

# A CNN-based Framework for Comparison of Contactless to Contact-based Fingerprints

Chenhao Lin, Ajay Kumar

**Abstract**—Accurate comparison of contactless 2D fingerprint images with contact-based fingerprints is critical for the success of emerging contactless 2D fingerprint technologies, which offer more hygienic and deformation-free acquisition of fingerprint features. Convolutional neural networks (CNN) have shown remarkable capabilities in biometrics recognition. However, there has been almost nil attempt to match fingerprint images using CNN-based approaches. This paper develops a CNN-based framework to accurately match contactless and contact-based fingerprint images. Our framework firstly trains a multi-Siamese CNN using fingerprint minutiae, respective ridge map and specific region of ridge map. This network is used to generate deep fingerprint representation using a distance-aware loss function. Deep fingerprint representations generated in such multi-Siamese network are concatenated for more accurate cross comparison. The proposed approach for cross-fingerprint comparison is evaluated on two publicly available databases containing contactless 2D fingerprints and respective contact-based fingerprints. Our experiments presented in this paper consistently achieve outperforming results, over several popular deep learning architectures and over contactless to contact-based fingerprints comparison methods in the literature.

**Index Terms**—Contactless and contact-based fingerprint, Sensor interoperability, Multi-Siamese CNN

## I. INTRODUCTION

FINGERPRINT recognition technology has become one of the most reliable approaches for human identification [1], [2]. Contactless 2D/3D fingerprint identification systems [3], [4] have been recently introduced to address the limitations of traditional contact-based fingerprint systems [5]. Development of advanced capabilities to accurately match contactless fingerprint images with contact-based fingerprints is critical for the success of such technologies as billions of fingerprints in legacy databases have been acquired using contact-based technologies. The studies in several publications [6]–[8] have indicated that the fingerprint comparison performance drops dramatically while matching the fingerprints acquired from different sensors, especially for the fingerprints acquired from contactless and contact-based sensors [9], [10].

Recent emergence of contactless fingerprint sensors and imaging convenience with smartphone sensors requires the development of specialized methods for matching contactless fingerprint images with those stored/acquired in legacy databases using contact-based fingerprint sensing technologies. These solutions have wide applications, especially in areas like forensics and e-business. Therefore, development of advanced algorithms to accurately match contactless to contact-based fingerprint images (cross-matching) has emerged as challenging research problem and is focus of our work in this paper.

The minutia feature is believed to be the most accurate fingerprint feature which has proved its efficiency and reli-

ability in fingerprint recognition [11], [12]. However due to significant differences in the nature of image acquisition technologies, between the contactless and contact-based fingerprint sensors, it is very difficult to accurately extract minutiae features and to ensure their correspondences in images from two such sensors. Contactless imaging itself is known to introduce significant intra-class variations which results from the high degree of freedom availed by fingers along the three axes during the sensing. In addition, different kinds of deformations in contactless (perspective distortion) and contact-based (elastic deformation) fingerprints also add to the challenges in such cross-fingerprint comparison problem than matching the fingerprints from different contact-based sensors. Although several fingerprint comparison methods [9], [13], [14] have been proposed to address such problem, the comparison performance remains far from the expectations for the deployments.

In recent years, the advancement in deep learning technologies have shown to offer remarkable success for the image recognition, classification and feature representation [15]–[17], particularly for the challenging biometrics problems [18], [52] such as for the face recognition [19], [20]. Matching biometric images using CNN approach shown to offer superior performance than traditional image-based or handcrafted feature based identification algorithms. In addition, recent publications [21], [22] have investigated fine-grained recognition problems using CNN approach. Fingerprint recognition, which also need to identify the images from same category, can be considered similar to fine-grained recognition problem.

### A. Related Work

In this section, we briefly review the related work on fingerprint sensor interoperability and CNN-based image recognition and representation. There have been many promising studies on the fingerprint sensor interoperability [7], [23]–[25]. Reference [24] has investigated fingerprint sensor interoperability using a multi-sensor fingerprint database acquired from different contact-based fingerprint sensors. In reference [7], the authors proposed an ‘average’ deformation model based on thin-plate model to address contact-based fingerprint sensor interoperability problem. Contact-based fingerprint sensor interoperability problem was also investigated in reference [25]. The experimental results in all above references have indicated degradation in the matching accuracy even for matching fingerprint images acquired from different contact-based fingerprint sensors.

In addition to investigating contact-based fingerprints sensor interoperability, the challenges in contactless to contact-based

fingerprint cross comparison have been detailed in recent references [9], [26]. Reference [27], [28] evaluated fingerprint sensor interoperability on commercial matchers using contact-based and contactless (2D/unwrapped 3D) fingerprints. Reference [14] investigated fingerprint mosaicking from three touchless fingerprint images and evaluated the interoperability between touchless and plain contact-based fingerprints using SIFT and minutiae features. The experimental results presented in all above references underline significant degradation in contactless to contact-based comparison accuracy. It may be important to underline that these studies have only evaluated the comparison performance instead of providing any new methods to address such problem.

More recently, reference [13] presented a new method to improve contactless to contact-based cross comparison performance by introducing a contact-based fingerprint deformation correction model. This is a more promising and efficient method to address the cross-comparison problem. However, the performance improvement is very limited and needs significant improvement for the real applications.

There have been many successful implementations of deep learning for the challenging pattern recognition problems and many CNN models have been introduced to achieve state-of-the-art performance for image recognition and classification [19], [29]. Siamese CNN consisting of two identical CNNs was introduced in [30] to learn the similarity metric from the face images. This approach has been quite promising in achieving higher performance for the face verification task. The Siamese CNN was further improved in [31] by combining two Siamese streams into two-channel with one stream network. The experimental results evaluated on image benchmark recognition validated this superior approach. The cross-domain sketch similarities using Siamese CNN were investigated in [32] and demonstrated significant performance improvement over the state-of-the-art approaches. The conference version of this paper [33] also investigated Siamese CNN architecture to address contact-based to contactless fingerprint cross comparison and illustrated preliminary results for improving the matching performance. Although these promising studies achieved high performance for image recognition and feature description using Siamese CNN, in the best of our knowledge, there have been very few attempts to address challenging problems in the fingerprint recognition using the deep learning algorithms.

In more recent years, fine-grained recognition problems have been investigated in the literature. Reference [21] proposed a triplet sampling algorithm to generate deep ranking model and learn the fine-grained image similarity. The experimental results demonstrated the superior performance of their approach than the other methods using hand-crafted visual features. A CNN-based bilinear model is proposed in reference [22] to generate fine-grained image descriptor. The network was firstly fine-tuned on ImageNet dataset to generate initialized parameters. The results demonstrated the effectiveness of their methods on various fine-grained recognition datasets. In reference [34], the authors proposed to train neural language models using CNN. The model incorporated raw text description with respective images and learned a



Fig. 1: Illustration of *high intra-class variation* between contactless fingerprints and respective contact-based fingerprints from different subjects due to intrinsic differences in the nature of imaging technologies and user habits/interaction.

scoring function between them. Their proposed approach significantly outperformed the state-of-art method on birds classification. A generic iterative framework was proposed in [35] for fine-grained categorization and dataset bootstrapping. A low dimensional feature embedding with anchor points was learned using deep metric learning. Based on these features, the authors retrained the proposed model. The experimental results on different datasets showed their approach achieved better performance than state-of-art methods. Inspired by all their successes in the above brief summary, it's meaningful and of significance to investigate the contactless to contact-based fingerprint cross comparison problem using deep learning.

### B. Open Challenges and Our Work

This paper develops a novel CNN-based framework to address the problem of accurately matching contactless 2D fingerprints with respective contact-based fingerprints. Our framework incorporates a robust multi-Siamese CNN to learn fingerprint minutiae feature correspondences. The proposed network improves Siamese CNN in both architecture and the loss function (as detailed in section II.D). The feature vectors generated from the multi-Siamese CNNs are concatenated to form more robust fingerprint deep feature representation, which is expected to incorporate more information to describe the similarities between contactless and contact-based fingerprints.

Although CNN-based approaches have achieved success in biometrics images recognition, matching cross-sensor fingerprint images using CNN is still a challenging problem, especially matching the fingerprints acquired from contactless and contact-based sensors. As compared with general biometrics recognition problem using same sensing technologies, relatively smaller inter-class variation and high intra-class variation among fingerprint images pose significant challenges for the comparison. The inter-class variation observed among fingerprint images is smaller, as compared to those from different objects used in general object recognition [29] problem, since these images are acquired from the same object/finger illustrating ridge-like pattern. On the other hand, due to significant differences in the nature of sensing technologies,

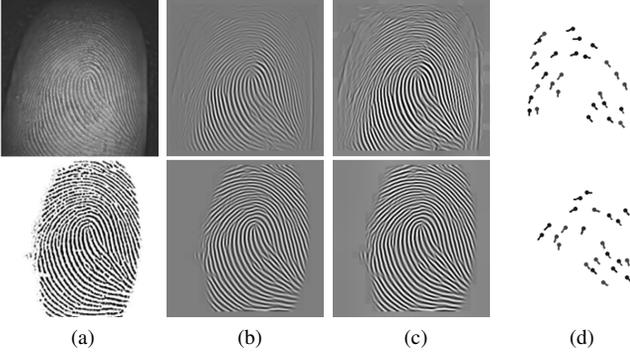


Fig. 2: (a) Contactless fingerprint and respective contact-based fingerprint from same subject (b) Enhanced fingerprints using Gabor filter (c) Enhanced fingerprints (ridge flow maps) using Gabor filter and adaptive histogram equalization (d) Minutiae maps generated from the fingerprints.

user interaction and habits, the cross-sensor images generally result in high intra-class variations. It makes the fingerprint cross-sensor comparison problem more challenging than the fingerprint recognition problem using the same sensor. Figure 1 illustrates the fingerprint image samples acquired from different fingers with *high intra-class variation*. Low inter-class variation and high intra-class variation among the cross-sensor fingerprint images also pose challenges for accurately matching fingerprint images using conventional CNN as the database available for training CNN is also quite limited. Therefore, this paper develops a specially designed CNN based framework to robustly extract deep feature representations that can more accurately describe the similarities between the contactless and contact-based fingerprint images.

