IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 6 June 2024, accepted 27 June 2024, date of publication 3 July 2024, date of current version 15 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3422807

RESEARCH ARTICLE

Real-Time Multi-Spectral Iris Extraction in Diversified Eye Images Utilizing Convolutional Neural Networks

RASANJALEE RATHNAYAKE[1], NIMANTHA MADHUSHAN[2], ASHMINI JEEVA[2], DHANUSHIKA DARSHANI[2], IMESH PATHIRANA[3], SOURIN GHOSH[4], AKILA SUBASINGHE[5], (Member, IEEE), BHAGYA NATHALI SILVA[6,7], (Member, IEEE), AND UDAYA WIJENAYAKE[1], (Member, IEEE)

[1]Department of Computer Engineering, University of Sri Jayewardenepura, Nugegoda 10250, Sri Lanka
[2]Department of Electrical and Electronic Engineering, University of Sri Jayewardenepura, Nugegoda 10250, Sri Lanka
[3]School of Computing, Informatics Institute of Technology, Colombo 00400, Sri Lanka
[4]Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India
[5]Department of Computer Science, University of Birmingham, Dubai, United Arab Emirates
[6]Department of Information Technology, Faculty of Computing, Sri Lanka Institute of Information Technology, Malabe 10115, Sri Lanka
[7]Center for Excellence in Intelligent Informatics, Electronics and Transmission (CIET), Sri Lanka Institute of Information Technology, Malabe 10115, Sri Lanka

Corresponding authors: Udaya Wijenayake (udayaw@sjp.ac.lk) and Bhagya Nathali Silva (nathali.s@sliit.lk)

This work was supported by the Science and Technology Human Resource Development Project, Ministry of Education, Sri Lanka, funded by Asian Development Bank under Grant STHRD/CRG/R1/SJ/06.

ABSTRACT Iris extraction has gained prominence due to its application versatility across many domains. However, achieving real-time iris extraction poses challenges due to several factors. Learning-based algorithms outperform non-learning-based iris extraction methods, delivering superior accuracy and performance. In response, this article proposes a Convolutional Neural Networks (CNN)-based, accurate direct iris extraction mechanism for a broad spectrum of eye images. The innovation of our approach lies in its proficiency with varied image types, including those where the iris is partially obscured by the eyelid. We enhance the method's reliability by introducing a modified Circular Hough Transform (CHT). Extensive testing demonstrates our method's excellent real-time performance across diverse image types, even under challenging conditions. These findings underscore the proposed method's potential as a cost-effective and computationally efficient solution for real-time iris extraction in varied application domains.

INDEX TERMS Convolutional neural networks, circular Hough transformation, Iris extraction, Iris recognition, human-computer-interaction.

I. INTRODUCTION

With recent advancements, iris localization (IL) and iris extraction are widely used in numerous emerging applications in various domains. IL aims to track the iris and its boundaries, and iris extraction is tracking the outer boundary of the iris. Cataract disease identification is one of the key applications in disease diagnosis with the help of iris extraction [1], [2]. Precise iris extraction is essential for such applications since the iris texture should be analyzed carefully to identify the disease.

Advancements in artificial intelligence, mathematical modeling, and electronic devices have upgraded most of the applications in different domains [4]. For example, recent improvements in iris detection, localization, segmentation, and tracking technologies, empowered identity

The associate editor coordinating the review of this manuscript and approving it for publication was Zhe Jin.

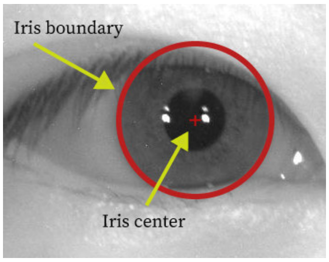


FIGURE 1. Objective of the study: tracking iris boundary and iris center [3].

authentication systems [5] to rely on iris-based person identification. Hence, systems that require strict security measurements have made a paradigm shift in biometric identification from fingerprint and facial identification to iris identification [6], [7]. In fact, iris identification fills the gap where other biometric identification mechanisms can go wrong due to various reasons, i.e. surgeries, cosmetics, etc.

Iris extraction is a crucial initial step in the gaze-tracking process. Identifying the outer boundary of the iris facilitates the extraction of inner details, such as the pupil, making the method more efficient by avoiding extraneous features like eyebrows and hair. This focused approach simplifies the analysis compared to evaluating the entire eye. Eye gaze tracking has diverse applications, including assessing the efficacy of teaching programs [8], [9], evaluating a person's physical state, and monitoring drivers' gaze [1], [2]. Recently, iris extraction has been utilized in diagnosing Autism Spectrum Disorder (ASD) [10], [11], as it aids in evaluating gaze shifting, a secondary parameter of attention tracking. Furthermore, tracking attention through gaze can assist in the early detection of Parkinson's disease [12], [13], allowing for more effective treatments [14], [15].

In addition to applications in the medical domain, gaze detection is widely used in transportation systems to track the gaze direction of drivers and pilots. This aims to ensure passenger safety as well as system safety by detecting the disturbed gaze or drowsiness of drivers and pilots [16], [17], [18], [19]. In the military, gaze detection assists pilots in targeting the exact location of certain military activities. Prominent works stated above confirm that iris extraction is a versatile notion that can be applied in various application domains. However, there is still a compelling demand for low-cost iris extraction schemes in multiple application facets.

The recent COVID-19 pandemic has highlighted the potential of gaze-controlled Automated Teller Machines (ATMs) as a novel application area for iris extraction technology [20]. However, this application domain presents several practical challenges. Iris extraction is particularly challenging in visible spectrum (VS) images captured in uncontrolled environments, where varying illumination conditions, fallen hair strips on eyes, eyeglasses, eye makeup, and partially opened eyes can all negatively impact the quality of the eye image [16], [21], [22].

The type of camera used in an eye-tracking system significantly affects the quality of the eye images and, consequently, the effectiveness of the iris extraction process. Near-infrared (NIR) cameras enhance image quality, but their higher cost increases the overall system expense, limiting their applicability in various domains. Despite this, there is a pressing need for low-cost gaze-tracking devices with high accuracy, especially given the large number of systems required for specific applications. While low-cost webcams produce lower-quality eye images, they can substantially reduce the total system cost. Therefore, utilizing VS webcams in eye-tracking systems can be beneficial for many applications by keeping costs low while meeting the demand for affordability.

As mentioned, extracting the iris from low-quality eye images captured in the visible spectrum (VS) domain is highly challenging. Despite this, there is a need for a hybrid iris extraction method that can handle both low-quality VS eye images and NIR eye images simultaneously. Considering the factors mentioned above, we plan to implement an iris extraction technique that can process real-time NIR and VS eye images. This study intends to extract the iris region from multi-spectral eye images, effectively supporting a wide variety of applications. The proposed iris extraction method results in identifying the iris's outer boundary and center, as illustrated in Figure 1. To facilitate iris extraction in difficult environments, our proposed work aims to achieve real-time iris extraction under varying image resolutions and camera types.

