

Yihao Qian

Team A: Aware

Teammates: Amit Agarwal Harry Golash

Menghan Zhang Zihao (Theo) Zhang

ILR06

February.2, 2016

Individual Progress

During the last week, I was in charge of researching how to improve the performance of object detection algorithm and transplant stereo vision program from Matlab to OpenCV

Objection Detection

Structure:

The algorithm we are going to use is called Single Shot MultiBox Detector (SSD)

Single Shot MultiBox Detector (SSD) is a convolutional neural network. The structure of this network are as follows:

1. Feature extraction network: VGG-16-Net. This is the common convolutional neural network that is used to extract the features. The convolutional neural network is the combination of the convolutional layers and pooling layers. The convolutional layers are used to extract the features. The pooling layers are used to decrease the size of the image. For example, the input of the pooling layer is a matrix [3,4;6,7], this is a 2*2 matrix, if we use the max pooling strategy, the output of the pooling layer is 7, since 7 is the biggest number in the matrix. The pooling layer is quite useful. It can present the special relationship between the features extracted from the convolutional layers, and change the features to a high level features. VGG-16-Net is a pre-trained network.
2. Region Proposal network: Unlike Faster R-CNN, which use the Region Proposal layer in order to generate the region proposal. SSD use a series of decreasing convolutional layer and regression boxes in order to generate the region proposal. You may find the structure of the SSD in Figure 1.
3. FC (Fully connected) layer. Fully connected layer takes both the input from the RPN and the feature map, outputs the bounding box, category and the possibility.
4. The last layer is called the Non-Maximum Suppression layer. The input of this layer is all the bounding box and the corresponding labels. However, for one object (human), there may be several bounding box around it, which makes the picture very untidy. Non-Maximum Suppression layer (NMS) is used to compute the most appropriate size for the object.

Improvement:

Based on our experiment, the object detection accuracy of SSD algorithm is around 60% and this algorithm could not detect small object in the image.

Based on my research, we could improve the performance of SSD using the following method:

1. Change feature extraction network for SSD: (Result: Improves accuracy a little bit) One of the main reason that may affect accuracy of object detection algorithm is the quality of feature provided by feature extraction network. However, changing the feature extraction network from VGG-16 to Inception-ResnetV-2 shows little improvement, which is the same as the author of SSD said. You may find the result in Figure 1.

Based on my research, we could improve the accuracy of object detection system using other method.

1. Use Faster-R-CNN algorithm with limited region proposals. (Result: Greatly improves accuracy, FPS decreases to 5 FPS). According to the experiment, if we combined Faster-R-CNN with Inception-ResnetV-2 feature extractor, the performance would increase a lot. You may find the result in Figure 1. The performance on detecting small object also increases a lot. One of the drawback is that Faster-R-CNN has low FPS. Which is due to the fact that the network generates a lot of region proposals. After the network generates region proposals, each ROI region needs to be sent to a fully connected neural network to classify which classes the region belongs to. This process would take a lot of time. If we could reduce the number of region proposals, the accuracy may drop 1-2 percent, however, it may greatly improve FPS. You may find the result in Figure 2. According to the result, Faster R-CNN with 50 region proposals has the best accuracy and speed trade-off.

Stereo Vision:

I am transplanting our stereo vision system from Matlab to C++. The stereo vision system algorithm I am going to use is called SGBM. Using a high performance CPU (I7), it is said that the algorithm could reach 10 FPS.

Challenges:

Stereo Vision :

Transplanting stereo vision from Matlab to C++ is not easy. There are two major different between those two program.

- (1) Matlab version of stereo vision only computes the depth in the region of interest. The C++ version computes the depth for whole image. The reason behind this is that apart from computing the moving object, we would also want the depth information of static object (Wall, Bridge, etc.)
- (2) The parameters for camera calibration are different in Matlab and C++. Due to the low performance of C++ and OPENCV camera calibrator. I need to use matlab to calibrate the camera first and try to re-compute the corresponding value that is used in C++ and OpenCV.

Teamwork:

Zihao built subscriber in PolySync to extract information from radar data. Amit read research papers about object tracking. Harry familiarized PolySync and did some work of radar focusing. According to what Menghan said, she installed CUDA on her computer.

Plans:

In the next week. We will finish transplanting our system from Matlab to C++. And start building our object tracking system.

