Interactive Fixation-to-AOI Mapping for Mobile Eye Tracking Data based on Few-Shot Image Classification

Michael Barz
michael.barz@dfki.de
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
University of Oldenburg
Oldenburg, Germany

Omair Shahzad Bhatti
ombh01@dfki.de
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany

Hasan Md Tusfiqur Alam
hasan_md_tusfiqur.alam@dfki.de
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany

Duy Minh Ho Nguyen
ho_minh_duy.nguyen@dfki.de
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
University of Stuttgart
Stuttgart, Germany
Max Planck Research School for Intelligent Systems
Stuttgart, Germany

Daniel Sonntag
daniel.sonntag@dfki.de
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
University of Oldenburg
Oldenburg, Germany

ABSTRACT

Mobile eye tracking is an important tool in psychology and human-centred interaction design for understanding how people process visual scenes and user interfaces. However, analysing recordings from mobile eye trackers, which typically include an egocentric video of the scene and a gaze signal, is a time-consuming and largely manual process. To address this challenge, we propose a web-based annotation tool that leverages few-shot image classification and interactive machine learning (IML) to accelerate the annotation process. The tool allows users to efficiently map fixations to areas of interest (AOI) in a video-editing-style interface. It includes an IML component that generates suggestions and learns from user feedback using a few-shot image classification model initialised with a small number of images per AOI. Our goal is to improve the efficiency and accuracy of fixation-to-AOI mapping in mobile eye tracking.

CCS CONCEPTS

• **Human-centered computing** → **Interactive systems and tools**; Empirical studies in HCI; • **Computing methodologies** → *Machine learning*.

KEYWORDS

eye tracking, interactive machine learning, area of interest, mobile eye tracking, visual attention, eye tracking data analysis, fixation to AOI mapping

ACM Reference Format:
Michael Barz, Omair Shahzad Bhatti, Hasan Md Tusfiqur Alam, Duy Minh Ho Nguyen, and Daniel Sonntag. 2023. Interactive Fixation-to-AOI Mapping for Mobile Eye Tracking Data based on Few-Shot Image Classification. In *28th International Conference on Intelligent User Interfaces (IUI '23 Companion), March 27–31, 2023, Sydney, NSW, Australia.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3581754.3584179

IUI '23 Companion, March 27–31, 2023, Sydney, NSW, Australia

ACM ISBN 979-8-4007-0107-8/23/03.
https://doi.org/10.1145/3581754.3584179

1 INTRODUCTION

Mobile eye tracking studies often use areas of interest (AOIs) and visual attention to these AOIs to analyse and understand how people process visual information. AOIs are specific regions in a scene or interface that are of interest to the researcher. Visual attention refers to the time a person pays attention to these regions. By measuring visual attention to and transitions between AOIs during a study, researchers can gain insights into which elements are most relevant or appealing and how they may influence decision-making. This is usually done based on fixation events because they approximate the time spent processing the visual scene [10]. However, accurately annotating mobile eye tracking data is a challenging and time-consuming task, because scene videos taken with a head-mounted eye tracking device are unique for every participant. Hence, efficient fixation-to-AOI mapping techniques from remote eye tracking, like keyframe-based annotation of dynamic AOIs in video-based stimuli [14], do not scale. In practice, one or more annotators decide, per fixation, whether an AOI was hit or not [13, 24]. A solution can be found in attaching fiducial markers to target stimuli [3, 18, 20, 29], but we aim at non-instrumented environments without obtrusive

