Research paper Summary

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

The main idea of this research is to create an object detection system with combining the region proposal with CNNs. It's mainly focus on two key factors, one is localizing the object and other on is train the model with a very small amount of data. The author proposed this object detection algorithm outperforms all existing methods on PASCAL VOC datasets. It achieved 53% of mean average precision (mAP) which is batter then the previous best result on PASCAL VOC 2012. Since this model combining the region proposal with CNNs they named it R-CNN (Regions with features).

This object detection system comprises of three module, category independent region proposals, a bigger CNN for extract a fixed-length feature vector from each region and the last one is set of class specific linear SVMs. For region proposal selective search methods is used to enable a controlled comparison with prior detection work. Using the Caffe implementation of Krizhevsky's CNN, this model extracts a 4096-dimensional feature vector from each region proposal. Each region proposal warped in tight bounding box and forward propagate through a large CNN. It extract a fixed length feature vector from each proposal. Then a set of class specific linear SVM's are used to classify the extracted feature vectors.

To deal with small amount of data they used a pre trained CNN on a larger auxiliary dataset without image level annotations and domain specific fine-tuning SGD to perform object detection as well as adopt new domain.

There are two main properties that makes this object detection efficient number one is CNN parameters are shared across all categories and second one is it compute low dimensional feature vector. They also use a simple bounding box regression stage to improve localization performance.

This model has several limitations, multi-stage training pipeline is one of them. At first CNN used for extract image feature. Then on this extracted feature support vector machine trained again and also box regression model trained separately. This multi-stage training process is time consuming. Because of CNN applied on each object region proposal it requires large memory for the model keep record which makes training process slow.