Intelligent Human Iris Recognition System Based on Deep Learning Models

Andreea Negoițescu

Faculty of Mathematics and Computer Science, "Babeș-Bolyai" University, Cluj-Napoca, Romania

Keywords: Deep Learning, Iris, Biometrics, Segmentation, Recognition.

Abstract: This research paper presents the development of an intelligent biometric system which performs human iris recognition. The software application that incorporates it is called KEYE. Deep learning models are implemented to segment and recognize the users' irises at authentication. Iris segmentation uses a modified version of the U-Net convolutional neural network, trained and validated on images from the I-SOCIAL-DB dataset. The experimental results prove a maximum validation accuracy of 98.98% and a Dice score of 0.93. The extraction of features from the segmented images is done using part of the layers of the pre-trained DenseNet-201 neural network. For classification, the KEYE-DB dataset with visible light spectrum images was created. The accuracy obtained after testing the recognition model is 99.98%. The precision, specificity, recall and F1 score exceed 0.9955, while the error and the false positive rate are almost zero, following the conducted experiments. The performance of the biometric system has proven to be gratifying.

1 INTRODUCTION

Due to the increasing interest in the development of science and technology worldwide, there is also an intense focus on security and, implicitly, on the development of intelligent systems that use biometric recognition for human identification and verification. Such an authentication system represents the basis of the KEYE mobile application developed from this study, which aims to keep users' credentials and photos safe from impostors.

Biometrics is defined, in (Tahir and Anghelus, 2019), as the technology that analyzes the physiological and behavioral features of people, with the aim of identifying and authorizing them. According to (Abdulkader et al., 2015), it is the most secure human authentication method among the existing ones: biometric, knowledge-based and possession-based. Biometrics is a vast field and is intensively studied by researchers, because it provides information used in the design and implementation of security technologies. It involves a wide range of human recognition techniques and portrays the unique and detailed characteristics of individuals.

The complexity and uniqueness of the human iris is fascinating compared to other biometric traits. The arrangement of pigments, the pattern of the collarette, the distribution of fibers and blood vessels, give this natural structure a huge potential for use in the field of security. Thus, the aim of this study is to demonstrate the reliability and accuracy of human iris feature recognition using artificial intelligence. The objectives of this study are: researching and implementing innovative methods in the field of iris biometrics, obtaining performant results after applying deep learning algorithms, demonstrating the uniqueness of the iris as a biometric characteristic and ensuring a high degree of personal data security.

From the first studies on iris recognition, there has been remarkable progress in the diversity and performance of the algorithms used for this purpose. Starting from images captured in infrared light, more and more emphasis has been placed on the use of datasets containing images from the visible light spectrum. These images are captured in uncontrolled environments, where iris region visibility conditions are not necessarily favorable, as the ones in the datasets used in this study. It is resorted to the development of methods with an increased degree of complexity for the purpose of iris segmentation and classification, based mainly on machine learning techniques, or to the improvement of already existing ones.

This paper focuses on the implementation of both segmentation and recognition deep learning methods. The unique contributions of this study consist, firstly, in using a deep learning model that consists of a vari-

15

ation of the U-Net architecture and trained on the I-SOCIAL-DB dataset, for iris segmentation. Secondly, the KEYE-DB dataset is created and used for the first time. Its scope is to help training and validating an iris recognition model to extract relevant features from irises and classify them, based on the DenseNet-201 neural network. Due to these implemented innovations, the biometric authentication system of KEYE mobile application proves outstanding performances.

Compared to the present study, others focus either on segmentation or recognition, or are simply not suitable for use on a mobile phone, either because of their computational complexity or due to the fact that most of them use images from the infrared light spectrum.

2 RELATED WORK

2.1 The Basis of Human Iris Recognition

The first patent on iris recognition was developed in 1987 (Flom and Safir, 1987). Most of the currently existing iris recognition methods have their basis on the algorithm of the British researcher John Daugman, patented in 1994. In his work, (Daugman, 1994), iris localization implies a integro-differential operator in order to demarcate the inner and the outer contours of the iris. Then, geometric normalization is performed and Gabor filters are used to represent the obtained rectangular image of the iris in binary code. In the iris code matching process, authenticity verification is performed by calculating the Hamming distance between pairs of codes. In order for two codes to define the iris of the same person, the value of the Hamming distance, scored between 0 and 1 inclusive, must be as close as possible to 0. The work (Wildes, 1997) investigates the application of the Hough transform for the purpose of detecting the iris and Gaussian filters for the representation of its code.

