

Received April 27, 2019, accepted May 22, 2019, date of publication June 4, 2019, date of current version June 21, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2920713

FD-GAN: Face De-Morphing Generative Adversarial Network for Restoring Accomplice's Facial Image

FEI PENG^{1,2}, (Member, IEEE), LE-BING ZHANG^{2,3}, AND MIN LONG⁴

¹School of Cyberspace Science, Dongguan University of Technology, Dongguan 523808, China

²College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China

³School of Computer Science and Engineering, Huaihua University, Huaihua 418000, China

⁴College of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha 410114, China

Corresponding author: Fei Peng (eepengf@gmail.com)

This work was supported by the National Natural Science Foundation of China under Grant 61572182 and Grant 61370225.

ABSTRACT Face morphing attack is proved to be a serious threat to the existing face recognition systems. Although a few face morphing detection methods have been put forward, the face morphing accomplice's facial restoration remains a challenging problem. In this paper, a face de-morphing generative adversarial network (FD-GAN) is proposed to restore the accomplice's facial image. It utilizes the symmetric dual network architecture and two levels of restoration losses to separate the identity feature of the morphing accomplice. By exploiting the captured facial image (containing the criminal's identity) from the face recognition system and the morphed image stored in the e-passport system (containing both criminal and accomplice's identities), the FD-GAN can effectively restore the accomplice's facial image. The experimental results and analysis demonstrate the effectiveness of the proposed scheme. It has great potential to be applied for tracing the identity of face morphing attack's accomplice in criminal investigation and judicial forensics.

INDEX TERMS Face de-morphing, face morphing attack, facial restoration, generative adversarial network.

I. INTRODUCTION

With the development of face biometrics, face recognition systems (FRS) with high accuracy are extensively used in our daily life. In real applications, a prevalently used FRS is Automatic Border Control (ABC) system. It can expediently verify a person's identity with one's electronic machine readable travel document (eMRTD) [1], which contains a facial reference image. In the past twelve years, more than 800 million eMRTD instances have been issued, following the specifications from the International Civil Aviation Organization (ICAO).

Recently, it was found that a novel identity theft scenario can easily spoof the FRS of ABC [2]. The idea of the attack is: a morphed facial image (a combination of two or more real facial images) is first generated, and it looks like multiple real persons (e.g. a criminal and an accomplice). Then the morphed image is enrolled as an identity template of the FRS. In a successful attack, both the criminal and the accomplice can

The associate editor coordinating the review of this manuscript and approving it for publication was Irene Amerini.

match the template stored in the FRS. It means that a wanted criminal is able to obtain a legitimate e-passport or eMRTD by morphing his facial image with an accomplice. Some samples of the morphed facial images are shown in Figure 1.

After this idea was put forward, some concerns about the vulnerability of commercial FRS with respect to morphed face attack have been investigated [3]–[7]. It was proved that face morphing attack is posing a serious threat to the existing FRS. To countermeasure this situation, some face morphing detection methods have been proposed to protect the security of FRS [8]–[16]. However, all of these methods only detected the presence of face morphing attack, the restoration of the accomplice's facial image has not been seriously discussed.

Recently, a face de-morphing method was proposed in [17]. It restored the facial image of the morphing accomplice by reversing the face morphing operation. Unfortunately, it depends on the prior knowledge on the face morphing operation and morphing parameters of the face morphing. If the de-morphing parameter is close to the morphing parameter, high morphing detection accuracy can be achieved. However, if the de-morphing parameter is considerable different



FIGURE 1. Example of the morphed images generated from two contributors.

from the morphing parameter, detection accuracy will be greatly affected.

Inspired by the popular use of generative adversarial networks (GAN) [18] in image synthesis, we consider the restoration of the accomplice's facial image from a perspective of a learning-based generation. There are some differences compared with previous work [17].

- The goals are different. The facial restoration method in [17] is used for face morphing attack detection. While our proposed FD-GAN focuses on how to restore the face morphing attack accomplice's facial image from a morphed facial image. It is a further forensics research based on the morphed facial image detection. It means that the input of FD-GAN contains the morphed facial image.
- The ideas for facial restoration are different. In [17], the reversion of face morphing process was used to realize the accomplice's facial restoration, while our proposed FD-GAN realizes the accomplice's facial restoration by disentangling the identity features of the morphing contributors hidden in the morphed facial image.

To accomplish this, FD-GAN is designed, and it utilizes a symmetric dual network architecture and two levels of restoration losses to extract the accomplice's identity feature, and restore the facial image of the accomplice. The main contributions are in the following.

- FD-GAN is proposed to restore the accomplice's facial image without prior knowledge of the morphed facial images (e.g. face morphing process and morphing parameters). To the best of our knowledge, it is the first attempt to exploit learning-based generation approach for facial image restoration in face morphing detection.
- A symmetric dual network architecture, pixel-level and feature-level restoration losses of FD-GAN are

designed to disentangle the identity features of the morphing contributors hidden in the morphed facial image.

- Experimental results and analysis illustrate it can effectively restore the facial image of the morphing accomplice.

The rest of the paper is organized as follows: the related work is introduced in Section II. The proposed FD-GAN is described in Section III. Experimental results are provided in Section IV. Finally, some conclusions are drawn in Section V.

II. RELATED WORK

Currently, the researches related to face morphing are mainly concentrated on the vulnerability of FRS to face morphing attack and the face morphing detection.

A. VULNERABILITY OF FRS TO FACE MORPHING ATTACK

Face morphing attack was first introduced in [2] by M. Ferrera *et al.* With a morphed facial image, which is similar to the appearance of multiple persons, it can match with multiple persons in FRS. Meanwhile, it indicates that a wanted criminal may acquire a legitimate eMRTD by using a morphed image, which is composed of his own facial image and an accomplice's facial image. However, the morphed facial images are manually generated, which is time-consuming and not suitable for generating large amounts of morphing images to verify the vulnerability of FRS. After that, an automatic generation of visually faultless facial morphing image was proposed in [9]. With this technique, a large number of morphed facial images (including complete morphing and splicing morphing images) can be generated. Furthermore, the quality of the morphed images is verified by human observers (including face recognition experts) and a commercial-off-the-shelf FRS (Luxand FaceSDK 6.1).

