MIPGAN - Generating Robust and High Quality Morph Attacks Using Identity Prior Driven GAN

Haoyu Zhang†, Sushma Venkatesh†, Raghavendra Ramachandra†
Kiran Raja†, Naser Damer‡, Christoph Busch†
† Norwegian University of Science and Technology (NTNU), Norway
‡Fraunhofer Institute for Computer Graphics Research IGD, Darmstadt, Germany.
{sushma.venkatesh; haoyu.zhang; raghavendra.ramachandra;kiran.raja;christoph.busch} @ntnu.no
{naser.damer}@igd.fraunhofer.de

Abstract—Face morphing attacks target to circumvent Face Recognition Systems (FRS) by employing face images derived from multiple data subjects (e.g., accomplices and malicious actors). Morphed images can be verified against contributing data subjects with a reasonable success rate, given they have a high degree of identity resemblance. The success of the morphing attacks is directly dependent on the quality of the generated morph images. We present a new approach for generating robust attacks extending our earlier framework for generating face morphs. We present a new approach using an Identity Prior Driven Generative Adversarial Network, which we refer to as *MIPGAN (Morphing through Identity Prior driven GAN)*. The proposed MIPGAN is derived from the StyleGAN with a newly formulated loss function exploiting perceptual quality and identity factor to generate a high quality morphed face image with minimal artefacts and with higher resolution. We demonstrate the proposed approach's applicability to generate robust morph attacks by evaluating it against both commercial and deep learning based Face Recognition System (FRS) and demonstrate the success rate of attacks. Extensive experiments are carried out to assess the FRS's vulnerability against the proposed morphed face generation technique on three types of data such as digital images, re-digitized (printed and scanned) images, and compressed images after re-digitization from newly generated *MIPGAN Face Morph Dataset*. The obtained results demonstrate that the proposed approach of morph generation poses high threat to FRS.

Index Terms—Morph Attacks, GAN, Attack detection, Face Recognition, Vulnerability, Deep Learning

1 Introduction

Face Recognition Systems (FRS) have provided ubiquitous ways of verifying the identity in many applications. FRS have been used in everyday applications from low-security applications such as smartphone unlocking to high-security applications such as identity verification in border control processes. Each of the applications mandate a chosen way of enrolment to FRS where either a supervised enrolment is carried out (for instance in on-boarding at bank premises) or unsupervised enrolment is requested (on-boarding for banking applications from home). While it provides a high degree of flexibility and convenience to users to initiate an enrolment process in an unsupervised manner, this potentially leads to a security risk: Without supervision, a data subject enrolling into the FRS can submit a face image which is manipulated, a printed face image, an image displayed from an electronic screen (e.g., iPad) or a silicone latex face mask [2]. In order to mitigate such attacks at the enrolment level, it is therefore essential to have a robust attack detection mechanism. While a number of works have been proposed on both conducting such attacks and detecting the attacks in a robust manner for printed attacks, display attacks and mask attacks in recent years, we focus our work on a new kind of attack referred popularly as *Morphing Attack*.

Haoyu Zhang, Sushma Venkatesh and Raghavendra Ramachandra contributed equally.

Face morphing is the process of combining two or more face images to generate a single face image that can resemble visually to all the contributing face images to a greater degree [3]. A good quality morphed face image is also effective in verifying against all contributing subjects by obtaining a comparison score that exceeds the pre-determined threshold (i.e., passes through FRS) [3], [4], [5], [6]. While morphing can be conducted using multiple face images of different subjects, the effectiveness of morphed images is reported when the face images of similar ethnicity, gender and age group are considered [6], [7], [8]. This is primarily due to the fact that a morphed image should not only defeat the FRS but should also provide high visual similarity, in order to convince a human expert in a visual comparison process.

Face morphing attacks threaten FRS due to the current practices in the ID-document application process, where the biometric enrolment is carried out in an unsupervised manner in many countries. Countries like the UK and New Zealand allow citizens to upload a digital face image for various applications such as passport renewal [9] and visa application [10]. The capture process for such images is unsupervised. In a similar manner, many Asian countries and European countries (e.g. in The Netherlands [11]) request the applicant to submit a scanned face image for passport/visa/identity-card applications. Given that the images are captured and submitted in an unsupervised setting, the applicant has vast opportunities to upload a

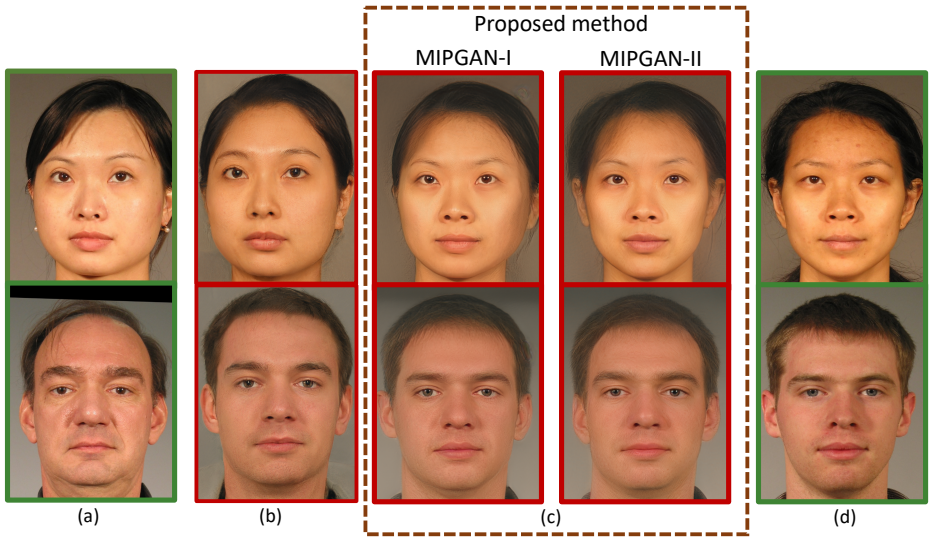


Fig. 1: Results from StyleGAN based face morphing [1] and the proposed MIPGAN (a) Contributing subject 1 (b) StyleGAN [1] (c) Proposed method (d) Contributing subject 2

morphed image with malicious intent underlining the need for robust Morphing Attack Detection (MAD) mechanisms.

1.1 Related Works on Face Morph Generation

While morphing attacks have been studied in recent years, most of the attacks are conducted using the morphed images created using facial landmarks-based approaches needing high a degree of supervision to first extract the facial landmarks, thereupon align them and then finally blend them to generate morphs. The common set of procedures for warping/blending includes Free Form Deformation (FFD) [12] [13], Deformation by moving least squares [14], deformation based on mass spring [15], Bayesian framework based morphing [16] and Delaunay triangulation based morphing [17] [18] [19] [20] [21]. Due to inadvertent artefacts caused by pixel/region-based morphing, the images need additional work in refining the signal to create highly realistic morph images. A set of post processing steps are usually included as illustrated in number of works [20] [22] [23]. Generally, some set of post processing techniques such as image smoothing, image sharpening, edge correction, histogram equalisation, manual retouching, image enhancement to improve the brightness and contrast are used to eliminate the artefacts generated during the morphing process. In a parallel direction, morphed face images can also be generated using landmarks-based methods available in open-source resources like GIMP/GAP and OpenCV. Morphs generated using GIMP/GAP technique are more efficient with respect to a good quality of the resulting image (i.e., less noticeable artefacts) as pixels are aligned manually. Despite the minimal amount of effort needed for creating morphs using such approaches, a significant amount of effort needs to be dedicated to correcting artefacts. Additionally, commercial solutions like Face Fusion [24] and FantaMorph [25] can also generate good quality morphed images with limited manual intervention. Although some steps can be excluded in creating the morphs, it is very critical to meet the face image quality standards laid out by the International Civil Aviation Organization (ICAO) [26] [27] for electronic Machine Readable Travel Document (eMRTD) and deployment of biometric identification applications.

1.2 GAN Based Face Morph Generation

In an attempt to overcome the cumbersome efforts of manually creating (semi-automated) morphed images, a fully automated approach using a Generative Adversarial Network (GAN) was proposed by Damer et al. [28]. Unlike the supervision required in extracting and aligning the face images in a manual morphing process, GAN-based techniques synthesise morphed images directly by merging two facial images in the latent space. In the work by Damer et al. [28], the proposed MorGAN architecture for morph generation basically employed a generator constituting encoders, decoders and a discriminator. The generator was trained to generate images with the dimension 64×64 pixels which is a key limiting factor of the attack, as most commercial FRS will reject images that do not meet the ICAO standard that requires a minimum Inter-Eye Distance (IED) of 90 pixels. The empirical evaluation of generated morph images using MorGAN in a vulnerability analysis against two commercial FRS indicated that those MorGAN morphs fail to meet both quality standards and the verification threshold of the FRS [1]. Motivated to address the deficiency of the MorGAN architecture, in our recent work [1] [1] we proposed an approach based on the StyleGAN architecture [29] to increase the spatial dimension to 1024×1024 and thus to improve the face quality. Unlike the previous approach of MorGAN [28], StyleGAN [1] achieves better spatial resolution by embedding the images in the intermediate latent space. With the increased spatial dimension of resulting

1. The preliminary work results were published at IWBF-2020 in April, 2020.

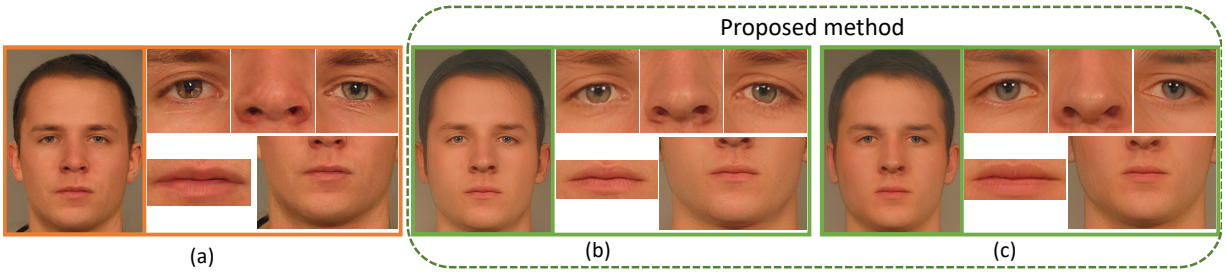


Fig. 2: Details of segmented components in morphs generated by earlier method based on StyleGAN [1] and proposed MIPGAN (a) StyleGAN [1] (b) MIPGAN-I (c) MIPGAN-II.

morphed images from our recently proposed architecture, we not only demonstrated that the images meet quality standards but also have a reasonable success rate when attacking commercial FRS [1].

1.3 Limitations of GAN Based Face Morph Generation and Our Contributions

While our earlier work [1] indicated that better GAN architectures could result in superior quality morphs and could attack an FRS in general, we also acknowledge the limited threats that exist for Commercial-Off-The-Shelf (COTS) FRS, as merely a subset of morphed images was accepted. Only approximately 50% of the generated morph images were verified successfully against probe images from a contributing subject. Thus the empirical evaluation in our earlier work has shown that the attack was yet not very effective [1] for a COTS FRS [30] and an open-source FRS based on ArcFace [31]. We must state that up to now FRS are not very vulnerable to GAN-based morphing attacks unlike to landmarks-based morphing attacks. With a clear introspection into this aspect, we notice that the resulting morphed images from our earlier work [1] does not retain a high degree of facial similarity to both contributing subjects. With lower similarity to contributing subjects in terms of facial structures, the FRS do not attribute a high comparison score, as anticipated. In other words, the missing enforcement of identity information of contributing subjects will lead to a high visual quality facial image but with lower face similarity to contributing face characteristics.

In an effort to make the attacks robust such that both subjects can be verified with a good success rate, in this work, we extend our previous architecture to generate morphs by including the identity priors before the generation of morphed faces. We now refer to this approach as *MIPGAN (Morphing through Identity Prior driven GAN)*. We propose two variants of our approach named as MIPGAN-I and MIPGAN-II based on the employed GAN being StyleGAN or StyleGAN2 respectively [29], [32]. With the inclusion of a new loss function in our proposed architecture, we increase the attack success rate against a commercial-off-the-shelf (COTS) FRS. Figure 1 shows the example of morphed face images generated using proposed MIPGAN along with outputs of both the variants. To further achieve superior quality face morphs, we also customize the newly designed loss function to account for ghosting and blurring artefacts in an end-to-end manner with no human or manual intervention eliminating the need for a high degree of interaction. As noted in Figure 2, the results from MIPGAN-I and MIPGAN-II is more coherent in retaining structural similarity as compared to our earlier architecture [1]. With the updated architecture to generate high-quality morphs which preserve both identity information and structural correspondence, we evaluate the applicability in creating stronger attacks by creating a large-scale dataset of morphed images by employing the face images derived from the FRGC-V2 face database [33]. The created dataset of 1270 bona fide images and 2500 morphed images is first evaluated to measure the attack success rate by verifying the morphed images against the contributing subjects using a commercial FRS from Cognitec [30]. Further to measuring the attack success rate from the digital images, we also extend our work by printing and scanning (re-digitizing) the dataset. We check the consistency of the attack success rate, unlike our earlier work which was limited to an investigation on digital images alone [1]. We also include the experiments on assessing the impact of compression (down to 15kb following ICAO guidelines) of printed and scanned face images that simulate the real-life e-passport application scenario. The key motivation to extend our work in this direction is, to mimic the passport application process that is operated in many European countries and Asian countries, which all accept printed-and-scanned face images in the application process for an identity document (e.g. passports).

