Realistic Face Super-Resolution via Generative Adversarial Networks: Enhancing Facial Recognition in Real-world Scenarios

Muhammad Nouman Ashraf[2], Muhammad Farooq[1], Muhammad Shahid Farid[3], Muhammad Hassan Khan[3], Shuja-ur-Rehman Baig[4]
[1]Department of Information Technology, University of the Punjab, Pakistan
[2]Department of Computer Science, The University of Lahore, Pakistan
[3]Department of Computer Science, University of the Punjab, Pakistan
[4]Department of Software Engineering, University of the Punjab, Pakistan
*Correspondence: muhammadfarooq44@gmail.com

Citation | Ashraf. M. N, Farooq. M, Farid. M. S, Khan. M. H, Rehman. S. U, "Realistic Face Super-Resolution via Generative Adversarial Networks: Enhancing Facial Recognition in Real-world Scenarios", IJIST, Vol. 7 Issue. 2 pp 675-688, April 2025
Received | March 12, 2025 **Revised |** April 10, 2025 **Accepted |** April 13, 2025 **Published |** April 14, 2025.

The accuracy of real-world facial recognition operations faces challenges because of the difficulties connected to Low-Resolution image quality. This indicates that super-resolution methods play a vital role in improving recognition outcomes. Currently, available SR techniques do not achieve generalization due to their dependence on synthetic LR data that uses basic down sampling processes. The proposed GAN-based approach establishes a solution to this challenge through its simulation of actual degradation algorithms which combine Gaussian blur with noise addition and color modification and JPEG compression. Random application of augmentation parameters allows the GAN model to acquire knowledge about diverse and realistic low-resolution data distribution patterns during training. A unique unaligned face image pair dataset was made specifically for research using Zoom-In and Zoom-Out methods to capture high-resolution and low-resolution images from the same individuals. The dataset presents authentic real-life scenarios better than conventional paired collection methods. Based on experimental results our method produces substantial gains in performance compared to other super-resolution methods across both self-created face data as well as established surveillance data. The proposed model achieves higher visual quality standards while improving facial recognition accuracy under different operational situations. In conclusion, this study implements an effective SR solution for facial recognition which tackles problems with standard training datasets while creating authentic face image data. The proposed method shows promise for enhancing SR applications which need high-quality facial recognition capability in surveillance systems and other security-based operations.

Keywords: Super-Resolution; Generative Adversarial Networks; Face Recognition; Surveillance Datasets; and Image Quality Enhancement

Introduction:

High-resolution (HR) facial images are increasingly in demand for applications in surveillance, identification, and medical diagnostics. However, many existing super-resolution (SR) techniques face practical limitations due to fundamental constraints, affecting their real-world effectiveness. A large number of SR models require synthetically produced Low-Resolution (LR) data as their base. However, this approach often fails to accurately replicate real-world degradation effects, such as blurring, noise, and compression artifacts, leading to potential inaccuracies in model performance [1]. Training based on mismatched data between training samples and real-world conditions produces ineffective SR models. Secondly, models trained on limited paired High-Resolution (HR) and Low-Resolution (LR) data from a single source struggle with generalization, making it challenging to adapt to diverse real-world environments. The SR models trained on studio portraits with close-up details often deteriorate performance when applied to surveillance footage captured from a distance [2]. Finally, a significant performance gap arises when training data does not align with the actual deployment environment of the model. The deployment of high-resolution professional images during training followed by the application of security camera grainy footage leads to performance degradation of the SR model [3]. The identification performance of facial recognition systems becomes substantially more challenging due to monitoring issues that lead to diminished image quality resolution and disused camera positioning.

Data degradation impacts images through blurring, downsampling, noise, and compression artifacts, all of which significantly reduce facial recognition accuracy in affected conditions. The current paper proposes a GAN-based system designed to address the practical challenges associated with LR facial images captured in real-world scenarios. The FR-SRGAN (Facial Recognition Enhancement Super-Resolution Generative Adversarial Network), built upon Wang et al.'s [4] GFP-GAN presents a distinct approach to resolving the current issues of super-resolution technologies.

The FR-SRGAN implements a novel data expansion method for reproducing real-life image degradations encountered in natural settings. The model is trained using a diverse range of image degradation techniques, including Gaussian blur, downsampling, noise addition, color jittering, and JPEG compression, to enhance its robustness in real-world scenarios. The FR-SRGAN training benefits from diverse realistic LR images when random parameters are added to different types of variations during training. This enhanced training technique enables the model to achieve superior performance in blind face restoration by effectively handling unknown degradation scenarios during the reconstruction process. The FR-SRGAN delivers highly detailed and noise-reducing face restoration which enables dependable performance of facial recognition systems throughout different operational settings.

Objectives:

This study aims to address these limitations by developing more effective SR methods for facial recognition. Specifically, our objectives are:

- **To develop** a novel unaligned facial dataset using Zoom-In and Zoom-Out capture methods, simulating real-world conditions more accurately.
- **To enhance** face recognition performance by testing our SR model on both the new dataset and existing surveillance datasets.
- **To evaluate** the effectiveness of face recognition techniques across varying image resolutions in real-world surveillance scenarios.
- **To contribute** to the field by providing insights into dataset alignment and model performance in low-resolution settings.

