Autonomous driving is undergoing intensive study in various fields of research. The number of research topics directly related to the success of an effective autonomous vehicle is proliferating. Online assessment of surroundings, one of the most important actions for an autonomous vehicle, has had a large number of research project attempting to tackle different aspects of driving; for instance, there is research geared toward control of the vehicle while other research is focused on understanding the environment. DeepROAD intends to leverage deep learning in a manner conducive to perception and understanding of surroundings in real-time. This research employs deep learning for detection and segmentation of surroundings for decision making as well as compression, allowing for smaller networks and quicker inference times.

Working closely with an automotive engineering school, we have been tasked with developing a system for use in an autonomous vehicle. This vehicle will perform a subset of the tasks needed for an effective autonomous vehicle as a proof of concept in a testing environment. Some of the goals of the project include: remaining in the correct lane while the vehicle is in motion, park at a specified location and pick up a passenger, transport that passenger to a different destination, and making sure to follow all traffic laws.

To accomplish the perception task to achieve this type of vehicle motion, we employ deep learning as our processing medium. Deep learning allows the system to learn all the objects that are necessary for effective motion including roadways, sidewalks, lane markings, traffic signs and signals, pedestrians, and of course other vehicles. Many different types of deep learning can be used for understanding vehicle surroundings including bounding box object detection, pixel-based image segmentation, and depth calculations.

Faster-RCNN [1] is a deep learning convolutional network where the convolutional layers produce a feature map. This feature map is then combined with region proposals for the image to provide bounding box coordinates and probability values for each region-of-interest in the image. SegNet [2] or some other variation of pixel-based segmentation networks classify each pixel in an image rather than trying to look for larger objects. This type of network removes all fully connected layers and replaces them with deconvolutional layer (mirrored convolutional layers). The advantage of this type of method is that larger objects that are differently shaped (such as roadways, sidewalks, and road markings) are easier to decipher. Depth data can also be added to this segmentation/detection allowing the subject to understand distance information of detected objects. Using multimodal data (e.g. RGB and depth simultaneously) has the potential to improve inference performance by make the network more robust to input representation.

All of these concepts provide a starting point for a deep learning suite for autonomous driving. However, one major flaw or downfall of these frameworks and deep learning techniques is the inherent longevity of the processing time. To process an image (or video frame) with the accuracy necessary for classification, the deep learning network has to be relatively large. However, as the network gets larger to increase accuracy, the processing time increases. This is counterintuitive for real-time situations (like autonomous driving) as processing time needs to be as fast as possible to make split-second decisions. To combat this problem, we have begun to look into network compression to reduce the size of the network while maintaining accuracy. This can be done by pruning, quantization, and Huffman coding. These techniques can reduce some network by ~35x, thereby leading to quicker inference times.

Following the three methods illustrated by this work (object detection, segmentation, and compression), we plan to utilize a combination of them to build an enhanced suite for autonomous driving which will be useful as a real-time system for perception.