IEICE TRANSACTIONS

on Fundamentals of Electronics, Communications and Computer Sciences

DOI:10.1587/transfun.2024EAP1043

Publicized:2024/04/05

This advance publication article will be replaced by the finalized version after proofreading.

A PUBLICATION OF THE ENGINEERING SCIENCES SOCIETY
The Institute of Electronics, Information and Communication Engineers
Kikai-Shinko-Kaikan Bldg., 5-8, Shibakoen 3 chome, Minato-ku, TOKYO, 105-0011 JAPAN

PAPER

Deep Learning-inspired Automatic Minutiae Extraction From Semi-automated Annotations

Hongtian ZHAO[†a)], *Member*, Hua YANG[††], *and* Shibao ZHENG[††], *Nonmembers*

SUMMARY Minutiae pattern extraction plays a crucial role in fingerprint registration and identification for electronic applications. However, the extraction accuracy is seriously compromised by the presence of contaminated ridge lines and complex background scenarios. General image processing-based methods, which rely on many prior hypotheses, fail to effectively handle minutiae extraction in complex scenarios. Previous works have shown that CNN-based methods can perform well in object detection tasks. However, the deep neural networks (DNNs)-based methods are restricted by the limitation of public labeled datasets due to legitimate privacy concerns. To address these challenges comprehensively, this paper presents a fully automated minutiae extraction method leveraging DNNs. Firstly, we create a fingerprint minutiae dataset using a semi-automated minutiae annotation algorithm. Subsequently, we propose a minutiae extraction model based on Residual Networks (Resnet) that enables end-to-end prediction of minutiae. Moreover, we introduce a novel non-maximal suppression (NMS) procedure, guided by the Generalized Intersection over Union (GIoU) metric, during the inference phase to effectively handle outliers. Experimental evaluations conducted on the NIST SD4 and FVC 2004 databases demonstrate the superiority of the proposed method over existing state-of-the-art minutiae extraction approaches.

key words: *Minutiae extraction, Fingerprint morphology processing, Resnet, GIoU-oriented NMS.*

1. Introduction

Despite the diverse representative features present in fingerprints, including grayscale maps, gradient fields, orientation fields, and orientation consistency, the majority of real-world recognition systems primarily depend on minutiae [1, 6]. Minutiae patterns generally consist of ridge endings and ridge bifurcations [7]. A ridge ending represents the start or end point of a ridgeline, while a bifurcation denotes the merging point of two ridgelines into one. Extensive theoretical proof and statistical analysis demonstrate that these two types of minutiae can effectively identify a fingerprint [18–20]. Therefore, the accurate and comprehensive extraction of minutiae serves as a fundamental problem in this field.

The problem attracts significant attention as it is crucial for various automated fingerprint applications, including ecommerce, phone unlock, crime identification, and intelligent security [2–5]. Minutiae extraction is a complex pattern recognition problem due to challenges posed by polluted areas and background noises, and there is still a lack of well-solved formulations and optimizations. To address this problem, researchers have proposed different approaches. For example, previous algorithms based on ridge tracing have been used to restore ridges on thinned fingerprint images before extracting minutiae. However, this approach is computationally intensive and involves trivial optimization processes [8]. Recent studies by Tang *et al.*[9] and Nguyen *et al.*[10] have explored using local shape structures and texture information for minutiae extraction, including position coordinates and ridge orientation angles. However, these techniques still have limitations, such as generating false minutiae or missing genuine ones. Traditional methods rely on artificial approximations or empirical fingerprint morphology processing [11–13, 30], which are often not flexible and struggle to handle complex fingerprints with noise or contamination. Conventional minutiae extraction methods, which are based on hand-designed or empirical approaches, are insufficient in accurately detecting minutiae in perturbed areas due to their limited adaptability and inability to handle various disturbances. This leads to information loss and errors. Furthermore, the complexity and diversity of perturbed fingerprint regions make it challenging for traditional approaches to address most cases. In summary, there are still numerous challenges in precisely formulating the problem of degraded fingerprints.

Significant progress has been made in simulated problem solver through the integration of domain knowledge with deep neural networks (DNNs). In complex scenarios, DNNs such as VGGNet [31], InceptionNet [32], MobileNet [33], and EfficientNet [34] demonstrate superior performance over handcrafted features by leveraging their varied hierarchical representation, adaptability, and non-linear processing characteristics to learn generic features. In the domain of fingerprint analysis, the adoption of an end-to-end inference paradigm by prevalent DNNs-based approaches facilitates efficient minutiae extraction, circumventing the need for iterative optimization strategies commonly associated with traditional methods. However, the drawback of DNNs is their reliance on well-defined training data. Standard fingerprint datasets, such as NIST SD27 [21], are no longer available in world wide web in contemporary time due to privacy security policies, which hinders the development of DNNs-based minutiae extraction. Although some approaches [9, 10, 14] have been proposed to handle this task, they have exhibited poor performance in detecting minutiae patterns in challenging fingerprints. Thus, the extraction of minutiae faces significant challenges in real-life scenarios.

To construct a comprehensive set of minutiae for train-

† The author is with College of Mathematics and System Science, Xinjiang University, 830046, Urumqi, China.
†† The author is with SEIEE of SJTU, 200240, Shanghai, China.
a) E-mail: zhaohongtian@xju.edu.cn

ing DNNs, we study and propose a complete fingerprint minutiae annotation pipeline. The extraction module consists of several crucial steps, including image normalization, segmentation, orientation and frequency estimation, enhancement, binarization, and thinning. Subsequently, the pipeline extracts minutiae points from the fingerprint skeleton, which retains only vital topological structures (e.g., the basic structure of the ridge) while removing redundant information in the image. To ensure reliable labeling, we incorporate a manual revision process as post-processing due to the underlying assumptions of the annotation approaches. Our minutiae annotation method is a semi-automatic technique that involves controlled user interactions. In comparison to manual labeling, our method offers ease of operation, time savings, reduces the workload for experts, and enhances human-machine cooperation. Importantly, we introduce an automatic minutiae extraction framework to enhance the effectiveness and robustness of minutiae extraction. In consideration of limited fingerprint databases, we first present a semi-automated labeled minutiae dataset. To simulate real-life fingerprint recognition scenarios, we develop an automatic minutiae extraction system based on DNNs for efficient prediction tasks, allowing for easy deployment and usability in authentic fingerprint application scenarios.

This work is an extension of our previous minutiae annotation algorithm [15], with several significant improvements: (1) We propose a ResNet-based neural network for automatic minutiae extraction. (2) To enhance the quality of detection, we introduce a novel generalized IoU (GIoU)-oriented NMS filter to correct falsely extracted minutiae. (3) Extensive validation experiments, along with discussions, demonstrate the effectiveness of the presented dataset and automatic minutiae extraction system. In summary, the main contributions of our method are as follows:

- To address the lack of available minutiae datasets, we propose a semi-automated annotation method for fingerprint minutiae. This method integrates automatic extraction and manual revision steps to ensure comprehensive and reliable training annotations. Based on it, we establish a dependable minutiae dataset that incorporates the expertise of human annotators.
- To adaptively detect fingerprint feature patterns, we propose a novel minutiae extraction model based on ResNet-based neural networks. This model simulates the fingerprint processing and minutiae extraction procedure, including orientation field estimation, fingerprint segmentation and minutiae extraction. Additionally, we introduce a GIoU-oriented NMS filter to enhance the quality of minutiae detection.
- Comprehensive experiments with analysis, discussions and comparisons verify the effectiveness of both the proposed dataset and the prediction method.

