Adversarial Learning of Mappings Onto Regularized Spaces for Biometric Authentication

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 17, 2020, accepted August 9, 2020, date of publication August 13, 2020, date of current version August 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3016599

Adversarial Learning of Mappings Onto Regularized Spaces for Biometric Authentication

ARSLAN ALI[1], (Graduate Student Member, IEEE), MATTEO TESTA[1], (Member, IEEE), LEV MARKHASIN[2], TIZIANO BIANCHI[1], (Member, IEEE), AND ENRICO MAGLI[1], (Fellow, IEEE)

[1]Department of Electronics and Telecommunication Engineering, Politecnico di Torino, 10129 Torino, Italy
[2]Sony Europe B.V., Stuttgart Technology Center, 70327 Stuttgart, Germany

Corresponding author: Arslan Ali (arslan.ali@polito.it)

This work was supported in part by the Sony Research and Development Center Europe Stuttgart Laboratory 1.

ABSTRACT We present AuthNet: a novel framework for generic biometric authentication which, by learning a regularized mapping instead of a classification boundary, leads to higher performance and improved robustness. The biometric traits are mapped onto a latent space in which authorized and unauthorized users follow simple and well-behaved distributions. In turn, this enables simple and tunable decision boundaries to be employed in order to make a decision. We show that, differently from the deep learning and traditional template-based authentication systems, regularizing the latent space to simple target distributions leads to improved performance as measured in terms of Equal Error Rate (EER), accuracy, False Acceptance Rate (FAR) and Genuine Acceptance Rate (GAR). Extensive experiments on publicly available datasets of faces and fingerprints confirm the superiority of AuthNet over existing methods.

INDEX TERMS AuthNet, adversarial learning, biometric authentication, face authentication, fingerprint authentication, latent space regularization.

I. INTRODUCTION

Biometric authentication systems are drawing increasing attention thanks to their convenience: the users are authenticated based on information they inherently own avoiding the need to remember passwords or provide keys. The typical approach followed by such systems is based on template matching: each biometric trait is associated with a template which should be able to embed its most discriminative features. Hence, all templates of a biometric trait belonging to the same user should be close in some suitable distance metric. Once the user's face, fingerprint or other biometric trait have been acquired through a dedicated sensor, they are processed in order to obtain the corresponding templates which are then stored in a secure fashion. This phase, which is referred to as the enrollment, prepares the system to grant access only to the enrolled users. At this point the system can be used in verification phase: a fresh biometric trait of a user requesting authentication is acquired, the associated template is computed and then matched with the stored ones. Depending on the outcome of the matching process, the user can be either granted or denied access to the system.

Focusing on authentication accuracy the most critical part of a biometric authentication system is the feature extraction. Indeed, the extracted features not only have to be the most discriminative ones but should also be embedded in a proper metric space, in order to enable the template matching. Traditionally, features were extracted by means of hand-crafted design. However, the advent of deep learning methods highlighted the great advantage of learning the best features from data instead of using a model-based design, in terms of learning complex mappings [1], [2] and addressing difficult classification tasks [3].

When considering deep learning approaches, the biometric authentication problems are usually addressed by learning a feature embedding in which a template is able to represent the most discriminative features of a specific biometric trait class in a suitable space. Similarly to standard biometric

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

authentication systems, the learned features are shared among different users and the template matching is based on a distance measure between two or more embeddings. In this work we follow a different path: with AuthNet we rely on a classification-based approach in which the neural network not only learns the most discriminative features of a specific user's biometric traits, but also learns the boundaries which can separate that specific user with respect to every other user.

As classification-based approaches require a per-user training, they trade off the added complexity with improved, user-specific features. Note that, the training process of embedding-based networks requires a large amount of labelled data as the network has to learn the very general features of the data class. The classification-based approach avoids this since, in a user-specific training the network has been trained on *that* specific user for which the most discriminative features have been learned. Conversely, embedding-based approaches learn specific features of the considered class, e.g. faces, and may fail on a specific user.

In this regard, it is important to underline that deep learning based classification learns highly non-linear boundaries with complex shapes in order to partition the feature space [4]. As shown in [4], the geometry of the decision boundaries heavily affects the robustness of the classifier. More specifically, as discussed in [5] most of the mass of the data points gathers close to the decision boundaries. As such, two similar biometric traits of a user may be assigned to different classes, leading to an error. Moreover, this undesirable behavior is an intrinsic property of the classifier structure and does not depend on the visual properties of the input data [5].

For the above reasons, we propose a novel *user-specific* classification strategy which does not explicitly enforce the network to learn complex classification boundaries. Instead, we envision a network design which learns a mapping of the input biometric traits onto a regularized and well-behaved latent space. By following this approach, the feature distributions are regularized so as to lead to simple and tunable boundaries between the classes, thereby reducing the probability of misclassification. In particular, we aim to obtain ''non-arbitrary'' boundaries which can lead to improved accuracy and increased robustness.

The first step consists in learning a compact and meaningful mapping of the input biometric traits onto the latent space. The latent space should be shaped in a simple and well-defined way: authorized and unauthorized users should cluster in two different and compact regions of the space leading to very regular boundaries. Then, a decision is made by employing a linear decision boundary to discriminate between the authorized user and everyone else. This system, which we will refer to as AuthNet, makes use of adversarial training in order to enforce a proper shaping of the latent space. With this paper we extend and improve our previous work [6] by introducing a new loss function via selection of better statistical parameters, provide an in-depth discussion on how AuthNet correctly maps users that are misclassified by other approaches and motivation behind higher misclassification rate by competing methods, introduce new architectural designs which come in two different flavors, based on ResNet [7] and DenseNet [8] with detailed performance comparison on the average values of the considered metrics computed independently on each user and aggregated scores. We further provide a detailed analysis of robustness on new datasets not seen during the training and on targeted perturbations, and verify how the regularization of latent space to simple target distributions leads to robust authentication compared to learning of the boundaries. Further, we add a discussion on the choice of the optimal system parameters.

II. RELATED WORK

Over the years, different methods have been proposed to address the biometric authentication task when dealing with different biometric traits such as faces, fingerprints, retinas and gait. With this work, we specifically focus on the most widely spread biometric modalities, namely face and fingerprint.

A. FACES

The face as a practical biometric modality has appeared only recently because of the inherent difficulty in handling far from ideal acquisition conditions. Indeed, standard model-based approaches tend to exhibit a high variance with respect to pose and/or illumination changes. A pioneer in this sense is the well-known eigenfaces approach [9]: the features used to describe the faces are obtained by projecting the test image onto the space spanned by the eigenvectors computed on the training data. Some of its weaknesses have been surpassed with the introduction of the Fisherfaces method [10] in which the projection operator is learned in a supervised fashion in order to maximize (minimize) inter (intra) class variance. This approach allowed a higher degree of invariance with respect to illumination changes. Other standard approaches are based on low-dimensional representations of the faces; examples include sparse representations [11], [12], linear subspace [13], [14] and manifold [15] representations. Following a different approach, [16], [17] attempt to overcome the limitations in handling facial changes by employing local features.

The largest performance improvement has been achieved by means of deep learning methods. These allowed to obtain excellent performance in far from ideal acquisition under different pose, expression and illumination conditions, see e.g. Deep face [18]. One of the most well-know methods is Facenet [19] which uses a triplet loss in order to learn embeddings of the input images. More specifically, the network is trained in such a way that the embeddings preserve the notion of image similarity in terms of ℓ_2 distance in the embedding space. However, because of the instability arising during a triplet-loss training, it is common to train the network with a softmax cross-entropy loss. Nevertheless, in this case the intra-class compactness and inter-class dispersion is not guaranteed. A more recent approach named ArcFace [20] introduced the additive angular margin loss to improve the

discriminative power of the learned embedding whilst leading to a stable training process. A few other works also adopt the same strategy, e.g. [21]–[25].

All the above works rely on the recent trend in face recognition based on embedding computation and matching. Indeed, most research efforts are spent on the design of novel loss functions which can lead to more effective and/or stable embeddings.

In this regard, let us better highlight the scope of AuthNet with respect to recent trends in unconstrained face *recognition*. With this work we are specifically focusing on the biometric *authentication* problem for which, apart from achieving a high recognition accuracy, it is even more crucial to reduce the number of wrongly authorized users. For the same reason, it is common to assume that the user puts him/herself in a controlled condition and as such, the face datasets we consider, are those commonly used for biometric authentication tasks, see [26], [27]. Conversely, recent "in-the-wild" face datasets, because of the large number of users and poses, are better suited for the evaluation of recognition and clustering tasks. Lastly, such datasets do not cope well with a user-specific training procedure as done in AuthNet since the number of samples per user is very limited.

B. FINGERPRINTS

This was one of the first biometric traits to be commonly used in practical systems. As such, most of the approaches rely on standard template matching based on hand-crafted features computed from minutiae, ridge and valleys patterns or global intensity image. In general, they can be categorized based on the use of either global or local features. Among the methods relying on global fingerprint features we mention the works in [28], [29]. Conversely, the approaches proposed in [30]–[34] rely on descriptors making use of local information of the minutiae and their neighbourhood. Additionally, in works such as [35] it has been shown that performance improvements can be achieved when additional information such as shape context and orientation is included. In the last few years, new approaches have been proposed in order to take advantage of the deep learning representational capabilities, for example, in order to improve the robustness of minutiae extraction and classification. Examples include [36], [37] in which Convolutional Neural Networks (CNN) are used to extract minutiae from raw fingerprint images and [38] where a stacked autoencoder is used to classify fingerprints into arch, left/right loop, and whorl. In [39] minutiae are filtered using a neural network to improve detection, whereas in [40] the authors use a neural network to extract the minutiae on thinned fingerprint images. Latent fingerprint minutiae extraction based on CNN has been also proposed in [41].

III. PROPOSED METHOD

In this section we introduce and describe the components of the proposed architecture for biometric authentication as shown in Fig. 1. As previously discussed, AuthNet strives to find a well-behaved representation of the input biometric traits in some latent space, which in turn enables simple decision boundaries to be used for the classification task. More specifically, as described in the following, we want to learn a mapping from a sample in the biometric space onto a sample of target probability distributions for authorized and unauthorized users. Ideally, the distance (in some suitable metric) between the probability distribution of the samples resulting from the mapping and the target one should be minimal. One of the most widely used approaches to tackle this kind of problem is by means of an adversarial game.

