Deep Age-Invariant Fingerprint Segmentation System

M.G. Sarwar Murshed[1, *], **Keivan Bahmani**[1], **Stephanie Schuckers**[1], **and Faraz Hussain**[1]

[1]Department of Electrical and Computer Engineering, Clarkson University, Potsdam, NY 13699, USA
[*]murshem@clarkson.edu

ABSTRACT

Fingerprint is one of the important modalities that have been used for biometric recognition applications such as border crossings, health benefits, criminal justice, electronic voting, etc. Fingerprint-based identification systems achieve higher accuracy when a slap containing multiple fingerprints of a subject is used instead of a single fingerprint. However, segmenting or auto-localizing all fingerprints in a slap image is a challenging task due to the different orientations of fingerprints, noisy backgrounds, and the smaller size of fingertip components. The presence of slap images in a real-world dataset where one or more fingerprints are rotated makes it challenging for a biometric recognition system to localize and label the fingerprints automatically. Improper fingerprint localization and finger labeling errors lead to poor matching performance. In this paper, we introduce a method to generate arbitrary angled bounding boxes using a deep learning-based algorithm that precisely localizes and labels fingerprints from both axis-aligned and over-rotated slap images. We built a fingerprint segmentation model named CRFSEG (Clarkson Rotated Fingerprint segmentation Model) by updating the previously proposed CFSEG model which was based on traditional Faster R-CNN architecture [21]. CRFSEG improves upon the Faster R-CNN algorithm with arbitrarily angled bounding boxes that allow the CRFSEG to perform better in challenging slap images. After training the CRFSEG algorithm on a new dataset containing slap images collected from both adult and children subjects, our results suggest that the CRFSEG model was invariant across different age groups and can handle over-rotated slap images successfully. In the *Combined* dataset containing both normal and rotated images of adult and children subjects, we achieved a matching accuracy of 97.17%, which outperformed state-of-the-art VeriFinger (94.25%) and NFSEG segmentation systems (80.58%). The results indicate that the deep learning-based slap segmentation system is more efficient for both children and adult slaps. Our pre-trained CRFSEG models and training codes for using our model are publicly available (https://github.com/sarwarmurshed/CRFSEG).

Introduction

Fingerprints are one of the most used modalities for biometric recognition systems. Many applications such as border-crossing, health benefits, and food distribution use slap images instead of single fingerprint images. A slap image contains multiple fingers of a person, usually four fingerprints other than the thumb of the left or right hand, or two fingers of two thumbs. Figure 1a shows a right-hand slap containing four fingerprints.

Previous work on fingerprint-based identification systems suggests higher accuracy when multiple fingers are used instead of a single finger [28]. The first step in a fingerprint identification system with multiple fingerprints is to segment the slap image into N images of individual fingerprints [5]; N is usually four for each of the slap of the left hand and right hand, and two for the thumb slap. Hand-crafted segmentation system NFSEG, created by NIST, is a popular open-source software to segment slap images [13]. In 2004 and 2008, the National Institute of Standards and Technology (NIST) organized two contests named SlapSeg04, and SlapsegII to evaluate the performance of segmentation algorithms [17]. NFSEG is the only publicly available segmentation algorithm published by NIST. Since then, this algorithm has been used as a benchmark to evaluate the performance of newly developed slap segmentation algorithms [10]. NFSEG has limitations such as poor accuracy and a high failure rate for rotated and in juvenile fingerprints [21]. NIST again conducted a contest called slap fingerprint segmentation evaluation II (SlapSegII) in 2008. The difference between the two contests is the metrics used for successful slap fingerprint segmentation .

Recently, Deep Learning (DL) based methods have made significant progress in different computer vision tasks such as object detection, classification, and segmentation [8]. Deep learning architectures such as Faster R-CNN, Mask R-CNN, and YOLO exhibit high performance in many aspects of computer vision applications [8, 20, 19]. Previous work suggests the Slap segmentation system developed using such architectures can potentially outperform NFSEG in terms of fingerprint localization and matching accuracy [21].

Fingerprint matching has become very accurate in recent years [9], but many assessments are based on already segmented fingerprints. If a fingerprint is not segmented properly or is mislabeled, it automatically leads to a matching error. Furthermore,

(a) Slap image of the right hand.

(b) Right index

(c) Right middle

(d) Right ring

(e) Right little

Figure 1. A sample slap image containing four fingerprints of the right hand is shown in Figure 1a. The separation of all fingerprints from a slap image to individual fingerprint images is called slap fingerprint segmentation. The output of segmented images is shown in Figure 1b to Figure 1e.

due to improvements in fingerprint recognition performance, performance is driven by the edge cases, such as fingers that have extreme rotation, improper orientation, and low quality.

The process of aging can impact the performance of fingerprint identification systems, as changes in skin elasticity and moisture can cause modifications to the patterns and ridges of our fingerprints. In a study conducted by Galbally et al. to investigate the effects of aging on fingerprint biometrics, they found that the accuracy of fingerprint identification systems was more negatively affected when they used fingerprints from children. Specifically, when testing the fingerprints of children aged 0 to 12 years, they reported a decrease of up to 50% in the accuracy of fingerprint identification systems. This leaves a gap in our understanding of how slap fingerprint segmentation algorithms perform in younger populations. Unfortunately, none of the previously published algorithms for slap segmentation tasks have been evaluated in a dataset containing juvenile and children subjects, as the current benchmark (SLAPSeg competition) only uses data from adult individuals.

Our approach to solving these problems involves the development of a deep-learning algorithm that can generate both axis-aligned and rotated bounding boxes for precise localization of fingerprints on slap images. To train our proposed algorithm and create a segmentation system called CRFSEG, we utilized a dataset of 11,844 axis-aligned slap images collected from both children and adult subjects that had been manually annotated. Furthermore, we generated arbitrarily angled slap images and their corresponding bounding boxes by rotating axis-aligned slap images and ground-truth bounding boxes from -90° to 90°. We constructed a large dataset named *Combined* slap dataset that consists of both axis-aligned and rotated slap images. This dataset is used to train the CRFSEG model in order to efficiently localize fingerprints even at a large angle of rotation. We created another dataset called the *Challenging* slap dataset which consists of challenging images and evaluated the performance of the CRFSEG model on that dataset. More details of our dataset are described in the slap dataset section.

Related Work

Different types of segmentation methods were used to precisely segment fingerprints and classify them. The backbone algorithms of those methods include convolutional neural network (CNN), Recurrent Adversarial Learning, Fuzzy C-Means and Genetic Algorithm, etc [21, 11, 16, 22].

A CNN-based fingerprint segmentation technique was proposed by Dai et al., where they consider fingerprint segmentation as a binary classification problem and used a CNN to find background (noise) and foreground (fingerprint) [3]. This method uses a pre-processing step to remove the background and enhance the ridge structure. After that, the image is divided into different ROIs (fingerprint patches) and fed to a CNN to predict patch categories. To reconstruct the whole fingerprint, a patch quilting method is used after receiving results from the CNN. Finally, morphological operations are used in the contracted segment to obtain a final foreground. Different pre and post-processing techniques are used in this method to get final results. This approach works well when an input image contains one fingerprint. This type of approach will likely fail in slap images because slap images not only capture the fingerprint but also captures other areas under the fingertips as shown in Figure 1a.

Serafim et al. proposed a segmentation technique that finds the regions of interest (ROI) in a fingerprint image using the CNN algorithm and performs patch-level classification of foreground (inside ROI) and background (outside ROI) [22].

Their method involves different pre-processing techniques before feeding input images to the CNN and performs different post-processing techniques to obtain the output ROIs. Pre-processing techniques involve dividing an input fingerprint image into different non-overlapping sub-blocks which they called patches. Post-processing involves block filtering, removing islands, pixel filling, and Border smoothing.

In other research, Stojanovic et al proposed an ROI segmentation method based on deep learning algorithms [23]. They used AlexNet and LeNet architectures for detecting ROI in the fingerprints. In this method, the input images are pre-processed which ensures the proper size and acceptable quality of the input image, and then divides the images into multiple non-overlapping patches. Those patches are then fed to the CNN model which generates ROI. After detecting ROIs, region mask margins and different smoothing techniques are used as post-processes methods to get the final results.

All the methods above rely on pre- and post-processing techniques. A few previous methods can not handle images with different shapes and sizes. Contrasting, our proposed fully automated method can handle images of different sizes and shapes and requires no involvement of humans throughout the entire process.

Novelties of this work

In this study, we focus on developing a deep learning-based age-invariant slap fingerprint segmentation system. We paid special attention to fingerprints collected from both children and adult subjects, because the shapes, spatial characteristics, and quality of children's fingerprints are different from adult fingerprints [4]. This paper describes the following novel contributions:

- Built two in-house large datasets named *Combined* and *Challenging* slap datasets that contain 133,611 slap fingerprint images of children and adults.
- Annotated all slap fingerprint images manually to establish a ground-truth baseline for the accuracy assessment of different fingerprint segmentation systems.
- Developed "a novel" age-invariant deep learning-based slap segmentation model (viz. CRFSEG), that can handle arbitrarily orientated fingerprints of adult and juvenile subjects.
- Evaluated the performance of state-of-the-art commercial and non-commercial fingerprint segmentation systems using both adult and children slap images and compared the performance of the CRFSEG model with other segmentation systems.
- Released a trained CRFSEG model through GitHub (`https://github.com/sarwarmurshed/CRFSEG`) for public use.

Research methods

In this section, we present in detail the method used for collecting slap image data from children and adult subjects, data annotation, data augmentation, and ground truth creation methods. Then, we describe the proposed neural network architectures. Finally, we discuss the metrics used to evaluate different slap segmentation algorithms.

Slap dataset

In this section, we discussed slap dataset generation processes including slap collection, annotation, augmentation, and ground truth creation.

Slap image collection

A clean, rich, and representative dataset is what drives complex and sophisticated deep-learning algorithms to deliver state-of-the-art performance and therefore an invaluable resource for building robust deep-learning models. Creating a balanced slap dataset containing fingerprints of both adult and children subjects is challenging since the data need to be captured using special scanners from real subjects. We collected children's slap fingerprint images from elementary and middle school under an approved IRB. Collection from children was approved by the Institutional Review Board (IRB) with parents' consent and child asset. We built our novel slap dataset by combining newly collected slap data with several adult slap datasets. This dataset contains a total of 15881 slaps (adults: 9131, children: 6750) from 339 adults and 260 children subjects. All slap images are captured at 500 PPI using the FBI-certified *Crossmatch L Scan Guardian (9000251)* fingerprint scanner [2].

A rich and representative slap image dataset should contain images that cover a wide variety of scenarios with various poses, illumination, size, brightness, and positions of fingerprints. Such datasets help to build a robust deep-learning-based fingerprint segmentation model and are compiled for a study to robustly evaluate the performance of the segmentation model under real-world conditions. After analyzing the data set, we found that many subjects, especially children, put their hands on

the scanner at different angles. This results in fingerprints that are often rotated in a slap image. Also, some subjects do not press all their fingers equally, resulting in not all fingers being equally visible in a slap image. Many times, due to the presence of sweat or dust on the hands of a subject, additional data such as lower areas of proximal phalanges which sometimes look like small fingerprints, are added to the slap image. Other common issues in a slap dataset include the presence of the "halo" effect, instances where two neighboring fingerprints touch each other, and amputated or partially captured slaps [1]. As a result, the commercial segmentation models that have not been trained with such examples fail in these cases.

