Boosting Single Image Super-Resolution via Partial Channel Shifting

Xiaoming Zhang, Tianrui Li, Xiaole Zhao*
Southwest Jiaotong University, China

1. Supplementary Material

In the supplementary material, we first determine the theoretical receptive fields that arise from applying Partial Channel Shifting (PCS).

Then we apply Local Attribution Maps (LAMs) [6] to investigate different shifting settings of PCS.

To further prove the generality of PCS, besides the representative SR models that we analyze in the manuscripts, *i.e.*, DRRN [18], EDSR(S) [14] and EDSR(L) [14], we also provide convergence comparisons for multiple state-of-the-art convolutional SR models when enhanced by PCS, ranging from lightweight to large-scale, including RFDN [15], BSRN [13], VAPSR [25], OISR [7], SAN [4] and NLSN [17]. We also visualize their Effective Receptive Fields [16] (ERFs) and LAMs.

Finally, additional qualitative visual comparisons demonstrate that PCS improves the reconstruction performance of SR models. We also provide pseudo-code to illustrate how to perform spatial-wise feature shifting.

2. Receptive Field with PCS

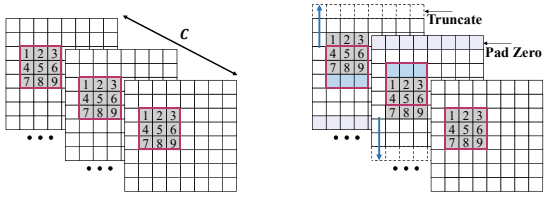


Figure 1: The receptive field is plausibly extended through shifting a part of channels. **Left**: The spatially aligned features provide a regular 3×3 receptive field for 3×3 conv kernels; **Right**: The conv kernels "see" a larger range via staggered feature maps.

In this section, we calculate the theoretical receptive field with the enlargement of PCS, and the calculation of the receptive field follows the guidance of Arujo *et al*. [2].

A 2D convolution with a kernel size of 3×3 and stride of 1 results in an initial receptive field of 3×3. By performing this convolutional operation N times, the theoretical receptive field on the original feature expands to $(2N + 1) \times (2N + 1)$.

*Corresponding author, email: zxlation@foxmail.com
Code is available at https://github.com/OwXiaoM/PCS

When considering a case of bi-directional shift as shown in Fig. 1, when applying a bi-directional shift with $|h| = 1$ and $|w| = 0$, the receptive field increases to 3×5. Upon applying this convolutional operation N times, the theoretical receptive field on the original feature expands to $(2N + 1) \times (4N + 1)$.

Generally, for a 2D convolution with a kernel size of $K \times K$ and stride of 1, when applying shifting operation on a single spatial direction with $|h| = P, |w| = 0$ or $|h| = 0, |w| = P (P \leq K)$ and repeating N times the convolution operation, the resulting receptive field will be $(N * K - N + 1) \times (N * K + N * P - N + 1)$.

Furthermore, Table 1 displays the theoretical receptive field sizes of different shift modes.

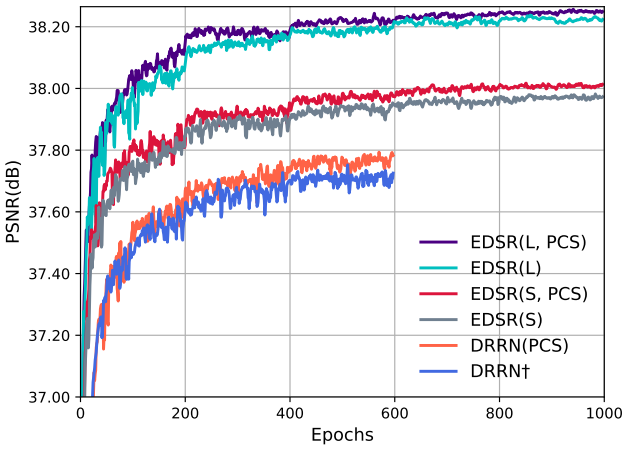


Figure 2: The training curves of representative models when applying PCS, and the results are tested on Set5 [3] (SR×2).

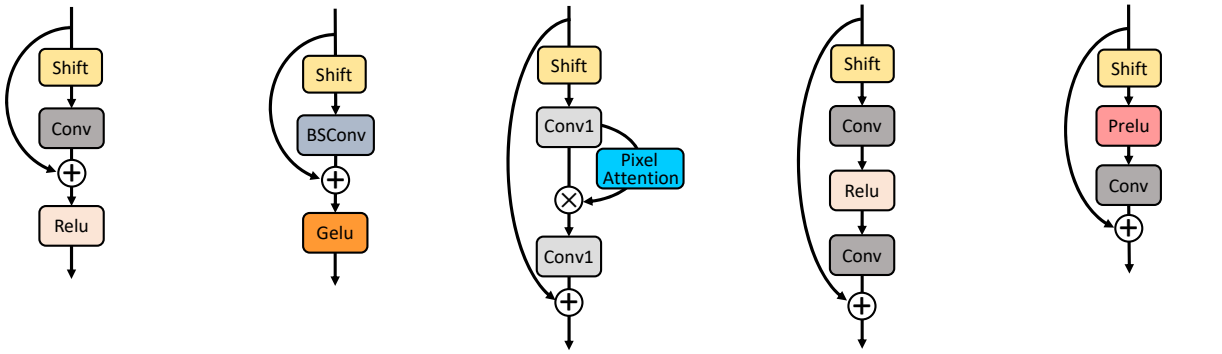
3. Local Attribution Maps

We apply Local Attribution Maps (LAMs) [6] to investigate the impact of different settings of PCS on the SR models.

As shown in Fig. 4, LAMs trace the involved range of utilized information when EDSR(S) and EDSR(S, PCS) reconstruct the red target area on the butterfly of Set5 [3]. The LAM attribution indicates the activated pixels during the process of reconstruction. The area of contribution is the visualized projection from the LAM attribution to the

Table 1: The theoretical receptive field is calculated based on the convolution with the kernel size of $K \times K$ and stride of 1, repeating convolution operation N times and shifting magnitude of $P(P \leq K)$.

Shift Mode	Receptive Field
Non-shift	$(N*K-N+1)\times(N*K-N+1)$
Uni-directional	$(N*K-N+1+N*P)\times(N*K-N+1)$
Bi-directional	$(N*K-N+1+2N*P)\times(N*K-N+1)$
Cross-directional	$(N*K-N+1+N*P)\times(N*K-N+1+N*P)$
Quad-directional	$(N*K-N+1+2N*P)\times(N*K-N+1+2N*P)$
Octa-directional	$(N*K-N+1+2N*P)\times(N*K-N+1+2N*P)$



(a) PCS SRB (b) PCS BSRB (c) PCS VAB (d) PCS ResBlock (e) PCS ResBlock-v2

Figure 3: The combinations of PCS with different residual components. (a) PCS ResBlock is with EDSR [14], SAN [4], and NLSN [17]. (b) PCS SRB is with RFDN [15]. (c) PCS BSRB is with BSRN [13]. (d) PCS Pixel Attention is with VAPSR [25]. (e) PCS ResBlock-v2 is with OISR-LF [7] and OISR-RK3 [7].

target image, and the shade of the red color simultaneously illustrates the proportion of the contribution on the spatial information.

Fig. 4(a) illustrates how the origin EDSR(S) activates the neighbor region in the target area. To further examine the activation process, we explore EDSR(S, PCS) with different shifting configurations, as shown in Fig. 4(b), Fig. 4(c), and Fig. 4(d).

