Report: Red Round Sign Detection

Part 3: Envision:

- **Will you change the NN you choose if you were targeting real time on an embedded device? Or if you had no time constraint during prediction?**

A: Traffic sign recognition is a very important computer vision task for a number of real-world applications such as intelligent transportation surveillance and analysis. A key challenge with deep neural networks is the very high computational and memory overhead, that prevents its deployment for blazing fast performance. Therefore, if we are targeting real time embedded devices, the previous simple CNN based neural net will have very poor performance during online inference, and is not an ideal choice for an embedded device. In this case, we have to explore various model compression techniques such Deep Compression, with pruning less important connections, trained quantization of the weights and biases to lower bits and use Huffman encoding to encode them. All these approaches combined has proven to reduce the model size drastically, with very little loss of performance. This kind of approach will also facilitate fitting the model into on chip SRAM cache, rather than off chip DRAM.

If we have no time constraint during inference, we can start by gathering more data on traffic signs, or use massive datasets such as CURE-TSR. Then we can design deep convolutional models such Inception network, Dense Net or Residual network with skip connections as the feature extraction pipeline, and use it with Region Proposal Networks (RPN) to generate region proposals that is used to predict bounding boxes over the image. This type of network will have very high computational overhead and memory requirement, but with sufficient data, can provide promising results.

- **If you were given the opportunity, what machine learning algorithm / method will you want to try?**

A: I would want to try Instance segmentation, which is basically pixel level classification of the entire image. Unlike object detection, that predicts bounding boxes, Instance segmentation provides a better alternative since it detects and delineates each distinct object of interest appearing in an image. In this context, I would like to explore Mask RCNN, which is one of the state-of-art paper for this task, and builds upon previous works such as Faster RCNN. The model uses a RPN to propose ROIs and use those ROIs to perform bounding box regression, classification and mask estimation simultaneously.

- **How does your result compare to the state of art in sign or object detection?**

A: We got a test accuracy of **0.9588**, on the reformulated binary classification task. However, I could not find any open source paper that reports application of deep neural networks for object detection or classification on the GTSDB dataset. Previous papers in the past 3 years have used hand crafted feature extraction algorithms such as ACF, Spatially pooled LBP or HOG, and reported 100% AUC performance on the prohibitory classes (red round signs), with various feature combinations.

- **What are the challenges you expect to encounter if you had to do classification along with detection? For example, if you had to detect and classify Chinese characters on traffic signs, what are the issues you expect to have and how do you plan to overcome them?**

A: Along with traffic sign detection, identifying and classifying characters on the traffic sign is indeed challenging, when there are hundreds of classes. We can design a data processing pipeline, that will generate the saliency map from each traffic sign in the original image, based on the idea of Image segmentation. The saliency map is a topographic representation of saliency which refers to visually dominant locations. It will contain information in terms of superpixels that outline the contours of the Chinese characters embedded on each traffic sign. The extracted saliency maps now represent new features for the classification head that will classify the inputs to hundreds of Chinese character. This head can be coupled with a standard CNN-RPN based detector head, that will operate on the generated feature maps of the whole traffic signs and use ROI pooling to predict bounding boxes. Both the heads can be trained end to end in parallel. The major issues and challenges will be identifying techniques to generate good saliency maps from the traffic signs and combining the classification and detection head and parallel training.