A Deep Learning based Unified Framework to Detect, Segment and Recognize Irises Using Spatially Corresponding Features

Zijing Zhao, Ajay Kumar

Department of Computing, The Hong Kong Polytechnic University

Abstract

This paper proposes a deep learning based unified and generalizable framework for accurate iris detection, segmentation and recognition. The proposed framework firstly exploits state-of-the-art and iris-specific Mask R-CNN, which performs highly reliable iris detection and primary segmentation i.e., identifying iris/non-iris pixels, followed by adopting an optimized fully convolutional network (FCN), which generates spatially corresponding iris feature descriptors. A specially designed Extended Triplet Loss (ETL) function is presented to incorporate the bit-shifting and non-iris masking, which are found necessary for learning meaningful and discriminative spatial iris features. Thorough experiments on four publicly available databases suggest that the proposed framework consistently outperforms several classic and state-of-the-art iris recognition approaches. More importantly, our model exhibits superior generalization capability as, unlike popular methods in the literature, it does not essentially require database-specific parameter tuning, which is another key advantage.

Keywords: iris recognition, deep learning, spatially corresponding features

1. Introduction

Iris recognition has emerged as one of the most accurate and reliable biometric approaches for the human recognition. Automated iris recognition systems therefore have been widely deployed for various applications from border control [1], citizen authentication [2], forensic [3] to commercial products [4]. The

usefulness of iris recognition has motivated increasing research effort in the past decades for exploring more accurate and robust iris matching algorithms under different circumstances [5, 6, 7, 8, 9, 10, 11].

In recent years, deep learning has gained tremendous success especially in the area of computer vision, and accomplished state-of-the-art performance for a number of tasks such as general image classification [12], object detection [13] and face recognition [14, 15]. However, unlike face, in the field of iris recognition, in the best of to our knowledge, there is almost nil attention to incorporate remarkable capabilities of the deep learning and achieve superior performance than popular or state-of-the-art iris recognition methods.

In the conference version of this paper [16], we proposed a new deep learning based iris recognition framework which not only achieves satisfactory matching accuracy but also exhibits outstanding generalization capability to different databases. With the design of effective fully convolutional network, our model is able to learn comprehensive *spatially corresponding* iris features which generalize well on different datasets. A newly developed Extended Triplet Loss (ETL) function provides meaningful and extensive supervision to the iris feature learning process with limited size of training data.

In the conference version [16], as the key focus was on learning effective iris feature representation, the approach relies on external and conventional method [17] for iris circle detection, which is parameter-sensitive and less generalizable to different datasets. This paper extends our previous work by integrating iris detection and segmentation from raw eye images into a unified framework, which is based on deep learning and referred to as *UniNet.v2* and is shown in Fig.1. Such an approach essentially improves detection and segmentation accuracy as well as robustness, and finally benefits the recognition performance as shown from the extensive experimental results. More importantly, by incorporating deep neural networks into the iris detection process, our framework can easily adapt to varying image qualities without additional parameter tuning. The high level of integration of our architecture also enables more consistent and smooth learning for the iris feature representation with respect to the deep learning based

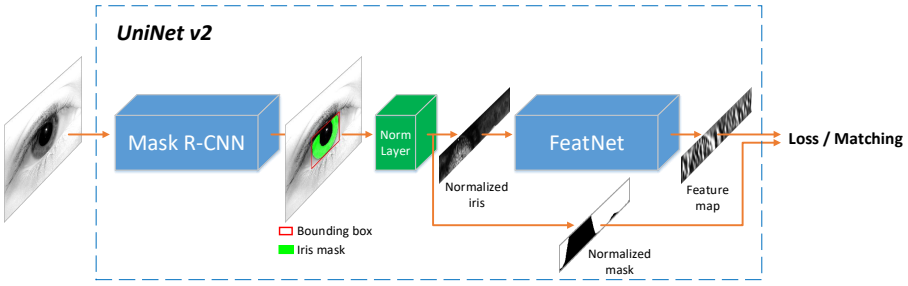


Figure 1: Overall framework of the proposed UniNet.v2. Raw acquired eye image is taken as input and parsed by an iris-specific Mask R-CNN for detecting iris location and segmenting iris region pixels. A normalization layer is developed and applied to fit circles and normalize the iris. Spatially corresponding feature map is then extracted by FeatNet for feature learning or matching.

detection and segmentation results. Kindly note that by "unified framework" we mainly refer to the fact that our framework utilizes deep learning techniques to cover all the major tasks in the workflow of an iris recognition system rather than implying an end-to-end optimization process for the framework.

The main contributions of this paper can be summarized as follows: (i) We develop a new deep learning based iris feature extractor which can generate highly effective deep iris representation. A new Extended Triplet Loss function has been developed to successfully address the nature of iris pattern for the feature learning. Significant advancement therefore has been made to bridge the gap between deep learning and iris recognition. (ii) Under fair comparison, our approach consistently outperforms several state-of-the-art methods on multiple datasets. Even under challenging scenario that without having any parameter tuning on the target dataset, our model can still achieve superior performance over state-of-the-art methods that have been extensively tuned. (iii) We present a deep learning based framework that can accurately detect, segment and recognize irises from raw eye images as input. Such unified framework provides higher robustness and consistency for the iris segmentation and feature optimization processes.

1.1. Related Work

One of the most classic and effective approaches for automated iris recognition was proposed by Daugman [5] in 2002. In his work, Gabor filter is applied on the segmented and normalized iris image, and the responses are then binarized as IrisCode. The hamming distance between two IrisCodes is used as the dissimilarity score for verification. Based on [1], 1D log-Gabor filter was proposed in [6] to replace 2D Gabor filter for more efficient iris feature extraction. A different approach, developed in [7] in 2007, has exploited discrete cosine transforms (DCT) for analyzing frequency information of image blocks and generating binary iris features. Another frequency information based approach was proposed in [9] in 2008, in which 2D discrete Fourier transforms (DFT) was employed. In 2009, the multi-lobe differential filter (MLDF), which is a specific kind of ordinal filters, was proposed in [8] as an alternative to the Gabor/log-Gabor filters for generating iris templates.

In addition to exploring various iris feature representations, researchers have devoted significant efforts to improving iris segmentation accuracy and robustness. In earlier years the integro-differential operator [5] and circular Hough transform [6] are adopted for detecting iris and pupil circles in eye images. These methods work well for high-quality iris images but are usually less reliable for noisy or blur images acquired under relaxed environments. An improved method proposed in [18] exploits an iterative approach to coarsely cluster the iris and non-iris region before applying integro-differential operator, which offers higher reliability for segmenting iris pixels. Following similar coarse-to-fine strategy, a competitive work detailed in [19] makes use of Random Walker algorithm [20] for coarsely locating the iris region, followed by a couple of gray-level statistics based thresholding to refine the boundaries. These thresholding operations enable pixel-level precision for the iris masks. Recent approaches include [17] which utilizes an improved total variation model to deal with undesired noise and artifacts in casually captured iris images, and [21] which relies on color/illumination correction and watershed transform for segmenting noisy iris images captured under visible wavelength.

Unlike the popularity of deep learning for various computer vision tasks, especially for face recognition, the literature so far has not yet fully exploited its potential for iris recognition. There has been very little attention on exploring iris recognition using deep learning. A deep representation for iris was proposed in [22] in 2015, but the purpose was for spoofing detection instead of iris recognition. A recent approach named DeepIrisNet in [23] has investigated deep learning based frameworks for general iris recognition. This work is essentially a direct application of typical convolutional neural networks (CNN) without much optimization for iris pattern. Our reproducible experimental comparison in section 5.3 further indicates that under fair comparison, this approach [23] cannot deliver superior performance even over other popular methods. Another recent work [24] has attempted to employ deep belief net (DBN) for iris recognition. Its core component, however, is the optimal Gabor filter selection, while the DBN is again a simple application on the IrisCode without iris-specific optimization. Above studies have made preliminary exploration but failed to establish substantial connections between iris recognition and deep learning.

