Rasterized Image Based Path Prediction Deep-Learning Model

Chaehyun Lee
Graduate School of Electronic and Electrical Engineering
Kyungpook National University
Daegu, Republic of Korea
hyeu330@knu.ac.kr

Dong Seog Han
Graduate School of Electronic and Electrical Engineering
Kyungpook National University
Daegu, Republic of Korea
dshan@knu.ac.kr

***Abstract*—In this paper, we design an ambient vehicle path prediction model based on deep learning. The most important goal of the autonomous driving system is to ensure the safety of the driver. Therefore, it is essential to predict changes in the surrounding environment of autonomous vehicles. This study generates a rasterized image of a HD(High-Definition) map to consider the actors and road conditions of the driving environment and uses it as an input to the deep learning model. In addition, speed, acceleration, and change of heading rate data are used together as inputs to deep learning models to provide status information on vehicles of interest to infer routes. Through this study, predicting the path of the surrounding vehicle of the autonomous vehicle, it is possible to contribute to securing the safety of the autonomous vehicle by applying the identification of the future driving intention of the surrounding vehicle to the risk evaluation.**

Keywords— autonomous vehicle, path prediction, trajectory forecasting

I. INTRODUCTION

Recently, various studies have been conducted for an autonomous driving system that guarantees the safety of passengers. In particular, research on predicting the driving environment of autonomous vehicles in advance is a field directly related to safety. Driver must understand the various actors around ego vehicle in real time and cope with the changing environment in an instant. According to the Korea Expressway Corporation, the negligence accounts for the largest proportion of traffic accidents in the past five years. This suggests that accident-related damages and costs can be saved if the system recognizes and predicts the driving environment in advance. Therefore, it is essential to infer the path of the surrounding vehicles and reflect it in the driving of the ego vehicle.

In [7] the authors proposed a Recurrent Neural Network (RNN)-based method for long-term predictions of multiple interacting agents given scene context. In [8] authors proposed a social Long Short- Term Memory (LSTM) to model human movement together with social interactions. Authors of [9] used LSTM to predict ball motion in billiards directly from image pixels using a sequence of visual glimpses. In [10] LSTM models were used to classify basketball plays, with overhead raster images as inputs. Similarly, the authors of [11], [12] used overhead rasters and RNNs to track multiple objects in a scene by predicting raster image in a next timestep, unlike our work where per-object trajectories are directly inferred. Due to strict time constraints of a deployed real-time system and the requirement to more easily debug and understand model decisions made on public roads, in this work we used simpler feed-forward CNN architectures for the prediction task is this paper.

II. MATERIALS AND METHODS

A. Training Data

The nuScene dataset[1] was used as the training data of the path prediction model. The data set was sampled at 0.5 seconds, and the path for 6 seconds is the label of the training data. That is, one path constant value has 12 coordinate values and is composed of x and y coordinates. As an image of the training data, an image obtained by rasterizing a cloud point-based HD map is used. This is generated from a HD map containing the surroundings of the vehicle of interest to predict the path. The image has a size of 500x500 pixels and a resolution of 1-pixel per 0.1m. RGB colors are assigned to highlight differences between map element vector layers. As shown in Fig. 1., vehicle of interest are marked in red and other vehicles are marked in yellow. The past movement of the vehicle is rasterized in the same color, but it dims the brightness of the box displaying the vehicle, giving the effect of fading. Roads rasterize the center of the lane, and opposite lanes are distinguished by giving complementary colors to each other.

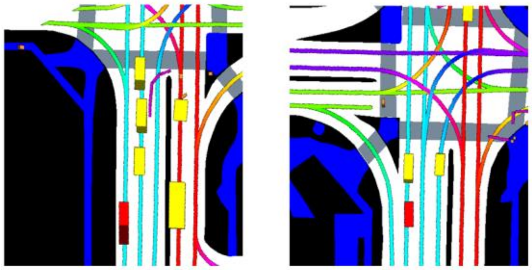

Fig. 1. Example of HD map rasterized images

B. Model architecture

We use the CNN architecture as a backbone network to extract features of input raster images[2]. It is also encoded as a vector containing speed, acceleration and changes of heading rate to input the state of the vehicle of interest, and associated with the output of the backbone network. It generates a path through a fully-connected layer. In order to update the error of the deep learning model, L2 distance between the label of the training data and the model inference value in the x-coordinate and y-coordinate are defined as errors.

