IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 20, 2021, accepted July 24, 2021, date of publication July 30, 2021, date of current version August 11, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3101256

Semantic Segmentation of Periocular Near-Infra-Red Eye Images Under Alcohol Effects

JUAN E. TAPIA[1,2], (Graduate Student Member, IEEE), ENRIQUE LOPEZ DROGUETT[3,4], ANDRES VALENZUELA[1], DANIEL P. BENALCAZAR[1], (Member, IEEE), LEONARDO CAUSA[5], AND CHRISTOPH BUSCH[2], (Senior Member, IEEE)

[1]Department of Mechanical Engineering, Universidad de Chile, Santiago 8370459, Chile
[2]da/sec-Biometrics and Internet Security Research Group, Hochschule Darmstadt, 64295 Darmstadt, Germany
[3]Department of Civil and Environmental Engineering, University of California at Los Angeles, Los Angeles, CA 90095, USA
[4]Garrick Institute for the Risk Sciences, University of California at Los Angeles, Los Angeles, CA 90095, USA
[5]R+D Center, TOC Biometrics, Santiago 7510099, Chile

Corresponding author: Juan E. Tapia (jtapiafarias@ing.uchile.cl)

This work was supported in part by the Agencia Nacional de Investigación y Desarrollo (ANID) by the FONDEF IDEA under Grant ID19I10118, in part by German Federal Ministry of Education and Research, in part by Hessen State Ministry for Higher Education, Research and the Arts, and in part by the National Research Center for Applied Cybersecurity ATHENE.

ABSTRACT This paper proposes a new framework to detect, segment, and estimate the localization of the eyes from a periocular Near-Infra-Red iris image under alcohol consumption. This stage will take part in the final solution to measure the fitness for duty. Fitness systems allow us to determine whether a person is physically or psychologically able to perform their tasks. Our segmentation framework is based on an object detector trained from scratch to detect both eyes from a single image. Then, two efficient networks were used for semantic segmentation; a Criss-Cross attention network and DenseNet10, with only 122,514 and 210,732 parameters, respectively. These networks can find the pupil, iris, and sclera. In the end, the binary output eye mask is used for pupil and iris diameter estimation with high precision. Five state-of-the-art algorithms were used for this purpose. A mixed proposal reached the best results. A second contribution is establishing an alcohol behavior curve to detect the alcohol presence utilizing a stream of images captured from an iris instance. Also, a manually labeled database with more than 20k images was created. Our best method obtains a mean Intersection-over-Union of 94.54% with DenseNet10 with only 210,732 parameters and an error of only 1-pixel on average.

INDEX TERMS Biometrics, fitness for duty, segmentation, iris, alcohol.

I. INTRODUCTION

The concept of biometrics is the set of intrinsical and behavioral characteristics that can be used to verify the identity of an individual. All human beings have unique morphological characteristics that differentiate and identify us, the shape of the face, the geometry of parts of our body, like hands, our eyes, and perhaps the best known, the fingerprint [1], [2].

Iris recognition systems have been used mainly to recognize the cooperative subjects in controlled environments for borderline demographics and to gain access to secure areas using near-infra-red capture devices [2]–[4]. With the improvements in iris performance and reduction in the cost of iris acquisition devices, the technology will witness broader applications and may be confronted with newer challenges. One kind of this new challenge is identifying if a captured subject is under alcohol, drug effects, or even in sleep deprivation and sleep restriction conditions. This area is known as "Fitness for Duty" [5], [6] and allows us to determine whether the person is physically or psychologically able to perform their task. [6]. This system delivers an answer based on the threshold of "fit" and "unfit" statistical people behavior threshold. It does not have a relationship with the automatic alcohol test that measures alcohol percentage on blood.

In state of the art, various performance tests have been proposed as FFD, including psycho-motor tasks, temperature sensor, electroencephalography, finger tapping, smart band wrist, in-cab monitoring, among others [7], [8].

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

It has been reported sensitivity problems to detect various job-related impairments. Also, some critical issues have been shown related to the type of device response and the possibility of impersonation. For instance, the wrist smartband can be shared with another partner or removed to avoid identification. Therefore, including an identification process is crucial to detect the FFD for each person. Biometrics modalities such as iris may help to identify the worker without removing the personal protection devices.

Several stages are required to build a new FFD device based on iris behavior to study alcohol's reaction, one of them is the semantic segmentation explored in this paper (etp). The whole process consists of the followings steps:

- Periocular NIR image acquisition(etp).
- Eyes detection (etp).
- Iris and pupil segmentation (etp).
- Iris recognition.
- FFD classification (fit/unfit).

In order to create a framework to understand the iris behavior and segment iris under alcohol presence, many algorithms are necessary to capture N frames, detect both eyes, and segment the images to localize the pupil and the iris to measure the changes over time. To segment this image is not a trivial task because the method needs to be efficient in the number of parameters to be implemented in a regular iris sensor. Most of these sensors are mobile devices self-integrated with a limited size of memory. See Figure 2.

Recently, automatic pupil segmentation is attracting many researchers to find the precise measurement of the pupil radii to apply to biometric analytics as part of medical, entertainment applications, virtual reality lenses, among others [9], [10]. In the biomedical field, there is a vital requirement in developing precise and automatic segmentation systems to capture saccade velocity, latency, the diameter of the iris, and pupil [11]–[13]. In this context, pupil measurement has been used to assess several cognitive functions, including fatigue, depression, and others. [14]. The iris under alcohol shows very large or small pupil sizes compared to the traditional iris in average condition. Therefore, the default parameters segmenter are not valuable for a proper segmentation of the eye under alcohol effects.

The paper aims to develop **an efficient framework to locate and segment the iris and pupil in multiple frames in subjects with and without alcohol consumption**. This work is ongoing research whose main objective is to estimate and extract the most relevant metrics from the iris and pupil to identify and know the behavior changes in the iris under alcohol influences. See Figure 1.

The iris alcohol segmentation present the following challenges:

- The people react differently with the same quantity of alcohol; some present pupil dilation, and other pupils constrict.
- The average constriction size of the pupil in alcohol presence is over the normal ranges. These changes do not allow used parametrical segmentation methods such as Osiris or commercial software.
- Most of the time, people present semi-closed eyes. This feature is an extra difficulty.
- In the presence of alcohol, the volunteer in front of the sensor shows an involuntary disbalance. This adds blurring to the captured images.

The article is organized as follows: Section II summarizes the related works on alcohol and semantic segmentation. The database is explained in Section III. The Eye detection methods proposed are in Section IV. The Iris and pupil localization methods are in Section V. The experimental framework and results are presented in Section VI, Ablation study is presented in Section VII and we conclude the article in Section VIII.

II. RELATED WORK

A. ALCOHOL DETECTION

The influence of alcohol, in particular in iris recognition, was reported by Arora *et al.* [13]. They presented a preliminary study of the impact of alcohol on an Iris recognition system. The experiments were performed on the 'Iris Under Alcohol Influence' database. Results show that when comparing pre and post-alcohol consumption images, the overlap between mated and non-mated compare distance score distributions increases by approximately 20%. These results were obtained using a relatively small database (220 pre-alcohol and 220 post-alcohol images obtained from 55 subjects). The subjects consumed about 200 ml of alcohol (with a 42% concentration level) in approximately 15 minutes, and the images were captured 15-20 minutes after alcohol consumption. This work suggests that about one in five subjects under the influence of alcohol may be able to evade identification by iris recognition systems.

Czajka [11] applied Hough transform operating on directional image (estimation of an image gradient delivering both a gradient value an its direction). After modeling the pupil size, It was shown as function of number of frames, representing the dynamic of pupil. In this way the pupil dynamic is modeled extracting liveness features.

Kumar and Passi [15] used a series of steps for preprocessing eye images based on replacing pixels over thresholding, median, Gaussian filtering, and a Canny edge detector. Using these steps, they performed segmentation over the Casia v3 and IITD datasets. This dataset does not present alcohol examples.

Bernstein *et al.* [16] used spectrogram images of size 224×224 from audio wave-forms to identify the presence of alcohol with Convolutional Neural Networks (CNN) and wearable sensors. They used 80 training images (40 positive, 40 negative) and 20 test images (10 positive, 10 negative) and achieved a test accuracy of 60% with Caffenet and 60% with Alexnet.

Koukiou and Anastassopoulos [17] proposed the use of thermal images to identify individuals under the influence of

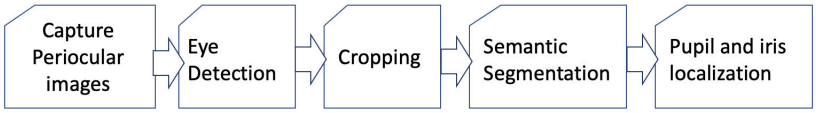

FIGURE 1. Block diagram of the proposed segmentation framework.

alcohol. They have shown that changes in the eye temperature distribution in intoxicated individuals can be detected using thermal imagery [18].