1.2. Limitations and Challenges

Despite the popularity of iris recognition in biometrics, conventional iris feature descriptors do have several limitations. The summaries of earlier work in [25, 26] reveal that existing methods can achieve satisfactory performance, but the performance needs to be further improved to meet the expectations for wider range of deployments. Besides, traditional iris features, such as IrisCode, are mostly based on empirical models which apply hand-crafted filters or feature generators. As a result, these models rely heavily on parameter selection when applied for different databases or imaging environments. Although there are some standards on iris image format [27], the selection of parameter for feature extraction remains empirical, or based on training methods such as boosting [28]. This situation can be observed from [8], where eight different combinations of parameters for ordinal filters delivered varying performance on three databases, or from [29] which employed two sets of parameters for log-Gabor

filter on two databases by extensive tuning. Another limitation is that due to the simplicity of conventional iris descriptors, they are less promising to fully exploit the underlying distribution from various types of iris data available today. Learning data distribution from large amount of samples to further advance performance is one of the key trends nowadays. Approaches for iris segmentation also suffer from similar challenges. Most of existing methods, such as [30], [18] and [17], rely on hand-crafted procedures for identifying iris pixel regions. These operators are usually defined by a set of empirically tuned parameters and less generalizable to different types of images.

Deep learning has the potential to address the above limitations, since the parameters in deep neural networks are learned from data instead of being empirically set, and deep architectures are known to have good generalization capability. However, new challenges emerge while incorporating typical deep learning architectures (e.g., CNN) for the iris recognition, which can be primarily attributed to the nature of iris patterns. Different from face, iris pattern is observed to reveal little structural information or meaningful hierarchies. Iris texture is believed to be random [31]. Earlier promising works on iris recognition [5, 6, 7, 8, 9] mainly employed small-size filters or block-based operations to obtain iris features. Therefore, we can infer that the most discriminative information in the iris pattern comes from the local intensity distribution of an iris image rather than the global features, if any. CNN is known as effective for extracting features from low level to high level, and from local to global, due to the combination of convolutional layers and fully connected layers [32]. However, as discussed above, high level and global features may not be the optimal for iris representation.

This paper aims to develop a unified framework for more accurate and robust iris detection, segmentation and recognition, making solid contributions towards fully discovering the potential of deep learning for the iris biometrics. Such objectives have not yet been pursued in the literature. Different from [23] and [24], this paper proposes a novel deep network and customized loss function, which are highly optimized for extracting discriminative iris features and have

been comparatively evaluated with several state-of-the-art methods on multiple iris image databases.

The rest of this paper is organized as follows: Section 2-4 detail the proposed approach in terms of network architecture, improved triplet loss function and feature encoding respectively; Section 5 presents the experimental configurations, results and analysis; finally, the key conclusions from this paper are presented in Section 6.

2. Network Architecture

In this paper we propose a unified and highly optimized deep learning framework, referred to as UniNet.v2, for iris detection, segmentation and feature extraction from raw eye images. As shown in Fig. 1, the unified architecture consists of several sub-components, which are an iris-specific Mask R-CNN [33], a normalization layer and a feature extraction network referred to as FeatNet. The technical specifications and optimization methods for these sub-networks are detailed in the following sections.

2.1. Mask R-CNN for Iris Detection and Segmentation

Accurate iris detection and segmentation are critical for achieving higher performance for the iris based personal identification. Inadequate segmentation can lead to severe degradation of the performance for automated iris recognition systems. We propose to exploit Mask R-CNN [33], one of state-of-the-art architectures designed for general instance segmentation, for improving iris detection and segmentation accuracy and reliability. A unified framework is enabled by the introduction of Mask R-CNN, which improves the stability and consistency between the iris masks and corresponding features. Kindly note that in modern iris recognition approaches, iris segmentation usually involves two parts: (a) pixel-level identification of iris and non-iris regions (e.g., excluding eyelash, sclera and reflection) and (b) fitting circles/ellipses or other geometric representations on iris structures to assist part (a) and to serve for normalization. In

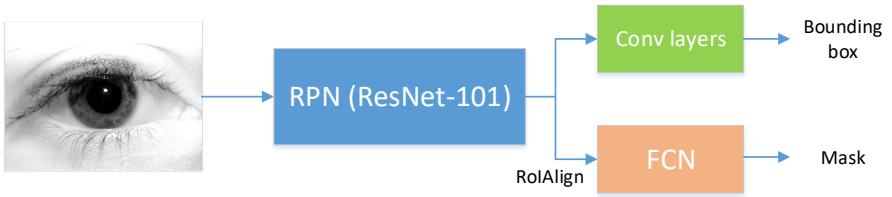


Figure 2: Summary of structure of Mask R-CNN employed in our work. A backbone CNN (we use ResNet-101 [35]) extracts features from the input image and proposes possible RoIs for the desired object. A head branch is used to evaluate objectiveness and regress bounding boxes, while another FCN branch predicts the object masks within each proposed RoI. Kindly note that prediction object class is reduced compared with the original implementation as only one class, i.e., iris, is of our interest.

earlier works [5, 6], part (b) is the main focus, while more recent and advanced methods [17, 18, 19] also heavily address part (a). In this paper, we employ Mask R-CNN for part (a) of the iris segmentation process, i.e., identifying iris and non-iris pixel regions from the input eye images.

2.1.1. Basic Introduction to Mask R-CNN

The overall structure of Mask R-CNN employed in our framework is illustrated in Fig. 2. Mask R-CNN is built on top of its predecessor, i.e., Faster R-CNN [34]. In this framework, the input eye image from iris sensor is firstly subjected to a backbone CNN, which serves as region proposal network (RPN), to obtain initial guesses of regions that may contain a desired object. The proposed regions are then sent to a branch classification network for identifying object classes within each region. In our approach, however, there is only one foreground class (i.e., iris) to be detected, therefore the classification branch is reduced. In addition, we assume each input eye image contains just a single iris, hence only the proposal with highest confidence will be processed subsequently. Such simplification of Mask R-CNN can better regularize the training process to avoid over-fitting when it is adopted to learn iris regions and masks.

Mask R-CNN includes one more branch, which is a fully convolutional network [14], to the Faster R-CNN in order to segment instance masks simulta-

neously inside the proposed regions. According to [33], RoIAlign operation is introduced to recover pixel-level segmentation accuracy on downsampled feature maps, and state-of-the-art performance was reported for the COCO segmentation challenge [36]. Due to its outstanding performance and built-in detection-segmentation design, Mask R-CNN is highly promising for addressing the reliability and generalizability for the iris segmentation as well as constructing a unified framework for iris recognition.

2.1.2. Training of Mask R-CNN for Detecting and Segmenting Irises

Adequate number of training samples along with their ground truth bounding box labels and instance masks are required to sufficiently train Mask R-CNN for the specific task, i.e., iris detection and segmentation in this paper. We adopted a semi-manual procedure to label images from multiple publicly available databases in order to enrich data variation in less available time. Firstly, a conventional iris segmentation approach [17] with well tuned parameters was applied on the training sets from each database. The training sets are subject-disjoint with the test sets as will be explained in more details in the experimental section. We then manually inspected the segmentation results and selected best ones, then incorporated some manual operations, such as filling holes and removing isolated pixels, to refined the segmentation results. Such filtered iris masks were regarded as ground truths and formed a training set of 1,700 images and a validation set of about 300 images. These samples were then used to fine-tune a Mask R-CNN model which has been pre-trained on the COCO dataset with some modification as discussed in previous section.

2.1.3. Iris Normalization Layer

A normalization layer is appended after Mask R-CNN, as shown in Fig. 1, to perform iris and mask normalization (unwrapping) before learning iris features. Input to this layer is the cropped image and mask from the full size image, where the crop region centers at the detected bounding box but is 1.2 times larger in order to accommodate marginal errors. Within this layer, simple circular Hough

transform which is similar to the one in [17] is applied for detecting iris and pupil circles. However, unlike in [17] where the circle detection is performed on the entire image as no prior information is known, in this framework the search region for the circle center is made near the center of the bounding box (especially the x-position), and the fitting range for the radius is also set to be around half of the width of the RoI. With such spatial constraints, the circle detection is least likely to generate erroneous output compared with the approach in [17].

Kindly note that unlike common iris segmentation approaches, there is no dataset-specific parameters required for the above circle detection step. Variables like possible range of radius are automatically inferred from the dynamically detected RoI from Mask R-CNN in order to achieve good generalizability. As will be shown from the experiments, the circle detection accuracy from the proposed framework is much higher than conventional methods. After the iris and pupil circles are detected, the iris region and mask are normalized using the classic rubber-sheet model [5] into a resolution of 512×64. The next step is to learn effective spatially corresponding features using our original FeatNet component and extended triplet loss (ETL) function, which will be detailed in the following sections.

