

Towards Complete and Accurate Iris Segmentation Using Deep Multi-Task Attention Network for Non-Cooperative Iris Recognition

Caiyong Wang^{ID}, Jawad Muhammad, Yunlong Wang^{ID}, Zhaofeng He^{ID}, and Zhenan Sun^{ID}, *Senior Member, IEEE*

Abstract—Iris images captured in non-cooperative environments often suffer from adverse noise, which challenges many existing iris segmentation methods. To address this problem, this paper proposes a high-efficiency deep learning based iris segmentation approach, named *IrisParseNet*. Different from many previous CNN-based iris segmentation methods, which only focus on predicting accurate iris masks by following popular semantic segmentation frameworks, the proposed approach is a complete iris segmentation solution, i.e., iris mask and parameterized inner and outer iris boundaries are jointly achieved by actively modeling them into a unified multi-task network. Moreover, an elaborately designed attention module is incorporated into it to improve the segmentation performance. To train and evaluate the proposed approach, we manually label three representative and challenging iris databases, i.e., CASIA.v4-distance, UBIRIS.v2, and MICHE-I, which involve multiple illumination (NIR, VIS) and imaging sensors (long-range and mobile iris cameras), along with various types of noises. Additionally, several unified evaluation protocols are built for fair comparisons. Extensive experiments are conducted on these newly annotated databases, and results show that the proposed approach achieves state-of-the-art performance on various benchmarks. Further, as a general drop-in replacement, the proposed iris segmentation method can be used for any iris recognition methodology, and would significantly improve the performance of non-cooperative iris recognition.

Index Terms—Iris segmentation, iris localization, attention mechanism, multi-task learning, iris recognition.

Manuscript received November 8, 2019; revised February 21, 2020; accepted March 7, 2020. Date of publication March 16, 2020; date of current version March 27, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant U1836217 and Grant 61427811 and in part by the National Key Research and Development Program of China under Grant 2017YFC0821602. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Vitomir Štruc. (*Corresponding author: Zhenan Sun*.)

Caiyong Wang and Jawad Muhammad are with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China, and also with the Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: caiyong.wang@cripac.ia.ac.cn; jawad@cripac.ia.ac.cn).

Yunlong Wang is with the Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: yunlong.wang@cripac.ia.ac.cn).

Zhaofeng He is with Beijing IrisKing Company Ltd., Beijing 100080, China (e-mail: hezhf@irisking.com).

Zhenan Sun is with the Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: znsun@nlpr.ia.ac.cn).

Digital Object Identifier 10.1109/TIFS.2020.2980791

I. INTRODUCTION

IR IS recognition has been considered as one of the most stable, accurate and reliable biometric identification technologies [1], hence it is widely applied in various biometric applications, including intelligent unlocking [2], border control [3], forensics [4], etc. A complete iris recognition system often consists of four sub-processes: iris image acquisition, iris pre-processing, feature extraction and matching. As part of iris pre-processing, iris segmentation defines the image region that is used for feature extraction and matching, hence directly affects the overall iris recognition performance [5].

Initially, iris segmentation algorithms such as Daugman's integro-differential operator [6] and Wilde's circular Hough transforms [7] were developed for controlled iris biometric set-up (e.g., close acquisition distance, stop-and-stare verification, high user cooperation, near infrared illumination). But with the increasing demands of iris recognition in our daily life, it has since been extended to non-cooperative environments (e.g. at-a-distance, on-the-move, limited user cooperation, visible illumination, using mobile devices). This has emerged as a hot research topic as it requires fewer constraints for user cooperation and imaging conditions. However, under these conditions, captured iris images often suffer from all kinds of noise factors, such as gaze deviation, iris rotation, absence of iris, specular reflections, motion/defocus blur, occlusions due to eyelid/eyelash/hair/glasses, as illustrated in Fig. 1, which make iris segmentation challenging.

To solve these problems, quite a lot of effort in the past decades has been made to improve the accuracy and robustness of iris segmentation. Approaches such as coarse iris clustering before applying integro-differential operator [8], Pulling and Pushing method [5], total-variation based iris denoise [9], geodesic active contours model [10], Polar Spline RANSAC and illumination compensation [11], and Graph Cut [12], [13] have been proposed. In addition to these traditional iris segmentation approaches, in recent years, popular deep learning technologies have also been widely used in iris segmentation and achieved new state-of-the-art performance in terms of iris mask error metrics, especially for highly non-ideal iris images [14]–[18]. However, there is one common problem with deep learning methods, related with the localization of iris boundaries. Usually, in the general iris recognition systems, the segmented iris has to be normalized in polar domain to decrease the influence of iris size and pupillary dilation on

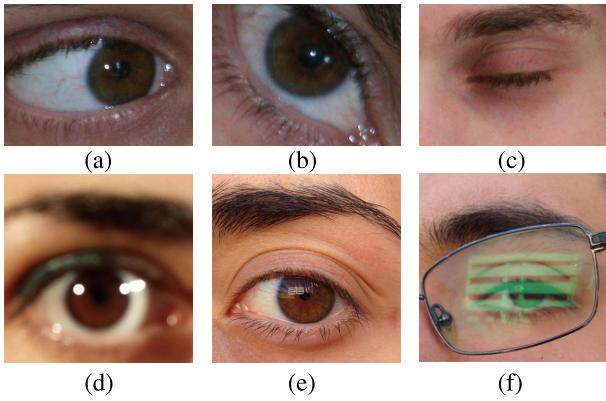


Fig. 1. Examples of degraded iris images with different types of noises. (a) gaze deviation, (b) iris rotation, (c) absence of iris, (d) blur, (e) specular reflections, and (f) glasses occlusion.

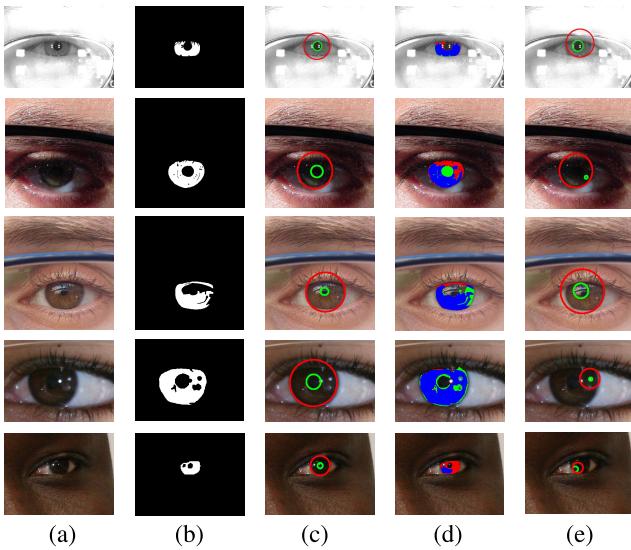


Fig. 2. Samples of incorrect segmentation results using CNNHT [20]. (a) Original iris images from CASIA.v4-distance (top one), MICHE-I (middle two), and UBIRIS.v2 (bottom two) iris databases, (b) ground truth iris mask, (c) ground truth iris boundaries (inner boundary (green), and outer boundary (red)), (d) segmentation result of CNNHT [20] (false positive error pixel (green), false negative error pixel (red), and true positive pixel (blue)), and (e) localization results of CNNHT [20] (inner boundary (green), and outer boundary (red)). Best viewed in color and with magnification.

feature extraction and matching, hence iris normalization is more-or-less an essential part of the whole iris recognition system [19]. But many of existing deep learning based iris segmentation methods only produce iris masks, not yielding the parameterized inner and outer iris boundaries, which means that original iris images and detected iris masks cannot be normalized for subsequent processes. Although some recent methods have proposed to impose circular Hough transforms on iris masks to obtain circular iris boundaries [20]–[22], the offered localization accuracy is still below the standard required for iris recognition.

In summary, in the general iris recognition systems, a complete iris segmentation should involve two parts: (1) pixel-level identification of iris and non-iris regions (*e.g.*, excluding eyelashes, eyelids, hair, glasses, reflections, shadows), which produces a binary iris mask, as illustrated in Fig. 2 (b), and (2) obtaining the parameterized (*e.g.*, circle,

ellipse, active contour, spline) inner and outer iris boundaries, as illustrated in Fig. 2 (c). In order to be consistent with the definition of general semantic segmentation problems, we will narrowly refer to the first part as *iris segmentation*, and the second part as *iris localization*. In this paper, we propose a simple yet effective multi-task attention network to simultaneously realize accurate iris segmentation and localization.

A. Related Work

Usually, in most of traditional iris segmentation methods, inner and outer iris boundaries are first detected and then refined iris masks are obtained by further locating the upper and lower eyelids, and excluding any superimposed occlusions of eyelashes, shadows, glasses, or reflections [23], which means *iris localization* is first, and narrowly-defined *iris segmentation* comes after.

For iris localization, many of the methods propose the approximation of iris boundaries as pre-defined geometric shapes, such as circles or ellipses. Two widely used baseline methods are Daugman's integro-differential operator [6] and Wilde's circular Hough transforms [7]. The integro-differential operator searched for the largest difference of intensity over the circle parameter space which normally corresponded to inner and outer iris boundaries, while circular Hough transforms found optimal circle parameters by a voting procedure in a binary edge image. Based on these baseline methods, many of the later proposed approaches made further improvements in accuracy, robustness, and efficiency, for example, RTV- L^1 [9] adopted a total-variation based formulation with L^1 norm regularization to robustly suppress noisy texture pixels before localization, [24] applied region clustering before localization for narrowing the parameter search range, [8] proposed integro-differential constellation to reduce the computation time of the integro-differential operator. Besides, in [25], the Viterbi algorithm was applied on gradient maps of iris images to find coarse low-resolution contours, which meant selecting as few noisy gradients points as possible, then least-squares circle fitting was used to obtain circular iris boundaries. In [5], a novel Pulling and Pushing iteration approach was proposed for robust and accurate iris localization. Considering that iris boundaries are not strictly circular, some works attempted to use more flexible curve shapes, such as elliptic contours [13], spline [11], and geodesic active contours [10] to approximate the actual iris boundaries.

In addition to obtain iris masks in the above indirect way, there are some pixel-based iris segmentation approaches to directly identify iris and non-iris regions. They usually exploit low-level visual information of individual pixel, such as intensity and color, to classify the pixels of interest from the background of the image. Specifically, promising methods usually use commonly known pixel-level techniques, such as Graph Cut [12], [13], to pre-process the image and then apply traditional classification methods, such as SVMs [26], to classify iris and non-iris pixels.

Recently, researchers have proposed deep learning models, especially Convolutional Neural Networks (CNNs), for iris segmentation. Similar to pixel-based approaches, deep learning based iris segmentation methods directly estimate

iris masks but utilize more high-level semantic features and automatically learn optimal features and classifiers in an end-to-end manner, hence higher segmentation accuracy is achieved. Two CNN-based iris segmentation models were first proposed by Liu *et al.* [14], namely hierarchical convolutional neural networks (HCNNs) and multi-stage fully convolutional networks (MFCNs). The HCNNs adopted the idea of image classification using CNNs, and took into consideration hierarchical patches around the pixel as input to classify iris and non-iris pixels. Unlike the HCNNs, the MFCNs were based on a Fully Convolutional Network (FCN), which took the entire iris image as input and then directly predicted the iris mask. Moreover, shallow and deep layers were fused to formulate a multi-scale network to better capture both the iris local texture and global structure. These made the MFCNs achieve state-of-the-art performance with the average segmentation error of 0.90% and 0.59% on UBIRIS.v2 and CASIA.v4-distance databases, respectively. After that, researchers have adopted off-the-shelf [20]–[22], [28], [29] or self-designed [16], [18], [30] FCN models for iris segmentation and obtained remarkable segmentation accuracy on different iris databases.

