NTIRE 2025 Image Super-Resolution ($\times 4$) Challenge Factsheet
-title of the contribution-

Yinxiang Zhang
Northwestern Polytechnical University
Xi'an, China
zhangyinxiang@mail.nwpu.edu.cn

Wenxuan Cai
Northwestern Polytechnical University
Xi'an, China
caiwenxuan@mail.nwpu.edu.cn

1. Introduction

This factsheet template is meant to structure the description of the contributions made by each participating team in the NTIRE 2025 challenge on image super-resolution ($\times 4$).

Ideally, all the aspects enumerated below should be addressed. The provided information, the codes/executables, and the achieved performance on the testing data are used to decide the awardees of the NTIRE 2025 challenge.

Reproducibility is a must and needs to be checked for the final test results in order to qualify for the NTIRE awards.

The main winners will be decided based on overall performance and a number of awards will go to novel, interesting solutions and to solutions that stand up as the best in a particular subcategory the judging committee will decide. Please check the competition webpage and forums for more details.

The winners, the awardees and the top-ranking teams will be invited to co-author the NTIRE 2025 challenge report and to submit papers with their solutions to the NTIRE 2025 workshop. Detailed descriptions are appreciated.

The factsheet, source codes/executables, trained models should be sent to **all of the NTIRE 2025 challenge organizers (Zheng Chen, Jue Gong, Jingkai Wang, Kai Liu, Lei Sun, Zongwei Wu, Yulun Zhang, and Radu Timofte)** by email.

2. Email final submission guide

To: zhengchen.cse@gmail.com
leosun0331@gmail.com
g1017325431@gmail.com
normal.kliu@gmail.com
jingkaiwang100@gmail.com
zongwei.wu@uni-wuerzburg.de
yulun100@gmail.com
Radu.Timofte@uni-wuerzburg.de
cc: your_team_members
Title: NTIRE 2025 Image Super-Resolution ($\times 4$) Challenge - TEAM_NAME - TEAM_ID

To get your TEAM_ID, please register at Google Sheet. Please fill in your Team Name, Contact Person, and Contact Email in the first empty row from the top of the sheet. Body contents should include:

a) team name
b) team leader's name and email address
c) rest of the team members
d) user names on NTIRE 2025 CodaLab competitions
e) Code, pre-trained model, and factsheet download command, e.g. `git clone ...`, `wget ...`
f) Result download command, e.g. `wget ...`
   - Please provide different URLs in e) and f)

Factsheet must be a compiled pdf file together with a zip with .tex factsheet source files. Please provide a detailed explanation.

3. Code Submission

The code and trained models should be organized according to the GitHub repository. This code repository provides the basis for comparing the various methods in the challenge. **Code scripts based on other repositories will not be accepted.** Specifically, you should follow the steps below.

1. Git clone the repository
2. Put your model script under the `models/[Team_ID]_[Model_Name]` folder
3. Put your pretrained model under the `model_zoo/[Team_ID]_[Model_Name]` folder
4. Modify `model_path` in `test.py`. Modify the imported models
5. `python test.py` (restore images, details in GitHub)
6. `python eval.py` (eval results, details in GitHub)

Please send us the command to download your code, e.g. `git clone [Your repository link]` When submitting the code, please remove the input and output images in the (any) `data` folder to save the bandwidth.

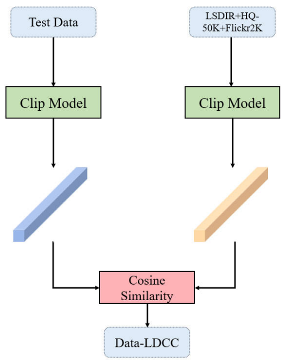

Figure 1. The flow chart of the dataset construction process

4. Factsheet Information

The factsheet should contain the following information. Most importantly, you should describe your method in detail. The training strategy (optimization method, learning rate schedule, and other parameters such as batch size, and patch size) and training data (information about the additional training data) should also be explained in detail.

4.1. Team details

- Team name: Endeavour
- Team leader name: Yinxiang Zhang Xi'an, China
- Team leader address: Xi'an, China
- Team leader phone number: 15904671917
- Team leader email: zhangyinxiang@mail.nwpu.edu.cn
- Rest of the team members:Wenxuan Cai, caiwenxuan@mail.nwpu.edu.cn
- Affiliation:Northwestern Polytechnical University, Xi'an, China
- Affiliation of the team and/or team members with NTIRE 2025 sponsors (check the workshop website)
- User names and entries on the NTIRE 2025 Codalab competitions (development/validation and testing phases)
  - User name : zyx.
  - development/validation entries : 8
  - User name : dream_it
  - testing entries : 3
- Best scoring entries of the team during the development/validation phase
  - development/validation phase Ranking : 4
  - PSNR : 31.45
  - SSIM : 0.85
- Link to the codes/executables of the solution(s): source codes/executables

4.2. Method details

We propsed an innovative approach based on the HAT[1] network model, enhanced with a frequency-domain fusion module, to improve the performance of image restoration tasks. By constructing a high-quality Data-LDCC dataset, the method incorporates the CLIP model to extract latent features and employs cosine similarity to ensure semantic consistency between the data and the test set, effectively bridging the domain gap between training and testing data. The training process is divided into three stages: initial pre-training on the ImageNet dataset, followed by optimization on the DF2K dataset, and finally fine-tuning on both the Data-LDCC and DF2K datasets to ensure high performance in the target domain. The introduction of the frequency-domain fusion module further enhances the model's ability to capture and reconstruct fine-grained details.