The key contributions of the proposed system are listed below:

- Provides accurate iris extraction that works with both NIR and VS images.
- Accurately extracts the iris in low-quality webcam images.
- Utilizes an updated Circular Hough Transform (CHT) method for iris boundary and center tracking to ensure system reliability.
- Offers an exhaustive solution that can be adopted by a variety of applications, including iris biometrics, gaze tracking, and medical applications.
- Introduces a visible spectrum dataset with nine subjects in a less controlled environment and a challenging webcam database with one subject.

The remainder of the paper is structured as follows: Section II provides a comprehensive discussion of the trending IL algorithms. Section III details the methodology, while Section IV presents the results and critical discussion. Finally, Section V concludes the paper and outlines future research directions.

II. OVERVIEW OF CURRENT IL ALGORITHMS

Current IL algorithms can be divided into two primary categories: learning-based schemes and non-learning-based schemes. Non-learning-based methods employ various image processing techniques, such as morphological operations and geometric features of the eye [21], [23]. These approaches are also less expensive since they do not require training or annotating large amounts of data. Despite these advantages, learning-based approaches have gained significant popularity due to their high accuracy and robustness in low-light and low-resolution conditions. The following section provides an in-depth discussion of learning-based IL techniques.

Convolutional Neural Networks (CNNs) represent a significant breakthrough in artificial intelligence, revolutionizing fields such as natural language processing and computer vision. CNNs are capable of identifying patterns, features, and even abstract concepts within large datasets with unprecedented accuracy and efficiency. This capability is due to their complex layers of interconnected neurons, which mimic the visual processing of the human brain. As a result, CNNs have paved the way for revolutionary advances in science and technology [24].

CNNs can efficiently and accurately segment the iris region; however, real-time iris localization in an unconstrained environment is challenging due to various factors, such as lighting conditions, pose variations, the distance between the camera and the user, and eye makeup, including thick eyelashes and mascara. To address these challenges, Muhammad et al. [25] proposed IrisGuideNet, which employs CNNs for unconstrained iris biometrics. IrisGuideNet features an encoder-decoder architecture with six stages in both the encoder and decoder. Among the three proposed networks, IrisGuideNet-Seg outputs the segmentation mask of the iris.

Mask-RCNN [26] is a deep-learning approach that can also be used in iris localization. The method proposed by Ahmad and Fuller [27] utilized Mask-RCNN to localize the iris within a rectangular bounding box and subsequently obtain the segmented iris mask. Although using a rectangular box for a circular object is not ideal, and Mask-RCNN is slower compared to other models, this method has achieved satisfactory results across various databases. Furthermore, Wang et al. [28] proposed a comprehensive iris segmentation approach utilizing a deep multitask attention network. This network includes both the iris mask and parameterized inner and outer iris boundaries, achieved jointly using a unified multitask network. This approach differs from conventional CNN-based iris segmentation methods that predict the iris masks by following semantic segmentation frameworks.

For automated iris segmentation in biometric identification, Sardar et al. [29] proposed a variant of UNet [30] with Squeeze and Expand modules for iris segmentation. This approach lowers the training time and improves storage efficiency by reducing the number of parameters used in the architecture. To address the issue of insufficient annotated samples, the study introduces an interactive component that automates the generation of ground truths.

Most studies focus on extracting the iris as a segmentation output and then applying circle fitting to determine the center and boundary of the iris. However, this approach requires high-resolution encoder-decoder networks, making the process expensive and susceptible to noise. To address these issues, the iris localization network (ILN) proposed in [31] returns the iris and pupil along with eight eyelid points. This method uses eyelid points, iris, and pupil circles (the iris inner boundary) for annotation, omitting the conventional iris segmentation task. Additionally, the authors proposed a Pupil Refinement Network (PRN) to enhance the accuracy of pupil localization.

According to the literature, many studies have proposed various methods to localize and segment the iris from eye images. However, extracting the iris from low-resolution eye images in the visible spectrum (VS) domain remains a challenge. To reduce the overall system cost, low-quality webcams can be used, but these result in low-quality eye images, leaving room for accuracy improvement. Despite these challenges, low-cost iris localization schemes can be beneficial for a wide range of applications. The next section discusses the proposed low-cost iris extraction strategy, which aims to extract the iris from challenging eye images in both the VS and NIR domains.

III. METHODOLOGY

This section outlines the detailed methodology used for developing and evaluating an iris extraction model that can process images from both VS and NIR imaging modalities. The methodology is divided into several key phases: data collection, sample selection, and the data training process, each essential for building a robust model.

A. DATA COLLECTION

The primary challenge addressed in this work is the development of a versatile iris extraction system that performs consistently across images captured in different spectral ranges. To achieve this, the model was trained on a diverse set of datasets collected under various environmental conditions using both NIR and VS cameras. This approach is designed to ensure the model's robust performance, regardless of the imaging spectrum or camera type used.

The NIR dataset employed in this study is the publicly accessible CASIA distance dataset [3], renowned for its comprehensive collection of NIR eye images. For the VS imagery, a novel dataset was generated using a Logitech C930E webcam in a relatively uncontrolled environment, comprising 105 instances of head movements across nine subjects. The Extended Yale Face Database B [32] is also utilized as an auxiliary source of VS images, further enriching the model's training material.

Additionally, a custom dataset captured with a low-quality laptop webcam featuring a single subject, along with the Chinese University of Hong Kong (CUHK) dataset [33], was employed exclusively for model accuracy testing.

The data collection for both self-made datasets was a non-invasive, non-destructive, and non-harmful procedure that captured participants' face images via webcam. All participants provided their informed consent for participation

