Link to data:
https://atreus.informatik.uni-tuebingen.de/seafile/d/8e2ab8c3fdd444e1a135/?p=%2FSegmentationEye&mode=list

# 500,000 images closer to eyelid and pupil segmentation

Wolfgang Fuhl[1], Wolfgang Rosenstiel[1], and Enkelejda Kasneci[1]

Eberhard Karls University Tübingen, Tübingen 72076 , Germany
{Wolfgang.Fuhl,Wolfgang.Rosenstiel,Enkelejda.Kasneci}@uni-tuebingen.de

**Abstract.** Human gaze behavior is not the only important aspect about eye tracking. The eyelids reveal additional important information; such as fatigue as well as the pupil size holds indications of the workload. The current state-of-the-art datasets focus on challenges in pupil center detection, whereas other aspects, such as the lid closure and pupil size, are neglected. Therefore, we propose a fully convolutional neural network for pupil and eyelid segmentation as well as eyelid landmark and pupil ellipsis regression. The network is jointly trained using the Log loss for segmentation and L1 loss for landmark and ellipsis regression. The application of the proposed network is the offline processing and creation of datasets. Which can be used to train resource-saving and real-time machine learning algorithms such as random forests. In addition, we will provide the worlds largest eye images dataset with more than 500,000 images. DOWNLOAD

**Keywords:** eye tracking, eyelid segmentation, eyelid opening, pupil segmentation, landmark detection, landmark regression, pupil ellipses regression, eyelid regression

## 1 Introduction

Psychology, medicine, marketing research, computer graphics, car industry, and many other disciplines are interested in the information contained in human eyes. For instance computer graphics need a robust gaze signal for foveated rendering [50,31] to be useful for the consumer market. With the upcoming of autonomous driving, the interest in information contained in the eye is also rising for the car industry [44,6]. They not only regard the gaze signal alone. The main interest is in the estimation of the drivers capability to take over the car in critical situations [44,6]. Therefore, other features such as the eyelids are important to extract information about the cognitive state of a person [29,46]. This information is also important for psychology and cognition science. Where the workload of a person [40,48], movement processes which are predictable based on the gaze [41], and also characteristic eye movements which identify diseases [5]. In medicine, eye tracking is not applied to disease classification [35,4]. Current research also regards novice training and expertise level classification [28,14] as well as surgical microscope steering [13]. Using the gaze signal as a control signal for human computer interaction is also used in virtual reality [12] and the eye is used as an identification characteristic [47]. Andrew T. Duchowski summarized most of the aforementioned application areas already in 2002 [11]. Further advancement in the field of eye tracking and its application areas is still limited by the amount of information that can be extracted out of the eye.

Current research in video based eye tracking mainly concerns a robust pupil signal [26,20,49]. The same is true for commercial systems [38] because robust gaze signal is the most important first step for eye tracking to be applicable. It needs to handle near infrared illuminated images containing heavy reflections, and rapid illumination changes trough sunlight etc. Stand alone algorithms for eyelid extraction and eyelid opening estimation were also proposed in [10,23]. This separation of pupil and eyelid detection limits research and progress in application areas that use all the information contained in the eye. Machine learning based image processing have made huge progress through the invention of local stationary features as used in convolution neural networks (CNN) [42]. Together with the advances in hardware allowing GPUs as massive parallel programming device, CNNs are already applicable in real time. Further advances in CNN architectures, like residual [34] and inception modules [57], allow to train deeper networks and improve the accuracy and robustness further. The current state-of-the-art in computer vision tasks (detection, classification, segmentation, image generation) is taken by the above mentioned architectures. For semantic image segmentation, the invention of transposed convolutions [45] lead to a breakthrough. They can train a fully convolution neural network [45] in a way that the loss function also contains spatial information compared to fully connected layers. Another approach is the encoder and decoder architecture [3]. Here, the pooling information is shared between corresponding layers in the encoder and the decoder for up-sampling. Other approaches for image segmentation stem from generative adversarial networks (GAN) [30]. They have the disadvantage that the training is difficult because the discriminator is likely to overfit. While this issue was solved with the cycle loss function [65] and unpaired training samples, they have not yet achieved the accuracy of fully convolutional neural networks trained on training sample pairs.