Figures:

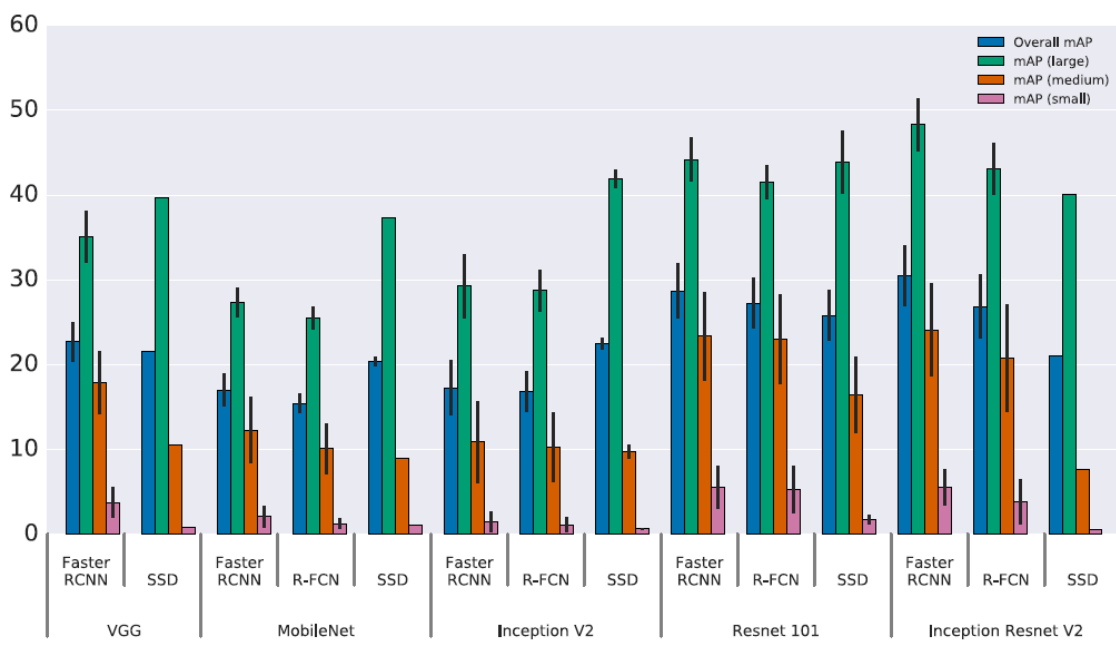


Figure 1. Object detection algorithm with different feature extractors

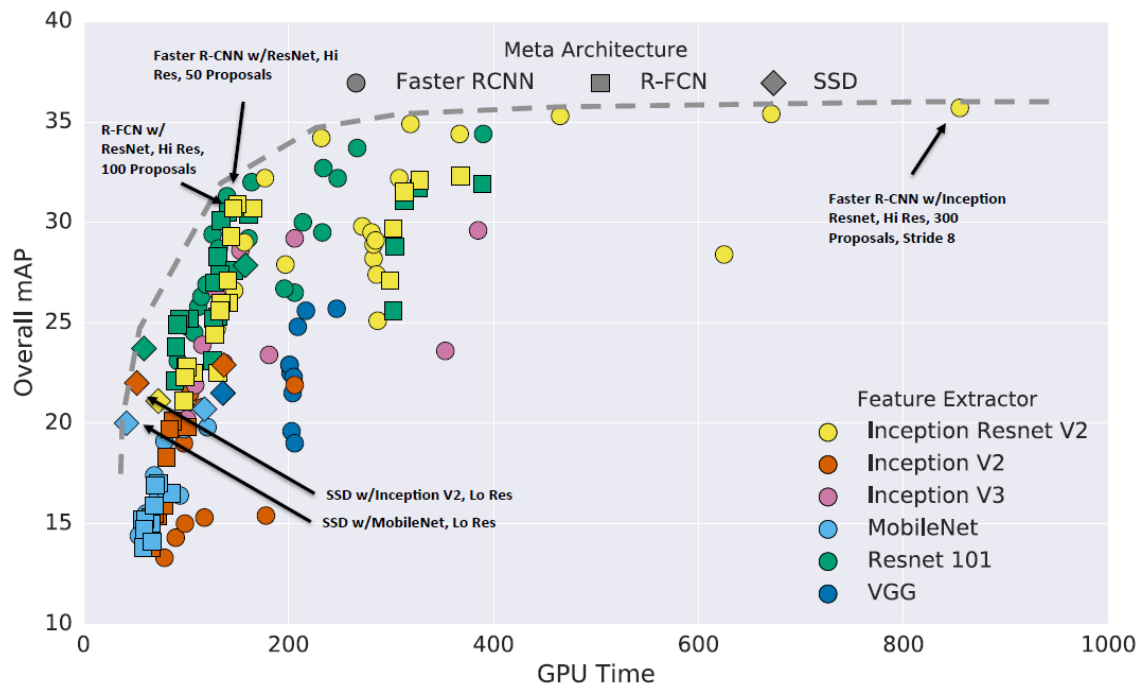


Figure 2. Speed & Accuracy Tradeoff