16th International Conference on Machine Vision Applications (MVA)
National Olympics Memorial Youth Center, Tokyo, Japan, May 27-31, 2019.

Pupil Localization for Ophthalmic Diagnosis Using Anchor Ellipse Regression

Horng-Horng Lin[1] Zheng-Yi Li[2] Min-Hsiu Shih[2,*] Yung-Nien Sun[2,*] Ting-Li Shen[2]

[1] Southern Taiwan University of Science and Technology, Tainan City, Taiwan
[2] National Cheng Kung University, Tainan City, Taiwan
{hhlin.tw, x36812}@gmail.com, [*]{mhshih, ynsun}@mail.ncku.edu.tw, tlshen3885@mail.mirdc.org.tw

Abstract

Recent developments of deep neural networks, such as Mask R-CNN, have shown significant advances in simultaneous object detection and segmentation. We thus apply deep learning to pupil localization for ophthalmic diagnosis and propose a novel anchor ellipse regression approach based on region proposal network and Mask R-CNN for detecting pupils, estimating pupil shape parameters, and segmenting pupil regions at the same time in infrared images. This new extension of anchor ellipse regression for Mask R-CNN is demonstrated to be effective in size and rotation estimations of elliptical objects, as well as in object detections and segmentations, by experiments. Temporal pupil size estimations by using the proposed approach for normal and abnormal subjects give meaningful indices of pupil size changes for ophthalmic diagnosis.

1. Introduction

Pupil reflexes with lighting changes are important diagnostic features of ocular diseases and neuropathies. For example, the diagnosis of relative afferent pupillary defect (RAPD), which is an important symptom for glaucoma disease, often includes an observation of slower or smaller pupil size changes subject to light simulations [1], [2]. In Adie's tonic syndrome, the pupil sizes of both eyes are uneven under normal brightness, and may be abnormally dilated with delayed constriction in response to light exposure [3], [4]. In clinical diagnosis, clinicians usually determine pupil normality subjectively, which may lead to different judgments among different clinicians, thereby rendering accurate assessment of pupil disease severity difficult. Therefore, an automatic and quantitative pupil assessment system is important for acquiring pupil parameters, i.e., pupil center locations and sizes, and for quantifying pupil diseases.

When designing such a pupil assessment system with video capturing for various clinical environments, several challenging conditions should be addressed. Figure 1(a)&(b) show the image acquisition system and an image sample of an infrared eye video, respectively, wherein the pupil regions are generally darker than surrounding iris regions under infrared video capturing. Nevertheless, in Figure 1(c)-(e), the pupils are partially or fully occluded by the eyelid and eyelashes. Besides, as shown in Figure 1(f), reflection spots of projection light in the pupil region will sometimes affect the pupil measurement. Accordingly, we aim to develop a robust pupil localization algorithm that can accurately estimate pupil center and size parameters in real time under various lighting conditions and interferences for the construction of a practical pupil size assessment system.

In recent years, deep learning has brought big advances in many challenging computer vision tasks, such as image object recognitions [5], [6], [7] and segmentations [8], [9], [10]. In this study, we adopt the framework of Mask R-CNN [11] for pupil region segmentation, and extend this image object segmentation framework to the estimation of pupil localization parameters by introducing *anchor ellipses* and *ellipse regressions*. An advanced computational scheme of deep neural network that is capable of simultaneous pupil detection, segmentation and localization parameter estimation is newly proposed.

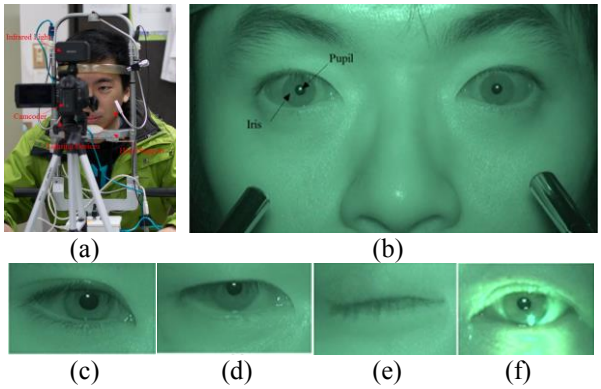


Figure 1. (a) Infrared eye imaging system and captured image samples (b) without eyelid occlusion, (c) with partial eyelashes occlusion, (d) with partial eyelid occlusion, (e) with full eyelid occlusion (eye blinking), and (f) with lighting reflection spots.

