deghosting: A survey and evaluation,'' *Comput. Graph. Forum*, vol. 34, no. 2, pp. 683–707, 2015.

[15] A. Srikantha and D. Sidib, ''Ghost detection and removal for high dynamic range images: Recent advances,'' *Signal Processing: Image Communication*, vol. 27, no. 6, pp. 650–662, 2012.

[16] H. Zimmer, A. Bruhn, and J. Weickert, ''Freehand HDR imaging of moving scenes with simultaneous resolution enhancement,'' in *Computer Graphics Forum*, 2011, pp. 405–414.

[17] N. K. Kalantari and R. Ramamoorthi, ''Deep high dynamic range imaging of dynamic scenes,'' *ACM Transactions on Graphics*, vol. 36, no. 4, pp. 1–12, 2017.

[18] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, ''Show, attend and tell: Neural image caption generation with visual attention,'' in *International Conference on Machine Learning (ICML)*, vol. 37, 2015, pp. 2048–2057.

[19] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, ''Generative image inpainting with contextual attention,'' *arXiv preprint arXiv:1801.07892*, 2018.

[20] I. K. Adam W Harley, Konstantinos G. Derpanis, ''Segmentation-aware convolutional networks using local attention masks,'' in *IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 5038–5047.

[21] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, ''Densely connected convolutional networks,'' in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708.

[22] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, ''Residual dense network for image super-resolution,'' in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 2472–2481.

[23] E. Reinhard, G. Ward, S. N. Pattanaik, and P. Debevec, *High dynamic range imaging, acquisition, display, and image-based lighting*, 2005.

[24] E. A. Khan, A. O. Akyuz, and E. Reinhard, ''Ghost removal in high dynamic range images,'' in *International Conference on Image Processing (ICIP)*, 2006, pp. 2005–2008.

[25] T. Jinno and M. Okuda, ''Motion blur free HDR image acquisition using multiple exposures,'' in *IEEE International Conference on Image Processing (ICIP)*, 2008, pp. 1304–1307.

[26] S. Raman and S. Chaudhuri, ''Reconstruction of high contrast images for dynamic scenes,'' *The Visual Computer*, vol. 27, no. 12, pp. 1099–1114, 2011.

[27] F. Pece and J. Kautz, ''Bitmap movement detection: HDR for dynamic scenes,'' in *Visual Media Production*, 2010, pp. 1–8.

[28] W. Zhang and W.-K. Cham, ''Gradient-directed multiexposure composition,'' *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2318–2323, 2012.

[29] Y. Heo, K. Lee, S. Lee, Y. Moon, and J. Cha, ''Ghost-free high dynamic range imaging,'' in *IEEE Conference on Asian Conference on Computer Vision (ACCV)*, 2011, pp. 486–500.

[30] J. Zheng, Z. Li, Z. Zhu, S. Wu, and S. Rahardja, ''Hybrid patching for a sequence of differently exposed images with moving objects,'' *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5190–201, 2013.

[31] C. Lee, Y. Li, and V. Monga, ''Ghost-free high dynamic range imaging via rank minimization,'' *IEEE signal processing letters*, vol. 21, no. 9, pp. 1045–1049, 2014.

[32] T.-H. Oh, J.-Y. Lee, Y.-W. Tai, and I. S. Kweon, ''Robust high dynamic range imaging by rank minimization,'' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 6, pp. 1219–1232, 2015.

[33] L. Bogoni, ''Extending dynamic range of mono-chrome and color images through fusion,'' in *IEEE International Conference on Pattern Recognition (ICPR)*, 2000, pp. 7–12.

[34] S. B. Kang, M. Uyttendaele, S. Winder, andR. Szeliski, ''High dynamic range video,'' *ACM Transactions on Graphics*, vol. 22, no. 3, pp. 319–325, 2003.

[35] P. Sen, K. K. Nima, Y. Maziar, D. Soheil, D. B. Goldman, and E. Shechtman, ''Robust patch-based HDR reconstruction of dynamic scenes,'' *ACM Transactions on Graphics*, vol. 31, no. 6, pp. 1–11, 2012.

[36] J. Hu, O. Gallo, K. Pulli, and X. Sun, ''HDR deghosting: How to deal with saturation?'' in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 1163–1170.

[37] D. Hafner, O. Demetz, and J. Weickert, ''Simultaneous HDR and optic flow computation,'' in *International Conference on Pattern Recognition (ICPR)*, 2014, pp. 2065–2070.

