

# 智能车目标检测方法综述

## 1 引言

### 1.1 智能车目标检测技术简介

智能车目标检测技术，是指在自动驾驶等应用场景中，使用计算机视觉技术和深度学习技术，对周围的环境进行感知、识别，自动识别和定位道路上的行人、车辆、道路标志等目标的技术[1]。智能车目标检测技术是自动驾驶汽车中最基础的技术之一，也是实现自动驾驶的关键技术之一。

智能车目标检测技术的主要应用场景是自动驾驶，目前，自动驾驶技术主要包括环境感知、定位导航、轨迹规划和执行控制 4 个方面[2]。其中，环境感知技术是智能车辆系统中的基础环节，为自动驾驶的决策和控制提供外界环境信息，图 2 展示了一种常见的环境感知系统[3]。准确、可靠的环境感知是智能车安全性、智能性的保障，而目标检测是环境感知的重要内容[4]。

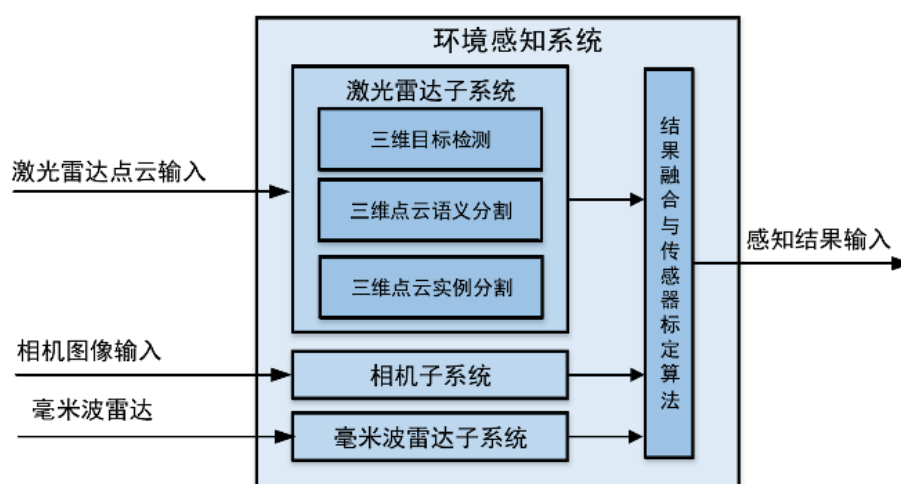


图 1 一种常见的环境感知系统[3]

### 1.2 智能车目标检测数据集与评估指标

智能车目标检测数据集是指用于训练和测试智能车目标检测算法的数据集，其中包含了车辆、行人、交通标志等多种道路上常见的物体，包括高分辨率立体图像、高精度的定位信息等多种类型的数据[5]。目标检测常用数据集有：COCO(Common Objects in Context)[6]、PASCAL VOC (Visual Object Classes)[7]、KITTI[8]等。

恰当的智能车目标检测评估标准，为不同算法之间的比较提供了平台，有助于智能车目标检测方法的成功[9]。通常采用的评估指标如表 1 所示。

表 1 目前常用目标检测性能评价指标[10]

评价指标	含义	公式	备注
准确率 (Accuracy)	正确预测的样本占所有样本的比重	$A=(TP+TN)/(TP+TN+FP+FN)$	TP: 被分类器正确检测为正例的数量; TN: 被分类器正确检测为负例的数量; FP: 被分类器错误地检测为正例的数量; FN: 被分类器错误地检测为负例的数量。
精确率 (Precision)	正确检测目标数占所有被检测目标数的比例	$P=TP/(TP+FP)$	
召回率 (Recall)	正确检测目标数占所有实际目标数的比例	$R=TP/(TP+FN)$	

平均精度 AP (Average Precision)	准确率-召回率曲线与 x 轴构成的面积, 值越高代表模型性能越好	$AP = \int_0^1 p dR$	对各类别的 AP 求平均值, 可得平均精度均值 mAP(Mean Average Precision)
F-指标 (F-Measure)	精确率和召回率的调和平均值	$F_\beta = (1 + \beta^2) \frac{P \times R}{\beta^2 \times P + R}$	$\beta=1$ 时, 为常用的 F1-指标
交并比 IoU (intersection-over-union)	表示产生的候选框和原标记框的重叠度, 即它们交集与并集的比值	$IoU = \frac{X \cap Y}{X \cup Y}$	X 为检测算法预测的候选框区域, Y 为原标记框区域。

智能车目标检测数据集与评估标准的意义在于提高智能车安全性能, 评估智能车目标检测算法的性能和可靠性。通过评估智能车目标检测算法的准确性、速度、鲁棒性、适用性和可扩展性等指标, 可以对算法的性能和可靠性进行评估, 为算法的改进和优化提供指导[5][11]。

## 2 智能车目标检测常见传感器

目标检测技术需要使用多种传感器来采集数据。无论是多传感器融合方案[12]还是深度学习方案[16], 都需要使用传感器来获取输入数据, 这些数据可以是图像、雷达点云等。因此, 传感器在自动驾驶目标检测中仍然起着至关重要的作用。图 2 展示了常用传感器在智能车上的相对作用距离和角度范围, 表 2 列举了常用传感器的功能特点和应用。

