分类号：TP242.6　　学校代码：10363
密　级：公开　　学　　号：2220320145

安徽工程大学
Anhui Polytechnic University

# 硕士学位论文

**题目**　基于多尺度特征融合网络的车辆目标检测与跟踪研究

论文作者　江自豪
校内导师　王冠凌（教授）
校外导师　毛德平

专业学位类别　电子信息（085400）
研究方向（领域）　人工智能

论文提交日期：2025 年 5 月 16 日

分类号：TP242.6　　单位代码：10363

密　级：公 开　　学　号:2220320145

# 题　目　基于多尺度特征融合网络的车辆目标检测与跟踪研究

# 题　目　Research on Vehicle Object Detection and Tracking Based on Multi-scale Feature Fusion Network

学生姓名：　江自豪

校内导师：　王冠凌

校外导师：　毛德平

专　　业：　电子信息

研究方向：　人工智能

论文答辩日期：　2025.5.12

# 安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律后果由本人承担。

学位论文作者签名： 

日期： 2025 年 5 月 16 日

# 安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 □，在 __ 年解密后适用本版权书。

本学位论文属于

不保密 ☑。

学位论文作者签名：江自豪　指导教师签名：

日期：2025 年 5 月 16 日　　日期：2025 年 5 月 16 日

# 基于多尺度特征融合网络的车辆目标检测与跟踪研究

## 摘 要

随着智能交通系统向智能化与数字化的深度演进，车辆目标检测与跟踪技术已成为智慧交通管理的关键核心技术，对于构建道路安全预警体系、提升交通管理效率以及保障道路交通安全具有至关重要的作用。传统的车辆检测和跟踪系统在面对目标遮挡、非线性运动、陡然显现等场景时跟踪误差较大，无法满足精确检测和跟踪的需要。针对上述问题，本文深入研究了一种基于多尺度特征融合网络的车辆目标检测与跟踪方法。该方法通过融合多尺度特征，有效提升了车辆目标检测与跟踪的精度与效率，能够更好地适应复杂多变的交通环境，为智能交通系统的发展提供了新的技术思路。本文的主要研究内容如下：

（1）针对车辆目标检测中精度低、速度慢、误识别等问题，本文基于 YOLOv7 模型改进了一种多尺度特征融合的 YOLOv7-PGC 网络。通过融合特征金字塔网络(Feature Pyramid Network, FPN)与路径聚合网络(Path Aggregation Network, PANet)，优化了一种多尺度特征融合结构，加强不同层级间的特征信息融合。引入 GE 注意力机制，增强了上下文特征关联，提高了目标检测精度。针对车辆目标检测实时性的需求，引入 CoordConv 增强对特征图的空间感知能力，减少了模型的参数量和运算量。改进后的 YOLOv7 平均精度提升 5.6 个百分点至 48.2%，召回率提高了 9.2%，验证了检测算法改进的有效

性和稳定性。

（2）针对动态交通场景中传统跟踪算法因特征鲁棒性不足与高速目标运动导致的线性模型预测误差问题，本文基于 BoT-SORT 多目标跟踪算法进行改进。首先对于线性运动模型适应性差的问题，引入了自适应卡尔曼滤波(Adaptive Kalman Filter, AKF)，增强模型对运动状态的适应性，减少复杂道路状况带来的影响。针对目标遮挡与密集环境下，预测框与目标匹配失败而导致的 ID 变化问题，采用 CIoU 匹配，对目标进行多维度的优化匹配，此外，引用聚合 SURF 描述符的 VLAD 特征提取方法，增强关键帧的特征聚合。改进后的 BoT-SORT 跟踪算法 MOTA 值提高了 5.1%，IDF1 提高了 9.8%，IDs 减少了 77 次。结果表明，该方法在快速移动、遮挡和形变等复杂交通场景下具备更好的适应性。

（3）针对传统多目标跟踪方法在复杂车辆行驶场景中的局限性，本文借助搭建的无人车集成系统，利用 ROS 操作系统进行模型部署和获取相机的视频图像，设计了搭载 RK3588S 和 GD32F4 处理核心的无人车硬件集成系统，通过远程终端控制无人车在实验场地进行实际行驶。在白天复杂的多目标车辆交通场景中，对车辆目标的跟踪 IDs 比原算法减少了 29，漏检率降低了 13.4%，结果表明，该系统算法可以达到预期效果，具备可行性和有效性。

**关键字**：无人车，YOLOv7，多目标跟踪，ROS，多尺度特征网络

# RESEARCH ON VEHICLE OBJECT DETECTION AND TRACKING BASED ON MULTI-SCALE FEATURE FUSION NETWORK

## ABSTRACT

With the deep evolution of intelligent transportation system to intelligence and digitalization, vehicle target detection and tracking technology has become the key core technology of intelligent traffic management, which plays a crucial role in building road safety early warning system, improving traffic management efficiency and ensuring road traffic safety. The traditional vehicle detection and tracking system can not meet the need of accurate detection and tracking because of its large tracking error in the face of target occlusion, nonlinear motion, sudden appearance and other scenes. Aiming at the above problems, a vehicle target detection and tracking method based on multi-scale feature fusion network is deeply studied in this paper. By integrating multi-scale features, this method effectively improves the accuracy and efficiency of vehicle target detection and tracking, and can better adapt to the complex and changeable traffic environment, providing a new technical idea for the development of intelligent transportation system. The main research contents of this paper are as follows:

(1) Aiming at the problems of low precision, slow speed and misrecognition in vehicle target detection, this paper improves a multi-scale feature fusion YOLOV7-PGC network based on YOLOv7 model. By fusing the Feature Pyramid Network (FPN) and the Path Aggregation Network (PANet),a multi-scale feature fusion structure is optimized to enhance the feature information fusion at different levels.The introduction of GE attention mechanism enhances the context feature correlation and improves the target detection accuracy. In order to meet the real-time requirement of vehicle object detection,CoordConv is introduced to enhance the spatial perception ability of feature maps and reduce the number of parameters and calculation of the model. After the improvement, the average accuracy of YOLOv7 is increased by 5.6 percentage points to 48.2%, and the recall rate is increased by 9.2%, which verifies the effectiveness and stability of the improved detection algorithm.

(2) Aiming at the linear model prediction error caused by the lack of feature robustness and high-speed target movement of traditional tracking algorithms in dynamic traffic scenes. This paper improves the multi-target tracking algorithm based on BoT-SORT. Firstly, to solve the problem of poor adaptability of linear motion model, Adaptive Kalman Filter (AKF) was introduced to enhance the adaptability of the model to motion state and reduce the influence of complex road conditions. In order to solve the problem of ID changes caused by the failure of prediction frame to match

the target in the environment of target occlusion and dense environment, CIoU matching is used to optimize the target in multiple dimensions. In addition, the VLAD feature extraction method of aggregate SURF descriptor is used to enhance the feature aggregation of key frames. The MOTA value of the improved BoT-SORT tracking algorithm increased by 5.1%, IDF1 increased by 9.8%, and IDs decreased by 77 times. The results show that the proposed method has better adaptability in complex traffic scenarios such as fast movement, occlusion and deformation.

(3) In view of the limitations of traditional multi-target tracking methods in complex vehicle driving scenes, this paper designed an unmanned vehicle hardware integration system equipped with RK3588S and GD32F4 processing cores by using the unmanned vehicle integration system built and ROS operating system for model deployment and camera video image acquisition. The unmanned vehicle is controlled by a remote terminal for actual driving in the experiment site. In the complex multi-target vehicle traffic scene during the day, the tracking IDs of vehicle targets is reduced by 29 % and the missing detection rate is reduced by 13.4%. The results show that the algorithm can achieve the expected effect and has feasibility and effectiveness.

**KEY WORDS:** Unmanned vehicles, YOLOv7, Multi-target tracking, ROS, Multi-scale feature network

# 目录

# 第 1 章 绪论

## 1.1 研究背景及意义

随着经济的持续增长和城市化进程的加快，汽车已成为人们日常生活中不可或缺的交通工具，但车辆数量的激增也带来了交通拥堵、环境污染以及交通事故频发等严重问题。为解决这些难题，城市智能交通系统(Intelligent Transportation System, ITS)通过集成人工智能与物联网技术，显著提升了交通效率与安全性[1]，旨在通过集成先进的智能化技术提升交通效率、保障交通安全并减少环境污染。车辆目标检测与跟踪技术作为智能交通系统的核心技术之一，其研究背景与意义深远，紧密关系着现代社会的交通发展、技术进步以及安全需求，在社会交通发展领域中扮演着至关重要的角色。

在智能交通系统中，车辆目标检测与跟踪技术能够实现车辆身份的实时识别、位置信息的精确追踪以及运动状态的动态分析，为交通监控、交通流量分析、交通信号控制等提供了有力支持[2]。通过实时监测交通状况，该技术能够及时发现并预警潜在的交通危险，如车辆碰撞、超速行驶等，有效减少交通事故的发生，提高道路通行能力。同时，该技术还能准确统计车辆数量，分析交通流量分布，为交通管理部门提供科学依据，帮助他们更好地规划和管理交通流量，优化交通网络结构。在交通信号控制方面，该技术能够根据实时交通信息智能调整交通信号灯的配时，提高交通效率，缓解交通拥堵[3]。

车辆目标检测与跟踪技术广泛应用在自动驾驶领域，该技术是实现车辆自主导航和自动避障的关键技术之一[4]，通过实时监测和分析车辆周围环境信息，自动驾驶系统能够做出准确决策，确保车辆安全、高效地行驶。在视频监控领域，该技术能够用于监控视频中的车辆识别和跟踪，为安全监控提供有力支持，特别是在公共场所、交通枢纽等关键区域，通过安装视频监控设备并应用车辆目标检测与跟踪技术[5]，可以实时监测车辆进出情况，有效防范和打击犯罪行为，提高社会安全性。在智能安防领域，该技术同样具有广泛的应用前景，可以实时监测车辆出入情况，及时发现异常行为，提高安全防范能力。

在自动驾驶领域，车辆与行人检测作为环境感知的关键环节，直接关系到

系统的安全性与可靠性。然而，复杂动态场景（如交通拥堵、极端光照、遮挡干扰）对检测算法的鲁棒性构成严峻挑战，这些问题使得车辆检测成为一项具有挑战性的任务。例如，CIDAS 对国内 4000 例事故的结果分析表明，约 31%的碰撞事故与感知延迟或目标跟踪丢失有关（如行人突然横穿时的漏检）；KITTI 数据集测试显示，遮挡场景下车辆检测精度下降约 23%[6]。传统的视频监控系统主要依赖人工观察，存在主观误差和误判的风险，尤其在监控点较多、背景复杂的情况下，容易造成目标的遗漏。现有的目标检测方法往往缺乏对于车辆的针对性，检测精度较低，尤其对于复杂环境下的小目标车辆，检测效果难以令人满意。针对此类问题，通过优化目标检测算法的多尺度处理能力与实时推理效率，可有效提升车辆在密集目标场景下的决策准确性，研究高效、鲁棒性强的车辆检测方法具有重要意义。

近年来，随着深度学习算法的不断优化和计算机硬件性能的不断提升，基于多尺度特征融合网络的车辆目标检测算法得以快速发展。该类算法能够跨层融合多个尺度的特征，并为不同尺度特征分配可学习的权重，从而增强特征的表示能力，该方法通过整合不同层次的特征表示，能够较好地保留易损失的细节特征，对处理具有复杂结构特征的输入数据有着显著的适应性。结合深度学习的目标跟踪算法，可以实现车辆的实时跟踪和测距，为车辆调度和管理提供重要数据支持和智能决策。

综上所述，车辆目标检测与跟踪技术的研究具有重要的理论意义和实际应用价值。它不仅为智能交通和自动驾驶提供了关键技术支持，还通过实时、准确的车辆检测和跟踪，提高了道路安全性，优化了交通流量管理。随着技术的不断进步和应用场景的不断拓展，车辆目标检测与跟踪技术将在未来的社会发展中发挥越来越重要的作用，推动智能交通和自动驾驶的持续发展。

## 1.2 国内外研究现状

车辆目标检测与跟踪技术的国内外正经历多维度的技术迭代与场景适配。在国内，研究者们不断优化算法，提高检测精度和速度，以适应复杂多变的交通环境。特别是在深度学习技术的推动下，基于深度学习的相关车辆目标检测系统不断涌现，并在城市级智慧交通管理平台中取得了显著成效。同时，国内

研究团队和学者们还积极探索车辆目标检测与跟踪技术在智能交通、自动驾驶等领域的应用，例如车路云协同感知框架的构建，有效降低多源异构数据的误差。而在国外，则呈现出跨模态融合与弱监督学习的前沿探索趋势。研究者们不仅关注算法的性能提升，产业界亦加速技术转化，如 Mobileye 的路网采集管理系统 (Road Experience Management, REM)系统已实现全球 2 亿公里道路要素的实时更新，其轨迹预测误差下降明显[7]。此外，国外研究机构还在不断探索新的技术路径，如基于多传感器融合的车辆目标检测、基于弱监督学习的车辆目标检测等，以进一步拓展车辆目标检测与跟踪技术的应用范围。

### 1.2.1 目标检测研究现状

计算机视觉技术已成为互联网应用领域中非常重要的方向，在众多视觉应用中发挥着基础性作用。近年来，基于卷积神经网络的特征提取方法显著提升了检测性能，推动了该领域的快速发展。车辆与行人检测是自动驾驶的重要组成部分之一。自动驾驶的安全性在很大程度上取决于对周围障碍物的目标检测，但是在复杂多变的实际环境下，目标检测需要面临大量艰巨的挑战，国内交通事故多发，特别是在拥堵的交通环境下，通过改进优化目标检测算法，将有助于解决这些困难。以深度学习为核心的目标检测技术呈现出蓬勃发展的态势，相较于传统手工设计特征的方法，深度神经网络能够自动学习更具判别性的特征表示，这一优势推动了该领域的革新。

早期的目标检测通过手动生成感兴趣位置，用滑动窗口扫描整个图像，通过将输入图像的大小调整为不同的比例，并使用多比例窗口在这些图像之间滑动，尽可能的捕获有关对象的多比例和不同纵横比的信息，然后从滑动窗口获得一个固定长度的特征向量，以获取所覆盖区域的判别性语义信息，例如 SIFT[8]、Haar[9]、HOG[10]和 SURF[11]，最后为覆盖区域分配分类标签。

近年来，随着深度学习技术的发展，基于深度学习的目标检测方法逐渐兴起。深度学习模型能够自动学习特征，并且在目标检测和识别任务中表现出色[12]。基于深度学习模型的目标检测算法主要为两类：一种是 Two-Stage 目标检测，典型代表有 R-CNN，Fast R-CNN[13]，Faster R-CNN[14]，这类方法使用启发式方法或者 CNN 网络产生候选区域做引导，然后再在区域上做分类与回归，另

一种是 One-Stage 目标检测算法，主要有 YOLO(You Only Look Once)系列[15][16][17]、SSD[18]、RetianNet[19]等做代表的算法，2016 年 Redmon 等提出 YOLO 算法，将目标的定位与分类问题转化为回归问题，把图像送入到网络中提取目标特征，直接输出目标的坐标位置与类别，这类算法旨在减少计算复杂度，增加鲁棒性，直接在整个图像上一次性预测所有目标的位置和类别，无需生成候选区域。在保持较高检测准确率的同时，显著提高检测速度。Cao 等[20]人将 Swin Transformer 结构[21]应用于 CNN 模型，同时插入坐标注意力机制 CA[22]，使网络能够专注于上下文空间信息，并将 PANet[23]路径聚合网络与 FPN 融合。Qin 等人[24]提出了一种预测尺度增大、网络深度减小的交通显著目标检测框架 YOLO(ID-YOLO)来预测驾驶员注视区域内的物体。盛等人[25]依托递归特征金字塔 REP[26]结构，设计基于反馈连接的特征金字塔 RFP-PAN，获得更丰富的定位信息，但是计算量很大。Wang 等人[27]提出了一种针对混合交通流环境的碰撞检测框架，引入 Retinex 图像增强算法来提高在低能见度条件下收集的图像质量，并提出一种基于决策树的框架，用于确定混合交通流环境中的各种崩溃场景，在各种能见度条件下混合交通流环境中有着不错的效果。

### 1.2.2 目标跟踪研究现状

随着新能源汽车行业的兴起，车辆智能驾驶系统也随之发展迅速，车辆目标跟踪技术已成为连接安全出行与高效交通管理的桥梁，是智能驾驶领域的重要组成部分。这一技术利用视频图像、雷达等传感器数据，实时捕捉并分析车辆的运动轨迹，为自动驾驶、交通监控等领域提供了有力支持。目标跟踪技术的核心研究内容在于对连续图像序列中的特定目标进行精确定位与持续追踪，与计算机视觉中的其他视觉任务相同，目标跟踪同样需要用相机摄像头来取代人眼进行目标搜索和观察，获取目标及其背景的数字图像信息，然后通过扮演人脑角色的人工智能来解析和处理数码信息[28]。

目标跟踪技术在这些年的发展历程中也经历着从传统方法到深度学习的演变，根据目标数量分类，目标跟踪可以大致的分为单目标跟踪(Single Object Tracking, SOT)与多目标跟踪(Multiple Object Tracking, MOT)[29]。单目标跟踪任务是跟踪视频序列中标记的对象。大多数现有的单目标跟踪器分 4 个步骤进行

跟踪。（1）输入图像和提取特征。特征对跟踪过程的性能影响很大，常用的特征是手动特征和卷积特征。（2）生成候选区域。常用方法是粒子过滤器预测[30]和滑动窗口，两者分别使用推理法和穷举法来预测候选区域。（3）构建跟踪模型。构建目标跟踪模型和准确选择候选区域是目标跟踪任务的核心，单跟踪模型包括生成式和判别式。（4）更新模型。环境和目标本身的变化需要实时进行更新。Bolme 等人[31]提出的 MOSSE 跟踪是将信号处理领域的相关滤波技术与跟踪领域问题结合起来，实现了超高速的跟踪性能。Henriques 等人[32]研究的 CSK 跟踪提出结合核方法解决跟踪问题，之后在 2014 年又提出 KCF 跟踪[33]，利用循环矩阵生成密集样本在保证高实时性的同时提高跟踪速度。Danelljan 等人[34]提出的 C-COT 跟踪在精度提升方面表现显著，但是模型实时性较差，导致速度降低。Lu 等人[35]通过通道正则化处理卷积特征的通道冗余，这提高了滤波器跟踪的速度和准确性。但是，随着深度学习的应用，传统方法具有很大的局限性。

基于相关滤波的目标跟踪方法在性能上展现出显著优势，其跟踪鲁棒性和计算效率均优于传统算法。然而，该方法在处理目标快速运动、严重遮挡等复杂场景时仍存在明显局限性。深度学习方法通过构建端到端的跟踪框架，能够更好地处理目标表观变化，为克服传统方法的局限性提供了新的解决思路。Bertinetto 等人[36]提出 SiamFC 网络通过相同权重的两个同框架网络衡量两个输入的相似度，使用深度卷积神经网络 AlexNet 的 SiamFC 首先开创了跟踪领域基于端到端的滤波器与深度学习结合的方法。Li 等人[37]基于区域跟踪的 SiamRPN 网络进一步提高了算法精度，核心思想是先大致地预测目标的尺度和位置作为先验预测。SiamRPN++[38]视觉跟踪算法在 SiamRPN 网络基础上引入 ResNet 网络和 RPN 网络等结构。2022 年 Yan 等人[39]提出的 Unicorn 算法通过替换局部搜索区域，提高了场景的适用性。2023 年 Chen 等人[40]提出 SeqTrack 跟踪，将坐标序列生成与动态模板更新结合，简化框架并提升低帧率场景性能。

多目标的跟踪策略主要有两种，一是基于检测的跟踪(Detection-Based Tracking, DBT)，另一种是基于初始框的跟踪(Detection Free Tracking, DFT)[41]，图 1-1 清晰地展现了两类算法的区别。

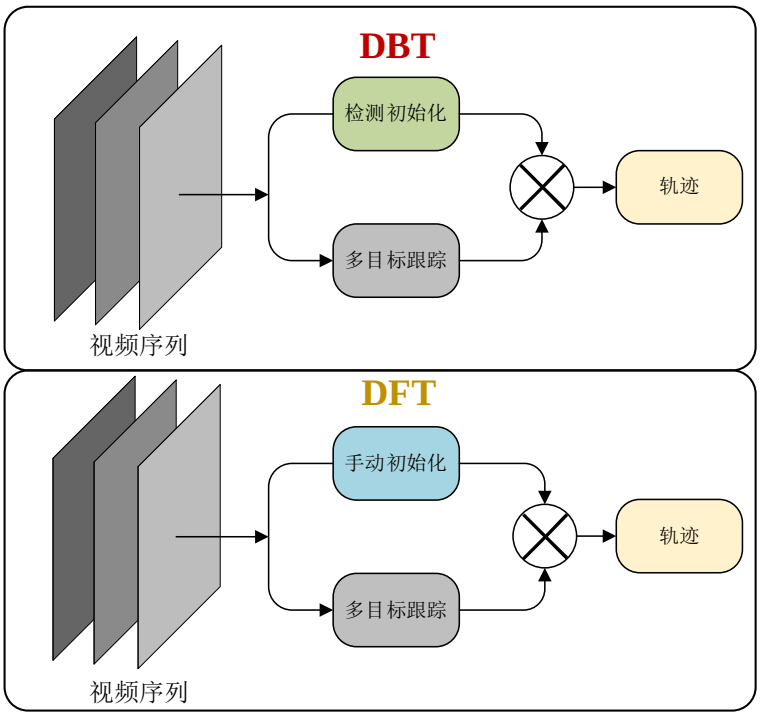



图 1-1 多目标跟踪分类示意图

DFT 与单目标跟踪存在显著共性特征，即两者均需在初始阶段通过人工标注对视频序列起始帧中的目标进行精确标注，随后根据初始化结果同步执行目标检测与轨迹追踪任务并更新。然而，这种人工初始化方式当视频场景中存在未在首帧出现的目标时，将无法有效跟踪。由于多目标跟踪任务本身涉及动态场景下目标的消失、重现及新生等复杂状况，人工标注的语义局限性将直接导致跟踪结果产生累积性误差，因而 DBT 比 DFT 更常用，2016 年提出的 SORT[42] 算法，通过匈牙利算法实现最优匹配，并使用卡尔曼滤波器对目标的位置进行预测和更新，使得 DBT 是目前学术界和工业界主流的多目标跟踪策略。

DBT 是指基于检测进行跟踪，基于 DBT 策略的 MOT 包括一个独立的检测过程、一个检测结果和跟踪器轨迹连接的过程。由于 DBT 的算法的跟踪效果与检测结果关联性过强，于是提出了降低检测不稳定性影响的基于深度学习候选人选择与再识别的实时多跟踪(Multiple Object Tracking with Deeply Learned Candidate Selection, MOTDT)[43]。

2017 年 Wojke 等人[44]提出 DeepSORT 算法结合了深度学习特征和卡尔曼滤波(Kalman Filter, KF)的运动模型。通过深度神经网络提取的外观特征，准确区分不同目标，即使在复杂场景下也能保持高鲁棒性。同时，整合的 KF 有效预测

目标位置，处理观测噪声，提高了追踪的稳定性和准确性。此外，DeepSORT 采用匈牙利算法进行数据关联，确保目标身份的一致性和追踪的连续性。2022 年 Aharon 等人提出的 BoT-SORT 算法优化了 KF 的状态向量，引入了相机运动模型 CMC 以补偿相机运动，并提出了 IoU-ReID 融合策略，通过融合位置、外观和运动特征来提高目标匹配的精度和鲁棒性。这些改进使得 BoT-SORT 在目标遮挡、快速移动等复杂场景下表现更佳，同时保持了算法的实时性。阎等人[45]基于 Bot-SORT 算法改进轻量化框架，同时采用 MobileNetV3 重识别网络，增强特征图中目标区域关注度。Yang 等人[46]提出 MR-SORT，通过减少特征信息离散并增加网络的感知场，更准确地匹配相邻视频帧中的目标，降低跟踪过程中 ID 切换的概率。2024 年，Tumanyan 等人[47]将 DINOv2 视觉 Transformer 的特征提取能力应用于密集跟踪，提出了一种无需监督训练的优化框架，通过特征匹配和运动观察调整，显著提升了长期遮挡场景下的跟踪精度。Liu 等人[48]提出了一种多匹配身份验证网络 (MIA-Net) 用于多无人机多目标跟踪任务，并设计了一个多设备目标关联分数 (MDA) 作为多设备跟踪中跨视图目标关联能力的评估标准。Cao 等人[49]提出了一种 OC-SORT，通过结合目标检测器的结果，计算虚拟轨迹，修复滤波器的误差累积。深度学习的应用，改变了传统滤波技术在跟踪中的统治地位，为广大研究学者提供了多种多样的技术思路，同时给目标跟踪带来了技术性的变革。

## 1.3 主要研究内容

本文对于复杂环境下自动驾驶目标检测效率低，环境感知困难等问题，根据深度学习特征网络提出了一种新的目标检测特征融合网络，对于目标采样更加多样，然后对车辆目标跟踪进行研究，通过改进多跟踪算法，最终提出了一种由 YOLOv7-PGC 和改进 BoT-SORT 结合的多尺度特征融合网络，最后通过搭建无人车实验平台，设计了无人车目标检测和跟踪系统，并对融合算法进行可行性测试。本文的主要研究内容如下：

