Journal of Computer Science
https://dergipark.org.tr/en/pub/bbd

ISSN,e-ISSN: 2548-1304
Volume:IDAP-2022, pp:185-191, 2022
https://doi.org/10.53070/bbd.1173482

Research Paper

Pupil Center Localization Based on Mini U-Net

Kenan DONUK*[1], Davut HANBAY[2]

[1]Department of Computer Programming, Sirnak University, Cizre Vocational School, Sirnak, Turkey

[2]Department of Computer Engineering, İnönü University, Malatya, Turkey

(kenandonuk@sirnak.edu.tr, davut.hanbay@inonu.edu.tr)

Received:Sep.08,2022 Accepted:Sep.16,2022 Published:Oct.10,2022

Abstract— Many methods have been used from past to present to determine the location of the pupil center, which has an important place in eye tracking algorithms. These methods are usually shape-feature and appearance-based. Shape-feature-based methods use morphological image processing techniques, invariant geometric features of the eye, and infrared light to locate the iris and pupil. These methods are affected by real world conditions such as light, low resolution. In contrast, appearance-based methods are less sensitive to these conditions. In this study, Mini U-Net network, which is one of the appearance-based methods that automatically learns eye features and performs pupil center localization, is proposed. The proposed network was evaluated using the publicly available GI4E dataset for pupil center localization. In the test results of the network, measurements were made according to the maximum normalized error criterion. Accordingly, the center of the pupil was determined with an accuracy of 98.40%. The proposed network is compared with the latest technological methods and the performance of the proposed network is shown.

Keywords : *U-Net, mini U-Net, pupil center localization, gi4e.*

1. Introduction

Pupil localization is an active area that has been researched. This area plays an important role especially in eye tracking systems. Driving safety, control of new generation devices, virtual and augmented reality, detection of attention deficits in individuals can be mentioned as examples of developing technologies with eye tracking systems. Correct pupil/center detection is the most important part of eye-tracking technology. Today's commercial eye trackers perform highly accurate pupil localization using infrared light and high-resolution cameras. With the infrared light used, the iris, pupil and sclera parts of the eye can be clearly distinguished. The center of the pupil can therefore be determined more accurately. These devices are less sensitive to different lighting conditions. While good localization is ensured in dark environments, localization accuracy is lower in bright environments than in dark environments. The high budget of these devices and the use of invasive methods in eye tracking under controlled conditions limit the user experience. To overcome these limitations, there is a need to obtain robust pupil localization/center for user experience in harsh real-world conditions, low hardware, and low resolution cameras.