2. Dataset Construction

In recent years, on-site fingerprint identification technology has advanced with the help of the NIST SD27 public fingerprint database. However, this database is no longer publicly available due to permission restrictions. Synthetic datasets from the Fingerprint Verification Competition (FVC) series [23], such as FVC2004 DB4, are limited in both scale and realism compared to real-world scenarios. The NIST SD04 dataset, composed of authentic fingerprint images, captures the nuanced details of fingerprint features such as skin texture and pores, which are essential for reliable experimental results. It offers a more challenging test-bed with diverse image quality, noise, and occlusions, thereby better evaluating model robustness and accuracy. In contrast, FVC2004 may oversimplify real-world complexities, potentially compromising recognition performance in practical applications. The NIST SD04 dataset's rich local and global features, critical for precise fingerprint recognition, are less effectively simulated in synthetic datasets, leading to a performance gap. Therefore, this study employs the NIST SD04 dataset for training and validating deep learning models, substantiating the superiority of semi-automatic, human-supervised annotation for enhancing accuracy. Additionally, manual minutiae extraction, known to boost performance in latent fingerprint images [16, 17], underscores the importance of human oversight in data annotation. To address this, we propose a novel and versatile minutiae dataset for investigating minutiae extraction. We describe the dataset in Appendix A and its implementation of a semi-automatic human-computer interaction labeling algorithm in the rest of this section.

Minutiae Database Annotation Workflow To ensure accurate and comprehensive minutiae annotation, the proposed algorithm involves two main steps: automated minutiae extraction and manual revision. Fig. 1 provides an overview of the complete workflow, while the revision step is illustrated in Fig. 2. The automated extraction process includes various stages, such as image segmentation, normalization, orientation estimation, frequency estimation, enhancement, binarization, thinning, minutiae extraction, and removal of pseudo minutiae points. The output results from automatic extraction are then carefully revised and checked to obtain the final ground truth.

Fig. 3 illustrates the intermediate outputs of the key steps in the aforementioned fingerprint minutiae annotation workflow. The algorithm for annotating the fingerprint minutiae database has been comprehensively explained in [15]. Therefore, this paper does not repeat the annotation algorithm in this module; instead, it directs readers to [15] for a comprehensive understanding of the algorithm.

3. Method

Deep learning methods in image recognition tasks depend on three key factors: data, algorithm, and model. Section 2 have introduced the created dataset, which includes fingerprint images and corresponding minutia labels with coordinates and orientations. We then provide an overview of our ResNet-based model and the proposed inference method for automated latent fingerprint minutiae extraction. Our

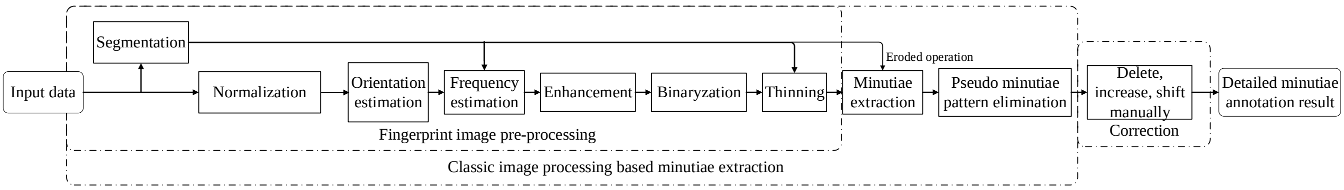
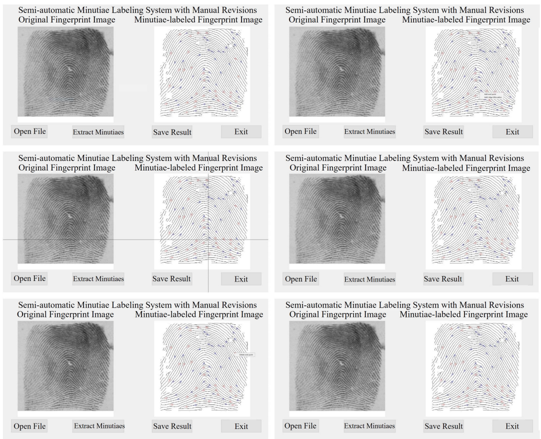


Fig. 1: Minutiae annotation workflow: automated extraction and manual correction.



The revision operations snapshots.

Fig. 2: Minutiae annotation refinement: process illustrations for addition and deletion of feature points.

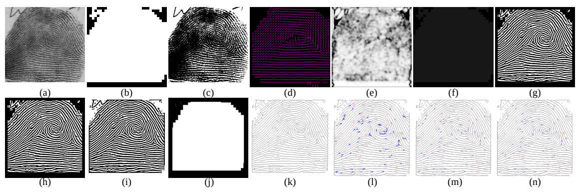


Fig. 3: Sample results of the presented annotation method, which are mainly corresponding to the output results of interim processes in Fig. 1: (a) Original image in NIST SD4; (b) Segmentation mask; (c) Normalized image; (d) Orientation distribution visualization image; (e) Reliability map; (f) Frequency map; (g) Enhanced image; (h) Binarized image; (i) Mask-operated binarized image; (j) Corroded segmented mask; (k) Refined skeleton image; (l) Refined skeleton image with extracted minutiae; (m) The skeleton image with extracted minutiae after post-processing; (n) The output image obtained by using manual editing from image (m).

focus is on accurately recognizing ending points and bifurcation points using a DNNs-based approach. While previous studies have integrated domain knowledge with CNN representation capabilities, such as FingerNet by Tang et al. [9], our observations and evaluations indicate limitations and instability in their minutiae extraction. To improve detection accuracy, we propose two optimization strategies: intrinsic feature extraction using ResNet-based structures and a GIoU-inspired NMS filter. The key implementation details of our method will be described in the following section.

3.1 Basic FingerNet-oriented Neural Network

FingerNet, as introduced in [9], is an innovative approach that combines domain knowledge and CNN's feature representation abilities to simplify minutiae detection. To simulate the classical minutiae extraction process in real-life applications, we refine FingerNet [9] and adopt it as the backbone network. Based on the residual structure's prominent fitting capability, we propose an enhanced network for fingerprint minutiae extraction, enabling the comprehensive utilization of morphological knowledge for learning effective features.

Fig. 4 shows the fundamental DNNs-based procedure, which encompasses common tasks such as image normalization, orientation estimation, segmentation, gabor enhancement, and minutiae extraction. Specifically, the input image undergoes initial normalization. Subsequently, the normalized image is directed into two pipelines. The first pipeline calculates gradients for orientation estimation and segmentation, while the second pipeline employs Gabor filters to compute group filters and shift the operation space from spatial to frequency domain, selecting suitable orientations for image enhancement. The final step of the method involves concatenating and merging the enhanced image with the segmented mask, followed by feature extraction from the objective-oriented enhanced fingerprint. In this paper, our emphasis is on orientation estimation and segmentation. We provide a brief discussion on the minutiae extraction module, with reference to Gabor filters and orientation selection from [9]. We will elaborate on the key implementations in the subsequent sections.

3.2 ResNet-based Orientation Estimation, Segmentation, and Minutiae Extraction

As one of the most crucial global features of fingerprints, the orientation field significantly impacts Automated Fingerprint Identification Systems (AFIS) and plays a substantial role in subsequent tasks such as feature point detection, fingerprint classification, and matching. In addition to the orientation field, the ROI in a fingerprint is essential for minutiae extraction, providing precise location and guiding information for morphological fingerprint processing and minutiae extraction steps. However, due to limitations imposed by collection devices, external environments, and human factors, captured fingerprints are often contaminated by unforeseen factors like equipment noise and uneven pressure during fingerprint collection. These contamination factors have a detrimental effect on both orientation estimation and segmentation tasks. Conventional orientation estimation methods typically rely on filtering operations, which exhibit robustness against noise. However, such methods may struggle to handle situations where ridge lines are heavily contaminated. To achieve accurate orientation field and segmentation maps, the proposed method utilizes an end-to-end trainable neural network to jointly estimate fingerprint orientation and extract foreground ridge/valley lines for subsequent tasks. We conduct a study to assess whether a statistically-driven skip connection neural network architecture can better approximate complex nonlinear transformation operations.