Recently, Robertson *et al.* examined the potential route of using facial morphing in fraudulent documents, and found that it is possible to forge identity by morphing facial images [3]. Meanwhile, Gomez-Barrero *et al.* also confirmed that not only face recognition system but also iris and fingerprint recognition systems are vulnerable to morphing attack [4], and a framework to evaluate the vulnerability of biometric systems to morphing attack is proposed. Furthermore, Scherhag *et al.* proposed new metrics including mated morph presentation match rate (MMPMR) and relative morph match rate (RMMR) to evaluate biometric systems' vulnerability to morphing attacks [5]. After that, some other metrics including attack presentation classification error rate (APCER) and the bona fide presentation classification error rate (BPCER) are complemented to evaluate the biometric systems' vulnerability to morphing attacks [6]. In addition, the vulnerability of deep learning based FRS to morphing attack was investigated in [7], and it is found that morphing attacks can degrade the performance of FRS to a certain extent.

B. FACE MORPHING DETECTION

The existing face morphing methods can be divided into two categories depending on whether the auxiliary image is used or not, and they are blind detection methods and non-blind detection methods [19]. Most face morphing detection methods belong to the first category.

1) BLIND DETECTION METHODS

Considering the texture discrepancy between morphed facial image and real facial image, the first work on automated face morphing detection was proposed in [8]. The micro-texture pattern represented by binarized statistical image features (BSIF) are extracted for detection. With linear support vector machine (SVM) classification, it can achieve good detection performance. As the morphed facial image is usually generated by the existing real facial images and stored in JPEG format, it will cause image quality degradation and JPEG artifact. Based on this fact, a morphed face detection method based on JPEG compression feature was proposed in [9] and [10], where the Benford feature extracted from quantized DCT coefficients is utilized to detect the morphed facial image. A new modeling approach for face morphing attacks was proposed by Kraetzer *et al.* [11]. Based on the modeling approaches, two different types face morphing attacks as well as a forensic morphing detector are implemented and evaluated. Eight features including scale invariant feature transform (SIFT), speed up robust feature (SURF), oriented BRIEF, features from accelerated segment test (FAST), adaptive and generic accelerated segment test (AGAST), Canny and Sobel edge operators are selected for the detector. At the same time, a detection approach for face morphing forgeries based on a continuous image JPEG degradation was proposed by Neubert [12]. It is similar to the method in [11], but only SIFT, AGAST and Shi-Tomasi features are used.

In [13], deep convolutional neural network (D-CNN) features were proposed to detect both digital and print-scanned morphed facial images. Meanwhile, a morphing attack detection approach based on deep learning was proposed in [14]. Three current widely-used network architectures are investigated, and it is found that the pre-trained VGG19 network can achieve the best performance among them.

Recently, Debiasi *et al.* utilized some statistical characteristics of sensor pattern noise (SPN) spectrum histogram (peak position, peak value, and the product of peak position and peak value) to detect face morphing attack [15]. It can achieve a fairly good detection performance. At the same time, inspired by the idea of image source identification, a face morphing detection method based on SPN was put forward by us [16]. The morphed facial image is viewed as a computer-generated image, and its SPN is different from that of the real facial image.

2) NON-BLIND DETECTION METHODS

Different from blind face morphing detection, non-blind face morphing detection usually needs an extra auxiliary image.

Ferrara *et al.* [17] first proposed a method for restoring the morphing attack accomplice's facial image, which is named as "face de-morphing". With an auxiliary image captured from FRS and a morphed image stored in the e-passport, it can restore the facial image of the accomplice by reversing the morphing process. However, it relies on some prior knowledge about the morphing operation process and the morphing parameters of the face morphing. Once an attacker uses other morphing operations or the morphing parameters are different from the recommended morphing parameters in [17], the restored effect will be greatly affected.

From the above analysis, face morphing detection is an emerging technology to protect biometrics. Although some work has been done to detect the existence of face morphing attack, there still has no reliable approach for restoring the facial image of the morphing accomplice.

In recent years, GAN [18] has been successfully applied in various face synthesis related tasks such as photorealistic frontal synthesis [20], anti-makeup [21], and identity preserving [22]. GAN based face synthesis shows powerful advantages in realistic face generation. To countermeasure the defects of the existing methods for restoring the morphing attack accomplice's facial image, a face de-morphing generative adversarial network FD-GAN is proposed for restoring accomplice's facial image from the perspective of learning-based generation in this paper.

III. THE PROPOSED FD-GAN

A. MOTIVATION

In recent years, face morphing detection has received extensive attention and a series of effective detection methods have been proposed. However, the research on facial restoration of face morphing accomplice is still in its infancy. Therefore, the goal of this paper is to achieve facial restoration of face morphing accomplice from the existing morphed facial images. It is a further forensics research based on face morphing detection.

As a morphed facial image is generally a synthesis of two different contributors (e.g. a criminal and an accomplice), it is reasonable to assume that the morphed image contains the identity features of two contributors at the same time. Therefore, if the contributor's identity feature hidden in the morphed facial image can be effectively separated, then the facial image of the morphing accomplice can be restored.

To effectively disentangle the identity features of the morphing contributors hidden in the morphed facial image, FD-GAN with symmetric dual network architecture is proposed.

B. OVERALL FRAMEWORK OF FD-GAN

As illustrated in Figure 2, the proposed FD-GAN framework is different from the previous GAN-based identity preserved face synthesis, the target of this paper is to restore the accomplice's facial image I_b^1 with a morphed facial image I_{ab} and the criminal's facial image I_a . Here, the notation of I_x^n

