Universidade Federal da Bahia
Instituto de Matemática e Estatística

Colegiado de Ciência da Computação

CNN-BASED PORE DETECTION AND DESCRIPTION FOR HIGH-RESOLUTION FINGERPRINT RECOGNITION

Gabriel Dahia Fernandes

TRABALHO DE CONCLUSÃO DE CURSO

Salvador
13 de janeiro de 2014

GABRIEL DAHIA FERNANDES

CNN-BASED PORE DETECTION AND DESCRIPTION FOR HIGH-RESOLUTION FINGERPRINT RECOGNITION

Este Trabalho de Conclusão de Curso foi apresentado ao Colegiado de Ciência da Computação da Universidade Federal da Bahia, como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Maurício Pamplona Segundo

Salvador
13 de janeiro de 2014

Sistema de Bibliotecas - UFBA

Dahia, Gabriel.

CNN-based pore detection and description for high-resolution fingerprint recognition / Gabriel Dahia Fernandes – Salvador, 2018.

33p.: il.

Orientador: Prof. Dr. Maurício Pamplona Segundo.

Trabalho de Graduação – Universidade Federal da Bahia, Instituto de Matemática e Estatística, 2018.

1. Fingerprint recognition. 2. Biometrics. 3. Computer vision. I. Pamplona Segundo, Maurício. II. Universidade Federal da Bahia. Instituto de Matemática e Estatística. III Título.

CDD – XXX.XX

CDU – XXX.XX.XXX

TERMO DE APROVAÇÃO

GABRIEL DAHIA FERNANDES

CNN-BASED PORE DETECTION AND DESCRIPTION FOR HIGH-RESOLUTION FINGERPRINT RECOGNITION

Este Trabalho de Conclusão de Curso foi julgado adequado à obtenção do título de Bacharel em Ciência da Computação e aprovado em sua forma final pelo Colegiado de Ciência da Computação da Universidade Federal da Bahia.

Salvador, 14 de Dezembro de 2018

Maurício Pamplona Segundo, Ph.D.
Universidade Federal da Bahia

André Brasil Vieira Wyzykowski, M.Sc.
Universidade Federal da Bahia

Leone da Silva de Jesus, M.Sc.
Universidade Federal da Bahia

ACKNOWLEDGEMENTS

I would first like to thank the friends I know longer, by supporting me when I was unhappy at med school and when I made an unexpected, baffling change in my life's direction. These friends include my family, who lovingly ensured that I felt confident about my choices, even when they were not, and listened to overlong, tiresome and mostly incomprehensible monologues about theoretical Computer Science. They also include my high and med school colleagues, for their affection and companionship over all these years, even if not always in touch.

I would also like to thank the friends I made in these last four years. Thank you for welcoming me so nicely and making my transition easier. I am lucky to count among these friends the people who have taught me much of what made me passionate about Computer Science. They did this in every capacity - teacher, colleague, advisor, lab mate - and manner - attending classes, answering persistent and annoying questions, chatting, sending insightful links, writing papers together, doing term projects - possible.

Without you all, I would probably would have not completed this journey.

RESUMO

Reconhecimento de impressões digitais em imagens de alta resolução geralmente depende de algoritmos de *matching* sofisticados para encontrar correspondências entre descritores *hand-crafted* para *keypoints*, geralmente poros. Nesse trabalho, tentou-se melhorar o reconhecimento de impressões digitais com base em poros e em imagens de alta resolução seguindo direção contrária ao dos trabalhos mais recentes sobre detecção de poros e *matching* de impressões digitais.

Primeiro, investigou-se se CNNs propostas previamente para detecção de poros em impressões digitais sobrestimaram o número necessário de parâmetros para essa tarefa. Demonstrou-se que esse era de fato o caso ao utilizar uma rede neural totalmente convolucional com significativamente menos parâmetros. Esse modelo foi avaliado usando um protocolo rigoroso e reprodutível que, antes deste trabalho, não existia para essa tarefa. Usando o protocolo proposto, mostrou-se que o modelo apresentado, quando combinado com pós-processamento, tem desempenho superior ao dos métodos anteriores, apesar de ser muito mais eficiente.

Em seguida, utilizou-se um simples algoritmo de *matching* baseado em descritores locais robustos para poros, descritores esses aprendidos a partir de dados de treino usando uma CNN. Para treinar essa CNN de maneira completamente supervisionada, descreve-se como o alinhamento automático de impressões digitais pode ser utilizado para obter as anotações necessárias - não disponibilizadas para as bases de dados públicas - para o treino. Isso é suficiente para melhorar o estado-da-arte para reconhecimento de impressões digitais parciais e completas em um *benchmark* público. Para confirmar que a melhoria observada ocorre por causa da adoção dos descritores propostos, conduziu-se um estudo de ablação com os dois descritores de poros mais bem-sucedidos utilizados anteriormente.

Todos os códigos relativos a esse trabalho estão disponíveis em ⟨https://github.com/gdahia/fingerprint-pore-detection⟩ e ⟨https://github.com/gdahia/high-res-fingerprint-recognition⟩.

Palavras-chave: Reconhecimento de impressões digitais, Biometria, Visão computacional

ABSTRACT

High-resolution fingerprint recognition often relies on sophisticated matching algorithms based on hand-crafted keypoint descriptors, with pores being the most common keypoint choice. In this work, we attempt to improve pore-based, high-resolution fingerprint recognition by following the opposite direction of recent works in both pore detection and fingerprint matching.

First, we investigate if previously proposed CNNs for fingerprint pore detection overestimate the number of required model parameters for this task. We show that this is indeed the case by proposing a fully convolutional neural network that has significantly fewer parameters. We evaluate this model using a rigorous and reproducible protocol, which was, prior to our work, not available to the community. Using our protocol, we show that the proposed model, when combined with post-processing, performs better than previous methods, albeit being much more efficient.

Second, we use instead a simple matching algorithm based on robust local pore descriptors that are learned from the data using a CNN. In order to train this CNN in a fully supervised manner, we describe how the automatic alignment of fingerprint images can be used to obtain the required training annotations, which are otherwise missing in all publicly available datasets. This improves the state-of-the-art recognition results for both partial and full fingerprints in a public benchmark. To confirm that the observed improvement is due to the adoption of learned descriptors, we conduct an ablation study using the most successful pore descriptors previously used in the literature.

All our code is available at ⟨https://github.com/gdahia/fingerprint-pore-detection⟩ and ⟨https://github.com/gdahia/high-res-fingerprint-recognition⟩.

Keywords: Fingerprint recognition, Biometrics, Computer vision

CONTENTS

1

INTRODUCTION

The aim of this work is to improve high-resolution fingerprint recognition. However, while we do propose a complete pipeline that comprises pore detection, description and matching to improve pore-based fingerprint recognition in high-resolution images, they are separated in two main modules, detection and description/matching. The corresponding work for the two modules is detailed in individual chapters. This is because each chapter contains work that was previously described in stand-alone papers. Therefore, to make the contents of this thesis correspond as best as possible to that of the papers, we keep them divided here as well.

The contents in chapter 2 are adapted from "*Improving fingerprint pore detection with a small FCN*" (DAHIA; SEGUNDO, 2018b). This work was submitted to IAPR International Conference on Biometrics (ICB), 2019, and is, as of this writing, under review. A pre-print is available at ⟨https://arxiv.org/abs/1811.06846⟩.

On the other hand, the contents in chapter 3 are adapted from "*Automatic dataset annotation to learn cnn pore description for fingerprint recognition*" (DAHIA; SEGUNDO, 2018a). This work was submitted to IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, and, like the previous one, is under review as of the writing of this thesis. A pre-print for it can be found at ⟨https://arxiv.org/abs/1809.10229⟩.

2

IMPROVING FINGERPRINT PORE DETECTION WITH A SMALL FCN

High-resolution fingerprint images allow the analysis of level 3 features (JAIN et al., 2007), with pores being a popular choice among researches (ZHAO et al., 2009; SEGUNDO; LEMES, 2015; LIU et al., 2017). The high number of pores per fingerprint, for example, allows recognizing fingerprints even when using only partial images (SEGUNDO; LEMES, 2015; LIU et al., 2011). In addition, combining them with traditional minutiae-based methods improves fingerprint recognition (ZHAO et al., 2009; LIU et al., 2017).

All automatic pore-based fingerprint recognition methods require detecting pores first, creating a demand for accurate and efficient pore detection algorithms. Recent works have tackled pore detection in high-resolution images with convolutional neural networks (CNN). We believe this is a promising approach since results from these methods are a substantial improvement over previous ones (WANG et al., 2017; SU et al., 2017).

However, it is also our opinion that the current use of CNNs overestimates the number of model parameters required to accomplish accurate pore detection, leading to inefficient networks. What is worse, the literature does not yet have a detailed protocol for evaluating pore detection. As authors of previous works do not detail the evaluation protocols of their works enough to reproduce them and do not make their code publicly available, this makes comparing the performance of previous methods impossible.

To solve both problems, we first propose a simple fully convolutional neural network (FCN) architecture, with many fewer parameters than previous CNN methods. To validate it, we devise and detail a protocol that uses the standard benchmark for high-resolution fingerprints, the Hong Kong Polytechnic University (PolyU-HRF) dataset (ZHAO et al., 2009). This protocol has the intention of standardizing pore detection evaluation, making it reproducible and directly comparable. In order to make this goal easier to attain, we make all of our code publicly available.

To point discrepancies between the previously used validations and the proposed evaluation protocol, we also employ it to evaluate the two pore detection methods that report

having the best results (SEGUNDO; LEMES, 2015; SU et al., 2017). In it, our FCN, combined with post-processing based on thresholding and non-maximum suppression (NMS), performs on par with the state-of-the-art, even with 200 times fewer parameters.

