

Fixed-length Dense Descriptor for Efficient Fingerprint Matching

Zhiyu Pan, Yongjie Duan, Jianjiang Feng[✉] and Jie Zhou

Department of Automation, BNRist, Tsinghua University, Beijing, China

psy20@mails.tsinghua.edu.cn, duanyj13@tsinghua.org.cn, {jfeng, jzhou}@tsinghua.edu.cn

Abstract—In fingerprint matching, fixed-length descriptors generally offer greater efficiency compared to minutiae set, but the recognition accuracy is not as good as that of the latter. Although much progress has been made in deep learning based fixed-length descriptors recently, they often fall short when dealing with incomplete or partial fingerprints, diverse fingerprint poses, and significant background noise. In this paper, we propose a three-dimensional representation called Fixed-length Dense Descriptor (FDD) for efficient fingerprint matching. FDD features great spatial properties, enabling it to capture the spatial relationships of the original fingerprints, thereby enhancing interpretability and robustness. Our experiments on various fingerprint datasets reveal that FDD outperforms other fixed-length descriptors, especially in matching fingerprints of different areas, cross-modal fingerprint matching, and fingerprint matching with background noise.

Index Terms—Fingerprint Matching, Fixed-length Dense Descriptors, Deep Learning.

I. INTRODUCTION

Fingerprint recognition has become a prevalent biometric technology used in various applications, such as access control, national ID cards, and law enforcement [1]. Typically, a fingerprint recognition system comprises three core modules: image acquisition, feature extraction, and matching. The feature extraction algorithm is designed to extract distinctive features, known as descriptors, from fingerprints. The matching algorithm then uses these descriptors to compute a matching score between samples. Numerous researchers have focused on developing automated and efficient algorithms for descriptor extraction and matching [2]–[5], aiming to meet the diverse demands of various fingerprint recognition applications.

Generally, fingerprint representations are categorized into two main types: minutiae set and fixed-length descriptors. Minutiae set based representations describe fingerprints by focusing on individual local minutiae, with each minutia having its associated descriptor. Most local minutiae descriptors [2], [4], [6] are generated by extracting patch images centered around the location and orientation of each minutia. This method provides a flexible representation of fingerprints by capturing detailed local features in the ridges and valleys, which enhances accuracy and adaptability to various fingerprint acquisition techniques and conditions. However, minutiae-based fingerprint representations pose several challenges, including high computational complexity [2], [4],

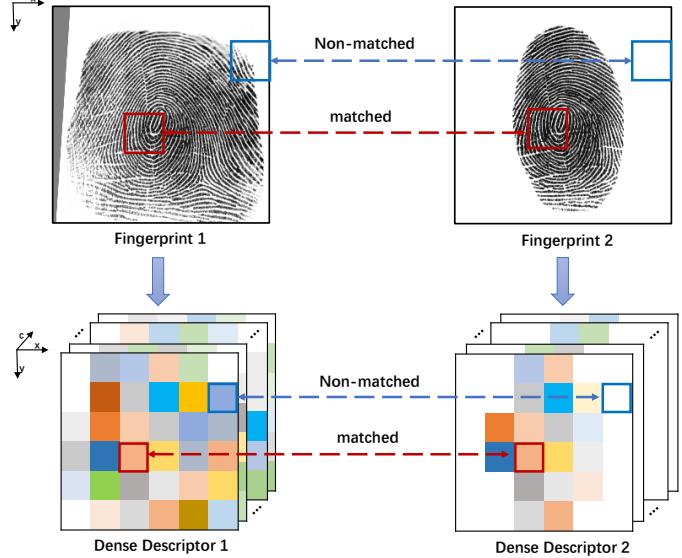


Fig. 1. The illustration of Fixed-length Dense Descriptors (FDD) of a rolled fingerprint and a mated plain fingerprint. Two corresponding regions and their features are marked.

difficulties in template encryption [7], and dependence on the precision of minutiae extraction [8], [9].

Fixed-length descriptors represent each fingerprint as a fixed-length tensor, allowing the similarity between multiple fingerprints to be efficiently calculated through parallel matrix multiplication. In the past, conventional fixed-length descriptors were typically defined by the local responses of multiple Gabor filters [10], parameters of the orientation field model [11], [12], local minutiae vicinities [13], histograms of minutiae triplet or pair features [14], [15], local ridge orientation and frequency [16], and local discretized texture descriptors [17]. These handcrafted representations often exhibited limited discriminating capability, leading to insufficient matching accuracy.

With the rapid development of deep learning technology, fixed-length representations are now extracted using deep networks from aligned fingerprints [18]–[20]. These methods demonstrate exceptional performance on high-quality large fingerprints, such as rolled fingerprints. However, most existing techniques represent a fingerprint image using a one-dimensional descriptor, which fails to sufficiently account for background areas. This limitation significantly decreases matching performance, especially when dealing with partial

[✉]Jianjiang Feng is the corresponding author.