Slap image annotation and augmentation

Data annotation is an important step in creating a structured and representative dataset which is a prerequisite for building a reliable image segmentation model. Annotation involves drawing a bounding box around each fingertip of a slap image and attaching a label to each fingertip. The fingerprint labels are Left-Index, Left-Middle, Left-Ring, Left-Little, Left-Thumb, Right-Index, Right-Middle, Right-Ring, Right-Little, and Right-Thumb. This annotation is used to train a deep learning algorithm to recognize an area as a distinct object or class in a slap image.

Manual annotation is the preferred way to create a ground truth dataset for evaluating the performance of the different slap fingerprint segmentation models. However, annotating thousands of images manually is time-consuming and tedious. As an alternative, we leveraged the pre-existing hand information available for all slap images in our dataset, indicating whether the images were from the right or left hand or thumb fingerprints, and utilized NFSEG, a widely used open-source slap fingerprint segmentation model developed by NIST, to segment all images. NFSEG requires hand information to execute the slap segmentation process. Slap images that NFSEG failed to segment accurately are labeled as *Difficult* images, while those that were successfully segmented were labeled as *Plain* images. However, even in the *Plain* images that are successfully segmented by NFSEG, errors may still exist, such as failure to detect all fingerprints on a slap, inaccurate bounding box positions, or incorrect rotation angles. Therefore, we manually reviewed each *Plain* image to ensure the accuracy of each bounding box and corrected any misplaced, mislabeled, or missing bounding boxes for fingerprints. To accomplish this manual inspection task, we used a GUI-based image visualization and annotating software developed using the `labelImg` [25]. Figure 2 illustrates our fingerprint annotation interface. A group of three human annotators examined each *Plain* slap visually using our annotation

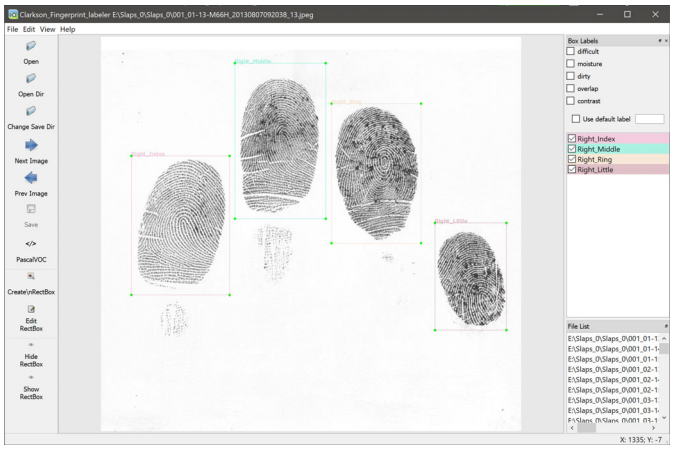


Figure 2. Graphical user interface of the slap annotating tool which is built using the Python-based *labelImg* package [25]. This tool also allows the human annotator to draw, examine and correct the position, orientation, and label of bounding boxes around fingerprints of slap images.

software, corrected all the errors, and made all the *Plain* slaps straight. At the end of this manual examination, we got two types of images, *Plain* images, and *Difficult* images. We then rotated *Plain* images from -90° to 90° to generate arbitrarily angled slap images containing rotated fingerprints.

1. **Plain images**: The slap images where NFSEG was able to correctly segment them, estimate the angle of all slap images and rotate slaps to the upright position. A human annotator cycled through slaps and corrected the position of bounding boxes around each finger that was incorrectly positioned and labeled by NFSEG, while the remaining bounding box(es) remained untouched. This process resulted in a cleaned and annotated 11844 slap images with a total of 40221 localized fingerprints (adult: 24234, children: 15987). All the *Plain* images were used for data augmentation. The distribution details of the *Plain* images can be found in Table 1.

2. **Difficult images**: A significant portion, around 25.4%, of the slap images in the dataset were not properly segmented by the NFSEG due to the inability to determine accurate rotational angles. These images were categorized as *Difficult*

Table 1. Description of each slap fingerprint dataset used for training and testing. The *Plain* dataset contains images that are successfully segmented by the state-of-the-art NIST NFSEG, while the *Challenging* dataset includes images that NFSEG failed to segment. Data augmentation techniques are applied to *Plain* annotated images to synthetically generated rotated images containing over-rotated fingerprints. The *Plain* and rotated images are mixed to build a *Combined* dataset containing both straight and over-rotated images. This *Combined* dataset is used to build the deep learning-based slap fingerprint segmentation model.

Dataset	Image types	Age groups	Total slaps	Left hand slaps	Right hand slaps	Thumb slaps
Combined	*Plain*: Successfully segmented by the NFSEG	Children	4642	1341	2075	1226
		Adult	7202	2148	2811	2243
	Augmented: Rotated images of *Plain* dataset	Children	45960	13190	20660	12110
		Adult	71770	21340	28070	22360
Challenging	*Difficult*: Failed to segment by the NFSEG	Children	2108	925	169	1014
		Adult	1929	909	229	791

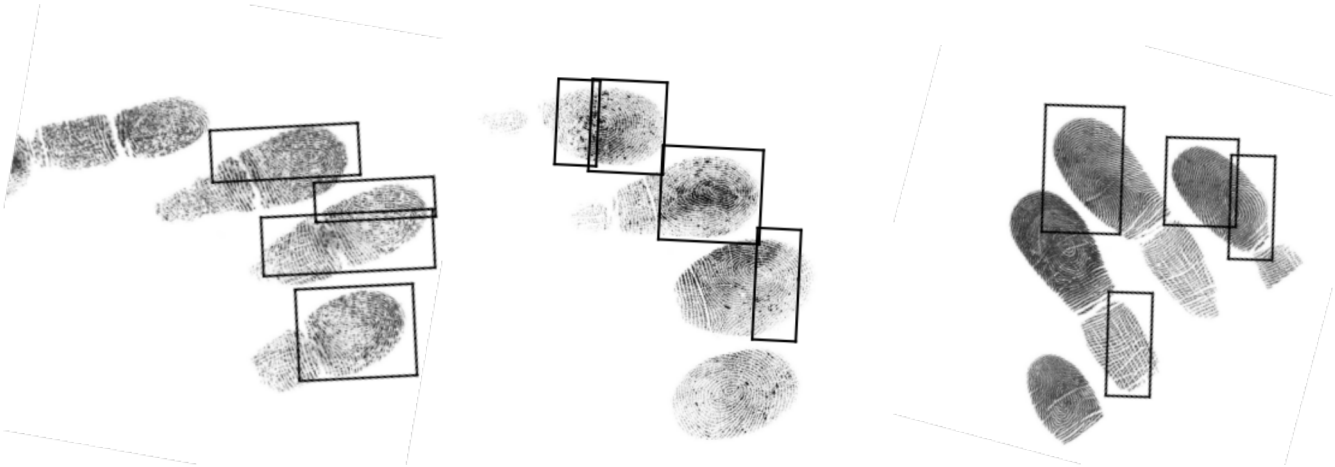


Figure 3. Examples of failed segmentation by NFSEG. This model predicted incorrect positions of bounding boxes if the slap images contained fingerprints that are rotated, i.e., not axis-aligned. These images are taken from the *Combined* dataset. We observed that the number of failures was higher in slaps collected from children subjects.

images. Upon manual examination, we found that NFSEG failed to segment an image when that image has fingers with weak and low resolution due to noise, deviation, variation, and particularly when fingerprints on a slap image are rotated more than $\pm 35^{\circ}$ from the upright position. The total number of *Difficult* slap images is 4037 containing 12370 fingerprints (adult: 6071, children: 6299). The *Difficult* slap images were not annotated manually, and therefore, they were not used for training or validating the deep learning model. Instead, these images were exclusively utilized for assessing different slap segmentation systems. Details of the *Difficult* images are shown in Table 1.

3. **Augmented images**: The NFSEG segmentation model fails to segment slaps when it fails to calculate the correct rotation angle of the slap. NFSEG predicts out-of-position bounding boxes around the fingerprints for most of these failure cases. Sometimes, predicted bounding boxes overlap each other. Example failure cases of the NFSEG model are shown in Figure 3.

   Additionally, the failure rate of the NFSEG model increases if the slap is not axis-aligned but rotated. VeriFinger segmentation software returns a failure message such as "TooFewObjects" or "BadObject" instead of returning the incorrect positions and incorrect labels of bounding boxes [18]. Our previously developed slap segmentation model CFSEG [21] generates axis-aligned straight bounding boxes with no rotational angles of the predicted bounding boxes resulting in the overlap between bounding boxes when a slap is rotated. Figure 4b shows a failure case by our previous segmentation model. To overcome the aforementioned segmentation failures in over-rotated slap images, we developed a deep learning-based robust slap segmentation model named CRFSEG model that predicts the angle of each fingerprint in a slap image individually. Training such models requires a dataset that contains both rotated and straight slap images. A minimum of four parameters are required to locate a fingerprint in a slap image using a bounding box. These parameters are (x_c, y_c, w, h), where (x_c, y_c) represents the center coordinates, w the width and h the height of a horizontal bounding box around a fingerprint as shown in Figure 4a. However, horizontal bounding boxes do not describe the fingerprint

(a) Straight bounding boxes around axis-aligned fingerprints.

(b) Straight bounding boxes around rotated fingerprints.

(c) Rotated bounding boxes around rotated fingerprints.

Figure 4. Three example slap images demonstrating various segmentation strategies. Figure 4a shows four bounding boxes surrounding four axis-aligned fingerprints with good precision. However, when the image is rotated Figure 4b, the straight bounding boxes are unable of capturing fingerprints precisely, resulting in more areas that sometimes contain regions of neighbor fingerprints. Moreover, some bounding boxes overlap with each other. Capturing more areas with less precise bounding boxes has a negative impact on fingerprint matching.Figure 4c illustrates rotated bounding boxes that enclose targeted fingerprints with high precision, which is an effective segmentation strategy for capturing fingerprints in a rotated slap image.

outline with high precision when the fingerprints are tilted or rotated relative to the horizontal axis which can happen when the full slap has not been axis aligned or individual fingers are rotated relative to each other. Figure 4b illustrates such a case, where some bounding boxes not only capture more areas than the actual fingerprint but also overlap each other. Capturing such areas reduces the preciseness of the segmentation capability of the slap fingerprint segmentation model, resulting in fingerprint-matching failure. An additional parameter is needed to accurately locate a rotated fingerprint and to reduce the unwanted area captured by the bounding boxes. Therefore, along with four original parameters, we added a new parameter θ to represent a rotated bounding box (x_c, y_c, w, h, θ), where, (x_c, y_c) is the center of the bounding box, w and h represent the width and height of the bounding box respective, and θ represents the angle (range of $[-90^\circ$ to $90^\circ]$) of the bounding box relative to the vertical axis. All parameters and rotated bounding boxes are shown in Figure 4c.

