Finger Vein Recognition and Intra-Subject Similarity Evaluation of Finger Veins using the CNN Triplet Loss

Georg Wimmer* and Bernhard Prommegger* and Andreas Uhl
Department of Computer Sciences
University of Salzburg
Salzburg, Austria
Email: {gwimmer, bprommeg, uhl}@cs.sbg.ac.at
*: These authors contributed equally

***Abstract*—Finger vein recognition deals with the identification of subjects based on their venous pattern within the fingers. There is a lot of prior work using hand crafted features, but only little work using CNN based recognition systems. This article proposes a new approach using CNNs that utilizes the triplet loss function together with hard triplet online selection for finger vein recognition. The CNNs are used for three different use cases: (1) the classical recognition use case, where every finger of a subject is considered as a separate class, (2) an evaluation of the similarity of left and right hand fingers from the same subject and (3) an evaluation of the similarity of different fingers of the same subject. The results show that the proposed nets achieve superior results compared to prior work on finger vein recognition using the triplet loss function. Furtherly, we show that different fingers of the same subject, especially symmetric fingers (same finger type but from different hand), show enough similarities to perform recognition. The last statement contradicts the current understanding in the literature for finger vein biometry, in which it is assumed that different fingers of the same subject are unique identities.**

I. Introduction

Vascular pattern based biometric systems, commonly denoted as vein biometrics, offer several advantages over other well-established biometric recognition systems. In particular, hand and finger vein systems have become a serious alternative to e.g. fingerprint based ones for several applications. Vein based systems use the structure of the blood vessels inside the human body, which becomes visible under near-infrared (NIR) light. As the vein structure is located inside the human body, it is resistant to abrasion and external influences on the skin. Furthermore, due to the bloodflow exhibited in NIR finger vein videos, liveness detection techniques can be applied to prevent presentation attacks [1], [2].

Most finger vein recognition systems use hand crafted features which are based on the pattern of the vascular network inside the finger. Classical correlation based systems, such as *Maximum Curvature* (MC) [3], *Principal Curvature* (PC) [4] or the *Wide Line Detector* (WLD) [5], use the binarized vein pattern as template and determine the similarity between two samples based on their correlation. More sophisticated methods, e.g. *Deformation Tolerant Feature Point Matching* [6] or *Finger Vein Recognition with Anatomy Structure Analysis* [7] are still based on the vein pattern, but apply matching schemes that are tolerant to misplacements of the finger (e.g. longitudinal finger rotation) to a certain extent. Other methods which utilize minutiae extracted from the vein pattern (e.g. [8]), are based on the texture of the finger vein image (e.g. LBP [9] or LDP [10]) or use keypoint based feature extraction schemes (e.g. SIFT based [11]). Nowadays, finger vein recognition systems using convolutional neural networks (CNN), e.g. [12], [13], [14], are getting more attention.

The key objective of this work is to refine the usage of the triplet loss function [15] in finger vein recognition. Xie and Kumar [13] already successfully applied the triplet loss using the *Light CNN* (LCNN) [16] architecture. They select the training samples (triplets) randomly. Our proposed approach improves on the approach in [13] by applying a more refined triplet selection to improve the effectiveness of CNN training, the so called hard triplet online selection [15], and by the utilization of better suited network architectures (SqNet [17] and ResNet [18]).

Another aim of this work is to examine if there are similarities between finger vein images of different fingers of the same subject. In finger vein recognition, each finger of each subject is considered as a separate class and it is assumed that different fingers of the same subject do not share many similarities since they (indeed) have quite different vein structures. In this work it is examined if this general understanding really holds true. Although the finger vein structures are definitely different for fingers of the same person, this not necessarily applies for other characteristics visible in finger vein images as e.g. thickness of the veins or the finger itself. Especially left and right hand fingers of the same type and subject (e.g. left index finger and right index finger, furtherly denoted as symmetric fingers) may share sufficient similarities to be applied in finger vein recognition systems where a subject can be identified using a finger, even though only the same finger of the other hand has been acquired during enrolment. It is not clear if vein images from symmetric fingers share sufficient similarities to apply recognition. For palm print identification, it was shown in [19] that a subject can be recognized through his/her left palm print even when only the other one has been enrolled. For finger vein biometry there is only one prior work [20] on this subject, with the outcome that there are hardly any similarities between symmetric fingers using several hand-crafted state-of-the-art approaches and one CNN approach. In this paper we will show that our proposed approach is able to recognize subjects using fingers that have not been enrolled.