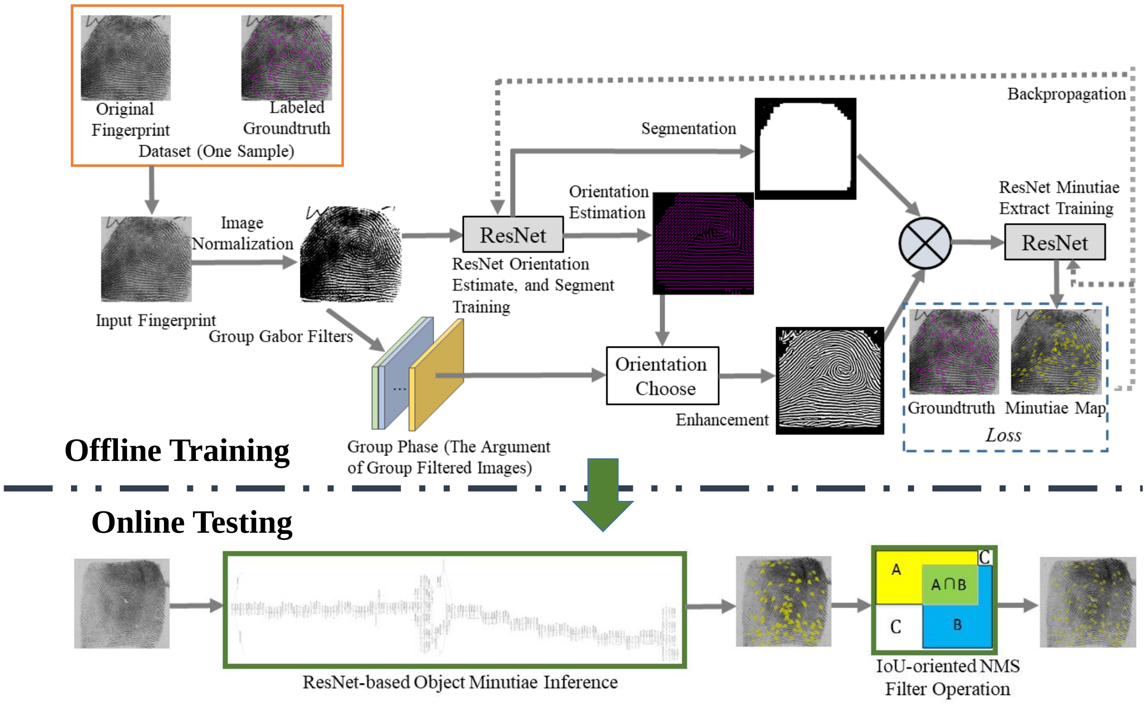


Fig. 4: Minutiae Extraction Framework: This approach entails deep network training offline and subsequent online testing on latent fingerprint images. It utilizes an expanded Resnet architecture for effective feature extraction, encompassing orientation and segmentation masks for fingerprint enhancement and minutiae extraction. Parameter optimization is achieved through backpropagation. Upon sufficient training, the network employs GIoU NMS for precise minutiae detection and pseudo minutiae elimination.

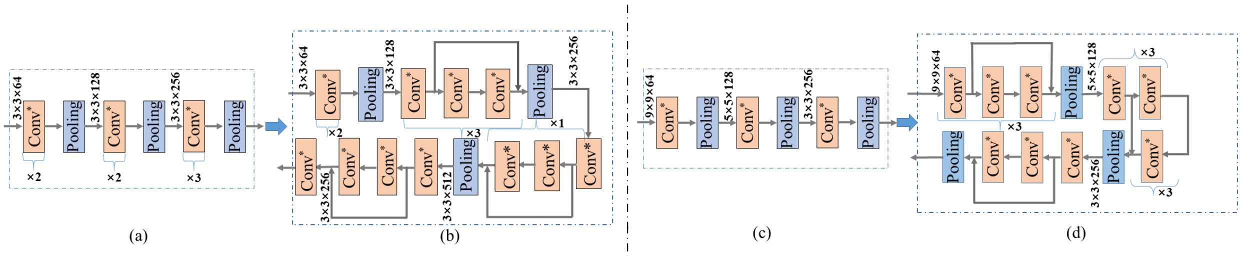


Fig. 5: (a), (c) depict the orientation estimation & segmentation, and minutiae extraction structures in FingerNet, while (b), (d) showcase our corresponding structures.

In contrast to the CNN architecture in [9], the proposed orientation estimation and segmentation module utilizes a ResNet structure to mitigate overfitting, address the issue of vanishing gradients, and augment the representational capacity of neural networks, with a comparative illustration provided in Fig. 5. ResNet [29] has demonstrated outstanding performance, particularly in scenarios requiring the extraction of deep features for image detection tasks. These structures intricately augment the topological graph derived from the original neural networks, and here, it facilitates the learning of discriminative orientation fields, segmentation information, and intrinsic minutiae features. In addition, to provide an objective assessment of ResNet, we will also compare its performance with that of established mainstream deep learning models such as InceptionNet [32], XceptionNet [35], and DenseNet [36] in the experimental section.

3.3 GIoU-oriented Non-Maximum Suppression for Outlier Removal

Because the predicted minutiae via DNNs may cluster together, the Non-maximum suppression (NMS) [24] is usually applied as the final step to remove redundant minutiae in automatic fingerprint minutiae prediction. Typically, Ref. [9] uses the spatial distance (Euclidean distance) between two points as a measurement for judging whether to delete the detected point. In the conventional NMS method for filtering outliers, the extracted minutiae points are sorted based on their scores. The point with the highest score is retained, and the algorithm compares the spatial distance and direction angle difference between each point and the subsequent points. If the distance and angle difference meet the preset thresholds, the point is labeled as redundant. However, this method may mistakenly filter out real minutiae points that are close and have small angle difference. To refine the NMS technique, a sophisticated selection algorithm that effectively discerns true minutiae from closely spaced predictions is required to prevent the erroneous exclusion of genuine features.

Thanks to Intersection over Union (IoU) [26], we are able to utilize a commonly used metric in object detection and tracking benchmarks. It measures the degree of overlap between predicted and ground-truth bounding boxes [27, 28]. While IoU has limitations when dealing with non-overlapping or irregular boundaries, particularly in small object detection tasks. To address this, a generalized version

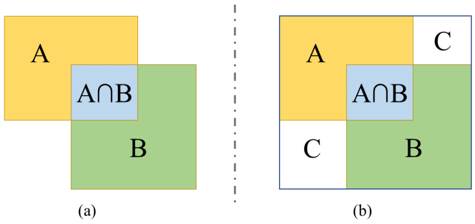


Fig. 6: GIoU Computation: (a) illustrates IoU, and (b) depicts GIoU.

called Generalized IoU (GIoU) has been introduced in [26]. In this research on minutiae point extraction, we propose to utilize the GIoU metric in NMS as it can compare arbitrary shapes and enhance detection quality by correcting false minutiae resulting from outliers and noisy entries. The IoU between two rectangular areas A and B, as depicted in Fig. 6a, can be computed as follows:

$$IoU = \frac{Intersection}{Union} = \frac{|A \cap B|}{|A \cup B|}. \tag{1}$$

Next, we describe the computation method of GIoU (depicted in Fig. 6b): as for the areas A and B, we first find the smallest enclosing convex object C, and compute IoU; based on the result, the GIoU is computed by:

$$GIoU = IoU - \frac{|C \backslash (A \cup B)|}{|C|}. \tag{2}$$

In the process of filtering false minutiae using the proposed NMS algorithm, each point in the final feature score map is considered as a specific region in the original input image. To achieve this, we expand a fixed-size rectangular area centered on each minutia point. All the extracted minutiae are sorted as a queue in descending order by score, denoted as $order$. We determine whether a minutia point in $order$ is deleted based on the GIoU evaluation metric. Specifically, loop to implement the following operation until $order$ is empty: first, the point with the highest score would be set to stored point; second, compute GIoUs between the rest points and the stored (chosen) point, and if GIoU value is bigger than the threshold, the corresponding point would be deleted; third, update $order$ by using remained points in the second step.

By integrating the spatial distance and orientation-based selection strategy with the GIoU-inspired selection strategy, we concatenate the key stages of the merged NMS algorithm to obtain the result with high precision. Alternatively, to enhance the recognition capability of the presented method, we can solely utilize the GIoU-inspired filter. Experimental comparisons of different NMS methods are available in Sec. 4.2.

4. Experiment

In this section, we present the experimental details of our study. We begin by validating the effectiveness of our novel fingerprint minutiae database through a primary verification experiment utilizing an online minutiae extraction algorithm [9]. Ablation studies are then conducted to assess the efficiency and effectiveness of each component in our method. Furthermore, we compare the performance of the proposed method with state-of-the-art algorithms on both the proposed dataset and public dataset FVC 2004 DB1 and DB2 [23], and analyze and discuss the obtained results.