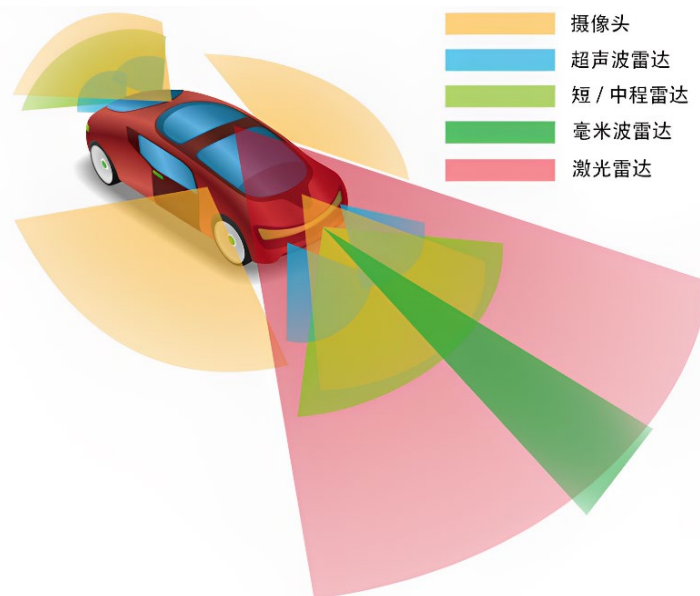


图 2 常用传感器相对作用距离和角度范围[13]

表 2 常用传感器的功能特点和应用[13]

传感器类型	功能特点	应用场景
摄像头	成本低、易于安装和使用; 可用于识别道路标志、车辆、行人等	路面信息获取、交通标志识别、行人检测、车辆检测等
超声波雷达	成本低、可靠性高; 可用于近距离避障、停车辅助	近距离避障、停车辅助
短/中程雷达	成本适中、精度较高、对天气等环境条件的影响较小; 可用于车辆检测、车道检测、障碍物检测等	车辆检测、车道检测、障碍物检测等

毫米波雷达	精度高、天气等环境条件对其影响较小；可用于长距离目标检测和跟踪，如车辆、行人和障碍物等	长距离目标检测和跟踪，如车辆、行人和障碍物等
激光雷达	精度高、分辨率高，能够应对多种场景下的目标检测和跟踪；可用于建立高精度的环境地图，提供车辆定位和路径规划	目标检测和跟踪、建立高精度的环境地图、提供车辆定位和路径规划等

### 3 智能车目标检测方法概述

智能车目标检测方法，是自动驾驶汽车、智能交通系统等应用场景中的重要组成部分，它可以通过计算机视觉技术，从图像或视频中检测出车辆、行人、路标等目标，实现对环境的感知和决策[1]。智能车目标检测方法大致可分为传统方法和深度学习方法两类[14]。

传统方法中，主要利用多传感器融合方案，弥补各传感器缺陷，实现优势互补。这些方法在早期应用中取得了一定的成功，但是在复杂场景下的检测准确率和鲁棒性不足[15]，无法满足实际应用需求。

近年来，深度学习方法[16]在智能车目标检测中得到了广泛应用。其中，卷积神经网络（Convolutional Neural Network, CNN）[17]在图像特征提取中表现优异，而区域提议网络（Region Proposal Network, 简称 RPN）[18]方法常用于生成目标候选框。这些方法利用深度神经网络[19]进行特征学习和目标检测，准确率和鲁棒性均有较大提高，但也存在一些局限性和挑战，需要进一步研究和改进。深度学习方法通常都是基于图像的[20]，因为图像是最常用的目标检测输入数据。不过，也有一些方法使用的不是图像数据，一个例子是基于点云数据的目标检测方法[21]。

目前，智能车目标检测方法的研究仍在不断发展中，一些新的方法不断涌现，如基于自监督学习的目标检测方法[22]等。

### 4 单传感器方案的目标检测技术实例

对于摄像头传感器，采集到的数据属于图像数据，其技术实例主要有基于视觉的道路检测技术[23]、车道线检测技术[24]、障碍物目标检测技术[25] [14]。对于各种雷达，主要有基于雷达的激光雷达检测[26]、毫米波雷达检测技术[27]等。图3展示了道路检测技术目标检测的实例，图4展示了车道线检测技术、障碍物目标检测的实例。图5展示了激光雷达道路点云数据图及毫米波雷达俯视点云图。单传感器方案的目标检测技术所获得的图像或点云，为多传感器融合方案和深度学习方案提供输入。

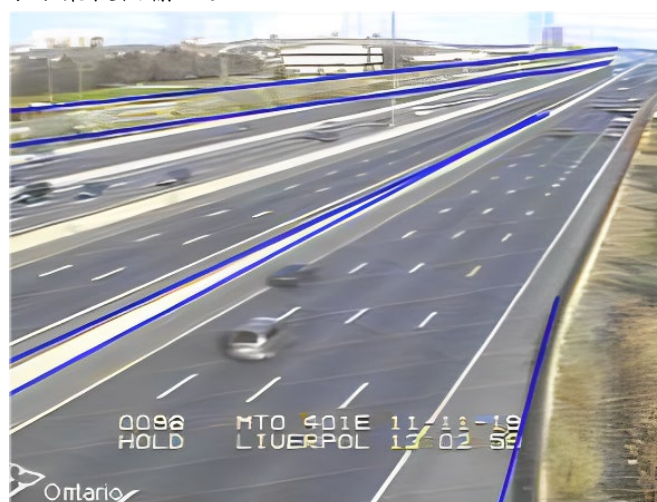


图3 道路检测技术目标检测实例[23]



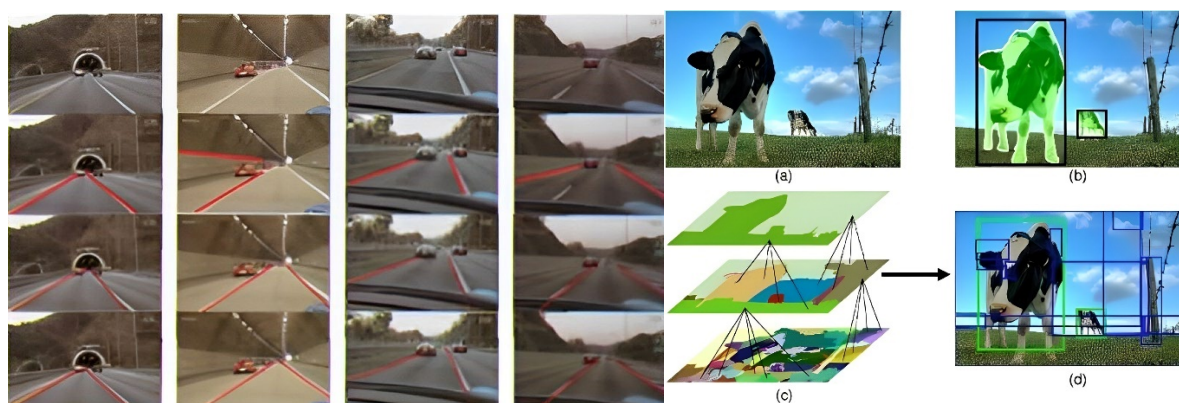


图4 车道线检测技术、障碍物目标检测实例[24][25]

