RESEARCH

KartalOl: a new deep neural network framework based on transfer learning for iris segmentation and localization task—new dataset for iris segmentation

Jalil Nourmohammadi Khiarak¹ · Amin Golzari Oskouei² · Samaneh Salehi Nasab³ · Farhang Jaryani⁴ · Seyed Naeim Moafinejad⁵ · Rana Pourmohamad⁶ · Yasin Amini⁷ · Morteza Noshad⁸

Abstract

Iris segmentation and localization in unconstrained environments are challenging due to long distances, illumination variations, limited user cooperation, and moving subjects. Some existing methods in the literature have somehow mitigated some of the above-mentioned issues. In this paper, motivated by these weaknesses, we propose a framework that employs a deep neural network-based approach to iris segmentation and localization. The proposed framework is based on a U-Net architecture initialized with a pre-trained MobileNetV2 model. In addition, to better study the detectors in iris recognition scenarios, we have collected 1000 images. The provided dataset (KartalOl) is made publicly available for the research community. In the proposed framework, to have better generalization, we fine-tuned the MobileNetV2 model on the provided data for NIR-ISL 2021 from the CASIA-Iris-Asia, CASIA-Iris-M1, and CASIA-Iris-Africa and our dataset. Likewise, data augmentation techniques are applied on images. We chose the binarization threshold for the binary masks by iterating over the images in the provided dataset. The proposed framework is trained and tested in CASIA-Iris-Asia, CASIA-Iris-M1, and CASIA-Iris-Africa, along with the KartalOl dataset. The experimental results highlight that our method surpasses state-of-the-art methods on mobile-based benchmarks. The implementation source code of KartalOl is made publicly available at https://github.com/Jalilnkh/KartalOl-NIR-ISL2021031301.

Keywords Iris segmentation · Localization · MobileNetV2 model · Unconstrained environment

✉ Jalil Nourmohammadi Khiarak
   Jalil.Nourmohammadi@elka.pw.edu.pl

✉ Samaneh Salehi Nasab
   salehinasab.s@lu.ac.ir

¹ Institute of Control and Computation Engineering, Warsaw University of Technology, Ul. Nowowiejska 15/19, 00-665 Warsaw, Poland

² Department of Computer Engineering, Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran

³ School of Engineering, Lorestan University, Aleshtar Campus, Aleshtar, Iran

⁴ Faculty of Engineering, Arak University, Arak, Iran

⁵ Department of Physics, Shahid Beheshti University, Tehran, Iran

⁶ School of Computing, Informatics and Decision Systems Engineering, Arizona State University, Arizona, USA

⁷ Faculty of Electrical and Computer Engineering, University of Kharazmi, Tehran, Iran

⁸ Stanford University, Stanford, USA

1 Introduction

Biometric technologies are automated methods of identifying or authenticating an individual's identity based on physiological or behavioral characteristics. Many biometric technologies have been developed that are used to authenticate a person's identity. Biometrics has distinct advantages in identifying people. Tokens and photo IDs can be lost, stolen, duplicated, or left at home, and passwords can be forgotten, shared, or observed. In contrast, biometrics promises fast, easy-to-use, accurate, reliable, and cost-effective authentication for a variety of applications. Biometric recognition techniques, such as fingerprint [1], face recognition, iris recognition, and periocular recognition, have been altered by

technological innovation. Among these biometric identification technologies, iris-based methods are highlighted as the most reliable and accurate technology [2, 3], which attracts the biometrics community [4].

The iris is unique to the human being. Even identical twins have diverse iris patterns; therefore, no two persons may have the same iris pattern. The iris pattern of the right eye differs from that of the left. As a rule, the left eye is used. The iris is irradiated with subtle infrared rays using camera technology. The image of the eye is captured, and the iris scanner analyzes the iris features, which have more than 200 points of comparison, such as rings, furrows, and freckles. Patterns are drawn and then compared with those of the user. The iris scanner is often used at airports as a passport substitute, logging on to computers and ATMs, entering buildings, and accessing databases.

Any iris recognition system usually consists of the following sub-processes: iris image acquisition, pre-processing, iris segmentation, iris feature extraction, and iris matching verification or identification [5]. Iris segmentation and localization involve determining the pixels of the iris region in an image that can be used for feature extraction and matching [6]. This is a crucial step to improve the accuracy of iris recognition [7]. By determining the correct iris region, we can obtain valuable information from the iris images and increase the efficiency of the iris recognition system [5]. Most acquired iris images contain noise, e.g., occlusions caused by eyelids or eyelashes, specular reflections, angular deviations, and blur. A noisy iris image is the most significant issue in iris segmentation and localization [8].

Moreover, most of the previous studies were performed in controlled environments. They are not very robust, and their goal is to segment the noise-free mask while ignoring the parameterization of the iris boundaries. This work addresses this problem by developing a new deep-learning-based framework [9]. The following are the main contributions of this work:

1. A new deep neural network based on transfer learning is proposed for iris segmentation and localization.
2. A new dataset for iris segmentation and localization, entitled KartalOl, is presented.
3. This research proposes an efficient method for performing NIR iris imaging in environments with unobstructed specular reflections, oblique angles, blur, and noise.

Although significant progress has been made in iris segmentation and localization, there are still some areas where knowledge is lacking. We focus on the following issues:

Robustness: Although many methods have been proposed to account for variations in iris shape, lighting conditions, and other environmental factors, there is still room for improvement in terms of robustness. Current methods do not work well in specific scenarios, such as low light conditions or when the iris is partially occluded.

Availability of datasets: for developing and evaluating iris segmentation and localization methods, the availability of large and diverse datasets is crucial. While several datasets are available, there is still a need for more comprehensive and diverse datasets that can capture a wide range of iris shapes, sizes, and variations.

Comparison of methods: There is a need for standardized evaluation metrics and benchmark datasets to enable fair and accurate comparisons of different iris segmentation and localization methods. This can help to identify the strengths and weaknesses of different methods and guide future research in this area.

Overall, while there has been significant progress in iris segmentation and localization, there are still several areas where current knowledge is lacking. Addressing these gaps can help to improve the accuracy, robustness, and efficiency of iris recognition systems and enable a broader range of applications in biometrics and computer vision.

Using a pre-trained MobileNetV2 deep neural network, we present a U-Net method for iris segmentation and localization. Our research differs from previous work in that we propose an efficient encoder–decoder framework to improve iris segmentation and localization performance. In addition to our collected dataset, we used three known publicly available NIR iris datasets. As shown in Table 2, the proposed method outperforms the other three state-of-the-art methods on the iris segmentation and localization tasks with a rank sum of 82.

The structure of the paper is as follows: Sect. 2 reviews the previous research; Sect. 3 presents the KartalOl dataset; Sect. 4 describes the proposed framework; Sect. 5 discusses the experimental results of the proposed method; and Sect. 6 provides the conclusions.