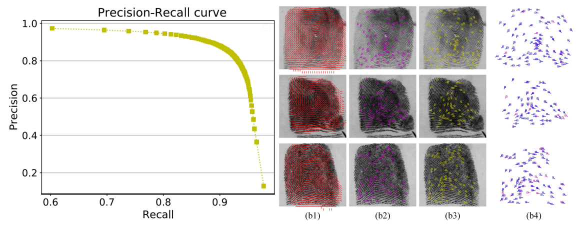


Fig. 7: Left: P-R curve for varying thresholds on the validation dataset. Right: Fingerprint image analysis−(b1) orientation field, (b2) ground-truth annotations, (b3) detected minutiae, and (b4) overlay of ground truth and detections. **Magenta points represent ground truth, while blue and yellow points represent extracted minutiae. The color scheme is consistent throughout the analysis.** The recall rates for the three test sample groups illustrated in (b) are 0.91, 0.85, and 0.86, respectively, with corresponding precision rates of 0.93, 0.90, and 0.85. The significant variability in the performance of the FingerNet model across different test samples primarily stems from the model's limited domain adaptability, variations in image quality, and the architecture's differential feature extraction capabilities in response to the diverse and complex patterns present within the test fingerprints. The extracted minutiae points largely correspond with the ground truth, albeit with some missed points and false positive detections.

The proposed method was implemented using Keras and TensorFlow and tested on a server equipped with an Xeon E7 v3 processor and GeForce GTX TITAN X GPU. Our experiments utilized a constructed dataset based on NIST SD04 [22] images, with a training-to-test set ratio of 3:1. Each input image had a size of 512×512, and a batch size of 1 was utilized to circumvent memory limitations. The neural network was trained end-to-end using the ADAM optimization method with a learning rate of 0.0001, first moment exponential decay rate β_1 of 0.9, second moment exponential decay rate β_2 of 0.999, and epsilon value of 1×10^{-8}. The model underwent training for 20 epochs. For objective performance assessment, we utilize precision, recall, F_1 score, location and orientation error, inference time, and the Precision-Recall (P-R) curve to evaluate the efficacy, efficiency, and robustness of the detection methods.

4.1 Experimental Evaluation of the Constructed Dataset

To assess the newly created minutiae dataset, we performed a 20-epoch training of the FingerNet model [15]. On the FingerNet, we follow the training settings in [15] to validate the effectiveness of the created dataset. We assess the P-R curve by testing FingerNet on our dataset (Fig.7 left part) and comparing the results with the ground truth. The minutiae detection threshold is adjusted across $[0.00001, 0.01, 0.02, ..., 0.98, 0.99, 0.99999]$. Lower thresholds increase recall but decrease precision, and vice versa for higher thresholds. Setting $thresh$ to 0.75 balances precision (0.8891) and recall (0.8915). Fig.7 (right part) shows three examples from our dataset with closely matched annotations and inferences, confirming the model-dataset synergy. The average inference time per image on a GPU is approximately 0.62 seconds.

4.2 Ablation Study

In the experiment, we utilize the ResNet-based backbone for orientation estimation, segmentation, and minutiae extraction, along with comparative experiments involving various neural network structures as backbones, which are detailed in Appendix B. In the preceding section, we have validated the efficacy of the proposed dataset by employing FingerNet. Henceforth, we will employ the well-trained FingerNet as the baseline. Here, we first conduct a comprehensive performance comparison of different NMS proposals. Next, we evaluate the method by analyzing the P-R curves across various neural networks and NMS combinations.

4.2.1 Performance Assessment of NMS Proposals

To assess the efficacy of GIoU-guided metrics, Euclidean distance heuristic metrics, and their combined schemes in reducing false minutiae, we perform ablation experiments using various non-maximum suppression (NMS) strategies. These strategies include position and orientation-based NMS, GIoU-guided NMS, and a hybrid approach that combines both methods. All experiments are performed under the same conditions. In the comparative experiment, we utilize the well-trained FingerNet [9] as the minutiae extractor and set the credibility threshold for all minutiae to 0.75 for fair comparison. Besides, we conduct comparison experiments using two sub-datasets formulated in Sec. 2, denoted as NISTSD 0406 and NISTSD 0407, which both consist of 258 images. Table 1 manifests the overall quantitative comparison results of the three methods. Meanwhile, two test examples from the test sets and results are shown in Fig. 8 with corresponding evaluation metrics. From the comparison results (including Table 1 and Fig. 8), we observe that there exists a trade-off between precision and recall and the GIoU-inspired operation slightly hinders the detection precision, while it indeed has some improvements over the others in terms of the recall evaluation on the two datasets, indicating that the GIoU has better ability to discern intricate features and cover more qualified minutiae points compared with spatial distance & orientation constrained method. That's mainly because GIoU comprehensively considers the overlap area, shape, and positional relationship of bounding boxes in removing redundancy, adapts to various shape variations, reduces the likelihood of erroneous deletions, and better balances precision and recall, as shown the F_1-value in Table 1. From the quantitative statistical results, we can also see that the precision is highly consistent with the location-coordinates and orientation-values errors, which are determined by the filter mechanisms, i.e., the combination method can significantly improve the precision and related two other indexes. In the combination approach, minutiae are extracted using dual filters and subsequently integrated via an iterative comparison process that retains candidates surpassing a defined credibility threshold, enhancing precision over single-filter methods. Meanwhile, minutiae within the 0 to $thresh$ distance range undergo deduplication, improving precision but potentially removing genuine minutiae and leading to decreased recall.

Table 1: The overall quantitative comparison results on NISTSD 0406 and 0407 using different types of NMS algorithms. In the context, "LE" refers to "Location-coordinates error", "OE" refers to "Orientation-value error" and "Combination" represents the combination of distance & orientation and GIoU constraints.

Dataset	NMS	D&O	GIoU	Combination
0406	Precision	0.8954	0.8841	**0.9178**
	Recall	0.8887	**0.9054**	0.8698
	F_1−value	0.8901	**0.8927**	0.8914
	LE	1.8075	2.0884	**1.7740**
	OE	0.0328	0.0350	**0.0313**
0407	Precision	0.8887	0.8757	**0.9127**
	Recall	0.8922	**0.9109**	0.8731
	F_1−value	0.8884	**0.8909**	0.8907
	LE	1.7781	2.0892	**1.7414**
	OE	0.0332	0.0355	**0.0315**

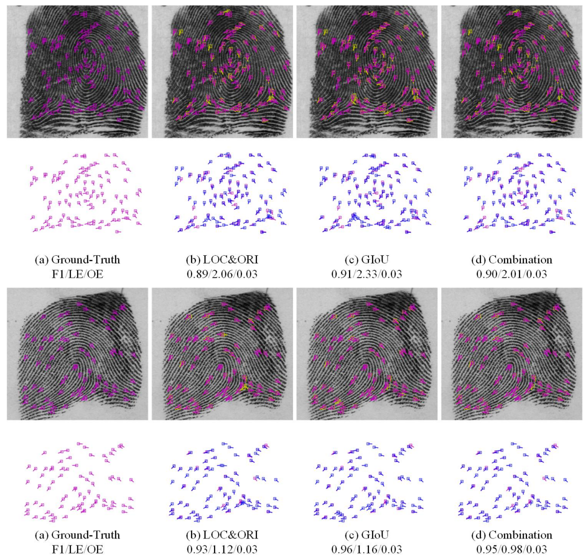


Fig. 8: Sample results obtained using different NMS algorithms. The evaluation metrics used are the F_1 score, location-coordinates error (LE), and orientation value error (OE), and **the abbreviations retain their meanings throughout the paper**. The abbreviations "LOC" and "ORI" represent location and orientation, respectively.

