# 目标检测文献综述

杜胜盛 1832957

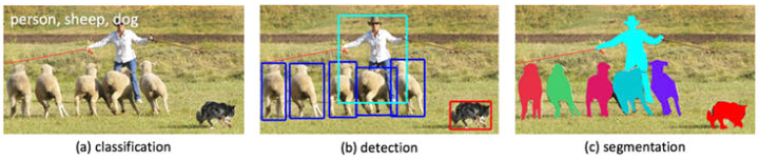
摘要：

关键词：

## 引言：

如何从图像中解析出可供计算机理解的信息，是机器视觉的中心问题。深度学习模型由于其强大的表示能力，加之数据量的积累和计算力的进步，成为机器视觉的热点研究方向。

那么，如何理解一张图片？根据后续任务的需要，有三个主要的层次。



一是分类（Classification），即是将图像结构化为某一类别的信息，用事先确定好的类别(string)或实例ID来描述图片。这一任务是最简单、最基础的图像理解任务，也是深度学习模型最先取得突破和实现大规模应用的任务。其中，ImageNet是最权威的评测集，每年的ILSVRC催生了大量的优秀深度网络结构，为其他任务提供了基础。在应用领域，人脸、场景的识别等都可以归为分类任务。

二是检测（Detection）。分类任务关心整体，给出的是整张图片的内容描述，而检测则关注特定的物体目标，要求同时获得这一目标的类别信息和位置信息。相比分类，检测给出的是对图片前景和背景的理解，我们需要从背景中分离出感兴趣的目标，并确定这一目标的描述（类别和位置），因而，检测模型的输出是一个列表，列表的每一项使用一个数据组给出检出目标的类别和位置（常用矩形检测框的坐标表示）。

三是分割（Segmentation）。分割包括语义分割（semantic segmentation）和实例分割（instance segmentation），前者是对前背景分离的拓展，要求分离开具有不同语义的图像部分，而后者是检测任务的拓展，要求描述出目标的轮廓（相比检测框更为精细）。分割是对图像的像素级描述，它赋予每个像素类别（实例）意义，适用于理解要求较高的场景，如无人驾驶中对道路和非道路的分割。

本文关注的领域是目标检测，即图像理解的中层次。

## 研究现状：

### 传统目标检测：



如上图所示，传统目标检测的方法一般分为三个阶段：首先在给定的图像上选择一些候选的区域，然后对这些区域提取特征，最后使用训练的分类器进行分类。下面我们对这三个阶段分别进行介绍。

(1) 区域选择：这一步是为了对目标的位置进行定位。由于目标可能出现在图像的任何位置，而且目标的大小、长宽比例也不确定，所以最初采用滑动窗口的策略对整幅图像进行遍历，而且需要设置不同的尺度，不同的长宽比。这种穷举的策略虽然包含了目标所有可能出现的位置，但是缺点也是显而易见的：时间复杂度太高，产生冗余窗口太多，这也严重影响后续特征提取和分类的速度和性能。（实际上由于受到时间复杂度的问题，滑动窗口的长宽比一般都是固定的设置几个，所以对于长宽比浮动较大的多类别目标检测，即便是滑动窗口遍历也不能得到很好的区域）

(2) 特征提取：由于目标的形态多样性，光照变化多样性，背景多样性等因素使得设计一个鲁棒的特征并不是那么容易。然而提取特征的好坏直接影响到分类的准确性。（这个阶段常用的特征有SIFT、HOG等）

 (3) 分类器：主要有SVM, Adaboost等。

总结：传统目标检测存在的两个主要问题： 一是基于滑动窗口的区域选择策略没有针对性，时间复杂度高，窗口冗余； 二是手工设计的特征对于多样性的变化并没有很好的鲁棒性。

### 基于深度学习的目标检测：

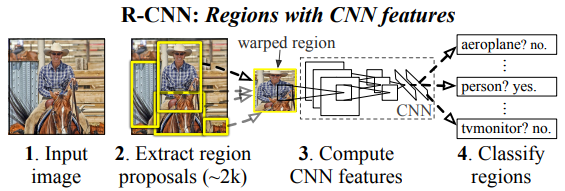
目前主流的目标检测算法主要是基于深度学习模型，其可以分成两大类：two-stage检测算法；one-stage检测算法。目标检测模型的主要性能指标是**检测准确度和速度**，对于准确度，目标检测要考虑物体的定位准确性，而不单单是分类准确度。一般情况下，two-stage算法在准确度上有优势，而one-stage算法在速度上有优势。不过，随着研究的发展，两类算法都在两个方面做改进。

two-stage:

two-stage检测算法将检测问题划分为两个阶段，首先产生候选区域（region proposals），然后对候选区域分类（一般还需要对位置精修），这类算法的典型代表是基于region proposal的R-CNN系算法，如R-CNN, SPPNet, Fast R-CNN, Faster R-CNN, OHEM,R-FCN, FPN, Mask R-CNN等。