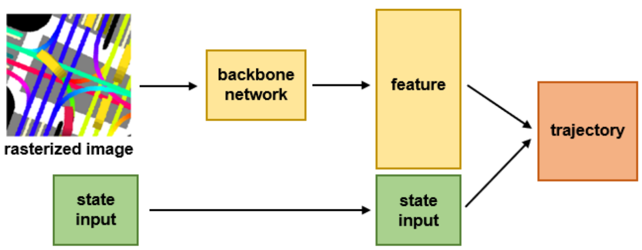

Fig. 2. Model architecture

III. Results

We compared the baseline to several variants of the proposed approach. We trained and evaluated the following base CNNs: ResNet-50 [4], and MobileNet-v2 [3]. Furthermore, to evaluate how varying input complexity affects performance we considered architectures that use raster images without fading.

Table 1 shows that the MobileNet-v2 architecture (e.g., bottleneck layer, residual connectivity, depth-specific convolution) that combines various deep learning ideas shows higher performance than ResNet-50. In addition, additional performance improvements can be obtained when fading is included in the Leicester image. This suggests that fading carries additional information that is not available through the state itself and that raster images and other external inputs can be seamlessly combined through the proposed architecture.

IV. Conclusion

In this study, the path of the vehicle of interest was predicted by extracting the characteristics of the raster image. In addition, it was found that the performance was further improved when the raster image and the vehicle state information were applied to learning together. Route inference reflecting the driving environment was possible by predicting the route based on extracting the characteristics of the vehicle driving environment.

TABLE I. Comparison for competing methods

method	input state	fading	ADE	FDE
ResNet-50	X	X	2.35	13.23
MobileNet-v2	X	X	2.28	12.82
MobileNet-v2	O	X	1.97	11.31
MobileNet-v2	O	O	1.82	11.02

Acknowledgment

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2022-2020-0-01808) supervised by the IITP(Institute of Information & Communications Technology Planning & Evaluation)

References

[1] Caesar, Holger, et al. “nuscenes: A multimodal dataset for autonomous driving." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

[2] Djuric, Nemanja, et al. “Short-term motion prediction of traffic actors for autonomous driving using deep convolutional networks." arXiv preprint arXiv:1808.05819 1.2 (2018): 6.

[3] Sandler, Mark, et al. "Mobilenetv2: Inverted residuals and linear bottlenecks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[4] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[5] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780.

[6] Djuric, Nemanja, et al. "Short-term motion prediction of traffic actors for autonomous driving using deep convolutional networks." arXiv preprint arXiv:1808.05819 1.2 (2018): 6.

[7] N. Lee, W. Choi, P. Vernaza, C. B. Choy, et al., DESIRE: Distant Future Prediction in Dynamic Scenes with Interacting Agents. IEEE, Jul 2017.

[8] A. Alahi, K. Goel, et al., Social LSTM: Human Trajectory Prediction in Crowded Spaces. IEEE, Jun 2016. [Online]. Available: http://dx.doi.org/10.1109/CVPR.2016.110

[9] K. Fragkiadaki, P. Agrawal, et al., “Learning visual predictive models of physics for playing billiards,” in International Conference on Learning Representations (ICLR), 2016.

[10] K.-C. Wang and R. Zemel, “Classifying nba offensive plays using neural networks,” in Proceedings of MIT Sloan Sports Analytics Conference, 2016, pp. 1094-1099.

[11] P. Ondrúška and I. Posner, “Deep tracking: Seeing beyond seeing using recurrent neural networks,” in Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. AAAI Press, 2016, pp. 3361-3367.

[12] P. Ondrúška, J. Dequaire, D. Z. Wang, and I. Posner, “End-to-end tracking and semantic segmentation using recurrent neural networks,” arXiv preprint arXiv:1604.05091, 2016.