2.2. FeatNet: Learning Spatially Corresponding Iris Features

2.2.1. Image Preprocessing

After the iris is detected and normalized, we apply a simple contrast enhancement process, which adjusts the image intensity so that 5% pixels are saturated at low and high intensities respectively. The enhanced images are the fed into the subsequent network, referred to as FeatNet, for extracting comprehensive features. Fig. 3 illustrates the key steps of image preprocessing.

2.2.2. Learning Spatially Corresponding Features

FeatNet is designed for extracting discriminative iris features and based on fully convolutional networks (FCN), which is originally developed for semantic

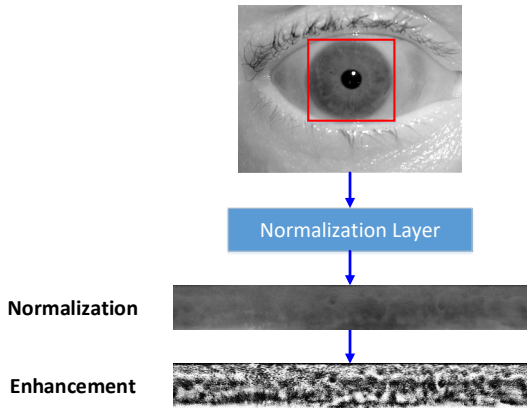


Figure 3: Illustration of enhancement effect for normalized iris image

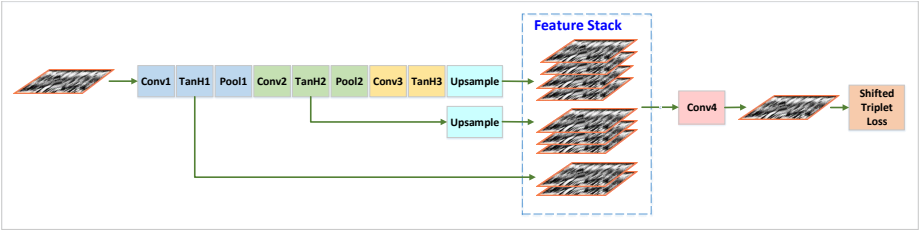


Figure 4: Detailed structures for FeatNet. This network gathers convolutional feature maps at different scales and resize them to the same resolution to form a feature stack. These features are then fused by a convolutional layer to generate a single-channel feature map which retains spatial correspondence with the original input.

segmentation [14]. Different from common convolutional neural network (CNN), the FCN does not have fully connected layer. The major components of FCN are convolutional layers, pooling layers, activation layers, etc. Since all these layers operate on local regions around pixels from their bottom map, the output map can preserve spatial correspondence with the original input image. By incorporating up-sampling layers, FCN is able to perform pixel-to-pixel prediction. The detailed structure of FeatNet is provided in Fig. 4 and Table 1.

As shown in Fig. 4, the input iris image is forwarded by several convolutional layers, activation layers and pooling layers. The network activations at different scales, i.e., TanH1-3, are then up-sampled if necessary to the size of original

Table 1: Layer configurations for FeatNet

Layer	Type	Kernel size	Stride	# Output channels
Conv1	Convolution	3×7	1	16
Conv2	Convolution	3×5	1	24
Conv3	Convolution	3×3	1	32
Conv4	Convolution	3×3	1	1
Tanh1, 2, 3	TanH activation	/	/	/
Pool1, 2, 3	Average pooling	2×2	2	/

input. These features form a multi-channel feature stack which contains rich information from different scales, and are finally convolved again to generate an integrated single-channel feature map.

The reason for selecting FCN instead of CNN for iris feature extraction primarily lies in the previous analysis on iris patterns in Section 1.2, i.e., the most discriminative information of an iris probably comes from small and local patterns. FCN is able to maintain local pixel-to-pixel correspondence between input and output, and therefore is a better candidate for the iris feature extraction. Regarding the format of the iris feature and depth of the network, there are trade-offs among the level of feature locality, complexity and compatibility with traditional iris recognition systems. When the network goes deeper, the receptive field, which describes how large the input area affects each output element, becomes larger and fine details can be more easily lost, resulting in higher level and more global feature descriptors [37]. As pointed out earlier, global features may not be suitable for iris recognition. On the other hand, networks with shallow layers can hardly capture enough or comprehensive information from the input images. After extensive exploration, we employed four convolutional layers for FeatNet as shown in Fig. 4 to achieve balance between maintaining feature locality as well as enabling comprehensive feature extraction. Another factor is the number of channels of the output iris feature map. In theory, multi-channel or multi-scale features can enrich the information in the descriptor and lead to higher recognition accuracy with more complex matching mechanisms [38, 8]. However, in this paper the primarily goal is to investigate fundamen-

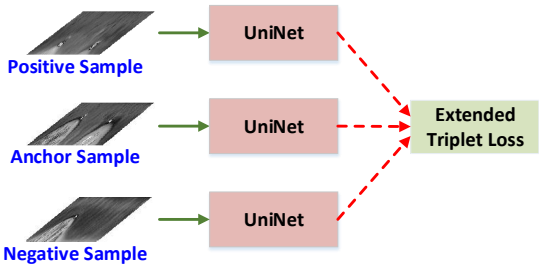


Figure 5: Triplet-based network organization for training

tal feature effectiveness, and also for the easier and fair comparison with other existing methods, we therefore only focus on single-channel iris feature map.

2.3. Triplet-based Network Architecture

A triplet network [39] was implemented for learning the convolutional kernels in FeatNet. The overall structure for the triplet network in the training stage is illustrated in Fig.5. As shown in the figure, three copies of Uninets, whose weights are kept identical during training, are placed in parallel to forward and back-propagate the data and gradients for anchor, positive and negative samples respectively. The anchor-positive (AP) pair should come from the same person while the anchor-negative (AN) pair comes from different persons. The triplet loss function in such architecture attempts to reduce the anchor-positive distance and meanwhile increase the anchor-negative distance. However, in order to ensure more appropriate and effective supervision in the generation of iris features by the FCN, we improve the original triplet loss by incorporating a bit-shifting operation. The improved loss function is referred to as Extended Triplet Loss (ETL), whose motivation and mechanism are detailed in the next section.

3. Extended Triplet Loss Function

In this paper we develop a problem-specific loss function for more effective iris feature learning, which is referred to as extended triplet loss (ETL) function.

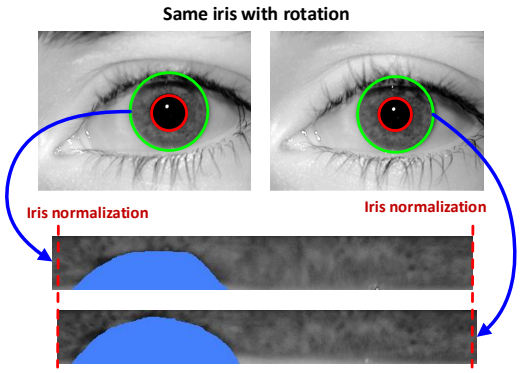


Figure 6: Illustration of occlusions (labeled in blue) and horizontal translation which usually exist between two normalized iris images even from a same iris.

The original loss function for a triplet network is defined as follows:

$$L = \frac{1}{N}\sum_{i=1}^{N}\left[\|\boldsymbol{f}_i^A - \boldsymbol{f}_i^P\| - \|\boldsymbol{f}_i^A - \boldsymbol{f}_i^N\| + \alpha\right]_+ \tag{1}$$

where N is the number of triplet samples in a mini-batch, $\boldsymbol{f}_i^A$, $\boldsymbol{f}_i^P$ and $\boldsymbol{f}_i^N$ are the feature maps of anchor, positive and negative images in the i-th triplet respectively. The symbol $[\cdot]_+$ is the as same as used in [39] and is equivalent to $\max(\cdot, 0)$. α is a preset parameter to control the desired margin between anchor-positive distance and anchor-negative distance. Optimizing above loss will lead to the anchor-positive distance being reduced and anchor-negative distance being enlarged until their margin is larger than a certain value.

In our case, however, using Euclidean distance as the dissimilarity metric is far from sufficient. As discussed earlier, we propose using spatial features which have the same resolution with the input, the matching process has to deal with non-iris region masking and horizontal shifting, which are frequently observed in iris samples as illustrated in Fig.6. Therefore in the following, we extend the original triplet loss function to address the above issues.