The reminder of this paper is organized as follows: After a short introduction on the usage of CNNs in finger vein biometry (section II), the triplet loss function and hard triplet online selection are presented in section III. The CNN architectures used for the experiments of this work are described in section III-A. Section IV describes the experimental setup. The results are shown in section V, followed by a short discussion on the results (section VI) and the conclusion together with an outlook on planned further work (section VII).

II. CNNs in Finger Vein Recognition

Convolutional neural networks (CNN) are gaining more and more interest in computer vision. The increase in computational power based on GPUs has led to more sophisticated and deeper architectures which have proven to be the state-of-the art in image classification in various challenges. In biometric applications like finger vein recognition, the problem with the common CNNs is that the CNNs are only able to identify those subjects which have been used for the training of the neural network. If new subjects are added in a biometric application system, then the nets need to be trained again or else a new subject can only be classified as one of the subjects that were used for training (the one that is most similar to the newly added subject with respect to the CNN). This of course makes the practical application of common CNNs impracticable for biometric applications. Some previous publications using CNNs for finger vein recognition just ignored this problem and used common CNNs for finger vein identification (e.g. [21], [22]). Other publications using CNNs used a more practicable approach by training the CNNs to not directly classify images but to compute a similarity measure between pairs of vein images, which also allows the identification of subjects that were not used for the training of the CNNs. For example the authors of [23] use difference images of pairs of finger vein images as inputs to train CNNs and the authors of [24] create 2-channel input images by combining two finger vein images (each channel is one image). Both approaches train with positive (2 images from the same class) and negative pairs (2 images from different classes) to enable the CNNs the distinction between genuine and imposter attempts. A more elegant approach to apply CNNs in practical applications is applied in [13] using the triplet loss function [15]. By using the triplet loss function, CNNs learn to quantify the similarity between images. As input three images are required, two images from identical classes and one of a different class. Then the net is trained to minimize the distance between images of same classes and maximize the distance between different classes.

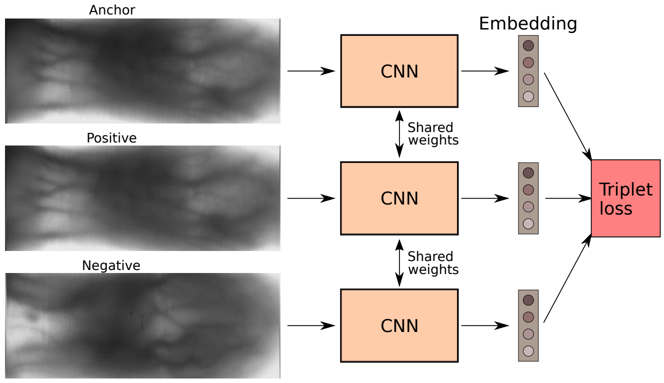


Fig. 1: CNN training using the triplet loss

III. The Triplet Loss Function and Hard Triplet Online Selection for CNNs

Contrary to more common loss functions like e.g. the SoftMax loss, the triplet loss does not directly learn the CNN to classify images to their corresponding classes. The triplet loss function does not even require to know the class affiliation of the training images. Per training step the triplet loss requires three input images at once (a so called triplet), where two images belong to the same class (the anchor image and a sample from the same class, further denoted as positive) and the third belongs to a different class (further denoted as negative). The triplet loss learns the network to minimize the distance between the anchor and the positive and maximize the distance between the anchor and the negative.

The triplet loss using the squared Euclidean distance is defined as follows:

$$\begin{aligned} L(A,P,N) = &\max(||f(A)-f(P)||^2 \\ &- ||f(A)-f(N)||^2+\alpha, 0), \end{aligned} \tag{1}$$

where A is the anchor, P the positive and N the negative. α is a margin that is enforced between positive and negative pairs and is set to $\alpha = 1$. $f(x)$ is an embedding (the CNN output of an input image x). Figure 1 shows the scheme of learning a CNN using the triplet loss. A triplet of training images (anchor, positive and negative) is fed through the CNN resulting in an embedding for each of the three images. The embeddings of the three images are then used to compute the triplet loss to update the CNN.

Summarized this means the CNN is trained to create an embedding $f(x)$, from an image to the feature space $\mathbb{R}^d$, such that the squared distances between all finger vein images of the same class (finger) is small, whereas the squared distance between any pairs of finger vein images from different classes is large. The function of the margin α is that the finger vein images of one class are not projected to only one single point in the embedding space but to live in a manifold, while still enforcing the distance and thus discriminability to other classes.