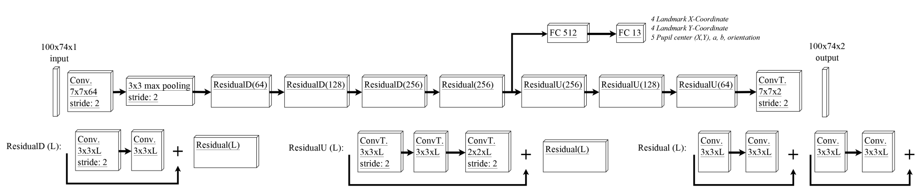


**Fig. 1.** The proposed architecture of the joint regression and segmentation CNN.

In this work, we propose a combined convolutional neural netowrk architecture for eyelid landmark, pupil ellipse regression together with pupil area and eyelid area segmentation. Our architecture is based on residual blocks [34], which allow us to train a deep network. We used the L1 loss and the log loss function for regression and segmentation respectively. The idea behind a combined approach is that multiple tasks performed by the same network improve the accuracy [64]. Additionally, the results and the landmarks can be compared against each other to detect invalid segmentations and correct them. Furthermore, we contribute a large dataset containing more than 500,000 images with segmented pupil and eyelid areas. These images were annotated using the proposed architecture, which will be described with the training procedure in the following.

## 2 Related works

Recent developments in video based eye tracking concern the improvement of the reliability of the pupil signal. Summarizations for head mounted and remote eye tracking can be found in [26,59] and [20,49] respectively. One difference between head and remote eye tracking is the resolution of the eye images and the necessity for remote eye tracking to detect the face of a person and estimate its head pose. In head mounted eye tracking, the most successful approaches are based on edge detection [56,21,25,54], which allows to extract the pupil ellipses. This process is important for the validation [25,54] as well as the precision, since the ellipsis allows sub pixel accuracy. While edge based approaches are continuously improving, other attempts based on thresholding also have their advantages if edge detection is not applicable, i.e. for blurred and out of focus images. These attempts range from adaptive thresholding [32] to segment selection and combination [36]. Multistage approaches based on CNNs where also applied already [24,60,2]. They have the advantage to be applicable for a wider variety of challenges with the drawback of higher computational demands. Other approaches out of the realm of machine learning are based on random ferns [19] and oriented edge boosting [18]. Currently, the main disadvantage of all machine learning approaches so far is the lack in available pupil outline annotations for segmentation.

In the field of eyelid segmentation, edge detection was one of the first applied methods [62]. After the eyelid edges where found, the structure was approximated using parabolas for the upper and lower lid. Another edge based approach was proposed in [9]. Here, the iris was initially detected and afterwards the eyelids based on their distortion of the circular structure of the iris. The final approximation of the eyelids was done using splines. Similar to first eyelid extraction approach, the largest edges were selected with the difference of an anisotropic diffusion preprocessing [1]. Since blurred and out of focus images affect edge detection, a thresholding approach was proposed in [55]. After the separation of the image into regions based on a threshold, a likelihood map was computed using texture patches [63]. This approach was further developed using additional statistics for the computation of the likelyhood map [16]. VASIR [43] uses a linear Hough transform for iris segmentation in the first step. In the second step, a third order polynomial is fitted to edges above and below the iris for eyelid approximation. Machine learning approaches where also applied for eyelid detection [53] as well as histograms of oriented gradients and support vector machines [17]. Similar to BORE [18], an optimization was formulated to extract the eyelids based on oriented edge values [23].

The field of computer vision also achieved considerable progress in image segmentation. On the one hand it comes by the further development of the CNN architecture [34,58] and on the other hand from the development of transposed convolution filters [45]. Other approaches where also developed using the pooling indices of an encoder for up sampling but with less success [3]. Both used the soft max loss function to train their networks. The fully convolution approach was further developed by applying a region loss function [33,8,51] to achieve a higher accuracy at the segmentation borders. Unpaired training of generative adversarial networks (GANs) [30] together with the cycle loss function [65] was also applied on image segmentation, but the current state-of-the-art is based on the fully convolution architecture together with conditional random fields for refinement [7].

