Deep-Learning Based Joint Iris and Sclera Recognition with YOLO Network for Identity Identification

Chia-Wei Chuang and Chih-Peng Fan
Department of Electrical Engineering, National Chung Hsing University, Taiwan
Email: yue841024@gmail.com, cpfan@nchu.edu.tw

***Abstract*—By jointly consideration of the partial iris and sclera region, no sclera and iris separation calculation is needed, and both of the sclera and iris information is used at the same time, and then the identity information is enhanced to avoid being forged. By the deep learning based YOLOv2 model, the visible-light eye images are marked with the jointly partial iris and sclera region, and the identity classifier is trained to inference the correct personal identity. By using the self-made and visible-light eye image database to evaluate the system performance, the proposed deep-learning based joint iris and sclera recognition reaches the mean Average Precision (mAP) up to 99%. Besides, compared with the previous works, the proposed design is more effective without using any iris and sclera segmentation process.**

***Index Terms*—biometric, iris/sclera recognition, deep learning, YOLO model, personal identifications**

I. Introduction

Biometrics [1] have been utilized to the personal recognition effectively, and the iris and sclera recognitions are the useful biometric identity identification technologies, and the biometric technology gives the accurate function of the personal recognition. Fig. 1 depicts the diagram of the periocular and eye regions. By the iris and sclera identification, the biometric information can be connected to the passport data database, and the personal identity will be checked. In recent years, the iris and sclera identifications are widely used for the personal identifications and security check. Many IoT devices also begin to use the iris and sclera identification technologies, and the biometric applications become more and more important.

In the iris recognition design [2], the system changes the features region of iris into a square matrix by the polar coordinate scheme, and the feature code is transformed from the square matrix, and then the signature code is applied for the features matching. In previous works [3]-[7], the spatial-domain based and frequency-domain based methods provide effective recognition performance for iris recognitions. The Scale-Invariant Feature Transform (SIFT) [8] utilizes the local features in images, and keeps the feature invariant for the changes of rotation, scaling, and brightness. Thus, the Scale-Invariant Feature Transform (SIFT) [9], [10] based methods are applied for matching iris features. Based on the SIFT features enhancement technologies, the iris matching process is effective for identity identification.

The scleral vascular pattern is unique between different persons. In addition, the scleral vascular features can also be achieved by the visible light camera. Thus, the blood vessel pattern in the sclera can be used as the personal identification. In recent years, the sclera identification methods in [11]-[15] have been developed for personal identifications. In general, the scleral identification system can be divided into several parts, which include the sclera segmentation, feature extraction, and feature matching processes. By experiments, the sclera segmentation process is important. In [15], the eyelid masking effect will affect the scleral vascular pattern, and the excess masking process will contain unwanted information, such as eyelashes and eyelids. Thus, the unwanted information will reduce the recognition accuracy. Because of the eyeball movement, the blood vessel pattern may be deformed or shifted. By the SIFT-based enhancement process, the effective rapid sclera matching scheme is proposed for identity identification.

Based on the YOLOv2 model [16]-[18], this work proposes a deep learning network model for the joint iris and sclera identification, and the deep learning algorithm process is divided into two phases, namely the training phase and the test phase. Before the training phase, the data augmentation technology is performed first, so that the data is more abundant and beneficial to the subsequent training process. It can also avoid the problem of overfitting. During the training phase, the input image is subjected to the data augmentation technology to increase the number of training samples. By using several repeated convolution and pooling operations to extract the features, the batch normalization is applied to avoid over-training. The weights are trained with a loss function until convergences, and the Softmax functions are used to calculate each category at the final detection layer. In addition, the concept of the R-CNN anchor box is introduced to predict the position of the bounding box of the image object. The proportion of each anchor box is calculated by the k-means method. During the testing

Manuscript received June 6, 2020; revised November 20, 2020.

doi: 10.12720/jait.12.1.60-65

phase, the trained model is used for performance evaluation.

