doi:10.3969/j.issn.1001-4616.2026.02.009

基于时空交互信息融合的车辆违规超车识别

巢　新[1,2],吉根林[2],赵　斌[2],麦丞程[2],王嘉琦[2,3]

(1.南京师范大学地理科学学院,江苏 南京 210023)
(2.南京师范大学计算机与电子信息学院/人工智能学院,江苏 南京 210023)
(3.南京师范大学 数学科学学院,江苏 南京 210023)

[摘要]　为提升车辆违规超车行为的识别精度,本文提出了基于时空交互信息融合的违规超车识别算法. 该算法以 TimeSformer 为主干模型,融合 RGB 图像、光流、深度图以及超车交互图四种模态信息,构建统一的超车信息图,从外观特征、运动特性、三维空间结构以及车辆间交互关系等多个维度对超车行为进行联合建模. 通过引入分离时空注意力机制以及多模态特征融合策略,有效刻画超车过程中目标车辆的动态演化特征及其与周围车辆之间的时空交互模式,从而弥补复杂交通场景下多车辆交互关系表述不足的问题. 在 PREVENTION 数据集上的实验结果表明,所提算法在违规超车识别任务中取得了 94.04%的识别准确率,较多种主流基准算法表现出更优的识别性能,验证了多模态时空交互信息融合策略在复杂交通行为识别中的有效性.

[关键词]　智能交通,车辆违规超车行为识别,时空交互信息融合,多车辆交互建模,TimeSformer
[中图分类号]U495　[文献标志码]A　[文章编号]1001-4616(2026)02-0085-13

Spatiotemporal Interaction Information Fusion for Vehicle Illegal Overtaking Recognition

Chao Xin[1,2],Ji Genlin[2],Zhao Bin[2],Mai Chengcheng[2],Wang Jiaqi[2,3]

(1.School of Geography,Nanjing Normal University,Nanjing 210023,China)
(2.School of Computer and Electronic Information/School of Artificial Intelligence,Nanjing Normal University,Nanjing 210023,China)
(3.School of Mathematical Sciences,Nanjing Normal University,Nanjing 210023,China)

Abstract:To improve the accuracy of vehicle illegal overtaking recognition,this paper proposes spatiotemporal interaction information fusion for vehicle illegal overtaking recognition algorithm. The algorithm is built upon the TimeSformer architecture as the backbone model. Four types of modality information,namely RGB images,optical flow,depth maps,and overtaking interaction graphs,are integrated to construct a unified overtaking information graph. From multiple perspectives,including appearance features,motion information,3D spatial structure,and inter-vehicle interaction relationships,the method performs joint modeling of overtaking behaviors. By introducing divided space-time attention mechanism and multi-modal feature fusion strategy,the proposed approach effectively captures the dynamic evolution of the target vehicle during the overtaking process as well as its spatiotemporal interactions with surrounding vehicles,thereby alleviating the insufficient representation of multi-vehicle interactions in complex traffic scenarios. Experimental results on the PREVENTION dataset show that the proposed algorithm achieves a recognition accuracy of 94.04% for illegal overtaking behaviors,outperforming several existing mainstream algorithms and validating the effectiveness of multimodal spatiotemporal interaction information fusion for complex traffic behavior recognition.

Key words:intelligent transportation,vehicle illegal overtaking recognition,spatiotemporal interaction information fusion,multi-vehicle interaction modeling,TimeSformer

收稿日期:2025-11-19.
基金项目:国家自然科学基金项目(41971343)、江苏省前沿技术研发计划项目(BF2024005).
通讯作者:吉根林,博士,教授,博士生导师,研究方向:大数据挖掘与人工智能. E-mail:glji@njnu.edu.cn

随着智能交通系统和地理信息技术的不断进步,道路安全问题愈发引起重视[1]. 尤其在高速公路和城市主干道上,车辆超车行为频繁,一旦驾驶者在实线、弯道区域、车流密集或视野受限的情况下违规超车,极易打破交通秩序,甚至引发严重事故. 因此,构建高效识别违规超车行为的智能交通系统,对于提升道路安全与交通管理水平具有重要意义.

图 1 展示了一个典型的超车行为过程. 具体而言,在 t0 时刻,小型车辆表现出超车意图并开始加速;至 t1 时,其向左侧车道实施变道操作,并逐步接近前方行驶的中型车辆;至 t2 时刻,小型车辆在纵向上已超越了中型车辆;随后在 t3 时刻,小型车辆继续保持较高速度,并向右侧变道,准备驶回原车道;最终于 t4 时刻,小型车辆成功并入原车道,与中型车辆之间保持相对安全的纵向距离,标志着一次合法的超车行为顺利完成. 此外,图中中型车辆在该过程中也以较高速度超过了其右侧车道上的大型车辆,并于 t4 时刻完成超越. 然而,由于二者不处于同一车道,该过程通常不被认定为标准意义上的超车行为. 需要注意的是,若小型车辆在变道过程中存在压越实线、穿越弯道区域,或在 t4 时刻与中型车辆之间的纵向距离明显不足,则该行为就构成违规超车. 因此,准确判断超车行为的合法性,需综合分析车道线类型、车辆变道轨迹与车距动态变化等多维信息.

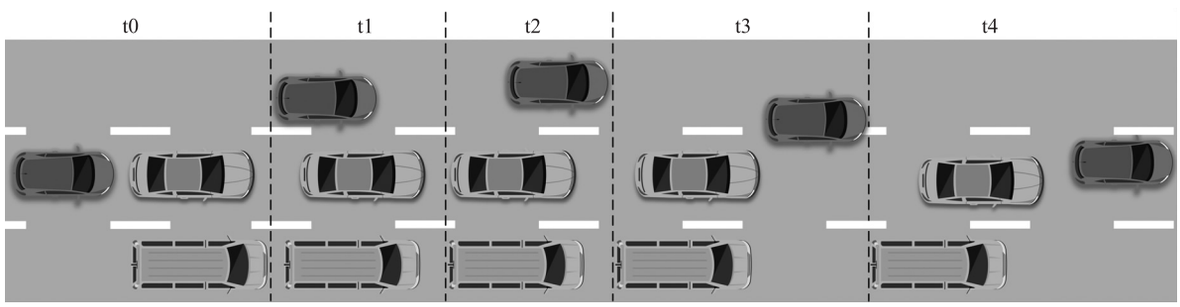

图 1　车辆超车行为示意图

Fig. 1　Schematic illustration of vehicle overtaking behavior

现有的违规超车行为识别算法多依赖于规则驱动的视觉检测[2-6]或轨迹分析[7-9],这类算法通常通过建模车辆轨迹与车道线之间的空间关系、分析车速变化与相对距离,以判断是否存在违规行为. 然而,在实际复杂道路环境中,该类算法面临诸多挑战. 首先,规则设计高度依赖人为经验,难以涵盖多样化的驾驶行为模式,适应性较差;其次,单一的视觉信息难以充分捕捉车辆动态行为中的时空模式关系,导致识别精度受限;此外,在面对遮挡、光照变化、摄像头抖动等干扰时,该类算法的鲁棒性不足,难以满足智能交通系统对高精度与高可靠性行为识别的现实需求.

随着计算机视觉和模式识别技术的迅速发展,基于数据驱动的行为识别算法逐渐成为智能交通领域的重要研究方向. 近年来,卷积神经网络[10](CNN)、循环神经网络[11](RNN)以及 Vision Transformer[12]架构广泛应用于视频分析与行为识别任务中,为复杂场景下的交通行为提供了强大的建模能力. 在车辆行为识别任务中,研究者逐步引入多模态[13-15]信息,通过融合 RGB 图像与光流特征,有效增强了对车辆运动轨迹及其与周围环境交互关系[16]的表达能力. 然而,现有算法在建模空间结构关系和处理车辆近距离动态交互方面仍存在不足,特别是在多车并行或变道等复杂场景中,违规超车行为识别精度仍有待提升.