The problem of missing minutiae or spurious minutiae is well known [1] to degrade fingerprint matching performance. One possible approach to increase the reliability of recovered minutiae features is to incorporate contextual information in the learning process for the minutiae feature correspondences from contactless and contact-based fingerprints. The fingerprint ridge patterns surrounding respective minutiae representations can provide such valuable contextual information. Different fingerprints may have similar structure but different ridge singularity which is embedded with texture details. Unlike face recognition, which largely benefits from the global high-level features [36], most fingerprint systems available today rely on the information extracted from the texture details (minutia) [1]. Any direct application of conventional convolution and pooling process on fingerprint ridge flow patterns can degrade important texture details. Therefore we propose to incorporate available ridge flow features with their singularity (minutiae features) in our framework, to enhance the learning process for the cross-fingerprint comparison. The effectiveness of such contextual information and the distance-aware loss function, introduced in section II.D, is also validated using separate experimentations. The key focus of this work has been to develop a specific deep learning based framework that can be used to accurately match a variety of *contactless* fingerprints with their corresponding *contact-based*

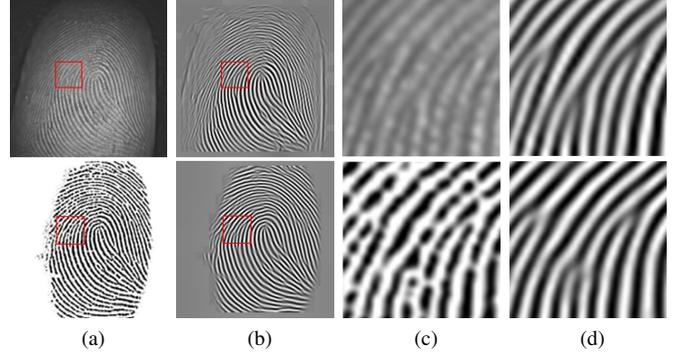


Fig. 3: (a) Original contactless and contact-based fingerprint (b) Enhanced fingerprints (c) Ridge details of original fingerprints (d) Ridge details of enhanced fingerprints.

fingerprints.

We present extensive experimental results on two different publicly available databases. Comparative experimental results presented in section III.E of this paper suggest that our specifically designed approach can significantly improve the performance over those from several popular CNN architectures and over state-of-the-art methods for the contactless to contact-based fingerprint cross comparison.

## II. CROSS FINGERPRINT COMPARISON FRAMEWORK

This section details proposed fingerprint cross-comparison framework. We briefly introduce the fingerprint image preprocessing and data augmentation approach for this framework in next subsection. We then elaborate the approach for simultaneously incorporating minutiae features and the respective ridge map and the approach for generating multiple inputs of proposed network is also described. The architecture and distance-aware loss function of proposed multi-Siamese CNN are described in following section. The details on generating fingerprint cross matching score are introduced in the last subsection.

### A. Data Preprocessing and Augmentation

Due to fundamental differences in the intrinsic nature of contact-based and contactless sensing technologies, the images from the same finger using such sensors appear quite different. Therefore it is necessary to incorporate fingerprint image preprocessing. The downsampling operation is further helpful to ensure same scale ratio or the size. The region of interest (ROI) in each of the images is automatically cropped using the center position of fingerprint impression. The sample fingerprint images acquired from the two sensors are illustrated in Figure 1. Histogram equalization is applied on all contactless fingerprints. In order to increase the similarity between the two cross sensor fingerprints, from the same subject/finger, Gabor filter in [11] and adaptive histogram equalization filter [37] are used to enhance these two kinds of fingerprint images. Figure 2 (a)-(c) illustrates such preprocessing of fingerprint images. As the ridge detail regions illustrated in Figure 3, we can see that such two steps enhancement operation improves the similarity of contactless and contact-based fingerprint images.

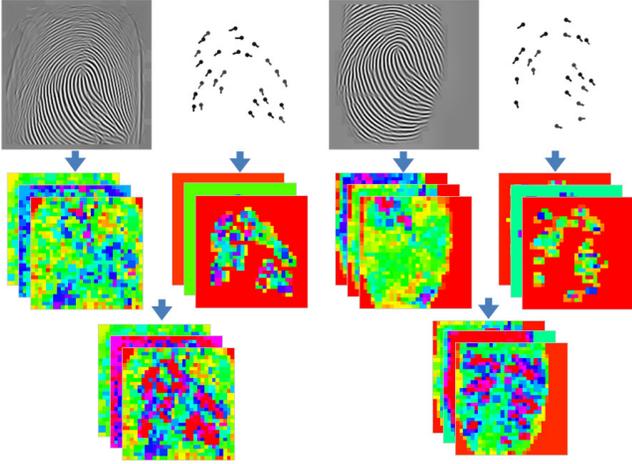


Fig. 4: Fingerprint ridge map and minutiae map fusion process and sample feature maps using HSV color space.

Data augmentation is a common strategy employed in the deep learning to increase the volume of training dataset. The contactless fingerprint images generally have larger angular rotations than the contact-based fingerprint images because of less restriction during the acquisition process. Hence, each contact-based training sample image is rotated by  $\pm 8$  and  $\pm 15$  degrees and each contactless training sample image is rotated by  $\pm 10$  and  $\pm 20$  degrees. This step increases the size of dataset five times and the further rotations of training sample images will not significantly improve the performance. The experimental results also indicate that the matching performance is improved with the fingerprint image enhancement operations and the data augmentation process.

### B. Fingerprint Ridge Map and Minutiae Features

The minutiae features are widely believed to be the most accurate [1] and employed in almost all the fingerprint technologies available today. However the degradation in matching accuracy due to missing or spurious minutiae is also well known in the literature [1]. This problem is more serious for the contactless fingerprints, largely due to the nature of contactless imaging that allows high degree of freedom in the acquisition under limited field of view and pronounced distortions in regions away from the image center. Several promising studies in the image recognition using CNN have also demonstrated the improved performance by incorporating contextual information or images with hand-crafted features [38], [39]. Therefore, it is judicious to incorporate (also experimentally justify) the minutiae feature and respective ridge feature to enhance the learning for the joint fingerprint feature correspondences from the two sensors.

The fingerprint minutiae extracted using the method in [11] can be represented as  $m = [x, y, \theta, q]$ , where  $x, y$  are minutiae location,  $\theta$  is the minutia direction and  $q$  is the minutia quality. For each fingerprint, minutiae map is generated by marking the minutiae location, direction (minutiae related ridge) and quality on the image. In order to avoid minutiae overlapping, each minutia point is represented using a solid circle with radius equal to 2 pixels. The minutiae



Fig. 5: Fingerprint ridge map, fingerprint core point region and fingerprint with blurred core point region.

direction is represented using 8 pixels' short line along the fingerprint ridge. Minutiae quality values from 0 to 1 are mapped into 0 to 255 so that it can be represented by using gray values. Figure 2 (c)-(d) illustrate the enhanced fingerprints (ridge flow maps) and related minutiae maps. The feature level combination in our architecture is expected to generate more accurate fingerprints feature correspondences for the cross-comparison. The ridge flow map and related minutiae map fusion process of the fingerprints from same subject is illustrated in Figure 4. The feature maps generated after first convolution layer and generated from concatenation layer using HSV color space are also shown in Figure 4. The feature map  $g(\cdot)$  generated from enhanced *ridge flow map* ( $rm$ ) samples of the contactless and corresponding contact-based fingerprints can be represented using  $g(rm)_{cl}$  and  $g(rm)_{cb}$  respectively. We use  $g(mm)_{cl}$  and  $g(rmm)_{cl}$  to respectively represent features generated from *minutiae map* ( $mm$ ) samples and our combination ( $rmm$ ) of contactless fingerprint samples. The average Euclidean distance  $D$  between each two feature maps among 6 samples from same finger (Fig.4 as example case) is used to represent the similarity of two feature maps. The average normalized distance among such 6 enhanced ridge flow map samples from this finger (Fig.4) is  $D(g(rm)_{cl}, g(rm)_{cb}) = 0.0429$  and the average distance among 6 minutiae map samples is  $D(g(mm)_{cl}, g(mm)_{cb}) = 0.0675$ . With the proposed combination, the average distance reduces to  $D(g(rmm)_{cl}, g(rmm)_{cb}) = 0.0226$  for the samples from this finger considered for illustration.