The *Plain* image set contains a total of 11844 annotated slap images that are oriented upright with respect to the y-axis. An image of such upright oriented slap is shown in Figure 4a. The data augmentation techniques are applied to all *Plain* annotated images and generated augmented images containing over-rotated fingerprints. Random rotations from -90° to 90° are applied to the *Plain* images to generate those rotated images. A total of 117730 rotated slap images are generated using augmentation techniques. The distribution details of the augmented images are shown in Table 1.

Dataset naming and ground truth

We mixed *Plain* and augmented images together and made a *Combined* dataset that contains both upright and rotated slap images. A total of 129,574 annotated images are included in this dataset, which was subsequently used as the ground truth to test different segmentation systems. We also created a separate dataset called the *Challenging* dataset, consisting of images that were labeled as *Difficult* images. The distribution details of the *Combined* and *Challenging* datasets are presented in Table 1. We used only the *Combined* dataset for training, validation, and testing of the CRFSEG model, and compared its performance with that of NFSEG and VeriFinger (a commercial slap segmentation software [18]) in the results section.

After training and validating the CRFSEG model on *Combined* dataset, we evaluated its performance on the *Challenging* dataset. We also evaluated VeriFinger on the the same *Challenging* dataset and compared its performance with the CRFSEG. The comparison results were reported in the results section.

Proposed deep learning-based network architecture

Fingerprints of adult and juvenile subjects have different shapes and sizes. To handle the size variations of fingerprints, we utilized a two-stage Faster R-CNN architecture consisting of a Feature Pyramid Network (FPN) which helps to detect objects at different scales [20]. The original Faster R-RCNN architecture is designed to detect objects with high accuracy and draw a horizontal axis-aligned bounding box around the detected object. However, this horizontal bounding box suffers several problems including capturing unnecessary areas, overlapping with close adjacent boxes, and less precise object localization.

To overcome such issues, we modified the original architecture of Faster R-CNN and included new layers to generate rotated bounding boxes around fingerprints. We named this new architecture the Clarkson Rotated Fingerprint segmentation (CRFSEG) system. This architecture has three building blocks namely, (i) Backbone network, (ii) Oriented region proposal network, and (iii) Box head.

Backbone network

The backbone network of CRFSEG is a ResNet-50 with FPN architecture that extracts feature maps from input slap images at different scales. The main components of ResNet-50 are a stem block and multiple bottleneck blocks. The stem block contains three main layers including two-dimensional convolution layers (Conv2D), rectified linear unit (ReLU), and two-dimensional max pooling layers. The Conv2D down-samples the input image twice by running the convolution with kernel size = 7, stride = 2, and max pooling with stride = 2. This type of steam block reduces the information loss when an input image is processed through the network, resulting in improved network expression ability without increasing the computational cost. On the other hand, bottleneck blocks, which are originally proposed in the ResNet paper, are used for efficient computation [7]. The bottleneck block has three main convolution layers with convolution kernel sizes of 1×1, 3×3, and 1×1 respectively. The first 1×1 is used to decrease the number of calculations required by decreasing the number of input channels. The next block with 3×3 convolution kernel is deployed to extract features for the network. The final block with 1×1 block increases the channels. The output of the backbone network is multi-scale semantic-rich feature maps which are used as input to the oriented region proposal network (O-RPN).

Oriented Region Proposal Network (O-RPN)

After obtaining the semantic-rich feature maps of the input image from the backbone network, the Oriented Region Proposal Network (O-RPN), which is a small network, uses those feature maps to propose the oriented regions of interest (ROIs) where the probability of having a targeted object is high. Proposing ROIs helps to generate final rotated bounding boxes resulting in precise bounding boxes for both axis-aligned and rotated objects. This type of regional proposal network is different from the conventional regional proposal network (RPN) used in the Faster R-CNN architecture where only axis-align regions are proposed by the RPN.

To propose the oriented regions, a 3×3 sliding window runs through the feature maps and generates a set of boxes referred to as anchors with different orientations, scales, and aspect ratios. If we have k_a different orientations, k_s different scales, and k_r different aspect rations, then $K = k_a \times k_s \times k_r$ numbers of anchor boxes are generated for each position in the feature map. Most of these anchor boxes might not have targeted objects in them. A sequence of convolution layers and two parallel output layers, which we call localizing and classifying layers, are used to separate anchors containing target objects from other anchors. The classifying layer calculates the intersection-over-union (IoU) score of the ground truth box with anchor boxes and classifies the anchors as foreground that contains targeted objects, or background that contains no objects. The localizing layer learns the offsets (x,y,w,h, θ) values for the foreground boxes. A (K$\times$5) parameterized encodings for regression offset are generated by the localizing layer and (K$\times$2) parameterized scores for region classification are generated by the classifying layer. The anchor generation strategy is illustrated in Figure 5.

For training O-RPN, seven different orientations (-π/4, -π/6, -π/12, 0, π/12, π/6, π/4), three different aspect ratios (1:1, 1:2, 2:1) and three different scales (128, 256 and 512) are used to generate anchors. Then, all the anchors are divided into three categories (i) positive anchors: anchors that have greater IoU overlap or an IoU overlap larger than 0.7 with respect to the ground-truth boxes, (ii) negative anchors: anchors that have an IoU overlap with a ground-truth box smaller than 0.3, (iii) neutral anchors: anchors that have IoU overlap between 0.3 and 0.7 with respect to ground-truth boxes. These types of anchors are removed from the anchor set and are not used during training. Note that, unlike Faster R-CNN, which uses horizontal anchor boxes during the calculation of IoU overlaps, we use oriented anchor boxes and oriented ground-truth boxes. The equations below show the loss functions that are used to train the O-RPN.

$$L_{o-rpn} = L_{cls}(p,u) + \lambda u L_{reg}(t,t^*) \tag{1}$$

Here, L_{cls} is the classification loss, p is the predicted probability over the foreground and background classes by the softmax function, u represents the class label for anchors, where u = 1 for foreground containing fingerprint and u = 0 for background; $t = (t_x, t_y, t_h, t_w, t_\theta)$ denotes the predicted regression offset value of an anchor calculated by the network, and $t^* = (t_x^*, t_y^*, t_h^*, t_w^*, t_\theta^*)$ denotes the ground truth. λ is a balancing parameter that controls the trade-off between class loss and regression loss. The regression loss is activated only if u = 1 for the foreground and there is no regression for the background.

We define the classification loss function as cross-entropy loss between the ground-truth label u and predicted probability p as:

$$L_{cls}(p,u) = -u \cdot \log(p) - (1-u) \cdot \log(1-p) \tag{2}$$

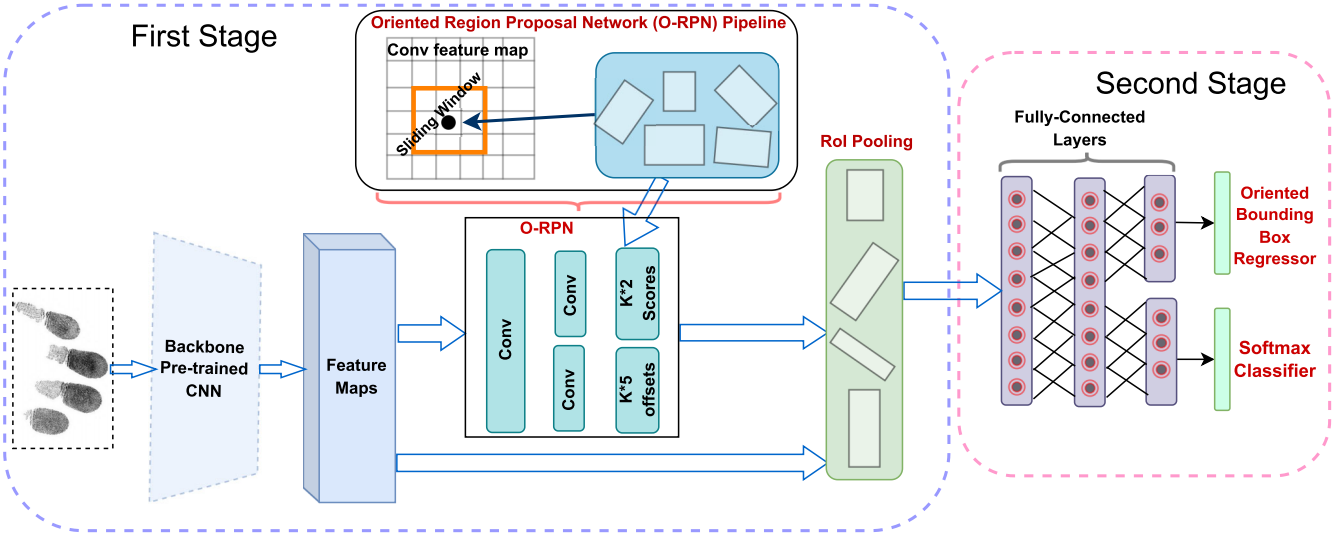


Figure 5. The complete architecture of Clarkson Rotated Fingerprint Segmentation (CRFSEG) system. An input fingerprint image is processed by a backbone CNN model that is pre-trained on the ImageNet dataset and generates feature maps. Then, O-RPN operates on all levels of the feature maps and generates oriented anchors. The RoI Pooling layer generates fixed-length feature vectors by selecting spatial features from the output of the backbone network and O-RPN. These fixed-length feature vectors are then passed through a sequence of fully connected layers. The output of fully-connected layers is fed to two parallel branches. One branch is named Oriented Bounding Box Regressor which contains regressors for bounding box regression and the other is named Softmax Classifier which contains softmax layers for multiclass classification.

The tuple t and t^* are calculated as follows:

$$t_x = (x - x_a)/w_a, t_y = (y - y_a)/h_a, \qquad t_w = \log(w/w_a), t_h = \log(h/h_a), \qquad t_\theta = \theta - \theta_a \tag{3}$$

$$t_x^* = (x^* - x_a)/w_a, t_y^* = (y^* - y_a)/h_a, \qquad t_w^* = \log(w^*/w_a), t_h^* = \log(h^*/h_a), \qquad t_\theta^* = \theta^* - \theta_a \tag{4}$$

Where, where x, x_a and x^* denote the predicted box, anchor and ground truth box, respectively; the same is for y, h, w and θ. The smooth-L1 loss is adopted for the bounding box regression as follows:

$$L_{reg}(t, t*) = \sum_{i \in x,y,w,h,\theta} u.\text{smooth}_{L1}(t_i^* - t_i) \tag{5}$$

$$\text{smooth}_{L1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases} \tag{6}$$

Box head

By default, the O-RPN network generates 1000 proposal boxes and objectness logits. An ROI pooling layer is used to project proposal boxes into feature space and the output of this layer is reshaped and fed to the fully connected (FC) layers. Then, the RoI vector is generated by the FC layers and passed through a predictor containing two branches named rotated bounding box regressor and classifier. Finally, the classification layer of the model predicts the object class, and the regressor layer regresses bounding box values.

