A. Chowdhury, S. Kirchgasser, A. Uhl, A. Ross,
"Can a CNN Automatically Learn the Significance of Minutiae Points for Fingerprint Matching?",
Proc. of IEEE Winter Conference on Applications of Computer Vision (WACV), (Aspen, USA), March 2020

Can a CNN Automatically Learn the Significance of Minutiae Points for Fingerprint Matching?

Anurag Chowdhury*, Simon Kirchgasser+, Andreas Uhl+, Arun Ross*
*Michigan State University, USA
+University of Salzburg, Austria
{chowdh51, rossarun}@cse.msu.edu, {skirch, uhl}@cs.sbg.ac.at

Abstract

Most automated fingerprint recognition systems use minutiae points for comparing fingerprints. In the parlance of Computer Vision, minutiae can be viewed as handcrafted features, i.e., features that have been proposed by human experts for the task of fingerprint recognition. In this work, we raise the following question: Can a machine learning system automatically determine the significance of minutiae points for fingerprint matching? To this effect, a patch-based Siamese Convolutional Neural Network (CNN), which does not explicitly rely on the extraction of minutiae points, is designed and trained from scratch. The purpose of this network is to learn the most effective features for matching fingerprint images. The features learned by this network are analyzed using Gradient-weighted Class Activation Mapping (Grad-CAM) to determine if they correlate with the locations of minutiae points. Our experiments suggest that the proposed network automatically learns to focus on minutiae points, when available, for fingerprint matching. Thus, an automated learner without any explicit domain knowledge establishes the significance of minutiae points for fingerprint matching.

1. Introduction

A fingerprint (FP) consists of ridges and valleys that form an oriented texture pattern on the tip of each finger. Fingerprint matching refers to the process of comparing two fingerprint patterns to determine if they have been acquired from the same finger or not [18]. This comparison is often accomplished using minutiae, i.e., ridge anomalies corresponding to ridge endings and ridge bifurcations. The spatial configuration of minutiae points has been assumed to be unique to every finger [21]; a hypothesis that has been borne out by large-scale experiments conducted on operational fingerprint data [33]. Several sophisticated minutiae descriptors have also been developed to improve the performance of minutiae-based fingerprint recognition [5, 7]. At the same time, non-minutiae-based features such as skin texture and ridge maps have also been used for fingerprint recognition [2, 4, 23, 32].

In this work, we raise the following question. Will a representation learning scheme automatically discover the significance of minutiae points for fingerprint matching? That is, will an automated feature extractor deduce the significance of minutiae points for matching even if no domain-specific knowledge is provided to it? In this paper, we answer this using a Convolutional Neural Network's (CNN's) capability for automated feature learning.

The advent of CNNs has resulted in the discovery of powerful features for performing a broad range of computer vision tasks [14]. Since automatic fingerprint recognition has been typically performed using minutiae features, CNNs developed for fingerprint matching have inevitably employed these features. In [12], a CNN-based fingerprint minutiae extraction approach was proposed that does not require any pre-processing before feature extraction. The authors use a dataset consisting of overlapping patches corresponding to 200 labelled fingerprint images. The patches were first processed by JudgeNet to detect those patches that contain exactly one minutia point. Then, the center area of each patch (27×27 pixels) containing a minutia point was processed by LocateNet to detect the exact position of the minutia point. A precision value of 94.59%, a recall value of 91.63% and an F1-score of 93.09% were reported. Another CNN, FingerNet [30], integrated some common processing methods (e.g. orientation estimation, segmentation), resulting in better matching accuracy compared to the state-of-the-art minutiae-based techniques on the NIST SD27 (latent) and the FVC 2004 (dab) datasets. The same datasets were used in [19] where the authors proposed a CNN named MinutiaeNet, consisting again of two different sub-networks. The first sub-network, CoarseNet, generated a minutiae score map by taking advantage of fingerprint domain knowledge (image enhancement, orientation field estimation, and segmentation map extraction), while the second sub-network, FineNet, determined the exact minutiae loca-

A. Chowdhury, S. Kirchgasser, A. Uhl, A. Ross,
"Can a CNN Automatically Learn the Significance of Minutiae Points for Fingerprint Matching?",
Proc. of IEEE Winter Conference on Applications of Computer Vision (WACV), (Aspen, USA), March 2020

tions using the previously generated minutiae score map. This approach was also observed to outperform state-of-the-art algorithms in terms of precision and recall.

Our goal is significantly different from the existing CNN-based studies for fingerprint recognition, viz., [12, 19, 30]. We are not trying to develop a new fingerprint recognition system. Instead we attempt to use a CNN's capability of *learning* the most effective features for a specific task based on the training samples, to answer the following question: *Can a CNN Automatically Learn the Significance of Minutiae Points for Fingerprint Matching?* Our goal is to harness the automatic feature learning capability of CNNs to determine if CNNs agree with human intuition about the significance of minutiae points for fingerprint matching.

In this regard, we design a CNN to learn local fingerprint features that are rotation and scale invariant. This is done by adopting a patch-based approach, where a pair of fingerprint patches is input to the CNN. No assumptions are made about the presence or absence of minutiae points in these patches. We allow the CNN to automatically learn salient and discriminatory features from these patches and then conduct a post-analysis to investigate the nature of the automatically learned features.

To conduct this research, we design our network to operate on local image *patch-pairs* and not on single patches. The corresponding idea will be introduced in Section 3 in detail. The rest of this paper is organized as follows: Section 2 provides a discussion of related work in other biometric domains. The description of the CNN is given in Section 3. Section 4 introduces the experimental protocol and details about the datasets and baseline fingerprint matchers employed in this study. A detailed discussion of the experimental results is provided in Section 5. The conclusions of this work and future work are discussed in Section 6.

2. Related Work

Over the past decade, various computer vision and pattern recognition tasks have been addressed effectively by harnessing the automated feature learning and feature classification prowess of CNNs [11]. In the context of biometrics, CNNs have proven to be valuable in multiple domains [22] including face, fingerprints, iris, voice, gait and vascular biometrics [28]. For the face modality, general-purpose CNNs have been successfully appropriated for matching purposes [6, 25, 31]. For the voice modality, CNNs have been used to extract features that are different from the classical MFCC, LPC or iVector representations [15]. For the iris modality, off-the-shelf CNNs have been used to demonstrate their efficacy in extracting useful information for iris recognition [20]. A detailed review of the use of CNNs in biometrics can be found in [28].

