
RAWENHANCER: ATTENTIVE MULTI-EXPOSURE SELECTION FOR BRACKETING IMAGE RECONSTRUCTION - SUPPLEMENTAL MATERIAL -

In this document, we first describe the used datasets in Section 1 and show the efficiency comparison across different iterations in Section 2. To provide an intuitive understanding of our proposed method, we present visualizations of the intermediate features of the EASA and IFS modules in Sections 3 and 4, respectively. Finally, we provide more visual comparison results in Section 5 and demonstrate the real-world comparisons in Section 6.

1 DATASET DESCRIPTION

We conduct experiments on Raw datasets from diverse scenarios, including BracketIRE-syn Zhang et al. (2025), MobileHDR Liu et al. (2023), and RawFusion Lee et al. (2025).

The BracketIRE-syn dataset Zhang et al. (2025) is synthetically generated from high-quality HDR video, simulating typical degradations such as noise, blur, uneven exposure, and dynamic range compression. It comprises 1,335 image pairs, each scene containing five distinct exposure levels. A total of 1,045 pairs are assigned for training, while the remaining 290 are reserved for testing.

The MobileHDR dataset Liu et al. (2023) is captured using three types of mobile phone in both daytime and nighttime scenes, covering a wide range of noise levels from ISO 100 to 6400. This dataset contains 251 samples, each containing three exposures. Among these, 223 sets are used for training, while the remaining 28 are used for testing.

The RawFusion dataset Lee et al. (2025) is generated via a virtual pipeline, introducing real-world degradations like noise, rotation, and motion blur. Each scene includes three exposure levels, resulting in a total of 340 image pairs with a resolution of 768×1536 . Among these, 300 pairs are used for training and 40 pairs are reserved for testing.

2 EFFICIENCY COMPARISON ACROSS DIFFERENT ITERATIONS

Table 1: efficiency comparison across different iterations.

Iterations	PSNR	SSIM	FLOPs (G)	Runtime (ms)
1	37.51	0.9392	191.59	57.76
2	37.58	0.9391	215.18	65.12
3	37.63	0.9403	238.76	72.79

Table 1 shows that as the number of iterations increases, PSNR and SSIM performance improves, but runtime also grows noticeably. To achieve a better balance between performance and efficiency, we ultimately choose 3 iterations as the default configuration.

3 VISUALIZATION OF HIDDEN FEATURES FROM THE EASA MODULE.

In this section, we visualize the intermediate features of the first EASA module to provide an intuitive illustration of how it leverages prior information to achieve effective representations. Figure 1 demonstrates that using gradient magnitude effectively preserves key information within the input features. By leveraging the selective attention mechanism, feature representations are further refined, leading to output features with enhanced structure and richer detail.

4 VISUALIZATION OF HIDDEN FEATURES FROM THE IFS BLOCK.

In this section, we visualize the intermediate features of the IFS block to intuitively demonstrate its process of dynamically merging valuable features from multi-exposure images. Figure 2 shows that

the two GRUs used with different kernel sizes, effectively fuse useful structural and detailed information from the multi-exposure features. By integrating these two complementary representations, the model achieves a more dedicated feature representation. Meanwhile, as the number of iterations increases, the model progressively suppresses irrelevant information and continuously enhances the learned features.

5 MORE VISUAL COMPARISONS

In this section, we provide additional visual comparisons between the proposed method and state-of-the-art approaches Dudhane et al. (2022; 2023); Kong et al. (2024); Liu et al. (2023); Zhang et al. (2025); Liu et al. (2022) on test datasets. Figures 3 and 4 present the comparison results, demonstrating that our method generates high-quality HDR images with more details.

6 VISUAL COMPARISON ON REAL-WORLD SCENES

In this section, we compare our method with TMRNet Zhang et al. (2025) and Joint-HDRDN Liu et al. (2023) on real-world scenes. As detailed in Section 3.1, these approaches rely on optical flow estimation and self-attention for inter-frame feature aggregation, which often fails under complex exposure conditions. Consequently, their reconstructions suffer from noticeable artifacts. In contrast, our proposed RawEnhancer consistently produces better results in these challenging scenarios, as shown in Figure 5.

