Learn from IoT: Pedestrian Detection and Intention Prediction for Autonomous Driving

Gürkan Solmaz*, Everton Luís Berz*, Marzieh Farahani Dolatabadi†, Samet Aytaç‡, Jonathan Fürst*, Bin Cheng*, Jos den Ouden†
*NEC Laboratories Europe, ‡Boğaziçi University, †Eindhoven University of Technology (TU/e)
*firstname.surname@neclab.eu, m.dolatabadi.farahani@tue.nl, samet.aytac@boun.edu.tr, j.h.v.d.ouden@tue.nl

ABSTRACT

This paper explores the potential of machine learning (ML) systems which use data from in-vehicle sensors as well as external IoT data sources to enhance autonomous driving for efficiency and safety in urban environments. We propose a system which combines sensor data from autonomous vehicles and IoT data collected from pedestrians' mobile devices. Our approach includes two methods for *vulnerable road user* (VRU) detection and pedestrian movement intention prediction, and a model to combine the two outputs for potentially improving the autonomous decision-making. The first method creates a *world model* (WM) and accurately localizes VRUs using in-vehicle cameras and external mobile device data. The second method is a deep learning model to predict pedestrian's next movement steps using real-time trajectory and training with historical mobile device data. To test the system, we conduct three pilot tests at a university campus with a custom-built autonomous car and mobile devices carried by pedestrians. The results from our controlled experiments show that VRU detection can more accurately distinguish locations of pedestrians using IoT data. Furthermore, up to five future steps of pedestrians can be predicted within 2 m.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in ubiquitous and mobile computing**; • **Computer systems organization** → Neural networks.

SMAS'19, October 21, 2019, Los Cabos, Mexico

ACM ISBN 978-1-4503-6930-5/19/10…$15.00
https://doi.org/10.1145/3349622.3355446

KEYWORDS

autonomous vehicles; vulnerable road user detection; deep neural networks; internet of things

ACM Reference Format:
Gürkan Solmaz*, Everton Luís Berz*, Marzieh Farahani Dolatabadi†, Samet Aytaç‡, Jonathan Fürst*, Bin Cheng*, Jos den Ouden†. 2019. Learn from IoT: Pedestrian Detection and Intention Prediction for Autonomous Driving. In *1st ACM Workshop on Emerging Smart Technologies and Infrastructures for Smart Mobility and Sustainability (SMAS'19), October 21, 2019, Los Cabos, Mexico.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3349622.3355446

1 INTRODUCTION

There have been major advancements in ML research due to the increased availability of computation capabilities as well as open datasets (e.g., popularly used labeled image datasets) which allow researchers to easily benchmark their approaches against the state of the art. Furthermore, the development of ML frameworks such as TensorFlow enable easy prototyping and experimenting with new ML systems. Simultaneously, in recent years, the Internet of Things (IoT) has been expanding to many domains such as smart cities and mobility. Access to resources such as IoT platforms (e.g., Microsoft Azure, AWS IoT) and large-scale experimental IoT testbeds [9] enables researchers in the field to conduct experiments.

The main motivation of this paper is to learn from IoT to improve the safety of future self-driving vehicles. Although today's autonomous cars have many sensors and also give control to the driver for improving safety, incidents such as the Uber accident (2018) show that both car sensors and drivers can easily fail to detect the pedestrians. IoT extends the connectivity in a way that many devices are connected. In addition, the devices become cheaper and more efficient. Through vehicle-to-vehicle (V2V) or vehicle-to-everything (V2X) communications in the 5G era and edge computing, road infrastructures in urban areas can provide support for traffic safety.

