EURASIP Journal on Image
and Video Processing

RESEARCH **Open Access**

Reconstruction and enhancement techniques for overcoming occlusion in facial recognition

Filip Pleško[1*], Tomáš Goldmann[1] and Kamil Malinka[1]

*Correspondence:
iplesko@fit.vut.cz

[1] Faculty of Information Technology, Brno University of Technology, Božetechova 2, 61200 Brno, CZ, Czech Republic

Abstract

Facial occlusions in surveillance footage can obscure important features, preventing facial recognition systems from identifying people. This work focuses on reconstructing these missing facial parts using Generative Adversarial Networks (GANs) to improve facial recognition accuracy while maintaining a low false acceptance rate. Additionally, we investigate how the generated images can be further enhanced using various image enhancement methods to boost recognition accuracy. To evaluate the results, we conduct experiments with widely used face embedding models, such as QMagFace and ArcFace, to determine whether image reconstruction and enhancement improve face recognition accuracy.

Keywords: Facial recognition, Facial reconstruction, Image enhancement, Image inpainting, ArcFace, MagFace, QMagFace, GAN

1 Introduction

Image processing technology has led to advancements in various fields, such as biometrics, security, surveillance, and personal identification. Among these applications, facial recognition (FR) systems have emerged as a crucial tool, offering both convenience and enhanced security. However, the robustness and accuracy of these systems are often challenged by various real-world conditions, such as facial expression or partially occluded faces [1, 2]. These scenarios can significantly degrade the performance of facial recognition systems, requiring innovative solutions to improve their resilience and effectiveness.

Generative Adversarial Networks (GANs) have shown remarkable potential in addressing these challenges through their ability to reconstruct missing or occluded parts of images [3–5]. This study explores the application of GANs to reconstruct damaged or partially obscured facial images and restore them to a state that is both visually plausible and suitable for processing by FR algorithms. The objective is to bridge the gap between the ideal conditions under which these systems are developed and the imperfect real-world scenarios in which they are deployed. Such pre-processing approaches for FR systems can bring significant benefits to public safety. In contrast to diffusion-based approaches, the significant advantage of the GAN-based solution is the inference time, which allows the reconstruction algorithm to be used in near-real time [6].

In addition to image reconstruction, this study employs state-of-the-art image enhancement approaches to enhance the quality of the reconstructed images. These enhancement methods are designed to refine the images' visual detail and overall clarity, which could potentially increase the accuracy of FR systems when dealing with compromised inputs.

This research proposes a comprehensive solution for improving the robustness of facial recognition technologies against a range of image impairments. The approach combines GAN-based reconstruction and advanced image enhancement. However, this approach has higher computational requirements, which necessitates a division into two independent parts.

In summary, the following research questions were defined:

- What impact do alterations to the neural network of a GAN have on the quality of facial image reconstruction?
- What is the impact of algorithms utilized for facial image reconstruction on the accuracy of facial recognition?
- Can state-of-the-art image enhancement methods improve FR accuracy?

2 Related works

The existing literature on facial recognition (FR) systems is rich with diverse approaches to overcoming the challenges posed by image imperfections [3, 7].

Research in this area ranges from sophisticated algorithms for reconstructing facial features in damaged or occluded images to advanced techniques for improving image quality. These solutions approach the problem from an image quality perspective, employing metrics for Image Quality Assessment (IQA) such as peak signal-to-noise ratio [8] and Structural Similarity Index [9] metrics.

However, this approach may not be sufficient for biometrics, which we address in a subsequent section of this study. Our focus is on facial image reconstruction, based on our previous paper, *Facial Image Reconstruction and Its Influence on Face Recognition* [10]. We extend this study by investigating whether enhancement techniques can be employed to improve the accuracy of facial recognition (FR) further, and we employ other algorithms for facial image reconstruction.

2.1 Facial image reconstruction

The reconstruction of facial images, also commonly referred to as facial image inpainting [11], is a challenging task that has been addressed by several innovative approaches documented in the scientific literature. However, for real-time facial recognition, inference time is a critical factor that has led to a reduction in the number of approaches. While diffusion-based methods offer impressive image reconstruction capabilities, this study focuses primarily on the feedforward neural network architecture, which allows for faster inference compared to diffusion-based approaches that suffer from computational cost [12]. During the course of our research, we encountered a number of different approaches, each of which offered a distinct perspective on the problem.

The first algorithm, based on the Generative Adversarial Network, G-NST [3] uses two discriminators alongside a semantic parsing network, where one discriminator functions to generate missing parts utilizing local loss, and the second ensures the coherence of these parts within the overall image structure. This algorithm incorporates neural style transfer to enhance visual coherence, involving a first step of image style clustering based on facial feature recognition, followed by applying style transfer via the VGG-16 network to ensure visually satisfying results. DFNet [4] uses the established U-net architecture and integrates a specialized fusion block connected to multiple decoder layers, emphasizing the filling in of missing image sections rather than generating a whole image, which distinguishes it from other techniques.

In contrast to previous approaches, the U-network architecture serves as the basis for the model proposed in [13], which is known as Image Inpainting via Conditional Texture and Structure Dual Generation. This architecture employs a two-stream network. In this architecture, the generator incorporates components that are responsible for both structure-constrained texture synthesis and texture-guided structure reconstruction. The output features are combined by bidirectional gated feature fusion (Bi-GFF) and a contextual feature aggregation (CFA) module [13]. The first module provides consistency enhancement, while the second module is designed to provide more vivid details.

In LaMa [7], the authors propose an architecture aimed at reconstructing large missing areas, complex geometric structures, and high-resolution images. The core of the architecture is Fast Fourier Convolution (FCC) [7], which allows the use of global context in early layers. This operator splits the channels into two parallel branches, the local branch and the global branch. For training, the design loss function was derived from the adversarial loss [14].

The last chosen model for reconstructing high-resolution images is called Aggregated Contextual-Transformation (AOT-GAN) [15], from a high-level perspective, the architecture is derived from GAN. However, unlike GAN, it uses the Aggregated Contextual-Transformation (AOT) instead of the Residual block in the generator, which is able to gather both informative distant contexts and rich patterns of interest.

2.2 Image enhancement

One possible method of increasing the effectiveness of FR systems is to enhance reconstructed facial images. In this study, four advanced image enhancement methods were employed: CodeFormer [16], DifFace [17], GFPGAN [18] and DFDNet [19], that were trained on FFHQ dataset [20]. Those are popular, state-of-the-art solutions, each contributing uniquely to improving the quality of facial images.

CodeFormer [16] uses a transformer-based architecture [21] to restore highly degraded images by modeling global interrelations and dependencies in the data. It includes a pre-trained quantized autoencoder with a discrete codebook [22], which, when combined with the Transformer, significantly reduces restoration uncertainty and facilitates the mapping of degraded to high-quality features.

The DifFace [17] method represents an innovative approach to blind face restoration (BFR), leveraging the capabilities of diffusion models without necessitating the training of these models under multiple constraints.

The DifFace method establishes a transition distribution that effectively models the transition from a low-quality (LQ) image to an intermediate diffused state of a pre-trained diffusion model, subsequently transitioning to the high-quality (HQ) target. This method primarily involves a neural network trained with L1 loss [23] on synthetic data, considerably simplifying the training process. The core advantage of DifFace lies in its error contraction mechanism, which systematically reduces the residual error during the transition phase, thus enhancing the robustness of the model against unknown and complex degradations. This approach streamlines the process by avoiding complex loss configurations and achieves notable robustness and efficiency, particularly in severe degradation scenarios.

GFPGAN [18], which stands for Generative Facial Prior Generative Adversarial Network, is a method designed to enhance facial images. It uses generative adversarial networks that are fine-tuned with a rich latent space of facial features to restore facial components in detail. The process begins by identifying degraded facial regions, which the network then enhances by drawing on its learned priors. This ensures that each reconstructed feature respects human faces' natural variations and structures [18]. The strength of GFPGAN lies in its capacity to reconstruct realistic textures and details often lost in damaged or low-quality images, particularly from a visual perspective.

