ENHANCED FACIAL RESTORATION WITH MISINFORMATION-FILTERED GUIDE-DENOISING DIFFUSION PROBABILISTIC MODELS

Wendi Liang[⋆†] *Yihan Wen*[⋆¶] *Zewei Wang*[⋆‡] *Jianuo Jiang* [†]
Tat-Ming Lok[‡] Guanchong Niu[§]

[†] The Chinese University of Hong Kong, Shenzhen [‡] The Chinese University of Hong Kong
[¶] The Hong Kong Polytechnic University [§] Xidian University

ABSTRACT

Most of the existing generation models encounter notable challenges in complex scenes, particularly with inaccuracies in facial organs and textures that do not align with actual conditions. Traditional face restoration methods, heavily dependent on facial geometry and reference priors, often generate incorrect facial images that contribute misleading prior information. In this study, a Misinformation-Flitered Guide-Denoising Diffusion Probabilistic Models (MF-GDDPM) is proposed to address these issues. Specifically, MF-GDDPM employs low-pass filtering to remove high-frequency details that contain misleading prior information. This process results in filtered low-dimensional facial contours that guide the diffusion model in generating high-quality facial images. To further enhance the fidelity of the generated results, a dual-stream encoder within the Denoising Unet is constructed to process facial contours and high-dimensional details separately, while the Attention Feature Fusion (AFF) attention mechanism ensures the fidelity of image restoration. We have also incorporated the Natural Image Quality Evaluator (NIQE), a deep learning-based image quality assessment tool, into our framework as a novel loss function to crucially ensure the naturalness of restored images. Overall, the proposed method marks a significant improvement in generating accurate and clear facial images using diffusion models.

Index Terms— Denoising Diffusion Probabilistic Model, Face Restoration, Generative Models

1. INTRODUCTION

The swift advancement in Generative Artificial Intelligence (GAI) has notably heightened interest in generating realistic human faces using generative models [1]. This attention is driven by the potential applications of GAI in various domains, including digital entertainment, and virtual reality. Despite these advancements, crafting accurate and lifelike facial images remains a formidable challenge. This difficulty is especially pronounced in intricate scenarios, where text-to-image generative models [2, 3, 4] often struggle with capturing the nuanced details of facial features and textures. Such limitations considerably impair the practical utility of GAI models, rendering the quality of the produced facial images inadequate for real-world deployment. Consequently, restoring the generated facial images and making them more natural hold considerable significance.

Traditional blind face restoration methods [5], while effective for damaged real photos, are limited by their reliance on facial geometry and reference priors, often leading to inaccuracies when processing generated faces that contain numerous incorrect textures and geometric features as misleading information. In response to these limitations, a groundbreaking approach, named Misinformation-Filtered Guide-Denoising Diffusion Probabilistic Models (MF-GDDPM), is designed to eliminate misleading prior information and reconstruct refined facial images with enhanced accuracy in this work.

Distinct from conventional strategies, the proposed method concentrates on the elimination of misleading high-frequency details through low-pass filtering [6]. The core breaktrhough involves distilling facial contours into a simpler, lower-dimensional state, which serves as an essential guide for the diffusion model. The process not only refines the facial contours but also aids the model in generating clearer and higher-quality facial images. Moreover, we employ an Attention Feature Fusion (AFF) method to integrate prior features within the original image feature and facial contour feature within a low-dimension facial image. This module effectively prevent the excessive loss of original features, ensuring the fidelity of the results from the proposed restoration method. Furthermore, we integrate a dual-stream encoder within the Denoising Unet architecture. This encoder processes facial contours and high-dimensional details independently, ensuring a more faithful restoration of facial images. Moreover, the integration of the Natural Image Quality Evaluator (NIQE) module [7], a cutting-edge tool in image quality evaluation, plays a pivotal role in preserving the naturalness of the generated images. This blend of advanced methodologies and

⋆ Authors contributed equally to this work.
Corresponding authors: Guanchong Niu (niuguanchong@xidian.edu.cn); Tat-Ming Lok (tmlok@ie.cuhk.edu.hk)

tools positions our approach as a significant step forward in the field of facial image generation using diffusion models, marking a leap in our capability to produce highly realistic and accurate facial representations. This model is designed for recovering and enhancing human facial images and is rooted in a series of innovative strategies. Our contributions are summarized below:

- **Misleading Prior Information Suppression:** In contrast to the traditional deep-learning reliance on facial geometry and reference priors, the proposed model leverages low-pass filtering to suppress the misleading prior information and obtain essential facial contours. These contours are then utilized to guide MF-GDDPM to generate high-quality facial images.
- **Dual-Stream Encoder:** By separately handling low-resolution facial contours and high-dimensional details, the proposed dual-stream encoder within Denoising Unet adapts the restoration process to different aspects of facial information. Enhanced by low-pass filtering, the low-resolution stream ensures that the fundamental shape and structure of the face are preserved, providing a reliable base for further detail addition. In parallel, the high-dimensional stream focuses on the fine details that contribute to the realism and texture of the facial image.
- **NIQE Module Integration:** To mitigate the instability and variability inherent in the MF-GDDPM, we incorporate the NIQE module as a loss function. This image quality assessment tool effectively evaluates and guides the generated images, ensuring the restored facial images meet high-quality standards and authentically represent human features.

2. SYSTEM MODEL

2.1. Overview

The MF-GDDPM framework, as shown in Fig. 1, is a novel system for facial image restoration, employing a two-stage diffusion process to enhance image fidelity and quality. Initially, it applies low-pass filtering to the input facial features to suppress high-frequency distortions while retaining essential details. These filtered features are adaptively fused to accentuate facial contours, preparing the image for restoration. In the forward diffusion stage, Gaussian noise is incrementally added to simulate corruption. The reverse diffusion stage then denoises the image, utilizing a dual-stream Guide encoder to ensure the balance between detail and fidelity. Concurrently, the NIQE module evaluates the authenticity of the facial features during the restoration, ensuring that the final, denoised image is not only high in quality but also true to natural human appearance. This comprehensive process establishes a new standard for restoring facial images in generative modeling, from initial low-pass filtering and dual-stream encoder to final quality assessment.

2.2. Diffusion Process

The MF-GDDPM framework addresses facial image restoration by orchestrating a series of processes that refine the original image x. Initially, the image is subjected to low-pass filtering (LPF), which serves to suppress high-frequency noise and enhance the essential low-frequency features. The LPF process is mathematically expressed as:

$$F_{lp} = \mathcal{F}^{-1}\{\mathcal{H}_{lp} \cdot \mathcal{F}(x)\}, \tag{1}$$

where $\mathcal{F}(x)$ is the Fourier transform of the original image x, $\mathcal{H}_{lp}$ represents the low-pass filter in the frequency domain, and $\mathcal{F}^{-1}$ is the inverse Fourier transform applied to obtain the filtered image F_{lp}.

Subsequently, in the forward diffusion process, the filtered image is fused with the original image's features F_i through AFF:

$$F_{fused} = \text{AFF}(F_i, F_{lp}), \tag{2}$$

where F_{fused} denotes the enhanced feature set ready for the diffusion stages.

The forward diffusion applies Gaussian noise to F_{fused}, progressively converting it into a noisier version x_T, described by the following Markov chain:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t\mathbf{I}), \tag{3}$$

where x_{t-1} is the image from the previous diffusion step, β_t is the noise variance at step t, and $\mathbf{I}$ is the identity matrix.

In the reverse diffusion stage, the model aims to restore the original image from the noisy state x_T by reversing the noise addition. This stage involves a neural network parameterized by θ that iteratively denoises the image:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \sigma^2_\theta(x_t, t)\mathbf{I}), \tag{4}$$

where $\mu_\theta(x_t, t)$ and $\sigma^2_\theta(x_t, t)$ are functions predicting the mean and variance for the distribution at each reverse step.

During this process, a dual-stream encoder combines the strengths of quality and fidelity guidance to inform the denoising process:

$$\epsilon_\theta(x_t, t) = D\left(E_q(x_t, F_{lp}, t) + E_f(x_t, F_{fused}, t), t\right), \tag{5}$$

where E_q is the quality-guidance encoder, E_f is the fidelity-guidance encoder, D is the decoder, and $\epsilon_\theta(x_t, t)$ represents the estimated noise at each reverse diffusion step t. The final output is the restored image $\hat{x}_0$, which is a denoised version of x_T.

This comprehensive approach, from low-pass filtering through feature fusion and dual-stream encoding, culminates in the reverse diffusion process that reconstructs a high-fidelity facial image from a corrupted state.

