CS 5100: Project Proposal
Authors: Xueyi Fan and Yuchen Xiao

Recognition of Object and Face

Object recognition technique has been widely applied in a variety of areas including robotics, surveillance, security system, self-driving car and human-computer interaction. Basically, it falls into object detection problem, which is to localize objects in an image and a video or during real-time recording, and classification problem, which is to predict the label of the target object in the predefined labels. Face recognition is one typical application of object recognition algorithms, meanwhile, the diversity and complexity of human face's expressions and features make this problem more interesting and challenging.

In the past several decades, PCA/SVM is the mainly algorithm applied on object/face recognition. Along with the research in neural network, conventional neural network (CNN) has been occupying the dominant position and significantly improving the speed and accuracy of detection and classification. In this project, we will learn and apply two advanced algorithms to solve the object and face recognition problem. One is to combine distinct datasets and train the object detector on both detection and classification data, which is based on YOLO9000 real-time framework that is faster than other detection systems across a variety of detection datasets and image sizes [1]. The other is a discriminative feature learning approach for deep face recognition based on the CNN model provided by Caffe library, which proposes a center loss function to improve the deep-learning features model and enhance the robustness of face recognition over different datasets, e.g., MegaFace Challenge, Labeled Faces in the Wild(LFW) and YouTube Faces(YTF)[2].

All background materials including algorithms and related libraries were collected by both of us. There are two parts of our project. Yunchen is mainly responsible for object recognition part and Xueyi is responsible for the work of face recognition. We will first install the two libraries and using the corresponding datasets to train the deep-learning models. During experiments, we will apply a 3D sensor to provide a real-time video stream, and simultaneously the recognizer will detect faces/objects, predict the specific types and show the results in the sensor's view.

Reference:

[1] Redmon, Joseph, and Ali Farhadi. "YOLO9000: Better, Faster, Stronger." arXiv preprint arXiv:1612.08242 (2016).
[2] Wen, Yandong, et al. "A discriminative feature learning approach for deep face recognition." European Conference on Computer Vision. Springer International Publishing, 2016.
[3] Caffe: http://caffe.berkeleyvision.org/
[4] YOLO: https://pjreddie.com/darknet/yolo/