Object Detection and Semantic Segmentation

CV Project - Team “Kuch bhi”

Team Members:
2018101033 - Jay Sharma
2018102021 - Tanmay Garg
2018102040 - Shantanu Agrawal
2020900019 - Anirudh Polatpally

Objective

- The paper proposes a simple and scalable object detection algorithm.
- We will finetune a convolutional neural networks (CNNs), pre-trained on classification tasks to bottom-up region proposals in order to localize objects.
- If the labeled training data is scarce then supervised pre-training for an auxiliary task, followed by domain-specific fine-tuning will yield a significant performance boost.
- If time permits, we would do small modifications to the implementation, to make it work for the task of semantic segmentation.

Method - Overview

- Selective search - Finding 2000 region proposals to further pass to higher object detection layers
- Feature Extraction with CNN - Making a compact representation vector for each region proposal (affine image scaling done)
- Classification with SVM - Classifying each feature vector obtained for each region proposal
- Bounding Box Regression - Modifying the coordinates and size of each region's bounding box to maximize IoU (intersection over union)

Method - Overview

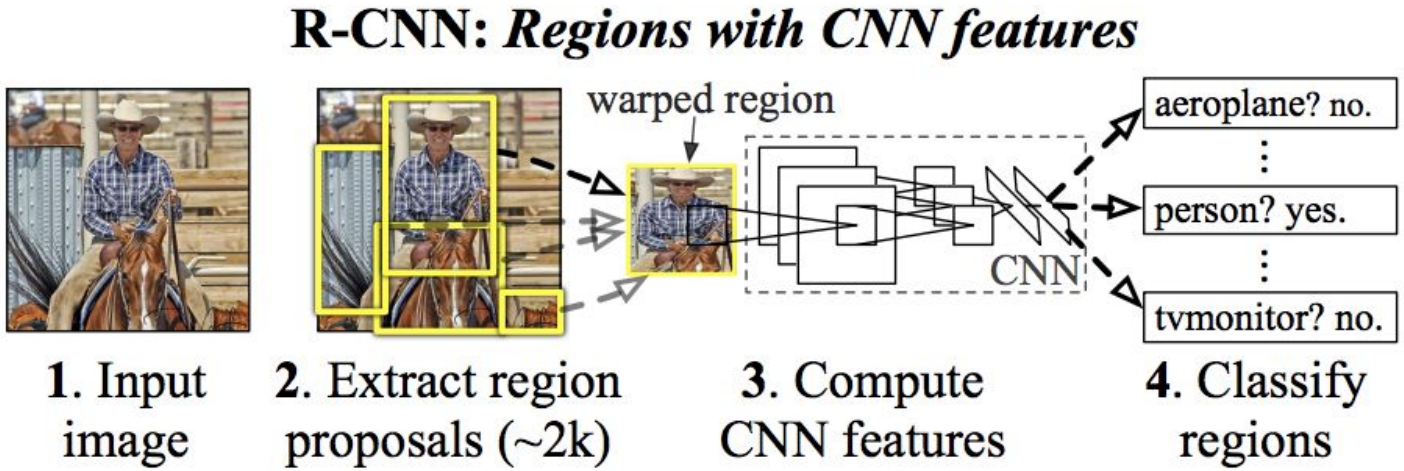

Fig 1: RCNN workflow
Source: [1]

Selective Search

1. Generate preliminary sub-segmentation of input image.
2. Recursively combine smaller regions based on their similarity
3. Generate 2000 regions
4. Output: 4096-dimensional feature vector

Fig 2: Selective Search at different scales
Source: [2]

Feature Extraction - CNN

- VGGNet or AlexNet, pre-trained on ImageNet object classification can be used.
- Last or latter few layers (specific to classification) can be replaced, or fine-tuned.
- Fine-tuning is required because resizing each region to a fixed shape for the CNN would warp the contents, making it tougher to identify features for the CNN.

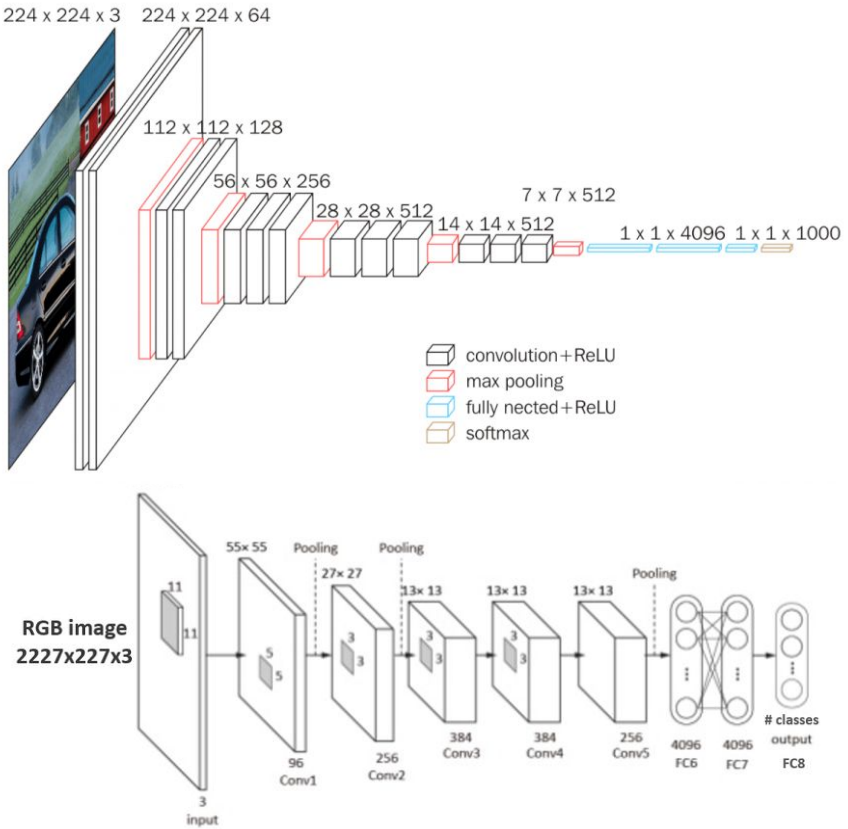

Fig 3: (a) VGG16, (b) AlexNet
Source: [3], [4]

Goals

- To achieve a decent accuracy (maP on classification, IoU on regression) on subset of PASCAL VOC dataset.
- To make the implementation modular enough to be easily modifiable for segmentation.

Deliverable(s)

The following are the deliverables for the mid-evaluation:

- Selective Search implementation
- Implementation of feature extraction using CNN [not trained]

Reference(s)

1. Rich feature hierarchies for accurate object detection and semantic segmentation - Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik (CVPR 2014) - Link
2. Selective Search for Object Recognition - Uijlings, Jasper & Sande, K. & Gevers, T. & Smeulders, A.W.M. (IJCV 2013) - Link
3. ImageNet Classification with Deep Convolutional Neural Networks - Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton (NIPS 2012) - Link
4. Very Deep Convolutional Networks for Large-Scale Image Recognition - Karen Simonyan, Andrew Zisserman (ICLR 2014) - Link

The End