This CVPR Workshop paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Recursive Multi-Exposure Alignment with Spatiotemporal Decoupling for Efficient Burst HDR and Restoration

Tianheng Qiu[*1,2] Qi Wu[*3] Yuchun Dong[3] Shenglin Ding[3] Xuan Huang[†2] Hu Wei[2] Guanghua Pan[†3]
[1]University of Science and Technology of China
[2]Hefei Institutes of Physical Science, Chinese Academy of Sciences
[3]Imvision

Abstract

Computing high dynamic range (HDR) RGB output from multi-frame low dynamic range (LDR) RAW input is a challenging task because it requires solving multiple subtasks including multi-frame fusion of different exposures, image restoration including denoising, deblurring, HDR imaging, and modeling RAW to RGB mapping. Solving the problem using a unified model is more difficult as these tasks need to be considered simultaneously.In this paper, in order to construct a generalized efficient Burst HDR and Restoration method, we propose the Recursive Multi-Exposure Alignment with Spatiotemporal Decoupling (RASD) algorithm. Specifically, in order to address the information discrepancy between multi-exposure data, we propose a recursive flow-guided alignment module based on multi-exposure alignment, which is used to provide more accurate multi-frame alignment. In addition, we introduce a spatiotemporal decoupling strategy to train the alignment and restoration tasks in stages to prevent possible optimization conflicts introduced between multiple tasks. Extensive experiments show that our proposed method obtains state-of-the-art performance, and we are the winner in the NTIRE 2025 Efficient Burst HDR and Restoration Challenge.

1. Introduction

HDR images are able to better simulate the human visual system, and the computational HDR which obtains results from multiple LDR frame inputs has gained a lot of attention in the last few years due to the fact that the hardware equipment for HDR imaging is too expensive. In addition, RAW images which provide more information to HDR imaging compared to RGB can give stronger ability to cope with inter-frame shifts, motion blur, etc., since they are not processed by non-linearity. Therefore, HDR imaging in the RAW domain has received a lot of attention over time [1, 23, 34, 43, 46]. In fact, burst RAW inputs with consecutive time series have multiple different degradations, and most of the previous work has been devoted to solving single restoration tasks, such as denoising [2, 12, 15], deblurring [5, 13, 19], and HDR imaging [17, 24, 35, 40], which fail to provide a unified solution and are limited by the multiple degradations present in real-world scenarios.

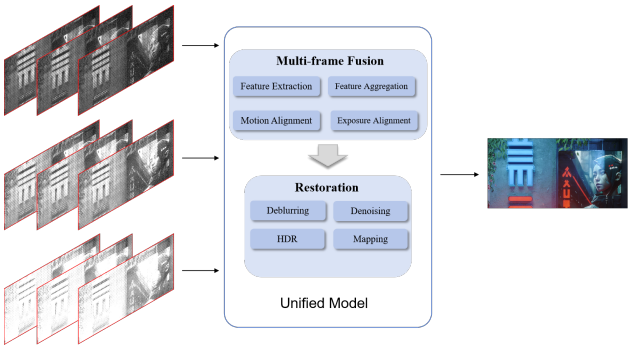


Figure 1. Brief task description. We need to construct an HDR RGB output from 9 frames of LDR RAW input, and mainly need to solve multi-frame fusion and restoration tasks.

Recently, [45] proposed Bracket Image Restoration and Enhancement (BIRE), the task of which can be summarized as shown in Fig. 1. Broadly speaking, this task can be divided into two sub-tasks, one is the fusion and feature extraction of multiple frames, the key lies in how to effectively extract the features of multiple input frames and efficiently perform the fusion, and the second is the restoration of the fused alignment features, such as denoising, deblurring, the key lies in how to design the restoration model to cope with a variety of degradations. Influenced by this, many works [7, 39, 42] started to focus on HDR imaging and unified restoration of burst RAW inputs.

However, limited by the number of input frames of the dataset and the intensity of degradation, it is still unable to produce much detail comparable to the real world. Com-

[*]These authors contributed equally to this work.
[†]Corresponding authors

pared to [45], the dataset provided by [21] has stronger noise simulation, more input frames, and more difficult data degradation, and its challenges come from two main sources: one is the information differences caused by brightness, noise, saturation, and dead-black regions between multi-exposure data domains, and the other is the disparity between alignment and restoration in the temporal and spatial domains. Ignoring these differences and contradictions would make the entire joint task inefficient.

Specifically, to solve the first issue, we first use the luminance prior for multi-exposure alignment. Unlike the previous method, we perform truncation-free luminance alignment toward the longest exposure to prevent alignment errors caused by too many saturated pixels(In a multi-exposure sequence, for irreversible numerical truncation, the probability of saturation occurring is much higher than that of dead black.). Then, based on the idea of divide-and-conquer, we use a recursive alignment scheme to align frames within the same exposure first, and then start the alignment recursively from short exposure. In addition, in order to provide more implicit prior to the alignment module, we introduce the pre-computed optical flow into the alignment module to produce more accurate alignment features and more efficient fusion features. To address the second issue, we propose a two-stage training strategy to decouple spatiotemporal learning and prevent performance loss caused by conflicting optimization objectives. First, the training of the alignment module is performed in the alignment stage, where we maximize the ability of the alignment module by using only a very small restoration module to construct the end-to-end loss. Then, in restoration stage, we freeze the alignment module in the previous stage and switch to a more complex restoration module to build clean HDR images based on the alignment features.

