Cancelable templates for secure face verification based on deep learning and random projections

Ali *et al.*
EURASIP Journal on Information Security (2024) 2024:7
https://doi.org/10.1186/s13635-023-00147-y

EURASIP Journal on Information Security

RESEARCH

Open Access

Cancelable templates for secure face verification based on deep learning and random projections

Arslan Ali[1], Andrea Migliorati[1*], Tiziano Bianchi[1] and Enrico Magli[1]

Abstract

Recently, biometric recognition has become a significant field of research. The concept of cancelable biometrics (CB) has been introduced to address security concerns related to the handling of sensitive data. In this paper, we address unconstrained face verification by proposing a deep cancelable framework called BiometricNet+ that employs random projections (RP) to conceal face images and compressive sensing (CS) to reconstruct measurements in the original domain. Our lightweight design enforces the properties of unlinkability, revocability, and non-invertibility of the templates while preserving face recognition accuracy. We compare facial features by learning a regularized metric: at training time, we jointly learn facial features and the metric such that matching and non-matching pairs are mapped onto latent target distributions; then, for biometric verification, features are randomly projected via random matrices changed at every enrollment and query and reconstructed before the latent space mapping is computed. We assess the face recognition accuracy of our framework on challenging datasets such as LFW, CALFW, CPLFW, AgeDB, YTF, CFP, and RFW, showing notable improvements over state-of-the-art techniques while meeting the criteria for secure cancelable template design. Since our method requires no fine-tuning of the learned features, it can be applied to pre-trained networks to increase sensitive data protection.

Keywords Biometrics, Face verification, Biometric authentication, Cancelable biometrics, Compressed sensing

1 Introduction

Biometric signals have long been used in applications such as surveillance, access control, and behavior analysis, which typically rely on the acquisition and analysis of faces, fingerprints, palmprints, irises, and speech. Biometric recognition is defined as the verification of the identity of a person through a chosen biometric signal. Recently, concerns have been raised on how to protect sensitive biometric data against malicious attackers or external agencies, for example in a transmitter-receiver scenario where information needs to be exchanged on an unsecured channel. To this end, the concept of *cancelable biometrics* (CB) has been introduced [1] and defined as repeatable, intentional distortions obtained using a specific transformation of biometric signals which are then compared in the transformed domain [2]. The goal of CB is to provide a framework to ensure the security of the transformed signals, referred to as *secure templates*, within existing biometric systems. Much like a traditional biometric system, the CB recognition process is composed of an enrollment and verification phase, which respectively consists of extracting significant features from the acquired biometric signal and comparing previously unseen templates to verify whether they match already enrolled ones. We refer to [2] and [3] for a survey of existing methods to generate CB templates. Most importantly, a CB design should exhibit the following characteristics: (i) *unlinkability*: there are no methods

*Correspondence:
Andrea Migliorati
andrea.migliorati@polito.it
[1] DET - Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy

to decide whether two templates are extracted from the same biometric instance or not; (ii) *revocability*: compromised templates need to be immediately deleted, and new secure ones issued; (iii) *non-invertibility*: the original biometric signal should be very difficult to recover by observing the correspondent template; (iv) *performance preservation*: the CB system needs to perform as close as possible to the initial, non-secured one.

In this scenario, considering face images are among the most widespread biometric signals, we specifically tackle the problem of unconstrained face verification, which consists of establishing whether a pair of properly aligned face images refer to the same person or not. In recent approaches based on deep learning, discriminative features (*embeddings*) are learned and then compared with a fixed distance metric within an end-to-end trainable framework. In this work, we propose a CB face verification framework based on deep learning, called BiometricNet+, which employs compressive sensing (CS) [4] techniques. More in detail, we use random projections (RP) to project learned features of the face images in dedicated random spaces, thus obtaining secure templates. Then, we operate reconstruction back to the original domain at the verification step via a very efficient deep learning CS technique. Indeed, RP has recently attracted interest due to its potential ability to provide an easily implemented security layer [5–8]. However, while works in the literature typically employ CS as a standalone framework, we effectively employ RP within the deep learning framework to implement a lightweight encryption method providing CB whose reconstruction performance can be tuned according to the shape of the projection matrices, provided they are different for each sensitive biometric signal and changed at every query and enrollment. The method we use to reconstruct the projected features is ISTA-Net [9], a fast and accurate deep algorithm developed for CS reconstruction of natural images here deployed to recover one-dimensional feature vectors, as further explained in Section 3.3.

In recent deep approaches for face verification, it has been shown that the choice of the distance metric is a very crucial aspect [10–14]. Typically, a fixed distance metric is employed to compute distances between embeddings, and a loss function is devised so that a large margin is enforced on non-matching pairs (NMP), while distances are low for matching pairs (MP). In this paper, we introduce a different approach by which the most discriminative features and the best distance metric are jointly learned. Specifically, we regularize the output of the distance metric so that the values follow two separate statistical distributions, one for NMP and one for MP. To achieve this, we employ a deep architecture composed of two parts: the first one, called FeatureNet+, is a Siamese network that learns discriminative features that are then combined and given as input to the second one, namely MetricNet+, which is, in turn, responsible for the mapping of the feature onto output points in the regularized latent space. This joint deep design allows for the learning of a distance metric by minimizing a novel loss function that takes into account triplets of input pairs [15]. As detailed in Section 4.4, the proposed approach is used to learn a metric for the face recognition task allowing for a significant improvement over the state of the art also in the case of large-scale, challenging datasets. We would like to highlight the fact that the idea behind BiometricNet+ can go beyond face verification and could be generally applied to any biometric signals such as fingerprints or retina images, as well as any other deep learning architectures. Further investigations of this matter are left as future work.

1.1 Contributions

The contributions we bring with this paper are as follows:

- We introduce the deep CB framework BiometricNet+ based on RP and CS; our novel contribution is bringing security to the field of unconstrained face verification where the handling of sensible data can be a major concern; our approach is robust against similarity attacks and ensures unlinkability, revocability, and non-invertibility of the templates, as highlighted by our robustness analysis (Section 5.1) where we also evaluate security in the stolen-token scenario; at the same time, BiometricNet+ significantly outperforms competing face verification methods that often do not offer techniques for the protection of the templates; moreover, the CB feature requires very little computational overhead, hence it can potentially be embedded into any pre-trained network when in need of protecting sensitive data;
- We expand on our previous contribution [16] where we presented the concept of learning output distributions by mapping features onto a regularized latent space. Specifically, we employ an improved deep architecture based on depthwise separable convolution layers [17] which enables us to use fewer fully connected layers than [16], ensuring faster convergence and also improved regularization of the latent space. Furthermore, during training, we use an improved selection strategy of the input sample triplets based on a different threshold in the input space. Finally, the discriminative features are combined differently, as we compute the difference between them instead of the concatenation, resulting in a lighter aggregated feature vector. As a result, BiometricNet+ ensures a significant verification per-

formance over [16], especially on the most challenging test datasets, as explained in the following;
- We extensively evaluate the face recognition performance of the proposed framework by experimenting on wide and challenging datasets such as LFW, CALFW, CPLFW, AgeDB, YTF, CFP, and RFW; we report notable improvements over existing unprotected methods, achieving state-of-the-art recognition accuracy while meeting the criteria for a secure cancelable template design.

2 Related works

2.1 Cancelable biometrics in deep learning

Over the years, the concept of projecting signals using a transformation matrix has been used on several occasions within the deep learning framework. Initially, deep networks have been employed as tools to improve over standard CS techniques for recovering measurements of sparse signals such as natural images [18–23]. However, recent works attempted to use RP to build privacy-preserving systems able to perform biometric recognition [3, 24], also in mobile face verification scenarios [25]. In [26], authors propose a system that relies on deep networks and error-correction coding to generate a secure template from each user's multiple biometrics fused at a feature level. [27] expands on the previous design by presenting an architecture composed of a deep hashing framework followed by a deep decoder used to refine the intermediate binary hashing codes and to compensate for the discrepancies between probe biometrics and previously enrolled data caused by illumination and pose variations. Also, [28] proposes a plug-and-play method to generate templates that employs a significant bit-based representation and a fuzzy commitment scheme. Finally, [29] presents a modular template design based on an angular distance metric that can be used in face verification systems. All these methods suffer from computational overhead and increased complexity brought by error-correction decoding, which also causes a decrease in recognition accuracy. In particular, they mainly work in domains limited to small-scale test datasets that are scarcely representative of real-life applications or within specific networks or loss functions. For this reason, existing techniques are not easily scalable to state-of-the-art, large-scale natural image datasets for face verification. Instead, our method can potentially be employed on any pre-trained network. Moreover, the chosen CS recovery technique (i.e., ISTA-Net [9] ensures high performance on very challenging datasets while enabling scalability of the template size thanks to the tunable reconstruction error which depends on the size of the random matrices, as better explained in the following.

2.2 Face verification

Face verification has a great number of possible applications, such as robotics [30–32], forensics [33, 34], and security [35–37], both with images [38] and videos [39–41]. The problem is still very far from being marked as solved, as its complexity also extends to image acquisition technological issues such as camera specifics and skin reflectance [42]. Early techniques employed handcrafted facial feature detectors that could find the most discriminative traits that are supposedly unique in a person's face [43]. These methods relied on heavy pre-processing and illumination normalization of the images, but could not handle the non-linear variations face pictures can exhibit. Recently, remarkable advancements have been made in the field thanks to new deep learning techniques such as DeepID [10] and DeepFace [11]. In deep methods, features are computed for each face image in a pair, and then the measure of the distance between them is computed (usually the ℓ_2 norm). The distance is used in the verification stage: if the value is over a certain threshold, the two images depict the same individual. However, deep techniques do not rely on ad-hoc knowledge of the input image distribution but instead try to minimize a given loss function (typically the softmax cross-entropy) with respect to the target output. Works such as [12, 13] discovered that deep learning generalization capacity could be boosted by minimizing the intra-class variance and maximizing the inter-class variance of the output distributions. In particular, they enforced large margins in the Euclidean distance space between output distributions using contrastive loss functions. The pivotal FaceNet made further advances [14], in which a triplet loss function was introduced to account for the distance between embeddings in relative terms rather than absolute. However, more recent works like [44] reported that the improvements in the feature space are often hindered by complex and demanding training. For this reason, the following research focused on the employment of distance metrics different from the ℓ_2 norm that could allow for even more strict margins and better embeddings. Yang [45] and Liu et al. [46] introduced an angular distance able to decrease the number of false positives by enforcing a large distance margin between non-matching pairs. Furthermore, [47] and [48] proposed novel angular distance metrics that confirmed the shift from Euclidean to angular ones, reporting a remarkable performance increase. However, differently from these approaches, we do not rely on a fixed metric but we instead learn the most suitable function within a deep end-to-end trainable framework.