Our contributions with this work are:

- we describe a simple FCN architecture and a post-processing step to enhance its results (section 2.2);
- we detail a reproducible evaluation protocol for pore detection using the PolyU-HRF dataset (section 2.3);
- we evaluate ours and two other methods with this new protocol (section 2.4). We find evidence that our protocol is stricter than previous ones. Also, these experiments provide strong evidence that our method is the state-of-the-art for pore detection.

2.1 RELATED WORK

Previously, methods approached pore detection with hand-crafted methods or heuristics. Pore extraction with adaptive pore modeling (ZHAO et al., 2010) convolves the input fingerprint images with a pore model, acting as a matched filter, and determines that locations with strong response are pores. The pore model is instantiated from a dynamic anisotropic pore model parametrized with local ridge orientation and pore scale. The response map is post-processed by removing pores that are not located on ridges, thresholding and converting connected components of detections into single proposals to obtain the final detections. We refer to this as the traditional post-processing approach for pore detection.

Adaptive pore modeling can be improved by considering the spatial distribution of pores (TEIXEIRA; LEITE, 2014). This approach computes the expected distance between adjacent pores as an estimate of the ridge width of non-overlapping blocks of the fingerprint image. This is input to a classifier that is used to discard false detections made by the standard approach.

Dynamic pore filtering (SEGUNDO; LEMES, 2015) consists of finding the nearest black (ridge) pixels above, below, to the right, and to the left of each white pixel. Then, it counts the number of black-white transitions in the circumference delimited by these four points to decide whether the pixel is a pore or not.

CNNs have recently been proposed as alternatives to handcrafted pore detection. Su et al. (2017) train a 7 layer CNN to classify image patches as centered on pores or not. Image patches are classified as centered on pores if they are centered within 3 pixels from a pore position in the ground truth. With the trained CNN, inference is performed by classifying the image at every possible patch location. To allow inferring for entire images, they propose to convert the last two fully connected layers into convolutional layers. The memory cost of doing this, however, is substantial, as the first fully connected layer has 4096 units.

Wang et al. (2017) use a U-Net to detect pores in fingerprint images. The U-Net is trained to classify each spatial location in the image into one of three categories: pore

centroid, ridge or background. To detect pores, 20 patches are extracted from the input image and each is forwarded through the trained CNN. This results in 3 probability maps per patch, one indicating the probability of pores, the other of ridges and the last of background region. The ridge probability map for each patch is used to post-process the one for pores. Afterward, the pore probability maps are binarized. To combine the predictions for each patch, a boolean "or" operation is performed with all of them.

2.2 FCN PORE DETECTION

To detect pores in fingerprint images, we train an FCN to determine which image patches are centered around a pore. This model is trained using image patches, instead of entire images. This is done to decorrelate examples in a batch and avoid overfitting because the datasets for this task have very few annotated images. However, inference can be done by performing a single forward pass per image in the model. We also propose using a post-processing method to enhance the FCN's outputs. Without this, our model proposes multiple detections for a single pore.

2.2.1 Model architecture

Table 2.1 shows our FCN architecture for pore detection. All four layers use batch normalization (IOFFE; SZEGEDY, 2015), valid padding, and stride equal to 1 in each dimension. In the first three layers, the kernel size is 3×3 and the activations are ReLU (JARRETT et al., 2009); each is followed by 3×3 max pooling. The number of filters is 32, 64 and 128 from first to third layer, respectively. The output layer is preceded by dropout (SRIVASTAVA et al., 2014), has kernel size of 5×5, a single filter, and sigmoid activation.

#	Layer	Size	Filters
1	Conv + ReLU	3×3	32
	BatchNorm	-	-
	Max Pooling	3×3	-
2	Conv + ReLU	3×3	64
	BatchNorm	-	-
	Max Pooling	3×3	-
3	Conv + ReLU	3×3	128
	BatchNorm	-	-
	Max Pooling	3×3	-
	Dropout	-	-
4	Conv	5×5	1
	BatchNorm	-	-
	Sigmoid	-	-

Table 2.1: Pore detection FCN architecture. Every layer uses valid padding and unit strides. The architecture is adequate for detecting pores because the required information to decide if a region is centered on a pore fits entirely in the network's receptive field.

Using an FCN allows us to perform inference for images of arbitrary size, without requiring adaptations, like networks with fully connected layers do. It also keeps the number of model parameters relatively low.

When applied to an $M \times N$ image, our network produces an $(M - 16) \times (N - 16)$ output that is a spatial map indicating the probability of each pixel being centered on a pore. If the input image has dimensions 17×17, the output is a scalar, indicating the probability of the image being centered on a pore. Thus, one can see it as a sliding window passing through overlapping image patches and outputting pore probabilities for each one of them.

Pore detection differs from general object detection, in which the sliding window approach we employ is no longer the state-of-the-art. We argue that this architecture suits the pore detection task for two reasons. First, region proposal in general object detection must consider object sizes that vary significantly since the objects to be detected can be as wide or as tall as the entire image. This is not the case for pore detection. Pores fit entirely in 17×17 patches since the area they occupy in 1200dpi images range from 5 to 40 pixels (SU et al., 2017). Second, though providing global context significantly improves general object detection (REDMON; FARHADI, 2017; REN et al., 2015), it does not has the same effect for pores. Most of the necessary context to determine if a patch is centered on a pore - the pixel intensities of ridges and valleys, the ridge orientation - is visible in the patch itself.

2.2.2 Training problem formulation

The labeling for the training of our detection FCN follows the formulation of a previous work (SU et al., 2017). The probability of an image patch I_p of dimensions 17×17 having a pore is 1 if there is a pore within a bounding box of dimensions 7×7 centered in I_p's center; the probability is 0, otherwise. Training, then, is formulated as minimizing the cross-entropy between the model's predictions and the training data pore probability distribution.

While it is possible to train the CNN end-to-end for whole images, this is suboptimal when the number of training images is very small. It is known that independent batch sampling is crucial for training neural nets (GOODFELLOW et al., 2016) and for the effectiveness of batch normalization (IOFFE, 2017).

In our model architecture, using an entire image of dimensions $M \times N$ in a mini-batch is equivalent to using $(M - 17)(N - 17)$ highly correlated image patches. The pixel intensities of the image, which can be roughly divided into two classes, ridge (darker pixels) and valley (lighter pixels), display little to no intraclass variation. Also, overlapping and neighboring patches differ from one another only by small translations and a few pixels.

We sidestep this problem by using patch sampling. Patch sampling consists of sampling patches randomly from all images of the training set to perform gradient-based learning, instead of using entire images in mini-batches. By doing this, the patches are as independent as possible. Figure 2.1 illustrates the problem and the patch sampling approach.

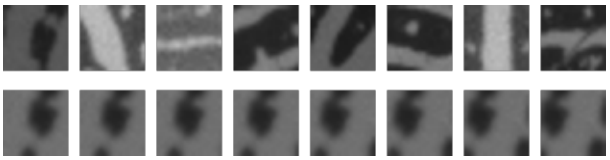


Figure 2.1: Image patch correlation and patch sampling. The top row displays patches sampled randomly across images, while the bottom row displays patches sampled sequentially from a single image, which is equivalent to use entire images in mini-batches. Notice the variations in intensities of ridges and valleys across patches in the top row, and how ridge orientations vary among them. These variations are absent from sequentially sampled patches. All patches are 17×17 and have been resized to improve visibility.

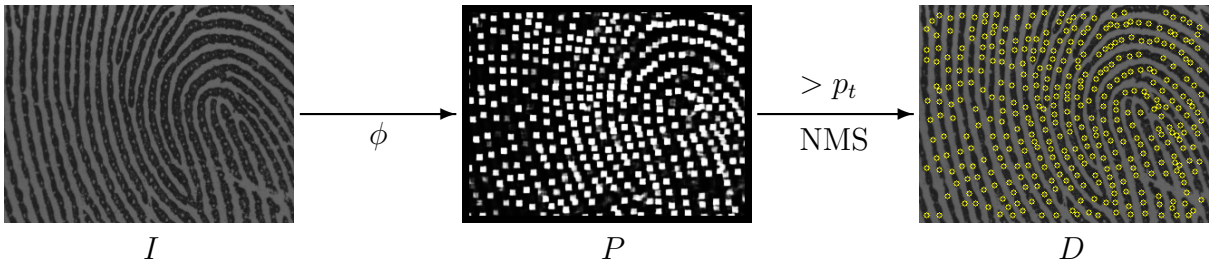


Figure 2.2: Pore detection pipeline. First, P is obtained from the input image I by zero-padding the output $\phi(I)$ of the detection CNN to input dimensions $M \times N$. Brighter pixels in P indicate higher probability. P is thresholded and the remaining points are converted to bounding boxes. We then merge multiple detections of the same pore with NMS and convert the bounding boxes back to coordinates to obtain D. Pore detections are circled in yellow. Best viewed in color.

2.2.3 Post-processing

Traditionally, models like ours propose every location in which the output indicates more than 50% probability as a pore detection. Visualizing these proposals showed us that there were multiple detections for each pore. To address that problem, we adapted traditional object-detection post-processing techniques to the pore detection task.

Figure 2.2 displays the complete pore detection pipeline when combined with post-processing. Let I be the input image of dimensions $M \times N$ and $P = \phi(I)$ be the CNN's output, zero-padded to the input dimensions. P represents the probability of each spatial location of I being centered on a pore.

We then construct the set B of 7×7 bounding boxes around each spatial coordinate (i, j) for which $P(i, j)$ is above a pre-determined threshold p_t. The construction of B is derived from how patches are labeled for training: probability 1 is assigned if a 7×7 bounding box around the center of the patch contains a pore.

The procedure outlined so far connects the bounding box detections of the same pore by the overlap in their areas. Hence, we apply NMS to convert them into a single detection. Converting the remaining bounding boxes to coordinates again results in the

final detections, D.

