SN Computer Science (2023) 4:619
https://doi.org/10.1007/s42979-023-01974-z

ORIGINAL RESEARCH

Learning Biometric Representations with Mutually Independent Features Using Convolutional Autoencoders

Riccardo Musto[1] · Ridvan Salih Kuzu[2] · Emanuele Maiorana[1] · Gabriel Emile Hine[1] · Patrizio Campisi[1]

Received: 11 July 2022 / Accepted: 29 May 2023

Abstract

Representations of biometric traits to be used in automatic recognition systems are typically learned with the goal of obtaining significant discriminative capabilities, that is, generating features that are notably different when produced by traits of different subjects, while maintaining an appropriate consistency for a given user. Nonetheless, discriminability is not the only desirable property of a biometric representation. For instance, the mutual independence of the coefficients in the employed templates is a valuable property when designing biometric template protection schemes. In fact, managing representations with independent coefficients allows to maximize the achievable security. In this paper we propose different learning strategies to obtain biometric representations with the property of statistical independence among coefficients, while preserving discriminability. In order to achieve this goal, different strategies are employed to train convolutional autoencoders. As a proof of concept, the effectiveness of the proposed approaches is tested by considering biometric recognition systems using both finger-vein and palm-vein patterns.

Keywords Biometric recognition · Statistical independence · Representation learning · Vein patterns

Introduction

Representation learning is a discipline concerned with the exploitation of machine learning algorithms to automatically obtain, from the analyzed data, a set of coefficients to be used for specific purposes, typically not attainable by directly exploiting the original raw form of the considered signals [2]. For example, a new data representation can be used to obtain discriminative characteristics, that is, features that can be effectively employed for classification purposes [19].

Among the others, a potential field of application of representation learning is automatic person recognition using biometric identifiers, where it is of paramount importance to handle templates that are significantly different when extracted from traits of distinct subjects, while as stable as possible when obtained from the same subject [3]. This allows to automatically recognize a legitimate person and grant physical or logical access to specific goods or services, as well as to reject potential impostors, by evaluating the similarity between the templates generated from the acquired data and those associated to the claimed identity. It is worth to remark that, while discriminability is a remarkable property needed to operate a biometric system, additional requirements are needed when addressing template security and privacy issues [6]. Actually, the compromise of a biometric identifier would imply severe consequences for the legitimate owner of the biometric data, such as the impossibility to further use the involved template, or the

This article is part of the topical collection "Advances on Pattern Recognition Applications and Methods 2022" guest edited by Ana Fred, Maria De Marsico and Gabriella Sanniti di Baja.

✉ Emanuele Maiorana
emanuele.maiorana@uniroma3.it

Riccardo Musto
ric.musto@stud.uniroma3.it

Ridvan Salih Kuzu
ridvan.kuzu@dlr.de

Gabriel Emile Hine
gabriel.hine@uniroma3.it

Patrizio Campisi
patrizio.campisi@uniroma3.it

[1] Department of Industrial, Electronic and Mechanical Engineering, Roma Tre University, Via V. Volterra 62, 00146 Rome, Italy

[2] Remote Sensing Technology Institute, German Aerospace Center (DLR), 82234 Wessling, Germany

Published online: 14 August 2023

possibility for an attacker to exploit the collected information for improper purposes, such as the tracking of the user's activity across multiple applications whose access control mechanisms rely on the same biometric trait [16]. In order to significantly mitigate such risks, several biometric template protection (BTP) approaches have been proposed in literature. Biometric cryptosystems are among the most effective BTP solutions. They are based on the combination of biometric representations with binary cryptographic keys, to generate template representations, namely *helper data*, which do not leak information about neither of the two original components [31]. When such methods are implemented, their robustness against attacks depends on the mutual statistical independence of the biometric template coefficients. Unfortunately, this aspect is often neglected when designing the feature extraction mechanisms, with the consequence that the actual security of the proposed applications is commonly much lower than the theoretical one [33].

Within this framework, this paper focuses on the design of methods to automatically learn biometric representations with mutual independent coefficients, while not affecting the discriminability of the considered representation. In more details, this paper stems from the preliminary work of the same authors in [21], where this aspect has been taken into account for the first time. Specifically, several metrics have been introduced in [21] to provide quantitative evaluations on the mutual statistical independence of the coefficients in a biometric template, and some attempts have been there made to improve such measurements for representations derived from finger-vein patterns. The present contribution advances the state of the art with respect to what in [21], in the following terms:

- the effectiveness of the proposed evaluation metrics and approaches to extract independent features, applied to finger-vein patterns in [21], is applied also to palm-vein patterns. Different databases are therefore here taken into account;
- an additional loss, already employed in literature for purposes different from the one here considered, is used to train convolutional autoencoders in the performed experimental tests;
- a novel loss, specifically designed to train autoencoders at learning mutually independent features, and based on the employed independence evaluation metrics, is here introduced and tested on the considered biometric modalities.

The paper is organized as follows: general information about BTP schemes, as well as specific information about the biometric cryptosystem here taken into account, are given in Sect. "Biometric Template Protection", where the importance of generating biometric representations with mutually independent coefficients is also highlighted. The metrics employed to quantitatively evaluate the independence of a representation, introduced in [21], are discussed in Sect. "Statistical Independence Metrics". The approaches here employed to generate the desired biometric representations are then introduced in Sect. "Biometric Representations". The tests performed to assess the effectiveness of the proposed methods are presented in Sect. "Experimental Tests", while some conclusions derived from the obtained results are eventually drawn in Sect. "Conclusions".

Biometric Template Protection

Along with the several advantages offered by biometric data, in comparison with traditional approaches relying on passwords or tokens, there are also several concerns that should be carefully taken into account when implementing an automatic recognition system based on personal traits. As already remarked, in case an attacker is able to fraudulently collect a biometric trait, it would be then possible to track the activities of its legitimate owner across different domains [29]. Moreover, the compromised data cannot be used anymore, and since the number of available biometric traits is limited, and it is impossible to revoke or reissue them, losing control over our own biometric data is a highly undesirable event. In addition, biometric traits could be also analyzed to reveal sensitive information about their owners, and therefore exploited for discriminatory purposes [8].

Actually, the EU General Data Protection Regulation (GDPR) states that biometric traits are sensitive and personal data, and should be therefore processed ensuring adequate levels of security. It is also worth observing that, for an attacker, collecting the templates stored in a database could be as effective as acquiring the original biometric data. In fact, it has been shown for several biometric identifiers that the original biometric traits can be adequately reconstructed from their representations, namely the templates [28], and that such reverse process can be also performed when the employed features are obtained through the use of neural networks [20].

When designing a biometric recognition system, it is therefore of paramount importance to take proper countermeasures in order to address the aforementioned issues.

A simple solution that could be employed to protect the templates stored in a database is to encrypt the data using some cryptographic algorithm. The downside of such approach is that the employed templates should be decrypted during the recognition process, which represents a system vulnerability [31]. Homomorphic encryption has been exploited to tackle the aforementioned disadvantage, by implementing the recognition step in the encrypted domain, without exposing the templates. While such solution could be effective in protecting the employed biometric data,

the computational complexity of the involved processing is usually quite demanding, and therefore not suitable for many applications. Furthermore, resorting to homomorphic encryption commonly involves the availability of secure servers which are responsible to manage the data exchange, which represent another constraint to be considered in practical scenarios.

