J. Vis. Commun. Image R. 56 (2018) 296–304

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Attention guided U-Net for accurate iris segmentation ☆

Sheng Lian [a], Zhiming Luo [b], Zhun Zhong [a], Xiang Lin [c,d], Songzhi Su [a], Shaozi Li [a,*]

[a] Cognitive Science Department, Xiamen University, China
[b] Postdoc Center of Information and Communication Engineering, Xiamen University, China
[c] Fujian Provincial Key Laboratory of Ophthalmology and Visual Science, Xiamen University, China
[d] Eye Institute of Xiamen University, China

ARTICLE INFO

Article history:
Received 20 March 2018
Revised 3 September 2018
Accepted 4 October 2018
Available online 5 October 2018

Keywords:
Iris segmentation
U-Net
Attention

ABSTRACT

Iris segmentation is a critical step for improving the accuracy of iris recognition, as well as for medical concerns. Existing methods generally use whole eye images as input for network learning, which do not consider the geometric constrain that iris only occur in a specific area in the eye. As a result, such methods can be easily affected by irrelevant noisy pixels outside iris region. In order to address this problem, we propose the ATTention U-Net (ATT-UNet) which guides the model to learn more discriminative features for separating the iris and non-iris pixels. The ATT-UNet firstly regress a bounding box of the potential iris region and generated an attention mask. Then, the mask is used as a weighted function to merge with discriminative feature maps in the model, making segmentation model pay more attention to iris region. We implement our approach on UBIRIS.v2 and CASIA.IrisV4-distance, and achieve mean error rates of 0.76% and 0.38%, respectively. Experimental results show that our method achieves consistent improvement in both visible wavelength and near-infrared iris images with challenging scenery, and surpass other representative iris segmentation approaches.

1. Introduction

Biometric recognition technologies, such as fingerprint, face and iris recognition, are adopted in various applications including smartphone unlocking, border control, and attendance record. Among these technologies, iris-based technologies have advantages of high distinctiveness, permanence, and performance [1]. The appearance of human eye is composed of sclera, iris, and pupil. Iris is a circular-like region between pupil and sclera, which contains richest features of texture. In fact, iris appearance is genetically determined and will remain unchanged since infancy. Further more, iris is a well protected internal organ, and even genetically identical individuals have completely independent iris textures. These characters make iris patterns to be among the most accurate and effective biometrics [2]. The task of iris segmentation is to determine pixels in given images that belong to iris region, which plays a critical role in iris recognition. Having a good segmentation means that we can extract more discriminative features and then get a higher final recognition accuracy [3].

Besides the usage on biometric recognition, iris segmentation can also play a significant role for medical concerns. Accurate iris segmentation is an essential pre-processing step for high performance computer-aided ocular disease diagnosis.

For iris images taken under ideal conditions (e.g., no occlusion, proper distance, visual angle, and good illumination), iris segmentation can be simplified as a simple image-processing problem [4]. This is because iris region shows transparent variation from both pupil and sclera in ideal conditions. However, in daily usage scenario, image acquisition is not always that ideal, which makes iris segmentation a challenging task. The factors that render iris segmentation a challenging task can be briefly summarized as follows:

- *Occlusion:* In practice, iris images usually face the problem of occlusion caused by eyelids and/or eyelashes. Moreover, some users wear glasses or contact lenses, which also introduce complex occlusion and reflection problems.
- *User cooperation:* In unsatisfactory situations when users are uncooperative, the iris regions in obtained images face the issues of blurring, position offset, inappropriate scale, hard perspective, etc.
- *Illumination:* Poor or overexposed illumination will reduce images' contract, which results in non-sharp boundary of iris region.
- *Camera equipment:* There are two main types of devices for taking iris images, one using near-infrared (NIR) spectrum and another using visible light (VW). In NIR images, the outer iris

☆ This paper has been recommended for acceptance by Caifeng Shan.
* Corresponding author.
E-mail address: szlig@xmu.edu.cn (S. Li).

https://doi.org/10.1016/j.jvcir.2018.10.001
1047-3203/

boundary is not as distinguishable as the inner one. While in VW images, the inner boundary also faces the problem of low-contrast.

Besides the four key challenging factors mentioned above, there are still some issues that introduce challenges to iris segmentation, including pupil zooming, abnormal iris samples, etc.

Facing all these challenges, former iris segmentation approaches can be roughly divided into boundary-based methods and pixel-based methods. Boundary-based methods [5–7] require prominent contrast of structure components, while pixel-based methods [8–10] rely highly on discriminative iris feature. Both types gave poor performance when dealing with complicated situations, such as occlusion caused by eyelashes and reflections. In recent years, deep learning-based iris segmentation methods [11,12] outperform former non-deep methods in segmentation accuracy. But existing deep-learning based iris segmentation methods take whole eye image as input, which do not consider the geometric constraint information of eye. As we know, the iris is with circle-like shape and only occur in a specific area in the eye, as marked with red rectangles in Fig. 1. As a consequence, these methods will be affected by irrelevant areas and are insufficient to distinguish the iris/non-iris pixels around the boundary area. As shown in the second column of Fig. 1, noisy pixels around the boundary are wrongly segmented as iris region by U-Net [13] model which is one of the most promising methods for binary segmentation tasks. Even in some cases, pixels far away from the iris region are segmented as iris pixels. This issue mainly caused by the U-Net didn't consider the region structure information of the eye, and treat each pixel inside and outside the iris region equally. As such, the U-Net model will easily affect by the noisy pixels.

For addressing this problem, in this work, we propose an accurate and robust iris segmentation model, namely ATTention U-Net (ATT-UNet). Our ATT-UNet is built upon the original UNet model, while we make a significant improvement by introducing an attention mechanism which guides the model to learn more discriminative features for separating the iris and non-iris pixels. The attention mechanism is achieved by adding a branch to estimate the position of the iris in the whole image at the end of the contracting path, and then we generate a weighted attention mask to guide the expanding path that should pay more attention on this specific region as well as reduce the influence of other irrelevant regions. Under such a strategy, the proposed model learns from iris-specific regions to avoid false segmentation in noisy background. And the last column of Fig. 1 indicates that after adding attention mechanism, not only noisy pixels far from iris region, but also low discrimination pixels around iris boundary, are sharply reduced.

