This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2018.2850320, IEEE Transactions on Information Forensics and Security

Convolutional Neural Network for Finger-Vein-based Biometric Identification

Rig Das, *Member, IEEE*, Emanuela Piciucco, *Student Member, IEEE*, Emanuele Maiorana, *Senior Member, IEEE*, and Patrizio Campisi, *Senior Member, IEEE*

***Abstract*—The use of human finger-vein traits for the purpose of automatic user recognition has gained a lot of attention in the recent years. Current state-of-the-art techniques can provide relatively good performance, yet they are strongly dependent upon the quality of the analyzed finger-vein images. In this paper, we propose a convolutional-neural-network-based finger-vein identification system and investigate the capabilities of the designed network over four publicly-available databases. The main purpose of this work is to propose a deep-learning method for finger-vein identification, able to achieve stable and highly-accurate performance when dealing with finger-vein images of different quality. The reported extensive set of experiments show that the accuracy achievable with the proposed approach can go beyond 95% correct identification rate for all the four considered publicly-available databases.**

***Index Terms*—Convolutional neural network, finger-vein, biometrics, identification.**

I. Introduction

THE design of efficient biometric identification systems, measuring unique physical or behavioral characteristics of individuals for their secure recognition, is nowadays a challenging and relevant task for both the scientific and the industrial communities. Commonly employed physical biometric traits include face, hand geometry, fingerprint, and iris among the others, whereas signature, voice, keystroke pattern, and gait are examples of behavioral modalities. As most of these modalities are prone to spoof attacks [1], [2], there is a high growth in demand for more user-friendly, yet secure, biometric modalities such as finger vein [3], hand vein [4] and palm vein [5], since they are harder to forge and difficult to acquire without the users' willingness. Vein images are usually captured using near-infrared-based optical imaging system. The illumination system is composed by infrared light that either passes through the hand, or it is reflected by it. Vein patterns are then acquired through an infrared camera and, since the haemoglobin in the blood absorbs infrared light, they appear as dark lines in the acquired image [6].

Despite the recent advances in finger-vein-based biometric recognition, finger-vein extraction approaches still remain broadly categorized into four sets such as vessel extraction [3], [6]–[10], subspace-learning-based approaches [11]–[15], statistical-based techniques [16]–[18], and local-invariant-based methods [19]–[22]. The vessel extraction technique relies on the fact that a vein pattern can be seen as a network of dark lines on a brighter background. Veins can be segmented and extracted as line-like structures [3], [6], [23], as curvature [7], [8], or minutiae [9], [10]. Vein patterns are also recognized using various thresholds, depending on neighborhood number, tracking times, and curvature values. Feature extraction techniques based on subspace learning exploit appearance-based methods, such as principal component analysis (PCA) [11], [12], two-dimensional PCA [13], [14] or linear discriminant analysis (LDA) [15], and consider the subspace coefficients as discriminative characteristics. Global or local statistical information such as the local binary histogram and invariant moments are employed by the statistical-based approaches. Local binary pattern (LBP) [16], [17] and local derivative pattern (LDP) [18] are examples of local-statistics-based methods, while the use of invariant moments is an instance of global statistics [24]. Finally, local-invariant-based methods are inspired by approaches stemming from computer vision. A typical application of these techniques is with the usage of key points for the scale invariant feature transform (SIFT) [19], [20].

Most of the current state-of-the-art models suffer from some shortcomings, mainly related to the associated feature extraction approaches. For example, some of the existing methods do not perform well on low-quality images, which can be originated by poor quality infrared light, ambient lighting conditions, light scattering in imaging finger tissues [21], fat finger, cold weather or poorly designed image capturing devices [22]. Besides that, most of the algorithms relies on parameters that cannot be set as standard values and may change while considering different databases. Moreover, for segmentation-based methods, as well as for techniques based on statistics, finger rotation and translation have a negative impact on recognition performance.

In order to overcome such limitations, in this paper we propose to perform finger-vein-based identification by exploiting deep-learning techniques. Deep learning is mainly inspired by the human brain and typically uses a multilayer perceptron (MLP) algorithm for classification. Deep-learning methods such as convolutional neural networks (CNNs) consist of a number of convolutional and sub-sampling layers producing a fully connected layer, which in turn can be used as a robust feature extractor and classifier module. The aim of our work is to achieve good and stable identification performance irrespective of the quality of the considered finger-vein images, their rotation, translation, and scaling. In order to verify the effectiveness of the designed CNN, we have tested our approach over four publicly-available finger-vein databases, characterized by different image quality levels. The achieved performance shows that the proposed method is able to guarantee stable and highly-accurate identification results,

R. Das, E. Piciucco, E. Maiorana, P. Campisi are with the Section of Applied Electronics, Department of Engineering, Roma Tre University, Via Vito Volterra 62, 00146, Rome, Italy e-mail: (rig.das@uniroma3.it, emanuela.piciucco@uniroma3.it, emanuele.maiorana@uniroma3.it, patrizio.campisi@uniroma3.it), Ph.: +39.0657337064, Fax: +39.0657337026.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2018.2850320, IEEE Transactions on Information Forensics and Security

TABLE I: Details of publicly-available finger-vein image databases.

Database	Subjects	No. of Fingers	Details of Fingers	Images per Finger	Sessions	Image Size	Total Images
HKPU [3]	156	2	Left hand index & middle finger	12	2	513×256	3132
FV-USM [25]	123	4	Left & right hand index & middle finger	12	2	640×480	5904
SDUMLA [26]	106	6	Left & right hand index, middle & ring finger	6	1	320×240	3816
UTFVP [27], [28]	60	6	Left & right hand index, middle & ring finger	4	2	672×380	1440

irrespective of the quality of the considered finger-vein images. Additionally, the proposed CNN-based identification system requires negligible manual effort for feature selection. In fact, it has been applied without variations to all the four considered databases, without using any application-dependent threshold or any manually-set parameter.

This paper is organized as follows: a detailed description of the four publicly-available finger-vein databases considered in most of the related literature is first provided in Section II. These datasets are exploited here in the experimental tests to facilitate reproducible research and future comparisons. Section III then provides a brief overview of state-of-the-art techniques specifically designed for finger-vein-based biometric recognition, identification, and verification scenarios separately, while deliberately focusing on approaches tested over the aforementioned four publicly-available databases. In addition, an overview of existing works exploiting CNNs in the field of finger-vein-based biometric recognition is also given. Section IV presents the topology of the adopted CNN, while Section V details the proposed finger-vein-based biometric identification system. Section VI then discusses about the obtained experimental results, while conclusions are eventually drawn in Section VII.

II. Finger-Vein Databases

The effectiveness of our proposed CNN-based identification system is evaluated on four publicly-available finger-vein databases, namely the Hong Kong Polytechnic University (HKPU) [3], the University Sains Malaysia (FV-USM) [25], the Shandong University (SDUMLA) [26], and the University of Twente Finger Vascular Pattern (UTFVP) [27], [28] database. The primary reason for using these specific databases is because most of the existing finger-vein-based recognition methods have been evaluated over one or more of these databases, and a fair comparison with these established methods can be therefore given. An overview of the four considered databases is given in Table I, and the following subsections provide more details about them.

1) HKPU database: The HKPU finger-vein image database [3] consists of images from 156 male and female volunteers. It has been acquired between April 2009 and March 2010 using a contact-less imaging device at the Hong Kong Polytechnic University campus. It is composed by 3132 images from the 156 subjects, all of them in BMP format with a resolution of 513×256 pixels. In this dataset about 93% of the subjects are younger than 30 years, and finger-vein images from 105 subjects have been acquired in two separate sessions with a minimum interval of one month and a maximum of over six months, with an average of 66.8 days. In each session, every subject has provided 6 image samples from index and middle finger of the left hand. Other 51 subjects have one single session of acquired data.

2) FV-USM database: The FV-USM database [25] is from University Sains Malaysia. It consists of left and right hand index and middle fingers' vein images from 123 subjects. Among them, 83 are male and 40 female, with an age range of $20 - 52$ years. Images have been acquired in two different sessions with six images per finger in every session. All images are in gray level BMP format with a resolution of 640×480 pixels.

3) SDUMLA database: The SDUMLA database [26] has been collected by Shandong University of China. It contains finger-vein images of 636 fingers from 106 subjects. Six images have been acquired 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.

4) UTFVP database: The UTFVP database [27], [28] has been collected by the University of Twente, Netherlands. It consists of 1440 PNG images with 672×380 resolution, taken from 60 subjects. Images have been acquired in two sessions from 6 fingers, i.e., left and right hand's index, middle and ring fingers, with every finger registered twice in each acquisition session. The images have a density of 126 pixels/cm and the width of the visible blood vessels is $4 - 20$ pixels [29].