Methodology Overview. Inspired by the remarkable advancements achieved through scaling and data optimization in tasks such as image restoration, this work proposes an enhanced approach by leveraging a high-quality dataset and integrating innovative architectural modifications. The proposed method adopts HAT as the base network model, which is further augmented with a frequency-domain fusion module to enhance feature representation and reconstruction capabilities. The training process is meticulously designed and divided into three distinct stages to ensure optimal performance.

Dataset Construction. A key contribution of this work is the construction of the Data-LDCC dataset, which is specifically tailored to bridge the domain gap between training and testing data. The dataset is carefully selected by first extracting latent features from a combined pool of LSDIR[3], HQ-50K, and Flickr2K[2] datasets, along with the test data, using the CLIP model. Subsequently, cosine similarity[4] is computed between the extracted features of the candidate data and the test data. Images with the highest similarity scores are selected to form the Data-LDCC dataset, ensuring semantic alignment and domain consistency.The process of data construction is shown in Fig. 1.

Training Pipeline. The training process is structured into three stages:

1. Stage 1: The model is initially trained on the ImageNet dataset to establish a robust foundational representation.
2. Stage 2: The pre-trained model is further optimized using the DF2K dataset, which includes diverse and high-quality images to enhance generalization.
3. Stage 3: Fine-tuning is performed on a combination of the Data-LDCC and DF2K datasets. This stage ensures that the model is specifically adapted to the target domain, leveraging the domain-aligned properties of the Data-LDCC dataset to improve performance on the test data.

In the first stage, the model is trained with a patch size of 64×64, a batch size of 256, and 70,000 iterations to establish a robust foundational representation; in the second stage, the patch size is increased to 96×96, the batch size

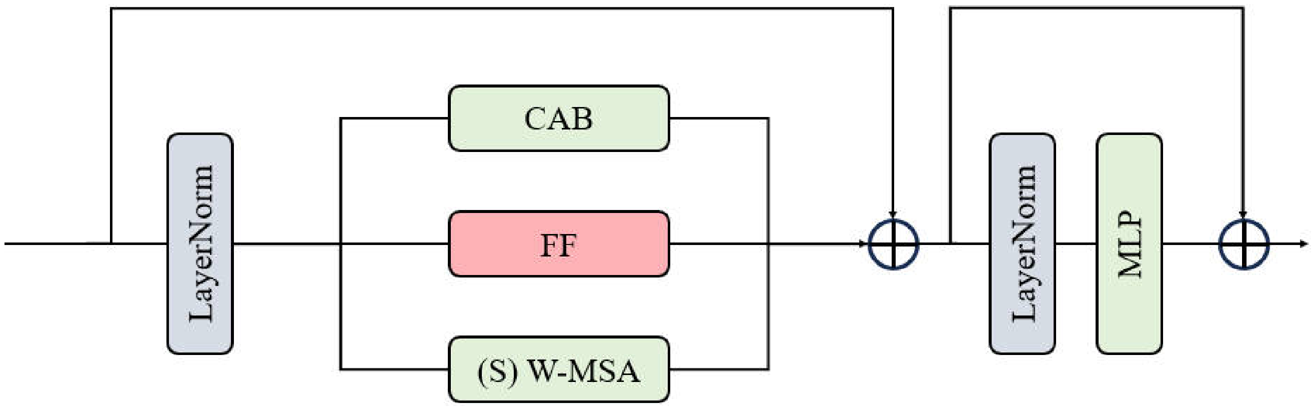

Figure 2. The proposed module

is reduced to 128, and the model undergoes 50,000 iterations for further optimization and improved generalization; finally, in the third stage, the patch size is further increased to 112×112, the batch size is reduced to 64, and the model is fine-tuned for 20,000 iterations to refine its performance on high-resolution patches and ensure domain-specific adaptation. This progressive training strategy enables the model to gradually adapt to increasingly complex data distributions, ensuring high-quality restoration results.

Network Architecture. The proposed model builds upon the HAT architecture, which integrates channel attention and window-based self-attention mechanisms. To further enhance the model's ability to capture and reconstruct fine details, a frequency-domain fusion module is incorporated. This module operates in the frequency domain to complement the spatial features extracted by the base network, enabling richer feature representation and improved restoration quality.The proposed module is shown in Fig. 2.

5. Other details

- Planned submission of a solution(s) description paper at NTIRE 2025 workshop.
- General comments and impressions of the NTIRE 2025 challenge.
- What do you expect from a new challenge in image restoration, enhancement and manipulation?
- Other comments: encountered difficulties, fairness of the challenge, proposed subcategories, proposed evaluation method(s), etc.

References

[1] Xiangyu Chen, Xintao Wang, Jiantao Zhou, and Chao Dong. Activating more pixels in image super-resolution transformer. *arXiv preprint arXiv:2205.04437*, 2022. 2

[2] R.Timofte et al. Ntire 2017 challenge on single image super-resolution: Methods and results. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1110–1121, 2017. 2

[3] 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. 2

[4] Amit Singhal et al. Modern information retrieval: A brief overview. *IEEE Data Eng. Bull.*, 24(4):35–43, 2001. 2