针对复杂交通场景下车辆违规超车行为识别中多源信息利用不足与车辆交互关系刻画不充分的问题,本文提出了基于时空交互信息融合的车辆违规超车行为识别算法. 该算法从后车视角采集的行车视频中分别提取 RGB 图像序列、光流场序列、深度图序列以及超车交互图序列,并将上述模态信息进行统一组织与表示,以综合刻画超车过程中车辆的外观特征、运动状态、空间结构及交互关系. 在此基础上,将构建的超车信息图输入至 TimeSformer 网络进行时空特征建模与分类预测,实现对违规超车、右侧超车、左侧超车、左变道未超车、右变道未超车和直行六类行为的精准判别.

1　相关工作

车辆违规超车行为识别的发展得益于多模态融合算法与深度学习模型的不断进步,为相关研究提供了坚实的技术支撑与发展基础.

1.1　车辆违规超车行为识别

近年来,车辆违规超车行为识别逐渐成为智能交通领域的研究热点,相关算法主要包括基于视觉图像[2-6]、车辆通信与轨迹信息[7-9]以及多模态感知[13-15]的数据驱动算法,并取得了诸多进展. Nalcakan 等人[2]构建了基于 YOLOv4 检测、DeepSort 跟踪与 LSTM 分类的三阶段算法,用于识别视频序列中的车辆超车行为. 文献[3]提出了基于自监督学习与对比损失的超车行为识别算法,通过语义分割提取车辆与自车道的高阶表示,结合编码器预训练与分类器微调,识别前方车辆违规超车行为. 文献[4]提出了基于注意力机制的交互感知模型,利用多头注意力与前馈网络构建轨迹编码器和社交交互模块,从自车视角预测违规超车行为. Ong 等人[5]使用 YOLO 和 DeepSORT 融合车道线检测与点线距离法,判断是否存在非法超车. Marcomini 等人[6]运用 YOLOv8 与 YOLOPv2,并结合手机摄像头和地理位置信息检测连续车道的非法超车,通过时序分析确认违法行为并提取发生时刻. 文献[7]提出基于数据驱动的分布式信息物理系统,通过历史行为建模与多时延侧向控制策略设计,识别车辆违规超车行为. 文献[8]提出基于 V2V 通信的侧向碰撞预警算法,融合变道意图识别、轨迹预测与碰撞风险判别模型,用于识别和预警违规超车. 文献[9]综合驾驶意图预测、轨迹规划与有限状态机模型,实现复杂队列中的超车行为判别. 文献[13]引入了用于预测车辆超过自行车后返回时机的概率行为模型,结合视觉与行为数据预测侧向碰撞风险. 文献[14]融合横纵向控制与认知建模,识别车辆在超车并入车队过程中的违规行为. Athree 等人[15]通过识别道路标线、交通标志以及车辆间距来检测超车行为是否违规.

然而,现有算法仍存在一定的局限性. 视觉单模态算法在复杂交通环境下对遮挡与细微行为差异的适应性有限;依赖轨迹或通信信息的技术对外部设备要求较高,难以在纯视觉场景中推广;基于轨迹或局部交互的模型难以细粒度建模车辆间的空间结构与动态关系;现有多模态融合算法多为浅层叠加,缺乏统一的时空语义建模机制;此外,多数算法未能捕捉超车行为的连续动态演化过程,导致对违规行为的判别能力受限. 因此,亟需一种能够同时整合外观、运动、空间结构与车辆交互关系的时空特征建模算法,以提升复杂场景下的违规超车识别性能.

1.2　多模态融合

近年来,视频行为识别在智能交通领域中受到广泛关注,尤其是在多模态信息融合方面取得了显著进展. 传统的时空建模算法如 C3D[17]和 I3D[18]采用三维卷积神经网络对视频序列进行建模,能够有效提取时序动态特征,在一定程度上提升了行为识别的准确性. 然而,此类算法在建模多车辆行为交互关系及多模态语义协同方面仍存在明显不足. 为增强模型对多模态特征的表征能力,研究者提出了特征融合(feature fusion)与分数融合(score fusion)等策略,广泛应用于基于多流 CNN[19]和 3D ResNet[20]等主干网络的多模态行为识别模型中. 特征融合算法通常在模型浅层或中间层对来自不同模态(如 RGB 图像、光流、深度图等)的特征进行拼接或加权整合,以提升模型对语义特征的联合建模能力. 分数融合[20]算法则在各模态子网络分别输出分类概率后,采用加权平均或多数投票机制进行决策级融合. 这类算法具有良好的可扩展性,能够灵活应对多种模态组合,在行为识别中取得了一定成效. 然而,传统融合策略多基于模态独立建模,缺乏统一的跨模态建模机制,难以有效捕捉不同模态间潜在的高阶依赖关系与结构化交互特征,限制了其在复杂交通场景下的行为识别性能提升.

为应对多模态特征建模能力不足的问题,基于 Transformer 架构的视频识别算法近年来展现出显著性能优势. Vision Transformer(ViT)通过将图像划分为 patch 并结合位置编码实现特征建模,但其在视频行为识别中的时序建模能力仍有限. TimeSformer 在 ViT 的基础上引入分离时空注意力机制,在提升时空建模能力的同时有效降低了计算复杂度. 相比于传统 CNN-LSTM 算法,该模型在建模复杂行为动态演化方面展现出更强表现力. 基于此,本文进一步提出融合 RGB 图像、光流、深度图与超车交互图四种模态的结构化输入策略,通过构建统一的超车信息图,综合建模目标车辆的外观、运动、空间结构与交互特征,实现对违规超车行为的统一时空语义建模与精确识别.

2　本文算法

2.1　总体框架

如图 2 所示,本文所提的车辆违规超车行为识别框架由多模态信息提取、超车信息图构建及时空特征

提取与分类三个阶段组成. 首先,输入连续行车视频,经帧级采样后形成图像序列. 接着,分别构建视觉表观流、运动信息流、空间结构感知流及交互事件流,以全面刻画超车过程中的多维语义信息. 其中,视觉表观流由原始 RGB 图像构成,用于描述车辆的外观特征;运动信息流基于光流估计算法金字塔、扭曲与代价体网络(Pyramid,Warping,and Cost volume Network,PWC-Net)[21]提取,用于刻画车辆在时间维度上的运动变化;空间结构感知流由 Depth Anything v2[22]生成,用于反映车辆与周围环境之间的三维空间关系;交互事件流则基于 YOLO v8[23]目标检测结果,结合图像二值化处理构建,用于显式表示车辆之间的交互状态. 上述多源信息在空间尺度上进行对齐,并以 2×2 的拼接方式进行融合,形成统一的超车信息图表示,从而实现对超车行为多维感知信息的集成表达. 随后,将超车信息图输入至引入分离式时空注意力机制的 TimeSformer[24]网络中进行时空特征提取,通过联合建模时间动态与空间结构关系,实现对违规超车、右侧超车、左侧超车、左变道未超车、右变道未超车以及直行六类行为的准确识别.