DFDNet (Deep Face Dictionary Network) [19] represents an innovative approach to restoring faces from low-quality images. Unlike traditional approaches, DFDNet does not require a reference image of the same identity. Instead, it leverages deep multi-scale component dictionaries constructed using K-means clustering on high-quality images to guide the restoration process. DFDNet matches the degraded input with the closest features from these dictionaries and employs a Dictionary Feature Transfer (DFT) block to enhance the input by transferring high-quality details. Component Adaptive Instance Normalization (CAdaIN) is employed to harmonize style differences (e.g., illumination, skin tone) between the input and dictionary features. Furthermore, the method incorporates a confidence score to adaptively fuse the dictionary features with the input, which is further refined through a progressive restoration approach from coarse to fine. Extensive testing demonstrates that DFDNet can effectively generate realistic and high-quality facial image restorations across various degraded conditions, substantially outperforming existing methods that rely on identity-specific reference images.

Each method enhances the reconstructed facial image's clarity, resolution, and fidelity. This provides a robust set of tools for improving facial image quality, which could increase FR systems' performance accuracy under various challenging conditions.

2.3 Facial recognition

Over the past decade, there have been notable advancements in the capabilities of facial recognition (FR) technology. In 2014, the introduction of DeepFace [24] represented a significant turning point, signifying the adoption of neural network architectures to address the complexities of facial recognition. The proposed results indicate that it is the first algorithm for FR to demonstrate performance that surpasses that of humans in an unconstrained scenario.

This approach was soon surpassed by FaceNet [25], which demonstrated higher accuracy. At the same time, the research community has been dedicated to collecting larger datasets of facial images to improve algorithmic accuracy further. One notable direction for improving these algorithms has been to modify the loss function used to train neural networks for FR.

The contemporary loss functions, including A-Softmax [26], AM-Softmax [27], CosFace [28], ArcFace [29], and SFace [30] use a modified Softmax loss based on the observation that feature vectors show angular distribution. These loss functions facilitate further enhancement of FR accuracy.

In addition to these methods, a novel approach called MagFace was introduced in 2021 [31]. This approach, like its predecessors, focuses on the feature distribution of the vectors. Moreover, MagFace uniquely used the size of the feature vectors to enforce higher diversity for inter-class samples and similarity for intra-class samples. The modifications to the loss function resulted in enhanced accuracy on LFW [32], CFP-FP [33], AgeDB-30 [34], and CPLFW [35] datasets when compared to the results obtained by ArcFace [31].

Besides tuning a loss function and data to train a neural network model, FR accuracy can be improved by using a different metric for comparison. This idea led the authors of [36] to create a new metric called Quality Aware Metric, which extends the capabilities of the algorithm by using a comparison metric based on both cosine similarity and quality weight function. This approach outperforms a combination of MagFace and cosine similarity used for matching between feature vectors.

Current research focuses on improving FR accuracy, especially in challenging scenarios involving damaged images, occlusions, difficult poses, and varying lighting conditions [37].

In our research, we use FR algorithms to find out how face reconstruction algorithms affect FR accuracy. Although the performance of these FR algorithms is affected by factors such as the backbone used, the loss function, the amount of training data, etc., these aspects are not critical for evaluating the impact of reconstruction on FR.

3 Facial image reconstruction

This section outlines approaches to reconstructing corrupted facial images. Initially, we developed a baseline model for facial image reconstruction. Subsequently, we conducted a detailed examination of various modifications designed to enhance the quality of the reconstructed images. These modifications were based on the work proposed in [38]. Each proposed modification was evaluated independently to ascertain its impact on the reconstruction process. This enabled the identification of the model that exhibited the most promising potential for subsequent FR experiments. Furthermore, a dataset was developed to simulate corrupted facial images, which served to train and assess the efficacy of our enhanced reconstruction process.

3.1 Architecture of proposed neural network

Firstly, a base net was constructed and trained as a model for subsequent extensions. Then, we systematically explored various modifications in the architecture and evaluated their impact on the reconstruction quality. The architecture was finalized by integrating modifications that were proven to enhance the quality of the result, thus creating a model optimized for high-quality image restoration.

The structure of the model, inspired by autoencoder architecture described in [39], consists of encoder and decoder blocks arranged symmetrically to process image data efficiently. An architecture overview of this base autoencoder is shown in Fig. 1. The encoder segment comprises four blocks, each with a convolutional layer for feature extraction and a MaxPooling layer for dimensionality reduction. The sequential arrangement concludes with two dense layers that serve as a bridge to the decoder segment. The decoder then uses a sequence of convolutional and transposed convolutional layers to reconstruct the image from the encoded representations effectively.

In our initial architectural improvement, we explored eliminating fully connected layers within the generator and replacing them with convolutional layers. This modification produced a model with significantly fewer parameters, allowing for the inclusion of more filters in the convolutional layers. Increasing the filter count was to improve the model's ability to extract spatial features, which is crucial for accurately reconstructing image details. This modification is a significant step towards optimizing the network's architecture for more efficient and effective image restoration.

The second architectural improvement involved replacing pooling layers with convolutional layers that use strategic stride settings. This modification maintains the model's down-sampling capabilities while preserving important information. Strided convolutions, unlike deterministic pooling, enable the network to learn optimal spatial down-sampling methods [40]. This modification further enhances the models' ability to extract spatial features, representing another step toward improving image reconstruction quality.

In another attempt to refine our model, we increased the number of convolutional layers within each encoder block. This was done to enhance the feature extraction capabilities. However, this adjustment unexpectedly resulted in a degradation of the quality of the reconstruction results. As a result, we decided not to include this modification in the final model design. This highlights the challenge of achieving an optimal architecture for image restoration tasks.

Previous designs have aimed to improve the generator's effectiveness by introducing various architectural modifications. These adjustments have been suggested as potential

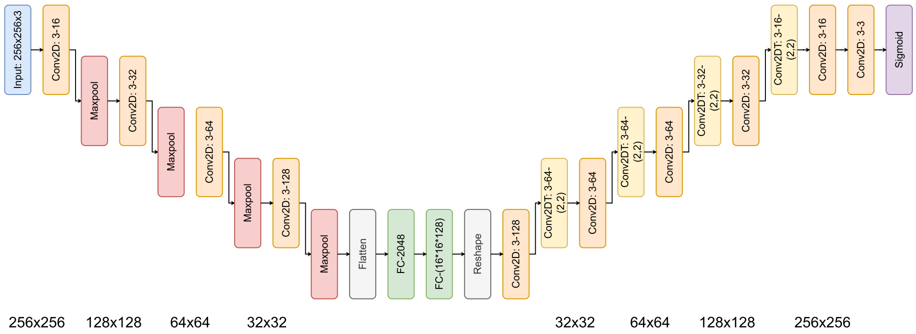


Fig. 1 Architecture of base autoencoder model used as a base model for all modifications

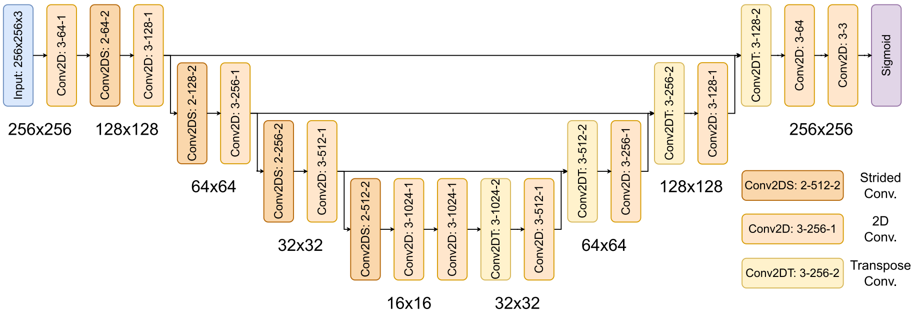


Fig. 2 An overview of a generator architecture that includes all modifications that improved the quality of the generated images [10]

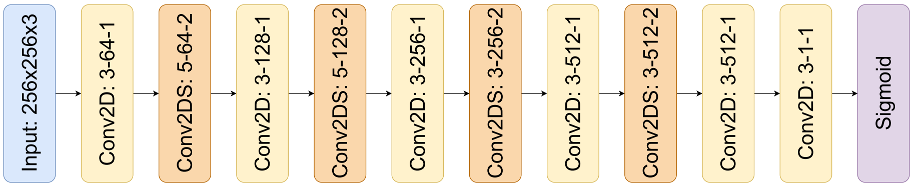


Fig. 3 Final discriminator design [10]

solutions to the challenges affecting the generator's performance. The final proposed model examines the possible advantages of integrating separate modifications into a single model. It also explores the synergy potential of these combined modifications when implemented together. As a result, we created the architecture of a combined model shown in Fig. 2.