Therefore, it can be observed that: (1) The direction in which the features shift has a direct impact on the activated spatial information; (2) Greater shifted proportions and larger shifting magnitudes promote activation of a wider range of pixels. However, as we explain in Sec. 4.3, excessively large shifting magnitudes and shifted proportion is probably detrimental to the spatial locality of the image, leading to the degradation of performance on SR models.

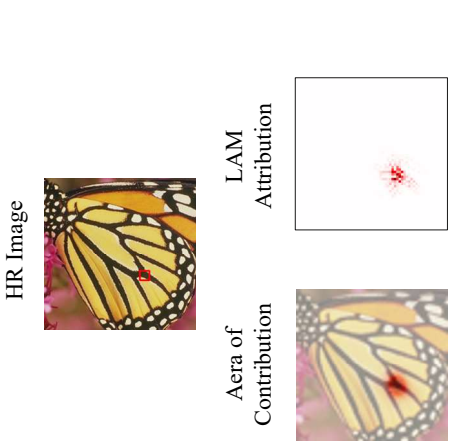
4. Convergence comparisons

In order to verify the general applicability of PCS on SISR models, we conduct additional experiments using highly effective and representative models that are widely used and extensively evaluated in the SR community.

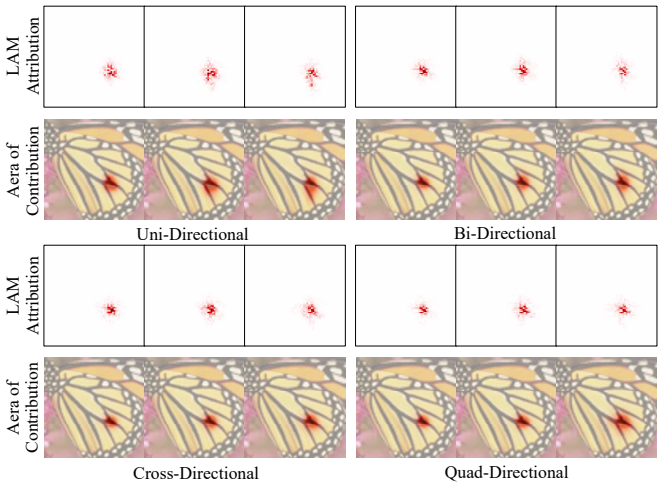
Besides the representative models (DRRN [18], EDSR [14]) that we analyze in the manuscript, we also investigated the impact of PCS on recently developed lightweight and large-scale state-of-the-art models, including RFDN [15], BSRN [13], VAPSR [25], OISR [7], SAN [4] and NLSN [17].

We integrate quad-directional shift with $|h_k| = |w_k| = 2, k \in \mathcal{I}, \gamma = 1/8$ for all lightweight and middle-scale models (except $|h_k| = |w_k| = 1$ for VAPSR [25]), and bi-directional shift with $|h_k| = 2, |w_k| = 0, k \in \mathcal{I}, \gamma = 1/16$ for all large-scale models. When designing the combination of PCS of each model, we follow our conclusion in Sec. 4.3.4 that always embed PCS towards the front of the Residual Connection, and the detailed implementations are in Fig. 3. For a fair comparison, all models are trained from scratch on DIV2K [1], The learning rate is set according to their original settings [18, 15, 13, 25, 7, 4, 17], and other training settings are consistent with Sec. 4.2 in the manuscript. We show the training curves of lightweight and middle-scale models of SR($\times$2,$\times$4), and the large-scale models of SR($\times$2).

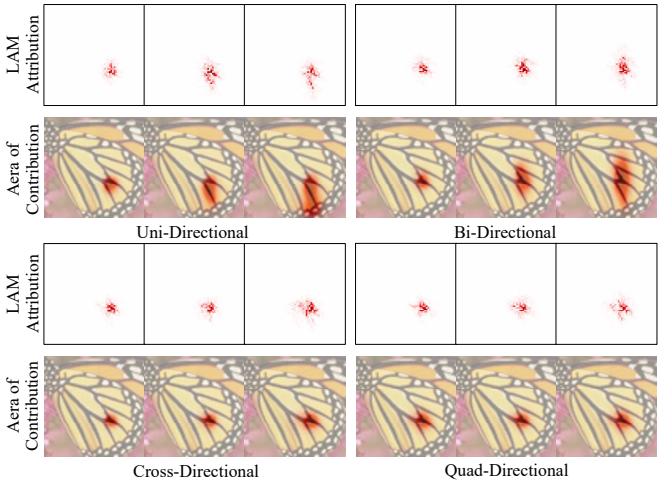
As the subsequent experiments show that all the PCS-enhanced models not only converge faster, but also have no difficulty in obtaining higher performance. These results empirically verify the effectiveness that PCS endows



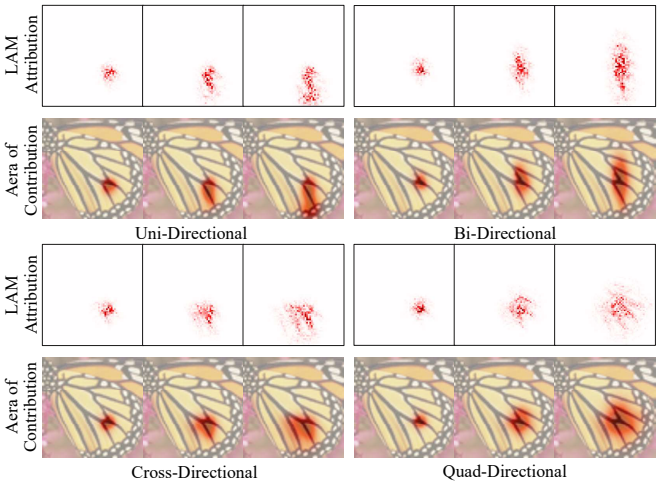
(a) LAM of EDSR(S)



(b) EDSR(S, PCS), $\gamma = 1/16$.



(c) EDSR(S, PCS), $\gamma = 1/8$.



(d) EDSR(S, PCS), $\gamma = 1/2$.

Figure 4: LAM results of different shifting settings. From left to right, uni-directional shift is with $h = -1, -2, -3$ and $|w| = 0$, bi-directional shift is with $h = 1, 2, 3$ and $|w| = 0$, cross-directional shift is with $h = -1, -2, -3$ and $w = -1, -2, -3$, quad-directional shift is with $|h| = 1, 2, 3$ and $|w| = 1, 2, 3$.

the models. PCS also showcases its outstanding plug-and-play capability and versatility when interacting with various state-of-the-art SR models.

4.1. Representative SISR Models

Fig. 2 illustrates the comparison of training curves between DRRN [18] and EDSR [14], as well as their PCS-enhanced counterparts. While the PCS-enhanced models exhibit better performance, it can be observed that as the model size increases, the extent of performance improvement provided by PCS diminishes. This phenomenon persists in subsequent experiments as well.

4.2. Lightweight SISR Models

For lightweight models, RFDN [15] is proposed with multiple feature distillation connections, which learns more characteristic feature representations, and won the champion of AIM 2020 Efficient SR Challenge [22]. RFDN contains 6 residual feature distillation blocks (RFDB), and in each RFDB, the feature is sequentially extracted into two branches composed of 1×1 convolutions and shallow residual blocks (SRB). We insert the PCS operation in the front of the convolution of SRB as shown in Fig. 3a. The convergence comparisons are in Fig. 5 and Fig. 8.

