Dorsal Hand Vein-Biometric Recognition Using Convolution Neural Network

Conference Paper · August 2020

DOI: 10.1007/978-981-15-5113-0_92

3 authors:

Rajendra Kumar
Sharda University
98 PUBLICATIONS **344** CITATIONS
SEE PROFILE

R C Singh
Sharda University, Greater Noida (Delhi-NCR), India
99 PUBLICATIONS **624** CITATIONS
SEE PROFILE

Shri Kant
Sharda University
41 PUBLICATIONS **202** CITATIONS
SEE PROFILE

Dorsal Hand Vein-Biometric Recognition Using Convolution Neural Network

Rajendra Kumar, Ram Chandra Singh, and Shri Kant

Abstract In this paper, a convolution neural network (CNN)-based vein recognition approach is used for dorsal hand vein patterns. Apart from using pre-trained version, VGG Net-16 model is fine-tuned on four datasets of dorsal hand vein images (good quality, medium quality, and low quality) and augmented images (between the two images from genuine matching or false matching). All four datasets consist of dorsal hand vein images of left and right hands. The comparison of results proposed model is done with other CNN models like VGG Face and VGG-19 (with and without fine-tuned) along with a recent work based on transfer learning. The accuracy of recognition of proposed work using fine-tuned VGG Net-16 model obtained is 99.60% for good quality images; for medium quality images, it is 98.46%, and 97.99%; for low-quality images, it is found to be.

Keywords Biometric recognition · Dorsal hand vein · CNN · FAR · FRR · GAR

1 Introduction

Comparing to other biometric traits like a fingerprint, face, iris, etc., the vein patterns are underneath the skin, so these are virtually impossible to replicate using existing methods. Also, the vein patterns are more secure in terms of a suspect to forgery, damage, or change with time; however, no such study is available to ensure whether the vein patterns of kids are changed with age from infants to 15 years of age because major physical changes take place during this period. The old and traditional biometric recognition systems based on behavior and physiological features like finger

R. Kumar (✉) · S. Kant
School of Engineering and Technology, Sharda University, Greater Noida, India
e-mail: rajendra04@gmail.com

S. Kant
e-mail: shri.kant@sharda.ac.in

R. C. Singh
School of Basic Sciences and Research, Sharda University, Greater Noida, India
e-mail: rcsingh_physics@yahoo.com

D. Gupta et al. (eds.), *International Conference on Innovative Computing and Communications*, Advances in Intelligent Systems and Computing 1165,
https://doi.org/10.1007/978-981-15-5113-0_92

[1–3], face [4–6], iris [7, 8] have several drawbacks. The traditional non-training-based methods have some drawbacks like recognition performance is compromised due to misalignment and shading in vein images due to changes in the environment. However, in good quality vein images, the recognition accuracy is improved using Gabor filter [8, 9], edge-preserving and high pass filter [10].

The dorsal vein patterns have received wide attention in recent years in the biometric domain due to properties being unique, contactless, stable, liveness detection, almost impossible to forge. Different approaches have been applied for improving performance of vein recognition system, like local binary pattern (LBP) [11], gray-level grouping [9], gradient-based learning technique [12], geometry structure [13] with LBP, circular difference, statistical directional pattern [14] (CDSDP), minutiae extraction [15, 16], CLAHE [17], SIFT [18–20], CNN followed by multi-class SVM, etc.

Recently, vein-based recognition systems [21–23] have been developed for palm [24], finger [8, 16–32], hand dorsal [20–22, 33–38], iris [39], etc. X. Zhu et al. [36] have used a hierarchical combination of texture and shape approach for hand vein pattern classification. Also, fast processing speed is achieved while entire data of region of interest (ROI) is used without vein-line detection in local binary patterns (LBP) [26]-based vein recognition systems. And, we achieve high recognition accuracy using the characteristics of blood vessels instead of whole patterns in ROI while applying LLBP [11–13] with a vein-line tracking approach. Vein pattern recognition has been applied to different biological objects of the human body using preprocessing steps like image enhancement, ROI, and post-processing steps like feature extraction, filtering, and matching. Sontakke et al. discussed the past and existing methods/practices used in dorsal hand vein recognition [34]. CNN has significant advantages over the conventional approach, as it has extended features, reduces image dimensions and classification in one network structure simultaneously. CNN is robust enough to use minimal image preprocessing and the least misalignment of the image acquired.

The traditional vein recognition systems are difficult to apply in a wide range of applications because of the complicated preprocessing of an input image and not providing the feature vectors. The CNN-based vein recognition system does not need complex image processing to eliminate the noise and feature enhancement before image classification. The CNN-based recognitions do not face the problem of illumination change, image rotation, and scale transformation.

Many environmental parameters affect the quality of image capturing like light scattering, optical blurring, motion blurring, etc. for improving the recognition rate. Different filters and image enhancement techniques like Gabor filter [8, 9, 27–29, 31, 32], edge-preserving filter [10], elliptic high pass filter, contrast enhancement [40], histogram equalization [10] have been used for better feature extraction. In recent years, the researchers used multimodal biometric systems [35, 41], biometric fusion [31, 32] techniques to improve the performance of the biometric system.

The training-based vein recognition system like CNN, support vector machine (SVM) [16–27] with suitable optimizers have high accuracy and performance as compared to traditional (non-training-based) systems [12, 14]. This work is focused

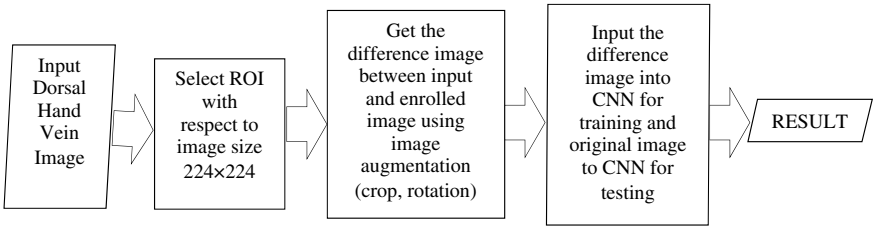

Fig. 1 The proposed model

on dorsal vein recognition using CNN and is categorized as (i) reduction of complexity of convolution neural network, (ii) use of CNN's fully connected layer (FCL) for results, (iii) handling of misalignment and shading, and (iv) comparison of accuracy of proposed work with some existing works.