2 Related work

Li et al. [7] proposed a deep-learning-based iris segmentation technique. The authors created a modified R-CNN with six layers to mark and identify eye location. Then, the Gaussian Mixture Model (GMM) was applied to find the pupil. Also, they used five crucial boundary points to identify the circular iris inner boundary. The boundary-point election procedure was used to determine the iris's outer boundary point. Their proposed method performed better on a dataset without thick or bunchy eyelashes. However, detecting the iris's inner boundary may be the scheme's biggest flaw. It may perform poorly in dimly lit areas (e.g., bunches of dark eyelashes, congested eyebrows, and hair strips).

Han et al. [10] focused on iris localization, proposing a practical approach for noisy iris images. The two steps in

their proposed iris localization method are pupil border and iris boundary localization. An efficient block-based minimum energy detection method is used to localize a pupil region, with specular reflection removal as a pre-processing step. The NICE.II dataset was used in these experiments. The dataset includes a variety of noisy photos captured in the real world. This method extracts features using the traditional methods while applying deep learning methods could increase the performance of the iris localization.

Gad et al. developed an iris-based recognition technique as a unimodal biometric using multi-biometric situations. During the segmentation phase, a novel method based on the masking approach was designed to find the iris. Two novel techniques (delta-mean and multi-algorithm-mean) were intended to extract iris feature vectors. CASIA v.1, CASIA v.4-Interval, UBIRIS v. 1, and SDUMLA-HMT were used to test the proposed approach. The results indicate that the proposed solution for authentication performs satisfactorily [11].

Another study proposes extracting region-based information from non-cooperative iris images using an adjustable filter bank [12]. The proposed approach is based on a 14th-order half-band polynomial. The model was trained in the CASIAv3, UBIRISv1, and IITD datasets. Furthermore, filter coefficients were collected from the polynomial rather than the z-domain. Texture features were extracted from the annular iris templates that were suitably localized via an integral differential operator using the tunable filter bank. The original iris template is initially separated into six equispaced sub-templates. A good regularity condition is not maintained because of the linear-phase property of other bi-orthogonal filters.

Using the Integer Wavelet Transform (IWT), Singh et al. [13] offer an iris identification technique based on feature extraction. The relative total variation model is used along with other models. Four-level IWT is used to normalize and deconstruct the segmented iris area. To speed up iris segmentation, simple filtering, Hough transforms, and edge detection are used to approximate the location. The input image is transformed into 256 sub-bands using a four-level IWT, of which only 192 lower sub-bands are considered. High-frequency sub-bands are ignored, since they introduce noise to the system and reduce precision. Energy is recovered from the 192 sub-bands, resulting in a 192-bit binary code. The energy of each sub-band is compared to a pre-computed individual-tailored threshold value to generate a unique iris code. The performance of the proposed algorithm needs to be improved by integrating the iris denoising method.

Current methods for iris segmentation and localization suffer from several limitations and challenges that impact the accuracy and robustness of the overall recognition system. To address these shortcomings, our research aims to create a more robust and accurate method for iris segmentation and localization. Specifically, we will explore U-Net architecture that has been initialized with a pre-trained MobileNetV2 model to automatically learn relevant features and adapt to a wide range of eye image conditions. We try to improve the accuracy and speed of iris segmentation and localization by leveraging the power of deep learning, ultimately contributing to the advancement of reliable biometric identification systems.

3 Collected dataset

Publicly available NIR-ISL 2021 datasets for human iris photographs serve and are essential in iris recognition research. The accessible datasets share characteristics, such as near-infrared imaging, and follow John Daugman's [14] requirements. We now have additional iris picture datasets [2, 15] thanks to advances in mobile computing and deep learning (DL) in biometric applications. Finding a suitable dataset to complete one's research activities can be difficult for new researchers in this field. Most dataset providers only allow noncommercial research to be performed with their datasets.

Furthermore, authors strictly adhere to the access terms and require a signature from the researcher or a legal representative of the research institution. This imposes additional limitations on the attractiveness of particular datasets among academics. This study presents a new dataset for visible spectrum iris research that includes iris photos taken with two different smartphones. Two of the latest and most popular phones have been added to the KartalOl dataset: the iPhone X and iPhone 11. The photos were taken with a smartphone's rear camera.

We have collected a dataset for visible spectrum iris photos and, more importantly, for using a smartphone camera, namely KartalOl dataset.

The images in this dataset were taken by volunteers from Central and South Asian countries. The proposed dataset helps evaluate the effectiveness of iris recognition algorithms among ethnic groups. The KartalOl dataset can be used to assess the robustness of several steps in an iris identification pipeline, such as segmentation and feature extraction, and thus reveals iris color and ethnicity dependencies. Another crucial consideration is the type of smartphone used to collect data for this dataset. This dataset was created using the most recent iPhones: the iPhone X and iPhone 11, to compare their iris recognition capabilities. The KartalOl data set is collected in different illuminations. This dataset was gathered from people worldwide and is divided into two distinct categories (the morning group and the evening group).

As shown in Fig. 1, 80% of the participants of the datasets are male, and only 20% of participants include women. Although most data were recorded from mid age people, there were participants between 20 years old to 60 years old.
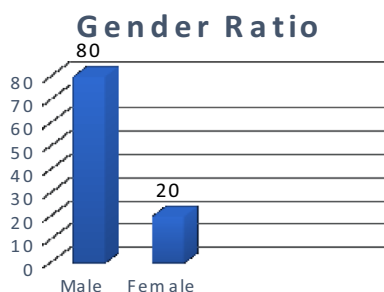
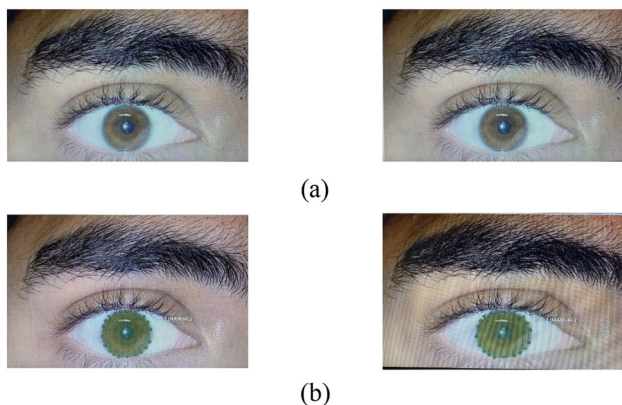**Fig. 1** Color and gender distribution of samples in the dataset

Fig. 2 Mask annotation for iris recognition and segmentation using CVAT tools. **a** Input images and **b** selected area as a mask

Also, different varieties of iris colors exist. In other words, there were brown eyes, blue eyes, green eyes, etc. This dataset supports color veracity, age differences, gender differences, and quality aspects.