2.2 Recent Studies Regarding Human Iris Segmentation and Recognition

A study that presents a complex approach is (Gangwar et al., 2019). From the visible light spectrum, it uses the UBIRIS.v1, UBIRIS.v2, UTIRIS V.1 and MICHE-1 datasets in various combinations. Iris segmentation is performed using a pair of convolutional neural networks. The first network, inspired by YOLO (Redmon et al., 2016), locates the iris and pupil. It receives as input an image of 448×448 pixels and the obtained accuracy is 96.78%. The second net-

work, similar to SegNet (Badrinarayanan et al., 2017), receives an input of size 100×100 pixels. It performs pixel-level segmentation of the localized region, resulting in an F1 score of 96.98%. For iris binary code generation, the paper proposes the DeepIrisNet2 architecture, with approximately 100 layers, which achieves remarkable results without the need for precise segmentation of images or their normalization. For the UBIRIS.v2 database, an error EER = 8.51% is obtained, while for MICHE-1 it varies between 1.05% and 3.98%. Such a model is extremely computational expensive to be used on a mobile phone.

Another approach that accepts segmented but non-normalized images is the ThirdEye system, described in (Ahmad and Fuller, 2019). It consists of triple convolutional neural networks, obtained by modifying the architecture of ResNet-50 (He et al., 2016). The model is trained using three input irises at once, each of size 200×200 pixels: two are from the same class and one is from a different class. The recognition error for the UBIRIS.v2 dataset is EER = 9.20% and the false rejection rate is FRR = 60%.

The (Ahmadi et al., 2019) study combines, for image feature extraction, two dimensional Gabor filters, step filtering and polynomial filtering. Then, for matching purposes, it uses a neural network with basic radial functions along with a genetic algorithm. Using the UBIRIS.v1 database, it achieves an accuracy of 99.9869% after only 10 iterations, with 10 neurons per layer and the following parameters: population of 150, maximum number of generations equal to 10, selection factor equal to 3, mutation of 0.35, crossover of 0.5 and recombination of 0.15. However, the process of locating, segmenting and normalizing the iris region is not specified.

The work (Yang et al., 2021) provides the encoder-decoder architecture of DualSANet. Being included in the pre-trained ResNet-18 (He et al., 2016) network, the encoder represents spatially corresponding features at multiple levels. For these features to fusion, a module based on spatial attention, integrated in the decoder, is introduced. It generates dual feature representations that contain complementary discriminative information. The described recognition model proves a great performance, having a minimum error EER = 0.27% and a rate FRR = 0.31%. It does not specify the behavior of the network on images in the visible light spectrum, as the experiments are performed on infrared images. They are first segmented, then normalized using the Daugman Rubber Sheet Model method and resized to 64×512 pixels, before feature extraction.

In the study (Lee et al., 2021), iris recognition is experimented with NICE.II and MICHE databases.

The iris region is detected then normalized along with two periocular regions. The reconstruction of the normalized blurred regions of the iris is done using the DeblurGAN model (Kupyn et al., 2018). Each triplet of normalized images becomes the input of a convolutional neural network that extracts a feature vector. 4096 dimensional features are distinguished across the layers. To check whether two irises correspond to the same person, the Euclidean distances between three pairs of feature vectors are calculated. This results in three scores are merged into one using a support vector machine. An error of 14.18% is obtained for images captured with a Samsung Galaxy S4 phone and 17.02% for those taken with an iPhone 5.

A variant of the algorithm that does not require the use of artificial neural networks is described in (Singh et al., 2020). The illumination and contrast of the images are improved, then the median filter is applied to reduce their noise. For iris localization and segmentation, the circular Hough transform and the total relative variation model measure and regularize the local variation of pixels. The obtained region is normalized with the Daugman Rubber Sheet Model method and decomposed using the four-level integer wavelet transform (IWT), which generates 256 frequency sub-bands. Only the lower 192 sub-bands, which produce a 192-bit binary code, are considered, by comparing their energies with a previously calculated corresponding threshold value. This is done by finding the Hamming distance between them. The algorithm achieves, on the UBIRIS.v2 dataset, an accuracy of 98.9% in segmentation and 98.02% in recognition.