The work in this paper is composed as: Section 2 discusses the structure and major components of the proposed CNN-based dorsal hand vein recognition system and Section 3 deals with data acquisition and training. Section 4 shows results and discussion and Sect. 5 outlines the conclusion of this work.

2 The Proposed Method

2.1 The Proposed System

Figure 1 shows the steps involved in the proposed dorsal hand vein recognition system. The captured dorsal vein images are resized in 224 × 224 concerning ROI. The first image does not undergo any filtering or quality enhancement. Then a difference image (augmented) is considered between the enrolled image and the input image. The difference image is applied to pre-trained CNN to enhance the training, and the original input dorsal vein images are recognized at CNN output as authentic matching or imposter matching. The dorsal vein images of self-created datasets are captured by a near-infrared (NIR) camera of 850 nm wavelength.

2.2 The NIR Camera

Figure 2 shows the NIR camera (VF620) with six infrared LEDs, which is used for capturing dorsal vein images at a fixed wavelength of 850 nm. Two datasets (first and third) used in this work are constructed with this device only.

Fig. 2 The NIR camera and user interface for vein image acquisition

2.3 CNN Based System for Dorsal Vein Recognition

CNN a class of deep neural networks is used for computer vision for the last three decades [42]. CNN uses mainly convolution layers, activation function, pooling layers, max-pooling, and fully connected layers. The convolution layers preserve the relationship among the pixel values by learning input image features in the grid. Convolution is a mathematical operation that works on input digital image/matrix and filter/kernel used for edge detection, blur, and shaping of image. Following layers and functions constitute a CNN architecture:

Convolution Layers—The convolution layers read the input image and produce a feature map using the activation function.

Filter—A small matrix of size generally 2×2 or 3×3 which is slide along the width and length of the input image to use the dot product as the activation map.

Padding—Padding represents the process of adding 0s to the input matrix symmetrically to fit the filter on an image.

Stride—Stride represents the number of steps (generally 1) the filter is slide in each step of convolution.

Activation Function—It is the function defined as node lying in between or at the end of the neural network, responsible for deciding if the neuron would fire or not. Rectified Linear Unit (ReLU) is a typical example of an activation function widely used.

Pooling Layer—Layer lying between the convolution layers is responsible for reducing the number of parameters and computational complexity, controlling overfitting problem. Max-pooling producing the maximum value from a pool on sliding of filters over the input is defined as $\frac{D-F}{S} + 1$, where D is the dimension of the pooling layer, F is the dimension of filter, and S is stride.

Fully Connected Layer—These layers are seen as the last phase of CNN as the neurons have full connectivity with all previous layer activities.
Softmax Function—The softmax function is used to classify the object's probability. The function returns 1 if the object is recognized otherwise 0. The softmax function is used finally to decide whether the compared image is genuine or imposter. The softmax function is defined as $\sigma(p)_i = \sum_{n=1}^{n=R} \frac{e^{pi}}{e^{pn}}$, where p is an array of output neurons, $\sigma(p)_i$ is the probability of neurons concerning ith class, and e^{pi} is the value of the ith element.

In this work, VGG Net-16 [43] is trained using four steps: (i) the fully connected nodes are removed just before the prediction; (ii) fully connected nodes are replaced with freshly initialized nodes; (iii) earlier convolution layers are freeze to preserve the robust features of model; and (iv) fully connected heads are trained. The fine-tuned model developed is trained with difference images (augmented images). In line with the fine-tuning of the VGG Net-16, the test performance is compared with different CNN structures over various training methods and various input types (dorsal vein images in different conditions like open and closed palm, hand kept at different temperatures, and one/two weeks' time interval). For accuracy of recognition of dorsal vein, VGG Net-16 fine-tuning with difference images (captured in different environments) as input is used.

The Rectified Linear Unit (ReLU) is the fastest nonlinear activation function [43] among currently being used. VGG Net-16 is configured on default parameters to get an input image of size 224 × 224 with 64 filters. As shown in Fig. 3, a total of 13 convolution layers, 13 ReLU layers, 5 pooling layers, and 3 fully connected layers (FCLs) in VGG Net-16 have been used to obtain a feature map of 7 × 7 × 512. And, 3 fully connected layers pass the output and the output nodes received at fully connected layers FCL_1, FCL_2, and FCL_3 are 4096, 4096 and 2, respectively.

A total of 64 filters of size 3 × 3 are used in the first convolution layer, i.e., in the first convolution layer, the size of the feature map is 224 × 224 × 64, where 224 is the length and width of the feature map. The length/width of the feature map is calculated as:

$$\text{Length} = \frac{\text{input length} - \text{filter length} + 2 \times \text{number of padding}}{\text{number of stride}} + 1,$$

where padding is the process of adding 0s to the input matrix symmetrically, and stride represents a number of steps the filter is slide in each step of convolution. The same formula is applied for calculating the width of the feature map. Hence, the input length is 224, filter length is 3, the number of padding is 1, and the number of strides is 1 in the image input layer and Conv_1_1. The input length obtained is $(224 - 3 + 2 \times 1)/1 + 1 = 224$. The CNN architecture of the proposed model is shown in Fig. 4.

The group-wise details of the layers are as follows:

Group-1: Conv_1_1, ReLU_1_1, Conv_1_2, ReLU_1_2, PL_1
Group-2: Conv_2_1, ReLU_2_1, Conv_2_2, ReLU_2_2, PL_2

Layer (type)	Output Shape	Param #
conv2d_27 (Conv2D)	(None, 224, 224, 64)	1792
conv2d_28 (Conv2D)	(None, 224, 224, 64)	36928
max_pooling2d_11	(MaxPooling (None, 112, 112, 64)	0
conv2d_29 (Conv2D)	(None, 112, 112, 128)	73856
conv2d_30 (Conv2D)	(None, 112, 112, 128)	147584
max_pooling2d_12	(MaxPooling (None, 56, 56, 128)	0
conv2d_31 (Conv2D)	(None, 56, 56, 256)	295168
conv2d_32 (Conv2D)	(None, 56, 56, 256)	590080
conv2d_33 (Conv2D)	(None, 56, 56, 256)	590080
max_pooling2d_13	(MaxPooling (None, 28, 28, 256)	0
conv2d_34 (Conv2D)	(None, 28, 28, 512)	1180160
conv2d_35 (Conv2D)	(None, 28, 28, 512)	2359808
conv2d_36 (Conv2D)	(None, 28, 28, 512)	2359808
max_pooling2d_14	(MaxPooling (None, 14, 14, 512)	0
conv2d_37 (Conv2D)	(None, 14, 14, 512)	2359808
conv2d_38 (Conv2D)	(None, 14, 14, 512)	2359808
conv2d_39 (Conv2D)	(None, 14, 14, 512)	2359808
max_pooling2d_15	(MaxPooling (None, 7, 7, 512)	0
flatten_4 (Flatten)	(None, 25088)	0
dense_8 (Dense)	(None, 256)	6422784
dense_9 (Dense)	(None, 60)	15420

