

计算机视觉课程报告

学号： 10185102144

姓名： 董辰尧

专业名称： 计算机科学与技术

学生年级： 2018 级本科生

课程性质： 专业选修

研修时间： 2020～2021 学年第 2 学期

计算机科学与技术学院

2021 年 6 月

课程内容统计

- 请自评你的项目完成情况，在表中相应位置划√。

课程学习自我评价

| 内容\评价 | 阅读文献 0—5 篇 | 阅读文献 5—10 篇 | 阅读文献 10 篇以上 | 代码 实现 |
|-------------|---------------|----------------|----------------|----------|
| 第14章 智能识别技术 | √ | | | |

总体课程学习情况自我评价

| 完成情况 | 尚 未 完 成 | 基 本 完 成 | 较 好 完 成 | 圆 满 完 成 |
|------|------------|------------|------------|------------|
| 总体情况 | | √ | | |

第 15 章 智能识别技术

一、这一章学习中你的工作

这一章我阅读了智能识别技术的相关文献，并且尝试实现了其中的算法。

二、查阅文献清单

格式：

[1] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.

[2] He K, Zhang X, Ren S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern

三、文献解读

1. 文献 1

(a) 文献名: Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.

(b) 主要创新思想

主流的目标检测方法在 detection network 已经取得了不错的运行时间方面的优化, 因此 region proposal algorithms 成为了一个计算瓶颈。作者本文的关键工作就是提出了 Region Proposal Network 用来进行候选区域提取, 这样可以大大缩短候选区域的提取时间。

(c) 主要原理剖析及说明

目标检测网络基本上可以分成两部分: 候选区域生成(Region proposals) + 对候选区域进行分类以及 bound boxes 回归。

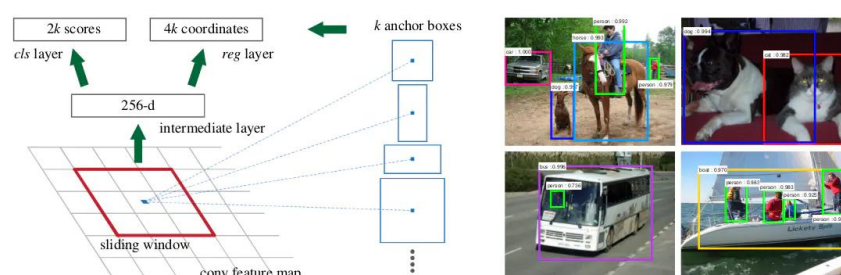


Figure 3: Left: Region Proposal Network (RPN). Right: Example detections using RPN proposals on PASCAL VOC 2007 test. Our method detects objects in a wide range of scales and aspect ratios.

对于一副输入图像, 将其送入 CNN 提取特征, 然后利用 Region Proposal Network 提取候选区域 (这个过程利用了刚才说的 Anchor 机制, 后面细说)。对每一个候选区域进行分类和 bounding box 回归, 就可以得到检测的结果。不过这里候选区域提取是在 CNN 的最后一个 feature map 上提取的, 那么问题来了, 提取的候选区域怎么映射对应回原图呢? feature map 上的框怎么对应上输入图像的框? 实际上, 就是很直接粗暴的方式: 根据特征图和原图的比例, 把特征图上的框按比例进行缩放。

具体的损失函数定义如下:

[2]. Our loss function for an image is defined as:

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*).$$

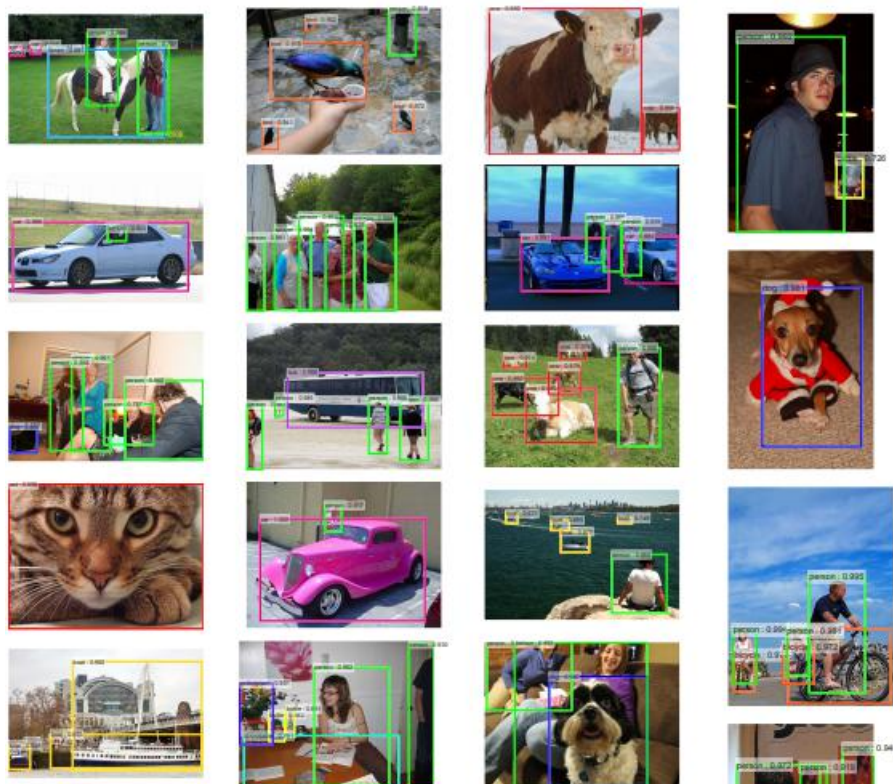
(d) 主要实验结果（现有原文章中的）

Table 6: Results on PASCAL VOC 2007 test set with Fast R-CNN detectors and VGG-16. For RPN, the train-time proposals for Fast R-CNN are 2000. RPN* denotes the unsharing feature version.

| method | # box | data | mAP | areo | bike | bird | boat | bottle | bus | car | cat | chair | cow | table | dog | horse | mbike | person | plant | sheep | sofa | train | tv |
|--------|-------|------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| SS | 2000 | 07 | 66.9 | 74.5 | 78.3 | 69.2 | 53.2 | 36.6 | 77.3 | 78.2 | 82.0 | 40.7 | 72.7 | 67.9 | 79.6 | 79.2 | 73.0 | 69.0 | 30.1 | 65.4 | 70.2 | 75.8 | 65.8 |
| SS | 2000 | 07+12 | 70.0 | 77.0 | 78.1 | 69.3 | 59.4 | 38.3 | 81.6 | 78.6 | 86.7 | 42.8 | 78.8 | 68.9 | 84.7 | 82.0 | 76.6 | 69.9 | 31.8 | 70.1 | 74.8 | 80.4 | 70.4 |
| RPN* | 300 | 07 | 68.5 | 74.1 | 77.2 | 67.7 | 53.9 | 51.0 | 75.1 | 79.2 | 78.9 | 50.7 | 78.0 | 61.1 | 79.1 | 81.9 | 72.2 | 75.9 | 37.2 | 71.4 | 62.5 | 77.4 | 66.4 |
| RPN | 300 | 07 | 69.9 | 70.0 | 80.6 | 70.1 | 57.3 | 49.9 | 78.2 | 80.4 | 82.0 | 52.2 | 75.3 | 67.2 | 80.3 | 79.8 | 75.0 | 76.3 | 39.1 | 68.3 | 67.3 | 81.1 | 67.6 |
| RPN | 300 | 07+12 | 73.2 | 76.5 | 79.0 | 70.9 | 65.5 | 52.1 | 83.1 | 84.7 | 86.4 | 52.0 | 81.9 | 65.7 | 84.8 | 84.6 | 77.5 | 76.7 | 38.8 | 73.6 | 73.9 | 83.0 | 72.6 |
| RPN | 300 | COCO+07+12 | 78.8 | 84.3 | 82.0 | 77.7 | 68.9 | 65.7 | 88.1 | 88.4 | 88.9 | 63.6 | 86.3 | 70.8 | 85.9 | 87.6 | 80.1 | 82.3 | 53.6 | 80.4 | 75.8 | 86.6 | 78.9 |

Table 7: Results on PASCAL VOC 2012 test set with Fast R-CNN detectors and VGG-16. For RPN, the train-time proposals for Fast R-CNN are 2000.