2.3 EXPERIMENTS

We conduct our experiments in the default benchmark for pore detection, the PolyU-HRF dataset (ZHAO et al., 2009). Polyu-HRF is divided into three subsets: *GroundTruth*, *DBI*, and *DBII*. *GroundTruth* consists of 30 1200dpi images annotated with coordinates for pores. *DBI* and *DBII* are for fingerprint recognition experiments and lack pore coordinate annotation. Therefore, we use *GroundTruth* for our experiments.

It is important to notice that there is no standard protocol to evaluate pore detection in *GroundTruth*. Because of this, authors follow their own protocols to validate their methods. Furthermore, these protocols usually do not have the required level of detail to be reproduced. Hence, we propose an evaluation protocol for pore detection in *GroundTruth*, describe it in detail and provide code to reproduce it. Our intention is that future works use our protocol, making reported results comparable.

Our evaluation protocol for pore detection in *GroundTruth* is as follows. Its 30 images should be split with the first 15 images forming the training set, the next 5 the validation set, and the last 10 the test set.

Evaluated methods should be compared by reporting true detection rates (TDR), false detection rates (FDR), and the resulting F-score. TDR is computed as the recall of detections and FDR is computed as the false discovery rate of detections. To compute these metrics, we further establish how to determine which detections are true detections and which are false detections. Given a set of detections D for an image and its corresponding ground truth G, first compute the distances of all pairs $(d, g) \in D \times G$. Consider true detections only the $d \in D$ for which the distance $||d - g||_2$ for some $g \in G$ is both the minimum of the distances of d and g, *i.e.*:

$$g = \arg\min_{g' \in G} ||d - g'||_2 \tag{2.1}$$

and

$$d = \arg\min_{d' \in D} ||d' - g||_2. \tag{2.2}$$

Every other detection is considered a false detection.

To allow a fair comparison between methods that use CNNs and the others, only pores that can be detected should be considered, *e.g.* for our method and Su et al. (2017)'s, pores in the first and last 8 rows and columns are not considered.

To validate our method, we conduct two experiments. The first experiment is an ablation study of the post-processing technique. We compare the results obtained when using the same FCN model with both post-processing as proposed and using the traditional approach (ZHAO et al., 2010). When using the FCN with traditional post-processing, a detection is proposed everywhere the output predicts more than 50% probability of a pore.

The second experiment consists of evaluating, using the proposed protocol, the two best methods in terms of originally reported F-score. This experiment allows not only

the direct comparison of our method to previous ones but also an indirect comparison of our evaluation protocol to theirs. If a method performs worse in the proposed protocol than it did originally, this is evidence that our protocol is stricter.

Segundo & Lemes (2015) provided an implementation of their method. Since we were unable to obtain an implementation, the required detections or a trained model for Su et al. (2017)'s method, we reimplemented it according to what is described in the paper. However, we were unable to reproduce their results. We attribute our failure to what we consider insufficient detailing of the method. Specifically, there is neither a description of how to convert the proposals into detections nor how to determine which detections are true and which are false. Since this conversion is not described, we assumed they used the traditional post-processing method (ZHAO et al., 2010). To improve transparency, we make our reimplementation of this method publicly available.

All the neural network models are optimized using Stochastic Gradient Descent (SGD) (ROBBINS; MONRO, 1985) with early stopping. Our code uses Tensorflow (ABADI et al., 2016), NumPy (OLIPHANT, 2006), and OpenCV (BRADSKI, 2000). We make the code required to reproduce our experiments, alongside trained models, publicly available.

We manually tune the learning rate and its decay schedule, the batch size, the dropout rate and the amount of weight decay by measuring the F-score of the patches classified as pores using the training labels. We set their values at 10^{-1}, exponentially decayed by 0.96 every 2000 steps, 256, 0.2 and 0, respectively. The post-processing parameters are obtained by performing a grid search, optimizing validation F-score with the trained model for p_t and the NMS intersection threshold i_t. The range of the search for p_t is $\{0.1, 0.2, ..., 0.9\}$ and for i_t is $\{0, 0.1, ..., 0.7\}$. The chosen values were $p_t = 0.6$ and $i_t = 0$. In our implementation of NMS, $i_t = 0$ corresponds to discarding bounding boxes that have any amount of intersection.

2.4 RESULTS

Our results for the post-processing ablation study are presented in Table 2.2. They reveal that using our method with the proposed post-processing is better than using the traditional one, according to the F-score metric. The FDR for traditionally post-processed detections is 0.06% lower in absolute than the proposed method, while its TDR is 16.85% lower in absolute values. This shows that the proposed post-processing technique merges detections of the same pore better than the traditional approach.

Post-processing	**TDR**	**FDR**	**F-score**
Proposed	**91.95%**	8.88%	**91.53%**
Traditional	75.10%	**8.82%**	82.36%

Table 2.2: Post-processing ablation study. Comparison of pore detection results with the proposed method, controlling for the post-processing method. This shows that the proposed method benefits from combination with post-processing method. Rows are sorted in order of decreasing F-score and the best results per column are in boldface.

Method	TDR	FDR	F-score
Proposed	**91.95%**	**8.88%**	**91.53%**
Su et al. (2017)	70.77%	11.58%	78.58%
Segundo & Lemes (2015)	89.31%	38.02%	73.17%

Table 2.3: Comparison of pore detection results using our evaluation protocol. We were unable to obtain an official implementation for Su et al. (2017)'s method, so these results are for our reimplementation. The drop in performance for the previous two best pore detection methods indicate that our evaluation protocol is stricter than previous ones. Also notice that our method is the state-of-the-art for all three metrics in this protocol. Rows are sorted in order of decreasing F-score and the best results per column are in boldface.

Method	TDR	FDR	F-score
Su et al. (2017)	88.6%	**0.4%**	**93.78%**
Proposed	**91.95%**	8.88%	91.53%
Segundo & Lemes (2015)	90.80%	11.10%	89.84%
Teixeira & Leite (2014)	86.10%	8.60%	88.67%
Wang et al. (2017)	83.65%	13.89%	85.88%
Zhao et al. (2010)	84.80%	17.60%	83.58%

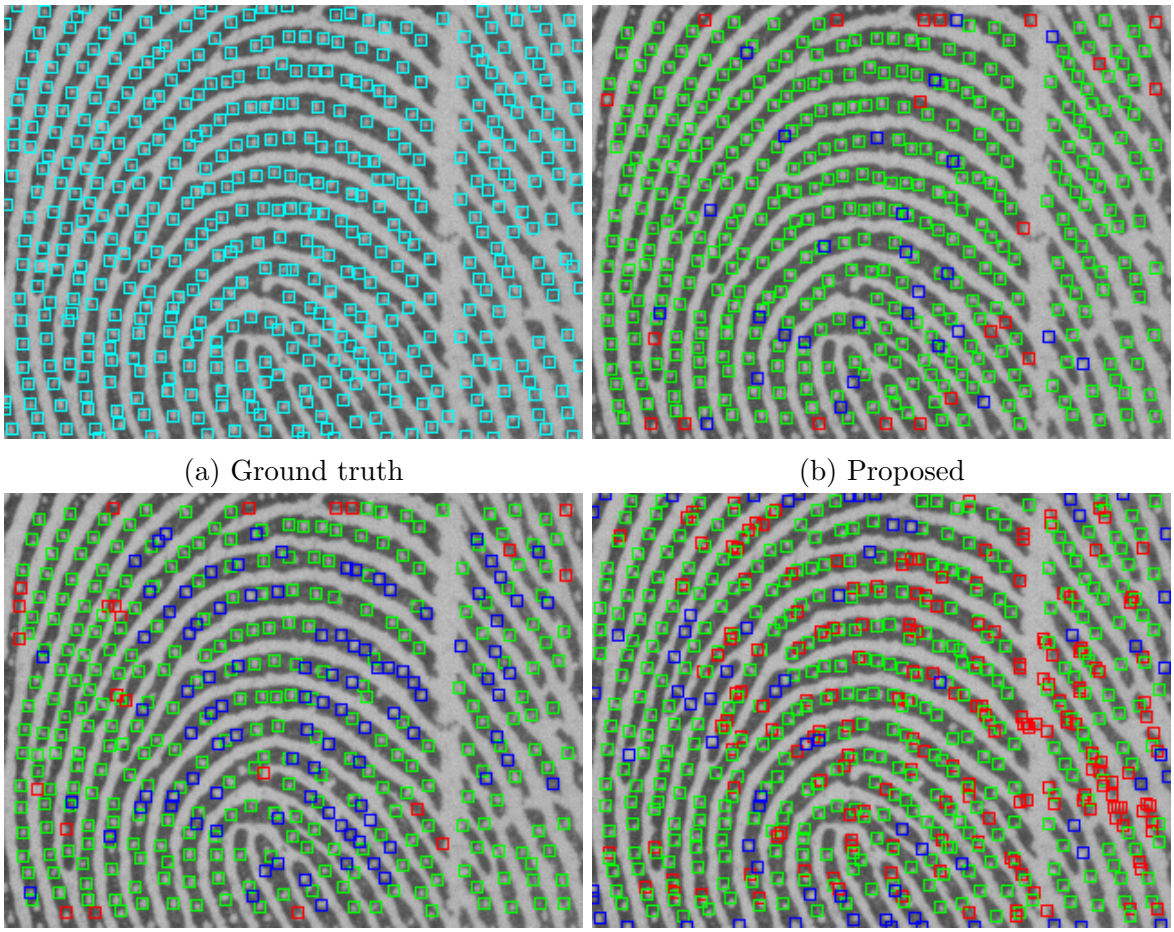
Table 2.4: Reported pore detection results. These results are not comparable because each work adopts its own validation method. Rows are sorted in order of decreasing F-score and the best results per column are in boldface.

Figure 2.3 presents qualitative results and Table 2.3 shows metrics for our second experiment. Table 2.4 provides reported results for comparison. The depicted fingerprint was taken from the test set and it was not seen before it was chosen. Nevertheless, we consider it representative of the quantitative results of our second experiment.