In order to generate more robust deep feature representation from the proposed network, besides of fingerprint ridge map and corresponding minutiae map, the specific regions of the fingerprint ridge map based on fingerprint core point are also extracted for the input of proposed network. The specific regions include fingerprint core point region and blurred core point region. This operation, i.e. incorporating the specific regions, can help the network to learn relative ridge pattern details that may be lost due to the convolution and pooling operation on fingerprint ridge map.

Fingerprint core point is automatically detected using the method in [40]. For the fingerprints, which have more than one core point, we always select the point with smaller horizontal value as the core point. For the fingerprints, which have no core point, we crop the center region of fingerprint images as core point region. Based on the core point, we crop the  $120 \times 120$  pixels surrounding region as the core point region from the fingerprint ridge map ( $192 \times 192$  pixels). The ridge map with blurred core point region ( $164 \times 164$  pixels) is generated by applying Gaussian blur (filter size equals to 13 and sigma

TABLE I: Configurations of each single CNN for each of three sub-networks

Operation	Sub-net1			Sub-net2			Sub-net3		
	Filter Size	Output Num	Stride	Filter Size	Output Num	Stride	Filter Size	Output Num	Stride
Conv1	11	64	3	7	64	3	5	64	1
Max Pool1	3	-	2	3	-	2	3	-	2
Concat	-	128	-	-	-	-	-	-	-
Conv2	5	256	1	3	256	1	3	256	1
Max Pool2	3	-	2	3	-	2	3	-	2
Conv3	3	512	1	3	512	1	3	512	1
Max Pool3	3	-	2	3	-	2	3	-	2
Conv4	3	1024	1	3	1024	1	3	512	1
FC	-	1024	-	-	1024	-	-	1024	-

equals to 2) to the resized ridge map image on  $80 \times 80$  pixels region surrounding the core point position. The contactless fingerprint image samples for core point region and the ridge map with blurred core point region are illustrated in Figure 5.

### C. Networks for Cross-Comparison Framework

In this work, our goal is to accurately match the fingerprints acquired from two different sensors. The contactless and contact-based fingerprints from same subject generally have different shapes, rotations and noises from range of sources (like residue of leftover latent, etc.). Therefore under such high inter-class and intra-class variations, it is difficult to achieve accurate feature correspondences using traditional CNN. Siamese CNN [30] includes twin CNN which benefits from the shared weights in these CNN and can offer better performance with enhanced learning capability. It learns a similarity metric between the two channels input data and is specifically incorporated to address inherent challenges for the contactless to contact-based fingerprint comparison problem.

The proposed fingerprint cross comparison framework takes the advantages of existing hand-crafted fingerprint features and Siamese CNN architecture, and differs from it in both architecture and loss function. This framework mainly consists of three sub-networks and each of them contains two single CNNs. The inputs of the three sub-networks are fingerprint ridge map and minutiae map (sub-net1), ridge map with blurred core point region (sub-net2) and core point region of ridge map (sub-net3) respectively. For each CNN in sub-net1, the network structure includes four convolution layers, each one followed by a max pooling layer, one concatenation layer, one fully connected layer. For each CNN in sub-net2, the network structure includes four convolution layers and max pooling layers and one fully connected layer. Each CNN in sub-net3 consists of four convolution layers and max pooling layers and one fully connected layer. The distance-aware loss function introduced in section II.D is jointly used for each two CNNs in each sub-nets. Each two corresponding convolution layers from two channels CNNs share the same weights. Parametric Rectified Linear Unit (PReLU) [16] is applied after the fully connected layer. The configurations of each single

CNN for each of three sub-networks are illustrated in Table I. The architecture of sub-net1 is shown in Figure 6. The deep feature representations generation process from each three sub-nets, for the contactless fingerprint, is shown in Figure 7.

### D. Distance-Aware Loss Function

The Siamese CNN framework consists of two CNNs and one contrastive loss function. We represent the given pair of input images as  $(X_1, X_2)$ , *i.e.* one contactless fingerprint and one contact-based fingerprint for our problem. Let  $W$  represent the learned parameters in this network,  $Y$  represents the label of the input pair images, if input images  $(X_1, X_2)$  belong to the same subject or finger (Genuine pair)  $Y = 0$  and otherwise (Imposter pair)  $Y = 1$ .  $d^n(W, X_1, X_2)$  represents the  $L_2$ -norm distance to measure the similarities between the output  $(J_W(X_1), J_W(X_2))$  generated by mapping  $(X_1, X_2)$ .  $n$  represents the  $n^{th}$  sample pair in each batch and  $N$  represents the batch size. The loss function includes two parts, genuine and imposter part, which can be represented as,

$$L(W, (Y, X_1, X_2)^n) = Y \max(M_1 - d^n(W, X_1, X_2), 0)^2 + (1 - Y) (d^n(W, X_1, X_2))^2 \quad (1)$$

$$d^n(W, X_1, X_2) = \|(J_W(X_1) - J_W(X_2))\| \quad (2)$$

The loss function requires the distance for imposter pairs larger than margin  $M_1$ . The training phase trains the Siamese network to reduce the distance among the genuine pairs and increase the distance for imposter pairs. The Siamese CNN has shown [30], [41] to offer better performance than the traditional CNN on image recognition and classification. However further improvement is required for challenging applications with high inter- and intra-class variations, particularly those with limited amount of training data. In general, more overlapping region between genuine and imposter distribution will result in higher probability of matching error. Our goal is to reduce this overlapping region. The imposter part  $\max(M_1 - d^n(W, X_1, X_2), 0)^2$  of loss function indicates that the network focuses on training the imposter pairs with distance smaller than margin  $M_1$ . It helps to reduce the

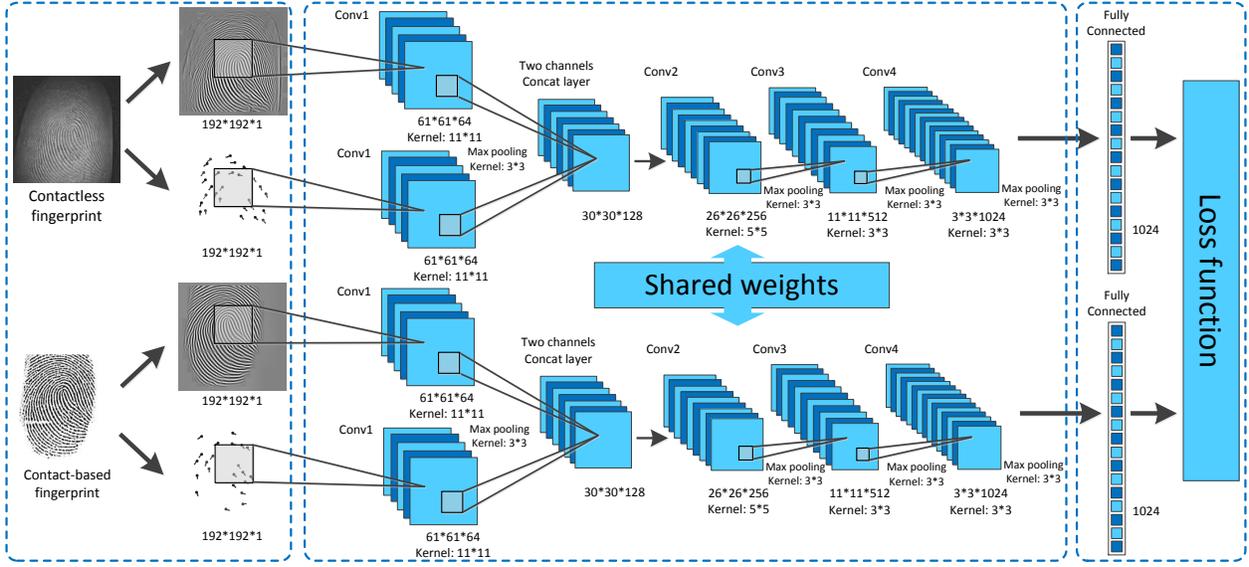


Fig. 6: Architecture of our contactless to contact-based fingerprint cross comparison network. Including three parts: fingerprint ridge map and minutiae map generation, the Siamese CNN and the distance-aware loss function.

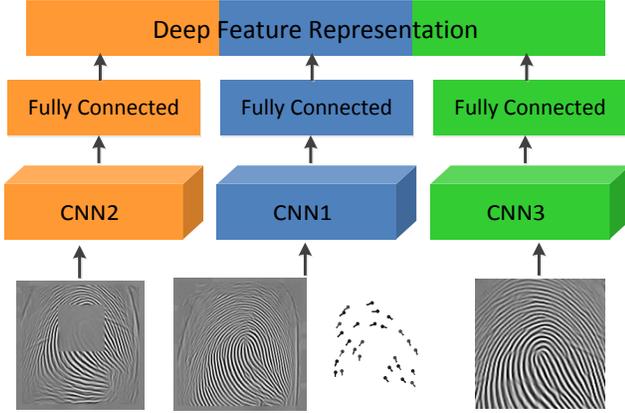


Fig. 7: Deep feature representation generation process for contactless fingerprint from each three sub-nets.

overlapping region of genuine and imposter distribution by increasing the distance between imposter pairs. However, the original contrastive loss function is not robust because there is no specific constraint on training genuine pairs. Therefore, we introduce an important parameter  $M_2$  to add constraint on the genuine part of the loss function. Such loss function can account for the distance between genuine pairs and is therefore referred as distance-aware loss function. This distance-aware loss function is motivated by the success of double-margin contrastive loss function in [54] and can be represented as follows:

$$\begin{aligned}
 L(W, (Y, X_1, X_2)^n) &= I(W, (Y, X_1, X_2)^n) \\
 &\quad + G(W, (Y, X_1, X_2)^n) \\
 &= Y \max(M_1 - d^n(W, X_1, X_2), 0)^2 \\
 &\quad + (1 - Y) \max(d^n(W, X_1, X_2) - M_2, 0)^2
 \end{aligned} \quad (3)$$

This genuine part  $\max(d^n(W, X_1, X_2) - M_2, 0)^2$  of this loss function can ensure that the network focus on training the

genuine pairs with larger distance  $M_2$ . It aims to focus on training the challenging genuine samples (genuine pairs with large distance or intra-class variations) to ensure robustness of the network in responding to intra-class variations. In such way, the overlapping region between genuine and imposter distribution can be further decreased. Hence further improvement in matching performance is expected.

