This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Towards Explainable and Unprecedented Accuracy in Matching Challenging Finger Crease Patterns

Zhenyu Zhou Chengdong Dong Ajay Kumar
Department of Data Science and Artificial Intelligence
The Hong Kong Polytechnic University, Hong Kong
{zhenyucs.zhou, chengdong.dong}@connect.polyu.hk, ajay.kumar@polyu.edu.hk

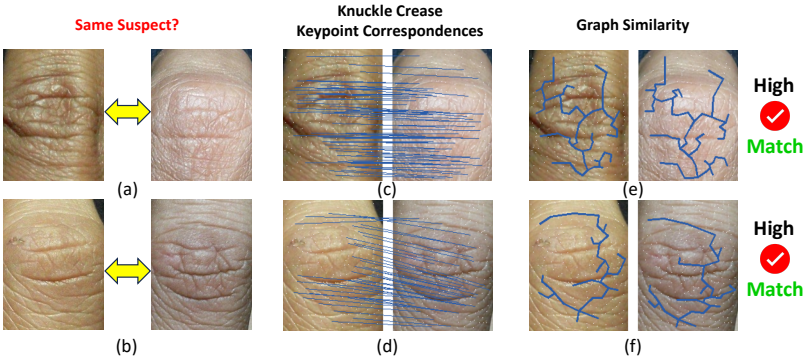


Figure 1. (a)-(b) sample matched knuckle image pairs from two identities, (c)-(d) uncovering knuckle crease point deep feature correspondences, (e)-(f) recovering respective graphs and computing respective graph similarity scores to identify persons.

Abstract

The primary obstacle in realizing the full potential of finger crease biometrics is the accurate identification of deformed knuckle patterns, often resulting from completely contactless imaging. Current methods struggle significantly with this task, yet accurate matching is crucial for applications ranging from forensic investigations, such as child abuse cases, to surveillance and mobile security. To address this challenge, our study introduces the largest publicly available dataset of deformed knuckle patterns, comprising 805,768 images from 351 subjects. We also propose a novel framework to accurately match knuckle patterns, even under severe pose deformations, by recovering interpretable knuckle crease keypoint feature templates. These templates can dynamically uncover graph structure and feature similarity among the matched correspondences. Our experiments, using the most challenging protocols, illustrate significantly outperforming results for matching such knuckle images. For the first time, we present and evaluate a theoretical model to estimate the uniqueness of 2D finger knuckle patterns, providing a more interpretable and accurate measure of distinctiveness, which is invaluable for forensic examiners in prosecuting suspects.

1. Introduction

The Victim Identification Programme in the Department of Homeland Security (DHS) has acquired millions of child abuse images and videos, and continues to receive thousands of such images every week. Finger Knuckle patterns are often the only evidence left in such images to identify the suspects, and law enforcement has successfully used such evidence for the prosecution [1–4]. However, such knuckle crease pattern matching is largely done manually [5] and is very challenging to match from videos and millions of images in surveillance patterns. Therefore, completely automated capabilities are required to accurately detect and match such finger knuckle images. Recent advancements in this area have indicated [6] high capabilities to detect knuckle patterns from complex videos and images. However, the accuracy for matching such completely contactless knuckle images with varying knuckle crease poses (referred here as challenging knuckle images) is extremely poor [1] and achieving such advancement is the key focus of this paper.

Completely contactless hand biometrics identification offers enhanced hygiene and higher user acceptance. Among various hand biometric traits, finger knuckle identification is the youngest [2, 8], and therefore, any large-scale business applications are yet to emerge or compete with other established traits like fingerprint or palmprint from centuries of identification using paper-based inked impressions. The success of contactless palmprint [9, 10] is

[1] Best methods for *such* cross-pose knuckle matching can only achieve 2.74% of GAR in [7] or 2.89% of GAR in [6] at FAR of 0.01%.

attributed to the presentation of nearly straight palm surfaces during image sensing, and similar (straight or fixed pose) presentation of finger knuckles has also been shown to achieve very high accuracy [6]. In such a *context* of completely contactless imaging, the proximal interphalangeal (PIP) finger knuckle joint has the highest or much higher agility than the palm. Therefore, for the success of completely contactless finger knuckle identification in civilian applications, e.g., mobile phones or access control in buildings, requires advanced capabilities to match cross-pose knuckle patterns and is another motivation for our work.

1.1. Our Work

Most accurate methods for finger knuckle identification available in the literature considered normalized knuckle patterns to be the texture. These methods can offer high accuracy for good quality knuckle images but have two fundamental limitations: (i) their performance on *cross-pose* knuckle matching or the dataset with deformed knuckle patterns is very poor, and (ii) these methods *lack explainability* or the matching methods are not interpretable or to be used in the courtroom arguments (like for the fingerprints) that can aid in the prosecution of suspects. Reference [11] is the first such attempt to articulate the challenges in deformable finger knuckle matching and also provides a database to advance this research. However, this database is of limited size, acquired in a controlled environment, and inadequate to evaluate much-needed *cross-database* performance for deep network-based biometric identification. Therefore, we also develop a challenging multi-pose finger knuckle video dataset by acquiring finger knuckle crease deformations under ambient and contactless mobile imaging.

To achieve explainable finger knuckle matching, we generate knuckle crease feature keypoint-based templates from unknown knuckle images. It is worth noting that the methods currently used for matching finger knuckle patterns during forensic investigations are largely manual and incorporate such visible keypoint-based patterns to serve as explainable pieces of evidence during the prosecution. It is observed that completely contactless imaging often introduces changes in illumination, viewpoints, and local deformations, while these factors do not easily change the knuckle crease points. Therefore, we extract such finger knuckle keypoints and their local features as templates using a convolutional neural network (CNN). These features are then used to generate mutually matched feature keypoint pairs and the feature graph. With the help of a graph neural network (GNN) that uses the graph structure and the keypoint feature templates, we generate a similarity score between two knuckle images.

Explainability and interpretation of finger knuckle matching also require us to ascertain the uniqueness of knuckle patterns. Such uniqueness can provide a theoretical upper bound on the number of finger knuckle patterns that can be reliably matched using the (interpretable) method introduced in this paper.

In summary, the main contributions are as follows:

- We design a novel framework to accurately match finger knuckle patterns, especially under severe pose deformations, which is built from the similarity of correspondence graph neural network (details in Sec. 3.3) to uncover the graph structure and keypoint feature similarity of the mutually matched correspondences in the knuckle images.
- One of the most important aspect of our knuckle matching framework is the *explainablity* and *interpretability* (in Sec. 3) in matching knuckle images, i.e., can provide visual graphs (e.g. Fig. 1 or Fig. 3) along with the list corresponded knuckle crease keypoint matches, which is consistent with practice by forensic examiners.
- We present the first attempt to theoretically evaluate the uniqueness of 2D finger knuckle patterns, which can help the law enforcement community pursue the successful prosecution of suspects during courtroom arguments.
- This paper introduces the first and most comprehensive *deformed* (multi-pose) finger knuckle video (4K UHD at 60fps) dataset (more details in Sec. 4.1), featuring 805,768 images from 351 different volunteers.

2. Related Work

Contactless Finger Knuckle Identification: Several studies on contactless finger knuckle identification, such as those in [12], [13], [14], [7], and [6], have demonstrated impressive performance across various publicly accessible 2D and 3D finger knuckle databases. However, they struggle with the particularly challenging deformable finger knuckle dataset [11]. To improve match performance, the [7] incorporated feature translations by shifting the window, and then the [6] also designed a new loss function to enhance the robustness of CNN in incorporating rotation and shift invariance capabilities. Reference [15] presents an impressive effort that also uses the STN [16] for getting learnable rigid transformation feature maps. Despite these advancements, there are two significant limitations in current methodologies: firstly, these methods lack explainability as the feature templates are not interpretable, and secondly, there is a dramatic degradation in performance when matching knuckle patterns with deformations, which is crucial for real-world applications, or using such database in [11].

