NTIRE 2025 Real-World Face Restoration Challeng Factsheet
A Face Image Restoration Method Applying Pre-trained Models and Test-time Adaptation

Wenjie An
Harbin University of Technology
92 Xidazhi Street, Nangang District, Harbin City, Heilongjiang Province
anwenjie1213@163.com

Kangmeng Yu
Harbin University of Technology
92 Xidazhi Street, Nangang District, Harbin City, Heilongjiang Province
2912422099@qq.com

1. Code Submission

The command to download our code:`git clone https://github.com/continue1213/02_faceRestoration.git`

2. Factsheet Information

2.1. Team details

- Team name:AIIA
- Team leader name: An Wenjie
- Team leader address, phone number, and email :92 Xidazhi Street, Nangang District, Harbin City, Heilongjiang Province, Harbin University of Technology; 15698236263; anwenjie1213@163.com
- Rest of the team members:Yu Kangmeng
- Team website URL (if any)
- Affiliation:Harbin University of Technology
- Affiliation of the team and/or team members with NTIRE 2025 sponsors (check the workshop website): No
- User names and entries on the NTIRE 2025 Codalab competitions (development/validation and testing phases):echooo,Anwenjie,Burningmuses
- Best scoring entries of the team during the testing phase: In the end, our NIQE value on all pictures is 5.186, CLIP_IQA value is 0.929, MANIQA value is 0.730, MUSIQ value is 77.481, QALIGN value is 4.646, FID value is 61.315, ID_sim value is 0.704, low_ID_Sim value is 5. Based on the final score calculation formula provided by the organizers, the score is 4.231.
- Link to the codes/executables of the solution(s):

2.2. Method details

We adopt a three-stage face restoration pipeline that first applies GFPGAN for coarse restoration, then uses TSD-SR for high-quality enhancement, and finally incorporates ZSSR for fine-tuning the restoration process. This enhanced pipeline aims to improve the face image quality by sequentially leveraging the strengths of a GAN-based face prior, a diffusion-based super-resolution model, and a zero-shot super-resolution approach that performs fine-tuning on a per-image basis.

Stage 1: GFPGAN for Coarse Restoration: GFPGAN [5] serves as the first stage of our pipeline, which performs coarse restoration of facial structures. It utilizes a generative facial prior network to recover identity, structure, and base details from a low-quality input image. This helps reconstruct the face's essential features, even when large degradations are present.

Stage 2: TSD-SR for High-Quality Enhancement: TSD-SR [1] is employed in the second stage to enhance the fine details of the face image. This model uses a diffusion-based super-resolution approach to inject realistic textures, refine facial details, and improve the overall quality of the restored image. By utilizing the powerful diffusion model, we can enhance the appearance of fine structures, such as wrinkles, skin textures, and other subtle features, which are difficult to recover using traditional methods.

Stage 3: ZSSR for Image-Specific Fine-Tuning: In the third stage, we apply the ZSSR (Zero-Shot Super-Resolution) approach [3] to perform image-specific fine-tuning. ZSSR is a powerful technique that operates on each image individually, allowing us to refine the restoration further based on the unique characteristics of the image.

This approach performs super-resolution without additional training on external datasets, utilizing the specific features of the input image to guide the restoration process.This step of the ZSSR method is optional. If you want to quickly generate the restored image, you can adopt the strategy of Stage 1 and Stage 2 only, which is also excellent.

Loss Function: To improve the perceptual quality of the restoration, we introduce the CLIP-IQA [4] loss function. This perceptual loss is designed to optimize the alignment of the restored image with human subjective preferences. CLIP-IQA uses CLIP embeddings to measure perceptual similarity between the ground truth image and the restored image.

The overall objective function $\mathcal{L}$ is defined as:

$$\mathcal{L} = \mathcal{L}_{\text{CLIP-IQA}}(I, \hat{I})$$

Where I is the ground truth high-quality image, and $\hat{I}$ is the predicted restored image. $\mathcal{L}_{\text{CLIP-IQA}}(I, \hat{I})$ is the CLIP-IQA perceptual loss, focusing on perceptual quality.