Evaluation metrics

We used Mean Absolute Error (MAE), fingerprint angle prediction error, fingerprint labeling accuracy, and matching score to evaluate and compare the performance of the CRFSEG model with the NIST fingerprint segmentation system, NFSEG, and a commercially used fingerprint segmentation software named Neurotechnology VeriFinger [18].

1. Mean Absolute Error (MAE): The MAE measures the capabilities of slap segmentation systems to correctly segment fingerprints within a certain geometric tolerance of human-annotated ground truth data. The minimum Geometric

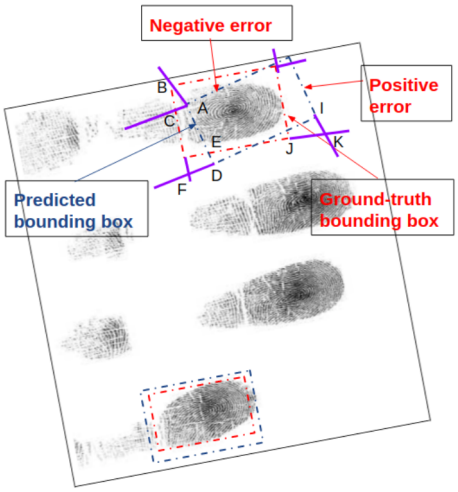


Figure 6. Example for calculating a positive and a negative error between the predicted and ground truth bounding box using Euclidean distance. To calculate the pixel error of each side of the predicted bounding boxes, we first calculate the distance of the endpoints of a side from the corresponding endpoints of the annotated ground-truth bounding box in terms of pixels. For instance, during calculating the pixel error of the side AD, we draw two perpendicular lines AC and DF with the line AD. The perpendicular line AC intersects the corresponding ground-truth line at point C. Another perpendicular line DF intersects the extension of the corresponding ground-truth line at point F. Finally, the Euclidean distances from point A to point C and point D to point F are calculated. The average of these two Euclidean distances is the pixel error for the side AD. We apply the same approaches to calculate all four sides of a bounding box. Finally, Equation 7 is applied separately on pixel errors of each size to calculate the MAE of that side.

Tolerance Limit (GTL) allowed in NIST Slapseg-II is -32 pixels for the left and right sides, -64 pixels for the top, and bottom sides [27]. Slap fingerprint segmentation models have to find the balance between over-segmentation (small bounding box covering less area than the ground truth fingerprint area) vs. under-segmentation (over-extending the predicted bounding box covering more area than the ground truth fingerprint area). Over-segmentation can lead to capturing the ridge-valley structure of close adjacent fingerprints during segmentation. This extra noise can potentially degrade the matching performance. On the other hand, under-segmentation can result in the loss of valuable parts of the fingerprints leading to the degradation of matching performance. MAE is a metric that is used to determine over-segmentation or under-segmentation by the model.

To measure the MAE of a detected bounding box, we calculated the distance of each side of a detected bounding box from the corresponding side of the annotated ground-truth bounding box in terms of pixels. [21]. A successful segmentation refers to finding a bounding box around a fingerprint within a certain geometric tolerance of the human-annotated ground truth bounding box.

In our previous study, we measured the MAE for the bounding boxes that are axis-aligned (straight from top to bottom) [21]. However, in this study, we need to measure MAE for rotated bounding boxes. For measuring the exact pixel loss, we separately measure the average Euclidean distance for all four sides of the detected bounding box. To measure the Euclidean distance of any side of a detected bounding box, first, we draw two perpendicular lines from two distance points of a side and find two intersect points of those perpendicular lines and the corresponding side of the ground-truth bounding box. If necessary we extend the length of the ground-truth side so that it intersects with the line drawn perpendicular to the detected side. Figure 6 illustrates more detail about calculating the MAE for a fingerprint. If any side of a detected bounding box captures more information than the side of the ground-truth bounding box, we consider it as a positive error. An example of a positive error is shown on the right side of the rightmost fingerprint in Figure 6. If any side of a detected bounding box captures less information than the ground-truth bounding box, we consider it as a negative error. An example of the negative error is shown in the top side of the left-most fingerprint in Figure 6.

Finally, the MAE for each side is calculated using Equation 7 separately.

$$\text{MAE} = \frac{1}{N}\sum_{i=0}^{N} |X\ error_i| \tag{7}$$

Where N is the total number of fingerprints in the test dataset. X represents the Euclidean distance error of any side such

as left, right, top, or bottom of the bounding box.

2. Error in Angle Prediction (EAP): To evaluate the capability of different fingerprint segmentation systems to predict the orientation of fingerprints, we calculate the deviation of the predicted angle compared to the ground truth using Equation 8.

$$\text{EAP} = \frac{1}{N}\sum_{i=0}^{N} |\theta - \theta^*| \tag{8}$$

Where N is the total number of fingerprints in the test dataset. θ is a ground truth angle and θ^* is a predicted angle by a fingerprint segmentation model.

3. Fingerprint classification accuracy: Hamming loss is used to evaluate the performance of multi-class classifiers [24]. Hamming loss measures the number of positions in which the corresponding symbols are different between two equal-length sets: a set of the ground-truth label and a set of the predicted label. Equation 9 is used to calculate the Hamming loss [24].

$$\text{Hamming loss} = \frac{1}{N}\sum_{i=1}^{N} \frac{|Y_i \Delta Z_i|}{|L|} \tag{9}$$

where N is the total number of samples in the dataset, and L is the number of labels. Z_i is the predicted value for the i-th label of a given sample, and Y_i is the corresponding ground true value. Δ stands for the symmetric difference between two sets of predicted and ground true value.

The accuracy of a multi-class classifier is related to Hamming loss [6, 15], which can be calculated using Equation 10.

$$\text{Accuracy} = 1 - \text{Hamming loss} \tag{10}$$

4. Fingerprint Matching: This is one of the most important aspects of slap segmentation systems. Fingerprint matching is performed to measure the capabilities of slap segmentation algorithms to correctly segment fingerprints within a certain tolerance. We use the true accept rate (TAR) and false accept rate (FAR) to report fingerprint matching. TAR is the percentage of instances a biometric recognition system verifies an authorized person correctly, which is calculated using Equation 11:

$$TAR = \frac{\text{Correct accepted fingerprints}}{\text{Total number of mated matching attempts}} \times 100\% \tag{11}$$

FAR is the percentage of instances that a biometric recognition system verifies an unauthorized user, calculated using Equation 12.

$$FAR = \frac{\text{Wrongly accepted fingerprints}}{\text{Total number of non-mated matching attempts}} \times 100\% \tag{12}$$

Experiments and results

In this section, we provide details of the experiments conducted to measure the current capabilities of different slap segmentation algorithms and demonstrate the effectiveness of our newly developed CRFSEG model. We describe the dataset used in our experiments and the training process of CRFSEG. We present the comparisons with existing segmentation algorithms and demonstrate that CRFSEG outperformed state-of-the-art segmentation systems in terms of segmentation, finger labeling, as well as matching accuracy.

Distribution of training dataset

We use the split ratio of 80:10:10 of the *Combined* dataset, where 80% slap images are used for training the deep learning model, 10% slap images are used for validating the model, and the rest of the 10% images are used for testing the model. We used the 10-fold cross-validation technique to build the model and evaluate its performance. All evaluation results are reported in the results section.

After building a robust segmentation model (CRFSEG), we used the separate *Challenging* dataset to test the performance of this deep-learning-based model. We also evaluate the performance of the VeriFinger slap segmentation software using this *Challenging* dataset. We observed that the VeriFinger slap segmentation software struggled to segment slap images in the *Challenging* dataset. On the contrary, our newly developed CRFSEG model performed better on that dataset and outperformed the VeriFinger segmentation system by achieving a 10% higher matching score.

Training

CRFSEG model was developed based on the Detectron2 software system developed by Facebook AI Research (FAIR) [29]. Detectron2 supports different state-of-the-art deep learning-based object detection algorithms such as Faster R-CNN, RetinaNet, etc. We adopted the Faster R-CNN algorithm and modified the code of Detectron2 software to implement the oriented regional proposal network (ORPN), to add new layers to the regressor in order to find the offset of the rotated bounding boxes and to incorporate all our requirements to segment fingerprints from a slap image accurately. The Faster R-CNN model used in our experiment was pre-trained on the MS-COCO dataset where the number of output classes is 81. We changed the output layers and reduced the class number from 81 to 10 in order to classify ten fingerprints from two hands. After that, the model was slowly fine-tuned using our novel slap image dataset.

We used the end-to-end training strategy, where loss values were obtained by comparing the predicted result with the true result. Training was done on a Linux-operated desktop machine with 20 cores Intel(R) Xeon(R) E5-2690 v2 @ 3.00GHz CPU, 64 GB RAM, and with a NVIDIA GeForce 1080 Ti 12-GB GPU. A total of 28000 iterations was used to train the fingerprint segmentation model with learning rates starting from 10^{-3}, and are multiplied by 0.1 after 4000, 8000, 12000, 18000, and 25000 iterations. Weight decays are 0.0005, and momentums are 0.9. All experiments used multi-scale training, hence we did not need to scale up or down any input before feeding it to the neural network.

Results

Mean Absolute Error (MAE) for segmented bounding boxes

MAE measures the preciseness of bounding boxes around fingerprints generated by slap segmentation algorithms. We used Equation 7 to calculate the MAE of different segmentation algorithms. Table 2 shows the MAE and its standard deviation for NFSEG, VeriFinger, and CRFSEG models calculated on the *Combined* slap dataset. The MAEs of the CRFSEG model in prediction bounding boxes on both adult and children subjects are significantly lower, which indicates less pixel error and more precise bounding boxes, compared to NFSEG and VeriFinger segmentation software. Additionally, our results confirm that compared to adults, NFSEG achieves lower performance for the children subjects in *Combined* dataset. Furthermore, NFSEG struggles to localize the bounding boxes when slap images are not axis-aligned but rotated. Compared to NFSEG, VeriFinger performs better, However, CRFSEG outperformed VeriFinger by reducing the MAE by 13.44, 12.21, 15.11, and 10.12 pixels in the left, right, top, and bottom sides, respectively, of detected bounding boxes on the *Combined* dataset.

We used histograms to show, analyze and evaluate the MAE results. To generate these histograms, we used the results obtained by subtracting the coordinate position of one side of all ground-truth bounding boxes from the coordinate position of the corresponding side of the detected bounding boxes. Figure 7 illustrates the histograms of MAE for all sides of bounding boxes generated by three segmentation models. Blue, orange, and green histograms represent the MAE of NFSEG, VeriFinger, and CRFSEG respectively. The first and second columns of Figure 7 show the results when we calculated MAE using children and adult subjects separately, and the third column shows results when we calculate the MAE using our entire *Combined* dataset. This separation helps us to analyze the results on children and adult subjects separately and make the model robust against age variants.