1.1. Related Work

Automatic image processing has been applied to diseases identifications [12]-[18] for many decades. In particular, the detection of circular image objects, such as pupils and irises, for medical diagnosis are conventionally performed by Hough transform [19], least squares fitting [20], active contour model (ACM) [21] or active shape model (ASM) [22]. For example, Sahmoud and Abuhaiba [23] adopted *k*-means clustering and Hough transform for iris detection in noisy images under unconstrained environments. Bastos *et al*. [13] combined

ACM with pulling-and-pushing and polar coordinate transform for pupil segmentation. Abdullah *et al*. [16] applied morphological operations and ACM to extracting pupil boundaries in eye images. Chen *et al*. [17] designed a pupil size and blink estimation method to detect pupil boundary points and eyelid occlusion states by convex hull and dual-ellipse fitting for infrared eye images. For the consideration of real-time processing speed, de Souza *et al*. [18] developed a pupil and blink detection method for infrared eye images, in which adaptive thresholding, morphological operator and the Canny edge filter were used to detect pupils. In [24], Shen *et al*. used *k*-means clustering to locate pupils and applied appearance-based circle matching to real-time pupil segmentation. These conventional methods for pupil detection, as mentioned above, despite their detection accuracies and computational efficiencies, often require a tedious work of manual parameter tuning. Recently, comparisons of several pupil detection algorithms for head-mount eye tracking are given in details in [25].

Different from the conventional approaches mentioned above, recent deep learning methods, e.g., U-NET [9], DenseUNET [10], Faster R-CNN [26] and Mask R-CNN [11], reveal a new framework of image object detection and/or segmentation in fully automatic manners by stacking convolutional neural network layers for the construction of feature maps for general image objects. In particular, to detect image objects of different size ratios efficiently, Fast R-CNN [27] and Faster R-CNN [26] adopted a region proposal network with anchor boxes of varied size ratios to locate and classify various image objects. Based on [26], Ho *et al*. [11] further combined object bounding box regression, object classification and segmentation in a unified computational scheme and achieved semi-real-time efficiency. In addition to 2D object detection and segmentation, 3D model estimation of human motion is recently proposed in QuaterNet [28] with Quaternion representations [29] for 3D joint localization and singularity treatment.

1.2. Our Approach for Pupil Localization

While a pupil is often modeled as a circle or an ellipse, it shape is actually not regular. However, we propose to adopt an elliptical shape model for pupil detection, because an ellipse is more general than a circle model and applicable for efficient localization task.

Based on the idea of extending anchor boxes to shape models in region proposal network (RPN) [26], we propose in this study *anchor ellipse regression* for RPN, which is a part of the Mask R-CNN deep network, and investigate the feasibility of estimating ellipse parameters for object localization. In order to treat parameter estimation ambiguity of ellipse fitting during regression, as will be elaborated in Sec. 2.3, we suggest an ellipse axis length constraint for the generations of anchor ellipse and training data. The proposed anchor ellipse regression allows a deep learning network to output not only object detection and segmentation, but also object localization.

Our extension of a deep neural network from object detection and segmentation to localization for ophthalmic diagnosis has several features, including:

1. A new development of a new deep learning application on ophthalmic medical diagnosis,
2. Efficient localization of elliptical objects with accurate parameter estimations, and
3. An effective treatment to the ambiguity of localization parameter estimation by axis length constraint.

2. Anchor Ellipse Regression

Based on region proposal network, a generalization from anchor boxes to anchor ellipses is newly proposed in this work for pupil localization. While anchor boxes have been proved effective for image object detection in RPN, the box representation somehow lacks further expression power of object shape changes, such as rotations, for object localization purposes and accurate parameter estimations. We therefore extend original anchor boxes to anchor ellipses of various major and minor axis lengths. Such an extension not only suits our application of pupil localization but also conveys a new perspective of unifying detection and localization in a consistent computational framework.

2.1 Anchor Box Regression in RPN

In RPN, the pre-specified anchor boxes are in 5 scales and 3 aspect ratios. To identify region proposals of target objects, two network layers of classification and regression are built upon deep convolutional feature maps and in charge of categorizing and tuning sliding anchor boxes. Thus, the loss function of RPN consists of a classification term L_{cls} and a regression term L_{reg} as

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i, p_i^*) + \lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(\boldsymbol{t}_i, \boldsymbol{t}_i^*),$$

where p_i and p_i^* denote the predicted and true class category of the *i*-th anchor, respectively, while $\boldsymbol{t}_i$ and $\boldsymbol{t}_i^*$ correspond to the vectors of estimated and labeled box parameters. The parameter λ here is weighting scalar. Particularly, the regression term plays an important role in locating an anchor box at target object position (x, y) and refining its box width/height $\{w, h\}$ to fit a true object size. Note that, in Faster R-CNN [26] and Mask R-CNN [11] that both adopt RPN as the first stage processing for deriving region proposals, the second stage regression layer is incorporated as well for fine-tuning the anchor box regression results.