In the case of fingerprint recognition, CNNs have been used for image pre-processing [8] as well as minutiae extraction [3, 9, 10]. Furthermore, as stated in the previous section, deep network architectures have been used to perform recognition directly on the raw fingerprint data as in [12, 19, 30]. However, there is no existing work in the literature that explores whether CNNs can *automatically learn* the significance of minutiae points in fingerprint matching. This paper seeks to determine if human intuition regarding the significance of minutiae points (as validated in many large-scale tests [33]) can be substantiated by a self-learning CNN.

3. Proposed Approach

To approach the stated goal, we develop a patch-based approach to fingerprint matching that relies on a CNN to deduce the features. If the features deduced by the CNN are correlated with minutiae points, it would reinforce the importance of minutiae points in fingerprint matching from an automated representation learning scheme.

We perform fingerprint matching based on tessellated local fingerprint patches that are independent of minutiae points, i.e., minutiae points are not used to determine patches of interest. To facilitate this, we design our own Multi-scale Dilated Siamese CNN (MD-CNN) architecture that is shown in Figure 1.

The proposed network design is based on Siamese network architecture where there are two identical CNNs with shared weights. The proposed network is trained on pairs of fingerprint patches called *fingerprint patch-pairs* (D_v). These patches are extracted from pairs of fingerprint images, explained in further detail in Section 3.2. For training the network, the data pair D_v has to be provided as input to the CNN. Here, $D_v = (S_1, S_2)$, where S_1 and S_2 are two fingerprint patch samples taken from fingers 'X' and 'Y', respectively. The *fingerprint patch-pair* (D_v) is deemed to be a genuine pair if and only if $X = Y$ and both the patches are taken from the same location (see Section 3.1). The loss function, described in Section 3.4, is used to help the proposed network learn the similarity between genuine *fingerprint patch-pairs* and the dissimilarity between impostor *fingerprint patch-pairs*. In the following sections, we will discuss our fingerprint tessellation process, network design choices and the rationale behind.

3.1. Patch Alignment

For training the proposed network architecture, we first arrange fingerprint images into genuine and impostor pairs. A genuine pair is defined as a pair of fingerprint images taken from the same finger of the same subject, while an impostor pair is defined as a pair of images taken from different fingers of the same subject, the same finger of different subjects, or different fingers of different subjects. The fingerprint image-pairs (D_t) are then aligned to each other using scale-invariant feature transform (SIFT) key points [16].

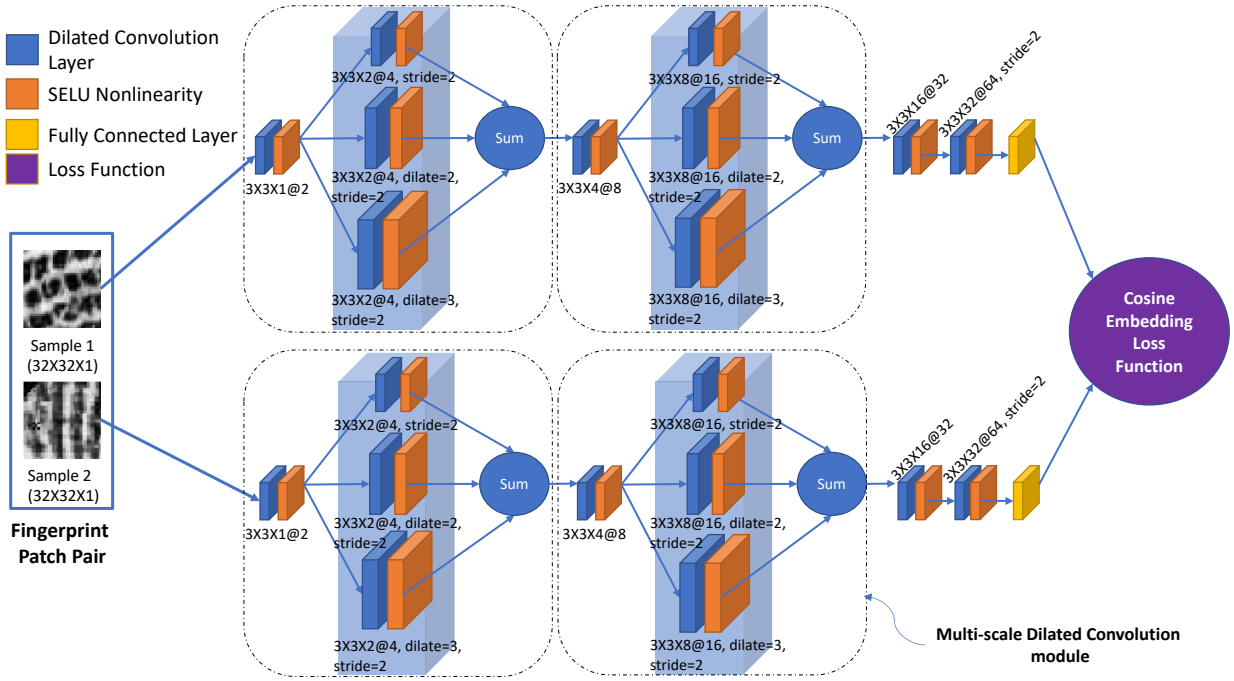


Figure 1: An illustration of the proposed Multi-scale Dilated Siamese CNN (MD-CNN) architecture during training.

Only the overlapping regions of the aligned image pairs are then retained for generating the *fingerprint patch-pairs*. It is important to note that a SIFT-based patch alignment method may include minutiae points as keypoints. Therefore, to account for the effect of minutiae points on fingerprint patch alignment, we implemented two different patch-alignment strategies. The first strategy, referred to as *SIFT with minutiae*, uses all the detected SIFT keypoints for performing fingerprint patch alignment, while the second strategy, referred to as *SIFT without minutiae*, discards all SIFT keypoints that are detected in a 10×10 pixel neighborhood of a minutiae point. Therefore, no minutiae points are involved in the second alignment strategy.