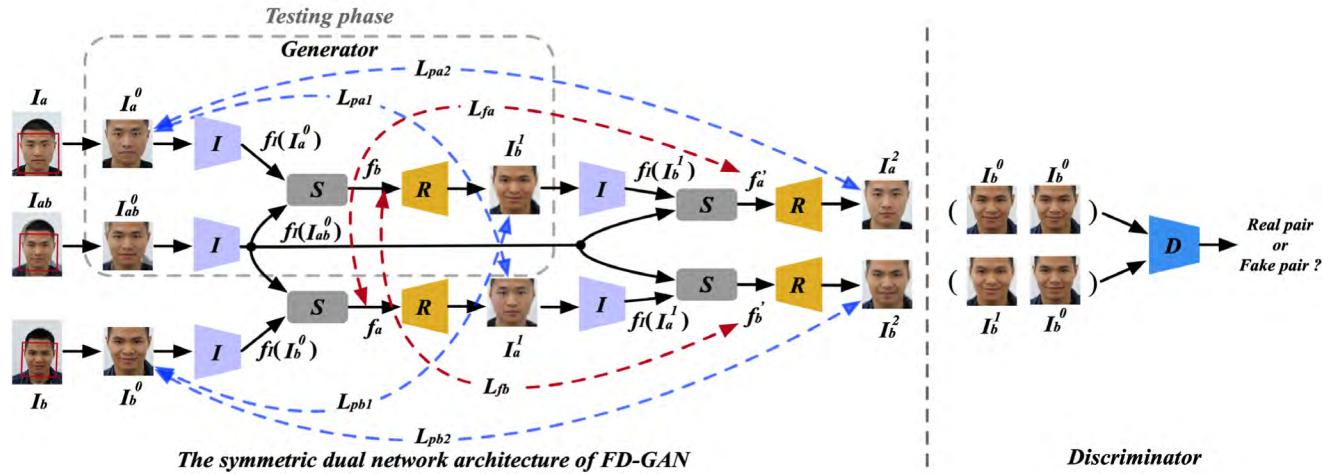


FIGURE 2. The overview of FD-GAN. The inputoutput are drawn with solid lines, The loss functions are drawn with the dashed lines, Blue dashed lines represent pixel-level ress, and red dashed lines represent feature-level restoration loss. Network modules with the same color indicate that they have the same network structure and share parameters in the training stage.

represents facial image, where $x \in \{a, b, ab\}$, and $n \in \{0, 1, 2\}$, a represents the identity of criminal, b represents the identity of accomplice, ab represents a morphed facial image generated by criminal a and accomplice b , 0 represents alignment facial image, I represents the restored image from previous-level generation network, and 2 represents the restored image from latter -level generation network.

Here, a symmetric dual network architecture is proposed to accomplish the separation of the morphing contributor's identity. It consists of four generators with shared weights in a symmetric dual structure. Each generator contains two identity encoder networks I , an identity separation network S ; and a facial restoration network R . In addition, an adversarial discriminator D is designed to guarantee the restored facial image I_b^1 not only resembles a realistic image but also keeps the identity of morphing accomplice I_b^0 . Therefore, it receives two facial image pairs, there are the real pair images $\{I_b^0, I_b^0\}$ and the fake (restored) pair images $\{I_b^0, I_b^1\}$.

In the training stage, FD-GAN receives three input facial images. There are criminal facial image I_a , accomplice facial image I_b , and morphed facial image I_{ab} . Through the training of symmetric dual network and discriminator D , FD-GAN learns to separate the identity features of face morphing contributors and restore the corresponding face images based on the separated identity features. In the testing stage, FD-GAN requires two input facial images. One is an image I_a of a criminal captured by an FRS, and the other is a morphed facial image I_{ab} synthesized from a criminal and an accomplice, which is stored in the e-passport system. The accomplice's facial image I_b^1 can be restored by the generator.

C. THE SYMMETRIC DUAL NETWORK ARCHITECTURE

As mentioned above, the morphed image contains the identity characteristics of two contributors (e.g. a criminal and an accomplice) at the same time. It is reasonable to assume that if the accomplice's identity feature f_b can be separated from a

criminal's facial image I_a and a morphed image I_{ab} , the criminal's identity feature f_a also can be separated from the same face morphing image I_{ab} and an accomplice's facial image I_b . It means the morphing contributors' identity recovery operations should be symmetrical. Therefore, the structure of FD-GAN should have a symmetrical network structure.

Furthermore, in order to guarantee the efficacy of the separated identity features, a dual network is intuitively considered. Assuming that once an accomplice's facial image I_b^1 is restored from I_a and I_{ab} , if this restoration is valid, I_b^1 can be further used to restore the criminal's facial image I_a^2 . In other words, with this dual architecture, the latter-level generation network can be used to verify the effectiveness of the forestage generation network, and guide the previous-level generation network to separate the identity in the right way.

Based on the above, FD-GAN with a symmetric dual network architecture is designed, and it is depicted in details in Figure 2. Unlike Cycle-GAN [23], which is used for domain-to-domain image style conversion by two reciprocal generators, our FD-GAN does not deal with the conversion problem from one image domain to another, and it have four same generators.

D. GENERATOR ARCHITECTURE AND GENERATOR LOSS

In FD-GAN, the generator aims to restore a desirable morphing accomplice's facial image from a morphed facial image and a criminal's facial image. The generator contains three parts: 1) an identity encoder network I ; 2) an identity separation network S ; and 3) a facial restoration network R .

Given a criminal's facial image I_a and a morphed facial image I_{ab} , face alignment is first performed to them, and then the identity encoder network I is used to extract the identity encoder features $f_I(I_a^0)$ and $f_I(I_{ab}^0)$, respectively. It needs to be mentioned that the identity encoder feature is different from the restored identity feature, e.g. $f_I(I_b^0)$ is different from f_b in Figure 2. After that, two identity encoder features

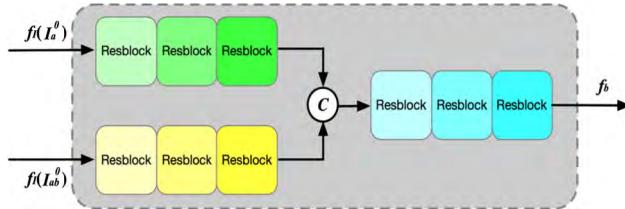


FIGURE 3. The structure of the identity separation network S .

are sent to the identity separation network S , which is used to separate the collaborator's identity feature. The structure of S is shown in Figure 3. The identity coding features $f_1(I_a^0)$ and $f_1(I_{ab}^0)$ are passed through three residual blocks [24], respectively, and then they are concatenated. After another three residual blocks, the restored identity feature f_b is obtained. Finally, the restored facial image I_b^1 can be obtained by facial restoration network R .

The details of the identity encoder network I and the facial restoration network R are shown in Table 1 and Table 2, respectively. For identity encoder network I , four convolution layers are used. Similar to DC-GAN [28], the pooling layer is replaced with a strided convolution layer. For the facial restoration network R , it is an inverse structure of I . The up-sampling layers are used, and the deconvolution layers are replaced with convolution layers. Using this “up-sampling + convolution layer” structure can effectively suppress the “checkerboard” artifact of the traditional de-convolution layer in the generated images. The ReLU activation is used in network I , S , and R .

TABLE 1. The architecture of identity encoder network I .

Group Name	Configuration (filter/stride)
Conv1	$7 \times 7/1$ Conv@64 + BN + ReLU
Conv2	$3 \times 3/2$ Conv@128 + BN + ReLU
Conv3	$3 \times 3/2$ Conv@256 + BN + ReLU
Conv4	$3 \times 3/2$ Conv@512 + BN + ReLU

TABLE 2. The architecture of facial restoration network R .

