Target identity attacks on facial recognition systems

Saheb Chhabra, Naman Banati, Gaurav Gupta, Garima Gupta

Chapter 13

TARGET IDENTITY ATTACKS ON FACIAL RECOGNITION SYSTEMS

Saheb Chhabra, Naman Banati, Gaurav Gupta and Garima Gupta

Abstract Advancements in digital technology have significantly increased the number of cases involving the counterfeiting of identity documents. One example is exam fraud, where a counterfeiter creates a composite morphed photograph of the real candidate and an imposter, and attaches it to the examination admit card. Automated facial recognition systems are beginning to be deployed at examination centers to match candidates' faces against their official facial images. While the need to perform manual matches is eliminated, the vulnerabilities of these automated systems are a major concern.

This chapter evaluates the vulnerability of an automated facial recognition system to input image manipulation via a target identity attack. The attack manipulates a facial image so that it looks similar to the real candidate, but outputs the identity feature representation of the imposter. This chapter also evaluates the performance of facial recognition models with regard to impersonator recognition. Experiments using image databases demonstrate the effectiveness of target identity attacks.

Keywords: Counterfeiting, facial recognition, target identity attacks

1. Introduction

Counterfeiting of identity documents is one of the fastest-growing frauds worldwide. Advancements in digital technology enable counterfeiters to perpetrate sophisticated frauds by creating fake identity cards and other related documents. An example is exam fraud, where a counterfeiter morphs the photograph of a real candidate with that of an imposter, and creates a tampered examination admit card with the morphed photograph. The quality of the counterfeit admit card makes it very difficult for a human examiner to determine that the person who

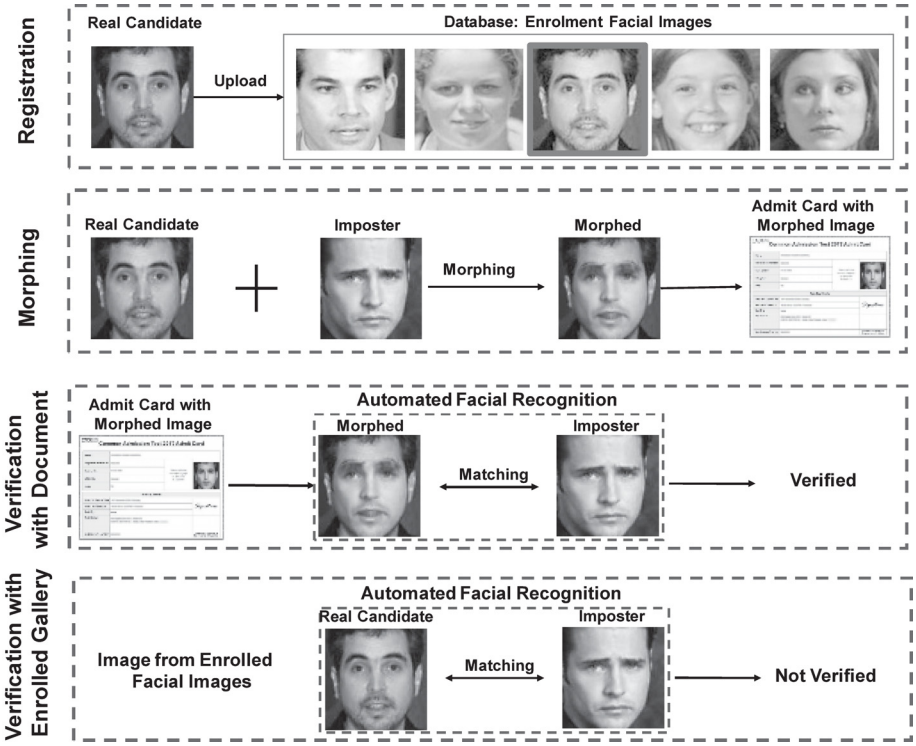


Figure 1. Identifying an imposter.

presents the card at an examination center is an imposter. The imposter then proceeds to take the exam on behalf of the real candidate.

Several instances of exam fraud involving the manipulation of facial images in admit cards have been reported by the international media. In 2018, law enforcement officers in Jodhpur, India arrested several members of a gang involved in a police examination cheating case [21]. The gang employed a team of 20 subject matter experts to take entry examinations on behalf of candidates who paid large sums of money. In another case [23], fifteen Chinese nationals were arrested for using counterfeit Chinese passports to take U.S. college entry tests such as the SAT, GRE and TOEFL.

Automated facial recognition systems are beginning to be deployed at examination centers to match candidates' faces against their official facial images. These systems perform a two-step verification procedure. The first step matches the facial image provided at the time of exam registration against the facial image of the supposed candidate who shows up at the examination center. The second step matches the photograph

