Iris Segmentation in the Wild using Encoder-Decoder based Deep Learning Techniques

Shreshth Saini` * Divij Gupta` Ranjeet Ranjan Jha` † Gaurav Jaswal` ‡
and Aditya Nigam `

Abstract

Iris recognition is considered to be one of the most widely used biometric modality, mainly due to its non-invasive nature and high reliability. However, in the whole process of authentication, segmentation of the iris is the most crucial one as being the second stage of the usual five-stage pipeline, the error introduced gets compounded in the subsequent stages. However, segmentation of the iris in non-ideal conditions is a challenging task owing to numerous artifacts such as occlusion by eyelids, off-angle rotations, irregular reflections, blurred boundaries, etc. Although the artifacts can be minimized up to a certain extent during the acquisition process, it requires a high level of control over the image capturing environment and also high user cooperation, which is not always feasible. For segmentation, quite a few methods have been put forward, but the ones using classical approaches usually have low generalisability. Over the past decade, various deep learning techniques have been proposed which have given satisfactory results. Since the problem at hand is that of an image-to-image generation(the input image and its corresponding segmentation mask), the most common similarity among them is the use of a standard encoder-decoder structure called the UNet. In this chapter, we discuss several such techniques and their intricate novelties, and shortcomings, while also throwing some light on the non-deep learning methods so as to get a wholesome comparison. We also discuss briefly about the various publicly available datasets and the artifacts they are riddled with while also discussing about the various metrics that are used by the scientific community to compare their works and establish the state-of-the-art. Lastly, we discuss a short implementation of the UNet done by ourselves on two of the available datasets and conclude this chapter with a thought on the future possibilities of the existing works.

Keywords — Deep Learning, Segmentation, UNet, Biometric ,Iris

---

*Shreshth Saini and Divij Gupta (equal contribution) - Department of Electrical Engineering, Indian Institute of Technology Jodhpur, India e-mail: saini.2@iitj.ac.in.

†Ranjeet Ranjan Jha and Aditya Nigam - School of Computing and Electrical Engineering, Indian Institute of Technology Mandi, India.

‡Gaurav Jaswal - Department of Electrical Engineering, Indian Institute of Technology Delhi, India.

Table 1: Comparison of some commonly used biometrics on the criteria as in [54]

Biometric	Univer.	Uniq.	Perm.	Collect.	Perf.	Accept.	Circum.
Iris	⇑	⇑	⇑	⇕	⇑	⇓	⇑
Fingerprint	⇕	⇑	⇑	⇕	⇑	⇕	⇑
Face	⇑	⇓	⇕	⇑	⇓	⇑	⇓
Gait	⇕	⇓	⇓	⇑	⇓	⇑	⇕
Hand Geometry	⇕	⇕	⇕	⇑	⇕	⇕	⇕
Retinal Scan	⇑	⇑	⇕	⇓	⇑	⇓	⇑
Voice Print	⇕	⇓	⇓	⇕	⇓	⇑	⇓

1 Introduction

In today's world, where technology has invaded almost every walk of our life, it has become more important now than ever to safeguard confidential information. While passwords were predominantly used for the task, they have become a thing of the past as they are weak and can be cracked using simple to sophisticated techniques by accomplished hackers. In comparison, biometrics provide a much more reliable and challenging to crack alternative in comparison to passwords. Biometrics can be broadly categorized into the categories, physiological and behavioral. Physiological identification is based on traits like face, hand geometry, iris, etc. whereas behavioral identification is made on the basis of characteristics such as signature, gait, etc. According to [54], any human behavioral or physiological characteristic may be utilized as a biometric given that it satisfies some specific properties:-

1) Universality i.e., it should be possessed by everyone.
2) Uniqueness i.e., it should be distinct among people.
3) Permanence i.e., it should remain the same throughout time.
4) Collectability i.e., it should be easily collectible and quantifiable.
5) Performance i.e., it should be efficient in the identification of the subject.
6) Acceptability i.e., it should be acceptable to the people in general.
7) Circumvention i.e., it should not be easy for the system using it to be fooled.

In, [54], the authors have scored the modalities on the above criteria in majorly three categories - $Low(\Downarrow), Medium(\Updownarrow), and High(\Uparrow)$.

The survey has extended or represented differently in many studies such as [96, 37, 103, 53, 20]. Table 1 represents that although Iris doesn't score 'High' in all the criteria, it is still by far the most suitable biometric modality. Owing to its accuracy and reliability [28], it is used in different biometric applications such as forensics [86], intelligent unlocking [21] etc.

The human eye is such that it provides not one, but two important biometric traits namely the retina and the iris. The iris is an annular structure positioned between the eye parts sclera, that is an off-white color region, and the pupil, which is the dark region at nearly the centre of the entire eye. The boundary where the iris and pupil meet is called the pupillary boundary similarly, where the iris and the sclera meet is called the limbic boundary. The iris comprises of elastic connective tissues which enriches the region with diverse patterns comprising crypts, ridges, radial and contraction furrows, freckles, and arching ligaments [18]. Unlike the rest of the biometric characteristics such as face or fingerprint etc., the iris is guarded by the cornea and aqueous humor which accounts for its high permanence. The iris starts developing in initial three months of the incubation period through forming and folding of the tissue membranes [41]. Moreover, only the pigmentation of the eye is genetically determined while its intricate structures are independent of genetics, which leads to it having more than 200 distinct features. Owing to its highly complex and random structure, the iris allows for its use as an efficient trait for
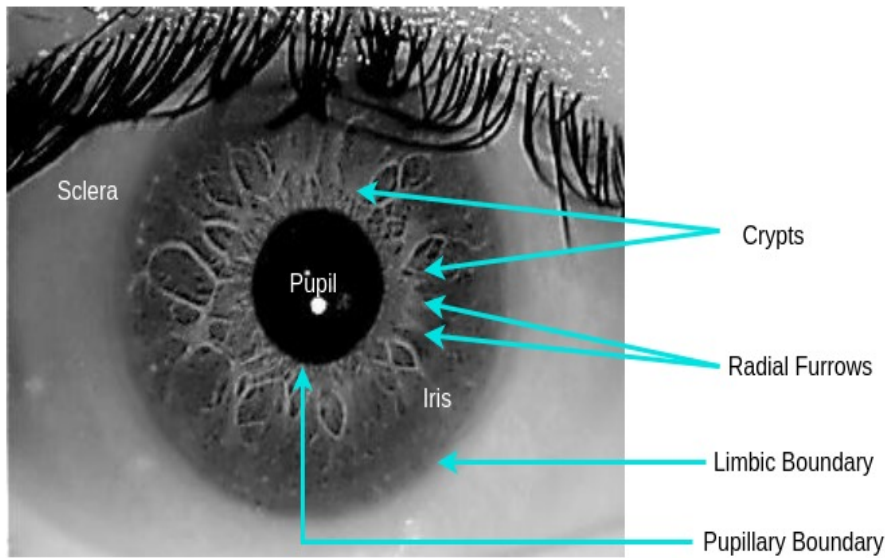
Figure 1: Anatomy of Human Eye.

distinguishing among different people. Moreover, it has also been established that the iris pattern among twins and even of the same person, right and the left eyes are different from each other.

Using iris as a biometric trait for human authentication involves the following main processes:-

- Acquisition of the Image from the subject
- Iris segmentation and Preprocessing
- Normalization of the segmented Iris
- Generation of Biometric Templates
- Matching of templates and subsequent authentication.

For proper image acquisition, the user must stay at a particular distance and look at the camera at a certain angle depending upon the camera specifications, which requires high user cooperation and constrained environment settings. After the image is captured, segmentation of iris is done by extracting it using the set iris segmentation technique. The next stage is normalization, wherein the segmented image is transformed into one with a preset dimension so as to maintain uniformity in all the segmented images making it easier to act upon in the further stages. After normalization, the next stage is feature extraction or biometric template generation. A biometric template is generated by using mathematical functions. The function is defined as such that the template represents the features in the best possible way for efficient representation. In the last stage, the template is attempted to match with the already existing template of the subject, and authentication is given based on a certain predefined threshold of matching accuracy. Also, if the template is meant to be added into the database against a new subject, then it is added in the 'enrollment' mode of the system, while the former is done in the 'identification' mode of the biometric system.