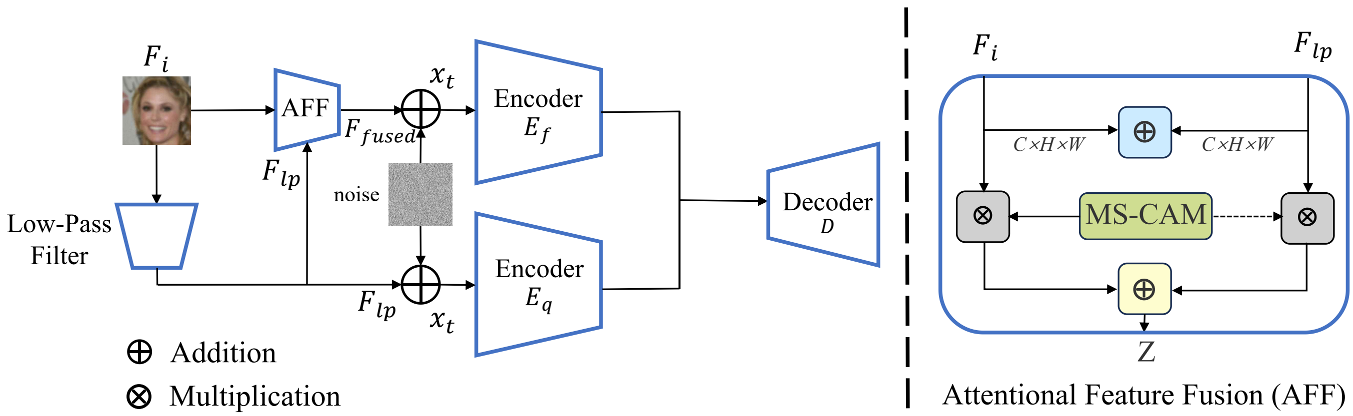

Fig. 1: System Overview

2.3. Training Stage

To incorporate the combination of Mean Squared Error (MSE) and Natural Image Quality Evaluator (NIQE) into the loss function for the training stage, we define the objective to minimize both the MSE between the estimated noise ϵ_θ and the actual noise ϵ added during the forward diffusion process and the NIQE value, which assesses the quality of the recovered image x_0. This combined loss function can be formulated as follows:

$$\mathcal{L}(\theta) = \lambda_{MSE} \left\| \epsilon - \epsilon_\theta(x_t, t) \right\|^2 + \lambda_{NIQE} \text{Ev}_{NIQE}(x_0), \quad (6)$$

where $\epsilon_\theta(x_t, t)$ is the noise predicted by the model for the noisy image x_t at timestep t,) is the actual noise added to the image. x_0 represents the recovered facial image, which can be formulated as

$$x_0 = \frac{x_t - \sqrt{1 - \overline{\alpha}_t}\epsilon_\theta}{\sqrt{\overline{\alpha}_t}}, \quad (7)$$

where $\overline{\alpha}_t$ is a coefficient related to the variance schedule of the diffusion process. The predicted facial image is further evaluated by NIQE module, which is employed as a loss function $\text{Ev}_{NIQE}(x_0)$. This loss function evaluates the quality of the recovered original image x_0. The coefficients λ_{MSE} and λ_{NIQE} are hyperparameters that balance the contributions of the MSE and NIQE loss components, respectively.

The estimated noise ϵ_θ is utilized to recover the original image according to the following relationship:

The NIQE loss for an image x_0 is defined as:

$$\text{loss}_{\text{NIQE}}(x_0) = D(f(x_0), M_{\text{natural}})$$

where the distance D is computed as:

$$D(f(x_0), M_{\text{natural}}) = \sqrt{(f(x_0) - \mu_{\text{natural}})^\top \Sigma_{\text{natural}}^{-1} (f(x_0) - \mu_{\text{natural}})}$$

- x_0: The image for which the quality is being evaluated. - $f(x_0)$: The feature vector extracted from x_0, typically using mean subtracted contrast normalized (MSCN) coefficients, which capture the local mean and variance of the image. - M_{natural}: The model of natural scene statistics, characterized by: - μ_{natural}: The mean vector of the feature space derived from a corpus of natural images. - Σ_{natural}: The covariance matrix of the feature space for the same corpus. - D: The statistical distance between the feature vector $f(x_0)$ and the model M_{natural}, measuring the deviation of the image's features from those typical of natural scenes. Lower values of D indicate a higher quality or more natural appearance of the image.