III. State-of-the-Art: Finger-vein Patterns and Biometrics

A relevant number of state-of-the-art finger-vein based biometric approaches are tested on in-house datasets, thus making it difficult to have a fair comparison with newly proposed method. Moreover, a significant performance variability can be often encountered when applying state-of-the-art approaches to different databases, especially if low-quality finger-vein images are present. For example, although the maximum-curvature-based (MC) approach proposed by Miura et. al [8] reaches a remarkable equal error rate (EER) of 0.0009% when applied over an in-house finger-vein database [8], its performance goes to a low correct identification rate (CIR) of 65.40% when the same algorithm is applied to the HKPU database [3]. A similar situation happens when the repeated line tracking (RLT) feature extraction algorithm is taken into account. In fact it leads to an EER of 0.145% using the settings proposed in [6], while only a CIR of 79.52% is obtained when the same method is applied to the HKPU finger-vein images. Performing tests on different large public databases and comparing the achieved results with state-of-the-art methods evaluated on the same datasets, is therefore of

TABLE II: State-of-the-art for finger-vein-based biometric identification.

Paper	Database	Subjects	Feature Extraction Method	Classifier	Performance
Kumar et al. [3]	HKPU	105	Gabor filter with morphological processing	X-OR based similarity score	CIR=90.08%
Liu et al. [10]	HKPU	156	SVD based minutiae extraction LEBP based false removing	Fusion of Euclidean and Hamming Distance	CIR=95.71%
Van et al. [30]	SDUMLA	106	MFRAT [31] & GridPCA	Euclidean distance	CIR=95.67%
Lu et al. [32]	SDUMLA	106	Polydirectional Local Line Binary Pattern	Histogram intersection	CIR=99.21%
Ong et al. [33]	SDUMLA	106	Minutiae	Genetic algorithm & k-modified Hausdorff distance (k-MHD)	CIR = 99.7%
Qui et al. [34]	SDUMLA	106	Dual-sliding window localization + Pseudo-elliptical transformer + 2D-PCA	Euclidean distance	CIR=97.61%
	FV-USM	123			CIR=97.02%
Xie et al. [35]	SDUMLA	106	Block-based average absolute deviation (AAD) features	Ensemble component-based extreme learning machines (EC-ELM) network	CIR=97.76%
Banerjee et al. [36]	SDUMLA	106	Images, after Fuzzy contrast enhancement + CLAHE + directional dilation (DD)	Affine registration based template matching algorithm (ARTeM)	CIR=90.72%

paramount importance for properly assessing the effectiveness of a proposed method.

In this regard, an overview of state-of-the-art finger-vein based biometric recognition systems tested over the publicly-available databases are described in Section II. Specifically, since identification and verification have different performance indicators, works related to these two modalities are discussed separately in Section III-A and III-B, respectively. Additionally, since we propose the adoption of CNNs to perform finger-vein-based biometric identification, an overview of the applications of CNNs in the field of biometric recognition using finger-vein patterns is given in Section III-C. In more detail, both approaches that uses CNNs as classifiers for biometric recognition purposes as well as those exploiting CNNs for tasks such as spoofing detection, image quality assessment, and vein segmentation, are discussed. A detailed explanation of the pros and cons of our proposed method over state-of-the-art approaches is provided in Section IV-C.

A. Finger-Vein Biometric Identification

Gabor filters with morphological processing have been used in [3] for feature extraction, with XOR-based similarity scores used for finding similarity between images and achieving a CIR of 90.08%. Liu et al. in [10] have used HKPU database's session-1's data for their system, while discarding session-2's biometric traits. Singular value decomposition (SVD) has been used for minutiae extraction, and local extensive binary pattern (LEBP) has been employed for removing false pairs. An accuracy of 95.71% has been achieved fusing Euclidean and Hamming distances of the compared templates. Van et al. in [30] have used the modified finite Radon transformation (MFRAT) [31] for discriminant orientation feature extraction over the SDUMLA dataset. GridPCA [39] has also been applied to remove further redundant information. Enlarging-training-set (ETS)-based comparison techniques [31] have been employed for overcoming translations thus achieving a CIR=95.67% by calculating Euclidean distances between test and training templates. In [32], authors have proposed a polydirectional local line binary pattern (PLLBP) method for extracting vein line patterns in any orientations. The discriminative ability of LLBP (local line binary pattern) [40] histograms from different orientations has been first exploited, with histogram intersection then employed to measure the similarity between two histograms, using a score-level fusion to provide the final similarity score. A CIR of 99.21% has been achieved for the SDUMLA database. Ong et al. have proposed a reliable two-stage multi-instance finger-vein identification system based on minutiae comparison [33]. For their research work, they have used the SDUMLA dataset and combined minutiae features extracted from multiple instances of finger veins. A genetic algorithm (GA) [41] has been used to select the most reliable minutiae points from the feature point pool-set. A K-modified Hausdorff distance (k-MHD) [42] has been employed to evaluate the closet point set of two minutiae templates for comparison. An identification rate of 99.7% has been achieved for a not-specified number of employed training images. Qui et al. in [34] have used dual-sliding window localization and pseudo-elliptical transformation, with a two-dimensional principal component analysis (2D-PCA) to project the transformed image for feature extraction. Euclidean distance has been used for measuring similarity between training and testing images. A CIR of 97.61% over the SDUMLA database and an accuracy of 97.02% is obtained for the FV-USM database. In [35], the authors have tested feature-component-based extreme learning machines (FC-ELMs) over the SDUMLA database, with enrollment sets of either 3 or 5 randomly-selected images are employed for training of each individual's. Features have been extracted by a guided filter using the eight block-based average absolute deviation (AAD) directional features from high-quality finger-vein contours without performing segmentation. An ensemble component-based ELM network (EC-ELM), which averages the eight FC-ELM outputs, has been employed for final decision. The best identification results have been obtained using 5 images for training, with a corresponding average accuracy of $97.76 \pm 0.048\%$. For the same $\{5\text{-}1\}$ strategy of SDUMLA database, Banerjee et al. in [36] have obtained an average percentage of correct classification (PCC) of 90.72%, using affine-registration-based

TABLE III: State-of-the-art for finger-vein-based biometric verification.

Paper	Database	Subjects	Feature extraction methods	Classifier	Performance
Gupta et al. [37]	HKPU	105	Variational approach for vein extraction	Overlapping pixels between registered and binarized templates	EER = 4.47%
Xi et al. [38]	HKPU	105	Discriminative binary codes (DBC)	SVM	EER = 1.44%
Bakhtiar et al. [25]	FV-USM	123	Modified Gaussian Filter (MGF) enhanced & displacement corrected images	Band Limited Phase Only Correlation (BLPOC)	EER = 2.34%
Yang et al. [22]	HKPU	105	Anatomy Structure Analysis based Vein Extraction(ASAVE)	Elastic Matching	EER = 0.38%
	SDUMLA	106			EER = 1.39%
Ton et al. [28]	UTFVP	60	Maximum curvature	Correlation based comparison	EER = 0.4%
Kauba et al. [29]	UTFVP	60	Different feature level fusion	Correlation based comparison	EER = 0.19%

template matching (ARTeM) algorithms. For their proposed method, the authors have first selected the region of interest (ROI) and, then, fuzzy contrast enhancement and contrast limited adaptive histogram equalization (CLAHE) have been performed, along with an average filtering and directional dilation (DD).

Table II reports a summary of the recent state-of-the-art finger-vein-based biometric identification techniques. It can be noted that most of these techniques have been tested on either one or maximum two publicly-available databases. A comprehensive testing of a proposed method on all the four major publicly-available databases, to prove its effectiveness under different conditions of available image quality, is still missing in state-of-the-art literature.

B. Finger-Vein Biometric Verification

Although our study focuses on biometric identification, we also provide an overview of the most relevant contributions on finger-vein-based verification systems. With the same rationale adopted in Section III-A, we explicitly review only papers tested on one or more databases out of the four publicly-available ones mentioned in Section II. The details of the reported works are summarized in Table III.

Gupta el al. [37] have used a fusion strategy named variational approach to combine enhanced vein images obtained from both multi-scale matched filtering and line tracking. Similarity scores have been obtained by first registering the two vein images to be compared and then computing the number of overlapping binary pixels between them. For their proposed method, the authors have been able to achieve an EER of 4.47% for index and middle finger combination, over the HKPU database. Xi et al. [38] have proposed a discriminative binary codes (DBC) learning method, building subject relation graph to capture correlations among subjects and, based on that, generating binary templates according to the graph transform. The distance between templates from different subjects has been maximized during the graph transform in order to ensure that templates are discriminative and representative. Eventually, support vector machines (SVMs) have trained as code learners for each bit. The proposed algorithm has obtained an EER of 1.44% on the HKPU database. Bakhtiar et al. [25] have enhanced finger-vein images of the FV-USM database using modified Gaussian filter (MGF) [52] and then correcting the image displacements. Band-limited phase only correlation (BLPOC) [53] has been used for measuring the similarity between registered and test images as it is resilient to noise, occlusions and rescaling factors. An EER of 2.34% has been achieved for unimodal finger veins. In [22] authors have used anatomy-structure-analysis-based vein extraction (ASAVE) and elastic matching, achieving an EER of 0.38% and 1.39% for the HKPU and SDUMLA databases respectively. In [28] and [29] authors have used maximum-curvature-based feature extraction and different feature level fusion techniques to achieve EERs of 0.4% and 0.19%, respectively, over the UTFVP database.

