



控制与决策

Control and Decision

ISSN 1001-0920, CN 21-1124/TP

## 《控制与决策》网络首发论文

题目: 多目标检测与跟踪算法在智能交通监控系统中的研究进展  
作者: 金沙沙, 龙伟, 胡灵犀, 王天宇, 潘华, 蒋林华  
DOI: 10.13195/j.kzyjc.2021.1763  
收稿日期: 2021-10-13  
网络首发日期: 2022-04-06  
引用格式: 金沙沙, 龙伟, 胡灵犀, 王天宇, 潘华, 蒋林华. 多目标检测与跟踪算法在智能交通监控系统中的研究进展[J/OL]. 控制与决策. <https://doi.org/10.13195/j.kzyjc.2021.1763>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

## 多目标检测与跟踪算法在智能交通监控系统中的研究进展

金沙沙<sup>1,2</sup>, 龙伟<sup>1</sup>, 胡灵犀<sup>1</sup>, 王天宇<sup>1,2</sup>, 潘华<sup>2†</sup>, 蒋林华<sup>1†</sup>

(1. 湖州师范学院 信息工程学院, 浙江省 湖州市 313000;

2. 浙江大学湖州研究院, 浙江省 湖州市 313000)

**摘要:** 为了构建人、路、车、云协同一体化的智能交通监控系统, 多目标跟踪的研究具有广泛的应用价值。传统的手工设计特征的方法对高层信息的表征能力有限, 较难进行复杂场景下的多目标跟踪。深度学习以其强大的学习能力, 逐渐渗透到各个行业和领域, 掀起了智能化浪潮。随着深度学习的发展, 多目标跟踪算法的性能也取得了较大的进展。为了宏观把握基于深度学习的多目标跟踪算法的研究进展, 首先, 比较了基于检测的跟踪算法、基于联合检测与跟踪算法、基于单目标跟踪器的多目标跟踪算法的优缺点。其次, 介绍了多目标跟踪算法在智能交通监控场景的应用。最后总结了目前多目标跟踪存在的问题与挑战, 对多目标跟踪算法未来在智能交通领域的发展进行了思考和展望。

**关键词:** 智能交通系统; 多目标跟踪; 深度学习; 智能化; 目标检测; 研究进展

中图分类号: TP391.4

文献标志码: A

DOI: 10.13195/j.kzyjc.2021.1763

## Research Progress of Detection and Multi-object Tracking Algorithm in Intelligent Traffic Monitoring System

JIN Sha-sha<sup>1,2</sup>, LONG Wei<sup>1</sup>, HU Ling-xi<sup>1</sup>, WANG Tian-yu<sup>1,2</sup>, PAN Hua<sup>2†</sup>, JIANG Lin-hua<sup>1†</sup>

(1. School of Information Engineering, Huzhou University, Zhejiang 313000, China; 2. Huzhou Institute of Zhejiang University, Zhejiang 313000, China)

**Abstract:** To build the integrated Intelligent Traffic Monitoring System based on the cooperation of human, road, vehicle and cloud, the research of multi-object tracking has wide application potentials. Traditional methods with handcrafted features are hard to fully represent high-level information, making it difficult to track multi-targets in complex scenes. Deep learning with its powerful learning ability, has gradually been used in various industries and fields, setting off a wave of smart technologies. To understand the research progress on multi-object tracking algorithm based on deep learning, firstly, the pros and cons of three tracking algorithms, namely Tracking by Detection, Joint Detection and Tracking as well as Multi-Object Tracking with Single Object Tracker, are compared. Secondly, the applications of multi-object tracking algorithm in Intelligent Traffic Monitoring System are introduced. Finally, the problems and challenges of multi-object tracking algorithm are concluded, and the growing trend of multi-object algorithms in Intelligent Transportation field is discussed and forecasted.

**Keywords:** Intelligent Transportation System; multi-object tracking; deep learning; smart technologies; object detection; research progress

## 0 引言

随着城市化进程的加快, 我国公路规模发展迅速。根据交通运输部的数据显示, 截至 2020 年底, 我国公路里程达 519.81 万公里, 其中高速公路 16.1 万公里。但是公路建设的速度仍不及汽车保有量的增长速度。这是导致汽车在道路上行驶时产生交通拥堵、事故频发、交通管理水平低等现象的原因之

一。智能交通系统集成了车联网、物联网、自动控制、智能计算、管理等技术, 充分利用智能化、网联化的技术进行信息融合<sup>[1]</sup>。云端平台汇集了车端、路端等信息后, 可以构建智能化的城市“交通大脑”, 做出交通系统的状态预测、交通的优化调度与控制决策<sup>[2]</sup>, 能够有效提升驾驶安全性和交通流稳定性<sup>[3]</sup>, 使车辆以更加节能的状态行驶, 有助于节能减排<sup>[4]</sup>。

收稿日期: 2021-10-13 录用日期: 2022-02-25.

基金项目: 国家自然科学基金 (61775139); 浙江省级重点研发计划 (2020C02020).

<sup>†</sup>通讯作者. 潘华 (E-mail: pan\_hua\_hzzj@hotmail.com); 蒋林华 (E-mail: 11594@zjhu.edu.cn).

交通监控系统是智能交通系统的子系统之一,在车辆检测与车流量统计、违章行为识别、交通事件预警、逃逸车辆跟踪等方面起到重要作用。

传统的交通监控系统需要大量人工参与,具有劳动强度大、实时性差、人工成本高等缺点。其在城市交通严峻的形势下暴露出一些弊端:(1)极端恶劣天气条件下(例如暴雨、浓雾、黑夜等),交通事故的防范能力不足;(2)“数据孤岛”现象突出,未能达到监控信息同步共享;(3)已有监控设施科技含量低,仍需更新技术,提高效率。基于手工设计特征的方法对高层信息的表征能力有限,难以在复杂场景下进行多目标跟踪。近年来,随着信息处理技术与深度学习技术的发展,基于计算机视觉的目标跟踪方法逐步应用于智能交通监控系统<sup>[5,6]</sup>。

## 1 多目标跟踪算法

视觉目标跟踪是计算机视觉领域的研究热点之一,被广泛应用于智能交通<sup>[7]</sup>、无人机目标侦查<sup>[8]</sup>、水面无人艇<sup>[9]</sup>等领域。鉴于大部分应用场景中需要跟踪的目标数量较多,本文聚焦于归纳和总结多目标跟踪算法。多目标跟踪的任务是在视频中随着时间的推移定位多个移动目标。现有的多目标跟踪算法,按跟踪技术大致可划分为:基于检测的跟踪算法、联合检测与跟踪算法、基于单目标跟踪器的多目标跟踪算法。

### 1.1 基于检测的跟踪算法

基于检测的跟踪(Tracking by Detection, TBD),是一种两阶段目标跟踪的方法(如图1所示),即采用检测器对视频序列的帧进行目标检测,再利用跟踪器提取特征并用数据关联的方法找出目标的轨迹对应关系。

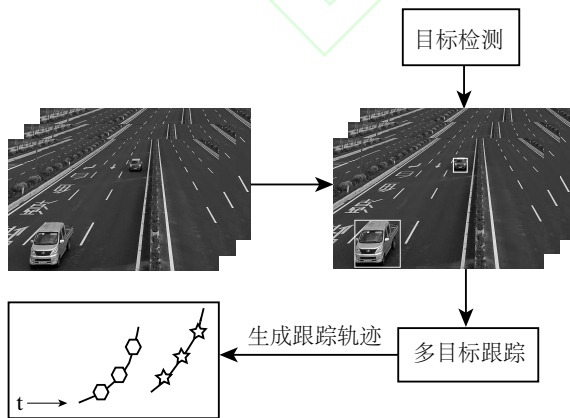


图1 基于检测的跟踪算法

#### 1.1.1 检测器

基于检测的多目标跟踪方法能够跟踪任意数量的目标,但该方法需要训练特定的目标检测器,而检测

结果对跟踪效果的影响很大<sup>[10,11]</sup>。根据是否生成候选区域,目标检测器分为两阶段目标检测方法(Two-stage)和单阶段目标检测方法(One-stage)。两阶段目标检测是指先单独生成目标候选区域,再对其进行分类。单阶段目标检测是指不生成候选区域,直接输出分类和定位信息(如图2所示)。