Group Name	Configuration (filter/stride)
Upconv1	Upsampling + $3 \times 3/1$ Conv@256 + BN + ReLU
Upconv2	Upsampling + $3 \times 3/1$ Conv@128 + BN + ReLU
Upconv3	Upsampling + $3 \times 3/1$ Conv@64 + BN + ReLU
Upconv4	$7 \times 7/1$ Conv@3 + Tanh

The generator receives three kinds of losses for updating parameters: pixel-level restoration loss L_{pxl} , feature-level restoration loss L_f and adversarial loss L_{adv} . The generator's total loss is

$$L_G = L_{pxl} + \lambda_1 L_f + \lambda_2 L_{adv}, \quad (1)$$

where λ_1, λ_2 represent the weights of the corresponding losses, respectively.

1) PIXEL-LEVEL RESTORATION LOSS

The pixel-level restoration loss L_{pxl} is composed of two parts: pixel-wise loss L_{pix} and symmetry loss L_{sym} , and it is defined as

$$L_{pxl} = L_{pix} + \beta_1 L_{sym}, \quad (2)$$

where β_1 represents the weight of symmetry loss L_{sym} .

Pixel-wise loss L_{pix} continuously pushes the restored facial image to the ground truth as close as possible. Considering that our proposed FD-GAN uses a symmetric dual network architecture during training stage, there are four pairs of pixel-wise losses L_{pb1} , L_{pb2} , L_{pa1} , and L_{pa2} . They are defined as

$$L_{pix} = L_{pb1} + L_{pb2} + L_{pa1} + L_{pa2}, \quad (3)$$

$$L_{pb1} = \|I_b^1 - I_b^0\|_1, \quad (4)$$

$$L_{pb2} = \|I_b^2 - I_b^0\|_1, \quad (5)$$

$$L_{pa1} = \|I_a^1 - I_a^0\|_1, \quad (6)$$

$$L_{pa2} = \|I_a^2 - I_a^0\|_1, \quad (7)$$

where $I_b^1 = G(I_a^0, I_{ab}^0)$, $I_a^1 = G(I_b^0, I_{ab}^0)$, $I_b^2 = G(G(I_b^0, I_{ab}^0), I_{ab}^0)$, $I_a^2 = G(G(I_a^0, I_{ab}^0), I_{ab}^0)$, $G(\cdot, \cdot)$ represents the generator of FD-GAN, and $\|\cdot\|_1$ represents L_1 -norm.

As symmetry is a prominent characteristic of human face, it is taken into account as a constraint to guarantee the rationality of the restored face. Similarly, there also have four symmetry losses $L_{symI_b^1}$, $L_{symI_b^2}$, $L_{symI_a^1}$, and $L_{symI_a^2}$, and they are defined as

$$L_{sym} = L_{symI_b^1} + L_{symI_b^2} + L_{symI_a^1} + L_{symI_a^2}, \quad (8)$$

$$L_{symI_b^1} = \frac{1}{h \times w/2} \sum_{i=1}^h \sum_{j=1}^w \|(I_b^1)_{i,j} - (I_b^1)_{i,w-j+1}\|_1, \quad (9)$$

$$L_{symI_b^2} = \frac{1}{h \times w/2} \sum_{i=1}^h \sum_{j=1}^w \|(I_b^2)_{i,j} - (I_b^2)_{i,w-j+1}\|_1, \quad (10)$$

$$L_{symI_a^1} = \frac{1}{h \times w/2} \sum_{i=1}^h \sum_{j=1}^w \|(I_a^1)_{i,j} - (I_a^1)_{i,w-j+1}\|_1, \quad (11)$$

$$L_{symI_a^2} = \frac{1}{h \times w/2} \sum_{i=1}^h \sum_{j=1}^w \|(I_a^2)_{i,j} - (I_a^2)_{i,w-j+1}\|_1, \quad (12)$$

2) FEATURE-LEVEL RESTORATION LOSS

Besides pixel-level restoration loss, generator loss also contains feature-level restoration loss. For the proposed symmetrical dual network architecture, the restored identity features in each stage should be as consistent as possible. There are

two pairs of feature-level restoration losses L_{fa} and L_{fb} , and they are defined as

$$L_f = L_{fa} + L_{fb}, \quad (13)$$

$$L_{fa} = \|f_a' - f_a\|_1, \quad (14)$$

$$L_{fb} = \|f_b' - f_b\|_1. \quad (15)$$

3) ADVERSARIAL LOSS

The adversarial loss of the generator for distinguishing the real pair facial images $\{I_b^0, I_b^0\}$ from the fake (restored) pair facial images $\{I_b^0, I_b^1\}$ is calculated as

$$L_{adv} = \mathbb{E}_{I_b^0 \sim P(I_b^0), I_a^0 \sim P(I_a^0), I_{ab}^0 \sim P(I_{ab}^0)} [(D(I_b^0, I_b^1) - 1)^2], \quad (16)$$

where $D(\cdot, \cdot)$ represents the discriminator of FD-GAN, and \mathbb{E} represents expectation.

E. DISCRIMINATOR ARCHITECTURE AND LOSS

Different from the traditional GAN discriminators, the discriminator of FD-GAN determines both the real pair facial images $\{I_b^0, I_b^0\}$ and the restored pair facial images $\{I_b^0, I_b^1\}$. The details of discriminator D are shown in Table 3. In the network D , three convolution layers are used, and the Leaky ReLU activation is used in network D .

TABLE 3. The architecture of discriminator network D.

Group Name	Configuration (filter/stride)
Conv1	$3 \times 3/2$ Conv@64 + Leaky ReLU
Conv2	$3 \times 3/2$ Conv@128 + BN + Leaky ReLU
Conv3	$3 \times 3/2$ Conv@256 + BN + Leaky ReLU
FC1	Fully Connected Layer+ Sigmoid

The discriminator requires the restored facial image I_b^1 should resemble a realistic facial image, and it should belong to the same identity as I_b^0 . Here, LS-GAN [25] loss instead of the negative log likelihood, which was used in the original GAN, is used. The discriminator loss is defined as

$$L_D = \mathbb{E}_{I_b^0 \sim P(I_b^0)} [(D(I_b^0, I_b^0) - 1)^2] + \mathbb{E}_{I_b^0 \sim P(I_b^0), I_a^0 \sim P(I_a^0), I_{ab}^0 \sim P(I_{ab}^0)} [(D(I_b^0, I_b^1))^2]. \quad (17)$$

IV. EXPERIMENTS AND ANALYSIS

As there is no publicly available facial image dataset for face morphing detection, we built a face morphing database (FM-database) and then it is used to evaluate the effectiveness of the proposed FD-GAN.