B. SEMANTIC SEGMENTATION

Semantic segmentation is the process of identifying each object in an image. This process is performed pixel by pixel to evaluate and assign a label to each pixel [19].

State-of-the-art algorithms such as semantic segmentation have been mainly trained to localize very complex objects from cities such as cars, buildings, and people and a few in biometric gaze applications. Even when we find more than one class object in the same image. Traditional pre-trained implementation models reach meager results when directly applied to eye segmentation and gaze estimation with alcohol presence. Another limitation is that most state of the art segmentation algorithms are based on deep convolutional networks with a large number of layers and parameters [20]–[22].

The U-Net [23] is a convolutional network architecture for the fast and precise segmentation of images. This network comprises two stages: the encoder (downsampling) and decoder (upsampling) paths using successive convolutional layers. The downsampling stage extracts the feature information from the images. The upsampling operators recover the image by replacing the pooling operators. Hence, these layers increase the resolution of the output. The high-resolution features from the path are localized and combined with the upsampled result. Successive convolution layers can then learn to assemble a more precise output based on this information. This U-Net network was improved in order to reduce the number of parameters using a pre-trained MobileNet network as encoder [24]. This kind of network has been used to perform semantic segmentation on medical images and iris.

Chen *et al.* [21] proposed a complex DeepLabv3+ as an extension of the previous DeepLabv3 by adding a simple yet effective encoder module to recover the object boundaries. The rich semantic information is encoded in the output of DeepLabv3 with atrous convolution allowing to control the density of the encoder features depending on a budget of computational resources. Furthermore, the decoder module allows detailed object boundary recovery. DeepLabv3+ is a cutting-edge architecture for semantic segmentation. This architecture can do multi-scale processing without increasing the number of parameters. DeepLabv3+ adds an intermediate decoder module on top. After processing the information via DeepLabv3+, the features are then up-sampled N times. This network improves the data load from the end of the network and provides a shortcut path from the feature extraction front-end to the near future of the network. This kind of network has been used to perform semantic segmentation on Virtual reality lens.

Badrinarayanan *et al.* [20] proposed the first Encoder-Decoder architecture for segmentation tasks called SegNet. SegNet has three main blocks: encoder, hidden vector, and decoder Network, followed by a pixel-wise classification layer. The encoder part will convert the input image into a single-dimensional vector (hidden vector). The decoder network will convert the hidden vector into the corresponding semantic segmentation. This kind of CNN was used to perform sclera segmentation on Multi-Angle Sclera Database (MASD) [25]. Hassan *et al.* [26] used the SegNet to segment pupil, iris, and sclera over several databases.

The Fully Convolutional Neural Networks (FCNN) are the natural evolution of the CNN for segmentation tasks [27], [28]. This kind of neural network delivers an image of equal size as the input image containing the segmented classes. To reach that goal, this adds an upsampling layer. Bezerra *et al.* [28] perform iris segmentation used the following databases: Casia v4, IITD, CrEye-Iris.

The DeepVOG [29] is a convolutional layer with 10×10 filters which output feature maps with the same size of the input by appropriate padding. The down-sampling path reduces the size of the feature maps and increases the size of receptive fields of convolutional filters each stage, such that more complex features in a larger context can be extracted. DeepVOG was developed as a prerequisite for many eye-tracking and Video-OculoGraphy (VOG) methods for accurate pupil localization. DeepVOG is based on Fully Convolutional Neural Network (FCNN). The output simultaneously enables to perform pupil center localization, elliptical contour estimation, and blink detection, all with a single network and with an assigned confidence value, at frame rates above 130 Hz on commercial workstations with GPU acceleration. Pupil center coordinates can be estimated with a median accuracy of around 1.0 pixel, and gaze estimation is accurate to within 0.5 degrees.

Valenzuela *et al.* [19] proposed an efficient DenseNet based on DenseNet56 and compared several implementation in number of parameters, scores and complexity using DeepLabv3, UNet, Mask-RCNN, DenseNet-56, DenseNet101 and, DenseNet10. These models were trained using the OpenEDS database.[1] The goal was to achieve an accurate and efficient segmentation algorithm for NIR eye Images taken from Virtual Reality (VR) lenses. Models with a

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

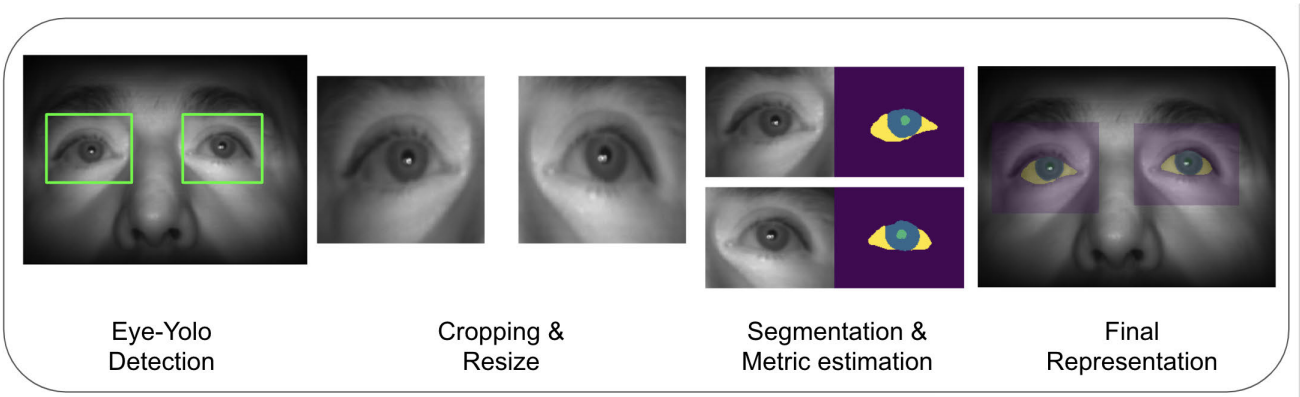

FIGURE 2. Graphics demonstration of all the stages used in the proposed framework method based on DenseNet10 and Eye-tiny-yolo.

low number of parameters were obtained. No alcohol images were used in this work. As a continuation of this work, this current paper is based on semantic segmentation of NIR under alcohol effect. The problem that we try to solve is different, and the previous version of DenseNet10 did not perform well with alcohol images. The new version of DenseNet10 used in this paper was trained from scratch with a new module of three levels of Data-Augmentation (basic, medium, and aggressive). Three layers of Transition down (TD) and three layers of Transition up (TU) were used. A new set-up value of K (grown rate factor) was selected $K = 5$. The best results were reached with an aggressive level of Data-Augmentation. The best parameters are described in the append section, Tables 6, 7 and 8.

Huang *et al.* [30], [31] proposed a semantic model of full-image dependencies over local feature representations using light-weight computation and memory. They introduce a criss-cross attention module (CCNet). The CCNet collects contextual information in horizontal and vertical directions to enhance pixel-wise representative capability. The CCNet can harvest the contextual information of its surrounding pixels on the criss-cross path through a novel criss-cross attention module for each pixel. By taking a further recurrent operation, each pixel can finally capture the long-range dependencies from all pixels.

It is essential to point out that previous work did not report results with alcohol presence.

III. DATABASES

For this paper, a new database of iris images under alcohol presence was created. Traditional databases available to test iris segmentation have not present images with iris captured under alcohol consumption on NIR or Visual Spectrum [32]. This new database was captured using two different iris NIR sensors: Iritech Gemini, and Iritech Venus.[2] Each image has a size of 1280 × 760 pixels. Example of images are shown in Figure 3.

Two-hundred and sixty-six subjects were included in the experiments under alcohol effect and 765 subjects with no

[2]https://www.iritech.com/products/hardware/gemini-camera

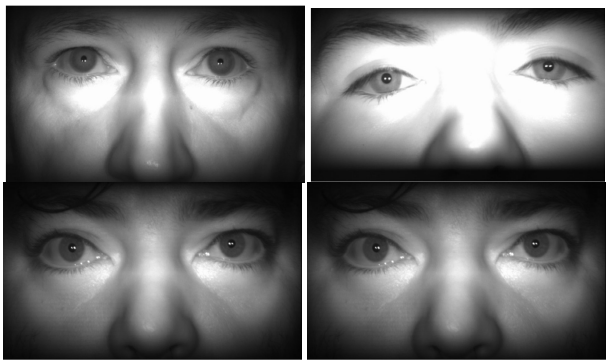

FIGURE 3. Example of binocular periocular images taken directly from NIR sensor at 30 cm.