In this paper, to increase the anti-counterfeiting ability without using the additional sclera segmentation, the YOLOv2-based deep learning design is developed for the joint iris and sclera based identity identification. The rest of the paper is described as follows. In Section II, the previous iris and sclera-based identity identification designs are briefly reviewed. In Section III, the design methodology of the proposed YOLOv2-based identify classifier is discussed. The experimental results and comparison are revealed in Section IV. Finally, a conclusion is stated.

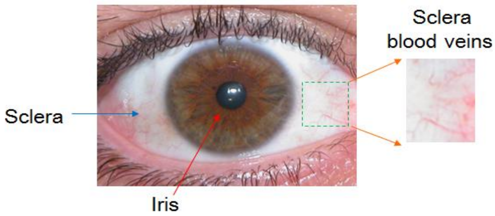


Figure 1. The diagram of the periocular and eye regions

II. Previous Iris and Sclera-Based Identity Identification Designs

A. Features-Based Iris Texture Recognition [10]

Fig. 2 illustrates the processing flow for the iris-based identity identification in [10]. The design has six computational processes, which include the pupil location, iris extraction, interference elimination, features enhancement, features extraction, and features matching. For pupil location, firstly the Sobel filtering process is used to compute the gradient of the first-order derivative of edges. Secondly, the circle-based Hough transform is applied to find the best possible circular contour from the boundary points. Finally, by the Hough-gradient algorithm, the system obtains the pupil's contour for the pupil location. To avoid the important iris area being covered by the upper and lower eyelids and eyelashes interferences, when the circular contour of pupil is estimated, the design locates a larger circle whose radius is 1.7 times of the estimated pupil circle.

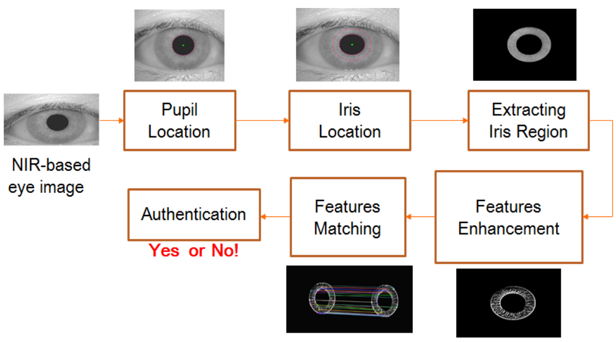


Figure 2. The processing flow for iris-based identity identification [10]

For features enhancement, firstly, the Contrast Limited Adaptive Histogram Equalization (CLAHE) [19] is used for features enhancement. To strengthen the texture information further, when the CLAHE process is finished, and then the cascaded Gabor filtering process is followed. For features extraction and features matching process, by the SIFT-based schemes, the iris features are extracted efficiently, and then the features matching process will be followed. By using the Fast Library for Approximate Nearest Neighbors (FLANN) with joint the RANSAC algorithm, the system gets the effective matching result and improves the matching efficiency simultaneously.

B. Feature-Based Sclera Vein Recognition [15]

Fig. 3 depicts the processing flow for the sclera-based identity identification [15]. The proposed method can be separated into five processes, which are the iris segmentation, the sclera segmentation, the scleral vascular enhancement, the features extraction, and the features matching. For iris segmentation, firstly, the process of reflective light elimination is active. Secondly, for the iris location and segmentation, the system uses the improved Daugman's algorithm. Based on the difference information of integral images, the fast searching method selects the largest circle in eye images, and the iris location is found correctly. For sclera segmentation, the eye image is processed by the color space transformation and the smoothing processing. Next, the gray-level image is binarized, and the contour of the sclera region is found, and then the convex hull by the contour matching is achieved. By building the contour mask, the sclera region is extracted successfully. For scleral vascular enhancement, firstly, the CLAHE scheme is used to the sclera features enhancement. Secondly, to strengthen the features texture further, the proposed system enables the Gabor filtering process continuously. For features extraction, the Dense-SIFT technology is applied to extract feature descriptors. For features matching, the system uses the K-d tree based nearest neighbor search method. The vascular sets with more feature points will be created as a K-d tree structure. Next, by the RANSAC process, the proposed design obtains the largest inner group, and excludes the outer groups that do not fit the spatial correspondence. Finally, the features matching efficiency is raised further.

