Supplementary Material
Burstormer: Burst Image Restoration and Enhancement Transformer

Akshay Dudhane[1] Syed Waqas Zamir[2] Salman Khan[1,3]
Fahad Shahbaz Khan[1] Ming-Hsuan Yang[4,5,6]
[1]Mohamed bin Zayed University of AI [2]Inception Institute of AI [3]Australian National University
[4]University of California, Merced [5]Yonsei University [6]Google Research

Here we provide more details on architectural design, additional ablations, and visual comparisons for burst SR, low-light image enhancement and denoising.

1. Network Architectural Details

In Burstormer, the EDA module is a 3-level encoder-decoder, where each level employs 1 FA (containing single deformable conv. layer) and 1 RBFE module. In the image reconstruction stage, we use 2 NRFE modules. The BFF unit both in RBFE and NRFE consists of 1 BFA module.

Figure 1 shows the BFA module that consists of multi-dconv head transposed attention (MDTA) and gated-Dconv feed-forward network (GDFN) [9]. MDTA encodes local and non-local context, and efficient enough to be applied to high-resolution images. Whereas, GDFN performs controlled feature transformation i.e., suppressing less informative features, and allowing only the useful information to pass further through the network.

2. Ablations on alignment and fusion modules

Table 1 compares the the properties of the proposed EDA and other existing alignment modules. Unlike existing explicit feature alignment approaches DBSR [1] and MFIR [2], the proposed EDA operates at multiple spatial scales and aligns burst features implicitly without any additional supervision. Overall, the proposed EDA module possesses required properties which makes it effective for the burst feature alignment.

Table 2 compares several feature fusion techniques. Our NRFE is flexible to taking as input the features of more than two frames. It extracts local and non-local burst features, enables long-range inter-frame interactions and aggregates the burst neighborhoods to obtain high-quality image.

3. Additional visual results

Burst Super-resolution. Figure 2, and Figure 3 show qualitative results of competing approaches on examples from the SyntheticBurst and (real) BurstSR datasets [1] for $4\times$ SR. The reproductions of our Burstormer are more detailed, sharper than those produced by the other methods.

Burst low-light image enhancement. Figure 4 depicts that Burstormer produces images that are visually more closer to the ground-truth than the other approaches.

Burst Denoising. Figure 5 shows that the proposed Burstormer is capable of removing noise, while preserving the desired texture and structural content.

References

[1] Goutam Bhat, Martin Danelljan, Luc Van Gool, and Radu Timofte. Deep burst super-resolution. In *CVPR*, 2021. 1, 2, 3, 4

[2] Goutam Bhat, Martin Danelljan, Fisher Yu, Luc Van Gool, and Radu Timofte. Deep reparametrization of multi-frame super-resolution and denoising. In *ICCV*, 2021. 1, 2

[3] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *CVPR*, 2018. 5

[4] Akshay Dudhane, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Burst image restoration and enhancement. In *CVPR*, 2022. 2

[5] Ben Mildenhall, Jonathan Barron, Jiawen Chen, Dillon Sharlet, Ren Ng, and Robert Carroll. Burst denoising with kernel prediction networks. In *CVPR*, 2018. 6

[6] Yapeng Tian, Yulun Zhang, Yun Fu, and Chenliang Xu. Tdan: Temporally-deformable alignment network for video super-resolution. In *CVPR*, 2020. 2

[7] Xintao Wang, Kelvin CK Chan, Ke Yu, Chao Dong, and Chen Change Loy. Edvr: Video restoration with enhanced deformable convolutional networks. In *CVPRW*, 2019. 2

[8] Zhihao Xia, Federico Perazzi, Michaël Gharbi, Kalyan Sunkavalli, and Ayan Chakrabarti. Basis prediction networks for effective burst denoising with large kernels. In *CVPR*, 2020. 6

[9] 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 *CVPR*, 2022. 1

	DBSR/MFIR [1,2]	TDAN [6]	PCD [7]	EBFA [4]	EDA (Ours)
Extra supervision	✓	×	×	×	×
Implicit alignment	×	✓	✓	✓	✓
Multi-scale hierarchy	✓	×	✓	×	✓
Attention for feature denoising	×	×	×	✓	✓
Reference-frame based refinement	×	×	×	×	✓

Table 1. Ablation on existing **Feature alignment strategies** with our EDA module.