In this work, we use fully convolutional neural networks which have been proven to be the best performers on image segmentation [7,27]. Our network is jointly trained for segmentation and landmark regression to further improve the accuracy [64]. This network was used to generate a huge dataset together with manual correction of the found segmentations.

## 3 Method

Figure 1 shows the architecture of our CNN. Our network starts with a convolution and pooling layer. Afterward, residual layers with downscaling (ResidualD(L)) are used for feature encoding. The parameter L is the amount of learned filters. For further feature encoding, we used a residual block without scaling (Residual(L)). Outgoing from the result of this block, we regress the landmarks and the parameters of the pupil ellipse. In addition, we use the results of the central residual block for segmentation. Therefore, we use transposed convolutions in residual blocks (ResidualU(L)). The "ConvT." is a transposed convolution and is computed by upscaling the input by the specified stride. The final transposed convolution outputs the segmentation in two separated layers.

The design of the architecture has two purposes. First it allows the usage of only the regression part of the CNN, which makes it real time capable on a GPU ($\approx 6ms$), this runtime can be further reduce using binarized weights. In addition, the segmentation using the transposed convolution layers allows to increase the loss for training due to the pixel-wise comparison. It allows longer training period and, therefore, more accurate results.

Since the pupil parameters are numerically in no relation to landmarks and the additional segmentation, the loss function of the ellipse was reformulated into landmark form. Therefore, one ellipse is interpreted as n points (eight in our implementation). It allows to calculate the euclidean distance for each point, which serves as a loss function and is equivalent to the eyelid landmarks. For this it is necessary that each landmark can be assigned to a different one, for which we have used the orientation to the elliptical center. It means that each landmark on the ellipse corresponds to an angle ($0°, 45°, 90°, 135°, 180°, \ldots$ in our implementation).

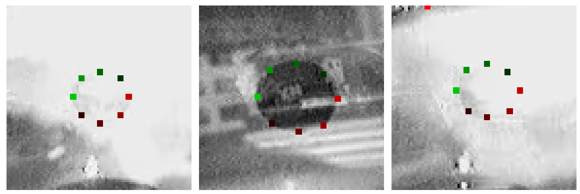

**Fig. 2.** The pupil ellipse as landmark representation. Landmarks in the same color belong to the same orientation.

Figure 2 shows the representation of the pupil ellipse as landmark. It means that our FC13 block in Figure 1 contains an internal block with sixteen inputs (elliptical landmarks) and five outputs (ellipse parameters). For the final ellipse parameter estimation we used the geometrical ellipse fitting approach proposed in [52].

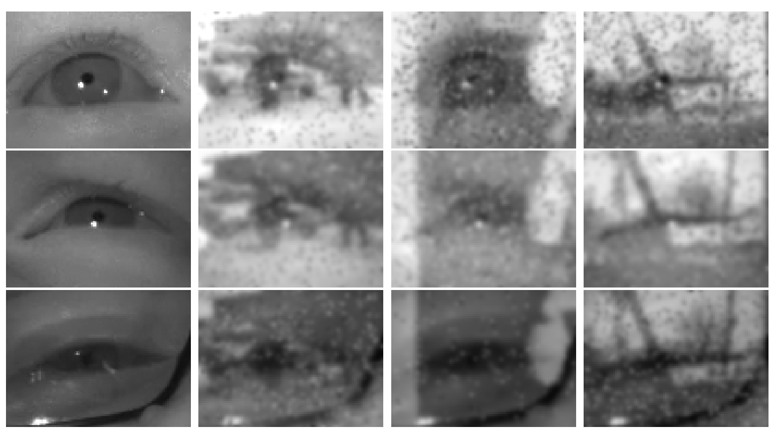

**Fig. 3.** Augmented training samples where the first column is the input image and the following rows are the image with added reflections, noise, adjusted contrast etc.

