

在计算机视觉领域中有几个基本的任务：图像分类、目标检测、实例分割及语义分割，其中目标检测作为计算机视觉中最基本的任务在近年来引起了广泛关注。某种意义上，它在过去二十年内的的发展也是计算机视觉发展史的缩影。

目标检测是一项计算机视觉任务。正如视觉对于人的作用一样，目标检测旨在解决计算机视觉应用中两个最基本的问题：1. 该物体是什么？2. 该物体在哪里？近年来的目标检测技术中也越来越被重视第三个深入性的问题——该物体在干什么？

在深度学习浪潮到来之前，目标检测精度的进步十分缓慢，靠传统依靠手工特征的方法来提高精度已是相当困难的事。而 ImageNet 分类大赛出现的卷积神经网络（CNN）——AlexNet 所展现的强大性能，吸引着学者们将 CNN 迁移到了其他的任务，这也包括着目标检测任务，近年来，出现了很多目标检测的方法，其中 YOLO，SSD，Retina Net 都是 one-stage 方法，而原始的 R-CNN 是 multi-stage 方法，它的延伸 Fast R-CNN 以及 Faster R-CNN 则是 two-stage 方法。RCNN 系列方法是先生成候选框，然后根据候选框来进行坐标回归预测，而 YOLO，SSD，Retina Net 则是直接进行回归生成坐标回归，没有经过候选框这一步。