Total params: 21,152,892
Trainable params: 21,152,892
Non-trainable params: 0

Fig. 3 Details of CNN configuration used in this work

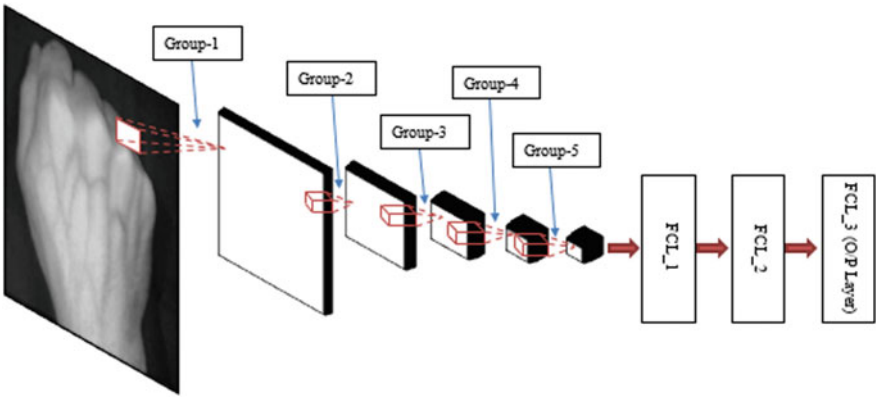

Fig. 4 The CNN architecture for proposed work

Group-3: Conv_3_1, ReLU_3_1, Conv_3_2, ReLU_3_2, Conv_3_3, ReLU_3_3, PL_3
Group-4: Conv_4_1, ReLU_4_1, Conv_4_2, ReLU_4_2, Conv_4_3, ReLU_4_3, PL_4
Group-5: Conv_5_1, ReLU_5_1, Conv_5_2, ReLU_5_2, Conv_5_3, ReLU_5_3, PL_5

A feature map of size $7 \times 7 \times 512$ is obtained after applying input image of size $512 \times 512 \times 3$ to 13 convolution layers, 13 ReLU layers, and 5 pooling layers. Additionally, it is passed through 03 fully connected layers. The output nodes obtained are 4096, 4096 and 2 after passing through first, second, and third fully connected layers. Generally, CNN undergoes overfitting problem which is a major reason for low recognition results over testing data, although the accuracy of training data is high. As a solution to this problem, we used data augmentation and dropout methods. In this study, the dropout probability [44] is set to 0.5 to disconnect two consecutive fully connected layers. The dropout layer is used twice, the first time between FCL_1 and FCL_2 and a second time between FCL_2 and ReLU_7 as shown in Fig. 3. For optimization of the model, Adam optimizer is used in pre-trained and fine-tuned CNN.

3 Data Acquisition and Training

3.1 The Datasets

Three types of dorsal hand vein datasets are used for experimental purposes. The first dataset used is self-constructed that consists of 200 persons including 100 male adults, 50 female adults, and 50 kids. A total of 20 images (10 of left-hand dorsal

and 10 of right-hand dorsal, as shown in Fig. 5) are captured using the NIR camera (VF620). Total 4000 images are captured in ten different conditions: (i) open hand; (ii) fist (iii) stretched; (iv) tilted hand; (v) rotated hand; (vi) low light; (vii) medium light; (viii) natural light; (ix) at temperature 05–10 °C; and (x) temperature 40–50 °C. Almost all environmental aspects are considered to construct this dataset like misalignment, translation, different shading, etc.

The second dataset used in this study is from SRD Laboratory [45] having 8000 dorsal hand vein images of 400 persons including 30 males and 10 females age between 18 and 22 years (as shown in Fig. 6). This dataset contains 10 images of left-hand dorsal and right-hand dorsal each. The input images of size 640 × 480 are resized in 224 × 224.

