NTIRE 2024 Image Super-Resolution ($\times 4$) Challenge Factsheet
Pre-trained Model with Ensemble Learning for Image Super-resolution

Zhiyuan Song
Sun Yat-sen University
Sun Yat-sen University, Guangzhou, China
songzhy29@mail2.sysu.edu.cn

Ziyue Dong
Xi'an Jiaotong University
Xi'an Jiaotong University, Xi'an, China
ziyuedong@stu.xjtu.edu.cn

Qing Zhao
Sun Yat-sen University
Sun Yat-sen University, Guangzhou, China
zhaoq78@mail2.sysu.edu.cn

Xiaogang Xu
Zhejiang University
Zhejiang University, Hangzhou, China
xiaogangxu@zju.edu.cn

Pengxu Wei
Sun Yat-sen University
Sun Yat-sen University, Guangzhou, China
weipx3@mail.sysu.edu.cn

1. Team details

- Team name: SYSU-SR
- Team leader name: Zhiyuan Song[1]
- Team leader address, phone number, and email:
  Sun Yat-sen University, Guangzhou, China
  15206389893
  songzhy29@mail2.sysu.edu.cn
- Rest of the team members: Ziyue Dong[2], Qing Zhao [1], Xiaogang Xu [3], Pengxu Wei[1]
- Affiliation: [1]Sun Yat-sen University, [2] Xi'an Jiaotong University, [3]Zhejiang University
- User names and entries on the NTIRE 2024 Codalab competitions (development/validation and testing phases)
  fan_g, m0NESY and Qing_Zhao
- Best scoring entries of the team during the development/validation phase:
  PSNR: 31.3954
  SSIM: 0.85
- Link to the codes/executables of the solution(s)
  https://github.com/Song-Zhiyuan/NTIRE2024-SYSU-SR

2. Method details

2.1. Network Architecture

The whole model workflow proposed by our team is shown in Fig. 1. Inspired by the excellent performance of large pre-trained model on the low-level computer vision tasks, we choose the HAT-L [1] pre-trained model as our main structure and we proposed enhancement in the two phases of training and testing, respectively.

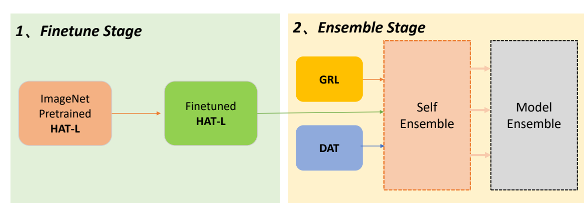

Figure 1. *SYSU-SR Team* : The flow chart of the proposed method

2.2. Training description

In order to further improve the performance of the pre-trained model, our team finetunes the pre-trained model from two main aspects: the loss function and the training strategy. The loss function uses L1loss and Gradient-Weighted (GW) loss [6], which constrains image's local structure and texture to generate more accurate details, which is efficient for the super-resolution task. The training strategy uses progressive training, and some works [4, 7] demonstrates its effectiveness.

2.3. Testing description

In order to further reduce the model's prediction bias on the super-resolution, our team fused two ensemble learning strategies, self-ensemble as well as model-ensemble, to enhance the model's test performance. Our team first employs the self-ensemble approach [5] to enhance all candidate models. Secondly, the model-ensemble approach proposed by ZZPM team [8]is applied to all enhanced models, where the average of the outputs of different models is calculated, and then the weights of the ensemble are assigned based on the MSE value between each model and the average. Our experiments show that the fusion of two ensemble strategies can achieve higher performance than each single strategy.

2.4. Implementation details

In the model finetuning phase, the training data contains DIV2K, Flickr2K and LSDIR and the training loss is $\mathcal{L}_{total} = \alpha \cdot \mathcal{L}_1 + \beta \cdot \mathcal{L}_{GW}$ where α and β are weights assigned to the two losses. We set $\alpha = 1, \beta = 3$ for our training setup and all experiments is conducted on 8 NVIDIA A100 GPUs using Adam optimizer. At the beginning of progressive training, the patch size is 64 and batchsize is 32, keeping the learning rate as 1×10^{-5} for $125k$ iterations. Finally setting the patch size as 128, batchsize as 16 with the learning rate as 5×10^{-6} for $60k$ iterations.

In the model testing phase, the pre-trained GRL [3], DAT [2] models and the finetuned HAT-L model are selected for fusion of outputs to obtain higher performance.

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. 2

[3] Yawei Li, Yuchen Fan, Xiaoyu Xiang, Denis Demandolx, Rakesh Ranjan, Radu Timofte, and Luc Van Gool. Efficient and explicit modelling of image hierarchies for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18278–18289, 2023. 2

[4] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021. 1

[5] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 2

[6] Pengxu Wei, Ziwei Xie, Hannan Lu, Zongyuan Zhan, Qixiang Ye, Wangmeng Zuo, and Liang Lin. Component divide-and-conquer for real-world image super-resolution. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16*, pages 101–117. Springer, 2020. 1

[7] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022. 1

[8] Yulun Zhang, Kai Zhang, Zheng Chen, Yawei Li, Radu Timofte, Junpei Zhang, Kexin Zhang, Rui Peng, Yanbiao Ma, Licheng Jia, et al. Ntire 2023 challenge on image super-resolution (x4): Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1864–1883, 2023. 2