

Finger Pose Estimation for Under-screen Fingerprint Sensor

Xiongjun Guan[✉], Zhiyu Pan[✉], Jianjiang Feng[✉], Member, IEEE, and Jie Zhou[✉], Fellow, IEEE

Abstract—Two-dimensional pose estimation plays a crucial role in fingerprint recognition by facilitating global alignment and reduce pose-induced variations. However, existing methods are still unsatisfactory when handling with large angle or small area inputs. These limitations are particularly pronounced on fingerprints captured by under-screen fingerprint sensors in smartphones. In this paper, we present a novel dual-modal input based network for under-screen fingerprint pose estimation. Our approach effectively integrates two distinct yet complementary modalities: texture details extracted from ridge patches through the under-screen fingerprint sensor, and rough contours derived from capacitive images obtained via the touch screen. This collaborative integration endows our network with more comprehensive and discriminative information, substantially improving the accuracy and stability of pose estimation. A decoupled probability distribution prediction task is designed, instead of the traditional supervised forms of numerical regression or heatmap voting, to facilitate the training process. Additionally, we incorporate a Mixture of Experts (MoE) based feature fusion mechanism and a relationship driven cross-domain knowledge transfer strategy to further strengthen feature extraction and fusion capabilities. Extensive experiments are conducted on several public datasets and two private datasets. The results indicate that our method is significantly superior to previous state-of-the-art (SOTA) methods and remarkably boosts the recognition ability of fingerprint recognition algorithms. Our code is available at <https://github.com/XiongjunGuan/DRACO>.

Index Terms—Fingerprint, pose estimation, fingerprint recognition, multimodal, decoupled probability distribution, feature fusion, knowledge distillation, knowledge transfer.

I. INTRODUCTION

Two-dimensional pose estimation has been extensively researched in the field of fingerprint [1]–[10]. This task aims to determine the fingerprint’s center position and rotation direction from an input image, enabling the effective alignment of heterogeneous data within a unified coordinate system [11]. Functioning as a robust global prior, fingerprint pose plays a pivotal role in fingerprint recognition systems, and is typically employed as an essential preprocessing stage [11]. For example, pose normalization can substantially reduce intra-class differences caused by varying geometric positions, thereby effectively enhancing the generalizability and discriminability of feature extraction [8], [12]–[15]. Besides, incorporating supplementary constraints on pose relationships inherently filters out imposter pairs with erroneous spatial correspondences while streamlining the search space, which can significantly improve the accuracy and efficiency of matching algorithms [10], [16]–[20].

This work was supported in part by the National Natural Science Foundation of China under Grant 62376132 and 62321005. (*Corresponding author: Jianjiang Feng*)

The authors are with Department of Automation, Tsinghua University, Beijing 100084, China (e-mail: gxj21@mails.tsinghua.edu.cn; pzy20@mails.tsinghua.edu.cn; jfeng@tsinghua.edu.cn; jzhou@tsinghua.edu.cn).

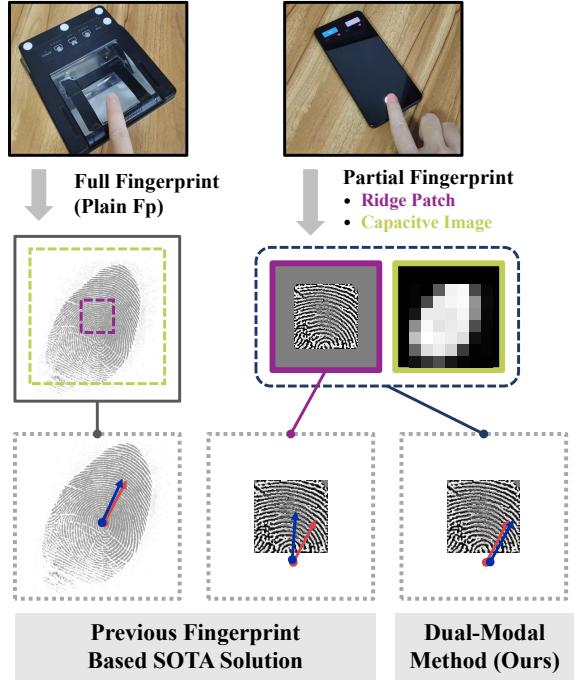


Fig. 1. Examples of fingerprint pose estimation under different input modalities. Among them, full fingerprint (plain fingerprint) is collected by a conventional optical fingerprint scanner, while the ridge patch and capacitive image are simultaneously collected from the under screen fingerprint sensor and touch screen of a smartphone (referred to as partial fingerprint collectively). All modalities, captured from the same finger with similar touch gestures, are marked in gray, purple, and green. For clarity, the dashed lines in the full fingerprint indicate the equivalent collection areas for the partial fingerprints. Subfigures in the last row represent the estimated result (blue) and ground truth (red) using corresponding modals. It can be observed that the performance of previous fingerprint based SOTA solution [10] declines significantly as the available area diminishes, while our dual-modal method achieves more accurate prediction.

Conventional approaches typically depend on special points [2], [4], [7] or specific areas [1], [5], [17], [21] to ascertain fingerprint pose. Motivated by the success of deep learning, researchers have gradually developed solutions based on neural network in recent years [6], [8]–[10], [14], [20], [22]. However, these methods are initially designed for rolled or plain fingerprints, which generally necessitate a sufficiently large effective area (about 512×512 px, 500 ppi) and small angle differences (usually less than 30°) to acquire adequate information for reliable pose estimation. In addition to fingerprints, some studies proposed predicting the three-dimensional angle of fingers from capacitive images [23]–[25], which has been proven effective when inputted at relatively small touch angles (usually within 45°). It is imperative to highlight that the mobile device recognition scenario unequivocally surpasses these constraints. Specifically, with existing under-screen fingerprint sensors, the size of captured image is substantially

reduced (about $132 \times 132\text{px}$, 500ppi), while users may attempt to unlock their devices from arbitrary touch poses, further exacerbating the challenges.

Fig. 1 shows examples of fingerprints in these mentioned modalities. To ensure terminological consistency, this paper adopts the following conventions: (1) *Full Fingerprint* refers to plain fingerprint, distinguishing it from those collected by mobile phones with limited size or resolution, (2) *Plain Fingerprint, Ridge Patch* and *Capacitive Image* denote the specific image types employed in distinct processing streams, and (3) *Partial Fingerprint* collectively describes the combined data in our experiments that includes the two modalities from smartphones. It can be intuitively seen that ridge patch and capacitive image exhibit a substantial degradation in terms of available information compared to plain fingerprint. Moreover, these two modalities from mobile devices exhibit significant differences: high resolution ridge patches contain rich local structures but limited receptive field, while low resolution capacitive images roughly represent the global contour but lacking localized details. This exciting complementarity motivates us to explore the enormous potential of modal fusion implementation. What's more, the widespread adoption of mobile devices equipped with under screen fingerprint sensors and touch screens has created an ideal ecosystem for applying this dual modal approach. These devices inherently support the simultaneous acquisition of ridge patches and capacitive images, making the proposed fusion method highly practical. Furthermore, this technology can be seamlessly integrated into existing devices through software updates, ensuring broad applicability across various scenarios. Overall, this innovative dual modal paradigm has convincing development value.

In this paper, we introduce a partial fingerprint pose estimation framework that effectively exploits such complementary strengths. The proposed approach leverages the collaborative potential of **Dual-modal** guidance from **Ridge patches And Capacitive images to Optimize** the feature extraction, fusion and representation, called **DRACO**. Different from the previous supervision forms of numerical regression [6], [8], [9], [14], [20] or heatmap voting [10], [22], we transform pose representations into decoupled quantized probability distribution, inspired by [26]–[29]. This upgrade enables our network to better grasp the relative relationships between adjacent pose spaces, resulting in impressive performance gains. We also apply the MoE mechanism to improve the feature fusion stage. A lightweight router is employed to dynamically generate adaptive weights for multiple separated feature branches, ensuring an appropriate balance of significance across different modal information. Furthermore, we leveraged the comparative relationships between groups to facilitate knowledge transfer from the high-performance plain fingerprint pose estimation domain to the target partial fingerprint domain, further strengthening the feature extraction part.

Extensive experiments were conducted on two public fingerprint databases and two private databases. The results strongly demonstrate the effectiveness of integrated strategies and mechanisms in DRACO. In addition, the proposed algorithm significantly outperforms existing SOTA pose estimation algorithms in terms of accuracy and stability. Moreover, we

also evaluated the assistance of incorporating pose information for fingerprint recognition, where our approach demonstrated consistent leading performance.

The main contributions of this work can be summarized as:

- We propose DRACO, a dual modal partial fingerprint pose estimation framework. The novel multimodality of ridge patch and capacitive image is explored and demonstrated to exhibit significant complementary advantages.
- Several simple but effective strategies and mechanisms are introduced to improve the feature extraction, fusion, and representation stages in pose estimation networks, including knowledge transfer, MoE, and decoupled probability distribution. We believe that these evolutions have substantial reference value and may provide potential inspiration for following studies.
- Extensive experiments were conducted to comprehensively evaluate the performance of DRACO and existing SOTA methods. The experimental results strongly demonstrated the superiority of our proposed approach in terms of both precision and robustness.

II. RELATED WORK

In this section, we first introduce the definition of fingerprint pose, and then review relevant finger pose estimation algorithms based on fingerprints (plain fingerprints or ridge patches) and capacitive images.

A. Definition of Fingerprint Pose

Owing to the absence of adequately distinct and consistent anatomical landmarks, the scientific community has yet to establish a clear and unified definition of fingerprint pose [11], [30]. Researchers have proposed multiple approaches to describe the center and direction of fingerprints in order to achieve approximate goals. Early studies employed the centroid and positive direction of foreground mask for pose estimation [3], [31], [32]. In addition, some approaches determined pose parameters through singular points [2], [33]. However, these approaches demonstrate unsatisfactory practicality, as their accuracy substantially depends on the completeness and quality of acquired fingerprint areas. To address these challenges, subsequent researchers proposed various fingerprint pose definitions based on special patterns along ridge orientation fields, such as points of maximum curvature [4], [34], points that match the reference templates [1], [31], or focal points perpendicular to the ridges [7], [35], [36]. Despite the improvement in accuracy, fingerprint centers under these definitions still cannot guarantee sufficient consistency for different impressions.

