NTIRE25-Single Image Reflection Removal (SIRR) in the Wild Challenge

Your Teamname Here
Your names*

March 24, 2025

This factsheet template is meant to structure the description of the contributions made by each participating team in the NTIRE25-Single Image Reflection Removal (SIRR) in the Wild Challenge. 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 NTIRE25-Single Image Reflection Removal (SIRR) in the Wild Challenge.

1. Reproducibility is a must and needs to be checked for the final test results in order to qualify for the NTIRE 2025 awards.

2. 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 decided. The final ranking will be decided based on the reported performance in terms of quantitative metrics, subjective scores, and solution efficiency. Please check the competition webpage and forums for more details.

3. The winners, the awardees and the top ranking teams will be invited to co-author the NTIRE25-Single Image Reflection Removal (SIRR) in the Wild Challenge report and to submit papers with their solutions to the NTIRE 2025 workshop. Detailed descriptions are much appreciated: brief description of all models and experiments tested, ablations, visualizations, things that did not work, external datasets, pre-trained models, ensembles, etc. You can share qualitative results (or even the raw images itself) via shared folder at google drive, dropbox, etc.

The factsheet, source codes/executables, and additional results (if required) should be sent to **all of the NTIRE2025 challenge organizers (Kangning Yang, Florin-Alexandru Vasluianu, Radu Timofte)** by email. The image results should be reproducible and match the **best** submission on Codalab. You still need to upload your results on Codalab! As the file size is expected to be large, we accept download links from your website, google drive, dropbox, etc. When using cloud services, please remember to allow sharing and consider the traffic limit of your service provider.

Organizer name	Email
Kangning Yang	kangning.yang@oppo.com
Florin-Alexandru Vasluianu	florin-alexandru.vasluianu@uni-wuerzburg.de
Radu Timofte	radu.timofte@uni-wuerzburg.de

*Affiliations

Email final submission guide

To: kangning.yang@oppo.com; florin-alexandru.vasluianu@uni-wuerzburg.de; radu.timofte@uni-wuerzburg.de
cc: your_team_members
Title: NTIRE25-Single Image Reflection Removal (SIRR) in the Wild Challenge - TEAM_NAME

Body contents should include:
a) the challenge name
b) team name
c) team leader's name and email address
d) rest of the team members
e) team members with NTIRE 2025 sponsors
f) team name and user names on NTIRE 2025 CodaLab competitions
g) executable/source code attached or download links.
h) factsheet attached
i) download link to the results (submitted images and additional results if requested).

The executable/source code should include trained models or necessary parameters so that we could run it and reproduce results. There should be a README or descriptions that explains how to execute the executable/code. Factsheet must be a compiled pdf file together with a zip with .tex factsheet source files (including figures). Please provide a detailed explanation.

1 Team details

- Team name

  AIIA

- Team leader name

  Mengru Yang

- Team leader institution and email (Please make sure is an active email)

  Harbin Institute of Technology ymr2200642844@163.com

- Rest of the team members

  Kui Jiang, Jin Guo, Yiang Chen, Junjun Jiang,

- Team website URL (if any)

- Affiliations

  Harbin Institute of Technology

- Usernames on the NTIRE25-Single Image Reflection Removal (SIRR) in the Wild Challenge Codalab leaderboard (development/validation and testing phases)

  koucy,brewie,Mr_Yang

- Link to the codes/executables of the solution(s)

  Our codes can be found from this

- Link to the enhancement results of all frames

  Our results can be downloaded from this

2 Contribution details

- DualPatchFusion-ReflectNet (DPF-ReflectNet)
- General method description In this study, an innovative reflection removal framework was proposed to improve the model performance through three technological breakthroughs: First, a two-stage fine-tuning strategy was designed, and a one-stage fine-tuning training was performed on the training set based on the pre-trained model of RDNet[1]. After the training was stable, patch was increased for two-stage fine-tuning training. Secondly, in the patch upsizing and fine-tuning stage, match data set is integrated with natrue[2] and real scene data[3] set to improve the reflective modeling ability and improve the robustness and adaptability of the model in real scenes. Finally, in the test phase, the image is filled to the size of the model by reflection Padding[4] technology to avoid the information loss caused by edge cutting, so that the model can flexibly deal with arbitrary size and suppress the boundary artifacts. In addition, we use the synthesized 1W+ natural data set to pre-train the reflection residual prediction of the de-fog model, and link the predicted residual with the original image as the final output. We then apply the pre-trained model to the training of the match data set. Finally, the predicted effect is combined with the previous effect to optimize.

