Inspecting Cycling Attention

Ivo Cornelis de Geus

ABSTRACT

As cycling is increasingly adopted as the future of sustainable human mobility, governments are increasingly interested in making it safer. During any traffic interaction, we use a variety of senses to direct us, of which the visual is the most dominant. Eye-movement in automotors is a mature field, while the amount research in cycling is relatively small. This project aimed to map a single wearable eye-trackers to a master video-track using computer vision. With test-footage, several methods of this mapping images together have been tested, of which a Siamese Network performed best.

KEYWORDS

Infrastructure, Smart Mobility, Computer Vision, Eye Tracking, Heatmaps

Student Ivo Cornelis de Geus
External Supervisor J. Schreuder, PhD
External Supervisor S. Buningh, Ir.
Internal Supervisor Prof. Dr. M. Worring
Github https://github.com/idegeus/msc-ds-thesis

1 INTRODUCTION

Biking is often touted as the future of individual mobility to replace cars, with good reason. It's healthy, cheap, fun and good for the environment [12, 14, 21]. From the perspective of a government, citizens get significantly healthier, combined with a more efficient spatial urban design []. With 68% of the world population expected to live in urban areas by 2050, biking is deemed such a promising fit in its sustainability goals that the United Nations declared the third of June as the day of the bike [38, 35, 34]. #BanCars

To get everyone on the bike and create a welcoming built environment for biking, governments are increasingly looking for advice to countries with an existing prominent biking culture, such as the Netherlands, Denmark and Germany [28]. An increasingly discussed, but equally controversial measure to improve the "biking climate" is a redistribution of urban space to prioritize one modality over the other [36, 41].

Equally important as linked to the urban decisions is the environment in which transport is taking place and how it is perceived. Since the visual perception is the most important factor in navigation and perceptual errors contribute 20% of european road accidents, it makes sense to see how we process this information [40]. ERSO claims that in 2020, biking is the only modality not decreasing in fatalities since 2010, therefore it makes sense to pay attention to where bikers are looking and how they are paying attention.

In a historical perspective of using eye movements, Gompel et al. refer to Du Laurens, a French anatomist and medical scientist in 1596, who described the eyes as *windowes of the mind*. Indeed, it seems clear today that eye movements reveal the workings of mind and brain [9]. The theory that eyes give information about what the brain is working on is sometimes referred to as the eye-mind assumption, and while it does not directly guarantee processing by the brain, it is still a robust and useful link [1, 30]. Because of this, eye-tracking has been widely deployed in a multiple of fields, ranging from tourism and user testing to software engineering and traffic evaluation [17]. While eye-tracking is has always been used in a fixed environment, such as a desktop, innovations in wearable electronics allow for researching eye-tracking in a more natural setting. With this comes the trouble of inspecting different participants in a single go, which is what this project will focus on.

In this research, an attempt is made to set up a method to map a wearable eye-tracker on one single master-image using a computer vision feature detection system. Normal eye-trackers in a controlled environment are mapped onto one single master-image in 360 degrees, called a map of Areas of Interest (AOI). In this way, it becomes clear what a group of participants are looking at. This is not possible in a more natural setting where eye-trackers are worn in a wearable form.

2 BACKGROUND

2.1 Eye-tracker

In navigating everyday traffic, using human visual information processing is our primary way of getting around safely [11, 9, 10]. It is an important factor in obstacle avoidance, safe navigation and risk perception, and is therefore a large part of the required workload [20, 18]. For this reason, eye-tracking has been used in traffic studies for a long time. Both for in car-driving and walking, eye-movements have been studied extensively to evaluate driver awareness [46], intersection design [18, 16], location, colors and font of signage [43] and the impact of information in advanced driver assistant systems [11, 42, 15]. Some studies have also tried using eye-tracking as a measure of workload and comfort [30]. While some studies argue eye-tracking is not sensitive enough to be used for workload measuring [15], it has sometimes been used to determine behavior and workload in traffic situations [27, 31, 26]. The amount of research in the visual behavior in bikers (as an urban transport) was somewhat limited, but has seen an increase in recent years [24, 29, 27, 33, 32, 37].