An important point for the training of the CNNs is the selection of the input triplets. Generating random triplets, as proposed in [13], results in many triplets that are easily satisfied (i.e. fulfill the constraint in Equation (1)). These triplets do not contribute to the training and result in slower convergence, as they would still be passed through the network. It is crucial to select so called hard triplets, that are active and can therefore contribute to improving the model. For the proposed approach, the training triplets are randomly selected to fulfill the following condition for a given anchor and margin α within a batch of images (denoted as hard triplet online selection [15]):

$$||f(A) - f(P)||^2 + \alpha > ||f(A) - f(N)||^2. \qquad (2)$$

As image representation (feature vector) for the classification of a finger vein image we employ the embedding $f(x)$ of a finger vein image x.

A. CNN Architectures

In this work, three different network architectures, SqNet [17], Light CNN (LCNN) [16] and ResNet50 [18] are employed. SqNet and LCNN are both small neural networks that were specifically created to have few parameters and only small memory requirements. The size of the nets and their memory requirements are essential for the training of CNNs using the triplet loss with hard triplet online selection since this kind of training requires big batch sizes as described later in Section VI. ResNet is a network architecture that utilizes skip connections, or shortcuts to jump over some layers. In that way the ResNet deals with the problem of vanishing gradients, a problem that occurs for deeper neural networks. This allows the construction of deeper architectures with more layers.

The LCNN is learned from scratch (as in [13]) and the two other nets are pre-trained on the ImageNet database (http://www.image-net.org/). The input images are resized to the required input sizes of the CNNs (SqNet and ResNet: $3\times224\times224$ (each color channel is the same as the grey-scale image), LCNN: $1\times256\times256$). For each net, the size of the last layers convolutional filter is adapted so that a 256-dimensional output (embedding) is produced.

IV. Experimental Setup

In this work, a 2-fold cross validation is employed. Each fold consists of the images from half of the subjects. First, one fold is used for CNN training and the other for evaluation. In the second iteration the training and evaluation folds are interchanged. We report the mean equal error rate (EER) over the two EERs from the two folds.

The employed similarity metric to measure the similarity between the 256 dimensional CNN outputs of different images (genuine and imposter scores) is derived from the Euclidean distance (which is a natural choice since the triplet loss function is based on the Euclidean distance). To transform the Euclidean distance to a similarity metric, the Euclidean distances are inversed ($d \rightarrow 1/d$) and normalized (for each fold separately) so that the resulting similarity values range from zero to one. For the computation of the EERs, we employ all genuine and imposter scores instead of using only a subset of the impostor scores like often done in other publications on finger vein recognition.

The nets are trained with the triplet loss for 400 epochs, starting with a learning rate of 0.001 that is divided by 10 every 120 epochs. Training is performed on batches of 128 images (images from 32 different classes with 4 images per class) except for the ResNet, where the batch size is reduced to 32 (images from 8 different classes with 4 images per class) because ResNet is the clearly biggest net and oversteps the available GPU memory (12 GB) for bigger batch sizes. The classes and the images per class in a batch are randomly selected. The embeddings of the images of a batch are computed and then 128 (32 for ResNet) hard triplets (triplets that fulfill Equation (2)) are selected within this batch to train and update the net. Each image of a batch is once employed as anchor for a triplet, the positive and negative samples of each triplet are randomly chosen within those samples of the batch that generate hard triplets together with the chosen anchor image.

The CNNs are implemented using the *PyTorch* framework [25].

The main experiment aims to find out how well suited the triplet loss is for finger vein recognition. For this experiment, the finger vein images of the data sets are grouped according to the common standard in finger vein recognition: each finger of a subject belongs to a separate class.

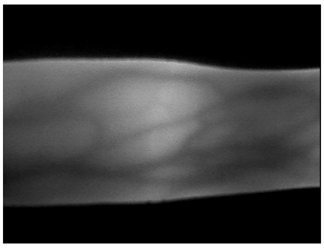
The other two experiments aim to investigate whether there are similarities between different fingers of the same person. For this, we change the class memberships of the finger vein images for training as well as evaluation: for the analysis of symmetric finger similarities (SFS), fingers of the same subject and type (index, middle and ring finger) but different hand (left, right) are grouped into one class (e.g. left and right index finger), for the subject based finger similarities (SBFS), all fingers of the same subject are grouped together, respectively. In this way, the nets are trained to build features that are shared by fingers of same subjects while still being able to distinguish between different subjects.