However, that is not always the case, which may lead to the inclusion of various artifacts such as off-angle gaze, eyelash/eyelid occlusion, motion blur, specular-reflections, etc. due to less user cooperation and non-ideal environments. All these will subsequently lead to poor segmentation results, and the error will be propagated and compounded when the information is passed through further in the system, ultimately leading to faulty results according to many studies [42, 78, 93, 83]. However, if the segmentation can be done accurately enough, then only the relevant information, although augmented, can be propagated further, making the system more accurate then before. A majority of iris
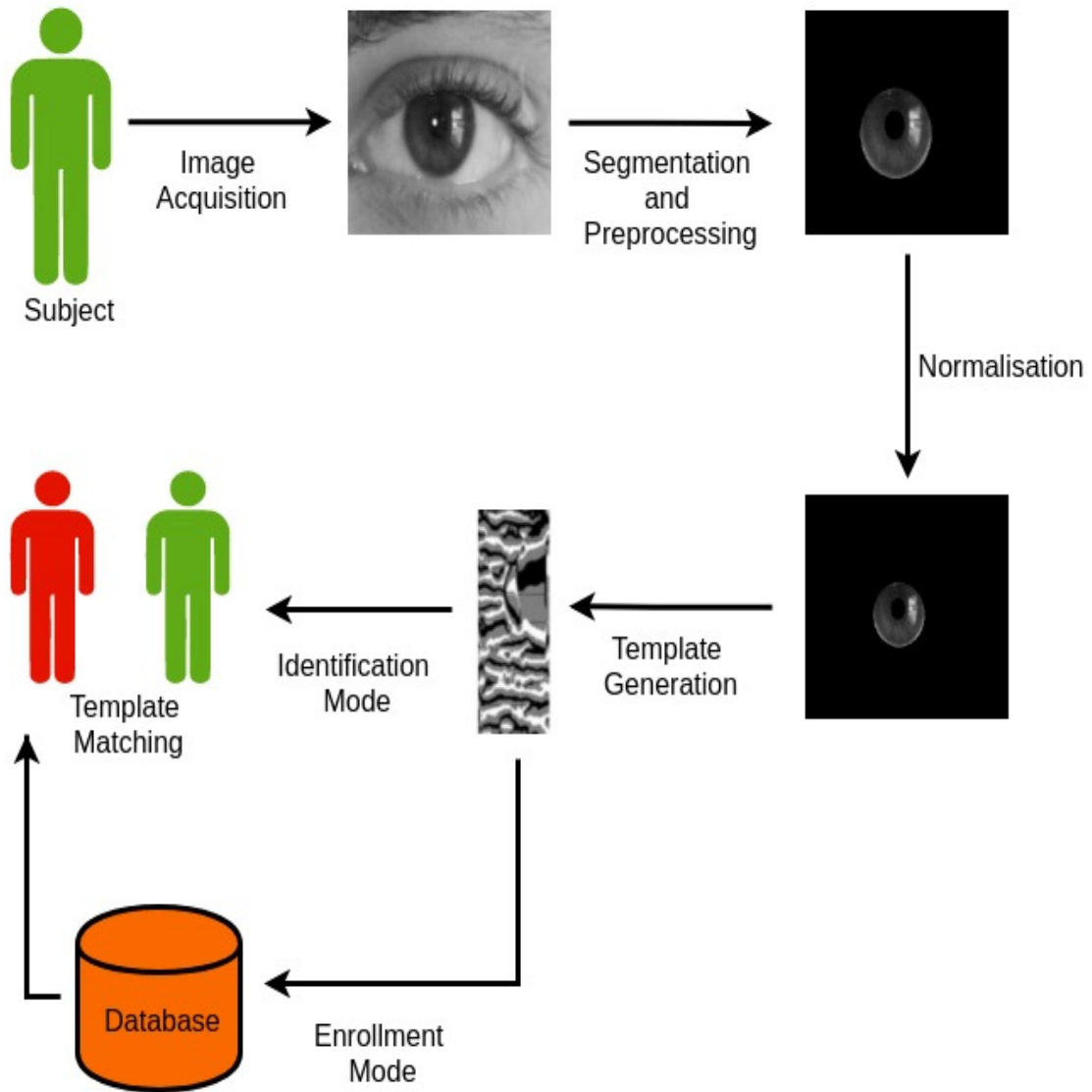
Figure 2: Iris Biometric Pipeline depicting the Identification and the Enrollment mode.
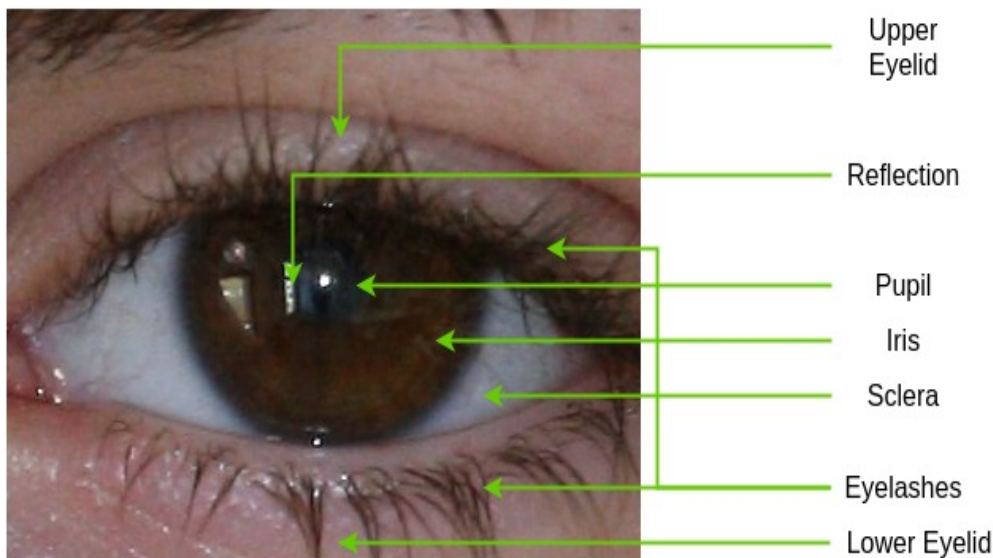
Figure 3: Some of the artifacts in iris image.

segmentation methodologies also assume iris shape to be circular, which is deviated from when the eye is partially closed, which further enhances the need for accurate segmentation techniques in non-ideal environments [27]. Moreover, with the rise in demand for the integration of biometric authentication into our daily lives, non-ideal conditions have to be factored in. In the further sections, we discuss the various segmentation techniques which involve both ideal and non-ideal environments.

2 Deep Learning for Segmentation

Segmentation is taken as the first step to process and understand an image, image segmentation is a process when a grayscale or color image is broken down into the clusters of pixels that contain similar meaningful attributes or homogeneity. Segmentation is titled as the highest domain-independent generalization of an image. For segmentation, a wider variety and in-depth research has been done, but due to task-dependent abstraction of pixels, those methods do not always produce good results.

Artificial neural networks were used way back in the 1940s initially, but it was only until the 1990s when research dwelled deep into this field [30]. With the availability of digital datasets and development towards computational power, deep learning made huge progress. The very first fully Convolutional Neural Network(CNN) was developed by LeCun et al. [63]. Following the work [63], many researchers started putting a huge effort into developing variations of CNN models, which could give higher performance. In the previous decade, deep learning has brought forth a revolution in the field of image analysis, computer vision, etc. Deep learning is being used for tasks like recognition, segmentation, detection, classification, and so on. Some most recognizable architectures which are used as base model for many derivative works are VGG [99], ResNet [39], GoogLeNet [101], MobileNet [44], DenseNet [46] (not exhaustive list).

Segmentation, being a necessary step in not just biometric but also in image processing, medical image analysis, vision tasks, augmented reality, etc.. Use of deep learning-based methods are segmentation tasks outperform classical approaches. Networks such as fully CNNs for pixel-level classification, encoder-decoder based models, attention modules, generative adversarial models, recurrent networks, etc. have been explored vastly for segmentation. Prior to deep learning, segmentation was done through

handcrafted features with rule-based algorithms such as Thresholding [70], K-Means Clustering [32], Watersheds [69], Contours based methods [57], Graph-Cut approaches [19] and Markov Random Fields [76].

Out of all the deep learning based approaches, encoder-decoder models produce the most promising results in the pipeline of image-to-image translation tasks, i.e. image segmentation. The encoder takes the input image and downsamples it by processing with CNN layers to obtain a compressed high-dimensional representation, which is then sent to the decoder, which upsamples the features to map them to the required output which in our case, is a segmentation map. Researchers adopted VGG architecture [99] by removing the fully connected layers of it and then deployed it as encoder while the decoder was developed, such as to mirror the encoder but with upsampling layers. Results obtained for segmentation tasks with such a network were better than other shallow, deep learning models. With modification and replacements in encoder part research kept developing architectures such as SegNet[13], HRNet [107], UNet[85],etc. which further helped the community to increase the performance over-segmentation tasks. UNet [85] architecture depicted in6 was initially developed for medical image segmentation, but later on, it was adopted in other domains as well. Ronneberger et al. [85] introduced novel connections between encoder and deocder layers to facilitate the better parameter updates from the decoder to encoder layers as an experimental setup, which accidentally gave them boosted performance while reducing the problem of vanishing gradients. Later on various modifications were introduced in the UNet, some noticeable works are nested UNet [110], 3D UNet [25] etc. Some more heavy architectures which use region proposal networks, for instance, segmentation are Faster RCNN[84], which was further extended to Mask-RCNN [38]. Many researchers tend to use the atrous-convolution at the lowest dimension of the encoder-decoder model(bottleneck) to increase the receptive field to get the global features better. Chen et al. [22] proposed the famous DeepLab, which uses spatial pyramid pooling, dilated convolution, probabilistic graph model, and deep CNNs for precise boundary detection in the image. The use of the ResNet model[39] as a backbone helps it increase the performance for segmentation tasks. Lastly, General Adversarial Networks(GANs) have shown great potential in segmentation tasks. For example, in work by Hung et al. [49], they proposed a segmentation network with an FCN as the discriminator, which takes predicted mask and ground truth mask as input in addition to the encoder-decoder model for segmentation stage. Recently researchers have been focusing on the network-in-network approach to better extract the features from bottleneck for the input image, which may allow the network to look beyond the obstacles in the image for which a more dense and deep network is required. In an attempt to tackle the issue of occlusion, noise, off-angle, and other non-idealities in iris images [55] proposed a stacked hourglass based model which sits at the bottleneck of an encoder-decoder model. They proposed a unique training strategy to introduce the optimum number of hourglass modules, which can effectively achieve the task of accurate and precise segmentation without the issue of gradient vanishing.

In the area of biometrics, all these deep learning based methods have been adopted, some with a little while some with significant modifications in them. In the subsequent section, we discuss in-depth about the related approaches which make use of deep learning for iris segmentation task.

3   Related Work

In this subsections, we shall discuss existing methods for iris segmentation. We cover both classical image processing techniques where algorithms rely on predefined rules and in-depth learning-based solutions that exploit the availability of massive paired databases for iris segmentation. There is no doubt that the results of deep learning-based methodologies surpass those of classical approaches. Recently data-driven deep learning approaches have proven to give exceptional results in the field of biometrics and beyond it. We provide a comprehensive and comparative study amongst the methods.

