Pedestrian Detection Using R-CNN

Group-17 Deepak Kumar(12228) and Mohit Singh Solanki(12419)

Advisor : Prof. Vinay P. Namboodiri

Introduction

Pedestrian detection has been an important problem for decades, given its relevance to a number of applications in robotics, including driver assistance systems, road scene understanding or surveillance systems. The task involves pedestrian detection in a given image using regions proposals along with CNN. For example:

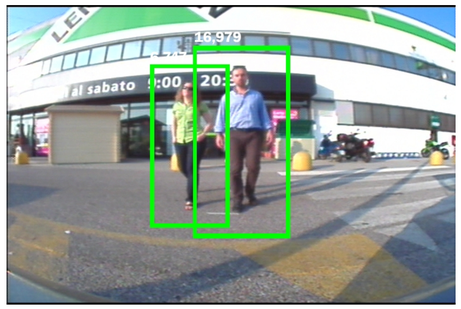
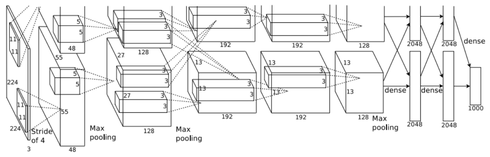


Figure 1: Detection Example

Previous Work

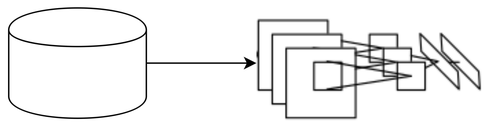
- Sliding window techniques.
- **DPM (Felzenswalb, 2008):** Uses HOG detector with variants
- Other HOG and SIFT feature based methods
- Classification using AlexNet



In all these previous work block-wise orientation histograms are used which can be costly in many cases.

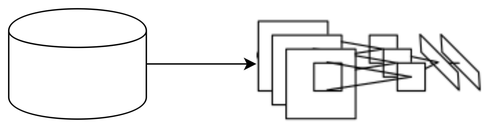
Training R-CNN

- **Pre-Training of CNN**



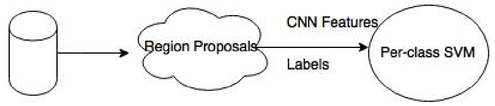
Training Using ILSVRC 2012 ~1.2 million images

- **Fine Tuning on INRIA dataset(around 1.2 million bounding boxes)**



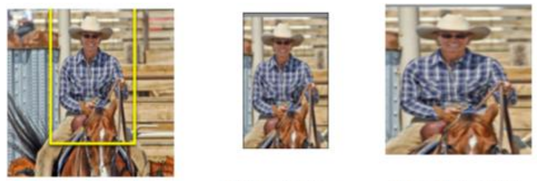
Fine tune on INRIA dataset

- **Training per class linear SVM**



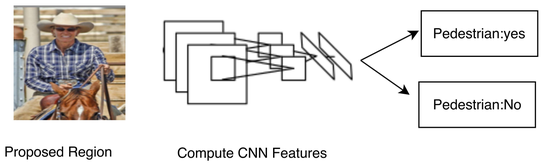
RCNN at test time

- Region Proposal using selective search
- Proposal refinement



1. bounding box 2. Cropped Image 3. Scale to 227 X 227

- Compute features and classify regions



Proposed Region Compute CNN Features

- Greedy non-maximal suppression that rejects a region if it has an (IoU) overlap with a higher scoring selected region larger than a learned threshold

Results

Figure 2: Example Results

Method	Accuracy
Discrimately trained DPM (based on HOG)	88%
Our RCNN based approach	88.2%

Table 1: Comparison of Mean precision on INRIA

- Trained number of region of interest = around 1.2 million
- Number of training Images = 614
- Total time taken for training = 4.7 hours
- Total no. of test images = 288

Selective Search

- Oversegment at small scale
- Group similar segment on different scale
- Each segment as a different region proposal

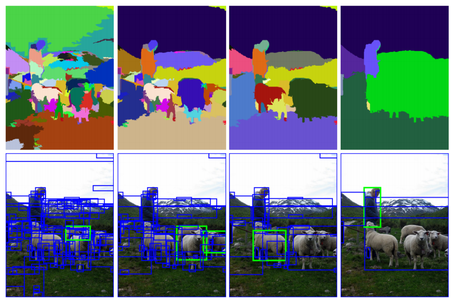


Figure 3: Selective Search

Future Work

- To test our architecture on different datasets (Caltech, ETH etc.)
- Test using different segmentation techniques(BING, MCG, CPMC etc.)
- In place of linear SVM try other classifiers

References

[1] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik.
Rich feature hierarchies for accurate object detection and semantic segmentation.
In *Computer Vision and Pattern Recognition*, 2014.

[2] Gevers2 J.R.R. Uijlings, van de Sande and A.W.M. Smeulders2.
Selective search for object recognition.
ICCV, 2011.

[3] Geoffrey E. Hinton Alex Krizhevsky, Ilya Sutskever.
Imagenet classification with deep convolutional neural networks.
NIPS, 2012.