Automated detection of e-scooter helmet use with deep learning

Felix W. Siebert[*], Christoffer Riis[*], Kira H. Janstrup[*], Jakob Kristensen[*], Oguzhan Gül[*], Hanhe Lin[#], Frederik B Hüttel[†*]

[*]Department of Technology, Management and Economics
Technical University of Denmark
Bygningstorvet 116, 2800, Lyngby, Denmark
email: felix@dtu.dk, chrrii@dtu.dk, kija@dtu.dk, s173174@student.dtu.dk, s194273@student.dtu.dk, fbohy@dtu.dk

[#]National Subsea Centre
Robert Gordon University
Garthdee Rd, Aberdeen AB10 7AQ, Scotland
email: h.lin2@rgu.ac.uk

[†]Institute for Data, Systems, and Society (IDSS)
Massachusetts Institute of Technology
E18-407A, 50 Ames St, 02142, Cambridge, Massachusetts, United States of America
email: fbohy@mit.edu

Keywords: e-scooter safety, injury prevention, helmets, deep learning, computer vision

1 INTRODUCTION

E-scooter riders have an increased crash risk compared to cyclists [1]. Hospital data finds increasing numbers of injured e-scooter riders, with head injuries as one of the most common injury types [2]. To decrease this high prevalence of head injuries, the use of e-scooter helmets could present a potential countermeasure [3]. Despite this, studies show a generally low rate of helmet use rates in countries without mandatory helmet use laws [4][5][6]. In countries with mandatory helmet use laws for e-scooter riders, helmet use rates are higher, but generally remain lower than bicycle use rates [7]. As the helmet use rate is a central factor for the safety of e-scooter riders in case of a crash and a key performance indicator in the European Commission's Road Safety Policy Framework 2021-2030 [8], efficient e-scooter helmet use data collection methods are needed. However, currently, human observers are used to register e-scooter helmet use either in direct roadside observations or in indirect video-based observation, which is time-consuming and costly. In this study, a deep learning-based method for the automated detection of e-scooter helmet use in video data was developed and tested, with the aim to provide an efficient data collection tool for road safety researchers and practitioners.

2 METHODS

Video data was collected from cycling lanes at two observation sites in Copenhagen, Denmark. In Denmark, it is mandatory for e-scooter riders to use the cycling infrastructure and wear a helmet. To only capture one instance of individual e-scooter riders passing the observation camera, the frame per second (fps) rate in the video data was reduced from 30 fps to 5 fps. As there is no trained object detection algorithm for detecting e-scooters including their riders (*active e-scooters*), an existing object detection algorithm [9] for the detection of persons was used as a first step. This process also detected persons on e-scooters (i.e. e-scooter riders). Once the algorithm identified all the frames with persons, frames with active e-scooters were manually selected for further processing. Then, a rectangular bounding box was drawn around e-scooters and their riders. For each bounding box, helmet use was registered (yes/no). Blurry or partly covered active e-scooters were not annotated. Using this approach, we identified a total of 558 samples of e-scooters (Table 1; n=244 with helmet; n=314 without helmet). This distribution does not represent the actual helmet use distribution in the observed cycle lanes, as not all collected video data was processed, and an effort was made to balance the number of helmet users and non-users in the dataset. To increase the robustness of our model, we added 114 creative commons images of e-scooters as a *robustness dataset* (consisting of 52 e-scooter riders with helmets, and 62 without). We then split the data up into a training (55%), validation (14%), and test set (26%) (Table 2). As implied by the name, we use the training set to train the detection model, and the validation set to validate the

model and prevent overfitting. After training, we assess the model's performance by evaluating it against a test set, which consists of frames that have not been part of the training and validation sets. We end with 374 objects in the training set, 99 objects in the validation set and 174 objects in the test set, some images can contain multiple objects. The overall helmet use percentage of the 174 riders in the test set was 65.5% (114 with helmet and 60 without helmet). We train the YOLO-R object detection model for 500 epochs and save the weights from the best performing model on the validation set. The model is trained with default hyperparameters [9].

Table 1: Class distribution of e-scooters in the dataset.

Class	Total Objects	Site 1	Site 2	Robustness Set
E-scooters without helmet	377	123	191	62
E-scooters with helmet	296	140	104	52
Total	673	263	295	114

To evaluate the detection accuracy, we compute descriptive statistics - true positive (TP), true negative (TN), false positive (FP), false negative (FN) - to calculate the Precision and Recall and the mean Average Precision (mAP) of the detection. Precision is calculated by TP / (TP + FP) and describes how accurate the model is, when it predicts an object to belong to a specific class. Recall is calculated by TP / (TP + FN) and describes how many of the actual objects the model detects within a specific class. The two metrics Precision and Recall create a trade-off between correctly classifying each object and capturing all objects, i.e. if precision is maximized, the model will only detect an objects if it is very confident, thus resulting in few but correct detections, whereas if recall is maximized, the model will not miss objects, but also detect many things that are not of interest, thus resulting in detecting all instance alongside many irrelevant objects. The model evaluation should be independent of this trade-off, and thus it is the standard within object detection to use the (mAP) as an evaluation metric. The AP is calculated as the area under the curve of the Precision Recall Curve, i.e. the average precision for all trade-offs. The mAP is the mean of the APs for the different classes. Accuracy measures are first calculated for detecting active e-scooters (i.e. e-scooter plus rider regardless of the rider's helmet use) and then calculated for helmet use detection.

