

Supplementary Materials: LD-BFR: Vector-Quantization-Based Face Restoration Model with Latent Diffusion Enhancement

Anonymous Authors

A OVERVIEW

This supplementary material includes:

- more details of the training process (Section B) ;
- more ablation studies (Section C);
- user study on real-world dataset (Section D);
- more test results on CelebA dataset (Section E).

Apart from this document, we have supplemented the *source code* of our LD-BFR to provide more implementation details.

B MORE DETAILS OF OUR LD-BFR

B.1 More Training Details

Due to space constraints, we introduce the settings of the training dataset and some training hyper-parameters in the Experiment Section. This section will show more training details of the VQ-GAN and the diffusion model.

At first, we follow the settings of the LDM [2] and train a VQ-GAN model and a latent diffusion model on a high-quality face dataset.

Secondly, we individually adjust the Encoder and the Decoder of the VQ-GAN. For the Encoder, we add a Dual Cross-Attention module after the original Encoder and finetune it with an HQ-LQ paired dataset. It should be noted that the parameters of the codebook and the Decoder are frozen when training the Encoder.

Finally, we train the HQI module of the diffusion model. The other modules of the diffusion model reuse the parameters of the pre-trained model and their parameters are frozen while training the HQI module. The HQ feature input of the HQI module is from the output of the LQ-Encoder with the cross-attention module. The pair feature for training is extracted from the HQ face image from the HQ-Encoder.

C MORE ABLATION STUDIES

In Section 4.2 of the main paper, we qualitatively study the effectiveness of the dual cross-attention module, the diffusion process, and the HQI module. Here, we present quantitative study of these modules.

For convenience, we show the quantitative ablation in one table. As is shown in Tab. 1, the DCAM is the Dual Cross-Attention module, the DQE represents the Diffusion Quality Enhancement module, and the conv stands for using the conv method to inject condition. We only use the FID metric and experiment on the CelebA-Test dataset.

D USER STUDY ON REAL-WORLD DATASET

In this Section, we conduct a user study in which users sort the results of three methods, *i.e.*, two state-of-the-art competitors (CodeFormer [6], DR2 [4]) and our method. In total 10 groups of images are evaluated in this user study. Note that all these 10 groups are randomly chosen from the test set. Our method outperforms the

Table 1: Quantitative results of ablation studies with FID↓ score.

DCAM	DQE	Conv	E-HQI	D-HQI	FID↓
					66.1826
✓					54.3621
✓	✓	✓			48.4688
✓	✓		✓		44.4849
✓	✓			✓	42.8421
✓	✓		✓	✓	38.9823

User Study

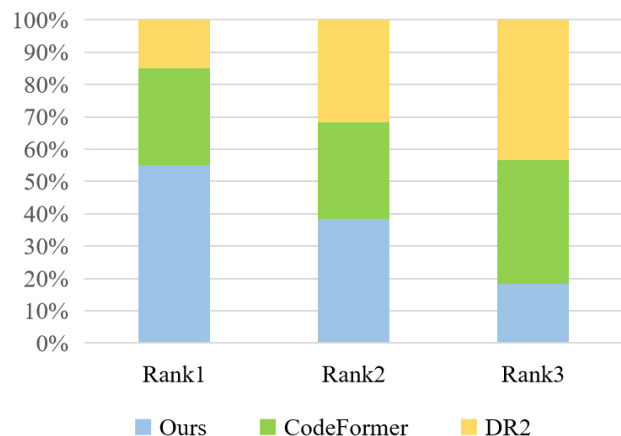


Figure 1: User study results of CodeFormer, DR2 and our method.

other comparison methods in most groups in terms of restoration quality.

We have invited 12 participants to complete our user study, and the results of the user study are summarized in Fig. 1. Results demonstrate the advantage of our method, where 55% votes chose our method as the best among the three methods.

E MORE TEST RESULTS ON CELEBA DATASET

For more comprehensive comparisons with previous methods [1, 3–6, 6] on different levels of degraded datasets, we provide qualitative results on each split of the CelebA-Test dataset in Figures 2, 3 and results on real-world dataset in Figure 4.



Figure 2: More results on the synthetic CelebA-Test dataset.