In achieving these objectives, our study makes significant following contributions.

- We developed a fresh unaligned dataset that contains matching face images from high

and low resolutions that were taken through the Zoom-In and Zoom-Out methods.

- We obtain exceptional results that apply to our developed facial test dataset together with present-day surveillance image datasets.

Related Work:

Computer vision relies heavily on image restoration for various applications. Super-resolution is a key task within image restoration, alongside noise and blur removal and the elimination of compression artifacts. These restoration methods enhance image quality, making them valuable for applications in facial recognition, medical imaging, and surveillance. Through deep learning technology, the field of image restoration has achieved major experimental development. It has witnessed significant growth through three main techniques commenting Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), Transformers, as well as Diffusion-based models. These diverse approaches enhance image restoration research by offering unique advantages, contributing to its overall progress and innovation.

Variational Autoencoders (VAEs):

The Variational Autoencoder (VAE) model generates large images from compressed lower-resolution data that resides in its latent space segment. While Variational Autoencoders (VAEs) exhibit a strong ability to analyze and model noise, their effectiveness in super-resolution tasks remains limited. Recent research advancements indicate that incorporating downscaled image information [5] alongside adversarial training of VAEs [6] significantly enhances image quality in super-resolution tasks. The data distribution focus of VAEs results in smooth image outputs, but it often overlooks fine image details, potentially resulting in issues such as posterior collapse in complex model architectures. High-resolution image generation benefits from models including SR-VAE [6] and VDVAE-SR [7] despite their performance limitations.

Generative Adversarial Networks (GANs):

GANs have revolutionized image restoration by the combination of a generator and a discriminator component. The generator synthesizes realistic images, while the discriminator evaluates them by comparing them against real photographs, enhancing the overall image quality and authenticity. This Research demonstrates that adversarial training reaches successful results in super-resolution applications as well as deblurring operations. The significant breakthroughs achieved by SRGAN [1] and ESRGN [8] came from their application of perceptual loss for improving fine image textures. The relativistic discriminator produces enhanced real-world effects in the output. GANs demonstrate exceptional performance in challenging image restoration tasks, particularly in blind face restoration. Notable examples in this field include GPEN [9] and Real-ESRGAN [10], which effectively enhance facial details in low-quality images. GLEAN [11] leverages pre-trained GANs for superior image super-resolution, while PSFR-GAN [12] employs a semantic progressive restoration approach enabling high-quality facial image recovery through a framework. The VQFR system [13] restores facial details using vector-quantized dictionaries while incorporating an information and style loss framework to enhance image quality and realism. However, GANs face challenges such as mode collapse, unrealistic artifacts, overfitting, and high computational costs, necessitating ongoing research for more stable, efficient, and generalizable super-resolution methods.

Transformer-Based Models:

Transformer-based models have emerged as powerful tools for image restoration, excelling in tasks like super-resolution and deblurring due to their ability to capture long-range dependencies. Models like Uformer [14], SwinIR [15] and DATSR [16] have achieved state-of-the-art results by incorporating advanced Transformer architectures and attention mechanisms. However, challenges such as computational intensity, difficulty in capturing fine details, and susceptibility to overfitting persist, necessitating further research to optimize transformer models for broader image restoration applications.

Diffusion-based models:

Diffusion models are a generative approach that progressively refines an image through iterative noise addition and removal. These models have demonstrated significant potential in super-resolution tasks, with approaches like SR3 [17] achieving remarkable results, especially in face restoration. However, limitations include computational intensity, the potential imbalance between noise removal and detail preservation, and sensitivity to training data quality. Despite these challenges, advancements like IDM [18], DiffPIR [19], and SRDiff [20] have addressed specific issues, demonstrating the potential of diffusion models in generating high-quality, diverse super-resolution images.

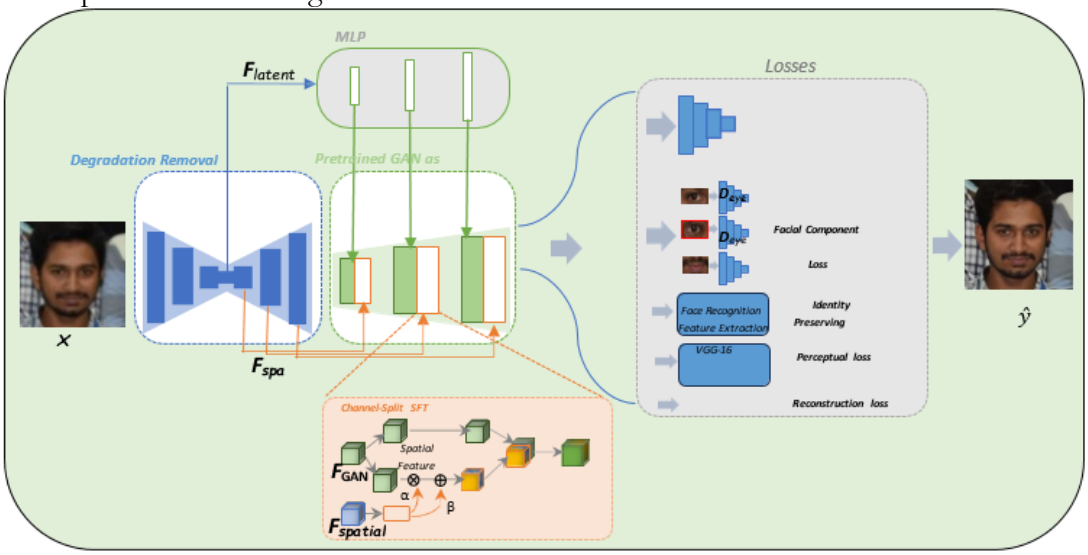

Figure 1: Overview of FR-SRGAN model framework. It consists of a degradation removal module (U-Net) and a pre-trained face GAN as facial prior. They are bridged by latent code mapping and several Channel-Split Spatial Feature Transform (CS-SFT) layers.

