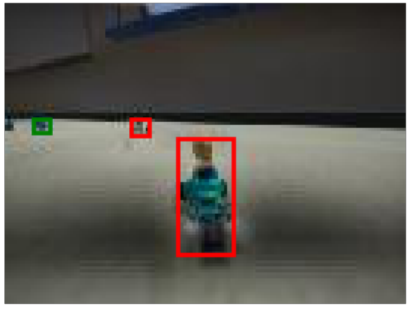
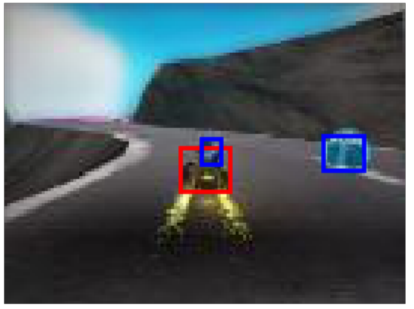
Object Detection

Object Detection

- Sparse labeling
  - Boxes around objects
  - Pose/Keypoint estimation
- Basis of other tasks





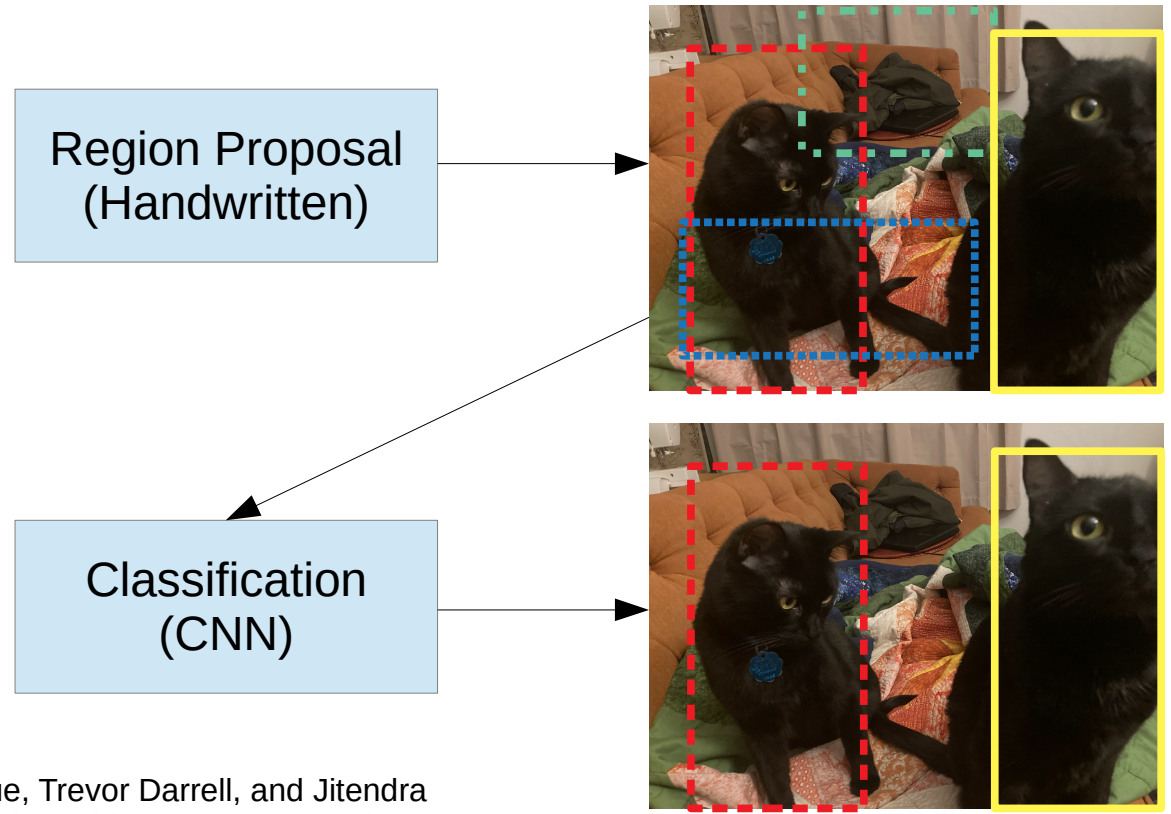
Datasets

- Pascal VOC (2007)
  - 11k images with 27k objects in 20 classes
- MS COCO (2014)
  - 200k images with 1.5m objects in 80 classes
  - Including 250k people with pose data
- Driving datasets
  - nuScenes, BDD100k, ApolloScape, etc.
  - Usually contain 3D and temporal data

Simulated Datasets

- GTA, Carla, Habitat
- Makes labeled data cheap
- Limited relation to real-world scenarios

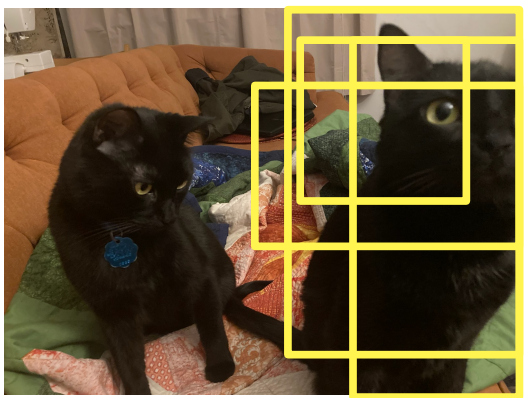
Region-Based Convolutional Neural Network (RCNN)