This training process involves sampling a random timestep t, adding noise to the original image x_0 to create a noisy version x_t using the forward process, predicting the added noise ϵ_{est} using the model, and then computing the combined loss that includes both the MSE between ϵ_{est} and ϵ, and the NIQE loss for the recovered image x_0.

2.4. Inference Stage

The inference stage involves reconstructing the original facial image x_0 from the noisy version x_T. This is done by iteratively applying the reverse diffusion process:

$$x_{t-1} = \mu_\theta(x_t, t) + \sigma_\theta(x_t, t) \cdot z, \quad (8)$$

where z is a sample from the standard normal distribution, and $\mu_\theta(x_t, t)$ and $\sigma_\theta(x_t, t)$ are the mean and standard deviation predicted by the model for timestep t.

The process starts with x_T sampled from a standard normal distribution and then applies the reverse diffusion step-by-step to generate $x_{T-1}, x_{T-2}, \ldots, x_0$. At each step, the model predicts the mean and standard deviation for the reverse diffusion transition, and a sample is drawn to generate the image for the next step. This iterative process continues until the model reconstructs the original image x_0, thus achieving facial image restoration.

3. METHODOLOGY

MF-GDDPM is designed to augment the fidelity and quality of restored facial images simultaneously. Specifically, the framework employs low-pass filter to suppress misinformation within the generated facial images, while AFF merges detail features from the original image with facial contour features. Moreover, the obtained pre-processed features are fed into a dual-stream encoder to improve the accuracy and quality of the restoration results. Additionally, NIQE is integrated as a performance metric, incorporated into our loss function to generate images with enhanced naturalness.

3.1. Low-Pass Filter Pre-Process

The low-pass filtering process, as mentioned in Equation 1, is pivotal in weakening high-frequency elements, thus retaining the low-frequency components that are essential for preserving the outline and structural integrity of facial features.

3.2. Attention Feature Fusion (AFF)

A significant component within the MF-GDDPM framework is the AFF, which is crucial for blending counter and high-frequency features from facial images to enhance the quality and fidelity of the restored face. The AFF mechanism is specifically designed to combine features derived from low-pass filtering and the original input image, thereby capitalizing on the complementary strengths of both features. Specifically, the AFF module employs a multi-scale channel attention (MS-CAM) strategy to attentively weigh the feature importance at different scales. This advanced attention scheme allows for the dynamic adjustment of feature contributions, ensuring that salient features are emphasized while redundant or misleading details are suppressed.

Given the input feature F_i and the low-pass filtered feature F_{lp}, the AFF module produces the fused feature F_{fused} which can be formulated as Equation 2, in which F_i is the original high-resolution input feature, F_{lp} is the feature after applying the low-pass filter, and F_{fused} is the synergized output which combines the information from both f_i and f_{lp} through the AFF mechanism.

3.3. Dual-Stream Guidance in Facial Image Restoration

Fig. 1 illustrates the dual-stream guidance approach for facial image restoration. We employ a combination of quality-guidance encoder and fidelity-guidance encoder within the DDPM framework to enhance the restoration of damaged facial images. The hybrid guidance architecture is expressed by:

$$\epsilon_\theta(x_t, F_{lp}, F_i, t) = D\left(x_{facial}, t\right) \tag{9}$$

$$x_{facial} = (E_q((x_t + F_{lp}), t) + E_f(x_t + F_{fused}), t)), t) \tag{10}$$