Keypoint and Descriptor Detection: The keypoints or interest points in an image can be reliably extracted and matched across images [17]. Traditional methods such as SIFT [18], SURF [19], and ORB [20] have been extensively documented in the literature. Building on this foundation, LIFT [21] detects the keypoint with a small image patch and the descriptor with a larger patch around the keypoint. LF-Net [22], with the help of STN [16], can detect the location,

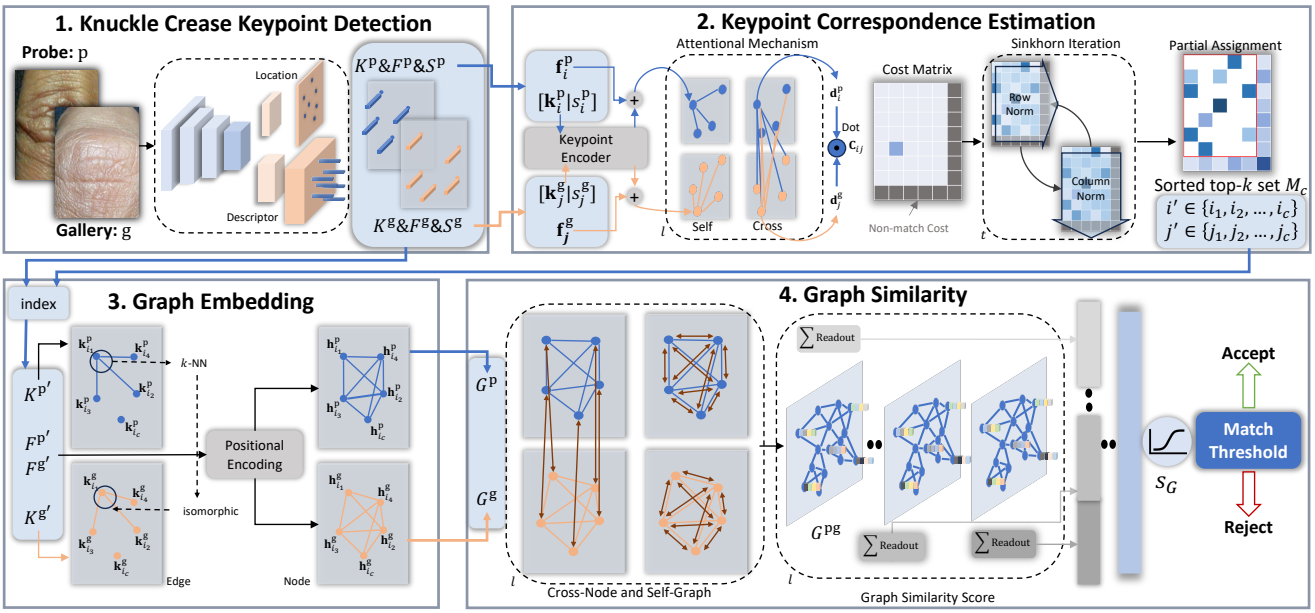


Figure 2. Visualization of our framework to match *cross-pose* knuckle images based on crease keypoint correspondences. Step 1: Detect keypoints and descriptors using a CNN; Step 2: Identify mutually matched correspondences by l attentional layers, on detected keypoints, optimized with t iterations; Step 3: Graph isomorphism-based formation of probe graph G^{p} and gallery graph G^{g} are formed; Step 4: Computing a match score using the similarity of graph structures built using correspondences. The Cross-Node and Self-Graph module, along with the tracked graph, are visualized on the Fig. 5 and discussed in Sec. 3.3.

orientation, and scales of a keypoint, and it also can detect the descriptor by extracting a patch around the keypoint. The D2-Net [23] is another advancement that uses a similar idea from the SIFT to detect the keypoint location and descriptor simultaneously. However, empirical evidence from the ASLFeat [24] indicates that both LF-Net [22] and D2-Net [23] exhibit low precision in tow-view matching due to their low-resolution feature maps. In contrast, the SuperPoint [25] can restore the spatial resolution by learning to up-sample, resulting in better performance.

Graph Similarity: Contemporary GNNs have been shown to efficiently compute the similarity between graph-structured objects, delivering consistently reliable performance. Firstly, a deep feature can represent a graph such as the GMN model [26] that learns graph-level representations by updating self-graph information with the graph convolution and cross-graph information with the attention mechanism under a Siamese architecture. In contrast, the GraphSim model [27] uses a CNN to convolve the pairwise node similarity matrix to get the similarity score. To improve the matching performance, the SimGNN model [28] and MGMN model [29] add the node similarity based on the graph-level representation to improve the matching performance. The former uses pairwise comparison, and the latter uses the attention mechanism for node similarity. Furthermore, the EGSC model [30] uses knowledge distillation to speed up the inference time, and the ERIC model [31] uses a simple regularization technique to replace the expensive node-to-node matching module to reduce complexity.

Uniqueness of Finger Knuckle Patterns: The distinctiveness of several established biometrics is well-studied, e.g., fingerprints [32], iris [33–37], or handwriting [38], and provides a theoretical upper limit on the expected performance from respective biometrics using 2D images. Such uniqueness is estimated from the probability of *false* random correspondences or matches generated from the respective biometric template. Most such studies on uniqueness are focused on fingerprints [32, 39, 40] and are widely considered helpful during courtroom arguments for the prosecution of suspects using fingerprint evidence. However, research on the uniqueness of 2D finger knuckle images remains unexplored and is highly sought, especially in the forensic community [2]. While some studies [41] provide promising evidence demonstrating the stability of knuckle crease patterns with age or gender, to our knowledge, no research has yet addressed their uniqueness.

3. Our Approach

Forensic examiners typically consolidate the match evidence from the finger knuckle creases and points because it is interpretable and explainable during courtroom arguments. However, such matching using real-world images is currently done manually and remains a key challenge in *timely* identifying the suspects (and ensuring justice). Also,

a key challenge in the realization of completely contactless knuckle identification for civilian or access control applications lies in the capability to accurately match finger knuckle patterns from the same user or fingers with different deformations. Such images are frequently acquired under such contactless imaging, while the currently available state-of-the-art (SOTA) models [6, 7, 13, 15] obtain unacceptably poor performance. Therefore, our framework for accurately matching such images is built on the graph structure similarity and localized deep features similarity formed from the corresponding keypoint *locations* between two matched images. This framework [42] is summarized in Fig. 2 and is organized under four key modules to enhance the interpretability of the method. More details on modules are in below sections.

3.1. Knuckle Crease Point Detection and Encoding

Among deep neural network-based interest point detection models, the SuperPoint [25] is a self-supervised learning model with a synthetic shapes dataset and a homographic adaptation detection strategy, which can be helpful with the limited availability of ground truth for the finger knuckle keypoints. This model was also employed (training protocol and justifications in Supplementary C.2) for our task and referred to as KnuckleCreasePoint. Its performance for the repeatability and reliability in detecting knuckle crease points is evaluated in Supplementary D.7. Given an image pair, the keypoint feature detection model in Fig. 2 is designed to detect the set of n keypoints with $F^{\mathrm{p}} = \{\boldsymbol{f}_i^{\mathrm{p}} \mid i \in \{1, 2, \ldots, n\}\}$, $K^{\mathrm{p}} = \{\boldsymbol{k}_i^{\mathrm{p}} \mid i \in \{1, 2, \ldots, n\}\}$, and $S^{\mathrm{p}} = \{s_i^{\mathrm{p}} \mid i \in \{1, 2, \ldots, n\}\}$ on the probe image (p), and to detect m keypoints set with F^{g}, K^{g}, and S^{g} on the gallery image (g).