This paper aims to combine IoT and ML for enhancing autonomous driving safety. The proposed system learns from data collected from in-vehicle sensors (e.g., cameras) and external IoT sources such as mobile devices of VRUs. The components included in the setup are the following: 1) autonomous car, 2) mobile devices with the mobility app, 3) in-vehicle IoT platform, 4) VRU detection and intention prediction components, and 5) cloud IoT platform. ML leverages data sources for training such as data collected by the vehicle at the Eindhoven University of Technology (TU/e) campus, our smartphone app for movement tracking, and mobile ITS-G5 location devices. We propose two methods for accurate VRU detection and pedestrian intention prediction. The first method is creating a WM by combining the vehicle and mobile device data and performs VRU detection and localization. The second method uses the latest trajectory data (coming from the vehicle's camera or mobile device) and predicts the pedestrians' intended movement steps based on pedestrian behaviors from historical measurements using a deep learning model. We also discuss a Petri Net model showing under which conditions which inputs can be combined for the autonomous decision-making.

To validate the proposed methods, we conduct three pilot tests at the TU/e campus using our custom-built autonomous car and the mobile devices (with the app) carried by pedestrians. The results from our controlled experiments show that the proposed approach produces promising results with minimum errors in VRU localization and prediction of future movements without much latency. We believe the proposed system can complement the existing autonomous driving systems to provide an additional layer of safety.

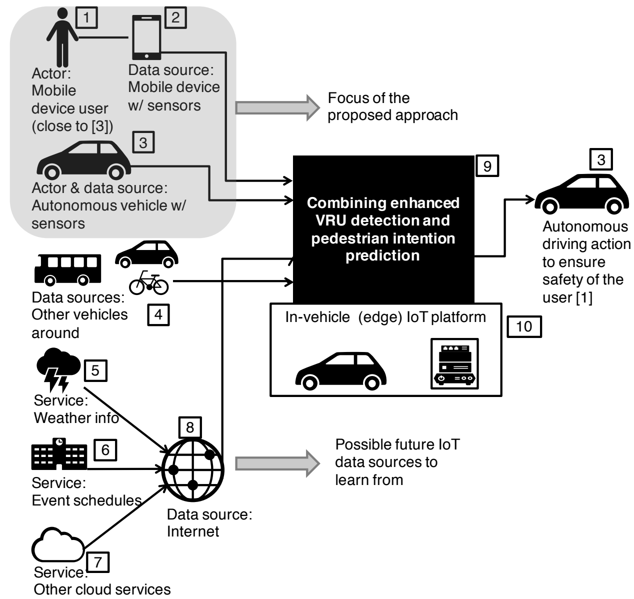


Figure 1: A perspective for improving autonomous driving decisions using IoT data sources. The proposed approach (the black box) combines enhanced VRU detection and pedestrian intention prediction.

2 RELATED WORK

Methods and technologies used for the VRU detection have different performance and constraints. Most of the previous works leverage computer vision to detect road users, whereas the recent works also propose predicting VRU intentions [10]. However, the computer vision approaches have problems such as occlusion and false positives. For instance, a pedestrian can be easily occluded by an obstacle or even by another person, decreasing the solution accuracy [7].

Goldhammer et al. [3] present movement models based on ML methods to classify the motion state and to predict the future trajectory of VRUs. Realistic data is captured by the vehicle's laser scanners and high-resolution cameras. Results show an accuracy of 88.6% for the motion state classification and a reduction of the trajectory prediction error by 41% on stopping motion scenarios. Bastani et al. [1] present a warning system using smartphones to protect VRUs. The system is activated by a geometric model and a fuzzy inference engine estimates the collision risk. In real-world evaluation samples, results show a 96% accuracy in six types of accidents. This system does not consider the pedestrian intention and behavior or intersection scenarios. Lastly, Murphey et al. [8] propose three methods to predict pedestrian positions. In this work, dead reckoning and neural networks models are evaluated on the trip data recorded from two pedestrians using smartphones. Results show an error between 0.57 m and 3.22 m for the best-case scenario. This study uses only smartphone data for the location of pedestrians, whereas our approach combines accurate detection from the autonomous vehicle sensors with real-time and offline mobile device data for pedestrian intention prediction.