| method | # box | data | mAP | areo | bike | bird | boat | bottle | bus | car | cat | chair | cow | table | dog | horse | mbike | person | plant | sheep | sofa | train | tv |
|--------|-------|------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| SS | 2000 | 12 | 65.7 | 80.3 | 74.7 | 66.9 | 46.9 | 37.7 | 73.9 | 68.6 | 87.7 | 41.7 | 71.1 | 51.1 | 86.0 | 77.8 | 79.8 | 69.8 | 32.1 | 65.5 | 63.8 | 76.4 | 61.7 |
| SS | 2000 | 07+12 | 68.4 | 82.3 | 78.4 | 70.8 | 52.3 | 38.7 | 77.8 | 71.6 | 89.3 | 44.2 | 73.0 | 55.0 | 87.5 | 80.5 | 80.8 | 72.0 | 35.1 | 68.3 | 65.7 | 80.4 | 64.2 |
| RPN | 300 | 12 | 67.0 | 82.3 | 76.4 | 71.0 | 48.4 | 45.2 | 72.1 | 72.3 | 87.3 | 42.2 | 73.7 | 50.0 | 86.8 | 78.7 | 78.4 | 77.4 | 34.5 | 70.1 | 57.1 | 77.1 | 58.9 |
| RPN | 300 | 07+12 | 70.4 | 84.9 | 79.8 | 74.3 | 53.9 | 49.8 | 77.5 | 75.9 | 88.5 | 45.6 | 77.1 | 55.3 | 86.9 | 81.7 | 80.9 | 79.6 | 40.1 | 72.6 | 60.9 | 81.2 | 61.5 |
| RPN | 300 | COCO+07+12 | 75.9 | 87.4 | 83.6 | 76.8 | 62.9 | 59.6 | 81.9 | 82.0 | 91.3 | 54.9 | 82.6 | 59.0 | 89.0 | 85.5 | 84.7 | 84.1 | 52.2 | 78.9 | 65.5 | 85.4 | 70.2 |



2. 文献 2

(a) 文献名: He K, Zhang X, Ren S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9):1904-16.

(b) 主要创新思想

spp 的思想如下图所示, 主要就是利用了金字塔的思想, 将原始 CNN 中得到的特征图复制三份, 第一份平均分割成 4×4 的网格, 每个网格计算平均值输出, 则会得到一个 16 维度的向量。第二份平均分成 2×2 的网格, 计算网格平均值, 得到 4 维向量, 最后一份直接计算平均值, 得到一维向量, 将三个向量拼接起来, 得到 21 维向量, 且不管图像尺寸为多大, 都能固定得到 21 维向量。其实这里存在一个问题, 如果 feature map 除以 4 的时候除不尽, 会导致定位精度的不准确, 在 mask rcnn 中用 align pool 改进了。

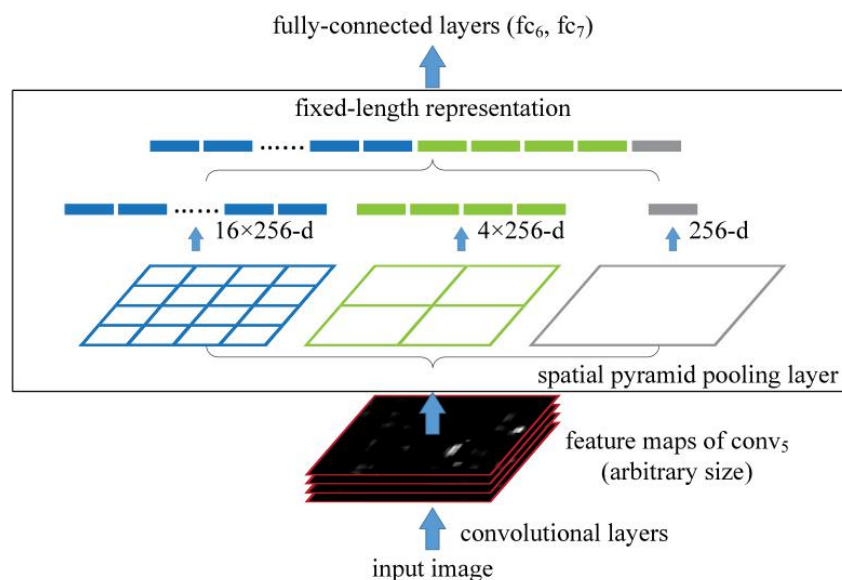
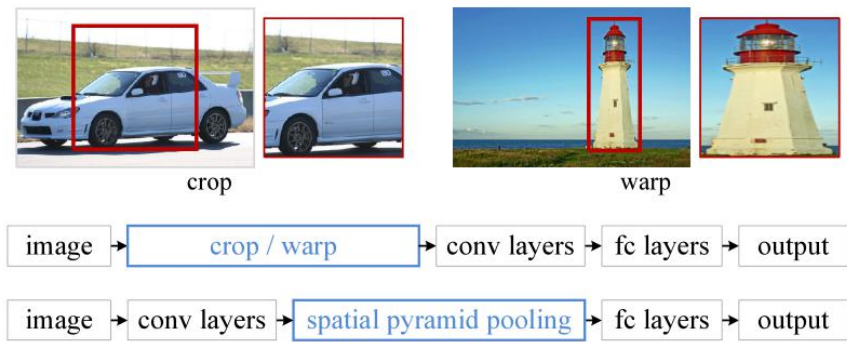