Running an eye-tracking experiment on the topic of traffic is possible in a variety of different ways. In the beginning, running such an experiment was mostly done in a controlled environment at a desk with a fixed eye-tracker such as in [5, 18, 11]. While this has many advantages such as internal validity, reliability and ethical advances, a possible lack of realism and ecological validity of a desk-mounted eye-tracker is a disadvantage, especially to research in behavior in a real-life [23, 25]. For this purpose, wearable eye-trackers have been developed, of which now exist a couple different types. This type of eye-tracker generates a scene-camera of the perspective of the participant, with the relative eye-movement projected on top of it. While fixed eye-trackers can generate an aggregated map of Areas of Interest (AOI) as the scene is always the same, the different head movements of different participants makes this more complicated as there is no single map to project on.

In previous studies, this has been resolved by analysing the scene frame-by-frame and fixation-by-fixation, which was was described by Duchowski as "rather tedious but surprisingly effective", and has been used successfully by several other researchers [8, 27, 23]. Creating a first example of automating this task, and mapping several of these relative AOI's on one single image is the topic where this project will focus on.

3 RELATED WORKS

3.1 Feature Detection & Siamese Networks

Mapping several different images onto one master image without having the ability to train can be considered a case of one-shot image recognition. This problem is defined as being able to learn information about an object from one, or only a few, training samples/images [47]. It is a problem is something humans are shown to be good at very quickly due to their ability to synthesize and learn new object classes from existing information about previously learned classes. This is the key motivation for one-shot learning techniques, where systems can, like humans, use prior knowledge to classify new objects [4, 6]. This section will base partly on the explanation from Koch in [22].

3.2 Traditional Computer Vision

While this topic in computer vision has been addressed earlier in the 1980's and the 1990's, the basis was laid by Fei-Fei, Fergus, and Perona [4]. In this paper, a variational Bayesian framework for one-shot image classification was created based on the idea that previously learned classes can help forecast future ones. Traditional following computer vision models for one-shot learning usually fall into two categories: feature learning and metric learning. Example of both these types are respectively the Bag of Features (BoF) [13] and Scale-Invariant Feature Transform (SIFT) [3, 19] or Features from Accelerated Segment Test [7].

3.3 Siamese Neural Networks

Another approach which uses the same core principle are Siamese Neural Networks (SNNs), which were first introduced by Bromley et al. to solve the signature matching problem. The core problem a SSN is aimed to solve is generating a robust representation in a multi-dimensional space, optimising for a low distance between same-class objects, and a high-distance between different objects. Training such a network is normally archieved by creating two augmentations of one image, subject to certain conditions to avoid collapsing solutions [39]. While many SNNs have been proposed and tried, a more typical SNN consists of two twin networks accepting different inputs, joined by an energy function at the top, see Figure 1. As visible, in this example, both networks use the same weight sets. This ensures both the consistency and symmetry of predictions, as both sides of the network will output the same function.

While many version exist, the architecture used in this project was introduced by Chen and He, called Simple Siamese Representation, short SimSiam. In this paper, Chen and He explores the effects of deliberately introducing a stop-gradient on the second "twin" of the SNN, which showed its effectiveness [39]. This version of a SNN showed an accuracy of 68.1%. The rest of this paragraph will briefly give an oversight of the structure of the used SNN as explained in [39] and Figure 2. The networks takes in two random augmentations x_1 and x_2 from image x. The two images are processed by an encoder network f, which consists of a backbone (in this case, ResNet) and a projection Multilayer Perceptron (MLP). The encoder f shares weights between the two views, as shown in Figure 2. Prediction head h transforms the output of f_1 and matches it to the other unprocessed view f_2. In training, their cost function is defined as the negative cosine similarity between the two views. See for a detailed explanation [39].

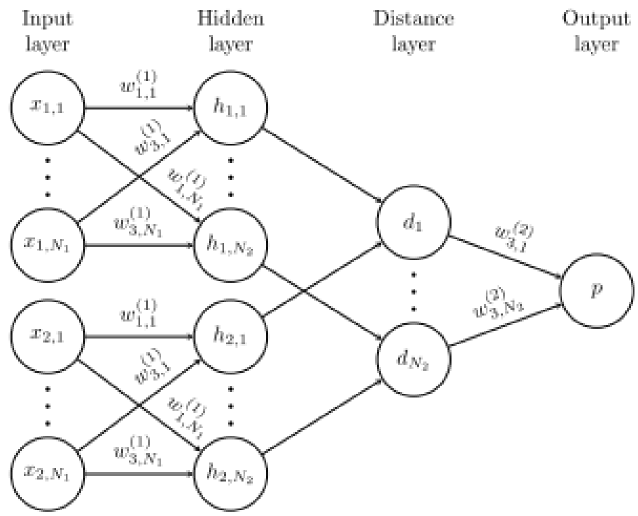

Figure 1: From Koch [22]: Simple 2-hidden-layer Siamese Network for binary classification with logistic prediction p. Top and bottom networks are twins with shared weight matrices.