Given this loss function, the stochastic gradient decent approach is used to optimize the Siamese CNN framework for cross-comparison. The gradient of the distance-aware loss function consists of two parts and can be computed from the following equations,

$$\frac{\partial L(W, (Y, X)^n)}{\partial w} = g(w, (Y, W)^n) + i(w, (Y, X)^n) \quad (4)$$

For each imposter pair,  $i(w, (Y, X)^n)$  is same as original imposter part of contrastive loss function. For each genuine pair, if  $d^n(w, x_1, x_2) < M_2$ ,

$$\frac{\partial L(W, (Y, X)^n)}{\partial w} = 0 \quad (5)$$

if  $d^n(w, x_1, x_2) > M_2$ ,

$$\begin{aligned}
 \frac{\partial L(W, (Y, X)^n)}{\partial w} &= \frac{\partial L}{\partial d^n(w, x_1, x_2)} \cdot \frac{\partial d^n(w, x_1, x_2)}{\partial w} \\
 &= 2 \left( \left| (J_w(x_1) - J_w(x_2)) \right| - M_2 \right) \\
 &\quad \cdot \frac{\partial \left| (J_w(x_1) - J_w(x_2)) \right|}{\partial w}
 \end{aligned} \quad (6)$$

The genuine and imposter pairs are expected to achieve more robust classification with the usage of this loss function. The experimental results (in section III.E) with test on statistical significance, also illustrate noticeable improvement in matching performance with the usage of this loss function. In addition, we follow the processes described in [30] and [31] to validate the effectiveness of proposed loss function respectively. The comparative experimental results in section III.A can further support/evaluate the loss function.

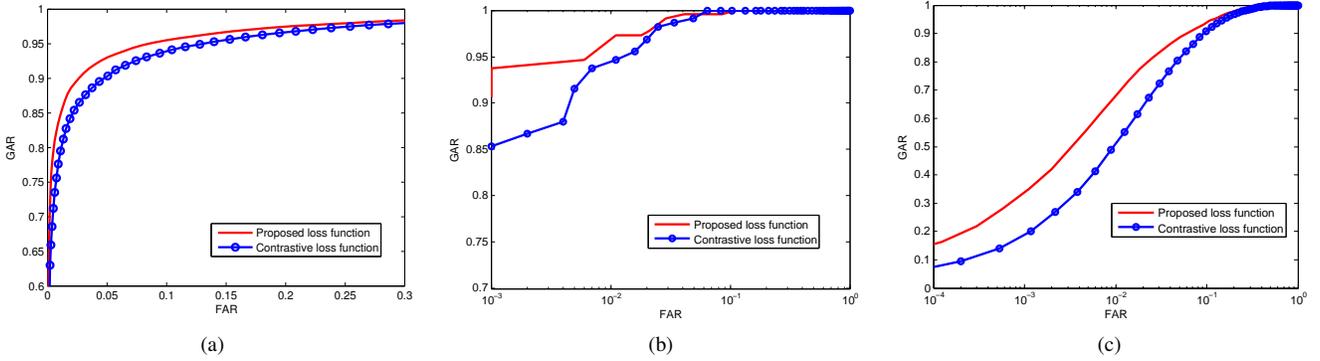


Fig. 8: (a) ROC curves using distance-aware loss function and contrastive loss function on UBC patch dataset (b) ROC curves using distance-aware loss function and contrastive loss function on AT&T dataset (c) ROC curves using distance-aware loss function and contrastive loss function on AR dataset.

TABLE II: Comparative experimental results using method/protocol in [31] and [43]

Dataset		False positive rate at 95% recall (FPR95)			
Train	Test	Distance-aware loss	Contrastive loss	Siamese with Euclidean loss in [31]	Siamese with Euclidean and Hamming loss in [43]
Notre Dame	Liberty	<b>8.04%</b>	12.34%	13.24%	8.50%

TABLE III: Comparative experimental results using method/protocol in [30]

Experiments		False Accept Rate (%)			
		10%	7.5%	5%	1%
AT&T Dataset	<b>Distance-aware loss function</b>	<b>0.00%</b>	<b>0.00%</b>	<b>0.44%</b>	<b>2.67%</b>
	Contrastive loss function	0.00%	0.00%	0.89%	5.67%
	Contrastive loss function reported in [30]	0.00%	1.00%	1.00%	N/A
AR Dataset	<b>Distance-aware loss function</b>	<b>6.15%</b>	<b>8.28%</b>	<b>11.52%</b>	<b>31.96%</b>
	Contrastive loss function	9.05%	12.86%	19.03%	48.51%
	Contrastive loss function reported in [30]	11.00%	14.60%	19.00%	N/A

### E. Generating Cross Comparison Score

Similar to the feature fusion processes in many publications [35], [36], during the testing phase, the extracted feature vectors for contactless/contact-based fingerprints from the fully connected layer of each three sub-nets are concatenated as one deep feature representation. Each test fingerprint image sample can be represented by three  $1-d$  vectors of length 1024. Let  $f(\cdot)$  be the function to generate feature vector from each sub-net of proposed network. Then the six feature vectors of contactless ( $cl$ ) and contact-based ( $cb$ ) fingerprint pairs generated from three sub-nets can be represented as  $(f(a_{cl}), f(a_{cb}))$ ,  $(f(b_{cl}), f(b_{cb}))$  and  $(f(c_{cl}), f(c_{cb}))$ , where  $a$  represents the input combination of fingerprint ridge map and minutiae map,  $b$  represents the input ridge map with blurred core point region and  $c$  represents the input core point region of ridge map. For each of the test or unknown fingerprint image pair, the corresponding enhanced ridge map and minutiae map, core point region of ridge map and ridge map with blurred core point region are generated automatically. After the feature fusion, the robust deep feature can be represented as  $(f((a, b, c)_{cl}), f((a, b, c)_{cb}))$ . Then the similarity between the deep feature representation of each contactless and contact-based fingerprint pair can be computed from the following

equation,

$$S = Dis(f((a, b, c)_{cl}), f((a, b, c)_{cb})) \quad (7)$$

where  $Dis$  represents the Euclidean distance between each two deep feature representations generated from the input pair.

## III. EXPERIMENTS

In this section, we detail on a range of experiments that were performed to evaluate the proposed approach for contactless to contact-based fingerprint cross comparison. Our fingerprint cross comparison network is implemented using the deep learning framework Caffe [42]. In the following experiments, stochastic gradient descent algorithm is used to optimize the network. The initial learning rate  $lr$  is set at 0.001 and decreases by following equation  $lr_{new} = lr * (1 + gamma * iter)^{-power}$ , where  $gamma$  is 0.0001 and  $power$  is 0.75. Weight decay and momentum we select for training the network are 0.005 and 0.9.  $M_1 = 10$  and  $M_2 = 5$  are selected as the parameters of distance-aware loss function. The imposter pairs are three times larger than the genuine pairs for effectively training the network. We employed two publicly available datasets in our experiments to evaluate the proposed contactless to contact-based fingerprint cross-comparison framework.



Fig. 9: Contactless fingerprint and corresponding contact-based fingerprint from same subject in dataset B.

#### A. Evaluation of Distance-aware Loss Function

In order to illustrate the robustness of proposed distance-aware loss function, before applying proposed method on fingerprint cross-sensor comparison problem, we first performed comparative experiments using Siamese-based network with traditional contrastive loss and distance-aware loss function on popular face recognition [30] and patch recognition [31], [43] problems.

In recent references [31] and [43], the authors have presented several experimental results using Siamese-based network with different loss functions on a benchmark dataset [44]. This dataset includes three subsets, Yosemite, Notre Dame, and Liberty and each of them contains more than 450,000 image patches. Since the Yosemite set is not available from [44], we use Notre Dame set for training and Liberty set for testing. The well-defined training set contains 500,000 patch pairs and the test set contains 100,000 patch pairs with equal number of matching and non-matching patch pairs. Reference [31] and [43] use the same training and testing protocol, we also use this protocol for making a fair comparison. We follow the network structure in [43] but using distance-aware loss function. The evaluation criterion is ensured to be the same as in [31] and [43], i.e. false positive rate at 95% recall (FPR95) and ROC curves. The comparative experimental results in Table II and Figure 8 (a) consistently confirm the effectiveness of the distance-aware loss function. Since the data augmentation operation in [43] is not described in details, we also attempted different rotation-based data augmentation and we achieved better performance by using distance-aware loss function. The network structures in [31] and [43] are specially designed for respective proposed structure and loss function, which may result in poor performance using Siamese network with contrastive loss, than the methods in [31] and [43].