3.1 Mask for our dataset

Intel develops a new program, Computer Vision Annotation Tool (CVAT), to speed up the annotation process. CVAT is applied to carry out the iris mask segmentation. We used the CVAT tool to annotate many objects with different properties. Figure 2 represents the annotation process in detail. Various methods exist to conduct data annotation, but applying a proper specific tool will increase the processing speed. As shown in Fig. 2, eyes were segmented using multi-point to draw a circle. This segmentation would allow us to draw an accurate circle around an eye, as shown in Fig. 2b.

4 Proposed method

In this section, the proposed framework is explained. In Sects. 4.1 and 4.2, the proposed segmentation and localization approaches are introduced, respectively.

4.1 *Iris segmentation*

We aim to study U-Net segmentation [16] and utilize it in iris image segmentation while using pre-trained MobileNetV2 [17] as the encoder for the U-Net architecture. The goal of combining the pre-trained MobileNetV2 with the U-Net is to improve iris localization and segmentation accuracy at the same time. We use the pre-trained MobileNetV2 model trained on the ImageNet dataset in the proposed framework.

Figure 3 presents the training process of our proposed iris segmentation model using a U-net architecture. The training phase consists of two main steps: data preparation and U-net training. In the data preparation step, we use iris and corresponding mask images to generate input–output pairs for the U-net model. The input images represent the iris images, while the related mask images represent the binary segmentation map of the iris region. We use the generated input–output pairs to train the U-net model using a stochastic gradient descent algorithm during the U-net training step. The U-net architecture consists of an encoder and a decoder network that perform feature extraction and reconstruction, respectively. The encoder network comprises convolutional and pooling layers that downsample the input images. In contrast, the decoder network is composed of convolutional and up-sampling layers that upsample the feature maps.

We compute the difference between the predicted segmentation map and the ground truth segmentation map using a binary cross-entropy loss function. The U-net model is trained to minimize this loss function by updating the network weights using backpropagation.

Once the U-net model is trained, we evaluate its performance on a separate set of iris images using the testing phase. The testing phase aims to measure the accuracy and generalizability of the trained model on previously unseen iris images.

A down-sampling part and an up-sampling part are two parts of the U-net model. High-level features are fused with low-level features via a shortcut between the two parts, improving the capability of segmenting image details. The feature extraction layers of pre-training MobileNetV2 substitute the down-sampling part. The presence of four inverted residual blocks and five deconvolution layers in the up-sampling part confirms the identical form of the input and output dimensions of Mobile-U-Net. Figure 4 depicts the architecture of Mobile-U-Net.

Table 1 provides the definition and operation of each layer of the MobileNetV2-U-Net architecture. The table contains seven columns: Layers, Input, Output, Output size, k, S, and T.

The first column, Layers, provides the name of each layer in the architecture. The second column, Input, provides the input for each layer. The third column, Output, provides the

Fig. 3 The pipeline of iris segmentation with two phases: **a** training phase and **b** testing phase
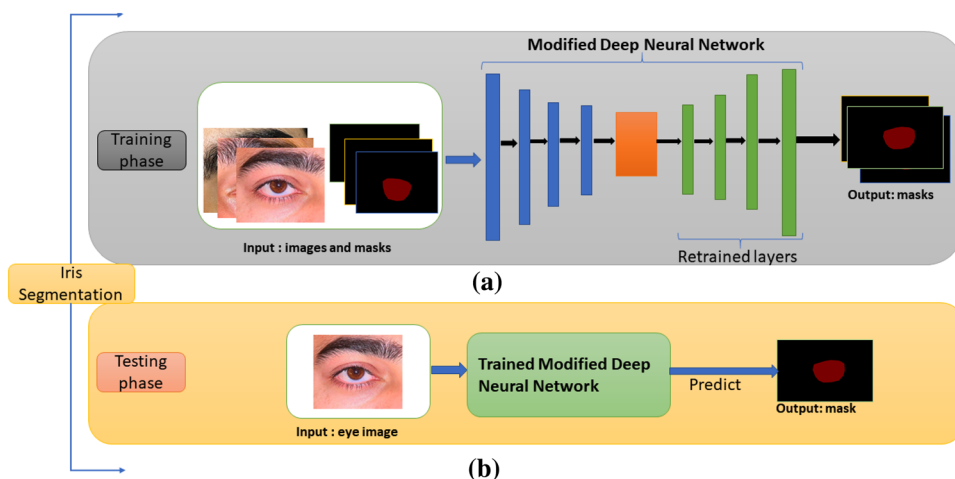

Table 1 Definition and operation of each layer of MobileNetV2-U-Net

Layers	Input	Output	Output size	k	S	T
Input image	–	–	(3, 224, 224)	–	–	–
D1	Input image	X1	(16, 128, 128)	–	–	–
D2	X1	X2	(24, 64, 64)	–	–	–
…	…	…	…	…	…	…
Inverted Residual4	L7 + X1	L8	(16, 128, 128)	–	6	1
ConvTranspose5	L8	Output	(1, 224, 224)	4	–	2
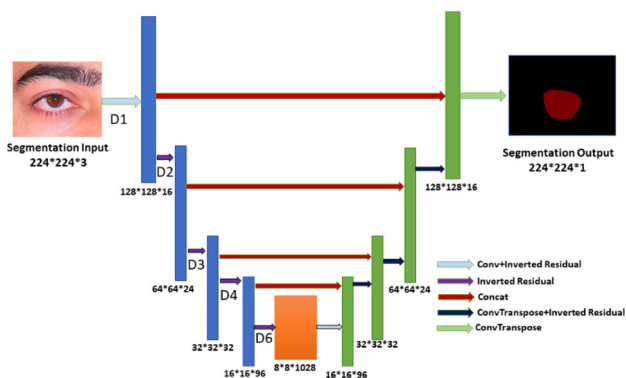
Fig. 4 The architecture of Mobile-U-Net

output of each layer. The fourth column, Output size, provides the size of the output of each layer. The fifth column, k, provides the kernel size of each layer. The sixth column, S, provides the stride size of each layer. The seventh column, T, provides the dilation rate of each layer.

Table 1 starts with the input image row, which provides the size of the input image, which is (3, 224, 224). The next row, D1, provides the input and output of the first layer, which is the input image and X1, respectively. The output size of this layer is (16, 128, 128). The next row, D2, provides the input and output of the second layer, which is X1 and X2, respectively. The output size of this layer is (24, 64, 64). The table

continues in this manner until the Inverted Residual4 row, which provides the fourth layer's input and output, which are L7 + X1 and L8, respectively. The output size of this layer is (16, 128, 128). The last row, ConvTranspose5, provides the input and output of the fifth layer, which is L8 and Output, respectively. The output size of this layer is (1, 224, 224).