BSRN [13] adapted the residual feature distillation connections with effective attention modules and re-

parameterized the redundant convolution by blueprint separable convolution (BSConv) of RFDN, and won the first place in the model complexity track in NTIRE 2022 Efficient SR Challenge [12]. As BSRN inherits the feature distillation topology of RFDN, the blueprint shallow residual block (BSRB) is introduced by replacing the standard 3×3 convolution of the RSB with BSConv. So we also insert the PCS operation in the front of BSConv as shown in Fig. 3b. The convergence comparisons are in Fig. 6 and Fig. 9.

Recently, VAPSR [25] is well-designed with pixel-attention [24], and re-parameterizes further the convolutions to a vast receptive field, which can achieve a similar performance of RFDN with only 28.18% parameters. Instead of the topology of feature distillation connections, VAPSR chose the residual architecture similar to VDSR [9], and 21 vast-receptive-field attention blocks (VABs) are cascaded together. Both ends of VAB are stacked with 1×1 convolutions, and its backbone is enhanced with pixel attention. We insert the PCS operation in front of the first convolution of VAB as shown in Fig. 3c. Considering the size of the foremost convolution in VAB is smaller compared to the 3×3 convolution, we opt for a smaller shifting magnitude as $|h_k| = |w_k| = 1$. The convergence comparisons are in Fig. 7 and Fig. 10.

4.3. Large-scale SISR Models

For large-scale models, He *et al*. are inspired by numerical analysis methods for ordinary differential equations to develop OISR [7], which subtly modifies the topological structure of the residual block in EDSR. The proposed modifications result in improved performance for EDSR, with a comparable number of parameters. OISR-LF is the medium-scale version, whose architecture is interpreted from the Leapfrog method, while OISR-RK3 is a large-scale version corresponding to the 3-stage Runge-Kutta method. In OISR, a residual block is formed by multiple residual blocks, we replace only the first residual block with PCS ResBlock-v2 as shown in Fig. 3e. The convergence comparisons of OISR-LF-based models are in Fig. 11 and Fig. 12, and the convergence comparisons of OISR-RK3-based models are in Fig. 13.

Dai *et al*. learned from channel-attention [8, 23], non-local [20] and matrix power normalized covariance [11], and proposed SAN [4], which combines the second-order channel attention and non-locally enhanced residual group (NLRG) to enhance the performance of SISR. There are 10 residual blocks in each NLRG, and they are identical to the residual blocks in EDSR [14]. We substitute all the residual blocks in SAN with the PCS ResBlock to form SAN(PCS), as shown in Fig. 3d. The convergence comparisons are in Fig. 14.

Mei *et al*. proposed NLSN [17], a state-of-the-art convolutional SISR model, by combining non-local [20] and sparse representation [21] with EDSR(L). Specifically, they improved the embedded Gaussian form of non-local operation by drawing inspiration from Reformer [10], and inserted the non-local sparse attention block in EDSR(L) after every 8 residual blocks, leading to the development of NLSN as one of the representative large-scale models at present. As NLSN shares the same network topology with EDSR, we also replace all the residual blocks with PCS ResBlock. The convergence comparisons are in Fig. 15.

5. Visualization of ERFs

The ERFs [16] calculate the pixels that contribute to the center of an output feature map in a neural network, thereby assisting us in comprehending the effective receptive field scale of each model. For SISR models, we randomly crop 200 patches of 64×64 resolution from DIV2K [1] SR($\times$2) as the input. Following [5, 19], we then calculated the score for the center of the output feature map of the layer before the up-scaling module (*e.g.* pixel-shuffle layer), and rescale the accumulated average output within a value between 0 and 1 for the whole to effective receptive fields. As shown in Fig. 16, the brighter and more influential a region is, the closer the value is to 1, while the darker the region, the closer the value is to 0. The size of ERFs is comparatively smaller in lightweight models, whereas it is larger in larger-scale models. Although the interactions between PCS and different attention are relatively complex and irregular, PCS is capable of expanding the receptive field of all the state-of-the-art SISR models.

6. Visualization of LAMs

We then present Local Attribution Map (LAM) analysis for both the non-PCS methods and PCS-enhanced methods in Fig. 17. The target area under consideration is identical to that in Fig. 4a. In lightweight models, the range of activated neighboring pixels is narrow due to the relatively small ERFs, while in large-scale models, PCS-enhanced methods exhibit a higher density of involved pixels, implying that they are likely to explore more local spatial correlation during reconstruction. Despite the improved representation capacity of PCS-enhanced models, as demonstrated by the training curves, the combinations of PCS and different attention mechanisms remain too complex and intricate to analyze.

7. More Qualitative Results

This section provides more qualitative results in Fig. 18 and Fig. 19. Compared with the original models, PCS-enhanced SR models significantly improve their representation abilities. Therefore, the integration of PCS into existing SR models is a promising approach to achieving higher-quality image reconstruction.

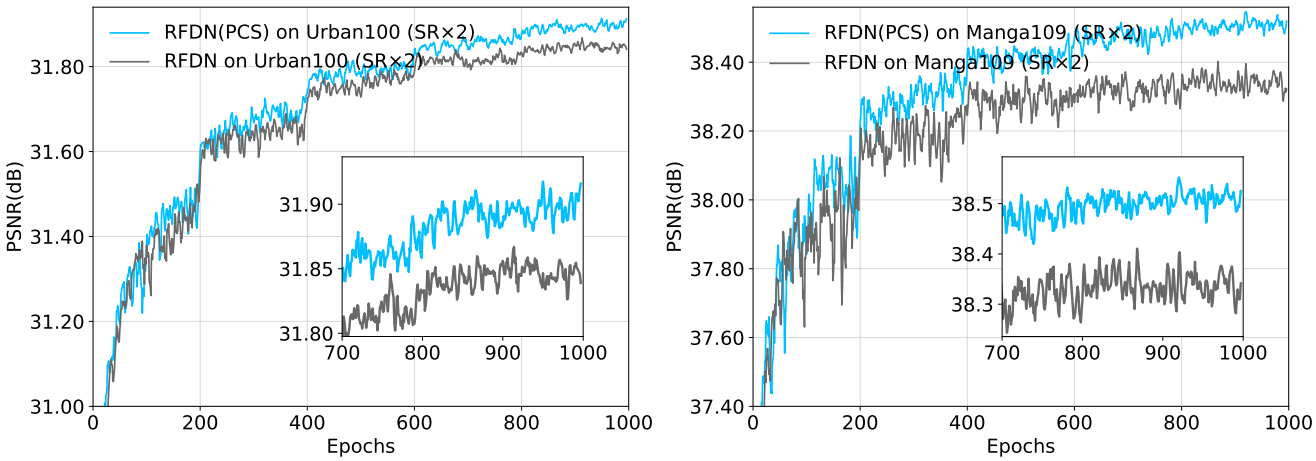


Figure 5: Convergence comparison between RFDN [15] and RFDN(PCS) on SR $\times 2$.

