

# 3D Priors-Guided Diffusion for Blind Face Restoration

## Supplementary Material

### 1 Overview

In this supplementary material, we present more qualitative analysis results with state-of-the-art algorithms on the CelebA-Test [Karras *et al.*, 2017], LFW-test [Huang *et al.*, 2008], WIDER-Test [Yang *et al.*, 2016], and WebPhoto dataset [Wang *et al.*, 2021]. Furthermore, we exhibit more visualization results of 3D reconstruction. Additionally, The detailed network architecture of the third section of our main paper is illustrated in Table 1.

#### 1.1 Qualitative analysis on the synthetic datasets.

The performance of various state-of-the-art methods on synthetic datasets is displayed in Fig. 1. Owing to the significant degradation of the synthetic data, GFPGAN [Wang *et al.*, 2021], PSFRGAN [Chen *et al.*, 2021], and DR2 [Wang *et al.*, 2023] fail to achieve satisfactory restoration results. Likewise, other methods exhibit deficiencies in fidelity, detail retention, and clarity, particularly in regions such as the eyes, mouth, and teeth. Our method performs best in both fidelity and quality. This robust performance underscores the efficacy of our approach in addressing the challenges posed by deteriorated synthetic data.

#### 1.2 Qualitative analysis on the real-world datasets.

Our approach better ensures facial identity preservation by incorporating 3D facial prior information into the diffusion model. Such as the first and second rows in Fig. 2 and Fig. 4, the photos will have reflections due to the shooting angle. Other methods cannot filter out such artifacts, but our method can recover faces better. As shown in Fig. 3, for facial images with severe degradation in the real world, other methods cannot well restore high-frequency areas of the face, such as the eyes, and may cause over-smoothing or artifacts. After performing 3D facial reconstruction, we can filter out artifacts that are not present on the face. Integrating 3D priors into the recovery network can effectively eliminate these artifacts. In contrast, our method utilizes 3D priors to effectively filter out such misleading information, resulting in more accurate and faithful face recovery.

#### 1.3 More visual results on the 3D face image rendered by our method.

We enumerate more results of 3D facial reconstruction in Fig. 5. The results demonstrate that our 3D facial reconstruc-

| Method          | Details                | Value                   |
|-----------------|------------------------|-------------------------|
| Diffusion model | input size             | $512 \times 512$        |
|                 | output size            | $512 \times 512$        |
|                 | time step              | 100                     |
|                 | loss type              | L1                      |
|                 | learned sigma          | True                    |
| UNet            | in channel             | 3                       |
|                 | out channel            | 6                       |
|                 | model channel          | 32                      |
|                 | attention resolutions  | [32, 16, 8]             |
|                 | TAFB channel           | [32, 64, 128, 256]      |
|                 | channel multiplier     | [1, 2, 4, 8, 8, 16, 16] |
|                 | the number of Resblock | [1, 2, 2, 2, 2, 3, 4]   |

Table 1: The details of our network architecture.

tion method effectively captures facial contours, expressions, identity features, and skin color with high fidelity. Even in the case of profile faces, our method excels in reconstructing 3D facial priors. Through the incorporation of 3D facial priors, our approach significantly enhances the restoration and reconstruction of faces in scenarios involving severe degradation.

#### 1.4 The details of our network architecture.

In Tab. 1, we present the specific details of our network structure, where the input and output resolutions are both  $512 \times 512$ . During the inference phase, the denoising iteration is set to 100 steps. The U-Net has an output channel number of 6, with the first three channels representing learned noise mean values, and the latter three channels representing learned noise standard deviations. The model channel denotes the fundamental number of channels in intermediate layers, where the specific layer’s channel count is determined by multiplying the base number of channels by the value listed in the channel multiplier. The TAFB channel indicates that the Time-Aware Fusion Block (TAFB) module for feature fusion is only incorporated when the number of input feature channels matches the specified value. Our approach comprises 180.51 million parameters (equivalent to 308.415 GFlops) and requires 6.687 seconds for processing a  $512 \times 512$  image on Nvidia A100.