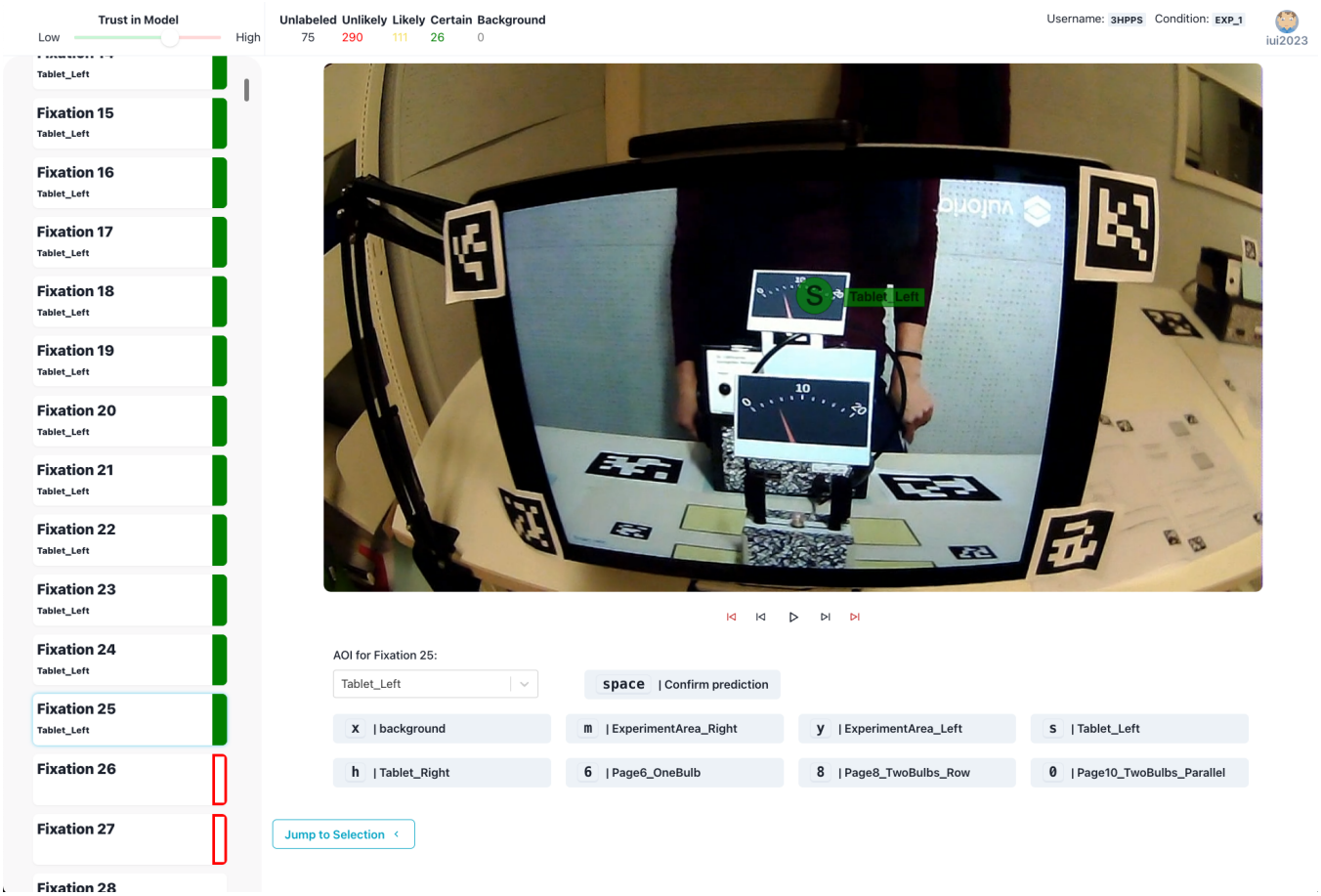

Figure 1: Screenshot of our intelligent user interface for semi-automatic annotation of mobile eye tracking data.

markers. Existing approaches for automatic or semi-automatic analysis of head-mounted eye tracking data use computer vision models to map fixations to AOIs [6–8, 11, 12, 17, 19, 21–24, 28]. But these approaches come with certain limitations. Most of them rely on pre-trained computer vision models that do not allow for adapting the underlying model to a certain target domain [6, 8, 17, 22, 24, 25]. These can be applied in very constrained settings only, i.e., if the dataset used for training the machine learning model matches the target domain. Other approaches suffer from a lack of flexibility. They are based on a single, a priori model training or fine-tuning step with no possibility to adapt the model during the annotation process [11, 19, 28]. Kurzhals [12] presented a promising approach that combines image clustering and human labour for annotating mobile eye tracking data. Annotators interact with a cluster representation of image patches extracted from the video stream for each fixation. In contrast, we combine few-shot image classification with human labour in a video-editing-style interface. We present a tool for fixation-to-AOI mapping that combines automation, based on a state-of-the-art few-shot image classification model and concepts from interactive machine learning (IML) [2], with human labour in an intelligent user interface.