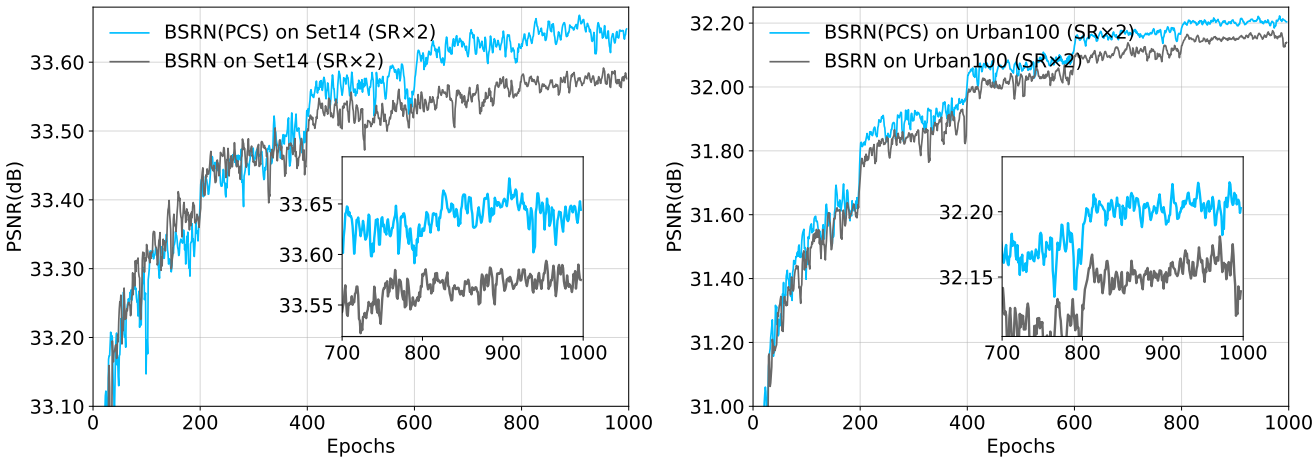


Figure 6: Convergence comparison between BSRN [13] and BSRN(PCS) on SR $\times 2$.

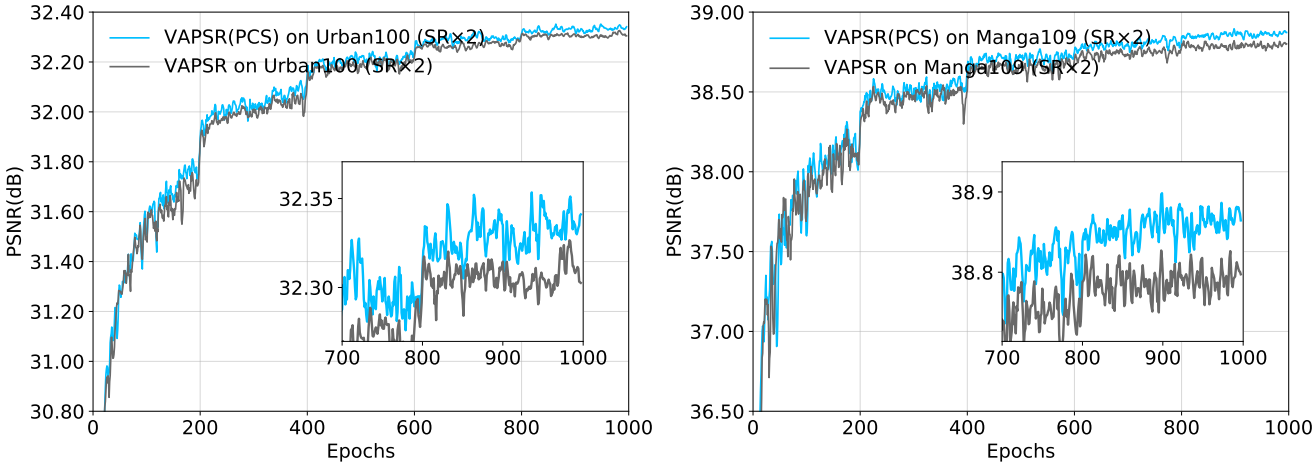


Figure 7: Convergence comparison between VAPSR [25] and VAPSR(PCS) on SR $\times 2$.

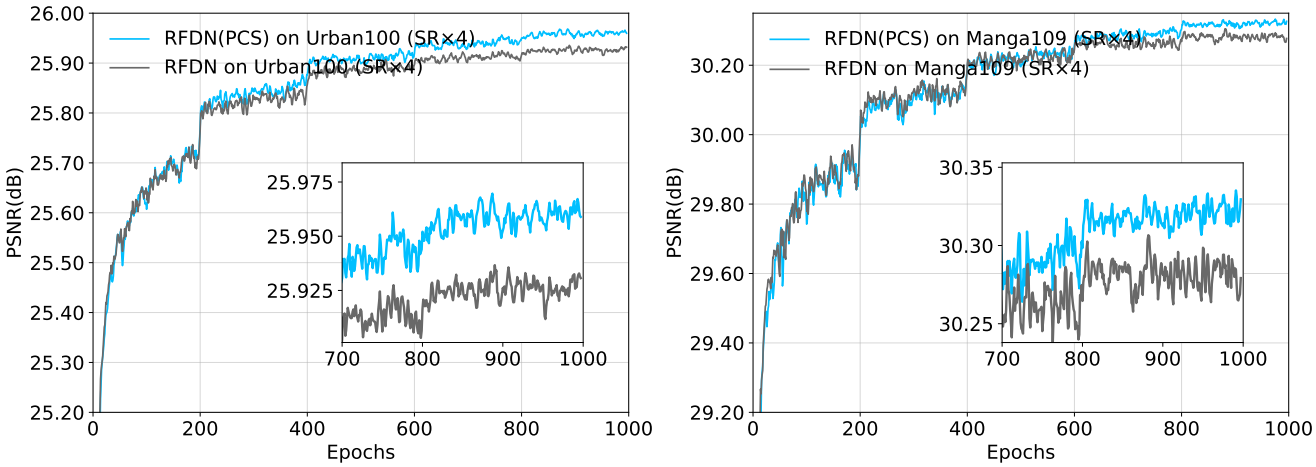


Figure 8: Convergence comparison between RFDN [13] and RFDN(PCS) on SR $\times 4$.

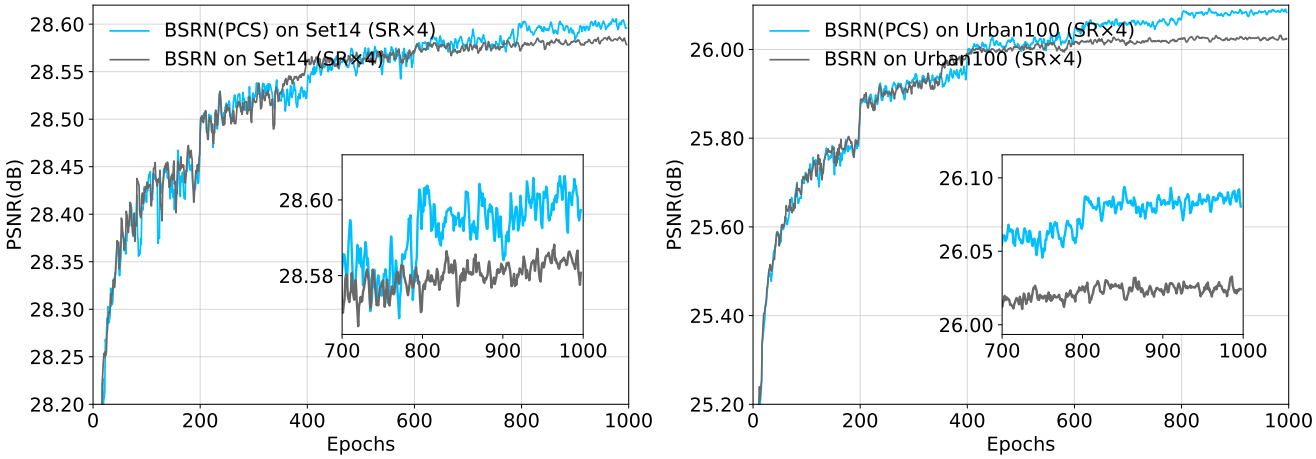


Figure 9: Convergence comparison between BSRN [13] and BSRN(PCS) on SR $\times 4$.