	DBSR/MFIR [1,2]	PBFF [4]	NRFE (Ours)
Flexible w.r.t multiple inputs	✓	×	✓
Long-range inter-frame interaction	×	✓	✓
Local and non-local feature extraction	×	✓	✓
Computational overhead	↓	↑	↓

Table 2. Ablation on existing **Feature fusion techniques** with our NRFE module.

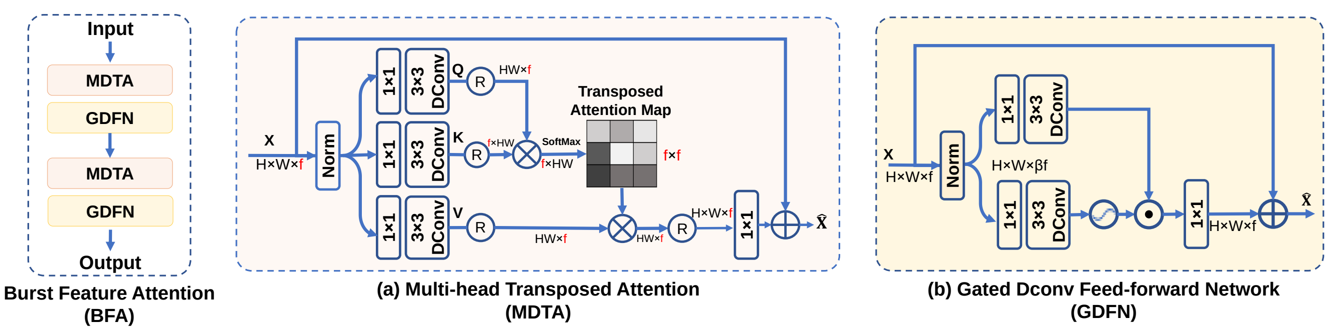

Figure 1. **Burst Feature Attention** (BFA) used in the proposed alignment and reconstruction stages to extract features encoding both local and non-local pixel interactions.