2 INTERACTIVE FIXATION-TO-AOI MAPPING

We demonstrate a web-based tool for fixation-to-AOI mapping which is an essential data processing step in research based on mobile eye trackers. It allows practitioners to efficiently annotate their recordings fixation-wise in a video-editing-like interface (see figure 1). We integrate an IML service that learns from prior annotations using a few-shot image classification model. It suggests AOI labels for fixations and visualises its certainty per suggestion using a colour-coding scheme. Annotators can easily confirm or correct these suggestions. This feedback is used to re-train the underlying model. Annotations and suggestions are stored in a database in our backend (see figure 2).

2.1 User Interface

The user interface includes three main components: the top bar displays information on the selected gaze recording and on the annotation progress, a list on the left that shows all fixations and their annotation state, and a video view on the right with a fixation overlay and buttons for manual annotation (see figure 1). When a fixation is selected from the list, the video view jumps to the corresponding frame and shows the fixation position and

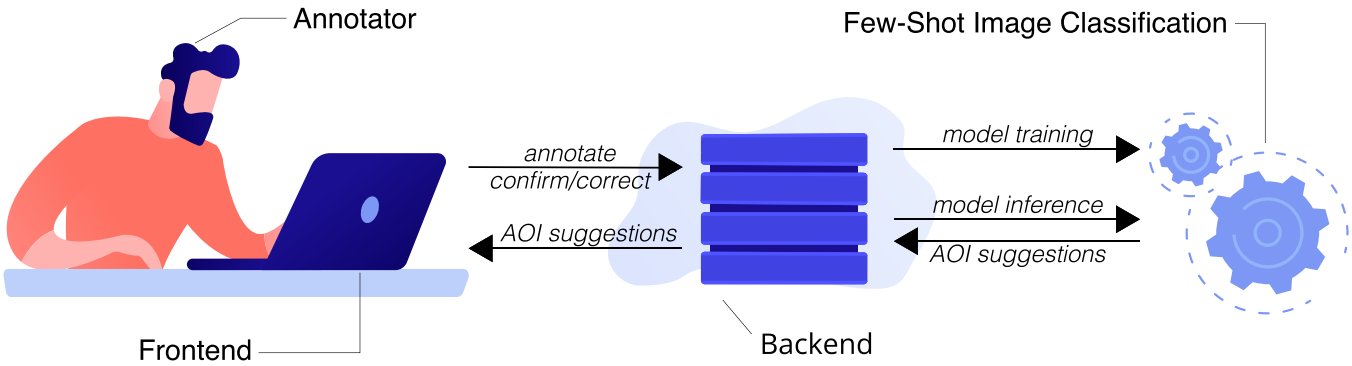

Figure 2: Overview of the architecture of our interactive annotation system including a web-based user interface, the backend that manages data storage, and the IML service.

the currently assigned AOI as an overlay. An AOI can be assigned to the fixation by clicking one of the buttons or by pressing the corresponding shortcut on the keyboard. A successful assignment is visually confirmed by adding a green badge on the right of the fixation's list item. Our tool allows navigation through fixations using the arrow keys and by video playback. The playback option can be used to quickly check AOI suggestions which are displayed as part of the fixation overlay. If multiple fixations hit the same AOI, these can be annotated simultaneously. For this, the annotator selects multiple fixations from the list using the shift and arrow keys, which is consistent with multi-item selection in many user interface frameworks and assigns them in the same way as a single fixation item. The video frames, the fixations, and the fixation-to-AOI mapping are retrieved from and managed by the backend of our system. For demonstration purposes, we use an existing mobile eye tracking dataset from the educational sciences domain which includes gaze recordings from 48 participants. The dataset was recorded at Saarland University with the aim of investigating the effects of augmented reality (AR) support in a laboratory-based learning scenario about electrical circuits on the learning outcomes and processes of primary school children [1, preregsitered at Open Science Framework]. Figure 1 shows a video frame from the head-mounted camera from the AR support condition with the tablet and the experiment setup in the foreground and the instructor in the background. A prototype of the eyeNotate Tool can be accessed via iml.dfki.de/demos/eyeNotate/.

