DLGSANet: Lightweight Dynamic Local and Global Self-Attention Network for Image Super-Resolution

Xiang Li, Jiangxin Dong, Jinhui Tang*, and Jinshan Pan*
School of Computer Science and Engineering, Nanjing University of Science and Technology

Overview

We appreciate the reviewers taking the time to look at the supplemental material. In this document, we continuously demonstrate the effectiveness of our method using more experimental results. In Section 1, we show the effectiveness of the HDTB with additional qualitative evaluations. In Section 2, we add some extra qualitative comparisons to demonstrate the effectiveness of the MHDLSA. In Section 3, we add some extra qualitative comparisons in more detail to show the effectiveness of the SparseGSA. In Section 4, we provide more quantitative and qualitative evaluations between the DLGSANet and state-of-the-art methods.

1. Effectiveness of the HDTB

In the submitted manuscript, we have shown the effectiveness of the HDTB in Table 3 of Section 5. To further illustrate the effectiveness of the proposed HDTB on image super-resolution, we provide more qualitative comparisons on the Urban100 dataset [2]. Figures 1-5 show that using the proposed HDTB generates better results. Additionally, Table 1 further demonstrates that when competing parameters are used, increasing the head number of the self-attention to 6 yields in superior PSNR values.

Table 1. Ablation study of the number of heads in the HDTB. The results ($\times 4$) are obtained from the Urban100 dataset.

heads	3	6	9	15
#Params	5.17M	4.76M	4.63M	4.52M
PSNR	26.89	27.17	27.13	27.10

2. Effectiveness of the MHDLSA

In the submitted manuscript, we compare the MHDLSA with the commonly used MHSA in Table 5 of Section 5. To further illustrate the effectiveness of the proposed MHDLSA on image super-resolution, we provide more qualitative comparisons on the Urban100 dataset [2]. Figures 6-10 show that using the proposed MHDLSA generates better results. By altering the kernel size of the Dynamic convolution, we further examine the effectiveness of the MHDLSA. Table 2 demonstrates that increasing the kernel size to 7 yields greater PSNR values.

Table 2. Ablation study of the Kernel Size of the Dynamic Convolution in the MHDLSA. The results ($\times 4$) are obtained from the Urban100 dataset.

kernel size	3	5	7	9
PSNR	26.98	27.16	27.17	27.16

*Corresponding author

3. Effectiveness of the SparseGSA

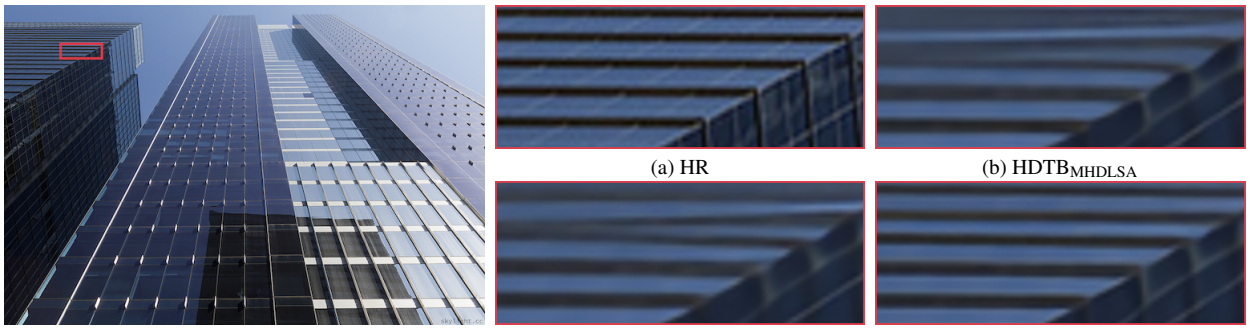
In the submitted manuscript, we compare the SparseGSA with the commonly used GSA in Table 4 and Figure 6 of Section 5. In this document, we further demonstrate the effectiveness of the SparseGSA by showing qualitative comparisons (Figures 11-13).

4. More Quantitative & Qualitative Evaluations

In this section, we provide more quantitative and qualitative evaluations of the proposed DLGSANet and state-of-the-art methods. As shown in Table 3, by stacking a deeper network, our proposed DLGSANet-XL outperforms state-of-the-art methods with similar parameters (8.97M) and lower computational costs. Figures 14-18 show that the proposed DLGSANet generates better images with finer structural details.