Numerous algorithms have been developed for pupil localization under real conditions. Some of them are the. In their study, Cai H. et al. (Cai et al., 2018) proposed a hierarchical adaptive convolution method to overcome the difficulties in localizing the eye center. In this method, they modeled different viewing angles of the iris with the new hierarchical kernels. The hierarchical kernels were adaptively selected depending on the 3D head pose. Thus, the localization of the eye center is performed accurately and quickly. The performance of the proposed method was tested using the GI4E dataset, which is the most commonly used dataset for eye localization. They achieved an accuracy of 85.7% and 99.5% for the normalized error criteria ($e \leq 0.025$, $e \leq 0.05$), respectively. Kitazumi K. and Nakazawa A. (Kitazumi and Nakazawa, 2019) performed pupil segmentation using the CNN-based U-Net architecture, which can be applied in dark iris and corneal reflectance scenarios. To train the proposed architecture, they used the UBIRIS.v2 eye dataset with data extensions such as color and corneal reflectance, as well as the eye images acquired in mobile environments. To verify the performance of the trained U-Net architecture, they tested the accuracy of pupil center detection using the GI4E face image set. As a result of the test, they achieved 96.28%, 98.62%, and 98.95% accuracy on the normalized error criteria ($e \leq 0.025$, $e \leq 0.05$, $e \leq 0.1$). Choi J. et al. (Choi et al., 2019) proposed an algorithm for pupil center localization using heterogeneous CNN models. BioID and GI4E datasets were used to train the proposed algorithm. For quantitative evaluation of these datasets, 5-fold cross-validation was applied. The proposed algorithm consists of two steps. Removal of the eye area and localization of the pupil center. After detecting the eye area, a fully convolutional based network (FCN) was used to segment the eye. In the pixel-based classification at the end of segmentation, the pixel position with the highest pixel density was accepted as the center of the pupil. To increase the accuracy of the network, a hopping link and the auxiliary network corresponding to the autoencoder decoder are used to compensate for the lost position information in the network FCN. An attempt was made to increase the accuracy by combining the losses of this auxiliary network and the network FCN. The proposed network was tested with the BioID and GI4E datasets. In the test performed with the GI4E dataset, the pupil center was detected with an accuracy of 90.4% and 99.6% for the normalized error criteria ($e \leq 0.025$, $e \leq 0.05$), respectively. Larumbe-Bergera A. et al. (Larumbe-Bergera et al., 2021) used a pre-trained ResNet-50 network with the ImageNet dataset to localize the eye center. They applied a procedure to fine-tune the ResNet-50 network in accordance with the eye center localization by using the PUPPIE dataset and augmenting the training with data augmentation. At the output of the network, they estimated a total of 6 landmarks, the corners of the eyes, and the centers of the two eyes. The performance of the proposed network was tested using the GI4E, I2Head, MPIIGaze, and U2Eyes datasets. As a result of the GI4E data set test, they obtained 98.46%, 100% and 100% accuracy in the normalized error criteria ($e \leq 0.025$, $e \leq 0.05$, $e \leq 0.1$), respectively.

2. U-Net Architecture

The U-Net architecture was developed in 2015 by Ronneberger O. et al. (Ronneberger et al., 2015) for biomedical image segmentation. This architecture has a "U" shaped structure. The first part of the architecture consists of encoder blocks and the second part consists of decoder blocks. The feature maps are obtained by doubling the number of channels of the image with convolution operations (3x3) in each of the 4 blocks of the first part. In each convolutional stage, all neurons are activated with the ReLU activation function, equipping the network for nonlinear situations and increasing the generalization capability of the network. By halving the spatial dimensions of the image at the end of convolution with maximum pooling (2x2), the parameters that can be trained are reduced, thus lowering the cost. In the decoder block, on the other hand, this process is symmetrically reversed, and in each of the 4 blocks the spatial dimensions are first doubled by upsampling, while the number of channels is halved. After upsampling (2x2), the tensor from the encoder block is combined with the jump connection with the upsampled tensor. The goal is to compensate for the features lost due to the deepening of the mesh and to make a better prediction using the information from the previous layers. After the fusion process, the number of channels is halved by convolution operations (3x3). After each convolution, the neurons are activated by the ReLU activation function. Segmentation or splitting is performed by pixel-level classification with a 1x1 convolution and finally sigmoid activation in the last convolutional layer in the last convolutional block of the decoder. The architecture of the U-Net is shown in Figure 1.

