There are several basic tasks in the field of computer vision: image classification, object detection, instance segmentation and semantic segmentation. Object detection, as the most basic task in computer vision, has attracted wide attention in recent years.In a sense, its development in the past two decades is also a microcosm of the development history of computer vision.

Just as vision is for people, object detection aims to solve two of the most fundamental problems in computer vision applications: 1.What is the object?2. Where is the object?In recent years, more and more attention has been paid to the third profound question—what is the object doing?

Before the advent of deep learning, the progress of object detection accuracy is very slow, and it is very difficult to improve the accuracy by traditional methods that rely on manual features.The powerful performance of AlexNet, the convolutional neural network which is called CNN emerging from ImageNet classification contest, attracts scholars to migrate CNN to other tasks, including target detection. In recent years, many target detection methods have appeared, among which, YOLO, SSD,The Retina Net are both one-stage methods, while the original R-CNN is a multi-stage method, and its extensions Fast R-CNN and Faster R-CNN are two-stage methods.The R-CNN series method is to convert the candidate box and then perform coordinate regression prediction according to the candidate box, while YOLO, SSD and Retina Net directly take the regression to generate coordinate regression without going through the candidate box.