A series of exploratory analyses were carried out using a set of models from the Keras U-Net collection [41]. These models were updated to incorporate the modifications outlined earlier in the text. This design aims to explore and determine the optimal architecture for facial image reconstruction.

Another critical component of the GAN model is the discriminator [42]. The discriminator receives both original and generated images as input, and its primary objective is to determine which image is real and which is fake [43]. The generator's effectiveness is inversely proportional to the discriminator's ability to discriminate between the two image types, so a less distinguishable output means an increase in generator output image quality. This evaluative feedback from the discriminator is essential for refining the generative process. To accomplish this task, a convolutional neural network (CNN) is used to extract features from the input and classify them into two distinct classes. The detailed architecture of this discriminator is shown in Fig. 3.

3.2 Dataset

For this work, we used the CelebA dataset [44] as a basis for image reconstruction algorithms and their evaluation in the context of facial recognition. Due to the lack of damaged images in this dataset, modifications were necessary to meet the research requirements.

To simulate occlusion in facial images, we created a modified version of the dataset called CelebA-C . This variant was generated by adding 30 randomly placed lines with widths between 8 and 13 pixels and lengths between 10 and 20 pixels, filled with RGB Gaussian noise $\mathcal{N}(128, 35)$, to the face region of each image from the original CelebA dataset. Algorithm 1 shows the pseudocode of how the drawn occlusion was created.

Algorithm 1 Draw random occlusion to an image

```
$num\_lines \leftarrow 30$
$start\_points\_xy \leftarrow [left\_eye, right\_eye, nose, mouth]$
$start\_point \leftarrow SelectRandom(start\_points\_xy)$
Draw square with center at start_point and size of side 20 px
$directions \leftarrow [left, right, up, down]$
$direction \leftarrow \emptyset$
for $line = 0$ to $num\_lines$ do
    $direction \leftarrow SelectRandom(directions - direction)$
    $distance \leftarrow SelectRandom([10, \ldots 20])\ px$
    $end\_point \leftarrow$ Select point in $direction$ and $distance$
    $line\_width \leftarrow SelectRandom([8, \ldots 13])$
    $DrawLine(start\_point, end\_point, line\_width)$
    $start\_point \leftarrow end\_point$
end for
Fill lines with Gaussian noise $\mathcal{N}(128, 35)$ and add them to original image.
```

Existing image reconstruction solutions use the NVIDIA Irregular Mask Dataset [45]. For a pertinent comparison of our solution with existing solutions, we also decided to use this dataset for the evaluation. From the dataset, we used a test set from which we selected masks that covered at most 50 % of the content. Since the test set itself contained only 12000 mask images, we had to duplicate some of them. For ease of replication, we performed the mask duplication by resetting the index after exhausting the original set and applying the masks from the beginning. Examples of each mask dataset and their combination with original data are shown in Fig. 4.

4 Experiments and results

Firstly, the performance of the facial image reconstruction algorithm was assessed and compared using the IQA metrics, as described in Subsection 4.2. While these metrics are widely recognized for evaluating the effectiveness of image reconstruction algorithms, they may not consider the potential impact on biometric recognition capabilities. With FR algorithms, the changes in latent space can be evaluated to examine the effect of reconstruction on FR accuracy. A more detailed examination of how FR is employed to assess image quality in generated images is presented in Subsection 4.2. The assessment of these two approaches can provide insight into the performance of the algorithm in terms of image quality, which has implications for biometric verification processes.

4.1 Training

In order to achieve the most accurate comparison results across architectures, uniform parameters were employed to train all models. For training purposes, an A40 GPU was utilized, with each model undergoing training for a total of 20 epochs with a batch size of 32 images.

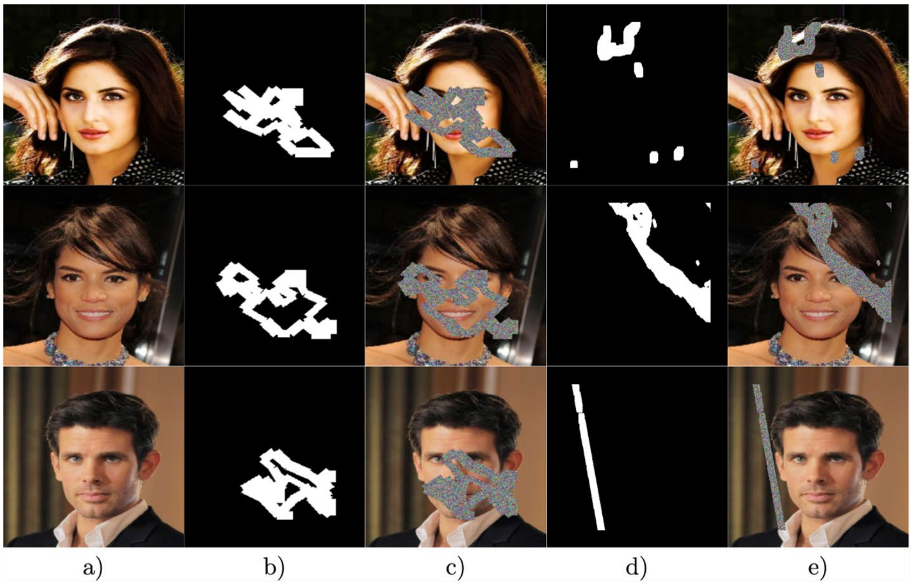


Fig. 4 (a) Example image from the CelebA dataset, (b) corresponding to our drawn masks, (c) combination of the drawn masks with the CelebA dataset, referred to as CelebA-C, (d) example of NVIDIA irregular masks, and (e) combination of NVIDIA irregular masks with the CelebA dataset

The generator was trained using a combination of binary cross-entropy and mean square error with weights of 1 and 100 as the loss function, and Adam was employed as the optimizer, with a learning rate of 0.0001 and β_1 of 0.5. The discriminator was trained using the binary cross-entropy loss function with the optimizer Adam, with the learning rate set to 0.00025 and β_1 set to 0.5.

4.2 Performance metrics

Although this study is primarily focused on evaluating the quality of reconstruction by FR algorithms and corresponding metrics, we also performed an evaluation with common metrics to calculate Image Quality Assessment (IQA). In order to evaluate the capability of the proposed neural network for image reconstruction, we have chosen commonly used metrics, namely peak signal-to-noise ratio (PSNR) [8], Structural Similarity Index (SSIM) [9] metrics, Learned Perceptual Image Patch Similarity (LPIPS) [46], Feature Similarity Index (FSIM) [47], Multi-scale Structural Similarity (MS-SSIM) [48], Fréchet Inception Distance (FID) [49], which facilitate direct comparison between the proposed solution and existing methods.

The first metric, PSNR, expresses the ratio between the maximum possible power of a signal and the power of the interfering noise. For a reference image f and a processed image g, it is given by [8]:

$$PSNR(f,g) = 10 \times log_{10}\left(\frac{(2^n-1)^2}{MSE(f,g)}\right),$$

where MSE denotes the mean squared error, which represents the average squared difference between the pixels of the original and the processed images. The variable n

denotes the number of bits per pixel, typically set to 8, reflecting the standard bit depth in image processing [8].