Table 1: The number of fingerprint patch-pairs containing (#present) or not containing (#absent) minutiae points across both the alignment strategies. The datasets correspond to subsets of the CASIA fingerprint Subject Ageing Version 1.0 dataset (finger1267, T2, uru4000_1, uru4000_2, uru4500_1 and uru4500_2) and the FVC-2000 dataset (db_1, db_2, db_3 and db_4).

Datasets	SIFT with minutiae			SIFT without minutiae		
	#all	#present	#absent	#all	#present	#absent
finger1267	13376	2398	10978	13376	3591	9785
T2	15232	1415	13817	15072	3535	11537
uru4000_1	16672	2164	14508	16224	2116	14108
uru4000_2	14240	2071	12169	13728	1941	11787
uru4500_1	16192	1919	14273	15520	1872	13648
uru4500_2	15200	2292	12908	14624	2261	12363
db_1	6656	301	6355	6208	476	5732
db_2	6656	1411	5245	6240	1303	4937
db_3	7392	461	6931	7072	436	6636
db_4	5216	1145	4071	5152	1117	4035

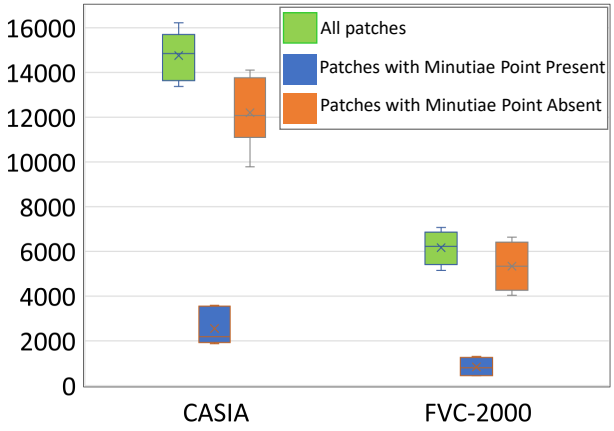


Figure 2: Aggregate number of fingerprint patch-pairs (with and without minutiae points) in that CASIA fingerprint Subject Ageing Version 1.0 and FVC-2000 datasets used in this work.

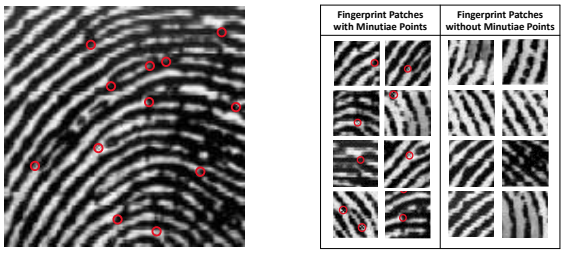
The number of extracted patch-pairs, reported in Table 1 and Figure 2, varies vastly across the different datasets (given in Section 4.1) and the applied alignment strategies. The observable differences can be explained as follows:

1. The two alignment strategies generate a similar number of fingerprint patches. This is because only a small proportion of the SIFT key points satisfy the elimination criteria in the second alignment strategy.

2. Furthermore, it is observed that dataset dependent variations (such as a low number of minutiae points, poor image quality, etc.) have a high impact on the number of success-

A. Chowdhury, S. Kirchgasser, A. Uhl, A. Ross,
"Can a CNN Automatically Learn the Significance of Minutiae Points for Fingerprint Matching?",
Proc. of IEEE Winter Conference on Applications of Computer Vision (WACV), (Aspen, USA), March 2020

fully extracted patch-pairs. This is demonstrated by the significantly low number of fingerprint patch-pairs extracted from the subsets of the "FVC-2000 dataset" compared to the "CASIA fingerprint Subject Ageing Version 1.0" dataset.



(a) fingerprint ROI (b) fingerprint patches

Figure 3: (a) A fingerprint region of interest (ROI) with minutiae points located and plotted (in red circles). (b) Fingerprint patches containing minutiae points (left column) and without any minutiae points (right column).

For this work, we limit the size of the overlapping regions of the aligned fingerprint pairs to 128×128 pixels (see Figure 3). These dimensions represent the largest region shared by all fingerprint pairs across all the datasets. The region of interest (ROI) is extracted by masking [13] the background and cropping the desired region in the foreground.

3.2. Fingerprint Tessellation

The cropped fingerprint images of dimension 128×128 are tessellated into non-overlapping fingerprint patches of dimension 32×32 pixels. The fingerprint patches are then arranged into fingerprint patch-pairs as shown in Figure 4. Fingerprint patches that are taken from the same corresponding location of the aligned genuine fingerprint pairs constitute the set of genuine fingerprint patch-pairs. Similarly, fingerprint patches taken from different locations of the aligned impostor fingerprint pairs constitute the set of impostor fingerprint patch-pairs.

3.3. Multi-scale Dilated Convolution (MD-CNN)

The design of our MD-CNN architecture is primarily focused on efficiently extracting rotation and scale-invariant fingerprint image features. Similar to the idea of inception modules introduced in GoogLeNet [29], we develop a *multi-scale dilated convolution* (MDC) module for extracting multi-scale fingerprint features in our MD-CNN architecture. The inception modules in GoogLeNet [29] learn multiple filters of different sizes, such as 3×3, 5×5, etc. and then use 1×1 convolutions for reducing the feature dimensionality, thereby allowing deeper networks to efficiently learn highly non-linear image features. The multi-scale dilated convolution module, on the other hand, learns

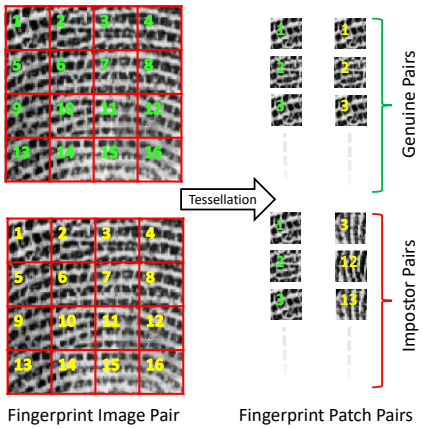


Figure 4: A visual representation of the fingerprint tessellation process for generating *fingerprint patch-pairs* D_v from a fingerprint image-pair D_t

