

Deep Neural Networks for Accurate Iris Recognition

Yuzheng Xu
National Tsing Hua Univ.
Hsinchu, Taiwan
blessxu@gmail.com

Tzu-Chan Chuang
National Tsing Hua Univ.
Hsinchu, Taiwan
alice20041025@gmail.com

Shang-Hong Lai
National Tsing Hua Univ.
Hsinchu, Taiwan
lai@cs.nthu.edu.tw

Abstract

Most prior iris recognition techniques based on the existing pipeline have already reached their limits. Therefore, this work explores the possibility of applying the deep learning technique to the field of iris recognition. We combine a novel segmentation network with a modified resnet-18 as the iris matching network. The segmentation network architecture consists of an iterative altered FCN (fully convolutional network) which contains a path of contracting layers to capture features and a symmetric upsampling path that gives precise pixel-to-pixel localization. The network not only generates visually implausible iris masks but also makes good use of data augmentation. We show that combining such networks outperforms the prior methods on several iris image datasets, including CASIA V3-interval and UBIRIS V2 datasets.

1. Introduction

Due to the complexity, uniqueness and stability of human iris, iris recognition has been recognized as one of the most accurate approaches for automated biometric identification. The very first complete and automated iris recognition system was introduced by Daugman [1] in 1993, who also devised the most classical algorithm, Integro-differential operator [2], for iris segmentation under NIR illumination. The accuracy of segmentation has a great impact on the performance of the subsequent verification, thus several researches focused on improving iris segmentation and have reached state-of-the-art results in recent years [3] [4] [5] [6]. In this paper, we propose both iris segmentation and verification methods that exceed previous approaches.

Most of the previous works on iris segmentation used traditional pipeline which mainly contains pre-processing, localization and post-processing. Traditional localization algorithms are based on integro-differential operators [2] or Hough transforms [7]. Integro-differential operator searches for the circular path where the maximal change in pixel values occurs by varying the radius and center of

the circular contour to achieve precise location of eyelids. Hough transforms find optimal curve parameters by a voting procedure in edge maps which is generated by calculating the first derivatives of intensity values in an eye image and then thresholding the result. However, these two traditional algorithms perform poorly in unconstrained iris images and had been improved or replaced by many other algorithms proposed recently [8] [4].

Different from traditional iris localization, pixel-to-pixel iris segmentation extracts discriminative appearance features in the neighborhood of pixels and builds classifiers for pixel-wise classification [9] [10]. Nowadays, deep learning has gained much attention and achieved great success in image segmentation. Convolutional Neural Networks (CNNs) [11] allow us to train end-to-end model which can learn optimal features and classifiers automatically and Fully Convolutional Networks (FCNs) [12] enable pixels-to-pixels prediction. In this paper, we modify FCNs to accurately generate pixel-wise iris segmentation prediction.

For iris verification or recognition tasks, most of the previous works are based on hand-crafted features [1][13][14], which reached state-of-the-art performance. Recently, deep learning was also utilized by many approaches and obtained significant improvements [15] [16]. However, even with a good segmentation result which directly gives the region of iris, they still transformed the iris region to a rectangle image. For example, [1] applied 2-D Gabor wavelet to a rectangle to obtain "iris code", and then compared them with hamming distance. [15] used Convolutional Neural Network (CNN) to compute a vector measuring the distance between different images. In this paper, we explore the utilization of iris region given by FCNs.