2.2 Interactive Machine Learning Component

Our tool has an interactive machine learning (IML) component that shall increase the efficiency of the fixation-to-AOI mapping process. It is based on a few-shot image classification model, which is initialised with a small number of images per AOI [26]. The model takes the fixation point and a corresponding video frame as input, crops an image patch around the fixation point, and classifies the image patch similar to Barz et al. [4], Barz and Sonntag [5, 6]. The model is used to generate AOI label suggestions for each fixation. The availability of an AOI suggestion is indicated by a non-filled badge at a fixation's list item (see figure 1). Its outline colour encodes the model's confidence: Green, yellow, or red representing high to low model confidence. The pre-configured thresholds can be adjusted by the user through a slider in the top bar. If the slider is moved towards *high* trust, the thresholds are decreased and more suggestions will appear in green (and vice versa). An overview with the number of items per confidence class is shown in the top bar. Selected suggestions can also be confirmed by pressing the return or space key. An incorrect suggestion can be corrected by manually assigning another class. All annotations by the user, including confirmative and corrective feedback, are used to re-train the underlying few-shot image classification model. We expect that the model will improve its performance in predicting the correct AOI over time. The model training and inference run in parallel.

At the core of our IML component, we run a few-shot image classification model that takes an image patch cropped around a fixation point as input to decide whether the fixation hits one of the defined AOIs or not. We employ a few-shot learning strategy [26] to enable fast model adaptation and improvements based on user-provided samples. Our approach is based on the idea of reconstruction [15, 16, 30] where the class membership task is framed as a problem of *reconstructing feature maps*. We have used a Feature Map Reconstruction Network (FRN) [27] which classifies a target image by reconstructing class associate feature maps of the image using a set of support features. The support features are learned from a set of images all belonging to the same class. For each query image, the FRN attempts to reconstruct the feature map as a weighted sum of the support features. The negative reconstruction error is used as the class score, with smaller errors indicating that the query image is more likely to belong to the same class as the support features. The backbone of the FRN architecture is ResNet50 [9]. The initial model is trained in a *10-shot-k-way* manner, with k being the number of classes and using 10 images per class. To update the classification model, we randomly select 10 images per class from the pool of user-annotated images and use them for re-training. Re-training is triggered whenever 10 new samples are available.

3 CONCLUSION

We demonstrated a tool for annotating mobile eye tracking data that combines machine learning with human input in a user-friendly interface. The tool provides label suggestions based on a few-shot

image classification model, which can be updated based on the user's feedback. Our goal is to make the annotation process more efficient and effective by reducing the time and effort required for manual annotation. We plan to conduct a user study to investigate the efficiency and effectiveness of our approach.

ACKNOWLEDGMENTS

This work was funded by the German Federal Ministry of Education and Research under grant number 01JD1811C (GeAR) and by the European Commission project MASTER (grant number 101093079; https://www.master-xr.eu/).

REFERENCES

[1] Kristin Altmeyer, Sebastian Kapp, Michael Barz, Luisa Lauer, Sarah Malone, Jochen Kuhn, and Roland Brünken. 2020. The effect of augmented reality on global coherence formation processes during STEM laboratory work in elementary school children. (Oct. 2020). https://doi.org/10.17605/OSF.IO/GWHU5

[2] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. 2014. Power to the People: The Role of Humans in Interactive Machine Learning. *AI Magazine* 35, 4 (Dec. 2014), 105–120. https://doi.org/10.1609/aimag.v35i4.2513

[3] Michael Barz, Florian Daiber, Daniel Sonntag, and Andreas Bulling. 2018. Error-aware gaze-based interfaces for robust mobile gaze interaction. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications, ETRA 2018, Warsaw, Poland, June 14-17, 2018*, Bonita Sharif and Krzysztof Krejtz (Eds.). ACM, 24:1–24:10. https://doi.org/10.1145/3204493.3204536