（1）在原有特 FPN 与 PANet 的基础上，结合原有 YOLOv7 网络结构提出了一种多尺度特征融合结构。在网络特征层中增加了适用于小物体检测任务的检测层。并添加多条水平链接以融合多尺度特征信息。在特征提取阶段添加 GE

注意力机制，加强复杂环境下的对上下文特征提取；在颈部和头部引入协调坐标卷积(Coordinate Convolution, CoordConv)，增强对特征图的空间感知能力，在几乎不增加参数量的同时，提高检测效果。

（2）提出一种基于改进 BoT-SORT 的车辆跟踪算法，通过匹配检测框与局部图像，并与卡尔曼滤波预测框相结合得到车辆框位置，采用 YOLOv7-PGC 算法作为检测器，通过采用迭代更新过程，引入自适应卡尔曼滤波器；优化 IoU 损失函数，采用完全交并比(Complete Intersection over Union, CIoU)作为跟踪两次关联匹配的损失函数；在第一次和第二次匹配之间添加引入 SURF 描述符的局部聚合描述符向量（Vector of Locally Aggregated Descriptors，VLAD）提取目标的外观特征并加强特征匹配，最后通过实验验证模型改进的有效性。

（3）将 YOLOv7-PGC 与改进 BoT-SORT 模框架相结合，搭建无人车软硬件集成平台，在系统集成平台上进行仿真测试，并设计了基于智能无人车的车辆目标检测和跟踪系统，同时结合优化后的车辆检测跟踪算法模型框架，将经过多尺度特征融合的检测跟踪模型框架应用于无人车嵌入式集成平台上，测试在实际环境中的检测跟踪效果。

## 1.4 论文结构安排

本文具体章节结构安排如下：

第一章：绪论。本章首先从目标检测与跟踪的背景出发，结合当前国内外相关技术研究现状，给出了切合实际的车辆目标检测和跟踪技术的研究意义，之后分别对基于深度学习的目标检测、目标跟踪相关背景技术进行介绍，阐述了现在车辆目标检测与跟踪所面临的难题。

第二章：相关理论基础。本章从四个部分详细介绍了本文所运用的技术知识及理论基础，首先介绍了深度学习中卷积神经网络的组成部分以及原理等，对多尺度特征网络这一经典网络结构进行简单的阐述，最后对于目标检测和目标跟踪领域中的一些方法进行原理介绍。

第三章：基于 YOLOv7-PGC 的车辆目标检测算法。本章针对复杂环境下 YOLOv7 目标检测算法精度差、误差高等问题，对原算法进行改进优化，最终提出了 YOLOv7-PGC 多尺度特征的目标检测算法，详细的介绍了 YOLOv7-PGC

算法的结构和创新部分，最后通过一系列的消融对比实验，验证改进后检测算法的有效性。

第四章：基于改进 BoT-SORT 的多目标跟踪算法。本章基于 BoT-SORT 多目标跟踪算法进行相关的创新，基于 BoT-SORT 的原理进行分析，采用自适应卡尔曼滤波和 CIoU 匹配对运动状态模型的局限性进行优化，同时还引入 VLAD 算法模块，最后在自制车辆数据集上对改进模型的性能进行分析，并根据分析结论进行总结。

第五章：车辆检测与跟踪系统实验。在本章中，根据智能无人车平台，首先对无人车的硬件设计和软件框架进行阐述，基于机器人操作系统(Robot Operating System, ROS)、OpenCV 库等工具对检测和跟踪融合的模型进行部署，通过远程控制无人车自动采集的图像与视频信息并进行复杂环境跟踪实验。

第六章：总结与展望。在本章内容中，对本文的总体内容进行概述，总结了论文的主要工作和研究内容，指出了尚且存在的一些不足和局限，对接下来的工作计划和方向进行展望。

# 第 2 章 相关理论基础

## 2.1 卷积神经网络

卷积神经网络(Convolutional Neural Networks, CNN)作为深度学习领域的核心技术，通过融合人工神经网络与深度学习，在计算机视觉任务中表现卓越。CNN 的核心特征在于其独特的特征提取机制，主要由卷积运算和下采样操作构成。与全连接神经网络不同，CNN 采用局部连接策略，即每个神经元仅与输入层的局部感受野相连，这种设计显著降低了网络复杂度。

典型的 CNN 架构包含输入层、隐含层和输出层三个主要部分，其中隐含层进一步细分为卷积层、池化层和全连接层，其具体结构如图 2-1 所示。这种层次化设计使得 CNN 能够逐级提取图像的低级到高级特征，为图像处理任务提供了强大的特征表示能力。

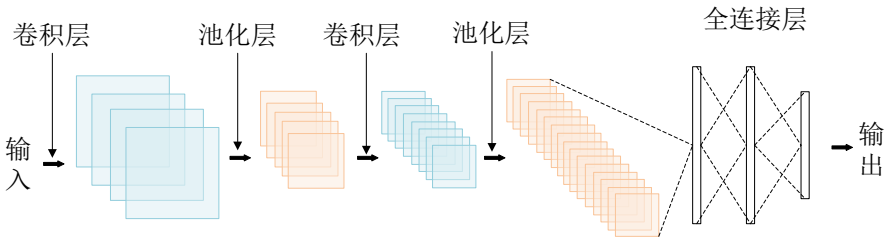



图 2-1 卷积神经网络结构图

### 2.1.1 卷积层

卷积层作为 CNN 的核心框架，在特征提取、参数优化和计算效率等方面展现出独特优势。该层由多个可学习的卷积核构成，其参数通过反向传播算法进行迭代优化。卷积运算的本质是通过局部连接和权值共享机制，从输入数据中提取多层次的特征表示。现代卷积神经网络通常采用多层卷积层级联的结构，随后连接池化层，这种设计能够逐步提取从低级到高级的特征表示。

卷积操作通过可学习的滤波器对输入图像进行处理，在增强有效特征的同时抑制噪声干扰。这一过程体现了两个重要特性：局部感受野机制限制了神经元只响应特定区域的信息，而权值共享则显著降低了模型的复杂度。与传统的

方法不同，卷积神经网络能够自动学习最优特征表示，极大地提高了特征提取的效率和效果。以具体实例说明，当输入特征图尺寸为 4×4，采用 2×2 的卷积核并以步长 1 进行滑动时，其计算过程如图 2-2 所示。

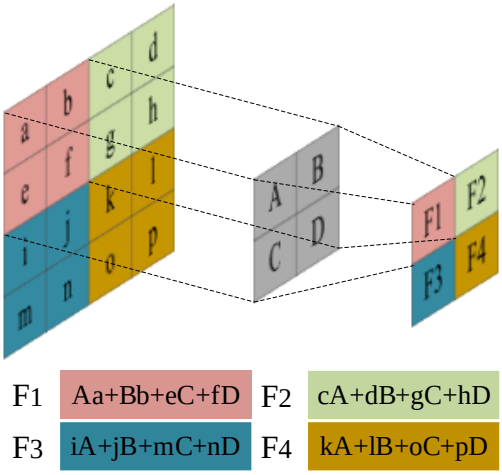



图 2-2 卷积运算过程

### 2.1.2 激活层和池化层

在卷积神经网络架构中，激活层作为关键的非线性转换模块，为网络赋予了处理复杂问题的能力。通过在神经元输出端引入非线性激活函数，网络能够学习并模拟更加复杂的特征表示。目前常用的激活函数主要分为饱和型和非饱和型两类：饱和型包括 sigmoid 函数和双曲正切(Tanh)函数，而非饱和型则以修正线性单元(ReLU)及其改进版本为代表。

其中，ReLU 函数因其计算简单且能有效缓解梯度消失问题而被广泛应用。然而，标准的 ReLU 函数在处理负值输入时存在神经元消失的问题。而 Leaky ReLU 函数通过在负区间引入一个小的斜率，有效解决了负值区域的梯度消失问题，从而提高了网络的训练稳定性和特征学习能力。这种改进使得神经网络在处理复杂非线性问题时表现出更好的性能。

$$Sigmoid: f(x) = \frac{1}{1+e^{-x}} \tag{2-1}$$

$$Tanh: f(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{2-2}$$

$$ReLU: f(x) = \max(0, x) \tag{2-3}$$

$$Leaky\ ReLU: \max(\alpha x, x) \tag{2-4}$$

而池化层是 CNN 中非常常见的隐藏层类型，专注于缩减特征图的空间维度，通过实施诸如最大池化、平均池化等操作，有效地降低了计算复杂度并减轻了过拟合的风险。因为镜像中的局部特征是相关的，所以池化镜像可以大大减少计算量，但不会丢失镜像的主要特征，假设图像的大小为一个 4×4、一个 2×2 大小采用池化，并将卷积核的滑动步长设为 2[50]。常见的池化方法如下图 2-3 所示。

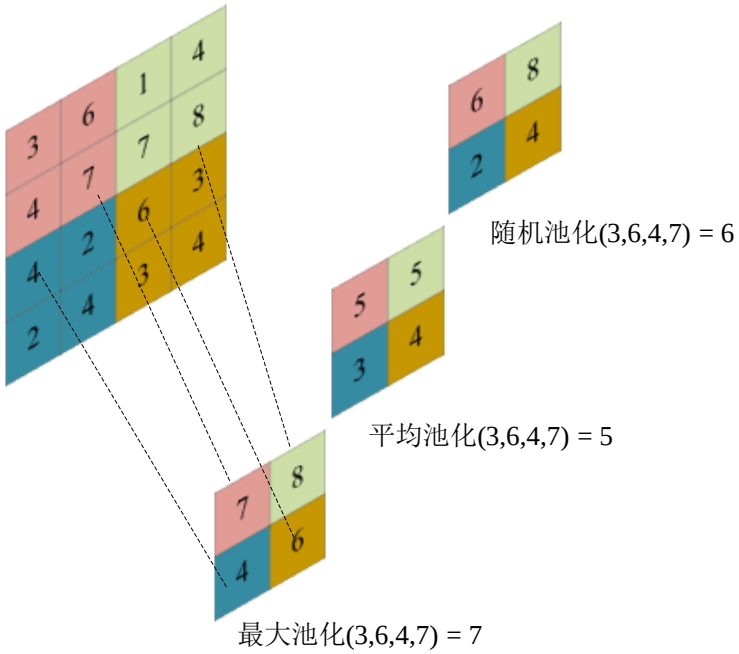



图 2-3 池化操作过程

### 2.1.3 全连接层

在卷积神经网络的层级结构中，全连接层作为特征处理的最终阶段，发挥着特征整合与分类决策的关键功能。该层通过对前序网络提取的局部特征进行全局性整合，实现特征的分类与识别。根据前序层的类型不同，全连接层的连接方式存在显著差异：当前序层为全连接层时，采用 1×1 卷积核进行处理；当前序层为卷积层时，则需使用与特征图空间维度相匹配的 $h \times w$ 全局卷积核，其中 $h$ 和 $w$ 分别代表输入特征图的高度和宽度。这种自适应连接机制确保了特征信息的有效传递，为最终的分类决策提供了可靠的特征表示。若全连接层位于整个网络的最后，当 CNN 网络输入了一张图片时，卷积层、激活层和池化层对图片信息进行局部特征提取，提取后的特征经过全连接层处理后，首先根据类

别分类成不同区域的信息，然后对这些向量或信息进行整合。

## 2.2 多尺度特征融合网络

多尺度特征融合网络是图像处理中一个被广泛使用的经典网络结构，它是在 CNN 的基础上进行改进和扩展的。低分辨率特征因其感受野比较广，所以具有足够的语义信息，但缺乏空间几何特征信息，而高分辨率部分包含丰富的空间信息，但全局语义上下文信息很少。因此，通过对特征的不同粒度采样，充分集成低分辨率和高分辨率语义特征，可以显著提高精度。多尺度特征融合网络通过 CNN 对特征信息进行多尺度提取，然后采用多尺度方法，构建多尺度的特征表达，对不同层次的特征信息进行融合。多尺度特征融合网络可以根据结构大致分为两种，一种是并行多分支网络，另外一种是串行跳跃式连接结构。多尺度特征融合网络分类如下图 2-4。

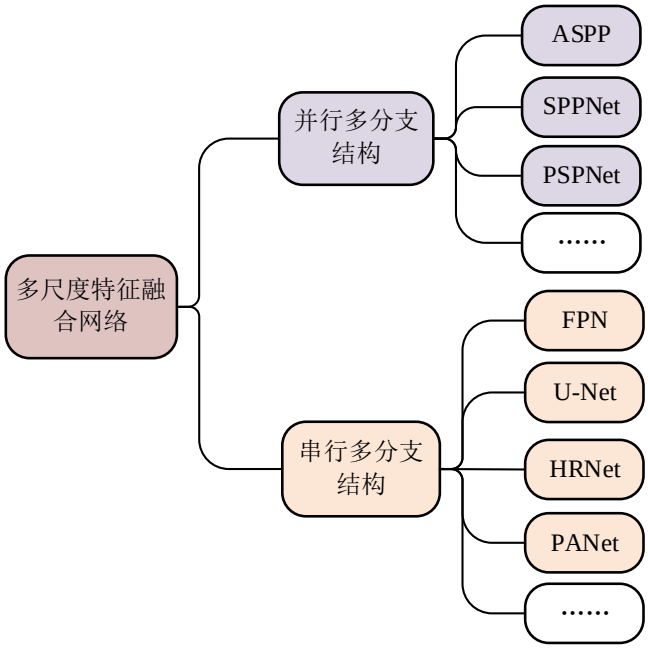



图 2-4 多尺度特征融合网络分类图

并行多分支网络的大致原理类似于电路中的并联结构，通过采用多组并行且尺寸多样的卷积核，网络模型能够从不同层级的感受野中提取特征信息，最后将不同的层次特征结合起来作为池化的全局特征。例如 PSPNet[51]结构如下图 2-5 所示。

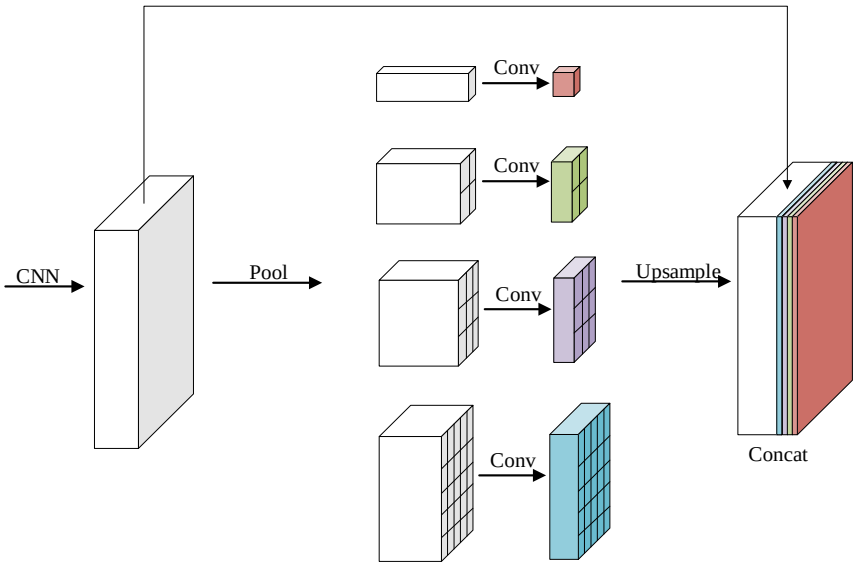



图 2-5 PSPNet 结构图

并行网络结构通过在同一层级集成多个具有不同感受野的特征提取分支，实现了多尺度特征的同步获取与融合。这种设计不仅增强了特征的多样性，还通过特征融合机制有效平衡了计算复杂度与模型表达能力，可以在某一层级中获取多尺度的特征信息。相比之下，串行结构采用层级递进的方式，通过一层一层的特征提取，再通过某一层或几层的跳层递进，最终获取到不同层级的信息，更与两个金字塔相互连接的结构相似，层层递进中穿插跳跃处理，在这种结构下，不同感受野下的图像特征信息可以被更精确的提取到并进行融合。例如常见的 FPN 网络结构如下图 2-6 所示。

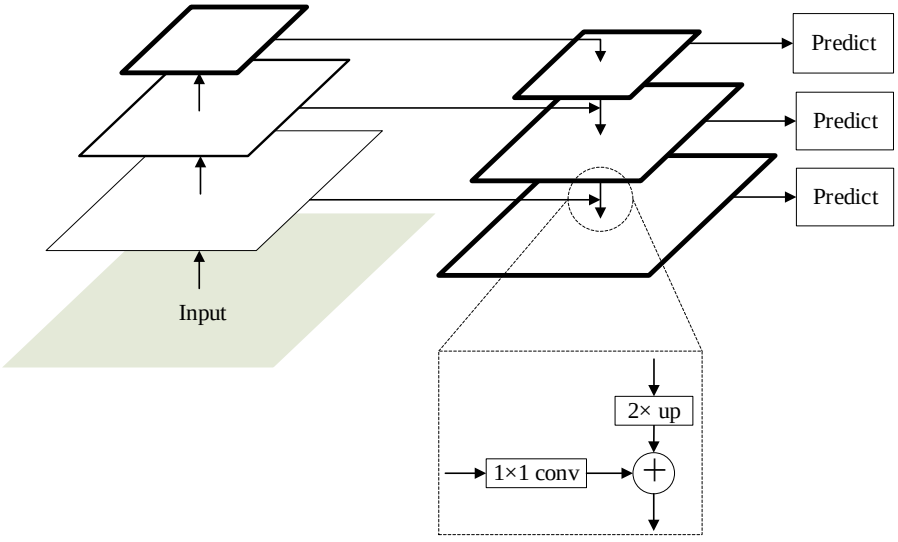



图 2-6 FPN 结构图

## 2.3 目标检测算法理论

随着用于检测任务的深度学习算法的快速发展，目标检测的性能得到了极大的提升。其主要任务是对图像或视频中的对象进行分类和定位。目标检测技术的发展已经有二十多年了，从早期的传统检测方法到现在的深度学习方法，目标检测精度和速度的提高源于深度学习技术的不断迭代。传统的目标检测技术有很多局限性，使用卷积神经网络作为目标检测的主要框架可以更高效的提取目标特征，降低手动特征提取带来的复杂性和错误率。随着视觉感知技术的不断发展和广泛应用，目标检测技术在各个领域上均取得了重大进步，其相关技术的发展可分为两个时期：传统检测方法时期(1998-2014 年)和深度学习时期(2014 年至今)。同时，现阶段涌现出多种技术手段和路线，在深度学习驱动的目标检测领域，研究方法主要分为两大方向：实时性优先的一阶段目标检测算法(One-Stage Object detector)和精度导向的二阶段目标检测框架(Two-Stage Object detector)。如下图 2-7 所示。

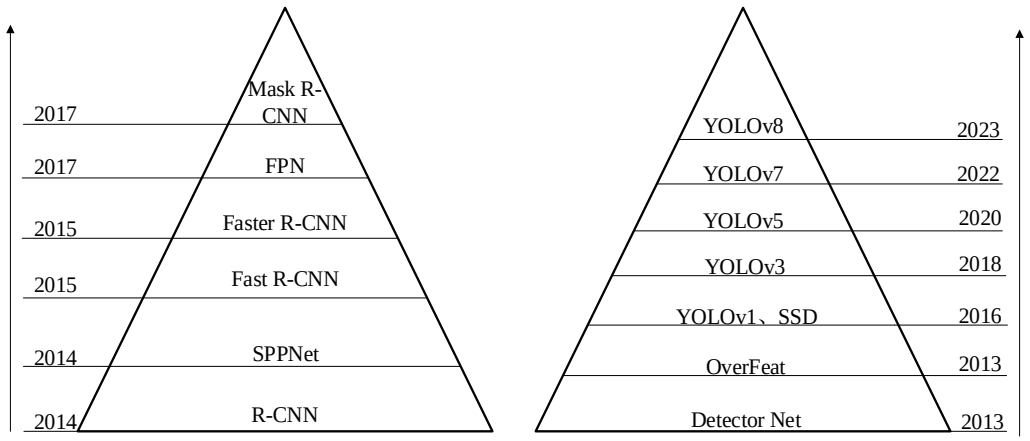



图 2-7 目标检测算法分类图

### 2.3.1 基于区域的两阶段目标检测

两阶段目标检测框架[52]划分了目标定位和目标分类的任务，首先在目标本地化的位置生成候选区域，然后根据其特定类别对该区域进行分类。这就是它被称为两阶段的原因。这些架构也称为基于区域的框架两阶段物体检测算法，其主要优点是检测精度高，但需要对特定类别区域进行重新分类，所以检测速

度和一阶段目标检测相比较慢。

基于区域的卷积神经网络[53](Region-based Convolutional Neural Networks, R-CNN)是使用深度学习方法检测目标的典型模型，该目标检测框架采用多阶段处理策略，首先，使用选择性搜索方法提取候选区域，然后使用 CNN 提取每个候选区域的特征，提取特征后，使用 SVM 分类器检测目标并进行判定。最后基于分类结果生成边界框。

尽管 R-CNN 在目标检测方面取得了很大的进步，但它仍然存在一些局限性，例如目标检测速度慢、多阶段管道训练和选择性搜索方法繁琐。由于 R-CNN 为每张图像生成 2000 个候选区域[54]，而全连接层使得输入大小固定，因此从这些区域提取 CNN 特征是主要难题。为了克服这个困难，引入了一种称为空间金字塔池化网络层(Spatial Pyramid Pooling Network, SPPNet)的新技术。SPPNet 层被添加到最终卷积层的顶部，为全连接层提取固定长度的特征，而不去约束 RoI 感兴趣区域的大小如何，虽然可能会导致部分信息丢失，但通过使用 SPPNet层，R-CNN 的检测速度在得到很大提升的同时，检测精度也仍良好。卷积层只需要在完整的测试图像上运行一次，即可为随机大小的区域获取固定长度的特征。

Mask R-CNN 框架[55]实现了在模型上进行目标检测与实例分割的统一训练，这是对 Faster R-CNN 的增强，用于解决执行目标检测和语义分割任务的实例分割问题，Mask R-CNN 的目标是执行像素级的分割。通过采用 RoIAlign 层替代传统的 RoI 池化层，有效解决了特征图与原始图像之间的空间错位问题，能够更好地保持目标的空间信息，为精确的实例分割提供了可靠保障。

如上所述，基于区域检测的框架由各个不同阶段组成，这些阶段彼此连接并单独训练。包括生成候选区域、使用 CNN 提取特征、分类和边界框回归，尽管这些方法能够实现检测的高精度，但在实时速度上存在一些问题。随着 CNN 的不断发展，这个问题可以通过统一的阶段检测器来克服。如删除区域预测阶段，并在单个 CNN 中实现特征提取和预测回归。

### 2.3.2 基于回归的一阶段目标检测

一阶段目标检测框架的深度卷积神经网络同时对目标进行定位和分类，而

无需将它们分为两部分，其核心思想是直接对输入图像进行网格划分，并在每个网格上预测目标的类别和位置，无需像两阶段算法那样先生成候选区域再进行分类和回归，在这种情况下，只需要通过神经网络进行一次传递。它们将图像像素直接映射到边界框坐标和类概率。如下列出了几个比较经典的一阶段目标检测框架如 YOLO 系列等。

YOLO 的核心思想是将检测任务转化为回归问题进行处理。该框架通过单次前向传播即可完成目标定位和分类，实现了端到端的优化，直接预测目标边界框的空间坐标及其类别概率，这种设计摒弃了传统方法中复杂的候选区域生成步骤。与基于区域建议的检测方法相比，YOLO 充分利用全局上下文信息，将整张图像的特征统一考虑。在 YOLO 目标检测中，图像分为 S×S 网格；每个网格由五个元组（$x$、$y$、$w$、$h$ 和置信度分数）组成。单个目标的置信度分数基于概率。此分数会给每个类，并且无论哪个类的概率较高，该类都会被赋予优先权。边界框的宽度 $w$ 和高度 $h$ 参数是根据对象的大小预测的。从重叠的边界框中，选择具有最高 IoU 的框，并删除其余框。YOLO 算法处理步骤如下图 2-8 所示。

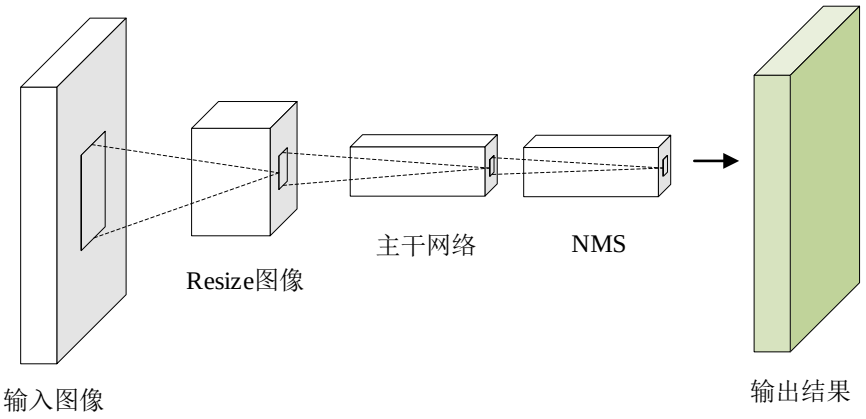



图 2-8 YOLO 算法示意图

单次多框检测器(Single Shot MultiBox Detector, SSD)是一种用于多个类别的快速单发多向检测算法，由 Liu 等人实现。它构建了一个统一的检测器框架，该框架比 YOLO 更快，比 Faster R-CNN 更加准确。SSD 的设计结合了 YOLO 模型的回归思想和 Faster R-CNN 算法的锚定过程。通过使用 YOLO 的回归方法，SSD 降低了神经网络的计算复杂性，从而保证了实时性能。通过设定锚点的办法，SSD 能够提取各种大小和纵横比的特征，以确保检测精度，SSD 使用 VGG-