3×3 convolution filters under different dilation levels, such as 2 and 3 in the proposed network. As shown in Figure 5, the dilated 3×3 filters effectively learn sparse convolution filters of size 5×5 and 7×7. Also, since the overlap between the effective receptive fields of the multiple dilated convolution filters is minimal, we choose to fuse their responses by summing them up, thereby obtaining a multi-scale dilated convolution response. The MDC module, therefore, helps in efficiently learning highly non-linear image features at multiple scales while keeping the size of the parameter space tractable. Such a module can be very effective when the training dataset is of limited size. For introducing rotation invariance into our model, we separately rotate the fingerprint patches in the patch-pair by a random value before inputting them to the proposed network.

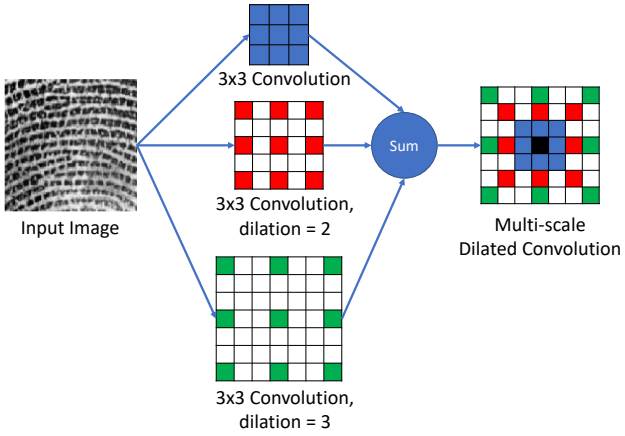


Figure 5: A visual representation of the proposed Multi-scale Dilated Convolution module.

A. Chowdhury, S. Kirchgasser, A. Uhl, A. Ross,
"Can a CNN Automatically Learn the Significance of Minutiae Points for Fingerprint Matching?",
Proc. of IEEE Winter Conference on Applications of Computer Vision (WACV), (Aspen, USA), March 2020

3.4. Cosine Embedding Loss Function

As described previously, we designed our network architecture to be trained on data pairs. Such an architecture aims to learn a feature embedding in which genuine pairs are embedded closer to each other than impostor pairs. To achieve such an embedding, we use the cosine embedding loss function for training our network:

$$loss(S_1, S_2, L) = \begin{cases} 1 - cos(S_1, S_2), & \text{if } L = 1 \\ max(0, cos(S_1, S_2) \\ \qquad -margin), & \text{if } L = -1 \end{cases} \quad (1)$$

Here, S_1 and S_2 are two fingerprint patch samples taken from fingers 'X' and 'Y', respectively. $margin$ is the minimum distance between genuine and impostor sample pairs and is a user-tunable hyper-parameter. L is the ground truth label. L is set to 1, if $X = Y$, and -1, otherwise.

3.5. Modified All-CNN architecture

We also modified the architecture (see Table 2) of the All-CNN model [27], referred to as modified All-CNN in this work, to serve as a baseline architecture for CNN-based fingerprint matching. The All-CNN architecture is chosen due to its compatibility with 32×32 input images and low computational complexity. However, we modified the original All-CNN architecture by reducing the maximum number of convolution channels from 192 to 64 across all the layers, thus, lowering the number of trainable parameters. We also replaced the RELU non-linearity with SELU non-linearity. These modifications helped the training process converge faster without over-fitting on the *limited* training data.

Table 2: Model description of the modified All-CNN architecture, based on the architecture of the All-CNN model [27], used for the fingerprint matching experiments discussed in Section 4.

Modified All-CNN Architecture	
Input Data	Input 32×32 fingerprint patch
Layer #1	3×3 conv. 2 SELU
Layer #2	3×3 conv. 4 SELU
Layer #3	3×3 conv. 8 SELU with stride r = 2
Layer #4	3×3 conv. 16 SELU
Layer #5	3×3 conv. 32 SELU with stride r = 2
Layer #6	3×3 conv. 64 SELU
Layer #7	1×1 conv. 64 SELU
Layer #8	Adaptive Avg. Pooling
Output Embedding	Output 1×64 embedding

4. Experiments

4.1. Datasets and Baseline Matcher

We have used the "CASIA fingerprint Subject Ageing Version 1.0" [1] and the "FVC-2000" [17] fingerprint datasets for experimentally evaluating both the proposed MD-CNN and the modified All-CNN architectures.

1. The "CASIA fingerprint Subject Ageing Version 1.0" dataset was assembled at the Center for Biometrics and Security Research (CBSR) at the Chinese Academy of Sciences Institute of Automation (CASIA) and is publicly available.[1] We used all 6 subsets (finger1267, T2, uru4000_1, uru4000_2, uru4500_1 and uru4500_2) of this dataset in our experiments (see Table 1). Each subset consists of images from 980 fingers corresponding to the forefinger and the middle finger of both hands of 49 subjects. For each finger, 5 impressions were captured.

2. We also used four subsets of the "FVC-2000 dataset" [17], referred to as db_1, db_2, db_3 and, db_4 (see Table 1), to evaluate the proposed algorithm. Each subset contains a total of 880 fingerprint images collected from 110 subjects.

3. A state-of-the-art minutiae recognition system (*VeriFinger SDK 11.1* [2]) was used to extract the minutiae positions from all the fingerprint images. The second alignment strategy uses these minutiae positions for eliminating those SIFT keypoints that are located in 10×10 neighborhoods of minutiae points.