The design of biometric template protection approaches have emerged as viable alternatives to the use of homomorphic encryption to ensure the secure and private handling of biometric traits during the recognition process. In general, these methods generate a protected template, that does not leak any information about the original data. The recognition process could be then carried out in such secure domain, thus protecting the data during the whole recognition process. The properties that a BTP scheme should satisfy, according to the ISO/IEC 24745 standard [15], are the following:

- *Irreversibility*: given a protected template, it should not be possible to reconstruct the original biometric sample, or any unprotected representation derived from it [22];
- *Renewability*: from a given biometric sample or representation, it should be possible to issue multiple protected templates;
- *Unlinkability*: given two protected templates, generated form the same biometric sample or representation, and stored in different applications, it should not be possible to determine that they belong to the same subject;
- *Performance*: using a BTP scheme should not significantly affect the system recognition performance [27].

Traditionally, BTP schemes are distinguished into two main categories: *cancelable biometrics* [23] and *biometric cryptosystems* [22] approaches.

The former class comprises methods applying a transformation to the biometric data or their representations for their protection. The use of invertible transformations leads to *salting* approaches, whose security relies on the secret storage of the parameters defining the employed transformation. Conversely, when it is assumed that an attacker can gain knowledge about the employed transformation, *non-invertible functions* have to be necessarily taken into account to properly define a BTP scheme [24]. While cancelable biometrics have been defined for several of the most used traits, such as fingerprint [34], face [5], and iris [26] among others, their irreversibility has been rarely evaluated through exhaustive and rigorous proofs, due to the intrinsic difficulties in proving the actual non-invertibility of a function against any possible kind of attack.

On the other hand, biometric cryptosystems could be distinguished into *key-generating* approaches, which extract cryptographic keys from the considered biometric data [32], and *key-binding* methods, whose aim is to secure a cryptographic key by means of biometric data and vice versa, combining the two information sources into a binary template commonly indicated as *helper data* [11]. The former approaches commonly fail at providing proper renewability and unlinkability, since information different from that of biometric data would be required to generate multiple representations of the same trait. Key-binding approaches are instead undoubtedly the most investigated approach among all possible BTP schemes. In more detail, the security of key-binding approaches has been the object of several rigorous evaluations [29], with in-depth information theoretic studies trying to exactly evaluate the amount of knowledge about the original secret sources leaked from the templates obtained when binding the considered biometric representations with cryptographic keys [14]. It has to be remarked that the robustness against attacks perpetrated against key-binding BTP schemes has been commonly investigated under the assumption that the employed biometric representations are made of mutually independent coefficients. Under such hypothesys, the binary helper data generated in key binding scheme have maximum entropy, and the analysis about the information leakage from the stored templates can be performed considering a single coefficient and extended to draw conclusions regarding the whole set of available features. Unfortunately, the parametric representations adopted in most biometric recognition systems typically consist of strongly correlated features, with the consequence of a loss in security of the designed cryptosystem the greater the further the available data are from the ideal condition of mutual independence [33].

For this reason, when designing a key-binding biometric cryptosystem, the generation of representations having mutually statistical independent features is as important as having highly-discriminative templates. In this paper we rely on a biometric cryptosystem proposed by some of the authors in [11] and detailed in the following paragraph, where the most important aspects evaluated to assess its effectiveness are also summarized.

Considered Biometric Cryptosystem

The key-binding scheme proposed in [11] is a code-offset method inspired by the digital modulation paradigm, specifically designed in order to guarantee, under proper circumstances, no information leakage about the employed cryptographic key from the knowledge of the stored helped data. The considered BTP scheme is graphically depicted in Fig. 1.

Within this biometric cryptosystem, a secret binary key is processed together with a biometric representation $\mathbf{f} \in \mathbb{R}^m$, comprising m coefficients, made available during

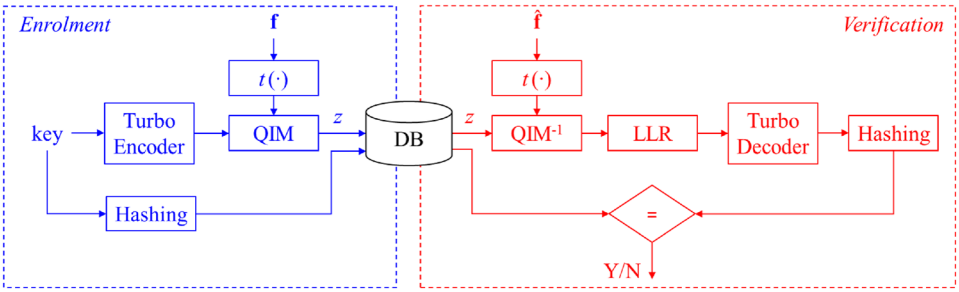


Fig. 1 Zero-leakage key-binding approach proposed in [11]

the enrolment of a legitimate user. The zero-leakage capability of the scheme is achieved through the use of a pointwise function $t(\cdot)$, applied to each coefficient of the biometric representation $\mathbf{f}$ in order to derive, from each of them, a variable with a probability density function following a raised cosine distribution, described by a roll-off parameter $\gamma \in [0, 1]$. It is actually provable that, under the assumption of mutual independence of the features in the employed representation $\mathbf{f}$, the use of such transformation guarantees that the stored helper data cannot reveal any information about the employed cryptographic key, thus implementing a zero-leakage protection scheme [11].

In order to produce the helper data to be securely stored in the system, a string q comprising m symbols belonging to a phase-shift keying (PSK) constellation of size $R \in \mathbb{N}^+$, with an alphabet $\{0, 2\pi/R, ..., (R-1)2\pi/R\}$, is obtained by encoding the input binary cryptographic key with an error correcting code. As reported in Fig. 1, turbo codes are exploited toward this aim in the employed implementation. A quantization index modulation (QIM) process is then performed to bind the considered binary key and biometric representation by computing the code-offset helper data $z = \left[t(\mathbf{f}) - q\right]_{2\pi}$.

When a user wants to be recognized by the system, a reverse process is performed. Specifically, an inverse QIM is applied to the stored helper data z, using the newly acquired biometric representation $\hat{\mathbf{f}}$. In case the new biometric representation is close to the one extracted during the user's registration, the obtained message $\hat{q}$ is similar to that of the enrolment, and could be therefore employed to retrieve the original binary key through a soft decoding process with decisions based on log-likelihood ratio (LLR) criteria. Conversely, if the verification process is carried out by an impostor, the original message, and the associated original key, cannot be reconstructed. Storing the employed keys in a hashed version allows to compare the reconstructed data with the original ones without having to reveal them, and thus taking decisions regarding the identity of the subjects involved in the recognition process.

As already mentioned, the considered scheme does not leak any information about the employed cryptographic key from the helper data z, in case of biometric representations $\mathbf{f}$ with mutually independent coefficients. Other aspects which have to be taken into account to evaluate the effectiveness of the employed key-binding approach include:

- the information leakage about the employed biometric representation from the knowledge of the stored helper data. This aspect, commonly indicated as *privacy* (P), can be evaluated by estimating the minimum reconstruction error an attacker can commit when trying to reconstruct the input biometric representation $\mathbf{f}$, given the stored helper data z. As shown in [11], the privacy of the proposed scheme improves with the use of larger values of the roll-off γ of the employed raised cosine distribution;
- the capacity (C) of the employed biometric representation. This aspect takes into account the maximum size of the cryptographic key which can be bound with the employed biometric representation. Obviously, the more bits can be embedded into the considered template, the higher the robustness of the system against brute-force attacks. Large values of the roll-off parameter γ negatively influence the achievable capacity.