3.1. Incorporating Masking and Shifting

As discussed earlier, in this paper we extend the original triplet loss function Eq.1 to deal with non-iris regions and horizontal translation, which we refer to

as the Extended Triplet Loss (ETL):

$$ETL = \frac{1}{N}\sum_{i=1}^{N}\left[\mathcal{D}\left(\boldsymbol{f}_i^A, \boldsymbol{f}_i^P\right) - \mathcal{D}\left(\boldsymbol{f}_i^A, \boldsymbol{f}_i^N\right) + \alpha\right]_+ \tag{2}$$

where $\mathcal{D}\left(\boldsymbol{f}^1, \boldsymbol{f}^2\right)$ is the Minimum Shifted and Masked Distance (MMSD) function, defined as follows:

$$\mathcal{D}\left(\boldsymbol{f}^1, \boldsymbol{f}^2\right) = \min_{-B\leq b\leq B}\left\{\mathcal{FD}\left(\boldsymbol{f}_b^1, \boldsymbol{f}^2\right)\right\} \tag{3}$$

$\boldsymbol{f}_b$ represents a shifted version of $\boldsymbol{f}$ obtained by horizontally shifting it by b pixels, and $\mathcal{FD}$ denotes the Fractional Distance. The shifted and the original feature maps have the following spatial correspondence:

$$\begin{gathered}\boldsymbol{f}_b[x_b, y] = \boldsymbol{f}[x, y] \\ x_b = (x - b + W) \bmod W\end{gathered} \tag{4}$$

where x, y are the spatial coordinates and x_b is obtained by shifting the pixel to the left by a step of b, assuming W is the width of the 2-D feature map. Note that when x is less than b, the pixel position will be directed to the right end of the map, as the iris map is normalized by unwrapping the original iris circularly and the left end is therefore physically connected with the right end. When b is negative, the bit-shifting operation would shift the map to the right by $-b$ pixels. The Fractional Distance $\mathcal{FD}$ in Eq.3 measures the relative difference between two feature maps within non-masked regions only and normalize it by the number of involved pixels:

$$\mathcal{FD}\left(\boldsymbol{f}^1, \boldsymbol{f}^2\right) = \frac{1}{|M|}\sum_{(x,y)\in M}\left(f_{x,y}^1 - f_{x,y}^2\right)^2 \tag{5}$$

where M is the common non-masked regions for the two feature maps.

Eq.3 and Eq.5 indicate that the new loss function will only evaluate the difference between features within non-masked areas and a shifting operation will be performed to address the horizontal translation, so that matching of the proposed spatially corresponding iris features is meaningful. In the following we will derive the gradients of the proposed ETL in order to perform back-propagation for the learning process. The cases of real-valued and binary versions ETL are quite different and therefore we will separately proceed.

3.2. Back-propagation for ETL

The components of the real-valued ETL are all differentiable and therefore the computation of gradients is quite straightforward. Firstly, in order to maintain simplicity of the notations for the upcoming derivation, we denote the offsets that fulfills the MMSD of AP-pair and AN-pair as follows:

$$\begin{aligned} b_{AP} &= \underset{-B\leq b\leq B}{\arg\min} \left\{\mathcal{FD}\left(\boldsymbol{f}_b^A, \boldsymbol{f}^P\right)\right\} \\ b_{AN} &= \underset{-B\leq b\leq B}{\arg\min} \left\{\mathcal{FD}\left(\boldsymbol{f}_b^A, \boldsymbol{f}^N\right)\right\} \end{aligned} \tag{6}$$

During the back-propagation (BP) of the training process, the gradients (or partial derivatives) of the new loss on the anchor, positive and negative feature maps need to be computed. For simplicity, let us firstly derive the partial derivative w.r.t the positive feature map $\boldsymbol{f}_A$. From Eq.2 it can be derived that for one sample in the batch:

$$\frac{\partial ETL}{\partial \boldsymbol{f}_P} = \begin{cases} 0, & \text{if } ETL = 0 \\ \dfrac{1}{N}\dfrac{\partial ETL}{\partial \mathcal{D}\left(\boldsymbol{f}^A, \boldsymbol{f}^P\right)}\dfrac{\partial \mathcal{D}\left(\boldsymbol{f}^A, \boldsymbol{f}^P\right)}{\partial \boldsymbol{f}^P}, & \text{otherwise} \end{cases} \tag{7}$$

Again from Eq.2 we can see that $ETL = 0$ is equivalent to $\mathcal{D}\left(\boldsymbol{f}_i^A, \boldsymbol{f}_i^P\right) - \mathcal{D}\left(\boldsymbol{f}_i^A, \boldsymbol{f}_i^N\right) + \alpha \leq 0$. We only need to show the derivation when ETL is not 0. Let us define the set of common valid iris pixel positions for AP pair as M_{AP}, from Eq.3 and Eq.4 we have the following pixel-wise derivatives:

$$\begin{aligned} \frac{\partial \mathcal{D}\left(\boldsymbol{f}^A, \boldsymbol{f}^P\right)}{\partial \boldsymbol{f}^P[x,y]} &= \frac{\partial \mathcal{FD}\left(\boldsymbol{f}_{b_{AP}}^A, \boldsymbol{f}^P\right)}{\partial \boldsymbol{f}^P[x,y]} \\ &= \begin{cases} 0, & \text{if } (x,y) \notin M_{AP} \text{ or } ETL = 0 \\ \dfrac{-2\left(\boldsymbol{f}^A[x_{b_{AP}}, y] - \boldsymbol{f}^P[x,y]\right)}{N|M_{AP}|}, & \text{otherwise} \end{cases} \end{aligned} \tag{8}$$

And apparently $\dfrac{\partial ETL}{\partial \mathcal{D}\left(\boldsymbol{f}^A, \boldsymbol{f}^P\right)} = 1$, thus combining Eq.7 and 8 we can obtain:

$$\frac{\partial ETL}{\partial \boldsymbol{f}^P[x,y]} = \begin{cases} 0, & \text{if } (x,y) \notin M_{AP} \text{ or } ETL = 0 \\ \dfrac{-2\left(\boldsymbol{f}^A[x_{b_{AP}}, y] - \boldsymbol{f}^P[x,y]\right)}{N|M_{AP}|}, & \text{otherwise} \end{cases} \tag{9}$$

Similarly, for the partial derivatives on the negative feature map, we have:

$$\frac{\partial ETL}{\partial \boldsymbol{f}^N[x,y]} = \begin{cases} 0, & \text{if } (x,y) \notin M_{AN} \text{ or } ETL = 0 \\ \dfrac{2\left(\boldsymbol{f}^A[x_{b_{AN}},y] - \boldsymbol{f}^N[x,y]\right)}{N|M_{AN}|}, & \text{otherwise} \end{cases} \tag{10}$$

The remaining step is to calculate the derivatives w.r.t the anchor feature map. It can be seen from Eq.3 - Eq.5 that shifting the first map to the left by b pixels is equivalent to shifting the second map to the right by b pixels when computing the distance. Making use of this property, we have $\mathcal{FD}\left(\boldsymbol{f}^A_{b_{AP}}, \boldsymbol{f}^P\right) = \mathcal{FD}\left(\boldsymbol{f}^A, \boldsymbol{f}^P_{-b_{AP}}\right)$ and $\mathcal{FD}\left(\boldsymbol{f}^A_{b_{AN}}, \boldsymbol{f}^N\right) = \mathcal{FD}\left(\boldsymbol{f}^A, \boldsymbol{f}^N_{-b_{AN}}\right)$. It is therefore quite straightforward to obtain from Eq.2, 3 and 5 that:

$$\frac{\partial ETL}{\partial \boldsymbol{f}^A[x,y]} = -\frac{\partial ETL}{\partial \boldsymbol{f}^P[x_{-b_{AP}},y]} + \frac{\partial ETL}{\partial \boldsymbol{f}^N[x_{-b_{AN}},y]} \tag{11}$$

After calculating the derivative maps w.r.t $\boldsymbol{f}^A$, $\boldsymbol{f}^P$ and $\boldsymbol{f}^N$ respectively, the rest of the BP process is the same as for common CNNs. Above derivation shows that gradients will be computed only for pixels that are not masked. In this way, features are learned only within valid iris regions, while non-iris regions will be ignored since they are not of our interest. After the last convolutional layer, a single-channel feature map is generated which can be used to measure similarities between the iris samples.

