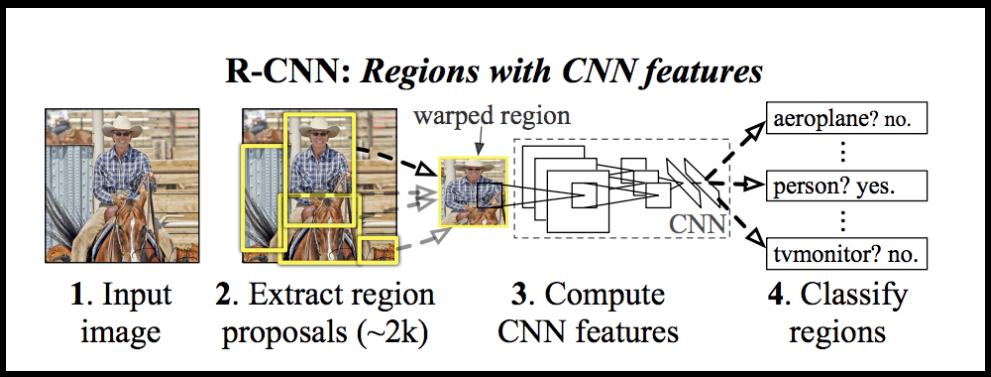
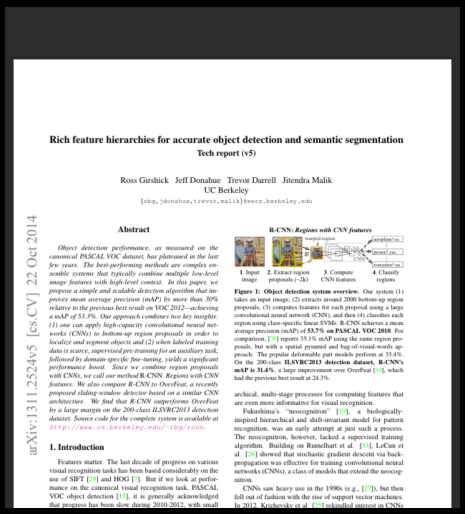
R-CNN (Regions with Convolutional Neural Networks) is a deep learning-based object detection framework introduced by Ross Girshick et al. in 2014. It was one of the first approaches to successfully apply Convolutional Neural Networks (CNNs) for object detection.




Rich feature hierarchies for accurate object detection and semantic segmentation
Tech report (v5)

Ross Girshick Jeff Donahue Trevor Darrell Jitendra Malik
UC Berkeley
{rbg,jdonahue,trevor,malik}@eecs.berkeley.edu

Abstract

1. Introduction

arXiv:1311.2524v5 [cs.CV] 22 Oct 2014

Old Object Detection Techniques

1) HOG
2) SIFT
3) Selective Search

* not very accurate nor efficient

Localization + Classification

1) Region Proposal Methods (Selective Search)
   ↳ generate potential candidate object locations
   2000
2) CNN Feature Extraction

Pipeline

1) Region Proposal Generation
   - Selective Search 2K

2) Feature Extraction by CNN
   - VGG / Alexnet

   Resize (224 x 224 x 3)

3) Classification
   - SVM

4) Bounding Box Regression
   - Fine Tune Bbox coordinates

Limitations

1) Slow Inferencing
   * Computationally expensive

   2000 Regions

   50 sec per image on GPU

2) Storage Overhead

* High Storage Requirements

3) Not End to End Trainable

3 components (Region Proposal, Feature Extraction, Classification)

Trained Separately

Improvements

1) FAST RCNN
2) FASTER RCNN