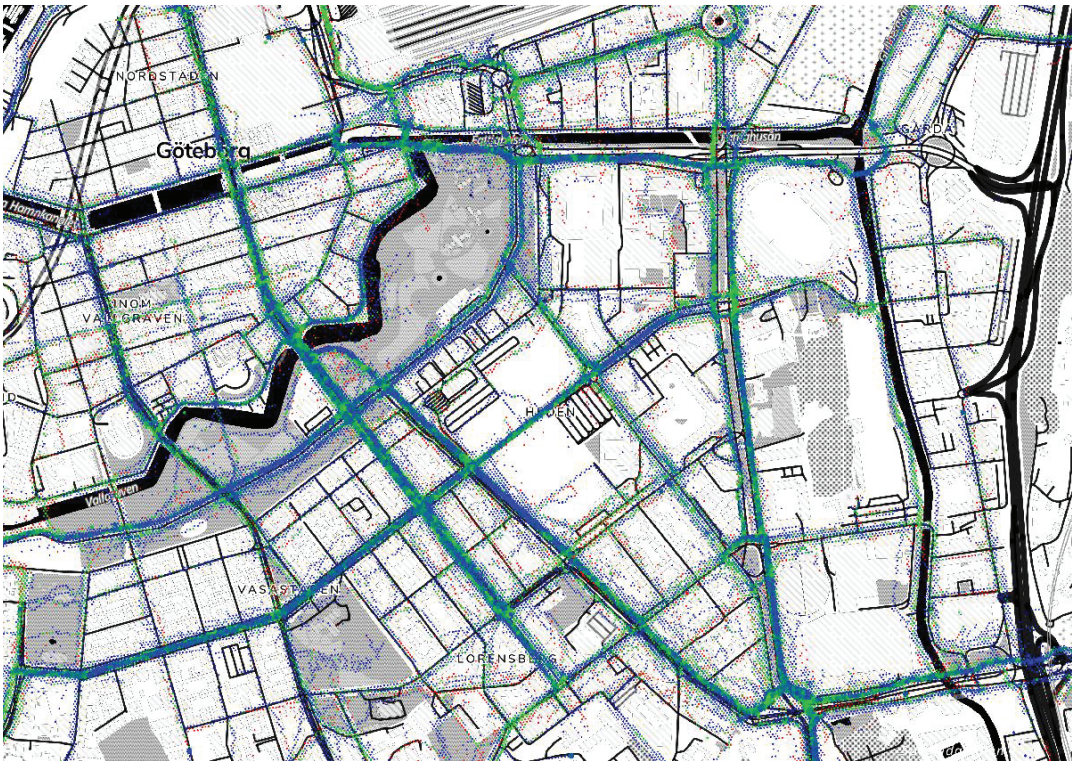
E-safe pre-study



Authors:	Christian-Nils Boda, Jawwad Ahmed, Raffaello Baluyot, Jolyon Carroll, Kristin Eklöf, Rahul Rajendra Pai, André Dankert
Date:	2023-01-14
Project within:	Trafiksäkerhet och automatiserade fordon

FFI Fordonsstrategisk Forskning och Innovation

VINNOVA Energimyndigheten TRAFIKVERKET FKG VOLVO SCANIA VOLVO

Table of Contents

1 Abstract **3**

2 Sammanfattning på Svenska **3**

3 Introduction **5**

3.1 Background 5

3.2 Objective and Research Questions 5

4 Methods **7**

4.1 Data Collection 7

4.2 Data Processing and Exploration 10

4.3 Safety-Critical Event Detection 17

4.4 Infrastructure and Rider Behaviours 19

4.5 Road-User Interaction Modelling 23

5 Results **24**

5.1 Data Collection Summary 24

5.2 Data Processing and Exploration 29

5.3 Safety-Critical Event Detection 30

5.4 Infrastructure and Rider Behaviours 32

5.5 Road-User Interaction Modelling 42

6 Discussion **52**

6.1 Research Question 1 52

6.2 Research Question 2 52

6.3 Research Question 3 54

7 Dissemination and Publications **56**

7.1 Dissemination 56

7.2 Publications 56

8 Conclusion and Future Work **57**

9 Bibliography **57**

10 Participating parties and contact persons **60**

1 Abstract

As shared e-scooter riding gains popularity, concerns about safety keep growing. While some site observation studies or crash statistics literature exist for e-scooter riding, naturalistic e-scooter riding data is rare. However, naturalistic riding studies are one of the few ways to study how people normally behave (or misbehave) in an ecologically valid environment. The e-safe pre-study instrumented seventeen e-scooters that were shared in the streets of Gothenburg to record riders' behaviour to better understand 1) how crashes could be reconstructed with information from basic sensors, 2) if unsafe behaviours could be detected with only inertial motion sensors and GNSS sensors, and 3) if the collected data were sufficient for modelling road-user interactions. The work in this pre-study was structured into three main parts: 1) data instrumentation and collection, 2) data processing, and 3) data analysis.

The data collected corresponds to 2315 distinct trips between July 2022 and November 2022. The safety-critical event detector devised in the project detected minor crashes and falls. No severe event was observed in the data which is in line with crash and insurance claim statistics as severe events are rather seldom. The collected data were not sufficient for a full crash reconstruction as no body motion was captured by the video camera during the events, but they still provide a lot of information that could be used in crash tests or simulations. Such information could be the initial speed or type of road surface when the fall occurred.

Results showed that the collected data could be used to identify some unsafe behaviours, for instance, unsafe speed behaviour given the road surface condition. The results also showed that shared e-scooters may be a means to estimate road surface condition which could be used by cities when planning or maintaining their road infrastructure. Additionally, video data were used to study hand placement on the handlebars, and it was shown that most of the time people hold their e-scooters with two hands and tend to anticipate braking when approaching risky spots (e.g., intersections).

The data were shown to be a rich resource for road-user interaction modelling, despite some technical limitations that should be accounted for in the future. A comfort boundary was estimated for e-scooter riders to pedestrians; preliminary results show that riders do not let pedestrians come within 3.5 m in the longitudinal direction and 2.5 m in lateral directions.

Some limitations were discovered during the analyses, and they will need to be addressed in the future.

2 Sammanfattning på Svenska

När delade elsparkcyklar ökar i popularitet så växer också oron kring deras säkerhet. Även om vissa observationsstudier och olycksstatistik för elsparkcykelåkning finns, så är naturalistisk data från elsparkcykelåkning ovanligt. Likafullt är naturalistiska studier ett av få sätt

att studera personers normala beteende (både bra och dåligt) i en ekologiskt valid miljö. I e-safe förstudien instrumenterades 17 hyrelsparkcyklar i Göteborg för att samla in data kring åkarnas beteenden för att bättre förstå 1) hur man kan återskapa olycksförlopp med hjälp av enkla sensorer, 2) om riskfyllda beteenden kan identifieras med endast rörelsesensorer och GNSS-mottagare, samt 3) om den insamlade datan var tillräcklig för att modellera interaktionen med andra trafikanter. Arbetet i denna förstudie var uppdelat i tre huvuddelar: 1) instrumentering och datainsamling, 2) databehandling, och 3) dataanalys.

Den insamlade datan består av 2315 olika resor gjorda mellan juli 2022 och november 2022. Den i projektet utvecklade metoden för att identifiera säkerhetskritiska händelser fann mindre krascher och fall. Inga allvarliga incidenter observerades vilket är i linje med olycks- och försäkringsstatistik som visar att allvarliga olyckor är relativt ovanliga. Den insamlade datan var inte tillräcklig för att fullt ut kunna rekonstruera olyckstillfällen eftersom kroppens rörelse inte fångades av den kamera som användes, men datan ger dock ändå mycket information som skulle kunna användas i kraschtester eller simuleringar. Sådan information skulle kunna vara initial hastighet samt typ av vägbeläggning vid olycksplatsen.

Resultaten visar att den insamlade datan kunde användas för att identifiera en del riskbeteenden, som till exempel olämplig hastighet givet typen av vägbeläggning och dess tillstånd. Resultaten visade också att hyrelsparkcyklar kan vara ett sätt för städer att estimera vägbeläggningens skick, vilket skulle kunna användas för planering av nybyggnation eller underhåll av befintligt vägnät. Videodata användes dessutom för att studera hur händerna placerades, och det visades att de flesta greppar styret med två händer samt att de tenderar att förutse behov att bromsa när de närmar sig riskfyllda platser (t.ex. korsningar).

Datan har visat sig vara en guldgruva när det kommer till att modellera interaktionen mellan trafikanter, trots vissa tekniska begränsningar som kan åtgärdas i framtiden. En komfortgräns när det gäller avstånd från elsparkcykelákare till fotgängare estimerades, och preliminära resultat visar att åkare föredrar att hålla minst 3,5 m till framförvarande fotgängare, samt 2,5 m i sidled.

Vissa begränsningar upptäcktes under analysen av datan, och de behöver adresseras i framtiden.

3 Introduction

3.1 Background

As shared e-scooter riding gains popularity, concerns about the safety of e-scooters keep growing. Several studies have shown that numbers of e-scooter riders involved in crashes have increased in the past years. Most of the crashes are single-vehicle crashes [27], and it has been shown that risky riding behaviours may be the main contributing factor in safety-critical events as well as inadequate traffic infrastructure design [15]. Some of the most concerning behaviours were identified as riding 'under the influence' or riding with multiple riders. Whilst single-vehicle crashes account for most cases requiring hospitalization, sometimes the injuries to the rider come from collisions with other road users (for instance cars). It has been hypothesized that e-scooters are similar to bicycles in the types of collisions that occur and the injury risk for the rider [5, 24]. Note that, in other cases, it is not only the e-scooter riders that sustain an injury but also other road users (e.g., cyclists or pedestrians). A survey among cyclists indicated that a majority perceived e-scooters as rather dangerous [16]. Particularly, the interaction of e-scooters with pedestrians was shown to be very critical for safe and attractive road use, as explained in the Tøi study where most of the pedestrians described their interaction with e-scooter riders as very uncomfortable opposed to most of the riders describing it as comfortable [15]. While such a mismatch in opinion can be identified via surveys, there is no available naturalistic riding data that can support researchers in analysing pedestrian-rider interactions. Similarly, risky behaviours could be identified retrospectively via hospital admission statistics, or surveys, but the lack of naturalistic riding data does not enable researchers or system designers to identify risky behaviours during the ride or to develop data-led countermeasures that could prevent them.

Moreover, we can see, by using hospital data, and/or insurance claims, that most of the riders involved in e-scooter crashes sustained head and facial injuries [4, 17, 21]. Although the actual crash mechanism is unknown due to the lack of naturalistic data suitable for crash reconstructions. Unknown naturalistic crash mechanisms mean that researchers are limited when developing passive safety systems and their evaluation tools, such as human body models.

Many naturalistic driving studies (NDS) have been carried out focusing on light vehicles (SHRP2 [3], UDRIVE [8, 30], 100-CAR [18]), heavy vehicles (UDRIVE) or powered two-wheelers (UDRIVE). Some NDS focused on bicycles [6] or e-bicycles [7] but none have focused on e-scooters. Some observational studies have focused on e-scooter riding [28, 13] but they do not enable researchers to run in-depth behavioural analyses.

3.2 Objective and Research Questions

The main objective of the pre-study was to collect naturalistic riding data to provide new information on how e-scooter operators ride or crash, and on how they interact with other road users (e.g., pedestrians, or cars). The pre-study analyses focused on addressing

three main research questions:

R.Q.1 Can data loggers, including cameras, provide sufficient data to support crash reconstruction?

R.Q.2 Can we identify unsafe riding behaviour (multiple riders, intoxicated riding, pavement riding, etc.) based on accelerometer, gyroscope, and GNSS data only?

R.Q.3 Can we use the collected data to model road user interactions?

4 Methods

4.1 Data Collection

Seventeen of Voi's shared e-scooters were instrumented and used in this pre-study. They were made available to any Voi customers in Gothenburg's streets from July 2022 to November 2022. An ethical application was sent to *Etikprövningsmyndigheten* (Dnr 2022-02273-01) and no concern about ethics were found by the ethical board for collecting data from the instrumented e-scooters. No personal data nor demographics data were recorded and, as the video camera was road-facing, the rider's face was never recorded.

The instrumentation is based on a prototype developed during a Master's Thesis supervised by Chalmers University and Voi Technology AB [20]. In addition to the usual internet-of-things hub (IoT) that equipped Voi's shared e-scooters, a data logger was built and added to collect data from an array of sensors on the e-scooter corresponding to 22 variables logged at a frequency of 10 Hz. Those variables are: 1) GNSS information, 2) wheel speed, 3) 3-axis acceleration, 4) 3-axis angular speed, 5) brake levers' and gas throttle positions, 6) battery information, 7) direction-indicator lights, and 8) time. The orientation of the accelerometer and gyroscope sensors (embedded in an inertial measurement unit (IMU)) is shown in Figure 2. All these data are uploaded to cloud storage at the end of each ride. In addition to these datasets, forward-facing video data are captured which provides visual information of the riding environment and riding behaviour (i.e., contextual information). The IoT triggers the recording of the video data during the locking and unlocking activities. The video data are stored in an encrypted format and are uploaded to the servers when the e-scooters are within the range of the preconfigured Wi-Fi network. The camera uses an Arducam LS-32220 fish-eye lens that has a field-of-view (FoV) of 220° which captures the road ahead of the e-scooter but also the handlebar, the front wheel, and the lateral sides (see Figure 2 for a visualization of the FOV). The video data are captured at 0.4 megapixels at 30 fps (an example of video frame is shown in Figure 1).

Figure 1: Frame from a video captured by the e-scooter.

In order to sync the e-scooter data with the video data, a timestamp is included in the message sent from the IoT to the camera unit during the unlock operation. The camera module saves this timestamp along with the meta data, each video frame will include a timestamp that can be synced with the other data afterwards since the initial timestamp is also marked on the file generated by the data logger. Once both the video data and the kinematic data files are available on the cloud storage, each of the video files is matched to the corresponding kinematic data file based on this timestamp.

The complete setup on the scooter is as shown in Figure 3. More detailed information on the e-scooter instrumentation can be found in [20].

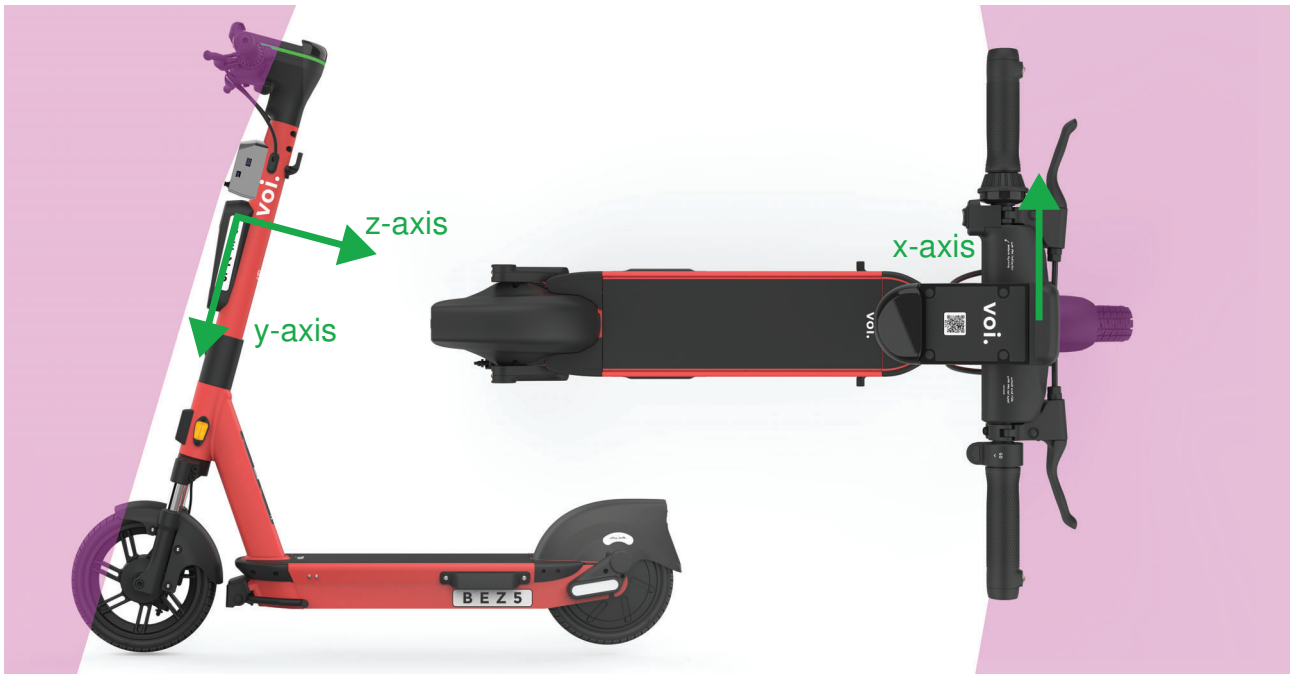


Figure 2: Visualization of the camera's field-of-view coloured in pink and orientation of the IMU reference frame.

Figure 3: Fleet of e-scooters equipped with camera modules.

4.2 Data Processing and Exploration

4.2.1 Data Transfer and Processing

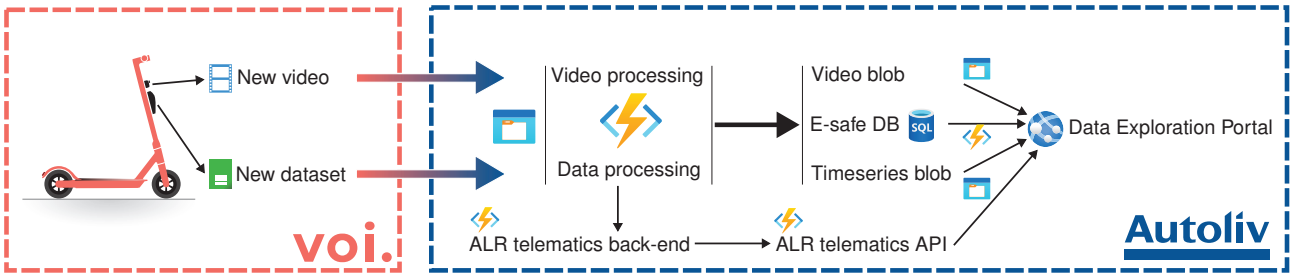


Figure 4: Visualization of the automated data flow processing from the e-scooter instrumentation to the data exploration portal.

When enough e-scooter data were collected and synced, they were manually transferred to Autoliv's storage. Upon the receipt of the new data batch, automatic processes are triggered to enrich the datasets and transform the video data to simplify further analyses and the data exploration. The overall processing flow is illustrated in Figure 4. When a video file or a data file was received, metadata related to the trip, the video, and the data file was saved on the E-safe database. After that, different processing was applied for both types of files. Regarding the video files, the videos were 1) reprojected in the equirectangular projection using FFmpeg (see https://ffmpeg.org/), and 2) automatically processed to evaluate the position of the hand on the handlebar (see Section 4.2.3). The data files were transformed into a format required for ingesting within Autoliv's in-house telematics cloud service to add contextual information (e.g., weather information, map-matching, filtered/processed kinematics, etc.). It is noteworthy that trips with high loss of data (>30%) or of short duration were not processed by Autoliv's telematics processing service. The resulting processed data were stored either in blob storage or on databases depending on their nature.