First, we remark that while Segundo & Lemes (2015)'s method finds many true detections (green squares in the figure), it also proposes many detections which are not pores (red squares). This result is observed not only in our experiments but in this work's original validation as well. The statistics obtained for this method are perfectly comparable to ours, because the authors provided us their implementation. These results show that the proposed method is an improvement over Pamplona Segundo & Lemes' and that the employed evaluation protocol is stricter.

As for Su et al. (2017)'s method, one can see in Figure 2.3 that it does not make many false detections. Our reimplementation, however, fails to detect a significant number of pores (blue squares in the figure), which impacts its TDR in our experiments. Despite reporting high TDR (88.6%), the qualitative results shown in Su et al. (2017)'s work display many undetected pores, which is consistent with our findings.

The decrease in Su et al. (2017)'s reported performance is more evidence that the



(a) Ground truth (b) Proposed

(c) Su et al. (2017) (d) Segundo & Lemes (2015)

Figure 2.3: Qualitative comparison of pore detection methods using the proposed evaluation protocol. The image is the last from the test set and was picked before the authors saw results for it. Green squares mark true positives (true detections), red squares mark false positives (false detections), and blue squares mark false negatives. For the proposed method (2.3b) and Su et al. (2017)'s (2.3c), only pores that can be detected are considered. Best viewed in color.

proposed protocol is stricter than previous validation methods. However, due to ours being a reimplementation and the lack of explanation in how detections are post-processed in that paper, we cannot guarantee this.

The proposed method not only outperforms our reimplementation of Su et al. (2017)'s method in the proposed evaluation protocol, it does so even when our method is traditionally post-processed. We consider this comparison fair because the optimization method, the effort put into optimizing the models, and the dataset split were the same for both methods. Moreover, we make the code for our method and the reimplementation of theirs publicly available.

We believe these arguments and the provided quantitative results are strong evidence to support our claim that the proposed evaluation protocol is stricter than previous validation methods. Accepting these evidence is enough to make the proposed method the state-of-the-art for the pore detection task. We further remark that our model has 96,548 parameters, in contrast to 21,211,074 (SU et al., 2017) or 1,696,512 (WANG et al., 2017) parameters of the previous neural network methods. We believe this improvement is due to the use of a more adequate number of parameters for the task and the effectiveness of our post-processing method.

Analyzing the visual results of the proposed method in Figure 2.3 provides some insight into what causes it to make a mistake. Most of the false positives of our method fall in two cases. The first is caused by pores that are visible in the region for which proposals are possible but whose centers are in the non-visible borders. This causes these pores to be discarded in our protocol. The same can be observed for our reimplementation of Su et al. (2017)'s method. The second case is for image regions in which it is difficult to distinguish if it is not, in fact, a pore. These can be, other than pores, lighter ridge regions, scars or bifurcations.

2.5 DISCUSSION AND FUTURE WORK

In this work, we trained an FCN to detect pores in high-resolution fingerprint images. It is our opinion that, as the writing of this paper, it is impossible to perfectly compare all the pore detection methods. This is because there is no default evaluation protocol in the standard pore detection benchmark, PolyU-HRF. In our attempt to establish that the proposed method is the state-of-the-art, we conducted a rigorous and extensive empirical evaluation. It includes an ablation study of our post-processing method, perfectly comparable results for one of the previous methods and our best efforts in reimplementing another. The evidence we gather point to our method indeed being the state-of-the-art.

Our second contribution, if effective, can solve the problem of non-comparable results. We propose a detailed and reproducible evaluation protocol, the first for this task, incredibly. Not only this, but we make the code for reproducing all of our experiments available.

We see potential of improving pore detection using semi-supervised learning techniques with the unlabeled training data in PolyU-HRF's *DBI* subset, because it has 14 times the images in the training split we propose. Previous work on semi-supervised learning indicates this is enough to offer substantial improvement (OLIVER et al., 2018).

The main limitation of our method is its possible susceptibility to domain variations (GANIN et al., 2016). Since PolyU-HRF's data is collected using a single sensor, it is possible that our pore detector fails to generalize to images collected with other sensors. The difference in fingerprint images collected with different sensors is significant, as can be seen when comparing images from PolyU-HRF to those from other datasets (ZHANG et al., 2017). If this is truly an issue, collecting more data and using domain adaptation might be necessary (GANIN et al., 2016).

The end objective of developing a pore detector is to use it to perform fingerprint recognition. We use the method described in this chapter in the next one to do so.

3

AUTOMATIC DATASET ANNOTATION TO LEARN CNN PORE DESCRIPTION

Automated fingerprint recognition systems mostly work by detecting keypoints, usually minutiae or pores, and matching them across images. Sometimes, the matching approach uses local descriptors to describe the keypoints' neighborhood and establish correspondences (CAO; JAIN, 2018; ZHAO et al., 2009; SEGUNDO; LEMES, 2015). These approaches usually provide excellent recognition results and can mostly be categorized into two settings. The first one uses low-resolution sensors but requires high-quality images comprising the entire fingerprint (JAIN et al., 2007; SEGUNDO; LEMES, 2015). The second setting requires high-resolution images and relies on sophisticated and computationally expensive algorithms to match keypoint descriptors (SEGUNDO; LEMES, 2015; SU et al., 2017; LIU et al., 2011). The latter usually provides better recognition results, thus allowing the use of partial fingerprints while maintaining the system secure. The focus of this work is on the second scenario, high-resolution fingerprint recognition. Specifically, we consider pores as keypoints, as many works in the literature do (ZHAO et al., 2009; SEGUNDO; LEMES, 2015; LIU et al., 2011; LIU et al., 2017; SU et al., 2017).

We attribute the previously mentioned need for sophisticated matching algorithms to the widespread use of hand-crafted keypoint descriptors. These descriptors do not require training data and are fast to compute (ZHAO et al., 2009; SEGUNDO; LEMES, 2015), but are not very robust to some of the variations to which fingerprint images are susceptible, *e.g.* the opening and closing of pores and distortions caused by the finger pressure applied to the sensor surface. This causes many spurious keypoint correspondences when using simple matching criteria.

There were two ways to solve this problem. The first one involves improving descriptors. While there are approaches that learn keypoint descriptors with better results than hand-crafted ones (TIAN et al., 2017; MISHCHUK et al., 2017; HE et al., 2018), their use is incipient in the fingerprint recognition literature (CAO; JAIN, 2018; ZHANG et al., 2017). This is probably caused by the lack of adequate keypoint annotations in publicly available fingerprint datasets, a requirement for the known approaches to learning local

descriptors. We believe this led researchers to focus on developing more and more sophisticated matching algorithms, the second way to address the spurious correspondences problem (LIU et al., 2011; SEGUNDO; LEMES, 2015; LIU et al., 2017). A detailed analysis of previous works is given in section 2.1.

In this work, we approach this problem using the former approach. First, we show that by aligning training images, we can automatically generate the required keypoint annotations for them. This allows learning local descriptors for pores with a convolutional neural network (CNN) in a fully supervised manner. While it is common to find previous works that align fingerprint images, they do so as a step in the matching stage to discard spurious correspondences (ZHAO et al., 2009; LIU et al., 2011). Our work, on the other hand, uses fingerprint alignment to generate training data. As we do this offline, and not as part of the matching method, it does not affect the computational cost to perform recognition. Second, we show that using the proposed descriptors with a simple matching algorithm improves the state-of-the-art performance in the standard high-resolution fingerprint recognition benchmark: the Hong Kong Polytechnic University High-Resolution-Fingerprint (PolyU-HRF) dataset (ZHAO et al., 2009).

Our method for fingerprint recognition consists of describing each pore with the learned descriptors, finding pore correspondences across images, and determining if the number of these correspondences is above a fixed threshold.

The contributions of this work are:

- we detail how aligning fingerprint images in a training set provides the required annotations to train a CNN that learns local keypoint descriptors (section 3.2);
- we describe our fingerprint recognition algorithm, which combines a simple matching criterion with the learned descriptors (section 3.3);
- we show that the proposed method yields state-of-the-art performance in the standard benchmark for the task (section 3.4). To verify our claim that the need for sophisticated matching is due to the use of hand-crafted descriptors, we conduct an ablation study with the most successful pore descriptors used in previous works.

3.1 RELATED WORK

We focus our review on fingerprint recognition works that use pores as keypoints, PolyU-HRF as benchmark, or that learn keypoint descriptors.

Direct Pore descriptors (hereon called DP descriptors for simplicity) are obtained by normalizing fingerprint patches around pores for contrast and rotation variations (ZHAO et al., 2009). This means that using such descriptors requires computing the fingerprint orientation at every pore location. In the first work that uses them, first, mutual nearest neighbors are used as correspondences. This initial correspondence set is filtered using Random Sample Consensus (RANSAC) (FISCHLER; BOLLES, 1981) and the final number of correspondences is used as score.