4.2.2 Comparative Analysis of Neural Network Inference and NMS Approaches

In this section, to comprehensively assess the efficacy of each component within the proposed method, we conducted an exhaustive evaluation of all possible module combinations. First, we evaluate the similar minutiae detection methods including FingerNet [9] accompanied with location-coordinate & orientation-inspired NMS, (referred to as *FingerNet+LO-NMS*) and joint FingerNet and GIoU-inspired post-processing (referred to as *FingerNet+GIoU-NMS*). Apart from comparison with the baseline algorithm, we also demonstrate the contribution of part of our GIoU-oriented NMS mechanism by replacing it with conventional location-coordinate and orientation-based mode, termed as *Ours+LO-NMS*, while our complete method is corresponding denoted as *Ours+GIoU-NMS*. Fig. 9 (right part) shows two fingerprint samples (including their groundtruth minutiae annotations obtained using the method described in Sec. 2) and four corresponding detection results by the afore-

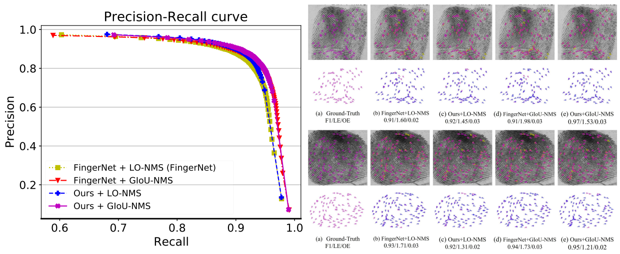


Fig. 9: P-R curves for minutiae detection using various methods. Yellow: FingerNet with coordinate and orientation-based NMS. Red: FingerNet with GIoU-based NMS. Blue: Proposed model with coordinate and orientation-based NMS. Magenta: Proposed model with GIoU-based NMS. The right panel displays detection samples from the proposed dataset under different processing combinations, with comparative results at optimal thresholds.

mentioned combination methods. Each visual result set is divided into two images: the upper fingerprint image contains the actual or actual and detected minutiae for visualization, while the lower image is reserved for showing actual minutiae or comparing actual minutiae against detected ones, displaying either solely the actual minutiae or both actual and detected minutiae to enable a detailed comparative analysis. From it, we can see that *FingerNet+LO-NMS* is capable of detecting main minutiae roughly, which also leaves out some minutiae or detects false minutiae. In comparison, *Ours+LO-NMS* can locate the minutiae more precisely in an unknown fingerprint image. A similar phenomenon also occurs in the comparison between *FingerNet+GIoU-NMS* and *Ours+GIoU-NMS*, which verifies the effectiveness of the developed end-to-end extractor module. The two groups of ablation experiments confirm that leveraging ResNet as the backbone network enhances the reliability and efficacy of presented detection method. This is mainly because residual connections add some value of their own, as well as allowing training of deeper networks, which may also make it easier to learn a good solution that generalizes well. Similarly, ablation studies demonstrate that incorporating the enhanced NMS module into our detection system yields an adaptive filtering effect and credible minutiae outcomes, with qualitative and quantitative comparisons affirming the superior performance of the GIoU-based NMS approach.

Fig. 9 (left part) presents the P-R curves, contrasting the detected minutiae against ground truth. The method assesses minutiae validity using orientation, location, and confidence score discrepancies. We observe that as the detection threshold varies, all the curves show a similar pattern. When the threshold is set higher, precision is higher while recall is lower. In this case, the curves of *FingerNet+LO-NMS*, *FingerNet+GIoU-NMS*, *Ours+LO-NMS* and *Ours+GIoU-NMS* mostly overlap. As the threshold decreases, recall increases while precision decreases. Notably, the performance ranking from high to low is *Ours+GIoU-NMS*, *Ours+LO-NMS*, *FingerNet+GIoU-NMS*, and *FingerNet+LO-NMS*, indicating that the neural network architecture plays a crucial role in improving detection accuracy. As the threshold decreases further, the curves of *FingerNet+LO-NMS* and *Ours+LO-NMS* as well as *FingerNet+GIoU-NMS* and *Ours+GIoU-NMS* overlap, indicating that a lower credibility threshold leads to more false positive detections. Additionally, the improved NMS shows better adaptability in

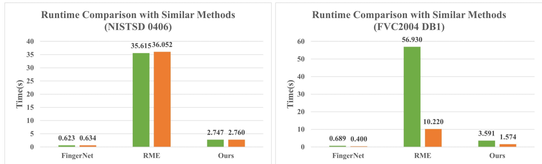


Fig. 10: Runtime performance comparison across NIST SD04 and FVC 2004 datasets.

removing false minutiae points.

4.3 Comparison with Other Methods

In this section, the overall performance of the proposed method will be validated through comparisons with several state-of-the-art methods. MINDTCT [30] will be included in the comparison, since it is a widely used open source NIST biometrics recognition software. Meanwhile, FingerNet [9] is a pioneering method in minutiae extraction using CNNs. It extracts fingerprint minutiae points by incorporating general prior knowledge of fingerprints, making it essential for comparison in this study. In addition, the robust minutiae extractor approach in [10] joins in the comparison since it carefully divides computing tasks among different neural networks under a novel architecture, denoted as *RME*. More specifically, *RME* uses a two-stage strategy for extracting the minutiae: first CoarseNet is applied to obtain both the minutiae score map and minutiae orientation results, and then FineNet is used to conduct candidate minutiae locations refinement processes. The algorithm implementation can be obtained from public project†. To ensure fairness in comparison, we retrain CoarseNet on the created dataset with its original settings and also utilize the FineNet model released in [10] as a classifier, as minutiae elements exhibit consistent patterns across different fingerprints, allowing direct usage of a pre-trained minutiae classification model.

Table 2 provides a comprehensive performance comparison on the NIST SD04, including precision, recall, location-coordinate error, and orientation error. The dataset consists of two sub-datasets, NISTSD 0406 and NISTSD 0407, each containing 258 images. The proposed method outperforms state-of-the-art techniques [9, 10, 30] in terms of precision and recall across both these sub-datasets. This is particularly crucial in the domain of personal identity verification. Furthermore, our method achieves the lowest orientation errors, while the location errors are comparable to FingerNet and significantly lower than the other two methods, demonstrating our approach's superiority. We also compare the run-time of the proposed method with two similar DNN methods [9, 10] in Fig. 10, using identical GPU parallel settings. Based on this comparison, our method outperforms *RME* and demonstrates significant speed improvements or approximation gains compared to FingerNet. However, testing on the NIST SD04 dataset alone is insufficient to validate the generalizability of the proposed method, thus, two additional datasets, FVC 2004 DB1 and DB2 [23] are exploited to evaluate our method, alongside a comparison with the aforementioned methods. The labeling method in Sec. 2

† https://github.com/luannd/MinutiaeNet

is applied to obtain the minutiae information as ground-truth. We conduct two statistical comparisons and show the overall test performance in Table. 2. Overall, the proposed method demonstrates superior performance in terms of minutiae extraction. Meanwhile, the speed of our method is also compared with two similar deep learning-based approaches [9, 10] in the same GPU parallel setting on the two datasets, as shown in Fig. 10. The figure demonstrates that the processing time for the first set of images is longer than that for the second set, which can be attributed to the disparity in image sizes between the two groups.

Fig. 11 provides a detailed comparative visual analysis of fingerprint samples from two benchmark databases, NIST SD4 and FVC 2004. The figure provides a side-by-side comparison of the raw fingerprint images with their corresponding detection results, showcasing the capabilities of four state-of-the-art detection algorithms. To facilitate a granular examination of the detection efficacy, the figure also features enlarged views of select regions, capturing the intricacies of the detection outcomes. These intricate visualizations are corroborated by the quantitative metrics enumerated in Table 2, ensuring a holistic understanding of the detection performance. It is evident from the visualization that MINDTCT exhibits limited accuracy in minutiae extraction due to its weaker representation power and its difficulty in dealing with blurry and noisy ridge areas. Tang *et al.*'s CNN-based method [9] shows improved performance but still suffers from false positives and missed detections due to the inadequate learning of distinctive minutiae features. The inadequate detection quality of such methods is further substantiated through experimental results on the NIST SD04 and FVC 2004 datasets. In contrast, the *RME* method [10], employing a two-stage deep learning approach, achieves impressive precision and recall. However, its performance in detecting complete minutiae is relatively poor, possibly due to less effective redundant point removal. The proposed method, benefiting from an advanced network architecture and GIoU-oriented NMS operation, demonstrates superior accuracy and completeness in the detection results of Fig. 11 and Table 2. Notably, the proposed method exhibits better detection performance, especially in areas with intricate details.