In summary, our contributions are as follows:

- We propose a truncation-free multi-exposure alignment strategy, which no longer aligns to short exposures, but aligns to the more information-rich long exposure frames, effectively providing for subsequent alignment and restoration.
- We propose recursive flow-guided alignment, which utilizes the recursive idea of partitioning each exposure group to simplify the complexity of the task and effectively improve the accuracy of the alignment.
- We propose a temporal spatial decoupling strategy, which decomposes the alignment that focuses more on the time domain and the restoration that focuses more on the spatial domain in two phases, preventing the conflict of optimization objectives and effectively improving the performance of alignment and restoration.
- We are the winner of the NTIRE 2025 Efficient Burst HDR and Restoration challenge, with a PSNR of 43.22 dB on the test set, outperforming the second place by 0.47 dB, and with a number of parameters less than 30 M and FLOPs less than 4 T.

2. Related Work

Single-frame Restoration Single-image restoration has been widely explored, with methods evolving from traditional model-based approaches to deep learning techniques. Early works leveraged handcrafted priors, such as sparse coding [10] and total variation regularization [3], to restore degraded images. With the advancement of deep learning, CNN-based models like NAFNet [4] introduced a simplified yet effective network design, achieving competitive results in image restoration. Meanwhile, transformer-based approaches, such as Restormer [44], have demonstrated superior performance by leveraging self-attention mechanisms to model long-range dependencies efficiently. More recently, diffusion models [18, 37] have been explored for image restoration, pushing the state of the art in handling complex degradations. Despite these advances, challenges remain in balancing restoration quality and computational efficiency.

Multi-frame Restoration Burst image restoration methods, such as multi-frame super-resolution approaches [11, 16, 36] and denoising techniques [14, 20, 27, 28] have largely focused on precise frame alignment and feature fusion. EBSR [25] employs explicit PCD alignment, while BSRT [26] uses optical flow-guided DCN along with a Swin Transformer for feature extraction. AFCNet [29] enhances feature extraction with Restormer. Both BIPNet [8] and Burstormer [9] adopt an implicit, shared feature alignment approach. Additionally, some kernel-based methods [30, 38] aim to establish a stable mapping model for better inter-frame feature processing. However, these approaches tend to overlook the distinct focuses of the two tasks—alignment primarily aggregates temporal information, whereas image restoration concentrates on spatial information. Training these tasks in an end-to-end manner can dilute their ability to effectively aggregate spatiotemporal features.

HDR Image Synthesis Existing HDR methods [17, 24, 35, 40] focus more on the design of data and network structures for fusion and deghosting. Some approaches [24, 35, 40] enhance the network's processing capability during the fusion stage using attention mechanisms, while others consider the impact of noise on HDR [22]. The latest TM-RNet [45] takes into account more complex data degradations affecting HDR, including noise and blur. However, these methods either ignore the degradation of pre-alignment accuracy caused by exposure differences or fail to perform a relatively reasonable luminance alignment modulation(aligns to short exposures), resulting in the inability to properly align over-saturated regions.

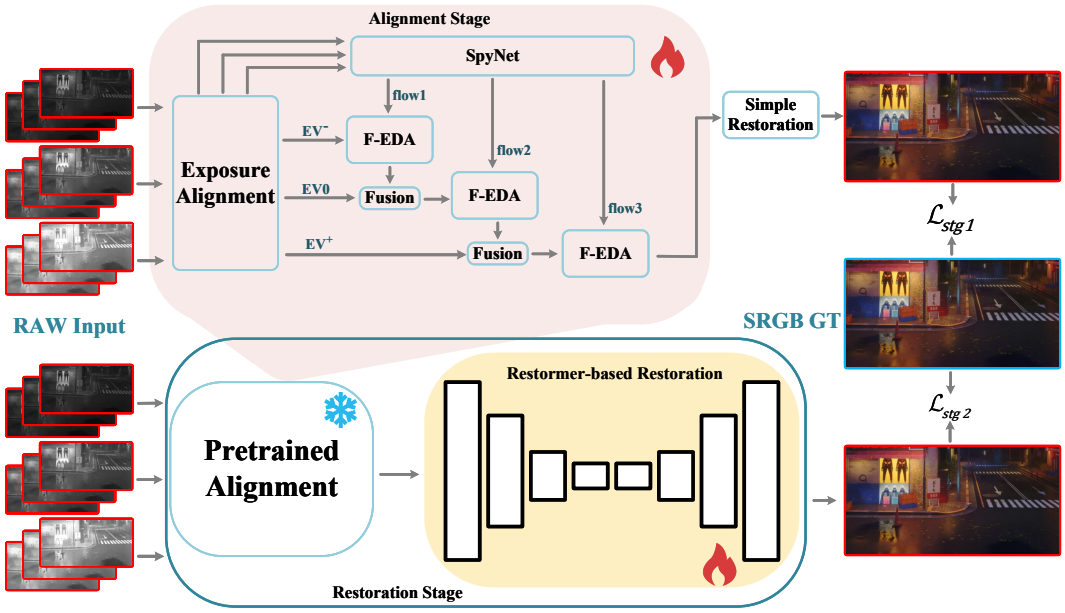


Figure 2. The overall architecture of our proposed algorithm RASD. Our algorithm is executed in two stages, the first stage is Alignment Stage, which is mainly used for efficient fusion and extraction of features, and its loss is computed as L_{stg1}, and the second stage is restoration stage, which is mainly used for the restoration of multiple degradations, and its loss is computed as L_{stg2}.

3. Proposed Method

We propose **R**ecursive Multi-Exposure **A**lignment with **S**patiotemporal **D**ecoupling (RASD), for solving end-to-end burst HDR and Restoration. Specifically, we divide the optimization process into alignment and restoration. In the alignment stage, we propose Multi-Exposure Alignment and Recursive Flow-based Enhanced Deformable Alignment for efficient feature alignment. In the restoration stage, we freeze the pre-trained alignment module and utilize the aligned features for the learning of the restoration module to cope with a variety of unknown degradations and enhance the overall resolution of the algorithm.