This work was supported in part by the National Natural Science Foundation of China under Grant 62376132 and 62321005.

fingerprints or significant background noise. Furthermore, the matching of fixed-length one-dimensional descriptors is not restricted to the overlapping foreground area, reducing their effectiveness for fingerprints with varying areas. Gu et al. [21] explored extracting pyramid features at different scales and positions. Despite this approach, matching performance remains hindered by the coarse division of local regions and the fusion weights for different fingerprints. Additionally, extracting these multi-scale features demands higher computational resources and storage space.

Given the strong performance of dense descriptors in characterizing neighborhood of minutiae [6], this paper extends the concept to the whole fingerprint and introduces the Fixed-length Dense Descriptor (FDD). Dense descriptors are three-dimensional representations, where two dimensions correspond to the spatial dimensions of the original fingerprint image (Fig. 1). Consequently, dense descriptors contain valid values only within the fingerprint foreground, effectively eliminating the background. Additionally, the spatial alignment of fingerprints to be matched ensures that the dense descriptor spaces are also aligned, thus restricting the matching process to the overlapping foreground regions.

Accordingly, the first step in our proposed method is to align the fingerprint using a robust fingerprint pose estimation approach [22]. Next, we extract the localized dense description of the fingerprint and identify the valid area. To add global constraints on the aligned space, we utilize a 2D sinusoidal positional embedding, introducing position information into the fixed-length representations. This guides the model to distinguish similar fingerprint textures at different locations, further enhancing matching performance.

To verify the effectiveness and generalization ability of FDD, we evaluate it on diverse datasets containing various fingerprint areas and impression types, including rolled, plain, latent, and contactless fingerprints. Some of these datasets present significant challenges for fixed-length descriptors, such as fingerprints with diverse finger poses and small image sizes. The experimental results demonstrate the superiority of FDD compared to other fixed-length descriptors. Additionally, we conduct experiments on binarized FDDs, obtained using a constant threshold, which still achieve good performance. Therefore, the Fixed-length Dense Descriptor shows great potential for application in automated fingerprint recognition systems.

II. RELATED WORK

Fixed-length fingerprint descriptor can be classified into two categories: alignment-based and alignment-free.

Currently, numerous high-performance and robust fixed-length descriptors leverage deep learning frameworks to extract more discriminative features. In alignment-based methods, fingerprints are initially aligned, followed by the extraction of fixed-length descriptors from these aligned images. Cao et al. [18] introduced the extraction of deep features from aligned fingerprints, while Engelsma et al. [19] developed an end-to-end network integrating alignment and feature

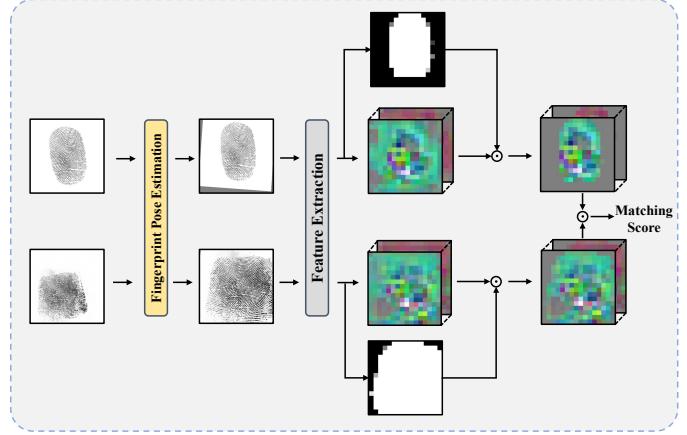


Fig. 2. Overview of fingerprint matching using the proposed FDD. First, fingerprint pose alignment is performed according to Duan et al. [22], followed by the extraction of fixed-length dense descriptors with the descriptor extraction network, which are then used for matching.

extraction. Grosz et al. [20] improved recognition by fusing attention-based and CNN-based embeddings. These techniques excel with large, complete fingerprint areas, like rolled prints, but may be compromised by background noise in incomplete prints, affecting matching accuracy. Song and Feng [3] and Gu et al. [21] proposed using pyramid features for global and local representations to address incomplete fingerprints, but this approach still has issues with region overlap and requires dataset-specific optimizations. Some researchers [23], [24] boost accuracy through concurrent alignment and scoring of two images, which, however, increases computational demands due to the dual-image input costs.

Alignment-free fingerprint descriptors typically utilize minutiae data. Song et al. [25] and Wu et al. [26] aggregate local features around minutiae into global descriptors, with the latter using multi-task learning. Li et al. [27] integrate minutiae via a fully convolutional network and global average pooling. However, these methods may be affected by minutiae extraction variability, and the loss of minutiae count and positions post-aggregation diminishes feature discriminability.