Further integrating these features, Yang et al. [5] defined the fingerprint direction as perpendicular to the ridge orientation around the knuckle region, and determined the center based on the type and number of singular points. On this basis, Si et al. [37] introduced a refined approach that utilizes solely the central singular point located at the northernmost as the fingerprint center, while maintaining the same directional definition. In cases where such a singular point is absent, the point exhibiting the highest curvature is designated as the

center. Subsequent researches [6], [9], [10], [17], [22] followed this form of definition. In the paper, we also adopt the same definition for consistency and comparability.

B. Pose Estimation Based on Fingerprint

Traditional methods typically estimate pose information through foreground mask information [3], [31], [32] or detecting special points [1], [2], [4], [7], [31], [33]–[36]. However, such approaches demonstrate substantial deficiencies when confronted with incomplete or highly noisy data. Yang et al. [5] introduced a pose estimation algorithm utilizing Hough voting. During the offline phase, orientation fields are extracted from high-quality image patches to build region-specific dictionaries, which are then matched with input fingerprints during the online phase, with voting in Hough space and selecting the maximum response as the result. Similarly, Yin et al. [21] constructed a dictionary of global orientation fields from aligned high-quality fingerprints and made decision through exhaustive search. Besides, Su et al. [17] employed Support Vector Machine (SVM) to build a set of classifiers for identifying fingerprint center and direction using orientation field histograms. Furthermore, Gu et al. [22] utilized the orientation field and periodic map of ridge patches as features, and subsequently predicted the center position and direction based on the Hough Forest model and SVM, respectively. Despite tangible improvements, these region-based conventional machine learning approaches still underutilize available data and achieve strong performance primarily on high-quality rolled fingerprints.

Over the past decade, deep learning based data-driven approaches have achieved impressive results across diverse domains. Ouyang et al. [6] decomposed the pose estimation task as object detection in position and classification in rotation, and introduced Faster-RCNN as the network structure. Schuch et al. [38] proposed an unsupervised learning paradigm, where a Siamese CNN is trained to predict the relative rotation between sample pairs and provide absolute angles during deployment. Yin et al. [9] proposed a multi-task network which simultaneously regresses the values of center, direction, and singular points of fingerprints. Furthermore, Arora et al. [39] suggested using two-stage prediction of core points through macro localization and micro-scale regression networks. Duan et al. [10] reformulated fingerprint pose estimation as dense prediction of grid offset vectors and employed a voting strategy. Moreover, some researchers developed several fingerprint descriptor extraction algorithms that include spatial transformation networks (STN) [8], [14], [15], [40], where the byproducts of affine transformation parameters can be considered as a form of pose representation. These deep learning approaches have demonstrated remarkable success in processing both plane and rolled fingerprint images. Nevertheless, in the scenario of partial fingerprints, the substantial loss of available information presents a critical challenge that necessitates innovative approaches for effective resolution. Fig. 1 present illustrative examples that offer qualitative validation of this perspective.

On the other hand, some studies proposed estimating relative pose from paired input fingerprints [20], [41], [42].

While the accuracy is notably improved, the trade-off involves additional relative alignment steps that must be executed separately for each comparison, significantly increasing the time cost of matching. Furthermore, these techniques necessitate the prior enrollment of sufficient fingerprint samples to facilitate the pose estimation of new impression through comparative analysis. Therefore, this paper focuses exclusively on absolute pose estimation methods which are more efficient and less dependent, leaving the discussion and research of such schemes for future studies.

C. Pose Estimation Based on Capacitive Image

There are many studies on predicting the three-dimensional angle (yaw and pitch, without roll) of fingers during touch from capacitance images. Zaliva et al. [43] introduced multiple descriptive characteristics, such as area, centroid, and average intensity, to calculate finger angles. Xiao et al. [23] further defined 42 features and used Gaussian models to regress pitch and yaw angles. Subsequent studies [24], [25] used neural networks to directly predict finger angles from capacitive images, achieving the current SOTA performance. Due to the lack of discriminative texture information, these approaches are primarily applicable for small angle inputs (less than $\pm 90^\circ$). In addition, inferring two-dimensional positions directly from low resolution contours (see Fig. 1) remains a huge challenge.

III. METHOD

In this section, we will specifically introduce our partial fingerprint pose estimation method, which utilizes a dual modal input consisting of ridge patches and capacitive images. The proposed network, named DRACO, is shown in Fig. 2. Our framework employs a three-stage pipeline: (1) initially, two parallel branches are utilized to extract distinctive features from each modality; (2) subsequently, these features undergo integration through the MoE mechanism for comprehensive fusion and collaborative guidance; (3) ultimately, DRACO generates decoupled pose representations in the form of independent one-dimensional probability distributions. Moreover, we design a knowledge transfer strategy from full fingerprints to partial fingerprints, as illustrated in Fig. 3, to help the network better comprehend and represent high-level semantic information as well as subtle differences between samples. These components will be sequentially introduced, and the corresponding loss functions are presented at the end. The construction of training data will be detailed in Section IV.

A. Feature Extraction

In this phase, two parallel encoders with homologous structures are utilized to separately derive modality-specific features from each input stream. Specifically, each branch begins with a stem that sequentially applies two consecutive groups of convolution, normalization, and activation operations. Given the success of ResNeXt-34 [44], we employed the same building blocks to construct a four-layer architecture with the configuration of [3, 4, 6, 3], while incorporating spatial and channel attention modules [45] between layers to enhance

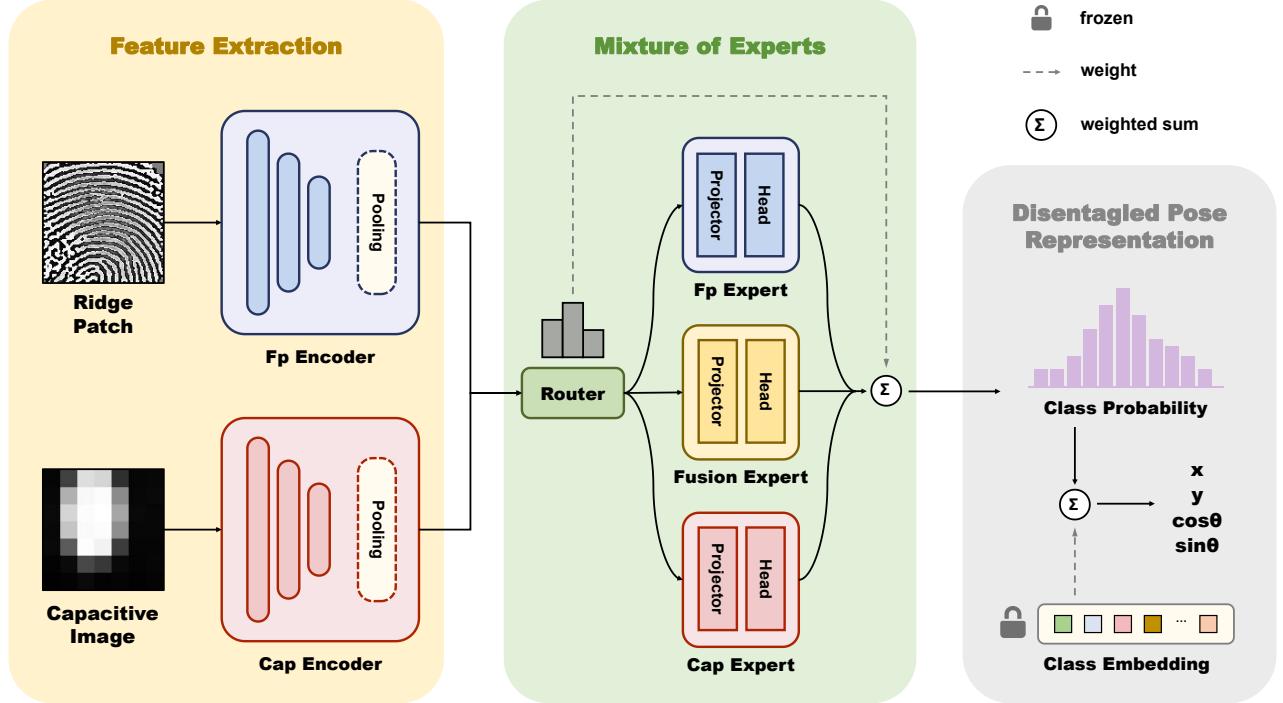


Fig. 2. An overview of our partial fingerprint pose estimation network DRACO. The ridge patch and capacitive image collected simultaneously from the touch device equipped with an under screen fingerprint sensor are input. The prediction results are represented by the horizontal and vertical coordinates of the center, as well as the sine and cosine values of the direction.

the feature representation capabilities. It should be noted that when processing high-resolution ridge patches, the stride of each layer is set to 2 to facilitate downsampling, whereas it remains at 1 for low-resolution capacitive images. The extracted features are ultimately converted into corresponding one-dimensional vectors via global average pooling.

B. Mixture of Experts

Inspired by [46]–[48], we introduced the MoE mechanism, which demonstrate significant efficacy in addressing the inherent heterogeneity across different domains. The proposed architecture integrates three specialized experts: two dedicated to processing independent feature representations and one focused on mixed feature interactions, thereby flexibly enhancing the ability of multimodal feature fusion through collaborative guidance. A straightforward router, composed of two fully connected layers, dynamically generates adaptive weights for the inference of each expert. Let f^P, f^C represent the feature vectors of corresponding branches, the processing flow can be represented as:

$$\begin{aligned} f^F &= \text{cat}(f^P, f^C), \\ w^P, w^F, w^C &= \text{Router}[f^F], \end{aligned} \quad (1)$$

where $\text{cat}(\cdot)$ corresponds to concatenation in the channel dimension. The subsequent feature enhancement and result assembly can be expressed as:

$$\begin{aligned} d^i &= \text{Head}^i[\text{Proj}^i[f^i]], \\ d &= \sum_{P,F,C} w^i \cdot d^i, \end{aligned} \quad (2)$$

where d is corresponding pose parameter. To effectively capture the high-level semantic features while alleviating gradient-related issues, we stack one linear layer and four Multilayer Perceptron (MLP) with residual connections [49] as projector. On the other hand, a single linear layer is serviced as corresponding task head.