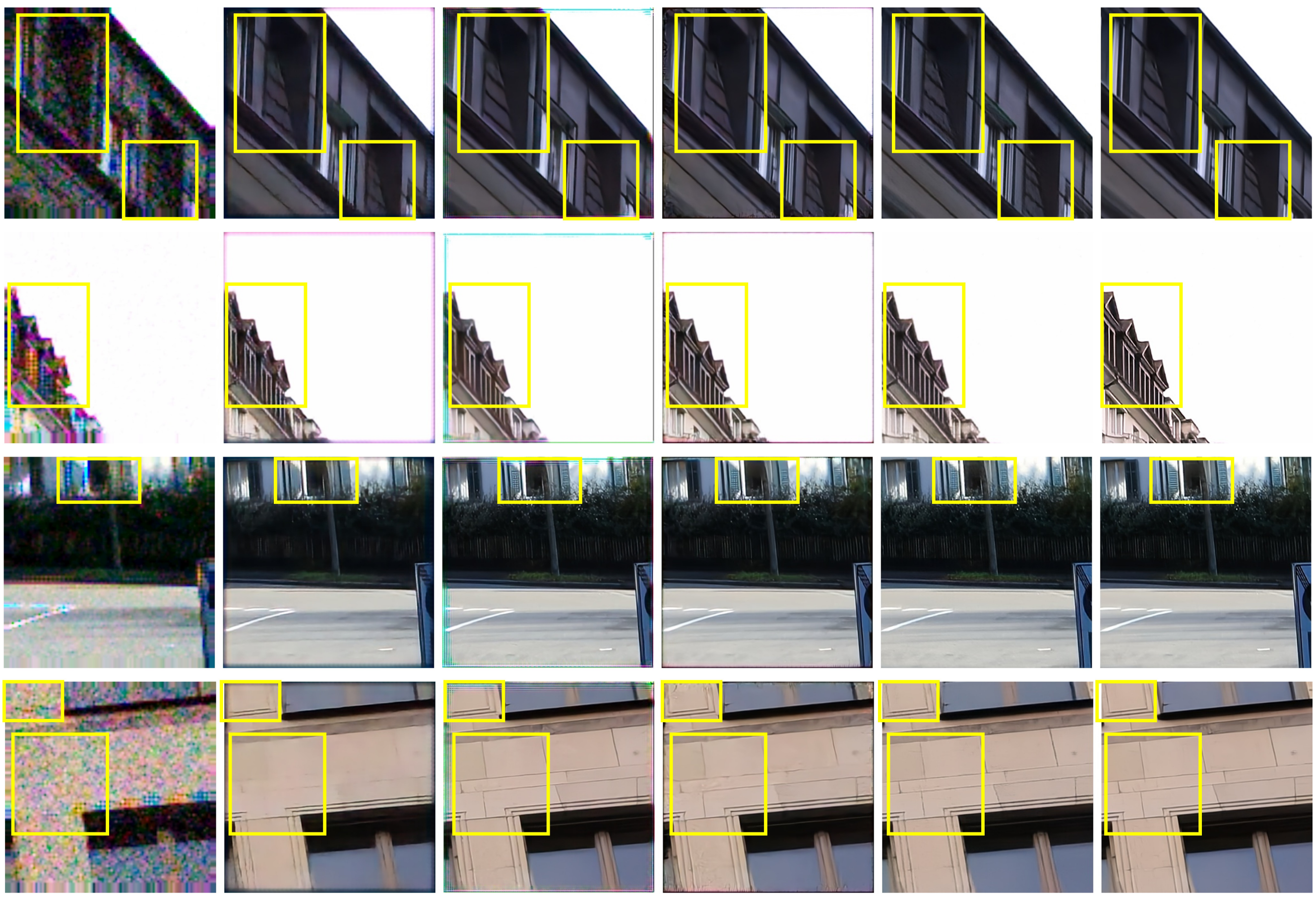

Figure 2. **Burst super-resolution** ($4\times$) results on SyntheticBurst dataset [1].

Figure 3. **Burst super-resolution** (4×) results on BurstSR (real) dataset [1].

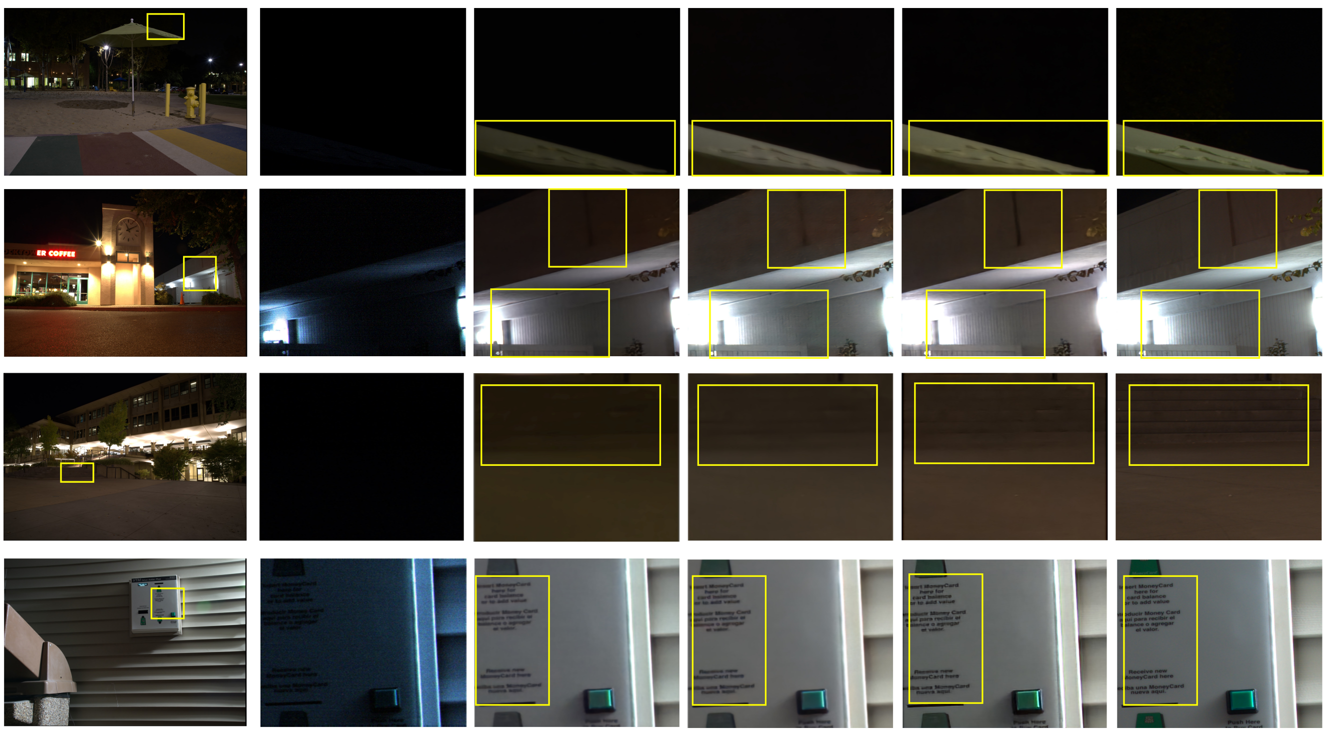

Figure 4. **Burst low-light image enhancement** comparisons on the Sony subset of SID dataset [3].