4.2.2 Data Exploration Tools

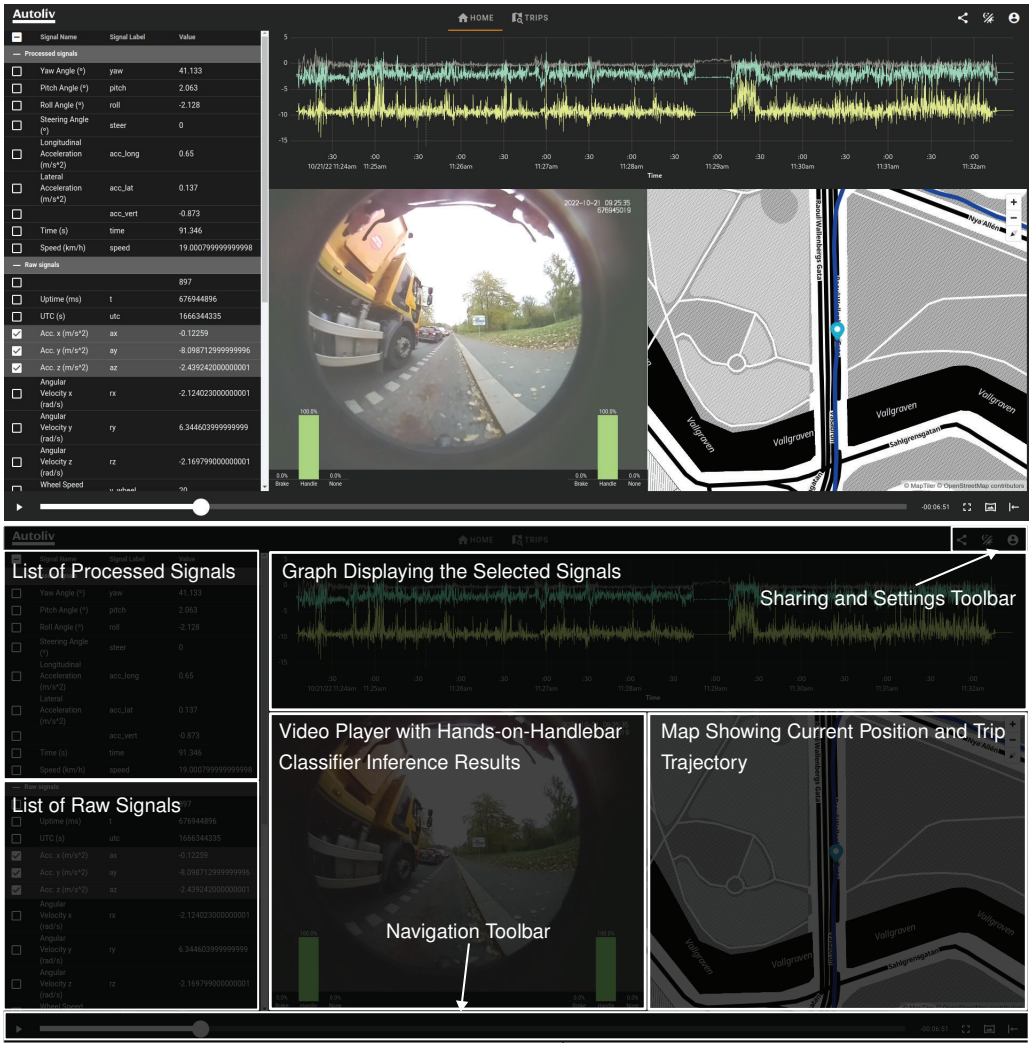


Figure 5: Description of the E-safe exploration tool interface.

A web app to facilitate the exploration of the data by the researchers was created using Nuxt.JS V.2 framework (see https://nuxtjs.org/). The framework enables developers to create web browser-based applications including both a client and a server side. In order to list, fetch, and update trip data with the web app (but also during the analyses), an application programming interface (API) was created using an Azure Function. A screen shot of the web app interface is reproduced in Figure 5. The web user can visualize the signals that are of interest by selecting them on the left panels. The selected signals are automatically rendered on the top graph. When playing the trip using the navigation toolbar at the bottom, the video (at the centre of the app) starts and a cursor synced to the video current frame shows the current timestamp on the signal graph. Similarly, the GNSS position and heading are displayed at the current timestamp, as well as the hands-on-handlebar classifier inference results (overlaying the video player).

The user interface of the web app was built using Vuetify (see https://vuetifyjs.com), the signal graph is using µPlot (see https://github.com/leeoniya/uPlot), the map is rendered

with MapLibre (see https://maplibre.org), the map tiles are generated and made available by MapTiler (see https://www.maptiler.com) using OpenStreetMap's map data (see https://www.openstreetmap.org), the video player uses VideoJS (see https://videojs.com) with its VR plugin to play "360" videos (see https://github.com/videojs/videojs-vr).

4.2.3 Hands-on-Handlebar Classifier

To facilitate the analyses for the researchers, a hands-on-handlebar classifier was created to automatically estimate the position of the riders' hands on the handlebar using the video data only. Video data were specially collected for this task by Voi before the e-scooters were available to their customers to make sure to have the classifier available as early as possible in the project timeline.

1) Training Data Preparation

The data need to be processed before labelling. First, key frames from videos were extracted. The key frames were then divided into four quadrants. The bottom quadrants were not used and were dropped while the top quadrants were saved separately as image files. The top quadrants were reshaped to 256 x 256 px and the colors were normalized from 0 to 255 to 0.0 to 1.0 (see Figure 6 illustrating the data preparation flow).

Each resulting image file was then labeled using an open-source data labeling platform called Label Studio (see https://labelstud.io). The labels used were *Absent*, *Handlebar*, and *Brake Lever*.

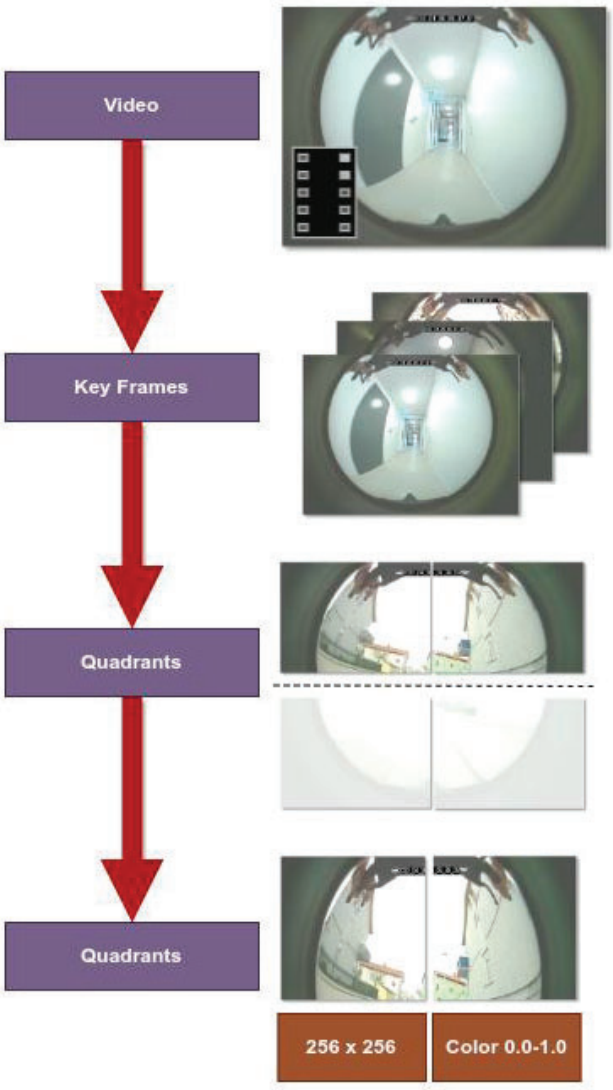


Figure 6: Training data preparation flow.

As the dataset available for this task was quite small, the data were augmented to increase generalisation (i.e., to account for variability between the e-scooters' cameras). Each quadrant of the image has a probability to be flipped horizontally and rotated from -18 to +18 degrees (see Figure 7).

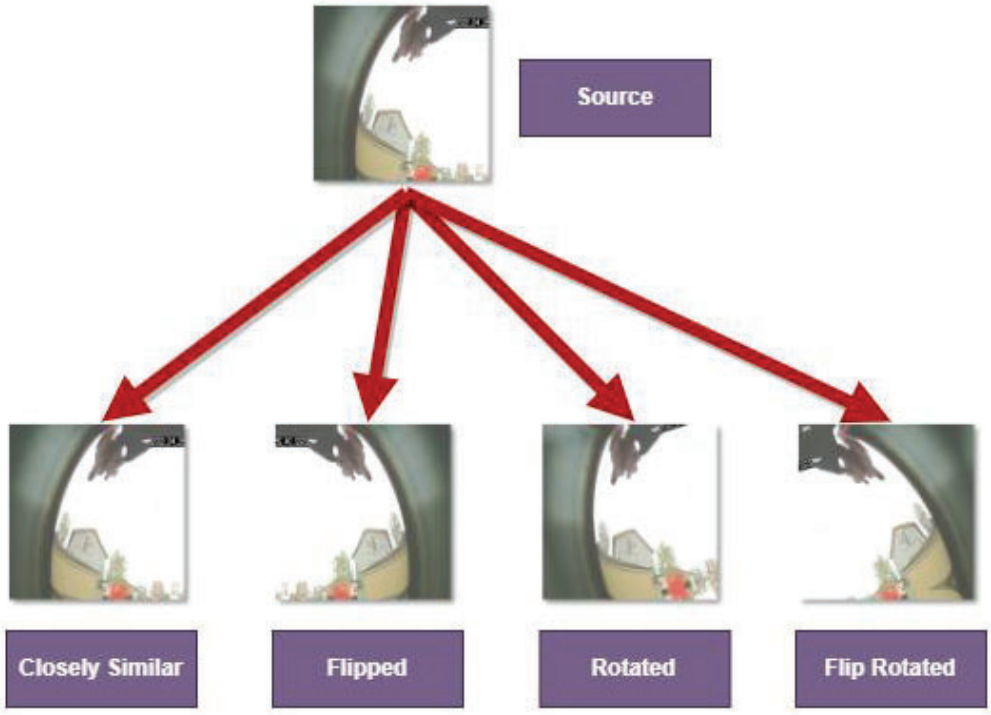


Figure 7: Training data augmentation flow.

2) Model Training

A pre-trained VGG16 model [26] was used as the base model. The last four layers of the model were unlocked and trained with an additional three layers (see Figure 8).

The data processing did not follow the same pre-processing as done on the VGG16 training data.

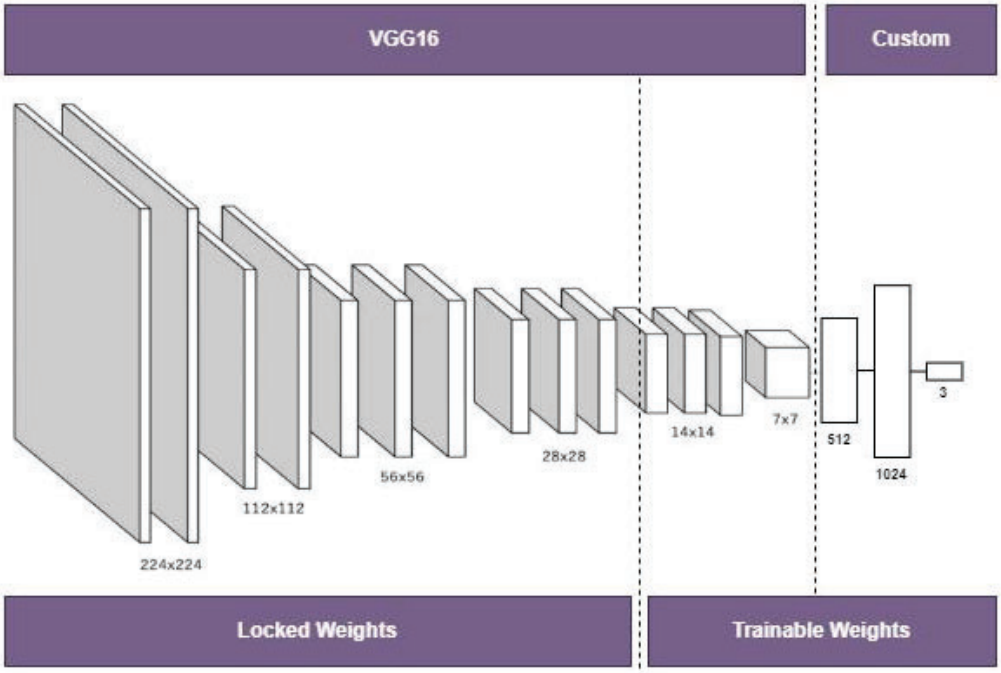


Figure 8: Illustration of the modified VGG16 neural network used for the hands-on-handlebar classifier.

3) Distance to Other Road Users

One of the most essential metrics when studying interactions between road users is the relative distance between them over time. In this pre-study, we focused on evaluating "off-the-shelf" algorithms that could be combined to estimate this metric. The major steps are divided into three parts: 1) object detection, 2) object tracking, and 3) distance estimation. For each step, we evaluated different algorithms as listed in the Figure 9.

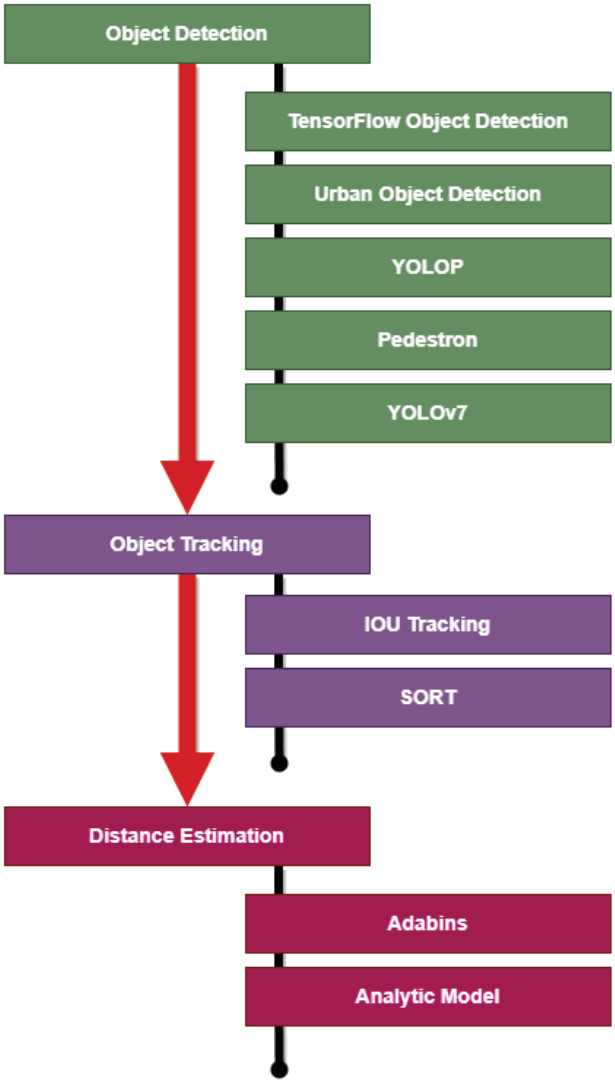


Figure 9: Distance to other road users data processing flow.

The *Analytic Model* differs from the other models as it is a mathematical model built in the project to find the corresponding real-world points given pixel coordinates of a point on a video frame. Calibration points were used to find correspondence between pixel positions and real-world positions. An e-scooter was placed vertically straight on horizontal and flat ground. For each key point on the ground, the pixel coordinates were stored together with the real-world point position expressed as a distance and an angle from the front wheel, where the forward direction corresponds to the angle set to 0 degree. Using the calibration points, a mathematical expression was devised to cover the entire field of view.

The videos in the present study were captured with a fish-eye lens camera (see video samples in Figure 10), however, most image processing algorithms evaluated here use a perspective projection. Ideally, the fish-eye videos should be reprojected into a perspective projection to use the algorithms to their maximal potential. This was not done in the study due to parameters missing to make such reprojection. As such, different algorithms were tested for each step of the methodology as their performance could be highly different

Figure 10: Some samples of the camera data.

when compared to established image benchmarks.

4.3 Safety-Critical Event Detection

To address **R.Q.1**, an algorithm to detect safety-critical events (SCEs) based on the non-visial data was devised. The data inputs chosen for the algorithm were the accelerometer and gyroscope data. The selected data were expected to offer information about the presence of high-severity kinematics (harsh braking or a fall).

Note that we do not concretely define SCE here as it is an open-ended concept without a precise definition. However, for the sake of this work and to limit scope we define the various types of SCE as follows from the perspective of motion activity.

SCE Type 1 Intuitively crashes and near-crashes (and other unsafe riding behaviour) which would result in high signal activity recorded on IMU sensors caused by abrupt motion. It may or may not involve braking.

SCE Type 2 Near-crashes as well as road user interactions which exhibit less abrupt motion but involve braking activity. It means a rider may try to actively control the situation just before the event happens using braking. For those scenarios we hypothesize that it will be combination of notable activity from braking sensors (just before the event) and notable activity on motion sensors (when the event happens).

SCE Type 3 Near-crashes as well as road user interactions which show no abrupt motion as well as non-existent braking activity. It means that the rider may not even be aware of an imminent near-crash event. Such events are near impossible to detect using only non-vision-based approaches.

In this pre-study, the focus was on identifying the SCE Type 1 only. Leaving the detection of SCE Types 2 and 3 to potential future work.

4.3.1 Approach

The accelerometer and gyroscope data were sanitised by removing outliers and normalising them by applying a min-max normalisation between 0 and 1 over the whole dataset.

The signal magnitude vector (SMV) was computed based on Euclidean vector-wise norm individually for the accelerometer and gyroscope data. For each time sample, we can express SMV as

$$SMV(i) = w_1 * \sqrt{\sum_{n=1}^{3} a(n,i)^2} + w_2 * \sqrt{\sum_{n=1}^{3} r(n,i)^2}, \quad (1)$$

where w_1 and w_2 represent the weights for the accelerometer and gyroscope signal respectively, i the discrete time index from 1 to the total number of data sample, n the axis component (n=1 $\rightarrow$ x-axis, 2 $\rightarrow$ y-axis, and 3 $\rightarrow$ z-axis), a the acceleration data components, and r the gyroscope data components.

However, we focus exclusively on leaning movement (gyroscope) and sideways movement (accelerometer). This is based on an observation that Type 1 SCEs usually involve a lot of sideways motion activity of the e-scooter. We can, therefore, simplify the SMV equation only considering the x-axis data from the accelerometer and the z-axis from the gyroscope after the normalization of their magnitude (i.e., between 0 and 1):

$$SMV(i) = w_1 * a(1,i) + w_2 * r(3,i). \quad (2)$$

The SMV signal was then time-windowed with a time-window length of 30 samples (3 s @ 10 Hz sampling rate) and stride/hop length of 1 sample (i.e., 100ms). The objective being to classify each time-window (i.e., sample) as *Normal* or *Not-Normal*. *Not-Normal* classification might indicate the occurrence of a SCE in that time-window.

A time-window was considered *Normal* if the standard deviation of the SMV in that window was lower than a specific threshold (Thrsh) and *Not-Normal* otherwise. After this step, a simple post-processing logic was applied. This logic made sure that the current time-window was detected and classified as *Not-Normal* for a pre-defined number of times consecutively before triggering an SCE signal to be registered. A high-level logic of the analytical SCE detection algorithm is shown in the Figure 11.

Pre-processed kinematics data were fed to the algorithm which executes an automated analysis and provides an output log of the suspected trips which contain SCEs and approximate timestamps of those events. This output log can then be manually analyzed to verify the events and to remove the false positives.

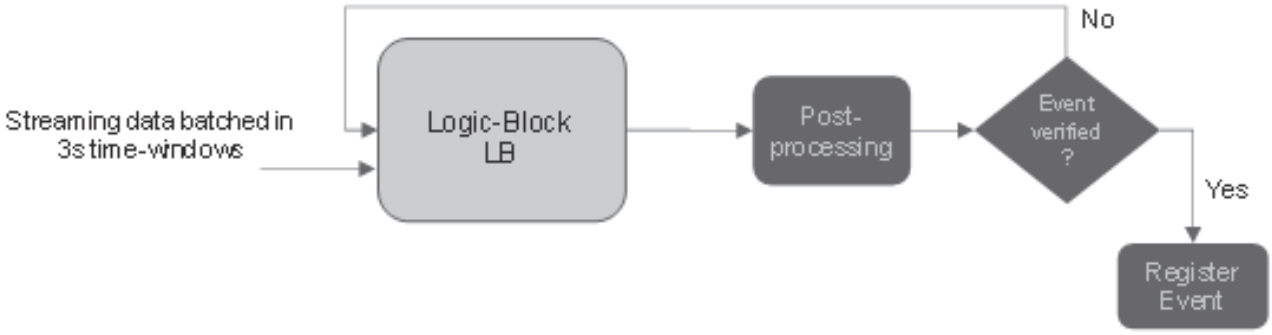


Figure 11: SCE detection algorithm logic.

4.4 Infrastructure and Rider Behaviours

There is no consensus on how to define safe or an unsafe behaviour. In this project, we consider an unsafe behaviour as a behaviour that is extraordinary or unsuitable for the current environment (e.g., infrastructure or other road users), but also a behaviour that does not follow common sense when riding an e-scooter (e.g., not holding the handlebar with 2 hands, riding while under the influence).

The data collected with the e-scooters enable researchers to not only study rider behaviour, but also to have data that can indicate how the traffic infrastructure is used and in which condition it is. We looked at two aspects of the infrastructure: 1) the road surface roughness, and 2) e-scooter speed regulation in relation to the infrastructure.