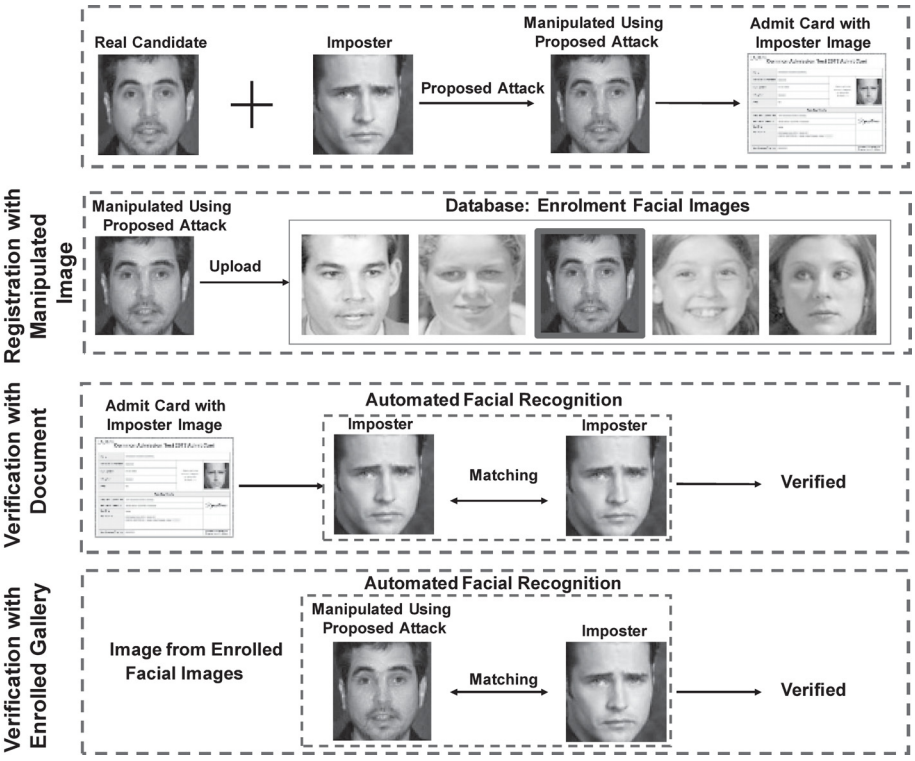


Figure 2. Successful target identity attack.

on the admit card against the facial image of the candidate that was provided at the time of exam registration.

Figure 1 shows how an automated facial recognition system identifies an imposter by comparing the photograph on the presented admit card against the facial image of the candidate that was provided at the time of exam registration.

The security of automated facial recognition systems is a major concern. Therefore, this research focuses on the vulnerability of an automated system to image manipulation using a novel target identity attack. This attack introduces perturbations in the facial image of the real candidate such that the manipulated image appears to be of the real candidate, but it outputs the identity features of an imposter (target). This manipulated image is submitted at the time of candidate registration. Thus, the imposter is able to masquerade as the real candidate and take the examination on his or her behalf. Figure 2 demonstrates a successful target identity attack.

This chapter focuses on the vulnerability of automated facial recognition systems to target identity attacks. It also evaluates the performance of facial recognition models with regard to impersonator recognition. Experiments using image databases demonstrate the effectiveness of target identity attacks.

2. Related Work

Several researchers have proposed morphing techniques that target facial recognition and other biometric systems [18]. Korshunova et al. [7] have proposed a morphing technique similar to style-transfer-based deep neural networks using a novel loss function. Othman and Ross [14] have used a morphing method to fool gender classifiers. The method morphs an input image with an image of a subject of the opposite gender; two parameters are used to control the appearance of the final image with gender suppression information. Mirjalili et al. [12] have extended the work of Othman and Ross so that the input candidate image is fused with another candidate image selected based on a correlation between facial landmark points. Delaunay triangulation is employed to identify the pixels to be modified.

Damer et al. [3] have proposed a generative adversarial network (GAN) method that employs representation loss to create morphed images. Ferrara et al. [5] have proposed a method that creates double identity fingerprints; the features of two fingers are combined to yield a new fingerprint that fools fingerprint recognition systems. Rathgeb and Busch [16] have developed a stability-based bit substitution method that morphs two iris codes.

Several researchers have proposed methods for detecting morphed images. Raghavendra et al. [15] have proposed an approach that employs fine-tuned deep convolutional neural networks and a probabilistic collaborative representation classifier (P-CRC). The approach extracts features from fully-connected layers of VGG19 [20] and AlexNet [8] models, and concatenates them before sending them to the classifier. Seibold et al. [19] compare the morphing detection performance of pre-trained and trained-from-scratch AlexNet, GoogLeNet and VGG19 models. They created a database using the triangle and mesh warping techniques.

Wandzik et al. [22] have proposed a technique for distinguishing between original and morphed images. The technique employs features extracted from four facial recognition models and classifies them using a support vector machine (linear classifier). Batskos et al. [1] have developed a distance-based approach for detecting morphed images. Euclidean distances are computed for probe and morphed, morphed and

e-pass, and probe and e-pass facial features, which yield 3D vectors. A support vector machine is used for linear classification.

Zhang et al. [24] have proposed a source identification scheme for detecting whether an image is *bona fide* or morphed. Scherhag et al. [17] have developed a technique for distinguishing morphed images using facial landmark points; the technique assumes that the intra-subject variance of landmarks extracted from *bona fide* images is less than the variance between landmarks of the morphed image and its contributing subjects. Debiasi et al. [4] have proposed a photo response non uniformity (PRNU) approach for detecting morphed images. Their approach assumes that the variance of PRNU signals increases across image cells when two images are morphed to create a single image.

Damer et al. [2] have employed two scenarios for morphing detection (i.e., with and without a probe). In the first scenario, facial landmarks are determined using an ensemble of regression trees, explicit shape regression and regressing local binary features; Euclidean distances are computed between the landmarks of live and previously-submitted images. In the second scenario, facial features are extracted using local binary pattern histograms and transferable deep convolutional neural networks.

Makrushin et al. [11] have proposed a morph detection algorithm, which assumes that some blocks in morphed images have undergone JPEG compression; however, blocks are not compressed in newly morphed images. The algorithm performs JPEG compression on a morphed image and determines nine Benford features that are used for classification. Finally, Neubert et al. [13] have employed frequency and spatial domain features for morph detection.