3.2. Knuckle Keypoint Correspondence Estimation

To enhance the reliability in the estimation of correspondences, similar to in [43], we use self-attention and cross-attention GNN model, referred as KnucklePointPair, to enhance the similarity between two sets. Firstly, a Multilayer Perceptron (MLP) is used to combine each keypoint descriptor matrix $\boldsymbol{F} \in \mathbb{R}^{n \times d}$, location matrix $\boldsymbol{K} \in \mathbb{R}^{n \times 2}$, and score matrix $\boldsymbol{S} \in \mathbb{R}^{n \times 1}$ of the probe image to generate an initial graph node feature matrix $\boldsymbol{D}^{\mathrm{p}} \in \mathbb{R}^{n \times d}$ as:

$$\boldsymbol{D}^{\mathrm{p}} = \boldsymbol{F}^{\mathrm{p}} + \mathrm{MLP}([\boldsymbol{K}^{\mathrm{p}} \mid \boldsymbol{S}^{\mathrm{p}}]) \tag{1}$$

$$\boldsymbol{C} = \begin{bmatrix} ({}^{(l)}\boldsymbol{D}^{\mathrm{p}})({}^{(l)}\boldsymbol{D}^{\mathrm{g}})^{\top} & \boldsymbol{U}_1 \\ \boldsymbol{U}_2 & u \end{bmatrix} \tag{2}$$

where the $[\cdot \mid \cdot]$ represents the concatenation operation. The gallery descriptor $\boldsymbol{D}^{\mathrm{g}}$ is also computed in a similar manner using Eq. (1). After l layers, the feature matrix can be represented by ${}^{(l)}\boldsymbol{D}^{\mathrm{p}}$ and ${}^{(l)}\boldsymbol{D}^{\mathrm{g}}$. The cost matrix $\boldsymbol{C} \in \mathbb{R}^{(n+1)\times(m+1)}$ is computed by the weighted product of $({}^{(l)}\boldsymbol{D}^{\mathrm{p}})({}^{(l)}\boldsymbol{D}^{\mathrm{g}})^{\top}$ and appended by the non-match scores in Eq. (2), while the $\boldsymbol{U}_1 \in \mathbb{R}^{n \times 1}$, $\boldsymbol{U}_2 \in \mathbb{R}^{1 \times m}$, and u are the learnable parameters to represent the non-match scores. From such mutually matched keypoint correspondence score matrix $\boldsymbol{M} \in \mathbb{R}^{n \times m}$ (discarding the non-match scores after t iterations [44]), we choose the top-k, with c, correspondence score set M_c based on index i' and j' as:

$$M_c = \{M_{i'j'} \mid \{(i', j') \in (i_1, j_1), (i_2, j_2), \ldots, (i_c, j_c)\}\} \tag{3}$$

where the $M_{i_1,j_1} > M_{i_2,j_2} > \cdots > M_{i_c,j_c}$. The top-$k$ $F^{\mathrm{p}'}$, $K^{\mathrm{p}'}$, $F^{\mathrm{g}'}$, and $K^{\mathrm{g}'}$ are selected by i' and j' index, respectively. For example, the $F^{\mathrm{p}'}$ and $K^{\mathrm{p}'}$ follows:

$$F^{\mathrm{p}'} = \{\boldsymbol{f}_{i'}^{\mathrm{p}} \mid i' \in \{i_1, i_2, \ldots, i_c\}\} \tag{4}$$

$$K^{\mathrm{p}'} = \{\boldsymbol{k}_{i'}^{\mathrm{p}} \mid i' \in \{i_1, i_2, \ldots, i_c\}\} \tag{5}$$

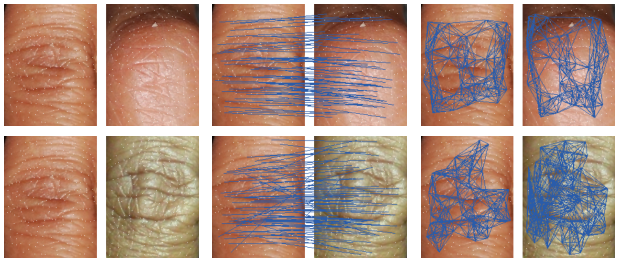


Figure 3. Visualization of detected keypoints (first two columns), matched correspondences (middle two column pairs), and composed graphs by k-NN (k=9) on the last two columns. The first row is a genuine, while the second is an imposter pair.

3.3. Corresponded Keypoints Graph and Similarity

The number or similarity of mutually matched correspondences (evaluated in Supplementary D.7.3) ignores the underlying graph structure or the complex relationship among such correspondences and fails to account for ambiguities such as repeated patterns, textures, or non-discriminative local appearance. It can be observed from sample image pairs in Fig. 3 or Fig. 1 that the high-order graph [45] shows higher visual similarity (scale tolerance) for genuine pairs as compared to imposter pairs. Therefore, we are motivated to explore the similarity of graph structure composed by correspondences and its deep features for evaluating the similarity between two knuckle images.

Graph Embedding: The graph $G^{\mathrm{p}} = (V^{\mathrm{p}}, E^{\mathrm{p}})$ from the probe image and $G^{\mathrm{g}} = (V^{\mathrm{g}}, E^{\mathrm{g}})$ from the gallery image with vertex sets $V^{\mathrm{p}} = \{\boldsymbol{k}_{i'}^{\mathrm{p}} \mid i' \in \{i_1, i_2, \ldots, i_c\}\}$ and $V^{\mathrm{g}} = \{\boldsymbol{k}_{j'}^{\mathrm{g}} \mid j' \in \{j_1, j_2, \ldots, j_c\}\}$, respectively, is firstly constructed. The graph G^{p} and G^{g}, are isomorphic if there exists a bijection function $f : V^{\mathrm{p}} \to V^{\mathrm{g}}$ such that every two vertices $\boldsymbol{k}_u^{\mathrm{p}}$ and $\boldsymbol{k}_v^{\mathrm{p}}$ in V^{p} meet the following:

$$(\boldsymbol{k}_u^{\mathrm{p}}, \boldsymbol{k}_v^{\mathrm{p}}) \in E^{\mathrm{p}} \leftrightarrow (f(\boldsymbol{k}_u^{\mathrm{p}}), f(\boldsymbol{k}_v^{\mathrm{p}})) \in E^{\mathrm{g}} \tag{6}$$

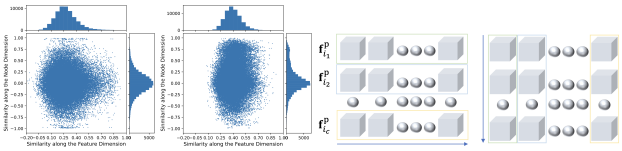
Between V^{p} and V^{g}, every vertex $\boldsymbol{k}_{i'}^{\mathrm{p}}$ meets the one-to-one bijection to the vertex $\boldsymbol{k}_{j'}^{\mathrm{g}}$. Following Eq. (6), we use the k-NN to link the $\boldsymbol{k}_{i'}^{\mathrm{p}}$ to their neighbor $\boldsymbol{k}_u^{\mathrm{p}} \in N(\boldsymbol{k}_{i'}^{\mathrm{p}})$ at location

$(x_{i'}, y_{i'})$ on a probe image, and use the same edge E^{p} to link the $\boldsymbol{k}^{\mathrm{g}}_{j'}$ to a graph G^{g} of a gallery image.