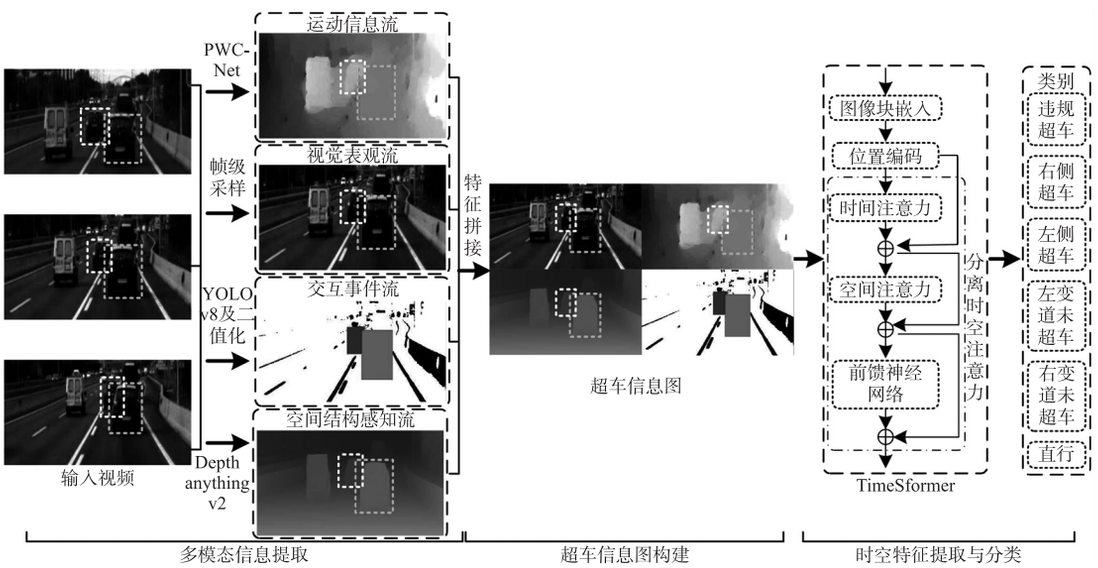

图 2　所提算法的流程图

Fig. 2　Flowchart of the proposed method

2.2　超车信息图构建

为全面捕捉车辆超车行为中的时空动态特征,本文提出超车信息图,通过融合视觉表观流、运动信息流、空间结构感知流与交互事件流,实现多模态特征的统一表征,从而综合反映目标车辆在复杂环境中的多维感知信息.

2.2.1　视觉表观流

对采集到的行车视频进行帧级采样,以提取目标车辆的 RGB 图像序列. 鉴于超车行为通常在数秒内完成,为了充分捕捉其时序特征,本文将视频采样率统一设定为固定的 FPS. 该设置在保证时序连续性的同时,能够有效保留超车过程中的关键动态信息. 随后,所提取的 RGB 图像序列作为后续运动信息流与空间结构感知流构建的基础输入.

2.2.2　运动信息流

在超车行为识别过程中,光流信息用于刻画目标车辆及其周围环境的动态运动特征. 通过分析连续图像帧之间的像素位移变化,光流能够反映车辆在时间维度上的位移趋势与速度变化,从而表征其运动轨迹及行为演化过程. 特别是在变道、加速等关键阶段,光流所包含的细粒度动态信息有助于增强模型对超车行为时序特征的感知能力. 本文采用 PWC-Net 算法提取光流信息,以刻画车辆在超车过程中的运动特征. 该算法基于多尺度特征金字塔结构,并结合特征对齐与代价体构建机制,对连续帧之间的像素位移进行估计. 在此基础上,通过逐层细化的方式获得更加精确的光流结果,从而有效应对超车过程中可能出现的较大位移变化. 得益于其多尺度建模与逐级优化策略,PWC-Net 在保证光流估计精度的同时兼顾计算效率,适用于复杂交通场景下的车辆运动信息提取.

2.2.3　空间结构感知流

深度信息在表征复杂交通环境中的空间结构方面具有重要作用. 深度图能够提供摄像头与各类目标之间的距离信息,从而刻画车辆与周围物体的三维相对位置关系与空间布局,为模型理解超车行为所处的时空环境提供重要支撑. 特别是在评估前后车间距、侧向距离以及道路几何结构等方面,深度信息有助于增强模型对行为合理性与安全性的判别能力. 本文采用 Depth Anything v2 提取场景的深度信息. 该算法在其原始框架中通过教师、学生学习策略进行训练,首先利用在大规模高质量合成数据上训练得到的教师模型为真实无标签图像生成伪深度标签,随后基于伪标注数据训练学生模型,以提升模型在真实复杂场景中的泛化能力与适应性. 在本文实验中,直接采用 Depth Anything v2 提供的预训练学生模型进行深度推理,不涉及额外的知识蒸馏过程或模型再训练.

2.2.4　交互事件流

为辅助模型理解车辆超车过程中的时空交互语义,引入了超车交互图,用于表述目标车辆与周边环境(如车道线、相邻车辆及交通要素)之间的交互关系. 具体构建过程如图 3 所示.

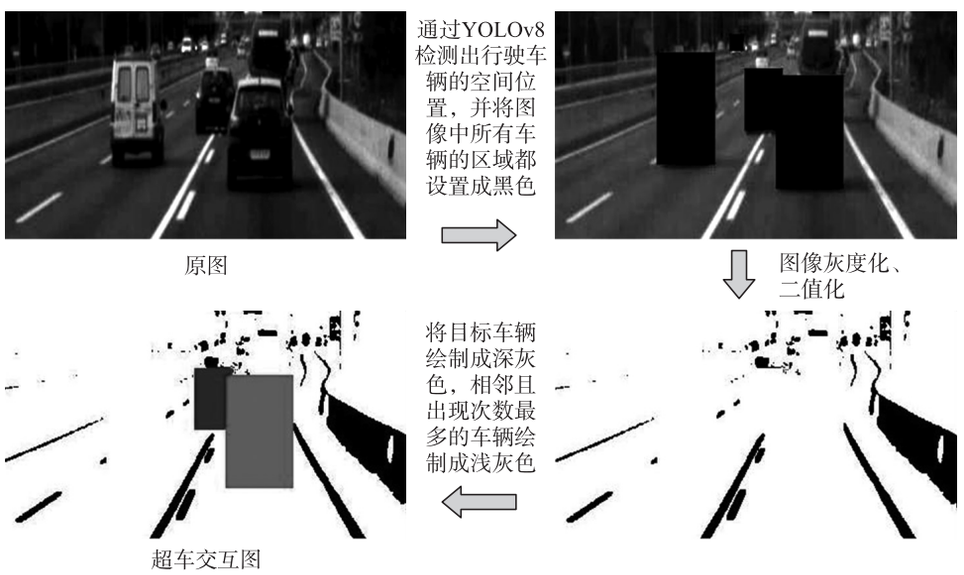

图 3　超车交互图构建过程

Fig. 3 Construction process of the overtaking interaction graph

如图 3 所示,首先采用目标检测算法 YOLO v8 对 RGB 图像序列中的所有车辆进行自动识别与跟踪,以获取其在每帧图像中的空间位置信息. 为进一步突出环境特征,随后对检测到的车辆区域进行遮蔽处理,即将其像素值设为黑色,从而保留图像中与驾驶环境相关的静态要素. 该策略旨在避免车辆轮廓对场景几何结构提取的干扰,提升环境特征的表达能力. 在此基础上,对图像进行灰度化与二值化处理,以有效提取车道线、交通标志及道路边缘等关键几何结构. 这一过程对于呈现目标车辆所处的行驶空间与周边道路环境具有重要意义,有助于模型更准确地理解其空间约束条件与可行驶区域. 为增强二值化后的图像在不同光照与场景条件下的鲁棒性,需对二值化参数进行合理设定,具体参数如下所示:

$$I'(x,y)=\begin{cases}0, & I(x,y)<\alpha\times(I_{\max}-I_{\min})\\ 255, & I(x,y)>\alpha\times(I_{\max}-I_{\min})\end{cases},\tag{1}$$

式中,$I(x,y)$表示原始图像的像素值,$I'(x,y)$表示处理后图像的像素值,$I_{\max}$表示图像中最大像素值,$I_{\min}$表示图像中最小像素值,α 表示调节参数,将其设置为$[0.35,0.75]$中的一个值.