The review of the literature reveals that morphing methods have limitations that introduce ghosting artifacts and change the visual appearances. Figure 3 shows the images obtained using four morphing methods (from left to right): (i) Neubert et al. [13]; (ii) Batskos et al. [1]; (iii) Scherhag et al. [17]; and (iv) Damer et al. [2]. Note that the first and second rows show the original images whereas the third row shows the morphed composite images.

3. Target Identity Attacks

A target identity attack introduces adversarial perturbations in the facial image of the real candidate (source image) such that the manipulated image appears to be similar to the real candidate, but it outputs the identity features of the facial image of the imposter (target image). The manipulated image is submitted at the time of candidate registra-

Figure 3. Morphed images generated using four methods.

tion. Thus, the imposter can fool an automated facial recognition system and masquerade as the real candidate.

Figure 4 shows a block diagram of a target identity attack. The adversarial perturbation **N** is initialized as zero in the first iteration. In subsequent iterations, the perturbed image obtained by adding perturbations **N** to the source image **S**, and the target image **T** is provided an input to the facial recognition model. Optimization is performed over perturbations **N** until the stopping criterion is satisfied.

The detailed steps involved in the target identity attack are discussed below. In the following, the terms real candidate image and source candidate image are used interchangeably. Likewise, the imposter candidate image and target candidate image are used interchangeably.

The fundamental problem is to manipulate a source candidate image so that the manipulated image outputs the identity features of the target candidate image while appearing to be similar to the source candidate image.

Let **S** be the source candidate image and **T** be the target candidate image in the range [0, 1]. Let **P** be the manipulated or perturbed image generated by adding perturbation **N** to the source candidate image **S**. In order for the perturbed image to be valid, it should be in the range [0, 1]. Mathematically, this is written as:

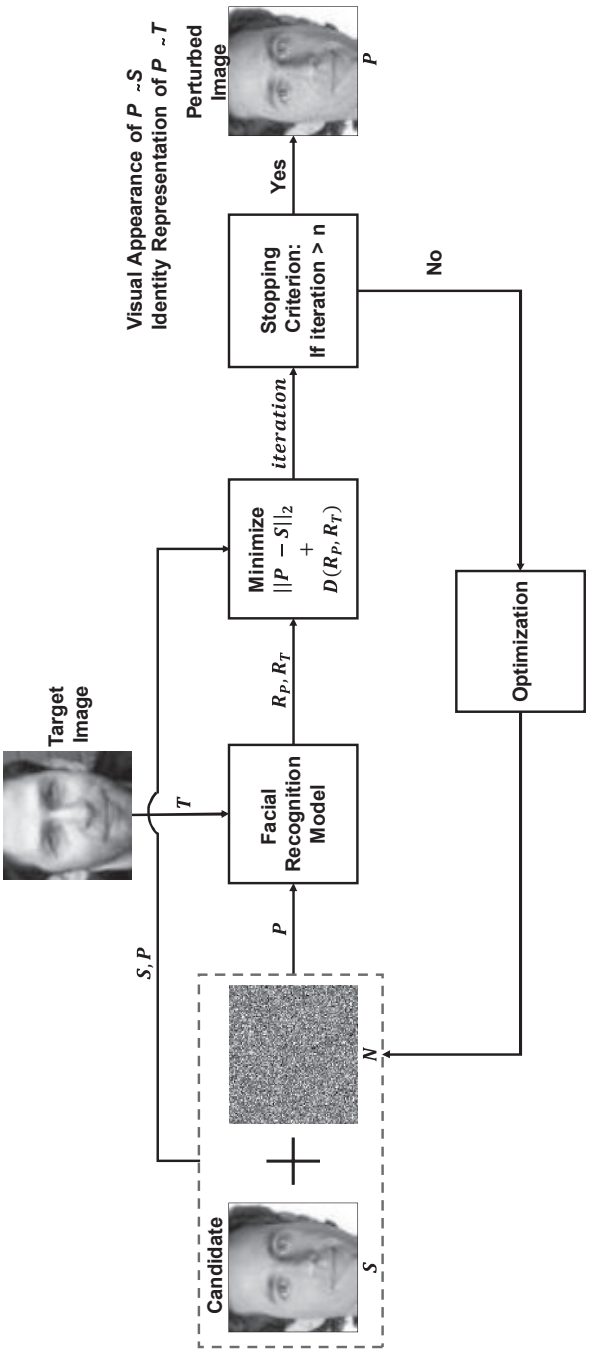


Figure 4. Target identity attack.

$$\mathbf{P} = \mathbf{S} + \mathbf{N} \tag{1}$$

where $\mathbf{P} \in [0, 1]$.

To satisfy this constraint, the following transformation is used to generate a perturbed image $\mathbf{P}$ in the range [0, 1]:

$$\mathbf{P} = \frac{1}{2}(\tanh(\mathbf{S} + \mathbf{N}) + 1) \tag{2}$$

Let V and I denote the visual appearance and identity of an image, respectively. The goal is to generate a perturbed image $\mathbf{P}$ such that its visual appearance is similar to the source candidate image $\mathbf{S}$ and its identity representation is similar to the target candidate $\mathbf{T}$. This is mathematically formulated as:

$$V_{\mathbf{P}} = V_{\mathbf{S}}; \quad I_{\mathbf{P}} = I_{\mathbf{T}} \tag{3}$$

where $V_{\mathbf{P}}$ and $V_{\mathbf{S}}$ are the visual appearances of the perturbed image $\mathbf{P}$ and source candidate image $\mathbf{S}$, respectively; and $I_{\mathbf{P}}$ and $I_{\mathbf{T}}$ are the identities of the perturbed image $\mathbf{P}$ and source candidate image $\mathbf{S}$, respectively.