C. CNNs in Finger-Vein Scenario

In recent years, applications of deep-learning-based methods, such as CNN, have been introduced in vein-based recognition scenarios, as summarized in Table IV.

Biometric identification using CNN has been studied by Radzi et al. in [43]. The employed network is based on the one proposed in [54], in which convolution and sub-sampling layers are fused into one layer, resulting in a reduced-complexity four-layer CNN. The CNN inputs are binary images obtained by thresholding the ROI of original finger-vein images. The proposed system has been tested on an in-house dataset.

Hong et al. in [44], exploited a pre-trained model of VGG-Net-16 [55] in order to perform biometric verification based on finger veins. VGG-Net-16 is composed of 13 convolutional layers, 5 pooling layers, and 3 fully-connected layers. The CNN pre-trained model has been fine-tuned with training images consisting of the differences between two finger-vein images. Experiments are performed on three different databases, namely the SDUMLA database and two other non-publicly-available datasets. Huang et al. [45] have proposed DeepVein, a finger-vein verification method based on a deep CNN (D-CNN) architecture inspired by the VGG-Net-16 model, and modified by removing some layers and reducing the number of filters in some convolutional layers. The resulting network consists of 26 layers: 10 convolutional layers, 4 pooling layers, and 2 fully-connected layers. The network is fed with two templates merged into a 2-channel image. Training and validation are carried out using a dataset collected by the authors, and three different publicly-available databases are used for testing.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2018.2850320, IEEE Transactions on Information Forensics and Security

TABLE IV: State-of-the-art for applications of CNN in the field of finger-vein-based biometric recognition.

Paper	Database	# Finger Classes	CNN Aim	Performance
Radzi et al. [43]	Own	300 (50 users)	Biometric Identification	CIR = 100%
Hong et al. [44]	Own (Good Quality)	120 (20 users)	Biometric Verification	EER = 0.804%
	Own (Middle Quality)	198 (33 users)		EER = 2.967%
	SDUMLA (Low Quality)	636 (106 users)		EER = 6.115%
Huang et al. [45]	Own (Training)	300.000	Biometric Verification	-
	FVRC2016 - DS1 [46] (Testing)	1000		EER = 0.42%
	FVRC2016 - DS2 [46] (Testing)	1000		EER = 1.41%
	FVRC2016 - DS3 [46] (Testing)	1000		EER = 2.14%
Raghavendra et al. [48]	Finger video [47]	300 (100 users)	PAD - Inkjet printed artefact	APCER = 3.48%
			PAD - Laserjet printed artefact	APCER = 0.00%
	Finger images [49]	300 (100 users)	PAD - Inkjet printed artefact	APCER = 3.20%
			PAD - Laserjet printed artefact	APCER = 0.40%
Qin et al. [50]	FV-USM	492 (123 users)	Finger-vein image quality assessment	EER = 0.80%
	HKPU	302 (156 users)		EER = 2.33%
Qin et al. [51]	FV-USM	492 (123 users)	Finger-vein segmentation and recovery	EER = 1.42%
	HKPU	302 (156 users)		EER = 2.70%

CNNs have been applied to finger-vein images in other works, although not explicitly for biometric classification purposes. Specifically, Raghavendra et al. [48] have proposed a finger-vein presentation attack detection (PAD) algorithm based on a D-CNN inspired by Alex-Net [56], yet with seven additional layers. The D-CNN model has been fine-tuned with finger-vein presentation attack samples. A majority voting rule has been exploited to classify images as either bona-fide or artefact. Two different attack databases have been taken into account for carrying out the experiments, with the proposed scheme able to guarantee high performance on both of them, improving the achievable results in comparison to other existing PAD methods. Qin et al. [50] proposed a deep neural network (DNN) consisting of three convolutional layers, three max-pooling layers, two fully-connected layers, and a softmax layer, to predict the quality of finger-vein images to be used in a biometric verification system. Specifically, the network's aim is to automatically label low- and high-quality images. Assuming that low-quality images are likely to yield more false non-matches, the authors have studied the impact on recognition performance of considering only images classified with high quality by the proposed DNN, while leveraging on state-of-the-art algorithms for feature extraction and comparison of finger-vein patterns [3]. Two publicly-available databases have been considered, with the proposed model outperforming the existing quality assessment approaches. Qin et al. have also proposed a deep learning model to extract and recover vein features. Specifically, their CNN has been exploited to segment vein pixels from the background, by predicting the probability of a pixel to belong to a vein pattern. The proposed CNN consists of two convolutional layers, two max-pooling layers, two local normalization layers, one fully-connected layer, and a softmax layer. Besides that, an approach for recovering vein patterns in the extracted finger-vein images, relying on a fully convolutional network (FCN), has been also proposed. The FCN consists of four layers: an input layer, two convolutional layers, and an output layer. The strategy adopted to compare either the features extracted through the CNN, or the ones obtained after the recovering procedure, is based on the computation of the amount of overlap between templates. Experimental results on two public finger-vein databases have shown an improvement in terms of finger-vein verification accuracy.

Table-V summarizes the details of the CNNs exploited in the aforementioned works. A detailed comparison among them and the network here proposed for finger-vein-based biometric identification is provided in Section IV-C, after the detailed description of our proposed CNN architecture.

IV. Convolutional Neural Network & Topology

In this section details of the employed CNN in the proposed finger-vein-based biometric identification system are given.

A. Convolutional Neural Network

A CNN is a multilayer perceptron (MLP) network with a special topology containing more than one hidden layer [56]. CNNs are primarily used for object recognition in image processing, handwritten character recognition and speech recognition, as they automatically extract discriminative features inside their layers from raw input information, without any specific normalization. This kind of model is advantageous for input data with an inner structure like images, and where invariant features have to be discovered. One of the main interest for using CNNs is to avoid hand-designed input features, which may not have been derived by considering the general problems. Following subsections will provide detailed description of different layers of a CNN.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2018.2850320, IEEE Transactions on Information Forensics and Security

TABLE V: State-of-the-art CNN architectures used for finger-vein-based biometric recognition.

Paper	Input size	Conv layers	Kernel size	Pooling layers	FC layers	Loss function	Reference CNN	Learning rate
Radzi et al. [43]	$55 \times 67 \times 1$	4	7×7	2	2	Mean square error	[54]	0.001
Hong et al. [44]	$224 \times 224 \times 3$	13	3×3	5	3	softmax	VGG-16 [55]	0.00001 0.0001 0.0005
Huang et al. [45]	$128 \times 128 \times 2$	10	3×3	4	2	Cross entropy error	VGG-16 [55]	0.01
Raghavendra et al. [48]	$224 \times 224 \times 3$	8	11×11 5×5 3×3	3	$2 + 3$ (extra)	softmax	Alex-Net [56]	0.00001 0.0001 0.001 0.01
Qin et al. [50]	$80 \times 240 \times 1$	4	5×5 3×3	3	2	softmax	-	0.0002
Qin et al. [51]	$15 \times 15 \times 1$	3	5×5 3×3	2	2	softmax	-	0.0002 0.01
	$39 \times 146 \times 1$	2	9×9 5×5	0	1	Mean square error	-	0.0001 0.01
Proposed	$65 \times 153 \times 1$	5	5×5	3	1	softmax	-	0.00001

1) Convolutional Layer: A set of two-dimensional convolutions is performed in the convolutional layer between the input maps x_m^l, with l and m being respectively the level and map indexes, and the filters represented through the kernels $w_{n,m}^l$, being n the filter index. The n-th output map y_n^l of layer l is computed as:

$$y_n^l = \sum_m^{M^{l-1}} w_{n,m}^l * x_m^l + b_n^l \tag{1}$$

where M^{l-1} is the number of input maps, $*$ denotes convolution, and b_n^l is the bias of the n-th output map in the l-th level. The values characterizing the kernels and the biases are set according to [57].

2) ReLU: The rectified linear unit (ReLU) is a nonlinear layer (or activation layer) which is usually applied immediately after the conv layer described by (1). The purpose of this layer is to introduce nonlinearity in the system. In the past, nonlinear functions like $\tanh$ and $sigmoid$ have been used, but researchers have found out that ReLU layers work far better, allowing networks to be trained faster without scarifying the accuracy. It also helps to alleviate the vanishing gradient problem, an issue making the lower layers of the network to be trained very slowly due to the exponential decrease of the gradient through the layers. The ReLU layer applies the function $f(y) = \max(0, y)$, changing all negative activations to 0. This layer increases the nonlinear properties of the model and the overall network, without affecting the receptive fields of the conv layer.