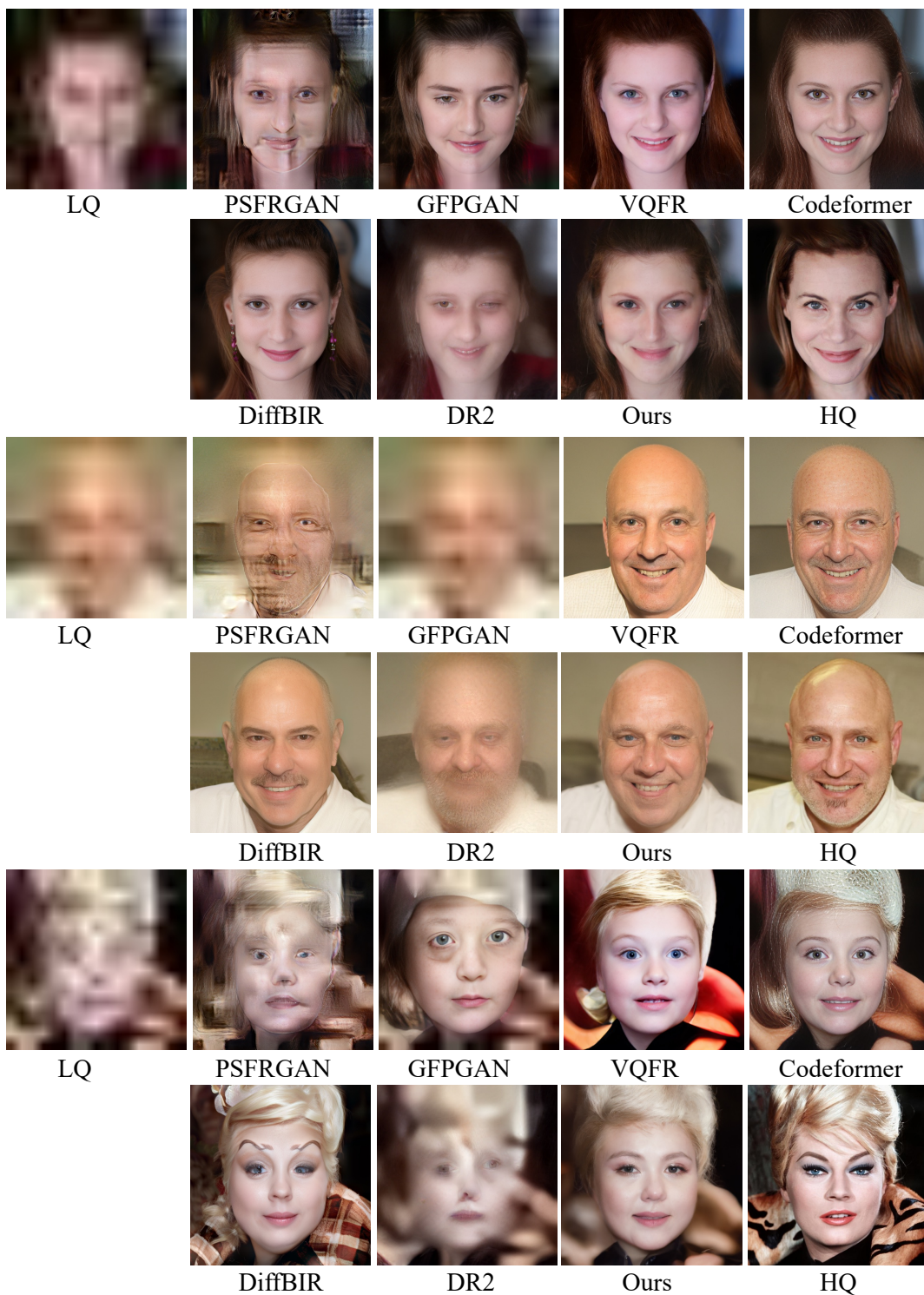


Figure 1: Qualitative comparisons of our methods with state-of-the-art methods PSFRGAN [Chen *et al.*, 2021], GFPGAN [Wang *et al.*, 2021], VQFR [Gu *et al.*, 2022], CodeFormer [Zhou *et al.*, 2022], DiffBIR [Lin *et al.*, 2023], and DR2 [Wang *et al.*, 2023] on the CelebA-Test [Karras *et al.*, 2017] dataset.