Pipeline Overview:

- GFPGAN [5] performs coarse restoration, recovering facial identity and removing common degradations.
- TSD-SR [1] refines details, adding realistic textures and fine structures.
- ZSSR [3] fine-tunes each image individually, further enhancing details specific to the input image.
- CLIP-IQA [4] is used as the perceptual loss function to guide the restoration and ensure high perceptual quality in the final result.

Training Strategy:

- GFPGAN uses official pretrained weights, without any additional fine-tuning, leveraging the strength of its generative facial prior network.
- TSD-SR was trained on diverse datasets (FFHQ, DIV2K, Flickr2K, LSDIR) using realistic image degradations. LoRA adapters were applied for efficient training and inference, guided by reconstruction loss, knowledge distillation, and ground-truth supervision. We use the official pretrained model for inference without further fine-tuning.
- ZSSR is applied in a zero-shot manner, meaning that it performs per-image enhancement directly, without the need for training on external datasets. This allows it to adapt to the specific characteristics of the input image, providing fine-tuning and high-quality restoration tailored to each case.

Testing Description:

- The proposed pipeline is tested sequentially. Low-quality face images are first processed by GFPGAN to generate an intermediate restored image. This intermediate result is then directly passed to TSD-SR for further detail enhancement, followed by application of ZSSR for fine-tuning the restoration based on each image's specific features. Finally, a wavelet-based color correction is applied to ensure consistency and realism in the final image.The third step of the ZSSR method is optional. If you want to quickly generate the restored image, you can adopt the strategy of Stage 1 and Stage 2 only, which is also excellent.

Qualitative and Quantitative Advantages:

- Our three-stage method effectively handles identity preservation, detail realism, and image-specific fine-tuning simultaneously, which is a challenge for traditional single-stage or two-stage methods. Qualitative evaluation shows that our approach produces visually appealing, high-fidelity facial textures that are more consistent with real-world images.
- Quantitative results demonstrate significant improvements in common evaluation metrics such as PSNR, SSIM, and perceptual quality scores, compared to standalone approaches. Our method shows robust performance in both restoration accuracy and perceptual realism, outperforming single-model methods in terms of both image fidelity and visual appeal.

Novelty and Reference to Existing Works:

- While our method builds upon existing models such as GFPGAN and TSD-SR, the integration of ZSSR into a sequential pipeline represents a novel extension of these models. Specifically, the introduction of per-image fine-tuning using ZSSR improves restoration performance for each input image, leveraging the unique features of each face. This three-stage approach of coarse restoration, detail enhancement, and per-image fine-tuning is unique in its design and its ability to balance fidelity and perceptual quality.
- Our pipeline is inspired by state-of-the-art strategies such as the DiffBIR [2] pipeline, which demonstrated excellent results in blind face restoration with a two-stage approach. Our method adds an additional layer of fine-tuning, which significantly enhances the final restoration quality. Proper citations are provided to all used models and techniques, ensuring transparent acknowledgment of prior work.

References

[1] Linwei Dong, Qingnan Fan, Yihong Guo, Zhonghao Wang, Qi Zhang, Jinwei Chen, Yawei Luo, and Changqing Zou. Tsd-sr: One-step diffusion with target score distillation for real-world image super-resolution. *arXiv preprint arXiv:2411.18263*, 2024. 1, 2

[2] Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Yu Qiao, Wanli Ouyang, and Chao Dong. Diffbir: Toward blind image restoration with generative diffusion

prior. In *European Conference on Computer Vision*, pages 430–448. Springer, 2024. 2

[3] Assaf Shocher, Nadav Cohen, and Michal Irani. "zero-shot" super-resolution using deep internal learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3118–3126, 2018. 1, 2

[4] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI conference on artificial intelligence*, pages 2555–2563, 2023. 2

[5] Xintao Wang, Yu Li, Honglun Zhang, and Ying Shan. Towards real-world blind face restoration with generative facial prior. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9168–9178, 2021. 1, 2