However, as stated previously, most of CNN-based iris segmentation methods only predict iris masks, but more important iris localization is not achieved, which causes the gap between CNN-based iris segmentation and Rubber-Sheet based iris normalization [27], and further affects the subsequent feature extraction and matching. To bridge this gap, CNNHT [20] first applied circular Hough transforms to iris masks to generate candidate circular iris boundaries, and then two quality measures were proposed to select the best inner and outer iris boundaries, respectively. Similar methods were also presented in [21], [22]. Although these methods are effective in normal cases, they often suffer from a serious performance drop when confronted with poorly segmented or highly irregular masks, as illustrated in Fig. 2. Therefore, a more accurate, robust, complete, and elegant iris segmentation solution is required. The reader is referred to [19] for a comprehensive survey of iris segmentation.

B. Motivation and This Work

Our work is motivated to further advance the complete iris segmentation using deep learning based approaches. To this end, the following three aspects are considered:

- 1) how to achieve iris localization without only relying on the iris mask?
- 2) how to obtain a simple yet effective baseline model?
- 3) how to take advantage of several spatial prior constraints to improve the segmentation performance?

For 1), from the former analysis, we can easily deduce that performing iris localization only by relying on iris masks is not optimal. Besides, it is extremely difficult to infer the pre-defined geometric shapes (*e.g.*, circle, ellipse) directly from the original iris images in an end-to-end regression model, due to the highly nonlinear characteristics in the model, and it may also be unreasonable, as some iris boundaries may deviate from predefined shapes. Therefore, an additional

intermediate iris localization representation is highly desirable. In fact, many related work has provided an insight. For instance, in [10], [11], to estimate iris inner circle, pupil area was separated from the original iris image via simple threshold segmentation, then circle fitting was performed on the extracted pupil boundary. In [25], the low-resolution iris contour was retrieved based on Viterbi algorithm, thus yielding a relatively complete iris outer boundary, which contained both occluded and non-occluded sampling points, then least-squares circle fitting was carried out for estimating iris outer circle. Based on the above inspiration, we propose to predict the pupil mask and iris outer boundary, and from there to fit the parameterized inner and outer iris boundaries. It should be noted that unlike the fine iris mask, pupil mask here is a coarse filled area of the iris inner boundary with any noise not considered, and iris outer boundary is a complete closed curve, as illustrated in Fig. 9.

For 2), according to 1), our prediction targets includes iris mask, pupil mask, and iris outer boundary, which have the same size as the input iris image. Hence, these three tasks could be regarded uniformly as a dense prediction task, which could be solved well by a FCN model. To accomplish this, we extend the binary output of the commonly used U-Net model [31] to three outputs, which correspond to our prediction targets, as shown in Fig. 4. We expect that our multi-task U-Net model could serve as a solid baseline and help ease further research in the complete iris segmentation.

For 3), there are some spatial prior constraints in the iris area, for instance, there is at most one iris area in an iris image, iris mask is surrounded by the inner and outer iris boundaries, the radius of iris inner boundary is no larger than 1/4th of the smallest image dimension, and the radius of iris outer boundary is greater than that of iris inner boundary [20]. In order to make full use of these constraints, on the one hand, our proposed multi-task U-Net model implicitly encodes the spatial relationship between the iris mask, pupil mask, and iris outer boundary by supervised learning, and an elaborately designed attention module further helps highlight the important signals about the prediction targets and suppresses the irrelevant noise signals, on the other hand, we explicitly utilize priori constraints to help improve post-processing and ensure the correctness of iris segmentation results, as presented in Section II-C, and Section III-D.

Main contributions of this paper are summarized as follows:

- 1) This paper proposes a simple yet effective multi-task network to simultaneously predict the iris mask, pupil mask and iris outer boundary. Further, based on the predicted pupil mask and iris outer boundary, parameterized inner and outer iris boundaries are achieved, which solves the problems of prediction inaccuracies and lacking robustness in only using the iris mask for iris localization. Moreover, attention mechanism is incorporated into the multi-task network to help improve the segmentation performance. Significant advancement therefore has been made to bridge the gap between deep learning and complete iris segmentation.

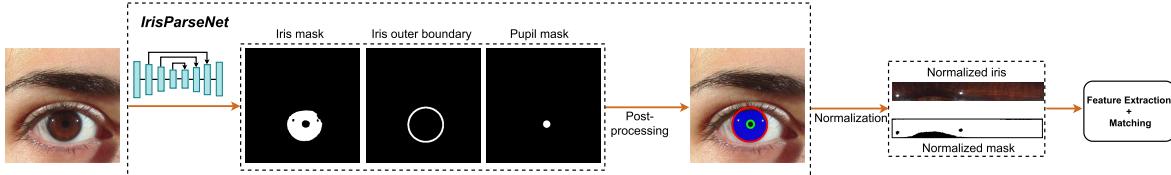


Fig. 3. Pipeline of a iris recognition system. Raw acquired iris image is taken as input and parsed by the proposed *IrisParseNet* for segmenting iris and pupil region pixels and extracting iris outer boundary, then followed by simple post-processing operations for obtaining the parameterized inner and outer iris boundaries. Daugman's rubber sheet model [27] is further applied to normalize the iris image and detected iris mask. Finally, any iris feature extraction and matching methods are employed to complete the identity identification (1:N) or verification (1:1).

- 2) Three representative and challenging iris databases, i.e., CASIA.v4-distance, UBIRIS.v2 and MICHE-I are selected and manually labeled with additional pupil mask, and iris outer boundary, to effectively train and evaluate the proposed approach. Besides, unified and comprehensive iris segmentation evaluation protocols are built for fair comparisons. Models, manual annotations and evaluation protocol codes will be made freely available to the community once requested.
- 3) The proposed method achieves state-of-the-art performance on various benchmarks. Moreover, it has strong robustness and generalization capability, providing a good foundation for non-cooperative iris recognition.

The paper is organized as follows: Our technical details are elaborated in Section II. Section III presents the experimental configurations, protocols, results and analysis. Finally, we conclude the paper and discuss future work in Section IV.

II. TECHNICAL DETAILS

In this section, we first introduce the whole pipeline of a iris recognition system, where the proposed deep learning framework, referred to as *IrisParseNet*, for complete iris segmentation. Then we elaborate on the details of the proposed approach, including a multi-task Fully Convolutional Network with attention mechanism assisted and simple yet effective post-processing process. Finally we describe our training objectives of the proposed model.

A. Pipeline

The pipeline of a complete iris recognition system is illustrated in Fig. 3, including iris segmentation, normalization, feature extraction and matching. Concretely, for iris segmentation, the proposed *IrisParseNet* model takes original iris image as input, then produces the iris mask, iris outer boundary, and pupil mask through the inference of the network. Besides, some simple yet effective post-processing operations are employed to estimate the parameterized inner and outer iris boundaries and exclude mispredicted results. It is evidently clear that, compared with previous CNN-based iris segmentation methods, our method can not only predict iris mask, but more importantly, can also predict parameterized iris boundaries, which are critical for subsequent iris normalization. To further decrease the influence of iris size and pupillary dilation on feature extraction and matching, the iris image and detected iris mask are normalized into a scale-invariant and pupil-dilation-invariant band with resolution of 512×64 pixels using the Daugman's Rubber-Sheet model [27].

Afterwards, discriminative iris features are extracted by using any effective methods such as traditional 2D Gabor filters [27], 1D log-Gabor filters [32], ordinal measures (OMs) [33], or deep learning based FeatNet [34], Maxout CNNs [35], and ultimately compared with that of iris template databases to complete the identity identification (1:N) or verification (1:1). Some of iris recognition experiments are presented in Section III-E3.

B. Multi-Task Attention Network

1) *Multi-Task U-Net Architecture*: Encouraged by the success of U-Net [31] in the biomedical image segmentation tasks, we employ U-Net for our iris segmentation task since iris is essentially a biological organ. The proposed network architecture is illustrated in Fig. 4. As can be observed, it consists of a contracting path and an expansive path. The contracting path adopts VGG-16 [36] as the encoding network with fully connected layers discarded. Specifically, it consists of five stages and each stage is composed of two or three 3×3 convolutions, Batch Normalization and ReLU layers. For the first four stages, each stage is followed by a 2×2 max pooling operation with stride 2 for downsampling. Overall, as the network goes deeper, the number of feature channels gradually increases meanwhile the size of feature maps gradually decreases. Furthermore, in lower stages of the network, feature maps capture more low-level spatial information such as edges due to the small receptive fields, while in higher stages, bigger receptive fields help extract more semantic information and embed it in feature maps.

As for the expansive path, every step consists of a 2×2 bilinear upsampling of the feature map, two 3×3 convolutions, each followed by a Batch Normalization and ReLU layer, and a skip connection (by concatenation) with the corresponding feature map from the contracting path of the network. It can be observed that, by multi-step operations, the expansive path gradually up-samples feature maps to the original size and reduces the feature channels. Skip connections further fuse the detailed spatial features from the lower stages and the rich contextual semantic features from the higher stages. Moreover, during the training of the network, skip connections also facilitate the fast propagation of gradients and accelerate the convergence of the network.

After fusing the feature maps of the final stage, we apply a sequence of 3×3 convolution with 32 filters, Batch Normalization and ReLU to summarize the final semantic feature. Next, a 1×1 convolutional layer with 3 filters and a per-pixel sigmoid function is adopted to map each 32-component feature vector to the desired number of classes (i.e., 3) and ultimately

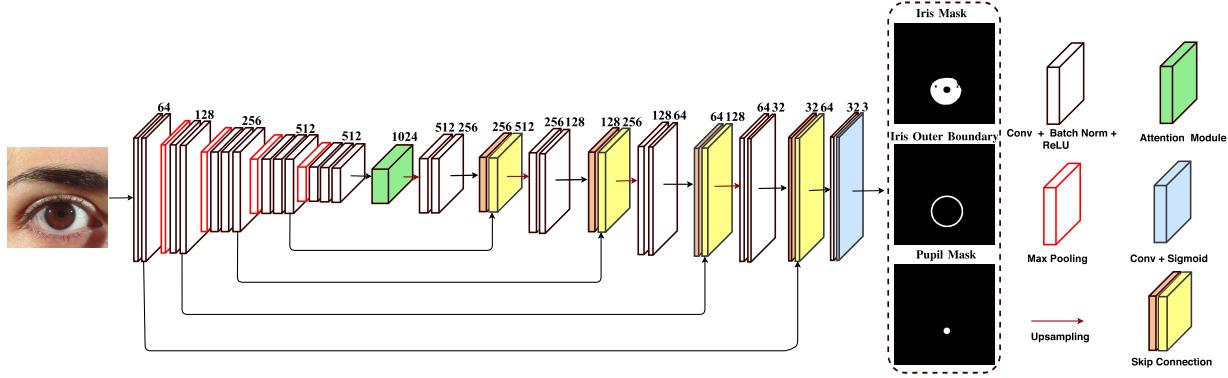


Fig. 4. Overview of the proposed multi-task attention network architecture, where each box corresponds to a series of network operations and the number of feature channel is annotated at the top of each box. This network adopts a symmetry encoder-decoder structure, which is also referred to as contracting path and expansive path, to simultaneously predict the iris mask, iris outer boundary, and pupil mask. To further focus on the more important information for prediction targets and suppress the irrelevant noise, we propose an attention module (highlighted in light green), which is embedded into the central bottleneck part between the contracting path and the expansive path.

generate probability maps of iris mask, iris outer boundary, and pupil mask, simultaneously.

It may be noticed that there is a central bottleneck part between the contracting path and the expansive path. During the forward propagation of the network, representative features collected from the contracting path propagate to the subsequent expansive path via the central bottleneck part. Similarly, during the backward propagation of the network, gradients from the expansive path also propagate to the former contracting path via it according to the chain rule. Therefore, the central bottleneck part has a critical impact on the optimization and prediction performance of the network.

In order to enhance the feature representation capabilities of the central bottleneck part, and further focus on the more important information for prediction targets and suppress the distracting noise, instead of simple convolutional units used in the original U-Net, we embed an attention module (highlighted in light green in Fig. 4) into the central bottleneck part. Attention mechanism allows us to adjust the weights of different channels in feature maps and also re-estimates the spatial distribution of feature maps according to [37], [38]. Hence, more discriminative features could be learned. The details of the proposed attention module are introduced in the following section.

