GroupGazer: A Tool to Compute the Gaze per Participant in Groups with integrated Calibration to Map the Gaze Online to a Screen or Beamer Projection

Wolfgang Fuhl
Department of Human Computer Interaction
University Tübingen
Tübingen, 72076
wolfgang.fuhl@uni-tuebingen.de

January 20, 2022

Abstract

In this paper we present GroupGaze. It is a tool that can be used to calculate the gaze direction and the gaze position of whole groups. GroupGazer calculates the gaze direction of every single person in the image and allows to map these gaze vectors to a projection like a projector. In addition to the person-specific gaze direction, the person affiliation of each gaze vector is stored based on the position in the image. Also, it is possible to save the group attention after a calibration. The software is free to use and requires a simple webcam as well as an NVIDIA GPU and the operating system Windows or Linux.
https://atreus.informatik.uni-tuebingen.de/seafile/d/8e2ab8c3fdd444e1a135/?p=%2FGG&mode=list.

1 Introduction

Eye tracking is an important input modality and information source in the modern world [12, 39, 19, 40, 42]. The signal itself is usually computed by first detecting the pupil [43, 54, 58, 53, 52, 30] and additional eye features [37, 28, 46, 24]. Based on the gaze signal it is possible to extract further information like the eye movement types [44, 55, 21, 22, 31, 34]. In the field of human-machine interaction [20], the gaze signal is used and further researched for interaction with robots [112] but also other technical devices [111, 18, 23, 45]. This involves not only simple control but also collaboration in which a human communicates complex behavior to a robot or system [95]. Interaction with the eyes based on pupil movements [56, 41, 29, 27, 25, 57] is also an interesting source of information in the field of computer games [2]. Eye interaction is many times faster than mouse interaction, which could revolutionize the professional computer gaming field [73]. In the field of virtual reality, gaze information can be used to render only small areas of the scene in high resolution, leading to a significant reduction in the resources consumption of the devices [91]. Another important area in which the gaze signal plays an important role is driver observation. Here it is necessary to assess whether the driver is able to control the vehicle or is too tired in the case of autonomous driving to take over the vehicle [118]. Of course, this also applies to car rental companies, for which it is important to know whether the driver is, for example, intoxicated or an unsafe driver [88, 49, 48, 50, 26, 51]. In the field of medicine, research is also being conducted into methods of self-diagnosis [10]. This involves, for example, the early detection of Alzheimer's disease [13], strokes [87], as well as eye

Figure 1: Detection results of the proposed tool GroupGazer on a group photo.

defects [16] or autism [6]. In the field of safety, the eye signal also gains increasingly more interest [75, 60]. This is due to the fact that personal behavior is reflected in the gaze signal, which can be used to identify the person [47, 60]. Other information contained in the eye is the cognitive load based on pupil dilation [7], attention [9], procedural strategies [71] and many others. A relatively new area in which the eye tracking signal is used is behavioral research [117, 14]. Here it is on the one hand about extracting expert knowledge from the eye signal and passing this knowledge to trainees [85, 65]. This concerns, all areas in which the training is only possible with expensive tools and training devices [109]. In the area of medicine the main interest is to distill the expert knowledge better [96, 85]. Another area of behavioral research which is also the subject of this thesis is group behavior [69, 100, 81]. Here there is research in the area of teaching [79, 105, 70] but also in dynamic environments like sports [94, 15, 102].

The current problems in the field of behavioral research for groups, is that there is no freely available software for this. Therefore, research groups have to resort to expensive solutions such as multiple worn eye trackers. This creates further issues like the assignment of the important areas between the different scene cameras. One way around this is to use virtual reality together with eye tracking. However, this also changes the behavior of the test persons and cannot be carried out over longer periods of time with regard to motion sickness. Alternatively to worn eye trackers, there is also the possibility to use external cameras. In this case, the researchers have to implement their technical solutions independently, which often leads to dependencies on other working groups and is also an expensive undertaking due to the image processing cameras which are usually used.

In this paper, we present software that allows anyone to use a simple webcam for gaze estimation of groups and calibration each subject in parallel. By doing so, we hope to enable anyone to conduct behavioral group research. Our contributions to the state of the art are:

1. A tool to record the gaze of groups and calibrate each individual in parallel.
2. The tool has no specialized hardware requirements and only needs an NVIDIA GPU with at least 4 GB memory (We used a 1050 ti with 4 GB).
3. Stores the gaze per person as well as the average gaze location of the group.

2 Related work

Since our software is the combination of several research fields, we have divided the related work into three categories. The first category is face recognition, the second category is appearance based gaze estimation, and the third category is gaze based group behavior research.

2.1 Face detection

Face recognition in arbitrary environments is still a very challenging field of research. Here, an arbitrarily large image is given, and all faces must be detected. This often involves occlusions, different head positions, changes in lighting conditions, and of course the faces in the image have different resolutions. The first very successful approach was presented by Viola and Jones [110]. This is based on hair features and trained using AdaBoost. The next major step was achieved with deformable part models (DPM) [116]. Compared to feature-based approaches, DPM is much more robust but requires significantly more computational effort. With the advent of deep neural networks, however, the state of the art was again significantly improved [4, 72, 82, 122]. The first extension of neural networks was the combination of face detection with face matching [119, 120]. Current methods for face detection follow two directions. The first direction is the multistage approach, which is based on a region proposal neural network followed by validation of the proposed faces. The most notable representatives of this direction of development are RCNN [63], almost RCNN [62], and faster RCNN [101]. The second direction of development is direct methods such as single shot multibox detector (SSD) [83] or you only look once (YOLO) [99]. For YOLO, there are already multiple versions, which consume even less resources at approximately the same detection rate. The advantage of the direct approaches, is the faster execution and the smaller resource consumption. The multilayer methods, on the other hand, provide a better detection rate and fewer misclassifications. Further ressource consumption approaches have also been proposed in the literature [38, 33, 35].

