ISSN (Online) 2582-9327

IJARSCMT

International Journal of Advanced Research in Science, Commerce, Management and Technology

Impact Factor: 5.781

Volume 2, Issue 5, May 2021

Face Recognition Systems Countenance Attacks

Yashvardhan SG

Standard Fireworks Rajaratnam College for Women, Sivakasi, Tamil Nadu, India

Abstract: *Due to its simplicity, face authentication is currently more frequently used than authentication using a personal identification number or an unlock pattern, especially on mobile devices. This has made it a seductive target for attackers who use a demonstration assault. Traditional presentation attacks employ the victim's face or victim footage. The existence of master faces—faces that match numerous enrolled templates in face recognition systems—has been demonstrated in earlier research, and their presence increases the effectiveness of presentation attacks. In this article, we present the results of a thorough investigation of latent variable evolution (LVE), a technique frequently employed to produce master faces. To determine the characteristics of master faces, an LVE algorithm was used in a variety of settings and with many databases and/or face recognition systems.*

Keywords: Master face, wolf attack, face recognition system, latent variable evolution

I. INTRODUCTION

Strong passwords, which can be challenging to remember, should be used, and they should be updated frequently to maintain security. Passwords are less handy than personal identification numbers and unlock patterns, yet the user is still needed to keep them in mind, and passersby might be able to sneak a glimpse at them. Biometric authentication, which makes use of a distinctive biometric feature, is an even more practical technique.

Fig. depicts the phases of master biometrics research. The following is a summary of our contributions:

We are the first to create master faces that can match many faces with various identities, building on our earlier work [4]. We extend our prior work by examining the impact of employing multiple databases (DBs) and/or several FR systems for the latent variable evolution (LVE) algorithm used to construct master faces. This capability makes FR systems susceptible to a master face attack. Attack performance was improved overall by some DB/FR system combinations, but not by others due to to disputes within components. Understanding the circumstances in which strong master faces can be formed and properly assessing the dangers require knowledge of the successful combinations.

II. RELATED WORK

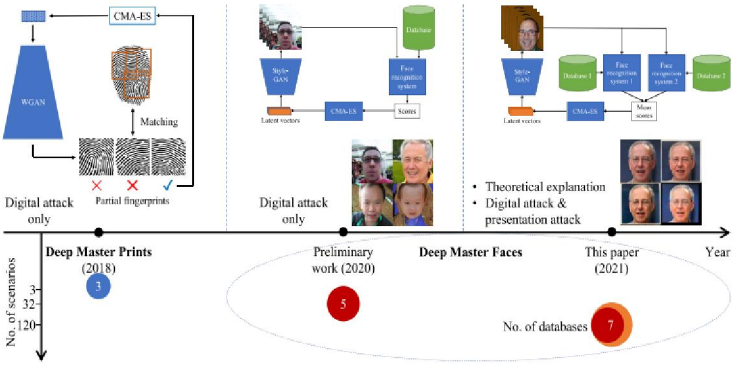
2.1 Facial Image Generation

The face is a frequent target in deep learning research, which is a key area of study. Variational autoencoders (VAEs) and generative adversarial networks are the two main methods for creating images.(GANs) . They could initially only produce little, poor-quality images. GANs were challenging to train, while VAEs frequently produced hazy images. The training problem was handled by later GAN advancements (WGAN)and WGAN Gradient Penalty (WGAN-GP)), which allowed GANs to be utilised to create master prints .High-resolution images can be produced using VAEs and GANs] in their most recent iterations. Karras et alprogressive .'s GAN was successful in producing images with a resolution of 1024 1024 pixels by gradually adding more layers during training to produce larger images .Later research improved the disentanglement network termed StyleGAN by fusing the concepts of progressive training and style transfer. StyleGAN employs a mapping network to convert a latent vector into intermediate style vectors that are used to synthesise images, in contrast to typical GANs that use a latent vector directly to generate images. The facial characteristics can be altered by manipulating these intermediary style vectors. The best techniques for creating master faces are Style GAN and its following version, which have the strengths of powerful disentanglement and high-quality facial image production.