Another metric, SSIM, is defined by the relationship between the means, variances, and covariance of the original and processed images [9]. This metric is formally defined as follows:

$$SSIM(I,I') = \frac{(2\mu_I\mu_{I'} + C_1)(2\sigma_{II'} + C_2)}{\left(\mu_I^2 + \mu_{I'}^2 + C_1\right)\left(\sigma_I^2 + \sigma_{I'}^2 + C_2\right)},$$

where μ_I and $\mu_{I'}$ represent the means of the original and modified images, respectively. The term $\sigma_{II'}$ denotes the covariance between the two images, while σ_I^2 and $\sigma_{I'}^2$ correspond to the variances of the original and processed images, illustrating the variability within each image. To normalize the measurement and control the stability of the division with weak denominators, constants C_1 and C_2 are introduced, defined as $C_1 = (k_1L)^2$ and $C_2 = (k_2L)^2$, where $k_1 = 0.01$ and $k_2 = 0.03$. The term L represents the dynamic range of pixel values, typically set to 255 for 8-bit images. This formulation emphasizes the composite evaluation of luminance, contrast, and structural similarity between the compared images, providing a comprehensive assessment beyond mere pixel-by-pixel differences [50].

A subsequent metric, designated as Multi-Scale Structural Similarity (MS-SSIM), was proposed in [48]. This metric, derived from SSIM, aims to determine contrast c and structure s similarity across varying scales. Luminance comparisons l are calculated only at Scale MM. The multi-scale structural similarity measurement is defined as follows:

$$\text{SSIM}(\mathbf{x},\mathbf{y}) = [l_M(\mathbf{x},\mathbf{y})]^{\alpha_M} \cdot \prod_{j=1}^{M} \left[c_j(\mathbf{x},\mathbf{y})\right]^{\beta_j} \left[s_j(\mathbf{x},\mathbf{y})\right]^{\gamma_j}, \tag{1}$$

where α_M, β_M and γ_M are parameters to control importance of different components.

In contrast to traditional methods that prioritize pixel differences, the metric Feature Similarity Index for Image Quality Assessment (FSIM) [47] is based on perceptual features such as edges and textures that closely align with human vision. The algorithm is divided into phases, with the first phase determining significant image features using the phase congruency (PC) model. Furthermore, the gradient magnitude (GM) is calculated as the secondary feature to encode contrast information. In the initial stage, a local similarity map is calculated, while the subsequent stage performs pooling of the similarity map to yield a single similarity score. Formally, the similarity between $PC_1(x)$ and $PC_2(x)$ is given by [47]:

$$S_{PC}(\mathbf{x}) = \frac{2PC_1(\mathbf{x}) \cdot PC_2(\mathbf{x}) + T_1}{PC_1^2(\mathbf{x}) + PC_2^2(\mathbf{x}) + T_1}, \tag{2}$$

where T_1 represents a positive constant that is employed to control stability.

In the case of GM, the similarity is defined by:

$$S_G(\mathbf{x}) = \frac{2G_1(\mathbf{x}) \cdot G_2(\mathbf{x}) + T_2}{G_1^2(\mathbf{x}) + G_2^2(\mathbf{x}) + T_2}, \tag{3}$$

where T_2 is a positive constant depending on the dynamic range of the GM value. The partial similarity $S_{PC}(x)$ and $S_G(x)$ are combined to obtain the similarity $S_L(x)$ of $f_1(x)$ and $f_2(x)$, which is expressed as follows:

$$S_L(\mathbf{x}) = [S_{PC}(\mathbf{x})]^{\alpha} \cdot [S_G(\mathbf{x})]^{\beta}, \tag{4}$$

where α and β are parameters to adjust the relative importance of PC and GM features. Overall, the FSIM index for grayscale or luminance components of color images is formally expressed as follows:

$$\mathrm{FSIM} = \frac{\sum_{\mathbf{x}\in\Omega} \mathrm{S_L}(\mathbf{x}) \cdot \mathrm{PC_m}(\mathbf{x})}{\sum_{\mathbf{x}\in\Omega} \mathrm{PC_m}(\mathbf{x})}, \tag{5}$$

where Ω denotes the entire spatial domain of a image, PC_m is defined as $max(PC_1(x), PC_2(x))$ to weight the importance of $S_L(x)$. However, this definition does not take color IQA into account.

In contrast to the preceding metric, which was designed for single-channel images, the authors proposed an approach for evaluating the similarity between color images [47]. This necessitates the conversion of the RGB channels to YIQ space [47], after which the similarity between the I and Q chromatic components is computed using the following equations:

$$\begin{aligned} S_I(\mathbf{x}) &= \frac{2I_1(\mathbf{x}) \cdot I_2(\mathbf{x}) + T_3}{I_1^2(\mathbf{x}) + I_2^2(\mathbf{x}) + T_3}, \\ S_Q(\mathbf{x}) &= \frac{2Q_1(\mathbf{x}) \cdot Q_2(\mathbf{x}) + T_4}{Q_1^2(\mathbf{x}) + Q_2^2(\mathbf{x}) + T_4}, \end{aligned} \tag{6}$$

where T_3 and T_4 are positive constants. Then, the similarity of I and Q can be modified to express the chrominance similarity:

$$S_C(\mathbf{x}) = S_I(\mathbf{x}) \cdot S_Q(\mathbf{x}). \tag{7}$$

Overall, the FSIM index extended for color images is defined by:

$$\mathrm{FSIM}_C = \frac{\sum_{\mathbf{x}\in\Omega} S_L(\mathbf{x}) \cdot [S_C(\mathbf{x})]^{\lambda} \cdot PC_m(\mathbf{x})}{\sum_{\mathbf{x}\in\Omega} PC_m(\mathbf{x})}, \tag{8}$$

where λ is a constant that controls the importance of the chrominance components.

Another possible approach to compare images by similarity is to use a learned neural network model. A representative of this group is Learned Perceptual Image Patch Similarity [46]. This is a metric that uses deep features to compute similarity metrics. The distance between the reference x and the distorted patches is given by the following equation [46]:

$$d(x, x_0) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \left\| w_l \odot \left(\hat{y}^l_{hw} - \hat{y}^l_{0hw} \right) \right\|_2^2, \tag{9}$$

where W_l and H_l are the width and height of the l^{th} layer, and C_l is the number of channels. The extracted features for the layer l are described as $\hat{y}^l, \hat{y}^l_0 \in \mathbb{R}^{H_l \times W_l \times C_l}$. Prior to calculating the L2 norm distance, the activations are scaled by $w^l \in \mathbb{R}^{C_l}$. Otherwise, the features are unit normalized in the channel dimension. The authors of this metric have proposed two models, the first of which is based on VGG and AlexNet architectures [46]. For training, the authors have proposed a dataset containing various image distortions, including those resulting from ghosting and compression.

In [49] another metric, called Fréchet Inception Distance, was introduced, which can be considered a widely used metric for evaluating the quality of generated images, usually generated by GAN. The mathematical representation of FID is based on the squared Wasserstein distance between two multidimensional Gaussian distributions, obtained by analyzing the activations of a pre-trained Inception v3 model without a classification layer. This pre-trained model was trained on a large dataset, resulting in the model's ability to capture different aspects of image features.

Each distribution represents the mean m and the covariance C of the activations for an image. Then (m, C) represents the distribution of the generated images and (m_w, C_w) represents the ground truth. The FID distance is given by:

$$d^2((\boldsymbol{m}, \boldsymbol{C}), (\boldsymbol{m}_w, \boldsymbol{C}_w)) = \|\boldsymbol{m} - \boldsymbol{m}_w\|_2^2 + \mathrm{Tr}\left(\boldsymbol{C} + \boldsymbol{C}_w - 2(\boldsymbol{C}\boldsymbol{C}_w)^{1/2}\right), \tag{10}$$

where T_r is the trace operator.

From a biometric perspective, the metrics used to measure visual similarity between images are less relevant than evaluating FR performance on reconstructed images. Our experiments are primarily focused on evaluating the impact of reconstruction on FR.