Methodology:

Our Proposed Model:

The FR-SRGAN develops an advanced solution that addresses current super-resolution technology weaknesses. The combination of GANs and U-Net provides FR-SRGAN with all key advantages and helps reduce their weaknesses to achieve enhanced resolution and quality enhancement. This model enables users to achieve both improved image resolution and enhanced quality while preserving the natural facial features and identity characteristics of the photographed individuals. Real-world testing confirms that FR-SRGAN delivers superior facial image quality results than other models in operation. By enhancing unclear or low-resolution faces, FR-SRGAN improves facial visibility, leading to greater accuracy and performance in face recognition systems. This innovation promises reliable applications especially for automated attendance systems because it enhances facial detail recognition to improve the system's performance in real-world scenarios.

The following sections examine the operational details of the proposed FR-SRGAN model. This section provides an in-depth evaluation of its structural configuration.

Overview of FR-SRGAN:

Figure 1 illustrates the overall architecture of FR-SRGAN, which combines a U-Net-based degradation removal module with a pre-trained face GAN, such as StyleGAN2, to generate high-quality facial images. A latent code mapping sequence connects these components using several Channel-Split Spatial Feature Transform (CS-SFT) layers for their operation.

The degradation removal module functions to erase sophisticated degradation while obtaining two distinct features which include latent features. F_{latent} for finding matching

StyleGAN2 latent codes in addition to multi-resolution spatial features F_{spatial} for the modification of StyleGAN2 features. The intermediate latent codes (W) emerge after F_{latent} undergo several linear operations. The latent code that matches the input image enables StyleGAN2 to create intermediate convolutional features named FmGAN which retrieve detailed facial attributes from its pre-trained GAN weights. The F_{spatial} features are then used to modulate the FmGAN features spatially using CS-SFT layers in a step-by-step manner, ensuring realistic results while maintaining high fidelity. During the training phase, besides the global discriminative loss, a facial component loss is introduced through discriminators to enhance the perceptually important face components such as eyes and mouth. Additionally, identity-preserving guidance is utilized to retain the person's original identity throughout the enhancement process.

Cleaning Up Blurry Faces using the Degradation Removal Module:

In real-world scenarios, the restoration of facial images often encounters complex quality challenges, such as low resolution, blurring, noise, and JPEG artifacts. Our degradation removal module is specifically crafted to tackle these issues and extract 'clean' features, denoted as F_{latent} and F_{spatial}, streamlining the subsequent processes.

To address the crucial task of degradation removal, we employed the U-Net architecture. U-Net serves a dual purpose: 1) it expands its receptive field to effectively handle significant blurring and 2) it generates features at multiple resolutions. Mathematically, this operation can be represented as:

$$F_{latent},\ F_{spatial} = \mathrm{U-Net}(x) \quad (1)$$

The F_{latent} latent features direct the input image toward its most compatible feature set in StyleGAN2. The multi-resolution spatial features F_{spatial} guide StyleGAN2 to better refine its features at the same time. The L1 restoration loss gets integrated for degradation removal during the initial training stage by applying it at each resolution scale. The system produces images at different U-Net decoder resolution levels and then enforces them to match corresponding levels of the ground-truth image's pyramid. The utilization of L1 restoration loss produces a more precise and effective degradation removal process through its deployment as an intermediate supervision mechanism. Each resolution output from the U-Net decoder is generated as images which enable direct image-by-image comparison against the ground-truth image's pyramid to achieve better alignment with the target results. This verification method improves the ultimate quality of degradation removal processing.

Leveraging a Pre-Trained Face Model for Rich Details (Generative Facial Prior and Latent Code Mapping):

The essential feature of FR-SRGAN involves incorporating a pre-trained face generation model known as a Generative Adversarial Network (GAN) typically based on StyleGAN2. The models function as "face experts" because they acquired extensive knowledge of facial elements from examining huge collections of face images. The model stores its acquired knowledge in layers which it refers to as "generative prior."

Closely related latent codes (Z) are found in the GAN's latent space by mapping the input image by typical methods to generate output images with a pre-trained GAN. The image restoration methods need extended optimization periods to maintain image fidelity.

FR-SRGAN provides an efficient solution to this challenge through its unique approach. The model operates through two stages by using U-Net generated "latent features" F_latent before proceeding to visualization of the final image (explained in Section 3.2: Cleaning Up Blurry Faces using the Degradation Removal Module). The latent features obtained from U-Net effectively retain the fundamental elements of blurry input images. FR-SRGAN analyzes latent features as input to identify the face from a database of faces contained within the pre-trained face generation model.

