NTIRE 2025 Image Super-Resolution ($\times$4) Challenge Factsheet
MDRCT: Multi-step Dense Residual Connected Transformer

Snehal Singh Tomar*
Stony Brook University
stomar@cs.stonybrook.edu

Rajarshi Ray*
Stony Brook University
rarray@cs.stonybrook.edu

Klaus Mueller
Stony Brook University
mueller@cs.stonybrook.edu

1. Introduction

This factsheet describes the contributions made by Team VAI-GM in the NTIRE 2025 Image Super-Resolution (SR) ($\times$4) Challenge. Our solution is built upon the DRCT architecture[4], which has demonstrated state-of-the-art performance on various SR tasks (e.g., BSD100$\times$4). Inspired by the methods proposed in the NTIRE 2024 challenge reports[2], we extend the original DRCT design by introducing several important modifications to the architecture and training pipeline. First, we introduce a Multi-step residual formulation in the architecture. Second, we adjust the learning rate. Third, we finetune on external datasets. These modifications enhance the gradient flow and help retain feature information across layers, leading to improved SR performance.

2. Factsheet Information

2.1. Team Details

- **Team name:** VAI-GM
- **Team leader:** Rajarshi Ray
  Email: rarray@cs.stonybrook.edu
  Phone: +1 631 949 3136
- **Other team members:** Snehal Singh Tomar, Prof. Klaus Mueller
- **Affiliation:** Stony Brook University
- **NTIRE 2025 Codalab Username:** RajarshiRay29
- **Best scoring entry:** 30.62 PSNR on the challenge's test datasetset.
- **Code/Executable:** https://github.com/rajarshi-ray29/NTIRE2025_ImageSR_x4_team06

2.2. Method Details

Our method builds on the DRCT architecture and introduces key modifications:

Architecture: Our method builds on the DRCT architecture and introduces key modifications. The baseline network is DRCT, which is already known for achieving state-of-the-art results in super-resolution (SR) tasks. We incorporate additional skip connections between every two RDG (Residual Dense Group) blocks. These connections ensure that important features from earlier layers are directly passed on to deeper layers, reducing the risk of vanishing gradients and helping the model preserve low-level details that are crucial for high-quality image reconstruction. This idea is partly inspired by traditional residual learning strategies (e.g., ResNet) [3], and recent transformer-based architectures like HAT (Hybrid Attention Transformer) [1], which demonstrate the benefit of combining hierarchical attention with skip connections for robust feature flow and high-quality reconstructions. By blending these principles into the DRCT framework, we improve upon the standard architecture by ensuring better feature flow and more stable training.

Representative Pipeline: The network comprises several RDG blocks interconnected through our extra skip links, as outlined in our model.py. A more detailed diagram is provided in our supplementary materials. In essence, the input travels through each RDG block, with these additional residual paths preventing the loss of vital information. This design enables the model to combine learned high-level features with essential fine-grained details, leading to more precise image reconstructions. Our modifications thus have a tangible impact: they help retain key texture details, allow for more robust convergence during training, and ultimately yield higher PSNR scores.

Training Strategy: Training was conducted using the DF2K dataset (a combination of DIV2K with 800 images and Flickr2K with 2650 images) for the training phase, and Unsplash2K (20 images) for validation. This diverse training set provides varied examples for the model to learn from, helping it generalize better. We further harnessed pre-training by starting from a DRCT-L model initialized on ImageNet, then fine-tuning it on our super-resolution task. This approach leverages the generic feature extraction capabilities gained from large-scale ImageNet training, adapting

* Denotes equal contribution.