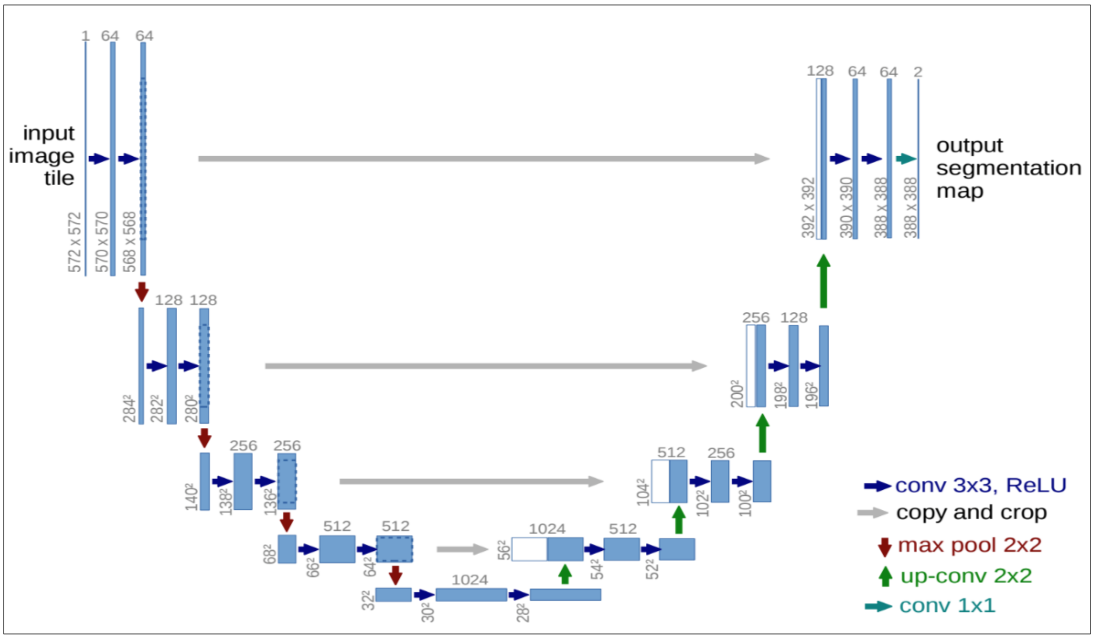

Figure 1. U-Net architecture

3. Dataset and Methodology

3.1. Dataset:

The GI4E dataset, which is widely used in pupil segmentation, was used for the study. The GI4E dataset is a database containing a total of 1236 face images corresponding to 12 images of 103 different users looking at different points on the screen with a standard webcam. The images are in png format and have a resolution of 800x600 pixels. In addition, the database contains a text file in which the points in the center of the iris and in the corners of the eyes in the face images are labeled with a total of 6 coordinate pairs for each image. Labeling was done manually. This process was performed by three different users, marking the iris center and eye corners on each image and averaging the marks of these three users (Villanueva et al., 2013). Some of the images in the GI4E dataset and the representation of the landmarks in the eye in the same dataset are shown in Figure 2 and Figure 3, respectively.

Figure 2. GI4E dataset images

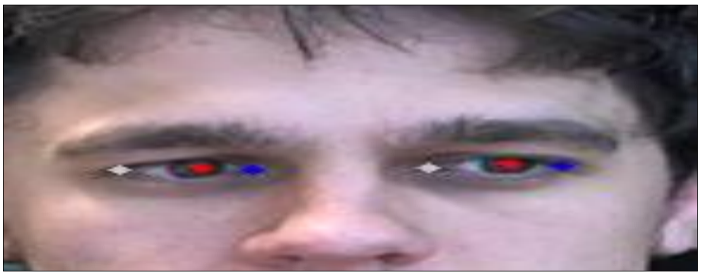

Figure 3. Implementation of GI4E dataset landmarks

3.2. Proposed Mini U-Net Architecture and Training:

The proposed Mini U-Net architecture was developed by fine-tuning the original U-Net architecture and making it suitable for pupil localization. In this architecture, the input layer is adapted to eye images with 32x32 spatial dimensions. It consists of 3 convolutional blocks instead of 4 on the encoder side of the architecture and ReLU activation functions following the convolutional operations in each block. At the end of each block, the number of channels was doubled and the spatial size was halved by applying maximum pooling. The spatial dimensions are kept constant by using padding in convolution operations. While the bottleneck layer ends with 1024 channels in the U-Net architecture, it ends with 256 channels in the proposed architecture. On the decoder side, 3 convolution blocks are used instead of 4. The spatial dimension has been doubled by upsampling at each block input. The incoming tensor with the hopping connection is combined with the upsampled tensor. The number of channels is halved by convolution operations at the end of each block. Convolution is applied to the last layer (1x1) of the last convolution block of the decoder. Unlike the original architecture, it has been changed to "swish" because it provides good localization results as the final output activation function. For segmentation, the number of channels was reduced to 1, and unlike the U-Net architecture with 32x32x1 spatial dimensions, one class output image is obtained instead of 2. The proposed Mini U-Net architecture is shown in Figure 4.