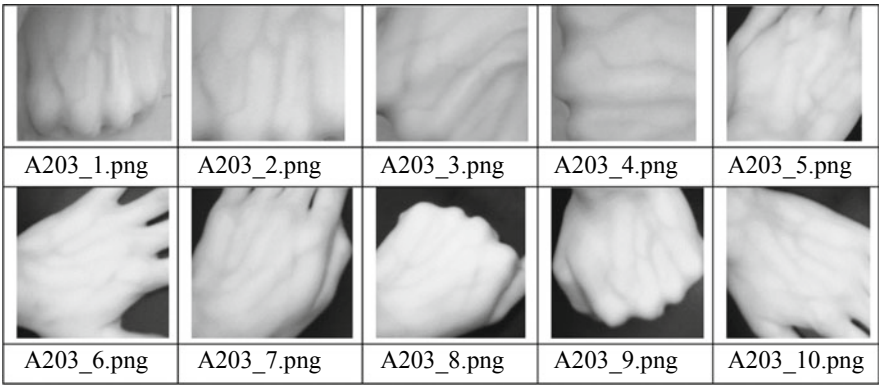

Fig. 5 Sample dataset of good quality images

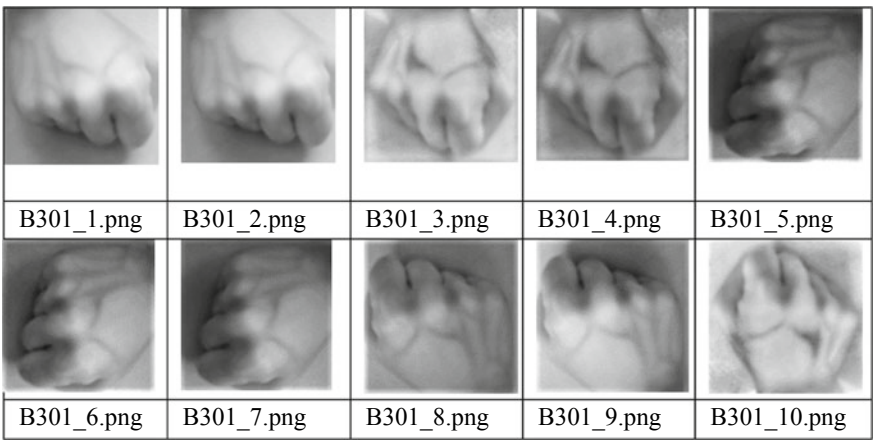

Fig. 6 Sample dataset of medium quality images

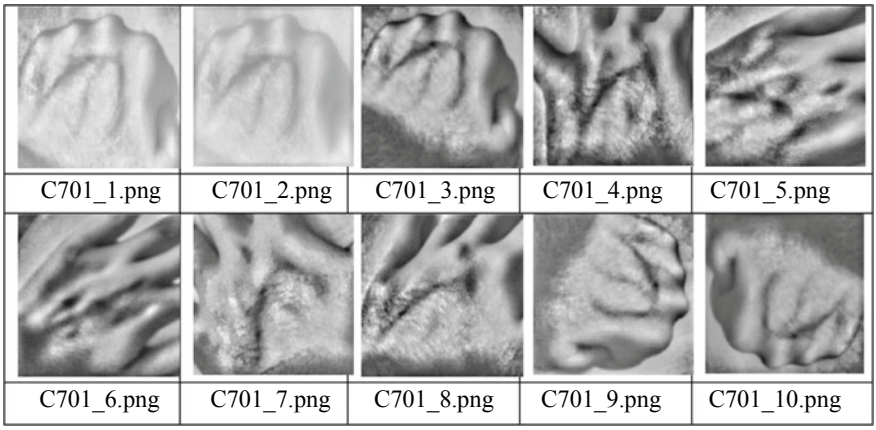

Fig. 7 Sample date set of low-quality images

The third dataset is also self-constructed in which images are captured in low light conditions with NIR camera VF620 with only three LEDs. The dataset has 2000 images of 100 persons (10 images of left-hand dorsal and 10 images of right-hand dorsal). Figure 7 shows a set of ten images of left-hand dorsal of a person.

The dataset with good quality (self-constructed) has 4000 hand vein images, the medium quality (SRD Lab) has 8000 hand vein images and BOSPPHORUS [29] dataset has 1575 hand vein images, and a low-quality dataset (self-constructed) has 2000 hand vein images. The datasets from SRD Laboratory and BOSPPHORUS are of medium quality. The classes from all four datasets are divided into two groups for training and testing purposes. For training and testing purpose, a group of 50% images of each class has been used. The average accuracy is calculated by applying twofold cross-validation by interchanging these groups of images for training and testing.

Figure 8 shows the augmented (difference) images from an original image. The augmented images are created from all four datasets, i.e., good quality, medium quality, and low-quality dorsal hand vein images. The augmented images are used only for training purposes to train CNN with a sufficient number of images. The

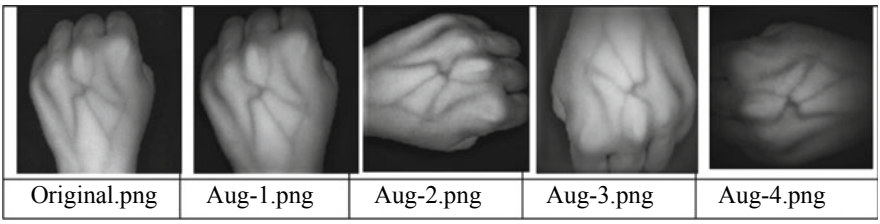

Fig. 8 Original and created augmented (difference) images

Table 1 Dataset acquisition details

Criteria			Good quality (*a*)	Low quality (*c*)
Original image		Number of images	4000	2000
		Number of people	200	100
		Number of hands	2	2
		Number of classes (number of images per class)	400 (10)	200 (10)
Augmented images for training	CNN applied over the original image as input (case-i)	Number of images	242,000 (200 classes × 10 images × 121 times)	121,000 (100 classes × 10 images × 121 times)
	CNN applied over difference image as input (case-ii)	Number of images	25,800[a] ((10 images × 13 times −1) × 200 classes)	12,900[a] ((10 images × 13 times −1) × 100 classes)

[a]Number of images in the dataset for authentic matching

testing is done with original images only. Table 1 shows the details of augmented images created for CNN training.

The performance of CNN-based dorsal hand vein recognition is measured for two cases: (i) with original images as input to CNN for testing, and (ii) with difference images obtained from two dorsal hand vein images of true matching (matching between input and enrolled dorsal hand vein images of same class) and the false matching (matching between input and enrolled dorsal hand vein images of different classes). Table 1 presents acquisition details of first and third datasets. The second and forth datasets of medium quality are from SRD Laboratory and BOSPPHORUS, respectively.

In case (i), the classes used for training purposes are different from those on which testing is performed, and the accuracy is calculated using distance matching. The testing is performed with 2000 images of 200 classes and training is also performed on rest 2000 images from the dataset. To have sufficient data for training, the image augmentation is done by translation, cropping 1–5 pixels in left/right/up/down directions depending upon flexibility in the original image. A total of 121 augmented images are created for each original image. In case (ii), a total of 12 augmented images are received from one original image as the coordinates of the original image are translated, cropped 1–4 pixels in left/right/up/down directions depending upon flexibility in the original image. The same criteria are applied for augmentation of medium and low-quality images.

3.2 Training the CNN Model

The proposed convolutional network training/testing is conducted using Intel® Core™ i5-7200 CPU @ 2.5 GHz with 8 GB RAM. The training/testing algorithms are implemented in Python with OpenCV, Keras, and Tensorflow. Table 2 shows the methods comparing two images using CNN models. Here, Conv_N represents a filter of $N \times N$. The comparisons are considered at FCL_7 with the original vein image and whether the image belongs to the same person is verified in the output layer FCL_8. In all cases, a pre-tuned or a fine-tuned model is applied. The models used are VGG Face [46], VGG Net-16, and VGG Net-19 [43] in different methods *A* to *C* (original Image) and *D* to *F* (difference image). Method *A* involves the comparison of two vein images by using features in FCL_7 with the original image as input in the VGG Face model. There are 2622 classes in the VGG Face model in FCL_8 as output. Therefore, the model is trained by approximately 2.6 million face images from 2622 persons. Methods *B* and *C* are involved in comparing two images by using features in FCL_7 with the original image as input to VGG Net-16 and VGG Net-19 models having 16 and 19 layers, respectively.