For training, we used a fixed learning rate of $10^{-8}$. The loss function for the fully connected layers regressing the landmarks and the ellipse was an L1 norm and for the segmentation we used a logarithmic loss. We trained our network for $\approx$ 2000 epochs, whereby we augmented the data online so that our network could not see the same image twice. For data augmentation, we used image shifts of up to 40% in each direction to cover images where the entire eye is not present. Scaling each axis between 0.8 and 1.2 was applied to cover more camera perspectives than present in the training dataset together with a rotation between -30 and 30 degrees. We also added random noise between 0 and 30% of the image. The contrast was changed between -30% and 30% of the intensity values. For occlusions, we added random patches which could cover up to 50% of the image. Reflections were added based on the approach from [61], where the reflection is assumed to be a blurred additive of a second image. Therefore, we used all the images from the PASCAL Visual Object Classes Challenge 2007 [15] as reflection pool from which we selected randomly. Those reflections were also shifted, rotated, and scaled with the same parameters as the original image. Examples of augmented data without shifting and scaling are shown in Figure 3. For the training dataset, we used the one published in [17], which consists of 16,200 hand-labeled images, where the eyelids and the pupil ellipsis was annotated. The recording system was a near-infrared remote camera in a driving simulator setting with a resolution of $1280 \times 752$ pixels.

## 4 Data set

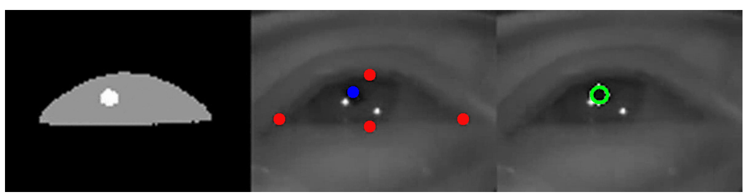

**Fig. 4.** Example result of our network with the segmentation, landmarks, and pupil ellipses in the same order.

The image source for our dataset was the same as for the ElSe [25] and ExCuSe [21] algorithm, which are the on-road recordings from [37]. The recording device during this study was a head-mounted camera system (Dikablis Mobile Eye Tracker by Ergoneers GmbH) with a frame rate of 25Hz and a resolution of $384 \times 288$. Our dataset contains 20 subjects and 501,230 images. Figure 5 shows some exemplary images with drawn eyelids (red) as well as the pupils ellipse (green). The contained challenges are noise, illumination changes, closed eyes, reflections, eyelashes, makeup, pupil dialation, motion blur, contact lens, and physiological characteristics (e.g. additional black dot on the iris). For the annotation, we used the proposed network (Figure 1). Afterwards, we used the Jaccard index ($\frac{A1 \cap A2}{A1 \cup A2}$) to evaluate the quality of the found annotations: 0.5 is

