Acadlore Transactions on AI and Machine Learning
https://www.acadlore.com/journals/ATAIML

A Lightweight Conditional Diffusion Model for Restoring Turbulence-Degraded Facial Images

Shaoyu Sun, Pinchao Meng*

School of Mathematics and Statistics, Changchun University of Science and Technology, 130022 Changchun, China

* Correspondence: Pinchao Meng (mengpc@cust.edu.cn)

Received: 11-27-2025 **Revised:** 12-15-2025 **Accepted:** 01-05-2026

Citation: S. Y. Sun and P. C. Meng, "A lightweight conditional diffusion model for restoring turbulence-degraded facial images," *Acadlore Trans. Mach. Learn.*, vol. 5, no. 1, pp. 1–10, 2026. https://doi.org/10.56578/ataiml050101.

Abstract: Atmospheric turbulence induces severe blurring and geometric distortions in facial imagery, critically compromising the performance of downstream tasks. To overcome this challenge, a lightweight conditional diffusion model was proposed for the restoration of single-frame turbulence-degraded facial images. Super-resolution techniques were integrated with the diffusion model, and high-frequency information was incorporated as a conditional constraint to enhance structural recovery and achieve high-fidelity generation. A simplified U-Net architecture was employed within the diffusion model to reduce computational complexity while maintaining high restoration quality. Comprehensive comparative evaluations and restoration experiments across multiple scenarios demonstrate that the proposed method produces results with reduced perceptual and distributional discrepancies from ground-truth images, while also exhibiting superior inference efficiency compared to existing approaches. The presented approach not only offers a practical solution for enhancing facial imagery in turbulent environments but also establishes a promising paradigm for applying efficient diffusion models to ill-posed image restoration problems, with potential applicability to other domains such as medical and astronomical imaging.

Keywords: Diffusion model; High-frequency information; Atmospheric turbulence; Facial image

1 Introduction

Atmospheric turbulence is a common form of atmospheric motion that exerts long-term and severe impacts on optical imaging. Its irregular temporal and spatial fluctuations cause random phase distortions of the optical wavefront, leading to phenomena such as beam drift and spreading. These effects result in non-uniform distortion and blurring in images, significantly degrading imaging quality. Such impacts are prevalent in practical tasks, including facial recognition and long-range surveillance. Furthermore, compared with other types of image degradation, the irregularity of turbulent degradation makes it more difficult to model and mitigate [1]. Consequently, image restoration methods specifically designed to alleviate turbulent degradation have emerged. In the early stages of research, lucky frame imaging [2–7] was one of the common approaches to mitigate turbulent degradation. This method captures a large number of short-exposure images, selects those least affected by atmospheric turbulence for stacking and synthesis, and typically yields high-resolution results close to the diffraction limit. However, lucky frame imaging is mainly suitable for weak turbulence distortion scenarios and struggles to achieve satisfactory outcomes under strong turbulence. Additionally, since this method takes sequential images as input, the reconstruction process usually incurs high time costs, which limits its application in many scenarios requiring real-time performance. With the vigorous development of deep learning methods, various neural network-based approaches for single-frame image restoration have emerged. Many researchers have focused on deterministic networks such as Convolutional Neural Networks (CNNs) and Transformers. For example, Mao et al. [8] proposed TurbNet, which utilizes self-attention mechanisms to extract dynamic turbulent distortion maps from degraded images, assisting the network in learning the spatial variation characteristics of turbulence and effectively alleviating turbulence effects in degraded images. Li et al. [9] designed Fourier Modulated Attention (FMA) and Dynamic Mixing Layers (DML) and proposed SRConvNet. By combining the capabilities of CNNs and Transformers, this network achieves excellent performance in tasks such as super-resolution imaging and motion deblurring. Further innovative architectures and methodologies in this domain are detailed [10–15]. Leveraging the strong inductive bias and global dependency capture capabilities of CNNs and Transformers, researchers have developed numerous network architectures with outstanding restoration