3 APPROACH

The goal of the proposed approach is to combine accurate VRU detection and pedestrian intention prediction with the enhancement of learning from IoT data sources. A simplistic view of this idea is given in Fig. 1. Possible data sources and actors are listed on the left side, given as inputs. These inputs are combined for autonomous driving actions of the car on the left side. The remainder of this section includes the following: 1) VRU detection, 2) pedestrian intention prediction, 3) how to combine the two methods to support autonomous decisions.

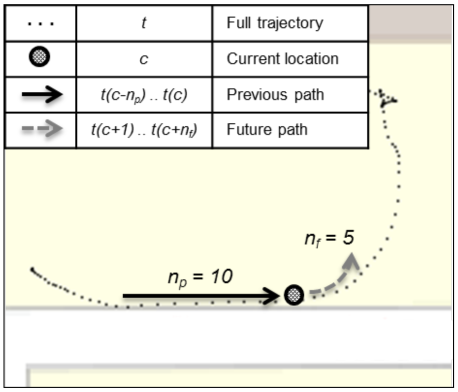


Figure 2: An example pedestrian trajectory and the defined variables.

3.1 VRU Detection

The VRU detection is based on the WM which contains the vehicle itself and the objects in its surrounding. The formalism adopted in this research is WIRE[1] where the WM aims to track semantic objects such as pedestrians and bicycles. To estimate features and track objects depending on the objects that are involved, an autonomous car should overcome various challenges. Incorporation of the Multiple Hypothesis Tracker (MHT) in an anchoring strategy is a solution applied in the WM. Applying this algorithm, it is possible to combine different forms of evidence into a common and updated world representation dynamically. Objects' attributes, classification, and prior knowledge are associated in the hypotheses tree. Every hypothesis contains a list of anchors and has a correctness probability. Each anchor on its turn contains an individual symbol, a set of measurements and a probabilistic signature that consists out of a mixture of probability density functions (PDFs) generated by a set of behavior models. The predicate attribute space represents predicate grounding relations that link attribute values and predicate symbols [2].

3.2 Pedestrian Intention Prediction

This method predicts the next location of pedestrians based on historical data and the current position. The applied ML model uses the mobile device data such as speed and GPS trajectory values of pedestrians to predict their next movements. The representation of pedestrian trajectory used on this work is inspired by the model in [8]. Fig. 2 shows the path modeling with an example pedestrian trajectory that appeared during the experiments at the campus.

The deep learning model consists of three input features, three concatenation layers, and n_f output features, where n_f is the number of future locations for every (x, y) coordinate.

[1] http://wiki.ros.org/wire

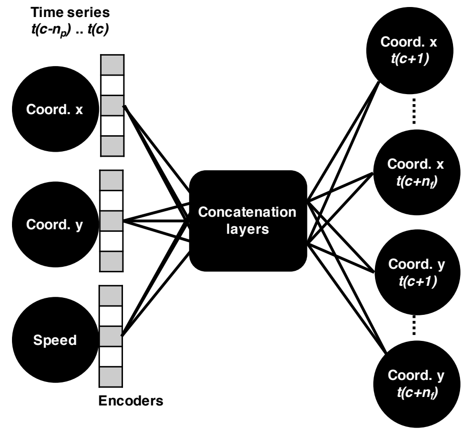


Figure 3: The deep learning model for pedestrian intention prediction.

The Adam algorithm [5] is applied for the optimization process with ReLU activation function. Fig. 3 shows the network model. The input layer consists of time-series values from the previous n_p positions (included in the latest n_p smartphone data) and their respective speed. Two input neurons represent x, y coordinates and one neuron contains the speed values. An embed encoder is applied to map the inputs into vectors and then forward to the concatenation layers. These intermediary layers concatenate all outputs of the feature encoders and pass the concatenation through fully connected layers. We have cross-validated the model with one, two and three concatenation layers. The model with three layers is selected as it has a better overall performance. In the output layer, each neuron represents the future location of the pedestrian, starting from the first next position $t(c + 1)$ until $t(c + n_f)$, where t represents the trajectory sequence and c is the current location.