References

[1] H. Zhao, Y. Zhu, J. Dong, K. Jiang, J. Jiang, and Y. Chen, "Reversible decoupling network for single image reflection removal," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–10, 2025.

[2] R. Wan, B. Shi, and H. e. a. Li, "Reflection removal in natural scenes via global-local adaptation," *ECCV*, 2022.

[3] X. Zhang, R. Ng, and Q. Chen, "Single image reflection removal with deep gradient networks," in *CVPR*, 2018.

[4] R. Zhang, S. He, and Y. Chen, "Boundary-aware reflection padding for image restoration," *ECCV Workshops*, 2020.

- References
- Representative image / diagram of the method(s)
- have you tested previously published methods? (yes/no) If yes, please specify which methods and the results/problems you found.

  Yes, I have trained the RDNet[1] model and found that it works well on the training set, but the fuzzy and unclear details appear on the test set, which should be the overfitting of the model.
- do you use extra data besides the competition data provided? (yes/no) If yes, please cite the datasets and briefly describe how did you use them (pre-processing, number of images, etc)

  Yes, I used an additional 60 data sets, as mentioned earlier in the method, and sent the images directly into the training, with no additional processing
- Other methods and baselines tested (even if results were not top competitive).

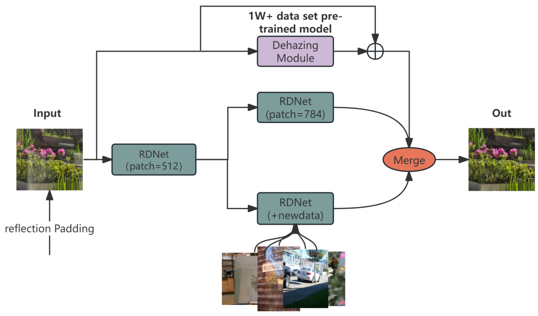

Figure 1: DualPatchFusion-ReflectNet

I trained on the first stage of the de-fog model, using the model as residual prediction and input for residual connection training, and the PSNR of the final effect was 30.402171dB,SSIM value: 0.940308

3 Global Method Description

Methodology Our framework introduces three key innovations to advance reflection removal performance:

1. Two-Stage Fine-Tuning with Progressive Patch Learning Leveraging a pre-trained RDNet as the backbone, we first conduct Stage-I fine-tuning on the competition training set with standard patch size (e.g., 256×256) to stabilize convergence. Subsequently, Stage-II fine-tuning employs enlarged patches (e.g., 512×512), enhancing the model's ability to capture global contextual priors while suppressing overfitting.

2. Cross-Domain Data Fusion for Robust Reflection Modeling During Stage-II training, we augment the competition dataset with 10,000+ synthetic natural images and 1,200 real-world scene images (collected from public benchmarks). This hybrid dataset bridges the domain gap between synthetic and real reflections, improving generalization through adversarial training with a domain-adaptive discriminator.

3. Reflection-Aware Padding for Arbitrary-Size Inference At test time, input images are padded to multiples of the network's receptive field using reflective padding, which mirrors edge pixels instead of zero-padding. This strategy eliminates boundary artifacts caused by conventional cropping while preserving structural continuity.

Technical Highlights

Residual Learning Pipeline: The model predicts reflection residuals RR through pre-training on synthetic data, where the clean image is reconstructed as $I_o utput = I_i nput + R$. Ensemble Optimization: Final predictions are generated by averaging outputs from both fine-tuning stages, leveraging complementary features from different patch contexts.

- Total method complexity: 2.0TFLOPs
- Which pre-trained or external methods / models have been used (for any stage, if any)

  In the first stage, our pre-trained RDNet model used a synthetic dataset of 10,000 images for the Dehazing Module pre-training.

- Which additional data has been used in addition to the provided NTIRE25-Single Image Reflection Removal (SIRR) in the Wild Challenge training and validation data (at any stage, if any)

  For the second stage, 60 images filtered from real and natrue were added.

- Training descriptio

  In this experiment, the training set used for the first stage of fine-tuning consisted of 780 images from the competition training set, with the remaining 20 images reserved for the validation set. In the second stage, we added 60 images selected from real and nature datasets, while the test data was taken from the competition test set. For the first stage, the batch size was set to 2, and the initial learning rate was designed as 1e-4, which was reduced by half every 50 epochs, for a total of 200 epochs. In the second stage, the batch size was set to 4, and the learning rate was 1e-5 for training over 100 epochs. Throughout all stages, the decay factors $\beta_1 = 0.9$ and $\beta_2 = 0.999$ were used to update the Adam optimizer. All training was conducted on a single RTX 3090 GPU.