4. Feature Encoding and Matching

For the real-valued features output from UniNet.v2, we perform a simple encoding scheme for the matching. The feature maps originally contain real values, and it is straightforward to measure the fractional Euclidean distance between the masked maps for matching, as the network is trained in this manner. As discussed earlier, however, binary features are more popular in most of the research works on iris recognition, since it is widely accepted by the community that binary features are more resistant to illumination change, blurring and other underlying noise. Besides, binary features consume smaller storage and enable faster matching. Therefore, we also investigated the feasibility of binarizing our features with a reasonable scheme as described in the following:

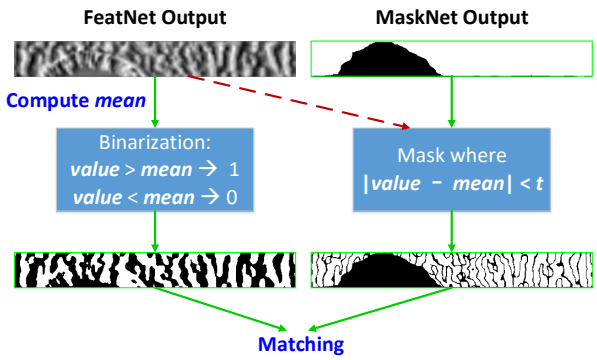


Figure 7: Illustration of feature binarization process

For each of the output feature map, the mean value of the elements within the non-masked iris regions is firstly computed as m. This mean value is then used as the threshold to binarize the original feature map. In order to avoid marginal errors, elements with feature values f close to m (i.e., $|f - m| < t$) are regarded as less reliable and will be masked together with the original mask output by MaskNet. Such a further masking step is inspired by the Fragile Bits [40], which discovered that some bits in IrisCode, with filtered responses near the axes of the complex space, are less consistent or unreliable. The range threshold t for masking unreliable bits is uniformly set to 0.6 for all the experiments. The feature encoding process can be demonstrated in Fig.7. For matching, we use the fractional Hamming distance [6] from the binarized feature maps and extended masks.

Fig. 8 presents the performance comparison between the original real-valued features and the binarized version as well as illustrating the effect of the additional masking operation, on ND-IRIS-0405 database [41]. Real-valued features are matched using Euclidean distance within the common valid iris regions while binary features are matched with Hamming distance, both averaged by the number of valid pixels. As shown in the receiver operating characteristic (ROC) curves, directly binarizing the real-valued features leads to performance degradation. After masking ambiguous feature points whose original value are close to the mean value, the performance from the binarized feature is improved and

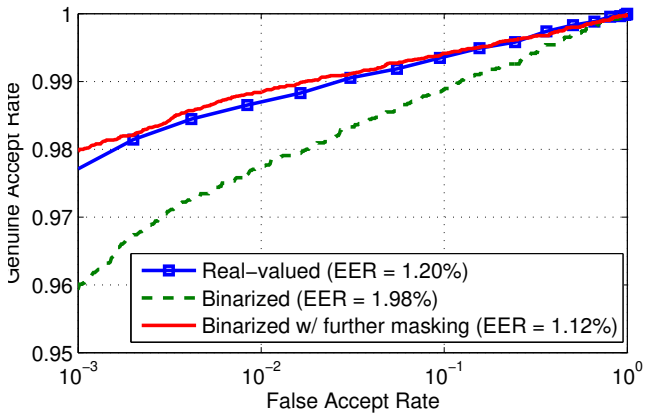


Figure 8: Comparison of ROCs from diverse deep learning architectures and configurations for the iris recognition problem.

becomes even slightly better than the real-valued version. Such improvement can be possibly attributed to the removal of less reliable features and relatively higher tolerance to noise in the binary feature representation. The threshold for selecting the fragile features, 0.6 as mentioned earlier, is determined from the feature value distribution and by extensive tuning.

5. Experiments and Results

Thorough experiments were conducted to evaluate the performance of the proposed approach from various aspects. The following sections detail the experimental settings along with the reproducible [42] results.

5.1. Databases and Protocols

We employed the following four publicly available databases our experiments:

- ND-IRIS-0405 Iris Image Dataset (ICE 2006)

This database [41] contains 64,980 iris samples from 356 subjects and is one of the most popular iris databases in the literature. The training set for this database is composed of the first 25 left eye images from all the subjects, and the test set consists of first 10 right eye images from all the subjects. The test

set, after removing some falsely segmented samples, contains 14,791 genuine pairs and 5,743,130 imposter pairs.

- CASIA Iris Image Database V4 distance

This database (subset) [43] includes 2,446 samples from 142 subjects. Each sample captures the upper part of face and therefore contain both left and right irises. The images were acquired from 3 meters away. An OpenCV-implemented eye detector [36] was applied to crop the eye regions from the original images. The training set consists of all the right eye images from all the subjects, and the test set comprises all the left eye images. The test set generates 20,702 genuine pairs and 2,969,533 imposter pairs.

- IITD Iris Database

The IITD database [44] contains 2,240 image samples from 224 subjects. All of the right eye iris images were used as training set while the first five left eye images were used as test set. The test set contains 2,240 genuine pairs and 624,400 imposter pairs.

- WVU Non-ideal Iris Database Release 1

The WVU Non-ideal database [45] (Rel1 subset) comprises 3,043 iris samples from 231 subjects which were acquired under different extends of off-angle, illumination change, occlusions, etc. The training set consists of all of the right eye images, and the test set was formed by the first five left eye images from all the subjects. The test set has 2,251 genuine pairs and 643,565 imposter pairs.

From the above introduction we can observe that the imaging conditions for these databases are quite different. Sample images from the four employed datasets are provided in Fig.9, where noticeable variation in image quality can be observed. It is therefore judicious to assume that these databases can represent diverse deployment environments.

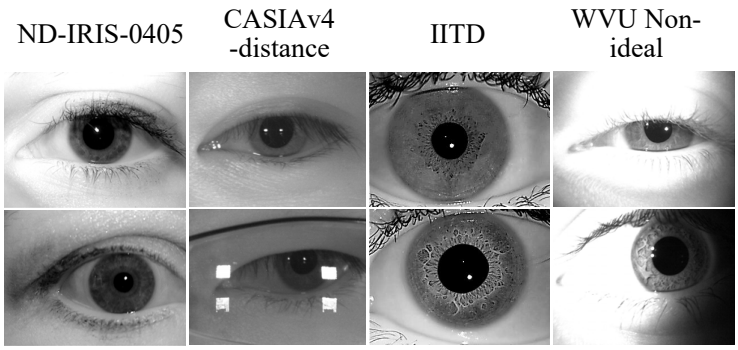


Figure 9: Sample raw images from four employed databases

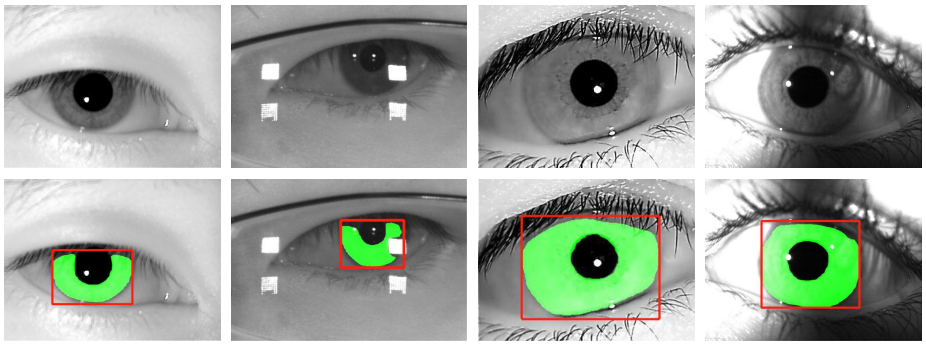


Figure 10: Sample results for the iris bounding box detection and mask segmentation from the proposed UniNet.v2.

5.2. Detection and Segmentation Accuracy

As the unified framework incorporating iris detection and segmentation is the key extension for our work compared with the conference version, we firstly evaluate the performance for this part. The iris detection and segmentation accuracy is of vital importance for the task of iris recognition as discussed earlier. Sample results of the iris bounding box detection and mask segmentation are provided in Fig. 10, from which it can be observed that the proposed framework can generalize well for varying image qualities. In this section we mainly compared our work with its earlier version, i.e., UniNet.v1 [16], and a recent promising work [17].