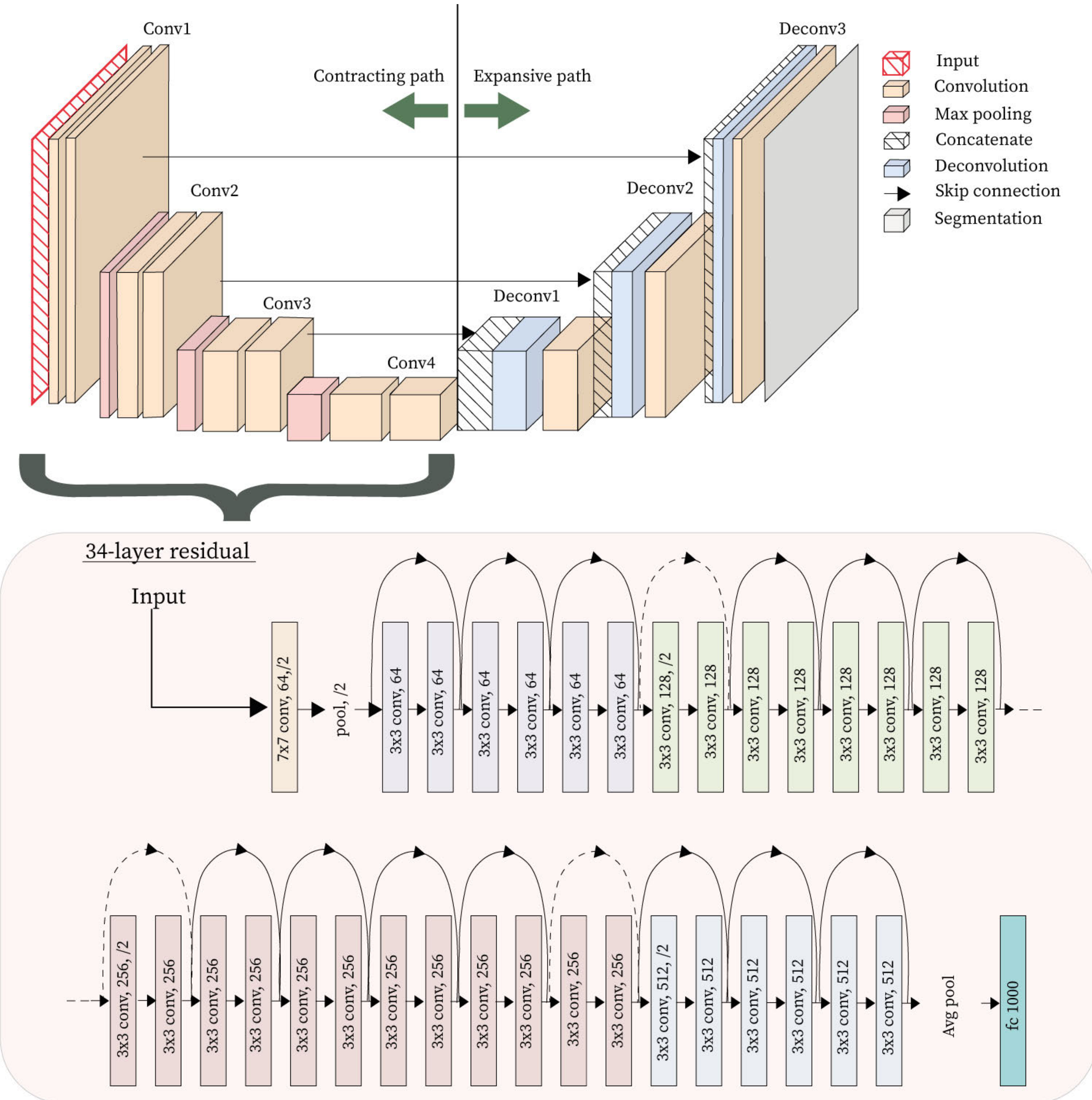


FIGURE 2. Machine learning model used in the proposed study: UNet architecture with the backbone of ResNet34.

in the face data collection procedure and the sharing of details of eye images.

B. SAMPLE SELECTION

The selection of samples is crucial in our study to ensure they accurately represent the target population. Among various sampling methods, stratified sampling was employed to minimize bias. This approach is particularly effective given that each database categorizes face images by subject, making stratified sampling the most suitable technique to ensure representative sampling across the population. For each database, we selected one sample, with Table 1 detailing the database names and the respective sample sizes utilized for training the CNN model.

For model validation, a distinct dataset was assembled, comprising the CUHK database [33], a general laptop webcam dataset, and the three primary training datasets. The CUHK database was subjected to stratified sampling to maintain consistency in methodology. Conversely, the

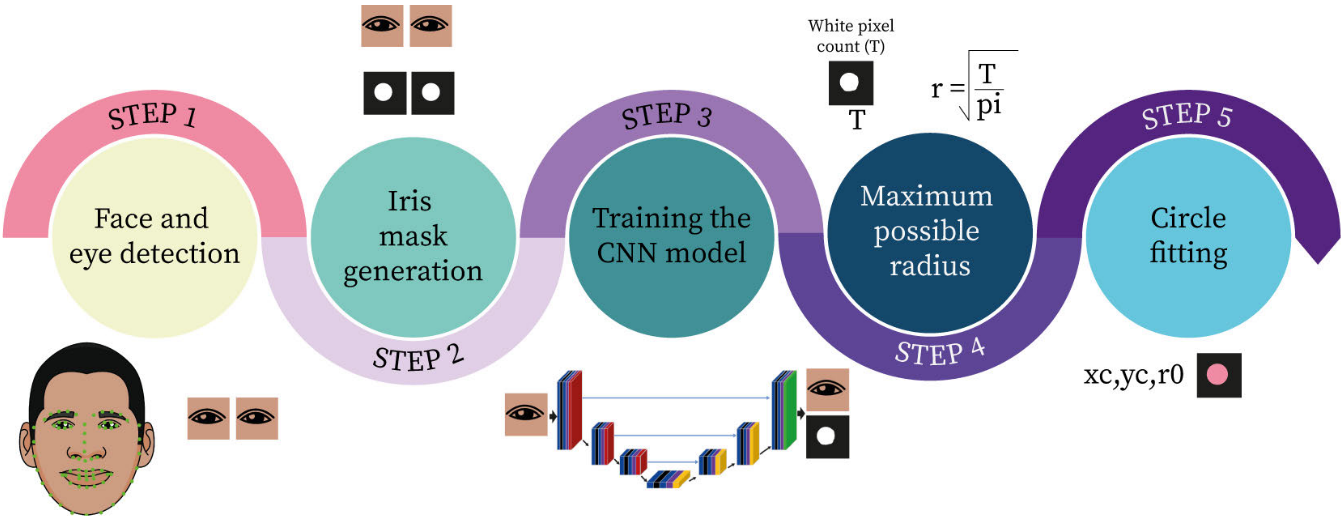


FIGURE 3. Detailed flowchart with five key steps of the iris extraction process in the proposed study.

