Rich feature hierarchies for accurate object detection and semantic segmentation

Abstract

This method combines two key discoveries: (1) high-capacity convolutional neural networks (CNNs) can be used to localize and segment objects from bottom-up region proposals; and (2) when labeled training data is scarce, supervised pre-training for an auxiliary coverage by domain-specific fine-tuning yields an importance performance.

In this study, region recommendations are combined with CNNs. A comparison of R-CNN and OverFeat, a sliding-window detector based on a similar CNN architecture that was recently proposed, is also available. R-CNN outperforms OverFeat by a significant margin on the 200-class ILSVRC2013 detection dataset.

Dataset

They utilized the ILSVRC2013 detection dataset, which has 200 classes. The dataset is divided into three groups: train (395,918), val (20,121), and test (40,152), with the number of pictures in each group shown in parentheses.

Architecture Model

- This system accepts a picture as input.
- Around 2000 bottom-up region ideas are extracted.
- Regions of distorted images
- CovNet should be used to forward each area.
- Uses class-specific linear SVMs to classify each area.

Working Procedure

Objects are localized and segmented by using-

- CNNS with a large capacity
- Proposals for regions from the ground up

High-capacity learning when labeled training data is sparse, CNNs are used-

- Pre-training for an auxiliary duty that was supervised
- Then there's domain-specific fine tuning

Limitations

- You'd have to categorize 2000 region ideas every image to train the network, which would take a long time. Because each test image takes around 47 seconds, it can't be done in real time.

- A predefined algorithm is used for selective search. As a result, there is no learning at that moment. As a result, a flurry of poor candidate region ideas might be generated.

Result

The success percentage of popular deformable part models is 33.4 percent. On the 200-class ILSVRC2013 detection dataset, R-mAP CNNs have a 31.4 percent accuracy, which is much higher than OverFeat's previous best result of 24.3 percent.

Fast R-CNN

Abstract

- In this study, the Fast Region-based Convolutional Network model (Fast R-CNN) is proposed for object detection.
- Fast R-CNN uses a shortened training procedure with only one fine-tuning, as opposed to R-multi-stage CNN's training procedure. When compared to previous R-CNN networks, fast R-CNN improves training and testing speed while also improving detection accuracy.
- Fast R-CNN trains the very deep VGG16 network 9 times faster than R-CNN, is 213 times faster at test time, and achieves a higher mAP on PASCAL VOC 2012.
- Fast R-CNN trains VGG16 three times faster, tests ten times faster, and outperforms SPPnet in accuracy.

Dataset

They evaluated the network using three datasets: VOC07, VOC 2010, and VOC 2012. Fast RCNN experiments all use single-scale training and testing.

Architecture Model

The fast R-CNN network employs both the entire input image and a list of item suggestions. The network first processes the entire image with many convolutional and max pooling layers to create a conv feature map. The feature map is then used by a region of interest (ROI) pooling layer to generate a fixed length feature vector for each object proposition. After that, each feature vector is routed through a series of fully connected layers.

Working Procedure

While training Fast R-CNN, the authors used pre-trained networks to initialize the architecture. The following modifications are made to the initialized pre-trained networks:

1. The previous max pooling layer has been replaced by the ROI pooling layer.
2. The network's final fully connected layer and softmax are replaced by two sibling layers: a fully connected layer with softmax across K + 1 categories (number of categories plus background) and category-specific bounding box regressors.
3. As input, a list of photos and a list of RoIs in those photos are sent to the network.

The ImageNet models listed below are used in the study.

1. The AlexNet Model (Small)
2. VGG CNN M 1024, which is deeper than S but wider. M (Medium): VGG CNN M 1024, which is deeper than S but wider.
3. L (Large): A deep VGG16 model.

Limitations

When we compare the performance of Fast R-CNN during testing, we notice that including region proposals significantly slows down the algorithm when compared to not including region proposals. As a result, region proposals act as roadblocks in the Fast R-CNN algorithm, slowing it down.

Result

The following primary experimental results back up the paper's statements.

- On VOC07, 2010, and 2012, cutting-edge mAP was used.
- When compared to R-CNN, SPPnet enables faster training and testing.
- Fine-tuning conv layers in VGG 16 improves mAP.

Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks

Abstract

- In this research, they present the Region Proposal Network (RPN), which shares full-image convolutional features with the detection network and so allows for almost cost-free region suggestions.
- An RPN is a fully convolutional network that simultaneously predicts object bounds and scores at each point. From start to end, the RPN is trained to provide high-quality region suggestions that Fast R-CNN employs for detection.
- They further merge RPN and Fast R-CNN into a single network by sharing their convolutional features, using the recently popular terminology of neural networks with "attention" processes. The RPN component tells the unified network where to seek for information.

Dataset

On the PASCAL VOC 2007 detection benchmark, they put their strategy to the test. There are around 5k trainval and 5k test images in this dataset, which spans 20 object categories. They also share results from the PASCAL VOC 2012 benchmark for a few models.

Model Architecture

Faster RCNN is built upon two modules:

1. The first module is a deep fully convolutional network that suggests regions (Region Proposal Network or 'RPN').
2. The Fast R-CNN detector, which uses the proposed regions to detect objects, is the second module.

Working Procedure

They create a Region Proposal Network (RPN) that shares full-image convolutional features with the detection network, allowing region proposals to be made for almost no money. An RPN is a fully convolutional network that predicts object limits and scores at each position at the same time. The RPN is trained from start to finish to create high-quality region proposals that Fast R-CNN uses for detection.

In this study, alternating optimization is utilized to learn shared features using a four-step training technique.

1. Train the RPN
2. Train the separate detection network by Fast RCNN using proposals generated by step 1 trained RPN

3. Use detector network to initialize RPN training, keeping the shared convolutional layers fixed, only fine tuning RPN exclusive layers.
4. Finally, keeping the shared convolutional layers fixed, they fine-tune the unique layers of Fast R-CNN.

Limitations

The RPN is trained using a single picture to extract all anchors in the mini-batch of size 256, which is one disadvantage of Faster R-CNN. Because all samples from a single picture may be associated, the network may take a long time to reach convergence.

Result

- The RPN technique beats Selective search and edgebox by 1.3 mAP on the PASCAL VOC 2007 testset (trained on VOC 2007 trainval) using Fast R-CNN with ZF detectors but separate region proposal methods.
  When using Fast RCNN and VGG16 as the detector, the mAP on the VOC 2007 test set is 78.8% when trained on COCO+07+12 (union set of VOC 2007 trainval and VOC 2012 trainval).
- Faster RCNN(on VGG-16) increases mAP@0.5 by 2.8 percent and mAP@[0.5, 0.95] by 2.2 percent on COCO test-dev when trained on the COCO train dataset.
  When trained on the COCO train dataset, the Faster R-CNN system improves the mAP@0.5/mAP@[0.5, 0.95] from 41.5 percent /21.2 percent (VGG-16) to 48.4 percent /27.2 percent (ResNet-101) on the COCO val set by simply substituting VGG-16 with a 101-layer residual net (ResNet-101).