Since the privacy and the capacity of the proposed scheme are in a trade-off relation depending on the choice of γ, an iterative selection strategy has been proposed in [11], with the used parameter determined as the one maximizing the achievable capacity while guaranteeing a minimum level of acceptable privacy, typically set in the range from 95 to 99%.

Statistical Independence Metrics

The statistical independence metrics proposed in [21] are defined according to the framework reported in Fig. 2. Specifically, the Hilbert–Schmidt Independence Criterion (HSIC) statistical test [10] is employed to compute measures associated to each pair of available features in the considered representation. Metrics derived from graph theory [4] are then employed to provide the desired quantitative

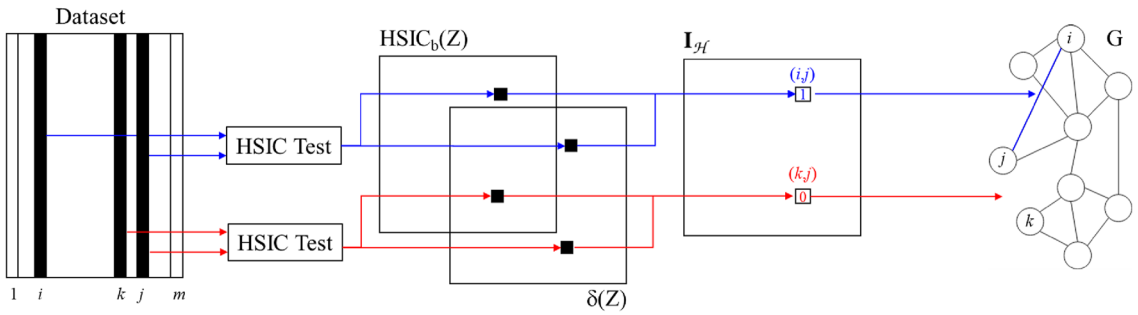


Fig. 2 Visual depiction of the employed framework for statistical independence evaluation, adapted from [21]

evaluations about the overall level of independence for the coefficients in the analyzed representation.

In more detail, to evaluate the independence of a given representation, it is assumed that a dataset of biometric templates, each expressed as a feature vector with length m and collected from u subjects, for a total of n samples, is available. As depicted in Fig. 2, such data can be arranged as an $n \times m$ matrix, with each row being a biometric template.

Given any pair of two features, represented by the random variables $\mathcal{X}$ and $\mathcal{Y}$, an HSIC test estimates the squared Hilbert-Schmidt norm of the population of interest, that is, HSIC($\mathbf{P}_{xy}, \mathcal{F}, \mathcal{G}$), where $\mathbf{P}_{xy}$ is the joint distribution of $\mathcal{Z} = (\mathcal{X}, \mathcal{Y})$, and $\mathcal{F}$ and $\mathcal{G}$ are two reproducing kernel Hilbert spaces (RKHS). The null and research hypotheses of the HSIC test are defined as follows:

$$\mathcal{J}(\mathcal{Z}) : (\mathcal{X} \times \mathcal{Y})^n \mapsto 0, 1, \tag{1}$$

$$\begin{aligned} \mathrm{H}_0 &: \mathbf{P}_{xy} = \mathbf{P}_x \mathbf{P}_y \\ \mathrm{H}_1 &: \mathbf{P}_{xy} \neq \mathbf{P}_x \mathbf{P}_y, \end{aligned} \tag{2}$$

that is, in case of two independent random variables $\mathcal{X}$ and $\mathcal{Y}$, the null hypothesis cannot be rejected.

Given a sample of observations $Z = (X, Y)$, the statistics employed in the used framework are estimated in a biased version as follows:

$$\mathrm{HSIC}_b(Z) = \frac{1}{n^2} \mathrm{trace}\,(\mathbf{KHLH}), \tag{3}$$

that is, through an operator that computes the sum of the elements on the main diagonal of a square matrix obtained as the product of the $n \times n$ matrices $\mathbf{H} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$, $\mathbf{K}$, and $\mathbf{L}$, where $\mathbf{1}$ represents a $n \times 1$ vector of ones, while the other two matrices are defined as follows:

$$\begin{aligned} \mathbf{K}[i,j] &:= \exp\left(-\sigma_x^{-2} \|x_i - x_j\|^2\right) \\ \mathbf{L}[i,j] &:= \exp\left(-\sigma_y^{-2} \|y_i - y_j\|^2\right); \quad i,j = 1, \ldots, n, \end{aligned} \tag{4}$$

with σ_x^2 and σ_y^2 representing the variances of the considered coefficients.

Having set a significance level α as upper bound of the type I error, the asymptotic distribution of the empirical estimate $\mathrm{HSIC}_b(Z)$ is derived under H_0, and the quantile $1 - \alpha$ of this distribution, indicated as $\delta(Z)$, can be used as a threshold to determine the test outcome. The null hypothesis H_0 cannot be rejected when $\mathrm{HSIC}_b(Z) < \delta(Z)$, with the two random variables $\mathcal{X}$ and $\mathcal{Y}$ assumed in this case as independent.

Once the HSIC test is performed for every possible couple of coefficients in the available m-dimensional biometric representation, an $m \times m$ square and symmetrical independence matrix $\mathbf{I}_{\mathcal{H}} \in \mathbb{Z}^{m \times m}$ is obtained as follows:

$$\mathbf{I}_{\mathcal{H}}[i,j] = \begin{cases} 1 \text{ if } \mathrm{HSIC}_b(F_i, F_j) < \delta(F_i, F_j) \\ 0 \text{ otherwise,} \end{cases} \tag{5}$$

where F_i and F_j represent any two possible features, $1 \leq i, j \leq m$.

The obtained binary independence matrix $\mathbf{I}_{\mathcal{H}}$ can be then interpreted as an adjacency matrix $\mathbf{A}_G$, associated to an independence undirected graph $G = \{V, E\}$, whose edges E connect the nodes V in case the corresponding coefficients are mutually independent. With the graph thus created, it is possible to define several metrics expressing the overall level of independence of a given representation, by exploiting concepts stemming from graph theory. The following metrics have been proposed in [21] and are employed in the following discussion to compare the approaches here proposed to generate representations with independent coefficients:

- *Normalized edge count*: number of edges in the graph G, normalized with respect to the maximum number of edges in a complete graph with the same number m of nodes as G, that is, $\mathrm{NEC}_m = \frac{1}{m(m-1)} \sum_{i,j} \mathrm{I}_{\mathcal{H}}[i,j]$. The computed value can be interpreted as a percentage of independent coefficients. This metric is quite simple to compute, yet it may convey information of little value. For instance, a value of $\mathrm{NEC}_m = 0.9$ does not mean that 90% of the features are mutually statistically independent, but

only that the independence matrix $\mathrm{I}_{\mathcal{H}}$ contains 90% of unitary entries;

- *Normalized maximum clique size*: a clique of G is defined as a complete subgraph such that every two distinct nodes in it are adjacent. A clique is also said to be *maximal* if it is not a subset of another clique, and *maximum* if it has the largest number of nodes. Given a graph G obtained from an independence matrix $\mathrm{I}_{\mathcal{H}}$, the size S of its maximum clique can be employed as a metric for the level of representation independence, once normalized with respect to the largest possible value, that is, $\mathrm{NMCS}_m = S/m$. This metric gives an effective measure of independence, since the features in the maximum clique are actually mutually independent. Nonetheless, it may often result in very low values, with consequent difficulties in performing comparisons through it among different approaches trying to maximize feature independence. Moreover, its computation could require significant processing time, especially when the number of considered features m is large. Furthermore, multiple maximum cliques can be derived from a single graph, making it hard to understand which of them is the best one;
- *Normalized degree centrality*: the degree centrality of a node i in a graph G is computed as the number $d_G(i)$ of edges incident to the node itself. This quantity expresses the importance of each node within the graph, and the level of independence of each coefficient from the others in the considered scenario. It can be computed in a normalized form dividing it by the maximum feasible degree of the graph, that is, $\mathrm{NDC}_m(i) = \frac{d_G(i)}{m-1}$. Instead of summing the values thus obtained for each node to obtain a single overall measure, in the performed tests the computed metrics are organized in a descending order to form a curve, thus showing the deviation from the ideal scenario with only mutually-independent features, whose normalized centrality is 1 for all the available nodes. Such expression makes it easier to have an indication on the number of the more important nodes, that is, the features independent of most other coefficients.