2.2 Appearance based gaze estimation

Here, the entire facial image or eye area of a person is used to directly determine the gaze vector via a neural network. The first work in this area is from 1994 [5] and was extended in [108] by linear projection functions. These methods require very expensive calibration, since the neural network was trained for each person individually with many training examples. The first extensions to reduce the effort in calibration were a Gaussian process regression [113], saliency maps [107], and optimal selection using a linear regression [84]. While all of these methods advanced the state of the art, the appearance based approach still had many limitations, such as a fixed head position and per-person calibration. With deep neural networks and the advent of big data, this has changed significantly. In [121] the first successful approach was presented, which realized appearance based gaze estimation with deep neural networks. The first extensions used, in addition to eye images, the subjects' faces, which resulted in a significant improvement [80, 76]. For extreme head positions and strongly deviating gaze angles to the head orientation, an asymmetric regression was presented [8].

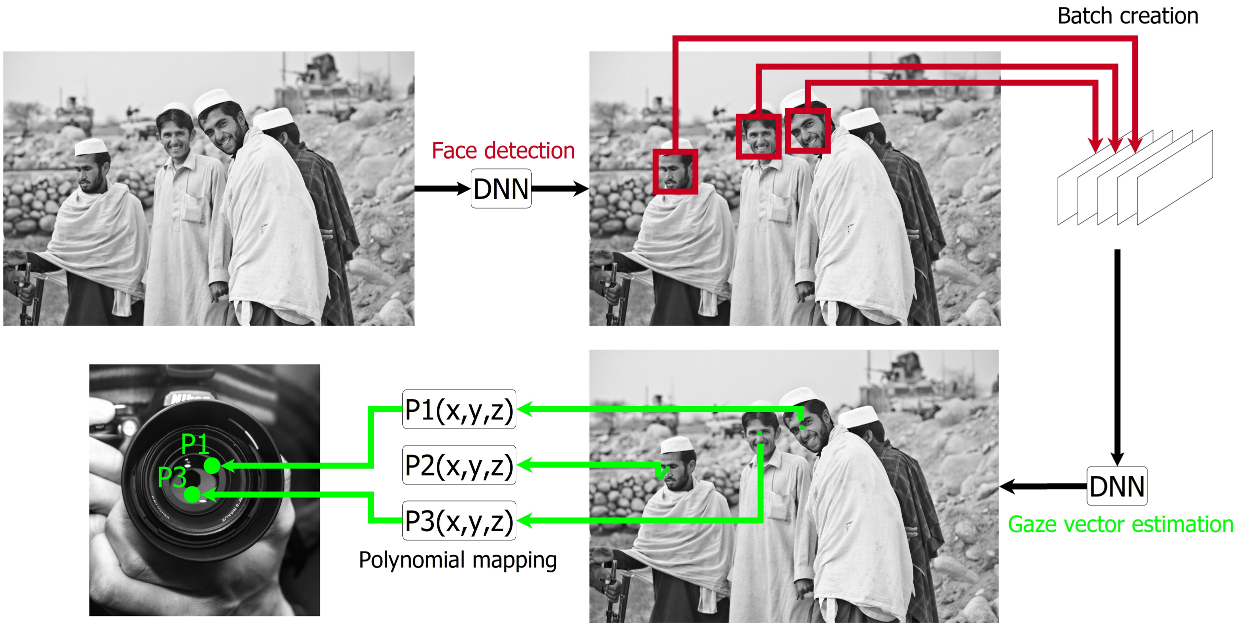

Figure 2: The workflow of our approach. We first detect the faces and compute the gaze vector using an appearance based approach. Each person is calibrated using mouse clicks on a projection in parallel. The fitted polynomials are afterwards used to map the gaze to the projection.

2.3 Gaze based group behavior research

In this section, we would like to mention and briefly explain only some works from this area, since our software is made for this purpose but does not perform a behavior research study.

The first area in behavioral research which can also be applied to groups is mind wandering [68, 67]. Mind wandering is a shift in attention to task-unrelated thoughts. This is an interesting effect for teaching since it negatively influences the learning performance of students [103, 106, 67]. Mind wandering itself is a special form of disengagement and has to be separated from boredom or off-task behaviors [11, 92, 67]. Another interesting social behavior is gaze following [3]. This gaze following is a form of communication and socializing. In some scenarios it has to be done only for a single person [74, 17] but in modern research entire scenes with multiple persons are evaluated and analyzed [3, 93, 86, 97, 98]. Nowadays, psychologists use behavior observation methods in classrooms as well as direct behavior ratings [114]. While both methods are valid and also used by teachers themselves, they are limited in effectiveness due to the attentional limits of the human observers as well as their induced biases [89]. Modern research focuses on establishing intelligent classroom technologies with eye tracking and voice recording [114, 89]. Those methods have their limitations due to the data security but deliver more insights and allow reducing the induced bias by humans [114, 89, 90].

3 Method

Table 1: Shows the architecture of our face detection deep neural network. The architecture is copied from dlib [77] and uses the max margin [78] training procedure. We modified the model in terms of tensor normalization [36] and gradient centralization [61] as well as convolution size and depth.