3) Pooling: Pooling layer is also referred to as a down-sampling layer. Maxpooling is the most popular layer option. It takes a filter, normally of size 2×2, and a stride of the same length. It is then applied to the input volume and outputs the maximum value in every subregion that the filter convolves around. Average pooling and L2-norm pooling are other options for pooling layers. The intuitive reasoning behind this layer is that once we know that a specific feature is in the original input volume. *i.e.* high activation value, its exact location is not as important as its relative location to the other features. This layer drastically reduces the spatial dimension of the input volume. As a consequence, the amount of parameters or weights is significantly reduced, thus lowering the computation cost and controlling over-fitting.

4) Fully Connected Layer: This layer takes the output of the preceding conv or pool layer or ReLU and generates an N dimensional vector, where N is the number of classes that the program has to choose from. A softmax classifier is typically employed to predict the probability of the input image belonging to a specific label. Let x_m be the m-th input map of the output layer, then the linear combination O_n is defined as:

$$O_n = \sum_{m=1}^{M} (w_{n,m} * x_m + b_n), \tag{2}$$

where $M = 1024$ in our case as shown in $L_5M_3R_1$ of Fig. 1. A fully connected layer looks for some high level features which are strongly correlated to a particular class, by computing probabilities for the available classes. The probability distribution of the input data over C different classes is predicted by the softmax function:

$$p_u = \frac{\exp(O_u)}{\sum_{n=1}^{C} \exp(O_n)}. \tag{3}$$

B. Network Topology

The CNN we have proposed is shown in Figure 1. The network has 5 convolutional layers, 3 max-pooling, 1 ReLU, and a softmaxloss layer. The detailed topology is described as follows:

- L_0: the input layer with an input data size of $[65 \times 153]$, which is the size of input images of finger veins. A

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2018.2850320, IEEE Transactions on Information Forensics and Security

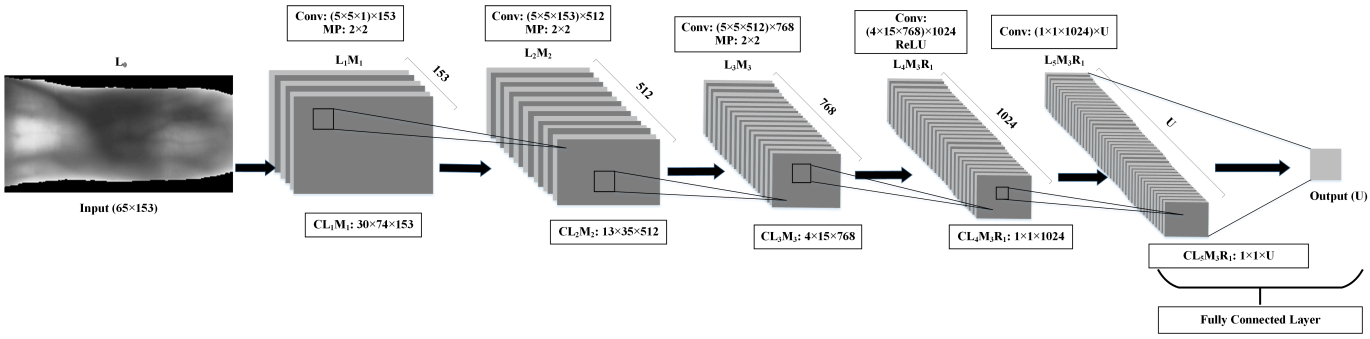


Fig. 1: Employed CNN architecture.

detailed description of how input data are processed to get to this size is provided in Section V-A and V-B;

- L_1M_1: first hidden layer, composed by 153 convolutional filter of size $[5 \times 5 \times 1]$ and a max-pooling (MP) layer of size $[2 \times 2]$. This layer transforms the input data into $CL_1M_1 = [30 \times 74 \times 153]$ low-level features after convolving and down-sampling;
- L_2M_2: second hidden layer, composed by 512 conv filter of size $[5 \times 5 \times 153]$ and a max-pooling layer of size $[2 \times 2]$. This layer transforms the first hidden layer's output into $CL_2M_2 = [13 \times 35 \times 512]$ features;
- L_3M_3: third hidden layer, composed by 768 conv filter of size $[5 \times 5 \times 512]$ and a max-pooling layer of size $[2 \times 2]$. This layer transforms the second hidden layer's output into $CL_3M_3 = [4 \times 15 \times 768]$ features;
- $L_4M_3R_1$: the fourth hidden layer is composed by 1024 convolutional filter of size $[4 \times 15 \times 768]$ and a ReLU layer. This layer changes the previous layer's activation map into a $CL_4M_3R_1 = [1 \times 1 \times 1024]$ feature map;
- $L_5M_3R_1$: the final hidden layer, or fully connected layer, is produced by convolving the previous layer's activation map with U convolutional filters of size $[1 \times 1 \times 1024]$. The U neurons of this layer represent the U classes/subjects. This layer generates a fully connected network with the input data and produces the probabilities of its belonging to one of the U classes. Softmaxloss function is used as loss function for back-propagation.

Table-VI shows the configuration details of our proposed CNN with kernel size, number of stride, and padding.

C. Comparison with state-of-the-art CNNs used for finger-vein images

As shown in Table-V, most state-of-the-art CNN architectures applied to finger-vein images require square inputs, as VGG-16 or Alex-net. Native biometric traits are instead typically acquired according to a rectangular shape, therefore implying the need for severe resizing operations, which can distort original features and loose vital information. Our proposed CNN instead takes input images of size $65 \times 153 \times 1$, therefore reducing the chances of distortions.

Moreover, very small kernels are used in approaches such as [44] and [45], therefore increasing the number of required conv layers and pooling layers, and resulting in longer time for training and testing. As our input image size is fixed into 65×153 along with the kernel of size 5×5, we are instead required of a total 5 conv layers and 3 max-pooling layers to dissolve the input image into a fully connected layer. This significantly reduces the training and testing time, and also increases the identification accuracy. Moreover, our learning rate is fixed into a very low value of 0.00001, that entails a very deep training and results into a very low testing error.

More importantly, the only work that has investigated so far the possible use of CNN for finger-vein-based biometric identification is [43], where the authors have used input images of size 55×67 with kernel size of 7×7. This means that, with every convolution, a very large block of information is processed, with a potentiality of leading into a very fast training saturation and a consequent high variability for the testing results. Also, there remains a high possibility of loosing some minute feature details of the finger-vein while training and testing. The results reported in [43] are also very hard to replicate, since they have been obtained from a small in-house dataset containing 50 subjects.

In summary, it can be stated that the advantages of the proposed CNN over the current state-of-the-art CNN architectures are in the use of more realistic input image size, with optimized kernel size that reducing the training and testing time and a very low learning rate for performing a very deep training, thus significantly lowering the testing error.

V. Employed Finger-Vein based Biometric System

Once finger-vein data are preprocessed, the corresponding templates are generated as described in Section V-B. The performed training and identification phases are described in Section V-C and V-D, respectively.

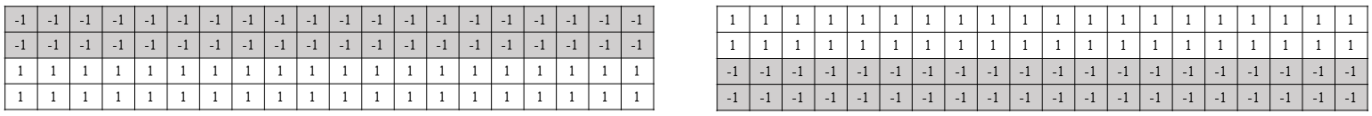
A. Preprocessing

The original images, gathered from four publicly-available databases, are pre-processed for ROI extraction and image enhancement. As a first step, the images from all the considered databases, having different sizes, are subsampled to 336×190 pixels in order to guarantee uniformity. Beside that, for the databases where the images show a ratio between number of rows and columns different from the target one, marginal background parts are removed by selecting a central area of the image. Eventually, the ROI, i.e. the part of the image which contains the interested finger, is then extracted and a binary mask in which the white pixels correspond to the finger region is obtained. Specifically, the ROI extraction

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2018.2850320, IEEE Transactions on Information Forensics and Security

TABLE VI: Proposed CNN configuration.