3.1. Multi-Exposure Alignment

Differences in feature representations between the luminance of multi-exposure images can degrade alignment accuracy, and previous methods [24, 31, 40, 41] used luminance normalization to align frames to the lowest exposure frame, then concatenate it with its gamma-transformed image. This process can be represented as shown below:

$$X^j_{EV_i} = Concat\left(\frac{X^j_{EV_i}}{EV_i/EV_1}, (\frac{X^j_{EV_i}}{EV_i/EV_1})^\gamma\right) \quad (1)$$

where EV_i means the exposure value in each level, and j means the index of frames in the current exposure value.

As the exposure time increases, the proportion of saturated pixels becomes larger, and the saturated pixels of the long-exposure image truncate much of the effective information when aligned to the short-exposure image because saturation is irreversible. It is also important to note that in a multi-exposure sequence, during the luminance alignment process, the probability of saturation in long exposure frames is significantly higher than the occurrence of dead black in short exposure frames under irreversible conditions. As shown in Fig. 3, the aligned long-exposure image has a lower contrast compared to the short-exposure image, which is caused by the saturated pixel truncation. Based on this finding, we consider reversing the previous alignment direction which aligns the short exposure to the long exposure as follows:

$$X^j_{EV_i} = X^j_{EV_i} \times \frac{EV_{max}}{EV_i} \quad (2)$$

We do not perform any truncation on $X^j_{EV_i}$, providing the alignment module with more search space. In this process, since we do not perform RAW to RGB conversion, we also disregard the subsequent ISP process (e.g., gamma transform). As shown in Fig. 3c, the image obtained by aligning accordingly has better luminance consistency.

3.2. Recursive Flow-based Enhanced Deformable Alignment

The key challenge in cross-exposure alignment is the difference in noise, saturation, and dead-black regions. Inspired by the idea of divide-and-conquer algorithm, unlike one-step alignment or parallel alignment, we recursively align



(a) Without EV Align (b) Align to EV- (c) Align to EV+

Figure 3. Different exposure alignment results (left) with corresponding optical flow alignment results (right). (a) indicates that no exposure alignment is performed and an abnormal alignment result occurs for different exposures. (b) indicates align to short exposure, long exposure images have bad alignment results due to too many saturated pixels. (c) indicates align to long exposure which we used, and all three different exposure images showed superior alignment results.

neighboring exposures as shown in Eq. 3.

$$y_i = \begin{cases} \xi_i(x^j_{EV_i}), & i = 0 \\ \xi_i(Concat(y_{i-1}, x^j_{EV_i})), & i < i_{max} \end{cases} \tag{3}$$

where ξ_i represents the alignment module currently exposure EV_i, and each alignment module contains a Flow-based Enhanced Deformable Alignment (F-EDA) module.

Each ξ_i performs intra alignment within the same exposure and then recursively aligns to the next stage, which can effectively alleviate the above problem. Because the recursive module is not considering the information differences between multiple exposure groups, and only needs to focus on the alignment between the output features of the previous stage and the features of this exposure group, so that each alignment module apportions the original solution space, and thus has a stronger alignment capability.

As shown in Fig. 2, EV- is used as the reference frame, and after alignment and aggregation of the same exposure information through F-EDA, it serves as the reference frame for alignment and aggregation of EV0 information in a recursive manner. Following [9], we additionally utilize the optical flow generated by the spynet [33] in the offset calculation as follows:

$$\mathcal{O}_i = Conv_{off}\left(f_i^{ref}, f_i^{nbr}, f_i^{flow}\right) \tag{4}$$

where f_i^{ref}, f_i^{nbr}, f_i^{flow} represent feature of reference frame, other frame and flow respectively, i means that we take EV_i alignment as example.

3.3. Spatiotemporal Decoupling Strategy

Essentially, we need to build an end-to-end Multiple-in-one model with the ability to handle tasks including but not limited to multi-frame alignment, denoising, deblurring, HDR imaging, RAW to RGB mapping. Compared to alignments which focus on the temporal domain, restoration prioritizes spatial information, and there is a conflict in the optimization space between the two, which leads to a loss of accuracy in joint training.

Based on this consideration, we design spatiotemporal decoupling which proposes a two-stage training strategy. As shown in Fig. 2, in the align stage (stage.1), the model focuses on the effective alignment between different frames without paying attention to the high-frequency information of the spatial level. Therefore, we used only a few layers of restormer blocks to ensure that the ability of align modules is maximized.

In contrast, the restoration stage (stage.2) uses the full restormer architecture to maximize the overall restoration of the model to potential blurring, noise, etc., and to provide a RAW to RGB mapping.

For the choice of restoration module, instead of using a more lightweight CNN model such as NAFNet, we used a transformer-based model which can model global relationships. The reason for this is when we defined the two-stage task as a pixel-to-pixel restoration task, the CNN model is less fault-tolerant for the alignment module, whereas the transformer-based model is more inclusive due to its ability to focus on the global information. Based on complexity considerations, we choose Restormer, an efficient architecture for applying transposed attention, and we achieve efficient restoration by reducing the number of blocks and model width.

3.4. Loss Function

Both for align stage loss L_{stg1} and restoration stage loss L_{stg2}, we just use PSNR loss [4] as follows:

$$\mathcal{L}(X, Y) = -PSNR(M(X), Y) \tag{5}$$

where X means the burst input RAW images, Y means the GT RGB image, M means the training model parameters, and $PSNR = 10 \times log_{10}\left(\frac{MAX^2}{MSE}\right)$.