Level	Gaze estimator
Input	RGB image any resolution
1	Pyramid layer with six stages
2	5×5 Conv, depth 8, 2×2 down scaling, BN, ReLu with tensor normalization
3	3×3 Conv, depth 8, 2×2 down scaling, BN, ReLu with tensor normalization
4	3×3 Conv, depth 8, 2×2 down scaling, BN, ReLu with tensor normalization
5	5×5 Conv, depth 16, BN, ReLu with tensor normalization
6	3×3 Conv, depth 16, BN, ReLu with tensor normalization
7	3×3 Conv, depth 16, BN, ReLu with tensor normalization
8	7×7 Conv, depth 1

Figure 2 shows the workflow of our approach. GroupGazer first opens a video stream on an available camera. Afterwards, all faces in the image are detected. If not all desired faces are detected, GroupGazer offers an upscaling factor, which can be set by the user. This upscaling factor resizes the input image to allow the face detection to detect

Table 2: Shows the architecture of our gaze estimation deep neural network. It has the structure of a ResNet-34 [64] and uses the leaky maximum propagation blocks [59], tensor normalization [36], as well as the weight and gradient centralization [61].

Level	Gaze estimator
Input	Gray scale image 100×100
1	5×5 Conv, depth 32
2	ReLu with tensor normalization
3	2×2 Max pooling
4	3 Maxium connection blocks, 2×2 down scaling, 3×3 Conv, depth 64, BN
5	ReLu with tensor normalization
6	3 Maxium connection blocks, 2×2 down scaling, 3×3 Conv, depth 128, BN
7	ReLu with tensor normalization
8	3 Maxium connection blocks, 2×2 down scaling, 3×3 Conv, depth 256, BN
9	ReLu with tensor normalization
10	Fully connected, 512 outputs
11	ReLu
12	Fully connected, 7 outputs (3 and 7 are the accuracy of the estimation [32])

even very small faces in the image. After the face detection, all detected faces are extracted from the image and resized to 100×100 pixels in a gray scale image. These images are grouped together to form a batch which is given to the gaze vector estimation DNN. The batch size can also be set by the user. This fixed batch size allows GroupGazer to have a static runtime and if there are fewer faces in the image, the rest of the batch is filled with black images. GroupGazer can be used with a 1050 ti graphics card for up to 40 faces in real time, which is also dependent on the input resolution to the face detection DNN. For newer GPUs more faces can be set by the user as well as larger input image resolutions for the face detection. The gaze estimation DNN processes the entire batch and computes a starting position (First two values), an accuracy of the starting position (Third value), the gaze vector (Forth to sixth value), as well as an accuracy of the gaze vector (Seventh value). With this information, each face has a gaze vector and an estimated accuracy. With the gaze vector and the starting position, a polynomial is used to map the gaze vector to a projection or monitor. The degree of the polynomial can be specified by the user, and the calibration procedure works as follows. The teacher or adviser tells the students to look at his mouse cursor position. On a left mouse click, all gaze vectors which are seen as valid and accurate are stored together with the click location. This is repeated multiple times. Afterwards, for each user, the polynomial is fitted in the least squares sense. With those polynomials, the mapping and therefore the gaze location is computed for each user. The reidentification of users is done by the smallest euclidean distance to the last detections, and the new position is not allowed to leaf the last face detection bounding box. This is a simple procedure but saves a lot of computational resources since no additional network has to be used. In addition, it is much more robust since fine-tuning a Network online usually needs multiple examples to deliver reliable results, even if we use the hypersphere approach [115] or siam networks [1].

The used model architectures can be seen in Table 1 and 2. Our face detection model is similar to the model from dlib [77] we only made some slight changes which improve the accuracy of the model and only impact the runtime slightly. For gaze estimation we used the architecture of a ResNet-34 [64] since it has a good accuracy and is resource saving in contrast to the other networks. We modified the ResNet-34 architecture only by adding some novel normalization [36, 61], the landmark validation loss [32], as well as leaky maximum propagations instead of the residual connections [59].

4 Evaluation

Gaze360 [76] is a huge data set with 3D gaze annotations recorded using multiple cameras covering 360 degree. The recordings were conducted indoor and outdoor with 238 subjects. The dataset contains large head variations as well as distances of the subjects to the camera. We only used approximately 80,000 images of this data set since the data set contains also human heads from behind as well as some partially covered heads which we removed from our data for training and evaluation. The train and test split was done by randomly selecting 20% for testing and 80% for training.

DLIB [77] data set contains images of various resolutions. Each image can have multiple faces which are annotated with bounding boxes. In total the data set contains 7213 images and 11480 annotated faces. We made a random 50% to 50% split and used the first half for training and the second half for evaluation (One image more for training due to the uneven number). The images in the dataset are taken from other public data sets and annotated by the authors of [77].

In Table 3 and 4 our models are compared with other approaches. For face detection (Table 3), it can be seen that we have chosen a tradeoff between detection rate and runtime. The recognition rate of our approach can be further increased via the upscaling factor. However, this also increases the computation time, which also increases the runtime per image. For Yolo this is not possible, because the memory usage for images larger than 300 becomes too large. For the backbone of the faster-RCNN, the memory consumption is also too high for a large resolution. Which is also the

Figure 3: Qualitative evaluation on different images from pixabay.com, which are free to use. Left original image, right detections. **The images have a high resolution in the pdf so you can zoom in to see everything.**

Table 3: Face detection results on DLIB data set [77] with percition and recall. We compare our model to other approaches in therms of detection percentage as well as runtime in milliseconds (ms) for one hundred images in average. OoM means out of memory exception.

Method	Percision	Recall	Runtime GPU (ms)	
			1920×1280	300×300
Proposed	0,99	0,89	67	3
dlib std. arch. [77]	0,99	0,88	175	8
ResNet-34 & Faster-RCNN [101]	0,99	0,91	OoM	22 & 1
Yolov5s [99]	0,99	0,89	OoM	10

Table 4: Appearance based gaze estimation results on the Gaze360 [76] dataset. We compared our model to other approaches and evaluated the gaze start estimation in average euclidean distance in pixel as well as the gaze vector estimation in degree. Time is measured for one face image as average over one thousand.