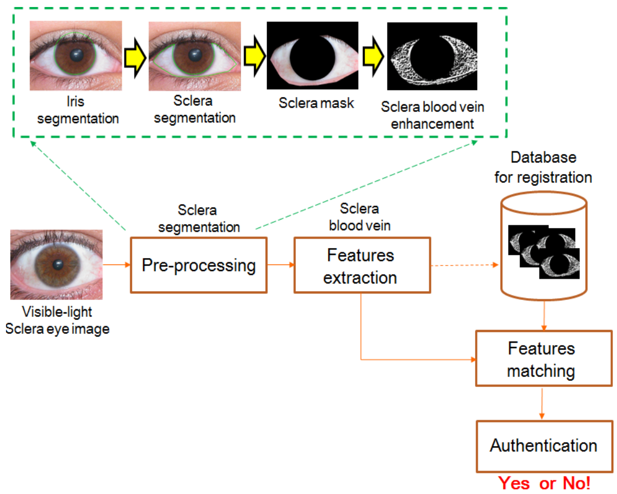


Figure 3. The processing flow for sclera-based identity identification [15]

III. Deep-Learning Based Identity Identification Design by YOLOv2 Model

A. *Brief Review of the YOLOv2 Model*

In [16]-[18], the deep-learning based YOLO model is effective to apply for image object detection. The traditional object detection schemes include three processing sections, which contain the object localization, the features extraction, and the image object classification. Compared to previous object detection works, the YOLO based model directly predict the bounding boxes and class probabilities of image objects by the cost-effective deep-learning neural network, and then the YOLO model can reduce the computing time effectively. During the training phase, the size of the YOLOv2 [17] input image is 416×416 pixels, and the size of the feature map output after five max pooling layers is 13×13. To detect target objects of different sizes, the fine-grained features is added and implemented with the structure of pass-through layers. The high-resolution feature map is added with the original low-resolution output feature map through the pass-through layer. The YOLOv2 network architecture is shown in Fig. 4. In addition, to enhance the robustness of the deep-learning network module, YOLOv2 uses a multi-scale training method. The size of the input image is randomly changed every ten batches. Since the total down-sampling factor of YOLOv2 is 32, therefore the size of the input image must be a multiple of 32, and its size is ranged between {320, 352, ..., and 608}.

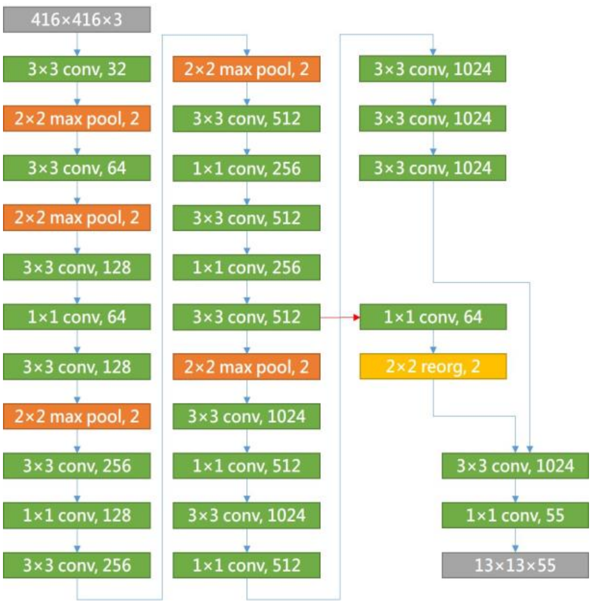


Figure 4. The network architecture of the YOLOv2 model [17]

B. *The Applied Design Flow*

For using the effective YOLOv2 based deep-learning network, the experimental process is described as follows:

At the first stage, the designers must prepare the labelled eye images. The database used in this study was taken through a macro lens. Fig. 5 depicts the visible-light eye image dataset for training and testing the YOLOv2-based model. After the data augmentation process, the amount of image data for one person is 300, and the category number of persons is 8. By using the data augmentation methodologies embedded in YOLO, the effects of the addition of Gaussian noise, the brightness variation, the scaling of images, the perspective transform within ±3~±5 degrees rotation angles about x and y axes, and image sharpening are applied for data augmentation of eye images. Fig. 6 shows the data augmentation results for one eye image in the dataset. Because the number of self-made eye image data is not large, the useful data augmentation process is applied before training to generate 300 augmented eye image data for each identity, and then all of the collected eye images in the database are labelled by LabelImg [20], which is a semi-assisted image labeling tool. Fig. 7 demonstrates the labelled results of visible-light eye images by LabelImg.

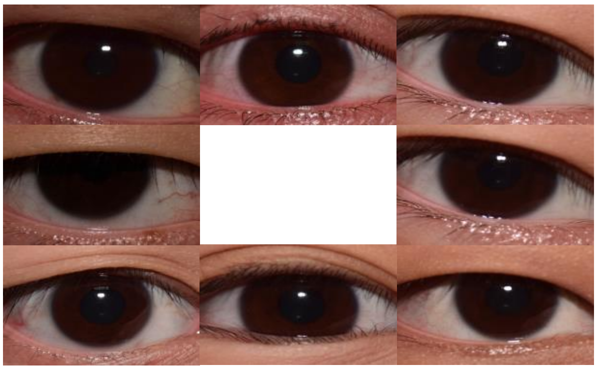


Figure 5. The applied eye image dataset for training and testing the YOLOv2-based model

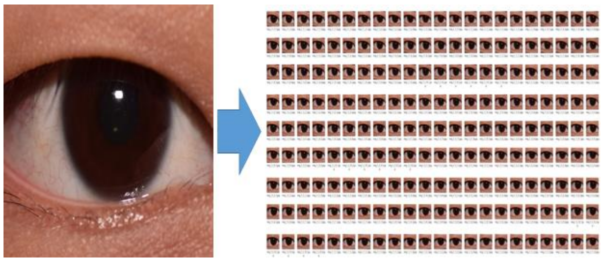
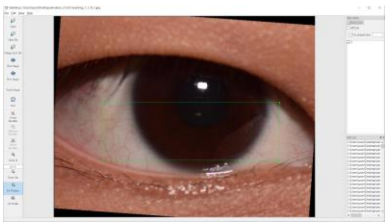
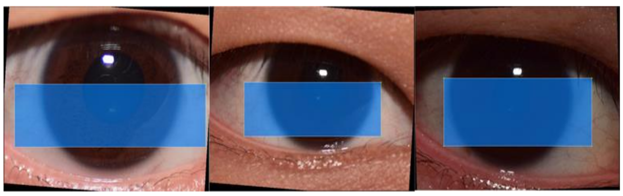


Figure 6. Data augmentation of the eye image dataset



(a)



(b)

Figure 7. The labelled results of eye images by the LabelImg tool (a) The labelling process (b) The labelled results

To make the prediction of the deep-learning network model have better generalization, the image data set is generally divided into three sub-sets, which are namely the training set, the validation set, and the test set. Since the amount of self-made image data in the validation and test sets is not large, this study uses two different allocation procedures to distribute the data sets, where the first allocation is 90%: 5%: 5%, and the second one is 80%: 10%: 10%. Besides, the training set, validation set, and test set are randomly divided.

