NTIRE 2025 Image Denoising ($\sigma = 50$) Challenge Factsheet
Deep ensemble for Image denoising

Donghun Ryou[2] Inju Ha[1] Bohyung Han[1,2]
Computer Vision Laboratory, [1]ECE & [2]IPAI, Seoul National University
{dhryou, hij1112, bhhan}@snu.ac.kr

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 denoising with noise level $\sigma = 50$.

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 decided. 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 (Lei Sun, Yawei Li, and Radu Timofte)** by email.

2. Email final submission guide

To: yawei.li@vision.ee.ethz.ch
yulun100@gmail.com
timofte.radu@gmail.com
cc: your_team_members
Title: NTIRE 2025 Image Denoising 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 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, pretrained 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 to compare 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 noisy and denoised images in `data` folder to save the bandwidth.

4. Factsheet Information

4.1. Team details

- **Team name**
  SNUCV

- **Team leader name**
  Donghun Ryou

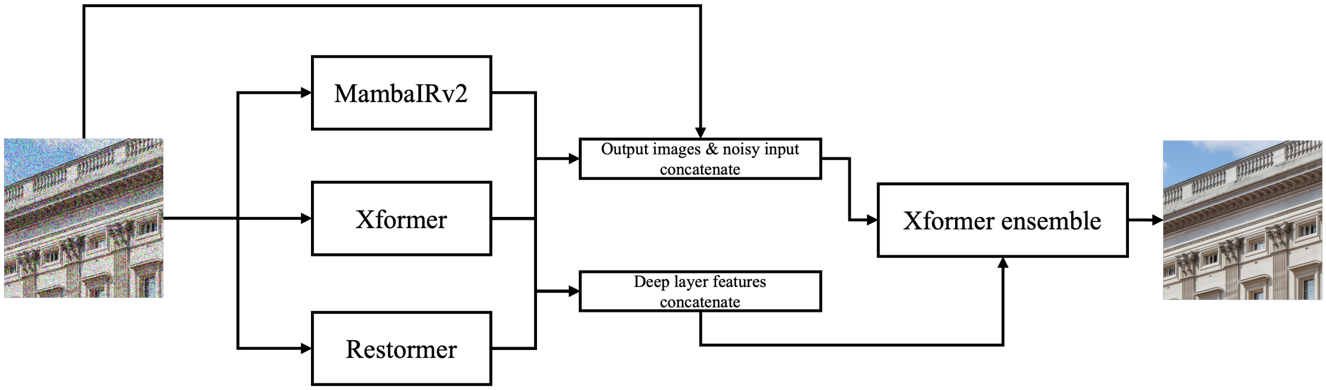

Figure 1. The architecture of our pipeline.

- **Team leader address, phone number, and email**
  Address: 70-15, Gwanak-ro 12-gil, Gwanak-gu, Seoul, Republic of Korea
  Phone number: +82 010-4879-3794
  Email: dhryou@snu.ac.kr

- **Rest of the team members**
  Inju Ha, Bohyung Han

- **Team website URL (if any)**
  https://cv.snu.ac.kr/

- **Affiliation**
  Seoul National University

- **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)**
  dhryou

- **Best scoring entries of the team during the development/validation phase**

- **Link to the codes/executables of the solution(s)** https://github.com/dhryougit/NTIRE2025_Dn50_SNUCV.git

4.2. Method details

- **General method description (How is the network designed.)**
  We endeavored to maximally leverage well-designed denoising models to enhance performance. We devised a two-stage training framework. In the first stage, we performed fine-tuning of existing models. In the subsequent stage, we developed an ensemble model that integrates the outputs of these models along with features derived from their deep layers for further training. The following section provides details of the network architectures and training methodology employed.
  As shown in Figure 1, the network architecture we utilized consists of MambaIRv2 [2], Xformer [6], and Restormer [5]. These networks are first pretrained on Gaussian noise with a standard deviation of 50. Subsequently, the outputs of these networks are concatenated with the noisy image, which is then used as input to the ensemble model. In addition to the outputs, the features from the deepest layers of these networks are also concatenated and integrated into the deepest layer features of the ensemble network. This approach ensures that the feature information from the previous networks is preserved and effectively transferred to the ensemble network without loss. The ensemble model is designed based on Xformer, accepting an input with 12 channels. Its deepest layer is structured to incorporate the concatenated features from the previous models. These concatenated features are then processed through a 1×1 convolution to reduce the channel dimension back to that of the original network, thereby alleviating the computational burden. Additionally, while Xformer and Restormer reduce the feature size in their deep layer, MambaIRv2 retains its original feature size without reduction. To align the sizes for concatenation, the features of MambaIRv2 were downscaled by a factor of 8 before being concatenated.

- **Training details**
  We first train the denoising network, and subsequently, we incorporate the frozen denoising network to train the ensemble model. Both the denoising models and the ensemble model were trained exclusively using the DIV2K [1] and LSDIR [3] datasets. Training was performed using the AdamW [4] optimizer with hyperparameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and a learning rate of 3×10^{-4}. All models were trained for a total of 300,000 iterations.

For the denoising models, Restormer [5] and Xformer [6] were trained using a progressive training strategy to enhance robustness and efficiency. The patch sizes were progressively increased as $[128, 160, 192, 256, 320, 384]$, with corresponding batch sizes of $[8, 5, 4, 2, 1, 1]$. In contrast, MambaIRv2 [2] was trained with a more constrained setup due to GPU memory limitations, utilizing patch sizes of $[128, 160]$ and batch sizes of $[2, 1]$.
The ensemble model was trained with a progressive patch size schedule of $[160, 192, 256, 320, 384, 448]$ and corresponding batch sizes of $[8, 5, 4, 2, 1, 1]$.
The denoising models were trained using L1 loss, while the ensemble model was trained using a combination of L1 loss, MSE loss, and high-frequency loss.

- **Inference details**
During the final inference stage to derive test results, we utilized a self-ensemble technique. Furthermore, inference was conducted using a patch-based sliding window approach. The patch sizes were set to [256, 384, 512], with corresponding overlap values of [48, 64, 96]. The resulting outputs were subsequently averaged to optimize performance

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2017. 2

[2] Hang Guo, Yong Guo, Yaohua Zha, Yulun Zhang, Wenbo Li, Tao Dai, Shu-Tao Xia, and Yawei Li. Mambairv2: Attentive state space restoration. *arXiv preprint arXiv:2411.15269*, 2024. 2, 3

[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. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1775–1787, 2023. 2

[4] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 2

[5] 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. 2, 3

[6] Jiale Zhang, Yulun Zhang, Jinjin Gu, Jiahua Dong, Linghe Kong, and Xiaokang Yang. Xformer: Hybrid x-shaped transformer for image denoising. *arXiv preprint arXiv:2303.06440*, 2023. 2, 3