Method	Gaze start	Gaze vector	Runtime GPU (ms)
Proposed	0,6	0,2	3
ResNet-34 [64]	0,9	0,5	8
ResNet-50 [64]	0,5	0,2	12
MobileNet [66]	1,8	1,6	7
MobileNetv2 [104]	1,7	1,6	7

main reason why we decided against YOLO and the faster-RCNN. In addition, both the faster-RCNN with backbone and the YOLO need a fixed input resolution with which they have to be trained. For our fully convolutional approach inspired by the dlib architecture, this is not necessary.

For the gaze direction determination, you can clearly see that our net runs significantly faster than the other nets. This is due to the fact that our layers use less depth than, for example, ResNet-34. The MobileNets cannot show their advantage on the GPU, since they cause cache conflicts here, whereby parts of the code are executed serialized. On a CPU, MobileNet would be significantly faster than our net, but with about 160 ms per face too slow for a real-time evaluation. In terms of results, ResNet-50 is the most accurate, closely followed by our network. In addition to accuracy, if we consider runtime on a GPU, our network is clearly ahead, which is why we chose our architecture.

The accuracy of our approach for different distances can be seen in Table 5. As can be seen, the distance has a huge impact on the accuracy. For a 6-meter distance, the error of 19 cm is approximately 20% of the projection area. For a beamer projection this would not be as crucial since the diagonal here is usually 2 to 3 meters or more. Therefore, our system is capable of computing useful gaze positions, but does not have the accuracy of a professional remote eye tracker which can only be used from persons in front of the screen. On the other side, our system can compute and map the gaze of multiple persons in parallel and is therefore made for classrooms or meetings.

In addition to the quantitative analysis from Table 3 and 4, we also made a qualitative analysis. The images in Figure 3 have a very high resolution in the PDF, so everything can be seen in detail when using the zoom function. Also, all images are in supplementary material. Looking at the first image in Figure 3 you can see that our mesh can also appreciate the head-only pose when the eyes are covered. On the second, third and last image, one can see that our network works with both color and gray scale images. Likewise, it is capable of good results on crowd gatherings, as can be seen in the fourth image.

5 Limitations

One limitation of the presented software is that persons are not recognized via a neural network, but are assigned based on their last position. This has the limitation that the persons cannot move freely, but the software is only suitable for slightly dynamic scenarios like a classroom. In the future, recognition, which is adapted online, will be integrated. Another limitation of our software is that currently only gaze data is extracted. Since other features such as the eyelids as well as a more accurate gaze determination via the pupils are also interesting for research, these will also be integrated in future updates. The final limitation of our software is the need for a GPU. This limitation will not be changed in the near future, since the problem of face recognition and fast gaze determination can be solved by cheaper

Table 5: Accuracy of the proposed tool for different distances to the camera. The results are the average accuracy over three subjects on a TV screen with a diagonal of 108cm.

1m	2m	3m	4m	5m	6m
4cm	5cm	8cm	12cm	15cm	19cm

methods, but these have strong losses in detection rate and accuracy.

6 Conclusion

In this paper, we have presented GroupGazer. This is a software that allows to determine the gaze direction of groups per person. This gaze determination is done online on a conventional computer with an NVIDIA GPU. GroupGazer allows each person in the group to be calibrated in parallel so that the individual gaze vectors can be mapped to a projection, such as that of a projector or large monitor. The software is intended to support behavioral research and thus make it possible to easily record the gaze positions of groups. Future enhancements to the software include person recognition for dynamic processes such as sports, pose determination, emotion determination, advanced feature extraction such as eyelids, and more accurate gaze determination based on eye features such as the pupil. We hope that the software will help other researchers in their work and further advance behavioral research as well as the application area of group-based gaze determination.

References

[1] Mohamed H Abdelpakey and Mohamed S Shehata. Dp-siam: Dynamic policy siamese network for robust object tracking. *IEEE Transactions on Image Processing*, 29:1479–1492, 2019.

[2] Serkan Alkan and Kursat Cagiltay. Studying computer game learning experience through eye tracking. *British Journal of Educational Technology*, 38(3):538–542, 2007.

[3] Arkar Min Aung, Anand Ramakrishnan, and Jacob R Whitehill. Who are they looking at? automatic eye gaze following for classroom observation video analysis. *International Educational Data Mining Society*, 2018.

[4] Yancheng Bai, Yongqiang Zhang, Mingli Ding, and Bernard Ghanem. Finding tiny faces in the wild with generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 21–30, 2018.

[5] Shumeet Baluja and Dean Pomerleau. Non-intrusive gaze tracking using artificial neural networks. Technical report, CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF COMPUTER SCIENCE, 1994.

[6] Zillah Boraston and Sarah-Jayne Blakemore. The application of eye-tracking technology in the study of autism. *The Journal of physiology*, 581(3):893–898, 2007.

[7] Margot Chauliac, Leen Catrysse, David Gijbels, and Vincent Donche. It is all in the" surv-eye": Can eye tracking data shed light on the internal consistency in self-report questionnaires on cognitive processing strategies?. *Frontline Learning Research*, 8(3):26–39, 2020.

[8] Yihua Cheng, Xucong Zhang, Feng Lu, and Yoichi Sato. Gaze estimation by exploring two-eye asymmetry. *IEEE Transactions on Image Processing*, 29:5259–5272, 2020.

[9] Meia Chita-Tegmark. Social attention in asd: A review and meta-analysis of eye-tracking studies. *Research in developmental disabilities*, 48:79–93, 2016.

[10] Rosie Clark, James Blundell, Matt J Dunn, Jonathan T Erichsen, Mario E Giardini, Irene Gottlob, Chris Harris, Helena Lee, Lee Mcilreavy, Andrew Olson, et al. The potential and value of objective eye tracking in the ophthalmology clinic. *Eye*, 33(8):1200–1202, 2019.