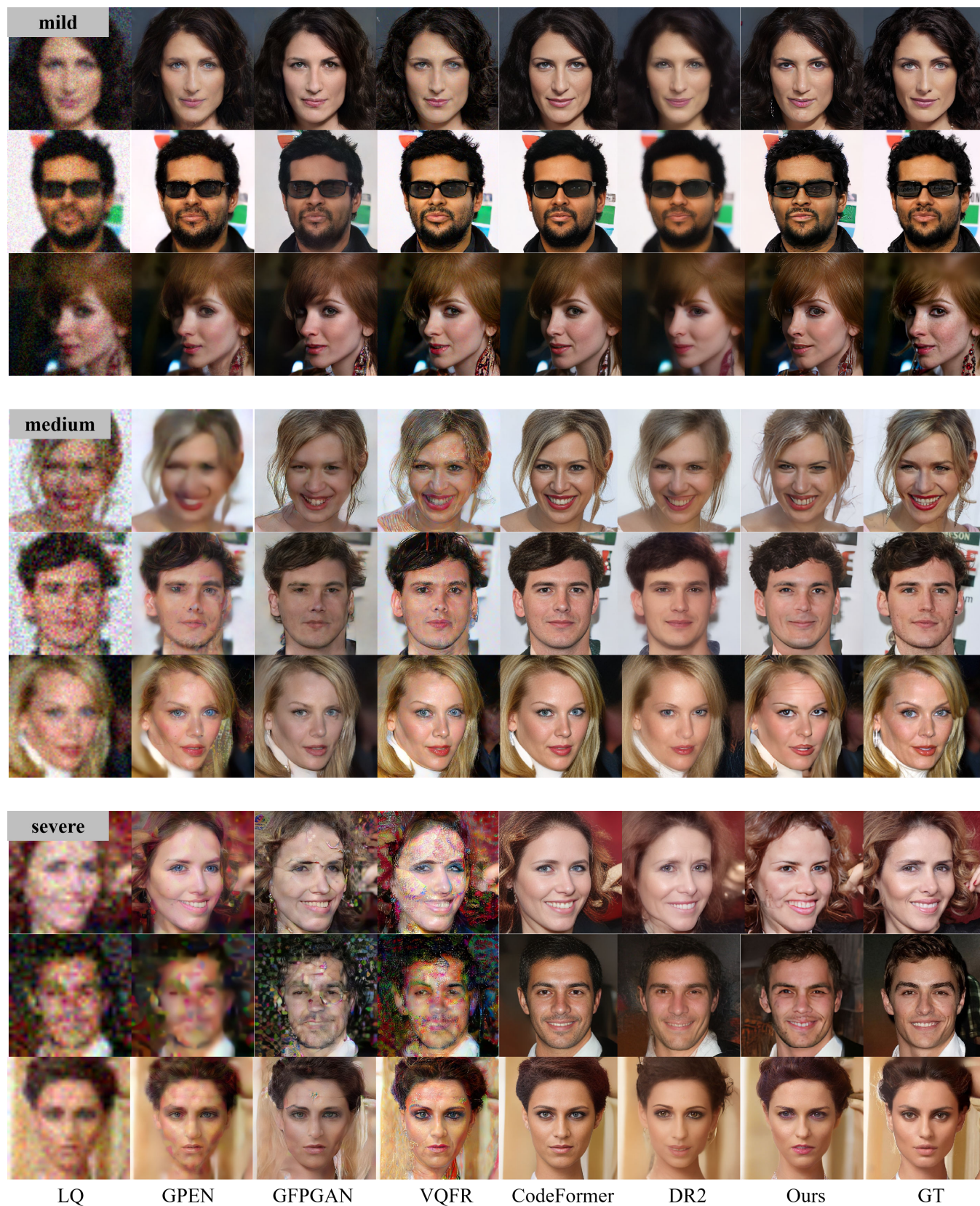


Figure 3: More results on the synthetic CelebA-Test dataset.