4.4.1 Road Surface Roughness

Can the data gathered by Voi e-scooters be used to infer information about road quality? This would be especially important for separate bike lanes, where traditional road measurements by car would not be feasible. One of the most common measures of road roughness is the International Roughness Index (IRI). It is measured in millimeters per meter and describes the longitudinal irregularities of the road. The measure was developed by the World Bank [22] in the mid-eighties and describes the road's ride comfort primarily when driving a car. Table 1 describes general driving experience at different IRI values.

Table 1: Ride experience for different IRI values.

IRI (mm/m)	Ride Experience
<1.5	Irregularities barely noticeable. Comfortable ride.
1.5-3.0	Moderate irregularities. At higher speeds, however, these can be clearly noticeable.
>3.0	The journey can be perceived as unsafe. Sharp heeling is common. Surface damages occur.
>6.0	The speed must be reduced to about 50 km/h for a pleasant ride. Severe surface damage may occur (cracks, potholes etc).

The IRI describes how much total vertical movement a standard passenger vehicle's body would experience if driven over a 1-kilometre segment at 80 km/h. The IRI is based on the concept of a "golden car" whose suspension properties are known. The IRI is calculated by simulating the response of this "golden car" to the road profile. The properties of the "golden car" were selected to provide high correlation with the ride response of a wide range of automobiles that might be instrumented to measure a slope statistic (mm/m) [11]. The damping in the IRI is higher than most vehicles, to prevent the model from tuning in to specific wavelengths and producing a sensitivity not shared by the vehicle population at large. IRI is primarily focused on wavelengths between 1.3 to 60 meters giving higher weight to wavelengths between 2.5 and 15 meters [23].

IRI was mainly developed to match the responses of passenger cars, but subsequent research has shown good correlation with light trucks and heavy trucks. The IRI has become a general-purpose roughness index. However, on many roads with a low-speed limit, the IRI can be high, even though the road is smooth to drive on at the maximum legal speed. This happens if the road has long wave irregularities that cause the car chassis to sway at 80 km/h, but not at low speed. In Sweden, the maintenance standard of paved roads has different IRI criteria for different speed limits, allowing for higher IRI at lower speeds [32].

The instrumented Voi e-scooters are not "golden e-scooters" and cannot be used to measure precisely an IRI, but they are all the same model (with the same suspension and the same instrumentation) which reduces the inter-vehicle variability when collecting data. We investigated two ways to classify the road surface quality: 1) segment trips and calculate the rolling energy of the longitudinal acceleration for each segment, and 2) aggregate all vertical accelerations by geo-location and calculate the standard deviation for each distribution.

Since we did not have access to precise measurements of IRI, we cannot estimate the global performance of the methods. However, we used the video camera to visually confirm the road surface roughness estimation.

a) Longitudinal Acceleration Energy over Trip Segments

A road surface quality estimate was developed by computing a windowed high-pass spectral signal energy of the a_z variable which measures longitudinal acceleration (in the backward direction). a_z was selected because the visualizations provided more consistent results on the same street than with any of the other accelerometer directions, while also correlating to visually observed cracks in the road. Running into a bump, even a small one, directly impacts the forward speed allowing the e-scooter to accelerate back up again, until it runs into the next bump. These variations in forward speed act as a proxy of road surface quality.

A simple high-pass filter H was applied by removing a windowed mean $H(x) = x - \bar{x}$. That is, for a road segment centred on a data point i, the surface q was estimated from a window of a_z of length w centred on the data point i. A rolling rectangular window of 100 samples ($w = 100$) was used. With a sample rate $f_s = 10\ Hz$ and a speed of 5 m/s this corresponds to a 50-meter long window. This captures wavelengths from 1 to 50 meters, which is equivalent to wavelengths of interest when calculating IRI. Some trips did not have

any single 50-meter long segment at the required speed. Each element associated to a window W would then be expressed as

$$W(i) = a_z(j), \text{ with } i - w/2 < j < i + w/2. \tag{3}$$

The rolling energy estimate would then be expressed as

$$E(i) = \sqrt{\sum H(W(i))^2 / f_s}. \tag{4}$$

Since we are interested in the relative values when comparing the different energy estimates, we do not need the absolute values of the energy. We can, therefore, simplify the calculation by using the standard deviation of the longitudinal acceleration over the segment instead since

$$q(i) = E(i) \times f_s/(w-1) = std(W(i)). \tag{5}$$

This surface quality estimates were presented in an interactive map with images from corresponding recordings for each road segment. Colours were assigned according to the distribution of quantiles across all segments. To improve visual quality, GNSS coordinates (sampled at $1\ Hz$) were interpolated, and segments were coloured with an average of 50 rolling window samples ($25\ m$).

b) Standard Deviation of Aggregated Vertical Acceleration Distribution

Since the trips location data are sampled at $1\ Hz$, we cannot have an accurate calculation of the acceleration energy for a given location as explained earlier. One proposition to address this issue was to aggregate all the vertical acceleration grouped by location data. For each location point, a distribution of vertical acceleration is thus created and the standard deviation (correlated to the energy as explained above) can give an indication of the vibrations experienced at this location.

The aggregation of the data (i.e., the vertical acceleration a_y) is made by associating them to the H3 index that their location data belongs to. The H3 index is part of Uber's Hexagonal Hierarchical Spatial Index (https://github.com/uber/h3). This spatial indexing library is an alternative to using triangle- or rectangle-based grid systems, it uses hexagons as the cell form which results in having an equal distance between a given cell and all its neighbouring cells (see Figure 12).

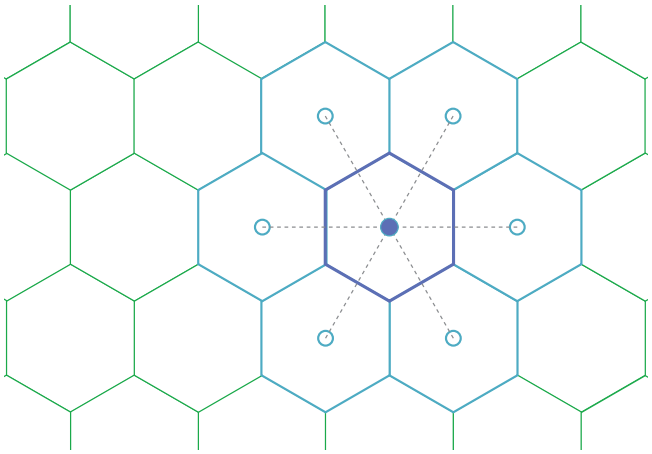


Figure 12: H3 hexagonal grid. The distances shown as gray dashed lines between a given cell and its neighbouring cells are identical.

After aggregating the vertical acceleration data points by indexing them using the H3 index, we calculated the average, and the standard deviation of each vertical acceleration distribution by H3 index.

The standard deviation was then plotted on a map to identify where high vertical acceleration standard deviations (i.e., high vibrations) occur. To simply the visualization, we set a standard deviation threshold to $1.5\ m/s^2$. Above this threshold, we consider the road surface to be rough.

4.4.2 Normal Rider Behaviour

1) Speed Regulation

The investigation of speed regulating behaviour used a similar approach to the one described above for the vertical acceleration aggregation. For each H3 index (or "H3-node"), statistics such as average, standard deviation, minimum, maximum, and count were calculated for the wheel speed variable. The speed statistics were plotted on a map (using tile images and map style created by https://www.maptiler.com/ generated with OpenStreetMaps' open map data https://www.openstreetmap.org) to investigate visually how riders usually regulate their speed given their location in the road infrastructure. Additionally, outliers were detected in the speed distribution of each H3-node; where all data points in each distribution that had a standard score (or zscore) higher than 3 were defined as outliers (which means that only high-speed outliers were considered interesting for this analysis). Each H3-node that contained at least one outlier was plotted with another colour on the map to locate easily where speed outliers happened.

2) Hand Placement Behaviour

For this analysis, the data generated by the hands-on-handlebar classifier presented earlier were used together with the other collected data. Only the data collected between 6:00 and 19:00 were kept as the hands-on-handlebar classifier was not reliable during darkness. Three analyses were done to characterise hand placement behaviour: 1) general statistics

of the hands' placement compared to the brake state (on/off), 2) hand placement distribution given the travelled speed, and 3) brake hovering behaviour in Gothenburg city.

1. For each brake state (on/off) and hand (left/right) the ratios of the hand placement (absent, handlebar, brake lever) were calculated and illustrated by pie charts.
2. For each hand (left/right), the ratios of hand placement were calculated for each 1-km/h travelled speed bin and plotted as a stacked histogram to visualize the variation of the ratios over the travelled speed.
3. As brake hovering behaviour may indicate anticipatory behaviours, we plotted the percentage of brake hovering for each H3-node and plotted them on a map for each hand (left/right). Brake hovering means that the hand is detected to be on the brake lever without activating it.

4.5 Road-User Interaction Modelling

We investigated the e-scooter riders' comfort boundaries when interacting with pedestrians as an illustration of road-user interaction modelling. A comfort boundary was defined as the minimum distance to pedestrians.

To estimate this comfort boundary, it was necessary to 1) detect and track pedestrians using the video data, 2) estimate the distance from the e-scooter to the pedestrian knowing the position of the pedestrian in the video frame, and 3) calculate the 5^{th} percentile of the distance distributions for each angle bin. The relative position of pedestrians was then expressed as an angle along the vertical axis and a distance which can be used to create distance distributions as a function of an angle.

The first step was carried out using off-the-shelf detection and tracking models available freely to the community. A comparison between the models found was made before selecting the best candidate.

5 Results

5.1 Data Collection Summary

5.1.1 General Statistics

A kinematics data snapshot was captured on Dec 05, 2022. There is a total of 2315 trips represented in the data. It equates to about 128 hours of data with a median of around 4 minutes per session length. A percentile distribution of the trip lengths can be seen on Table 2. Out of those 2315 trips, only 768 had no data loss.

Table 2: Distribution of trip duration.

Percentile	Trip duration (s)
0	1.0
1	1.9
5	47.4
10	80.1
25	142.4
50 (Median)	255.2
75	433.2
90	644.9
95	816.8
99	1283.8
100	2625.6

The summary statistics for the IMU data on the entire dataset were reported on Table

Table 3: Summary statistics for motion sensor data (ax, ay, and az in g, and rx, ry, and rz in rad/s).

	ax	ay	az	rx	ry	rz
count	7.582742e+06	7.582742e+06	7.582742e+06	7.582742e+06	7.582742e+06	7.582742e+06
mean	-1.161391e-02	-9.418160e-01	-2.316013e-01	-1.142559e-02	2.521844e-04	4.379022e-03
std	9.116612e-02	1.018044e-01	9.380385e-02	9.340948e-02	3.852527e-01	1.479942e-01
min	-1.378756e+00	-2.240120e+00	-1.902626e+00	-3.449220e+00	-5.474234e+00	-3.120213e+00
25%	-4.698571e-02	-9.840761e-01	-2.874671e-01	-3.195792e-02	-1.270328e-01	-6.087986e-02
50%	-8.847959e-03	-9.488673e-01	-2.390780e-01	6.391570e-04	1.917471e-03	4.154529e-03
75%	2.733714e-02	-9.162214e-01	-1.803153e-01	2.109223e-02	1.302286e-01	6.966829e-02
max	1.671349e+00	1.639130e+00	1.375461e+00	2.446859e+00	5.235349e+00	3.508981e+00

We can observe that accelerometer activity is constrained in the relatively narrow range of $-2.24\ g$ to $1.70\ g$ while the gyroscope activity is constrained within $-5.5\ rad/s$ to $+5.5\ rad/s$.

The summary statistics of the brake levers and gas throttle are reported in Table 4. One can notice that the standard deviation (std) value is higher on the left brake ($brake_l$) compared with the right brake ($brake_r$). Also, the mean is slightly higher on the $brake_l$ than $brake_r$ even though the difference in standard deviation is more noticeable. It seems riders are using the left brake (corresponding to the rear-wheel brake) more than the right brake on average as it exhibits more variation over time.

Table 4: Summary statistics for brakes and throttle signals.

	brake_l	brake_r	throttle
count	7.582742e+06	7.582742e+06	7.582742e+06
mean	5.221852e+01	5.103576e+01	1.428998e+02
std	1.697959e+01	1.234999e+01	8.371460e+01
min	-6.200000e+01	0.000000e+00	-6.530000e+02
25%	4.800000e+01	4.800000e+01	4.900000e+01
50%	4.900000e+01	4.900000e+01	1.490000e+02
75%	4.900000e+01	4.900000e+01	2.280000e+02
max	2.560000e+02	2.120000e+02	5.900000e+02

Columns showing the Pearson correlations between different motion related features such as accelerometer, gyroscope, e-scooter throttle and brake control are reported in Table 5. Note that Pearson correlation only tracks the linear correlation between the features. Various features might still be correlated in a non-linear way.

- There is some correlation between the ax and rz representing sideways movement
- ay shows some correlation with both az and rx
- rx also shows correlation with the az
- ry is not correlated to any of the other features here
- $throttle$ intuitively shows inverse correlation with both brake features
- Both left and right brake features also intuitively show positive correlation with each other

Table 5: Linear correlation of features related to motion, speed and brakes.

	ax	ay	az	rx	ry	rz	throttle	brake_l	brake_r
ax									
ay	-0.02								
az	-0.02	-0.20							
rx	0.03	-0.16	0.16						
ry	-0.02	0.00	-0.02	0.07					
rz	0.11	0.00	-0.01	0.04	0.06				
throttle	0.05	0.04	0.06	-0.02	0.00	-0.01			
brake_l	0.01	-0.06	0.13	-0.00	0.00	0.00	-0.23		
brake_r	-0.00	-0.05	0.11	0.01	0.01	0.00	-0.20	0.25	

5.1.2 Observations on Sensor Data

Acceleration Data

We found that acceleration values along all three axes are confined within the $+2.0\ g$ to $-2.0\ g$ (Figure 13). Peaks and troughs are not very high in terms of intensity for any axis, particularly compared with the noise on the signals. a_y on average stays close to $-1.0\ g$ which is expected because it is pointing downwards to the ground. Acceleration values along the heading direction (a_z) and sideways movement (a_x) have about the same intensity overall.

This information can be used to put in place an automated check to flag any trips which do not conform to this magnitude range to suggest manual handling of that trip for manual diagnostics. This may be caused from a crash event or a data anomaly due to sensor/data-collection issues.

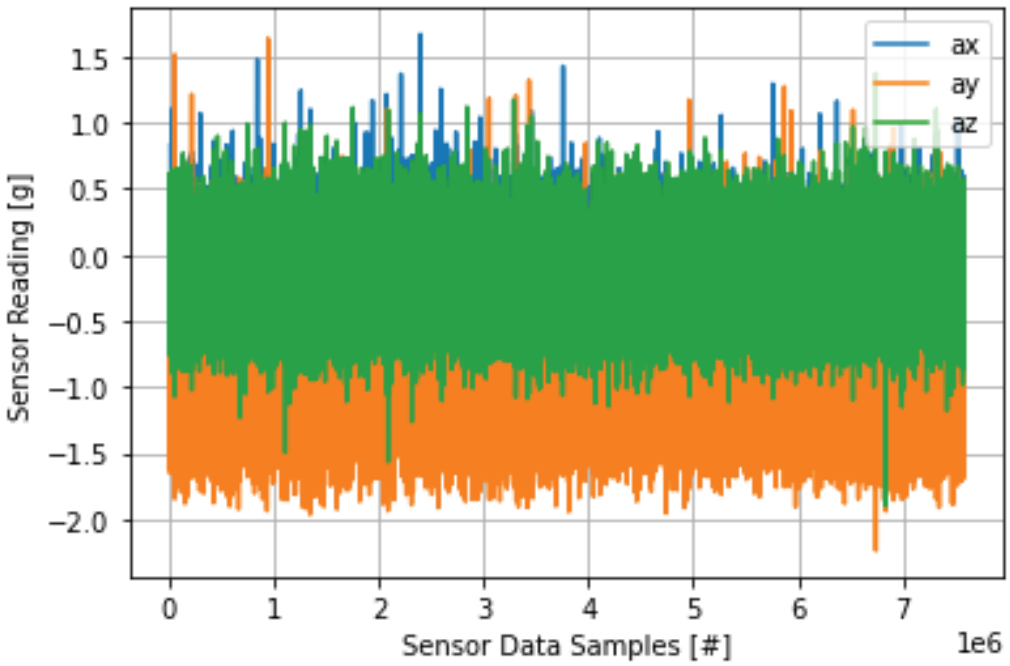


Figure 13: Acceleration plot for all trips in data.

Angular speed data

In general, measured angular speed values are confined within the band of -5.0 to $+5.0\ rad/s$ (Figure 14). r_y has significantly more variance and more pronounced peaks and troughs. It is expected because it represents the motion of handlebar to left/right to navigate/steer along the path (i.e., yaw). r_z shows the noticeable rotational activity representing the tilting of e-scooter sideways (i.e., roll or lean). r_x shows the least rotational activity which is expected (i.e., pitch). In other words, $r_y > r_z > r_x$. This could also be implemented as an automated data validation check.

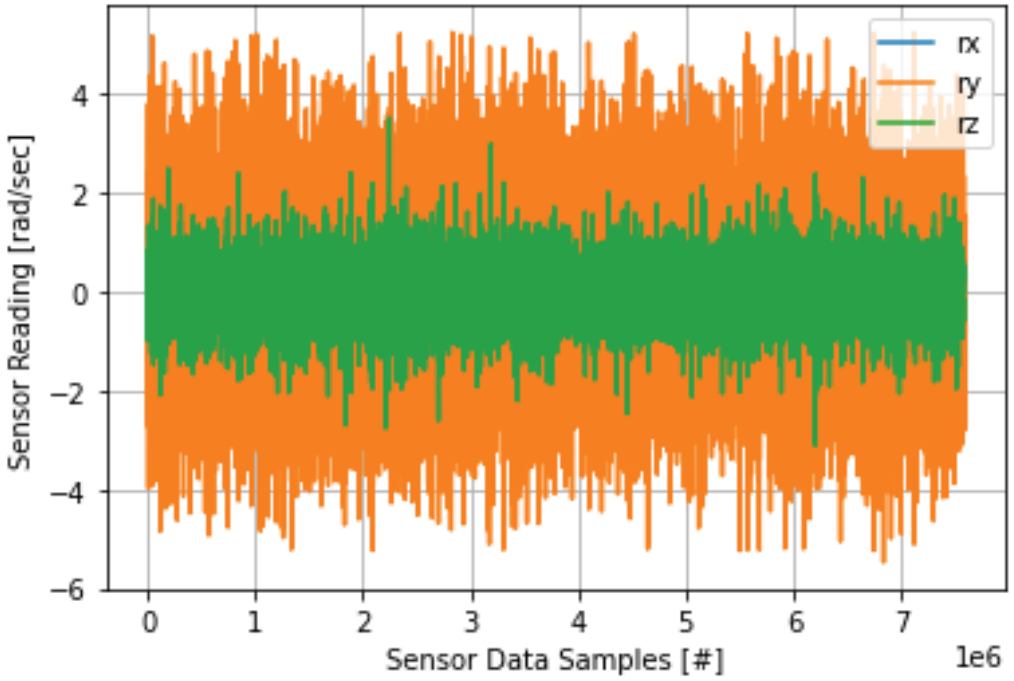


Figure 14: Angular speed plot for all trips in data.

Video Data Limitations

The roadside camera in general performed well however, under some circumstances, its visibility can be hindered. The observed issues were:

- Blurriness in some trips (lens not correctly tuned)
- Sun shining into the cameras - sometimes part of the screen where the hand is placed gets overblown
- In nighttime, the view was very dim sometimes due to limited lighting and visibility
- Other objects and clothes hanging on e-scooter obscuring the camera view
- Rain droplets when it was raining

Figure 15: Examples of visual issues.

5.2 Data Processing and Exploration

5.2.1 Hands-on-Handlebar Classifier

After data preparation, the number of quadrants used for each label and dataset are reported in Table 6.

Table 6: Occurences of labelled data.

Dataset	Label	Number of Quadrants
Train	Brake Lever	302
Train	Handlebar	398
Train	Absent	126
Validation	Brake Lever	19
Validation	Handlebar	169
Validation	Absent	8
Test	Brake Lever	224
Test	Handlebar	381
Test	Absent	23

The model was able to classify the image on the Test data with good performance since the lowest F1-score for all classes is 0.88. Although there is probably a bias from the number of data samples in the training set as the performance metrics for each label are related with the number of samples in training (see Table 7).