[38] M. Gharbi, J. Chen, J. T. Barron, S. W. Hasinoff, and F. Durand, ''Deep bilateral learning for real-time image enhancement,'' *ACM Transactions on Graphics*, vol. 36, no. 4, pp. 118–130, 2017.

[39] G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk, and J. Unger, ''HDR image reconstruction from a single exposure using deep cnns,'' *ACM Transactions on Graphics*, vol. 36, no. 6, pp. 178–193, 2017.
[40] Y. Endo, Y. Kanamori, and J. Mitani, ''Deep reverse tone mapping,'' *ACM Transactions on Graphics*, vol. 36, no. 6, pp. 1–10, 2017.
[41] J. Cai, S. Gu, and L. Zhang, ''Learning a deep single image contrast enhancer from multi-exposure images,'' *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 2049–2062, 2018.
[42] X. Yang, K. Xu, Y. Song, Q. Zhang, X. Wei, and L. Rynson, ''Image correction via deep reciprocating HDR transformation,'' in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
[43] K. R. Prabhakar, V. S. Srikar, and R. V. Babu, ''Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs,'' in *IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 4724–4732.
[44] Y. Tai, J. Yang, and X. Liu, ''Image super-resolution via deep recursive residual network,'' in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
[45] X. Glorot and Y. Bengio, ''Understanding the difficulty of training deep feedforward neural networks,'' *Journal of Machine Learning Research*, vol. 9, pp. 249–256, 2010.
[46] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, ''HDR-VDP-2:a calibrated visual metric for visibility and quality predictions in all luminance conditions,'' in *ACM Siggraph*, 2011, pp. 1–14.
[47] C. Liu, *Beyond pixels: exploring new representations and applications for motion analysis*. Massachusetts Institute of Technology, 2009.
[48] O. T. Tursun, A. O. Akyüz, A. Erdem, and E. Erdem, ''An objective deghosting quality metric for HDR images,'' *Comput. Graph. Forum*, vol. 35, no. 2, pp. 139–152, 2016.
[49] A. Serrano, F. Heide, D. Gutierrez, G. Wetzstein, and B. Masia, ''Convolutional sparse coding for high dynamic range imaging,'' *Comput. Graph. Forum*, vol. 35, no. 2, pp. 153–163, 2016.
[50] R. Wan, B. Shi, L.-Y. Duan, A.-H. Tan, and A. C. Kot, ''CRRN: Multi-scale guided concurrent reflection removal network,'' in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 05 2018.
[51] G. Ward, ''Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures,'' *Journal of Graphics Tools*, vol. 8, 2012.
[52] A. Tomaszewska and R. Mantiuk, ''Image registration for multi-exposure high dynamic range image acquisition,'' in *International Conference in Central Europe on Computer Graphics and Visualization*, 2007.
[53] F. Yu and V. Koltun, ''Multi-scale context aggregation by dilated convolutions,'' in *International Conference on Machine Learning (ICML)*, 2016.
[54] T. Mertens, J. Kautz, and F. V. Reeth, ''Exposure fusion,'' in *Pacific Conference on Computer Graphics and Applications*, 2007, pp. 382–390.
[55] B. Funt and L. Shi, ''The rehabilitation of MaxRGB,'' in *Color and Imaging Conference*, 2010, pp. 256–259.
[56] S. Ferradans, M. Bertalmío, E. Provenzi, and V. Caselles, ''Generation of HDR images in non-static conditions based on gradient fusion.'' in *International Conference on Vision Theory and Applications*, 2012, pp. 31–37.
[57] Q. Yan, Y. Zhu, Y. Zhou, J. Sun, L. Zhang, and Y. Zhang, ''Enhancing image visuality by multi-exposure fusion,'' *Pattern Recognition Letters*, vol. 127, pp. 66–75, 2019.
[58] Q. Yan, J. Sun, H. Li, Y. Zhu, and Y. Zhang, ''High dynamic range imaging by sparse representation,'' *Neurocomputing*, vol. 269, pp. 160–169, 2017.
[59] W. Zhang and W.-K. Cham, ''Gradient-directed multiexposure composition,'' *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2318–2323, 2011.
[60] K. Ma, L. Hui, Y. Hongwei, W. Zhou, M. Deyu, and Z. Lei, ''Robust multi-exposure image fusion: A structural patch decomposition approach,'' *IEEE Transactions on Image Processing*, vol. 26, no. 5, pp. 2519–2532, 2017.
[61] S. Wu, X. Jiarui, T. Yu-Wing, and T. Chi-Keung, ''Deep high dynamic range imaging with large foreground motions,'' in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 117–132.