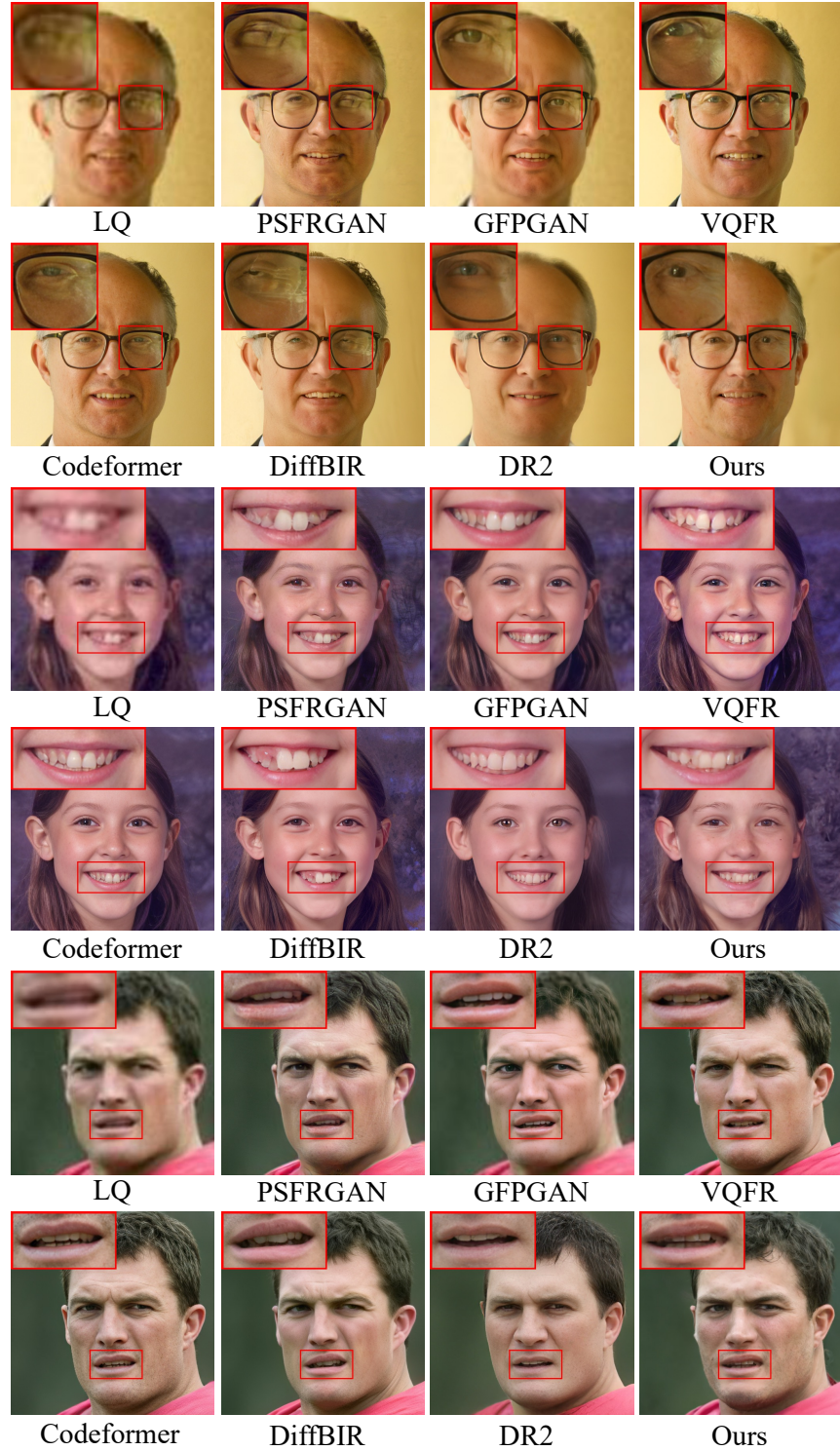


Figure 2: Qualitative comparisons of our methods with state-of-the-art methods PSFRGAN [Chen *et al.*, 2021], GFPGAN [Wang *et al.*, 2021], VQFR [Gu *et al.*, 2022], CodeFormer [Zhou *et al.*, 2022], DiffBIR [Lin *et al.*, 2023], and DR2 [Wang *et al.*, 2023] on the LFW-test [Huang *et al.*, 2008] dataset.





Figure 3: Qualitative comparisons of our methods with state-of-the-art methods PSFRGAN [Chen *et al.*, 2021], GFPGAN [Wang *et al.*, 2021], VQFR [Gu *et al.*, 2022], CodeFormer [Zhou *et al.*, 2022], DiffBIR [Lin *et al.*, 2023], and DR2 [Wang *et al.*, 2023] on the WIDER-Test [Yang *et al.*, 2016] dataset.