Fixed-length Dense Descriptor proposed in this paper is alignment-based. To alleviate the negative impact of inaccurate fingerprint alignment, a fingerprint pose estimation algorithm, which achieves state-of-the-art performance across various fingerprint datasets [22], was employed to align fingerprints.

III. METHOD

We propose the Fixed-length Dense Descriptor (FDD) for fingerprint matching with efficiency and interpretability. Fingerprints are first aligned using the estimated fingerprint poses [22]. Subsequently, localized representations are extracted, and the segmentation mask is determined simultaneously. The final FDD is obtained by element-wise multiplying the local representations with the predicted mask. And the matching score is computed based on the cosine similarity of two normalized flattened FDD. The overall schematic illustration of fingerprint matching using FDD is shown in Fig. 2.

A. Network Design

We draw upon the method of extracting Dense Minutiae Descriptor [6], employing ResNet-34 [28] as the backbone of our network, and modify it to a dual-branch structure. To retain more details of ridges, we remove the first max pooling layer, resulting in an output size that is 1/16 of the input size. Considering the compromise between computational efficiency and performance, we cropped the aligned fingerprint with 500 ppi to 512×512 and subsequently scaled it down to 256×256 as input for the network. Therefore, the spatial resolution of FDD in our case is 16×16 . The overall network structure is shown in Fig. 3.

Except for global semantic information, local positional information is also essential for incorporating global constraints due to the alignment among the fingerprint spaces. We utilize the classic 2D sinusoidal positional embedding module [29] to enable the network focus on distinct characteristics across different local regions, resulting in a more refined feature representation.

We design a dual-branch structure consisting of a minutiae branch and a texture branch. Each branch follows a multi-task approach, incorporating dense descriptor extraction along with a corresponding auxiliary task. Minutiae branch outputs the minutiae distribution in a heatmap form called minutiae map [19] denoted as M . The extraction of the minutiae map $M \in \mathbb{R}^{6 \times 128 \times 128}$ and minutiae descriptor $f_m \in \mathbb{R}^{C \times 16 \times 16}$ share features from Minutiae Feature Encoder. Similarly, for the texture branch, mask $h \in \mathbb{R}^{16 \times 16}$ prediction and texture descriptor $f_t \in \mathbb{R}^{C \times 16 \times 16}$ deriving also share the same features. The final Fixed-length Dense Descriptor $f \in \mathbb{R}^{2C \times 16 \times 16}$ is obtained by concatenating two branch's descriptors and multiplying the mask as

$$f = (f_t \oplus f_m) \odot h, \quad (1)$$

where \oplus denotes the concatenation and \odot denotes the Hadamard product.

B. Fingerprint Matching

Given a pair of fingerprints (p, g) , the extracted FDDs by Eq. (1) are denoted as $f^q \in \mathbb{R}^{2C \times 16 \times 16}$ and $f^g \in \mathbb{R}^{2C \times 16 \times 16}$, and the estimated mask area as $h^q \in \mathbb{R}^{1 \times 16 \times 16}$ and $h^g \in \mathbb{R}^{1 \times 16 \times 16}$. The matching score is obtained by calculating the cosine similarity between the two flattened dense descriptors $f^{q'}$ and $f^{g'}$. Only the feature embeddings within the overlapping area are considered during comparison. Specifically, the matching score of fingerprint pair (p, g) is computed by

$$s(q, g) = \frac{1}{2} \cdot \frac{f^{q'}^T f^{g'}}{\|f^q \odot h^g\|_F \|f^g \odot h^q\|_F} + \frac{1}{2}, \quad (2)$$

where $f^{q'} \in \mathbb{R}^{512C}$, $f^{g'} \in \mathbb{R}^{512C}$, and $\|\cdot\|_F$ represents Frobenius norm.

C. Objective Functions

The proposed network aims to extract robust and discriminative representations, and the following provides detailed descriptions of the objective functions.

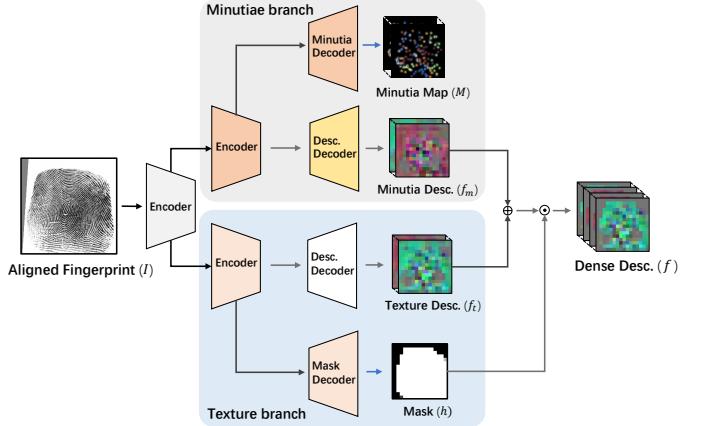


Fig. 3. The overall network architecture of FDD consists of the minutiae branch and the texture branch, which extract minutiae maps and masks, respectively, along with their corresponding descriptors. “Desc.” is an abbreviation for “Descriptor”.