4.2. MD-CNN based Experiments

For the experiments on each subset of the CASIA fingerprint Subject Ageing Version 1.0 dataset, 40 subjects are randomly sampled to constitute the training set and the remaining 9 subjects constituted the test set. Similarly, for each subset of the FVC-2000 dataset, 81 subjects are randomly sampled to constitute the training set and the remaining 29 subjects constituted the test set. Therefore, we maintain disjoint subjects in the training and test sets in all experiments. As illustrated in Figure 1, the proposed CNN is trained using fingerprint patch-pairs extracted from the fingerprints of subjects in the training set. The trained CNN models are then evaluated using fingerprint patch-pairs extracted from the fingerprints in the test set.

4.3. Modified All-CNN based Experiments

The modified All-CNN architecture, similar to the proposed MD-CNN architecture, is trained and evaluated as a Siamese network in verification mode. All the experiments done with the proposed MD-CNN model were re-

[1] http://biometrics.idealtest.org/dbDetailForUser.do?id=15
[2] http://www.neurotechnology.com/verifinger.html

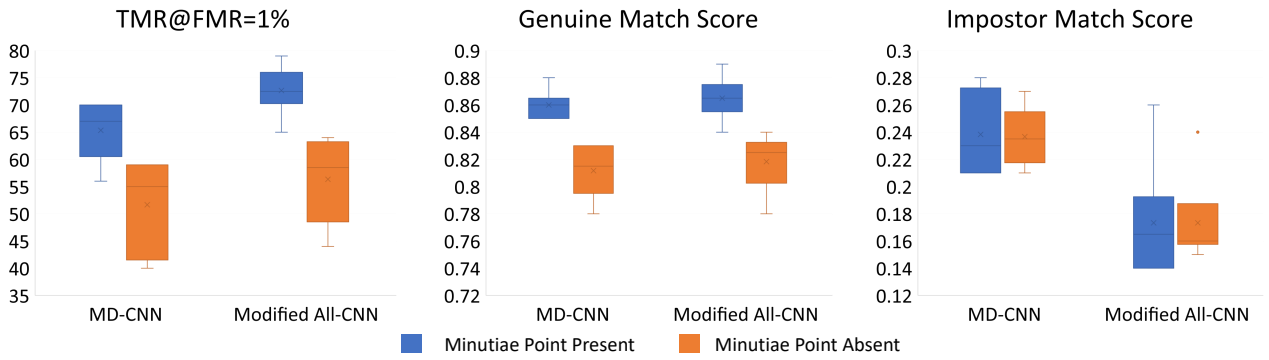


Figure 6: Aggregate performance of proposed MD-CNN and Modified All CNN models on the six subsets of the CASIA fingerprint Subject Ageing Version 1.0 dataset.

peated with the modified All-CNN model. These experiments were performed to examine whether two different CNN architectures, trained from scratch, can independently deduce the importance of minutiae points in the context of fingerprint matching.

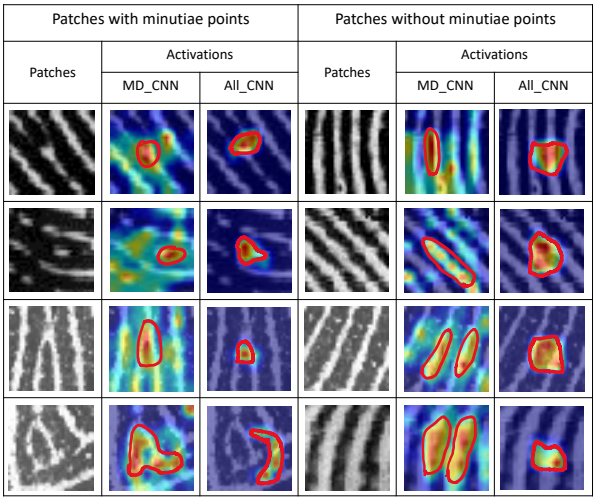


Figure 7: Example pairs of fingerprint patches and their corresponding Grad-CAM [26] visualizations showing the localization (contours manually annotated in red color) of the proposed-CNN's and modified All-CNN's attention around minutiae points (when present). Note that in case of the MD-CNN, the attention is focused along the ridges of the fingerprint when minutiae points are absent.

5. Results and Analysis

The results of all the experiments are presented in Tables 3, 4 and Figure 6. We report the fingerprint matching performance using True Match Rate at False Match Rates of 1% and 0.1% (TMR@FMR) in Table 3. We also report the average genuine and impostor match scores in Table 4.

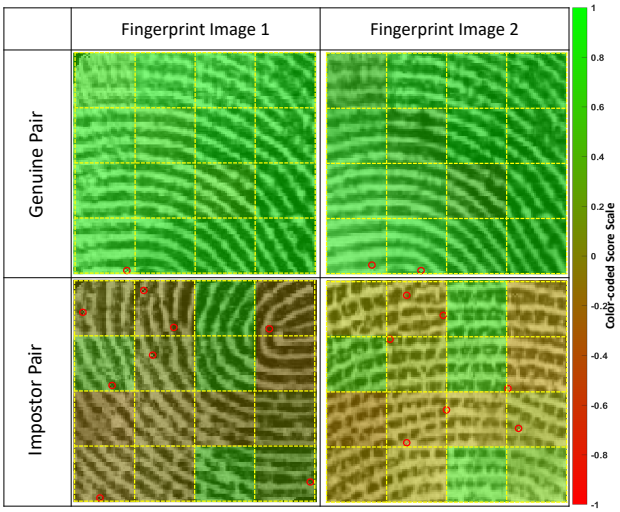


Figure 8: A visual analysis of patch based fingerprint matching with respect to presence of minutiae points in corresponding patches. Note, impostor patch-pairs containing minutiae points show lower match scores (red color) than patch-pairs without minutiae points. The color gradient between red and green correspond to match score in the range $[-1, 1]$.

The match scores are reported in the range $[-1, 1]$, where a value of 1 indicates a perfect match while a value of -1 indicates a perfect non-match. The matching performance of patch-pairs is reported for both alignment strategies described in Section 3.1. For each alignment strategy, three types pf evaluations were done: *#all*, *#present*, and *#absent*. The first one (*#all*) uses all patch-pairs irrespective of the presence or absence of minutiae points when comparing fingerprint patches. *#present* only considers patch-pairs containing at least one minutiae point in each of the patches constituting the pair. *#absent* only takes into account patch-pairs devoid of any minutiae points.