2.3 Latent space regularization

The notion of improving the discriminative capability of the distance metric between learned features by analyzing their distributions has first been discussed in [49]. The overlap between the histogram of the distances of MP and NMP is minimized to obtain more regularized embeddings. However, while working well with clustering problems, this approach is unsuited for face verification as the decision boundary between the histograms depends on the specific dataset and tends not to generalize well to other input data distributions. Conversely, our BiometricNet+ approach aims at regularizing the latent feature space by imposing a desired target behavior to the output distributions, based on a highly non-linear learned metric. In turn, we enable a straightforward decision boundary between matching and non-matching embeddings ensuring better generalization. The idea of enforcing target distributions was introduced in [50, 51] to solve one-against-all classification problems. In these works, the latent space is regularized to map the biometrics of a *single* individual onto a given output distribution, while all the other possible users fall onto another one. However, this design suffers from the major drawback of requiring user-specific training for each enrolled user. On the contrary, the proposed BiometricNet+ avoids any user-specific training by mapping feature differences, and not single features, on different distributions in a latent space.

3 Proposed method

The proposed BiometricNet+ CB design employs RP to create the secured templates from the discriminative representation of the input biometrics, jointly learned together with a suitable distance metric used to compare pairs of features. Specifically, we project feature vectors via the multiplication with a random matrix changed at every iteration, as further illustrated in Section 3.1. Our design enjoys several advantages. The enforcing of target distributions ensures the decision boundaries are straightforward and easily interpretable, in contrast with the typical behavior of deep networks which tend to yield highly complex, non-linear boundaries. By enforcing same-variance, different-mean Gaussian distributions, the optimal decision boundary in the latent space is exactly a hyperplane. Hence, the decision threshold can be tuned to obtain the required level of genuine acceptance rate (GAR) or false alarm rate (FAR), as the most difficult pairs of images are mapped onto the tails of the target distributions. Furthermore, Gaussian distributions are also amenable to writing the loss function in closed form which is a key point in the definition of the employed loss function.

As reported in Fig. 1, our structure is composed of two jointly trained sub-networks, namely FeatureNet+ and MetricNet+. The former is a siamese neural network [52] that receives as input pairs of face images $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2]$ and outputs pairs of features $[\mathbf{f}_1, \mathbf{f}_2]$ ($\mathbf{f}_1 \in \mathbb{R}^d, \mathbf{f}_2 \in \mathbb{R}^d$). Then, MetricNet+ maps the difference between the features $\mathbf{f} = \mathbf{f}_1 - \mathbf{f}_2$ ($\mathbf{f} \in \mathbb{R}^d$) onto a $\mathbf{z}$ point in a p-dimensional, regularized latent space where the output decision is made. The distance metric between $\mathbf{f}_1$ and $\mathbf{f}_2$ is learned through the loss function, which forces it to take values onto two latent target distributions ($\mathbf{z}$) according to whether the input pair is made of matching or non-matching features. In the test phase, we assume that a secure biometric template $\mathbf{f}_{1-proj}$ has been enrolled in the system by projecting the features $\mathbf{f}_1$ through a random matrix $\mathbf{W}_1$. A query template $\mathbf{f}_{2-proj}$ is obtained similarly by projecting the query features $\mathbf{f}_2$ via a random matrix $\mathbf{W}_2$. The two secure templates are then reconstructed at a trusted matching device that has access to both $\mathbf{W}_1$ and $\mathbf{W}_2$, and the reconstructed features $\mathbf{f}_{1-rec}$, $\mathbf{f}_{2-rec}$ are fed to the trained MetricNet+ to compute the decision score. In the following sections, we describe the different design choices of the proposed scheme.

3.1 Cancelable templates and reconstruction design

As anticipated, we generate cancelable templates by employing independent random matrices, re-generated at every enrollment to project the learned feature vectors. At testing time, query templates are computed similarly by using a different set of independent random matrices generated at every query. We consider a scenario where enrolled and query feature vectors are transmitted over an unsecured channel that is located between the FeatureNet+ and MetricNet+ subnetworks. More specifically, protected templates are computed in a secure area within the sensor and then either stored in an unprotected database or sent over an insecure channel to a remote server where the verification phase takes place in a secure area. In order to provide independent random matrices, a shared secret key may be available both at the sensor and at the remote server, and a new random matrix can be generated by combining the secret key with a random public initialization vector, in a way similar to common modes of operation of standard block ciphers. In this setting, the proposed design is necessary as unprotected templates are vulnerable to inversion attacks that might enable attackers to recover a version of the face image showing relevant features of the biometric signal, leading to privacy concerns. With the proposed solution, every query is independent of previous queries, even when associated with the same identity. In such a fashion, collecting multiple queries from the same

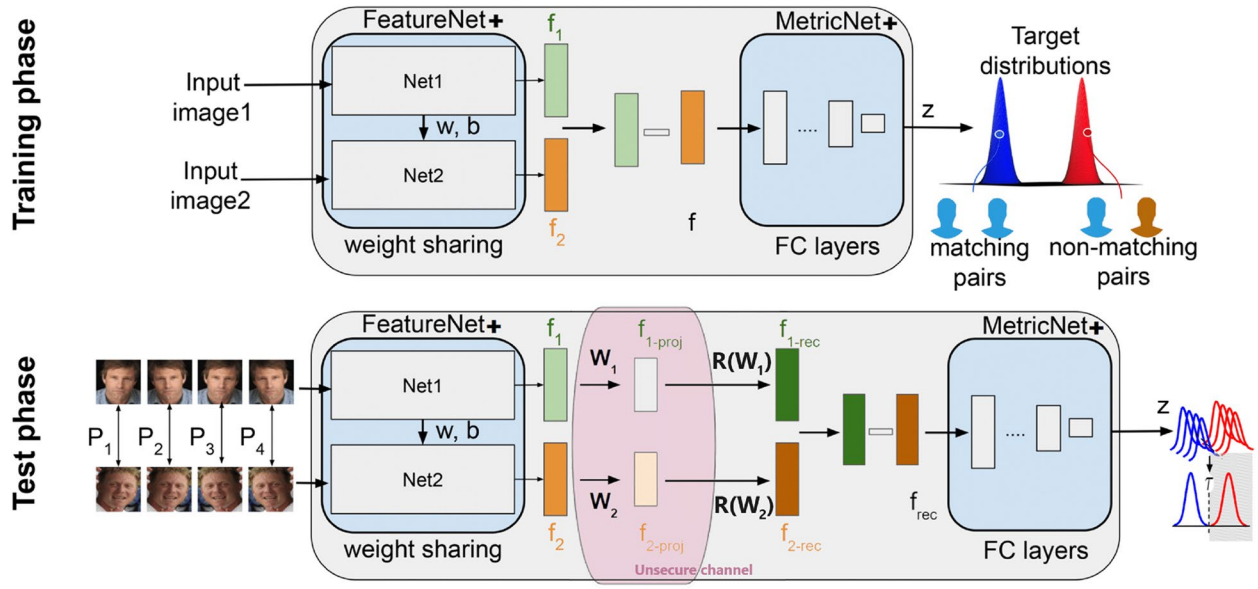

Fig. 1 *(Top)* BiometricNet+ during training: after face detection and alignment, pairs are given as input to FeatureNet+ which extracts discriminative features $\mathbf{f}_i \in \mathbb{R}^d$. The feature vectors are subtracted $\mathbf{f} = \mathbf{f}_1 - \mathbf{f}_2$ ($\mathbf{f} \in \mathbb{R}^d$) and passed to MetricNet+ which maps $\mathbf{f}$ onto the target distributions $\mathbf{z} \in \mathbb{R}^p$ in the latent space. *(Bottom)* BiometricNet+ during the test (i.e., authentication) phase: given a pair of aligned face images, we obtain 4 image pairs, i.e., P_1, P_2, P_3 and P_4 by accounting for all the possible horizontal flip combinations; then, features are projected in separated random spaces and reconstructed using ISTA-Net [9], simulating the transmission of sensitive data over an unsecured channel; the corresponding output vectors in the latent space are computed and then aggregated to $\bar{\mathbf{z}}$; finally, aggregated features are compared to a threshold τ

identity does not leak any information on the protected identity [53].

According to our design choice, our CB scheme projects one-dimensional feature vectors $\mathbf{f}_k$ at the output of FeatureNet+ ($\mathbf{f}_k \in \mathbb{R}^d$) via the multiplication with a $r \times d$ random matrix $\mathbf{W}_k \in \mathbb{R}^{r \times d}$, such that:

$$\mathbf{f}_{k-proj} = \mathbf{W}_k \cdot \mathbf{f}_k, \tag{1}$$

where $\mathbf{f}_{k-proj}$ indicates the projected features ($\mathbf{f}_{k-proj} \in \mathbb{R}^r$), and $\cdot$ indicates matrix multiplication. For generating the entries of the random matrices $\mathbf{W}_k$, we sample a random Gaussian distribution $\mathcal{N}(0, 1)$ and then normalize the columns to unitary energy, ensuring the security of the projected templates as further illustrated in Section 5.1. When an enrolled template must be compared with a query template, the original feature vectors can be approximately recovered from $\mathbf{f}_{k-proj}$ using a reconstruction function $R(\mathbf{W}_k)$, such that:

$$\mathbf{f}_{k-rec} = R(\mathbf{W}_k) \cdot \mathbf{f}_{k-proj}, \tag{2}$$

where $\mathbf{f}_{k-rec}$ is the reconstructed vector ($\mathbf{f}_{k-rec} \in \mathbb{R}^d$). In our implementation, we employed as R a modified ISTA-Net [9] network with np steps previously trained on 50000 $\mathbf{f}_k$ feature vectors at the output of FeatureNet+.

Our approach allows for independent random matrices at enrollment and query time, which is instrumental in providing the unlinkability and non-invertibility of the proposed templates. In our experimental evaluation, we assess performance by changing the number of rows of the random matrices r. We define the compression ratio as r/d, i.e., the ratio between the projection size and the feature dimensionality. The compression ratio of the framework determines a trade-off between the accuracy and the complexity of the system. In more detail, lower compression ratios correspond to secured templates of smaller size but lower accuracy.

3.2 Target distributions and loss selection

Let us refer to the chosen target distributions for MP and NMP as $\mathbb{P}_m$ and $\mathbb{P}_n$, respectively. We set $\mathbb{P}_m$ and $\mathbb{P}_n$ to be p-variate Gaussian distributions:

$$\mathbb{P}_m = \mathcal{N}(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m), \ \mathbb{P}_n = \mathcal{N}(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n), \tag{3}$$

where $\boldsymbol{\Sigma}_m = \sigma_m^2 \mathbb{I}_p$ and $\boldsymbol{\Sigma}_n = \sigma_n^2 \mathbb{I}_p$ are the diagonal covariance matrices, and $\boldsymbol{\mu}_m = \mu_m \mathbf{1}_p^T$, $\boldsymbol{\mu}_n = \mu_n \mathbf{1}_p^T$ denote the expected values. We enforce two same-variance Gaussian distributions with far enough expected values so that separability is ensured. Our choice is motivated by the fact that different variances could cause their optimal values to be dependent on the considered input dataset, as the training would need to match the specific intra-class and inter-class variances.