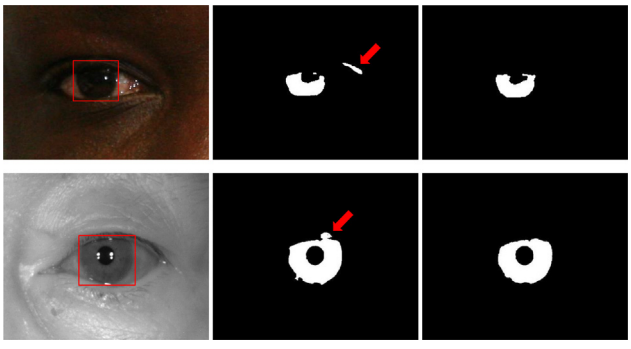

Fig. 1. Examples of segmentation results with/without attention mechanism. The first column displays original iris images selected from *UBIRIS.v2* and *CASIA-IrisV4-distance*. Areas marked with red rectangles indicate regions that need to pay more attention for accurate segmentation. The second column display segmentation predictions using U-Net without attention mechanism, and the locations indicated by red arrows show obvious errors in such situation. The last column displays the segmentation results using our proposed ATT-UNet. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We conduct several experiments of our method on two representative iris datasets, UBIRIS.v2 and CASIA.IrisV4-distance. Experiments results demonstrate that our proposed method can outperform former representative approaches. Besides, comparisons between U-Net and our ATT-UNet, with/without post-processing, shows the superiority of our method. The visual examples of zoomed-in output segmentation masks indicate that our model can deal with challenging situations.

The remainder of this paper is organized as follows. Section 2 introduces related works of iris segmentation in recent years. In Section 3, we describe the proposed ATT-UNet and the training pipeline. Section 4 talks about the details of experiment results. Finally, we conclude our paper in Section 5.

2. Related work

Iris segmentation was first developed from the requirement of iris recognition. As far as we know, the idea of distinguishing identity by iris texture was proposed by an ophthalmologist named Frank Burch in 1936 [14]. In 1994, Daugman [15] got a patent on iris recognition algorithm. Such algorithm first locates iris in an image, and encode its texture into 'iris code'. This method becomes the basis of existing iris recognition algorithm. Most of existing iris recognition algorithms follow 4 steps described below: (**a**) image acquisition; (**b**) iris segmentation; (**c**) feature extraction; (**d**) similarity discrimination. In recent years, most iris recognition approaches, including traditional approaches [16–18], and deep learning based approaches [19–21], still follow the above steps, in which iris segmentation is a critical step for subsequent operations like normalization, feature extraction, etc.

Because inaccuracy in segmentation can cause different mapping of the iris pattern in its extracted description, which can cause failure in recognition [22].

Proença et al. [23] proved that as more inaccuracies were introduced to segmentation results, the error rates of iris recognition dramatically raised.

There are three main types of existing iris segmentation methods: boundary-based methods, pixel-based methods and deep learning based methods. Boundary-based methods try to locate iris region by finding pupil boundary and limbus boundary. While pixel-based methods directly determine whether each pixel belongs to iris region. Deep learning-based methods can be roughly classified as pixel-based method. But due to their outstanding performance, we list them separately.

Boundary-based methods. Boundary-based methods try to locate iris region by finding pupil boundary and limbus boundary. It's natural to have the idea that iris region is annulus. But in reality, the inner and outer boundary of iris are usually not concentric. To avoid such mistake, Daugman's early work [5] simply regard iris region as circulars, which can be represented by two circles. He proposed a two-step iris segmentation method by repeatedly using an integro-differential operator to search over the image domain to find first the outer boundary, and then the inner boundary of iris region. Wildes et al. [6] adopted a gradient-based edge detection and then located outer and inner circles by using Hough transform [24,25]. Nearly all early methods were based on the assumption that iris had annulus-like boundary. However, usually, due to the reason of noise and occlusion, iris boundary, especially outer boundary, are not circular. So, Wildes et al. [6] parameterized upper and lower eyelids as parabolic arcs and located them by a gradient-based edge detector that favors the horizontal (based on

the assumption that head is upright). Methods proposed by Shah [7] and Daugman [22] used active contours to enhance iris segmentation, because such method allowed for non-circular boundaries and enabled flexible coordinate systems. Moreover, many approaches such as reflection removal [3], illumination normalization and coarse iris localization [26], boundary fitting model [27] have been introduced to improve the performance of boundary-based iris segmentation methods.

Pixel-based methods. Essentially, pixel-based methods try to construct classifiers for determining if each pixel is within iris region. Early promising pixel-based methods proposed by Pundlik et al. [8] adopted a step-wise procedure based on image intensities. In this method, a graph cut based energy minimization algorithm was used to separate first eyelash, then pupil, iris, and background. Tan and Kumar [9] exploited the localized Zernike moments feature [28,29] at different radii to classify each pixel into iris or non-iris category using support vector machines (SVMs). Proença [10] first introduced neural network to classify iris pixels. In this method, a shallow neural network, which contained only one hidden layer, was adopted for first sclera and then iris training/classification. The idea of first stage comes from the insight that sclera is the most distinguishable region in non-ideal images, and the mandatory adjacency of the sclera and the iris was exploited to detect noise-free iris regions.

Basically, boundary-based iris segmentation methods require prominent contrast of structure components. Gradient and contour information was concerned more in such methods. While pixel-based methods rely highly on discriminative features such as images' texture, color and intensity. Also, approaches such as [30,31] integrated these two kind of methods. [30] first roughly cluster image pixels into iris and non-iris regions by setting threshold on brightness, and then on the obtained coarse iris location image, an integro-differential model was adopted to locate iris boundary. While [31] first used Random Walker to locate coarse boundary circle of iris region, then a series of operations based on statistical gray level intensity information were adopted for pixel-level boundary refine.

Deep learning based methods. Benefited from large-scale data collection, rapid development on computing performance and fast GPU implementations of artificial neural networks [32–35], since 2010s, deep learning-based method dramatically boost in the field of computer vision, as well as the field of image segmentation [36–39]. Unlike traditional patch classification-based CNN models that using fully connected layers after convolutional layers to get fixed length feature vectors, FCN [39] allow arbitrary input image size and adopt deconvolution layer for upsampling the different convolutional layers' feature maps to target size. Compared with former approaches, FCN avoided separately running network for each patch, and boosted segmentation speed. Papers including [11,12] introduce modified FCN to the task of iris segmentation. However, because the feature maps in FCN for upsampling is too coarse, FCN's segmentation results are not fine enough. Unlike FCN that upsampling different size coarse feature maps to target resolution, U-Net [13] reform a lot in upsampling stage. U-Net adopt an encoder-decoder structure, and in U-Net's successive layers, pooling operators are replaced by a serial of upsampling operators, which makes the whole network a symmetrical U-shaped model. U-Net has been proven good performance in the field of biomedical [40–42], and can work with relatively few training images and yields more precise segmentations.