此时,图像中的车道线以白色区域呈现,其余部分为黑色背景. 为进一步增强车道线特征的显著性,对图像进行了像素反转操作,即将原图中的黑色区域转换为白色,白色区域转换为黑色. 该操作有效提高了车道线在图像中的对比度,从而增强了其在后续处理中的可识别性,有助于模型更加准确地感知道路结构与车辆行驶轨迹.

在此基础上,结合变道行为识别算法[20]与人工筛选算法共同确定执行超车行为的目标车辆. 由于超

车行为通常伴随车辆变道操作,可通过变道识别结果初步筛选出候选目标车辆. 为进一步刻画超车过程中的交互关系,需选取与目标车辆存在紧密交互的相邻车辆. 在超车行为过程中,相邻交互车辆定义为所有非目标车辆中,与目标车辆平均空间距离最小且共同出现帧数最多的车辆. 记非目标车辆与目标车辆之间的平均距离为 D,共同出现的帧数为 N. 相邻交互对象的选取依据如下:

$$s_{id}=0.5\times\frac{D_{\max}-D_{id}}{D_{\max}-D_{\min}+\varepsilon_1}+0.5\times\frac{N_{id}-N_{\min}}{N_{\max}-N_{\min}+\varepsilon_2},\tag{2}$$

式中,s_{id}表示某相邻车辆 id 所对应的综合得分,D_{id}表示该车辆在目标车辆整个超车过程中与其之间的平均欧式距离,$D_{\max}$和 $D_{\min}$分别表示所有候选车辆中与目标车辆平均距离最远和最近的车辆所对应的距离值,用于归一化处理. N_{id}表示该车辆在目标车辆执行超车行为过程中共同出现的帧数. $N_{\max}$和 $N_{\min}$分别表示所有候选车辆中出现次数最多与最少的帧数. ε_1 和 ε_2 用于防止分母为 0. 该评分机制综合考虑了空间接近性与时间共现性,旨在精确选取与目标车辆交互最显著的相邻车辆.

选取得分最高的车辆作为交互图中的相邻交互车辆,并与目标车辆共同绘制在二值图像中. 图中通过不同颜色进行角色区分,黑色图框表示目标车辆,灰色图框表示相邻交互车辆. 该交互图有效增强了模型对关键语义要素的感知能力,准确表征了目标车辆与周围交通参与者之间的交互关系,有助于模型更全面地理解超车行为的合理性与安全性,从而为行为判别提供重要的语义支持.

为实现多模态感知信息的统一表征,本文采用特征融合策略构建结构化的超车信息图,将视觉表观流、运动信息流、空间结构感知流与交互事件流四类信息以 2×2 格式进行拼接整合. 除光流信息外,其余模态均取自同一时间帧;光流序列的首帧由前后两帧图像计算获得,后续帧由相邻帧图像对逐帧生成. 所构建的超车信息图在保证多源特征时空对齐性的同时,增强了信息表达的紧凑性与完整性,为后续模型提供了更加全面且结构化的输入表示,其构建过程如图 2 所示.

2.3　违规超车行为识别算法

本文采用 TimeSformer 架构作为超车行为识别的主干模型,通过分离式时空注意力机制对超车信息图中的空间结构特征与时间动态特征进行联合建模,从而挖掘车辆运动过程中的时空交互信息,实现对违规超车行为的精确识别. 作为一种面向视频理解任务设计的 Transformer 架构,TimeSformer 通过解耦时间维度与空间维度的建模过程,在保持全局时空建模能力的同时有效控制计算复杂度. 该架构依托自注意力机制避免了基于循环结构的递归计算所带来的梯度传播限制,更适于刻画长时间跨度内的时空依赖关系,从而能够较好地适应车辆违规超车等具有连续演化特征的交通行为识别任务. 通过分别建模时间演化过程与空间结构关系,分离式时空注意力在时序动态表达能力与空间结构感知能力之间实现了较为合理的平衡,为复杂交通场景下的细粒度行为分析提供了稳定而高效的网络结构支撑.

如图 2 所示,融合得到的超车信息图序列被输入至 TimeSformer 网络进行特征建模. 模型首先按照固定采样策略从输入序列中选取 F 帧超车信息图,并通过图像块嵌入(Patch Embedding)模块将每一帧图像划分为若干不重叠的图像块(patch),将二维图像表示映射为 patch 级 token 序列. 随后,通过线性投影方式将各 patch 映射至统一的高维特征空间. 在进入 Transformer 编码器之前,对 token 序列引入位置编码(Position Encoding)以增强位置信息表征,其中空间位置编码用于刻画不同 patch 的空间位置信息,时间位置编码用于表征帧序列中的时间顺序关系.

将位置编码增强后的 token 序列输入至由多层分离式时空注意力块堆叠构成的 TimeSformer 编码器中,用于进行时空特征提取. 在每一层编码块中,模型采用分离式时空注意力机制,依次对时间维度和空间维度进行特征提取. 具体而言,时间注意力(Temporal Attention)模块首先作用于时间维度,对同一空间位置在连续帧中的 token 序列进行特征提取,通过多头自注意力机制捕捉不同时刻特征之间的依赖关系,从而表征视频序列在时间维度上的动态变化. 随后,空间注意力(Spatial Attention)模块在单帧范围内对不同空间位置之间的 token 关系进行特征提取,以捕捉图像中各空间区域之间的结构关联.

在完成时空注意力特征提取后,通过前馈神经网络(feed-forward network,FFN)对时空特征进行非线性映射与重整,以进一步提升高层语义表示能力. 通过多层 TimeSformer 编码块的逐级堆叠,逐步形成具有较强判别性的时空语义特征表示. 最终,经分类头输出车辆违规超车行为的识别结果. 具体执行流程如算法 1 所示.

算法 1　违规超车行为识别算法

输入:超车信息图序列 X,其中每帧图像尺寸为 $H\times W$.

输出:车辆行为类别 $\hat{y}$.

1. 从超车信息图序列中提取 F 帧的图像片段 $\{X_t\}_{t=1}^{F}$.

2. for $t=1$ to F do

通过 Patch Embedding 将第 t 帧图像划分为 N 个大小为 $P\times P$ 的不重叠图像块,$N=H\times W/P\times P$.对每个图像块进行线性映射,生成对应的 Patch 嵌入向量 $x_{(p,t)}\in\mathbb{R}^{3\times P\times P}$,空间索引 $p=1,\cdots,N$,时间索引 $t=1,\cdots,F$.

end for

3. 为 $x_{(p,t)}$ 引入空间位置编码与时间位置编码,构建时序 *Token* 序列 T.

4. for $p=1$ to N do

构建该位置在 F 帧上的时间特征序列 T_p.

通过时间注意力机制建模 T_p 的动态演化特征.

end for

5. for $t=1$ to F do

构建第 t 帧内的空间特征集合 S_t.

通过空间注意力机制建模 S_t 中不同空间区域之间的依赖关系.

end for

6. 通过 FFN 对时空特征进行非线性映射与语义融合.