Biometric Representations

The scenario taken into account to test the strategies designed to generate biometric representations with mutually independent coefficients involves the use of deep learning strategies to process hand vein patterns. In more detail, as specifically illustrated in Sect. "Experimental Tests", biometric representations from both finger-vein and palm-vein patterns are here generated through the approaches described in this section.

Hand-vein-based biometric recognition systems rely on the uniqueness of the vessel patterns of our wrists, palms, fingers, and hand dorsa [37]. Thanks to the absorption properties of the haemoglobin, it is in fact possible to acquire images depicting subcutaneous vein patterns through non-invasive and contactless devices, by simply illuminating them with near-infrared (NIR) light comprising wavelengths between 700 and 900 nm [30]. Capturing devices working in either transmission or reflection modality could thus produce images where blood vessels appear dark, with the surrounding tissue, that let the light passing, being instead brighter [25].

The approach proposed to create representations with mutually independent coefficients, still maintaining proper discriminative capabilities is depicted in Fig. 3, and relies on a cascade neural network, composed by the following:

- a baseline system, whose aim is to generate representations suitable to be used for verification tasks in open-set conditions;
- a densely-connected convolutional autoencoder (DCCAE), whose purpose is to estimate an inner representation of the features produced by the baseline system, while maximizing the mutual independence of the derived coefficients.

In more detail, since the baseline system should process vein patterns to create discriminative templates to be used in a verification application, this component is designed following the approach proposed in [17], that is, using a convolutional neural network (CNN) derived from DenseNet-161 [12], with the addition of a custom set of layers as reported in Table 1, creating a biometric representation **v** comprising

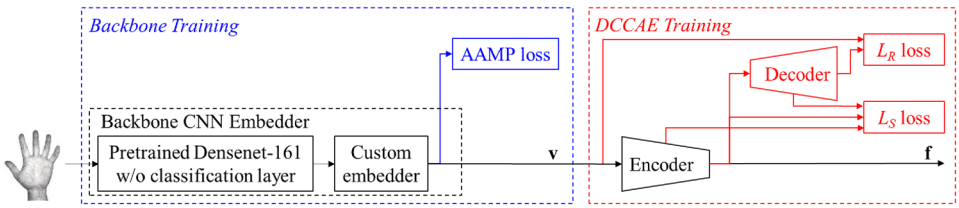


Fig. 3 Proposed approach based on autoencoders to create biometric representations with mutually-independent coefficient while maintaining discriminative capabilities

Table 1 Backbone CNN embedder derived from Densenet-161

Layers		Input size	Output size
Convolution	7×7 conv, str.2	$224 \times 224 \times N_c$	$112 \times 112 \times 96$
Pooling	3×3 max pool, str.2	$112 \times 112 \times 96$	$56 \times 56 \times 96$
Dense block 1	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$56 \times 56 \times 96$	$56 \times 56 \times 384$
Transition 1	1×1 conv	$56 \times 56 \times 384$	$28 \times 28 \times 192$
	2×2 avg pool, str.2		
Dense block 2	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$28 \times 28 \times 192$	$28 \times 28 \times 768$
Transition 2	1×1 conv	$28 \times 28 \times 768$	$14 \times 14 \times 384$
	2×2 avg pool, str.2		
Dense block 3	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 36$	$14 \times 14 \times 384$	$14 \times 14 \times 2112$
Transition 3	1×1 conv	$14 \times 14 \times 2112$	$7 \times 7 \times 1056$
	2×2 avg pool, str.2		
Dense block 4	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$	$7 \times 7 \times 1056$	$7 \times 7 \times 2208$
Custom embedder	7×7 global avg pool	$7 \times 7 \times 2208$	1×2208
	Batch normalization		
	Dropout (50%)		
	Fully connected layer	1×2208	1×1024
	Batch normalization		

1024 features. The baseline system is trained using a cross-entropy function with additive angular margin penalty (AAMP) [9] as objective loss, in order to define representations with proper discriminative capabilities, that could be used also for subjects distinct from those whose traits are employed during the training process [17].

Once features capable of discriminating among different subjects have been defined, the proposed approach tries to estimate an alternative representation with the additional characteristic of independence. Autoencoders are employed for this task, leveraging on their ability to automatically learn efficient encodings of the input data, with the aim of guaranteeing that the inner representations have some specific property, such as sparsity or compactness, while keeping all the informative content of the considered input. Specifically, the autoencoder employed in the performed tests is the DCCAE proposed in [18]. As reported in Table 2, it consists of a total of 55 layers, and has an inner encoding **f**

Table 2 Employed DCCAE

h	Layer		Input size	Output size
ENCODER				
1	Input layer	1×3 Conv	1×1024	16×1024
2-6	Dense block 1	$\begin{bmatrix} \text{Batchnorm, ReLU} \\ 1 \times 3 \text{ Conv} \end{bmatrix} \times 5$	16×1024	80×1024
7	Transition 1	Batchnorm, ReLU 1×3 Conv, str.2	80×1024	32×512
8-12	Dense block 2	$\begin{bmatrix} \text{Batchnorm, ReLU} \\ 1 \times 3 \text{ Conv} \end{bmatrix} \times 5$	32×512	80×512
13	Transition 2	Batchnorm, ReLU 1×3 Conv, str.1	80×512	64×512
14-18	Dense block 3	$\begin{bmatrix} \text{Batchnorm, ReLU} \\ 1 \times 3 \text{ Conv} \end{bmatrix} \times 5$	64×512	80×512
19	Transition 3	Batchnorm, ReLU 1×3 Conv, str.2	80×512	32×256
20-24	Dense block 4	$\begin{bmatrix} \text{Batchnorm, ReLU} \\ 1 \times 3 \text{ Conv} \end{bmatrix} \times 5$	32×256	80×256
25	Transition 4	Batchnorm, ReLU 1×3 Conv str.2	80×256	16×128
26	Hidden encoder	Fully-connected Layer Batchnorm, Sigmoid	1×2048	1×512
27	Latent encoder	Fully-connected Layer Batchnorm, Sigmoid	1×512	1×256
DECODER				
28	Latent Decoder	Fully-connected Layer Batchnorm, Sigmoid	1×256	1×512
29	Hidden Decoder	Fully-connected Layer Batchnorm, Sigmoid	1×512	1×2048
30	Transition 1	Batchnorm, ReLU 1×3 Tran-conv, str.2	16×128	32×256
31-35	Dense block 1	$\begin{bmatrix} \text{Batchnorm, ReLU} \\ 1 \times 3 \text{ Tran-conv} \end{bmatrix} \times 5$	32×256	80×256
36	Transition 2	Batchnorm, ReLU 1×3 Tran-conv, str.2	80×256	64×512
37-41	Dense block 2	$\begin{bmatrix} \text{Batchnorm, ReLU} \\ 1 \times 3 \text{ Tran-conv} \end{bmatrix} \times 5$	64×512	80×512
42	Transition 3	Batchnorm, ReLU 1×3 Tran-conv, str.1	80×512	32×512
43-47	Dense block 3	$\begin{bmatrix} \text{Batchnorm, ReLU} \\ 1 \times 3 \text{ Tran-conv} \end{bmatrix} \times 5$	32×512	80×512
48	Transition 4	Batchnorm, ReLU 1×3 Tran-conv, str.2	80×512	16×1024
49-53	Dense block 4	$\begin{bmatrix} \text{Batchnorm, ReLU} \\ 1 \times 3 \text{ Tran-conv} \end{bmatrix} \times 5$	16×1024	80×1024
54	Transition 5	Batchnorm, ReLU 1×3 Tran-conv, str.2	80×1024	16×1024
55	Output layer	1×3 Tran-conv	16×1024	1×1024

consisting of 256 coefficients, derived from the 1024 coefficients of the input $\mathbf{v}$.