In general, any desired target distribution could be enforced. However, Gaussian distributions are arguably the best fit for multiple reasons. First, according to the central limit theorem, the output of the fully connected layers typically employed in deep networks tends to follow a Gaussian distribution [54]. Secondly, when $\mathbf{\Sigma}_m = \mathbf{\Sigma}_n$, Gaussian distributions ensure the optimal decision boundary is a linear hyperplane in the decision space. Finally, Gaussian distributions enable the loss function to be written in closed form. Hence, in the following, we will refer only to the Gaussian case.

Let us define $\mathbf{x}_m$ and $\mathbf{x}_n$ as the generic pairs of matching and non-matching face images, respectively. Similarly, let us denote by $\mathbf{f}_m$ and $\mathbf{f}_n$ the corresponding FeatureNet+ output features. MetricNet+ can be represented as a generic encoding function $H(\cdot)$ of the input feature pairs, i.e., $\mathbf{z} = H(\mathbf{f})$, where $\mathbf{z} \in \mathbb{R}^p$ such that $\mathbf{z}_m \sim \mathbb{P}_m$ if $\mathbf{f} = \mathbf{f}_m$, and $\mathbf{z}_n \sim \mathbb{P}_n$ if $\mathbf{f} = \mathbf{f}_n$. By our design, the Kullback-Leibler (KL) divergence between the input and target distributions can be written in closed form and easily minimized. More specifically, the KL divergence for multivariate Gaussian distributions is a function of only first-order and second-order statistics:

$$\mathcal{L}_m = \frac{1}{2}\left[\log\frac{|\mathbf{\Sigma}_m|}{|\mathbf{\Sigma}_{Sm}|} - p + \mathrm{tr}(\mathbf{\Sigma}_{\mathrm{m}}^{-1}\mathbf{\Sigma}_{\mathrm{Sm}}) + (\boldsymbol{\mu}_m - \boldsymbol{\mu}_{Sm})^{\intercal}\mathbf{\Sigma}_m^{-1}(\boldsymbol{\mu}_m - \boldsymbol{\mu}_{Sm})\right], \tag{4}$$

where the S indicates the input statistics.

In particular, we can compute the KL divergence batch-wise. During training, the network receives as input a set of b image face pairs (where b is the batch size). From this set, we extract two subsets of $b/2$ difficult matching and $b/2$ difficult non-matching face pairs, as explained in detail in Section 3.4. Furthermore, given b pairs in the batch $\mathbf{X} \in \mathbb{R}^{b\times r}$, where r is the size of a face pair, these are mapped onto the set of latent space points $\mathbf{Z} \in \mathbb{R}^{b\times p}$. Then, we can compute the 1st and 2nd order statistics of the embeddings $\mathbf{Z}_m, \mathbf{Z}_n$, respectively referring to matching ($\boldsymbol{\mu}_{Sm}$,$\mathbf{\Sigma}_{Sm}$) and non-matching ($\boldsymbol{\mu}_{Sn}$,$\mathbf{\Sigma}_{Sn}$) input faces. The loss as written in Eq. (4) captures the MP statistics and enforces the target distribution $\mathbb{P}_m$. In a similar fashion, we can derive $\mathcal{L}_n$ for NMP, enforcing the target distribution $\mathbb{P}_n$. Hence, the final loss function $\mathcal{L}$ minimized end-to-end across the whole network is defined as:

$$\mathcal{L} = \mathcal{L}_m + \mathcal{L}_n. \tag{5}$$

As explained, the loss as in Eq. 5 is based on the Kullback-Leibler (KL) divergence and offers the advantage of being written in closed form, efficiently enforcing the target distribution onto the MP and NMP one, and ensuring fast convergence. However, a potential drawback might appear in situations where training with very large batch sizes is required. Specifically, the complexity of the batch-wise computation as in Eq. 4 increases linearly with the batch size. Even if it does not affect the proposed method in the considered scenarios, further investigations on how to efficiently deal with large batch sizes are left as future work.

3.3 Architecture

In the following, we discuss the architecture and the implementation of FeatureNet+ and MetricNet+.

3.3.1 FeatureNet+

As said, obtaining discriminative features from input face pairs is of pivotal importance. Hence, we employ a siamese version of the state-of-the-art architecture Inception-ResNet-V1 [55], which also exhibits fast convergence. In more detail, the Inception-ResNet stem block has an output size of $35 \times 35 \times 256$, and it is followed by 5 blocks of *Inception-ResNet-A*, 10 blocks of *Inception-ResNet-B*, and 5 blocks of *Inception-ResNet-C*. After the Inception-ResNet blocks, we employ a fully connected layer with output size d, i.e., the feature vector dimensionality.

3.3.2 MetricNet+

MetricNet+ receives as input the difference between the FeatureNet+ feature vectors $\mathbf{f} = \mathbf{f}_1 - \mathbf{f}_2 \in \mathbb{R}^d$ of size d. To learn the best metric according to the target distributions in the latent space, we employ 7 fully connected layers and ReLU activations at the output of the first 6 layer, while the last one is left without an activation function. The feature output size decreases by a factor of 2 for each layer, starting from the first fully connected layer with an output size equal to d. In particular, our design ensures the final layer exhibits an output size equal to the latent space dimensionality p.

3.4 Pairs selection during training

To improve convergence during training, we select the most difficult MP and NMP. These pairs are far from the mean values of the target distributions and therefore close to the decision boundary and prone to misclassification. At the end of the forward pass, we identify at the mini-batch level the subset of MP defined as $||\mathbf{z}_m - \boldsymbol{\mu}_m||_\infty \geq (\boldsymbol{\mu}_m + \boldsymbol{\mu}_n)/2$, which represents the matching pairs whose output $\mathbf{z}_m$ is sufficiently distant from the $\mathbb{P}_m$ center of mass. In a similar fashion, we choose NMP such that $||\mathbf{z}_n - \boldsymbol{\mu}_n||_\infty \leq (\boldsymbol{\mu}_m + \boldsymbol{\mu}_n)/2$. Consequently, in the backward pass, we minimize the

loss as in Eq. (5) over a subset of $b/2$ difficult matching and $b/2$ difficult non-matching pairs. However, to ensure stable training, the backward pass is executed only when $b/2$ difficult pairs can be found for both matching and non-matching pairs. If not, the mini-batch is discarded. This procedure ensures that difficult pairs are mapped onto regions of the latent space that are progressively separated as the training proceeds, contributing to a decrease in the uncertainty for pairs lying near the decision boundary in the latent space.

3.5 Verification

During the verification phase, an enrollment template and a query template are given as input to the network which outputs the corresponding metric value $\mathbf{z}$, which is then compared against the threshold τ. The choice of Gaussian target distributions ensures a hyperplane is the optimal decision boundary. Hence, the following conditions can be evaluated:

$$(\boldsymbol{\mu}_m - \boldsymbol{\mu}_n)^T \mathbf{z} \lessgtr (\boldsymbol{\mu}_m - \boldsymbol{\mu}_n)^T (\boldsymbol{\mu}_m + \boldsymbol{\mu}_n)/2. \tag{6}$$

In particular, when $p = 1$, Eq. (6) boils down to comparing the scalar $\mathbf{z}$ with the threshold $\tau = (\mu_m + \mu_n)/2$. However, to improve upon this design, we employ flipped images to improve the ability of the network to capture discriminative information, inspired by recent approaches in the literature such as [46, 47]. In more detail, given a pair of aligned face images, we compute the corresponding metric output $\mathbf{z}$ for all the 4 pairs resulting from all possible combinations of horizontally flipped and non-flipped images. In practice, this means that for every face we enroll two different templates obtained from the original image and the flipped version, while in the verification phase, we compute two different query templates using the same strategy. The system then computes all possible distances between an enrollment template and a query template.

We employ horizontal flipping defined as $(x, y) \rightarrow (width - x - 1, y)$, where *width* is the width of the aligned face image in the pair, and (x, y) represents the spatial coordinates. Thus, given the 4 obtained metric outputs, the decision is performed based on the value $\bar{\mathbf{z}} = \frac{1}{4}\left(\sum_{i=1}^{4} \mathbf{z}_i\right)$, where $\mathbf{z}_i$ refers to the output of the *i-th* image flip combination. The expected value of $\bar{\mathbf{z}}$ for matching and non-matching pairs is respectively equal to $\boldsymbol{\mu}_m$ and $\boldsymbol{\mu}_n$. Hence, Eq. (6) still holds and we can compare the $\bar{\mathbf{z}}$ metric output against the hyperplane decision boundary in the latent space. Figure 1 (*Bottom*) illustrates the proposed BiometricNet+ architecture during testing, where P1 represents the input image pair, and P2, P3, and P4 represent the three horizontal flip combinations.

4 Experiments

In this section, we evaluate the proposed approach over multiple challenging datasets and compare it against state-of-the-art face verification techniques. First, we establish a baseline temporarily disregarding the protection of the templates and analyze the effect of the parameters on verification accuracy. We refer to the baseline as *unsecured baseline* or *BiometricNet+NOCB*. Secondly, we introduce the protection of the templates in the testing phase and show how performance is preserved while ensuring the required security properties (Section 5.1).

4.1 Pre-processing and datasets

To validate our method, we use different datasets for training and testing. Specifically, we train on Casia [56], which is composed of $0.49M$ images for a total of $10k$ individuals, and MS1M-DeepGlint [57], made of $3.9M$ images referring to $87k$ individuals. On the other hand, we test our design over six popular unconstrained face datasets built for 1:1 verification (i.e., only a single image template per subject is available), which is the specific problem we tackle with BiometricNet+. For this reason, we do not consider large-scale datasets such as MegaFace [58] and IJB [59] that are instead used for set-based face recognition (i.e., to verify whether two sets of face images refer to the same person). The most widespread datasets for unconstrained face verification are Labeled Faces in the Wild (LFW) [60] and YouTube Faces (YTF) [61], respectively composed of 13233 face images collected from 5749 people and 3425 videos of 1595 individuals. However, late state-of-the-art techniques have reached almost perfect accuracy on these two datasets. For this reason, we also evaluate four more challenging datasets: (i) Cross-Age LFW (CALFW) [62], constructed by selecting 3000 positive face pairs from LFW and including images of the same individual taken at different moments in time: the introduced age gap increases the intra-class variance of the image set belonging to the person; (ii) Cross-Pose LFW (CPLFW) [63], based on 3000 face pairs from LFW taken with different facial poses, hence also accounting for increased intra-class variance; (iii) Celebrity Dataset in Frontal and Profile Views (CFP) [64], mounting up to 7000 images from 500 individuals; (iv) In-the-wild age database (AgeDB) [65], containing 16488 images of 568 different people.

As a first step, we pre-process input images by employing the MTCNN [66] alignment technique to generate normalized facial crops of size 160×160. Then, following [46–48], we mean-normalize the cropped images and constrain them in the range $[-1, 1]$. Finally, for all the

considered datasets, we test on 3000 matching and 3000 non-matching pairs, for a total of 6000 test image pairs. We report the performance of the proposed approach following the standard evaluation protocol *unrestricted with labeled outside data* used in [14, 47, 48].