A. THE SETUP OF THE FM-DATABASE

The FM-database contains 63 subjects (27 females and 36 males). It is divided into three disjoint subsets: training set, development set and testing set, and they consist of 28, 7 and 28 subjects, respectively. All subjects were photographed

twice with an interval of one week and 10~12 frontal images were taken each time. Thus, there are about 23 images for each of 63 subjects. One image with ideal quality of each subject is chosen for creating morphed facial images and the rest are used to construct the bona fide facial images (with a total of 1378 images). The facial images of all subjects are captured by Canon 550D SLR camera following the ICAO guidelines on eMRTD. To guarantee the disjoint nature of the database as mentioned above, the morphed facial images are independently generated in training, development, and testing set. In each subset, face morphing images are automatically generated according to the workflow proposed in [9].

To effectively evaluate the validity of the FM database, a commercial FRS Megvii Face++ Compare API [26] is exploited. A successful morphing attack to the FRS is defined as that both subjects of the morphed image can be successfully matched with morphed image by the FRS [2]. It means that each morphed facial image needs to be compared with the images of all subjects, which have contributed to the morphing process. By using Face++, the Compare API requires to be called twice, and only two returned scores are both greater than the recommended threshold can the morphed facial image be considered to be a successful face morphing attack. The recommended threshold of Face++ Compare API is 62.37 when FAR is 0.1% (FRONTEX recommends that the operation point of FRS in border control scenarios is at FAR=0.1%).

From previous literature [2], [17], pixel fusion factor of face morphing has significant influence on the recognition accuracy of the FRS. Therefore, 9 groups of 777 morphed facial images were generated according to pixel fusion factor varied from 0.1 to 0.9. Here, the standardized ISO metrics [27], impostor attack presentation match rate (IAPMR) is used for evaluating the validity of the generated face morphing images. The IAPMRs of the morphed facial images under different pixel fusion factors are listed in Table 4.

According to different pixel fusion factors, some morphed facial images which can successfully attack Face++ are randomly selected (considering the balance of the sample sizes between the real face images and morphed face images). Finally, 1378 morphed facial images are collected into the FM-database. Among them, 584 morphed facial images are used for restoration testing. The detail of FM-database is listed in Table 5. FM-database is subdivided into disjoint training set, development set, and testing set. The training set is used to train the network model, and the development set is used to adjust some super-parameters and prevent under-fitting or over-fitting of training models. The testing set is used for the final test (The following experimental results were performed on the testing set.)

B. IMPLEMENTATION DETAILS

In order to make the training phase effective, for all input images, facial detection toolkit named as Dlib programming library [28] is implemented for face detection and alignment according to 5 facial landmarks models provided

TABLE 4. IAPMR of morphed facial image on face++.

Pixel fusion factor	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
IAPMP	8.49%	31.01%	69.36%	97.29%	99.87%	93.82%	61.13%	27.02%	7.72%

TABLE 5. The summary of the FM-database.

Image size	Subset	#Subject (Male, Female)	# Real facial image	# Morphed facial image
360 × 480	Training set	28 (16, 2)	613	632
	Development set	7 (4, 3)	138	101
	Testing set	28 (16, 12)	627	584

by the library. Each aligned facial image is resized to $128 \times 128 \times 3$ pixels, and the restored facial images from the generator are of the same size.

The FD-GAN is implemented using the deep learning toolbox Tensorflow 1.11.0 [30]. For the loss weights, they are empirically set as $\lambda_1 = 10$, $\lambda_2 = 1$ and $\beta_1 = 1$.

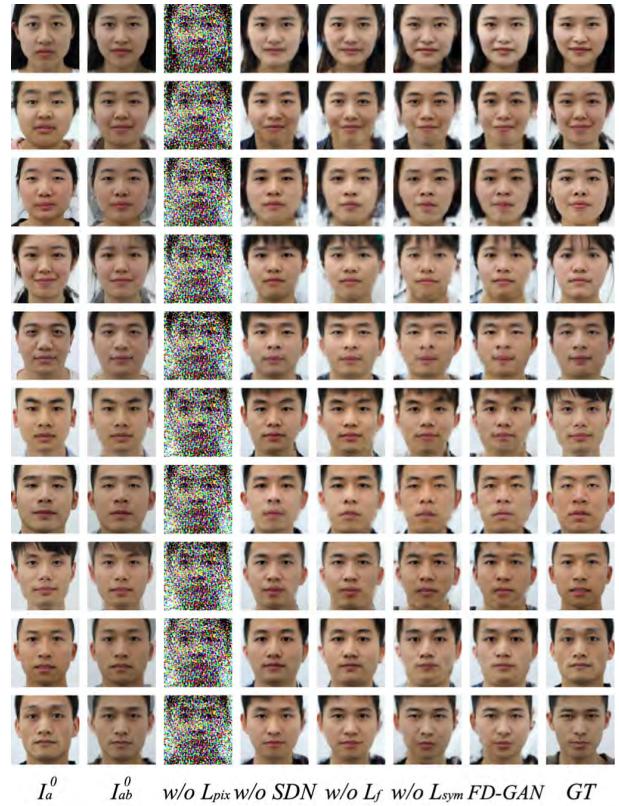
C. EXPERIMENTAL RESULTS

In the experiments, two scenarios are considered. In the first scenario, in order to verify the feasibility of detangling the morphing contributor's identity, we simplify the identity feature of facial image by ignoring the expression, occlusion and other changes of the subjects. For the sake of simplicity, the real facial images used to generate the morphed facial images are utilized as the auxiliary image of FD-GAN for training and testing. The second scenario simulates the real restoration situation. The auxiliary facial images of FD-GAN may contain various expressions, makeup and occlusion during training and testing.

1) ANALYSIS OF THE PROPOSED FRAMEWORK

For the first scenario, an ablation study is performed. Different loss combinations and without use of symmetric dual network are done to the FD-GAN to evaluate its performance. Here, four variations of FD-GAN are compared, and they are 1) removing the loss L_{pix} (denoted as w/o L_{pix}); 2) removing the loss L_f (denoted as w/o L_f); 3) removing the loss L_{sym} (denoted as w/o L_{sym}); and 4) training without symmetric dual network architecture (denoted as w/o SDN). The visualization samples of each variant are shown in Figure 4. Where GT represents the ground truth of the accomplice I_b^0 , and I_b^0 , I_{ab}^0 represent the criminal facial image and morphed facial image, respectively. As expected, FD-GAN including all losses achieves the best visual effects. Without the loss L_{pix} , it is almost impossible to generate the real facial images. Without symmetric dual network architecture, the restored facial images are obviously different from the ground truth. After removing the loss L_f , the restored facial images are also different from the ground truth, but there is a little improvement comparing with those without symmetrical structure network. When the loss L_{sym} is removed, the facial asymmetry is apparent in nostrils and eyes' size.