The employed DCCAE is trained by trying to minimize a loss function defined as $L = L_R + \beta \cdot L_S$, with β being an hyperparameter, and L_R representing the reconstruction loss, computed through the cosine dissimilarity

$$L_R = \frac{1}{B}\sum_{i=1}^{B}[1 - cos(\mathbf{v}_i, \hat{\mathbf{v}}_i)], \tag{6}$$

being $\mathbf{v}_i$ the i-th feature representation generated by the baseline CNN, $\hat{\mathbf{v}}_i$ its counterpart reconstructed by the autoencoder, and B the employed batch size.

The component L_S of the autoencoder loss is instead defined with the aim of learning inner representations with mutually independent coefficients. In the performed tests, such purpose is sought through the use of several different approaches, relying on the following:

- a loss L_S based on the Kullback–Leibler divergence (KLD), defined as

$$L_S^{\text{KLD}} = \sum_{h\in\mathcal{L}}\sum_{j=1}^{N^{(h)}} D_{KL}\left(\rho|\hat{\rho}_j^{(h)}\right),\ \hat{\rho}_j^{(h)} = \frac{1}{B}\sum_{i=1}^{B}\left[a_j^{(h)}(\mathbf{v}_i)\right], \tag{7}$$

  where $a_j^{(h)}$ is the j-th activation output of the h-th hidden layer of the DCCAE when $\mathbf{v}_i$ is fed as input to the DCCAE, with $j = 1, \ldots, N^{(h)}$, being $N^{(h)}$ the number of activation units in the h-th hidden layer, and $\rho \in [0, 1]$ is the sparsity parameter. The set $\mathcal{L}$ represents the layers of the DCCAE dedicated to the inner encoder and decoder, with $\mathcal{L} = \{26 - 29\}$ for the DCCAE reported in Table 2;
- a loss L_S based on the *L1* distance, defined as

$$L_S^{L1} = \sum_{h\in\mathcal{L}}\sum_{j=1}^{N^{(h)}} |a_j^{(h)}|, \tag{8}$$

  where, similarly to L_S^{KLD}, only the activation outputs of the last two layers in the encoder, and the first two in the decoder, are considered to compute the desired loss;
- a loss L_S based on the *spectral restricted isometry property* (SRIP) [1], computed on the weights of each convolutional layer of the proposed DCCAE as

$$L_S^{\text{SRIP}} = \sum_{h=1}^{55}\sigma(W^{(h)})\left(W^{(h)\top}W^{(h)} - I\right), \tag{9}$$

  where $W^{(h)\top}$ is a matrix with the weights of the h-th layer, I is the identity matrix and σ is the spectral norm, defined as the largest singular value of $W^{(h)}$. Such loss forces the weights of the network to be near-orthogonal, with the possibility to thus make mutually independent the coefficients of the learned encoding;
- a loss L_S based on *DeCov* [7] regularization, that is, an approach whose aim is to minimize the cross-covariance of hidden activations through a regularization operation. Considering the h-th layer of the employed DCCAE that generates the inner encodings, and its activations $a_j^{(h)}$, the interested cross-covariance $\mathbf{C}$ is obtained by computing, for all the possible pairs of activations j and k,

$$\mathbf{C}[j,k] = \frac{1}{B}\sum_{i=1}^{B}\left(a_j^{(h)}(\mathbf{v}_i) - \mu_j\right)\left(a_k^{(h)}(\mathbf{v}_i) - \mu_k\right), \tag{10}$$

  being μ_j the sample mean of activation j over the batch, that is,

$$\mu_j = \frac{1}{B}\sum_{i=1}^{B} a_j^{(h)}(\mathbf{v}_i). \tag{11}$$

  The DeCov loss is then computed as

$$L_S^{DeCov} = \frac{1}{2}(\|\mathbf{C}\|_F^2 - \|\text{diag}(\mathbf{C})\|_2^2), \tag{12}$$

  where $\|\cdot\|_F$ is the Frobenius norm. This loss should allow to learn non-redundant representations, therefore possibly improving mutual independence;
- a loss L_S based on the HSIC statistical test discussed in Sect. "Statistical Independence Metrics". This novel loss is here specifically proposed with the aim to generate representations optimizing the independence metrics presented in Sect. "Statistical Independence Metrics", by including them into the employed loss functions. For a batch of samples considered during network training, the HSIC global statistics in Eq. (3) is computed for all possible pair of coefficients in the inner encoding created within the autoencoder, and a loss is computed as

$$L_{\text{HSIC}} = \sum_{j=1}^{256}\sum_{k=1,k\neq j}^{256}\text{HSIC}_b(\mathbf{f}_j, \mathbf{f}_k), \tag{13}$$

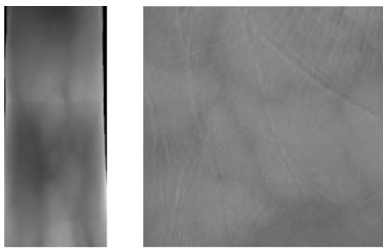
  with the inner encodings of the considered DCCAE autoencoder corresponding to the activations of the 27th layer, that is, $\mathbf{f}_j = a_j^{(27)}$. In the employed implementation of this loss, the variances of the coefficients required to compute the HSIC statistics as reported in Eq. (4) are set as hyperparameters, due to the inaccurate estimates that would be otherwise obtained when considered limited batch sizes during the training of the autoencoder.

Experimental Tests

As already remarked, the approaches here proposed for the generation of biometric representations with mutually independent features have been tested considering recognition systems relying on hand vein patterns. In more details, both palm-vein and finger-vein traits have been taken into account, exploiting, respectively, the palm vein samples from the PolyU-P multispectral dataset [36] and the finger vein data from the SDUMLA dataset [35]. The PolyU-P database contains left and right palm vein images collected from 250 subjects, with 6 samples collected during each of 2 recording sessions for each user. The SDUMLA database contains finger-vein images of 636 fingers from 106 subjects. Six images have been acquired during a single session from each of the left and right hand's index, middle and ring fingers in gray level BMP format with a resolution of 320× 240 pixels. For each database, different fingers and different hands of the users are treated as distinct classes. The available data have been divided into two disjoint datasets of equal size, one used to train the considered architectures and the other one to perform the required evaluations, with 20% of the training data reserved for validation in all datasets. Recognition results have been computed considering an open-set verification scenario, using half of the available classes during training, and the remaining half for testing. Since the PolyU-P database contains data collected in multiple sessions, data for enrolment and verification have been taken from different sessions to avoid the bias effect. A summary of the characteristics of the employed databases is reported in Table 3, together with details on the applied experimental protocols. Examples of samples taken from the two datasets are depicted in Fig. 4.