TABLE 1. Database description. Time is reported in minutes.

Session	Condition	Time	Alcohol
S0	Pre-alcohol	0	0
S1	Post-alcohol	15	200 ml
S2	Post-alcohol	30	0
S3	Post-alcohol	45	0
S4	Post-alcohol	60	0

alcohol. Each volunteer was requested to step on a floor mark (30 cm from the camera for Gemini and 50 cm for Venus Iritech capture devices) and to look at the NIR capture device after having consumed 200 ml of alcohol according to the protocol explained as follows:

The data capture process was organized in 5 sessions, according to the following protocol:

The room temperature and lighting were controlled and kept constant during the data capturing process. A total of 100 images(average) per eye were captured for each individual per session.

Overall a total of 21,315 subject-disjoint images were used to train our proposed method. All the images were manually labeled using the VIA tools [33]. For left and right eyes, pupil, iris, and sclera were labeled. This was a very demanding and time-consuming process that took over one year. See Figure 4.

After the eye detector was applied to separate left and right eyes, the images were divided into Train (70%), Validation (20%), and Test (10%). That is 14,918 (alcohol:

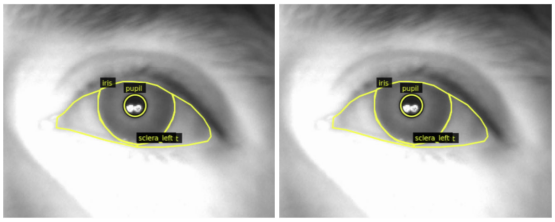

FIGURE 4. Example images with both eyes manually labeled.

TABLE 2. Division database on train, validation test subsets.

Class	Train (70%)	Val. (20%)	Test (10%)
Alcohol	13,126	3,735	1,874
Non Alcohol	1,790	528	256
Total	14,918	4,263	2,134

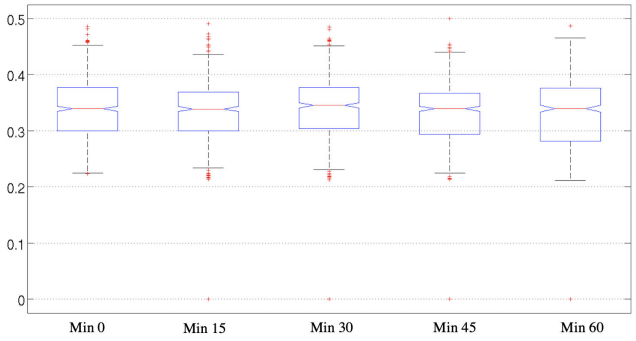

FIGURE 5. Y-axis shows Pupil-Iris Ratio for each session separately. X-axis show the time consumption of alcohol. From left to right: 0, 15, 30, 45 and 60 minutes.

13,126, and 1,790 No_alcohol), 4,263 (alcohol: 3.725, and 528 No_alcohol), 2,134 (alcohol: 1,874, and 256 No_alcohol) images respectively. The subset separation takes into consideration the amount of real levels of alcohol and non-alcohol cases in a working environment. On average, 10% of workers present some level of alcohol consumption. See Table 2.

The manually labeled pupils contain pupil radii ranging from 7 to 18 pixels with an average radio of 9 pixels, making both databases complicated for performing segmentation over the pupil.

Figure 5 shows a statistical representation of each group present in the dataset according to pupil-iris ratio. In this database, no gender analysis was performed.

As we mentioned before, a stream of images containing both irises were captured - 100 frames in average. This process, which takes five seconds, was repeated for all the participants. During that time, the aperture of the subject's iris adjusts to NIR light changes of the sensor's LEDs. This velocity change, measured across all the frames, helps estimate alcohol's influence on the individuals since alcohol directly affects the velocity of iris adjustment to direct light.

A. DATA AUGMENTATION

An aggressive Data Augmentation (DA) was performed using the imgaug library [34]. DA was applied using non-geometrical transformations organized on three groups: Common, Noise/Blur, Image corruption. The parameters for each modification are reported in the append section. Figure 6 shows examples of the output images.

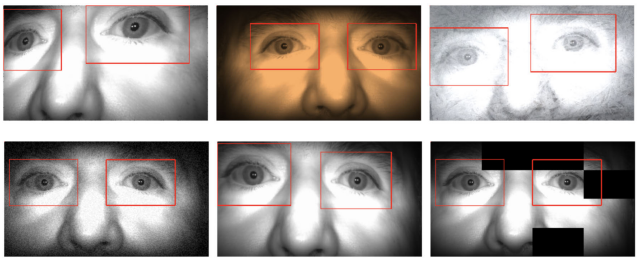

FIGURE 6. Example of aggressive data augmentation from semantic periocular NIR. In red the eye-yolo detection.

IV. EYE DETECTION IN PERIOCULAR IMAGES

The capture process is cooperative and delivers periocular images with both eyes on it. The images are usually not centralized because of the effect the alcohol had for the volunteer. On average, one hundred frames are captured per subject. This capture process takes five seconds. Then, an eye detector was implemented in order to find both eyes in the images and crop them to be segmented in a later step. A traditional eye detector based on haar-cascade [35] was evaluated as a baseline and discarded because it presents a deficient performance (78,60% of accuracy) and also needs a left and right eye version. These results are not reliable for our purpose. In the end, a new version of eye detector based on tiny-yolo was trained to detect simultaneously both eyes with high precision, called Eye-tiny-yolo. This new version can detect the eyes on periocular NIR images.

A. EYE DETECTORS

You Only Look Once (YOLO) [36] is a state-of-the-art detector and predicts multiple bounding boxes per grid cell. At training time, we only want one bounding box predictor to be responsible for each object. We assign one predictor to be "responsible" for predicting an object-based (Eye) prediction with the highest current Intersection Over Union with the ground truth. This metric leads to specialization between the bounding box predictors. Each predictor better predicts specific sizes, aspect ratios, or object classes, improving overall recall. However, the pre-trained tiny-yolo did not perform well for our proposed method based on periocular NIR eyes images. This result is because no eye images are used while training it. Therefore, regular tiny-yolo implementation was retrained from scratch using images of size 416×416, the Learning rate of $1e-3$, Batch size of 32, Optimizer Adam, custom loss function, and Data-Augmentation (DA). A dataset of 1,400 periocular images (both eyes are present) was divided into 1,120 for the train and 380 for the test. All the images are manually labeled. The performance improved substantially; the two eyes are detected simultaneously in bounding boxes of different sizes. Therefore, they are all resized to 320×320 as an input to the segmentation stage. This new version can detect both NIR eyes simultaneously and is called "Eye-tiny-yolo".

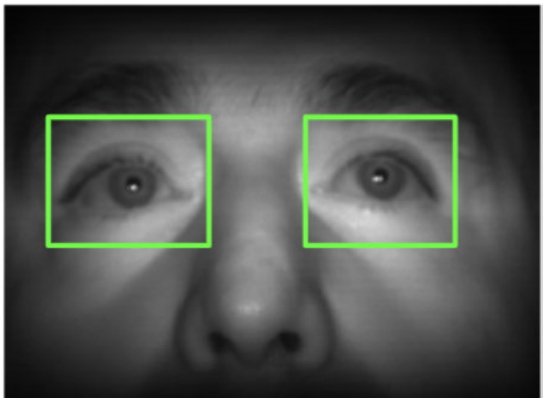

FIGURE 7. Example of Eye-tiny-yolo detection with both eyes automatically detected. The nostrils also represent a challenge in this image because the segmenter could have confused it with pupils. The green squares show a correct detection of the left and right eyes.

Figure 7 shows a high confidence detection of both eyes in the same picture simultaneously. It is not unnecessary to pass the image two times to detect the left and right eye. This method reached 0.9860 +/− 0.021 of IoU and standard deviation. See Table 3.

Table 3 shows the performance of Eye-tiny-yolo with different thresholds. The state-of-the-art value reported is a threshold of 0.50. The performance of Eye-tiny-yolo was evaluated by comparing the coordinates (x,y) of the ground truth for each eye and the detection coordinates of the Eye-tiny-yolo with several thresholds as is shown in Table 3. For all the segmenters, DeepVOGv2, DenseNet10, and CCNet, the intersection Over Union (IOU) was reported according to the state of the art.

