Towards Real-World Burst Image Super-Resolution: Benchmark and Method

Pengxu Wei[1] Yujing Sun[2] Xingbei Guo[1] Chang Liu[3] Guanbin Li[1] Jie Chen[2] Xiangyang Ji[3] Liang Lin[1]
[1]Sun Yat-sen University
[2]School of Electronic and Computer Engineering, Peking University, Shenzhen, China
[3]Tsinghua University
{weipx3,guoxb7,liguanbin}@mail.sysu.edu.cn, yujingsun1999@gmail.com,
{liuchang2022,xyji}@tsinghua.edu.cn, chenj@pcl.ac.cn, linliang@ieee.org

This supplementary file provides additional dataset analysis and more qualitative visualizations of burst Super-Resolution (SR) results.

A. Additional Dataset Analysis

Pixel Shift in Burst Datasets.

Real-world vs. Synthetic. Due to the difficulty of collecting paired real data, Multi-frame Super-Resolution (MFSR) has been limited by synthetic SR data. To figure out the differences between synthetic and real-world data, we conduct an empirical comparison on our RealBSR. A synthetic burst SR dataset, denoted as RealBSR (synthetic), is hand-crafted. Its pixel shift and subpixel shift are shown in Fig. A (b) and (e). It is observed that it has a relatively uniform distribution of pixel shift. In particular, its pixel shift in the range of (0,1) has a low percentage. The main reason is possible that the hand-craft manner is hard to simulate the real-world subpixel shift. On the contrary, real-world datasets, including DBSR and our RealBSR, have a high percentage of subpixel shifts (Fig. A (a) and (c)) and their subpixel distributions are similar (Fig. A (d) and (f)).

RealBSR vs. BurstSR. In the BurstSR dataset, its burst LR images and HR images are captured with different devices, *i.e.*, cellphone and DSLR. Thus, it has to consider simultaneously the real-world burst image super-resolution and enhancement as a coupled task in a model. Instead, our RealBSR avoids this issue and provides a well-prepared benchmark for the research on

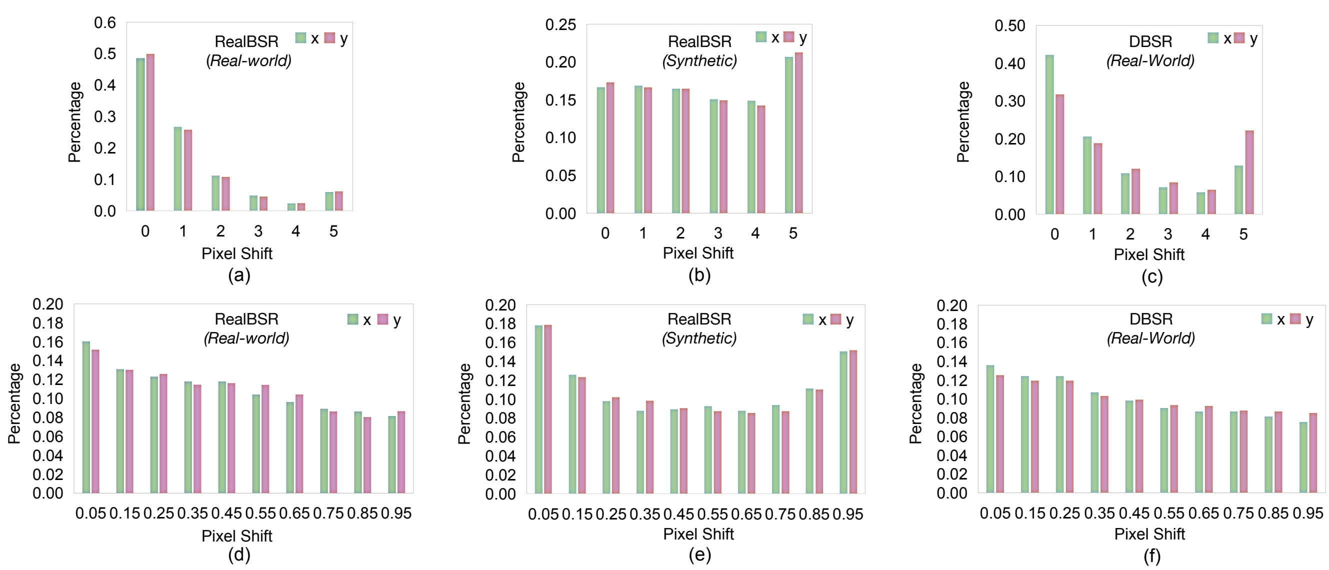

Figure A: Pixel shifts ($a \sim c$) and sub-pixel shifts ($d \sim f$) in burst datasets. We compare our RealBSR (a and d) with its synthetic burst version (b and e) and real-world BurstSR dataset (c and f).