The performance of VGG Face, VGG Net-16, and VGG Net-19 models using methods *A*, *B*, and *C,* respectively, is measured with testing data without using any additional training on vein images. Fine-tuning is applied for methods *A*-1, *B*-1, and *C*-1 to pre-trained models VGG Face, VGG Net-16, and VGG Net-19, respectively. The method *D* decodes whether the image belongs to the same person or not by verifying the output of the FCL_8 layer. Presently, filters of small size (5×5 and 3×3) in *D*, for method *E* the filter size designed is 3×3, and the max-pooling count is changed from 3 to 5. The same feature of output 2048×1 is applied to FCL_6 and FCL_7. A verification structure with method *D* is created with two classes, i.e., authenticate and imposter by using difference image as input to be trained. Here, we propose a method in which the CNN model VGG Net-16 is fine-tuned with original and difference dorsal hand vein images. VGG Net-16 is fine-tuned with dorsal hand vein images from four different datasets with 1 to 10 epochs. Epoch is a measure of the number of times the training vector is used to update the weights. Table 3 represents the average loss and training accuracy in twofold cross-validation for proposed method training *F*. VGG Net-16 has converged loss close to 0% and training accuracy close to 100%. A 100% training accuracy is achieved after completion of very first epoch.

4 Results and Discussion

4.1 Matching of Samples with Trained Model

Figure 9 shows dorsal hand vein patterns as the original image, enhanced image, and Gabor filtered images from (a) good, (b) medium, and (c) low-quality images.

Table 2 Comparison of various CNN models

Input image	Original image input to CNN			Difference image input to CNN		
Network configuration	VGG Face	VGG Net-16	VGG Net-19	Trained AlexNet-1	Trained AlexNet-2	Fine-tuned VGG Net-16 proposed method
Method	*A*/*A*-1[a]	*B*/*B*-1[a]	*C*/*C*-1[a]	*D*	*E*	*F*
No. of layers	16	16	19	8	8	16
No. of filters (filter size)[b]	F1 F1	F1 F1	F1 F1	F2	F1	F1
Type of pooling	Maximum	Maximum	Maximum	Maximum	Maximum	Maximum
No. of filters (filter size)[b]	F3 F3	F3 F3	F3 F3	F4	F3	F3
Type of pooling	Maximum	Maximum	Maximum	Maximum	Maximum	Maximum
No. of filters (filter size)[b]	F5 F5 F5	F5 F5 F5	F5 F5 F5 F5	F5	F5	F5 F5 F5
Type of pooling	Maximum	Maximum	Maximum	Maximum	Maximum	Maximum
No. of filters (filter size)[b]	F6 F6 F6	F6 F6 F6	F6 F6 F6 F6	F5	F5	F6 F6 F6
Type of pooling	Maximum	Maximum	Maximum	Maximum	Maximum	Maximum
No. of filters (filter size)[b]	F6 F6 F6	F6 F6 F6	F6 F6 F6 F6	F3	F3	F6 F6 F6
Type of pooling FCL_6 (1st FCL) FCL_7 (2nd FCL) FCL_8 (3rd FCL)	MAX 4096 4096 2622/No. of class	MAX 4096 4096 2622/No. of class	MAX 4096 4096 2700/No. of class	MAX 4096 1024 2	MAX 2048 2048 2	MAX 4096 4096 2

[a]Method *A*, *B*, and *C*: without fine-tuned, and method *A*-1, *B*-1 and *C*-1: fine-tuned
[b]F1: 64 (Conv_3), F2: 96 (Conv_11), F3: 128 (Conv_3), F4: 128 (Conv_5), F5: 256 (Conv_3), F6: 512 (Conv_3)

Table 3 Average loss and training accuracy in twofold cross-validation

Epoch		0	1	2	3	4	5	6	7	8	9	10
Loss	Good quality dataset	0.55	0	0	0	0	0	0	0	0	0	0
	Medium quality dataset	0.65	0	0	0	0	0	0	0	0	0	0
	Low-quality dataset	0.60	0	0	0	0	0	0	0	0	0	0
Accuracy	Good quality dataset	0	100	100	100	100	100	100	100	100	100	100
	Medium quality dataset	0	100	100	100	100	100	100	100	100	100	100
	Low-quality dataset	0	100	100	100	100	100	100	100	100	100	100

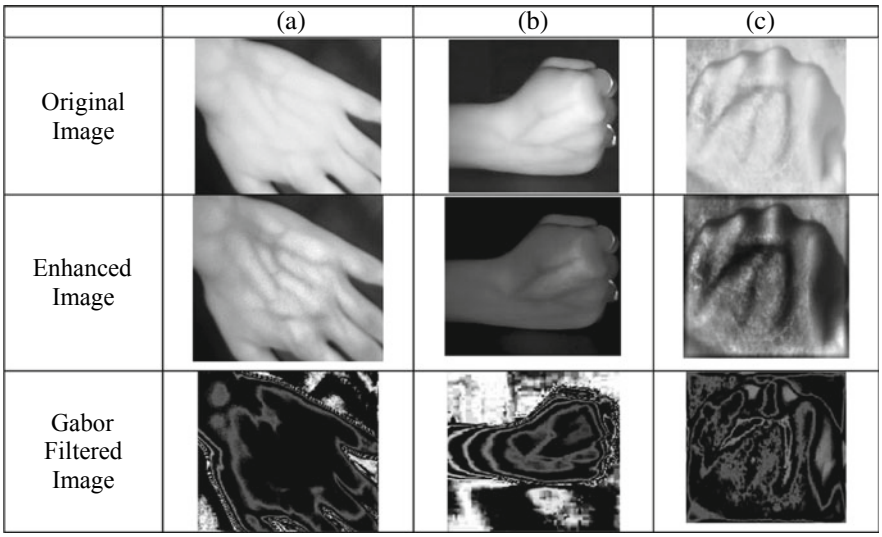

Fig. 9 Original, enhanced and Gabor filtered images

Table 4 shows the accuracy of the recognition of original images and Gabor filtered images in terms of equal error rate (EER) from all four types of datasets used on VGG Net-16 model with fine-tuning and without fine-tuning. The equal error rate is the measure of equality of FAR (false accept rate, i.e., rate of accepting of imposter matching) and FRR (false reject rate, i.e., rate of reject of authentic matching). It is observed that better EER is obtained with Gabor filtered images in the case when there is no fine-tuning and with fine-tuning, the results of original images are better than the Gabor filtered images. The possible reason for having better results of fine-tuning over Gabor filtering could be better extraction of features from good quality input images. However, it has not given better results with low-quality input images as shown in Table 4. The Gabor filter is applied on the first three datasets only.