2. Iris Segmentation

In this section, a brief introduction of Fully convolutional networks (FCNs) [12] is presented first. Afterwards, we provide clear details of the proposed iris segmentation network, including generic layout of our network architecture and the training procedures. The experimental results and the discussion will be presented along with the recognition

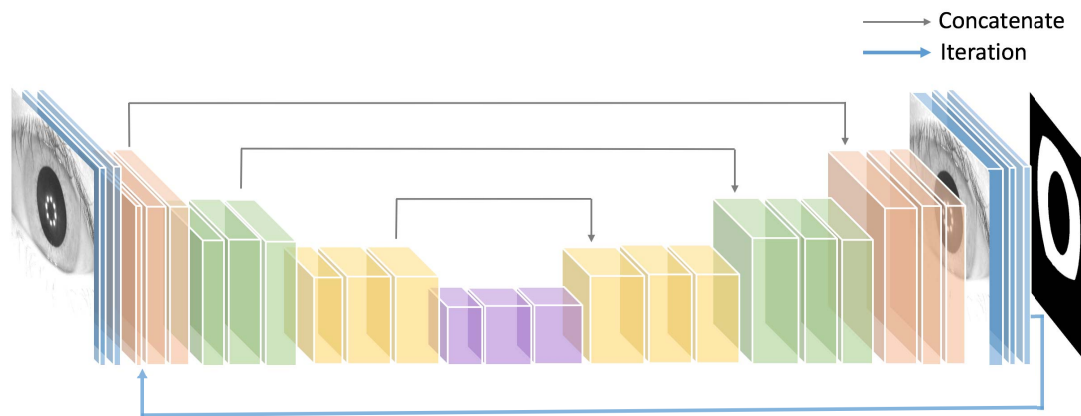


Figure 1. The architecture of our proposed segmentation network. Each colored cube is a multi-channel feature map. There are totally 44 convolutional layers and 8 pooling layers.

results in section 4.

2.1. Network Architecture

The precise architecture of the proposed iris segmentation network is shown in Figure 1. Our segmentation architecture is built upon FCNs. We decompose the prediction process into two steps: we modified and combined FCNs with iterative process such that it generates more precise image segmentation. The main idea of our work is to integrate the benefits of some network architecture and utilize them into the iris segmentation domain.

Fully convolutional networks can be trained end-to-end specifically for generating detection proposals. FCNs were originally introduced for the task of semantic segmentation which predicts dense outputs from arbitrary-sized inputs. They modified deep classification architectures by using image classification as supervised pre-training and fine-tune fully convolutionally to learn segmentation map efficiently from whole image inputs and ground truths [12]. The key contributions in [12] are the extension of the pixel-to-pixel convolutional networks to the whole network architecture and replacing the pooling operators by upsampling operators. These layers increase the resolution of the output. One advantage of FCNs is that it allows for input of arbitrary sizes and outputs the segmentation map of the same size. Therefore, we take advantage of the fully-convolutional architecture, and our segmentation method allows for input with arbitrary size and produces the iris map of the same size.

Unlike the upsampling approach in [12], we also altered the upsampling method by retaining the number of feature maps in the corresponding downsampling layers. The original network of FCNs combines coarse, high layer information with fine, low layer information in the prediction

layers. They adopt 1x1 convolution with fixed channel dimension to predict classification scores for each class and bilinearly upsample the coarse outputs to dense outputs by deconvolution layer. However, our upsampling approach which merges current feature channels with the corresponding feature channels brings more information of the down-sampling layers to the higher resolution layers. Moreover, in order to predict the missing pixels around the border of iris, we combine the input image with the last upsampling layer in each iteration.

After adapting and amending the FCNs, we found that it was still not enough to precisely predict the border region of iris. To this end, we devise the method from the inspiration of prior iterative approaches, such as iterative instance segmentation [17] and iterative error feedback (IEF) [18]. Consequently, instead of learning to predict the result in one single step, we developed an iterative version of FCNs. It shows that the results were further improved by the iterative refinement. With the iterative step, more accurate results can be generated with border region refined.