[11] Mihaela Cocea and Stephan Weibelzahl. Disengagement detection in online learning: Validation studies and perspectives. *IEEE transactions on learning technologies*, 4(2):114–124, 2010.

[12] Matteo Cognolato, Manfredo Atzori, and Henning Müller. Head-mounted eye gaze tracking devices: An overview of modern devices and recent advances. *Journal of rehabilitation and assistive technologies engineering*, 5:2055668318773991, 2018.

[13] Trevor J Crawford. The disengagement of visual attention in alzheimer's disease: a longitudinal eye-tracking study. *Frontiers in aging neuroscience*, 7:118, 2015.

[14] Marcel Das, Peter Ester, and Lars Kaczmirek. *Social and behavioral research and the internet: Advances in applied methods and research strategies.* Routledge, 2018.

[15] Peet J Du Toit, Pieter Ernst Kruger, NZ Chamane, Jolene Campher, and Dalene Crafford. Sport vision assessment in soccer players and sport science. *African Journal for Physical Health Education, Recreation and Dance*, 15(4):594–604, 2009.

[16] Mads Gjerstad Eide, Ruben Watanabe, Ilona Heldal, Carsten Helgesen, Atle Geitung, and Harald Soleim. Detecting oculomotor problems using eye tracking: Comparing eyex and tx300. In *2019 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pages 381–388. IEEE, 2019.

[17] Alircza Fathi, Jessica K Hodgins, and James M Rehg. Social interactions: A first-person perspective. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1226–1233. IEEE, 2012.

[18] E. Fuhl, W.and Bozkir, B. Hosp, N. Castner, D. Geisler, T. Santini, and E. Kasneci. Encodji: encoding gaze data into emoji space for an amusing scanpath classification approach. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, pages 1–4, 2019.

[19] W. Fuhl. *Image-based extraction of eye features for robust eye tracking.* PhD thesis, University of Tübingen, 04 2019.

[20] W. Fuhl. From perception to action using observed actions to learn gestures. *User Modeling and User-Adapted Interaction*, pages 1–18, 08 2020.

[21] W. Fuhl, N. Castner, and E. Kasneci. Histogram of oriented velocities for eye movement detection. In *International Conference on Multimodal Interaction Workshops, ICMIW*, 2018.

[22] W. Fuhl, N. Castner, and E. Kasneci. Rule based learning for eye movement type detection. In *International Conference on Multimodal Interaction Workshops, ICMIW*, 2018.

[23] W. Fuhl, N. Castner, T. C. Kübler, A. Lotz, W. Rosenstiel, and E. Kasneci. Ferns for area of interest free scanpath classification. In *Proceedings of the 2019 ACM Symposium on Eye Tracking Research & Applications (ETRA)*, 06 2019.

[24] W. Fuhl, N. Castner, L. Zhuang, M. Holzer, W. Rosenstiel, and E. Kasneci. Mam: Transfer learning for fully automatic video annotation and specialized detector creation. In *International Conference on Computer Vision Workshops, ICCVW*, 2018.

[25] W. Fuhl, S. Eivazi, B. Hosp, A. Eivazi, W. Rosenstiel, and E. Kasneci. Bore: Boosted-oriented edge optimization for robust, real time remote pupil center detection. In *Eye Tracking Research and Applications, ETRA*, 2018.

[26] W. Fuhl, H. Gao, and E. Kasneci. Neural networks for optical vector and eye ball parameter estimation. In *ACM Symposium on Eye Tracking Research & Applications, ETRA 2020.* ACM, 01 2020.

[27] W. Fuhl, H. Gao, and E. Kasneci. Tiny convolution, decision tree, and binary neuronal networks for robust and real time pupil outline estimation. In *ACM Symposium on Eye Tracking Research & Applications, ETRA 2020.* ACM, 01 2020.

[28] W. Fuhl, D. Geisler, W. Rosenstiel, and E. Kasneci. The applicability of cycle gans for pupil and eyelid segmentation, data generation and image refinement. In *International Conference on Computer Vision Workshops, ICCVW*, 11 2019.

[29] W. Fuhl, D. Geisler, T. Santini, T. Appel, W. Rosenstiel, and E. Kasneci. Cbf:circular binary features for robust and real-time pupil center detection. In *ACM Symposium on Eye Tracking Research & Applications*, 06 2018.

[30] W. Fuhl, D. Geisler, T. Santini, and E. Kasneci. Evaluation of state-of-the-art pupil detection algorithms on remote eye images. In *ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct publication – PETMEI 2016*, 09 2016.

[31] W. Fuhl and E. Kasneci. Eye movement velocity and gaze data generator for evaluation, robustness testing and assess of eye tracking software and visualization tools. In *Poster at Egocentric Perception, Interaction and Computing, EPIC*, 2018.

[32] W. Fuhl and E. Kasneci. Learning to validate the quality of detected landmarks. In *International Conference on Machine Vision, ICMV*, nov 2019.

[33] W. Fuhl and E. Kasneci. Multi layer neural networks as replacement for pooling operations. *arXiv preprint arXiv:2006.06969*, 08 2020.

[34] W. Fuhl and E. Kasneci. A multimodal eye movement dataset and a multimodal eye movement segmentation analysis. In *Proceedings of the ACM Symposium on Eye Tracking Research & Applications (ETRA)*, 2021.

[35] W. Fuhl and E. Kasneci. Rotated ring, radial and depth wise separable radial convolutions. In *Proceedings of the International Joint Conference on Neural Networks*. IEEE, 2021.

[36] W. Fuhl and E. Kasneci. Tensor normalization and full distribution training. In *Conference on Artificial Intelligence Workshops, AAAI Workshop Adversarial Machine Learning and Beyond*, 02 2022.