Regarding the semantic segmentation of the iris, one of the most recent approaches is mentioned in (Pourafkham and Khotanlou, 2023). This presents the ES-Net architecture, which uses an ESP (Efficient Spatial Pyramid) block (Mehta et al., 2018) to minimize the time complexity of a network model inspired by the U-Net architecture (Ronneberger et al., 2015), but also an attention mechanism (Vaswani et al., 2017) to enhance performance. Through the experiments, a MIOU (Mean Intersection Over Union) score of 93.61% and an F1 score of 97.03% are obtained for the UBIRIS.v2 dataset.

The study (Nourmohammadi Khiarak et al., 2023) proposes a new dataset, called KartalOl. It contains images from the visible light spectrum, captured using a mobile phone camera. As a segmentation architecture, Mobile-Unet is built, consisting of the pretrained MobileNetV2 model (Sandler et al., 2018), integrated into the encoder part of the U-Net network. It achieves 98% accuracy on validation data.

3 PROPOSED IRIS SEGMENTATION APPROACH

3.1 I-SOCIAL-DB Dataset

For segmentation, the I-SOCIAL-DB, namely Iris Social Database dataset (Donida Labati et al., 2021) was used. It contains 3286 color images from the visible light spectrum, collected from a sample of 400 subjects, in uncontrolled environments. These were obtained by extracting two eye regions of 300×350 pixels each, corresponding to the left and right eye, from 1643 high-resolution portrait images. Because they were collected from various online public sources, both the devices that captured the images and the distances from which they were taken are unknown.
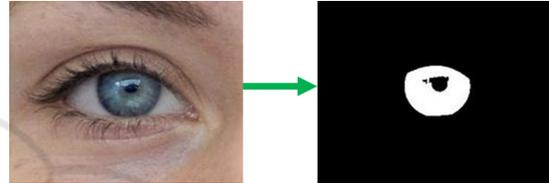
Figure 1: I-SOCIAL-DB ocular region sample.

Each image in this dataset corresponds to a manually constructed segmentation mask at the pixel level, as in the example in Figure 1. The mask highlights, through white pixels, the iris as the region of interest, excluding reflections and other possible occlusions. The portion of the iris after segmentation represents, on average, 71.4% of the total area of the ring formed by the circles that approximate the inner and outer border of the iris.

3.2 U-Net Architecture

The architecture of the U-Net convolutional neural network was firstly introduced in the paper (Ronneberger et al., 2015). Even though it was originally intended for the processing of microscopic biomedical images, it also proves extraordinary results in the case of semantic segmentation of human irises photographed under various conditions. The advantage of this network lies, in addition to the speed of segmentation, in the useful ability to learn from a relatively small set of data. This is proven by the great performances shown by the network through the experiments conducted in this research, as the used dataset for learning contains only 3286 images. Also, being a fully convolutional network, the sizes of the outputs adapt to those of the input image, so their resolutions and number of channels can vary. The U-Net architecture is of encoder-decoder type, being formed of

a contraction path, followed by an expansion path. They are connected to each other symmetrically, to preserve information lost by contracting. The network contains 23 convolutional layers.

The contraction path consists of repetitive steps which extract relevant features from images. Each step involves the application of two convolutions with a filter of 3×3. Each of them is followed by a ReLU function for activation, at the end of which a 2×2 max-pooling operation is performed with a step of 2. To compensate for this reduction in spatial dimensionality caused by subsampling, the number of channels of the feature maps is doubled at each iteration.

The expansion path is relatively symmetrical and achieves a precise localization, at pixel level, of the region of interest. Each step represents an upsampling of the feature map. Then, a 2×2 filter convolution that halves the number of channels is applied, a concatenation with a copy of the clipped feature map from the corresponding step of the shrinking path and two 3×3 convolutions, followed by one ReLU activation function each. The final layer maps each 64-element feature vector to the desired number of classes.

In this study, the U-Net neural network is adapted to act as a binary classifier, assigning each pixel in each input image a corresponding class, iris or non-iris. The difference from the original model described in (Ronneberger et al., 2015) consists, firstly, in the use of padding in the case of convolutions and the BatchNorm2d layer (Ioffe and Szegedy, 2015), which has the role of normalizing the activations between network layers. Since the bias will be canceled by this normalization layer, its existence is no longer necessary. Also, because color input images are provided to the network, the input layer contains 3 channels. For semantic segmentation, a binary classification of the pixels is performed, so the existence of a single output channel is sufficient.