4.4 Discussion

In Sec. 4.3, we have compared our method with leading techniques, including MINDTCT, FingerNet and *RME*. The experimental results reported in the previous sections indicate that the proposed method surpasses the compared techniques in terms of precision and recall across the NIST SD04 dataset (including NISTSD 0406 and NISTSD 0407 two sub-datasets). The effectiveness of the proposed method is crucial for applications demanding high accuracy, such as criminal investigation, access control systems and financial transactions. Furthermore, our approach achieves relatively lower orientation errors and shows comparable location errors against FingerNet, implying an overall superior performance in prediction.

We have also evaluated the run-time efficiency of the proposed method against similar DNN-based methods. The observed significant speed improvements over *RME* and competitive performance compared to FingerNet demonstrate the efficiency of our approach. We further validate the generalizability on the FVC 2004 DB1 and DB2 datasets, where our method consistently delivers robust experimental outcomes. While on DB1 test set, the *RME* method [10] achieves impressive result, because the patch based minutiae classifier applied can compact embedding of minutiae features, which is particularly suitable for scenes with concentrated ROI and fingerprint patterns. Our method also demonstrates good performance on DB1, particularly in terms of detection integrity, surpassing other methods. These results affirm our method's robustness across varied datasets.

The primary strength of the proposal lies in its ability to detect minutiae with enhanced accuracy and completeness. This is facilitated by the innovative network architecture and the implementation of GIoU-oriented NMS operation. The latter contributes to a better detection performance due to its flexible adaptivity, particularly in challenging areas with intricate details. The experimental results, supported by quantitative data and visual analysis, demonstrate the robustness of our method across a range of fingerprint image qualities. Despite achieving high precision and low orientation errors, the need for wider dataset validation, refined location accuracy in noisy conditions, and improved run-time efficiency for real-time application persists, pointing towards future work in model optimization and lightweight design.

5. Conclusion

This paper proposes an effective automatic minutiae extraction method. To address the lack of comprehensive minutiae datasets, we propose a semi-automated annotation algorithm based on explicit knowledge of morphology to label fingerprint images. Our method effectively fills the gap in the availability of minutiae datasets. We propose a novel end-to-end detection model for AFIS that leverages the ResNet structure and adopts the Highway networks strategy to enhance the extraction of minutiae with higher accuracy. Moreover, we incorporate the GIoU-oriented NMS filter to adaptively remove pseudo minutiae points. Experimental results on different datasets demonstrate that our method achieves competitive performance compared to state-of-the-art approaches for small-scale minutiae detection. Additionally, our method is versatile and applicable to diverse types of minutiae, making it suitable for various real-world fingerprint-related tasks.

References

[1] A. K. Hrechak *et al.*, "Automated fingerprint recognition using structural matching," *Pattern Recognit.*, vol. 23, no. 8, pp. 893-904, 1990.

[2] J. Sang, H. Wang, Q. Qian, H. Wu, and Y. Chen, "An

Table 2: Comparative performance on NISTSD 0406, 0407, and FVC2004 DB1 and DB2 datasets, evaluated by Precision, Recall, Location Error, and Orientation Error. All results are derived from uniform quantitative testing protocols.

Methods	Datasets	Precision	Recall	LE	OE	Datasets	Precision	Recall	LE	OE
MINDTCT [30]	0406	0.1637	0.2210	3.7482	0.2746	DB1	0.2288	0.2732	4.4265	0.2895
FingerNet [9]		0.8997	0.8838	**1.7958**	0.0326		0.2768	0.3790	5.4717	0.1181
RME [10]		0.7846	0.3255	2.5400	0.0521		**0.6804**	0.3881	**4.1607**	**0.0747**
Ours		**0.9006**	**0.9009**	1.8708	**0.0303**		0.5859	**0.4326**	4.2298	0.0858
MINDTCT [30]	0407	0.1566	0.2127	3.6692	0.2755	DB2	0.2973	0.3765	4.5108	0.2489
FingerNet [9]		0.8934	0.8880	**1.7711**	0.0331		0.4027	0.5418	3.8472	0.0746
RME [10]		0.7920	0.3216	2.5312	0.0526		0.6854	0.4840	4.7409	**0.0698**
Ours		**0.8987**	**0.9056**	1.8430	**0.0298**		**0.7201**	**0.7277**	**3.6379**	0.0714

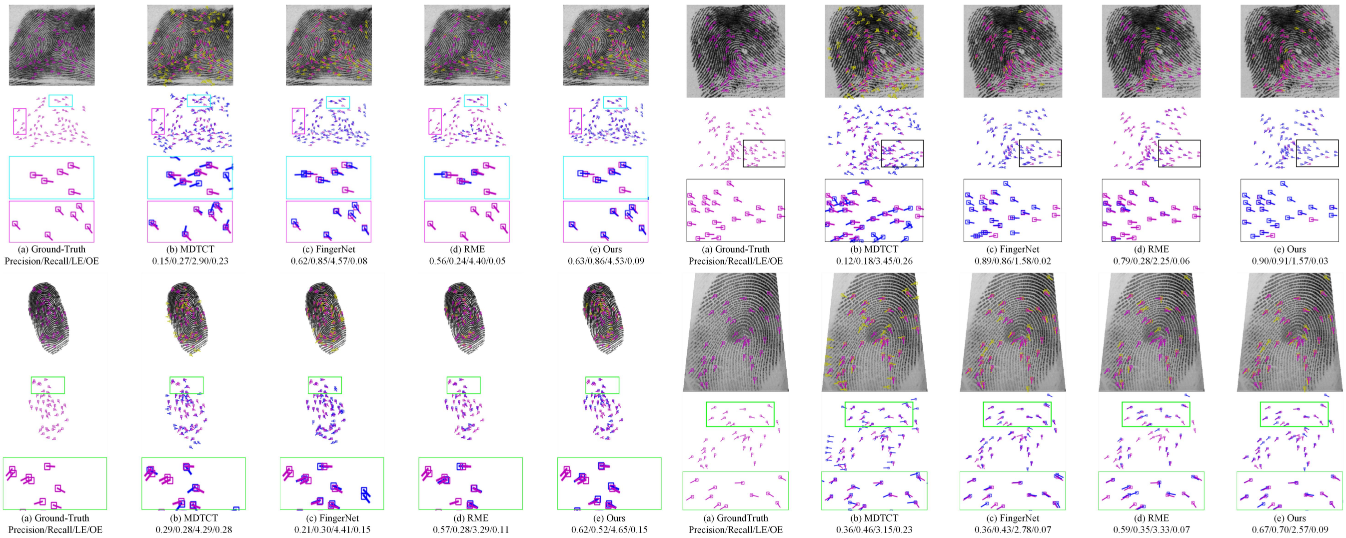


Fig. 11: Comparative analysis of fingerprint minutiae detection on NISTSD 0406, NISTSD 0407, FVC2004 DB1, and FVC2004 DB2 datasets. The first row presents experimental results for NIST SD4, and the second row for FVC 2004 datasets.

efficient fingerprint identification algorithm based on minutiae and invariant moment," *Pers Ubiquit Comput.*, vol. 22, no. 1, pp. 71-80, 2018.

[3] T. Chugh, K. Cao, and A. K. Jain, "Fingerprint Spoof Buster: Use of Minutiae-Centered Patches," *IEEE Trans. Inf. Forensics Secur.*, vol. 13, no. 9, pp. 2190-2202, 2018.

[4] J. J. Engelsma, K. Cao and A. K. Jain, "Learning a Fixed-Length Fingerprint Representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 6, pp. 1981-1997, 2021.

[5] A. K. Jain, D. Deb and J. J. Engelsma, "Biometrics: Trust, But Verify," *IEEE Trans. Inf. Forensics Secur.*, vol. 4, no. 3, pp. 303-323, 2022.

[6] D. Maltoni *et al.*, "Handbook of fingerprint recognition," *2nd ed. Springer*, 2009.

