N-shot Palm Vein Verification Using Siamese Networks

Felix Marattukalam[1] , Waleed H. Abdulla[2] , Akshya Swain[3]

Abstract: The use of deep learning methods to extract vascular biometric patterns from the palm surface has been of interest among researchers in recent years. In many biometric recognition tasks, there is a limit in the number of training samples. This is because of limited vein biometric databases being available for research. This restricts the application of deep learning methods to design algorithms that can effectively identify or authenticate people for vein recognition. This paper proposes an architecture using Siamese neural network structure for few shot palm vein verification. The proposed network uses images from both the palms and consists of two sub-nets that share weights to identify a person. The architecture's performance was tested on the HK PolyU multi spectral palm vein database with limited samples. The results suggest that the method is effective since it has 91.9% precision, 91.1% recall, 92.2% specificity, 91.5% F1-Score, and 90.5% accuracy values.

Keywords: Palm Vein Verification, biometrics, Siamese neural network, few- shot learning.

1 Introduction

The need for contactless biometric systems have significantly increased due to the onset of the Covid-19 global pandemic. Although various *extrinsic* modalities like face, iris, and palmprint [Fe18] which are tangible part of the body are successfully being used, now there is a need for *intrinsic* systems like finger vein, hand vein, and palm vein [MA19] which are subcutaneous [Uh20] and not visible to the naked eye. The features in these systems are the veins which helps liveliness detection, and is slightly more robust to spoof attacks that has been a challenge among researchers [Uh20]. Palm vein systems, which is the focus of the present investigation, is an*intrinsic* biometric system and preferred due to the ease of interaction with palm vein scanners. A palm vein biometric system has to go through several stages: *acquisition*, *pre-processing*, *feature extraction*, *decision making*. Here, we briefly look into the functions of every stage. A vein scanner captures the palm vein image using a near infrared camera in the acquisition stage. The veins are visible to the camera when illuminated under infrared light with wavelengths of 760-800 nm [MA19]. The acquisition process suffers from issues due to uneven illumination. These issues are addressed in the pre-processing and feature extraction stages. The incoming images are cropped to a region of interest (ROI), essentially the palm region having maximum vein information to perform recognition. Then the ROI image is processed using image processing methods, and passed on to matching algorithms for decision making. The methods for matching can essentially be classified as traditional and deep learning methods. Since palm vein recognition is a classification problem and deep learning methods have been successful on such tasks, researchers are inclined towards its use.

[1] Faculty of Engineering, The University of Auckland, New Zealand, felix.marattukalam@auckland.ac.nz
[2] Faculty of Engineering, The University of Auckland, New Zealand, w.abdulla@auckland.ac.nz
[3] Faculty of Engineering, The University of Auckland, New Zealand, a.swain@auckland.ac.nz

However, the limitation that deep learning has for palm vein recognition is the need for massive databases with high quality labeled images [Th19]. This is often scarce in the case of palm vein recognition systems. Therefore, our work showcases the use of deep learning networks using limited samples for vein verification. The architecture used is inspired from the Siamese neural network structure, and specifically addresses the verification setting in the recognition system.

The contributions of this research paper are: 1) The advantages of Siamese neural network architecture is exploited for palm-vein recognition by sharing the information from both the palms. As a result, a unique Siamese neural network architecture is developed for palm vein verification 2) the proposed architecture is tested on the HK PolyU multi-spectral palm vein database [Zh09], and its performance is evaluated. The performance evaluation show that this Siamese neural network setting is effective for palm vein verification and useful for palm vein recognition systems.

2   Related work

Numerous methods have been proposed to extract and match vein patterns from vein images. These patterns are used for biometric recognition using different approaches. In this section a few methods are presented for sake of completeness.

The extraction methods can be categorised into subspace learning, local descriptor, vessel geometry, and deep learning. The recent inclination is in using Siamese networks to curb the need for large databases [Th19]. [ES14] elaborates how subspace learning uses obtained coefficients as unique features for recognition. Local descriptor approaches are better described in [XYY17]. A detailed comparison on vessel geometry is discussed in one of our previous works [MA20]. Finally, deep learning approaches such as convolutional neural network(CNN), deep belief network (DBN), and auto-encoders (AE) [Qi21] are used for feature extraction and subject recognition.