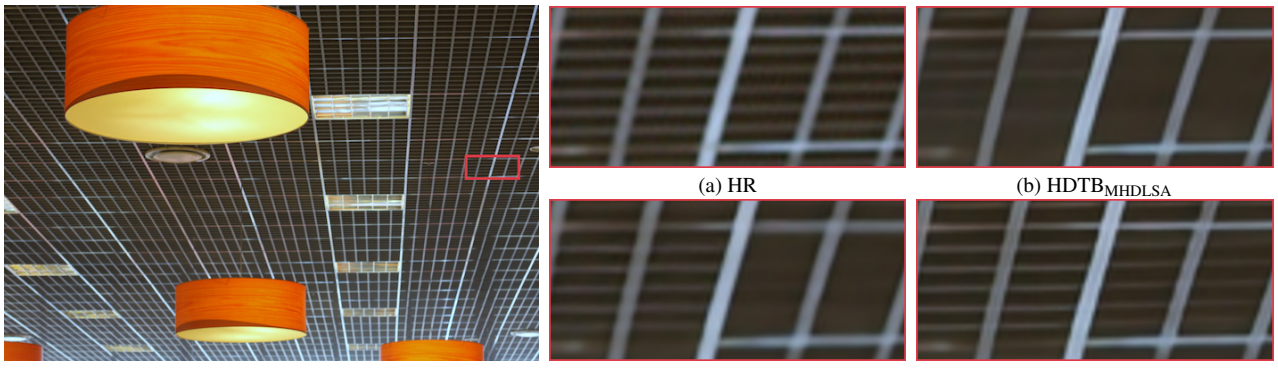
Table 3. Quantitative evaluations of the proposed DLGSANet against state-of-the-art methods on the Urban100 dataset ($\times 4$).

Model	#Params(/M)	#FLOPs(/G)	PSNR
RCAN	15.59	918	26.82
SwinIR	11.90	584	27.07
ELAN	8.31	494	27.13
DLGSANet (Ours)	4.76	274	27.17
DLGSANet-XL (Ours)	8.97	418	27.28



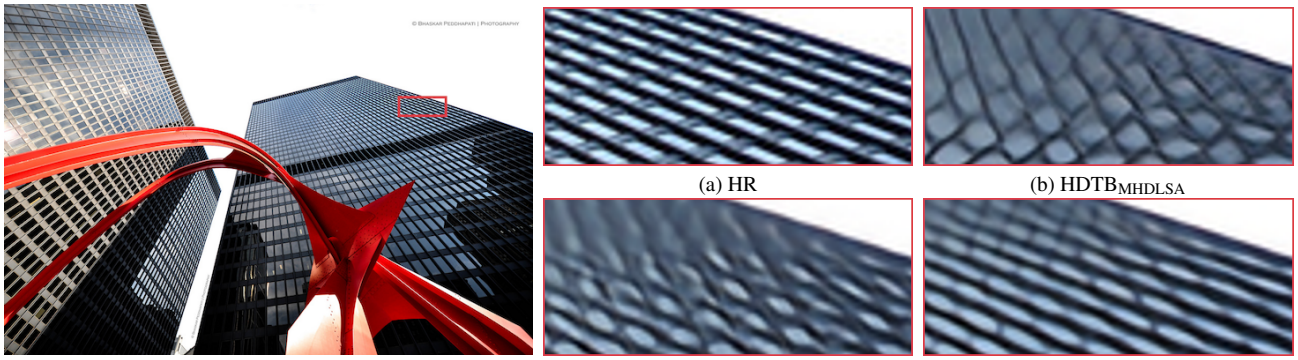
(a) HR (b) $\text{HDTB}_{\text{MHDLSA}}$ Urban-img033 (c) $\text{HDTB}_{\text{SparseGSA}}$ (d) HDTB

Figure 1. Effect of the MHDLSA and the SparseGSA in the HDTB for SISR. The results ($\times 4$) are obtained from the "img033" image of the Urban100 dataset. (b) and (c) show that only using the MHDLSA or the SparseGSA in the HDTB does not restore the structures well. In contrast, using both the MHDLSA and the SparseGSA in the HDTB leads to finer structural details (d).



(a) HR (b) $\text{HDTB}_{\text{MHDLSA}}$ Urban-img044 (c) $\text{HDTB}_{\text{SparseGSA}}$ (d) HDTB