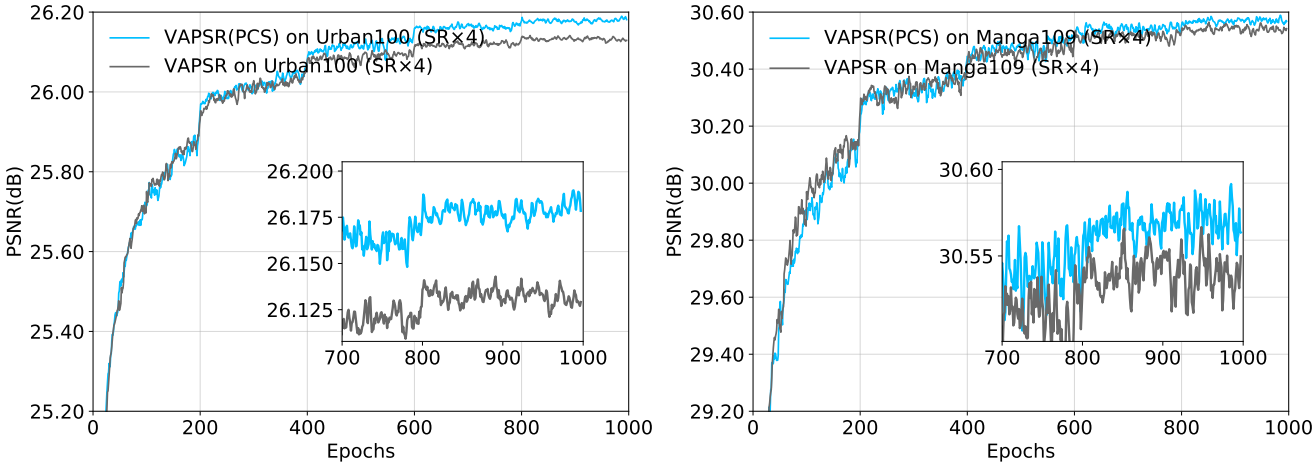


Figure 10: Convergence comparison between VAPSR [25] and VAPSR(PCS) on SR $\times 4$.

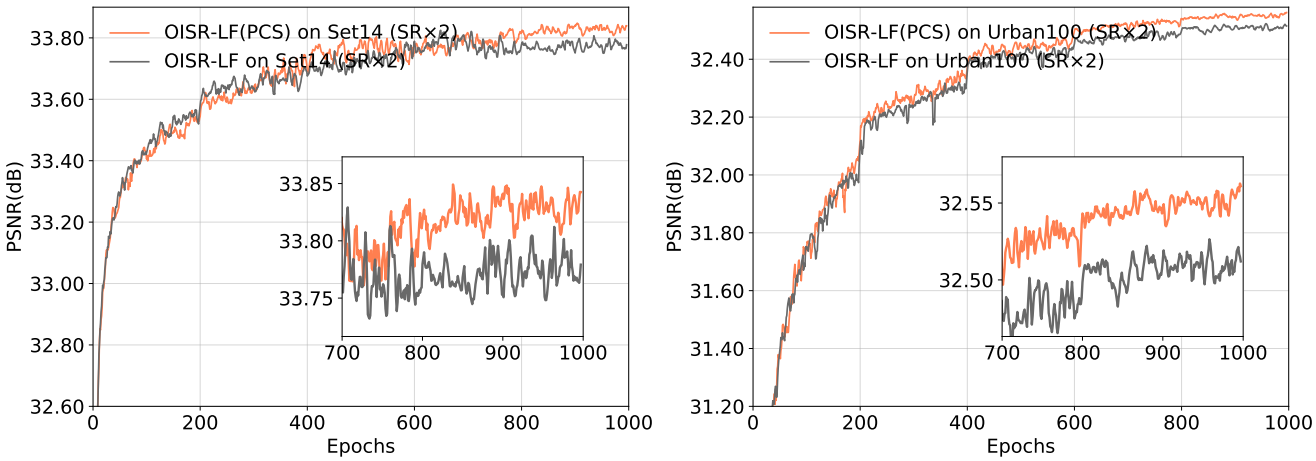


Figure 11: Convergence comparison between OISR-LF [7] and OISR-LF(PCS) on SR ×2..

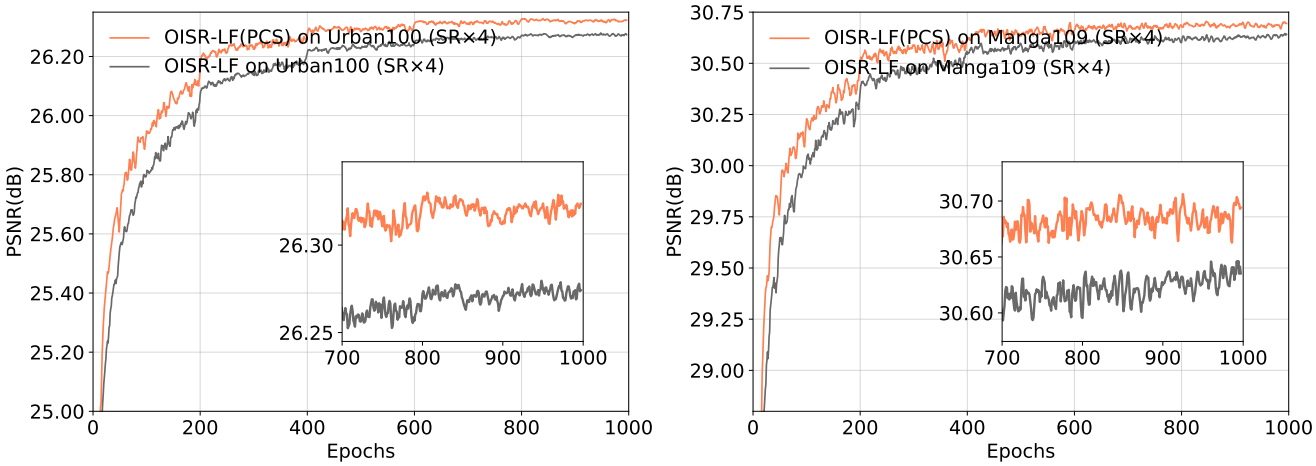


Figure 12: Convergence comparison between OISR-LF [7] and OISR-LF(PCS) on SR ×4..

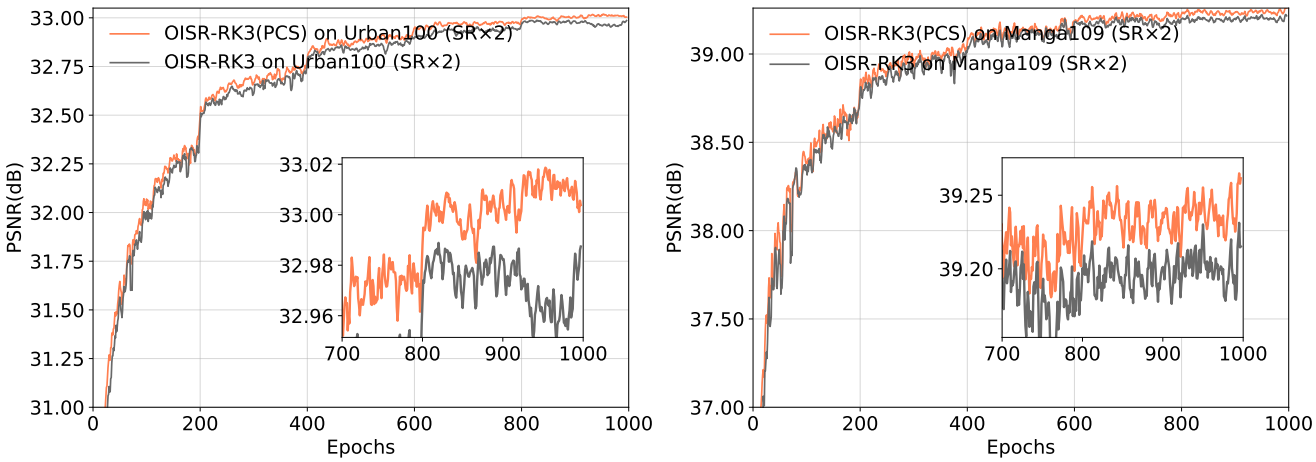


Figure 13: Convergence comparison between OISR-RK3 [7] and OISR-RK3(PCS) on SR ×2.