16 作为主干网络。SSD 的过程基于前馈 CNN，该 CNN 为这些框中存在的目标类实例生成固定大小的边界框和对象性分数，然后利用非极大值抑制 NMS 来输出最终检测效果，还通过使用 RPN 来提高检测准确度和速度，使用一些辅助数据增强模型性能，SSD 在各种基础数据集上检测效果良好。

## 2.4 多目标跟踪算法理论

多目标跟踪算法的主要任务是在连续的视频帧中，对多个目标进行实时、准确的跟踪，并赋予每个目标唯一的标识(Track ID)。其基本原理是通过目标检测算法获取每帧图像中的目标位置信息，然后利用这些信息以及目标之间的关联性实现目标的跟踪。多目标跟踪算法主要分为基于检测的跟踪(DBT)和基于初始框的跟踪(DFT)两大类。DFT 是单目标跟踪领域的常用初始化方法，即每当新目标出现时，人为告诉算法新目标的位置，过程繁琐。所以 DBT 相对来说更受欢迎。

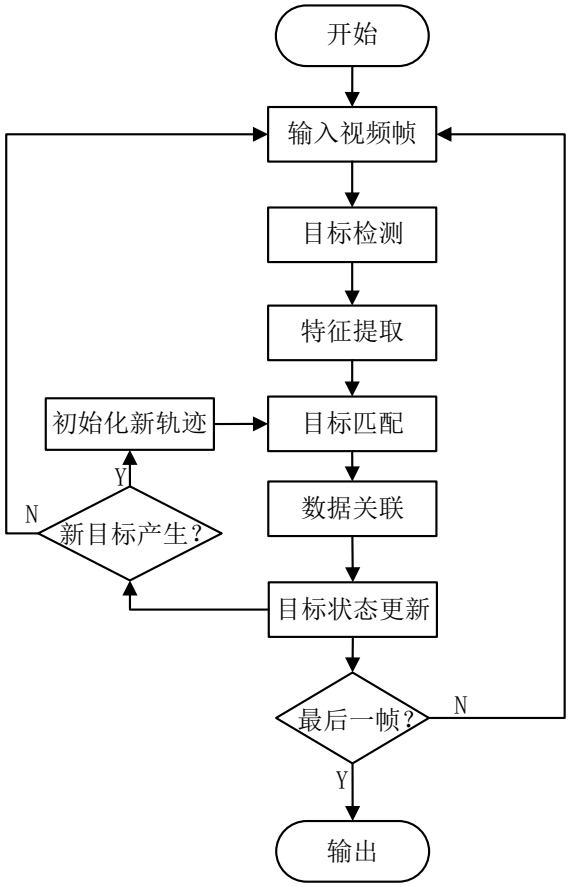



图 2-9 基于检测的跟踪流程图

DBT 的方式就是典型的 tracking-by-detection 模式，即先检测目标，然后将目标关联进入跟踪轨迹中。首先利用深度学习模型（如 YOLO、SSD 等）对视频帧中的目标进行检测，获取每个目标的位置信息。然后对于每个检测到的目标，提取其外观特征和运动特征，这些特征将用于后续的目标匹配和追踪。之后，通过计算相邻帧中目标之间的距离和速度差异，或者利用特征相似度进行匹配，将当前帧中的目标与上一帧中的目标关联起来。常用的匹配算法包括匈牙利算法、级联匹配等。接着根据匹配结果，更新目标的状态信息，包括位置、速度等，并初始化新的目标轨迹（如果有新目标出现）。最后，输出跟踪结果，包括目标的轨迹、位置信息等，供后续分析或应用使用。

在 DBT 中，现已涌出各种各样类型的跟踪模型，如比较经典常用的 SORT、DeepSORT、FairMOT、BoT-SORT 和 ByteTrack 算法，接下来对 SORT 算法进行简单阐述。

SORT 算法框架是一种典型的基于深度学习的多目标跟踪算法，它融合了 Faster R-CNN 检测器与卡尔曼滤波技术的优势。具体而言，该算法首先借助 Faster R-CNN 目标检测方法从视频帧中精准地识别并提取目标特征，生成初始检测框。随后，利用卡尔曼滤波器对目标在后续帧中的潜在位置进行预测。为了有效地将预测结果与新的检测框进行关联，SORT 算法引入了匈牙利匹配算法，该算法通过计算 IoU 值这一关键指标，通过计算距离度量函数来完成匹配过程，确保目标轨迹的连续性。SORT 算法的主要工作流程是：

（1）通过视频序列输入第一帧，通过第一帧结果初始化卡尔曼滤波目标检测器得到目标框，同时卡尔曼滤波器预测当前的帧的轨迹，然后将目标框与轨迹进行 IoU 匹配，未达到 IoU 阈值的删除。

（2）数据关联后的轨迹筛选结果可归纳三种情况：首先，对于未能成功匹配的轨迹，将其标记为失配状态。若某条轨迹连续 T 帧未能完成匹配，则判定该目标丢失，系统将移除其对应的目标 ID。其次，当检测到未匹配的目标框时，则该目标尚未建立对应轨迹，系统将其初始化新的轨迹。最后，对于成功匹配的轨迹和目标对，系统采用卡尔曼滤波算法对目标状态进行更新，通过融合观测值和预测值来优化轨迹估计精度。

（3）重复上述步骤直至视频最后一帧结束。

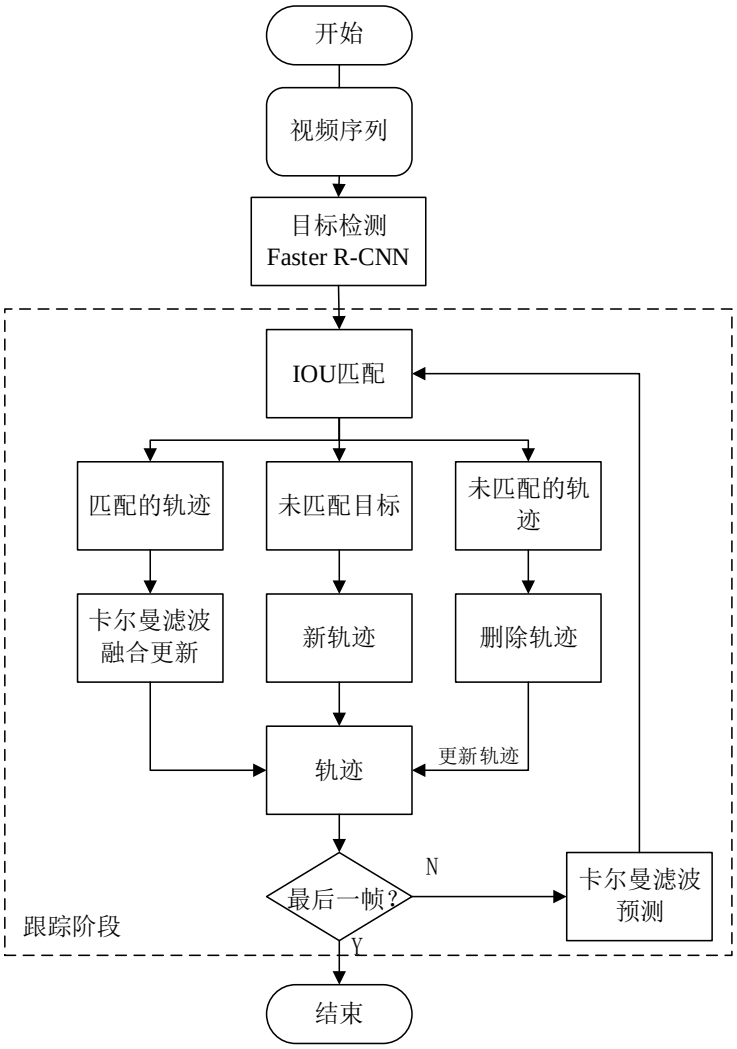



图 2-10 SORT 算法流程图

## 2.5 本章小结

本章主要阐述了本文所使用的部分算法模型的理论基础，为接下来的研究内容提供了充分的技术路线和理论支持。首先，对深度学习中经典的卷积神经网络进行了介绍，说明了它的主要结构及其作用，其次，对多尺度特征融合网络进行了阐述，介绍了几种常见的网络模型，然后，对目标检测算法的两种分类分别列举了几种常见的算法，并对其优缺点进行分析。最后，对目标跟踪算法进行简单介绍，引出相关的多目标跟踪算法，介绍了常见基于检测的跟踪算法及其工作流程，为后面的车辆多目标跟踪算法提供了技术支撑。

# 第 3 章 基于 YOLOv7-PGC 的车辆目标检测算法

YOLOv7 是一种被广泛应用的深度学习算法，常被用于实时目标检测任务中。其核心思想是将目标检测任务转换为一种类似于回归问题的方式进行处理，通过单一的神经网络对图像进行前向传播，同时预测所有边界框坐标和类别概率，从而极大地提高了目标检测的实时性。总的来说，YOLOv7 在保留了 YOLO 系列算法实时性优点的同时，通过改进模型结构和训练策略等手段，提高了目标检测的准确性和鲁棒性，成为了目标检测领域的经典算法之一。本章通过对网络结构的优化，构建了一个多尺度特征融合网络，增强模型对复杂特征信息的提取与融合，同时，通过引入协调坐标卷积使得模型的参数量与计算量获得显著的降低，结果表明，优化后的模型在复杂道路环境数据集中获得更优异的效果。

## 3.1 YOLOv7 目标检测算法

相比于之前的 YOLO 算法，YOLOv7 在模型结构中引入了模型重参数化思想，将原本的卷积网络中的卷积层、Batch Normalization 层、激活函数 SiLU 等模块放在 CBS 模块中，提高了模型的表达能力和训练效率。对于 CBS 模块，可以看从图中可以看出它是由一个 Conv 层，也就是卷积层，一个 BN 层，也就是 Batch normalization 层，还有一个 SiLU 层，这是一个激活函数。SiLU 激活函数是 Swish 激活函数的变体，两者的公式如下所示：

$$Silu(x) = x \cdot sigmoid(x) = \frac{x}{1+e^{-x}} \tag{3-1}$$

$$Swish(x) = x \cdot sigmoid(\beta x) \tag{3-2}$$

从下图 3-1 CBS 模块架构图中可以看出，CBS 模块这里有三种不同的卷积核。第一个 CBS 模块，它是一个 1x1 的卷积，stride(步长为 1)。主要用来改变通道数，第二个 CBS 模块，它是一个 3x3 的卷积，stride(步长为 1)。主要用来特征提取，第三个 CBS 模块，它是一个 3x3 的卷积，stride(步长为 2)。主要用来下采样。

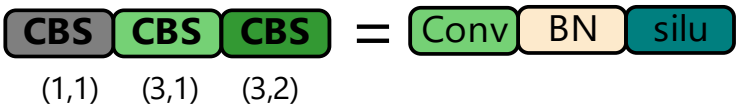



图 3-1 CBS 模块架构图

相对于 ReLU 函数，SiLU[56]函数在接近零时具有更平滑的曲线，并且由于其使用了 sigmoid 函数，可以使网络的输出范围在 0 和 1 之间。如下图 3-2 所示。

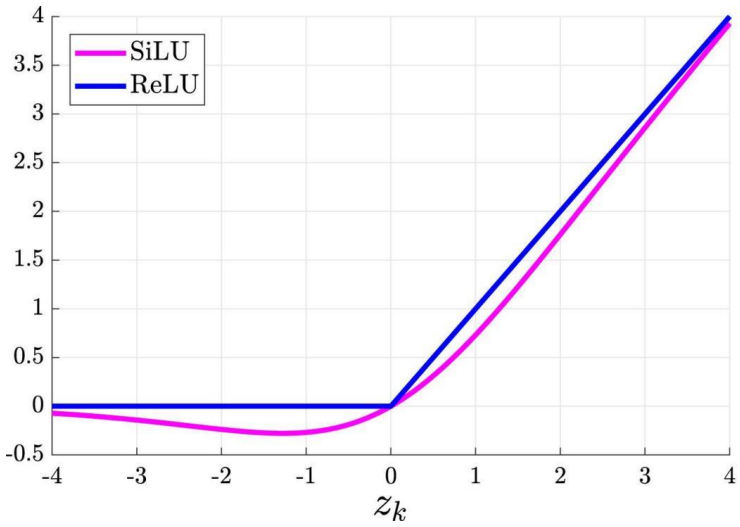



图 3-2 SiLU 函数曲线图

在标签分配方面采用了 YOLOv5 的跨网格搜索以及 YOLOX 的匹配策略进行标签分配[57]，使得模型在训练过程中能够更好地处理不同大小、形状的目标，提高了模型的鲁棒性。同时 YOLOv7 提出了一个新的网络架构 ELAN，以高效为主，通过对特征图进行多次下采样，使得模型能够对不同大小的目标进行有效的特征提取，进而提高模型的准确性。通过引入辅助头(auxiliary head)进行训练，使得主模型和辅助头可以相互辅助，进一步提高了模型的准确性和泛化能力。

YOLOv7 的主干网络(Backbone)、颈部(Neck)、检测头部(Head)三部分构成了目标检测的主干网络，其中：

Backbone 用于特征提取，通过多尺度特征融合，使得模型对不同尺度的目标都具有较好的检测能力。在 YOLOv7 中，Backbone 网络主要由 CSPDarknet53 和 SPP 模块构成，CSPDarknet53 相对于常规的 Darknet53 在计算量和特征提取能力上有所提升，SPP 模块则能够进行多尺度特征融合。

Neck 部分在 Backbone 网络之后，用于进一步提取和融合特征，通常采用一

些结构如 FPN、PANet 等。

在 YOLOv7 中，Neck 部分采用了类似于 YOLOv5 的融合策略，即将不同大小的特征图进行上下采样并相加以获得更全面的目标信息。Head 部分用于最终的目标检测，一般包含分类器、回归器和非极大值抑制等模块。在 YOLOv7 中，Head 部分与 YOLOv5 有所不同，将 Neck 层与 Head 层合称为 Head 层，但功能上保持一致，即对不同任务进行预测和分类。

在整体的网络结构上，YOLOv7 首先对输入的图片格式进行预处理，使其图片大小统一为 640×640，然后再将其输入到 Backbone 网络中进行特征提取操作。根据 Backbone 网络的三层输出，在 Neck 层进行特征融合。最后，通过 Head 层对不同大小的特征图进行处理，输出最终的目标检测结果。

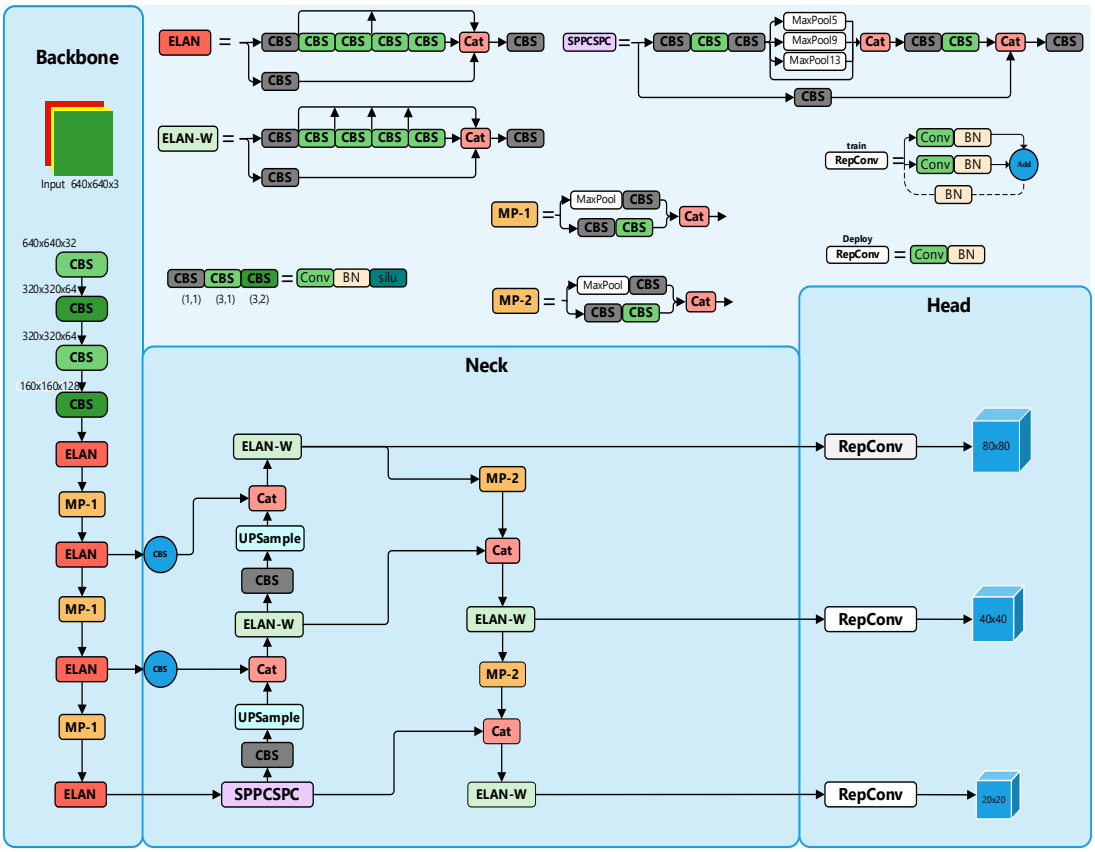



图 3-3 YOLOv7 网络结构图

## 3.2 YOLOv7-PGC 目标检测算法模型

### 3.2.1 改进的 YOLOv7 网络

针对复杂的交通环境，自动驾驶系统的感知和检测精度低、原始 YOLOv7 算法难以满足多目标检测需求的问题，提出一种 YOLOv7-PGC 模型。通过采用特征增强技术，对原网络结构进行优化，实现了多尺度特征的融合，显著增强了模型的特征表达能力。此外，引入聚集-激励(Gather-Excite, GE)注意力机制，进一步强化多尺度特征的提取能力，有效提升目标检测的精度。同时，在模型的颈部和检测头部分，融入了协调坐标卷积(CoordConv)，显著增强网络对空间信息的捕捉能力，优化网络的学习能力和效果。YOLOv7-PGC 的网络结构图如下图 3-4 所示。

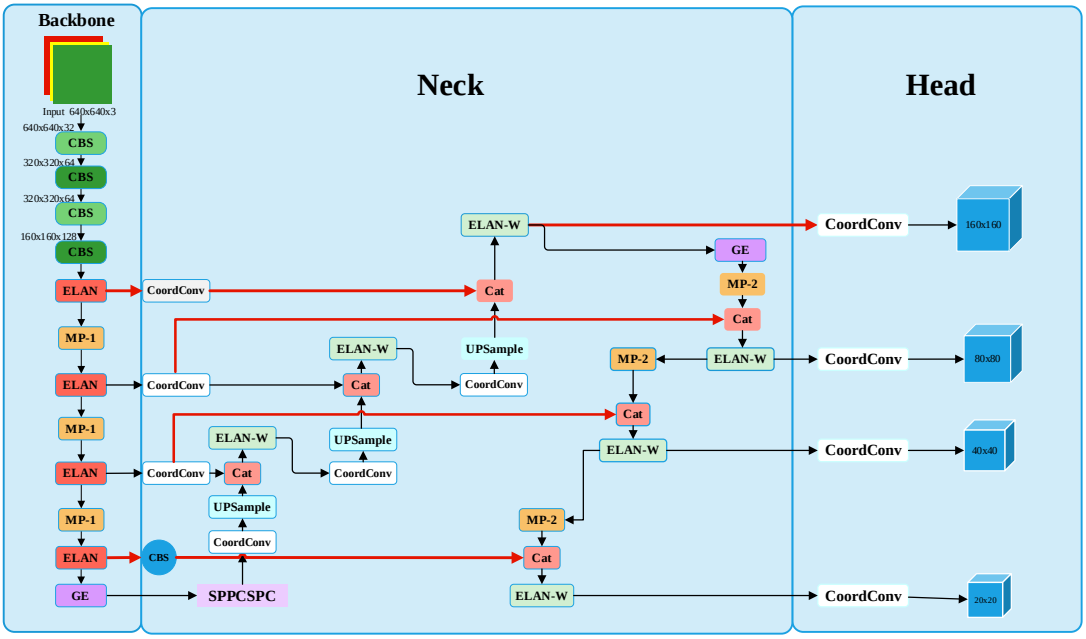



图 3-4 YOLOv7-PGC 结构图

改进后的网络结构将原来连接颈部与骨干网络结构的普通卷积和头部的 RepConv 替换成了协调坐标卷积。本文在 YOLOv7 骨干结构中融合了 SPPF 层之前的 GE 注意力机制模块，以分配不同的权重。融合 GE 注意力机制模块的 YOLOv7 网络结构可以着重关注道路环境中小目标的特征信息，并且过滤掉其他无关信息，同时提取目标的弱小特征，以解决信息复杂而导致的效率差等问题，从而提高目标检测的精确度和速度。在自动驾驶环境感知的实际场景中，在出行高峰期或拥挤的街道上，小而密的物体更多，检测密集小目标的准确性和及时性对交通行车安全起着重要性的作用[58]。针对 YOLOv7 在处理小目标检测时由于网络深层特征图的下采样率较高，导致小目标的特征信息在传播过程中逐渐丢失。本文在原有架构中新增了专门的小目标检测层，该层能够有效保

留小目标的细节特征。同时，考虑到深层网络在特征提取过程中会损失浅层的位置信息，基此在网络设计中引入了跨层连接机制，将主干网的低级特征直接传递到颈部网络。这种改进不仅增强了位置信息的保留能力，还实现了多尺度特征的融合，从而显著提升了模型对小目标的检测性能。

### 3.2.2 多尺度特征融合结构

车辆行驶中道路环境比较复杂，特别是在一些拥堵路段，感受野更广则可以捕捉到更多的物体信息。在 Yolov7 中，FPN[59]和 PANet 被融合在一起得到 PAFPN 结构，它在 PANet 的基础上增加了级联注意力机制，可以融合不同层次的特征图，更好地捕捉不同尺度下的目标特征，从而提高目标检测的精度。PANet 是一种用于实例分割的先进神经网络模型，它通过引入自上而下和自下而上的路径聚合，有效地提高了对小目标和遮挡目标的处理能力。如下图 3-5 为 PANet 的结构图。

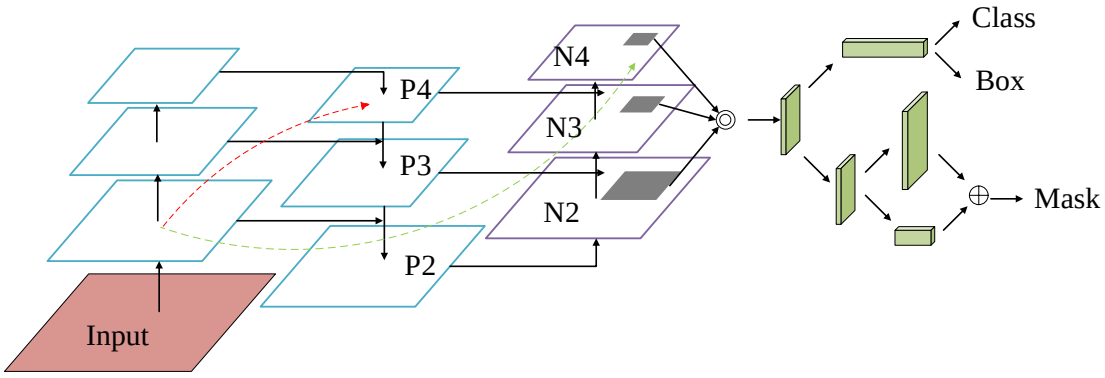



图 3-5 PANet 结构图

其中在特征融合方面，PANet 结构采取张量连接(concat)代替了原来的捷径连接，即从原来的加法的特征融合方式改进为鲁棒性更高的特征堆叠。

改进后的 YOLOv7 结构在原算法融合 FPN 与 PANet 结构的基础上增加了一个小目标检测层可用于生成 160×160 特征图，拓宽网络的检测范围，虽然这种方法会带来计算量和检测速度的小幅提高，但可以有效提高检测精度，这种结构可以更有效地利用特征信息，从而提高目标检测的精度。去除将原来 P5 与 N5 的级联，同时在骨干网络与颈部之间添加了一个从浅到深的跨越连接如图 3-6 虚线所示，这样就可以在空间上补充特征图的详细信息，在随后的采样过程中可以提取出更准确的特征，有利于密集小目标的检测。本文在特征图上执行

卷积操作、特征提取和上采样操作 P3 层以进一步扩展特征图。然后将输出与特征图融合以生成大小为 160×160 的 P2 层。通过 ELAN 模块提取特征后，N2 层用作输出层，相当于 P2 层。如下图 3-6 为设计的多尺度特征融合结构。