After identifying a suitable match FR-SRGAN starts creating "intermediate features" (F_GAN) from that match. The intermediate features in the system contain expanded details beyond basic code thus enabling a more precise depiction of the original facial features. FR-SRGAN adds Multi-Layer Perceptron (MLP) layers to its architecture for fine-tuning as displayed below.

$$W = MLP(F_{\text{latent}}). \quad (2)$$

The optimization adjustments maintain the semantic content of image information. New features created from W move through multiple convolution stages present in the pre-trained GAN to create features at each dimension level as shown below.

$$F_{GAN} = StyleGAN(\text{W}). \quad (3)$$

This two-step approach, leveraging the pre-trained face generation model's knowledge and then fine-tuning with the U-Net extracted features, allows FR-SRGAN to achieve superior image quality with a more efficient computational process compared to traditional methods.

Refining Details with Channel-Split Spatial Feature Transform (CS-SFT):

To enhance image quality, the spatial features of the input, denoted as F_{spatial} (generated by U-Net as per Equation 1), were employed to adjust the GAN features F_{GAN} As defined in Equation 3. Preserving spatial information is crucial for facial image restoration as it involves retaining local features and customizing restoration for different facial regions. To accomplish this, FR-SRGAN employs the Spatial Feature Transform (SFT) technique, which is renowned for its effectiveness in spatial feature modulation across various imaging tasks. At each resolution level, a pair of parameters (α, β) scale and shift the GAN features F_{GAN} based on F_{spatial} through several convolution layers. Mathematically, this operation can be expressed as:

$$\alpha, \beta = \text{Conv}(F_{\text{spatial}}), F_{\text{output}} = SFT(F_{GAN}|\alpha, \beta) = \alpha \odot F_{GAN} + \beta. \quad (4)$$

To maintain a balance between realism and detail preservation, Channel-Split Spatial Feature Transform (CS-SFT) layers were incorporated into the model. These layers selectively adjust a subset of the GAN features. F_{GAN} using the input features F_{spatial}, while allowing the remaining GAN features to pass through unchanged. Formally, this is represented as:

$$F_{\text{output}} = \text{CS} - \text{SFT}(F_{GAN}|\alpha, \beta), = \text{Concat}\left[\text{Identity}(F_{\text{GAN}}^{\text{split0}}), \alpha \odot F_{\text{GAN}}^{\text{split1}} + \beta\right]. \quad (5)$$

where $F_{\text{split0,GAN}}$ and $F_{\text{split1,GAN}}$ represent partitioned features from F_{GAN}.

The utilization of CS-SFT offers a dual advantage: it effectively integrates prior knowledge and enables adaptive image modulation, resulting in a well-balanced output in terms of texture and detail. Moreover, CS-SFT reduces computational complexity by requiring fewer channels for modulation.

In summary, these channel-split SFT layers were applied at each resolution level to produce the final restored facial image denoted by $\hat{y}$.

Training FR-SRGAN:

Our FR-SRGAN model is designed with a comprehensive learning objective, incorporating four distinct types of losses: reconstruction loss for accuracy, adversarial loss for texture realism, facial component loss for enhancing facial features, and identity-preserving loss for maintaining facial uniqueness.

Reconstruction Loss: FR-SRGAN utilizes both L1 loss and perceptual loss as part of its reconstruction loss to ensure an accurate reconstruction L_{rec}:

$$L_{\text{rec}} = \lambda_{l1} \parallel \hat{y} - y \parallel_1 + \lambda_{\text{per}} \parallel \phi(\hat{y}) - \phi(y) \parallel_1 \quad (6)$$

where ϕ represents a pre-trained VGG-19 network and the weights λ_{l1} and λ_{per} determine the importance of L1 and perceptual losses.

Adversarial Loss: For realistic textures, FR-SRGAN introduces an adversarial loss. L_{adv} to encourage the model to generate images indistinguishable from real ones. This loss function is similar to StyleGAN2 and is defined as:

$$L_{\text{adv}} = -\lambda_{\text{adv}} E_{\hat{y}}[\text{softplus}(D(\hat{y}))] (7)$$

where D is the discriminator, $E_{\hat{y}}$ Denotes the expectation over the generated images, soft plus is a smooth approximation of the rectified linear unit (ReLU) activation function, commonly used in adversarial networks, and λ_{adv} is the weight for the adversarial loss, controlling its influence in the training process.

Facial Component Loss: To refine specific facial regions like the eyes and mouth, a specialized facial component loss is introduced. This loss involves local discriminators trained on distinct facial areas, extracted through ROI alignment. Additionally, a feature style loss is integrated to match Gram matrix statistics between real and restored patches, enhancing facial details:

$$L_{\text{comp}} = \sum_{\text{ROI}} \left(\lambda_{\text{local}} E_{\hat{y}_{\text{ROI}}}[\log(1 - D_{\text{ROI}}(\hat{y}_{\text{ROI}}))] + \lambda_{fs} \parallel \text{Gram}(\psi(\hat{y}_{\text{ROI}})) - \text{Gram}(\psi(y_{\text{ROI}})) \parallel_1\right) (8)$$

where λ_{local} and λ_{fs} Represent the weights for local discriminative loss and feature style loss, respectively.