For SFS and SBFS, the objective is to find out if the trained nets are able to identify:

- the finger type and subject of a finger vein image even though the considered finger was not enrolled but only its symmetric counterpart on the other hand (SFS).
- the subject using finger vein images of a finger that was not enrolled but only the other fingers of the subject (SBFS).

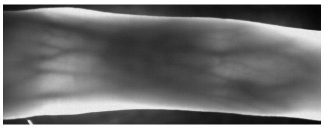
For the computation of the EERs this means that we use the changed class membership assignment and the genuine scores consist of:

- only the similarity measures between symmetric fingers (same finger type and subject, different hand side) but not the similarity measures between images of identical fingers (SFS).
- only the similarity measures between images of different fingers from the same subject (SBFS).

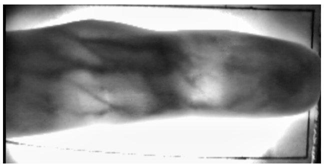


(a) SDUMLA

(b) PLUS



(c) UTFVP



(d) HKPU

Fig. 2: Example images from the employed datasets

A. Databases

For the experiments, four publicly available finger vein databases were used. The data sets under investigation are:

- **SDUMLA [26]** is a multimodal biometric database that contains samples for face, gait, iris, fingerprint and finger veins from 106 individuals. The finger vein subset contains six fingers (ring, middle and index finger from both hands) per subject, captured in one session taking six images of each finger.
- The *University of Twente Finger Vascular Pattern Database* (**UTFVP**) [27] contains six fingers (ring, middle and index finger from both hands) from 60 volunteers in two sessions. At each session two samples per finger were captured.
- The third data set (**PLUS**) is a combination of the *PLUSVein-FV3 Finger Vein Data Set* [28] and the *PROTECT Multimodal DB* [29]. Both data sets contain palmar and dorsal images from the ring, middle and index finger of the left and right hand and have been acquired using the same capturing devices [30]. In this paper only the palmar images acquired by the LED version are used.
- *The Hong Kong Polytechnic University Finger Image Database* (**HKPU**) [31] contains finger vein and finger texture images of 156 subjects from two fingers (index and middle finger of one hand) acquired in two separate sessions.

Sample images of the vein images contained in the chosen data sets are depicted in Fig. 2.

The finger detection, finger alignment and ROI extraction for SDUMLA, PLUS and UTFVP is done as described in [32]. For HKPU, where the finger detection is more challenging, the finger masks provided in [31] are used to align the images. The ROI extraction was again carried out as described in [32].

The four examined data sets have different properties with regard to the ease of their processing. UTFVP and PLUS do not pose any major problems: in both, the NIR illumination and the positioning of the fingers are consistently good over the entire data set. The HKPU contains some partially overexposed images and therefore, the segmentation of the fingers is more difficult. The SDUMLA suffers mainly from finger misplacements, especially due to longitudinal finger rotation (see [33]). This difficulties might be the reason why the recognition results for UTFVP and PLUS are superior to those of HKPU and especially SDUMLA in most publications, especially if classical hand crafted features are used.

V. Results

Table I depicts the recognition results (EER) for the experiments carried out using the classical finger vein recognition scenario (each finger of a subject is considered as a separate class). The used CNN architectures are the ones described in section III-A: Triplet-SqNet, Triplet-ResNet and Triplet-LCNN. In order to be able to better assess the results, two classic vein pattern-based methods, namely *Maximum Curvature* (MC) [3] and *Principal Curvature* (PC) [4], a texture-based method (*local binary patterns*, LBP [34]) and a *SIFT* based approach [35] are applied to the same data sets.

Methods	SDUMLA	UTFVP	PLUS	HKPU
Triplet-SqNet	**2.7**	2.5	2.4	3.7
Triplet-ResNet	3.1	3.6	3.2	5.6
Triplet-LCNN	4.9	4.6	4.7	10.0
MC	4.0	**0.2**	0.5	**1.0**
PC	4.9	0.4	**0.2**	1.3
LBP	7.3	1.5	3.6	4.0
SIFT	5.4	1.5	0.8	1.8

TABLE I: Recognition performance (EER in [%]) of the proposed method using three different CNN architectures, as well as four approaches using classical hand crafted features, on four publicly available finger vein databases.