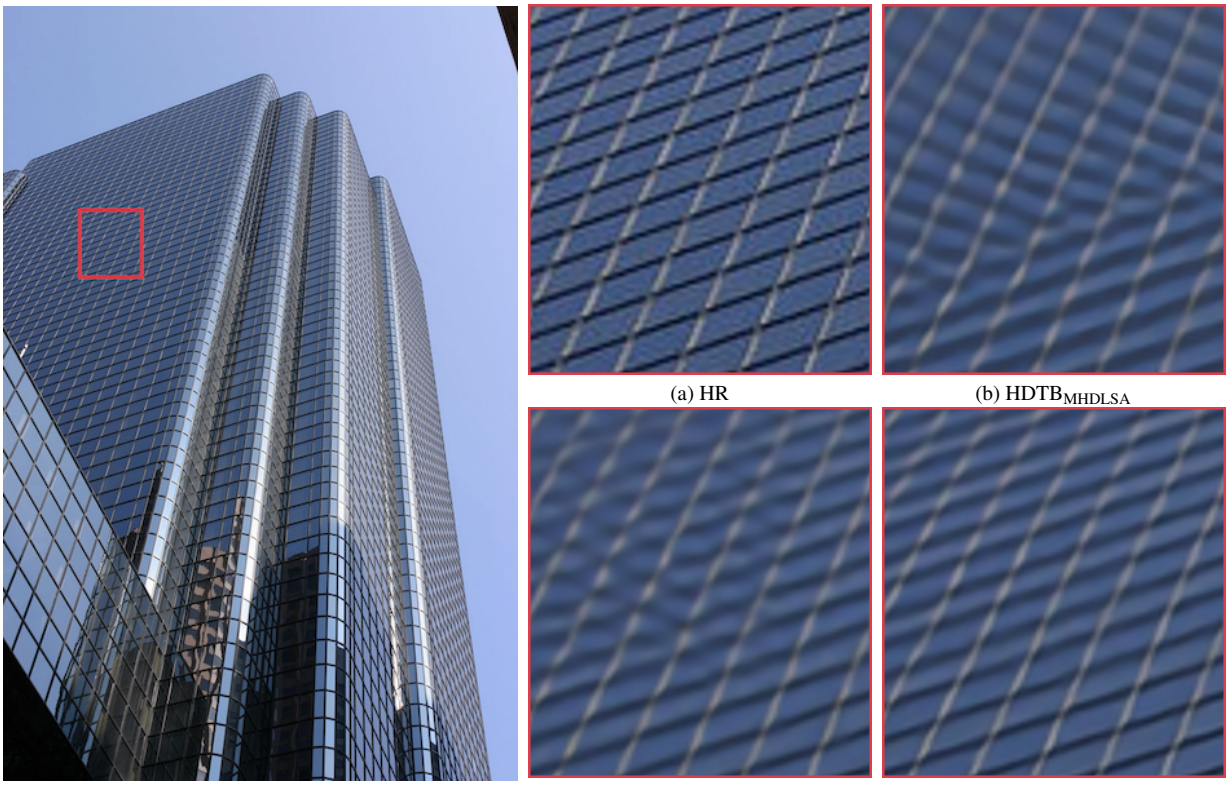
Figure 2. Effect of the MHDLSA and the SparseGSA in the HDTB for SISR. The results ($\times 4$) are obtained from the "img044" image of the Urban100 dataset. (b) and (c) show that only using the MHDLSA or the SparseGSA in the HDTB produces blurred structural details of the ceiling. In contrast, using both the MHDLSA and the SparseGSA in the HDTB leads to a better super-resolved image with finer structural details (d).



(a) HR (b) $\text{HDTB}_{\text{MHDLSA}}$

Urban-img062 (c) $\text{HDTB}_{\text{SparseGSA}}$ (d) HDTB

Figure 3. Effect of the MHDLSA and the SparseGSA in the HDTB for SISR. The results ($\times 4$) are obtained from the "img062" image of the Urban100 dataset. (b) and (c) show that only using the MHDLSA or the SparseGSA in the HDTB does not restore the structures well. In contrast, using both the MHDLSA and the SparseGSA in the HDTB leads to a clearer image, where the structures of windows are recovered well (d).



(a) HR (b) $\text{HDTB}_{\text{MHDLSA}}$

Urban-img074 (c) $\text{HDTB}_{\text{SparseGSA}}$ (d) HDTB

Figure 4. Effect of the MHDLSA and the SparseGSA in the HDTB for SISR. The results ($\times 4$) are obtained from the "img074" image of the Urban100 dataset. (b) and (c) demonstrate that the correct structures are not restored when the HDTB merely uses the MHDLSA or the SparseGSA. The MHDLSA and the SparseGSA in the HDTB (d), however, produce a clear picture with the better structural details.

References

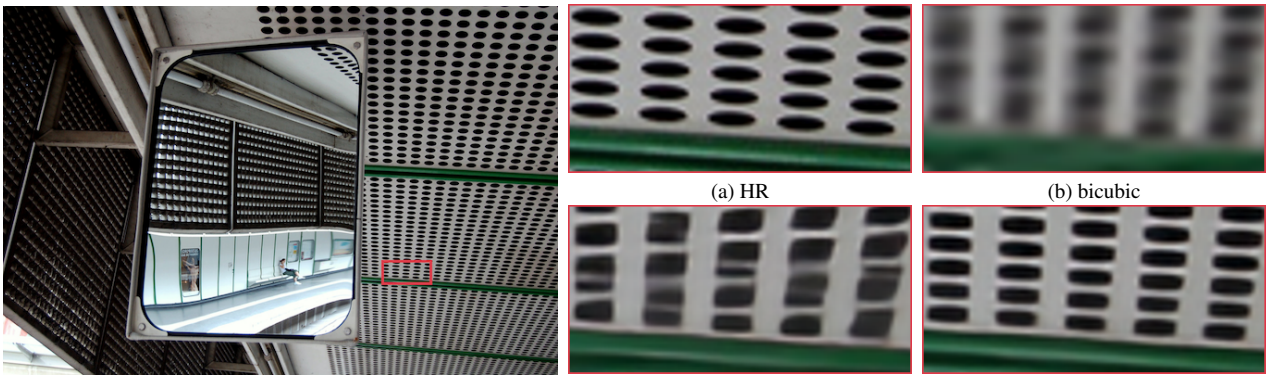
[1] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *CVPR*, 2019. 8, 9