7. 将融合后的特征输入分类器,输出车辆行为类别 $\hat{y}$.

8. return $\hat{y}$.

车辆行为可总体划分为行驶与停止两类;在行驶状态下,进一步细分为直行与变道两类,其中变道依据方向性划分为左变道与右变道. 在实际交通环境中,变道行为通常伴随与前后或邻近车辆的交互,可进一步区分为超车行为与非超车行为. 结合交通规则约束、车道线类型(如实线/虚线)、车辆间距及道路几何结构(如弯道)等因素,超车行为可细化为合法超车与违规超车. 综合交通行为语义与规则约束,将车辆在行驶过程中的动态行为划分为六类:违规超车、右侧超车、左侧超车、左变道未超车、右变道未超车以及直行行为,以实现对复杂交通场景中车辆行驶行为的细粒度建模与识别.

3　实验与结果分析

本文中的所有实验均在一台华硕台式计算机上完成,具体硬件配置包括:华硕 B760M 主板、Intel Core i7-13700F 处理器、64 GB 内存以及配备 24 GB 显存的 NVIDIA RTX 4090 显卡. 实验使用了 Python 3.8、PyTorch 2.0 和 MATLAB R2020b.

3.1　数据集与实验设置

采用由西班牙马德里阿尔卡拉大学采集的 PREVENTION 数据集[25]. 该数据集提供了丰富的交通行为信息,包括车辆轨迹、类别、车道属性以及事件标注,涵盖诸如变道(插入、驶出、左/右变道)及危险驾驶行为等多种典型场景. 整个数据集覆盖了 356 min 的真实高速公路行驶数据,约合 540 km 的行驶里程. 基于文献[20]中对左变道与右变道行为的划分标准,进一步采用人工筛选的算法,从原始数据中选取了违规超车样本 25 个、右侧超车样本 113 个、左侧超车样本 271 个、左变道未超车样本 249 个以及右变道未超车样本 172 个,并保留原有的直行样本 5 487 个,最终构建了一个涵盖六类典型车辆行为的评估数据集.

为缓解违规超车类样本数量不足所带来的类别不平衡问题,采用了系统性的图像增强策略. 具体而言,对原始样本进行了多种单一及组合增强操作,包括:添加随机噪声以提高鲁棒性、调整亮度以模拟不同光照条件、灰度化处理以减弱颜色依赖、模糊化以模拟运动或成像模糊,以及旋转变换以增强模型对视角变化的适应性. 通过上述增强算法,违规超车类样本数量由 25 个扩展至 1 025 个,实现了约 40 倍的提升. 同时,右侧超车、左侧超车、左变道未超车及右变道未超车四类样本均扩展至 1000 个以上,使得其总量超过了直行类样本,从而有效缓解了数据分布不均衡的问题,并为后续模型训练提供了更加均衡的数据基础.

按照各类行为样本的 75%用于训练、10%用于验证、15%用于测试的比例对数据集进行划分. 各类别

在训练集、验证集与测试集中的样本数量如表 1 所示. 具体而言,违规超车样本共计 1 025 个,其中训练样本 768 个,验证样本 102 个,测试样本 155 个;右侧超车样本共 1 130 个,其中训练 847 个,验证 113 个,测试 170 个;左侧超车样本共 1 355 个,其中训练 1 016 个,验证 135 个,测试 204 个;左变道未超车样本共 1 245 个,其中训练 933 个,验证 124 个,测试 188 个;右变道未超车样本共 1 032 个,其中训练 774 个,验证 103 个,测试 155 个;直行样本共 5 487 个,其中训练样本 4 115 个,验证样本 548 个,测试样本 824 个.

表 1　各类别样本数量

Table 1　Number of samples for each category

类别	训练	验证	测试	合计	类别	训练	验证	测试	合计
违规超车	768	102	155	1 025	左变道未超车	933	124	188	1 245
右侧超车	847	113	170	1 130	右变道未超车	774	103	155	1 032
左侧超车	1 016	135	204	1 355	直行	4 115	548	824	5 487

对于 RGB 序列,首先将原始视频进行转化,并统一设置采样率为 10 FPS,即每秒提取 10 帧图像;对于光流序列,则基于相邻 RGB 帧计算帧间差分以获得运动信息;对于深度图序列与超车交互图序列,则在 RGB 序列的对应帧上分别提取深度信息与交互关系特征,以保证多模态数据在时序上的对齐与一致性.

TimeSformer 模型训练过程中,训练总轮数为 100,批量大小设置为 12. 每段视频序列采样 8 帧图像,采样间隔为 16. 图像输入在训练阶段采用随机尺度抖动,尺寸范围为 256 至 320,裁剪大小统一为 224,输入图像为三通道. 优化器采用随机梯度下降法,基础学习率设为 0.01,结合多阶段学习率衰减策略,动量系数为 0.9,权重衰减系数为 1e-4,丢弃率设置为 0.5.

3.2　评价指标

为全面评估模型性能,本文选用 Top1 准确率、Top2 准确率、精确度、召回率和 F1 分数作为评价指标,其中 Top1 准确率是主要评价指标.

Top1 准确率表示模型预测得分最高的类别与真实标签完全一致的样本在总样本中的占比,反映模型最直接的分类能力. Top2 准确率表示真实类别出现在模型预测得分前两位中的样本占总样本的比例,体现模型的次优判断能力.

精确率衡量的是模型在所有预测为正类的样本中,实际为正类的比例,反映模型对正类的判别精度. 召回率衡量的是模型在所有真实正类样本中,成功识别出的比例,反映其对正类样本的检出能力. F1 分数为精确率与召回率的调和平均值,综合考虑分类的准确性与完整性,尤其适用于样本类别不平衡的识别任务.

3.3　消融实验

从主干网络的选择、分离时空注意力机制的作用、特征融合策略的设计以及各模态数据的贡献四个方面对所提算法进行系统性分析与验证. 以 Top1 和 Top2 作为本节实验的评估指标.

3.3.1　主干网络的选择

对 TimeSformer 与 3D ResNet[20] 两类时空建模网络的行为识别性能进行对比分析,分别基于 RGB 图像、光流、深度图、超车交互图以及融合后的超车信息图进行识别实验. 各模态下的识别结果如图 4 所示.

如图 4(a)所示,在不同模态输入下,3D ResNet 与 TimeSformer 在违规超车识别任务中的 Top1 准确率存在较大差异. TimeSformer 在深度图、超车交互图及超车信息图模态下的准确率分别为 84.38%、93.34% 和 94.04%,均高于 3D ResNet 的 65.53%、85.01%和 87.43%;而在 RGB 和光流模态下,两者差距较小,分别为 60.32%对 56.26%以及 76.21%对 74.76%. 图 4(b)所示的 Top2 准确率对比亦保持一致趋势,TimeSformer 在深度图、超车交互图及超车信息图模态下分别达到 88.56%、98.00%和 98.05%,优于 3D ResNet 的 75.68%、93.74%和 94.33%;在 RGB 和光流模态下,两者差异依旧较小. 从结果可见,当输入模态包含车辆空间结构与交互关系等高层信息时,TimeSformer 依托其时空特征建模机制能够更有效地表征超车行为,从而获得更高识别性能;而 3D ResNet 在主要依赖低层视觉特征的 RGB 和光流模态下表现相对接近. 总体而言,TimeSformer 在结构化与语义化程度较高的模态输入下具有更优的判别能力和泛化性能,更适用于复杂交通场景下的行为识别任务.