3. Proposed method

In this section, we describe the proposed ATT-UNet for addressing the task of iris segmentation in detail.

The proposed ATT-UNet is built upon the original UNet model, and we make a significant improvement by introducing an attention mechanism which guides the model to learn more discriminative features for separating the iris and non-iris pixels. We will first illustrate the architecture of the proposed ATT-UNet in Section 3.1, then we describe the novel training pipeline of ATT-UNet in Section 3.2.

3.1. Structure of ATT-UNet

The overall structure of our ATT-UNet is shown in Fig. 2, which contains a contracting path (marked as gray in Fig. 2) and an expanding path (marked as blue[1] in Fig. 2). The contacting path encodes feature maps of the CNN models, and at the end we add a regression module to estimate the bounding box of iris which is served as attention mask. The expanding path decodes the feature maps, and the attention mask is integrated into the model at the final prediction step. First we will introduce the U-Net Architecture in 3.1.1. Detailed introductions for regression part and attention-based segmentation part of our proposed ATT-UNet are described in 3.1.2 and 3.1.3.

3.1.1. U-Net architecture

U-Net [13] is a novel segmentation architecture built upon FCN. Unlike FCN, U-Net adopts a symmetry encoder-decoder structure, which also referred as contraction path and expanding path. The contracting path of U-Net (the gray-colored part without the bounding box regression part, in Fig. 2) follows typical CNN architecture, which contains a series of convolutional layers and max pooling layers. It gradually reduces feature maps' size and meanwhile increases the number of feature channels, that encourage the model to learn global and non-local features. While the expanding path of U-Net (the blue-color part in Fig. 2) contains a series of convolution and deconvolution operations, which can step-wise up-sampling the feature maps to the original size and reduces the feature channels. The skip connections between contacting and expanding path concatenate features from both sides which force the model to capture local and global information. Finally, a 1×1 convolutional layer is adopted to map feature vector to the desired number of classes. Such a neat symmetry architecture of U-Net has already achieved very promising performance on various biomedical segmentation applications.

3.1.2. Attention mask generation

As we know, our iris only occur in a specific area in the eyes, as marked with red rectangles in Fig. 1. In order to modeling this geometric constraint, we add an attention mask generation step to estimate the potential area where the iris most likely to appear. We use a bounding box regression module to estimate the coordinates as shown in the most right part in Fig. 2.

In our ATT-UNet, we add a pooling layer and a fully connected layer at the end the contracting path as a regression module. And the bounding box is determined by the coordinates of the upper left corner (x_1, y_1) and the bottom right corner (x_2, y_2) of the iris region. So the regression part of our model predict and output *rectangle coordinates arrays* consisting of (x_1, y_1, x_2, y_2). We adopt *Mean Squared Error (MSE)* as loss function in this step, and the equation goes as Eq. (1), in which n represents the number of rectangle array, Y_i and $\hat{Y}_i$ represent the i_{th} rectangle array of groundtruth and prediction, respectively.

$$MSE = \frac{1}{n} * \left(Y_i - \hat{Y}_i\right)^2 \tag{1}$$

[1] For interpretation of color in Figs. 2 and 3, the reader is referred to the web version of this article.

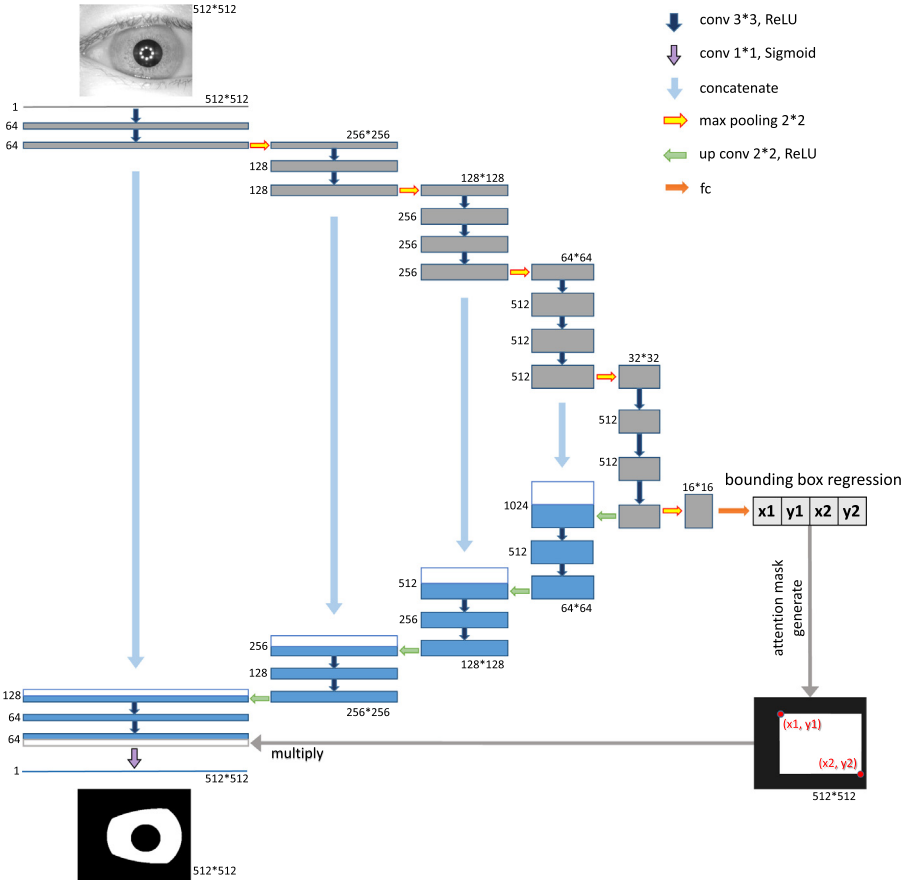

Fig. 2. ATT-UNet architecture. All the boxes correspond to multi-channel feature map. The contracting path of ATT-UNet use the same architecture as VGG16 (*without fc top*). The gray-colored part of ATT-UNet is also act as the bounding box regression model and is used for attention mask generation. The generated attention mask is merged back to ATT-UNet and make our model focus on the segmentation of iris region.

After rectangle arrays are predicted, in attention mask generation, we create a attention mask M.

3.1.3. Attention guided segmentation

After getting the predicted attention mask M, we use it to guide the final segmentation which forces the model to focus on this specific region. Instead of doing a hard attention that only segmenting pixels inside the mask, we utilize a soft attention scheme by setting the weights inside mask as 1.0 and outside as δ. By doing so, we can alleviate the mistake in the attention mask generation step to a certain degree.

