Object Detection Proposal Convolutional Neural Networks +3

... does the region proposal netwo... Faster R-CNN work?

This question previously had details. They are now in a comment.

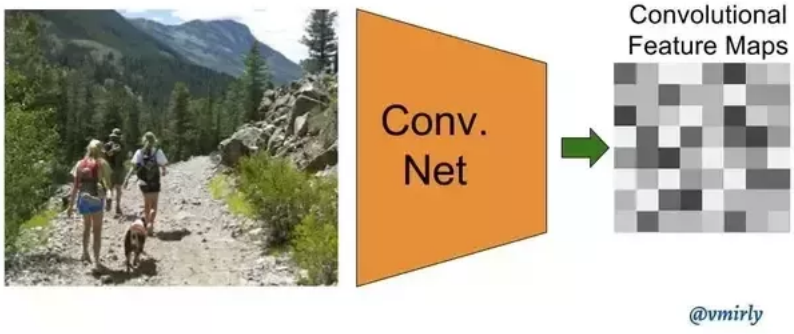
3 Answers

Vahid Mirjalili, Ph.D. from Michigan State University (2019)
Answered Apr 9, 2017 · Upvoted by Elijah Philpotts, M.S. Computer Science & Machine Learning, Georgia Institute of Technology (2016)

Originally Answered: How does RPN work on the Faster R-CNN?

Based on the original paper, Towards Real-Time Object Detection with Region Proposal Networks , I have summarized the RPN in 3 steps.

1. In the first step, the input image goes through a convolution network which will output a set of convlutional feature maps on the last convolutional layer:



2. Then a sliding window is run spatially on these feature maps. The size of sliding window is $n \times n$ (here 3×3). For each sliding window, a set of 9 anchors are generated which all have the same center (x_a, y_a) but with 3 different aspect ratios and 3 different scales as shown below. Note that all these coordinates are computed with respect to the original image.

Related Questions

Will capsule networks replace neur

How is Fully Convolutional Networ different from the original Convolut Network (CNN)?

What is the advantage of combinin Convolutional Neural Network (CN Recurrent Neural Network (RNN)?

Convolutional Neural Networks: W bounding box regressors doing in

What is a good Deep Learning fran implement R-CNN based Object D Tracking?

What are the differences between segmentation, instance detection a proposal?

Is Fast R-CNN a deep learning bas

Is R-CNN convenient for text detec

Why are deep learning architecture Faster R-CNN, or SSD open to all? these people get these architecture

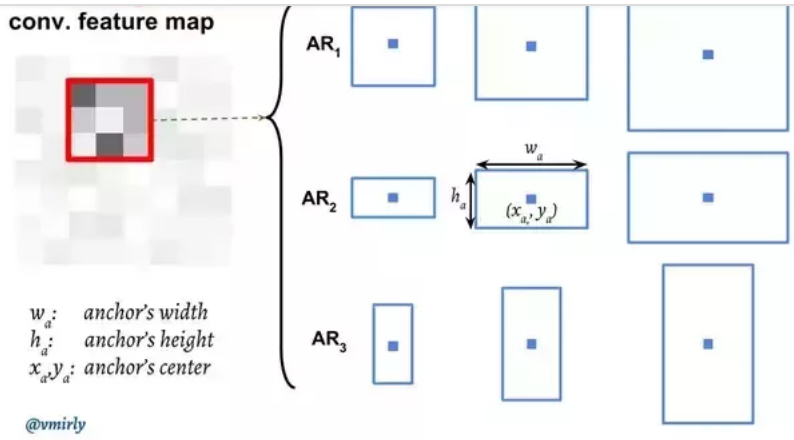
How important is it to share the co layers between the region proposa detection network in the faster R-C architectu...

