Object Localization and Detection

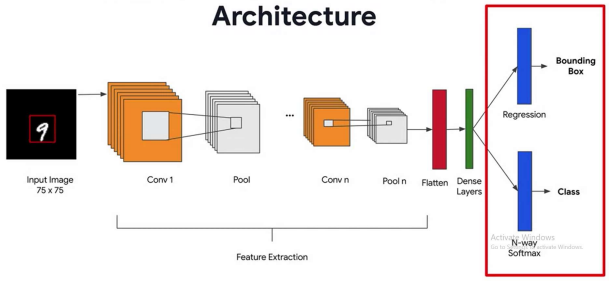
Augmenting MNIST for Object Localization

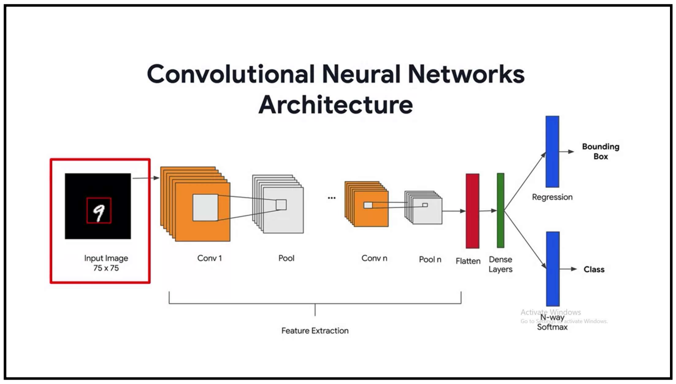
Sample pasted on a 75 x 75 black canvas

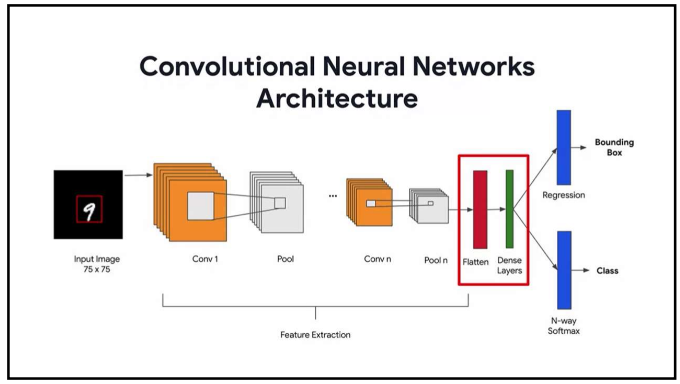
Augmenting MNIST for Object Localization

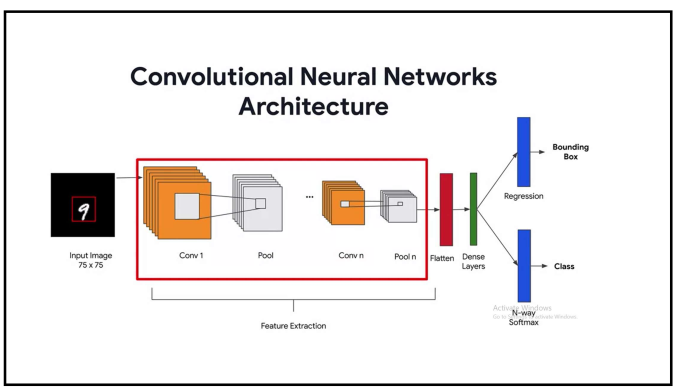
Sample pasted on a 75 x 75 black canvas

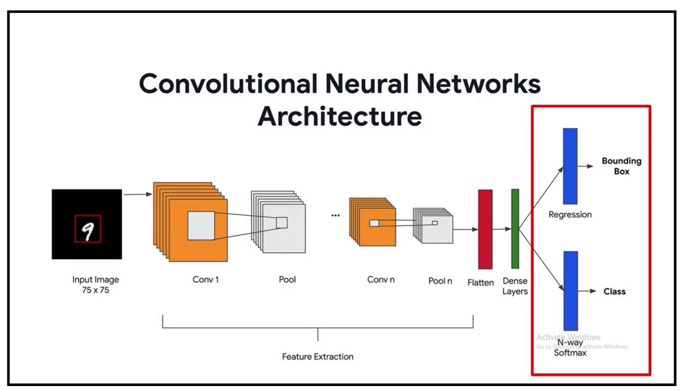
Convolutional Neural Networks Architecture