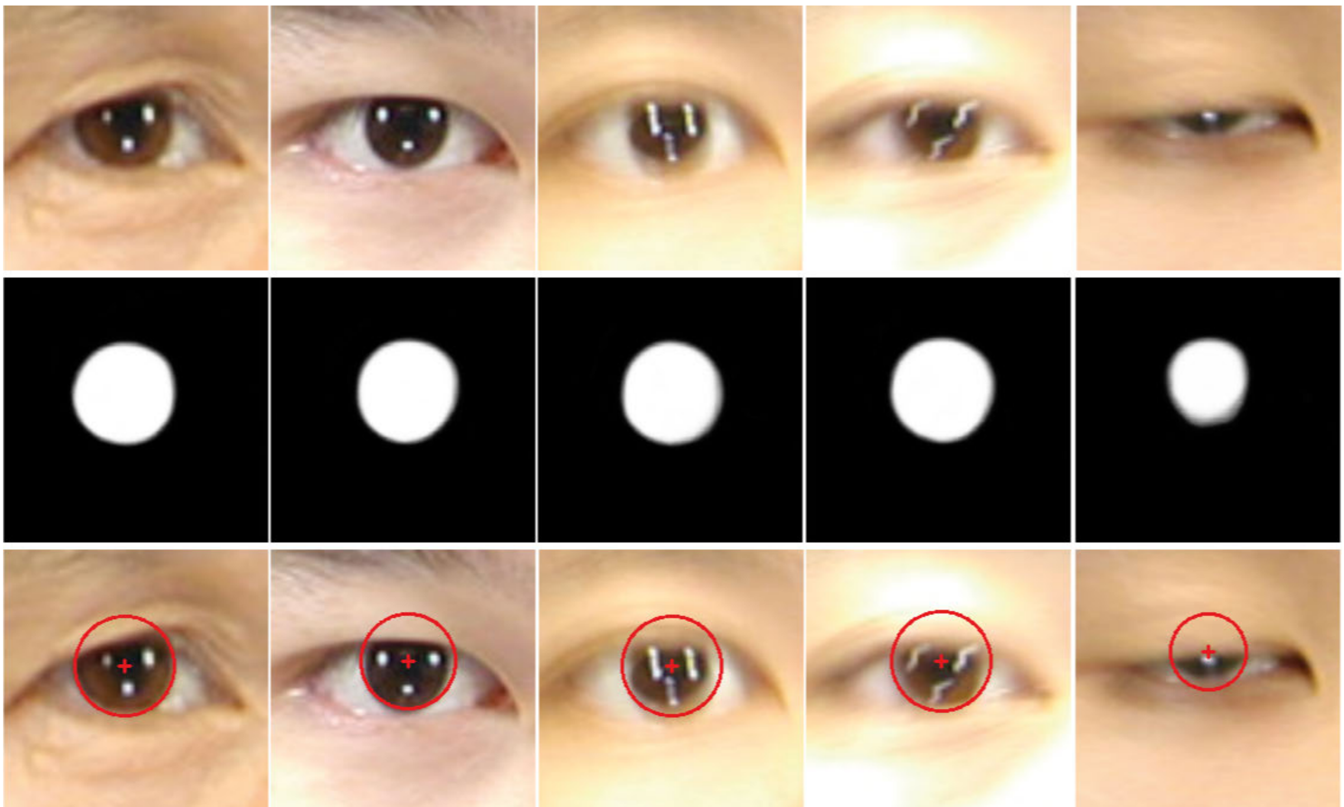


FIGURE 4. Original eye images (first row), predicted iris mask (second row), iris extraction using modified CHT (last row) [33].

laptop webcam dataset, which encompasses images from a single subject, was sampled using a simple random sampling approach, reflecting its unique collection context.

C. DATA TRAINING PROCESS

The comprehensive training process for the dataset is intricately designed to refine the input data and enhance the CNN model's performance. It unfolds through a sequence of critical steps, including face and eye detection, iris segmentation, and training the CNN model.

1) FACE AND EYE DETECTION

Face tracking serves as the foundational step for isolating the eye region from facial images. To enhance image quality,

TABLE 1. Summary of databases and sample sizes utilized for CNN model training.

Database name	Number of eye images in sample	NIR/ VS
CASIA Distance dataset	492	NIR
Custom made dataset with Logitech C930E webcam	813	VS
Extended YALE face database B	512	VS

each face image underwent Contrast Limited Adaptive Histogram Equalization (CLAHE) and gamma correction techniques. Subsequently, the refined images were employed to identify faces utilizing the Dlib 68 facial landmarks detector, achieving a remarkable accuracy rate of 99.97% across diverse databases.

Following face detection, the output was utilized to pinpoint the eye region. The Dlib 68 facial landmarks detector, specifying landmarks 36 to 42 for the left eye and 43 to 49 for the right eye, facilitated the tracking and extraction of the eye from the image. This method of eye tracking demonstrated an accuracy of 98.03% when tested on various databases.

2) IRIS MASK GENERATION

In the next phase of the experiment, square eye images with a size of 224 × 224 were employed to facilitate the generation process of iris masks using an annotation tool. With the use of this tool, the iris area was defined as a circular white mask on a 224 × 224 black mask using the brush tool. Following the segmentation process, the final iris masks were transformed into binary images, which was an important preliminary step before the data training procedure started.

3) TRAINING THE CNN MODEL

The main goal of this study is to create a model that can adapt to changes in the environment and work well with different types of cameras while also being very accurate in finding the iris in images. To achieve this, the model was trained using three different sets of images: the Logitech C930E webcam database (for VS images), the Extended YALE Face Database B (also for VS images), and the CASIA distance database (for NIR images). These databases were chosen to make sure the model can handle various lighting conditions and camera qualities.

Since the U-Net architecture excels at semantic segmentation tasks, it was selected to provide a strong basis for the development of a flexible and precise model, which is needed to harness the adaptability and precision required for iris localization across various situations and camera types.

U-Net architecture [30] is a CNN designed for the semantic segmentation of images. It is a U-shaped structure that consists of a contracting path and an expansive path. While the contracting path involves convolutional and pooling operations to capture context and reduce spatial dimensions, the bottleneck layer captures high-level features. Further, U-Net includes skip connections between layers of the contracting and expansive paths, and it allows for the utilization of high-resolution features, which helps preserve spatial information.

Even though Convolutional Neural Networks (CNNs) are highly effective for processing image data, they can face challenges as the network depth increases. These challenges often relate to selecting the right parameters, which can lead to reduced model performance. However, the ResNet architecture [34] offers a solution to this issue with its skip connections feature. These connections help to bypass some layers in the network, allowing for easier training of deeper models without sacrificing performance. Additionally, ResNet helps to address the vanishing gradient problem [35], enabling the construction of networks with thousands of layers that perform better than shallower ones [34]. It is also important to note that using pre-trained models often yields better results than training from scratch. In this study, we utilized a pre-trained ResNet34 model as the underlying structure for our UNet architecture. See Figure 2 for a visual representation of this network architecture.

Even though we used the pre-trained ResNet34 model, hyperparameter tuning was done. The learning rate was set to the default, but the optimizer used was Adamax, as it surpassed the accuracy of the model more than using Adam or other optimizers. We keep the batch size low to ensure lower computational complexity.

Figure 3 illustrates the flowchart of the method proposed. Following the completion of the data training phase, a series of steps are undertaken to predict the center of the iris based on the mask generated by the model.