The values for the parameters of an algorithm are selected based on their ability to improve accuracy and performance. The abstract's Table 2 shows that each layer has a set of parameters such as K, S, and T. These parameters govern how much information each layer processes during training or inference. In this case, these values were chosen, because they are effective at improving iris segmentation and localization when used with U-Net architectures.

4.2 Iris localization

We proposed a deep neural network-based localization method using pre-trained MobileNetV2 alone with U-Net to obtain inner and outer boundaries in iris images. The pipeline of the proposed method is shown in Fig. 5.

Figure 5 illustrates the iris localization process, which involves two main phases: training and testing. In the training phase, we use a pre-trained iris segmentation method to locate the iris area within an eye image. Once the iris area is

Table 2 Performance comparison between the proposed approach and existing methods: EyeCool, Lao Yang Sprint Team, and U-net (baseline) on iris localization and segmentation tasks

Methods	Iris segmentation			Iris localization			Rank sum
	$E1$ (rank)	$E2$ (rank)	Rank sum	$mDice$ (rank)	$mHdie$ (rank)	Rank sum	
EyeCool [8]	34	34	68	30	28	58	112
Lao Yang Sprint Team [8]	7	7	14	20	22	42	56
U-net (base-line)	92	92	184	106	115	221	405
Proposed method	19	19	38	21	23	44	82

The best values are bolded based on their rank sum on all evaluation measures

Fig. 5 The pipeline of iris localization with two phases: **a** training phase and **b** testing phase
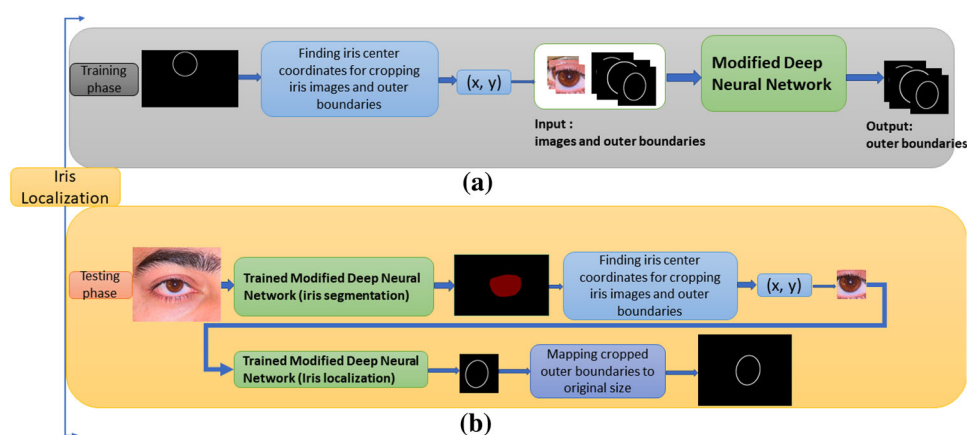
identified, we find the center of the iris and crop it to create a specific area of the eye image. This cropped image is then used as input to train a deep neural network (DNN) model. By providing the DNN model with a specific area of the eye image that includes most of the iris area, we can focus the model's attention on the desired object in the image and remove any extraneous information. This approach has been shown to improve the accuracy and reliability of the DNN model for iris localization.

In the testing phase, we start with the original eye image and use the pre-trained iris segmentation method to determine the iris area. Once the iris area is located, we map it to the original image to identify the precise location of the iris. This allows us to evaluate the performance of our iris localization method and compare it to other existing methods.

4.2.1 Center of iris outer boundary

The following equation shows how to calculate and find the center of the iris's outer boundary per image. We have the coordinates of multiple points on the circumference of the circle. Therefore, we use the following formula to find the center of the circle:

$$C_x = (x_1 + x_2 + x_3)/3 \qquad (1)$$

$$C_y = (y_1 + y_2 + y_3)/3, \qquad (2)$$

where (x_1, y_1), (x_2, y_2), and (x_3, y_3) are the coordinates of any three points on the circumference of the circle. This formula works, because the center of the circle is the intersection point of the perpendicular bisectors of any two chords of the circle. Therefore, by finding the intersection point of the perpendicular bisectors of three chords formed by three points on the circumference of the circle, we can find the circle's center. Note that this formula assumes that the three points are not collinear (i.e., they do not lie in a straight line). If the three points are collinear, the circle cannot be uniquely determined, and this formula will not work. In that case, we would need to use more points to find the circle's center.

To solve the collinear problem, we use more points to find the center of the circle. We use the least-squares method to fit a circle to a set of points. Given a set of points(x_1, y_1), $(x_2,$

y_2), ..., (x_n, y_n), the least-squares method finds the circle that best fits the data in the sense that it minimizes the sum of the squared distances between the points and the circle. The equation of a circle can be written as

$$(x - \text{Center}_x)^2 + (y - \text{Center}_y)^2 = R^2, \tag{3}$$

where (Center_x, Center_y) are the coordinates of the center of the circle, and R is the radius of the circle. To fit a circle to a set of points, you can use the following algorithm:

Calculate the mean values of x and y

$$x_{\text{mean}} = (x_1 + x_2 + ... + x_n)/n \tag{4}$$

$$y_{\text{mean}} = (y_1 + y_2 + ... + y_n)/n. \tag{5}$$

Define the matrices A and b as follows:

$$\begin{aligned} A = [&2(x_1 - x_{\text{mean}})2(y_1 - y_{\text{mean}})1; \\ &2(x_2 - x_{\text{mean}})2(y_2 - y_{\text{mean}})1; \\ &\cdots \\ &2(x_n - x_{\text{mean}})2(y_n - y_{\text{mean}})1] \end{aligned} \tag{6}$$

$$b = \left[x_1^2 + y_1^2; x_2^2 + y_2^2; \ldots x_n^2 + y_n^2 \right] \tag{7}$$

Solve the linear system $A * x = b$ for the vector x, where $x = [\text{Center}_x; \text{Center}_y; R^2]$. The center of the circle is given by (Center_x, Center_y), and the radius of the circle is given by $R = \text{sqrt}(x(3) + \text{Center}_x^2 + \text{Center}_y^2)$. $x(3)$ represents the squared radius of the circle, and we need to take the square root of $x(3)$ to get the actual radius of the circle.

This algorithm can be extended to fit a circle to any number of points (not just three), giving you the best-fitting circle in the least-squares sense. Note that this method assumes that the points lie on a perfect circle; if there is any noise or error in the data, the result may not be accurate.

5 Experiments

5.1 Datasets

To train and evaluate the model's performance and compare it with other methods, we use four datasets, including our collected dataset called KartalOl, CASIA-Iris-ASIA [8], CASIA-Iris-Mobile [8], and CASIA-Iris-Africa [8]. In this section, each of them is introduced.