2) *Attention Module*: In the proposed attention module, we first adopt Atrous Spatial Pyramid Pooling (ASPP) module introduced in DeepLab V3 [39] to extract multi-scale contextual features, then under the framework of attention mechanism, important feature signals for prediction targets are highlighted, and irrelevant noise signals are simultaneously suppressed. The detailed structure is illustrated in Fig. 5.

Specifically, given the final feature map F in the contracting path as input, a 3×3 pooling layer with stride 1 is first used to get the same sized refined feature map P as the new input. Then, five parallel modules are used, including one 1×1 convolution with 256 filters (as in Eq. (1)), three dilated convolution with 256 filters and dilation rate set to 6, 12, 18, respectively (as in Eq. (2)-Eq. (4)), and one global average pooling layer followed by one 1×1 convolution with 256 filters and a upsampling layer, mapping the feature map

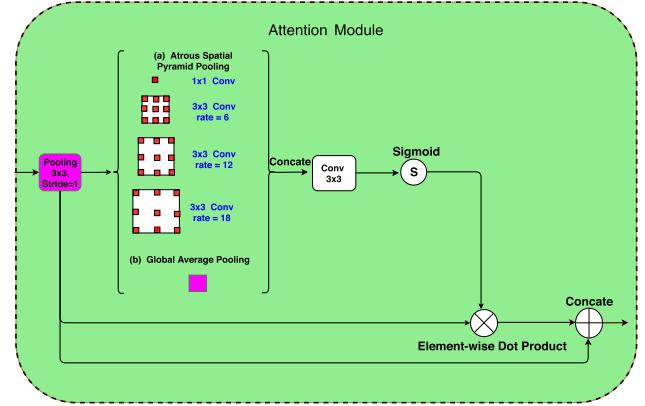


Fig. 5. An illustration of Attention Module. We extract multi-scale context features using multiple parallel filters with different dilation rates along with global average pooling. Afterwards, the visual attention map is computed to highlight critical signals of the input feature map and suppress irrelevant noise signals.

back to the desired dimension (as in Eq. (5)). It is worth mentioning that all the convolutional layers are followed by a Batch Normalization and ReLU layer sequentially. These five modules could be mathematically described as follows:

$$D_1(P) = \text{ReLU}(BN(\text{Conv}_{1 \times 1}(P))) \quad (1)$$

$$D_2(P) = \text{ReLU}(BN(\text{Conv}_{3 \times 3}^{(6)}(P))) \quad (2)$$

$$D_3(P) = \text{ReLU}(BN(\text{Conv}_{3 \times 3}^{(12)}(P))) \quad (3)$$

$$D_4(P) = \text{ReLU}(BN(\text{Conv}_{3 \times 3}^{(18)}(P))) \quad (4)$$

$$G(P) = Up(\text{ReLU}(BN(\text{Conv}_{1 \times 1}(\text{AvgPool}(P)))))) \quad (5)$$

The above feature maps are fused as:

$$H = D_1(P) \oplus D_2(P) \oplus D_3(P) \oplus D_4(P) \oplus G(P) \quad (6)$$

where \oplus represents channel-wise concatenation. Subsequently, we apply one single 3×3 convolution to refine the fused feature maps and reduce the number of output channel to 512 to match with the input feature map F . Further, the 3D attention map $M(F)$ is produced by applying a per-pixel sigmoid operation to the refined feature maps. As a result,

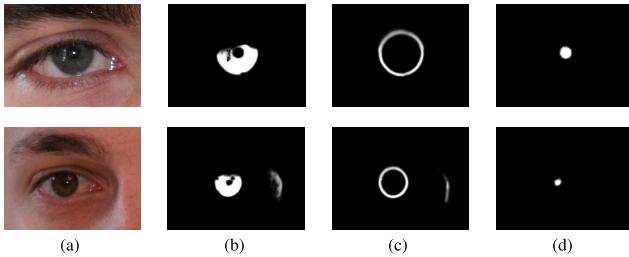


Fig. 6. Iris images and corresponding network outputs. (a) iris images, (b-d) predicted iris mask, iris outer boundary, and pupil mask.

values of the attention map $M(F)$ are bounded in $[0,1]$, where the bigger value indicates the higher importance.

To focus on important signals of an input feature and suppress irrelevant noise signals, the attention operation is defined as:

$$P' = P \otimes (M(F)) \quad (7)$$

where \otimes represents element-wise dot product operation.

Finally, the original refined input is concatenated to the feature map after attention to keep other valuable information in the original input signal, as followed:

$$F' = P \oplus P' \quad (8)$$

C. Post-Processing

Probability images¹ of iris mask, iris outer boundary, and pupil mask are predicted by forwarding the iris image through the network. Generally, predicted results are clean and accurate, as shown in the first row of Fig. 6. However, in a few cases, predicted results may contain some adverse noise regions, which affect the accuracy of iris segmentation, as shown in the second row of Fig. 6. To reduce the interference of noise, we make use of spatial priori constraints among iris mask, iris outer boundary, and pupil mask to denoise the predicted results. To be specific, probability images of iris mask, iris outer boundary, and pupil mask are first binarized, where iris mask adopts a fixed threshold value 127 while iris outer boundary and pupil mask employ an adaptive thresholding way based on Otsu's method [40]. The reason for two different thresholding approaches is that valid iris texture region consistently has high pixel values in the iris mask image while for iris outer boundary and pupil mask images, the occluded part from eyelids, specular reflections, etc., tends to have relatively lower pixel values than the exposed part due to different prediction difficulty, hence it is difficult to decide a globally generic threshold. Fig. 7(b-d) shows iris mask, iris outer boundary, and pupil mask images after binarization.

As expected, apart from the valid prediction targets, other noise objects are also in the binary images. To further exclude these noise, all objects are performed 8-neighborhood connected-component analysis, and valid and noisy connected objects are therefore separated from each other, as shown in Fig. 7(e-g). Considering the iris mask is surrounded by the inner and outer iris boundaries, hence if the valid iris mask

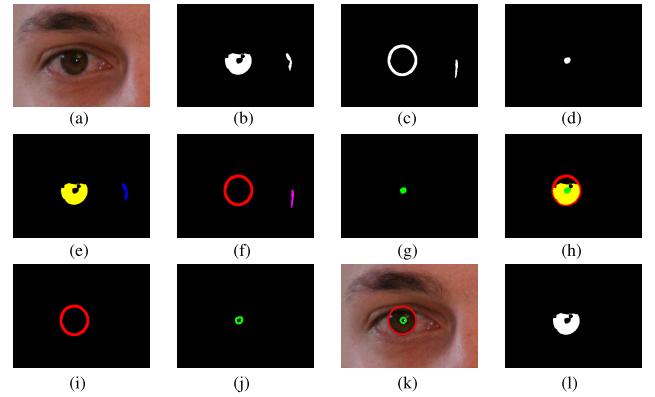


Fig. 7. Post-processing pipeline of noisy network outputs, which corresponds to the second row of Fig. 6. (a) iris image, (b-d) thresholded binary iris mask, iris outer boundary, and pupil mask images, (e-g) 8-neighborhood connected objects separated from (b-d) foreground regions, which are labeled with different colours (h) the largest triplet set extracted from (e-g), where a triplet set is composed of the iris mask connected object and its adjacent iris outer boundary and pupil mask connected objects, (i) candidate iris outer boundary sampling points (red), (j) candidate iris inner boundary sampling points (green), (k) iris localization results after curving fitting (inner boundary (green) and outer boundary (red)), and (l) refined iris mask image.

connected object is detected, its adjacent iris outer boundary and pupil mask connected objects are considered to be valid. It is empirically assumed that a region is adjacent to another region means that the chessboard distance between two regions is less than or equal to 10 pixels. Since there may be multiple connected objects in the binary iris mask image, we traverse each object separately and then search for its adjacent iris outer boundary and pupil mask connected objects, which collectively constitute a triplet set. Those remaining connected objects that cannot form a triplet set would be treated as noise and discarded. Of these detected triplet sets, the largest triplet set is deemed to be valid and contains noise-free iris mask, iris outer boundary, and pupil mask connected objects, as illustrated in Fig. 7(h).

Pixel points of the detected iris outer boundary connected object are then identified as candidate iris outer boundary sampling points, as shown in Fig. 7(i), and a morphological boundary extraction operation is also imposed on the detected pupil mask connected object to extract the candidate iris inner boundary sampling points, as shown in Fig. 7(j). Subsequently, any form of parameterized iris boundaries (*e.g.*, circle, ellipse, spline) are achieved by fitting candidate inner and outer boundary sampling points. In this work, without special instructions, we apply least-squares circle fitting algorithm [41] to produce circular inner and outer iris boundaries, as illustrated in Fig. 7(k). Besides, we further refine the initially predicted iris mask by making an intersection of the annular region encompassed by predicted parameterized inner and outer iris boundaries and the thresholded binary iris mask region. Fig. 7(l) presents the refined iris mask image. The preceding operations are also applicable to iris images with clean and accurate network outputs, as shown in Fig. 8.

Finally, to ensure the accuracy of segmentation results, inspired by [20] and based on the prior knowledge, we further check the segmentation results and think that a segmentation is correct only when all of the following quality measures are

¹Initial pixel value in a probability image is bounded in the $[0, 1]$ through the Sigmoid layer, and then rescaled to $[0, 255]$.

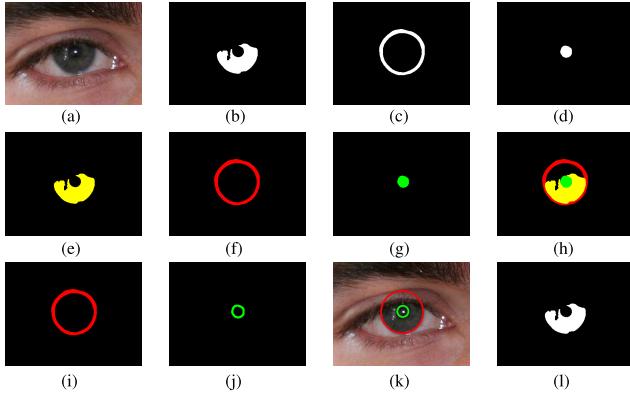


Fig. 8. Post-processing pipeline of clean network outputs, which corresponds to the first row of Fig. 6. (a-l) is in accordance with that in Fig. 7.

satisfied: 1) After threshold segmentation, all the foreground iris mask, iris outer boundary, and pupil mask must exist, 2) The radius of the inner boundary cannot be larger than 1/4th of the smallest image size, 3) The radius of the outer boundary should be 1.2 times greater than or equal to that of the inner boundary, and 4) The inner iris boundary should be contained in the outer iris boundary.

D. Training Objectives

We simultaneously optimize all the outputs of the network in an end-to-end manner. Specially, since the predictions of iris mask, iris outer boundary, and pupil mask could be regarded as a pixel-wise binary classification problem, a standard binary cross-entropy loss is employed. The overall loss function is expressed as follows:

$$\begin{aligned} \mathcal{L}(h(X|W), G) \\ = & \lambda_1 \mathcal{L}_{\text{iris_mask}} + \lambda_2 \mathcal{L}_{\text{outer_boundary}} + \lambda_3 \mathcal{L}_{\text{pupil_mask}} \\ = & \lambda_1 l(M_{\text{iris}}, \bar{M}_{\text{iris}}) + \lambda_2 l(B_{\text{outer}}, \bar{B}_{\text{outer}}) \\ & + \lambda_3 l(M_{\text{pupil}}, \bar{M}_{\text{pupil}}) \end{aligned} \quad (9)$$

where $h(X|W) = \{M_{\text{iris}}, B_{\text{outer}}, M_{\text{pupil}}\}$ represents the iris mask, iris outer boundary and pupil mask predicted by IrisParseNet, $G = \{\bar{M}_{\text{iris}}, \bar{B}_{\text{outer}}, \bar{M}_{\text{pupil}}\}$ is the corresponding ground truth, $\mathcal{L}_{\text{iris_mask}}$, $\mathcal{L}_{\text{outer_boundary}}$ and $\mathcal{L}_{\text{pupil_mask}}$ denote the binary cross-entropy loss for iris mask, iris outer boundary and pupil mask. $h(X|W)$ is the model hypothesis taking image X as input, parameterized by W . We can obtain the optimal parameters by minimizing the overall loss function as follow:

$$(W)^* = \arg \min \mathcal{L}. \quad (10)$$

The hyper-parameters λ_1 , λ_2 and λ_3 are all set to 1 in our experiments, making these loss value ranges comparable.