Positional Encoding: The similarity between such feature descriptors $\boldsymbol{f}^{\mathrm{p}}_{i'}$ and $\boldsymbol{f}^{\mathrm{g}}_{j'}$ for corresponded keypoints between the knuckle image pair is significant, as also shown from Fig. 4a. Furthermore, we also incorporate positional encoding $\boldsymbol{e}_{i'}$ [46] to enhance the similarity of outstanding correspondence, and can be described as follows:

$$\boldsymbol{e}(i', \alpha) = \begin{cases} \sin(i'/10000^{2\alpha/d}), & \text{for } 2\alpha \\ \cos(i'/10000^{2\alpha/d}), & \text{for } 2\alpha + 1 \end{cases} \tag{7}$$

where i' is the position of the feature vector $\boldsymbol{f}_{i'}$, α is the dimension of the positional encoding, and the d is the dimension of $\boldsymbol{f}_{i'}$. The position of feature vector $\boldsymbol{f}^{\mathrm{p}}_{i'}$ and $\boldsymbol{f}^{\mathrm{g}}_{j'}$ is already sorted by similarity score by Eq. (3). The graph node feature vectors $\boldsymbol{h}^{\mathrm{p}}_{i'}$ and $\boldsymbol{h}^{\mathrm{g}}_{j'}$ are computed by $\boldsymbol{h}^{\mathrm{p}}_{i'} = \boldsymbol{f}^{\mathrm{p}}_{i'} + \boldsymbol{e}_{i'}$ and $\boldsymbol{h}^{\mathrm{g}}_{j'} = \boldsymbol{f}^{\mathrm{g}}_{j'} + \boldsymbol{e}_{j'}$, respectively. Positional encoding can enhance feature similarity, as can also be observed by comparing the plots in Fig. 4b and Fig. 4a.



(a) (b) (c) (d)

Figure 4. Feature correlation between $\boldsymbol{f}^{\mathrm{p}}_{i'}$ and $\boldsymbol{f}^{\mathrm{g}}_{j'}$ as shown in (a), and between $\boldsymbol{h}^{\mathrm{p}}_{i'}$ and $\boldsymbol{h}^{\mathrm{g}}_{j'}$ as shown in (b) computed *along* the feature dimension (c) and the node dimension (d).

Convolution along the Node Dimension: Spatial-based GNNs can update node feature templates with their linked nodes as follows:

$$^{(l)}\boldsymbol{a}^{\mathrm{p}}_{i'} = f_a(\{^{(l-1)}\boldsymbol{h}^{\mathrm{p}}_{u} : \boldsymbol{k}^{\mathrm{p}}_{u} \in N(\boldsymbol{k}^{\mathrm{p}}_{i'})\}) \tag{8}$$

$$^{(l)}\boldsymbol{h}^{\mathrm{p}}_{i'} = \phi(f_c(^{(l-1)}\boldsymbol{h}^{\mathrm{p}}_{i'}, ^{(l)}\boldsymbol{a}^{\mathrm{p}}_{i'})) \tag{9}$$

It can aggregate its neighbor node feature template $^{(l-1)}\boldsymbol{h}^{\mathrm{p}}_{u}$ at node $\boldsymbol{k}^{\mathrm{p}}_{u}$ with the f_a aggregation function, then update the current node $\boldsymbol{k}^{\mathrm{p}}_{i'}$ by combining the neighborhood information $^{(l)}\boldsymbol{a}^{\mathrm{p}}_{i'}$ and the node self feature template $^{(l-1)}\boldsymbol{h}^{\mathrm{p}}_{i'}$ to $^{(l)}\boldsymbol{h}^{\mathrm{p}}_{i'}$ using the combination function f_c and an activation function ϕ.

However, along the node dimension Fig. 4, the cosine similarity of these vectors is smaller than such similarity observed along the feature dimension which can be largely attributed to the difference in the pattern kernels. Therefore, we change the graph convolution dimension to the node dimension, instead of the feature dimension, to retain the discriminant properties of the feature vectors from the CNN model KnuckleCreasePoint. This new aggregation function concatenates the neighboring features $\boldsymbol{h}^{\mathrm{p}}_{u}$ ($\boldsymbol{k}^{\mathrm{p}}_{u} \in N(\boldsymbol{k}^{\mathrm{p}}_{i'})$) of $\boldsymbol{k}^{\mathrm{p}}_{i'}$ to $^{(l)}\boldsymbol{A}^{\mathrm{p}}_{i'} \in \mathbb{R}^{d_{in} \times (k+1)}$ at the layer l and can be defined as follows:

$$^{(l)}\boldsymbol{A}^{\mathrm{p}}_{i'} = [(^{(l-1)}\boldsymbol{h}^{\mathrm{p}}_{i'})^{\top} \mid (^{(l-1)}\boldsymbol{h}^{\mathrm{p}}_{u_1})^{\top} \mid \cdots \mid (^{(l-1)}\boldsymbol{h}^{\mathrm{p}}_{u_k})^{\top}] \tag{10}$$

$$^{(l)}h^{\mathrm{p}}_{i'd_i} = \phi(bias + tr((^{(l)}\boldsymbol{W}_h[d_i, :, :])(^{(l)}\boldsymbol{A}^{\mathrm{p}}_{i'})^{\top})) \tag{11}$$

$$^{(l)}\boldsymbol{h}^{\mathrm{p}}_{i'} = (^{(l)}h^{\mathrm{p}}_{i'1}, \ldots, ^{(l)}h^{\mathrm{p}}_{i'd_i}, \ldots, ^{(l)}h^{\mathrm{p}}_{i'd_{out}}) \tag{12}$$

where $tr()$ represented trace of square matrix, and the $\boldsymbol{W}_h[d_i \ :, :]$ is the slicing operation. We use graph convolution as in Eq. (11) to the $^{(l)}\boldsymbol{A}^{\mathrm{p}}_{i'}$ with the learnable weights $^{(l)}\boldsymbol{W}_h \in \mathbb{R}^{d_{out} \times d_{in} \times (k+1)}$, and concatenate to the $^{(l)}\boldsymbol{h}^{\mathrm{p}}_{i'} \in \mathbb{R}^{d_{out}}$. The node feature $^{(l)}\boldsymbol{h}^{\mathrm{g}}_{j'}$ of G^{g} also follow above graph convolution function.

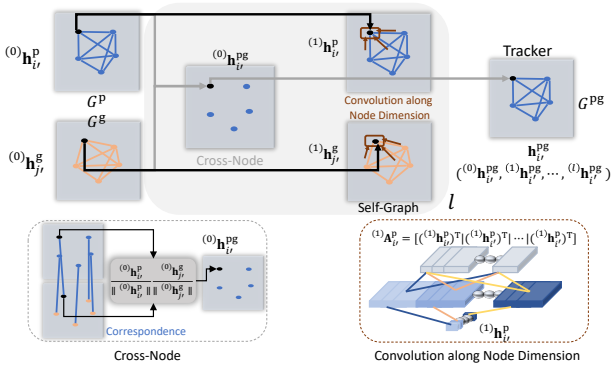


Figure 5. Tracker records change in one-to-one cross-similarity scores along l layers of Self-Graph and Cross-Node.

Cross-Node and Self-Graph: Current graph similarity models [28, 29, 31] use computationally expansive attention to compute node similarity along graph-level representation. However, the keypoint correspondences and their neighbors are expected to have high similarity in Fig. 4. Also, the one-to-one node and visual graph structure similarity of genuine pairs is greater than imposter pairs. Therefore (more details in Supplementary B), we compute the cross-node similarity as follows:

$$^{(l)}h^{\mathrm{pg}}_{i'} = \cos^{-1}\left(\frac{^{(l)}\boldsymbol{h}^{\mathrm{p}}_{i'} \cdot ^{(l)}\boldsymbol{h}^{\mathrm{g}}_{j'}}{\|^{(l)}\boldsymbol{h}^{\mathrm{p}}_{i'}\| \|^{(l)}\boldsymbol{h}^{\mathrm{g}}_{j'}\|}\right) \tag{13}$$

$$\boldsymbol{h}^{\mathrm{pg}}_{i'} = (^{(0)}h^{\mathrm{pg}}_{i'}, ^{(1)}h^{\mathrm{pg}}_{i'}, \cdots, ^{(l)}h^{\mathrm{pg}}_{i'}) \tag{14}$$

As shown in Fig. 5, we use the l layer (we use $l = 3$) of Self-Graph to update the node feature of $\boldsymbol{k}^{\mathrm{p}}_{i'}$ and $\boldsymbol{k}^{\mathrm{g}}_{j'}$, and use the Cross-Node to compute the changes in one-to-one similarity score between $\boldsymbol{k}^{\mathrm{p}}_{i'}$ and $\boldsymbol{k}^{\mathrm{g}}_{j'}$ on each layer based on Eq. 13. Finally, the Tracker is employed to concatenate the changes in Cross-Node similarity scores to $\boldsymbol{h}^{\mathrm{pg}}_{i'} \in \mathbb{R}^{l+1}$ as same in Eq. (14) and construct new tracked graph G^{pg} that has the same graph structure with G^{p} and G^{g}.