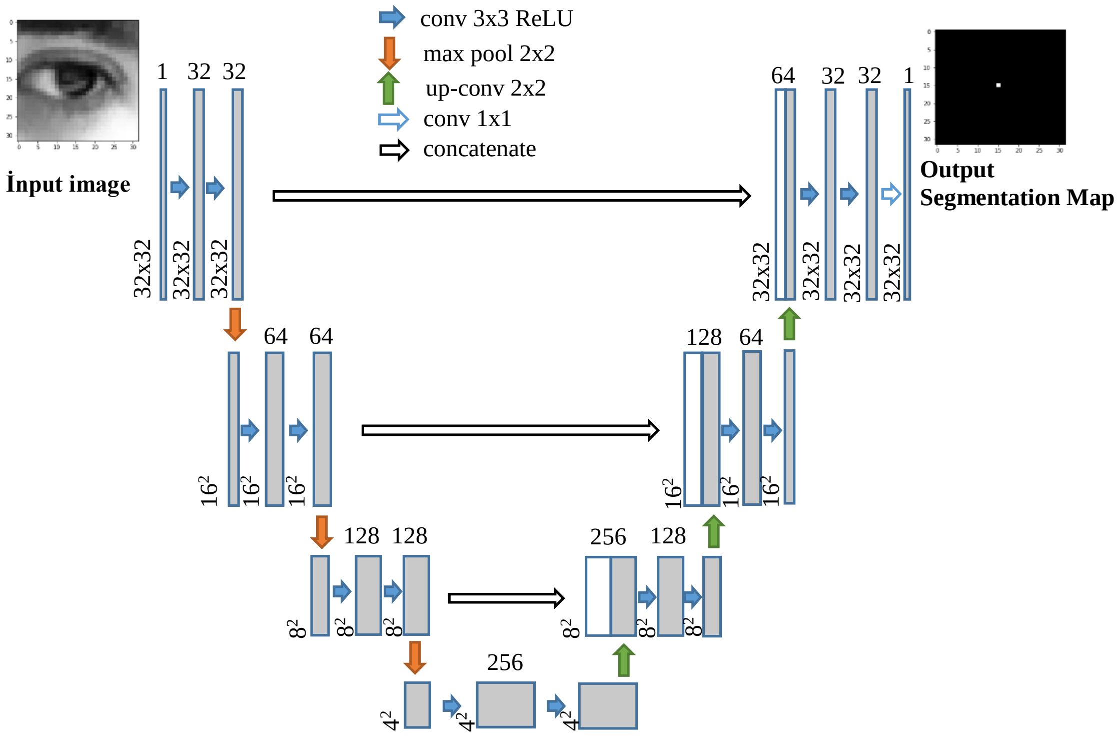

Figure 4. Mini U-Net architecture

Mini U-Net architecture needs supervised training to perform pupil localization. For this, 2472 eye images with 3 channels, right and left, were obtained from 1236 face images in the GI4E dataset. These eye images were cropped in 32x32 grayscale format using the OpenCV (OpenCV, 2022) library with eye corner coordinate tags in the text file of the GI4E dataset. Then, normalization was performed on the gray images. Obtained eye images are reserved for 80% training and 20% testing. In order to perform the pupil center localization with the Mini U-Net, the network must be fed with masked eye center images to be used in training. For this, masked images corresponding to training and test images with spatial dimensions of 32x32 were created. As shown in Figure 5, the center of the pupil is masked as white and the rest as black.

