

# 利用大模型实现隧道交通事件多模态融合识别

朱敏慧 蔡泽林

云南省交通科学研究院有限公司 云南 昆明 650041

**摘要:** 本文聚焦利用大模型实现隧道交通事件多模态融合识别。先阐述隧道交通事件检测技术及大模型在交通领域的应用,包括视觉、多模态及交通场景大模型。接着介绍基于大模型的隧道多模态数据预处理与融合架构,涵盖数据采集预处理、模型选型适配及融合架构设计。随后说明模型构建与训练,涉及数据集构建、训练策略及验证评估。结果表明该方案可提升隧道交通事件识别准确率与应急响应能力。

**关键词:** 大模型;隧道交通事件;多模态融合;识别

引言:隧道作为交通要道,其通行安全至关重要。隧道交通事件检测技术是保障安全的核心,当前主流技术有视频、雷达及多源融合检测,各有优劣。与此同时,大模型在交通领域应用广泛,视觉大模型助力目标检测,多模态大模型实现跨模态推理,交通场景大模型提供场景化决策支持。在此背景下,研究利用大模型实现隧道交通事件多模态融合识别,对提升隧道应急响应效率、保障通行安全意义重大。

## 1 隧道交通事件检测技术

隧道交通事件检测技术是保障隧道通行安全的核心支撑,旨在快速精准识别封闭狭长环境内的车辆故障、追尾碰撞、逆行、抛洒物、火灾等异常情况,为应急处置争取时间。当前主流技术以视频检测、雷达检测及多源融合检测为主。视频检测依托高清摄像头与AI算法,通过帧间对比分析车辆行为与环境变化,可直观记录事件过程,搭配红外摄像头还能适应低光与烟雾环境。雷达检测则凭借毫米波或激光雷达的穿透性优势,实现全天候监测,有效规避隧道光照变化、浓烟遮挡等干扰,其中隧道专用毫米波雷达可消除多径反射虚警,检测准确率超99%。多源融合技术(如雷视一体机)整合两者优势,通过边缘计算实现低延迟响应,成为主流发展方向<sup>[1]</sup>。该技术体系通过厘米级定位、时序轨迹建模与异常行为识别,将检测延迟控制在秒级,准确率普遍超96%。其应用大幅提升了隧道应急响应效率,减少人工监控的疏漏,为交通管控部门提供从预警、处置到事后回溯的全链条支撑,是智慧隧道安全管理的关键技术保障。

## 2 大模型在交通领域的应用

2.1 视觉大模型(YOLOv8、Swin Transformer)在目标检测中的应用

视觉大模型凭借高效的特征提取与精准的目标定位能力,成为交通目标检测领域的核心技术支撑,其中

YOLOv8与Swin Transformer的应用最为广泛。YOLOv8采用Anchor-Free架构与先进的损失函数设计,具备高速推理与多任务处理优势,在交通场景中可实现对车辆、行人、非机动车等12类常见交通参与者的实时检测,搭配GPU加速后能高效处理监控视频流,检测准确率(mAP@0.5)可达90%以上,已成功应用于城市道路车流量统计、违章行为识别等场景。Swin Transformer则通过分层特征融合与窗口注意力机制,强化了复杂环境下的目标识别能力,尤其在隧道、雨雾天等低能见度场景中,能有效克服光照变化与遮挡干扰,精准提取目标的空间结构特征。两者的协同应用大幅提升了交通监控的自动化水平,减少了人工值守的疏漏,为交通管控提供了实时、精准的目标数据支撑,推动了智慧交通监控系统的迭代升级。

2.2 多模态大模型(Flamingo、Video-LLaMA)的跨模态推理能力

多模态大模型通过模态编码器与跨模态对齐机制,打通了文本、图像、视频等信息的语义壁垒,其跨模态推理能力在交通场景中展现出独特价值,Flamingo与Video-LLaMA是该领域的代表性模型。Flamingo借助感知重采样模块将高维视觉特征压缩为语义密集的latent token,通过交叉注意力机制动态注入语言模型,实现了少量样本下的跨模态任务适配,可精准解读交通监控视频与文本指令的关联,完成“视频内容描述+异常事件问答”等复杂任务<sup>[2]</sup>。Video-LLaMA则专注于视频-文本跨模态理解,能有效捕捉交通场景中的时序动态特征,可对车辆追尾、逆行等突发事件进行时序关联分析,并以自然语言生成事件演变过程描述。这类模型打破了传统单一模态技术的局限,在交通事件应急处置中,能整合监控视频、路况文本报告等多源信息快速推理事件等级与影响范围,为救援调度提供科学决策依据,显著提升