[37] W. Fuhl, G. Kasneci, and E. Kasneci. Teyed: Over 20 million real-world eye images with pupil, eyelid, and iris 2d and 3d segmentations, 2d and 3d landmarks, 3d eyeball, gaze vector, and eye movement types. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2021.

[38] W. Fuhl, G. Kasneci, W. Rosenstiel, and E. Kasneci. Training decision trees as replacement for convolution layers. In *Conference on Artificial Intelligence, AAAI*, 02 2020.

[39] W. Fuhl, T. Kübler, T. Santini, and E. Kasneci. Automatic generation of saliency-based areas of interest for the visualization and analysis of eye-tracking data. In *VMV*, pages 47–54, 2018.

[40] W. Fuhl, T. C. Kübler, H. Brinkmann, R. Rosenberg, W. Rosenstiel, and E. Kasneci. Region of interest generation algorithms for eye tracking data. In *Third Workshop on Eye Tracking and Visualization (ETVIS), in conjunction with ACM ETRA*, 06 2018.

[41] W. Fuhl, T. C. Kübler, D. Hospach, O. Bringmann, W. Rosenstiel, and E. Kasneci. Ways of improving the precision of eye tracking data: Controlling the influence of dirt and dust on pupil detection. *Journal of Eye Movement Research*, 10(3), 05 2017.

[42] W. Fuhl, T. C. Kübler, K. Sippel, W. Rosenstiel, and E. Kasneci. Arbitrarily shaped areas of interest based on gaze density gradient. In *European Conference on Eye Movements, ECEM 2015*, 08 2015.

[43] W. Fuhl, T. C. Kübler, K. Sippel, W. Rosenstiel, and E. Kasneci. Excuse: Robust pupil detection in real-world scenarios. In *16th International Conference on Computer Analysis of Images and Patterns (CAIP 2015)*, 09 2015.

[44] W. Fuhl, Y. Rong, and K. Enkelejda. Fully convolutional neural networks for raw eye tracking data segmentation, generation, and reconstruction. In *Proceedings of the International Conference on Pattern Recognition*, pages 0–0, 2020.

[45] W. Fuhl, Y. Rong, T. Motz, M. Scheidt, A. Hartel, A. Koch, and E. Kasneci. Explainable online validation of machine learning models for practical applications. In *Proceedings of the International Conference on Pattern Recognition*, pages 0–0, 2020.

[46] W. Fuhl, W. Rosenstiel, and E. Kasneci. 500,000 images closer to eyelid and pupil segmentation. In *Computer Analysis of Images and Patterns, CAIP*, 11 2019.

[47] W Fuhl, N Sanamrad, and E Kasneci. The gaze and mouse signal as additional source for user fingerprints in browser applications. *arXiv preprint arXiv:2101.03793*, 01 2021.

[48] W. Fuhl, T. Santini, D. Geisler, T. C. Kübler, and E. Kasneci. 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.

[49] W. Fuhl, T. Santini, D. Geisler, T. C. Kübler, W. Rosenstiel, and E. Kasneci. Eyes wide open? eyelid location and eye aperture estimation for pervasive eye tracking in real-world scenarios. In *ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct publication – PETMEI 2016*, 09 2016.

[50] W. Fuhl, T. Santini, and E. Kasneci. Fast and robust eyelid outline and aperture detection in real-world scenarios. In *IEEE Winter Conference on Applications of Computer Vision (WACV 2017)*, 03 2017.

[51] W. Fuhl, T. Santini, and E. Kasneci. Fast camera focus estimation for gaze-based focus control. *arXiv preprint arXiv:1711.03306*, 2017.

[52] W. Fuhl, T. Santini, G. Kasneci, and E. Kasneci. Pupilnet: Convolutional neural networks for robust pupil detection. *arXiv preprint arXiv:1601.04902*, 2016.

[53] W. Fuhl, T. Santini, G. Kasneci, W. Rosenstiel, and E. Kasneci. Pupilnet v2.0: Convolutional neural networks for cpu based real time robust pupil detection. *arXiv preprint arXiv:1711.00112*, 2017.

[54] W. Fuhl, T. Santini, T. C. Kübler, and E. Kasneci. Else: Ellipse selection for robust pupil detection in real-world environments. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications (ETRA)*, pages 123–130, 03 2016.

[55] W. Fuhl, T. Santini, T. Kuebler, N. Castner, W. Rosenstiel, and E. Kasneci. Eye movement simulation and detector creation to reduce laborious parameter adjustments. *arXiv preprint arXiv:1804.00970*, 2018.

[56] W. Fuhl, T. Santini, C. Reichert, D. Claus, A. Herkommer, H. Bahmani, K. Rifai, S. Wahl, and E. Kasneci. Non-intrusive practitioner pupil detection for unmodified microscope oculars. *Elsevier Computers in Biology and Medicine*, 79:36–44, 12 2016.

[57] W. Fuhl, J. Schneider, and E. Kasneci. 1000 pupil segmentations in a second using haar like features and statistical learning. In *International Conference on Computer Vision Workshops, ICCVW*, 2021.

[58] 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. In *Machine Vision and Applications*, pages 1–14, 06 2016.

[59] Wolfgang Fuhl. Maximum and leaky maximum propagation. *arXiv preprint arXiv:2105.10277*, 2021.

[60] Wolfgang Fuhl, Efe Bozkir, and Enkelejda Kasneci. Reinforcement learning for the privacy preservation and manipulation of eye tracking data. In *Proceedings of IEEE International Joint Conference on Neural Networks*, 2021.

[61] Wolfgang Fuhl and Enkelejda Kasneci. Weight and gradient centralization in deep neural networks. In *Proceedings of IEEE International Joint Conference on Neural Networks*, 2021.