The matching algorithm of this method was improved by using tangent distance and sparse representations to find the initial set of correspondences and replacing RANSAC

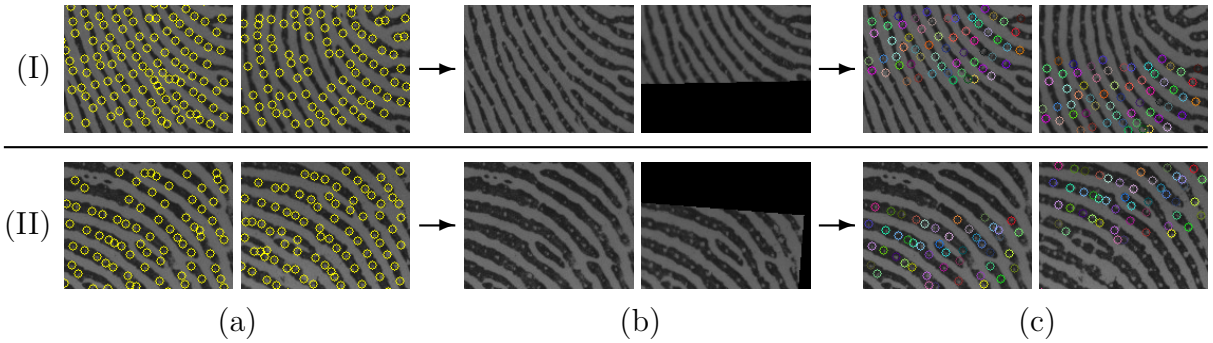


Figure 3.1: Automatic pore annotation pipeline. (a) A fingerprint dataset for which images have subject labels is the input to our method. Images in the same row belong to the same subject, *i.e.* images in the first row belong to subject (I), images in the second row belong to subject (II). (b) For each subject, a reference image is picked arbitrarily and every other image of that subject is aligned to it. The left images are the references for their subjects. The right images are transformed only to show the found alignment. (c) Pores that are not visible in some image of its subject are discarded. The remaining pores are annotated with unique labels. In these images, circles of different colors correspond to different, unique pore labels. Though there are colors in one image that look the same, they represent different labels. Notice how the alignment is used only to find pore correspondences; the images are unchanged by the end of the annotation process. Best viewed in color.

with weighted RANSAC (WRANSAC) (ZHANG et al., 2008). The overall method is called TDSWR (LIU et al., 2011). While it greatly improves recognition results, the use of sparse representations in the correspondence establishment step requires solving an NP-Hard optimization problem (DAVIS et al., 1997). The authors approach it by solving instead an ℓ_1-regularized least squares problem. This surrogate problem can be solved polynomially, but it forfeits theoretical guarantees of sparse representations and still involves solving a linear program with many variables and constraints.

The latest improvement of DP-based approaches is the use of pores in the neighborhood of minutiae. This method consists of hierarchically combining the matching scores of TDSWR obtained using two different subsets of pores. The first set consists of all pores; the second one only considers pores that are spatially close to minutiae, which the authors call distinctive pores (LIU et al., 2017). This approach indirectly uses minutiae to recognize fingerprints, as local descriptors of pores that are close to minutiae can be seen as descriptors of the minutiae themselves. Other methods also report better scores when combining pore-only matching scores with that of minutiae (ZHAO et al., 2009; LIU et al., 2011).

Affine Fourier Moment-Matching (AFMM) (SU et al., 2017) considers solving a different optimization problem to match fingerprints. It aims to minimize the energy cost function between the second-degree Fourier moments of the input fingerprint images using gradient descent. The first variant of this method, AFMM-G, considers the sums of the energy cost functions for ridge patterns, minutiae, and pores. The second and better variant is called AFMM-B. It proposes to first divide the gallery fingerprint image into

non-overlapping blocks. Then, for each block, a sliding window iterates over the query fingerprint to find the image region with the smallest energy cost function. If this match is below a certain threshold, a correspondence is established. The matching score is the total number of correspondences. AFMM-B, then, requires that for every block and every possible sliding window position, an optimization problem be solved with gradient descent. For partial fingerprints, the authors use 12 blocks. They do not report results for full fingerprint images.

Another approach altogether involves describing pores using SIFT descriptors (LOWE, 1999) to establish pore correspondences. This method discards spurious correspondences by reconstructing fingerprint ridges using pore detections. Matches are scored, then, using a spatially informed metric over pairs of correspondences (SEGUNDO; LEMES, 2015).

The first approach, to the best of our knowledge, to learn local keypoint descriptors for fingerprint images was used for latent fingerprints and uses minutiae as keypoints (CAO; JAIN, 2018). It trains a CNN patch descriptor in a dataset of rolled fingerprint images. To establish minutiae correspondences for training, it uses traditional fingerprint minutiae matching methods, discarding low-confidence correspondences as spurious. One disadvantage of this approach, when compared to aligning fingerprints, is that might allow false minutiae correspondences that have high-confidence scores to be in the training set. We also believe that this approach might not add hard minutiae correspondences to the training set because traditional minutiae matching algorithms are sub-optimal.

GlobalNet (ZHANG et al., 2017) is a CNN trained to generate global image descriptors for 2000dpi partial fingerprint images. Fingerprint recognition is done by determining if image descriptors for the input images are sufficiently similar.

MinutiaeNet (ZHANG et al., 2017) claims to do something very similar to what we do in this work: describing local neighborhoods of keypoints and match partial fingerprints by finding correspondences amongst the generated descriptors. What it actually does is, first, cropping 320×320 patches around the center of each minutia. The patches are then described using a CNN; recognition is done using all the resulting descriptors. We believe this setup is better framed as performing recognition using an ensemble of global image descriptors, in which each descriptor is obtained from random crops of the original image, than matching with local keypoint descriptors. In the global image descriptor setting, almost the entirety of the image is visible in each described sub-image. Each of the MinutiaNet patches contains more than a fourth of the entire source image, which has dimensions 480×800, and there are more than four minutiae per image. For the local keypoint descriptor setting, on the other hand, sub-images are restricted to small image regions around the keypoints. While we admit that there is no exact definition of what "small image regions" means in this context, we also remark that using an ensemble of global image descriptors is known to improve recognition (PARKHI et al., 2015).

Another similarity to our work is that MinutiaNet is trained with triplet loss, albeit with modifications proposed by its authors. The authors defend that these modifications are necessary to improve training convergence, but we did not find that that was necessary to train our CNN.

Finally, the results of the last two methods, GlobalNet and MinutiaNet, are not

reproducible. This is so because the only dataset for which they report results is not publicly available. Besides that, we argue that the evaluation protocol we follow is more rigorous. While the dataset used in those experiments has 1,800 images, more than PolyU-HRF, the authors split it in a way that 90% of the images are in the training set. This split is considerably easier than PolyU-HRF's because it not only provides a much bigger training set - 1620 against 210 - but it also tests in fewer images and subject identities - 180 against 1480. Furthermore, the number of subjects exclusively in the test set in PolyU-HRF, at least 113, is greater than the total number of subjects in their test set, 30.

3.2 CNN PORE DESCRIPTION

In this work, we aim to train a CNN to describe pores such that multiple instances of a same pore have more similar descriptors than two distinct pores. It is important to notice that subject labels do not coincide with pore labels: a single subject has many pores, and, therefore, many pore labels per image.

Since the learning approach we employ requires supervision, we must have an annotated dataset. In this situation, this means that every pore must receive a unique label. No fingerprint image dataset provides such annotations, to the best of our knowledge. We believe this is because such labeling task would require workers to annotate hundreds of pores per image and a few images per subject, which would be expensive, time-consuming and burdensome.

Thus, we devised an automated procedure for annotating pores that works by aligning fingerprint images. Figure 3.1 illustrates its overall structure. Images are iteratively aligned to a reference image that belongs to the same subject. This alignment process uses hand-crafted descriptors to establish pore correspondences and it also enforces spatial coherence using the alignment of the previous iteration. The final alignment allows discarding pores outside the overlap area and assigning unique labels to the remaining ones.

As we stated before, establishing correspondences based on hand-crafted descriptors causes many spurious correspondences. Since we are specifically annotating a dataset to learn a robust descriptor, such a descriptor is not yet available to be used for alignment. However, the previous methods are not unsuited to alignment, because it can be more costly to run and is an easier task than matching. Alignment can be more computationally costly because it is only executed once, to annotate the training data for the CNN. It is the trained CNN that will be used for matching. Matching is a harder task than alignment because it is required to work for both positive image pairs, *i.e.* images that belong to the same subject, and negative pairs, *i.e.* images that do not belong to the same subject. Alignment, on the other hand, is required to work only on positive pairs. We describe the annotation process in detail next.

3.2.1 Automatic pore annotation

Since for alignment we are using hand-crafted descriptors, preprocessing is essential. First, we apply a median blur, to denoise the image, and we equalize the images' contrast with CLAHE (ZUIDERVELD, 1994). Both are established preprocessing steps in the fingerprint community (SEGUNDO; LEMES, 2015).

After, we compute SIFT (LOWE, 1999) descriptors for each pore. SIFT descriptors have been previously used to perform pore-based fingerprint recognition with good results (SEGUNDO; LEMES, 2015). While there are other domains in which SIFT was sometimes shown to be inferior to other techniques, it was the first descriptor we tried. Since we were successful and aligning fingerprint images is not the purpose of this paper, we only tried SIFT.

First, we find correspondences between SIFT descriptors in images I_1 and I_2 using the original SIFT criterion: descriptors must be mutual nearest neighbors and pass a distance ratio check for their second nearest neighbors (LOWE, 1999).