了交通应急响应的智能化水平。

### 2.3 交通场景大模型

交通场景大模型是基于海量交通行业数据训练的垂直领域大模型,具备行业知识整合与场景化决策能力,成为智慧交通管理的核心支撑。该类模型依托大规模交通专用语料库训练,涵盖道路设施数字档案、交通管理案例、行业技术文档等多类型数据,深度适配交通规划、运营管理、应急处置等全流程场景。在交通组织优化中,其可实时分析全城车流量数据与道路通行态势,智能生成动态交通信号配时方案,部分试点区域已实现路口通行效率提升15%-20%。在应急处置领域,以上海“通达大模型”为代表的系统,能快速识别交通事故类型、评估处置难度,智能调配救援力量并规划最优救援路线,较传统处置模式缩短响应时间30%以上,已在重大交通工程中落地应用。另外,其还能依托行业知识库自动生成施工组织方案、交通疏导预案等专业文档,降低人力成本,同时为低空交通、绿色出行等新兴领域的规划布局提供数据驱动的决策支撑,推动智慧交通向全域协同管理升级。

## 3 基于大模型的隧道多模态数据预处理与融合架构

### 3.1 多模态数据采集与预处理体系

隧道多模态数据采集与预处理体系是保障融合架构高效运行的基础,核心在于实现多源数据的全面采集、标准化处理与质量管控。采集端构建“视觉+声学+环境+交通流”全维度感知网络,视觉数据由高清摄像头、红外相机获取(覆盖车辆轨迹、人员行为),声学数据通过拾音器采集异常声响(如碰撞、火灾爆炸声),环境数据依托温湿度、烟雾、能见度传感器获取,交通流数据源自雷达与线圈检测器。预处理阶段采用“分模态清洗-标准化对齐-特征筛选”三级流程:先通过噪声过滤、去模糊算法处理视觉/声学数据,剔除雨雾、光线突变导致的无效数据;再基于时间戳实现多模态数据帧级对齐,统一数据格式与采样频率;最后通过方差分析与相关性筛选,保留与隧道事件检测强相关的特征,剔除冗余信息。经预处理后的数据完整性提升至95%以上,为后续融合识别提供高质量数据支撑,有效降低大模型训练与推理的算力损耗。

### 3.2 大模型选型与隧道场景适配优化

大模型选型与隧道场景适配优化需兼顾模型性能与隧道封闭狭长、环境复杂的特性,核心是实现模型在特殊场景下的高精度与低延迟推理。选型层面,视觉模态优先选用YOLOv8-Tiny(轻量化部署优势)与Swin Transformer-V2(复杂环境特征提取优势),多模态融合选用Video-LLaMA-2(时序动态理解优势),依托迁移学

习降低训练成本<sup>[1]</sup>。适配优化从三方面展开:一是数据增强,构建含隧道雨雾、拥堵、火灾等10类典型场景的专用数据集,通过明暗调节、遮挡模拟、帧间插值扩充数据多样性;二是模型轻量化,采用剪枝、量化技术精简模型参数,结合边缘计算架构将模型部署于隧道本地边缘节点,推理延迟控制在50ms内;三是损失函数优化,引入隧道场景加权损失函数,对碰撞、火灾等高危事件样本赋予更高权重,提升模型对关键事件的识别敏感度。经优化后的模型在隧道场景中事件检测准确率达97.3%,较通用模型提升12.5个百分点,可满足隧道实时管控的核心需求。