To perform face detection, we choose the RetinaFace detector [51]. Two selected face recognition neural networks were used to generate embeddings: Arcface, which was chosen as a representative of angle-based FR approaches, and QMagFace, which is based on the MagFace neural network and uses a quality-aware metric to match two embeddings. This algorithm was chosen under the assumption that face quality can be affected by reconstruction algorithms. In this study, ArcFace uses Resnet50 [52], and QMagFace uses IResnet100 [53] to extract features from a face image. Overall, the L2 norm distance is used to compare the embeddings provided by ArcFace, and a custom similarity metric is used for QMagFace with coefficients $\alpha = 0.077428$ and $\beta = 0.125926$. Unlike [10], the face images were aligned by mapping the five detected face keypoints to reference points through affine transformations [54].

4.3 Evaluation of image quality assessment

Our approach to testing individual modifications against the base network architecture (base net) resulted in filtering out insufficient modifications. Table 1 compares each modification against the baseline model, clearly assessing their impact on model performance. Firstly, we utilized only PSNR and SSIM image quality assessment metrics to compare our designs with existing solutions presented in Table 3.

A novel model was developed that demonstrated superior quality in the generated images compared to a model based on the base net. Furthermore, these modifications were applied to U-net architectures from the Keras library to evaluate their effectiveness in facial image reconstruction tasks. The results are summarized in Table 2. For both the PSNR and SSIM metrics, the V-net achieved the best results.

In order to conduct a more detailed investigation, we selected the most recent models from previous evaluations. When utilizing existing solutions, we endeavored to leverage

Table 1 A comparison of influence models with individual modifications against a model based on the base net [10]

Model	PSNR	SSIM
Base net	22.641	0.710
No dense layers	28.814	0.893
Strided convolutions	24.015	0.751
Skip connections	25.227	0.916

Table 2 A comparison of image quality metrics (PSNR and SSIM) for various modified U-Net architectures from the Keras U-Net library

Base model	PSNR	SSIM
Swin-UNet (2022) [55]	33.405	0.965
Unet3plus (2020) [56]	33.714	0.971
U-Net (2015) [57]	33.737	0.969
V-Net (2016) [58]	**34.326**	**0.972**
U-net++ (2018) [59]	29.744	0.905

Table 3 Comparison of performance of our model with existing solutions [10] on CelebA-C dataset. The first three rows show the performance of our final models

Model	PSNR	SSIM
Modified U-Net	33.737	0.969
Modified V-Net	34.326	0.972
Modified base net - MBNet	33.659	0.969
Generative face completion (2017) [5]	19.500	0.784
G-NST (2020) [3]	29.655	0.937
DFNet (2019) [4]	31.662	0.965
CTSDG (2021) [13]	27.920	0.925
AOT-GAN (2020) [15]	23.606	0.905
LaMa (2022) [7]	33.209	0.969

their pre-trained models to achieve the most accurate outcomes. In cases where a given solution lacked a pre-trained model on the CelebA dataset, we fine-tuned the model on that specific dataset. The parameters utilized in the fine-tuning were consistent with those described in the original paper, ensuring the highest degree of consistency. We then proceeded to compute the reconstruction on these models for two different mask datasets: our CelebA-C and NVIDIA Irregular Mask Dataset. We further computed more detailed metrics such as MS-SSIM, FSIM, LPIPS, and FID for the images reconstructed on these datasets. All of those metrics were initialized with default parameters as described in their papers. The performance of each solution displayed using these metrics on both datasets can be seen in Tables 4 and 5.

A comparative analysis of the performance of the above architectures, including existing solutions, is presented in Tables 3, 4 and 5. We used standard IQA metrics to compare the quality of our designs with other approaches. As we can see from the results, all

Table 4 Additional metrics comparison of existing solutions trained and tested on CelebA-C dataset

Model	MS-SSIM	FSIM	LPIPS	FID
Modified V-Net	**0.984**	0.905	0.026	1.704
CTSDG	0.934	0.881	0.066	16.159
AOT-GAN	0.878	0.828	0.075	14.282
LaMa	0.980	**0.935**	**0.015**	**0.508**

Table 5 Additional metrics comparison of existing solutions trained and tested on NVIDIA Irregular Mask Dataset

Model	MS-SSIM	FSIM	LPIPS	FID
Modified V-Net	0.952	0.822	0.068	4.161
CTSDG	0.911	0.824	0.106	13.783
AOT-GAN	0.771	0.733	0.215	29.963
LaMa	0.952	0.872	0.038	1.990

of our designs surpassed the performance of existing solutions when comparing PSNR and SSIM metrics, with the modified V-Net leading the way. A comparison of our best model with other solutions using MS-SSIM, FSIM, LPIPS, and FID metrics on our CelebA-C, see Subsection 3.2, as shown in Table 4, reveals that our model outperforms the majority of the solutions and is comparable to the LaMa model. In addition, our model was compared to other solutions using the NVIDIA irregular mask dataset. The comparison results are presented in Table 5. As can be observed, the results are comparable to those presented in Table 4. The results demonstrated that our model outperformed the majority of existing solutions. A comparative analysis of our model with LaMa indicates a slight decrease in performance. However, our model exhibits several advantages. Firstly, it is half the size of LaMa, comprising 26 million parameters, whereas LaMa has 51 million parameters. Secondly, our model uses only the corrupted image as input, whereas LaMa requires both the corrupted image and the mask to reconstruct.

For a visual representation of the outputs from our models and a comparison with ground truth, please refer to Fig. 5. Moreover, Fig. 6 illustrates the difference in reconstruction quality between our Modified V-Net model and the LaMa model. Additionally, the figure illustrates the difference in reconstruction quality between the CelebA-C dataset and the NVIDIA irregular mask dataset.

4.4 Evaluation of face recognition performance

The current state of image reconstruction solutions employs IQA metrics to assess the accuracy of image reconstruction. However, this approach may be inadequate for biometric purposes. This paper primarily aims to investigate the impact of facial image reconstruction on facial recognition. The following results were obtained to address this question.

Initially, based on previous findings, we selected our and LaMa models to assess the impact of image reconstruction on the accuracy of face recognition algorithms. The subsequent experiments were conducted on the CelebA-C dataset and a dataset

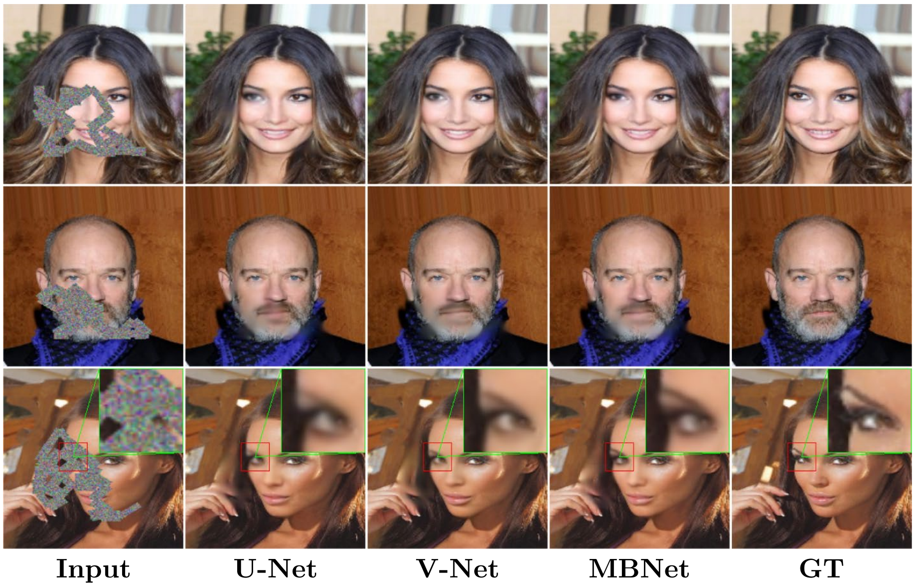


Fig. 5 Comparison of three different generators for generating damaged facial parts. We compare our combined model with tested modifications implemented into U-Net and V-Net architectures [10]