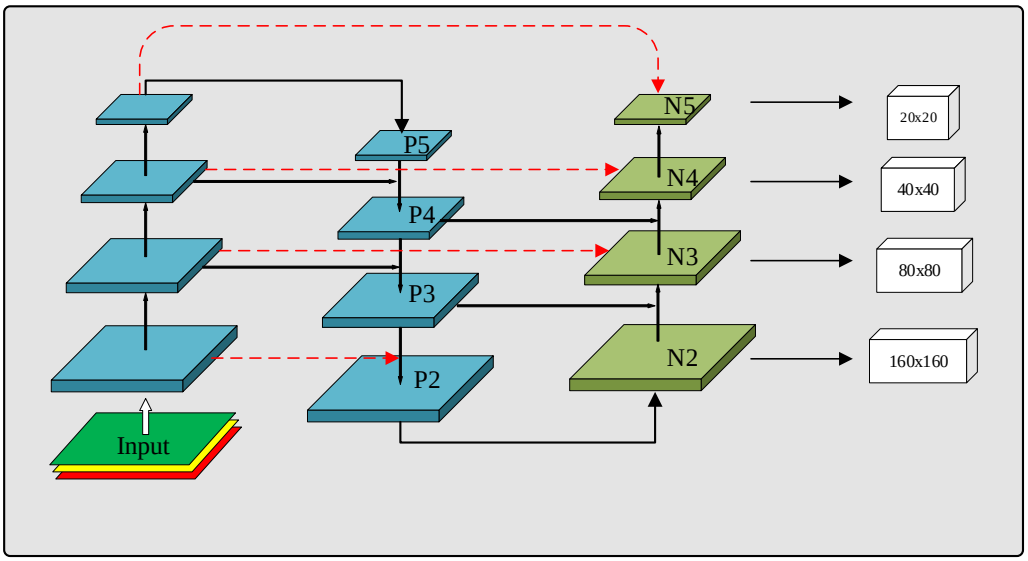



图 3-6 设计结构图

综上所述，通过对多尺度特征融合结构的优化，增强对上下文信息的整合，融合不同尺度的特征来增强目标的边缘、纹理和形状等特征，能够更好地适应尺度变化，提高模型整体的准确性和鲁棒性。

### 3.2.3 添加 GE 注意力机制

本研究将GE注意力机制模块添加到特征提取层之后，通过利用空间注意力挖掘特征之间的上下文信息，以便帮助神经网络在处理输入序列时更有效地辨识关键信息，从而提升模型的总体效能与表现，引入了一种轻量级的模块。这一模块可以灵活地融入每个残差单元中，类似于 SE 块的作用，但具有更高的效率与适应性。通过这一创新设计，神经网络能够更好地理解和利用输入数据的内在特征，从而在各种任务中展现更卓越的性能。它在抑制噪音的同时强调重要特征。这种机制从上下文角度出发，提出了 SE 的更一般的形式 GE，即 Gather 和 Excite，并利用空间注意力来更好的挖掘特征之间的上下文信息。其中，Gather 操作用于从局部的空间位置上提取特征，Excite 操作是一种与 Encoder-Decoder 模型运行很相似的编解码过程，用于将其进行缩放还原回原始尺寸，可以以很小的参数量和计算量来提升网络的性能。通过对整体图像信息

的相关性建模，来捕获长距离的统计信息，充分利用到了图像的上下文信息。GE 注意力机制如下图 3-7 所示。

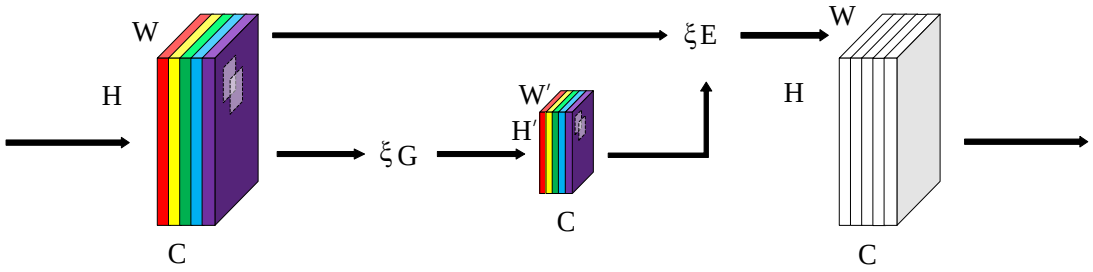



图 3-7 GE 注意力机制

Gather-Excite[60]通过引入一对运算符 ξ G（逐步深度卷积）和 ξ E，进一步利用了模型中的特征上下文。GE 模块的核心是在特征图上逐层使用不同的过滤器，Gather 可以在大空间尺度上有效地聚合特征响应，然后使用 excite 将聚合信息重新分发到局部特征中。GE 模块的结构如图 3-7 所示。经 GE 模块处理后，改进 YOLOv7 模型能够更精确地聚焦局部特征，大大提高了特征提取能力。ξG 定义得到相关公式：

$$\xi \mathrm{G}: \mathrm{R}^{H\times W\times C} \rightarrow \mathrm{R}^{H'\times W'\times C} \tag{3-3}$$

其中 $H$、$W$ 和 $C$ 表示图像输入 $x$ 的高度、宽度和通道，$e$ 表示范围比率，H′=H/e，W′=W/e。使用全局平均池化作为 GE 模块的全局范围比率。ξE 的定义公式为：

$$\xi \mathrm{E}（\mathrm{x}, \hat{\mathrm{x}}） = \mathrm{x} \odot f(\hat{x}) \tag{3-4}$$

$$f: \mathrm{R}^{H'\times W'\times C} \rightarrow [0,1]^{H\times W\times C} \tag{3-5}$$

$\hat{x}$表示处理后的输出 ξ G，⊙表示 Hadamard 乘积，$f$表示映射关系，是负责重新缩放和分发聚合中的信号的映射。将 $V$ 设为输出，则 GE 模块后的结果 $G$ 其定义公式为：

$$\mathrm{V} \in \mathrm{R}^{\left[\frac{\mathrm{H}}{\sigma}\right]\times\left[\frac{\mathrm{W}}{\alpha}\right]\times \mathrm{C}}, \mathrm{v} \in \mathrm{R}^{\left[\frac{\mathrm{H}}{\alpha}\right]\times\left[\frac{\mathrm{W}}{\sigma}\right]} \tag{3-6}$$

$$\mathrm{G} = \sum_{\mathrm{s}=1}^{\mathrm{C}} \xi \mathrm{E}(\mathrm{u}^{\mathrm{s}}, \xi \mathrm{G}(\mathrm{u})_{\mathrm{v}}^{\mathrm{s}}) = \sum_{\mathrm{s}=1}^{\mathrm{C}} \mathrm{u}^{\mathrm{s}} \odot \mathrm{g}\left(\xi \mathrm{G}\left(\mathrm{u}^{\mathrm{s}} \odot \mathrm{T}_{\tau(\mathrm{v},\mathrm{x})}^{\mathrm{s}}\right)\right) \tag{3-7}$$

其中 $\mathrm{T}(\mathrm{v}, \alpha) = \left\{\alpha \mathrm{v} + \theta : \theta \in \frac{[-[\alpha-1], [2\alpha-1]]^2}{4}\right\}$，$\alpha$ 为所选范围比。$T_{\{\cdot\}}$表示张量，$g$是明确定义的映射。

### 3.2.4 检测头部改进

传统的卷积神经网络具有平移等变性，这意味着当输入特征图发生平移时，卷积神经网络的输出不会发生改变。然而，当需要感知位置信息时，传统卷积神经网络就满足不了需求。与普通卷积相比，CoordConv 在输入中添加了两个坐标通道来表示每个像素点的坐标信息：一个用于 $x$ 坐标，一个用于 $y$ 坐标。将两个坐标通道与输入通道拼接，然后进行卷积运算，为网络提供空间信息，这有助于网络了解图像中特征的空间对应关系。借助空间信息，网络可以构建更强的空间建模能力，增强对局部和全局位置的理解，并改进基于位置的推理。

普通卷积的计算公式可以定义为：

$$O = \left\lfloor \frac{n+2p-f}{s} \right\rfloor + 1 \tag{3-8}$$

其中，用 n×n 表示输入图片的尺寸大小，$f \times f$表示卷积核的大小，$p$是填充大小，$s$用来表示步长大小，当计算结果不是整数时，向下取整，最后得到结果 $O$。

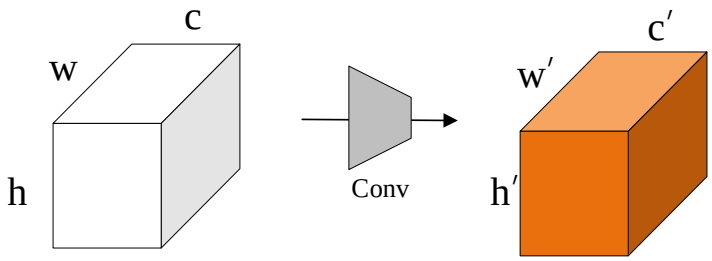



图 3-8 常规卷积

协调坐标卷积能够在保持卷积的计算效率和参数效率的同时，根据具体任务需求学习到完全的平移不变性或不同程度的平移依赖关系。这一特性使得它在处理坐标变换问题时表现出色，并且相较于传统卷积方法，协同坐标卷积的参数数量大幅减少，从而提高了模型的效率和泛化能力。本文在网络模型结构中引入一种新的卷积，其名称为协调坐标卷积，作用是在减少参数量同时提高检测效率。本文将协调坐标卷积替换 FPN 中的上采样前后位置的卷积和检测头前的 REP 卷积，CoordConv 结构如下图 3-9 所示。

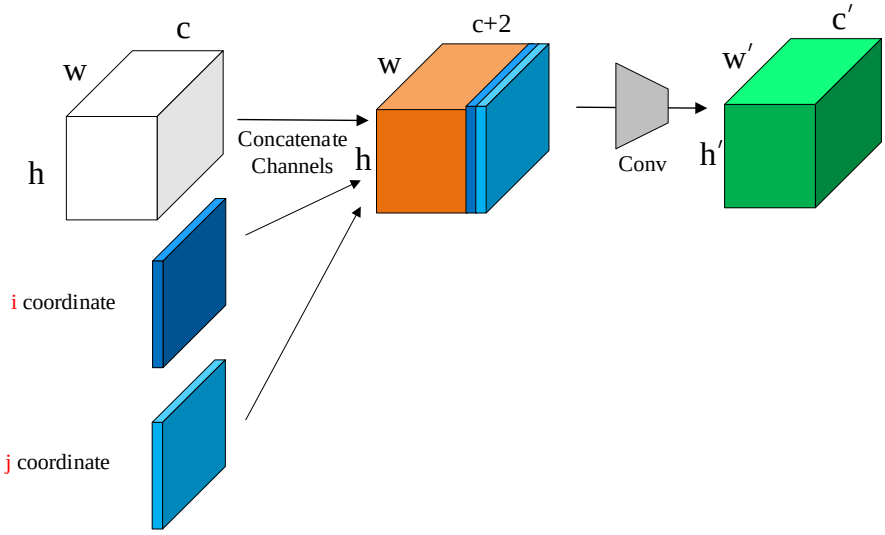



图 3-9 协调坐标卷积

在每次通过卷积提取参数化特征之前，输入张量的每个像素都会被赋予相应的坐标信息。有 $x$ 和 $y$ 坐标信息。由于每次特征提取期间张量的大小都不同，因此 $x$ 和 $y$ 会进一步缩放到[-1,1]之间。生成的两个二维坐标矩阵被分配给相应的张量通道层。通过卷积层操作，进一步获得相应的具有像素级坐标信息的高级特征，为头部层的回归定位提供了更丰富的特征信息。重要的是，提取的高级特征具有相应的坐标信息，对提高车辆道路环境下小目标的检测有很大帮助。

## 3.3 实验结果与分析

### 3.3.1 实验环境与数据集

（1）检测实验数据集

为了验证改进的 YOLOv7 算法在自动驾驶环境感知过程中的有效性和真实性，使用权威的公共数据集 BDD100K 进行评估实验，BDD100K 数据集[61]是伯克利大学 AI 实验室(BAIR)发布的目前规模最大、内容最具多样性的公开驾驶数据集。它是在现实生活中收集的，并且对十类目标进行标记。BDD100K 数据集庞大而多样，包含 100,000 张图像。此外，数据集的背景包含六种不同的天气条件：晴天、多云、多云、雨天、下雪和有雾。为了更好地验证模型的性能，本文随机选取了数据集中包含标签的 2003 张图像，并以 8:2 的比例对选中的图像

数据进行了重新分区。训练集数为 1603，验证集数为 400。部分数据集图如下图 3-10 所示。

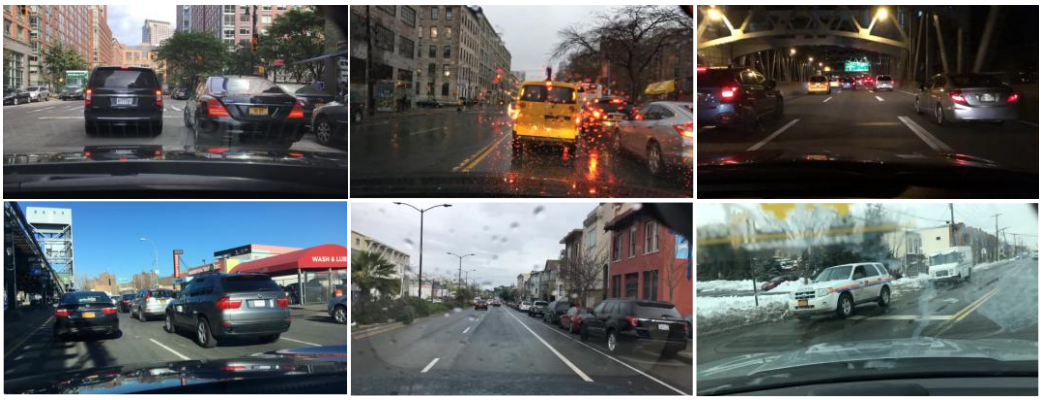


图 3-10 检测数据集示意图

（2）实验环境设定

本文对 BDD100K 数据集进行了消融实验和对比测试，验证了 YOLOv7-PGC 算法的有效性。表 3-1 为实验环境等参数，用于执行实验的计算机操作系统为 Ubuntu 18.04 系统。CPU 配置类型型号为 12th Gen Intel(R) Core(TM) i7-12700 2.10 GHz。GPU 型号为 NVIDIA RTX A5000(24G)×2。编程语言结构为 Python 3.9.10。加速环境是 CUDA 12.1，深度学习框架是 Pytorch 1.12.1。设置模型训练的参数为：网络的输入图像大小为 640×640×3，优化器为 Adam，初始学习率设置为 0.01。此外，动量因子设置为 0.937，权重衰减设置为 0.0005，周期数为 300。而且，训练阶段分为冻结阶段和解冻阶段，其中特征提取网络在冻结阶段没有变化，从而减少内存占用，提高训练效率。

表 3-1 实验环境参数配置表

| 配置名称 | 版本及型号 |
|---|---|
| 操作平台 | Ubuntu18.04 |
| GPU | NVIDIA RTX A5000(24G)×2 |
| CPU | 12th Gen Intel(R) Core(TM) i7-12700 2.10 GHz |
| 内存 | 128G |
| 深度学习框架 | Pytorch1.12.1 |
| CUDA | v12.1 |

### 3.3.2 评估指标

本文采用一系列的指标来评估所提方法的有效性。这些指标包括精确率(Precision)、参数量(Parameters)和平均精度(mAP)。平均精度(mAP)是精度-召回率曲线下区域和坐标轴上的精度值的平均值。此外运算量(FLOPs)、每秒帧数(FPS)用于评估模型的大小、模型的检测速度和计算成本：

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{3-9}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{3-10}$$

$$\text{FPS} = \frac{\text{Frames}}{\text{Time}} \tag{3-11}$$

$$\text{mAP} = \frac{\sum_{i=1}^{S} \text{AP}_i}{\text{S}} \tag{3-12}$$

其中 TP 代表正确检测框，FP 代表误检框，FN 代表漏检框，AP 代表一个目标的检测精度，S 代表检测类别数，Frames 代表帧数，Time 代表检测时间。精确率 P 表示分类器判定为正类的样本中，确实为正类的样本所占的比例。召回率 R 又称查全率，即分类器认为是正类并且确实是正类的部分占所有确实是正类的比例，召回率越高，代表模型能将准确目标找出的能力强。FPS 则是衡量模型能找到准确目标的检测速度，越快说明模型 FPS 越高，模型性能越好。mAP 则是比较常用的评价模型准确率，mAP 越高，说明准确率越好，模型检测效果也越好。

### 3.3.3 结果分析

（1）消融实验

本文通过消融实验，对比了不同改进方法的检测指标，深入分析了各个改进部分对网络性能的具体提升情况，从而验证了这些改进措施的有效性。

表 3-2 不同改进方法消融实验结果

| 组别 | PAFPN | GE | CoordConv | Params/M | mAP/% | FLOPs/G |
|---|---|---|---|---|---|---|
| 1 | | | | 37.2 | 42.6 | 105.3 |
| 2 | * | | | 37.7 | 43.8 | 119.4 |
| 3 | * | * | | 37.8 | 47.2 | 119.8 |
| **4** | * | * | * | **37.1** | **48.2** | **117.5** |

表 3-2 为不同改进方法对模型性能的改进结果，并用“ * ”代表添加改进方

法。在本章所实验的 4 组消融实验中，组别 1 代表原 YOLOv7 算法在数据集上的结果，组别 2 代表在 YOLOv7 算法上改进 PAFPN 结构，融合多尺度特征网络，由表 3-2 可以看出 YOLOv7 原始算法的 mAP 为 42.6%，而在本文改进多尺度特征融合网络后，第 2 组实验的 mAP 达到 43.8%，相比较原来提高 1.2%，这说明改进 PAFPN 结构可以融合不同层次的特征图，更好地捕捉不同尺度下的目标特征。第 3 组实验表示在第 1 组基础上改进 PAFPN 结构和在网络中添加 GE 注意力机制，添加改进后的 mAP 达到 47.2%，对比第 1 组提高了 4.6%，同时对比第 2 组提高了 3.4%，可以看出本文添加 GE 注意力机制可以有效联系上下文信息，捕捉特征信息，提高模型能力。第 4 组是在 YOLOv7 的基础上改进 PAFPN 结构、添加 GE 注意力机制以及引入 CoordConv，mAP 达到 48.2%，对比前面几组实验，mAP 分别提升了 5.6%、4.4%、1%。同时参数量与运算量相比第 3 组下降了 0.7M、2.3G，由此可以看出协调坐标卷积可以在更好地感知空间位置信息的同时，减小模型的参数量和运算量大小，提高模型的效率。对比原算法提高了 5.6%，说明本文改进的算法在自动驾驶环境中的有效性。

表 3-3 改进前后对比

| 模型 | Car | Bus | Person | Bike | Truck | Motor | Train | Params/M | mAP/% | FPS/s |
|---|---|---|---|---|---|---|---|---|---|---|
| YOLOv7 | 72.9 | 47.7 | 51.9 | 29.2 | 42.4 | 6.8 | 47.0 | 37.2 | 42.6 | 143 |
| **Ours** | **75.1** | **54.6** | **53.7** | **29.5** | **46.7** | **29.8** | **47.8** | **37.1** | **48.2** | **86** |

为了简单直接的验证所改进的方法能够在自动驾驶实验数据集上产生期望的效果，在表 3-3 中将改进前后的实验结果数据对比出来，可以看出，改进后的 YOLOv7 模型在各类目标上的效果整体精度上全都有所提升，在对车辆类别 Car、Bus、Truck 检测结果对比提升效果明显，精度分别提升了 2.2%、6.9%、4.3%。表明改进模型对于车辆检测具有显著优势。同时对比 Person、Motor 类别可以看出，改进后的模型在小目标检测上更具有优势，在提高检测精度的同时，模型大小也降低，检测帧率达到 86 帧/s，能够满足自动驾驶在复杂的道路环境下的检测。


(a)原图　　(b)YOLOv7 算法　　(c)YOLOv7-PGC 算法

图 3-11 可视化结果对比图

为了更直观地感受改进效果，分别用改进前后的模型在部分验证集进行测试，测试结果如图 3-11 所示。由图 3-11 可知，在夜间和多目标复杂环境下，改进后的网络模型不仅整体精度有所提高，且针对漏检情况有所改善，能够准确对遮挡以及小目标进行检测，更能全面的满足不同环境下自动驾驶对多种目标的精确检测。

（2）注意力机制对比

本章在实验阶段为了提高对目标的检测精确度，通过在相同位置进行添加多种不同注意力机制模块的对比过程，在此过程中，对添加的注意力模块进行了多次的对比实验，最终选择添加 GE 注意力机制，同时验证了添加 GE 注意力模块的优越性和有效性。

如下表 3-4 所示，其中 1 表示 YOLOv7 改进 PAFPN 结构的多尺度特征网络，ACmix 注意力机制[62]是最近较为新颖的自注意力机制，在许多环境下都有不错的效果，CBAM 与 SE 注意力机制都是比较经典的注意力机制。由表 3-4 的实验结果可以看出本文添加的 GE 模块检测后 mAP 达到 47.2%，与其他三种模块 mAP 比较分别提高了 3.4%、1.7%、1.3%，同时 GE 注意力模块也是一种比较轻量化的注意力模块，模型大小相比 CBAM 和 ACmix 模块，分别减少了 0.2M、3.1M，在添加后比原 YOLOv7 算法仅添加了 0.5M，可以看出 GE 注意力模块在提高精确性的同时减少了模型大小与运算量，检测速度提升。

表 3-4 注意力机制模块对比

| 注意力模块 | Params/M | mAP/% |
|---|---|---|
| 1 | 37.7 | 43.8 |
| 1+CBAM | 38.0 | 45.5 |
| 1+ACmix | 40.9 | 45.9 |
| 1+GE | 37.8 | 47.2 |

（3）对比实验

为了更好地验证改进后的模型效果以及选取 YOLOv7 作为改进原型算法的原因。本文选择 YOLO 系列多个检测网络以及常见的检测算法进行比较，分别对 Faster R-CNN、EfficientDet-D2、YOLOv5s 和 YOLOv7 算法在 BDD100K 数据集上进行实验，并记录每个模型的帧率和准确度，以便与改进后的算法进行比较。结果如下表 3-5 所示。

表 3-5 不同算法检测结果对比

| 算法 | FPS/s | mAP/% |
|---|---|---|
| Faster R-CNN | 12 | 39.2 |
| EfficientDet-D2 | 23 | 40.1 |
| YOLOv3 | 20 | 32.9 |
| YOLOv5s | 61 | 39.4 |
| YOLOv7 | 143 | 42.6 |
| Ours | 86 | 48.2 |

表 3-5 展示了不同目标检测算法的性能对比。在 mAP 和 FPS 等关键指标上，YOLOv7 作为 YOLO 系列性能最为出色的算法之一，显著超越了 YOLOv3、YOLOv5s、EfficientDet-D2 以及 Faster R-CNN 等算法。值得一提的是，经过改进后的 YOLOv7 算法相较于原版的 YOLOv7，mAP 提升了 5.6%。实验数据充分证明了本文研究的改进算法在检测精度上的卓越表现，明显优于原算法。这一改进不仅验证了算法的有效性，还为实际应用提供了更高的准确性和效率，并且检测速度满足检测要求，具有适用性。

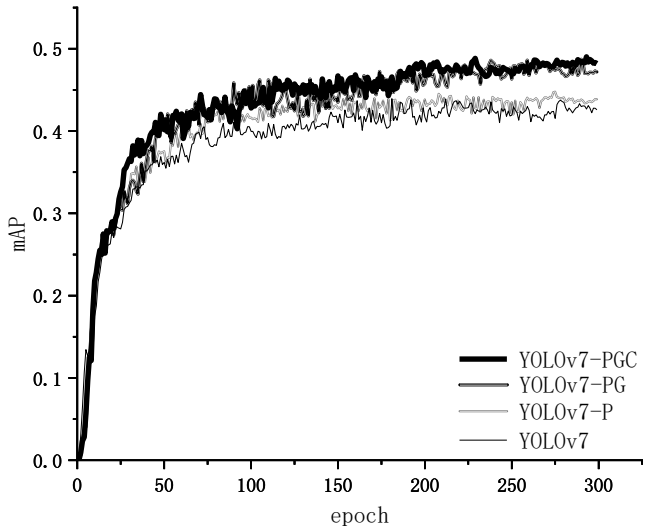



图 3-12 迭代曲线对比

另外，为了方便对比，实验将本文改进后的算法与原算法在平均精确度上进行可视化对比，如图 3-12 所示。从图 3-12 可以看出，相较于原 YOLOv7 算法，本文算法在检测精确度上均高于原算法。在损失差异上，从图 3-13 中可以看出，改进后的算法收敛速度更快，并且曲线快速趋于平稳。