Table 7: Performance indicators of the classifier.

Label	Precision	Recall	Quadrants
Brake Lever	0.98	0.84	0.91
Handlebar	0.91	0.99	0.95
Absent	0.85	0.83	0.88

The model was not retrained with the data collected later which resulted in good performance but not as robust (i.e., high confidence) as one could expect. For instance, when trees or high buildings are cluttering the view behind the hands the classification results will be unstable jumping from one label to another. Additionally, the classifier will perform poorly when the video is dark (e.g., nighttime or e-scooter passing under a bridge) which may be addressed if the camera were to be equipped with night vision.

Despite those limitations and a somewhat low robustness, the classifier showed that the methodology could be used for automatically labelling the hand placement for this pre-study but would benefit from being retrained in the future to improve its robustness.

5.3 Safety-Critical Event Detection

Based on our execution of the algorithm with 2315 trips of data we generated a log file with 111 trips sessions with a suspected SCE. Manual inspection of the log file allowed us to confirm the correct detection of several crashes as well as other types of SCEs. The algorithm also provides a list of timestamps for each of the suspected SCEs on a trip basis, so that the data and videos can be found corresponding to that event. Note that algorithm sensitivity can be increased or decreased by tuning the "Thrsh" value. This allows to achieve a suitable trade-off between the false positive (FP) and false negative (FN) performance of the algorithm.

Regarding the results, here is a brief description of some SCEs we found after analysing the output log from the algorithm and categorising the events:

- **Playful/Unstable Steering**: Usually for fun by the rider during normal e-scooter drive
  - 9 playful events detected
- **Crash/Fall**: Unintentional collision/e-scooter falling
  - 8 crashes/falls detected (single rider)
  - 1 known crash **not** detected due to the low intensity (i.e., FN)
- **Twin Riding**: 2 people riding on the same e-scooter
  - 3 crashes detected
  - 2 events where 2 people were riding the bike (without a crash)
- **Playful Crash**: Rider bumping others with the e-scooter for fun
  - 1 crash detected

- **Abuse/Misuse**: Rider bumping the e-scooter into different objects
  - 2 abuse or misuse events detected
- **Intentional Harsh Braking**: Rider going to full deceleration deliberately (and without a collision threat)
  - 1 intentional harsh braking events
- **Rapid Shaking**: Quick shaking of the e-scooter without anyone on it
  - 1 rapid shaking event

Out of the 8 crashes/falls that were detected, a series of images from two of them are reproduced in Figure 16. Both crashes happened at night.

Crash A happened at a very slow speed on the side of the bike lane. The rider was riding slowly but zigzagging, until riding out of the bike lane. Visual analysis of the video suggests that the person had a hard time keeping balance in the grass path and fell. It is noteworthy that the rider had a bag hung on the left handle of the handlebar which may have contributed to instability and the fall.

Crash B happened at a higher speed than Crash A, and this was mainly due to a loss of balance by the rider. Before the crash happened, the rider was riding only with one hand as their left hand was occupied holding their phone (they were interacting with their phone screen some time before the fall happened). When turning to the right, the rider lost balance and had to jump onto the road to avoid falling, the e-scooter is the only object that fell.

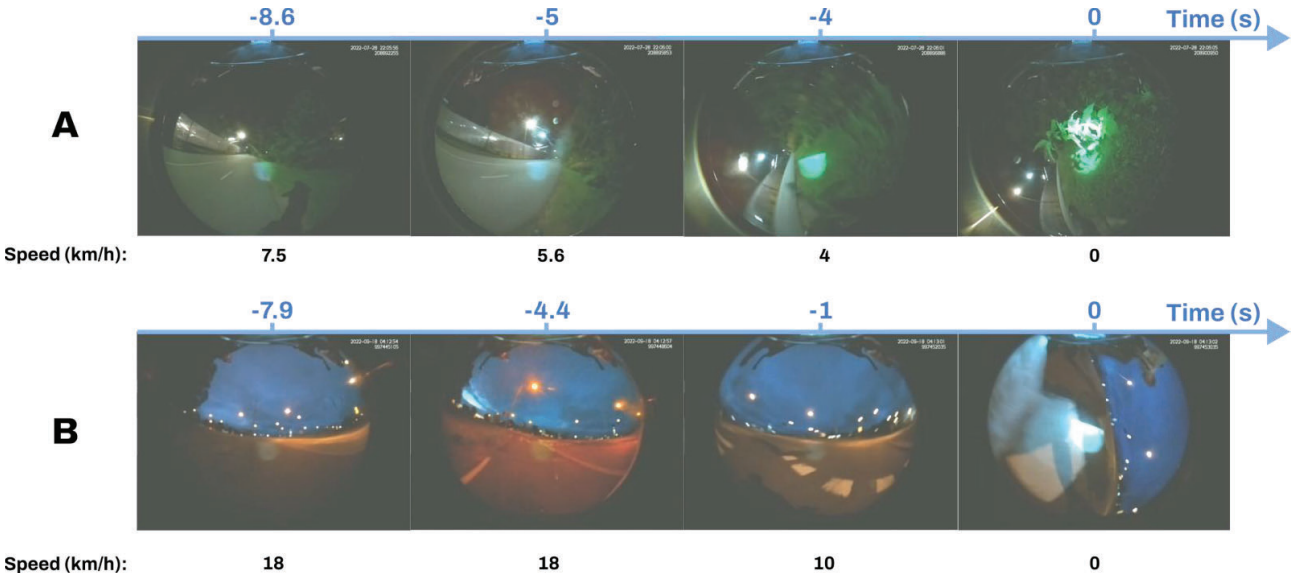


Figure 16: Two examples of fall events.

All the crashes detected in the collected data happened at low speed and are single-vehicle crashes (mostly falls due to loss of balance as exemplified above).

The kinematics data, and the video data are a great source of information for crash reconstruction, initial conditions and contextual information can be given by those data.

Those initial conditions could be used in crash tests and simulations. However, as all the observed crashes in the database are non-severe crashes/falls (people usually could ride right away after their fall) the current dataset cannot provide initial conditions for crash tests or simulations that would aim at evaluating injury prediction. Additionally, we do not see the rider falling as the video camera is oriented forwards. This limits the possibility to reconstruct body movements during a crash or a fall.

5.4 Infrastructure and Rider Behaviours

5.4.1 Road Surface Roughness

The results of the first method to determine road surface using the longitudinal acceleration are illustrated on the map reproduced in the Figure 17. The red stretches of road are the ones detected as having a rougher surface compared to the yellower ones.

Figure 17: Overview of ride comfort in Gothenburg. Yellow represents smoother surfaces; red represents rougher surfaces.

Restricting sampled data to stretches of $3—5m/s$ proved to be an efficient way of removing a whole range of irrelevant data such as crossing over curbs and other obstacles that could otherwise skew the results when identifying surface quality.

Figure 18: Examples of two different segments identified as rough (dark red) vs two segments identified as smooth (yellow).

The surface type was clearly visible in many of the recordings, and it was possible to discern cobblestone, sidewalks, gravel and repair patches (examples in Figure 18). However, the video quality was typically not high enough to clearly discern pavement quality unless there were obvious cracks present. The videos could be used to qualitatively check for an apparent correlation with surface type as expected from a road quality estimate. Cracks seemed to be found more often in segments with poorer estimated road quality; therefore, a full categorization of visually identified cracks into poor pavement or good pavement sections would perhaps be possible. However, the current algorithm provides an estimate of ride comfort but cannot conclusively and automatically distinguish whether poor ride comfort is due to lacking surface quality, obstacles or the surface material.

The second road surface assessment method gave a more gradual map where vibrations can be clearly pin pointed, it is also less dependent on the riding's braking behaviour that could affect the first method. The calculated standard deviation of the vertical acceleration distribution for each H3-node was represented by circles on the map of Gothenburg (see Figure 19). The areas where a lot of pink circles are displayed represent places where the surface roughness is estimated to be high. For instance, the old city centre, *Haga*, has many streets made of cobblestones, which can be seen on the map. Many other areas, where cobblestones are placed, or tramway rails are present show large vibrations as well.

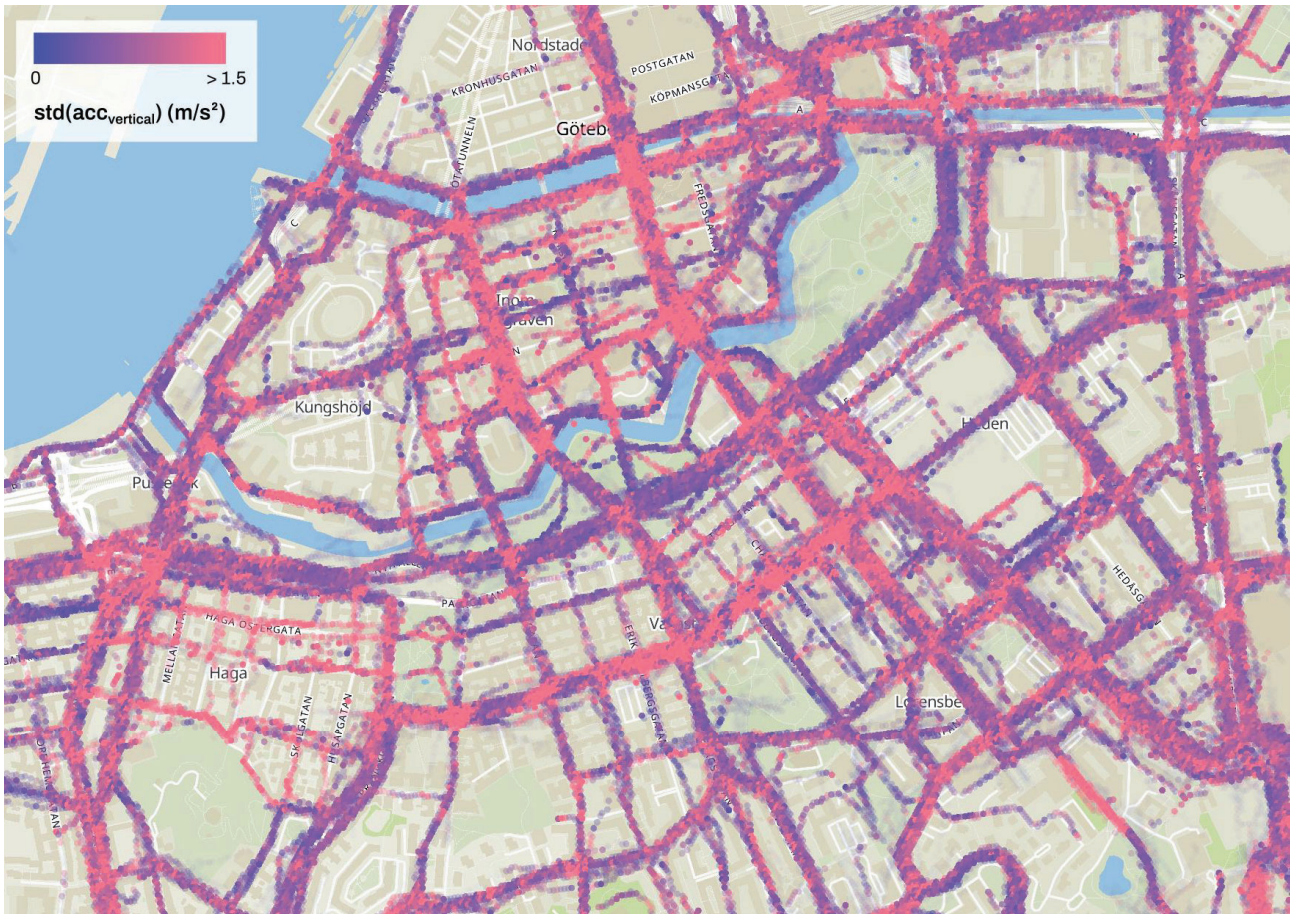


Figure 19: Overview of the estimated road roughness in Gothenburg. Yellow represents smoother surfaces; red represents rougher surfaces.

When zooming on the map, the viewer can see more precisely where the high vibrations were observed. For instance, Figure 20 shows the roads in front of *Trädgårdsföreningen*'s entrance, where there are cobblestones laid in front of the entrance. Similarly, cobblestones were placed before pedestrian crossings (at the top of the view), to help road users slow down before potential road conflicts.

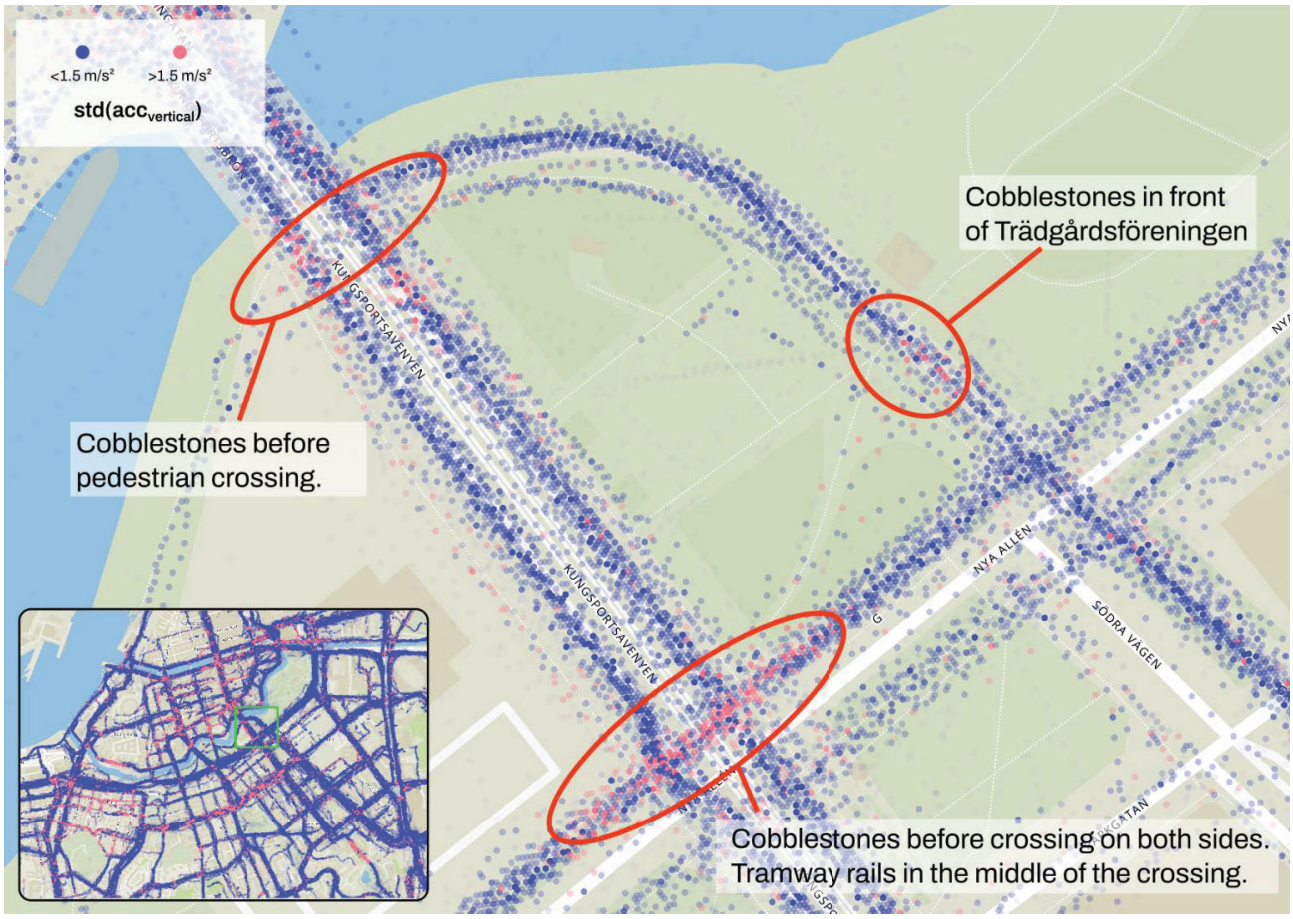


Figure 20: Estimated road surface roughness by *Trädgårdsföreningen*'s entrance. Blue circles show low surface roughness; pink high surface roughness.

It becomes interesting to look at the riders' speed behaviour as a function of the vertical vibration (i.e., rough road surface). This is illustrated in Figure 21 with the plot of the average speed against the vertical acceleration standard deviation for each H3-node. This is represented in three different forms: A) a scatter plot to indicate the spread of the data, B) a heatmap to represent the overall bivariate distribution, and C) another heatmap to represent the same distribution but normalized over the vertical acceleration standard deviation.

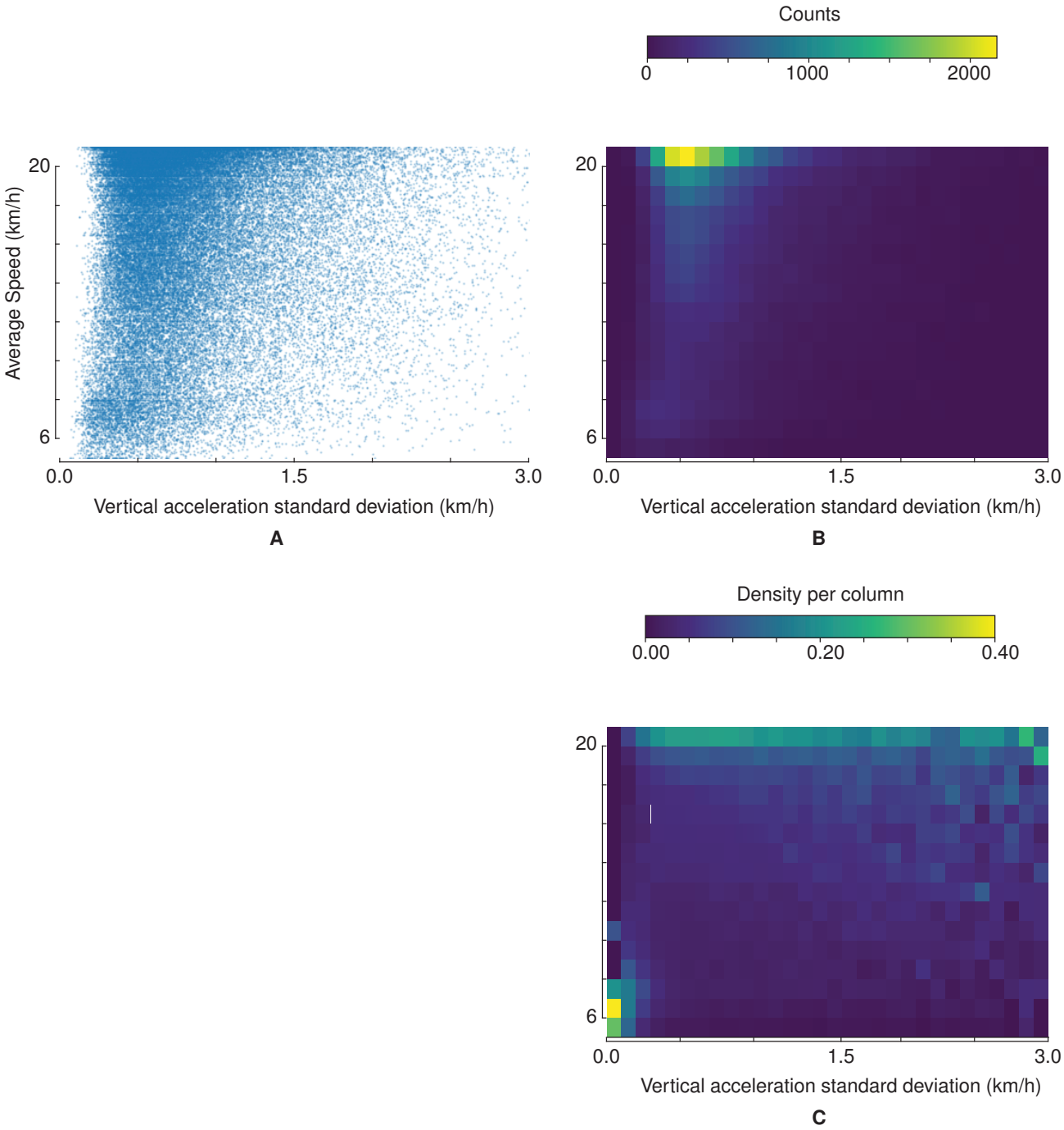


Figure 21: A) Average speed for each H3-index versus vertical acceleration standard deviation, the e-scooter's maximum speed is shown by a horizontal black line, B) Count 2-D histogram of the scatter plot A), and C) 2-D histogram of scatter plot A) after normalizing by column (i.e., by vertical acceleration standard deviation bin).