To quantitatively evaluate the performance of FD-GAN, Face++ is used to compare the restored facial image I_b^1 with

**FIGURE 4.** Restoration results of FD-GAN and its variants.**TABLE 6.** The restoration accuracy with ablation in scenario 1.

Method	w/o L_{pix}	w/o SDN	w/o L_f	w/o L_{sym}	FD-GAN
Accuracy	-	45.55%	49.32%	79.11%	85.97%

I_b^0 and I_a^0 , respectively. When the system determines that I_b^0 matches I_b^0 but it does not match I_a^0 , the restoration is regarded as successful. The restoration accuracy is defined as

$$Accuracy = \frac{N}{T}, \quad (18)$$

where N represents the number of successfully restored facial images, and T represents the total number of restored facial images.

The restoration accuracy of scenario 1 is listed in Table 6. For the second scenario, the same hyper-parameters (loss weight and learning rate etc.) of the first scenario are used for simplicity. Some synthetic samples are illustrated in Figure 6, and the restoration accuracy is listed in Table 7.

Comparing the results listed in Table 6 and Table 7, the restoration accuracy of the second scenario is much lower than α that of the first situation. This is mainly because the facial images captured by FRS in the second situation

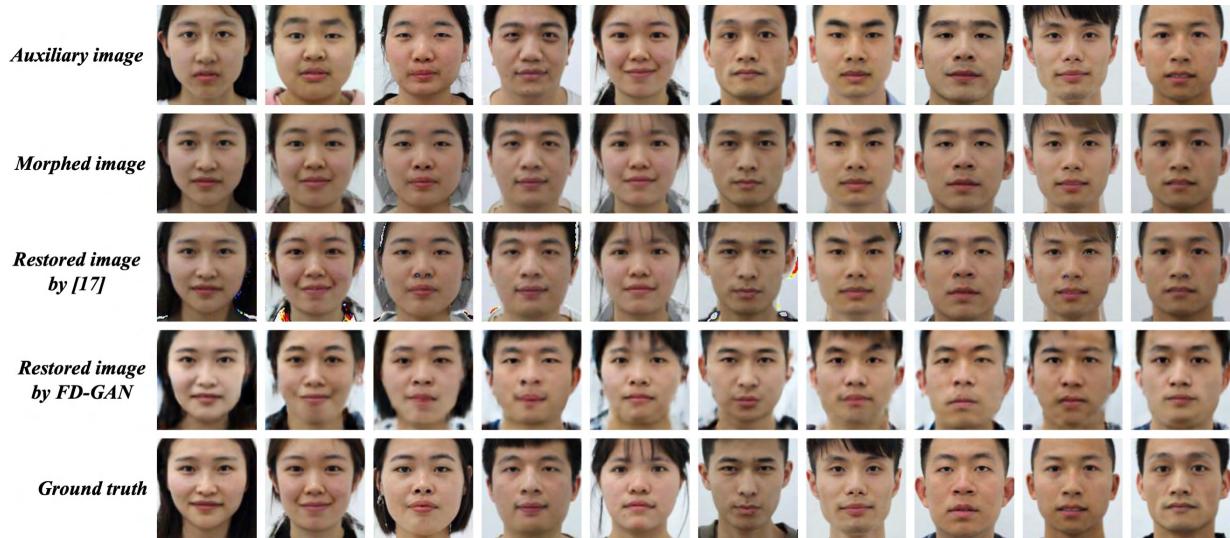


FIGURE 5. Example of restoration comparison in scenario 1.

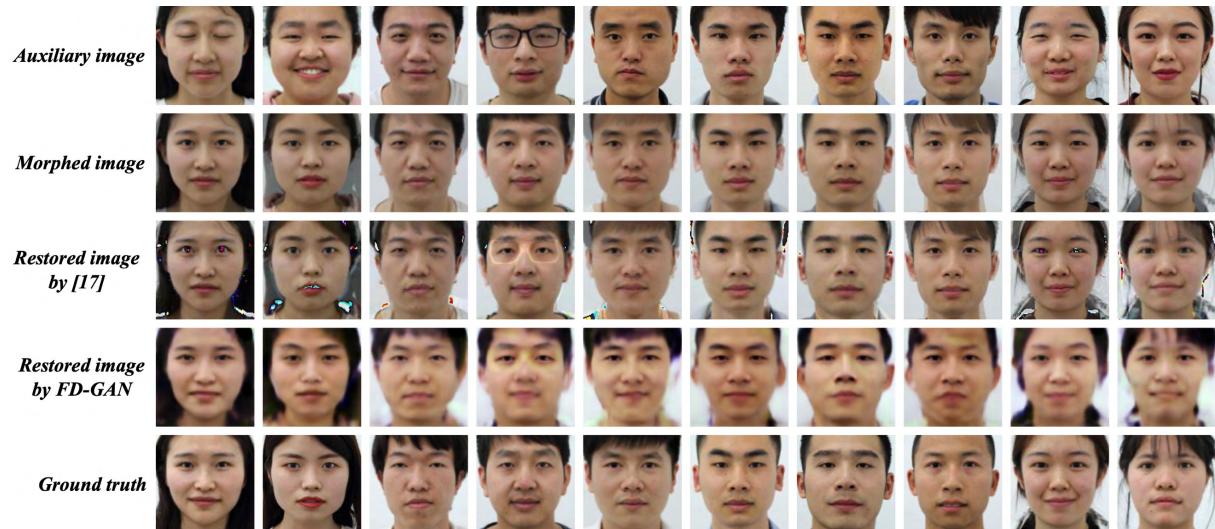


FIGURE 6. Example of restoration comparison in scenario 2 (auxiliary images contain expression, make-up and occlusion changes).

TABLE 7. The restoration accuracy with ablation in scenario 2.

Method	w/o L_{pix}	w/o SDN	w/o L_f	w/o L_{sym}	FD-GAN
Accuracy	-	33.56%	49.82%	61.13%	64.90%

have more variants in expression, make-up, occlusion, etc. Therefore, it is more difficult to restore the face morphing accomplice.

