Object Recognition with Neural Networks

Rahul Handa,[1] Rohini Kapoor,[2]

[1]Department of Computer Science, University of Massachusetts, Amherst

Introduction

Object detection is a growing field of research in Computer Vision and machine learning with applications ranging from self-driving cars to making surveillance monitoring smarter.
In the recent years, deep learning (Convolutional networks) based methods have become the state of the art in object detection in image and videos. They construct a representation in a hierarchical manner with increasing order of abstraction from lower to higher levels of neural network.
We look to perform object detection in videos, and if time and resources permit, we plan to extend the project by looking for dangerous objects(knives, guns, unmonitored bag etc) in the video frames. In our project, we will implement a system similar to the one in "Rich feature hierarchies for accurate object detection and semantic segmentation" by Girshick et al. We follow the following approach for this:

- Extract possible objects using a region proposal method (the most popular one being Selective Search).

- Extract features from each region using a CNN.

- Classify each region with SVMs.
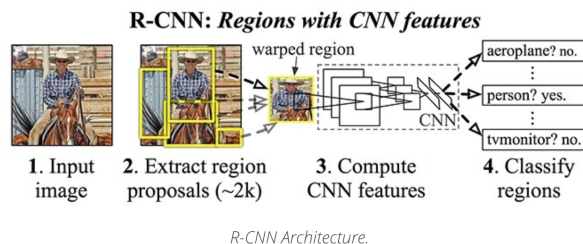

R-CNN Architecture.

Figure 1: Architecture of R-CNN from the paper

We extract a feature vector from each region proposal using implementation of the CNN described by Krizhevsky et al.[2]. Features are computed by forward propagating a mean-subtracted RGB image through five convolutional layers and two fully connected layers.
We might need to experiment with pre-trained CNNs (such as ILSVRC2012 classification or IMAGENET) and modify layers to improve features. The CNN will be adapted further with domain specific fine tuning as per the selected localization method.

For now the evaluation method will be to check the accuracy with which different objects are classified within any frame of the video.

We plan to divide the work as follows:

- Rahul will be working on detection of region from the frames

- Rahul will also work on detection of specific objects from the frames.

- Rohini will be working on extracting features and training a network to classify the images

- Rohini will also working on experimenting with other pretrained models

References and Notes

1. Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik [*Rich feature hierarchies for accurate object detection and semantic segmentation*].

2. A. Krizhevsky, I. Sutskever, and G. Hinton. [*ImageNet classification with deep convolutional neural networks.*]