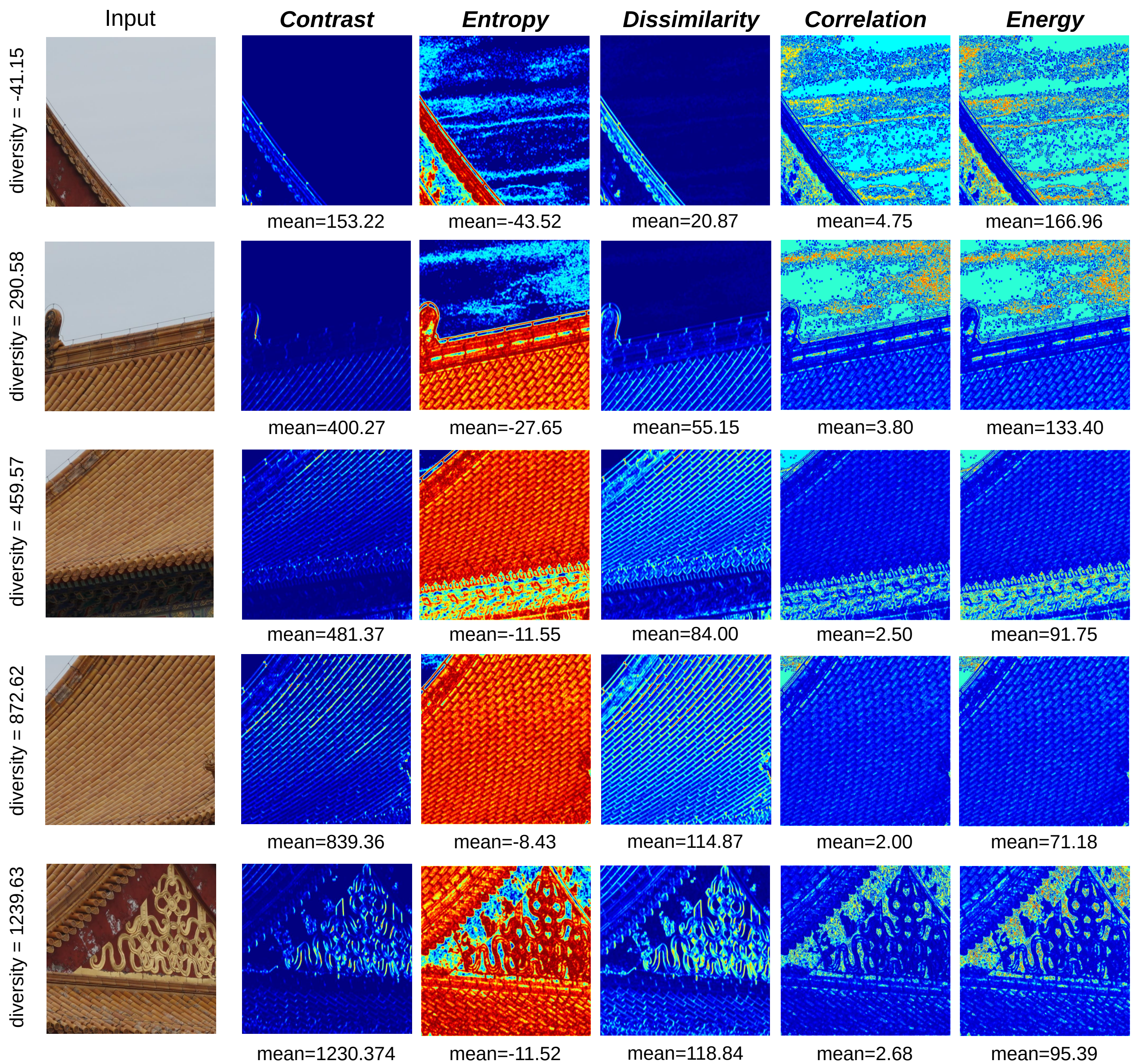

Figure B: Visualization of image diversity analyses in terms of image *contrast*, *entropy*, *dissimilarity*, *correlation* and *energy*. '*mean*' indicates the average of all the elements in a feature matrix. Following [3], '*diversity*' is the sum of the means of five feature matrices.

real-world burst image super-resolution. We compare the pixel shift distribution of BurstSR and RealBSR in Fig. A. Larger pixel shifts beyond the range of (0,1) are observed in BurstSR. This does not indicate the larger difficulty of real-world burst SR in BurstSR, but it instead reflects a problematic phenomenon: BurstSR has a distinct image misalignment and image style change between LR and HR, as claimed in Sec. 3 in the main paper.