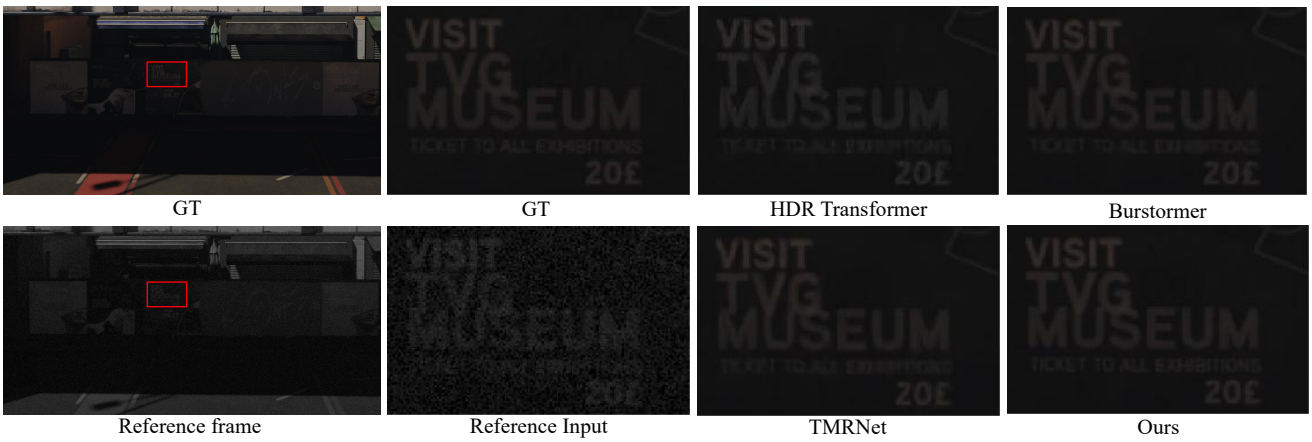


Figure 4. Visualization results for dark text details. Our proposed algorithm achieves the best results, recovering text with clear outlines. Zooming in allows better viewing.

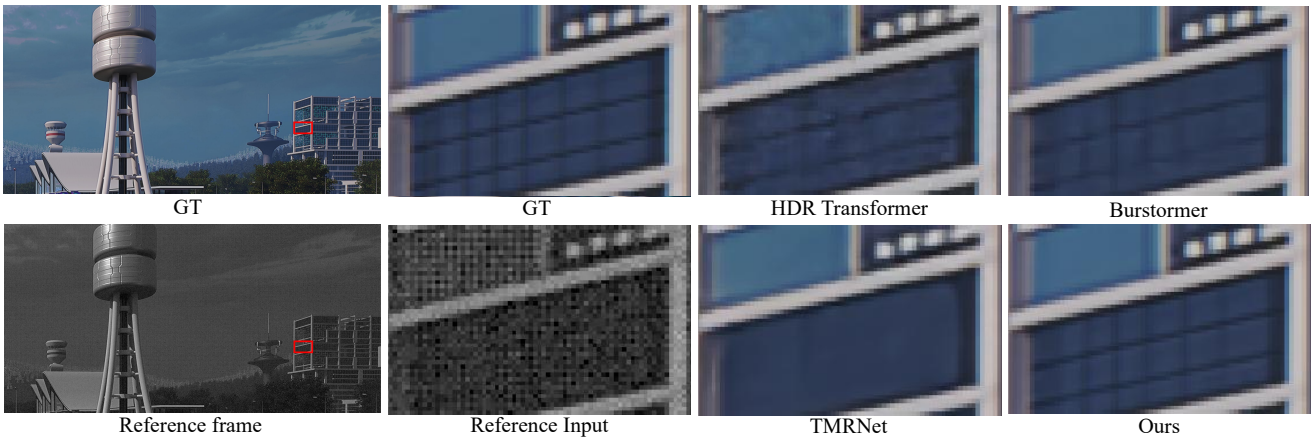


Figure 5. Visualization results for building details. Our proposed algorithm restores clear boundaries of windows and contains no other degradation. Zooming in allows better viewing.

4. Experimental Results

4.1. Dataset

We use a dataset with the resolution of $768(H) \times 1536(W)$ provided by NTIRE 2025 Efficient Burst HDR and Restoration Challenge. The training set contains 300 scenes and provides multi-frame RAW images and ground truth (GT) RGB images, while the validation and test sets contain 20 scenes with hidden GT. In particular, each scene contains three different exposure times (low, middle and high), and each exposure contains three frames of input with a variety of different degradations. The first frame of the short exposure is aligned with the GT, and the rest of the frames have a different degree of shift from the GT.

4.2. Implementation Details

During training we use AdamW optimizer ($\beta_1 = 0.9$ and $\beta_2 = 0.9$) with 1000 training epochs each for the alignment stage and the restoration stage. The initial value of the learning rate was 8×10^{-4} and is updated by a cosine annealing schedule and then restarted at half the learning rate in the second stage. The training patch size is set to 128 × 128 pixels. For data augmentation, we only use the bayeraug [22] for random flips and transposition. In the testing phase, we use TLC [6] for full image inference, and in our approach, the addition of TLC brings no complexity gain. All methods at all stages use 9 frames of RAW images as input and construct RGB outputs. All experiments are conducted with Pytorch-2.1.1 [32] and cuda-11.8 on 8 Nvidia Tesla V100 GPUs.

4.3. Comparison with State-of-the-Art Methods

We compared our method with other advanced methods and evaluate the quality of restored images with PSNR and SSIM, and measure the complexity of the model with the number of parameters and FLOPs.