With the extensive experimental results indicating a highly satisfactory attack success rate, we also evaluate a set of MAD algorithms to benchmark the detection capabilities. To this extent, we evaluate two state-of-the-art MAD approaches on digital morphed images, re-digitized and compressed morphed images after re-digitizing. Thus, we comprehensively cover the potential morphing attacks in the digital domain and the re-digitized domain. While we note the earlier works [1] arguing that attacks in the digital domain can be detected by studying the cues such as residual noise in morphing [34], patterns of noise from morphed images, histogram features of textures or the deep features [4], we also investigate the MAD capabilities for re-digitized images which do not exhibit the similar features (residual noise) as the print-scan process eliminates the digital cues and presents another set of variations. Specifically, given the nature of the dataset in which we have only a single suspected morphed image, for which we must determine

either the morph or the bona fide class, we resort to Single Image based MAD (S-MAD) approaches using two recent but robust approaches using hybrid and ensemble features [34], [35], [36], [37].

We therefore present a summary of contributions of this work as listed below:

- We present a novel approach of generating morphed face images through GAN architecture with enforced identity priors and a customized novel loss function to generate highly realistic images which we refer as *MIPGAN (Morphing through Identity Prior driven GAN).* We present two variants of the proposed approach for generating attacks with a high success rate.
- The proposed approach (both variants) is benchmarked to measure the attack success rate by verifying a COTS and deep learning based FRS through studying the vulnerability using a newly generated dataset from our proposed architecture which is referred as *MIPGAN Face Morph Dataset*.
- Human observer analysis for detecting morphs generated by the proposed and existing morphing attack methods is presented.
- Analysis of the perceptual quality metrics to illustrate the visual quality of the generated morph images is presented.
- Extensive experiments on three different data types such as (a) digital morphed images (b) print-scan morphed image (c) print-scan morphed images with compression are presented to cover the full spectrum of passport application process under morphing attacks.
- The generated images are also benchmarked against the existing MAD approaches both in digital form and the re-digitized form to provide the insights on detection challenges of SOTA approaches. We also present a generalizability study on MAD schemes by training one kind of morph generation and testing on a different kind of morph generation approach to indicate directions to future works.

In the rest of the paper, Section 2 describes the new architecture along with the newly designed loss function to generate high-quality morphs. Section 3 provides the details on the quantitative experiments indicating the vulnerability of FRS and the detection challenge. With the set of remarks and future works in this direction, we draw the conclusion in Section 5.

2 Proposed Morphed Face Generation

Figure 3 presents the block diagram of the proposed morphed face image generation using MIPGAN. The proposed method is based on the end-to-end optimisation using a new loss function that can preserve the identity of the generated morphed face image through enforced identity priors. The proposed MIPGAN framework is designed independently on two different GAN models based on StyleGAN [29] and StyleGAN2 [32] model. We refer the proposed scheme with StyleGAN as MIPGAN-I and with StyleGAN2 as MIPGAN-II respectively. Given the face images from the accomplice (I_1) (contributing subject 1) and the malicious (I_2) (contributing subject 2) data subjects, we predict the corresponding latent vectors $L_1^{'}$ and $L_2^{'}$ in the first step. In this work, we have employed the open-source pre-trained prediction models trained to predict the corresponding latent vector given an input image. Hence, L_1' and L_2' are predictions from the final output layer of the model, which is further reshaped. Since MIPGAN-I and MIPGAN-II are based on pre-trained StyleGAN [29] and StyleGAN2 [38] model respectively, we used two different open-source pre-trained models for prediction. Both of the prediction models employ $ResNet50$ [39] as backbone. The model for MIPGAN-I (StyleGAN) uses one convolution layer and two tree-connected layers [40] to map the output of $ResNet50$ into the final latent vector with the size of $(18, 512)$. In comparison, the model for MIPGAN-II (StyleGAN2) just uses one fully-connected layer to achieve the mapping. The predicted latent vectors thus provide the initialization for the morphed face generation that is obtained using a weighted linear average of $L_1^{'}$ and $L_2^{'}$ as follows:

$$L_M^{'} = \frac{w_1 * L_1^{'} + w_2 * L_2^{'}}{2}, \tag{1}$$

where w_1 and w_2 indicate the weights, which we have chosen to be $w_1 = w_2 = 1$. Equal weights are selected as shown in earlier work [41] where the morphing images generated with equal weights pose higher vulnerability to COTS FRS. Finally, $L_M^{'}$ is passed through the synthesis network (independently from StyleGAN [29] and StyleGAN2 [32] model) to generate the corresponding morphed image $I_M^{'}$ that has a resolution of 1024×1024 pixels. The generated morphed face image $I_M^{'}$ is then optimised using the proposed loss function to generate the high quality morphed face image. In the following section, we discuss the loss function to optimise the latent vector obtained using Equation 1.

2.1 Proposed Loss Function

The proposed loss function is based on both perceptual fidelity, quality and identity factors that can facilitate high-quality face morph generation. The common issue with the GAN-based morph generation is the presence of the ghost artefacts and the blurring issues. We employ the perceptual loss with multiple layers to eliminate such effects as given by Eqn 2.

$$\begin{aligned} Loss_{Perceptual} = & \frac{1}{2}\sum_i \frac{1}{N_i}||F_i(I_1) - F_i(I_M^{'})||_2^2 \\ & + \frac{1}{2}\sum_i \frac{1}{N_i}||F_i(I_2) - F_i(I_M^{'})||_2^2, \end{aligned} \tag{2}$$

where N_i denotes the number of features in layer i and F_i denotes features in layer i of the perceptual network (VGG-16 in our case). For the combination of perceptual layers, we choose $conv1_1$, $conv1_2$, $conv2_2$, $conv3_3$ inspired by [42]. Compared with the original combination of layers $conv1_2$, $conv2_2$, $conv3_3$, $conv4_3$ [43], our design measures low-level features instead of high-level features like style of an image and is closer to our goal of morphing faces with high quality.

The main goal of this paper is to generate the morphed face images that can significantly attack FRS. In order to achieve this, we have introduced the identity loss function based on the feedback from FRS. We employ Arcface [31] - a deep learning based FRS because of its robust and accurate

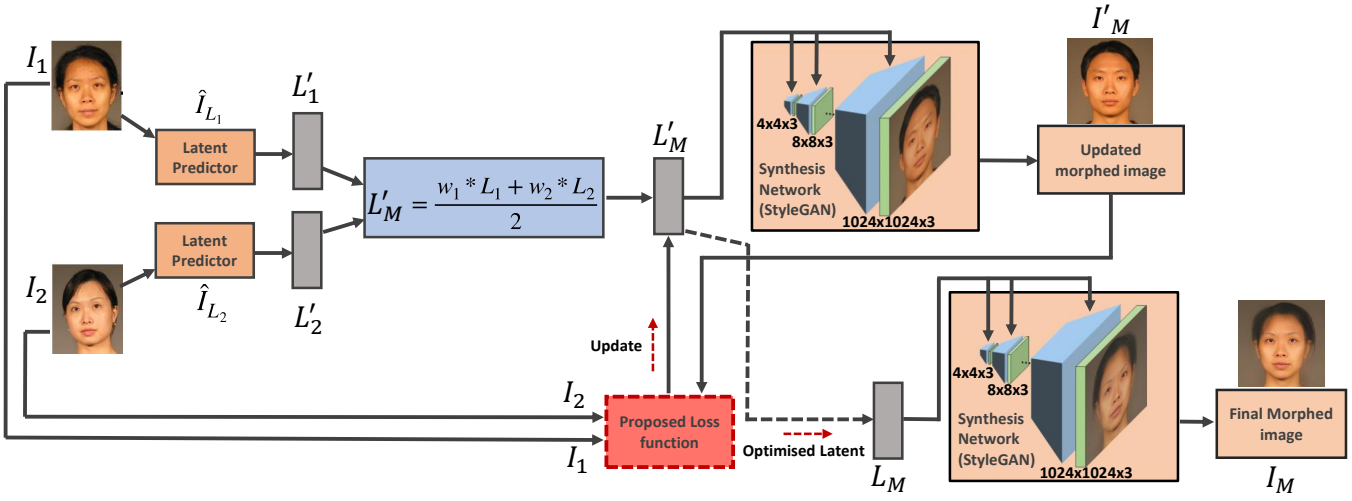


Fig. 3: Block diagram of the proposed MIPGAN for generating high quality morphed face images

performance to obtain the feedback on generated morphed face. Specifically, we employ a pre-trained embedding extractor with $ResNet50$ as the backbone to extract the unit embedding vectors and define the identity loss by their cosine distance to improve the morph generation process as given by Eqn 3.

$$Loss_{Identity} = \frac{(1 - \frac{\vec{v}_1 \cdot \vec{v}_M}{\|\vec{v}_1\|\|\vec{v}_M\|}) + (1 - \frac{\vec{v}_2 \cdot \vec{v}_M}{\|\vec{v}_2\|\|\vec{v}_M\|})}{2}, \quad (3)$$

where $\vec{v}_1$, $\vec{v}_2$, $\vec{v}_M$ respectively denotes the embedding vectors which are extracted from image $I_1, I_2, I_M^{'}$ respectively.

To further prove the loss function is differential for the morphed embedding vector $\vec{v}_M$, we define x_d, y_d, z_d to be the value of vector $\vec{v}_1$, $\vec{v}_2$, $\vec{v}_M$ in dimension d respectively and $d' \neq d$ to be other dimensions except d. The expanded identity loss function and its partial derivative are:

$$Loss_{Identity} = \frac{(1 - \frac{\sum_d x_d z_d}{\|\vec{v}_1\|\|\vec{v}_M\|}) + (1 - \frac{\sum_d y_d z_d}{\|\vec{v}_2\|\|\vec{v}_M\|})}{2}, \quad (4)$$

$$\begin{aligned}\frac{\partial Loss_{Identity}}{\partial z_d} &= 1 - \frac{x_d}{2\|\vec{v}_1\|}\frac{\partial}{\partial z_d}\left(\frac{z_d}{\sqrt{z_d^2 + \sum_{d' \neq d} z_{d'}^2}}\right) \\ &\quad - \frac{y_d}{2\|\vec{v}_2\|}\frac{\partial}{\partial z_d}\left(\frac{z_d}{\sqrt{z_d^2 + \sum_{d' \neq d} z_{d'}^2}}\right),\end{aligned} \quad (5)$$

$$\begin{aligned}\frac{\partial}{\partial z_d}\left(\frac{z_d}{\sqrt{z_d^2 + \sum_{d' \neq d} z_{d'}^2}}\right) &= \frac{1}{\sqrt{z_d^2 + \sum_{d' \neq d} z_{d'}^2}} \\ &\quad + \frac{2z_d^2}{-2(z_d^2 + \sum_{d' \neq d} z_{d'}^2)^{\frac{3}{2}}} \\ &= \frac{\sum_{d' \neq d} z_{d'}^2}{(z_d^2 + \sum_{d' \neq d} z_{d'}^2)^{\frac{3}{2}}},\end{aligned}$$

$$\frac{\partial Loss_{Identity}}{\partial z_d} = 1 - \frac{(\frac{x_d}{2\|\vec{v}_1\|} + \frac{y_d}{2\|\vec{v}_2\|})\sum_{d' \neq d} z_{d'}^2}{(z_d^2 + \sum_{d' \neq d} z_{d'}^2)^{\frac{3}{2}}}. \quad (6)$$

For any value $z_d = z'_d$, it is obvious that:

$$\begin{aligned}&\lim_{\Delta z_d \to 0} \frac{\partial Loss_{Identity}(z'_d + \Delta z_d)}{\partial z_d} \\ &= \lim_{\Delta z_d \to 0}(1 - \frac{(\frac{x_d}{2\|\vec{v}_1\|} + \frac{y_d}{2\|\vec{v}_2\|})\sum_{d' \neq d} z_{d'}^2}{((z'_d + \Delta z_d)^2 + \sum_{d' \neq d} z_{d'}^2)^{\frac{3}{2}}}) \\ &= 1 - \frac{(\frac{x_d}{2\|\vec{v}_1\|} + \frac{y_d}{2\|\vec{v}_2\|})\sum_{d' \neq d} z_{d'}^2}{(z_d'^2 + \sum_{d' \neq d} z_{d'}^2)^{\frac{3}{2}}} \\ &= \frac{\partial Loss_{Identity}(z'_d)}{\partial z_d}.\end{aligned}$$

Hence, for any dimension of d, the partial derivative of the identity loss function is continuous.

It is interesting to note that the identity loss based on the Arcface feature extractor model is trained to maximize the face class separability and thus is more sensitive to face attributes. Hence, only optimising the identity loss cannot achieve the same reconstruction performance as the perceptual loss but applying it on the face region can effectively control the generated attributes to be recognized as both subjects.