REFERENCES

- Akshay Dudhane, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Burst image restoration and enhancement. In *CVPR*, 2022.
- Akshay Dudhane, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Burstormer: Burst image restoration and enhancement transformer. In *CVPR*, 2023.
- Lingtong Kong, Bo Li, Yike Xiong, Hao Zhang, Hong Gu, and Jinwei Chen. Safnet: Selective alignment fusion network for efficient hdr imaging. In *ECCV*, 2024.
- Sangmin Lee, Eunpil Park, Angel Canelo, Hyunhee Park, Youngjo Kim, Hyungju Chun, Xin Jin, Chongyi Li, Chun-Le Guo, Radu Timofte, and et al. Ntire 2025 challenge on efficient burst hdr and restoration: Datasets, methods, and results. In *CVPR Workshops*, 2025.
- 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 *CVPR*, 2023.
- Zhen Liu, Yinglong Wang, Bing Zeng, and Shuaicheng Liu. Ghost-free high dynamic range imaging with context-aware transformer. In *ECCV*, 2022.
- Zhilu Zhang, Shuohao Zhang, Renlong Wu, Zifei Yan, and Wangmeng Zuo. Exposure bracketing is all you need for a high-quality image. In *ICLR*, 2025.

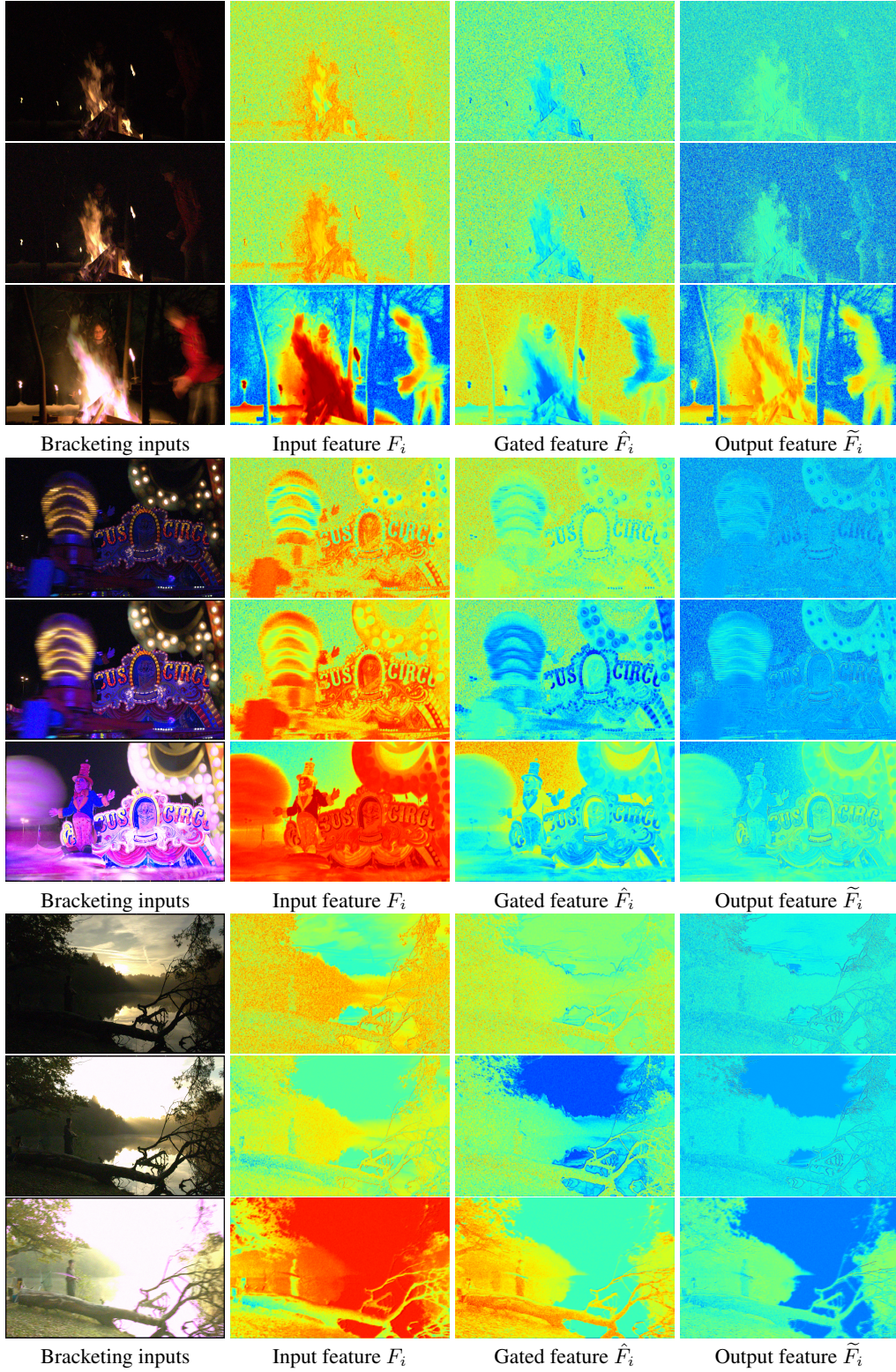


Figure 1: Illustration of the learned feature from the first EASA module. The EASA module can effectively utilize valuable information within images across various scenarios and exposures.