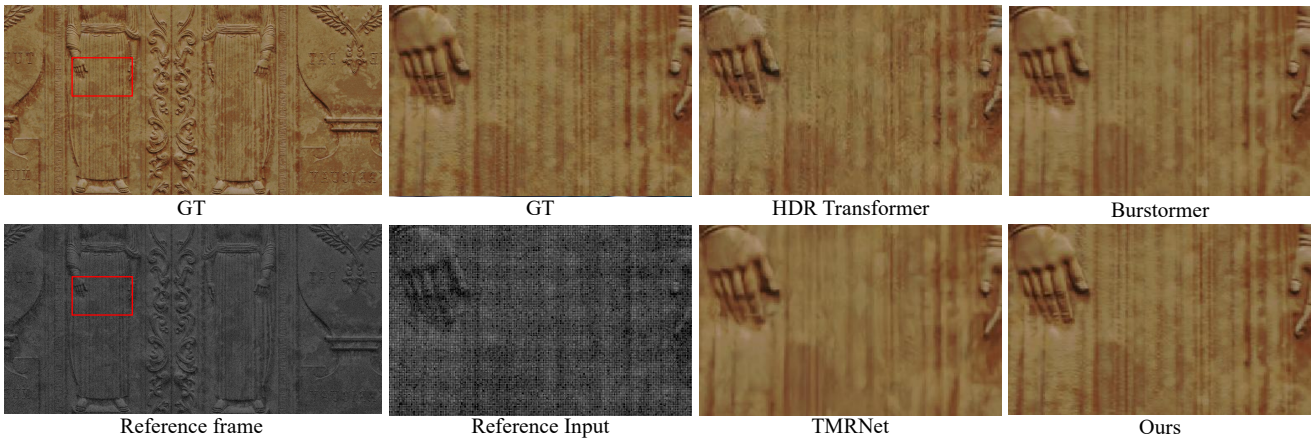


Figure 6. Visualization results for brick details. Our proposed algorithm yields the richest details, especially for the restoration of the hand in the upper right corner. Zooming in allows better viewing.

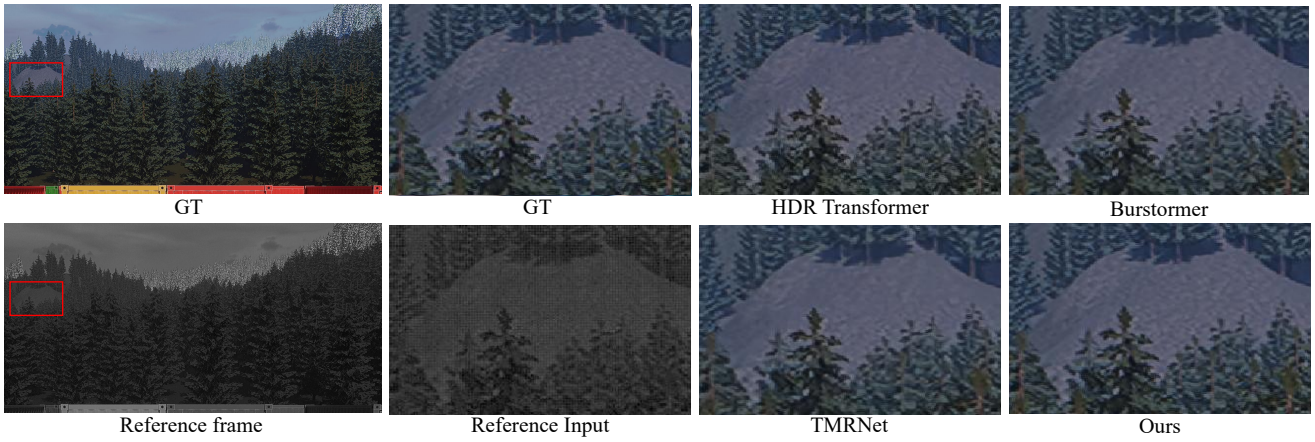


Figure 7. Visualization results for wild details. Our proposed algorithm can greatly cope with wild scenes with complex illumination and texture. Zooming in allows better viewing.

Method	PSNR	SSIM	Params (M)	FLOPs (T)
BSRT [26]	33.50	0.942	4.636	14.44
Burstormer [9]	42.037	0.978	3.138	22.531
HDR-Trans [24]	37.09	0.959	**3.036**	3.691
TMRNet [45]	38.59	0.963	20.966	11.025
Ours*	42.43	0.980	3.949	**2.695**
Ours	**43.38**	**0.983**	29.051	3.965

Table 1. Quantitative comparison results. Our proposed algorithm achieves state-of-the-art performance, and * indicates that only the alignment stage is used.

Quantitative Results. Since the GT images in the validation and test sets are hidden, we selected some images with complex textures and rich scene variations in the training set as the validation set. Tab. 4 shows the results of the quantitative comparison, where we pick several widely used burst fusion and restoration methods for the comparison test, including TMRNet [45], BSRT [26], HDR-Tran [24], and Burstormer [9]. As shown in Tab. 4, our method obtains SOTA PSNR and SSIM on this new dataset, outperforming the ranked second Burstormer [9] with 1.34 dB higher PSNR and 0.0005 higher SSIM, proving the excellent performance of our proposed method. Also, we show that the results trained using only the alignment stage can still outperform the previous best Burstormer by 0.39 dB with minimal complexity with the parameter of 3.949 M and the FLOPs of 2.695 T. The results of quantitative comparisons with various advanced algorithms demonstrate the superior performance of our proposed algorithm, both in terms of accuracy and complexity metrics.

Visualization Results. We selected several typical scenes from multi-frame image restoration and HDR evaluation for subjective comparison, such as dark text details, building details, rock details, and wild details. Fig. 4 demonstrates

Team name	PSNR↑	SSIM↑	Params (M)	FLOPs (T)	Time (s)
ImvisionAI	**43.22**	**0.992**	29.051	3.965	7.12
DeepTrans	42.75	0.991	12.839	3.864	0.24
SimonLarsen	41.38	0.989	18.651	1.685	0.74
E_Group	40.64	0.988	8.386	0.288	0.62
CidautAI	37.51	0.979	3.346	0.279	0.064
X-L	33.85	0.934	1.73	0.346	0.088
UNet	31.29	0.960	7.701	0.666	-

Table 2. The final results of participants. Our team ImvisionAI achieved the highest PSNR value of 43.22 dB using maximal resources close to the competition restriction. The inference time represented in the rightmost column is measured by NVIDA H100 GPU.