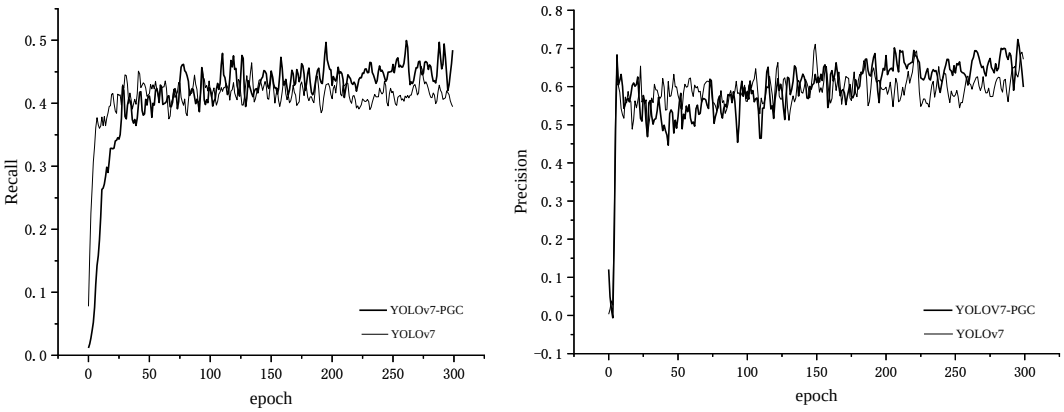



图 3-13 改进前后指数对比图

图 3-13 是改进前后的召回率与精确率的对比，改进前的算法召回率为 0.386，改进后的召回率达到 0.478，同时 P 值在改进后提升了 1.13%，由此看出改进后的算法能够更准确地识别正负实际样本，从而提高了召回率与精确率。对比改进前后的准确率，可以观察观察改进后的算法模型能够更好地提取数据特征，减少主要特征信息的损失，提高模型检测准确度。

## 3.4 本章小结

为了解决自动驾驶道路环境下推理速度慢、目标检测精度低的问题。本章提出了一种改进的多尺度道路目标检测算法，本章设计了一种多尺度小目标检测结构，允许网络对大、中、小和小目标进行多尺度检测。本章首先对YOLOv7 算法网络结构进行较为详细的介绍，分析了选取 YOLOv7 算法的原因。其次，本章引出对 YOLOv7 算法改进的介绍。本章在骨干和颈部部分引入了 GE 注意力机制，以增强网络聚合特征的能力。本章将协调坐标卷积应用到网络模型中，可以获得准确的上下文特征信息。在BDD100K数据集上进行消融实验。实验结果表明，改进后的 YOLOv7 算法在测试集上的平均检测准确率为48.2%，提高了 5.6%。改进后模型在测试场景中达到 86 帧/s，满足自动驾驶道路环境检测的准确性和实时性要求。证明了本章算法在速度和精度上的平衡性，更适合应用于自动驾驶环境感知任务，体现算法的有效性和可行性。为下一章的目标跟踪改进模型，提供了一种优化的目标检测算法作为其检测部分。

# 第 4 章 基于改进 BoT-SORT 的多目标跟踪算法

目标跟踪在自动驾驶以及智慧交通中拥有重要的地位，然而车辆目标跟踪仍然面临着许多挑战，比如复杂多变的交通环境中跟踪准确率低，车辆间会出现遮挡和重叠等问题。本章将围绕车辆目标跟踪技术展开详细的探讨，介绍了一种基于 Bot-SORT 多目标跟踪算法基础上的改进 BoT-SORT 多目标跟踪算法。

## 4.1 BoT-SORT 多目标跟踪算法

### 4.1.1 卡尔曼滤波器

BoT-SORT 多目标跟踪算法主要由优化后的 KF、相机运动模型、IoU-ReID 三个核心优化结构组成，其中 KF 根据系统的数学模型和测量数据，计算出当前时刻系统的最优状态即动态线性系统的隐藏状态。这种滤波器基于线性动态系统的状态空间模型，主要通过预测和更新两个核心步骤，不断迭代地优化对多个目标的预测。

动态过程分为两个步骤，第一步使用自身的动力学模型估计状态变量（预测阶段）；第二步使用来自可观察变量的信息改进第一次估计（更新阶段）[63]，如下图 4-1 为卡尔曼滤波结构图。

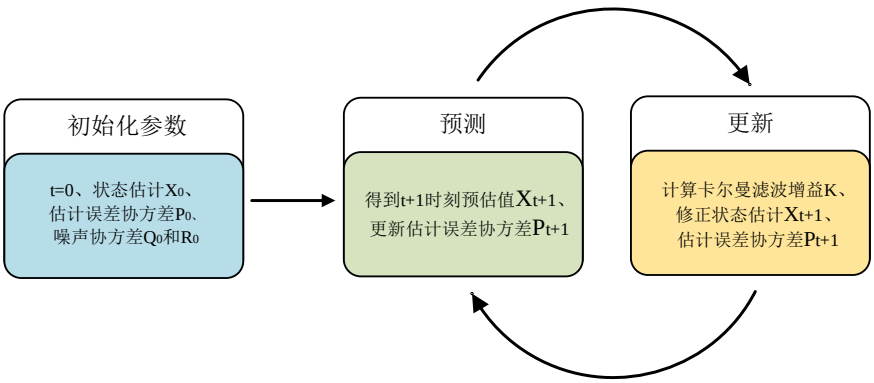



图 4-1 为卡尔曼滤波结构图

基于算法的递归模型，状态变量的估计使用了截至该时刻的所有可用信息，而不仅仅是估计前的信息。给出线性过程公式表示为：

$$x_t = Ax_{t-1} + Bu_{t-1} + w_t \tag{4-1}$$

测量方程为：

$$z_t = Hx_t + v_t \tag{4-2}$$

其中，$w_t, v_t$为独立高斯分布的白噪声，分别为$p(w)\sim N(0,Q)$，$p(v)\sim N(0,R)$即均值为 0，协方差矩阵为 $Q$ 和 $R$。$x_t$表示 $t$ 时刻状态变量，位置、速度等还处于待估计状态，$x_{t-1}$则表示 t-1 时刻的变量，$u_{t-1}$表示 t-1 时刻的控制输入量，$z_t$表示 t 时刻的环境测量值，$A$ 表示状态转移矩阵，$B$ 表示控制矩阵，$H$表示观测转移矩阵。因为 t 时刻的噪声是随机的，这就导致无法精确计算出 t 时刻的状态量即$x_t$。此时因为估计误差较大，还需要进行校正。从测量过程角度出发，校正过程及其公式如下：

首先，因为噪声随机，先将其忽略，并且假设状态变量为$\hat{x}_t^-$，环境测量值设为$\acute{z}_t$则得到：

$$\hat{x}_t^- = A\hat{x}_{t-1} + Bu_{t-1} \tag{4-3}$$

$$\hat{z}_t = Hx_t^- \tag{4-4}$$

$$\hat{x}_t = \hat{x}_t^- + K_t(z_t - Hx_t^-) \tag{4-5}$$

忽略噪声影响后的环境测量值$\acute{z}_t$维度相同，其中，误差由$z_t$-$\hat{z}_t$得到，$K_t$为 KF 增益，而$\hat{x}_t$则表示 t 时刻的最优估计。实际的卡尔曼滤波过程公式如下：

（1）预测阶段

在预测阶段，初始化开始的状态估计，同过卡尔曼滤波将 $t$-1 时刻的状态变量作为 $t$ 时刻的预测估计。

$$\hat{x}_t^- = A_t\hat{x}_{t-1} + B_tu_{t-1} \tag{4-6}$$

$$P_t^- = A_tP_{t-1}A_t^T + Q_t \tag{4-7}$$

其中，$\hat{x}_t^-$为预测估计是根据最后的估计值获得的，即通过上一时刻的最优估计值得到的，$A_t$和 $A$ 类似，在没有控制信号前提下，反应 t 与 t-1 时刻的状态关系，$B_t$则是将控制信号与 $t$ 时刻关联。$P_t^-$是时刻 $t$ 的状态估计的协方差矩阵，描述了状态估计的不确定性，$P_{t-1}$是 t-1 时刻状态估计的协方差矩阵，$A_t^T$是 t 时刻状态转移矩阵的转置矩阵，$Q_t$表示的是过程噪声协方差矩阵。

（2）更新阶段

使用在预测阶段获得的估计值，对系统进行校正并对系统当前上个目标信

息更新。

$$K_t = P_t^- H_t^T (H_t P_t H_t^T + R_t)^{-1} \tag{4-8}$$

$$\hat{x}_t = \hat{x}_t^- + K_t(z_t - H x_t^-) \tag{4-9}$$

$$P_t = (I - K_t H_t) P_t^- \tag{4-11}$$

$K_t$表示卡尔曼增益的$n \times m$矩阵，卡尔曼增益表示对观察到的特征的置信度。$P_t^-$表示 $t$ 时刻状态的先验估计的协方差矩阵，$P_t$后验估计的协方差矩阵。总而言之，KF 通过将测量和预测相结合以找到最佳值，然后不断递归下去。

在基于检测的目标跟踪算法中，KF 是最常用的滤波方法之一。通过 KF 的递归模型，可以跟踪模型预测目标下一帧或下一时刻的位置参数，同时下一帧时通过检测到的目标结果与先验估计匹配，并对目标状态进行更新。而 BoT-SORT 模型相比于 DeepSORT，状态向量中直接输出宽高，如下公式所示：

$$DeepSORT: x = \left[x_c, y_c, a, h, \acute{x_c}, \acute{y_c}, \acute{a}, \acute{h}\right]^T \tag{4-12}$$

$$BoTSORT: x = \left[x_c, y_c, w, h, \acute{x_c}, \acute{y_c}, \acute{w}, \acute{h}\right]^T \tag{4-13}$$

将宽高比改成了预测宽度 $w$，这样在预测过程中预测框可以更好的匹配目标轮廓，减少了目标特征信息的丢失，同时在接下来 IoU 阈值匹配也更加优越。除此之外，还修改初始化的矩阵参数过程噪声 $Q$、测量噪声 $R$，从而提高跟踪的精确度。

### 4.1.2 相机运动模型 CMC

在检测式的目标跟踪模型算法中，跟踪效果很大一部分上取决于预测框和检测框之间的重叠程度。但实际场景中，相机极大可能会因运动发生偏移、颠簸，这时如果再继续按照原来方法进行预测，预测框与真实的检测框就会有误差，如下图4-2所示，运动相机发生变化，相当于图像二维平面上坐标系发生变换第t+1帧图中的预测框位置与真实检测框会有一定误差，而修正后的检测框则将误差减少甚至消除。误差会造成匹配过程中边界框的偏移，从而影响 ID 连续性，导致 ID 跳变，而匀线性加速模型的卡尔曼滤波器，对于实际环境中非线性的运动模型效果很差。如下图 4-2 所示:

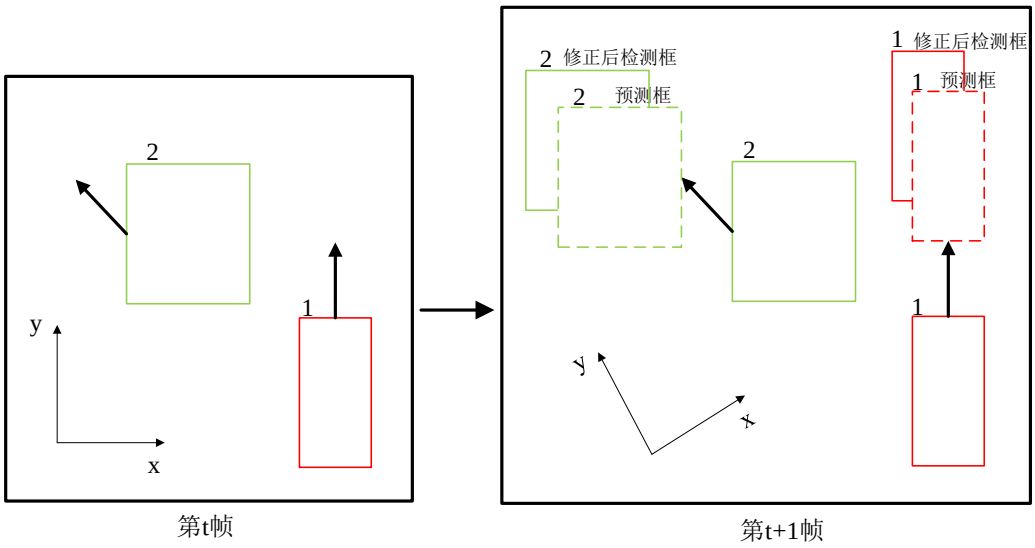



图 4-2 边界框偏移示意图

针对上述问题，BoT-SORT 跟踪模型根据全局运动补偿（GMC）[64]技术在 OpenCV 库上实现的视频序列稳定模块与仿射变换方法，首先，使用提取图像关键点技术对背景提取特征描述子，然后，将稀疏光流运用于特征跟踪，最后，使用 RANSAC 计算运动的仿射矩阵。利用稀疏配准获得的前后帧运动信息，计算得到仿射矩阵，从而在目标检测基础上对动态目标进行忽略，以更准确的预测边界框，相机运动补偿流程如下公式所示：

$$A_{t-1}^{t} = [M_{2\times2}|T_{2\times1}] = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \tag{4-14}$$

$$\widetilde{M}_{t-1}^{t} = \begin{bmatrix} M & 0 & 0 & 0 \\ 0 & M & 0 & 0 \\ 0 & 0 & M & 0 \\ 0 & 0 & 0 & M \end{bmatrix}, \quad \tilde{T}_{t-1}^{t} = \begin{bmatrix} a_{13} \\ a_{23} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{4-15}$$

$$\hat{x}_{t|t-1}^{\cdot} = \widetilde{M}_{t-1}^{t}\hat{x}_{t|t-1} + \tilde{T}_{t-1}^{t} \tag{4-16}$$

$$P_{t|t-1}^{\cdot} = \widetilde{M}_{t-1}^{t}P_{t|t-1}\widetilde{M}_{t-1}^{t^T} \tag{4-17}$$

因此，利用 RANSAC 算法得到平面坐标变化的仿射矩阵 A，描述二维平面旋转和缩放变化矩阵为$M \in \mathbb{R}^{2\times2}$，平移变化向量矩阵为$T \in \mathbb{R}^{2}$，$\hat{x}_{t|t-1}^{\cdot}$、$\hat{x}_{t|t-1}$分别表示相机运动补偿前后 K 时刻卡尔曼滤波的预测估计。$P_{t|t-1}^{\cdot}$、$P_{t|t-1}$分别是修正前后 KF 的预测协方差矩阵。将经过相机运动模型补偿后的$\hat{x}_{t|t-1}^{\cdot}$和$P_{t|t-1}^{\cdot}$

带入卡尔曼滤波更新阶段中则可以得到修正后的预测估计，可以定义为：

$$K_t = P_{t|t-1}^{\cdot} H_t^T \left(H_t P_{t|t-1}^{\cdot} H_t^T + R_t\right)^{-1} \tag{4-18}$$

$$\hat{x}_{t|t} = \hat{x}_{t|t-1}^{\cdot} + K_t\left(z_t - H_t \hat{x}_{t|t-1}^{\cdot}\right) \tag{4-19}$$

$$P_{t|t} = (I - K_t H_t) P_{t|t-1}^{\cdot} \tag{4-20}$$

在高速运动下，对速度在内的状态向量的矫正是至关重要的，同时再通过对变化缓慢下的状态矢量的校正，可以使模型算法对于相机运动下的跟踪鲁棒性更强。

### 4.1.3 IoU-ReID 融合

在匹配过程中，模型将深度学习特征提取融合到跟踪过程，以ResNest50作为主干网络，以 SBS 作为训练 ReID[65]的基线网络，损失函数使用了经典的TripletLoss，并且采用指数移动平均(EMA)来更新第 t 帧中第 k 个预测框的匹配外观轨迹状态，如下面公式：

$$e_k^t = \alpha e_k^{t-1} + (1-\alpha) f_k^t \tag{4-21}$$

$$C = \lambda A_a + (1-\lambda) A_m \tag{4-22}$$

其中$f_k^t$表示当前t帧的特征输入，由于复杂环境可能会造成的遮挡、模糊等问题，所以提高对高置信度目标的检测，从而能过滤部分复杂环境造成的影响，运用余弦距离进行度量平均轨迹特征状态$e_k^t$与新的$f_k^t$的匹配，C 为代价矩阵，$\lambda$为权重系数，通常设为 0.98。

提出 IoU 加 ReID 融合机制的方法来进行特征匹配，即对于余弦距离较小或较远的目标通过 IoU 阈值进行删除。IoU-ReID 融合方法可以定义为下面公式：

$$\hat{d}_{i,j}^{cos} = \begin{Bmatrix} 0.5 \cdot d_{i,j}^{cos}, (d_{i,j}^{cos} < \theta_{emb}) \wedge (d_{i,j}^{iou} < \theta_{iou}) \\ 1, otherwise \end{Bmatrix} \tag{4-23}$$

$$C_{i,j} = min\left\{d_{i,j}^{iou},\ \hat{d}_{i,j}^{cos}\right\} \tag{4-24}$$

根据实际情况设置$\theta_{iou}$，$\theta_{emb}$阈值，分别用来过滤不符合的轨迹和轨迹特征向量与检测输入向量的匹配结果。

### 4.1.4 BoT-SORT 算法流程

传统的基于检测的多目标跟算法如 DeepSORT、ByteTrack 等模型，对于背景环境简单下的较少数量的目标跟踪效果较为理想，由于 KF 的线性运动模型，以及实际环境中相机的抖动，在跟踪过程中造成轨迹逐渐偏离，重叠、遮挡情况下的输出重复等问题。BoT-SORT 算法得益于一系列改进，在跟踪匹配中跟踪连续性和鲁棒性都更好。为了提高对检测轨迹的特征匹配，BoT-SORT 采用更高效的级联匹配，在减少计算量的同时，提高了筛选匹配广度，有效的提高了跟踪的实时性。相比于其他 SORT 算法的跟踪模型，因为其工作流程中的分离模块化设计，可以更加方便的对目标检测网络和外观特征提取器进行测试对比。主要流程如下：

（1）输入视频序列，以 YOLOX 算法为目标检测模型，对视频的每一帧图像进行目标检测，获取当前帧的检测框及其位置、置信度等参数。

（2）根据检测框置信度对目标进行判断，大于阈值则使用 SBS-ResNest50 重识别网络对目标进行外观特征提取，小于阈值进入第二次关联匹配。

（3）通过利用上一帧轨迹结果，KF 对当前帧目标轨迹状态进行预测，得到当前帧预测框的位置及目标状态向量，相机运动补偿减少误差。

（4）将第（2）步的结果与第（3）步结果进行第一次关联匹配，通过匈牙利算法，根据检测框的 IoU 信息先进行初步匹配，然后根据提取的外观特征补充匹配。

（5）根据第一次匹配得到两种结果，即成功匹配和未成功匹配，其中成功匹配的目标在经过 KF 更新和外观特征更新后，进入轨迹管理中，为成功匹配的目标，与第（2）步中置信度低的检测框进行第二次关联匹配。

（6）第二次关联匹配通过匈牙利算法，利用 IoU 交互比，对轨迹进行进一步匹配，当成功匹配时，通过 KF 更新和外观特征更新后将轨迹归入轨迹集合汇总，未成功匹配时，将其输入到轨迹集合中。

（7）在轨迹集合中，从多级联匹配中得到的轨迹进行管理，成功匹配的轨迹输出为此帧的轨迹结果，未成功匹配的轨迹根据连续帧的丢失次数进行分配，如果丢失帧数小于设定的阈值，则将其初始化为新轨迹，重新进行算法流程，大于阈值，则将轨迹删除。