Graph Similarity: We follow the graph-level task to classify the graph G^{pg} by using the same graph convolution inEq. (11). After l GNN layers (we use $l = 3$), each layer's graph-level details $^{(l)}\boldsymbol{h}^{\mathrm{pg}}_{G}$ from Eq. (15) is concatenated to a feature vector $\boldsymbol{h}^{\mathrm{pg}}_{G}$. An MLP with sigmoid function σ is used to compute similarity score s_G from $\boldsymbol{h}^{\mathrm{pg}}_{G}$.

$$^{(l)}\boldsymbol{h}^{\mathrm{pg}}_{G} = ^{(l)}\boldsymbol{h}^{\mathrm{pg}}_{i_1} + ^{(l)}\boldsymbol{h}^{\mathrm{pg}}_{i_2} + \cdots + ^{(l)}\boldsymbol{h}^{\mathrm{pg}}_{i_c} \tag{15}$$

$$\boldsymbol{h}_G^{\text{pg}} = [^{(1)}\boldsymbol{h}_G^{\text{pg}} \,|^{(2)}\, \boldsymbol{h}_G^{\text{pg}} \,|\, \cdots \,|^{(l)}\, \boldsymbol{h}_G^{\text{pg}}] \tag{16}$$

$$s_G = \sigma(\text{MLP}(\boldsymbol{h}_G^{\text{pg}})) \tag{17}$$

For convenience, our proposed method, which relies on the similarity of correspondence graph neural network, is abbreviated as CGN in this paper.

3.4. Uniqueness of 2D Finger Knuckle Patterns

The highly interpretable knuckle crease keypoint-based matching method is suitable to ascertain the uniqueness of finger knuckle patterns. Such estimation of the uniqueness can be determined using the probability of the false random correspondence (FRC) for two randomly matched 2D finger knuckle patterns known to belong to two different fingers/subjects. Therefore a joint mixture multivariate Gaussian model is incorporated to estimate such dependency between locations and keypoint features.

Multivariate Distribution Model: We use $\boldsymbol{k} = (x, y) \in K$ and $\boldsymbol{f} = (f_1, f_2, \ldots, f_d) \in F$ to represent the possible configurations for K (possible location) and F (feature). A multivariate model with N Gaussian distribution components is employed to estimate the correlation between the locations and features, the weights $\epsilon_1, \epsilon_2, \ldots, \epsilon_N$ for each component such that $\epsilon_i \geq 0$ and $\sum_i^N \epsilon_i = 1$. In each of the n components, the locations of keypoints are modeled using bivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}_n^{(1)}, \boldsymbol{\Sigma}_n^{(1)})$ while the features follow $\mathcal{N}(\boldsymbol{\mu}_n^{(2)}, \boldsymbol{\Sigma}_n^{(2)})$, where $\boldsymbol{\mu}_n^{(1)} \in \mathbb{R}^{2\times 1}$, $\boldsymbol{\Sigma}_n^{(1)} \in \mathbb{R}^{2\times 2}$, $\boldsymbol{\mu}_n^{(2)} \in \mathbb{R}^{256\times 1}$, $\boldsymbol{\Sigma}_n^{(2)} \in \mathbb{R}^{256\times 256}$. Therefore, the distribution of keypoint feature in knuckle templates can be represented as follows:

$$f(\boldsymbol{k}, \boldsymbol{f}|\Omega) = \sum_{n=1}^{N} \epsilon_n f_n^K(\boldsymbol{k}|\boldsymbol{\mu}_n^{(1)}, \boldsymbol{\Sigma}_n^{(1)}) \cdot f_n^F(\boldsymbol{f}|\boldsymbol{\mu}_n^{(2)}, \boldsymbol{\Sigma}_n^{(2)}) \tag{18}$$

where the Ω denotes all the unknown parameters of the model including N, $\epsilon_1 \cdots \epsilon_N$, $\boldsymbol{\mu}_1^{(1)} \cdots \boldsymbol{\mu}_N^{(1)}$, $\boldsymbol{\Sigma}_1^{(1)} \cdots \boldsymbol{\Sigma}_N^{(1)}$, $\boldsymbol{\mu}_1^{(2)} \cdots \boldsymbol{\mu}_N^{(2)}$, $\boldsymbol{\Sigma}_1^{(2)} \cdots \boldsymbol{\Sigma}_N^{(2)}$. Similar to [32, 40], the Expectation Maximization (EM) algorithm can be used to estimate parameters Ω_N. The optimal number of components N^* is selected using Bayes Information Criteria (BIC).

Estimating FRC: Let T^{g} denotes the gallery template containing m keypoints and T^{p} denotes the probe template containing n keypoints. Following [32, 40], the probability $p(w; T^{\text{p}}, T^{\text{g}})$ of observing at least w matches between two templates T^{p} and T^{g} can be modeled using the following Poisson distribution:

$$p(w; T^{\text{p}}, T^{\text{g}}) = \sum_{w}^{\infty} \frac{e^{-\lambda(T^{\text{p}}, T^{\text{g}})} \lambda(T^{\text{p}}, T^{\text{g}})^w}{w!} \tag{19}$$

where parameter $\lambda(T^{\text{p}}, T^{\text{g}}) = mnp(T^{\text{p}}, T^{\text{g}})$. $p(T^{\text{p}}, T^{\text{g}})$ denotes the probability of a pair of acquiring a random match between a keypoint in T^{p} and another keypoint in T^{g}. The statistical tolerance or threshold for $\Delta_{\boldsymbol{k}}$ and $\Delta_{\boldsymbol{f}}$ to authenticate *match* from knuckle keypoint templates ground truth is empirically estimated (as detailed in Supplementary E). It is estimated as $p(\delta(\boldsymbol{k}^{\text{p}}, \boldsymbol{k}^{\text{g}}) \leq \Delta_{\boldsymbol{k}}) = 0.975$ and $p(\delta(\boldsymbol{f}^{\text{p}}, \boldsymbol{f}^{\text{g}}) \leq \Delta_{\boldsymbol{f}}) = 0.975$ so that top 2.5% genuine matches are excluded following same protocol in [39]. The match probability for a pair of random keypoints can be computed as follows:

$$p(T^{\text{p}}, T^{\text{g}}) = \int\int_{(\boldsymbol{k}^{\text{p}}, \boldsymbol{f}^{\text{p}}) \in B(\boldsymbol{k}^{\text{g}}, \boldsymbol{f}^{\text{g}})} f_{T^{\text{p}}}((\boldsymbol{k}^{\text{p}}, \boldsymbol{f}^{\text{p}})) f_{T^{\text{g}}}((\boldsymbol{k}^{\text{g}}, \boldsymbol{f}^{\text{g}})) d\boldsymbol{k}^{\text{p}} d\boldsymbol{f}^{\text{p}} d\boldsymbol{k}^{\text{g}} d\boldsymbol{f}^{\text{g}} \tag{20}$$

where $(\boldsymbol{k}^{\text{p}}, \boldsymbol{f}^{\text{p}}) \in B(\boldsymbol{k}^{\text{g}}, \boldsymbol{f}^{\text{g}})$ denotes the set of all $(\boldsymbol{k}^{\text{p}}, \boldsymbol{f}^{\text{p}})$ with locational and feature distances from $(\boldsymbol{k}^{\text{g}}, \boldsymbol{f}^{\text{g}})$ less than $\Delta_{\boldsymbol{k}}, \Delta_{\boldsymbol{f}}$ respectively. In a finger knuckle database with z identities, the average FRC index can be computed from a total of $z(z-1)/2$ imposter pairs, as follows:

$$\overline{FRC} = \frac{2}{z(z-1)} \sum_{T^{\text{p}} \neq T^{\text{g}}} p(w; T^{\text{p}}, T^{\text{g}}) \tag{21}$$

In addition, to avoid the adverse influence from the extremely large or small value, only the $p(w; T^{\text{p}}, T^{\text{g}})$ between the lower $100\alpha/2$ percentile $p(w; \alpha/2)$ and upper $100\alpha/2$ percentile $p(w; 1-\alpha/2)$ are retained. Finally, the respective mean FRC index can be represented by:

$$\overline{FRC}_\alpha = \frac{2}{z(z-1)(1-\alpha)} \sum_{T^{\text{p}} \neq T^{\text{g}}} p_\alpha(w; T^{\text{p}}, T^{\text{g}}) \tag{22}$$

where $p(w; \alpha/2) \leq p_\alpha(w; T^{\text{p}}, T^{\text{g}}) \leq p(w; 1-\alpha/2)$.