The weight map is merged by multiplying with the ATT-UNet's second last layer's discriminative feature map, and the merged features is used for back propagation and act as key step for attention mechanism. The whole multiply operation can be expressed as (2):

$$\mathcal{F}(x,y) = \begin{cases} \mathcal{V}(x,y) * 1.0 & (x,y) \in M \\ \mathcal{V}(x,y) * \delta & (x,y) \notin M \end{cases} \tag{2}$$

in which $\mathcal{V}(x,y)$ represents the features in position (x,y), δ is the soft attention weight in range of $[0,1]$. In Section 4.4.2, we evaluate the effect of δ with different value on the segmentation performance, and found that the best performance was achieved by setting δ equals to 0.2 for UBIRIS.v2 and 0.6 for CASIA-IrisV4-Distance.

The whole architecture of our proposed ATT-UNet is shown in Fig. 2. Zero padding was adopted for all the convolutional layers to maintain the feature map. Binary cross entropy is used as segmentation loss function, which goes as Eq. (3),

$$L_H(x,z) = -\frac{1}{n}\sum_{k=1}^{n} z_k \log x_k + (1-z_k)\log(1-x_k) \tag{3}$$

where n represents the number of pixels in each image, x and z represent estimated foreground probability and its corresponded groundtruth label.

3.2. Training pipeline

In this section, we describe the detailed training pipeline of our ATT-UNet. Since the ATT-UNet has a contracting path and expanding path, we utilize a two-step training strategy to train our model. But it is also possible to train the whole network end-to-end.

During the first step, we only train the weights in contracting path (*the gray part in Fig. 2*) for computing the bounding box of attention mask. Firstly, we generate the groundtruth coordinates (x_1, y_1, x_2, y_2) by computing the upper left and bottom right corners of the external rectangle. Then we use the mean square error loss function as shown in Eq. (1) to train these weights.

For the second step, we only train the weights in the expanding path by keeping the weights in the contracting path fixed. For each image, we use the estimated coordinates (x_1, y_1, x_2, y_2) produced by the contracting path to compute the attention mask, then multiply with features from the penultimate layer. Finally, the binary cross-entropy loss function in Eq. (3) is used for training.

4. Experiments

4.1. Datasets

We use UBIRIS.v2 [43] and CASIA-IrisV4-Distance [44] dataset to evaluate the performance of our proposed method. Both datasets are representative dataset in the task of iris segmentation

and recognition. The detail parameters of these two datasets are listed in Table 1.

UBIRIS.v2 is an iris dataset which contains visible wavelength iris images captured on-the-move and at a distance, with corresponding more realistic noise factors. The pixel-wise segmentation groundtruth is given by NICE.I competition [45]. We use a subset of 500 images in UBIRIS.v2 for training and 500 for testing, following the same setting as the NICE.I competition.

CASIA-IrisV4-Distance is a subset of CASIA-IrisV4 which was collected by the Chinese Academy of Sciences' Institute of Automation (CASIA) using CASIA long-range iris camera. All the images are captured indoor with a distance of more than 2 m. In this dataset, a subset of 400 iris images are manually fine labeled by [11]. We use a subset of 300 images in this dataset for training and 100 for testing.

4.2. Evaluation metrics

Intersection over Union (IoU) is an evaluation metric to compare the predicted attention mask with the external bounding box of the groundtruth iris region. The equation of IoU is

$$IoU = \frac{area(P) \cap area(G)}{area(P) \cup area(G)}, \tag{4}$$

where P represent the predicted attention rectangle, and G is the groundtruth bounding box.

Mean Error Rate (MER) is a widely used evaluation metric for the task of binary segmentation. Error rate can be regarded as the ratio of all false pixel prediction in the whole image. Eq. (5) is used for calculating mean error rate on test set:

$$error = \frac{1}{N} \times \frac{I}{w \times h} \sum_{x=1}^{w} \sum_{y=1}^{h} M(x,y) \oplus G(x,y), \tag{5}$$

where N is the number of testing images, and w, h are the length and width of test images. M and G are the predicted segmentation mask and the groundtruth mask, and x, y are the coordinates of each pixel. The $\oplus$ represents the XOR operator, which calculate the dissimilar pixels between M and G.

Mean True Positive Rate (mTPR) is another commonly used evaluation metric for segmentation which computes the average ratio of predicted groundtruth pixels against the total groundtruth foreground pixels. The equation goes as Eq. (6),

$$mTPR = \frac{1}{N} \times \frac{TP}{TP + FN} \tag{6}$$

where N is the total number of testing images, TP and FN are the *true positive* and *false negative*.

4.3. Implementation

We implement our ATT-UNet on Keras with TensorFlow backend. The Adam optimizer is used to training our model with an initial learning rate equals to 0.0001. We set the number of training epochs to be 30. The contracting path in our ATT-UNet has the same structure as VGG16 (without all the fully connected layers), and we initialized the weights with the pre-trained model on ImageNet provided by Keras.

All the models were trained and tested by the pipeline discussed in Section 3.2 on a machine with Intel i7-7700K cpu and an NVIDIA 1080Ti GPU. We also apply data augmentation for pre-processing during training, including horizontal flip, width shift and height shift.

Table 1
Comparisons of two datasets used in our experiment.

Dataset	UBIRIS.v2	CASIA.v4-distance
Sensor	Canon EOS 5D	CASIA long-range iris cam
Environment	Indoor	Indoor
Type	VW	NIR
Resolution	$400 * 300$	$640 * 480$
Color	RGB	gray-level
No. of train	500	300
No. of test	500	100

4.4. Experimental results

4.4.1. Evaluation of the estimated attention masks

For the first part of experiment, we reported the quantitative IoU metric of estimated attention mask compared with the groundtruth in Table 2. As can be seen, the attention mask generate from our model can reach the IoU of 91.37% and 90.88% in UBIRIS.v2 and CASIA-IrisV4-distance respectively, which means the proposed ATT-UNet can accurately locate iris regions for attention.

We also plot some estimated attention masks in Fig. 3, which includes several challenging situations such as reflection, position offset, poor illumination and wearing glasses. As showed in Fig. 3, the red bounding box generated by our model is well aligned with the blue groundtruth bounding box, but there are some of small errors in some challenge cases. As a consequence, we erode the estimate attention masks by 5 pixels, which can get a better localization of the iris regions for the next segmentation step.