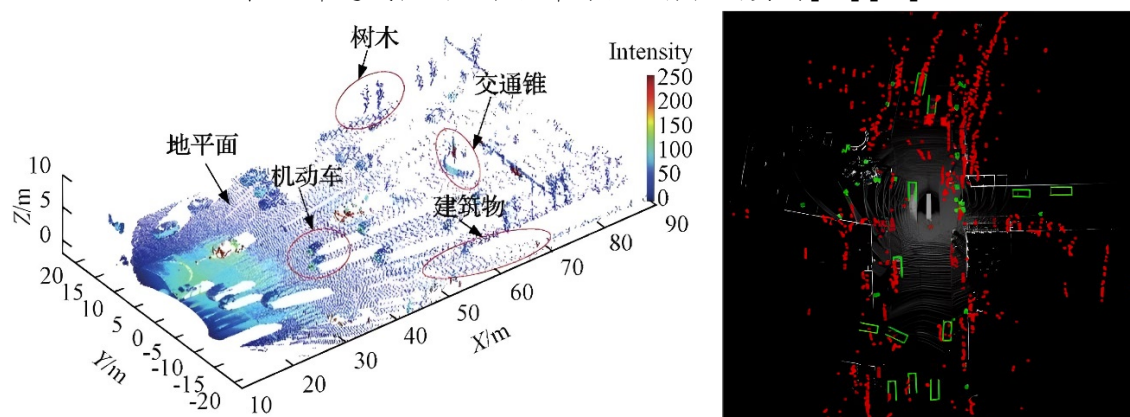


图5 激光雷达道路点云数据[26]、毫米波雷达俯视点云图[27]

## 5 基于多传感器融合方案的目标检测方法

对于单传感器方案的目标检测方法，其传感器自身存在局限，比如视觉传感器无法获得目标的深度信息，易受光照等外部因素影响等。基于多传感器融合的目标检测方法可以较好的弥补各传感器缺陷，实现优势互补。

根据传感器信息在不同信息层次上的融合，可将多传感器融合分为基于数据融合的目标检测方法[27]，基于多模态特征融合的目标检测方法[27]，基于传感器互补（目标融合）的目标检测方法[27]。三种传感器融合方式示意图如图4所示。

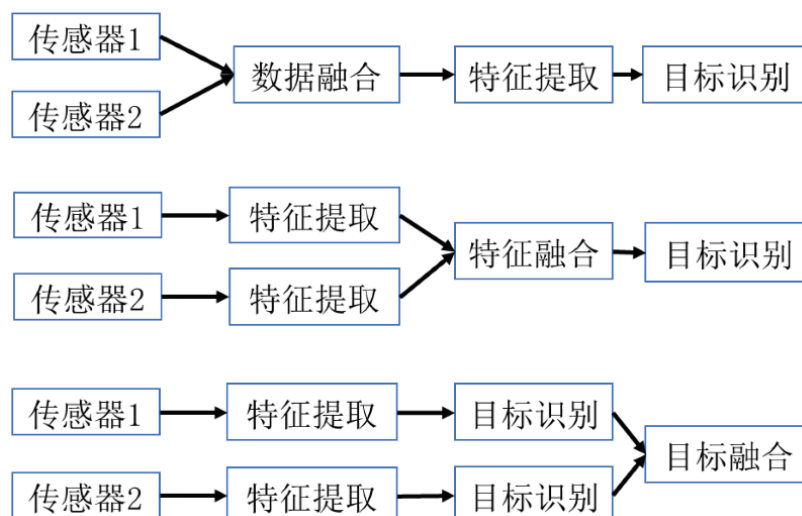


图6 三种传感器融合方式[27]

### 5.1 基于数据融合的目标检测方法

数据融合是一种低层次、集中式融合的结构，它直接将来自不同传感器的原始数据进行整合，然后对融合后的数据进行特征提取和目标检测[28]。数据融合发生在传感器输出端，数据损失较少，因此获得的检测结果也更加精确。但数据层信息量大、处理数据的时间成本巨大，不利于实时处理[28]。

### 5.2 基于多模态特征融合的目标检测方法

多模态特征融合，即先从多个传感器输出的数据进行特征提取，再对获得的特征向量进行融合，从而实现对目标的多维度描述。即先对原始数据进行特征提取、压缩信息量，然后融合各传感器数据的特征向量，并对融合后的数据进行目标检测[29]。

相比于数据融合，特征融合所消耗的内存、时间和数量级相对减少，算法实时性有所提高。

### 5.3 基于传感器互补（目标融合）的目标检测方法

基于传感器互补的目标检测方法，是指通过多个传感器之间的协同作用，将来自多个传感器的信息进行整合和协同处理，从而实现准确的目标检测。在目标融合中，每个传感器有各自独立的检测算法，将各传感器检测后的目标信息输入到融合算法进行智能优化、组合、推理，获得最终的检测结构[30]。