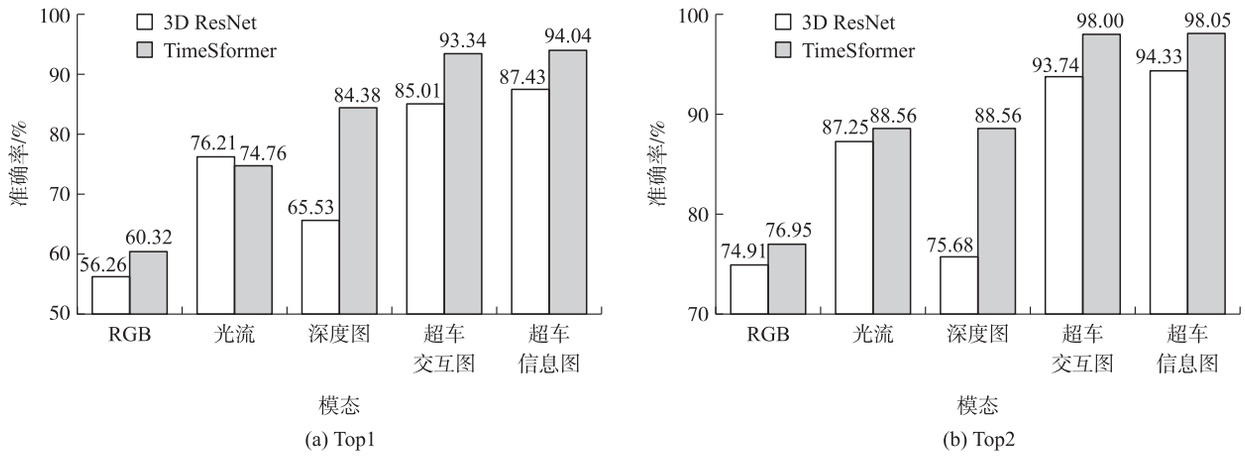

图 4　各模态数据在不同主干网络下的准确率

Fig. 4　Accuracy of different modalities under various backbone networks

3.3.2　分离时空注意力模型的作用

为了验证分离时空注意力机制在违规超车识别任务中的有效性,进行了对比实验,分别评估了空间注意力[24](Space Attention)、时空联合注意力[24](Joint Space-Time Attention)和分离时空注意力三种注意力机制在模型中的表现,实验结果如图 5 所示.

如图 5 所示,在不同模态输入下,分离时空注意力在整体表现上最优. 在 Top1 准确率方面,分离时空注意力在深度图、超车交互图及超车信息图模态下分别达到 84.38%、93.34%和 94.04%,均高于空间注意力的 75.35%、80.07%、87.21%和时空联合注意力的 79.42%、84.20%、88.09%. 在 RGB 与光流模态下,三种注意力策略的差异相对较小. Top2 准确率对比结果保持一致趋势. 分离时空注意力在深度图、超车交互图及超车信息图模态下分别达到 95.22%、98.00%和 98.05%,同样优于空间注意力的 88.97%、92.33%、

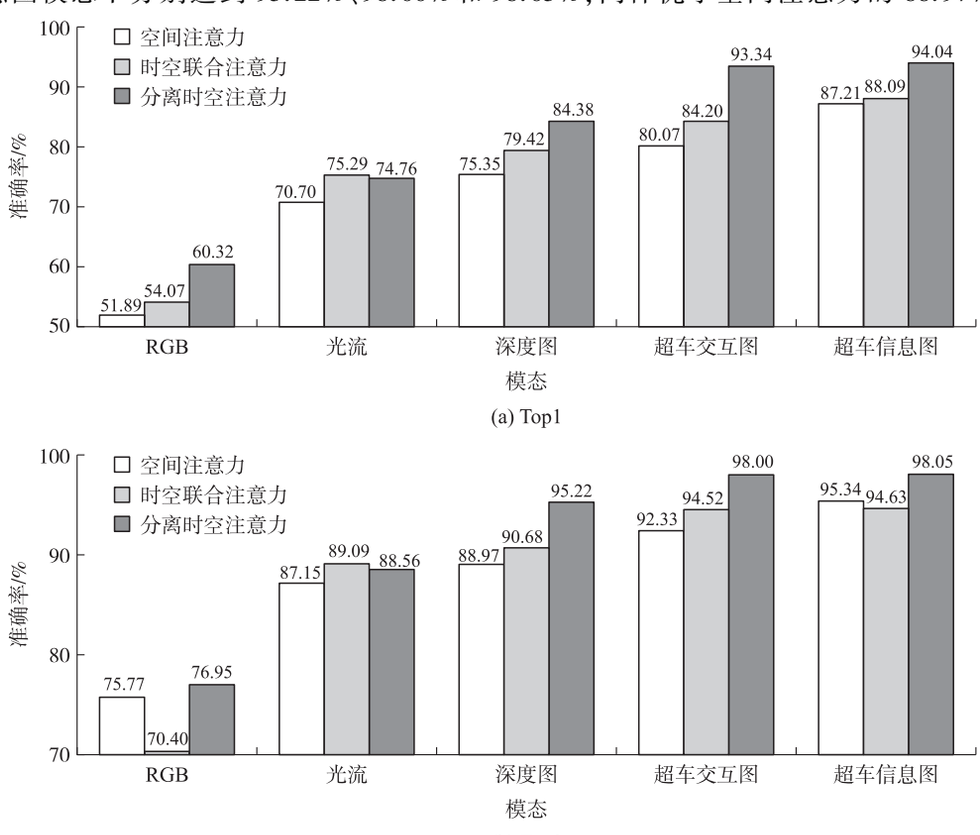

图 5　各模态数据在不同注意力模型下的准确率

Fig. 5　Accuracy of different modalities under various attention mechanisms

95.34%与时空联合注意力 90.68%、94.52%、94.63%. 整体而言,分离时空注意力在包含交互结构与时序关系的模态中优势最为显著,说明将空间特征与时间特征解耦建模能够更有效捕捉超车行为的语义动态与车辆交互关系,提升行为识别判别能力. 空间注意力仅对单帧内的空间关系进行表征,难以刻画行为随时间演化的动态特征,因而整体识别性能相对较低;时空联合注意力在一定程度上融合了空间与时间信息,但其联合建模方式在复杂时序关系刻画及结构化模态表达方面仍存在局限. 相比之下,分离式时空建模策略在融合结构化模态时表现更为有效,进一步验证了分离式时空注意力机制在违规超车行为识别任务中的优势.

3.3.3　融合策略的选择

为验证融合策略对模型性能的影响,分别对比了多模态特征融合策略与分数融合策略在车辆违规超车行为识别任务中的效果. 实验结果如表 2 所示,展示了不同融合方式下模型的识别性能差异.

表 2　不同融合方式下模型的准确率

Table 2　Accuracy of model under different fusion strategies　%

	Top1	Top2
特征融合	94.04	98.05
分数融合	93.93	98.29

其中,分数融合策略是指分别对 RGB 图像、光流、深度图和超车交互图各模态子网络进行独立分类预测,再对其输出的分类概率进行加权平均,以实现决策级的多模态融合. 特征融合策略则在特征提取阶段对 RGB 图像、光流、深度图和超车交互图各模态的特征进行拼接形成超车信息图,构建统一的特征表示,并将其输入主干网络以获得最终预测结果.