如下图 4-3 为 BoT-SORT 算法的流程图。

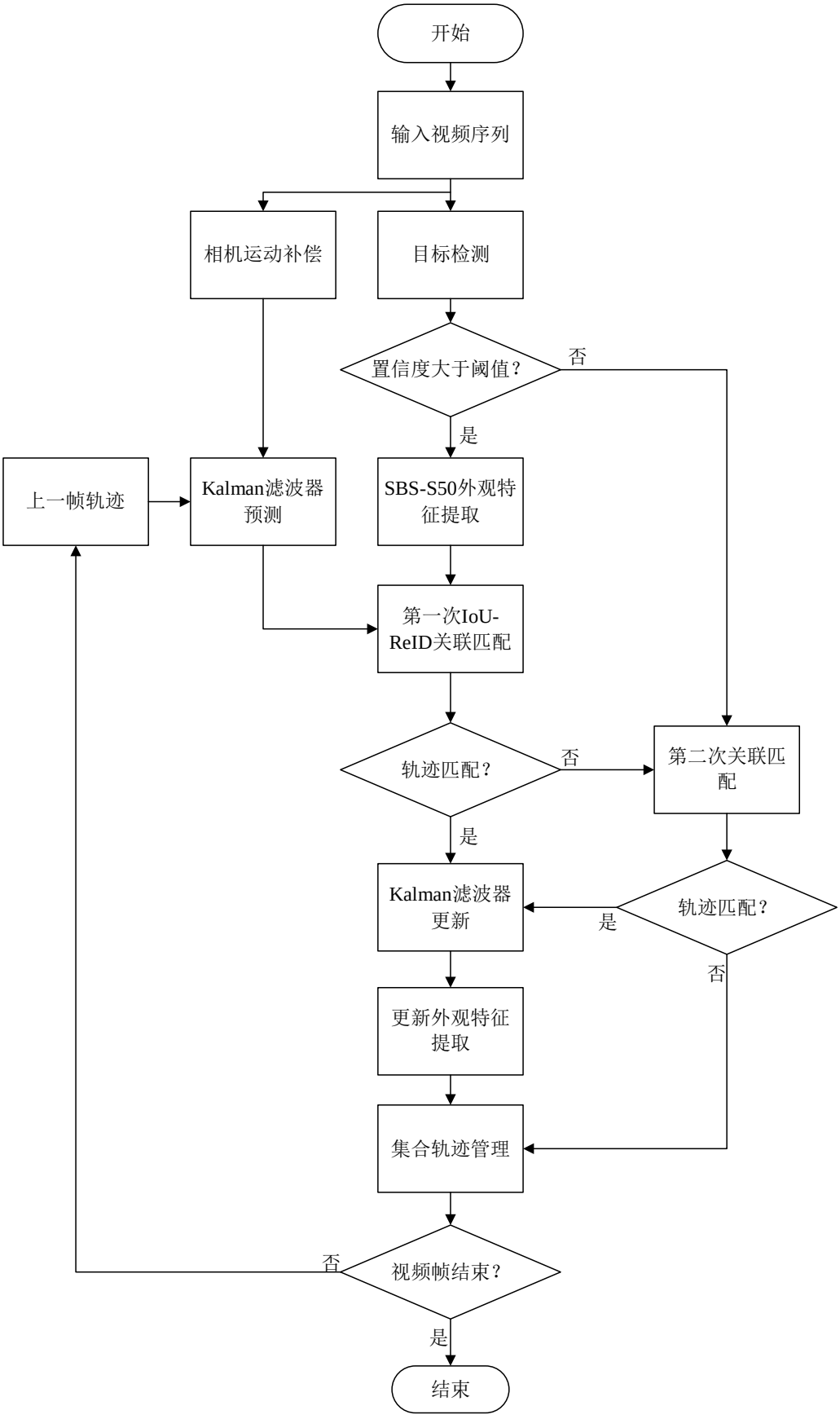



图 4-3 BoT-SORT 算法流程图

## 4.2 BoT-SORT 多目标算法优化

### 4.2.1 改进 BoT-SORT 算法框架

传统 BoT-SORT 在多目标的道路场景下对运动状态多变的多个车辆进行目标轨迹跟踪任务艰巨，基此进行优化 BoT-SORT 跟踪算法。如下图 4-4 所示。

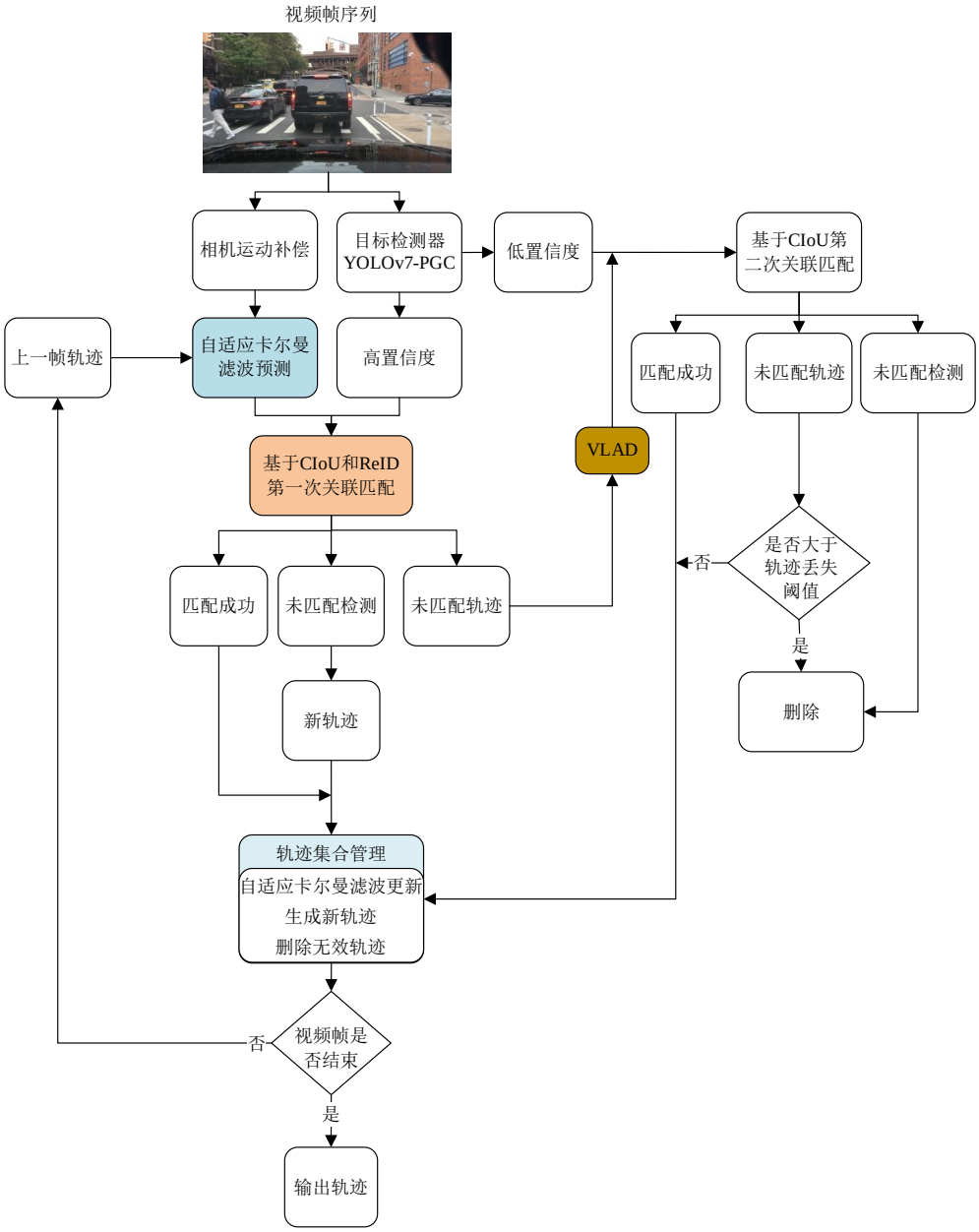



图 4-4 改进 BoT-SORT 算法框架图

为了减小物体运动状态突变对于跟踪效果的影响，使用一种在动态系统下适应性更强的 AKF，提高轨迹连续帧的稳定性。引用 CIoU 匹配算法替换原框

架中的 IoU 匹配算法，改善身份跳变状况。在级联匹配间添加 SURF 描述符的 VLAD 图像特征匹配算法，加强对特征提取，提高模型的实时性和高效性。

表 4-1 改进 BoT-SORT 算法流程

| |
|---|
| Input：视频帧 V，目标检测器 Det，高分置信度阈值 m，新轨迹置信度阈值$\eta$ |
| Output：目标轨迹 T |
| 1：初始化：T←0 |
| 2： for frame $f_t$ in V do |
| 3: $D_t$ ←Det($f_t$) ;$D_{high}$←0;$D_{low}$←0 |
| 4: for d in $D_t$ do |
| 5: if d.score > $m$ then |
| 6: $D_{high} \leftarrow D_{high} \cup \{d\}$ |
| 7: else |
| 8: $D_{low} \leftarrow D_{low} \cup \{d\}$ |
| 9: end if |
| 10: end |
| 11: $A_{t-1}^{t}$←计算前后帧仿射矩阵 |
| 12: for t in T do |
| 13: T←AKF（t） |
| 14: end for |
| 15: 第一次关联匹配$A_t$←$CIoU\&ReID\ Dist$($D_{high}$,T) |
| 16: $D_{remain}$← ($D_{high}$, $D_{low}$)， $T_{remain}$←T |
| 17: 外观匹配←VLAD（SURF） |
| 18： 更新$D_{remain}$， $T_{remain}$ |
| 19: 第二次关联匹配$T_{re-remain}$ |
| 20: $A_t$←CIoU Dist（$D_{remain}$,$T_{remain}$） |
| 21: 更新 AKF 轨迹匹配 |
| 22: 更新外观特征 |
| 23: 删除未匹配轨迹 T←$T/T_{re-remain}$ |
| 24: for d in $D_{remain}$ do |
| 25: if d(score) > $\eta$ then |
| 26: T←T ∪ {$d$} |
| 27: end if |
| 28: end |
| 29: end |

改进后的算法流程如上表 4-1 所示，算法的主要步骤分为四个步骤：

（1）根据相机的运动状态和前一帧的检测结果，使用卡尔曼滤波预测当前帧每条轨迹的新位置。

（2）其次，使用 Hungarian 算法将当前帧中的高分（高置信度）框与 IoU 和外观特征的预测位置进行匹配，其中外观特征匹配基于使用 VLAD 方法的 SURF 描述符。

（3）使用 IoU 和外观特征作为相似性标准，将低分框与不匹配的轨迹相关联。

（4）更新匹配轨迹的状态，为新出现的目标生成新的轨迹，删除一段时间未出现的轨迹，更新卡尔曼滤波。

### 4.2.2 自适应卡尔曼滤波

虽然在 BoT-SORT 算法中，作者利用了相机运动补偿的方法以尽可能的消除物体运动状态下发生的颠簸、遮挡对于跟踪的影响。结果也表明这中方法可以在一定条件下对动态背景如车辆跟踪产生不错效果，但是，现如今基于检测的跟踪算法普遍采用 KF 作为预测更新模型，而 KF 是一种线性的匀速模型算法，虽然可以在运动状态下准确预测目标的下一个轨迹，但它也有局限性，原始 KF 输入状态矩阵以恒定速度建模，导致平台运动期间跟踪帧的不匹配，而在车辆目标跟踪任务中，道路环境中的车辆目标速度非常复杂，这就容易导致在递归模型下检测位置与预测位置误差越来越大，针对这一问题，本文对 KF 进行优化，采用了一种自适应卡尔曼滤波算法对目标进行预测和更新，因为其在动态系统适应上的优异表现，AKF 相比于 KF 在车辆目标跟踪上拥有更好的稳定性和鲁棒性。

原始跟踪框是通过计算目标的纵横比得到的，从 4.1.1 小节中可以知道，预测框的准确性会直接影响跟踪和 ID 的精确度。位置紧密的目标边界框可能会重叠并相互干扰，从而导致跟踪失败和严重的 ID 跳变问题，错误识别的可能性增加，导致其他物体被误识别，针对上面问题在式(4-25)和式(4-26)中重新定义了 KF 的状态向量，以直接估计轨迹的宽度和高度。

$$x_k = [x_c, y_c, \omega, h, \dot{x}_c, \dot{y}_c, \omega, \mathrm{h}]^T \tag{4-25}$$

$$z_k = \left[z_{xc}, z_{yc}, z_\omega, z_h\right]^T \tag{4-26}$$

其中$x_c(t), y_c(t)$检测框的中心坐标;$\omega(t), h(t)$检测框的宽度和高度;$\dot{x}_c, \dot{y}_c, \omega, \mathrm{h}$是相应参数的速度变化值，$z_k$是测量的中心坐标及宽度和高度。调整测量噪声$R(t)$和过程噪声$Q(t)$以适应特定场景。然而，当相机移动或光线或环境发生变化时，原始$R(t)$和$Q(t)$参数无法适应实时场景。这会导致跟踪丢失和准确性下降。因此，通过微调 sum 的权重来动态调整关联模块的增益，具体方式如下。第一步是预测过程，针对状态向量$x$和伴随的协方差$p$，如式(4-27)和(4-28)所示：

$$\hat{X}_{(t|t-1)} = \Phi_{(t,t-1)}\hat{X}_{(t-1)} \tag{4-27}$$

$$P_{(\mathrm{t}|\mathrm{t}-1)} = \Phi_{(t,t-1)}P_{(t-1)}\widehat{\Phi}^{T}_{(t,t-1)} + \Gamma_{(t,t-1)}Q_{(t)}\Gamma^{T}_{(t,t-1)} \tag{4-28}$$

这里$P_{(\mathrm{t}|\mathrm{t}-1)}$和$P_{(t-1)}$表示预测误差协方差和误差协方差；$\Phi_{(t,t-1)}$表示状态传递矩阵，$\Gamma_{(t,t-1)}$表示噪声矩阵，同时还引入加权系数$d_t$，以便于$R_{(t)}$可以根据$d_t$与$R_{(t)}$和$R_{(t-1)}$关于帧数 t 的变化，用于表示$R_{(t-1)}$的稳定性，如下式所示：

$$d_t = \frac{1-b}{1-b^{t+1}}, 0 < b < 1 \tag{4-29}$$

$$v_{(t)} = Z_{(t)} - H_{(t)}\hat{X}_{(t|t-1)} \tag{4-30}$$

$$\mathrm{Y}_t = \left[I - H_{(t)}K_{(t-1)}\right]v_{(t)}V^{T}_{(t)}\left(I - H_{(t)}K_{(t-1)}\right) \tag{4-31}$$

$$R_{(t)} = (1-d_t)R_{(t-1)} + d_t\left[(\mathrm{Y}^{T}_{t}) + H_{(t)}P_{(t-1)}H^{T}_{(t)}\right] \tag{4-32}$$

其中$b$是遗忘因子，取值范围为 0.95 到 0.99，过小会导致噪声影响占比增大;$H_{(t)}$是观测矩阵；和$v_{(t)}$是系统观测向量；

$$S_{(t)} = H_{(t)}P_{(t,t-1)}H^{T}_{(t)} + R_{(t)} \tag{4-33}$$

下一步是更新过程：

$$K_{(t-1)} = P_{(t|t-1)}H^{T}S^{-1}_{(t)} \tag{4-34}$$

$$P_{(t)} = \left[I - K_{(t)}H_{(t)}\right]P_{(t,t-1)} \tag{4-35}$$

$$\hat{X}_{(t)} = \hat{X}_{(t|t-1)} + K_{(t)}v_{(\mathrm{t})} \tag{4-36}$$

更新过程与 KF 类似，主要区别在于观测误差 R 基于$d_t$调节，而随着帧数 t 的变化，$R_{(t)}$越来越趋于$R_{(t-1)}$，即观测误差 R 趋向稳定。

### 4.2.3 改进损失函数

在 BoT-SORT 算法的多级联匹配阶段，都有使用到 IoU 作为一个参量来计算距离值，把匹配失败的检测框与未确认状态的预测框轨迹通过 IoU 算法来判断两者的匹配程度。IoU 又称交互比，其计算公式如下：

$$\mathrm{IoU}=\frac{|A\cap B|}{|A\cup B|} \tag{4-37}$$

所以，IoU 指标可以很好的反映预测框与检测框之间的匹配情况。但是，在某些特殊情况下会使得 IoU=0，这样就无法反映检测效果。如下图 4-5 所示：

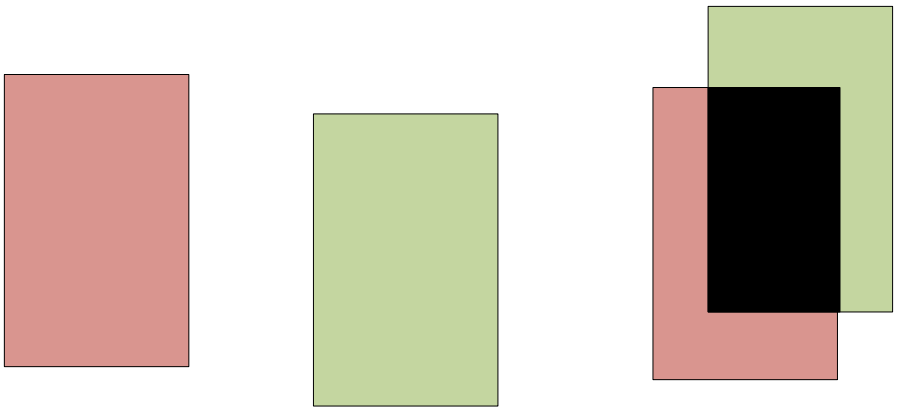



图 4-5 交互比情况对比图

在对车辆目标轨迹跟踪的过程中，由于视野范围小，道路环境复杂，时常会发生车辆突然出现、多个目标相互重叠而导致的身份跳变。为了更好地适应车辆轨迹跟踪的复杂环境，本文引用完全交并比(Complete IoU，CIoU)[66]匹配算法作为原框架中的 IoU 匹配算法。相比于 DIoU，CIoU 是在 DIoU 的优化基础上再次进行改进的，由于 CIoU 考虑了更多的影响因素，如图像相似性和长宽比等，使得 BoT-SORT 模型能够多尺度地考虑到车辆目标的差异性，从而能够更好地提取到车辆目标的特征。CIoU 的这些优点可以有效提高 BoT-SORT 模型应对复杂道路环境的能力。CIoU 的计算公式如下：

$$CIoU=\frac{\rho^2(b,b^{gt})}{c^2}+\alpha v \tag{4-38}$$

其中，$b$表示第 t 帧的轨迹预测框，$b^{gt}$表示第 t 帧的轨迹检测框，$\rho$表示这两个框中心点的欧几里德距离，$c$是同时包含轨迹测量框和预测框的最小矩阵的

斜率距离，值越高，测量框和预测框之间的距离越远，CIoU 在 DIoU 的基础上加了一个影响系数$\alpha v$，$\alpha$是权重函数，而$v$用来度量匹配过程中，轨迹预测框与测量框之间长宽比的相似性，可以定义为：

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{4-39}$$

$w$，$h$分别表示轨迹预测框的宽和高，$w^{gt}$，$h^{gt}$则分别表示检测框的宽和高。如下图 4-6 所示。

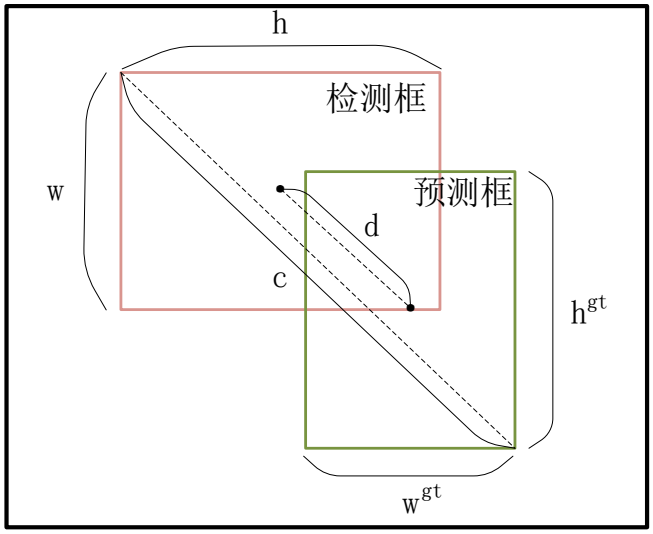



图 4-6 CIoU 示意图

从 CIoU 的计算公式可以看出，当检测框和预测框的 IoU 值为 0 时，并不会直接将其删除，而是通过计算得到$c$的值。这一步骤改善了 IoU 无法匹配未相交的目标框的问题。完整的 CIoU 损失函数的定义为：

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b,b^{gt})}{c^2} + \alpha v \tag{4-40}$$

将 CIoU 的损失函数运用到模型匹配过程中，使得预测框与检测框的相似度更高，为特征匹配提供了多方面的特征信息，在跟踪中对目标轨迹 ID 持续跟踪的效果显著。

### 4.2.4 添加 VLAD 模块

由于车辆轨迹跟踪中，外观特征提取网络需要对车辆的特征信息进行特征提取，而在复杂情况下，首先相机会被环境因素等影响，其次，会出现多个车辆目标外观特征信息差异性较少，导致在特征匹配中无法根据特征信息对轨迹

进行连续性跟踪。本文使用 SURF 模块来提取目标的外观特征，在保证一定实时性的前提下实现特征提取。然后，采用 VLAD 方法测量不同目标之间的相似性，以实现基于外观特征的目标再匹配。

局部聚合描述符向量[67]是一种基于积聚的图像特征匹配方法，通过将局部特征聚合为全局表示来，能有效的提取图像中的关键特征信息，相比于其他图像特征表示方法，VLAD 算法的计算处理效率更高，所以对于复杂背景下的多个车辆目标轨迹可以更好地进行特征的提取和匹配任务。其框架如下图 4-7 所示。

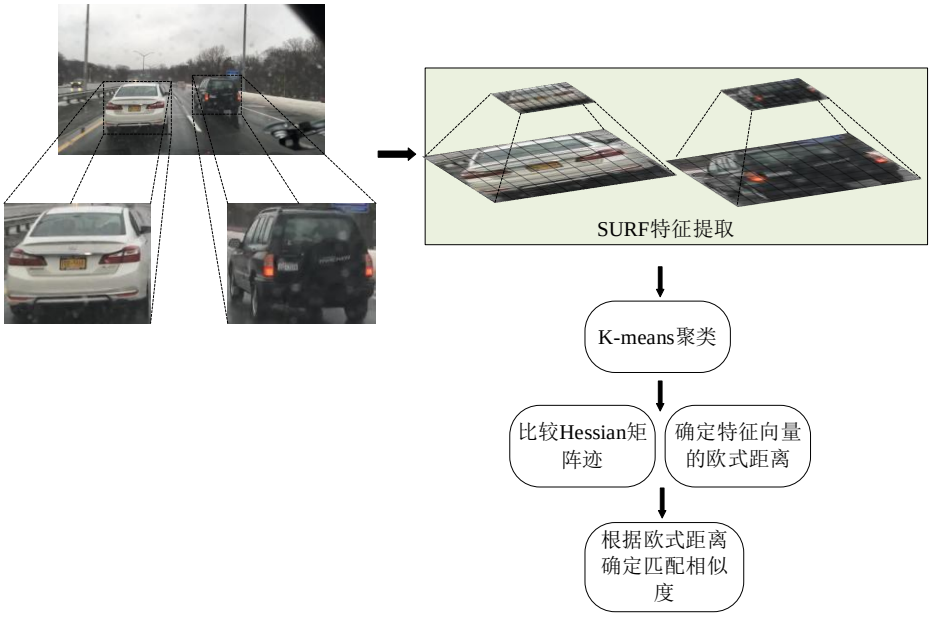



图 4-7 VLAD 特征匹配框架

VLAD 算法常用 SIFT 描述子对特征进行提取，但效率不高，不能满足车辆轨迹跟踪任务的实时性要求，本文使用提取和表述特征更高效的 SURF 描述符。VLAD 的整体流程如下：首先，对于需要比较的两张图像（车辆目标），使用 SURF 描述符来提取图像特征。然后，使用聚类算法（如 K-means 聚类）将特征描述符划分为多个聚类。接下来，对于每个提取的局部特征描述符，找到最近的聚类并将其编码为与该聚类相关的向量。随后，聚合所有编码的局部特征描述符以生成全局特征表示。最后，计算两个图像的特征表示之间的相似度值，并使用余弦相似度确定两个特征向量之间的距离，距离越小，两个图像越相似，匹配效果越显著。

这里在车辆多目标跟踪中，本文将一个目标轨迹的不同帧作为 VLAD 算法

的匹配对象，从而可以根据 SURF 描述符对于图像特征提取高效的优势，对轨迹进行精确的预测，实现轨迹连续具有鲁棒性的效果。

## 4.3 实验结果与分析

### 4.3.1 跟踪环境配置

本章在 BDD100K 跟踪数据集的基础上，结合其他视频，融合选取了十个视频作为自建数据集，同时针对实验环境，分别选取雾天、雨天、晴天等复杂道路环境作为实验测试数据集，然后，本文使用 Darklabel 标注软件对视频序列进行标注，视频帧率为 30FPS/秒，像素为 1920*1080。该数据集包含了 10 段视频，分别将其按照 MOT01、MOT02…MOT10 的顺序命名。视频时间长短不一，按照每帧图像进行划分后，总计 3214 帧图像。如下图 4-8 所示。

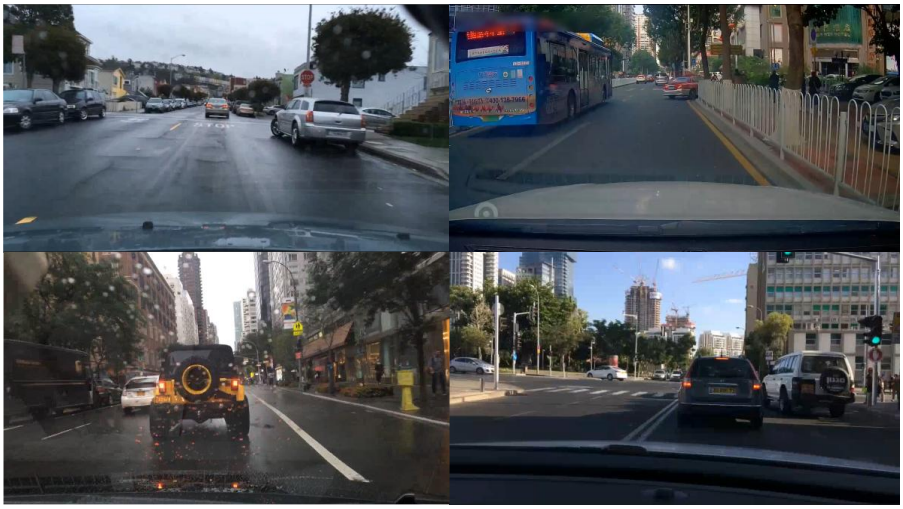


图 4-8 部分数据集示意图

标注完成后生成的跟踪数据集格式为 MOT 标签格式，其标签信息包含当前帧 frame，当前帧中目标的数量，以及目标预测框的位置信息 box(x1,y1,x2,y2)，目标类别 class 等，同时在标注时会为每个目标分配一个唯一的 id。为了方便使用，将其格式转换成跟踪训练的 MOT 跟踪数据集结构。为了验证跟踪模型的改进的有效性，在本文自建数据集进行了消融实验和对比测试，验证了本实验车辆跟踪算法的有效性。实验使用第 3 章 YOLOv7-PGC 算法作为检测器部分，重识别部分为 ResNet-50 框架作为主干网络，训练周期为 100 轮。操作系统为

Ubuntu 18.04 系统。CPU 配置类型型号为 AMD Ryzen 7 7840H。GPU 型号为 NVIDIA RTX A5000(24G)×2，采用 Adam 优化器。

### 4.3.2 评估指标

为了对跟踪实验性能进行分析，本小节对跟踪中的多个实验参考数据进行选择对比，从跟踪精准性和连续性方面的相关指标对模型跟踪性能进行评估。

其中多目标跟踪准确度(Multiple Object Tracking Accuracy, MOTA)表示正确的 *TP*（True Positive）样本占所有跟踪目标数量的比率，除了衡量模型对正确预测的 *FP*、*FN*、*IDs* 之外，还考虑了正确预测的样本数量，因此 MOTA 可以作为一个全方面评估跟踪模型准确度和稳定性的指标[68]，计算公式如下式(4-41)所示：

$$MOTA = 1 - \frac{\Sigma(FN_t + FP_t + IDS_t)}{\Sigma GT_t} \tag{4-41}$$

其中，*FN* 为视频中所有的漏检目标数，*FP* 则是误检目标数，*IDs*（ID switches）则表示在跟踪目标过程中该目标的 ID 切换的次数，*GT* 表示真实目标数，*t* 是视频帧的帧数，*GT* 是真实跟踪目标的数量，MOAT 可以表示漏检率、误检率和 ID 错误率三种类型的错误率。

通过公式(4-11)可以看出，MOTA 收到三个关键错误参数影响：错误匹配的数量、目标丢失的次数以及身份标识的混淆程度，因此能够全面评估跟踪算法的性能表现。

多目标跟踪精确度(Multiple Object Tracking Precision, MOTP)在跟踪算法中用来衡量跟踪目标位置估计的精确程度。和 MOTA 类似，数值越大表示精确度越高，但 MOTP 仅反映匹配成功轨迹的位置误差大小。计算公式如式(4-42)所示：

$$MOTP = \frac{\Sigma_{t,j} d_{t,j}}{\Sigma TP_t} \tag{4-42}$$

其中，$d_{t,j}$表示第 t 帧中第 j 个目标的检测框与预测框之间的重合率，重合率用 IoU 表示；$TP_t$表示第 t 帧中检测框和预测框匹配成功的个数。由此可以分析得到，MOTP 的范围在(0, 1)之间，其值越大表示定位的越准确。

跟踪准确率（Identification F1-Score, IDF1）作为多目标跟踪评估的重要指

标，反映了算法在保持目标身份一致性方面的性能。该指标通过计算 ID 精度和 ID 召回率的调和平均数，综合评估了目标身份识别的准确性和完整性，是作为衡量跟踪 ID 连续性的重要指标，其公式如式(4-43)所示。

$$IDF1 = \frac{2IDTP}{2IDTP+IDFP+IDFN} \tag{4-43}$$

其中 IDTP 表示正确检测到的目标数量，IDFP 表示误检的目标数量，IDFN 表示漏检的目标数量。IDF1 越接近 1 越好，数值越高代表跟踪真实目标的精确度越好。该指标侧重于跟踪算法跟踪某个目标的时间长短，对跟踪模型在复杂环境下的身份连续性性能提供了依据。

### 4.3.3 实验结果

（1）消融实验

为了提高对复杂环境下车辆的多目标跟踪模型性能，对 BoT-SORT 跟踪模型进行了多次不同的优化方法，在本节所完成的 4 组消融实验中，将原算法和改进 AKF、CIoU 和 VLAD 算法分别记为第 1、2、3、4 组。本组实验使用的检测器均为 YOLOv7-PGC 模型，保持检测器的一致性。如下表 4-2 所示。