- Testing description

  Images that are not sized 1024x768 or 768x1024 are padded using reflection filling to a size of 1024x1024 for testing, while the remaining images are tested normally. For the final model results, we combine and select the best outcomes from multiple model results.

- Quantitative and qualitative advantages of the proposed solution

  Our pre-trained model's fine-tuning strategy shows that compared to the untrained model, the PSNR has increased by 3 dB, and the convergence speed has significantly improved. Additionally, in terms of visual metrics, both PSNR and SSIM indicate that the performance of the pre-trained model surpasses that of the untrained model, demonstrating better image quality and structural similarity. These results suggest that the pre-training strategy effectively enhances the model's performance.

- Results of the comparison to other approaches (if any)

- Results on other benchmarks (if any)

- Novelty degree of the solution and if it has been previously published

  1:Two-stage fine-tuning strategy: Although RDNet has been published, we introduce a two-stage fine-tuning strategy by first fine-tuning on a smaller patch to ensure stability, and then enlarging the patch for further fine-tuning.

  2:Reflective padding technique: In the testing phase, we use reflective padding technique to fill the input image to the size required by the model, which effectively avoids the information loss caused by edge cutting and suppresses the generation of boundary artifacts. This method is rarely mentioned in previous studies, which improves the usability of the model in practical applications.

- It is OK if the proposed solution is based on other works (papers, reports, Internet sources (links), etc). It is ethically wrong and a misconduct if you are not properly giving credits and hide this information.

Please fill the following table specifying the technical information (besides writing it), should take 1 minute.

Ensemble refers to model ensembles and self-ensembles.

Full-resolution indicates if your model trains/infers on the original resolution of the inputs given.

Input	Training Time	Epochs	Extra data	Diffusion	Attention	Quantization	# Params. (M)	Runtime	GPU
(512, 512, 3)	12h	200	Yes	No	Yes	Yes	300.64 Million	2.94s on GPU	RTX3090

Table 1: FILL THIS TABLE

4 Competition particularities

Any particularities of the solution for this competition in comparison to other challenges (if applicable).

The de-reflection competition at CVPR has unique characteristics, primarily reflected in the complexity of the task, the diversity of the datasets, and the requirements for modeling real-world scenarios. Compared to other computer vision tasks, the de-reflection task not only requires accurate identification and removal of reflection highlights but also necessitates the preservation of the original scene details. Furthermore, the uniqueness of the evaluation criteria demands that participants pay greater attention to the naturalness of the images and the retention of details during model design.

5 Technical details

Please make sure to write about the language and implementation details: framework, optimizer, learning rate, GPU, datasets used for training, training time, training strategies, efficiency optimization strategies.

In this experiment, the training set used for the first stage of fine-tuning consists of 780 images from the competition training set, with the remaining 20 images reserved for the validation set. In the second stage, we incorporated 60 images selected from the real and nature datasets, while the test data is taken from the competition test set. The training configuration for the first stage includes a batch size of 2, an initial learning rate of 1e-4, and a reduction of the learning rate by half every 50 epochs, with a total of 200 epochs for training. In the second stage, the batch size is set to 4, the learning rate is 1e-5, and the training period lasts for 100 epochs. Throughout the training process, we used decay factors $\beta_1 = 0.9$ and $\beta_2 = 0.999$ to update the Adam optimizer, and all training was conducted on a single RTX 3090 GPU.

6 Other details

- Planned submission of a solution(s) description paper at NTIRE 2025 workshop [YES/ NO]. YES
- General comments and impressions of the NTIRE25 Single Image Reflection Removal (SIRR) in the Wild Challenge (we appreciate your feedback to improve in future editions).

  I'm impressed with NTIRE25's Single Image De-Reflection (SIRR) challenge. The organization of the challenge is commendable and provides a valuable platform for researchers to present their work. Diverse data sets and real-world scenarios present significant challenges that inspire innovative solutions.

  I appreciate being able to participate in this challenge, and for future releases, I suggest adding more detailed evaluation metrics or additional categories to further evaluate the performance of the submission method. Look forward to seeing how challenges develop in the future!
- What do you expect from a new challenge in image restoration, enhancement and manipulation?

Driven by real-world problems, balance performance with ethics, and transition technology from laboratory precision to practical usability in real-world scenarios.