Another important step is to ensure that the network works properly for any input image dimensions by performing resizing when concatenating the feature maps. Otherwise, if the input dimensions were not divisible by 2 at each of the four steps at which the max-pooling operation is performed within the contracting path, some pixels would be lost. For example, if max-pooling is performed on an image of size 175×175, an image of 87×87 pixels will result. In the expansion path, when oversampled, it will end up being only 174×174 pixels in size. In order to be concatenated with the original image, they must be brought to the same size.

3.3 Training and Validation

The described model was trained and validated on 3-channel color images from the I-SOCIAL-DB dataset, both original size of 300×350 pixels and resized to 160×240 pixels. The first 3000 images, respectively masks, were kept for training and the next 286 images, respectively masks, for validation. In this way, the train : validation ratio is approximately 90 : 10.

By feeding the network batches of 16 images, 188 steps are performed in each training epoch. The chosen loss function is Binary Cross Entropy. For optimization, the Adam algorithm is used, with a constant learning rate of 0.001. The model is saved in a checkpoint every time the validation accuracy increases following the completion of an epoch.

3.4 Experimental Results

For images of 300×350 pixels, the model proves outstanding performance even from the first epoch. On the training dataset, an accuracy of 98.01%, an average loss of 0.67731, a final loss of 0.276 and a Dice score of 0.83326 are obtained. Upon validation, an accuracy of 97.85%, and an average loss equal to 0.67600 is obtained. The Dice score increases by approximately 0.00222, reaching a value of 0.83548. The model is trained over 18 epochs, each taking between 2 and 4 hours to run. This number was chosen because after epoch 18 the model performance does not improve anymore. The results can be observed in the graphics from Figure 2 and Figure 3, where the color blue is used for training evolution and the orange color corresponds to validation evolution.
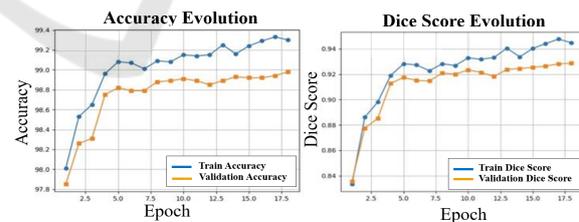
Figure 2: Analysis of accuracy and Dice score after training and validating the model on 300×350 pixels images.
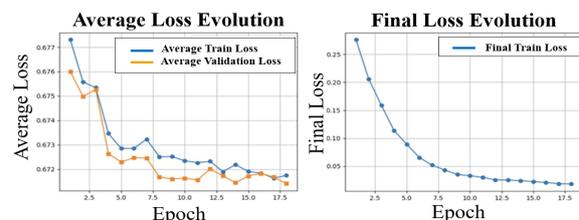
Figure 3: Average and final loss progress of the model during training and validation on 300×350 pixels images.

Table 1: Comparison of model performance for images of different sizes.

Criterion		Training		Validation	
		300×350	160×240	300×350	160×240
Accuracy	first epoch	98.01%	98.33%	97.85%	98.07%
	last epoch	99.30%	99.26%	98.98%	98.97%
	maximum	99.33%	99.27%	98.98%	98.97%
Dice score	first epoch	0.83326	0.86242	0.83548	0.85592
	last epoch	0.94474	0.94195	0.92860	0.92833
	maximum	0.94765	0.94195	0.92860	0.92833
Medium loss	first epoch	0.67731	0.67570	0.67600	0.67488
	last epoch	0.67174	0.67185	0.67141	0.67138
	minimum	0.67163	0.67179	0.67141	0.67138
Final loss	first epoch	0.276	0.184	-	-
	last epoch	0.0182	0.0188	-	-
	minimum	0.0182	0.0173	-	-

At the end of the 18th epoch, an accuracy of 99.30% is noted on the training data, 0.03% lower than the maximum, which was achieved in the 17th epoch. The training Dice score is maximum in the penultimate epoch, reaching 0.94765. In epoch 18, it drops to 0.94474. The average loss in epoch 18 is equal to 0.67174, while in the previous epoch it is lower, reaching the minimum value of 0.67163. Considering the value of the final loss at the end of each of the 18 epochs, without calculating the average of its values within them, its decrease during training is achieved progressively, from 0.276 in the first epoch to 0.0182 in the last. Upon validation, the accuracy and the Dice score reach the maximum of 98.98%, respectively 0.92860, in the last epoch, in which the average loss is also minimal, being equal to 0.67141. From the first to the last epoch, a 1.13% increase in accuracy is reported. The Dice score also increases by about 0.09312 and the average loss decreases by about 0.0046.