V. IRIS AND PUPIL LOCALIZATION

Once both eyes are isolated, we employ convolutional neural networks to find a mask that segments the iris and the pupil. The networks used in this paper are described in the Section VI. However, those networks only output a mask that highlights pixels that belong to the iris and the pupil in the images. In order to use this information, we employ pupil and iris localization algorithms, which find the centers and the radii of the circles that best adapt the pupil and iris contours, using the mask as input. The ratio between the pupil and iris radii is essential to assert alcohol consumption [13]. This section describes the localization algorithms that were used in this study as follows:

A. MASS CENTER

This localization method is straightforward and effective. First, a binary image of the pupil and iris is complemented using the *XOR* operation. As a result, a contour area is obtained for the pupil and iris. Afterward, the most significant area is filled in order to search and estimate the boundaries of the iris and the pupil. This method explores the vertical and horizontal boundaries (edges) of the pixels. With this information, we calculate the radii of the pupil and the iris separately, as shown in Figure 8.

TABLE 3. Performance of Eye-tiny-yolo with different threshold levels and the standard deviation (Std).

Threshold	IoU	Std
0.50	0.996	0.033
0.75	0.991	0.023
0.85	0.986	0.021
0.95	0.959	0.019

B. LEAST-MEAN-SQUARES BASED SEGMENTATION

This localization algorithm is conceived to be as lightweight as possible. A binary and morphological operations [37] were used instead of computationally expensive algorithms that have been applied in traditional approaches, such as Canny edge detection, the Hough transform, and RANSAC [2], [38]–[41]. First, A single-channel prediction of the iris region is computed. Then, it was isolated the iris region employing a hole-filling operation [37], which removes the pupil. The pupil is then separated by applying a *XOR* operation between the raw prediction and the iris region. After that, the contour of the iris and the pupil regions was obtained. For this purpose, the *XOR* operation between the binary image and the erosion of the same binary image was used, thus recovering only the removed pixels during the erosion operation. In the next step, horizontal lines from the iris contour, using Sobel filters [37] were removed to eliminate eyelids and eyelashes. Finally, the coordinates of the contour pixels and estimation of the best fitting circle using a Least-Mean-Squares (LMS) algorithm [37] was calculated. As a result, the coordinates of the center and radius of both pupil and iris are obtained. See Figure 5.

C. MIXED APPROACH

The ''Mixed Algorithm'' is a combination of the previous methods (Mass Center and LMS). First, the erosion algorithm is applied to find boundaries. After that, the Mass Center is used to find the pupil circle. If the pupil circle was found correctly, the LMS algorithm is used to find the iris circle. However, when eyelids intersect the pupil, the Mass-Center algorithm fails to report a pupil. In such cases, the Hough transform is applied to find both pupil and iris circles. The combination of these three algorithms allows us to improve the results.

VI. EXPERIMENTS AND RESULTS

A. METRICS

This section describes all the metrics used to evaluate the whole framework: eye detection, segmentation stage, localization of the iris and pupils. Figure 2 shows the results of the framework with all stages applied.

B. INTERSECTION OVER UNION

In order to evaluate Eye-tiny-yolo detector, the Intersection over Union (IoU) metric was used. The IoU measures the overlap between two boundary images. This is used to measure how much the boundary predicted by the algorithm

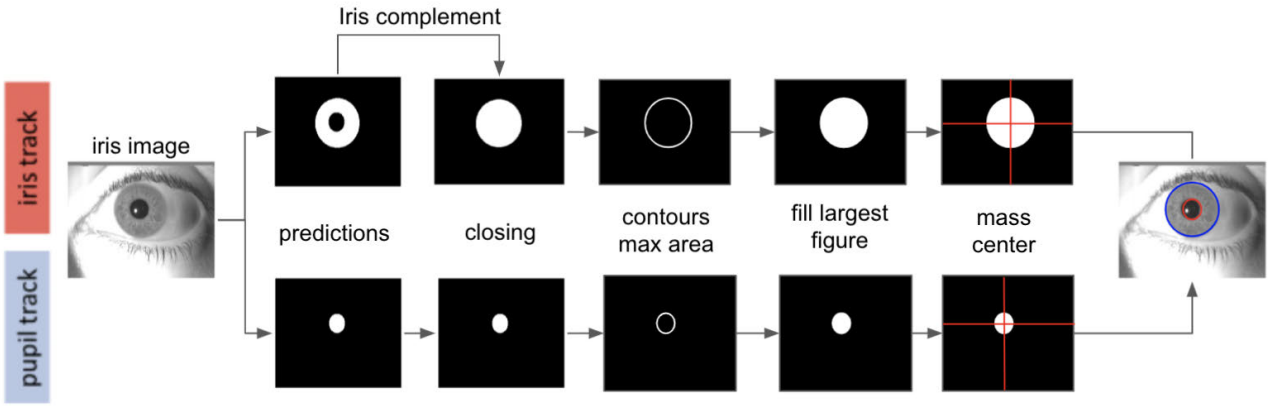

FIGURE 8. Mass center framework implementation. letters a) Represents the Iris track and b) Represent Pupil track.

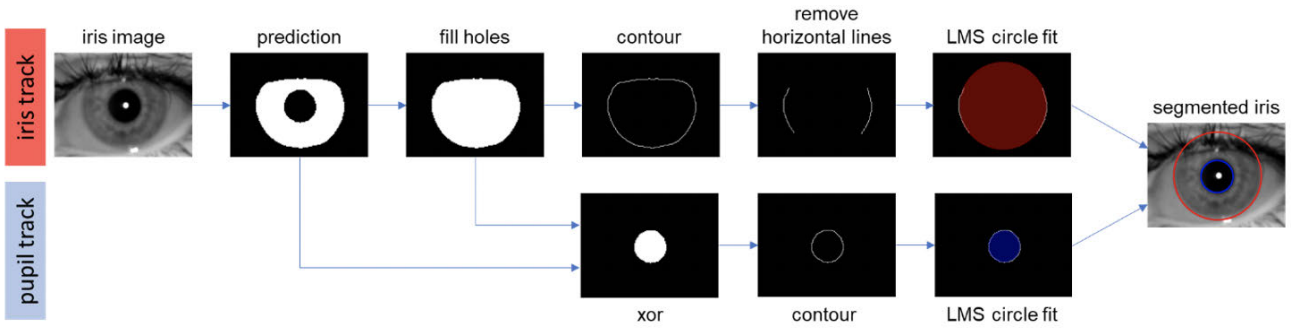

FIGURE 9. Least-Mean-Squares framework implementation.

overlaps with the ground truth (the real object). Traditionally, state-of-the-art datasets reported an IoU threshold equal to or greater than 0.5. Figure 10 show the graphical effect to apply IoU. The IoU is calculated using the following equation:

$$IoU = \frac{A \cap B}{A \cup B} \tag{1}$$

On the other hand, for the evaluation of the segmentation networks we use the bit-wise version of IoU. This IoU assesses how much of the predicted binary mask correlates with the ground-truth mask pixel by pixel. The following equation is used for its calculation:

$$IoU = \frac{\sum and(A, B)}{\sum or(A, B) + c}, \tag{2}$$

where A and B are binary images, and c is a small constant that prevents the IoU from taking an infinite value when there is no overlap between A and B.

C. EXPERIMENT 1:OSIRIS

As a baseline, the Osiris software was used to segment the pupil and the iris from alcohol images. However, the resulting segmentation performance was low. The alcohol images show the size of pupil very large or small compared to traditional images in normal condition. Therefore, the default parameters are not valuable for the proper segmentation of the image. Figure 11 Show examples of wrong segmentation for alcohol images. The method fails in both pupil and iris localization.

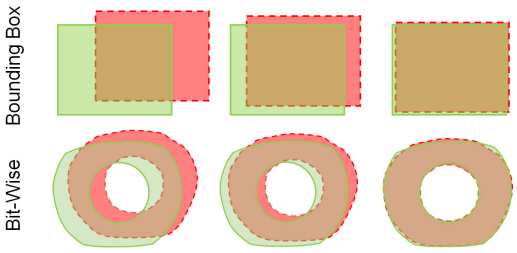

FIGURE 10. A visual example of intersection over union and bitwise quality metric. The dashed red box is the predicted detection, the continuous green box is the ground truth, and the gray area is the overlap between the two. The example shows three different IOU scores from left to the right, with the rightmost being the best. Based on [42].