1) Classification Loss: Inspired by face recognition deep models, we utilize the CosFace loss [30] to supervise the descriptors' extraction. Specifically, the extracted dense descriptors are first flatten into one dimension and followed by a Fully Connected (FC) layer with K classes. The CosFace loss function is modified based on SoftMax

$$\mathcal{L}_{cls}^i = -\frac{1}{N} \sum_{n=1}^N \log \frac{e^{A(\cos(\theta_{y_n}^i) - b)}}{e^{A(\cos(\theta_{y_n}^i) - b)} + \sum_{k=1, k \neq y_n}^K e^{A \cos(\theta_k)}}, \quad (3)$$

where $\cos(\theta_k)$ is calculated by

$$\cos(\theta_k^i) = W_k^T f_i, \quad W_k = \frac{W_k}{\|W_k\|}, \quad f_i = \frac{f_i}{\|f_i\|}. \quad (4)$$

The hyper-parameter A is utilized for magnitude normalization and $i \in [t, m]$ denotes the descriptor's type, b denotes the margin, N is the number of samples within a batch, and y_n is the class label of sample f_i^n . The W_k represents the learnable weight vector of the k -th class.

2) Auxiliary Loss: The auxiliary tasks contain the minutiae map extraction and mask prediction, therefore the auxiliary loss includes the binary cross entropy loss for calculating \mathcal{L}_{mask} and mean square error for calculating minutiae heatmap loss \mathcal{L}_{mnt} . We utilize the VeriFinger v12.0¹ to extract the target minutiae and masks.

3) Local Similarity Loss: Given that \mathcal{L}_{cls} considers all elements of the extracted localized representations, the features from incomplete fingerprints cannot be directly optimized. Consequently, we proposed a “teaching-mimicking” tactic, wherein features extracted from incomplete fingerprints mimic those found in complete fingerprints to uphold consistency in representations across fingerprints of various visible areas, intensity distributions, and distortion patterns originating from the same finger. We simulate the incomplete fingerprints by

¹<https://neurotechnology.com/verifinger.html>

adding masks on the complete ones. Specifically, the loss function is defined by

$$\mathcal{L}_{\text{sim}} = \frac{1}{|h_{q \cap g}|} \sum_{(i,j) \in h_{q \cap g}} \|f^q(i,j) - f^g(i,j)\|^2, \quad (5)$$

where f_q and f_g denote the representations extracted from incomplete and complete fingerprints respectively.

Consequently, the proposed network was supervised by optimizing the overall objective function

$$\mathcal{L} = \sum_{i \in \{t,m\}} \mathcal{L}_{\text{cls}}^i + \lambda_{\text{mask}} \mathcal{L}_{\text{mask}} + \lambda_{\text{minu}} \mathcal{L}_{\text{minu}} + \lambda_{\text{sim}} \mathcal{L}_{\text{sim}}, \quad (6)$$

where λ_{mask} , λ_{minu} , and λ_{sim} are trade-off parameters for L_{mask} , L_{minu} , and L_{sim} respectively. In this paper, we set $\lambda_{\text{mask}} = 1$, $\lambda_{\text{minu}} = 0.01$, and $\lambda_{\text{sim}} = 0.00125$ to achieve roughly the same decreasing rate for those loss functions.

D. Implementation Details

We designed a data augmentation strategy to simulate multiple impressions from the same finger, thus increasing the diversity of training data to avoid over-fitting. Specifically, we utilized the fingerprint distortion model in Si et al. [31] and randomly generated different distortion fields, which were then applied on real fingerprints to simulate impressions with different distortion patterns. Histogram matching of intensity was then employed to adjust the intensity distribution of synthetic fingerprints with arbitrary plausible distributions, which aims to increase the variations of intensity distribution in synthetic fingerprints. Moreover, additional data augmentation was also adopted, including random translation in range of 10 pixels and random rotation to follow a uniform distribution $[-5^\circ, 5^\circ]$. The feature dimension C of our Fixed-length Dense Descriptor is set to 6, therefore $f \in \mathbb{R}^{12 \times 16 \times 16}$. The normalization parameter A and margin b in Eq. (3) are set to 30 and 0.4, respectively. AdamW optimizer with initial learning rate of 0.00035 was utilized for network training. Learning rate decays by 0.1 after 10 epochs without performance improvement on validation dataset. And early stopping was performed after three times learning rate decays. Furthermore, L2 regularization on trainable parameters was also utilized to avoid over-fitting.

IV. EXPERIMENTS

In this section, we first introduce the datasets utilized in this paper and the comparing methods, then the evaluation of matching performance on these datasets.