We observed that while VeriFinger and CRFSEG maintain performance on *Combined* slap datasets, NFSEG is particularly susceptible to this type of MAE error, especially when slap images are over-rotated or from children subjects. The long blue tail of the error in histograms of the NFSEG model confirms that NFSEG over-segmented many samples beyond the Minimum Tolerance Limit (MTL) of 64 pixels suggested by Slapseg-II [27]. Figure 8 depicts an example of this problem, where the red bounding boxes indicate the ground truth and the green bounding boxes present the bounding box predicted by NFSEG. NFSEG failed to correctly rotate some slap images causing the identified bounding boxes to encompass additional areas beyond the intended fingerprints. This reduces the matching performance of fingerprints segmented by the NFSEG segmentation system.

Error in angle prediction (EAP)

The EAP is the difference between the angle predicted by the segmentation algorithms and the ground-truth angle of a slap image. EAP is directly related to the preciseness of bounding boxes detected by the algorithms. Less deviation between predicated angles and ground-truth angles produces more accurate segmentation results. Equation 8 is used to calculate EAP. Table 3 shows the EAP and its standard deviation which is calculated by using the predicted angle output of different segmentation algorithms on the *Combined* dataset. The EAP values of CRFSEG are lower than the EAP values of NFSEG and VeriFinger, indicating less error during angle prediction, resulting in better performance.

Figure 9 shows the histograms of the error in angle prediction of three segmentation algorithms. The standard deviation and the area of the CRFSEG histogram is significantly smaller than the NFSEG and VeriFinger histograms, which indicates better performance of the CRFSEG model during angle prediction. In Figure 10, the boxplot graphs are used to visually display the minimum, the lower quartile, the median, the upper quartile, and the maximum quartile of the EAP values produced by NFSEG, VeriFinger, and CRFSEG. The line that is drawn across the box represents the median value of EAP. The boxplot graph shows that the absolute median value of EAP produced by the CRFSEG and VeriFinger models is lower than the absolute median

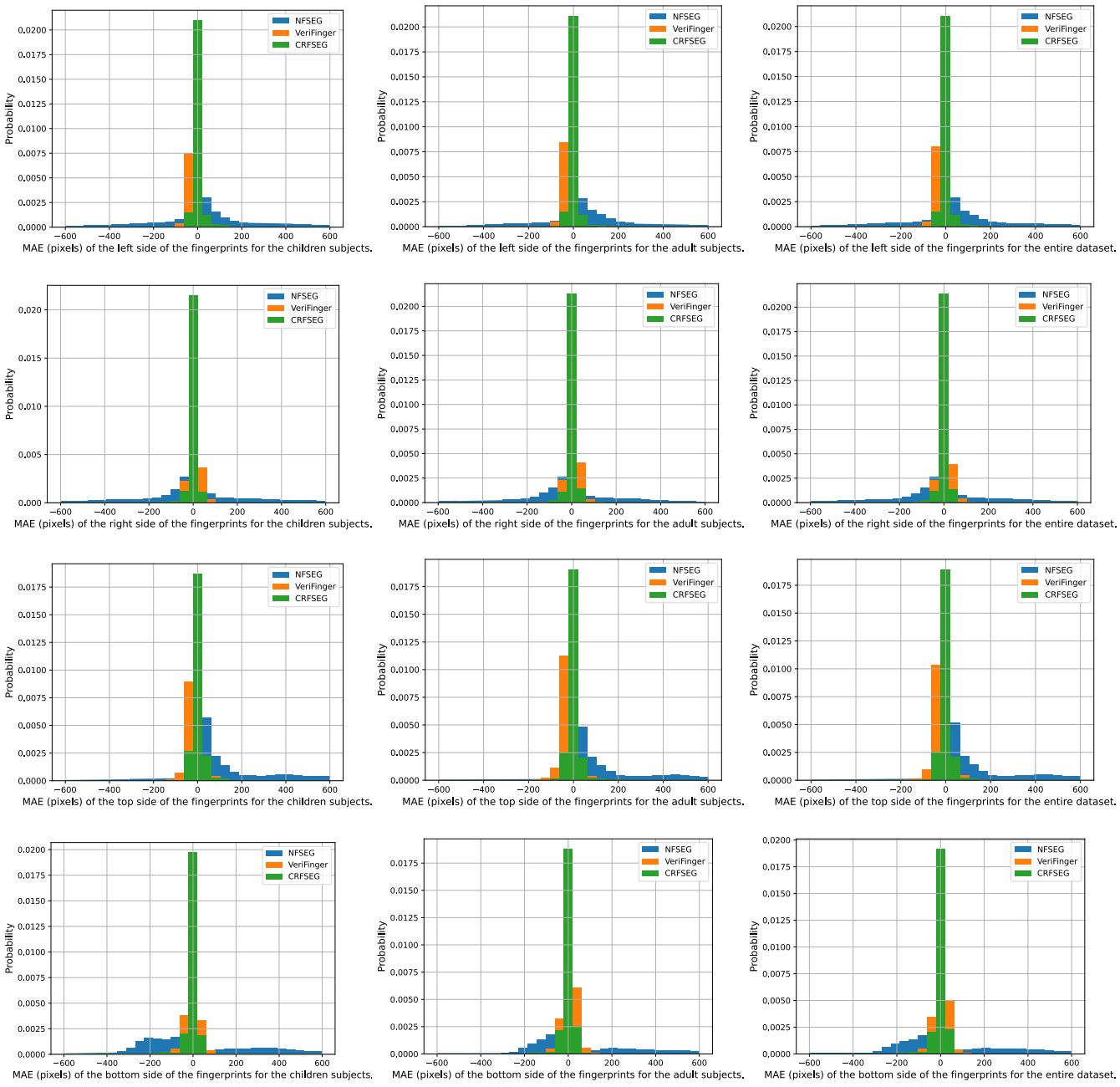


Figure 7. The histograms of the MAE generated from the output of different slap segmentation algorithms using Equation 7. Blue, orange, and green colors represent the MAE of NFSEG, VeriFinger, and CRFSEG, respectively, on the *Combined* dataset. To evaluate segmentation algorithms on different age groups, we report MAE in children and adult groups separately. The four images in the first column depict the MAEs of each side of the bounding box generated using slap images of child subjects, the second column depicts MAEs generated using adult slaps, and the third column shows MAEs generated using the entire *Combined* dataset (containing both adult and child subjects). The MAE values for the sides of the bounding boxes are presented in four rows with the first row representing the left side, the second row representing the right side, the third row representing the top side, and the fourth row representing the bottom side. The long tails of the blue histograms indicate poor performance of NFSEG. WVeriFinger performs better than NFSEG on the slaps of children and adult subjects. However, CRFSEG outperformed both NFSEG and VeriFinger in terms of MAE.

value of EAP produced by NFSEG. Hence, it is experimentally shown that the NFSEG gives lower accuracy in estimating the slap angles as compared to the VeriFinger and CRFSEG.

Table 2. Mean Absolute Error (MAE) and its standard deviation for NFSEG, VeriFinger, CRFSEG on our *Combined* slap dataset. It is calculated by taking the average of the absolute differences between each side of the detected bounding box and the corresponding side of the ground-truth bounding box in terms of pixels. The smaller MAE indicates better performance.

		Model		
		NFSEG	VeriFinger	CRFSEG
Dataset	Side	MAE (Std.)	MAE (Std.)	MAE (Std.)
Children	Left	184.74 (274.35)	28.63 (30.53)	**17.58 (20.55)**
	Right	182.08 (272.16)	27.27 (31.34)	**16.54 (17.31)**
	Top	240.22 (344.53)	30.16 (35.60)	**18.06 (24.89)**
	Bottom	260.73 (382.92)	25.87 (35.64)	**17.19 (26.56)**
Adult	Left	145.29 (215.80)	34.11 (26.81)	**19.08 (20.20)**
	Right	144.13 (215.97)	31.79 (27.11)	**18.58 (18.56)**
	Top	193.61 (301.27)	35.59 (36.74)	**18.47 (26.36)**
	Bottom	184.24 (306.34)	30.05 (36.73)	**18.96 (28.24)**
Entire	Left	161.13 (241.05)	31.92 (28.37)	**18.48 (20.35)**
	Right	159.36 (240.18)	29.98 (28.88)	**17.77 (18.08)**
	Top	212.32 (319.78)	33.42 (36.36)	**18.31 (25.78)**
	Bottom	214.95 (339.21)	28.37 (36.50)	**18.25 (27.59)**

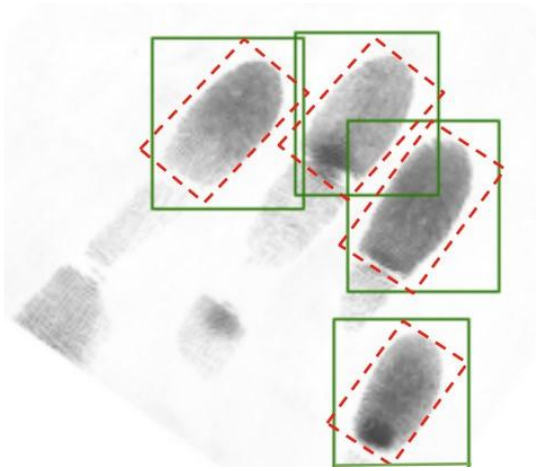


Figure 8. Green rectangles show the detected bounding boxes by NFSEG and red rectangles show the ground-truth bounding box. NFSEG encountered issues with correctly rotating certain slap images, causing the bounding boxes to overlap and encompass extraneous regions beyond the intended fingerprints, which subsequently leads to lower fingerprint-matching performance.

Label prediction accuracy

In order to calculate the fingerprint label prediction accuracy of NFSEG, VeriFinger, and CRFSEG models, we use the 1-Hamming Loss metric, a common measure for multi-label classifiers [14]. We individually study and analyze how the prediction performance varies with age group by separating children and adult subjects on our *Combined* dataset. Table 4 shows the overall label prediction accuracy of NFSEG, VeriFinger, and CRFSEG.

CRFSEG achieves a label accuracy of 98.90% on our *Combined* dataset which is slightly lower than the accuracy (99.11%) of NFSEG. However, it is important to mention that, the NFSEG and VeriFinger need hand information such as left hand, right hand, or thumb to segment a slap image. Without this hand information, the accuracy of NFSEG is very poor (around 15%). On the contrary, the CRFSEG model segment slap images without any hand information. This characteristic of CRFSEG makes this model more suitable to use in situations where additional information such as hand information is not available. It is worth mentioning that someone can use NFSEG and VeriFinger for labeling slap images if the hand information is available.

Furthermore, we calculate the label prediction accuracy for the adult and children's fingerprints separately. Our results show that the label prediction accuracy of CRFSEG is consistently good for adult and children subjects. We also observe that the label prediction accuracy of NFSEG and VeriFinger is also good on adult and children slap images. However, this is because we provide hand information to NFSEG and VeriFinger during label-predicting experiments.