- In a majority of the 10 fingerprint data-subsets given in Tables 3(a) and 3(b), both the MD-CNN and the modified

Table 3: Results of matching fingerprint patch-pairs using (a) the proposed MD-CNN model and (b) the modified All-CNN model both evaluated at TMR@FMR=1% and TMR@FMR=0.1%. Higher verification performance (TMR@FMR=$\{1\%, 0.1\%\}$) is observed when matching fingerprint patch-pairs containing minutiae points (*#present*) as compared to matching fingerprint patch-pairs devoid of any minutiae points (*#absent*). Red colored font is used to highlight the entries where higher matching performance is observed in presence of minutiae points. All other entries are given in blue colored font.

	TMR@FMR=1%						**TMR@FMR=0.1%**					
	SIFT with minutiae			**SIFT without minutiae**			**SIFT with minutiae**			**SIFT without minutiae**		
Datasets	*#all*	*#present*	*#absent*	*#all*	*#present*	*#absent*	*#all*	*#present*	*#absent*	*#all*	*#present*	*#absent*
finger1267	0.57	0.70	0.53	0.49	0.56	0.42	0.32	0.21	0.30	0.14	0.17	0.19
T2	0.48	0.60	0.47	0.46	0.62	0.40	0.24	0.15	0.26	0.21	0.19	0.18
uru4000_1	0.57	0.67	0.55	0.56	0.67	0.55	0.34	0.53	0.32	0.36	0.53	0.34
uru4000_2	0.58	0.62	0.56	0.61	0.70	0.59	0.31	0.35	0.30	0.34	0.52	0.34
uru4500_1	0.57	0.73	0.54	0.60	0.70	0.59	0.32	0.50	0.31	0.41	0.37	0.43
uru4500_2	0.60	0.67	0.58	0.56	0.67	0.55	0.43	0.50	0.37	0.40	0.54	0.38
db_1	0.21	0.01	0.21	0.15	0.18	0.14	0.01	0.01	0.01	0.01	0.18	0.01
db_2	0.02	0.01	0.01	0.20	0.04	0.62	0.01	0.01	0.01	0.01	0.02	0.01
db_3	0.56	0.79	0.53	0.63	0.65	0.62	0.32	0.79	0.31	0.45	0.65	0.44
db_4	0.84	0.87	0.83	0.72	0.05	0.74	0.73	0.56	0.75	0.03	0.01	0.04

(a) Results using the MD-CNN model

	TMR@FMR=1%						**TMR@FMR=0.1%**					
	SIFT with minutiae			**SIFT without minutiae**			**SIFT with minutiae**			**SIFT without minutiae**		
Datasets	*#all*	*#present*	*#absent*	*#all*	*#present*	*#absent*	*#all*	*#present*	*#absent*	*#all*	*#present*	*#absent*
finger1267	0.60	0.71	0.57	0.59	0.79	0.50	0.37	0.29	0.35	0.33	0.38	0.27
T2	0.54	0.65	0.52	0.51	0.73	0.44	0.26	0.00	0.22	0.21	0.52	0.16
uru4000_1	0.57	0.68	0.55	0.58	0.72	0.56	0.34	0.55	0.33	0.35	0.49	0.31
uru4000_2	0.62	0.66	0.61	0.65	0.72	0.64	0.35	0.38	0.33	0.44	0.36	0.43
uru4500_1	0.62	0.73	0.60	0.65	0.75	0.63	0.32	0.54	0.30	0.45	0.47	0.43
uru4500_2	0.59	0.71	0.57	0.62	0.65	0.61	0.42	0.59	0.40	0.45	0.57	0.42
db_1	0.83	0.00	0.82	0.85	0.97	0.83	0.49	0.00	0.48	0.51	0.00	0.50
db_2	0.49	0.62	0.46	0.76	0.83	0.73	0.14	0.00	0.13	0.61	0.00	0.59
db_3	0.54	0.80	0.53	0.59	0.07	0.57	0.14	0.00	0.14	0.26	0.00	0.25
db_4	0.83	0.86	0.82	0.87	0.91	0.85	0.74	0.00	0.63	0.77	0.00	0.66

(b) Results using the modified All-CNN model

All-CNN models match fingerprint pairs containing minutiae points (*#present*) at a higher TMR across both the alignment strategies, compared to fingerprint patch-pairs devoid of any minutiae points (*#absent*).

- In a majority of the 10 fingerprint data subsets given in Table 4, both the MD-CNN and the modified All-CNN models match fingerprint pairs containing minutiae points (*#present*) with a higher genuine match score (*gen*) and a lower impostor match score (*imp*) across both the alignment strategies, compared to fingerprint patch-pairs devoid of any minutiae points (*#absent*).

- In a majority of the 10 fingerprint data-subsets given in Tables 4(a), 4(b), 3(a), and 3(b) the *SIFT without minutiae* alignment strategy appears to have only a marginal detrimental effect on the overall matching performance when compared against the *SIFT with minutiae* alignment strategy. This marginal reduction in overall matching performance is explained by the reduction in the overall number of patch-pairs in the *SIFT without minutiae* alignment strategy, as given in Table 1. However, it should be noted, that irrespective of the alignment strategy used, both the methods perform better on patch-pairs containing minutiae points (*#present*). This shows that the type of fingerprint features learned by both the MD-CNN and the modified All-CNN models are independent of the fingerprint alignment strategy used to generate the fingerprint image-pairs.

- The higher matching performance on the subsets of CASIA fingerprint Subject Ageing Version 1.0 dataset compared to that of the FVC-2000 subsets can be attributed to the larger number of training samples available for the former due to the larger number of extracted fingerprint patch-pairs, as reported in Table 1.

We also performed Grad-CAM [26] based qualitative analysis of the image features extracted by the CNNs for performing fingerprint matching at the patch level. Grad-CAM uses gradient information flowing through the CNN to iden-

Table 4: Average comparison scores of fingerprint patch-pairs (both genuine and impostor pairs) from all datasets using the proposed the proposed MD-CNN model(top) and the modified All-CNN model(bottom). On an average, a higher genuine match score (*gen*) and a lower impostor match score (*imp*) is observed when matching fingerprint patch-pairs containing minutiae points (*#present*) as compared to matching fingerprint patch-pairs devoid of any minutiae points (*#absent*). The scores are in the $[-1, 1]$ range.

