

# CS766 Project Proposal

## Topic: Object Detection via YOLO approach

Members: Ya-Chun Yang, Hui-Shun Hung

### Overview

With the development of autonomous vehicles, smart video surveillance, robotics and the facial detection in real-world applications, accurate and fast object detection systems are rising in demand. We need a system that can recognize and classify every object in an image, as well as localize a bounding box around it. The ideal detection system thus enable assistive devices to convey real-time scene information to human users, allowing computers to drive cars in any environment, and applying in responsive robotic systems.

### Problem

Real-time object detection is essential for self-driving cars and traffic monitoring. Recent detection approach, R-CNN (Region-Based Convolutional Neural Network), use region proposal methods to first generate potential bounding boxes in an image and then run a classifier on these proposed boxes, followed by post-processing to refine the bounding box. These complex pipelines are slow and hard to optimize [1]. Although the Fast and Faster R-CNN were shown to provide speed and accuracy improvements over R-CNN, both models still fall short of real-time performance [2, 3].

### Solution

Rather than repurposing classifiers to perform object detection, YOLO (You Only Look Once) reframe object detection as a single regression problem to spatially separated bounding boxes and associated class probabilities (fig. 1)[4]. In the YOLO approach, only a single convolutional network was trained to simultaneously predict multiple bounding boxes and class probabilities for those boxes. Thus, it not only is extremely fast (processes 155 frames per second), but can be optimized end-to-end directly on detection performance to achieve high mean average precision (mAP). Previous analysis indicated that YOLO model outperforms all other detection methods by low background errors and wide margin when generalizing detection from artwork [4].

In this project, we will follow and re-implement the state-of-the-art YOLO approach, and evaluate the model with existing datasets. We will further run the model with streaming video in real-time. Next, we will measure the YOLO detection performance on a various range of testing parameters to come up a general strategy for parameter design in YOLO approach. The work is summarized as follows:

- 1) **Re-implement the YOLO algorithm, train and build an object detection pipeline**  
We will implement the YOLO convolution network [4] and load the pre-trained weights, followed by further training using VOC 2007 and VOC 2012 train/val dataset.
- 2) **Test our YOLO model on VOC 2007 test data set**  
We will evaluate our model by measuring mAP using VOC 2007 test dataset.
- 3) **Run YOLO on streaming video in real-time**

To further test YOLO in computer vision applications, we will connect YOLO to a webcam and measure its real-time performance, including the time to fetch images from the camera and display the detections.

#### 4) **Examine the test parameters**

We will measure the YOLO detection performance on a various range of testing parameters, threshold and iou-threshold, to come up a general strategy for parameter design in YOLO approach.

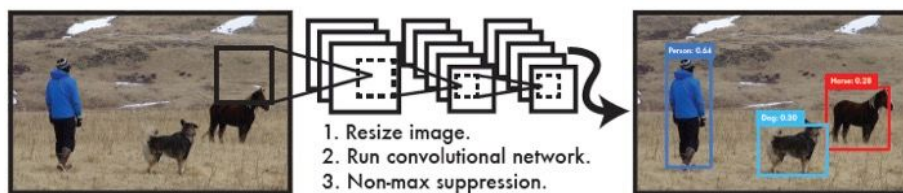


Fig. 1.The YOLO detection system

## Milestones

Time	Goal
2/14	Proposal due
2/18	Papar review and understanding hte concept of YOLO
3/4	Finish implementing YOLO network architechture
3/18	Finish training and tuning of YOLO network
3/31	Test YOLO model on VOC 2007 test datasets
4/2	Midterm report due
4/8	Run YOLO on streaming video in real-time
4/18	Examine the test parameters
4/30	Project presentation
5/7	Project webpage due

## References

- [1] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *In Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, page 580-587.
- [2] R. B. Girshick. Fast R-CNN. *CoRR*, abs.1504.08083, 2015.
- [3] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *arXiv preprint arXiv: 1506.01497*, 2016.
- [4] J. Redmon, S. Divvala, R. Girshick, A.Farhadi. You only look once: unified, real-time object detection. *arXiv preprint arXiv: 1506.02640v5*, 2016.