3.1   Non-Deep learning Based Methodologies

In iris segmentation, the steps followed in general involve: first, the extraction of the Region Of Interest(ROI) from the complete image, followed by the approximation of two circles which separates the iris region from pupil and sclera [5]. In this subsection, we have tried to cover all categories of iris

segmentation and boundary detection. Classical approaches take advantage of either the pixel-based features or the boundary-based features [61, 14]. The first noticeable work was done by Daugman [29] back in 1993, his work formed the basis of all the work thereafter. In the eye, pupil along with iris are taken as non-concentric circles; an integrodifferential operator localizes the boundary to segment out the iris. The method avoided the use of images, which included any type of occlusions like eyelids, eyelashes, and reflection, etc. Overall, we can say that classical methods followed by Daugman focused on extracting the edges of pupil and iris to localize the to-be segmented area more precisely. Authors relied on rule-based

Table 2: Non Deep learning Based Approaches.

Methods	Novelty	Strength	Performance	Dataset Used
Sardar et al. [95]	Circular sector analysis (CSA) and use of rough entropy for Iris localization	Less computational expensive	E1 error rate : 0.08%, Acc. : 97.12%	IITD, MMU, CASIA
Parikh et al. [71]	Color clustering along with curve fitting	Eyelids detected reduces the chances of error in performance	Acc. : 93.3% (UBIRIS), 94.9% (NICE-II), 92.8% (NICE-I)	NICE-I, NICE-II and UBIRIS v2
Khan et al. [5]	Gradient based approach for localization of iris with sclera boundary points	Iris borders were identified using novel gradient approach	Acc. : 98.22%(MMU), 100%(CASIA)	MMU, CASIA
Pundlik et al. [82]	Graph Cut approach for iris segmentation	Markov random field removes the eyelashes to boost the performance	None	None
Ibrahim et al. [50]	A two stage Hierarchical approach with moving circular window for iris segmentation	Intensity of pupil makes it easier to separate using probability approximation	Acc. : 98.28% (CASIA IrisV3 Lamp), 99.90% (CASIA IrisV1), 99.77% (MMU)	MMU, CASIA
Hu et al. [45]	Fusion of three models to extract the iris with the help of Daugman's method	Integral derivative to detect the iris boundary is time-efficient	E1 error rate : 1.75% (MICHE), 1.30% (UBIRIS v2)	MICHE, UBIRIS v2
Huang et al. [4]	Radial suppression after the thresholding of detected edges	For ideal cases of iris images with assumption of circular iris boundary radial suppression gives good performance	E1 error rate : 0.32%	CASIA
Abate et al. [69]	Novel watershed method in addition to seed selection	Sclera along with eyelash/eyelid are separated using Limbus detection approach	None	MICHE
Jeong et al. [3]	Eye detection with Adaboost and moving edge detector of shape circular	Detecting the obstructions in iris images decreases the chances of error in segmentation map	E1, E2 error rates : 2.8%, 14.4% respectively	UBIRIS.v2
Ibrahim et al. [51]	First derivative based iris detection with adaptive thresholding	Boundary detection takes lesser time	Acc. : 99.13% (MMU)	MMU v1.0
Patel et al. [72]	Binary integrated intensity curve with region growing approach	Detection of eyelash and eyelid increases the performance of segmentation task	E1 error rate : 5.14%	CASIA

lgorithmic approaches, which in some sense limited their methods to work on vast variations of images of iris. We can say that those methods could not perform well on non-ideal images of iris. Realizing the limitation of methods and instead of using simple edge detection steps, they started to use more statistical approaches [50]. With the advancements in research, researchers started to model the anatomy

8

of the eye realistically such as, taking the boundaries of pupil and iris as non-circular. While some of the work did handle the problems of obstruction, specular reflection, and eyelash, etc., their limited feature modeling and extraction approaches were not too vast to cover all the variations in the iris images. In this subsection, we dive deep into the classical work done so far and provide a comparative analysis.

After Daugman, Wilde [108] in 1997 proposed a new approach in which he used a LED point source in addition to a camera for capturing eye images. He identified iris boundaries by gradient dependent binary edge map in addition to the circular Hough Transform. Paper also presented an in-depth comparative study with Daugman's work. While Wilde's work is considerably complex than Daugman's, the segmentation approach proposed by Wilde was better as it detected the eyelids as well as it worked better with noisy images. In [16] Boles et al. proposed a circular edge detection method.

The authors of [95] proposed an iris localization technique, namely, circular sector analysis(CSA), before applying rough entropy for segmentation. Their localization methods decreased the overall uncertainty in the segmentation mask. Another work [5] proposed the iris localization by assuming that shapes of the pupil and iris are circular wherein they first localized the pupil through eccentricity dependent bisection approach then for iris a region totally free from noise was obtained with directional segmentation followed by obtaining the gradients of direction lines to localize iris.

It was not until 2001 when Kong and Zhang in their work [59] incorporated the noisy and occluded images for iris segmentation. They used Hough Transform to isolate the iris followed by 1-d Gabor filters for eyelids detection and thresholding to identify specular reflection. Their work gave better results for segmentation as well as in the final recognition task. Lim et al. [64], segmented the iris images by edge detection method through finding virtual circles where the pupil was detected first by the center-point-detection method. They acquired eye images but from a distance, and to reduce the reflections; they used halogen lamps. Their dataset consisted of both eyes, with and without lens and glasses. Daugman, in his work [26], proposed the algorithm where he detected eyelid occlusion while segmenting the iris. Huang et al. [48] applied a median filter prior to canny operator for edge detection. Outer boundary was detected using a voting scheme on the maximum circle, similarly for an inner boundary was identified using a rectangular inter interval. Localized iris was then segmented with the help of an integrodifferential operator. They, too, handled the eyelid occlusion using thresholding of histogram-based Hough Transform. Huang et al. [47] again proposed a novel segmentation technique that also eliminated the noise to improve the results. They localized the iris using a simple filtering step with edge detection and Hough Transform; occlusion factors were then eliminated using a Gabor filter.

Dorairaj et al. [33] developed a approach to deal with off-angle iris image. In this work, he used PCA and global ICA for the encoding of off-angle iris images, while applying PCA/ICA they first estimated the gazing angle by using Hamming distance followed by a simple integrodifferential operator for segmentation. Daugman in [27] developed an algorithm to tackle off-angle images similar to that of [33] with the elimination of occlusion caused by eyelashes. Abiyev et al., in their work [6], came up with the neural network-based method for iris recognition, a rectangular area of size 10x10 were used to identify the pupil region. For removal of noise, they utilized the standard Linear Hough Transform for eyelids.

Authors of [50] proposed a multi-stage technique. First, a moving window of circular shape was used for pupil estimation, following which the estimation of the pupil was done through the standard-deviation peaks in both x as well as y directions, and after that, a median-filter reduced the eyelash effects. In [3], authors proposed AdaBoost for eye detection for further segmentation. [82] present an unsupervised approach where images were modeled as Markov Random Field. Graph-Cut method extracted the texture region, and for iris segmentation image, intensities were exploited. Roy et al. [90] proposed a non-ideal iris recognition method, in which they used a Mumford-shah segmentation method.

All these classical approaches claim to handle various noises, distortion, and non-ideal iris images, but all being rule-based feature-driven approaches are limited in handling the variation of a non-ideal iris image. In table 2, some classical approaches are compared bases on their novelty and performance.

3.2 Deep Learning Based Methodologies

Deep learning based approaches are the primary source of majority of state-of-the-arts solutions. In segmentation, researchers have come up with numerous modifications of simple autoencoder based convolutional neural networks (CNNs). UNet, as mentioned earlier, has now become the basis network for segmentation tasks. In iris segmentation, researchers shifted towards the deep neural network based approaches gradually as the availability of the iris databases increased, and classical methods were challenged with increasing non-ideality in iris images. CNN remove pre/post-processing steps. CNN models were introduced to increase the accuracy of segmentation masks generated over non-ideal iris images.

Table 3: Deep Learning Based Approaches.

Methods	Novelty	Strength	Performance	Dataset Used
Lakra et al. [61]	CNN architecture with patched input of iris images	Post and pre-cataract surgery the CNN based methods helps in segmentation of Region of interest	E1 error rate : 0.98%	IITD, CASIA
Liu et al. [65]	CNN architecture trained and deployed an end-to-end method	Give significantly good results over non-ideal iris images	E1 error rate : 0.59% (CASIA v4), 0.90% (UBIRIS v2)	CASIA v4, and UBIRIS v2
Wang et al. [106]	Robust single end-to-end model for segmentation	Worked with mobile images of eye, where dataset was generated in nearly wild environment	E1 error rate : 0.72% (CASIA-iris-M1), 0.82% (MICHE-I)	CASIA-Iris-M1, MICHE-I
Arsalan et al. [12]	Inspired from DenseNet for iris segmentation task which does not use any handcrafted features	DensNet architecture allows to increase the network depth which supports in giving improved results	E1 error rate : 0.695%	UBIRIS.v
Wang et al. [105]	Multi-task CNN based model	Use of attention module with novel architecture resulted in increased performance over segmentation task	E1 error rate : 0.41% (CASIA v4), 0.84% (UBIRIS v2), 0.66% (MICHE-I)	CASIA v4-distance, UBIRIS v2, and MICHE-I
Arsalan et al. [10]	Combination of CNNs based model with modified Hough Transform for iris segmentation	Region of interest extracted with HT improves the performance of CNN to produce mask	E1 error rate : 0.82% (UBIRIS v2), 0.345% (MICHE)	UBIRIS v2, MICHE
Chen et al. [23]	Dense block based CNN with dropout and Batch normalisation	Use Labelme software package to label the occlusion in iris images to work with images takes in non-ideal environment	Acc. : 99.05% (CASIA-interval-v4), 98.84% (IITD), 99.47% (UBIRIS v2)	CASIA-Interval-v4, UBIRIS v2, IITD

Liu et al. [65] developed two modalities with CNN, where a multi-scale CNN along with hierarchical CNN was deployed to detect the iris boundaries in non-ideal cases. In [61], authors fed the input iris

images to a series of four dense convolution blocks, feature maps extracted from blocks fused with a weighted sum gave coarse as well as fine features to produce the required segmentation mask finally. Authors of [73] trained two different CNN architectures, which were derived from networks such as Faster RCNN [84] and SSD [66] which localize the circular region of the pupil along with iris. Similarly in [12], authors combined two existing CNN networks the DenseNet [46] and the SegNet [13] for iris segmentation. Another such work which incorporates the features of one CNN architecture into another is [11], they developed a CNN model with residual connections in SegNet which allowed the authors to develope deeper network while reducing the chances of vanishing gradient. In [65], authors developed two CNNs, the first one based on Hierarchical CNNs and the second based on Multi-Scale CNNs. Authors of [15] used the data augmentation approach along with the CNN model to virtually increase the non-ideal iris database for segmentation.

