Mamba-based Light Field Super-Resolution with Efficient Subspace Scanning Supplementary Material

Ruisheng Gao, Zeyu Xiao, and Zhiwei Xiong (✉)

University of Science and Technology of China, Hefei, China
{grsmc4180, zeyuxiao}@mail.ustc.edu.cn, zwxiong@ustc.edu.cn

1 Relation to MambaIR and Vim

In this section, we compare our method with MambaIR [1] and Vim [4] to gain deeper insights into the design. MambaIR is designed for single-image restoration, while our method focuses on 4D Light Field Super-Resolution (LFSR). We propose a bidirectional subspace scanning technique as an efficient approach for 4D light fields, whereas MambaIR adopts the quad-directional scanning method introduced in Vmamba [2].

We believe that the scanning method for high-dimensional modalities plays a crucial role in improving performance for low-level tasks. To this end, we incorporated the bidirectional scanning method from Vim to implement subspace scanning in our framework. Compared to the quad-directional scanning in Vmamba, the bidirectional approach in Vim is more efficient and better suited for interactions among subspaces. This suggests that increasing the number of scanning operations is not always advantageous.

Moreover, we observed that the number of scanning operations significantly impacts the network runtime. Therefore, balancing performance and efficiency by optimizing the number of scanning operations is an important area for further investigation.

2 Additional Experiments

2.1 Angular Consistency

To further evaluate the angular consistency of SR results generated by our method, we provide depth estimation results ($4\times$ scale on the Tarot scene of the STFGantry testset for example) compared with baseline methods in Fig. 1. We adopt OCC [3] as the LF depth estimator. It is clear that our method achieves comparable depth estimation results with baseline methods quantitatively. With full-image inference, the result produced by our method has the least mean square error and is competitive with baseline methods qualitatively.

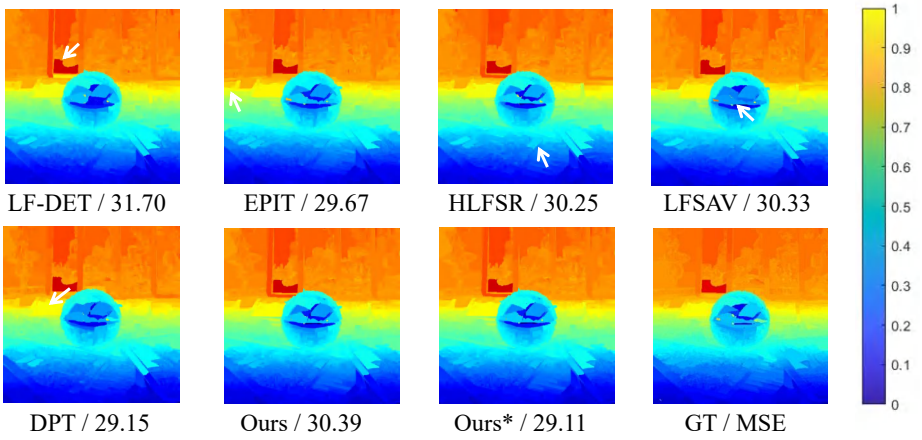

Fig. 1: Quantitative and qualitative (MSE) comparisons of depth estimation results using OCC [3].

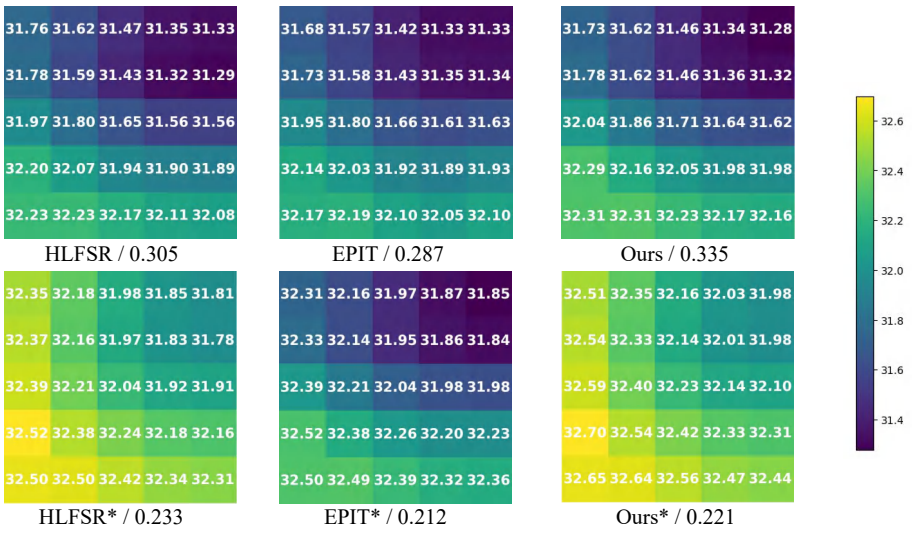

Fig. 2: Perspective performance comparison (PSNR) on the EPFL-Bikes scene (method / standard deviation). * denotes result using full-resolution inference.