The proposed segmentation network consists of an iterative two-step FCNs which has one downsampling path that follows the architecture of a typical convolutional neural network (CNNs) [11] and one upsampling path modified from the original FCNs in each step. In the downsampling path, a 2x2 max pooling operations with stride 2 is applied after every two convolutional layers of 3x3 convolutions that are followed by a Rectified Linear Unit (ReLU) in each downsampling step. After each max pooling operation, the feature channels doubles. On the other hand, in the upsampling path, each step contains an upsampling operation followed by a 2x2 convolution along with a concatenation of the symmetrically corresponding feature map from the downsampling path, and two convolutional layers

of 3x3 convolutions with a ReLU after that. Each 2x2 convolution in the upsampling step halves the feature channels. On the last layer of the whole network, we use a 1x1 convolution and sigmoid activation to output the final prediction. Our network does not contain fully connected layers.

2.2. Training

The implementation of our segmentation network is based on Keras. During training, optimization is performed using adam algorithm [19]. The input to the network is first resized to 256 x 256 x original feature channels, the output segmentation image is equal to 256 x 256 x 1 before resizing the image back to the input image size. We set the batch size into 16 patches with a learning rate of 1×10^{-5} . Finally, The loss at each pixel for training is defined using negative of dice coefficient inspired by the evaluation formula of ultrasound nerve segmentation competition on kaggle [20]:

$$L(X, Y) = -\frac{2 \times |X \cap Y|}{|X| + |Y|} \quad (1)$$

where X, Y are the predicted set of pixels and the ground truth, respectively. The dice coefficient is defined to be 1 when both X and Y are empty.

3. Iris Verification

3.1. Verification Architecture and Training

For most of the existing methods, iris images are normalized to polar coordinates (i.e., Daugman's rubber sheet model). In this paper, instead of normalizing the images, we also directly segment the iris part. We treat the iris verification as a patch match task. Using a resnet-18 [21] based 2-channel siamese network, we receive a 90.85% intra-class accuracy and a 99.59% inter-class accuracy. In the case of using normalized image, we receive a 94.31% intra-class accuracy and a 98.83% inter-class accuracy. We finally combine them to get a 95.26% intra-class accuracy and a 99.33% inter-class accuracy.

3.2. Verification Data Preparation

The original image, the result of segmentation and the iris part image are illustrated in figure 2. By treating the segmentation image as a mask, we can directly extract the iris region from the original eye image. First, we remove the useless black part to obtain a minimum square that contains the iris and then resize the image to 224x224.

Initially, We have 2,639 images with 498 eyes and the training of network needs intra-class pairs and inter-class pairs, thus we remove all the eye classes with only 1 image since they cannot construct intra-class pairs. Besides, considering the classes with the number of images less than 6, they are selected as testing data.

We start with 2,252 images for training and 364 images

for testing. To generate the training data, we utilize all the intra-class pairs, because the intra-class pairs are usually far less than inter-class pairs. Then, we apply a random rotation 10 times with angles between -20 and 20 degrees for up-sampling, and we use the first three images of each class to generate inter-class pairs. The number of inter-class pairs is almost 2 times more than intra-class pairs. In our experiment, this ratio leads to good result. The centers of pupil are aligned to a fixed point in the pre-processing step. For the testing data, we cut some additional part to reduce shape difference, omitting the black part, to make the two images in each pair have the same length. Finally, we have 213,510 images for training and 64,261 images for testing. An example image pair is depicted in figure 3.

By using the same pre-processing step in [1], we obtain the normalized iris images in the log-polar coordinate as the input images to our second CNN model. The model architecture for the second model is the same as that of the first one. It is easier to normalize the eye image after obtaining the iris segmentation results 4.

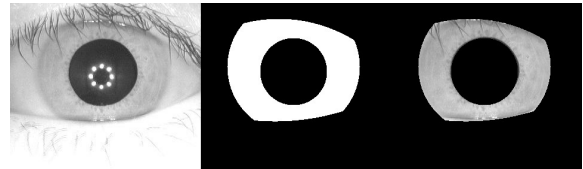


Figure 2. Images from left to right are the original image, generated segmentation mask and the extracted iris region.