2.2 Anchor Ellipse Extension

As shown in Figure 2(a), an anchor box can be transferred to an ellipse by assigning box width and height as elliptical major and minor axes. Accordingly, the regression variable $\boldsymbol{t}_i$ that originally contains four tuples $\{x, y, w, h\}$ needs to be concatenated with major axis length A, minor axis length B, and ellipse rotation angle θ. The new regression variable $\boldsymbol{t}_i$ is hence extended to seven tuples $\{x, y, w, h, A, B, \theta\}$.

In anchor generation, anchor ellipses can be conven-

iently derived from original anchor boxes, having the same 5 scales and 3 aspect ratios. The major and minor axes of an anchor ellipse are actually set as halves of the long and short sides, l_a and s_a, of an anchor box, respectively. The rotation angles of all anchor ellipses are assigned 0 degree as well.

For anchor ellipse regression, the ellipse parameters in $\boldsymbol{t}$ are computed by

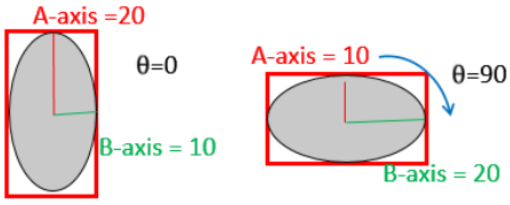
$$\begin{aligned} t_A &= \log(A/l_a) \quad & t_A^* &= \log(A^*/l_a) \\ t_B &= \log(B/s_a) \quad & t_B^* &= \log(B^*/s_a) \\ t_\theta &= \theta/180 \quad & t_\theta^* &= \theta^*/180 \end{aligned} \tag{1}$$

where A^*, B^* and θ^* are ground-truth parameters of a target elliptical object. Here we arrange rotation angles in [0,179] degrees to accommodate all possible rotation conditions due to the symmetry of elliptical shapes.

2.3 Treatment of Ellipse Estimation Ambiguity by Axis Length Constraint

When estimating ellipse parameters for localization, there exists an estimation ambiguity that a target ellipse can be fitted from more than one anchor ellipse. As depicted in Figure 2(b), a vertical ellipse can be fitted by either a horizontal ellipse of $A = 10$, $B = 20$ under $\theta = 90$ rotation, or a vertical ellipse of $A = 20$, $B = 10$ under $\theta = 0$ rotation. Such an ambiguity cannot be resolved by RPN regression and needs to be amended by an extra treatment.

To deal with this ambiguity, we propose to impose an axis length constraint, i.e., $A > B$, on the generations of anchor ellipse and training ellipses. With this constraint, the fitting of the vertical ellipse, as shown in Figure 2(a), will be $A = 20$, $B = 10$ and $\theta = 0$. Such a treatment confines the RPN regressor to learn preferable, consistent ellipse localization parameters from data.



(a) (b)

Figure 2. Example anchor ellipses of (a) a vertical one, and (b) a horizontal one.

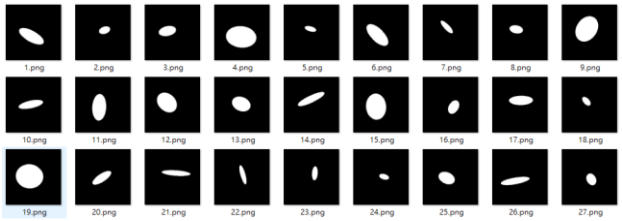


Figure 3. Samples of binary ellipse images.

3. Data Preparation and Implementations

We modify the RPN of an alternative Mask R-CNN implementation [30] to incorporate the proposed anchor ellipse models. To verify the feasibility of anchor ellipse regression, binary ellipse images are firstly generated as a basic training and testing dataset for performance test.

As image samples shown in Figure 3, we randomly generate, in 256×256 images, 9000 and 300 ellipses of different sizes and rotation angles for training and testing respectively**.** In these binary ellipse images, the major-axies As are all parallel to the Y-axis and the minor-axes Bs are along the X-axis. To prevent degenerate cases from ellipse to circle, we manually set $\mathrm{A} > \mathrm{B} + 5$. Experimental results of this basic test are given in the next section.