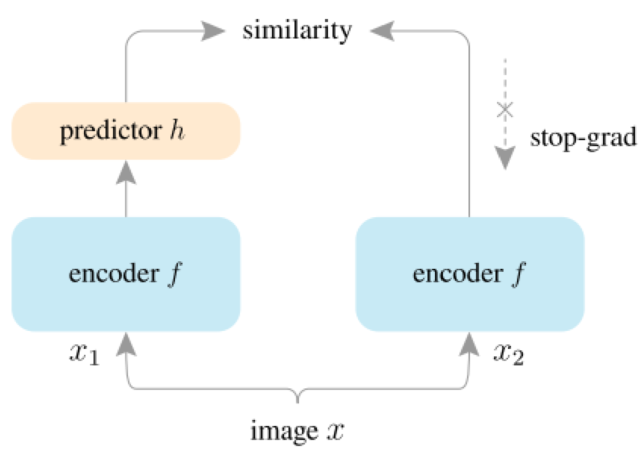

Figure 2: SimSiam Architecture, from Chen and He [39], page 1

Figure 3: Dual-fisheye (raw) and equirectangular footage

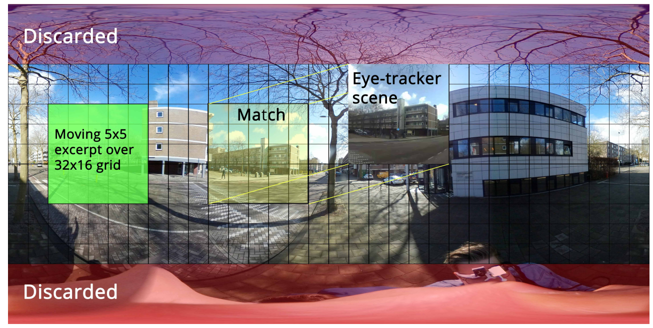

Figure 4: Visual explanation of comparison check.

4 RESEARCH QUESTION

The question that is aimed to answer is as follows:

> How can an aggregated AOI 360-degree map of several wearable eye-trackers be created?

4.1 Sub questions

- How well does FAST (expand) detect fragment location?
- How well can a SNN detect fragment location?

5 METHODOLOGY

To develop a proof of concept (PoC), the initial concept is taking one master-image in 360 degrees and using video footage from the wearable eye-tracker to map on this image. A 360-degree image mainly exists in two forms: dual fish-eye, which is the raw output of two lenses on the camera, and equirectangular, which is a stitched and rectangular projection of the source. See an example of this behavior on Figure 3 on page 3. In order to compare the eye-tracker to this image, we will use the equirectangular projection.

In this stage, mapping will only work when the participant wearing the eye-tracker is standing at the same point as the master-image was created. As the eye-tracker returns a JSON-file with the estimated coordinates of the tracked pupil per frame, this can be mapped on the master image when its location is determined, see Figure 4. In order to compare and evaluate the different methods which will be tried, the footage from the eye-tracker will have to be hand-labeled to establish a ground truth. While traditional computer vision models can output an estimated location directly, a SNN needs to compare two images. In this case, every frame will be compared to an extracted grid of unrolled footage of the 360-degree camera, from which the most similar will be saved. The performance of the network will be measured in distance between the predicted position to the target (hand-labeled) position in pixels, where some margin can be taken in labelling a guess as a correct as the base frames will not always be a perfect fit.

The workflow and procedure in this experiment is therefore as follows:

(1) Grab image or video in 360-degree as master-track.
(2) Unroll, stabilize and crop 360-degree footage.
(3) Collect eye-tracker footage and data.
(4) Hand-label correct location of eye-tracker footage.
(5) In case of SNN: Divide 360-degree footage in excerpts to compare to eye-tracker.
(6) Compare accuracy of different mapping methods.