D. EXPERIMENT 2:DeepVOG

Two experiments were developed for DeepVOG. For the first one, pre-trained models were used. For the second one, a DeepVOGv2 was trained from scratch. DeepVog is only focused on gaze estimation and pupil estimation. The input images of the original version used video with a frame rate above 130 Hz of 320 × 320 × 3 image size. Our DeepVOGv2 implementation can segment four classes: Iris, the pupil, left and right sclera. DeepVOGv2 was trained from scratch with a low frame rate of images (less than 20 fps) using alcohol presence images of $640 \times 480 \times 3$ pixels, Adam optimizer with a learning rate of 1e-5 during 400 epochs. A batch size equals 8 with a step size of 64 and an input shape of 320 × 240 × 3. An additional convolutional layer with

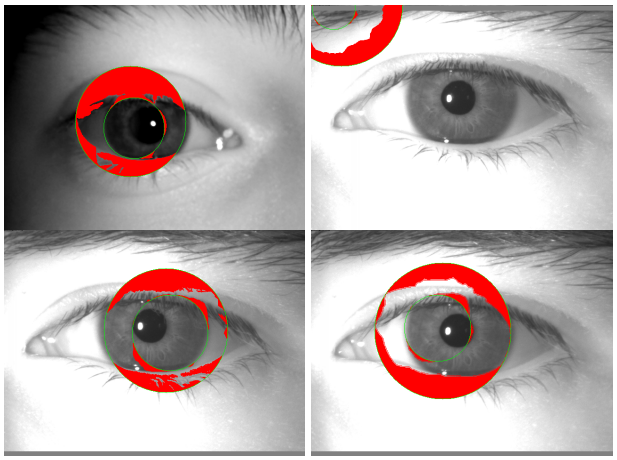

FIGURE 11. Example of wrong segmentation under presence of alcohol for traditional Osiris software. The green circle represents the iris boundaries. Red color represents the mask area detected by Osiris. The software masks all the areas not belonging to the iris texture(such as highlights, eyelish, eyelashes).

TABLE 4. Summary of the segmentation methods under alcohol influences. TFL: represent transfers learning models.

Method	Mean IoU	Std	Nº Param
Osiris	0.6336	0.2780	N/A
DeepVOG	0.9003	0.0710	2,058,979
DeepVOGv2	0.9130	0.0170	2,058,979
DenseNet10	0.9458	0.0160	210,732
DenseNet10 TFL-1	0.9055	0.0167	156,235
MobileNet-UNet	0.9297	0.0307	5,105,924
MobileNet-UNet (TFL-2)	0.9290	0.0275	6,968,580
CCNET	0.9190	0.0400	122.514
CCNET (TFL-3)	0.9080	0.0490	85,760

a filter of 3 × 3 and Tversky Loss was added to generalize the loss function. The Tversky index addresses data imbalance and achieves a better trade-off between precision and recall in iris images of convolutional deep neural networks. In addition, we specify our camera parameters such as focal length because we used a different capture device with a lower resolution and fps.

The DeepVOGv2 was also tested with the same database to check their performance under a complex scenario (very low pupil size radio). As we mentioned before, the database was manually labeled. The coordinate allows us to crop the eyes from the periocular images manually or automatically using Eye-tiny-yolo.

For the manually cropped images, they were cropped, splitting the 1280×760 image into two images of 640×480 pixels; each one of them has one eye. The second cropping was made using Eye-tiny-yolo. The output image was resized to its original size and placed again in a black image of 1280 × 760 to compute the IoU. Table 4 shows the results for DeepVOG and DeepVOGv2.

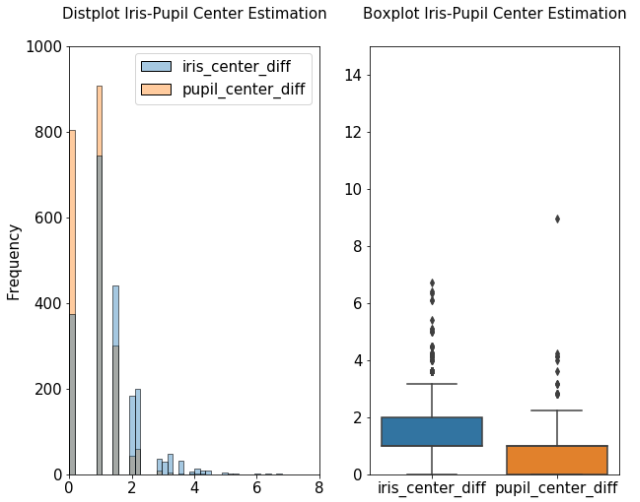

FIGURE 12. Histogram of the distribution error for the pupil and the iris center localization using DenseNet10 for the test dataset. Left. Show a histogram of segmentation results with the error in pixels. Right. Box-plot to describe the average error in pixels from pupil and iris centers.

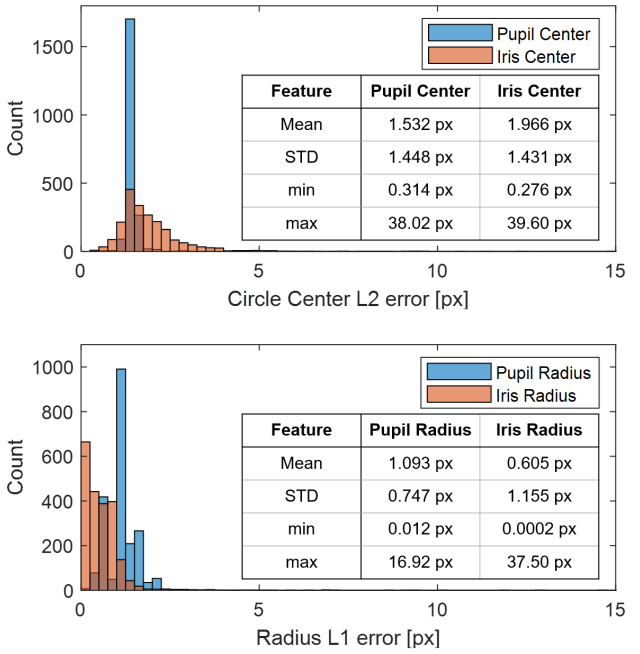

FIGURE 13. Histogram of the distribution error for pupil and iris center localization using CCNET for the test dataset.

E. EXPERIMENT 3:DenseNet10

One of the goals of this work is to reduce the layers and the concatenation matrix's size from the original implementation of DenseNet56 and DenseNet101. For doing so, a feature extractor and two paths (Down-sampling and one Up-sampling) were modified. The down-sampling path has 1 Transition Down (TD), and an up-sampling course has 1 Transition Up (TU) instead of the four Transitions (2TD+2TU) used in the traditional approach. For each layer i, the number of feature maps k obtained is given by the following equation: $k0 + k \times (i - 1)$, where $k0$ is the number of channels in the input layer. In order to define the best number of filters k to use in each layer, a grid

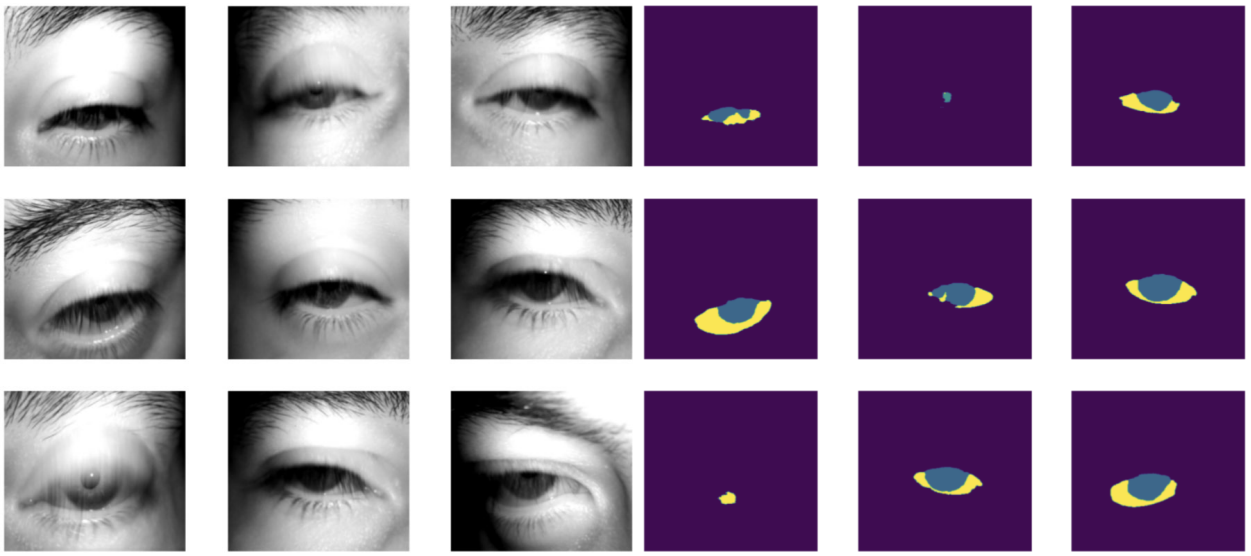

FIGURE 14. Example of semi-closed challenging images. Left original images under alcohol consumption. Right: DeepVOGv2 results.

search from $k = 3$ to $k = 15$ was used. The best result was obtained with $k = 5$. The following parameters were used to train our DenseNet10: Epochs: 200, Batch Size: 32, Optimizer: RMSprop, LR: $1e-4$, Decay: $1e-6$, Dropout: 0.15, with aggressive Data Augmentation and an Input shape: $320 \times 240 \times 3$. This model was trained and tested from scratch with a GPU-1080-ti with 32GB RAM.