The segmentation results of 4 validation images, along with their original masks above, are illustrated in Figure 4. After training the model for 18 epochs, it recognizes iris reflections with high accuracy.
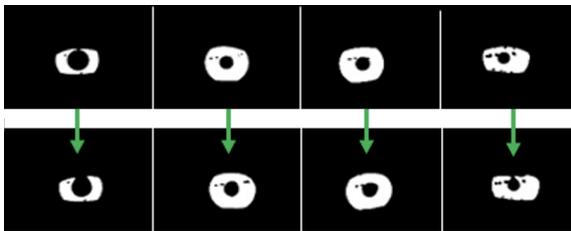
Figure 4: Original vs. predicted iris masks.

To reduce the execution time of an epoch to a maximum of one hour, it is experimented with images resized to 160×240 pixels. For these, the model proves, in the first of the 17 total epochs, better results

compared to the previously described approach. This number was chosen because after epoch 17 the model performance does not improve anymore. A comparison of the results from the first and final epochs and the best values of the metrics obtained for both dimensions of the images, is made in Table 1.

4 PROPOSED IRIS RECOGNITION APPROACH

4.1 KEYE-DB Dataset Creation

The KEYE-DB dataset contains 1370 3-channel color images from the visible light spectrum captured by various mobile phones with high-resolution cameras. To capture them, both the front and back cameras of the devices were used, with and without flash. They were positioned at distances between 7 and 10 centimeters from the eyeball of the subjects, who were in various environments with natural or artificial light. This study involved 36 subjects, 22 women and 14 men, with irises of various colors and shades. The subjects belong to several age categories: 5-20 years (14%), 20-35 years (33%), 35-50 years (25%), 50-65 years (20%) and 65-80 years (8%). Most of them are between 20 and 35 years old. Both the left and the right iris were photographed for each. Between 25 and 50 photographs were collected for each individual, with an average of approximately 38 photographs per person. As the privacy of the subjects is prioritized, the dataset is not made publicly accessible.

Most images have been cropped to approximate 3:4 or 4:3 aspect ratios. They were then resized to 300×350 pixels height×width for further segmentation. In the case of older subjects, a cropping of the images that more closely frames the eye region than

in the case of the others was considered. Thus, a precise segmentation was ensured, which is not disturbed by the uneven distribution of light on the skin folds.

Each binary mask obtained after segmentation was transformed back to the original image dimensions. To avoid false positive regions in the predicted mask as much as possible, the largest area of white pixels is found, as it is most likely to describe the iris region. Then, the radius and the center of the smallest enclosing circle are calculated to simulate the outer boundaries of the iris. With their help, the coordinates of the square that inscribes this circle are determined. Next, a multiplication of the pixel values of the original image with those of the mask pixels is performed, to obtain an image in which the white pixels in the mask are replaced by the corresponding ones in the original image. The resulting image is cropped based on the coordinates of the previously obtained square, resized to 300×300 pixels and saved in the folder corresponding to the subject to which the iris belongs. These stages are summarized in Figure 5.
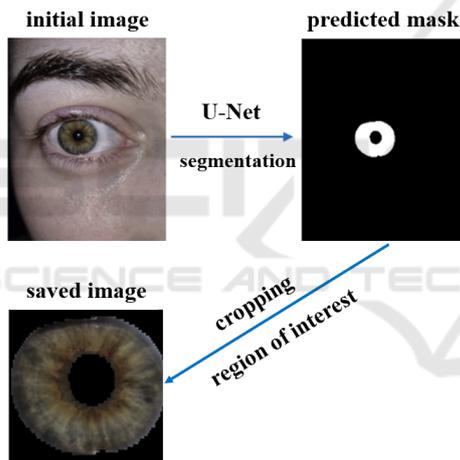
Figure 5: General steps of obtaining KEYE-DB images.
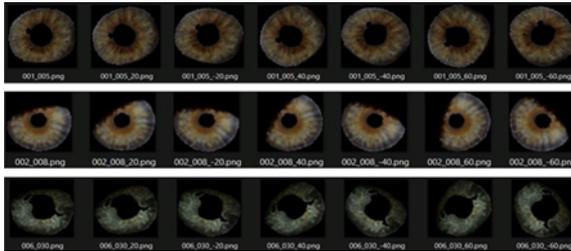
Figure 6: Iris images augmentations.