4.2 Parameters evaluation

Initially, we investigate how the design parameters influence performance. First, we explore the impact of the feature vector dimensionality d as given as output by FeatureNet+ while keeping the latent space dimensionality fixed at $p = 1$. Secondly, we set d to the best value found in the previous step and vary p to determine the pair of values leading to maximum verification accuracy. Finally, we analyze the effect of the enforced target distribution parameters, namely σ_m and σ_n. Having two same-variance target distributions ($\sigma_m = \sigma_n = \sigma = 1$), and assuming $\mu_m = 0$, we can restrict the analysis to the single free parameter μ_n, which indicates as to how far apart the learned distributions are in the latent space. Despite the sensible improvements over our previous contribution [16], the proposed deep network design leads to the same optimal values of d, p, and μ_n well across the considered datasets. Hence, for the sake of brevity, we do not report this experimental validation. In the following, if not differently specified, we employ $d = 512$, $p = 1$, $\mu_n = 40$. This particular choice, while it enables to capture of the most discriminative facial features and to avoid overfitting, also ensures the balance between the enforceability of the target distributions and latent space complexity. Performance comparison between [16] and the proposed BiometricNet+ is presented later in the manuscript.

4.3 Experimental settings

We train BiometricNet+ using stochastic gradient descent (SGD) [67, 68] with the Adam optimizer [69]. Each epoch runs over images of 720 different faces, with at least 5 images per individual, to ensure enough MP and NMP are available at training time. We set the batch size b to 210 pairs, chosen as explained in Section 3.4. Our choice ensures a sufficiently high amount of pairs per batch so that significant first-order and second-order statistics can be computed by the network. Also, we balance the batches so that they each contain the same amount of MP and NMP. The learning rate is first set to 0.01, and then it decreases with an exponential decay factor of 0.98 every 5 epochs. The network is trained for a total of 500000 iterations, with a weight decay fixed to 2×10^{-4}. We also employ dropout with a keep probability of 0.8. Finally, we employ data augmentation via horizontal flipping of the input images. The experiments have been implemented using TensorFlow [70] and run on NVIDIA GeForce GTX Titan X GPUs.

4.4 Cancelable template performance assessment

In this section, we proceed to evaluate performance when deploying the secured templates in the verification phase and also compare results with the unsecured framework and other competing methods for unconstrained face verification. Besides verification accuracy, we also take into account the genuine acceptance rate (GAR) as a function of the false acceptance rate (FAR), which is respectively the relative amount of correctly accepted MP and the relative amount of incorrectly accepted NMP. In such a fashion, we analyze how BiometricNet+ can generalize across different datasets on the 1:1 verification task. Specifically, we compute verification accuracy, GAR at FAR $= 10^{-2}$, and GAR at FAR $= 10^{-3}$ at test time when employing RP with random matrices $\mathbf{W}_i \in \mathbf{R}^{r \times d}$ and ISTA-Net reconstruction with $np = 3$, where the feature vector dimensionality is $d = 512$, and np denotes the number of ISTA-Net phases [9]. Unlinkability and non-invertibility are ensured by assigning a different random matrix $\mathbf{W}_k$ to each feature vector and re-generating them at every encryption. The results, shown in Tables 1 and 2, are computed for projection size $r \in 64, 128, 256$, corresponding to a compression ratio equal to 1/8, 1/4, and 1/2 of the size of the embeddings ($d = 512$), and compared against the unsecured baseline. To perform the reconstruction of projected signals, we employ a pre-trained ISTA-Net network with $np = 3$ and 4 channels of 1×3 kernels, accounting for a total number of parameters just short of $1K$. This design allows the reconstruction process to be extremely lightweight as it can run on GPU with very limited resource consumption. For the sake of completeness, it is worth mentioning that we tested our framework using ISTA-Net reconstruction with a number of phases rp taking integer values in the range $1 - 10$, with no significant improvement in reconstruction performance when $np > 3$. For this reason, we only report results for $np = 3$.

Results in Tables 1 and 2 show that verification accuracy and GAR obtained when employing the cancelable templates are comparable to the ones yielded by an unsecured framework. As a consequence of the improved separability of the learned output distributions in the latent space, our method achieves high GAR values at low FAR values even for the most complex datasets. In particular, the gap between the unsecured baseline and the secured design is almost negligible when $r = 256$ (i.e., compression ratio equal to 1/2, given $d = 512$) and reasonably low when $r = 64$ (i.e., compression ratio equal to 1/8). The choice of r determines a trade-off between verification performance and compression. Hence, it is of pivotal importance in determining the most suitable configuration for the security layer according to the application and its required work point, typically specified by

Table 1 Verification accuracy, GAR @ FAR = 10^{-2}, and GAR @ FAR = 10^{-3}at test time when employing the CB framework with random matrices $\mathbf{W}_i \in \mathbf{R}^{r\times d}$and ISTA-Net reconstruction with $np = 3$, where the feature vector dimensionality is $d = 512$, and np denotes the number of ISTA-Net phases [9]. Results are computed for projection size $r \in 64, 128, 256$, corresponding to a compression ratio equal to 1/8, 1/4, and 1/2 of the size of the embeddings ($d = 512$), and compared against the unsecured baseline (NOCB)

Proj. size	Measure	LFW	YTF	CALFW	CPLFW	CFP	AgeDB
$r = 64$	Verification Acc.	99.28	97.08	96.17	94.18	97.73	94.45
	GAR@FAR=10^{-2}	99.53	93.97	92.00	83.25	94.83	82.20
	GAR@FAR=10^{-3}	92.83	82.77	79.03	61.12	85.07	63.53
$r = 128$	Verification Acc.	99.62	97.78	96.67	95.33	99.13	95.57
	GAR@FAR=10^{-2}	99.80	96.17	93.43	86.70	99.17	87.40
	GAR@FAR=10^{-3}	98.37	90.53	86.77	65.87	96.83	73.67
$r = 256$	Verification Acc.	99.72	97.85	96.70	95.67	99.18	95.82
	GAR@FAR=10^{-2}	99.83	96.47	93.63	87.63	99.33	88.20
	GAR@FAR=10^{-3}	98.50	91.43	87.57	66.73	97.07	74.07
NOCB	Verification Acc.	99.82	98.16	97.10	95.85	99.43	96.24
	GAR@FAR=10^{-2}	99.80	96.93	94.63	88.53	99.43	89.23
	GAR@FAR=10^{-3}	99.20	92.20	89.50	68.27	97.57	74.70

Table 2 Equal error rate (EER, %, the lower the better) for BiometricNet+ as a function of the template projection size (r). For comparison, we also report the best EER obtained by the competing method SoftmaxOut [24]

Proj. size	LFW	YTF	CALFW	CPLFW	CFP	AgeDB
$r = 64$	1.13	4.20	6.17	8.56	3.00	10.33
$r = 128$	0.87	3.54	5.40	6.69	1.78	8.11
$r = 256$	0.76	3.31	4.22	6.19	1.33	6.67
NOCB	0.56	3.02	4.00	5.72	1.03	6.13
SoftmaxOut [24]	3.73	9.12	-	-	-	-

FAR requirements. Our security framework allows for maximum flexibility while preserving the discriminative power of the underlying deep structure, at the expense of limited computational overhead.

Table 3 reports the maximum verification accuracy achieved by several state-of-the-art techniques over the six considered large-scale datasets. We compared BiometricNet+ with the competing methods that reach the highest accuracy in the face verification field but often do not offer the protection of the templates. The comparison is however fair since unprotected models offer an upper bound on the verification performance of the system. As shown, the BiometricNet+ method significantly outperforms the other methods, especially for the most challenging datasets where we remarkably improve over the baseline methods CosFace, ArcFace, and SphereFace. In particular, the performance gain we obtain over other competitors is the highest on the most challenging CPLFW dataset. To summarize, Table 3 shows how the proposed technique consistently obtains higher accuracy, hence proving that the discriminative capacity of deep networks for facial verification can be boosted by learning the best metric to compare facial features in a regularized latent space. It is also important to notice how the proposed method is extremely competitive with the state of the art even for very low projection ratios ($r = 64$).

To further substantiate our claims, we also computed and compared the verification accuracy of the proposed BiometricNet+ over the RFW dataset [79], recently introduced to study the effects of racial bias in the field of deep face recognition. Results in Table 4 show that the proposed approach outperforms competing methods by a large margin.

5 Security and robustness analysis

In this section, we discuss the properties of non-invertibility and unlinkability ensured by the proposed CB design. Also, we analyze the characteristics of the learned distributions in the latent space and their robustness against perturbations. Compared to other CB approaches, the proposed design has different

Table 3 Verification accuracy (%) of different methods on LFW, YTF, CALFW, CPLFW, CFP, and AgeDB. BiometricNet+ outperforms the state of the art on all the considered datasets. The # Images column indicates the dimension of the training set for each approach. The numbers in brackets for the BiometricNet+ entries represent the verification accuracy obtained without the four flips at inference time

Method	# Images	LFW	YTF	CALFW	CPLFW	CFP	AgeDB
DeepID [10]	0.2M	99.47	93.20	-	-	-	-
SphereFace [46]	0.5M	99.42	95.0	90.30	81.40	94.38	91.70
SphereFace+ [71]	0.5M	99.47	-	-	-	-	-
CenterLoss [72]	0.7M	99.28	94.9	85.48	77.48	-	-
Baidu [73]	1.3M	99.13	-	-	-	-	-
UniformFace [74]	2.21M	99.80	97.70	-	-	-	-
VGGFace [11]	2.6M	98.95	97.30	90.57	84.00	-	-
MarginalLoss [75]	3.8M	99.48	95.98	-	-	-	-
DeepFace [76]	4.4M	97.35	91.4	-	-	-	-
RangeLoss [77]	5M	99.52	93.7	-	-	-	-
CosFace [47]	5M	99.73	97.6	-	-	95.44	-
ArcFace [48]	5.8M	99.82	98.02	95.45	92.08	98.37	95.15
FaceNet [14]	200M	99.63	95.10	-	-	-	89.98
BiometricNet [16]	3.8M	99.80	98.06	97.07	95.60	99.35	96.12
BiometricNet+ r=64	4.4M	99.28	97.08	96.17	94.18	97.73	94.45
		(99.11)	(96.98)	(96.03)	(94.02)	(97.59)	(94.21)
BiometricNet+ r=128	4.4M	99.62	97.78	96.67	95.33	99.13	95.57
		(99.32)	(97.39)	(96.29)	(95.11)	(99.01)	(95.27)
BiometricNet+ r=256	4.4M	99.72	97.85	96.70	95.67	99.18	95.82
		(99.41)	(97.56)	(96.41)	(95.27)	(99.04)	(95.43)
BiometricNet+ NOCB	4.4M	99.82	98.16	97.10	95.85	99.43	96.24
		(99.75)	(98.01)	(96.88)	(95.30)	(99.37)	(96.17)

security requirements, since the matching of protected templates is assumed to be performed in a secure area. Nevertheless, the different matching domain does not affect the security of the scheme in terms of non-invertibility and unlinkability of the protected templates, as shown in the following.

