Scene-Adaptive Driving Area Prediction based on Automatic Label Acquisition from Driving Information

Takuya Migishima[1], Haruya Kyutoku[2], and Daisuke Deguchi[1]
Yasutomo Kawanishi[1] Ichiro Ide[1] Hiroshi Murase[1]

[1] Nagoya University, Nagoya, Aichi, Japan
migishimat@murase.is.i.nagoya-u.ac.jp
[2] Toyota Technological Institute, Nagoya, Aichi, Japan

Abstract. Technology for autonomous vehicles has attracted much attention for reducing traffic accidents, and the demand for its realization is increasing year-by-year. For safety driving on urban roads by an autonomous vehicle, it is indispensable to predict an appropriate driving path even if various objects exist in the environment. For predicting the appropriate driving path, it is necessary to recognize the surrounding environment. Semantic segmentation is widely studied as one of the surrounding environment recognition methods and has been utilized for drivable area prediction. However, the driver's operation, that is important for predicting the preferred drivable area (scene-adaptive driving area), is not considered in these methods. In addition, it is important to consider the movement of surrounding dynamic objects for predicting the scene-adaptive driving area. In this paper, we propose an automatic label assignment method from actual driving information, and scene-adaptive driving area prediction method using semantic segmentation and Convolutional LSTM (Long Short-Term Memory). Experiments on actual driving information demonstrate that the proposed methods could both acquire the labels automatically and predict the scene-adaptive driving area successfully.

Keywords: Semantic segmentation · Path prediction · Autonomous vehicle.

1 Introduction

Technology for autonomous vehicle has attracted much attention for reducing traffic accidents, and the demand for its realization is increasing year-by-year. Although some automotive manufacturers are already providing autonomous driving functions on expressways, it is still difficult to provide such functions on urban roads due to the diversity of the surrounding environment. To drive safely on urban roads, it is indispensable to predict an appropriate driving path even if various objects exist in the environment. Here, the path can be considered as a trajectory of future vehicle positions, which should be predicted appropriately since it will affect vehicle control during autonomous driving. Since vehicle

where fully connected layers in LSTM are replaced with convolutional layers. It can be applied to the semantic segmentation task for predicting labels of a future frame [4]. The proposed method incorporates ConvLSTM in the prediction of the scene-adaptive driving area to learn the movement of other objects.

Based on the above concept, we propose a method to automatically acquire training data from actual driving information and predict a scene-adaptive driving area. The main contributions of this paper can be summarized as follows:

- Proposal of the concept of "scene-adaptive driving area" considering the necessity of braking.
- Automatic label assignment for training a model to predict a scene-adaptive driving area referring to the own vehicle's speed.
- Prediction of a scene-adaptive driving area that considers the movement of other objects using ConvLSTM.

In the following, Sec. 2 proposes the automatic label assignment method, Sec. 3 describes the construction of the scene-adaptive driving area predictor using ConvLSTM, Sec. 4 reports the evaluation experiments, and Sec. 5 concludes this paper.

2 Automatic label assignment using driving information

Fig. 3. Process flow of the automatic label assignment.

Figure 3 shows the process flow of the automatic label assignment. To output the training data automatically, the proposed automatic label assignment method receives driving trajectory, speed, and in-vehicle camera images of the own vehicle simultaneously. A state-of-the-art semantic segmentation model is used for assigning labels to the training data, such as roads, pedestrians, vehicles, etc. To assign the scene-adaptive driving area label, the actual driving trajectories are projected considering the reduction of the vehicle speed. Then, the labeled images of training data are generated by integrating the scene-adaptive driving area label and other labels. Figure 4 shows an example of an automatically generated ground-truth label image by the proposed method.

Fig. 4. Examples of scene-adaptive driving areas.

2.1 Assigning labels: Pedestrian, vehicle, roads, etc.

To train the semantic segmentation model, it is necessary to prepare images and pixel-wise annotation to each of them. Since drivers usually determine their driving path from the relationship between the own vehicle and its surrounding objects (e.g. pedestrians, vehicles, and roads), it is important to recognize those objects and scene-adaptive driving area simultaneously. Here, the labels of pedestrians, vehicles, and roads can be easily and accurately extracted by applying a state-of-the-art semantic segmentation model.