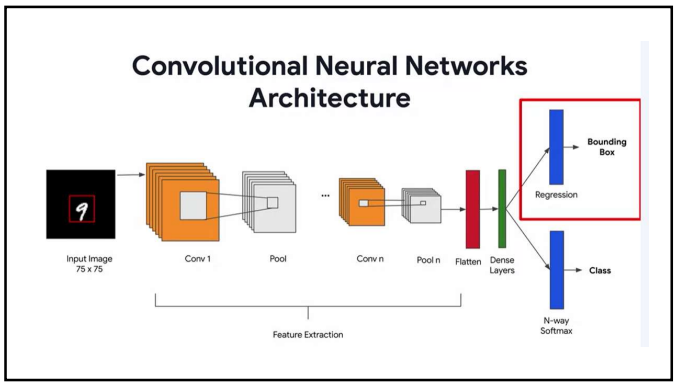


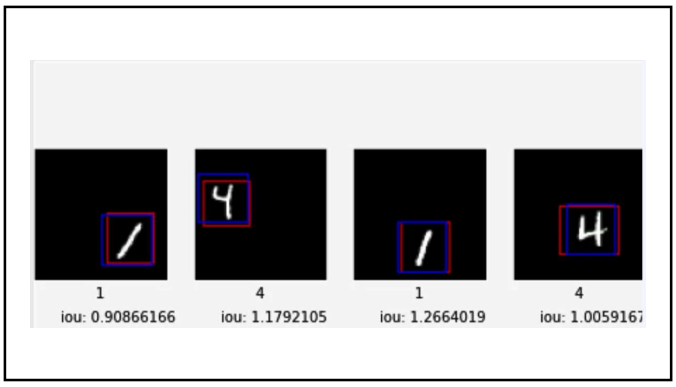

Convolutional Neural Networks
Architecture
Input Image
75 x 75
Conv 1
Pool
...
Conv n
Pool n
Flatten
Dense
Layers
Regression
Bounding
Box
Class
N-way
Softmax
Feature Extraction


Convolutional Neural Networks
Architecture
Input Image
75 x 75
Conv 1
Pool
...
Conv n
Pool n
Flatten
Dense
Layers
Regression
Bounding
Box
Class
N-way
Softmax
Feature Extraction

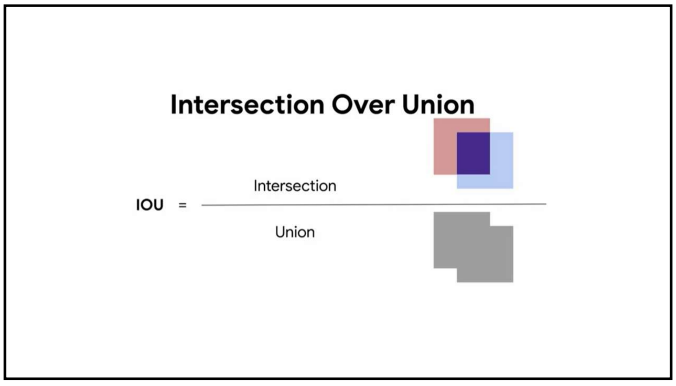

Convolutional Neural Networks
Architecture
Input Image
75 x 75
Conv 1
Pool
...
Conv n
Pool n
Flatten
Dense
Layers
Regression
Bounding
Box
Class
N-way
Softmax
Feature Extraction


Convolutional Neural Networks
Architecture
Input Image
75 x 75
Conv 1
Pool
...
Conv n
Pool n
Flatten
Dense
Layers
Regression
Bounding
Box
Class
N-way
Softmax
Feature Extraction





