Semantic Segmentation for Autonomous Vehicles Progress Report

Group 10

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

Semantic segmentation has a multitude of applications, specifically in medical imaging and autonomous driving. In order to increase understanding of how best to implement these models, this project will survey different values for hyperparameters and network architectures.

The dataset being studied is from a Kaggle competition [1], which contains image pairs of street scenes taken in Germany, with original image on the left and ground truth on the right.

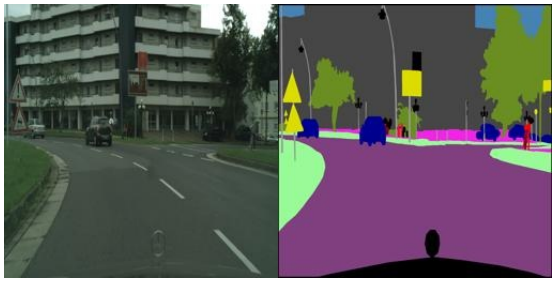

Figure 1: Dataset Example

Feature extraction is not going to be used as the models used are CNNs. Feature extraction is done manually within the architecture using convolutional layers. The regularization that would be handled by feature engineering will be done internal to the network as well. Examples include dropout, batch normalization, and pooling to name a few.

The models that are being used are U-Nets [1] and FCNs [2]. To improve upon baseline implementations, the initial U-Net architecture will be changed to that similar to a U-Net++ [3] - using more skip connections. Additionally, a pixel-to-pixel loss function will be used in place of MSE. Modifications will also be done for the FCN to tune the performance.

Current progress is that the baseline U-Net implementation is completed [2] and the results for 5 epochs of training are included below in Figure 2. The average IoU score across the first 5 images in the validation set of the U-Net is 0.95. Further experimentation is ongoing, but limited due to the provided Datahub server computational capacity.

[1] https://www.kaggle.com/dansbecker/cityscapes-image-pairs
[2] https://github.com/samuel-cowin/ECE228_Group10_Project

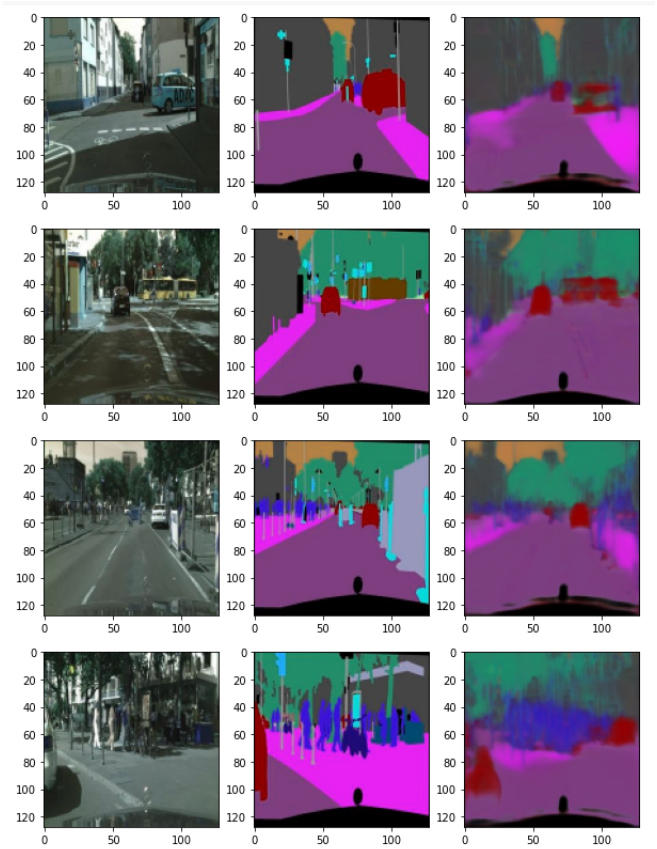

Figure 2: Output from U-Net

Some progress has been made using the Fully-Convolutional-Network model here [3] as a starting point. Due to limited computational resources, the model was only able to train two validation images. Both output scores were over 0.974 in IoU, with an average of 0.977.

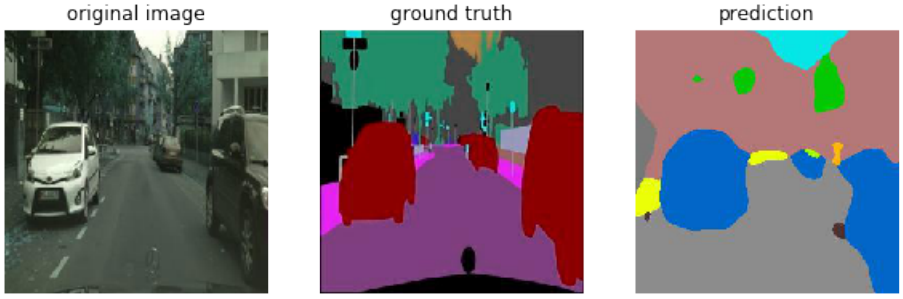

Figure 3: FCN output

[3] https://github.com/hellochick/semantic-segmentation-tensorflow

An overview of the group member contributions are included here:

Member	Current Contributions	To Do
Vivaswat	Edited baseline U-Net to find optimal hyperparameters and training epochs. Read U-Net++ paper for implementation.	U-Net++ architecture and tuning; Presentation
Jianghong	Tested FCN [4] with the environment under computational resource constraint through 1 image trainer	Improving referenced FCN; Presentation
Sam	Created baseline U-Net model and training script	U-Net++ architecture and tuning; Presentation

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.

[2] 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.

[3] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*, Springer, 2018, pp. 3–11.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.