In addition, In reference [30], the authors evaluated Siamese network on two face datasets [45], [46]. For the first dataset [45], it includes 400 images from 40 subjects, with variations in lighting, facial expression accessories, and head position. The training set consists of 350 images for 35 subjects and test set consists of 50 images for 5 subjects. The second dataset [46] contains 3536 images for 136 subjects with 26 images per subject. The images with expression variations, lighting variations, dark sunglasses and face-obscuring scarves make the dataset extremely challenging. It is combined with a subset (randomly selected) of [47] as the training dataset

that includes 2496 images from 96 subjects in [46] and 1122 images from 187 subjects in [47]. The test set contains 1040 images from 40 subjects in [45]. In order to ensure a fair comparison, we used the same training and testing protocol and the network structure described in [30] but using distance-aware loss function. The evaluation criterion is ensured to be the same as in [30], i.e. false reject rate for different false accept rates. In addition, we plot ROC curves for performance evaluation. The comparative experimental results in Figure 8 and Table III consistently illustrate significant performance improvement and confirm the effectiveness of the distance-aware loss function.

#### B. Training and Testing Dataset A

The first dataset [48] consists of 5760 contactless and respective contact-based fingerprint images acquired from 320 fingers. A total of 3840 fingerprints from 160 fingers were used for training the network and remaining fingerprints were used for the performance testing. Each training finger contained 12 contact-based and contactless fingerprint samples while each testing finger had 6 contact-based and contactless fingerprint samples. Each fingerprint image was resized into the same resolution ( $192 \times 192$ ). The training dataset contained 19200 ( $160 \times 60 \times 2$ ) fingerprint images after implementing data augmentation approach detailed in section II.A.

#### C. Fine-tuning Proposed Network on Dataset B

In order to validate the robustness of proposed cross comparison approach, another public available dataset [10] is also utilized. This dataset contains 1500 fingers data with 3000 contactless fingerprint samples and 6000 corresponding flat contact-based fingerprint samples. In this experiment first 1000 fingers containing 6000 fingerprint samples were used to evaluate the proposed method. A total of 500 fingers images from the dataset B were used for fine-tuning the fully connected layer of the proposed network that has been trained using dataset A. The rest of 500 fingers images were used for the testing. In order to fine-tune this network, each fingerprint image was also resized into the same resolution ( $192 \times 192$ ). Each of the training fingerprint ridge map and related minutiae map were rotated by  $\pm 3$  and  $\pm 6$  degrees. The image samples of contactless and contact-based fingerprints from the dataset B are illustrated in Figure 9.

#### D. Evaluation Protocol

In order to demonstrate the effectiveness of proposed method, we not only compare the proposed method with the other popular CNN architectures, but also present comparison with more accurate minutiae based method recently presented for the same problem in [13]. Since all the image samples in both the databases are complete or acquired in advance under supervision, the failure to capture or extract case [57] is not considered in both databases. We use ROC (Receiver Operating Characteristic) and EER (equal error rate) to ascertain the performance from the proposed method. The CMC (cumulative match characteristics) and rank-one accuracy are

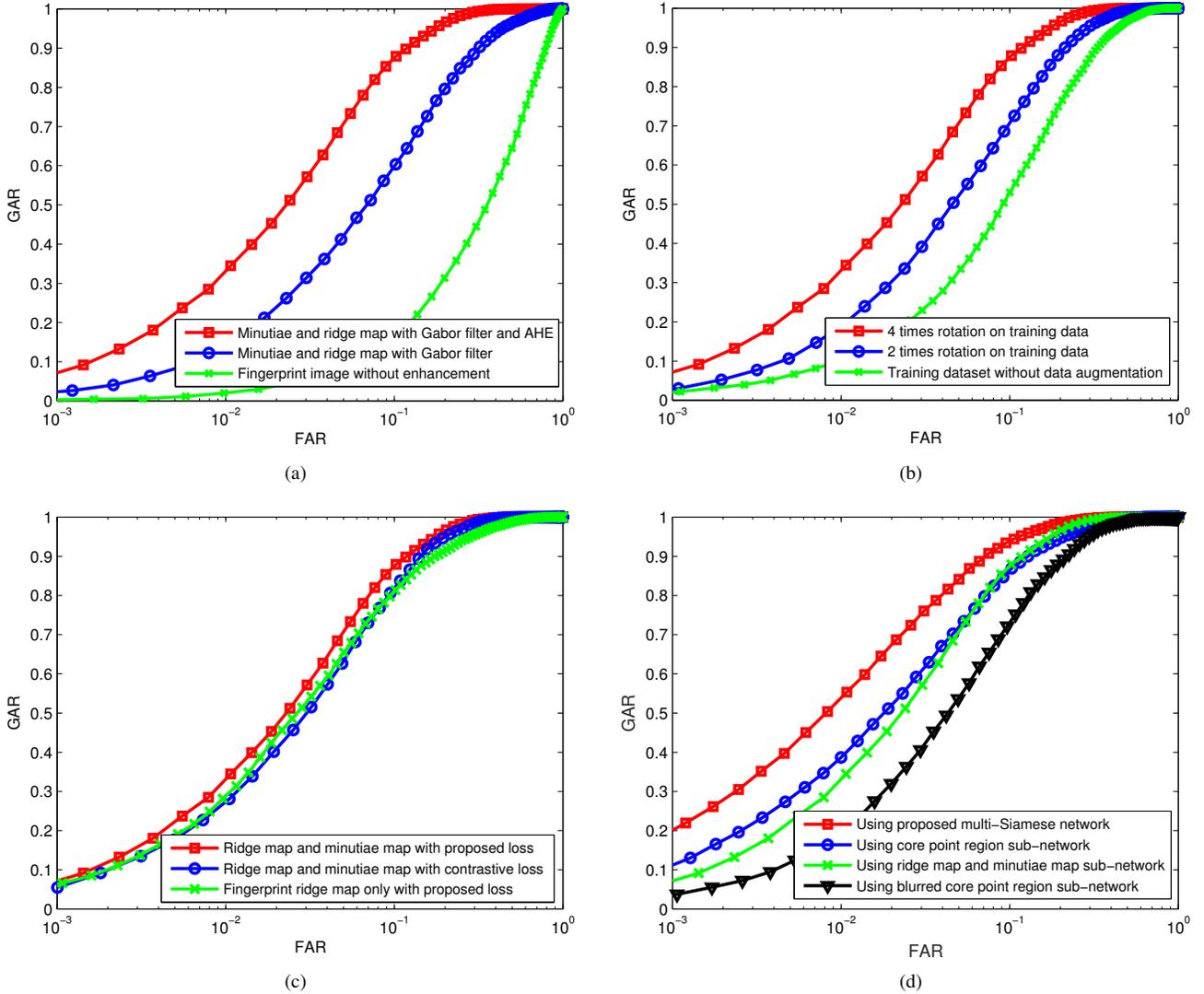


Fig. 10: (a) The ROC curves for cross comparison using minutiae map (MM) and ridge map (RM) with Gabor filter and adaptive histogram equalization(AHE), with Gabor filter only and fingerprint image without enhancement (b) ROC curves for cross comparison using 4 times rotated training data, 2 times rotated training data and training data without augmentation (c) ROC curves of our method using MM and RM with distance-aware loss function, MM and RM with contrastive loss function and RM with distance-aware loss function (d) ROC curves for cross comparison using whole multi-Siamese CNN and three subnetworks.

also employed to ascertain the cross comparison performance for the recognition problem. In addition, we also perform test for the statistical significance ( $p$ -value as also employed in [49], [51]) to ascertain significance of the performance improvement using the ROC. For two ROC curves, it can be reasonably established that the area under the ROC curve 1 is significantly different from the area under ROC curve 2 if the respective  $p$  value is smaller than 0.05 ( $p$ -value < 0.05) [49].

### E. Experimental Evaluations

This section provides a range of comparative experimental results. For contactless to contact-based fingerprint cross comparison using dataset A, 5760 ( $160 \times 36$ ) genuine matching scores and 915840 ( $160 \times 6 \times 159 \times 6$ ) imposter matching scores were generated. We generated 4000 ( $500 \times 8$ ) genuine matching

scores and 1996000 ( $500 \times 2 \times 499 \times 4$ ) imposter matching scores for the cross comparison experiments using dataset B.