the restoration ability of our proposed algorithm for dark text. It can be seen that our proposed algorithm obtains the best text restoration in the selected region with clear boundaries, especially on the small strings of text that are hidden underneath. Fig. 5 demonstrates the restoration ability of our proposed algorithm for the building details. It can be seen that our proposed algorithm restores clearly visible windows in the selected region and eliminates all degradation. In contrast, other algorithms either do not restore the window texture well (e.g., TMRNet completely smoothes out the window details) or fail to do so by removing all the degradation (e.g., HDR trans restores the image with a not-insignificant amount of noise residue). Fig. 6 demonstrates the ability of our proposed algorithm to restore the brick texture details. It can be seen that our proposed algorithm restores the best and richest brick texture. In the upper left corner of the selected region, the other algorithms show more or less weak rendering of the hand painting, while our algorithm performs well overall and does not have any off-color issues. Fig. 7 demonstrates the restoration ability of our proposed algorithm for wild scenes. It can be seen that our proposed algorithm is optimal for texture restoration of hills and trees, proving that the algorithm can cope with field scenes with complex illumination and texture, as well as fine-grained HDR imaging.

The outstanding performance in restoring textures across these broad scene categories demonstrates the effectiveness of our approach in enhancing the stability and generalization of existing baselines.

4.4. NTIRE 2025 Efficient Burst HDR and Restoration Challenge Result

Our proposed algorithm RASD achieves championship performance on the NTIRE 2025 Efficient Burst HDR and Restoration Challenge, achieving a PSNR of 43.22 dB on the test set, outperforming the second place by 0.47 dB. The results of the Challenge are shown in Tab. 2, which demonstrates the advanced performance of RASD.

Parallel	Recursive	F-EDA	EV-Align	PSNR	SSIM
✓				42.04	0.976
	✓			42.25	0.977
	✓	✓		42.24	0.977
	✓		✓	42.44	0.979
	✓	✓	✓	43.00	0.982

Table 3. Ablation study on components of the proposed methods. We set the baseline network to use parallel in its entirety, and models that do not use F-EDA will represent simple alignment using EDA.

Method	PSNR	SSIM	Params (M)	FLOPs (T)
Align Stage	42.43	0.980	3.949	2.695
+ Restoration Stage	43.38	0.983	29.051	3.965

Table 4. Comparison of the various metrics in the two stages. The restoration stage must be based on the alignment stage, and the complexity is calculated in the same way as in [21].

Method	Params (M)	FLOPs (T)	time (s)
w/ Parallel	26.201	2.866	1.93
w/ Recursive	27.607	3.144	2.74

Table 5. Complexity of parallel alignment vs. recursive alignment. Inference time is computed for 9 frames of input at 768×1536 resolution and synchronization time is used.

4.5. Analysis and Discussion

In this section, we provide a more in-depth analysis and show the contribution of each component of the proposed method. Tab. 3 shows the detailed ablation experiments. In this process, we keep the restoration module unchanged and conduct ablation only on the design of the alignment module. We use the parallel alignment and fusion module based on the same EDA architecture, where all exposures are aligned with the reference frame simultaneously, as the

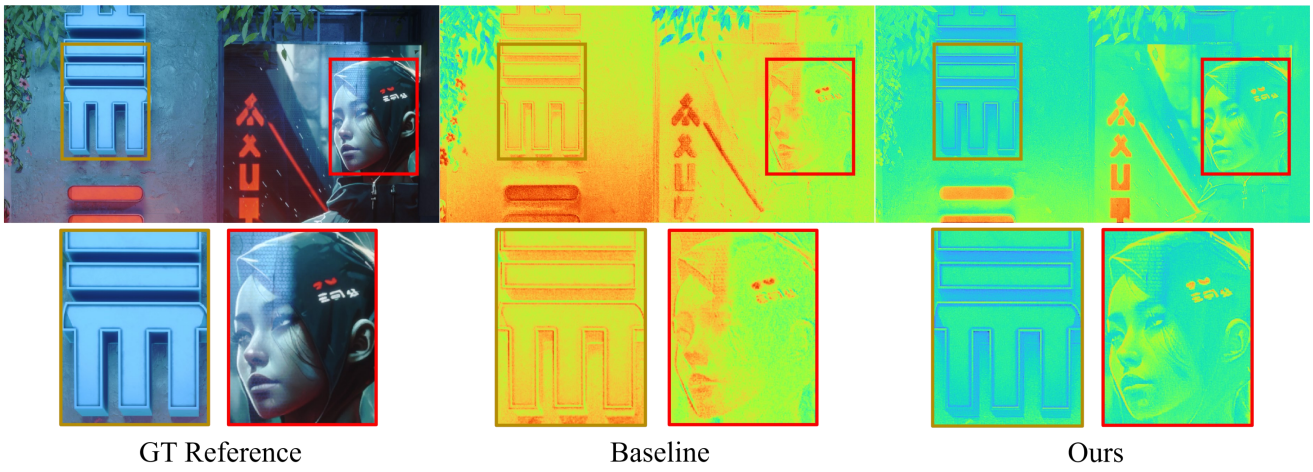


Figure 8. Visualization of aligned feature map. From left to right are the GT image, the baseline aligned feature visualization, the aligned feature visualization of our proposed method, and then we intercepted two contrast locations, text and face, as a way to improve the contrast. Zooming in allows better viewing.

baseline for developing the experimental analysis. All experiments in Tab. 3 use the one stage (Alignment + Restoration) training strategy instead of the proposed one for easy comparison.

Recursive or Parallel Alignment. As we analyzed in Sec. 3.1, using a recursive alignment strategy is able to decompose the solution space of the problem and provide a better learning path for the alignment model. Tab. 3 shows that using recursive alignment is able to obtain PSNR of 42.25 dB, which is 0.19 dB more than using parallel alignment, proving the effectiveness of recursive alignment. Additionally, we show the complexity comparison of the two schemes in Tab. 5, and it can be seen that the recursive alignment scheme brings only a small complexity gain, with the number of parameters increasing by only 1.406M and FLOPs increasing by only 0.248T.