Layer Type	Number of Filter	Size of Feature Map	Size of Kernel	Number of Stride	Number of Padding
Image input layer	-	$65 \times 153 \times 1$	-	-	-
CL_1 (Convolutional layer-1)	153	$65 \times 153 \times 1$	5×5	1×1	0×0
M_1 (Max-Pooling Layer-1)	1	$61 \times 149 \times 153$	2×2	2×2	0×0
CL_2 (Convolutional layer-2)	512	$30 \times 74 \times 153$	5×5	1×1	0×0
M_2 (Max-Pooling Layer-2)	1	$26 \times 70 \times 512$	2×2	2×2	0×0
CL_3 (Convolutional layer-3)	768	$13 \times 35 \times 512$	5×5	1×1	0×0
M_3 (Max-Pooling Layer-3)	1	$9 \times 31 \times 768$	2×2	2×2	0×0
CL_4 (Convolutional layer-4)	1024	$4 \times 15 \times 768$	4×15	1×1	0×0
R_1 (ReLu Layer-1)	-	$1 \times 1 \times 1024$	-	-	-
CL_5 (Convolutional layer-5)	U (number of classes)	$1 \times 1 \times 1024$	1×1	1×1	0×0
Softmax Layer	-	$U \times 1$	-	-	-



(a) Masks for detection of upper region of finger

(b) Masks for detection of lower region of finger

Fig. 2: Masks for ROI selection of finger-vein images.

is based on the method proposed by Lee et al. [16], where two different masks, as shown in Figure 2, are used to extract the upper and lower finger's edges respectively. For the HKPU database the aforementioned masks are provided, whereas the aforementioned procedure is applied to the other databases. Starting from the extracted edges and masks, a normalization step is performed in order to compensate rotation and vertical translation during the acquisition step. In our work, we use the approach proposed in [58], which attempts to fit a straight line between the edges detected in the previous step and estimate the parameters of rotation and vertical translation which are later used to perform an affine transformation. If required, the normalized images may be then enhanced through contrast limited adaptive histogram equalization (CLAHE) [59], which is an adaptive histogram equalization (AHE) method whose aim is to improve the contrast of the image by limiting the contrast amplification in the different considered parts of the image. The preprocessed images are then transposed and resized into 65×153 pixels.

Figure 3 shows a comparison of the final images with and without performing the CLAHE enhancement for the four different databases.

B. Template Generation

Bigger images usually lead to a larger CNN with more hidden layers. Hence, in order to have a feasible size network, the images are first resized into 65×153. In our approach, the training and testing templates of our network are either generated by selecting the images from a single session, as proposed by existing state-of-the-art methods, or by selecting a combination of images from all available sessions. The reason behind this latter strategy is that, as can be seen from Fig. 4, the same data can be acquired in different sessions under diverse illumination conditions. Hence, the network may require images captured in different settings for its proper training, in order to not to affect the identification accuracy. To find the best combination of templates for training, we have investigated 1, 2, 3, and 4 images' combinations from all available sessions for training. The obtained results allowed us to find the best possible combination of templates to be used for person identification.

C. CNN Training

The generated templates are passed through the designed CNN and a set of very low-level features are extracted in the first hidden layer. The network gradually builds up over these low-level features in the subsequent convolutional layers, in order to create a set of high-level features for the fully connected layer.

For our experiments we have considered each finger of every person as a separate class. For the HKPU dataset, since 105 subjects have contributed with their index and middle fingers to two sessions, there is a total of 210 classes available for training. The remaining 51 subjects have contributed only to session 1, so they have not been considered for training in our tests and they have been instead only used as imposters while testing. Similarly, for FV-USM database we have considered 492 classes (123 subjects with 4 fingers each), 636 classes for the SDUMLA database (106 subjects with 6 fingers each), and 360 classes for the UTFVP database (60 subjects with 6 fingers each).

For CNN designing and training we have used the MatConvNet-1.0-beta24 tool [60]. For training purposes 90% finger-vein images are considered, with the remaining 10% used for validation. The learning rate of the CNN is set at 0.00001 with a batch size of 3 samples for HKPU and FV-USM, 4 for SDUMLA and 2 for UTFVP, so that the loss can

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2018.2850320, IEEE Transactions on Information Forensics and Security

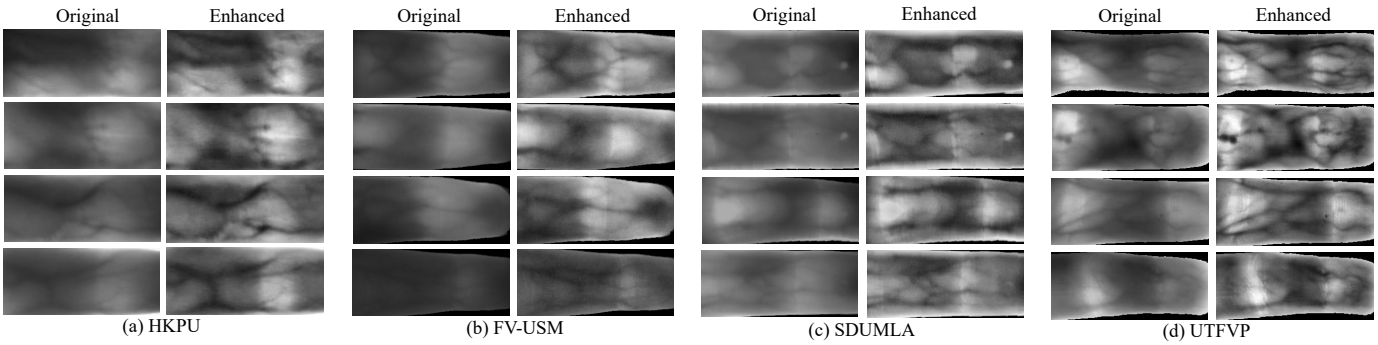


Fig. 3: Original and CLAHE [59] enhanced finger-vein image from four publicly-available databases.

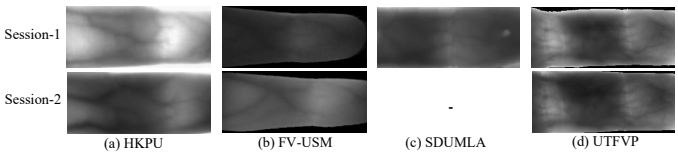


Fig. 4: Different luminosity images from different sessions of four publicly-available databases.

be minimized with higher precision through the execution of every epoch or iteration. As for the number of epochs, higher numbers usually allow the network to be well-trained, so that the weights of different layers are updated with precision. For our experiment, we have considered 2500 epochs for all the experiments. The main purpose of using such a low learning rate and high number of epochs is that it is typically preferable to let a network learn very slowly and converge into the smallest details of every class.

D. Identification

In the identification stage, the testing templates are generated as described in Section V-B from the remaining images. For each testing sample, the trained CNN returns a probability value for all the available classes/fingers. The maximum probability value identifies the most similar finger to the testing sample. As we have considered each and every finger of an individual as a different class, we are able to identify the particular finger with which it is matched and the corresponding subject to whom it belongs to.

It is worth specifying that, similar to what has been proposed in [3], for our experimental setup we have introduced a threshold for matching probability of a test image, below which we consider the test image as "not-identified". This is for the purpose of genuine imposter testing where no sample images are trained for that particular subject, as they are not associated with any of the enrolled identities. For a given testing sample, if the matching probability value returned by the proposed network is less than 50% for its comparisons with any trained class, then that test image is classified as "not-identified" or "not-present" in the database. For example, we have tested this scenario with the finger-vein images of the 51 subjects captured during a single session in the HKPU dataset. Such images have not been ever employed for training purposes and have been instead used only as testing probes. Each time, when a test sample's result reaches into a maximum matching probability value of lower than 50% for all the trained classes, then it is possible to declare them as "not-identified".

VI. Results & Discussion

In order to evaluate the proposed network, we have first compared its performance with several state-of-the-art identification techniques in Section VI-A, by using the training and testing strategies adopted in referenced papers for our proposed network as well.

We have then designed an optimal training strategy for our proposed network in Section VI-B. Most of the state-of-the-art techniques have used either a single image or images from a single session for their network's training, which may not be ideal for our CNN-based approach. It is in fact well-known that the availability of a single sample of every class, here individual fingers, does not allow a CNN to get trained properly.

Eventually, we have also evaluated the utility of exploiting image enhancement preprocessing techniques together with the proposed network in Section VI-C, to understand if further performance improvement could be achieved.

All the experiments have been performed in MATLAB® (R2017a) with a system configuration of 64Gb RAM; Titan X^{TM}(Pascal) graphics card; i7, 3.40GHz processor and Windows® 10 operating system.

A. Performance Comparisons

The identification accuracies achieved by the state-of-the-art finger-vein-based biometric systems that are discussed in Table II are reported in Table VII, together with the obtained performance with our proposed CNN-based approach, when using the same training and testing strategies. The results obtainable while exploiting two of the most-commonly employed methods for finger-vein recognition, i.e., MC [8] and RLT [6], under all the considered settings, are additionally reported for further comparisons.

