Omni Aggregation Networks for Lightweight Image Super-Resolution
Supplementary Material

Hang Wang[2*] Xuanhong Chen[1*] Bingbing Ni[1,2†] Yutian Liu[1] Jinfan Liu[1]
[1]Shanghai Jiao Tong University, Shanghai 200240, China [2] Huawei
francis970625@gmail.com {chen19910528,nibingbing,stau7001,ljflnjz}@sjtu.edu.cn

In this part, we provide the following material: (1) Sec. 1 describes detailed architectures of Omni-SR; (2) Sec. 2 presents ablation study for aggregation order in OSAG; (3) Sec. 3 demonstrates LAM [3] comparisons of different lightweight methods.

1. Detailed Network Architecture

Our OSA blocks follow typical Transformer designs with Feedforward network (FFN) and LayerNorm, and the only difference is that self-attention operation is replaced with our proposed OSA operator. For FFN, we adopt the GDFN proposed by Restormer [7]. The detailed architecture of OSA block is shown in Figure 1.

We use Local-Conv block, Meso-OSA block and Global-OSA block to build the OSAG for omni-scale aggregation. In experiments, we set the number of OSAG as 5 to make the model size around 800K for a fair comparison with other methods. We can further reduce the OSAG number to get smaller Omni-SR models. When OSAG number is reduced from 5 to 1, model parameters are reduced from 792K to 211k (792K → 647K → 502K → 356K → 211K).

2. Impact of aggregation order.

We adopt three different interaction modules (i.e, local, meso and global) to build the Omni-Scale Aggregation Group (OSAG), aiming for omni-scale information aggregation from local to global. In order to explore the most effective combination, we conduct ablation experiments with different combination orders for ×4 SR on Urban100 and present the results in Table 1. From the table, we can observe that placing Local-Conv at the front achieves pleasant result compared to other variant models, which demonstrates that it is more suitable to perform local feature aggregations and then aggregate them globally. This phenomenon is consistent with the results of some recent hybrid models in image classification [2,6] and image restoration [5]. The Local-Conv → Meso-OSA → Global-OSA setting obtains the best SR performance, and therefore we employ this combination order in all experiments.

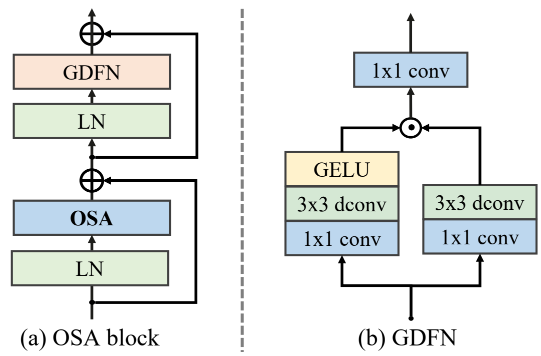

Figure 1. Detailed architecture of the proposed OSA block.

Aggregation Order	PSNR / SSIM
Global-OSA → Meso-OSA → Local-Conv	26.56 / 0.8017
Global-OSA → Local-Conv → Meso-OSA	26.55 / 0.8016
Meso-OSA → Local-Conv → Global-OSA	26.49 / 0.8014
Meso-OSA → global-Conv → Local-Conv	26.52 / 0.8015
Local-Conv → Global-OSA → Meso-OSA	26.60 / 0.8018
Local-Conv → Meso-OSA → Global-OSA	26.64 / 0.8018

Table 1. Aggregation Order study in OSAG.

3. LAM comparison.

To further evaluate the effectiveness of Omni-SR, we compare the LAM [3] results of our model with other advanced lightweight methods (i.e., CARN [1], ESRT [5] and SwinIR [4]) for ×4 SR. LAM tool is proposed to study the interactive capabilities of models, and diffusion index (DI) [3] is introduced to quantify the above-mentioned interaction range. The larger LAM range and DI value indicate a larger receptive field (i.e., interaction scope) for the model, which yields better performance. From Figure 2, it can be observed that our Omni-SR achieves largest LAM range and DI value compared with CNN-based or transformer-based models, verifying the effectiveness of the proposed omni-dimensional and omni-scale aggregation scheme. This phenomenon is also in line with the better performance of our method on five benchmark datasets.

*Equal Contribution.
†Corresponding author: Bingbing Ni.