results. However, due to the lack of randomness, deterministic networks tend to suffer from the mean regression effect when addressing the ill-posed problem of image restoration, resulting in reconstructed images lacking fine details [12]. Generative models aim to learn the distribution of target images and then obtain restoration results through distribution sampling. The randomness introduced during sampling can avoid the mean regression effect. As one of the generative models, Generative Adversarial Networks (GANs) benefit from the adversarial game between the generator and the discriminator, which guides the generator to produce results close to the distribution of clear images, achieving excellent performance in single-frame image restoration tasks. For example, Awasthi and Sharma [16] proposed a comprehensive loss function combining perceptual consistency and pixel-level fidelity in GANs, improving the overall restoration quality. Further developments in this area have been documented [17, 18]. However, during the training of GANs, they may suffer from mode collapse, making it impossible to learn the complexity and diversity of real data. Recently, diffusion models have achieved exciting results in the field of image generation due to their powerful distribution learning capabilities. Researchers have extended the application of diffusion models to image restoration tasks and obtained promising restoration outcomes. Özdenizci and Legenstein [19] utilized diffusion models to achieve model restoration independent of image size, achieving favorable results in weather-degraded images. Nair et al. [20] started denoising from degraded images rather than Gaussian noise, reducing the inference time of diffusion models and improving the fidelity of restoration results. Considering that diffusion models usually require high computational costs, Zhu et al. [21] integrated traditional plug-and-play methods into the diffusion sampling framework, maintaining high reconstruction perceptual quality while achieving high computational efficiency. Several other single-frame image restoration methods based on diffusion models have been explored [22–25]. Starting from the Gaussian distribution, diffusion models predict the true distribution of clear images and can output reasonable restoration results with rich details. Despite the great potential demonstrated by diffusion models, the limited information contained in single-frame degraded facial images (such as contours and facial feature positions) may lead the models to generate results that are seriously inconsistent with real images. Based on the above considerations, a conditional diffusion model integrated with high-frequency information, named HFDM, was proposed in this study. Super-resolution technology was utilized to capture high-frequency information from degraded images, which serves as a conditional constraint for the diffusion model, and the improved diffusion model was leveraged to restore single-frame turbulence-degraded facial images. Additionally, considering the complex inference steps of diffusion models, lightweight processing on the U-Net within the diffusion model was performed. Experiments demonstrate that HFDM achieves higher-quality restoration results while maintaining faster model inference efficiency. The main research contributions of this study are as follows:

1. Proposal of a conditional diffusion model with high-frequency information constraints to achieve high-fidelity restoration results.
2. Simplification of the U-Net, significantly reducing the model's computational cost.
3. Full demonstration of the model's restoration performance by conducting extensive experiments under scenarios with different turbulence degradation intensities.

The rest of this study is organized below. Section 2 details the proposed turbulence-degraded image restoration method, including the reconstruction of high-frequency information and the lightweight processing of the U-Net. Section 3 introduces the experimental data, implementation details, and various experimental results to demonstrate the restoration performance of the proposed method. Section 4 presents the conclusions of this study, summarizing its contributions.

2 Method

In this paper, a diffusion model is adopted as the restoration network, and super-resolution imaging technology is utilized to extract high-frequency information from degraded images, which serves as the conditional input to the diffusion model. Since high-frequency information contains details such as edges and textures of the image, it can constrain the diffusion model to generate restoration results closer to the original image, thereby improving the fidelity of the restored outcomes. The overall restoration workflow of our method is illustrated in Figure 1. In this section, we first elaborate on the acquisition of high-frequency information, and then further describe the implementation details of HFDM.

2.1 High-Frequency Information

Local regions of facial images typically exhibit significant structural characteristics and correlations. Transformers capture the dependencies between arbitrary positions in sequences directly through self-attention mechanisms, enabling global dependency modeling. This makes them highly suitable for the task of extracting high-frequency information from facial images, yet they also incur high computational costs. In the super-resolution network proposed by Li et al. [9], self-attention mechanisms are simulated via convolutions and Fourier transforms, which not only accurately represent facial image features but also achieve a more lightweight effect compared to the original Transformer. Therefore, we retain the main feature extraction structure of this network while removing the

final residual structure to serve as our high-frequency information extraction network. The process of extracting high-frequency information is illustrated in Figure 2.

Let x_{turb} denote a turbulence-degraded facial image. First, a 3×3 convolution is used to extract preliminary facial features, and the result is denoted as f_S.

$$f_S = conv\left(x_{turb}\right) \tag{1}$$

where, $conv\left(\cdot\right)$ denotes a convolutional operation. Subsequently, f_S undergoes M Attentive Convolutional Blocks for feature extraction, yielding f_M. A residual connection is performed between f_M and f_S to obtain deep facial features:

$$f_m = ACB\left(f_{m-1}\right), m = 1, 2, \cdots M \tag{2}$$

$$f_D = f_S + f_M \tag{3}$$

in the above formula, f_m denotes the result of m rounds of feature extraction, i.e., $f_0 = f_S$. $ACB\left(\cdot\right)$ represents the Attentive Convolutional Block, which is the core component of the network. Within this block, f_{m-1} is primarily processed through a 1×1 convolution to simulate the calculation of values in the self-attention mechanism. Additionally, fast Fourier transform (FFT) is utilized on f_{m-1} to efficiently generate global attention weights, and the values are fused with the weights in a multi-head attention manner. This process reconstructs clear facial features and the correlations between them.