III. EXPERIMENTS AND ANALYSIS

In this section, extensive experiments are conducted to evaluate the proposed model from various aspects. The detailed experimental settings and results are presented in the following sections.

A. Databases

We employ the following three representative and publicly available databases for our experiments:

- 1) **CASIA.v4-distance (CASIA)** [42] contains 2,567 images from 142 subjects, captured at-a-distance with a CASIA long-range iris camera under near infrared illumination (NIR). Each image includes the upper part of face, hence both left and right irises are contained. For segmentation experiments, we use a subset from [14], which consists of 400 iris images with resolution of 640×480 pixels from the first 40 subjects and iris masks are manually labeled by the authors. Same as [14], the first 300 images from the first 30 subjects are used for training, and the last 100 images from the last 10 subjects are left for testing.
- 2) **UBIRIS.v2 (UBIRIS)** [43] consists of 11,102 images from 259 subjects, captured on-the-move and at-a-distance with a Canon EOS 5D camera under visible illumination (VIS). In NICE. I competition [44], a subset of 1,000 UBIRIS.v2 images with resolution of 400×300 pixels is used. Moreover, iris masks are manually labeled. We use the same protocol as NICE.I competition for experiments, i.e., 500 images are used for training and another disjoint 500 images are used for testing. However, it should be noted that the actual testing set has only 445 images when we download the test images from the organizers.
- 3) **MICHE-I (MICHE)** [45] is created to evaluate and develop algorithms for colour iris images captured by mobile devices. Images in MICHE-I are captured by three mobile devices including iPhone5 (1,262 images), Samsung Galaxy S4 (1,297 images), and Samsung Galaxy Tab2 (632 images) in uncontrolled conditions with visible illumination (VIS) and without the assistance of any operator. For our experiments, two MICHE subsets from Hu *et al.* [46] and Bezerra *et al.* [47] with ground truth iris masks are used to form a large database with 871 images, where 680 images are randomly selected for training and the rest 191 images are used for testing.

Images from these adopted databases are acquired under different types of non-cooperative environments, involving different illumination (NIR, VIS) and imaging sensors (long-range and mobile iris cameras), thus various kinds of noise factors such as out-of-focus, motion blur, gaze deviation, occlusions, specular reflections, illumination variation, iris deformation, and rotation are contained. Hence, it is convincing and reasonable to evaluate the performance of the proposed approach using these databases.

In addition to ground truth iris masks which were labeled by other researchers, we manually label the pupil mask (boundary) and iris outer boundary for each selected iris image. The initially labeled ground-truth iris outer boundary only has a width of one pixel, but the predicted boundary from deep models is rather thick. To tackle this inconsistency, inspired by [48], ground-truth iris outer boundary images of training set are dilated using morphologic dilation operator with a circular structuring element of radius 3.

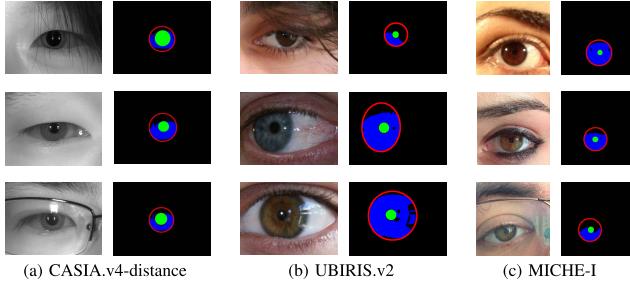


Fig. 9. Example images and corresponding ground truths (including pupil mask (green), iris outer boundary (red), and iris mask (blue)) of three iris databases. Best viewed in color and with magnification.

Some examples with manual labeled ground truths are shown in Fig. 9. We seek to accurately locate the inner and outer boundaries of the iris as well as eliminate all noise presented to separate the valid iris pixels.

B. Implementation Details

The proposed model is implemented in *Caffe* [49] and initialized with the VGG-16 model [50] pretrained on ImageNet. We train the network using mini-batch stochastic gradient descent (SGD) [51] with batch size of 4, momentum of 0.9 and weight decay of 0.0005. Inspired by [52], we use the “poly” learning rate policy where the learning rate is multiplied by $(1 - \frac{\text{iter}}{\max_iter})^{\text{power}}$ with *power* set to 0.9, initial learning rate set to $1e^{-3}$ and maximal iteration of 30,000.

Data augmentation is a simple yet effective way to enrich training data. During training, we augment each training sample with random combination of different geometric transformations (scaling, translation, flip, rotation, cropping) and image variations (blur) on-the-fly. Detailed augmentation operations are: (1) randomly resize the image (and gt maps) to 7 scales (0.5, 0.75, 1, 1.25, 1.5, 1.75, 2.0); (2) randomly blur the image (mean filter, gaussian blur, median blur, bilateral filter, box blur); (3) randomly translate the image (and gt maps) in x and y axis by a uniform factor between -30 and 30; (4) randomly left or right flip the image (and gt maps); (5) randomly rotate the image (and gt maps) by a uniform factor between -60 and 60; and (6) randomly crop the image (and gt maps) to a fixed size (321×321) at last. Besides, all training images (and gt maps) are shuffled when reaching the end of an epoch. For testing, we drop all augmentation operations and directly apply the model on the original image.

C. Evaluation Protocols

To quantitatively evaluate the proposed approach, we introduce several evaluation protocols for narrowly-defined iris segmentation, iris localization and iris recognition. The details are described as follows:

- 1) Iris segmentation: The NICE. I competition [44] provides two metrics to evaluate the accuracy of iris segmentation. The first measurement is the average segmentation error rate, which could be formulated as follows:

$$E1 = \frac{1}{n \times c \times r} \sum_{c'} \sum_{r'} G(c', r') \otimes M(c', r') \quad (11)$$

where n is the number of test images of r rows and c columns. In addition, G and M are the ground truth iris mask and the predicted iris mask, respectively, and c', r' are the column and row coordinates of pixels in G and M . The operator \otimes represents the XOR operation to evaluate the inconsistent pixels between G and M .

The second error measure aims to compensate the disproportion between the priori probabilities of “iris” and “non-iris” pixels in the images. To be specific, it averages the false positives (fp) and false negatives (fn) rates as follows:

$$E2 = \frac{1}{2 \times n} \sum_i (fp + fn) \quad (12)$$

where n is the number of testing images.

We also report the F-Measure (*F1*) (the harmonic mean of precision and recall) [53] to provide a comprehensive analysis of the propose method.

Values of *E1* and *E2* are bounded in $[0, 1]$, where the smaller value indicates the better result. The value of *F1* also falls in the same interval, but the greater value suggests the higher performance in this case.

- 2) Iris localization: Inspired by [54], we compute the Hausdorff distance between the predicted iris inner or outer boundary (denoted as B) and the corresponding ground truth boundary (denoted as G) to measure their shape similarity, which could be defined as:

$$H(G, B) = \max\{\sup_{x \in G} \inf_{y \in B} \|x - y\|, \sup_{y \in B} \inf_{x \in G} \|x - y\|\} \quad (13)$$

Furthermore, the Hausdorff distance is normalized by horizontal eye width to make the localization results at different image scale and iris size comparable.

Smaller normalized Hausdorff distances correspond to higher shape similarity between the predicted iris boundary and its ground truth, suggesting higher localization accuracy. We report the mean normalized Hausdorff distance (mHdis) for iris inner or outer boundary to evaluate their own localization accuracy. Additionally, the average value of two mean normalized Hausdorff distances demonstrates the overall accuracy of iris localization, thus we include it in the evaluation protocol.

Besides, inspired by [55], we report the varying iris localization success rate with respect to an error threshold given by both normalized Hausdorff distances between predicted inner and outer iris boundaries and their ground truths, and draw the corresponding Percentage of Correct Localization (PCL) curve. The further statistic, i.e., the area-under-the-curve (AUC) corresponding to the maximum threshold (τ), expressed as $AUC@ \tau$ is also presented.

In summary, the proposed normalized Hausdorff distance is not affected by image scale, iris size, and parameterization form of iris boundaries (e.g., circle, ellipse, spline), hence it is a unified evaluation criterion.

- 3) Iris recognition: To verify that the proposed method is able to improve the performance of iris recognition,

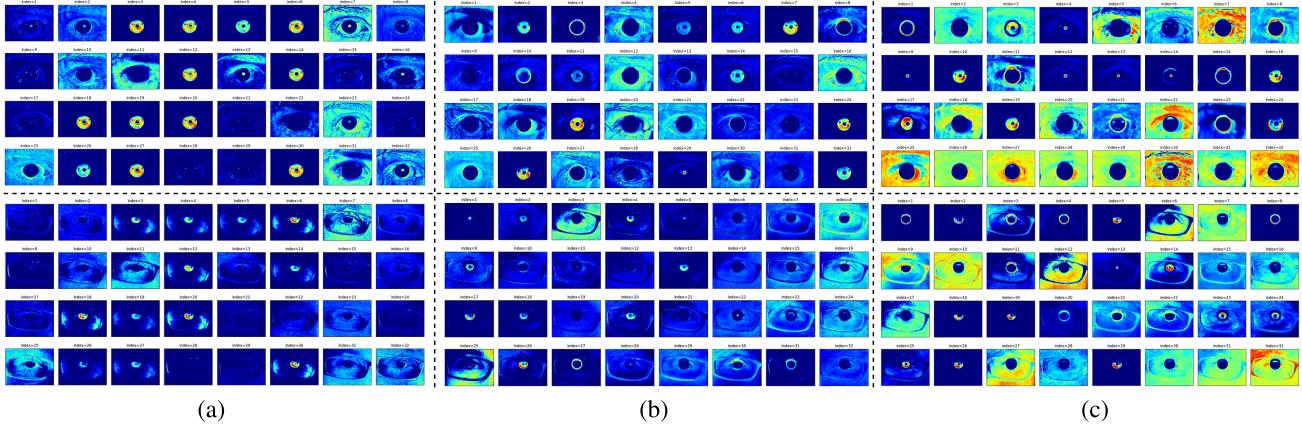


Fig. 10. Visualization of feature maps obtained by different IrisParseNet variants, i.e., (a) IrisParseNet (only Mask, w/o attention), (b) IrisParseNet (Multi-task, w/o attention), and (c) IrisParseNet (Multi-task, w/ attention), which corresponds to sample images of Fig. 12 from top to bottom.