In the first two experiments, we evaluated the effectiveness of fingerprint preprocessing and data augmentation approach using fingerprint ridge map and minutiae map. The Euclidean distance was computed as the matching score between each of the two test data representations. As we can observe from the ROC curve in Figure 10 (a), the fingerprint cross comparison performance has significant improvement with the usage of Gabor filter and the adaptive histogram equalization. The EER achieved 10.95% when using enhanced fingerprint ridge map with two filters and the related minutiae map. It is 19.19% when using enhanced fingerprint ridge map with Gabor filter and related minutiae map. We also evaluated the results of data augmentation in the second experiment. Although Siamese CNN architecture generated much more input data pairs than

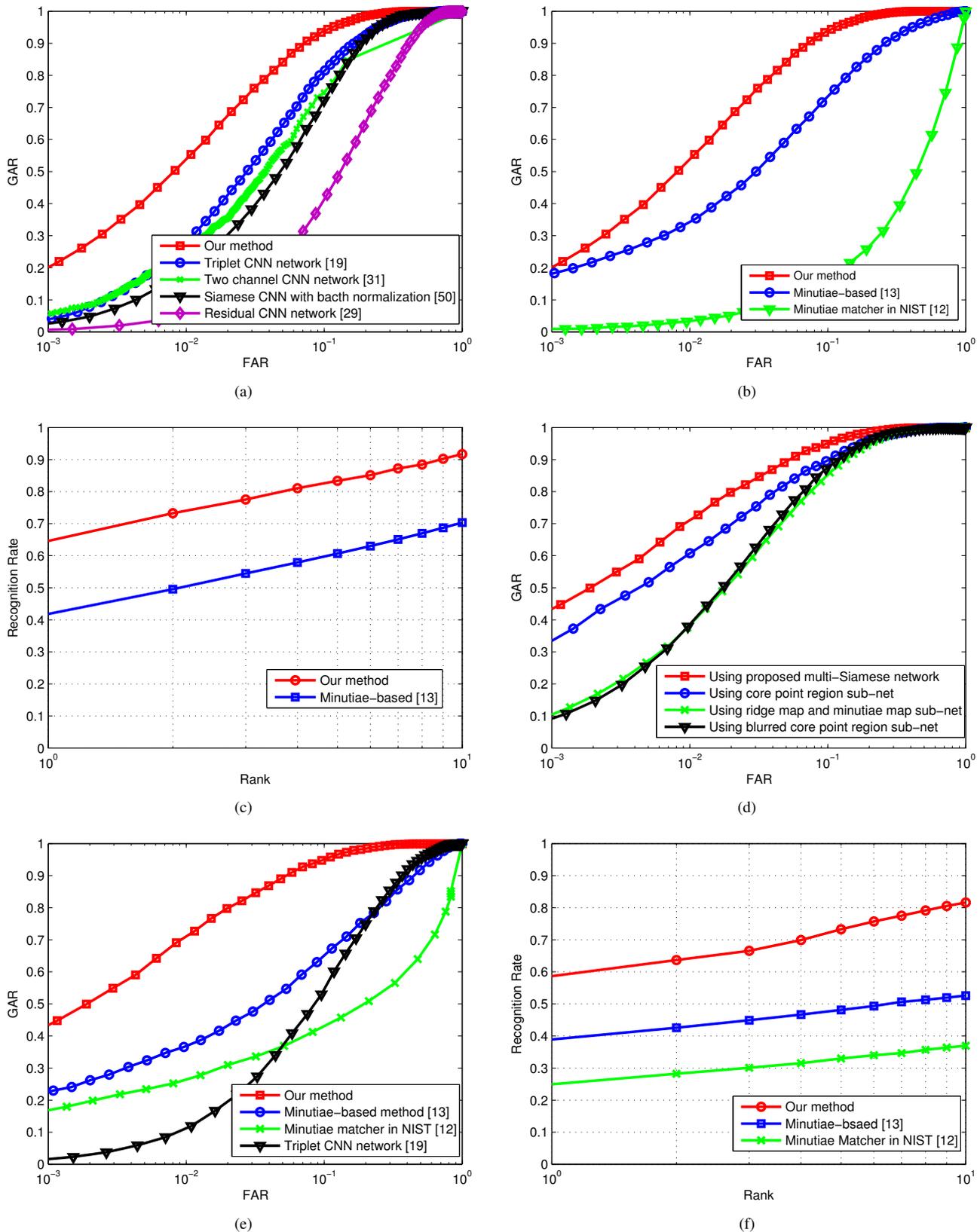


Fig. 11: (a) ROC curves for cross comparison using our method and other CNN-based methods. (b) ROC curves for cross comparison using our method and minutiae-based methods. (c) CMC curves for cross comparison using our method and minutiae-based method. (d) ROC curves on dataset B for cross comparison using whole multi-Siamese CNN and three subnetworks. (e) ROC curves on dataset B for cross comparison using our method and other methods. (f) CMC curves on dataset B for cross comparison using our method and other methods.

TABLE IV: Experimental evaluations on two datasets

Experiments	Equal Error Rate	Rank-one accuracy
Deformation correction model [13] on dataset A	16.17%	41.82%
Minutiae matcher in NIST [12] on dataset A	43.83%	10.99%
<b>Our method</b> on dataset A	<b>7.93%</b>	<b>64.59%</b>
Deformation correction model [13] on dataset B	21.60%	38.90%
Minutiae matcher in NIST [12] on dataset B	38.01%	24.92%
<b>Our method</b> on dataset B	<b>7.11%</b>	<b>58.87%</b>

traditional CNN, in our particular case, the size of the data was not large enough to train a robust deep network. As illustrated in Figure 10 (b), matching performance is improved by incorporating data augmentation.

Then the efficiency of proposed approach to incorporate fingerprint ridge maps and related minutiae maps was evaluated. The architecture of designed network for training fingerprint images only is similar to the network in Figure 6 but by removing Concat layers. The comparative experiment to evaluate the proposed distance-aware loss function was also performed. The proposed cross comparison method was used for training and testing on dataset A. It can be observed from the ROC in Figure 10 (c) that training the network using proposed approaches, *i.e.* incorporating fingerprint ridge map with the related minutiae map, with distance-aware loss function achieved better performance than only using the fingerprint ridge maps. The EER values of using proposed approach and only using fingerprint ridge map are 10.95% and 13.46% respectively. Our comparative results also indicated that fingerprint cross comparison using the distance-aware loss function achieved better performance than that using the traditional contrastive loss function ( $EER = 12.91\%$ ). As compared with the approach using contrastive loss function and the approach using fingerprint ridge map only, the improvement using distance-aware loss function was statistically significant ( $p\text{-value} < 0.0001$ ).

In addition, several experiments were performed to evaluate the effectiveness of proposed multi-Siamese network. The ROC curves of the results for cross-fingerprint comparison by separately using three sub-networks and the whole network were illustrated in Figure 10 (d). Our results demonstrate that more robust/effective deep fingerprint feature representations can be learned from the proposed multi-Siamese network. From the ROC curves, we can see that better performance for the cross-fingerprint comparison can be achieved by using the proposed multi-Siamese network framework than using individual sub-network. The EER is further decreased to 7.93% using the proposed approach.

Several experiments were performed to evaluate the proposed fingerprint cross comparison approach by comparing with other competing or popular methods. We firstly compared the proposed approaches with four promising CNN-based methods such as *Resnet* [29], *triplet* network [19] and in [50] or [31]. According to the network structure in [31], a

TABLE V: Results of statistical significance test ( $p\text{-value}$ ) for comparison of ROCs

Comparison	$p\text{-value}$	
	Database A	Database B
Distance-aware loss function & Contrastive loss function	< 0.0001	/
Our method & Method in [13]	< 0.0001	< 0.0001
Our method & Method in [19]	< 0.0001	< 0.0001

6-layer 2-channel network was implemented using fingerprint ridge map and related minutiae map as the input. In order to implement Residual network [29], contactless and contact-based fingerprints from the same subject were combined as one class. Then fingerprints were classified into 160 classes by this network. Since the size of fingerprint dataset is not large enough for training very deep network, we implemented a 6-layer Resnet following the network structure in [29]. In order to ascertain the effectiveness of batch normalization [50], we added batch normalization layer after each convolution layer of our network. We also generated the fingerprint dataset for training a 6-layer triplet network [19]. The testing fingerprint deep representations were extracted from the fully connected layer of above networks for the cross comparison. The matching performances are evaluated using the ROC and shown in Figure 11 (a). Although these methods achieve great success in image recognition or classification, in our particular case the proposed method achieves outperforming results than from these competing methods.

Another competing method appears in [13] as it provides comparison with some methods in the literature. Therefore, this approach and the method in [12] were also evaluated for the cross fingerprint comparison performance. In order to ensure fairness in comparison and to achieve the best performance of conventional minutiae based method, the contactless fingerprint images with  $512 \times 384$  resolution and the contact-based fingerprint with  $328 \times 356$  resolution in database A were used for minutiae-based matching. The resolution of contactless fingerprint images and contact-based fingerprints used for minutiae-based matching in database B is  $512 \times 640$  and  $640 \times 480$  respectively. Such comparative results in Figure 11 (b) illustrate that the proposed approach ( $EER = 7.93\%$ ) can significantly improves the state-of-the-art method ( $EER = 16.17\%$ ) in [13]. Contactless to contact-based fingerprint cross comparison performance was also evaluated using the CMC curve and rank-one accuracy in Figure 11 (c) and Table IV. The rank-one accuracy is improved to 64.59% compared with the best performing method (41.82%). The statistical significance levels ( $p\text{-values}$ ) were also computed for the above experiments. Comparing the ROC of proposed method with the ROC of other methods, the improvement of proposed method was shown to be statistically significant ( $p\text{-values} < 0.0001$ ).