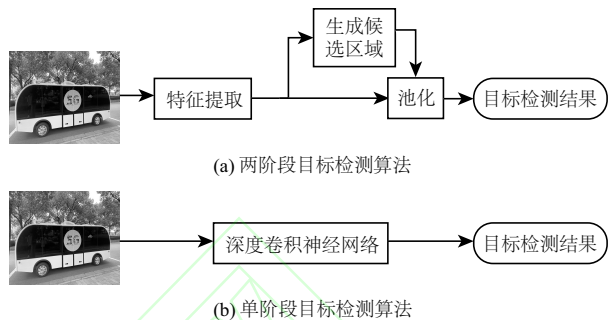


图2 目标检测算法

代表性的两阶段目标检测算法有 R-CNN<sup>[12-15]</sup>系列、SPPNet<sup>[16]</sup>等。R-CNN<sup>[12]</sup>首次将卷积神经网络(Convolutional Neural Network, CNN)引入目标检测任务,提高了目标检测的准确率,搭建了图像分类和目标检测之间的桥梁。然而 CNN 的全连接层限制了输入图像的尺寸,通过裁剪或拉伸的方法使输入图像变成 CNN 所需要的尺寸会导致图像内容丢失或失真。鉴于此, SPPNet 直接将整张图像输入 CNN 中进行特征提取得到特征图,再将选择性搜索算法(Selective search, SS)获取的约 2k 个候选区域映射到特征图上,然后用 4 个尺度的空间金字塔滤波器对每个候选区域的特征进行最大池化,并将池化的结果拼接起来以输出固定长度的特征向量,从而满足全连接层的固定尺寸约束。虽然 SPPNet 通过共享计算的方式,比 R-CNN 的速度提升了几十倍,但 SPPNet 沿用了 R-CNN 中分阶段训练网络的方法,致使训练的速度较慢。针对这个问题, Fast R-CNN<sup>[13]</sup>不仅简化了空间金字塔池化层,提出了感兴趣区域池化层(ROI pooling),还采用多任务损失将训练过程整合为一个阶段,从而使使得在相同配置和同一数据集上, Fast R-CNN 比 SPPNet 训练速度快 3 倍、测试速度快 10 倍,比 R-CNN 训练速度快 9 倍、测试速度快 213 倍。

以上算法均采用了 SS 算法来生成候选区域,影响了模型整体的速度。因此, Faster R-CNN<sup>[14]</sup>提出了区域建议网络替代 SS 算法,进一步将测试时间较 Fast R-CNN 提升了约 9.3 倍。Mask R-CNN<sup>[15]</sup>采用感兴趣区域匹配(ROI Align)来代替 ROI pooling,以解决 ROI pooling 两次将坐标值取整的量化操作

造成的区域不匹配问题,并在 Faster R-CNN 的基础上增加了掩膜损失,整体较 Faster R-CNN 的检测精度提升了 3.6%。Cai 等<sup>[17]</sup>将检测器的检测质量定义为检测框与真实框之间的交并比 (Intersection over union, IoU) 大小。提高检测器训练时的 IoU 阈值虽然能提高检测质量但是会降低检测器的准确率。针对这个问题, Cascade R-CNN<sup>[17]</sup>将多个检测器级联在一起,使前一个 IoU 阈值低的检测器的输出作为后一个 IoU 阈值高的检测器的输入,从而在优化检测质量的同时提高检测精度。经过测试,级联结构能使 R-CNN 系列算法的精度提升 2.4%。Mask R-CNN 与 Cascade R-CNN 均在不同层面上对 Faster R-CNN 进行增强与拓展,但均未解决训练过程中存在的平衡

问题。为了缓解这个问题,Libra R-CNN<sup>[18]</sup>优化了采样方式、改进了特征融合方法,平衡了多任务的损失,使得检测精度较 Faster R-CNN 提升了 2.5%。在两阶段目标检测中,准确地对候选框进行排序是提升检测器性能的关键之一。VFNet<sup>[19]</sup>提出了交并比感知分类评分法 (IoU-Aware Classification Score, IACS) 来同时表示目标存在的置信度和定位精度,并设计了可动态缩放的二元交叉熵损失来训练检测器以预测 IACS。当使用相同的主干网络时, VFNet 的检测精度比 Cascade R-CNN 高 3.9%。

不同算法各有优势,表 1 汇总了典型目标检测算法在不同测试集上的检测时间与平均精度 (mean Average Precision, mAP)

表 1 典型检测方法在不同数据集上的运行速度与平均精度

模式	算法	发布年份	数据集	时间 / ms	mAP / %
Two-stage	R-CNN <sup>[12]</sup>	2014	VOC2012	≥18300	53.3
	SPPNet <sup>[16]</sup>	2015	VOC2007	18300	59.2
	Fast R-CNN <sup>[13]</sup>	2015	VOC2012	1830	65.7
	Faster R-CNN <sup>[14]</sup>	2015	VOC2012	59~198	67.0
	Faster R-CNN <sup>[14]</sup>	2015	COCO test - dev	59~198	36.2
	Mask R-CNN <sup>[15]</sup>	2017	COCO test - dev	200	39.8
	Cascade-R-CNN <sup>[17]</sup>	2018	COCO test - dev	—	42.8
	Libra R-CNN <sup>[18]</sup>	2019	COCO test - dev	—	38.7
	VFNet <sup>[19]</sup>	2021	COCO test - dev	64.1	46.1
One-stage	VFNet <sup>[19]</sup>	2021	COCO test - dev	238.1	55.1
	SSD <sup>[26]</sup>	2016	COCO	61	28
	YOLOv3 <sup>[22]</sup>	2018	COCO test - dev	22	28.2
	YOLOv3 <sup>[22]</sup>	2018	COCO test - dev	51	33
	EfficientDet-D0 <sup>[24]</sup>	2020	COCO test - dev	10.2	34.6
	EfficientDet-D2 <sup>[24]</sup>	2020	COCO test - dev	17.7	43.0
	EfficientDet-D5 <sup>[24]</sup>	2020	COCO test - dev	72.5	51.5
	YOLOv4 <sup>[23]</sup>	2020	COCO test - dev	15.4	43.5
	YOLOv4-P5 <sup>[30]</sup>	2021	COCO test - dev	23.3	51.8
	YOLOv5-l5	2020	COCO test - dev	11.6	48.2
	YOLOv5-l6	2020	COCO test - dev	10.9	48.8
	YOLOX-L <sup>[31]</sup>	2021	COCO test - dev	14.5	50.0

其中 EfficientDet 与 YOLOv5 发布于 2020 年,其最佳检测结果更新于 2021 年。

典型的单阶段目标检测算法有 YOLO 系列<sup>[20-23]</sup>、EfficientDet<sup>[24]</sup>等。YOLOv3<sup>[22]</sup>拓展了主干网络 Darknet 的深度,并借鉴了 ResNet<sup>[25]</sup>的思想,加入了残差跳跃连接,解决了深层次网络的梯度问题。在与 SSD<sup>[26]</sup>达到同样的精度时, YOLOv3 的速度大约是 SSD 的三倍。考虑到实际应用中计算资源与软件支持不足, YOLOv3 仍是目前工业上应用最广泛的检测器之一。为了克服资源限制的难题,许多研究人员都致力于优化模型——在提高检测器的准确性的同时保证检测器运行的效率。EfficientDet<sup>[24]</sup>改进了特征金字塔网络的结构,提出了一种加权的双向金字塔网络以高效融合多尺度特征。受 EfficientNet<sup>[27]</sup>启发, EfficientDet 设计了将模型深度、宽度、图像分

辨率等进行复合缩放的方法。从 EfficientDet-D0 到 EfficientDet-D7,检测器的速度逐渐变慢,但精度逐渐提升。在与 YOLOv3 达到相同精度时, EfficientDet-D0 的运行时间仅为 YOLOv3 的 1/3 左右。

YOLOv4<sup>[23]</sup>相较于 YOLOv3,增加了很多训练技巧,如马赛克数据增强、自对抗训练,并改进了空间注意力机制,融合了特征金字塔<sup>[28]</sup>与路径聚合网络<sup>[29]</sup>等。多种先进算法的集成使 YOLOv4 在相同检测速度时,检测精度比 YOLOv3 提高了约 10%,在相同检测精度时,检测速度是 EfficientDet 的 2 倍左右。为了满足实际工程部署的需要, YOLOv4 原作者团队也对模型缩放技术进行了研究,提出了一套模型——Scaled-YOLOv4<sup>[30]</sup>,进一步实现了检测



速度与检测精度的平衡, 如其中的 YOLOv4-P5 达到了与 EfficientDet-D5 相同的精度, 而检测速度是 EfficientDet-D5 的 3 倍。

在 YOLOv4 发布之后不到两个月, Ultralytics LLC 公司在 Github 上发布了 YOLOv5 的开源项目。YOLOv5 除了沿用了 YOLOv4 中的马赛克数据增强方法之外, 还在算法中嵌入了自适应锚框计算功能。截至目前, YOLOv5 已经更新到了 V6.0 版本, 既减小了模型尺寸又提高了目标检测的准确率与推理速度, 更适合用于工业项目落地应用。但以上 YOLO 系列算法采用的锚框设计增加了检测器的计算负荷。基于此, 旷视科技结合头部解耦、无锚框设计、高级标签分配等目标检测先进技术对 YOLO 算法进行改进, 提出了无锚框的高性能检测器——YOLOX<sup>[31]</sup>。将检测头进行解耦, 可以在仅增加一点参数数量的情况下, 大大提升 YOLO 系列算法的收敛速度。YOLOX 将标签分配建模成最优传输问题, 通过求近似解的方法减少了训练时间并有效避免了求解额外超参数的状况发生。