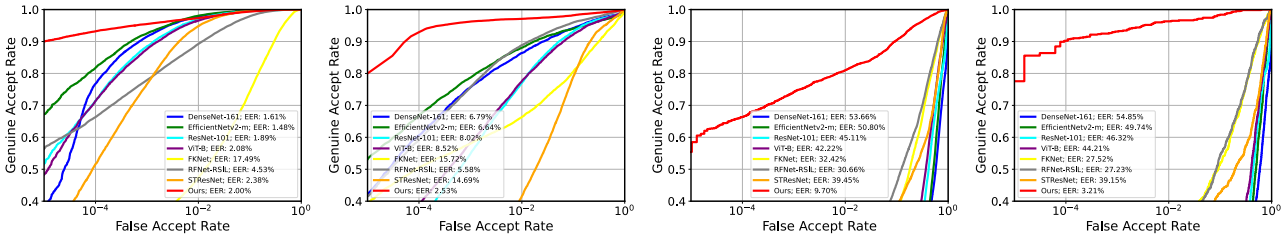
4. Experiments and Results

We present within-database and cross-database (without *any* fine-tuning) experimental results under a challenging *all-to-all* protocol [15]. This section also presents results from our study to ascertain the uniqueness of 2D finger knuckle patterns. Please see Supplementary C and D for additional details on training protocols, justifications, and results.

Choice of baseline matching methods: To ascertain the merit of the matching performance from our method, we choose SOTA CNN models as baseline methods, such as ResNet-101 [48], DenseNet-161 [49], and EfficientNetv2-m [50], and even with the transformer ViT-B model [51]. We also chose SOTA models that have been shown to offer impressive performance in the literature as baseline methods, e.g., FKNet [7], RFNet [6], and STResNet [15].

Table 1. Comparative summary of 2D finger knuckle databases.

Dataset	Completely Contactless	Pose	Format	Resolution w×h	Image Samples
[41]	✗	Single-Pose	Image	1600×1200	2,515
[47]	✗	Single-Pose	Image	1600×1200	3,560
[12]	✗	Single-Pose	Image	N/A	4,872
[13]	✗	Single-Pose	Image	720×576	5,016
[11]	✓	Six-Pose	Image	3072×2304	1,950
Ours	✓	Multi-Pose	Video	3840×2160	805,768



(a) (b) (c) (d)

Figure 6. Comparative ROC plots using *all-to-all* protocol on (a) our dataset, (b) the dataset in [47], and (c) the most challenging dataset in [11]. Such plots using the same (best score) protocols in [13] are in (d).

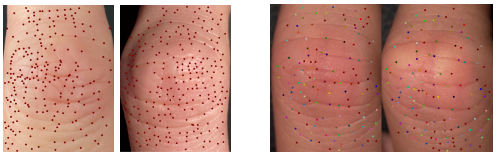


Figure 7. Visualization of keypoint (first two images) and correspondence (last ones) ground truth (more in Supplementary C).

4.1. Multipose Finger Knuckle Video Dataset

We developed a contactless and largest finger knuckle video dataset (351 subjects) with multi-pose finger deformations. A comparative summary with other open databases is presented in Tab. 1. We fully followed the guidelines for the data acquisition from Institutional Review Board (IRB). Volunteers consented to support our research. They were not paid any honorarium, and we did not record any of their personal details. This dataset is accordingly made available for further research via [52] (details in Supplementary D.1). We select an average of 25 frames per video of the left hand (about two frames per second) for training (in a total of 25,000 images) and 10 frames per video (about one frame per second) for the performance evaluation. Therefore, the finger knuckle from 351 subjects generates 31,590 ($351 \times 10 \times 9$) genuine and 12,285,000 ($351 \times 350 \times 10 \times 10$) imposter match scores. The comparative performance on this dataset is shown in Fig. 6a and summarized in Tab. 2. Our method significantly outperforms other methods, especially for the lower values of FAR.

Table 2. Comparative performance summary for our dataset.

Model	GAR (FAR=10^{-5})	GAR (FAR=10^{-4})	GAR (FAR=10^{-2})
DenseNet-161 [49]	38.19%	76.77%	97.71%
EfficientNetv2-m [50]	67.22%	81.78%	**98.06%**
ResNet-101 [48]	51.67%	71.46%	97.01%
ViT-B [51]	48.30%	71.38%	96.52%
FKNet [7]	17.41%	23.21%	47.39%
RFNet-RSIL [6]	56.39%	66.33%	89.28%
STResNet [15]	27.07%	51.63%	94.81%
Ours	**89.99%**	**93.19%**	97.32%

4.2. Cross-Database Performance

Hand Dorsal Image Dataset: This public dataset [47] was firstly chosen for the cross-database performance evaluation as it was acquired from 712 subjects, which is the largest for ascertaining generalization capability (14,240 ($712 \times 5 \times 4$) genuine and 12,655,800 ($712 \times 711 \times 5 \times 5$) imposter match scores). It can be observed from comparative experimental results in Fig. 6b or summary in Tab. 3 that even for less challenging (in the context of *intra-class* variations), there is significant performance improvement as compared to other methods.

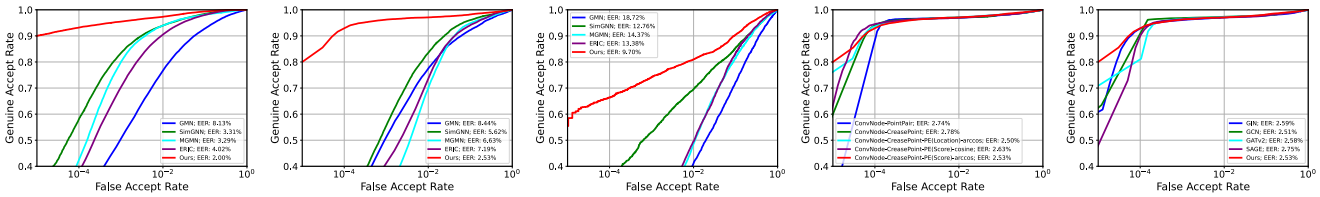
Table 3. Comparative performance summary for dataset in [47].

Model	GAR (FAR=10^{-5})	GAR (FAR=10^{-4})	GAR (FAR=10^{-2})
DenseNet-161 [49]	42.20%	60.65%	86.38%
EfficientNetv2-m [50]	53.11%	66.84%	88.05%
ResNet-101 [48]	17.71%	33.37%	77.02%
ViT-B [51]	24.84%	38.74%	77.62%
FKNet [7]	38.54%	50.63%	66.39%
RFNet-RSIL [6]	41.53%	59.23%	88.88%
STResNet [15]	3.01%	7.26%	41.59%
Ours	**80.19%**	**92.88%**	**97.09%**

Finger Knuckle v3.0 Dataset: This is the *most* challenging knuckle image dataset available to date in the public domain [11]. This dataset was acquired from 221 different subjects, but only 104 subjects provided images for the second session (with 3,744 ($104 \times 6 \times 6$) genuine and 385,632 ($104 \times 103 \times 6 \times 6$) imposter scores). Our comparative experimental results are shown in Fig. 6c and summarized in Tab. 4. It can be observed from the experimental results in Fig. 6c and Tab. 4 that our method can show dramatically enhanced performance as compared to the rest of the available methods on this most challenging dataset. Following the same protocol in [13], performance is also presented in Fig. 6d.