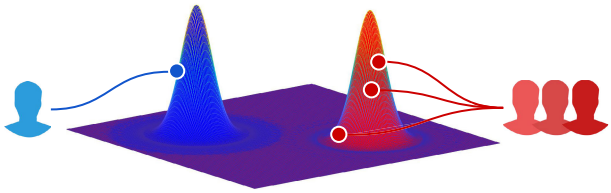


FIGURE 1. The goal of AuthNet is to map the input biometric traits onto target distributions in the latent space. Authorized users (blue) are mapped to a target distribution whose mean value is far from that of the unauthorized users (red).

A. ADVERSARIAL LEARNING

Adversarial models are now a very widespread approach to generative models. The first generative model trained by means of an adversarial loss, the Generative Adversarial Network (GAN) [1], gained immediate popularity and opened the path to the field of adversarial training.

A GAN tries to implicitly learn the probability distribution of the input data in such a way that the network is then able to *generate* samples similar to the input data. In other words, the network learns to minimize a distance metric between the distribution of the generated samples and that of the real data. The distance metric employed by GAN is Jensen-Shannon (JS) divergence which, interestingly, is the optimal solution of a two-player adversarial game. The main idea behind adversarial models is to reach the minimum of a functional defined as a minimax game where two entities have adversarial (opposite) goals. The global optimum corresponds to the equilibrium solution between the locally optimal solutions of the single entities. Within the deep learning framework, the two entities called generator and discriminator are modeled as neural networks and the minimax game is introduced in the loss function in order to make the two networks compete against each other during the training process. In more detail, the discriminator should be able to correctly discriminate between generated and real samples, while the generator should be able to generate samples which are realistic enough to fool the discriminator.

In AuthNet, as described in detail in the following, samples of the data distribution are mapped onto a latent representation which follows a target distribution. This can be considered as the inverse mapping of a conventional GAN, in which samples of a fixed distribution are mapped onto the captured distribution of the data.

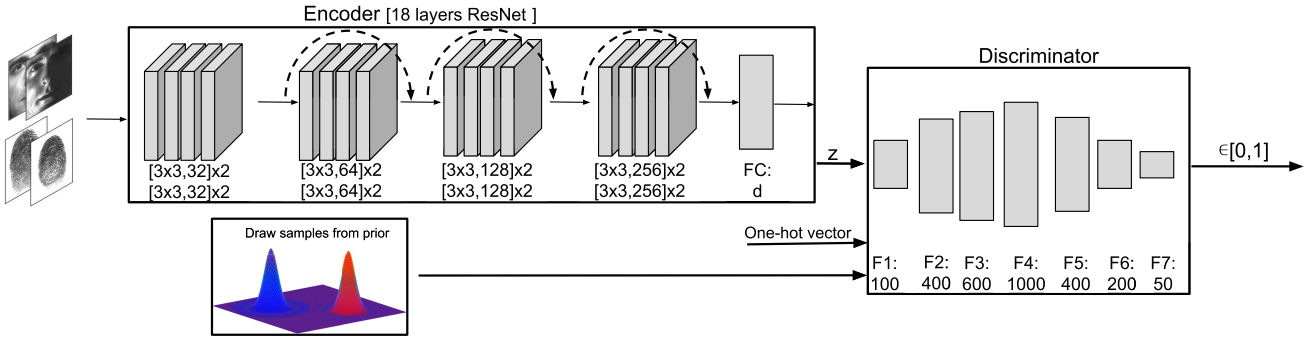


FIGURE 2. AuthNet-R architecture at enrollment phase. Training biometric traits are given as input to the encoder which consists of an 18-layered residual network followed by a fully connected layer. The output of the encoder, together with a one-hot vector and samples of the target distributions, is given as input to the discriminator which is made of 6 fully connected layers.

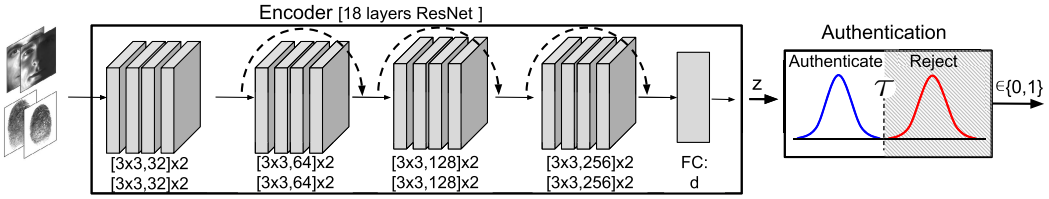


FIGURE 3. AuthNet-R architecture at authentication phase. In this phase the biometric trait of a user requesting access is given to the pre-trained encoder which will output a sample z coming from either $\mathbb{P}_0$ or $\mathbb{P}_1$. Then, the thresholding decision is made and a binary output (accept or reject) is returned.

B. LATENT MAPPING

We are now ready to provide the details of AuthNet whose main concept is depicted in Fig. 1.

Let $\mathcal{B} = \{\mathcal{B}_{a=0}, \mathcal{B}_{a=1}\}$ denote the set of all possible biometric traits and $a \in \{0, 1\}$ an indicator variable such that $a = 1$ represents the authorized user and $a = 0$ represents all other unauthorized users. Moreover, let us define as $\mathbf{x} \in \mathbb{R}^n$ a generic biometric trait in $\mathcal{B}$ and as $\mathbf{z} \in \mathbb{R}^d$ its latent representation with $d < n$. The goal is to learn an encoding function $\mathbf{z} = H(\mathbf{x})$ of the input biometric trait such that $\mathbf{z} \sim \mathbb{P}_1$ if $\mathbf{x} \in \mathcal{B}_{a=1}$ and $\mathbf{z} \sim \mathbb{P}_0$ if $\mathbf{x} \in \mathcal{B}_{a=0}$, with $\mathbb{P}_1$ and $\mathbb{P}_0$ the target distributions in the latent space. If the distributions $\mathbb{P}_1$ and $\mathbb{P}_0$ are well-behaved, a simple distance-based thresholding approach can be employed to determine whether the user with its associated biometric trait $\mathbf{x}$ is authorized or not.

Let us set $\mathbb{P}_1 = \mathcal{N}(\boldsymbol{\mu}_1, \sigma_1\mathbf{I})$ and $\mathbb{P}_0 = \mathcal{N}(\boldsymbol{\mu}_0, \sigma_0\mathbf{I})$ to be Gaussian, this amounts to enclosing the energy of the latent representation of authorized and unauthorized users within hyperspheres whose radius depends on both d and the distribution parameters. For the sake of simplicity and without loss of generality, we set $\mathbb{E}[\mathbf{z}_1] < \mathbb{E}[\mathbf{z}_0]$ with $\mathbf{z}_1 \sim \mathbb{P}_1$ and $\mathbf{z}_0 \sim \mathbb{P}_0$ having $\sigma_1 = \sigma_0$. If the distributions are taken as Gaussian with the same variance, a hyperplane is the optimal decision boundary, which further boils down to a simple threshold when z is a scalar. This leads to a very simple classifier, which learns a complex mapping to a high-dimensional latent space, in a way that mimics kernel-based methods.

Modes of Operation: AuthNet operates in two phases, an enrollment phase and an authentication phase. During the enrollment phase (see Fig. 2), based on the training data users are registered in the system. Latent representation of authorized users are forced to follow $\mathbb{P}_1$, whereas latent representations of unauthorized users are forced to follow $\mathbb{P}_0$ based on the one-hot label vector. Once the enrollment phase is completed, in the following authentication phase (see Fig. 3), the latent representations of the input biometric traits are tested against the target distributions, to find out whether the test biometry belongs to the authorized user class, or to the class of unauthorized users class. For $d = 1$, if the metric value is less than the threshold i.e. $\mathbf{z} \sim \mathbb{P}_1$, the user is categorized as an authorized user, else the user is categorized as an unauthorized user.

C. ENROLLMENT

During the enrollment phase the goal is to learn an encoding function $H(\mathbf{x})$ which maps the user biometric traits onto the target distributions. The optimal $H(\mathbf{x})$ is the one for which a distance metric between $H(\mathbf{x}) : \mathbf{x} \in \mathcal{B}_{a=1}$ and $\mathbb{P}_1$, and between $H(\mathbf{x}) : \mathbf{x} \in \mathcal{B}_{a=0}$ and $\mathbb{P}_0$ is minimized. To address this problem we propose to employ an adversarial model whose optimum is reached when the JS divergence between the latent mapping and target distribution is minimized.

The AuthNet architecture at enrollment phase is depicted in Fig. 2. It is made of two competing neural networks: an encoding function $H(\mathbf{x}, \boldsymbol{\theta}_h)$ having parameters $\boldsymbol{\theta}_h$ and a

discriminator $D(\mathbf{p}, \boldsymbol{\theta}_d)$ with parameters $\boldsymbol{\theta}_d$. For the sake of readability, unless needed, we will drop the parameters in the notation of the encoding and discriminator networks.

The encoding function $H(\cdot)$ takes as input the biometric traits $\mathbf{x}$ and output their encoded latent representation $\mathbf{z}$. The discriminator $D(\mathbf{p})$ takes as input the vector $\mathbf{p} \in \{\mathbf{s}, \mathbf{z}\}$, namely it is given in an alternate fashion either a sample from one of the target distributions $\mathbf{s}$ or the encoded latent representation $\mathbf{z}$. The vector $\mathbf{s} \in \mathbb{R}^d$ is made of randomly drawn samples from the target distributions $\mathbb{P}_1$ if $\mathbf{x} \in \mathcal{B}_{a=1}$ or $\mathbb{P}_0$ if $\mathbf{x} \in \mathcal{B}_{a=0}$, respectively. In order to improve the stability and performance of the training process, the input biometric trait label a is given to the discriminator as an additional information which, a acts as a switch to select a ''sub-discriminator'' function for either authorized or unauthorized users.

The discriminator $D(\mathbf{p})$ outputs a scalar value which can be interpreted as the probability of given input coming either from the encoding function or the target distribution.

The loss function we consider to address the above-defined adversarial setting is given by

$$V(H, D) = \mathbb{E}_{\mathbf{s}\sim\mathbb{P}}\left[\log(D(\mathbf{s}, a))\right] + \mathbb{E}_{\mathbf{x}\sim\mathcal{B}}\left[\log(1 - D(H(\mathbf{x}), a))\right], \quad (1)$$

which is optimized as a minimax two-player game according to

$$\min_{\boldsymbol{\theta}_h} \max_{\boldsymbol{\theta}_d} V(H, D), \quad (2)$$

where the optimization is carried over the parameters $\boldsymbol{\theta}_h$ and $\boldsymbol{\theta}_d$ in an alternate fashion.