A visual observation of Figure 21.C suggests that the most frequent average speed is constant (between 20 and 21 km/h) regardless of the vertical acceleration standard deviation value. A simple linear regression was applied between the observed speed average and the vertical acceleration standard deviation, and the results suggest that the vertical acceleration standard deviation does not affect the speed average ($p - value > 0.05$ with its 95th

percentile confidence interval containing 0).

5.4.2 Normal Rider Behaviour

Speed Regulating Behaviour

The average speed for each H3-node was plotted on a map. Only excerpts are reproduced in the figures below. Each circle on the map represents a speed distribution, the radius being the standard deviation and the colour corresponding to the average. Figure 22 shows that the riders typically ride as fast as they can with the e-scooter (circles in pink) except at intersections (where riders may have to give the way or may slow down before crossing), and in low-speed zones set by the city of Gothenburg (e.g., the blue stretches in the centre of the map, in front of the central station, or in front of *Trädgårdsföreningen*).

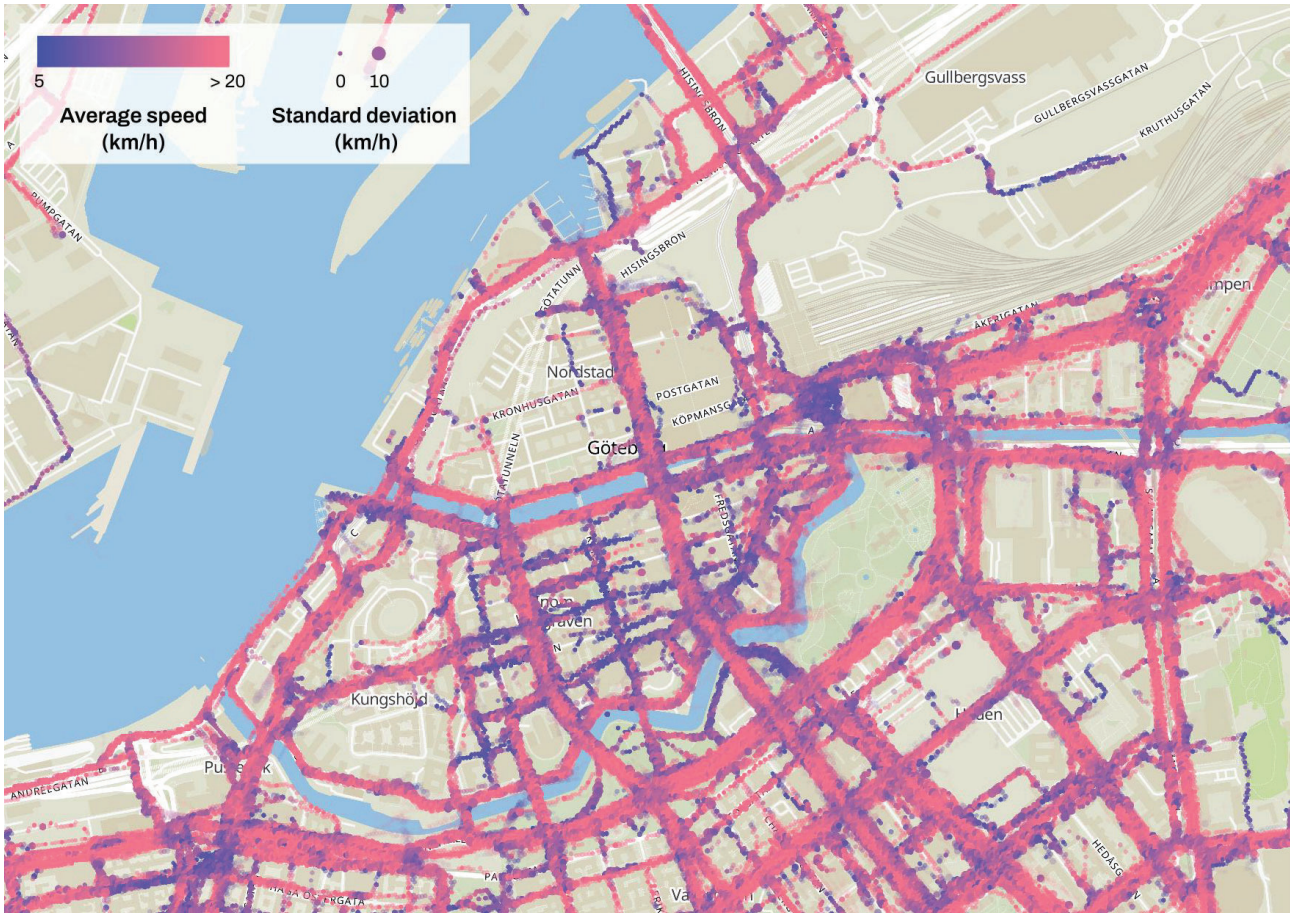


Figure 22: Observed average (blue: low value, pink: high value) and standard deviation (low: small circle radius, high: large radius) for each speed distribution for each H3-node.

The outliers seem to be more frequent in the low-speed zones (see Figure 23). But it also happens scarcely in other places, and that could indicate an inappropriate speed choice given the road and the usual (i.e., normal) rider speed regulating behaviour.

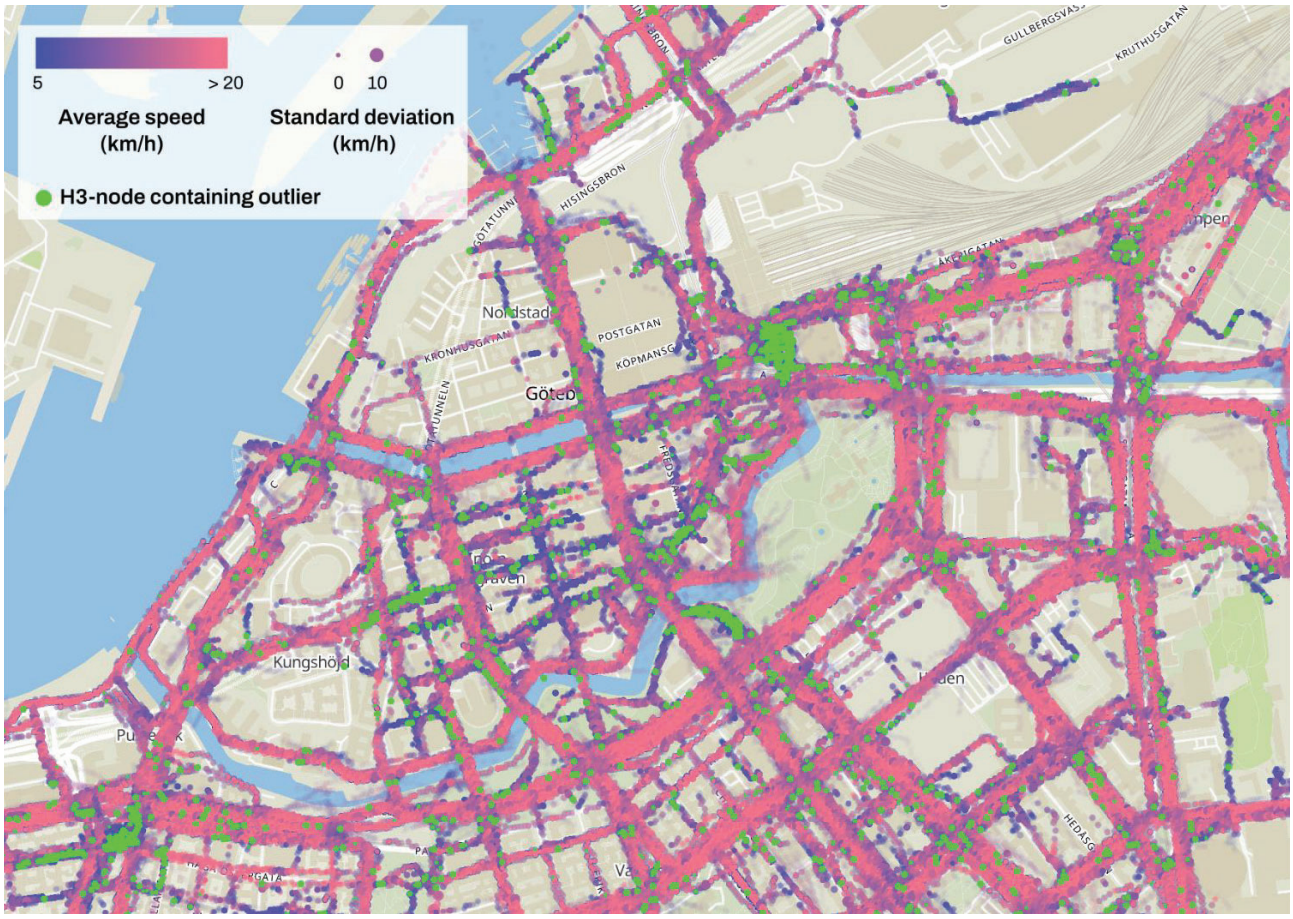


Figure 23: Observed average (blue: low value, pink: high value), standard deviation (low: small circle radius, high: large radius), and outliers (green dots) for each speed distribution for each H3-node.

Hand Placement Behaviour

Hand placement observations resulted in a total number of data points equal to 1,234,528. Figure 24 illustrates the summarized statistics of the hand placement depending on if the left or right brake activated or not. During the retained trip data, the left brake was used 5.67% of the time, and the right brake, 4.45%. The video-based hands-on-handlebar classifier detected that 0.91% of the time no hands were on the handlebar, 1.06% only the left hand, 2.37% only the right hand, and the rest of the time (95.66%) both hands. As e-scooters are not self-stabilizing the value of 0.91% of "no hands" might be mainly a misclassification from the classifier since the speed is higher than $5\ km/h$ for this analysis, but in some few cases (e.g., falls) this could be true.

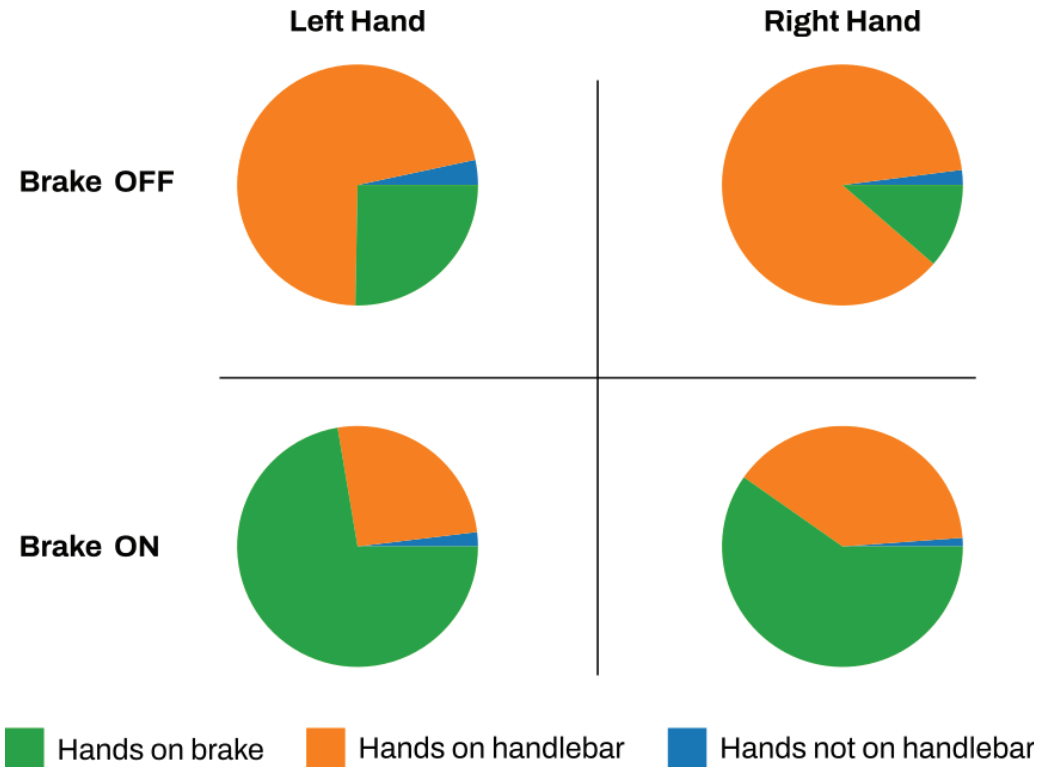


Figure 24: General statistics of hand placement given the left or right hand, and if the corresponding brakes are activated or not.

For each of the brakes, we looked at the proportions of hands placement given the brakes being activated or not (ON/OFF on the figure above). According to the numbers, when the brakes are not activated, the hands are usually placed on the handlebar, and only 27.96% of the time the hands are hovering the left brake lever and around 13.50% for the right hand. When the brakes are activated, the values show that most of the hand placements are on the brake levers, which indicates that the video-based classifier does a good job in estimating that the hands are on the brake lever. But it is not perfect as we can see that the rest of the time, the classifier estimated that the hands were on the handlebar although the brake levers were activated. This error could also be caused by noise in the brake levers' sensor where the brake levers are assumed to be activated when they are not.

Finally, the distributions of the hand placements for the left and right hands given the travelled speed are shown in Figure 25. As mentioned earlier, we can see that the left hand was usually on the brake lever more often compared to the right hand. The histograms show a slight decrease of hands placed on the brake levers when the usual cruising speed (around $20\ km/h$) is reached. This may be explained by the fact that at the cruising speed, riders do not expect to have to brake for an evasive manoeuvre (e.g., a long, straight, and free-of-other-road-user path).

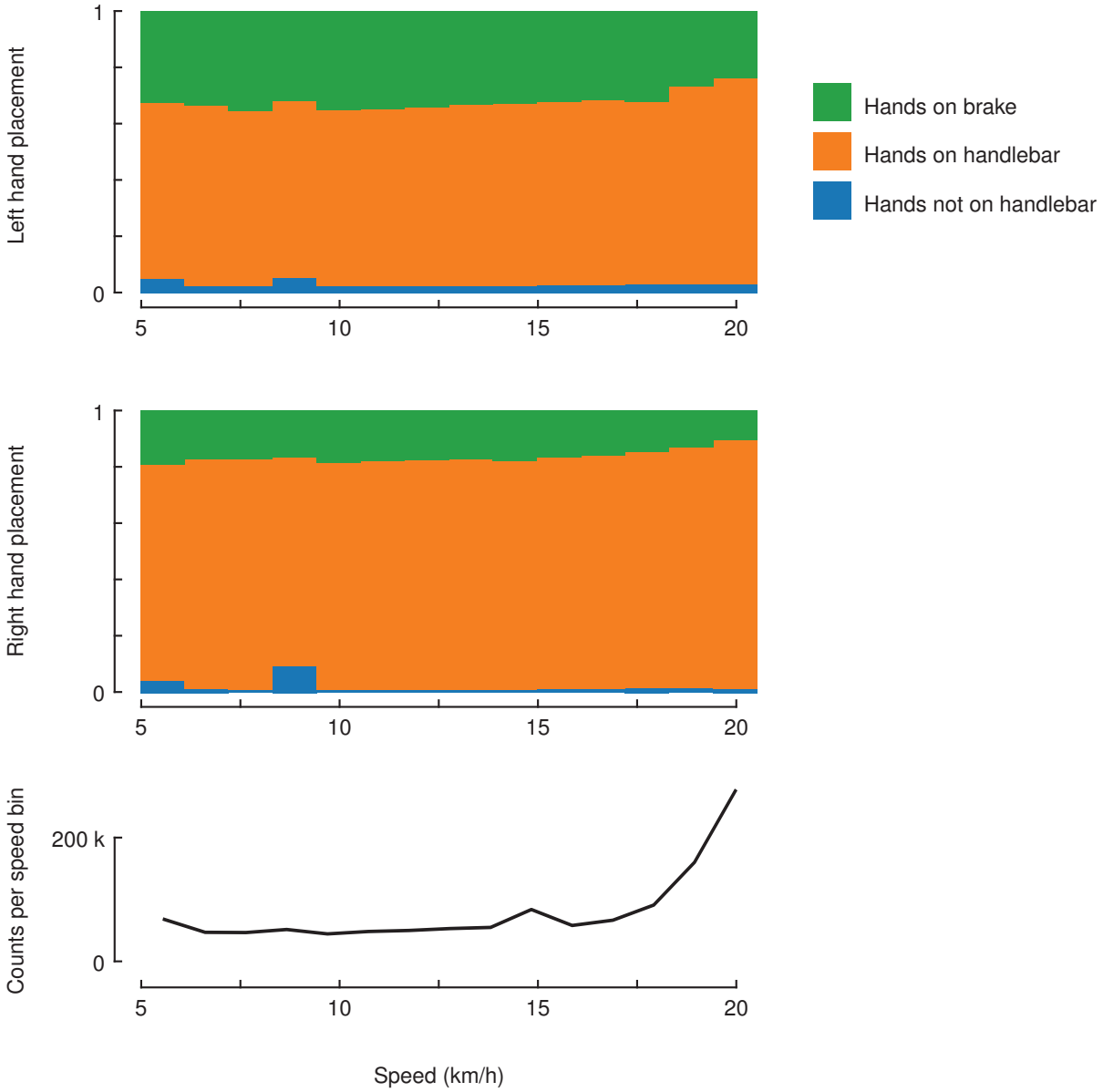
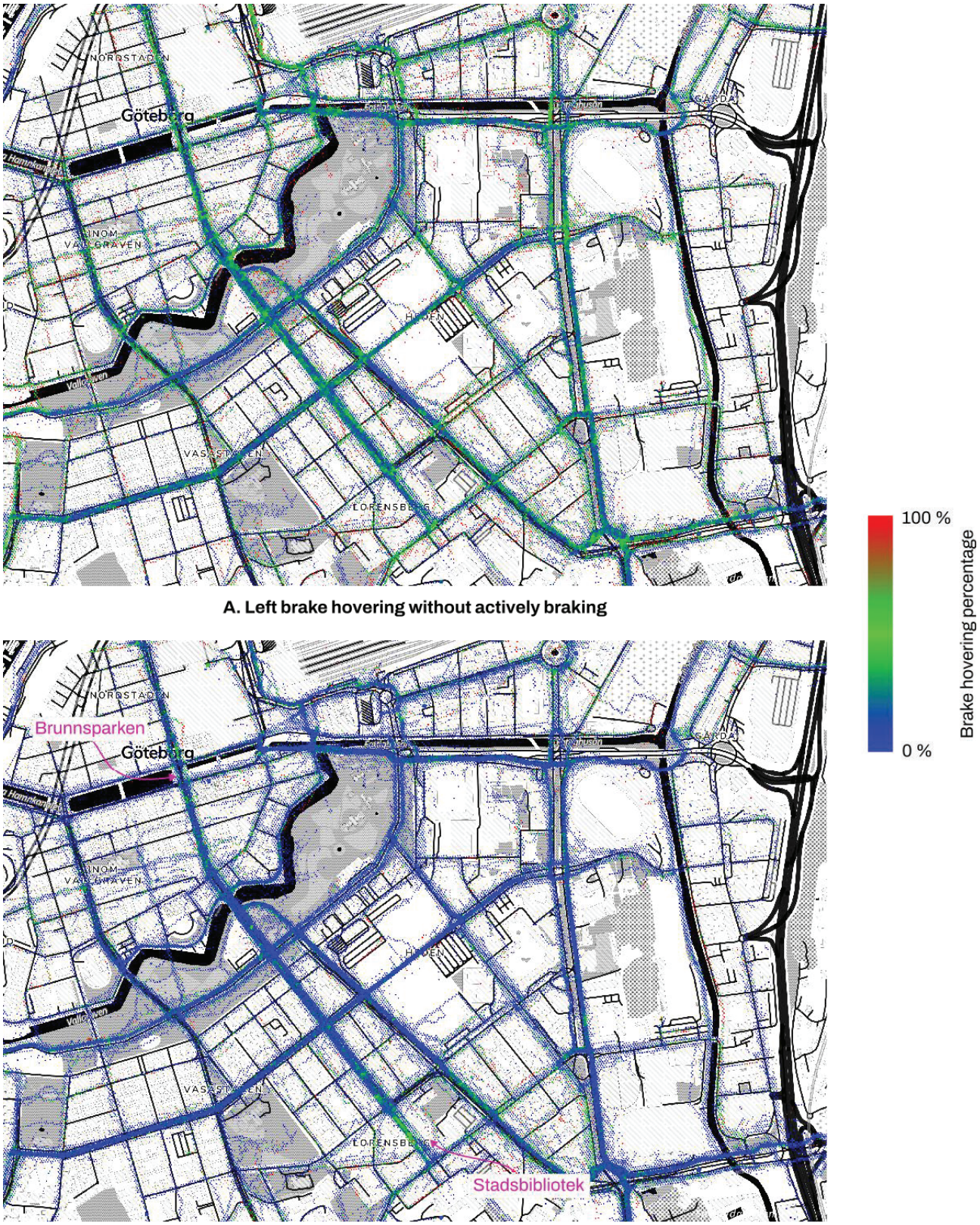


Figure 25: Left and right hand placement distribution as a function of speed.