Table 4. Comparative performance summary for dataset in [11].

Model	GAR (FAR=10^{-5})	GAR (FAR=10^{-4})	GAR (FAR=10^{-2})
DenseNet-161 [49]	0.00%	0.00%	0.00%
EfficientNetv2-m [50]	0.00%	0.00%	0.03%
ResNet-101 [48]	0.00%	0.00%	0.05%
ViT-B [51]	0.00%	0.00%	0.78%
FKNet [7]	1.04%	2.74%	15.07%
RFNet-RSIL [6]	0.96%	2.89%	17.03%
STResNet [15]	0.08%	0.58%	14.26%
Ours	**55.41%**	**66.35%**	**80.95%**



(a) (b) (c) (d) (e)

Figure 8. Comparative ROC plots using other graph similarity models on (a) our dataset, (b) dataset in [47], and (c) dataset in [11]. Other ablation study results from our method are shown in (d)-(e), whose summary and details are in Supplementary D.

4.3. Estimation of Finger Knuckle Uniqueness

We adopt the KnuckleCreasePoint to extract keypoint templates from knuckle crease images, using the hand dorsal database [47]. Firstly, we estimate the position $\Delta_{\boldsymbol{k}}$ and feature $\Delta_{\boldsymbol{f}}$ threshold from the ground truth of correspondences to ascertain whether two random keypoints (one is from probe template, the other is from the gallery template) are matched or not. Such an estimation using dataset in [47] reveals that the tolerance parameter $\Delta_{\boldsymbol{k}}$ and $\Delta_{\boldsymbol{f}}$ should be set as 6.750 and 0.726, respectively. Therefore, we empirically set w, $m = n$, $\Delta_{\boldsymbol{k}}$, and $\Delta_{\boldsymbol{f}}$ to 17, 69, 6.750, 0.726 respectively, and set $\alpha = 0.05$. These parameters aim to reveal the variations in FRC score with the Poisson parameter λ_α. Our results are summarized in Tab. 5. Supplementary E explains parameter choice and presents additional results. The numbers in Tab. 5 should be interpreted in the context of the resolution of images acquired or used in this study, which is far below the resolution (about 500 dpi) widely employed for the fingerprint.

Table 5. Uniqueness analysis for finger knuckle patterns.

$\Delta_{\boldsymbol{k}}$	$\Delta_{\boldsymbol{f}}$	(m, n, w, α)	$\overline{p(T^{\mathrm{p}}, T^{\mathrm{g}})}$	$\overline{\lambda}_\alpha$	$\overline{\mathrm{FRC}}_\alpha$
6.750	0.726	(69,69,15,0.05)	2.980×10^{-4}	6.167	6.780×10^{-4}
6.750	0.726	(69,69,17,0.05)	2.980×10^{-4}	6.613	1.860×10^{-4}
6.750	0.726	(69,69,19,0.05)	2.980×10^{-4}	6.996	4.408×10^{-5}
6.750	0.726	(69,69,21,0.05)	2.980×10^{-4}	7.325	9.055×10^{-6}
3.750	0.726	(69,69,17,0.05)	1.031×10^{-4}	2.962	2.948×10^{-9}
5.250	0.726	(69,69,17,0.05)	1.914×10^{-4}	4.950	4.731×10^{-6}
8.250	0.726	(69,69,17,0.05)	4.119×10^{-4}	8.061	1.729×10^{-3}
9.750	0.726	(69,69,17,0.05)	5.273×10^{-4}	9.124	6.080×10^{-3}
6.750	0.626	(69,69,17,0.05)	6.210×10^{-6}	N/A	N/A
6.750	0.676	(69,69,17,0.05)	5.309×10^{-5}	1.367	1.546×10^{-14}
6.750	0.776	(69,69,17,0.05)	1.194×10^{-3}	12.476	8.305×10^{-2}
6.750	0.826	(69,69,17,0.05)	3.620×10^{-3}	14.893	2.421×10^{-1}

4.4. Ablation Study

Graph Similarity Model: A range of SOTA graph similarity models, GMN [26], SimGMN [28], MGMN [29], and ERIC [31], are employed to ascertain the effectiveness of our CGN. It can be observed from within-database and cross-database performance results in Fig. 8, that our CGN can significantly outperform these graph similarity models.
Module Efficiency: To ascertain the effectiveness of the proposed "convolution along the node dimension" (ConvNode) strategy, incorporating "positional encoding", and the arccos (cos^{-1}) efficiency. We comparatively evaluate the cross-database performance *with* and *without* these modules, and the efficiency of our modules can be proved in Fig. 8 on the hand dorsal dataset [47]. ConvNode is compared with other graph convolutions methods, including GIN [53], SAGE [54], GATv2 [55], and GCN [56].
Keypoint and Descriptor Statistics: Here, we follow the same approach as in SIFT [18] to evaluate the repeatability of detected knuckle crease keypoints and the reliability of descriptors of finger knuckles computed using KnuckleCreasePoint model, under 152×200 image resolution that is used for the CGN model.
More detailed explanation for the ablation study and the corresponding summary tables are in Supplementary D.6

5. Conclusions and Further Work

This paper has presented a new [2] and interpretable approach that has been shown to offer dramatic enhancement in match accuracy. Compared to the existing CNN-based, attention-based, and graph-based models, our models can significantly outperform baseline methods (both for within-dataset and cross-dataset evaluations (Sec. 4) under the challenging all-to-all protocols, indicating high generalization ability. Additionally, we developed and publicly release [52] a new multi-pose dataset from 351 different subjects. This work also marks the first attempt to establish the uniqueness (Sec. 4.3) of contactless 2D finger knuckle patterns, providing a theoretical upper limit on the expected performance of this biometric. Contributions from this paper are valuable not only for the forensic community but also for advancing contactless biometrics for civilian use.

Despite impressive results and contributions, there are several aspects to enhance this work. The GNN architecture considered in our work effectively processes structural cues from the knuckle crease keypoint correspondences but falls short in spatial awareness. Therefore, further extension of this model to incorporate spatial awareness is expected to enhance performance further and is suggested in future work.

[2] Algorithms introduced in this paper are part of U.S. Patent [42].

References

[1] Tim Ring. How the biometrics industry can help stop child abuse. *Biometric Technology Today*, 2016(11):9–11, 2016. 1

[2] Marco Santopietro, Richard Guest, Kathryn C Seigfried-Spellar, and Stephen J Elliott. A multi-factor knuckle and nail bed verification tool for forensic imagery analysis. *Child Abuse & Neglect*, 154:106910, 2024. 1, 3, 10

[3] Lilly Dan, Lucina Hackman, Vincenzo Rinaldi, and Emanuele Trucco. The persistence of knuckle creases during finger flexion for the identification of perpetrators from digital images of their hands. *Forensic Science International*, 348:111603, 2023.