目标层融合的优势在于融合的信息量较小，融合算法运行速度快且可靠性高，其缺点是各传感器都有独立算法对计算平台要求较高[30]。

## 6 基于深度学习方案的智能车目标检测方法

### 6.1 深度学习方案概述

基于深度学习的目标检测方法[16]，是指利用深度学习模型实现目标检测的方法。这种方法通过对目标的外观特征进行学习，并结合运动信息来进行目标跟踪。

目前，基于深度学习的目标检测方法已经成为智能车目标检测的主流方法[16]。其中，基于卷积神经网络（CNN）的目标检测算法，如 Faster R-CNN[18]、YOLO[31]、SSD[32]等，已经得到广泛应用。这些算法通过多层卷积神经网络来提取图像特征，并通过回归或分类的方式来实现目标检测。

卷积神经网络（Convolutional Neural Network, CNN）[17]是一种基于深度学习的神经网络模型，用于图像处理和计算机视觉领域，可以有效地提取图像特征，实现图像分类、目标检测、图像分割等任务，主要结构包括输入层、卷积层、池化层、全连接层和输出层[33]，如图 5 所示。其中，卷积层主要用于提取图像特征，通过卷积核对输入图像进行卷积运算，得到特征图像，从而实现特征提取的功能；池化层则用于降低特征图像的维度，减少计算量，并且可以提高模型的泛化能力；全连接层用于将特征图像转换为分类结果，最终输出分类概率[34]。

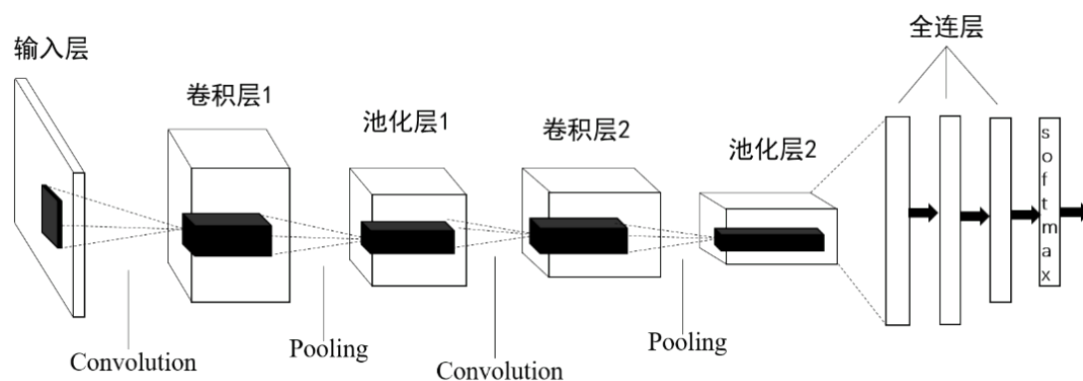


图 7 卷积神经网络结构[33]

最近几年出现一些新的些深度学习目标检测方法,使用的不是图像数据,而是点云数据,例如 PointRCNN[35]和 PV-RCNN[36]。这些方法主要用于处理三维点云数据,可以用于自动驾驶、机器人导航和无人机等应用场景。

依据目标检测流程的不同,目标检测算法可以分为两阶段检测算法和一阶段检测算法[16],以下分别介绍。

## 6.2 两阶段检测算法

两阶段检测算法的典型代表是 RCNN (Region-based Convolutional Neural Network) [20] 系列算法,包括 Fast R-CNN[37]、Faster R-CNN[18]和 Mask R-CNN[38]等。PointRCNN 和 PV-RCNN 与 RCNN 系列算法没有直接的关系,它们不属于 RCNN 系列算法。

两阶段检测算法通常具有更好的定位性能,但需要多次前向传播,处理速度相对较慢。这些算法主要分为两个阶段:候选框提取和分类回归[20]。

在候选框提取阶段,算法会使用一些方法(如 Selective Search[25]、区域提取网络 RPN 等)生成多个可能包含目标的候选框。这些候选框会被送入 CNN 进行特征提取,然后通过 SVM[39]等分类器进行分类和回归,以得到目标的位置和类别信息。

### 6.2.1 R-CNN 目标检测算法

R-CNN (Regions with CNN features) [20]是目标检测领域的经典算法之一。R-CNN 算法的主要思想是将图像中的候选区域提取出来,然后对每个候选区域进行特征提取和分类,从而实现目标检测。R-CNN 算法使用实例如图 8 所示。

R-CNN 算法主要包括候选区域生成、特征提取、分类、位置精修四个步骤[20]:

- 1) 输入图像经过选择性搜索 (Selective Search) 等算法进行候选区域生成;
- 2) 对每个候选区域进行卷积神经网络特征提取;
- 3) 将提取的特征送入支持向量机 (SVM) 进行分类和回归;
- 4) 使用回归器精细修正候选框位置。

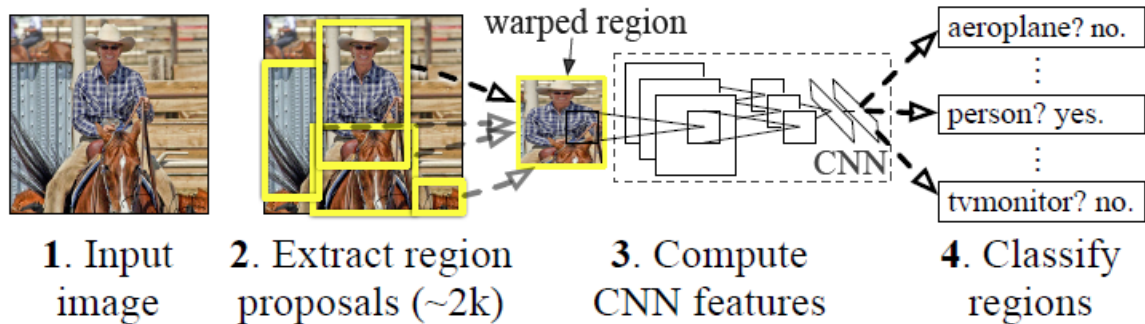


图 8 R-CNN 算法使用实例[20]