To solve the imbalance between different subjects, we introduce an identity difference loss as given by Eqn 7.

$$Loss_{ID-Diff} = |(1 - \frac{\vec{v}_1 \cdot \vec{v}_M}{\|\vec{v}_1\|\|\vec{v}_M\|}) - (1 - \frac{\vec{v}_2 \cdot \vec{v}_M}{\|\vec{v}_2\|\|\vec{v}_M\|})|. \quad (7)$$

With the idea of the Lagrange multiplier, it adds a constraint to the optimisation process to force the cosine distance between morph embedding and each of the two reference embeddings to be the same. Since $Loss_{ID-Diff}$ is usually small with a value less than 1, we apply $L1$ loss on the difference of two cosine distance terms to avoid the vanishing gradient problem.

Finally, in order to improve the structural visibility of the generated morphed face image, we also apply the Multi-Scale Structural Similarity (MS-SSIM) loss $L_{MS-SSIM}$ to measure the structure similarity [45]. Given two discrete non-negative signals (images in our case) x and y, lumi-

nance, contrast and structure comparison measures were given by l, c, s as computed using Eqn 8.

$$\begin{aligned} l(x,y) &= \frac{(2\mu_x 2\mu_y + (K_1L)^2)}{\mu_x^2 + \mu_y^2 + (K_1L)^2}, \\ c(x,y) &= \frac{(2\sigma_x 2\sigma_y + (K_2L)^2)}{\sigma_x^2 + \sigma_y^2 + (K_2L)^2}, \\ s(x,y) &= \frac{(\sigma_{xy} + \frac{(K_2L)^2}{2})}{\sigma_x\sigma_y + \frac{(K_2L)^2}{2}}, \end{aligned} \tag{8}$$

where μ_x, σ_x and σ_{xy} denotes the mean of x, the variance of x and the covariance of x and y respectively. L is the dynamic range of the signal and $K_1 \ll 1, K_2 \ll 1$ are two constant scalars. The MSSSIM loss $L_{MS-SSIM}$ is further defined by Eqn 9.

$$\begin{aligned} MSSSIM(x,y) =& [l_J(x,y)]^{\alpha_J} \cdot \prod_{j=1}^{J} [c_j(x,y)]^{\beta_j} [s_j(x,y)]^{\gamma_j}, \\ L_{MS-SSIM} =& \frac{1}{2}(1 - MSSSIM(I_1, I'_M)) \\ &+ \frac{1}{2}(1 - MSSSIM(I_2, I'_M)), \end{aligned} \tag{9}$$

where $j = 1, 2, \ldots, J$ represents the j^{th} scale and α_j, β_j and γ_j are the factors of relative importance. As suggested in [45], we also set $\alpha_j = \beta_j = \gamma_j$, $\sum_{j=1}^{J} \gamma_j = 1$ and use the resulting parameters $\beta_1 = \gamma_1 = 0.0448, \beta_2 = \gamma_2 = 0.2856, \beta_3 = \gamma_3 = 0.3001, \beta_4 = \gamma_4 = 0.2363, \alpha_5 = \beta_5 = \gamma_5 = 0.1333$.

Thus, the proposed loss function can be formulated as:

$$\begin{aligned} Loss =& \lambda_1 Loss_{Perceptual} + \lambda_2 Loss_{Identity} \\ &+ \lambda_3 Loss_{MS-SSIM} + \lambda_4 Loss_{ID-Diff}, \end{aligned} \tag{10}$$

where λ_1, λ_2, λ_3 and λ_4 are the hyper-parameters that are set to achieve both stable and generalised convergence. In this work, we empirically set $\lambda_1 = 0.0002$, $\lambda_2 = 10$, $\lambda_3 = 1$ and $\lambda_4 = 1$.

2.2 Training and Optimisation

The training and optimisation of the proposed method are carried out on Tensorflow version 1.13 and version 1.14 for StyleGAN and StyleGAN2, respectively. The optimisation is carried out using NVIDIA GTX 1070 8 GB GPU with CUDA version 10.0 and CUDNN version 7.5 and NVIDIA Tesla P100 PCIE 16 GB GPU. The Adam optimiser with hyper-parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 1 \times 10^{-8}$ as recommended in the original paper [46] is employed on this work. The list of morphing pairs is generated in advance with careful considerations to gender. During each optimisation process of 150 iterations, the learning rate is initially set to $\eta = 0.03$ with an exponential decay per 6 iterations of $\eta_{new} = \eta * 0.95$.

Figure 4 illustrates the qualitative results of the proposed MIPGAN framework based on StyleGAN and StyleGAN2. Further, the qualitative results of the existing methods based on StyleGAN [1] and MorGAN [28] is provided alongside for the convenience of the reader in the same figure. It is interesting to note that the proposed MIPGAN generated face morph images indicate both perceptual and geometric features correspondence to both contributing subjects (for instance, malicious actor and accomplice).

3 Experiments and Results

This section presents and discusses the experimental protocols, datasets, and quantitative results of the proposed face morphing technique. The images generated from proposed MIPGAN-I and MIPGAN-II architectures are compared with the state-of-the-art techniques based on both facial landmarks [7] and StyleGAN based morph generation [1]. The effectiveness of the face morphing generation is quantitatively evaluated by benchmarking the vulnerability of the COTS FRS for generated morphed face images. Further, we also evaluate the morph attack detection potential by evaluating the generated morphed face images using the most recent and robust MAD techniques.

3.1 MIPGAN Face Morph Dataset

We employ the face images from FRGC-V2 face database [33] to generate the *MIPGAN Face Morph Dataset* consisting of morphed face images using both state-of-the-art and the proposed MIPGAN technique. We have selected 140 unique data subjects from the FRGC dataset by considering the high-quality face images captured in constrained conditions that resemble the passport image quality. Among 140 data subjects, 47 data subjects are female and 93 data subjects are male. Each data subject has a variable size of 7-21 additional

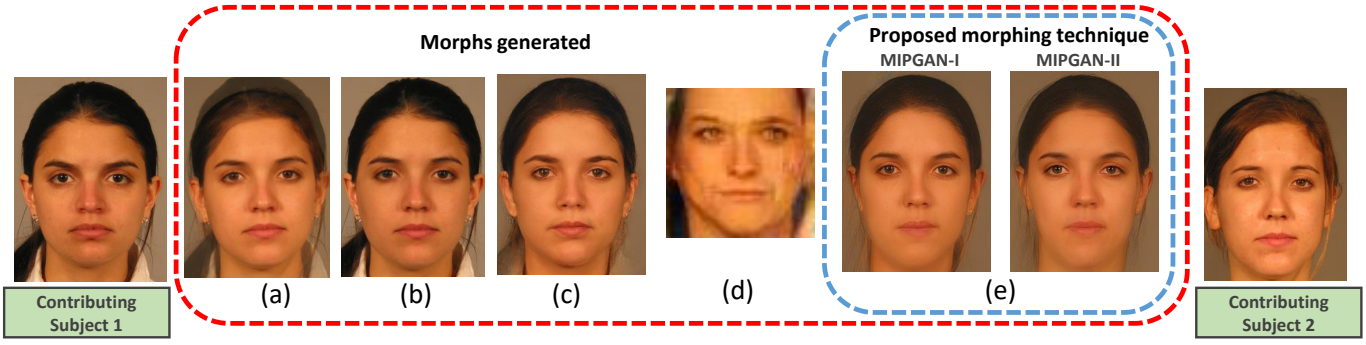


Fig. 4: Qualitative results of proposed MIPGAN together with existing GAN based face morph generation methods (a) Landmark-I [7] (b) Landmark-II [44] (c) StyleGAN [1] (d) MorGAN [28] (e) Proposed method

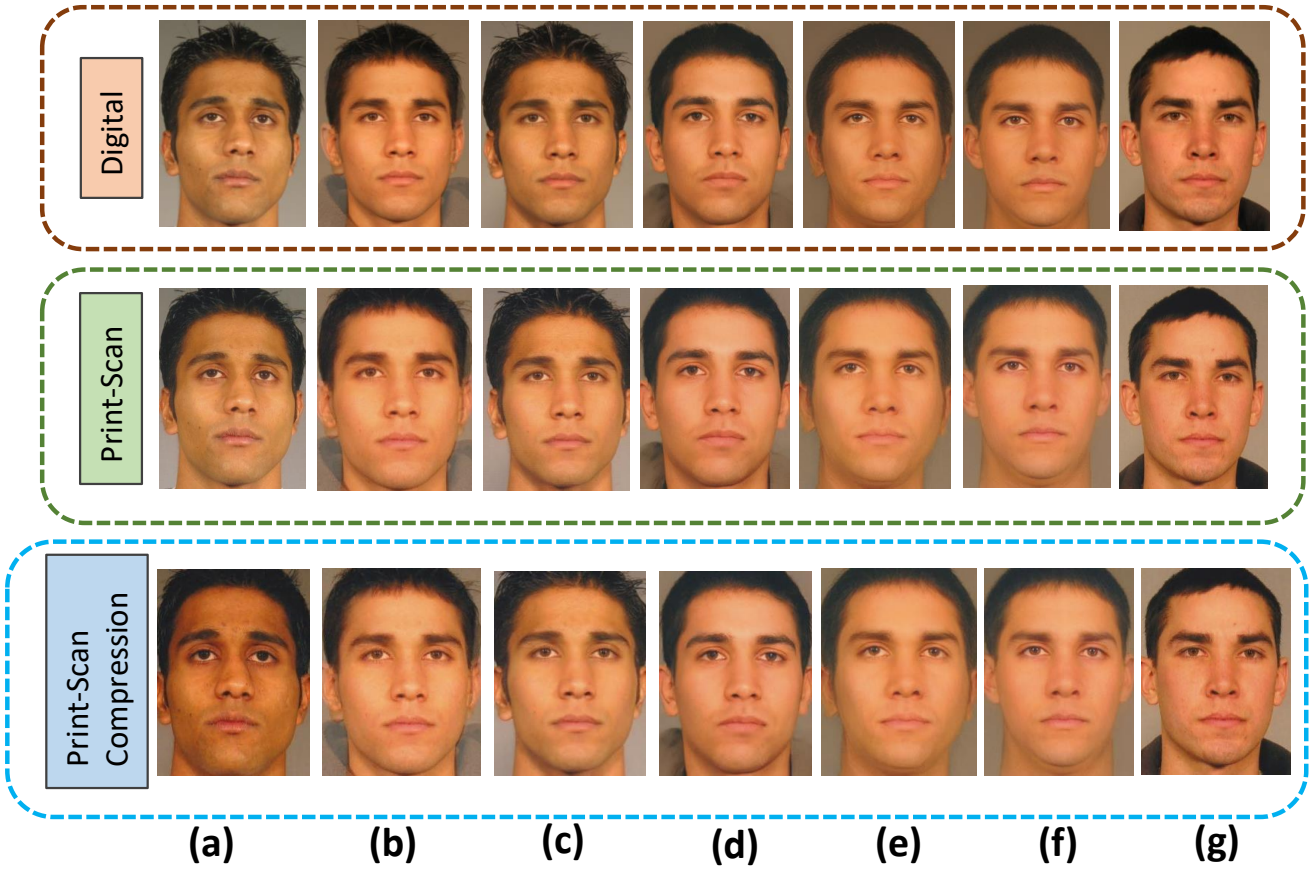


Fig. 5: Illustration of morphing in digital, print-scan and print-scan compression data (a) Contributing subject 1 (b) Landmark-I [7] (c) Landmark-II [44] (d) StyleGAN [1] (e) MIPGAN-I (f) MIPGAN-II (g) Contributing subject 2

captures making the whole dataset have 1270 samples corresponding to 140 data subjects. We employ three different face morph generation techniques based on facial landmarks constrained by Delaunay triangles with blending [7] we term this as Landmarks-I, landmarks-based techniques with automatic post processing and color equalisation [44], we term this as Landmarks-II and StyleGAN [1]. We do not consider MorGAN [28] [47] based face morph generation as it was earlier demonstrated that MorGAN does not generate ICAO compliant images and thus makes COTS FRS not vulnerable [1]. All the samples are pre-processed to meet the ICAO standards [27] and morphing is carried out by following the guidelines outlined earlier [7] [8], i.e, careful selection of subjects based on gender and comparison score using FRS to make attacks realistic.

To effectively evaluate the proposed method's quantitative performance and the existing techniques, we create three different types of attacks from morphed images, such as **Digital morphed images**: Morphed face images that are obtained from the morph generation process in the digital domain. **Print-scanned morphed images**: The digital morphed and bona fide images are printed and then scanned (or re-digitized) to simulate the passport application process. We have employed DNP-DS820 [48] dye-sublimation photo printer to generate the print of the digital morphed and bona fide face images in this work. The use of a dye-sublimation photo printer guarantees high-quality photo printing (generally used for a passport application) and makes sure that printed photos are free from dotted patterns (or individual droplets of ink) that are resulting from the printing process of conventional printers. Each of these printed photos is then scanned (or re-digitized) using the Canon office scanner to have 300 dpi as suggested in ICAO standards [27]. **Print-scanned compressed morphed images**: The printed and scanned images (both morphed and bona fide) are compressed to have a size of 15kb that makes it suitable to store in the e-passport. This process reflects the real-life scenario of face image storage in passport systems. Thus, the overall dataset has 2500×3 (types of morph data) $\times$ 4 types of morph generation technique = $30,000$ morph samples and 1270×3 (types of morph data) $\times$ 4 types of morph generation technique = $15,240$ bona fide samples. Figure 5 illustrates the three data types of attacks that are used to evaluate the effectiveness of the proposed method and the existing methods of face morph generation. It is evident that the visual quality of the images vary largely for different attack types (for instance, the digital data attack indicates the best quality and print-scan with compression indicates the lowest quality).