[2] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *CVPR*, 2015. 1

[3] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image restoration using swin transformer. In *ICCV Workshops*, 2021. 8, 9

Urban-img076 (a) HR (b) $\text{HDTB}_{\text{MHDLSA}}$ (c) $\text{HDTB}_{\text{SparseGSA}}$ (d) HDTB

Figure 5. Effect of the MHDLSA and the SparseGSA in the HDTB for SISR. The results ($\times 4$) are obtained from the "img076" image of the Urban100 dataset. (b) and (c) demonstrate that the building's structures are not restored by using simply the MHDLSA or the SparseGSA in the HDTB. On the other hand, using both the MHDLSA and the SparseGSA in the HDTB yields a clear recognition of the structural elements of the building (d).



Urban-img004 (a) HR (b) bicubic (c) $\text{HDTB}_{\text{MHSA}}$ (d) $\text{HDTB}_{\text{MHDLSA}}$

Figure 6. Effect of the MHDLSA for SISR. The results ($\times 4$) are obtained from the "img004" image of the Urban100 dataset. (c) shows that using the MHSA in the HDTB does not effectively restore the structural details of the ceiling. In contrast, using the MHDLSA in the HDTB results in more distinct structural details of the ceiling (d).

[4] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *CVPR Workshops*, 2017. 8, 9

[5] Yiqun Mei, Yuchen Fan, and Yuqian Zhou. Image super-resolution with non-local sparse attention. In *CVPR*, 2021. 8, 9

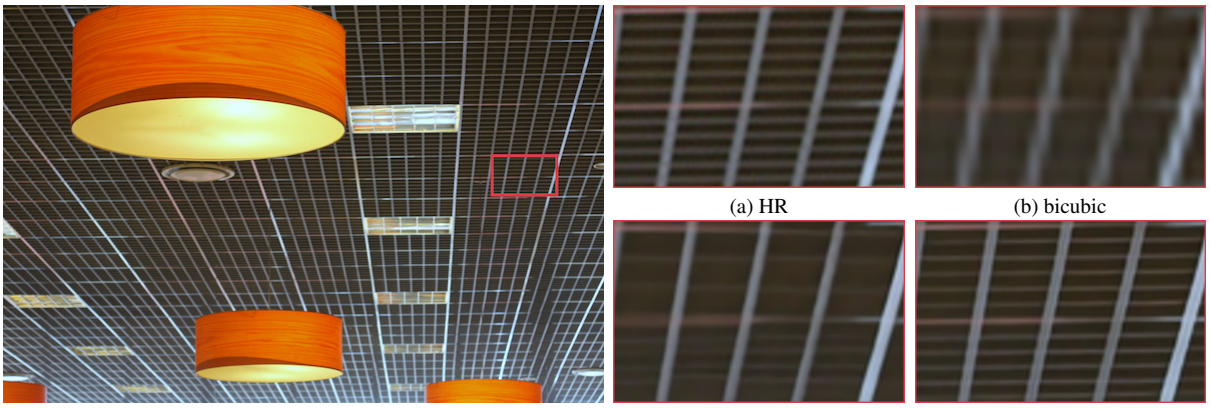
[6] Ben Niu, Weilei Wen, Wenqi Ren, Xiangde Zhang, Lianping Yang, Shuzhen Wang, Kaihao Zhang, Xiaochun Cao, and Haifeng Shen. Single image super-resolution via a holistic attention network. In *ECCV*, 2020. 8, 9