As highlighted already,deep learning in biometrics needs for large datasets . In one current public database, the images for each subject or class are limited to twelve [Qi21] and not accounting dynamic class change. This led to the need for alternate approaches. Some researchers proposed to augment the available data. They explored generative adversarial networks (GANs) with data augmentation to improve classification performance [Qi21]. This did achieve reasonable performance but did not solve the problem of system speed, data privacy and storage space associated with duplication of input data. Also, data augmentation can easily lead to overfitting. Researchers are inclined towards more effective methods like similarity learning and few-shot learning. [Sh21] discusses these latest approaches and proposes a few shot learning approach for palmprint recognition, which displays good accuracy. We propose in this paper a novel combined Siamese structure for palm vein verification.

3   Methodology

This section discusses the database used, the general Siamese neural network architecture, and the network structure implemented with the loss functions that have been used to evaluate the network performance.

3.1 Database

The palm vein image database used in our research is the HK PolyU Multispectral Palmprint and Palm Vein database (publicly available) [Zh09] released by Hong Kong Polytechnic University (PolyU) Biometric Research Centre. A Near Infrared Region (NIR) scanner is used to capture the palm vein images. The database released has images from 250 subjects (195 male and 55 female) 20-60 years of age. The total database comprises 6000 images from 500 palms collected in two separate illumination sessions. The images captured are of resolution 352×288 pixels.
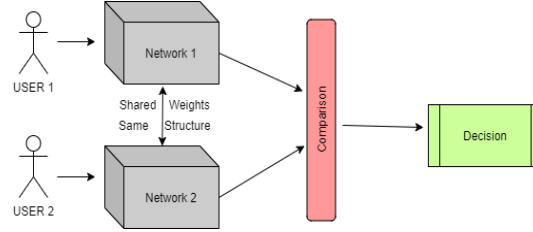
3.2 Siamese Neural Networks

Fig. 1: A typical Siamese neural network structure for biometric system

Even though deep learning algorithms have proven their ability to produce exceptional results, the performance of the designed algorithm is often dependant on the number of data samples available to train the network [Sh21]. The performance of the network improves with the increasing number of data samples. In biometric systems, especially palm vein systems, suitably labelled datasets are not readily available. One-shot or few-shot learning is the appropriate approach when only a few training examples are available for the network to train. The few-shot learning approach uses Siamese neural network. As shown in Fig. 1, a typical Siamese neural network has two paths and aims to find the similarity between its inputs. It has identical parallel networks which share the same architecture and weights. Siamese neural networks was first proposed by Bromley et al [Br93] for signature verification in biometrics and is widely used in face verification tasks. Profound details about Siamese networks for image recognition are available in [Ch21].

3.3 Network Structure

This paper proposes to develop two identical networks that process two images simultaneously and compute the similarity or difference between the two images. If the images are from two different candidates, the network essentially needs to compute the similarity function and increase the distance between them.

Fig.2 shows the overview of the proposed network structure which is based on Siamese architecture. As introduced briefly in section 3.2, the network consists of multiple sub-networks having the same structure and share weights. Here, there are two sub-networks, subnet 1 and subnet 2. Each of them has sub subnets within them to process the input image. The left and right palm images pass through a spatial feature extractor with the convolutional neural network network structure shown in Table 1.
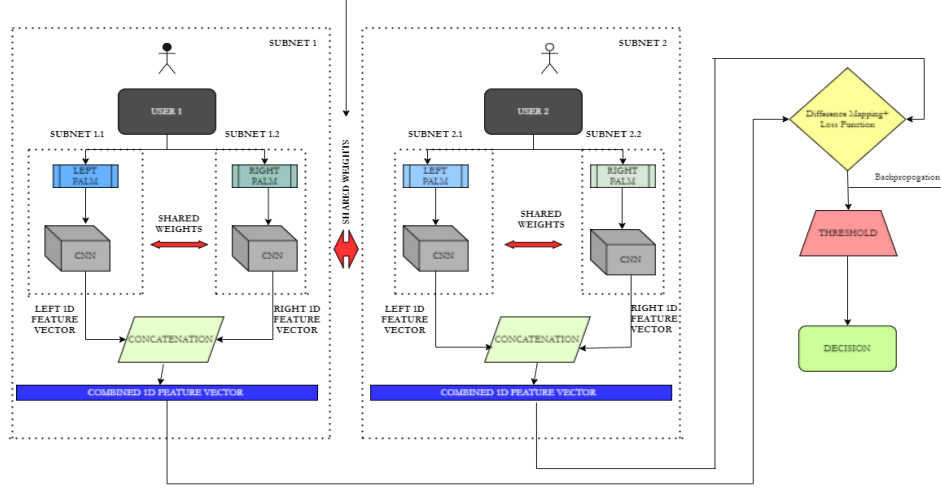
Fig. 2: Overview of the proposed network structure based on Siamese architecture. The red arrows indicate the weights being shared between sub-networks. The sub-networks outputs are 1-D feature vectors. The distance between the feature vectors are then calculated