5.1 Computer vision

Regarding the computer vision architectures, both SIFT and FAST were selected as a baseline, using the existing open implementation by OpenCV2 [44, 45]. For the SNN, an open source implementation on Github with pretrained weights was found and used[1].

6 EXPERIMENTS

6.1 Kexxu OpenEye

The eye-tracker used in this project is a beta-version of OpenEye, a prototype wearable eye-tracker made by Kexxu, see **??**. This version uses a pre-trained neural network on a wearable Raspberry PI to interpret pupil location in real time[2]. While it is normal for eye-trackers to incorporate and distinguish between saccades and fixations [9], this eye-tracker was not equipped with this capability. A 3D-printed wearable frame with one pupil-facing camera and one scene-facing camera are combined directly in one combined MP4 video-file and a JSON-file with relative focus positions. Every frame was center-cropped to 720x720 pixels in order to be used in the different image recognition methods.

The footage grabbed for this Proof of Concept was 18 seconds of footage, a total of 245 frames, at a single location with an accurate embedded eye-tracking registration. The location for this initial test was in Amsterdam, near the office of Kexxu at the A. J. Ernststraat

[1] See https://github.com/taoyang1122/pytorch-SimSiam for this implementation
[2] See https://kexxu.com for more details about the eye-tracker used.

in Amsterdam. This is an urban location with plenty of possible features to be extracted.

6.2 360-Degree Camera

The camera used for grabbing the master-track, in this case a 360-degree picture, is the Samsung Gear 360 II which can grab both images and videos in 360-degrees[3]. The footage generated by this camera was pre-processed using Cyberlink ActionDirector.

6.3 Image labelling

To label the correct position of each frame of the eye-tracker camera, a simple Flask-React service was built to hand-label the correct position and validate the accuracy of several methods[4].

7 RESULTS

7.1 SIFT and FAST

For trying out the functionality of SIFT and FAST, the recommended code by OpenCV was used. [44, 45]. A couple of samples were attempted, as seen in Figure 7 on 7. While these algorithms do get some points correct, these directions are not consistent enough to provide any real information. This method has not been attempted further than these samples.

7.1.1 Discussion. The lower scores generated by these methods can be explained by their designed nature. These size-invariant feature detection methods are great at detecting similar items based on corners, but the warped images generated by both the eye-tracker and the 360-degree camera could have been a reason for this malfunction.

7.2 Siamese Network

The SNN used has been pre-trained using the resources in the github-repository. As explained in the methodology, all 245 frames of the eye-tracker were run through the SNN, as well as the grid of 31x5 extracts of the base image. These images were compared using the negative cosine similarity metric as proposed in the original paper. Using this method, an accuracy of estimation within 2 frames around the center track was created of **38.6%**, see some examples and their scores (expressed as deviation of 181px, 1 frame) in Figure 6 on page 5. See a compiled estimated location of eye-tracker frames on the master-track in the video on https://youtu.be/x9i05IzH_Cs. A distribution of the deviation from the target point in pixels is visible in Figure 5.

7.2.1 Discussion. While this SNN showed potential, its accuracy does seem too low. As visible in Figure 5, a big part is within the 2 frames of deviation from the target position. This is the part that could be explained, as similar features exist both in the source and target frame. The parts determined beyond the first spike are created by noise, and are incorrect. These are, for example, the third, fourth and fifth image in the examples in Figure 6.

[3]For specifications, see https://www.samsung.com/global/galaxy/gear-360/.

[4]See the GitHub repository for this service.

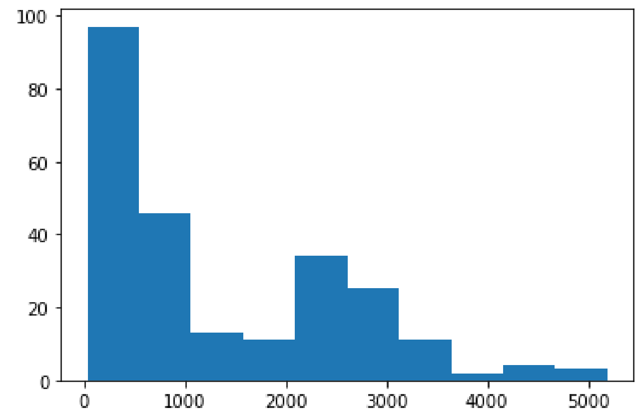

Figure 5: Binned amount of deviation from the target point.