In addition to the binary ellipse images for feasibility test of anchor ellipse regression, we also acquired 2436 infrared eye images for medical diagnosis test by Sony HDR-PJ790 camcorder with built-in night vision, where the camcorder was positioned 10~20cm behind a head support, as shown in Figure 1. Two controllable lighting devices attached to both sides of the head support provide light stimulation, with each side alternatively turning on and off a given number of times. Subjects were asked to place their chin on the head support and stare at a pre-specified spot. The pupil is then recorded for about 10 seconds with an image size of 1280×720 at 30 fps. In this manner, the resultant pixel resolution of the acquired image is 0.116mm/pixel.

In the adopted Mask-RCNN implementation, the computational backbone can be ResNet50 or ResNet101 [7], together with FPN feature map layers of 256 channels on P2~P5 [31]. As the RPN slides anchor windows on the feature map, each feature map convolves a 3×3 kernel with ReLU activation for increasing the number of channels to 512, which will be used as share information of later classification and regression. We train on a NVIDIA GTX1060 GPU for 15k iterations with learning rate 0.001. We set a weight decay of 0.0001, the momentum 0.9 and batch size 10 in the experiments.

4. Experimental Results

To quantify the pupil segmentation accuracy of in our experimental comparisons, the Dice coefficient [32] defined as $2|R \cap T|/(|R| + |T|)$, where R and T are the segmented pupil region and the ground truth, respectively, is used as a performance measure. In the 2436 infrared eye images, 100 frames are manually labeled as a test set to mark left and right eye regions, in which some challenging conditions, e.g., eye blinking, pupil shifting, and overexposure, are included. The execution speeds of our anchor ellipse regression on Mask R-CNN are about 7.6 fps and 6.97 fps for the backbone of ResNet50 and ResNet101, respectively.

4.1 Ellipse Localization Results

We use the generated binary ellipse images, as shown in Figure 3, to verify the feasibility of the proposed anchor ellipse regression by localizing target ellipses and estimating their major and minor axes and rotation angles. As a result, the average major- and minor-axis estimation errors in length are both below 1 pixel, while the average error of rotation angle estimation is about 4

degrees. Some localization results using the proposed anchor ellipse regression are given in Figure 4. The experimental results of this basic test are rather promising and validate the effectiveness of the proposed approach of ellipse parameter estimation using anchor regression based on deep feature maps.

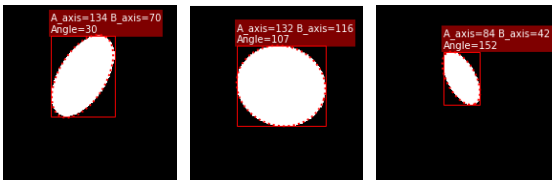


Figure 4. Sample results of binary ellipse localization.

4.2 Pupil Localization for Medical Diagnosis

Because the positions of left and right pupils acquired by the imaging device are mostly within fixed ranges in an 1280×720 image, we therefore crop two 384×384 image regions corresponding to left and right eyes in each image frame for further pupil localization.

In our extension of anchor ellipse regression for Mask R-CNN, pupil regions can be quantified by rectangular bounding boxes (BBoxes), the proposed elliptical shapes and mask segments. By averaging the width and height of a BBox on a pupil as a circle radius, one can easily extract a circular region round a pupil. Such a circular region derived from BBox is used as a baseline for pupil localization evaluations. Besides, a segmentation mask on pupil, which may be in non-regular shape, can also be obtained by Mask R-CNN. The mask segments of pupils are also adopted as an index for experimental comparisons. Additionally, the adopted deep learning-based approach is also compared to the conventional adaptive thresholding method [24], which is based on *k*-means clustering and thresholding, for pupil localization.

As depicted in Table 1 for quantitative evaluations, the proposed elliptical localizations is compared with the adaptive thresholding [24], the baseline BBox circles and the original mask segments from Mask R-CNN for pupil region extraction. The proposed ellipse regression results are better than BBox circles, but a little worse than pupil mask segmentations, due to pupils being not in perfect elliptical shapes. Note also that mask segmentations derived from the original Mask R-CNN, if without any further processing, cannot output localization parameters of pupils, such as, axes and rotation angles, which can be directly derived by the proposed approach. Moreover, the proposed anchor ellipse regression is shown in Figure 5 to be effective on pupil localization under slightly partial occlusion.