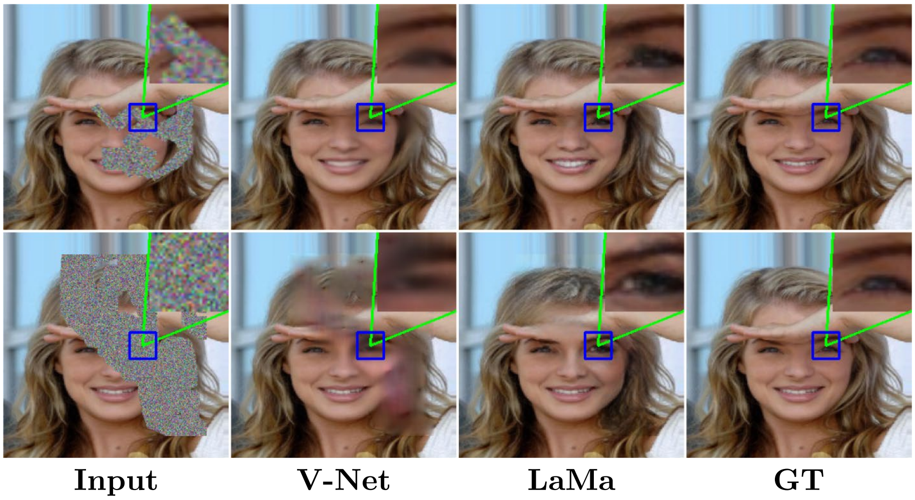


Fig. 6 Comparison of results on different mask datasets. The first row shows LaMa and Modified V-Net reconstruction on the CelebA-C dataset, and the second row shows reconstruction on the NVIDIA irregular masks dataset

generated using NVIDIA masks. Firstly, experiments were conducted utilizing the CelebA-C dataset. The results are presented in Table 6. It is obvious that facial image reconstruction has a significant impact on facial recognition. From the perspective of the false positive rate (FPR), facial recognition of damaged images shows a higher value. After reconstruction, the true positive rate (TPR) shows a slight decrease compared to undamaged facial images. Furthermore, a comparison of the LaMa model with the Modified V-Net model reveals that, despite the LaMa model's superior performance in terms of the metrics presented in Table 4, it exhibited a notable

Table 6 ArcFace and QMagFace performance comparison on CelebA-C dataset to show how facial image reconstruction using V-Net and LaMa models affects facial recognition accuracy

	ArcFace			**QMagFace**		
	ACC	**FPR**	**TPR**	**ACC**	**FPR**	**TPR**
Original	0.9864	0.0039	0.9029	0.9950	0.0033	0.9565
Damaged	0.9225	0.0172	0.3983	0.9620	0.0101	0.7193
Modified V-Net	0.9747	0.0074	0.8194	0.9875	0.0035	0.9089
Inpainted V-Net	0.9747	0.0074	0.8194	0.9876	0.0033	0.9089
LaMa	0.9408	0.0332	0.8958	0.9632	0.0148	0.9250

Table 7 ArcFace and QMagFace performance comparison on NVIDIA irregular mask dataset to show how facial image reconstruction using V-Net and LaMa models affects facial recognition accuracy

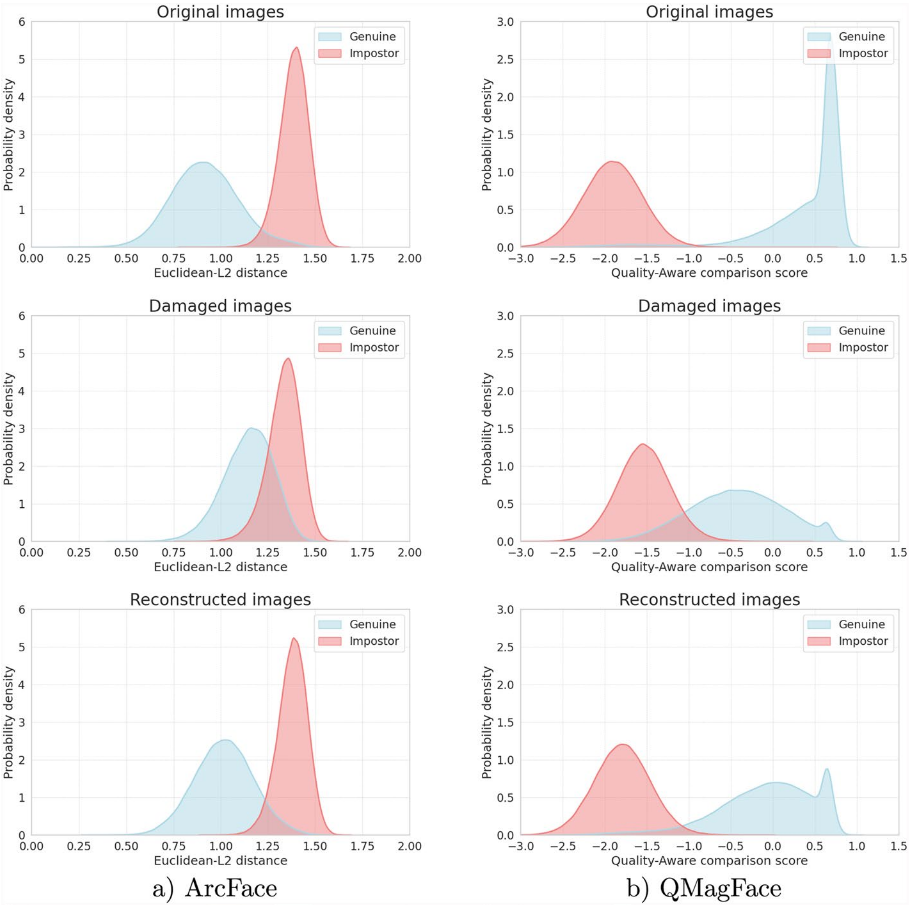
	ArcFace			**QMagFace**		
	ACC	**FPR**	**TPR**	**ACC**	**FPR**	**TPR**
Original	0.9864	0.0039	0.9029	0.9950	0.0033	0.9565
Damaged	0.7877	0.1183	0.6245	0.8727	0.0602	0.7562
Modified V-Net	0.9090	0.0411	0.8223	0.9363	0.0224	0.8647
LaMa	0.9210	0.0454	0.8628	0.9379	0.0276	0.8780

deficiency in this experiment. This indicates that, from a biometrics perspective, the IQA metrics are inadequate and that a greater emphasis should be placed on facial image reconstruction from a face recognition perspective.

As part of the experiments, we also aimed to determine whether our solution for facial image reconstruction focuses only on the missing parts and does not damage the known areas, which could lead to a reduction in recognition accuracy. To accomplish this, we took the initially corrupted area from the reconstructed photo and inserted it back into the original photo. A comparison of the Modified V-Net and Inpainted V-Net in Table 6 shows little to no difference in accuracy between facial recognition systems when using reconstructed or inpainted images. Therefore, there is no need to remove the reconstructed area from the generated image and reinsert it into the original.

When compared to the more extensive corruptions from the NVIDIA irregular mask database, as shown in Table 7, it is evident that the LaMa model exhibits an improvement in accuracy. However, it is crucial to highlight that despite this, when we examine the false positive rate, it is observed to be worse than that of the Modified V-Net model and across all experiments.

In Fig. 7, we utilized the modified V-Net model to reconstruct occluded images and calculated the probability density function to illustrate the improvement in FR accuracy relative to occluded images and the deviation from ideal conditions represented by original images. It should be noted that, based on previous results, the QMagFace model performs significantly more effectively in facial recognition on images of reduced quality than the ArcFace model.



a) ArcFace b) QMagFace

Fig. 7 Genuine and impostor score distributions obtained using ArcFace (a) and QMagFace (b). Distributions were obtained for original, damaged, and reconstructed data. Graphs from top to bottom accordingly

4.5 Image enhancement

Additionally, we investigated whether image enhancement of reconstructed images could improve FR results. We used pre-trained models from advanced image enhancement techniques—CodeFormer [16], DifFace [17], GFPGAN [18], and DFD-Net [19]—to improve the quality of the reconstructed images. All mentioned methods were trained on the FFHQ dataset [20]. Using those methods allowed us to directly assess the impact of existing state-of-the-art image enhancement techniques on identity changes in enhanced images. By applying these pre-trained enhancement models to our reconstructed images, we aimed to refine the visual details of the images and improve their suitability for face recognition tasks.