When applying the performed processing, all the considered samples have been re-sized into 224×224 pixels before being fed to the employed networks, whose first stage is given by a Densenet-161 architecture, and normalized to zero mean and unit variance. The employed networks have been trained using stochastic gradient descent with momentum (SGDM) and a batch size of 64. Tests have been performed using PyTorch 1.1.0, with a system configuration of 32Gb RAM, two NVIDIA^{TM} Titan V graphics cards, i7-3.4GHz processors, Windows^{TM} 10 OS.

The results of a comparative analysis among the methods presented in Sect. "Biometric Representations", in terms of capability to create templates with independent coefficients expressed through the NEC_m and $NMCS_m$ metrics, are reported in Tables 4 and 5, respectively for tests performed on the SDUMLA and the PolyU-P databases. The metrics



(a) (b)

Fig. 4 Examples of considered vein images. **a** Finger-vein sample from SDUMLA; **b** Palm-vein sample from PolyU-P

Table 3 Used vein databases and corresponding experimental protocols

Benchmark database	Vein modality	Database statistics		Capturing conditions	Capturing parts	Summary of the experimental protocol
SDUMLA [35]	Finger	# of Subjects	106	Grayscale single channel	Index-, middle- and ring-fingers from left and right hand	Available classes divided into two equal-size subsets, respectively employed for training and testing
		# of Classes	636			
		# of Sessions	1			
		Samples per session	6			
		Total samples	3.816			
PolyU-P [36]	Pam	# of Subjects	250	4 different spectral channels	Left and right hand palm	Available classes divided into two equal-size subsets, respectively employed for training and testing. Enrolment performed on session-1 data, verification on session-2 data. Samples in the NIR channel used in the experiments
		# of Classes	500			
		# of Sessions	2			
		# per Session	6			
		Total Samples	24.000			

NDC_m are instead reported in Fig. 5, which shows the behaviors obtained for all the considered nodes (coefficients), in order to better illustrate the deviation from the ideal conditions with all values set to 1.

All the aforementioned results have been obtained by choosing the hyperparameters of the employed DCCAE autoencorders, for each considered loss L_S, with the aim of guaranteeing the best achievable performance in terms of independence of the generated representations. A significance level $\alpha = 2.5\%$ has been employed when performing the HSIC tests required to compute the independence metrics described in Sect. "Statistical Independence Metrics". The employed independence metrics have been evaluated also for templates obtained applying independent component analysis (ICA) [13] to the features **v** generated from the used backbone CNN embedder, with a standard approach such as ICA considered as an alternative to the proposed autoencoder-based method for the creation of biometric representations with independent coefficients. In more detail, the FastICA approach, which applies an orthogonal rotation to prewhitened data in order to maximize a measure of non-Gaussianity, is employed in the performed tests.

It is possible to observe that the proposed approaches relying on KLD, SRIP, and HSIC losses are able to provide an improvement in terms of NEC_m, NDC_m, and $NMCS_m$ with respect to the use of representations obtained through the baseline networks, for both SDUMLA and PolyU-P. Nonetheless, an ICA transformation still guarantees a slightly better independency. Yet, the use of DeCov or L1 losses within the proposed DCCAE-based feature generation approach allows to reach even further improvements over ICA, with the method relying on DeCov representing by far the best solution to create templates with

Table 4 Comparative analysis of the statistical independence for biometric representations created from finger-vein samples of the SDUMLA database

Metric	Baseline	ICA	DCCAE (**f**)				
	(**v**)		KLD	L1	SRIP	DeCov	HSIC
NEC_m	0.152	0.407	0.378	0.418	0.385	**0.777**	0.165
$NMCS_m$	6	9	9	10	9	**26**	6

Best results reported in bold

Table 5 Comparative analysis of the statistical independence for biometric representations created from palm-vein samples of the PolyU-P database

Metric	Baseline	ICA	DCCAE (**f**)				
	(**v**)		KLD	L1	SRIP	DeCov	HSIC
NEC_m	0.161	0.322	0.238	**0.610**	0.181	0.604	0.330
$NMCS_m$	7	9	7	15	6	**16**	8

Best results reported in bold

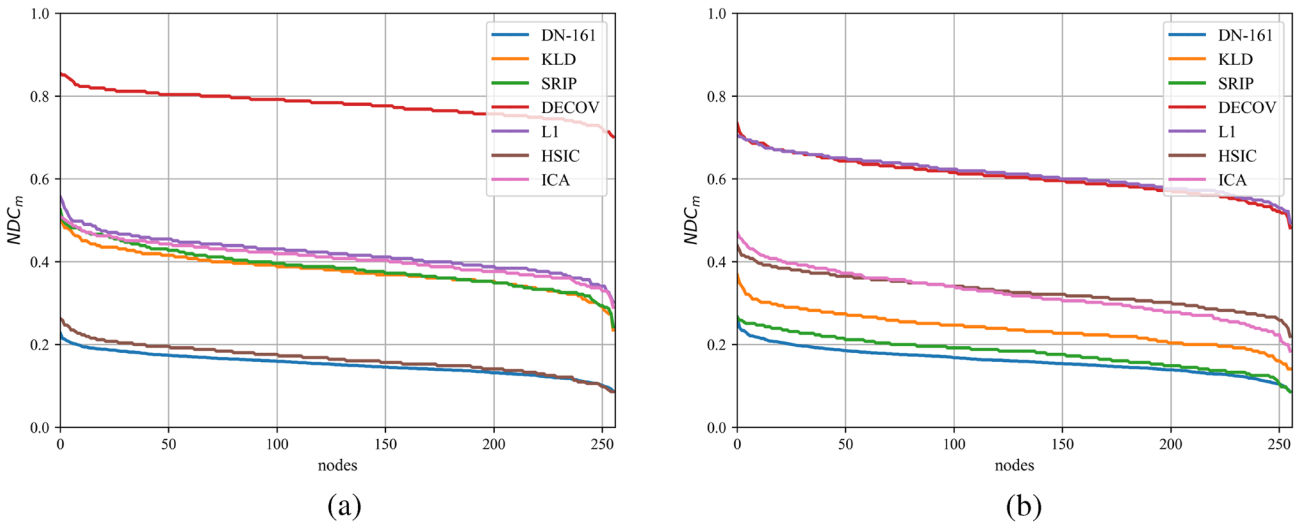


Fig. 5 Normalized degree centrality (NDC_m) computed for the considered representations. **a** SDMULA; **b** PolyU-P

independent coefficients for SDUMLA, and on par with L1 on PolyU-P.

In addition to the analysis of the proposed template generation methods in terms of achievable feature independence, other aspects relevant to the design of biometric cryptosystems are also evaluated. The recognition performance, expressed in terms of false rejection rate (FRR) and false acceptance rate (FAR), attainable exploiting the biometric representations generated through the proposed approaches, are shown by the detection error trade-offs (DET) curves of Fig. 6. It can be noticed that the use of SRIP, KLD, and HSIC losses in the proposed DCCAE guarantees results similar to those of the baseline network, while resorting to DeCov and L1 losses, and even to an ICA transformation, may notably affect the achievable recognition rates. It is worth remarking that the reported results have been obtained when training the proposed DCCAE-based architectures with the aim of maximizing the feature independence of the created representations, with the attainable performance therefore not involved when learning the parameters of the employed DCCAE. The observed behavior confirms the trade-off commonly present in a biometric cryptosystem, where an improvement in terms of security (here expressed through the achieved independence) can be achieved at the cost of a worsening in terms of recognition performance.