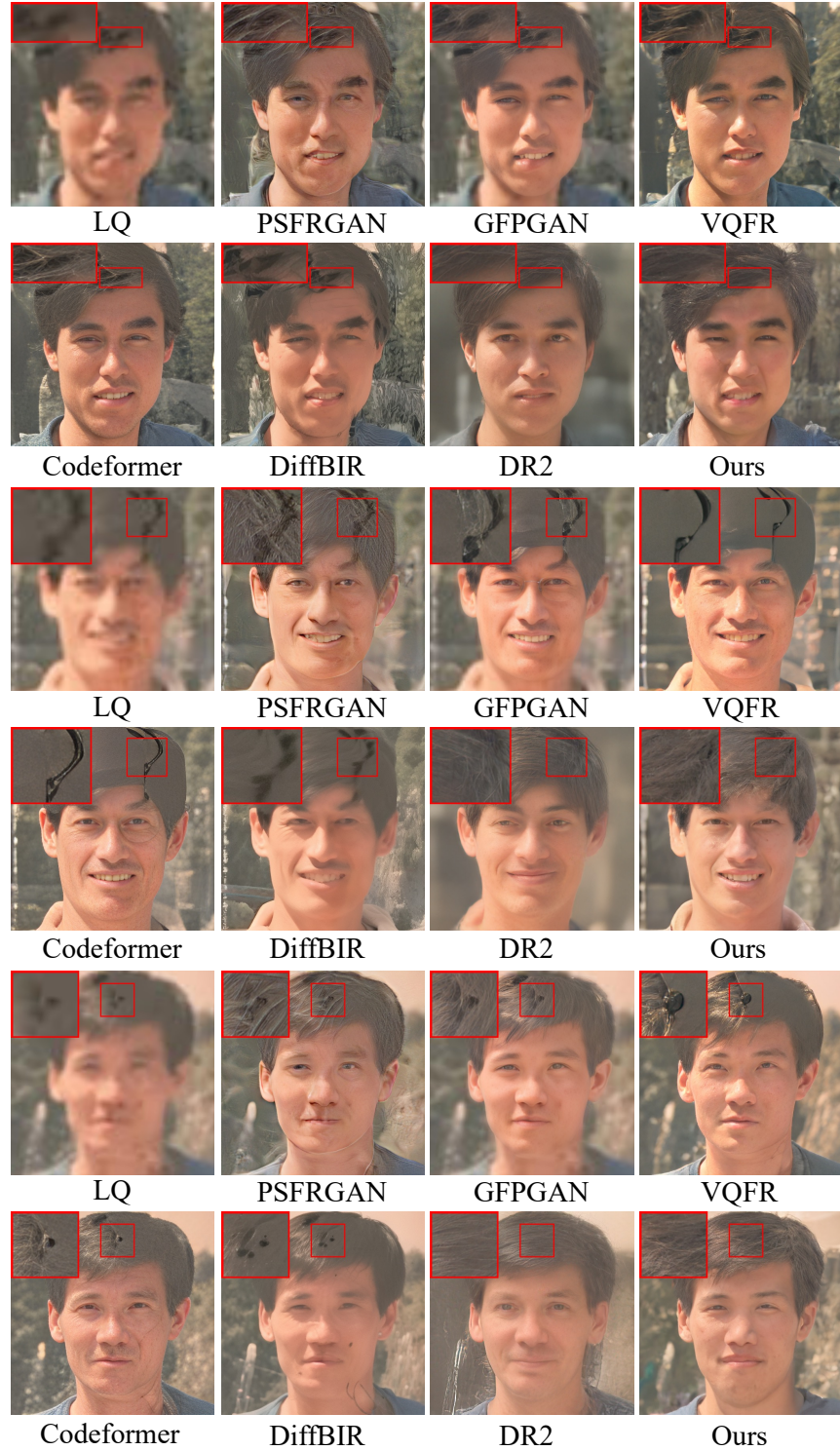


Figure 4: Qualitative comparisons of our methods with state-of-the-art methods PSFRGAN [Chen *et al.*, 2021], GFPGAN [Wang *et al.*, 2021], VQFR [Gu *et al.*, 2022], CodeFormer [Zhou *et al.*, 2022], DiffBIR [Lin *et al.*, 2023], and DR2 [Wang *et al.*, 2023] on the WebPhoto [Wang *et al.*, 2021] dataset.