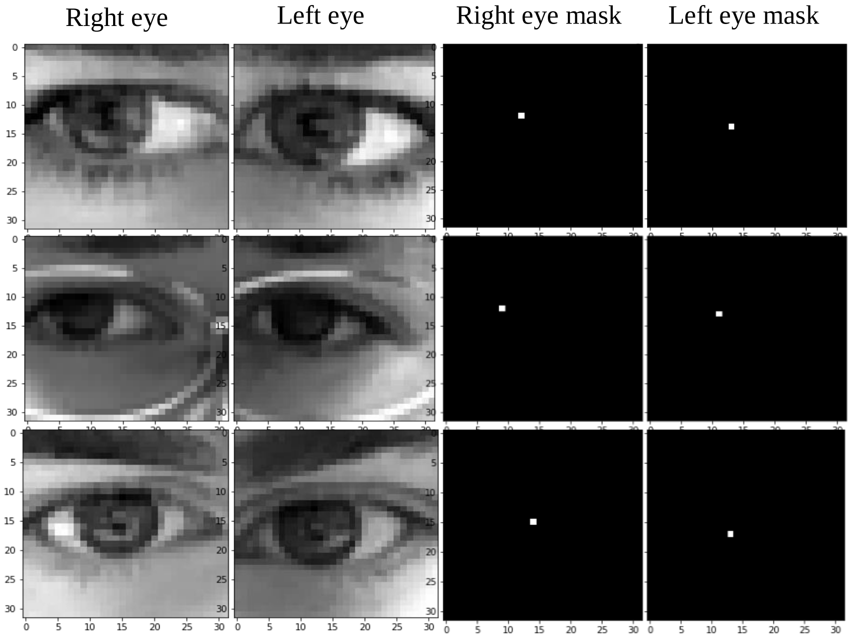

Figure 5. Masking of eye images

Loss function and optimization algorithms are needed to realize the learning process in the training phase of the network for which training and tag information is prepared. For the loss function, Mean Square Error (MSE) was used, which calculates the difference between the estimated and actual values. The aim in MSE is to reveal the similarity of the two images. Therefore, the error value is obtained by dividing the square of the difference between the pixel values between the two images by the total number of pixels. The MSE equation is given in Equation 1. As the optimization algorithm, the "Adam" optimization algorithm, which performs the best optimization suitable for the problem, was used.

$$MSE = \frac{1}{W*H}\sum_{i=1}^{W}\sum_{j=1}^{H}(y_{i,j} - \acute{y}_{i,j})^2 \tag{1}$$

W and H given in Equation 1 are the dimensions of the image, $y_{i,j}$ is the actual value of the relevant pixel, and $\acute{y}_{i,j}$ is the estimated related pixel value. Mini U-Net network is trained with 1972 training images and 1972 masked eye images. The training reached the minimum MSE value in 250 cycles. At the end of the training, the performance of the network was tested with 500 test data that had not been used before. Pupil center localization was realized by giving test images to Mini U-Net network. On the masked image estimated at the end of the segmentation performed by the network, the pixel with the most intense value was accepted as the eye center by means of the OpenCV library. In Figure 6, the positions of the localization estimates on some test data of the proposed network on the real images are marked with a red dot.