5.1.1 KartalOl

This study utilizes a dataset from KartalOl that contains 300 photographs taken by 65 people throughout two separate sessions. These pictures include a variety of noise components,

which are intended to simulate less restrictive acquisition settings when capturing photographs.

5.1.2 CASIA-Iris-ASIA

Many of the images in the CASIA-Iris-ASIA collection are taken from the participants' left and right eyes. Through uniformity and cross-validation, we determined that each subject in CASIA-Iris-Africa is applied if it has an equal number of iris photographs or more than five images on either the left or right eye branch on each side of the iris.

5.1.3 CASIA-Iris-Africa

CASIA-Iris-Africa includes a total of five pictures for each subject. More importantly, when the left and right eye branches meet the criterion, we select the left eye branch to represent the subject. After doing a requirement analysis, it was discovered that only 50 individuals from CASIA-Iris-Africa fulfilled the study's qualifying criteria. Each study participant is referred to as a class throughout the experiment.

5.2 Experimental setup

In the experiments, five iris images are randomly selected from each of the mentioned subjects. One image is used for the testing sample, while the other four were used for training samples. In most real-world applications, obtaining enough images per subject for training and testing is difficult. Therefore, this setting was made to simulate the practical application. The following section shows the all-recognition rate, the average result of five cross-verification experiments. There are 300 iris images from 100 subjects in the CASIA-Iris-ASIA dataset. Eyeglasses and specular reflections are the primary sources of intra-class variations. These variations are applied to examine images in a harsh environment. It is also used to calculate similarities. A cosine similarity metric is utilized for similarity calculation. As highlighted by the previous research, Cosine similarity metrics is a powerful method for image recognition problems [18]. It is used to classify test samples into the same class as the nearest sample from the training set. In general, it is applied to measure the distance when the magnitude of the vectors is not essential. Usually, this occurs during working with image data characterized by some features. We can presume that when a feature (e.g., length of edges) happens more frequently in image one than in image two, image one is more related to the length of edges. Then, the length of edges probably happened frequently in image one as it was way longer than in image two. Cosine similarity corrects this.

The appropriate example of using this metric would be image data, although cosine similarity might be applied for

other cases where instances properties cause larger weights without any distinct difference.

5.3 Evaluation metrics

We considered the various components of the datasets and application situations based on four distinct types of training data, CASIA-Iris-Africa, CASIA-Iris-Africa, and KartalOl datasets, and trained the algorithm in this paper. Following the algorithm's training step, a model tailored to the specific target was obtained. Finally, during the testing step, we conducted a regular within-dataset assessment of the models that had been trained. The testing results include binary iris segmentation masks that matched the test images, with non-zero-valued pixels. This indicated the foreground (excellent iris texture or border) area and zero-valued pixels representing everything else in the images.

Iris segmentation: When it came to segmenting and localizing the iris, each method was evaluated for effectiveness. Based on the binary segmentation predictions and associated ground truths for all techniques, the following assessment metrics were generated:

To evaluate the performance of iris segmentation, we used the two criteria provided by the NICE-I competition. The most significant statistic, $E1$, is the average proportion of the matching disagreeing pixels of overall pictures, which may be determined using a pixel-wise XOR operator across the anticipated and ground truth iris masks

$$E1 = \frac{1}{n * h * w} \sum_i \sum_j M(i, j) \otimes G(i, j) \qquad (8)$$

where i and j are the pixel coordinates in the predicted iris mask M and the ground truth iris mask G, respectively, h and w are the width and height of the testing image, and n refers to the number of testing photos.

The equation is written as E1 and comprises two parts: $M(i, j)$ and $G(i, j)$. $M(i, j)$ is a binary mask that is used to compare the predicted and ground truth segmentation results. The binary mask is a matrix of 0 s and 1 s, where 0 s represent the background and 1 s represent the foreground. $G(i, j)$ is the ground truth segmentation results. It is also a matrix of 0 s and 1 s, where 0 s represent the background and 1 s represent the foreground.

The equation is used to calculate the accuracy of the proposed framework by taking the sum of the product of $M(i, j)$ and $G(i, j)$ and dividing it by the total number of pixels ($n * h * w$). The accuracy is calculated by comparing the predicted segmentation results with the ground truth segmentation results.

This equation is used to calculate the proposed framework's accuracy by comparing predicted segmentation results to ground truth segmentation results.

The second metric, $E2$, compensates for the difference between the prior probability of "iris" and "non-iris" pixels in the images' prior probabilities. In other words, the average of the false positives (FP) and false negatives (FN) rates is the rate of false positives

$$E2 = \frac{1}{n * 2} \sum_i (f_p + f_n). \qquad (9)$$

In this equation, (E2) is the evaluation metric, (n) is the number of testing images, and (F_P) and (F_n) are two terms that are used to calculate the evaluation metric.

(F_P) is the balance of background pixels that are wrongly retrieved as iris pixels. This means that it is the proportion of background pixels that are incorrectly classified as iris pixels. (F_n) is the ratio of iris pixels inaccurately recovered as background pixels. This means that it is the proportion of iris pixels that are incorrectly classified as background pixels. The evaluation metric (E2) is calculated by taking the average of (F_p) and (F_n) and dividing it by (n) multiplied by 2. This equation is used to calculate the evaluation metric of the proposed framework.

$E1$ and $E2$ are defined in [0; 1], where the lower and higher values serve sequentially better and worse iris segmentation results.

Iris localization: In this study, we used two standard metrics for assessing iris localization performance: the Dice index and the Hausdorff distance. The Dice index and the Hausdorff distance are used to evaluate iris localization performance.

The dice index is the calculated area of the overlap in two images. It means that each recognized utter boundary or inner boundary image is compared with the grand truth.

Hausdorff distance is the longest of all the distances from any points in one set (utter boundary or inner boundary in iris images) to the closest points in the other utter boundary or inner boundary of iris images.

5.4 Experiment 1: quantitative evaluation

The proposed method is evaluated using a variety of metrics across a range of testing sets and experimental settings. Multiple computed metric scores and associated rankings are combined to compare the proposed overall performance. We first show the prediction results for iris segmentation and localization, and then synthesize the findings to analyze the suggested method's performance. Table 2 summarizes the overall assessment findings obtained by the proposed approach across all testing sets.

Table 2 compares the performance of the proposed approach with existing methods on iris segmentation and localization tasks. The first column is the name of the methods, which are EyeCool, Lao Yang Sprint Team, U-net

(baseline), and the proposed method. The following three columns are for Iris Segmentation, which are E1 (Rank), E2 (Rank), and Rank Sum. The following three columns are for Iris Localization, which are mDice (Rank), mHdie (Rank), and Rank Sum. The last column is the Rank Sum of both Iris Segmentation and Localization.

