

---

# Project Proposal

---

Jinyi Lu  
email

Xiaoqing Tao  
xtao@andrew.cmu.edu

## 1 Literature Survey

### 1.1 Rich feature hierarchies for accurate object detection and semantic segmentation

Traditional methods for object detection usually utilize image features, such as SIFT and HOG. The first challenge for object detection is how to implement localizing within an image. The new method proposed in this paper(1) combines CNN with region proposals. So it is called R-CNN. The general detection process is to extract about 2000 region proposals for each input image. Then utilize CNN to compute a fixed length feature for each region proposal. Finally utilize linear SVM to classify each region.

The second challenge is how to train a large CNN using limited labeled data. This paper proposed to pre-train CNN on a large auxiliary data set, then continually train on a small data set. This method significantly improved the accuracy of objection detection compared with feature learning models. But training is expensive in space and time, and detection is also slow at test time.

### 1.2 Fast R-CNN

This paper(2) talked about how to train a detection network faster. R-CNN is very slow because it extracts feature for each region proposal and there are many duplicate computations. The process could be faster if we share computation.

This paper proposed to modify R-CNN's architecture by taking an image and multiple regions of interests as input. Region proposal method usually depends on Selective Search. Each region of interest is pooled into a fixed-size feature map and fully connected layers are used to extract features. There two output vectors: softmax probabilities and per-class bounding-box regression offsets. The second one is to reduce mislocalization.

### 1.3 Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks

Fast-CNN achieves near real-time rates using very deep networks, when ignoring the time spent on region proposals. Select Search takes about 2 seconds per image on the CPU. Faster R-CNN(3) adds a fully convolutional network to compute region proposals directly. Then use the same detector on the proposed regions as what Fast R-CNN does.

## 2 Project Plan

We plan to implement faster R-CNN on the dataset by the second milestone. After that, we will try YOLO(4) method and see if there are any improvement. Detailed plan depends on time and compute power.

## References

- [1] Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik. *Rich feature hierarchies for accurate object detection and semantic segmentation*. In computer vision and pattern recognition(CVPR),

2014

- [2] Ross Girshick. *Fast R-CNN*. In IEEE International Conference on Computer Vision (ICCV), 2015
- [3] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun. *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*. In neural information processing systems (NIPS), 2015
- [4] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. *You Only Look Once: Unified, Real-Time Object Detection*. In arXiv preprint arXiv:1506.02640, 2015