[7] A. Jain, L. Hong, R. Bolle "On-line fingerprint verification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 4, pp. 302-314, 1997.

[8] Y. L. Yin *et al.*, "An Improved Algorithm for Minutiae Extraction in Fingerprint Images," *Journal of Image and Graphics*, 2002.

[9] Y. Tang, F. Gao, J. Feng *et al.*, "FingerNet: An unified deep network for fingerprint minutiae extraction," in *Proc. IJCB.*, Oct. 2017, pp. 108-116.

[10] D. Nguyen, K. Cao, and A. K. Jain, "Robust Minutiae Extractor: Integrating Deep Networks and Fingerprint Domain Knowledge," in *Proc. ICB.*, Feb. 2018, pp. 9-16.

[11] Neurotechnology, VeriFinger, 2010.

[12] X. Jiang, W.-Y. Yau, and W. Ser, "Detecting the fingerprint minutiae by adaptive tracing the gray-level ridge," *Pattern Recognit.*, vol. 34, no. 5, pp. 999-1013, 2001.

[13] F. Zhao and X. Tang, "Preprocessing and postprocessing for skeleton-based fingerprint minutiae extraction," *Pattern Recognit.*, vol. 40, no. 4, pp. 1270-1281, 2007.

[14] B. Zhou, C. Han, Y. Liu, T. Guo, and J. Qin, "Fast minutiae extractor using neural network," *Pattern Recognit.*, vol. 103, p. 107273, 2020.

[15] H. Zhao, S. Zheng, "A Morphological Fingerprint Minutiae Annotation Algorithm for Deep Learning Datasets," in *Proc. ISCAS.*, May 2022, pp. 1-5.

[16] A. A. Paulino, A. K. Jain and J. Feng, "Latent Fingerprint Matching: Fusion of Manually Marked and Derived Minutiae," in *Proc. SIBGRAPI*, Sep. 2010, pp. 63-70.

[17] M. Kayaoglu, B. Topcu and U. Uludag, "Standard Fingerprint Databases: Manual Minutiae Labeling and Matcher Performance Analyses," *arXiv:1305.1443*, 2013.

[18] R. Bansal *et al.*, "Minutiae extraction from fingerprint images-a review," *arXiv:1201.1422*, 2011.

[19] A. Chowdhury *et al.*, "Can a CNN Automatically Learn the Significance of Minutiae Points for Fingerprint Matching?," in *Proc. WACV.*, Mar. 2020, pp.351-359.

[20] H. Lin, W. Yifei, and A. Jain, "Fingerprint image enhancement: algorithm and performance evaluation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 777-789, 1998.

[21] M. D. Garris and R. M. McCabe, "NIST Special Database 27: fingerprint minutiae from latent and matching tenprint images," *National Institute of Standards & Technology*, 2000.

[22] NIST special database 4, Aug. 27, 2010. [Online]. https://www.nist.gov/srd/nist-special-database-4.

[23] FVC2004: The Third International Fingerprint Verification Competition. http://bias.csr.unibo.it/ fvc2004/.

[24] A. Neubeck and L. V. Gool, "Efficient Non-Maximum Suppression," in *Proc. ICPR.*, Aug. 2006, pp. 850-855.

[25] D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv:1412.6980*, 2014.

[26] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression," in *Proc. CVPR.*, Jun. 2019, pp. 658-666.

[27] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," *Int. J. Comput. Vision*, vol. 88, no. 2, pp. 303-338, 2010.

[28] T.-Y. Lin et al., "Microsoft COCO: Common Objects in Context," in *Proc. ECCV.*, Sep. 2014, pp. 740-755.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. CVPR.*, Jun. 2016, pp. 770-778.

[30] Ko, K. (2007), User's Guide to NIST Biometric Image Software (NBIS), *National Institute of Standards and Technology*, Gaithersburg, MD.

[31] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv:1409.1556*, 2014.

[32] C. Szegedy, W. Liu, Y. Jia, *et al.* "Going deeper with convolutions," in *Proc. CVPR.*, Jun. 2015, pp. 1-9.

[33] A. G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv:1704.04861*, 2017.

[34] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proc. ICML.*, Jun. 2019, pp. 6105-6114.

[35] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in *Proc. CVPR*, Jul. 2017, pp. 1800-1807.

[36] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *Proc. CVPR*, Jul. 2017, pp. 2261-2269.

Appendix A: Introduction of Fingerprint Dataset

The dataset, named the Fingerprint Minutiae Dataset (FMD), comprises the location coordinates and orientation values of each minutia within every fingerprint image. For this dataset, we select 2910 images from the publicly available NIST SD04 dataset [22]. The NIST SD04 dataset is specifically distributed for the fingerprint classification task and contains 4000 8-bit encoded images. The selected images are categorized into five classes (Arch, Left Loop, Right Loop, Tented Arch, and Whorl) based on the pattern near the singularity points. Each image has a size of 512×512, with 32 rows of pixel blanks at the bottom. The labeling algorithm is implemented in MATLAB, and during the labeling process, we manually review and correct any inaccuracies in the minutiae annotations, involving at least two annotators. The aligned minutiae points in minutiae dataset are stable and representative, as they can be used to determine the uniqueness of a fingerprint [18–20].

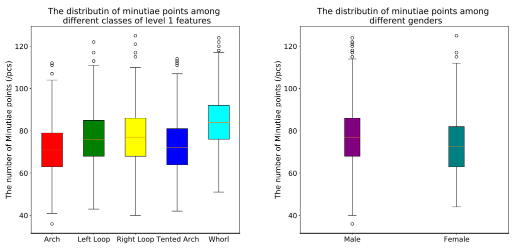
The dataset comprises 2910 fingerprint images with a total of 223,207 minutiae, averaging 76.7 minutiae points per image. We conducted statistical analysis on the distribution of images and minutiae based on original classes and gender divisions. The results are summarized in Table A· 1. For visualization and analysis purposes, a boxplot (Fig. A· 1) is generated, where the red line represents the median value. The boxplots demonstrate that the interquartile ranges (IQRs) for minutiae counts, when considering various Level-1 features and across genders, are primarily concentrated within the 60 to 90 interval. This distribution is consistent with the established fingerprint quality standards, which suggest an acceptable range of 40 to 100 [18]. The distribution of minutiae points demonstrates a relatively balanced distribution among different classes and genders, with slightly higher median values for Whorl and Male images. Additionally, a few outliers (e.g., minutiae Num $\geq$ 110.5 for Left Loop) are identified outside of the main intervals. However, these minor deviations are not expected to significantly affect the labeling results, as the overall distribution of minutiae points remains fairly consistent. Therefore, the annotated minutiae dataset meets the requirements for subsequent training applications in theory.

Table A· 1: The statistics of minutiae points in different categories and genders

Tuple count	Categories of level-1 feature classification					Genders	
	Arch	Left Loop	Right Loop	Tented Arch	Whorl	Male	Female
Image	534	589	606	572	609	2410	500
Minutiae	38123	45299	46971	41572	51242	186551	36656

Table A· 2: The detailed attributes comparison of different datasets

Dataset	Image Size	Number of Minutiae		
		Avg	Max	Min
FVC2004DB1A	640×480	40.96	80	11
FVC2004DB3A	300×480	40.76	76	11
NISTSD0406	512×512	78.01	114	45
NISTSD0407	512×512	79.96	121	49



(a) Level-1 features (b) Genders

Fig. A· 1: Distribution of minutiae points across Level-1 Features and Genders.

Compared with the revoked NIST SD27, we use more fingerprint images for testing, which include 516 fingerprints and more than 258 fingerprints in the NIST SD27 dataset. We compare our minutiae dataset with the FVC 2004 dataset [23], a benchmark for fingerprint recognition. Table A· 2 shows the comparison results, including statistical information on the FVC 2004 dataset obtained from [14]. Compared to the standard fingerprint distribution of [40, 100], the proposed dataset exhibits a more reasonable distribution of minutiae counts in each fingerprint.