4.4.2. Comparison of different soft attention weight

In this section, we compare the effect of different soft attention weight of our ATT-UNet model. Instead of hard attention that only segment pixels inside the mask, we test the background attention weight from 0.0 to 1.0 with an interval of 0.2, and fix the foreground attention weight as 1.0. The MER and mTPR performance of different attention weight of the dataset of UBIRIS.v2 and CASIA-IrisV4-Distance are displayed in Fig. 4. For better display, we plot the (1.0-MER) instead. As can be seen from Fig. 4, since we can get a very high MER, different attention weights have a very small influence of the MER. For the mTPR metric, we can get a significant improvement at the attention weight of 0.2 for UBIRIS.v2 dataset, and 0.6 for CASIA-IrisV4-Distance dataset. It is worth noting that even we choose the worst performing weight in both datasets, the proposed ATT-UNet can achieve better MER and mTPR than other methods.

4.4.3. Evaluation of the iris segmentation

In this section, we evaluate the segmentation performance of our proposed ATT-UNet by using the best attention weights of each dataset (0.2 for UBIRIS.v2 and 0.6 for CASIA-IrisV4-Distance). Firstly, we implemented a U-Net model without the bounding box regression step as showed in Fig. 2 as the baseline model. Besides we also implement an FCN for comparison. All these three models use the backbone of VGG16 and initialized with the weights pre-trained on ImageNet. Same data augmentation procedures are adopted during the training. We adopt mean error rate and mTPR as evaluation metrics, as is introduced in Section 4.2. Table 3 list a summary of segmentation performance between the proposed ATT-UNet, U-Net and FCN. The proposed ATT-UNet achieves a mean error rate (MER) of 0.764% and 0.381% on UBI-

Table 2
Attention masks' IoU results between predicted attention rectangle and groundtruth rectangle on both datasets.

Dataset	UBIRIS.v2	CASIA.v4-distance
IoU (%)	91.37	90.88

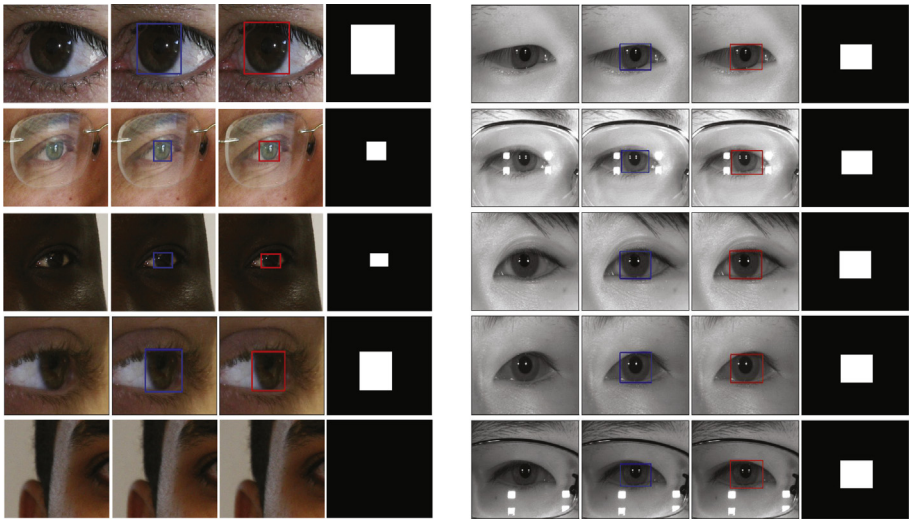

Fig. 3. Examples of attention mask generation results. The left part is example results from UBIRIS.v2 and the right part is from CASIA-IrisV4-Distance. The first column of each part is original iris images from each dataset, and the second column is the iris region bounding boxes given by finding max contour on segmentation masks. The example bounding boxes predicted by gray part of the proposed ATT-UNet are shown in the third column. The fourth column is examples of generated attention mask according to coordinates of bounding box, following the setting of Eq. (2). For better locating iris region, we expand the predicted attention rectangles by 5 pixels.

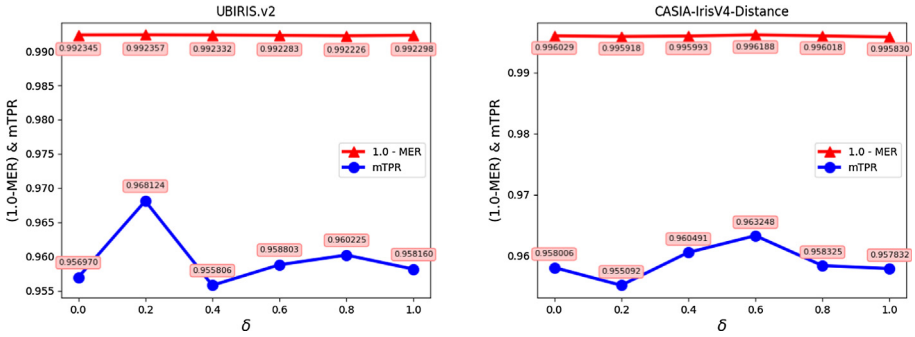

Fig. 4. The MER & mTPR metrics of the segmentation results under different soft attention weight for the UBIRIS.v2 and CASIA-IrisV4-Distance datasets. For better display, we use (1.0-MER) instead.

Table 3
Mean error rate and mTPR comparisons on proposed ATT-UNet and two promising network in the field of segmentation, including U-Net and FCN. Bold values represent the best results in the comparison methods.

Method	UBIRIS.v2		CASIA.V4- distance	
	MER (%)	mTPR (%)	MER (%)	mTPR (%)
U-Net	0.898	94.804	0.478	95.026
FCN	1.004	95.482	0.528	95.014
Ours ATT-UNet	**0.764**	**96.812**	**0.381**	**96.325**

RIS.v2 and CASIA.IrisV4-Distance, which are lower than the U-Net and FCN. And for the mTPR, our ATT-UNet can get a higher performance of 96.812% and 96.325% respectively. It shows that our proposed ATT-UNet surpass U-Net and FCN in both datasets on two evaluate metrics.

In Fig. 5, we plot some typical segmentation results produced by our proposed ATT-UNet compared and the original U-Net. Compared with the U-Net, the proposed ATT-UNet model can avoid some obvious mistakes as indicated by yellow arrow in sample ①, as well as some hard mistakes such as the indistinguishable edges as indicated in ②, ③, ④.

In Fig. 6, we gave examples of segmentation results from UBIRIS.v2 and CASIA.IrisV4-distance datasets. The first column is the original images, the second column is the groundtruths, and the third column is the segmentation results of our ATT-UNet. In the last column, we labeled those false segmented pixels in red.