In [106], authors took the non-ideality to the next level, where iris images were taken using the mobile images. They developed a lightweight deep CNN as a complete end-to-end segmentation method. Followed by previous work Wang et al., in their work [105], came up with a Multi-task CNN architecture that also incorporates the attention module for iris segmentation and boundary localization. Similarly, in [56], authors developed EyeNet, an attention based CNN for eye region segmentation.

While numerous work have been published on iris segmentation with CNN both over ideal and non-ideal images, more or less, they propose a few new additional features or modules towards the existing models. Challenges that the deep learning solutions are trying to solve are a variety of textural complexity and shape of iris from person to person, non-ideality such as distortion, non-regular illumination, motion blur, digital noise, and poor image quality, etc. Algorithms are not robust enough to extract the region of interest for segmentation. One such work that handles the majority of the above issues while giving state-of-the-art segmentation maps is [55], wherein authors proposed a three-stage trained novel deep CNN architecture for non-ideal iris segmentation. In addition to novel models, they used a combination of multiple loss functions to give precise segmentation maps. Table 3 compares some selected deep learning based methods.

4 Datasets and Evaluation Metrics

For any work, it is very important to have preset dataset(s) and an appropriate evaluation metric(s) to validate the work and the results. Further, this allows for establishing fixed guidelines for comparison with other methods and establish the standard as well as best methods.

4.1 Datasets

Here, we discuss the various datasets briefly [109] used by the researchers in their work.

4.1.1 CASIA* :

Compilation done by, the Chinese Academy of Science - Institute of Automation(CASIA), it is the first freely available iris database for research purposes [9]. To date, it has four versions CASIA-Iris V1, V2, V3, and V4, wherein each of the datasets have their own subsets[1]. The first version i.e., CASIA-Iris V1, comprises 756 iris images(320x280) that were acquired from 108 subjects using a home-made iris camera. The second version comprises of two equal subsets, each comprising of 1200 iris images(640x480) acquired through OKI IRISPASS-h device and CASIA-IrisCamV2. As compared to its predecessors, the third version introduced important noise factors and comprised nearly 22,034 images of iris of 700 subjects divided unequally among three sets. The Interval set has 2,639 images(320x280), the Lamp subset has 16,212 images(640x480) and the Twins subset has 3,183 images(640x480) collected from 100 pair of twins. The latest version, i.e., CASIA-Iris V4, which is an extended version of CASIA-Iris V3, consisting of the addition of three new subsets. The first subset CASIA-Iris-Distance comprises of 2576 images(2352x1728) while the second subset CASIA-Iris-Thousand comprises of 20,000 images(640x480),

---

*Link-http://www.cbsr.ia.ac.cn/english/IrisDatabase.asp

and the last subset CASIA-Iris-Syn comprises of 10,000(640x480) generated images from CASIA-Iris V1. The versions were released in the order in 2002, 2004, 2010, and 2010.
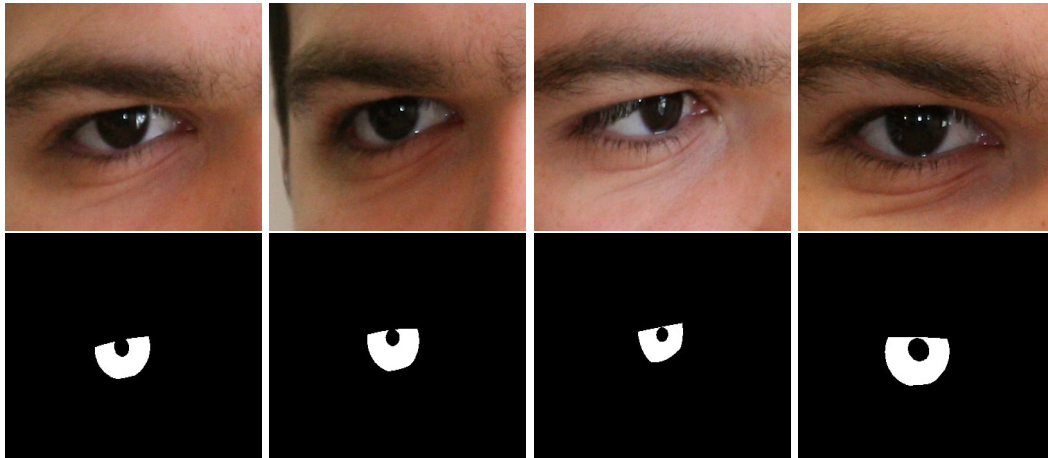

Figure 4: Some images and their corresponding groudtruth segmentation masks from the UBIRIS-v2 dataset[80].

4.1.2 UBIris v1[†] and UBIris v2[‡] :

The UBIris datasets [81, 80] were compiled by the Soft Computing and Image Analysis Group(SOCIA), University of Beira Interior, Portugal. V1 which comprises nearly 1877 images from 241 subjects whereas V2 comprises of 11,102 images from 259 subjects. V1 was captured Nikon E5700 camera in two parts wherein first part; the noise elements were controlled by having the image acquire setup in a unilluminated room, while in the second part, the images were under normal light which simulated the images captured with minimal active participation and introduced several noise factors such as contrast, focus, reflections, etc. V2 was acquired using a Canon EOS 5D camera in unconstrained environments such as, on the visible-wavelength and on-the-go, which simulated more realistic noise factors as compared to V1. V1 was released in 2004, and V2 was released in 2010. While v1 is available in multiple resolutions such as 800x600 pixels, and 200x150 pixels, v2 is available in 400x300.

4.1.3 NICE-I[§] and NICE-II[¶] :

Both these datasets [77, 79], part of the Noisy Iris Challenge Evaluation, have been distributed by the same group as UBIris. Both the datasets are subsets of UBIris v2. NICE-I was held in 2008, and NICE-II was held in 2010.

4.1.4 ND-Iris-0405[‖] :

This dataset[17] include more than 64,979 iris images acquired from nearly 356 subjects taken from 2004 to 2005. It's subset [75] is associated with the iris challenge evaluation, which was organized by the

---

[†]Link-http://iris.di.ubi.pt/ubiris1.html
[‡]Link-http://iris.di.ubi.pt/ubiris2.html
[§]Link-http://nice1.di.ubi.pt/
[¶]Link-http://nice2.di.ubi.pt/
[‖]Link-https://cvrl.nd.edu/projects/data/

National Institute of Standards and Technology, the USA in 2005. It comprises of 2953 iris images of resolution 480x640 acquired from 132 subjects under NIR illumination using an LG EOU 2200 acquisition system.

4.1.5   IITD** :

It was compiled by the Biometrics Research Laboratory of Indian Institute of Technology Delhi, India, and contains 2240 images of resolution 320x240 of nearly 224 subjects acquired using a fully-digital CMOS, JPC1000, JIRIS camera. The subjects comprised of students and staff at IIT D itself having age between 14-55 years out of which 48 subjects were female, and 176 were male. The dataset [60] was published in 2007.
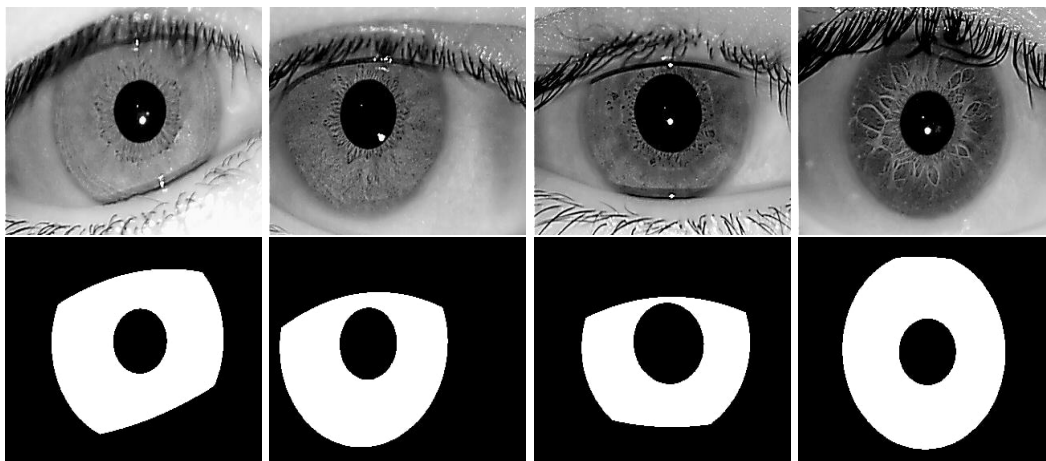

Figure 5: Some images and their corresponding groudtruth segmentation masks from the IITD dataset[60].