Two loss functions $f(\mathbf{P})_V$ and $f(\mathbf{P})_I$ are specified to incorporate the constraints mentioned above in the attack. The first loss function $f(\mathbf{P})_V$ deals with the visual appearance of the perturbed image and the second loss function $f(\mathbf{P})_I$ deals with the identity representation of the perturbed image. Both the functions have to be minimized as follows:

$$\text{Min}\{f(\mathbf{P})_V + f(\mathbf{P})_I\} \tag{4}$$

In order to make the visual appearance of the perturbed image $\mathbf{P}$ similar to the source candidate image $\mathbf{S}$, the distance between $\mathbf{S}$ and $\mathbf{P}$ must be minimized. Thus, the function $f(\mathbf{P})_V$ is written as:

$$f(\mathbf{P})_V = D(\mathbf{P}, \mathbf{S}) \tag{5}$$

where D is the distance metric.

Since the Euclidean distance is used as the metric, Equation (5) is written as:

$$f(\mathbf{P})_V = ||\mathbf{P} - \mathbf{S}||_2 \tag{6}$$

The next task is to make the identity representation of the perturbed image P similar to the identity representation of the target image $\mathbf{T}$.

Let ϕ be a pre-trained facial recognition model with weights $\mathbf{W}$ and bias b. This model takes an image as input and outputs its identity representation. Therefore, the identity representation $\mathbf{R_T}$ of an input target candidate image $\mathbf{T}$ is computed as:

$$\mathbf{R_T} = \phi(\mathbf{WT} + b) \tag{7}$$

The corresponding identity representation $\mathbf{R_P}$ of the perturbed image $\mathbf{P}$ is computed as:

$$\mathbf{R_P} = \phi(\mathbf{WP} + b) \tag{8}$$

The next task is make the identity representation of the perturbed image $\mathbf{P}$ similar to the identity representation of the target candidate image $\mathbf{T}$. The distance between the identity representation of the perturbed image $\mathbf{P}$ and the target candidate image $\mathbf{T}$ is minimized. Thus, the function $f(\mathbf{P})_I$ is written as:

$$f(\mathbf{P})_I = D(\mathbf{R_P}, \mathbf{R_T}) \tag{9}$$

The Euclidean and cosine distance metrics are used to minimize the distance between the identity representations of the perturbed image $\mathbf{P}$ and target candidate image $\mathbf{T}$. Thus, the overall objective function is written as:

$$\text{Min}\{||\mathbf{P} - \mathbf{S}||_2 + D(\mathbf{R_P}, \mathbf{R_T})\} \tag{10}$$

which is optimized over the perturbation variable $\mathbf{N}$.

4. Experiments and Results

Experiments were performed to evaluate the effectiveness of target identity attacks. One set of experiments was performed under two scenarios, one involving white-box attacks and the other involving black-box attacks. Another set of experiments, involving impersonator recognition using pre-trained models, was performed to evaluate the performance of pre-trained facial recognition models in impersonator recognition.

The first set of experiments was performed on the Labeled Faces in the Wild (LFW) dataset [6]. The second set of experiments was performed on the Disguised Faces in the Wild (DFW) dataset [9].

The following dataset details and evaluation protocols are pertinent:

- **Labeled Faces in the Wild (LFW) Dataset:** This dataset contains 13,233 facial images of 5,749 subjects. The evaluation of target identity attacks in the white-box and black-box scenarios employed View 2 of the LFW dataset, which comprises 6,000 pairs of images. Of the 6,000 pairs of images, 3,000 pairs are genuine images while the remaining 3,000 pairs are imposter images.

  The target identity attacks were performed by perturbing one image from each imposter pair so that its identity representation becomes similar to the other image in the imposter pair.

Table 1. Summary of experiments.

Experiment	Dataset	Model	Distance Metric
Target Identity Attacks	LFW	VGGFace, ResNet50	Euclidean, Cosine
Impersonator Recognition with Pre-Trained Models	DFW	VGGFace, ResNet50, LCNN-29	Euclidean, Cosine

- **Disguised Faces in the Wild (DFW) Dataset:** This dataset contains 11,157 facial images of 1,000 subjects. Four types of images – normal, validation, disguised and impersonator – are included for each subject.

  The DFW dataset provides three protocols. This research employed Protocol 1 (impersonation) to evaluate facial recognition models. Specifically, Protocol 1 is used to distinguish impersonators from legitimate subjects. In the protocol, the combination of a normal image with a validation image of the same subject corresponds to a genuine pair. The combination of an impersonator image with the normal, validation and disguised images of the same subject corresponds to an imposter pair.

Table 1 provides details about the two sets of experiments. The first set of experiments employed the VGGFace facial recognition model (pre-trained with the VGGFace dataset) and the ResNet50 facial recognition model (pre-trained with the VGGFace2 dataset); target identity attacks on the two facial recognition models were evaluated using the LFW dataset. The second set of experiments employed the VGGFace, ResNet50 and LCNN-29 facial recognition models for impostor recognition; the DFW dataset was used in the evaluation.

4.1 Implementation Details

The experiments were performed on an NVIDIA Tesla P100 server with 96 GB RAM and 16 GB GPU memory. All the images were resized to 224×224 pixels.

The target identity attacks were performed by learning the perturbation corresponding to each image to be attacked. The attacks were implemented in Tensorflow v1.9.0. The learning rate was set to 0.1 during the training phase. The perturbations were adjusted over 15 iterations.

