

Efficient Single Image Super-Resolution with Entropy Attention and Receptive Field Augmentation

Anonymous Authors

This is the supplementary materials for **Efficient Single Image Super-Resolution with Entropy Attention and Receptive Field Augmentation**. We mainly provide further analysis of the effective receptive field and more visualization results here.

1 COMPARISON OF EFFECTIVE RECEPTIVE FIELD IMAGE

A feature point in convolutional networks depends on a region of the input, which is termed as receptive field (RF) for this feature point. RF can be described as the maximum amount of information a feature point can capture, while the effective receptive field (ERF) [7] can be viewed as the efficiency of information capture for each feature point. Here, we compared the ERF of several advanced methods, including EDSR [6], SwinIR [5], and SRFormer [11].

Fig. 1 demonstrates the comparison between the ERFs of these methods. It can be observed that the ERF of our **EARFA** covers a larger area than other methods, covering almost the entire input image. This indicates that our **SLKA** endows the proposed method with the ability to perceive a larger area and output more precise SR results.

The superiority of our method in terms of ERF shown in Fig. 1 is inapparent. Therefore, we further illustrate the Area Ratio and Rectangle Side Length of covered by the ERF with the same contribution threshold, as shown in Fig. 3, where the horizontal axis represents the proportion of included weights to total weights, the points in the plot are positioned at [0.2, 0.3, 0.5, 0.95]. The vertical axis on the left side of the graph indicates the area of the central window required to include the specified proportion of weights. The vertical axis on the right side of the graph indicates the side length of the central window required to include the specified proportion of weights. The image size used for testing ERF [7] is 128×128 . When the threshold is set to 0.95, we find that the side length of the central window for EDSR [6] is 45, indicating that it only requires a window size of 45×45 to cover 95% of the weights. This suggests that its ERF [7] is small, with less concentrated and diverse effective information, and insufficient feature diversity. SwinIR [5] requires a central window size of 99×99 to cover 95% of the weights, while SRFormer [11] requires a central window size of 119×119 , indicating that SRFormer [11] contains more information richness compared to SwinIR [5], possessing certain feature diversity. In contrast, our method requires a central window size of 125×125 to cover 95% of the weights. At this point, the relationship between the side length of the central window and the covered weight range is almost linear, suggesting that our method has a global receptive field, containing a significant amount of effective information.

2 EARFA-LIGHT'S VISION RESULTS

In the main text, due to space constraints, we did not provide visual comparisons between EARFA-light and other state-of-the-art methods. In this case, we visualize the graph of **EARFA-light** and

Bicubic [4], FSRCNN [1], IMDN [3], PAN [10], ShuffleMixer [9], SAFMN [8] on Urban100 [2]. The visual comparison between our models and other compared methods on Urban100 [2] with SR $\times 4$ is shown in Fig. 2. In some challenging scenarios, the previous methods may suffer blurring artifacts, distortions, or inaccurate texture restoration. In contrary, the proposed **EARFA-light** can effectively mitigate these artifacts and preserve more structures and finer details. For example, in testing image "img_078", the reconstructed images of the previous methods are mostly have some problems such as blur, distortion, and poor detail recovery. And our proposed **EARFA-light** can restore the correct structure. We also observed this phenomenon in other images. This is primary because **EA** and **SLKA** provide our models with more informative inference and enhanced effective receptive fields.

REFERENCES

- [1] Chao Dong, Chen Change Loy, and Xiaoou Tang. 2016. Accelerating the super-resolution convolutional neural network. In *Proceedings of the European conference on computer vision*. Springer, 391–407.
- [2] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. 2015. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 5197–5206.
- [3] Zheng Hui, Xinbo Gao, Yunchu Yang, and Xiumei Wang. 2019. Lightweight image super-resolution with information multi-distillation network. In *Proceedings of the 27th acm international conference on multimedia*. 2024–2032.
- [4] Robert Keys. 1981. Cubic convolution interpolation for digital image processing. *IEEE Transactions On Acoustics, Speech, and Signal Processing* 29, 6 (1981), 1153–1160.
- [5] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. 2021. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*. 1833–1844.
- [6] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. 2017. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 136–144.
- [7] Wenjie Luo, Yujia Li, Raquel Urtasun, and Richard Zemel. 2016. Understanding the effective receptive field in deep convolutional neural networks. *Advances in neural information processing systems* 29 (2016).
- [8] Long Sun, Jiangxin Dong, Jinhui Tang, and Jinshan Pan. 2023. Spatially-adaptive feature modulation for efficient image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 13190–13199.
- [9] Long Sun, Jinshan Pan, and Jinhui Tang. 2022. Shufflemixer: An efficient convnet for image super-resolution. *Advances in Neural Information Processing Systems* 35 (2022), 17314–17326.
- [10] Hengyuan Zhao, Xiangtao Kong, Jingwen He, Yu Qiao, and Chao Dong. 2020. Efficient image super-resolution using pixel attention. In *Proceedings of the European Conference on Computer Vision Workshops*. Springer, 56–72.
- [11] Yupeng Zhou, Zhen Li, Chun-Le Guo, Song Bai, Ming-Ming Cheng, and Qibin Hou. 2023. Sformer: Permuted self-attention for single image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 12780–12791.