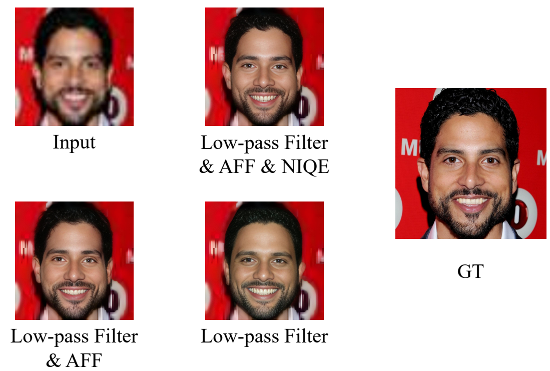

Fig. 2: Comparison of the Outputs from MF-GDDPM under Different Modules

where x_{facial} is facial-related noise feature. This feature is derived from a combination of outputs from two distinct encoders: the quality-guidance encoder E_q and the fidelity-guidance encoder E_f. The quality-guidance encoder, E_q, focuses on enhancing the overall visual quality of the restored image by processing the noise-added image x_t combined with the low-pass filtered features F_{lp}. The fidelity-guidance encoder, E_f, prioritizes the fidelity of the restoration, ensuring that critical details and features of the original damaged facial image F_i are accurately preserved and reconstructed. The term AFF(F_i, F_{lp}) represents the Attention Fusion mechanism that intelligently merges the high-dimensional features of the damaged image F_i with the stable low-frequency features of F_{lp}. This fusion process enriches the input to the fidelity-guidance encoder, allowing for a more nuanced restoration that balances high-quality visual appearance with faithful reconstruction of the original features. The output of both encoders is then combined and fed into the DDPM Decoder D, which reconstructs the facial image at each reverse diffusion step t. The decoder leverages the strengths of both streams, integrating the quality-enhanced and fidelity-preserved features to produce a high-resolution, restored facial image that is both visually appealing and true to the original.

Specifically, in a dual-stream framework, for E_q, the DDPM-based super-resolution technique is utilized for extracting detailed facial features from images with limited resolution, ensuring that the resultant high-resolution images are rich in detail and quality. To further refine the fidelity of facial restorations, E_f is incorporated within the Denoising Unet of the diffusion model. The dual-stream encoder processes two distinct types of feature maps: those derived from facial contours in low-resolution images and those pertaining to the high-dimensional details of facial features. The integration of these inputs is Guide by the AFF attention mechanism, formulated as Equation 2. AFF is utilized to ensure that the Denoising Unet can restore high-quality facial images while maintaining a high degree of faithfulness to the original images.

These features is empolyed into the dual-encoder of DDPM, which is specially designed to process the fused counter feature and detail feature respectively, which encapsulate both the original and low-pass filtered facial features. The incorporation of this Guide information into the encoder can be mathematically expressed as x_{facial}, which is the attention-fused latent representation. The x_{facial} is the output of the encoder, representing the Guide version of the input image.

These enhancements mark a significant stride in facial image restored field. By suppressing the misleading information, employing AFF to integrate features derived from low-pass filtering and the original input image, thereby capitalizing on the com- plementary strengths of both feature

3.4. Integration of NIQE for Quality Control

In addition to the Mean Squared Error (MSE), which trains the denoising capability of the Noise Predictor, the NIQE score is employed as a complementary loss function. This enables a comprehensive evaluation of the quality of the restoration results.

A lower score indicates a closer resemblance to natural images. The NIQE scoring process is as follows. During the training process, the estimated noise ϵ_θ is utilized to recover the original image, which is represented by Equation 7

To evaluate the quality of noise-suppressed images, NIQE is employed as Equation **??** The scoring mechanism of NIQE is based on comparing the statistical features of the generated image with those of a corpus of natural images, thus quantifying the degree of naturalness. The $loss_{NIQE}$ significantly enhances the authenticity and fidelity of the restored images, addressing the issue of generated images deviating from real features in both the generated and damaged pictures.

4. EXPERIMENT

4.1. Datasets and Experimental Metrics

We assessed the performance of our method using the CelebA dataset, LFW and WebPhoto datasets. For the CelebA dataset, we randomly selected 2500 images, which were then manually degraded. Additionally, we employed real-world datasets, namely the LFW-Test dataset and the WebPhoto dataset. These datasets consist of 1,711 low-quality face images and 407 low-quality face images respectively.

For the assessment of CelebA with ground truth, we employ traditional and widely-used image quality metrics, namely PSNR and SSIM. Additionally, we utilize CLIP-IQA [8], FID, and MUSIQ to comprehensively evaluate the image quality. In the evaluation of real-world datasets lacking ground truth, we employ FID to compare the restored images with the FFHQ dataset.