At the end of the network, a pixel shuffle module (PS) and a 3×3 convolution operation are employed to restore feature f_D to the same size as the turbulence-degraded facial image x_{turb}, yielding high-frequency feature x_{HFI}. The results of high-frequency information extraction are presented in Figure 3.

$$x_{HFI} = conv\left(PS\left(f_D\right)\right) \tag{4}$$

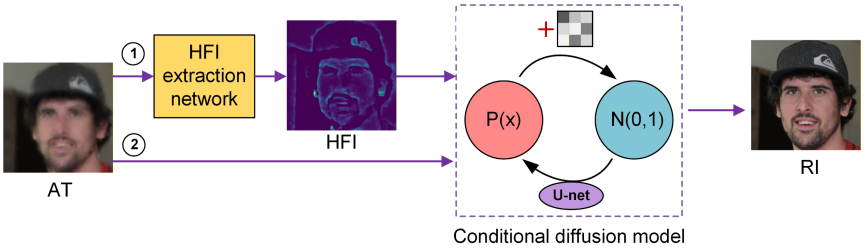

Figure 1. Overall restoration workflow. High-frequency information is used as the conditional input to the diffusion model to generate high-fidelity restoration results, AT denotes a turbulent degraded image, HFI represents high-frequency information, and RI stands for a restored image

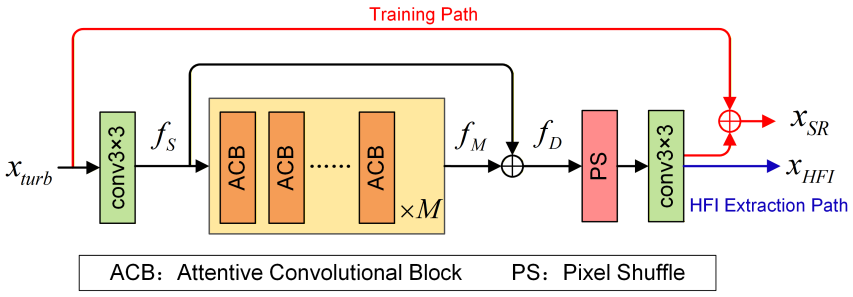

Figure 2. The process of extracting high-frequency information. Using lightweight convolutions and Fourier transforms to simulate the computational process of self-attention mechanisms, and achieving efficient capture of high-frequency features

Figure 3. Presentation of high-frequency information results. The super-resolution network [9] is retained during model inference, with the residual connections removed and the backbone network utilized for feature extraction to obtain high-frequency information

2.2 Conditional Diffusion Model

Compared with other generative models such as GANs and VAEs, diffusion models exhibit a more stable training process and higher fidelity. From a distribution perspective, our diffusion model maps the unknown distribution of clear facial images in the training set to a standard normal distribution through a noise addition process during training. Subsequently, it leverages a U-Net to perform a gradual denoising process starting from randomly generated Gaussian noise, predicts the distribution of clear facial images from the standard normal distribution, and obtains restored facial images by sampling from this distribution. In this process, the turbulence-degraded facial image to be restored and the corresponding high-frequency information serve as conditional inputs to guide the distribution in generating more realistic facial results.

The clear facial images in the training set are denoted as x_0, and the noise addition process for x_0 is expressed by the following formula:

$$x_t = \bar{\alpha}_t x_0 + \bar{\beta}_t \varepsilon_t, t = 1, 2, \cdots T \tag{5}$$

where, x_t denotes the result after t rounds of noise addition, and t here is typically referred to as the time step. ε_t is randomly generated Gaussian noise, while $\bar{\alpha}_t$ and $\bar{\beta}_t$ are two one-dimensional constants. Specifically, $\bar{\alpha}_t$ gradually decreases as the time step increases, eventually approaching 0, whereas $\bar{\beta}_t$ behaves conversely—it starts near 0 and gradually approaches 1. The purpose of this setting is to control the proportion of the clear image and Gaussian noise in the final noised result, enabling the distribution of the noised result to gradually skew toward the Gaussian distribution. The noise addition process is illustrated in Figure 4.

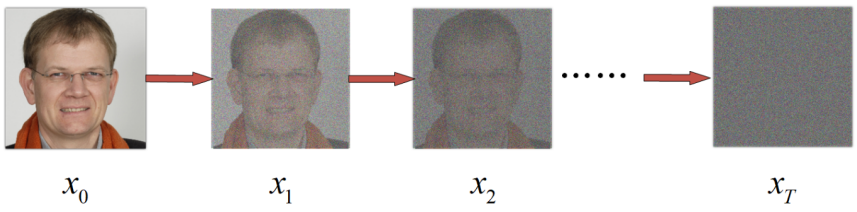

Figure 4. Noise addition process of diffusion models. Adding Gaussian noise to clean facial images to realize the transformation of the true facial distribution toward the Gaussian distribution

