Towards Understanding Pedestrian Behavior Patterns from LiDAR Data

Chi Zhang[1], Christian Berger[1], Marco Dozza[2]

[1] University of Gothenburg, Sweden, Department of Computer Science and Engineering
[2] Chalmers University of Technology, Sweden, Department of Mechanics and Maritime Sciences

Abstract

Pedestrian behavior prediction is an essential task for automated driving (AD) systems. Research of this topic in recent years has been primarily stimulated by significant achievements of deep learning (DL). However, few researches exploit 3D LiDAR point cloud data to include human body posture information while predicting the pedestrian's behavior. In this research, we aim to propose a novel approach which exploits 3D LiDAR point cloud data and uses pedestrian posture information in addition to pedestrians' history trajectories, social interaction, and scene information to identify possible behavior patterns. DL methods heavily depend on the amount and quality of data that is used for training and validation. Recently, Waymo has released a large-scale real-world dataset, which consists of 1,150 scenes that each span 20 seconds, containing well synchronized and calibrated 3D LiDAR and 2D camera data with high quality labeled bounding boxes. Our approach will be trained and evaluated on the Waymo Open Dataset. The results shall contribute to our European research project SHAPE-IT[1] that investigates the influence of human factor on the design and evaluation of AD.

1. Introduction

Pedestrian behavior prediction is essential for automated driving (AD) systems as it can help Automated Vehicles (AVs) to make better decisions and to prevent hazardous situations. As pedestrians are very agile that can change both their direction and velocity without reducing the speed [1,2], it is difficult to reliably predict their intentions. Hence, much work has been devoted to improve the pedestrian behavior prediction performance by deep learning (DL) methods. Long-Short Term Memory (LSTM) networks have showed their ability on pedestrian sequence prediction tasks, and it is adopted by many researchers to predict behavior sequences from pedestrians [3,4,5].

A camera is often a preferred sensor to predict the pedestrians' intention [2]. However, cameras usually have a very limited field of view and lack depth information. On contrast, 3D data such as point cloud data collected by Light Detection and Ranging (LiDAR) can provide depth information as well as rich geometric and shape information [6], which can improve the understanding of a pedestrian's behavior during complicated traffic conditions. Besides, as the image captured by cameras lack shape information, the postures of a pedestrian and people within their surroundings are usually not taken into consideration when using camera data for training. For this reason, we propose a method that uses LiDAR data as input, and to include the posture information of pedestrians.

In this research, we aim to design, implement, and systematically evaluate a network to predict pedestrian behavior by using an LSTM-based network with LiDAR point cloud data. The posture information will be considered in our network. A history-aware, context-aware, interaction-aware, and posture-aware trajectory prediction network is targeted. Features are learned from raw point cloud data by our network instead of being designed manually. The proposed network will be validated and tested on the largest released urban scenario dataset – Waymo Open Dataset [7]. Our focus area in this research is to infer a pedestrian's intention to cross the street and to predict their future trajectories.

2. Related Work

LSTM-Based Trajectory Prediction: LSTM-based methods, an improved version of RNNs, are preferred by many researchers for pedestrian trajectory prediction. Alahi *et al.* [3] proposed "Social LSTM" to predict pedestrian trajectories by learning social interaction from network. Xue *et al.* [5] introduced scene information to the LSTM-based framework. A combined attention model is applied over LSTM by Fernando *et al.* [4] that utilizes "soft" and "hard-wired" attention.

[1] Cf. www.shape-it.eu

However, these methods are all based on camera 2D image data and did not include pedestrians' posture information.

Pedestrian's Intention Prediction from LiDAR Data: Due to drawbacks such as illumination impacts and the lack of depth and shape information, some researchers exploited the LiDAR data for pedestrian's intention prediction. Volz *et al.* [1,2] and Zhao *et al.* [8] tried to involve LiDAR data to predict pedestrians' crossing road intention. However, these approaches did not fully exploit the advantages of the sensor. For example, the shape information is not part of the prediction. Besides, they only predicted the crossing intention, without forecasting the future trajectory of a pedestrian.

Datasets: High-quality, large-scale datasets are crucial for data-driven machine learning algorithms. KITTI [9] is a widely used multi-sensor dataset, which comprises 389 stereo and optical flow image pairs, and over 200,000 3D object annotations of synchronized LiDAR and stereo images. Recently, Waymo released a large scale, high quality, diverse dataset [7], which consists of 1,150 scenes that each span 20 seconds, containing LiDAR and camera data labeled by 2D and 3D bounding boxes and unique track ids. This diverse and abundant dataset attracted the attention in academia of 3D detection and tracking and is used for our research.

