Project Proposal

MIS 382N - Professor Joydeep Ghosh, UT Austin, Fall 2021

Section B Group 2 - Namit Agrawal, Christian Alfonso, Brett Bartol, Vishal Gupta, Troy Richard

Problem Outline:

In this project, our group will attempt to segment pictures into different classes. The goal of the network being utilized will be to find edges and curves within the images, and utilize these edges to break the image into different segments.

Business Applications:

This project will allow for further development in the autonomous vehicle industry by utilizing Machine Learning to segment real-life images of cities, creating an opportunity for self-driving automobiles to increase accuracy in avoidance of objects typically found on and around roads. Image segmentation, also known as scene understanding, is an imperative component of the autonomous vehicle industry, and given the increasing popularity of products offering this technology, the need to continually improve accuracy is ever-growing. The risks associated with poor-performing scene understanding are extremely high. Potential physical dangers can be introduced to the driver of an autonomous vehicle as well as other drivers or pedestrians around

the vehicle. Further, damage to the vehicle or property in its vicinity is another risk that increases when image segmentation is not properly developed. Lastly, the improvement of scene understanding is vital to reach wide-spread acceptance among consumers of autonomous vehicles. As seen in recent coverage and discussions surrounding the few traffic accidents associated with Tesla's Autopilot feature, consumer trust in this technology is not yet high enough to allow image segmentation to be adopted as an industry standard.

Data Available:

Our dataset consists of roughly 30 classes:

Group	Classes
flat	road · sidewalk · parking+ · rail track+
human	person* · rider*
vehicle	car* · truck* · bus* · on rails* · motorcycle* · bicycle* · caravan*+ · trailer*+
construction	building · wall · fence · guard rail+ · bridge+ · tunnel+
object	pole · pole group+ · traffic sign · traffic light
nature	vegetation · terrain
sky	sky
void	ground+ · dynamic+ · static+

Our dataset consists of 5000 images with 3475 training images and 1525 test images. The data was captured in 50 cities over the span of several months during the daytime and in good weather conditions. The images are screenshots that were manually selected from video taken in various German towns and cities, and each picture includes a large number of dynamic objects, a variety of scenery, and different backgrounds. In addition to this, we are also provided with json files that specify the labels within the images along with coordinates of those labels within the image.

Approach:

To start, we will need to shrink the image sizes because the images in the dataset are unnecessarily large. Then we will start working on our model. To accomplish this, we will be building a convolutional neural network using Python and the PyTorch package. The neural network will have convolutional layers that use both max pooling and dilated convolution in order to find these edges and boundaries. A lot of these models use a fully convolutional network. The goal after finding these edges will be to assign each pixel to a segmentation group. The model will be trained on our City Scapes data set, and then will be tested on real data.

List of References:

https://paperswithcode.com/dataset/ade20k

https://www.cityscapes-dataset.com/dataset-overview/

https://divamgupta.com/image-segmentation/2019/06/06/deep-learning-semantic-segmentation-keras.html

https://neptune.ai/blog/image-segmentation-in-2020

https://nanonets.com/blog/semantic-image-segmentation-2020/

https://arxiv.org/pdf/1704.06857.pdf