Tab. 1: CNN feature extractor structure

	Input image	
Layer 1	Convolution 1, 64 x 3 x 3, Stride 1, Padding 0, ReLU	Batch Norm+Max Pool
Layer 2	Convolution 2, 64 x 3 x 3, Stride 1, Padding 0, ReLU	Batch Norm+Max Pool
Layer 3	Convolution 3, 64 x 3 x 3, Stride 1, Padding 1, ReLU	Batch Norm+Max Pool
Layer 4	Convolution 4, 64 x 3 x 3, Stride 1, Padding 1, ReLU	Batch Norm+Max Pool
Layer 5	Fully Connected, 1000 hidden units, ReLU	
Layer 6	Fully Connected, 128 hidden units, Sigmoid	
	Extracted Features	

Consider two users, where each user submits the left and right palm image to the network. Hence, the network receives four images in total, namely, x_1, x_2 and y_1, y_2. The spatial feature extractor network shown in Table 1 generates the feature embeddings $f(x_1)$ and $f(x_2)$ respectively, which are one-dimensional vectors of length 128×1. These vectors are then concatenated together to form $F(X)$. Similar process is followed to obtain $F(Y)$. Then the feature embeddings are subjected to a function E which computes the L_1 distance. The function is given by eqn (1):

$$E(X,Y) = d(X,Y) = ||F(X) - F(Y)|| \tag{1}$$

The function E will be smaller if the concatenated feature vector $F(X)$ is similar to $F(Y)$. This distance value is used to fine-tune the network weights using back propagation. A sigmoid activation function is used to convert the distance to probability P.

3.4 Loss function

Siamese networks classify the inputs into binary classes ie. "1" being same inputs and "0" being different. Contrastive loss and binary cross-entropy function loss are the two common loss options in binary classification.

- Contrastive loss: It requires pairs of input samples. The encoder is penalized by the loss function based on the class of the input image. If the input images are from the same class, the model produces similar feature embeddings. Mathematically it is given by eqn (2):

$$Loss = (1-y) * \frac{1}{2}(d)^2 + (y) * \frac{1}{2}[max(0, m-d)]^2 \tag{2}$$

  Here y is the actual label and will be zero when the embeddings of combined input images (left palm and right palm) and one if they are not same, d is the distance measure between the feature embeddings and the input images, m is the hyper parameter margin which is maximized if the input images are similar. If the input pairs are dissimilar and the distance d is greater than the margin m, no loss is incurred.

- Binary cross-entropy loss: It is also known as log loss and is used to calculate the classifier performance which is in the range between 0 and 1. If the predicted probability varies from actual class, the loss increases. Mathematically it is given by eqn (3):

$$Loss = -ylogp + (1-y)log(1-p)^2 \tag{3}$$

  Here y is the class label and p is the prediction probability. It is used to differentiate between similar and different images by providing the aggregate of positive and negative loss probability.

4 Experimental Results and Analysis

This section discusses about the results based on the experimental setting and the proposed architecture. The performance analysis of the results is done using the matrices namely, accuracy, precision, recall, specificity and F1-score.

4.1 Implementation

This is a k-way n shot classification problem. The dataset D with a data split of 70:30 was used. The training set contains n samples from k-classes adding upto $k \times n$ samples in the training dataset and a query set in the testing dataset. Predominantly there were only two classes i.e. genuine and imposter subjects, and hence, k=2 and n varying from one to five depending on the sample set considered from the existing database. The model was trained using batch of training tasks to ultimately categorize the image during the testing task. At the end of each epoch, the model parameters were updated through back-propagation as per the loss calculated.