8 CONCLUSION

While object detection by feature extraction is increasingly powerful, we have not been able to create a correctly functioning version in this project. While the SIFT and FAST networks showed potential, their output did not robustly show one position. The used Siamese Neural Network has a showed accuracy of 68.1%, we reached an accuracy of 38.6% within a margin of 200 pixels around the focus point. While it should be possible to make adjustments, such as integrating the dimension of time to improve these numbers, this was beyond the scope of this project.

9 ACKNOWLEDGEMENTS

Thank you to Jurriaan Schreuder for providing resources and guidance for this thesis, and an inspirational space to work with colleagues. I learned a lot while working on this project, for which I am thankful. A special thanks to the people I have discussed this idea with and have provided me with guidance, such as Sander Buningh and Marco te Brommelstroet. Lastly, thank you to my supervisor at the University of Amsterdam: Marcel Worring for his feedback and supervision.

10 REFERENCES

[1] Marcel Adam Just and Patricia A Carpenter. "A theory of reading: From eye fixations to comprehension". In: *Psychological Review* 87.4 (1980), pp. 329–354. ISSN: 0033295X. DOI: 10.1037/0033-295X.87.4.329.

[2] Jane Bromley et al. "Signature verification using a "Siamese" Time-delay Neural Network". In: *International Journal of Pattern Recognition and Artificial Intelligence* 07.04 (1993), pp. 669–688. ISSN: 0218-0014. DOI: 10.1142/s0218001493000339.

[3] David G. Lowe. "Object recognition from local scale-invariant features". In: *Proceedings of the IEEE International Conference on Computer Vision.* Vol. 2. 1999, pp. 1150–1157. DOI: 10.1109/iccv.1999.790410.

[4] Li Fei-Fei, Rob Fergus, and Pietro Perona. "A Bayesian approach to unsupervised one-shot learning of object categories". In: *Proceedings of the IEEE International Conference*

Figure 6: Frames and distance (in frames) from the target coordinates.

on Computer Vision. Vol. 2. 2003, pp. 1134–1141. doi: 10.1109/iccv.2003.1238476.

[5] Boris M Velichkovsky et al. "Visual fixations as a rapid indicator of hazard perception". In: *Operator functional state : the assessment and prediction of human performance degradation in complex tasks* (2003), pp. 313–321. issn: 1566-7693. url: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.569.8939&rep=rep1&type=pdf.

[6] L Fei-Fei. "Knowledge transfer in learning to recognize visual objects classes". In: *Proceedings of the Fifth International Conference ...* (2006). url: http://www-cs.stanford.edu/groups/vision/documents/Fei-Fei_ICDL2006.pdf.

[7] Edward Rosten and Tom Drummond. "Machine learning for high-speed corner detection". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 3951 LNCS. 2006, pp. 430–443. isbn: 3540338322. doi: 10.1007/11744023{_}34.

[8] Andrew Duchowski. *Eye tracking methodology: Theory and practice*. 2007, pp. 1–328. isbn: 9781846286087. doi: 10.1007/978-1-84628-609-4.

[9] Roger PG van Gompel et al. *Eye-movement research: An overview of current and past developments*. 2007, pp. 1–28. doi: https://doi.org/10.1016/B978-008044980-7/50003-3. url: http://marefateadyan.nashriyat.ir/node/150.

[10] G. Underwood. "Visual attention and the transition from novice to advanced driver". In: *Ergonomics* 50.8 (2007), pp. 1235–1249. issn: 00140139. doi: 10.1080/00140130701318707.

[11] Michelle L. Reyes and John D. Lee. "Effects of cognitive load presence and duration on driver eye movements and event detection performance". In: *Transportation Research Part F: Traffic Psychology and Behaviour* 11.6 (2008), pp. 391–402. issn: 13698478. doi: 10.1016/j.trf.2008.03.004. url: http://dx.doi.org/10.1016/j.trf.2008.03.004.

[12] B. De Geus, J. Joncheere, and R. Meeusen. "Commuter cycling: Effect on physical performance in untrained men and women in Flanders: Minimum dose to improve indexes of fitness". In: *Scandinavian Journal of Medicine and Science in Sports* 19.2 (2009), pp. 179–187. issn: 09057188. doi: 10.1111/j.1600-0838.2008.00776.x.