[7] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *ECCV*, 2018. 8, 9

(a) HR (b) bicubic

Urban-img012 (c) HDTB$_{\text{MHSA}}$ (d) HDTB$_{\text{MHDLSA}}$

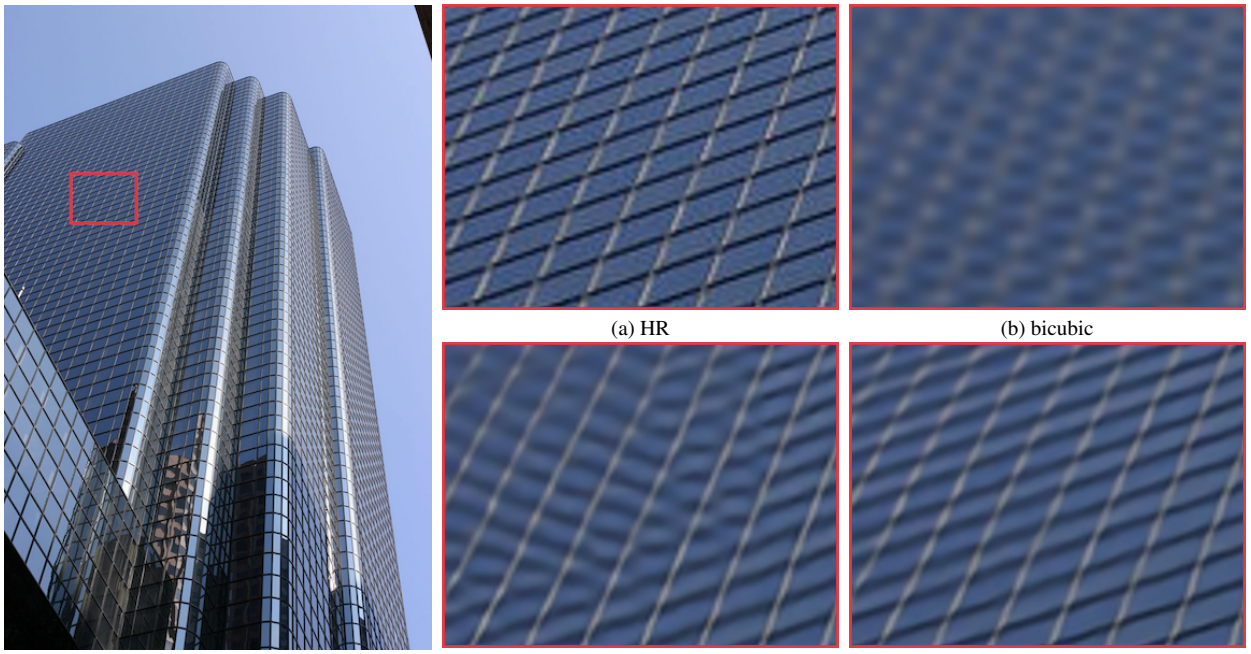
Figure 7. Effect of the MHDLSA for SISR. The results ($\times 4$) are obtained from the "img012" image of the Urban100 dataset. (c) shows that using the MHSA in the HDTB does not restore the windows of the building well. On the other hand, using the MHDLSA in the HDTB yields a clear image with finer window structural details (d).



(a) HR (b) bicubic

Urban-img044 (c) HDTB$_{\text{MHSA}}$ (d) HDTB$_{\text{MHDLSA}}$

Figure 8. Effect of the MHDLSA for SISR. The results ($\times 4$) are obtained from the "img044" image of the Urban100 dataset. (c) shows that using the MHSA in the HDTB produces a blurred ceiling. In contrast, using the MHDLSA in the HDTB leads to a better super-resolved image with finer structural details (d).



(a) HR (b) bicubic

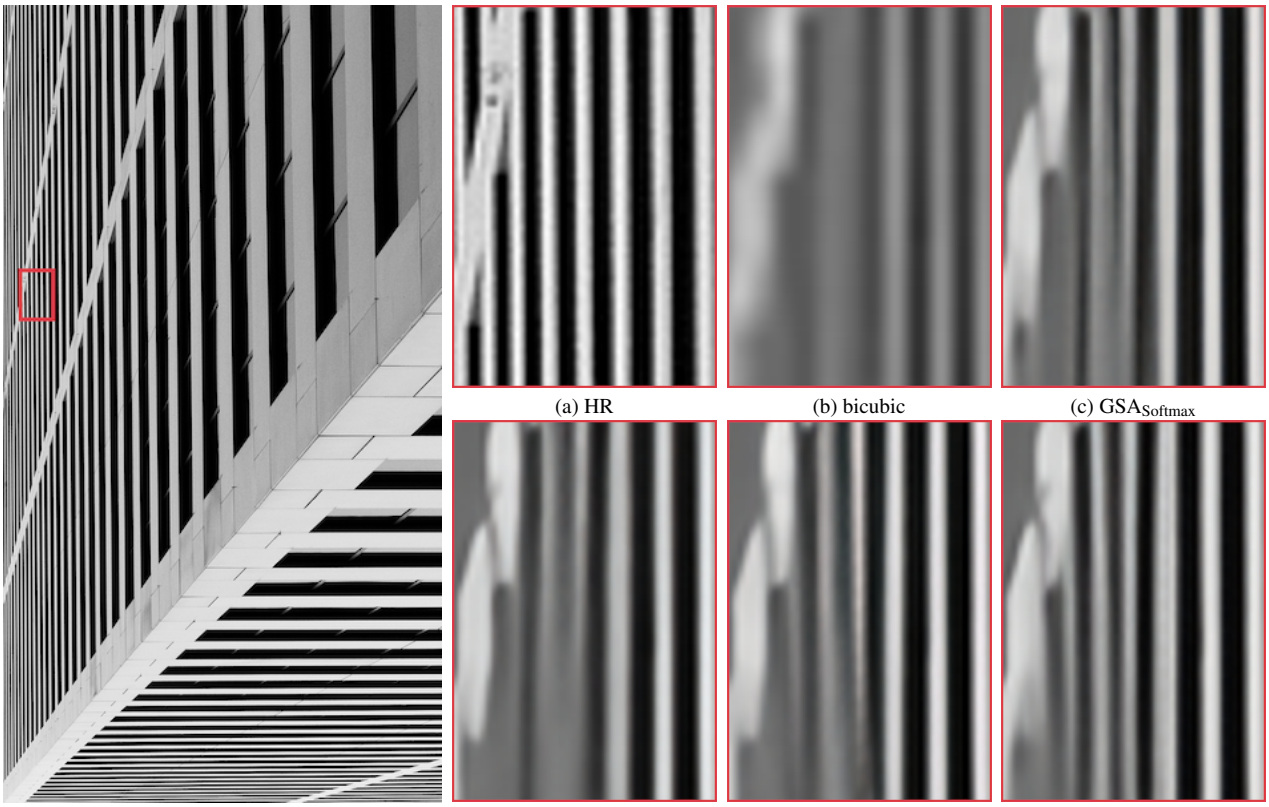
Urban-img074 (c) $\text{HDTB}_{\text{MHSA}}$ (d) $\text{HDTB}_{\text{MHDLSA}}$