The denoising process proceeds in the reverse direction, starting from the final noised result x_T and progressing step-by-step toward $T \to 0$. Assuming that x_t is obtained after $T - t$ denoising steps from x_T, the process of denoising x_t to get x_{t-1} can be described by the following formula:

$$x_{t-1} = \mu_t(x_t) + \sigma_t^2 * \varepsilon \tag{6}$$

where, ε is a randomly sampled Gaussian noise, μ_t and σ_t^2 denote the mean and variance of x_{t-1}, respectively. Their calculation formulas are as follows:

$$\mu_t(x_t) = \frac{1}{\alpha_t}(x_t - \frac{\beta_t^2}{\bar{\beta}_t} \times \varepsilon') \tag{7}$$

$$\frac{1}{\sigma_t^2} = \frac{\alpha_t^2}{\beta_t^2} + \frac{1}{1 - \bar{\alpha}_{t-1}^2} \tag{8}$$

where, α_t and β_t are one-dimensional constants, and their relationships with $\bar{\alpha}_t$ and $\bar{\beta}_t$ are given by $\bar{\alpha}_t = \alpha_t \alpha_{t-1} \cdots \alpha_1$ and $\bar{\beta}_t = \beta_t \beta_{t-1} \cdots \beta_1$, respectively. ε' denotes the unknown Gaussian noise to be removed in this denoising step, which is predicted by feeding x_t and the degraded facial image x_{turb} into the U-Net as inputs. Additionally, to guide the denoising process in generating restored results that are more consistent with human visual perception and closer to the true facial distribution, we extract the corresponding high-frequency information x_{HFI} from the degraded image x_{turb} as an additional conditional input, assisting the diffusion model in generating more realistic facial textures and details.

2.3 U-Net Structure

In our U-Net, the U-Net structure proposed by Nair et al. [20] is lightweighted to reduce inference costs. We retain the main residual block-based feature extraction structure while removing the self-attention mechanism. The core of self-attention lies in calculating global similarity, and its time complexity is proportional to the square of the spatial dimension ($O\left(N^2\right)$), where N denotes the number of feature patches). On high-dimensional feature maps, this module generates substantial redundant computations, leading to slow inference speed. However, the U-Net itself already possesses strong capabilities in local feature capture and contextual information fusion. After removing the self-attention module, the model can significantly reduce the computational burden while maintaining core performance. The specific structure of our U-Net is illustrated in Figure 5. Where, X denotes the input to the U-Net, which is constructed by concatenating $\{x_{turb}, \bar{x}_t, x_{HFI}\}$ along the channel dimension. In the U-Net, a 3×3 convolution is first used to increase the image depth and extract preliminary features f.

$$f = conv\left(X\right) \tag{9}$$

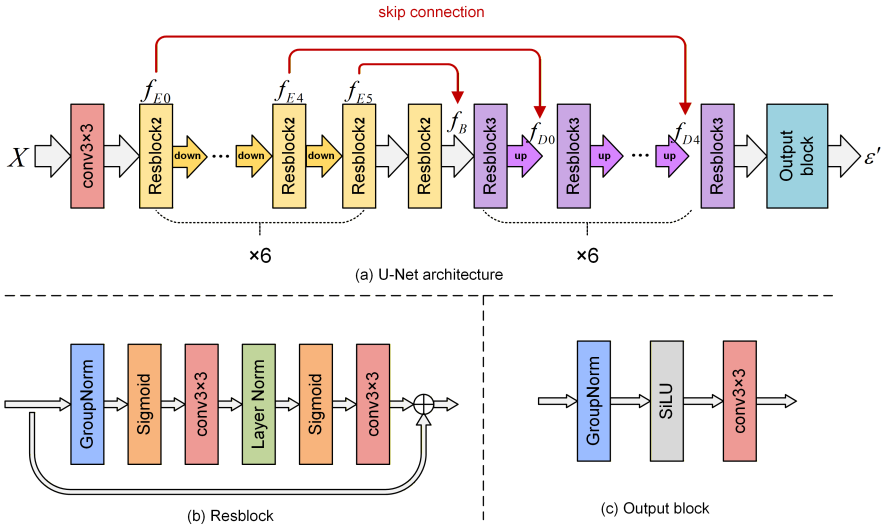

Figure 5. U-net network structure: (a) The overall architecture is mainly composed of residual modules, where the yellow part represents the encoder and the purple part represents the decoder; (b) Structure of the residual module; (c) Structure of the output module

In the encoder path, a 6-layer feature extraction structure is designed, where each layer consists of two consecutive residual blocks. A downsampling operation is added between layers to reduce the spatial size of the feature maps and increase the depth. Eventually, the shallow features $f_{E0} \sim f_{E5}$ of the encoder are obtained.