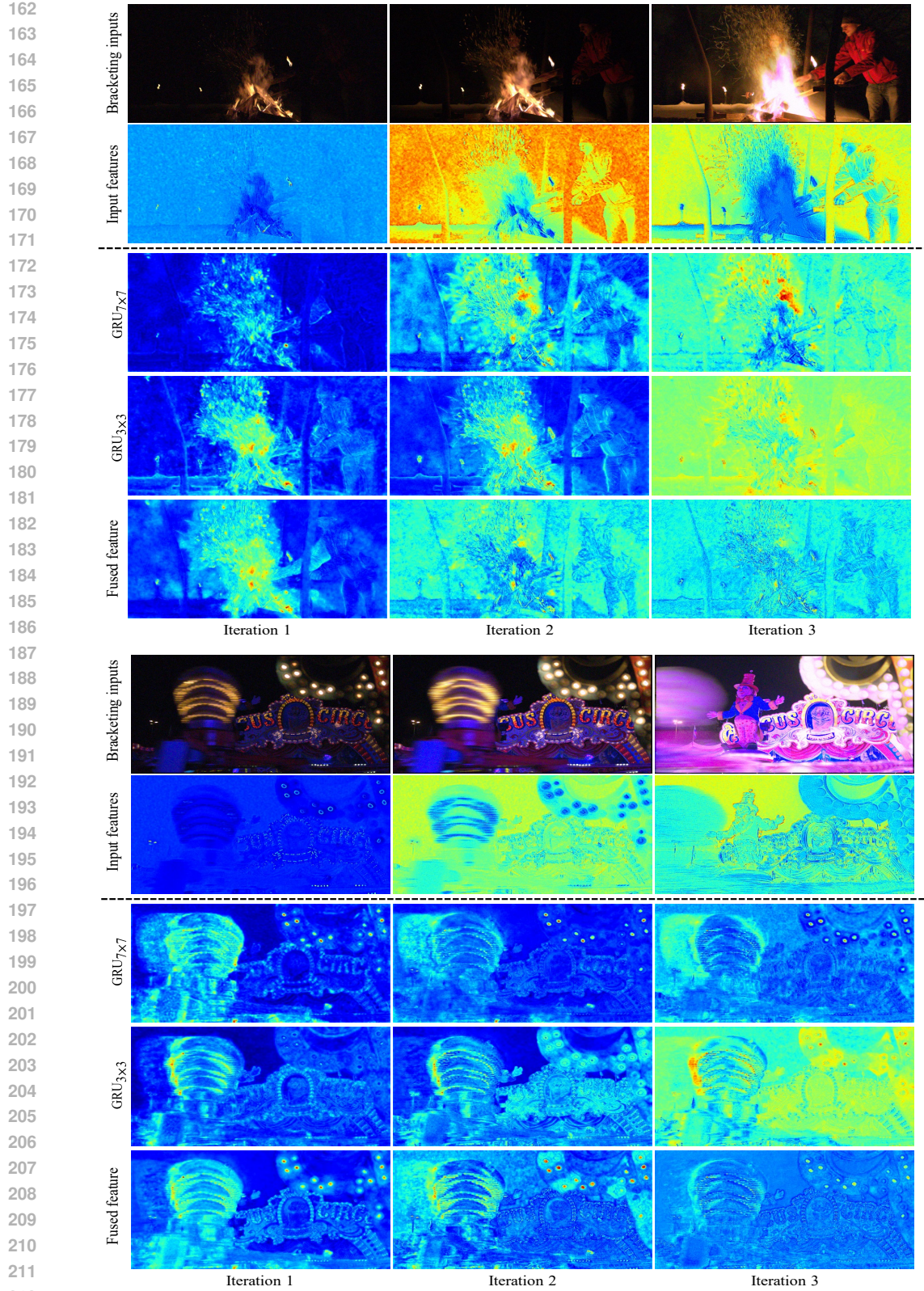


Figure 2: Illustration of the learned features from the IFS block at the $\times 2$ downsampled feature space. As the number of iterations increases, the model progressively suppresses irrelevant information and achieves a more discriminative representation.