从表 1 中可以看出大部分单阶段目标检测方法比两阶段目标检测方法检测速度快。同时期的目标检测算法中, 两阶段的目标检测算法的平均检测精度较高。随着时间的发展, 单阶段目标检测算法在提高了平均检测精度的情况下仍保持了较快的检测速度。

### 1.1.2 跟踪器

在目标跟踪中, 两阶段是指检测与跟踪独立, 先检测再跟踪的方法。依据处理视频帧的模式的不同, 基于检测的跟踪又可以分为离线跟踪 (Offline Tracking) 与在线跟踪 (Online Tracking)。

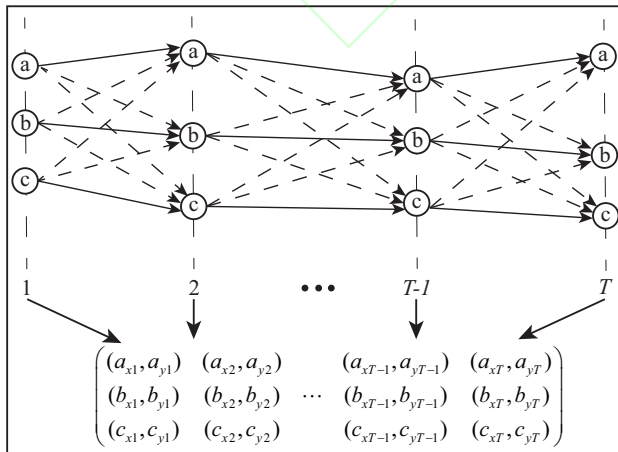


图 3 离线跟踪

离线跟踪, 又称为批量跟踪, 是一种利用全局信息来进行数据关联的跟踪方式 (如图 3 所示), 但是

受算力和内存的限制, 离线跟踪一般将视频分段分批处理。得益于批量帧处理, 离线跟踪理论上可获得全局最优跟踪结果, 然而其实时性较差, 应用场景有限。目前, 离线跟踪一般应用于运动分析、医学图像处理等场景。

MCMOT<sup>[32]</sup> 是一种基于变点检测的多类别多目标跟踪算法。该法在检测到遮挡时, 只跟踪可见目标, 并将整体轨迹分成几个轨迹段。若关联目标存在歧义, 则终止轨迹段; 若相同目标再次出现, 则在终止的跟踪点附近重新开始跟踪。结合了层次数据模型的 MCMOT\_HDM 离线跟踪方法, 实现了跟踪速度与跟踪准确率的平衡。为了进一步提升跟踪的准确率, 部分学者专注于数据关联做了一些研究。KDNT<sup>[33]</sup> 构造了基于外观、运动和平滑的密集近邻 (Dense Neighbors, DN) 关联矩阵, 将跟踪视频分为多个不相交的片段, 再用 DN 搜索每个片段中的小轨迹, 以实现小轨迹间的关联。LMP<sup>[34]</sup> 将多人跟踪抽象为一个最小成本提升多割问题, 通过对图的分解来确定其可行解。针对长期遮挡的问题, GCRA<sup>[35]</sup> 提出基于深度孪生双向门控循环单元的轨迹重连接方法, 将隶属于同一目标的子轨迹连接起来, 以形成一个完整的轨迹。不相交路径问题, 是一种经典的组合优化问题。为了解决图中不相交路径的表达能力有限的问题, Lif\_T<sup>[36]</sup> 引入额外的提升边 (Lifted edges) 来提供路径连通先验, 并从多面体的角度研究优化问题, 提出了一种多面体松弛 (Polyhedral relaxation), 以从全局求解多目标跟踪问题。

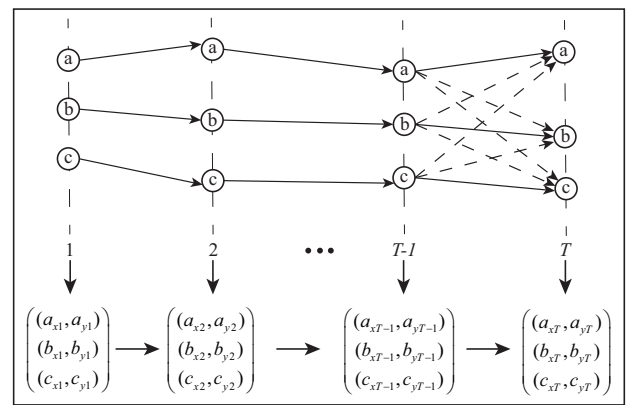


图 4 在线跟踪

在线跟踪, 又称为顺序跟踪, 是一种逐帧渐进的跟踪方式 (如图 4 所示), 在处理当前帧时仅利用此帧与先前帧的信息进行数据关联。该法实时性强, 但是易受目标遮挡和检测器漏检、误检的影响。在线跟踪的特点使其广泛应用于自动驾驶、智能视频监控等领域。SORT<sup>[37]</sup> 以目标之间的 IoU 作为衡量两

个目标之间距离的标量, 并采用匈牙利算法构建含 IoU 的关联矩阵。DeepSORT<sup>[38]</sup> 在 SORT 的框架基础上, 增加了外观信息提升了算法性能, 以实现存在较长时间遮挡的目标跟踪。SORT 和 DeepSORT 是目前在工业界比较流行的在线跟踪算法。POI<sup>[33]</sup> 采用卡尔曼滤波进行轨迹预测, 并使用 Kuhn-Munkres 算法将轨迹跟新的检测结果相关联。EAMTT<sup>[39]</sup> 在概率假设密度粒子滤波框架中设计了强、弱检测器, 其中强检测器具有较高的置信度, 用于初始化和跟踪; 弱检测的置信度较低, 用于处理漏检。HOGM<sup>[40]</sup> 将

多目标跟踪归结为一个高阶图匹配问题, 用一个代表关联元组之间关系的张量来定义多线性目标函数。CNNMTT<sup>[41]</sup> 观察到许多行人倾向于集体行走, 同一集体的人彼此靠得很近, 且具有速度相关性, 而不同集体之间存在排斥, 倾向于保持一定的空间, 利用这种社会先验知识, 可以提升跟踪算法的性能。

部分基于检测的多目标跟踪方法在 MOT16 数据集上的跟踪时间、整体框架预估速度与准确率 (Multi-object tracking accuracy, MOTA) 如表 2 所示。

表 2 基于检测的多目标跟踪方法在 MOT16 数据集上的运行速度与准确率

模式	算法	检测器	时间/ms	FPS	MOTA/%
Offline	MCMOT_HDM <sup>[32]</sup>	Private	28.7	<11.4	62.4
	KDNT <sup>[33]</sup>	Private	1429	<0.7	68.2
	LMP_p <sup>[26]</sup>	Private	2000	<0.5	71.0
	LMP <sup>[34]</sup>	Public	2000	<0.5	48.8
	GCRA <sup>[35]</sup>	Private	357.1	<2.4	48.2
	Lif_T <sup>[36]</sup>	Public	2000	<0.5	61.3
Online	SORT <sup>[37]</sup>	Private	16.8	13.2	59.8
	POI <sup>[33]</sup>	Private	101	<6.9	66.1
	EAMTT <sup>[39]</sup>	Private	82.0	<7.1	52.5
	EAMTT_pub <sup>[39]</sup>	Public	84.7	<7.0	38.8
	DeepSORT <sup>[38]</sup>	Private	57.5	<8.6	61.4
	HOGM <sup>[40]</sup>	Private	54.9	<8.8	64.8
	CNNMTT <sup>[41]</sup>	Private	89.3	<6.7	65.2

数据来源于 MOT Challenge 网页及论文。

表 2 中的 FPS 是指帧/秒, 是由跟踪时间与检测时间叠加所推算出来的。推算过程中的目标检测时间统一采用的是 MOT Challenge 中较为流行的检测器 Faster R-CNN 的最快运行时间 (59ms)。从表 2 中可以看出, 在线跟踪的速度普遍比较快。改善同一跟踪方法的检测器, 能使跟踪性能得到较大提升, 如 LMP 和 EAMTT 在公开检测器下的性能不如使用私人检测器。离线跟踪可以结合过去和未来的帧来进行数据关联, 利用了比在线跟踪更多的信息, 因此准确率比较高。如 KDNT 是 POI 的离线版本, 其采用离线训练的方法, 使得跟踪精度高于 POI。

## 1.2 联合检测与跟踪算法

基于检测的跟踪虽然跟踪性能处于领先水平, 但是模型比较复杂, 计算代价较高, 难以均衡算法精度与跟踪速度。与基于检测的跟踪的两阶段模式不同, 联合检测与跟踪 (Joint Detection and Tracking, JDT) 是多个模块联合学习的单阶段跟踪算法 (如图 5)。JDT 减轻了框架的复杂度, 满足了车辆跟踪的实时性要求。近些年出现的 JDT 算法在 MOT Challenge 中的不同数据集上的运行速度与算法精度如表 3 所示。