表 4-2 消融实验结果

| 组别 | AKF | CIoU | VLAD | MOTA(%) | MOTP(%) | IDs(次) | IDF1(%) |
|---|---|---|---|---|---|---|---|
| 1 | | | | 57.7 | 77.1 | 142 | 67.3 |
| 2 | √ | | | 58.3 | 78.4 | 103 | 68.4 |
| 3 | √ | √ | | 60.5 | 77.2 | 86 | 69.7 |
| **4** | √ | √ | √ | **61.3** | **78.6** | **78** | **71.4** |

通过第 1 组 BoT-SORT 算法实验与第 2 组组别实验数据对比可以看出，通过将跟踪模型 BoT-SORT 的卡尔曼滤波改进为自适应卡尔曼滤波后，算法模型对于动态遮挡场景下的车辆检测更加稳定，BoT-SORT 原始算法的 MOTA 为 57.7%，IDs 为 142，而对卡尔曼滤波进行改进后，第 2 组模型的 MOTA 为 58.3%，IDs 为 103，相比原始算法准确度提高了 0.6%，IDs 则降低了 39 次，这说明改进 AKF 后的模型弥补了在复杂环境下的运动预测的局限性，对于减少 ID 跳变次数有着显著的效果。第 3 组是在改进 AKF 的基础上对 IoU 匹配函数进行

优化，采用参数更全面的CIoU匹配，相比第1组原模型的MOTA提高了2.8%，相比第 2 组提高了 2.2%，IDF1 提高了 1.3%，通过这一结果可以分析得到，通过采用CIoU匹配，显著提高了跟踪连续性和准确性，模型可以更加有效的匹配目标真实位置与预测框，有效减少了复杂环境的中误差导致的误匹配，第 4 组是本文所综合改进的多尺度特征融合算法，即基于 YOLOv7-PGC 和 BoT-SORT 模型，在多目标跟踪器中进行多处创新，即优化自适应卡尔曼滤波、优化 CIoU 匹配以及添加 VLAD 特征匹配方法，最后，模型的 MOTA 值为 61.3%，与第 1 组结果相比，提高了 3.6%。MOTP 值为 78.6%，提升了 1.5%。IDs 从 142 减少至 78，降低了 64。

综上所述，通过在第一次匹配与第二次匹配之间添加 VLAD 算法模块，提高了跟踪的准确度，同时通过局部特征聚合，使跟踪中相似目标的特征差异更显著，IDF1 显著提高，反映出改进后模型的误匹配率降低，通过实验验证了改进后的 BoT-SORT 算法模型弥补了复杂环境下遮挡和运动的局限性，有效的解决 ID 跳变次数多的问题，加强了轨迹跟踪的连续性。结果表明，在车辆跟踪的场景中，改进后的模型显著的提高了跟踪的准确性和稳定性，验证了改进方法的有效性。

（2）对比实验

为了验证基于 BoT-SORT 改进的多目标跟踪算法框架具备改进有效性，本文基于自制车辆数据集对模型进行相关的消融实验和对比实验，实验数据如下表 4-3 和表 4-4 所示。

表 4-3 不同检测器对比实验

| 检测器 | 跟踪器 | MOTA/% | MOTP/% | IDs/次 | IDF1/% |
|---|---|---|---|---|---|
| YOLOX | BoT-SORT | 56.2 | 76.3 | 155 | 61.6 |
| YOLOv7-PGC | | 57.7 | 77.1 | 142 | 67.3 |
| YOLOX | 改进 BoT-SORT | 58.1 | 76.5 | 97 | 68.2 |
| YOLOv7-PGC | | 61.3 | 78.6 | 78 | 71.4 |

表 4-3 为 YOLOX 和 YOLOv7-PGC 与跟踪模型 BoT-SORT 以及本文改进后

BoT-SORT 与不同检测器的跟踪效果，在本文跟踪数据集上，由两个检测器产生的跟踪效果可知，在引入 YOLOv7-PGC 替换 BoT-SORT 中的目标检测器 YOLOX 后，跟踪模型对于车辆的的跟踪性能有所提高，其中 MOTA 提升 1.5%，IDs 降低了 13。可以看出目标检测检测器的性能对于 BoT-SORT 跟踪模型性能有所影响，目标检测模型性能更好，整体跟踪模型的性能也更强，说明了本文采用 YOLOv7-PGC 的改进检测器在车辆多目标跟踪中比原检测器具有更好的效果和鲁棒性，最后将不同检测器与改进后 BoT-SORT 对比，融合后的跟踪算法效果最优。综上所述，本文最终融合的跟踪算法准确度为 61.3%，对比原 BoT-SORT 算法 MOTA 值提高了 5.1%，跟踪连续性效果更良好，ID 跳变为 78，减少了 77 次。

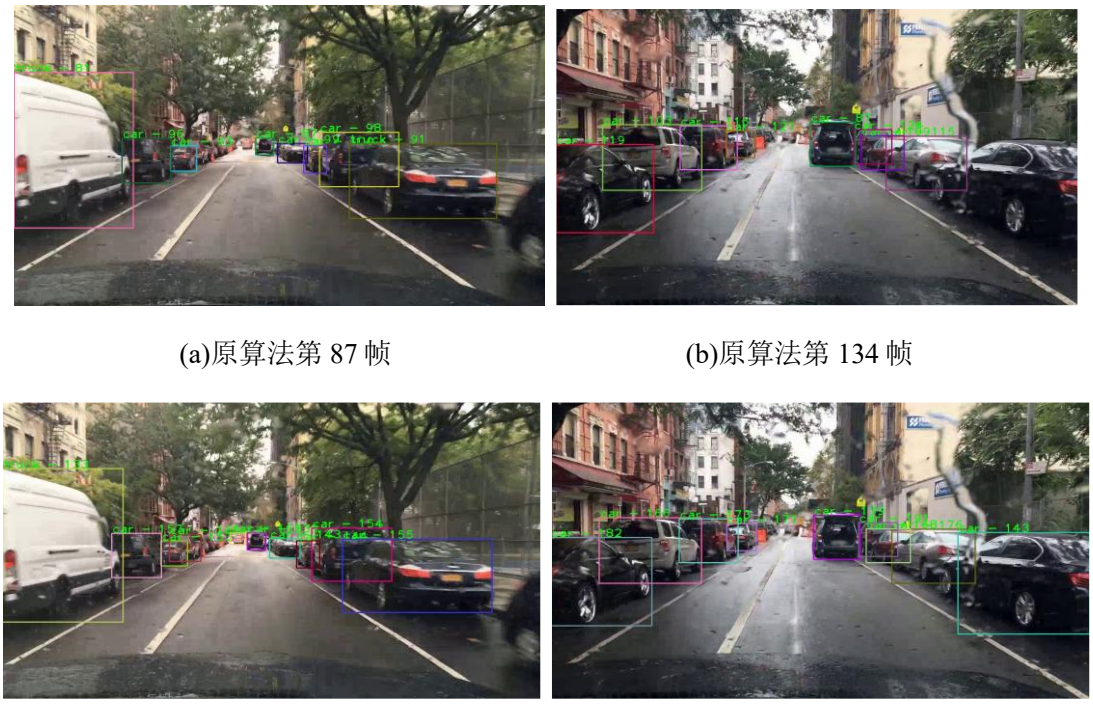
将改进的方法与传统的数字图像处理方法和几种主流的 MOT 方法进行比较。跟踪性能实验对各种 MOT 算法和原始模型进行试验对比，检测器部分统一为 YOLOv7-PGC 模型，实验选取了五种被广泛认可的 MOT 算法和本文改进的跟踪模型进行对比实验：SORT、StrongSORT、DeepSORT、ByteTrack、BoT-SORT 和本文算法。实验开始前，每个模型都利用了从公开 MOT20 数据集上训练中获得的预训练权重，并在跟踪数据集上进行了 100 个周期的训练阶段。

表 4-4 不同模型对比实验

| 模型 | 准确度 MOTA（%） | 精确度 MOTP（%） |
|---|---|---|
| SORT | 35.4 | 72.3 |
| StrongSORT | 44.5 | 74.6 |
| DeepSORT | 42.7 | 74.8 |
| ByteTrack | 54.8 | 76.5 |
| BoT-SORT | 57.7 | 77.1 |
| ours | 61.3 | 78.6 |

对比实验结果如上表 4-4 所示，在这里选取了 MOTA 和 MOTP 作为综合衡量模型跟踪性能的指标，由上表可以看出 BoT-SORT 的跟踪准确度和精确度相比较其他模型都更高。而本文改进后的 BoT-SORT 模型在所有指标上都优于其他方法，在自建车辆数据集上的跟踪效果达到了 61.3%的 MOTA 值，同时

MOTP 值也比其他模型更好为 78.6%。相比与原来的 BoT-SORT 模型，跟踪的准确度提高了 3.6%，跟踪精确度提升了 1.5%。SORT 和 DeepSORT 模型虽然适用性广泛，但在本实验中表现出的多目标跟踪效果较差。它们的 MOTA 分别为 35.4%、42.7%。虽然 StrongSort 利用 NSA 非线性 KF 模块来预测复杂运动目标的状态，但在实验的极端环境中难以发挥作用，导致在本实验中的跟踪效果不佳，表现在 MOTA 为 44.5%，ByteTrack 模型展示了其分段匹配的特征关联能力，这归因于其类似于 BoT-SORT 算法的分阶段匹配高低置信度检测框，增强 MOT 模型对实时环境的车辆跟踪适用性。BoT-SORT 算法通过采用运动相机补偿和卡尔曼滤波直接输出的方式显著提高了算法的跟踪准确度，其多特征融合方法更是对在复杂场景下实现了检测精度与 ID 连续性产生了重要作用，从 MOTA 和 MOTP 分别达到了 57.7%、77.1%可以验证，在本文的复杂场景车辆数据集的实验背景下，BoT-SORT 相比较其他 MOT 跟踪模型综合性能更好。因此，选用 BoT-SORT 作为本文的改进基础跟踪器更具鲁棒性和适应性。



(a)原算法第 87 帧　　(b)原算法第 134 帧

(c)本文算法第 87 帧　　(d)本文算法第 134 帧

图 4-9 跟踪对比效果图

本章改进的算法与原算法如上图 4-9 所示，通过上面几张测试视频帧的跟踪结果图可以更加直观、便捷的对实验改进效果有更加明显的对比，在 MOT8 视频中的雨天环境下，第 87 帧中，对比原 BoT-SORT 算法，改进后的算法在车辆

跟踪过程中，面对复杂环境如雨天环境下，改进后算法的准确度更高，车辆误检率更低，连续性有明显的提高。在视频的第 134 帧跟踪结果图中，由图中可以看出，原跟踪算法对于车辆跟踪丢失，且对于遮挡情况下，跟踪效果较差，而改进后的跟踪效果更好，能够成功的识别遮挡等场景下目标车辆。在复杂环境下改进后的跟踪模型对多目标车辆跟踪具有更强的稳定性，对比原算法在各种环境下的多目标跟踪更有优势。

## 4.4 本章小节

本章对于 BoT-SORT 算法在车辆跟踪环境下的实验进行研究，引用多种优化方法，对原算法模型 ID 跳变、跟踪易丢失等问题进行更有目的性的解决，通过改进对原算法的结构，实现了更优良的综合跟踪效果。首先，介绍了 Bot-SORT 多目标跟踪算法的基本组成和工作流程，然后，对本章改进的自适应卡尔曼滤波、CIoU 匹配进行介绍，优化后的 CIoU 和 AKF 有效地减少了 ID 跳变次数，同时提高了跟踪准确度，同时融合带有 SURF 特征描述符的 VLAD 方法进行外观特征匹配，对于非线性状态目标跟踪连续性表现出色，并且对改进后的模型进行对比消融实验。改进的 YOLOv7-PGC 和 BoT-SORT 跟踪算法对比原 YOLOX 和 BoT-SORT 算法 MOTA 值提高了 5.1%，MOTP 提升了 2.3%，IDF1 提高了 9.8%，IDs 减少了 77 次。综上，得出改进方法具有有效性的结论。

# 第 5 章 车辆检测与跟踪系统实验

为了测试多尺度特征融合的模型在实际环境中的适应性和可信性，本章对改进模型进行部署实验，基于 RK3588S 和 GD32F4 处理核心集成软硬件框架搭建了无人车目标检测和跟踪实验平台，对无人车完成算法部署后，选择多中复杂道路环境进行对比实验，最终跟踪结果跟踪 IDs 比原算法减少了 29，漏检率降低了 13.4%。

## 5.1 无人车集成系统设计

### 5.1.1 实验平台硬件设计

上面章节已经对车辆检测跟踪算法进行了多种程度的实验与验证，并根据多次实验结果对模型进行准确性、稳定性等性能参数的分析。为了更进一步的验证基于多尺度特征融合的车辆检测跟踪系统的在实际智能驾驶无人车上搭建架构的部署效果，本章在智能车机器人的基础上，搭建了一个无人车仿真测试平台，如下图 5-1 所示。

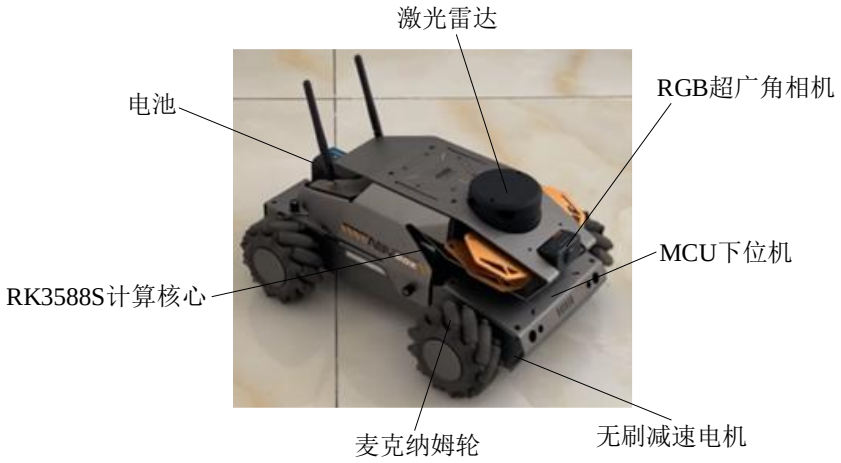



图 5-1 无人车集成平台

U-CAR-02 是科大讯飞公司设计研制的智能移动开发平台，集成了多模态传感器与高性能计算核心，支持机器视觉、SLAM 导航、全向运动控制等主要核心功能。在智能移动机器人平台中，处理器被分为上位机和下位机。

上面所述的硬件部分通过互相协同，组成了智能的高度集成系统，其中RK3588S 上位机作为主要计算处理核心，通过其 USB 接口对相机、激光雷达等传感器进行数据反馈，RGB 超广角相机为实验采集了大量的视觉感知信息，接着通过 CPU、GPU 核心处理单元对复杂的图像处理等任务进行计算处理。同时通过 UART 通信协议与 GD32F4 下位机连接，GD32F4 下位机在接口上与 IMU 和电机实现数据反馈任务，实现对实验平台的控制。

### 5.1.2 系统软件框架

本文的实验集成设计以Ubuntu 20.04操作系统为基础，深度集成ROS Noetic框架，形成一套模块化、高实时性的机器人开发平台。上位机搭载 RK3588S 高性能处理器，依托 Ubuntu 20.04 的稳定性和 ROS Noetic 的分布式架构，实现多传感器数据融合与复杂视觉信息的处理。集成系统的软件框架如下图5-2所示。

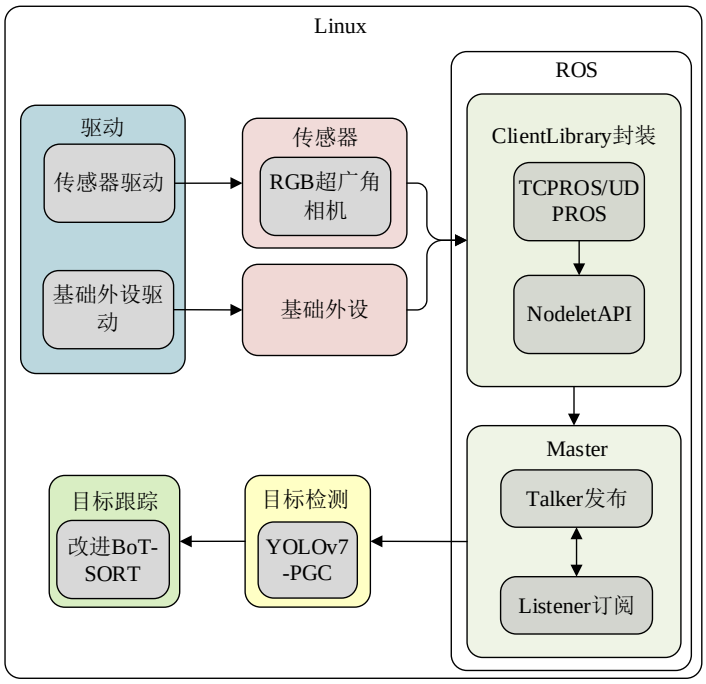



图 5-2 系统软件框架结构图

基于移动机器人的集成实验平台下的系统软件框架可以大致分为一下几个主要部分：

数据采集:由驱动器对传感器和基础设备进行驱动，RGB 超广角相机感知周围环境，采集视觉信息。

数据传输:TCPROS/UDPROS：ROS 的通信协议，分别用于可靠传输和低延

迟传输。Nodelet API：高效节点内通信接口，减少数据拷贝开销，用于高频率相机数据处理，减小节点通信延迟。

节点通信：通过 Talker 将目标检测发布到指定的话题，通过 Listener 订阅话题不间断接受相机传输的数据

数据处理与算法应用：图像数据被传输至 YOLOv7-PGC 中，输出检测结果，改进 BoT-SORT 获取结果，匹配检测框，生成轨迹 ID，输出轨迹结果。

在驱动 RGB 相机前，需要在系统中安装配置 OpenCV 库，以便获取相机数据，通过 cv_bridge 库将 OpenCV 图像数据转换为 ROS 消息格式，实现与 ROS 节点的无缝对接，加载已经训练完成的模型框架，从而对检测目标进行分类和识别。在 Linux 基础下采用 Ubuntu20.04 作为底层操作系统，负责对硬件资源进行管理分配，搭建稳定性可靠的运行环境。

## 5.2 融合多目标检测和跟踪实验

### 5.2.1 实验环境和模型部署

本文的部分实验环境如下图 5-3 所示。在保证实验的安全性和防止其他干扰，通过相机获取多个道路环境下的车辆场景视频和图片，并在无人车上对图像进行检测和跟踪实验。

(a)实验场地前方　　(b)实验场地后方

图 5-3 部分实验场地

本文所用的无人车平台为 ROS 系统框架，同时需要在工作空间中提前完成 OpenCV 安装和环境部署。本文无人车平台检测跟踪系统的 ROS 中话题之间的通信结构如下图 5-4 所示。

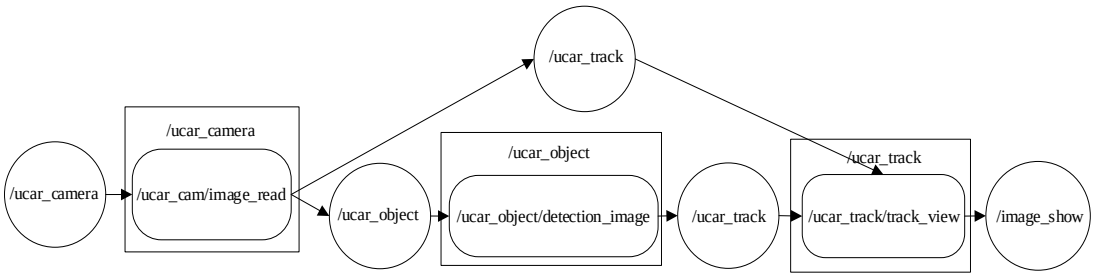



图 5-4 系统 ROS 通信结构图

为了将改进模型应用到无人车实验平台上，并在无人车系统中成功完成目标检测和跟踪任务，需要对整体实验进行预先部署，首先连接数据 wifi，使用 MobaXterm 连接无人车远程终端，然后对无人车进行启动操作，执行 roslaunch 指令，启动 base_driver 节点发布话题控制无人车的底盘，使用键盘控制无人车移动，运行 ucar_camera.py 文件进行视频采集和图像获取，输入目标检测和跟踪模型脚本文件的启动命令，使无人车进行车辆目标检测和跟踪任务。

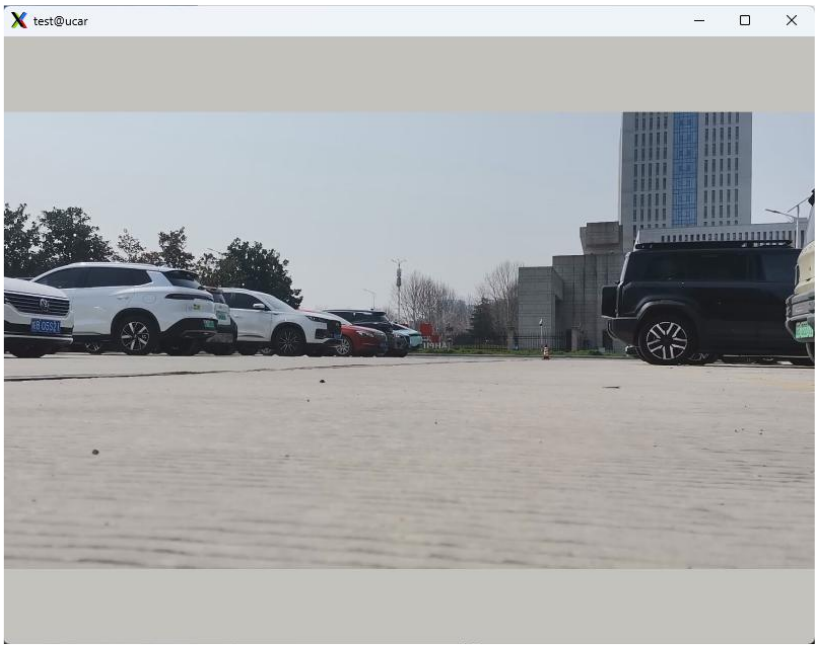



图 5-5 MobaXterm 摄像头界面

ROS 系统中，通过 ucar_camera 节点订阅 ucar_camera 功能包中的 /ucar_camera/image_read 话题，从而获取摄像头信息，从相机中获取采集的视频和图像信息，然后通过话题发布出去。ucar_object 节点订阅 RGB 相机发布的图像视频信息，然后对图像进行目标检测任务，RKNN 框架对训练的 YOLOv7-PGC 模型权重进行加载，OpenCV 的 DNN 模块可以进行目标检测和特征提取。

```
    # 目标检测
    ucar_object = Node(
        package='ucar_object',
        executable='detection_node',
        name='object_detector',
        remappings=[
            ('camera/image_processed', '/ucar_camera/image_processed'),
            ('detection_result', '/ucar_object/detection_image')
        ],
        parameters=[{'model_path': 'weights/yolov7-pgc.rknn'}]
    )
    # 跟踪
    ucar_track = Node(
            package='ucar_track',
            executable='tracker_node',
            name='primary_tracker',
            remappings=[
                ('detection_input', '/ucar_object/detection_image'),
                ('track_output', '/ucar_track/track')
            ]
        )
```

图 5-6 部分节点文件

ucar_track 节点订阅/ucar_object/detection_image 话题获取检测结果和特征，包括检测框位置，置信度等信息，根据 BoT-SORT 算法的运动相机补偿特点，通过 ucar_track 节点订阅相机的图像信息。最后通过/ucar_track/track_view 话题将跟踪的结果发布出去，最后将视频保存至结果文件夹中。

### 5.2.2 无人车目标检测实验

本文在前两章中已在数据集上对 YOLOv7-PGC 模型和改进的 BoT-SORT 算法进行多种实验，为了更好的验证在真实道路环境中模型对于车辆检测和跟踪的可行性和稳定性，实验选取了校园内的真实道路作为实验场地，在检测实验中，通过摄像头获取了白天、夜晚环境下的共计 258 张图片，本实验中对于检测结果使用置信度(Confidence Score)来评估检测实际效果，不仅决定了实际检测性能的优劣，更对跟踪效果产生直接影响，既反映了目标存在的概率，另又体现类别的分布概率。

（1） 白天环境

在下图 5-7 中图 a 至图 d 分别为车辆检测实验中不同状态下的检测效果。可以由检测结果看出有多个车辆目标被分类，且分配的类别标签未发生误识别的情况，图 a 和图 b 分别为不同距离的检测效果，图 a 中对与远处的目标，模型也能够准确的识别。图 c 的存在车辆遮挡的情况，目标的置信度仍较高，检测效果依旧良好，图 d 对于多个目标检测效果稳定。在白天环境的目标检测中，对于简单车辆检测置信度均大于 0.9，同时在多目标场景下未出现漏检误检情况。


(a)远距离　　(b)近距离


(c)有遮挡　　(d)多目标

图 5-7 白天环境检测图

（2）夜间环境

本文的目标检测实际实验中，除了选取了白天环境下多个场景下车辆检测环境，还针对夜晚光线暗的环境基于无人车平台进行实验。


图 5-8 夜间环境检测图

如上图 5-8 所示，无人车部署 YOLOv7-PGC 模型后对夜晚车辆的检测效果良好，图 b 中对光线暗的远处车辆也能达到 0.81 的置信度，检测效果稳定。图 c 和图 d 中，对夜间环境下车流量大、环境复杂的路口进行检测实验，由图可以看出对于远距离小目标的检测效果依旧良好，图 c 中左侧车辆被行人的障碍遮