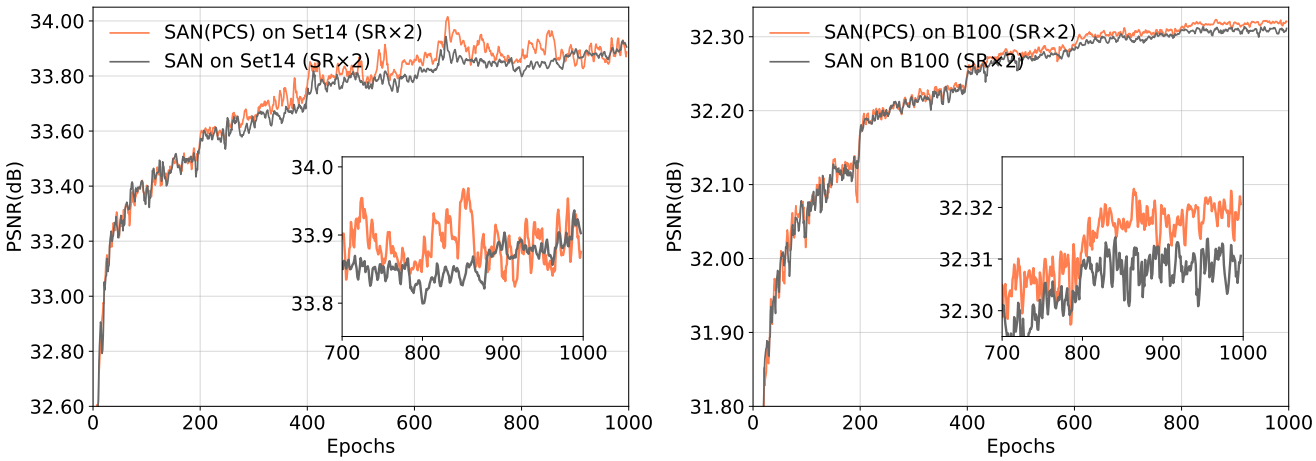


Figure 14: Convergence comparison between SAN [4] and SAN(PCS) on SR $\times 2$.

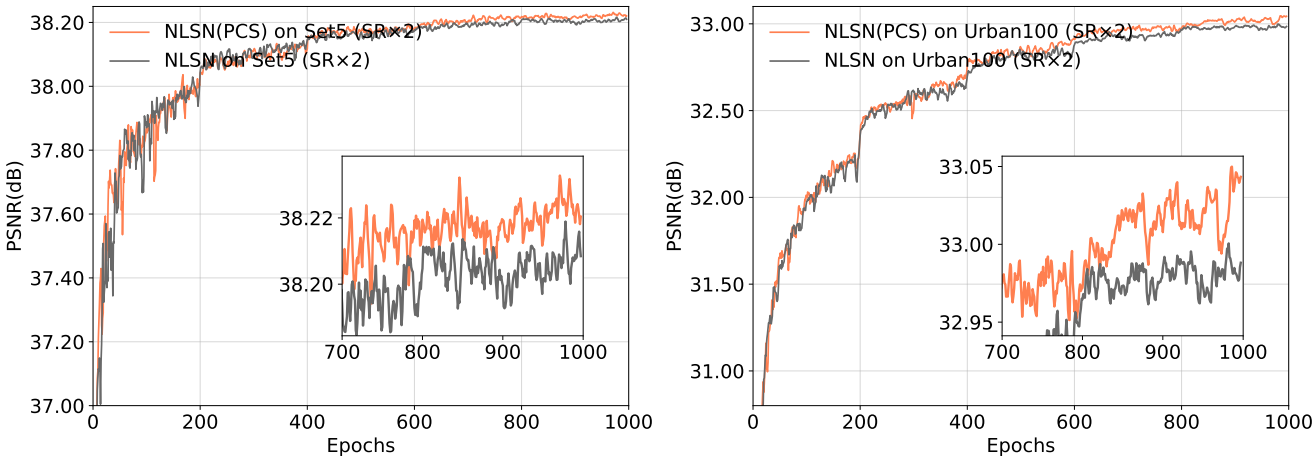


Figure 15: Convergence comparison between NLSN [17] and NLSN(PCS) on SR $\times 2$.

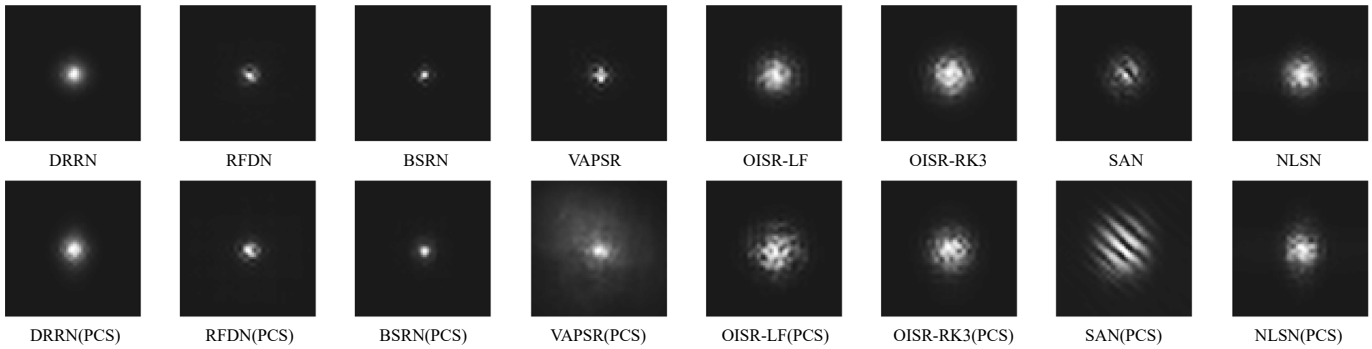


Figure 16: ERFs [16] of non-PCS models and PCS-enhanced models.