D. CIRCLE FITTING

After the model has been trained, it uses eye images to predict the iris mask, which appears as a circular shape even when part of the iris is obscured by the eyelid, as shown in Figure 4. The predicted iris mask is then utilized to identify the iris boundary and center using a modified Circular Hough Transform (CHT) method. Traditional CHT relies on several parameters, including the minimum and maximum radii, which are crucial for accurately locating circles within specified limits. However, setting a fixed maximum radius is impractical due to variations in iris size among different individuals. To overcome this, two strategies are proposed: one sets a high constant value for the maximum radius, such as 500 pixels, and the other uses a mathematical formula within the model to calculate the radius dynamically. For the purposes of this study, the latter approach was adopted to determine the maximum radius based on the mathematical framework.

The process begins with counting the total number of white pixels (T) in the predicted mask. This count is then used in (1) to compute the maximum possible radius for the iris mask. With this calculated maximum radius, the CHT algorithm is updated and executed, resulting in the determination of the

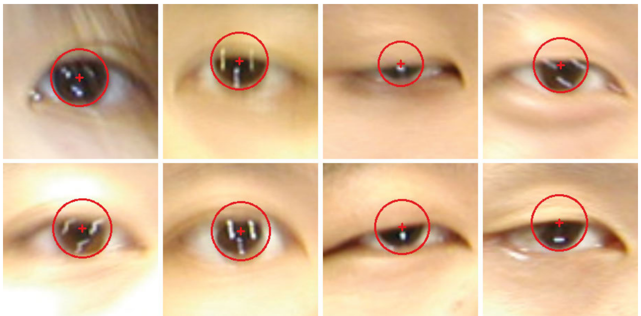


FIGURE 5. Accurate predictions by the model on the CUHK dataset.

FIGURE 6. Accurate predictions by the model on the low-quality laptop webcam image sample.

x and y coordinates as well as the radius of the predicted circle, which correspond to the iris center and boundary, respectively. The key advantage of this method is its ability to adapt the maximum radius parameter dynamically to each input image, enhancing the precision of the iris detection process.

Table 2 showcases a comparative analysis of the average execution time between the two approaches for determining the maximum radius within the CHT method. While the updated method demonstrates a marginally increased execution time compared to the original approach, the average time discrepancy is notably minor, amounting to just 0.2165 milliseconds. Despite this slight delay, the updated method significantly enhances the accuracy and reliability of iris localization. Consequently, the minor increase in latency introduced by the new method is considered negligible, given its substantial contributions to the precision of the localization process.

$$T = pi \times r^2 \tag{1}$$

IV. ACCURACY METRIC

Pixel error was used as the accuracy metric to measure the accuracy of the model. Pixel error calculates the Euclidean distance between the actual iris and the predicted iris. Equation 2 shows the formulae for pixel error calculation. A successful iris extraction is recorded when the pixel error is less than 5 pixels.

$$Pixel\ error = \sqrt{(x_c - x_p)^2 + (y_c - y_p)^2} \tag{2}$$

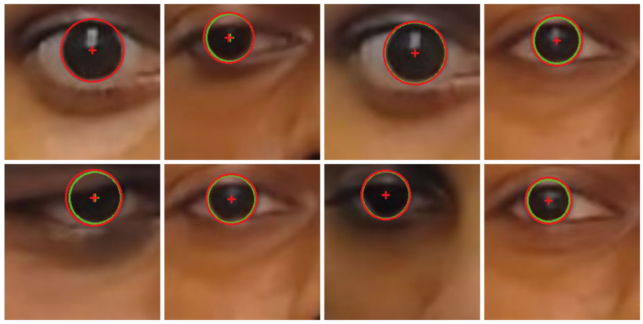


FIGURE 7. Accurate predictions given by the model on the self-made dataset (Logitech C930E webcam image dataset) green: ground truth; red: predicted.

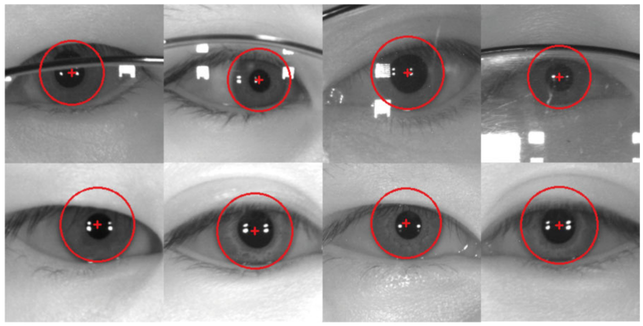


FIGURE 8. Accurate predictions given by the model on the CASIA distance dataset.

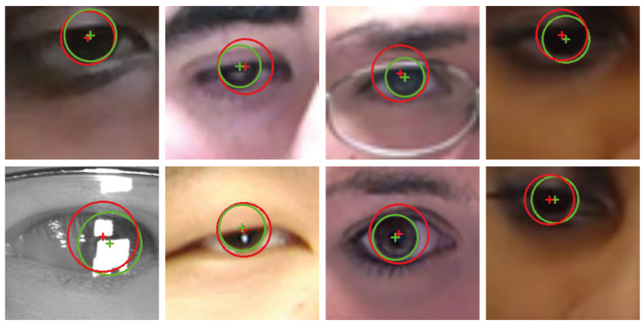


FIGURE 9. Erroneous results given by the model on different databases (green: ground truth, red: predicted).

where x_c, y_c are actual iris center coordinates extracted manually and x_p, y_p are predicted iris center coordinates.

V. RESULTS

To analyze and compare the results of different models, samples of challenging situations from different databases were selected. All the training datasets, including the CASIA distance database [3], Extended YALE face database B images [32], and the self-made dataset collected with a Logitech C930E webcam, were used for the accuracy testing. In addition to the CUHK database [33], a laptop webcam database created using one subject in a less controlled environment was used as a sample to validate the accuracy of the model.

TABLE 2. Average time taken by original and updated CHT algorithms.

	Original CHT	Updated CHT
Average time (milliseconds)	1.2791	1.4956
Standard deviation of the time (milliseconds)	0.6746	0.6864

TABLE 3. Model accuracy over different databases.

Database name	Accuracy (error < 5 pixels)
CASIA Distance dataset	97.76%
Extended Yale face database B	92.79%
Self-made dataset with LogitechC930E webcam	93.00%
CUHK dataset	87.86%
General laptop webcam database with one subject	82.81%

TABLE 4. Accuracy comparison among the different learning models proposed in different studies and our model.