Examples of enhanced images, including comparisons with the original images, are shown in Fig. 8. Previous evaluations have shown that the IQA metrics and face recognition metrics are not correlated. Therefore, we have only utilized the FR metrics in this comparison, allowing us to observe how these enhancement methods affect the accuracy of the FR algorithms. In Table 8, we can see that the image quality enhancement methods do not contribute to increasing the recognition accuracy of the FR algorithms. Although the performance of the DFDNet output is comparable to that of the unenhanced image, none of the TPR, FPR, or ACC metrics showed superior

Table 8 The enhanced images were also evaluated by FR algorithms to determine whether the enhancement techniques can increase their accuracy

	ArcFace			QMagFace		
	ACC	**FPR**	**TPR**	**ACC**	**FPR**	**TPR**
CodeFormer	0.9667	0.0104	0.7710	0.9742	0.0032	0.9183
DFDNet	0.9717	0.0093	0.8069	0.9855	0.0008	0.9533
DifFace	0.9454	0.0156	0.6074	0.9433	0.0111	0.7419
GFPGAN	0.9653	0.0096	0.7475	0.9729	0.0032	0.9076

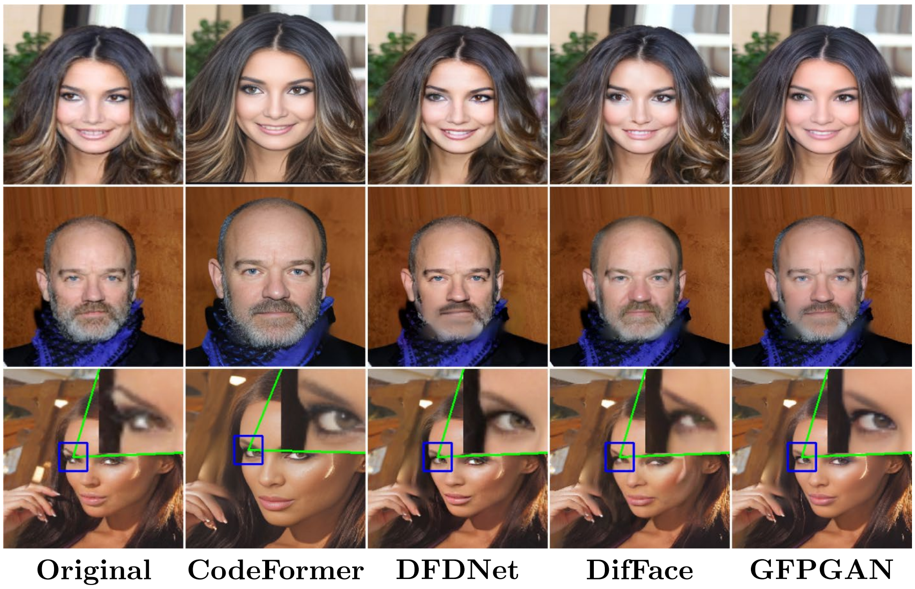


Fig. 8 Comparison of the results of four different image enhancement methods

performance. Therefore, the use of these enhancers does not have a positive impact on FR performance.

5 Conclusion

This paper primarily assesses the impact of neural network-based facial image reconstruction on facial recognition. While image quality assessment (IQA) metrics are commonly used to evaluate algorithm performance, the evaluation based on facial recognition is of greater importance from a biometric perspective. The results demonstrate that these two approaches do not correlate.

In order to evaluate the performance of our models for the reconstruction of facial images, experiments were conducted with the existing solution and with image enhancers. The model proposed in this study is based on advanced generative adversarial networks (GANs). With regard to the assumptions underlying the theoretical framework, the results are significantly affected by the training process. The approaches employed in this study were trained from scratch using a modified CelebA dataset, which was named CalebA-C. With regard to the existing solutions, we employed pre-trained models that were subsequently fine-tuned with the CalebA-C dataset.

Experiments conducted on the test subset of the CalebA-C dataset demonstrate that our models yield significant enhancements. In this case, the facial recognition metrics demonstrate superior results compared to the second model, LaMa. However, when the dataset was generated using NVIDIA masks, the LaMa model exhibited a higher accuracy than our approach. It is noteworthy that the LaMa model demonstrated a higher false acceptance rate in all cases than our proposed model based on V-Net. A higher false positive rate can potentially compromise the security of a biometric system.

Furthermore, a study was conducted to determine the influence of image enhancement on the accuracy of face-based recognition. The results indicated that the application of image enhancers after the reconstruction phase was not beneficial. The highest accuracy was achieved with DFDNet, although the results were slightly worse than before the enhancement.

Another noteworthy outcome of the experiments is that the QMagFace algorithm exhibits superior performance in facial recognition on corrupted images in comparison to the ArcFace model.

In conclusion, GAN-based approaches present a promising avenue for facial image reconstruction. However, our findings indicate that the use of facial recognition metrics, as opposed to IQA metrics, is of paramount importance when assessing performance.

Future research could focus on approaches that utilize identity loss, which would require the assumption of the perseverance of identity for reconstruction. One such approach could be the diffusion model, which could be used to solve this task.

Acknowledgements
This work was supported by the internal project of Brno University of Technology FIT-S-23-8151. Computational resources were provided by the e-INFRA CZ project (ID:90254), supported by the Ministry of Education, Youth and Sports of the Czech Republic.

Author contributions
FP designed and executed experiments, a major contributor to the manuscript. TG helped with the experiment design and writing the manuscript. KM oversaw the experiment design and execution.

Funding
This work was supported by the internal project of Brno University of Technology FIT-S-23-8151.

Data availability
The dataset used in our experiments is publicly available at: CelebA: https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html NVIDIA: https://research.nvidia.com/labs/adlr/publication/partialconv-inpainting/ Models used for image enhancement are available at: https://github.com/sczhou/Awesome-Face-Restoration

Declarations

Competing interests
The authors declare that they have no conflict of interest.

Received: 30 April 2024 Accepted: 5 May 2025
Published online: 20 May 2025

References

1. B. Lahasan, S.L. Lutfi, R. San-Segundo, A survey on techniques to handle face recognition challenges: occlusion, single sample per subject and expression. Artif. Intell. Rev **52**(2), 949–979 (2017). https://doi.org/10.1007/s10462-017-9578-y
2. D. Zeng, R. Veldhuis, L. Spreeuwers, A survey of face recognition techniques under occlusion. IET Biom **10**(6), 581–606 (2021)
3. Y. Zhao, J. Hu, X. Zhang, Face restoration based on gans and nst. In: Proceedings of the 2020 5th International Conference on Mathematics and Artificial Intelligence. ICMAI 2020, pp. 198–203. Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3395260.3395304
4. X. Hong, P. Xiong, R. Ji, H. Fan, Deep Fusion Network for Image Completion (2019)