The entire process is applied to all captured images. After completion, the dataset can be augmented by rotating each image by -60, -40, -20, 20, 40, and 60 degrees, respectively. For each image among the approximately 30 of a subject, 6 more images are ob-

tained, as in Figure 6. In this way, each subject will have 7 times more images of their own irises than originally. This type of augmentation is necessary because various factors can obscure the iris region and cause significant areas of black pixels, whose orientation is not relevant.

4.2 DenseNet-201 Architecture

The DenseNet-201 convolutional neural network architecture was first introduced in the study (Huang et al., 2017). Its major advantage is the presence of dense blocks, where each layer has direct connections to all the others. With their help, the risk of losing information through the network layers is considerably reduced, while the direction of its transmission remains constant. Dense connections reduce the number of parameters and avoid possible overfitting tendencies, which is why DenseNet-201 was chosen to be used in this research.

Considering that each Conv layer corresponds to the triplet of BN (Batch Normalization) (Ioffe and Szegedy, 2015), ReLU and Conv layers, an example of the DenseNet-201 network architecture is shown in Table 2.

Table 2: DenseNet-201 architecture with a growth rate of 32 (Huang et al., 2017).

Layers	Output dimension	DenseNet-201
Convolution	112×112	7×7 Conv, step=2
Pooling	56×56	3×3 max-pooling, step=2
Dense Block (1)	56×56	[1x1 Conv and 3x3 Conv] x 6
Transition layer (1)	56×56	1×1 Conv
	28×28	2×2 average pooling, step=2
Dense block (2)	28×28	[1x1 Conv and 3x3 Conv] x 12
Transition layer (2)	28×28	1×1 Conv
	14×14	2×2 average pooling, step=2
Dense block (3)	14×14	[1x1 Conv and 3x3 Conv] x 48
Transition layer (3)	14×14	1×1 Conv
	7×7	2×2 average pooling, step=2
Dense Block (4)	7×7	[1x1 Conv and 3x3 Conv] x 32
Classification layer	1×1	7×7 global average pooling
		completely connected and Softmax

The inputs of each layer are represented by the concatenation of feature maps received from previous layers and the total number of connections between L layers has a value equal to c_L.

$$c_L = \frac{L \cdot (L+1)}{2} \qquad (1)$$

The layer l that performs the nonlinear transformations denoted by H_l receives, at input, the concatenated outputs x_0, x_1, up to x_{l-1}, of the previous layers. Thus, its output is given by x_l.

$$x_l = H_l(x_{l-1}) + x_{l-1} = H_l([x_0, \ldots, x_{l-1}]) \qquad (2)$$

If each function H_l produces k feature maps, which represent the growth rate of the network, layer l will have, at input, m_l feature maps.

$$m_l = k_0 + k \cdot (l-1) \qquad (3)$$

4.3 Feature Extraction and Classification

In order to extract the relevant features from the previously segmented iris images, the DenseNet-201 network is used, from which the last 58 layers are removed. It is pre-trained on the ImageNet dataset, which contains 1281167 images, divided into 1000 classes. The features are extracted using the weights resulting from learning based on the data in that set and become the inputs of a multi-class classifier represented by a simple artificial neural network using a Flatten and a Dense layer with Softmax activation. For each feature, the prediction of the classifier is a probability distribution, indicating the percentage match of the image to each of the existing classes.

4.4 Training and Validation

The classification model was trained and validated over 50 epochs using batches of 8 images from the KEYE-DB dataset, with and without augmentation. This batch size was chosen as it proved the best validation results. It is small enough to provide frequent gradient update and great generalization potential, but also large enough to be used on a relative small dataset like KEYE-DB. The images were randomly split so that 30% of them were dedicated to validation and 70% to training. The used loss is Categorical Cross Entropy and, as an optimizer, the Adam algorithm was chosen, with a learning rate equal to 0.001. The model is saved in a checkpoint each time the validation accuracy increases following the completion of an epoch. This technique is very useful, as it provides the possibility of continuing the training starting from the iteration that proved the best result previously.