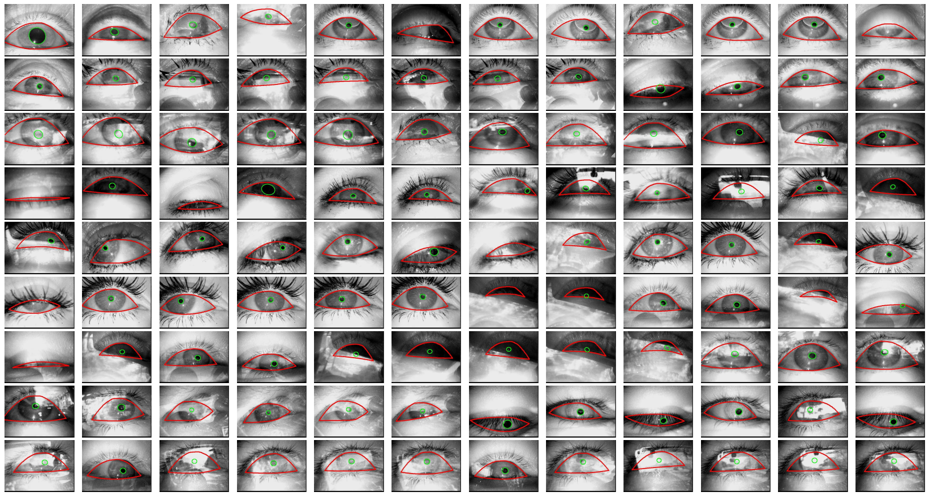

**Fig. 5.** Examples from the proposed dataset.

common as a good value for segmentations. A1, in our case, is the area of the segmentation and A2 is the area segmented by our landmarks and pupil ellipses. For interpolation of the landmarks we used natural Bezier splines. Figure 4 shows a segmentation on the left, where white is the pupil area and grey the eyelid area. The central image in Figure 4 shows the detected landmarks in red and the pupil center in blue. The pupil ellipses is drawn in green on the right of Figure 4. For finding false detections and false segmentations, we used the averaged sum of the Jaccard index of the segmented eyelid area and the area enclosed by the landmarks (red dots in Figure 4) as well as the pupil segment (white) and the area of the pupil ellipses. If the average Jaccard index was below 0.8, we inspected and corrected the image. Afterwards, we analysed all frames as video for mistakes. Overall, we had to correct $\approx 2\%$ of the eyelid and $\approx 4\%$ of the pupil annotations. The manual annotiation itself was done using a modified version of EyeLad [22], which allows to annotate the pupil ellipses and the eyelids with many supporting features like relative normalization, zooming, and tracking.

**Table 1.** Average Jaccard index per algorithm on the dataset from [25].

| Algorithm | Eyelid | Pupil |
|---|---|---|
| [39] | 41,45 | - |
| [23] | 52.89 | - |
| Proposed | 87.34 | 81.00 |

The runtime of our trained network is 30ms per image on an NVIDIA 1050Ti for the entire architecture. Since our network is capable of regressing landmarks and segment the image, the segmentation part can be omitted for a faster detection. It improves the runtime to 6ms per image, which allows a realtime usage with 160Hz using a GPU. The average Jaccard index per subject cross validated on the dataset proposed in [17] is 87.34 for the eyelid area and 81.00 for the pupil area (Table 1). The ExCuSe [21], ElSe [25], and PupilNet [24], Swirski [56], and

**Table 2.** Average detection result over all subjects on the publicly available datasets [21,25,24,56,59].

| Algo. | ElSe | ExCuSe | [56] | PURE | CBF | PNET | Prop. |
|---|---|---|---|---|---|---|---|
| [21,25,24] | 0.67 | 0.54 | 0.30 | 0.72 | 0.91 | 0.76 | **0.92** |
| [56] | 0.81 | 0.86 | 0.77 | 0.78 | - | - | **0.96** |
| [59] | 0.54 | 0.50 | 0.49 | 0.73 | - | - | **0.92** |

Labeled Pupils in the Wild [59] datasets were recorded using a head mounted eye tracker, we reach 92%, 96%, and 92% respectively for the pixel error up to 5 pixels which is the suggested pixel tolerance by the authors to compensate for inaccurate annotations (Table 2).

## 5 Conclusion

We propose a dataset for eyelid and pupil detection wherein the pupil ellipse as well as the enclosed eyelid area is segmented. The dataset contains more than 500,000 images from challenging real world recordings. In addition, we proposed a combined architecture for landmark regression and segmentation with a novel ellipse-to-landmark loss transformation. This network can also be used partially with a runtime of 6ms on the GPU for real time usage of up to 160 Hz. In addition, it can be used for dataset generation and data post processing. The dataset itself will help training and evaluating pupil and landmark detection algorithms, such as [39,19,18], which use less computational resources.

**Acknowledgments:** Work of the authors is supported by the Institutional Strategy of the University of Tübingen (Deutsche Forschungsgemeinschaft, ZUK 63). This research was supported by an IBM Shared University Research Grant including an IBM PowerAI environment. We especially thank our partners Benedikt Rombach, Martin Mähler and Hildegard Gerhardy from IBM for their expertise and support.