The HK PolyU multispectral database consists of uniform images. We used the region of interest (ROI) images of resolution 128×128 pixels by using the method in [LXY16]. The

database was prepared into small subsets of classes containing two, three, four and five ROI images for training depending on the value of *n*. This was subjected to the spatial feature extractor described section 3.3. The experiment was carried out using the Keras framework with Tensorflow on a NVIDIA GTX 2080 8GB GPU with i7 3.3 GHz processor supported with 16 GBs of RAM. The learning rate was set at 0.0001 and Adam Optimizer was used. The one dimensional feature vector extractor which uses the structure shown in Table 1 has fully connected layers having 128 hidden units followed by the sigmoid function.

The model was evaluated using the metrics: accuracy, precision, recall, specificity and F1- score. These parameters were preferred based on our previous study for comparison between SVM with CNN. These parameters are briefly summarised in Table 2.

Tab. 2: Performance metrics using confusion matrix

Accuracy (A)	Precision (P)	Recall (R)	Specificity (S)	F1-Score
$\dfrac{(TP+TN)}{(TP+TN+FP+FN)}$	$\dfrac{(TP)}{(TP+FP)}$	$\dfrac{(TP)}{(TP+FN)}$	$\dfrac{(TN)}{(TN+FP)}$	$\dfrac{(2\times TP)}{(2\times TP+FP+FN)}$

Here, the number of predictions where the classifier correctly predicts the positive class as positive is True Positive (TP), the negative class as negative is True Negative (TN), the negative class as positive is False Positive (FP), and the positive class as negative is False Negative (FN).

4.2   Results and Discussion

An ideal result for this experiment would be obtaining a classification accuracy of 100 %, precision, recall and specificity values of unity, and a 100 % F1 score. Based on literature it can be said that such results are rare for deep learning models with small datasets. The goal of this study specifically is to exploit the benefits of Siamese neural network in palm-vein verification by discussing its performance parameters and establish its use in an end-to-end palm vein recognition system.
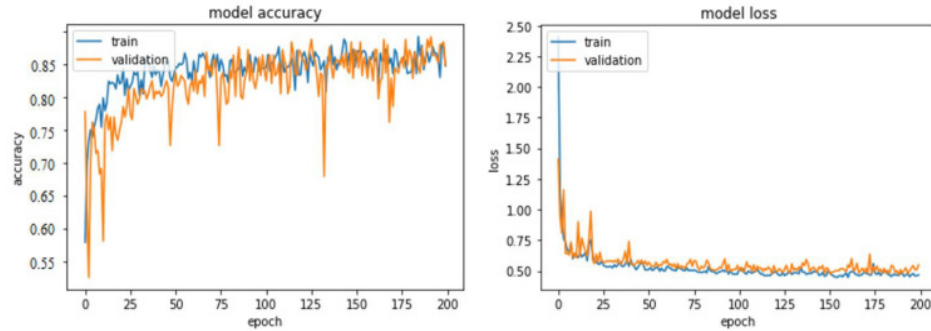

Fig. 3: Accuracy and loss plots: Contrastive loss k=2, n=5

Experiments were performed to evaluate the network performance for different k-way,n-shot learning iterations using both contrastive and cross-entropy losses for classification. However, the results graphically represented in fig.3 are for contrastive loss function as it was seen to be more effective than cross-entropy loss and is in line with what has been

Tab. 3: Results using contrastive loss for 2-way, n-shot settings using both palm images with n varying from 2 to 5.

Model	Accuracy	Recall	Precision	Specificity	F1-Score
k=2, n=2	0.862	0.867	0.874	0.881	0.871
k=2, n=3	0.881	0.885	0.892	0.899	0.889
k=2, n=4	0.892	0.897	0.906	0.911	0.903
k=2, n=5	0.905	0.911	0.919	0.922	0.915

reported in [Li18]. The results obtained in the experiment show that the model training and validation accuracy and loss using contrastive loss function were more stable and merged better than the plots generated for cross-entropy (even though cross-entropy used lower epochs than contrastive function). Also, Siamese neural networks use the principle of similarity between image pairs. As the experiment revolves around n-shot learning, the observations were based on how the model performance varied for different n values. As mentioned in section 4.1, the main results discussed here are for k=2 and n varying from 2 to 5.

Table 3 shows the performance metrics used. Here, contrastive loss was used and n varies from 2 to 5. The results show that the proposed network model performance metrics increase steadily as the number of shots/ support samples is varied sequentially. This is justified because the model takes maximum benefit of the increased number of available palm vein image pairs which helps to differentiate a similar image from non-similar ones. Adam optimizer was used along with dropout prevention techniques and relevant learning rate reduction classes in Keras to pause the training process as soon as stagnancy is detected thus reducing overfitting.