The significance of multi-exposure alignment. We believe that exposure alignment plays a crucial role in two aspects: 1) it enables more accurate optical flow priors from the pre-trained optical flow network. As shown in Tab. 3, introducing additional flow in the F-EDA experiment actually leads to a performance drop. This is due to severe optical flow prediction errors caused by differences in exposure brightness domains, as illustrated in Fig.3. However, when applying EV-Align in advance, the performance improves directly by 0.76 dB. 2) it facilitates feature domain alignment within the implicit alignment module of EDA, significantly enhancing the accuracy of the multi-scale DCN in EDA. As shown in Tab.3, even without optical flow and only using the recursive EDA baseline, applying EV-Align in advance still yields a 0.40 dB improvement. This indirectly demonstrates that EV-align can effectively mitigate the degradation of implicit alignment accuracy caused by exposure domain differences.

Visualization of aligned feature map. Fig. 8 shows a visual comparison of the alignment features. From Fig. 8, we can see that more accurate alignment features can be obtained by using our proposed alignment scheme. Specifically, compared to the baseline features, firstly, our extracted aggregated features have better boundaries, such as the eye part of the avatar shown in the red box, where baseline obviously does not merge the multi-frame features well, and thus has fuzzy ghosting. Second, as can be seen from the contrast of the feature maps, our extracted features are more sensitive to the perception of different physical brightnesses in the scene, and thus are more suitable for HDR tasks.

5. Conclusion

In this work, we present Recursive Multi-Exposure Alignment with Spatiotemporal Decoupling (RASD). To address the challenging task of Efficient Burst HDR and Restoration from multiple RAW inputs to RGB output, we propose a recursive flow guided alignment module based on multi-exposure alignment, which is capable of providing accurate multi-frame alignment to provide a more efficient implicit prior to the restoration module. In addition, we innovatively propose a spatiotemporal decoupling training strategy to prevent the conflict of optimization objectives, which greatly improves the performance of the algorithm, and since the strategy is not limited by specific algorithms, we hope to provide performance enhancement for more burst image restoration tasks. In conclusion, our algorithm achieves state-of-the-art performance and wins the NTIRE 2025 Efficient Burst HDR and Restoration Challenge, being able to restores the finest details in this task.

References

[1] Antoni Buades, Onofre Martorell, and M Sánchez-Beeckman. Joint denoising and hdr for raw image sequences. *IEEE Transactions on Computational Imaging*, 10:277–290, 2024. 1

[2] Yuanhao Cai, Xiaowan Hu, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, and Donglai Wei. Learning to generate realistic noisy images via pixel-level noise-aware adversarial training. *Advances in Neural Information Processing Systems*, 34:3259–3270, 2021. 1

[3] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20:89–97, 2004. 2

[4] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European conference on computer vision*, pages 17–33. Springer, 2022. 2, 4

[5] Zheng Chen, Yulun Zhang, Ding Liu, Jinjin Gu, Linghe Kong, Xin Yuan, et al. Hierarchical integration diffusion model for realistic image deblurring. *Advances in neural information processing systems*, 36, 2024. 1

[6] Xiaojie Chu, Liangyu Chen, Chengpeng Chen, and Xin Lu. Improving image restoration by revisiting global information aggregation. In *ECCV*, pages 53–71. Springer, 2022. 5

[7] Xinwei Dai, Yuanbo Zhou, Xintao Qiu, Hui Tang, Wei Deng, Qinquan Gao, and Tong Tong. Learnable global spatio-temporal adaptive aggregation for bracketing image restoration and enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6235–6245, 2024. 1

[8] Akshay Dudhane, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Burst image restoration and enhancement. In *Proceedings of the ieee/cvf Conference on Computer Vision and Pattern Recognition*, pages 5759–5768, 2022. 2

[9] Akshay Dudhane, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Burstormer: Burst image restoration and enhancement transformer. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5703–5712. IEEE, 2023. 2, 4, 6

[10] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing*, 15(12):3736–3745, 2006. 2

[11] Michael Elad and Yacov Hel-Or. A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur. *IEEE Transactions on image Processing*, 10(8):1187–1193, 2001. 2

[12] Chi-Mao Fan, Tsung-Jung Liu, Kuan-Hsien Liu, and Ching-Hsiang Chiu. Selective residual m-net for real image denoising. In *2022 30th European Signal Processing Conference (EUSIPCO)*, pages 469–473. IEEE, 2022. 1

[13] Zhenxuan Fang, Fangfang Wu, Weisheng Dong, Xin Li, Jinjian Wu, and Guangming Shi. Self-supervised non-uniform kernel estimation with flow-based motion prior for blind image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18105–18114, 2023. 1

[14] Samuel W Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics (ToG)*, 35(6):1–12, 2016. 2

[15] Tao Huang, Songjiang Li, Xu Jia, Huchuan Lu, and Jianzhuang Liu. Neighbor2neighbor: Self-supervised denoising from single noisy images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14781–14790, 2021. 1

[16] Michal Irani and Shmuel Peleg. Improving resolution by image registration. *CVGIP: Graphical models and image processing*, 53(3):231–239, 1991. 2

[17] Nima Khademi Kalantari, Ravi Ramamoorthi, et al. Deep high dynamic range imaging of dynamic scenes. *ACM Trans. Graph.*, 36(4):144–1, 2017. 1, 2

[18] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Advances in Neural Information Processing Systems*, 35:23593–23606, 2022. 2

[19] Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5886–5895, 2023. 1

[20] Dabov Kostadin, Foi Alessandro, and Egiazarian Karen. Video denoising by sparse 3d transform-domain collaborative filtering. In *European signal processing conference*, 2007. 2