Datasets	**Average Comparison Score**											
	SIFT with minutiae						**SIFT without minutiae**					
	#all		**#present**		**#absent**		**#all**		**#present**		**#absent**	
	gen	*imp*	*gen*	*imp*	*gen*	*imp*	*gen*	*imp*	*gen*	*imp*	*gen*	*imp*
finger1267	0.84	0.23	0.87	0.23	0.83	0.23	0.83	0.25	0.88	0.27	0.80	0.25
T2	0.80	0.26	0.85	0.26	0.79	0.26	0.80	0.27	0.86	0.28	0.78	0.27
uru4000_1	0.82	0.22	0.85	0.21	0.82	0.21	0.83	0.23	0.85	0.24	0.83	0.23
uru4000_2	0.84	0.21	0.85	0.21	0.83	0.21	0.84	0.21	0.86	0.21	0.83	0.21
uru4500_1	0.83	0.22	0.87	0.21	0.82	0.22	0.82	0.22	0.86	0.21	0.81	0.22
uru4500_2	0.81	0.20	0.84	0.18	0.80	0.21	0.83	0.23	0.85	0.22	0.82	0.24
db_1	0.88	0.31	0.98	0.25	0.85	0.32	0.88	0.29	0.98	0.26	0.86	0.29
db_2	0.94	0.82	0.96	0.80	0.93	0.82	0.79	0.27	0.84	0.24	0.77	0.27
db_3	0.79	0.22	0.86	0.21	0.78	0.22	0.80	0.21	0.86	0.21	0.79	0.21
db_4	0.89	0.12	0.92	0.11	0.88	0.12	0.89	0.10	0.94	0.11	0.87	0.87

(a) Results using the modified MD-CNN model

Datasets	**Average Comparison Score**											
	SIFT with minutiae						**SIFT without minutiae**					
	#all		#present		#absent		#all		#present		#absent	
	gen	imp	gen	imp	gen	imp	gen	imp	gen	imp	gen	imp
finger1267	0.84	0.16	0.88	0.18	0.83	0.16	0.84	0.16	0.89	0.16	0.81	0.15
T2	0.81	0.20	0.85	0.20	0.80	0.20	0.80	0.25	0.86	0.26	0.78	0.24
uru4000_1	0.83	0.16	0.85	0.15	0.83	0.16	0.84	0.17	0.86	0.17	0.83	0.17
uru4000_2	0.85	0.19	0.86	0.19	0.85	0.19	0.84	0.16	0.87	0.17	0.84	0.16
uru4500_1	0.84	0.16	0.87	0.16	0.83	0.16	0.83	0.16	0.87	0.14	0.83	0.16
uru4500_2	0.82	0.13	0.85	0.12	0.81	0.13	0.82	0.16	0.84	0.14	0.82	0.16
db_1	0.87	0.15	0.96	0.13	0.86	0.15	0.89	0.14	0.97	0.17	0.87	0.14
db_2	0.78	0.34	0.83	0.35	0.77	0.34	0.86	0.16	0.90	0.17	0.85	0.16
db_3	0.80	0.19	0.85	0.21	0.79	0.19	0.80	0.20	0.86	0.21	0.79	0.20
db_4	0.88	0.14	0.91	0.14	0.86	0.14	0.89	0.15	0.93	0.16	0.86	0.15

(b) Results using the modified All-CNN model

tify the local regions in an input image crucial for performing the task. Grad-CAM analysis revealed that both the proposed MD-CNN and the modified All-CNN localized their attention on the vicinity of minutiae points for extracting fingerprint features, as shown in Figure 7. This indicates that both the CNNs focused automatically on minutiae points (when present), even when they were never explicitly made aware of its importance in the task. The same analysis on fingerprint patches devoid of minutiae points shows that the proposed MD-CNN network's attention to concentrate along the ridges (see Figure 7). However, such an observation is not made in case of the modified All-CNN network (see Figure 7). We believe this to be an effect of the Multi-scale Dilated Convolution modules in the proposed MD-CNN network that allows the network to focus on ridges in absence of minutiae points for performing fingerprint matching. Furthermore, the Grad-CAM results are obtained on fingerprint images in a cross-dataset and cross-sensor setup, thereby minimizing any dataset or preprocessing based bias in the overall results. Additionally, the observations from the Grad-CAM analysis are substantiated by a quantitative analysis of the patch-based fingerprint matching process, as illustrated in Figure 8. A genuine and an impostor pair of an exemplary set of fingerprint images are tessellated and organized into fingerprint patch-pairs, as previously illustrated in Figure 4. Fingerprint patch-pairs are then compared using the MD-CNN. The resultant match scores in the range [-1,1] are indicated using a color-coded overlay in Figure 8. All the minutiae points are also annotated using red circles at their corresponding locations. As seen in Figure 8, all the genuine patch-pairs (top row) are overlaid with shades of green color, indicating a positive match result with match scores mostly in the range $(0, 1]$. The impostor patch-pairs (bottom row), on the other hand,

are observed to be overlaid with shades of red, yellow, and green, representing corresponding match scores spanning the entire range $[-1, 1]$. This observation, combined with the location of minutiae points, leads to the following inferences:

- The presence of minutiae points in an impostor patch pair yields a lower match score in the range $[-1, 0]$, as indicated by the red color.
- The absence of minutiae points in an impostor patch pair yields a higher match score in the range $(0, 1]$, as indicated by the green color.
- This indicates the positive effect of the the presence of minutiae points on fingerprint patch pair comparison.

