Fast R-CNN Paper Summary

Title: Fast R-CNN

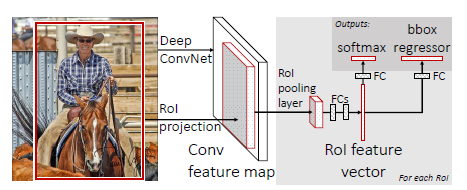
Initial Goal

1. Instead of a multi-stage training process, propose a single-stage training process.
2. Save storage, and fasten training.

Contributions

1. Single-stage training.
2. The resultant approach is 9 times quicker than R-CNN and 3 times faster than SPPnet in training a very deep detection network (VGG16).
3. In VGG16, fine-tuning Conv layers enhances mAP. mAP increases from 66.0 to 70.0 for VOC-2007. The 2010 mAP for VOC has increased from 62.2 to 68.8. 62.4 to 68.4 for VOC 2012.

Model Architecture



RoI pooling layer: The RoI pooling layer employs max pooling to create a small feature map from the features inside any eligible region of interest.

Initializing from pre-trained networks: They tested three pre-trained ImageNet networks, each having five maximum pooling layers and five to thirteen convolution layers.

How it works

1. A fully convolutional network is used to process an input picture and numerous regions of interest (RoIs).
2. Fully connected layers pool each RoI into a fixed-size feature map, which is subsequently mapped to a feature vector (FCs).
3. Softmax probabilities and per-class bounding-box regression offsets are the network's two output vectors per RoI.

Advantages

1. Single-stage training.
2. Faster than R-CNN and SPPnet.
3. Increased mAP.

Limitations

Still uses the Selective Search model for region proposal which is slow.