Identity-Preserving Loss: The identity-preserving loss (L_{id}) ensures that the restored image faithfully preserves the individual's unique facial identity. Facial feature comparison is performed through an ArcFace model (η) operating in its compact feature space to accomplish this process.

$$L_{\text{id}} = \lambda_{\text{id}} \parallel \eta(\hat{y}) - \eta(y) \parallel_1 (9)$$

The ArcFace model (η) represents the face feature extraction component while λ_id controls its weight when calculating L_id.

The total objective function combines these individual losses:

$$L_{\text{total}} = L_{\text{rec}} + L_{\text{adv}} + L_{\text{comp}} + L_{\text{id}} (10)$$

The weights assigned to these losses are as follows: $\lambda_{l1} = 0.1, \lambda_{\text{per}} = 1, \lambda_{\text{adv}} = 0.1,\ \lambda_{fs} = 200, \lambda_{id} = 10.$

Experiments:

Datasets:

To build a comprehensive dataset, we employed multiple data acquisition methods to collect facial images for training and evaluating restoration models. We implemented the data capture methodology described in "Zoom to Learn, Learn to Zoom" by Zhang et al. [21] using a Nikon D3500 camera to obtain high-resolution (HR) and low-resolution (LR) image pairs. A combination between zoom-in and zoom-out methods was used to create image pairs as illustrated in Figure 2. The dataset achieves greater generalizability by incorporating multiple data collection settings across two distinct locations: a controlled indoor environment and an outdoor setting with varying lighting conditions. After image acquisition, the next step involved a structured data pre-processing procedure. The Haar Cascade model was used for automatic face detection within the images as an initial step. The system proceeded to extract the face region from the identified areas in the pictures. Finally, to ensure consistency and meet computational requirements, we resized the cropped HR images to 512 × 512 pixels and the LR images to 128 × 128 pixels. The resulting dataset comprises 1,240 paired HR and LR images for training and 160 paired images for testing. Figure 3 shows the sample LR and HR images in our dataset. It's important to note that while the images within each pair are highly similar, they are not perfectly aligned, making them " weakly aligned" data.

Additional Dataset from Surveillance Scenario:

In addition to the primary dataset, experiments were also performed on Dataset-IV. Originally used in the SR-CGAN study by [2], Dataset-IV is a highly degraded dataset designed for super-resolution research.

Training Details:

Training is carried out through the Adam optimization algorithm for a cumulative total of 80,000 iterations. The initial learning rate is fixed at 2 × 10−3, which is subsequently reduced

by half at the 70,000th and 75,000th - operation milestones. All implementations are executed using the PyTorch framework and deployed on a computing environment equipped with a Single NVIDIA GeForce RTX 3060 GPU.

Experiments and Results:

In this section, we evaluated the effectiveness of our proposed model and compared it with state-of-the-art methods. The comparison is based on two categories of criteria:

Figure 2: Sample images of low- and high-resolution image pair capturing Using Zoom-In and Zoom-Out Techniques.

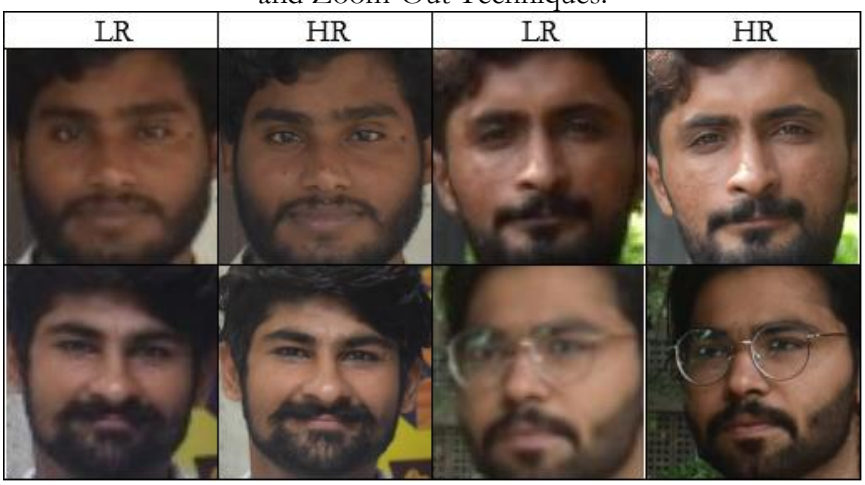

Figure 3: The first and third columns show sample real low-resolution face images, while the second and fourth column shows corresponding high-resolution face images.

Qualitative evaluation: We visually compared the restored images generated by our model with those produced by other leading methods and the original High-Resolution (HR) ground truth images. This visual inspection allows us to identify which method generates restorations that most accurately resemble the actual high-resolution (HR) images. Quantitative evaluation: For quantitative evaluation, we acknowledged that a definitive ground truth for super-resolution (SR) reconstruction is unavailable, as the HR-LR pairs in the test set are not perfectly aligned. However, their similarity is sufficient to allow comparison using a perceptual similarity measure. To quantitatively assess SR reconstructions against their corresponding HR images, we use the

Fr´echet Inception Distance (FID) as an evaluation metric. We employed the implementation by Seitzer[1], which utilizes an Inception model to extract features from an intermediate layer for comparing SR reconstructions with their paired HR images.