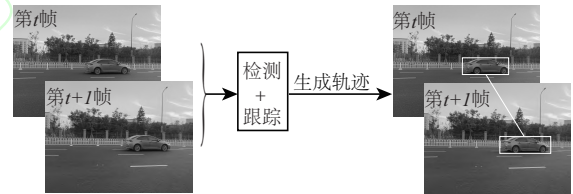


图 5 联合检测与跟踪算法

JDE<sup>[42]</sup> 框架将目标检测与外观模型嵌入到共享网络中进行学习, 提高了目标跟踪算法的运行速度, 且准确率高于 DeepSort。Zhang 等<sup>[43]</sup> 研究发现使用锚框可能给网络造成歧义, 导致跟踪效果下降, 因此在 FairMOT 中摒弃了锚框, 进一步提高了运行速度与跟踪精度。CenterTrack<sup>[44]</sup> 是基于对象中心点的联合检测与跟踪方法, 简化了跟踪框架的复杂性。该法采用 CenterNet<sup>[45]</sup> 定位对象中心, 并以先前小段轨迹的帧的热图作为先验, 对相邻两帧进行检测, 输出当前物体中心点与其前一帧检测中心点之间的距离。

Chained-Tracker<sup>[46]</sup> 首次将目标检测、特征提取、数据关联集成到一个端到端的解决方案, 但跟踪部分耗时较以往的算法更长。Sun 等<sup>[47]</sup> 利用 transformer 和注意力中的查询-键机制提出一种在线联合检测与跟踪方法——TransTrack, 在 MOT-16 数

数据集上能够达到 74.5% 的跟踪准确率。TransTrack 的运行速度在输入图像的短边分辨率为 800 像素时,运行速度约为 10FPS。TraDeS<sup>[48]</sup> 利用跟踪线索来辅助检测,通过代价容量函数推断目标跟踪偏移量,以改进当前的目标检测。TubeTK<sup>[49]</sup> 引入了“边界管道”来指示目标的时空位置,具有一定的抗遮挡能力。CorrTracker<sup>[50]</sup> 利用局部相关模块来构建目标与其周围环境之间的拓扑关系,增强了模型在拥挤场景中的识别能力。

表 3 联合检测与跟踪算法在不同测试集上的速度和准确率

算法	数据集	FPS	MOTA / %
JDE <sup>[42]</sup>	MOT16	22.2	64.4
FairMOT <sup>[43]</sup>	MOT16	25.9	68.7
FairMOT <sup>[43]</sup>	MOT17	25.9	67.5
FairMOT <sup>[43]</sup>	MOT20	13.2	58.7
CenterTrack <sup>[36]</sup>	MOT16	17.5	69.6
Chained-Tracker <sup>[46]</sup>	MOT16	6.8	67.6
Chained-Tracker <sup>[46]</sup>	MOT17	6.8	66.6
TransTrack <sup>[47]</sup>	MOT16	10	74.5
TransTrack <sup>[47]</sup>	MOT17	10	64.5
TraDeS <sup>[48]</sup>	MOT16	15	70.1
TraDeS <sup>[48]</sup>	MOT17	15	69.1
TubeTK <sup>[49]</sup>	MOT15	5.8	58.4
TubeTK <sup>[49]</sup>	MOT16	1.0	66.9
TubeTK <sup>[49]</sup>	MOT17	1.0	63.0
CorrTracker <sup>[50]</sup>	MOT16	15.9	76.6
CorrTracker <sup>[50]</sup>	MOT17	14.8	76.5
CorrTracker <sup>[50]</sup>	MOT20	8.5	65.2

基于私人检测器。

从表 2 和表 3 可以发现,大部分联合检测与跟踪方法的跟踪速度要快于基于检测的跟踪。这是因为基于检测的跟踪框架将检测与跟踪看作两个独立的部分学习,计算量比较大。而联合检测与跟踪算法是多任务学习,降低了框架的复杂度,减少了计算代价。

### 1.3 基于单目标跟踪器的多目标跟踪算法

单目标跟踪器 (Single Object Tracker, SOT), 在过去的几十年中取得了长足的发展。虽然单目标跟踪器仅能得到视频中一个物体的跟踪轨迹,但其与多目标跟踪任务有很强的联系。事实上,一个多目标跟踪任务确实可以用多个单目标跟踪器来实现。如图 6 所示,基于单目标跟踪器的多目标跟踪算法首先对视频序列进行目标检测,其次对所有目标进行单目标跟踪以获得表观特征和运动信息等新的特征,然后再进行多目标跟踪从而生成跟踪轨迹。

目前利用单目标跟踪器的多目标跟踪算法较少,表 4 汇总了部分基于单目标跟踪器的多目标跟踪算法在 MOT Challenge 中基于公开检测赛道的不同数据集上的运行速度与算法精度。

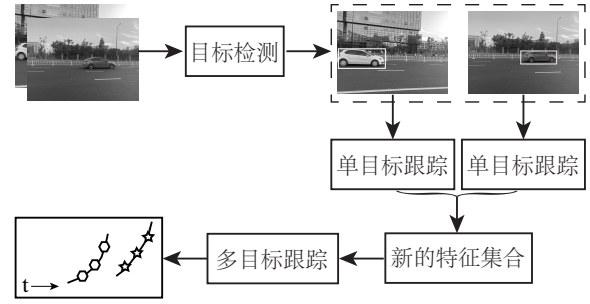


图 6 基于单目标跟踪器的多目标跟踪算法

DMAN<sup>[51]</sup> 将单目标跟踪和数据关联的优点集成在一个统一的框架中,并提出了具有时空注意机制的双匹配注意网络以适应对多目标跟踪的遮挡和噪声检测。Chu 等<sup>[52]</sup> 将基于 CNN 的单目标跟踪器应用于 MOT,通过在多个目标之间共享计算来解决计算复杂度问题,并引入时空注意力机制处理遮挡和模板间相互作用引起的漂移问题。但该模型定位输出与检测模块相互独立,每个单目标跟踪器需要处理大量背景,做了许多重复的计算,计算开销比较大。针对这个问题,Yin 等<sup>[53]</sup> 将物体运动模型与关联模型集成到一个网络,设计了一种面向在线的多目标跟踪算法 UMA。该法简化了训练过程,只需要一个训练阶段,提高了计算效率。即便如此,该法的运行速度仅为 5FPS,难以满足车辆跟踪实时性的要求。FAMNet<sup>[54]</sup> 是以孪生网络为单目标跟踪器的多目标跟踪方法。该法将多目标跟踪视为多维分配问题,融合了特征子网络和相似性度量子网络进行全局优化,获得最优分配。Feng<sup>[55]</sup> 等充分利用长期、短期的线索来处理多目标跟踪中的复杂案例,并提出了切换感知分类器以防止同一目标的 ID 发生变化。

为了提高 SOT 类 MOT 算法的跟踪速度,Zheng 等<sup>[56]</sup> 基于可微且具有封闭解的判别式单目标跟踪模型 DCFST<sup>[57]</sup>,利用 CenterNet<sup>[45]</sup> 作为检测器,提出了 SOTMOT 算法。该法跟踪速度较快,能够同时跟踪几十个目标。

表 4 基于单目标跟踪器的多目标跟踪算法在不同测试集上的速度和准确率

算法	数据集	时间 / ms	FPS	MOTA / %
STAM <sup>[52]</sup>	MOT15	2000	<0.5	34.3
STAM <sup>[52]</sup>	MOT16	5000	<0.2	46.0
DMAN <sup>[51]</sup>	MOT16	3333	<0.3	46.1
DMAN <sup>[51]</sup>	MOT17	3333	<0.3	48.2
FAMNet <sup>[54]</sup>	MOT17	1667	<0.6	52.0
LSST <sup>[55]</sup>	MOT16	500	<2.0	49.2
LSST <sup>[55]</sup>	MOT17	667	<1.5	54.7
UMA <sup>[53]</sup>	MOT16	200	<5.0	50.5
UMA <sup>[53]</sup>	MOT17	200	<5.0	53.1
SOTMOT <sup>[56]</sup>	MOT17	62.5	<16.0	62.8

基于公开检测器。



从表2、表3、表4,不难看出,基于单目标跟踪器的多目标跟踪方法仍在起步研究阶段,其运行速度与准确率仍有待提高。其中速度最快的是SOTMOT算法,能够在1秒钟内跟踪8帧左右的图像。若采用私人检测器,SOTMOT在MOT17测试集上的跟踪速度可以达到16FPS。基于SOT的MOT算法采用了性能强大的单目标跟踪模型,对于跟踪目标数量较少时具有一定的鲁棒性。但是基于单目标跟踪器的多目标跟踪方法是针对每个目标进行一次跟踪,随着跟踪目标数量增加,计算增长的消耗比较大。如何平衡跟踪精度与跟踪速度,提高多目标跟踪的效率,是基于单目标跟踪器的多目标跟踪算法所要解决的主要问题之一。

