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

Summary written by Vedant Jain
April 21, 2021

Summary: This paper describes a novel approach to object detection based on a region marking algorithm. The Region Proposal Network makes use of existing convolutional network features and then simultaneously predicts the boundaries of an object as well as the detection of it. The algorithm is written so that it can work with the Fast R-CNN which involves actual object identification. The authors show that their method is very powerful at detecting object bounds as well as objectiveness scores. Furthermore, the network is much faster than existing methods because it only requires one pass per pixel and makes use of anchor points that help streamline the process. Testing with the PASCAL VOC 2007 dateset they had a object detection accuracy of 73.2% mAP.

Related work: CNN's are highly used in image classification, but fall sort on object detection datasets. As such many new methods have been proposed. Specifically, the question of object detection can't be solved with simple CNN's because it would be too computationally expensive to check every pixel.[1] One method proposed was the Overfeat method, which makes use of fully connected layers that are used to determine the coordinates of object boxes for a single object. In the case of multiple objects convolutional networks are used. Overfeat outputs many bounding boxes for objects which are then used in the R-CNN. A big issue in this method is that it doesn't share any convolutional layers with R-CNN and so causes an increase in the training time. [2]

Approach: The Region Proposal network outputs a group of rectangular object proposals, where each rectangle has an objectiveness score. The network shares it's convolutional networks with the Fast R-CNN object detection network. The convolutional layers are pulled from the fully trained VGG model and ZG model. The Region Proposal Network essentially involves sliding a smaller window that is of different shapes based on the a prior k anchor boxes. Then each of the measurements is fed into a box regression layer which is used for finding the coordinates of the boxes, and a box classification layer which is used to describe the probability of an object in the proposed area. Training of the network involves a modified cost function that involves the log loss over object vs no object, ie two classes and for regression it focuses on individual k regressors, allowing for calculations based on a specific size input.

Datasets, Experiments and Results: The authors of this paper attempted their detection network on the PASCAL VOC 2007 detection benchmark. They used a fast version of ZF net using 5 convolutional layers and a VGG-16 model that has 13 conv layers. They used the detection mean Average Precision measurement as the criteria for measuring the success of the measurement. The RPN achieves competitive measurements at roughly 59.9% mAP. But more importantly it results in much faster measurements then other methods like Selective Search.

Strengths: This paper describes a very clever approach to working with object detection. The strength in their method comes with maintaining similar accuracy to the state of the art methods while significantly lowering the amount of time required for testing. The author also does a good job of going through why their method works and looking back at exactly why it is the case. For example, the last section of the paper deals with ablation experiments where certain parts of the algorithm are excluded to determine their contribution.

Weaknesses: A weakness in this paper was how thorough they were in explaining their algorithm. The article makes sizable jumps in knowledge from the basics of detection networks to complicated applications using multiple loss functions. The authors could strengthen the paper by systematically going through their methods while also providing enough context to make the work easy to decipher.

Reflections: Ren et al, have created a new algorithm which they call the region proposal network. The model makes use of the same convolutional layers as the fast-RCNN leading to faster calculations maintaining the same if not better accuracy. Finally, they describe how different abelations to the network result in changes in performance leading to findings that describe the importance of the classification and regression layers.

References

[1] A. T. C. Szegedy and D. Erhan. Deep neural networks for object detection. 2013.

[2] X. Z. M. M. R. F. P. Sermanet, D. Eigen and Y. LeCun. Overfeat:integratedrecognition, localization and detection using convolutional networks. 2014.