- Detection Accuracy

Table 2: Comparison of correct rates of iris detection obtained from this approach and a competitive baseline

	ND_IRIS_0405	Casia.v4	IITD	WVU Non-ideal
UniNet.v2	94.4%	96%	96.8%	89.6%
RTV-L^1 [17]	92.8%	92%	96%	85.6%

The term detection accuracy here refers to the accuracy in automatically detecting the iris and pupil circle positions compared with manually labeled ground truths. To generate ground truth circle positions, we randomly selected 500 sample images from test sets of the four employed databases which do not have overlapping subjects with the training sets. Then we manually labeled the positions of both pupil and iris circles as the ground truths. Let us represent a circle as $C = \{x, y, r\}$ where x and y are coordinates of the center and r is the radius, and assume we have a automatically detected circle C_d and the ground truth circle C_g. The detection is considered as accurate if the following conditions persist:

$$\begin{cases} \dfrac{\sqrt{(x_d - x_g)^2 + (y_d - y_g)^2}}{r_g} < 5\%, \\ \dfrac{|r_d - r_g|}{r_g} < 10\% \end{cases} \tag{12}$$

which considers the distance between two centers and the difference between the radii. An iris is then considered correctly detected if both iris and pupil circles are accurately found. Table 2 shows the comparison of iris detection accuracy from UniNet.v2 and a recent method [17] which appeared in the conference version of this paper.

The comparative results shown in Table 2 indicate that consistent improvements on the iris detection accuracy have been achieved by exploiting Mask R-CNN in place of parameter-dependent hand-crafted approach [17]. Note that the results of our approach is obtained from one model without fine-tuning whereas the parameters of [17] have been extensively tuned for each of the employed database separately. Therefore we can infer that our new model offers

superior generalization capability, which has been the key motivation for the work in this paper.

- Segmentation Accuracy

Apart from detection accuracy which evaluates the correctness of the location of detected iris, we should also examine the segmentation accuracy that measures pixel-level precision for the iris mask. Manually labeled ground truth masks are necessary for such evaluation. The IRISSEG-EP [46] has provided manually labeled iris masks for part of images from ND-IRIS-0405 and IITD databases, and another research work [19] has released ground truth masks for the Casia.v4-distance dataset. We utilize these masks as ground truths for the evaluation. After removing duplicated samples in the training set for training our model, we obtain 819, 1,890 and 437 ground truth masks for the images from ND-IRIS-0405, IITD and Casia.v4-distance databases respectively.

The segmentation accuracy is evaluated using the NICE.I protocol [47] which is widely adopted in the literature:

$$\begin{aligned} E &= \frac{1}{N_s}\sum_{i=1}^{N_s} E_i \\ E_i &= \frac{1}{W_i \times H_i}\sum_{x=1}^{W_i}\sum_{y=1}^{H_i} O_i[x,y] \oplus G_i[x,y] \end{aligned} \tag{13}$$

where O_i and G_i are the output binary mask from the algorithm and the ground truth binary mask respectively for the i-th sample, and with a size of $W_i \times H_i$, while $\oplus$ denotes the exclusive-or operation. The above formulation evaluates the number of inconsistent pixels between the predicted and ground truth masks, and normalized by the resolution of the image as the segmentation error rate. We compare the results using this metric (13) from our approach with that of [17] and our conference version, MaskNet [16] The results are shown in Table 3.

As shown from the segmentation results, iris masks generated from our framework using Mask R-CNN achieve consistently higher accuracy as compared with those from hand-crafted or post-normalization segmentation approaches.

Table 3: Comparison of segmentation error rates from different approaches

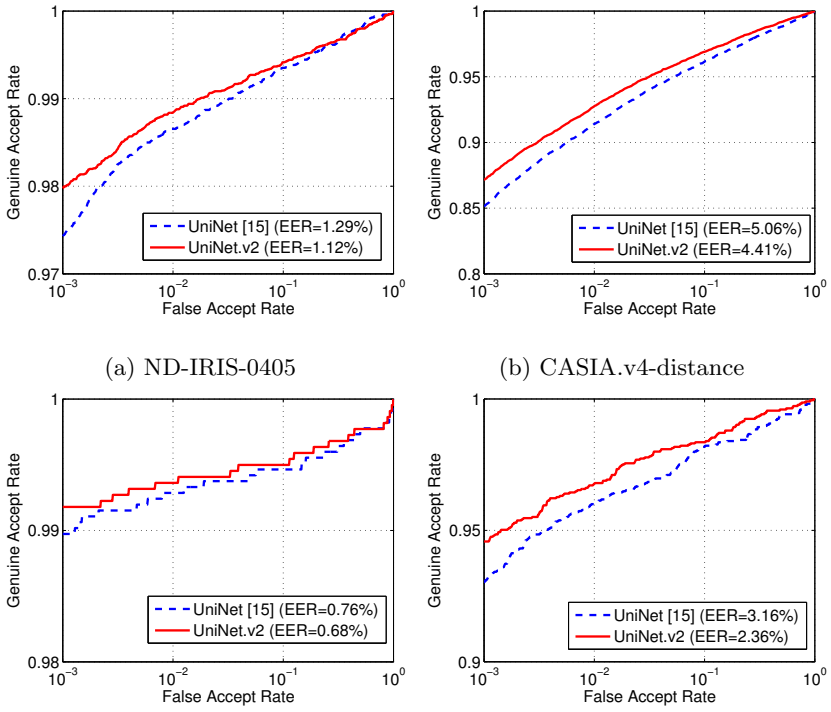
	ND_IRIS_0405	Casia.v4	IITD
UniNet.v2	1.68%	0.67%	5.34%
UniNet [16]	1.74%	0.83%	6.63%
RTV-L^1 [17]	1.93%	0.70%	5.89%

Such observation underlines the usefulness of Mask R-CNN is addressing problem of iris segmentation with superior generalization capability. It is again important to note that there is no dataset-specific parameter tuning for our approach for the automated iris segmentation.

5.3. Ablation Study for Recognition

In this section we will compare the recognition accuracy of UniNet.v2 with the results in the conference version, to investigate possible performance improvements by incorporating the Mask R-CNN for unified iris detection and segmentation. However, it is important to underline a key step to ensure fair comparison. In our conference paper, as the key focus was on learning effective iris features for the matching only, we *manually* removed some samples with badly detected iris circles, or corrected some of the detection results from both training set and test set to avoid learning meaningless information. This setup has been stated in the original paper [16] and such manual filtering for test images was identically performed for each baseline to ensure fairness in the comparison. In the experiment presented in this section, however, as automated iris detection is also part of our new framework, we skip such human intervention on the iris detection results to eliminate bias on the earlier version [16]. Comparative receiver operating characteristic (ROC) curves for the matching are shown in Fig. 11.

It can be observed from the ROCs shown in Fig. 11 that the matching accuracy has been consistently improved as compared with the results in conference version of this paper. Our work has implemented fully automated iris



(a) ND-IRIS-0405 (b) CASIA.v4-distance

(c) IITD (d) WVU Non-ideal

Figure 11: ROCs for comparison with the conference version of this paper [16] *Best viewed in color.*

segmentation with no human intervention, and results have ensured fairness in the comparison. Such results indicate that the addition of iris detection and segmentation module using Mask R-CNN offers encouraging usefulness on the recognition performance.

5.4. Comparison with Earlier Works

In this section we present comparative experimental results using several earlier and highly competitive baselines.

5.4.1. Test Configurations

During the comparison, we incorporated following two configurations in the test phase for extensive evaluation.

- CrossDB

In the CrossDB configuration, we use the ND-IRIS-0405 as the training set. During testing, the trained model was directly applied on CASIA.v4-distance and IITD without any further tuning. The purpose of the CrossDB setting is to examine the generalization capability of the proposed framework under challenging scenario that few training samples are available.

- WithinDB

In this configuration we use the network trained on ND-IRIS-0405 as the initial model, then fine-tune it using the independent training set from the target database. The fine-tuned network is then evaluated on the respective test set. Being capable of learning from data is the key advantage of deep learning, therefore it is judicious to examine the best possible performance from the proposed model by fine-tuning it with some samples from the target database. The fine-tuned models from the WithinDB configuration are expected to perform better than the one with CrossDB, due to higher consistency of image quality between the training set and test set.