Furthermore, to verify the validity of restored accomplice's facial image, the average FRS scores between morphed facial image - restored accomplice's facial image, and between ground truth image (accomplice) - restored accomplice's facial image, grouped by the pixel fusion factor α used for face morphing, are listed in Table 8. As seen from Table 8, it shows that the pixel fusion factor has a certain impact on the restoration effect, but the average FRS scores of the ground truth-restored pair is higher than that

of the morphed-restored pair. It indicates that the restored accomplice's facial image is closer to the ground truth image (accomplice) for FRS, and it also indicates the effectiveness of FD-GAN in restoring accomplice's facial image from the morphed facial image.

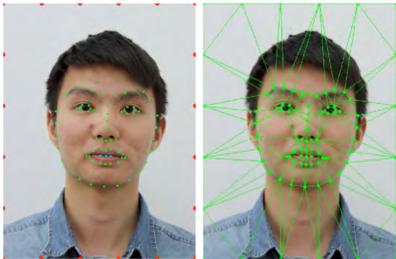
2) PERFORMANCE COMPARISON

In order to evaluate the effectiveness of the proposed scheme, the proposed FD-GAN is also compared with the latest face de-morphing method [17] in two scenarios (The de-morphing parameter was selected as 0.3, which is within the range recommended in [17]). The restoration accuracy is listed in Table 9, and some synthetic samples are also illustrated in Figure 5 and Figure 6.

From Table 9, it can be found that the accuracy of restoring accomplice's facial image of the proposed method outperforms that of the method in [17]. In scenario 1,

TABLE 8. Average score of different face pairs on Face++.

Score \ α	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Morphed - Restored	62.36	63.57	62.45	62.28	63.89	66.66	67.54	70.63	73.90
Ground truth - Restored	62.37	63.94	67.48	70.58	74.96	75.73	75.71	76.55	78.08

**FIGURE 7.** Example of facial landmarks and corresponding Delaunay triangles (green and blue landmarks are localized by Dlib, blue landmarks are removed, and red points are additional landmarks (red)).**TABLE 9.** The restoration accuracy comparison in two scenarios.

Method	Scenario 1	Scenario 2
Face de-morphing [17]	49.82%	46.91%
FD-GAN (Proposed)	85.97%	64.90%

FD-GAN achieve the best restored accuracy 85.97%, while the method [17] is only 49.82%. As seen from Figure 5 and Figure 6, the restored facial images with the method in [17] have different degrees of restoration defects, such as abnormal color spots (e.g. the second and the sixth examples in Figure 5), edge artifacts, (e.g. the first and the fourth examples in Figure 5), ghosting artifacts (e.g. the third example in Figure 5), etc. The most obvious is that in the last four examples in Figure 5. The restored facial images are almost the same as the morphed facial images, and the facial images of the accomplices are not effectively recovered. The main reason is that the method in [17] depends on the prior knowledge of the generation of the morphed face, such as the morphing process and the morphing parameters. This makes the face restoration effect affected by the correlation between the morphing parameter and the de-morphing parameter. In this experiment, the morphing parameter varies from 0.1 to 0.9, which is used in morphed facial image generation. However, the de-morphing parameter is fixed, and this leads to limited restoration performance. It also indicates that the method in [17] is sensitive to the morphing parameter and limited performance for accomplice's facial restoration. In contrast, the proposed FD-GAN doesn't rely on prior knowledge of the generation of the morphed face (e.g. face morphing process and morphing parameter), and it realizes the accomplice's facial restoration by separating the identity of the accomplice hidden in morphed facial images.

The experimental results show the superiority of the proposed FD-GAN in restoring the face morphing attack accomplice's facial image from a morphed facial image. It verifies that it is feasible to restore the morphing

accomplice's facial image by separating the hidden identity of face morphing contributors.

V. CONCLUSION

In this paper, a FD-GAN is proposed to restore the face morphing accomplice's facial image. To the best of our knowledge, it is the first work to restore the face morphing accomplice's facial image using learning-based generation approach. The symmetric dual network is specially designed to effectively disentangle the identity features of the morphing contributors. Experimental results demonstrate that the proposed FD-GAN can achieve fairly good restoration performance. It has great potential to investigate the identity of face morphing attack's accomplice in criminal investigation and judicial forensics, which can help the police lock the identity of criminal's accomplice, and crack down on the crime of identity theft (face morphing attack). The research on the facial restoration of morphing accomplice from a morphed facial image is an open question, and many problems remain to be solved. Our future work will be concentrated on researching the effects of proposed FD-GAN on a larger dataset, which contains more subjects, printing/scanning images, other types of morphing faces (e.g. non-ghosting artifacts, non-Asian subjects, etc.), and improving the morphing contributor's facial restoration accuracy involving expression, posture, make-up and occlusion variants (e.g. with glasses, beard, mask, etc.).

APPENDIX

AUTOMATIC FACE MORPHING PROCESS

According to the workflow of [9], the general steps for automatic generation of the morphed facial image are as follows.

Step 1: Locate the key points of the morphed facial image. Given two facial images I_1 and I_2 , K_1 and K_2 are the locations of the corresponding facial landmarks. Then the location K_M of the key points of the morphed facial image can be calculated as

$$K_M = (1 - \beta)K_1 + \beta K_2, \quad (19)$$

where β represents a location fusion factor.

To avoid the failure of morphing face generation, 85 landmarks are used for generating morphed facial image, as shown in Figure 7. Among them, 68 facial landmarks are localized by using Dlib programming library [28]. However, three facial landmarks depicting the lower contour of the upper lip are omitted, because they may result in morphing errors. The other 20 landmarks added on the image borders are used for triangulation in the outer region of the convex hull, which is composed by 65 facial landmarks. With the help of facial landmark points of both contributing images,

a morphed facial image using triangulation based tight morphing is generated.

Step 2: Warping. Triangular mesh T_M can be derived from K_M via Delaunay triangulation. Similarly, two triangular mesh T_1 and T_2 can also be derived from K_1 and K_2 , respectively. For each triangle in T_1 , all pixels inside the triangle are transformed to the corresponding triangle in T_M by affine transform, and then the warped image I'_1 from I_1 is obtained. Similarly, the warped image I'_2 from I_2 can be also obtained.

Step 3: Blending. A morphed facial image I_M can be obtained as

$$I_M = (1 - \alpha) I'_1 + \alpha I'_2, \quad (20)$$

where α represents a pixel fusion factor.

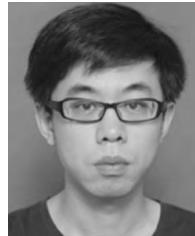
In the experiments, the pixel fusion factor α is varied from 0.1 to 0.9, and the location fusion factor β is fixed as 0.5.