2.2 Assinging labels: Scene-adaptive driving area

The proposed method assigns the ground-truth label for scene-adaptive driving area from the own vehicle's speed and the actual trajectory of the own vehicle estimated by a LIDAR-based localization method. Here, let X_t be the vehicle center position at time $t \in T$ in the world coordinate system and $F_{t'}$ be a transformation matrix that converts the vehicle position from the world coordinate system to a coordinate system whose origin is $X_{t'}$ (t'-th coordinate system). Based on these notations, the vehicle position $\widetilde{X}_t$ in the t'-th coordinate system at time t is calculated as

$$\widetilde{X}_t = F_{t'} X_t. \tag{1}$$

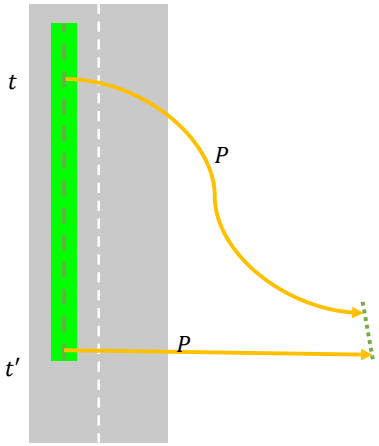


Fig. 5. Projection of vehicle positions onto an image at time t'.

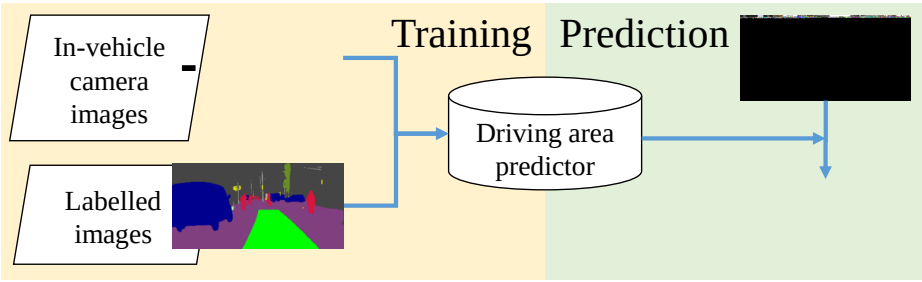


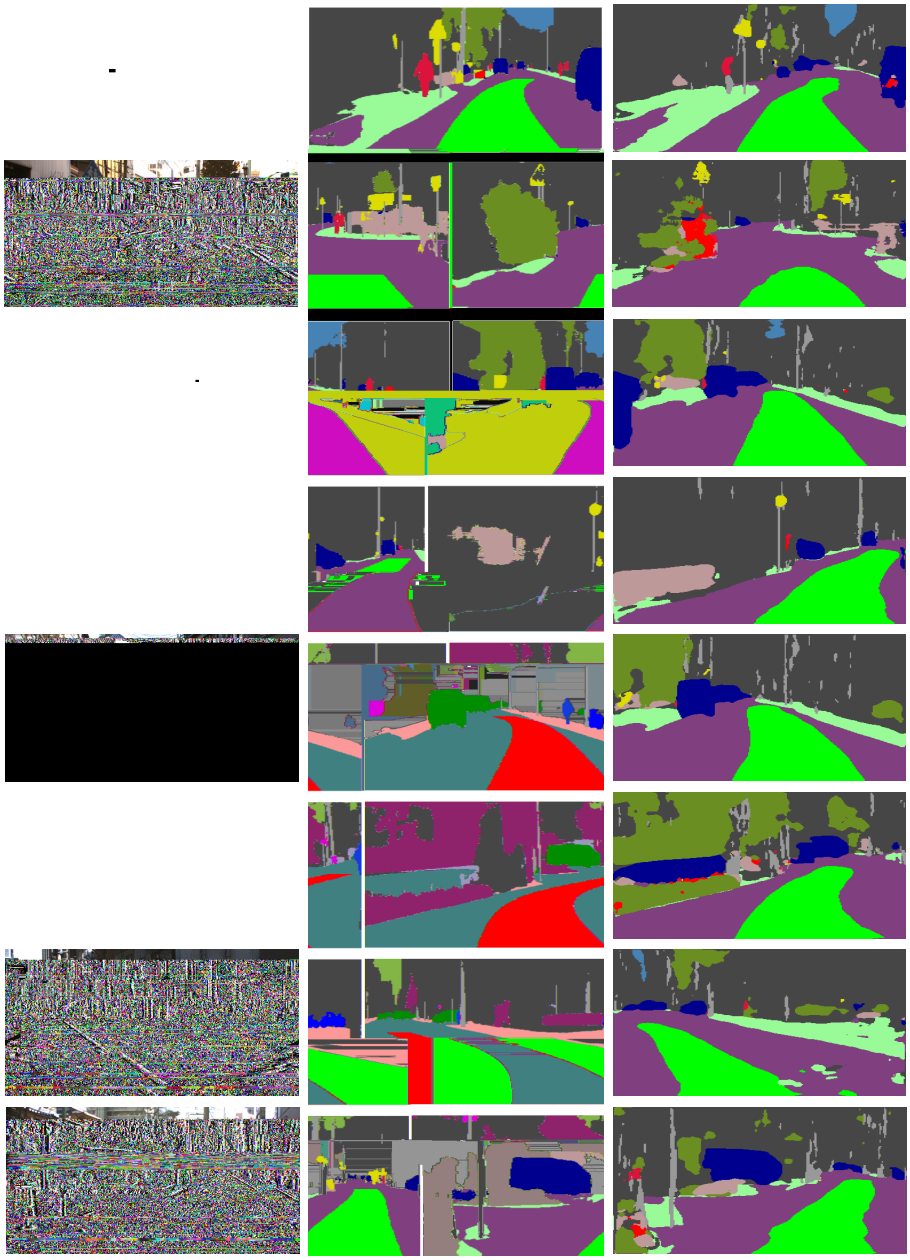
Fig. 6. Overview of the proposed scene-adaptive driving area predictor.