4.1.6   CSIP†† :

Short for Cross-Sensor Iris and Periocular dataset [94], it was too compiled by the SOCIA group, and was acquired using four different mobile devices, i.e., W200 (THL), Xperia Arc S (Sony Ericsson), U8510 (Huawei), and iPhone 4 (Apple). The images were taken by choosing both the front and the rear cameras with flash, which led to 10 combinations and their corresponding setups. Also, the lighting condition was varied between natural, artificial, and mixed. Owing to all the factors, several noises were incorporated in the dataset. The dataset comprises of 2004 iris images of multiple resolutions from 50 subjects and was released in 2014.

4.1.7   MICHE-I and MICHE-II‡‡ :

Both MICHE-I and MICHE-II datasets [31] have been created specifically for mobile biometric applications. Part of the Mobile Iris Challenge Evaluation and compiled by the Biometric and Image Processing Lab, University of Salerno, Italy, it[31] comprises of images captured solely from mobile phones in non-restrained environments without the use of any sophisticated equipment to model real-world

---

**Link-http://www4.comp.polyu.edu.hk/ csajaykr/IITD/Database$_I$ris.htm

††Link-http://csip.di.ubi.pt/

‡‡Link-http://biplab.unisa.it/MICHE/MICHE-II/

image acquisition thereby incorporating various noise factors into the dataset. Images were captured using three mobile devices, namely Samsung Galaxy S4(2322x4128), Samsung Galaxy Tab2(640x480), and iPhone5(1536x2048) where each captured 1,297, 632, and 1,262 images respectively. MICHE-II SPECIFICATIONS MICHE-I was released in 2015, and MICHE-II was released in 2016.

4.1.8 SBVPI[§§] :

It is distributed by the Faculty of Computer and Information Science, University of Ljubljana, and comprises of 1858 high resolution(3000x1700) eye images acquired from 55 subjects. Each subject contributed 32 images, which comprise of the person looking at four different gaze-directions i.e., straight, up, left, and right. As the name suggests, corresponding to each image, there is a separate binary mask for the sclera, pupil, iris, and periocular region. It is a fairly new dataset [104, 88, 87] released in 2018.

4.1.9 IRISSEG-CC[¶¶] :

The dataset [43, 8] has been compiled by the Halmstad University wherein the ground truths for the subset of or the whole dataset of three other iris datasets are generated. The first is the BioSec Multimodal Biometric Database Baseline [35], which was acquired through an LG IrisAccess EOU3000 close-up infrared iris camera for 3200 images(640x480) from 200 subjects. The IRISSEG-CC comprises of ground truth for 75 of them. Next is the CASIA Iris v3 Interval Database [1] which comprises of 2655 iris images(320x280) from 249 subjects acquired using a close-up infrared iris camera. The whole ground truths for this one was compiled into the IRISSEG-CC. Lastly is the MobBIO Database [97], which contains 800 iris images(240x200) from 100 subjects through an Asus Eee Pad Transformer TE300T Tablet. There were two distinct illumination conditions while varying the orientation of the eye with considerable occlusion. For this, too, the ground truth for all the images was compiled.

4.1.10 IRISSEG-EP[***] :

Compiled along with IRISSEG-CC [43, 8], the dataset [43] was compiled by the Multimedia Signal Processing and Security Lab, University of Salzburg. It comprises ground truths of other iris datasets, namely UBIRIS v2[80], IIT D[60], Notredame 0405 Iris Image Dataset[17], and CASIA Iris v4 Interval[1]. The ground truths for the Notredame dataset consists of 837 images(640x480) whose original images were acquired using LG 2200 close-up near-infrared camera in indoor lighting but with noises such as occlusion, off angle, and blur. The ground truths for the Casia dataset consist of 2639 images(320x280) whose original images were taken using CASIA close-up near-infrared camera.

4.1.11 MMU1 and MMU2[†††] :

Courtesy of Multimedia University, MMU1[2] comprises about 450 images of nearly 45 subjects captured using LG IrisAccess 2200. MMU2 comprises 995 images from 100 subjects acquired using the Panasonic BM-ET100US camera. Images in the dataset are of poor resolution and were taken in NIR lighting.

4.1.12 OpenEDS[‡‡‡] :

The dataset [36] was compiled by Facebook Research comprising of 356,649 eye images of resolution 400x640 were collected from 152 subjects, wherein only 12,759 images have pixel-level annotations of the pupil, the iris, the sclera, and the background. The images were acquired under controlled illumination through a head-mounted display.

[§§]Link-http://sclera.fri.uni-lj.si/database.html

[¶¶]Link-http://islab.hh.se/mediawiki/Iris$_S$egmentation$_G$roundtruth

[***]Link-http://www.wavelab.at/sources/Hofbauer14b/

[†††]Link-http://pesona.mmu.edu.my/ ccteo/

[‡‡‡]Link-https://research.fb.com/programs/openeds-challenge/

4.1.13 iBUG[§§§] :

This dataset was compiled by the Intelligent Behavior Understanding Group for their work [67]. They compiled their own non-ideal iris dataset through manual annotation of nearly 4461 face images picked individually from IMDB [89], HELEN [62], UTDallas Face database [68] 300VW [98], CVL [74], 300W [92], and Columbia Gaze database [100] to finally obtain 8882 iris images.

Here, we discussed not only the old and explored datasets but also about some new datasets that are yet to be extensively experimented upon.

4.2 Performance Metrics

In this subsection, we will discuss the metrics that are used for quantifying the results of the various works.

4.2.1 Jaccard Index (JI):

It signifies the overlap i.e., intersection over the combined area i.e., a union of the segmentation maps for each class. It is calculated over each class and averaged as given by the formula :

$$JI = \frac{1}{N} \sum_{i=1}^{N} \frac{C_{ii}}{GT_i + P_i - C_{ii}} \tag{1}$$

Here, N is 2 i.e. binary classes. C_{ii} is the common pixels i.e., all the pixels having both the ground truth and predicted label as i. Here P_i and GT_i are the number of pixels where predicted label is i and the other whose ground truth label is i, respectively. The final value is reported after averaging over all the images.

4.2.2 Mean Segmentation Error:

Also termed as E_1, it is the overall pixel-wise classification error (PCL) calculated as the exclusive-OR(XOR) ($\oplus$) between given segmentation map(M^{gt}) and the predicted segmentation map(M^p). The equation is given as:

$$PCL(M^p, M^{gt}) = \frac{1}{\#ofpixels} \sum_{i=1}^{\#ofpixels} M^p(pix_i) \oplus M^{gt}(pix_i) \tag{2}$$

Thereafter, PCL is calculated for all the testing images and averaged to report the overall mean segmentation error i.e. E_1.

4.2.3 Nice2 error:

Nice2 error or E_2 is another measure to evaluate the disparity of the two regions i.e., non-iris and iris pixels. E_2^i, the error for i^{th} image is computed by taking mean of the False-Positive Rate (FPR) along with False-Negative Rate (FNR), which itself is computed at pixel level. Formulas for all stated are given by :

$$FNR = \frac{1}{\#ofpixels} \sum_{i=1}^{\#ofpixels} [((M^{gt}(pix_i) \cdot M^p(pix_i)) \oplus M^{gt}(pix_i)] \tag{3}$$

[§§§]Link-https://ibug.doc.ic.ac.uk/resources/ ibug-eye-segmentation-dataset/

15

$$FPR = \frac{1}{\#ofpixels} \sum_{i=1}^{\#ofpixels} [((M^{gt}(pix_i) \cdot M^p(pix_i)) \oplus M^p(pix_i)] \tag{4}$$

$$E_2^i = \frac{FNR + FPR}{2} \tag{5}$$

Thereafter, E_2^i is computed for every test image and averaged to report the final E_2. E_1 and E_2 are bounded between [0, 1], and since are errors, the closer their values are to "0" the better while the opposite holds for values closer to "1". However, the opposite is true for IOU.

Next, we briefly state some metrics standard in the case of classification, in our case, binary classification. For that, we first define some terms:

- True-Positive (T_p): total foreground-pixels classified correctly as iris pixels.
- False-Positive (F_p): total pixels incorrectly classified as foreground-pixels.
- True-Negative (T_n): total background-pixels classified correctly as non-iris pixels.
- False-Negative (F_n): total pixels incorrectly classified as background-pixels.

With the knowledge of the above terminology, the following metrics are defined :

1. **Accuracy:** Fraction of all pixels classified correctly irrespective of the class upon all the pixels in a dataset.

$$Accuracy = \frac{T_p + T_n}{T_p + F_p + T_n + F_p} \tag{6}$$

2. **Precision:** Fraction of all positive-class pixels predicted correctly upon all the pixels predicted as positive.

$$P = \frac{T_p}{F_p + T_p} \tag{7}$$

3. **Recall:** It is the fraction of all the positive-class pixels classified correctly upon all the positive-class pixels.

$$R = \frac{T_p}{F_n + T_p} \tag{8}$$

4. **F-score:** It optimises both the recall and precision as it is the harmonic mean of both.

$$F - score = \frac{2RP}{P + R} \tag{9}$$

5   Experimentation

Here, we briefly discuss an implementation of the UNet[85] done by us on two majorly used datasets, namely, UBIris v2[80] along with CASIA v4 Interval[1]. We have already discussed the UNet in the above sections whose implementation has been open sourced by the authors [34] ¶¶¶. It is a simple encoder-decoder CNN based architecture wherein novel skip connections joining encoder layers to the decoder layers which provide a global context to the already processed local information for generation of location-precise maps. For each of the datasets, a similar procedure as described below, was followed. Firstly, the images in the dataset were reshaped into 256x256 and were normalised between [0,1] by dividing it by 255. Next, the dataset was divided in to a 30% and 70% into testing-set and training-set