Table 4 Comparison of original and Gabor filtered image over VGG Net-16

Name of method	Input image	EER (%)		
		With good quality images dataset (a)	With medium quality images datasets (b)	With low-quality images dataset (c)
B (VGG Net-16 without fine-tuning)	Original image	1.481	4.928	7.278
	Gabor filtered image	1.065	4.112	7.895
B-1 (VGG Net-16 with fine-tuning)	Original image	1.029	3.021	7.006
	Gabor filtered image	1.201	3.392	7.511

4.2 Comparison of the Proposed Model with Some Previous Models

The work presented in this study is compared with the work of N. A. Al-Johania et al. [29], which is based on CNN for dorsal hand vein recognition using feature learning and transfer learning approach. Al-Johania et al. used two datasets; in the first dataset is Dr. Badawi hand vein dataset [35] containing hand vein images of 50 persons (10 images per person, 5 for left hand and 5 for right hand). This dataset contains images of persons between 16 and 65 years of age including males and females. The second dataset is BOSPPHOURS that contains 1575 hand vein images from 100 persons. A total of 219 images out of 1575 are captured after a period of 2–5 months in different positions. The same dataset is also used in this study as mentioned earlier. As a specific situation, the users are asked to carry a bag of 3 kg for 1 min and the images were captured and then the hand was cooled and again the images were captured. Unlike our approach as 50–50 split ratio, Al-Johania et al. used 80–20, 70–30, and 60–40 split ratios for training and testing. The best accuracy of recognition among all three split ratios has been considered for comparisons.

While comparing the accuracy of our work with Al-Johania et al. on VGG Net-16 model at FCL_7 and FCL_8, our method shows the better result as can be seen from Table 5. There could be a possibility that our proposed system is fine-tuned with three qualities of image datasets and augmented dataset or may be due to small size datasets used by Norah et al. Dr. Badawi dataset has vein images of 50 persons (having five images of left and five images of right hand) with total of 500 images and BOSPPHOURS dataset has vein images of 100 persons with total 1575 images. But, while using the transfer learning approach, Al-Johania et al. produced 100% accuracy in training VGG Net-16 model on BOSPPHOURS dataset with split ratios 80–20 and 70–30. The transfer learning is used with 30–50 epochs to achieve a 100% result. With the increase of epochs, the system accuracy is increased but the system takes time to update itself and the time complexity of the system is increased. The accuracy of our proposed method is 99.60% at 10 epochs while Al-Johania et al. have accuracy 84.40% with the same number of epochs.

Figures 10, 11, and 12 represent the accuracy of recognition for different methods. Our proposed method (represented by *F*) shows the higher accuracy in all cases, i.e., good quality dataset, medium quality dataset, and low-quality dataset as GAR 99.6% at FAR 0.4%, GAR 98.5% at FAR 1.5%, and GAR 96.0% at FAR 4.0%, respectively. Genuine accept rate (GAR) is the measure of the percentage of genuine users accepted by a system. Among all cases, the best accuracy is obtained with good quality of dorsal hand vein images. The lowest accuracy with good quality datasets is obtained from VGG Net-19 without fine-tuning. The lowest accuracy with medium and low-quality datasets is obtained from AlexNet-1. The possible reasons could be (i) it has 60 million parameters causing overfitting problem, (ii) it is not deep enough like VGG Net, and (iii) it uses comparatively large convolution filter (5×5).

Table 5 Comparative accuracy of various methods/CNN models with the proposed method

Method	Input image type	Fully connected layer used for recognition	Equal error rate (%)							
			Proposed approach				N. A. Al-Johania et al. [29]			
			Good quality dataset	Medium quality dataset	Low-quality dataset	BOSPPHORUS	Dr. Badawi dataset		BOSPPHOURS dataset	
							K-NN	SVM	K-NN	SVM
A (**VGG Face** without fine-tuning)	Original image	FCL_7	4.001	5.003	7.002	–	–	–	–	–
A-1 (fine-tuned **VGG Face**)		FCL_7	1.002	3.003	7.501	–	–	–	–	–
B (**VGG Net-16** without fine-tuning)		FCL_7	1.481	4.928	7.278	4.920	6.500 A	3.000 D	10.500 G	3.250 J
B-1 (fine-tuned **VGG Net-16**)		FCL_7	1.029	3.021	7.006	–				
C (**VGG Net-19** without fine-tuning)		FCL_7	1.481	4.928	7.278	–	8.000 B	2.000 E	13.000 H	2.500 K
C-1 (fine-tuned **VGG Net-19**)		FCL_7	1.006	6.502	6.502	–				

(continued)

Table 5 (continued)

Method	Input image type	Fully connected layer used for recognition	Equal error rate (%)							
			Proposed approach				N. A. Al-Johania et al. [29]			
			Good quality dataset	Medium quality dataset	Low-quality dataset	BOSPPHORUS	Dr. Badawi dataset		BOSPPHOURS dataset	
							K-NN	SVM	K-NN	SVM
D (Trained **AlexNet-1**)	Difference image	FCL_8	0.901	9.502	8.751	–	9.000 C	5.000 F	27.230 I	13.000 L
E (Trained **AlexNet-2**)		FCL_8	0.503	4.502	6.252	–				
F (fine-tuned **VGG Net-16**)		FCL_8	0.402	1.540	4.001	1.540	–	–	–	–

Split ratios A: 60-40, B: 80-20 and 70-30, C: 80-20 and 70-30, D: 80-20, 70-30 and 60-40, E: 80-20 and 70-30, F: 80-20 and 70-30, G: 80-20, H: 80-20, I: 80-20, J: 70-30, K: 80-20, L: 80-20