$$f_{E0} = Resblock_2\left(f\right), f_{Ei} = Resblock_2\left(down(f_{Ei-1})\right), i = 1, 2, 3, 4, 5 \tag{10}$$

where, $down\left(\cdot\right)$ denotes the downsampling operation, and the subscript of $Resblock_2\left(\cdot\right)$ indicates the number of residual blocks. Two consecutive residual blocks are also used in the bottleneck layer:

$$f_B = Resblock_2\left(f_{E5}\right) \tag{11}$$

In the decoder path, a 6-layer feature extraction structure is used symmetrically with the encoder, and 3 residual modules are arranged in each layer. An upsampling operation is employed between layers, and before feature extraction in each layer, a skip connection is performed with the corresponding encoder features. The formula is expressed as follows:

$$f_{D0} = up\left(Resblock_3\left(SC\left(f_B, f_{E5}\right)\right)\right) \tag{12}$$

$$f_{Di} = up\left(Resblock_3\left(SC\left(f_{Di-1}, f_{E5-i}\right)\right)\right), i = 1, 2, 3, 4 \tag{13}$$

$$f_{D5} = Resblock_3\left(SC\left(f_{D4}, f_{E0}\right)\right) \tag{14}$$

where, $up\left(\cdot\right)$ denotes the upsampling operation and $SC\left(\cdot\right)$ stands for skip connection. At the end of the U-Net, in the output module, normalization and an activation function are performed sequentially, and a 3×3 convolution is used to restore the result to 3 channels:

$$\varepsilon' = conv\left(SiLU\left(GN\left(f_{D5}\right)\right)\right) \tag{15}$$

where, $GN\left(\cdot\right)$ denotes the group normalization operation, and $SiLU\left(\cdot\right)$ stands for the Sigmoid Linear Unit activation function.

3 Experiments

In this section, we will demonstrate the lightweight achievements of the HFDM, enabling the model to complete inference in a shorter time. Subsequently, we will conduct qualitative and quantitative comparisons between our model and other diffusion models as well as super-resolution models to verify the effectiveness in alleviating turbulence degradation. Finally, we will validate the restoration performance of our model under turbulence degradation scenarios of various intensities.

3.1 Data Preparation

In this paper, the Flickr-Faces-High-Quality (FFHQ) dataset is used for all model training and experiments, with an image size of 256 × 256 × 3. The P2S method [26] is employed to generate turbulence-degraded facial images. In this simulation method, the ratio of the camera aperture D to the atmospheric coherence length r_0 is used to control the turbulence intensity of the degraded images.

3.2 Implementation Details

All model training and experiments are implemented using the PyTorch framework on an NVIDIA GeForce RTX 3090. In the high-frequency information extraction network, 3,000 facial images are used as the training set, with the batch size set to 16 and the number of model training iterations set to 10,000. When training the diffusion model, 5,000 facial images are used as the training set and 1,000 as the test set. The batch size is set to 8, the number of model training iterations is set to 50,000, and the time step during training is set to 1,000. Similar to the approach proposed by Nair et al. [20], during model inference, the total number of time steps is 100, with 40 of them skipped.

3.3 Model Complexity Comparison Experiment

We trained ATDDPM [20] using the same training configurations as the proposed method and conducted comparative experiments. The quantitative analysis results of model complexity and restoration performance are presented in Table 1 (where HFI denotes our high-frequency information extraction network). The experiments evaluate model complexity from three core dimensions: computational complexity (FLOPs), number of parameters (Parameters), and inference time per image (Seconds per image).

Meanwhile, the LPIPS and FID metrics are adopted to quantitatively assess restoration performance. The results show that, on the premise of achieving better restoration performance, the proposed method reduces the computational complexity by approximately 1.1%, decreases the number of parameters by 9.3%, and shortens the inference time per image by 20.6%, successfully achieving a better balance between complexity and performance. Notably, although our network additionally incorporates high-frequency information extraction, the overall restoration time per image is still significantly shorter, fully verifying the overall efficiency of the model.

Table 1. Model complexity comparison results

Method	FLOPs	Parameters	Seconds per Image	LPIPS↓	FID↓
ATDDPM [20]	627.39 G	311.03 M	6.1800 s	0.1626	55.2916
HFI	22.48 G	0.98 M	0.8107 s	-	-
Ours	620.54 G	282.07 M	4.9078 s	0.1324	35.3613

Note: Our method and ATDDPM [20] were trained with the same configurations on the FFHQ dataset, and the results show that our method achieves more efficient restoration

3.4 Restoration Results Comparison Experiment

We conducted comparative experiments on the FFHQ dataset with the turbulence intensity set to $D/r_0 = 4$. In the experiments, various image restoration methods (ATDDPM [20], DSRNet [10], and SRConvNet [9]) were trained with the same configurations as our model (5,000 images for the training set and 1,000 for the test set) to ensure the fairness of the comparison. The qualitative and quantitative comparison results are shown in Figure 6 and Table 2, respectively. In Table 2, we used two commonly used evaluation metrics, PSNR and SSIM, to measure the differences between the restoration results and the ground-truth images. The results indicate that our restoration results have the smallest deviations from the ground-truth images in terms of brightness, contrast, structure, and other aspects, making them closer to the real images.

