

# Supplementary Materials: Reference-based Burst Super-resolution

Anonymous Authors

This supplementary material discusses further details of the reference-based burst super-resolution (RefBSR). In Section A, we explain more details of the proposed architecture. In Section B, we provide more analysis of our proposed method and dataset. Section C shows more results of our model with the state-of-the-art methods.

## A DETAIL OF REFBSR ARCHITECTURE

In this section, we describe our proposed network architecture in detail. Firstly, our model contains an encoder that takes the burst frames  $I_b$  as input. The encoder extracts the feature of the burst frames and aligns each burst frame  $I_{b_i}$  to the base frame  $I_{b_1}$  grid. Specifically, we extract each burst feature using a Swin transformer group comprising six Swin transformer blocks [5]. Since burst features are not aligned, we estimate the flow and then align each burst feature as following in [6]. Accordingly, we obtain the aligned burst features  $\{\tilde{F}_{b_i \rightarrow b_1}\}_{i=1}^N$ . On the other hand, our network contains a decoder to aggregate all fused Ref-burst features  $\tilde{F}_f^i$  and generate the final output. All Ref-burst features  $\{\tilde{F}_f^i\}_{i=1}^N$  are concatenated and then fused via point-wise convolution. After that, the pixelshuffle upsampling operation following five Swin transformer groups reconstructs the final output.

## B ADDITIONAL ANALYSIS

**More Analysis of BRTT.** We discuss more detailed results to generate the burst feature-aware aligned Ref feature in the BRTT module. we build the three variations to further verify the burst feature-aware aligned Ref feature. In Table 4, ‘ $+F_{r \rightarrow b_1}$ ’ model indicates the model only without the aligned burst feature  $F_{b_i \rightarrow b_1}$  in the main manuscript (4). In other words, there is no awareness of the burst features to align the Ref feature. Another variant, ‘ $+F_{b_i \rightarrow b_1}$ ’ is the model without the flow-based aligned Ref feature  $F_{r \rightarrow b_1}$  in (4). Finally, ‘ $+F_{r \rightarrow b_1} + F_{b_i \rightarrow b_1}$ ’ is the final model. Our model demonstrates enhanced performance when employing the aligned burst and the flow-based aligned Ref feature.

**More Analysis of RBAF.** We provide additional analysis on the adaptive fusion weight map  $A$  in the RBAF module. We build the variant that explicitly utilizes the confidence map instead of applying the weight map  $A$  in the RBAF module. The Ref confidence map can be derived by applying a max operation to the correlations in the base and reference matching step. In contrast, the Ref confidence map is inverted for the burst confidence map. Table 5 shows the ‘Confidence’ model, which replaces the weight map  $A$  in (6) with two confidence maps. Another variant, the ‘Confidence+Conv’ model, applies a convolutional layer to each confidence map. The last method is our proposed method with the adaptive fusion weight map  $A$ . As indicated in Table 5, the RBAF module with the adaptive fusion weight map  $A$  demonstrates better performance.

**Effectiveness of RefBSR Dataset.** To further substantiate the effectiveness of our dataset, we train our model with each dataset

Table 1: Effectiveness of multiple reference images.

Dataset	RefBSR Dataset		
Method	PNSR	SSIM	LPIPS
1 Ref	42.63	0.960	0.034
2 Refs	43.09	0.964	0.033
3 Refs	<b>43.22</b>	<b>0.965</b>	<b>0.033</b>

Table 2: Effectiveness of burst frames.

Dataset	RefBSR Dataset		
Method	PNSR	SSIM	LPIPS
RefSR	43.64	0.965	0.066
Burst 3 + Ref	44.92	0.972	0.043
Burst 7 + Ref	46.08	0.976	0.032
Burst 14 + Ref	<b>46.49</b>	<b>0.980</b>	<b>0.030</b>

Table 3: Computational cost.

Method	Params(M)	GMac	RT(ms)
C2-Matching [4]	7.6	220	56
LMR [7]	23.7	647	73
DBSR [1]	13.0	236	430
MFIR [2]	12.1	220	420
Burstormer [3]	3.5	185	70
Ours	10.6	881	188

and test it on the RefBSR dataset. In Table 6, ‘Only synthetic’ indicates our model is trained using only synthetic dataset and tested on the RefBSR dataset. In contrast, ‘Only RefBSR’ is our model trained with only the RefBSR dataset and evaluated on the RefBSR dataset. The last row is our proposed training strategy that trains our model with the synthetic dataset and then finetunes with the RefBSR dataset. The results indicate that training solely on the RefBSR dataset yields considerable improvement. Notably, our proposed training strategy shows the most superior performance. **Effectiveness of Burst Frames.** We further investigate the performance of our model based on the number of burst frames. As indicated in Table 2, ‘RefSR’ denotes the utilization of the base frame with the Ref image. ‘Burst3+Ref’ means the usage of three burst frames and the Ref image as inputs. Similarly, ‘Burst7+Ref’ indicates the input comprises seven burst frames and the Ref image. Lastly, ‘Burst14+Ref’ denotes the model utilizing all burst frames and the Ref image. Table 2 shows that as the number of frames increases, generally the performance also increases.