3.2 Vulnerability Analysis

This section presents the vulnerability analysis of the proposed morphed face generation techniques to quantify the

Morph generation type	MMPMR/RMMR(%)	FMMPMR/RMMR(%)	MMPMR/RMMR(%)	FMMPMR/RMMR(%)	MMPMR/RMMR(%)	FMMPMR/RMMR(%)
	Digital		**Print-Scan**		**Print-Scan with compression**	
	Male		Male		Male	
Landmark-I [7]	100	98.77	97.23	97.34	97.38	96.95
Landmark-II [44]	87.29	76.86	90.32	78.23	88.78	77.14
StyleGAN [1]	63.51	41.27	60.59	39.51	57.12	35.05
MIPGAN-I	93.35	83.08	91.72	80.55	91.07	77.89
MIPGAN-II	92.22	80.45	90.74	77.67	89.16	73.47
	Female		Female		Female	
Landmark-I [7]	100	99.26	99.37	99.02	99.78	99.24
Landmark-II [44]	94.28	88.67	98.22	91.48	98.16	90.97
StyleGAN [1]	68.75	42.62	66.45	42.01	66.45	40.49
MIPGAN-I	98.57	93.11	98.16	91.22	96.12	90.52
MIPGAN-II	95.91	87.66	95.30	86.26	94.69	84.47
	Combined		Combined		Combined	
Landmark-I [7]	**100**	**98.84**	**97.64**	**97.60**	**97.84**	**97.30**
Landmark-II [44]	88.65	78.72	91.85	81.56	90.61	79.33
StyleGAN [1]	64.68	41.49	61.72	39.90	58.92	35.89
MIPGAN-I	94.36	84.65	92.97	82.23	92.29	79.88
MIPGAN-II	92.93	81.59	80.56	79.02	90.24	75.20

TABLE 1: Quantitative evaluation of vulnerability of COTS Cognitec-FRS [30] from various morph generation approaches. Note that, since FNMR = 0 @ FMR = 0.1% for Cognitec-FRS [30] following Eq. 12 and 13, the value of RMMR is equal to MMPMR/FMMPMR. Therefore, we have not entered RMMR separately in the Table above.

Morph generation type	MMPMR/RMMR(%)	FMMPMR/RMMR(%)	MMPMR/RMMR(%)	FMMPMR/RMMR(%)	MMPMR/RMMR(%)	FMMPMR/RMMR(%)
	Digital		**Print-Scan**		**Print-Scan with compression**	
	Male		Male		Male	
Landmark-I [7]	85.59	70.80	83.91	68.20	83.86	67.73
Landmark-II [44]	63.27	46.55	63.12	46.37	63.72	46.80
StyleGAN [1]	61.19	41.01	61.68	41.43	61.68	41.04
MIPGAN-I	76.96	59.24	76.96	57.16	76.07	57.31
MIPGAN-II	75.73	56.97	72.87	54.57	72.87	54.43
	Female		Female		Female	
Landmark-I [7]	96.03	83.55	93.95	82.02	93.32	81.39
Landmark-II [44]	87.76	71.85	89.39	73.82	89.80	74.27
StyleGAN [1]	80.42	59.19	79.79	59.10	78.54	58.83
MIPGAN-I	90.41	76.68	89.39	75.95	89.18	75.85
MIPGAN-II	88.98	75.42	87.96	74.54	88.37	74.90
	Combined		Combined		Combined	
Landmark-I [7]	**87.64**	**72.82**	**85.87**	**70.39**	**85.71**	**69.90**
Landmark-II [44]	68.07	50.64	68.27	50.80	68.86	51.28
StyleGAN [1]	64.92	43.91	65.20	44.25	64.96	43.88
MIPGAN-I	79.61	62.06	79.41	60.19	78.66	60.30
MIPGAN-II	78.34	59.95	75.84	57.80	75.92	57.73

TABLE 2: Quantitative evaluation of vulnerability of VGGFace2 [49] FRS from various morph generation approaches. Note that, since FNMR = 0 @ FMR = 0.1% for VGGFace2 [49] following Eq. 12 and 13, the value of RMMR is equal to MMPMR/FMMPMR. Therefore, we have not entered RMMR separately in the Table above.

impact of our efficient attacks on FRS. We quantify the attack success for five different FRS including two Commercial-off-the-Shelf (COTS) FRS and three deep-learning-based open-source FRS. The COTS FRS include the Cognitec FRS (Version 9.4.2) [30] and Neurotechnology (Version 10) [50] and the set of open-source FRS includes Arcface [31], VGGFace [49] and LCNN-29 [51]. The operational threshold for all 5 FRS is set at False Match Rate (FMR) of 0.1% following the guidelines of Frontex [52].

The vulnerability is assessed using two metrics Mated Morphed Presentation Match Rate (MMPMR) [8] and Fully Mated Morphed Presentation Match Rate (FMMPMR) [1] based on the threshold provided by Cognitec FRS. For a given morph image $M_{I_{1,2}}$ obtained using two subjects, we compute the vulnerability by enrolling $M_{I_{1,2}}$ and verifying it against probe images from the corresponding contributing subjects I_1 and I_2. The obtained comparison scores S_1 and S_2 for both probe images I_1 and I_2 against the morphed image $M_{I_{1,2}}$ indicates the threat to FRS, if and only if both S_1 and S_2 cross the actual verification threshold at FMR = 0.1%. The corresponding metric *FMMPMR* [1] [41] is therefore

Morph generation type	MMPMR/RMMR(%)	FMMPMR/RMMR(%)	MMPMR/RMMR(%)	FMMPMR/RMMR(%)	MMPMR/RMMR(%)	FMMPMR/RMMR(%)
	Digital		**Print-Scan**		**Print-Scan with compression**	
	Male		Male		Male	
Landmark-I [7]	99.60	98.19	97.38	96.88	97.33	96.70
Landmark-II [44]	91.09	84.62	93.45	86.42	93.60	86.02
StyleGAN [1]	70.99	55.76	73.86	58.67	73.32	58.26
MIPGAN-I	93.70	85.17	92.76	84.39	93.01	84.41
MIPGAN-II	93.65	86.45	93.55	85.30	93.25	85.06
	Female		Female		Female	
Landmark-I [7]	99.79	97.01	99.79	96.91	99.79	97.01
Landmark-II [44]	94.49	86.71	97.76	89.76	98.16	89.17
StyleGAN [1]	80.21	63.22	82.71	65.70	82.71	66.05
MIPGAN-I	97.35	89.53	97.96	91.02	97.76	91.02
MIPGAN-II	96.33	89.47	95.92	89.33	96.12	89.42
	Combined		Combined		Combined	
Landmark-I [7]	**99.68**	**98.00**	**97.88**	**96.89**	**97.84**	**96.75**
Landmark-II [44]	91.79	84.96	94.33	86.96	94.53	86.54
StyleGAN [1]	72.80	56.95	75.60	59.79	75.16	59.51
MIPGAN-I	94.45	85.94	93.81	85.46	93.97	85.48
MIPGAN-II	94.21	86.94	94.05	85.95	93.85	85.77

TABLE 3: Quantitative evaluation of vulnerability of Arcface [31] FRS from various morph generation approaches. Note that, since FNMR = 0 @ FMR = 0.1% for Arcface [31] following Eq. 12 and 13, the value of RMMR is equal to MMPMR/FMMPMR. Therefore, we have not entered RMMR separately in the Table above.

Morph generation type	MMPMR/RMMR(%)	FMMPMR/RMMR(%)	MMPMR/RMMR(%)	FMMPMR/RMMR(%)	MMPMR/RMMR(%)	FMMPMR/RMMR(%)
	Digital		**Print-Scan**		**Print-Scan with compression**	
	Male		Male		Male	
Landmark-I [7]	99.40	94.70	95.45	83.71	93.23	77.16
Landmark-II [44]	88.99	68.51	88.92	63.31	80.62	53.63
StyleGAN [1]	52.26	26.47	31.88	12.98	31.60	12.15
MIPGAN-I	58.18	32.56	32.59	25.33	57.6	53.52
MIPGAN-II	53.16	29.65	47.41	20.71	50.73	23.72
	Female		Female		Female	
Landmark-I [7]	100	99.25	100	98.11	98.74	91.18
Landmark-II [44]	94.69	85.96	97.49	84.92	95.40	78.89
StyleGAN [1]	70.60	50.13	55.20	25.72	52.39	26.19
MIPGAN-I	80.98	56.29	73.06	46.87	77.89	30.50
MIPGAN-II	74.79	49.45	69.59	42.17	70.73	46.18
	Combined		Combined		Combined	
Landmark-I [7]	**99.51**	**95.37**	**96.32**	**85.43**	**94.30**	**79.25**
Landmark-II [44]	90.16	71.17	90.59	66.67	83.50	57.38
StyleGAN [1]	55.06	29.39	36.36	14.83	35.62	14.28
MIPGAN-I	63.22	35.73	40.46	28.71	61.66	34.14
MIPGAN-II	57.47	31.45	51.72	23.54	54.94	27.46

TABLE 4: Quantitative evaluation of vulnerability of COTS Neurotec [50] FRS from various morph generation approaches. Note that, since FNMR = 0 @ FMR = 0.1% for COTS Neurotec [50] following Eq. 12 and 13, the value of RMMR is equal to MMPMR/FMMPMR. Therefore, we have not entered RMMR separately in the Table above.

computed as:

$$FMMPMR = \frac{1}{P}\sum_{M,P}(S1_M^P > \tau)\&\&(S2_M^P > \tau)\ldots\&\&(Sk_M^P > \tau), \tag{11}$$

where $P = 1, 2, \ldots, p$ represent the number of attempts made by presenting all probe images of the contributing subjects against the M^{th} morphed image, $K = 1, 2, \ldots, k$ represents the number of composite image constitute to generate the morphing image (in our case $K = 2$), Sk_M^P represents the comparison score of the K^{th} contributing subject obtained with P^{th} attempt corresponding to M^{th} morphing image and τ represents the threshold value corresponding to FMR = 0.1%. When compared to MMPMR, the FMMPMR will consider both pair-wise comparison of contributory subjects and the number of attempts. In order to also establish the relationship with respect to earlier metrics, we also report the vulnerability using MMPMR [8].

Further, to effectively analyse the vulnerability, we also present the results using Relative Morph Match Rate (RMMR) defined as follows [53]:

$$\begin{aligned} RMMR(\tau)_{MMPMR} =& 1 + (MMPMR(\tau)) \\ & - [1 - FNMR(\tau)] \end{aligned} \tag{12}$$

Morph generation type	MMPMR/RMMR(%)	FMMPMR/RMMR(%)	MMPMR/RMMR(%)	FMMPMR/RMMR(%)	MMPMR/RMMR(%)	FMMPMR/RMMR(%)
	Digital		**Print-Scan**		**Print-Scan with compression**	
	Male		Male		Male	
Landmark-I [7]	96.63	89.28	95.25	89.36	94.80	88.62
Landmark-II [44]	75.09	60.72	74.64	57.81	82.43	68.32
StyleGAN [1]	83.12	66.44	85.20	69.54	84.85	68.88
MIPGAN-I	95.13	86.35	94.04	84.39	94.09	84.30
MIPGAN-II	94.93	85.14	93.94	83.14	93.75	82.63
	Female		Female		Female	
Landmark-I [7]	99.16	95.00	98.75	94.26	98.96	94.49
Landmark-II [44]	92.04	82.28	94.69	82.85	95.92	86.98
StyleGAN [1]	93.33	80.08	92.92	83.06	92.92	82.76
MIPGAN-I	97.76	92.27	96.94	91.59	96.94	91.44
MIPGAN-II	95.71	90.72	95.31	89.85	95.71	89.58
	Combined		Combined		Combined	
Landmark-I [7]	**97.16**	**90.19**	**95.96**	**90.14**	**95.64**	**89.55**
Landmark-II [44]	78.42	64.20	78.58	61.85	85.11	71.36
StyleGAN [1]	85.12	68.61	86.72	71.69	86.44	71.09
MIPGAN-I	95.68	87.30	94.64	85.55	94.68	85.45
MIPGAN-II	95.12	86.05	94.25	84.23	94.17	83.75

TABLE 5: Quantitative evaluation of vulnerability of LCNN-29 [51] FRS from various morph generation approaches. Note that, since FNMR = 0 @ FMR = 0.1% for LCNN-29 [51] following Eq. 12 and 13, the value of RMMR is equal to MMPMR/FMMPMR. Therefore, we have not entered RMMR separately in the Table above.

$$RMMR(\tau)_{FMMPMR} = 1 + (FMMPMR(\tau)) - [1 - FNMR(\tau)] \quad (13)$$

Where, FNMR indicates the False Reject Rate (FNMR) of the FRS under consideration obtained at the threshold τ. In this work, τ represents the value corresponding to FMR = 0.1%. Since we have evaluated 5 different FRS systems, we have computed FNMR corresponding to these FRS to calculate the RMMR. Note that, in Equation 12 and 13 if FNMR = 0 then RMMR corresponds to MMPMR/FMMPMR.