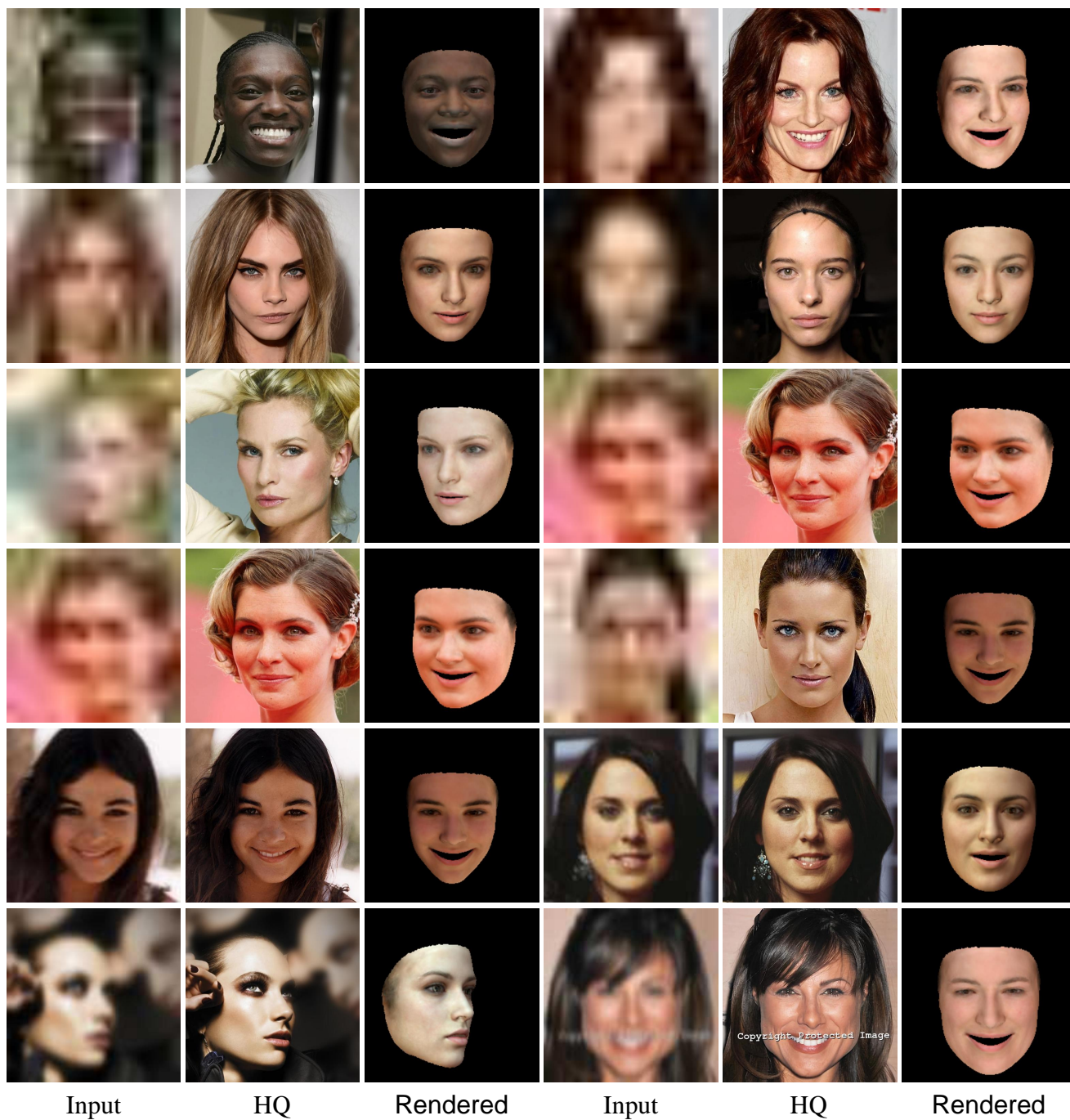


Figure 5: The 3D face image rendered by our method. The 3D face reconstructed by our method can better provide facial structure information and identity information.

## References

- [Chen *et al.*, 2021] Chaofeng Chen, Xiaoming Li, Lingbo Yang, Xianhui Lin, Lei Zhang, and Kwan-Yee K Wong. Progressive semantic-aware style transformation for blind face restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11896–11905, 2021.
- [Gu *et al.*, 2022] Yuchao Gu, Xintao Wang, Liangbin Xie, Chao Dong, Gen Li, Ying Shan, and Ming-Ming Cheng. Vqfr: Blind face restoration with vector-quantized dictionary and parallel decoder. In *European Conference on Computer Vision*, pages 126–143. Springer, 2022.
- [Huang *et al.*, 2008] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In *Workshop on faces in 'Real-Life' Images: detection, alignment, and recognition*, 2008.
- [Karras *et al.*, 2017] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.