With this initial set of correspondences, we align I_1 and I_2 using the closed form solution to the absolute orientation problem (HORN et al., 1988). The overlap between images is determined using the alignment and pores outside of it are discarded. We repeat this process, but in iteration $i+1$ we consider the following metric for pores p_1 of I_1 and p_2 of I_2 to find descriptor correspondences:

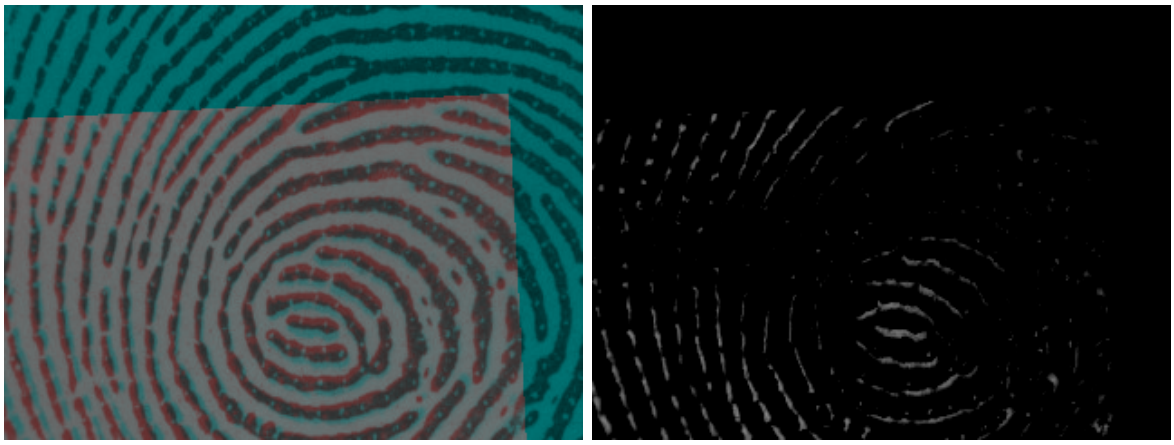
$$d_{i+1}(p_1, p_2) = d_{SIFT}(p_1, p_2) + \frac{\lambda}{w_i + \epsilon}||p_1 - T_i(p_2)||^2, \tag{3.1}$$

where d_{SIFT} is the distance between SIFT descriptors, T_i is the alignment transformation in iteration i, w_i is the mean squared error (MSE) of the correspondences found in iteration i, λ is a user specified parameter, and ϵ is used for numerical stability. This process stops in convergence or when the maximum number of iterations t is reached.

The second term in Equation 3.1 represents the Euclidean distance, in aligned space, between the 2D coordinates of pores p_1 and p_2. We use this term to determine how good the current alignment is. The intuition behind this thinking is that a perfect alignment should have zero MSE. So, when the alignment has high MSE, the images are not yet well aligned. This means that the similarity of pore descriptors will outweigh the Euclidean distance between the pores when establishing correspondences. Similarly, when the alignment has low MSE, the images are better aligned. Thus, correspondences with similar descriptors that do not fit the current alignment are unlikely to be created, reducing spurious correspondences. This rationale is encoded in Equation 3.1 by w_i, with λ serving to control the proportion between d_{SIFT} and Euclidean distance.

Having the images aligned does not solve the dataset annotation problem in its entirety. One must also be able to determine what is the minimum distance between two keypoints so that they should have a different label. This is an issue because two fingerprint images are often not perfectly alignable with rigid transformations. Figure 3.2 shows images of the same fingerprint that are not perfectly alignable due to non-rigid transformations.

To address this issue, we assume that the proposed alignment is capable of perfectly aligning images, *i.e.* the final transformation T maps coordinates between aligned im-



(a) (b)

Figure 3.2: Aligned fingerprint images. (a) Aligned images are superimposed in different color channels. (b) Only the alignment residues for one of the images are shown. These images show that fingerprints are not perfectly alignable by rigid transformations. Best viewed in color.

ages without distortion. Hence, to generate keypoint identity annotations, we choose an arbitrary image for each subject as a reference image. Each pore in reference images is then assigned a unique label. Afterward, every other image is aligned to its corresponding reference image. The coordinates for the labeled pores are then found in the other images belonging to the same subject using the alignment. If a pore's coordinates are outside the image boundaries for some image, then this pore is discarded. Finally, a patch of dimensions 33×33 is extracted around each of the remaining pores. The set of pore patches and their corresponding labels form the training set. Notice that the training images are aligned only in the sense that we find transformations to map coordinates from one of them into the other. We use the transformations to annotate the data, not to transform the images themselves.

Allowing the direct use of the alignment transformation to generate pore labels is one of the benefits of aligning images to annotate the dataset. This means that if the alignment is successful, even pores for which the hand-crafted descriptors were not able to find a confident correspondence are correctly labeled. The other benefit is that this approach assures that the pore annotations are spatially coherent, *i.e.* no label is assigned to a pore in a way that is spatially incompatible with the other labels assigned by the method.

3.2.2 Training problem formulation

Given image patches of dimensions 33×33 centered on pores with corresponding pore labels, our task is to train a CNN to learn pore descriptors such that descriptors for pores with the same label are closer in ℓ_2 space than those with different ones. HardNet (MISHCHUK et al., 2017) showed state-of-the-art performance in the wide baseline

stereo, patch verification, and instance retrieval tasks until recently (HE et al., 2018). We can frame the fingerprint recognition problem based on finding pore descriptor correspondences as multiple instance retrieval problems. This allows training HardNet to learn pore descriptors in ℓ_2 space. Since its architecture takes as input patches with dimensions 32×32, we discard the last row and the last column of the pore patches. Doing this enables us to use HardNet's architecture without modifications.

While we do not modify the network's architecture, we train it differently. HardNet is supposed to be trained with a mini-batch containing a single positive pair of patches. However, Schroff et al. (2015), which proposes a very similar approach to learn face embeddings, state that the approach taken by HardNet is sub-optimal. None of these works provide evidence that their approaches are better than the other, but Schroff et al. (2015) state that unreported experiments show that triplet loss works better with more examples of the same identity in each batch. With these considerations in mind, we employ Schroff et al. (2015)'s approach, which is called triplet semi-hard loss.

Also, to improve the robustness of our descriptors, we augment the dataset as we sample mini-batches from it. This is done by randomly sampling a transformation that includes translating, rotating, and perturbing the contrast and brightness of each patch from a multivariate normal distribution with zero covariance. Since we augment after we crop the patches from the image, we treat translations and rotations as if the patch is zero-padded. We also linearly interpolate the image when the sampled transformation requires it.

It is important to notice that our procedure for augmentation is able to simulate patches undergoing rigid transformations. This corresponds to a user translating and changing his/her finger's orientation with respect to the sensor from one capture session to another. If rigid transformations were the only concern, we could forfeit pore identity annotation altogether and augmentation alone would suffice to provide same identity patches. However, if a user presses his/her finger against the sensor with a different amount of pressure in a subsequent session, then there is no rigid transformation that can map the earlier image into the new one. This is not to mention other sources of variation for fingerprint images, such as those caused by age or the presence of sweat or dirt in the finger. This is why pore identity annotation is essential to our work.

3.3 FINGERPRINT MATCHING

Given two fingerprint images I_1 and I_2 annotated with pore coordinates, the proposed method requires descriptors for each pore. They are described using a CNN trained following the steps in section 3.2. This results in descriptor set D_1, for the pores of image I_1, and descriptor set D_2, for the pores in image I_2.

It is important to notice that the alignment in subsection 3.2.1 is not part of our fingerprint matching method, nor is the use of SIFT descriptors. These are used only in steps prior to the training of the CNN, which takes place once and is separate from the fingerprint matching method.

Next, we establish correspondences by finding mutual nearest neighbors between the descriptors in D_1 and in D_2. This means that nearest neighbors for descriptors in D_1 are

searched in D_2 and vice-versa. We further filter these correspondences using a distance ratio check for the second nearest neighbors (LOWE, 1999).

The recognition score, then, is the final number of established correspondences. If they are above a given threshold, the pair is deemed genuine, *i.e.* the images belong to the same fingerprint; otherwise, the pair is considered an impostor pair, *i.e.* they do not belong to the same fingerprint.

Our approach to fingerprint matching is considerably simpler than previous works. The proposed method requires aligning only training images and this only happens once, prior to the training of the CNN. Previous methods require alignment, either explicitly or implicitly, *e.g.* with RANSAC, for matching (ZHAO et al., 2009; LIU et al., 2011). Other methods require solving optimization problems or reconstructing fingerprint ridges (SU et al., 2017; LIU et al., 2017; SEGUNDO; LEMES, 2015).

3.4 EXPERIMENTS AND RESULTS

We use the PolyU-HRF dataset (ZHAO et al., 2009), the public benchmark for high-resolution fingerprint recognition, for our experiments.

We conduct experiments to (1) compare our results to previous methods, and (2) determine if found differences are due to improved descriptors. To assess (2), we conduct an ablation study, keeping the entire matching method unchanged except for the used pore descriptors. We replace the proposed descriptors for each descriptor previously used in the literature, comparing the observed changes.

The pore description CNN is optimized using Stochastic Gradient Descent (SGD) (ROBBINS; MONRO, 1985) with early stopping. Our implementation uses Tensorflow (ABADI et al., 2016), NumPy (OLIPHANT, 2006), and OpenCV (BRADSKI, 2000). We make all the code required to train the CNN and to reproduce our experiments, alongside the trained model, publicly available.

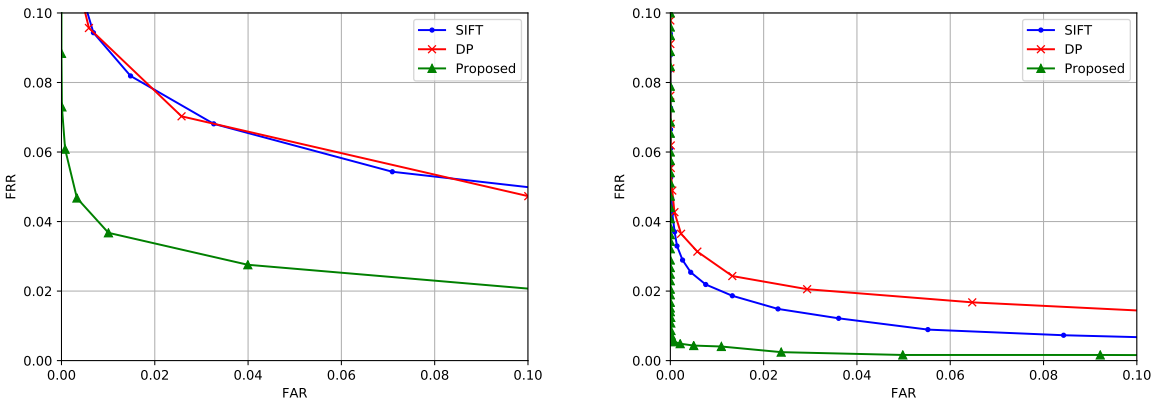
3.4.1 PolyU-HRF dataset

Polyu-HRF is divided into two subsets: *DBI* and *DBII*.