Figure 12 shows a histogram with the distribution error for pupil and iris localization with images under alcohol presence using the DenseNet10 method. The left figure shows a histogram error rate of 1 pixel for the test set. The right figure shows the plot distribution for the iris and pupil.

F. EXPERIMENT 4:CCNet

Criss-Cross attention network (CCNet) was used in this approach to get an efficient model with a low number of parameters. Our model was trained from scratch with the presented dataset. The input image used is $360 \times 240 \times 1$. The best parameters used were: Batch size: 60, LR: $1e-3$, Epochs: 200. This network was trained and tested with a 6GB GTX-1660 GPU and 16 GB of RAM.

Figure 13 shows a histogram with the distribution error for pupil and iris localization with images under alcohol presence using CCNet model and the LMS algorithm.

Table 4 shows the comparison results among the Osiris as a baseline, DeepVOG, DeepVOGv2, DenseNet10, DenseNet10 with transfer learning (TFL-1), MobilNet-UNet, Mobil-UNet with transfer learning (TFL-2), CCNet, and CCNet with transfer-learning (TFL-3). The Osiris system performs over 63.4 % Mean IoU, while the Original DeepVOG obtained a 90.03%. DeepVOGv2 trained from scratch to reach 91.30%. A fine-tuning was applied to DenseNet10, MobileNet-UNet, and CCNet using pre-trained models in order to explore the efficiency and the ability to segment eyes. The results are competitive but do not improve the previous version trained from scratch; even the parameters numbers are reduced. The proposed method DenseNet10 trained under the database alcohol presents better results 94,58 % of Mean IoU.

Table 5 shows the results for the five different metrics. Traditional Hough and Mass center and three proposed metrics to measure the center of the pupil and radii.

Figure 14 and Figure 15 show a set of challenging images with semi-closed eyes under alcohol effects. Left: The images in grayscale show semi-closed eyes where the pupil is not visible and presents some eyelashes occlusions. Figure 14 Right: figures show the results for DeepVOGv2. This method reaches good results, but it cannot deal very well with semi-closed eyes. Figure 15 right, the DenseNet10 method shows it can detect these semi-closed eyes with high precision.

G. EXPERIMENT 5: GRAND-MEAN

The grand mean [43] of a set of multiple subsamples is the mean of all observations: each data point, divided by the pooled sample size. This analysis type is used to analyze the differences in the data series acquired between different trials. The pooled mean allows to analyze the general behavior of the studied group and therefore provides guidelines of the ''waveform'' for each group. Grand mean is widely used in the processing of EEG signals, especially in cognitive evoked potentials, since it allows to obtain the fundamental component of the EEG signal using individual signals that by themselves can be very noisy [44]. In this way, the grand mean enables us to estimate the tendency. This tendency capacity was used to analyze the measures obtained from alcohol and no-alcohol images for each subject using the pupil and iris radii.

Figure 16 shows the population behavioral study as such a time sequence of 5 seconds of pupil size recordings. This analysis was conducted to determine whether there were differences in pupil radius size under alcohol and non-alcohol conditions.

TABLE 5. Comparison metrics for all the methods to estimate the localization of the iris and pupil using CCNet. The table shows the error in pixels and the standard deviation. The best results are highlighted in bold.

Metric	Hough	Mass Center	LMS	Mixed
Pupil Center (L2 error)	3.9 +/− 4.58	**1.43 +/− 0.11**	1.86 +/− 3.77	1.53 +/− 1.45
Iris Center (L2 error)	4.6 +/− 4.42	7.30 +/− 5.43	2.28 +/− 300	**1.96 +/− 1.43**
Pupil Radius on x (L1 error)	2.28 +/− 4.28	1.96 +/− 0.36	1.17 +/− 1.02	**1.09 +/− 0.75**
Pupil Radius on y (L1 error)	2.34 +/− 3.80	1.64 +/− 0.57	1.19 +/− 2.24	**1.12 +/− 1.99**
Iris Radius on x (L1 error)	2.67 +/− 4.09	3.03 +/− 1.12	0.72 +/− 1.50	**0.604 +/− 1.16**
Iris Radius on y (L1 error)	14.53 +/− 7.16	4.27 +/− 1.35	0.16 +/− 0.29	**0.157 +/− 0.289**

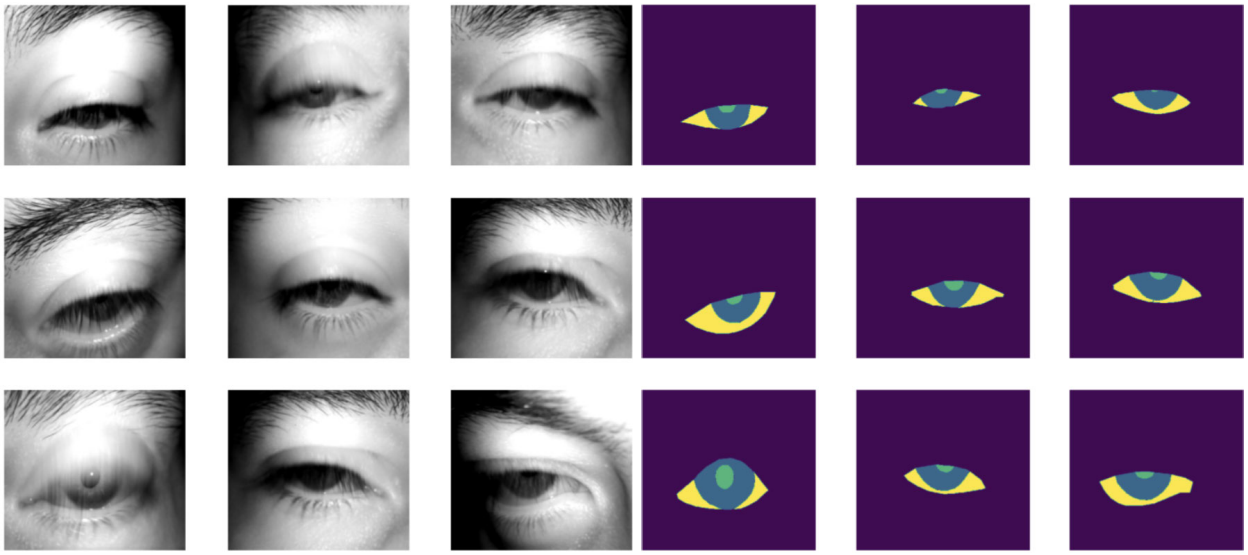

FIGURE 15. Example of semi-closed challenging images. Left original images under alcohol consumption. Right: Our proposed DenseNet10 results.

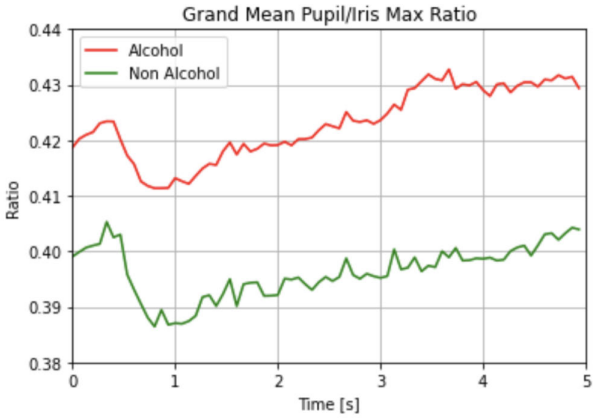

FIGURE 16. Grand mean curves of the alcohol (Red) and no alcohol (Green) population.

In order to obtain the alcohol and non-alcohol curves, the grand mean was estimated for the pupil radii at each time instant for all the subjects in each one of the groups. Thus, it is possible to define the baseline behavior curve for people in the presence and absence of alcohol consumption. The analysis shows a difference in pupil size's temporal behavior, with larger size (more significant dilation) in subjects under the effects of alcohol. It is important to note that this analysis shows the average behavior of the population, so it is not possible to use it as a single variable to separate both groups.