With our data augmentation, it is also possible to plot the data on a map and see where people usually hover the brake levers. To remind the reader, brake hovering means that the hands are detected to be on the brake lever without activating it. Figure 26 shows the percentage of brake hovering in a part of Gothenburg for the left and right hands (top and bottom, respectively).



A. Left brake hovering without actively braking

B. Right brake hovering without actively braking

Figure 26: Brake hovering of A) the left hand, and B) the right hand.

The maps show that usually the brake hovering happens at intersections or infrastructure crowded with multiple types of road users, this is quite clear for the left-hand behaviour. When it comes to the right hand, this seems to happen at very specific places. For instance, in front of Gothenburg's *Stadsbibliotek* or on the main road passing through *Brunnsparken*.

5.5 Road-User Interaction Modelling

5.5.1 Object Detection

TensorFlow Object Detection

TensorFlow provides an object detection API with out-of-the-box models for generating object detection results [14]. One of its models was trained on the KITTI Dataset [10], a dataset that is composed of data for autonomous driving. Examples of detections are shown below.

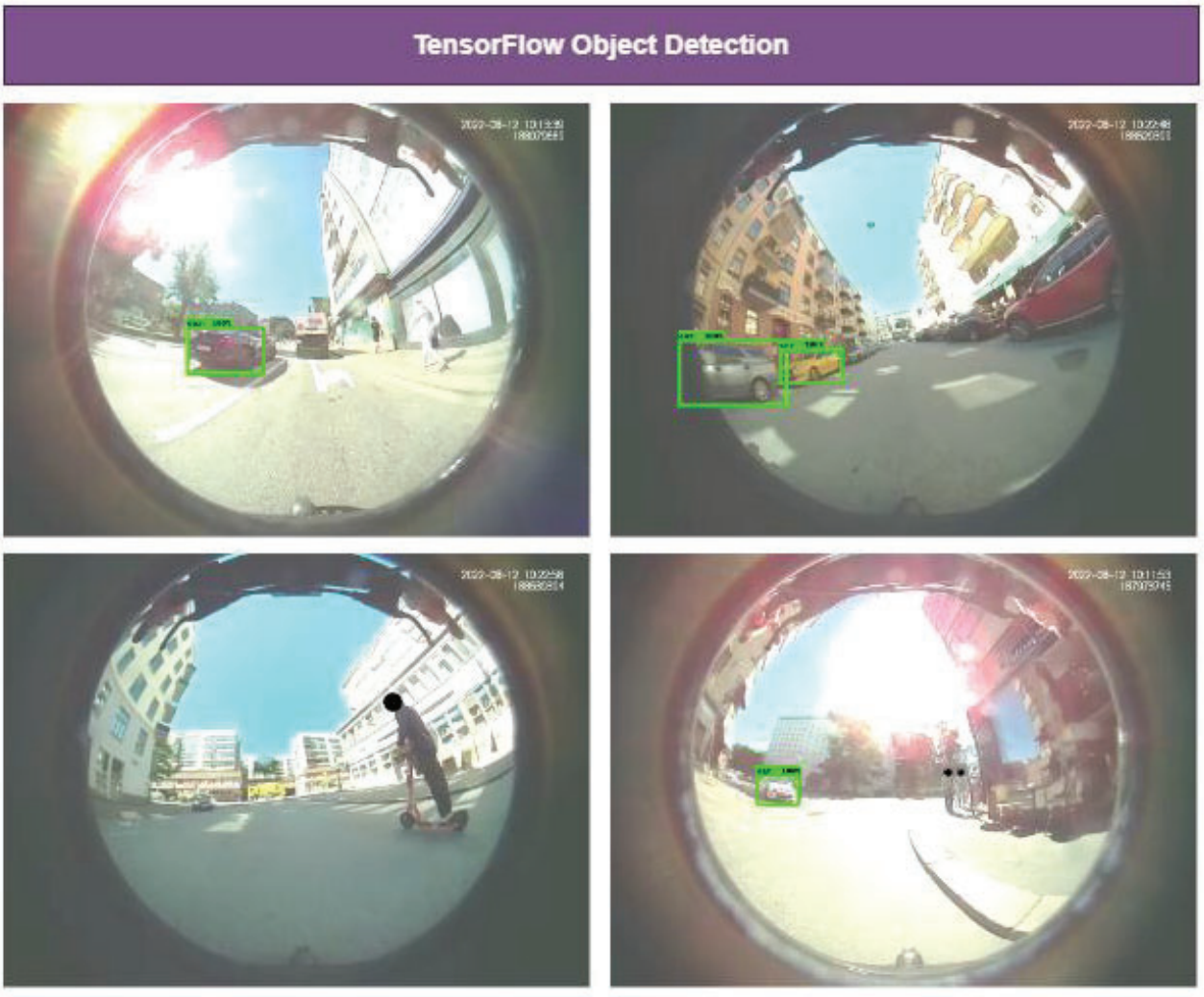


Figure 27: Examples of detections with TensorFlow Object Detection.

The model did detect vehicles and pedestrians. However, some frames did not get detection at all. Lowering the prediction confidence threshold was an option to increase the number of detections. However, given the age of the repository, the researchers explored the other options instead.

Urban Object Detection

Urban Object Detection is an open-source project developed from an initiative from the City of Montréal [29]. The project used traffic cameras in order to detect pedestrians, vehicles, and other road objects. Examples of detections are shown below.

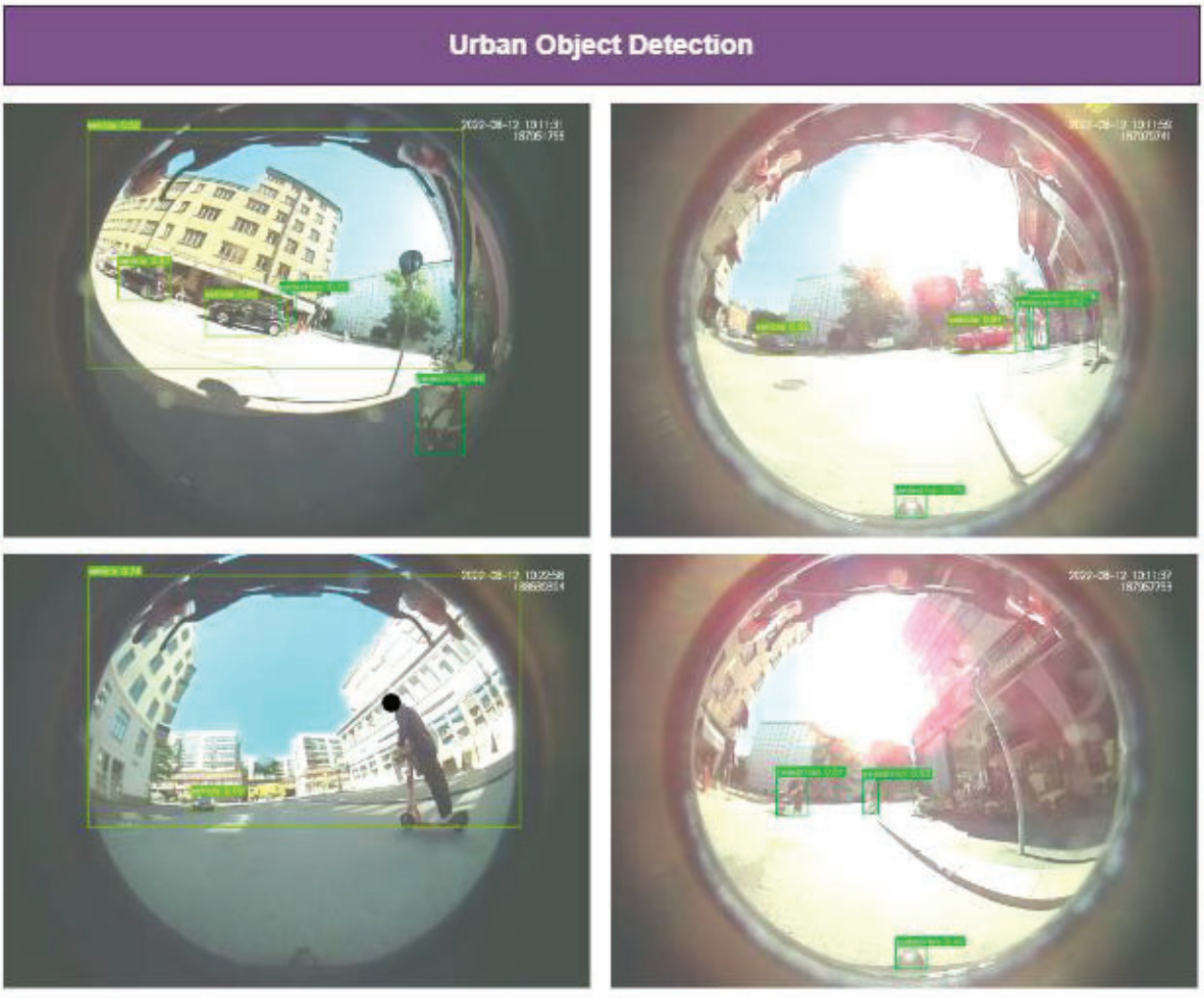


Figure 28: Examples of detections with Urban Object Detection.

The model was good at predicting objects. However, given the nature of the input images, it looked like some of the areas around the edge of the e-scooter videos caused false positives. Decreasing the threshold did not improve this aspect, as some of the false positives have high confidence.

YOLOP

YOLOP is a variant of YOLO (You Only Look Once), created for driving perception [33]. It was trained to perform different kinds of driving related tasks, however, in this application, it was only utilised for its object detection capabilities. The limitation of YOLOP's detection was that it only detects vehicles. Examples of detections are shown below.

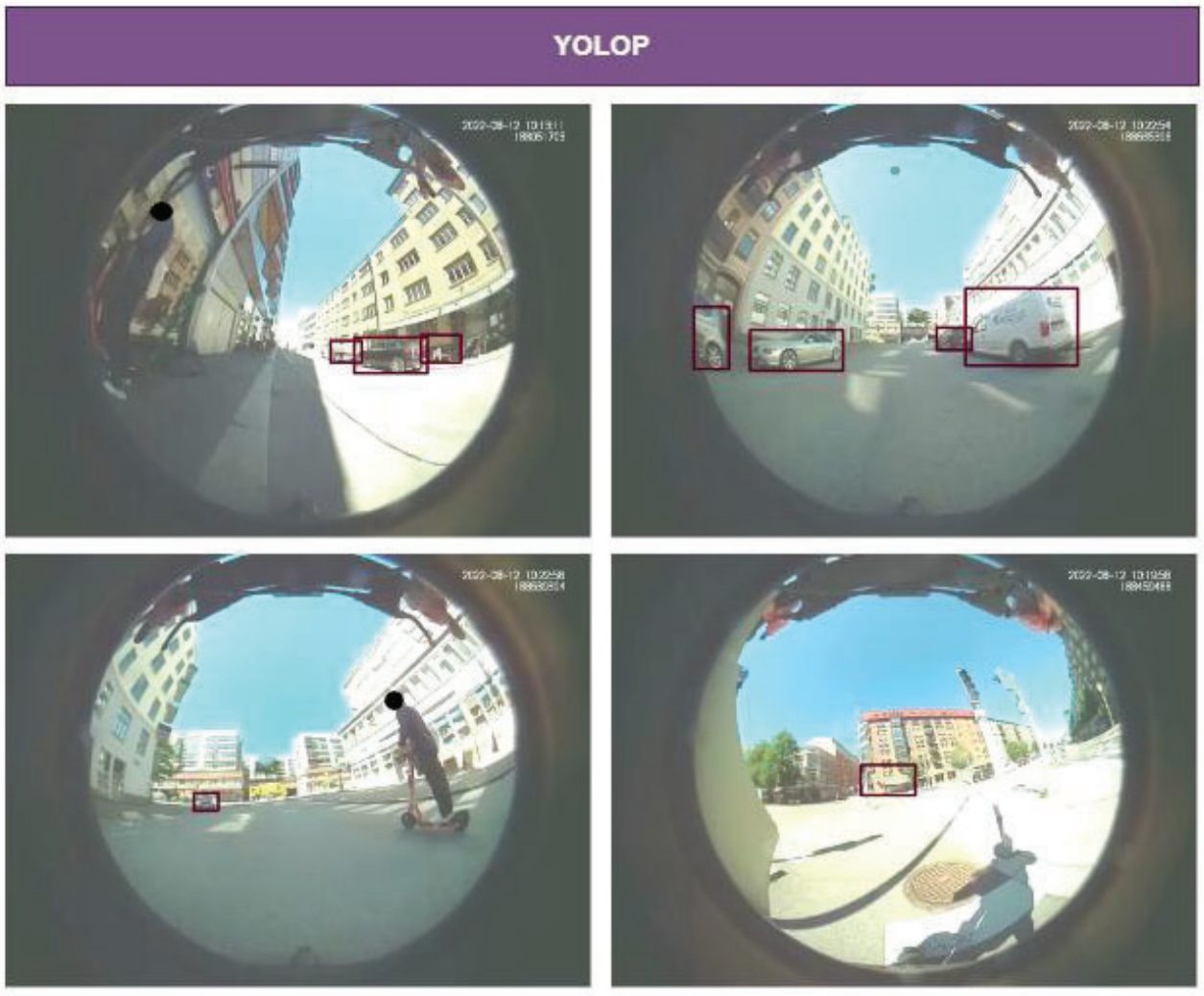


Figure 29: Examples of detections with YOLOP.

YOLOP did a good job at detecting vehicles. It can also detect them even if it is through a mirror. It did generate some false positives on buildings' windows, but overall, the detection performance was consistent.

Pedestron

Pedestron is a model that is specialised for pedestrian detection [12]. It also has a variant that works for crowds and people. Given the good performance of YOLOP, the idea was to explore if specialized models perform better than general ones. Examples of detections are shown below.

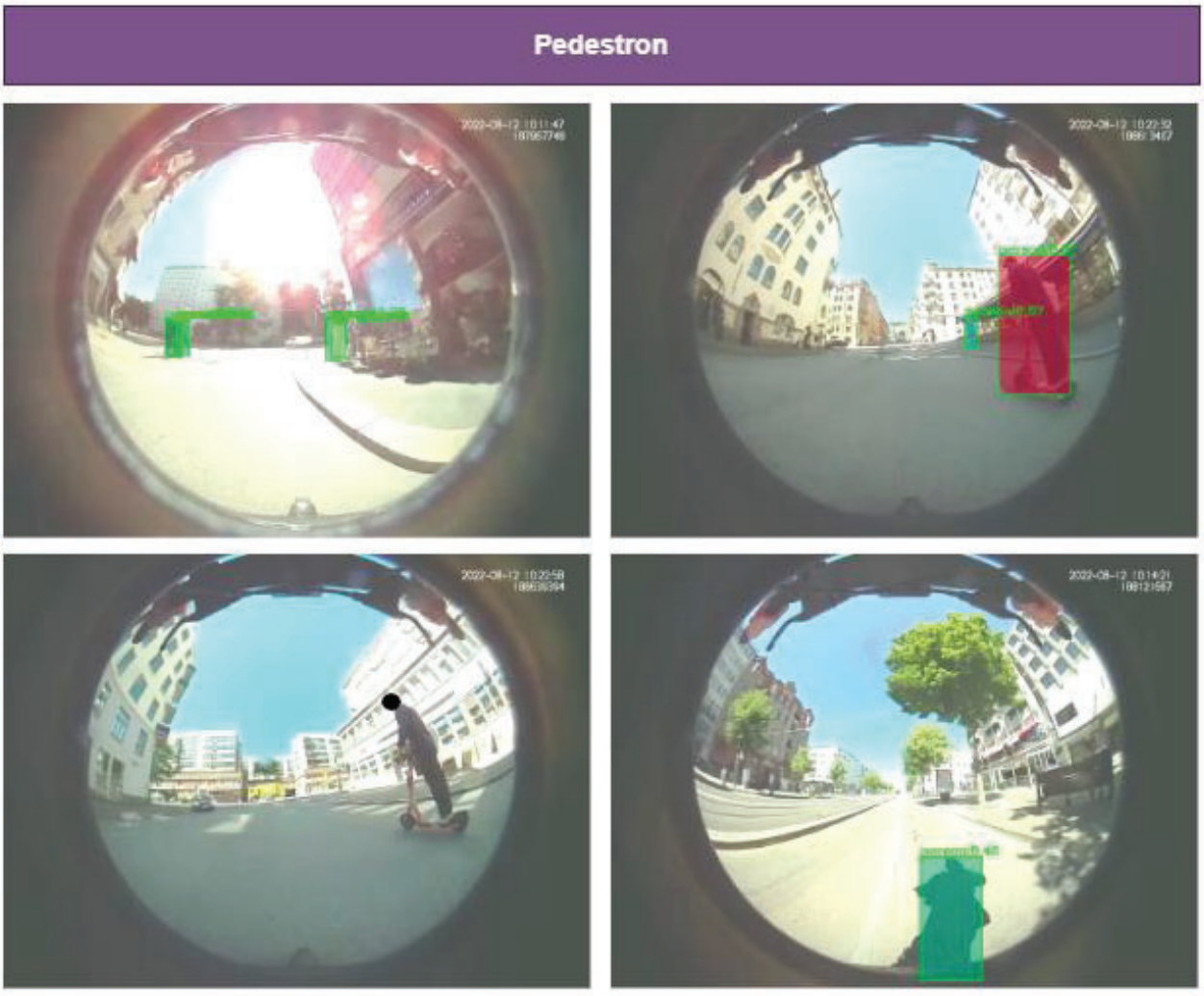


Figure 30: Examples of detections with Pedestron.

While the performance was good in general, the pedestrian detection also missed some people. There were also some inconsistencies in how it handles people on bikes and e-scooters. Another issue was that sometimes even shadows were detected as pedestrians.

YOLOv7

YOLOv7 is an improvement over the YOLO family of models [31]. Compared to the more specific YOLOP, YOLOv7 is a general model that also had better performance across different object detection benchmarks. Examples of detections are shown below.