C. Disentangled Pose Representation

Departing from conventional numerical regression [6], [8], [9], [14], [20] or heatmap voting [10], [22] approaches, we reformulate pose parameters and supervision as decoupled probability distributions, thereby providing a more robust and interpretable representation. In other words, each sample yields four pose representations as output, corresponding to the horizontal and vertical coordinates of the fingerprint center, as well as the sine and cosine of the angle. Four sets of frozen category embeddings is pre-set to provide all spatial information. With this assistance, the model only needs to describe the similarity between pose information and each category, rather than directly predicting the specific encoding. This evolution can significantly alleviate the learning complexity and enhances the generalization capability [28], [29].

In this paper, we divide the horizontal and vertical displacements (from -256 px to 256 px) into 256 segments at equal intervals, and the sine and cosine (from -1 to 1) of the angle into 120 uniform segments as frozen class embeddings. It is worth noting that trigonometric functions are used to represent fingerprint direction, instead of angle, to avoid confusion that may occur when approaching the two synonymous ends of 0° and 360° . Motivated by [26], [27], we consider the class

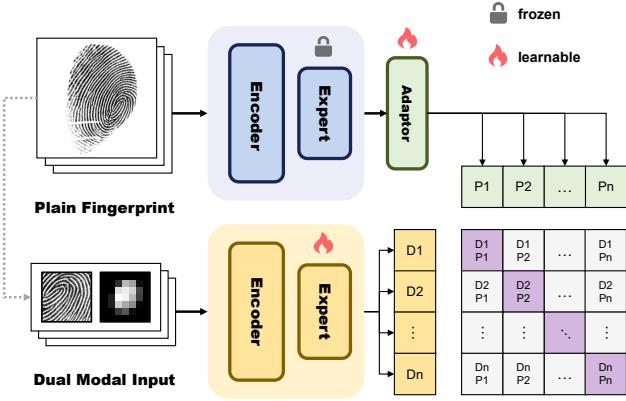


Fig. 3. The main process of knowledge transfer during network training. The blue and yellow modules corresponds to modules of the same color in Fig. 2. As shown in the bottom right corner, the features of dual modal inputs are progressively aligned with the depiction in plain fingerprints through contrastive learning techniques. The gray dashed line represents the process of data synthesis.

probability as a quantified distribution and perform weighted summation as the estimation result, rather than one-hot hard classification. For example, given the estimated probability distribution $\{d_t\}$ and class embedding $\{e_t\}$, the horizontal coordinate x of the fingerprint center can be calculated as

$$x = \frac{1}{\sum_{t=1}^n d_t} \sum_{t=1}^n d_t \cdot e_t, \quad (3)$$

where t and n correspond to the index and total number of segments, respectively. Taking this expectation helps avoid quantization errors caused by segment partitioning. The other pose parameters $y, \cos \theta$ and $\sin \theta$ are also obtained according to Equation 3. Additionally, the final direction θ is obtained by calculating the arctangent of these two trigonometric functions.

D. Knowledge Transfer

During the experiments, we observed that the same model demonstrated significantly superior performance on full fingerprint compared to partial fingerprint modalities. This phenomenon is reasonable because full fingerprints contain almost all the available one-sided features from both ridge patches and capacitive images, as well as richer and more comprehensive texture information. Therefore, we employ contrastive learning techniques [50], [51] to infuse global prior correlations (in plain fingerprint) into the current model applied to degraded modalities (partial fingerprint), leveraging a relationship-based knowledge transfer strategy [52], [53]. As shown in Fig. 3, a teacher model (blue) is pre-trained on plain fingerprints and its parameters are frozen. Next, a dual-modal student (yellow) is trained with the objective of extracting features whose relative relationships align as closely as possible with the judgments of the teacher. It should be noted that there is a strong correspondence between the dual modal input and plain fingerprint in each group, as the former is simulated from the latter and its process details are introduced in Section IV. In this paper, we use a network with the same structure as the fingerprint part in Fig. 2 as teacher, and reinforce the

corresponding feature extraction and fusion expert through this task. Specifically, a stacked 2-layer MLP is used as adapter to appropriately adjust the inherent information gap between teacher and student. Under this relationship based supervision, features from different modals of the same impression are brought as close as possible, while features from distinct impressions are intentionally separated to maximize their distance. By leveraging these relational insights, the model gains a deeper understanding of data nuances, resulting in better outcomes without any additional cost during the testing stage.

E. Loss Function

For convenience, we jointly optimize the pose estimation and knowledge transfer process in one training process. As outlined in Section III-C, we employ the distance between quantitative probability distributions associated with each parameter as the supervisory signal for pose estimation task. Let Φ represent a certain pose component, d and \tilde{d} represent the predicted result and ground truth of corresponding probability distribution, the loss function of pose estimation \mathcal{H} is defined as:

$$\mathcal{H} = \sum_{\Phi \in \{x, y, \cos, \sin\}} \lambda_\Phi \cdot \text{dist}(d_\Phi, \tilde{d}_\Phi), \quad (4)$$

where cross entropy (CE) serves as the distance metric $\text{dist}()$ between two distributions. All balance factors λ are empirically set to 1.0. The value \tilde{v} of ground truth is converted into discrete probability \tilde{d} using gaussian distribution:

$$\tilde{d}_t = \exp\left(-\frac{\tilde{v} - e_t}{2\sigma^2}\right) / \sum_t \tilde{d}_t, \quad (5)$$

where the definition of e_t is consistent with Equation 3, and the hyperparameter σ is set to 3.5 and 2.5 for position and angle sub-losses. To further augment the individual capabilities of each expert, we integrate classification heads with the same structure after each expert branch as subtasks. The total loss of the pose estimation part is

$$\mathcal{L}_{\text{pose}} = \sum_e \lambda_e \cdot \mathcal{H}_e, \quad (6)$$

where the hyperparameters λ_e corresponding to the three experts (Fp, Fusion and Cap in Fig. 2) and the comprehensive results, fixed as 0.2, 0.2, 0.4, and 1.0 respectively.

On the other hand, the Information Noise Contrastive Estimation (InfoNCE) with temperature coefficients is used to optimize the knowledge distillation process in Section III-D. Referring to Fig. 3, a plain fingerprint and the corresponding simulated dual modal images are used as the input group during training. For the features $\{P\}$ and $\{D\}$ extracted by the teacher and student networks, this relationship-based supervision is computed as:

$$z(D_i, P) = \frac{\exp(\text{sim}(D_i, D_i^+) / \tau)}{\sum_{j=1}^B \exp(\text{sim}(D_i, P_j) / \tau)}, \quad (7)$$

$$\mathcal{L}_{\text{KT}} = -\frac{1}{2B} \sum_i (\log z(D_i, P) + \log z(P_i, D)),$$

TABLE I

ALL FINGERPRINT DATASETS USED IN EXPERIMENTS. AMONG THEM, PARTIAL FINGERPRINT REFERS TO TWO MODALITIES: RIDGE PATCHES AND CAPACITIVE IMAGES.

Dataset	Type	Description	Usage	Genuine pairs ^a	Impostor pairs ^a
FVC2002 DB1_A [54]	Plain fingerprints with front pose	100 fingers × 8 impressions	test	n/a	n/a
FVC2004 DB1_A [55]	Plain fingerprints with front pose	100 fingers × 8 impressions	test	n/a	n/a
DPF [10]	Rolled fingerprints	933 fingers × 1 impression	calibration	n/a	n/a
PCF	Plain fingerprints with diverse poses ^b	933 fingers × 3.1 impressions ^c	train & test	40,579	4,157,822
	Rolled fingerprints	100 fingers × 1 impression	calibration	n/a	n/a
	Partial fingerprints with diverse poses	100 fingers × 32 impression	finetune & test	46,338	3,413,262

^a Effective pairs in matching experiments.

^b Simultaneously, partial fingerprints are simulated based on plain fingerprints.

^c Average number after screening.

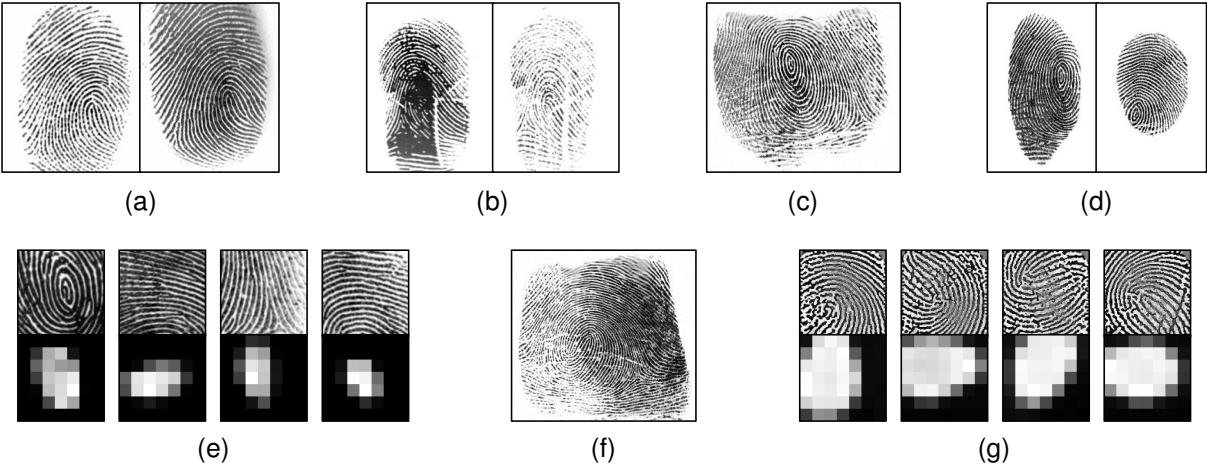


Fig. 4. Image examples from different datasets: (a) FVC2002 DB1_A [54], (b) FVC2004 DB1_A [55], (c) DPF [10], rolled fingerprint, (d) DPF [10], plain fingerprints, (e) DPF [10], simulated partial fingerprints, (f) PCF, rolled fingerprints, (g) PCF, partial fingerprints. The ‘partial fingerprint’ is a general term used to represent two modalities: ridge patch and capacitive image.

where B is the batch size, sim is the cosine similarity. The temperature τ is set to 8.0 according to the results of small-scale parameter search.