## References


1. Adam, M., Rossant, F., Amiel, F., Mikovikova, B., Ea, T.: Eyelid localization for iris identification. Radioengineering **17**(4), 82–85 (2008)
2. Anas, E.R., Henríquez, P., Matuszewski, B.J.: Online eye status detection in the wild with convolutional neural networks. In: VISIGRAPP (6: VISAPP). pp. 88–95 (2017)
3. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. arXiv preprint arXiv:1511.00561 (2015)
4. Benitezy, J.T.: Eye-tracking and optokinetic tests: Diagnostic significance in peripheral and central vestibular disorders. The Laryngoscope **80**(6), 834–848 (1970)
5. Boraston, Z., Blakemore, S.J.: The application of eye-tracking technology in the study of autism. The Journal of physiology **581**(3), 893–898 (2007)
6. Braunagel, C., Rosenstiel, W., Kasneci, E.: Ready for take-over? a new driver assistance system for an automated classification of driver take-over readiness. IEEE Intelligent Transportation Systems Magazine **9**(4), 10–22 (2017)
7. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence **40**(4), 834–848 (2017)
8. Dai, J., He, K., Li, Y., Ren, S., Sun, J.: Instance-sensitive fully convolutional networks. In: European Conference on Computer Vision. pp. 534–549. Springer (2016)
9. Daugman, J.: How iris recognition works. In: The essential guide to image processing, pp. 715–739. Elsevier (2009)
10. Dong, W., Qu, P.: Eye state classification based on multi-feature fusion. In: Control and Decision Conference, 2009. CCDC'09. Chinese. pp. 231–234. IEEE (2009)
11. Duchowski, A.T.: A breadth-first survey of eye-tracking applications. Behavior Research Methods, Instruments, & Computers **34**(4), 455–470 (2002)
12. Duchowski, A.T., Shivashankaraiah, V., Rawls, T., Gramopadhye, A.K., Melloy, B.J., Kanki, B.: Binocular eye tracking in virtual reality for inspection training. In: Proceedings of the 2000 symposium on Eye tracking research & applications. pp. 89–96. ACM (2000)
13. Eivazi, S., Bednarik, R., Leinonen, V., von und zu Fraunberg, M., Jääskeläinen, J.E.: Embedding an eye tracker into a surgical microscope: requirements, design, and implementation. IEEE Sensors Journal **16**(7), 2070–2078 (2016)
14. Eivazi, S., Bednarik, R., Tukiainen, M., von und zu Fraunberg, M., Leinonen, V., Jääskeläinen, J.E.: Gaze behaviour of expert and novice microneurosurgeons differs during observations of tumor removal recordings. In: Proceedings of the Symposium on Eye Tracking Research and Applications. pp. 377–380. ACM (2012)
15. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. International journal of computer vision **88**(2), 303 (2010)
16. Fuhl, W., Santini, T., Geisler, D., Kübler, T., Rosenstiel, W., Kasneci, E.: Eyes wide open? eyelid location and eye aperture estimation for pervasive eye tracking in real-world scenarios. In: PETMEI (09 2016)
17. Fuhl, W., Castner, N., Zhuang, L., Holzer, M., Rosenstiel, W., Kasneci, E.: Mam: Transfer learning for fully automatic video annotation and specialized detector creation. In: Egocentric Perception, Interaction and Computing Workshop (EPIC@ECCV) (09 2018)

