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

Zonghao Chen
KUNBYTE
Hangzhou, Zhejiang Province, China
chenzonghao@k-fashionshop.com

Yang Ji
KUNBYTE
Hangzhou, Zhejiang Province, China
jiyang_qut@163.com

Xi Wang
KUNBYTE
Hangzhou, Zhejiang Province, China
wangxi@k-fashionshop.com

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 much appreciated.

The factsheet, source codes/executables, trained models should be sent to **all of the NTIRE 2025 challenge organizers (Zheng Chen, Zongwei Wu, Eduard-Sebastian Zamfir, Kai Zhang, Yulun Zhang, Radu Timofte, and Xiaokang Yang)** by email.

2. Email final submission guide

To: zhengchen.cse@gmail.com
zongwei.wu@uni-wuerzburg.de
Eduard-Sebastian.Zamfir@uni-wuerzburg.de
cskaizhang@gmail.com
yulun100@gmail.com
Radu.Timofte@uni-wuerzburg.de
xkyang@sjtu.edu.cn
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` folder. Name your model script as `[Your_Team_ID]_[Your_Model_Name].py`.

3. Put your pretrained model under the `model_zoo` folder. Name your model checkpoint as `[Your_Team_ID]_[Your_Model_Name].[pth or pt or ckpt]`
4. Modify `model_path` in `test_demo.py`. Modify the imported models.
5. `python test_demo.py`

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

4. Factsheet Information

4.1. Team details

- Team name: Aimanga
- Team leader name: Zonghao Chen
- Team leader address: Hangzhou, Zhejiang Province, China
  phone number: 18817558184
  email: chenzonghao@k-fashionshop.com
- Rest of the team members: Yang Ji, Xi Wang
- Team website URL: https://discord.com/servers/mira-1256955108718284800
- Affiliation: KUNBYTE
- User names: yangyi1221
- Best scoring: 31.214
- Link to the codes:
  https://github.com/jiyang0315/NTIRE2025_ImageSR_x4.git

4.2. Method details

This model is trained based on the RealESRGAN[5]. We have done in-depth thinking in cleaning data and image degradation models to achieve higher image reconstruction effects and better generation quality.

As shown in Figure 1, the generator is based on an RRDBNet architecture, which primarily consists of multiple stacked Residual-in-Residual Dense Blocks (RRDB). Each RRDB integrates dense connections and residual skip connections, with residual scaling applied to stabilize training. The network removes batch normalization (BN) layers to avoid artifacts and incorporates pixel attention or channel attention mechanisms to enhance detail recovery. The input image first undergoes shallow feature extraction via convolutional layers, then passes through multiple RRDB blocks for deep feature refinement. Finally, PixelShuffle[4] upsampling reconstructs the high-resolution output. This design prioritizes preserving fine details and natural textures, making it effective for super-resolution tasks.

As shown in Figure 2, the discriminator adopts a U-Net-based[3] architecture to better capture both global and local details. It consists of an encoder with strided convolutions for downsampling and a decoder with transposed convolutions for upsampling, connected by skip connections to preserve spatial information. The network employs spectral normalization for stable training and uses PatchGAN-style[2] prediction to distinguish real vs. fake patches at multiple scales. Additionally, LeakyReLU activation and instance normalization help improve discrimination ability while maintaining gradient flow. This design enables effective adversarial training by providing detailed feedback to the generator on both high-level structures and fine textures.

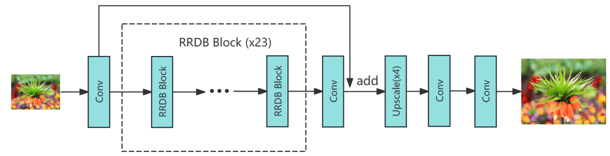

Figure 1. Architecture of Generator.

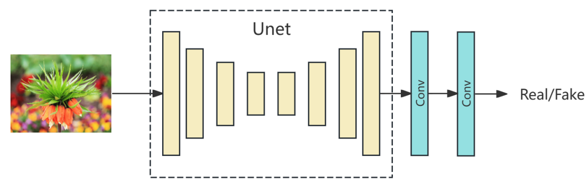

Figure 2. Architecture of Discriminator.

We used nomos2k as the training set. To extract the detailed areas in the image, we designed an effective data screening process. Finally, we selected 1342 images from 2436 nomos2k data. We carefully analyzed and tuned the image degradation model. Through continuous experiments, we found very effective image degradation models and parameters. The training process takes over 100k iterations with a batch size of 16 and a fixed learning rate of 1e-4.

To evaluate the performance of our trained model, we created a synthetic evaluation set of 125 images. The evaluation set contains 10 categories such as human, animals, animation, and AI. This can more accurately evaluate the model's capabilities for various scenes. We further conducted experiments on the real-world dataset RealSR[1], which contains 100 images captured by Canon 5D3 and Nikon D810 cameras.

We compared our model with RealESRGAN and the latest work InvSR[6] in CVPR2025. We used niqe, brisque, nrqm, pi, clipiqa and musiq as evaluation indicators. The evaluation results are shown in the following table, from which we can see that our model is better than these two models in most indicators.

Table 1. Performance comparison on the Synthetic testset

Synthetic testset						
Models	niqe↓	brisque↓	nrqm↑	pi↓	clip-iqa↑	musiq↑
Realesrgan_x4plus[5]	3.443	20.604	6.333	3.628	0.563	63.923
InvSR(CVPR2025)[6]	3.234	13.382	6.751	3.255	0.706	**69.114**
Ours	**2.937**	**11.902**	**6.854**	**3.081**	**0.713**	66.984

Table 2. Performance comparison on the Realworld testset

Realworld testset						
Methods	niqe↓	brisque↓	nrqm↑	pi↓	clip-iqa↑	musiq↑
Realesrgan_x4plus[5]	4.668	29.14	5.864	4.446	0.491	59.669
InvSR(CVPR2025)[6]	4.223	16.484	6.694	3.78	0.692	**67.473**
Ours	**3.896**	**15.881**	**6.831**	**3.537**	**0.7**	64.874

5. Other details

- We are preparing to submit a solution description paper to the NTIRE 2025 workshop, focusing on advancements in AI-based image and video enhancement techniques. The paper will present our novel model architectures designed to address real-world degradation challenges, alongside a comparison with existing state-of-the-art methods. We will also discuss the performance of our models on benchmark datasets, showcasing improvements in efficiency and quality. Additionally, we will explore potential applications and outline future directions for further development. This submission aims to contribute to the ongoing research and development in the field of image and video restoration.

References

[1] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3086–3095, 2019. 2

[2] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. 2

[3] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015. 2

[4] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016. 2

[5] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1905–1914, 2021. 2, 3

[6] Zongsheng Yue, Kang Liao, and Chen Change Loy. Arbitrary-steps image super-resolution via diffusion inversion. *arXiv preprint arXiv:2412.09013*, 2024. 2, 3