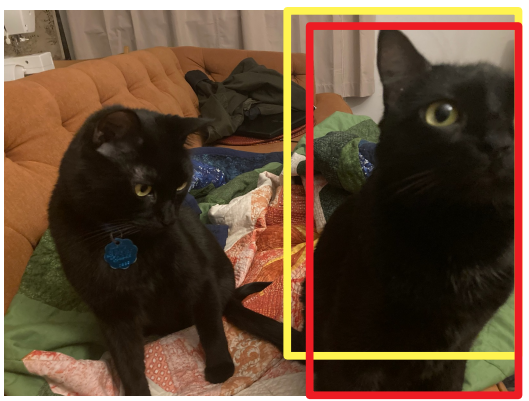


Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2014. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. CVPR 2014.

RCNN

Region Proposal (Handwritten)	Classification (CNN)	Non-maxima Suppression	Bounding Box Regression (CNN)



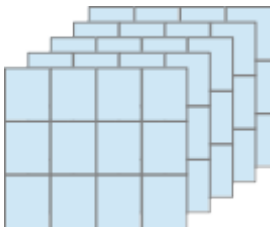


Better performance than existing models in 2013
Very slow – ~1 min per image

Fast RCNN

RCNN: pick regions here

Convolutional Layers



Fast RCNN: pick regions here

Linear Layers

Needs some processing to fit cropped activations into the linear layers: RoIPooling and RoIAlign

50-100x faster than RCNN

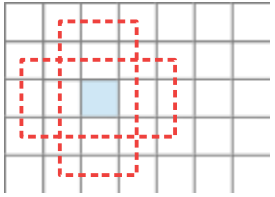
Ross Girshick. 2015. Fast R-CNN. ICCV 2015.

Region Proposal Networks

Use a neural network for region proposal

For several predetermined box sizes NxM, train a classifier to predict whether an NxM box is interesting or not based on only the center location

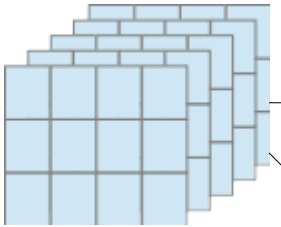
Run those classifiers at every spatial location



Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. NeurIPS 2015.

Faster RCNN

Convolutional Layers



Proposal Network

RoIPooling

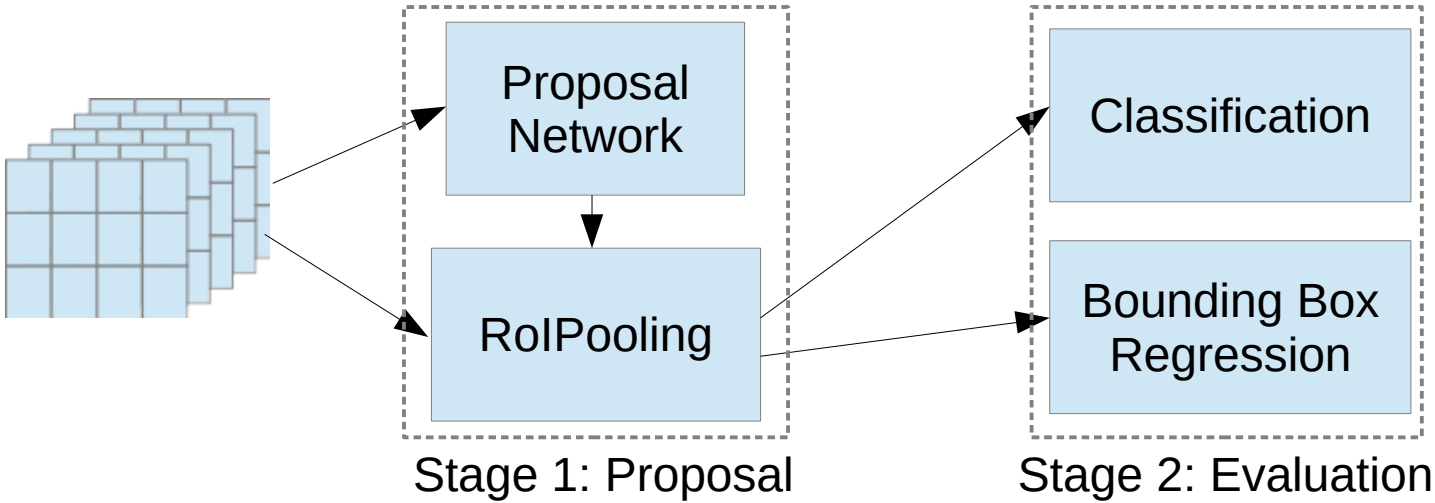
Classification

Bounding Box Regression

Non-Maxima Suppression

All the networks can be trained together

RetinaNet



RetinaNet: Combine these two stages

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2017. Focal Loss for Dense Object Detection. ICCV 2017.

Focal Loss

Reduces the weight
Of high-confidence
samples

$$p_t = \begin{cases} p & if\ y=1 \\ 1-p & otherwise \end{cases}$$

$$\ell_{CE}(p_t) = -\log(p_t)$$

$$\ell_{FL}(p_t) = -(1-p_t)^{\gamma}\log(p_t)$$