Secondly, we employed the LPIPS metric, which is more consistent with human visual perception, to verify the restoration performance. It can be seen that our method achieved the optimal LPIPS result, demonstrating that there is a smaller perceptual difference between our restoration results and the ground-truth images. Finally, FID was used to compare image distributions, which is one of the important metrics for evaluating the generation quality of generative models. As shown in Table 2, compared with other methods, our diffusion model can generate results that are most similar to the distribution of real images. On the other hand, in the qualitative result Figure 6, compared with DSRNet [10] and SRConvNet [9], our restoration results have clearer facial details and image backgrounds. In the restoration result of ATDDPM [20] (Figure 6b), obvious distortion effects appear at the image contours. However, in our restoration result (Figure 6e), this issue is well improved by using high-frequency information as a conditional constraint.

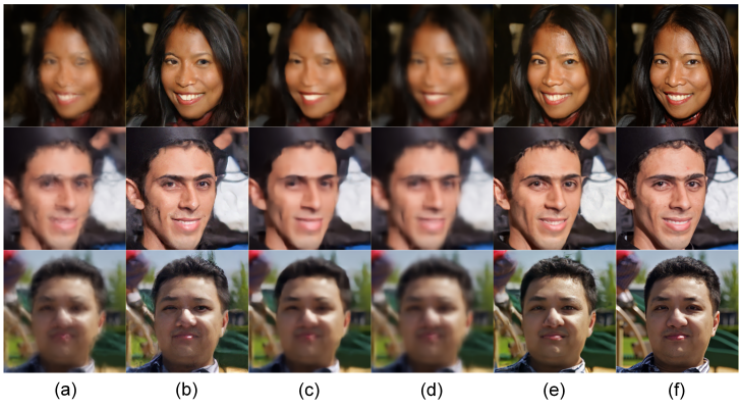

Figure 6. Qualitative restoration results comparison: (a) Turbulence-degraded images; (b) Restoration results of ATDDPM [20]; (c) Restoration results of SRConvNet [9]; (d) Restoration results of DSRNet [10]; (e) Restoration results of our model; (f) Ground-truth sharp images

Table 2. Quantitative comparison of the proposed method with other methods

	ATDDPM [20]	**DSRNet [10]**	**SRConvNet [9]**	**Ours**
PSNR↑	21.7669	21.2985	21.3787	**22.7135**
SSIM↑	0.6211	0.6048	0.6245	**0.6950**
LPIPS↓	0.1626	0.5041	0.3800	**0.1324**
FID↓	55.2916	100.2637	76.4782	**35.3613**

Note: The bold values in each row indicate the best results for that metric among all methods, and it can be seen that the proposed method exhibits significant advantages across all four metrics

3.5 Mixed Turbulence Intensity Restoration Experiment

We retrained the diffusion model on facial image data with mixed turbulence intensities. The training set consists of 1,600 facial images with weak turbulence degradation ($D/r_0 = 2$), 1,600 facial images with moderate turbulence degradation ($D/r_0 = 3.5$), and 1,600 facial images with strong turbulence degradation ($D/r_0 = 5$). The restoration results are shown in Figure 7. In scenarios with weak turbulence (Figure 7a) and moderate turbulence (Figure 7b), we can restore relatively clear facial results that are close to the ground-truth images. In the scenario with strong turbulence degradation, although the restoration results (the first row of Figure 7c) can achieve a level of clarity consistent with human visual perception, compared with the ground-truth images, inconsistent results with the original images may be generated in the local details of facial features.

Moreover, due to the excessive turbulence distortion, a small amount of distortion and blurriness may still exist in some restoration results (as shown in the second row of Figure 7c). However, overall, the proposed method can still effectively eliminate most turbulence effects and generate results close to real faces, providing reliable technical support for alleviating turbulence in different scenarios.

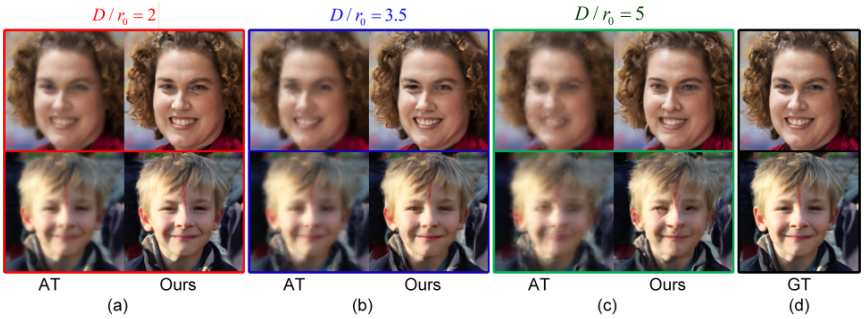

Figure 7. Restoration results under various turbulence intensities: (a) Restoration results under weak turbulence intensity; (b) Restoration results under moderate turbulence intensity; (c) Restoration results under strong turbulence intensity; (d) Ground-truth sharp images