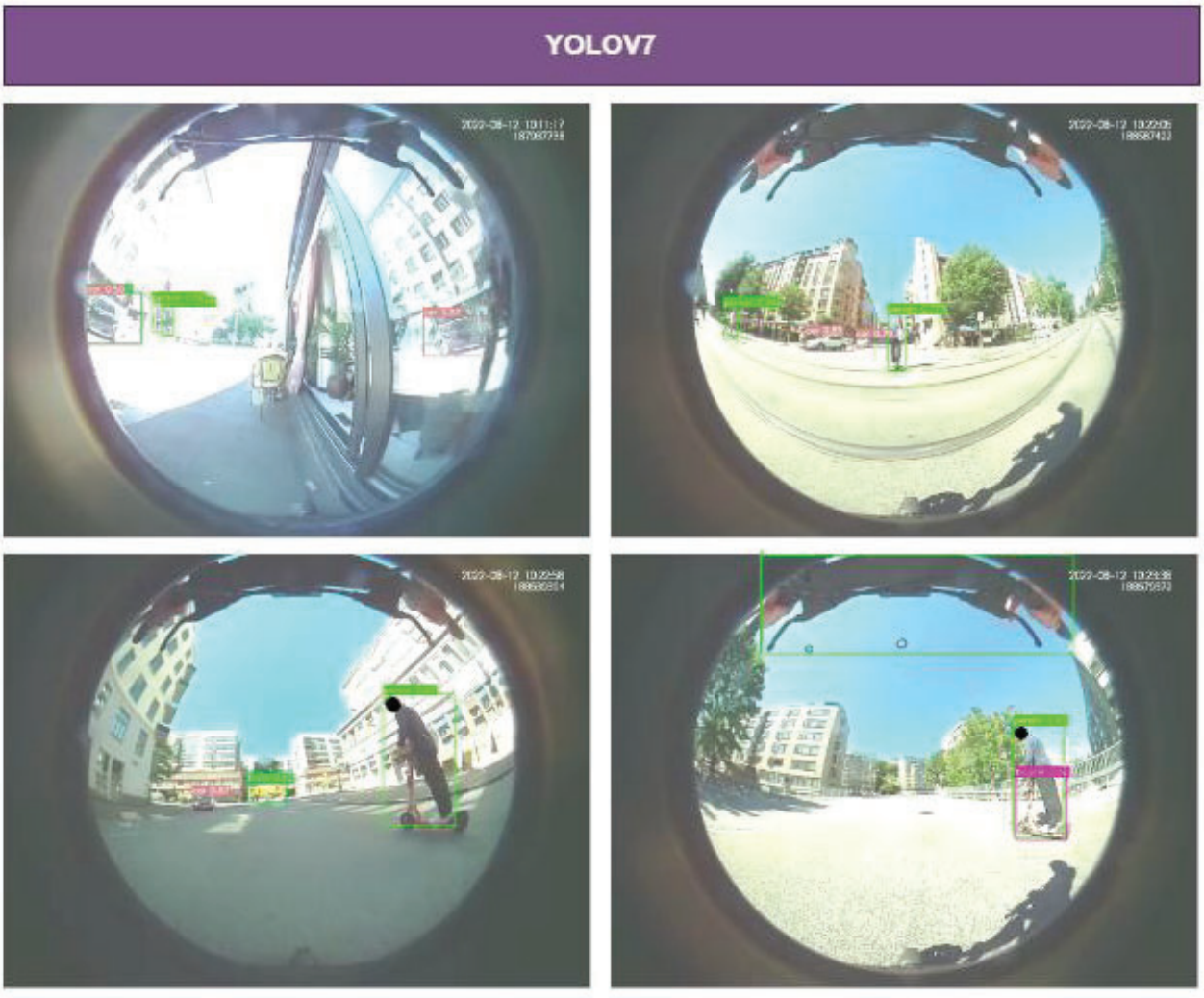


Figure 31: Examples of detections with YOLOv7

The performance of the model was good. But similar to other models, there were also false positives on the video edges and shadows being detected. The model also provided detection on partial objects, which may or may not be ideal depending on the use case.

5.5.2 Object Tracking

Intersection Over Union Tracking

To track the detected object across the video, a basic implementation of tracking algorithm was incorporated using the bounding box overlaps by calculating the intersection over union (IOU) metric.

For each frame, all detections were merged into a single object that had the same class label and met the IOU threshold. The detection that was selected amongst the ones with overlap was the one with the highest confidence.

To create a tracked object, the detections on the previous frame were compared using the

class label and IOU threshold. This was the same condition as the detection merging. If there were two new objects that matched with the previous frame, the one with the higher IOU was tracked together with the previous detection.

To make the tracking more robust to detection drops, a threshold for the number of frames missing was defined. At the end of the tracking, if there were tracked objects that had the same class and met the missing frames and IOU thresholds, they were merged into a tracked object. Figure 32 summarizes the algorithm processing flow.

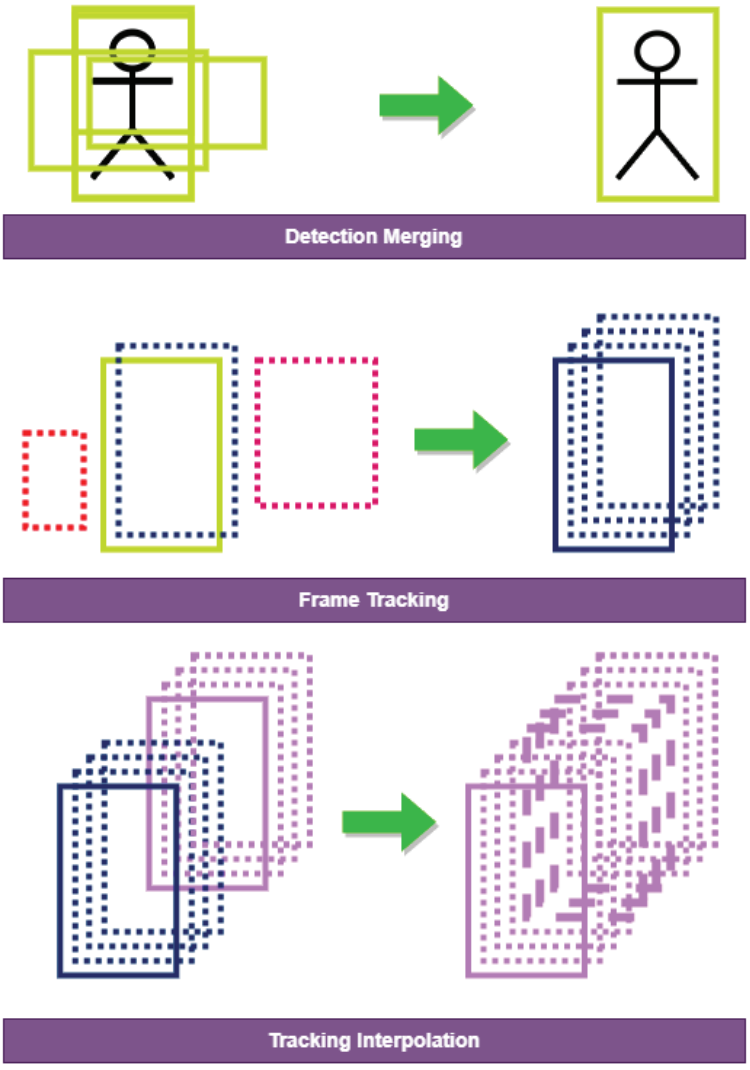


Figure 32: Intersection Over Union tracking's processing flow.

SORT

SORT was one of the top object-tracking open-source solutions at the time of the pre-study [1]. Since SORT only does object tracking, its overall performance was reliant on the performance of the detection model paired with it.

The implementation of SORT that was used does not interpolate the missing bounding

boxes when the detection was missed by several frames. Instead, the missing frames were interpolated linearly as a simple solution. See examples of tracking below.

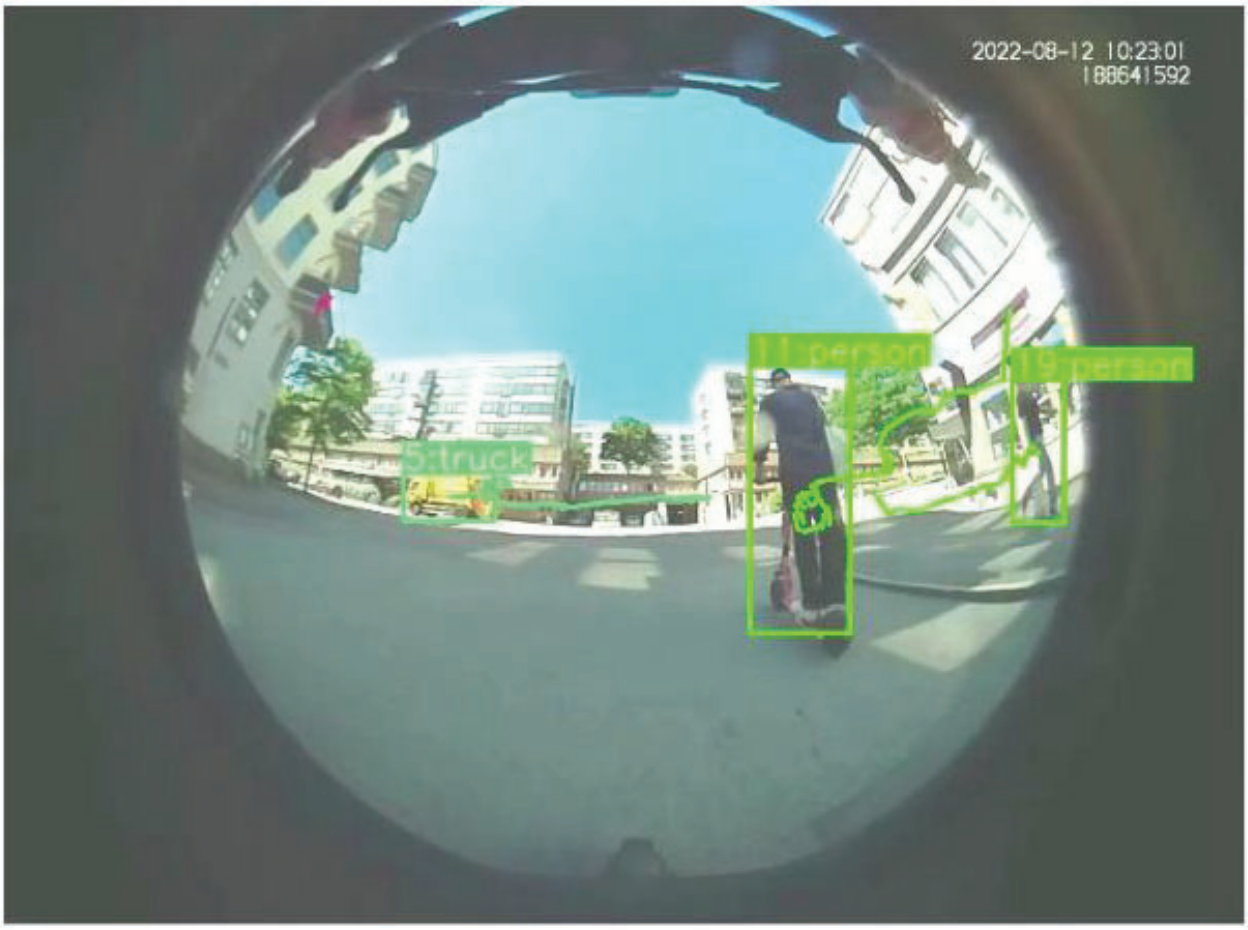


Figure 33: Examples of tracking with SORT.

5.5.3 Distance Estimation

Adabins

Adabins is a model used for estimating distances from an image [2]. Adabins offered two models trained on different datasets: NYU Depth Dataset for indoor images [25] and KITTI Dataset for outdoor images [10].

Visually, Adabins provided good depth maps for the images, however when used in the videos, the predicted depth maps for the frames were inconsistent causing some objects to increase in depth with less movement.

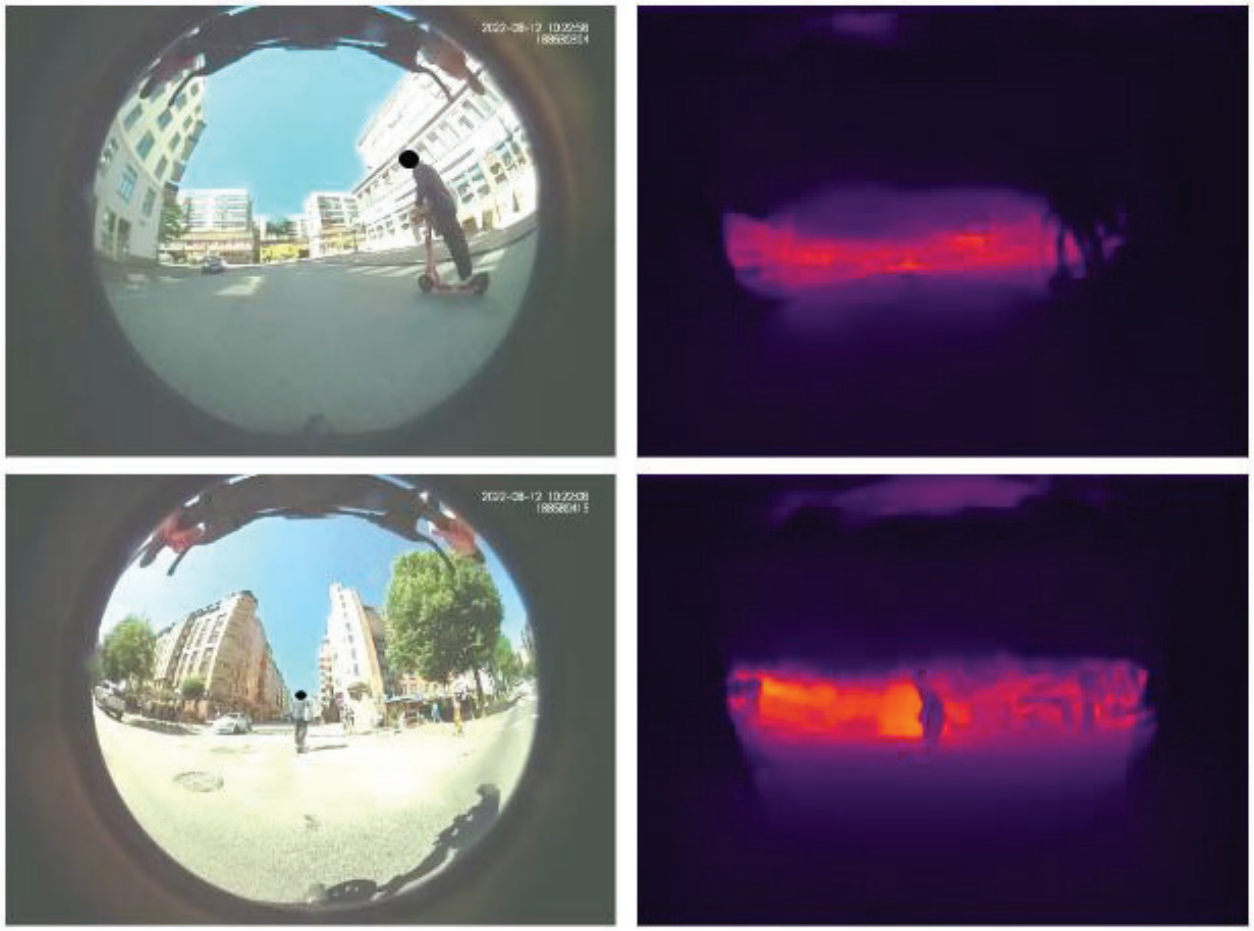


Figure 34: Examples of depth estimation with Adabins.

Analytic Estimation

Fifteen calibration points were used to create the analytic estimation. The calibration points lie on a grid defined by angles of -60, -30, 0, 30, and 60 degrees and distances of 1, 2, and 3 meters (see Figure 35).

Figure 35: Examples of depth estimation with Adabins.

To cover the entire field of view (i.e., to use the analytical model outside the grid rendered

on the rightmost picture on Figure 35), some assumptions were made. First, given an angle, there is a half ellipse in which the points at that angle lie. The vertical axis of the ellipse was calculated by laying down the points of the same angle in an ellipse, then minimizing the error. The horizontal axis was mapped to the angle values (see Figure 36).

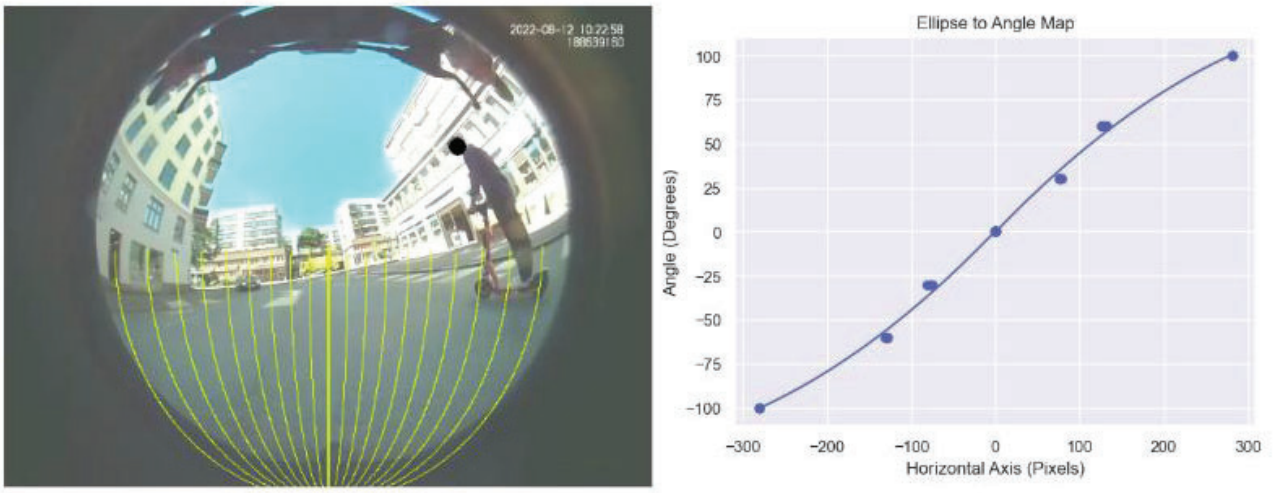


Figure 36: Angle estimation analytic model.

For the distance values, a series of hyperbola was used. For each point below the horizon, points of the same distance belonged to the same hyperbola. The focus of the hyperbola was calculated similarly to the angle ellipse vertical axis. The vertex position was mapped to the distance values (see Figure 37).

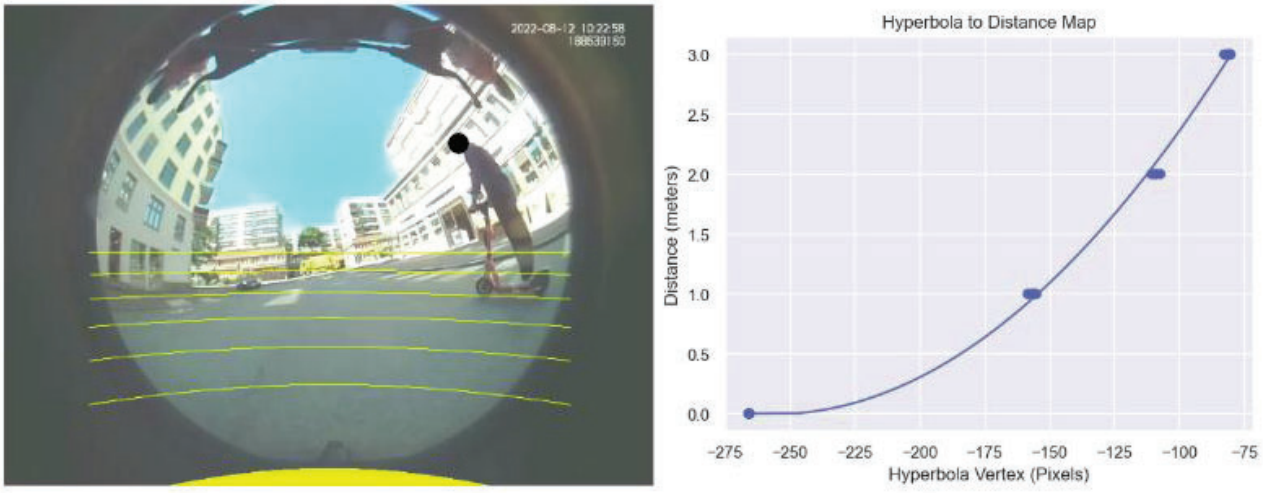


Figure 37: Distance estimation analytic model.

5.5.4 Comfort Boundary Modelling

After exploring the different kinds of algorithms, the detection and distance estimations of other road users with the e-scooter were performed by utilizing the YOLOv7 model for ob-

ject detection, SORT for object tracking, and the analytic model was used to extract the distance of the objects.

The data from 47 trips were used for this analysis. All the relative positions of the detected and tracked pedestrians are plotted on Figure 38. The 5th percentile of the position distributions binned every 6 degrees (i.e., 20 bins between -60 and 60º), are plotted on the thick black line (this could be interpreted as a comfort zone boundary). The subset of the collected data, represented by those 47 trips, shows that the riders tend to allow longer distances in the longitudinal direction (5th percentile around 3.5 m), compared to the lateral directions (around 2.5 m). This could be explained by the fact that the riders keep a comfortable distance of 3.5 m or more to ensure enough distance for an evasive manoeuvre. In the lateral aspects, the users may allow shorter distances because they are overtaking other road users and they are not on a collision path. Of course, it could also be artefacts created by the distance estimation algorithm not accounting for lean or steer angles, or that the flat road surface assumption is broken.