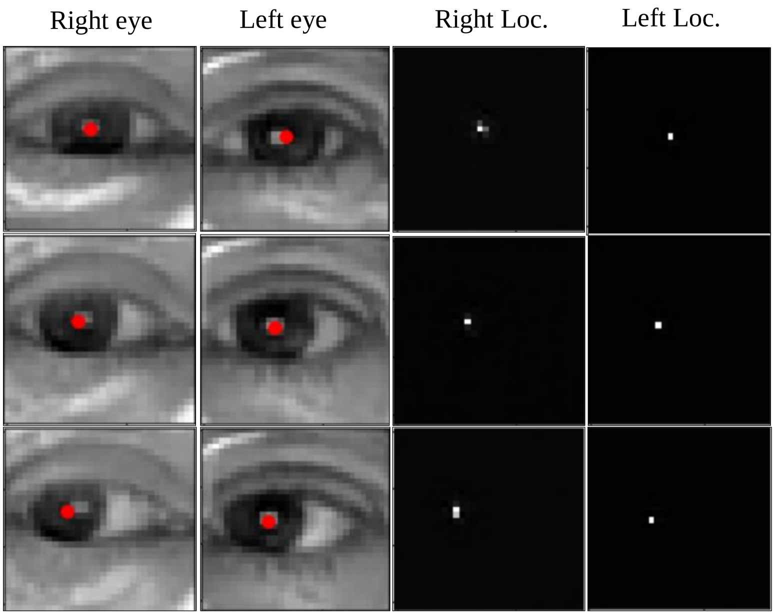

Figure 6. Test data localization estimates

4. Experimental Results

Jesorsky O. et al., a widely used evaluation criterion in pupil localization, to compare the performance of the trained mini U-Net network with state-of-the-art algorithms. The normalized error value presented by (Jesorsky et al., 2001) was used. If the normalized error value is less than or equal to 0.025, the estimated eye center is the position closest to the actual eye center. If the normalized error value is less than or equal to 0.05, the estimated eye center position is within the limits of the actual pupil. The normalized error value indicates a location within the borders of the iris if it is less than or equal to 0.10, and finally at the distance between the center of the eye and the corner of the eye if it is less than or equal to 0.25. The normalized error equation is given in equations 2 and 3. As a normalized error criterion, an accuracy of 98.40% was obtained from the Mini U-Net network in localization estimations less than 0.025 error, which refers to the position closest to the center of the pupil. In the test images given in Figure 7, the red dots represent the actual pupil center location and the yellow dots represent the estimated pupil center localization of the proposed mesh.

$$error_{max} = \frac{max(d_{left}, d_{right})}{aid} \tag{2}$$

$$aid = \sqrt{(x_{i,left} - x_{i,right})^2 + (y_{i,left} - y_{i,right})^2} \tag{3}$$

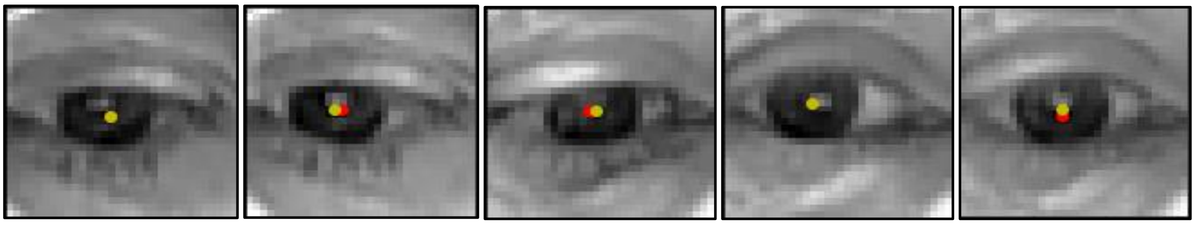

Figure 7. Comparison of actual and predicted localizations

The *error_{max}* maximum error measure in Equation 2, d_{left} is the distance between the estimated left pupil center position and the actual left eye center position, d_{right} is the distance between the estimated right pupil center position and the true right eye center position. "*aid*" represents the actual distance between the center of the left pupil and the center of the right pupil. In the "aid" equation in Equation 3, $x_{i,left}$, $x_{i,right}$ represent the x coordinates of the left and right real pupil centers, $y_{i,left}$, $y_{i,right}$, respectively, the y coordinates of the real left and right pupil centers. The accuracy comparison of the pupil center localization made with the GI4E dataset with the latest technology algorithms is given in Table 1.

Table 1. GI4E dataset pupil center localization performance comparison

Method	**$error_{max} \leq 0.025$**	**$error_{max} \leq 0.05$**	**$error_{max} \leq 0.1$**	**$error_{max} \leq 0.25$**
(Zhang et al., 2016)	--------	%97.90	%99.60	%99.92
(Gou et al., 2016)	--------	%98.20	%99.80	%99.80
(Gou et al., 2017)	--------	%94.20	%99.10	%99.80
(Levinshtein et al., 2018)	%88.34	%99.27	%99.92	%100
(Cai et al., 2018)	%85.70	%99.50	-------	--------
(Xiao et al., 2018)	--------	%97.90	%100	%100
(Kitazumi ve Nakazawa, 2019)	%96.28	%98.62	%98.95	--------
(Choi et al., 2019)	%90.40	%99.60	-------	--------
(Xia et al., 2019)	--------	%99.10	%100	%100
(Kim et al., 2020)	%79.50	%99.30	%99.90	--------
(Lee et al., 2020)	--------	%99.84	%99.84	%100
(Larumbe-Bergera et al., 2021)	%98.46	%100	%100	--------
Proposed network (Mini U-Net)	**%98.40**	**%99.40**	**%99.60**	**%99.80**

When Table 1 is examined, it is seen that the algorithms exhibited head-to-head results in the 0.1 and 0.25 normalized error comparison. However, it is seen that the proposed Mini U-Net network performs well in cases of 0.025 and 0.05, which are the closest error rates to the location of the real pupil center.