Author	Database Type		GPU Requirement		Real-time Processing		Performance metric	Accuracy for each database
	NIR	VS	Yes	No	Yes	No		
Pasha et al. [36]	✓		✓		✓		Accuracy [20]	• CASIA Iris dataset - 99.89%
Ma et al. [23]	✓			✓	✓		Pixel error [20]	• Self-made dataset - 95.24% • MMUv1 - 98.18% • IITDv1 - 98.30% • CASIA V3 Interval - 98.56%
Muhammad et al. [25]	✓	✓	✓			✓	E1 and E2 error rates	For all the 8 datasets Iris GuideNet-Seg • Average E1 - 0.3955 • Average E2 - 0.1978 Iris GuideNet-Join • Average E1 - 0.4064 • Average E2 - 0.2032
Toizumi et al. [31]	✓	✓	Tested with both CPU ad GPU		✓		L2 loss [20]	• CAISIA V4 thousand - 0.0044 • CASIA V4 distance - 0.0107 • IITD - 0.005 • MMUv1 - 0.0064 • MMUv2 - 0.0069 • CASIA V4 twins - 0.0085
Sardar et al. [29]	✓	✓	✓			✓	• Mean Error Rate (MER) • Dice similarity coefficient (DSC) • Mean True Positive Rate (mTPR)	• CASIA-Irisv4-Interval (MER - 0.176, DSC - 0.987, mTPR - 0.972) • IITD (MER - 0.249, DSC - 0.985, mTPR - 0.980) • NICE.I (MER - 0.261, mTPR - 0.983)
Proposed study	✓	✓	Tested with both CPU ad GPU		✓		Pixel error [20]	• Casia Distance - 97.76% • Extended YALE face database B - 92.79% • Self-made dataset - 93.00% • CUHK dataset - 87.86% • General laptop webcam database - 82.81%

A. OBTAINING THE GROUND TRUTH

To determine the accuracy, the actual iris center (ground truth) must first be calculated. This was achieved by extracting the actual iris centers using manually created iris masks.

B. RESULTS OVER DIFFERENT DATABASES

This section focuses on the accuracy of different databases to validate the model's performance. The CUHK dataset contains challenging images due to varying illumination conditions, blurred images caused by eye saccades, off-axis eyes, and partially opened eyes. Although the database lacks ethnic diversity, it includes challenging eye images from individuals of different age groups. Some of these challenging eye images and the model outcomes are shown in Figure 5.

Using a low-quality webcam is beneficial for reducing costs in many applications involving iris localization. Typically, laptop webcams cannot capture high-quality images. To test the model's accuracy with low-quality webcam images, we used a dataset consisting of images from a general laptop webcam for one subject. The predicted iris localizations are shown in Figure 6.

Further, an untrained sample of images from the self-made dataset was also used to test the model's accuracy. It was done to analyze the behavior of the model with unseen data, like in real-time scenarios. Figure 7 shows the predicted iris boundary and center in red and ground truths in green on the same image. Moreover, model accuracy was tested using NIR images obtained from the CASIA Distance dataset. Figure 8 shows the correct predictions given. Table 3 summarizes the accuracy over different databases for the proposed method.

The gaze point, where a person is looking, depends on the center of the pupil [14]. As mentioned, for the visible spectrum dark eye images, it is challenging to track the center of the pupil, as the pupil and iris can be barely identified in low-quality VS eye images. In this study, we analyzed the error of treating pupil and iris as concentric. The error was an average of 3 pixels with a 1.47-pixel standard deviation. In applications where pinpoint accuracy is not required, such as gaze-based graphical user interfaces, this assumption is acceptable even if it increases gaze error.

C. PERFORMANCE COMPARISON

The proposed model's accuracy was compared with other contemporary SOTA methods, and the results are presented in Table 4. The proposed methods by Ma et al. [23] and Pasha et al. [36] deal only with NIR images, even though those methods achieve real-time performance. Further, Toizumi et al. [31] ensured real-time performance without using a GPU; the system works fine with both NIR and VS images. However, the result could deviate when the iris region is highly irregular because the proposed approach works with bounding boxes. Even though the methods suggested by Muhammad et al. [25] and Sardar et al. [29] work fine with both NIR and VS images, they do not ensure real-time performance. Refer to Table 4 for a summary of the SOTA methods. Even though some models require demanding resources to execute, the proposed model runs smoothly on a CPU in real time. Further, along with the outer boundary of the iris, the proposed method extracts the center of the iris accurately. The extracted iris center can be used in specific gaze tracking applications, as discussed in the latter part of the results section, and therefore, this proposed study can be used to extend the number of applications in the domain of iris extraction.

The size of the images selected for the training affects the time complexity, and in this study, the input image size selected was 224 × 224. Space complexity depends on the total number of parameters used in the architecture, batch size, and size of the training dataset. Even though the larger batch size improves the model's performance, it increases the computational cost [37]. Therefore, the batch size was set to a lower number, which is 20. However, the total number of parameters used is 24 million. Further, the total number of images used for the training and testing was 1817, which is in the same range as [31], and the average frames per second in the proposed study is 10 fps.

Even though the model achieved high accuracy over different illumination and environmental conditions, in some cases, it deviates from the ground truth. Figure 9 shows some examples of erroneous predictions over different databases under highly challenging conditions.

VI. CONCLUSION

This article proposes a learning-based iris extraction model with UNet architecture and circle fitting on the segmented iris to extract the iris boundary and the center. The breakthrough of the proposed algorithm is its ability to run in real-time regardless of the types of images, i.e., NIR or VS images. Further, the algorithm possesses the ability to extract the iris area completely, even if the eyes are partially opened or hidden underneath the eyelashes. According to the performance analysis, the model offers a higher level of accuracy over challenging databases. The time consumption was negligibly increased due to the use of the updated CHT algorithm for maximum radius calculation. However, this updated calculation confirms the dynamic adjustment of the maximum radius for iris localization instead of relying on an arbitrary static value. Thus, it confirms the proposed model is bound by a mathematical framework and efficiently works in real-time with various image types in challenging environments without using a GPU. These characteristics will make this a proper fit for many applications that require real-time, accurate, and cost-efficient iris localization solutions.

The potential for improving communication for people with disabilities through the development of an eye-gaze-controlled assistive communication system is evident. Our future research will focus on enhancing and advancing this technology to overcome current constraints, particularly by exploring alternatives to head-mounted eye trackers. Developing a more effective and accessible communication tool that enables people with disabilities to express themselves more naturally and independently through emerging technologies and interdisciplinary collaboration is a promising avenue to explore.

Furthermore, extracting the outer boundary of the iris can aid in identifying the pupil in both VS and NIR challenging images. This study will benefit many domains, as it will expand the range of low-cost applications that require the pupil center as an input.

REFERENCES

[1] Q. Zhuang, Z. Kehua, J. Wang, and Q. Chen, "Driver fatigue detection method based on eye states with pupil and iris segmentation," *IEEE Access*, vol. 8, pp. 173440–173449, 2020.

[2] C.-H. Weng, Y.-H. Lai, and S.-H. Lai, "Driver drowsiness detection via a hierarchical temporal deep belief network," in *Computer Vision—ACCV 2016 Workshops* (Lecture Notes in Computer Science), C.-S. Chen, J. Lu, and K.-K. Ma, Eds. Cham, Switzerland: Springer, 2016, pp. 117–133.