Table 4 Verification accuracy (%) of different state-of-the-art methods on RFW. BiometricNet+ outperforms the state of the art on all the considered datasets. The numbers in brackets for the BiometricNet+ entries represent the verification accuracy without the four flips at inference time

Method	Caucasian	Indian	Asian	African
CenterLoss [72]	87.18	81.92	79.32	78.00
SphereFace [46]	90.80	87.02	82.95	82.28
VGGFace2 [78]	89.90	86.13	84.93	83.38
ArcFace [48]	97.37	95.68	94.55	93.87
CosFace [47]	96.63	94.68	93.50	92.17
IMAN-A [79]	-	94.15	91.15	91.42
RL-RBN [80]	97.08	95.63	95.57	94.87
BiometricNet+ NOCB	99.33	98.75	98.33	98.12
	(99.12)	(98.41)	(98.13)	(97.98)

5.1 Cancelable templates security evaluation

The security of the proposed CB scheme is evaluated according to the two following properties:

- *Unlinkability*: it is computationally infeasible to determine whether two protected templates $\mathbf{f}_{1-proj}$ and $\mathbf{f}_{2-proj}$ are derived from two images of the same subject or images of different subjects;
- *Non-invertibility*: it is computationally infeasible to reconstruct the enrollment image $\mathbf{x}$ from the protected template $\mathbf{f}_{proj}$ without the corresponding key $\mathbf{W}$.

To evaluate security, we can refer to the literature on the security properties of random projections. Starting from the seminal work of [5], authors in [6] investigate the secrecy properties of RP matrices with entries drawn from Gaussian distributions. They demonstrate that, if the sensing matrix is independently drawn at each encryption and the acquired signals have the same energy, then it is possible to achieve perfect secrecy assuming that the matrix is unknown to the adversary, i.e.:

$$P\{\mathbf{x}|\mathbf{f}_{proj}\} = P\{\mathbf{x}\}. \tag{7}$$

In our framework, since different random matrices are generated for every enrollment and every query, the above property immediately implies non-invertibility under ideal generation of random matrices. This is a strong theoretical guarantee, often not available in current CB schemes. Moreover, it is easy to show that, for a generic observed template $\mathbf{f}_{proj}$ and any possible pair of enrollment images $\mathbf{x}_1$ and $\mathbf{x}_2$, we have:

$$P\{\mathbf{f}_{proj}|\mathbf{x}_1\} = P\{\mathbf{f}_{proj}|\mathbf{x}_2\}, \tag{8}$$

which also implies perfect unlinkability.

5.1.1 Unlinkability analysis

However, practical systems cannot deal with infinite precision representations for which the above theoretical results would hold. Also, feature vectors do not have constant energy, meaning that an adversary may have a small advantage in solving the unlinkability problem. To address the first concerns, different works such as [7] and [8] consider more practical projection matrices such as those with entries drawn from Bernoulli distributions (the former) and quantized Gaussian distributions (the latter). In [8], it is shown that the probability of distinguishing RP obtained with different quantized sensing matrices decreases exponentially with the number of bits used to represent the entries of the projection matrix. Hence, by choosing a suitable precision for the random matrix entries, we can achieve any desired level of unlinkability. Regarding the second concern, even if feature vectors do not have constant energy, we can safely assume that they have the same energy on average. For feature vectors having a large number of dimensions, this implies that most vectors have very similar energy. In this case, an adversary has a very small advantage in distinguishing templates of different subjects from templates of the same subject.

As in most recent CB schemes [24, 25], we employ the evaluation framework proposed in [81] to measure the unlinkability of the templates, so that results can be directly compared with other CB schemes. In particular, [81] introduces two unlinkability metrics for the evaluation of biometric template protection systems. The first is a local score-wise measure, $D_{\leftrightarrow}(s)$, based on the likelihood ratio between the score distributions obtained with a given score function computed between matching and non-matching distributions of secure templates, and the second is a global measure, $\mathrm{D}_{\leftrightarrow}^{\mathrm{sys}}$, independent of the score domain, which allows for a fair evaluation of the unlinkability of each system. Both measures yield values in the [0, 1] range, where 0 indicates perfect unlinkability and 1 perfect separability of the secure template distributions. We first generated two distributions of 100000 matching and 100000 non-matching pairs of secure templates and computed the scores for each of the pairs using the normalized crossed-correlation function. The considered number of pairs is high enough to give a statistically meaningful estimate of the two unlinkability metrics [81]. We repeated the experiment for different projection sizes $r \in 64, 128, 256$. The results are reported in Fig. 2, which presents the local unlinkability plots and the global unlinkability measure. As no significant difference between different r values has been observed, only the $r = 128$ plot has been reported.

The plots of the matching (green) and non-matching (red) scores distribution between the secure template pairs are completely overlapped, making it impossible for an attacker to assume whether two secure templates are matching or not. One may notice that the $D_{\leftrightarrow}(s)$ measure across the normalized cross-correlation score range (blue curve) exhibits two spikes in the correspondence of the tails of the distributions. However, this does not have to be attributed to a decreased local unlinkability of the templates but rather to the low number of pairs with that particular score which causes the $D_{\leftrightarrow}(s)$ computation to be noisy in that region. Overall, the unlinkability of the secure templates generated by our method is conclusively confirmed by the value of the global unlinkability measure $\mathrm{D}_{\leftrightarrow}^{\mathrm{sys}} = 0.01$ which is very close to perfect unlinkability ($\mathrm{D}_{\leftrightarrow}^{\mathrm{sys}} = 0$).

5.1.2 Non-invertibility in the stolen-token scenario

We now assess the security of our approach when the random matrix and the projected features are disclosed to the public. Only a handful of works [82–86] have previously tackled the problem of reconstructing face images from deep features. These methods are usually based on convolutional neural networks, reformulated minimization problems, or generative adversarial network (GAN) mappings. Specifically, GANs are showing promise in recent works in the development of new inversion attacks like in the case of [87] where authors reformulate the face reconstruction task as a constrained optimization problem solved using the genetic algorithm. In this paper, we evaluate the non-invertibility of the templates by simulating a worst-case white-box pre-image attack based on the state-of-the-art NbNet [88] network. In particular, NbNet is based on specific de-convolution blocks aimed at reconstructing face images from deep templates. In NbNet, the output channel of each layer consists of fewer repeated and noisy channels, hence containing more details for face image reconstruction than the typical de-convolution blocks. In our evaluation, we assume that the attacker has access to the fully trained BiometricNet+ model and a pre-trained ISTA-Net model. The attacker is also able to observe a significant amount of protected templates $f_{i,\ldots,k\ -proj}$ and their corresponding

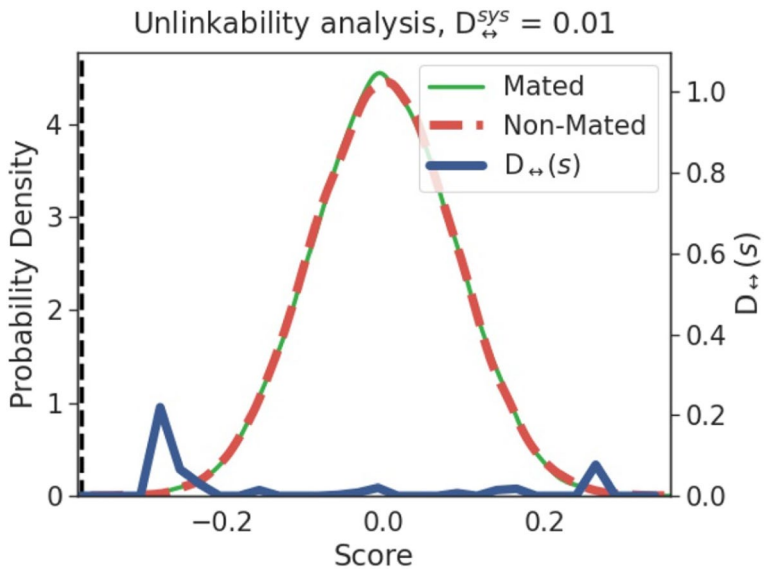

Fig. 2 Unlinkability plots for 100000 matching and 100000 non-matching secure templates obtained with BiometricNet+; the scores between each of the pairs of templates are computed as the normalized cross-correlation between the one-dimensional template vectors of projection size $r = 128$; there are no significant differences when varying the value of r

projection matrices $W_{i,\ldots,k}$. The attacker is then able to estimate a certain amount of FeatureNet+ embeddings $f_{i,\ldots,k\ -est}$ reconstructed using ISTA-Net from the secure templates. Furthermore, these estimated embeddings are used to train NbNet [88] to recover face pre-images that would possibly allow the attacker to obtain access to the system, effectively breaking security. In such a fashion, we bypass NbNet's shortcoming of requiring a GAN to generate images for training by relying directly on the feature vectors extracted from the training dataset at the output of MetricNet+. We simulate the attack by training NbNet on 1000000 embeddings $f_{i,\ldots,k\ -rec}$ reconstructed with ISTA-Net from secure templates $f_{i,\ldots,k\ -proj}$. Once convergence is reached, we are now able to obtain a significant number of NbNet [0, 1] normalized pre-images that will be used by the attacker to attempt to access the BiometricNet+ system. For this experiment, we considered random matrices $W_{i,\ldots,k}$ with projection size $r = 128$, i.e., we trained NbNet on embeddings that are reconstructed via ISTA-Net from 128-dimensional templates. However, our findings also extend to $r = 64$ and $r = 256$.

Figure 3 illustrates some examples of pre-images generated with an NbNet model trained on $r = 128$ reconstructed embeddings, compared against the original image in the pixel domain (first column). In particular, the figure shows pre-images generated from reconstructed embeddings with different projection sizes (third to fifth column). We evaluate how the different pre-images correlate to the original images from which the secured templates have been derived. We measure the average distance between the pre-images and the original images using the LPIPS perceptual similarity metric [89], which is known to be more informative in evaluating how face images correlate in the pixel domain than traditional metrics such as MSE or PSNR. Results in Table 5 show that matching and non-matching pairs of original images and pre-images generated from NbNet inferred embeddings of size $r = 64, 128, 256$ are not separable from one another. For reference, Fig. 3 and Table 5 also include the pre-images generated from *clean* embeddings taken at the output of FeatureNet+ (second column), which are assumed to be unavailable to the attacker. In this *clean* case, the LPIPS similarity yields a lower average score for matching pairs than for non-matching pairs (0.37624 against 0.52557). Indeed, these pre-images appear to be structurally similar to their original corresponding images. Hence, it would be possible for an attacker to have an advantage in inferring key features of the original face image from the reconstructed pre-images. This result confirms the necessity of employing secure templates to avoid a potential privacy breach.

