自动驾驶调研报告

**摘要**：自动驾驶是未来汽车发展的趋势，具有良好的发展前景。实时准确的检测是保证自动驾驶安全稳定的基础，这包含了许多计算机视觉相关的应用。本文调研了目前主流的目标检测方案，对比不同技术的优缺点，最后总结了未来技术的发展方向。

**关键词**：自动驾驶、目标检测、深度学习、雷达点云

# 1 引言

交通是一个国家经济发展的重要基础产业。随着人们生活水平的提高，汽车已经成为人们日常出行的重要交通工具。但是在汽车方便大众的出行的同时，也带来了交通事故数量的居高不下。传统的汽车驾驶者需要集中注意力面对周遭环境的变换，而且疲劳、天气、心理等因素也会影响驾驶者的判断和反应速度，进而产生危险。

作为未来汽车的发展方向，自动驾驶汽车拥有自主判断能力，能较大程度地减少人为失误。同时，自动驾驶汽车能更好的节能减排、减少污染[1]，有良好的应用前景。为了让自动驾驶安全运行，需要对道路上的各种物体进行识别和判断，例如车辆、行人、交通标志、灯光、车道线等。自动驾驶车辆需要对这些做出实时准确的判断。

# 2 调研内容

## 2.1 自动驾驶简述

自动驾驶系统是多种不同技术融合的产物。自动驾驶即通过多种传感器（如摄像头、激光雷达、毫米波雷达、GPS、惯性传感器等）来识别车辆所处的周边环境和状态，并根据所获得的环境信息（包括道路信息、交通信息、车辆位置和障碍物信息等）自主做出分析和判断，从而自主地控制车辆运动，最终实现自动驾驶[2]。

根据车辆的智能性程度，2021年8月20日，工信部批准发布了GB/T 40429-2021《汽车驾驶自动化分级》标准[3]。该标准将自动驾驶划分为L0-L5等级，具体分级标准如表1所示。

目前自动驾驶获取周围环境信息的方式大致分为2类：摄像头、雷达。相对应的处理方式也分为基于深度学习的方式处理摄像头获取的数据和基于激光雷达点云的方式处理雷达获取的数据。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 等级 | 名称 | 定义 | 车辆横向和纵向运动控制 | 目标和事件探测与响应 | 动态驾驶任务接管 | 设计运行条件 |
| L0 | 应急辅助 | 系统不能持续执行动态驾驶任务中的车辆横向或纵向运动控制，但具备持续执行动态驾驶任务中的部分目标和事件探测与响应的能力 | 驾驶员 | 驾驶员+系统 | 驾驶员 | 有限制 |
| L1 | 部分驾驶辅助 | 系统在其设计运行条件下持续地执行动态驾驶任务中的车辆横向或纵向运动控制，且具备与所执行的车辆横向或纵向运动控制相适应的部分目标和事件探测与响应的能力。 | 驾驶员+系统 | 驾驶员+系统 | 驾驶员 | 有限制 |
| L2 | 组合驾驶辅助 | 系统在其设计运行条件下持续地执行动态驾驶任务中的车辆横向和纵向运动控制，且具备与所执行的车辆横向和纵向运动控制相适应的部分目标和事件探测与响应的能力。 | 系统 | 驾驶员+系统 | 驾驶员 | 有限制 |
| L3 | 有条件自动驾驶 | 系统在其设计运行条件下持续地执行全部动态驾驶任务。 | 系统 | 系统 | 动态驾驶任务接管用户（接管后成为驾驶员） | 有限制 |
| L4 | 高度自动驾驶 | 系统在其设计运行条件下持续地执行全部动态驾驶任务并自动执行最小风险策略。 | 系统 | 系统 | 系统 | 有限制 |
| L5 | 全自动驾驶 | 系统在任何可行驶条件下持续地执行全部动态驾驶任务并自动执行最小风险策略。 | 系统 | 系统 | 系统 | 无限制 |

表1《汽车驾驶自动化分级》标准

### 2.2 基于深度学习

### 2.2.1 YOLO算法在自动驾驶的应用

2016年Redmon等人提出YOLO算法[4]，将检测问题转化为回归问题，使用卷积神经网络来直接完成目标类别的判定和边界的预测。真正意义上实现了目标的实时检测，开启了目标检测单阶段算法的新纪元。