4 Conclusion

Super-resolution technology was combined with diffusion models in this study. Reconstruction of the high-frequency information of facial images through super-resolution technology provided effective assistance for diffusion models, enabling them to generate smoother facial contours and clearer facial feature details. Comparative experiment results show that the proposed method achieved optimal performance among all four quantitative evaluation metrics and possessed efficient computational performance. In qualitative comparisons, this method effectively eliminated most random distortions and blurriness in images, significantly alleviating the negative impacts caused by turbulence degradation. Additionally, in the restoration experiment with mixed turbulence intensities, the proposed method maintained stable and excellent restoration performance under scenarios with moderate or lower turbulence intensities. Even when facing extremely severe turbulence degradation, although the restoration performance was subject to certain limitations, it still effectively eliminated most degradation features. This also reflects the applicability limitation of the current method in severely degraded scenarios. To address the aforementioned shortcomings, future work will focus on optimizations in the following two aspects:

1. Performance enhancement of the high-frequency information extraction network: For scenarios with stronger turbulence degradation, the network's ability could be improved to mine effective high-frequency information, providing strong support for diffusion models to generate more stable and realistic detailed features.
2. Improvement of high-frequency information fusion: In the current method, high-frequency information is only used as a simple conditional input for diffusion models, and its role has not been fully exerted. Subsequent work will optimize this fusion strategy, enabling high-frequency information to be integrated into the overall inference process of the network more deeply and efficiently.

Author Contributions

Conceptualization, S.Y.S.; methodology, S.Y.S.; validation, S.Y.S.; resources, P.C.M.; data curation, S.Y.S.; writing—original draft preparation, S.Y.S.; writing—review and editing, P.C.M.; visualization, S.Y.S.; supervision, P.C.M.; project administration, P.C.M.; funding acquisition, P.C.M. All authors have read and agreed to the published version of the manuscript.

Data Availability

The data used to support the research findings are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Z. Mao, A. Jaiswal, Z. Wang, and S. H. Chan, "Single-frame atmospheric turbulence mitigation: A benchmark study and a new physics-inspired transformer model," in *European Conference on Computer Vision, (ECCV 2022), Tel Aviv, Israel*, 2022, pp. 430–446. https://doi.org/10.1007/978-3-031-19800-7_25

[2] M. Aubailly, M. A. Vorontsov, G. W. Carhart, and M. T. Valley, "Automated video enhancement from a stream of atmospherically-distorted images: The lucky-region fusion approach," in *Atmospheric Optics: Models, Measurements, and Target-in-the-Loop Propagation III, San Diego, California, USA*, vol. 7463, 2009, pp. 104–113. https://doi.org/10.1117/12.828332

[3] X. Zhu and P. Milanfar, "Removing atmospheric turbulence via space-invariant deconvolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 157–170, 2012. https://doi.org/10.1109/TPAMI.2012.82

[4] M. Hirsch, S. Sra, B. Schölkopf, and S. Harmeling, "Efficient filter flow for space-variant multiframe blind deconvolution," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (CVPR 2010), San Francisco, CA, USA*, 2010, pp. 607–614. https://doi.org/10.1109/CVPR.2010.5540158

[5] C. P. Lau, Y. H. Lai, and L. M. Lui, "Restoration of atmospheric turbulence-distorted images via rpca and quasiconformal maps," *Inverse Probl.*, vol. 35, no. 7, 2019. https://doi.org/10.1088/1361-6420/ab0e4b

[6] Y. Lou, S. H. Kang, S. Soatto, and A. L. Bertozzi, "Video stabilization of atmospheric turbulence distortion," *Inverse Probl. Imaging*, vol. 7, no. 3, pp. 839–861, 2013. https://doi.org/10.3934/IPI.2013.7.839

[7] Y. Xie, W. Zhang, D. Tao, W. Hu, Y. Qu, and H. Wang, "Removing turbulence effect via hybrid total variation and deformation-guided kernel regression," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4943–4958, 2016. https://doi.org/10.1109/TIP.2016.2598638

[8] Z. Mao, A. Jaiswal, Z. Wang, and S. H. Chan, "Single-frame atmospheric turbulence mitigation: A benchmark study and a new physics-inspired transformer model," in *European Conference on Computer Vision, (ECCV 2022), Tel Aviv, Israel*, 2022, pp. 430–446. https://doi.org/10.48550/arXiv.2207.10040

[9] F. Li, R. Cong, J. Wu, H. Bai, M. Wang, and Y. Zhao, "Srconvnet: A transformer-style convnet for lightweight image super-resolution," *Int. J. Comput. Vis.*, vol. 133, no. 1, pp. 173–189, 2025. https://doi.org/10.1007/s11263-024-02147-y

[10] C. Tian, X. Zhang, Q. Zhang, M. Yang, and Z. Ju, "Image super-resolution via dynamic network," *CAAI Trans. Intell. Technol.*, vol. 9, no. 4, pp. 837–849, 2024. https://doi.org/10.1049/cit2.12297

[11] J. Qiu, R. Jiang, W. Meng, D. Shi, B. Hu, and Y. Wang, "Dual-domain cooperative recovery of atmospheric turbulence degradation images," *Remote Sens.*, vol. 16, no. 16, 2024. https://doi.org/10.3390/rs16162972