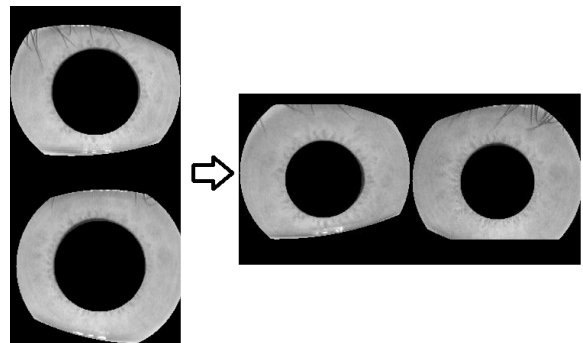


Figure 3. An example of image pair before and after the pre-processing step.

3.3. Network

We apply a 2-channel network[22] with resnet-18[21] as our verification network. The two eye images are stacked as 2 channels in one image, i.e. every channel is one eye image, because our iris image is gray scale. The architecture is illustrated in figure 5. For images from the same eye, the label is 1, otherwise it is -1. The loss function used for training the iris verification network is the soft-margin Loss,

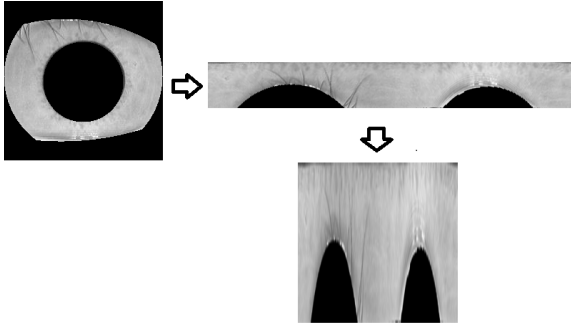


Figure 4. An example of image normalization as the pre-processing step for preparing training data.

given by

$$\text{loss}(x, y) = \log(1 + \exp(-y * x)) \quad (2)$$

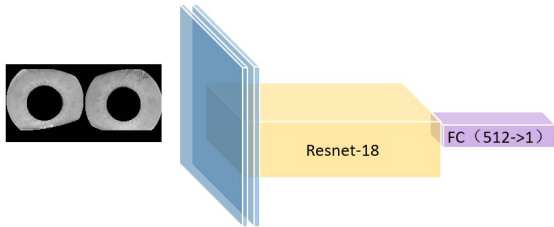


Figure 5. The network architecture of the iris verification network

4. Experimental Results

4.1. Experimental Datasets

In our experiments, we used two datasets, i.e., UBIRIS.v2 [23] and CASIA-Iris-Interval[24]. UBIRIS.v2 is an iris dataset which is acquired with visible light illumination. It consists of 11,102 images from 261 subjects. The images are captured on non-constrained conditions with subjects at-a-distance and on-the-move. Therefore, the images in UBIRIS.v2 often contain serious occlusions, specular reflections, off-axis and blur. Iris images in CASIA-Iris-Interval dataset were captured by Center for Biometrics and Security Research's self-developed close-up iris camera. There are totally 2,639 images collected from 249 people. The images of CASIA-Iris-Interval are very clear iris images with detailed texture since they designed a circular NIR LED array, with suitable luminous flux for iris imaging.

4.2. Performance Evaluation

4.2.1 Iris Segmentation

The segmentation ground truth of 2,250 images collected from 104 people in the UBIRIS.v2 dataset are manually la-

Table 1. Average iris segmentation error (%) for each of the state-of-the-art approaches. The symbol ‘-’ in the table indicates that the accuracy of applying the method in the corresponding dataset is not available.

Average iris segmentation error (%)		
Approach	UBIRIS.v2	CASIA-Iris-Interval
Proposed method	0.81	1.60
MFCNs [3]	0.90	-
Z. Zhao and A. Kumar, ICCV [4], 2015	1.21	-
T-IP 2013[5]	1.72	-
T-PAMI 2013[6]	1.92	-

beled and publicly provided by WaveLab, Salzburg University [25], and the hand-made ground truth for the total 2,639 images in the CASIA-Iris-Interval dataset are made publicly available by Halmstad University [26].