The final loss is computed as the weighted sum of aforementioned two tasks:

$$\mathcal{L} = \mathcal{L}_{\text{pose}} + \lambda \cdot \mathcal{L}_{\text{KT}}. \quad (8)$$

The balance term λ is configured to 1.0.

IV. DATASET

A. Dataset Introduction

The utilization information of all datasets are listed in Table I. Moreover, Fig. 4 shows corresponding image examples. Considering that most existing public datasets are predominantly collected in frontal poses and lack diversity in terms of location, we only used two common datasets (*FVC2002 DB1_A* [54] and *FVC2004 DB1_A* [55]) as representatives. Private *DPF* database [10] explicitly require subjects to adopt diverse poses during collection, making it more suitable for this task. Specifically, 8,479 samples from 744 fingers were used for training, while the remaining 2,049 samples (from other 189 fingers) were allocated for testing. In addition, we collected and established a real partial fingerprint dataset using

a smartphone equipped with an underscreen fingerprint sensor. In this paper, it referred to as the Phone Captured Fingerprint and Rolled Fingerprint Database (*PCF*). Ridge patches with a size of 132×132 px in 500 ppi, as well as capacitive images with 7×7 px (effective area) in 10 ppi, can be obtained simultaneously. A total of 640 images from 20 fingers were used to fine-tune the model, and 2,560 images from other 80 fingers were used for testing. In addition, subjects in *DPF* and *PCF* were also required to collect rolled fingerprints, which were used for subsequent calibration to obtain pose ground truth.

B. Training Set Construction

We use the subset of *DPF* mentioned above for training. For full fingerprint scenarios, we directly utilize the samples displayed in Fig. 4(d) for training. The ground truth is obtained by minutiae matching between the input plain fingerprint and its corresponding pose-standardized rolled fingerprint. In this paper, we employ VeriFinger SDK 12.0 [56] for minutiae extraction and matching, and adopt the method proposed by Duan et al. [10] for rolled fingerprint pose rectification. These approaches are proven to be sufficiently reliable under conditions of high image quality and small pose variation.

Therefore, we approximately consider it as an unbiased and precise calibration.

Due to the scarcity of readily available large-scale partial fingerprint data and capacitive images, we generated approximate samples from the plain fingerprints of *DPF*. Following [41], [42], [57], we randomly cropped square regions from the plain fingerprint to simulate ridge patches. On the other hand, window-based uniform filtering and interpolation are applied to downsample the original image to 10 ppi, emulating the capacitive image [57]. The setting of cropping size and target resolution is to maintain consistency with the real data in *PCF*. Nevertheless, we fine-tuned our model using a small amount of local data before testing on *PCF* to minimize domain bias as much as possible.

C. Test Set Protocol

To comprehensively evaluate the accuracy and robustness of pose estimation algorithms, we generated four distinct test sets for each scenarios. Specifically, the pose of samples in *FVC2002 & FVC2004 DB1_A* and *DPF* are first standardized and then randomly rotated within the four ranges of $\pm 45^\circ$, $\pm 90^\circ$, $\pm 135^\circ$ and $\pm 180^\circ$. Each evaluated method will be trained and tested separately based on the range of direction angle under full fingerprint or partial fingerprint scenario. For simplicity, the training conditions of each model will not be specifically declared in experiments. Additionally, the real partial fingerprint data *PCF* remains unchanged in order to accurately provide feedback on the real environment. The ground truth of its pose information is obtained through the same calibration process introduced in Section IV-B. In the matching experiments, pairs with the same identity will be checked in advance to exclude situations where there is no overlapping area. The effective number of genuine and impostor pairs is presented in Table I.

V. EXPERIMENTS

In experiments, we compare the proposed DRACO with SOTA finger pose estimation algorithms on full fingerprints (plain fingerprints) and partial fingerprints (ridge patches and capacitive images). The implementation details of DRACO are provided first, followed by an introduction to the representative methods used for comparison. The performance of these algorithms is then thoroughly evaluated in terms of pose estimation and matching capabilities across various rotation ranges. This assessment ensures a comprehensive understanding of how each algorithm performs under different conditions, highlighting their strengths and weaknesses in handling pose variations. In addition, extensive ablation experiments are conducted to validate the effectiveness of our proposed modules and strategies, while also offering potential inspiration for future works. Finally, the efficiency of different algorithms is reported to assess their deployment costs.

A. Implementation Details

Our proposed DRACO are trained under the corresponding modality in *DPF* with an initial learning rate of $1e-3$ (end of

$1e-6$), cosine annealing scheduler, default AdamW optimizer and batch size of 256 for 80 epochs. Data augmentation is used, including random translation within ± 40 pixels and random rotation of $\pm 45^\circ$, $\pm 90^\circ$, $\pm 135^\circ$, and $\pm 180^\circ$ (according to the corresponding test scenario). Specifically, when incorporating contrastive learning, we increase the batch size to 512 to ensure sample richness. The learning rate and epoch number is adjusted to $4 \times$ and 200 to roughly maintain the original optimization iterations. Before testing on *PCF*, the parameters under $\pm 180^\circ$ augmentation are loaded and further fine-tuned with an initial learning rate of $1e-4$ (end of $1e-5$) for 200 epochs, and keep other parameters consistent. When DRACO is applied in a single modal, only the corresponding feature extractor and expert shown in Fig. 2 are activated. In following experiments, we used suffixes ‘fp’ and ‘cap’ to distinguish models with DRACO structures that only use branch of ridge patch and capacitive image, respectively.

B. Compared Methods

For plain fingerprint and ridge patch, we reproduced four representative methods, including:

- **Faster-RCNN**: Object detection network proposed by Ouyang et al. [6];
- **STN**: STN module under indirect supervision of fixed-length representation task [8], [14];
- **JointNet**: Numerical regression network proposed by Yin et al. [9];
- **GridNet**: Heatmap voting network recently proposed by Duan et al. [10].

Additionally, for the capacitive image modality, since existing researches focus on predicting angles in 3D space, we reimplement the following approaches and adjust the output head to estimate the center and rotation value of 2D pose:

- **Cap-MLP**: Regressor based on manually defined features and multi-layer MLP, inspired by Xiao et al. [6];
- **Cap-CNN**: Numerical regression network inspired by recent works [24], [25].

Above methods are also trained under the combinations of settings in DRACO to ensure sufficient fairness in the comparison.

C. Pose Estimation Performance

Although the focus of this paper is on partial fingerprint pose estimation, our disentangled pose representation strategy is universal. Therefore, we first compare representative methods with our proposed DRACO in full fingerprint scenario, and then provide detailed evaluation on partial fingerprint (including partial fingerprint and capacitive image) pose estimation.

1) *Evaluation on Full Fingerprints*: Consistent with previous works [9], [10], the deviation in location and direction of mated minutiae pairs on *FVC2002 DB1_A* and *FVC2004 DB1_A* is presented to characterize the accuracy of pose estimation. Specifically, each method infers the 2D pose of single plain fingerprint, which is then utilized to execute a rigid transformation and align the image to standard coordinate system. Subsequently, VeriFinger [56] is used to extract and

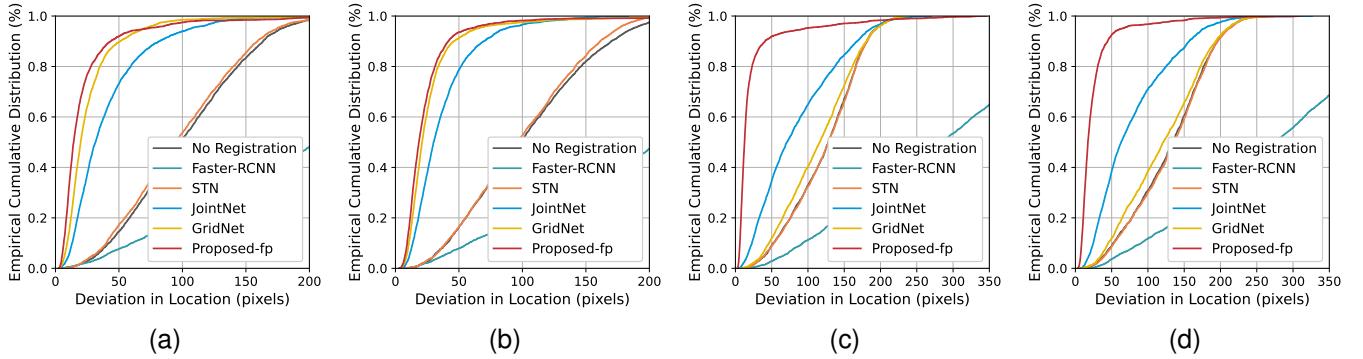


Fig. 5. The empirical cumulative distribution of location deviations on full fingerprints from (a) FVC2002 DB1_A $[-90^\circ, 90^\circ]$, (b) FVC2004 DB1_A $[-90^\circ, 90^\circ]$, (c) FVC2002 DB1_A $[-180^\circ, 180^\circ]$, (d) FVC2004 DB1_A $[-180^\circ, 180^\circ]$. Suffixes ‘fp’ indicate that only the corresponding branch is used.