Table 2. Imposter mean distance scores for the white-box and black-box scenarios.

	Euclidean Distance				**Cosine Distance**			
	ResNet50		**VGGFace**		**ResNet50**		**VGGFace**	
	Before	**After**	**Before**	**After**	**Before**	**After**	**Before**	**After**
ResNet50	1.20	0.30	0.91	0.77	0.72	0.08	0.42	0.31
VGGFace	1.20	1.06	0.91	0.30	0.72	0.62	0.42	0.08

4.2 Attack Performance Evaluation

This section discusses the performance of target identity attacks. The Euclidean distance and cosine distance were used as performance metrics.

Target Identity Attacks. This set of experiments evaluated a scenario where perturbed images are presented in place of real candidate images during the exam registration process. The perturbed image would look similar to the real candidate image, but it would output the imposter or target identity representation.

In a real-world scenario, the counterfeiter would not know the facial recognition model that is used to authenticate supposed candidates. Therefore, the target identity attacks are evaluated for white-box and black-box facial recognition scenarios. In a white-box facial recognition scenario, the counterfeiter knows the facial recognition model used to authenticate candidates and generates perturbed images corresponding to the same facial recognition model. In a black-box facial recognition scenario, the counterfeiter does not know the facial recognition model used to authenticate candidates and, therefore, generates perturbed images corresponding to a different facial recognition model.

As mentioned above, 3,000 imposter pairs were considered when implementing the target identity attacks. One image in each pair was perturbed to obtain an identity representation similar to that of the other image (target identity) in the pair. To evaluate the performance, the distance between the target identity representation and the identity representation of the perturbed image was computed for each pair. In the ideal case, this distance should be zero for a successful attack.

Table 2 shows the mean distance scores for 3,000 imposter pairs before and after performing the target identify attacks. Note that good results are obtained for the white-box and black-box scenarios. For example, when perturbed images are generated for the VGGFace model and evaluated using the same model and the cosine distance metric, the mean

distance is reduced by 0.34 ($= 0.42 - 0.08$). Similarly, when perturbed images are generated for the VGGFace model and evaluated using the ResNet50 model and the Euclidean metric, the mean distance score is reduced by 0.14 ($= 1.20 - 1.06$).

Figure 5 compares the imposter distance score distributions for 3,000 imposter pairs before and after target identity attacks in the white-box and black-box scenarios. Figure 5(a) shows the score distributions obtained using the VGGFace and ResNet50 models when the images were perturbed based on the ResNet50 model. Figure 5(b) shows the score distributions obtained using the VGGFace and ResNet50 models when the images were perturbed based on the VGGFace model.

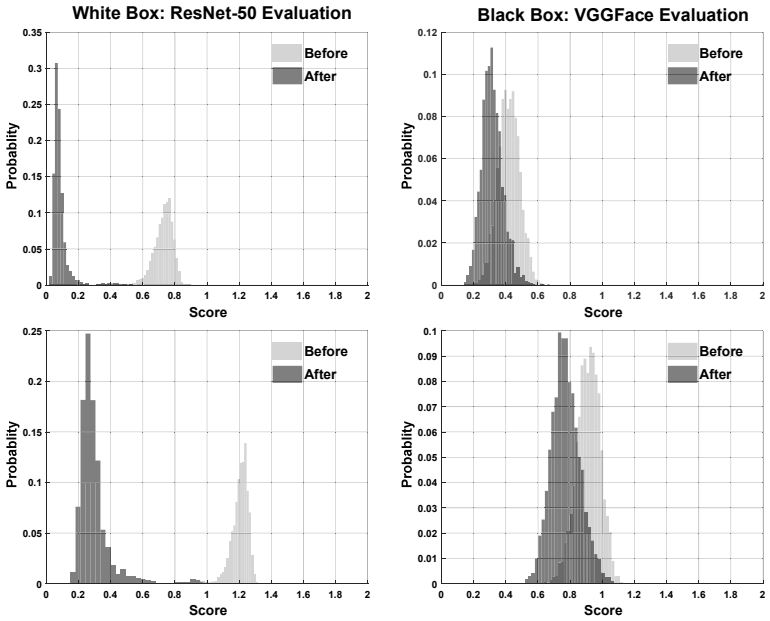
Figure 5(a) and 5(b) consistently show that the distributions are shifted towards the left or to zero after the target identity attacks. This demonstrates that the target identity attacks are effective.

Figures 6 and 7 compare the genuine and imposter distance score distributions before and after identity target attacks in the white-box and black-box scenarios. The images in Figure 6 are perturbed based on the VGGFace model whereas the images in Figure 7 are perturbed based on the ResNet50 model.