## 2 多目标跟踪算法在智能交通监控场景中的应用

智能交通监控系统是指通过在城市各个节点部署高清网络摄像机和光纤传输模块以采集交通信息,从而构建涵盖道路流量监控、车辆运行状态监控的综合分析与科学决策系统。该系统对于目标跟踪的需求主要包括:

(1) 快速响应突发事件:系统应及时反应流量监控情况、发出形势异常警告,促进减少偶发性交通拥堵状况,便于智能调度或应急救援,为管理决策提供科学依据。

(2) 跟踪具有可靠性:智能交通监控系统需要目标跟踪为公众、企业和管理部门提供准确的信息,如交通流信息、车辆定位信息、交通事件信息、逃逸车辆行驶轨迹等。

(3) 促进合作,信息共享:智能交通系统要求将跟踪数据有效整合与分析,提高智能视频监控覆盖面,建立视频监控统一管理平台,使交通监控与城市安防等业务相融合。

多目标跟踪在智能交通监控中有许多应用场景,如车流量检测、跨摄像头车辆跟踪、车辆异常行为检测与跟踪等。

### 2.1 车流量检测

车流量检测有助于对交通做出诱导控制,缓解交通拥堵状况,提升交通流效率和车辆吞吐量<sup>[58]</sup>,提高交通信号灯的使用效率<sup>[59]</sup>。获取车辆事故发生时的交通速度和密度也有利于事故预测<sup>[60]</sup>。常见的车流量检测技术有视频检测<sup>[61]</sup>、环形线圈检测<sup>[62]</sup>、无线传感器检测<sup>[63]</sup>、超声波检测<sup>[64]</sup>等。其中基于视频的车辆检测技术近年来发展迅速,该法融合了视频图像处理与计算机视觉技术,通过对车辆的识别、跟

踪获得交通流参数。目前基于视频的检测方法主要应用在道路条件复杂的地段,如城市道路的十字路口、高速公路立交、隧道等。

陈佳倩等<sup>[65]</sup>提出一种基于YOLOv3和DeepSort的车流量检测方法,实现了车辆视频实时监控与跟踪计数。此法是一种基于检测的跟踪方法,其使用YOLOv3训练车辆检测模型,再将高精度的车辆检测结果作为DeepSort跟踪算法的输入,从而弥补了DeepSort本身假阳率高的缺陷,提高了车流量检测的精度。百度公司和中山大学合作,提出了一种在线运动车辆计数系统<sup>[66]</sup>,以实现对接挤的十字路口进行交通流分析。该系统以PP-YOLO<sup>[67]</sup>作为车辆检测器,结合DeepSort算法进行多目标跟踪,采用了基于形状的运动分配策略和精心设计的空间约束提高了检测效果。

车流量检测技术的发展与交叉路口智能信号引导系统相结合,能够帮助合理调整信号灯的时长,从而缓解交叉路口的交通流矛盾。

### 2.2 跨摄像头车辆跟踪

近年来,在建设智慧城市的热潮下,多目标跨摄像头跟踪(Multi-Target Multi-Camera Tracking, MTMCT)引起了研究人员和政府机构的广泛关注。跨摄像头车辆跟踪是MTMCT的子任务,可以更好地助力于交通设计和交通流优化,因此在现代交通管理系统中扮演着重要的角色。一般来说,MTMCT可以分解为单摄像头多目标跟踪、车辆重识别、跨摄像头车辆轨迹关联等子任务。车辆重识别是其中的重要一环,旨在构建一个跨摄像头的车辆图像检索模型<sup>[68]</sup>,能够实现对目标车辆的快速查找、跟踪与定位。

受道路结构和交通规则限制,车辆在道路上的行驶轨迹通常有迹可循。Qian等<sup>[69]</sup>提出了一种名为ELECTRICITY的跨摄像头车辆跟踪系统。该系统采用Mask R-CNN作为帧级车辆检测器,联合DeepSort与JDE算法中的数据关联方法实现单摄像头多目标跟踪,将各单摄像头多目标跟踪获得的小段轨迹用聚合损失训练得到的模型进行重识别,并利用几何信息实现跨摄像头多视角跟踪结果的同步,加速跟踪过程。为了解决数据量不足、合成数据与真实数据存在偏差的问题,Zheng等<sup>[70]</sup>分析了制约车辆重识别性能的瓶颈,设计了一种高效的车辆重识别方法。该法提出了新颖的数据增强策略,即采用DG-Net<sup>[71]</sup>把不同结构和外观的车辆合成新的样本,使用生成对抗网络对合成车辆图像进行风格迁移,令生成的



图像更接近真实的图像, 并通过将真实车辆图像粘贴到合成的背景中以增强数据。

在多目标跟踪任务中, 由于车辆具有相似的外观, 所以简单的重识别模型可能无法区分外观相似的不同车辆。此外, 即使是同一辆车, 也有可能发生较大的外观变化。针对这两个问题, Liu 等<sup>[72]</sup>提出了基于方向的时间掩膜以限制匹配空间和相邻摄像头的子聚类方法以匹配相邻摄像头中的小段轨迹。目前, 较多研究仍然集中在行人重识别上, 有关车辆重识别的研究相对较少。现阶段的一些联合检测与重识别的任务是相互独立的, 导致模型的内存和运行时长加倍。Munjaj 等<sup>[73]</sup>提出了一种新的多任务损失函数, 将车辆检测、跟踪、重识别联合优化, 构建了一个端到端的跨帧跟踪匹配框架, 降低了内存和运行时长的需求。该法虽然是基于单摄像头跟踪提出的, 但对于跨摄像头跟踪有一定的借鉴意义。

跨摄像头车辆跟踪算法除了有助于优化交通流设计外, 还有利于快速侦破交通事故逃逸事件, 减少视频分析的人力成本, 提高侦查效率。

### 2.3 车辆异常行为检测与跟踪

随着车流密度的逐渐增大, 车辆异常行为检测与跟踪对改善交通状况、降低交通事故发生率起着越来越重要的作用。

为了克服数据稀缺、场景多变等困难, Zhao 等<sup>[74]</sup>对基于无监督学习的交通异常检测进行了研究, 提出了一种基于车辆轨迹的框架, 并利用单目标跟踪器 (SASiamRPN<sup>[75]</sup>) 来提高多目标跟踪算法的性能, 降低漏检率。同时, 他们提出了一种自适应无监督道路掩膜生成方法, 以使车辆异常行为检测框架专注于道路区域的检测。通过分析不同的车辆监控视频, Li 等<sup>[76]</sup>得出当发生事故时, 相关车辆会突然停车并持续一段时间的结论, 并提出一个假设, 即如果车辆停止时间超出交通信号灯持续时间即为有可能发生交通事故。据此, 他们设计了一种联合帧差与车辆跟踪轨迹的方法来提取假设异常区域的掩膜, 提出了包含检测框跟踪和像素级回溯的多粒度跟踪算法。为了提高检测框跟踪的准确性, Wu 等<sup>[77]</sup>设计了新的检测框改进方案, 在方案中融入了空间融合、静态滤波、时间融合和前馈优化方法。针对复杂场景的交通异常行为, 例如车辆异常移动, Chen 等<sup>[78]</sup>首次提出了双模态双边轨迹跟踪方法, 以定位车祸发生的时间。

车辆异常行为检测与跟踪算法能够在监控视频中根据车辆行驶轨迹的突变对发生的各种交通违章

与交通事故事件采取及时的应对措施, 保障公路安全、高效运行。

## 3 多目标跟踪算法在智能交通监控场景中面临的挑战

尽管多目标跟踪技术已经取得不错的进展, 但仍难以切实满足智能交通场景的实际需求。如何提高多目标跟踪的性能, 应对目标跟踪中的挑战, 是计算机视觉研究者们目前所要攻克的难点。具体来说, 这些挑战主要包括:

(1) 遮挡与再现: 在车辆行驶的过程中, 不止车辆之间因为超车或变道会发生彼此遮挡的现象, 路边的行道树也有可能对车辆产生遮挡效果。在拥挤的车流中, 频繁的遮挡会导致目标跟踪丢失部分信息, 从而降低跟踪性能。针对跟踪中的遮挡与再现问题, 有的模型采用背景建模<sup>[78]</sup>、约束混合序列蒙特卡罗 (CMSMC) 方法<sup>[79]</sup>、基于 3D 数据深度排序的匹配算法<sup>[80]</sup>进行了一些尝试, 取得了良好的效果, 但在车辆离摄像头较远的情况下性能较差。

(2) 视场角度: 在跨摄像头跟踪任务中, 由于摄像头布置时视场角度不一致, 在视频中, 车辆的外观可能会发生变化。而车辆外观是跟踪目标的重要特征, 一旦外观信息丢失, 会对目标跟踪产生很大的困扰。一些算法采用聚合损失<sup>[69]</sup>、相邻镜头子聚类<sup>[72]</sup>、感兴趣区域回溯<sup>[78]</sup>等方法挖掘视频帧间的信息, 来应对视场角度变化, 但仍依赖于数据驱动。