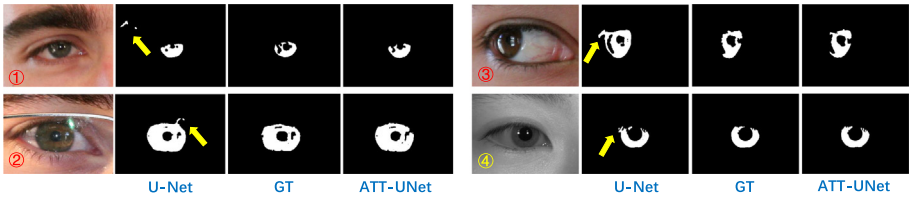

Fig. 5. We compare some typical segmentation results from both dataset, under original U-Net and the proposed ATT-UNet. The groundtruths are listed in the third column. The yellow arrows point out some obvious errors in U-Net segmentation results. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

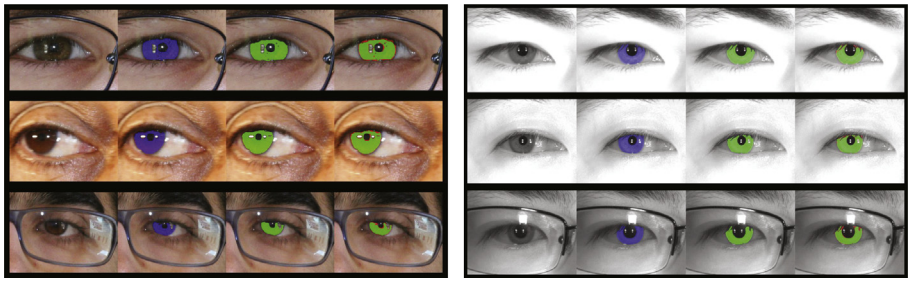

Fig. 6. Examples of segmentation results from UBIRIS.v2 and CASIA-IrisV4-distance. The first columns of both parts are original iris images selected from UBIRIS.v2 and CASIS. IrisV4-distance, respectively. The second and the third column respectively display the groundtruths (marked with blue color) and segmentation predictions (marked with green color). The correct and false segmented pixels are labeled as green and red color, respectively in column 4. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In order to have a better observation of those segmentation errors, we also select some challenging samples and plot the zoomed segmentation results from two datasets in Fig. 7. Here, challenging samples with hair occlusion (see sample ① ②), eyelashes occlusion (see sample ①⑤⑥⑦⑧⑨⑩), strong reflection (see sample ③④⑤⑥⑦⑨⑩), eyeglasses occlusion (see sample ③④ ⑦⑨⑩), non-cooperative (see sample ③⑤), blurring (see sample ⑤), are included. Even in sample ②, our model successfully distinguish a very slim and inconspicuous hair, as indicated by yellow arrow. Results show that the proposed ATT-UNet can deal with non-ideal and non-cooperative images and give pretty good segmentation results. Additionally, although time efficiency is not the main concern of our task, we achieved segmentation speed of 17 fps on $512 * 512$ resolution iris images, with one NVIDIA 1080Ti GPU.

However, there are still some very extreme samples that our model cannot perfectly segment. In Fig. 8, we select some weakly segmented samples and display them in zoomed-in manner. Samples in UBIRIS.v2 dataset are VW images, so there are many problems caused by low contrast. In the area indicated by yellow arrow in sample ①, the light reflection is very weak and is understandable to be segmented as foreground. In sample ②, our model fail to segment a very thin hair. In sample ③, the mis-segmented areas have very low contrast. Samples in CASIA-IrisV4-distance dataset are NIR images with high contrast, but many images still suffer from the effects of dense drooping eyelashes, as indicated by yellow arrow in sample ④. Even if we relabel the dataset with human eye, those samples are still questionable and worth discussing.

Since iris is only one large connected region, after getting the segmentation results, it's possible to remove those isolated false

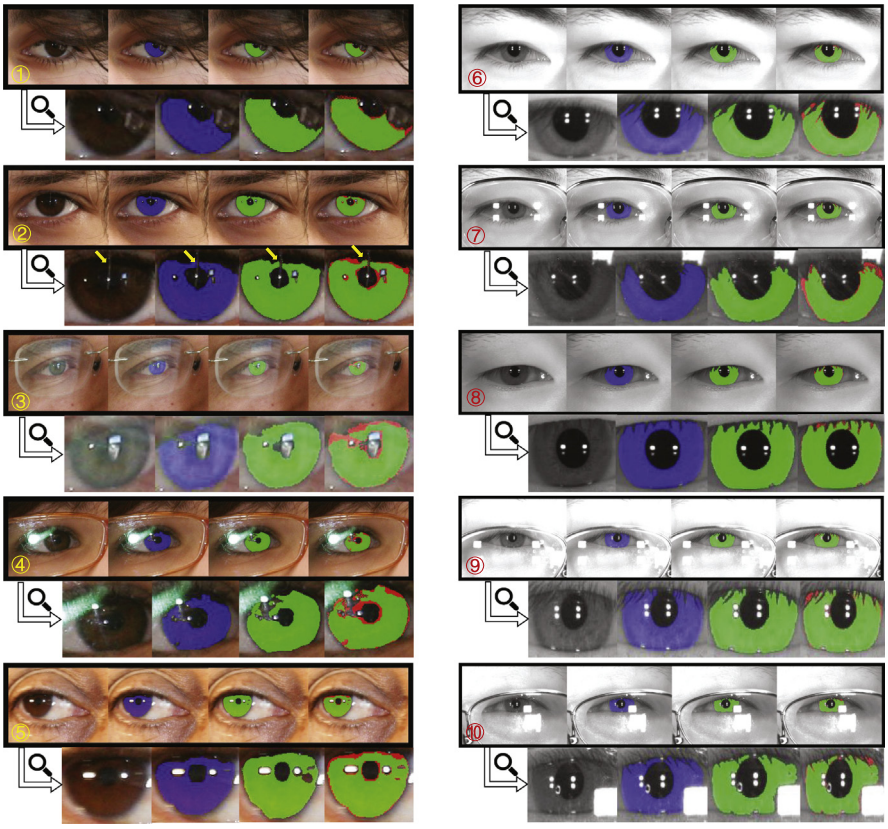

Fig. 7. Some challenging samples and their segmentation results are selected and displayed from two datasets. The first columns of both side are original iris images. The second and the third column respectively display the groundtruths (marked in blue) and segmentation predictions (marked in green). The false segmented pixels are labeled as red color in forth column. All the images are zoomed in for detail. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

