Urban Scene Segmentation for Autonomous Vehicles

Hsiao-Chen Huang
hsh030@eng.ucsd.edu

Eddie Tseng
edtseng@eng.ucsd.edu

Ping-Chun Chiang
p3chiang@eng.ucsd.edu

Chih-Yen Lin
chl886@eng.ucsd.edu

1 Introduction

Nowadays, the continuing evolution of autonomous vehicles aims to deliver even greater safety benefits. One day, we can handle all kinds of tasks of driving when we don't want to or can't do it ourselves.
Considering the safety factor, the first challenge that stuck out to us is how to tell the difference between people and signs. It's a hugely important, but typically very simple, distinction that you would make reflexively. However, autonomous vehicles can't do this effortlessly. Therefore, we will help vehicles how to classify objects.
Though there have been much work in image segmentation [1], we would like to implement and compare the error rate of several algorithms such as SVM, MLP, and CNN to find out the most suitable classification algorithm for this dataset. [10] [5] [4] [9] [7]

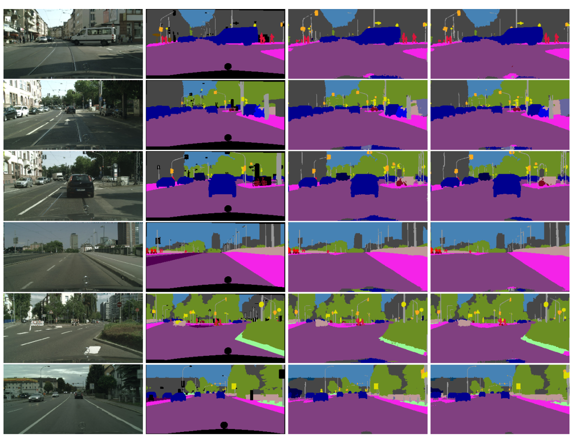

Figure 1: Example of urban scene segmentation [2]

2 Datasets

The model will be trained and evaluated on Kaggle's CVPR 2018 workshop on autonomous driving (WAD) Video Segmentation Challenge[2]. We will predict segmentations of different movable objects appearing in the view of a car camera.
This dataset contains a large number of segmented and original driving images. There are multiple labels, but we only choose seven different classes, which are car, motorcycle, bicycle, pedestrian, truck, bus, and tricycle to evaluate the result. Since this is the latest challenge, we can not find the error rate from different teams to compare. However, the main purpose for our team is to learn and implement what we have learned from this course.

3 Proposed Framework

Recently, there has been a surge of designing networks and try to segment images or videos, e.g.,human segmented natural images[6] and Efficient hierarchical graph-based video segmentation[3]. In this project, we will take the advantages of those networks and observe different responses from tuning the hyperparameters such as number of layers and the size of receptive fields.
We are going to implement Convolutional Neural Networks, which are powerful visual models that yield hierarchies of features [8], to achieve our goal. By using contemporary classification networks, like AlexNet, and VGG net, trained with image-level labels for the task.

4 Implementation and Expected Results

We anticipate that we can help autonomous vehicles to identify objects immediately on the road. In our project, we will find out the pros and cons of networks in several architectures, and we will look for the lowest error rate, which is lower than 10% (expected) by using Python and TensorFlow. To give autonomous vehicles the ability to segment the scene instantly, the first challenge we have to tackle is do scene segmentation precisely in images. Since images segmentation is a crucial and fundamentally concept for video segmentation, it is useful for us to go through this obstacle. Finally, we can obtain our final goal to make autonomous vehicles safe and feasible.

References

[1] J. M. Alvarez, T. Gevers, Y. LeCun, and A. M. Lopez. Road scene segmentation from a single image. In *Proceedings of the 12th European Conference on Computer Vision - Volume Part VII*, ECCV'12, pages 376–389, Berlin, Heidelberg, 2012. Springer-Verlag.

[2] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. *CoRR*, abs/1604.01685, 2016.

[3] M. Grundmann, V. Kwatra, M. Han, and I. Essa. Efficient hierarchical graph-based video segmentation. In *2010 IEEE Computer Soci-*

ety Conference on Computer Vision and Pattern Recognition, pages 2141–2148, June 2010.

[4] A. Khoreva, R. Benenson, J. H. Hosang, M. Hein, and B. Schiele. Weakly supervised semantic labelling and instance segmentation. *CoRR*, abs/1603.07485, 2016.

[5] D. Lin, J. Dai, J. Jia, K. He, and J. Sun. Scribblesup: Scribble-supervised convolutional networks for semantic segmentation. *CoRR*, abs/1604.05144, 2016.

[6] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 416–423 vol.2, 2001.

[7] H. Noh, S. Hong, and B. Han. Learning deconvolution network for semantic segmentation. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ICCV '15, pages 1520–1528, Washington, DC, USA, 2015. IEEE Computer Society.

[8] E. Shelhamer, J. Long, and T. Darrell. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):640–651, April 2017.

[9] S. Tsutsui, T. Kerola, and S. Saito. Distantly supervised road segmentation. *CoRR*, abs/1708.06118, 2017.

[10] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia. Pyramid scene parsing network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6230–6239, July 2017.