2.2 Face Recognition

Large databases have been made available, such as the CASIA-WebFace database [19] and the MS-Celeb database , and recent developments in convolutional neural networks (CNNs) have greatly enhanced FR system performance and made it possible for them to function well in diverse domains . A network design that performed well in the ImageNet Challenge is used by the majority of cutting-edge FR systems, including the VGG (Visual Geometry Group) network architecture and the inception network architecture. The VGG-Face network was developed by Parkhi et al. using training data from a large-scale database they created themselves. Ten times fewer parameters than the VGG-Face network are proposed by Wu et al. in their lightweight CNN . De Freitas Pereira et al. built heterogeneous FR networks using the inception design , and Schroff et al. created the FaceNet network . FaceNet was re-implemented by Sandberg as an open-source platform . DeepFace is a system developed by Taigman et al. that uses explicit 3D face modelling to enhance the facial alignment stage and a CNN to extract face representation . Contrary to other approaches that employ discriminative classifiers, Tran et algenerative .'s classifier, dubbed DR-GAN, learns a disentangled representation .The embedding distribution optimization is the main emphasis of more recent methods. To increase the FR model's ability to discriminate between different inputs and to stabilise the training process, Deng et al. suggested employing the additive angular margin loss (ArcFace) rather than the more widely utilised cosine distance loss . In their UniformFace FR system, Duan et al. advocated adopting a uniform loss to train equally distributed representations since they believed that the distribution of the features is crucial .

The biometric (FR) system's intended policy by the (facial) capturing subsystem. 1 A photo attack is a presentation attack in which the attacker shows the FR system sensor a picture of the victim. This image can be printed on paper or viewed on a screen device (such as a laptop, tablet, or smartphone) . Another presentation assault is a replay attack, which plays the victim's video instead of presenting a picture . An FR system can incorporate a presentation attack detector to reduce presentation attacks .



Building on some of the innovations mentioned above, we carried out rigorous experiments with four modern (and conceptually dissimilar) state-of-the-art master faces to explore the security danger posed by master faces.

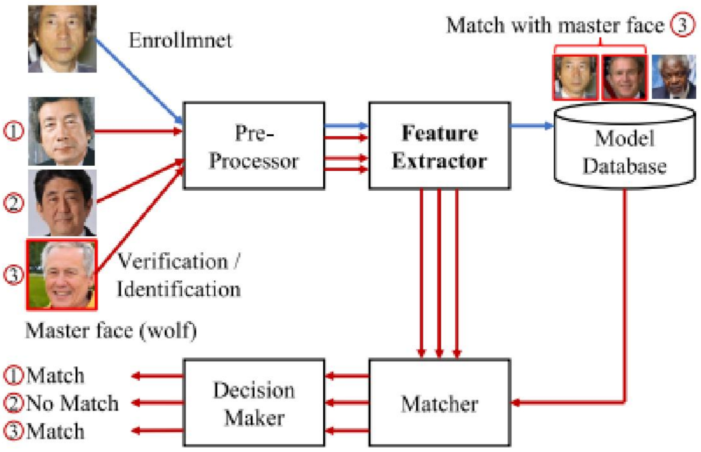


Fig. 3. Operation of typical FR system. There are two phases: enrollment (blue path) and verification/identification (red path). The master face (face 3) was falsely matched with the two faces of two enrolled subjects. Best viewed in color.

2.4 Latent Variable Evolution

Since they do not require any assumptions about the underlying fitness landscape, evolution algorithms are frequently employed in artificial intelligence applications to approximate complicated, multimodal, and non-differentiable functions. Designed for non-linear and non-convex problems, the covariance matrix adaptation evolution strategy (CMA-ES) is a potent method functions . Bontrager et al.