Evaluating Object Localization



Intersection Over Union

IOU = Intersection / Union

Object Detection and Sliding Windows

Object Detection examples

Facial Detection / Recognition
Amazon Rekognition

Counting vehicles
PowerAI

Analyzing aerial imagery
DIGITS

Object Detection examples

Facial Detection / Recognition
Amazon Rekognition

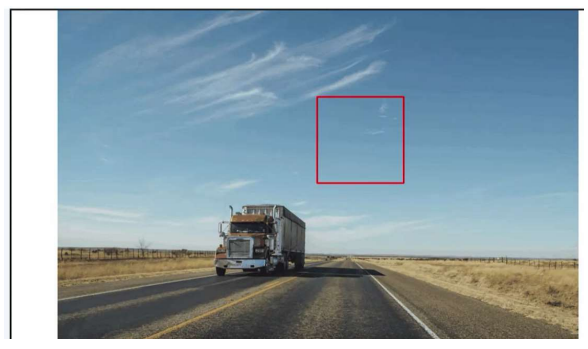
https://docs.aws.amazon.com/rekognition/latest/dg/faces.html

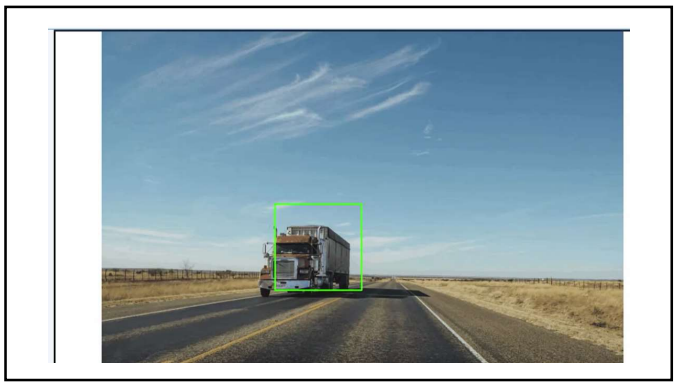
Counting vehicles
PowerAI

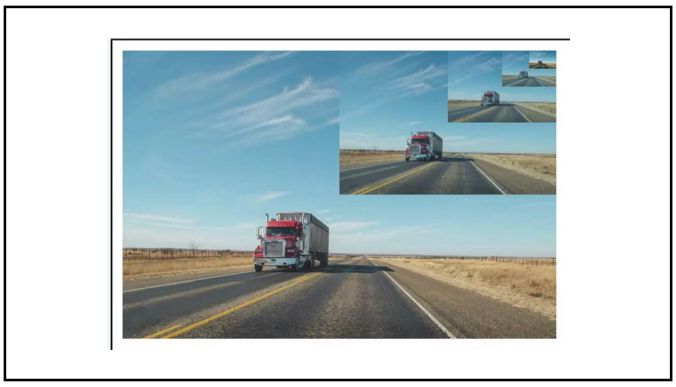
https://github.com/IBM/powerai-counting-cars

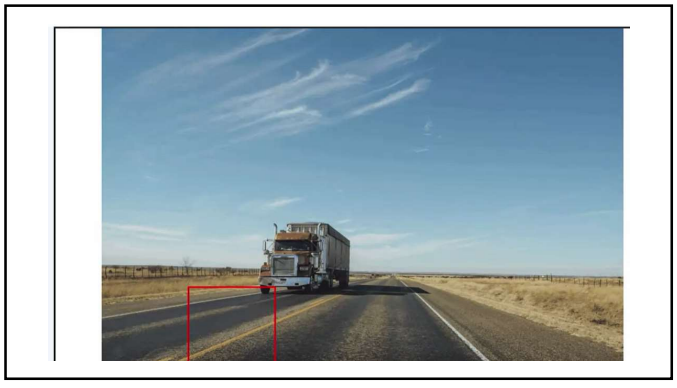
Analyzing aerial imagery
DIGITS

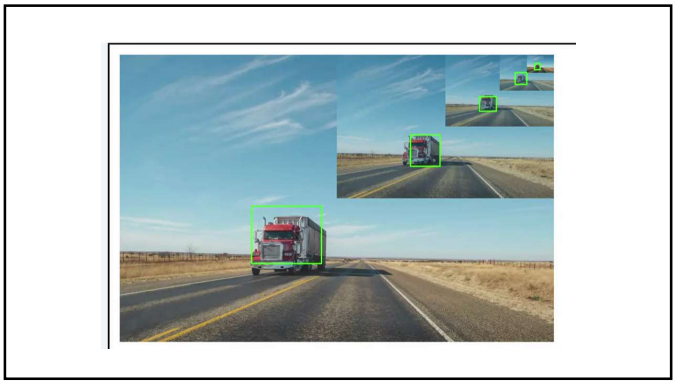
https://devblogs.nvidia.com/detectnet-deep-neural-network-object-detection-digits/