As expected, the hand-crafted approaches perform best for UTFVP and PLUS (consistently good vein images) followed by HKPU (issues with overexposed images). The results for SDUMLA (problem with finger misplacement) are noticeable inferior. The same holds true for LBP and SIFT, although the LBP results on PLUS are close to those on HKPU. For all four data sets, the best recognition results of the proposed CNN architectures are attained with SqNet followed by ResNet, whose results are slightly worse. The LCNN results are clearly inferior. In general, the recognition performance of the three CNN architectures are quite similar for UTFVP, PLUS and SDUMLA. For HKPU, the results are noticeable worse. The overall best results for UTFVP, PLUS and HKPU are achieved utilizing the classical vein pattern based methods MC and PC. For SDUMLA, the most challenging data set, the best results are attained using Triplet-SqNet. Both, Triplet-SqNet and Triplet-ResNet outperform all hand-crafted systems, even the vein pattern based approaches.

As mentioned in section II, there is one prior publication [13] that applied the triplet loss for finger vein recognition using the LCNN network. The experiments in [13] were applied to the HKPU database using the enhanced and ROI extracted images of the database (the HKPU data set provides the original images, ROI masks (which are not always correct) and ROI extracted images processed with image enhancement methods), whereas for the experiments in this paper, the ROIs as described in section IV-A are used. To have a fair comparison of the results of both papers, the best performing net (Triplet-SqNet) of this paper and the net used in [13] (Triplet-LCNN) are applied to the same image database, the enhanced images as provided in [31]. The results on the enhanced images of the HKPU data set are presented in Table II. The results from [13] are shown with and without supervised discrete hashing (SDH). SDH was applied to the CNN output in order to reduce the template/storage size. It can be observed, that the results of the proposed approach (triplet loss together with hard triplet online selection) are clearly superior to those in [13].

[13] with SDH	[13] without SDH	Triplet-LCNN	Triplet-SqNet
9.8	13.1	7.2	5.3

TABLE II: Recognition performance (EER in [%]) on the HKPU database using the contrast enhanced ROIs provided in [31]

Contrary to classical finger vein recognition, the second part of the experiments examines if different fingers of the same subject contain enough similarities to identify the correct person. In detail, the similarity between symmetric fingers (SFS) and between different fingers of the same person (SBFS) in general is examined. Since the HKPU database consists only of images from two different fingers of the same hand, it is not possible to apply these experiments on this data set.

Table III presents the results for the experiments evaluating SFS. The best recognition results are attained for the PLUS database. The best performing net, Triplet-ResNet, achieves an EER of 10%. Triplet-SqNet, reaching an EER of 11.3% is only slightly worse. The evaluation on the SDUMLA data set results in EERs just above 15%, and for UTFVP above 20%, respectively. Again, Triplet-LCNN gives the worst results. This results indicate, that there are indeed clear similarities between symmetric fingers. If not, the resulting EERs would be close to 50%.

Methods	SDUMLA	UTFVP	PLUS
Triplet-SqNet	15.2	23.0	11.3
Triplet-ResNet	16.4	20.4	10.1
Triplet-LCNN	18.2	25.5	17.1

TABLE III: Recognition performance (EER in [%]) for symmetric finger similarity (SFS)

Piciucco et al [20] applied a similar experiment on the SDUMLA database (Test-4 in [20]) using three hand-crafted finger vein detection methods and a CNN (DenseNet) using the cross-entropy loss. The experimental setup of Test-4 in [20] is quite similar to the one for SFS in this work: Symmetric fingers belong to the same class and for the genuine scores only the similarity scores between symmetric fingers were used. Since the CNN in [20] uses the cross-entropy loss, using different subjects for training and evaluation is impossible and a less practice-oriented approach had to be employed using 83% of the images per class for training and the remaining images for evaluation.

The three hand-crafted methods in [20] achieved EERs between 45 and 47%. These results are not surprising, since the three methods are solely focusing on the vein structure, which is clearly different for symmetric fingers. However, also the CNN approach achieved an EER of only 32.6 %, although specifically trained for symmetric finger recognition. Hence the authors in [20] concluded that there are no significant similarities between symmetric fingers. In our experiments however we achieve EERs down to almost 15% on the SDUMLA database and EERs down to 10% on the PLUS database, which leads to a contrary conclusion.

The results for the subject based finger similarity (SBFS) are listed in Table IV. The observed EERs, which are in the range of 20% to 30%, are clearly worse than those for SFS. Still, the results indicate that there are distinct subject specific similarities in finger vein images. However, the similarities between images from different fingers of the same subject are not high enough to be used for a recognition systems where the subject is identified using fingers that were not used for enrolment. So, the similarities between symmetric fingers are clearly higher as between random fingers of one subject. This of course is not surprising since symmetric fingers are supposed to share characteristics such as finger thickness, vein visibility, vein width or finger shape, which clearly does not hold true to the same extent for different finger types of the same person.