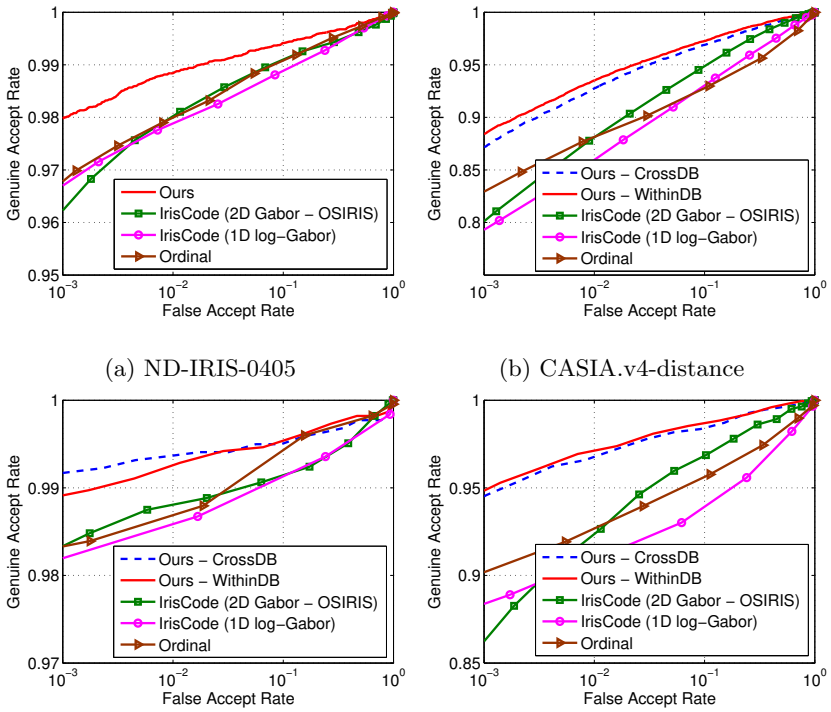
It should be noted that in both of the above configurations, training set and test set are totally separated, i.e., none of the iris images are overlapping between

the training set and test set. All the experimental results were generated under all-to-all matching protocol, i.e., the scores of every image pair in the test set have been counted.

5.4.2. Comparative Results

We employ several highly competitive baselines for the comparison. Gabor filter based IrisCode [5] has been the most widely deployed iris feature descriptor, largely due to the fact that few alternative iris features in the literature are universally accepted as better than IrisCodes. Instead, the majority of recent works on iris biometrics are more on improving segmentation and/or normalization models [17, 48], applying multi-score fusion [29] or feature bits selection [40]. In other words, in the context of iris feature representations, IrisCode is still the most popular and highly competitive approach, and therefore is definitely a fair benchmark for the performance evaluation. IrisCode has a number of advanced versions. From the publicly available ones, we selected OSIRIS [49], which is an open source tool for iris recognition. It implements a band of multiple tunable 2D Gabor filters that can encode iris patterns at different scales, therefore is a highly credible competitor. Another classic implementation of IrisCode is based on 1D log-Gabor filter(s) [6], which is claimed to encode iris patterns more efficiently, and is also widely chosen as benchmark in a variety of research works (e.g., [10, 17]). Therefore, this approach is also investigated. Apart from the Gabor series filters, ordinal filters proposed in [8] can serve as a different type of iris feature extractors to complement the comparisons. The aforementioned benchmarks have been extensively tuned on target databases during testing to ensure as good performance as possible.

The comparison results are shown in Fig. 12 and Table 4. Significant and consistent improvements from our deep learning based methods over other baselines have been shown on all of the four databases, under both WithinDB and CrossDB configurations. Such results suggest that the proposed iris feature representation not only achieves superior accuracy but also exhibits outstanding generalization capability. Even without additional parameter tuning, the well-



(a) ND-IRIS-0405 (b) CASIA.v4-distance

(c) IITD (d) WVU Non-ideal

Figure 12: ROCs for comparison with other state-of-the-art methods on for the employed databases. *Best viewed in color.*

trained model from our framework is promising to be directly used in deployment environments with varying image qualities. The relaxation of parameter tuning is apparently a highly desirable property for many real-life applications. An interesting finding is that on IITD database, the CrossDB model performs better even than the fine-tuned one. This is possibly because most of the images in IITD are with high qualities and less challenging, and its training set is not large enough, which causes slight over-fitting problem.

5.5. Comparison with Other Deep Learning Configurations

In order to ascertain the effectiveness of the proposed network architecture for spatial feature extraction and the extended triplet loss, we also compared our method against typical deep learning architectures that are widely employed

Table 4: Summary of false reject rates (FRR) at 0.1% false accept rate (FAR) and equal error rates (EER) for the comparison.

	ND-IRIS-0405		**CASIA.v4-distance**		**IITD**		**WVU Non-ideal**	
	FRR	**EER**	**FRR**	**EER**	**FRR**	**EER**	**FRR**	**EER**
IrisCode (OSIRIS)	3.73%	1.70%	19.93%	6.39%	1.61%	1.11%	13.70%	4.43%
IrisCode (log-Gabor)	3.31%	1.88%	20.72%	7.71%	1.81%	1.38%	11.63%	6.82%
Ordinal	3.22%	1.74%	16.93%	7.89%	1.70%	1.25%	9.89%	5.19%
Ours-CrossDB	/	/	12.83%	4.41%	**0.82%**	**0.68%**	5.42%	2.36%
Ours-WithinDB	**2.02%**	**1.12%**	**11.61%**	**4.07%**	1.12%	0.76%	**5.06%**	**2.20%**

in various recognition tasks. The tested configurations are introduced in the following:

(a) CNN+softmax/triplet loss

CNN+softmax is the most widely employed deep learning configurations in the community, such as in [12] and [32]. Besides, CNN+triplet loss is gaining increasing popularity after it was proposed in [39], and therefore may also be interesting and worth evaluating. For the CNN model, we have chosen the popular VGG-16 which has achieved superior performance in face recognition.

(b) FCN+triplet loss

Comparative evaluation has also been performed on using the proposed FCN (FeatNet only) and the original triplet loss function without incorporating bit-shifting and masking. Such comparison may assert the necessity of extending the original triplet loss.

(c) DeepIrisNet [23]

We also compared our method against the recent deep learning based iris recognition framework, DeepIrisNet, which reports promising results. This architecture actually belongs to the CNN+softmax category, but we separately inspected it as it is directly proposed for iris recognition. Since the original model

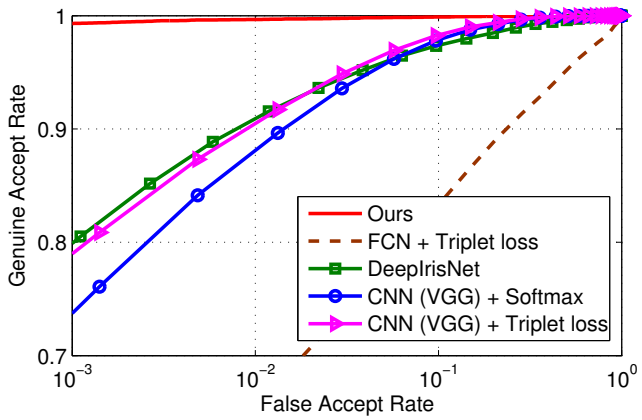


Figure 13: Comparison of ROCs from diverse deep learning architectures and configurations for the iris recognition problem.

in their paper is not publicly available, we carefully implemented and trained the CNN according to all the details in [23].

The comparison with aforementioned configurations was performed on ND-IRIS-0405 dataset, which has the largest number of training images among employed ones. The test set is kept consistent during the comparison. Hyper-parameters of the training processes for above architectures have been carefully investigated to achieve best possible performance. It should be noted that the iris segmentation and normalization procedures were made exactly the same for the different configurations introduced above, and therefore the only factor impacting the final recognition accuracy is the performance of the extracted features from these networks. The results are presented in Fig.13.

It can be observed from Fig.13 that our newly developed architecture significantly outperforms other deep learning configurations. CNN based configurations have failed to deliver satisfactory results especially at lower FAR. Such results support our previous analysis that global and high level features extracted by CNN may not be suitable for iris recognition. The poor performance from FCN+triplet loss strongly suggests that it is necessary to account for bit-shifting and non-iris region masking when learning spatially corresponding features through FCN.

Table 5: Execution time of the proposed framework and other methods per eye image. The resolutions for the original image and normalized template are 640×480 and 512×64 respectively.

	Detection/Segmentation Time		Feature Extration Time	
	GPU	CPU	GPU	CPU
UniNet.v2	271ms	4.3s	5.9ms	193ms
RTV-L^1 [17]	/	763ms	/	/
OSIRIS [49]	/	93ms	/	17ms

5.6. Execution Speed

We also evaluated the computational efficiency of the proposed framework by measuring the execution time. The programs were executed on a desktop computer with Intel Core 3.4GHz i7-4770 CPU, 16GB RAM and NVIDIA GTX 1080 GPU. The summary is provided in Table 5.