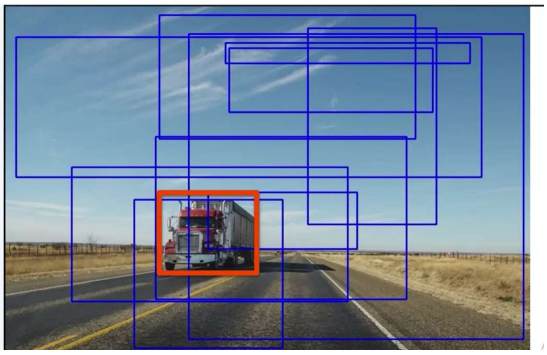




Non-maximum suppression (NMS)

Before NMS

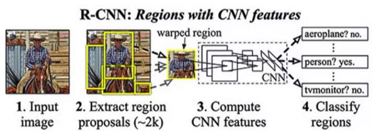
After NMS



Two Stages to Object Detection:

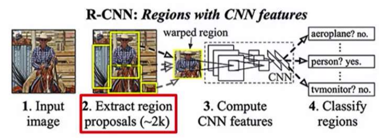
1. Region proposal
2. Object detection and classification

R-CNN

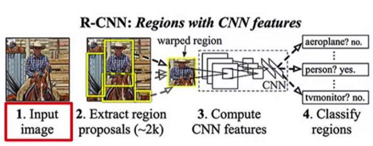


The R-CNN architecture (source: Girshick et al., 2013)

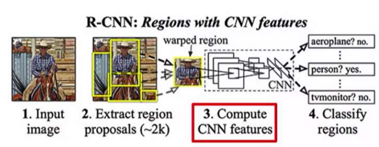
R-CNN



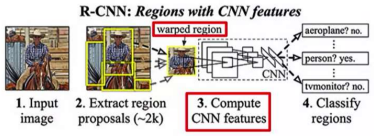
R-CNN



R-CNN



R-CNN



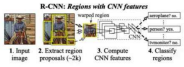
https://arxiv.org/pdf/1311.2524.pdf

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

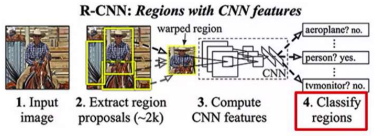
Ross Girshick Jeff Donahue Trevor Darrell Jitendra Malik
UC Berkeley
{rbg,jdonahue,trevor,malik}@eecs.berkeley.edu

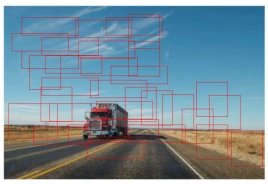
Abstract

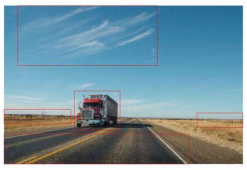
Object detection performance, as measured on the canonical PASCAL VOC dataset, has plateaued in the last few years. The best-performing methods are complex ensemble systems that typically combine multiple low-level image features with high-level context. In this paper, we

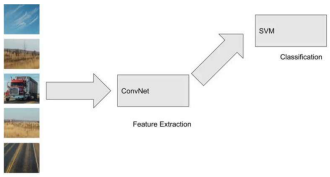


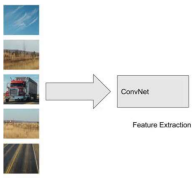
R-CNN

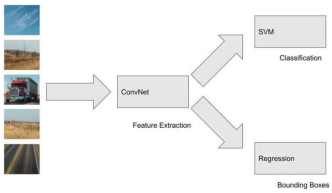







SVM
Classification
ConvNet
Feature Extraction


ConvNet
Feature Extraction


SVM
Classification
ConvNet
Feature Extraction
Regression
Bounding Boxes

Transfer Learning for R-CNN

Pre-train: auxiliary task

Large auxiliary Data Set

Fine-tune: domain specific task

Warped region proposals

R-CNN disadvantages

Slow

memory

Transfer Learning for R-CNN

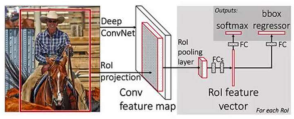
Pre-train: auxiliary task

Large auxiliary Data Set

Fine-tune: domain specific task

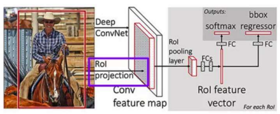
Warped region proposals

Fast R-CNN



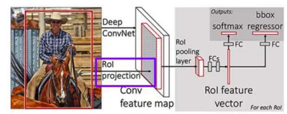
The Fast R-CNN architecture (source: Girshick et al., 2015).