By iterating the above transformation until the following conditions (2) are met, a set of vehicle positions $\widetilde{\mathcal{X}}_{t'} = \{\widetilde{X}_{t'}, \widetilde{X}_{t'+1}, ...\}$ in the t'-th coordinate system is obtained.

$$\|X_t - X_{t'}\|_2 > D \quad \text{or} \quad \alpha_t \leq -2 \quad \text{or} \quad v_t = 0, \tag{2}$$

where α_t is the acceleration [m/s^2] at time t, and v_t is the velocity [m/s] at time t.

Here, since a scene-adaptive driving area far from the own vehicle cannot be observed in the image plane, the proposed method calculates the distance between the vehicle position $X_{t'}$ and each point X_t, and terminates the transformation process if the distance becomes larger than a threshold D. In addition, it is necessary to consider traffic signs and traffic signals requesting the vehicle to stop for safety. Through observation of actual driving data, we found that these situations are observed with a specific driving behavior like the reduction of the vehicle speed. From these points-of-view, the proposed method terminates the transformation when the vehicle's speed decreases.



(a) In-vehicle camera images. (b) Ground-truth labelling. (c) Predicted labelling.

Fig. 10. Examples of prediction results.

5 Conclusion

In this paper, we proposed a method to acquire training labels of scene-adaptive driving area automatically from driving information of the in-vehicle camera image, own vehicle's speed, and location. Here, the training labels are acquired by using the actual trajectory of the own vehicle. According to the own vehicle speed, the proposed method generates the label of the scene-adaptive driving area reflecting the driving context. We also proposed a scene-adaptive driving area predictor based on the semantic segmentation model introducing ConvLSTM trained with the acquired training data.

To evaluate the proposed method, we created 5,650 labels and predicted the scene-adaptive driving area by the proposed method with 65.3%. We also confirmed the effectiveness of considering the movement of the surrounding object.

Future work will include an improvement of scene-adaptive driving area prediction, an enhancement of the network structure, and experiments using a larger dataset including various patterns of the scene-adaptive driving area.

Acknowledgements. Parts of this research were supported by MEXT, Grant-in-Aid for Scientific Research 17H00745, and JST-Mirai Program Grant Number JPMJMI17C6, Japan.

References

1. Barnes, D., Maddern, W., Posner, I.: Find your own way: Weakly-supervised segmentation of path proposals for urban autonomy. In: Proceedings of 2017 IEEE International Conference on Robotics and Automation. pp. 203–210 (2017)
2. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of 2018 European Conference on Computer Vision. pp. 833–851 (2018)
3. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B.: The Cityscapes dataset for semantic urban scene understanding. In: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. pp. 3213–3223 (2016)
4. Nabavi, S., Rochan, M., Wang, Y.: Future semantic segmentation with convolutional LSTM. In: Proceedings of 2018 British Machine Vision Conference. pp. 137-1–137-12 (2018)
5. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Proceedings of 2015 International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 234–241 (2015)
6. Shi, X., Chen, Z., Wang, H., Yeung, D., Wong, W., Woo, W.: Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In: Advances in Neural Information Processing Systems 28. pp. 802–810 (2015)
7. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. pp. 6230–6239 (2017)
8. Zhou, W., Worrall, S., Zyner, A., Nebot, E.M.: Automated process for incorporating drivable path into real-time semantic segmentation. In: Proceedings of 2018 IEEE International Conference on Robotics and Automation. pp. 1–6 (2018)