Being an adversarial model, the specific goal of the encoding function $H(\mathbf{x})$ is to generate samples which, when given to the discriminator, minimise the probability of D making a correct choice, i.e. generate samples $\mathbf{z}$ which will fool the discriminator. The task of the discriminator $D(\mathbf{p})$ is to **maximize** the probability of assigning the correct label to both latent representations $\mathbf{z}$ and samples from the target distribution $\mathbf{s}$.

At the beginning of the learning phase, the discriminator quickly learns how to distinguish the latent representation $\mathbf{z}$ and the samples from the target distribution $\mathbf{s}$. After some iterations, the encoder learns to generate samples which are closer to the target distributions. Eventually, the encoder will start to generate samples $\mathbf{z}$ which are close enough to $\mathbf{s}$ so that the discriminator is not able to distinguish between them.

In the case of AuthNet, as commonly done for adversarial models, these two objectives are optimized in an alternate fashion: one step for the discriminator followed by one for the encoder.

D. AUTHENTICATION

For AuthNet, during the authentication phase only the trained encoder network is utilized. This network computes the latent representation $\mathbf{z}$ of the input biometric trait. Then, a decision is made according to this value. As said, for our choice of target distributions a hyperplane can be used for the optimal decision, i.e., we can use the test

$$(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)^T \mathbf{z} \lessgtr (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)^T (\boldsymbol{\mu}_0 + \boldsymbol{\mu}_1)/2. \quad (3)$$

For $d = 1$, this boils down to comparing the scalar $\mathbf{z}$ with a threshold $\tau = (\mu_1 + \mu_0)/2$, (see Fig. 3).

IV. TRAINING AND IMPLEMENTATION DETAILS

A. NETWORK INSIGHT

1) ENCODER SUB-NETWORK

A biometric trait in the form of either a RGB or a gray-scale image with size depending on the employed dataset is given as an input to the encoder sub-network. The choice of the encoder is a crucial task. In general, one may employ any state-of-the-art neural network architecture able to learn good features. To prove the idea, we conducted experiments on several neural network architectures such as plain CNN, ResNet [7] and DenseNet [8] with different number of layers. For the considered datasets, it was empirically found that either ResNet-18 or DenseNet-50, followed by a fully connected layer having an output of size d, are sufficient to effectively learn the latent mapping. It is important to notice that in this last layer of the encoder network we do not use any non-linear activation as the output should be mapped to a sample of the target distributions. Further, it was found that if a network with too many parameters, like ResNet-101/152 or DenseNet-121/169 is employed for a small/medium sized datasets, it leads to slower training without performance improvement. This motivates us to use ResNet-18 / DenseNet-50 as the encoder sub-network.

In the following sections we will refer to AuthNet with ResNet encoder sub-network as AuthNet-R and to AuthNet employing DenseNet encoder as AuthNet-D.

2) DISCRIMINATOR SUB-NETWORK

The discriminator sub-network has three main inputs: i) samples from target prior distributions, ii) latent vector output from the encoder sub-network $\mathbf{z}$ having size d, and iii) one-hot vector a used during the training process to tell the discriminator whether the sample is authorized or unauthorized. The discriminator is a fully connected network consisting of 8 layers with the ReLU activation function employed at the output of each layer. This number was chosen empirically so that the discriminator has enough capacity to compete with the encoder sub-network. We found from empirical testing that the chosen network sizes worked well across different d-values, and they make the discriminator strong enough and with enough capacity to compete with the given encoder (i.e. ResNet-18 or DenseNet-50) and thus lead to a stable training. Indeed, the layer size depends on the structure of the encoder sub-network: if the discriminator layers are properly sized the encoder loss might quickly drop to zero, thus stopping the training. We found that 8 discriminator layers are enough to cope with the ''capacity'' (or the number of parameters) of the encoder sub-network.

The input of the discriminator sub-network is the concatenation of latent vector $\mathbf{z}$ from the encoder sub-network and the one hot vector a indicating the class to which the corresponding user belongs to. The first fully connected layer has an output size equal to 100. This size gradually increases to a maximum of 1000. After this, the size gradually decreases with the final layer having an output of size equal to 1 to which a sigmoid activation is applied estimating the probability that the sample is coming from the encoder or the target prior distribution.

3) PREPROCESSING AND TRAINING PARAMETERS

The network is trained using Adam optimizer [42] using an iterative algorithm as discussed in [1]; the optimization is carried out one step for the encoder and one for the discriminator. Weight decay is set to 0.0004 and a dropout of 0.7 is used. The learning rate is set to 0.01 for first 5000 iterations and it is then decreased by a factor of 10 after every 5000 iterations. In total, the network is trained for 30000 iterations. The only pre-processing employed for AuthNet on all considered datasets is energy normalization of the input images.

B. DATA AUGMENTATION

Having a diverse and large dataset is crucial for deep neural networks training. The performance of a neural network depends upon the features learned from the training data. In the case of biometric authentication the acquisition process should be fast and usually the number of acquired samples during the enrollment is very limited. An efficient augmentation strategy is hence needed, so that enough data are provided to the network. In addition, we aim to have a general purpose augmentation strategy which could work for different biometric traits.

As summarized in Fig. 4, our augmentation process is based on both image crops and samples mixup. For each sample of size $m \times m$, all possible crops of size $n \times n$ are extracted. Since the number of positive samples (authorized users) is much less than the number of available negative samples (unauthorized users), we employ two different augmentation factors, namely F and F_1. The former refers to the augmentation factor for the positive samples, the latter for the negative ones. Clearly, in our case $F > F_1$.

After obtaining multiple crops of the samples, positive and negative training samples are mixed using a convex combination as described in [43] in order to create more diverse training samples. As a side advantage, as shown in [43], the mixup also helps to regularize and improve the network generalization. Given a positive and a negative sample, respectively denoted as $\mathbf{x}_{a=1}$ and $\mathbf{x}_{a=0}$, a new sample is fabricated as $\mathbf{x}'_m = \lambda \mathbf{x}_{a=1} + (1-\lambda)\mathbf{x}_{a=0}$, where $\lambda \in [0, 1]$ follows a Beta distribution with parameters α and β that in our case are both fixed to 0.4. This parameter choice results in a distribution peaked at 0 and 1 and achieves the lowest probability for $\lambda = 0.5$. This avoids creating augmented samples that are too distant from the centroid of either class. To associate a label to a newly created sample, we use $l = \text{round}(\lambda)$.

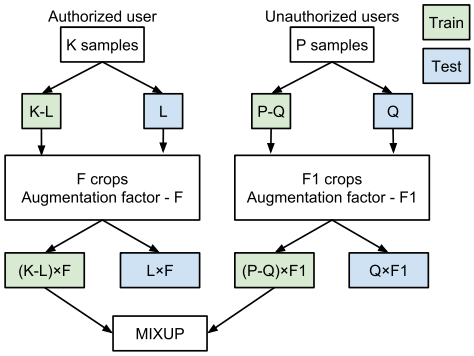


FIGURE 4. The data augmentation makes use of random crops to increase the number of input samples to a factor of F and F_1 for authorized and unauthorized users respectively. Then, positive and negative samples are mixed by means of a convex combination.

V. PERFORMANCE ANALYSIS

AuthNet is a general purpose network designed to seamlessly work on different types of biometric traits. We have conducted experiments on **faces** and **fingerprints**. In biometric authentication systems it is common to assume that the user puts him/herself in a controlled condition for the biometric traits acquisition. In this regard, the datasets we consider are among the biggest ones acquired in such conditions.

A. DATASETS

For **face** authentication, we evaluate our method on CMU Multi-PIE [44] and Yale Face database DB2 [45].

CMU Multi-PIE consists of 750,000 images of 337 candidates. The dataset is acquired over a span of 5 months in four different sessions. The dataset consists of images having 15 view points and 19 illumination conditions. It contains images with different poses, illuminations and expressions. We consider the frontal posed images with different expressions and illuminations to highlight the robustness of the algorithm. Indeed, as hinted above we assume to have controlled acquisitions which lead us to consider only the frontal pose. However, to keep high intra-class variability, we do not fix other sources of noise such as facial expressions and illumunation conditions.

For each user enrollment 75% of the samples are employed for the training and remaining 25% are left for testing. For unauthorized users, out of 128, 96 users samples are drawn for the training and remaining 32 users samples are left for testing. Further, train and test splits are made in such a way to avoid the sharing of the same facial expressions or illumination conditions and thus reducing the probability of overfitting.

Samples are resized to $144 \times 192 \times 3$ maintaining the aspect ratio. To create more diverse samples, positive and negative users samples are combined through a mixup strategy as discussed in Sec. IV-B.

The second dataset we employ is the cropped version of *extended Yale Face Database B*. It contains the frontal pose of 38 subjects with varying illumination conditions. For each authorized user enrollment 75% of the samples are drawn for training, and remaining 25% are left for testing. For unauthorized users, 31 users samples are used for training and 6 users samples are left for testing. Further, by employing crops of size 184×160, the samples are augmented as described in Sec. IV-B by an augmentation factor of $F = 81$ and $F_1 = 25$ respectively. As a last step, for each training batch of size b we randomly select b samples from both authorized and unauthorized users datasets. Then, positive and negative samples are combined through mixup as explained in Sec. IV-B resulting in b new samples.

The **fingerprint** authentication experiments are performed on *Fingerprint Verification Competition (FVC 2006) DB2* [46] dataset. Albeit old, this is still an actively used dataset [47], [48]. It consists of 150 users samples acquired through an optical sensor. Maintaining the aspect ratio, the samples are resized to 202×149. For each authorized user enrollment 75% of the samples are used for training and remaining 25% are left for testing. For the case of unauthorized users, 124 users samples are used for training and 25 users samples are left for testing. Finally the dataset is augmented by factors of $F = 289$ and $F_1 = 25$ using the crops of sizes 186×133 pixels and mixup augmentation as done for the faces dataset is employed.

B. EVALUATION METRICS

The main metric we will use in our experiments is Equal Error Rate (EER) defined as the value at which the False Acceptance Rate (FAR) equals the False Rejection Rate (FRR). Given a threshold τ, the FAR indicates the number of accepted samples that should have been rejected over the total number of samples. Conversely, the FRR indicates the number of rejected samples which should have been accepted over the total number of samples.