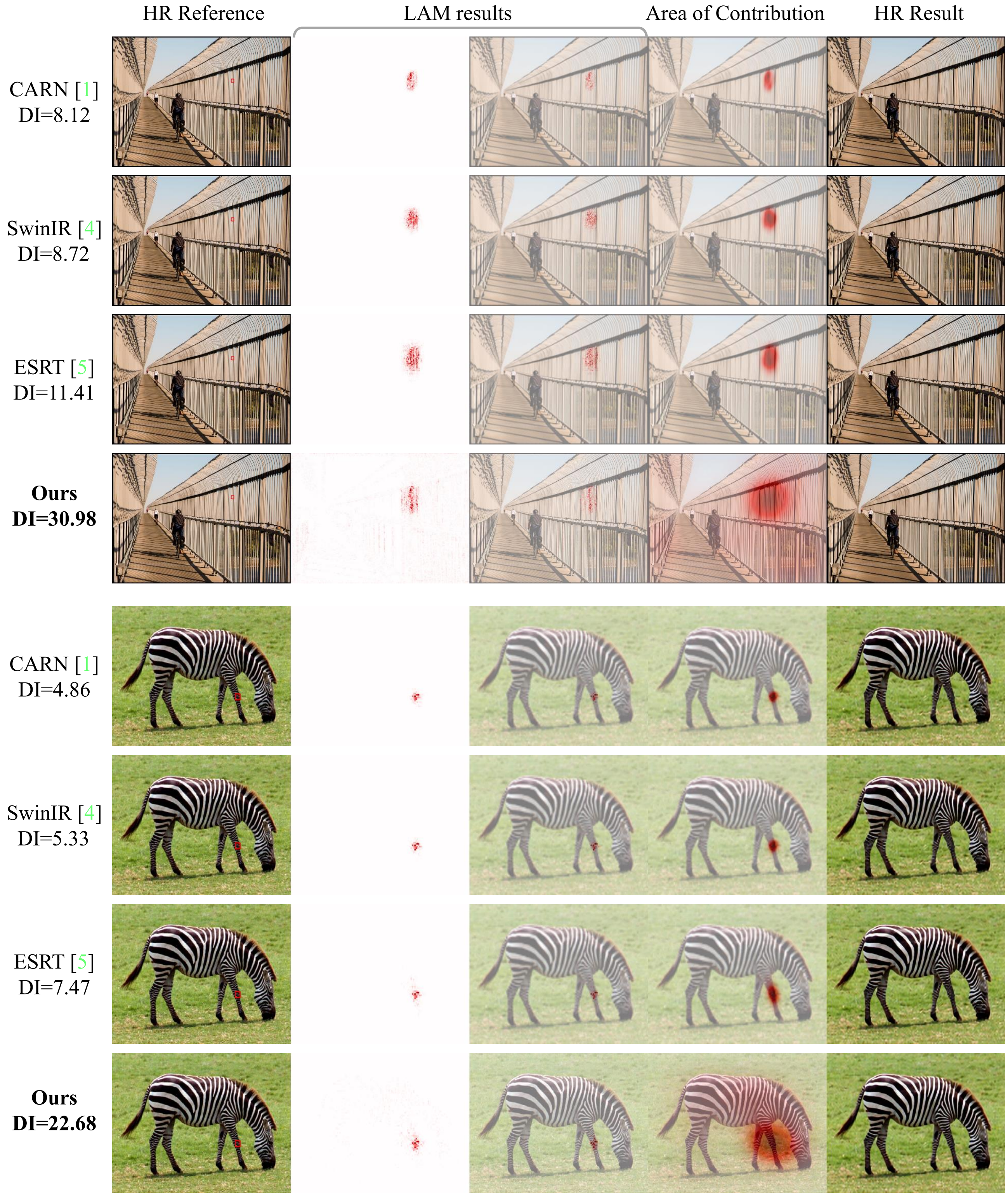

Figure 2. LAM [3] comparison for $\times 4$ SR on Urban100: `img_012.png` and Set14: `zebra.png`. LAM tool is proposed to study the interactive capabilities of models, and diffusion index (DI) [3] is introduced to quantify the above-mentioned interaction range. The larger LAM range and DI value indicate a larger receptive field (i.e., interaction scope), which yields for better performance. As shown in the figure, our Omni-SR achieves the largest LAM range and DI value, thus obtaining the best performance.

References

[1] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *ECCV*, 2018. 1

[2] Zihang Dai, Hanxiao Liu, Quoc V. Le, and Mingxing Tan. Coatnet: Marrying convolution and attention for all data sizes. In *NeurIPS*, pages 3965–3977, 2021. 1

[3] Jinjin Gu and Chao Dong. Interpreting super-resolution networks with local attribution maps. In *CVPR*, 2021. 1, 2

[4] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *ICCV*, 2021. 1

[5] Zhisheng Lu, Juncheng Li, Hong Liu, Chaoyan Huang, Linlin Zhang, and Tieyong Zeng. Transformer for single image super-resolution. In *CVPR Workshops*, 2022. 1

[6] Tete Xiao, Mannat Singh, Eric Mintun, Trevor Darrell, Piotr Dollár, and Ross B. Girshick. Early convolutions help transformers see better. In *NeurIPS*, pages 30392–30400, 2021. 1

[7] 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