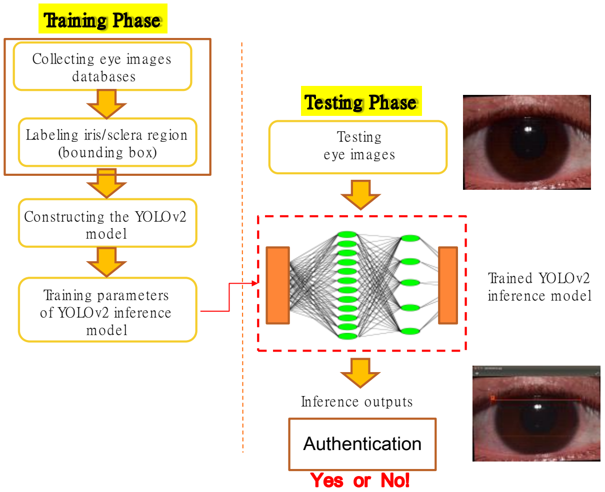


Figure 8. The training/testing flow to obtain the YOLOv2 inference model for identity identification

At the second stage, to obtain a fine inference model, the YOLOv2 network will be trained and tested. Fig. 8 demonstrates the training/testing flows to obtain the YOLOv2 inference model. The used procedures have three steps, which involve the pre-processing for data augmentation and labelling, the training process for obtaining the inference model, and the testing process for the inference model. Based on the darknet framework proposed by YOLOv2, to build a model in the darknet environment, the creation of a configuration file is required. The training procedure uses the configuration file to adapt various layers of deep-learning networks and concatenate to generate a complete model. The configuration file for the network architecture can perform the data augmentation and enhance the training process, and the network parameters are achieved by using the Stochastic Gradient Descent (SGD) training scheme. Finally, the loss function determines and selects the best training model. When the inference model is well trained, then the testing process will be active. In the testing phase, the highest final confidence score is calculated, and the result of detection are predicted.

IV. Experimental Results and Comparisons

The visible-light eye image databases for training and testing are captured by the still camera. In the eye image database, 8 persons are included, and each person own 500 sub-images by using the data augmentation technology. A total of 4000 eye images are contained with a resolution of 800×600 pixels. To train and test the proposed deep-learning based design, a personal computer, which includes a GeForce® GTX 1080 GPU, 24GB memory, and a CPU by the 3.7GHz operational frequency, is applied for the evaluation platform. Due to the blurred effect without focus, the eye images with serious blurred are not selected in our experiments. Fig. 9 shows the dropped image without fine focus. After testing the trained inference model, Fig. 10 demonstrates the inference results by the applied YOLO-based classifier.

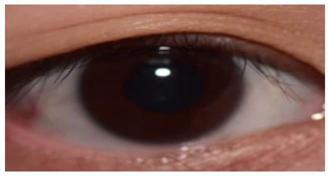


Figure 9. An example of the dropped image

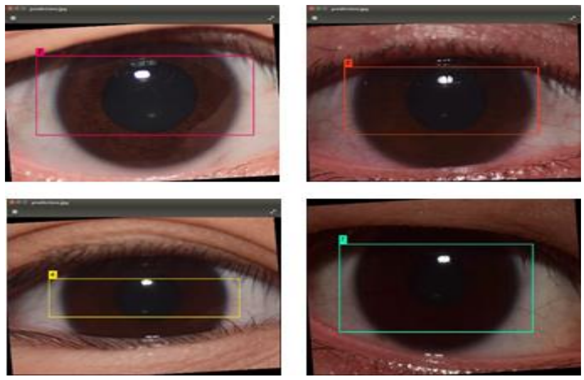


Figure 10. The inference results by the applied YOLO-based model

Train/Valid/Test (90%/5%/5%)	
Category	AP
0	100%
1	100%
2	100%
3	100%
4	100%
5	93%
6	100%
7	100%
mAP	99.17%
Average IoU	89.75%

(a)

Train/Valid/Test(80%/10%/10%)	
Category	AP
0	100%
1	100%
2	100%
3	100%
4	100%
5	96.67%
6	100%
7	100%
mAP	99.58%
Average IoU	88.48%

(b)

Figure 11. The inference results by the applied YOLOv2-based model

In this study, two different allocation procedures of the data sets are used, where the first allocation is 90%: 5%: 5%, and the second one is 80%: 10%: 10%. Fig. 11 shows the inference results by the applied YOLOv2-based model. In Fig. 11, by applying the first allocation for the dataset, the mAP is up to 99.17%. By applying the second allocation for the data set, the mAP can be up to 99.58%. Table I shows the performance comparisons among four different sclera/iris based identity identification designs. By the labelled eye image database for performance evaluations, the proposed deep–learning based joint iris and sclera recognition approaches the mAP up to 99.58%. Besides, compared with the previous designs in [10], [13], [15], the proposed design is efficient and does not need any iris and sclera segmentation computations.