[4] Michael Barz, Sebastian Kapp, Jochen Kuhn, and Daniel Sonntag. 2021. Automatic Recognition and Augmentation of Attended Objects in Real-time using Eye Tracking and a Head-mounted Display. In *ACM Symposium on Eye Tracking Research and Applications (ETRA '21 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 1–4. https://doi.org/10.1145/3450341.3458766

[5] Michael Barz and Daniel Sonntag. 2016. Gaze-guided object classification using deep neural networks for attention-based computing. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp Adjunct 2016, Heidelberg, Germany, September 12-16, 2016*, Paul Lukowicz, Antonio Krüger, Andreas Bulling, Youn-Kyung Lim, and Shwetak N. Patel (Eds.). ACM, 253–256. https://doi.org/10.1145/2968219.2971389

[6] Michael Barz and Daniel Sonntag. 2021. Automatic Visual Attention Detection for Mobile Eye Tracking Using Pre-Trained Computer Vision Models and Human Gaze. *Sensors* 21, 12 (Jan. 2021), 4143. https://doi.org/10.3390/s21124143 Number: 12 Publisher: Multidisciplinary Digital Publishing Institute.

[7] Stijn De Beugher, Geert Brône, and Toon Goedemé. 2014. Automatic analysis of in-the-wild mobile eye-tracking experiments using object, face and person detection. In *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*, Vol. 1. 625–633.

[8] Oliver Deane, Eszter Toth, and Sang-Hoon Yeo. 2022. Deep-SAGA: a deep-learning-based system for automatic gaze annotation from eye-tracking data. *Behavior Research Methods* (June 2022). https://doi.org/10.3758/s13428-022-01833-4

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 770–778. https://doi.org/10.1109/CVPR.2016.90

[10] Marcel A Just and Patricia A Carpenter. 1980. A theory of reading: from eye fixations to comprehension. *Psychological review* 87, 4 (1980), 329. Publisher: American Psychological Association.

[11] Niharika Kumari, Verena Ruf, Sergey Mukhametov, Albrecht Schmidt, Jochen Kuhn, and Stefan Küchemann. 2021. Mobile Eye-Tracking Data Analysis Using Object Detection via YOLO v4. *Sensors* 21, 22 (2021). https://doi.org/10.3390/s21227668

[12] Kuno Kurzhals. 2021. Image-Based Projection Labeling for Mobile Eye Tracking. In *ACM Symposium on Eye Tracking Research and Applications*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3448017.3457382

[13] Kuno Kurzhals, Cyrill Fabian Bopp, Jochen Bässler, Felix Ebinger, and Daniel Weiskopf. 2014. Benchmark Data for Evaluating Visualization and Analysis Techniques for Eye Tracking for Video Stimuli. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization (BELIV '14)*. Association for Computing Machinery, New York, NY, USA, 54–60. https://doi.org/10.1145/2669557.2669558 event-place: Paris, France.

[14] Kuno Kurzhals, Florian Heimerl, and Daniel Weiskopf. 2014. ISeeCube: Visual Analysis of Gaze Data for Video. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '14)*. Association for Computing Machinery, New York, NY, USA, 43–50. https://doi.org/10.1145/2578153.2578158 event-place: Safety Harbor, Florida.

[15] Dong Hoon Lee and Sae-Young Chung. 2021. Unsupervised embedding adaptation via early-stage feature reconstruction for few-shot classification. In *International Conference on Machine Learning*. PMLR, 6098–6108.

[16] Yuewen Li, Wenquan Feng, Shuchang Lyu, and Qi Zhao. 2023. Feature reconstruction and metric based network for few-shot object detection. *Computer Vision and Image Understanding* 227 (2023), 103600. Publisher: Elsevier.

[17] Eduardo Manuel Silva Machado, Ivan Carrillo, Miguel Collado, and Liming Chen. 2019. Visual Attention-Based Object Detection in Cluttered Environments. In *2019 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. 133–139. https://doi.org/10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00064

[18] Gregor Mehlmann, Markus Häring, Kathrin Janowski, Tobias Baur, Patrick Gebhard, and Elisabeth André. 2014. Exploring a Model of Gaze for Grounding in Multimodal HRI. In *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI '14)*. Association for Computing Machinery, New York, NY, USA, 247–254. https://doi.org/10.1145/2663204.2663275 event-place: Istanbul, Turkey.