[13] Lior Wolf, Tal Hassner, and Yaniv Taigman. "The one-shot similarity kernel". In: *Proceedings of the IEEE International Conference on Computer Vision* (2009), pp. 897–902. doi: 10.1109/ICCV.2009.5459323.

[14] Ingrid J.M. Hendriksen et al. "The association between commuter cycling and sickness absence". In: *Preventive Medicine* 51.2 (2010), pp. 132–135. issn: 00917435. doi: 10.1016/j.ypmed.2010.05.007. url: http://dx.doi.org/10.1016/j.ypmed.2010.05.007.

[15] Nils Osbeck Emelie; Åkerman. "Information Hold: Ways of preventing information overload in Scania vehicles in critical traffic situations". PhD thesis. KTH, 2010.

[16] David Crundall and Geoffrey Underwood. *Visual attention while driving: Measures of eye movements used in driving research*. Elsevier, 2011, pp. 137–148. isbn: 9780123819840. doi: 10.1016/B978-0-12-381984-0.10011-6. url: http://dx.doi.org/10.1016/B978-0-12-381984-0.10011-6.

[17] Bo Hua Liu, Li Shan Sun, and Jian Rong. "Driver's visual cognition behaviors of traffic signs based on eye movement parameters". In: *Jiaotong Yunshu Xitong Gongcheng Yu Xinxi/Journal of Transportation Systems Engineering and Information Technology* 11.4 (2011), pp. 22–27. issn: 10096744. doi: 10.1016/s1570-6672(10)60129-8. url: http://dx.doi.org/10.1016/S1570-6672(10)60129-8.

[18] Julia Werneke and Mark Vollrath. "What does the driver look at? the influence of intersection characteristics on attention allocation and driving behavior". In: *Accident Analysis and Prevention* 45 (2012), pp. 610–619. issn: 00014575. doi: 10.1016/j.aap.2011.09.048. url: http://dx.doi.org/10.1016/j.aap.2011.09.048.

[19] Jun Wan et al. "One-shot learning gesture recognition from RGB-D data using bag of features". In: *Journal of Machine Learning Research* 14 (2013), pp. 2549–2582. issn: 15324435. doi: 10.1007/978-3-319-57021-1{_}11.

[20] Esko Lehtonen et al. "Effect of driving experience on anticipatory look-ahead fixations in real curve driving". In: *Accident Analysis and Prevention* 70 (2014), pp. 195–208. issn: 00014575. doi: 10.1016/j.aap.2014.04.002. url: http://dx.doi.org/10.1016/j.aap.2014.04.002.

[21] Evelyne St-Louis et al. "The happy commuter: A comparison of commuter satisfaction across modes". In: *Transportation Research Part F: Traffic Psychology and Behaviour* 26.PART A (2014), pp. 160–170. issn: 13698478. doi: 10.1016/j.trf.2014.07.004. url: http://dx.doi.org/10.1016/j.trf.2014.07.004.

[22] Gregory Koch. “Siamese Thesis”. In: *Cs.Toronto.Edu* 2 (2015). URL: http://www.cs.toronto.edu/~gkoch/files/msc-thesis.pdf.

[23] Pieter Vansteenkiste et al. “Measuring dwell time percentage from head-mounted eye-tracking data – comparison of a frame-by-frame and a fixation-by-fixation analysis”. In: *Ergonomics* 58.5 (2015), pp. 712–721. ISSN: 13665847. DOI: 10.1080/00140139.2014.990524. URL: https://doi.org/10.1080/00140139.2014.990524.

[24] Esko Lehtonen et al. “Evaluating bicyclists’ risk perception using video clips: Comparison of frequent and infrequent city cyclists”. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 41 (2016), pp. 195–203. ISSN: 13698478. DOI: 10.1016/j.trf.2015.04.006. URL: http://dx.doi.org/10.1016/j.trf.2015.04.006.

[25] Linus Zeuwts et al. “Is gaze behaviour in a laboratory context similar to that in real-life? A study in bicyclists”. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 43 (2016), pp. 131–140. ISSN: 13698478. DOI: 10.1016/j.trf.2016.10.010. URL: http://dx.doi.org/10.1016/j.trf.2016.10.010.