We also performed comparative experiments on dataset B to ascertain the robustness of proposed framework. The comparative ROC curves, CMC curves and rank-one accuracy were respectively generated. Figure 11 (d) illustrated the

TABLE VI: Comparative time and storage complexity between different cross fingerprint comparison methods

Approaches	Number of parameters	Feature extraction	Matching	Template size
Minutiae-based	<10	1.682s	1.256s	292 - $d$ vector
Triplet CNN	~32.5 M	0.008s	$7.98 * 10^{-5} s$	1024 - $d$ vector
Two channel CNN	~15.7 M	0.009s	$8.68 * 10^{-5} s$	1024 - $d$ vector
Residual CNN	~10.1 M	0.012s	$7.12 * 10^{-5} s$	1024 - $d$ vector
<b>Proposed method</b>	<b>~22.6 M</b>	<b>0.014s</b>	<b><math>1.49 * 10^{-4} s</math></b>	<b>3072 - <math>d</math> vector</b>

results for cross-fingerprint comparison by separately using three sub-networks and the whole network. We also attempted to incorporate different weights to concatenate the feature representations generated from three subnets. When weight ratio of sub-net1 (ridge and minutiae map), sub-net2 (blurred core point region) and sub-net3 (core point region) is 2:3:5, we achieved the best performance and respective EER value is 7.11%. It can be observed from Figure 11 (e) and Figure 11 (f), the proposed cross comparison method also achieved outperforming results than the other methods on dataset B. This again validated the robustness and efficiency of the proposed method. The EER is 7.11% using proposed model and it is 21.60% using traditional minutiae-based matching and it is 22.10% using triplet CNN-based matching. The rank-one accuracy in Table IV is also improved from 38.9% to 58.87% by incorporating proposed approach. The proposed method on this dataset also shows a significant improvement ( $p$ -value < 0.0001) of traditional methods. Table V illustrated the statistical significance ( $p$ -value) of the performance improvement using distance-aware loss function and proposed approach. The results indicate that, by the commonly used confidence level of 95%, the improvement of distance-aware loss function is statistically significant and our method significantly outperforms other two methods [13], [19] on two databases ( $p$ -value is expressed as < 0.0001 because the computed  $z$ -statistics are too large that the corresponding  $p$ -value exceed double precision).

#### F. Time Complexity Analysis

The proposed implementation runs on an Intel i5-2500 3.3GHz CPU with NVIDIA 980Ti GPU. It takes around 4 hours to train the sub-net1 for around 20k iterations, around 5 hour to train the sub-net2 for around 40k iterations and around 4 hours to train the sub-net3 for around 30k iterations. Although training the whole networks takes several hours, during the testing phase, the deep feature representations are generated by simply forwarding operations. In addition, the matching scores are generated by simply computing the Euclidean distance between each two feature representations, which takes less than 0.10 second for feature extraction and comparison. Comparing with minutiae-based approaches, the proposed method does not need complicated feature extraction and time-consuming minutiae alignment processes and therefore offers faster fingerprint cross comparison. Table VI illustrates comparative

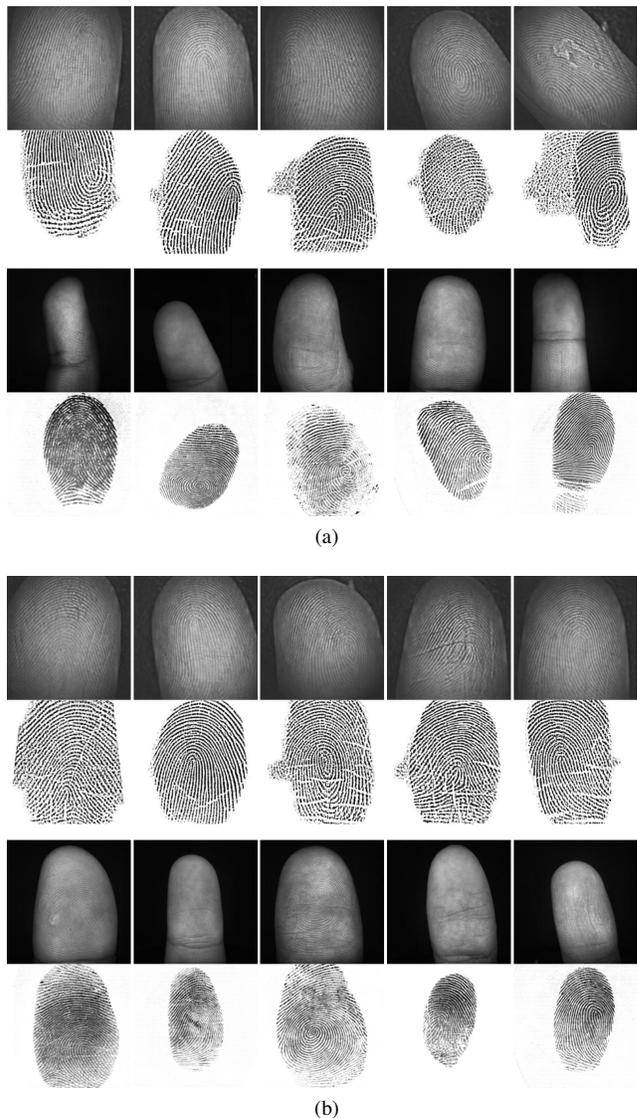


Fig. 12: (a) Falsely rejected image sample pairs from test dataset A (first two rows) and dataset B (last two rows); (b) Falsely accepted image sample pairs from test dataset A (first two rows) and dataset B (last two rows).

storage requirements for the respective templates and the computational complexity between the methods considered in this work.

#### G. Discussion

Despite significant improvement in the accuracy for contactless to contact-based fingerprint cross comparison, further improvement is required to meet the expectations for the deployment. The limitations in the matching accuracy from the proposed method can be observed from the falsely rejected genuine fingerprints and falsely accepted imposter fingerprints. Figure 12 presents such falsely rejected image samples and falsely accepted image samples from the two public databases used in this work. As can be observed from images in Figure 12 (a), most false rejects in both datasets can be attributed to three key reasons. Firstly, the quality of these contactless or

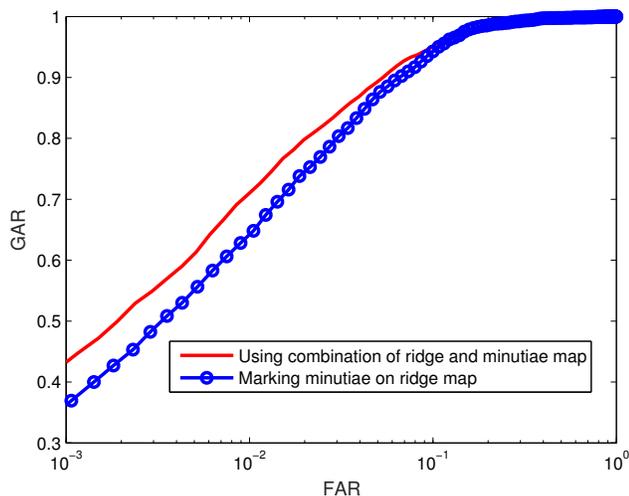


Fig. 13: ROC curves for cross comparison using combination of ridge and minutiae map, and for direct marking the minutiae on ridge map.

contact-based fingerprint images is quite poor. Secondly, contactless or contact-based fingerprints have very large rotation. Thirdly, there is very little overlap between the contactless and contact-based fingerprints due to significant differences in the image sensing and/or acquisition approaches. False acceptance of image pairs in both datasets, as can also be appreciated from sample images in Figure 12 (b), can be attributed to two key reasons. Firstly, the contactless and contact-based fingerprints illustrate similar contour or structure. Secondly, the contactless and respective falsely accepted contact-based fingerprints present high similarities in image details including similar core point region and ridge flow patterns. The limitations in accurately matching these images can be addressed by incorporating larger database, (using more challenging data for the training), and further advancing the network architecture to more accurately learn local ridge map from the fingerprints, especially under the deformation.

We also performed additional experiments to ascertain the performance improvement, from the combination of ridge and minutiae map, over the usage of marking minutiae on the ridge map as we used in [33]. Such comparative experimental results using the database B are shown in Figure 13. These experimental results indicate further improvement in the performance when combination of ridge and minutiae map is used in our approach for the contactless to contact-based cross fingerprint comparison problem. The EER value is reduced from 8.14% to 7.11% using such combination.

#### IV. CONCLUSIONS AND FURTHER WORK

In this paper, we have presented a specially designed multi-Siamese fingerprint cross comparison framework to accurately match contactless to contact-based fingerprints. In the best of our knowledge, this is the first such attempt to address challenging cross-fingerprint comparison problem using deep learning. Our framework detailed in this paper incorporates most reliable minutiae features along with the respective ridge flow maps to ensure robustness in the learning minutiae

feature correspondences. A multi-Siamese CNN with distance-aware loss function is used to generate the fingerprint feature representation vectors. In addition, the hand-craft features (fingerprint core point) is utilized to generate more robust fingerprint deep feature representation. Each three fingerprint representation vectors from three sub-nets are concatenated for more accurate cross comparison. The experimental results from two public datasets presented in section III.E illustrated that proposed method can achieve outperforming results than many other promising deep learning methods. Our experiments also indicated that the proposed method can achieve significantly improved performance over other minutiae based fingerprint cross comparison methods.