The values in the table are the performance scores of the methods on the iris segmentation and localization tasks. The best values are bolded based on their rank sum on all evaluation measures. The rank-sum decreases in direct proportion to the overall performance. The higher the rank-sum, the better the performance of the method.

The E1 (Rank) and E2 (Rank) columns refer to the evaluation measures for Iris Segmentation. The mDice (Rank) and mHdie (Rank) columns refer to the evaluation measures for Iris Localization. The Rank Sum columns refer to the sum of the ranks of the evaluation measures for each task.

The table shows that the proposed method performs best on the iris segmentation and localization tasks, with a rank sum of 82. This is followed by the EyeCool method with a rank sum of 112, the Lao Yang Sprint Team with a rank sum of 56, and the U-net (baseline) with a rank sum of 405.

Result of training the proposed method: 3530 images are used to train the model. The dataset was divided into three sets: training, validation, and testing sets. We used 3178 items of training data, 176 pieces of validation data, and 176 pieces of testing data. Hyper-parameter optimization is a critical step for training an effective ML model. It involves selecting the appropriate parameters to provide well-fit weights and biases for the deep learning U-net model. Compiler (Adam), Learning rate (0.0001), and accuracy metric were all used. Negative dice coefficient [9] (to be maximized) and loss metric based on dice coefficients (to be minimized) were used as well.

The model complexity is another parameter that needs to be tweaked (for instance, the training complexity forced by the size of the U-net). As a result, training on deep U-nets (depth = 4) and shallow U-nets (depth = 3) is required. For "small data challenges," the model with fewer parameters is usually the best option. The loss curves at the end of model training illustrate how effective the training procedure was. Figures 6 and 7 demonstrate the loss curves for our implementation.

The model mentioned above managed to predict with an accuracy of 98% on the test set. There are just two publicly available open-source results and code of the competition, namely EyeCool and Lao Yang Sprint Team.

EyeCool: A newly developed U-Net model is used to segment the iris mask as well as the inner and outer boundaries of the iris at the same time. There are two types of enhancements. The EfficientNet-B5 is utilized for feature extraction, and a similar technique is suggested to improve the encoder's ability to consider iris borders. The goal of this technique is
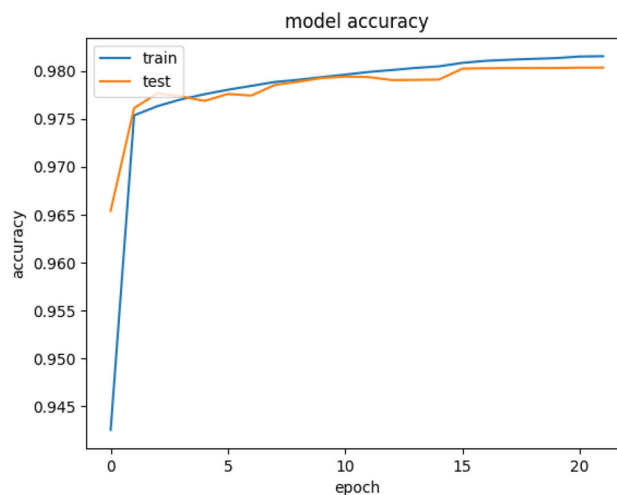
Fig. 6 Accuracy of our proposed method on the dataset
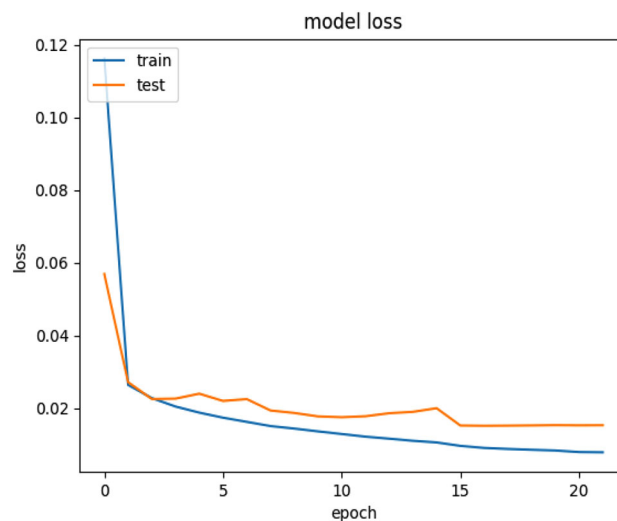
Fig. 7 Loss value of our proposed method on the dataset

to insert many side-out layers after the decoder convolutional layers to forecast the iris boundary heat maps and then use these heat maps as the weight to force the decoder to consider the iris borders [7].

Lao Yang Sprint Team: ResNet-34 is a backbone but with a U-Net-based design. As an extra job, this group examines iris localization to tackle the iris segmentation problem by segmenting the mask within the iris border. U-Net in two separate models for iris segmentation and localization solves the poor performance of the multi-tasking model for these two tasks where the only variation between the segmentation and localization networks is the final head layer. Two parallel convolution layers are used by the localization head to produce the inner and outer border masks. In contrast, just one convolution layer is used by the segmentation head to create the iris mask. The Lao Yang Sprint Team used dice loss

with binary cross-entropy loss as a part of this technique, in addition to significant data augmentation [8].

*Results of iris segmentation and localiza*tion: It is shown in Table 1 that the metric rankings on all testing sets are totaled to represent the overall segmentation and localization performance on various datasets, which are decided based on the respective metric score. The total Dice index for the inner or outer iris margin was calculated by averaging similar Dice indices across all test pictures. Finally, the challenge utilized the mean value of the total mean Dice Coefficient ($mDice$) iris Dice indexes for ranking. Like the Dice index, the challenge calculated the mean normalized Hausdorff distance ($mHdis$) for the iris' inner and outer margins. Finally, the ranking relied on the average of two normalized Hausdorff distances referred to as $mHdis$.

As can be seen from Table 2, different methods demonstrate almost identical rankings on the $E1$, $E2$, $mDice$, and $mHdis$, which reflect that the functionality of the metric is equivalent to some extent. This may be why the Noisy Iris Challenge Evaluation and Part I (NICE-I) competition only adopted the $E1$ as the measure for ranking the methods. Further observations reveal the diverse results of the two methods. Lao Yang's method, with the lowest sum rank of 14 for $E1$ and $E2$ in iris segmentation, is the best solution. Moreover, that method has achieved the best result on iris localization with a sum rank of 56. Even though our proposed method has achieved the second rank in overall comparison, our proposed method achieved better results on CASIA-Iris-M1 during evaluation. Using our collected dataset based on a mobile device might be the reason we have achieved better performance on the CASIA-Iris-M1 test set.