It is important to notice that in biometric authentication systems the FAR is a critical parameter: a large value indicates a high number of unauthorized users wrongly authorized by the system. This situation is indeed more dangerous with respect to having high false rejections of authorized users (large FRR). For good biometric systems minimum FAR is desired. For this reason we also test the systems at small values of FAR: we report the Genuine Acceptance Rate (GAR), namely the relative number of correctly accepted users at FAR equal to 10^{-2} and 10^{-3}. Finally, we report the maximum accuracy, defined as the value at which the number of correctly classified samples is maximized.

In the results section, the metrics are first computed independently for each of the considered users, reporting the resulting average values and their relative standard deviations. This will give insights on how the system performs, on average, on a per-user basis. Additionally, to gain a better understanding of the overall performance, we also report the aggregated results on all the users scores and illustrate the Receiver Operating Characteristic (ROC) curve computed on the aggregated scores of the considered users.

TABLE 1. GAR comparison of randomly selected users from Yale DB2 and CMU MultiPIE when considering different dimensionality of the latent space d. The best case is obtained for $d = 1$: highest GAR for a fixed FAR $= 10^{-3}$.

d	GAR@10^{-3}FAR% Yale face B	GAR@10^{-3}FAR% CMU Multi-PIE
1	100	100
3	97.663	99.88
64	91.001	96.53

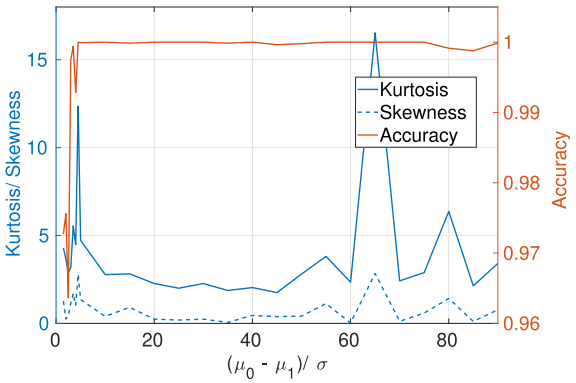


FIGURE 5. Accuracy, kurtosis and skewness comparison of a randomly selected user from CMU-MultiPIE having $\mathbb{P}_0 = \mathcal{N}(k, 1)$ where $k = [0.5, 90]$, and $\mathbb{P}_1 = \mathcal{N}(0, 1)$. If the means of the two distribution are too far apart the training process gets unstable, hence it effects the accuracy, kurtosis and skewness of the imposed distributions.

C. DIMENSIONALITY OF LATENT SPACE

An important parameter in the design of AuthNet is the choice of the latent space dimensionality d. The datasets we are considering are medium sized, thus it is not surprising that a smaller d achieves better results. In case of large datasets, a larger latent space improves the data separation and leads to improved performance.

In our tests, we fixed the hyperparameter $d = 1$, since in our experiments this choice gave us better results as can be seen in Tab. 1. Intuitively, as the latent space grows in dimensionality, a larger number of training samples are required to avoid overfitting. As an example MultiPIE has a relatively larger size compared to Yale dataset, it can be observed from Tab. 1 that for MultiPIE higher GAR is achieved at larger values of d compared to Yale dataset.

D. PARAMETERS OF AUTHORIZED AND UNAUTHORIZED USERS DISTRIBUTIONS

In AuthNet the authorized and unauthorized target distributions are set to be Gaussian. This choice comes from the fact that the output of a (large enough) fully connected layer, by the central limit theorem, will naturally tend to a Gaussian distributed output [49], [50]. We set the distributions to be $\mathbb{P}_1 = \mathcal{N}(0, 1)$ and $\mathbb{P}_0 = \mathcal{N}(40, 1)$. We choose $\mu_1 = 0$ and

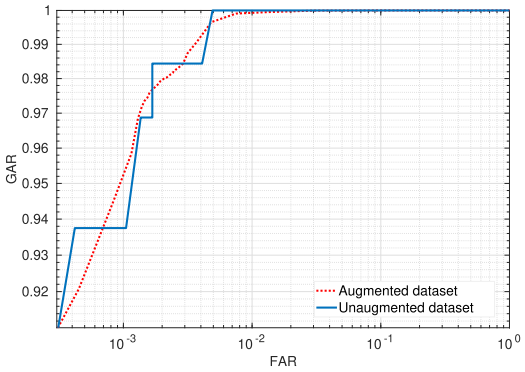


FIGURE 6. ROC comparison on overall results of the different users from FVC2006 DB2 for augmented and unaugmented datasets. The augmented dataset shows the same performance without quantization of probability values.

$\mu_0 = 40$ to be different enough to keep the distributions far apart from each other. Further, we set $\sigma_1 = \sigma_0 = 1$ as the choice lead to simple decision boundaries. As for the Gaussian discrimination problem if $\sigma_1 = \sigma_0$, then a linear decision boundary (hyperplane) is optimal. In more detail, in Fig. 5 we show the maximum accuracy obtained by AuthNet, together with skewness and kurtosis of the latent representation as a function of $(\mu_0 - \mu_1)/\sigma$ for a randomly selected CMU-MultiPIE user. It can be seen that the region for which the accuracy is maximum, corresponds roughly to $15 \leq (\mu_0 - \mu_1)/\sigma \leq 45$; in this region, skewness and kurtosis are close to 0 and 3 respectively, showing that the training indeed converges to Gaussian distributions. Further, if $(\mu_0 - \mu_1)/\sigma$ is too large, the training process becomes unstable and the distributions become far from Gaussian.

E. RESULTS

Before presenting the results it is important to consider that the precision of the performance metrics we consider is proportional to the number of test samples. The maximum precision which can be obtained for the considered metrics (explained in Sec. V-B) is given by $1/c$ with $c = \min\{L \times F, Q \times F_1\}$. Therefore, we will verify that the proposed augmentation strategy does not introduce any bias on the measured performance. As can be seen in Fig. 6, augmentation avoids coarse quantization of probability values without introducing any bias. For this reason, the metrics we will consider from now on will be computed on the augmented dataset.

For our results, in addition to biometric-related methods, we also include the comparison with the Encoder network of AuthNet-R used as a classifier and trained with sigmoid cross entropy loss. In Sec. I we discussed the issue of classifiers having highly non-linear and complex to analyze boundaries. Therefore, we evaluate the behavior of a deep learning classifier based on the same architecture as the AuthNet-R encoder but which is not trained in an adversarial way, in order to assess the benefits of the adversarial scheme employed in AuthNet.

Tab. 2 presents the results achieved by AuthNet and benchmarking methods in terms of EER, GAR values at FAR $= \{10^{-2}, 10^{-3}\}$ and maximum accuracies on the individual and aggregated scores. Fig. 7-9 depict the histogram of the aggregated scores obtained by different methods. The ROC comparison for different benchmarking methods is depicted in Fig. 10. Lastly, for the sake of readability, unless differently specified from now on we will refer to both AuthNet-R and AuthNet-D as "AuthNet".

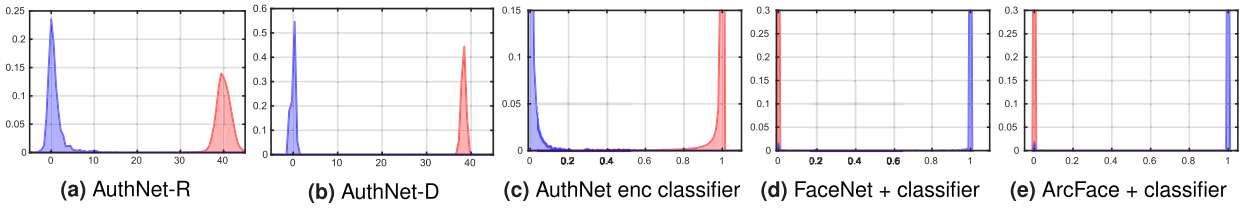


FIGURE 7. MultiPie authentication scores for authorized users (blue) and unauthorized users (red). (a), (b) Histogram of z decision statistics of AuthNet; (c) Histogram of the sigmoid outputs of AuthNet encoder classifier; (d) Histogram of the sigmoid outputs of FaceNet embeddings classifier; (e) Histogram of the sigmoid outputs of ArcFace embeddings classifier. The plots in (c)-(e) depict a detailed view to better appreciate the leakage effects.

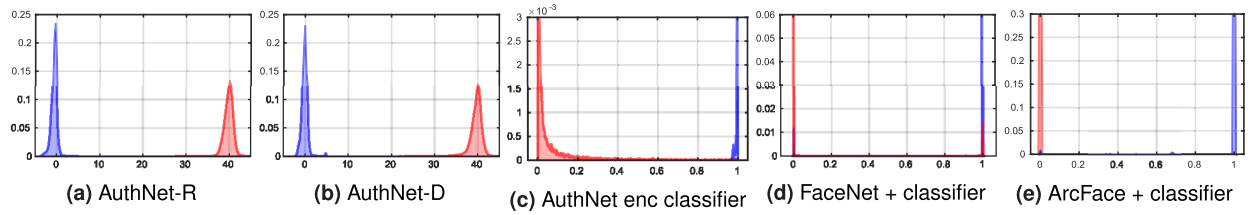


FIGURE 8. Face Yale authentication scores for authorized users (blue) and unauthorized users (red). (a), (b) Histogram of z decision statistics of AuthNet; (c) Histogram of the sigmoid outputs of AuthNet encoder classifier; (d) Histogram of the sigmoid outputs of FaceNet embeddings classifier; (e) Histogram of the sigmoid outputs of ArcFace embeddings classifier. The plots in (c)-(e) depict a detailed view to better appreciate the leakage effects.

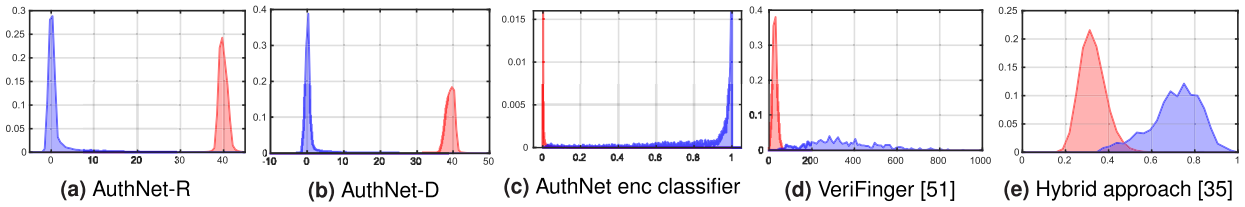


FIGURE 9. Fingerprint authentication scores for authorized users (blue) and unauthorized users (red). (a), (b) Histogram of z decision statistics for AuthNet; (c) Histogram of the sigmoid outputs of the AuthNet encoder classifier; (d) histogram of the matching scores of Verifinger; (e) histogram of the matching scores of the hybrid approach. The plot in (c) depicts a detailed view to better appreciate the leakage effects.