216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269

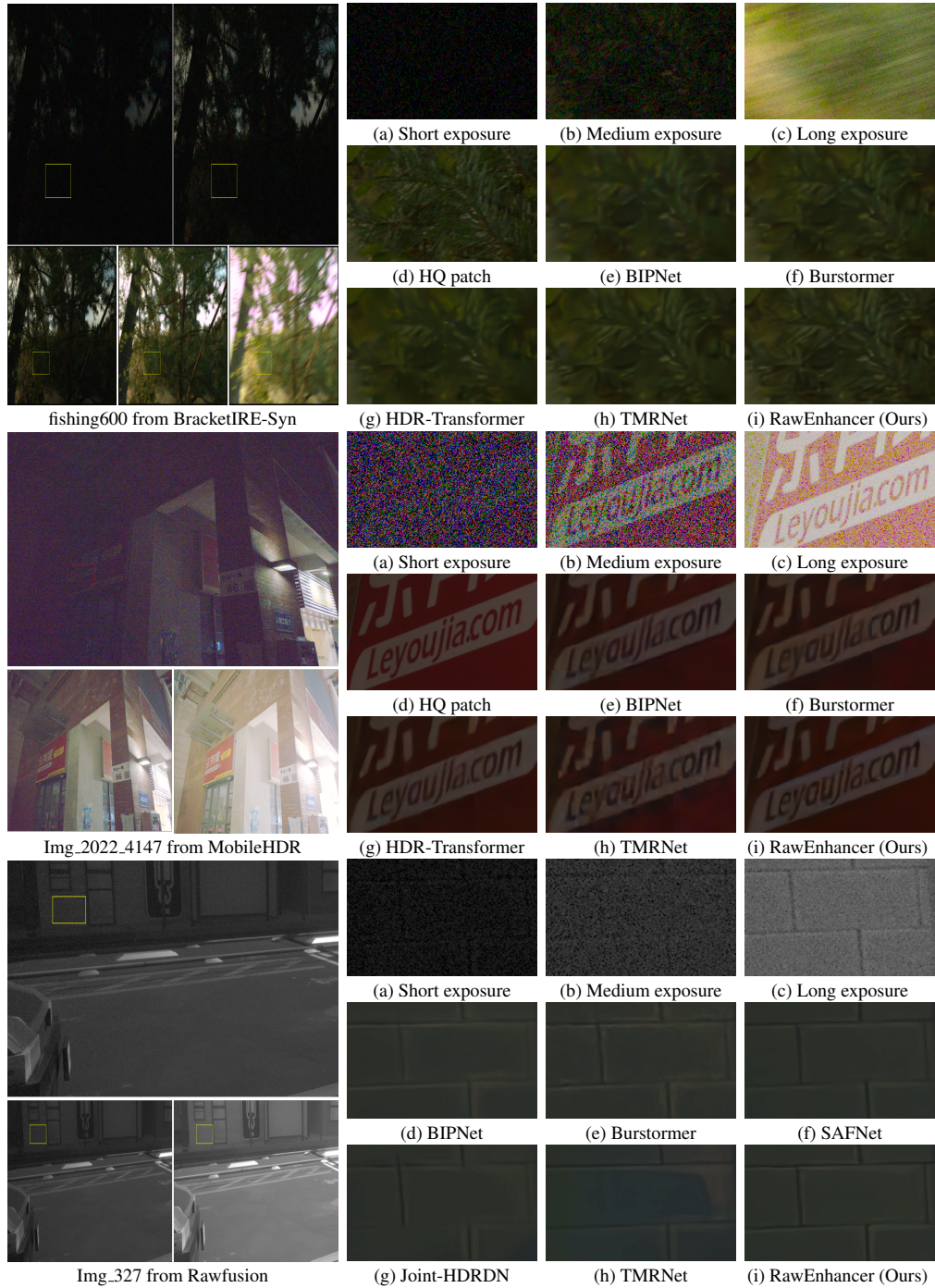


Figure 3: Visual comparisons on different scenes. The reconstructed results of the listed methods exhibit noticeable artifacts, whereas our RawEnhancer produces sharper and cleaner images.