.

2.2 Perspective Performance Distribution

We provide perspective performance distribution comparison between HLFSR, EPIT and our method using both path-based inference and whole-resolution inference, as shown in Fig. 2. We adopt the standard deviation among views to evaluate the performance distribution. With extra information supplemented from whole resolution inputs, the PSNR values among views become more balanced

which shows the merit of full resolution inference. Our method has comparable balanced PSNR distribution while achieves better PSNR performance.

2.3 Additional Visual Results

We provide additional visual results on 4× LFSR and 2× LFSR in this section, as shown in Fig. 3 and Fig. 4.

References

1. Guo, H., Li, J., Dai, T., Ouyang, Z., Ren, X., Xia, S.T.: Mambair: A simple baseline for image restoration with state-space model. In: ECCV (2024)
2. Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., Liu, Y.: Vmamba: Visual state space model. Advances in Neural Information Processing Systems (2024)
3. Wang, T.C., Efros, A.A., Ramamoorthi, R.: Occlusion-aware depth estimation using light-field cameras. In: Proceedings of the IEEE international conference on computer vision. pp. 3487–3495 (2015)
4. Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., Wang, X.: Vision mamba: Efficient visual representation learning with bidirectional state space model. In: Forty-first International Conference on Machine Learning (2024)

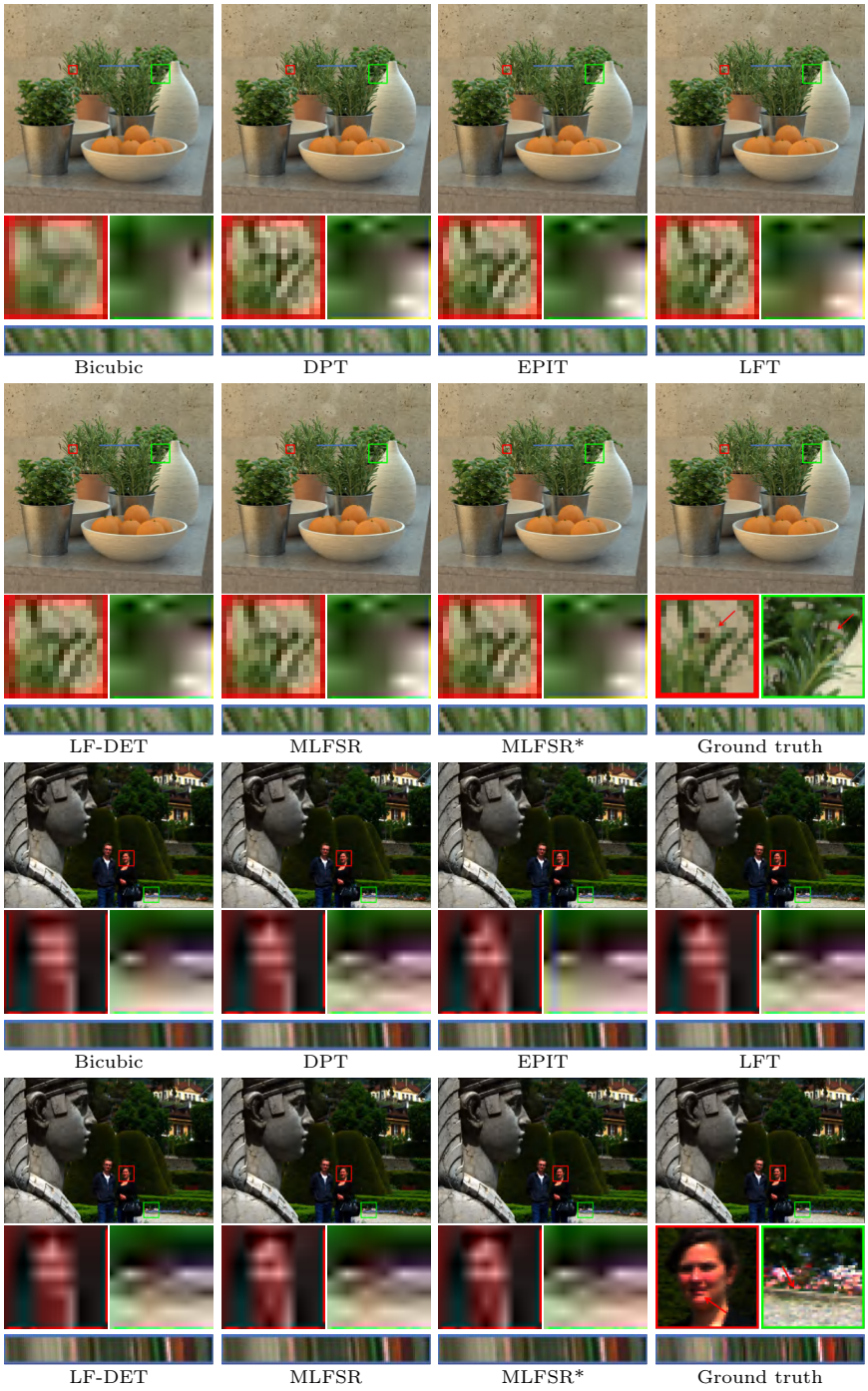

Fig. 3: Visual comparisons of different methods on 2× SR (view coordinates: (2, 2)). Please zoom in for better visualization and best viewing on screen.

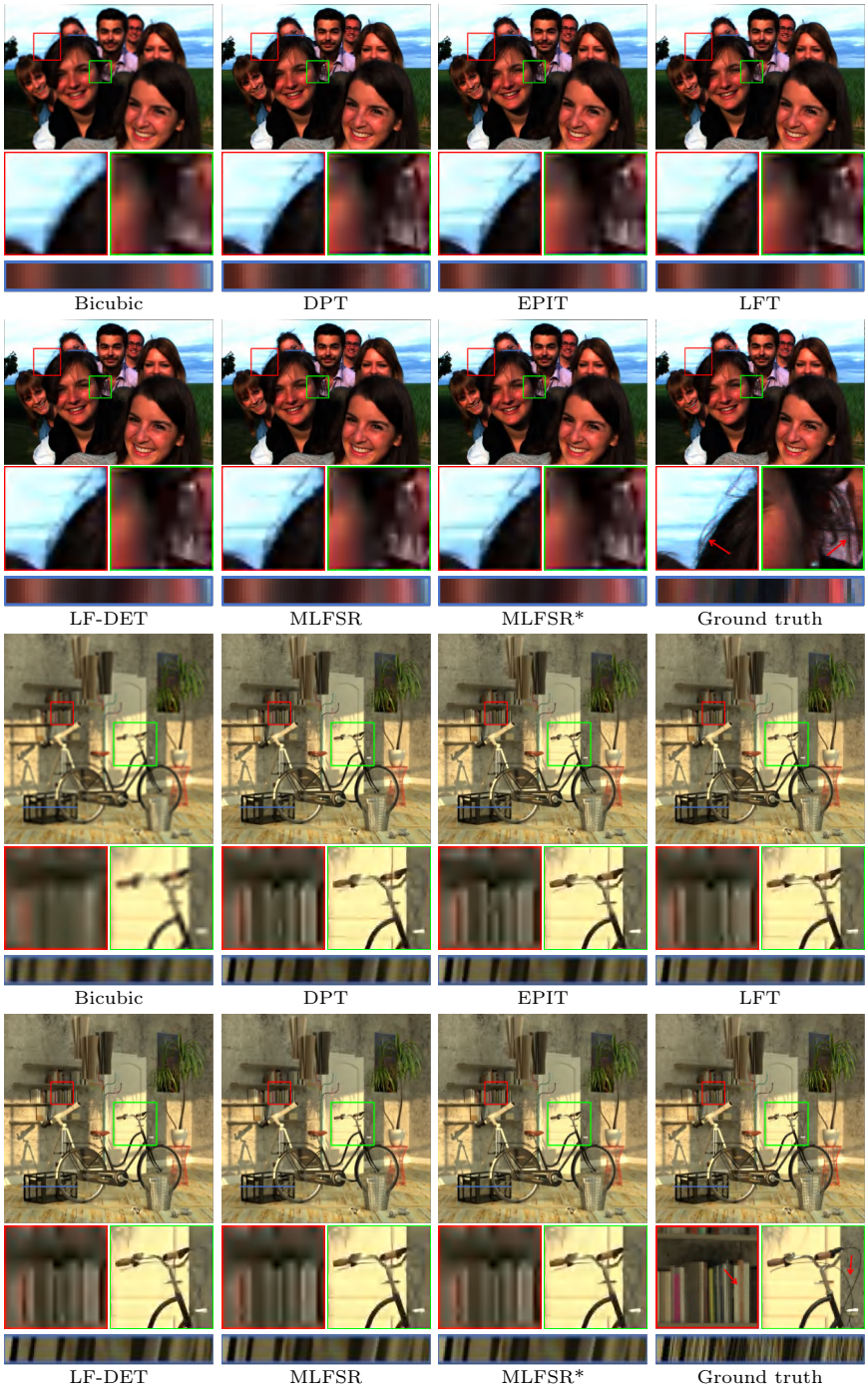

Fig. 4: Visual comparisons of different methods on 4× SR (view coordinates: (2, 2)). Please zoom in for better visualization and best viewing on screen.