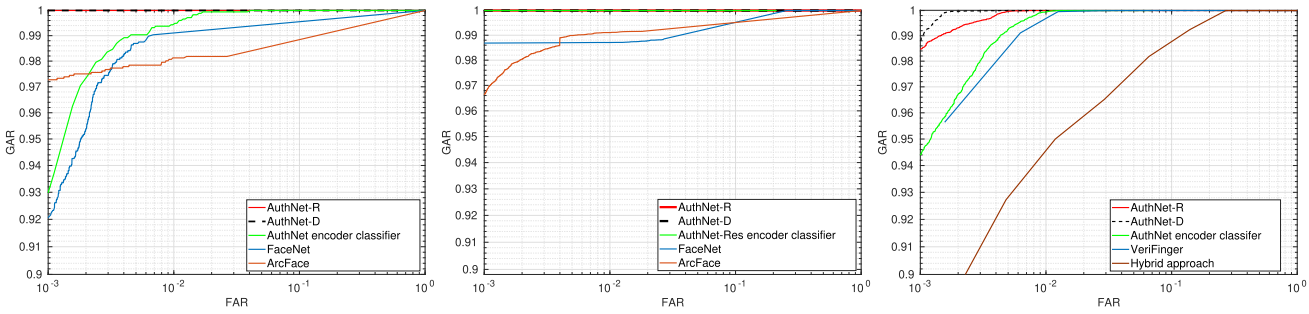


FIGURE 10. ROC comparison on aggregated results of users for faces (a) CMU Multi-PIE, (b) Face Yale database B, - fingerprint (c) FVC 2006 DB2 datasets. (a-b); AuthNet is compared with the AuthNet encoder classifier, FaceNet [19] and ArcFace [20] in (c); with AuthNet encoder classifier, VeriFinger [51] and the hybrid approach [35] in (b). In all the cases, AuthNet (red) and (black) achieves higher GAR with respect to other authentication schemes at different values of FAR.

TABLE 2. Performance comparison of AuthNet with respect to other face authentication schemes. Average values of the considered metrics computed independently on each user and on the aggregated scores (shown in parenthesis) are reported. We mark as 0* and 100* values below the minimum achievable precision, i.e. smaller than $4.1 \cdot 10{-}5$ and $5.68 \cdot 10{-}4$ for Yale and MultiPIE datasets respectively.

Dataset	Method	EER%	GAR@10^{-2}FAR%	GAR@10^{-3}FAR%	Max accuracy
Face-Yale	**AuthNet-D**	**0* ± 0 (0.009)**	**100* ± 0 (100*)**	**100* ± 0 (100*)**	**100* ± 0 (99.991)**
	AuthNet-R	**0.003 ± 0.011 (0.019)**	**100* ± 0 (100*)**	**99.687 ± 1.767 (100*)**	**99.997 ± 0.011 (99.983)**
	AuthNet enc. classifier	0.013 ± 0.054 (0.040)	100* ± 0 (100*)	99.011 ± 0.031 (100*)	99.987± 0.054 (99.961)
	FaceNet	1.258 ± 2.084 (1.288)	98.793 ± 3.114 (98.707)	98.781 ± 3.150 (98.683)	98.825 ± 1.869 (99.300)
	ArcFace	0.696 ± 1.533 (0.893)	99.024 ± 2.798 (99.108)	97.762 ± 4.063 (96.630)	99.367 ± 1.434 (99.229)
Face Multi-Pie	**AuthNet-D**	**0* ± 0 (0.005)**	**100* ± 0 (100*)**	**100* ± 0 (100*)**	**100*± 0 (99.993)**
	AuthNet-R	**0* ± 0 (0.001)**	**100* ± 0 (100*)**	**100* ± 0 (100*)**	**100*± 0 (99.998)**
	AuthNet enc. classifier	0.009 ± 0.034 (0.676)	100* ± 0 (99.432)	99.886 ± 0.447 (92.897)	99.991 ± 0.034 (99.325)
	FaceNet	0.770 ± 1.080 (0.930)	98.466 ± 3.490 (99.201)	90.513 ± 21.582 (92.045)	99.368± 0.981 (99.197)
	ArcFace	1.727 ± 0.164 (1.811)	98.124 ± 0.321 (98.125)	97.897 ± 0.814 (97.272)	99.058 ± 0.103 (98.871)

1) FACE AUTHENTICATION

The datasets we employ for face authentication are CMU Multi-PIE and Yale Face database B, as detailed in Sec. V-A. For benchmarking with state-of-the-art deep learning techniques, we compare with ArcFace [20] and FaceNet [19]. FaceNet and ArcFace tend to work better on aligned face patches. For CMU-Multi-PIE, we pre-process the dataset by aligning and cropping the input faces using the well-known approach of joint face detection and alignment using Multi-task Cascaded Convolutional networks (MTCNN) [52]. Yale Face database already consists of frontal face images of the subjects, so face alignment and crop are not needed.

Regarding the training process of Facenet and ArcFace, we employ the standard architecture as described in their respective papers using 512-dimensional embeddings. Since the above methods are meant to learn a generic face embedding to be used for either face recognition, verification, or clustering, they 1) require a very large training dataset, and 2) cannot learn a user-specific embedding. This will result in an unfair comparison with AuthNet. To alleviate this issue and make the comparison fair, we follow a two-step approach. At first we train FaceNet and Arcface on the large CASIA WebFace dataset [53] in such a way that we can obtain 512 dimensional embeddings from given input face images. Then, given the embeddings, we train two-class FC classfiers (one for each user) which have to classify the embeddings as either authorized or unauthorized.

Tab. 2 presents a comparison of EER, GAR at FAR $= \{10^{-2}, 10^{-3}\}$, and maximum accuracy for CMU Multi-PIE and Yale Face Database B, calculated on the individual and

TABLE 3. Performance comparison of AuthNet with respect to other fingerprint authentication schemes on FVC 2006 DB2 dataset. Average values of the considered metrics computed independently on each user and on the aggregated scores (shown in parenthesis) are reported. We mark as 0* and 100* values below the minimum achievable precision, i.e. smaller than $5.5 \cdot 10{-}5$.

Dataset	Method	EER%	GAR@10^{-2}FAR%	GAR@10^{-3}FAR%	Max accuracy
Fingerprint	**AuthNet-D**	**0* ± 0 (0.147)**	**100* ± 0 (100*)**	**100* ± 0 (98.817)**	**100*± 0 (99.895)**
	AuthNet-R	**0.058 ± 0.217 (0.339)**	**99.955 ± 0.248 (100*)**	**99.400 ± 2.949 (98.476)**	**99.957± 0.173 (99.740)**
	AuthNet enc. classifier	0.188 ± 0.722 (0.565)	98.448 ± 8.621 (99.845)	95.148 ± 9.742 (94.384)	99.812 ± 0.722 (99.435)
	Verifinger	0.163 ± 0.697 (0.758)	99.680 ± 1.229 (99.796)	99.375 ± 2.459 (95.638)	99.902 ± 0.373 (99.398)
	Hybrid approach [35]	1.515 ± 2.651 (3.200)	95.937 ± 7.452 (95.000)	90.001 ± 19.261 (85.909)	98.868 ± 2.178 (96.906)

the aggregated scores of the users. From the results it can be observed that, in terms of EER, AuthNet achieves the lowest value outperforming other methods. Further, a very small advantage of AuthNet-R with respect to AuthNet-D can also be observed. Nevertheless, as shown in later experiments the performance of the these two AuthNet flavors is comparable and a clear winner cannot be identified.

It is also interesting to observe that for the AuthNet, even for very small values of FAR, high GAR values are obtained. The high performance for Multi-PIE compared to Yale Face database B is understandable since the former has a significantly larger number of high-quality samples per user compared to other datasets. Further, AuthNet outperforms the competing methods in terms of maximum accuracy achieved. It can be observed that for AuthNet encoder classifier the performance in terms of EER is an order of magnitude less than that of AuthNet. In more detail, we can exclude that this is due to AuthNet encoder classifier overfitting on the negative samples. Indeed, this case be seen by looking at Fig. 13 where it is depicted the ROC for the considered approaches when tested on out-of-domain or never-seen negative examples. It can be noticed that the performance drop of AuthNet encoder classifier is mostly bounded, and thus the poorer performance is due to the lack of regularization of the decision space. Indeed, the results of this comparison imply that by regularizing the latent space through well-behaved distributions, it is possible to increase the accuracy of the system by decreasing the number of false positives. This highlights the superiority of the proposed latent space regularization over a traditional classifier. Additionally, the achieved EER by FaceNet and ArcFace is also an order of magnitude less than that of AuthNet. Furthermore, for small values of FAR, the genuine acceptance for these methods significantly reduces, which is not the case with AuthNet. This indicates a high variablity of the results on a per-user basis, which can be observed from both individual and aggregated user scores in Tab. 2.

Furthermore, to better appreciate the effective regularization of the latent space of AuthNet, in Fig. 7 and 8 the face authentication scores for authorized and unauthorized users are depicted for different benchmarking algorithms. The blue curve in the figure depicts the histogram of the score obtained for the authorized users, and the red curve depicts the histogram of the scores obtained for unauthorized users. The histogram of the $\mathbf{z}$ scores obtained from AuthNet-R and AuthNet-D are depicted in Fig. 7a, 7b for Multi-PIE and Fig. 8a, 8b for Yale Face database B respectively. It can be observed that for both datasets, AuthNet very effectively separates authorized and unauthorized users samples and there is no mixing of authorized and unauthorized users distributions. The scores of the sigmoid output obtained from the AuthNet encoder classifiers are depicted in Fig. 7c and 8c. It can be observed that, being the output a sigmoid activation, the distributions are mainly peaked at 0 and 1; however there is noticeable spillover in the area in between. This is the reason for lower EER and GAR at small values of FAR. The histogram of the sigmoid output obtained from FaceNet and ArcFace embeddings classifiers is depicted in Fig. 7d, 7e for Multi-PIE and Fig. 8d, 8e for Yale Face database B, respectively. In both cases it is possible to appreciate a non-perfect separation of the scores: these misclassified users eventually lead to lower performance.