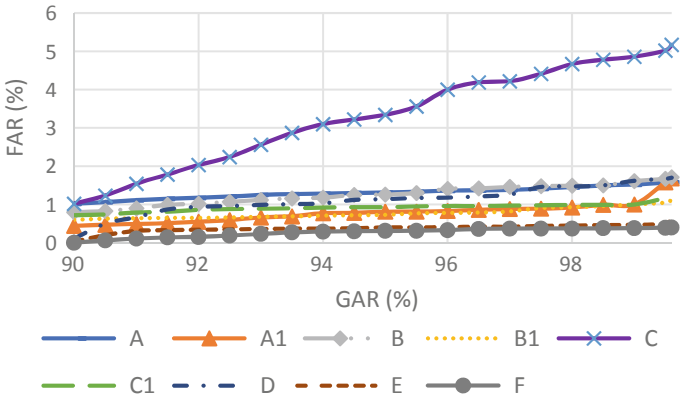

Fig. 10 Genuine accept rate versus false accept rate graph for good quality dataset

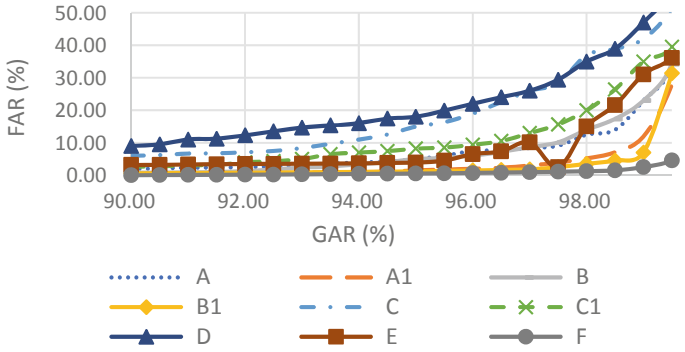

Fig. 11 Genuine accept rate versus false accept rate graph for medium quality dataset

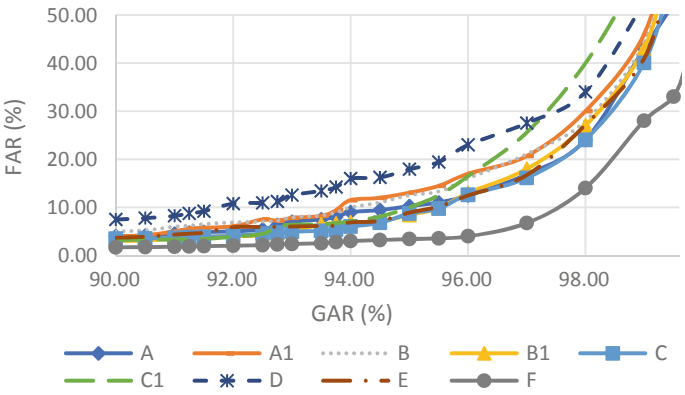

Fig. 12 Genuine accept rate versus false accept rate graph for low-quality dataset

5 Conclusion

Dorsal hand vein recognition based on CNN is robust to dorsal hand vein misalignment as the shading problem is taken care of. In this work, the CNN model VGG Net-16 is fine-tuned with datasets having vein images of three qualities (a) good, (b and d) medium, and (c) low quality. CNN is trained to be robust regarding various environmental changes like temperature, alignment, and different postures. The recognition accuracies of various CNN models are compared in this work. OpenCV, Keras, and Tensorflow with Python 3.7 are used for experimental analysis. We have compared our approach with previous recent work and produces better results in terms of EER. The results of experiments show that the proposed approach (fine-tune of VGG Net-16 on difference image) achieves higher accuracy of recognition (99.60%) as compared to some other CNN structure on different datasets. However, sufficiently high identification rates have been attained due to advances in the biometric system but convenience, robustness, and reliability are the major parameters that can be improved by developing new approaches. The split ratio used in this work is 50–50. As future work, it can be extended with experiments on other split ratios (other than [29] ratios). Instead of dropout, this model can also be trained using batch normalization for better standardization and stability.

References

1. J.K. Gupta, R. Kumar, An efficient ANN based approach for latent fingerprint matching. Int. J. Comput. Appl. **7**(10), 18–21 (2010)
2. V. Jain, A.K. Singh, R. Kumar, Refinement of latent fingerprint. Int. J. Eng. Res. Online **3**(6), 20–24 (2015)
3. M.M. Askarin, K.S. Wong, R.C.-W. Phan, Reduced contact lifting of latent fingerprint, in *Proceedings of IEEE Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, pp. 1406–1410 (2017)
4. W. Liao, Contextual patch feature learning for face recognition. JSW **9**, 1827–1832 (2014)
5. L.C. Jain, U. Halici, I. Hayashi, S.B. Lee, S. Tsutsui, *Intelligent Biometric Techniques in Fingerprint and Face Recognition* (CRC Press Inc., Boca Raton, FL, 1999)
6. F. Jan, I. Usman, S. Agha, Reliable iris localization using Hough transform histogram-bisection and eccentricity. Sig. Process. **9**(93), 230–241 (2013)
7. W. Hao, *Iris Identification Algorithm Study Based on Feature Points* (Peking University, Beijing, 2004)
8. N. Kaur, M. Juneja, A novel approach for Iris recognition in unconstrained environment. J. Emerg. Technol. Web Intell. **6**(2), 243–246 (2014)
9. J. Zhang, J. Yang, Finger-vein image enhancement based on combination of gray-level grouping and circular Gabor filter, in *Proceedings of the International Conference on Information Engineering and Computer Science*, Wuhan, China, 19–20 Dec 2009, pp. 1–4
10. W. Pi, J. Shin, D. Park, An effective quality improvement approach for low quality finger vein image, in *Proceedings of the International Conference on Electronics and Information Engineering*, Kyoto, Japan, 1–3 Aug 2010, pp. 424–427
11. R. Kumar, R.C. Singh, A.K. Sahoo, SIFT based dorsal vein recognition system for cashless treatment through medical insurance. IJITEE **8**(10s), 444–451 (2019)