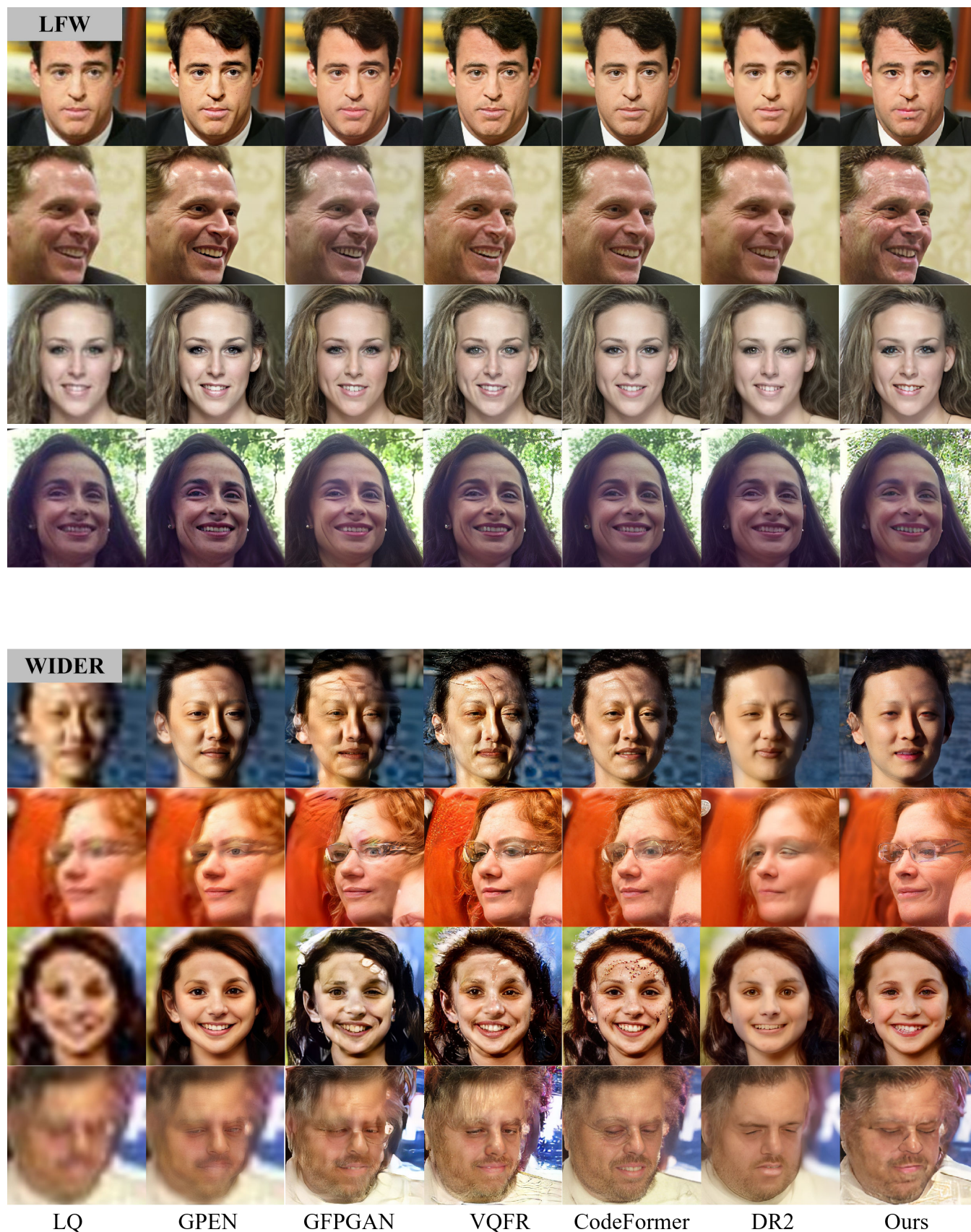


Figure 4: More results on the real-world dataset.

REFERENCES

[1] Yuchao Gu, Xintao Wang, Liangbin Xie, Chao Dong, Gen Li, Ying Shan, and Ming-Ming Cheng. 2022. Vqfr: Blind face restoration with vector-quantized dictionary and parallel decoder. In *ECCV*. Springer, 126–143.

[2] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10684–10695.

[3] Xintao Wang, Yu Li, Honglun Zhang, and Ying Shan. 2021. Towards real-world blind face restoration with generative facial prior. In *CVPR*. 9168–9178.

[4] Zhixin Wang, Xiaoyun Zhang, Ziyang Zhang, Huangjie Zheng, Mingyuan Zhou, Ya Zhang, and Yanfeng Wang. 2023. DR2: Diffusion-based Robust Degradation Remover for Blind Face Restoration. *arXiv preprint arXiv:2303.06885* (2023).

[5] Tao Yang, Peiran Ren, Xuansong Xie, and Lei Zhang. 2021. Gan prior embedded network for blind face restoration in the wild. In *CVPR*. 672–681.

[6] Shangchen Zhou, Kelvin Chan, Chongyi Li, and Chen Change Loy. 2022. Towards robust blind face restoration with codebook lookup transformer. *NIPS* 35 (2022), 30599–30611.