The obtained success rate, or alternatively the vulnerability of FRS is provided in Table 1, 2, 3, 4 and 5 corresponding to to Cognitec [30], VGGFace [49], Arcface [31], Neurotechnology (Version 10) [50] and LCNN-29 [51] respectively. The vulnerability analysis is carried out on 5 different morph generation methods that include facial landmarks (Landmarks-I) with image smoothing as the post-processing operation [7], Facial landmarks (Landmarks-II) with automatic image retouching and colour equalisation [44], existing GAN based face morphing method based on StyleGAN [1] and proposed MIPGAN variants (MIPGAN-I and MIPGAN-II). Based on the obtained results, the following are the concrete observations:

- The FNMR corresponding to five different FRS is equal to 0. Therefore, the value of the RMMR is equal to MMPMR or FMMPMR. This indicates that the FRS systems are accurate on our face datasets employed in this work.
- Among the five FRS, the highest vulnerability is noted for Arcface [31], which is vulnerable to all five kinds of face morphing attack methods.
- Among COTS FRS, the Cognitec FRS indicates a higher vulnerability on all five types of face morphing attack methods compared to Neurotechnology FRS.
- Among five different morph generation methods, Landmark-I indicates the highest vulnerability on all five other FRS.
- The proposed face morphing methods MIPGAN-I and MIPGAN-II consistently indicate the highest vulnerability, when compared to the existing method based on StyleGAN [1]. This indicates the high quality of morphs generated using the proposed MIPGAN-I and MIPGAN-II methods.
- The proposed MIPGAN-I and MIPGAN-II methods also indicate a higher vulnerability than the Landmark-II technique for morph generation with four different FRS.
- Among the two different metrics (MMPMR and FMMPMR), the proposed FMMPMR indicates a lower vulnerability than MMPMR consistently as FMMPMR imposes a strict selection of attack images, unlike MMPMR.
- MIPGAN-I based morphed images show a marginally better performance in attacking FRS than images generated by MIPGAN-II.

3.3 Perceptual Image Quality Analysis

This section presents quantitative results of the proposed morphed image generation techniques using the perceptual image quality metrics PSNR and SSIM. Both of these metrics are computed based on the reference image. Morphed face images are generated based on parent face images from two contributory data subjects. Therefore, we used the parent face images from both contributory data subjects as the reference image against which the given morphed image is assessed and we average the obtained image quality scores for both parent images. Table 6 indicates the quantitative results of both PSNR and SSIM on four different types of face morph generation mechanism in the digital format. Based on the obtained results, it can be observed that:

- There is little deviation in the perceptual image quality metrics computed on all four different types of face morph generation mechanisms.

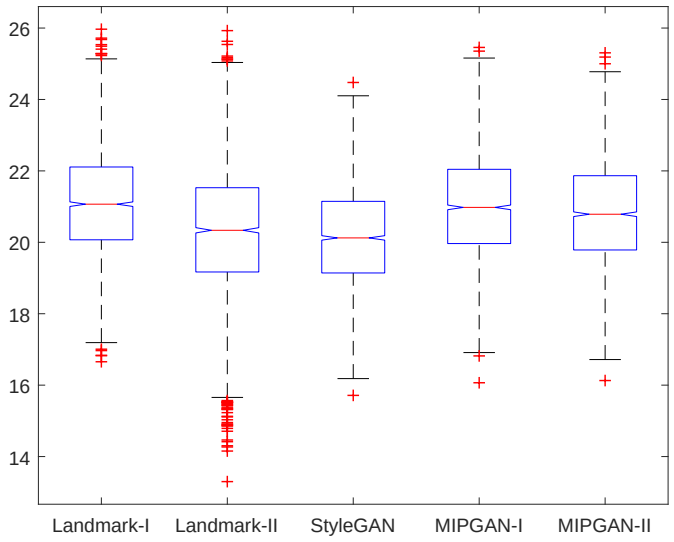


Fig. 6: Box plots of PSNR values computed from different face morph generation methods (digital version)

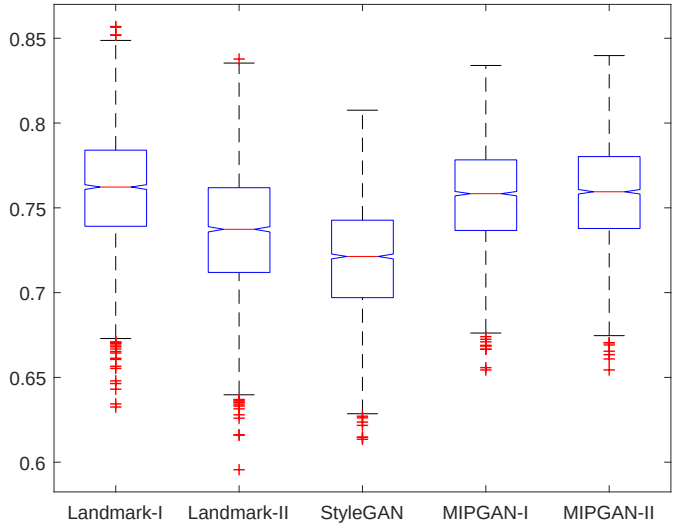


Fig. 7: Box plots of SSIM values computed from different face morph generation methods (digital version)

- The proposed MIPGAN-I and MIPGAN-II methods indicate a slightly better image quality when compared to the StyleGAN [1] based face morphing method.
- The proposed MIPGAN-I and facial landmarks-based methods [44] indicate a similar image quality.
- Figure 6 and 7 indicate the box plots of the PSNR and SSIM quality scores. These results further indicate that the perceptual quality of the proposed MIPGAN-I and MIPGAN-II is superior to the existing state-of-the-art method based on StyleGAN [1].

3.4 Human Observer Analysis

In this section, we discuss the quantitative detection performance of human observations regarding morphed face images, which are generated using MIPGAN-I and MIPGAN-II. To this extent, we have designed and developed a web-portal to evaluate the human morph detection performance

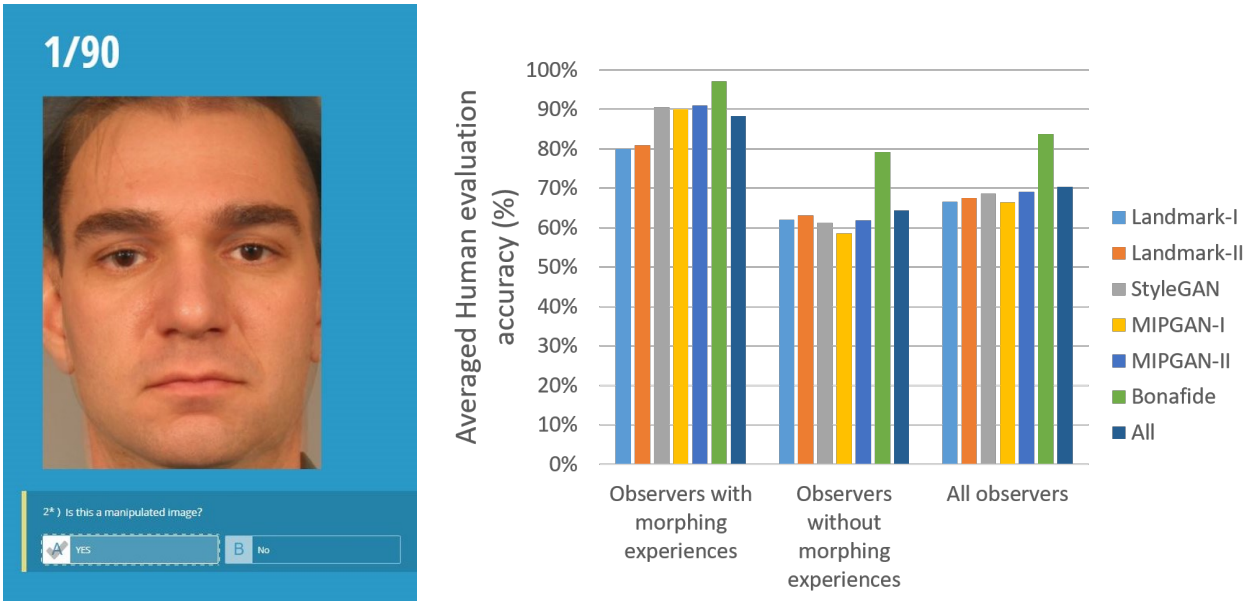
Morph generation Methods	PSNR	SSIM
Landmark-I [7]	21.1111± 0.0415	0.7609±0.0009
Landmark-II [44]	20.2737±0.0523	0.7363±0.0010
StyleGAN [1]	20.1347±0.0383	0.7199±0.0008
MIPGAN-I	21.0133±0.0409	0.7573±0.0008
MIPGAN-II	20.8306±0.0409	0.7586±0.0008

TABLE 6: Morph image quality analysis using PSNR and SSIM with 95% confidence interval

reflecting both single image-based morphing attack detection scenario (S-MAD) and differential morphing attack detection scenario (D-MAD). We have used only digital samples of both bona fide and morphed face images as the proposed MIPGAN is used to generate the images in the digital domain. Figure 8 (a) shows the screenshot of the web-portal for S-MAD in which the human observer needs to decide whether the displayed image is a morphed face image or a bona fide image by looking at one single image at a time. Correspondingly, Figure 9 (a) presents the screenshot for D-MAD experiment where the observer needs to detect whether the unknown image is morphed given a trusted bona fide image as a reference. We have selected a total of 90 images where 15 images are from each group corresponding to bona fide, two different types of facial landmarks based morphing such as Landmarks-I [7] and Landmarks-II [44], StyleGAN [1] based face morphing, MIPGAN-I and MIPGAN-II based face morphing. To make the testing robust, all 90 chosen images correspond to unique data subjects and there is no repetition of data subjects. To avoid gender bias by participants, we have selected a near equal distribution of male and female data subjects in each group. We have chosen 90 images considering the time constraints required to assess these images for human observers. It was important that observers do not loose focus while conducting the detection experiments.

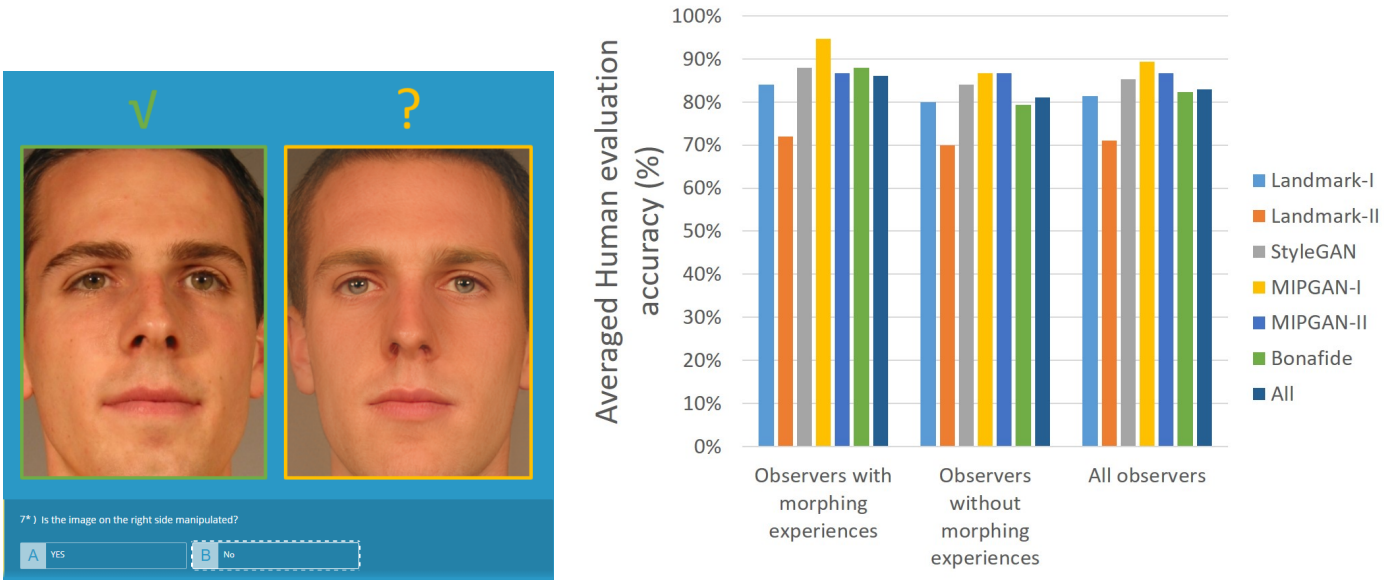
Figure 8 (b) shows the quantitative results of S-MAD obtained from 56 human observers, including 14 experienced and 42 inexperienced observers. The experienced observers' group consists of researchers working in face morphing attack detection and as ID expert's in border control, while the non-experienced group consists of students and other computer science professionals. As noticed from the Figure 8 (b) following are the main observations:

- Detection performance of the bona fide images indicates better detection performance by both experienced and non-experienced group when compared to the morphed face image. The experienced group indicates the detection performance with an accuracy of 97.14%, while the non-experienced group indicates the detection performance with an accuracy of 79.21%.
- Human observers with experience in face morphing demonstrate higher detection accuracy on four different face morph generation mechanisms than the inexperienced group.
- Among the four different morphing types, the experienced group indicates that the detection of the landmarks-based morphing is challenging compared to other morphing mechanisms (deep learning-based).
- Human observers with no experience in face morphing



(a) (b)

Fig. 8: (a)Example of screen shot used for human observer study (b) Quantitative results



(a) (b)

Fig. 9: (a)Example of screen shot used for differential human observer study (b) Quantitative results

$Loss_{ID-Diff}$	$Loss_{Identity}$	$Loss_{MS-SSIM}$	$Loss_{Perceptual}$	MIPGAN-I				MIPGAN-II			
				FMMPMR		MMPMR		FMMPMR		MMPMR	
				Cognitec	ArcFace	Cognitec	ArcFace	Cognitec	ArcFace	Cognitec	ArcFace
×	✓	✓	✓	81.82	75.87	90.69	93.47	77.83	71.98	90.1	91.18
✓	×	✓	✓	78.07	62.15	89.17	83.77	78.39	64.51	90.04	82.54
✓	✓	×	✓	80.82	73.33	91.81	92.66	78.73	71.79	89.58	90.55
✓	✓	✓	×	21.37	47.85	44.18	71.95	11.92	33.12	29.47	59.56
✓	✓	✓	✓	**84.65**	**85.94**	**94.36**	**94.45**	**81.59**	**86.24**	**92.93**	**94.21**

TABLE 7: Vulnerability - Ablation study on the proposed loss function. Here, ✓ indicates the selected and × indicates the not selected loss function in the ablation study

are marginally good in detecting the landmarks-based face morph images compared to other types of face morphing techniques. MIPGAN-I exhibits more challenging morph images to detect as compared to other morph generation methods.
- Based on the obtained results, it can be noted that the human observers with good experience in face morphing can detect morphed images with an accuracy of 88.25% while the human observer with no knowledge of face morphing shows the challenge to detect the morphed face images with a detection accuracy of 64.31%.
- The overall results from 56 human observers indicate that detecting morphed face images is challenging. Further, it is also interesting to note that detecting different face morphing types is also challenging.