(3) 天气影响: 在暴雨、浓雾、大风等极端恶劣的天气条件下, 目标跟踪算法的速度与精度都会受到较大影响。暴雨形成的雨水条纹遮挡了部分视频图像背景外观。附着在镜头前的雨珠也会影响成像的效果。雾是由水汽凝结物、细小灰尘颗粒等组成的天气现象。由于悬浮颗粒对光的吸收和折射作用, 雾天能见度低, 背景光强度大于目标光, 图像对比度与清晰度下降。风是空气流动引起的自然现象。当风力强劲时, 户外摄像头不可避免地会抖动。恶劣的天气条件很大程度上会造成误跟踪、漏跟踪等问题。一些视频去雾<sup>[81-83]</sup>、视频去雨<sup>[84, 85]</sup>、视频去抖动<sup>[86]</sup>算法虽然有利于视频图像增强但是同时保证精度和实时性。在跟踪算法中嵌入此类图像增强算法时, 可能会降低跟踪速度。

(4) 尺度变化: 随着车辆与摄像头距离的不同, 摄像头获取的视频中, 车辆的尺度也会发生相应的变化。如何提高尺度自适应性是目标跟踪算法在智能交通场景落地所要研究的关键点之一。

(5) 公开可用数据稀缺: 虽然城市中摄像头很多,

但是公开的带标注数据集较少。目前车辆跟踪的数据集绝大部分采集的是相对简单的道路场景视频和图像。获取车流量、人流量密集,交通情况复杂,天气条件恶劣的交通场景数据集更有利于解决车辆跟踪的实际应用问题。一些工作探索了不同方法来合成图像以进行数据增强<sup>[70, 71, 80]</sup>,而另一些工作则着重于无监督学习<sup>[63]</sup>以即时利用无标注的数据集,但无法消除数据中正负样本不平衡所带来的影响。

(6) 模型的鲁棒性:人类智能具有高度的鲁棒性和多功能性,而人工智能是通过学习来模拟人类在某些任务上的认知能力。深度学习技术的特点是需要大量高质量训练数据和计算资源来学习模型的一系列参数。因此模型的性能往往存在着较大的局限性,只能适应与已知数据非常接近的新情况,泛化能力较差。

(7) 模型的可解释性:目前基于深度学习的模型的可解释性不足,分析、决策过程被当作一个“黑盒”,也缺乏可交互性和操作性。当数据中出现随机扰动,就会使模型的输出结果发生较大的变化,其行为和表现难以合理解释。科学的决策离不开对目标的跟踪、对视觉场景的理解。如何有效提取交通对象与交通环境的交互行为信息,在建模时整合专家知识来充分理解交通行为和交通规律以定义决策规则,构建人、车、路、环境协同关联分析模型是当前智能交通监控系统亟需解决的问题之一。

## 4 结语

本文将近年来的视觉多目标跟踪算法分为基于检测的跟踪、联合检测跟踪与基于单目标跟踪器的多目标跟踪三大类算法进行分析,结合跟踪速度与跟踪准确率可以看出,基于检测的跟踪方法由于检测与跟踪模块分离,模型框架较为复杂,计算代价较大,检测与跟踪总体花费的时间比联合检测跟踪的时间长;联合检测跟踪算法由于优化了目标跟踪的框架,跟踪速度和检测准确度均有所提高;基于单目标跟踪器的多目标跟踪方法是近几年来比较新的研究方向,大部分算法跟踪速度与检测准确度仍不如前面两类多目标跟踪算法。

视觉多目标跟踪算法在智能交通场景,如车流量检测、跨摄像头追踪、车辆异常行为检测与跟踪等领域均有较好的辅助作用。但由于跟踪目标相互遮挡、视场角度不一致、极端天气影响、跟踪目标尺度变化、公开可用数据集稀缺、模型泛化能力不足、算法可解释性较差,多目标跟踪算法的跟踪仍面临着一些挑战。通过数据增强、迁移学习、多模态融

合、增强上下文特征分析,在建模过程中引入专家知识,令人工智能学习专家的决策标准,从而提高算法的普适性与鲁棒性,能使多目标跟踪算法更好地应用于智能交通监控场景。

## 参考文献 (References)

- [1] 郭戈, 许阳光, 徐涛, 等. 网联共享车路协同智能交通系统综述 [J]. 控制与决策, 2019, 34(11): 2375-2389.  
(Guo G, Xu Y G, Xu T, et al. A survey of connected shared vehicle-road cooperative intelligent transportation systems[J]. Control and Decision, 2019, 34(11): 2375-2389.)
- [2] 郭戈, 徐涛, 韩英华, 等. 电动汽车时代的电网-交通网协同优化综述 [J]. 控制与决策, 2021, 36(9): 2049-2062.  
(Guo G, Xu T, Han Y H, et al. A survey of cooperative optimization of traffic-grid networks in the era of electric vehicles[J]. Control and Decision, 2021, 36(9): 2049-2062.)
- [3] Guo G, Li P, Hao L Y. Adaptive fault-tolerant control of platoons with guaranteed traffic flow stability[J]. IEEE Transactions on Vehicular Technology, 2020, 69(7): 6916-6927.
- [4] Guo G, Wang Q. Fuel-efficient en route speed planning and tracking control of truck platoons[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 20(8): 3091-3103.
- [5] 杨红红, 曲仕茹. 基于压缩感知尺度自适应的多示例交通目标跟踪算法 [J]. 中国公路学报, 2018, 31(06): 281-290+316.  
(Yang H H, Qu S R. Traffic target tracking algorithm based on scale adaptive multiple instance learning with compressive sensing[J]. China Journal of Highway and Transport, 2018, 31(06): 281-290+316.)
- [6] 赖见辉, 王扬, 罗甜甜, 等. 基于 YOLO\_V3 的侧视视频交通流量统计方法与验证 [J]. 公路交通科技, 2021, 38(01): 135-142.  
(Lai J H, Wang Y, Luo T T, et al. A YOLO\_V3-based road-side video traffic volume counting method and verification[J]. Journal of Highway and Transportation Research and Development, 2021, 38(01): 135-142.)
- [7] 李震霄, 孙伟, 刘明明, 等. 交通监控场景中的车辆检测与跟踪算法研究 [J]. 计算机工程与应用, 2021, 57(08): 103-111.  
(Li Z X, Sun W, Liu M M, et al. Research on vehicle detection and tracking algorithms in traffic monitoring scenes[J]. Computer Engineering and Applications, 2021, 57(08): 103-111.)
- [8] 王旭辰, 韩煜祺, 唐林波, 等. 基于深度学习的无人机载平台多目标检测和跟踪算法研究 [J/OL]. 信号处理, (2021-05-12)[2021-09-27]. <http://kns.cnki.net/kcms/detail/11.2406.TN.20210512.0841.006.html>.  
(Wang X C, Han Y Q, Tang L B, et al. Multi target detection and tracking algorithm for

- UAV platform based on deep learning[J/OL]. Journal of Signal Processing, 2021[2021-09-27]. <http://kns.cnki.net/kcms/detail/11.2406.TN.20210512.0841.006.html>
- [9] 毕思凯, 张宁, 吕文喜. 水面无人艇自主航线偏航路径跟踪控制方法[J]. 舰船科学技术, 2021, 43(14): 79-81. (Bi S K, Zhang N, Lv W X. Yaw path tracking control method for autonomous route of surface unmanned boat vehicle[J]. Ship Science and Technology, 2021, 43(14): 79-81.)
  - [10] L W H, Xing J L, Milan A, et al. Multiple object tracking: A literature review[J]. Artificial Intelligence, 2020: 103448.
  - [11] Sun Z H, Chen J, Chao L, et al. A survey of multiple pedestrian tracking based on tracking-by-detection framework[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 31(5): 1819-1833.
  - [12] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Proc of the 2014 IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2014: 580-587.
  - [13] Girshick R. Fast r-cnn[C]. Proc of the 2015 IEEE Int Conf on Computer Vision (ICCV). Piscataway, NJ: IEEE, 2015: 1440-1448.
  - [14] Ren S Q, He K M, Girshick R, et al. Faster r-cnn: towards real-time object detection with region proposal networks[J]. Advances in Neural Information Processing Systems, 2015, 28: 91-99.
  - [15] He K M, Gkioxari G, Dollár P, et al. Mask r-cnn[C]. Proc of the 2017 IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2017: 2961-2969.
  - [16] He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
  - [17] Cai Z W, Vasconcelos N. Cascade r-cnn: Delving into high quality object detection[C]. Proc of the 2018 IEEE Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2018: 6154-6162.
  - [18] Pang J M, Chen K, Shi J P, et al. Libra r-cnn: Towards balanced learning for object detection[C]. Proc of the 2019 IEEE Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2019: 821-830.
  - [19] Zhang H Y, Wang Y, Dayoub F, et al. Varifocalnet: An iou-aware dense object detector[C]. Proc of the 2021 IEEE Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2021: 8514-8523.
  - [20] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]. Proc of the 2016 IEEE/CVF Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2016: 779-788.
  - [21] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]. Proc of the 2017 IEEE Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2017: 7263-7271.
  - [22] Redmon J, Farhadi A. Yolov3: an incremental improvement[J]. arXiv preprint, arXiv:1804.02767, 2018.
  - [23] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint, arXiv:2004.10934, 2020.
  - [24] Tan M X, Pang R M, Le Q V. Efficientdet: Scalable and efficient object detection[C]. Proc of the 2020 IEEE/CVF Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2020: 10781-10790.
  - [25] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]. Proc of the 2016 IEEE Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2016: 770-778.
  - [26] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]. Computer Vision - ECCV 2016. Cham, Switzerland: Springer, 2016: 21-37.
  - [27] Tan M X, Le Q V. Efficientnet: Rethinking model scaling for convolutional neural networks[C]. Int Conf on Machine Learning (ICML). New York, NY: ACM, 2019: 6105-6114.
  - [28] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]. Proc of the 2017 IEEE Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2017: 2117-2125.
  - [29] Liu S, Qi L, Qin H F, et al. Path aggregation network for instance segmentation[C]. Proc of the 2018 IEEE Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2018: 8759-8768.
  - [30] Wang C Y, Alexey B, Liao H Y M. Scaled-yolov4: Scaling cross stage partial network[C]. Proc of the 2021 IEEE Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2021: 13029-13038.
  - [31] Ge Z, Liu S T, Wang F, et al. YoloX: Exceeding yolo series in 2021[J]. arXiv preprint arXiv:2107.08430, 2021.
  - [32] Lee B, Erdene E, Jin S G, et al. Multi-class multi-object tracking using changing point detection[C]. Computer Vision - ECCV 2016. Cham, Switzerland: Springer, 2016: 68-83.
  - [33] Yu F W, Li W B, Li Q Q, et al. Poi: multiple object tracking with high performance detection and appearance feature[C]. Computer Vision - ECCV 2016. Cham, Switzerland: Springer, 2016: 36-42.
  - [34] Tang S Y, Andriluka M, Andres B, et al. Multiple people tracking by lifted multicut and person re-identification[C]. Proc of the 2017 IEEE Conf