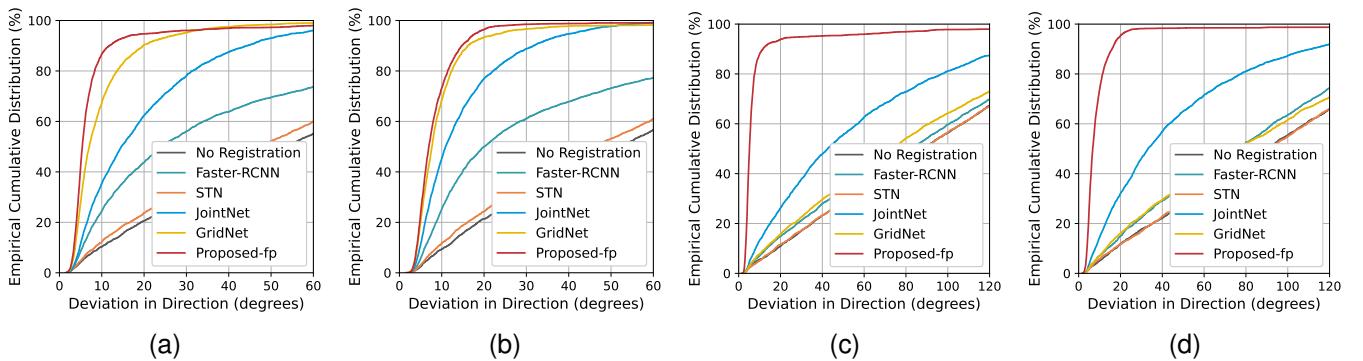


Fig. 6. The empirical cumulative distribution of direction deviations on full fingerprints from (a) FVC2002 DB1_A $[-90^\circ, 90^\circ]$, (b) FVC2004 DB1_A $[-90^\circ, 90^\circ]$, (c) FVC2002 DB1_A $[-180^\circ, 180^\circ]$, (d) FVC2004 DB1_A $[-180^\circ, 180^\circ]$. Suffixes ‘fp’ indicate that only the corresponding branch is used.

match minutiae between impressions of the same identity. The average deviation of Euclidean distance and absolute rotation error between paired points is visualized through empirical cumulative score function. As shown in Fig. 5 and Fig. 6, the proposed DRACO exhibits significant advantages across all test datasets, especially when dealing with large rotation angles. The performance of *STN* [8], [14] is close to no registration, primarily because pose estimation is not directly supervised. This absence of explicit guidance may impede the network’s ability to learn accurate transformations. In addition, the object detection based *Faster-RCNN* [6] shows limited performance in estimating fingerprint position. A compelling explanation is that in this task, the network primarily focuses on the outer contours, which may lead to neglect of the precise center.

Furthermore, we evaluated the performance of these algorithms across four different rotation ranges on plain fingerprints of *DPF*. The mean absolute error of translation and rotation is reported in Table A.1. It can be seen that the performance of previous SOTA methods significantly declines as the rotation range increases. In contrast, our method demonstrates exceptional precision and robustness across the entire rotation range. Based on these experimental results, we selected the best-performing algorithms, *JointNet* [9] and *GridNet* [10], for the subsequent experiments on partial fingerprints.

2) Evaluation on Partial Fingerprints: Similarly, we assessed the performance of pose estimation algorithms on partial fingerprints (partial fingerprints and capacitive images) using the simulated test set from *DPF*. As demonstrated in Table II, our proposed solution outperforms previous methods in respective unimodal groups. This result strongly emphasizes the advantages of our pose representation scheme once again. In addition, the full version of DRACO showcases comprehensive and notable leadership in both accuracy and stability following the integration of dual modal information. This substantial improvement clearly demonstrates the considerable complementarity between ridge patches and capacitive images, affirming the effectiveness of the collaborative dual-modal guidance paradigm. Table III presents the key comparative results on *PCF*, which also highlight the impressive success of our proposed algorithm. The overall performance on *PCF* is somewhat inferior to the simulation dataset of *DPF*, possibly due to domain bias and the limited fine-tuning data (640 samples during finetuning, which is significantly smaller than the 8,479 samples in training stage). However, this experiment still provides valuable insights into evaluating the relative performance of pose estimation algorithms, which is our primary concern.

3) Visual Analysis: Several intuitive examples are provided to qualitatively compare different pose estimation algorithms. Fig. 7 shows three representative visualization results. It can be

TABLE II

ALIGNMENT ERROR UNDER DIFFERENT FINGERPRINT POSES ON DPF (PARTIAL FINGERPRINTS WITH DIFFERENT ROTATION RANGES). THE FOUR GROUPS, FROM TOP TO BOTTOM, REPRESENT DEFAULT CONFIGURATION AND METHODS THAT USE ONLY RIDGE PATCHES, ONLY CAPACITIVE IMAGES, AND A COMBINATION OF BOTH. SUFFIXES ‘FP’ AND ‘CAP’ INDICATE THAT ONLY THE CORRESPONDING BRANCH IS USED. BOLD AND UNDERLINED NUMBER REPRESENT THE CORRESPONDING GLOBAL OPTIMAL RESULT AND OPTIMAL METHOD IN EACH GROUP RESPECTIVELY.

Method	[-45°, 45°]		[-90°, 90°]		[-135°, 135°]		[-180°, 180°]	
	trans (px)	rot (°)	trans (px)	rot (°)	trans (px)	rot (°)	trans (px)	rot (°)
No Registration	83.4	22.7	85.5	45.5	85.1	67.4	88.3	90.8
JointNet [9]	34.3	17.3	36.1	18.7	37.3	27.9	41.4	34.5
GridNet [10]	35.2	12.0	44.2	19.6	47.3	37.8	74.8	64.5
Proposed-fp	<u>23.4</u>	<u>10.3</u>	<u>22.6</u>	<u>11.7</u>	<u>26.8</u>	<u>15.1</u>	<u>25.4</u>	<u>14.1</u>
Cap-MLP [23]	79.7	13.3	79.3	39.2	83.5	65.5	86.2	91.1
Cap-CNN [24], [25]	67.8	7.8	<u>66.2</u>	16.0	71.7	58.5	78.3	70.4
Proposed-cap	<u>65.0</u>	<u>7.1</u>	67.8	<u>12.9</u>	<u>70.8</u>	<u>49.7</u>	<u>75.7</u>	<u>68.1</u>
Proposed	18.4	4.8	19.5	5.4	20.3	5.5	19.8	5.5

TABLE III

ALIGNMENT ERROR UNDER DIFFERENT FINGERPRINT POSES ON PARTIAL FINGERPRINTS FROM PCF. THE GROUPING RULES ARE THE SAME AS TABLE II.

Method	trans (px)	rot (°)
No Registration	98.4	82.0
JointNet [9]	51.5	38.0
GridNet [10]	87.2	63.2
Cap-MLP [23]	94.8	75.9
Cap-CNN [24], [25]	90.6	68.2
Proposed	31.2	16.3

observed that fingerprint based methods (*JointNet* [9], *GridNet* [10]) function effectively when the texture features of ridge patches possess sufficient discriminability (line 1). On the other hand, it is possible to accurately infer angles using only capacitive images (*Cap-MLP* [23], *Cap-CNN* [24], [25]), which is significantly ahead of relying solely on ridge patches (line 2). However, capacitive image based methods expose obvious deficiencies in localization. The respective advantages and limitations of these two modals effectively highlight their complementarity. Naturally, our method, guided by dual-modal collaboration, demonstrates more precise performance (line 1 & 2). Even when previous solutions have completely failed, it can still provide accurate predictions (line 3).

We further illustrate some failure cases of DRACO in Fig. 8. Under certain extreme pressing postures, such as those involving fingertips (column 1 & 2), the model may encounter substantial pose estimation errors. Additionally, in rare instances where both ridge patches and capacitive images lack sufficient recognition, DRACO may occasionally experience considerable confusion and misjudgment. (column 3).

D. Matching Performance

Fingerprint pose estimation serves as a valuable source of auxiliary information in matching tasks [10], [11], [17]. For instance, an impression taken from the left side of a finger should not be considered a successful match with any impression from the right side, no matter how similar they are. According to this logic, recognition systems can swiftly identify and

discard candidate samples that exhibit incompatible positions and orientations. This selective filtering significantly reduces the search space, leading to enhanced matching accuracy and efficiency in the overall process. In experiments, we use VeriFinger [56] as a representative keypoint based matcher, which provides high quality minutiae extraction and matching functions. Samples with pose differences greater than an optimal threshold (determined through exhaustive parameter search) will be excluded in advance. In other words, the comparison scores of these detected abnormal situations are set to infinitely small.

On the other hand, we also assessed the impact of pose rectification on recognition schemes using fixed-length representations. The current SOTA method FDD [15] is utilized on behalf of these approaches. In the recognition process, the input image is first rigidly transformed based on corresponding estimated pose. Subsequently, one-dimensional representation vectors are extracted and the matching similarity between pairs is calculated.

1) *Evaluation on Fingerprint Verification:* In line with Section V-C, we evaluate the performance of pose estimation methods on both plain and partial fingerprints. This enables us to thoroughly evaluate their effectiveness across different situations. The results at different rotation ranges on *DPF* are reported in Table A.2 and IV, respectively. Additionally, Table V presents a further comparison on *PCF*. When the pose difference is small, *GridNet* [10] exhibits certain advantages. As the rotation angle increases, the performance of *JointNet* [9] becomes more stable. In more challenging scenarios involving partial fingerprints, previous works, whether based on ridge patches or capacitive images, have demonstrated unsatisfactory performance. At the same time, our proposed DRACO showcases impressive and comprehensive advancements. We attribute it to the decoupled pose representation and collaborative dual-modal guidance developed in this paper. It is worth mentioning that fixed-length representation based matcher [15] still has a gap in partial fingerprints compared to minutiae based solutions [56]. Nevertheless, the introduction of our pose estimation method shows a significant relative improvement and shows attractive potential for future refinements.