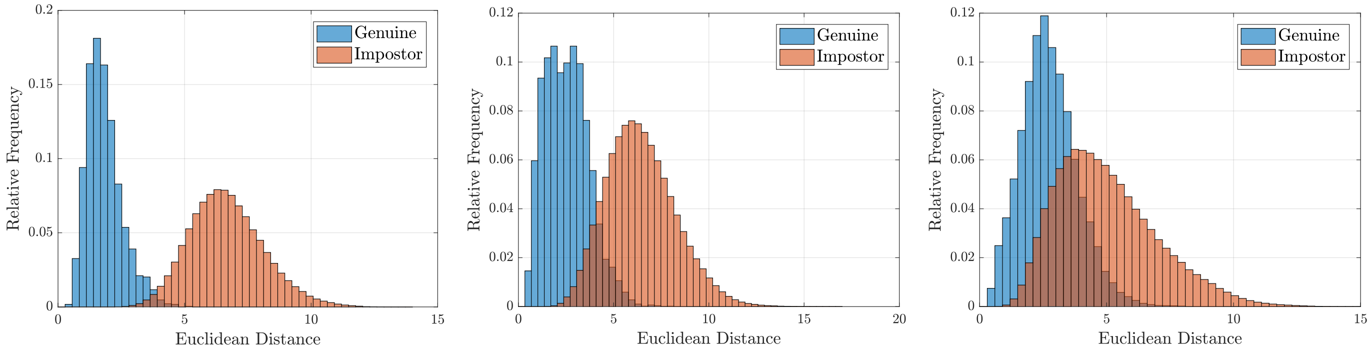
Methods	SDUMLA	UTFVP	PLUS
Triplet-SqNet	22.8	26.4	23.3
Triplet-ResNet	21.9	24.3	26.9
Triplet-LCNN	25.3	30.4	30.1

TABLE IV: Recognition performance (EER in [%]) for subject based finger similarity (SBFS)

Fig. 3 depicts the distribution of the distances of the three employed experiments, the classical finger vein recognition scenario (recognition), symmetric finger similarity (SFS) and subject based finger similarity (SBFS) for the best performing net, Triplet-SqNet. For classical recognition, the two distributions are nicely separated. For SFS, the two distributions can still be clearly distinguished. However, the overlapping area is already noticeable larger. For SBFS, the distributions are getting closer to each other and the overlapping part is distinctly larger than for SFS.

VI. Discussion

From the results in Table I we can clearly observe that the best results are achieved for the Triplet-SqNet. However, it is not clear if the results of SqNet are better because it is the best suited network architecture to identify finger vein images using the triplet loss or if it is because of the different training



(a) Recognition (b) SFS (c) SBFS

Fig. 3: Distributions of the resulting distances for the three experiments (Recognition, SFS, SBFS) using the best performing net (Triplet-SqNet)

configurations. There are two main differences between the training configurations of the three nets. LCNN is the only net that was not pre-trained on another database. SqNet and ResNet have been pre-trained on the ImageNet database. ResNet is the only net with a smaller batch size for training (because of GPU memory issues). The smaller batch size has an impact on the triplet selection since the hard triplets are chosen within a batch of images. For a higher batch size, there is a higher chance that hard triplets can be found which still properly help to train a net especially on the later stages of training, when nearly all possible combinations of images as triplets does not produce a positive loss value (> 0) and hence do not contribute to the training of the net.

To find out if the SqNet is really the best choice or if the training configurations (batch size and pre-training) are the reason for its superior results, the results of the SqNet using three different training configurations are compared to the results of ResNet and LCNN. SqNet is trained once with the standard parameters (batch size 128, pre-trained) like in Table I, once with a batch size of 32 using the pre-trained version of the net (like ResNet) and once we train the net without any pre-training (from scratch) with batch size 128 (like LCNN). The results on the four data sets are presented in Table V.

Net	Configuration	SDUMLA	UTFVP	PLUS	HKPU
	standard	2.7	2.5	2.4	3.7
SqNet	batch size 32	3.6	4.0	4.1	5.9
	no pre-training	3.0	3.7	3.2	7.5
ResNet	batch size 32	3.1	3.6	3.2	5.6
LCNN	no pre-training	4.9	4.6	4.7	10.0

TABLE V: Recognition performance (EER in [%]) for different training configurations of the Triplet-SqNet compared to the results of the other two nets. Standard configuration means a batch size of 128 images and a pre-trained net.