在 R-CNN 算法中,候选区域生成是一个比较耗时的过程,因为需要对整张图像进行大量的计算。为了解决这个问题,后来的算法如 Fast R-CNN 和 Faster R-CNN 都使用了 RPN 网络来代替选择性搜索等算法进行候选区域的生成。

### 6.2.2 Fast R-CNN 目标检测算法

Fast R-CNN[37]是 RCNN 系列算法的第二个版本,结构图见图 9。相对于 R-CNN 算法, Fast R-CNN 算法具有更快的检测速度和更高的准确率。

Fast R-CNN 算法的主要特点是将整个图像输入到 CNN 中进行特征提取,而不是对候选区域进行单独的特征提取。其主要流程如下[37]:

- 1) 输入图像经过卷积层和池化层进行特征提取;
- 2) 候选区域通过 RoI 池化层进行特征提取;
- 3) 将候选区域的特征和全局特征进行融合;
- 4) 将融合后的特征送入全连接层进行分类和回归。



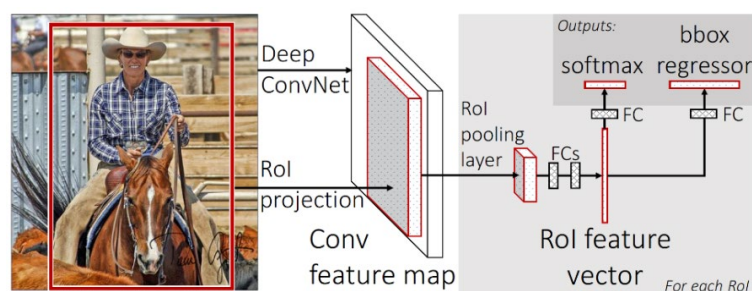


图9 Fast R-CNN 完整结构图[37]

其中，RoI 池化层（Region of Interest Pooling Layer）[37]是一种用于目标检测的神经网络层。在 Fast R-CNN 算法中，RoI 池化层是一个非常重要的部分，用于将任意尺寸的特征图中的区域（候选区域）转换为固定尺寸的特征图，便于后续的分类和回归操作。RoI 池化层的具体操作流程如下[40]：

- 1) 对每个候选区域进行适当的缩放，使其对应到特征图的大小；
- 2) 将候选区域划分为  $N \times N$  个子区域；
- 3) 对每个子区域进行最大池化操作，得到对应的特征值。

相对于 RCNN 算法，Fast R-CNN 算法在 COCO、PASCAL VOC 等数据集上的检测速度有较大进步。但训练 Fast R-CNN 算法时，需要对每个候选区域进行特征提取，计算量较大，而且需要使用负样本挖掘等技巧来平衡正负样本的数量[37]。因此，后续的研究还在继续改进 Fast R-CNN 算法。

### 6.2.3 Faster R-CNN 目标检测算法

Faster R-CNN[18]是一种基于卷积神经网络的目标检测算法，是 R-CNN 系列算法的第三个版本。其核心思想是在卷积神经网络的基础上引入区域提取网络（RPN）[18]，通过 RPN 来生成候选目标区域，然后再通过分类网络和回归网络对这些候选区域进行分类和精确定位。

Faster R-CNN 算法主要包括区域提议网络和检测网络两个部分，其主要流程如下[34]：

- 1) 输入图像经过卷积层和池化层进行特征提取；
- 2) RPN 网络对特征图进行卷积操作，得到多个候选区域；
- 3) 对每个候选区域进行 RoI 池化操作，得到固定大小的特征图；
- 4) 将特征图送入检测网络进行分类和回归。

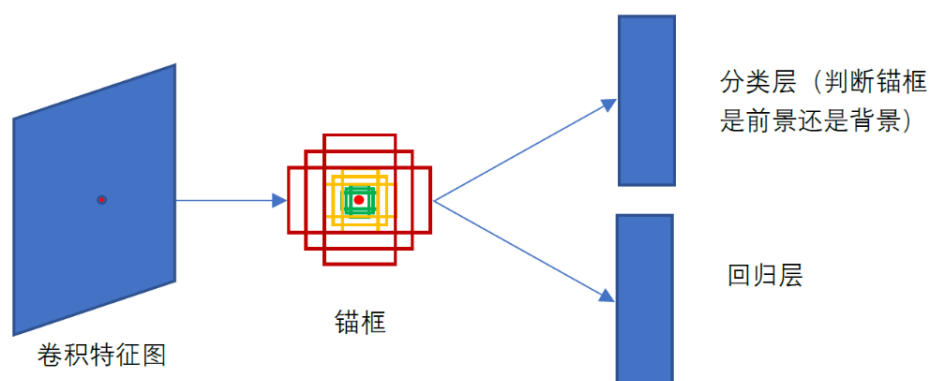


图10 RPN 的结构框架[34]

其中，RPN 网络[34]是 Faster R-CNN 算法的一个重要组成部分，结构框架如图 8 所示。它可以在特征图上生成多个候选区域，并且可以共享特征提取层和卷积层的计算结果。由于 RPN 和分类回归网络共享卷积特征，因此 Faster R-CNN 能够在较快的速度下实现准确的目标检测，从而大大提高了算法的效率。RPN 网络主要通过滑动窗口的方式在特征图上生成多个候选区域，然后使用两个全连接层进行分类和回归，得到每个候选区域的置信度和坐标信息。

相对于 Fast R-CNN 算法, Faster R-CNN 算法在目标检测的速度和准确率方面都有了较大的提高。同时, Faster R-CNN 算法还具有较好的可扩展性, 可以适用于不同的目标检测任务和不同的数据集。