18. Fuhl, W., Eivazi, S., Hosp, B., Eivazi, A., Rosenstiel, W., Kasneci, E.: Bore: boosted-oriented edge optimization for robust, real time remote pupil center detection. In: Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications. p. 48. ACM (2018)
19. Fuhl, W., Geisler, D., Santini, T., Appel, T., Rosenstiel, W., Kasneci, E.: Cbf: circular binary features for robust and real-time pupil center detection. In: Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications. p. 8. ACM (2018)
20. Fuhl, W., Geisler, D., Santini, T., Rosenstiel, W., Kasneci, E.: Evaluation of state-of-the-art pupil detection algorithms on remote eye images. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. pp. 1716–1725. ACM (2016)
21. Fuhl, W., Kübler, T., Sippel, K., Rosenstiel, W., Kasneci, E.: Excuse: Robust pupil detection in real-world scenarios. In: International Conference on Computer Analysis of Images and Patterns. pp. 39–51. Springer (2015)
22. Fuhl, W., Santini, T., Geisler, D., Kübler, T., Kasneci, E.: Eyelad: Remote eye tracking image labeling tool. In: 12th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017) (02 2017)
23. Fuhl, W., Santini, T., Kasneci, E.: Fast and robust eyelid outline and aperture detection in real-world scenarios. In: Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on. pp. 1089–1097. IEEE (2017)
24. Fuhl, W., Santini, T., Kasneci, G., Kasneci, E.: Pupilnet: convolutional neural networks for robust pupil detection. arXiv preprint arXiv:1601.04902 (2016)
25. Fuhl, W., Santini, T.C., Kübler, T., Kasneci, E.: Else: Ellipse selection for robust pupil detection in real-world environments. In: Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications. pp. 123–130. ACM (2016)
26. Fuhl, W., Tonsen, M., Bulling, A., Kasneci, E.: Pupil detection for head-mounted eye tracking in the wild: an evaluation of the state of the art. Machine Vision and Applications **27**(8), 1275–1288 (2016)
27. Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Garcia-Rodriguez, J.: A review on deep learning techniques applied to semantic segmentation. arXiv preprint arXiv:1704.06857 (2017)
28. Gegenfurtner, A., Lehtinen, E., Säljö, R.: Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. Educational Psychology Review **23**(4), 523–552 (2011)
29. Gilzenrat, M.S., Nieuwenhuis, S., Jepma, M., Cohen, J.D.: Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. Cognitive, Affective, & Behavioral Neuroscience **10**(2), 252–269 (2010)
30. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems. pp. 2672–2680 (2014)
31. Guenter, B., Finch, M., Drucker, S., Tan, D., Snyder, J.: Foveated 3d graphics. ACM Transactions on Graphics (TOG) **31**(6), 164 (2012)
32. Haro, A., Flickner, M., Essa, I.: Detecting and tracking eyes by using their physiological properties, dynamics, and appearance. In: Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on. vol. 1, pp. 163–168. IEEE (2000)
33. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Computer Vision (ICCV), 2017 IEEE International Conference on. pp. 2980–2988. IEEE (2017)

34. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
35. Holzman, P.S., Proctor, L.R., Levy, D.L., Yasillo, N.J., Meltzer, H.Y., Hurt, S.W.: Eye-tracking dysfunctions in schizophrenic patients and their relatives. Archives of general psychiatry **31**(2), 143–151 (1974)
36. Javadi, A.H., Hakimi, Z., Barati, M., Walsh, V., Tcheang, L.: Set: a pupil detection method using sinusoidal approximation. Frontiers in neuroengineering **8**, 4 (2015)
37. Kasneci, E., Sippel, K., Aehling, K., Heister, M., Rosenstiel, W., Schiefer, U., Papageorgiou, E.: Driving with binocular visual field loss? a study on a supervised on-road parcours with simultaneous eye and head tracking. PloS one **9**(2), e87470 (2014)
38. Kassner, M., Patera, W., Bulling, A.: Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In: Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: Adjunct publication. pp. 1151–1160. ACM (2014)
39. Kazemi, V., Sullivan, J.: One millisecond face alignment with an ensemble of regression trees. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1867–1874 (2014)
40. Krumpe, T., Scharinger, C., Gerjets, P., Rosenstiel, W., Spüler, M.: Disentangeling working memory loadfinding inhibition and updating components in eeg data. In: Proceedings of the 6th International Brain-Computer Interface Meeting: BCI Past, Present, and Future. p. 174 (2016)
41. Lappi, O.: Eye movements in the wild: oculomotor control, gaze behavior & frames of reference. Neuroscience & Biobehavioral Reviews **69**, 49–68 (2016)
42. LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: Backpropagation applied to handwritten zip code recognition. Neural computation **1**(4), 541–551 (1989)
43. Lee, Y., Micheals, R.J., Filliben, J.J., Phillips, P.J.: Vasir: an open-source research platform for advanced iris recognition technologies. Journal of research of the National Institute of Standards and Technology **118**, 218 (2013)
44. Liu, X., Xu, F., Fujimura, K.: Real-time eye detection and tracking for driver observation under various light conditions. In: Intelligent Vehicle Symposium, 2002. IEEE. vol. 2, pp. 344–351. IEEE (2002)
45. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440 (2015)
46. Marshall, S.P.: Identifying cognitive state from eye metrics. Aviation, space, and environmental medicine **78**(5), B165–B175 (2007)
47. Matsushita, M.: Iris identification system and iris identification method (May 4 1999), uS Patent 5,901,238
48. Palinko, O., Kun, A.L., Shyrokov, A., Heeman, P.: Estimating cognitive load using remote eye tracking in a driving simulator. In: Proceedings of the 2010 symposium on eye-tracking research & applications. pp. 141–144. ACM (2010)
49. Park, S., Zhang, X., Bulling, A., Hilliges, O.: Learning to find eye region landmarks for remote gaze estimation in unconstrained settings. In: Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications. p. 21. ACM (2018)
50. Patney, A., Salvi, M., Kim, J., Kaplanyan, A., Wyman, C., Benty, N., Luebke, D., Lefohn, A.: Towards foveated rendering for gaze-tracked virtual reality. ACM Transactions on Graphics (TOG) **35**(6), 179 (2016)