[62] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

[63] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[64] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[65] Shahab Hoghooghi, Vesna Popovic, and Levi Swann. Novice to expert real-time knowledge transition in the context of x-ray airport security. In *Proceedings of DRS 2020 International Conference: Synergy. Vol. 4*. Design Research Society (UK), 2020.

[66] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[67] Stephen Hutt, Kristina Krasich, Caitlin Mills, Nigel Bosch, Shelby White, James R Brockmole, and Sidney K D'Mello. Automated gaze-based mind wandering detection during computerized learning in classrooms. *User Modeling and User-Adapted Interaction*, 29(4):821–867, 2019.

[68] Stephen Hutt, Caitlin Mills, Nigel Bosch, Kristina Krasich, James Brockmole, and Sidney D'mello. " out of the fr-eye-ing pan" towards gaze-based models of attention during learning with technology in the classroom. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, pages 94–103, 2017.

[69] Yoon Min Hwang and Kun Chang Lee. An eye-tracking paradigm to explore the effect of online consumers' emotion on their visual behaviour between desktop screen and mobile screen. *Behaviour & Information Technology*, pages 1–12, 2020.

[70] Halszka Jarodzka, Irene Skuballa, and Hans Gruber. Eye-tracking in educational practice: Investigating visual perception underlying teaching and learning in the classroom. *Educational Psychology Review*, pages 1–10, 2020.

[71] Libby Jenke, Kirk Bansak, Jens Hainmueller, and Dominik Hangartner. Using eye-tracking to understand decision-making in conjoint experiments. *Political Analysis*, 29(1):75–101, 2021.

[72] Huaizu Jiang and Erik Learned-Miller. Face detection with the faster r-cnn. In *2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017)*, pages 650–657. IEEE, 2017.

[73] Erika Jönsson. If looks could kill–an evaluation of eye tracking in computer games. *Unpublished Master's Thesis, Royal Institute of Technology (KTH), Stockholm, Sweden*, 2005.

[74] Tilke Judd, Krista Ehinger, Frédo Durand, and Antonio Torralba. Learning to predict where humans look. In *2009 IEEE 12th international conference on computer vision*, pages 2106–2113. IEEE, 2009.

[75] Christina Katsini, Yasmeen Abdrabou, George E Raptis, Mohamed Khamis, and Florian Alt. The role of eye gaze in security and privacy applications: Survey and future hci research directions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–21, 2020.

[76] Petr Kellnhofer, Adria Recasens, Simon Stent, Wojciech Matusik, and Antonio Torralba. Gaze360: Physically unconstrained gaze estimation in the wild. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6912–6921, 2019.

[77] Davis E King. Dlib-ml: A machine learning toolkit. *The Journal of Machine Learning Research*, 10:1755–1758, 2009.

[78] Davis E King. Max-margin object detection. *arXiv preprint arXiv:1502.00046*, 2015.

[79] Andreas Korbach, Paul Ginns, Roland Brünken, and Babette Park. Should learners use their hands for learning? results from an eye-tracking study. *Journal of Computer Assisted Learning*, 36(1):102–113, 2020.

[80] Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba. Eye tracking for everyone. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2176–2184, 2016.

[81] Ralf Kredel, Christian Vater, André Klostermann, and Ernst-Joachim Hossner. Eye-tracking technology and the dynamics of natural gaze behavior in sports: A systematic review of 40 years of research. *Frontiers in psychology*, 8:1845, 2017.

[82] Zhihang Li, Xu Tang, Junyu Han, Jingtuo Liu, and Ran He. Pyramidbox++: high performance detector for finding tiny face. *arXiv preprint arXiv:1904.00386*, 2019.

[83] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016.

[84] Feng Lu, Yusuke Sugano, Takahiro Okabe, and Yoichi Sato. Adaptive linear regression for appearance-based gaze estimation. *IEEE transactions on pattern analysis and machine intelligence*, 36(10):2033–2046, 2014.

[85] David Manning, Susan C Ethell, and Trevor Crawford. Eye-tracking afroc study of the influence of experience and training on chest x-ray interpretation. In *Medical Imaging 2003: Image Perception, Observer Performance, and Technology Assessment*, volume 5034, pages 257–266. International Society for Optics and Photonics, 2003.

[86] Manuel Jesús Marin-Jimenez, Andrew Zisserman, Marcin Eichner, and Vittorio Ferrari. Detecting people looking at each other in videos. *International Journal of Computer Vision*, 106(3):282–296, 2014.

[87] Hideyuki Matsumoto, Yasuo Terao, Akihiro Yugeta, Hideki Fukuda, Masaki Emoto, Toshiaki Furubayashi, Tomoko Okano, Ritsuko Hanajima, and Yoshikazu Ugawa. Where do neurologists look when viewing brain ct images? an eye-tracking study involving stroke cases. *PloS one*, 6(12):e28928, 2011.

[88] Pierre Maurage, Nicolas Masson, Zoé Bollen, and Fabien D'Hondt. Eye tracking correlates of acute alcohol consumption: A systematic and critical review. *Neuroscience & Biobehavioral Reviews*, 108:400–422, 2020.

[89] Nora A McIntyre and Tom Foulsham. Scanpath analysis of expertise and culture in teacher gaze in real-world classrooms. *Instructional Science*, 46(3):435–455, 2018.

[90] Aideen McParland, Stephen Gallagher, and Mickey Keenan. Investigating gaze behaviour of children diagnosed with autism spectrum disorders in a classroom setting. *Journal of Autism and Developmental Disorders*, pages 1–16, 2021.

[91] Xiaoxu Meng, Ruofei Du, Matthias Zwicker, and Amitabh Varshney. Kernel foveated rendering. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 1(1):1–20, 2018.