---

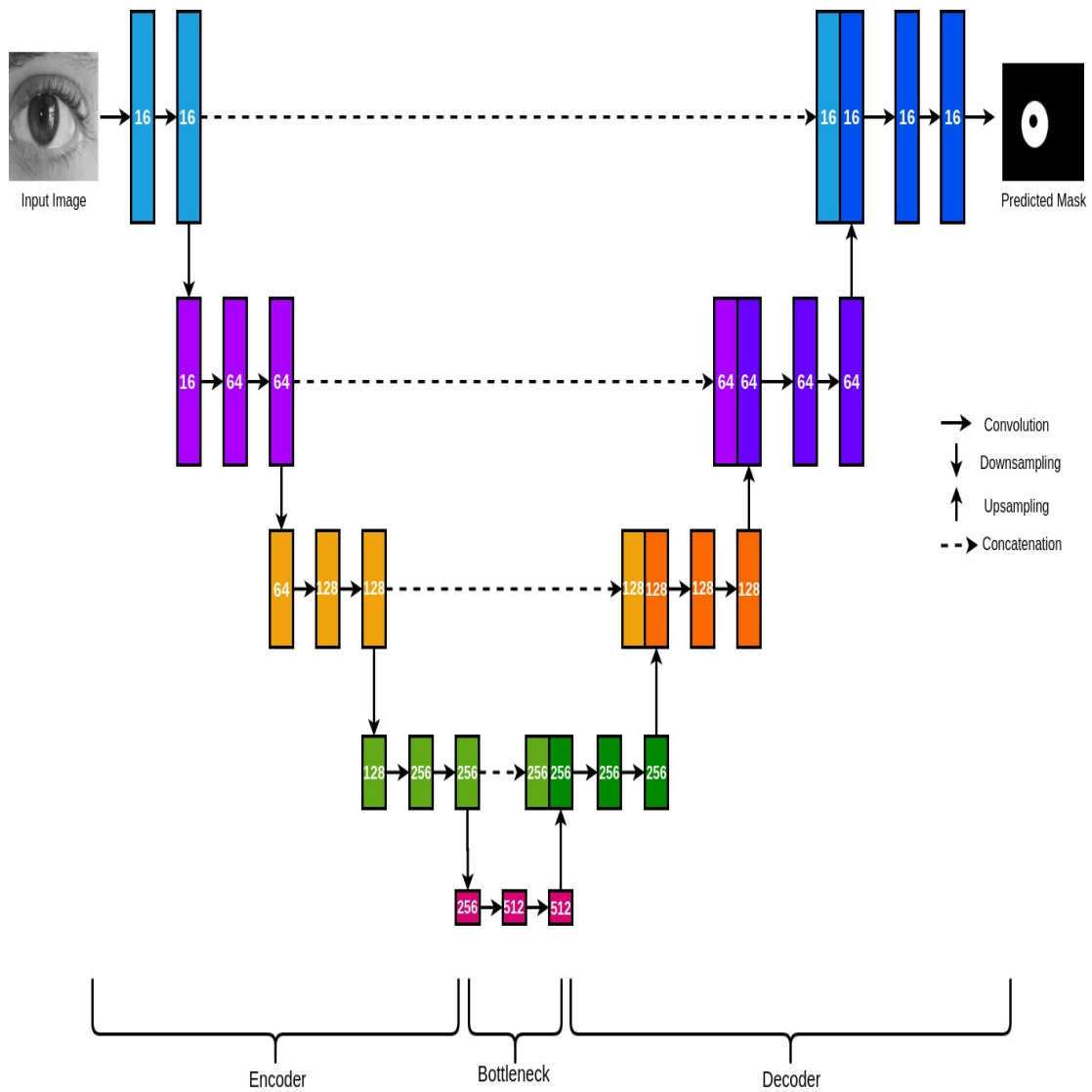¶¶¶Link-https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/

Figure 6: UNet[85].

respectively. We coded our model using Keras[24], and trained on the NVIDIA GeForce GTX 1080 Ti for 100 epochs.

Table 4: Architecture of the implemented UNet.

Bock Name	Layer Name	No. of Filters	Strides	Output Shape
-	Input	-	-	256x256x1
Encoder	Conv1_1	16	(1,1)	256x256x16
	Conv1_2	16	(1,1)	256x256x16
	Pool1	16	(2,2)	128x128x16
	Conv2_1	64	(1,1)	128x128x64
	Conv2_2	64	(1,1)	128x128x64
	Pool2	64	(2,2)	64x64x64
	Conv3_1	128	(1,1)	64x64x128
	Conv3_2	128	(1,1)	64x64x128
	Pool3	128	(2,2)	32x32x128
	Conv4_1	256	(1,1)	32x32x256
	Conv4_2	256	(1,1)	32x32x256
	Pool4	256	(2,2)	16x16x256
BottleNeck	Conv1_1	512	(1,1)	16x16x512
	Conv1_2	512	(1,1)	16x16x512
Decoder	Up1	256	(2,2)	32x32x256
	Concat1	-	-	32x32x512
	Conv1_1	256	(1,1)	32x32x256
	Conv1_2	256	(1,1)	32x32x256
	Up2	128	(2,2)	64x64x128
	Concat2	-	-	64x64x256
	Conv2_1	128	(1,1)	64x64x128
	Conv2_2	128	(1,1)	64x64x128
	Up3	64	(2,2)	128x128x64
	Concat3	-	-	128x128x128
	Conv3_1	64	(1,1)	128x128x64
	Conv3_2	64	(1,1)	128x128x64
	Up4	16	(2,2)	256x256x16
	Concat4	-	-	256x256x32
	Conv4_1	16	(1,1)	256x256x16
	Conv4_2	16	(1,1)	256x256x16
-	Output	1	(1,1)	256x256x1

The loss function was Binary-cross-entropy, and E_1 and E_2 were taken as the validation metrics.

Table 5: Testing results on the implementation of UNet.

Dataset	E1 % error	E2 % error
UBIris v2	0.54	3.12
Casia-Iris v4 (Interval)	0.67	1.36

The kernel size was kept 3x3 with stride 1 for convolution and 3x3 with stride 2 for both downsampling and upsampling. Each convolution was followed by a ReLu activation[7] and Batch Normalization[52] for better generalisation. Different optimisers such as SGD[91], RMSProp[102], and Adam[58] were used, and the best results reported below were obtained using Adam. All weights were initialised according to the He et al. [40] initialisation. In accordance with the above nomenclature, the loss function is defined as:

$$L_{bce}(M^p, M^{gt}) = \frac{-1}{\#ofpixels} \sum_{i=1}^{\#ofpixels} \left[M^{gt}(pix_i) * \log(M^p(pix_i)) + (1 - M^{gt}(pix_i)) * \log(1 - M^p(pix_i)) \right] \tag{10}$$

6 Challenges identified and further direction

Before discussing about the future work, we again describe the various non-idealities introduced in the dataset and also some future non-idealities that may be encountered in the future datasets.

- **Occlusion** Eyelids, Eyelashes, Hair are the leading cause of occlusion in Iris images with massively varying levels of occlusions sometimes to the tune of 80-90%.

- **Blurring** In many cases, either the subject is on the move in unconstrained environments, or even in the case of constrained environments, and they may move a little causing the image to blur. Also, in some cases, if the equipment is not correctly setup, the camera itself may move, adding to the blurriness.

- **Alignment** In some cases, the subject may move their eye or even their entire head, which causes the iris to appear more oval-shaped and not occupy the center of the image as intended due to misalignment of the face and the equipment. Needless to say, in the case of an unconstrained environment, this is highly prevalent.

- **Resolution** For proper segmentation and subsequent verification, it is always the best to have high resolution images so that the relevant features are easily extracted and represented. However, this is not always the case since the resolution of the acquired image is solely dependent on the camera equipment, which may vary from a high-end imaging setup to a mobile camera.

- **Adulteration** This is an artifact that has the potential to appear in many ways. For example, subjects wearing lenses or glasses alter the natural appearance of the iris. Although indistinguishable to the naked eye, they alter the intensity of the original iris at the pixel level, which may cause substantial hindrance to the segmentation algorithms. In the future, more "artificialness" might be introduced in presently unknown ways.
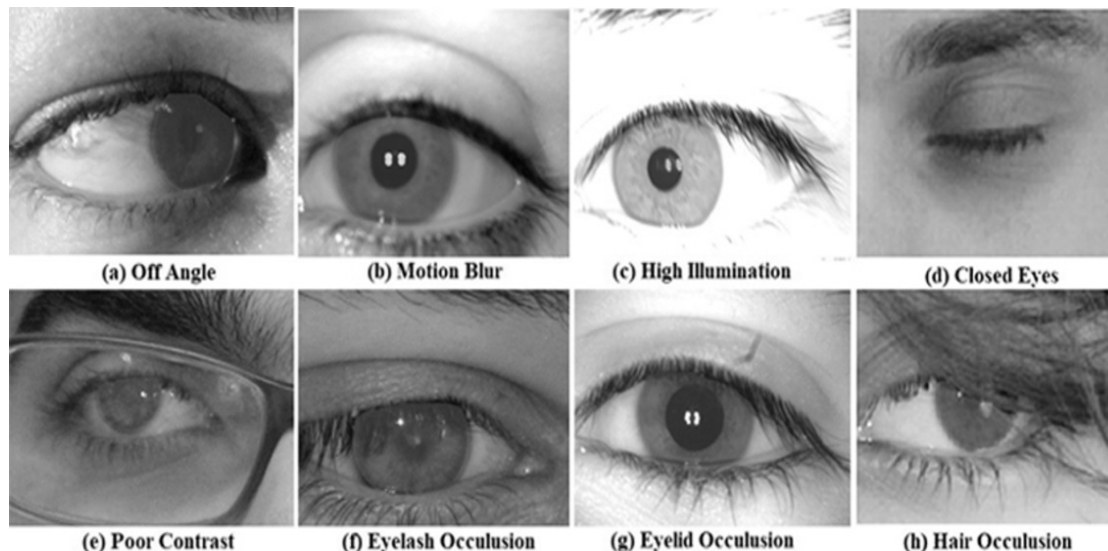
19

Figure 7: Some examples of the artifacts present in Non-Ideal Iris images.