6. Conclusion and Future Goals

The purpose of this work was to determine whether a representation learning scheme would automatically deduce the significance of minutiae points for fingerprint matching. In this regard, we designed a Multi-scale Dilated Siamese CNN architecture capable of extracting scale and rotation invariant image features for comparing fingerprint patches. The proposed CNN was trained to perform fingerprint matching without explicitly being made aware of the concept of minutiae points. The experimental results, augmented with extended qualitative and quantitative analysis, show that the proposed CNN automatically learns and extracts fingerprint features from the vicinity of minutiae points (when available) for performing fingerprint matching. We performed the experiments using two different alignment strategies to verify the effect of alignment, with and without minutiae, on the type of fingerprint features extracted by the proposed CNN. We also repeated the experiments using a modified All-CNN model and confirmed the significance of the minutiae points for fingerprint matching regardless of the choice of CNN architecture used.

We plan to extend this work by evaluating the proposed method on different types of fingerprints such as latent fingerprints, partial fingerprints, moist fingerprints, etc. This will help establish the importance of minutiae points in fingerprint matching as a function of the type of fingerprints. We are also looking at ways in which information extracted by the proposed network can be better exploited by traditional fingerprint matchers, especially those used to process partial fingerprints [24]. In summary, the results of this work suggest that automatic feature learning schemes also determine minutiae points to be of importance for fingerprint matching, therefore, establishing the significance of minutiae points from a machine learning perspective.

Acknowledgement

Anurag Chowdhury and Arun Ross were partially supported by the United States National Science Foundation under Grant 1617466.

References

[1] Casia fingerprint subject ageing version 1.0 dataset. `http://biometrics.idealtest.org/dbDetailForUser.do?id=15`. Accessed: 2020-01-14.

[2] J. Abraham, P. Kwan, and J. Gao. Fingerprint matching using a hybrid shape and orientation descriptor. In *State of the art in Biometrics*. Intechopen, 2011.

[3] R. Abrishambaf, H. Demirel, and I. Kale. A fully CNN based fingerprint recognition system. In *IEEE IWCNNA*, 2008.

[4] K. Cao and A. K. Jain. Automated latent fingerprint recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.

[5] R. Cappelli, M. Ferrara, and D. Maltoni. Minutia cylinder-code: A new representation and matching technique for fingerprint recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2128–2141, 2010.

[6] C. Ding and D. Tao. Robust face recognition via multimodal deep face representation. *IEEE Transactions on Multimedia*, 17(11):2049–2058, 2015.

[7] J. Feng. Combining minutiae descriptors for fingerprint matching. *Pattern Recognition*, 41(1):342–352, 2008.

[8] Q. Gao, P. Forster, K. R. Mobus, and G. S. Moschytz. Fingerprint recognition using CNNs: Fingerprint preprocessing. In *IEEE ISCAS*, 2001.

[9] Q. Gao and G. Moschytz. A CNN-based fingerprint verification system. In *Complex Computing-Networks*, pages 243–256. Springer, 2006.

[10] Q. Gao and G. S. Moschytz. Fingerprint feature extraction using CNNs. In *IEEE ECCTD*, 2001.

[11] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[12] L. Jiang, T. Zhao, C. Bai, A. Yong, and M. Wu. A direct fingerprint minutiae extraction approach based on convolutional neural networks. In *IEEE IJCNN*, 2016.

[13] S. Kirchgasser and A. Uhl. Fingerprint template ageing vs. template changes revisited. In *BIOSIG*, 2017.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*. 2012.

[15] C. Li, X. Ma, B. Jiang, X. Li, X. Zhang, X. Liu, Y. Cao, A. Kannan, and Z. Zhu. Deep speaker: an end-to-end neural speaker embedding system. *arXiv preprint arXiv:1705.02304*, 2017.

[16] D. G. Lowe. Object recognition from local scale-invariant features. In *IEEE CVPR*, 1999.

[17] D. Maio et al. FVC2000: Fingerprint verification competition. *IEEE TPAMI*, 2002.

[18] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar. *Handbook of fingerprint recognition*. Springer Science & Business Media, 2009.

[19] D.-L. Nguyen, K. Cao, and A. K. Jain. Robust minutiae extractor: Integrating deep networks and fingerprint domain knowledge. In *IEEE ICB*, 2018.

[20] K. Nguyen, C. Fookes, A. Ross, and S. Sridharan. Iris recognition with off-the-shelf CNN features: A deep learning perspective. *IEEE Access*, 6:18848–18855, 2018.

[21] S. Pankanti, S. Prabhakar, and A. K. Jain. On the individuality of fingerprints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1010–1025, 2002.

[22] A. Ross, S. Banerjee, C. Chen, A. Chowdhury, V. Mirjalili, R. Sharma, T. Swearingen, and S. Yadav. Some research problems in biometrics: The future beckons. *arXiv preprint arXiv:1905.04717*, 2019.

[23] A. Ross, A. Jain, and J. Reisman. A hybrid fingerprint matcher. *Pattern Recognition*, 36(7):1661–1673, 2003.

[24] A. Roy, N. Memon, and A. Ross. Masterprint: Exploring the vulnerability of partial fingerprint-based authentication systems. *IEEE Transactions on Information Forensics and Security*, 12(9):2013 – 2025, 2017.

[25] F. Schroff, D. Kalenichenko, and J. Philbin. FaceNet: A unified embedding for face recognition and clustering. In *IEEE CVPR*, 2015.

[26] R. Selvaraju et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *IEEE ICCV*, 2017.

[27] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller. Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806*, 2014.

[28] K. Sundararajan and D. L. Woodard. Deep learning for biometrics: A survey. *ACM Comput. Surv.*, 51(3):65:1–65:34, May 2018.

[29] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *IEEE CVPR*, 2015.

[30] Y. Tang, F. Gao, J. Feng, and Y. Liu. Fingernet: An unified deep network for fingerprint minutiae extraction. In *IEEE IJCB*, 2017.

[31] Y. Wen, K. Zhang, Z. Li, and Y. Qiao. A discriminative feature learning approach for deep face recognition. In *ECCV*. Springer, 2016.

[32] J. Yang. Non-minutiae based fingerprint descriptor. *Biometrics*, page 79, 2011.

[33] S. Yoon and A. K. Jain. Longitudinal study of fingerprint recognition. *Proceedings of the National Academy of Sciences*, 112(28):8555–8560, 2015.