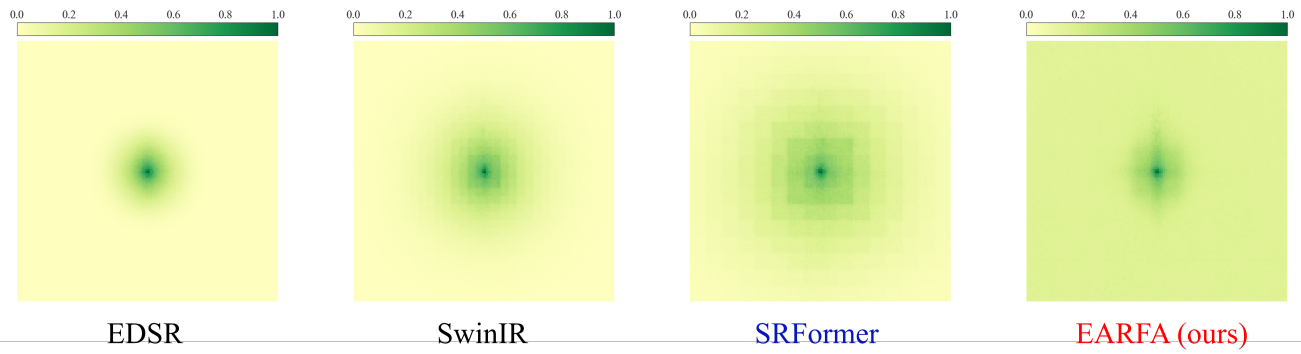


Figure 1: The Effective Receptive Field (ERF) of EDSR [6] and SwinIR [5], SRFormer [11], EARFA (ours). The best and second best results are marked in **red** and **blue**, respectively.

