Springboard Machine Learning Career Track
Capstone Project Proposal - Aidan Dunlop

Traffic Light Recognition

With the potential to improve road safety, reduce congestion and free up time for human drivers, the impact that the self driving car might have is huge, and needs to be taken seriously. McKinsey & Co. consulting firm reported that widespread embrace of self-driving vehicles could eliminate 90% of all auto accidents in the U.S [1].

In order for self driving cars to become a reality, the car needs to be able to identify it's environment in order to navigate around it safely. One clear example of this is the detection and classification of traffic light signals. The car will need to obey the signals, and act accordingly based on the state of the traffic light. A system will have to be developed that, based on the cameras or sensors on the car, determines whether there are any traffic lights near by and what to do when approaching them.

This problem encompasses a huge range of typical Computer Vision problems. To begin with, the system will need to detect whether there are any traffic lights in close proximity to the car. This is tolerable at a simple junction, but a more complex junction with a number of different traffic lights makes the problem harder, as it would have to determine the correct traffic light to follow. Also, it's highly likely that occlusion could occur - a traffic light may be out of sight, or partially hidden, making it difficult to determine whether it is a traffic light, and if it is, what its current state is. The system would also have to cope with different light conditions - e.g. the camera will produce deviating data from a traffic light at night versus a traffic light in bright sunshine.

This is a well known and essentially solved problem, and many car manufacturers have started to produce systems that solve this problem. The LISA data set [2] is an industry standard traffic light dataset, and provides four day-time and two night-time sequences primarily used for testing, providing 23 minutes and 25 seconds of driving in Pacific Beach and La Jolla, San Diego. The training clips consists of 13 day-time clips and 5 night-time clips. The location of the traffic light in the frame and the state of the traffic light is given for each frame.

There are a number of different approaches one could take in attempting to solve this problem. In the past, it has been solved with traditional Computer Vision and Image Processing techniques, such as edge finding, spotlight detection and adaptive templates. More recently, Machine Learning algorithms have been utilized to develop classifiers which determine the location and state of a traffic light.

In this project, I'm going to attempt to use Deep Learning techniques, such as a Convolutional Neural Network. This is a supervised classification problem - first I will need to determine whether or not a traffic light is in a frame, and then determine what state a traffic light is in within that frame. The dataset has labelled data which can be used to learn a model of traffic light images: each frame in the dataset has the pixel coordinates of each visible traffic light, and the state of the traffic light (e.g. "RED").

To begin with, I'd focus on building a basic machine learning pipeline, which takes in a single image, and outputs whether or not there is a traffic light in the image. I'll then extend this to output the pixel coordinates of all the traffic lights in a frame, and then output the state of the traffic light if one is found. With this pipeline, I can then extend to allowing a video as input, and running the pipeline on each frame of the video.

The final deliverable will be an API that a user can upload an image or video to, and the response will be the location of any traffic lights in the frame(s) and their state. If I have time, I'll deploy a simple web app which allows you to upload an image or video to. The UI will show the bounding box of the traffic light overlaid on top of the original image, and showing its state.
Given more time, I'll also attempt to use traditional image processing techniques to solve this problem, and then give the option to select which technique to use, allowing a comparison between the two techniques.

Computational resources
Deep learning techniques have been shown to work best on GPUs, so I'll attempt to utilise GPUs with AWS. Amazon EC2 P3 instances deliver up to 8 NVIDIA® V100 Tensor Core GPUs and are recommended for deep learning applications. In my early research stages I'll determine whether or not to break down the data set into smaller pieces, to simplify the process and reduce running costs, which will affect how much memory and processing power will be needed.

References

1. "Self-Driving Cars Could Cut Down on Accidents, Study Says", https://www.wsj.com/articles/self-driving-cars-could-cut-down-on-accidents-study-says-1425567905
2. LISA traffic light dataset, https://www.kaggle.com/mbornoe/lisa-traffic-light-dataset