总的来说, Faster R-CNN 算法是目前目标检测领域的主流算法之一, 其在实际应用中取得了广泛的应用, 如自动驾驶等。

#### 6.2.4 Mask R-CNN 目标检测算法

Mask R-CNN[38]是一种基于深度学习的目标检测算法, 是 R-CNN 系列算法的一种改进版。相比于传统的 RCNN 算法, Mask R-CNN 不仅能够实现目标检测, 还能够实现物体实例分割。

Mask R-CNN 算法主要包括三个部分: 特征提取、区域提议网络和 Mask 网络[38]。其主要流程如下:

- 1) 输入图像经过卷积层和池化层进行特征提取;
- 2) 区域提议网络对特征图进行卷积操作, 得到多个候选区域;
- 3) 对每个候选区域进行 RoI 池化操作, 得到固定大小的特征图;
- 4) 将特征图送入 Mask 网络进行物体实例分割。

与 Faster R-CNN 类似, Mask R-CNN 算法也使用区域提议网络 (RPN) 来生成候选区域, 但是在 RPN 之后, Mask R-CNN 还会使用 Mask 网络[38]来对每个候选区域进行像素级别的分割。Mask 网络主要是一种全卷积网络, 它可以将每个候选区域映射到一个固定大小的特征图上, 并对特征图中的每个像素点进行分类, 从而实现物体实例分割。

相对于其他目标检测算法, Mask R-CNN 算法具有较高的准确率和较好的物体实例分割效果[35]。同时, Mask R-CNN 算法还能够处理多个物体之间的重叠和遮挡问题, 具有较好的鲁棒性和可靠性。因此, Mask R-CNN 算法在实际应用中取得了广泛的应用, 如自动驾驶、医学图像分析等领域。

#### 6.2.5 RCNN 系列算法小结

R-CNN 对图像选取若干候选区域, 然后用卷积神经网络对每个候选区域进行特征提取和分类, 再用这些特征预测候选区域的类别和边界框。

Fast R-CNN 相比 R-CNN 的主要改进在于, 使用 RoI 池化层代替原来的 SVM 分类器, 从而加快了目标检测的速度。

Faster R-CNN 将 Fast R-CNN 中的选择性搜索替换成区域提取网络, 从而减少候选区域的生成数量, 并保证目标检测的精度。

Mask R-CNN 在 Faster R-CNN 基础上引入一个全卷积网络, 实现了目标实例分割, 从而借助目标的像素级位置进一步提升目标检测的精度。

### 6.3 一阶段检测算法

一阶段检测算法, 基于图像的典型代表是 YOLO (You Only Look Once) 和 SSD (Single Shot MultiBox Detector) 等, 基于点云数据的典型代表是 PointRCNN 和 PV-RCNN 等。

这些算法只需要一次前向传递就可以同时完成目标的检测和分类, 相比两阶段检测算法, 速度更快、效率更高, 但检测精度可能会稍低[16]。但是, 随着技术的发展和算法的改进, 一阶段检测算法的检测精度也在不断提高。

#### 6.3.1 YOLO 算法

YOLO (You Only Look Once) [31]是一种基于深度学习的实时目标检测方法, 其核心思想是将目标检测问题转化为一个回归问题, 通过卷积神经网络对输入图像进行划分, 同时对每个划分框进行分类和位置回归, 从而实现目标检测。

YOLO 算法将整张图像分成  $S \times S$  个网格, 每个网格预测  $B$  个边界框和每个边界框的置信度和类别概率。对于每个边界框, YOLO 算法使用卷积神经网络 (CNN) 来提取其特征, 并计算其包含物体的概率。同时, YOLO 算法还会对边界框进行位置和尺度的回归, 进一步提高检测



精度。

YOLO 算法的主要优点是速度快、准确率高、可以同时检测多个物体，并且可以处理不同尺度的物体。YOLO 算法的缺点是定位精度不如其他算法高、对于小物体的检测效果较差、对于重叠的物体检测效果也不理想。针对这些问题，后续的研究工作提出了一系列改进的 YOLO 算法，如 YOLOv2、YOLOv3、YOLOv4 等。

总之，YOLO 算法是一种快速且准确的目标检测算法，适用于实时场景下的物体检测任务，具有广泛的应用前景。

### 6.3.2 SSD 算法

SSD (Single Shot MultiBox Detector) [32] 是一种基于深度学习的目标检测方法，其核心思想是在卷积神经网络的基础上引入多尺度特征图，并在每个特征图上进行目标检测，通过对不同尺度的特征图进行融合，来提高目标检测的准确性。

SSD 算法的主要思想是将目标检测视为一个回归问题，通过在每个特征层上预测不同尺度和长宽比的边界框，同时预测每个边界框内包含物体的类别概率。同时，SSD 算法还使用多尺度特征图来检测不同大小的物体，从而提高检测精度。

具体地，SSD 算法的主要流程包括：

1) 在特征图上生成多个默认框 (default boxes) [41]，这些默认框具有不同的尺度和长宽比，并且覆盖了整个图像；2) 对每个默认框，通过卷积神经网络提取其特征，并分别预测其包含物体的概率和位置信息；3) 对每个默认框的类别概率和位置信息进行筛选和调整，得到最终的物体检测结果。

SSD 算法的优点是具有较高的检测精度和较快的检测速度，可以实现实时目标检测。同时，SSD 算法还能够处理不同尺度和长宽比的物体，并且对于小物体的检测效果较好。然而，SSD 算法在处理物体遮挡和重叠等复杂情况时，仍然存在一定的局限性。

### 6.3.3 PointRCNN 算法

PointRCNN 是一种基于深度学习的点云目标检测算法，首次出现在 2019 年。它是第一个在 KITTI 数据集上，实现 3D 目标检测性能超越了基于深度学习的二维图像目标检测的算法[35]。PointRCNN 使用点云数据作为输入，将点云数据转换成三维体素网格[21]，然后使用三维卷积神经网络 (3D CNN) [42] 对体素网格进行处理，最后使用分类器和回归器对目标进行检测和定位。