As it can be seen from the reported accuracies, our CNN-based identification system cannot be properly trained under the experimental setup employed in [3], where only session-1 images from HKPU dataset are used for training, and session-2 images for testing. A similar situation is encountered when comparing the proposed system against the one in [34], where tests over the FV-USM dataset have been performed by considering only the first image of every finger from session-1 for training and the 6 images per finger of session-2 for

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2018.2850320, IEEE Transactions on Information Forensics and Security

TABLE VII: Identification accuracy comparison for the four considered publicly-available databases.

Database	Training	Testing	State-of-the-art comparison methods									Proposed CNN
			Kumar et al. [3]	Qui et al. [34]	Van et al. [39]	Jia et al. [31]	Van et al. [30]	Xie et al. [35]	Banerjee et al. [36]	MC [8]	RLT [6]	
HKPU	6 images from session 1	6 images from session 2	90.08%	-	-	-	-	-	-	85.24%	78.28%	71.11%
FV-USM	first image from session 1	6 images from session 2	-	97.02%	-	-	-	-	-	90.34%	78.28%	72.97%
SDUMLA	4 images	remaining 2 images	-	-	91.83%	92.50%	95.67%	-	-	86.01%	87.11%	97.48%
	5 images	remaining image	-	-	-	-	-	97.76%	90.72%	97.95%	96.06%	98.90%
UTFVP	1 image from each session	remaining image from each session	-	-	-	-	-	-	-	92.22%	93.05%	95.56%

TABLE VIII: Identification accuracy for different training strategies over original images.

Database	Training (images from each available sessions)			
	1	2	3	4
HKPU	82.19%	92.02%	95.32%	**96.55%**
FV-USM	91.75%	94.82%	97.53%	**98.58%**
SDUMLA	75.25%	77.99%	80.27%	**97.48%**
UTFVP	**95.56%**	-	-	-

testing. Again, the reason behind such a low performance depends upon the number of training samples, along with the different quality of finger-vein images that exists in two distinct sessions, as shown by the examples in Fig. 4.

A different behavior of our network has been observed when considering the training/testing settings employed in [30] and the SDUMLA database, which contains images taken from a single session. In this scenario, the method here proposed is able to achieve identification performance better than those obtained in [35] and [36]. It is therefore reasonable to observe that the proposed CNN-based identification system can work properly when images of similar quality are used for both training and testing purposes, regardless the absolute quality level of the considered images. This assumption is confirmed by the results in Table VII, referred to the comparison of the proposed approach against MC and RLT, while taking one finger-vein image from each session of the UTFVP dataset, and using the remaining ones for testing purposes. The proposed CNN-based identification system easily outperforms both MC and RLT.

It is worth remarking that the aforementioned results have been obtained with the proposed CNN-based identification system without performing any kind of enhancement on finger-vein images. Conversely, all the methods we are comparing with use some image enhancement technique and feature selection processes. Therefore, the use of original images without any preprocessing and automatic feature extraction are among the advantages of our proposed network.

B. Training Strategy Selection

As mentioned in the previous subsection, the use of a single image or images from a single session for training purposes may not be enough to produce the desired accuracy. Therefore, we have analyzed the full potentiality of our proposed CNN architecture by investigating the variations in identification accuracies that can be achieved while adopting different numbers of images for training. Hence, wherever possible, 1, 2, 3, and 4 original images from all the available sessions of each of the finger-vein patterns from the four databases are considered for the training of our network.

Table VIII summarizes the obtained results and clearly shows that, if the number of training samples from each finger is increased, then the achieved accuracy also increases significantly. Comparing these results with those reported in Table VII, a notable improvement in terms of achievable performance can be seen for both HKPU and FV-USM databases. From Table VIII it can also be noted that there is not much difference in accuracy, for HKPU and FV-USM database, when 3 or 4 images are used for training. Hence, for further experiments we employ 3 images from each session of these two databases for training. When considering the SDUMLA database, 4 images from each session are needed for training, since low quality images are present in this database, as shown in Figure 3. As for the UTFVP database, we can choose at most one image from each session for training, since only 2 images are available for each finger's acquisition sessions. The obtained results show that identification accuracy greater than 95% can be achieved for all the four considered publicly-available databases.

C. Proposed Network's Accuracy for Best Training Strategies

In this subsection we report the results of experiments performed in order to further evaluate the effectiveness of our proposed approach. Specifically, we first have assessed the improvements that can be obtained when the proposed network is trained according to the settings described in Section VI-B. A performance comparison with standard MC and RLT methods is also provided here. Moreover, we have also investigated whether the proposed method needs any image enhancement technique to further improve the attainable identification performance.

Table IX shows rank-1 identification accuracy for four publicly-available databases, when exploiting the original finger-vein images with MC, RLT, and proposed CNN-based approach. Identification performance obtained using the proposed system, when the network is fed with CLAHE [59]-enhanced finger-vein images, is also reported here.

According to the obtained results, the proposed CNN-based identification system systematically outperforms MC and RLT approaches. It can also be seen that the contrast-enhanced images achieve performance better than the original ones only when considering the UTFVP dataset. It is evident from Fig. 3 that vein patters in images from UTFVP database are significantly more prominent and clearly distinguishable

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2018.2850320, IEEE Transactions on Information Forensics and Security

TABLE IX: CNN-based identification accuracy over the considered publicly-available databases.

Database	Training	Testing	Methods			
			MC	RLT	CNN with original images	CNN with CLAHE [59] enhanced images
HKPU	3 images from each session	remaining 3 images from each session	83.33%	83.81%	**95.32%**	94.37%
FV-USM	3 images from each sessions	remaining 3 images from each session	92.60%	94.44%	**97.53%**	97.05%
SDUMLA	4 images	remaining 2 images	86.01%	87.11%	**97.48%**	95.13%
UTFVP	1 image from each session	remaining 1 image from each session	92.22%	93.05%	95.56%	**98.33%**

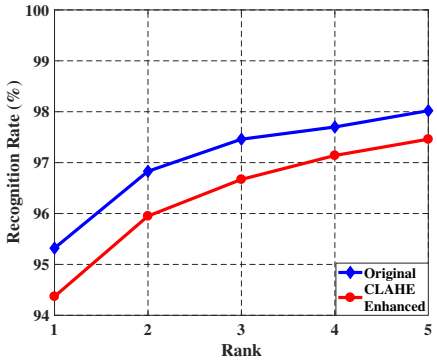


Fig. 5: HKPU database: CMCs for original and contrast enhanced finger-vein images while 3 images of each finger from both sessions are used for training and testing.

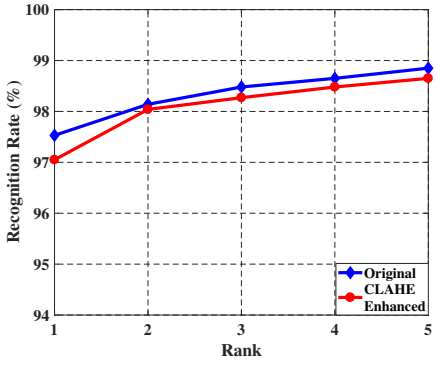


Fig. 6: FV-USM database: CMCs for original and contrast enhanced finger-vein images while 3 images of each finger from both sessions are used for training and testing.

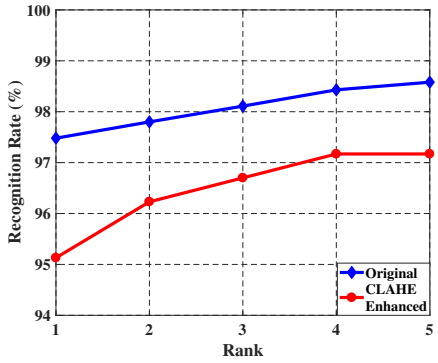


Fig. 7: SDUMLA database: CMCs for original and contrast enhanced finger-vein images while 4 images of each finger are used for training and 2 images for testing.

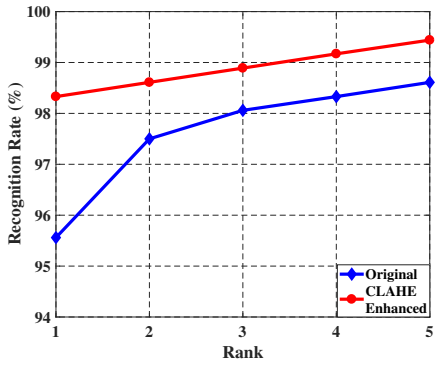


Fig. 8: UTFVP database: CMCs for original and contrast enhanced finger-vein images while 1 image of each finger from both sessions are used for training and testing.

in their enhanced versions rather than in their original ones. Nevertheless, the proposed CNN-based approach is typically able to guarantee a very-high identification rate without using any image enhancement technique.

Figures 5-8 provide further details regarding the proposed method, by reporting the cumulative match curves (CMC) obtained for different ranks, while using both the original and the contrast-enhanced versions of the finger-vein images. Rank-5 identification accuracy greater than 98% can be achieved for all the four considered publicly-available databases, while a maximum of 99.4% Rank-5 identification accuracy can be achieved for the UTFVP database.