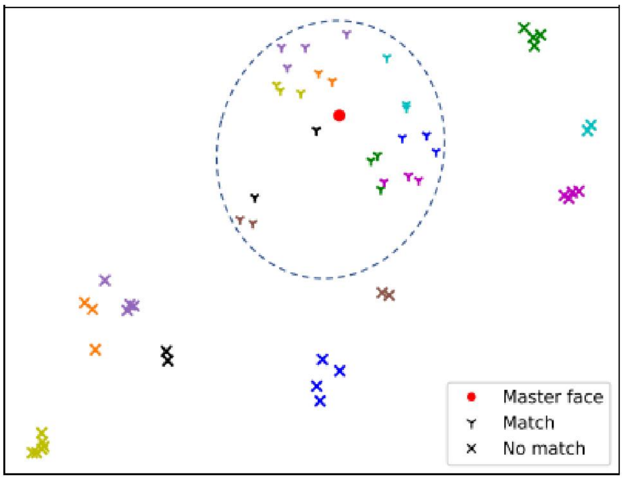


Fig. 4. UMAP visualization of identity space containing embeddings of a master face and of "match" and "no-match" faces of 18 enrolled subjects. For each cluster (match or no match), symbols with the same color correspond to the same subject. Best viewed in color.

III. DEEP MASTER FACES

3.1 Existence of Master Faces

We briefly describe the existence of master faces before going into the suggested master face

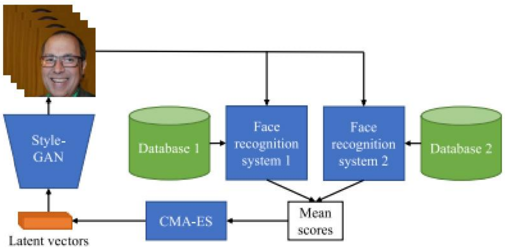
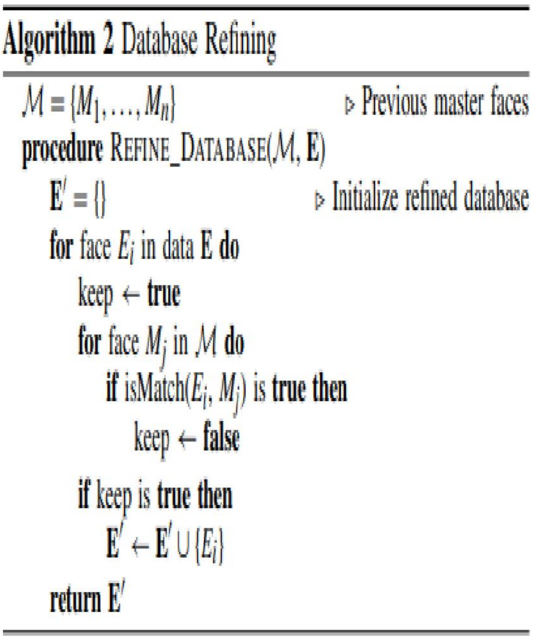


Fig. 6. Overview of extended LVE algorithm. Latent vectors are fed into StyleGAN [17] to generate facial images. One or more surrogate FR system(s) then calculates mean score for each image on the basis of the subjects in one or more database(s). For example, for the *combination 3* setting described in Table III, **database 1** is LFW - Fold 1, **database 2** is mobile biometry (MOBIO), **FR system 1** is Inception-ResNet-v2 network (trained on MS-Celeb database), and **FR system 2** is DR-GAN network. The CMA-ES [35] algorithm uses these scores to generate new latent vectors.