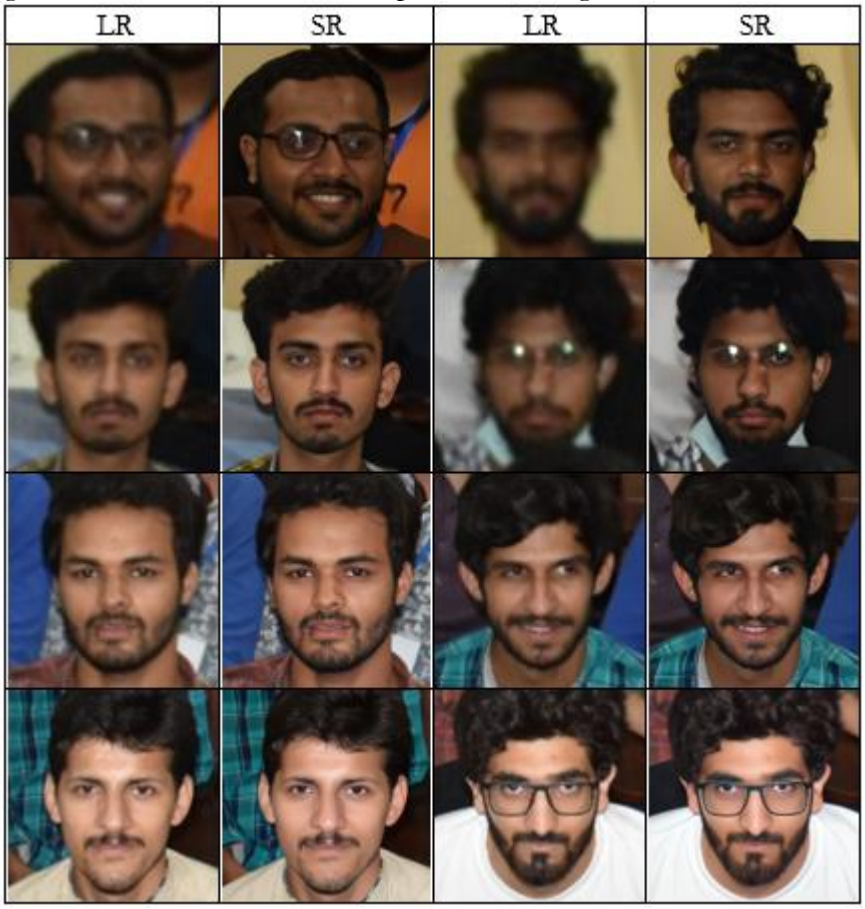

Figure 4: Figure showcasing the outcomes of our proposed super-resolution method (FR-SRGAN). Low-resolution images are displayed in the first and third columns, with their enhanced, super-resolved versions depicted in the second and fourth columns..

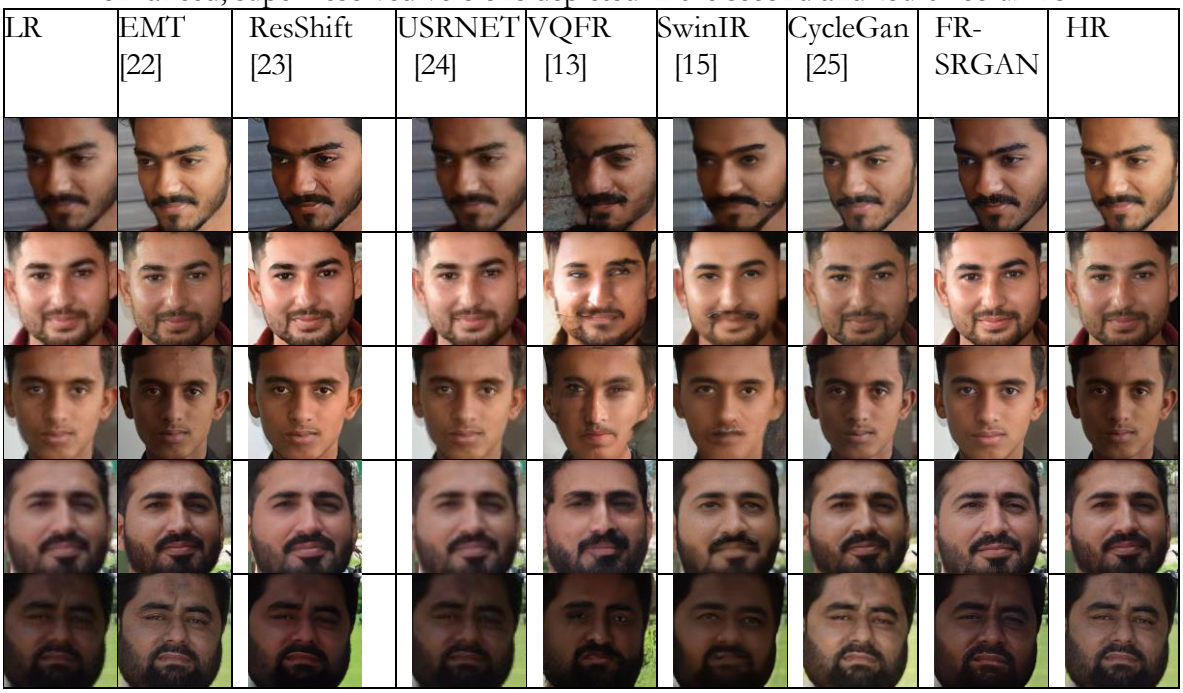

Figure 5: Comparision of state-of-the-art Super-Resolution methods

Our proposed model demonstrated remarkable performance. It particularly excelled in preserving facial details and identity while minimizing artifacts when applied to real-world, low-quality images. For optimal detail visualization, zooming in is recommended.