A. Datasets

Multiple datasets with various impression types, including rolled, plain, latent, and contactless fingerprints, are utilized for evaluation. Table I provides the details of all datasets and several fingerprint examples from these datasets are shown in Fig. 4.

We utilize the first 24,000 pairs of rolled fingerprints from NIST SD14 for training. Note that no fine-tuning is applied on any other dataset. Diverse Pose Fingerprint dataset (DPF)

TABLE I
FINGERPRINT DATASETS USED IN EXPERIMENTS.

Type	Dataset	Sensor	Quantity	Usage
Rolled	NIST SD4	Inking	4,000	test
	NIST SD14	Inking	48,000	train
Plain	FVC2002 DB3A	Capacitive	800	test
	FVC2004 DB1A	Optical	800	test
	FVC2006 DB1A	Electric field	1,680	test
	N2N Plain ^a	Optical	4,000	test
	DPF	Optical	1,287	test
Latent	NIST SD27 ^b	-	516	test
Contactless	PolyU CL2CB ^c	Optical/Camera	2,016	test

^a Subset R and S in NIST SD302b.

^b 10,458 fingerprints (plain or rolled) from our internal dataset THU Contact10K were added as gallery.

^c Original contactless fingerprints are scaled to match the same mean ridge period of contact-based fingerprints [32], and no warping was performed.

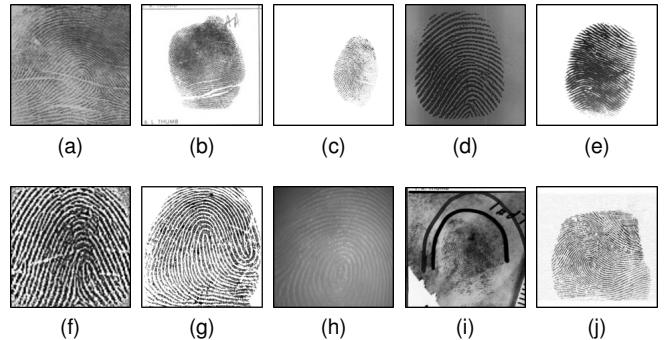


Fig. 4. Image examples from different fingerprint datasets (a) NIST SD4, (b) NIST SD14, (c) DPF, (d) FVC2002 DB3A, (e) FVC2004 DB1A, (f) FVC2006 DB1A, (g) N2N Plain, (h) PolyU CL2CB, (i) NIST SD27, (j) THU Contact10K. Images are input at 500 ppi, and they have been rescaled here to ensure optimal visualization.

[22], which is collected from fingers with various 3D poses, consists of 143 rolled fingerprints, each accompanied by 8 plain fingerprints of various poses. For the FVC series dataset, we selected three of the most challenging ones: FVC2002 DB3A, FVC2004 DB1A, and FVC2006 DB1A.

B. Compared Methods

We compared our method with previous high-performing fixed-length fingerprint descriptors, including DeepPrint [19], AFRNet [20], and MultiScale [21]. Since the code of DeepPrint and AFRNet is not available to us, we reimplemented them using the same training data to comprehensively evaluate their performance on multiple datasets. Specifically, we excluded the feature realignment strategy in AFRNet because its usage significantly increases computational complexity, thereby losing the significance of fixed-length representation. MultiScale [21] and our method employed identical pose estimation results obtained from [22]. Additionally, DeepPrint and AFRNet were evaluated using their original STN modules and the pose estimation method from [22], respectively. This ensures consistency across all compared methods, emphasizing that the observed matching performance solely reflects the efficacy of the fingerprint representations. In addition, we evaluated two state-of-the-art minutiae-based descriptor methods, VeriFinger and DMD [6], for comparison.

TABLE II
VERIFICATION AND RECOGNITION ACCURACY (%) ON SEVERAL FINGERPRINT DATASETS. TAR@FAR=0.1% IS REPORTED.

Method ¹	Type	NIST SD4		FVC02 DB3	FVC04 DB1	FVC06 DB1	N2N Plain		DPF	PolyU	NIST SD27	
		TAR ²	Rank-1	TAR			TAR ²	Rank-1	TAR		TAR	Rank-1
VeriFinger v12.0	Minutia-based	99.80	99.65	99.43	98.93	92.65	99.30	99.20	98.51	97.21	53.10	59.69
DMD [6]	Minutia-based	99.80	99.80	99.36	98.96	93.89	99.15	99.35	98.78	90.43	81.78	79.84
DeepPrint [19] ³	Fixed-length	96.25	98.60	70.96	94.82	83.02	93.35	96.00	64.86	56.05	34.88	25.58
V-DeepPrint [19] ⁴	Fixed-length	96.55	99.00	69.85	94.00	82.39	95.00	97.15	67.05	54.20	39.53	28.29
AFRNet [20] ⁵	Fixed-length	94.75	97.40	80.25	98.14	83.31	96.90	97.40	77.36	52.61	45.30	35.66
V-AFRNet [20] ⁶	Fixed-length	97.90	99.30	49.82	87.43	59.49	95.70	97.15	77.45	10.69	36.05	25.58
MultiScale [21]	Fixed-length	97.80	98.85	83.36	97.36	47.37	97.80	98.10	87.68	64.13	44.57	43.80
FDD (binary)	Fixed-length	99.50	99.60	95.50	98.71	86.99	98.35	98.30	88.90	81.21	43.80	44.19
FDD	Fixed-length	99.60	99.55	96.71	98.79	90.87	98.50	98.40	91.61	86.62	53.49	50.00
Fusion ⁷	–	99.85	99.80	99.39	99.36	94.33	99.30	99.35	98.78	93.62	82.95	82.56