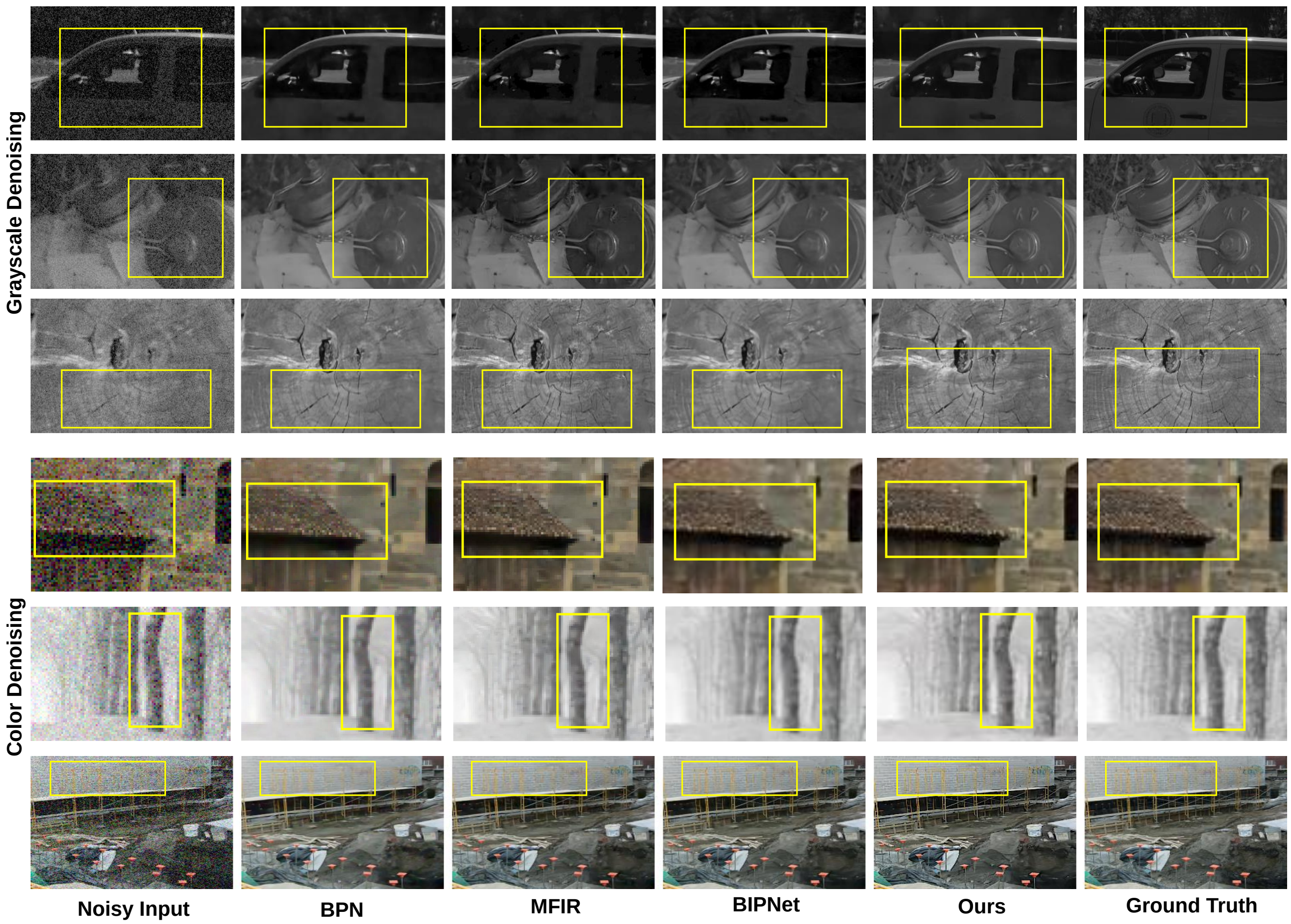

Figure 5. **Burst denoising** results on burst images from the grayscale [5] and color datasets [8].