3. Methodology

In our research, 3D point cloud data is used to predict intentions and trajectories of multiple pedestrians simultaneously. We are aiming at developing and evaluating a network structure that is: **a) History aware:** Inspired by SS-LSTM [5], we consider using an LSTM-based approach to process history trajectories of pedestrians and predict their future behaviors. The history trajectories for target pedestrians are labeled manually. **b) Context aware:** To involve the local scene context, an additional LSTM network is used to process the context feature. Here we use a CNN feature extractor, which encodes the scene feature from the bird-eye-view frame. **c) Interaction aware:** The influence of social neighborhood is also considered for trajectory prediction. Soft and hard-wired attention [4] can be used to deal with the pedestrian of interest and the neighbors in different distances. **d) Posture aware:** Human body pose can be important for predicting future intention, since different postures of pedestrians usually indicate different behaviors. We can use a subnetwork to learn the body pose feature of the pedestrian, and then the feature is fed into an LSTM network.

Currently, we are using the Waymo Open Dataset to perform Social-LSTM method to create a baseline of our developed approach. The scene context and posture are then to be added into the network.

Acknowledgements

The research is funded by the European Commission within the Marie Skłodowska-Curie Action (MSCA) Innovative Training Network (ITN) programme.

References:

[1] B. Volz, K. Behrendt, H. Mielenz, I. Gilitschenski, R. Siegwart, and J. Nieto. A data-driven approach for pedestrian intention estimation. In 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), pages 2607–2612. IEEE, 2016.

[2] B. Volz, H. Mielenz, G. Agamennoni, and R. Siegwart. Feature relevance estimation for learning pedestrian behavior at crosswalks. In 2015 IEEE 18th International Conference on Intelligent Transportation Systems, pages 854–860. IEEE, 2015.

[3] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, F.-F. Li, and S. Savarese. Social lstm: Human trajectory prediction in crowded spaces. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pages 961–971, 2016.

[4] T. Fernando, S. Denman, S. Sridharan, and C. Fookes. Soft+ hardwired attention: An lstm framework for human trajectory prediction and abnormal event detection. Neural networks, 108:466–478, 2018.

[5] H. Xue, D. Q. Huynh, and M. Reynolds. Ss-lstm: A hierarchical lstm model for pedestrian trajectory prediction. In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), pages 1186–1194. IEEE, 2018.

[6] Y. Guo, H.Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun. Deep learning for 3d point clouds: A survey. arXiv preprint arXiv:1912.12033, 2019.

[7] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, V. Vasudevan, W. Han, J. Ngiam, H. Zhao, A. Timofeev, S. Ettinger, M. Krivokon, A. Gao, A. Joshi, Y. Zhang, J. Shlens, Z. Chen, and D. Anguelov. Scalability in perception for autonomous driving: Waymo open dataset, 2019.

[8] J. Zhao, Y. Li, H. Xu, and H. Liu. Probabilistic prediction of pedestrian crossing intention using roadside lidar data. IEEE Access, 7:93781–93790, 2019.

[9] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In Conference on Computer Vision and Pattern Recognition (CVPR), 2012.

Motivations for joining mentor sessions

The European research project “SHAPE-IT – Supporting the Interaction of Humans and Automated Vehicles: Preparing for the Environment of Tomorrow” is a Marie Skłodowska-Curie Action (MSCA) Innovative Training Network (ITN) project, which focuses on reliable development of safe and user-centered automated vehicles (AV) operating in urban environments. As a part of the "SHAPE-IT" project, my research topic is "Classifying and Predicting Interactions Between AVs and VRUs Using AI", which aims at using AI to better understand human behaviors. I would like to join mentor sessions to get some valuable feedbacks from other experienced researchers, and to improve my research skills.

As a new PhD student started from this year, I am outlining my research plan for my PhD period:

The entire process about VRU behavior prediction usually includes: (a) obstacle detection (to detect obstacles from raw sensor data), (b) tracking (associate the same obstacles of different timestamp), and (c) prediction (classify VRUs’ intentions and predict future trajectories). We are striving at developing an end-to-end method to directly predict trajectories and intentions of pedestrians instead of solving several separate problems, because the cascade approaches restrict the information that the behavior prediction module has access to, which may lead to sub-optimal solutions. There are several kinds of sensors on a self-driving vehicle, e.g. camera, LiDAR, and radar. We plan to use multi-sensors, as they can provide more information than only one sensor so that can be more accurate. To achieve the goal, we plan to take several steps for the next coming years:

1) Using one sensor, e.g. 3D LiDAR, to design and develop the ML methods to predict the behaviors of pedestrians. The inputs are manually labeled pedestrians’ history trajectories, as well as other information from the raw sensor data, and the outputs are trajectories and intentions of pedestrians in the future.

2) Using one sensor, e.g. 3D LiDAR, to develop the entire end-to-end prediction process, including detection, tracking, and prediction. The inputs are only raw single-sensor data, and the outputs are trajectories and intentions of pedestrians in the future.

3) Fusing multi-sensor information, e.g. 3D LiDAR and 2D cameras, to develop the entire end-to-end prediction process, including detection, tracking, and prediction. The inputs are raw data from multiple sensors, and the outputs are trajectories and intentions of pedestrians in the future.

4) Including not only the information of the pedestrians and environments, but also the interaction with the ego-vehicle and the VRUs.

Above are my current plans. Hope to have the chance to join the mentor sessions. I would be more than grateful to get any feedbacks and comments about my research.