NTIRE 2025 Image Super-Resolution ($\times$4) Challenge Factsheet -Pre-trained Models with Ensemble Learning for Image Super-resolution

Weijun Yuan
Jinan University
Guangzhou, China
yweijun@stu2022.jnu.edu.cn

Zhan Li
Jinan University
Guangzhou, China
lizhan@jnu.edu.cn

Zhanglu Chen
Jinan University
Guangzhou, China
czhanglu@stu2024.jnu.edu.cn

Boyang Yao
Jinan University
Guangzhou, China
yaoboy@stu.jnu.edu.cn

1. Description

Inspired by the remarkable performance of Transformer-based models and Mamba-based models, we introduce a framework for image super-resolution based on pre-trained models and ensemble learning (PMELSR). As shown in Fig 1, our pipeline can be divided into two stages. We employed the idea of ensemble learning to design our pipeline. In the first stage, we have chosen the DAT [2], HAT-L [1], and MambaIRv2 [3] which have achieved outstanding results in the field of image super-resolution as the backbone to process the low-resolution separately, and then fusion the results. In the second stage, we used the RRDBNet [4] as the refinement module to optimize the previous fused results and further improve the image quality. The design of the entire pipeline aims to maximize image quality although it increases a lot of computational complexity.

2. Implementation Details

During the training phase, the parameters of pre-trained models (DAT, HAT-L, and MambaIRv2) are fixed, and only the RRDBNet refinement module is trained. The PMELSR is optimized using the Adam optimizer, with L1 loss function, and data augmentations including flips and rotations. The DIV2K dataset is used for training, where HR patches of size 512×512 are randomly cropped from the HR images. The mini-batch size is set to 4, and the total number of training iterations is 400K. The initial learning rate is set to 2e-4, with the learning rate decayed using the cosine annealing strategy.

In the testing phase, to further mitigate the prediction bias of the super-resolution model, two ensemble learning strategies are employed: self-ensemble and model ensemble. Specifically, the self-ensemble method is first applied to enhance the performance of all pre-trained models. Subsequently, the model-ensemble strategy is employed to combine the outputs of all pre-trained models. Finally, the fused results are further optimized by applying RRDBNet in conjunction with the self-ensemble approach to enhance image quality.

3. Team details

- Team name: JNU620
- Team leader name: Weijun Yuan
- Team leader address, phone number, and email: Jinan University, No.601 Huangpu Dadao Xi, Tianhe District, Guangzhou, Guangdong, China; 17817975121; yweijun@stu2022.jnu.edu.cn
- Rest of the team members: Zhan Li, Zhanglu Chen, Boyang Yao
- Team website URL (if any): None
- Affiliation: Jinan University
- Affiliation of the team and/or team members with NTIRE 2025 sponsors (check the workshop website): None
- User names and entries on the NTIRE 2025 Codalab competitions (development/validation and testing phases): Wedream
- Best scoring entries of the team during the development/validation phase: Validation phase: PSNR: **31.5664**; Test phase: PSNR: **31.1192**;
- Link to the codes/executables of the solution(s): https://github.com/Wedream-wj/NTIRE2025_ImageSR_x4

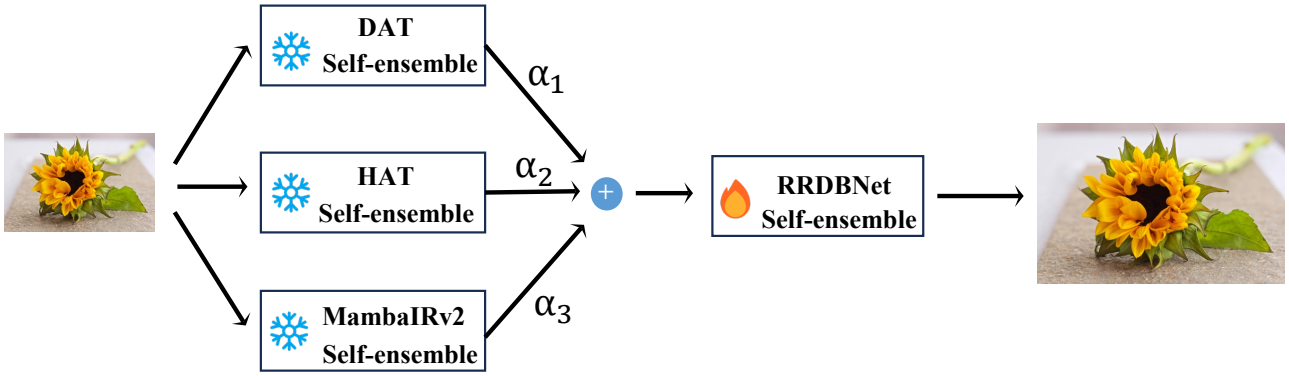

Figure 1. The architecture overview of the proposed PMELSR

References

[1] Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22367–22377, 2023. 1

[2] Zheng Chen, Yulun Zhang, Jinjin Gu, Linghe Kong, Xiaokang Yang, and Fisher Yu. Dual aggregation transformer for image super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 12312–12321, 2023. 1

[3] Hang Guo, Yong Guo, Yaohua Zha, Yulun Zhang, Wenbo Li, Tao Dai, Shu-Tao Xia, and Yawei Li. Mambairv2: Attentive state space restoration. *arXiv preprint arXiv:2411.15269*, 2024. 1

[4] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018. 1