VII. Conclusions

In this paper, we have proposed a CNN-based finger-vein identification system which can perform an effective identification irrespective of the environmental conditions. We have presented an exhaustive set of experimental tests performed over the four commonly-used and publicly-available databases. The obtained results show that it is possible to achieve a rank-1 identification accuracy greater than 95% for all the four databases, using our proposed CNN architecture. The present work is one of the first comprehensive study analyzing a finger-vein-based biometric identification system with more than two publicly-available databases, to assess the effectiveness of the proposed network under different image quality conditions, with minimum human intervention. It can also be seen that the identification accuracy of the proposed network significantly increases with the employed number of training images. Moreover, if the finger-vein images are not acquired with the same illumination intensity and ambient lighting

conditions for different sessions of acquisition, then the use of multiple session's data for training can be considered as an effective strategy for improving the achievable identification accuracy.

ACKNOWLEDGMENT

We gratefully acknowledge the support of NVIDIA® Corporation for providing us a Titan X™ Pascal GPU.

This research was partially supported by the Grant PRIN 2015 COSMOS: "COntactlesS Multibiometric mObile System in the wild", funded by the Italian Ministry of Education, University and Research.

REFERENCES

[1] S. Marcel, M. S. Nixon, and S. Z. Li, *Handbook of Biometric Anti-Spoofing: Trusted Biometrics Under Spoofing Attacks*. Springer Publishing Company, Incorporated, 2014.

[2] D. Menotti, G. Chiachia, A. Pinto, W. R. Schwartz, H. Pedrini, A. X. Falco, and A. Rocha, "Deep representations for iris, face, and fingerprint spoofing detection," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 864–879, April 2015.

[3] A. Kumar and Y. Zhou, "Human identification using finger images," *IEEE Trans. on Image Proc.*, vol. 21, no. 4, pp. 2228–2244, April 2012.

[4] A. Kumar and K. V. Prathyusha, "Personal authentication using hand vein triangulation and knuckle shape," *IEEE Transactions on Image Processing*, vol. 18, no. 9, pp. 2127–2136, Sept 2009.

[5] Y. Zhou and A. Kumar, "Human identification using palm-vein images," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 4, pp. 1259–1274, Dec 2011.

[6] N. Miura, A. Nagasaka, and T. Miyatake, "Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification," *Machine Vision and Applications*, vol. 15, no. 4, pp. 194–203, 2004.

[7] W. Song, T. Kim, H. C. Kim, J. H. Choi, H.-J. Kong, and S.-R. Lee, "A finger-vein verification system using mean curvature," *Pattern Recognition Letters*, vol. 32, no. 11, pp. 1541–1547, 2011.

[8] N. Miura, A. Nagasaka, and T. Miyatake, "Extraction of finger-vein patterns using maximum curvature points in image profiles," *IEICE - Trans. Inf. Syst.*, vol. E90-D, no. 8, pp. 1185–1194, aug 2007.

[9] C.-B. Yu, H.-F. Qin, Y.-Z. Cui, and X.-Q. Hu, "Finger-vein image recognition combining modified hausdorff distance with minutiae feature matching," *Interdisciplinary Sciences: Computational Life Sciences*, vol. 1, no. 4, pp. 280–289, Dec 2009.

[10] F. Liu, G. Yang, Y. Yin, and S. Wang, "Singular value decomposition based minutiae matching method for finger vein recognition," *Neurocomputing*, vol. 145, pp. 75 – 89, 2014.

[11] Z. Liu, Y. Yin, H. Wang, S. Song, and Q. Li, "Finger vein recognition with manifold learning," *Journal of Network and Computer Applications*, vol. 33, no. 3, pp. 275–282, 2010.

[12] J.-D. Wu and C.-T. Liu, "Finger-vein pattern identification using principal component analysis and the neural network technique," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5423–5427, 2011.

[13] G. Yang, X. Xi, and Y. Yin, "Finger vein recognition based on (2d) 2 pca and metric learning," *BioMed Research Int.*, vol. 2012, 2012.

[14] F.-X. Guan, K. Wang, J. Liu, and H. MA, "Bi-direction weighted (2d) 2pca with eigenvalue normalization one for finger vein recognition," *Pattern Recogn. and Artificial Int.*, vol. 24, no. 3, pp. 417–424, 2011.

[15] J.-D. Wu and C.-T. Liu, "Finger-vein pattern identification using svm and neural network technique," *Expert Systems with Applications*, vol. 38, no. 11, pp. 14 284–14 289, 2011.

[16] E. C. Lee, H. C. Lee, and K. R. Park, "Finger vein recognition using minutia-based alignment and local binary pattern-based feature extraction," *International Journal of Imaging Systems and Technology*, vol. 19, no. 3, pp. 179–186, 2009.

[17] B. A. Rosdi, C. W. Shing, and S. A. Suandi, "Finger vein recognition using local line binary pattern," *Sensors*, vol. 11, no. 12, pp. 11 357–11 371, 2011.

[18] E. C. Lee, H. Jung, and D. Kim, "New finger biometric method using near infrared imaging," *Sensors*, vol. 11, no. 3, pp. 2319–2333, 2011.

[19] J. Peng, N. Wang, A. A. A. El-Latif, Q. Li, and X. Niu, "Finger-vein verification using gabor filter and sift feature matching," in *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on*. IEEE, 2012, pp. 45–48.

[20] H. Qin, L. Qin, L. Xue, X. He, C. Yu, and X. Liang, "Finger-vein verification based on multi-features fusion," *Sensors*, vol. 13, no. 11, pp. 15 048–15 067, 2013.

[21] E. C. Lee and K. R. Park, "Image restoration of skin scattering and optical blurring for finger vein recognition," *Optics and Lasers in Engineering*, vol. 49, no. 7, pp. 816 – 828, 2011.

[22] L. Yang, G. Yang, Y. Yin, and X. Xi, "Finger vein recognition with anatomy structure analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, no. 99, pp. 1–1, March 2017.

[23] J. Yang, Y. Shi, and G. Jia, "Finger-vein image matching based on adaptive curve transformation," *Patt. Rec.*, vol. 66, pp. 34–43, 2017.

[24] X. Li, S. Guo, F. Gao, and Y. Li, "Vein pattern recognitions by moment invariants," in *Bioinformatics and Biomedical Engineering, 2007. ICBBE 2007. The 1st International Conference on*. IEEE, 2007, pp. 612–615.

[25] M. S. M. Asaari, S. A. Suandi, and B. A. Rosdi, "Fusion of band limited phase only correlation and width centroid contour distance for finger based biometrics," *Expert Systems with Applications*, vol. 41, no. 7, pp. 3367 – 3382, 2014.

[26] Y. Yin, L. Liu, and X. Sun, *SDUMLA-HMT: A Multimodal Biometric Database*, Z. Sun, J. Lai, X. Chen, and T. Tan, Eds. Springer Berlin Heidelberg, 2011.

[27] P. Tome, M. Vanoni, and S. Marcel, "On the vulnerability of finger vein recognition to spoofing," in *IEEE International Conference of the Biometrics Special Interest Group (BIOSIG)*, Sep. 2014.

[28] B. T. Ton and R. N. J. Veldhuis, "A high quality finger vascular pattern dataset collected using a custom designed capturing device," in *Proceedings of the 2013 International Conference on Biometrics (ICB), Madrid, Spain*, 2013, pp. 1–5.

[29] C. Kauba, E. Piciucco, E. Maiorana, P. Campisi, and A. Uhl, "Advanced variants of feature level fusion for finger vein recognition," in *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*, Sept 2016, pp. 1–7.

[30] H. T. Van, T. T. Thai, and T. H. Le, "Robust finger vein identification base on discriminant orientation feature," in *2015 7th Int. Conf. on Knowledge and Systems Engineering (KSE)*, Oct 2015, pp. 348–353.

[31] W. Jia, D.-S. Huang, and D. Zhang, "Palmprint verification based on robust line orientation code," *Pattern Recognition*, vol. 41, no. 5, pp. 1504 – 1513, 2008.

[32] Y. Lu, S. J. Xie, S. Yoon, and D. S. Park, "Finger vein identification using polydirectional local line binary pattern," in *2013 International Conference on ICT Convergence (ICTC)*, Oct 2013, pp. 61–65.

[33] T. S. Ong, J. H. Teng, K. S. Muthu, and A. B. J. Teoh, "Multi-instance finger vein recognition using minutiae matching," in *2013 6th International Congress on Image and Signal Processing (CISP)*, vol. 03, Dec 2013, pp. 1730–1735.

[34] S. Qiu, Y. Liu, Y. Zhou, J. Huang, and Y. Nie, "Finger-vein recognition based on dual-sliding window localization and pseudo-elliptical transformer," *Expert Systems with Applications*, vol. 64, pp. 618 – 632, 2016.