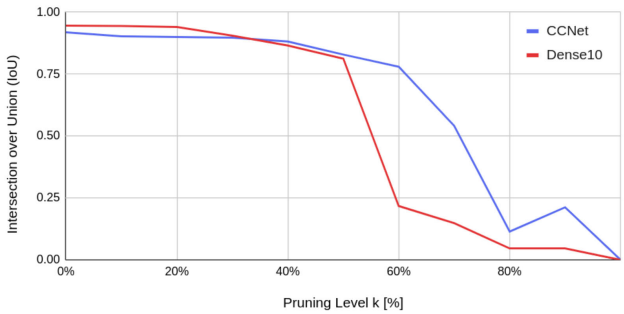

FIGURE 17. Performance analysis using pruning drop weights.

VII. ABLATION STUDY

In order to study the true impact of individually trained weights in the presented networks, we performed an Ablation study. For this study, we pruned an amount of k-percent of the best-trained network weights (CCNet and DenseNet10) and evaluated the IoU in the 2.134 images of the test set. The pruning level k was increased from 10% in steps of 10% up to 100% of the weights were pruned. For CCNet and DenseNet10, we employed the L1-unstructured pruning algorithm of PyTorch, which eliminates the k-percent of weights with the smallest activation values in the entire model. The results of both networks are shown in Fig. 17. For CCNet and DenseNet, up to 30% of the weights could be eliminated without experiencing a major drop in performance since the

TABLE 6. Aggressive-Common data augmentation parameters.

Common		
Augmentation Name	Probability (0.0 - 1.0)	Parameters
Rotation	0.5	degrees_range (-30, 30)
Zoom in-out	0.5	scale (0.75, 1.5)
Add (brightness)	1.0	value (-30, 30)
Translation axis x-y (%)	0.5	axis x-y (-0.5, 0.5)
Flip Up-Down / Left-Right	0.5	-

TABLE 7. Aggressive- Noise/Blur data augmentation parameters.

Noise / Blur		
Augmentation Name	Probability	Parameters
Additive Gaussian Noise	0.33	loc = (2.5, 25.5)
Additive Laplace Noise	0.33	loc = (2.5, 25.5)
Additive Poisson Noise	0.33	lam = (0.01, 16.0)
Gaussian Blur	0.1	sigma = (0.01, 2.5)
Average Blur	0.1	k = (1, 5)
Motion Blur	0.1	k = (3, 5)
Gamma Contrast	0.1	gamma = (0.01, 1.0)
Zoom Blur	0.1	severity = (1, 3)
Glass Blur	0.1	severity = (1, 3)
Defocus Blur	0.1	severity = (1, 3)

IoU would only decrease from 91,9% to 89,7% and 94.54% to 91.0%, respectively. Therefore, the significance of those weights is very low in the segmentation of iris images.

VIII. CONCLUSION

According to the results, it is possible to measure the alcohol presence and changes in the behavior using the frame of the eyes. The diameters of the pupil and iris present abnormal sizes that confuse traditional approaches and do not allow us to use a parametric method such as Osiris. The semi-closed eyes and eyelashes present a real challenge to efficiency and high accuracy about IoU. A mixed localization method is more suitable for the measure; nevertheless, we reach a high precision of less than one pixel. The number of parameters for the best semantic segmentation approach is also relevant because it allows us to implement this framework in a mobile device or commercial hardware. The proposed system involves two efficient methods: CCNet and DenseNet. DenseNet10 obtained a high score but a higher number of parameters in comparison to CCNet. The database capture for this project is also a significant contribution. This database

TABLE 8. Aggressive- Image corruption data augmentation parameters.

Image Corruption		
Augmentation Name	Probability	Parameters
Rain	0.25	speed (0.01, 0.1)
Frost	0.25	severity (1, 2)
Spatter	0.25	severity (1, 2)
Snowflakes	0.25	flake_size (0.01, 0.2) speed (0.01. 0.05)
Piecewise	0.33	scale (0.01, 0.05)
Coarse Dropout	0.35	p = (0.05, 0.25)
JPEG Compression	0.5	compression (90, 98)
Elastic Transformation	0.33	alpha = (0, 2.5) sigma = 0.25
Perspective Transformation	0.33	scale = (0.01, 0.2) keep size = True

will be available for a research community by the end of 2022. As future work, we need to study the feasibility of measuring fitness for duties using a regular NIR iris sensor. The results obtained show that it is possible to accomplish this goal.

APPEND

See Tables 6–8.

ACKNOWLEDGMENT

The database will be available for a research community by the end of 2021.

Further thanks to all the volunteers that participated in this research.

REFERENCES

[1] A. K. Jain, D. Deb, and J. J. Engelsma, "Biometrics: Trust, but verify," *CoRR*, vol. abs/2105.06625, pp. 1–20, May 2021. [Online]. Available: https://arxiv.org/abs/2105.06625

[2] J. Daugman, "How iris recognition works," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 21–30, Jan. 2004.

[3] K. W. Bowyer, E. Ortiz, and A. Sgroi, "Trial Somaliland voting register de-duplication using iris recognition," in *Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG)*, May 2015, pp. 1–8.

[4] J. Tapia, C. Perez, and K. Bowyer, "Gender classification from the same iris code used for recognition," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 8, pp. 1760–1770, Aug. 2016.

[5] A. J. MacQuarrie, C. Robertson, P. Micalos, J. Crane, R. High, E. Drinkwater, and J. Wickham, "Fit for duty: The health status of New South Wales paramedics," *Irish J. Paramedicine*, vol. 3, no. 2, pp. 1–10, Dec. 2018.

[6] S. E. Murphy and T. E. Fleming, "Fitness for duty in the nuclear power industry: The effects of local characteristics," in *Proc. Conf. Rec. 5th Conf. Hum. Factors Power Plants*, 1992, pp. 127–132.

[7] R. W. Allen, M. Silverman, and M. Itkonen, "Real world experience in fitness-for-duty testing," in *Proc. SAE Tech. Paper*, Oct. 1992, pp. 1527–1534, doi: 10.4271/921908.

[8] L. Peter, R. Reindl, S. Zauter, T. Hillemacher, and K. Richter, "Effectiveness of an online CBT-I intervention and a face-to-face treatment for shift work sleep disorder: A comparison of sleep diary data," *Int. J. Environ. Res. Public Health*, vol. 16, no. 17, p. 3081, Aug. 2019.

[9] M. Vitek *et al.*, "SSBC 2020: Sclera segmentation benchmarking competition in the mobile environment," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Sep. 2020, pp. 1–10.

[10] L. Alzubaidi, M. A. Fadhel, O. Al-Shamma, J. Zhang, J. Santamaría, Y. Duan, and S. R. Oleiwi, "Towards a better understanding of transfer learning for medical imaging: A case study," *Appl. Sci.*, vol. 10, no. 13, p. 4523, Jun. 2020.

[11] A. Czajka, "Pupil dynamics for iris liveness detection," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 726–735, Apr. 2015.

[12] H. M. Pinheiro, R. M. da Costa, E. N. R. Camilo, A. da Silva Soares, R. Salvini, G. T. Laureano, F. A. Soares, and G. Hua, "A new approach to detect use of alcohol through iris videos using computer vision," in *Image Analysis and Processing—ICIAP 2015*, V. Murino and E. Puppo, Eds. Cham, Switzerland: Springer, 2015, pp. 598–608.

[13] S. S. Arora, M. Vatsa, R. Singh, and A. Jain, "Iris recognition under alcohol influence: A preliminary study," in *Proc. 5th IAPR Int. Conf. Biometrics (ICB)*, Mar. 2012, pp. 336–341.

[14] J. R. Bergstrom, S. Duda, D. Hawkins, and M. McGill, "4—Physiological response measurements," in *Eye Tracking in User Experience Design*, J. Romano Bergstrom and A. J. Schall, Eds. Boston, MA, USA: Morgan Kaufmann, 2014, pp. 81–108.

[15] A. Kumar and A. Passi, "Comparison and combination of iris matchers for reliable personal authentication," *Pattern Recognit.*, vol. 43, no. 3, pp. 1016–1026, 2010.

[16] J. P. Bernstein, B. J. Mendez, P. Sun, Y. Liu, and Y. Shang, "Using deep learning for alcohol consumption recognition," in *Proc. 14th IEEE Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2017, pp. 1020–1021.

[17] G. Koukiou and V. Anastassopoulos, "Eye temperature distribution in drunk persons using thermal imagery," in *Proc. Int. Conf. BIOSIG Special Interest Group (BIOSIG)*, Sep. 2013, pp. 1–8.

[18] G. Koukiou and V. Anastassopoulos, "Drunk person screening using eye thermal signatures," *J. Forensic Sci.*, vol. 61, no. 1, pp. 259–264, Jan. 2016.

[19] A. Valenzuela, C. Arellano, and J. E. Tapia, "Towards an efficient segmentation algorithm for near-infrared eyes images," *IEEE Access*, vol. 8, pp. 171598–171607, 2020.