For the quantitative results of D-MAD, 5 experienced observers and 10 inexperienced observers have participated. As shown in Figure 9 (b), following observations are illustrated:

- In the scenario of D-MAD, the group with relevant experiences achieved an overall 86% accuracy, which is better than 81% for the inexperienced group. However, this difference is much less than the difference in S-MAD, which means that the reference image can help inexperienced observers to identify the morphs.
- Morphs generated by Landmark-II present significant challenge than other morph generation mechanisms in D-MAD. This may be attributed to a more natural skin texture appearance (comparing with GAN-based mechanisms) and fewer artefacts (comparing with Landmark-I) and observers focusing less on its minor artefacts in the pairwise comparison.
- It is also interesting to see that the performances of experienced observers on detecting Landmark-II (80.95% and 72.00%), StyleGAN (90.48% and 88.00%), MIPGAN-II (90.95% and 86.67%), and bona fide images (90% and 88.00%) are lower than their performance in S-MAD. We believe this is because the experienced observers do not pay critical attention to tolerable difference between the trusted reference image and the unknown comparison image.

3.5 Ablation Study

In order to measure the impact of the loss functions in the proposed approach, we conduct an extensive ablation study. The proposed loss function combines four different entities such as: perceptual loss ($Loss_{Perceptual}$), identity loss ($Loss_{Identity}$), identity difference ($Loss_{ID-Diff}$) and Multi-Scale Structural Similarity (MS-SSIM) loss ($Loss_{MS-SSIM}$). The main contribution of our work is to use identity information, which can be considered as a specific high-level feature, to measure the loss. However, high-level features also mean that it is hard for the gradient descent algorithm to ensure a good convergence during the optimisation process. Therefore, we have introduced the perceptual loss that can measure relatively low-level features in addition to MS-SSIM and identity difference loss to effectively control the optimisation process to generate a high-quality morphed image. We perform the ablation study by discarding each term in the loss function iteratively. We benchmark the vulnerability using COTS FRS (Cognitec FRS (Version 9.4.2)) and the open-source ArcFace FRS, as the proposed approach is dedicated to generating high-quality morphed images.

Table 7 indicates the quantitative performance of the ablation study using a vulnerability analysis for both the COTS-FRS from Cognitec and for the open-source Arcface FRS with the proposed MIPGAN-I and MIPGAN-II methods. The ablation study is carried out on the digital morphed images generated using both MIPGAN-I and MIPGAN-II Methods. Figure 10 and 11 shows the qualitative performance of the ablation study on both MIPGAN-I and MIPGAN-II, respectively. Based on the obtained results, the following is the main observations:

- Each term in our proposed loss function (see Eq. 10) contributes to posing a greater challenge to FRS from the both proposed MIPGAN-I and MIPGAN-II morph generation framework.
- Among the four other losses that we have used, the $Loss_{Perceptual}$ is critical in improving the proposed method's performance. Discarding the perceptual loss has resulted in a degrading performance in both qualitative (see Figure 10 (d) and 11 (d)) and quantitative results.
- The use of identity loss ($Loss_{Identity}$) also indicates the importance of improving the quantitative performance of the proposed method.
- The $Loss_{MS-SSIM}$ also contributes to both qualitative and quantitative improvements of the morphs generated by the proposed method.

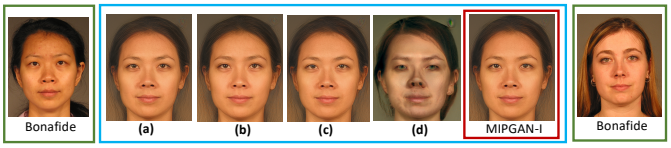


Fig. 10: Qualitative results of ablation study using proposed MIPGAN-I (a)$Loss_{ID-Diff}$ (b) $Loss_{Identity}$ (c) $Loss_{MS-SSIM}$ (d) $Loss_{Perceptual}$

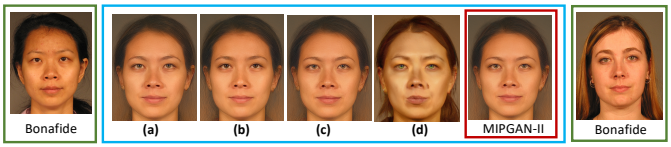


Fig. 11: Qualitative results of ablation study using proposed MIPGAN-II (a)$Loss_{ID-Diff}$ (b) $Loss_{Identity}$ (c) $Loss_{MS-SSIM}$ (d) $Loss_{Perceptual}$

3.6 Hyper-parameters Study

This section presents both qualitative and quantitative results on the selection of hyper-parameters (λ_1, λ_2, λ_3, and λ_4) in the proposed loss function employed in both MIPGAN-I and MIPGAN-II. Based on the ablation study reported in Section 3.5, we have noticed that the perceptual loss is the vital component of our loss function (see Eq. 10) and the other three terms can be used as constraints during

the optimisation. Therefore, the first step is to study the generated morphed face images' attack strength by increasing and decreasing the value of λ_1. Among the remaining three terms, we have also noticed from the ablation study that the identity loss ($Loss_{Identity}$) is contributing more towards generating a high-quality morph compared to the other two-loss functions ($loss_{MS-SSIM}$, $Loss_{ID-Diff}$). We analyze the importance of identity loss ($Loss_{Identity}$) with respect to the other two loss functions ($Loss_{MS-SSIM}$, $Loss_{ID-Diff}$) by increasing the value of λ_3 and/or λ_3 and decreasing the value of λ_2. Further, we have also noticed from the ablation study that the loss functions $loss_{MS-SSIM}$ and $Loss_{ID-Diff}$ are less important and numerically very small. Therefore, we did not conduct studies on decreasing the values of λ_3 and λ_4. Altogether, we have tested four different cases of changing the hyper-parameter values to generate the morphed face images. These generated morphed face images are benchmarked against the proposed hyper-parameter values through the vulnerability analysis using both COTS FRS (Cognitec FRS (Version 9.4.2)) and open-source ArcFace FRS.

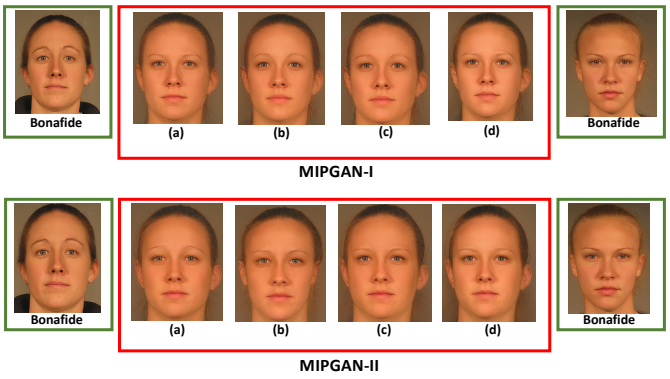


Fig. 12: Qualitative results of Hyper-parameters study on both MIPGAN-I and MIPGAN-II (a)λ_1 (b) λ_2 (c) λ_3 (d) λ_4

Table 8 shows the qualitative performance and Figure 12 shows the qualitative performance of the hyper-parameter study. Based on the obtained results, it can be noted that the increase in the value of λ_1 and λ_3 shows comparable results with the proposed weighting schemes. However, based on our empirical study on hyper-parameters, we noted that: if we set λ_1 and λ_2 with equal weights, then, during the optimisation, the generated morph image will soon become roughly similar to both contributing subjects. This will quickly reduce identity loss ($Loss_{Identity}$) to a minimal value and loose its importance in the optimisation. Hence, we set a larger factor to the identity loss compared with other loss terms measuring high-level features to ensure our most important constraint term is still effective in the later stage of optimisation. Further, both λ_3 and λ_4 can make the optimisation goal more comprehensive but setting a large factor will obstruct the convergence. Especially setting high values to λ_4 will end up with an image not similar to both subjects. Therefore, the selection of the proposed hyper-parameters confirms the generation of a high-quality morphed image but also aids for effective and comprehensive optimisation.

3.7 Morph Attack Detection Potential

Considering the success rate of the newly generated dataset, we naturally choose to evaluate the morphing attack detection performance to also validate the robustness of the existing MAD mechanisms. Additionally, we investigate recent works about general face manipulation detection [54] [55] and some results are shown in the supplementary material. In this work, we focus on single image based morphing attack detection (S-MAD) as it perfectly suits our dataset. MAD has been widely addressed in the literature by developing the techniques based on both deep learning [56], [57], [58] [59] [60] and non-deep learning [61] [19] [62] [63] approaches. Readers can refer to [64] for an exclusive survey on face MAD. Owing to the recent works detailing the applicability of Hybrid features [35] and Ensemble features [36] in detecting morphing attacks, we choose to benchmark both Hybrid features [35] and Ensemble features [36]. While the Hybrid features [35] resort to extracting features using both scale space and color space combined with multiple classifiers, Ensemble features [36] employ a variety of textural features in conjunction with a set of classifiers. In common both approaches evaluate a wide variety of MAD mechanisms in a holistic manner supported by empirical results [35], [36]. In addition, the Hybrid features [35] mechanisms are also validated against the ongoing NIST FRVT morph challenge dataset [37] with the best performance in detecting printed and scanned morph images justifying our selection of algorithm to benchmark the newly composed database.

The reporting of MAD performance is following the ISO/IEC metrics [65] namely the Attack Presentation Classification Error Rate (APCER (%)) which defines the proportion of attack images (morph images) incorrectly classified as bona fide images and the Bona fide Presentation Classification Error Rate (BPCER (%)) in which bona fide images incorrectly classified as attack images are counted [65] along with the Detection Equal Error Rate (D-EER (%)). To evaluate the generated morphed face image's attack potential, we have sub-divided the newly generated database into two sets for training and testing that consists of independent data subjects with no overlap between the splits. The training set includes 690 bona fide images and 1190 morphed images. The testing set consists of 580 bona fide and 1310 morphed images. To effectively evaluate the performance of the MAD reflecting a real-life scenario, we report the results on both intra (training and testing dataset from the same morph generation approach) and inter (training on one type of morphing techniques and testing on another type of morphing techniques) evaluation of MAD mechanisms. Extensive experiments are performed on digital, print-scan and print-scan with compression data types to provide an in-depth analysis of the S-MAD performance. Table 9, 10, 11, 12 and 13 presents the quantitative results of MAD mechanisms on morph generation methods together with the SOTA morph generation techniques. Based on the results obtained from the intra-dataset experiments, we make some concrete observations as listed below:

- The intra-dataset evaluation indicates that the morphing attacks are detected with a good success rate irrespective of the type of generation.