As shown, the face pre-images in the third to fifth columns in Fig. 3, generated from reconstructed embeddings ($r = 64$, $r = 128$, $r = 256$) appear to be very distant from the original image, both from a qualitative point of view and in terms of LPIPS similarity. However, it is necessary to quantitatively assess whether the pre-images would allow the attacker to access BiometricNet+ as a legitimate user. To verify whether the attack will prove successful, we assess performance over a set

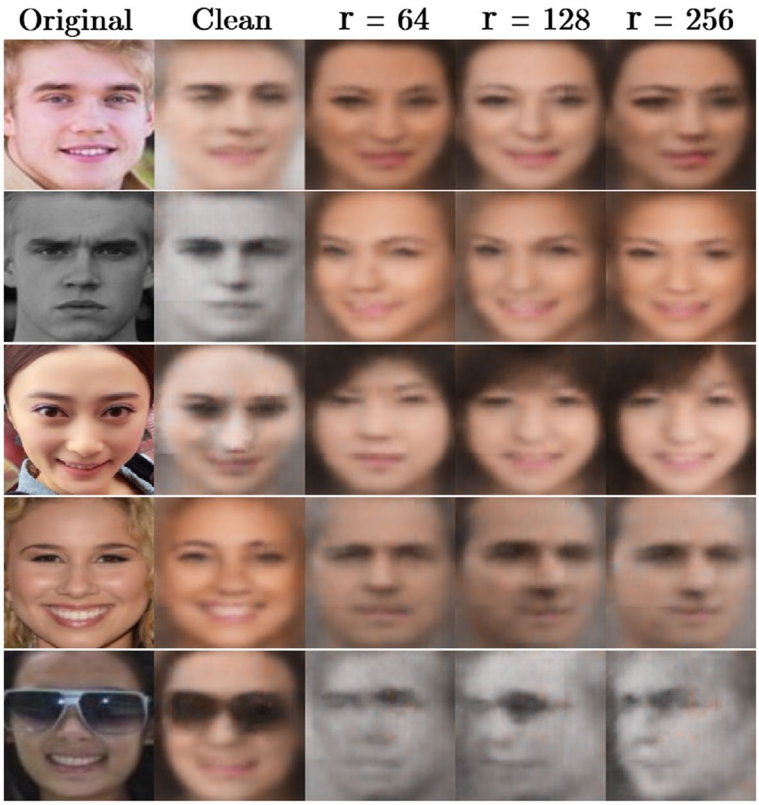

Fig. 3 Examples of pre-images generated with a NbNet model (trained on $r = 128$ reconstructed embeddings) from reconstructed embeddings with different projection sizes (third to fifth column) and from *clean* embeddings taken at the output of FeatureNet+ (second column), compared against the original image in the pixel domain (first column)

Table 5 Average LPIPS perceptual similarity over 30000 pairs of images composed of original images and their corresponding NbNet pre-image (*matching pairs*) and 30000 pairs of images and pre-images generated from unrelated templates (*non-matching pairs*), as a function of the size of the secured templates from which the pre-images have been derived

Pairs	Clean	$r = 64$	$r = 128$	$r = 256$
Matching	0.37624	0.51812	0.50773	0.50107
Non-matching	0.52557	0.50843	0.50893	0.504596

of 30000 matching face image pairs composed of the original image and the corresponding NbNet pre-image. Similarly, we run BiometricNet+ over the same amount of non-matching face image pairs generated by pairing NbNet pre-images with non-matching original images. Hence, we can produce matching accuracy, FAR, and GAR over the pairs to verify whether the attacker could use NbNet pre-images to access the system. Results in Fig. 4 show how a pre-trained BiometricNet+ network reaches approximately 50% accuracy (random guess) on the considered dataset and close to zero FAR and GAR. Specifically, the verification accuracy sets to 50.55%, while the GAR@FAR=10^{-1} and the GAR@FAR=10^{-2} are equal to 8.33% and 0.67%, respectively. In other words, BiometricNet+ is unable to separate matching original/pre-image pairs from non-matching ones as the learned distributions in the latent space do not allow for pairs of features that are not discriminative enough to be mapped onto the matching distribution, hence strictly limiting the success of the attack. Our findings demonstrate the attacker would not be able to effectively access the system as a legitimate user even in the case of the considered worst-case white-box attack scenario. We also argue that conventional encryption would not provide an alternative solution to the proposed protected templates. Biometric templates are similar to passwords hashed with a random salt. A biometric template should guarantee that the original biometric data cannot be retrieved even if the template and the random token used to generate the template are available to the adversary. Simply encrypting the biometric template using the token as the secret key does not work in this scenario. As shown in our results, an adversary who is able to recover the projection matrix

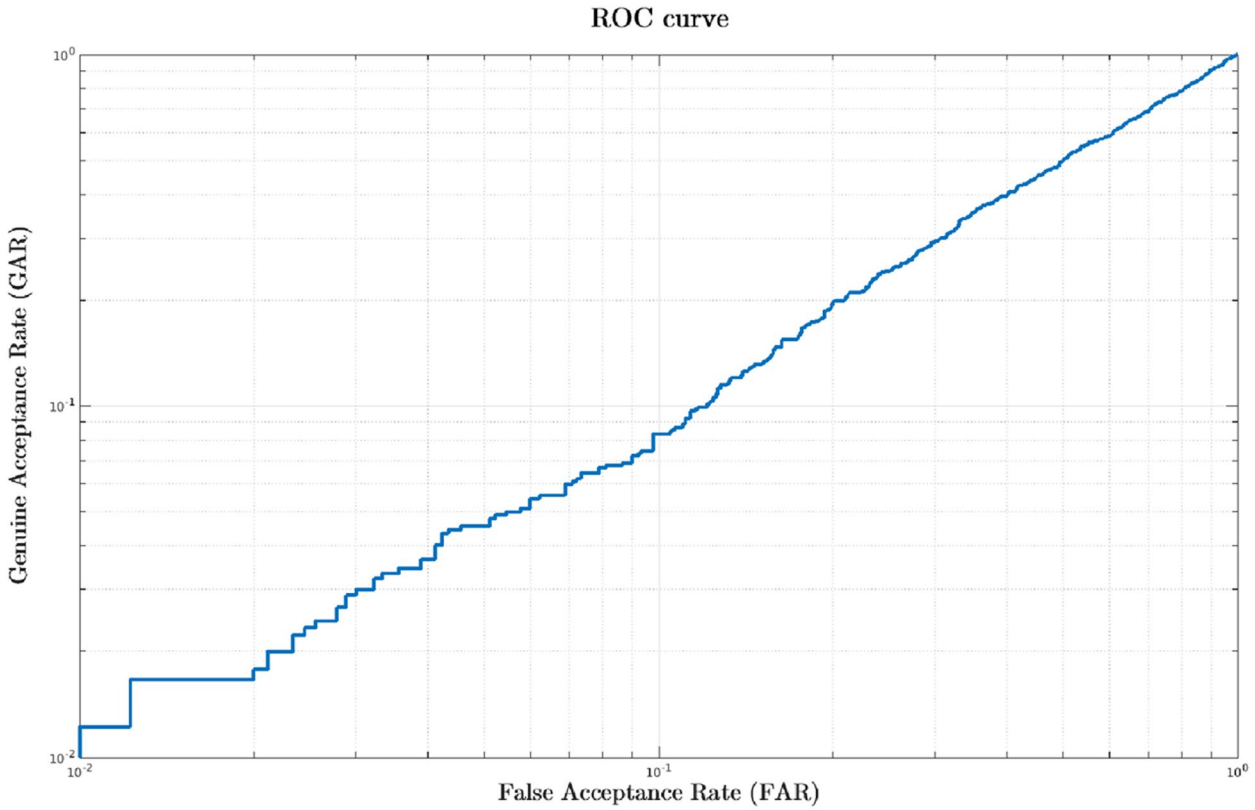

Fig. 4 ROC curve for the BiometricNet+ performance at inference time on 30000 matching and 30000 non-matching pairs of original and NbNet pre-images recovered from $r = 128$ embeddings; the curve, which exhibits an almost perfect diagonal behavior (e.g., *GAR* = *FAR*), confirms that an attacker would not be able to use NbNet pre-images to access the system even when the projection matrix and the secure templates are disclosed to the public

would still be unable to reconstruct face images that are similar to the original template. Conversely, an adversary who can decrypt a template that has been protected with conventional encryption will have access to clean embeddings and hence be able to obtain a high-quality estimate of the original template.

5.2 Analysis of the feature distribution in the latent space

In this section, we analyze the effects of the regularization of the latent space on the learned distributions. For the sake of brevity, we refer to the unsecured baseline only but all our findings straightforwardly extend also to the CB design.

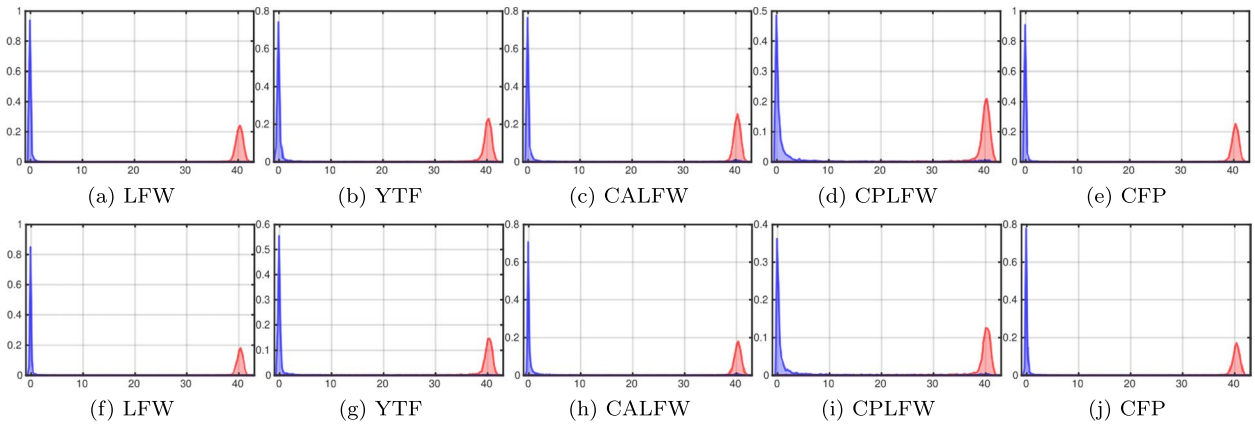

Fig. 5 (*Top*) **a**–**e** Histograms of the distributions of the **z** decision variable for matching (blue) and non-matching (red) pairs, obtained with the proposed method, over different datasets. (*Bottom*) **f**–**j** Histogram of the distribution of the $\bar{\mathbf{z}}$ decision variable for matching (blue) and non-matching (red) pairs, obtained with the proposed BiometricNet+, over different datasets