TABLE I
COMPARISON OF IRISPARSENET UNDER DIFFERENT ABLATION SETTINGS

Dataset	Method	E1 (%)	E2 (%)	F1 (%)	Overall mHdis(%)	AUC@.1
CASIA.v4-distance	IrisParseNet (only Mask, w/o attention)	0.43	0.22	93.91	1.1794	0.0908
	IrisParseNet (Multi-task, w/o attention)	0.42	0.21	94.08	1.0840	0.0900
	IrisParseNet (Multi-task, w/ attention)	0.41	0.20	94.25	0.7768	0.0966
UBIRIS.v2 (NICE.I)	IrisParseNet (only Mask, w/o attention)	0.94	0.47	90.55	2.8277	0.0741
	IrisParseNet (Multi-task, w/o attention)	0.85	0.42	91.74	1.5953	0.0850
	IrisParseNet (Multi-task, w/ attention)	0.84	0.42	91.78	1.4837	0.0869
MICHE-I	IrisParseNet (only Mask, w/o attention)	0.72	0.36	92.48	3.4121	0.0654
	IrisParseNet (Multi-task, w/o attention)	0.70	0.35	92.56	1.5079	0.0862
	IrisParseNet (Multi-task, w/ attention)	0.66	0.33	93.05	1.2416	0.0903

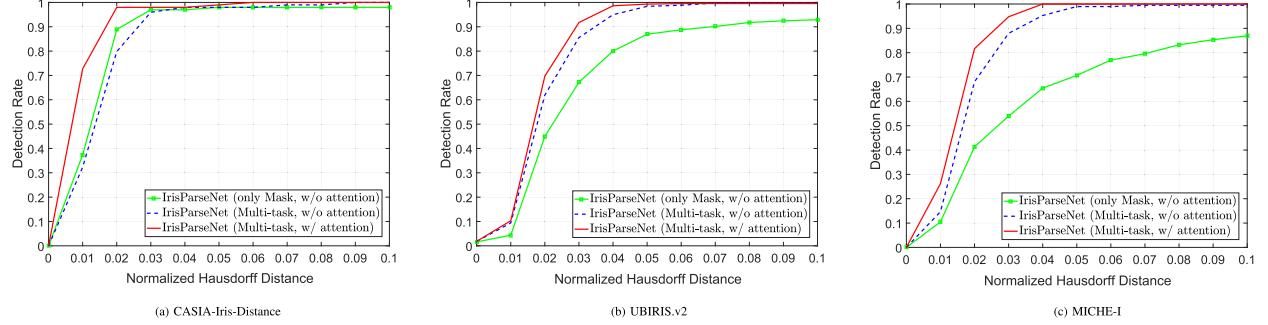


Fig. 11. Performance comparison of iris localization under different ablation settings using the proposed PCL curve. Detection success rate is thresholded on the Hausdorff distance error normalized by horizontal eye width. Best viewed in color and with magnification.

we conduct iris recognition experiments with all components but iris segmentation methods fixed. We use the equal error rate (EER) and Daugman's decidability index (DI) [6] to quantitatively evaluate the performance of iris recognition. Higher DI values correspond to better discriminative ability of iris recognition systems, meanwhile the iris recognition system with the lowest EER is considered the most accurate. Besides, a new metric introduced in [20], i.e., segmentation errors (SE), denoting the number of incorrect iris segmentation samples, is used to analyze the impact of segmentation exceptions on the iris recognition.

D. Ablation Study

In this section, we do ablation study to further verify the effectiveness of multi-task learning and attention module for improving the performance of iris segmentation.

Specifically, we compare three types of IrisParseNet models: IrisParseNet (Only Mask, w/o attention) represents the original U-Net model with only iris segmentation mask as prediction target, IrisParseNet (Multi-task, w/o attention) adds iris outer boundary and pupil mask as additional prediction targets on the basis of the former, IrisParseNet (Multi-task, w/ attention) is the proposed model, which further adopts attention mechanism to enhance the feature representation capabilities of the network, compared with IrisParseNet (Multi-task, w/o attention). Therefore, these three IrisParseNet variants are gradually improved. Since IrisParseNet (only Mask, w/o attention) could only predict iris masks, to achieve iris localization, we use the same method as CNNHT [20] to produce the circular inner and outer iris boundaries.

Experimental results are shown in Table I and Fig. 11. It can be observed that, compared with the baseline IrisParseNet (only Mask, w/o attention) model, IrisParseNet

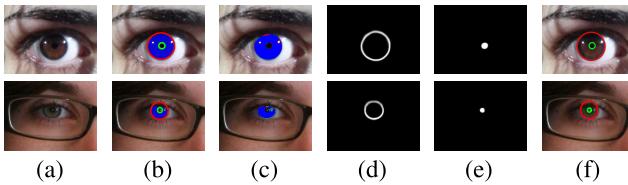


Fig. 12. Samples of good segmentation results using the proposed IrisParseNet. (a) Original iris images from MICHE-I (top row), and UBIRIS.v2 (bottom row) iris databases, (b) ground truth iris mask (blue), and inner (green) and outer (red) boundaries of the iris, (c) segmentation results of IrisParseNet (false positive error pixel (green), false negative error pixel (red), and true positive pixel (blue)), (d) iris outer boundary predicted by IrisParseNet, (e) pupil mask predicted by IrisParseNet, and (f) localization results of IrisParseNet after post-processing (inner boundary (green) and outer boundary (red)). Best viewed in color and with magnification.

(Multi-task, w/o attention) improve the iris segmentation and localization performance, especially for more challenging UBIRIS.v2 and MICHE-I databases. Further, with the assistance of attention module, IrisParseNet (Multi-task, w/ attention) achieves the higher segmentation and localization performance than IrisParseNet (Multi-task, w/o attention).

In addition to above quantitative comparisons, we have also conducted qualitative studies, to investigate possible performance improvements by incorporating the multi-task learning and attention mechanism. To this end, 32 feature maps obtained before the final 1×1 convolution in Fig. 4 are visualized and illustrated in Fig. 10. As can be seen, there are a large number of redundant feature maps related to iris mask and background in the IrisParseNet (only Mask, w/o attention). Besides, in the bottom row of Fig. 10, some of glasses areas are misidentified as iris pixels. Contrarily, by applying multi-task learning, IrisParseNet (Multi-task, w/o attention) produces more kinds of useful feature maps, such as iris outer boundary and pupil mask. And the former misidentification problem is greatly alleviated, we believe that such a big improvement because that multiple kinds of outputs implicitly impose spatial constrains that iris only occurs in the annular region surrounded by inner and outer iris boundaries, which reduces the segmentation errors. At the same time, more accurate parameterized iris boundaries are also obtained on the basis of additional predicted iris outer boundary and pupil mask, without the interference from irregular iris segmentation masks occurred in the those methods like CNNHT [20]. Moreover, the proposed attention module also helps IrisParseNet (Multi-task, w/ attention) further highlight those informative features and suppress adverse noise signals for both foreground and background. Based on these highly discriminative features, the proposed IrisParseNet, i.e., IrisParseNet (Multi-task, w/ attention), achieves accurate iris segmentation and localization results, as shown in Fig. 12. Overall, the above analysis indicates that the use of multi-task learning and attention mechanism helps improve the iris segmentation performance, and realize the complete iris segmentation.

E. Method Comparison

1) *Benchmarks:* We select five representative iris segmentation approaches, including traditional and deep learning based methods, as the benchmark. In particular,

Tan *et al.* [8] proposed an efficient and robust segmentation method to deal with noisy iris images and it could be roughly divided into four processes: clustering based coarse iris localization, pupillary and limbic boundary localization based on a novel integrodifferential constellation, eyelid localization and eyelash/shadow detection. The method was ranked the first place in NICE.I competition [44]. Since there is no source code available, we only report the result presented in the paper.

RTV- L^1 [9] proposed a novel total-variation based segmentation framework which used L^1 norm regularization to robustly suppress noisy texture pixels to obtain clear iris images. Then, an improved circular Hough transform was used to detect inner and outer circles on noise-free iris images. Finally, the authors developed a series of robust post-processing operations to accurately identify valid iris pixels based on local gray level analysis. We use the source codes² provided by the authors for experiments.

Haindl and Krupička [56] proposed an unsupervised segmentation method for colored eye images obtained through mobile devices. The method was ranked *first* in the Mobile Iris Challenge Evaluation (MICHE)-I [45] and also outperformed the NICE.I competition winning algorithm, namely Tan *et al.* [8], with average segmentation error rate $E1$ of 1.24% on UBIRIS.v2 database. We directly use the executable program³ provided by the authors to test on UBIRIS.v2 and MICHE-I databases except CASIA.v4-distance, as images in CASIA.v4-distance are grayscale images.

In addition to above traditional iris segmentation methods, MFCNs [14] were the first method that applied Fully Convolutional Network for iris segmentation to the best of our knowledge and achieved better results than previous state-of-the-art methods on CASIA.v4-distance and UBIRIS.v2 database. We reproduce this method as the authors suggested and apply it to our labeled three databases.

As stated in Section I-A, CNNHT [20] was specifically designed to solve the parameterization of the iris (i.e., iris localization) based on iris masks generated by CNN. Following the paper, we first apply the PyTorch implementation⁴ of Refinenet [57] to generate the iris mask, then use their proposed post-processing method⁵ to obtain the final circular iris boundaries.

Note: that except for RTV- L^1 and CNNHT, other baseline methods only provide the comparison of iris masks due to lack of the output of the inner and outer iris boundaries. Furthermore, all methods, including the proposed IrisParseNet, are trained or tuned on the training set from each single target database, then evaluated on the respective testing set for fair comparisons.

2) *Evaluation of Iris Segmentation and Localization:* Table II and Table III, and Fig. 13 provide summaries of the performance comparison of the proposed approach with baseline methods on iris segmentation and iris localization using the proposed evaluation protocols. We also report the

²<https://www4.comp.polyu.edu.hk/~csajaykr/tvmiris.htm>

³http://biplab.unisa.it/MICHE/MICHE-II/PRL_Haindl_Krupicka.zip

⁴<https://github.com/DrSleep/refinenet-pytorch>

⁵<http://wavelab.at/sources/Hofbauer19a/>

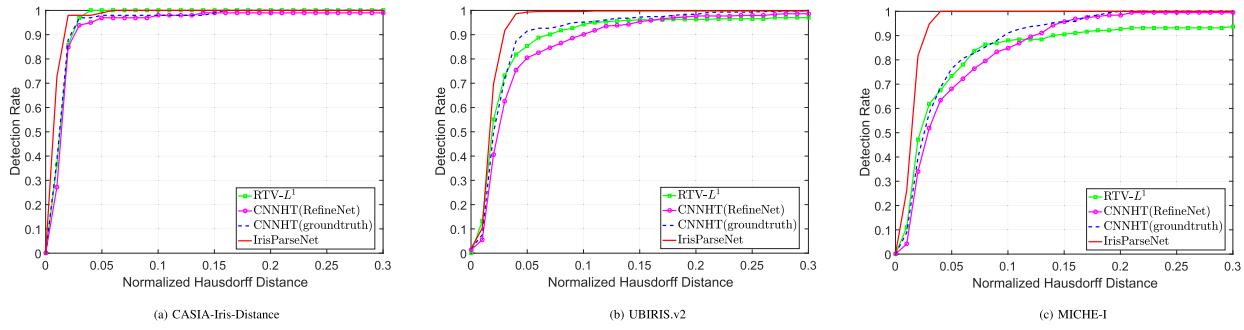


Fig. 13. Performance comparison of iris localization for different methods using the proposed PCL curve. Detection success rate is thresholded on the Hausdorff distance error normalized by horizontal eye width. Best viewed in color and with magnification.

TABLE II

COMPARISON OF DIFFERENT APPROACHES ON THE IRIS SEGMENTATION USING THE PROPOSED PROTOCOLS

Method	Database	E1 (%)	E2 (%)	F1 (%)
T. Tan <i>et. al.</i> [8]	UBIRIS.v2 (NICE.I)	1.31	N/A	N/A
RTV- L^1 [9]	CASIA.v4-distance	0.68	0.44	87.55
	UBIRIS.v2 (NICE.I)	1.21	0.83	85.97
	MICHE-I	2.42	1.21	79.24
Haindl and Krupička [56]	UBIRIS.v2 (NICE.I)	3.24	1.62	77.03
	MICHE-I	3.86	1.93	70.17
MFCNs [14]	CASIA.v4-distance	0.59	0.24	93.09
	UBIRIS.v2 (NICE.I)	0.90	0.49	91.04
	MICHE-I	0.74	0.37	92.01
CNNHT [20] (RefineNet)	CASIA.v4-distance	0.56	0.28	92.27
	UBIRIS.v2 (NICE.I)	0.97	0.48	90.34
	MICHE-I	0.80	0.40	91.41
IrisParseNet	CASIA.v4-distance	0.41	0.20	94.25
	UBIRIS.v2 (NICE.I)	0.84	0.42	91.78
	MICHE-I	0.66	0.33	93.05

parameter amount, FLOPs, storage space and runtime of the model in order to further evaluate the practicability of the proposed approach.