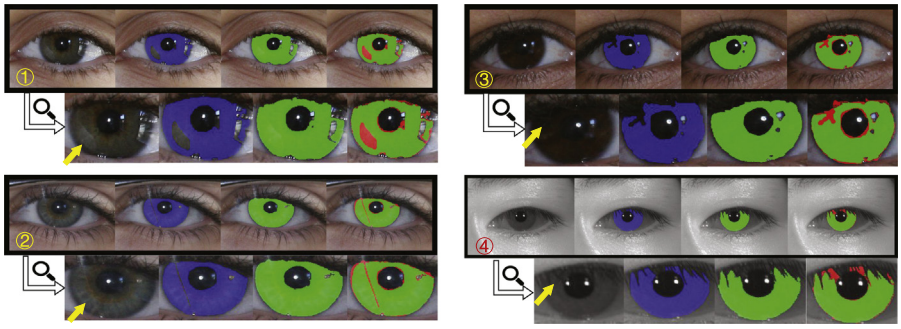

Fig. 8. Some weakly segmented samples are selected and displayed from both dataset. Just like 7, the first column are original iris images. The second and the third column respectively display the groundtruths (marked in blue) and segmentation predictions (marked in green). The false segmented pixels are labeled as red color in forth column. All the images are zoomed in for detail. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

positive pixels far away from iris region by only keeping the largest connect component. In this part, we also did a comparison of the U-Net and ATT-UNet with/without the post-processing step. The MER and mTPR are list in Table 4. From the result, we can find that using the post-processing step don't have much effect of the experiment results. For the U-Net, the post-processing slightly improves the performance. While for ATT-UNet, post-processing even decreases the segmentation performance. The main reason is that our ATT-UNet already handle this problem by ignoring those false positive pixels outside iris region in the attention step.

4.4.4. Comparison with other methods

We also compare the proposed method with several representative and excellent iris segmentation approaches in recent years, including boundary-based approaches [26], pixel-based approaches [9,10] and deep learning based approaches [11]. In particular, [31] is a composite model that integrated boundary-based and pixel-based methods. Specifically, [11] proposed a multi-scale fully convolutional network for iris segmentation, and achieved former state-of-the-art results in both datasets. Table 5 list a summary of mean error rate, as is given in 4.2, for our proposed method and other approaches. Results show that the proposed ATT-UNet achieves the state-of-the-art performance with mean error rate 0.76% and 0.38% on UBIRIS.v2 and CASIA.IrisV4-Distance dataset respectively. Moreover, the improvements of proposed method over former state-of-the-art method [11] is respectively 15.56% and 35.59%.

Table 4
Comparison of with/without post-processing (pp) step in the original U-Net and proposed ATT-UNet. Bold values represent the best results in the comparison methods.

Method	UBIRIS.v2		CASIA.v4- distance	
	MER (%)	mTPR (%)	MER (%)	mTPR (%)
U-Net	0.8977	94.8045	0.4779	95.0263
U-Net + PP	0.8851	94.6778	0.4722	95.0202
ATT-UNet	**0.7643**	**96.8124**	**0.3812**	**96.3248**
ATT-UNet + PP	0.7679	96.7839	0.3848	96.3201

Table 5
Error rate comparisons on proposed method and other relatively approaches. The '–' means the method has not been implemented on that dataset. Bold values represent the best results in the comparison methods.

Method	UBIRIS.v2 MER (%)	CASIA.v4-distance MER (%)
Ours ATT-UNet	**0.76**	**0.38**
MFCNs [11]	0.90	0.59
RTV-L^1 [26]	1.21	0.68
Tan et al. [31]	1.72	0.81
Proença et al. [10]	1.87	–
Tan et al. [9]	1.90	1.13

5. Conclusions

Accurate and effective iris region segmentation play as a significant step in iris recognition, as well as computer-aided ocular disease diagnosis. For better iris segmentation, challenges such as occlusion, reflection, poor illumination should be overcome. Former learning-based methods generally use global iris images as input for learning. Given that iris region's appearance is a unitary area, pixels outside iris-specific region is typically of no use for segmentation results. Training deep models with global iris images can be easily effected by noisy pixels in complicated iris images. In this paper, we present an accurate network model for iris segmentation, namely ATT-UNet. The proposed model learns from global information and iris-specific local region to avoid false segmentation caused by noisy pixels. We design a novel training strategy. First, the information of labeled segmentation masks is fully utilized for ROI groundtruth generation. Then, adopting the contracting path of proposed ATT-UNet as bounding box regression model, we generate attention masks to merge with discriminative feature maps in model, making the proposed model pay more attention on iris-specific region and avoid false segmentation in noisy background.

We adopt mean IoU for measuring the accuracy of generated attention mask. Also, mean error rate and mTPR are used for measuring segmentation results. Experimental results show that iris-specific bounding box in complicated iris images can be correctly located. And iris images in VW and NIR, including non-cooperative and non-ideal samples, can be well segmented with ATT-UNet. Further experiments show that the proposed model surpasses representative models in accuracy. Results of U-Net, with or without attention mechanism, indicate that the segmentation performance is improved after adding attention mechanism to U-Net. Since the proposed approach has not been set as end-to-end mode, further research will focus on optimizing model pipeline and making models more robust in different scenery.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 61572409, No. U1705286 & No. 61571188), Fujian Province 2011 Collaborative Innovation Center of TCM Health Management and Collaborative Innovation Center of Chinese Oolong Tea IndustryCollaborative Innovation Center (2011) of Fujian Province, Fund for Integration of Cloud Computing and Big Data, Innovation of Science and Education.