Lastly, Fig. 10a and 10b illustrates the ROC comparison of AuthNet with respect to other benchmark techniques on the aggregated scores of the users. It can be clearly observed that the ROC curves for AuthNet lie above all other methods and consistently achieve higher GAR even at very low values of FAR proving its superiority.

2) FINGERPRINT AUTHENTICATION

For the fingerprints, we employ the FVC 2006 DB2 dataset, detailed in Sec. V-A.

For benchmarking, we compare AuthNet with AuthNet encoder classifier, Verifinger [51] and the hybrid approach described in [35]. Verifinger is a well-known and commercially available system commonly used for minutiae extraction and fingerprint matching achieving state-of-the-art performance in fingerprint identification [54].

Tab. 3 depicts the comparison of EER, maximum accuracy, and GAR of AuthNet at small values of FAR with the benchmarking methods. From Tab. 3, it can be observed that AuthNet achieves the lowest EER and highest accuracy, outperforming all benchmark methods. However, differently from the previous results, it can be observed that AuthNet-D has a slight performance advantage over AuthNet-R. In general it is difficult to state which of the two AuthNet flavors achieves higher performance. Indeed, the performance of AuthNet is to some extent independent of the encoder network architecture. As long as the encoder network has

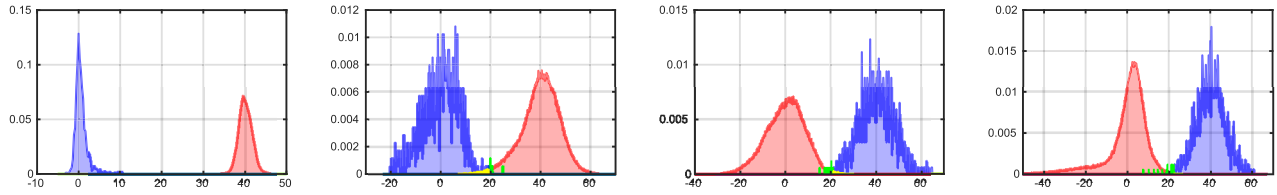


FIGURE 11. Normalized logits scores for correctly accepted authorized users (blue), wrongly rejected authorized users (green), correctly rejected unauthorized users (red), and wrongly accepted unauthorized users (yellow). (a), (b) Histogram of z decision statistics of AuthNet; (c) Histogram of the logits scores of AuthNet encoder classifier; (d) Histogram of the logits scores of FaceNet embeddings classifier; (e) Histogram of the logits scores of ArcFace embeddings classifier.

enough capacity, any recent CNN architecture will be able to reach, on average, high performance.

Additionally, for small values of FAR, both AuthNet-R and AuthNet-D achieve high values of GAR. Verifinger, AuthNet encoder classifier and hybrid approach, also achieve small EER values; however, it can be observed that the GAR values significantly drop as the FAR values are decreased, which is not the case with AuthNet.

Further, it can be seen from Fig. 9a and 9b that the proposed method separates the authorized and unauthorized users very effectively. Conversely, in the case of non-deep learning approaches such as Verifinger in Fig. 9d, and the hybrid approach in Fig. 9e, the authorized and unauthorized users do not have a clear scores separation and the related regions are not well-behaved. Moreover, it can be noticed in Fig. 9c that similarly to the case of face datasets, while AuthNet encoder classifier provides a separation between the scores it also introduces some ''leakage''.

Lastly, in Fig. 10c the ROC comparison of AuthNet with respect to other fingerprint authentication schemes is depicted. The red curve depicts the GAR at different FAR values obtained by AuthNet. It can be seen that AuthNet ROC curve lies above other benchmarking methods. Furthermore, it can be clearly observed here that at small values of FARs, AuthNet clearly outperforms all the other competing algorithms, maintaining highest GAR values.

VI. IN-DEPTH ANALYSIS OF AuthNet

In order to better understand the performance improvement of AuthNet with respect to competing methods, a deeper technical insight is provided with the purpose of explaining how the regularization of the distributions performed by AuthNet yields fewer misclassifications compared to existing methods. Further, it is shown how Authnet is able to correctly classify samples that are misclassified by competing approaches.

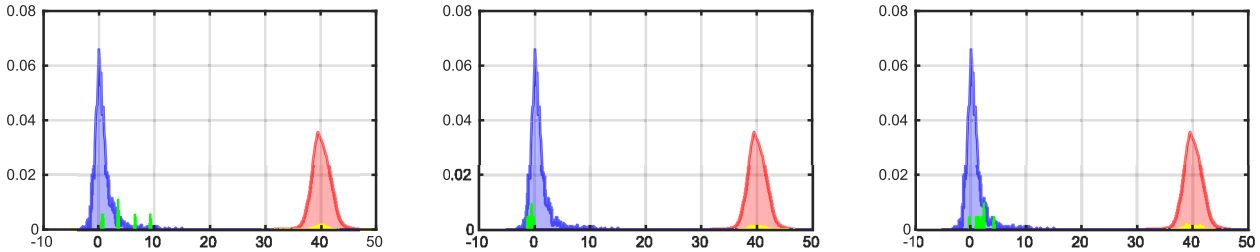
A. MOTIVATION BEHIND HIGHER MISCLASSIFICATION RATE BY COMPETING METHODS

In the first set of experiments, shown in Fig. 11 the latent space outputs of AuthNet and the logit scores obtained by the competing methods, normalized to the target means of $\mu = 0$ for the authorized users and $\mu = 40$ for unauthorized users are presented; this normalization allows us to directly compare these methods with AuthNet. It can be observed that the logit scores of the other methods naturally tends to be Gaussian, from the central limit theorem [49], [50]. During AuthNet training, the target distributions are enforced to follow Gaussian distributions that are well separated, with predefined mean and standard deviation. However, for traditional classification methods this is not specifically enforced which results in distributions with unpredictable mean and standard deviation. As a result, it can be observed in Fig. 11 that the normalized logit score distributions of the competing methods exhibit higher variance with heavier tails, compared to that of AuthNet which instead obtains distributions which are well-separated in the latent space. Moreover, in Fig. 11 normalized logit scores for correctly accepted authorized users (blue), wrongly rejected authorized users (green), correctly rejected unauthorized users (red), and wrongly accepted unauthorized users (yellow) are highlighted. It can be clearly observed that for AuthNet the authorized and unauthorized users scores are well separated based on the predefined target distributions, yielding very few misclassifications, i.e. false rejections of authorized users (green) and false acceptance of unauthorized users (yellow) area. On the other side, in the competing methods, the logit scores distributions of the authorized and unauthorized users are broader, which results in a higher number of misclassifications as can be observed from the green and yellow areas.

Tab. 4 reports the standard deviation σ and kurtosis β_2 of the latent space features of AuthNet and the normalized logit scores obtained by different methods. It can be observed from the table that the lack of regularization of the distributions in the competing methods tends to have much higher σ. Similarly, the distributions obtained by the competing methods are heavy-tailed as can be seen from the measured values of β_2. This points out a higher spread of the authorized and unauthorized user distributions with respect to the mass center, resulting in a higher number of misclassifications.

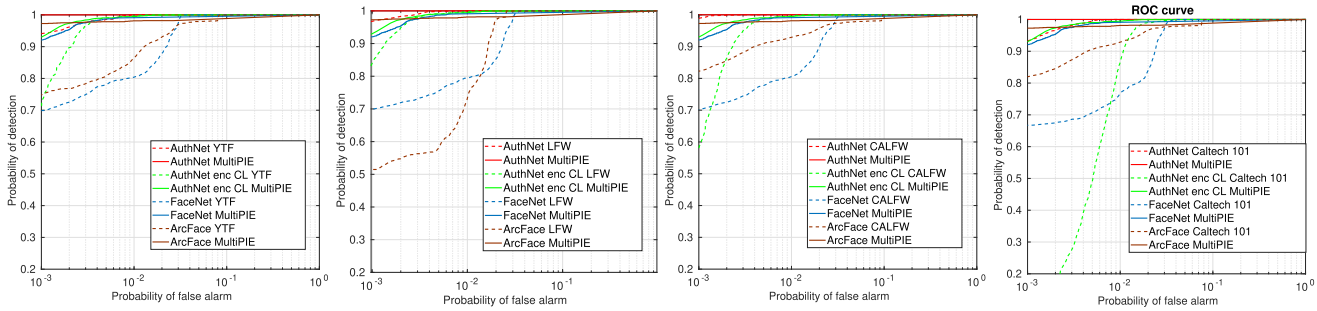
B. HOW AuthNet CORRECTLY MAPS USERS THAT ARE MISCLASSIFIED BY OTHER METHODS

In Fig. 13, depicting latent features obtained by Authnet, the latent feature outputs corresponding to authorized users that are wrongly rejected by competing networks are highlighted



(a) AuthNet enc classifier (b) FaceNet + classifier (c) ArcFace + classifier

FIGURE 12. Correct mapping of the misclassified users by other methods using AuthNet: mapping of the wrongly rejected authorized users (green) and wrongly accepted unauthorized users (yellow) by competing methods on AuthNet, (a) misclassified users of AuthNet encoder classifier mapped on AuthNet; (b) misclassified users of FaceNet embeddings classifier mapped on AuthNet; (c) misclassified users of ArcFace embeddings classifier mapped on AuthNet.



(a) YTF dataset (b) LFW dataset (c) CALFW dataset (d) Caltech 101 dataset

FIGURE 13. ROC comparison of AuthNet and benchmarking methods when tested on the same dataset used during training (MultiPIE) and on face (YTF, LFW, CALFW) and non-face (Caltech 101) datasets that have not been used during training. In all the cases, AuthNet performs consistently and give stable GAR at different FAR values.

TABLE 4. Standard deviation σ and Kurtosis β_2 of normalized test logit scores for authorized and unauthorized users.

method	$\sigma_{a=1}$	$\sigma_{a=0}$	$\beta_{2,a=1}$	$\beta_{2,a=0}$
AuthNet	1.108	1.451	3.369	3.282
AuthNet enc CL	7.176	7.739	3.985	3.809
FaceNet	7.446	8.517	3.843	3.918
ArcFace	6.919	10.183	3.514	4.215

in green, whereas features corresponding to wrongly accepted unauthorized users are highlighted in yellow. It can be observed that in all the cases AuthNet maps the wrongly accepted unauthorized users near the mass center of correctly rejected unauthorized users i.e. the red area. Similarly, AuthNet properly maps the wrongly rejected authorized users in the right class in the blue area.