Figure 9. Effect of the MHDLSA for SISR. The results ($\times 4$) are obtained from the "img074" image of the Urban100 dataset. (c) shows that using the MHSA in the HDTB produces blurred structural details. However, using the MHDLSA in the HDTB leads to a clear picture with the better structural details (d).

(a) HR (b) bicubic

Urban-img076 (c) $\text{HDTB}_{\text{MHSA}}$ (d) $\text{HDTB}_{\text{MHDLSA}}$

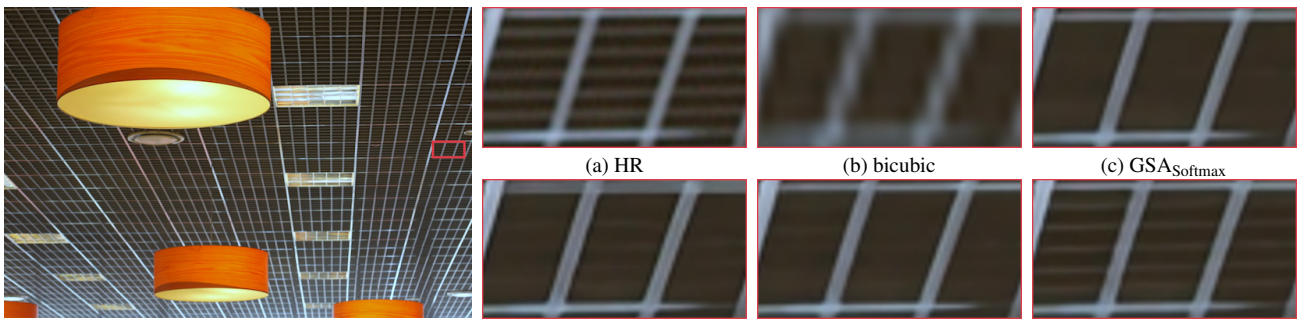
Figure 10. Effect of the MHDLSA for SISR. The results ($\times 4$) are obtained from the "img076" image of the Urban100 dataset. (c) shows that using the MHSA in the HDTB produces blurred LED screens. In contrast, using the MHDLSA in the HDTB leads to clearer LED screens with finer details (d).



(a) HR (b) bicubic (c) $GSA_{Softmax}$

Urban-img011 (d) $GSA_{Sigmoid}$ (e) GSA_{GELU} (f) SparseGSA

Figure 11. Effect of the SparseGSA for SISR. The results ($\times 4$) are obtained from the "img011" image of the Urban100 dataset. (c), (d) and (e) show that the $GSA_{Softmax}$, the $GSA_{Sigmoid}$ and the GSA_{GELU} do not restore the structures well. In contrast, using the SparseGSA leads to finer structural details (f).



(a) HR (b) bicubic (c) $GSA_{Softmax}$

Urban-img044 (d) $GSA_{Sigmoid}$ (e) GSA_{GELU} (f) SparseGSA

Figure 12. Effect of the SparseGSA for SISR. The results ($\times 4$) are obtained from the "img044" image of the Urban100 dataset. (c), (d) and (e) show that the $GSA_{Softmax}$, the $GSA_{Sigmoid}$ and the GSA_{GELU} produce blurred structures of the ceiling. In contrast, using the SparseGSA leads to a better super-resolved image with finer structural details (f).