¹ All methods were trained using the same fingerprint data.

² TAR@FAR=0.01%

³ Performance of reimplemented DeepPrint.

⁴ Performance of DeepPrint reimplemented by replacing STN with the pose estimation method in [22].

⁵ Performance of reimplemented AFRNet.

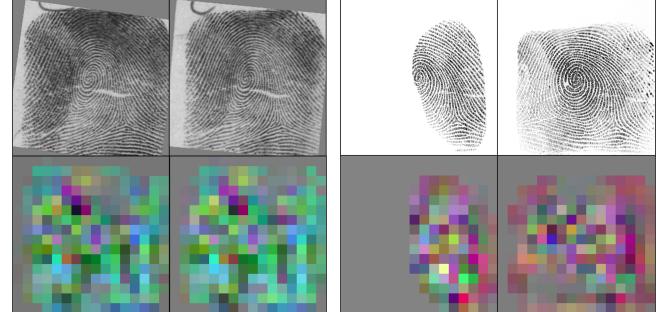
⁶ Performance of AFRNet reimplemented by replacing STN with the pose estimation method in [22].

⁷ The score-level fusion of FDD and DMD employs fusion weights learned from our other private fingerprint datasets.

TABLE III

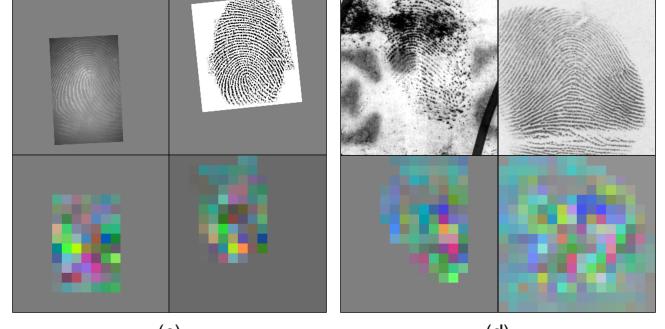
EVALUATION OF FDD f WITH DIFFERENT FEATURE DIMENSION SIZE C ON SEVERAL CHALLENGING DATASETS. AND $f \in \mathbb{R}^{2C \times 16 \times 16}$.

# Dimension	TAR@FAR=0.1% (%)				
	FVC2002 DB3A	FVC2006 DB1A	DPF	PolyU CL2CB	NIST SD27
$C = 1$	94.79	81.37	90.82	86.14	41.47
$C = 3$	96.00	88.59	90.82	89.15	51.16
$C = 6$	96.71	90.87	91.61	86.62	53.49
$C = 9$	94.56	90.50	90.91	86.40	56.98
$C = 12$	96.64	89.31	91.60	87.44	59.69



(a)

(b)



(c)

(d)

TABLE IV
EFFICIENCY (SECONDS PER SAMPLE OR PAIR) OF DIFFERENT METHODS.

	Pose Estimation	Descriptor Extraction	Descriptor Matching
VeriFinger	–	1.98	5.48×10^{-3}
DMD	–	4.60×10^{-2}	1.25×10^{-3}
DeepPrint	–	1.47×10^{-3}	5.12×10^{-9}
AFRNet	–	1.15×10^{-3}	2.14×10^{-8}
MultiScale	2.81×10^{-3}	3.18×10^{-3}	2.63×10^{-7}
FDD(binary)	2.81×10^{-3}	6.13×10^{-4}	5.16×10^{-8}
FDD	2.81×10^{-3}	6.13×10^{-4}	6.53×10^{-8}

* The efficiency results are calculated based on NIST SD4.