REFERENCES

- [1] *9303-Machine Readable Travel Documents—Part 9: Deployment of Biometric Identification and Electronic Storage of Data in eMRTDs*, Int. Civil Aviation Org., Montreal, QC, Canada, 2015.
- [2] M. Ferrara, A. Franco, and D. Maltoni, “The magic passport,” in *Proc. IJCB*, Sep. 2014, pp. 1–7.
- [3] D. J. Robertson, R. S. Kramer, and A. M. Burton, “Fraudulent ID using face morphs: Experiments on human and automatic recognition,” *PLoS ONE*, vol. 12, no. 3, 2017, Art. no. e0173319.
- [4] M. Gomez-Barrero, C. Rathgeb, U. Scherhag, and C. Busch, “Is your biometric system robust to morphing attacks?” in *Proc. IWBF*, Apr. 2017, pp. 1–6.
- [5] U. Scherhag, R. Raghavendra, K. B. Raja, M. Gomez-Barrero, C. Rathgeb, and C. Busch, “On the vulnerability of face recognition systems towards morphed face attacks,” in *Proc. IWBF*, Apr. 2017, pp. 1–6.
- [6] U. Scherhag, A. Nautsch, C. Rathgeb, M. Gomez-Barrero, R. N. J. Veldhuis, L. Spreeuwerts, 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 *Proc. BIOSIG*, Sep. 2017, pp. 1–7.
- [7] L. Wandzik, R. V. Garcia, G. Kaeding, and X. Chen, “CNNs under attack: On the vulnerability of deep neural networks based face recognition to image morphing,” in *Proc. IWDW*, 2017, pp. 121–135.
- [8] R. Raghavendra, K. B. Raja, and C. Busch, “Detecting morphed face images,” in *Proc. BTAS*, Sep. 2016, pp. 1–7.
- [9] A. Makrushin, T. Neubert, and J. Dittmann, “Automatic generation and detection of visually faultless facial morphs,” in *Proc. VISIGRAPP*, 2017, pp. 39–50.
- [10] 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 *Proc. IWBF*, Apr. 2017, pp. 1–6.
- [11] 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 *Proc. IHMMSec*, 2017, pp. 21–32.
- [12] T. Neubert, “Face morphing detection: An approach based on image degradation analysis,” in *Proc. IWDW*, 2017, pp. 93–106.
- [13] R. Raghavendra, K. B. Raja, S. Venkatesh, and C. Busch, “Transferable deep-CNN features for detecting digital and print-scanned morphed face images,” in *Proc. CVPRW*, Jul. 2017, pp. 1822–1830.
- [14] C. Seibold, W. Samek, A. Hilsmann, and P. Eisert, “Detection of face morphing attacks by deep learning,” in *Proc. IWDW*, 2017, pp. 107–120.
- [15] L. Debiasi, U. Scherhag, C. Rathgeb, A. Uhl, and C. Busch, “PRNU-based detection of morphed face images,” in *Proc. IWBF*, Jun. 2018, pp. 1–7.
- [16] L.-B. Zhang, F. Peng, and M. Long, “Face morphing detection using Fourier spectrum of sensor pattern noise,” in *Proc. ICME*, Jul. 2018, pp. 1–6.
- [17] M. Ferrara, A. Franco, and D. Maltoni, “Face demorphing,” *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 4, pp. 1008–1017, Nov. 2018.
- [18] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in *Proc. NIPS*, 2014, pp. 2672–2680.
- [19] U. Scherhag, C. Rathgeb, J. Merkle, R. Breithaupt, and C. Busch, “Face recognition systems under morphing attacks: A survey,” *IEEE Access*, vol. 7, pp. 23012–23026, 2019.
- [20] R. Huang, S. Zhang, T. Li, and R. He, “Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis,” in *Proc. ICCV*, Oct. 2017, pp. 2439–2448.
- [21] Y. Li, L. Song, X. Wu, R. He, and T. Tan, “Anti-Makeup: Learning a bi-level adversarial network for makeup-invariant face verification,” in *Proc. AAAI*, 2018, pp. 1–8.
- [22] J. Bao, D. Chen, F. Wen, H. Li, and G. Hua, “Towards open-set identity preserving face synthesis,” in *Proc. CVPR*, Jun. 2018, pp. 6713–6722.
- [23] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proc. ICCV*, Oct. 2017, pp. 2223–2232.
- [24] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. CVPR*, Jun. 2016, pp. 770–778.
- [25] G.-J. Qi, “Loss-sensitive generative adversarial networks on Lipschitz densities,” *arXiv:1701.06264*. [Online]. Available: <https://arxiv.org/abs/1701.06264>
- [26] *Face++ Compare API*. Accessed: Nov. 12, 2018. [Online]. Available: <https://www.faceplusplus.com/face-comparing/>
- [27] *Information Technology—Biometric Presentation Attack Detection—Part 3: Testing and Reporting*, document ISO/IEC JTC1 SC37 Biometrics, ISO/IEC FDIS 30107-3:2017, Geneva, Switzerland, International Organization for Standardization, 2017.
- [28] *Dlib C++ Library*. Accessed: Nov. 12, 2018. [Online]. Available: <http://dlib.net/>
- [29] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” in *Proc. ICLR*, 2016, pp. 1–16.
- [30] *Tensorflow*. Accessed: Nov. 12, 2018. [Online]. Available: <https://tensorflow.google.cn/>



FEI PENG received the Ph.D. degree in circuits and systems from the South China University of Science and Technology, Guangzhou, China, in 2006. He was a Visiting Fellow with the Department of Computer Science, University of Warwick, U.K., from 2009 to 2010. He was a Visiting Professor with the SeSaMe Centre, College of Computing, National University of Singapore, in 2016. He is currently a Professor with the College of Computer Science and Electronic Engineering, Hunan University, Changsha. He is also a Visiting Professor with the Dongguan University of Technology, Dongguan. His research interests include digital watermarking and digital forensics.



LE-BING ZHANG received the B.S. and M.S. degrees from Hunan Normal University, in 2003 and 2009, respectively. He is currently pursuing the Ph.D. degree with the College of Computer Science and Electronic Engineering, Hunan University, Changsha. He is also an Assistant Professor with Huaihua University, Huaihua, China. His research interests include digital forensics, multimedia security, and deep learning.



MIN LONG received the Ph.D. degree in circuits and systems from the South China University of Science and Technology, Guangzhou, China, in 2006. She was a Visiting Fellow with the Department of Computer Science, University of Warwick, U.K., from 2009 to 2010. She is currently a Professor with the College of Computer and Communication, Changsha University of Science and Technology. Her research interests include digital watermarking and chaos-based secure communication.