[12] C. Wang, G. Sun, C. Wang, Z. Gao, and H. Wang, "Monte carlo-based restoration of images degraded by atmospheric turbulence," *IEEE Trans. Syst., Man, Cybern.: Syst.*, vol. 54, no. 11, pp. 6610–6620, 2024. https://doi.org/10.1109/TSMC.2024.3399464

[13] R. Yasarla and V. M. Patel, "Learning to restore images degraded by atmospheric turbulence using uncertainty," in *2021 IEEE International Conference on Image Processing, (ICIP 2021), Anchorage, AK, USA*, 2021, pp. 1694–1698. https://doi.org/10.1109/ICIP42928.2021.9506614

[14] B. Xia, Y. Zhang, S. Wang, Y. Wang, X. Wu, and Y. Tian, "Diffir: Efficient diffusion model for image restoration," in *Proceedings of the IEEE/CVF International Conference on Computer Vision, (ICCV 2023), Paris, France*, 2023, pp. 13 095–13 105. https://doi.org/10.1109/ICCV51070.2023.01204

[15] C. Wang, J. Jiang, Z. Zhong, D. Zhai, and X. Liu, "Super-resolving face image by facial parsing information," *IEEE Trans. Biom., Behav., Identity Sci.*, vol. 5, no. 4, pp. 435–448, 2023. https://doi.org/10.1109/TBIOM.2023.3264223

[16] R. Awasthi and B. K. Sharma, "Image restoration using optimized generative adversarial networks for superior visual quality," *ICTACT J. Image Video Process.*, vol. 15, no. 3, 2025. https://doi.org/10.21917/ijivp.2025.0501

[17] M. K. Ngo-Huu, V. Ngo, D. T. Luu, B. N. Pham, and V. T. Nguyen, "Sterr-gan: Spatio-temporal re-rendering for facial video restoration," *IEEE MultiMedia*, pp. 1–12, 2025. https://doi.org/10.1109/MMUL.2025.3611072

[18] F. S. Khan, J. Ebenezer, H. Sheikh, and S. J. Lee, "Mfsr-gan: Multi-frame super-resolution with handheld motion modeling," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, (CVPRW 2025), Nashville, TN, USA*, 2025, pp. 800–809. https://doi.org/10.1109/CVPRW67362.2025.00084

[19] O. Özdenizci and R. Legenstein, "Restoring vision in adverse weather conditions with patch-based denoising diffusion models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 8, pp. 10 346–10 357, 2023. https://doi.org/10.1109/TPAMI.2023.3238179

[20] N. G. Nair, K. Mei, and V. M. Patel, "At-ddpm: Restoring faces degraded by atmospheric turbulence using denoising diffusion probabilistic models," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, (WACV 2023), Waikoloa, HI, USA*, 2023, pp. 3434–3443. https://doi.org/10.1109/WACV56688.2023.00343

[21] Y. Zhu, K. Zhang, J. Liang, J. Cao, B. Wen, and R. Timofte, "Denoising diffusion models for plug-and-play image restoration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, (CVPRW 2023), Vancouver, BC, Canada*, 2023, pp. 1219–1229. https://doi.org/10.1109/CVPRW59228.2023.00129

[22] M. Ren, M. Delbracio, H. Talebi, G. Gerig, and P. Milanfar, "Multiscale structure guided diffusion for image deblurring," in *Proceedings of the IEEE/CVF International Conference on Computer Vision, (ICCV 2023), Paris, France*, 2023, pp. 10 721–10 733. https://doi.org/10.1109/ICCV51070.2023.00984

[23] C. Saharia, W. Chan, H. Chang, C. Lee, J. Ho, T. Salimans, D. Fleet, and M. Norouzi, "Palette: Image-to-image diffusion models," in *ACM SIGGRAPH 2022 Conference Proceedings, Vancouver, BC, Canada*, 2022, pp. 1–10. https://doi.org/10.1145/3528233.3530757

[24] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, (CVPR 2022), New Orleans, LA, USA*, 2022, pp. 10 684–10 695. https://doi.org/10.1109/CVPR52688.2022.01042

[25] J. Whang, M. Delbracio, H. Talebi, C. Saharia, A. G. Dimakis, and P. Milanfar, "Deblurring via stochastic refinement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, (CVPR 2022), New Orleans, LA, USA*, 2022, pp. 16 293–16 303. https://doi.org/10.1109/CVPR52688.2022.01581

[26] Z. Mao, N. Chimitt, and S. H. Chan, "Accelerating atmospheric turbulence simulation via learned phase-to-space transform," in *Proceedings of the IEEE/CVF International Conference on Computer Vision, (ICCV 2021), Montreal, QC, Canada*, 2021, pp. 14 759–14 768. https://doi.org/10.1109/ICCV48922.2021.01449