4.5 Experimental Results

After training the model on the KEYE-DB dataset without augmentation, a maximum accuracy of 100% is achieved on the training dataset after only 4 epochs, which is maintained up to epoch 50. The maximum validation accuracy is 96.594% and results at the end of epoch 20, when the validation loss is 0.1590 and the training loss is 0.0016. It then oscillates for 30 epochs, without exceeding the mentioned maximum percentage and having, in the final epoch, a value of 96.11%. The minimum training, respectively validation loss, is recorded in the last epoch, with the values of 1.9594e-04, respectively 0.1428. The model performance over epochs is shown in Figure 7.
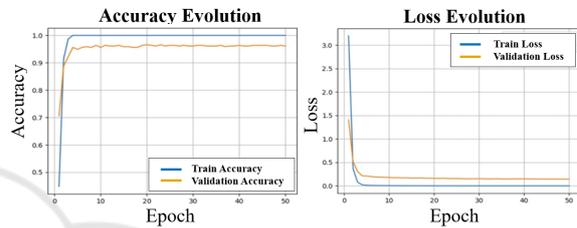
Figure 7: Analysis of accuracy and loss after training and validating on the KEYE-DB dataset without augmentation.

To increase the performance of the recognition model, the images from the KEYE-DB dataset are augmented as previously specified. Thus, after training the model, a maximum accuracy of 100% is obtained over 50 epochs on the training dataset, which is maintained from epoch 23 to the end, and 99.583% on the validation dataset, at epoch 47. In epoch 50 the same accuracy values are recorded as in epoch 47, but the loss decreases, from 6.1620e-09 to 4.4750e-09, respectively from 0.0299 to 0.0295. The model performance over epochs is shown in Figure 8.

Table 3 compares the performances of the model considering the average value of each metric for all classes. It proves the advantage of data augmentation.

Table 3: Comparative analysis of the validation performance of the model before and after data augmentation.

Performance metric	Before data augmentation	After data augmentation
Accuracy	0.998107597	0.999768277
Error	0.001892403	0.000231723
Precision	0.967860534	0.995949288
Recall	0.967048978	0.995510867
Specificity	0.999024939	0.999880640
False Positive Rate	0.000975061	0.000119360
F1 score	0.966338857	0.995705300

Table 4: Comparative analysis of proposed segmentation and recognition approaches with the existing ones in literature.

Study	Segmentation performance	Recognition performance
proposed	accuracy = 98.98% Dice = 0.93	accuracy = 99.98% F1 ≈ 99.57% EER, FRR ≈ 0
(Gangwar et al., 2019)	accuracy = 96.78% YOLO F1 = 96.98% SegNet	EER = 8.51% UBIRIS.v1 $1.05\% \leq$ EER $\leq 3.98\%$ MICHE-1
(Ahmad and Fuller, 2019)	-	EER = 9.20% FRR = 60%
(Ahmadi et al., 2019)	-	accuracy = 99.9869%
(Yang et al., 2021)	-	EER = 0.27% FRR = 0.31%
(Lee et al., 2021)	-	EER = 14.18% Samsung Galaxy S4 EER = 17.02% iPhone 5
(Singh et al., 2020)	accuracy = 98.9%	accuracy = 98.02%
(Pourafkham and Khotanlou, 2023)	MIOU = 93.61% F1 = 97.03%	-
(Nourmohammadi Khiarak et al., 2023)	accuracy = 98%	-
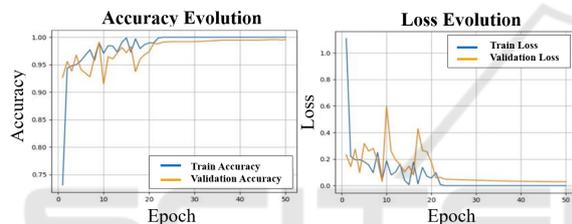
Figure 8: Analysis of accuracy and loss after training and validating on the KEYE-DB dataset after augmentation.

5 DISCUSSION

After analyzing the obtained experimental results, in the case of segmentation, it was concluded that the model trained on images of size 300×350 is the most suitable. Despite the longer training time, the decision was made considering the increased performance of the model. It proves a maximum validation accuracy equal to 98.98% and a Dice score of about 0.93.

For iris classification based on the features extracted from the segmented images, data augmentation was chosen, with the model obtaining an accuracy of 99.98% and an almost insignificant error, as well as the false positive rate. The precision, specificity, recall and F1 score all exceed the value of 0.9955. Thus, the chance of unauthorized persons logging into the application is almost zero. This statement is made considering that no photos of digital or printed iris pictures are used, as these cases have not yet been extensively tested to reach a firm conclusion.

As seen in Table 4, this study shows promising results and even competitive with those obtained in other studies in existing literature. The study

(Gangwar et al., 2019) performs recognition using iris matching, as well as classification, while (Ahmad and Fuller, 2019), (Ahmadi et al., 2019), (Yang et al., 2021) and (Lee et al., 2021) describe, in essence, iris matching approaches.