- on Computer Vision and Pattern Recognition(CVPR). Piscataway, NJ: IEEE, 2017: 3539-3548.
- [35] Ma C, Yang C S, Yang F, et al. Trajectory factory: tracklet cleaving and re-connection by deep siamese bi-gru for multiple object tracking[C]. 2018 IEEE Int Conf on Multimedia and Expo(ICME). Piscataway, NJ: IEEE, 2018: 1-6.
- [36] Hornakova A, Henschel R, Rosenhahn B, et al. Lifted disjoint paths with application in multiple object tracking[C]. Int Conf on Machine Learning(ICML). New York, NY: ACM, 2020: 4364-4375.
- [37] Bewley A, Ge Z Y, Ott L, et al. Simple online and realtime tracking[C]. 2016 IEEE Int Conf on image processing(ICIP). Piscataway, NJ: IEEE, 2016: 3464-3468.
- [38] Wojke N, Bewley A, Paulus D. Simple online and realtime tracking with a deep association metric[C]. 2017 IEEE Int Conf on image processing(ICIP). Piscataway, NJ: IEEE, 2017: 3645-3649.
- [39] Sanchez-Matilla R, Poiesi F, Cavallaro A. Online multi-target tracking with strong and weak detections[C]. Computer Vision - ECCV 2016. Cham, Switzerland: Springer, 2016: 84-99.
- [40] Zhou Z W, Xing J L, Zhang M D, et al. Online multi-target tracking with tensor-based high-order graph matching[C]. 2018 24th Int Conf on Pattern Recognition(ICPR). Piscataway, NJ: IEEE, 2018: 1809-1814.
- [41] Mahmoudi N, Ahadi S M, Rahmati M. Multi-target tracking using CNN-based features: CNNMTT[J]. Multimedia Tools and Applications, 2019, 78(6): 7077-7096.
- [42] Wang Z D, Zheng L, Liu Y X, et al. Towards real-time multi-object tracking[C]. Computer Vision -ECCV 2020: 16th European Conf, Glasgow, UK, August 23-28, 2020, Proc, Part XI 16, 2020: 107-122.
- [43] Zhang Y F, Wang C Y, Wang X G, et al. A simple baseline for multi-object tracking[J]. arXiv preprint, arXiv:2004.01888, 2020, 3(4): 6.
- [44] Zhou X Y, Koltun V, Krähenbühl P. Tracking objects as points[C]. Computer Vision - ECCV 2020, Cham, Switzerland: Springer, 2020: 474-490.
- [45] Zhou X Y, Wang D Q, Krähenbühl P. Objects as points[J]. arXiv preprint, arXiv:1904.07850, 2019.
- [46] Peng J L, Wang C G, Wan F B, et al. Chained-tracker: chaining paired attentive regression results for end-to-end joint multiple-object detection and tracking[C]. Computer Vision - ECCV 2020, Cham, Switzerland: Springer, 2020: 145-161.
- [47] Sun P Z, Jiang Y, Zhang R F, et al. Transtrack: multiple-object tracking with transformer[J]. arXiv preprint arXiv:2012.15460, 2020.
- [48] Wu J L, Cao J L, Song L C, et al. Track to detect and segment: an online multi-object tracker[C/OL]. Proc of the 2021 IEEE/CVF Conf on Computer Vision and Pattern Recognition(CVPR). Piscataway, NJ: IEEE, 2021[2021-09-27]. [https://openaccess.thecvf.com/content/CVPR2021/html/Wu\\_Track\\_To\\_Detect\\_and\\_Segment\\_An\\_Online\\_Multi-Object\\_Tracker\\_CVPR\\_2021\\_paper.html](https://openaccess.thecvf.com/content/CVPR2021/html/Wu_Track_To_Detect_and_Segment_An_Online_Multi-Object_Tracker_CVPR_2021_paper.html).
- [49] Pang B, Li Y Z, Zhang Y F, et al. Tubetk: adopting tubes to track multi-object in a one-step training model[C]. Proc of the 2020 IEEE/CVF Conf on Computer Vision and Pattern Recognition(CVPR). Piscataway, NJ: IEEE, 2020: 6308-6318.
- [50] Wang Q, Zheng Y, Pan P, et al. Multiple object tracking with correlation learning[C/OL]. Proc of the 2021 IEEE/CVF Conf on Computer Vision and Pattern Recognition(CVPR). Piscataway, NJ: IEEE, 2021[2021-09-27]. [https://openaccess.thecvf.com/content/CVPR2021/html/Wang\\_Multiple\\_Object\\_Tracking\\_With\\_Correlation\\_Learning\\_CVPR\\_2021\\_paper.html](https://openaccess.thecvf.com/content/CVPR2021/html/Wang_Multiple_Object_Tracking_With_Correlation_Learning_CVPR_2021_paper.html).
- [51] Zhu J, Yang H, Liu N, et al. Online multi-object tracking with dual matching attention networks[C]. Proc of the Computer Vision - ECCV 2018. Cham, Switzerland: Springer, 2018: 366-382.
- [52] Chu Q, Ouyang W L, Li H S, et al. Online multi-object tracking using CNN-based single object tracker with spatial-temporal attention mechanism[C]. Proc of the IEEE Int Conf on Computer Vision(ICCV). Piscataway, NJ: IEEE, 2017: 4836-4845.
- [53] Yin J B, Wang W G, Meng Q H, et al. A unified object motion and affinity model for online multi-object tracking[C]. Proc of the 2020 IEEE/CVF Conf on Computer Vision and Pattern Recognition(CVPR). Piscataway, NJ: IEEE, 2020: 6768-6777.
- [54] Chu P, Ling H B. Famnet: Joint learning of feature, affinity and multi-dimensional assignment for online multiple object tracking[C]. Proc of the 2019 IEEE/CVF Int Conf on Computer Vision(ICCV). Piscataway, NJ: IEEE, 2019: 6172-6181.
- [55] Feng W T, Hu Z H, Wu W, et al. Multi-object tracking with multiple cues and switcher-aware classification[J]. arXiv preprint, arXiv:1901.06129, 2019.
- [56] Zheng L Y, Tang M, Chen Y Y, et al. Improving multiple object tracking with single object tracking[C]. Proc of the 2021 IEEE/CVF Conf on Computer Vision and Pattern Recognition(CVPR). Piscataway, NJ: IEEE, 2021[2021-09-27]. [https://openaccess.thecvf.com/content/CVPR2021/html/Zheng\\_Improving\\_Multiple\\_Object\\_Tracking\\_With\\_Single\\_Object\\_Tracking\\_CVPR\\_2021\\_paper.html](https://openaccess.thecvf.com/content/CVPR2021/html/Zheng_Improving_Multiple_Object_Tracking_With_Single_Object_Tracking_CVPR_2021_paper.html).
- [57] Zheng L Y, Tang M, Chen Y Y, et al. Learning feature embeddings for discriminant model based tracking[C]. Computer Vision - ECCV 2020. Cham, Switzerland: Springer, 2020: 759-775.
- [58] Guo G, Zhang T Q. A residual spatio-temporal architecture for travel demand forecasting[J]. Transportation Research Part C: Emerging Technologies,