Table 3. The Error in Angle Prediction (EAP) is obtained by finding the difference between the angle predicted by the segmentation algorithms and the ground-truth angle of a slap image. This table presents the average EAP and its standard deviation for three algorithms, NFSEG, VeriFinger, and CRFSEG, calculated on the *Combined* slap dataset. The smaller EAP and lower standard deviation are the indicators of better segmentation results.

	Model		
	NFSEG	VeriFinger	CRFSEG
Dataset	EAP (Std.)	EAP (Std.)	EAP (Std.)
Children	33.18° (47.85)	9.88° (14.14)	**6.66**° (9.81)
Adult	29.3° (43.83)	7.65° (11.13)	**5.27**° (7.34)
Entire	31.62° (46.56)	8.61° (12.23)	**5.86**° (8.05)

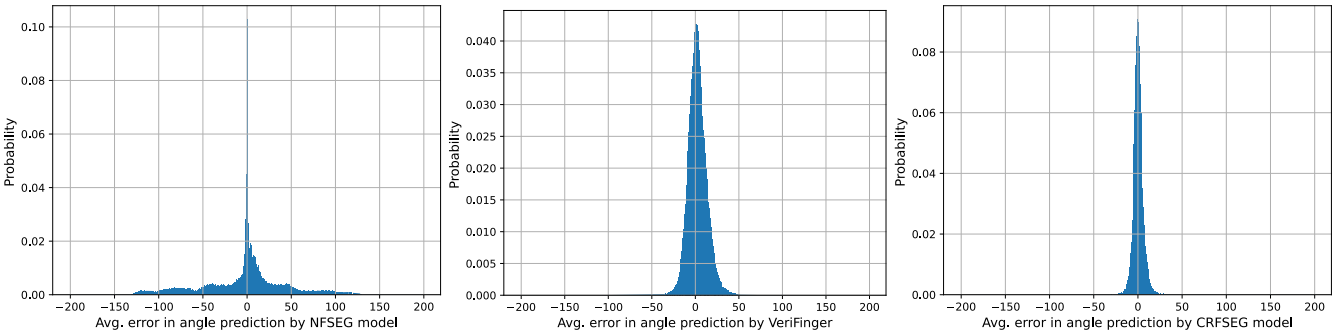


Figure 9. The histograms of the error in slap angle prediction by different segmentation models on the *Combined* dataset. This error is calculated by subtracting the angles predicted by the models from the ground-truth angles of the slap images. Low standard deviation values indicate better results.

Fingerprint matching

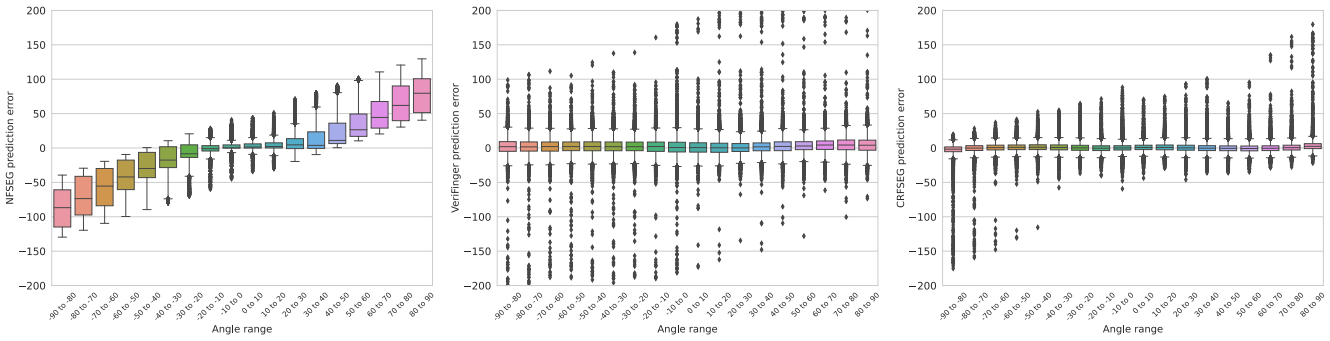
The main goal of a fingerprint segmentation algorithm is to improve the matching performance. In our experiments, we used the VeriFinger fingerprint matcher (version 10, compliant with NIST MINEX [26]) to measure the matching accuracy and quantitatively evaluate the contributions of the segmentation algorithms on the overall matching performance. We then compared the matching accuracy of CRFSEG with the state-of-the-art NFSEG and VeriFinger segmentation software.

In order to calculate the matching accuracy, four sets of segmented fingerprint images are generated from slap images of the *Combined* dataset. The bounding box data generated by human annotators (ground truth), NFSEG, Verifigner, and CRFSEG are used to generate those sets of segment fingerprint images. For every fingerprint in a segmented fingerprint set, we evaluated all the mated comparisons for our genuine distribution (3,188,232), while randomly selecting 20 non-mated fingerprints to construct an imposter distribution (32,281,840 imposter comparisons). A total of 35,470,072 comparisons using all ten fingers are performed in our experiment to estimate the matching accuracy. To evaluate the performance of different segmentation models on child and adult fingerprints, we calculated fingerprint-matching accuracy on child and adult subjects separately.

Table 5 lists the True Positive Rate (TPR) at False Positive Rate (FPR) of 0.001 for fingerprints segmented using bounding box information of ground-truth, NFSEG, VeriFinger, and CRFSEG. Ground truth achieves higher accuracy on both adult and children subjects. Among the segmentation algorithms, CRFSEG achieves better results compared to NFSEG and VeriFinger in terms of matching accuracy. Figure 11, Figure 12, and Figure 13 are three Receiver Operating Characteristics (ROC) curves that show the matching scores of different models along with ground-truth on children, adult, and entire slap dataset. The ROC curve is a graphical plot that represents the tradeoff between the true positive rate (TPR) and the false positive rate (FPR) at various threshold values. Our results indicate that for both adult and children subjects, the fingerprints segmented with CRFSEG provide higher accuracy among different segmentation algorithms across various threshold values.

Label prediction accuracy on Challenging dataset

As we mentioned before, we created a dataset of slap images that NFSEG failed to segment, which we named the *Challenging* dataset. We have finger information for each slap of that dataset but no bounding boxes around each fingerprint. Using finger information we calculated the label prediction accuracy of VeriFinger and CRFSEG on the *Challenging* dataset. Table 6 shows the results of label prediction accuracy, where CRFSEG outperformed VeriFinger by 9% in terms of identifying fingers on the slap images.



(a) EAP of NFSEG. **(b)** EAP of VeriFinger. **(c)** EAP of CRFSEG.

Figure 10. Boxplots of the error in angle prediction (EAP) of different segmentation algorithms on the *Combined* dataset. The boxplot statistical analysis of the mean with $\pm 10^{\circ}$ (standard errors) for the EAP values of VeriFinger and CRFSEG indicates that these algorithms are invariant to slap rotation. On the contrary, the mean of NFSEG output is high and not consistent with the rotation of the slap images which indicates the low performance of NFSEG when slap images are overly rotated.

Table 4. Label prediction accuracy of NFSEG, VeriFinger, and CRFSEG on the *Combined* dataset. We also calculate label prediction accuracy separately for adult and child subjects which helps us evaluate the performance of different slap segmentation models across different age groups. Although NFSEG had slightly better label accuracy than CRFSEG, but this was due to the use of hand information (right, left, or thumb) during segmentation, without which its performance drops significantly to around 15%. VeriFinger also requires hand information to segment a slap image, while CRFSEG can perform segmentation without the need for any additional hand information.

Model	Input	Dataset	Accuracy (%)
NFSEG	Slap & Hand Information	Children	98.26
		Adult	**99.95**
		Entire	**99.11**
VeriFinger	Slap & Hand Information	Children	90.34
		Adult	91.03
		Entire	90.69
CRFSEG	Only Slap	Children	99.57
		Adult	98.48
		Entire	98.90

Fingerprint matching results for Challenging dataset

As we do not have ground-truth bounding box information for the *Challenging* dataset, we were not able to calculate the MAE or EAP of the different segmentation algorithms. However, we are able to calculate matching accuracy using the finger label information of the *Challenging* dataset shown in Table 7.

Out of 12370 fingerprints, VeriFinger segmented 10142 (82%) fingerprints successfully, much less than the number of fingerprints segmented by CRFSEG which was 11338 (91%). However, the VeriFinger matcher failed to process all the images segmented by the VeriFinger segmentation software and CRFSEG model. The reasons behind this failure are bad image quality, lack of fingerprint area, and fewer minutiae found on the segmented fingerprints. The number of failure images during matching was 2403 (21%) for the VeriFinger segmentation software. On the other hand, the number of failure images during matching was 1485 (13%) for the CRFSEG segmenter. This type of failure is considered a False Reject. The VeriFinger segmentation software had 2403 failure images during matching, which is 21% of the total segmented fingerprint images of the *Challenging* dataset, while the CRFSEG had 1485 failure images or 13% of the total segmented fingerprint images. Figure 14 shows ROC for VeriFinger, CRFSEG on the entire *Challenging* image dataset.

Discussion

This paper presents a highly accurate slap segmentation system constructed with a deep learning algorithm, which is capable of handling challenges such as rotation and other types of noise commonly found in slap images. The convolutional networks used in our system are trained using the end-to-end approach to achieve higher performance and are more generalized across different orientations of slap images. The advantage of our system is that once the entire segmentation system is built, it can

Table 5. True Positive Rate (TPR) at False Positive Rate (FPR) of 0.001 for fingerprints segmented using ground-truth, NFSEG, VeriFinger, and CFSEG on *Combined* dataset. To evaluate model performance across different age groups, we calculate matching accuracies separately for adult and children subjects. The results indicate that CRFSEG performed close to the ground-truth level and outperformed VeriFinger and NFSEG in terms of fingerprint matching.