Since the evaluation protocol provided in NICE I competition [27] is widely used to measure the performance of iris segmentation method, the accuracy of our iris segmentation is computed using the same protocol as that in the NICE I competition as follows:

$$\bar{e} = \frac{1}{N \times w \times h} \sum_{p \in w} \sum_{q \in h} G(p, q) \oplus M(p, q) \quad (3)$$

where w and h are width and height of an iris image, N is the total number of images, and G and M are the ground truth mask and the generated mask, respectively. The \oplus in the equation is the exclusive OR operation used to evaluate the disagreeing pixels between G and M .

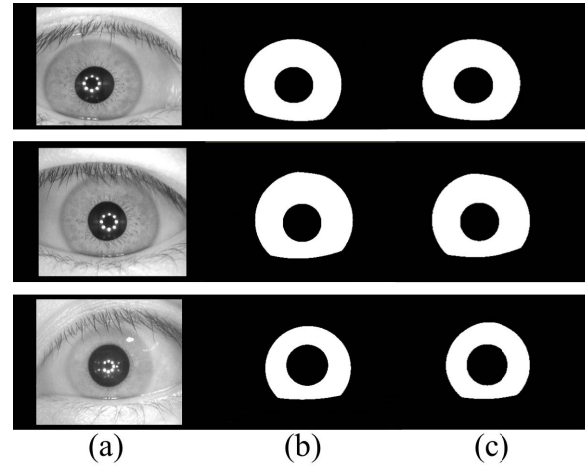


Figure 6. Column (a) shows the original images from CASIA-Iris-Interval database, column (b) are the ground truths and (c) are our segmentation results.

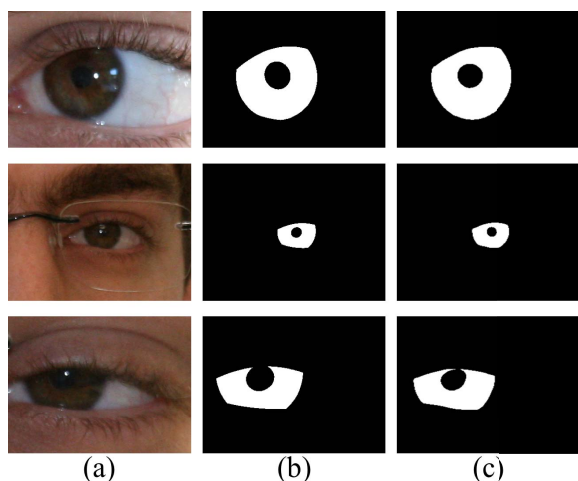


Figure 7. Column (a), (b) and (c) show the original images from UBIRIS.v2, the ground truths and our segmentation results, respectively.

We compare the performance of the proposed iris segmentation approach with some state-of-the-art methods, as shown in Table 1. It is clear from the table that our iris segmentation accuracy on UBIRIS.v2 outperforms all the other recent methods.

Our experimental setting of CASIA-Iris-Interval includes 12,000 images for training and 635 for testing. Note that we use data augmentation to increase the original 2,000 training images to 12,000 images by mirroring and adjusting the brightness of the original images. Furthermore, due to some obvious hand-made ground truth error, we removed 4 images and ground truth masks from the 639 testing images, thus the remaining 635 images are used for testing. Several example images in our experiment are shown in Figure 6. Column (a) are the original images from the database, column (b) are the ground truths and (c) are our segmentation results. It can be seen that the circle border regions are well segmented.