It can be observed from the summary that the overall execution time of the proposed framework is within reasonable range. It should be note that: (a) in practical systems, the iris detection, segmentation and feature extraction are one-time effort for the on-line probe sample and template generation for the gallery subjects can be done off-line in the registration process, and (b) the generated iris templates from the proposed framework are in the same format as those from conventional methods, e.g., iriscodes. Consequently, there is no additional computational cost for matching/searching iris templates for the probe to gallery samples compared with existing approaches.

6. Conclusions

This paper has developed a novel deep learning based framework, referred to as UniNet.v2, for effective iris detection, segmentation and recognition, which can offer superior performance and generalization capability for the focused problems. Higher iris detection and segmentation accuracy has been achieved by introducing Mask R-CNN compared with the earlier version of UniNet. As

for the feature learning, the specially designed Extended Triplet Loss function can provide effective supervision for learning comprehensive and spatially corresponding iris features through the fully convolutional network. Further extension of this work should focus on developing more effective algorithms for the simultaneous optimization for the iris segmentation and feature learning processes through the deep networks, including more advanced and backpropagation-complaint iris contour fitting and normalization, which is expected to further exploit the spatially corresponding features for the problem of iris recognition.

References

[1] J. Daugman, Iris recognition border-crossing system in the uae, International Airport Review 8 (2).

[2] J. Daugman, 600 million citizens of india are now enrolled with biometric id, SPIE newsroom 7.

[3] Nist presentation: Forensic data for face & iris, `http://biometrics.nist.gov/cs_links/standard/ansi-overview_2010/presentations/Forensic_data_for_Face_Iris.pdf`.

[4] CNNMoney, Galaxy note 7 is first samsung device with iris scanner, `http://money.cnn.com/2016/08/02/technology/samsung-note-7/index.html`.

[5] J. Daugman, How iris recognition works, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY 14 (1) (2004) 21.

[6] L. Masek, et al., Recognition of human iris patterns for biometric identification.

[7] D. M. Monro, S. Rakshit, D. Zhang, Dct-based iris recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (4) (2007) 586–595.

[8] Z. Sun, T. Tan, Ordinal measures for iris recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (12) (2009) 2211–2226.

[9] K. Miyazawa, K. Ito, T. Aoki, K. Kobayashi, H. Nakajima, An effective approach for iris recognition using phase-based image matching, IEEE transactions on pattern analysis and machine intelligence 30 (10) (2008) 1741–1756.

[10] J. K. Pillai, V. M. Patel, R. Chellappa, N. K. Ratha, Secure and robust iris recognition using random projections and sparse representations, IEEE transactions on pattern analysis and machine intelligence 33 (9) (2011) 1877–1893.

[11] K. Wang, A. Kumar, Cross-spectral iris recognition using cnn and supervised discrete hashing, Pattern Recognition 86 (2019) 85–98.

[12] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[13] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580–587.

[14] E. Shelhamer, J. Long, T. Darrell, Fully convolutional networks for semantic segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (4) (2017) 640–651.

[15] Y. Sun, X. Wang, X. Tang, Deeply learned face representations are sparse, selective, and robust, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[16] Z. Zhao, A. Kumar, Towards more accurate iris recognition using deeply learned spatially corresponding features, in: Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 2017, pp. 22–29.

[17] Z. Zhao, A. Kumar, An accurate iris segmentation framework under relaxed imaging constraints using total variation model, in: Computer Vision (ICCV), 2015 IEEE International Conference on, IEEE, 2015, pp. 3828–3836.

[18] T. Tan, Z. He, Z. Sun, Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition, Image and vision computing 28 (2) (2010) 223–230.

[19] C.-W. Tan, A. Kumar, Towards online iris and periocular recognition under relaxed imaging constraints, IEEE Transactions on Image Processing 22 (10) (2013) 3751–3765.

[20] L. Grady, Random walks for image segmentation, IEEE transactions on pattern analysis and machine intelligence 28 (11) (2006) 1768–1783.

[21] M. Frucci, M. Nappi, D. Riccio, G. S. di Baja, Wire: Watershed based iris recognition, Pattern Recognition 52 (2016) 148–159.

[22] D. Menotti, G. Chiachia, A. Pinto, W. R. Schwartz, H. Pedrini, A. X. Falcão, A. Rocha, Deep representations for iris, face, and fingerprint spoofing detection, IEEE Transactions on Information Forensics and Security 10 (4) (2015) 864–879.

[23] A. Gangwar, A. Joshi, Deepirisnet: Deep iris representation with applications in iris recognition and cross-sensor iris recognition, in: Image Processing (ICIP), 2016 IEEE International Conference on, IEEE, 2016, pp. 2301–2305.

[24] F. He, Y. Han, H. Wang, J. Ji, Y. Liu, Z. Ma, Deep learning architecture for iris recognition based on optimal gabor filters and deep belief network, Journal of Electronic Imaging 26 (2) (2017) 023005.

[25] K. W. Bowyer, M. J. Burge, Handbook of iris recognition, Springer, 2016.

[26] G. W. Quinn, P. J. Grother, M. L. Ngan, J. R. Matey, Irex iv: part 1, evaluation of iris identification algorithms, Tech. rep. (2013).

[27] I. O. for Standardization, Information Technology, Biometric Data Interchange Formats, ISO/IEC, 2005.

[28] Z. He, Z. Sun, T. Tan, X. Qiu, C. Zhong, W. Dong, Boosting ordinal features for accurate and fast iris recognition, in: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, IEEE, 2008, pp. 1–8.

[29] A. Kumar, A. Passi, Comparison and combination of iris matchers for reliable personal authentication, Pattern recognition 43 (3) (2010) 1016–1026.

[30] Y.-H. Li, M. Savvides, An automatic iris occlusion estimation method based on high-dimensional density estimation, IEEE transactions on pattern analysis and machine intelligence 35 (4) (2013) 784–796.

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

[32] Y. Sun, X. Wang, X. Tang, Deep learning face representation from predicting 10,000 classes, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1891–1898.

[33] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, in: Computer Vision (ICCV), 2017 IEEE International Conference on, IEEE, 2017, pp. 2980–2988.

[34] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: towards real-time object detection with region proposal networks, IEEE transactions on pattern analysis and machine intelligence 39 (6) (2017) 1137–1149.

[35] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[36] Coco challenges, https://places-coco2017.github.io/.

[37] W. Luo, Y. Li, R. Urtasun, R. Zemel, Understanding the effective receptive field in deep convolutional neural networks, in: Advances in neural information processing systems, 2016, pp. 4898–4906.

[38] L. Ma, Y. Wang, T. Tan, Iris recognition based on multichannel gabor filtering, in: Proc. Fifth Asian Conf. Computer Vision, Vol. 1, 2002, pp. 279–283.

[39] F. Schroff, D. Kalenichenko, 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.

[40] K. P. Hollingsworth, K. W. Bowyer, P. J. Flynn, The best bits in an iris code, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (6) (2009) 964–973.

[41] K. W. Bowyer, P. J. Flynn, The nd-iris-0405 iris image dataset, Notre Dame CVRL Technical Report.

[42] Web link to download the source code and executable files for the approach detailed in this paper, http://www.comp.polyu.edu.hk/~csajaykr/deepiris.htm.

[43] Casia.v4 iris database, http://biometrics.idealtest.org/dbDetailForUser.do?id=4.

[44] Iitd iris database, http://www.comp.polyu.edu.hk/~csajaykr/IITD/Database_Iris.htm.

[45] S. Crihalmeanu, A. Ross, S. Schuckers, L. Hornak, A protocol for multi-
biometric data acquisition, storage and dissemination, Technical Report, WVU, Lane Department of Computer Science and Electrical Engineering.

[46] H. Hofbauer, F. Alonso-Fernandez, P. Wild, J. Bigun, A. Uhl, A ground truth for iris segmentation, in: Pattern Recognition (ICPR), 2014 22nd International Conference on, IEEE, 2014, pp. 527–532.

[47] Noisy iris challenge evaluation - part i, `http://nice1.di.ubi.pt/index.html`.

[48] H. Proenca, Iris recognition: On the segmentation of degraded images acquired in the visible wavelength, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (8) (2010) 1502–1516.

[49] N. Othman, B. Dorizzi, S. Garcia-Salicetti, Osiris: An open source iris recognition software, Pattern Recognition Letters 82 (2016) 124–131.