Proposed Morph Generators	Case-study	Hyper-parameters weights				MMPMR (%)		FMMPMR (%)	
		λ_1	λ_2	λ_3	λ_4	Cognitec	ArcFace	Cognitec	ArcFace
MIPGAN -I	1	0.0004	10	1	1	93.94	93.49	84.39	75.61
	2	0.0001	10	1	1	92.66	91.15	79.66	72.94
	3	0.0002	1	10	1	94.17	91.9	84.34	75.66
	4	0.0002	1	1	10	83.16	82.14	67.19	59.46
	Proposed weights	0.0002	10	1	1	**94.36**	**94.45**	**84.65**	**85.94**
MIPGAN -II	1	0.0004	10	1	1	91.36	91.98	81.29	76.18
	2	0.0001	10	1	1	91.69	88.29	73.91	68.16
	3	0.0002	1	10	1	90.63	90.91	80.76	75.87
	4	0.0002	1	1	10	87.22	74.33	57.43	51.91
	Proposed weights	0.0002	10	1	1	**92.93**	**94.21**	**81.59**	**86.94**

TABLE 8: Quantitative results of hyper-parameters study

Morph Generation Type: Training	Morph Generation Type: Testing	MAD Algorithms	Digital			Print-scan			Print-scan with compression		
			D-EER(%)	BPCER @ APCER =		D-EER(%)	BPCER @ APCER =		D-EER(%)	BPCER @ APCER =	
				5%	10%		5%	10%		5%	10%
Landmarks-I [7]	Landmarks-I [7]	Ensemble Features [36]	0	0	0	2.35	1.45	0.96	2.58	1.71	1.54
		Hybrid Features [35]	0.16	0	0	1.85	0.85	0.34	2.25	1.12	0.51
	Landmarks-II [44]	Ensemble Features [36]	49.55	92.22	88.85	41.93	81.45	76.25	42.15	83.88	77.64
		Hybrid Features [35]	49.16	99.31	97.59	44.17	86.48	80.24	46.49	88.38	81.95
	StyleGAN [1]	Ensemble Features [36]	0.22	0	0	13.36	27.44	16.46	14.77	27.27	19.38
		Hybrid Features [35]	0.16	0	0	44.96	83.7	75.47	9.44	14.57	9.14
	MIPGAN-I	Ensemble Features [36]	39.16	73.14	65.35	9.45	14.57	8.74	8.95	15.26	9.26
		Hybrid Features [35]	46.82	86.62	81.64	12.32	19.72	13.2	9.74	15.95	8.91
	MIPGAN-II	Ensemble Features [36]	34.13	70.49	61.57	5.32	6.68	2.57	6.72	8.16	4.14
		Hybrid Features [35]	44.96	83.7	75.47	5.9	8.42	3.23	5.67	6.18	2.91

TABLE 9: Quantitative performance of MAD - Training- Landmarks-I [7]

Morph Generation Type: Training	Morph Generation Type: Testing	MAD Algorithms	Digital			Print-scan			Print-scan with compression		
			D-EER(%)	BPCER @ APCER =		D-EER(%)	BPCER @ APCER =		D-EER(%)	BPCER @ APCER =	
				5%	10%		5%	10%		5%	10%
Landmarks-II [44]	Landmarks-I [7]	Ensemble Features [36]	48.57	97.77	95.36	24.19	52.48	43.22	21.64	47.51	36.19
		Hybrid Features [35]	45.67	96.91	94.16	32.26	77.87	66.55	24.51	50.94	40.65
	Landmarks-II [44]	Ensemble Features [36]	3.62	2.22	0.68	6.32	7.97	2.42	5.57	6.41	2.42
		Hybrid Features [35]	1.53	0.17	0	5.21	5.19	3.14	5.37	5.71	3.46
	StyleGAN [1]	Ensemble Features [36]	29.67	61.92	52.48	27.18	61.57	50.6	29.18	62.14	52.48
		Hybrid Features [35]	34.76	74.44	62.95	34.8	67.23	58.14	23.17	49.22	38.25
	MIPGAN-I	Ensemble Features [36]	30.23	65.35	53.17	43.92	87.65	79.24	44.24	89.23	82.33
		Hybrid Features [35]	46.29	84.04	77.01	34.16	71.18	64.66	35.5	76.84	65.52
	MIPGAN-II	Ensemble Features [36]	27.13	58.83	45.45	33.57	77.35	65.52	40.46	84.9	75.47
		Hybrid Features [35]	46.82	83.53	75.81	35.91	77.18	65.24	36.5	79.24	68.78

TABLE 10: Quantitative performance of MAD - Training- Landmarks-II [44]

Morph Generation Type: Training	Morph Generation Type: Testing	MAD Algorithms	Digital			Print-scan			Print-scan with compression		
			D-EER(%)	BPCER @ APCER =		D-EER(%)	BPCER @ APCER =		D-EER(%)	BPCER @ APCER =	
				5%	10%		5%	10%		5%	10%
StyleGAN [1]	Landmarks-I [7]	Ensemble Features [36]	0.32	0	0	16.6	28.13	19.89	13.89	22.12	17.66
		Hybrid Features [35]	0.42	0	0	15.26	26.41	17.66	14.37	22.81	16.92
	Landmarks-II [44]	Ensemble Features [36]	44.72	89.53	80.61	38.31	78.5	69.15	38.84	83.7	74.17
		Hybrid Features [35]	45.65	90.22	84.56	34.18	81.95	70.53	32.93	78.5	64.12
	StyleGAN [1]	Ensemble Features [36]	0	0	0	0	0	0	0	0	0
		Hybrid Features [35]	0	0	0	0	0	0	0	0	0
	MIPGAN-I	Ensemble Features [36]	39.97	75.98	68.78	20.21	42.14	33.44	20.73	45.28	36.53
		Hybrid Features [35]	46.45	86.79	77.87	29.34	59.19	47.51	24.87	51.62	41.18
	MIPGAN-II	Ensemble Features [36]	39.93	73.58	66.89	15.78	28.14	19.38	13.72	28.98	16.63
		Hybrid Features [35]	44.72	82.16	73.75	19.36	43.22	28.64	16.98	32.93	23.84

TABLE 11: Quantitative performance of MAD - Training- StyleGAN [1]

Morph Generation Type: Training	Morph Generation Type: Testing	MAD Algorithms	Digital			Print-scan			Print-scan with compression		
			D-EER(%)	BPCER @ APCER =		D-EER(%)	BPCER @ APCER =		D-EER(%)	BPCER @ APCER =	
				5%	10%		5%	10%		5%	10%
MIPGAN-I	Landmarks-I [7]	Ensemble Features [36]	23.66	51.45	39.96	5.82	7.22	2.92	6.17	7.54	3.94
		Hybrid Features [35]	47.15	87.16	79.41	6.5	8.23	4.15	7.91	10.29	6.34
	Landmarks-II [44]	Ensemble Features [36]	35.38	82.33	68.95	41.67	95.14	83.53	43.68	96.01	85.44
		Hybrid Features [35]	28.62	75.64	61.4	44.38	95.66	85.78	38.18	90.46	78.16
	StyleGAN [1]	Ensemble Features [36]	17.72	37.22	26.58	12.19	26.24	15.26	11.82	24.69	14.23
		Hybrid Features [35]	31.16	64.32	53.85	11.99	19.2	13.72	9.93	18.15	9.94
	MIPGAN-I	Ensemble Features [36]	0	0	0	0	0	0	0	0	0
		Hybrid Features [35]	0	0	0	0	0	0	0	0	0
	MIPGAN-II	Ensemble Features [36]	2.15	0.17	0	0.68	0	0	0.64	0	0
		Hybrid Features [35]	1.36	0.34	0	0.86	0	0	0.8461	0	0

TABLE 12: Quantitative performance of MAD - Training- MIPGAN-I

Morph Generation Type: Training	Morph Generation Type: Testing	MAD Algorithms	Digital			Print-scan			Print-scan with compression		
			D-EER(%)	BPCER @ APCER =		D-EER(%)	BPCER @ APCER =		D-EER(%)	BPCER @ APCER =	
				5%	10%		5%	10%		5%	10%
MIPGAN-II	Landmarks-I [7]	Ensemble Features [36]	13.08	29.15	15.78	4.28	3.94	2.22	4.28	3.61	2.22
		Hybrid Features [35]	40.14	77.7	67.23	5.49	5.48	2.4	7.21	10.98	4.15
	Landmarks-II [44]	Ensemble Features [36]	32.37	84.9	70.32	39.2	90.12	82.32	44.17	95.49	88.73
		Hybrid Features [35]	23.88	63.8	45.62	40.22	88.9	79.2	38.96	94.28	82.14
	StyleGAN [1]	Ensemble Features [36]	12.51	22.29	15.78	13.72	29.67	18.18	14.25	31.73	20.41
		Hybrid Features [35]	24.7	49.74	41.85	12.87	26.58	14.75	11.86	26.92	15.09
	MIPGAN-I	Ensemble Features [36]	1.56	0.68	0.34	2.14	1.22	0.53	2.57	0.85	0.34
		Hybrid Features [35]	2.27	0.85	0.17	4.79	4.8	3.43	4.3	3.6	2.22
	MIPGAN-II	Ensemble Features [36]	0	0	0	0	0	0	0	0	0
		Hybrid Features [35]	0	0	0	0	0	0	0	0	0

TABLE 13: Quantitative performance of MAD - Training- MIPGAN-II

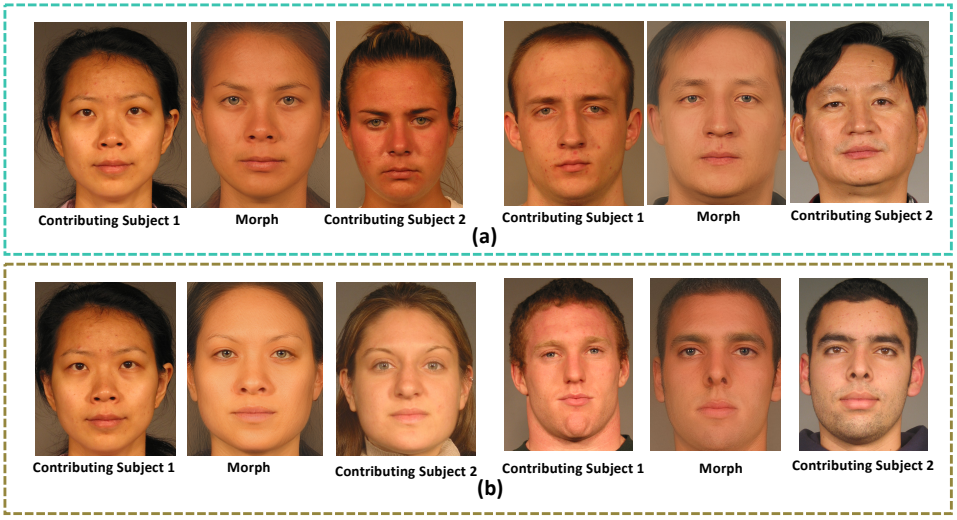


Fig. 13: Examples of morphed images that failed to attack FRS (a) morphed face images generated using proposed MIPGAN-I (b) morphed face images generated using proposed MIPGAN-II

- In general, the attack detection success rate is high with digital data when compared to print-scan and print-scan compression.
- Among the different types of morph generation techniques, the Landmark-II based morph generation shows the highest error rates. The attack images created using StyleGAN and proposed MIPGAN can be efficiently detected using both the employed approaches with high accuracy. This can be attributed to the noises that are synthesised using GANs due to the computational modifications performed on the latent space in GAN-based morph generation methods.

In the following, we discuss the important observations based on the results obtained from inter-dataset MAD analysis:

- The performance of the MAD techniques are degraded on all five different case studies as indicated in the Table 9, 10, 11, 12 and 13.
- Training MAD algorithms with one type of landmarks-based method did not show the improvement in detection performance of another kind of landmarks-based morph generation method.
- When MAD mechanisms are trained using the Landmarks-I [7] method, the degraded performance is

noted for all other morph generation methods except for the StyleGAN [1] based approach. This fact is also noted when we train the MAD techniques using StyleGAN [1] generated samples and test it with Landmarks-I [7] samples. Thus, the StyleGAN [1] based morph generation is easy to detect even when MAD mechanisms are not trained using the images from same morph generation scheme.

- When MAD algorithms are trained using Landmarks-II [44] samples, MAD algorithms indicate degraded performance on all other morph generation techniques.
- When MAD mechanisms are trained using the proposed MIPGAN-I generated samples. The MAD mechanisms indicate an excellent detection performance on MIPGAN-II samples. However, the detection performance of MAD methods is deceived with other morph generation techniques.
- It is interesting to note that when MAD mechanisms are trained using MIPGAN-I/MIPGAN-II, higher detection accuracy can be observed for print-scan and print-scan with compression data when compared to digital morph data. A possible reason is that the noise generated together with the morphed images using the proposed MIPGAN-I/MIPGAN-II can approximate the generated noise resulting from the print-scan and print-scan compression process.
- Based on the results of the inter-database MAD analysis, the detection of Landmarks-II [44] samples are challenging.

4 Limitations of Current Work and Potential Future Works

Despite the work presenting a new and robust approach to generate the morph attacks which is empirically evaluated using COTS FRS, this work has a few noted limitations. In the current scope of work, we evaluate the impact of print and scan (re-digitizing) using one printer reflecting a realistic scenario. The MAD mechanism employed in this work has not been investigated with a wide range of printers and scanners that may impact the MAD performance. While we assert that the MAD performance may not vary extremely, when tested with a wider combination of printers and scanners, that empirical evaluation is yet to be conducted in future works.

A second aspect is that the proposed approach needs pre-selection of ethnicity for generating stronger attacks. Figure 13 shows the example morphed face images generated using the proposed method using MIPGAN-I and MIPGAN-II that fail to get verified to contributing subjects when ethnicity pre-selection is not performed [7]. We notice that the selection of contributing subjects plays an important role with the proposed method to generate stronger attacks with MIPGAN. It is our assertion that the selection of contributing subjects with similar geometric structures (particularly ethnicity and age) can improve the performance of the proposed system, but that aspect needs further investigation.