As listed in Table V, pre-trained nets perform better than nets without pre-training and a higher batch size leads to better results. When applying the same training configuration (batch size = 32, pre-trained net), then ResNet performs slightly better than SqNet. That means that deeper net architectures seem to achieve superior results. However, since higher batch sizes are not possible for the ResNet using the triplet loss because of limited GPU memory, the SqNet is the better choice if one does not have access to very expensive, deep learning specific hardware for the training of the nets. The LCNN is definitely the worst performing net of the three employed nets.

VII. Conclusion

In the experiments we showed, that CNNs using the triplet loss function combined with hard triplet online selection are perfectly suited for finger vein recognition. Compared to previously proposed CNNs using triplet loss function without hard triplet online selection, i.e. proposed in [13], we could noticeable improve the recognition results. We have also shown that symmetrical fingers (same finger type but different hand, e.g. left and right index finger) share enough similarities to identify people. This disproves the results that were presented in [20]. Furthermore, we showed that different fingers of the same person also exhibit similarities, but these similarities (at least in our experimental setup) are not sufficient for recognition.

The results for the standard finger vein recognition use case presented in Table I, especially those of SDUMLA, indicate that CNN-based methods may be less prone to finger misplacement, including longitudinal finger rotation. In our future work we plan to examine the robustness of the proposed CNN architectures (using the triplet less function together with hard triplet online selection) to longitudinal finger rotation. Furthermore, driven by the surprising results for symmetric finger identification, we will investigate whether other similarities, such as e.g. sex or age, can be predicted using CNNs for finger vein input images.

Acknowledgments

This project was partly funded from the FFG KIRAS project AUTFingerATM under grant No. 864785 and the FWF project "Advanced Methods and Applications for Fingervein Recognition" under grant No. P 32201-NBL.

References

[1] J. Y. Bok, K. H. Suh, and E. C. Lee, "Detecting fake finger-vein data using remote photoplethysmography," *Electronics*, vol. 8, no. 9, p. 1016, 2019.

[2] R. Raghavendra and C. Busch, "Exploring dorsal finger vein pattern for robust person recognition," in *2015 International Conference on Biometrics (ICB)*, May 2015, pp. 341–348.

[3] N. Miura, A. Nagasaka, and T. Miyatake, "Extraction of finger-vein patterns using maximum curvature points in image profiles," *IEICE transactions on information and systems*, vol. 90, no. 8, pp. 1185–1194, 2007.

[4] J. H. Choi, W. Song, T. Kim, S.-R. Lee, and H. C. Kim, "Finger vein extraction using gradient normalization and principal curvature," *Proc.SPIE*, vol. 7251, p. 9, 2009. [Online]. Available: https://doi.org/10.1117/12.810458

[5] B. Huang, Y. Dai, R. Li, D. Tang, and W. Li, "Finger-vein authentication based on wide line detector and pattern normalization," in *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE, 2010, pp. 1269–1272.

[6] Y. Matsuda, N. Miura, A. Nagasaka, H. Kiyomiu, and T. Miyatake, "Finger-vein authentication based on deformation-tolerant feature-point matching," *Machine Vision and Applications*, vol. 27, no. 2, pp. 237–250, 2016.

[7] 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*, pp. 1–1, 2017.

[8] 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. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231214007735

[9] 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. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/ima.20193

[10] L. Feng, W. Chao, and P. Jialiang, "Finger vein recognition using log gabor filter and local derivative pattern," *International Journal of Signal Processing, Image Processing and Pattern Recognition (IJSIP)*, vol. 9, no. 12, pp. 231–242, 2016.

[11] 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, p. 15048–15067, Nov 2013. [Online]. Available: http://dx.doi.org/10.3390/s131115048

[12] H. G. Hong, M. B. Lee, and K. R. Park, "Convolutional neural network-based finger-vein recognition using nir image sensors," *Sensors*, vol. 17, no. 6, p. 1297, 2017.

[13] C. Xie and A. Kumar, "Finger vein identification using convolutional neural network and supervised discrete hashing," *Pattern Recognition Letters*, vol. 119, pp. 148 – 156, 2019, deep Learning for Pattern Recognition.

[14] W. Kang, H. Liu, W. Luo, and F. Deng, "Study of a full-view 3D finger vein verification technique," *IEEE Transactions on Information Forensics and Security*, pp. 1–1, 2019.