PointRCNN 的主要贡献在于提出了一种基于体素网格的点云处理方法[21]，将点云数据转换成三维体素网格，并使用体素网格中的点云坐标和特征信息来训练 3D CNN 模型，以实现点云的高效处理和目标检测。此外，PointRCNN 还使用了一种新的点云特征提取方法，称为 RoI-aware 点特征提取 (RoI-aware Point Feature Extractor) [43]，能够有效地提取点云中的目标信息，从而提高目标检测性能。

在 KITTI 数据集上的实验结果表明，PointRCNN 在三维目标检测任务上取得了优异的性能，超越了其他基于深度学习的二维图像目标检测算法[35]。PointRCNN 的成功表明，基于点云数据的目标检测算法具有重要的应用价值，可以应用于自动驾驶、机器人视觉和工业自动化等领域。

### 6.3.4 PV-RCNN 算法

PV-RCNN 是一种基于深度学习的点云目标检测算法，首次出现在 2020 年。与 PointRCNN 类似，PV-RCNN 也使用点云数据作为输入，但它不是将点云数据转换成体素网格，而是将点云数据投影到二维图像上，然后在二维图像上进行目标检测，最后将检测结果映射回到原始点云中。

具体地说，PV-RCNN 首先使用激光雷达的固有属性（如反射强度和距离）将点云数据投影到二维图像上，然后使用二维卷积神经网络 (2D CNN) [44] 对图像进行处理，提取目标的视觉

特征。接着, PV-RCNN 使用双向点特征聚合 (Bilateral Point Feature Aggregation) [45]的方法, 将二维图像中的特征映射回原始点云中, 同时将点云的属性信息 (如法向量和曲率) 与目标的二维特征进行融合。最后, PV-RCNN 使用分类器和回归器对目标进行检测和定位。

PV-RCNN 的主要贡献在于将点云数据和二维图像数据相结合, 充分利用了二维图像处理的优势, 同时保留了点云数据的三维信息。此外, PV-RCNN 使用了新的点云特征聚合方法, 能够有效地提取点云中的目标信息, 从而提高目标检测性能。

在 KITTI 数据集上的实验结果表明, PV-RCNN 在三维目标检测任务上取得了优异的性能, 超越了其他基于深度学习的点云目标检测算法[36]。PV-RCNN 的成功表明, 将点云数据和二维图像数据相结合, 是一种有效的点云目标检测方法, 具有重要的应用价值。

### 6.3.5 一阶段检测算法总结

YOLO 算法采用单一网络结构, 将目标检测问题转化为回归问题, 并通过对图像进行划分来检测目标。YOLO 算法速度快, 可以实现实时检测, 但是在小目标检测和目标密集的情况下, 精度不如 SSD 算法。SSD 算法通过多尺度的特征图来检测目标, 提高了对小目标和密集目标的检测精度, 但相对于 YOLO 算法来说, 速度稍慢。

PointRCNN 和 PV-RCNN 都是基于深度学习的点云目标检测算法, 但在处理点云数据和提取特征方面有所不同。PointRCNN 使用三维体素网格将点云数据转换成三维数据, 然后使用 3D CNN 进行处理, 利用 RoI-aware 点特征提取方法提取目标信息。PV-RCNN 则将点云数据投影到二维图像上, 使用 2D CNN 提取目标的视觉特征, 然后使用双向点特征聚合方法将特征映射回原始点云中, 同时融合点云的属性信息, 最后使用分类器和回归器进行目标检测和定位。

## 7 基于自监督学习的目标检测方法

基于自监督学习的目标检测方法[22]是一种使用自监督学习技术来训练目标检测模型的方法, 是近年来计算机视觉领域的研究热点。

这种方法不需要使用手工标注的训练数据, 而是使用自监督任务来训练模型, 可以生成大量的无标注数据, 从而可以用于训练目标检测模型。在这种方法中, 一般会使用一个预训练的卷积神经网络来提取图像特征, 并使用这些特征来生成候选检测框。然后, 使用一些技术[22] (如半监督学习、迁移学习等) 利用生成的无标注数据来训练目标检测模型。

基于自监督学习的目标检测方法大大减少了数据标注的工作量, 并且可以提高目标检测的准确性, 但性能还没有达到基于有监督学习的方法的水平, 需要进一步提升性能和稳定性。

## 8 总结

本文以“智能车目标检测方法”为主题, 首先简要介绍智能车目标检测方法的概概念, 智能车的系统框架及环境感知功能, 罗列了智能车目前常用数据集及目标检测评估方法, 介绍了目标检测常用的传感器, 让读者对智能车目标检测有初步的认知。接着给出智能车目标检测方法概述, 并展示单传感器目标检测应用实例, 为后文详细介绍多传感器融合方案、深度学习方案做准备。然后详细介绍了多传感器融合方案和深度学习方案的智能车目标检测方法, 文章着重介绍了深度学习方案的目标检测方法, 其中偏重于介绍基于图像的方法。最后简要介绍基于自监督学习的目标检测方法, 使读者了解智能车目标检测方法的未来发展。

## 参考文献

- [1] Zou Z, Chen K, Shi Z, Guo Y, Ye J. Object Detection in 20 Years: A Survey[J]. Proceedings of the IEEE, 2023, 111(3): 257-276.
- [2] 王艺帆. 自动驾驶汽车感知系统关键技术综述[J]. 汽车电器, 2016, No.340(12): 12-16.
- [3] 王博. 基于激光雷达的自动驾驶三维环境感知系统关键技术研究[D]. 中国科学院大学(中国