As can be seen from Table II, CNN-based iris segmentation methods consistently outperforms all traditional methods across all metrics on all databases. Furthermore, RTV- L^1 [9] achieves the state-of-the-art segmentation performance on all databases among the traditional methods. Similarly, the proposed IrisParseNet also presents the best segmentation results among the CNN-based methods. By Carefully comparing these two state-of-the-arts, IrisParseNet achieves the average improvement over RTV- L^1 of 39.71%, 30.58%, and 72.73% in the $E1$ metric, 54.55%, 49.40%, and 72.73% in the $E2$ metric, and 7.65%, 6.76%, and 17.43% in the $F1$ metric for CASIA.v4-distance, UBIRIS.v2 (NICE.I) and MICHE-I, respectively. We believe that, these results are attained first due to the discriminative power of the proposed deep learning model by using the multi-task learning and attention module, and also because our model is trained specifically for each database. It may be noted that the performance of Haindl and Krupička [56] is inconsistent with the presented results in their paper despite that we used their provided executable program for all experiments.

For iris localization, since RTV- L^1 [9] and CNNHT [20] can produce the circular iris boundaries, for fair comparisons with these methods, we adopt least-squares circle fitting

TABLE III

COMPARISON OF DIFFERENT APPROACHES ON THE IRIS LOCALIZATION (CIRCULAR BOUNDARY) USING THE PROPOSED PROTOCOLS

Method	Database	Inner	Outer	Overall	
		mHdis (%)	mHdis (%)	mHdis (%)	AUC@0.3
RTV- L^1 [9]	CASIA.v4-distance	0.7046	1.2457	0.9751	0.2919
	UBIRIS.v2 (NICE.I)	2.0911	2.8387	2.4649	0.2693
	MICHE-I	3.5578	5.2127	4.3852	0.2522
CNNHT [20] (RefineNet)	CASIA.v4-distance	1.1973	2.0251	1.6112	0.2861
	UBIRIS.v2 (NICE.I)	2.9887	3.2382	3.1135	0.2633
	MICHE-I	3.7298	3.6351	3.6824	0.2559
CNNHT [20] (groundtruth)	CASIA.v4-distance	0.9050	1.2649	1.0850	0.2899
	UBIRIS.v2 (NICE.I)	1.6668	2.6476	2.1572	0.2746
	MICHE-I	2.7750	3.1327	2.9538	0.2643
IrisParseNet	CASIA.v4-distance	0.6874	0.8662	0.7768	0.2966
	UBIRIS.v2 (NICE.I)	1.4644	1.503	1.4837	0.2865
	MICHE-I	1.1533	1.3299	1.2416	0.2903

method to obtain the circular inner and outer iris boundaries, as described in Section II-C. Moreover, for CNNHT, we use the ground-truth iris masks and predicted iris masks from RefineNet [57] to obtain the circular iris boundaries, respectively. Results of iris localization are shown in Table III and Fig. 13. It can be observed that the proposed approach consistently outperforms the other two methods across all metrics on all databases. In particular, in two noisy VIS databases, i.e., UBIRIS.v2 and MICHE-I, our method achieves the highest AUC@0.3 values of 0.2865, and 0.2903, respectively, showing significant superiority over the other two methods. Further, when comparing two CNNHT variants, we can see that by using the ground truth iris masks, CNNHT can achieve the better localization results than using the predicted iris masks from RefineNet on all databases, which reflects that the localization accuracy of CNNHT depends on the quality of segmentation masks. Additionally, CNNHT does not show a clear advantage over traditional RTV- L^1 , even for less noisy CASIA.v4-distance, its localization performance (AUC@0.3: 0.2899) is inferior to RTV- L^1 (AUC@0.3: 0.2919). Hence, methods of locating iris boundaries only relying on the iris mask are neither accurate or robust, which in turn proves the importance of our approach.

Although the proposed approach achieves superior segmentation and localization performance, it consumes relatively large computing and storage resources. Specifically, it has 31.28M parameters, 263.56G FLOPs relative to the resolution

TABLE IV
COMPARISON OF DIFFERENT SEGMENTATION APPROACHES ON THE IRIS RECOGNITION USING THE PROPOSED PROTOCOLS

Encoding	Method	CASIA.v4-distance					CASIA-Iris-M1-S1					MMU2				
		EER	DI	SE	# genuine	# imposter	EER	DI	SE	# genuine	# imposter	EER	DI	SE	# genuine	# imposter
1D log-Gabor	RTV- L^1 [9]	17.98%	1.79	1,226	21,143	4,832,027	27.66%	1.26	1,932	4,374	371,904	N/A	N/A	N/A	N/A	N/A
	CNNHT [20]	12.47%	2.25	12	38,078	9,334,207	18.35%	1.72	0	26,600	3,892,000	14.57%	2.03	0	1,990	492,525
	IrisParseNet	4.93%	3.17	11	38,083	9,338,532	11.25%	2.30	11	26,415	3,861,451	5.84%	3.36	6	1,966	486,600
FeatNet	RTV- L^1 [9]	16.91%	1.89	1,226	18,520	4,001,499	20.70%	1.70	1,932	2,774	178,722	N/A	N/A	N/A	N/A	N/A
	CNNHT [20]	9.07%	2.66	12	34,891	8,480,631	17.25%	1.83	0	24,698	3,619,889	14.88%	2.00	0	1,888	467,773
	IrisParseNet	2.71%	4.03	11	35,047	8,241,736	11.39%	2.37	11	25,674	3,683,747	5.70%	3.65	6	1,898	445,302

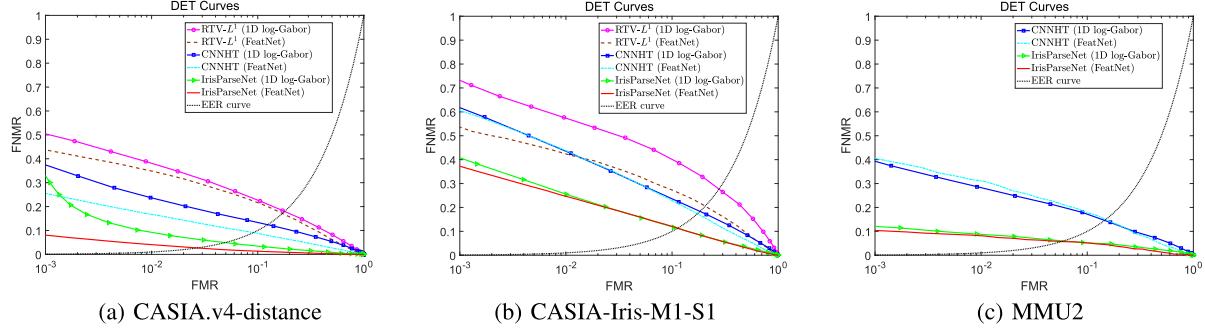


Fig. 14. DET curves of iris recognition experiments on CASIA.v4-distance, CASIA-Iris-M1-S1, and MMU2 iris databases using different iris segmentation and feature extraction methods. Best viewed in color and with magnification.

of 640×480 pixels, and occupies 119 MB storage space. As for runtime, for a 640×480 input image, the average inference time of IrisParseNet is 0.15 second on a NVIDIA TITAN Xp GPU with 12GB memory, and the average post-processing (including iris localization, normalization, IO operations) time is 0.12 second on a desktop computer with an Intel(R) Core(TM) i7-6700 CPU. Overall, our initially proposed model is still difficult to meet deployment requirements on real-world devices, hence further model optimization is required.

3) *Evaluation of Iris Recognition*: To perform iris recognition (more precisely, iris verification) experiments, we select three challenging NIR databases, i.e., CASIA.v4-distance [42], CASIA-Iris-M1-S1 [58] and MMU2 [59], and left and right iris images are treated as different identities in the experiment. The evaluation is performed on the all possible genuine and imposter pairs. Specifically, for CASIA.v4-distance database, the testing set contains 4,342 cropped left and right iris images from the first 119 subjects with resolution of 400×400 pixels. CASIA-Iris-M1-S1 is a mobile Asian iris database, which contains 1,400 images from 70 Asians. Since original samples capture the partial or complete face, we manually crop the left and right iris images and resize them to 400×400 pixels. MMU2 is a NIR database provided by the Malaysian Multimedia University, which comprises of 995 iris images collected at-a-distance from 100 subjects with resolution of 320×238 pixels. For CASIA-Iris-M1-S1 and MMU2, we use their full sets for recognition experiments. Overall, these three databases are acquired in non-cooperative environments, hence much realistic noise is contained, bringing a great challenge to iris segmentation.

As described in Section II-A, during the experiments, RTV- L^1 [9], CNNHT [20], and the proposed IrisParseNet

pre-trained on the CASIA.v4-distance database are first used for complete iris segmentation, respectively, then Daugman's Rubber-Sheet model [27] is used to produce normalized iris image and iris mask for feature extraction and matching. We adopt the traditional 1D log-Gabor filters [32]⁶ and deep FeatNet [34] model pre-trained on the ND-IRIS-0405 database [60]⁷ to extract *IrisCodes* [27] and compute fractional Hamming Distance of iris codes to verify whether two iris are from the same class. Therefore, the only two factors that impact the recognition performance are the iris segmentation and feature extraction methods. It may be noted that that VIS databases are not evaluated, this is because they usually contain many noisy colored iris images that are not compliant with the ISO/IEC 29794-6:2015 standard on the iris image quality, thus are not suitable to be used with a traditional iris recognition algorithm [22], and for deep FeatNet, no specific pre-trained models for VIS images are available.

Evaluation results of iris recognition are shown in Table IV and Fig. 14. First, in terms of feature encoding, FeatNet can achieve better recognition accuracy than 1D log-Gabor filters for any iris segmentation methods in most cases. This on the one hand reflects the strong generalization capability of deep FeatNet, since the testing model is directly from the pre-trained model on the ND-IRIS-0405 database without any further tuning, and more importantly, deep FeatNet can extract more discriminative features than traditional 1D log-Gabor filters. On the other hand, from the number of genuine and

⁶It can be implemented using the open source USIT 2.4.2, which is made available via <http://www.wavelab.at/sources/Rathgeb16a/>

⁷Official models and codes are made available via <https://www4.comp.polyu.edu.hk/~csajaykr/deepiris.htm>

impostor pairs (# genuine, # imposter), we can see that for any iris segmentation methods, under the same SE, which means the initial possible genuine and imposter pairs are the same, deep FeatNet obtains the fewer matching pairs than 1D log-Gabor filters at last. This is because that in the matching process of FeatNet, only when the overlap of the two iris masks from a matching pair is larger than 20%, a matching score is produced, which also further implies that more accurate iris segmentation methods could produce more matching pairs, especially for genuine pairs. Similar restrictions are not applied to 1D log-Gabor filters. Hence, by filtering invalid matches, FeatNet may also yield better recognition performance than 1D log-Gabor filters.

Second, further comparisons of three segmentation methods can observe that the experiments using the proposed IrisParseNet model can achieve the lower EER and FNMR@FMR = 10^{-3} , higher DI, and produce the better DET curves than those using other two segmentation methods for both two feature encoding techniques. Specially, for CASIA.v4-distance and MMU2, on the basis of *IrisCodes* produced by FeatNet, IrisParseNet obtains the extremely low EER values of 2.71%, and 5.70%, respectively, reaching state-of-the-art recognition performance. Moreover, in terms of SE, CNNHT almost produces the lowest values on all databases, followed by the proposed IrisParseNet, and the worst is RTV-L¹, which is even not applicable to MMU2 according to our many experiments. The reason may be that CNNHT always try to generate a parameterization based on the iris mask, while the proposed IrisParseNet sets several quality measures to check the accuracy of iris segmentation, hence some poorly segmented results are filtered. As for RTV-L¹, it is not robust for challenging noisy iris images, hence leads to a large number of incorrect segmentations. Further, according to the previous analysis, we compare the number of genuine pairs when FeatNet is used as a feature extractor. We can clearly see that IrisParseNet can produce more genuine pairs than other two segmentation methods on all databases, even when its SE value is larger than CNNHT on CASIA-Iris-M1-S1 and MMU2 databases. We believe that these results are attained because the proposed model is more accurate and robust, and has a strong generalization capability.

In summary, experiment results illustrate that the proposed IrisParseNet greatly improves the performance of iris recognition. Further, in conjunction with state-of-the-art feature extraction and matching techniques, it can achieve state-of-the-art iris recognition accuracy.