**Effectiveness of Multiple References.** We investigate the effectiveness of multiple Ref images. We manually construct a subset containing multiple Ref images from the RefBSR dataset. To build the subset, multiple Ref images are selected from a similar scene in

**Table 4: More analysis of BRTT.**

Dataset	RefBSR Dataset		
Method	PNSR	SSIM	LPIPS
$+F_{r \rightarrow b_1}$	46.37	0.977	0.031
$+F_{b_i \rightarrow b_1}$	46.40	0.978	0.032
$+F_{r \rightarrow b_1} + F_{b_i \rightarrow b_1}$	<b>46.49</b>	<b>0.980</b>	<b>0.030</b>

**Table 5: More analysis of RBAF.**

Dataset	RefBSR Dataset		
Method	PNSR	SSIM	LPIPS
Confidence	46.36	0.978	0.035
Confidence + Conv	46.37	0.979	0.034
Weight map $A$	<b>46.49</b>	<b>0.980</b>	<b>0.030</b>

**Table 6: Effectiveness of RefBSR dataset.**

Dataset	RefBSR Dataset		
Method	PNSR	SSIM	LPIPS
Only synthetic	38.98	0.922	0.081
Only RefBSR	44.40	0.969	0.074
Ours	<b>46.49</b>	<b>0.980</b>	<b>0.030</b>

the RefBSR dataset. Thus, a single set consists of 14 burst frames, 3 Ref images and a ground-truth image. The subset provides a total of 10 sets. The multiple Ref images in the single set are concatenated as the input. We evaluate the performance improvements based on the number of Ref images. As shown in Table 1, ‘1 Ref’ is only the usage of the single Ref image and 14 burst frames. Similarly, ‘2 Refs’ denotes that 2 Ref images and 14 burst frames are the inputs. Finally, ‘3 Refs’ indicates that 3 Ref images and 14 burst frames are the inputs. It is observed that as the number of Ref images increases, performance also improves.

**Computational Cost.** Furthermore, we compare computational cost with  $C^2$ -Matching [4], LMR [7], DBSR [1], MFIR [2], Burstormer [3]. We measured the complexity of the model in terms of parameters, MACs, and runtime. As shown in Table 3, our model requires fewer parameters compared to the state-of-the-art RefSR models such as LMR and MFIR.

## C MORE QUALITATIVE COMPARISON

In this part, we visualize more results with state-of-the-art models. We compare our model with SwinIR [5], MRefSR [7], MFIR [2] and Burstormer [3]. As shown in Figure 1 Figure 2 Figure 3, it

can be observed that the base frame and the Ref image differ in brightness and viewpoint. In a comparison between the outputs of our model and other models, our results demonstrate a relatively higher detail of the textures. These results prove that our model leverages information from the burst frames and the Ref image.

## REFERENCES

- [1] Goutam Bhat, Martin Danelljan, Luc Van Gool, and Radu Timofte. 2021. Deep burst super-resolution. In *CVPR*. 9209–9218.
- [2] Goutam Bhat, Martin Danelljan, Fisher Yu, Luc Van Gool, and Radu Timofte. 2021. Deep reparametrization of multi-frame super-resolution and denoising. In *ICCV*. 2460–2470.
- [3] Akshay Dudhane, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Hsuan Yang. 2023. Burstormer: Burst Image Restoration and Enhancement Transformer. In *CVPR*. 5703–5712.
- [4] Yuming Jiang, Kelvin CK Chan, Xintao Wang, Chen Change Loy, and Ziwei Liu. 2021. Robust Reference-based Super-Resolution via  $C^2$ -Matching. In *CVPR*. 2103–2112.
- [5] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. 2021. Swinir: Image restoration using swin transformer. In *ICCV*. 1833–1844.
- [6] Ziwei Luo, Youwei Li, Shen Cheng, Lei Yu, Qi Wu, Zhihong Wen, Haoqiang Fan, Jian Sun, and Shuaicheng Liu. 2022. BSRT: Improving burst super-resolution with swin transformer and flow-guided deformable alignment. In *CVPR*. 998–1008.
- [7] Lin Zhang, Xin Li, Dongliang He, Fu Li, Errui Ding, and Zhaoxiang Zhang. 2023. LMR: A Large-Scale Multi-Reference Dataset for Reference-based Super-Resolution. In *ICCV*. 13118–13127.



Figure 1: More qualitative comparison with SwinIR [5], LMR [7], MFIR [2] and Burstormer [3] on RefBSR dataset.





Figure 2: More qualitative comparison with SwinIR [5], MRefSR [7], MFIR [2] and Burstormer [3] on RefBSR dataset.



Figure 3: Qualitative comparison with SwinIR [5], MRefSR [7], MFIR [2] and Burstormer [3] on RefBSR dataset.