- 2020, 115: 102639.
- [59] Guo G, Wang Y P. An integrated MPC and deep reinforcement learning approach to trams-priority active signal control[J]. *Control Engineering Practice*, 2021, 110: 104758.
- [60] 张力天, 孔嘉漪, 樊一航, 等. 基于宏微观因素的概率级别的车辆事故预测 [J]. *计算机研究与发展*, 2021, 58(09):2052-2061.  
(Zhang L T, Kong J Y, Fan Y H, et al. Car accident prediction based on marco and micro factors in probability level[J]. *Journal of Computer Research and Development*, 2021, 58(09):2052-2061.)
- [61] Tian B, Yao Q M, Gu Y, et al. Video processing techniques for traffic flow monitoring: A survey[C]// 2011 14th Int IEEE Conf on Intelligent Transportation Systems(ITSC), 2011: 1103-1108.
- [62] 郑建湖, 郭银岁, 叶润真, 等. 基于环形线圈检测器信息的交通状态模糊识别方法 [J]. *昆明理工大学学报(理工版)*, 2009, 34(2): 71-75.  
(Zheng J H, Guo Y S, Ye R Z. Fuzzy identification method of traffic condition based on loop detector information[J]. *Journal of Kunming University of Science and Technology(Natural Sciences)*, 2009, 34(2): 71-75.)
- [63] Chao K H, Chen P Y. An intelligent traffic flow control system based on radio frequency identification and wireless sensor networks[J]. *Int journal of distributed sensor networks*, 2014, 10(5): 694545.
- [64] Jo Y, Choi J, Jung I. Traffic information acquisition system with ultrasonic sensors in wireless sensor networks[J]. *Int Journal of Distributed Sensor Networks*, 2014, 10(5): 961073.
- [65] 陈佳倩, 金晖宏, 王文远, 等. 基于 YOLOv3 和 DeepSort 的车流量检测 [J]. *计量学报*, 2021, 42(06): 718-723.  
(Chen J Q, Jin X H, Wang W Y, et al. Vehicle flow detection based on YOLOv3 and DeepSort[J]. *Acta Metrologica Sinica*, 2021, 42(06): 718-723.)
- [66] Lu J C, Xia M, Gao X, et al. Robust and online vehicle counting at crowded intersections[C]. *Proc of the 2021 IEEE/CVF Conf on Computer Vision and Pattern Recognition Workshops(CVPRW)*. Piscataway, NJ: IEEE, 2021: 4002-4008.
- [67] Long X, Deng K P, Wang G Z, et al. PP-YOLO: an effective and efficient implementation of object detector[J]. *arXiv preprint arXiv:2007.12099*, 2020.
- [68] 谢秀珍, 罗志明, 连盛, 等. 一种融合表观与属性信息的车辆重识别方法 [J]. *厦门大学学报(自然科学版)*, 2021, 60(01): 72-79+150.  
(Xie X Z, Luo Z M, Lian S, et al. A vehicle re-identification method by fusing the vehicle appearance and attribute information[J]. *Journal of Xiamen University(Natural Science)*, 2021, 60(01): 72-79+150.)
- [69] Qian Y J, Yu L J, Liu W H, et al. Electricity: an efficient multi-camera vehicle tracking system for intelligent city[C]. *Proc of the 2020 IEEE/CVF Conf on Computer Vision and Pattern Recognition Workshops(CVPRW)*, 2020: 588-589.
- [70] Zheng Z D, Jiang M Y, Wang Z G, et al. Going beyond real data: A robust visual representation for vehicle re-identification[C]. *Proc of the 2020 IEEE/CVF Conf on Computer Vision and Pattern Recognition Workshops(CVPRW)*, 2020: 598-599.
- [71] Zheng Z D, Yang X D, Yu Z D, et al. Joint discriminative and generative learning for person re-identification[C]. *Proc of the 2019 IEEE/CVF Conf on Computer Vision and Pattern Recognition(CVPR)*. Piscataway, NJ: IEEE, 2019: 2138-2147.
- [72] Liu C, Zhang Y Q, Luo H, et al. City-scale multi-camera vehicle tracking guided by crossroad zones[C]. *Proc of the 2021 IEEE/CVF Conf on Computer Vision and Pattern Recognition Workshops(CVPRW)*. Piscataway, NJ: IEEE, 2021: 4129-4137.
- [73] Munjal B, Aftab A R, Amin S, et al. Joint detection and tracking in videos with identification features[J].
- [74] Zhao J, Yi Z, Pan S, et al. Unsupervised traffic anomaly detection using trajectories[C]. *Proc of the IEEE Conf on Computer Vision and Pattern Recognition Workshops(CVPRW)*. Piscataway, NJ: IEEE, 2019: 133-140.
- [75] Zhu Z, Wang Q, Li B, et al. Distractor-aware siamese networks for visual object tracking[C]. *Computer Vision - ECCV 2018*. Cham, Switzerland: Springer, 2018: 101-117.
- [76] Li Y Y, Wu J, Bai X, et al. Multi-granularity tracking with modularized components for unsupervised vehicles anomaly detection[C]. *Proc of the 2020 IEEE/CVF Conf on Computer Vision and Pattern Recognition Workshops(CVPRW)*, 2020: 586-587.
- [77] Wu J, Wang X H, Xiao X F, et al. Box-level tube tracking and refinement for vehicles anomaly detection[C]. *Proc of the 2021 IEEE/CVF Conf on Computer Vision and Pattern Recognition Workshops(CVPRW)*. Piscataway, NJ: IEEE, 2021: 4112-4118.
- [78] Chen J Y, Ding G C, Yang Y C, et al. Dual-modality vehicle anomaly detection via bilateral trajectory tracing[C]. *Proc of the 2021 IEEE/CVF Conf on Computer Vision and Pattern Recognition Workshops(CVPRW)*. Piscataway, NJ: IEEE, 2021: 4016-4025.
- [79] Li J C, Zhan W, Hu Y P, et al. Generic tracking and probabilistic prediction framework and its application in autonomous driving[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2019, 21(9): 3634-3649.
- [80] Hu H N, Cai Q Z, Wang D Q, et al. Joint monocular 3D vehicle detection and tracking[C]. *Proc of the 2019 IEEE/CVF Int Conf on Computer Vision(ICCV)*. Piscataway, NJ: IEEE, 2019: 5390-5399.
- [81] Ren W Q, Zhang J G, Xu X Y, et al. Deep video dehazing

- with semantic segmentation[J]. IEEE Transactions on Image Processing, 2018, 28 (4): 1895-1908.
- [82] Chen Z Y, Wang Y C, Yang Y, et al. PSD: Principled Synthetic-to-Real Dehazing Guided by Physical Priors[C]. Proc of the 2021 IEEE/CVF Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2021: 7180-7189.
- [83] Zhang X Y, Dong H, Pan J S, et al. Learning To Restore Hazy Video: A New Real-World Dataset and a New Method[C]. Proc of the 2021 IEEE/CVF Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2021: 9239-9248.
- [84] Yan W D, T.Tan R, Yang W H, et al. Self-Aligned Video Deraining With Transmission-Depth Consistency[C]. Proc of the 2021 IEEE/CVF Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2021: 11966-11976.
- [85] Yue Z S, Xie J W, Zhao Q, et al. Semi-Supervised Video Deraining With Dynamical Rain Generator[C]. Proc of the 2021 IEEE/CVF Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2021: 642-652.
- [86] 尹雅楠. 视频监控图像去抖动视觉监测算法优化仿真[J]. 计算机仿真, 2017, 34 (09): 450-454.  
(Yin Y N. Video monitoring image de jitter visual monitoring algorithm optimization simulation[J]. Computer Simulation, 2017, 34 (09): 450-454.)

### 作者简介

金沙沙(1994—), 女, 硕士生, 从事计算机视觉和图像处理等研究, E-mail: christinejin@qq.com;

龙伟(1978—), 男, 讲师, 博士, 从事计算机应用、图像处理、人工智能等研究, E-mail: lw@zjhu.edu.cn;

胡灵犀(1990—), 女, 博士, 学历, 从事光子准粒子和结构物质的相互作用、光电子学太赫兹辐射源以及图像处理等研究, E-mail: 03078@zjhu.edu.cn;

王天宇(1996—), 男, 硕士生, 学历, 从事计算机图形学、VR/AR 的研究, E-mail: 18362985169@163.com;

潘华(1977—), 女, 高级工程师, 博士, 从事智能工业数据解析与工艺流程优化、数学建模、项目管理等研究, E-mail: pan\_hua\_hzzj@hotmail.com;

蒋林华(1977—), 男, 教授, 博士 (博士生导师), 从事计算机视觉和人工智能、自动驾驶等研究, E-mail: 11594@zjhu.edu.cn.