挡下，对车辆仍能进行准确识别。

（3）实验结果对比


(a)YOLOv7　　(b)YOLOv7-PGC

图 5-9 实际检测对比图

同时，为了验证实际环境下 YOLOv7-PGC 改进算法的有效性，实验在无人车平台进行检测对比实验，由图 5-9 的图 a1 和图 b1 对比可以看出，在无人车实验平台上对模型部署后，改进后的 YOLOv7-PGC 模型检测准确度更高，车辆目标漏检降低，尤其对小目标检测效果更好。图 a2 和图 b2 中，对夜间复杂环境进行检测对比实验，由图 a2 可以看出，YOLOv7 算法对于遮挡的车辆小目标产生了漏检，同时对近距离的同一个目标产生了误检，对比图 b2 本文的 YOLOv7-PGC 算法检测精度更好，在夜间环境下无人车检测实验的平均置信度为 0.76，误检个数 FP 为 15，漏检个数 FN 为 17。

表 5-1 实际检测对比结果

| 模型 | 白天环境 | | 夜间环境 | | 平均置信度/socre |
|---|---|---|---|---|---|
| | 误检/个 | 漏检个 | 误检/个 | 漏检/个 | |
| YOLOv7 | 16 | 22 | 29 | 38 | 0.73 |
| YOLOv7-PGC | 8 | 11 | 15 | 17 | 0.81 |

在夜间环境下，由于光线昏暗，同时本文所选实验环境情况复杂，导致整体实际实验的平均置信度不高，同时受摄像头硬件限制，检测实时帧率保持平均 30 帧/s。两组实验结果表明，YOLOv7-PGC 算法对比原算法平均置信度提高

了 10.9%，误检和漏检分别减少了 22、32。综上，本文所采用的目标检测改进方法在无人车上进行了多场景复杂实验，根据实验结果，验证了改进的有效性和可行性，实时帧率满足在车辆检测过程中的需求。

### 5.2.3 无人车道路环境跟踪实验

为了验证无人车集成系统部署多尺度特征融合目标跟踪模型的可操作性和实效性，本文使用无人车在实验场地对车辆进行跟踪实际测试，对复杂道路多场景保存了 7 个跟踪结果视频，共 1134 张视频帧图片。

（1）多目标静态场景

无人车在多个静态车辆目标的实际实验效果图如下图 5-10 所示，由图 a 到图 b 的跟踪效果可以看出，箭头所指目标在被遮挡较为严重时，算法对于真实目标跟踪的效果依旧稳定，未定义为新的 ID，图 c 与图 d 中当无人车处于不规则运动状态时，跟踪的状态并没有随着非线性运动而产生较大的误差，车辆的跟踪效果稳定，实验效果良好，跟踪的实验过程中 IDs 切换次数为 13 次，实时跟踪帧率 29 帧/s，验证了本文改进后模型在无人车集成平台上也具有适用的稳定性和有效性。


图 5-10 无人车跟踪实验图

（2）白天动态场景

为了验证融合模型在复杂交通环境下的适应性，对无人车实验平台进行复杂环境实验，选择一个交通车流量繁杂的路口，对本文模型和原模型进行实

验，并总结实验结果。如下图 5-11。


(a)BoT-SORT　　(b)YOLOv7-PGC+改进 BoTSORT

图 5-11 复杂环境对比

由图 a1 和图 b1 对比可以看出，由于本文跟踪模型引入 YOLOv7-PGC 检测器，对比原 BoT-SORT 算法对于远处小目标跟踪效果更好，由图 a2 可以看出，原模型在复杂环境下因为环境干扰和运动状态，导致原算法在复杂车辆环境下的跟踪误检率较高，同时还出现了漏检的问题。而且根据表 5-2 的实验结果表明，本文改进后的融合跟踪算法对于复杂环境下的目标跟踪 IDS 更小为 24 次。同时漏检率更低，实时帧率为 28/s 略低于原算法，但满足车辆跟踪的实时性需求。

表 5-2 白天复杂环境对比

| 模型 | 漏检率/% | IDs/次 | FPS/s |
|---|---|---|---|
| BoT-SORT | 35.1 | 53 | 30 |
| YOLOv7-PGC+改进 BoTSORT | 21.7 | 24 | 28 |

（3）夜间动态场景

因目标检测在夜晚环境下的效果良好，实验另外选取了夜晚下交通复杂环境的场景进行跟踪实验。

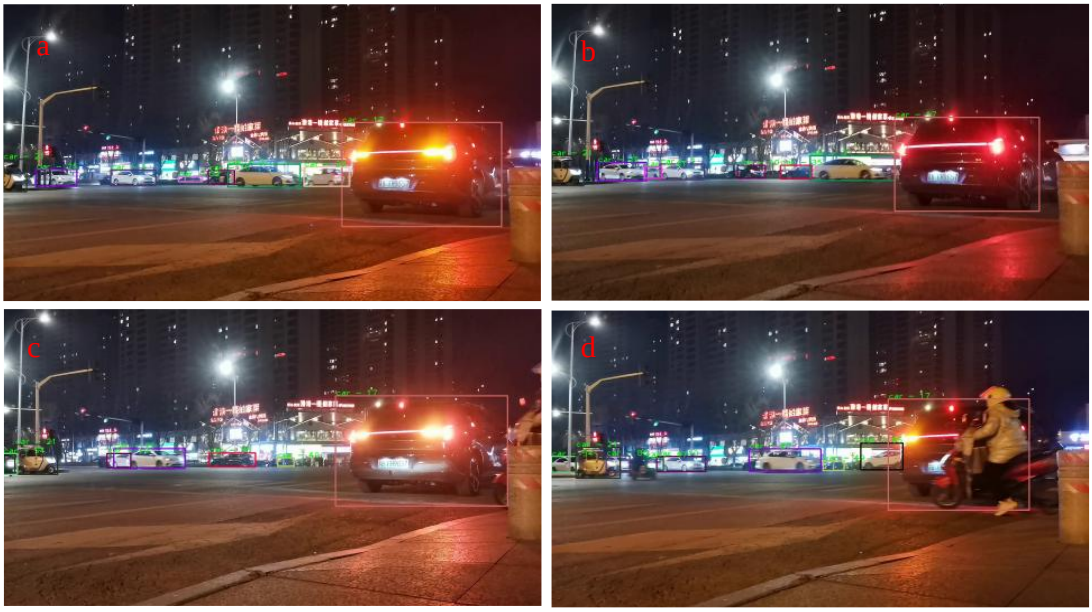



图 5-12 夜间场景跟踪

上图 5-12 为无人车实验平台在夜晚环境的跟踪效果，由图 a 和图 b 可以看出，改进模型对于运动车辆的跟踪效果稳定，轨迹框贴近真实目标位置，且不会因为目标运动而产生 ID 切换，图 c 和图 b 中电动车对车辆产生大面积遮挡，而改进模型依然能对被遮挡的车辆目标进行持续跟踪，并且没有产生 ID 变换，左边白车远距离遮挡下也能进行准确检测，并进行 ID 保持不变的持续跟踪任务。模型在夜间环境下跟踪中，整个过程的车辆 ID 切换次数为 37，实验结果表明，改进后的模型跟踪效果更好，改进后的融合检测和跟踪模型在复杂环境下具有鲁棒性和稳定性。

## 5.3 本章小结

本章基于无人车移动机器人搭建检测和跟踪实验平台，从硬件集成设计和软件框架对无人车进行介绍，接着对真实道路环境视频进行测试，通过 ROS 系统对目标检测和跟踪算法进行部署，远程控制无人车在校园安全道路环境下进行实际环境实验，验证了本文基于 YOLOv7 和 BoT-SORT 算法改进的多尺度特征融合模型满足可行性条件，无人车部署改进的检测和跟踪模型后，在白天复杂的多目标车辆交通场景中，对目标的跟踪 IDs 为 24，比原算法减少了 29，漏检率降低了 13.4%，且实时帧率满足车辆跟踪实际需求。实验结果表明，无人车集成系统具有准确的跟踪效果，可以满足在复杂环境中的车辆跟踪条件。

# 第 6 章 总结与展望

## 6.1 总结

智能驾驶在现代社会的交通领域中发展迅速，多目标检测和跟踪在自动驾驶领域的应用是环境感知系统的核心模块，其通过动态关联与预测周围目标的轨迹，为车辆感知环境信息提供了关键支持。但是面临极端天气、动物闯入等复杂场景的模型泛化能力较弱，对于复杂环境下的轨迹预测误差较大。针对这些问题，本文对 YOLOv7 检测算法和 BoT-SORT 跟踪算法进行深入研究，提出了一种改进后的多尺度特征融合算法。本文主要的研究内容如下：

（1）介绍了目标检测与跟踪领域的研究现状，阐明车辆目标感知技术在智能交通与自动驾驶中的核心价值。通过分析 YOLO 系列、SORT 等算法的理论优势与场景适配性，构建了融合多尺度架构与动态特征优化的技术体系，确立了多尺度特征融合检测网络与自适应跟踪协同优化的技术路线，为复杂交通场景下的车辆感知研究提供了坚实的理论支撑与方法论框架。

（2）在特征金字塔网络与路径聚合网络融合架构的基础上，创新性构建了多尺度特征融合网络。通过增设小目标专用检测层与多层级横向跨连接，实现跨尺度特征交互强化，在骨干网络嵌入GE注意力模块，增强复杂背景下的语义特征聚焦能力；并且在颈部及检测头部分引入协调坐标卷积，通过构建更强的空间模型，提升空间信息感知精度。优化后的 YOLOv7 模型在 BDD100K 测试集上实现 48.2%的平均精度，较原模型提升 5.6 个百分点，同时维持 86FPS 的实时推理速度。验证了算法在速度精度平衡上的优越性，满足车辆检测和跟踪的实时性、精确性需要。

（3）针对道路行驶场景下 BoT-SORT 算法存在的轨迹关联鲁棒性不足及身份连续性不稳定的问题，系统性的提出一种基于 BoT-SORT 算法的改进框架，替换卡尔曼滤波为适应性更强的自适应卡尔曼滤波，引用动态噪声估计机制提升非线性运动目标的轨迹预测精度，其次融合改进 CIoU 匹配机制，通过 VLAD 特征聚合，增强目标特征匹配能力。实验表明，改进算法在密集遮挡及高速变向场景中展现出优越的轨迹连续性，显著降低身份切换频次并改善目标识别准

确率。

（4）搭建了由硬件架构 RK3588S 和 GD32F4 为核心以及 ROS 操作系统为软件框架基础的无人车集成系统，对本文优化后的 YOLOv7-PGC 和改进 BoT-SORT 算法进行协同部署，实验采用实验道路场景数据集离线测试与校园封闭场地实车验证双结构验证的方法，通过远程终端获取车载相机实时感知图像信息并同步传感器数据，实验充分验证了多尺度特征融合算法系统部署的有效性，满足车辆检测和跟踪的实时性和轨迹连续性需要。

## 6.2 展望

本文面向复杂道路环境对车辆的多目标检测和跟踪进行研究，深度学习技术在智能驾驶尚未成熟的智慧交通领域展现出巨大的发展潜力。本文提出的融合算法通过实验验证了其改进的有效性，提高了跟踪和检测准确度，在无人车平台上也能够成功部署并实验，但是面对更复杂的场景仍然具有一些局限性，将来还可以从下面几个方面进行下一步改进：

（1）多传感器融合。依靠视觉进行检测和跟踪任务具有一定的局限性，融合多传感器数据和时序信息，提升复杂场景下的鲁棒性。例如通过结合多个红外图像、激光雷达、RGB-D 相机等传感器，可以加强模型对目标特征的多维提取，提高检测和跟踪的鲁棒性。

（2）搭建轻量化框架。本文模型对多目标的检测和跟踪准确度有了很大的提升，虽然满足自动驾驶领域中大于 25FPS 的实时率要求，但是当面对例如雨雪、雾霾、夜晚等极端环境下，受限于硬件条件可能会导致实时性急剧下降，所以下一步需要对模型进行轻量化设计，减少模型算力成本，提高整体的泛化性。

（3）应用场景扩展。本文改进的算法及设计的无人车系统均是基于车辆视角进行实验的，将来可以采用无人机-无人车协同的系统对农业、医疗场景进行应用，例如利用多目标跟踪实现作物生长状态的实时监控，运用目标检测对于小目标检测的适用性，对医疗中的微小细胞进行长期检测。

# 参考文献

[1] Chan-Edmiston S, Fischer S, Sloan S, et al. Intelligent transportation systems (its) joint program office: strategic plan 2020–2025[R]. United States: Intelligent Transportation Systems Joint Program Office, 2020.

[2] 马永杰, 程时升, 马芸婷, 等. 卷积神经网络及其在智能交通系统中的应用综述[J]. 交通运输工程学报, 2021, 21(04): 48-71.

[3] Qadri S S S M, Gökçe M A, Öner E. State-of-art review of traffic signal control methods: challenges and opportunities[J]. European transport research review, 2020, 12: 1-23.

[4] Grigorescu S, Trasnea B, Cocias T, et al. A survey of deep learning techniques for autonomous driving[J]. Journal of field robotics, 2020, 37(3): 362-386.

[5] Tsakanikas V, Dagiuklas T. Video surveillance systems-current status and future trends[J]. Computers & Electrical Engineering, 2018, 70: 736-753.

[6] Hu L, Bao X, Lin M, et al. Research on risky driving behavior evaluation model based on CIDAS real data[J]. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2021, 235(8): 2176-2187.

[7] Wang S, Li C, Ng D W K, et al. Federated deep learning meets autonomous vehicle perception: Design and verification[J]. IEEE network, 2022, 37(3): 16-25.

[8] Lowe D G. Object Recognition from Local Scale-Invariant Features[C]//Proceedings of the 7th IEEE International Conference on Computer Vision. Corfu, Greece: IEEE, 1999: 1150-1157.

[9] Viola P, Jones M J. Rapid object detection using a boosted cascade of simple features[C]//Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Kauai, HI, USA: IEEE, 2001: 511-518.

[10] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, CA, USA: IEEE, 2005: 886-893.

[11] Bay H, Tuytelaars T, Van Gool L. Surf: Speeded up robust features[C]//Computer Vision–ECCV 2006: 9th European Conference on Computer Vision. Graz, Austria: Springer Berlin

Heidelberg, 2006: 404-417.

[12] 杜娟, 崔少华, 晋美娟, 等. 改进 YOLOv7 的复杂道路场景目标检测算法[J]. 计算机工程与应用, 2024, 60(01): 96-103.

[13] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. Santiago, Chile: IEEE, 2015: 1440-1448.

[14] 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 and machine intelligence, 2016, 39(6): 1137-1149.

[15] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition,Las Vegas. NV, USA,2016: 779-788.

[16] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Honolulu, HI, USA, 2017: 7263-7271.

[17] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Vancouver, BC, Canada, 2023: 7464-7475.

[18] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]//European Conference on Computer Vision, Amsterdam. The Netherlands, 2016: 21-37.

[19] 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. Venice, Italy, 2017: 2980-2988.

[20] Cao Y, Li C, Peng Y, et al. MCS-YOLO: A multiscale object detection method for autonomous driving road environment recognition[J]. IEEE Access, 2023, 11: 22342-22354.

[21] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]// Proceedings of the IEEE/CVF international conference on computer vision, Montreal, QC, Canada, 2021: 10012-10022.

[22] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Nashville, TN, USA, 2021: 13713-13722.

[23] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City, UT, USA,

2018: 8759-8768.

[24] Qin L, Shi Y, He Y, et al. ID-YOLO: Real-time salient object detection based on the driver's fixation region[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(9): 15898-15908.

[25] 盛博莹, 侯进, 李嘉新, 等. 面向复杂交通场景的道路目标检测方法[J]. 计算机工程与应用, 2023, 59(15): 87-96.

[26] Qiao S, Chen L C, Yuille A. Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Nashville, TN, USA, 2021: 10213-10224.

[27] Wang C , Dai Y , Zhou W ,et al. A Vision-Based Video Crash Detection Framework for Mixed Traffic Flow Environment Considering Low-Visibility Condition[J].Journal of Advanced Transportation, 2020, 2020:1-11.

[28] Szeliski R. Computer Vision: Algorithms and Applications[M]. London: Springer Nature, 2022: 45-68.

[29] Zhang Y C, Wang T, Liu K X, et al. Recent advances of single-object tracking methods: A brief survey[J]. Neurocomputing, 2021, 455: 1-11.

[30] Arulampalam M S, Maskell S, Gordon N, et al. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking[J]. IEEE Transactions on Signal Processing, 2002, 50(2): 174-188.

[31] Bolme D S, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters[C]//Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA: IEEE, 2010: 2544-2550.

[32] Henriques J F, Caseiro R, Martins P, et al. Exploiting the circulant structure of tracking-by-detection with kernels[C]//Proceedings of the 12th European Conference on Computer Vision. Florence, Italy: Springer, 2012: 702-715.

[33] Henriques J F, Caseiro R, Martins P, et al. High-speed tracking with kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 583-596.

[34] Danelljan M, Hager G, Khan F S, et al. Beyond correlation filters: Learning continuous convolution operators for visual tracking[C]//Proceedings of the 14th European Conference on

Computer Vision. Amsterdam, Netherlands: Springer, 2016: 472-488.

[35] Lu X, Ma C, Ni B, et al. Adaptive region proposal with channel regularization for robust object tracking[J]. IEEE transactions on circuits and systems for video technology, 2019, 31(4): 1268-1282.

[36] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-Convolutional Siamese Networks for Object Tracking[C]//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, Netherlands: Springer, 2016: 850-865.

[37] Li B, Yan J, Wu W, et al. High performance visual tracking with siamese region proposal network[C]//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer, 2018: 120-135.

[38] Li B, Wu W, Wang Q, et al. SiamRPN++: Evolution of siamese visual tracking with very deep networks[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019: 4277-4286.

[39] Yan B, Zhao H, Wang D, et al. Towards grand unification of object tracking[C]//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer, 2022: 733-750.

[40] Chen X, Peng H, Wang D, et al. Sequence to sequence learning for visual object tracking[C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE, 2023: 14572-14581.

[41] Fragkiadaki K, Shi J. Detection free tracking: Exploiting motion and topology for segmenting and tracking under entanglement[C]//CVPR 2011. Colorado Springs, CO, USA: IEEE, 2011: 2073-2080.

[42] Bewley A, Ge Z, Ott L, et al. Simple Online and Realtime Tracking[C]//Proceedings of the 2016 IEEE International Conference on Image Processing. Phoenix, AZ, USA: IEEE, 2016: 3464-3468.

[43] Chen L, Ai H, Zhuang Z, et al. Real-time multiple people tracking with deeply learned candidate selection and person re-identification[C]//2018 IEEE international conference on multimedia and expo (ICME). San Diego, CA, USA：IEEE, 2018: 1-6.

[44] Wojke N, Bewley A, Paulus D. Simple online and realtime tracking with a deep association metric[C]//2017 IEEE international conference on image processing (ICIP). Beijing, China:

IEEE, 2017: 3645-3649.

[45] 阎婧宇, 谢永华. 基于 YOLOv7-Tiny 算法的无人机实时跟踪野生动物方法[J]. 野生动物学报, 2024, 45(02): 251-261.

[46] Yang X, Gao Y, Yin M, et al. Automatic Apple Detection and Counting with AD-YOLO and MR-SORT[J]. Sensors, 2024, 24(21): 7012.

[47] Tumanyan N, Singer A, Bagon S, et al. Dino-tracker: Taming dino for self-supervised point tracking in a single video[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024: 367-385.

[48] Liu Z, Shang Y, Li T, et al. Robust multi-drone multi-target tracking to resolve target occlusion: A benchmark[J]. IEEE Transactions on Multimedia, 2023, 25: 1462-1476.

[49] Cao J, Pang J, Weng X, et al. Observation-centric sort: Rethinking sort for robust multi-object tracking[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Vancouver, Canada: IEEE, 2023: 9686-9696.

[50] Zhang Y D, Dong Z, Chen X, et al. Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation[J]. Multimedia Tools and Applications, 2019, 78: 3613-3632.

[51] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Honolulu, HI, USA: IEEE,2017: 2881-2890.

[52] Liu L, Ouyang W, Wang X, et al. Deep learning for generic object detection: A survey[J]. International journal of computer vision, 2020, 128: 261-318.

[53] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Columbus. USA： IEEE, 2014: 580-587.

[54] Boukerche A, Hou Z. Object detection using deep learning methods in traffic scenarios[J]. ACM Computing Surveys (CSUR), 2021, 54(2): 1-35.

[55] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]//Proceedings of the IEEE international conference on computer vision. Venice, Italy: IEEE 2017: 2961-2969.

[56] Sharma S, Sharma S, Athaiya A. Activation functions in neural networks[J]. Towards Data Sci, 2017, 6(12): 310-316.

[57] 宋绍剑, 夏海姐, 李刚, 等. YOLOv5 的改进算法及其在自动驾驶多目标检测的应用研究[J]. 计算机工程与应用, 2023, 59(15): 68-75.

[58] 刘昌华. 复杂交通场景下自动驾驶道路目标检测[D]. 大连: 大连理工大学, 2022.

[59] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Honolulu, HI, USA, 2017: 2117-2125.

[60] Chen Z, Li Q, Feng H, et al. Nonuniformly dehaze network for visible remote sensing images[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA: IEEE, 2022: 447-456.

[61] Yu F, Chen H, Wang X, et al. Bdd100k: A diverse driving dataset for heterogeneous multitask learning[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. New York, USA: IEEE, 2020: 2636-2645.

[62] Pan X, Ge C, Lu R, et al. On the integration of self-attention and convolution[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. New Orleans, LA, USA: IEEE,2022: 815-825.

[63] Khodarahmi M, Maihami V. A review on Kalman filter models[J]. Archives of Computational Methods in Engineering, 2023, 30(1): 727-747.

[64] Dufaux F, Konrad J. Efficient, robust, and fast global motion estimation for video coding[J]. IEEE transactions on image processing, 2000, 9(3): 497-501.

[65] Zhang H, Cao H, Yang X, et al. Self-training with progressive representation enhancement for unsupervised cross-domain person re-identification[J]. IEEE Transactions on Image Processing, 2021, 30: 5287-5298.

[66] Zheng Z, Wang P, Liu W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]//Proceedings of the AAAI conference on artificial intelligence. New York, USA: AAAI, 2020, 34(07): 12993-13000.

[67] Jégou H, Douze M, Schmid C, et al. Aggregating local descriptors into a compact image representation[C]//2010 IEEE computer society conference on computer vision and pattern recognition. San Francisco, CA, USA: IEEE, 2010: 3304-3311.

[68] Dendorfer P, Osep A, Milan A, et al. Motchallenge: A benchmark for single-camera multiple target tracking[J]. International Journal of Computer Vision, 2021, 129: 845-881.

# 攻读学位期间发表的学术论文目录

（1）发表论文

[1] 江自豪、杨思远、王世康、王坤相、何宇豪、王冠凌. 改进 YOLOv7 的自动驾驶目标检测算法 [J]. 传感技术学报,（已录用，第一作者）。

[2] 杨思远、江自豪、王坤相、王世康、何宇豪、易方旭、王冠凌. 基于改进 YOLOV7 的多场景火灾识别算法[J].红外技术（已录用，第二作者）。

（2）获奖情况

[1] 2022 年第十七届全国大学生智能汽车竞赛创意组讯飞服务机器人安徽赛区选拔赛三等奖。

[2] 2023 年第十八届全国大学生智能汽车竞赛创意组航天智慧物流线下选拔赛一等奖

[3] 2023 年第十八届全国大学生智能汽车竞赛创意组航天室外 ROS 无人车竞速赛全国总决赛二等奖。

[4] 2023 年第十八届全国大学生智能汽车竞赛创意组航天室外 ROS 无人车竞速赛南部赛区选拔赛二等奖。

[5] 2023 年第十七届“西门子杯”中国智能制造挑战赛离散行业工程实践全国初赛一等奖。

[6] 2024 年第十八届“西门子杯”中国智能制造挑战赛离散行业工程实践全国初赛特等奖。

[7] 2024 年第十九届全国大学生智能汽车竞赛创意组讯飞智慧救援安徽赛区选拔赛二等奖。

# 致 谢

在研究生的三年期间，通过在研究生的学习和生活中，我学到了许多知识，掌握了许多技能，在三年的充实生活中，成功的提升了学术素养，在此过程中，我要感谢家人、老师、同学和许多学校长辈们给予的支持。

我要感谢我的家人，在大四时就给予我考研的支持，在读研期间也一直给予我肯定，每当在学习和生活上有所困扰，他们都会安慰我，不断的支持我，帮助我度过难关，感谢父母对我的支持，在给我生命的同时，教会我做人，给我一个温馨的家庭。寸草之心，虽无三春之报，然血脉所系，此生铭怀。

感谢在研究生学习期间一直给予我帮助的导师王冠凌教授和毛德平老师，不论在科研和学习上还是在毕业论文中，都一直对我悉心指导和耐心包容，给予我宝贵的建议。同时还要感谢实验室的同门和同学，感谢他们的帮助，同时为我的研究生生活增添了许多色彩。

今当临别，万绪难陈。惟愿父母椿龄永茂，恩师桃李长繁，故友前程似月。深恩厚意，结草衔环，惟以余生勤勉，不负所望。

尚德教学 唯实惟新