5. Y. Li, S. Liu, J. Yang, M.-H. Yang, Generative face completion. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5892–5900 (2017). https://doi.org/10.1109/CVPR.2017.624
6. S.N. Uddin, Y.J. Jung, Global and local attention-based free-form image inpainting. Sensors **20**(11), 3204 (2020)
7. R. Suvorov, E. Logacheva, A. Mashikhin, A. Remizova, A. Ashukha, A. Silvestrov, N. Kong, H. Goka, K. Park, V. Lempitsky, Resolution-robust large mask inpainting with fourier convolutions. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 2149–2159 (2022)
8. A. Hore, D. Ziou, Image quality metrics: Psnr vs. ssim. In: 2010 20th International Conference on Pattern Recognition, pp. 2366–2369 (2010). IEEE
9. Z. Wang, E. Simoncelli, A. Bovik, Multis. Struct. Sim. Imag. Qual. Assess. **2**, 1398–14022 (2003). https://doi.org/10.1109/ACSSC.2003.1292216
10. F. Pleško, T. Goldmann, K. Malinka, Facial image reconstruction and its influence to face recognition. In: 2023 International Conference of the Biometrics Special Interest Group (BIOSIG), pp. 1–5 (2023). https://doi.org/10.1109/BIOSIG58226.2023.10346000
11. X. Gao, M. Nguyen, W.Q. Yan, Face image inpainting based on generative adversarial network. In: 2021 36th International Conference on Image and Vision Computing New Zealand (IVCNZ), pp. 1–6 (2021). https://doi.org/10.1109/IVCNZ54163.2021.9653347
12. M.N. Yeğin, M.F. Amasyalı, Theoretical research on generative diffusion models: an overview. arXiv preprint arXiv:2404.09016 (2024)
13. X. Guo, H. Yang, D. Huang, Image inpainting via conditional texture and structure dual generation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 14134–14143 (2021)
14. H.-W. Dong, Y.-H. Yang, Towards a deeper understanding of adversarial losses under a discriminative adversarial network setting. arXiv preprint arXiv:1901.08753 (2019)
15. Y. Zeng, J. Fu, H. Chao, B. Guo, Aggregated contextual transformations for high-resolution image inpainting. IEEE Trans. Vis. Comput. Gr. **29**(7), 3266–3280 (2022)
16. S. Zhou, K.C.K. Chan, C. Li, C.C. Loy, Towards Robust Blind Face Restoration with Codebook Lookup Transformer (2022)
17. Z. Yue, C.C. Loy, DifFace: Blind Face Restoration with Diffused Error Contraction (2023)
18. X. Wang, Y. Li, H. Zhang, Y. Shan, Towards Real-World Blind Face Restoration with Generative Facial Prior (2021)
19. X. Li, C. Chen, S. Zhou, X. Lin, W. Zuo, L. Zhang, Blind Face Restoration via Deep Multi-scale Component Dictionaries (2020)
20. T. Karras, S. Laine, T. Aila, A Style-Based Generator Architecture for Generative Adversarial Networks (2019)
21. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention Is All You Need (2023)
22. A. Tamkin, M. Taufeeque, N.D. Goodman, Codebook Features: Sparse and Discrete Interpretability for Neural Networks (2023)
23. Gao Huang, Hua Lan, Deep learning for super-resolution in a field emission scanning electron microscope. Ai **1**(1), 9 (2019)
24. Y. Taigman, M. Yang, M. Ranzato, L. Wolf, Deepface: Closing the gap to human-level performance in face verification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701–1708 (2014)
25. F. Schroff, D. Kalenichenko, J. Philbin, Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 815–823 (2015)
26. W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, L. Song, Sphereface: Deep hypersphere embedding for face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 212–220 (2017)
27. F. Wang, J. Cheng, W. Liu, H. Liu, Additive margin softmax for face verification. IEEE Signal Process. Lett. **25**(7), 926–930 (2018)
28. H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, W. Liu, Cosface: Large margin cosine loss for deep face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5265–5274 (2018)
29. J. Deng, J. Guo, J. Yang, N. Xue, I. Kotsia, S. Zafeiriou, Arcface: Additive angular margin loss for deep face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **44**(10), 5962–5979 (2022). https://doi.org/10.1109/tpami.2021.3087709
30. Y. Zhong, W. Deng, J. Hu, D. Zhao, X. Li, D. Wen, Sface: Sigmoid-constrained hypersphere loss for robust face recognition. IEEE Trans. Image Process. **30**, 2587–2598 (2021)
31. Q. Meng, S. Zhao, Z. Huang, F. Zhou, Magface: A universal representation for face recognition and quality assessment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14225–14234 (2021)
32. G.B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Tech. Rep. **1**, 7–49 (2007)
33. S. Sengupta, J.C. Cheng, C.D. Castillo, V.M. Patel, R. Chellappa, D.W. Jacobs, Frontal to profile face verification in the wild. In: IEEE Conference on Applications of Computer Vision (2016)
34. S. Moschoglou, A. Papaioannou, C. Sagonas, J. Deng, I. Kotsia, S. Zafeiriou, Agedb: the first manually collected, in-the-wild age database. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop, 2017; 2 5
35. T. Zheng, W. Deng, Cross-pose lfw: A database for studying cross-pose face recognition in unconstrained environments. Tech. Rep. **5**, 5 (2018)
36. P. Terhörst, M. Ihlefeld, M. Huber, N. Damer, F. Kirchbuchner, K. Raja, A. Kuijper, QMagFace: Simple and accurate quality-aware face recognition. CoRR arXiv:abs/2111.13475 (2021)
37. C.R. Kavita, Face recognition challenges and solutions using machine learning. Int. J. Intell. Syst. Appl. Eng. **10**(3), 471–476 (2022)
38. A. Radford, L. Metz, S. Chintala, Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks (2016)
39. M. Tripathi, Facial image denoising using autoencoder and unet. Herit. Sustain. Dev **3**(2), 89–96 (2021). https://doi.org/10.37868/hsd.v3i2.71

40. J.T. Springenberg, A. Dosovitskiy, T. Brox, M. Riedmiller, Striving for Simplicity: The All Convolutional Net (2015)
41. Y. Sha, Keras-unet-collection. GitHub (2021). https://doi.org/10.5281/zenodo.5449801
42. K. Wang, C. Gou, Y. Duan, Y. Lin, X. Zheng, F.-Y. Wang, Generative adversarial networks: introduction and outlook. IEEE/CAA J. Autom. Sinica **4**(4), 588–598 (2017)
43. I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative Adversarial Networks (2014)
44. Z. Liu, P. Luo, X. Wang, X. Tang, Deep learning face attributes in the wild. In: Proceedings of International Conference on Computer Vision (ICCV) (2015)
45. G. Liu, F.A. Reda, K.J. Shih, T.-C. Wang, A. Tao, B. Catanzaro, Nvidia irregular mask dataset. (2018). https://nv-adlr.github.io/publication/partialconv-inpainting
46. R. Zhang, P. Isola, A.A. Efros, E. Shechtman, O. Wang, The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 586–595 (2018)
47. L. Zhang, L. Zhang, X. Mou, D. Zhang, Fsim: A feature similarity index for image quality assessment. IEEE Trans. Imag. Process. **20**(8), 2378–2386 (2011)
48. Z. Wang, E.P. Simoncelli, A.C. Bovik, Multiscale structural similarity for image quality assessment. In: The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003; 2; 1398–1402
49. M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, Gans trained by a two time-scale update rule converge to a local nash equilibrium. Adv. Neural. Inf. Process. Syst. **30**, 1 (2017)
50. A. Hore, D. Ziou, Image quality metrics: Psnr vs. ssim. In: 2010 20th International Conference on Pattern Recognition, pp. 2366–2369 (2010). IEEE
51. J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia, S. Zafeiriou, Retinaface: Single-stage dense face localisation in the wild. arXiv preprint arXiv:1905.00641 (2019)
52. K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition (2015)
53. I.C. Duta, L. Liu, F. Zhu, L. Shao, Improved Residual Networks for Image and Video Recognition (2020)
54. B. Adhikari, X. Ni, E. Rahtu, H. Huttunen, Towards a real-time facial analysis system. In: 2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP), pp. 1–6 (2021). IEEE
55. H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, M. Wang, Swin-unet: Unet-like pure transformer for medical image segmentation. In: European Conference on Computer Vision, pp. 205–218 (2022). Springer
56. H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, J. Wu, Unet 3+: A full-scale connected unet for medical image segmentation. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1055–1059 (2020). IEEE
57. O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-assisted intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp. 234–241 (2015). Springer
58. F. Milletari, N. Navab, S.-A. Ahmadi, V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation (2016). https://arxiv.org/abs/1606.04797
59. Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4, pp. 3–11 (2018). Springer

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.