51. Pinheiro, P.O., Lin, T.Y., Collobert, R., Dollár, P.: Learning to refine object segments. In: European Conference on Computer Vision. pp. 75–91. Springer (2016)
52. Prasad, D.K., Leung, M.K., Quek, C.: Ellifit: an unconstrained, non-iterative, least squares based geometric ellipse fitting method. Pattern Recognition **46**(5), 1449–1465 (2013)
53. Ren, S., Cao, X., Wei, Y., Sun, J.: Face alignment at 3000 fps via regressing local binary features. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1685–1692 (2014)
54. Santini, T., Fuhl, W., Kasneci, E.: Pure: Robust pupil detection for real-time pervasive eye tracking. Computer Vision and Image Understanding (2018)
55. Suzuki, M., Yamamoto, N., Yamamoto, O., Nakano, T., Yamamoto, S.: Measurement of driver's consciousness by image processing-a method for presuming driver's drowsiness by eye-blinks coping with individual differences. In: SMC. vol. 4, pp. 2891–2896. IEEE (2006)
56. Świrski, L., Bulling, A., Dodgson, N.: Robust real-time pupil tracking in highly off-axis images. In: Proceedings of the Symposium on Eye Tracking Research and Applications. pp. 173–176. ACM (2012)
57. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: AAAI. vol. 4, p. 12 (2017)
58. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1–9 (2015)
59. Tonsen, M., Zhang, X., Sugano, Y., Bulling, A.: Labelled pupils in the wild: a dataset for studying pupil detection in unconstrained environments. In: Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications. pp. 139–142. ACM (2016)
60. Vera-Olmos, F., Malpica, N.: Deconvolutional neural network for pupil detection in real-world environments. In: International Work-Conference on the Interplay Between Natural and Artificial Computation. pp. 223–231. Springer (2017)
61. Wan, R., Shi, B., Duan, L.Y., Tan, A.H., Kot, A.C.: Benchmarking single-image reflection removal algorithms. In: Proc. ICCV (2017)
62. Wildes, R.P.: Iris recognition: an emerging biometric technology. Proceedings of the IEEE **85**(9), 1348–1363 (1997)
63. Yang, F., Yu, X., Huang, J., Yang, P., Metaxas, D.: Robust eyelid tracking for fatigue detection. In: ICIP. pp. 1829–1832 (Sept 2012)
64. Zhang, K., Zhang, Z., Li, Z., Qiao, Y.: Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters **23**(10), 1499–1503 (2016)
65. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. arXiv preprint (2017)