如表 2 所示,在不同的多模态信息融合策略比较中,特征融合在 Top1 识别准确率上达到 94.04%,略高于分数融合的 93.93%,提升幅度为 0.11%;在 Top2 准确率方面,特征融合为 98.05%,略低于分数融合的 98.29%. 总体而言,特征融合通过在特征层对多模态信息进行深度整合,有助于充分挖掘不同模态之间的互补语义特征,从而增强模型对违规超车行为的判别能力. 该结果进一步验证了在 RGB 图像、光流、深度图及超车交互图等多模态数据上进行特征级融合的有效性,不仅能够提升整体识别性能,也有助于增强模型的泛化能力.

3.3.4　各模态数据的作用

接下来,对各单一模态与多模态融合后的识别性能进行了对比分析,实验结果如表 3 所示. 本节中的多模态融合策略皆为特征融合.

表 3　各模态特征准确率的对比

Table 3　Comparison of accuracy across different modality features

模态				Top1/%	Top2/%	模态				Top1/%	Top2/%
RGB	光流	深度图	超车交互图			RGB	光流	深度图	超车交互图		
√				60.32	76.95		√	√		86.73	94.99
	√			74.76	88.56		√		√	93.40	97.76
		√		84.38	95.22			√	√	93.63	98.05
			√	93.34	98.00	√	√	√		86.56	94.99
√	√			74.94	86.73	√	√		√	93.63	97.29
√		√		84.67	91.51		√	√	√	93.87	97.88
√			√	93.40	96.70	√	√	√	√	94.04	98.05

如表 3 所示,融合 RGB 图像、光流、深度图及超车交互图的四流特征实现了最优识别性能,Top1 和 Top2 准确率分别达到 94.04%和 98.05%,明显优于任何单一模态,充分体现了超车信息图在视觉表观、运动轨迹、空间结构及车辆交互关系等模态信息上的语义互补性. 在单模态中,超车交互图表现最为突出,Top1 和 Top2 准确率分别为 93.34%和 98.00%,显著高于 RGB、光流及深度图,表明车辆间的结构化交互信息在复杂行为表征与判别中具有关键作用. 进一步分析表明,包含交互图的多模态组合识别性能普遍优于不含交互图的组合,验证了超车交互信息在建模目标车辆与周围车辆动态关系中的重要性. 总体而言,多模态特征融合不仅显著提升了违规超车行为的判别能力,也通过引入超车交互信息增强了模型对复杂时空行为语义的建模能力,从而有效提高整体识别性能.

3.4　与其他算法的对比

为全面评估所提算法的识别性能,本文将其与多种现有主流方法进行了对比实验,包括 3D ResNet、2D CNN 以及 Video Vision Transformer(ViViT),并进一步引入仅包含空间注意力和时空联合注意力的特征融合模型作为对比,以分析不同注意力建模策略对违规超车识别性能的影响. 各方法的识别结果如表 4 所示.

表 4　不同算法识别性能的对比

Table 4　Comparison of recognition performance across different methods　%

算法	Top1	Top2	精确率	召回率	F1 分数
3D ResNet[20]-分数融合	86.84	94.27	80.54	78.58	78.39
3D ResNet[20]-特征融合	87.43	94.33	82.29	79.12	78.96
2D CNN[19]-特征融合	89.67	97.17	85.41	85.41	84.09
2D CNN[19]-分数融合	91.50	99.53	89.03	87.75	87.28
ViViT[26]-特征融合	64.68	76.59	50.36	44.38	41.23
ViViT[26]-分数融合	64.56	76.12	47.63	44.09	41.41
空间注意力-特征融合	87.21	95.34	81.10	83.24	81.63
时空联合注意力-特征融合	88.09	94.63	82.29	83.12	82.42
TimeSformer-特征融合(本文算法)	94.04	98.05	91.00	93.81	92.30

如表 4 所示,本文算法在 Top1 准确率、精确率、召回率以及 F1 分数等指标上均优于对比算法. 具体而言,该算法在 Top1 准确率上达到 94.04%,优于 3D ResNet、2D CNN 以及 ViViT 在两种融合策略下的表现. 在 Top2 准确率方面,虽然略低于 2D CNN 分数融合策略的 99.53%,但仍保持在较高水平,并优于除该方法外的其他对比模型. 在精确率、召回率与 F1 分数方面,本文算法分别达到 91.00%、93.81%和 92.30%,体现了其在整体分类性能和违规超车行为检测能力上的优势.

从模型特性来看,2D CNN 仅能提取静态空间特征,难以捕捉行为的时间演化过程;3D ResNet 虽能同时表征时空信息,但其基于卷积的局部感受野限制了对车辆间长程依赖关系的建模能力;ViViT 使用端到端 Transformer 进行统一时空建模,但计算代价较高、对大规模训练数据依赖明显,导致整体表现受限. 相比之下,TimeSformer 有效降低建模复杂度,同时增强了对长时程动态与细粒度交互关系的表达能力,使其在多模态输入下表现尤为突出.

从注意力角度分析,分离式时空注意力在特征建模能力上整体优于空间注意力与时空联合注意力. 通过对空间结构与时间动态进行解耦建模,该机制更有利于刻画违规超车行为的长时程动态变化及车辆间交互关系,从而在多模态输入条件下获得更优的识别性能.

从融合策略的角度来看,特征融合在召回率与 F1 分数上整体优于分数融合,该趋势在 3D ResNet 和 2D CNN 架构中均得到验证. 这表明特征级融合能够更充分地整合多模态互补信息,有助于增强模型对关键行为语义和时序线索的利用能力.

结合图 6 所示的混淆矩阵分析,可进一步发现模型在右侧超车、左侧超车、违规超车以及直行类别上保持稳定的高识别率,均超过 93%. 同时,在结构相似且易混淆的类别,左变道未超车与右变道未超车中,模型的误判比例较低. F1 分数的大幅提升表明本文算法能够有效缓解类别间结构相似性带来的干扰,提高对边界模糊行为的判别能力.

综上所述,本文提出的时空交互信息融合算法在复杂道路场景下的违规超车行为识别任务中展现出优异的性能. 该算法通过融合多模态语义特征并有效建模车辆间的交互关系,显著提升了模型对复杂行为模式的理解与判别能力,充分验

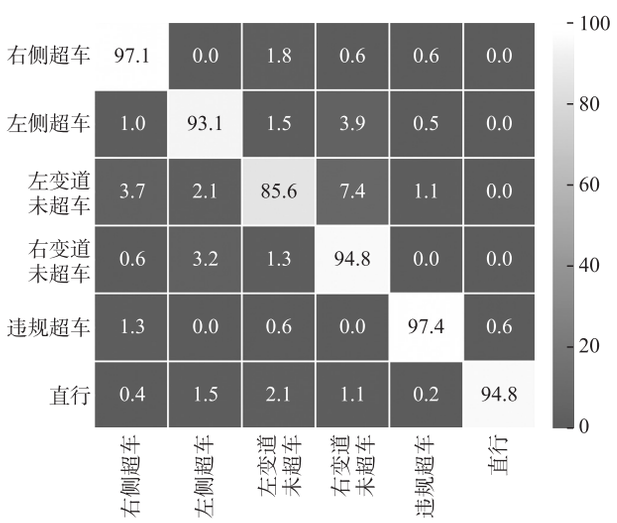

图 6　所提算法识别结果的混淆矩阵

Fig. 6　Confusion matrix of the recognition results of the proposed method

证了其在识别精度与鲁棒性方面的优势,体现出较高的实用价值与应用前景.

3.5　算法复杂度对比

为全面评估所提出模型在实际交通监测系统中的部署可行性,本文在比较各算法识别性能的基础上,进一步对比了不同模型的参数规模与浮点运算量(floating point operations,FLOPs),结果如表 5 所示.