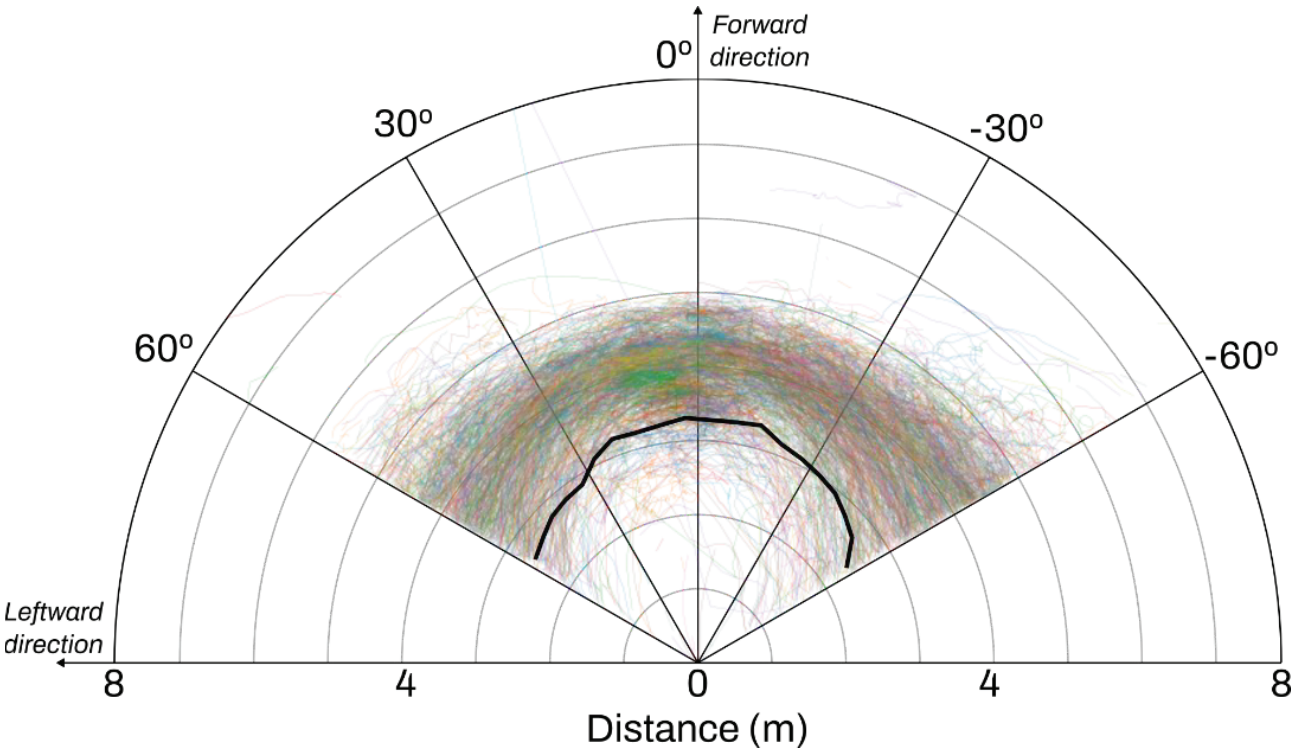


Figure 38: All estimated position of pedestrians with respect to the e-scooter front wheel. The dark thick line represents the 5th percentile of the distance distributions binned along the angle.

6 Discussion

The pre-study collected more than 2000 trips from riders using shared e-scooters in Gothenburg, Sweden. On-cloud data processing and exploration tools were built to minimize the effort needed to transfer data, process them and explore them. But this was only the first step towards preliminary analyses to address the three research questions that guided the project.

6.1 Research Question 1

To address the first research question (**R.Q.1**), a SCE detection algorithm was developed using only the data captured with the 6-axis IMU mounted on the e-scooter's handlebar stem. As the total duration of the collected data was too high to manually review the trips one by one to identify crashes, this was the preferred option. Of the 111 SCEs detected by the algorithm only 25 were crash, near crash, and unsafe behaviours. While the precision of the algorithm is not high (Precision = TP/(TP + FP) = 25/111 = 22.5%), this is not really a problem since the algorithm was mainly used to detect SCEs in a large dataset, any false positive could be discarded quite easily using the developed exploration tools for instance. Regarding false negatives, their total number would be more difficult to evaluate as it would require checking manually all the collected data to find SCEs that were not detected by the algorithm, this evaluation was, therefore, out of scope of the pre-study. Of the 25 detected SCEs, 8 were crashes or falls. Reviewing the events showed that kinematics and video data were definitively bringing information that could be used for event reconstruction. More specifically, the kinematics data could give information of initial conditions for crash tests and simulation, some data (such as time and position) could give some more information about the environment (weather, darkness, etc.). The video data were, however, essential to be able to interpret the kinematics data, without the video it would be quite difficult to replace the kinematics data in the environmental context. However, as the video camera was oriented forwards, no body motion during the crashes or falls was captured and this is of course an essential piece of information if one wants to validate the rider's trajectory during a reconstructed crash or fall. We can therefore address the **R.Q.1** concluding that data gathered during the observed crashes and falls bring necessary information for reconstruction methods. Unfortunately, even more data might be needed to reconstruct a crash or fall to the point of recreating the rider's motion. However, as no severe crash or fall where a rider would fly over the e-scooter was observed, it is difficult to say if some part of the body motion would have been captured or not with the current configuration of the instrumentation. Ideally, having another camera, rider-facing, would bring the other half of the sphere that is missing in the 360º video, but this was not an option for privacy issues.

6.2 Research Question 2

The second research question, **R.Q.2**, was mainly focused on unsafe behaviours that may lead to crashes or falls. In the project, we focused on unsafe behaviours related to 1) SCEs

with high kinematics, 2) speed regulating behaviour with respect to the infrastructure, and 3) hands-on-handlebar behaviour.

For the first type of category, a SCE detector was created in the project. Due to the nature of the motion sensors, one main limitation is the high number of false positives which stem from the fact that rider might do certain activities on the e-scooter which do not necessarily relate to a safety event but nevertheless result in abrupt motion being recorded - resulting in false positives being generated by the algorithm. Despite this, even in its current form the SCE detection algorithm can still provide a useful list of suspected trips and time info for faster investigation of SCEs instead of directly analysing hundreds of trips manually. Humans in the loop are still required for effective use of this approach. This also means that such a system is not suitable for real-time triggering of an active safety system. One potential use case of the algorithm may involve a soft real-time deployment on an e-scooter which issues a visual signal and/or haptic feedback to the rider when unsafe behaviours are detected. Future investigations will need to be carried out to account for the other types of SCE. Detecting Type 2 SCEs will be possible by considering the braking data; however, SCEs of Type 3 will still be missed because they do not result in enough motion activity recorded on the sensors resulting in false negatives. These limitations can be addressed by combining the motion sensor-based SCE detection with a vision-based approach to add extra robustness.

Regarding the speed regulation with respect to the infrastructure, it was first observed that the geofencing and speed limits per zone works perfectly. The second observation was that the average speed of the e-scooter riders is often as high as the speed limit for the given geographical location (which is often 20km/h). This may show that people tend to go as fast as they can regardless of the environment in which they move. This was especially observed where cobblestone paths were used which could cause more instability for e-scooter riders. While cobblestones may have been used in Gothenburg to attempt to slow down cyclists when approaching potential crossing interactions (e.g., zebra crossing, road intersection, etc.), it was not proven that it had any effect other than reduce comfort and surface friction [19], resulting in a potentially riskier surface for the riders. Nowadays, the road construction guidelines in Gothenburg recommend the use of asphalt for the whole cycle path surface [9]. Visual analyses of the maps show, however, that the average speeds are lowered when approaching major intersections. Future work should include a systematic analysis between speed behaviour at intersections compared to other parts of the infrastructure. Another aspect that was observed was that the e-scooter data could provide information about the surface roughness, to some extent. We did not have ground-truth data test the surface roughness estimations but using the video data and online services such as Google Street View, we could confirm that high roughness was usually correlated with the presence of cobblestone paths, tramway rails, or speed bumps. The validation of the estimation could be done in future work with access to the right reference measurements. Nonetheless, the results of the present pre-study strengthen the case of using shared e-scooters to sense meaningful infrastructure data that could be informing cities on the quality of their infrastructure and potentially help them with the road maintenance (or planning). This would benefit all types of mobility (e-scooters—of course—but also pedestrians, cyclists, etc.).

The third type of category was the very interesting aspect of hand placement behaviour which required some consequent data processing work before carrying out its analysis. A hands-on-handlebar classifier was developed using video camera data and, every 100 ms, the left and right hands' placement was estimated (absent, on handlebar, on brake lever). This enabled us to have an estimation of the hands' placement for more than 2000 trips. As we knew that the classifier had poor performance in darkness, we excluded the parts of the trips that were at night. Other limitations of the classifier are partly due to the hardware which could be fixed in the future by choosing adequate cameras, and partly due to the small training dataset used for the machine-learning model which could be also trained with new labelled data to improve the classification performance in the future. The aggregated statistics of the hands-on-handlebar behaviour showed that people usually had both their hands on the handlebar. They hovered over or used more often the left hand's brake lever compared to the right. When looking at these data as a function of speed, we saw that people had the tendency to have their hands on the brake lever less once the e-scooter top speed was reached. This reinforces the idea that e-scooterists tend to accelerate to the maximum speed and other speeds are only intermediate speeds where they do not plan to stay long (either by braking hence hands-on-brake-lever or by accelerating). Looking at the brake hovering behaviour on a map, we could see that people tended to anticipate when approaching intersections or potential unsafe locations which tends to demonstrate that, for the most part, people rode e-scooters in a safe way.

The analyses and results show that it seems possible to identify unsafe riding behaviours (e.g., speed regulation behaviour) using only kinematics data (i.e., without the video data) but some specific metrics such as hand placements require video camera or, potentially, specialised sensors or systems to capture those metrics.

6.3 Research Question 3

On the instrumented e-scooters used for the pre-study, the video camera was the sensor providing the analysts with the most contextual information which is important to better understand how riders behave. For the third research question, **R.Q.3**, the video data were heavily used as the camera was the only installed sensor that could inform about the vicinity of the e-scooter to other objects and road-users. After evaluating multiple off-the-shelf algorithms, a data processing pipeline using the best-performing algorithms was created to 1) detect the type of road users in interaction with the e-scooter riders, 2) track the detected road users, and 3) estimate the position of the tracked road users. The estimates from the pipeline are still to be validated, but manual verification showed that when the assumptions were not broken, the estimates seemed to be in the right order of magnitude. However, as soon as the e-scooter was leaning or steering, or the ground was not perfectly flat, the estimates seemed to be clearly off. This is one point that will need to be addressed in future work, how can we account for steering and leaning during the estimation of the distance to other road users? Another point that could be improved in the future is the detection step that may not be optimal with the video data input as it was a fish-eye projection while the object detection algorithms evaluated use perspective-projected images.

When aggregating the distances to all detected and tracked pedestrians for 47 trips, we

could see a pattern that e-scooter riders seemed to follow: that being, to keep some minimum distance (or “comfort distance”) to the pedestrians in interaction with them. We can clearly say that the dataset is, therefore, a valuable resource for studying and modelling road user interaction with e-scooter riders although some limitations still exist due to technical issues. Some of those issues could be addressed in future work.

7 Dissemination and Publications

7.1 Dissemination

How are the project results planned to be used and disseminated?	Mark with X	Comment
Increase knowledge in the field	X	Preliminary results of the project were shared with SAFER's Safety Performance Evaluation Research Area in November 2022. In a more general way, the naturalistic riding data analyses within this pre-study increase the knowledge in the field and the collected data is a rich dataset for future projects to better understand what can be done to improve micro-mobility safety.
Be passed on to other advanced technological development projects		
Be passed on to product development projects		
Introduced on the market		
Used in investigations / regulatory / licensing / political decisions		

The research work will be extended in another project called "*e-SAFER - Datamodeller för säkra interaktioner mellan (automatiserade) fordon och elsparkcyklar*" partially funded by VINNOVA's FFI programme (project number 2022-01641).

7.2 Publications

Except for the present report, no public publications were made. The work will be carried over the next FFI project (2022-01641) and is expected to generate multiple scientific publications.

8 Conclusion and Future Work

Naturalistic e-scooter riding data from more than 2000 trips were collected during this pre-study and multiple axes of research were investigated: 1) crash investigation and reconstruction, 2) unsafe riding behaviours, and 3) road-user interaction modelling. The results of the analyses presented in this report are promising and are either in line with previous literature or bring new knowledge that helps to better understand how e-scooter riders operate their shared e-scooter in safe or unsafe ways and how they interact with other road users. Results may also help industry to make e-scooter riding safer, and the cities to better understand how they can make their road infrastructures adapted to micro-mobility transportation as e-scooters.

Insights from the results demonstrate that the data collection strategy, processing, and augmentation are valuable and commend continuation which will be done in another project called "*e-SAFER - Datamodeller för säkra interaktioner mellan (automatiserade) fordon och elsparkcyklar*" partially funded by VINNOVA's FFI programme (project number 2022-01641). More data will be collected during this next project and the data processing algorithms presented in the present report will be improved to tackle down limitations that were observed during the pre-study.

9 Bibliography

[1] Alex Bewley et al. "Simple Online and Realtime Tracking". In: *2016 IEEE International Conference on Image Processing (ICIP)*. Sept. 2016, pp. 3464–3468. DOI: 10.1109/ICIP.2016.7533003. arXiv: 1602.00763 [cs].

[2] Shariq Farooq Bhat, Ibraheem Alhashim, and Peter Wonka. "AdaBins: Depth Estimation Using Adaptive Bins". In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2021, pp. 4008–4017. DOI: 10.1109/CVPR46437.2021.00400. arXiv: 2011.14141 [cs].

[3] Alan Blatt et al. "Naturalistic Driving Study: Field Data Collection". In: *SHRP 2 Report* SHRP 2 Report S2-S07-RW-1 (2015).

[4] Stig Nikolaj Fasmer Blomberg et al. "Injury from Electric Scooters in Copenhagen: A Retrospective Cohort Study". In: *BMJ Open* 9.12 (Dec. 2019), e033988. ISSN: 2044-6055, 2044-6055. DOI: 10.1136/bmjopen-2019-033988.

[5] Jessica B. Cicchino, Paige E. Kulie, and Melissa L. McCarthy. "Injuries Related to Electric Scooter and Bicycle Use in a Washington, DC, Emergency Department". In: *Traffic Injury Prevention* 22.5 (July 2021), pp. 401–406. ISSN: 1538-9588. DOI: 10.1080/15389588.2021.1913280.

[6] Marco Dozza. "BikeSAFER: Final Report". In: (2013).

[7] Marco Dozza, Julia Werneke, and Michael Mackenzie. "e-BikeSAFE: A Naturalistic Cycling Study to Understand How Electrical Bicycles Change Cycling Behaviour and Influence Safety." In: *International Cycling Safety Conference 2013*. Helmond, The Netherlands, 20-21 November 2013.

[8] R. Eenink et al. *UDRIVE: The European Naturalistic Driving Study*. Proceedings Paper. Paris, France, Apr. 2014.

[9] Olle Evenäs. *Interaktion mellan fotgängare och cyklister - En kunskapssammanställning och utgångspunkter för utformning i Göteborg*. Tech. rep. Göteborgs Trafikkontoret, Mar. 2020.

[10] A Geiger et al. "Vision Meets Robotics: The KITTI Dataset". In: *The International Journal of Robotics Research* 32.11 (Sept. 2013), pp. 1231–1237. ISSN: 0278-3649. DOI: 10.1177/0278364913491297.

[11] T. D. Gillespie, M. W. Sayers, and L. Segel. "CALIBRATION OF RESPONSE-TYPE ROAD ROUGHNESS MEASURING SYSTEMS". In: *NCHRP Report* 228 (Dec. 1980). ISSN: 0077-5614.

[12] Irtiza Hasan et al. *Generalizable Pedestrian Detection: The Elephant In The Room*. Dec. 2020. DOI: 10.48550/arXiv.2003.08799. arXiv: 2003.08799 [cs].

[13] Narelle Haworth, Amy Schramm, and Divera Twisk. "Comparing the Risky Behaviours of Shared and Private E-Scooter and Bicycle Riders in Downtown Brisbane, Australia". In: *Accident Analysis & Prevention* 152 (Mar. 2021), p. 105981. ISSN: 0001-4575. DOI: 10.1016/j.aap.2021.105981.

[14] Jonathan Huang et al. *Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors*. Apr. 2017. DOI: 10.48550/arXiv.1611.10012. arXiv: 1611.10012 [cs].

[15] Katrine Karlsen and Torkel Bjørnskau. *Road User Interactions: A Survey from Nine Urban Areas in Norway*. Tech. rep. 1771-2020. 2020.

[16] Ma (fh) Ernestine Mayer and Klaus Robatsch. "5D.001 E-Scooters: What Do They Mean for the Safety of Cyclists and Pedestrians?" In: *Injury Prevention* 27.Suppl 2 (Mar. 2021), A43–A43. ISSN: 1353-8047, 1475-5785. DOI: 10.1136/injuryprev-2021-safety.130.

[17] Nikan K. Namiri et al. "Electric Scooter Injuries and Hospital Admissions in the United States, 2014-2018". In: *JAMA Surgery* 155.4 (Apr. 2020), pp. 357–359. ISSN: 2168-6254. DOI: 10.1001/jamasurg.2019.5423.

[18] Vicki L Neale et al. "An Overview of the 100-Car Naturalistic Study and Findings". In: *National Highway Traffic Safety Administration, Paper* 5 (2005), p. 0400.

[19] Anna Niska. *Service levels of cycleways – state-of-the-art report focusing on maintenance and operation*. Tech. rep. VTI, 2011.

[20] Rahul Rajendra Pai. "Logging Data From E-Scooters To Improve Traffic Safety". MA thesis. Gothenburg: Chalmers University of Technology, 2022.

[21] Thaddeus J. Puzio et al. "The Electric Scooter: A Surging New Mode of Transportation That Comes with Risk to Riders". In: *Traffic Injury Prevention* 21.2 (Feb. 2020), pp. 175–178. ISSN: 1538-9588. DOI: 10.1080/15389588.2019.1709176.

[22] M. W. Sayers, T. D. Gillespie, and W. D. O. Paterson. "GUIDELINES FOR CONDUCTING AND CALIBRATING ROAD ROUGHNESS MEASUREMENTS". In: *World Bank Technical Paper* Technical Paper 46 (1986). ISSN: 0253-7494.

[23] Michael W. Sayers. *The Little Book of Profiling: Basic Information about Measuring and Interpreting Road Profiles*. Technical Report. University of Michigan, Ann Arbor, Transportation Research Institute, Sept. 1998.

[24] Nitesh R. Shah et al. “Comparison of Motor Vehicle-Involved e-Scooter and Bicycle Crashes Using Standardized Crash Typology”. In: *Journal of Safety Research* 77 (June 2021), pp. 217–228. ISSN: 0022-4375. DOI: 10.1016/j.jsr.2021.03.005.

[25] Nathan Silberman et al. “Indoor Segmentation and Support Inference from RGBD Images”. In: *Computer Vision – ECCV 2012*. Ed. by Andrew Fitzgibbon et al. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2012, pp. 746–760. ISBN: 978-3-642-33715-4. DOI: 10.1007/978-3-642-33715-4_54.

[26] Karen Simonyan and Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. Apr. 2015. DOI: 10.48550/arXiv.1409.1556. arXiv: 1409.1556 [cs].

[27] H. Stigson, I. Malakuti, and M. Klingegård. “Electric Scooters Accidents: Analyses of Two Swedish Accident Data Sets”. In: *Accident Analysis & Prevention* 163 (Dec. 2021), p. 106466. ISSN: 0001-4575. DOI: 10.1016/j.aap.2021.106466.

[28] Jay Todd et al. *Behavior of Electric Scooter Operators in Naturalistic Environments*. SAE Technical Paper 2019-01-1007. Warrendale, PA: SAE International, Apr. 2019. DOI: 10.4271/2019-01-1007.

[29] *Urban Object Detection*. Ville de Montréal. Apr. 2022.

[30] Nicole van Nes et al. “The Potential of Naturalistic Driving for In-Depth Understanding of Driver Behavior: UDRIVE Results and Beyond”. In: *Safety Science* 119 (Nov. 2019), pp. 11–20. ISSN: 0925-7535. DOI: 10.1016/j.ssci.2018.12.029.

[31] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. *YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors*. July 2022. DOI: 10.48550/arXiv.2207.02696. arXiv: 2207.02696 [cs].

[32] Mats Wendel and Kenneth Natanaelsson. *Underhållsstandard Belagd Väg 2011*. Tech. rep. 2012:074. Trafikverket, 2012.

[33] Dong Wu et al. “YOLOP: You Only Look Once for Panoptic Driving Perception”. In: *Machine Intelligence Research* 19.6 (Dec. 2022), pp. 550–562. ISSN: 2731-5398. DOI: 10.1007/s11633-022-1339-y.

10 Participating parties and contact persons

Christian-Nils Boda, Ph.D.
Senior Research Engineer
*firstname.lastname*_AT_ autoliv.com

André Dankert, Ph.D.
Embedded Machine Learning Developer
*firstname.lastname*_AT_ voiapp.io