CS766 Project Proposal
Computer Vision in Self-driving Car

Chao Hsiang Chung (cchung49), Shang-Yen Yeh (syeh6)

0.1 Briefly explain what problem you are trying to solve

We like to study how Computer Vision can be used in object detection system in autonomous vehicle. This can be split into two division, the first thing is that we need to detect where is passable lane or area. Second, we also need to know is there any objects that will affect the driving decision. For example, other vehicles, pedestrians or any other obstacles. Our goal is to let computer distinguishes the scene as human as possible.

0.2 Why is this problem important? Why are you interested in it

As transportation become a essential part of daily life, people now are empowered to move in high speed unprecedentedly. However, people could make mistakes, such as inappropriate driver behavior or wrong decision-making. In our country, one of the serious reason of car accident is drunk- driving. With self-driving car, people can rely on the computer and reduce all these human factors accident. Driving after drinking will no longer be a life threatening behavior with the technology. To make self-driving car feasible, the fundamental thing is to let the self-driving car understand the surroundings, and this is based on object detection system with Computer Vision.

0.3 What is the current state-of-the-art?

- **Machine Learning approaches:** First define features using SIFT, HOG, etc..., then using technique like SVM to classify objects.

– **Deep Learning approaches:** Traditional way to detect objects is by region proposals (i.e., R-CNN, Fast R-CNN, Faster R-CNN, etc.). There is also some using You Only Look Once (YOLO). Developed by these approaches, there are some models available such as VGG, ResNet.

0.4 Are you planning on re-implementing an existing solution, or propose a new approach

We plan to re-implement the existing object recognition methods (e.g., convolutional neural networks, etc.) as there are a few of techniques available, and this is a better way to get started and familiarise ourselves with object detection. Additionally, we will apply some existing models such as VGG, ResNet, GoogLeNet to ensemble the result in order to reduce the bias from any certain model. Once we are able to identify objects and achieve certain accuracy rate and if time allows, we would also like to experiment some new approaches.

0.5 If you are proposing your own approach, why do you think existing approaches cannot adequately solve this problem? Why do you think you solution will work better?

If time allowed, we plan to implement several solutions together in order to get a more objective result(e.g., YOLO, Rol-Pooling, etc.). Since every kind of approaches might have its own bias, it could generate its result in a certain perspective. The better way is to combine every kind of method and get a fairer output.

Furthermore, the result we are having now is the current position of objects, but the objects in the real world will move from their current place. We like to try if we can predict our future position and the possible position for the objects in the scene to know the potential danger zone.

0.6 How will you evaluate the performance of your solution? What results and comparisons are you eventually planning to show? Include a time-line that you would like to follow.

- **Evaluation** First we will have some labeled data (pictures). After detecting objects in the picture, we will cross validate if the number and types of objects match and the centres of objects are within a certain error value. For instance, in the labeled picture, we have 2 trees and 3 cars. The prediction result also has to be 2 trees and 3 cars and the corresponding centres have to be the same. Also, we will draw lines at the edge of lane.

- **Result** We will either split a video to several frames or use a few of pictures as data. We will identify barriers such as human, trees, cars, etc. by plotting a rectangle on the corresponding object and show the label (i.e., tree). Also we will show the direction or road where is accessible and safe.

- **Time-line** Project Mid-Term Report (Due: March 27):
  Able to detect objects in the frame and draw rectangle to indicate the position of all objects.

  Final Project Presentations:
  Draw the lines along lane and recognize the object type in the rectangle.

0.7 Reference:

- **"ImageNet Classification with Deep Convolutional Neural Networks", Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton**

- **"You Only Look Once: Unified, Real-Time Object Detection", J. Redmon, S. Divvala, R. Girshick, A. Farhadi**

- **"Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning", Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, Alexander A Alemi**