Image Diversity.

As claimed in Sec. 3.2, the grey-level co-occurrence matrix (GLCM) is used to analyze the image diversity [2]. Accordingly, based on GLCM, we have five second-order statistic features from all the training images, *i.e.*, Haralick features [2], including image *contrast*, *entropy*, *dissimilarity*, *correlation* and *energy*. We follow their definitions of [3] and provide the illustration

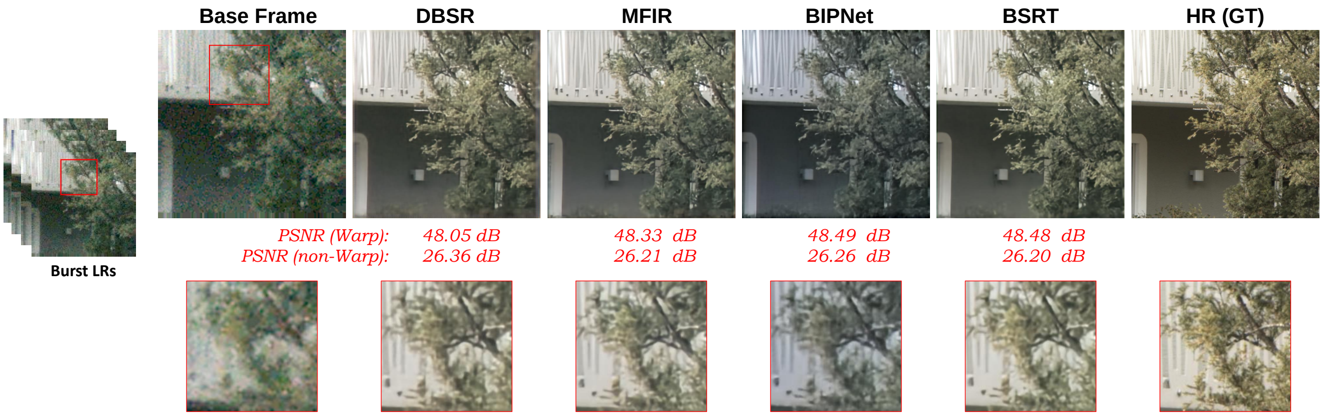

Figure C: Model evaluation of the *0120_0009* image in the BurstSR dataset. 'Warp' means that the primary SR prediction is warped with respect to the ground-truth HR image and then used for evaluation, which is evaluation manner in [1]. 'non-Warp' means the primary SR prediction is directly used for model evaluation. It is observed that the evaluation with warped results has a very high performance, *e.g.*, PSNR, while the SR results do not have a very high image quality. Instead, those SR prediction results have a large gap from the ground-truth HR image in terms of the image sharpness and the image color.

Method	Warp			Non-Warp		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
DBSR	48.05	0.984	0.029	26.36	0.758	0.108
MFIR	48.33	0.985	0.023	26.21	0.751	0.111
BIPNet	48.49	0.985	0.050	26.26	0.758	0.100
BSRT	48.48	0.985	0.021	26.20	0.760	0.110

Table A: The quantitative evaluation for the *0120_0009* image in the BurstSR dataset.

results of images in terms of these five features, as shown in Fig. B.

Model Evaluation in the BurstSR dataset.

As claimed in Sec. 3, To evaluate models on the BurstSR dataset [1], it has an evaluation issue that the predicted SR images are firstly warped with respect to the ground-truth image and then evaluated with the ground-truth image. In C, we provide the SR results of different existing methods following the same evaluation strategy and in Tab. A, we also present detailed quantitative results. It is observed that although models can achieve very high performance, *e.g.*, PSNR, they have a large gap from the ground-truth HR image in terms of the image sharpness and the image color.

References

[1] Goutam Bhat, Martin Danelljan, Luc Van Gool, and Radu Timofte. Deep burst super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9209–9218, 2021. 3

[2] Robert M Haralick, Karthikeyan Shanmugam, and Its' Hak Dinstein. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6):610–621, 1973. 2

[3] Luis Moya, Homa Zakeri, Fumio Yamazaki, Wen Liu, Erick Mas, and Shunichi Koshimura. 3d gray level co-occurrence matrix and its application to identifying collapsed buildings. *ISPRS journal of photogrammetry and remote sensing*, 149:14–28, 2019. 2