Table 6 Scores for the high confidence region for matching pairs ($\mathbf{z}_m$) and non-matching pairs ($\mathbf{z}_n$) together with the maximum accuracy obtained for LFW, YTF, CALFW, CPLFW, CFP, and AgeDB datasets

Dataset	$(\mu - 3\sigma \leq \mathbf{z}_m \leq \mu + 3\sigma)$	$(\mu - 3\sigma \leq \mathbf{z}_n \leq \mu + 3\sigma)$	Max accuracy
LFW	99.33%	95.60%	99.82
YTF	95.37%	86.23%	98.16
CALFW	91.67%	95.80%	97.10
CPLFW	85.27%	84.30%	95.85
CFP	98.40%	94.73%	99.43
AgeDB	86.27%	83.03%	96.24

5.2.1 Histograms of the $\mathbf{z}$ *and* $\bar{\mathbf{z}}$ *distributions*

We start by looking at the histograms of the distributions of $\mathbf{z}$ and $\bar{\mathbf{z}}$ calculated over different test datasets. Results are respectively shown in Fig. 5 (*Top*) (a–e) and Fig. 5 (*Bottom*) (f–j), where blue curves refer to the matching pairs, and red curve to the non-matching pairs. The curves exhibit the enforced Gaussian shape for both MP and NMP. As expected, the histograms of the NMP distributions are centered around $\mu_n = 40$, i.e., the μ_n target value we set initially for $\mathbb{P}_n$. The different pairs are well-separated in the latent space with almost no overlap of the histograms, accounting for the reported improved accuracy over the state-of-the-art. It can be observed that the MP distributions tend to exhibit a lower variance than the target one. This is particularly evident when testing on more difficult datasets such as CALFW (Fig. 5c–h) and CPLFW (Fig. 5d–i), where MP distributions exhibit heavier tails than the target. A way to explain this behavior lies in the large difference in variability between MP and NMP input distributions, which can lead to difficulties in enforcing the same variance on both of them at the same time. Indeed, given a fixed number of individuals, the amount of possible NMP is much larger than the amount of possible MP. Moreover, due to the lack of symmetry of the KL divergence employed in the loss function as in Eq. (4), the training phase tends to promote the learning of output distributions with a smaller variance than the target one. Hence, the target variance should be considered as an upper bound and the training ends up being conservative.

Furthermore, by comparing the histograms for the $\bar{\mathbf{z}}$ scores as in Fig. 5 (*Bottom*) against the $\mathbf{z}$ histograms in Fig. 5 (*Top*), we can observe that the variance of both the MP and NMP distributions is slightly lower compared to that of $\mathbf{z}$. This substantiates that it is indeed preferable to employ the averaged output $\bar{\mathbf{z}}$ over the simple $\mathbf{z}$. At the same time, since the optimal decision boundaries in the latent space depend only on the 1st order statistics which are indeed preserved, we can conclude that the slight decrease in the variance of the learned output distributions in the latent space does not affect the accuracy of the proposed method.

5.2.2 Confidence Intervals

We now proceed to analyze the confidence intervals of the learned distributions. In particular, we are interested in the percentage of pairs belonging to the *high-confidence* region. We define as belonging to this region those pairs that fall within an interval of 3σ around the target mean, both for MP and NMP, with respective decision variables z_m and z_n. The greater the percentage of pairs falling in this region, the lower the probability of having pairs that are mapped onto the tails of the distributions in the latent space, allowing for lower misclassification rates. We compute these percentages on the LFW dataset in testing and report the results in Table 6 which also shows GAR values at FAR=$\{10^{-2}, 10^{-3}\}$). As previously shown, the proposed method achieves very high accuracy on LFW, learning an output distribution that is very close to the target one. Our method maps 99.33% of the matching pairs and 95.60% of the non-matching pairs onto the high confidence region within the 3σ band around the mean. The same conclusions can be drawn for the CFP and YTF datasets, where the output distributions exhibit very light tails ensuring high confidence. Considering instead the more challenging CALFW and CPLFW datasets, the percentage of MP mapped onto the high confidence region of the latent space is considerably lower as compared to NMP. This behavior accounts for the great intra-class variance introduced within the datasets by adding age and pose variations. The above results suggest that the position of a pair in the latent space directly relates to the confidence the network has in such a choice. This is particularly useful for biometric authentication systems as the authentication threshold has to be set according to some expected confidence measure.

6 Conclusions

In this work, we propose a CB framework based on deep learning with application to the problem of unconstrained face verification. Our method employs random projections to project face images onto dedicated random spaces and compressive sensing to reconstruct the signals. Our lightweight technique can potentially be employed

on any pre-trained network in need of securing sensitive data. To generate templates, we use different random matrices for each signal, re-generated at each encryption. In this way, our design ensures the unlinkability, revocability, and non-invertibility of the templates, while preserving the face recognition accuracy of the unsecured design. Our method relies on the mapping of discriminative learned facial features onto a regularized latent space, where an effective distance metric between matching and non-matching pairs can be learned. For this reason, we enable the learning of a complex metric as opposed to the typical complex partitioning of the feature space achieved by deep methods, so that BiometricNet+ employs simple linear decision boundaries in the latent space. Significant accuracy improvements over the state of the art have been found for the resulting face verification system, both in secure and unsecured cases. As reported in our extensive experimental evaluation, BiometricNet+ consistently outperforms existing techniques over large-scale challenging datasets, while ensuring the security of the templates.

Abbreviations

CB	Cancelable biometrics
CS	Compressive sensing
RP	Random projections
MP	Matching pairs
NMP	Non-matching pairs
SGD	Stochastic gradient descent

Acknowledgements
This research has been carried out in collaboration with Sony R &D Center Europe, Stuttgart, Laboratory 1.

Authors' contributions
AA performed the initial set of experiments and drafted the first version of the manuscript. AM curated the security and robustness analysis. AM, TB, and EM updated the manuscript to the final version. All authors read and approved the final manuscript.

Availability of data and materials
Model evaluation was performed on publicly available datasets such Labeled Faces in the Wild (LFW) [60], YouTube Faces (YTF) [61], Cross-Age LFW (CALFW) [62], Cross-Pose LFW (CPLFW) [63], Celebrity Dataset in Frontal and Profile Views (CFP) [64], in-the-wild age database (AgeDB) [65], and RFW [79].

Declarations

Competing interests
The authors declare that they have no competing interests.

Received: 8 June 2023 Accepted: 20 November 2023
Published online: 08 March 2024

References

1. C. Soutar, D. Roberge, A. Stoianov, R. Gilroy, B.V. Kumar, in *Optical Security and Counterfeit Deterrence Techniques II*. Biometric encryption using image processing (1998)
2. Patel, V.M., Ratha, N.K., Chellappa, R. (2015). "Cancelable biometrics: A review." *IEEE signal processing magazine*, *32*(5), 54-65.
3. N. Kumar, et al., Cancelable biometrics: a comprehensive survey. Artif. Intell. Rev. (2020)
4. Donoho, D.L. (2006). "Compressed sensing." *IEEE Transactions on information theory*, *52*(4), 1289-1306.
5. Yaron R., Baron D. (2008). "The secrecy of compressed sensing measurements." 2008 46th Annual Allerton conference on communication, control, and computing. IEEE.
6. Tiziano, B., Bioglio, V., Magli E. (2015). "Analysis of one-time random projections for privacy preserving compressed sensing." *IEEE Transactions on Information Forensics and Security*, *11*(2), 313-327.
7. Cambareri, V., et al. (2015). "Low-complexity multiclass encryption by compressed sensing." *IEEE transactions on signal processing*, *63*(9), 2183-2195.
8. Testa, M., Tiziano, B., Magli, E. (2019). "Secrecy Analysis of Finite-Precision Compressive Cryptosystems." *IEEE transactions on information forensics and security*, *15*, 1-13.
9. Zhang, J., Ghanem B. (2018). "ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing." Proceedings of the IEEE conference on computer vision and pattern recognition.
10. Sun, Y., et al. (2014). "Deep learning face representation by joint identification-verification." *Advances in neural information processing systems*, *27*.
11. Omkar, P, Vedaldi, A., Zisserman A. (2015). "Deep face recognition." BMVC 2015-Proceedings of the British Machine Vision Conference 2015. British Machine Vision Association
12. Sun, Yi., et al. (2015). "Deepid3: Face recognition with very deep neural networks." arXiv preprint arXiv, 1502.00873.
13. Sun, Y., Wang, X., & Tang, X. (2015). Deeply learned face representations are sparse, selective, and robust. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2892-2900).
14. Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 815-823).
15. Chechik, G., Sharma, V., Shalit, U., & Bengio, S. (2010). Large scale online learning of image similarity through ranking. *Journal of Machine Learning Research*, *11*(3).
16. Ali, A., Testa, M., Bianchi, T., & Magli, E. (2020). Biometricnet: deep unconstrained face verification through learning of metrics regularized onto gaussian distributions. In European Conference on Computer Vision (pp. 133-149). Springer International Publishing, Cham.
17. L. Sifre, S. Mallat, Rigid-motion scattering for image classification. Ph. D. thesis (2014)
18. Mousavi, A., Patel, A. B., & Baraniuk, R. G. (2015). A deep learning approach to structured signal recovery. In 2015 53rd annual allerton conference on communication, control, and computing (Allerton) (pp. 1336-1343). IEEE.
19. Adler, A., Boublil, D., Elad, M., & Zibulevsky, M. (2016). A deep learning approach to block-based compressed sensing of images. arXiv preprint arXiv:1606.01519.
20. Nguyen, D. M., Tsiligianni, E., & Deligiannis, N. (2017). Deep learning sparse ternary projections for compressed sensing of images. In 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP) (pp. 1125-1129). IEEE.
21. Xu, K., Zhang, Z., & Ren, F. (2018). Lapran: A scalable laplacian pyramid reconstructive adversarial network for flexible compressive sensing reconstruction. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 485-500).
22. Wu, Y., Rosca, M., & Lillicrap, T. (2019). Deep compressed sensing. In International Conference on Machine Learning (pp. 6850-6860). PMLR.
23. Shi, W., Jiang, F., Liu, S., & Zhao, D. (2019). Scalable convolutional neural network for image compressed sensing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 12290-12299).
24. Lee, H., Low, C. Y., & Teoh, A. B. J. (2021). SoftmaxOut transformation-permutation network for facial template protection. In 2020 25th International Conference on Pattern Recognition (ICPR) (pp. 7558-7565). IEEE.
25. V.K. Hahn, S. Marcel, Towards protecting face embeddings in mobile face verification scenarios. IEEE Trans. Biom. Behav. Identity Sci. **4**(1), 117–134 (2022)
26. Talreja, V., Valenti, M. C., & Nasrabadi, N. M. (2017). Multibiometric secure system based on deep learning. In 2017 IEEE Global conference on signal and information processing (globalSIP) (pp. 298-302). IEEE.
27. Talreja, V., Soleymani, S., Valenti, M. C., & Nasrabadi, N. M. (2019). Learning to authenticate with deep multibiometric hashing and neural network