[3] *Center for Biometrics and Security Research*. Accessed: Feb. 20, 2024. [Online]. Available: http://www.cbsr.ia.ac.cn/english/IrisDatabase.asp

[4] X. Zhu, K. Guo, T. Qiu, H. Fang, Z. Wu, X. Tan, and C. Liu, "Stereoscopic image super-resolution with interactive memory learning," *Expert Syst. Appl.*, vol. 227, Oct. 2023, Art. no. 120143. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417423006450

[5] K. W. Bowyer, K. Hollingsworth, and P. J. Flynn, "Image understanding for iris biometrics: A survey," *Comput. Vis. Image Understand.*, vol. 110, no. 2, pp. 281–307, May 2008, doi: 10.1016/j.cviu.2007.08.005.

[6] Q. Guo and J. Zheng, "An iris recognition algorithm for identity authentication," in *Proc. Int. Conf. Intell. Transp., Big Data Smart City (ICITBS)*, Jan. 2018, pp. 621–624. [Online]. Available: https://ieeexplore.ieee.org/document/8332846

[7] C.-W. Chuang and C.-P. Fan, "Biometric authentication with combined iris and sclera information by YOLO-based deep-learning network," in *Proc. IEEE Int. Conf. Consum. Electron. Taiwan (ICCE-Taiwan)*, Sep. 2020, pp. 1–2. [Online]. Available: https://ieeexplore.ieee.org/document/9258253

[8] A. Susac, A. Bubic, M. Planinic, M. Movre, and M. Palmovic, "Role of diagrams in problem solving: An evaluation of eye-tracking parameters as a measure of visual attention," *Phys. Rev. Phys. Educ. Res.*, vol. 15, Jan. 2019, Art. no. 013101. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.15.013101

[9] Y. Wang, S. Lu, and D. Harter, "Multi-sensor eye-tracking systems and tools for capturing Student attention and understanding engagement in learning: A review," *IEEE Sensors J.*, vol. 21, no. 20, pp. 22402–22413, Oct. 2021.

[10] V. Yaneva, L. A. Ha, S. Eraslan, Y. Yesilada, and R. Mitkov, "Detecting high-functioning autism in adults using eye tracking and machine learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 6, pp. 1254–1261, Jun. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9082703/

[11] C. E. Venker, R. Pomper, T. Mahr, J. Edwards, J. Saffran, and S. Ellis Weismer, "Comparing automatic eye tracking and manual gaze coding methods in young children with autism spectrum disorder," *Autism Res., Off. J. Int. Soc. Autism Res.*, vol. 13, no. 2, pp. 271–283, Feb. 2020.

[12] C. Lu, P. Chakravarthula, K. Liu, X. Liu, S. Li, and H. Fuchs, "Neural 3D gaze: 3D pupil localization and gaze tracking based on anatomical eye model and neural refraction correction," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Singapore, Oct. 2022, pp. 375–383. [Online]. Available: https://ieeexplore.ieee.org/document/9995692/

[13] T. Tabashum, A. Zaffer, R. Yousefzai, K. Colletta, M. B. Jost, Y. Park, J. Chawla, B. Gaynes, M. V. Albert, and T. Xiao, "Detection of Parkinson's disease through automated pupil tracking of the post-illumination pupillary response," *Frontiers Med.*, vol. 8, Mar. 2021, Art. no. 645293. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8026862/

[14] B. Massé, S. Ba, and R. Horaud, "Tracking gaze and visual focus of attention of people involved in social interaction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 11, pp. 2711–2724, Nov. 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8194910/

[15] M. Puurtinen, U. Hoppu, S. Puputti, S. Mattila, and M. Sandell, "Investigating visual attention toward foods in a salad buffet with mobile eye tracking," *Food Quality Preference*, vol. 93, Oct. 2021, Art. no. 104290. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950329321001737

[16] L. Shi, C. Wang, F. Tian, and H. Jia, "An integrated neural network model for pupil detection and tracking," *Soft Comput.*, vol. 25, no. 15, pp. 10117–10127, Aug. 2021, doi: 10.1007/s00500-021-05984-y.

[17] E. Q. Wu, G.-R. Zhou, L.-M. Zhu, C.-F. Wei, H. Ren, and R. S. F. Sheng, "Rotated sphere Haar wavelet and deep contractive auto-encoder network with fuzzy Gaussian SVM for pilot's pupil center detection," *IEEE Trans. Cybern.*, vol. 51, no. 1, pp. 332–345, Jan. 2021.

[18] P. Kanade, F. David, and S. Kanade, "Convolutional neural networks(CNN) based eye-gaze tracking system using machine learning algorithm," *Eur. J. Electr. Eng. Comput. Sci.*, vol. 5, no. 2, pp. 36–40, Apr. 2021. [Online]. Available: https://ejece.org/index.php/ejece/article/view/314

[19] D. Kang and H. S. Chang, "Low-complexity pupil tracking for sunglasses-wearing faces for glasses-free 3D HUDs," *Appl. Sci.*, vol. 11, no. 10, p. 4366, May 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/10/4366

[20] R. Rathnayake, N. Madhushan, A. Jeeva, D. Darshani, A. Subasinghe, B. N. Silva, L. P. Wijesinghe, and U. Wijenayake, "Current trends in human pupil localization: A review," *IEEE Access*, vol. 11, pp. 115836–115853, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10286541

[21] K. Donuk and D. Hanbay, "Pupil center localization based on mini U-Net," in *Proc. Comput. Sci., Volume, Int. Artif. Intell. Data Process. Symp. (IDAP)*, 2022, pp. 185–191.

[22] F. Jan and N. Min-Allah, "An effective iris segmentation scheme for noisy images," *Biocybern. Biomed. Eng.*, vol. 40, no. 3, pp. 1064–1080, Jul. 2020.

[23] L. Ma, H. Li, and K. Yu, "Fast iris localization algorithm on noisy images based on conformal geometric algebra," *Digit. Signal Process.*, vol. 100, May 2020, Art. no. 102682. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1051200420300270

[24] X. Zhu, K. Guo, S. Ren, B. Hu, M. Hu, and H. Fang, "Lightweight image super-resolution with expectation-maximization attention mechanism," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1273–1284, Mar. 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9427111

[25] J. Muhammad, C. Wang, Y. Wang, K. Zhang, and Z. Sun, "Iris-GuideNet: Guided localization and segmentation network for unconstrained iris biometrics," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 2723–2736, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10105641/

[26] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," Jan. 2018, *arXiv:1703.06870*.

[27] S. Ahmad and B. Fuller, "Unconstrained iris segmentation using convolutional neural networks," Dec. 2018, *arXiv:1812.08245*.