Unlike contact-based fingerprint cross-sensor comparison, the cross comparison using contact-based to contactless fingerprints is more challenging. In practice, lack of sufficient training data, i.e. contact-based and respective contactless fingerprints, in proposed framework can significantly degrade the matching accuracy. Incorporating different data augmentation strategies like scale-based data augmentation and using more promising learning strategies including different stochastic optimization algorithms like Adagrad [55] or Adam [56], are expected to improve the cross comparison performance and should be considered in the extension of this work. Deep learning based architecture which is specially designed to learn inter and intra class variations for other cross-sensor problems, e.g. contact-based to rolled fingerprints, offers significant potential to improve the accuracy and is suggested for further work. Further work is also required to develop and incorporate larger dataset, better architecture and strategy to achieve performance improvement to meet the expectations from the biometrics community for the deployments.

#### REFERENCES

- [1] D. Maltoni, D. Maio, A. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*. 2nd Ed., Springer, 2009.
- [2] D. Alonso-Fernandez, J. Bigun, J. Fierrez, H. Fronthaler, K. Kollreider, and J. Ortega-Garcia, "Fingerprint recognition," in *Guide to biometric reference systems and performance evaluation*. Springer, 2009, 51–88.
- [3] A. Kumar and Y. Zhou, "Contactless fingerprint identification using level zero features," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*. IEEE, 2011, pp. 114–119.
- [4] G. Parziale and Y. Chen, "Advanced technologies for touchless fingerprint recognition," in *Handbook of Remote Biometrics*. Springer, 2009, pp. 83–109.
- [5] P. Krishnasamy, S. Belongie, and D. Kriegman, "Wet fingerprint recognition: Challenges and opportunities," in *Biometrics (IJCB), International Joint Conference on*. IEEE, 2011, pp. 1–7.
- [6] A. Ross and R. Nadgir, "A calibration model for fingerprint sensor interoperability," in *Defense and Security Symposium*. International Society for Optics and Photonics, 2006, pp. 62 020B–62 020B–12.
- [7] A. Ross and R. Nadgir, "A thin-plate spline calibration model for fingerprint sensor interoperability," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 20, no. 8, pp. 1097–1110, 2008.
- [8] S. S. Wood and C. L. Wilson, *Studies of plain-to-rolled fingerprint matching using the NIST algorithmic test bed (ATB)*. US Department of Commerce, Technology Administration, National Institute of Standards and Technology, 2004.
- [9] Y. Chen, G. Parziale, E. Diaz-Santana, and A. K. Jain, "3d touchless fingerprints: compatibility with legacy rolled images," in *Biometric Consortium Conference. Biometrics Symposium: Special Session on Research at the*. IEEE, 2006, pp. 1–6.
- [10] W. Zhou, J. Hu, I. Petersen, S. Wang, and M. Bennamoun, "A benchmark 3d fingerprint database," in *Fuzzy Systems and Knowledge Discovery (FSKD), 11th International Conference on*. IEEE, 2014, pp. 935–940.

- [11] A. Jain, L. Hong, and R. Bolle, "On-line fingerprint verification," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 4, pp. 302–314, 1997.
- [12] C. I. Watson, M. D. Garris, E. Tabassi, C. L. Wilson, R. M. McCabe, S. Janet, and K. Ko, "User's guide to nist biometric image software (nbis)," NIST, 2007.
- [13] C. Lin and A. Kumar, "Improving cross sensor interoperability for fingerprint identification," in *International Conference on Pattern Recognition (ICPR)*, 2016, pp. 943–948.
- [14] F. Liu, D. Zhang, C. Song, and G. Lu, "Touchless multiview fingerprint acquisition and mosaicking," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 9, pp. 2492–2502, 2013.
- [15] Y.-L. Boureau, J. Ponce, and Y. LeCun, "A theoretical analysis of feature pooling in visual recognition," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 111–118.
- [16] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1026–1034.
- [17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1106–1114.
- [18] R. F. Nogueira, R. de Alencar Lotufo, and R. C. Machado, "Fingerprint liveness detection using convolutional neural networks," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 6, pp. 1206–1213, 2016.
- [19] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 815–823.
- [20] Y. Sun, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2892–2900.
- [21] J. Wang, Y. Song, T. Leung, C. Rosenberg, J. Wang, J. Philbin, B. Chen, and Y. Wu, "Learning fine-grained image similarity with deep ranking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1386–1393.
- [22] T.-Y. Lin, A. RoyChowdhury, and S. Maji, "Bilinear cnn models for fine-grained visual recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1449–1457.
- [23] A. Ross and A. Jain, "Biometric sensor interoperability: A case study in fingerprints," *Proc. BioAW, LNCS* 3086, Springer, 2004, pp. 134–145.
- [24] F. Alonso-Fernandez, R. N. Veldhuis, A. M. Bazen, J. Fierrez-Aguilar, and J. Ortega-Garcia, "Sensor interoperability and fusion in fingerprint verification: A case study using minutiae-and ridge-based matchers," in *Control, Automation, Robotics and Vision, 9th International Conference on*. IEEE, 2006, pp. 1–6.
- [25] S. K. Modi, "Analysis of fingerprint sensor interoperability on system performance," Purdue University West Lafayette, Tech. Rep., 2008.
- [26] R. D. Labati, A. Genovesi, V. Piuri, and F. Scotti, "Toward unconstrained fingerprint recognition: A fully touchless 3-d system based on two views on the move," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, no. 2, pp. 202–219, 2016.
- [27] L. Ericson and S. Shine, *Report on Evaluation of Contactless Versus Contact Fingerprint Data, Phase 2 (Version 1.1)*. Report No. 249552, I. ManTech Advanced Systems International, DOJ Office of Justice Programs, 2015. <https://www.ncjrs.gov/pdffiles1/nij/grants/249552.pdf>.
- [28] Azimuth. Inc., *Report on Evaluation of Contact versus Contactless Fingerprint Data (Final Report v2)*. Report No. 245146, I. ManTech Advanced Systems International, DOJ Office of Justice Programs, 2014. <https://www.ncjrs.gov/pdffiles1/nij/grants/245146.pdf>.
- [29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [30] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1. IEEE, 2005, pp. 539–546.
- [31] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4353–4361.
- [32] F. Wang, L. Kang, and Y. Li, "Sketch-based 3d shape retrieval using convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1875–1883.
- [33] C. Lin and A. Kumar, "Multi-Siamese Networks to Accurately Match Contactless to Contact-based Fingerprint Images," in *International Joint Conference on Biometrics (IJCB)*, 2017.
- [34] S. Reed, Z. Akata, H. Lee, and B. Schiele, "Learning deep representations of fine-grained visual descriptions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 49–58.
- [35] Y. Cui, F. Zhou, Y. Lin, and S. Belongie, "Fine-grained categorization and dataset bootstrapping using deep metric learning with humans in the loop," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1153–1162.
- [36] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Advances in neural information processing systems*, 2014, pp. 1988–1996.
- [37] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Computer vision, graphics, and image processing*, vol. 39, no. 3, pp. 355–368, 1987.
- [38] Z. Lan, S.-I. Yu, M. Lin, B. Raj, and A. G. Hauptmann, "Hand-crafted local features are convolutional neural networks," *arXiv preprint arXiv:1511.05045*, 2015.
- [39] L. Wang, Y. Qiao, and X. Tang, "Action recognition with trajectory-pooled deep-convolutional descriptors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4305–4314.
- [40] "Verifinger," <http://www.neurotechnology.com/verifinger.html>, accessed 2017.
- [41] J. Hu, J. Lu, and Y.-P. Tan, "Discriminative deep metric learning for face verification in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1875–1882.
- [42] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 675–678.
- [43] Z. Liu, Z. Li, J. Zhang, and L. Liu, "Euclidean and hamming embedding for image patch description with convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 72–78.
- [44] "Ubc patch dataset," <http://www.cs.ubc.ca/~mbrown/patchdata/patchdata.html>, accessed 2016.
- [45] "At&t face dataset," <http://www.uk.research.att.com/facedatabase.html>, accessed 2016.
- [46] A. M. Martinez, "The ar face database," *CVC technical report*, 1998.
- [47] "Feret dataset," <http://www.itl.nist.gov/iad/humanid/feret/>, accessed 2016.
- [48] Weblink for downloading *PolyU Contactless to Contact-based Fingerprint Database*, <http://www.comp.polyu.edu.hk/~csajaykr/fingerprint.html/>.
- [49] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach," *Biometrics*, pp. 837–845, 1988.
- [50] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of The 32nd International Conference on Machine Learning*, 2015, pp. 448–456.
- [51] Z. Zhao and A. Kumar, "Accurate Periocular Recognition Under Less Constrained Environment Using Semantics-Assisted Convolutional Neural Network," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 5, pp. 1017–1030, 2017.
- [52] P. Schuch, S. Schulz, and C. Busch "Deep Expectation for Estimation of Fingerprint Orientation Fields," in *International Joint Conference on Biometrics (IJCB)*, 2017.
- [53] Weblink for downloading *codes for the algorithms in this paper*, <http://www.comp.polyu.edu.hk/~csajaykr/tifs18.zip/>.
- [54] F. adegghi, C. L. Zitnick, and A. Farhadi "Visalogy: Answering visual analogy questions," in *Advances in Neural Information Processing Systems*, pp. 1882–1890, 2015.
- [55] J. Duchi, E. Hazan, and Y. Singer "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of Machine Learning Research*, vol. 12, pp. 2121–2159, 2011.
- [56] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [57] International Standard ISO/IEC 19795-1:2006. <https://www.iso.org/obp/ui/#iso:std:iso-iec:19795-1:ed-1:v1:en>.