[20] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[21] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Computer Vision— ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham, Switzerland: Springer, 2018, pp. 833–851.

[22] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, pp. 2–74, Dec. 2021.

[23] 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*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham, Switzerland: Springer, 2015, pp. 234–241.

[24] J. Jing, Z. Wang, M. Rätsch, and H. Zhang, "Mobile-Unet: An efficient convolutional neural network for fabric defect detection," *Textile Res. J.*, vol. 1948, May 2020, Art. no. 004051752092860.

[25] P. Rot, Ž. Emeršič, V. Struc, and P. Peer, "Deep multi-class eye segmentation for ocular biometrics," in *Proc. IEEE Int. Work Conf. Bioinspired Intell. (IWOBI)*, Jul. 2018, pp. 1–8.

[26] B. Hassan, R. Ahmed, T. Hassan, and N. Werghi, "SIP-SegNet: A deep convolutional encoder-decoder network for joint semantic segmentation and extraction of sclera, iris and pupil based on periocular region suppression," *ArXiv*, vol. abs/2003.00825, 2020.

[27] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[28] C. S. Bezerra, R. Laroca, D. R. Lucio, E. Severo, L. F. Oliveira, A. S. Britto, and D. Menotti, "Robust iris segmentation based on fully convolutional networks and generative adversarial networks," in *Proc. 31st SIBGRAPI Conf. Graph., Patterns Images (SIBGRAPI)*, Oct. 2018, pp. 281–288.

[29] Y.-H. Yiu, M. Aboulatta, T. Raiser, L. Ophey, V. L. Flanagin, P. Z. Eulenburg, and S.-A. Ahmadi, "DeepVOG: Open-source pupil segmentation and gaze estimation in neuroscience using deep learning," *J. Neurosci. Methods*, vol. 324, Aug. 2019, Art. no. 108307.

[30] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "CCNet: Criss-cross attention for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 603–612, doi: 10.1109/ICCV.2019.00069.

[31] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "CCNet: Criss-cross attention for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2019, pp. 603–612.

[32] 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 Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1529–1535, Aug. 2010.

[33] A. Dutta and A. Zisserman, "The VIA annotation software for images, audio and video," in *Proc. 27th ACM Int. Conf. Multimedia*. New York, NY, USA: ACM, Oct. 2019, pp. 2276–2279.

[34] A. B. Jung *et al.* (2020). *Imgaug*. Accessed: Feb. 1, 2020. [Online]. Available: https://github.com/aleju/imgaug

[35] P. Viola and M. Jones, "Robust real-time face detection," in *Proc. 8th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 2, 2001, pp. 747–747, doi: 10.1109/ICCV.2001.937709.

[36] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[37] P. I. Corke, *Robotics, Vision & Control: Fundamental Algorithms in MATLAB*, 2nd ed. Springer, 2017.

[38] N. Othman, B. Dorizzi, and S. Garcia-Salicetti, "OSIRIS: An open source iris recognition software," *Pattern Recognit. Lett.*, vol. 82, pp. 124–131, Oct. 2016. [Online]. Available: https://app.dimensions.ai/details/publication/pub.1008694265

[39] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[40] R. O. Duda and R. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Commun. ACM*, vol. 15, no. 1, pp. 11–15, Jan. 1972.

[41] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[42] S. Gonzalez, C. Arellano, and J. E. Tapia, "Deepblueberry: Quantification of blueberries in the wild using instance segmentation," *IEEE Access*, vol. 7, pp. 105776–105788, 2019.

[43] P. A. G. D. Smith and M. Egger, "Beyond the grand mean?" *Nat. Library Med.*, vol. 4, no. 315, pp. 1610–1614, 1997.

[44] T. Nezam, R. Boostani, V. Abootalebi, and K. Rastegar, "A novel classification strategy to distinguish five levels of pain using the EEG signal features," *IEEE Trans. Affect. Comput.*, vol. 12, no. 1, pp. 131–140, Jan. 2021.

JUAN E. TAPIA (Graduate Student Member, IEEE) received the P.E. degree in electronics engineering from Universidad Mayor, in 2004, the M.S. degree in electrical engineering from the Universidad de Chile, in 2012, and the Ph.D. degree from the Department of Electrical Engineering, Universidad de Chile, in 2016. In addition, he spent one year of internship at the University of Notre Dame, USA. From 2016 to 2017, he was an Assistant Professor with Universidad Andres Bello. From 2018 to 2020, he was the Research and Development Director for the area of electricity and electronics with the Universidad Tecnologica de Chile—INACAP. He is currently a Senior Researcher at Hochschule Darmstadt (HDA) and the Research and Development Director of TOC Biometrics. His main research interests include pattern recognition and deep learning applied to iris/face biometrics, vulnerability analysis, morphing, feature fusion, and feature selection. In 2016, he received the award for best Ph.D. thesis.

ENRIQUE LOPEZ DROGUETT is currently a Professor with the Civil and Environmental Engineering Department and the Garrick Institute for the Risk Sciences, University of California, Los Angeles (UCLA), USA. He conducts research on Bayesian inference and artificial intelligence supported digital twins and prognostics and health management based on physics informed deep learning for reliability, risk, and safety assessment of structural and mechanical systems. His most recent focus has been on quantum computing and quantum machine learning for developing solutions for risk and reliability quantification and energy efficiency of complex systems, particularly those involved in renewable energy production. He has led many major studies on these topics for a broad range of industries, including oil and gas, nuclear energy, defense, civil aviation, mining, renewable, and hydro energy production and distribution networks. He has authored more than 250 papers in archival journals and conference proceedings. He serves in the Board of Directors of the International Association for Probabilistic Safety Assessment and Management (IAPSAM). He is an Associate Editor for the *Journal of Risk and Reliability* and the *International Journal of Reliability and Safety*.

ANDRES VALENZUELA received the B.S. degree in computer engineering from the Faculty of Engineering, Universidad Andres Bello, Santiago, Chile, in 2020. He is currently working as a Researcher with the Universidad de Chile and TOC Biometrics, Chile. His main interests include computer vision, pattern recognition, and deep learning applied to semantic segmentation problems, focusing in NIR and RGB eyes images.

DANIEL P. BENALCAZAR (Member, IEEE) was born in Quito, Ecuador, in 1987. He received the B.S. degree in electronics and control engineering from Escuela Politecnica Nacional, Quito, in 2012, the M.S. degree in electrical engineering from The University of Queensland, Brisbane, Australia, in 2014, with a minor in biomedical engineering, and the Ph.D. degree in electrical engineering from the Universidad de Chile, Santiago, Chile, in 2020. From 2015 to 2016, he worked as a Professor at the Central University of Ecuador. Ever since, he has participated in various research projects in biomedical engineering and biometrics. He is currently working as a Researcher with the Universidad de Chile and TOC Biometrics, Chile.

LEONARDO CAUSA received the P.E. degree in electrical engineering from the Universidad de Chile, in 2012, and the M.S. degree in biomedical engineering (BME) from the Universidad de Chile, in 2012. He is a Ph.D. Candidate in electrical engineering and medical informatics by co-tutelage with the Universidad de Chile and Universite Claude Bernard Lyon 1. His research interests include sleep pattern recognition, signal and image processing, neuro-fuzzy systems applied to classify physiological data, and machine and deep learning. His primary interest is automated sleep-pattern detection and respiratory signal analysis, fitness for duty, human fatigue, drowsiness, alertness, and performance. He is currently a Researcher with the R+D Center TOC Biometrics Company.

CHRISTOPH BUSCH (Senior Member, IEEE) is currently a member of the Department of Information Security and Communication Technology (IIK), Norwegian University of Science and Technology (NTNU), Norway. He holds a joint appointment with the Faculty of Computer Science, Hochschule Darmstadt (HDA), Germany. Further, he has been lecturing the course biometric systems at Denmark's DTU, since 2007. On behalf of the German BSI, he has been the coordinator for the project series BioIS, BioFace, BioFinger, BioKeyS Pilot-DB, KBEinweg, and NFIQ2.0. In the European research program, he was an Initiator of the Integrated Project 3D-Face, FIDELITY, and iMARS. Further, he was/is a Partner in the projects TURBINE, BEST Network, ORIGINS, INGRESS, PIDaaS, SOTAMD, RESPECT, and TReSPAsS. He is also a Principal Investigator with the German National Research Center for Applied Cybersecurity (ATHENE). Moreover, he is a Co-Founder and a member of board of the European Association for Biometrics (www.eab.org) that was established, in 2011, and assembles in the meantime more than 200 institutional members. He coauthored more than 500 technical papers and has been a speaker at international conferences. He is a member of the editorial board of the IET journal.