[92] Caitlin Mills, Nigel Bosch, Art Graesser, and Sidney D'Mello. To quit or not to quit: predicting future behavioral disengagement from reading patterns. In *International Conference on Intelligent Tutoring Systems*, pages 19–28. Springer, 2014.

[93] Sankha S Mukherjee and Neil Martin Robertson. Deep head pose: Gaze-direction estimation in multimodal video. *IEEE Transactions on Multimedia*, 17(11):2094–2107, 2015.

[94] Jessie R Oldham, Christina L Master, Gregory A Walker, William P Meehan III, and David R Howell. The association between baseline eye tracking performance and concussion assessments in high school football players. *Optometry and Vision Science*, 98(7):826–832, 2021.

[95] Oskar Palinko, Francesco Rea, Giulio Sandini, and Alessandra Sciutti. Robot reading human gaze: Why eye tracking is better than head tracking for human-robot collaboration. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5048–5054. IEEE, 2016.

[96] Mathieson Tan Zui Quen, J Mountstephens, Yong Guang Teh, and J Teo. Medical image interpretation training with a low-cost eye tracking and feedback system: A preliminary study. *Healthcare Technology Letters*, 2021.

[97] Adria Recasens, Carl Vondrick, Aditya Khosla, and Antonio Torralba. Following gaze in video. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1435–1443, 2017.

[98] Adriá Recasens Continente Recasens. *Where are they looking?* PhD thesis, Massachusetts Institute of Technology, 2016.

[99] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[100] Jonas Reichenberger, Michael Pfaller, and Andreas Mühlberger. Gaze behavior in social fear conditioning: An eye-tracking study in virtual reality. *Frontiers in psychology*, 11:35, 2020.

[101] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28:91–99, 2015.

[102] Jennifer C Reneker, W Cody Pannell, Ryan M Babl, Yunxi Zhang, Seth T Lirette, Felix Adah, and Matthew R Reneker. Virtual immersive sensorimotor training (vist) in collegiate soccer athletes: A quasi-experimental study. *Heliyon*, 6(7):e04527, 2020.

[103] Ian H Robertson, Tom Manly, Jackie Andrade, Bart T Baddeley, and Jenny Yiend. Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, 35(6):747–758, 1997.

[104] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018.

[105] Michael Schneider, Angela Heine, Verena Thaler, Joke Torbeyns, Bert De Smedt, Lieven Verschaffel, Arthur M Jacobs, and Elsbeth Stern. A validation of eye movements as a measure of elementary school children's developing number sense. *Cognitive Development*, 23(3):409–422, 2008.

[106] Jonathan Smallwood, Merrill McSpadden, and Jonathan W Schooler. When attention matters: The curious incident of the wandering mind. *Memory & Cognition*, 36(6):1144–1150, 2008.

[107] Yusuke Sugano, Yasuyuki Matsushita, and Yoichi Sato. Appearance-based gaze estimation using visual saliency. *IEEE transactions on pattern analysis and machine intelligence*, 35(2):329–341, 2012.

[108] Kar-Han Tan, David J Kriegman, and Narendra Ahuja. Appearance-based eye gaze estimation. In *Sixth IEEE Workshop on Applications of Computer Vision, 2002.(WACV 2002). Proceedings.*, pages 191–195. IEEE, 2002.

[109] Kavin Kathiresh Vijayan, Ola Jon Mork, and Irina Emily Hansen. Eye tracker as a tool for engineering education. *Universal Journal of Educational Research*, 6(11):2647–2655, 2018.

[110] Paul Viola and Michael J Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.

[111] Nutthanan Wanluk, Sarinporn Visitsattapongse, Aniwat Juhong, and C Pintavirooj. Smart wheelchair based on eye tracking. In *2016 9th Biomedical Engineering International Conference (BMEiCON)*, pages 1–4. IEEE, 2016.

[112] Cesco Willemse and Agnieszka Wykowska. In natural interaction with embodied robots, we prefer it when they follow our gaze: a gaze-contingent mobile eyetracking study. *Philosophical Transactions of the Royal Society B*, 374(1771):20180036, 2019.

[113] Oliver Williams, Andrew Blake, and Roberto Cipolla. Sparse and semi-supervised visual mapping with the s^ 3gp. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 230–237. IEEE, 2006.

[114] Genevieve Alice Woolverton and Alisha R Pollastri. An exploration and critical examination of how "intelligent classroom technologies" can improve specific uses of direct student behavior observation methods. *Educational Measurement: Issues and Practice*, 2021.

[115] Renjie Xie, Yanzhi Chen, Yan Wo, and Qiao Wang. A deep, information-theoretic framework for robust biometric recognition. *arXiv preprint arXiv:1902.08785*, 2019.

[116] Junjie Yan, Zhen Lei, Longyin Wen, and Stan Z Li. The fastest deformable part model for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2497–2504, 2014.

[117] Xiaozhi Yang and Ian Krajbich. Webcam-based online eye-tracking for behavioral research. *Judgment and Decision Making*, 16(6):1486, 2021.

[118] Ali Shahidi Zandi, Azhar Quddus, Laura Prest, and Felix JE Comeau. Non-intrusive detection of drowsy driving based on eye tracking data. *Transportation research record*, 2673(6):247–257, 2019.

[119] Cha Zhang and Zhengyou Zhang. Improving multiview face detection with multi-task deep convolutional neural networks. In *IEEE Winter Conference on Applications of Computer Vision*, pages 1036–1041. IEEE, 2014.

[120] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016.

[121] Xucong Zhang, Yusuke Sugano, Mario Fritz, and Andreas Bulling. Appearance-based gaze estimation in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4511–4520, 2015.

[122] Zhishuai Zhang, Wei Shen, Siyuan Qiao, Yan Wang, Bo Wang, and Alan Yuille. Robust face detection via learning small faces on hard images. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1361–1370, 2020.