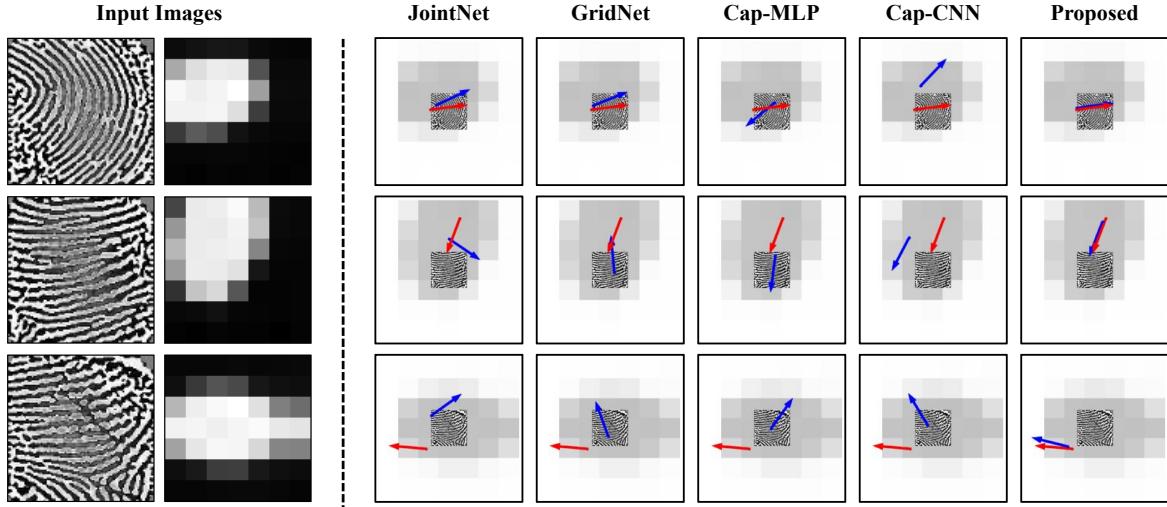


Fig. 7. Examples of different pose estimation methods on **partial fingerprints** from *PCF*. To facilitate observation, the capacitive image is inverted and overlaid as background onto the corresponding ridge patch, displaying both the prediction result (blue arrow) and ground truth (red arrow).

TABLE IV

VERIFICATION PERFORMANCE (%) ON DPF (**PARTIAL FINGERPRINTS**) WITH DIFFERENT ROTATION RANGES) USING DIFFERENT FINGERPRINT POSES. THE FOUR GROUPS, FROM TOP TO BOTTOM, REPRESENT DEFAULT CONFIGURATION AND METHODS THAT USE ONLY RIDGE PATCHES, ONLY CAPACITIVE IMAGES, AND A COMBINATION OF BOTH.

Matcher	Method	[-45°, 45°]			[-90°, 90°]			[-135°, 135°]			[-180°, 180°]		
		EER	FNMR ¹	FNMR ²	EER	FNMR ¹	FNMR ²	EER	FNMR ¹	FNMR ²	EER	FNMR ¹	FNMR ²
VeriFinger [56]	No Registration	2.24	4.39	13.94	2.06	3.88	14.81	2.24	4.74	15.14	2.32	5.19	17.17
	JointNet [9]	2.13	3.84	13.08	2.07	3.70	14.19	2.60	4.81	15.74	2.87	5.71	18.08
	GridNet [10]	2.06	3.42	13.12	2.02	3.59	14.10	2.78	5.23	16.08	3.71	7.01	19.19
	Cap-MLP [23]	2.24	3.80	13.44	2.35	4.50	15.37	6.56	10.82	20.68	6.31	10.21	20.93
	Cap-CNN [24], [25]	2.04	3.40	13.11	2.12	4.05	14.97	4.80	8.64	18.74	4.56	8.04	20.16
FDD [15]	Proposed	1.27	1.61	9.50	1.11	1.30	8.41	1.13	1.26	8.92	0.96	0.92	9.09
	No Registration	25.43	67.77	85.32	36.67	84.20	93.53	44.30	90.29	96.03	44.05	92.49	97.46
	JointNet [9]	20.72	53.97	74.40	17.42	42.98	62.33	23.74	56.50	74.82	28.90	67.27	83.11
	GridNet [10]	14.34	38.54	58.24	19.04	47.84	66.17	31.50	69.64	82.82	40.46	85.99	93.74
	Cap-MLP [23]	23.82	58.48	76.88	36.05	82.12	91.37	44.62	88.56	95.10	44.15	90.60	96.58
	Cap-CNN [24], [25]	23.66	59.14	78.27	27.12	61.97	79.13	43.33	88.71	95.31	44.14	92.18	97.34
	Proposed	5.70	16.71	37.41	6.40	19.48	40.47	6.75	20.39	44.14	6.87	21.37	46.42

¹ FNMR@FMR=1e-3, ² FNMR@FMR=1e-4.

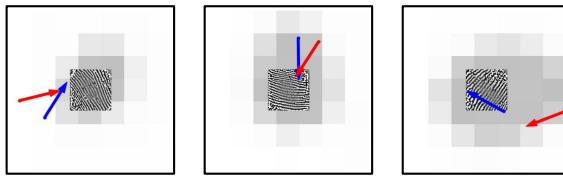


Fig. 8. Failure cases of DRACO on **partial fingerprints** from *PCF*. The visualization protocol is consistent with Fig. 7.

2) *Evaluation on Fingerprint Indexing*: Similarly, we further examine the role of fingerprint pose estimation algorithms in the indexing system. Experiments are conducted solely on the minutiae based matcher, as this approach demonstrates superior performance and is the most commonly used in related works [11]. Fig. A.1 and 9 illustrate the indexing performance for full fingerprint and partial fingerprint scenarios

on *DPF*, respectively, across different rotation ranges. The results on *PCF* are shown in Fig. 10. Comparisons in these curves indicate that appropriate pose estimation can effectively improve the indexing accuracy. In scenarios with a large rotation range or restricted effective areas, previous works face greater challenges and even have negative impacts in some cases. Notably, our method excels as the most stable and accurate across all scenarios, which shows sufficient positive effects in all tests.

E. Ablation Study

1) *Pose Representation Form*: We begin by thoroughly investigating the impact of pose representation and supervision forms on plain fingerprints of *DPF*. Table VI presents the ablation study for rotation as a representative example. For the regression task head, two specific expressions are compared: (1) *ang*: directly predicting the angle, and (2) *tan*:

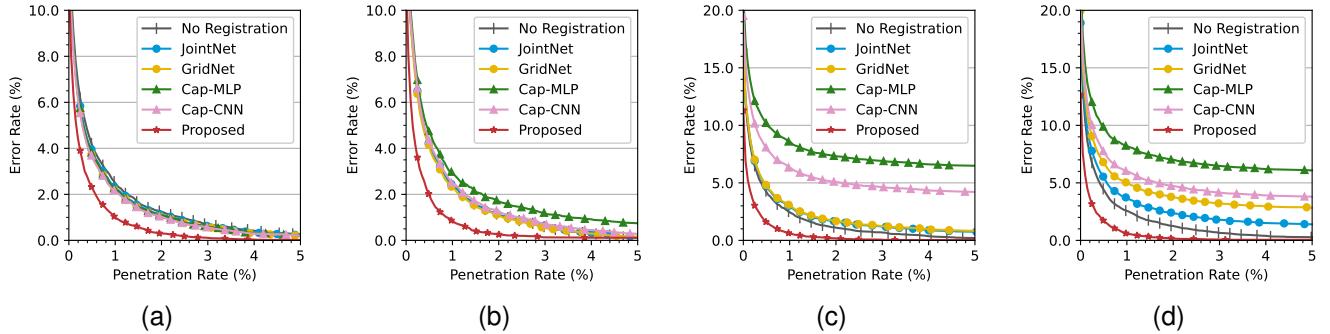


Fig. 9. The VeriFinger [56] based fingerprint indexing performance with corresponding pose constraint on partial fingerprints from DPF under different rotation ranges: (a) $[45^\circ, 45^\circ]$, (b) $[90^\circ, 90^\circ]$, (c) $[135^\circ, 135^\circ]$, (d) $[180^\circ, 180^\circ]$. Different input modalities are distinguished by the shape of markers.

TABLE V
VERIFICATION PERFORMANCE (%) ON PARTIAL FINGERPRINTS FROM PCF. THE GROUPING RULES ARE THE SAME AS TABLE IV.

Matcher	Method	EER	FNMR ¹	FNMR ²
VeriFinger [56]	No Registration	5.69	17.51	33.40
	JointNet [9]	5.90	16.53	33.65
	GridNet [10]	7.21	18.77	35.31
	Cap-MLP [23]	8.18	20.33	36.75
	Cap-CNN [24], [25]	7.85	19.72	36.19
	Proposed	4.09	12.07	30.08
FDD [15]	No Registration	45.62	89.55	95.04
	JointNet [9]	34.08	82.27	92.66
	GridNet [10]	41.27	89.65	95.50
	Cap-MLP [23]	44.01	91.53	96.80
	Cap-CNN [24], [25]	43.81	91.31	96.15
	Proposed	17.92	56.01	76.54

¹ FNMR@FMR=1e-3, ² FNMR@FMR=1e-4.

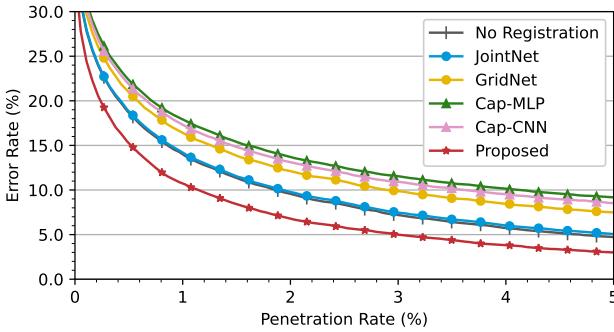


Fig. 10. The VeriFinger [56] based fingerprint indexing performance with corresponding pose constraint on partial fingerprints from PCF. Different input modalities are distinguished by the shape of markers.

predicting sine and cosine values and indirectly calculating the angle through the arctangent function. Similarly, two types of mean squared error (MSE) are used as available losses. The results indicate that representing angles using trigonometric functions yields improved performance, which is consistent with the explanation in Section III-C. On this basis, we further transform the task into probability estimation and explore two classification heads for prediction: (1) *max*: the

TABLE VI
ABLATION STUDY (°) OF POSE REPRESENTATION FORM ON DPF (FULL FINGERPRINTS WITH ROTATION RANGE OF $[-180^\circ, 180^\circ]$).

