NTIRE 2025 Efficient SR Challenge Factsheet
-PAEDN-

Simone Bianco
Department of Informatics Systems and Communication
University of Milano-Bicocca
Viale Sarca 336, Building U14, Milan, Italy
simone.bianco@unimib.it

Luca Cogo
Department of Informatics Systems and Communication
University of Milano-Bicocca
Viale Sarca 336, Building U14, Milan, Italy
luca.cogo@unimib.it

Gianmarco Corti
Department of Informatics Systems and Communication
University of Milano-Bicocca
Viale Sarca 336, Building U14, Milan, Italy
g.corti1967@campus.unimib.it

1. Team details

- Team name: IVL
- Team Leader's name: Simone Bianco
- Team Leader's email: simone.bianco@unimib.it
- Team Leader's address: Milan, Viale Sarca 336, Building U14, Room 1011,
- Team Leader's phone number: +390264487827
- Rest of the team members: Gianmarco Corti, Luca Cogo
- Team's website www.ivl.disco.unimib.it
- Affiliation of the team: DISCo (Department of Informatics, Systems and Communication), University of Milano-Bicocca
- User names and entries on the NTIRE 2025 Codalab competitions (development/validation and testing phases): gianmarco.corti
- Best scoring: 26.66 (valid), 26.76 (test)
- Link to the codes/executables of the solution.

2. Method details

Method Our approach builds upon the strategy used in SPAN [6], last year's winning method, to extract attention maps and integrates it into the proposed baseline architecture, EFDN [7], aiming to enhance feature extraction and structural representation in image processing tasks.

Specifically, as illustrated in Figure 1, this strategy is incorporated within the EDBB blocks of EFDN, which are designed to capture fundamental structural features of an image by applying Sobel and Laplacian filters. These filters emphasize edge and texture information, contributing to improved representation learning. During the inference phase, the EDBB blocks are reparameterized into 3x3 convolutions to maintain computational efficiency while preserving learned feature representations.

The attention maps are derived following the approach implemented in SPAN, leveraging an activation function that is both odd and symmetric to effectively highlight essential regions of the image. These attention maps serve as a direct substitute for the ESA block present in the original EFDN model, aiming to refine feature selection and enhance the model's overall performance.

As a result of the applied modifications, the final architecture has a lower parameter count and requires fewer