F. Qualitative Evaluation

Here we perform a qualitative evaluation to further analyze the advantages and disadvantages of our method. We select several representative and challenging iris images from CASIA.v4-distance, MICHE-I, and UBIRIS.v2 databases, respectively, where part of iris images are from Fig. 2 and suffer from incorrect iris segmentation with CNNHT [20]. In contrast, the proposed IrisParseNet shows good segmentation results, as shown in Fig. 15. It can be seen that our model predicts accurate iris mask, iris outer boundary, and pupil mask in face of all kinds of noise such as occlusions,

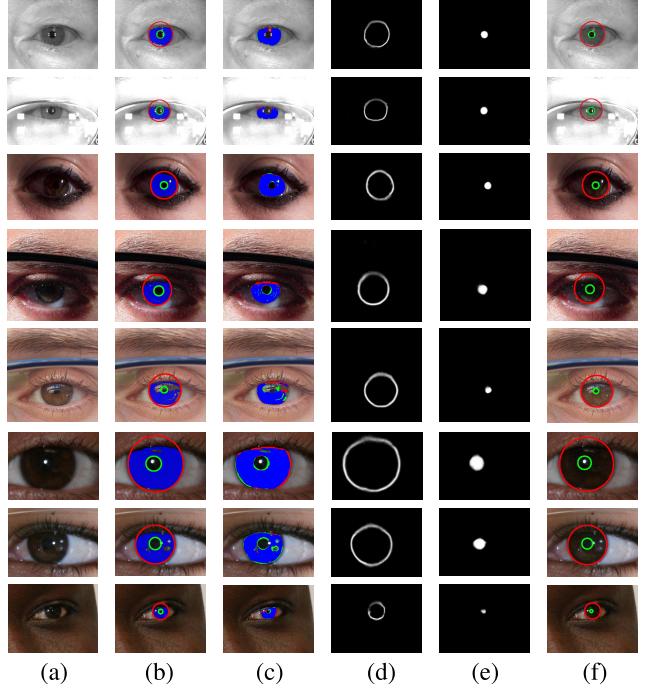


Fig. 15. Samples of good segmentation results using the proposed IrisParseNet. (a) Original iris images from CASIA.v4-distance (top two), MICHE-I (middle three), and UBIRIS.v2 (bottom three) iris databases, where part of iris images are from Fig. 2, (b-f) is in accordance with that in Fig. 12.

specular reflections, dark iris, uneven illumination as well as different sized iris. In particular, for iris images with serious occlusions due to eyelids or glasses, the predicted iris outer boundary is relatively complete and highlighted. Additionally, the predicted pupil mask is almost noiseless and accurately segment the full pupil region. Hence, based on these accurate predictions, the proposed IrisParseNet successfully locates the inner and outer iris boundaries, as illustrated in Fig. 15(f).

Generally, glasses are considered as an important noise source in non-cooperative iris recognition, especially possible noise factors, e.g., specular reflections, shadows, motion blur, are more pronounced in iris images with glasses [61], [62]. Besides, glasses frames/edges may also be misidentified as pupillary or limbic boundaries by general iris segmentation algorithms such as open-source OSIRIS [63]. As a consequence, these issues often lead to incorrect iris segmentation, and further deteriorate the iris recognition performance. For this reason, a special investigation about the impact of glasses on the proposed IrisParseNet is made. Fig. 16 shows iris segmentation results of some representative iris images with glasses. As can be observed, previously mentioned specular reflections, shadows, etc., are presented in both NIR and VIS iris images with glasses, which poses a critical challenge to iris segmentation. Nevertheless, as discussed in Section III-D, due to the proposed multi-task attention Network, the proposed IrisParseNet still predicts the fairly accurate iris mask, iris outer boundary, and pupil mask as expected, except for partially incomplete iris outer boundary and tiny pupil mask noise, whose errors would be further eliminated by the proposed post-processing, as stated in Section II-C. In summary, visual

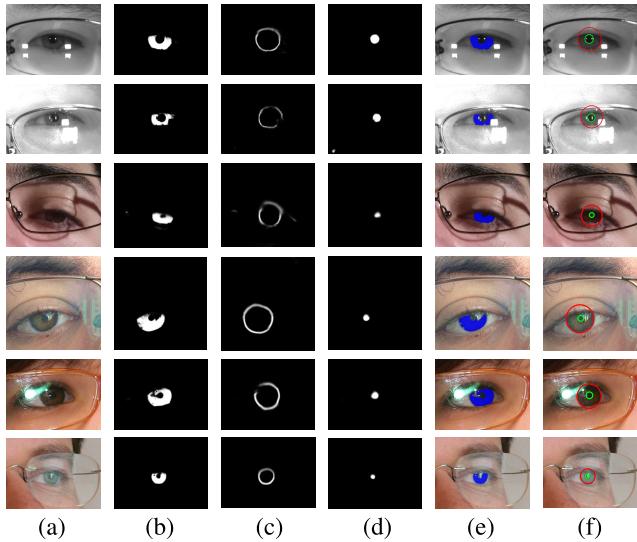


Fig. 16. Samples of iris segmentation results with glasses using the proposed IrisParseNet. (a) Original iris images from CASIA.v4-distance (top two), MICHE-I (middle two), and UBIRIS.v2 (bottom two) databases, (b-d) predicted iris mask, iris outer boundary and pupil mask, (e) iris images with the refined iris mask labeled in blue, and (f) localization results of IrisParseNet after post-processing (inner boundary (green) and outer boundary (red)).

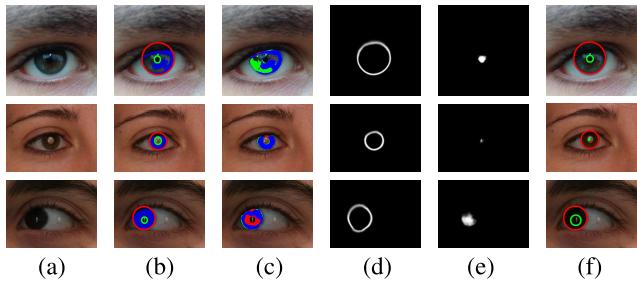


Fig. 17. Samples of incorrect segmentation results using the proposed IrisParseNet. (a) Original iris images from MICHE-I (top one) and UBIRIS.v2 (bottom two) databases, (b-f) is in accordance with that in Fig. 12.

results demonstrate that glasses do not cause a clearly negative impact on the proposed IrisParseNet in most cases.

There are a few incorrect segmentation samples with our method, as exemplified in Fig. 17. Concretely, for the first row, the false positive errors are mainly caused by reflection noise whose pixel values are similar to those of the iris region. For the second row, the errors are mainly in the pupil region, which presents a noticeable inconsistency in the pupil color with iris images of the training set. For the last row, dark iris negatively affects the iris-pupil contrast and makes iris and pupil pixels hard to distinguish, resulting in a poor segmentation of iris and pupil masks. To sum up, these errors may be caused by a lack of sufficient training samples, hence in case more annotated training images with similar appearance are provided, the segmentation performance would be expected to improve significantly.

In addition to within-database evaluation, we further conduct a cross-database testing to examine the generalization capability of the proposed approach. For this purpose, we select three additional iris databases: Bath [64],

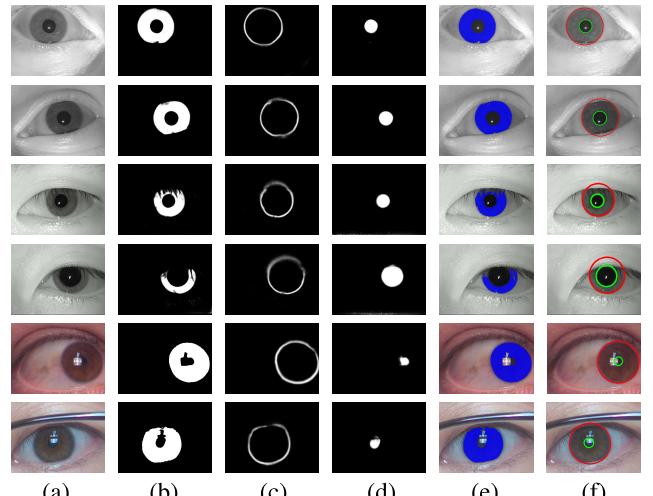


Fig. 18. Cross-database iris segmentation results using the proposed IrisParseNet. (a) Original iris images from Bath (top two rows), MMU1 (middle two rows), and Cross-Eyed (bottom two rows), (b-f) is in accordance with that in Fig. 16.

MMU1 [65], and Cross-Eyed [66], where the former two iris databases are acquired with traditional iris scanners in controlled NIR light source environments, while the last iris database is captured with a custom developed dual spectrum imaging sensor (NIR and VIS) in uncontrolled realistic indoor environments. During testing, according to the imaging light source (NIR or VIS), the trained model on CASIA.v4-distance is directly applied to Bath and MMU1, and the trained model on UBIRIS.v2 is directly applied to Cross-Eyed, both of which are without any further tuning. Fig. 18 presents examples of cross-database iris segmentation results. It is observed that the proposed approach consistently achieves quite good iris segmentation and localization results in dealing with iris images captured in both controlled and uncontrolled environments, suggesting its outstanding generalization capability to different databases.

IV. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a novel multi-task attention network for jointly learning iris mask, iris outer boundary, and pupil mask. Further, parameterized inner and outer iris boundaries are achieved based on the learned pupil mask and iris outer boundary. In this way, two major tasks of a complete iris segmentation, i.e., segmentation of iris mask and localization of parameterized inner and outer iris boundaries, are jointly completed. Meanwhile, we have taken efforts to label three representative and challenging iris databases, and establish comprehensive and unified evaluation protocols. The proposed approach is compared with state-of-the-art iris segmentation methods on the newly annotated iris databases, and shows a leading performance on various benchmarks. As for future work, we would explore more efficient strategies to explicitly exploit the spatial relationship among iris mask, inner and outer iris boundaries to improve the segmentation performance. Moreover, lightweight iris segmentation networks are also expected to be developed for practical deployment.