C. Results

As shown in Tab. II, FDD demonstrates superior performance compared to the other fixed-length representation methods on all evaluated datasets, particularly on incomplete or partial fingerprints, contactless fingerprints, and latent fingerprints. Specifically, our approach outperforms the previous best performance by 16.01% on FVC2002 DB3A, 9.07% on FVC2006 DB1A, 35.07% on PolyU CL2CB, and 18.07% on NIST SD27 in terms of TAR@FAR=0.1%. It clearly highlights the advantages of dense descriptors in matching partial fingerprints, isolating background noise, and facilitating cross-modal fingerprint matching. Additionally, we binarize the descriptor features with a threshold of 0 to obtain a binary version of FDD, and the similarity score are calculated

Fig. 5. Examples of the FDD extracted from genuine pairs of (a) NIST SD4, (b) DPF, (c) PolyU CL2CB, and (d) NIST SD27. The fingerprint images shown in the figure have been aligned based on their estimated poses.

using XOR. It still exhibits great matching performance. Although the matching performance of FDD currently falls short of minutiae-based methods, score fusion between FDD and minutiae-based representation like DMD [6] yields superior results compared to either method alone, highlighting their complementarity. Furthermore, as shown in Tab. IV, FDD significantly outperforms minutiae-based methods in efficiency.

Fig. 5 shows genuine fingerprint pairs with their corresponding dense descriptors (three channels displayed). For high-quality rolled fingerprints (Fig. 5a), the dense features are highly consistent. For partial fingerprints (Fig. 5b), the features

at corresponding positions remain correlated. In contactless-to-contact fingerprint matching (Fig. 5c), the descriptors exhibit cross-modal robustness. For latent-to-rolled fingerprint matching fingerprints (Fig. 5d), the descriptors effectively reduce background noise, and have high consistency in the overlapping regions.

Furthermore, we conduct matching experiments with different representation dimensions C for FDD ($f \in \mathbb{R}^{2C \times 16 \times 16}$). In Tab. III, the overall trend shows that matching accuracy decreases as the feature dimension is reduced. Despite this, the low-dimensional versions of FDD still exhibit good matching performance, even surpassing other fixed-length methods on some datasets.

V. CONCLUSION

In this paper, we propose a fixed-length representation called Fixed-length Dense Descriptor (FDD), which enables efficient and interpretable fingerprint matching in large-scale datasets. FDD is a three-dimensional representation where the descriptor space coincides with the fingerprint space. This ensures that after aligning fingerprints using pose estimation, descriptor-level matching occurs only within overlapping foreground. To enhance representational capability, we design a dual-branch network structure to capture texture-related and minutiae-related descriptors. Experiments on various fingerprint datasets show that FDD outperforms other fixed-length descriptors and its binary and low dimensional versions still achieve good matching accuracy. However, the current extraction of dense descriptors exhibits a certain dependency on fingerprint pose accuracy. Future research will aim to develop pose-independent descriptors to enhance matching performance.