Fast R-CNN



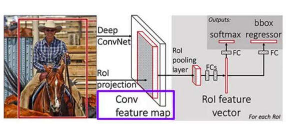
The Fast R-CNN architecture (source: Girshick et al., 2015).

Fast R-CNN



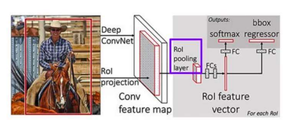
The Fast R-CNN architecture (source: Girshick et al., 2015).

Fast R-CNN



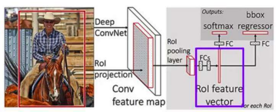
The Fast R-CNN architecture (source: Girshick et al., 2015).

Fast R-CNN



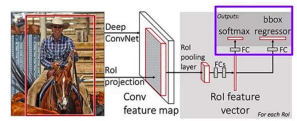
The Fast R-CNN architecture (source: Girshick et al., 2015).

Fast R-CNN



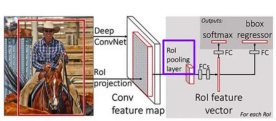
The Fast R-CNN architecture (source: Girshick et al., 2015).

Fast R-CNN



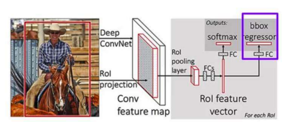
The Fast R-CNN architecture (source: Girshick et al., 2015).

Fast R-CNN



The Fast R-CNN architecture (source: Girshick et al., 2015).

Fast R-CNN



The Fast R-CNN architecture (source: Girshick et al., 2015).

https://arxiv.org/pdf/1504.08083.pdf

Fast R-CNN

Ross Girshick
Microsoft Research
rbg@microsoft.com

Abstract

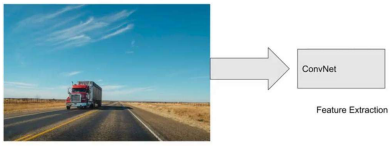
This paper proposes a Fast Region-based Convolutional Network method (Fast R-CNN) for object detection. Fast R-CNN builds on previous work to efficiently classify object proposals using deep convolutional networks. Compared to previous work, Fast R-CNN employs several innovations to improve training and testing speed while also increasing detection accuracy. Fast R-CNN trains the very deep VGG16 network 9× faster than R-CNN, is 213× faster at test-time, and achieves a higher mAP on PASCAL VOC

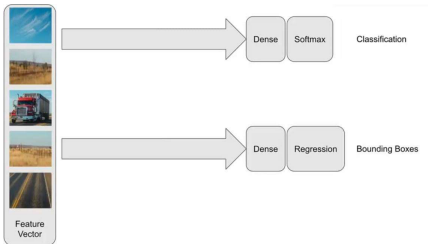
while achieving top accuracy on PASCAL VOC 2012 [7] with a mAP of 66% (vs. 62% for R-CNN).[1]

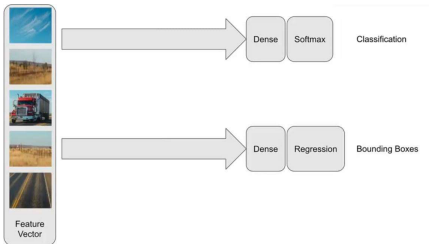
1.1. R-CNN and SPPnet

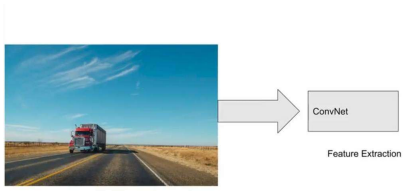
The Region-based Convolutional Network method (R-CNN) [9] achieves excellent object detection accuracy by using a deep ConvNet to classify object proposals. R-CNN, however, has notable drawbacks:

1. **Training is a multi-stage pipeline.** R-CNN first fine-tunes a ConvNet on object proposals using log loss.