[19] Karen Panetta, Qianwen Wan, Aleksandra Kaszowska, Holly A. Taylor, and Sos Agaian. 2019. Software Architecture for Automating Cognitive Science Eye-Tracking Data Analysis and Object Annotation. *IEEE Transactions on Human-Machine Systems* 49, 3 (2019), 268–277. https://doi.org/10.1109/THMS.2019.2892919

[20] Thies Pfeiffer, Patrick Renner, and Nadine Pfeiffer-Leßmann. 2016. EyeSee3D 2.0: Model-Based Real-Time Analysis of Mobile Eye-Tracking in Static and Dynamic Three-Dimensional Scenes. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications (ETRA '16)*. Association for Computing Machinery, New York, NY, USA, 189–196. https://doi.org/10.1145/2857491.2857532 event-place: Charleston, South Carolina.

[21] Daniel F. Pontillo, Thomas B. Kinsman, and Jeff B. Pelz. 2010. SemantiCode: Using Content Similarity and Database-Driven Matching to Code Wearable Eyetracker Gaze Data. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications (ETRA '10)*. Association for Computing Machinery, New York, NY, USA, 267–270. https://doi.org/10.1145/1743666.1743729 event-place: Austin, Texas.

[22] Ömer Sümer, Patricia Goldberg, Kathleen Stürmer, Tina Seidel, Peter Gerjets, Ulrich Trautwein, and Enkelejda Kasneci. 2018. Teachers' Perception in the Classroom. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. 2315–2324. https://openaccess.thecvf.com/content_cvpr_2018_workshops/w47/html/Sumer_Teachers_Perception_in_CVPR_2018_paper.html

[23] Takumi Toyama, Thomas Kieninger, Faisal Shafait, and Andreas Dengel. 2012. Gaze Guided Object Recognition Using a Head-Mounted Eye Tracker. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12)*. Association for Computing Machinery, New York, NY, USA, 91–98. https://doi.org/10.1145/2168556.2168570 event-place: Santa Barbara, California.

[24] Karan Uppal, Jaeah Kim, and Shashank Singh. 2022. Decoding Attention from Gaze: A Benchmark Dataset and End-to-End Models. In *NeuRIPS 2022 Workshop on Gaze Meets ML*. https://openreview.net/forum?id=1Ty3Xd9HUQv

[25] Pranav Venuprasad, Li Xu, Enoch Huang, Andrew Gilman, Leanne Chukoskie Ph.D., and Pamela Cosman. 2020. Analyzing Gaze Behavior Using Object Detection and Unsupervised Clustering. In *ACM Symposium on Eye Tracking Research and Applications (ETRA '20 Full Papers)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3379155.3391316 event-place: Stuttgart, Germany.

[26] Yaqing Wang, Quanming Yao, James Kwok, and Lionel Ni. 2020. Generalizing from a Few Examples: A Survey on Few-shot Learning. *Comput. Surveys* 53 (June 2020), 1–34. https://doi.org/10.1145/3386252

[27] Davis Wertheimer, Luming Tang, and Bharath Hariharan. 2021. Few-Shot Classification With Feature Map Reconstruction Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 8012–8021.

[28] Julian Wolf, Stephan Hess, David Bachmann, Quentin Lohmeyer, and Mirko Meboldt. 2018. Automating areas of interest analysis in mobile eye tracking experiments based on machine learning. *Journal of Eye Movement Research* 11, 6 (Dec. 2018). https://doi.org/10.16910/jemr.11.6.6 Section: Articles.

[29] L.H. Yu and M. Eizenman. 2004. A new methodology for determining point-of-gaze in head-mounted eye tracking systems. *IEEE Transactions on Biomedical Engineering* 51, 10 (Oct. 2004), 1765–1773. https://doi.org/10.1109/TBME.2004.831523

[30] Chi Zhang, Yujun Cai, Guosheng Lin, and Chunhua Shen. 2020. DeepEMD: Few-Shot Image Classification With Differentiable Earth Mover's Distance and Structured Classifiers. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 12200–12210. https://doi.org/10.1109/CVPR42600.2020.01222