On the other hand, the setting of UBIRIS.v2 experiment contains 2,000 images for training and 250 images for testing without any data augmentation. Although the results cannot be directly compared to other methods because the training and testing settings of other methods and ours are different, it is still effective to show our improvement. Figure 7 shows some segmentation results by using our method on UBIRIS.v2 dataset.

We also use IriShield-USB MK 2120U which is a iris auto-capture camera to collect our own testing data. And we convert all the training images in UBIRIS.v2 to grayscale as training data. Since the images in UBIRIS.v2 are variant in the scale of iris, it is suitable for cross-dataset experiment. Figure 8 depicts some segmentation results on our own testing data.

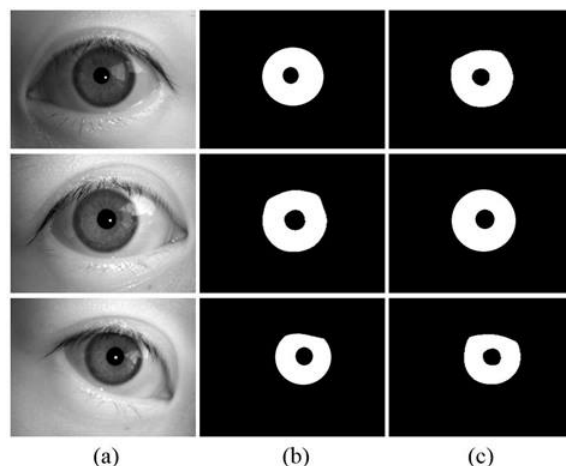


Figure 8. Column (a), (b) and (c) show the original images from IriShield-USB MK 2120U, the ground truths and our segmentation results, respectively.

4.2.2 Iris Verification

During the training, we used stochastic gradient descent (SGD) with momentum set to 0.9 for optimization, the constant learning rate was set to 1×10^{-4} , and the training was performed in mini-batches of size 8. The data augmentation procedure was described in section 4. If the output of an image pair to the proposed iris verification network is larger than 0, we consider them as images of the same eye, otherwise they are decided to be from different eyes. Finally, we obtain 90.85% intra-class accuracy and 99.59% inter-class accuracy for the iris verification experiment on CASIA-Iris-Interval dataset. If we use the normalized images as input, we can reach 94.31% intra-class accuracy and 98.83% inter-class accuracy. With appropriate combination of these two networks, we can reach 95.26% intra-class accuracy and 99.33% inter-class accuracy. The two network models are combined as follows: if the output score of the second model falls between -5 and 5, then we use the first model to make the final decision.

5. Conclusion

In this paper, we propose a deep learning approach to iris recognition. We combine a novel FCN model for iris segmentation along with a modified resnet-18 model for the iris matching. With the fully convolutional network, we are able to generate accurate iris segmentation. Considering all the state-of-the-art methods for iris recognition, the proposed FCNs model performs better iris segmentation, especially for some hard cases. In the future, we would like to refine the iris verification network to improve the robustness of iris matching, especially when the segmentation masks of the two iris images for verification are quite different.