Regarding pupil localization for medical diagnosis, we report light stimulation experiments for a normal and an abnormal case in this paper. Considering that the pupil sizes of both eyes were different in some subjects, we compute the pupil expansion and contraction sizes in terms of circle radii. Here, two ellipse axes are averaged to estimate a pupil radius for fast analysis. Figure 6(a) and (b) show the curve diagrams of the detected pupil radii for a normal and an abnormal case, respectively. The blue and red curves respectively indicate the radii of the left and right pupils, and the blue and pink marks in the plots correspond to the detected lighting periods on the left and right sides. The trends of pupil radius changes of both the left and right pupils are nearly identical in the normal case. On the other hand, the trend difference between both eyes can be easily discriminated in the abnormal subject. This experiment demonstrates the effectiveness of the proposed anchor ellipse regression on practical ophthalmic diagnosis.

Table 1. Experimental comparisons of pupil segmentation results using Dice coefficients.

Dice Coefficient	BBox Estimation	Anchor Ellipse Regression	Mask Segmentation
[24]	-	-	0.892
Our approach w/ ResNet50	0.931	0.940	0.958
Our approach w/ ResNet101	0.938	0.945	0.959

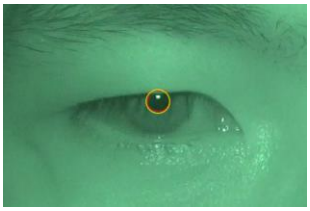
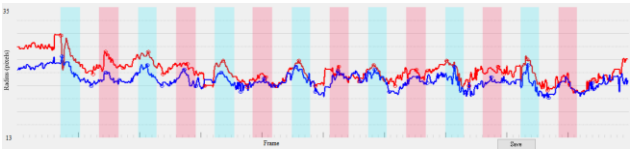


Figure 5. Result of pupil localization using anchor ellipse regression (in yellow) and mask segmentation (in red) under partial eyelid occlusion.


(a)

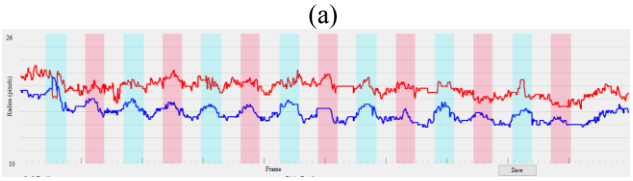

(b)

Figure 6. Pupil radii curves for (a) a normal subject and (b) an abnormal subject for ophthalmic diagnosis.

5. Conclusion

We propose a new computational scheme of anchor ellipse regression for Mask R-CNN-based deep network for simultaneous derivations of object detection, localization parameter estimation and object mask segmentation. Experimental results demonstrate the effectiveness of the proposed approach on pupil localization parameter estimation and on ophthalmic diagnosis.

In the future, a larger number of subject image sequences may be acquired for further evaluation and improvement of the proposed system. The clinical diagnostic applications for different eye diseases will also be investigated.

Acknowledgement: This work is supported in part by the Ministry of Science and Technology, Taiwan, under MOST 108-2634-F-006-005 and MOST 107-2637-E-218-004, and the A^{2}IBRC, STUST, under MOE Higher Education Sprout Project, Taiwan.

References

[1] J. Pearce: "The Marcus Gunn Pupil," *Journal of Neurology, Neurosurgery & Psychiatry*, vol.61, no.5, p. 520, 1996.

[2] C. David: "How to test for a relative afferent pupillary defect (RAPD)," *Community Eye Health Journal*, vol.29, no.96, pp. 68-69, 2016.

[3] A. A. Siddiqui, J. C. Clarke, and A. Grzybowski; "William John Adie: the man behind the syndrome," *Clinical & Experimental Ophthalmology*, vol.42, no.8, pp. 778-784, 2014.

[4] D. G. F. Harriman, and H. Garland: "The pathology of Adie's syndrome," *Brain*, vol.91, no.3, pp. 401-418, 1968.

[5] A. Krizhevsky, I. Sutskever, and G. Hinton: "Imagenet classification with deep convolutional neural networks," *NIPS*, 2012.

[6] K. Simonyan and A. Zisserman: "Very deep convolutional networks for large-scale image recognition," *ICLR*, 2015.

[7] K. He, X. Zhang, S. Ren, and J. Sun: "Deep residual learning for image recognition," *CVPR*, 2016.

[8] J. Long, E. Shelhamer, and T. Darrell: "Fully convolutional networks for semantic segmentation," *CVPR*, 2015.