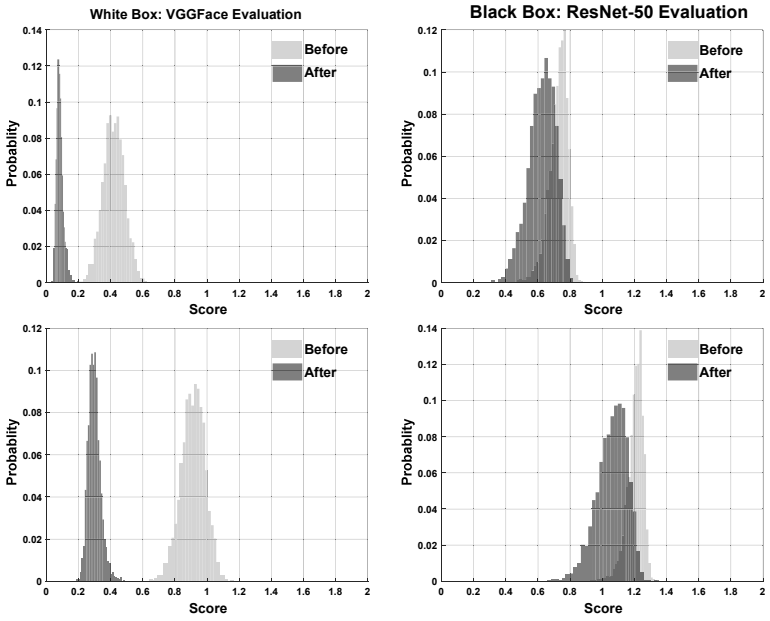
Figures 6 and 7 consistently show that the imposter distance score distributions are shifted closer towards the genuine distance score distributions in the black-box scenario. Moreover, the overlaps between the genuine and imposter distance score distributions are increased. These results demonstrate that the target attacks are effective at fooling the facial recognition models.

Figure 8 shows three sets of images generated via morphing and via target identity attacks. The first and second columns in the figure show the images of two different people, denoted as identity A and identity B, respectively. The third column shows the images generated via morphing whereas the fourth column shows the images generated via target image attacks.

In the case of the morphed images, the visual appearances and the identities correspond to both A and B. However, in the case of the target identity attack images, the visual appearances correspond to A whereas the identities correspond to B. Clearly, the target identity attack images in the fourth column preserve the original A appearances to a greater degree than the morphed images in the third column. However, because their identities correspond to B, the target identity attack images could be used to successfully perpetrate exam fraud.



(a) VGGFace and ResNet50 models with ResNet50 model perturbations.



(b) VGGFace and ResNet50 models with VGGFace model perturbations.

Figure 5. Comparison of impostor distance score distributions.

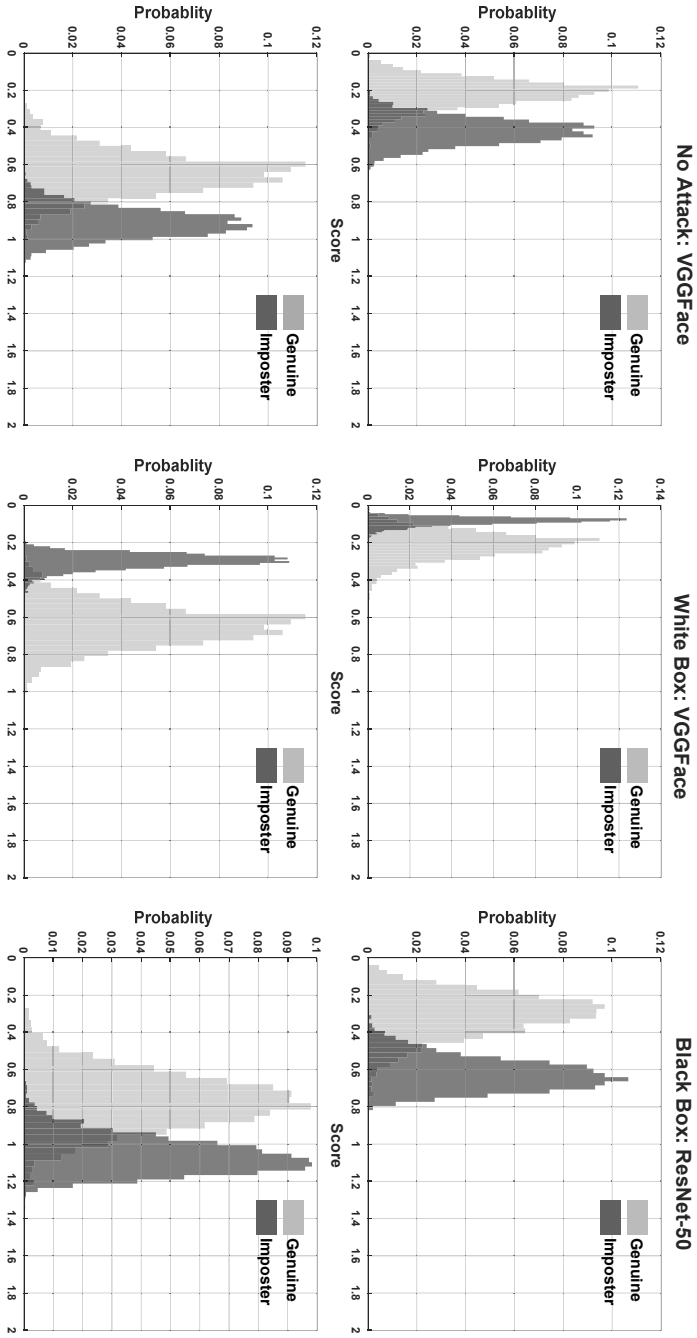


Figure 6. Comparison of genuine and imposter score distributions with VGGFace model perturbations.