To feed the model with past and future positions, we define $n_p = 10$ and $n_f = 5$, which are approximately equivalent to 11 m and 5.5 m in a straight-walking distance, respectively. The data used to train and optimize the model is randomly partitioned into training (70%), validation (20%), and testing (10%) subsets.

3.3 Combining VRU Detection and Intention Prediction

We propose combining the inputs from the two previous parts to support the autonomous decision-making. Here, we use a simple model based on stochastic priority Petri Nets [6]. Fig. 4 shows the model with the places (big circles), transitions (rectangles), and tokens at the initial stage. λ denotes the probabilistic variables based on the transition step, whereas the curly-braced numbers indicate the priority labels and numbers w/o curly braces denote the number of

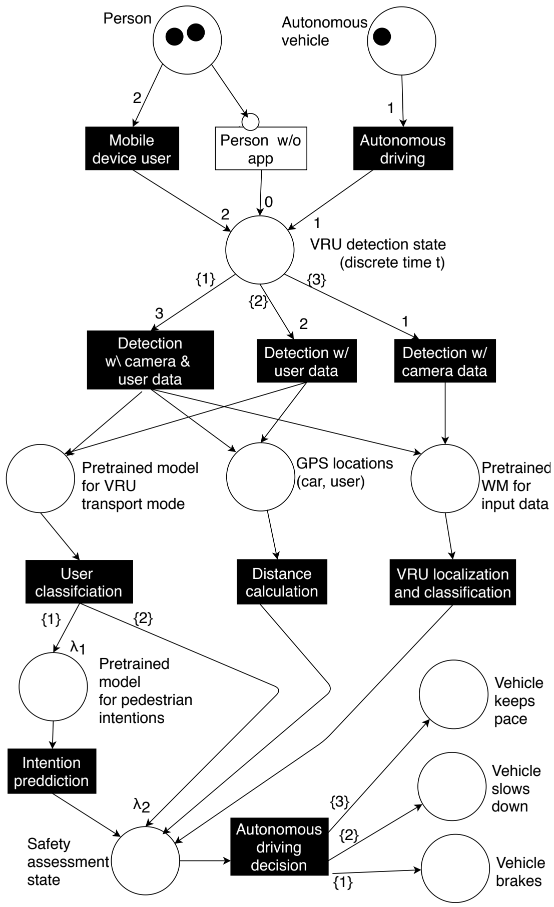
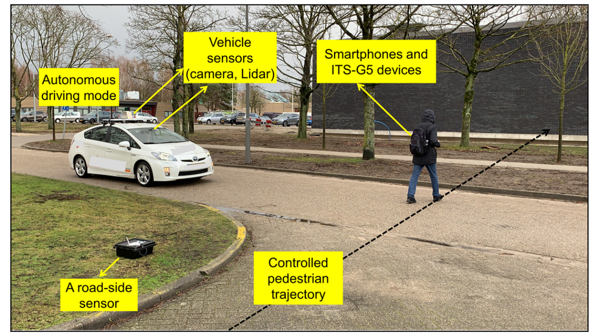
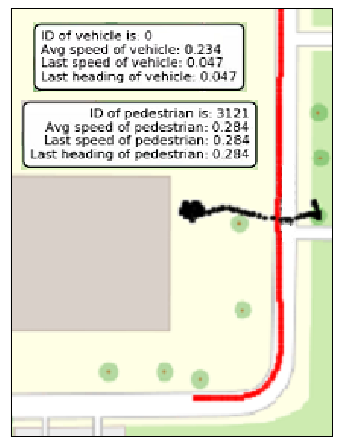


Figure 4: The Petri Net model showing the cases when the proposed methods are applicable to support autonomous decision-making.