[28] C. Wang, J. Muhammad, Y. Wang, Z. He, and Z. Sun, "Towards complete and accurate iris segmentation using deep multi-task attention network for non-cooperative iris recognition," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2944–2959, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9036930

[29] M. Sardar, S. Banerjee, and S. Mitra, "Iris segmentation using interactive deep learning," *IEEE Access*, vol. 8, pp. 219322–219330, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9274419/

[30] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015* (Lecture Notes in Computer Science), N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham, Switzerland: Springer, 2015, pp. 234–241.

[31] T. Toizumi, K. Takahashi, and M. Tsukada, "Segmentation-free direct iris localization networks," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Waikoloa, HI, USA, Jan. 2023, pp. 991–1000. [Online]. Available: https://ieeexplore.ieee.org/document/10030850/

[32] *Yale Face Database*. Accessed: Mar. 12, 2024. [Online]. Available: http://cvc.cs.yale.edu/cvc/projects/yalefacesB/yalefacesB.html

[33] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 1955–1967, Nov. 2009. [Online]. Available: https://ieeexplore.ieee.org/document/4624272

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Dec. 2015, *arXiv:1512.03385*.

[35] H. Li, Z. Xu, G. Taylor, C. Studer, and T. Goldstein, "Visualizing the loss landscape of neural nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018. [Online]. Available: https://papers.nips.cc/paper_files/paper/2018/hash/a41b3bb3e6b050b6c9067c67f663b915-Abstract.html

[36] S. N. Pasha, D. Anitha, R. Anbunathan, C. Gobinath, and G. K. Madhura, "An efficient novel approach for iris recognition and segmentation based on the utilization of deep learning," in *Proc. Int. Conf. Evol. Algorithms Soft Comput. Techn. (EASCT)*, Bengaluru, India, Oct. 2023, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/10392672/

[37] B. Shah and H. Bhavsar, "Time complexity in deep learning models," *Proc. Comput. Sci.*, vol. 215, pp. 202–210, Jan. 2022. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1877050922020944

RASANJALEE RATHNAYAKE received the B.Sc. degree in electrical and electronic engineering from the South Eastern University of Sri Lanka, in 2020. She is currently pursuing the M.Phil. degree in image processing and machine learning with the University of Sri Jayewardenepura (USJ), Sri Lanka. She started her career as an Instructor with USJ, in 2020. She is a Research Assistant with the Department of Computer Engineering, Faculty of Engineering, USJ. Her research interests include image processing and pattern recognition, machine learning, gaze tracking, brain sensing, statistics, and their applications in human–computer interaction.

NIMANTHA MADHUSHAN received the B.Sc. degree in electrical and information engineering from the University of Ruhuna, Matara, Sri Lanka, in 2019. He is currently pursuing the M.Phil. degree in electrical and electronic engineering with the University of Sri Jayewardenepura (USJ), Ratmalana, Sri Lanka. He is a Research Assistant with the Department of Electrical and Electronic Engineering, Faculty of Engineering, USJ. His research interests include non-intrusive load monitoring and power system measurement.

ASHMINI JEEVA received the B.Sc. degree in electrical and electronic engineering from the South Eastern University of Sri Lanka, in 2021. She started her career as a Research and Development Engineer with Ceylon Agro Food Technologies (Pvt.) Ltd., in 2022. She is currently a Research Assistant with the Department of Electrical and Electronic Engineering, Faculty of Engineering, University of Sri Jayewardenepura (USJ). Her research interests include unmanned aerial vehicles, image processing, machine learning, and their applications in disaster management.

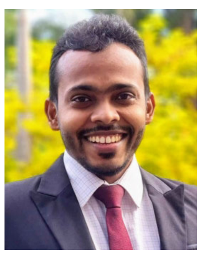
DHANUSHIKA DARSHANI received the B.Sc. (Eng.) degree (Hons.) in electrical and electronics engineering from the South Eastern University of Sri Lanka. She is currently a Research Assistant with the Department of Computer Engineering, Faculty of Engineering, University of Sri Jayewardenepura, Sri Lanka. Her research interests include the IoT, power electronics, artificial intelligence, and machine learning-related projects.

IMESH PATHIRANA received the Bachelor of Science degree (Hons.) in applied mathematics from the Department of Mathematics, University of Sri Jayewardenepura, Sri Lanka, in 2023. Currently, he is an Assistant Lecturer with the Informatics Institute of Technology, Sri Lanka. His areas of research interests include graph theory, image processing, computational discrete mathematics, operational research, and data analytics.

SOURIN GHOSH received the bachelor's degree in electronics and telecommunications engineering from the University of Pune, in 2012, and the M.Tech. degree specializing in signal processing from the Government College of Engineering, Pune, India, in 2017. Currently, he is pursuing the Ph.D. degree with the Department of Electrical Engineering, Indian Institute of Technology (IIT) Bombay. His professional journey began with Tata Consultancy Services Ltd., Pune, from 2012 to 2018. During this period, he played a pivotal role in the development of Advanced Driver Assistance Systems, leveraging his expertise in computer vision. Since 2018, he has been contributing as the Lead Robotics Engineer with John Deere India Pvt. Ltd., Pune. In this capacity, he is actively involved in advancing precision agriculture technologies aimed at empowering farmers to enhance food production.

AKILA SUBASINGHE (Member, IEEE) received the B.Sc. (Eng.) degree in electrical engineering from the University of Moratuwa, Sri Lanka, and the M.E.Sc. and Ph.D. degrees in radiation bio dosimetry involving image analysis and machine learning from Western University, Canada. He is currently an Assistant Professor with the Department of Computer Science, University of Birmingham, Dubai. His research interests include artificial intelligence (AI) (including deep learning (DL) and machine learning (ML)) and Image Analysis. Within this scope, he has been actively engaged in many research projects spanning the areas of medical image analysis, pulmonary disease diagnosis, biomedical engineering, drones, disaster management, and security surveillance.

BHAGYA NATHALI SILVA (Member, IEEE) received the B.S. and M.S. degrees in information technology from Sri Lanka Institute of Information Technology, Sri Lanka, in 2011 and 2015, respectively, and the Ph.D. degree from Kyungpook National University, South Korea, in 2018. She is currently an Assistant Professor with the Faculty of Computing, Sri Lanka Institute of Information Technology. Her research interests include the Internet of Things, data analytics, deep learning, optimization, and sustainable energy management. She serves as a reviewer for numerous reputed journals and conferences.



UDAYA WIJENAYAKE (Member, IEEE) received the B.Sc. degree in computer science from the University of Colombo, Sri Lanka, in 2010, and the M.Sc. degree in electronic engineering and computer science and the Ph.D. degree in computer science and engineering from Kyungpook National University, South Korea, in 2013 and 2018, respectively. He is currently a Senior Lecturer with the Department of Computer Engineering, University of Sri Jayewardenepura. His research interests include 3D reconstruction, image processing, machine learning, human–computer interaction, and 3D motion estimation.