TABLE I. PERFORMANCE COMPARISONS AMONG FOUR DIFFERENT SCLERA/IRIS BASED IDENTITY IDENTIFICATION DESIGNS

Methods	Sclera/Iris segmentation	Sclera/Iris information	Core classification technologies	Accuracy for identity identification
Li, *et al.* [10]	Yes	Iris only	SIFT features and matching based design	Up to 97%
Alkassar, *et al.* [13]	Yes	Sclera only	Sclera features and template registration learning based design	Up to 98%
Gu, *et al.* [15]	Yes	Sclera only	SIFT features and matching based design	Up to 96%
Proposed design	**Don't need**	**Sclera & Iris**	Deep-learning based design (YOLOv2)	**Up to 99% (mAP)**

V. CONCLUSION

For identity identification, the effective deep-learning and joint iris and sclera information based recognition design is proposed in the paper. By using both of the partial iris and sclera region, no additional sclera and iris separation process is required, and the security of identity identification can be increased for the applications. By the YOLOv2 based network, the identity inference model is trained to inference the correct personal identity. By the self-made visible-light eye images dataset for performance evaluations, the proposed deep-learning based joint iris and sclera recognition approaches the mAP up to 99%. Besides, compared with the previous designs in [10], [13], [15], the proposed design is more efficient and does not need any iris and sclera segmentation computations. In future works, the other suitable CNN models and the low-complexity YOLO-based models are studied to evaluate for the enhancement of real-time performance at the edge device.

CONFLICT OF INTEREST

To the best of our knowledge, the named authors have no conflict of interest, financial or otherwise.

AUTHOR CONTRIBUTIONS

Chia-Wei Chuang conducted the deep-learning model design, and the functional simulation and verification; Chih-Peng Fan proposed the experiment method and wrote the paper; all authors had approved the final version.

ACKNOWLEDGEMENT

This work was financially supported by the Ministry of Science and Technology (MOST) under Grant No. MOST 108-2218-E-005-017.

REFERENCES

[1] P. M. Corcoran, "Biometrics and consumer electronics: A brave new world or the road to dystopia? [Soapbox]," *IEEE Consumer Electronics Magazine*, vol. 2, no. 2, pp. 22-33, April 2013.

[2] J. Daugman, "How iris recognition works," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 21-30, 2004.

[3] F. Nadia and K. Hamrouni, "An efficient and reliable algorithm for iris recognition based on Gabor filter," *International Multi-Conference on Systems Signals and Devices*, April 2009.

[4] H. Mehrotra, B. Majhi, and P. K. Sa, "Unconstrained iris recognition using F-SIFT," in *Proc. 8th International Conference on Information, Communications & Signal Processing*, Singapore, Dec. 13-16, 2011.

[5] C. C. Tsai, J. S. Taur, and C. W. Tao, "Iris recognition using Gabor filters and the fractal dimension," *Journal of Information Science and Engineering*, vol. 25, pp. 633-648, 2009.

[6] P. S. Patil, S. R. Kolhe, R. V. Patil, and P. M. Patil, "The comparison of iris recognition using principal component analysis, log Gabor and Gabor wavelets," *International Journal of Computer Applications*, vol. 43, no. 1, pp. 29-33, April 2012.

[7] R. M. Gad, M. A. Mohamed, and N. A. El-Fishawy, "Iris recognition based on log-Gabor and discrete cosine transform coding," *Journal of Computer and Engineering*, vol. 5, no. 2, pp. 19-26, Feb. 2011.

[8] Y. Nakashima and Y. Kuroki, "SIFT feature point selection by using image segmentation," in *Proc. International Symposium on Intelligent Signal Processing and Communication Systems*, 2017, pp. 275-280.