Figure 4 shows the results of using our proposed method for super-resolution. The images in the first and third columns are low-resolution, while those in the second and fourth columns are their enhanced, super-resolved versions.

Comparison with State-of-the-Art Methods:

We conducted a comprehensive comparison of our proposed model with cutting-edge facial restoration and super-resolution techniques both numerically and qualitatively.

1. Unpaired Image-to-Image Translation using Cycle Consistent Adversarial Networks CycleGAN[25]
2. Blind Face Restoration with Vector-Quantized Dictionary and Parallel Decoder[13] (VQFR)
3. Designing a Practical Degradation Model for Deep Blind Image Super-Resolution [26] (BSRGAN)
4. Deep Unfolding Network for Image Super-Resolution [24] (USENET)
5. Efficient Mixed Transformer for Single Image Super Resolution [22] (EMT)
6. Image Restoration Using Swin Transformer [15] (SWNIR)
7. Efficient Diffusion Model for Image Super-resolution by Residual Shifting[23] (ResShift)

Additionally, we trained these models using our face-specific training dataset to ensure an equitable assessment. The official implementation codes were employed for all models to maintain consistency and reliability in the evaluation process. Importantly, following rigorous training and testing procedures, it is observed that our proposed model outperformed the competing methodologies both visually as well as quantitatively. In terms of FID, our proposed method largely outperformed all other methods as shown in Table 1. Moreover, as shown in Figure 5, our method yields the most visually compelling results.

Performance Comparison on Highly Degraded Images:

To evaluate the effectiveness of our proposed model (FR-SRGAN) on severely degraded images, we compared its performance with Farooq et al. [2] proposed model (SRCGAN) using its most challenging test set of LR Dataset IV due to its high level of degradation. As illustrated in Figure 6, our model demonstrably outperforms SRCGAN on this particularly degraded dataset.

Table 1: Comparison with state of the art. Lower FID values are better. Our proposed model results are superior to those of other methods.

Method	FID
Our proposed model (FR-SRGAN)	**135.02**
CycleGAN [25]	284.7
VQFR [13]	135.09
USENET [24]	221.97
SWNIR [15]	159.38
EMT [22]	228.98
ResShift [23]	175.83

Discussion:

The core difference between SR-CGAN [2] and our proposed model (FR-SRGAN) lies in the training data used. SR-CGAN relies on paired real high-resolution (HR) and real low-resolution (LR) images. In contrast, our proposed model leverages real HR images from Dataset IV and generates its own corresponding LR images. This approach allows our model to learn from a wider range of potential LR image degradations compared to the fixed real LR counterparts used by SR-CGAN. As a consequence, when tested on real LR images, our proposed model outperformed SR-CGAN. This is because our proposed model's training incorporates a more diverse set of realistic LR image variations, potentially better equipping it to handle real-world scenarios. The model excels notably in the context of facial super-resolution, demonstrating an enhanced ability to preserve the.

Figure 6: Outcomes of our proposed super-resolution method. Low-resolution images are displayed in the first and third columns, with their enhanced, super-resolved versions depicted in the second and fourth columns.

Conclusion and Future Work:

Our proposed model presented in this paper addresses the limitations of existing super-resolution techniques for facial recognition by leveraging a data augmentation approach that simulates real-world image degradation processes. The approach enables the model to perform substantially better image restoration of low-resolution images and deliver superior facial recognition outcomes beyond current methods. The approach for future development would extend data augmentation methods to incorporate multiple types of real-world image degradation effects including motion blur and various degrees of lighting change. Additionally, future development efforts should both study different discriminator and generator architectures in GAN frameworks and create an entire system integration framework for real face recognition deployments to improve model functionality. Finally, research must explore methods to enhance the scalability of resource-limited environments and to improve efficiency along with ethical safeguards for high-quality facial recognition systems implementation.

References:

[1] W. S. Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," *arXiv:1609.04802*, 2017, doi: https://doi.org/10.48550/arXiv.1609.04802 Focus to learn more.

[2] M. Farooq, M. N. Dailey, A. Mahmood, J. Moonrinta, and M. Ekpanyapong, "Human face super-resolution on poor quality surveillance video footage," *Neural Comput. Appl.*, vol. 33, no. 20, pp. 13505–13523, Apr. 2021, doi: 10.1007/S00521-021-05973-0/METRICS.

[3] J. Y. & G. T. Adrian Bulat, "To Learn Image Super-Resolution, Use a GAN to Learn How to Do Image Degradation First," *Comput. Vis. – ECCV 2018*, pp. 187–202, 2018, doi: https://doi.org/10.1007/978-3-030-01231-1_12.

[4] H. Z. and Y. S. X. Wang, Y. Li, "Towards Real-World Blind Face Restoration with Generative Facial Prior," *2021 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Nashville, TN, USA*, pp. 9164–9174, 2021, doi: 10.1109/CVPR46437.2021.00905.