tokens. The model considers two types of VRUs: 1) users of mobile devices and our app (w/ IoT data) and 2) people w/o the app. Three possible cases exist when a person is in the vicinity: 1) both the mobile device and vehicle sensor data, 2) only mobile device data, 3) only vehicle sensor (camera) data is available in the in-vehicle platform. For cases 1 and 3, the WM creates a list of VRUs and classifies them (i.e., pedestrian/cyclist). For case 1, the detected pedestrians are matched with the the mobile device w/ available data. The outputs of the pretrained intention-prediction model is used for assessing the *safety* (bottom-left). For case 2, the distance is calculated and given to safety assessment using the vehicle and person location data. For case 1, a state-of-the-art approach [4] can be used to predict the transport mode to filter out passengers and cyclists. In the case of pedestrian existence, the pretrained model for pedestrian intentions provides additional input for the safety assessment state. Other than the shown inputs, other IoT data from the vehicle, environment, or people can be leveraged by the autonomous



(a) A controlled experiment run: The walking-straight scenario.



(b) Animation of the above experiment run.

Figure 5: Visualizing the experiments through animations. Pedestrian: the short trajectory (from left to right), car: the long trajectory (from bottom to top).

decision-making for a set of final actions (bottom-right). For simplicity, we include three probabilistic actions: 1) keeping same pace, 2) slowing down, and 3) brake, whereas more complex maneuvers can be considered.

4 EVALUATION

4.1 Pilot Setup

We conduct the pilot tests mainly at the TU/e campus. The campus has a 2 km road network and tests are executed with a speed limit of 15 km/h. The custom-built autonomous car prototype (Toyota Prius) is used. The car has a custom mobile ITS-G5 device connected to the in-vehicle IoT platform. Two pedestrians also carry these ITS-G5 devices. The IoT Gateway in the in-vehicle platform is connected to two cloud IoT platforms via MQTT and HTTP using the cellular 4G connection. The Robot Operating System (ROS) collects

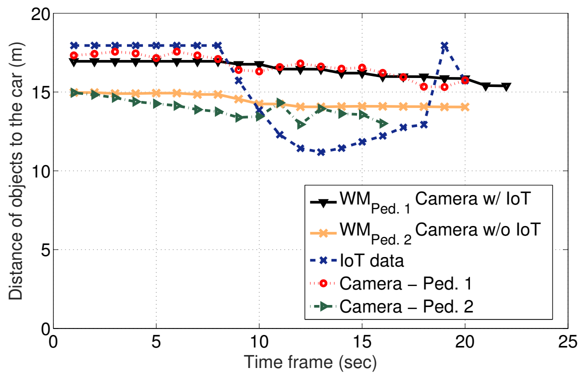


Figure 6: The output of the WM. The relative pedestrians-car distance using camera and IoT data.

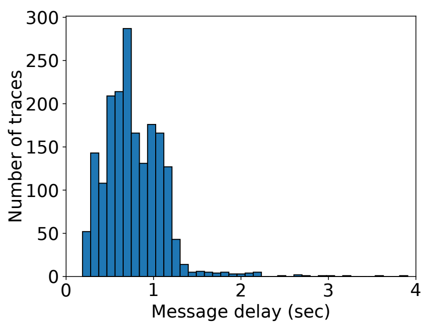


Figure 7: Message delays from smartphone to vehicle ROS during the initial pilot tests.

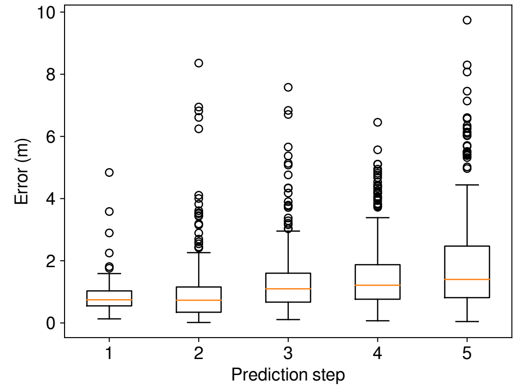


Figure 8: Distribution of the error for every step.