5 Conclusion

Addressing the limitations of generating the robust morphing attacks using GAN, we have proposed a new architecture for generating face morphed images in this work. The proposed approach (MIPGAN with two variants) for devising robust morphing attacks uses identity prior driven GAN with a customized loss exploiting perceptual quality and identity factors to generate realistic images that can strongly threaten FRS. In order to validate the attack potential of the proposed morph generation method, we have created a new dataset consisting of $30,000$ morphed images and $15,240$ bona fide images. Both COTS and deep learning based FRS was evaluated empirically to measure the success rate of the new approach and vulnerability was reported indicating the applicability of the new approach and newly generated database. In a similar direction, the dataset is also validated for detection performance by studying two state-of-art MAD mechanisms. Despite the high attack detection success rate by employed MAD, we note that the morphed images generated by MIPGAN can severely threaten FRS in a present state without MAD in FRS.

References

[1] S. Venkatesh, H. Zhang, R. Raghavendra, K. Raja, N. Damer, and C. Busch, "Can gan generated morphs threaten face recognition systems equally as landmark based morphs? - vulnerability and detection," in *2020 International Workshop on Biometrics and Forensics (IWBF)*. IEEE, 2020, pp. 1–6.

[2] R. Raghavendra and C. Busch, "Presentation attack detection methods for face recognition systems: A comprehensive survey," *ACM Computing Surveys (CSUR)*, vol. 50, no. 1, pp. 1–37, 2017.

[3] M. Ferrara, A. Franco, and D. Maltoni, "The magic passport," in *IEEE International Joint Conference on Biometrics*. IEEE, 2014, pp. 1–7.

[4] R. Raghavendra, K. B. Raja, S. Venkatesh, and C. Busch, "Transferable deep-cnn features for detecting digital and print-scanned morphed face images," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2017, pp. 1822–1830.

[5] R. Raghavendra, K. B. Raja, and C. Busch, "Detecting morphed face images," in *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 2016, pp. 1–7.

[6] K. Raja, M. Ferrara, A. Franco, L. Spreeuwers, I. Batskos, F. de Wit Marta Gomez-Barrero, U. Scherhag, D. Fischer, S. Venkatesh, J. M. Singh, G. Li, L. Bergeron, S. Isadskiy, R. Ramachandra, C. Rathgeb, D. Frings, U. Seidel, F. Knopjes, R. Veldhuis, D. Maltoni, and C. Busch, "Morphing attack detection – database, evaluation platform and benchmarking," 2020.

[7] R. Raghavendra, K. B. Raja, S. Venkatesh, and C. Busch, "Face morphing versus face averaging: Vulnerability and detection," in *IEEE International Joint Conference on Biometrics (IJCB)*, 2017, pp. 555–563.

[8] U. Scherhag, A. Nautsch, C. Rathgeb, M. Gomez-Barrero, R. N. Veldhuis, L. Spreeuwers, M. Schils, D. Maltoni, P. Grother, S. Marcel, B. Ralph, R. Raghavendra, and C. Busch, "Biometric systems under morphing attacks: Assessment of morphing techniques and vulnerability reporting," in *2017 International Conference of the Biometrics Special Interest Group (BIOSIG)*. IEEE, 2017, pp. 1–7.

[9] Gov.uk. [Online]. Available: https://www.gov.uk/photos-for-passports/photo-requirements

[10] Department of Internal Affairs (DIA), NZ. [Online]. Available: https://www.passports.govt.nz/passport-photos/passport-photo-requirements/

[11] Photo for a passport or identity-card, netherlands. [Online]. Available: https://www.netherlandsworldwide.nl/countries/iran/living-and-working/photo-for-a-passport-or-identity-card

[12] S.-Y. Lee, K.-Y. Chwa, S. Y. Shin, and G. Wolberg, "Image metamorphosis using snakes and free-form deformations," in *SIGGRAPH*, vol. 95. Citeseer, 1995.

[13] T. Ucicr, "Feature-based image metamorphosis," *Computer graphics*, vol. 26, p. 2, 1992.

[14] S. Schaefer, T. McPhail, and J. Warren, "Image deformation using moving least squares," in *ACM transactions on graphics (TOG)*. ACM, 2006, pp. 533–540.

[15] D. W. Choi and C. J. Hwang, "Image morphing using mass-spring system," 2011.

[16] M. Bichsel, "Automatic interpolation and recognition of face images by morphing," in *Proc. of the Second Intl. Conf. on Automatic Face and Gesture Recognition*. IEEE Comput. Soc. Press, 1996.

[17] A. Makrushin, T. Neubert, and J. Dittmann, "Automatic generation and detection of visually faultless facial morphs," in *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 6: VISAPP, (VISIGRAPP 2017)*, 2017, pp. 39–50.

[18] J. Wu, "Face recognition jammer using image morphing," *Dept. Elect. Comput. Eng., Boston Univ., Boston, MA, USA, Tech. Rep. ECE-2011*, 2011.

[19] M. Hildebrandt, T. Neubert, A. Makrushin, and J. Dittmann, "Benchmarking face morphing forgery detection: Application of stirtrace for impact simulation of different processing steps," in *International Workshop on Biometrics and Forensics (IWBF 2017)*, 2017, pp. 1–6.

[20] C. Seibold, W. Samek, A. Hilsmann, and P. Eisert, "Detection of face morphing attacks by deep learning," in *International Workshop on Digital Watermarking*, 2017, pp. 107–120.

[21] U. Scherhag, R. Raghavendra, K. Raja, M. Gomez-Barrero, C. Rathgeb, and C. Busch, "On the vulnerability of face recognition systems towards morphed face attack," in *International Workshop on Biometrics and Forensics (IWBF 2017)*, 2017, pp. 1–6.

[22] D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, and S. K. Nayar, "Face swapping: Automatically replacing faces in photographs," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 39:1–39:8, 2008. [Online]. Available: http://doi.acm.org/10.1145/1360612.1360638

[23] Y. Weng, L. Wang, X. Li, M. Chai, and K. Zhou, "Hair interpolation for portrait morphing," *Computer Graphics Forum*, vol. 32, no. 7, pp. 79–84, October 2013.

[24] U. Scherhag, C. Rathgeb, J. Merkle, R. Breithaupt, and C. Busch, "Face recognition systems under morphing attacks: A survey," *IEEE Access*, vol. 7, pp. 23 012–23 026, 2019.

[25] "Abrosoft fantamorph," FantaMorph,Abrasoft:http://www.fantamorph.com/, 2020, accessed: May 2020.

[26] I. C. A. Organization, "Machine readable passports – part 1 – introduction," http://www.icao.int/publications/Documents/9303_p1_cons_en.pdf, International Civil Aviation Organization (ICAO), 2015.

[27] International Civil Aviation Organization, "Machine readable passports – part 9 – deployment of biometric identification and electronic storage of data in eMRTDs," http://www.icao.int/publications/Documents/9303_p9_cons_en.pdf, International Civil Aviation Organization (ICAO), 2015.

[28] N. Damer, A. M. Saladié, A. Braun, and A. Kuijper, "Morgan: Recognition vulnerability and attack detectability of face morphing attacks created by generative adversarial network," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, Oct 2018, pp. 1–10.

[29] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4401–4410.

[30] C. S. GmbH. (2020) Facevacs technology - version 9.4.2. [Online]. Available: https://www.cognitec.com/facevacs-technology.html

[31] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4690–4699.

[32] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of stylegan," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8110–8119.

[33] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, Jin Chang, K. Hoffman, J. Marques, Jaesik Min, and W. Worek, "Overview of the face recognition grand challenge," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, June 2005, pp. 947–954 vol. 1.

[34] S. Venkatesh, R. Raghavendra, K. Raja, L. Spreeuwers, R. Veldhuis, and C. Busch, "Morphed face detection based on deep color residual noise," in *ninth International Conference on Image Processing Theory, Tools and Applications (IPTA 2019)*. IEEE, 2019, pp. 1–5.

[35] R. Ramachandra, S. Venkatesh, K. Raja, and C. Busch, "Towards making morphing attack detection robust using hybrid scale-space colour texture features," in *2019 IEEE 5th International Conference on Identity, Security, and Behavior Analysis (ISBA)*. IEEE, 2019, pp. 1–8.

[36] S. Venkatesh, R. Raghavendra, K. Raja, and C. Busch, "Single image face morphing attack detection using ensemble of features," in *23rd International Conference on Information Fusion*, 2020, pp. 1–5.

[37] NIST. (2020) Frvt morph web site. [Online]. Available: https://pages.nist.gov/frvt/html/frvt_morph.html

[38] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of stylegan," *arXiv preprint arXiv:1912.04958*, 2019.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[40] O. Richter and R. Wattenhofer, "Treeconnect: A sparse alternative to fully connected layers," in *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, 2018, pp. 924–931.

[41] S. Venkatesh, K. Raja, R. Raghavendra, and C. Busch, "On the influence of ageing on face morph attacks: Vulnerability and detection," in *International Joint Conference on Biometrics (IJCB)*, September 2020, pp. 1–8.

[42] R. Abdal, Y. Qin, and P. Wonka, "Image2stylegan++: How to edit the embedded images?" *arXiv preprint arXiv:1911.11544*, 2019.

[43] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*. Springer, 2016, pp. 694–711.

[44] M. Ferrara, A. Franco, and D. Maltoni, "Decoupling texture blending and shape warping in face morphing," in *2019 International Conference of the Biometrics Special Interest Group (BIOSIG)*. IEEE, 2019, pp. 1–5.

[45] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, vol. 2. Ieee, 2003, pp. 1398–1402.

[46] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[47] N. Damer, F. Boutros, A. Moseguí Saladié, F. Kirchbuchner, and A. Kuijper, "Realistic dreams: Cascaded enhancement of gan-generated images with an example in face morphing attacks," 10 2019.

[48] Dnp printer. [Online]. Available: http://dnpphoto.com/en-us/Products/Printers/DS820A

[49] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, pp. 67–74.

[50] F. COTS, "Verilook cots," http://www.neurotechnology.com/verilook.html, 2015, accessed: 2015-02-08.

[51] X. Wu, R. He, Z. Sun, and T. Tan, "A light cnn for deep face representation with noisy labels," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, pp. 2884–2896, 2018.

[52] FRONTEX, "Best practice technical guidelines for automated border control ABC systems," 2015.

[53] U. Scherhag, A. Nautsch, C. Rathgeb, M. Gomez-Barrero, R. N. J. Veldhuis, L. Spreeuwers, M. Schils, D. Maltoni, P. Grother, S. Marcel, R. Breithaupt, R. Ramachandra, and C. Busch, "Biometric systems under morphing attacks: Assessment of morphing techniques and vulnerability reporting," in *Intl. Conf. of the Biometrics Special Interest Group BIOSIG 2017*, 2017, pp. 1–7.

[54] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "Faceforensics++: Learning to detect manipulated facial images," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 1–11.

[55] A. Jain, P. Majumdar, R. Singh, and M. Vatsa, "Detecting gans and retouching based digital alterations via dad-hcnn," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 672–673.

[56] M. Ferrara, A. Franco, and D. Maltoni, "Face morphing detection in the presence of printing/scanning and heterogeneous image sources," *CoRR*, vol. abs/1901.08811, 2019. [Online]. Available: http://arxiv.org/abs/1901.08811

[57] C. Seibold, A. Hilsmann, and P. Eisert, "Style your face morph and improve your face morphing attack detector," in *2019 International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2019, pp. 1–6.
[58] A. Makrushin, C. Kraetzer, J. Dittmann, C. Seibold, A. Hilsmann, and P. Eisert, "Dempster-shafer theory for fusing face morphing detectors," in *2019 27th European Signal Processing Conference (EUSIPCO)*, 2019, pp. 1–5.
[59] D. Ortega-Delcampo, C. Conde, D. Palacios-Alonso, and E. Cabello, "Border control morphing attack detection with a convolutional neural network de-morphing approach," *IEEE Access*, pp. 1–1, 2020.
[60] F. Peng, L.-B. Zhang, and M. Long, "FD-GAN: Face de-morphing generative adversarial network for restoring accomplice's facial image," *IEEE Access*, vol. 7, pp. 75 122–75 131, 2019. [Online]. Available: https://doi.org/10.1109%2Faccess.2019.2920713
[61] M. Hildebrandt, T. Neubert, A. Makrushin, and J. Dittmann, "Benchmarking face morphing forgery detection: Application of stirtrace for impact simulation of different processing steps," in *2017 5th International Workshop on Biometrics and Forensics (IWBF)*, 2017, pp. 1–6.
[62] C. Seibold, A. Hilsmann, and P. Eisert, "Reflection analysis for face morphing attack detection," *arXiv preprint arXiv:1807.02030*, 2018.
[63] C. Kraetzer, A. Makrushin, T. Neubert, M. Hildebrandt, and J. Dittmann, "Modeling attacks on photo-id documents and applying media forensics for the detection of facial morphing," in *Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security*, 2017, pp. 21–32.
[64] S. Venkatesh, R. Ramachandra, K. Raja, and C. Busch, "Face morphing attack generation & detection: A comprehensive survey," *IEEE Transactions on Technology and Society*, 2021.
[65] ISO/IEC JTC1 SC37 Biometrics, *ISO/IEC 30107-3. Information Technology - Biometric presentation attack detection - Part 3: Testing and Reporting*, International Organization for Standardization, 2017.