Loss\Head	Regressor		Classifier	
	ang	tan	max	sum
MSE (ang)	13.2	7.9	-	7.3
MSE (tan)	12.3	7.1	-	3.3
JS	\	\	2.2	2.0
CE	\	\	2.0	1.8

'\ indicates that corresponding item is not applicable.

'-' indicates that corresponding process does not converge.

highest response across all categories is directly selected as the result, and (2) *sum*: the quantified probability distribution and class embedding are weighted and summed as Equation 3. Two classic distribution metrics, cross entropy (CE) and Jensen-Shannon divergence (JS), are supplemented as candidate supervisors. The comparison strongly demonstrates the superiority of our proposed pose representation of probability distribution forms and verifies the rationality of the analysis in Section III-C.

2) *Feature Extraction & Fusion*: Furthermore, the mechanisms and modules proposed in this paper for modal fusion and feature extraction enhancement are verified in Table VII. In the first group of experiments, the significant complementarity between ridge patches and capacitive images is reconfirmed. On this basis, we compare different fusion strategies in the second group. Three typical fusion schemes are evaluated, including (1) *E. MoE*: treating each expert as equally important, with their results directly summed, (2) *F. MoE*: establishing a set of learnable parameters as globally fixed weights assigned to the experts, (3) *A. MoE*: dynamically assigning sample-specific adaptive weights to different experts through the router depicted in Fig. 2. Strategy (3) outperforms the others, which is understandable as it clearly exhibits the highest flexibility and generalization ability.

The aim of the last comparison group is to assess the effectiveness of knowledge transfer concept introduced in Section III-D. The supervision scheme based on features and responses [58] is additionally examined. From the third and fourth lines to the bottom, it can be seen that directly approximating

TABLE VII

ABLATION STUDY OF FEATURE EXTRACTION AND FUSION ON DPF
 PARTIAL FINGERPRINTS WITH ROTATION RANGE OF $[-180^\circ, 180^\circ]$.
 THE GRAY BACKGROUND INDICATES THE OPTIMAL STRATEGY
 COMBINATION WITHIN EACH GROUP, WHICH SERVES AS THE BASIS FOR
 THE FOLLOWING GROUPS.

Used Modal	Fusion	Knowledge	Avg. Error		
Ridge	Cap	Strategy [†]	Transfer	trans (px)	rot (°)
✓		\	\	25.4	14.1
	✓	\	\	75.7	68.1
✓	✓	\	\	23.5	8.7
✓	✓	E. MoE	\	22.4	7.7
✓	✓	F. MoE	\	22.9	7.3
✓	✓	A. MoE	\	21.3	6.8
✓	✓	A. MoE	Feature	21.0	6.5
✓	✓	A. MoE	Response	21.9	6.8
✓	✓	A. MoE	Relation	19.8	5.5

[†] Abbreviations ‘E.’, ‘F.’ and ‘A.’ respectively represent Equal, Fixed-Weight and Adaptive.

TABLE VIII

MODEL SIZE AND AVERAGE TIME COST OF DIFFERENT FINGERPRINT POSE ESTIMATION ALGORITHMS WHEN PROCESSING PCF. METHODS IN EACH GROUP USE DIFFERENT MODAL INPUT.

Method	Param (M)	Times (ms)
JointNet [9]	5.30	19.7
GridNet [10]	14.2	16.8
CNN-MLP [23]	0.17	3.6
CNN-Cap [24], [25]	1.70	4.4
Proposed	9.65	26.1

numerical values from the feature space proved ineffective, likely due to the significant domain difference between full and partial fingerprints. The response based approach even has adverse effects. A convincing explanation is that it disrupts the probability distribution introduced by Equation 6, which may lead to inaccurate representations. Finally, knowledge transfer based on comparative relationships highlights the structured connections between samples, effectively incorporating higher-level semantic information and enhancing model performance.

F. Efficiency

Model size and inference speed of different fingerprint pose estimation algorithms on *PCF* are listed in Table VIII. The time covers a complete process from inputting a sample to outputting the corresponding pose information, which is measured on a single NVIDIA GeForce RTX 3090 GPU by setting the batch size to 1, with an Intel Xeon E5-2680 v4 CPU @ 2.4 GHz. All algorithms are implemented in Python (Pytorch). It can be seen that our method exhibits comparable efficiency while delivering high estimation performance, thereby highlighting its attractive practical value.

VI. CONCLUSION

In this paper, we propose DRACO, a novel partial fingerprint pose estimation method under dual-modal collaborative guidance of ridge patches and capacitive images, which are

captured by under-screen fingerprint sensor and touch sensors of smartphones. Unlike previous single modal based approaches, we demonstrate the strong complementarity between these two modalities and present an effective framework to integrate and leverage their combined strengths. Specifically, relationship based knowledge transfer and MoE strategies are employed to enhance the network’s feature extraction and fusion capabilities. Furthermore, we reformulate fingerprint pose representation as a decoupled probability distribution, significantly improving prediction accuracy. Extensive experiments on multiple databases show that DRACO surpasses state-of-the-art methods in both precision and robustness. Future work will investigate deeper integration of fingerprint pose estimation with other related tasks and downstream processes, particularly improving its synergy with feature extraction and matching algorithms.

REFERENCES

- [1] A. Jain, S. Prabhakar, L. Hong, and S. Pankanti, “Filterbank-based fingerprint matching,” *IEEE Transactions on Image Processing*, vol. 9, no. 5, pp. 846–859, 2000.
- [2] K. Nilsson and J. Bigun, “Localization of corresponding points in fingerprints by complex filtering,” *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2135–2144, 2003, audio- and Video-based Biometric Person Authentication (AVBPA 2001).
- [3] R. Cappelli and D. Maltoni, “On the spatial distribution of fingerprint singularities,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 742–448, 2009.
- [4] S. Yoon, K. Cao, E. Liu, and A. K. Jain, “LFIQ: Latent fingerprint image quality,” in *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2013, pp. 1–8.
- [5] X. Yang, J. Feng, and J. Zhou, “Localized dictionaries based orientation field estimation for latent fingerprints,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 5, pp. 955–969, 2014.
- [6] J. Ouyang, J. Feng, J. Lu, Z. Guo, and J. Zhou, “Fingerprint pose estimation based on faster R-CNN,” in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, 2017, pp. 268–276.
- [7] C. Deerada, K. Phromsuthirak, A. Rungchokanun, and V. Areekul, “Progressive focusing algorithm for reliable pose estimation of latent fingerprints,” *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 1232–1247, 2020.
- [8] 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.
- [9] Q. Yin, J. Feng, J. Lu, and J. Zhou, “Joint estimation of pose and singular points of fingerprints,” *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1467–1479, 2021.
- [10] 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.
- [11] D. Maltoni, D. Maio, A. K. Jain, and J. Feng, *Handbook of Fingerprint Recognition*. Cham: Springer International Publishing, 2022.
- [12] 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.
- [13] X. Guan, Y. Duan, J. Feng, and J. Zhou, “Regression of dense distortion field from a single fingerprint image,” *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 4377–4390, 2023.
- [14] S. A. Grosz and A. K. Jain, “AFR-Net: Attention-driven fingerprint recognition network,” *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 6, no. 1, pp. 30–42, 2024.
- [15] Z. Pan, Y. Duan, J. Feng, and J. Zhou, “Fixed-length dense descriptor for efficient fingerprint matching,” in *2024 IEEE International Workshop on Information Forensics and Security (WIFS)*, 2024, pp. 1–6.
- [16] 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.
- [17] Y. Su, J. Feng, and J. Zhou, “Fingerprint indexing with pose constraint,” *Pattern Recognition*, vol. 54, pp. 1–13, 2016.