+ Ask New Question

More Related Questions

Question Stats

54 Public Followers
22,010 Views
Last Asked Jul 19, 2017
3 Merged Questions
Edits



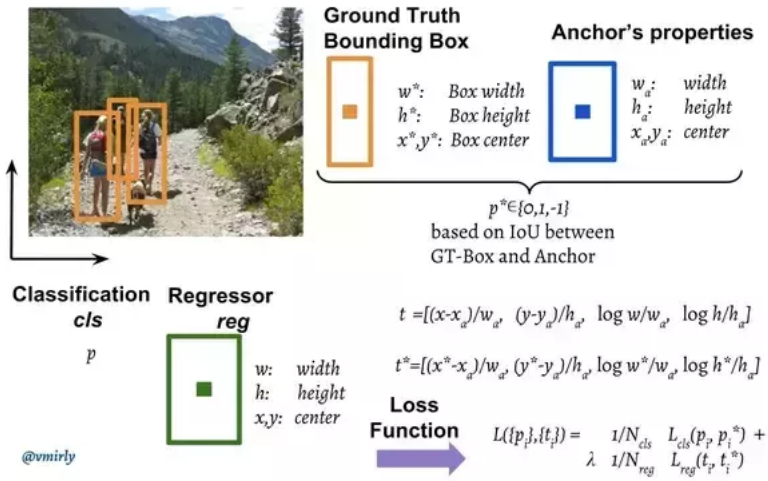
Furthermore, for each of these anchors, a value p^* is computed which indicated how much these anchors overlap with the ground-truth bounding boxes.

$$p^* = \begin{cases} 1 & if \quad IoU > 0.7 \\ -1 & if \quad IoU < 0.3 \\ 0 & otherwise \end{cases}$$

where IoU is intersection over union and is defined below:

$$IoU = \frac{Anchor \cap GTBox}{Anchor \cup GTBox}$$

3. Finally, the 3×3 spatial features extracted from those convolution feature maps (shown above within **red box**) are fed to a smaller network which has two tasks: classification (cls) and regression (reg). The output of regressor determines a a predicted bounding-box (x, y, w, h), The output of classification sub-network is a probability p indicating whether the the predicted box contains an object (1) or it is from background (0 for no object).



Chomba Bupe, develops machine learning algorithms
Answered Mar 29, 2017 · Upvoted by Elijah Philpotts, M.S. Computer Science & Machine Learning, Georgia Institute of Technology (2016)

The region proposal network (RPN) in the faster region-based convolutional neural network (Faster R-CNN) is used to decide “where” to look in order to reduce the computational requirements of the overall inference process. The RPN quickly and efficiently scans every location in order to assess whether further processing needs to be carried out in a given region. It does that by outputting *k* bounding box proposals each with 2 scores representing probability of object or not at each location.

The anchor boxes are just references, they are selected to have different aspect ratios and scales in order to accommodate different types of objects, elongated objects like buses, for example, cannot be properly represented by a square bounding box. In Faster R-CNN they used k = 9 representing 3 scales and 3 aspect ratios. Each regressor in the RPN only computes 4 offset values (w, h, x, y) to the corresponding reference anchor box.

where w = width, h = height, (x,y) = center

The RPN uses a 3x3 window that slides over a high-level conv feature map, the effective size of that small window is actually 177x177 when reprojected back to the input layer, so the RPN is actually using a lot of context when making the proposals. This 3x3 window is resampled to a 256 dimensional vector before feeding into two fully connected layers, a box regression layer (*reg*), that computes the box offsets, and the box classification layer (*cls*) that computes the confidence scores that are related to probability of objectness. The *reg* layer has *4k* outputs while the *cls* layer has *2k* outputs making the total RPN output per position to *4k + 2k*.

So at each location of the conv layer, the bounding box regression heads outputs the bounding box offsets for each anchor box while the classification layer outputs the confidence scores that represents whether an object is present or not within each anchor box. Only those boxes with a corresponding high

probability of containing an object.

Hope this helps.

10.1k Views · View Upvoters · Answer requested by Feras Almasri

Upvote 51 Downvote

Abhishek Shivkumar
Chomba Bupe can you please clarify what is meant by resampling 3x3 window to a 25...

EstherRietmann, Tech and Machine Learning at Tryolabs
Answered Jan 18

Hi!

As already mentioned by Vahid Mirjalili, the RPN takes all the reference boxes (anchors) and outputs a set of good proposals for objects. It does this by having two different outputs for each of the anchors.

If you are interested in a detailed explanation of the RPN (and other parts of the Faster R-CNN model), you might find this blog post helpful:

Faster R-CNN: Down the rabbit hole of modern object detection

946 Views · View Upvoters

Upvote 4 Downvote

Add a comment... Recommended All

Top Stories from Your Feed