In summary defining well separated target Gaussian distributions having specified mean and standard deviation during training avoids spread of the authorized and unauthorized users samples yielding a lower number of misclassifications.

VII. ROBUSTNESS ANALYSIS

In this second set of experiments, we show that regularizing the latent space to simple target distributions leads not only to improved accuracy, but also to more robust authentication. In particular, we test AuthNet and the benchmark methods trained on MultiPIE on datasets that the network has not seen during the training. Further, we also test the robustness of the proposed approach against targeted perturbations.

A. EVALUATION ON NEW DATASETS NOT SEEN DURING TRAINING

To show the robustness and resilience of AuthNet against the *face* datasets that the network has never seen during training, we test AuthNet-R and competing methods trained on MultiPIE on LFW [55], YTF [56], and CALFW [57] datasets. Fig. 13 shows the ROC comparison of methods trained and tested on MultiPIE versus the same methods trained on MultiPIE and tested on YTF, LFW, and CALFW datasets, for the class of unauthorized users (note that in this setup the unauthorized users are not present in the test dataset). The solid curves depict the results when methods are trained and tested on the same dataset, the dotted curves

TABLE 5. Absolute performance drop comparison of AuthNet and benchmarking methods when trained on MultiPIE and tested on different datasets. We mark as 0* values below the minimum achievable precision, i.e. smaller than $5.6 \cdot 10{-}4$.

Dataset	Method	ΔGAR@10^{-1}FAR%	ΔGAR@10^{-2}FAR%	ΔGAR@10^{-3}FAR%	ΔMax accuracy
YTF	**AuthNet-R**	**0***	**0.056**	**5.852**	**0.487**
	AuthNet enc. classifier	0*	0.227	22.273	0.658
	FaceNet	0.632	18.309	22.038	0.897
	ArcFace	1.132	11.989	22.501	1.842
LFW	**AuthNet-R**	**0***	**0***	**3.068**	**0.309**
	AuthNet enc. classifier	0*	0.283	9.773	0.621
	FaceNet	0.611	19.762	22.102	0.960
	ArcFace	1.189	24.829	45.965	0.387
CALFW	**AuthNet-R**	**0***	**0***	**1.080**	**0.136**
	AuthNet enc. classifier	0*	0.227	35.568	0.648
	FaceNet	0.622	18.521	21.725	1.100
	ArcFace	1.330	8.057	15.001	1.770
Caltech 101	**AuthNet-R**	**0***	**0***	**6.762**	**0.381**
	AuthNet enc. classifier	0*	13.068	88.470	1.468
	FaceNet	0.609	22.497	25.341	1.190
	ArcFace	1.130	5.398	15.342	1.421

depict the test results on the datasets which the network has not seen during training. The robustness is measured in terms of the performance drop on the datasets that have not been seen during the training. It can be observed that AuthNet is robust against the datasets which were not presented at training time: it correctly maps the samples from these datasets to the unauthorized target distribution. This effect is more significant at small FAR (10^{-3}) where a large performance drop can be observed for the competing methods, whereas AuthNet maintains high GAR value, outperforming them by a big margin.

For a more detailed analysis, Tab. 5 reports the absolute difference in GAR at different values of FAR and the maximum accuracy difference achieved by different methods when tested on MultiPIE versus the other datasets. It can be seen that AuthNet consistently outperforms all competing methods, yielding a very small performance drop when tested on different datasets. The effect is very evident at small values of FAR.

To further evaluate the robustness of AuthNet, we also considered a non-face dataset: we test AuthNet and competing methods trained on MultiPIE on Caltech 101 [58] dataset. This dataset does not include faces and it is made of images of objects belonging to 101 different categories. From both Fig. 13d and Tab. 5 it can be observed that the performance drop is very significant for the competing methods. Conversely, AuthNet still maps the images of Caltech 101 to the unauthorized distribution giving stable results even at small FAR values.

The results in this section show that regularizing the latent space using well-behaved target distributions leads to robust authentication against features that have never been seen before. Furthermore, the behavior of the non-authorized region of AuthNet is consistent across different datasets.

B. EVALUATION ON TARGETED PERTURBATIONS

We further analyse the robustness of the AuthNet approach against the targeted perturbations. We consider white-box Fast Gradient Sign Method (FGSM) [59] due its simplicity and speed in crafting the perturbations. In FGSM the input samples are adjusted to maximize the loss based on the back propagated gradients. The model back propagates to the input data to calculate $\nabla_x J(\theta, \mathbf{x}, a)$, then the input samples are adjusted by a step of ϵ in the direction of $sign(\nabla_x J(\theta, \mathbf{x}, a))$ that will maximize the loss.

For this experiment, we compare AuthNet-R with the AuthNet encoder classifier trained on Multi-PIE in order to highlight the advantages of learning the mapping instead of the boundaries. The rationale is to show that for traditional methods producing arbitrary boundaries, it is usually possible to craft samples that result in incorrect classification with a minimal perturbation, whereas for the proposed method this is much more difficult, leading to improved robustness.

For both AuthNet and AuthNet encoder classifier, every test sample is perturbed with ℓ_∞ bounded perturbation and the results are aggregated. We define $\mathbf{n}$ as the noise vector such that $\|(\mathbf{n})\|_\infty \leq \epsilon$ where noise strength ϵ is defined as the ratio $\|(\mathbf{n})\|_\infty / \|(\mathbf{x})\|_\infty$. As an example 100% noise strength means the model is able to corrupt the image with noise values within the full range of the input image.

In Fig. 14 we depict the probability of success of FGSM as a function of the noise strength. For AuthNet it can be noticed that trying to move the authorized users into the unauthorized region ($\mathbf{z} < 20$) has a high probability of success for large noise strength, i.e. larger than 10% of the maximum pixel values of the input images. However, by lowering ϵ, the probability of success decreases accordingly. Conversely, granting access to unauthorized users is a much harder task. The maximum probability of success, reached at 2% noise strength, is 0.27. Furthermore, the probability of success in such setting is close to zero even for very large perturbations. This can be explained by the way AuthNet regularizes the latent space: authorized users are strictly enclosed within the high mass region of $\mathbb{P}_1$. If the perturbation is too strong, the likelihood that the perturbed samples are treated as unauthorized users increases. We further study this effect in Fig. 15 where we

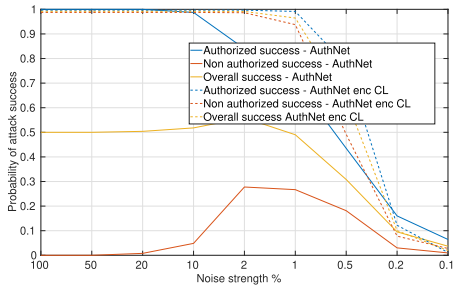


FIGURE 14. Probability of success of FGSM for authorized, unauthorized users and overall success as a function of the noise strength for AuthNet-R and AuthNet-R encoder classifier.

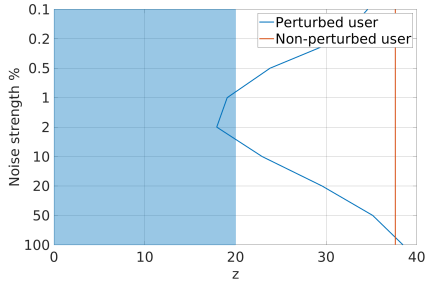


FIGURE 15. Trajectory of decision statistics for a perturbed sample (from an unauthorized user) in the latent space at different noise strength levels.

show the trajectory of $\mathbf{z}$ in the latent space: for a perturbed sample coming from an unauthorized user as a function of the noise level. It can be seen that for large perturbations, $\mathbf{z}$ stays within the high mass region of $\mathbb{P}_0$. Similarly, if ϵ is limited to less than 1%, the value of $\mathbf{z}$ remains close to 40. Between these limits we have a region which may lead to misclassification of unauthorized users. An interpretation of this behavior is that the regularized decision boundary provided by AuthNet does not allow to choose an easy path for crossing the boundary from a generic point within the decision region, i.e., every point on the other side of the boundary tends to be equally far away. If we compare these results with those of the AuthNet encoder classifier in Fig. 14, it is immediate to notice that overall FGSM is much more successful, especially for large noise strength. Also in this case FGSM targeting authorized users is more successful. This confirms our conjecture that the highly complex boundaries learned through a classifier are more vulnerable to adversarial perturbations. Conversely, the proposed AuthNet architecture, by properly regularizing the latent space is able to greatly reduce such effects and thus reduce the likelihod of targeted perturbations to succeed.

VIII. CONCLUSION

We presented a novel approach for biometric authentication based on adversarial learning in which the latent space regularization leads to improved robustness and accuracy of the biometric classification. Our intuition behind this behavior is that the non-linear boundaries learned by standard deep learning classifiers indeed become very complex as they try to closely fit the training data, leaving room for misclassification. Conversely, the adversarial learning of AuthNet enables much simpler boundaries to be used as it does not learn how to partition the space but rather how to map the input space into the latent space. With extensive experimentation, on multiple large biometric datasets with several state-of-the-art benchmark methods, we showed that AuthNet consistently outperforms other existing techniques. We further show that regularizing the latent space makes the architecture less vulnerable to targeted and non targeted perturbations.

Future work will consider adding new users to a pre-trained AuthNet and to handle user revocation.

REFERENCES

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[2] A. Almahairi, S. Rajeswar, A. Sordoni, P. Bachman, and A. Courville, "Augmented CycleGAN: Learning many-to-many mappings from unpaired data," 2018, *arXiv:1802.10151*. [Online]. Available: http://arxiv.org/abs/1802.10151

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[4] A. Fawzi, S.-M. Moosavi-Dezfooli, P. Frossard, and S. Soatto, "Classification regions of deep neural networks," 2017, *arXiv:1705.09552*. [Online]. Available: http://arxiv.org/abs/1705.09552

[5] A. Fawzi, S.-M. Moosavi-Dezfooli, and P. Frossard, "The robustness of deep networks: A geometrical perspective," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 50–62, Nov. 2017.

[6] A. Ali, M. Testa, T. Bianchi, and E. Magli, "Authnet: Biometric authentication through adversarial learning," in *Proc. IEEE 29th Int. Workshop Mach. Learn. Signal Process. (MLSP)*, Oct. 2019, p. 1–6.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[8] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[9] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognit. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.

[10] D. L. Swets and J. J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 831–836, Aug. 1996.