### 2.2.1.1 交通标志检测

Zhang等人[5]通过添加新图像和变换图像扩展了中国交通标志数据集（CTSD），形成新的数据集CCTSDB,并在YOLOv2算法网络的中间层采用多个1×1卷积层，在顶层减少卷积层，降低计算复杂度。

Yang等[6]同样选择了CCTSDB数据集，分别使用YOLOv3和YOLOv4训练训练集。通过测试对比为其在数据集测试中IoU的变化，YOLOv4在目标检测方面优于YOLOv3。

Dewi等[7]则使用生成式对抗网络生成训练图片，将合成图像与原始图像相结合，以增强数据集并验证合成数据集的有效性。图像混合后识别性能得到了提升，在 Yolo v3 上mAP(平均精度值)为 84.9%，在 Yolo v4 mAP为 89.33%。

### 2.2.1.2 交通灯检测

Jensen等[8]使用YOLO，检测公共LISA交通灯数据集，该数据集包含大量在不同光照和天气条件下捕获的带标签的交通灯，达到了90%的准确率。

Possatti等[9]使用YOLOv3模型，与智能汽车使用的先验地图相结合，识别预定义路线的相关交通灯。但是数据集分类只考虑了两类对象：红、黄和绿交通灯。通过混合红色和黄色信号灯的方式，克服黄色样品的不足，实现实时的交通信号灯的检测。

### 2.2.1.3 交通车辆检测

叶佳林等[10]在YOLO v3框架下，设计特征融合结构和采用GIOU损失函数，降低非机动车漏检率，提高定位准确度。

Zhou等[11]将毫米波雷达与摄像机信息融合，利用时空同步关联多传感器数据，通过YOLO v2算法实现深度融合对车辆的目标检测识别。

### 2.2.1.4 行人检测

高宗等人[12]在原有的YOLO网络结构中，结合行人在图像中表现出小纵横比的特点，通过聚类选择合适数量和规格的候选帧，改变YOLO网络结构，调整候选帧在X轴和Y轴方向的分布密度，形成适合行人检测的网络结构。该方法将行人检测视为二分类问题，在行人动态变化的检测方面具有较大局限性。

Zhang等[13]的 Caps-YOLO检测模型则基于YOLOv3，该模型采用dense connection代替了原有网络中的shortcut connection，构造了dense block组件，提高了feature map的利用率。采用向量神经元结合动态路由机制来实现该模型的检测功能，降低漏检和误检姿态复杂的行人的概率，在不同的数据集上精度提升1.81%-6.63%不等。

### 2.2.1.5 车道线的检测

Zhang等人[14]构建了一个基于YOLO v3的两阶段学习网络，对YOLO v3算法的结构参数进行了修改，采用自适应边缘检测算法对车道进行重新定位，并将处理的图像作为标签数据用于第二阶段模型的训练，提高复杂场景下车道检测的准确性。

崔文靓等人[15]在原YOLO v3的基础上，利用K-means聚类算法优化网络anchor参数, 并改进YOLOv3算法卷积层结构，使最终结果较原始算法平均检测准确率提升了11%。

### 2.2.2 CycleGAN模型

CycleGAN模型是一种无监督的模型[16]，用于生成训练集。它可以更好的提升模型跨域检测的精度和模型的迁移能力，可以完成从源数据域到目标数据域的风格转换。

CycleGAN模型包含2个生成器网络和2个判别器网络。2个生成器网络的作用是将输入的源数据域、目标数据域中的图像分别得到其对应的合成图像，通过不断地训练生成器使其合成的图像与对应目标域中的真实图像相似度尽可能高，以骗过判别器使其无法区分该图像是合成图还是真实图。

2个判别器网络的作用是将生成器得到的合成图与原数据域中的真实图像区分开，对于真实图像尽可能输出一个较高的分数(接近于1),对于合成图尽可能输出一个较低的分数(接近于0),因此，判别器不会被生成器生成的‘假图’所蒙骗，能够更好地区分真实图和合成图。