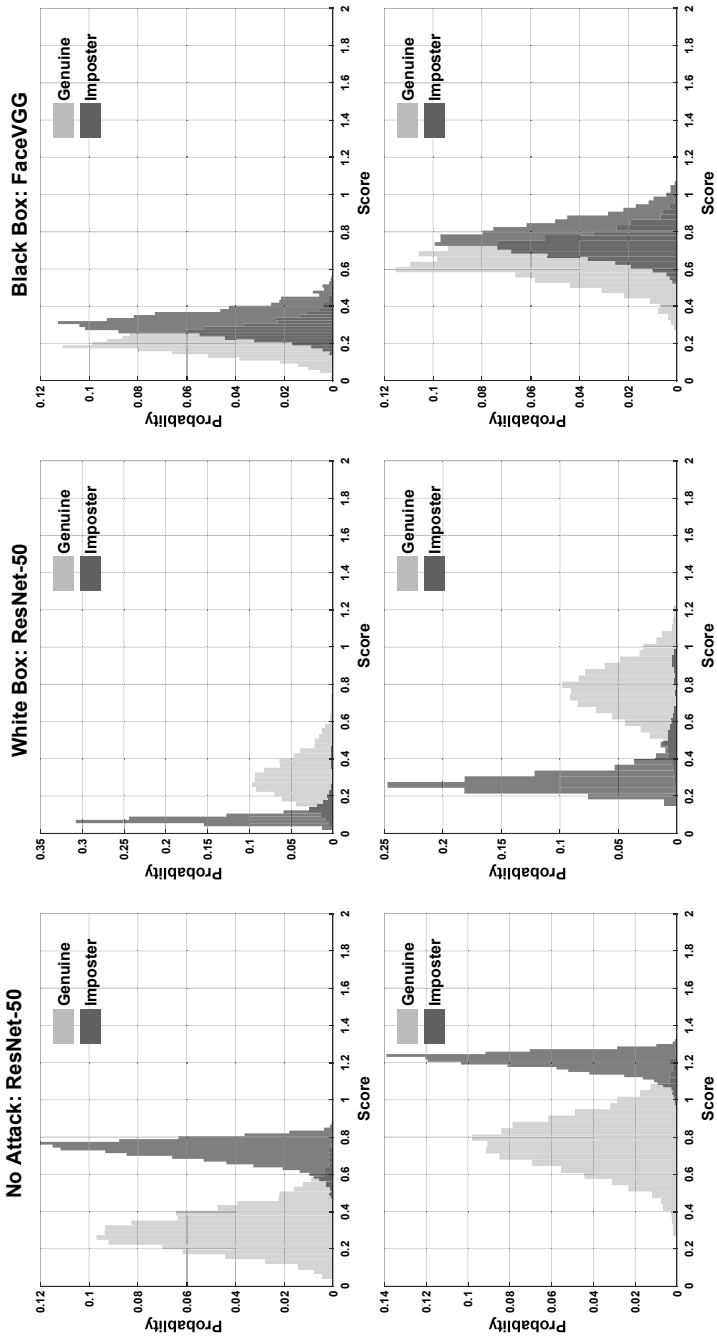
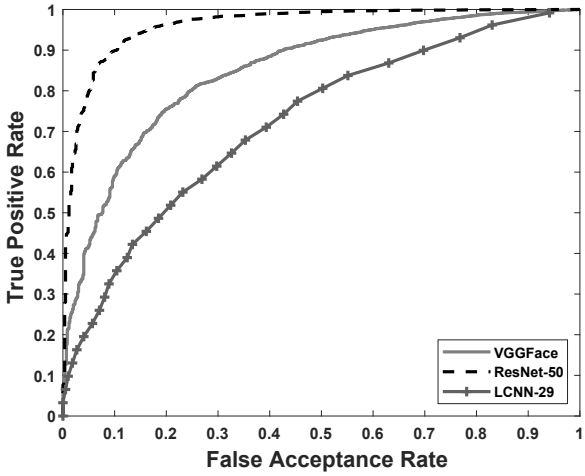


Figure 7. Comparison of genuine and imposter score distributions with ResNet50 model perturbations.

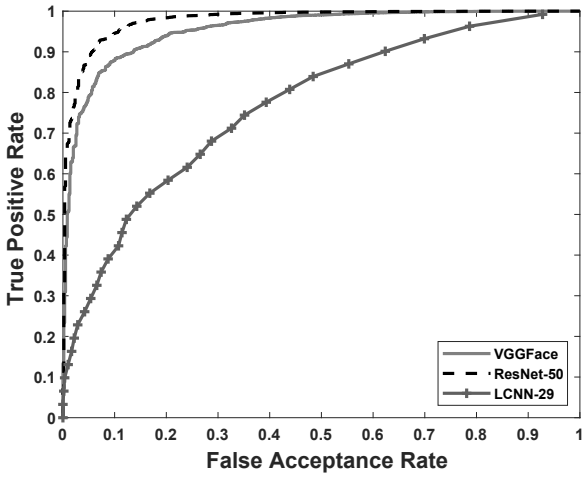
Figure 8. Images generated via morphing and target identity attacks.

Impersonator Recognition with Pre-Trained Models. This set of experiments evaluated a scenario where an impersonator uses a target identity attack image to take an examination on behalf the real candidate. The target identity attack image would be submitted to the authorities when the real candidate registers for the examination. By fooling the automated facial recognition system, the imposter would be able to masquerade as the real candidate and take the examination on behalf of the candidate.

Figure 9 shows the receiver operating characteristic (ROC) curves obtained using pre-trained VGGFace, ResNet50 and LCNN-29 facial recognition models with the Euclidean and cosine distance metrics. The experiments used the DFW dataset and Protocol 1. The ROC curves demonstrate that the three facial recognition models are not effective at identifying impersonators who use images generated by target identity attacks.



(a) ROC plots for Euclidean distance.



(b) ROC plots for cosine distance.

Figure 9. ROC plots for the ResNet50, VGGFace and LCNN-29 models.

5. Conclusions

Advancements in digital technology have significantly increased the number of cases involving the counterfeiting of exam identity documents. As a result, automated facial recognition systems are deployed at examination centers to match the registered facial images of candidates against the facial images of prospective examinees.