References

- [1] John G Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE transactions on pattern analysis and machine intelligence*, 15(11):1148–1161, 1993. 1, 3
- [2] John Daugman. How iris recognition works. *IEEE Transactions on circuits and systems for video technology*, 14(1):21–30, 2004. 1
- [3] Nianfeng Liu, Haiqing Li, Man Zhang, Jing Liu, Zhenan Sun, and Tieniu Tan. Accurate iris segmentation in non-cooperative environments using fully convolutional networks. In *International Conference on Biometrics*, pages 1–8. IEEE, 2016. 1, 4
- [4] Zijiang Zhao and Kumar Ajay. An accurate iris segmentation framework under relaxed imaging constraints using total variation model. In *IEEE International Conference on Computer Vision*, pages 3828–3836, 2015. 1, 4
- [5] Chun-Wei Tan and Ajay Kumar. Towards online iris and periocular recognition under relaxed imaging constraints. *IEEE Transactions on Image Processing*, 22(10):3751–3765, 2013. 1, 4
- [6] Yung-Hui Li and Marios Savvides. An automatic iris occlusion estimation method based on high-dimensional density estimation. *IEEE transactions on pattern analysis and machine intelligence*, 35(4):784–796, 2013. 1, 4
- [7] Richard P Wildes. Iris recognition: an emerging biometric technology. *Proceedings of the IEEE*, 85(9):1348–1363, 1997. 1
- [8] Haiqing Li, Zhenan Sun, and Tieniu Tan. Robust iris segmentation based on learned boundary detectors. In *International Conference on Biometrics*, pages 317–322. IEEE, 2012. 1
- [9] Hugo Proenca. Iris recognition: On the segmentation of degraded images acquired in the visible wavelength. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8):1502–1516, 2010. 1
- [10] Yung-Hui Li and Marios Savvides. An automatic iris occlusion estimation method based on high-dimensional density estimation. *IEEE transactions on pattern analysis and machine intelligence*, 35(4):784–796, 2013. 1
- [11] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989. 1, 2
- [12] Evan Shelhamer, Jonathan Long, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(4):640–651, 2017. 1, 2
- [13] Richard P Wildes, Jane C Asmuth, Gilbert L Green, Steven C Hsu, Raymond J Kolczynski, James R Matey, and Sterling E McBride. A machine-vision system for iris recognition. *Machine vision and Applications*, 9(1):1–8, 1996. 1
- [14] Li Ma, Tieniu Tan, Yunhong Wang, and Dexin Zhang. Personal identification based on iris texture analysis. *IEEE transactions on pattern analysis and machine intelligence*, 25(12):1519–1533, 2003. 1
- [15] Abhishek Gangwar and Akanksha Joshi. Deepirisnet: Deep iris representation with applications in iris recognition and cross-sensor iris recognition. In *IEEE International Conference on Image Processing*, pages 2301–2305. IEEE, 2016. 1
- [16] Nianfeng Liu, Man Zhang, Haiqing Li, Zhenan Sun, and Tieniu Tan. Deepiris: Learning pairwise filter bank for heterogeneous iris verification. *Pattern Recognition Letters*, 82:154–161, 2016. 1
- [17] Ke Li, Bharath Hariharan, and Jitendra Malik. Iterative instance segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3659–3667, 2016. 2
- [18] Joao Carreira, Pulkit Agrawal, Katerina Fragkiadaki, and Jitendra Malik. Human pose estimation with iterative error feedback. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4733–4742, 2016. 2
- [19] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 3
- [20] Ultrasound nerve segmentation evaluation on kaggle: <https://www.kaggle.com/c/ultrasound-nerve-segmentation>. 2016. 3
- [21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 3
- [22] Sergey Zagoruyko and Nikos Komodakis. Learning to compare image patches via convolutional neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4353–4361, 2015. 3
- [23] Hugo Proenca, Silvio Filipe, Ricardo Santos, Joao Oliveira, and Luis A Alexandre. The ubiris. v2: A database of visible wavelength iris images captured on-the-move and at-a-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8):1529, 2010. 4
- [24] Biometrics ideal test, casia.v4 database: <http://www.idealtest.org/dbdetailforuser.do?id=4>. 4
- [25] Heinz Hofbauer, Fernando Alonso-Fernandez, Peter Wild, Josef Bigun, and Andreas Uhl. A ground truth for iris segmentation. In *International Conference on Pattern Recognition*, pages 527–532. IEEE, 2014. 4
- [26] Fernando Alonso-Fernandez and Josef Bigun. Near-infrared and visible-light periocular recognition with gabor features using frequency-adaptive automatic eye detection. *IET Biometrics*, 4(2):74–89, 2015. 4
- [27] Hugo Proença and Luís A Alexandre. The nice. i: noisy iris challenge evaluation-part i. In *IEEE International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–4. IEEE, 2007. 4