### 2.3 基于激光雷达点云

### 2.3.1 单阶段检测算法

单阶段算法指的是从原始的点云中提取特征，使用这些提取的特征直接进行预测并输出结果。这种算法一般来说运行速度快，但是精度欠佳。

### 2.3.1.1 基于体素的检测算法

基本思想是将原始的点云通过变换后通过体素进行表示，之后通过骨干网络对体素进行特征提取。

例如，在VoxelNet[17]网络中，原始的点云数据首先根据点云的(X,Y,Z)坐标被划分为若干体素，对划分后的点云进行特征编码，然后生成对应的特征图，通过在骨干网络中使用3D卷积对特征图进行特征提取，最后3D检测头将根据获得到的特征进行目标预测并生成相应的边界框。

### 2.3.1.2 基于点的检测算法

基本思想是直接使用原始的点云作为网络的输入并提取出点特征，用于目标检测。

3DSSD[18]提出了一种融合采样方法，通过Distance-FPS和Feature-FPS代替了传统的特征传播层从而提高模型的运行效率。之后获得的特征将作为候选生成层的输入，最后使用无锚框检测头生成预测结果。

### 2.3.1.3 基于图的检测算法

Point-GNN[19]网络使用图来表示点云，首先使用体素对点云进行下采样，利用下采样后的点云特征作为节点，通过多层感知机来提取边特征并迭代更新节点的特征，最后利用迭代后获取的点云特征作为检测头的输入来生成检测结果。

### 2.3.2 两阶段检测算法

两阶段检测方法是使用第二阶段的网络从第一阶段生成的区域提案中进行更精确的检测[20]，因此两阶段检测算法也被称为基于区域提案的方法。两阶段检测方法在运行速度上较慢，但是却可以获得相较单阶段检测算法更好的检测性能。

PointVoxel-RCNN[21](PV-RCNN）算法首先将点云划分为若干球形体素，之后使用3D稀疏卷积网络以及基于PointNet[22]的网络来提取点云特征。过3D稀疏卷积网络将体素化的点云通生成多尺度的语义信息，并生成3D目标提议。通过体素集抽象模块将把学习到的体素特征编码为一组关键点。关键点特征将作为输入，来获得更加丰富的信息。这些信息被用于优化之前生成的3D目标提议，提升检测的精确度。

# 3 结语

目前基于神经网络的目标检测技术，主要是从YOLO算法为基础，使用卷积神经网络来直接进行检测和判断。常见的优化方式有修改模型和修改训练集。对于自动驾驶的过程中需要检测的不同物体，有不同的模型进行处理，目标检测的准确率有90%以上，但是处理速度不高，尤其是要对每个模型分别处理一次，没有通用的模型来识别。对于自动驾驶的设备性能有着一定的要求。

基于激光雷达的点云的目标识别技术，分为单阶段和两阶段算法，单阶段的速度快但是精度低，两阶段则是在单阶段的基础上再次进行精确检测。点云包含了物体的深度信息，对比摄像头获取的照片多出了一个维度，获取信息的难度降低。但是点云只包含了离散的数据，相比于照片获取的点的数量更少。

未来自动驾驶的潜在发展方向要结合摄像头和雷达，对大体的情况使用摄像头的方案，对于局部需要精确判断的地方用雷达的方案，将两者结合可以增加自动驾驶的安全性和稳定性。同时也要注意两种方式的判断结果如果相反，那么有一个决策的问题。而且自动驾驶的设备性能对于安全性也有极大的提升。

参考文献：

[1]. LUETTEL T, HIMMELSBACH M, WUENSCHE H-J. Autonomous ground vehicles—Concepts and a path to e[J]. Proceedings of the IEEE, 2012, 100(Special Centennial Issue):1831-1839.

[2]. 申泽邦，雍宾宾，周庆国，李良. ⽆⼈驾驶原理与实践[M]. 北京：机械⼯业出版社, 2019.

[3]. 全国汽车标准化技术委员会. GB/T 40429-2021《汽车驾驶自动化分级》正式发布 [EB/OL].(2021-09-11) [2021-12-05]. http://www.catarc.org. cn/xinwen/show-3334.html.National Automotive Standardization Technical Committee. GB/T 40429-2021 "Automotive Driving Automation Classification" officially released[EB/OL]. (2021-09-11) [2021-12-05].