It has been widely discussed that models based on Deep learning are heavily dependent on data characteristics. Iris images acquired in an unregulated condition represent nearly all the characteristics of the real world scenarios. Developed robust and complex deep learning models such as PixISegNet, Iris-DenseNet, etc. are now able to process and can be trained on the datasets generated in an unconstrained environment and hence are bound to give excellent performance when deployed to the real world. However, there is still much of room for development where the researchers can improve upon the robustness of the model to look beyond occlusion and identify the features of anatomy nearby iris. A lot of work is yet to be done GANs, a generative model wherein an encoder-decoder may act as an active generator and a separate discriminator to differentiate amongst the predicted map and the ground truth maps may result in the improved performance. Similar to this, the use of attention modules, dictionary learning, recurrent neural network, and many other existing ideas are not thoroughly explored by the researchers. Moreover, most of the models have been trained upon existing loss functions, as discussed above, leaving room for improvement there too. Hence, there is still a vast space for the development of novel and state-of-the-art techniques, and overcome all the drawbacks of previous works.

7   Conclusion

In the previous decades, numerous researchers have devised several biometric systems based on various biometric traits depending upon the specific need and application of the system such as retina, fingerprints, palm, voice, etc. However, one thing common with most of the biometric traits is them being invasive or having any form of contact for acquisition, e.g. - touching the scanners for fingerprints, hand veins, signing a phrase, or requiring some form of cooperation from the user. Due to this, non-invasive data acquisition and methods for biometric systems have started surfacing more and more. Moreover, with the recent spread of the highly transmissible nCovid-19, the need for non-invasive biometrics is bound to increase exponentially. While the face and iris are suitable non-invasive biometric candidates, the face lacks permanence unlike the iris. In this manner, iris can be said to emerge as the leader in non-invasive biometrics and making its study all the more relevant. In accordance, in this study, we

discussed various deep learning and non deep learning based methodology concerning the extraction of the iris portion(a crucial step in Iris Biometric). Also, we provided information about the various iris datasets available to the public, along with various metrics to compare the works and help the research community.

8   Acknowledgements

References

[1] Casia iris image database, http://biometrics.idealtest.org.

[2] Multimedia-university. mmu database [online]. available: http://pesona.mmu.edu.my/ ccteo/.

[3] A new iris segmentation method for non-ideal iris images.

[4] A novel iris segmentation using radial-suppression edge detection. Dec. 2009.

[5] Automatic localization of pupil using eccentricity and iris using gradient based method. *Optics and Lasers in Engineering*, 49(2):177 – 187, 2011.

[6] R. H. Abiyev and K. Altunkaya. Personal iris recognition using neural network. *Intl. J. Security its Appl 2*, 37(2):41–50, 2008.

[7] A. F. Agarap. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018.

[8] F. Alonso-Fernandez and J. Bigun. Near-infrared and visible-light periocular recognition with gabor features using frequency-adaptive automatic eye detection. *IET Biometrics*, 4(2):74–89, 2015.

[9] M. M. Alrifaee, M. M. Abdallah, and B. G. Al Okush. Ashort survey of iris images databases.

[10] M. Arsalan, H. G. Hong, R. A. Naqvi, M. B. Lee, M.-C. Kim, D. S. Kim, C. S. Kim, and K. R. Park. Deep learning-based iris segmentation for iris recognition in visible light environment. *Symmetry*, 9:263, 2017.

[11] M. Arsalan, D. S. Kim, M. B. Lee, M. Owais, and K. R. Park. Fred-net: Fully residual encoder–decoder network for accurate iris segmentation. *Expert Systems with Applications*, 122:217 – 241, 2019.

[12] M. Arsalan, R. A. Naqvi, D. S. Kim, P. H. Nguyen, M. Owais, and K. R. Park. Irisdensenet: Robust iris segmentation using densely connected fully convolutional networks in the images by visible light and near-infrared light camera sensors. *Sensors (Basel, Switzerland)*, 18(5), May 2018.

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

[14] S. BARPANDA, B. Majhi, P. Sa, A. Kumar, and S. Bakshi. Iris feature extraction through wavelet mel-frequency cepstrum coefficients. *Optics Laser Technology*, 110, 03 2018.

[15] S. Bazrafkan, S. Thavalengal, and P. Corcoran. An end to end deep neural network for iris segmentation in unconstrained scenarios. *Neural Networks*, 106:79 – 95, 2018.

[16] W. W. Boles and B. Boashash. A human identification technique using images of the iris and wavelet transform. *IEEE Transactions on Signal Processing*, 46(4):1185–1188, 1998.

[17] K. W. Bowyer and P. J. Flynn. The nd-iris-0405 iris image dataset. *arXiv preprint arXiv:1606.04853*, 2016.

[18] K. W. Bowyer, K. Hollingsworth, and P. J. Flynn. Image understanding for iris biometrics: A survey. *Computer vision and image understanding*, 110(2):281–307, 2008.

[19] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.

[20] T. Burghardt. A brief review of biometric identification. *Retrieved August*, 6:2010, 2009.

[21] J. L. Cambier and J. E. Siedlarz. Portable authentication device and method using iris patterns, Mar. 11 2003. US Patent 6,532,298.

[22] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP, 06 2016.

[23] Y. Chen, W. Wang, Z. Zeng, and Y. Wang. An adaptive cnns technology for robust iris segmentation. *IEEE Access*, PP:1–1, 05 2019.

[24] F. Chollet et al. Keras. `https://keras.io`, 2015.

[25] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3d u-net: Learning dense volumetric segmentation from sparse annotation. In S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pages 424–432, Cham, 2016. Springer International Publishing.

[26] J. Daugman. How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):21–30, 2004.

[27] J. Daugman. New methods in iris recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 37(5):1167–1175, 2007.

[28] J. Daugman. Information theory and the iriscode. *IEEE transactions on information forensics and security*, 11(2):400–409, 2015.

[29] J. G. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1148–1161, 1993.

[30] T. D. D.D. Cox. Neural networks and neuroscience-inspired computer vision. *Curr. Biol. 24 (18)*, 2014.

[31] M. De Marsico, M. Nappi, D. Riccio, and H. Wechsler. Mobile iris challenge evaluation (miche)-i, biometric iris dataset and protocols. *Pattern Recognition Letters*, 57:17–23, 2015.

[32] N. Dhanachandra, K. Manglem, and Y. J. Chanu. Image segmentation using k -means clustering algorithm and subtractive clustering algorithm. *Procedia Computer Science*, 54:764 – 771, 2015.

[33] V. Dorairaj, N. A. Schmid, and G. Fahmy. Performance evaluation of non-ideal iris based recognition system implementing global ica encoding. In *IEEE International Conference on Image Processing 2005*, volume 3, pages III–285, 2005.

[34] T. Falk, D. Mai, R. Bensch, Ö. Çiçek, A. Abdulkadir, Y. Marrakchi, A. Böhm, J. Deubner, Z. Jäckel, K. Seiwald, et al. U-net: deep learning for cell counting, detection, and morphometry. *Nature methods*, 16(1):67–70, 2019.

[35] J. Fierrez, J. Ortega-Garcia, D. T. Toledano, and J. Gonzalez-Rodriguez. Biosec baseline corpus: A multimodal biometric database. *Pattern Recognition*, 40(4):1389–1392, 2007.

[36] S. J. Garbin, Y. Shen, I. Schuetz, R. Cavin, G. Hughes, and S. S. Talathi. Openeds: Open eye dataset. *arXiv preprint arXiv:1905.03702*, 2019.

[37] S. N. Garg, R. Vig, and S. Gupta. A critical study and comparative analysis of multibiometric systems using iris and fingerprints. *International Journal of Computer Science and Information Security*, 15(1):549, 2017.

[38] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988, 2017.

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

[40] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[41] M. Hill. Anat2310: Eye development. *The University of South Wales*, 2003.

[42] H. Hofbauer, F. Alonso-Fernandez, J. Bigun, and A. Uhl. Experimental analysis regarding the influence of iris segmentation on the recognition rate. *Iet Biometrics*, 5(3):200–211, 2016.

[43] H. Hofbauer, F. Alonso-Fernandez, P. Wild, J. Bigun, and A. Uhl. A ground truth for iris segmentation. In *2014 22nd international conference on pattern recognition*, pages 527–532. IEEE, 2014.

[44] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *ArXiv*, abs/1704.04861, 2017.

[45] Y. Hu, K. Sirlantzis, and G. Howells. Improving colour iris segmentation using a model selection technique. *Pattern Recognition Letters*, 57:24 – 32, 2015. Mobile Iris CHallenge Evaluation part I (MICHE I).

[46] G. Huang, Z. Liu, and K. Q. Weinberger. Densely connected convolutional networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, 2017.

[47] J. Huang, Y. Wang, T. Tan, and J. Cui. A new iris segmentation method for recognition. In *Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 3 - Volume 03*, ICPR '04, page 554–557, USA, 2004. IEEE Computer Society.

[48] Y.-P. Huang, S.-W. Luo, and E.-Y. Chen. volume 1, pages 450 – 454 vol.1, 02 2002.

[49] W. Hung, Y. Tsai, Y. Liou, Y. Lin, and M. Yang. Adversarial learning for semi-supervised semantic segmentation. Jan. 2019. 29th British Machine Vision Conference, BMVC 2018 ; Conference date: 03-09-2018 Through 06-09-2018.

[50] M. T. Ibrahim, T. M. Khan, S. A. Khan, M. A. Khan, and L. Guan. Iris localization using local histogram and other image statistics. *Optics and Lasers in Engineering*, 50(5):645 – 654, 2012.