[26] Nicola Bongiorno et al. “How is the Driver’s Workload Influenced by the Road Environment?” In: *Procedia Engineering* 187 (2017), pp. 5–13. ISSN: 18777058. DOI: 10.1016/j.proeng.2017.04.343. URL: http://dx.doi.org/10.1016/j.proeng.2017.04.343.

[27] Alessandra Mantuano, Silvia Bernardi, and Federico Rupi. “Cyclist gaze behavior in urban space: An eye-tracking experiment on the bicycle network of Bologna”. In: *Case Studies on Transport Policy* 5.2 (2017), pp. 408–416. ISSN: 22136258. DOI: 10.1016/j.cstp.2016.06.001. URL: http://dx.doi.org/10.1016/j.cstp.2016.06.001.

[28] P. Schepers et al. “The Dutch road to a high level of cycling safety”. In: *Safety Science* 92 (2017), pp. 264–273. ISSN: 18791042. DOI: 10.1016/j.ssci.2015.06.005. URL: http://dx.doi.org/10.1016/j.ssci.2015.06.005.

[29] S de Vries. “Using a wearable eye-tracking device on bicyclists to explore the possibity of measuring motorcyclist eye movements.” In: (2017), pp. 1–37. URL: http://essay.utwente.nl/73485/.

[30] Martin Berger and Linda Dörrzapf. “Sensing comfort in bicycling in addition to travel data”. In: *Transportation Research Procedia* 32 (2018), pp. 524–534. ISSN: 23521465. DOI: 10.1016/j.trpro.2018.10.034. URL: https://doi.org/10.1016/j.trpro.2018.10.034.

[31] Tomaž Čegovnik et al. “An analysis of the suitability of a low-cost eye tracker for assessing the cognitive load of drivers”. In: *Applied Ergonomics* 68.September 2017 (2018), pp. 1–11. ISSN: 18729126. DOI: 10.1016/j.apergo.2017.10.011.

[32] N. Kovácsová et al. “Cyclists’ eye movements and crossing judgments at uncontrolled intersections: An eye-tracking study using animated video clips”. In: *Accident Analysis and Prevention* 120.July (2018), pp. 270–280. ISSN: 00014575. DOI: 10.1016/j.aap.2018.08.024. URL: https://doi.org/10.1016/j.aap.2018.08.024.

[33] Mathias Trefzger et al. “A visual comparison of gaze behavior from pedestrians and cyclists”. In: *Eye Tracking Research and Applications Symposium (ETRA)* (2018). DOI: 10.1145/3204493.3204553.

[34] United Nations. *68% of the world population projected to live in urban areas by 2050*. 2018. URL: https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html.

[35] United Nations. *Resolution 72/272 adopted by the General Assembly*. Tech. rep. April. 2018, pp. 71–73. URL: https://undocs.org/Home/Mobile?FinalSymbol=A%2FRES%2F72%2F272&Language=E&DeviceType=Desktop.

[36] Samuel Nello-Deakin. “Is there such a thing as a ‘fair’ distribution of road space?” In: *Journal of Urban Design* 24.5 (2019), pp. 698–714. ISSN: 14699664. DOI: 10.1080/13574809.2019.1592664. URL: https://doi.org/10.1080/13574809.2019.1592664.

[37] Federico Rupi and Kevin J. Krizek. “Visual eye gaze while cycling: Analyzing eye tracking at signalized intersections in urban conditions”. In: *Sustainability (Switzerland)* 11.21 (2019). ISSN: 20711050. DOI: 10.3390/su11216089.

[38] United Nations. *Special edition: progress towards the Sustainable Development Goals*. 2019. URL: https://sustainabledevelopment.un.org/sdg11.

[39] Xinlei Chen and Kaiming He. “Exploring Simple Siamese Representation Learning”. In: Figure 1 (2020). URL: http://arxiv.org/abs/2011.10566.

[40] ERSO. *European Road Safety Observatory: Facts and Figures - Cyclists*. Tech. rep. 20206. 2020, pp. 1–23.

[41] Stefan Gössling. “Why cities need to take road space from cars - and how this could be done”. In: *Journal of Urban Design* 25.4 (2020), pp. 443–448. ISSN: 14699664. DOI: 10.1080/13574809.2020.1727318. URL: https://doi.org/10.1080/13574809.2020.1727318.