[35] S. J. Xie, S. Yoon, J. Yang, Y. Lu, D. S. Park, and B. Zhou, "Feature component-based extreme learning machines for finger vein recognition," *Cognitive Computation*, vol. 6, no. 3, pp. 446–461, Sep 2014.

[36] A. Banerjee, S. Basu, S. Basu, and M. Nasipuri, "Artem: a new system for human authentication using finger vein images," *Multimedia Tools and Applications*, Mar 2017.

[37] P. Gupta and P. Gupta, "An accurate finger vein based verification system," *Digital Signal Processing*, vol. 38, pp. 43 – 52, 2015.

[38] X. Xi, L. Yang, and Y. Yin, "Learning discriminative binary codes for finger vein recognition," *Pattern Recognition*, vol. 66, pp. 26–33, 2017.

[39] H. T. Van, P. Q. Tat, and T. H. Le, "Palmprint verification using gridpca for gabor features," in *Proceedings of the Second Symposium on Information and Communication Technology*, ser. SoICT '11. New York, NY, USA: ACM, 2011, pp. 217–225.

[40] A. Petpon and S. Srisuk, "Face recognition with local line binary pattern," in *2009 Fifth International Conference on Image and Graphics*, Sept 2009, pp. 533–539.

[41] K. F. Man, K. S. Tang, and S. Kwong, "Genetic algorithms: concepts and applications [in engineering design]," *IEEE Transactions on Industrial Electronics*, vol. 43, no. 5, pp. 519–534, Oct 1996.

[42] M.-P. Dubuisson and A. K. Jain, "A modified hausdorff distance for object matching," in *Proceedings of the 12th IAPR international Conference on Pattern Recognition*, vol. 1, 1994, pp. 566–568.

[43] S. Radzi, M. Khalil-Hani, and R. Bakhteri, "Finger-vein biometric identification using convolutional neural network," *Turkish Journal of Electrical Eng. & Comp. Sciences*, vol. 24, no. 3, pp. 1863–1878, 2016.
[44] H. Hong, M. Lee, and K. Park, "Convolutional neural network-based finger-vein recognition using nir image sensors," *Sensors*, vol. 17, no. 6, pp. 1–21, 2017.
[45] H. Huang, S. Liu, H. Zheng, L. Ni, Y. Zhang, and W. Li, "Deepvein: Novel finger vein verification methods based on deep convolutional neural networks," in *IEEE International Conference on Identity, Security and Behavior Analysis*, 2017.
[46] Y. Ye, L. Ni, H. Zheng, S. Liu, Y. Zhu, D. Zhang, W. Xiang, and W. Li, "Fvrc2016: The 2nd finger vein recognition competition," in *Biometrics (ICB), 2016 International Conference on*. IEEE, 2016, pp. 1–6.
[47] R. Raghavendra, M. Avinash, S. Marcel, and C. Busch, "Finger vein liveness detection using motion magnification," in *Biometrics Theory, Applications and Systems (BTAS), 2015 IEEE 7th International Conference on*. IEEE, 2015, pp. 1–7.
[48] R. Raghavendra, S. Venkatesh, K. B. Raja, and C. Busch, "Transferable deep convolutional neural network features for fingervein presentation attack detection," in *5th International Workshop on Biometrics and Forensics (IWBF)*. IEEE, 2017, pp. 1–5.
[49] R. Raghavendra and C. Busch, "Presentation attack detection algorithms for finger vein biometrics: A comprehensive study," in *11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. IEEE, 2015, pp. 628–632.
[50] H. Qin and M. A. El Yacoubi, "Deep representation for finger-vein image quality assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, 2017.
[51] H. Qin and M. A. El-Yacoubi, "Deep representation-based feature extraction and recovering for finger-vein verification," *IEEE Trans. on Inf. Forensics and Security*, vol. 12, no. 8, pp. 1816–1829, Aug 2017.
[52] E. C. Lee, H. Jung, and D. Kim, "New finger biometric method using near infrared imaging," *Sensors*, vol. 11, no. 3, pp. 2319–2333, 2011.
[53] K. Takita, T. Aoki, Y. Sasaki, T. Higuchi, and K. Kobayashi, "High-accuracy subpixel image registration based on phase-only correlation," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E86-A, no. 8, pp. 1925–1934, 2003.
[54] P. Y. Simard, D. Steinkraus, J. C. Platt *et al.*, "Best practices for convolutional neural networks applied to visual document analysis." in *ICDAR*, vol. 3, 2003, pp. 958–962.
[55] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
[56] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Inf. Proc. Systems 25*. Curran Associates, Inc., 2012, pp. 1097–1105.
[57] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, *Efficient BackProp*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 9–48.
[58] B. Huang, Y. Dai, R. Li, D. Tang, and W. Li, "Finger-vein authentication based on wide line detector and pattern normalization," in *2010 20th Int. Conf. on Pattern Recognition*, Aug 2010, pp. 1269–1272.
[59] K. Zuiderveld, *Contrast Limited Adaptive Histograph Equalization*, P. S. Heckbert, Ed. Academic Press Professional, Inc., 1994.
[60] A. Vedaldi and K. Lenc, "Matconvnet – convolutional neural networks for matlab," in *Proceeding of the ACM Int. Conf. on Multimedia*, 2015.

Rig Das (M'17) received his Ph.D. degree in Applied Electronics Engineering with European Doctorate Label from Roma Tre University, Rome, Italy, in 2018. His primary research goals are directed towards the application of deep learning methods in biometrics. As a Doctoral student with Roma Tre University, he has contributed towards the physiological biometric traits such as EEG signal, finger vein and face. His other research interests are image and video processing, steganography and steganalysis, signal processing, pattern recognition and machine learning. He was a visiting researcher with the Bar-Ilan University, Ramat Gan, Israel, in 2016. He is currently working as a research assistant with the Section of Applied Electronics, Department of Engineering of Roma Tre University, Rome, Italy. Rig has received both his Bachelor and Master of Technology degree in Computer Science and Engineering, in 2007 and 2012 respectively from India.

Emanuela Piciucco received the bachelors degree in Electronic Engineering (cum laude) in 2013 and the masters degree in Information and Communication Technology Engineering (cum laude) in 2016 at the Roma Tre University, Rome, Italy, where she is currently a PhD student in Applied Electronics. She was a Visiting Researcher at University of Salzburg, Salzburg, Austria, in 2015, in the framework of the European project ICT COST Action IC1206, and at Telefnica I+D, Barcelona, Spain, in 2017 and 2018, in the framework of the European project ENCASE. Her current research areas are biometric recognition, mainly focusing on vein pattern and EEG biometric identifiers, and physiological signal processing.

Emanuele Maiorana (IEEE SM) received the Ph.D. degree in biomedical, electromagnetism, and telecommunication engineering with European Doctorate Label from Roma Tre University, Rome, Italy, in 2009. He is currently a Research Engineer with the Section of Applied Electronics, Department of Engineering, of Roma Tre University, Rome, Italy. His research interests are in the area of digital signal and image processing with applications to multimedia communications and security of telecommunication systems. Specifically, he worked on biometric recognition and protection of biometric templates, high dynamic range images imaging and watermarking, synthesis of video textures, and stereo image analysis and enhancement. He is the recipient of the Lockheed Martin Best Paper Award for the Poster Track at the IEEE Biometric Symposium 2007, and the Honeywell Student Best Paper Award at the IEEE Biometrics: Theory, Applications and Systems conference 2008.

Patrizio Campisi (IEEE SM) received the Ph.D. degree in electrical engineering from Roma Tre University, Rome, Italy, where he is currently a Full Professor with the Section of Applied Electronics, Department of Engineering. His current research interests are in the area of secure multimedia communications and biometrics. He is the Chair of the IEEE Technical Committee on Information Forensics and Security (2017-2018) and was the IEEE SPS Director Student Services (2015-2017), and a member of the IEEE Certified Biometric Program Learning System Committee. He is the General Chair of the 26th European Signal Processing Conference EUSIPCO 2018, Rome, Italy. He has been the General Chair of the 7th IEEE Workshop on Information Forensics and Security (WIFS) 2015, Rome, Italy, and of the 12th ACM Workshop on Multimedia and Security, Rome, Italy in 2010. He is the Editor of the book Security and Privacy in Biometrics (Springer, 2013). He is a Co-Editor of the books Blind Image Deconvolution: Theory and Applications (CRC press, 2007), and High Dynamic Range Video, Concepts, Technologies and Applications (Academic Press, 2016).He has been an Associate Editor and a Senior Associate Editor of the IEEE Signal Processing Letters and an Associate Editor of the IEEE Transactions on Information Forensics and Security. He is currently Editor-in-Chief of the IEEE Transactions on Information Forensics and Security.