论文图片中定位公式位置

张浩然 1500010684

Contents

Abstract

1 背景介绍

在处理论文图片中定位公式位置这个问题上，我们准备了两个方向上的方法。一是将论文图片切割为段落图片后使用目标检测方法来定位公式的位置，二是在预处理上更进一步将论文图片切割为单词图片，再将单词图片分类。

在目标检测这个问题上有许多的经典算法。基于卷积神经网络的目标检测开始于 2013 年 RBG 的论文 [2] 提出的 RCNN。RCNN 的算法过程大致为生成候选区域后使用 CNN 进行特征提取，将提取的特征通过 SVM 分类，最后通过边框回归 (bounding-box regression) 得到精确的目标区域。此算法的主要问题在于候选区域过多，大量的区域重复和无效造成了巨大的计算浪费。另一个问题在于使用 CNN 需要输入固定尺寸的图片，而图片的截取和拉伸等操作造成了输入信息的丢失。之后有许多算法以此为基础进行了改进。

首先是空间金字塔池化 SPP-Net [4]，在全连接层前加入了一层将输入的特征图池化为特定尺寸的输出的特殊池化层，通过输入的尺寸与需要的输出尺寸计算出所需的池化核和步长从而实现了输出固定尺寸至全连接层。而之前的卷积层并不依赖于输入图片的尺寸，从而实现了任意尺寸的输入。实际上是将原图片多尺度采样输入带 SPP 层的 CNN 进行训练，也是被称为金字塔的原因。

之后 RBG 又提出了新的 Fast-RCNN [1]，借鉴了 SPP 的思路提出了 ROI 池化层，以及将 SVM 分类改为使用 softmax 进行分类，并将分类和边框回归整合，不再独立进行训练。这个算法将除了候选框提取的所有步骤整合在一起进行训练，并引入类似 SPP 的池化层解决了不同尺寸的输入问题，使得训练过程大大提高了。

之后 Faster-RCNN [7] 又更进一步，提出了 RPN 解决了候选框提取的问题。RPN 的特点在于不是在原图上进行候选框提取，而是在特征图上进行。原图通过 CNN 后首先在特征图上进行候选框提取，并将候选框进行分类，只将感兴趣的区域输入到 ROI 池化并进行下一步的分类学习。这样做可以让网络自己学习生成候选区域，大大减少选取候选区域的冗余，提高了预测时间，使得预测可以做到实时。至此候选框选取，CNN，ROI 池化，分类与边框回归都整合到一起训练。

YOLO [6] 则使用了另外一个思路，直接将整个图像进行训练，不预先进行候选框提取。将整个图像分为 $S \times S$ 的网格，物体的中心所在的网格负责该物体的检测，直接经过神经网络得到输出，输出包含物体位置、类别和置信度信息。YOLO 全称为 You Only Look Once，体现了该算法的简介和迅速。该算法相对于 RCNN 系列的算法拥有检测速度快和背景误检率低等优势，但在准确率和物体位置精度上较差。而且 YOLO 只在一个网格尺度上进行回归，缺乏多尺度信息，容易丢失小目标。

SSD [5] 在 YOLO 之上做了许多改进，采用了多尺度特征图的检测来适应不同大小的物体，最后的输出不是使用全连接层而是用卷积来取得检测结果，同时引入了 Faster R-CNN 中 anchor 的概念，设置不同长宽比和尺寸的先验框。这些改进使得 SSD 同时获得了较高的准确率和速度。

2 数据处理

3 网络结构与算法

relu 激活 conv1, pool1, conv2, pool2, spp, fc 阶梯衰减学习率过拟合，正则化滑动平均 spp lrn batch_normalization 批标准化

sigmoid cross entropy loss

[2] [4] [1] [7] [3] [6] [5] [8]

4 结果分析

5 改进

References

[1] Ross B. Girshick. Fast R-CNN. *CoRR*, abs/1504.08083, 2015.

[2] Ross B. Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *CoRR*, abs/1311.2524, 2013.

[3] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B. Girshick. Mask R-CNN. *CoRR*, abs/1703.06870, 2017.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. *CoRR*, abs/1406.4729, 2014.

[5] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott E. Reed, Cheng-Yang Fu, and Alexander C. Berg. SSD: single shot multibox detector. *CoRR*, abs/1512.02325, 2015.

[6] Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. *CoRR*, abs/1506.02640, 2015.

[7] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster R-CNN: towards real-time object detection with region proposal networks. *CoRR*, abs/1506.01497, 2015.

[8] Zhi Tian, Weilin Huang, Tong He, Pan He, and Yu Qiao. Detecting text in natural image with connectionist text proposal network. *CoRR*, abs/1609.03605, 2016.