[4] Freckle clue child abuser jailed. *BBC News* http://news.bbc.co.uk/1/hi/england/london/8014103.stm, April 2009. Accessed: 2024-11-11. 1, 10

[5] Ajay Kumar, Personal Communications, FBI, Quantico, USA, 2023. 1

[6] Zhenyu Zhou and Ajay Kumar. Completely contactless and online finger knuckle identification for real world applications. *IEEE Journal of Selected Topics in Signal Processing*, 17(3):648–661, 2023. 1, 2, 4, 6, 7

[7] Kevin HM Cheng and Ajay Kumar. Deep feature collaboration for challenging 3d finger knuckle identification. *IEEE Transactions on Information Forensics and Security*, 16: 1158–1173, 2020. 1, 2, 4, 6, 7, 3

[8] Gaurav Jaswal, Amit Kaul, and Ravinder Nath. Knuckle print biometrics and fusion schemes–overview, challenges, and solutions. *ACM Computing Surveys (CSUR)*, 49(2):1–46, 2016. 1

[9] Amazon One. https://one.amazon.com/how-it-works, 2024. Accessed: 2024-11-08. 1

[10] Palm Payment used in Beijing Subway Station. https://www.ourchinastory.com/en/11071/Palm%20payment%20used%20in%20Beijing%20subway%20station, 2024. Accessed: 2024-11-08. 1

[11] Ajay Kumar. Toward pose invariant and completely contactless finger knuckle recognition. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 1(3):201–209, 2019. 2, 6, 7, 8, 1, 3, 4, 5

[12] Wenming Yang, Zhiquan Chen, Chuan Qin, and Qingmin Liao. α-trimmed weber representation and cross section asymmetrical coding for human identification using finger images. *IEEE Transactions on Information Forensics and Security*, 14(1):90–101, 2018. 2, 6

[13] Kevin HM Cheng and Ajay Kumar. Contactless biometric identification using 3d finger knuckle patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8): 1868–1883, 2019. 2, 4, 6, 7

[14] Kevin HM Cheng and Ajay Kumar. Efficient and accurate 3d finger knuckle matching using surface key points. *IEEE Transactions on Image Processing*, 29:8903–8915, 2020. 2

[15] Zhenyu Zhou and Ajay Kumar. Finger-knuckle assisted slap fingerprint identification system for higher security and convenience. *IEEE Transactions on Information Forensics and Security*, 19:441–454, 2024. 2, 4, 6, 7

[16] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. *Advances in Neural Information Processing Systems*, 28, 2015. 2

[17] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 2

[18] David G Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004. 2, 8, 6

[19] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346–359, 2008. 2

[20] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *2011 International Conference on Computer Vision*, pages 2564–2571. Ieee, 2011. 2

[21] Kwang Moo Yi, Eduard Trulls, Vincent Lepetit, and Pascal Fua. Lift: Learned invariant feature transform. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI 14*, pages 467–483. Springer, 2016. 2

[22] Yuki Ono, Eduard Trulls, Pascal Fua, and Kwang Moo Yi. Lf-net: Learning local features from images. *Advances in Neural Information Processing Systems*, 31, 2018. 2, 3

[23] Mihai Dusmanu, Ignacio Rocco, Tomas Pajdla, Marc Pollefeys, Josef Sivic, Akihiko Torii, and Torsten Sattler. D2-net: A trainable cnn for joint description and detection of local features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8092–8101, 2019. 3

[24] Zixin Luo, Lei Zhou, Xuyang Bai, Hongkai Chen, Jiahui Zhang, Yao Yao, Shiwei Li, Tian Fang, and Long Quan. Aslfeat: Learning local features of accurate shape and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6589–6598, 2020. 3

[25] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition workshops*, pages 224–236, 2018. 3, 4, 1, 2

[26] Yujia Li, Chenjie Gu, Thomas Dullien, Oriol Vinyals, and Pushmeet Kohli. Graph matching networks for learning the similarity of graph structured objects. In *International Conference on Machine Learning*, pages 3835–3845. PMLR, 2019. 3, 8, 1, 5

[27] Yunsheng Bai, Hao Ding, Ken Gu, Yizhou Sun, and Wei Wang. Learning-based efficient graph similarity computation via multi-scale convolutional set matching. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3219–3226, 2020. 3

[28] Yunsheng Bai, Hao Ding, Song Bian, Ting Chen, Yizhou Sun, and Wei Wang. Simgnn: A neural network approach to fast graph similarity computation. In *Proceedings of the twelfth ACM International Conference on Web Search and Data Mining*, pages 384–392, 2019. 3, 5, 8, 1

[29] Xiang Ling, Lingfei Wu, Saizhuo Wang, Tengfei Ma, Fangli Xu, Alex X Liu, Chunming Wu, and Shouling Ji. Multilevel

graph matching networks for deep graph similarity learning. *IEEE Transactions on Neural Networks and Learning Systems*, 34(2):799, 2023. 3, 5, 8, 1

[30] Can Qin, Handong Zhao, Lichen Wang, Huan Wang, Yulun Zhang, and Yun Fu. Slow learning and fast inference: Efficient graph similarity computation via knowledge distillation. *Advances in Neural Information Processing Systems*, 34:14110–14121, 2021. 3

[31] Wei Zhuo and Guang Tan. Efficient graph similarity computation with alignment regularization. *Advances in Neural Information Processing Systems*, 35:30181–30193, 2022. 3, 5, 8, 1

[32] Yongfang Zhu, Sarat C Dass, and Anil K Jain. Statistical models for assessing the individuality of fingerprints. *IEEE Transactions on Information Forensics and Security*, 2(3): 391–401, 2007. 3, 6, 8

[33] John G Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1148–1161, 1993. 3

[34] John Daugman. The importance of being random: statistical principles of iris recognition. *Pattern Recognition*, 36(2): 279–291, 2003.

[35] John Daugman. Information theory and the iriscode. *IEEE Transactions on Information Forensics and Security*, 11(2): 400–409, 2015.

[36] Ruud M Bolle, Sharath Pankanti, Jonathan H Connell, and Nalini K Ratha. Iris individuality: A partial iris model. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 2, pages 927–930. IEEE, 2004.

[37] Bin Li, Zifei Yan, Wangmeng Zuo, and Feng Yue. Modeling the individuality of iris pattern and the effectiveness of inconsistent bit masking strategy. In *IEEE International Conference on Identity, Security and Behavior Analysis (ISBA 2015)*, pages 1–6. IEEE, 2015. 3

[38] Sargur N Srihari, Sung-Hyuk Cha, Hina Arora, and Sangjik Lee. Individuality of handwriting. *Journal of Forensic Sciences*, 47(4):856–872, 2002. 3

[39] Sharath Pankanti, Salil Prabhakar, and Anil K Jain. On the individuality of fingerprints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1010–1025, 2002. 3, 6, 7

[40] Sarat C Dass. Assessing fingerprint individuality in presence of noisy minutiae. *IEEE Transactions on Information Forensics and Security*, 5(1):62–70, 2010. 3, 6

[41] Ajay Kumar. Importance of being unique from finger dorsal patterns: Exploring minor finger knuckle patterns in verifying human identities. *IEEE Transactions on Information Forensics and Security*, 9(8):1288–1298, 2014. 3, 6, 2

[42] Ajay Kumar and Zhenyu Zhou. A method and device for biometric identification using finger images. *U.S. Patent No. 63/774,168*, 2025. 4, 8

[43] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superglue: Learning feature matching with graph neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4938–4947, 2020. 4, 2

[44] Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019. 4, 2

[45] Olivier Duchenne, Francis Bach, In-So Kweon, and Jean Ponce. A tensor-based algorithm for high-order graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2383–2395, 2011. 4

[46] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017. 5

[47] Ajay Kumar and Zhihuan Xu. Personal identification using minor knuckle patterns from palm dorsal surface. *IEEE Transactions on Information Forensics and Security*, 11(10): 2338–2348, 2016. 6, 7, 8, 1, 2, 3, 4, 5, 9, 10

[48] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 6, 7, 4

[49] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017. 6, 7, 4

[50] Mingxing Tan and Quoc Le. Efficientnetv2: Smaller models and faster training. In *International Conference on Machine Learning*, pages 10096–10106. PMLR, 2021. 6, 7, 4

[51] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 6, 7, 3, 4

[52] The Hong Kong Polytechnic University Completely Contactless Multipose Finger Knuckle Video Database. https://www4.comp.polyu.edu.hk/~csajaykr/fkvideo.html, 2025. Accessed: 2025-03-20. 7, 8, 3, 4, 6, 9, 10

[53] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018. 8, 1, 5

[54] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems*, 30, 2017. 8, 5

[55] Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks? *arXiv preprint arXiv:2105.14491*, 2021. 8, 5

[56] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016. 8, 5