[9] T. D. Yang and C. P. Fan, "Effective scale-invariant feature transform based iris matching technology for identity identification," in *Proc. IEEE International Conference on Consumer Electronics*, Taichung City, Taiwan, May 2018.

[10] K. C. Li and C. P. Fan, "Effective local features matching based rapid iris recognition with interference elimination pre-processing for identity identification," in *Proc. the 11th International Conference on Future Computer and Communication*, Yangon (Rangoon), Myanmar (Burma), February 2019.

[11] Z. Zhou, E. Y. Du, N. L. Thomas, and E. J. Delp, "A new human identification method: Sclera recognition," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 42, no. 3, pp. 571-583, November 2011.

[12] Y. Lin, E. Y. Du, Z. Zhou, and N. L. Thomas, "An efficient parallel approach for sclera vein recognition," *IEEE Trans, Inf. Forensics Security*, vol. 9, no. 2, pp. 147-157, Feb. 2014.

[13] S. Alkassar, W. L. Woo, S. S. Dlay, and J. A. Chambers, "Robust sclera recognition system with novel sclera segmentation and validation techniques," *IEEE Transactions on Systems*, vol. 47, pp. 474-486, 2017.

[14] D. U. Pal, M. A. F. Ballester, and M. Blumenstein, "Sclera recognition using Dense-SIFT," in *Proc. 13th International Conference on Intelligent Systems Design and Applications*, 2013, pp. 74-79.

[15] T. W. Gu and C. P. Fan, "Sclera recognition by density sampling features based vascular structure rapid matching for identity identification," in *Proc. IEEE 7th Global Conference on Consumer Electronics*, Nara, Japan, October 2018.

[16] J. Redmon, S. Divvala, and R. Girshick, "You only look once - Unified real-time object detection," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779-788.

[17] J. Redmon and A. Farhadi, "YOLO9000 better faster stronger," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7263-7271.

[18] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," arXiv:1804.02767, 2018.

[19] G. Yadav, S. Maheshwari, and A. Agarwal, “Contrast Limited Adaptive Histogram Equalization (CLAHE) based enhancement for real time video system,” in *Proc. International Conference on Advances in Computing, Communications and Informatics*, 2014.
[20] Tzutalin/LabelImg. [Online]. Available: https://github.com/tzutalin/labelImg

Chia-Wei Chuang was born in Nantou, Taiwan, in 1995. He received his B.S. degree in Department of Electronic Engineering from Feng Chia University, Taiwan, in 2018, and the M.S. degree in Electrical Engineering from National Chung Hsing University, Taichung, Taiwan, in 2020. His research interests include AI image recognition and VLSI designs.

Chih-Peng Fan received the B.S., M.S., and, and Ph.D. degrees, all in electrical engineering, from the National Cheng Kung University, Tainan, Taiwan, in 1991, 1993 and 1998, respectively. During October 1998 to January 2003, he was a design engineer for the cable modem and multimedia transmission project at the CPE and Interface Technology Department (N100), Computer and Communications Research Laboratories (CCL), Industrial Technology Research Institute (ITRI), Hsinchu, Taiwan. In 2003, he joined the Department of Electrical Engineering at National Chung Hsing University in Taiwan as an Assistant Professor. He became an Associate Professor in 2007, and a full Professor in 2013. He has more than 110 publications, including technical journals, technical reports, book chapters, and conference papers. His teaching and research interests include digital image processing and pattern recognition, digital video coding and processing, digital baseband transceiver design, VLSI signal processing, and fast prototype of DSP systems with FPGA and embedded SOC platform. He served and is serving as: TPC Chair of ICCE-TW 2017; General Chair of ICCE-TW 2018; General Co-Chair of ICCE-TW 2019; International Coordinator of GCCE 2019; TPC Co-Chair of ICCCI 2020; Publicity Co-Chair of ICIET 2020; He is a member of TICD, IEICE, and IEEE.