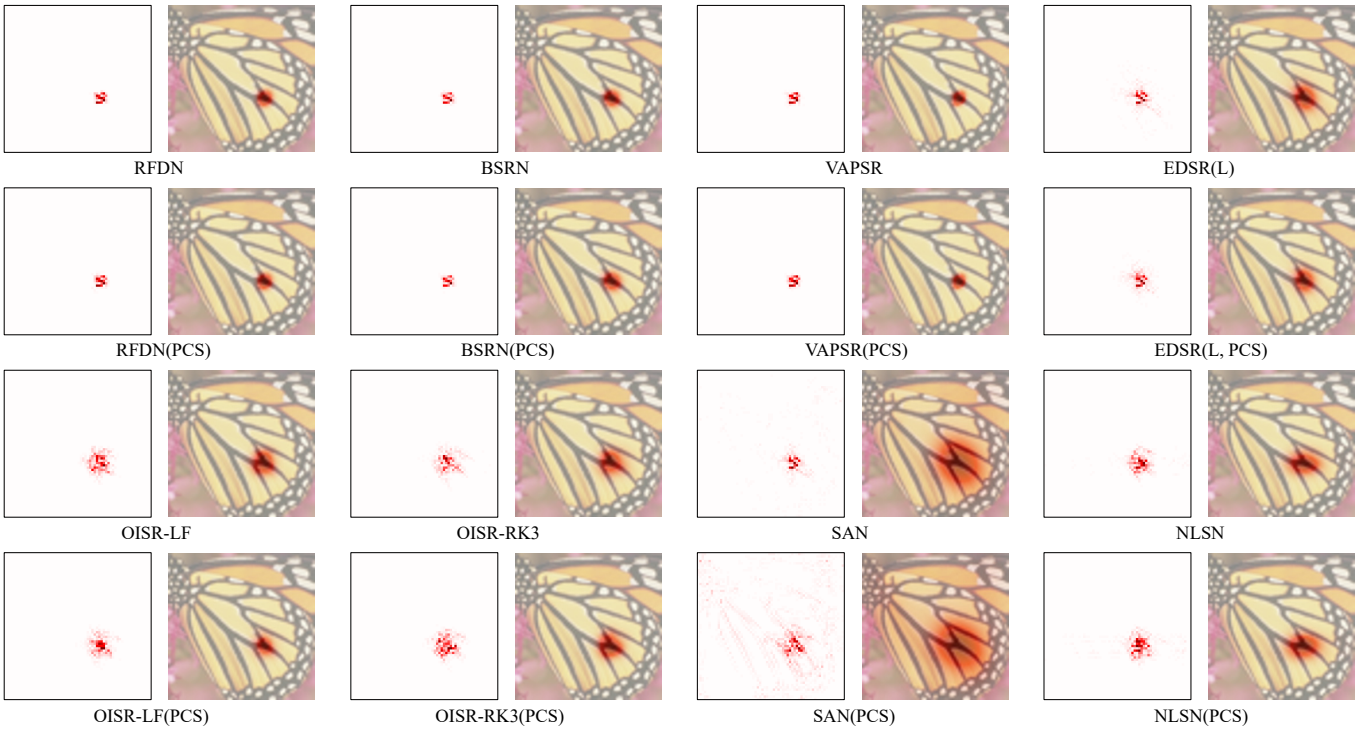


Figure 17: LAM attributions (left) and area of contributions (right) of non-PCS models and PCS-enhanced models.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *CVPRW*, pages 126–135, 2017. 2, 4

[2] André Araujo, Wade Norris, and Jack Sim. Computing receptive fields of convolutional neural networks. *Distill*, 2019. https://distill.pub/2019/computing-receptive-fields. 1

[3] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In *BMVC*, pages 1–10, 2012. 1

[4] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *CVPR*, pages 11065–11074, 2019. 1, 2, 4, 8

[5] Xiaohan Ding, Xiangyu Zhang, Jungong Han, and Guiguang Ding. Scaling up your kernels to 31x31: Revisiting large kernel design in cnns. In *CVPR*, pages 11963–11975, 2022. 4

[6] Jinjin Gu and Chao Dong. Interpreting super-resolution networks with local attribution maps. In *CVPR*, pages 9199–9208, 2021. 1

[7] Xiangyu He, Zitao Mo, Peisong Wang, Yang Liu, Mingyuan Yang, and Jian Cheng. Ode-inspired network design for single image super-resolution. In *CVPR*, pages 1732–1741, 2019. 1, 2, 4, 7

[8] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *CVPR*, pages 7132–7141, 2018. 4

[9] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *CVPR*, pages 1646–1654, 2016. 4

[10] Nikita Kitaev, Lukasz Kaiser, and Anselm Levskaya. Reformer: The efficient transformer. In *ICLR*, 2019. 4

[11] Peihua Li, Jiangtao Xie, Qilong Wang, and Wangmeng Zuo. Is second-order information helpful for large-scale visual recognition? In *ICCV*, pages 2070–2078, 2017. 4

[12] Yawei Li, Kai Zhang, Radu Timofte, Luc Van Gool, Fangyuan Kong, Mingxi Li, Songwei Liu, Zongcai Du, Ding Liu, Chenhui Zhou, et al. Ntire 2022 challenge on efficient super-resolution: Methods and results. In *CVPRW*, pages 1062–1102, 2022. 4

[13] Zheyuan Li, Yingqi Liu, Xiangyu Chen, Haoming Cai, Jinjin Gu, Yu Qiao, and Chao Dong. Blueprint separable residual network for efficient image super-resolution. In *CVPRW*, pages 833–843, 2022. 1, 2, 3, 5, 6

[14] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *CVPRW*, pages 136–144, 2017. 1, 2, 3, 4

[15] Jie Liu, Jie Tang, and Gangshan Wu. Residual feature distillation network for lightweight image super-resolution. In *ECCVW*, pages 41–55, 2020. 1, 2, 3, 5

[16] Wenjie Luo, Yujia Li, Raquel Urtasun, and Richard Zemel. Understanding the effective receptive field in deep convolutional neural networks. *NeurIPS*, 29, 2016. 1, 4, 8

[17] Yiqun Mei, Yuchen Fan, and Yuqian Zhou. Image super-resolution with non-local sparse attention. In *CVPR*, pages 3517–3526, 2021. 1, 2, 4, 8

[18] Ying Tai, Jian Yang, and Xiaoming Liu. Image super-resolution via deep recursive residual network. In *CVPR*, pages 3147–3155, 2017. 1, 2, 3

[19] Yuki Tatsunami and Masato Taki. Sequencer: Deep lstm for image classification. *NeurIPS*, 2022. 4

[20] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *CVPR*, pages 7794–7803, 2018. 4

[21] Jianchao Yang, John Wright, Thomas S. Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE TIP*, 19(11):2861–2873, 2010. 4

[22] Kai Zhang, Martin Danelljan, Yawei Li, Radu Timofte, Jie Liu, Jie Tang, Gangshan Wu, Yu Zhu, Xiangyu He, Wenjie Xu, et al. Aim 2020 challenge on efficient super-resolution: Methods and results. In *ECCVW*, pages 5–40, 2020. 3

[23] 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*, pages 286–301, 2018. 4

[24] Hengyuan Zhao, Xiangtao Kong, Jingwen He, Yu Qiao, and Chao Dong. Efficient image super-resolution using pixel attention. In *ECCVW*, pages 56–72, 2020. 4

[25] Lin Zhou, Haoming Cai, Jinjin Gu, Zheyuan Li, Yingqi Liu, Xiangyu Chen, Yu Qiao, and Chao Dong. Efficient image super-resolution using vast-receptive-field attention. In *ECCVW*, pages 256–272, 2023. 1, 2, 4, 5, 6