decoding. In ICC 2019-2019 IEEE International Conference on Communications (ICC) (pp. 1-7). IEEE.
28. Mohan, D. D., Sankaran, N., Tulyakov, S., Setlur, S., & Govindaraju, V. (2019). Significant Feature Based Representation for Template Protection. In CVPR Workshops (pp. 2389-2396).
29. Kim, S., Jeong, Y., Kim, J., Kim, J., Lee, H. T., & Seo, J. H. (2021). IronMask: Modular architecture for protecting deep face template. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 16125-16134).
30. Wang, J., Zheng, J., Zhang, S., He, J., Liang, X., & Feng, S. (2016). A face recognition system based on local binary patterns and support vector machine for home security service robot. In 2016 9th international symposium on computational intelligence and design (ISCID) (Vol. 2, pp. 303-307). IEEE.
31. Li, X., Zhao, H., Zhao, H., Wang, J., & Xia, P. (2017). Face Recognition for Intelligent Robot Safety Verification System. In Proceedings of the 2017 International Conference on Computer Science and Artificial Intelligence (pp. 10-13).
32. Dua, I., Nambi, A. U., Jawahar, C. V., & Padmanabhan, V. N. (2019). Evaluation and visualization of driver inattention rating from facial features. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, *2*(2), 98-108.
33. Banerjee, S., & Ross, A. (2020). Face phylogeny tree using basis functions. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, *2*(4), 310-325.
34. Suri, A., Vatsa, M., & Singh, R. (2020). A2-LINK: recognizing disguised faces via active learning and adversarial noise based inter-domain knowledge. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, *2*(4), 326-336.
35. Scherhag, U., Debiasi, L., Rathgeb, C., Busch, C., & Uhl, A. (2019). Detection of face morphing attacks based on PRNU analysis. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, *1*(4), 302-317.
36. Heusch, G., George, A., Geissbühler, D., Mostaani, Z., & Marcel, S. (2020). Deep models and shortwave infrared information to detect face presentation attacks. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, *2*(4), 399-409.
37. Kotwal, K., Bhattacharjee, S., & Marcel, S. (2019). Multispectral deep embeddings as a countermeasure to custom silicone mask presentation attacks. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, *1*(4), 238-251.
38. Sae-Bae, N., Wu, J., Memon, N., Konrad, J., & Ishwar, P. (2019). Emerging NUI-based methods for user authentication: A new taxonomy and survey. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, *1*(1), 5-31.
39. Zheng, J., Ranjan, R., Chen, C. H., Chen, J. C., Castillo, C. D., & Chellappa, R. (2020). An automatic system for unconstrained video-based face recognition. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, *2*(3), 194-209.
40. Mokhayeri, F., & Granger, E. (2019). Video face recognition using siamese networks with block-sparsity matching. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, *2*(2), 133-144.
41. Sharma, V., Tapaswi, M., Sarfraz, M. S., & Stiefelhagen, R. (2019). Video face clustering with self-supervised representation learning. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, *2*(2), 145-157.
42. Cook, C. M., Howard, J. J., Sirotin, Y. B., Tipton, J. L., & Vemury, A. R. (2019). Demographic effects in facial recognition and their dependence on image acquisition: An evaluation of eleven commercial systems. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, *1*(1), 32-41.
43. Kas, M., El-merabet, Y., Ruichek, Y., & Messoussi, R. (2020). A comprehensive comparative study of handcrafted methods for face recognition LBP-like and non LBP operators. *Multimedia Tools and Applications*, *79*, 375-413.
44. Wang, J., Zhou, F., Wen, S., Liu, X., & Lin, Y. (2017). Deep metric learning with angular loss. In Proceedings of the IEEE international conference on computer vision (pp. 2593-2601).
45. Liu, W., Wen, Y., Yu, Z., & Yang, M. (2016). Large-margin softmax loss for convolutional neural networks. arXiv preprint arXiv:1612.02295.
46. Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., & Song, L. (2017). Sphereface: Deep hypersphere embedding for face recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 212-220).
47. Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., ... & Liu, W. (2018). Cosface: Large margin cosine loss for deep face recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5265-5274).
48. Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4690-4699).
49. Ustinova, E., & Lempitsky, V. (2016). Learning deep embeddings with histogram loss. Advances in neural information processing systems, 29.
50. Testa, M., Ali, A., Bianchi, T., & Magli, E. (2019). Learning mappings onto regularized latent spaces for biometric authentication. In 2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP) (pp. 1-6). IEEE.
51. Ali, A., Testa, M., Bianchi, T., & Magli, E. (2019). Authnet: Biometric authentication through adversarial learning. In 2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP) (pp. 1-6). IEEE.
52. Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., & Shah, R. (1993). Signature verification using a" siamese" time delay neural network. *Advances in neural information processing systems*, *6*.
53. Zhou, X., & Kalker, T. (2010, January). On the security of biohashing. In Media forensics and security II (Vol. 7541, pp. 266-273). SPIE.
54. R.M. Neal, *Bayesian learning for neural networks*, vol. 118 (2012)
55. Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the AAAI conference on artificial intelligence (Vol. 31, No. 1).
56. Yi, D., Lei, Z., Liao, S., Li, S.Z. (2014). "Learning face representation from scratch." arXiv preprint arXiv:1411.7923.
57. Deng, J., Guo, J., Zhang, D., Deng, Y., Lu, X., & Shi, S. (2019). Lightweight face recognition challenge. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (pp. 0-0).
58. Kemelmacher-Shlizerman, I., Seitz, S. M., Miller, D., & Brossard, E. (2016). The megaface benchmark: 1 million faces for recognition at scale. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4873-4882).
59. Brianna, M., Adams, J., Duncan, J.A., Kalka, N., Miller, T., Otto, C., Jain A.K., et al. (2018). "Iarpa janus benchmark-c: Face dataset and protocol." In 2018 international conference on biometrics (ICB), IEEE, pp. 158-165.
60. Huang, GB., Mattar, M., Berg, T., Learned-Miller, E. (2008). "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments." In Workshop on faces in'Real-Life'Images: detection, alignment, and recognition.
61. Wolf, L., Hassner, T., Maoz, I. (2011). "Face recognition in unconstrained videos with matched background similarity." In CVPR 2011, IEEE, pp. 529-534.
62. Tianyue, Z., Deng, W., Hu, J. (2017). "Cross-age lfw: A database for studying cross-age face recognition in unconstrained environments." arXiv preprint arXiv:1708.08197.
63. T. Zheng, W. Deng, Cross-Pose LFW: a database for studying crosspose face recognition in unconstrained environments. Beijing University of Posts and Telecommunications, Tech. Rep (2018)
64. Jun-Cheng C., Patel, V.M., Chellappa, R. (2016) "Unconstrained face verification using deep cnn features." In 2016 IEEE winter conference on applications of computer vision (WACV), IEEE, pp. 1-9.
65. Stylianos, M., Papaioannou, A., Sagonas, C., Deng, J., Kotsia, I., Zafeiriou, S. (2017) "Agedb: the first manually collected, in-the-wild age database." In proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 51-59.
66. Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE signal processing letters*, *23*(10), 1499-1503.
67. Yann, L., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel L.D. (1989) "Backpropagation applied to handwritten zip code recognition." *Neural computation*, *1*(4), 541-551.
68. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533-536.
69. Diederik P.Kingma., Ba, J. (2014) "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980.
70. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). {TensorFlow}: a system for {Large-Scale} machine learning. In 12th USENIX symposium on operating systems design and implementation (OSDI 16) (pp. 265-283).
71. Liu, W., Lin, R., Liu, Z., Liu, L., Yu, Z., Dai, B., & Song, L. (2018). Learning towards minimum hyperspherical energy. Advances in neural information processing systems, 31.
72. Wen, Y., Zhang, K., Li, Z., & Qiao, Y. (2016). A discriminative feature learning approach for deep face recognition. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14 (pp. 499-515). Springer International Publishing.

73. Liu, J., Deng, Y., Bai, T., Wei, Z., & Huang, C. (2015). Targeting ultimate accuracy: Face recognition via deep embedding. arXiv preprint arXiv:1506.07310.
74. Y. Duan, J. Lu, J. Zhou, in *Proceedings of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*. Uniformface: learning deep equidistributed representation for face recognition (2019)
75. J. Deng, Y. Zhou, S. Zafeiriou, in *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops*. Marginal loss for deep face recognition (2017)
76. Y. Taigman, M. Yang, M. Ranzato, L. Wolf, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Deepface: closing the gap to human-level performance in face verification (2014)
77. X. Zhang, Z. Fang, Y. Wen, Z. Li, Y. Qiao, in *Proc. of the IEEE Int. Conf. on Computer Vision*. Range loss for deep face recognition with long-tailed training data (2017)
78. Q. Cao, L. Shen, W. Xie, O.M. Parkhi, A. Zisserman, in *13th IEEE Int. Conf. on Automatic Face & Gesture Recognition*. Vggface2: a dataset for recognising faces across pose and age (2018)
79. M. Wang, W. Deng, J. Hu, X. Tao, Y. Huang, in *IEEE/CVF Int. Conf. on Computer Vision (ICCV)*. Racial faces in the wild: reducing racial bias by information maximization adaptation network (2019)
80. WANG, Mei; DENG, Weihong. Mitigating bias in face recognition using skewness-aware reinforcement learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. p. 9322-9331.
81. Gomez-Barrero, M., Galbally, J., Rathgeb, C., Busch, C. (2017) General framework to evaluate unlinkability in biometric template protection systems. *IEEE Trans Inf Forensics Secur*, *13*(6), 1406-1420
82. Zhmoginov, A., & Sandler, M. (2016). Inverting face embeddings with convolutional neural networks. arXiv preprint arXiv:1606.04189.
83. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). Analyzing and improving the image quality of stylegan. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 8110-8119).
84. Ahmad, S., & Fuller, B. (2020, September). Resist: Reconstruction of irises from templates. In 2020 IEEE International Joint Conference on Biometrics (IJCB) (pp. 1-10). IEEE.
85. Dong, X., Jin, Z., Guo, Z., & Teoh, A. B. J. (2021, September). Towards generating high definition face images from deep templates. In 2021 International Conference of the Biometrics Special Interest Group (BIOSIG) (pp. 1-11). IEEE.
86. K. et al, Inverting binarizations of facial templates produced by deep learning (and its implications). IEEE Trans. Inf. Forensics Secur. *16*, 4184–4196 (2021)
87. Dong, X., Miao, Z., Ma, L., Shen, J., Jin, Z., Guo, Z., & Teoh, A. B. J. (2022). Reconstruct Face from Features Using GAN Generator as a Distribution Constraint. arXiv preprint arXiv:2206.04295.
88. Mai, G., Cao, K., Yuen, P. C., & Jain, A. K. (2018). On the reconstruction of face images from deep face templates. *IEEE transactions on pattern analysis and machine intelligence*, *41*(5), 1188-1202.
89. Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 586-595).

Publisher's Note