[4]. REDMON J, DIVVALA S, GIRSHICK R, et al.You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE con- ference on computer vision and pattern recognition. 2016:779-788.

[5]. ZHANG J, HUANG M, JIN X, et al. A real- time chinese traffic sign detection algorithm based on modified YOLOv2[J]. Algorithms, 2017, 10(4):127.

[6]. YANG W, ZHANG W. Real-time Traffic Signs Detection Based on YOLO Network Model[C] //2020 International Conference on Cyber- Enabled Distributed Computing and Knowledge Discovery (CyberC). Chongqing, China.29-30 Oct. 2020 IEEE, 2020:354-357.

[7]. DEWI C, CHEN R-C, LIU Y-T, et al. Yolo V4 for advanced traffic sign recognition with synthetic training data generated by various GAN[J]. IEEE Access, 2021, 9:97228-97242

[8]. DICKMANNS E D, ZAPP A. Autonomous high speed road vehicle guidance by computer vision[J]. IFAC Proceedings Volumes, 1987, 20(5): 221-226.

[9]. 穆康乐. 无人驾驶汽车发展现状及未来展望[J]. 电子技术与软件工程, 2017, (21):112-114. MU K L. Development Status and Future Prospects of Autonomous Vehicles [J]. Electronic Technology and Software Engineering, 2017, (21): 112-114.

[10]. 叶佳林, 苏子毅, 马浩炎,等. 改进 YOLOv3 的 非机动车检测与识别方法[J]. 计算机工程与应用, 2021, 57(1):194-199.YE J L, SU Z Y, MA H Y, et al. Improved YOLOv3 Non-motor Vehicles Detection and Recognition Method[J]. Computer Engineering and Applications, 2021, 57(1):194-199.

[11]. ZHOU T, JIANG K, XIAO Z, et al. Object Detection Using Multi-Sensor Fusion Based on Deep Learning [M]. CICTP 2019. 2019: 5770- 5782.

[12]. 高宗, 李少波, 陈济楠,等. 基于 YOLO 网络的 行人检测方法[J]. 计算机工程, 2018, 44(5):215-219+226.GAO Z, LI S B, CHEN J N, et al. Pedestrian detection method based on YOLO network [J]. Computer Engineering, 2018, 44(5): 215-219+ 226.

[13]. ZHANG C, LUO K, GU S, et al. Caps-YOLO: Pedestrian Detection Method of Complex Posture Combined with Capsules Network[J]. Journal of Flow Visualization and Image Processing, 2021, 28(3):41-69.

[14]. ZHANG X, YANG W, TANG X, et al. A fast learning method for accurate and robust lane detection using two-stage feature extraction with YOLO v3[J]. Sensors, 2018, 18(12):4308.

[15]. 崔文靓, 王玉静, 康守强, 等. 基于改进 YOLOv3 算法的公路车道线检测方法[J/OL]自动化学报:1-9 [2022-03-24]. DOI:10.16383/j.aas.c190178.

[16]. 范佳琦,李鑫,霍天娇,洪金龙,高炳钊,陈虹.基于单阶段算法的智能汽车跨域检测研究[J].中国公路学报,2022,35(03):249-262.DOI:10.19721/j.cnki.1001-7372.2022.03.021.

[17]. ZHOU Y，TUZEL O.Voxelnet：End-to-end learning for pointcloud based 3d object detection[C]//Proceedings of the IEEEconference on computer vision and pattern recognition，2018.

[18]. YANG Z，SUN Y，LIU S，et al.3dssd:Point-based 3d single stage object detector[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition，2020.

[19]. SHI W，RAJKUMAR R.Point-gnn:Graph neural network for 3d object detection in a point cloud[C]//Proceedings of the IEEE /CVF conference on computer vision and pattern recognition，2020.

[20]. 彭玉旭，董胜超 . 基于注意力机制的三维点云车辆目标检 测 [J]. 计算机系统应用，2021，30(12):211-217.

[21]. SHI S，GUO C，JIANG L，et al.Pv-rcnn:Point-voxel feature set abstraction for 3d object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2020.

[22]. 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.