Rich feature hierarchies for accurate object detection and semantic segmentation

Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik

Presented by Nick Bergh and Tom Molinari

Problem Background

- Object recognition in PASCAL VOC dataset plateaued 2010-2012
- General approach was SIFT and HOG
- Fukushima’s neocognitron attempted to use a hierarchical and shift invariant approach
- CNNs work well on ImageNet
- This approach tries to use CNNs in conjunction with object detection in order to boost performance

Key Metrics

Model Evaluation

- mAP - Mean Average Precision

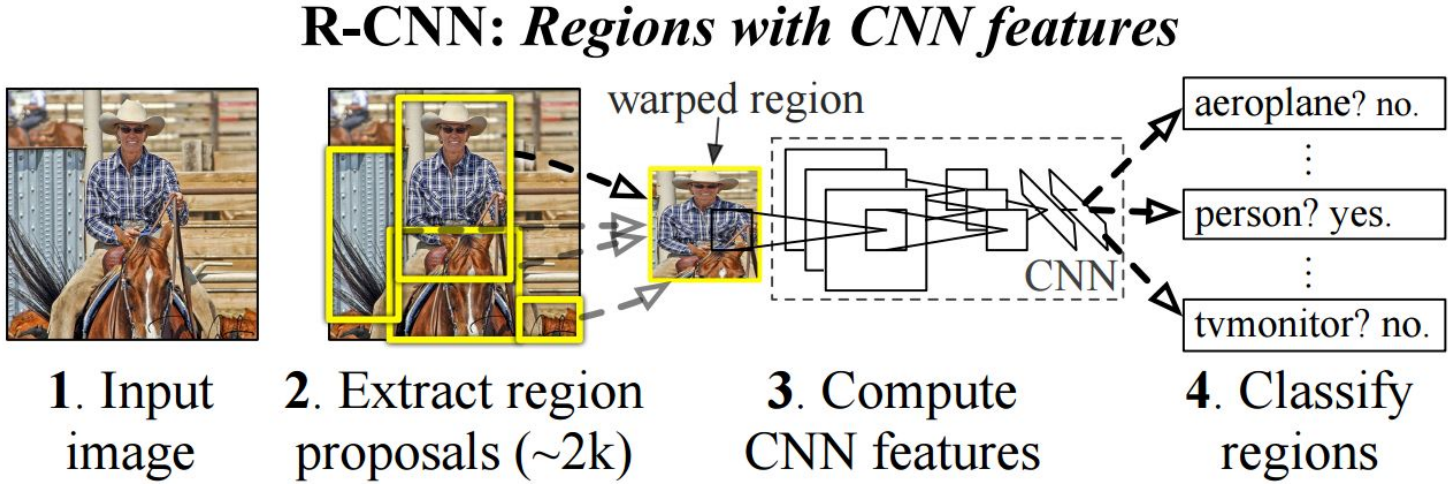
Region Evaluation

- IoU - Intersection over Union

Train/Test Splitting

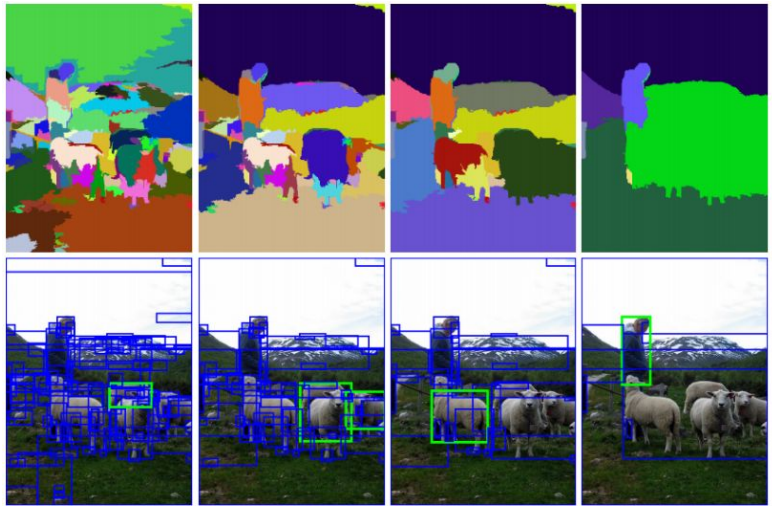
- Relative Imbalance

Model overview



Region Proposals

- Calculated using Selective Search for Object Recognition (Uijlings, et. al.)



CNN Features

- Used the region selections as input into the canonical Krizhevsky, Sutskever, Hinton paper
  - Also known as Alex net
  - Utilized most of the popular techniques which we use: convolutional layers -> FC layers, dropout, momentum, weight decay
- Input regions were stretched to fit the 227x227 input size of Alex net

Clever Optimizations

Supervised Pre-training

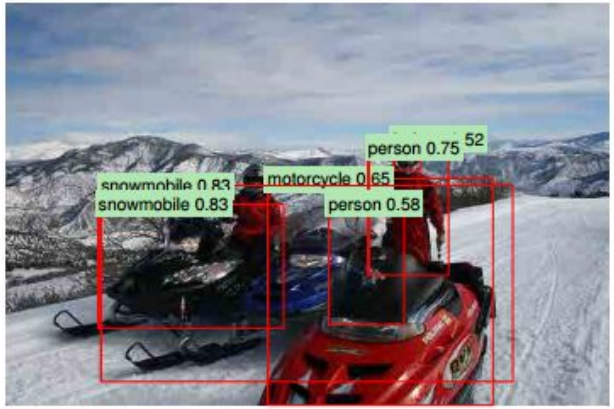
Domain Specific Tuning

- Bounding Box Regressors
- SVM (as classifier)

Mining Hard Negatives

Results

- Only submitted results for evaluation twice (once with and once without bounding box regression)
- PASCAL VOC 2010
  - 53.3% mAP
  - Compared to 35.1% mAP in Uijlings, et. al.
- ILSVRC2013 (Imagenet)



Conclusion

Preselection of regions of interest helps

Alex net can be adapted to different data sets

Domain specific training helps

- bounding box regression

Questions?