DBI consists of 210 images of partial fingerprints, with 6 images per subject, in a training set and 1,480 images, with 10 images per subject, in a test set. It provides the corresponding identity for each of those images and separates them in acquisition sessions.

DBII has no training set and its test set has the same number of images and subjects of *DBI*'s. The difference is that, in *DBII*, the images are of full-size fingerprints, instead of partial ones. Its label information is like that of *DBI*.

PolyU-HRF's *DBI* and *DBII* do not provide pore coordinate annotations, and, therefore, lack pore identity annotations for its images. We believe no fingerprint dataset provides this information and that this is the reason why hand-crafted features are so prevalent for this task. This also means that we must detect pores in the images to conduct fingerprint recognition experiments.



(a) (b)

Figure 3.3: ROC curves for descriptor ablation study in (a) *DBI* and (b) *DBII*. Keeping the matching method unchanged and changing the descriptor shows that the proposed descriptor is responsible for the improvement in recognition results. This is observed in the entirety of the plot. Best viewed in color.

3.4.2 Experimental setup

We conducted fingerprint recognition experiments for the proposed method in both subsets of PolyU-HRF destined for testing. The test set of *DBI* is destined to evaluate performance in the partial fingerprints scenario, while *DBII* is for evaluating the method with full fingerprint images.

We followed the established protocol for validating fingerprint recognition methods in PolyU-HRF, which was proposed by its authors and is the same for both *DBI* and *DBII* (ZHAO et al., 2009). This protocol consists of performing all 3,700 comparisons of images with the same identity but across sessions. These are the only genuine comparisons to be performed. The impostor comparisons are the 21,756 comparisons of images with different identities obtained by matching, for each subject, the first image of the first session with the first image of the second session of every other subject.

In order to determine if observed differences in our results arise from the robustness of the proposed descriptor, we conduct an ablation study. In it, we repeat our experiment, keeping the same protocol and overall fingerprint recognition method, but replace the proposed descriptors with every other local descriptor used by previous works to describe pores. In one experiment, we use SIFT descriptors (SEGUNDO; LEMES, 2015) and in the other we use DP descriptors (ZHAO et al., 2009; LIU et al., 2011; LIU et al., 2017).

For these other descriptors, we attempt to reproduce the local description pipeline as close as we can interpret from the work that proposes to use them as pore descriptors. To allow a fair comparison, we optimize the parameters for these experiments with other descriptors observing the recognition Equal Error Rate (EER) in the training set. The SIFT parameter values were scale of 8, median blur kernel dimensions of 3×3, and CLAHE clip limit of 3. For DP descriptors, we used 32×32 patches, obtained discarding

the last row and column of the 33×33 patches centered at each pore detection. The distance ratio check threshold is 0.7 for both of them.

PolyU-HRF's images in *DBI* and *DBII* are not annotated with pore coordinates. So, to conduct our experiments, we must first detect pores in these images. To this end, we use the pore detection FCN described in the previous chapter (DAHIA; SEGUNDO, 2018b) with its default parameters.

To generate the training set annotations, we empirically determined the alignment parameters by observing the results in a few training images. Their values are $\lambda = 500$, $\epsilon = 10^{-5}$ and $t = 10$. SIFT related parameters were used as determined previously, except for the distance ratio check threhsold, which was of 0.8.

We randomly split *DBI*'s training images, reserving 60% of them for an effective training set, *i.e.* images we actually use for training the description CNN. The remaining 40% of the training images form a validation set. Importantly, we make this split subject-independent, *i.e.* there are no subject identities both in the training set and in the validation set. Another important aspect of validation is that we perform manual hyper-parameter tuning and early stopping based directly on the end-task. That is, we observe the recognition EER using descriptors from the model in the current training stage. The distance ratio check is fixed at 0.7 during training to avoid unforeseen variations in the training dynamics.

The best trained model obtained recognition EER of 0% in the validation set. Its parameters are constant learning rate of 10^{-1}, batches of size 252, with exactly 6 patches per pore, dropout rate of 0.3, triplet semi-hard loss margin of 2.0, and no weight decay.

While we take steps to ensure that there is no identity that appears in both the training and validation sets to improve generalization, the default protocol of PolyU-HRF has subjects that appear in both its training and test sets. This setting is sub-optimal to evaluate methods based on deep neural networks, as it is known that they are able to memorize the training set (ZHANG et al., 2016).

We believe it is unlikely that dataset memorization represents a major issue in our evaluation. We argue this based on three facts. First, while there are subjects in both training and test sets, there is no image in the test set that is also in the training set. Second, our method obtains 0% recognition EER in a validation set that has no subject overlap with the training set. Third, despite this overlap, most subjects in the test set are not available for training.

3.4.3 Results

Figure 3.3 presents Receiver Operating Characteristic (ROC) curves for our ablation study in *DBI* and *DBII*. Table 3.1 displays EERs for these experiments.

Our ablation study provides strong evidence that the effectiveness of our approach is due to the improved descriptors. Replacing the proposed descriptors with SIFT ones results in 5.87% EER in *DBI*; with DP ones, it results in 5.98%. This shows that, contrary to what was previously reported (ZHAO et al., 2009), at least for this simple matching algorithm and partial fingerprints, the performance for SIFT and DP descriptors are very similar. In contrast, when using the proposed descriptors, the EER value is only 3.05%,

Descriptor	**DBI**	**DBII**
Proposed	**3.05%**	**0.44%**
SIFT	5.87%	1.71%
DP	5.98%	2.22%

Table 3.1: EERs for *DBI* and *DBII* for the descriptor ablation study. Using the proposed descriptors while keeping the matching method otherwise unchanged reduces EER to almost half of the best hand-crafted descriptors' in *DBI* and to almost a quarter in *DBII*. Best results are in boldface.

cutting the error almost in half. The effect of using learned descriptors is even more significant in *DBII*: the EER, 0.44%, is almost a quarter of SIFT's, 1.71%, and less than a fifth of DP's, which is 2.22%. The ROC curves also show that for the entire plot, our results are better.

Table 3.2 compare our results with previously reported methods in terms of EER for both subsets. We note that Su et al. (2017) do not report results for *DBII* and that is why its entry is blank in Table 3.2. Our results are only worse than methods that leverage the use of minutiae when comparing EER for both the partial fingerprints in *DBI* and the full fingerprints in *DBII*. This confirms that the proposed method is the state-of-the-art in pore-based fingerprint recognition for high-resolution images, according to the PolyU-HRF benchmark.

Method	**DBI**	**DBII**
Liu et al. (2017)*	**2.17%**	**0.17%**
Proposed	**3.05%**	**0.44%**
Liu et al. (2011)	3.25%	0.53%
Su et al. (2017)*	3.66%	-
Segundo & Lemes (2015)	3.74%	0.76%
Zhao et al. (2009)*	12.40%	5.93%
Zhao et al. (2009)	20.49%	7.05%

Table 3.2: EERs for fingerprint recognition in *DBI* and *DBII*. * marks methods that consider both minutiae and pores as keypoints. Our method outperforms all previous methods that use only on pores for recognition. We note that Su et al. (2017) does not report results for *DBII* and that is why its entry is blank. Best results for methods using only pores and overall methods are in boldface.

Observe that combining minutiae information with that of pores improves the result of fingerprint recognition. In *DBI*, Zhao et al. (2009) results go from 20.49% to 12.40% and Liu et al. (2017) goes from 3.25% to 2.17%. The same effect occurs in *DBII*. This is evidence that combining the proposed descriptors somehow with minutiae could yield even better results.

This comparison also confirms our hypothesis that more robust descriptors do not

need the sophisticated matching methods previously used in the literature. The previous methods either use the way in which we establish correspondences as an intermediate step (ZHAO et al., 2009; LIU et al., 2011; LIU et al., 2017), involve solving complex optimization problems (LIU et al., 2011; LIU et al., 2017; SU et al., 2017), or require reconstructing fingerprint ridges (SEGUNDO; LEMES, 2015). We believe this shows that they are more complex than the algorithm presented in section 3.3.

It is also interesting to remark that the only method to present better results quantified by EER than ours does so when combining already sophisticated matching methods with minutiae information. Other methods present higher EER, even when using minutiae information for matching.

3.5 DISCUSSION AND FUTURE WORK

In this work, we detailed how aligning fingerprint images allows us to automatically annotate a training set with keypoint labels. Our method iteratively establishes correspondences based on the similarity of hand-crafted keypoint descriptors and aligns positive pairs of images with them. This process is more computationally expensive than our matching algorithm. However, it has no impact in the running time of recognition since it only takes place once, prior to training the CNN. Finally, to show the usefulness of that approach, we annotated the training set of the default benchmark for high-resolution fingerprint recognition with pore labels and used those labels for training the CNN we used in our experiments.

We see as our main contribution the substantial empirical evidence to corroborate our hypothesis that improving local descriptors allows the use of simple matching rules to improve high-resolution fingerprint recognition. We both (1) obtain the state-of-the-art performance in the default benchmark for high-resolution, pore-based fingerprint recognition using a method that fulfills these criteria, (2) and conduct an ablation study with the most successful descriptors used in the literature to confirm that the improvement is due to better descriptors.

Our method, expectedly, has limitations. The most significant, in our opinion, has to do with using the data to learn the descriptor. It is known that deep learning approaches suffer from domain variability (GANIN et al., 2016). Fingerprint images obtained from different sensors vary considerably. This can be seen by comparing images in PolyU-HRF (ZHAO et al., 2009) and in the proprietary dataset used by MinutiaNet (ZHANG et al., 2017). The question, then, is if this could impact the performance of using a CNN trained with images from one sensor in images obtained from a different sensor.