[11] W. Deng, J. Hu, and J. Guo, "Extended SRC: Undersampled face recognition via intraclass variant dictionary," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1864–1870, Sep. 2012.

[12] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[13] B. Moghaddam, W. Wahid, and A. Pentland, "Beyond eigenfaces: Probabilistic matching for face recognition," in *Proc. 3rd IEEE Int. Conf. Autom. Face Gesture Recognit.*, Apr. 1998, p. 30.

[14] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, 1997.

[15] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, Mar. 2005.

[16] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 467–476, Apr. 2002.

[17] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[18] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1701–1708.

[19] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 815–823.
[20] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," 2018, *arXiv:1801.07698*. [Online]. Available: http://arxiv.org/abs/1801.07698
[21] S. Sankaranarayanan, A. Alavi, C. D. Castillo, and R. Chellappa, "Triplet probabilistic embedding for face verification and clustering," in *Proc. IEEE 8th Int. Conf. Biometrics Theory, Appl. Syst. (BTAS)*, Sep. 2016, pp. 1–8.
[22] H. Xu, J. Zheng, A. Alavi, and R. Chellappa, "Template regularized sparse coding for face verification," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 1448–1454.
[23] Y. Sun, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2892–2900.
[24] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *Proc. Eur. Conf. Comput. Vis.* Amsterdam, The Netherlands: Springer, 2016, pp. 499–515.
[25] R. Ranjan, A. Bansal, J. Zheng, H. Xu, J. Gleason, B. Lu, A. Nanduri, J.-C. Chen, C. Castillo, and R. Chellappa, "A fast and accurate system for face detection, identification, and verification," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 1, no. 2, pp. 82–96, Apr. 2019.
[26] R. K. Pandey, Y. Zhou, B. U. Kota, and V. Govindaraju, "Deep secure encoding for face template protection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2016, pp. 77–83.
[27] A. K. Jindal, S. Chalamala, and S. K. Jami, "Face template protection using deep convolutional neural network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 5755–5758.
[28] C. Liu, T. Xia, and H. Li, "A hierarchical hough transform for fingerprint matching," in *Proc. Int. Conf. Biometric Authentication*. Berlin, Germany: Springer, 2004, pp. 373–379.
[29] N. K. Ratha, K. Karu, S. Chen, and A. K. Jain, "A real-time matching system for large fingerprint databases," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 799–813, Aug. 1996.
[30] Y. He, J. Tian, X. Luo, and T. Zhang, "Image enhancement and minutiae matching in fingerprint verification," *Pattern Recognit. Lett.*, vol. 24, nos. 9–10, pp. 1349–1360, 2003.
[31] L. Sha and X. Tang, "Orientation-improved minutiae for fingerprint matching," in *Proc. 17th Int. Conf. Pattern Recognit. (ICPR)*, vol. 4, Aug. 2004, pp. 432–435.
[32] D. Lee, K. Choi, and J. Kim, "A robust fingerprint matching algorithm using local alignment," in *Proc. Object Recognit. Supported User Interact. Service Robots*, vol. 3, Aug. 2002, pp. 803–806.
[33] J. Cha, H. Jang, G. Kim, and H. Choi, "Fingerprint matching based on linking information structure of minutiae," in *Proc. Int. Conf. Comput. Sci. Appl.* Berlin, Germany: Springer, 2004, pp. 41–48.
[34] K. D. Yu, S. Na, and T. Y. Choi, "A fingerprint matching algorithm based on radial structure and a structure-rewarding scoring strategy," in *Proc. Int. Conf. Audio Video-Based Biometric Person Authentication*. Berlin, Germany: Springer, 2005, pp. 656–664.
[35] J. Abraham, P. Kwan, and J. Gao, "Fingerprint matching using a hybrid shape and orientation descriptor," in *State of the Art in Biometrics*. Rijeka, Croatia: InTech, 2011.
[36] L. Jiang, T. Zhao, C. Bai, A. Yong, and M. Wu, "A direct fingerprint minutiae extraction approach based on convolutional neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 571–578.
[37] L. N. Darlow and B. Rosman, "Fingerprint minutiae extraction using deep learning," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Oct. 2017, pp. 22–30.
[38] R. Wang, C. Han, and T. Guo, "A novel fingerprint classification method based on deep learning," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 931–936.
[39] D. Maio and D. Maltoni, "Neural network based minutiae filtering in fingerprints," in *Proc. 14th Int. Conf. Pattern Recognit.*, vol. 2, Aug. 1998, pp. 1654–1658.
[40] W. F. Leung, S. H. Leung, W. H. Lau, and A. Luk, "Fingerprint recognition using neural network," in *Proc. Neural Netw. Signal Process.*, Sep./Oct. 1991, pp. 226–235.
[41] Y. Tang, F. Gao, and J. Feng, "Latent fingerprint minutia extraction using fully convolutional network," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Oct. 2017, pp. 117–123.
[42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980
[43] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: Beyond empirical risk minimization," 2017, *arXiv:1710.09412*. [Online]. Available: http://arxiv.org/abs/1710.09412
[44] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image Vis. Comput.*, vol. 28, no. 5, pp. 807–813, May 2010.
[45] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.
[46] R. Cappelli, M. Ferrara, A. Franco, and D. Maltoni, "Fingerprint verification competition 2006," *Biometric Technol. Today*, vol. 15, nos. 7–8, pp. 7–9, Jul. 2007.
[47] M. A. Alsmirat, F. Al-Alem, M. Al-Ayyoub, Y. Jararweh, and B. Gupta, "Impact of digital fingerprint image quality on the fingerprint recognition accuracy," *Multimedia Tools Appl.*, vol. 78, no. 3, pp. 3649–3688, Feb. 2019.
[48] M. Sabri, M.-S. Moin, and F. Razzazi, "A new framework for match on card and match on host quality based multimodal biometric authentication," *J. Signal Process. Syst.*, vol. 91, no. 2, pp. 163–177, Feb. 2019.
[49] R. M. Neal, *Bayesian Learning for Neural Networks*, vol. 118. New York, NY, USA: Springer, 2012.
[50] A. Borovykh, "A Gaussian process perspective on convolutional neural networks," 2018, *arXiv:1810.10798*. [Online]. Available: http://arxiv.org/abs/1810.10798
[51] "Neuro technology (2010)," VeriFinger, SDK Neuro Technol., Tech. Rep.
[52] J. Xiang and G. Zhu, "Joint face detection and facial expression recognition with MTCNN," in *Proc. 4th Int. Conf. Inf. Sci. Control Eng. (ICISCE)*, Jul. 2017, pp. 424–427.
[53] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," 2014, *arXiv:1411.7923*. [Online]. Available: http://arxiv.org/abs/1411.7923
[54] H. AlShehri, M. Hussain, H. AboAlSamh, and M. AlZuair, "A large-scale study of fingerprint matching systems for sensor interoperability problem," *Sensors*, vol. 18, no. 4, p. 1008, Mar. 2018.
[55] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments," in *Proc. Workshop Faces Real-Life Images, Detection, Alignment, Recognit.*, 2008, pp. 1–15.
[56] L. Wolf, T. Hassner, and I. Maoz, *Face Recognition in Unconstrained Videos With Matched Background Similarity*. Piscataway, NJ, USA: IEEE, 2011.
[57] T. Zheng, W. Deng, and J. Hu, "Cross-age LFW: A database for studying cross-age face recognition in unconstrained environments," 2017, *arXiv:1708.08197*. [Online]. Available: http://arxiv.org/abs/1708.08197
[58] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories," *Comput. Vis. Image Understand.*, vol. 106, no. 1, pp. 59–70, Apr. 2007.
[59] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," 2014, *arXiv:1412.6572*. [Online]. Available: http://arxiv.org/abs/1412.6572

ARSLAN ALI (Graduate Student Member, IEEE) received the M.Sc. degree in electrical engineering and information technology from the University of Stuttgart, Germany, in 2016. He is currently pursuing the Ph.D. degree in electrical engineering with the Politecnico di Torino, Turin, Italy.

His research interests include biometrics, face recognition, deep learning, image processing, and computer vision. He received several awards and honors, including the Gold Medal in the Faculty of Electrical and Computer Engineering and the Best Research Thesis Award at the Center for Advanced Studies in Engineering.

MATTEO TESTA (Member, IEEE) received the B.Sc. and M.Sc. degrees in telecommunications engineering from the Politecnico di Torino, Turin, Italy, in 2011 and 2012, respectively, and the Ph.D. degree in electronic and communications engineering from the Electronics Department, Politecnico di Torino, in 2016, under the supervision of Prof. E. Magli.

His research interests include deep learning and computer vision, random projections, and bayesian inference.

LEV MARKHASIN received the M.S. and Ph.D. degrees in mathematics from Friedrich Schiller University Jena, in 2009 and 2012, respectively.

He is currently a Senior Engineer with the Sony Research and Development Center Europe Stuttgart Laboratory 1. He has authored or coauthored over ten articles in the area of wavelets, numerical integration, computer vision, and deep learning. His research interests include neural network compression, RGB-D fusion, and sensor security.

TIZIANO BIANCHI (Member, IEEE) received the M.Sc. degree (Laurea) in electronic engineering and the Ph.D. degree in information and telecommunication engineering from the University of Florence, Italy, in 2001 and 2005, respectively. From 2005 to 2012, he was a Research Assistant with the Department of Electronics and Telecommunications, University of Florence. He is currently an Associate Professor with the Politecnico di Torino. He has authored over 100 papers in international journals and conference proceedings. His research interests include multimedia security technologies, signal processing in the encrypted domain, and security aspects of compressed sensing. He is a member of the IEEE SPS Technical Committee on Information Forensics and Security. He is an Associate Editor of the *Journal of Visual Communication and Image Representation*.

ENRICO MAGLI (Fellow, IEEE) received the M.Sc. and Ph.D. degrees from the Politecnico di Torino, Turin, Italy, in 1997 and 2001, respectively.

He is currently a Full Professor with the Politecnico di Torino. His research interests include deep learning, compressive sensing, image and video coding, and vision. He was a recipient of the 2011 Transactions Prize Paper Award by the IEEE Geoscience and Remote Sensing Society, the 2015 Best Student Paper Award (as the Senior Author), the 2019 Best Paper Award by the IEEE International Conference on Image Processing (IEEE ICIP), and the 2010 and 2014 Best Associate Editor Award by the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. He was an Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA. He is also an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and the *EURASIP Journal on Image and Video Processing*. From 2015 to 2016, he was an IEEE Distinguished Lecturer.