[5] J. M. T. Ioannis Gatopoulos, "Self-Supervised Variational Auto-Encoders," *Entropy*, vol. 23, no. 6, p. 747, 2021, doi: https://doi.org/10.3390/e23060747.

[6] A. M. A. Ali Heydari, "SRVAE: super resolution using variational autoencoders," *Pattern Recognit. Track.*, 2020, doi: https://doi.org/10.1117/12.2559808.

[7] D. Chira, I. Haralampiev, O. Winther, A. Dittadi, and V. Liévin, "Image Super-Resolution with Deep Variational Autoencoders," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 13802 LNCS, pp. 395–411, 2023, doi: 10.1007/978-3-031-25063-7_24.

[8] Y. Q. & C. C. L. Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks," *Comput. Vis. – ECCV 2018 Work.*, pp. 63–79, 2019, doi: https://doi.org/10.1007/978-3-030-11021-5_5.

[9] L. Z. Tao Yang, Peiran Ren, Xuansong Xie, "GAN Prior Embedded Network for Blind Face Restoration in the Wild," *arXiv:2105.06070*, 2021, doi: https://doi.org/10.48550/arXiv.2105.06070.

[10] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2021-October, pp. 1905–1914, 2021, doi: 10.1109/ICCVW54120.2021.00217.

[11] C. C. L. Kelvin C.K. Chan, Xiangyu Xu, Xintao Wang, Jinwei Gu, "GLEAN: Generative Latent Bank for Image Super-Resolution and Beyond," *arXiv:2207.14812*, 2022, doi: https://doi.org/10.48550/arXiv.2207.14812.

[12] K.-Y. K. W. Chaofeng Chen, Xiaoming Li, Lingbo Yang, Xianhui Lin, Lei Zhang, "Progressive Semantic-Aware Style Transformation for Blind Face Restoration," *arXiv:2009.08709*, 2020, doi: https://doi.org/10.48550/arXiv.2009.08709.

[13] Y. Gu *et al.*, "VQFR: Blind Face Restoration with Vector-Quantized Dictionary and Parallel Decoder," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 13678 LNCS, pp. 126–143, 2022, doi: 10.1007/978-3-031-19797-0_8.

[14] H. L. Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, "Uformer: A General U-Shaped Transformer for Image Restoration," *arXiv:2106.03106*, 2021, doi: https://doi.org/10.48550/arXiv.2106.03106.

[15] R. T. Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, "SwinIR: Image Restoration Using Swin Transformer," *arXiv:2108.10257*, 2021, doi: https://doi.org/10.48550/arXiv.2108.10257.

[16] L. V. G. Jiezhang Cao, Jingyun Liang, Kai Zhang, Yawei Li, Yulun Zhang, Wenguan Wang, "Reference-based Image Super-Resolution with Deformable Attention Transformer," *arXiv:2207.11938*, 2022, doi: https://doi.org/10.48550/arXiv.2207.11938.

[17] M. N. Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J. Fleet, "Image Super-Resolution via Iterative Refinement," *arXiv:2104.07636*, 2021, doi: https://doi.org/10.48550/arXiv.2104.07636.

[18] S. Gao et al, "Implicit Diffusion Models for Continuous Super-Resolution," *2023 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Vancouver, BC, Canada*, pp. 10021–10030, 2023, doi: 10.1109/CVPR52729.2023.00966.

[19] Y. Zhu et al, "Denoising Diffusion Models for Plug-and-Play Image Restoration," *2023 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW), Vancouver, BC, Canada*, pp. 219–1229, 2023, doi: 10.1109/CVPRW59228.2023.00129.

[20] Y. C. Haoying Li, Yifan Yang, Meng Chang, Shiqi Chen, Huajun Feng, Zhihai Xu, Qi Li, "SRDiff: Single image super-resolution with diffusion probabilistic models," *Neurocomputing*, vol. 479, pp. 47–59, 2022, doi: https://doi.org/10.1016/j.neucom.2022.01.029.

[21] X. Zhang, Q. Chen, R. Ng, and V. Koltun, "Zoom to learn, learn to zoom," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 3757–3765, Jun. 2019, doi: 10.1109/CVPR.2019.00388.

[22] L. Zheng, J. Zhu, J. Shi, and S. Weng, "Efficient mixed transformer for single image super-resolution," *Eng. Appl. Artif. Intell.*, vol. 133, p. 108035, 2024, doi: https://doi.org/10.1016/j.engappai.2024.108035.

[23] C. C. L. Zongsheng Yue, Jianyi Wang, "ResShift: Efficient Diffusion Model for Image Super-resolution by Residual Shifting," *arXiv:2307.12348*, 2023, doi: https://doi.org/10.48550/arXiv.2307.12348.

[24] K. Zhang, L. Van Gool, and R. Timofte, "Deep unfolding network for image super-resolution," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 3214–3223, 2020, doi: 10.1109/CVPR42600.2020.00328.

[25] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017-October, pp. 2242–2251, Dec. 2017, doi: 10.1109/ICCV.2017.244.

[26] S. H. Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium," *arXiv:1706.08500*, 2018, doi: https://doi.org/10.48550/arXiv.1706.08500.