REFERENCES

- [1] A. K. Jain, P. Flynn, and A. A. Ross, *Handbook of Biometrics*. New York, NY, USA: Springer, 2007.
- [2] J. L. Cambier and J. E. Siedlarz, "Portable authentication device and method using iris patterns," U.S. Patent 6 532 298, Mar. 11, 2003.
- [3] A. F. Sequeira *et al.*, "PROTECT multimodal DB: Fusion evaluation on a novel multimodal biometrics dataset envisaging border control," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2018, pp. 1–5.
- [4] A. Ross. (2009). *Iris as a Forensic Modality: The Path Forward*. [Online]. Available: <http://www.nist.gov/forensics/upload/Ross-Presentation.pdf>
- [5] Z. He, T. Tan, Z. Sun, and X. Qiu, "Toward accurate and fast iris segmentation for iris biometrics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 9, pp. 1670–1684, Sep. 2009.
- [6] J. G. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 11, pp. 1148–1161, Nov. 1993.
- [7] R. P. Wildes, "Iris recognition: An emerging biometric technology," *Proc. IEEE*, vol. 85, no. 9, pp. 1348–1363, Sep. 1997.
- [8] T. Tan, Z. He, and Z. Sun, "Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition," *Image Vis. Comput.*, vol. 28, no. 2, pp. 223–230, Feb. 2010.
- [9] Z. Zhao and A. Kumar, "An accurate iris segmentation framework under relaxed imaging constraints using total variation model," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3828–3836.
- [10] S. Shah and A. Ross, "Iris segmentation using geodesic active contours," *IEEE Trans. Inf. Forensics Secur.*, vol. 4, no. 4, pp. 824–836, Dec. 2009.
- [11] R. Donida Labati, E. Muñoz, V. Piuri, A. Ross, and F. Scotti, "Non-ideal iris segmentation using polar spline RANSAC and illumination compensation," *Comput. Vis. Image Understand.*, vol. 188, Nov. 2019, Art. no. 102787.
- [12] A. Radman, N. Zainal, and S. A. Suandi, "Automated segmentation of iris images acquired in an unconstrained environment using HOG-SVM and GrowCut," *Digit. Signal Process.*, vol. 64, pp. 60–70, May 2017.
- [13] S. Banerjee and D. Mery, "Iris segmentation using geodesic active contours and grabcut," in *Proc. Pacific-Rim Symp. Image Video Technol.* Cham, Switzerland: Springer, 2016, pp. 48–60.
- [14] N. Liu, H. Li, M. Zhang, J. Liu, Z. Sun, and T. Tan, "Accurate iris segmentation in non-cooperative environments using fully convolutional networks," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2016, pp. 1–8.
- [15] E. Jalilian, A. Uhl, and R. Kwitt, "Domain adaptation for CNN based iris segmentation," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2017, pp. 1–6.
- [16] S. Bazrafkan, S. Thavalengal, and P. Corcoran, "An end to end deep neural network for iris segmentation in unconstrained scenarios," *Neural Netw.*, vol. 106, pp. 79–95, Oct. 2018.
- [17] D. Osorio-Roig, C. Rathgeb, M. Gomez-Barrero, A. Morales-Gonzalez, E. Garea-Llano, and C. Busch, "Visible wavelength iris segmentation: A multi-class approach using fully convolutional neuronal networks," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2018, pp. 1–5.
- [18] M. Arsalan, D. S. Kim, M. B. Lee, M. Owais, and K. R. Park, "Fred-net: Fully residual encoder-decoder network for accurate iris segmentation," *Expert Syst. Appl.*, vol. 122, pp. 217–241, 2019.
- [19] R. Vyas, T. Kanumuri, G. Sheoran, and P. Dubey, "Recent trends of ROI segmentation in iris biometrics: A survey," *Int. J. Biometrics*, vol. 11, no. 3, p. 274, 2019.
- [20] H. Hofbauer, E. Jalilian, and A. Uhl, "Exploiting superior CNN-based iris segmentation for better recognition accuracy," *Pattern Recognit. Lett.*, vol. 120, pp. 17–23, Apr. 2019.
- [21] Z. Zhao and A. Kumar, "A deep learning based unified framework to detect, segment and recognize irises using spatially corresponding features," *Pattern Recognit.*, vol. 93, pp. 546–557, Sep. 2019.
- [22] D. Kerrigan, M. Trokielewicz, A. Czajka, and K. W. Bowyer, "Iris recognition with image segmentation employing retrained Off-the-Shelf deep neural networks," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2019, pp. 1–7.
- [23] J. Daugman, "New methods in iris recognition," *IEEE Trans. Syst., Man, Cybern., B (Cybern.)*, vol. 37, no. 5, pp. 1167–1175, Oct. 2007.
- [24] H. Proença and L. A. Alexandre, "Iris segmentation methodology for non-cooperative recognition," *IEE Proc.-Vis., Image, Signal Process.*, vol. 153, no. 2, pp. 199–205, 2006.
- [25] G. Sutra, S. Garcia-Salicetti, and B. Dorizzi, "The viterbi algorithm at different resolutions for enhanced iris segmentation," in *Proc. 5th IAPR Int. Conf. Biometrics (ICB)*, Mar. 2012, pp. 310–316.
- [26] T. Rongnian and W. Shaojie, "Improving iris segmentation performance via borders recognition," in *Proc. 4th Int. Conf. Intell. Comput. Technol. Autom.*, Mar. 2011, pp. 580–583.
- [27] J. Daugman, "How iris recognition works," in *The Essential Guide to Image Processing*. Amsterdam, The Netherlands: Elsevier, 2009, pp. 715–739.
- [28] E. Jalilian and A. Uhl, "Iris segmentation using fully convolutional encoder-decoder networks," in *Deep Learning for Biometrics*. Cham, Switzerland: Springer, 2017, pp. 133–155.
- [29] S. Lian, Z. Luo, Z. Zhong, X. Lin, S. Su, and S. Li, "Attention guided U-Net for accurate iris segmentation," *J. Vis. Commun. Image Represent.*, vol. 56, pp. 296–304, Oct. 2018.
- [30] M. Arsalan, R. Naqvi, D. Kim, P. Nguyen, M. Owais, and K. Park, "IrisDenseNet: Robust iris segmentation using densely connected fully convolutional networks in the images by visible light and near-infrared light camera sensors," *Sensors*, vol. 18, no. 5, p. 1501, 2018.
- [31] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*. Cham, Switzerland: Springer, 2015, pp. 234–241.
- [32] L. Masek *et al.*, "Recognition of human iris patterns for biometric identification," Ph.D. dissertation, Univ. Western Australia, Perth, WA, Australia, 2003.
- [33] Z. Sun and T. Tan, "Ordinal measures for iris recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2211–2226, Dec. 2009.
- [34] Z. Zhao and A. Kumar, "Towards more accurate iris recognition using deeply learned spatially corresponding features," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3809–3818.
- [35] Q. Zhang, H. Li, Z. Sun, and T. Tan, "Deep feature fusion for iris and periocular biometrics on mobile devices," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 11, pp. 2897–2912, Nov. 2018.
- [36] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [37] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 3–19.
- [38] J. Park, S. Woo, J.-Y. Lee, and I. So Kweon, "BAM: Bottleneck attention module," 2018, *arXiv:1807.06514*. [Online]. Available: <http://arxiv.org/abs/1807.06514>
- [39] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*. [Online]. Available: <http://arxiv.org/abs/1706.05587>
- [40] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.
- [41] N. Chernov and C. Lesort, "Least squares fitting of circles," *J. Math. Imag. Vis.*, vol. 23, no. 3, pp. 239–252, Nov. 2005.
- [42] B. I. Test. *Casia.v4 Database*. Accessed: Feb. 2020. [Online]. Available: <http://www.idealtest.org/dbDetailForUser.do?id=4>
- [43] H. Proenca, S. Filipe, R. Santos, J. Oliveira, and L. A. Alexandre, "The UBIRIS.v2: A database of visible wavelength iris images captured On-the-Move and At-a-Distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1529–1535, Aug. 2010.
- [44] H. Proença and L. A. Alexandre, "The NICE.I: Noisy Iris challenge evaluation—Part I," in *Proc. 1st IEEE Int. Conf. Biometrics: Theory, Appl., Syst.*, Sep. 2007, pp. 1–4.
- [45] M. De Marsico, M. Nappi, D. Riccio, and H. Wechsler, "Mobile iris challenge evaluation (MICHE)-I, biometric iris dataset and protocols," *Pattern Recognit. Lett.*, vol. 57, pp. 17–23, May 2015.
- [46] Y. Hu, K. Sirlantzis, and G. Howells, "Improving colour iris segmentation using a model selection technique," *Pattern Recognit. Lett.*, vol. 57, pp. 24–32, May 2015.
- [47] C. S. Bezerra *et al.*, "Robust iris segmentation based on fully convolutional networks and generative adversarial networks," in *Proc. 31st SIBGRAPI Conf. Graph., Patterns Images (SIBGRAPI)*, Oct. 2018, pp. 281–288.
- [48] Y. Liu and M. S. Lew, "Learning relaxed deep supervision for better edge detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 231–240.
- [49] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 675–678.
- [50] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
- [51] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

- [52] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," 2014, *arXiv:1412.7062*. [Online]. Available: <http://arxiv.org/abs/1412.7062>
- [53] H. Hofbauer, F. Alonso-Fernandez, P. Wild, J. Bigun, and A. Uhl, "A ground truth for iris segmentation," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 527–532.
- [54] K. Sirinukunwattana *et al.*, "Gland segmentation in colon histology images: The glas challenge contest," *Med. Image Anal.*, vol. 35, pp. 489–502, Jan. 2017.
- [55] W. Fuhl, D. Geisler, T. Santini, W. Rosenstiel, and E. Kasneci, "Evaluation of state-of-the-art pupil detection algorithms on remote eye images," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput. Adjunct (UbiComp)*, 2016, pp. 1716–1725.
- [56] M. Hajndl and M. Krupička, "Unsupervised detection of non-iris occlusions," *Pattern Recognit. Lett.*, vol. 57, pp. 60–65, May 2015.
- [57] G. Lin, A. Milan, C. Shen, and I. Reid, "RefineNet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1925–1934.
- [58] Q. Zhang, H. Li, M. Zhang, Z. He, Z. Sun, and T. Tan, "Fusion of face and iris biometrics on mobile devices using near-infrared images," in *Proc. Chin. Conf. Biometric Recognit.* Springer, 2015, pp. 569–578.
- [59] M. University. *Multimedia University Version 2 Iris Database*. Accessed: Oct. 2019. [Online]. Available: <https://www.cs.princeton.edu/~andyz/downloads/MMU2IrisDatabase.zip>
- [60] K. W. Bowyer and P. J. Flynn, "The ND-IRIS-0405 iris image dataset," 2016, *arXiv:1606.04853*. [Online]. Available: <http://arxiv.org/abs/1606.04853>
- [61] P. Drozdowski, F. Struck, C. Rathgeb, and C. Busch, "Detection of glasses in near-infrared ocular images," in *Proc. Int. Conf. Biometrics (ICB)*, Feb. 2018, pp. 202–208.
- [62] D. Osorio-Roig, P. Drozdowski, C. Rathgeb, A. Morales-Gonzalez, E. Garea-Llano, and C. Busch, "Iris recognition in visible wavelength: Impact and automated detection of glasses," in *Proc. 14th Int. Conf. Signal-Image Technol. Internet-Based Syst. (SITIS)*, Nov. 2018, pp. 542–546.
- [63] N. Othman, B. Dorizzi, and S. Garcia-Salicetti, "OSIRIS: An open source iris recognition software," *Pattern Recognit. Lett.*, vol. 82, pp. 124–131, Oct. 2016.
- [64] Smart-Sensor-Ltd. *Bath Iris Database*. Accessed: Dec. 2016. [Online]. Available: <http://www.smartsensors.co.uk/products/iris-database/>
- [65] M. University. *Multimedia University Version 1 Iris Database*. Accessed: Oct. 2019. [Online]. Available: <https://www.cs.princeton.edu/~andyz/downloads/MMU1IrisDatabase.zip>
- [66] A. Sequeira *et al.*, "Cross-eyed-cross-spectral iris/periocular recognition database and competition," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2016, pp. 1–5.



Caiyong Wang received the B.E. degree in applied mathematics from Xinjiang University, China, in 2013, and the M.S. degree in computational mathematics from Xiamen University, China, in 2016. He is currently pursuing the Ph.D. degree with the School of Artificial Intelligence, University of Chinese Academy of Sciences, China, and the Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. His current research interests include biometrics, computer vision, and deep learning. He has received an Honorable Mention Paper Award at IAPR ICB 2019.



Jawad Muhammad received the B.E. degree in computer engineering from Bayero University, Kano, Nigeria, in 2010, and the M.S. degree in computer engineering from Selcuk University, Konya, Turkey, in 2014. He is currently pursuing the Ph.D. degree with the School of Artificial Intelligence, University of Chinese Academy of Sciences, China, and the Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. His current research interests include biometrics, computer vision, and robot vision.



Yunlong Wang received the B.E. and Ph.D. degrees from the Department of Automation, University of Science and Technology of China. He is currently an Assistant Professor with the Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. His research focuses on pattern recognition, machine learning, light field photography, and biometrics.



Zhaofeng He received the B.E. degree from the University of Science and Technology of China in 2005 and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, in 2010. He is currently the Vice-General Manager with Beijing IrisKing Company Ltd. His research interests include biometrics and image understanding.



Zhenan Sun (Senior Member, IEEE) received the B.E. degree in industrial automation from the Dalian University of Technology, China, in 1999, the M.S. degree in system engineering from the Huazhong University of Science and Technology, China, in 2002, and the Ph.D. degree in pattern recognition and intelligent systems from the Institute of Automation, Chinese Academy of Sciences, China, in 2006. He is currently a Professor with the Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, and also with the School of Artificial Intelligence, University of Chinese Academy of Sciences, China. He has authored/coauthored over 200 technical papers. His research interests include biometrics, pattern recognition, and computer vision. He is the Chair of Technical Committee on Biometrics, International Association for Pattern Recognition (IAPR) and an IAPR fellow. He serves as an Associate Editor for the IEEE TRANSACTIONS ON BIOMETRICS, BEHAVIOR, AND IDENTITY SCIENCE.