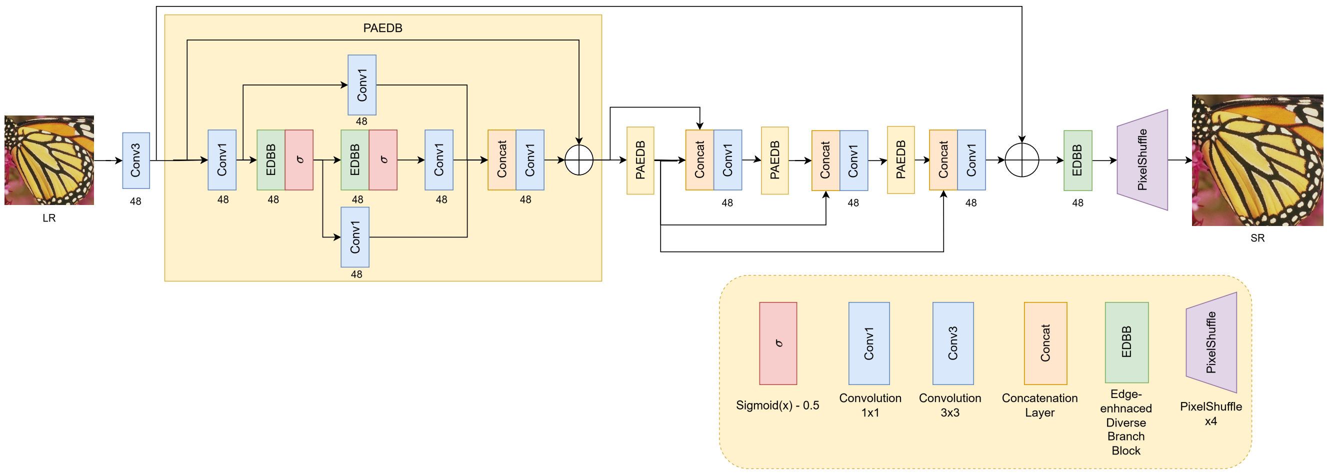

Figure 1. Schematic diagram of the proposed method

Method	PSNR/SSIM						# Params (M)	Runtime (ms)	FLOPs (G)
	DIV2K_LSDIR		Set5	Set14	Urban100	BSD100			
	Val	Test							
EFDN	**26.96/0.7965**	**27.01/-**	**30.18/0.8931**	**25.65/0.7507**	**24.48/0.7815**	**24.99/0.7099**	0.276	29.71	16.70
Ours	26.66/0.7890	26.76/0.80	29.74/0.8868	25.50/0.7472	24.07/0.7472	24.94/0.7061	**0.240**	**28.42**	**15.64**

Table 1. Results obtrained by the proposed method for datasets DIV2K_LSDIR [1] [4], Set5 [2], Set14 [8], BSD100 [5], Urban100 [3], compared with the baseline method from the challenge, EFDN. Best results are reported in **bold**.

floating-point operations compared to the proposed baseline method, EFDN.

Training details The training process is structured into three progressive phases to optimize performance and stability:

- Pre-training: The model undergoes an initial training phase using the DIV2K dataset, incorporating data augmentation techniques such as random rotations, horizontal flipping, and random cropping to generate patches of size 64×64. Training is conducted over 30,000 iterations with a batch size of 32, utilizing the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$). The learning rate is initially set to 1e-3 for the first 20,000 iterations and subsequently reduced to 1e-4 for the remaining 10,000 iterations. L1 loss is used throughout this phase.

- First training stage: The model is further refined using the DIV2K_LSDIR dataset, while maintaining the same augmentation strategies as in the pre-training phase. The patch size is increased to 256×256, and training is extended to 100,000 iterations with a batch size of 64. The Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$) is employed, starting with a learning rate of 5e-4, which undergoes a decay by a factor of 0.5 every 20,000 iterations. L1 loss remains the chosen loss function for this stage.

- Second training stage: In the final phase, training continues on the DIV2K_LSDIR dataset with an expanded patch size of 512×512 for an additional 40,000 iterations. The same augmentation methods are retained, and most hyperparameters remain unchanged. However, to ensure stable convergence and fine-tune performance, the learning rate is reduced to 5e-5. During this stage, L1 loss is applied for the first 10,000 iterations, after which L2 loss is utilized to enhance final model performance.

  All the training phases were performed of the model a single NVIDIA RTX 4070 Super GPU and required approximately 20 hours.

Experimental results We evaluate our model in terms of Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) and compare it against the baseline EFDN model. The obtained results, as indicated in Table 1, show that while our proposed approach does not surpass the baseline in terms of these performance metrics, it achieves comparable results while reducing computational complexity.

Specifically, our modified architecture requires fewer floating-point operations (GFLOPs), has a reduced parameter count, and demonstrates lower runtime (both runtime results were computed locally on our machine).

Considering the relatively short training duration and the observed efficiency improvements, our approach presents a viable and competitive alternative to the baseline model.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 126–135, 2017. 2

[2] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. 2012. 2

[3] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5197–5206, 2015. 2

[4] Yawei Li, Kai Zhang, Jingyun Liang, Jiezhang Cao, Ce Liu, Rui Gong, Yulun Zhang, Hao Tang, Yun Liu, Denis Demandolx, et al. Lsdir: A large scale dataset for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1775–1787, 2023. 2

[5] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings eighth IEEE international conference on computer vision. ICCV 2001*, volume 2, pages 416–423. IEEE, 2001. 2

[6] Cheng Wan, Hongyuan Yu, Zhiqi Li, Yihang Chen, Yajun Zou, Yuqing Liu, Xuanwu Yin, and Kunlong Zuo. Swift parameter-free attention network for efficient super-resolution, 2024. 1

[7] Yan Wang. Edge-enhanced feature distillation network for efficient super-resolution, 2022. 1

[8] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *Curves and Surfaces: 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers 7*, pages 711–730. Springer, 2012. 2