[62] Y. Niu, W. Jianbin, L. Wenxi, G. Wenzhong, and W. L. Rynson, ''Hdr-gan: Hdr image reconstruction from multi-exposed ldr images with large motions,'' *IEEE Transactions on Image Processing*, vol. 30, pp. 3885–3896, 2021.
[63] K. R. Prabhakar, A. Susmit, K. S. Durgesh, A. Balraj, and B. R. Venkatesh, ''Towards practical and efficient high-resolution hdr deghosting with cnn,'' in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020, pp. 497–513.
[64] P. Xiong and C. Yu, ''Hierarchical fusion for practical ghost-free high dynamic range imaging,'' in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 4025–4033.
[65] Q. Ye, J. Xiao, K.-m. Lam, and T. Okatani, ''Progressive and selective fusion network for high dynamic range imaging,'' in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 5290–5297.
[66] Z. Liu, L. Wenjie, L. Xinpeng, R. Qing, J. Ting, H. Mingyan, F. Haoqiang, S. Jian, and L. Shuaicheng, ''Adnet: Attention-guided deformable convolutional network for high dynamic range imaging,'' in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshop*, 2021, pp. 463–470.
[67] J. Chen, Y. Zaifeng, C. Tsz Nam, L. Hui, H. Junhui, and C. Lap-Pui, ''Attention-guided progressive neural texture fusion for high dynamic range image restoration,'' *IEEE Transactions on Image Processing*, vol. 31, pp. 2661–2670, 2022.
[68] Z. L. Szpak, C. Wojciech, E. Anders, and V. D. H. Anton, ''Sampson distance based joint estimation of multiple homographies with uncalibrated cameras,'' *Computer Vision and Image Understanding*, vol. 125, pp. 200–203, 2014.
[69] G. Orazio, T. Alejandro, H. Jun, P. Kari, and K. Jan, ''Locally non-rigid registration for mobile hdr photography,'' in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshop*, 2015, pp. 49–56.
[70] J. Liang, C. Jiezhang, S. Guolei, Z. Kai, V. G. Luc, and T. Radu, ''Swinir: Image restoration using swin transformer,'' in *Proceedings of the IEEE/CVF International Conference on Computer Vision (CVPR) Workshop*, 2021, pp. 1833–1844.
[71] Z. Xia, X. Pan, S. Song, E. L. Li, and G. Huang, ''Vision transformer with deformable attention,'' in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 4794–4803.
[72] N. Park and K. Songkuk, ''How do vision transformers work,'' in *International Conference on Learning Representations (ICLR)*, 2022.
[73] J. Hu, G. Choe, Z. Nadir, O. Nabil, S.-J. Lee, H. Sheikh, Y. Yoo, and M. 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 (CVPR) Workshops*, 2020, pp. 516–517.
[74] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, ''Swin transformer: Hierarchical vision transformer using shifted windows,'' in *Proceedings of the IEEE/CVF International Conference on Computer Vision (CVPR)*, 2021, pp. 10 012–10 022.
[75] Q. Yan, G. Dong, S. Qinfeng, v. d. H. Anton, S. Chunhua, R. Ian, and Z. Yanning, ''Attention-guided network for ghost-free high dynamic range imaging,'' in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 1751–1760.
[76] Q. Yan, Z. Lei, L. Yu, Z. Yu, S. Jinqiu, S. Qinfeng, and Z. Yanning, ''Deep hdr imaging via a nonlocal network,'' *IEEE Transactions on Image Processing*, vol. 29, pp. 4308–4322, 2020.
[77] Q. Yan, S. Zhang, W. Chen, Y. Liu, Z. Zhang, Y. Zhang, J. Q. Shi, and D. Gong, ''A lightweight network for high dynamic range imaging,'' in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) workshop*, 2022, pp. 824–832.
[78] Q. Yan, D. Gong, J. Q. Shi, A. van den Hengel, C. Shen, I. Reid, and Y. Zhang, ''Dual-attention-guided network for ghost-free high dynamic range imaging,'' *International Journal of Computer Vision*, vol. 130, no. 1, pp. 76–94, 2022.
[79] A. Mittal, R. Soundararajan, and A. C. Bovik, ''Making a ''completely blind'' image quality analyzer,'' *IEEE Signal processing letters*, vol. 20, no. 3, pp. 209–212, 2012.
[80] A. Mittal, A. K. Moorthy, and A. C. Bovik, ''No-reference image quality assessment in the spatial domain,'' *IEEE Transactions on image processing*, vol. 21, no. 12, pp. 4695–4708, 2012.
[81] M. R. Luo, G. Cui, and B. Rigg, ''The development of the cie 2000 colour-difference formula: Ciede2000,'' *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Societyfor Color, Color Science Association*