According to the authors, the segmentation and localization of the iris are essential techniques in iris pre-processing when it comes to conventional iris identification. As a result, to determine whether a given entity is suitable for use as a pre-processing plug-in in an existing iris recognition system, iris segmentation, and localization performance must be thoroughly evaluated. As indicated in Table 2, the entries are arranged in descending order based on their rank sums on the iris segmentation and localization tasks. The rank-sum decreases in direct proportion to the overall performance. The research demonstrates that the suggested method provides excellent performance. We received the highest ranking on CASIA-Iris-M1 because the two-independent single-tasking model setup makes model training easier than the multi-tasking model used in the U-net approach. Our method uses a single-tasking model setup that is less complicated than the U-net approach. We propose a new transfer learning training approach to improve the model's generalization capabilities, as well as a model ensemble-based Test-Time Augmentation strategy to refine the prediction results, which are both lacking in our proposed method. As the last point, adopting MobileNetV2 as the backbone of our proposed approach

results in a more robust feature extraction capacity than U-net.

5.5 Experiment 2: qualitative evaluation

This part compares and contrasts several approaches regarding anticipated binary iris segmentation masks and iris inner and outer boundaries. Two representative and challenging samples from the CASIA-Iris-Distance, CASIA-Iris-Complex-Occlusion, CASIA-Iris-Complex-Off-angle, CASIA-Iris-M1, and CASIA-Iris-Africa datasets were chosen separately to demonstrate the methods' iris segmentation and localization performance in different non-cooperative environments (Figs. 8, 9, and 10). The best of the applied methods' iris segmentation and localization results are shown in Figs. 8, 9, and 10. They indicate that the best approach of CASIA-Iris-M1 is robust with various types of degraded iris pictures in non-cooperative contexts; however, further enhancements are needed to better cope with some complex cases.

6 Discussion of the findings

In this study, we proposed a method for iris segmentation and localization using U-Net architecture and introduced a new dataset. When compared to previous work, our approach performed better, with an overall rank of 82.

One possible weakness of our method is that it may not perform well if the center of the iris is not selected correctly. This is a common issue with iris localization and is often caused by variations in iris shape and pupil dilation. To overcome this limitation, we used a thresholding technique to select the iris region and data augmentation to train the model on various iris shapes.

Our new dataset consists of 1000 iris images, which we collected from different people and annotated with ground truth labels. We believe that this dataset will be helpful for further research on iris recognition and can also be used to evaluate the performance of other segmentation and localization methods.

Our results demonstrate the effectiveness of U-Net architecture for iris segmentation and localization, and we believe that this method has potential for real-world applications, such as biometric identification and access control systems. However, there is still room for improvement, and future research can explore other architectures and techniques further to enhance the accuracy and robustness of iris recognition systems.

In conclusion, our study provides a valuable contribution to the field of iris recognition by proposing a new method and dataset for iris segmentation and localization. Our results show that U-Net architecture can achieve high accuracy in
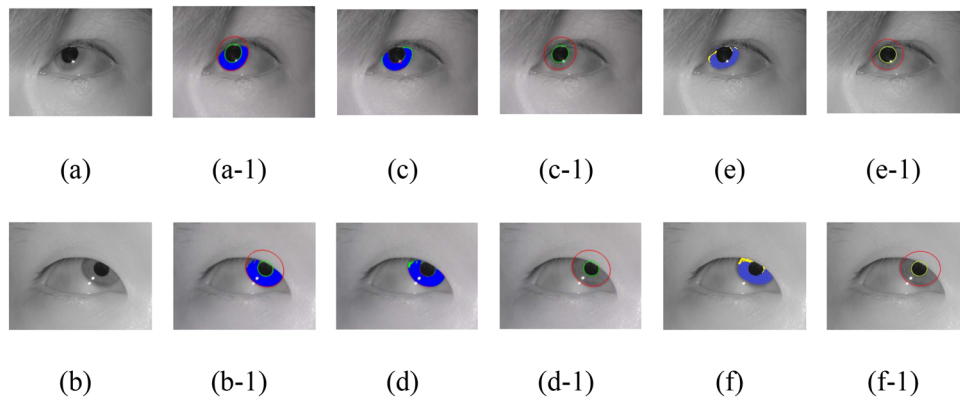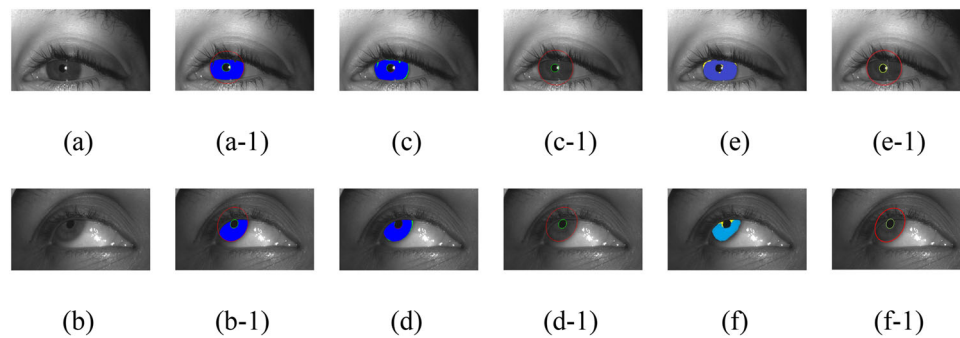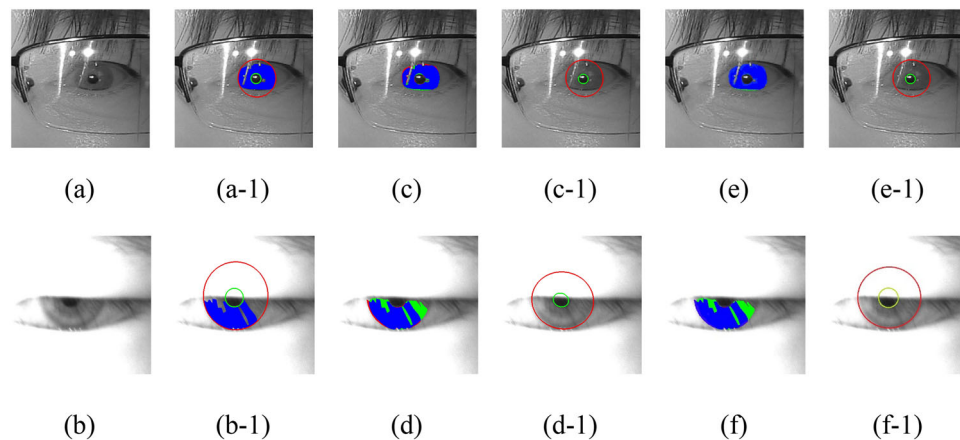
Fig. 8 Samples of iris segmentation and localization results from the best two methods (two related works and our proposed method) on the CASIA-Iris-Complex-Off angle. **a** real sample, **a-1**) grand truth for (**a**), **b** real sample, **b-1**) grand truth for (**b**), **c** Lao Yang Sprint Team's iris segmentation, **c-1**) Lao Yang Sprint Team's iris localization, **e** our proposed method iris recognition, and **e-1**) our proposed method's iris localization. (**d**), (**d-1**), (**f**), and (**f-1**) results for methods like the sample (**a**)