5. Conclusion

In this study, a lower level Mini U-Net algorithm is proposed based on the deep learning-based U-Net architecture that has proven its success in image segmentation in the localization of the pupil center. Pupil center coordinates are estimated from the eye images of the GI4E dataset, which is given as input to the proposed algorithm, with an accuracy of 98.40% according to the 0.025 normalized error criterion. The results were compared with other algorithms suggested in the literature. At the end of the comparison, the success of the proposed algorithm is clearly seen. The proposed algorithm can be implemented in real time on eye images obtained using the dlib (Dlib, 2022) library. It can also be used to accurately detect the eye center in real-time eye tracking applications.

References

Cai H, Liu B, Ju Z, Thill S, Belpaeme T, Vanderborght B, Liu H. (2018) Accurate Eye Center Localization via Hierarchical Adaptive Convolution. In Proceedings of the 29th British Machine Vision Conference, BMVC, pp.284.

Choi J. H, il Lee K, Kim Y. C, Cheol Song B. (2019) Accurate Eye Pupil Localization Using Heterogeneous CNN Models. Proceedings - International Conference on Image Processing, ICIP, pp.2179-2183.

Dlib C++ Library (2022). http://www.dlib.net. Accessed 25 July 2022

Gou C, Wu Y, Wang K, Wang F. Y, Ji Q. (2016) Learning-by-synthesis for accurate eye detection. Proceedings - International Conference on Pattern Recognition, pp.3362-3367.

Gou C, Wu Y, Wang K, Wang K, Wang F. Y, Ji Q (2017) A joint cascaded framework for simultaneous eye detection and eye state estimation. *Pattern Recognition* 67:23–31.

Jesorsky O, Kirchberg K. J, Frischholz R. W (2001) Robust face detection using the Hausdorff distance. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 2091:90-95.

Kim S, Jeong M, Ko B. C (2020) Energy Efficient Pupil Tracking Based on Rule Distillation of Cascade Regression Forest. *Sensors* 20(18):5141.

Kitazumi K, Nakazawa A. (2019) Robust Pupil Segmentation and Center Detection from Visible Light Images Using Convolutional Neural Network. Proceedings - 2018 IEEE International Conference on Systems, Man, and Cybernetics, SMC, pp.862–868.

Larumbe-Bergera A, Garde G, Porta S, Cabeza R, Villanueva A (2021) Accurate pupil center detection in off-the-shelf eye tracking systems using convolutional neural networks. *Sensors* 21(20).

Lee K. Il, Jeon J. H, Song B. C (2020) Deep Learning-Based Pupil Center Detection for Fast and Accurate Eye Tracking System. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12364 LNCS, 36-52.

Levinshtein A, Phung E, Aarabi P (2018) Hybrid eye center localization using cascaded regression and hand-crafted model fitting. *Image and Vision Computing* 71:17–24.

OpenCV (2022). https://www.opencv.org. Accessed 25 July 2022

Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9351:234-241.

Villanueva A, Ponz V, Sesma-Sanchez L, Ariz M, Porta S, Cabeza R (2013) Hybrid method based on topography for robust detection of iris center and eye corners. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 9(4).

Xia Y, Yu H, Wang F. Y (2019) Accurate and robust eye center localization via fully convolutional networks. *IEEE/CAA Journal of Automatica Sinica* 6(5):1127–1138.

Xiao F, Huang K, Qiu Y, Shen H (2018) Accurate iris center localization method using facial landmark, snakuscule, circle fitting and binary connected component. *Multimedia Tools and Applications* 77(19):25333-25353.

Zhang W, Smith M. L, Smith L. N, Farooq A (2016) Eye center localization and gaze gesture recognition for human-computer interaction. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision* 33(3):314-325.