REFERENCES

- [1] D. Maltoni, D. Maio, A. K. Jain, and J. Feng, *Handbook of Fingerprint Recognition*, 3rd ed. Springer Nature, Cham, Switzerland, 2022.
- [2] 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*, vol. 32, no. 12, pp. 2128–2141, 2010.
- [3] D. Song and J. Feng, “Fingerprint indexing based on pyramid deep convolutional feature,” in *2017 IEEE International Joint Conference on Biometrics*. IEEE, 2017, pp. 200–207.
- [4] K. Cao and A. K. Jain, “Automated latent fingerprint recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 4, pp. 788–800, 2019.
- [5] S. A. Grosz and A. K. Jain, “Latent fingerprint recognition: Fusion of local and global embeddings,” *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 5691–5705, 2023.
- [6] Z. Pan, Y. Duan, X. Guan, J. Feng, and J. Zhou, “Latent fingerprint matching via dense minutia descriptor,” *arXiv preprint arXiv:2405.01199*, 2024.
- [7] S. Rane, Y. Wang, S. C. Draper, and P. Ishwar, “Secure biometrics: Concepts, authentication architectures, and challenges,” *IEEE Signal Processing Magazine*, vol. 30, no. 5, pp. 51–64, 2013.
- [8] A. Sankaran, P. Pandey, M. Vatsa, and R. Singh, “On latent fingerprint minutiae extraction using stacked denoising sparse autoencoders,” in *2014 IEEE International Joint Conference on Biometrics*, pp. 1–7.
- [9] Y. Tang, F. Gao, J. Feng, and Y. Liu, “Fingernet: An unified deep network for fingerprint minutiae extraction,” in *2017 IEEE International Joint Conference on Biometrics*. IEEE, 2017, pp. 108–116.
- [10] A. K. Jain, S. Prabhakar, L. Hong, and S. Pankanti, “FingerCode: A filterbank for fingerprint representation and matching,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, vol. 2. IEEE, 1999, pp. 187–193.
- [11] Y. Wang, J. Hu, and D. Phillips, “A fingerprint orientation model based on 2D Fourier expansion (FOMFE) and its application to singular-point detection and fingerprint indexing,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 573–585, 2007.
- [12] A. Kumar and Y. Zhou, “Contactless fingerprint identification using level zero features,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition WorkShop*, 2011, pp. 114–119.
- [13] J. Bringer and V. Despiegel, “Binary feature vector fingerprint representation from minutiae vicinities,” in *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems*, pp. 1–6.
- [14] F. Farooq, R. M. Bolle, T.-Y. Jea, and N. Ratha, “Anonymous and revocable fingerprint recognition,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–7.
- [15] Z. Jin, A. B. J. Teoh, T. S. Ong, and C. Tee, “A revocable fingerprint template for security and privacy preserving,” *KSII Transactions on Internet and Information Systems*, vol. 4, no. 6, pp. 1327–1342, 2010.
- [16] R. Cappelli, “Fast and accurate fingerprint indexing based on ridge orientation and frequency,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, no. 6, pp. 1511–1521, 2011.
- [17] Y. Imamverdiyev, A. B. J. Teoh, and J. Kim, “Biometric cryptosystem based on discretized fingerprint texture descriptors,” *Expert Systems with Applications*, vol. 40, no. 5, pp. 1888–1901, 2013.
- [18] K. Cao and A. K. Jain, “Fingerprint indexing and matching: An integrated approach,” in *2017 IEEE International Joint Conference on Biometrics*, pp. 437–445.
- [19] J. J. Engelsma, K. Cao, and A. K. Jain, “Learning a fixed-length fingerprint representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 6, pp. 1981–1997, 2021.
- [20] S. A. Grosz and A. K. Jain, “Afr-net: Attention-driven fingerprint recognition network,” *IEEE Transactions on Biometrics, Behavior, and Identity Science*, pp. 1–1, 2023.
- [21] S. Gu, J. Feng, J. Lu, and J. Zhou, “Latent fingerprint indexing: Robust representation and adaptive candidate list,” *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 908–923, 2022.
- [22] Y. Duan, J. Feng, J. Lu, and J. Zhou, “Estimating fingerprint pose via dense voting,” *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 2493–2507, 2023.
- [23] Z. He, J. Zhang, L. Pang, and E. Liu, “PFVNet: A partial fingerprint verification network learned from large fingerprint matching,” *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 3706–3719, 2022.
- [24] Y. Qiu, H. Chen, X. Dong, Z. Lin, I. Y. Liao, M. Tistarelli, and Z. Jin, “IFViT: Interpretable fixed-length representation for fingerprint matching via vision transformer,” *arXiv preprint arXiv:2404.08237*, 2024.
- [25] D. Song, Y. Tang, and J. Feng, “Aggregating minutia-centred deep convolutional features for fingerprint indexing,” *Pattern Recognition*, vol. 88, pp. 397–408, 2019.
- [26] S. Wu, B. Liu, Z. Wang, Z. Jia, and J. Feng, “Minutiae-aware learning fingerprint representation for fingerprint indexing,” in *2022 IEEE International Joint Conference on Biometrics*, pp. 1–8.
- [27] R. Li, D. Song, Y. Liu, and J. Feng, “Learning global fingerprint features by training a fully convolutional network with local patches,” in *2019 International Conference on Biometrics*. IEEE, 2019, pp. 1–8.
- [28] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [29] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [30] F. Wang, X. Cao, X. Ren, and P. Irani, “Detecting and leveraging finger orientation for interaction with direct-touch surfaces,” in *Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology*, 2009, pp. 23–32.
- [31] X. Si, J. Feng, J. Zhou, and Y. Luo, “Detection and rectification of distorted fingerprints,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 555–568, 2015.
- [32] Z. Cui, J. Feng, and J. Zhou, “Monocular 3D fingerprint reconstruction and unwarping,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 7, pp. 8679–8695, 2023.

Supplementary Material

TABLE V
DETAILED NETWORK STRUCTURE OF FDD EXTRACTION NETWORK

Encoder			
Type	Output Size (spatial scales, channels)		# of Layer
7 × 7 Conv. Layer	/2, 64		1
3 × 3 Residue Block	1, 64		3
3 × 3 Residue Block	/2, 128		4
Encoder in Minutia\Texture Branch			
Type	Output Size (spatial scales, channels)		# of Layer
3 × 3 Residue Block ^a	/2, 256		6
3 × 3 Residue Block ^b	/2, 512		3
Minutia Decoder			
Type	Output Size (spatial scales, channels)		# of Layer
3 × 3 Conv. Layer	1, 128		6
4 × 4 Deconv. Layer	2, 64		2
3 × 3 Conv. Layer	1, 6		1
Mask Decoder			
Type	Output Size (spatial scales, channels)		# of Layer
3 × 3 Conv. Layer	1, 512		1
1 × 1 Conv. Layer	1, 1		1
Descriptor Decoder in Minutia\Texture Branch			
Type	Output Size (spatial scales, channels)		# of Layer
3 × 3 Conv. Layer	1, 512		1
1 × 1 Conv. Layer	1, 6		1

^a Minutia Map Decoder connects from here.

^b The branches for Mask Decoder and Minutia\Texture Descriptor Decoder connect from here.