[15] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 815–823.

[16] X. Wu, R. He, and Z. Sun, "A lightened CNN for deep face representation," *CoRR*, vol. abs/1511.02683, 2015. [Online]. Available: http://arxiv.org/abs/1511.02683

[17] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and $<$1mb model size," *CoRR*, vol. abs/1602.07360, 2016. [Online]. Available: http://arxiv.org/abs/1602.07360

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385

[19] A. Kumar and K. Wang, "Identifying humans by matching their left palmprint with right palmprint images using convolutional neural network," in *First International Workshop on Deep Learning and Pattern Recognition*, 2016.

[20] E. Piciucco, R. S. Kuzu, E. Maiorana, and P. Campisi, "On the cross-finger similarity of vein patterns," in *New Trends in Image Analysis and Processing – ICIAP 2019*. Springer International Publishing, 2019, pp. 12–20.

[21] R. Das, E. Piciucco, E. Maiorana, and P. Campisi, "Convolutional neural network for finger-vein-based biometric identification," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 2, pp. 360–373, Feb 2019.

[22] S. A. Radzi, M. Khalil-Hani, and R. Bakhteri, "Finger-vein biometric identification using convolutional neural network," *Turkish Journal of Electrical Engineering and Computer Science*, vol. 24, pp. 1863 – 1878, 2016.

[23] H. G. Hong, M. B. Lee, and K. R. Park, "Convolutional neural network-based finger-vein recognition using nir image sensors," *Sensors*, vol. 17, no. 6, 2017. [Online]. Available: https://www.mdpi.com/1424-8220/17/6/1297

[24] Houjun Huang, S. Liu, H. Zheng, L. Ni, Yi Zhang, and W. Li, "Deepvein: Novel finger vein verification methods based on deep convolutional neural networks," in *2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, Feb 2017, pp. 1–8.

[25] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," in *NIPS-W*, 2017.

[26] Y. Yin, L. Liu, and X. Sun, "Sdumla-hmt: a multimodal biometric database," *Biometric Recognition*, pp. 260–268, 2011.

[27] B. Ton and R. Veldhuis, "A high quality finger vascular pattern dataset collected using a custom designed capturing device," in *International Conference on Biometrics, ICB 2013*. IEEE, 2013. [Online]. Available: http://doc.utwente.nl/87790/

[28] C. Kauba, B. Prommegger, and A. Uhl, "Focussing the beam - a new laser illumination based data set providing insights to finger-vein recognition," in *Proceedings of the IEEE 9th International Conference on Biometrics: Theory, Applications, and Systems (BTAS2018)*, Los Angeles, California, USA, 2018, pp. 1–9, accepted.

[29] A. F. Sequeira, J. Ferryman, L. Chen, C. Galdi, J.-L. Dugelay, V. Chiesa, A. Uhl, B. Prommegger, C. Kauba, S. Kirchgasser, A. Grudzien, M. Kowalski, L. Szklarski, P. Maik, and P. Gmitrowicz, "Protect multimodal db: a multimodal biometrics dataset envisaging border control," in *Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG'18)*, Darmstadt, Germany, 2018, pp. 1–8. [Online]. Available: https://doi.org/10.23919/BIOSIG.2018.8552926

[30] C. Kauba, B. Prommegger, and A. Uhl, "Openvein - an open-source modular multipurpose finger vein scanner design," in *Handbook of Vascular Biometrics*, A. Uhl, C. Busch, S. Marcel, and R. Veldhuis, Eds. Cham, Switzerland: Springer Nature Switzerland AG, 2019, ch. 3, pp. 77–111.

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

[32] Y. Lu, S. J. Xie, S. Yoon, J. Yang, and D. S. Park, "Robust finger vein roi localization based on flexible segmentation," *Sensors*, vol. 13, no. 11, pp. 14 339–14 366, 2013.

[33] B. Prommegger, C. Kauba, and A. Uhl, "On the extent of longitudinal finger rotation in publicly available finger vein data sets," in *Proceedings of the 12th IAPR/IEEE International Conference on Biometrics (ICB'19)*, Crete, Greece, 2019, pp. 1–8.

[34] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51 – 59, 1996.

[35] C. Kauba, J. Reissig, and A. Uhl, "Pre-processing cascades and fusion in finger vein recognition," in *International Conference of the Biometrics Special Interest Group (BIOSIG)*, Darmstadt, Germany, sep 2014.