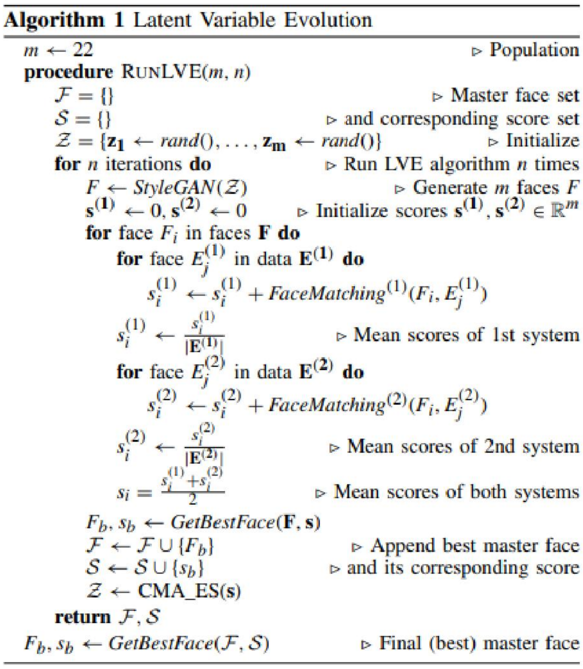
We extended our previous work by using one more database and/or FR system to generate master faces, which requires support from the LVE algorithm. The extended LVE algorithm is formalised in Algorithm 1 and is depicted in Fig. 6. M latent vectors, z1,..., zm, are first initialised at random first. They are then used as input into a style-trained GAN network to produce faces, m. The similarity between all subject faces in databases E(1) j and E(2) j is determined by two face matching functions, FaceMatching(1) (,), and FaceMatching(2) (,) (corresponding to two FR systems). The outcomes of FaceMatching(1) (,) and FaceMatching(2) (,) yield two m dimension mean score vectors, s(1) and s(2).

The best local master face Fb is chosen among the m produced faces using the mean s of these two vectors. Finally, s is used to feed fresh latent vectors (z1,..., zm) into the CMA-ES algorithm.



Algorithm 2 Database Refining

```
$\mathcal{M} = \{M_1, \ldots, M_n\}$                ▷ Previous master faces
procedure REFINE_DATABASE($\mathcal{M}$, E)
    $E' = \{\}$                          ▷ Initialize refined database
    for face $E_i$ in data E do
        keep ← true
        for face $M_j$ in $\mathcal{M}$ do
            if isMatch($E_i$, $M_j$) is true then
                keep ← false
        if keep is true then
            $E' \leftarrow E' \cup \{E_i\}$
    return $E'$
```



Algorithm 1 Latent Variable Evolution

```
$m \leftarrow 22$                                        ▷ Population
procedure RUNLVE($m$, $n$)
    $\mathcal{F} = \{\}$                                 ▷ Master face set
    $\mathcal{S} = \{\}$                     ▷ and corresponding score set
    $\mathcal{Z} = \{\mathbf{z_1} \leftarrow rand(), \ldots, \mathbf{z_m} \leftarrow rand()\}$    ▷ Initialize
    for $n$ iterations do                 ▷ Run LVE algorithm $n$ times
        $F \leftarrow StyleGAN(\mathcal{Z})$              ▷ Generate $m$ faces $F$
        $\mathbf{s}^{(1)} \leftarrow 0, \mathbf{s}^{(2)} \leftarrow 0$   ▷ Initialize scores $\mathbf{s}^{(1)}, \mathbf{s}^{(2)} \in \mathbb{R}^m$
        for face $F_i$ in faces $\mathbf{F}$ do
            for face $E_j^{(1)}$ in data $\mathbf{E}^{(1)}$ do
                $s_i^{(1)} \leftarrow s_i^{(1)} + FaceMatching^{(1)}(F_i, E_j^{(1)})$
            $s_i^{(1)} \leftarrow \frac{s_i^{(1)}}{|\mathbf{E}^{(1)}|}$           ▷ Mean scores of 1st system
            for face $E_j^{(2)}$ in data $\mathbf{E}^{(2)}$ do
                $s_i^{(2)} \leftarrow s_i^{(2)} + FaceMatching^{(2)}(F_i, E_j^{(2)})$
            $s_i^{(2)} \leftarrow \frac{s_i^{(2)}}{|\mathbf{E}^{(2)}|}$           ▷ Mean scores of 2nd system
            $s_i = \frac{s_i^{(1)} + s_i^{(2)}}{2}$           ▷ Mean scores of both systems
        $F_b, s_b \leftarrow GetBestFace(\mathbf{F}, \mathbf{s})$
        $\mathcal{F} \leftarrow \mathcal{F} \cup \{F_b\}$            ▷ Append best master face
        $\mathcal{S} \leftarrow \mathcal{S} \cup \{s_b\}$       ▷ and its corresponding score
        $\mathcal{Z} \leftarrow$ CMA_ES($\mathbf{s}$)
    return $\mathcal{F}, \mathcal{S}$
$F_b, s_b \leftarrow GetBestFace(\mathcal{F}, \mathcal{S})$       ▷ Final (best) master face
```