References

[1] S. Prabhakar, S. Pankanti, A.K. Jain, Biometric recognition: security and privacy concerns, IEEE Symp. Secur. Privacy 1 (2003) 33–42.
[2] A.K. Jain, Biometric recognition: how do i know who you are? in: International Conference on Image Analysis and Processing, 2005, pp. 1–5.
[3] Z. He, T. Tan, Z. Sun, X. Qiu, Toward accurate and fast iris segmentation for iris biometrics, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2009) 1670–1684.
[4] K.W. Bowyer, M.J. Burge, Handbook of Iris Recognition, Springer, 2016.
[5] J.G. Daugman, High confidence visual recognition of persons by a test of statistical independence, IEEE Trans. Pattern Anal. Mach. Intell. 15 (1993) 1148–1161.
[6] R.P. Wildes, J.C. Asmuth, G.L. Green, S.C. Hsu, R.J. Kolczynski, J.R. Matey, S.E. McBride, A machine-vision system for iris recognition, Mach. Vision Appl. 9 (1996) 1–8.
[7] S. Shah, A. Ross, Iris segmentation using geodesic active contours, IEEE Trans. Inf. Forensics Secur. 4 (2009) 824–836.
[8] S.J. Pundlik, D.L. Woodard, S.T. Birchfield, Non-ideal iris segmentation using graph cuts, in: IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2008, pp. 1–6.
[9] C.-W. Tan, A. Kumar, Unified framework for automated iris segmentation using distantly acquired face images, IEEE Trans. Image Process. 21 (2012) 4068–4079.
[10] H. Proenca, Iris recognition: on the segmentation of degraded images acquired in the visible wavelength, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 1502–1516.
[11] N. Liu, H. Li, M. Zhang, J. Liu, Z. Sun, T. Tan, Accurate iris segmentation in non-cooperative environments using fully convolutional networks, in: IEEE International Conference on Biometrics, 2016, pp. 1–8.
[12] S. Bazrafkan, S. Thavalengal, P. Corcoran, An end to end deep neural network for iris segmentation in unconstraint scenarios, 2017. Available from: arXiv preprint <arXiv:1712.02877>.
[13] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.
[14] K. Irsch, D.L. Guyton, Anatomy of eyes, Encycl. Biometr. (2009) 11–16.
[15] J.G. Daugman, Biometric personal identification system based on iris analysis, 1994. US Patent 5,291,560.
[16] P.R. Nalla, A. Kumar, Toward more accurate iris recognition using cross-spectral matching, IEEE Trans. Image Process. 26 (2017) 208–221.
[17] J. Chen, F. Shen, D.Z. Chen, P.J. Flynn, Iris recognition based on human-interpretable features, IEEE Trans. Inf. Forensics Secur. 11 (2016) 1476–1485.
[18] N. Othman, B. Dorizzi, S. Garcia-Salicetti, Osiris: an open source iris recognition software, Pattern Recogn. Lett. 82 (2016) 124–131.
[19] F. He, Y. Han, H. Wang, J. Ji, Y. Liu, Z. Ma, Deep learning architecture for iris recognition based on optimal gabor filters and deep belief network, J. Electron. Imaging 26 (2017) 023005.
[20] K. Nguyen, C. Fookes, A. Ross, S. Sridharan, Iris recognition with off-the-shelf cnn features: a deep learning perspective, IEEE Access 6 (2018) 18848–18855.
[21] A. Gangwar, A. Joshi, Deepirisnet: deep iris representation with applications in iris recognition and cross-sensor iris recognition, in: IEEE International Conference on Image Processing, 2016, pp. 2301–2305.
[22] J. Daugman, New methods in iris recognition, IEEE Trans. Syst., Man, Cybernet., Part B (Cybernet.) 37 (2007) 1167–1175.
[23] H. Proença, L.A. Alexandre, Iris recognition: analysis of the error rates regarding the accuracy of the segmentation stage, Image Vis. Comput. 28 (2010) 202–206.
[24] H.P. VC, Method and means for recognizing complex patterns, 1962. US Patent 3,069,654.
[25] D.H. Ballard, Generalizing the hough transform to detect arbitrary shapes, Pattern Recogn. 13 (1981) 111–122.
[26] Z. Zhao, K. Ajay, An accurate iris segmentation framework under relaxed imaging constraints using total variation model, in: IEEE International Conference on Computer Vision, 2015, pp. 3828–3836.
[27] A. Uhl, P. Wild, Weighted adaptive hough and ellipsopolar transforms for real-time iris segmentation, in: IEEE International Conference on Biometrics, 2012, pp. 283–290.
[28] A. Khotanzad, Y.H. Hong, Invariant image recognition by zernike moments, IEEE Trans. Pattern Anal. Mach. Intell. 12 (1990) 489–497.
[29] A. Fathi, P. Alirezazadeh, F. Abdali-Mohammadi, A new global-gabor-zernike feature descriptor and its application to face recognition, J. Vis. Commun. Image Represent. 38 (2016) 65–72.
[30] T. Tan, Z. He, Z. Sun, Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition, Image Vis. Comput. 28 (2010) 223–230.
[31] C.-W. Tan, A. Kumar, Towards online iris and periocular recognition under relaxed imaging constraints, IEEE Trans. Image Process. 22 (2013) 3751–3765.
[32] K.-S. Oh, K. Jung, Gpu implementation of neural networks, Pattern Recogn. 37 (2004) 1311–1314.
[33] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
[34] D.C. Ciresan, U. Meier, J. Masci, L. Maria Gambardella, J. Schmidhuber, Flexible, high performance convolutional neural networks for image classification, in: International Joint Conference on Artificial Intelligence, vol. 22, 2011, p. 1237.
[35] S. Yu, Y. Cheng, L. Xie, Z. Luo, M. Huang, S. Li, A novel recurrent hybrid network for feature fusion in action recognition, J. Vis. Commun. Image Represent. 49 (2017) 192–203.
[36] D. Ciresan, A. Giusti, L.M. Gambardella, J. Schmidhuber, Deep neural networks segment neuronal membranes in electron microscopy images, in: Advances in Neural Information Processing Systems, 2012, pp. 2843–2851.
[37] C. Farabet, C. Couprie, L. Najman, Y. LeCun, Learning hierarchical features for scene labeling, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2013) 1915–1929.
[38] P. Pinheiro, R. Collobert, Recurrent convolutional neural networks for scene labeling, in: International Conference on Machine Learning, 2014, pp. 82–90.
[39] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.
[40] H. Dong, G. Yang, F. Liu, Y. Mo, Y. Guo, Automatic brain tumor detection and segmentation using u-net based fully convolutional networks, in: Annual Conference on Medical Image Understanding and Analysis, Springer, 2017, pp. 506–517.
[41] R. Mehta, J. Sivaswamy, M-net: a convolutional neural network for deep brain structure segmentation, in: IEEE International Symposium on Biomedical Imaging, 2017, pp. 437–440.
[42] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3d u-net: learning dense volumetric segmentation from sparse annotation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 424–432.
[43] H. Proenca, S. Filipe, R. Santos, J. Oliveira, 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. 32 (2010) 1529–1535.
[44] CASIA Iris Image Database, 2010. <http://biometrics.idealtest.org/>.
[45] H. Proença, L.A. Alexandre, The nice. i: noisy iris challenge evaluation-part i, in: IEEE International Conference on Biometrics: Theory, Applications, and Systems, 2007, pp. 1–4.