It is worth remarking that, in case the produced representations are employed within a biometric recognition system such as the one here considered and summarized in Sect. "Considered Biometric Cryptosystem", the zero-leakage requirements necessarily sets the system at the ZeroFAR, that is, the operative condition at which FAR = 0%. Given the plots in 6, and the independence results in Fig. 5 and in Tables 4 and 5, the DeCov loss can be held as the most reliable choice to create templates with independent coefficients while guaranteeing acceptable FRR values when FAR = 0%.

The achievable system security can be also evaluated by estimating the capacities of the employed representations, reported in Fig. 7 in terms of the average number of bits of the binary string that can be embedded within each coefficient of the generated representation. As detailed in [11] and mentioned in Sect. "Considered Biometric Cryptosystem", the achievable capacity depends on the roll-off parameter γ of the raised cosine distribution employed in the considered biometric cryptosystem. A trade-off between different aspects can be also seen in the graphs, with the methods providing the best performance in terms of achievable independence having the lower capacity values. Such relationship can be clearly exposed by plotting in Fig. 8 the average capacity measured for $\gamma = 0$ (highest possible capacity values) against the achievable independence, expressed in terms of NEC_m, for all the proposed DCCAE-based approaches for representation learning. As can be seen, there is a monotone decreasing trend for the achievable capacity as long as the independence of the generated features increases.

Conclusions

In this paper, an analysis on the possibility of generating biometric representations with independent coefficients has been conducted. To this aim, an approach relying on autoencoders, which could be trained according to different loss functions, has been proposed. The performance of the proposed methods have been evaluated according

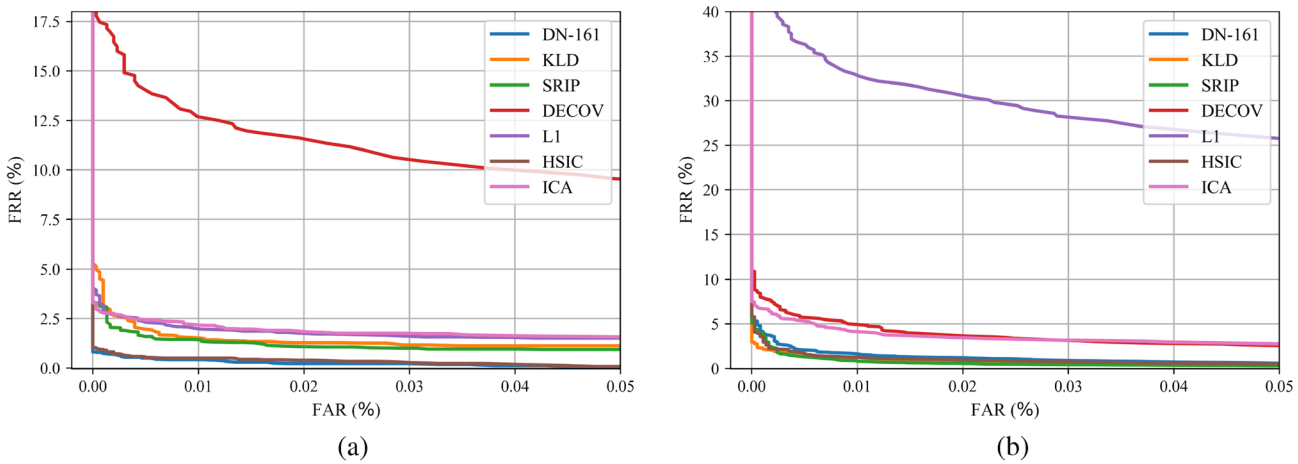


Fig. 6 DET curves reporting the recognition performance achievable with the considered representations. **a** SDUMLA; **b** PolyU-P

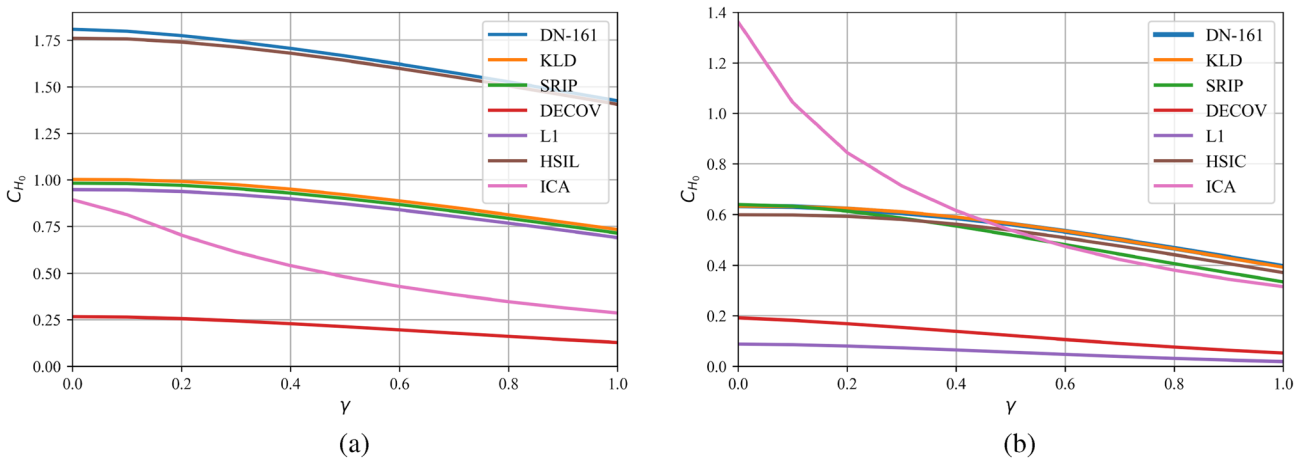


Fig. 7 Average embedding capacity achievable with the considered representations. **a** SDUMLA; **b** PolyU-P

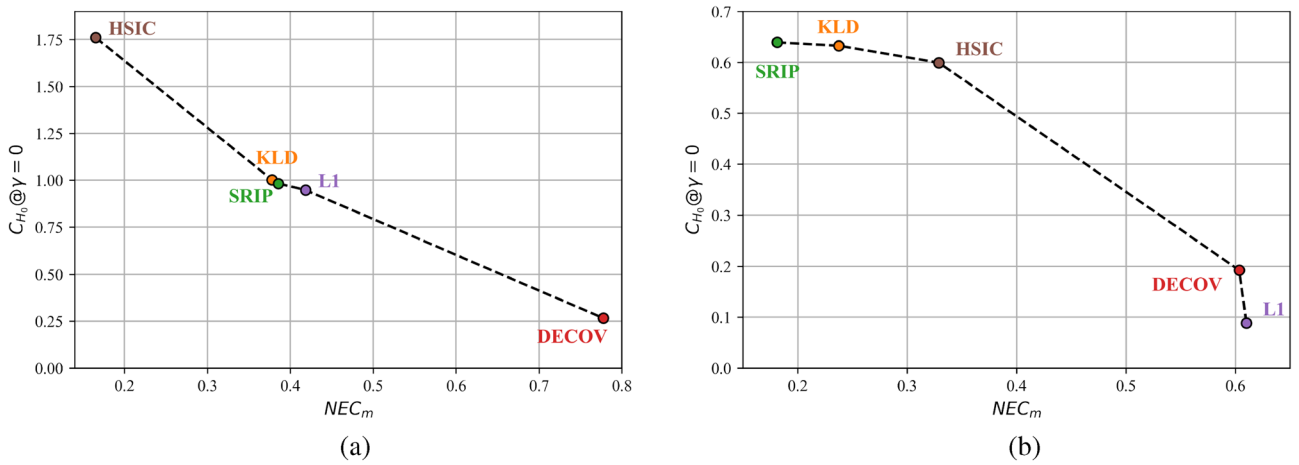


Fig. 8 Trade-off between channel capacity (average number of bits embedded for $\gamma = 0$) and independence (expressed in terms of NEC_m). **a** SDUMLA; **b** PolyU-P

to metrics specifically defined to evaluate the achievable feature independence. Experimental tests conducted over two distinct biometric databases, containing samples of finger-vein and palm-vein patterns, have highlighted that the proposed approaches are actually able to notably increase the independence of the employed representations while maintaining proper discriminative capabilities. In more detail, two of the considered loss functions allow to generate biometric representation with higher independence than what could be achieved when resorting to ICA transformations. It has yet been observed that the sought feature independence is typically in a trade-off relationship with both the attainable recognition rates and the average embedding capacity.