[51] M. T. Ibrahim, T. Mehmood, M. A. Khan, and L. Guan. A novel and efficient feedback method for pupil and iris localization. In *ICIAR*, 2011.

[52] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[53] A. K. Jain, P. Flynn, and A. A. Ross. *Handbook of biometrics*. Springer Science & Business Media, 2007.

[54] A. K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology*, 14(1):4–20, 2004.

[55] R. R. Jha, G. Jaswal, D. Gupta, S. Saini, and A. Nigam. Pixisegnet: pixel-level iris segmentation network using convolutional encoder–decoder with stacked hourglass bottleneck. *IET Biometrics*, 9(1):11–24, 2019.

[56] P. Kansal and S. Nathan. Eyenet: Attention based convolutional encoder-decoder network for eye region segmentation, 10 2019.

[57] W. A. . T. D. S. Kass, M. Active contour models. *Int J Comput Vision 1*, 1998.

[58] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[59] W. K. Kong and D. Zhang. Accurate iris segmentation based on novel reflection and eyelash detection model. In *Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing. ISIMP 2001 (IEEE Cat. No.01EX489)*, pages 263–266, 2001.

[60] A. Kumar and A. Passi. Comparison and combination of iris matchers for reliable personal authentication. *Pattern recognition*, 43(3):1016–1026, 2010.

[61] A. Lakra, P. Tripathi, R. Keshari, M. Vatsa, and R. Singh. Segdensenet: Iris segmentation for pre- and-post cataract surgery. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 3150–3155, 2018.

[62] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang. Interactive facial feature localization. In *European conference on computer vision*, pages 679–692. Springer, 2012.

[63] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[64] S. Lim, K. Lee, O. Byeon, and T. Kim. Efficient iris recognition through improvement of feature vector and classifier. *ETRI Journal*, 23(2):61–70, 2001.

[65] N. Liu, H. Li, M. Zhang, J. Liu, Z. Sun, and T. Tan. Accurate iris segmentation in non-cooperative environments using fully convolutional networks. *2016 International Conference on Biometrics (ICB)*, pages 1–8, 2016.

[66] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. Ssd: Single shot multibox detector. In B. Leibe, J. Matas, N. Sebe, and M. Welling, editors, *Computer Vision – ECCV 2016*, pages 21–37, Cham, 2016. Springer International Publishing.

[67] B. Luo, J. Shen, S. Cheng, Y. Wang, and M. Pantic. Shape constrained network for eye segmentation in the wild. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 1952–1960, 2020.

[68] M. Minear and D. C. Park. A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, 36(4):630–633, 2004.

[69] L. Najman and M. Schmitt. Watershed of a continuous function. *Signal Processing*, 38(1):99 – 112, 1994. Mathematical Morphology and its Applications to Signal Processing.

[70] N. Otsu. A threshold selection method from gray-level histograms. 1979.

[71] Y. Parikh, U. Chaskar, and H. Khakole. Effective approach for iris localization in nonideal imaging conditions. In *Proceedings of the 2014 IEEE Students' Technology Symposium*, pages 239–246, 2014.

[72] H. Patel, C. K. Modi, M. Paunwala, and S. Patnaik. Human identification by partial iris segmentation using pupil circle growing based on binary integrated edge intensity curve. *2011 International Conference on Communication Systems and Network Technologies*, pages 333–338, 2011.

[73] S. M. Patil, R. R. Jha, and A. Nigam. Ipsegnet : Deep convolutional neural network based segmentation framework for iris and pupil. In *2017 13th International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*, pages 184–191, 2017.

[74] P. Peer. Cvl face database. *Computer vision lab., faculty of computer and information science, University of Ljubljana, Slovenia. Available at http://www. lrv. fri. uni-lj. si/facedb. html*, 2005.

[75] P. J. Phillips, K. W. Bowyer, P. J. Flynn, X. Liu, and W. T. Scruggs. The iris challenge evaluation 2005. In *2008 IEEE Second International Conference on Biometrics: Theory, Applications and Systems*, pages 1–8. IEEE, 2008.

[76] N. Plath, M. Toussaint, and S. Nakajima. Multi-class image segmentation using conditional random fields and global classification. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, page 817–824, New York, NY, USA, 2009. Association for Computing Machinery.

[77] H. Proença and L. A. Alexandre. The nice. i: noisy iris challenge evaluation-part i. In *2007 First IEEE International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–4. IEEE, 2007.

[78] H. Proença and L. A. Alexandre. Introduction to the special issue on the segmentation of visible wavelength iris images captured at-a-distance and on-the-move. *Image and Vision Computing*, 28(2):213–214, 2010.

[79] H. Proenca and L. A. Alexandre. Toward covert iris biometric recognition: Experimental results from the nice contests. *IEEE Transactions on Information Forensics and Security*, 7(2):798–808, 2011.

[80] H. Proenca, S. Filipe, R. Santos, J. Oliveira, and L. A. Alexandre. The ubiris. v2: A database of visible wavelength iris images captured on-the-move and at-a-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8):1529–1535, 2009.

[81] H. Proença and L. Alexandre. UBIRIS: A noisy iris image database. In *13th International Conference on Image Analysis and Processing - ICIAP 2005*, volume LNCS 3617, pages 970–977, Cagliari, Italy, September 2005. Springer.

[82] S. Pundlik, D. Woodard, and S. Birchfield. Non-ideal iris segmentation using graph cuts. pages 1–6, 06 2008.

[83] A. Radman, K. Jumari, and N. Zainal. Fast and reliable iris segmentation algorithm. *IET Image Processing*, 7(1):42–49, 2013.

[84] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 91–99. Curran Associates, Inc., 2015.

[85] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[86] A. Ross. Iris as a forensic modality: The path forward. *URL http://www. nist. gov/forensics/upload/Ross-Presentation. pdf*, 2009.

[87] P. Rot, Ž. Emeršič, V. Struc, and P. Peer. Deep multi-class eye segmentation for ocular biometrics. In *2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI)*, pages 1–8. IEEE, 2018.

[88] P. Rot, M. Vitek, K. Grm, Ž. Emeršič, P. Peer, and V. Štruc. Deep sclera segmentation and recognition. In *Handbook of Vascular Biometrics*, pages 395–432. Springer, 2020.

[89] R. Rothe, R. Timofte, and L. Van Gool. Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision*, 126(2-4):144–157, 2018.

[90] K. Roy, P. Bhattacharya, and C. Y. Suen. Towards nonideal iris recognition based on level set method, genetic algorithms and adaptive asymmetrical svms. *Eng. Appl. Artif. Intell.*, 24(3):458–475, Apr. 2011.

[91] S. Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.

[92] C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. 300 faces in-the-wild challenge: Database and results. *Image and vision computing*, 47:3–18, 2016.

[93] W. Sankowski, K. Grabowski, M. Napieralska, M. Zubert, and A. Napieralski. Reliable algorithm for iris segmentation in eye image. *Image and vision computing*, 28(2):231–237, 2010.

[94] G. Santos, E. Grancho, M. V. Bernardo, and P. T. Fiadeiro. Fusing iris and periocular information for cross-sensor recognition. *Pattern Recognition Letters*, 57:52–59, 2015.

[95] M. Sardar, S. Mitra, and B. U. Shankar. Iris localization using rough entropy and csa: A soft computing approach. *Applied Soft Computing*, 67:61–69, 06 2018.

[96] M. Schuckers. Some statistical aspects of biometric identification device performance. *Stats Magazine*, 3, 2001.

[97] A. F. Sequeira, J. C. Monteiro, A. Rebelo, and H. P. Oliveira. Mobbio: a multimodal database captured with a portable handheld device. In *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*, volume 3, pages 133–139. IEEE, 2014.

[98] J. Shen, S. Zafeiriou, G. G. Chrysos, J. Kossaifi, G. Tzimiropoulos, and M. Pantic. The first facial landmark tracking in-the-wild challenge: Benchmark and results. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 50–58, 2015.

[99] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2015.

[100] B. A. Smith, Q. Yin, S. K. Feiner, and S. K. Nayar. Gaze locking: passive eye contact detection for human-object interaction. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*, pages 271–280, 2013.

[101] C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015.

[102] T. Tieleman and G. Hinton. Rmsprop gradient optimization. *URL http://www. cs. toronto. edu/tijmen/csc321/slides/lecture_ slides_ lec6. pdf*, 2014.

[103] K. Tripathi. A comparative study of biometric technologies with reference to human interface. *International Journal of Computer Applications*, 14(5):10–15, 2011.

[104] M. Vitek, P. Rot, V. Štruc, and P. Peer. A comprehensive investigation into sclera biometrics: a novel dataset and performance study. *Neural Computing and Applications*, pages 1–15, 2020.

[105] 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 Transactions on Information Forensics and Security*, 15:2944–2959, 2020.

[106] C. Wang, Y. Wang, B. Xu, Y. He, Z. Dong, and Z. Sun. A lightweight multi-label segmentation network for mobile iris biometrics. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1006–1010, 2020.

[107] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, and B. Xiao. Deep high-resolution representation learning for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2020.

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

[109] L. A. Zanlorensi, R. Laroca, E. Luz, A. S. Britto Jr, L. S. Oliveira, and D. Menotti. Ocular recognition databases and competitions: A survey. *arXiv preprint arXiv:1911.09646*, 2019.

[110] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang. Unet++: A nested u-net architecture for medical image segmentation. In D. Stoyanov, Z. Taylor, G. Carneiro, T. Syeda-Mahmood, A. Martel, L. Maier-Hein, J. M. R. Tavares, A. Bradley, J. P. Papa, V. Belagiannis, J. C. Nascimento, Z. Lu, S. Conjeti, M. Moradi, H. Greenspan, and A. Madabhushi, editors, *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 3–11, Cham, 2018. Springer International Publishing.