**R-CNN**

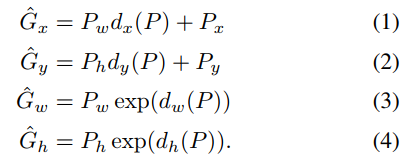
**网络结构：**

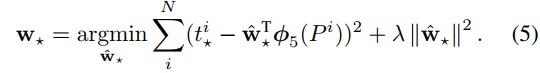


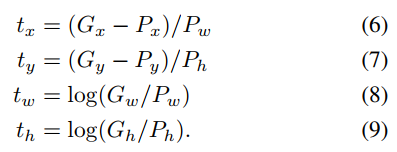
（TODO：对图片展开讲解）

**主要贡献：**（1）用CNN提取建议框的特征。避免了人工设计特征算法的盲目性。（2）任务数据不充足时，可在大容量的公共数据集上做有监督预训练，然后在任务数据上微调，效果也不差。

**损失函数：**



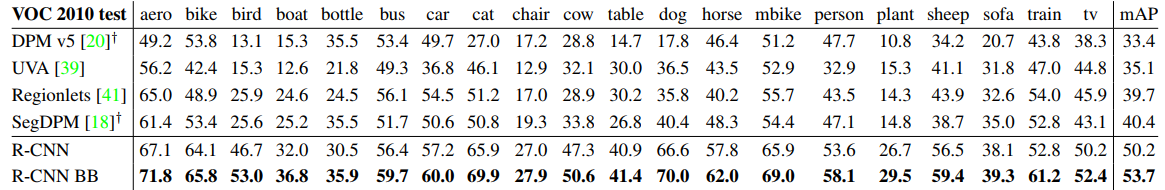


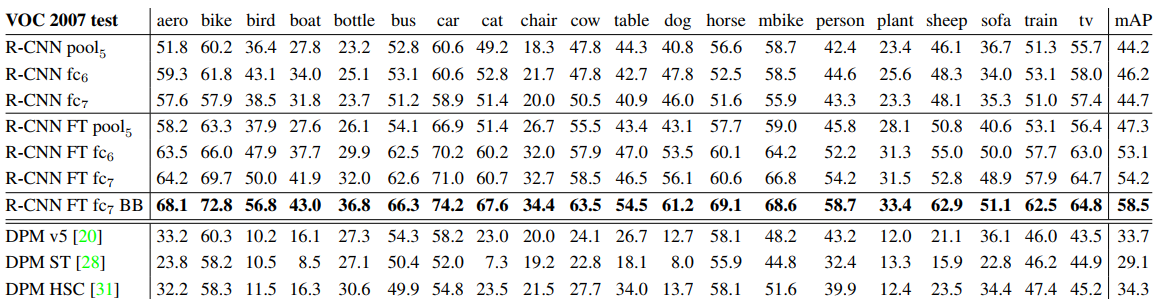


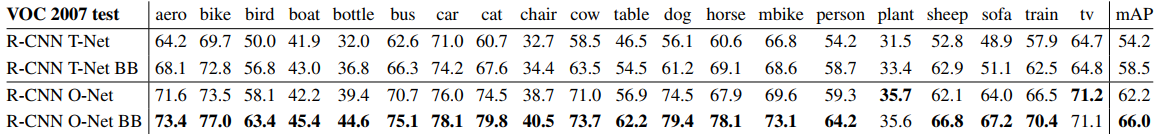


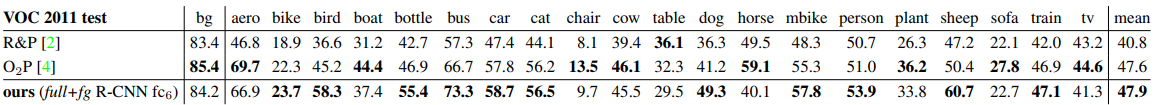
**结果比较：**

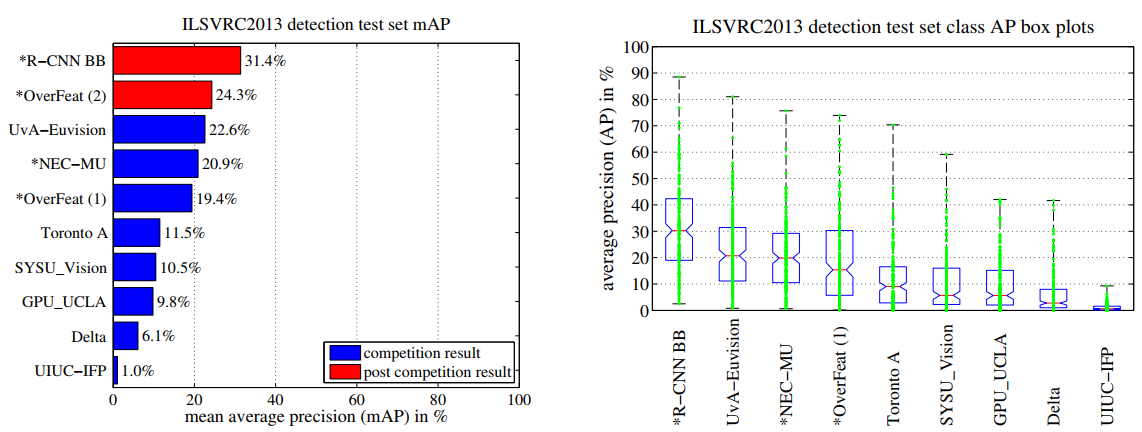
(13s/image on a GPU or 53s/image on a CPU)











one-stage:

one-stage检测算法，其不需要region proposal阶段，直接产生物体的类别概率和位置坐标值，经过单次检测即可直接得到最终的检测结果，因此有着更快的检测速度，比较典型的算法如OverFeat，YOLO，YOLOv2，YOLOv3,SSD，DSSD，DSOD，RetinaNet, M2Det。

RefineDet (结合两者)

### 常用的专业名称（评估标准）：

### 网络结构：

### 损失函数：

### 结果比较：

## 面临的挑战：

## 结论：