12. T. Jain, R. Kumar, R.C. Singh, An algorithm for determining SIFT matching score for dorsal vein recognition system. IJITEE **8**(10s), 468–471 (2019)
13. T. Jain, R. Kumar, A study of vein recognition system. Acta Informatica Malaysia (AIM) **3**(1), 13–15 (2019)
14. G. Neha, S. Tuly, Palmprint recognition: a selected review. Int. J. Eng. Sci. **6**(6), 6776–6780 (2016)
15. H.G. Hong, M.B. Lee, K.R. Park, Convolutional neural network-based finger-vein recognition using NIR image sensors. Sensors **17** (2017). https://doi.org/10.3390/s17061297
16. B.A. Rosdi, C.W. Shing, S.A. Suandi, Finger vein recognition using local line binary pattern. Sensors **11**, 11357–11371 (2011)
17. J. Peng, N. Wang, A.A. Abd El-Latif, Q. Li, X. Niu, Finger-vein verification using Gabor filter and SIFT feature matching, in *Proceedings of the 8th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Piraeus-Athens, Greece, 18–20 July 2012, pp. 45–48
18. J.F. Yang, J.L. Yang, Multi-channel gabor filter design for finger-vein image enhancement, in *Proceedings of the Fifth International Conference on Image and Graphics*, Xi'an, China, 20–23 Sept 2009, pp. 87–91
19. K.Y. Shin, Y.H. Park, D.T. Nguyen, K.R. Park, Finger-vein image enhancement using a fuzzy-based fusion method with Gabor and Retinex filtering. Sensors **14**, 3095–3129 (2014)
20. Y.H. Park, K.R. Park, Image quality enhancement using the direction and thickness of vein lines for finger-vein recognition. Int. J. Adv. Robot. Syst. **9**, 1–10 (2012)
21. J. Yang, X. Zhang, Feature-level fusion of global and local features for finger-vein recognition, in *Proceedings of the IEEE 10th International Conference on Signal Processing*, Beijing, China, 24–28 Oct 2010, pp. 1702–1705
22. Y. Lu, S. Yoon, D.S. Park, Finger vein recognition based on matching score-level fusion of Gabor features. J. Korean Inst. Commun. Inf. Sci. **38A**, 174–182 (2013)
23. X. Qian, S. Guo, X. Li, F. Zhong, X. Shao, Finger-vein recognition based on the score level moment invariants fusion, in *Proceedings of the International Conference on Computational Intelligence and Software Engineering*, Wuhan, China, 11–13 Dec 2009, pp. 1–4
24. G. Yang, X. Xi, Y. Yin, Finger vein recognition based on a personalized best bit map. Sensors **12**, 1738–1757 (2012)
25. N. Miura, A. Nagasaka, T. Miyatake, Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification. Mach. Vis. Appl. **15**, 194–203 (2004)
26. J. Yang, Y. Shi, Finger-vein ROI localization and vein ridge enhancement. Pattern Recogn. Lett. **33**, 1569–1579 (2012)
27. Y. Lu, S. Yoon, S.J. Xie, J. Yang, Z. Wang, D.S. Park, Finger vein recognition using generalized local line binary pattern. KSII Trans. Internet Inf. Syst. **8**, 1766–1784 (2014)
28. S.R. Cho, Y.H. Park, G.P. Nam, K.Y. Shin, H.C. Lee, K.R. Park, S.M. Kim, H.C. Kim, Enhancement of finger-vein image by vein line tracking and adaptive Gabor filtering for finger-vein recognition. Appl. Mech. Mater. **145**, 219–223 (2012)
29. N.A. Al-Johania, L.A. Elrefaei, Dorsal hand vein recognition by convolutional neural networks: feature learning and transfer learning approaches. Int. J. Intell. Eng. Syst. **12**(3), 178–191 (2019)
30. B. Sontakke, V. Humbe, P. Yannawar, Dorsal hand vein authentication system: a review. Int. J. Sci. Res. Eng. Technol. **6**(5), 511–514 (2017)
31. M. Shahin, A. Badawi, M. Rasmy, Multimodal biometric system based on near-infra-red dorsal hand geometry and fingerprints for single and whole hands. World Acad. Sci. Eng. Technol. **4**(4), 268–283 (2010)
32. D. Huang, X. Zhu, Y. Wang, D. Zhang, Dorsal hand vein recognition via hierarchical combination of texture and shape clues. Neurocomputing **214**, 815–828
33. J. Lee, T. Lo, C. Chang, Dorsal hand vein recognition based on directional filter bank. Signal Image Video Process **10**(1), 145–152 (2016)
34. R. Trabelsi, A. Masmoudi, D. Masmoudi, Hand vein recognition system with circular difference and statistical directional patterns based on an artificial neural network. Multimedia Tools Appl. **75**(2), 687–707 (2014)

35. Y. Hu, Z. Wang, X. Yang, Y. Xue, Hand vein recognition based on the connection lines of reference point and feature point. Infrared Phys. Technol. **62**(1), 110–114 (2013)
36. X. Zhu, D. Huang, Hand dorsal vein recognition based on hierarchically structured texture and geometry features, in *Proceedings of Chinese Conference on Biometric Recognition*, pp. 157–164 (2012)
37. H. Wan, L. Chen, H. Song, J. Yang, Dorsal hand vein recognition based on convolutional neural networks, in *Proceedings of the IEEE International Conference on Bioinformatics Biomedicine*, pp. 1215–1221 (2017)
38. M. Alaslani, L. Elrefaei, Convolutional neural network based feature extraction for iris recognition. Int. J. Comput. Sci. Inf. Tech. **10**(2), 65–78 (2018)
39. G. Yadav, S. Maheshwari, A. Agarwal, Contrast limited adaptive histogram equalization based enhancement for real time video system, in *Proceedings of the International Conference on Advances in Computing, Communications and Informatics*, pp. 1–6 (2014)
40. M.A. Olsen, D. Hartung, C. Busch, R. Larsen, Contrast enhancement and metrics for biometric vein pattern recognition, in *Proceedings of 6th International Conference on Intelligent Computing, ICIC 2010*, Changsha, China, 18–21 Aug 2010
41. R.S. Choras, Multimodal biometrics for person authentication. IntechOpen (2019). https://doi.org/10.5772/intechopen.85003
42. A.G. Waibel, T.H. Hinton, K. Shikano, K. Lang, Phoneme recognition using time-delay neural networks. IEEE Trans. Acoust. Speech Signal Process. **37**, 328–339 (1989)
43. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in *Proceedings of the 3rd International Conference on Learning Representations*, San Diego, CA, USA, 7–9 May 2015, pp. 1–14
44. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting. J Mach. Learn. Res. **15**, 1929–1958 (2014)
45. https://socrd.org/?page_id=88
46. O.M. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition, in *Proceedings of the British Machine Vision Conference*, Swansea, UK, 7–10 Sept 2015, pp. 1–12