The novel identity target attack described in this chapter introduces perturbations in the facial image of the real candidate to create a ma-

nipulated image that looks just like the real candidate, but tricks an automated facial recognition system by outputting the identity features of the imposter who plans to take the exam on behalf of the candidate. Experiments using 3,000 image pairs from the Labeled Faces in the Wild (LFW) dataset demonstrate the effectiveness of target identity attacks in white-box as well as black-box scenarios.

Future research will focus on developing an algorithm that detects manipulated images and localizes the manipulated regions. Additionally, research will attempt to characterize the properties of manipulated images in the forensic context to identify the specific technique used for manipulation. This will help develop advanced facial authentication systems that are robust to attacks and provide forensically-sound evidence of manipulation.

References

[1] I. Batskos, A. Macarulla Rodriguez and Z. Geradts, Face morphing detection, *Proceedings of the Twentieth Irish Machine Vision and Image Processing Conference*, pp. 162–172, 2018.

[2] N. Damer, V. Boller, Y. Wainakh, F. Boutros, P. Terhorst, A. Braun and A. Kuijper, Detecting face morphing attacks by analyzing the directed distances of facial landmark shifts, *Proceedings of the German Conference on Pattern Recognition*, pp. 518–534, 2018.

[3] N. Damer, A. Saladie, A. Braun and A. Kuijper, MorGAN: Recognition vulnerability and attack detectability of face morphing attacks created by generative adversarial networks, *Proceedings of the Ninth IEEE International Conference on Biometrics Theory, Applications and Systems*, 2018.

[4] L. Debiasi, C. Rathgeb, U. Scherhag, A. Uhl and C. Busch, PRNU variance analysis for morphed face image detection, *Proceedings of the Ninth IEEE International Conference on Biometrics Theory, Applications and Systems*, 2018.

[5] M. Ferrara, R. Cappelli and D. Maltoni, On the feasibility of creating double-identity fingerprints, *IEEE Transactions on Information Forensics and Security*, vol. 12(4), pp. 892–900, 2017.

[6] G. Huang, M. Mattar, T. Berg and E. Learned-Miller, Labeled faces in the wild: A database for studying face recognition in unconstrained environments, presented at the *Workshop on Faces in Real-Life Images: Detection, Alignment and Recognition*, 2008.

[7] I. Korshunova, W. Shi, J. Dambre and L. Theis, Fast face-swap using convolutional neural networks, *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3697–3705, 2017.

[8] A. Krizhevsky, I. Sutskever and G. Hinton, ImageNet classification with deep convolutional neural networks, *Proceedings of the Twenty-Sixth Annual Conference on Neural Information Processing Systems*, pp. 1106–1114, 2012.

[9] V. Kushwaha, M. Singh, R. Singh, M. Vatsa, N. Ratha and R. Chellappa, Disguised faces in the wild, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1–9, 2018.

[10] A. Makrushin, C. Kraetzer, T. Neubert and J. Dittmann, Generalized Benford's law for blind detection of morphed face images, *Proceedings of the Sixth ACM Workshop on Information Hiding and Multimedia Security*, pp. 49–54, 2018.

[11] A. Makrushin, T. Neubert and J. Dittmann, Automatic generation and detection of visually faultless facial morphs, *Proceedings of the Twelfth International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pp. 39–50, 2017.

[12] V. Mirjalili, S. Raschka and A. Ross, Gender privacy: An ensemble of semi adversarial networks for confounding arbitrary gender classifiers, *Proceedings of the Ninth IEEE International Conference on Biometrics Theory, Applications and Systems*, 2018.

[13] T. Neubert, C. Kraetzer and J. Dittmann, A face morphing detection concept with a frequency and spatial domain feature space for images on eMRTD, *Proceedings of the Seventh ACM Workshop on Information Hiding and Multimedia Security*, pp. 95–100, 2019.

[14] A. Othman and A. Ross, Privacy of facial soft biometrics: Suppressing gender but retaining identity, *Proceedings of the Computer Vision – European Conference on Computer Vision 2014 Workshops*, pp. 682–696, 2014.

[15] R. Raghavendra, K. Raja, S. Venkatesh and C. Busch, Transferable deep-CNN features for detecting digital and print-scanned morphed face images, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1822–1830, 2017.

[16] C. Rathgeb and C. Busch, On the feasibility of creating morphed iris codes, *Proceedings of the IEEE International Joint Conference on Biometrics*, pp. 152–157, 2017.

[17] U. Scherhag, D. Budhrani, M. Gomez-Barrero and C. Busch, Detecting morphed face images using facial landmarks, *Proceedings of the Eighth International Conference on Image and Signal Processing*, pp. 444–452, 2018.

[18] 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.

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

[20] K. Simonyan and A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, *arXiv*: 1409.1556v6, 2015.

[21] Staff Writer, 11 members of police exam cheating gang arrested in Jodhpur, *The Pink City Post*, July 12, 2018.

[22] L. Wandzik, G. Kaeding and R. Vicente-Garcia, Morphing detection using a general purpose face recognition system, *Proceedings of the Twenty-Sixth European Signal Processing Conference*, pp. 1012–1016, 2018.

[23] A. Yuhas, Chinese nationals charged with cheating by impersonation on US college tests, *The Guardian*, May 28, 2015.

[24] L. Zhang, F. Peng and M. Long, Face morphing detection using the Fourier spectrum of sensor pattern noise, *Proceedings of the IEEE International Conference on Multimedia and Expo*, 2018.