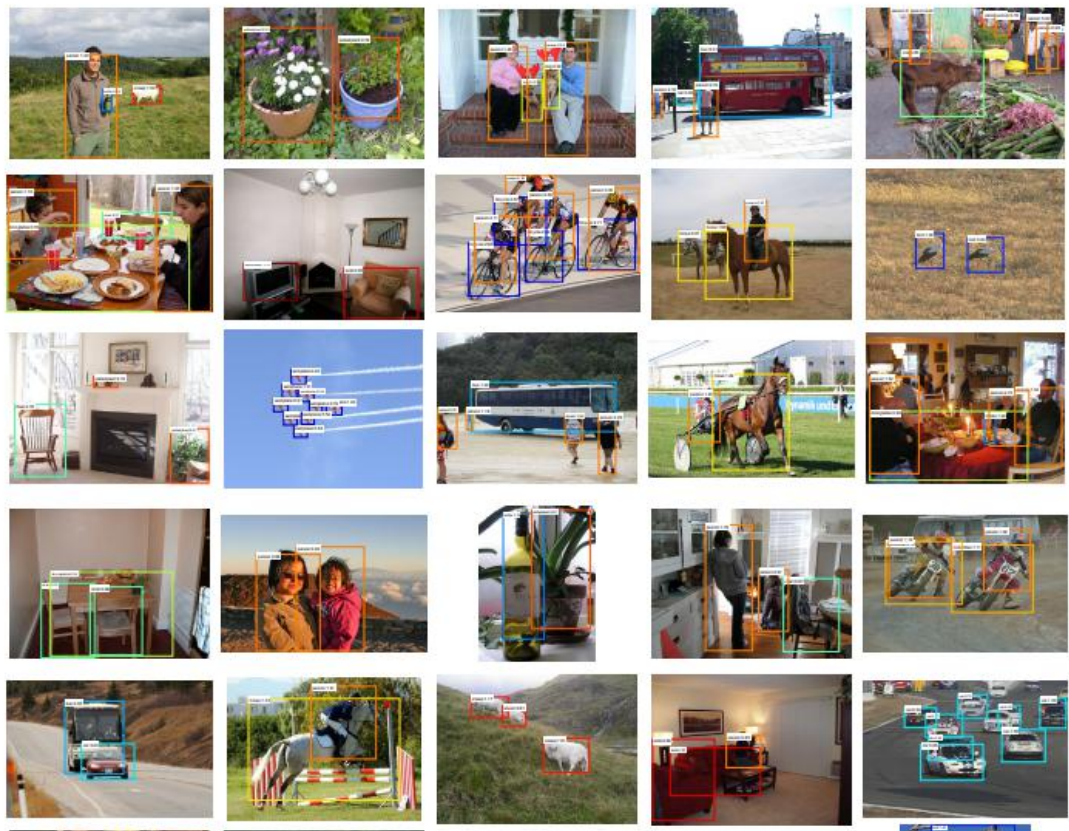
Figure 3: A network structure with a **spatial pyramid pooling layer**. Here 256 is the filter number of the conv₅ layer, and conv₅ is the last convolutional layer.

(c) 主要原理剖析及说明

主要流程图如下所示:



(d) 主要实验结果（现有原文章中的）



三、本章学习小结

这一章我阅读了智能识别技术的相关文献，了解到很多智能分割领域的前沿知识，大概了解了整个领域的研究情况。

第 11 章 运动跟踪

一、这一章学习中你的工作

二、查阅文献清单

格式:

[1] Jalal D K , Ganesan R , Merline A . Fuzzy-C-Means Clustering Based Segmentation and CNN-Classification for Accurate Segmentation of Lung Nodules[J]. asian pacific journal of cancer prevention apjcp, 2017, 18(7):1869-1874.

[2].....

三、文献解读

1. 文献 1

(a) 文献名: Jalal D K , Ganesan R , Merline A . Fuzzy-C-Means Clustering Based Segmentation and CNN-Classification for Accurate Segmentation of Lung Nodules[J]. asian pacific journal of cancer prevention apjcp, 2017, 18(7):1869-1874.

(b) 主要创新思想

.....

(c) 主要原理剖析及说明

.....包括框架图及原理图、损失说明等

(d) 主要实验结果（现有原文章中的）

(e) 代码（如果有写一下具体内容：实现、复现、配置运行？）

2. 文献 2

.....

3. 文献 3

三、本章学习小结

你学到的知识、你的收获和体会

四、附录