4.2. Comparisons with State-of-the-Art Methods

We compare the proposed MF-GDDPM against state-of-the-art methods, including DFDNet, GFP-GAN, GPEN, and CodeFormer. Our comprehensive comparisons encompass evaluations on both synthetic and real-world datasets. As shown in TABLE 1, Our method attains superior scores compared to existing methods, particularly excelling in CLIP-IQA and FID, with notably lower FID scores than its counterparts. MF-GDDPM also holds the second position in both MUSIQ and PSNR. Notably, the previously advanced method CodeFormer exhibits poor performance in PSNR and SSIM, falling significantly below mainstream methods. In contrast, MF-GDDPM demonstrates relatively high performance compared to other approaches. Furthermore, a qualitative comparison is presented in Fig. 3, illustrating although the compared methods struggle to produce satisfactory restoration results, our method successfully restores a natural and clear facial appearance.

Table 1: Quantitative Comparison on CelebA

Metrics	CLIP-IQA↑	FID↓	MUSIQ↑	PSNR↑	SSIM↑
Input	0.655	152.63	16.92	24.60	0.665
DFDNet [9]	0.331	48.82	26.31	22.88	0.683
GFP-GAN [10]	0.486	32.77	54.74	24.18	0.681
GPEN [11]	0.345	48.82	46.40	23.96	0.676
CodeFormer [12]	0.662	57.45	74.97	18.33	0.569
Guide-DDPM (ours)	0.688	31.00	69.40	24.52	0.665

Table 2

Dataset Metric	LFW-Test FID↓	WebPhoto-Test FID↓	Stable Diffusion FID↓
Input	128.08	173.55	438.51
DFDNet [9]	251.63	361.65	346.80
GFP-GAN [10]	105.30	153.30	366.26
GPEN [11]	57.68	91.88	359.01
CodeFormer [12]	55.82	86.17	332.36
Guide-DDPM (ours)	54.31	92.84	319.68

As shown in TABLE 2, MF-GDDPM attains the highest score on the Stable Diffusion and LFW dataset. Fig. 3 visually conveys that a comparable perceptual quality when compared to CodeFormer on the real-world faces, and clearly better than the rest. Fig. 3 shows the exceptional robustness of our method to Stable Diffusion output with heavy degradation.

4.3. Ablation Study

Effectiveness of Fidelity Encoder. We investigate the Importance of the fidelity encoder. The removal of the fidelity encoder results in a substantial loss of facial details in the reconstructed images, leading to a reduction in fidelity. The results in Ablation (a) of TABLE 3 suggest that the fidelity

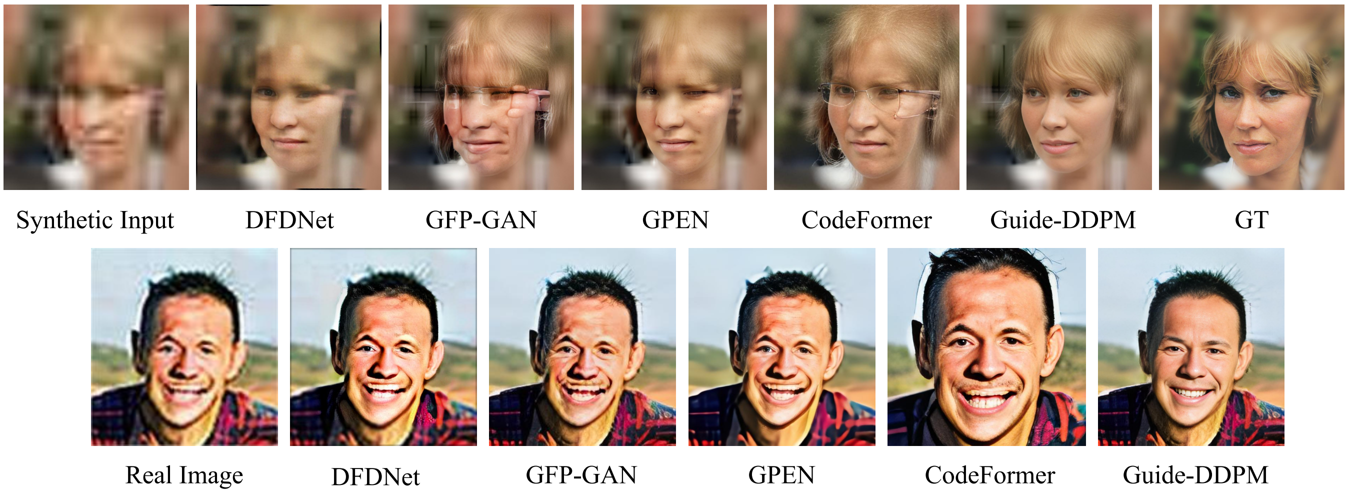

Fig. 3: Qualitative comparison on the CelebA image and stable diffusion image

encoder plays a crucial role in maintaining overall model quality and its removal adversely affects multiple aspects of performance.