[42] Julia Kohl et al. “Driver glance behavior towards displayed images on in-vehicle information systems under real driving conditions”. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 70 (2020), pp. 163–174. ISSN: 13698478. DOI: 10.1016/j.trf.2020.01.017. URL: https://doi.org/10.1016/j.trf.2020.01.017.

[43] Kevin J. Krizek, Bert Otten, and Federico Rupi. “EMERGING TRANSPORT FUTURES FOR STREETS AND HOW EYE TRACKING CAN HELP IMPROVE SAFETY AMD DESIGN”. In: *Urban Experience and Design: Contemporary Perspectives on Improving the Public Realm*. 2020, pp. 140–144. ISBN: 9781000178357. DOI: 10.4324/9780367435585.

[44] OpenCV. *FAST Algorithm for Corner Detection*. 2020. URL: https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_feature2d/py_fast/py_fast.html.

[45] OpenCV. *Introduction to SIFT (Scale-Invariant Feature Transform)*. 2020. URL: https://docs.opencv.org/master/da/df5/tutorial_py_sift_intro.html.

[46] Jork Stapel, Mounir El Hassnaoui, and Riender Happee. “Measuring Driver Perception: Combining Eye-Tracking and Automated Road Scene Perception”. In: *Human Factors* (2020). ISSN: 15478181. DOI: 10.1177/0018720820959958.

[47] Wikipedia. *One-Shot Learning*. 2021. URL: https://en.wikipedia.org/wiki/One-shot_learning.

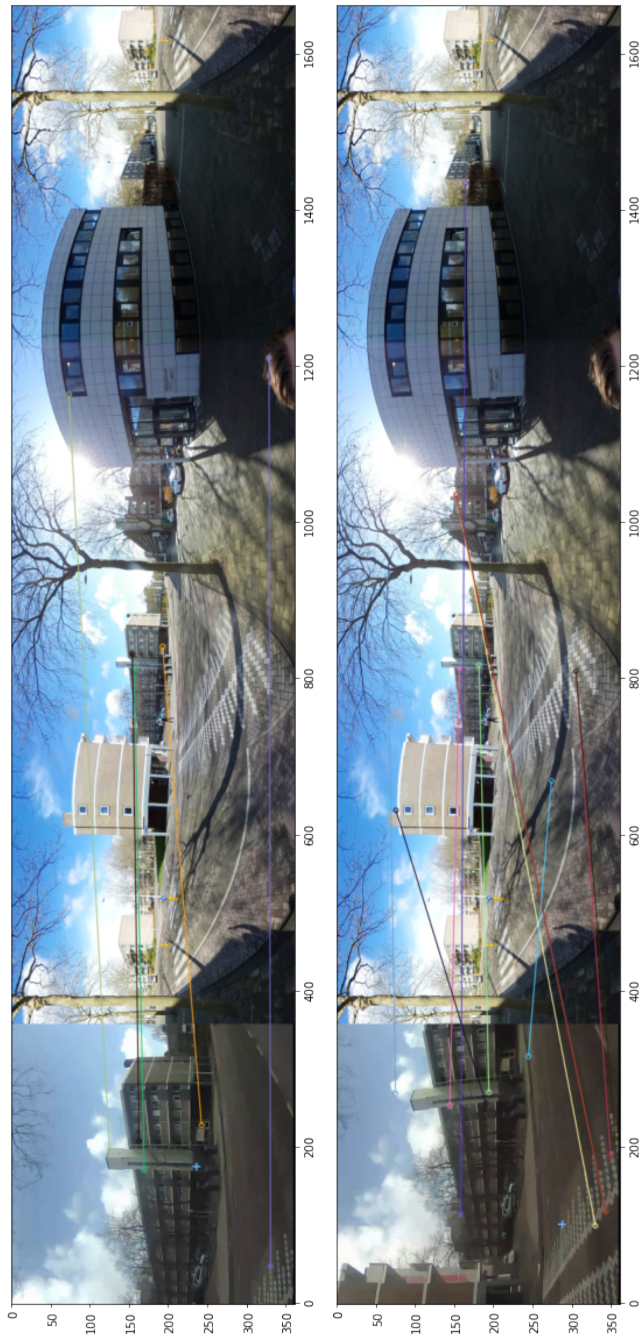

Figure 7: Two sample explorations using the SIFT feature extraction.