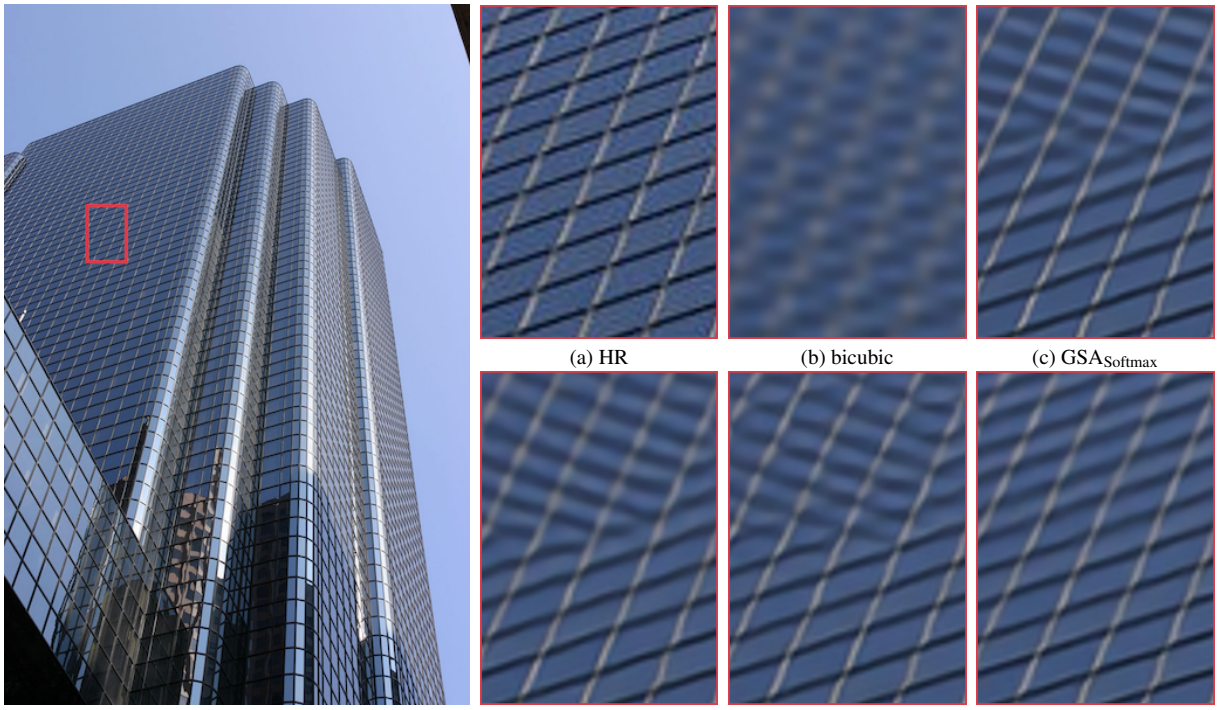


Figure 13. Effect of the SparseGSA for SISR. The results ($\times 4$) are obtained from the "img074" image of the Urban100 dataset. (c), (d) and (e) show that the GSA$_{\text{Softmax}}$, the GSA$_{\text{Sigmoid}}$ and the GSA$_{\text{GELU}}$ do not restore the structures well. In contrast, using the SparseGSA leads to a clearer image with finer structural details (f).

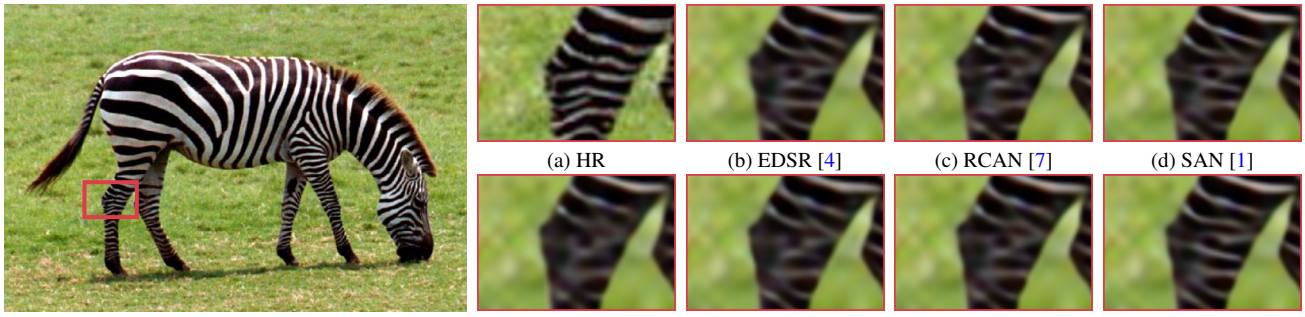


Figure 14. Super-resolution results ($\times 4$) on the "zebra" image from the Set14 dataset. The stripes on the zebra's leg are not recovered well by the evaluated methods.