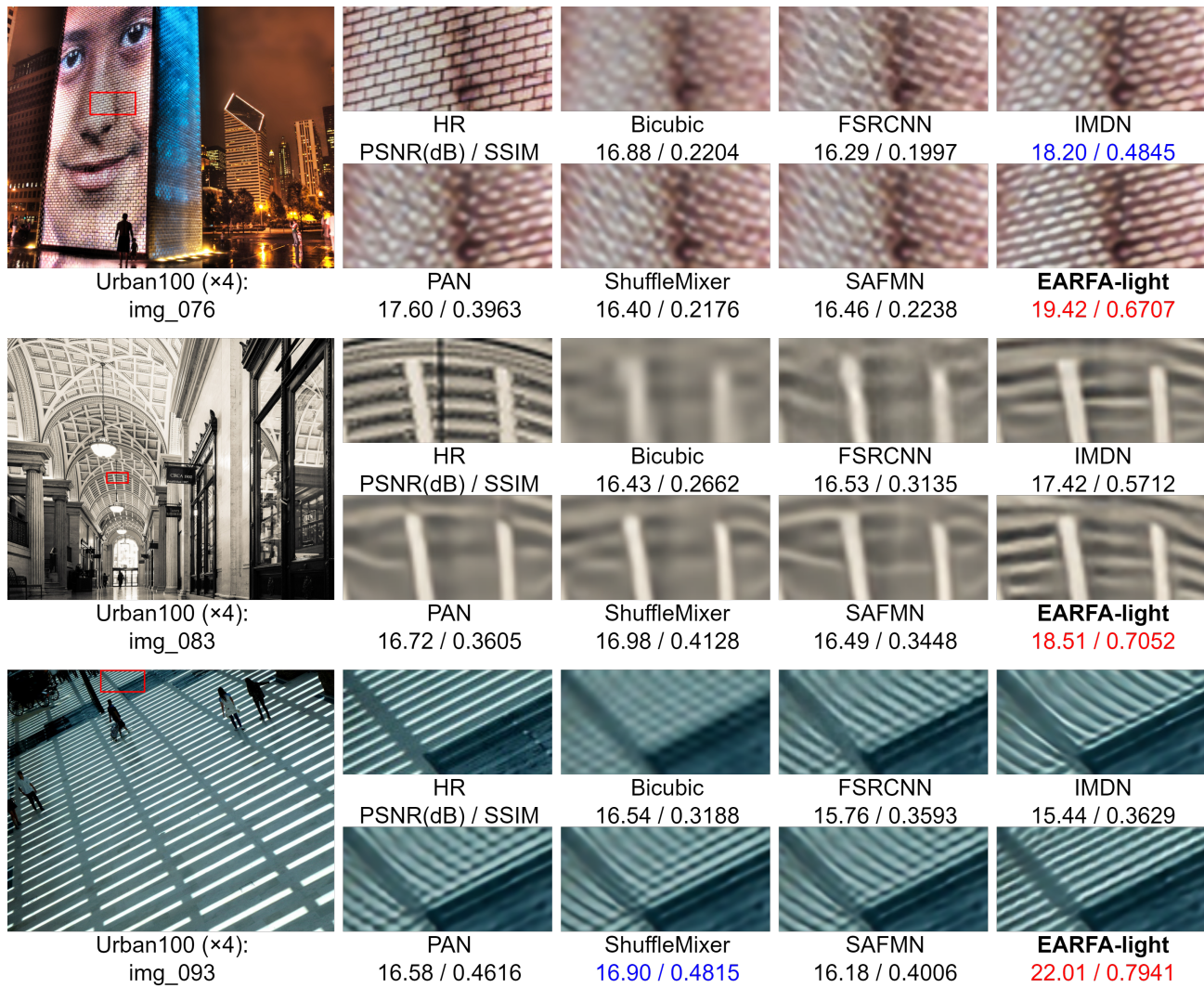


Figure 2: Visual comparison between the proposed models and other advanced SISR methods on Urban100 [2] with SR×4. The best and second best results are marked in **red** and **blue**, respectively.

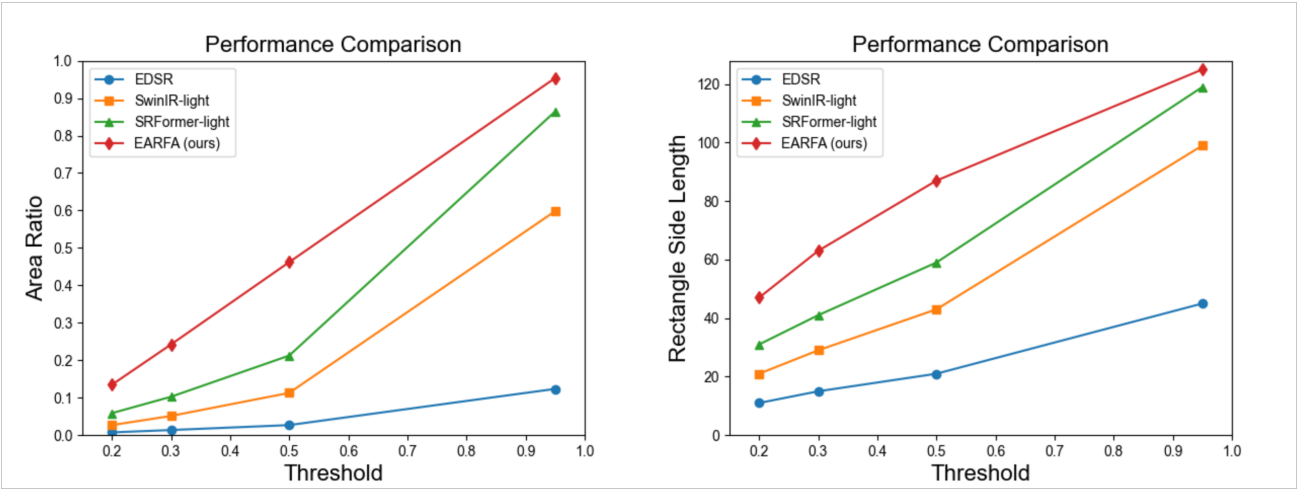


Figure 3: Comparative graph of the performance of different methods. The vertical axis on the left side of the graph indicates the area of the central window required to include the specified proportion of weights. The vertical axis on the right side of the graph indicates the side length of the central window required to include the specified proportion of weights.