Fig. 9 Samples of iris segmentation and localization result from the best two methods (two related works and our proposed method) on the CASIA-Iris-Africa. **a** real sample, **a-1**) grand truth for (**a**), **b** real sample, **b-1**) grand truth for (**b**), **c** Lao Yang Sprint Team's iris segmentation, **c-1**) Lao Yang Sprint Team's iris localization, **e** our proposed method iris recognition, and **e-1**) our proposed method's iris localization. (**d**), (**d-1**), (**f**), and (**f-1**) results for methods like the sample (**a**)

Fig. 10 Samples of iris segmentation and localization result from the best two methods (two related works and our proposed method) on the CASIA-Iris-M1. **a** real sample, **a-1**) grand truth for (**a**), **b** real sample, **b-1**) grand truth for (**b**), **c** Lao Yang Sprint Team's iris segmentation, **c-1**) Lao Yang Sprint Team's iris localization, **e** our proposed method iris recognition, and **e-1**) our proposed method's iris localization. (**d**), (**d-1**), (**f**), and (**f-1**) results for methods like the sample (**a**)

this task, and we hope that our work will inspire further research and development in this area.

7 Conclusion

This paper presents a U-Net method with a pre-trained MobileNetV2 deep neural network method for iris segmentation and localization. In this method, to extract better geometric constraints of inner and outer boundaries for both pupil and iris, we propose an efficient encoder–decoder framework to improve iris segmentation and localization performance. Meanwhile, we use three well-known publicly available NIR iris datasets, along with our collected dataset, which is also publicly available. We compare our proposed method with state-of-the-art methods on the three iris datasets that prove the top performance of our proposed method. As for future research, we will enhance the iris segmentation and localization efficiency to improve state-of-the-art NIR iris image datasets.

Availability of data and materials Not applicable.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical approval and consent to participate Forasmuch as we need participants (volunteers) to collect datasets for investigation of the iris segmentation system's performance, the ethics committee of Warsaw University of Technology has approved the experiment protocol to collect the data.

Consent for publication I, the undersigned, give my consent for the publication of identifiable details, which can include photographs to be published in the Journal of Big Data.

References

1. Iwasokun, G.B., Aladesaye, A.: Noise attenuation and ridge processing technique for fingerprint bit minimization. Iran J. Comput. Sci. **5**(1), 1–16 (2022). https://doi.org/10.1007/s42044-021-00081-7
2. De Marsico, M., Nappi, M., Proença, H.: Results from miche ii—mobile iris challenge evaluation ii. Pattern Recogn. Lett. **91**, 3–10 (2017). https://doi.org/10.1016/j.patrec.2016.12.013
3. Singh, S., Prasad, S.: Techniques and challenges of face recognition: a critical review. Procedia Comput. Sci. **143**, 536–543 (2018). https://doi.org/10.1016/j.procs.2018.10.427
4. Trokielewicz, M., Czajka, A., Maciejewicz, P.: Post-mortem iris recognition with deep-learning-based image segmentation. Image Vis. Comput. **94**, 103866 (2020). https://doi.org/10.1016/j.imavis.2019.103866
5. Zhang, W., Lu, X., Gu, Y., Liu, Y., Meng, X., Li, J.: A robust iris segmentation scheme based on improved U-net. IEEE Access **7**, 85082–85089 (2019). https://doi.org/10.1109/access.2019.2924464
6. He, Z., Tan, T., Sun, Z., Qiu, X.: Toward accurate and fast iris segmentation for iris biometrics. IEEE Trans. Pattern Anal. Mach. Intell. **31**(9), 1670–1684 (2008). https://doi.org/10.1109/TPAMI.2008.183
7. Lian, S., Luo, Z., Zhong, Z., Lin, X., Su, S., Li, S.: Attention guided U-Net for accurate iris segmentation. J. Vis. Commun. Image Represent. **56**, 296–304 (2018). https://doi.org/10.1016/j.jvcir.2018.10.001
8. Wang, C. et al.: NIR iris challenge evaluation in non-cooperative environments: Segmentation and localization. In: 2021 IEEE International Joint Conference on Biometrics (IJCB), 2021: IEEE, pp. 1–10, https://doi.org/10.1109/IJCB52358.2021.9484336
9. Bagherzadeh, J., Asil, H.: A review of various semi-supervised learning models with a deep learning and memory approach. Iran J. Comput. Sci. **2**(2), 65–80 (2019). https://doi.org/10.1007/s42044-018-00027-6
10. Han, Y.L., Min, T.H., Park, R.H.: Efficient iris localisation using a guided filter. IET Image Proc. **9**(5), 405–412 (2015). https://doi.org/10.1049/iet-ipr.2014.0496
11. Gad, R., et al.: Iris recognition using multi-algorithmic approaches for cognitive internet of things (CIoT) framework. Futur. Gener. Comput. Syst. **89**, 178–191 (2018). https://doi.org/10.1016/j.future.2018.06.020
12. Barpanda, S.S., Sa, P.K., Marques, O., Majhi, B., Bakshi, S.: Iris recognition with tunable filter bank based feature. Multimed. Tools Appl. **77**(6), 7637–7674 (2018). https://doi.org/10.1007/s11042-017-4668-z
13. Singh, G., Singh, R.K., Saha, R., Agarwal, N.: IWT based iris recognition for image authentication. Procedia Comput. Sci. **171**, 1868–1876 (2020). https://doi.org/10.1016/j.procs.2020.04.200
14. Daugman, J.: How iris recognition works. In: The essential guide to image processing, pp. 715–739. Elsevier (2009)
15. De Marsico, M., Nappi, M., Proença, H.: Mobile iris challenge evaluation part II (miche II). Pattern Recogn. Lett. **91**, 1–2 (2017). https://doi.org/10.1016/j.patrec.2017.04.006
16. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International

Conference on Medical image computing and computer-assisted intervention, 2015: Springer, pp. 234–241, https://doi.org/10.1007/978-3-319-24574-4_28.

17. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4510–4520, https://doi.org/10.1109/CVPR.2018.00474

18. Nguyen, H.V., Bai, L.: Cosine similarity metric learning for face verification. In: Asian conference on computer vision, 2010: Springer, pp. 709–720, https://doi.org/10.1007/978-3-642-19309-5_55

Springer