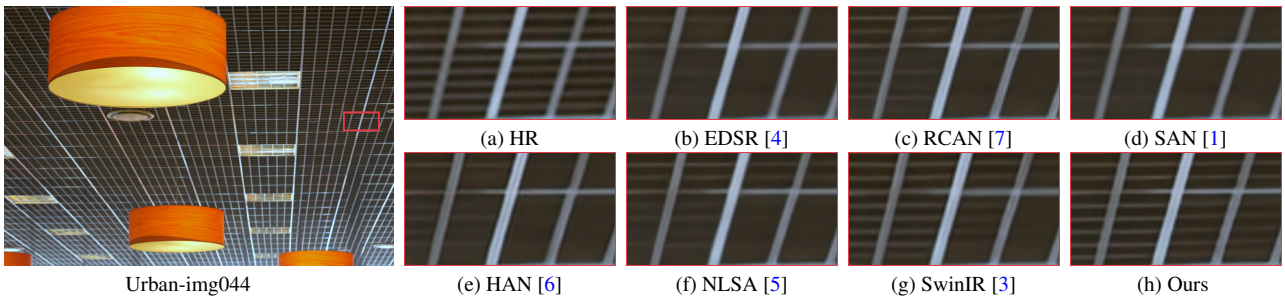


Figure 15. Super-resolution results ($\times 4$) on the "img044" image from the Urban100 dataset. The structures of the ceiling are not recovered well by the evaluated methods. Ours can efficiently produce a better super-resolved image with finer structural details (h).

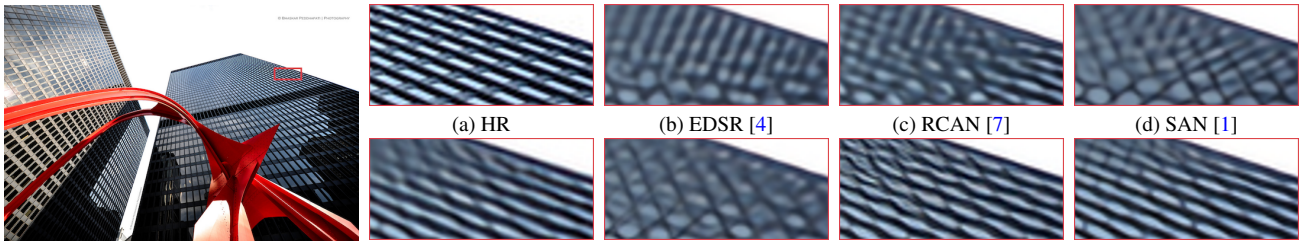


Figure 16. Super-resolution results ($\times 4$) on the "img062" image from the Urban100 dataset. Ours produces a clearer image with finer structural details (h).

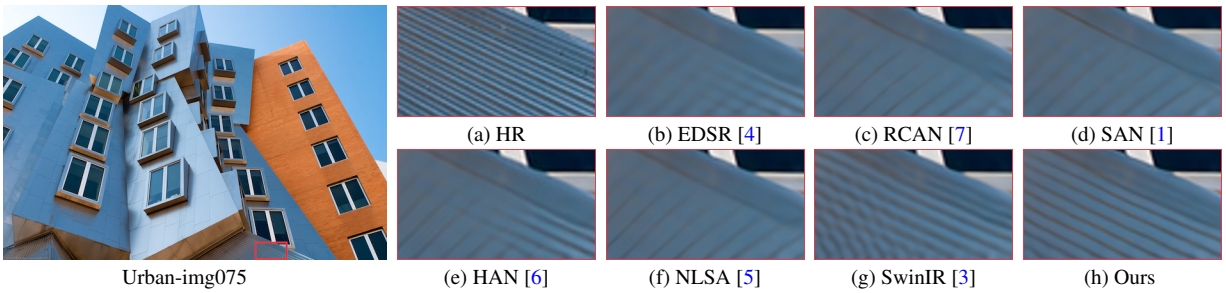


Figure 17. Super-resolution results ($\times 4$) on the "img075" image from the Urban100 dataset. Ours produces a clearer image with more structural details (h).

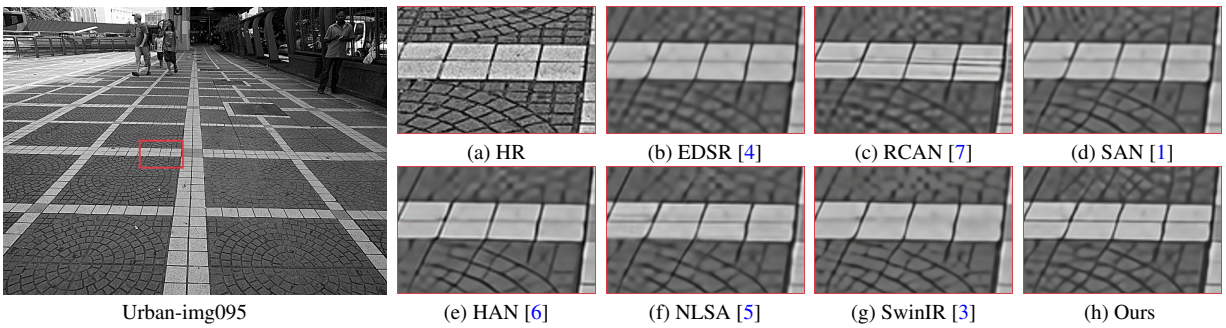


Figure 18. Super-resolution results ($\times 4$) on the "img095" image from the Urban100 dataset. Due to the better local aggregation ability of the MHDLSA, Our model is able to produce clearer road bricks with well-restored patterns (h).