Future work should investigate if domain variability represents a limitation of our method. This would involve testing the trained CNN in a dataset of images from a different sensor. As of the writing of this work, this would require collecting a new dataset or obtaining approval of use from the maintainers of a proprietary high-resolution fingerprint dataset. PolyU-HRF is the only publicly available dataset for high-resolution fingerprints, to the best of our knowledge. If such an investigation finds that domain variability impacts the method significantly, one could solve it in either a supervised manner, which would involve labeling fingerprints from several sensors, or an unsupervised

manner, *e.g.* using domain adversarial training (GANIN et al., 2016).

A straightforward way to improve our method would be to replace the triplet-based loss we use to train our CNN with histogram loss (USTINOVA; LEMPITSKY, 2016). Descriptors that use this loss (HE et al., 2018) have recently surpassed HardNet's performance in instance retrieval and are the new state-of-the-art for this task.

Using minutiae somehow in the proposed fingerprint matching method also appears likely to be able to improve our results. Previous work (LIU et al., 2017; LIU et al., 2011) show that this holds potential to make recognition significantly more discriminative. This would probably involve adding steps to the matching algorithm to detect and find correspondences for minutiae but the performance improvement could outweigh the increased computational cost for high-security environments.

Another promising line of work is the use of a global fingerprint texture descriptor, like GlobalNet (ZHANG et al., 2017), either combined with or side-stepping keypoint-based recognition altogether. This may prove difficult for the proposed PolyU-HRF protocol since it only allows 210 images to be used for training. The original dataset does not reserve images for validation, which must then be taken from the training set.

Finally, we made the code required to reproduce our experiments publicly available. We believe this substantially improves the transparency, reproducibility, and the usefulness of our work.

4

CONCLUSION

This work detailed a complete high-resolution, pore-based fingerprint recognition method, except for image acquistion.

We first provided evidence that training an FCN to detect pores in high-resolution fingerprint images as proposed can improve the state-of-the-art for this task, even with significantly less parameters than previous methods. The reason why we cannot be sure of this result is related to our second contribution to the pore detection task. Incredibly, prior to our work, there was no detailed and reproducible evaluation protocol for pore detection. We change that by proposing such a protocol for PolyU-HRF's *GroundTruth* subset. In this protocol, our method obtains 91.53% F-score, a significant improvement over what is obtained by the previous state-of-the-art in the same conditions, 78.58%.

Another contribution of our work is the detailing of how aligning fingerprint images allows us to automatically annotate a training set with keypoint labels. We used this approach to annotate PolyU-HRF's training set with pore labels. Those labels are then used for training the pore description CNN based on the HardNet architecture (MISHCHUK et al., 2017). The learned local descriptors allows the use of simple matching rules to improve high-resolution fingerprint recognition. Our method reduces the best EER for the pore-based setting from 3.25% to 3.05%, in *DBI*, and from 0.53% to 0.44%, in *DBII*.

We also conduct an ablation study with the most successful pore descriptors used in the literature to confirm that the improvement is due to the improved descriptors. The proposed descriptors is almost half the second best descriptor's EER in *DBI* - 3.05% *v.s.* 5.87% - and almost a quarter in *DBII* - 0.44% *v.s.* 1.71%.

Both the proposed methods, however, have limitations. As they both use data to learn representations, they are susceptible to the domain variability problem (GANIN et al., 2016). This is specially significant because fingerprint images obtained from different sensors vary considerably. Future work should, then, investigate if domain variability represents a limitation of our method, as the only public dataset, PolyU-HRF, was collected with a single sensor and is not appropriate to assess this.

Finally, we made the code for pore detection, pore description and the fingerprint recognition method overall publicly available.

BIBLIOGRAPHY

ABADI, M. et al. Tensorflow: a system for large-scale machine learning. In: *OSDI*. [S.l.: s.n.], 2016. v. 16, p. 265–283.

BRADSKI, G. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

CAO, K.; JAIN, A. K. Automated latent fingerprint recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, 2018.

DAHIA, G.; SEGUNDO, M. P. Automatic dataset annotation to learn cnn pore description for fingerprint recognition. *arXiv preprint arXiv:1809.10229*, 2018.

DAHIA, G.; SEGUNDO, M. P. Improving fingerprint pore detection with a small fcn. *arXiv preprint arXiv:1811.06846*, 2018.

DAVIS, G.; MALLAT, S.; AVELLANEDA, M. Adaptive greedy approximations. *Constructive Approximation*, v. 13, n. 1, p. 57–98, Mar 1997. ISSN 1432-0940.

FISCHLER, M. A.; BOLLES, R. C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, ACM, v. 24, n. 6, p. 381–395, 1981.

GANIN, Y. et al. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, JMLR. org, v. 17, n. 1, p. 2096–2030, 2016.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016.

HE, K.; LU, Y.; SCLAROFF, S. Local descriptors optimized for average precision. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2018.

HORN, B. K.; HILDEN, H. M.; NEGAHDARIPOUR, S. Closed-form solution of absolute orientation using orthonormal matrices. *JOSA A*, Optical Society of America, v. 5, n. 7, p. 1127–1135, 1988.

IOFFE, S. Batch renormalization: Towards reducing minibatch dependence in batch-normalized models. In: *Advances in Neural Information Processing Systems*. [S.l.: s.n.], 2017. p. 1945–1953.

IOFFE, S.; SZEGEDY, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: *International Conference on Machine Learning*. [S.l.: s.n.], 2015. p. 448–456.

JAIN, A. K.; CHEN, Y.; DEMIRKUS, M. Pores and ridges: High-resolution fingerprint matching using level 3 features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 29, n. 1, p. 15–27, Jan 2007. ISSN 0162-8828.

JARRETT, K. et al. What is the best multi-stage architecture for object recognition? In: IEEE. *Computer Vision, 2009 IEEE 12th International Conference on*. [S.l.], 2009. p. 2146–2153.

LIU, F.; ZHAO, Q.; ZHANG, D. A novel hierarchical fingerprint matching approach. *Pattern Recognition*, Elsevier, v. 44, n. 8, p. 1604–1613, 2011.

LIU, F.; ZHAO, Y.; SHEN, L. Feature guided fingerprint pore matching. In: SPRINGER. *Chinese Conference on Biometric Recognition*. [S.l.], 2017. p. 334–343.

LOWE, D. G. Object recognition from local scale-invariant features. In: IEEE. *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*. [S.l.], 1999. v. 2, p. 1150–1157.

MISHCHUK, A. et al. Working hard to know your neighbor's margins: Local descriptor learning loss. In: *Advances in Neural Information Processing Systems*. [S.l.: s.n.], 2017. p. 4826–4837.

OLIPHANT, T. E. *A guide to NumPy*. [S.l.]: Trelgol Publishing USA, 2006.

OLIVER, A. et al. Realistic evaluation of deep semi-supervised learning algorithms. *arXiv preprint arXiv:1804.09170*, 2018.

PARKHI, O. M.; VEDALDI, A.; ZISSERMAN, A. Deep face recognition. In: *British Machine Vision Conference*. [S.l.: s.n.], 2015.

REDMON, J.; FARHADI, A. YOLO9000: Better, faster, stronger. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2017. p. 6517–6525. ISSN 1063-6919.

REN, S. et al. Faster R-CNN: Towards real-time object detection with region proposal networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2015. p. 91–99.

ROBBINS, H.; MONRO, S. A stochastic approximation method. In: *Herbert Robbins Selected Papers*. [S.l.]: Springer, 1985. p. 102–109.

SCHROFF, F.; KALENICHENKO, D.; PHILBIN, J. Facenet: A unified embedding for face recognition and clustering. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2015. p. 815–823.

SEGUNDO, M. P.; LEMES, R. de P. Pore-based ridge reconstruction for fingerprint recognition. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. [S.l.: s.n.], 2015.

SRIVASTAVA, N. et al. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, JMLR. org, v. 15, n. 1, p. 1929–1958, 2014.

SU, H.-R. et al. A deep learning approach towards pore extraction for high-resolution fingerprint recognition. In: IEEE. *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. [S.l.], 2017. p. 2057–2061.

TEIXEIRA, R. F.; LEITE, N. J. Improving pore extraction in high resolution fingerprint images using spatial analysis. In: IEEE. *Image Processing (ICIP), 2014 IEEE International Conference on*. [S.l.], 2014. p. 4962–4966.

TIAN, Y.; FAN, B.; WU, F. L2-net: Deep learning of discriminative patch descriptor in euclidean space. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2017. p. 6128–6136. ISSN 1063-6919.

USTINOVA, E.; LEMPITSKY, V. Learning deep embeddings with histogram loss. In: *Advances in Neural Information Processing Systems*. [S.l.: s.n.], 2016. p. 4170–4178.

WANG, H. et al. Fingerprint pore extraction using u-net based fully convolutional network. In: SPRINGER. *Chinese Conference on Biometric Recognition*. [S.l.], 2017. p. 279–287.

ZHANG, C. et al. Understanding deep learning requires rethinking generalization. *arXiv preprint arXiv:1611.03530*, 2016.

ZHANG, D. et al. Matching images more efficiently with local descriptors. In: CITESEER. *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. [S.l.], 2008. p. 1–4.

ZHANG, F.; XIN, S.; FENG, J. Combining global and minutia deep features for partial high-resolution fingerprint matching. *Pattern Recognition Letters*, Elsevier, 2017.

ZHAO, Q. et al. Adaptive fingerprint pore modeling and extraction. *Pattern Recognition*, Elsevier, v. 43, n. 8, p. 2833–2844, 2010.

ZHAO, Q. et al. Direct pore matching for fingerprint recognition. In: SPRINGER. *International Conference on Biometrics*. [S.l.], 2009. p. 597–606.

ZUIDERVELD, K. Contrast limited adaptive histogram equalization. *Graphics gems*, Academic Press, p. 474–485, 1994.