Appendix B: Comparative Analysis of Deep Learning Models for Minutiae Extraction

To objectively evaluate the backbone network, we benchmarked ResNet and other prevalent architectures, including VGGNet [31], InceptionNet [32], XceptionNet [35], DenseNet [36], MobileNet [33], and EfficientNet [34], in our experiments. We conducted experiments on publicly available fingerprint datasets, including NIST SD04 and FVC 2004. For each model, fingerprint feature extraction modules were implemented based on their respective core ideas. To ensure fair comparison, consistent preprocessing and augmentation were performed on all models. In this study, we employ the aforementioned F_1 Score, LE and OE as the evaluation metrics. Because the overall model size remains relatively consistent, the difference in inference time can be considered negligible. Therefore, we focus solely on presenting the F_1 Score, LE, and OE in our experiments. The F_1 Score, being the harmonic mean of precision and recall, offers a comprehensive representation of the overall performance of the detector. The mean localization error is a metric that quantifies the average Euclidean distance between the predicted and ground truth positions of fingerprint minutiae. The mean error of angle is a metric that assesses the average angular deviation between the predicted and actual orientations of fingerprint minutiae.

Table A· 3 manifests performance comparison of different models on fingerprint minutiae extraction task. Fig. A· 2 shows two fingerprint image samples obtained from the NISTSD 04 and FVC 2004 datasets, along with the corresponding minutiae detection results of several state-of-the-art models. In each set, the upper image represents the deep model's detection results and the corresponding ground truth, while the lower image compares the model's pure detections with the ground truth minutiae. We observe that although VGGNet performs well in image classification tasks, its performance in fingerprint minutiae extraction is slightly inferior to ResNet, possibly due to its deep hierarchical structure not being suitable for capturing subtle detailed features. The Inception model, with its multi-scale convolutional kernel, is capable of capturing details at different levels. The Xception model, utilizing depthwise separable convolution, improves parameter efficiency and helps in learning finer features with limited data, achieving relatively better performance on both datasets compared to InceptionNet. However, their overall performance is not as good as ResNet. DenseNet facilitates feature propagation and detailed feature acquisition via feature reuse; nevertheless, as network depth grows, the potential for suboptimal feature reuse may arise, possibly impeding generalization. Moreover, deeper networks are usually harder to train due to issues like noisy gradient updates, which can affect the learning process. Therefore, models that perform well on the NIST SD04 dataset may have poor generalization ability. MobileNet is designed for mobile and embedded devices, and its lightweight structure may be beneficial for deploying fingerprint recognition systems in resource-constrained environments. Nonetheless, its accuracy on NIST SD04 is relatively low. EfficientNet exhibits excellent capability in extracting complex fingerprint features, which contributes to the generation of well-generalized trained models. The findings of our study reveal that while EfficientNet generally outperforms other models in generalization, ResNet has been adopted as the baseline for our investigation. This decision is informed by ResNet's exemplary proficiency in feature extraction, the ease with which it can be implemented and deployed, and its demonstrated robustness in accurately extracting a diverse range of fingerprint minutiae.

Table A· 3: Comparative ablation study of backbone networks for fingerprint feature estimation: evaluating the impact of VGG, Inception, Xception, DenseNet, MobileNet, EfficientNet, and ResNet on F_1 Score, location-coordinates error (LE), and orientation-values error (OE).

F_1/LE/OE / Model	NISTSD		FVC2004	
	0406	0407	DB1	DB2
VGGNet	0.8917/**1.7958**/0.0326	0.8907/**1.7711**/0.0331	0.3199/5.4717/0.1181	0.4620/3.8472/0.0746
Inception	0.8868/1.9996/0.0356	0.8884/1.9816/0.0354	0.1207/3.9802/0.0792	0.5403/3.7707/0.0674
Xception	0.8913/1.9612/0.0340	0.8966/1.9338/0.0330	**0.5650**/4.4865/0.0886	0.6054/3.7494/0.0672
DenseNet	0.8946/1.9749/0.0327	0.8979/1.9706/0.0332	0.1642/**3.8622**/**0.0763**	0.5035/3.8680/**0.0630**
MobileNet	0.8881/2.0644/0.0342	0.8873/2.0422/0.0341	0.4674/4.6422/0.0937	**0.7799**/3.6671/0.0665
EfficientNet	0.8821/2.3007/0.0355	0.8835/2.2956/0.0361	0.5002/5.1321/0.1011	0.7547/3.7846/0.0700
ResNet	**0.9007**/1.8708/**0.0303**	**0.9021**/1.8430/**0.0298**	0.4977/4.2298/0.0858	0.7239/**3.6379**/0.0714

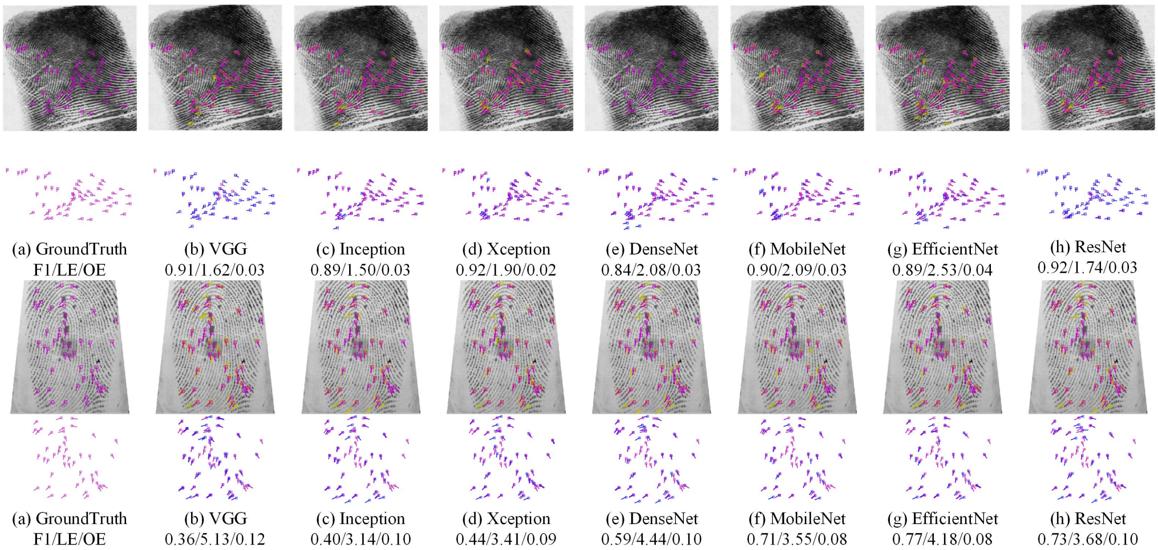
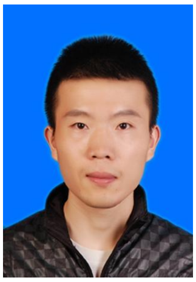
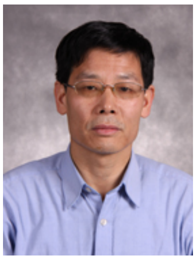


Fig. A· 2: Evaluation of fingerprint minutiae detection on sample images from NIST SD4 and FVC 2004 datasets. In the conducted experiments, the employment of ResNet as the backbone network demonstrates superior robustness in fingerprint minutiae detection across varied image inputs.



Hongtian Zhao received B.S., M.S., and Ph.D from Shandong University of Science and Technology, Sichuan University, and Shanghai Jiao Tong University in 2015, 2018 and 2023, respectively. He is currently working in College of Mathematics and Systems from Xinjiang University. His current research interests include adversarial robustness in deep learning, intelligent video analysis, and biometric recognition.

Hua Yang received the Ph.D. degree in communication and information from Shanghai Jiao Tong University, in 2004, and both the B.S. and M.S. degrees in communication and information from Haerbin Engineering University, China in 1998 and 2001, respectively. She is currently a Professor with the Electronic Engineering Department SJTU, Shanghai, China. Her current research interests include computer vision, machine learning, and smart video surveillance applications.



Shibao Zheng received the B.S. degree in communication engineering from Xidian University, Xi'an, and the M.S. degree in the signal and information processing from the 54th institute of CETC, Shijiazhuang, China, in 1983 and 1986, respectively. He is currently a Professor with the Electronic Engineering Department SJTU, Shanghai, China. His current research interests include urban video surveillance system, intelligent video analysis.