Out of the n best master faces F obtained in the n iterations, the overall (global) best master face is selected.When creating a new master face, all faces that match the As seen in Algorithm 2, it is necessary to eliminate any previously generated master face(s) from the training database(s). As a result, the new master face won't cover the old master face (s). displays an illustration of a second master face alongside the first master face, the real wolf face, and their associated FMRs. The second master face's FMR is lower than the first one's, and any succeeding master face often follows suit.

IV. GENERATING MASTER FACES

We created a number of LVE algorithm settings and a number of attack scenarios that cover white-box, gray-box, and black-box attacks in order to assess the dangers and threats posed by a master face attack. While only one of the target FR system's architecture or its training database is known for gray-box attacks, both are known for white-box attacks. There is no information available regarding the target FR system for black-box assaults.

4.1 Experiment Materials

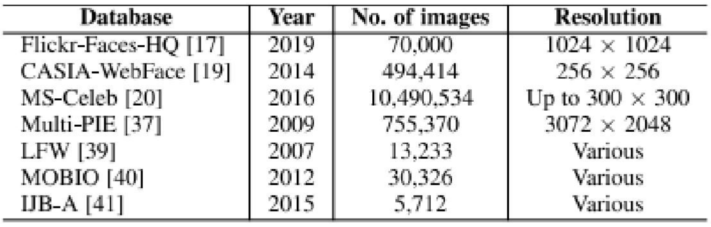
A. Face Recognition Systems

We used five mainstream publicly available high-performance FR systems in our experiments:

One trained on the CASIA-WebFace database and one trained on the MS-Celeb database by de Freitas Pereira et al. are Inception-ResNet-v2 based FR systems.

Sandberg implemented and trained an open-source version of FaceNet on the MS-Celeb database.

TABLE I
DETAILS OF DATABASES USED IN OUR EXPERIMENTS



Database	Year	No. of images	Resolution
Flickr-Faces-HQ [17]	2019	70,000	1024 × 1024
CASIA-WebFace [19]	2014	494,414	256 × 256
MS-Celeb [20]	2016	10,490,534	Up to 300 × 300
Multi-PIE [37]	2009	755,370	3072 × 2048
LFW [39]	2007	13,233	Various
MOBIO [40]	2012	30,326	Various
IJB-A [41]	2015	5,712	Various

The CASIA-WebFace database and the Multi-PIE database were used to train the DR-GAN.

The MS-Celeb database was used to train ArcFace .

The two Inception-ResNet-v2 based FR systems, DR-GAN, and all of the FR systems were utilised to generate master faces and assess master face attacks.

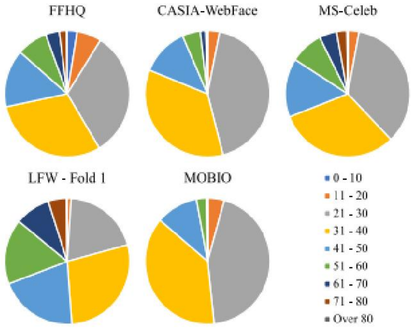


Fig. 7. Estimated age distribution of five databases used for training StyleGAN, FR systems, and generation of master faces. Best viewed in color.