Funding Open access funding provided by Università degli Studi Roma Tre within the CRUI-CARE Agreement.

Data availability The raw data employed for the experimental tests are taken from two public databases, referenced in the manuscript.

Declarations

Conflict of Interest The authors declare that no funds, grants, or other support were received during the preparation of this manuscript. The authors also declare that there are no conflicts of interest. This article does not contain any studies with human participants performed by any of the authors.

References

1. Bansal N, Chen X, Wang Z. Can we gain more from orthogonality regularizations in training deep cnns? In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. Curran Associates Inc., Red Hook, NIPS'18, 2018; pp. 266–4276
2. Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell. 2013;35(8):1798–828.
3. Bhanu B, Kumar A. Deep learning for biometrics. Cham: Springer; 2017.
4. Bondy J, Murty U. Graph theory. 1st ed. Incorporated: Springer Publishing Company; 2008.
5. Boult TE, Scheirer WJ, Woodworth R. Revocable fingerprint biotokens: accuracy and security analysis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2007; pp 1–8
6. Campisi P. Security and privacy in biometrics. Cham: Springer; 2013.
7. Cogswell M, Ahmed F, Girshick R, et al. Reducing overfitting in deep networks by decorrelating representations. 1511.06068, 2016.
8. Dantcheva A, Elia P, Ross A. What else does your biometric data reveal? a survey on soft biometrics. IEEE Trans Inf Forensics Secur. 2016;11(3):441–67.
9. Deng J, Guo J, Xue N, et al. Arcface: Additive angular margin loss for deep face recognition. 1801.07698, 2019.
10. Gretton A, Fukumizu K, Teo CH, et al. A kernel statistical test of independence. In: Proceedings of the 2007 Conference on Advances in Neural Information Processing Systems 2007.
11. Hine G, Maiorana E, Campisi P. A zero-leakage fuzzy embedder from the theoretical formulation to real data. IEEE Trans Inf Forens Secur. 2017;12(7):1724–34.
12. Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks. 1608.06993, 2018.
13. Hyvarinen A, Oja E. Independent component analysis: algorithms and applications. Neural Netw. 2000;13(4–5):411–30.
14. Ignatenko T, Willems F. Fundamental limits for privacy-preserving biometric identification systems that support authentication. IEEE Trans Inform Theory. 2015;61(10):5583–94.
15. ISO, IEC JTC1 SC27 Security Techniques,. ISO/IEC 24745:2011. International Organization for Standardization: Information Technology - Security Techniques—Biometric Information Protection; 2011.
16. Jain AK, Nandakumar K, Nagar A. Biometric template security. EURASIP J Adv Signal Process, Special Issue on Biometrics 2008; 1–17
17. Kuzu R, Maiorana E, Campisi P. Loss functions for cnn-based biometric vein recognition. In: European Signal Processing Conference (EUSIPCO), 2020a.
18. Kuzu RS, Maiorana E, Campisi P. Vein-based biometric verification using densely-connected convolutional autoencoder. IEEE Signal Process Lett. 2020;27:1869–73. https://doi.org/10.1109/LSP.2020.3030533.
19. Lavrac N, Podpecan V, Robnik-Sikonja M. Representation Learning. Cham: Springer; 2021.
20. Mai G, Cao K, Yuen PC, et al. On the reconstruction of face images from deep face templates. IEEE Trans Pattern Anal Mach Intell. 2019;41(5):1188–202.
21. Musto R, Kuzu R, Maiorana E, et al. On the statistical independence of parametric representations in biometric cryptosystems: Evaluation and improvement. In: 11th International Conference on Pattern Recognition Applications and Methods (ICPRAM), 2022.
22. Nandakumar K, Jain AK. Biometric template protection: bridging the performance gap between theory and practice. IEEE Signal Process Mag. 2015;32(5):88–100. https://doi.org/10.1109/MSP.2015.2427849.
23. Patel VM, Ratha NK, Chellappa R. Cancelable biometrics: a review. IEEE Signal Process Mag. 2015;32(5):54–65 (**Special Issue on Biometric Security and Privacy**).
24. Piciucco E, Maiorana E, Kauba C, et al. Cancelable biometrics for finger vein recognition. In: IEEE International Workshop on Sensing, Processing and Learning for Intelligent Machines (SPLINE), 2016.
25. Piciucco E, Maiorana E, Campisi P. Palm vein recognition using a high dynamic range approach. IET Biom. 2018;7(5):439–46.
26. Pillai JK, Patel VM, Chellappa R, et al. Secure and robust iris recognition using random projections and sparse representations. IEEE Trans Pattern Anal Mach Intell. 2011;33(9):1877–93.
27. Rathgeb C, Uhl A. A survey on biometric cryptosystems and cancelable biometrics. EURASIP J Inf Secur. 2011;1:1–25.
28. Ross A, Shah J, Jain AK. From template to image: reconstructing fingerprints from minutiae points. IEEE Trans Pattern Anal Mach Intell. 2007;29(4):544–60.
29. Simoens K, Tuyls P, Preneel B. Privacy weaknesses in biometric sketches. In: 2009 30th IEEE Symposium on Security and Privacy, 2009; pp 188–203, https://doi.org/10.1109/SP.2009.24.
30. Smith AM, Mancini MC, Nie S. Bioimaging: second window for in vivo imaging. Nat Nanotechnol. 2009;4(11):710.
31. Tuyls P, Skoric B, Kevenaar T. Security with noisy data. London: Springer; 2007.
32. Uludag U, Pankanti S, Prabhakar S, et al. Biometric cryptosystems: issues and challenges. Proc IEEE. 2004;92(6):948–60.
33. Van Hamme T, Rúa EA, Preuveneers D, et al. On the security of biometrics and fuzzy commitment cryptosystems: a study on gait authentication. IEEE Trans Inf Forens Secur. 2021;16:5211–24.
34. Wang S, Hu J. Design of alignment-free cancelable fingerprint templates via curtailed circular convolution. Pattern Recogn. 2014;47(3):1321–9.
35. Yin Y, Liu L, Sun X. SDUMLA-HMT: a multimodal biometric database. In: Chinese Conference on Biometric Recognition, Springer, 2011; pp 260–268
36. Zhang D, Guo Z, Lu G, et al. An Online System of Multispectral Palmprint Verification. IEEE Trans Instrum Meas. 2009;59(2):480–90.
37. Zhou X, Kuijper A, Veldhuis R, et al. Quantifying privacy and security of biometric fuzzy commitment. In: International Joint Conference on Biometrics (IJCB), 2011; pp 1–8.