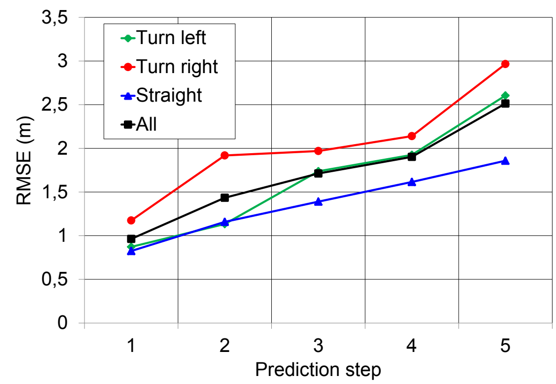


Figure 9: RMSE of the pedestrian intention prediction model for each walking scenario.

data from the vehicle and pedestrians (via direct communication with ITS-G5 or indirectly from cloud platforms). It operates on the in-vehicle IoT platform. So far, we conducted prototype integration tests and 3 pilot tests (each 1-2 weeks long).

We store the pilot tests data in *rosbag* and CSV files. Each in-vehicle or cloud component publishes data to ROS in real-time (via an IoT Gateway). The controlled experiments data is analyzed for the performance of the methods. We conducted 21 controlled experiments (consists of a total of 70 runs) where the pedestrian movements are predefined, whereas the autonomous-driving behaviors are mostly uncontrolled and the driving is affected by the vehicle sensors or other factors. The controlled experiments include autonomous and manual driving modes. We visualize the experiments by various animation tools. Fig. 5 shows one of the controlled experiment runs and a view from the visualization of the trajectories through discrete-time animation.

4.2 Experimental Results

VRU detection results: As a proof of concept of the VRU detection for autonomous cars, the method is applied on the data streams from the vehicle sensors and IoT data. Table 1 lists the input data employed in the WM. A constant velocity Kalman filter and a uniform distribution of objects movement are used as the initial conditions. In the experiments, we associate the camera data with the IoT data. This is the same as it is being done in anchoring algorithms and using prior knowledge to prediction. Fig. 6 includes results from an experiment with two pedestrians, one with ITS-G5 and smartphone devices and one w/o devices. The mobile devices representing the pedestrian send the IoT data to the vehicle. As it has shown in this figure, the VRU detection has a better prediction for pedestrian 1 with the IoT data, since the vehicle receives the global location of a pedestrian a few seconds before the camera detection. After a few seconds, the IoT error becomes significantly large compared to the more accurate camera detection.

While ITS-G5 device has message latency measured in msec, using smartphones (with 4G) has longer delays as the data comes through the cloud IoT platform. Fig. 7 shows the latency of smartphone location data arrival to in-vehicle IoT platform, more specifically to the ROS component. Most of

Devices	Type	Input for the WM	Frequency	Error (m)
Camera	Vehicle sensor	Number of objects, relative distance from each object to the car (extracted from the image frame)	8 frame/sec	<2
Smartphone	IoT	ID, latitude, longitude	1 frame/sec	<3
ITS-G5 device	IoT	ID, latitude and longitude of the mobile device	1 frame/sec	<3

Table 1: The world model inputs.

the smartphone location data arrive in less than 1 sec with an average of ~0.6 sec.

Intention prediction results: The deep learning model proposed in Section 3.2 is trained with data from all 70 runs. We use Ludwig[2] and TensorFlow™ to train and validate the model. After the training, three new sets of experiments in distinct scenarios are run to evaluate the method. In the first scenario, a pedestrian crosses a street walking straight. In the second and third, instead of crossing, a pedestrian leaves a building and then turns right or left on the sidewalk. Each scenario is run 6 times. Fig. 5 shows a photo and GPS traces collected during a run of the “walking-straight" scenario.