- [Lin *et al.*, 2023] Xinqi Lin, Jingwen He, Ziyang Chen, Zhaoyang Lyu, Ben Fei, Bo Dai, Wanli Ouyang, Yu Qiao, and Chao Dong. Diffbir: Towards blind image restoration with generative diffusion prior. *arXiv preprint arXiv:2308.15070*, 2023.
- [Wang *et al.*, 2021] Xintao Wang, Yu Li, Honglun Zhang, and Ying Shan. Towards real-world blind face restoration with generative facial prior. In *CVPR*, pages 9168–9178, 2021.
- [Wang *et al.*, 2023] Zhixin Wang, Ziyang Zhang, Xiaoyun Zhang, Huangjie Zheng, Mingyuan Zhou, Ya Zhang, and Yanfeng Wang. Dr2: Diffusion-based robust degradation remover for blind face restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1704–1713, 2023.
- [Yang *et al.*, 2016] Shuo Yang, Ping Luo, Chen-Change Loy, and Xiaoou Tang. Wider face: A face detection benchmark. In *CVPR*, pages 5525–5533, 2016.
- [Zhou *et al.*, 2022] Shangchen Zhou, Kelvin Chan, Chongyi Li, and Chen Change Loy. Towards robust blind face restoration with codebook lookup transformer. *Advances in Neural Information Processing Systems*, 35:30599–30611, 2022.