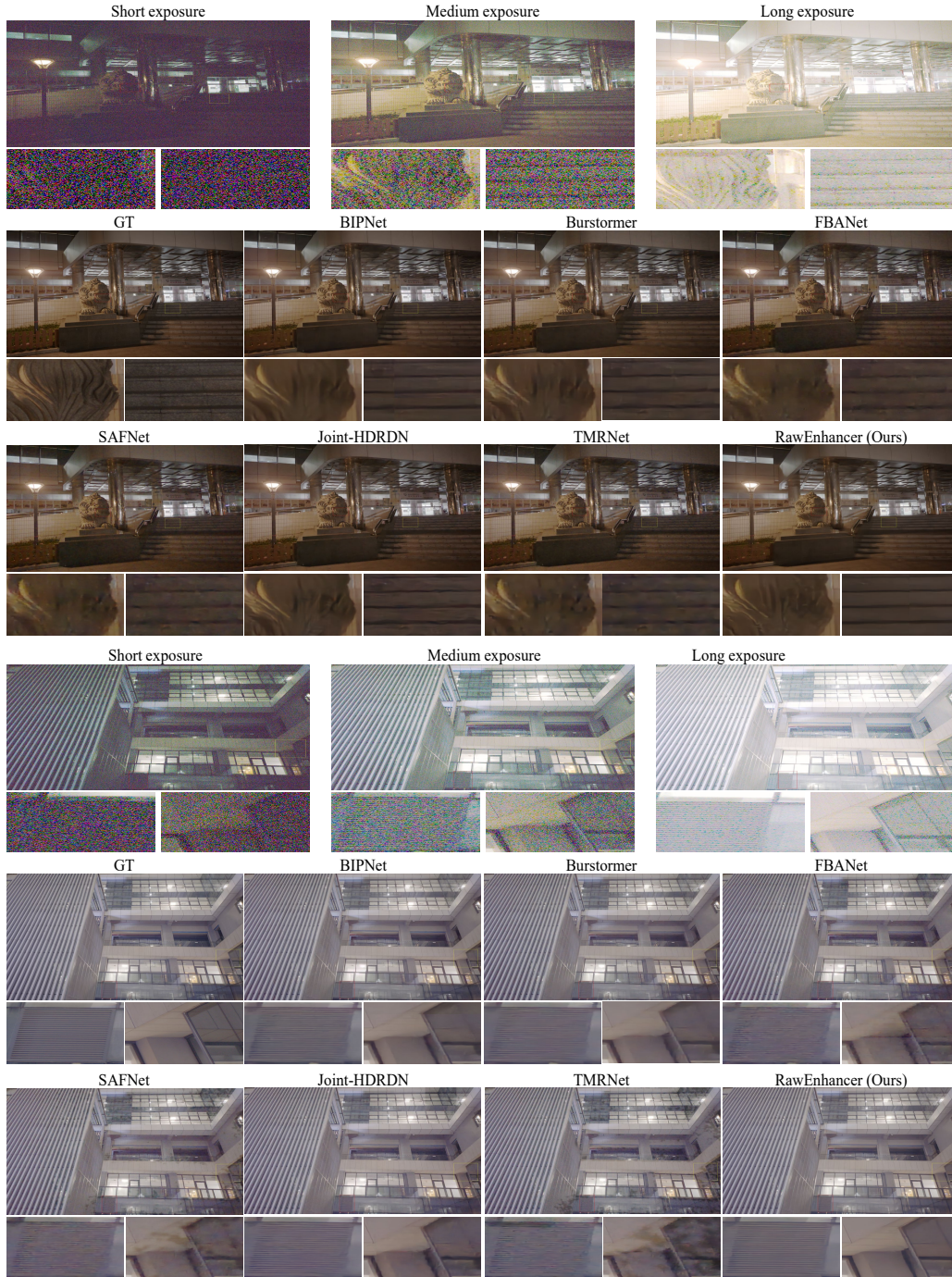


Figure 4: Visual comparison on MobileHDR Liu et al. (2023) dataset. The reconstructed results of the listed methods exhibit noticeable artifacts, whereas our RawEnhancer produces sharper and cleaner images.



Figure 5: Visual comparison on real multi-exposure scenarios Liu et al. (2023). The reconstructed results of the listed methods exhibit noticeable artifacts, whereas our RawEnhancer consistently reconstructs images with better detail and structure across various complex exposures.