5 Conclusion

This paper discusses the dynamics and benefits of palm vein verification and proposes a state-of-the-art deep learning Siamese neural network that can be used in contactless biometric systems. This is achieved by integrating a k-way n-shot learning network model with contrastive loss and an optimized CNN feature encoder. The results highlight that the model for k=2, n-shot learning settings using contrastive loss function is effective. The best case amongst the experiments was the 5-shot learning setting that provides an accuracy of 90.5% in verifying the palm vein image with good recall (91.1%) and specificity (92.2%). These results are critical performance estimates in medical and biometric applications. The results obtained are promising owing to the fact that this model is trained with only a limited sample set of five samples from each palm for a given training class/subset.

References

[Br93]   Bromley, Jane; Guyon, Isabelle; LeCun, Yann; Säckinger, Eduard; Shah, Roopak: Signature verification using a" siamese" time delay neural network. Advances in neural information processing systems, 6:737–744, 1993.

[Ch21]   Chicco, Davide: Siamese neural networks: An overview. Artificial Neural Networks, pp. 73–94, 2021.

[ES14]   Elnasir, Selma; Shamsuddin, Siti Mariyam: Proposed scheme for palm vein recognition based on Linear Discrimination Analysis and nearest neighbour classifier. In: 2014 International Symposium on Biometrics and Security Technologies (ISBAST). IEEE, pp. 67–72, 2014.

[Fe18]   Fei, Lunke; Lu, Guangming; Jia, Wei; Teng, Shaohua; Zhang, David: Feature extraction methods for palmprint recognition: A survey and evaluation. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 49(2):346–363, 2018.

[Li18]   Lian, Zheng; Li, Ya; Tao, Jianhua; Huang, Jian: Speech emotion recognition via contrastive loss under siamese networks. In: Proceedings of the Joint Workshop of the 4th Workshop on Affective Social Multimedia Computing and First Multi-Modal Affective Computing of Large-Scale Multimedia Data. pp. 21–26, 2018.

[LXY16]   Lin, Sen; Xu, Tianyang; Yin, Xinyong: Region of interest extraction for palmprint and palm vein recognition. In: 2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). IEEE, pp. 538–542, 2016.

[MA19]   Marattukalam, Felix; Abdulla, Waleed H: On palm vein as a contactless identification technology. In: 2019 Australian & New Zealand Control Conference (ANZCC). IEEE, pp. 270–275, 2019.

[MA20]   Marattukalam, Felix; Abdulla, Waleed H: Segmentation of Palm Vein Images Using U-Net. In: 2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). IEEE, pp. 64–70, 2020.

[Qi21]   Qin, Huafeng; El-Yacoubi, Mounim A; Li, Yantao; Liu, Chongwen: Multi-Scale and Multi-Direction GAN for CNN-Based Single Palm-Vein Identification. IEEE Transactions on Information Forensics and Security, 16:2652–2666, 2021.

[Sh21]   Shao, Huikai; Zhong, Dexing; Du, Xuefeng; Du, Shaoyi; Veldhuis, Raymond NJ: Few-Shot Learning for Palmprint Recognition via Meta-Siamese Network. IEEE Transactions on Instrumentation and Measurement, 2021.

[Th19]   Thapar, Daksh; Jaswal, Gaurav; Nigam, Aditya; Kanhangad, Vivek: PVSNet: Palm vein authentication siamese network trained using triplet loss and adaptive hard mining by learning enforced domain specific features. In: 2019 IEEE 5th International Conference on Identity, Security, and Behavior Analysis (ISBA). IEEE, pp. 1–8, 2019.

[Uh20]   Uhl, Andreas; Busch, Christoph; Marcel, Sébastien; Veldhuis, Raymond: Handbook of vascular biometrics. Springer Nature, 2020.

[XYY17]   Xi, Xiaoming; Yang, Lu; Yin, Yilong: Learning discriminative binary codes for finger vein recognition. Pattern Recognition, 66:26–33, 2017.

[Zh09]   Zhang, David; Guo, Zhenhua; Lu, Guangming; Zhang, Lei; Zuo, Wangmeng: An online system of multispectral palmprint verification. IEEE transactions on instrumentation and measurement, 59(2):480–490, 2009.