of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur, vol. 26, no. 5, pp. 340–350, 2001.
[82] Q. Yan, D. Gong, J. Q. Shi, A. van den Hengel, J. Sun, Y. Zhu, and Y. Zhang, ''High dynamic range imaging via gradient-aware context aggregation network,'' *Pattern Recognition*, vol. 122, p. 108342, 2022.
[83] Q. Yan, B. Wang, P. Li, X. Li, A. Zhang, Q. Shi, Z. You, Y. Zhu, J. Sun, and Y. Zhang, ''Ghost removal via channel attention in exposure fusion,'' *Computer Vision and Image Understanding*, vol. 201, p. 103079, 2020.
[84] Z. Liu, Y. Wang, B. Zeng, and S. Liu, ''Ghost-free high dynamic range imaging with context-aware transformer,'' in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022, pp. 344–360.
[85] Y. S. Heo, K. M. Lee, S. U. Lee, Y. Moon, and J. Cha, ''Ghost-free high dynamic range imaging,'' in *Asian Conference on Computer Vision*. Springer, 2010, pp. 486–500.
[86] T.-H. Min, R.-H. Park, and S. Chang, ''Histogram based ghost removal in high dynamic range images,'' in *2009 IEEE International Conference on Multimedia and Expo*. IEEE, 2009, pp. 530–533.
[87] K. R. Prabhakar and R. V. Babu, ''Ghosting-free multi-exposure image fusion in gradient domain,'' in *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2016, pp. 1766–1770.
[88] S. Wu, S. Xie, S. Rahardja, and Z. Li, ''A robust and fast anti-ghosting algorithm for high dynamic range imaging,'' in *2010 IEEE International Conference on Image Processing*. IEEE, 2010, pp. 397–400.
[89] W. Zhang and W.-K. Cham, ''Reference-guided exposure fusion in dynamic scenes,'' *Journal of Visual Communication and Image Representation*, vol. 23, no. 3, pp. 467–475, 2012.
[90] J. Hu, O. Gallo, and K. Pulli, ''Exposure stacks of live scenes with hand-held cameras,'' in *European Conference on Computer Vision*. Springer, 2012, pp. 499–512.
[91] S. B. Kang, M. Uyttendaele, S. Winder, andR. Szeliski, ''High dynamic range video,'' *ACM Transactions on Graphics (TOG)*, vol. 22, no. 3, pp. 319–325, 2003.
[92] G. Ward, ''Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures,'' *Journal of graphics tools*, vol. 8, no. 2, pp. 17–30, 2003.
[93] Q. Yan, B. Wang, L. Zhang, J. Zhang, Z. You, Q. Shi, and Y. Zhang, ''Towards accurate hdr imaging with learning generator constraints,'' *Neurocomputing*, vol. 428, pp. 79–91, 2021.
[94] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, ''Ghostnet: More features from cheap operations,'' in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1580–1589.
[95] E. Pérez-Pellitero, S. Catley-Chandar, R. Shaw, A. Leonardis, R. Timofte *et al.*, ''NTIRE 2022 challenge on high dynamic range imaging: Methods and results,'' in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2022.
[96] X. Chen, Y. Liu, Z. Zhang, Y. Qiao, and C. Dong, ''Hdrunet: Single image hdr reconstruction with denoising and dequantization,'' in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 354–363.
[97] Q. Yan, Y. Zhu, and Y. Zhang, ''Robust artifact-free high dynamic range imaging of dynamic scenes,'' *Multimedia Tools and Applications*, vol. 78, no. 9, pp. 11 487–11 505, 2019.

ZEBIN WEN currently enrolled at a university in the Greater Bay Area, has established a strong research focus on visual recognition and deep learning. As the first author, he has authored over 10 publications in leading academic journals and conferences within the field. Additionally, he has actively participated in two provincial-level scientific research projects, contributing to cutting-edge advancements in his discipline. In terms of intellectual property, he holds one domestic utility model patent, one invention patent, and four software copyrights, demonstrating his capacity for translating research into practical innovations.