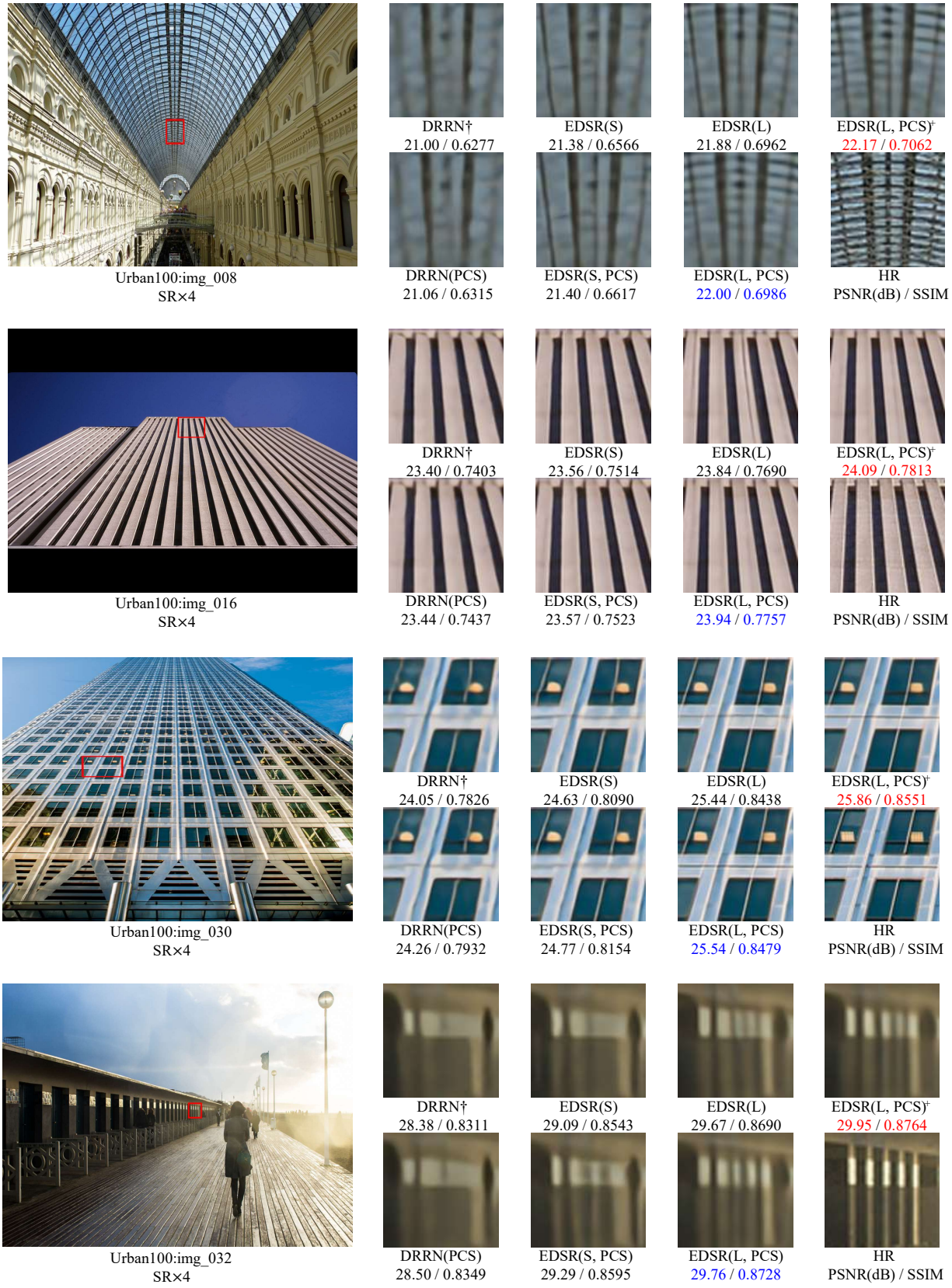


Figure 18: Qualitative comparisons between SR methods. The best and second best results are in red and blue respectively.

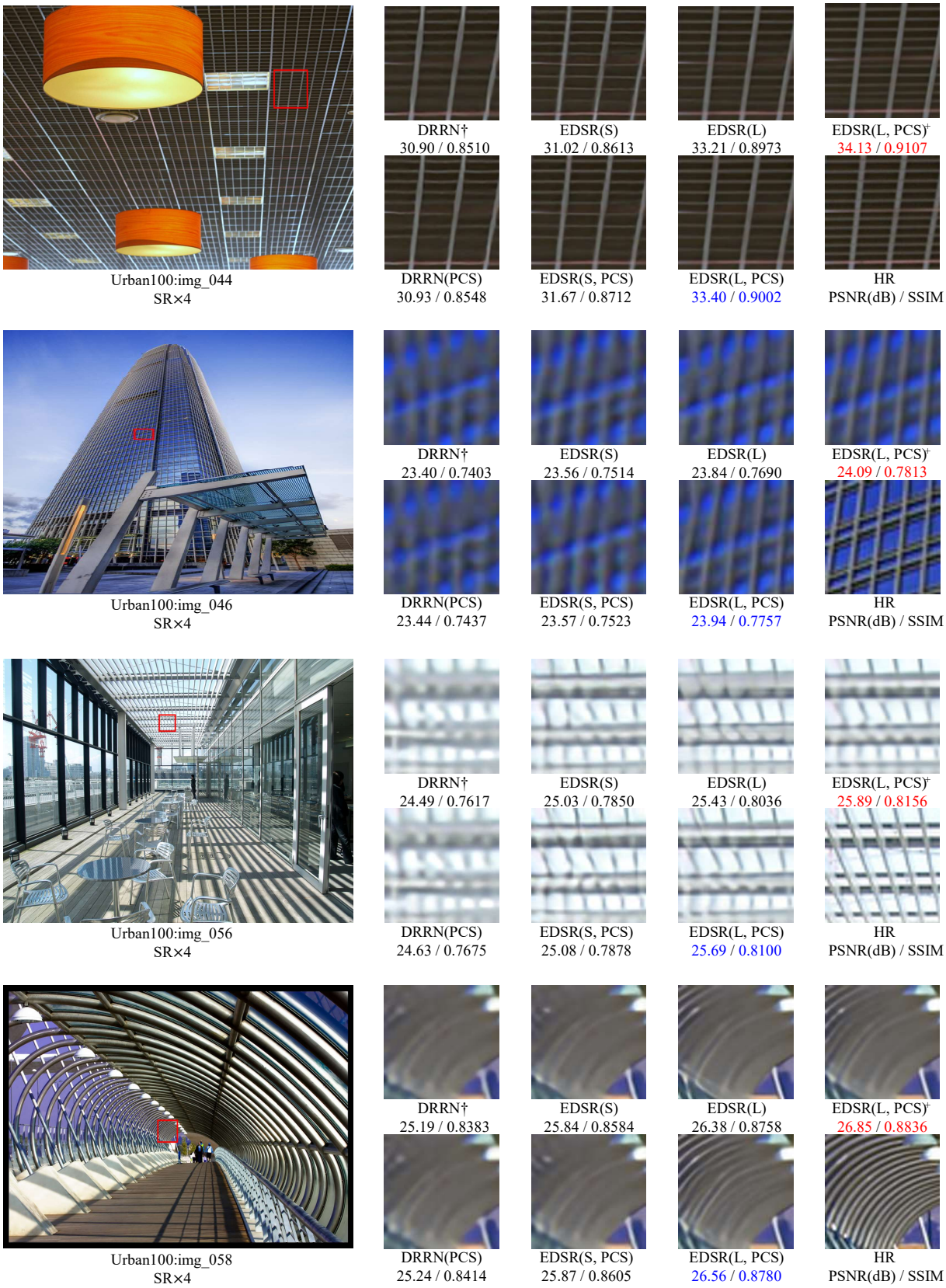


Figure 19: Qualitative comparisons between SR methods. The best and second best results are in red and blue respectively.