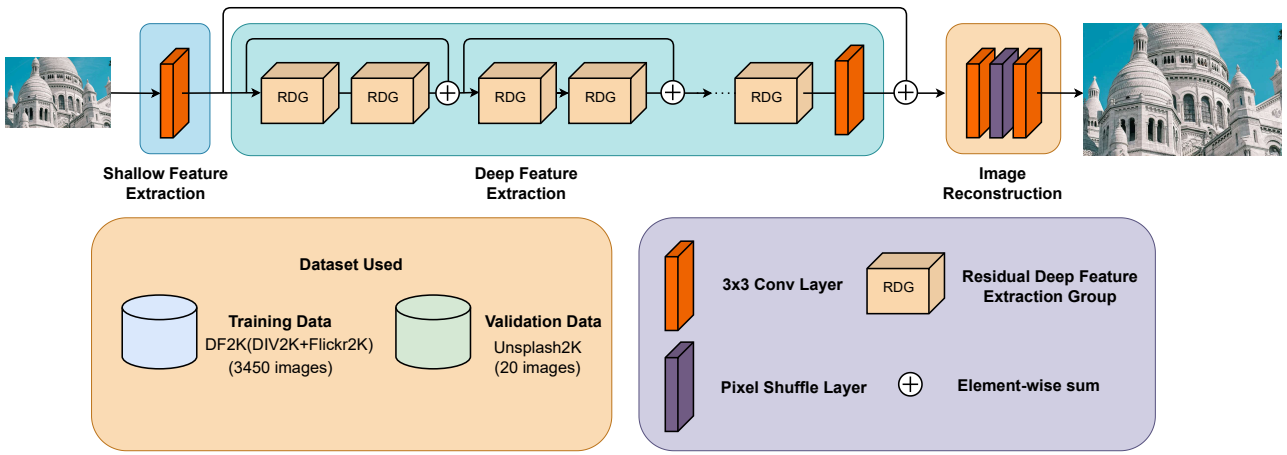

Figure 1. Team VAI-GM DSRCT architecture

Method	PSNR Val	SSIM	Time	Training Dataset	Validation Data
DSRCT	30.86	0.8942	2 days 3 hours	DIV2K (800)+Flickr2K (2650)	Unsplash2K (20)

Table 1. Quantitative Results from Training implemented on Tesla V100-SXM2 GPU w

them specifically to the nuances of high-resolution image synthesis. Key training parameters are as follows:

- **Optimizer:** Adam with a learning rate of 5×10^{-4}, weight decay set to zero, and betas [0.9, 0.99].
- **Scheduler:** MultiStepLR with milestones at [125000, 200000, 225000, 240000] iterations and a decay factor of 0.5.
- **Iterations:** Total of 250,000 iterations.
- **Data Augmentation:** Random horizontal flips and rotations were applied.
- **Batch Size and Patch Size:** Batch size per GPU is 4 with a ground truth patch size of 256.

Experimental Results: Our experiments show a consistent PSNR above 30 on the validation set during training, with our best test entry achieving 30.86 PSNR.

These results emphasize the effectiveness of our added skip connections and overall network design. By carefully fine-tuning a strong baseline, integrating established ideas of residual learning, and utilizing a rich training set, our approach pushes the DRCT architecture forward and offers a more robust solution for super-resolution tasks.

2.3. Additional Remarks

We plan to submit a detailed solution paper to the NTIRE 2025 workshop, highlighting our method's focus on integrating skip connections in deep transformer-based architectures for image super-resolution. The review paper of last year was especially insightful and helped us grasp recent advancements and trends in the field. Feedback from previous NTIRE challenges emphasized the importance of innovative training strategies and effective data augmentation to boost performance.

We are grateful to the organizers for providing this platform and the opportunity to learn throughout the challenge. Regardless of acceptance, any feedback on our submission would be immensely valuable. We believe that further discussions on future directions, such as new categories, fair evaluation methods, and broader tasks in image restoration, enhancement, and manipulation, will continue to push the boundaries of this research area.

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 (CVPR)*, pages 22367–22377, 2023. 1

[2] Zheng Chen, Zongwei Wu, Eduard Zamfir, Kai Zhang, Yulun Zhang, Radu Timofte, Xiaokang Yang, et al. NTIRE 2024 challenge on image super-resolution (×4): Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 6108–6127, 2024. 1

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 1

[4] Chih-Chung Hsu, Chia-Ming Lee, and Yi-Shiuan Chou. DRCT: Saving image super-resolution away from information bottleneck. *arXiv preprint arXiv:2404.00722*, 2024. 1