Table 2: Training, validation, and test set split of the dataset.

Class	Total Objects	Training set	Validation set	Test set
E-scooters without helmet	377	257	60	60
E-scooters with helmet	296	143	39	114
Total	673	374	99	174

Table 3: Model performance on the test set, with a confidence threshold of 0.5.

Class	Total Objects	Precision	Recall	AP@.5
Active E-scooter	174	0.946	0.709	0.981
E-scooter with Helmet	114	0.729	0.833	0.882
E-scooter without Helmet	60	0.503	0.900	0.693
Average over the classes	174	0.616	0.867	0.787

3 RESULTS

Of all 174 e-scooters in the test dataset, 122 (TP) were detected as e-scooters (regardless of rider helmet use). 52 instances of e-scooters were not detected (FN) and 7 e-scooter detections did not actually identify an e-scooter (FP). This represents a Precision of 0.946, a Recall of 0.709, and an *mAP* of 0.981 for the detection of active e-scooter riders (Table 3). For all the riders with helmets, 83 instances of helmet use were correctly identified. In 30 instances, a rider was registered as using a helmet although they were not using one (FP), and in 29 instances a rider was registered as not using a helmet although they were using one (FN). For the calculation of precision, recall and *mAP*, these instances of misclassified helmet use as well as instances, in

which the whole e-scooter was either missed or falsely detected are taken into account. Precision for helmet use detection was 0.616, and Recall was 0.867, representing an *mAP* of 0.787. Algorithm-registered helmet use for the test set data was 47.7% (see Figure 1 for examples of detections on test set).

Fig 1: Sample frames from detections on the test set (anonymization added for privacy data protection).

4 DISCUSSION

We investigated the feasibility of a deep learning-based object detection algorithm for the detection of e-scooter riders' helmet use in video data. For the general detection of active e-scooter riders, our study revealed an accuracy of *mAP 0.981*. This indicates that deep learning-based e-scooter detection can be useful tool to detect and extract occurrences of e-scooter riders in video data. For e-scooter riders' helmet use, the trained algorithm registered a helmet use rate of 47.7% for the test set data, which is considerably lower than the actual helmet use in the test data (65.5%). Further analysis of the accuracy measures reveals comparatively high inaccuracies in the precision of the detection of e-scooters without helmets. Further research is needed to increase accuracy in helmet use registration.

REFERENCES

[1] L. Gebhardt, C. Wolf, S. Ehrenberger, R. Seiffert, D. Krajzewicz and R. Cyganski, E-Scooter-Potentiale, Herausforderungen und Implikationen für das Verkehrssystem: Abschlussbericht Kurzstudie E-Scooter. https://elib.dlr.de/141837/1/ArbeitsberichteVF_Nr4_2021.pdf, 2021 (accessed 28.04.2022).

[2] M. Aizpuru, K.X. Farley, J.C. Rojas, R.S. Crawford, T.J. Moore, E.R. Wagner, Motorized scooter injuries in the era of scooter-shares: A review of the National Electronic Surveillance System, *The American Journal of Emergency Medicine*. 37 (2019) 1133–1138. doi:10.1016/j.ajem.2019.03.049.

[3] D. Uluk, T. Lindner, M. Dahne, J.W. Bickelmayer, K. Beyer, A. Slagman, et al., E-scooter incidents in Berlin: An evaluation of risk factors and injury patterns, *Emergency Medicine Journal*. 39 (2021) 295–300. doi:10.1136/emermed-2020-210268.

[4] T.K. Trivedi, C. Liu, A.L. Antonio, N. Wheaton, V. Kreger, A. Yap, et al., Injuries associated with Standing Electric Scooter use, *JAMA Network Open*. 2 (2019). doi:10.1001/jamanetworkopen.2018.7381.

[5] T. Petzoldt, M. Ringhand, J. Anke, N. Schekatz, Do German (non)users of e-scooters know the rules (and do they agree with them)?, *HCI in Mobility, Transport, and Automotive Systems*. (2021) 425–435. doi:10.1007/978-3-030-78358-7_29.

[6] F.W. Siebert, M. Ringhand, F. Englert, M. Hoffknecht, T. Edwards, M. Rötting, Braking bad – ergonomic design and implications for the safe use of shared e-scooters, *Safety Science*. 140 (2021) 105294. doi:10.1016/j.ssci.2021.105294.

[7] N. Haworth, A. Schramm, D. Twisk, Comparing the risky behaviours of shared and private e-scooter and bicycle riders in downtown Brisbane, Australia, *Accident Analysis & Prevention*. 152 (2021) 105981. doi:10.1016/j.aap.2021.105981.

[8] European Commission, EU Road Safety Policy Framework 2021-2030 - Next steps towards "Vision Zero". https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/move-2019-01178-01-00-en-tra-00_3.pdf, 2019 (accessed 28.04.2022).

[9] C.-Y. Wang, I. Yeh, H.-Y. M. Liao, You only learn one Representation: unified network for multiple Tasks, *Arxiv Preprint*. doi:10.48550/arXiv.2105.04206.