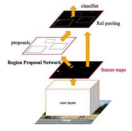




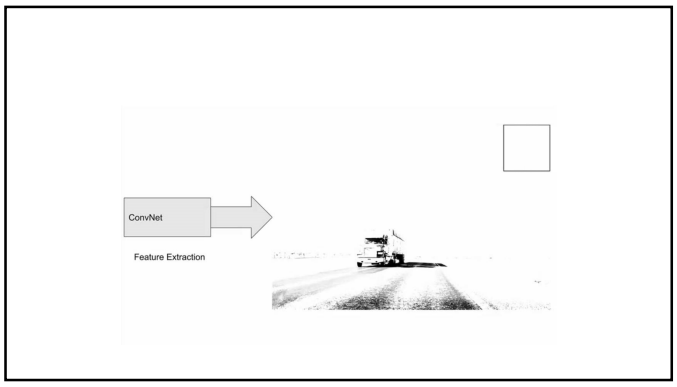


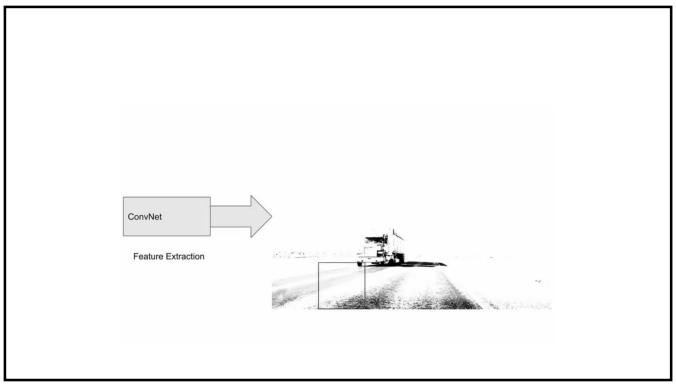


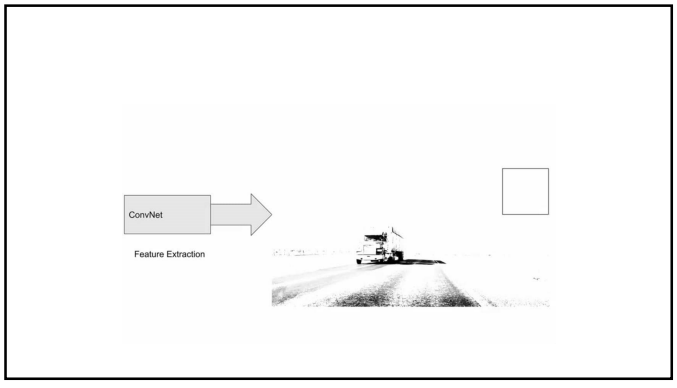
Faster R-CNN

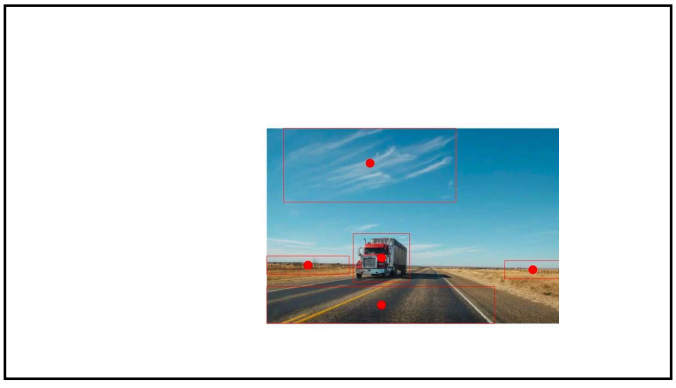


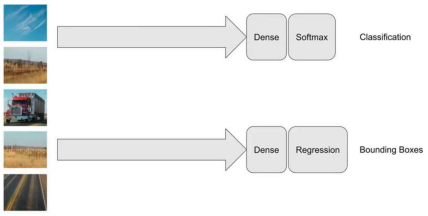
source: Deep Learning for Computer Vision with Python


ConvNet
Feature Extraction


ConvNet
Feature Extraction


ConvNet
Feature Extraction




Dense
Softmax
Classification
Dense
Regression
Bounding Boxes