		Model		
	Ground-truth	NFSEG	VeriFinger	CRFSEG
Dataset	Accuracy	Accuracy	Accuracy	Accuracy
Children	**97.43**%	78.39%	93.26%	96.47%
Adult	**98.75**%	82.31%	94.43%	97.30%
Entire	**98.34**%	80.58%	94.25%	97.17%

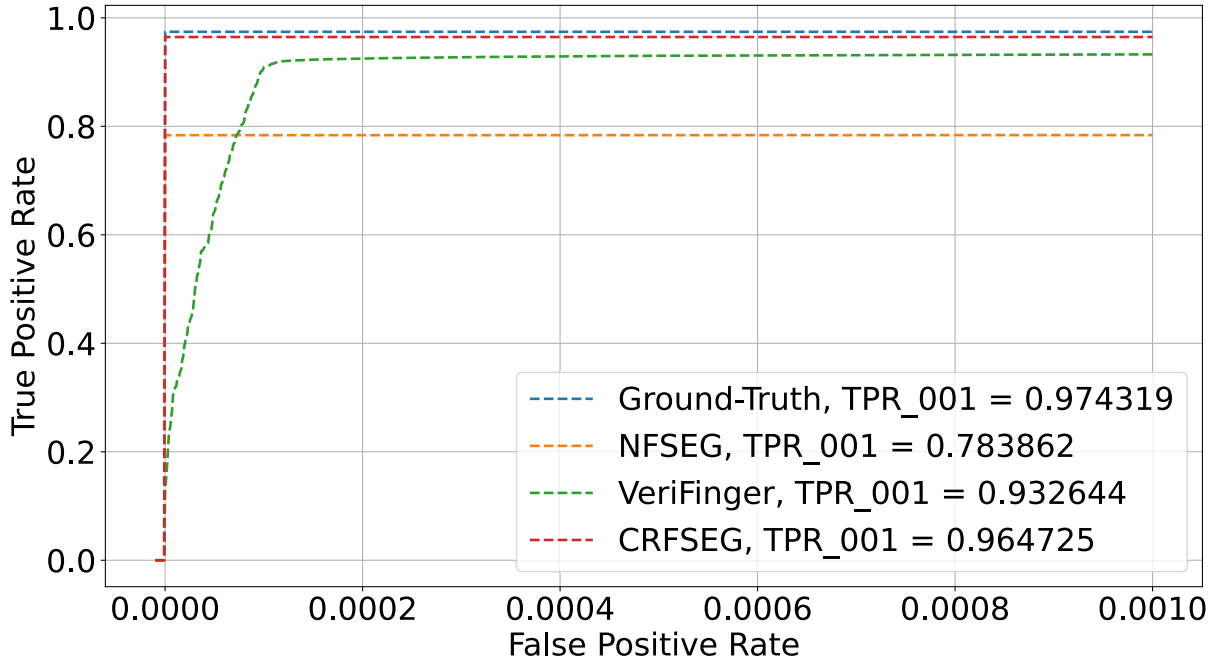


Figure 11. The Receiver Operating Characteristics (ROC) for the fingerprint matching performance of Ground Truth, NFSEG, VeriFinger, and CRFSEG on children subjects in the *Combined* dataset.

be fine-tuned using additional datasets with fewer data points to facilitate the diversity of newer datasets. To the best of our knowledge, most commercial slap segmentation systems cannot easily be fine-tuned on different slap datasets.

In the slap dataset, we observed the presence of rotated slap images that posed a challenge for NFSEG to accurately calculate the rotational angle, resulting in less precise segmentation. This is a limiting factor to improving fingerprint-matching performance. To address this issue, we introduced augmented rotated images to our testing dataset and evaluated angle prediction performance. Compared to commercially available slap segmentation methods like VeriFinger, our approach achieves a lower error of 5.86° in slap angle prediction, lower mean absolute error, and the highest fingerprint matching accuracy of 97.17%. If we compare this matching accuracy to the human-annotated ground truth matching accuracy, our method achieves a very high score, indicating it performs at a human level.

We observed that the label prediction accuracy of our CRFSEG is 0.21% less than NFSEG on the *Combined* dataset. However, NFSEG and VeriFinger systems require information about the hand, such as whether the slap image is of the left, right, or thumb, in order to accurately predict the labels of fingerprints within a slap image. On the contrary, our model predicts labels without any additional hand information. This property is particularly important when dealing with millions of fingerprints without additional information or fingerprints from a crime scene or any other source where visually it is not clear whether it is from the right hand or from the left hand. The ability of CRFSEG to label unknown fingers automatically, through its auto-labeling feature, results in a huge reduction in the time necessary for fingerprint matching.

An important point to consider in our novel segmentation model is that it was trained with fingerprints of different age groups. The main advantage of this is that it is robust toward different characteristics of fingerprints of different age groups. Indeed, our experimental results showed that our model is invariant to adult and children datasets. Moreover, we particularly focus on predicting arbitrarily oriented bounding boxes to increase the preciseness of segmented fingerprints. To achieve

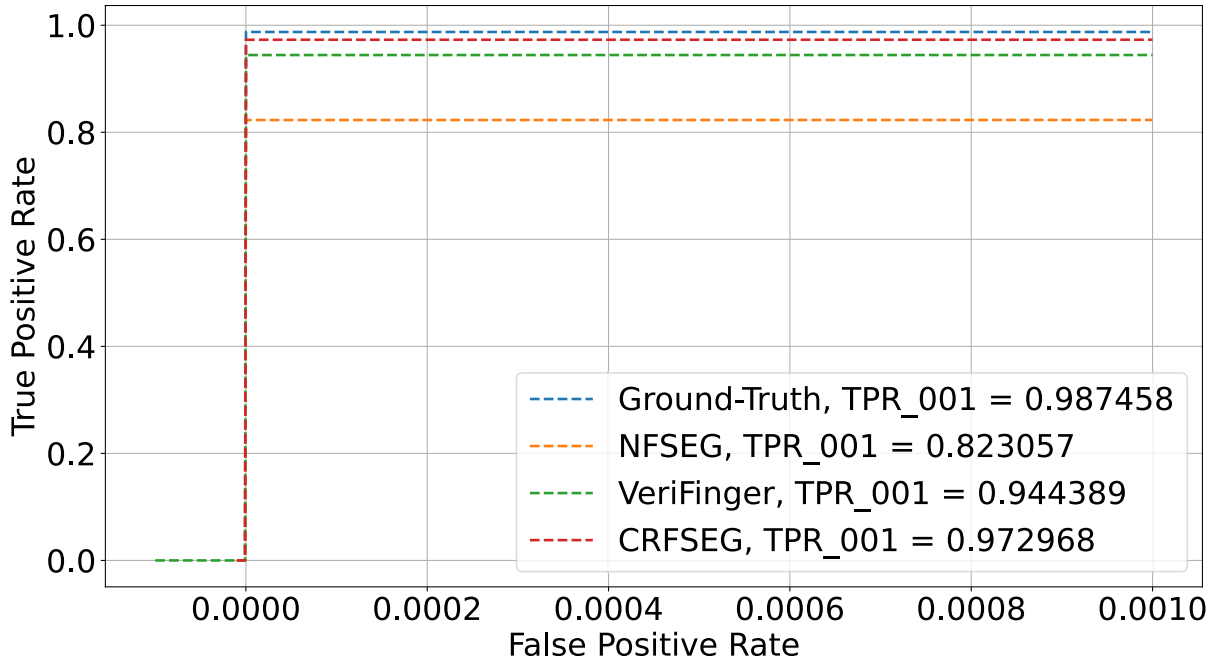


Figure 12. The Receiver Operating Characteristics (ROC) for Ground Truth, NFSEG, VeriFinger, CRFSEG on adult subjects in the *Combined* dataset.

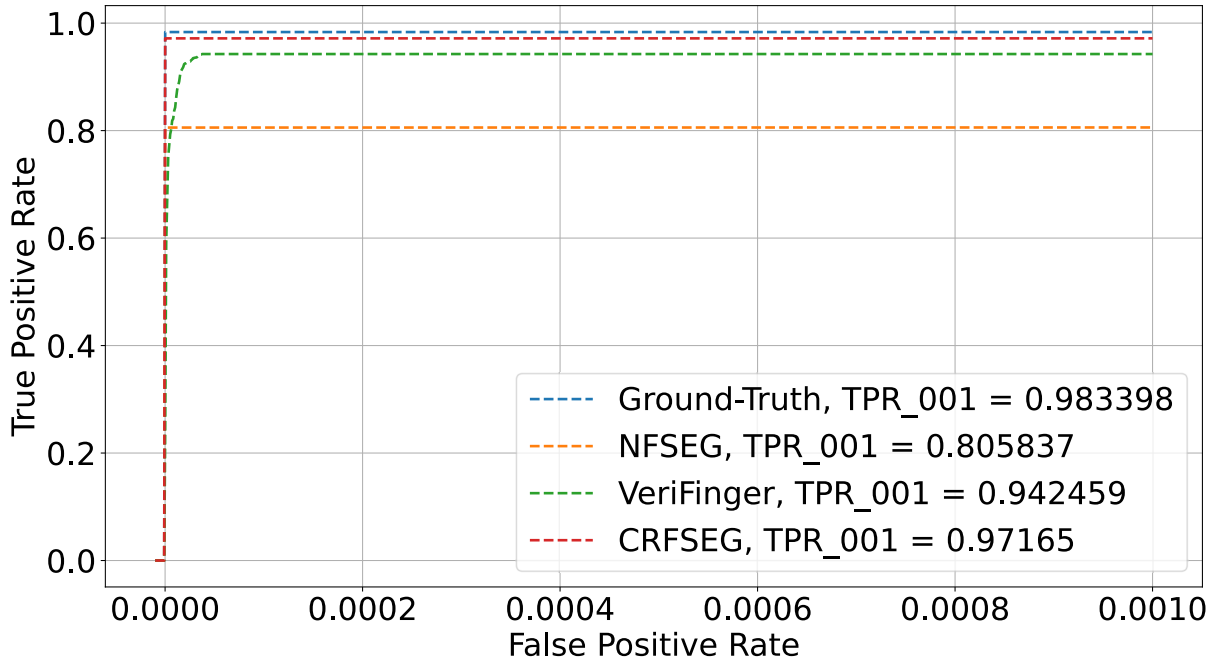


Figure 13. The Receiver Operating Characteristics (ROC) for Ground Truth, NFSEG, VeriFinger, CRFSEG on both adult and children subjects in the *Combined* dataset.

this objective, we designed an oriented region proposal network (ORPN) in our CNN architecture. The main advantage of this design is the model's ability to learn fingerprint features with different orientations, resulting in better performance. The results of Mean Absolute Error (MAE) and Error in Angle Prediction (EAP) confirmed the notable superiority of our proposed approach in terms of bounding box predictions.

Our deep learning-based segmentation system has some limitations, including its dependency on large amounts of labeled data for training, as well as the need for significant computational resources for training and inference, which may hinder its scalability and accessibility.

Table 6. Label prediction accuracy of VeriFinger and CRFSEG on the *Challenging* dataset. In this dataset, compared to VeriFinger, the CRFSEG model achieved better results by successfully segmenting all images with a label prediction accuracy of 91.83%. This dataset is composed of slap images where NFSEG failed and thus did not have label accuracy.

Model	Dataset	Accuracy
VeriFinger	*Challenging* Dataset	82.37%
CRFSEG	*Challenging* Dataset	91.83%

Table 7. TPR at FPR of 0.001 for VeriFinger, and CFSEG on the *Challenging* dataset. The CRFSEG model demonstrated improved segmentation performance, as evidenced by its higher matching results compared to VeriFinger segmentation algorithm.

