Semantic Segmentation for Autonomous Vehicles Project Proposal

Samuel Cowin
Department of Electrical and Computer Engineering
UC San Diego
scowin@ucsd.edu

Jianghong Wan
Department of Electrical and Computer Engineering
UC San Diego
j6wan@ucsd.edu

Vivaswat Suresh
Department of Mathematics
UC San Diego
visuresh@ucsd.edu

One of the most challenging problems in the area of computer vision has been sematic segmentation, which is to segment images based on objects. With the rise in Deep Learning and the success of Autonomous Vehicles both in utilizing this technology and commercially, investigations into this field for furthering these developments are crucial. One realm that has garnered much attention is the application of Deep Learning for Semantic Segmentation of scenes for understanding. Various approaches has been taken in the research community. For example, Fully Convolutional Network(FCN) by Long et al[1], and Deep Convolutional Networks with unsampled filters and atrous spatial pyramid pooling to explicitly control the resolution of the computed features, and robustly segment objects at multiple scales by Chen et al[2].

Within this study, we will examine various approaches to Semantic Segmentation [3] drawing from implementations completed on the popular Cityscapes dataset. To do this, we start with popular CNN architectures (i.e., VGG, AlexNet, ResNet), tweaking and modifying them to change the output architecture from classification to do per-pixel identification as outlined in FCN by Long et al [1]. This will be done on the Kaggle version of the dataset taken from Berkeley [4], and, time permitting, extended to larger versions of the data for further understanding [5]. The Kaggle version of the dataset [1] consists of 2975 training images and 500 validation images taken from vehicles driven in Germany, all of which are 256x512 pixels, and each image contains the original image on the left half, and the segmented output(with labels) on the right.

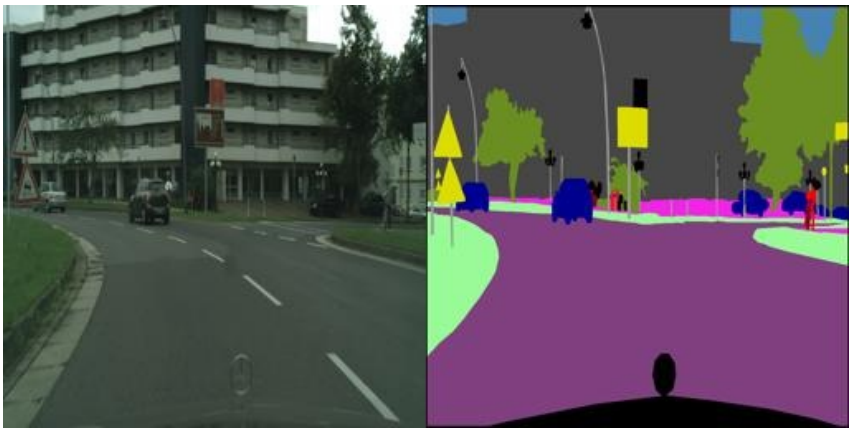

Figure 1: An Example of the dataset

[1] https://www.kaggle.com/dansbecker/cityscapes-image-pairs

Project for ECE 228/SIOC 228: Machine Learning for Physical Applications, Spring 2021, Group 10.

References

[1] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[2] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[3] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," *arXiv preprint arXiv:1704.06857*, 2017.

[4] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[5] 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," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213–3223.