The choice for using a classification approach in this study has been made due to the fact that by trying several variants of iris matching algorithms, no satisfactory results were obtained, considering the limited public datasets resources that contain images from the visible light spectrum. The comparison in Table 4 is made with the mention that the described studies do not use the same datasets nor approaches as in the present work. An exact comparison cannot be made because there are no relevant studies in the literature that address the problem in this paper using the I-SOCIAL-DB dataset for segmentation. Also, there are no studies that use the KEYE-DB dataset since it is created within this work. However, an attempt was made to select related studies that use similar datasets.

The limitations of this study consist in the number of subjects who agreed to participate in the research by providing photos of their irises to the KEYE-DB dataset. As a future improvement, it is desired to expand the sample of users recognized by the KEYE application. Also, additional verification at authentication should be implemented to confirm the physical presence of the user who tries to access the application, such as the live recording of subtle but continuous movements of the pupil. This should stop fraudulent intents of authentication using photos of iris pictures or artificial irises.

6 CONCLUSIONS

In conclusion, following the extensive research carried out in this paper, it is confirmed with certainty that the iris is a biometric feature that meets all the necessary conditions to be used in the implementation of a reliable biometric recognition system. The experimental results are gratifying for the development of the KEYE mobile application, so the objectives of this research were achieved.

REFERENCES

Abdulkader, S., Atia, A., and Mostafa, M.-S. (2015). Authentication systems: principles and threats. *Computer and Information Science*, 8.

Ahmad, S. and Fuller, B. (2019). Thirdeye: Triplet based iris recognition without normalization. *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–9.

Ahmadi, N., Nilashi, M., Samad, S., Rashid, T., and Ahmadi, H. (2019). An intelligent method for iris recognition using supervised machine learning techniques. *Optics & Laser Technology*, 120.

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495.

Daugman, J. (1994). Biometric personal identification system based on iris analysis.

Donida Labati, R., Genovese, A., Piuri, V., Scotti, F., and Vishwakarma, S. (2021). I-social-db: A labeled database of images collected from websites and social media for iris recognition. *Image and Vision Computing*, 105(104058):1–9. 0262-8856.

Flom, L. and Safir, A. (1987). Iris recognition system.

Gangwar, A., Joshi, A., Joshi, P., and Ramachandra, R. (2019). Deepirisnet2: Learning deep-iriscodes from scratch for segmentation-robust visible wavelength and near infrared iris recognition.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269.

Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 448–456, Lille, France. PMLR.

Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., and Matas, J. (2018). Deblurgan: Blind motion deblurring using conditional adversarial networks. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8183–8192.

Lee, M. B., Kang, J. K., Yoon, H. S., and Park, K. R. (2021). Enhanced iris recognition method by generative adversarial network-based image reconstruction. *IEEE Access*, 9:10120–10135.

Mehta, S., Rastegari, M., Caspi, A., Shapiro, L. G., and Hajishirzi, H. (2018). Espnet: Efficient spatial pyramid of dilated convolutions for semantic segmentation. In *European Conference on Computer Vision*.

Nourmohammadi Khiarak, J., Golzari Oskouei, A., Salehi Nasab, S., Jaryani, F., Moafinejad, N., Pourmohamad, R., Amini, Y., and Noshad, M. (2023). Kartalol: Transfer learning using deep neural network for iris segmentation and localization: New dataset for iris segmentation. *Iran Journal of Computer Science*, 6:1–13.

Pourafkham, B. and Khotanlou, H. (2023). Es-net: Unet-based model for the semantic segmentation of iris.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham. Springer International Publishing.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4510–4520.

Singh, G., Singh, R., Saha, R., and Agarwal, N. (2020). Iwt based iris recognition for image authentication. *Procedia Computer Science*, 171:1868–1876.

Tahir, A. A. K. and Angheluș, S. (2019). Human biometrics and biometric recognition systems; an overview. In *The 19th International Multidisciplinary Conference "Professor Dorin Pavel - the founder of Romanian hydropower"*, volume 35, pages 431–446.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. volume 30, pages 5998–6008. Curran Associates Inc.

Wildes, R. (1997). Iris recognition: an emerging biometric technology. *Proceedings of the IEEE*, 85(9):1348–1363.

Yang, K., Xu, Z., and Fei, J. (2021). Dualsanet: Dual spatial attention network for iris recognition. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 888–896.