Model	Dataset	Matching Accuracy
VeriFinger	*Challenging* Dataset	79.65%
CRFSEG	*Challenging* Dataset	89.93%

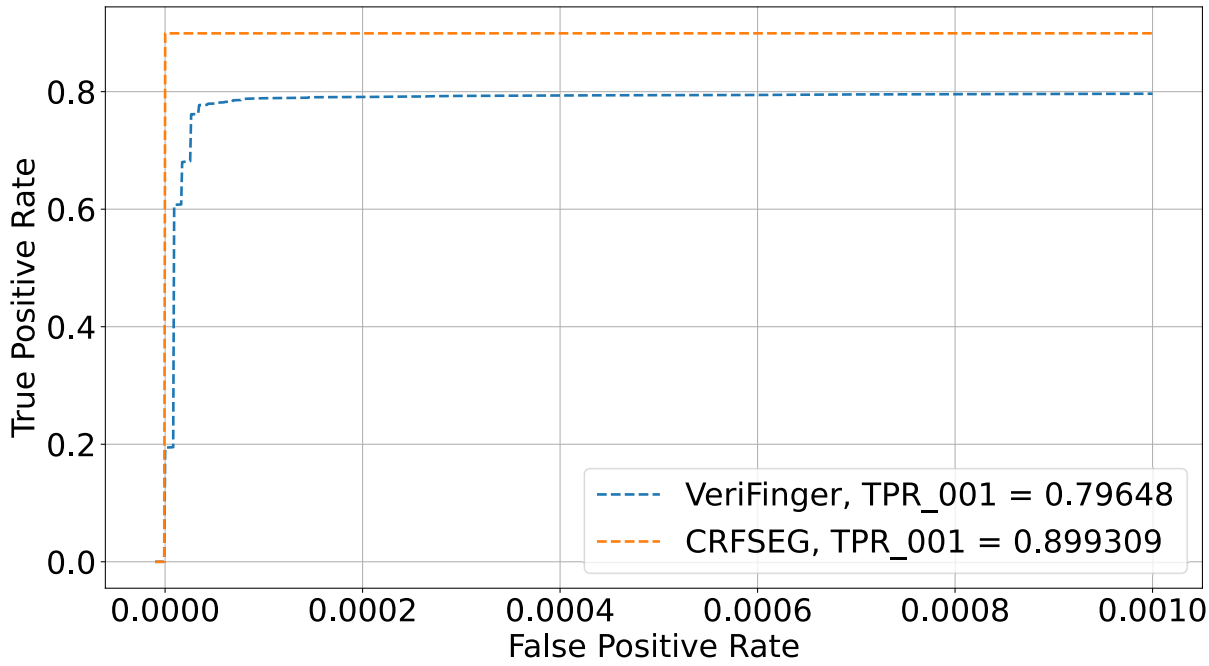


Figure 14. Receiver Operating Characteristics (ROC) for VeriFinger, CRFSEG on the *Challenging* dataset.

During testing on the challenging dataset, we discovered that the training dataset used to train CRFSEG did not include any annotated slap images with missing fingers, which were present in the challenging dataset. As a result of being trained on a dataset without missing fingerprints, CRFSEG occasionally mislabels fingerprints in slap images, even though it can accurately detect the fingerprints of such images. This problem can be addressed by adding images with missing fingers to the training dataset. In summary, CRFSEG may struggle to segment new and diverse data beyond the scope of the training dataset.

Conclusions

We designed an oriented bounding box-based deep learning method to segment overly rotated slap fingerprint images of different age groups. Unlike conventional image detection and segmentation algorithms, this algorithm uses five parameters to determine a rotated or axis-aligned object, where four parameters are used to determine the position of a bounding box around the object, and the last one is used for the angle of that bounding box.

We used a modified version of Faster R-CNN to achieve this goal. The two-stage object detection architecture depicted in Figure 5 utilizes a deep learning feature extractor model, pre-trained on the Imagenet dataset, to extract semantic-rich feature maps from input slap images at different scales. These feature maps are then processed by the O-RPN, a modified regional proposal network specially designed for detecting both axis-aligned and rotated objects, to precisely segment objects in an input image. Another branch of the detection architecture uses the same feature maps but outputs the label of objects.

In the training stage, the dataset was initially augmented to increase its size and to incorporate common scenarios found in a real-world slap dataset, such as overly rotated slap images. Then, the Faster R-CNN based search approach was used to find the positive ROIs by calculating the IOU between detected bounding boxes and ground-truth bounding boxes. Those positive ROIs were used to learn the fingerprint features in a slap image. A scale-up search procedure is used during training to find positive ROIs and to ensure that all potential fingerprint regions are found and processed correctly. In the inference stage, all the test slap images are processed by the feature extraction network, and hundreds of ROIs are proposed by the O-RPN. After that, the cross-entropy is performed to generate class labels, and regression is performed to locate the precise bounding box around fingerprints. A threshold (≥ 0.7) of the class score is used to filter the proposal regions and determine whether it is a fingerprint. Finally, post-processing, such as non-maximum suppression, is performed to obtain the final segmentation.

We segmented all the slap images using ground-truth information, NFSEG, VeriFinger, and CRFSEG, resulting in four sets of segmented fingerprint images. The success of various fingerprint segmentation systems was evaluated by comparing the matching scores of the segmented images produced by each system. The CRFSEG model showed a 97.17% match accuracy on the *Combined* dataset, which was higher compared to the NFSEG and Verifinger systems, which had 80.58% and 94.25% match accuracy respectively. We also calculated the MEA, EAP, and label prediction accuracy to evaluate all the slap segmentation models. The experimental result showed that our CRFSEG model segmented fingerprints robustly across over-rotated slap images for both children and adult subjects. We outperformed state-of-the-art NFSEG and VeriFinger in terms of MAE, EAP, and matching performance. A pre-trained CRFSEG model and corresponding training codes are publicly accessible (https://github.com/sarwarmurshed/CRFSEG).

References

[1] Brian Cochran Craig Watson Patricia Flanagan. *SlapSegII – Slap Fingerprint Segmentation Evaluation II*. en. 2010-11-16 2010. URL: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=901467.

[2] *Cross Match L SCAN Patrol and Patrol ID fingerprint scanners*. URL: https://neurotechnology.com/fingerprint-scanner-cross-match-l-scan-patrol.html (visited on 07/24/2021).

[3] Xiaowei Dai et al. "Fingerprint Segmentation via Convolutional Neural Networks". In: *Biometric Recognition*. Ed. by Jie Zhou et al. Cham: Springer International Publishing, 2017, pp. 324–333. ISBN: 978-3-319-69923-3.

[4] Javier Galbally, Rudolf Haraksim, and Laurent Beslay. "A Study of Age and Ageing in Fingerprint Biometrics". In: *IEEE Transactions on Information Forensics and Security* 14 (2019), pp. 1351–1365.

[5] Patrick Grother, Wayne Salamon, Ramaswamy Chandramouli, et al. "Biometric specifications for personal identity verification". In: *NIST Special Publication* 800 (2013), pp. 76–2.

[6] Sooji Ha et al. "Topic classification of electric vehicle consumer experiences with transformer-based deep learning". In: *Patterns* 2.2 (2021), p. 100195.

[7] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[8] Kaiming He et al. "Mask r-cnn". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2961–2969.

[9] Young-Hoo Jo et al. "Security Analysis and Improvement of Fingerprint Authentication for Smartphones". In: *Mobile Information Systems* 2016 (Jan. 2016), pp. 1–11. DOI: 10.1155/2016/8973828.

[10] Derek Johnson. "Segmentation of slap fingerprints". In: 2010.

[11] Indu Joshi et al. "Sensor-invariant Fingerprint ROI Segmentation Using Recurrent Adversarial Learning". In: *2021 International Joint Conference on Neural Networks (IJCNN)*. 2021, pp. 1–8. DOI: 10.1109/IJCNN52387.2021.9533712.

[12] Kate Keahey et al. "Lessons learned from the chameleon testbed". In: *Proceedings of the 2020 USENIX Annual Technical Conference (USENIX ATC '20)*. USENIX Association, 2020, pp. 219–233.

[13] Kenneth Ko, Wayne J. Salamon. *NIST Biometric Image Software (NBIS)*. https://www.nist.gov/services-resources/software/nist-biometric-image-software-nbis. Last Accessed: Dec 12, 2022. Feb. 2010.

[14] Sameh Khamis and Larry S. Davis. "Walking and talking: A bilinear approach to multi-label action recognition". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2015, pp. 1–8. DOI: 10.1109/CVPRW.2015.7301277.

[15] Oluwasanmi O Koyejo et al. “Consistent multilabel classification”. In: *Advances in Neural Information Processing Systems* 28 (2015).

[16] Wu Lei and You Lin. “A Novel Dynamic Fingerprint Segmentation Method Based on Fuzzy C-Means and Genetic Algorithm”. In: *IEEE Access* 8 (2020), pp. 132694–132702. DOI: 10.1109/ACCESS.2020.3011025.

[17] Nalla et al. “De-Duplication Complexity of Fingerprint Data in Large-Scale Applications”. In: 2014.

[18] NeuroTechnology. *VeriFinger*. http://www.neurotechnology.com/verifinger.html. Last Accessed: Jan 02, 2023.

[19] Joseph Redmon et al. “You only look once: Unified, real-time object detection”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 779–788.

[20] Shaoqing Ren et al. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”. In: *Advances in Neural Information Processing Systems*. Ed. by C. Cortes et al. Vol. 28. Curran Associates, Inc., 2015. URL: https://proceedings.neurips.cc/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf.

[21] M. G. Sarwar Murshed et al. “Deep Slap Fingerprint Segmentation for Juveniles and Adults”. In: *2021 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*. 2021, pp. 1–4. DOI: 10.1109/ICCE-Asia53811.2021.9641980.

[22] Paulo Bruno S. Serafim et al. “A Method based on Convolutional Neural Networks for Fingerprint Segmentation”. In: *2019 International Joint Conference on Neural Networks (IJCNN)*. 2019, pp. 1–8. DOI: 10.1109/IJCNN.2019.8852236.

[23] Branka Stojanović et al. “Fingerprint ROI segmentation based on deep learning”. In: *2016 24th Telecommunications Forum (TELFOR)*. 2016, pp. 1–4. DOI: 10.1109/TELFOR.2016.7818799.

[24] Grigorios Tsoumakas and Ioannis Manousos Katakis. “Multi-Label Classification: An Overview”. In: *Int. J. Data Warehous. Min.* 3 (2007), pp. 1–13.

[25] Tzutalin. *LabelImg*. https://github.com/tzutalin/labelImg. 2015.

[26] CI Watson et al. *Fingerprint vendor technology evaluation, nist interagency/internal report 8034: 2015*. 2014.

[27] Craig Watson. *SlapSegII Analysis: Matching Segmented Fingerprint Images*. en. 2010-11-16 2010. URL: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=906389.

[28] Charles Wilson et al. “Fingerprint vendor technology evaluation 2003: Summary of results and analysis report”. In: *NIST Technical Report NISTIR* 7123 (2004).

[29] Yuxin Wu et al. *Detectron2*. https://github.com/facebookresearch/detectron2. 2019.

Acknowledgements

This material is based upon work supported by the Center for Identification Technology Research and the National Science Foundation (NSF) under Grant No.1650503. We would like to thank Winnie Liu and Heidi Walko for their contribution in preparing the dataset. We also extend their gratitude to Precise Biometrics and the Potsdam Elementary and Middle School administration, staff, students, and the parents of the participants for supporting our research and the greater goal of scientific contribution to society. The authors would also like to thank the *Chameleon* cloud for providing the computational infrastructure required for this work [12].