[9] O. Ronneberger, P. Fischer, and T. Brox: "U-net: Convolutional networks for biomedical image segmentation," *MICCAI*, 2015.

[10] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng: "H-denseunet: Hybrid densely connected unet for liver and tumor segmentation from CT volumes," *IEEE Trans. Medical Imaging*, vol. 37, no. 12, pp. 2663-2674, 2018.

[11] K. He, G. Gkioxari, P. Dollar,and R. Girshick: "Mask R-CNN," *CVPR*, 2017.

[12] Z, He, T. Tan, and Z. Sun: "Iris localization via pulling and pushing." *ICPR,* 2006.

[13] C. A. Bastos, R. Tsang, and G. D. Calvalcanti: "A combined pulling & pushing and active contour method for pupil segmentation," *ICASSP*, 2010.

[14] E. T. Mahnaz, C. Lucas, S. Sadri, and E. Y. K. Ng: "Analysis of breast thermography using fractal dimension to establish possible difference between malignant and benign patterns," *Journal of Healthcare Engineering* vol.1, no.1, pp. 27-43, 2010.

[15] F. Fahmi, H. Marquering, G. Streekstra, L. Beenen, N. Janssen, C. Majoie, and E. vanBavel: "Automatic Detection of CT Perfusion Datasets Unsuitable for Analysis due to Head Movement of Acute Ischemic Stroke Patients," *Journal of Healthcare Engineering*, vol.5, no.1, pp. 67-78, 2014.

[16] M. A. Abdullah, S. S. Dlay, and W. L. Woo: "Fast and accurate pupil isolation based on morphology and active contour," *Proc. ICSIPA*, 2014.

[17] S. Chen, and J. Epps: "Efficient and robust pupil size and blink estimation from near-field video sequences for human–machine interaction," *IEEE Trans. Cybernetics*, vol.44, no.12, pp. 2356-2367, 2014.

[18] J. K. S. de Souza, M. A. da Silva Pinto, P. G. Vieira, J. Baron, and C. J. Tierra-Criollo: "An open-source, firewire camera-based, Labview-controlled image acquisition system for automated, dynamic pupillometry and blink detection." *Computer Methods and Programs in Biomedicine*, vol.112, no.3, pp. 607-623, 2013.

[19] M. Rizon, Y. Haniza, S. Puteh, A. Yeon, M. Shakaff, S. Abdul Rahman, and M. Karthigayan: "Object detection using circular Hough transform," *American Journal of Applied Sciences*, vol. 2, no. 12, pp. 1606-1609, 2005.

[20] W. Gander, G. H. Golub, and R. Strebel: "Least-squares fitting of circles and ellipses," *BIT Numerical Mathematics*, vol.34, no.4, pp. 558-578, 1994.

[21] M. Kass, A. Witkin, and D. Terzopoulos: "Snakes: Active contour models," *International Journal of Computer Vision*, vol.1, no.4, pp. 321-331, 1988.

[22] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham: "Active shape models - their training and application," *Computer Vision and Image Understanding*, vol. 61, pp. 38-59, 1995.

[23] S. A. Sahmoud, and I. S. Abuhaiba: "Efficient iris segmentation method in unconstrained environments," *Pattern Recognition*, vol.46, no,12, pp. 3174-3185, 2013.

[24] T.-L. Shen, B.-I Chuang, M.-H. Shih, and Y.-N. Sun: "Fast pupil assessment for sensory evaluation from infrared video," *Proc. CVGIP*, Taiwan, 2015.

[25] W. Fuhl, M. Tonsen, A. Bulling, and E. Kasneci, "Pupil detection for head-mounted eye tracking in the wild: an evaluation of the state of the art," *Machine Vision and Applications*, vol. 27, no. 8, pp. 1275-1288, 2016.

[26] S. Ren, K. He, R. Girshick, and J. Sun: "Faster R-CNN: Towards real-time object detection with region proposal networks," *NIPS*, 2015.

[27] R. Girshick: "Fast R-CNN," *ICCV*, 2015.

[28] D. Pavllo, D. Grangier, and M. Auli: "Quaternet: A quaternion-based recurrent model for human motion," *BMVC*, 2018.

[29] Edward Pervin and Jon Webb: "Quaternions for computer vision and robotics," *CVPR*, 1983.

[30] https://github.com/matterport/Mask_RCNN

[31] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie: "Feature pyramid networks for object detection," *CVPR*, 2017.

[32] L. R. Dice: "Measures of the amount of ecologic association between species," *Ecology*, vol.26, no.3, pp. 297-302, 1945.