[21] Sangmin Lee, Eunpil Park, Angel Canelo, Hyunhee Park, Youngjo Kim, Hyung-Ju Chun, Xin Jin, Chongyi Li, Chun-Le Guo, and Radu Timofte. Ntire 2025 challenge on efficient burst hdr and restoration: Datasets, methods, and results. *CVPRW*, 2025. 2, 7

[22] Jiaming Liu, Chi-Hao Wu, Yuzhi Wang, Qin Xu, Yuqian Zhou, Haibin Huang, Chuan Wang, Shaofan Cai, Yifan Ding, Haoqiang Fan, et al. Learning raw image denoising with bayer pattern unification and bayer preserving augmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 2, 5

[23] Shuaizheng Liu, Xindong Zhang, Lingchen Sun, Zhetong Liang, Hui Zeng, and Lei Zhang. Joint hdr denoising and fusion: A real-world mobile hdr image dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13966–13975, 2023. 1

[24] Zhen Liu, Yinglong Wang, Bing Zeng, and Shuaicheng Liu. Ghost-free high dynamic range imaging with context-aware transformer. In *European Conference on computer vision*, pages 344–360. Springer, 2022. 1, 2, 3, 6

[25] Ziwei Luo, Lei Yu, Xuan Mo, Youwei Li, Lanpeng Jia, Haoqiang Fan, Jian Sun, and Shuaicheng Liu. Ebsr: Feature enhanced burst super-resolution with deformable alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 471–478, 2021. 2

[26] Ziwei Luo, Youwei Li, Shen Cheng, Lei Yu, Qi Wu, Zhihong Wen, Haoqiang Fan, Jian Sun, and Shuaicheng Liu. Bsrt: Improving burst super-resolution with swin transformer and flow-guided deformable alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 998–1008, 2022. 2, 6

[27] Matteo Maggioni, Giacomo Boracchi, Alessandro Foi, and Karen Egiazarian. Video denoising using separable 4d nonlocal spatiotemporal transforms. In *Image Processing: Algorithms and Systems IX*, pages 9–19. SPIE, 2011. 2

[28] Matteo Maggioni, Giacomo Boracchi, Alessandro Foi, and Karen Egiazarian. Video denoising, deblocking, and enhancement through separable 4-d nonlocal spatiotemporal transforms. *IEEE Transactions on image processing*, 21(9): 3952–3966, 2012. 2

[29] Nancy Mehta, Akshay Dudhane, Subrahmanyam Murala, Syed Waqas Zamir, Salman Khan, and Fahad Shahbaz Khan. Adaptive feature consolidation network for burst super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1279–1286, 2022. 2

[30] Ben Mildenhall, Jonathan T Barron, Jiawen Chen, Dillon Sharlet, Ren Ng, and Robert Carroll. Burst denoising with kernel prediction networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2502–2510, 2018. 2

[31] 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. 3

[32] A Paszke. Pytorch: An imperative style, high-performance deep learning library. *arXiv preprint arXiv:1912.01703*, 2019. 5

[33] Anurag Ranjan and Michael J Black. Optical flow estimation using a spatial pyramid network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4161–4170, 2017. 4

[34] Yong Shu, Liquan Shen, Xiangyu Hu, and Zihao Zhou. High dynamic range imaging from raw domain with a new benchmark dataset and a multi-scale network. In *2024 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pages 1–6. IEEE, 2024. 1

[35] Steven Tel, Zongwei Wu, Yulun Zhang, Barthélémy Heyrman, Cédric Demonceaux, Radu Timofte, and Dominique Ginhac. Alignment-free hdr deghosting with semantics consistent transformer. *arXiv preprint arXiv:2305.18135*, 2023. 1, 2

[36] Roger Y Tsai and Thomas S Huang. Multiframe image restoration and registration. *Multiframe image restoration and registration*, 1:317–339, 1984. 2

[37] Qi Wu, Mingyan Han, Ting Jiang, Haoqiang Fan, Bing Zeng, and Shuaicheng Liu. Realistic noise synthesis with diffusion models. *arXiv preprint arXiv:2305.14022*, 2023. 2

[38] Zhihao Xia, Federico Perazzi, Michaël Gharbi, Kalyan Sunkavalli, and Ayan Chakrabarti. Basis prediction networks for effective burst denoising with large kernels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11844–11853, 2020. 2

[39] Xiaoxia Xing, Hyunhee Park, Fan Wang, Ying Zhang, Sejun Song, Changho Kim, and Xiangyu Kong. High quality reference feature for two stage bracketing image restoration and enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6267–6276, 2024. 1

[40] Qingsen Yan, Dong Gong, Qinfeng Shi, Anton van den Hengel, Chunhua Shen, Ian Reid, and Yanning Zhang. Attention-guided network for ghost-free high dynamic range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1751–1760, 2019. 1, 2, 3

[41] Qingsen Yan, Weiye Chen, Song Zhang, Yu Zhu, Jinqiu Sun, and Yanning Zhang. A unified hdr imaging method with pixel and patch level. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22211–22220, 2023. 3

[42] Kangzhen Yang, Tao Hu, Kexin Dai, Genggeng Chen, Yu Cao, Wei Dong, Peng Wu, Yanning Zhang, and Qingsen Yan. Crnet: A detail-preserving network for unified image restoration and enhancement task. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6086–6096, 2024. 1

[43] Qirui Yang, Yihao Liu, Qihua Chen, Huanjing Yue, Kun Li, and Jingyu Yang. Efficient hdr reconstruction from real-world raw images. *arXiv preprint arXiv:2306.10311*, 2023. 1

[44] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022. 2

[45] Zhilu Zhang, Shuohao Zhang, Renlong Wu, Zifei Yan, and Wangmeng Zuo. Exposure bracketing is all you need for a high-quality image. In *The Thirteenth International Conference on Learning Representations*, 2025. 1, 2, 6

[46] Yunhao Zou, Chenggang Yan, and Ying Fu. Rawhdr: High dynamic range image reconstruction from a single raw image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12334–12344, 2023. 1