The trained model predicts the next five possible steps of the pedestrian. Each data created by the app represents a step. The time between two consecutive data is in the range of [0.5, 1] sec. The position error is given by the Haversine distance between the predicted and actual locations. Fig. 8 shows the error distance for all experiments. The lowest error for the first predicted step is 0.1 m whereas most of the predictions result in less than 1 m errors. Moreover, the latest (fifth-step) prediction on the path sequence has an average error of 1.9 m.

Fig. 9 shows the root mean square error (RMSE) results which are calculated as

$$RMSE = \sqrt{\frac{\sum_{t=1}^{k} (\hat{x}_t - x_t)^2 + (\hat{y}_t - y_t)^2}{k}}, \quad (1)$$

where $\hat{x_t}$,$\hat{y_t}$ and x_t,y_t represent the predicted and actual locations, and k is the number of runs in the experiment. The walking-straight scenario has the best accuracy with a max error of 1.9 m in the furthest prediction step, while the turn-right movement has the lowest accuracy with errors between 1.2 m and 3 m. These results show that the straight-walk intention has a better correlation with the previous steps. In comparison to the existing MLP network model [8], our model performs 0.2 m better on average, considering that our longest prediction step is equivalent to the longest time measured on that work. Moreover, the results also show a decrease of 18% and 48% in the error compared to the NARX and Dead Reckoning approaches, respectively.

[2] http://uber.github.io/ludwig/

5 CONCLUSION AND FUTURE WORK

This paper proposes learning from IoT data for the key pedestrian detection and intention prediction problems to improve the safety of autonomous driving. The proposed system achieves promising performance in the experiments. We believe the VRU detection and pedestrian intention prediction can complement the existing safety systems. Our future work includes extracting features from other IoT data sources such as OpenStreetMap and students' lecture schedules which may further increase the prediction accuracy.

Acknowledgment: This work has been partially funded by the EU Horizon 2020 Programme under Grant Agreements No. 731993 AUTOPILOT (Automated Driving Progressed by Internet Of Things) project. The content of this paper does not reflect the official opinion of the EU. Responsibility for the information and views expressed therein lies entirely with the authors.

REFERENCES

[1] Roya Bastani Zadeh, Mehdi Ghatee, and Hamid Reza Eftekhari. 2018. Three-Phases Smartphone-Based Warning System to Protect Vulnerable Road Users Under Fuzzy Conditions. *IEEE Trans. on Intelligent Transportation Systems* 19, 7 (2018), 2086–2098.

[2] Jos Elfring et al. 2013. Semantic world modeling using probabilistic multiple hypothesis anchoring. *Robotics and Autonomous Systems* 61, 2 (2013), 95–105.

[3] Michael Goldhammer et al. 2018. Intentions of vulnerable road users-detection and forecasting by means of machine learning. *arXiv preprint arXiv:1803.03577* (2018).

[4] Samuli Hemminki, Petteri Nurmi, and Sasu Tarkoma. 2013. Accelerometer-based transportation mode detection on smartphones. In *Proc. of the ACM SenSys*. 1–14.

[5] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint:1412.6980* (2014).

[6] Irina A Lomazova and Louchka Popova-Zeugmann. 2016. Controlling Petri Net Behavior Using Priorities for Transitions. *Fundam. Inform.* 143, 1-2 (2016), 101–112.

[7] Patrick Mannion. 2019. Vulnerable road user detection: state-of-the-art and open challenges. *arXiv preprint:1902.03601* (2019), 1–5.

[8] Yi Lu Murphey, Chang Liu, Muhammad Tayyab, and Divyendu Narayan. 2017. Accurate pedestrian path prediction using neural networks. In *Proc. of the IEEE SSCI*. 1–7.

[9] Luis Sanchez et al. 2014. SmartSantander: IoT experimentation over a smart city testbed. *Computer Networks* 61 (2014), 217–238.

[10] B. Völz et al. 2016. A data-driven approach for pedestrian intention estimation. In *Proc. of the IEEE ITSC*. 2607–2612.