表 5　各类模型的参数量和 FLOPS 对比

Table 5　Comparison of parameter counts and FLOPs across rdifferent models

融合策略	算法	参数量	FLOPS
特征融合	2D CNN	23.52M	6.36G
	3D ResNet	46.21M	10.10G
	空间注意力	85.80M	140.65G
	时空联合注意力	85.81M	179.71G
	分离时空注意力	121.26M	196.05G
分数融合	分离时空注意力	485.05M	196.05G

如表 5 所示,在特征融合策略中,传统模型如 2D CNN 与 3D ResNet 分别仅包含 23.52M 与 46.21M 参数,FLOP 较低,计算开销小,但由于缺乏有效的时空建模能力,其识别性能受限. 分离时空注意力模型虽然参数量 121.26M 和 FLOPs 196.05G 相较传统模型略高,但其在识别准确率方面显著优于其他模型,兼具效率与性能优势. 相比之下,空间注意力与时空联合注意力虽然参数相近约 85.8M,但时空建模能力仍有限,无法充分捕捉行为的动态特征. 分数融合策略虽在识别效果上有所提升,但需分别训练多个模态子网络,导致模型整体参数膨胀至 485.05M,显著增加了计算开销,难以满足实际部署需求. 因此,特征融合下的分离时空注意力模型在精度与计算复杂度之间取得了良好平衡,展现出更优的工程可行性.

4　结论

本文围绕复杂交通场景下车辆违规超车行为识别问题,系统分析了多模态时空信息与车辆交互关系对行为判别性能的影响,并提出了基于时空交互信息融合的违规超车识别算法. 实验结果表明,通过联合建模外观特征、运动信息、空间结构及交互语义等多源信息,并引入分离式时空注意力机制,模型能够更准确地刻画超车行为的时空演化过程,从而在 PREVENTION 数据集上取得了 94.04%的识别准确率,整体性能稳定且优于多种主流算法.

尽管所提算法在违规超车识别任务中展现出较好的有效性与稳定性,但仍存在一定局限. 一方面,违规超车行为样本数量相对有限,在一定程度上制约了模型性能的进一步提升;另一方面,多模态信息融合与 Transformer 结构引入了较高的计算复杂度,仍需在实际交通监测系统部署中加以优化. 未来研究将重点围绕轻量化网络结构设计、违规行为样本规模扩展以及复杂交通场景下模型泛化能力提升等方面展开,以进一步增强算法在真实交通环境中的适用性与工程实用价值.

[参考文献]

[1] Hong S,Yue T,Liu H. Vehicle energy system active defense:A health assessment of lithium-ion batteries[J]. International Journal of Intelligent System,2022;37:10081-10099.

[2] Nalcakan Y,Bastanlar Y. Monocular vision-based prediction of cut-in manoeuvres with LSTM networks[C]//Proceedings of Science,Engineering Management and Information Technology. Ankara,Turkey:Springer,2022:111-123.

[3] Nalcakan Y,Bastanlar Y. Cut-in maneuver detection with self-supervised contrastive video representation learning[J]. Signal,Image and Video Processing,2023,17:2915-2923.

[4] Li Z,Wang Y,Zuo Z. Interaction-aware prediction for cut-in trajectories with limited observable neighboring vehicles[J]. IEEE Transactions on Intelligent Vehicles,2023,8(3):2148-2161.

[5] Ong C S,Connie T,Goh M K O. Vehicle overtaking detection using computer vision techniques[C]//Proceedings of International Symposium on Intelligent Robotics and Systems. Changsha,China:IEEE,2024.

[6] Marcomini K D,Brito V d C,Balestra G d C,et al. A novel approach to road safety:detecting illegal overtaking using smartphone cameras and deep learning for vehicle auditing[J]. Journal of Sensor and Actuator Networks,2025,14(1):10.

[7] Zhang T,Zou Y,Zhang X,et al. Data-driven based cruise control of connected and automated vehicles under cyber-physical system framework[J]. IEEE Transactions on Intelligent Transportation Systems,2021,22(10):6307-6319.

[8] Lyu N,Wen J,Duan Z,et al. Vehicle trajectory prediction and cut-in collision warning model in a connected vehicle

environment[J]. IEEE Transactions on Intelligent Transportation Systems,2022,23(2):966-981.
[9] Lu Y, Huang L, Yao J, et al. Intention prediction-based control for vehicle platoon to handle driver cut-in[J]. IEEE Transactions on Intelligent Transportation Systems,2023,24(5):5489-5501.
[10] 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(4):48-71.
[11] 杨建喜,郁超顺,李韧,等. 基于多周期组件时空神经网络的路网通行速度预测[J]. 交通运输系统工程与信息,2021,21(3):112-119+139.
[12] 齐浩轩,曹弋,赵斌. 基于强化邻接图卷积的交叉口车辆轨迹预测模型[J]. 地球信息科学学报,2025,27(3):623-635.
[13] Rasch A, Flannagan C, Dozza M. When is it safe to complete an overtaking maneuver? modeling drivers' decision to return after passing a cyclist[J]. IEEE Transactions on Intelligent Transportation Systems,2024,25(11):15587-15599.
[14] Lu Y, Wang B, Huang L, et al. Modeling of driver cut-in behavior towards a platoon[J]. IEEE Transactions on Intelligent Transportation Systems,2022,23(12):24636-24648.
[15] Athree M, Jayasiri A. Vision-based automatic warning system to prevent dangerous and illegal vehicle overtaking[C]// Proceedings of International Research Conference on Smart Computing and Systems Engineering. Colombo, Sri Lanka: IEEE,2020.
[16] Wang P, Wu X, He X. Vibration-theoretic approach to vulnerability analysis of nonlinear vehicle platoons[J]. IEEE Transactions on Intelligent Transportation Systems,2023,24(10):11334-11344.
[17] Tran D, Bourdev L, Fergus R, et al. Learning spatio-temporal features with 3d convolutional networks[C]//Proceedings of the IEEE Conference on International Conference on Computer Vision. Santiago, Chile: IEEE,2015:4489-4497.
[18] Carreira J, Zisserman A. Quo vadis, action recognition? A new model and the kinetics dataset[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, Hawaii, USA: IEEE,2017:6299-6308.
[19] Wang L, Xiong Y, Wang Z, et al. Towards good practices for very deep two-stream convNets[PP]. 2015, arXiv:1507.02159.
[20] Chao X, Qi X, Ding R, et al. Vehicle lane change behavior recognition based on multi-scale three-stream 3D ResNets[J]. Multimedia Systems,2025,31:129.
[21] Sun D, Yang X, Liu M, et al. PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, Utah, USA: IEEE,2018.
[22] Yang L, Kang B, Huang Z, et al. Depth anything V2[C]//Proceedings of Neural Information Processing Systems. Vancouver, Canada,2024:21875-21911.
[23] Pandey S, Chen K, Dam E B. Comprehensive multimodal segmentation in medical imaging: combining yolov8 with sam and hq-sam models[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops. Paris, France: IEEE, 2023:2592-2598.
[24] Bertasius G, Wang H, Torresani L. Is space-time attention all you need for video understanding? [C]//Proceedings of the International Conference on Machine Learning. Vienna, Austria: PMLR,2021.
[25] Izquierdo R, Quintanar A, Parra I, et al. The prevention dataset: a novel benchmark for prediction of vehicles intentions[C]// Proceedings of the IEEE Intelligent Transportation Systems Conference. Auckland, New Zealand: IEEE,2019.
[26] Arnab A, Dehghani M, Heigold G, et al. ViViT: a video vision transformer[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Quebec, Canada: IEEE,2021:6836-6846.

[责任编辑:陆炳新]