- [18] K. Cao and A. K. Jain, "Fingerprint indexing and matching: An integrated approach," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, 2017, pp. 437–445.
- [19] 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.
- [20] 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.
- [21] Q. Yin, J. Feng, J. Lu, and J. Zhou, "Orientation field estimation for latent fingerprints by exhaustive search of large database," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2018, pp. 1–9.
- [22] S. Gu, J. Feng, J. Lu, and J. Zhou, "Efficient rectification of distorted fingerprints," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 1, pp. 156–169, 2018.
- [23] R. Xiao, J. Schwarz, and C. Harrison, "Estimating 3D finger angle on commodity touchscreens," in *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces*, ser. ITS '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 47–50.
- [24] S. Mayer, H. V. Le, and N. Henze, "Estimating the finger orientation on capacitive touchscreens using convolutional neural networks," in *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*, ser. ISS '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 220–229.
- [25] K. He, C. Li, Y. Duan, J. Feng, and J. Zhou, "TrackPose: Towards stable and user adaptive finger pose estimation on capacitive touchscreens," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 7, no. 4, p. 1–22, Jan. 2024.
- [26] Y. Li, S. Zhang, Z. Wang, S. Yang, W. Yang, S.-T. Xia, and E. Zhou, "TokenPose: Learning keypoint tokens for human pose estimation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021, pp. 11 313–11 322.
- [27] J. Li, S. Bian, A. Zeng, C. Wang, B. Pang, W. Liu, and C. Lu, "Human pose regression with residual log-likelihood estimation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021, pp. 11 025–11 034.
- [28] Y. Li, S. Yang, P. Liu, S. Zhang, Y. Wang, Z. Wang, W. Yang, and S.-T. Xia, "SimCC: A simple coordinate classification perspective for human pose estimation," in *Computer Vision – ECCV 2022*, S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, Eds. Cham: Springer Nature Switzerland, 2022, pp. 89–106.
- [29] T. Jiang, P. Lu, L. Zhang, N. Ma, R. Han, C. Lyu, Y. Li, and K. Chen, "RTMPose: Real-time multi-person pose estimation based on mmpose," *arXiv preprint arXiv:2303.07399*, 2023.
- [30] K. C. Mangold, "Data format for the interchange of fingerprint, facial & other biometric information ansi/nist-itl 1-2011 nist special publication 500-290 edition 3," 2016-08-21 04:08:00 2016.
- [31] G. T. Candela, "PCASYS - A pattern-level classification automation system for fingerprints," *National Institute of Standards and Technology, NISTIR 5647*, 1995.
- [32] C. Watson, P. Flanagan, and B. Cochran, *SlapsSegII-Slap Fingerprint Segmentation Evaluation II*. US Department of Commerce, National Institute of Standards and Technology, 2009.
- [33] A. M. Bazeen and S. H. Gerez, "Segmentation of fingerprint images," in *ProRISC 2001 Workshop on Circuits, Systems and Signal Processing*. Citeseer, 2001, pp. 276–280.
- [34] M. Liu, X. Jiang, and A. C. Kot, "Fingerprint reference-point detection," *EURASIP Journal on Advances in Signal Processing*, vol. 2005, pp. 1–12, 2005.
- [35] K. Kerkrui and V. Areekul, "A new reference point for fingerprint recognition," in *Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101)*, vol. 2, 2000, pp. 499–502 vol.2.
- [36] V. Areekul, K. Suppasriwaseth, and S. Jirachawang, "The new focal point localization algorithm for fingerprint registration," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 4, 2006, pp. 497–500.
- [37] 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.
- [38] P. Schuch, J. M. May, and C. Busch, "Unsupervised learning of fingerprint rotations," in *2018 International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2018, pp. 1–6.
- [39] G. Arora, A. Kumbhat, A. Bhatia, and K. Tiwari, "CP-Net: Multi-scale core point localization in fingerprints using hourglass network," in *2023 11th International Workshop on Biometrics and Forensics (IWBF)*, 2023, pp. 1–6.
- [40] S. A. Grosz, J. J. Engelsma, E. Liu, and A. K. Jain, "C2CL: Contact to contactless fingerprint matching," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 196–210, 2022.
- [41] 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.
- [42] X. Guan, Z. Pan, J. Feng, and J. Zhou, "Joint identity verification and pose alignment for partial fingerprints," *IEEE Transactions on Information Forensics and Security*, vol. 20, pp. 249–263, 2025.
- [43] V. Zaliva, "3D finger posture detection and gesture recognition on touch surfaces," Jan. 10 2013, uS Patent App. 13/544,960.
- [44] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [45] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.
- [46] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive mixtures of local experts," *Neural Computation*, vol. 3, no. 1, pp. 79–87, 1991.
- [47] W. Fedus, B. Zoph, and N. Shazeer, "Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity," *Journal of Machine Learning Research*, vol. 23, no. 120, pp. 1–39, 2022.
- [48] A. Liu, B. Feng, B. Xue, B. Wang, B. Wu, C. Lu, C. Zhao, C. Deng, C. Zhang, C. Ruan *et al.*, "Deepseek-v3 technical report," *arXiv preprint arXiv:2412.19437*, 2024.
- [49] H. Touvron, P. Bojanowski, M. Caron, M. Cord, A. El-Nouby, E. Grave, G. Izard, A. Joulin, G. Synnaeve, J. Verbeek, and H. Jégou, "ResMLP: Feedforward networks for image classification with data-efficient training," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 4, pp. 5314–5321, 2023.
- [50] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Guo, M. Gheshlaghi Azar, B. Piot, k. kavukcuoglu, R. Munos, and M. Valko, "Bootstrap your own latent - a new approach to self-supervised learning," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 21 271–21 284.
- [51] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable visual models from natural language supervision," in *Proceedings of the 38th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M. Meila and T. Zhang, Eds., vol. 139. PMLR, 18–24 Jul 2021, pp. 8748–8763.
- [52] S. Zhou, W. Liu, C. Hu, S. Zhou, and C. Ma, "UniDistill: A universal cross-modality knowledge distillation framework for 3D object detection in bird's-eye view," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 5116–5125.
- [53] C. Yang, Z. An, L. Huang, J. Bi, X. Yu, H. Yang, B. Diao, and Y. Xu, "CLIP-KD: An empirical study of clip model distillation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2024, pp. 15 952–15 962.
- [54] FVC 2002: the Second International Competition for Fingerprint Verification Algorithms. Accessed: Apr. 17, 2025. [Online]. Available: <http://bias.csr.unibo.it/fvc2002/default.asp>
- [55] FVC 2004: the Third International Fingerprint Verification Competition. Accessed: Apr. 17, 2025. [Online]. Available: <http://bias.csr.unibo.it/fvc2004/default.asp>
- [56] Neurotechnology. (2024) VeriFinger SDK. [Online]. Available: <https://www.neurotechnology.com/verifinger.html>
- [57] Y. Duan, J. Yu, J. Feng, K. He, J. Lu, and J. Zhou, "3D finger rotation estimation from fingerprint images," *Proc. ACM Hum.-Comput. Interact.*, vol. 7, no. ISS, Nov. 2023.
- [58] C. Yang, X. Yu, Z. An, and Y. Xu, *Categories of Response-Based, Feature-Based, and Relation-Based Knowledge Distillation*. Cham: Springer International Publishing, 2023, pp. 1–32.

SUPPLEMENTARY MATERIALS

TABLE A.1

ALIGNMENT ERROR UNDER DIFFERENT FINGERPRINT POSES ON DPF ([FULL FINGERPRINTS] WITH DIFFERENT ROTATION RANGES). SUFFIXES ‘FP’ INDICATE THAT ONLY THE CORRESPONDING BRANCH IS USED.

Method	[−45°, 45°]		[−90°, 90°]		[−135°, 135°]		[−180°, 180°]	
	trans (px)	rot (°)	trans (px)	rot (°)	trans (px)	rot (°)	trans (px)	rot (°)
No Registration	83.4	22.7	85.5	45.5	85.1	67.4	88.3	90.8
Faster-RCNN [6]	218.4	22.3	224.4	27.0	118.3	66.6	221.5	100.1
STN [8], [14]	90.1	16.2	85.9	41.8	85.7	67.6	89.2	92.0
JointNet [9]	16.3	4.6	18.8	7.5	23.0	13.5	22.8	19.9
GridNet [10]	14.9	4.2	19.7	6.2	70.9	28.6	83.7	62.3
Proposed-fp	14.5	1.8	15.0	2.0	14.1	1.8	14.2	1.8

TABLE A.2

VERIFICATION PERFORMANCE (%) ON DPF ([FULL FINGERPRINTS] WITH DIFFERENT ROTATION RANGES) USING DIFFERENT FINGERPRINT POSES. SUFFIXES ‘FP’ INDICATE THAT ONLY THE CORRESPONDING BRANCH IS USED.

Matcher	Method	[−45°, 45°]			[−90°, 90°]			[−135°, 135°]			[−180°, 180°]		
		EER	FNMR ¹	FNMR ²	EER	FNMR ¹	FNMR ²	EER	FNMR ¹	FNMR ²	EER	FNMR ¹	FNMR ²
VeriFinger [56]	No Registration	0.71	0.42	1.90	0.71	0.55	1.95	0.64	0.34	1.84	0.66	0.43	1.84
	JointNet [9]	0.64	0.32	1.90	0.64	0.32	1.95	0.60	0.23	1.66	0.85	0.79	2.20
	GridNet [10]	0.63	0.32	1.90	0.63	0.32	1.95	0.70	0.48	1.98	4.07	4.43	5.81
	Proposed-fp	0.59	0.26	1.82	0.61	0.24	1.76	0.54	0.15	1.66	0.54	0.13	1.70
FDD [15]	No Registration	30.66	61.02	69.73	40.14	83.44	89.03	44.41	90.05	94.45	44.95	92.80	96.40
	JointNet [9]	1.53	1.72	4.34	2.67	3.79	7.06	6.72	11.66	19.05	12.19	22.13	30.36
	GridNet [10]	1.34	1.49	3.62	2.57	3.48	5.84	30.00	55.44	63.11	41.40	85.19	90.73
	Proposed-fp	1.32	1.50	3.48	1.19	1.35	3.67	1.18	1.29	3.65	1.07	1.14	3.34

¹ FNMR@FMR=1e-3, ² FNMR@FMR=1e-4.

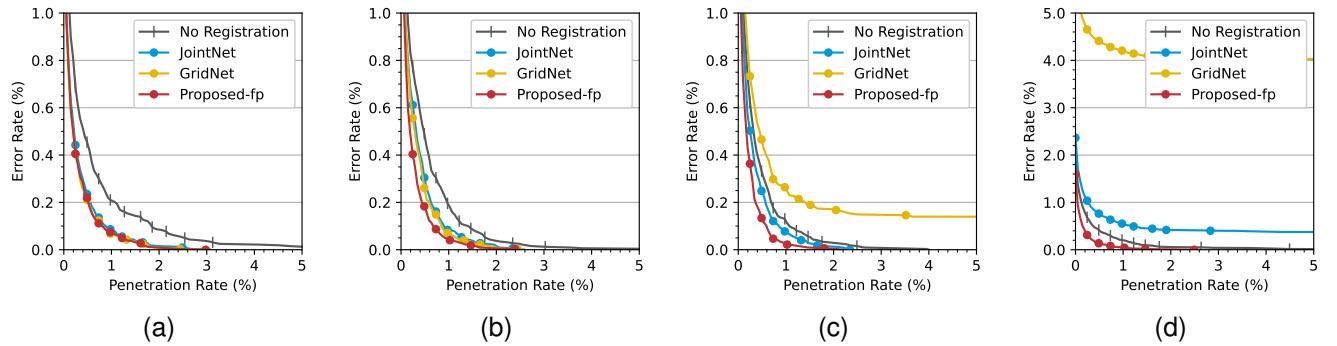


Fig. A.1. The VeriFinger [56] based fingerprint indexing performance with corresponding pose constraint on [full fingerprints] from DPF under different rotation ranges: (a) [45°, 45°], (b) [90°, 90°], (c) [135°, 135°], (d) [180°, 180°]. Different input modalities are distinguished by the shape of markers.