Research paper Summary

Fast R-CNN

This Fast RCNN is build top of previous RCNN and SPPnet. It addresses the drawback of past RCNN and SPPnet. The main goal of this experiments was to reduce the model complexity of those previous SPPnet. SPPnet was build top of RCNN. This study suggests a single-stage training system that simultaneously learns to categorize object suggestions and improve their spatial positions. Previous R CNN training had several stages, additionally, this model's object detection took a long time. SPPnet accelerates RCNN by 10 to 100× during testing, but it also has some significant downsides. Similar to R-CNN, training is a multi-stage process that includes feature extraction, log loss network fine-tuning, SVM training, and bounding-box regressor fitting.

At first a whole image and a list of object proposals are taken by Fast R-CNN. The network then applies multiple CNN and max pooling layers to the entire image to create a feature map. Then the ROI pooling layer extracts a fixed-length feature vector from the feature map and fed into a sequence of fully connected layers that provide two output layers. One is Softmax probability and other one is for bounding box regression. An end-to-end multi-task loss is used to train in this architecture. This is a single-stage training and the model is comparatively fast to training and test that's why it is named Fast R-CNN.

The key difference between SPPnet and Fast R-CNN is in the output layer and pooling layer. In SPPnet they did three or four levels of pooling and Fast R-CNN used single level of 7 × 7 grid pooling, and they named it RoI pooling. In output layer instead of using both Softmax and SVM classification they used only Softmax classification, because it’s provide the same level of accuracy. Additionally, they choose to employ smooth L1 loss for bounding box rather than L2 loss. Instead of merely using classification loss to backpropagate via the network in this model, they additionally chose to add together classification loss and bounding box regression loss. It provides a higher level of accuracy, and also reduce the training stages.

Though this model solves some of previous issues it also have some drawbacks. This algorithm requires many passes through a single image to extract all the objects, this creates too complications. There is also no such improvement on accuracy, but it is 146× faster than the previous R-CNN.