- 科学院长春光学精密机械与物理研究所),2022.
- [4] Wang G, Wu J, He R, Yang S. A Point Cloud-Based Robust Road Curb Detection and Tracking Method[J]. IEEE Access, 2019, 7: 24611-24625.
  - [5] Huang X, Wang P, Cheng X, Zhou D, Geng Q, Yang R. The ApolloScape Open Dataset for Autonomous Driving and Its Application[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(10): 2702-2719.
  - [6] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context[C]//Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer International Publishing, 2014: 740-755.
  - [7] Everingham M, Van Gool L, Williams C K I, et al. The pascal visual object classes (voc) challenge[J]. International journal of computer vision, 2010, 88: 303-338.
  - [8] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI, USA, 2012: 3354-3361.
  - [9] 曹健,陈怡梅,李海生等.基于深度学习的道路小目标检测综述[J/OL].计算机工程:1-17[2023-04-10].
  - [10] CSDN. 目标检测-语义分割-实例分割模型常用性能评价指标 [EB/OL].(2019-12-29)[2023.4.24]. <https://blog.csdn.net/sdkjfk/article/details/103759401>
  - [11] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]. 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 2012: 3354-3361.
  - [12] 刘明亮. 基于多传感器融合的智能车道路目标检测[D].江苏大学,2021.
  - [13] nidec-components. 自动驾驶上所需要的各种传感器 [EB/OL].(2023-04-01)[2023.5.1]. <https://www.nidec-components.com/cn/featuring/lidar-polygon/various-sensors/>
  - [14] 任凤雷. 基于智能车辆视觉导航的环境感知技术研究[D].中国科学院大学(中国科学院长春光学精密机械与物理研究所),2020.
  - [15] 周昕. 复杂道路场景下多车辆的检测与跟踪算法研究[D].南京邮电大学,2022.
  - [16] 黄卫. 基于深度学习的目标检测应用研究[D].南京邮电大学,2022.
  - [17] Li Z, Liu F, Yang W, Peng S, Zhou J. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(12): 6999-7019.
  - [18] 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.
  - [19] Oliveira G L, Burgard W, Brox T. Efficient deep models for monocular road segmentation[C]. 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Korea (South), 2016: 4885-4891.
  - [20] Girshick R, Donahue J, Darrell T, Malik J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014: 580-587.
  - [21] Qi C R, Su H, Mo K, et al. Pointnet: Deep learning on point sets for 3d classification and segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 652-660.
  - [22] He K, Fan H, Wu Y, Xie S, Girshick R. Momentum Contrast for Unsupervised Visual

- Representation Learning[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020: 9726-9735.
- [23] Helala M A, Pu K Q, Qureshi F Z. Road Boundary Detection in Challenging Scenarios[C]. 2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance, Beijing, China, 2012: 428-433.
- [24] 王永忠, 王晓云, 文成林. 梯度点对约束的结构化车道检测[J]. 中国图象图形学报, 2012, 17(06): 657-663.
- [25] van de Sande K E A, Uijlings J R R, Gevers T, Smeulders A W M. Segmentation as selective search for object recognition[C]. 2011 International Conference on Computer Vision, Barcelona, Spain, 2011: 1879-1886.
- [26] 杨建华, 赵轩, 郭全民等. 基于路侧激光雷达的障碍物目标检测方法[J]. 国外电子测量技术, 2023, 42(01): 13-19.
- [27] 李朝. 基于激光雷达和毫米波雷达融合的目标检测方法研究[D]. 中北大学, 2021.
- [28] Nabati R, Qi H. RRPN: Radar Region Proposal Network for Object Detection in Autonomous Vehicles[C]. 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019: 3093-3097.
- [29] 郝艺铭. 多尺度特征融合的轻量深度学习目标检测算法研究[D]. 内蒙古大学, 2021.
- [30] Kunz F, Bader M, Hermann D, Lutz P, Michel T, Pfeiffer D, Schuhmacher D, Schütt J, Stiller C. Autonomous driving at Ulm University: A modular, robust, and sensor-independent fusion approach[J]. 2015 IEEE Intelligent Vehicles Symposium (IV), Seoul, Korea (South), 2015: 666-673.
- [31] Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016: 779-788.
- [32] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot Multibox Detector[C]//Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I. Springer International Publishing, 2016: 21-37.
- [33] sohu. 深度学习入门 | 第四章: 卷积神经网络基础 (1) [EB/OL]. (2020-08-18)[2023.4.24]. [https://www.sohu.com/a/413639327\\_455817](https://www.sohu.com/a/413639327_455817)
- [34] 刘琳. 基于改进的 Faster R-CNN 目标检测算法研究[D]. 吉林大学, 2021.
- [35] Shi S, Wang X, Li H. PointRCNN: 3D Object Proposal Generation and Detection From Point Cloud[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA, 2019: 770-779.
- [36] Shi S, Wang X, Li H, Li L. PV-RCNN: Point-voxel feature set abstraction for 3D object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA, 2020: 10526-10535.
- [37] Girshick R. Fast R-CNN[C]. 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015: 1440-1448.
- [38] He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN[C]. 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017: 2980-2988.
- [39] Salcedo-Sanz S, Rojo-Álvarez JL, Martínez-Ramón M, et al. Support vector machines in engineering: an overview[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2014, 4(3): 234-267.
- [40] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the



- IEEE international conference on computer vision. 2017: 2980-2988.
- [41] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot Multibox Detector[C]//Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I. Springer International Publishing, 2016: 21-37.
- [42] Tran D, Bourdev L, Fergus R, et al. Learning spatiotemporal features with 3d convolutional networks[C]//Proceedings of the IEEE international conference on computer vision. 2015: 4489-4497.
- [43] Shi S, Wang Z, Wang X, et al. Part-a<sup>2</sup> net: 3d part-aware and aggregation neural network for object detection from point cloud[J]. arXiv preprint arXiv:1907.03670, 2019, 2(3).
- [44] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [45] Qin H, Cai Z, Zhang M, et al. Bipointnet: Binary neural network for point clouds[J]. arXiv preprint arXiv:2010.05501, 2020.