Table 3: Ablation Study

	Metrics			Remove
	CLIP-IQA↑	FID↓	MUSIQ↑	
Benchmark	**0.67**	**45.37**	**59.47**	-
Ablation (a)	0.52	50.23	24.8	Fidelity Encoder
Ablation (b)	0.48	55.17	46.2	AFF-Module
Ablation (c)	0.44	57.30	45.4	Low-Pass Filter

Importance of Low-Pass Filter and AFF Module. In the ablation experiment conducted, we observed a noticeable reduction in quality and fidelity upon removing the AFF-module. To verify the superiority of the low-pass filter in our model, we compare it with different σ value, which represents the standard deviation of a Gaussian kernel used for convolution. It controls the smoothness of the filter and determines the extent to which high-frequency components are attenuated. A higher σ value results in a smoother filter response, effectively reducing fine-scale details and noise in the image, while a lower σ value preserves more fine-scale features. A suitable σ value is able to moderately weaken high-frequency elements and retain the low-frequency outline and face structure, which is beneficial to face restoration. Ablation (b) of TABLE 3 also indicates the importance of a low-pass filter, it significantly improves the quality and fidelity of the output image.

5. CONCLUSION

In this work, the proposed MF-GDDPM framework introduces a novel approach to facial image restoration, combining low-pass filtering, dual-stream guidance, and the NIQE module for enhanced fidelity and naturalness in restored images. In quantitative terms, our model has demonstrated outstanding performance across multiple datasets, surpassing existing methods. Qualitatively, it is evident that our approach produces images that are not only clearer but also more natural in appearance. The effectiveness of each component, validated through numerical experiments, highlights the robustness of the proposed framework.

6. REFERENCES

[1] Keng-Boon Ooi, Garry Wei-Han Tan, Mostafa Al-Emran, Mohammed A Al-Sharafi, Alexandru Capatina, Amrita Chakraborty, Yogesh K Dwivedi, Tzu-Ling Huang, Arpan Kumar Kar, Voon-Hsien Lee, et al., "The potential of generative artificial intelligence across disciplines: perspectives and future directions," *Journal of Computer Information Systems*, pp. 1–32, 2023.

[2] Vivian Liu and Lydia B Chilton, "Design guidelines for prompt engineering text-to-image generative models," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–23.

[3] Zijie J Wang, Evan Montoya, David Munechika, Haoyang Yang, Benjamin Hoover, and Duen Horng Chau, "Diffusiondb: A large-scale prompt gallery dataset for text-to-image generative models," *arXiv preprint arXiv:2210.14896*, 2022.

[4] Chenshuang Zhang, Chaoning Zhang, Mengchun Zhang, and In So Kweon, "Text-to-image diffusion model in generative ai: A survey," *arXiv preprint arXiv:2303.07909*, 2023.

[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*, 2021, pp. 9168–9178.

[6] Jiangxin Dong, Jinshan Pan, Zhongbao Yang, and Jinhui Tang, "Multi-scale residual low-pass filter network for image deblurring," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 12345–12354.

[7] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.

[8] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy, "Exploring clip for assessing the look and feel of images," in *AAAI*, 2023.

[9] Xiaoming Li, Chaofeng Chen, Shangchen Zhou, Xianhui Lin, Wangmeng Zuo, and Lei Zhang, "Blind face restoration via deep multi-scale component dictionaries," in *ECCV*, 2020.

[10] Xintao Wang, Yu Li, Honglun Zhang, and Ying Shan, "Towards real-world blind face restoration with generative facial prior," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[11] Xuansong Xie Tao Yang, Peiran Ren and Lei Zhang, "Gan prior embedded network for blind face restoration in the wild," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[12] Shangchen Zhou, Kelvin C.K. Chan, Chongyi Li, and Chen Change Loy, "Towards robust blind face restoration with codebook lookup transformer," in *NeurIPS*, 2022.