B. Latent Variable Evolution Configurations

With the current computation and temporal resources, it is not possible to evaluate all conceivable combinations due to the large number of FR systems and databases. So, in order to provide the broadest feasible coverage, we chose a subset. We created eight parameters (Table II) for the LVE method utilising three FR systems, two databases (LFW - Fold 1 and MOBIO), and two versions of Inception-ResNet-v2 (one trained on the CASIA-WebFace database and one trained on the MS-Celeb database). One FR system and one database are used in five settings (single 1 to single 5) while more than one FR system and/or database are utilised in three settings (combination 1, combination 2, and combination 3).

Each combination setting, which merged two single settings, was chosen for having adequate case coverage.Table III highlights the primary variations between the three combination settings. The databases used to train the FR systems were comparable, and just one database was used with the LVE algorithm in the combination 1 setting.

V. PRESENTATION ATTACKS

Finally, we assessed the threat and risk of presentation attacks on FR systems utilising master faces. We selected two master face candidates: one produced using the combination 1 setting and another using the single 2 setting. We selected two attack scenarios from the IJB-A database where the two master faces were mistakenly accepted by the Inception-ResNet-v2 based FR system (CASIA-WebFace version) and the DR-GAN FR system as candidates for digital attacks.

VI. DEFENSE AGAINST MASTER FACE ATTACKS

What fundamental flaw in the current FR architecture gives rise to master faces? We postulated that it originates from embedding space distributions where the retrieved features are not evenly distributed. As a result, clusters develop, including multi-identity clusters as well as ones based on age and gender. The training data and the objective function design are two potential causes of this issue. The training data was uneven in terms of age and gender, as can be seen in Figs. 7 and 8. This might change how the FR systems distribute the embeddings so that faces in the majority group are discriminated against more effectively than those in the minority group. For instance, the face embeddings from 30 to 60 years old.

When it comes to objective function design, the major goal is to keep same-identity embeddings close together while keeping different-identity embeddings apart.

This is made better by the inclusion of the angular margin loss whereas the embeddings are forced to have a uniform distribution by the uniform loss. These upgrades mostly concentrate on identity, despite the fact that they lessen the chance of master face attacks.

The attack is successful in some instances because gender, age, and race are also significant factors. This shows that there is room for improvement in the design of the objective functions used to train the FR systems.

VII. CONCLUSION

We have one more shown that master face attacks constitute a serious security danger if the FR systems are not well safeguarded, particularly in our presentation attack experiment. Our thorough analysis of the LIVE algorithm's effectiveness in a variety of circumstances, including single and combination settings, has revealed various characteristics of master faces as well as the LVE algorithm itself. While some of the combination settings led to intra-component disputes, others had intriguingly favourable outcomes. To increase the robustness of FR systems, it is essential to understand the existence of master faces and their characteristics. Master face attacks could be mitigated by combining the employment of a FR system with a well-designed goal function trained on a sizable balanced database and a false picture detector. since master face attacks continue to improve, these attacks cannot be taken lightly. Future work will focus on designing a better method to generate master faces and one to detect master face attacks.

REFERENCES

[1]. “The Goode Intelligence Biometric Survey 2021.” Goode Intelligence. Apr. 2021. [Online]. Available:https://www.goodeintelligence.com/ report/the-goode-intelligence-biometric-survey-2021/

[2]. S. Bhattacharjee, A. Mohammadi, A. Anjos, and S. Marcel, “Recent advances in face presentation attack detection,” in Handbook of Biometric Anti-Spoofing. Cham, Switzerland: Springer, 2019, pp. 207–228.

[3]. P. Bontrager, W. Lin, J. Togelius, and S. Risi, “Deep interactive evolution,” in Proc. Int. Conf. Comput. Intell. Music Sound Art Des., 2018, pp. 267–282.

[4]. H. H. Nguyen, J. Yamagishi, I. Echizen, and S. Marcel, “Generating master faces for use in performing wolf attacks on face recognition systems,” in Proc. IJCB, 2020, pp. 1–10.

[5]. 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.