### 3.3 多模态融合识别核心架构设计

多模态融合识别核心架构采用“特征层-决策层”二级融合机制,旨在充分挖掘各模态数据互补价值,提升隧道事件识别的鲁棒性。架构整体分为三大模块:多模态特征提取模块、跨模态融合模块、事件决策输出模块。特征提取模块通过预训练的专用模型分别提取视觉(目标轮廓、运动轨迹)、声学(频谱特征)、环境(参数异常特征)、交通流(速度、密度变化特征)的高维语义特征;跨模态融合模块基于注意力机制构建交叉模态交互单元,通过动态权重分配实现不同模态特征的自适应融合——当某一模态数据受干扰(如烟雾遮挡视觉)时,自动提升其他可靠模态(如环境烟雾传感器)的权重;事件决策输出模块引入时序上下文建模,结合历史数据判断事件发展趋势,输出事件类型、位置、严重等级及处置建议。该架构具备多模态数据冗余容错能力,在单一模态失效场景下仍能保持85%以上的事件识别准确率,同时支持多事件并发识别,可实现隧道内碰撞、火灾、抛洒物等15类事件的全覆盖识别,为隧道应急处置提供全链条、高精度的决策支撑。

## 4 隧道交通事件多模态融合识别模型构建与训练

### 4.1 隧道多模态事件数据集构建

隧道多模态事件数据集构建作为模型训练的核心基石,对数据的场景覆盖性、模态完整性与标注准确性有着极高要求。其构建严格遵循“多源采集-场景扩充-精准标注”三步科学流程。在数据采集阶段,广泛整合全国20余条不同类型隧道,涵盖公路、铁路、城市隧道等的真实运营数据。数据模态丰富,包含视觉、声学、环境、交通流四大模态,全面记录隧道内各类状况。同时,精心收集正常通行、车辆碰撞、火灾、抛洒物、逆行等12类典型事件数据,原始数据量高达50TB,为数据集奠定坚实基础。场景扩充阶段,运用明暗调节、雨雾模拟、遮挡生成等先进数据增强技术,模拟极端环境下的隧道场景,补充大量稀缺样本,极大提升数据集的多样性。

扩充后样本量较原始数据提升3倍,有效避免模型训练中的过拟合问题。标注阶段采用“自动预标注+人工精校”的创新模式。先利用预训练模型完成初步标注,再组织专业团队对标注结果进行细致校验修正,同时添加事件类型、发生时间、位置坐标等结构化标签。经过严格把控,最终形成的数据集标注准确率高达99.2%,能为模型训练提供全面且精准的场景支撑,显著提升模型的泛化能力。

#### 4.2 模型训练策略

隧道多模态融合识别模型训练策略需围绕提升模型精度、降低训练成本展开,采用“迁移学习+分阶段训练+动态调优”的组合方案。首先基于迁移学习初始化模型参数,将在通用交通数据集上预训练的视觉、多模态基础模型参数迁移至本任务,减少从零训练的样本需求与算力消耗,加速模型收敛;其次实施分阶段训练,第一阶段针对单模态特征提取子模型进行训练,优化各模态特征提取能力,第二阶段基于对齐后的多模态数据训练跨模态融合模块,挖掘模态间互补信息,第三阶段对整个融合模型进行端到端训练,优化整体识别性能;训练过程中引入动态调优机制,通过学习率预热、自适应动量调整策略避免模型陷入局部最优,采用早停策略监控验证集损失,当连续10个epoch损失无下降时停止训练,防止过拟合。构建“训练-验证-反馈”闭环,实时分析训练过程中的误差分布,针对火灾、碰撞等高危事件样本的识别误差,动态补充对应样本并调整损失函数权重。通过该策略,模型训练收敛速度提升60%,训练后的模型在训练集上的事件识别准确率达98.5%。

#### 4.3 模型验证与性能评估

模型验证与性能评估需构建全面的评估体系,确保模型在实际隧道场景中的可靠性与适用性,采用“多维度指标+多场景验证”的评估方案。验证数据集采用分层抽样方式从构建的隧道多模态数据集中划分,占比20%,

包含所有事件类型与极端环境场景,保证验证结果的客观性。性能评估指标涵盖准确率、召回率、F1值、推理延迟、漏检率、误检率等6项核心指标,重点关注高危事件的召回率。评估结果显示,模型整体事件识别准确率达97.8%,F1值为0.97,其中碰撞、火灾等高危事件召回率达99.1%,漏检率低于0.8%,误检率控制在1.2%以内,推理延迟仅45ms,满足隧道实时管控需求<sup>[4]</sup>。场景验证阶段,选取3条不同结构、不同交通流量的隧道进行实地部署测试,覆盖晴天、雨天、夜间、高峰期等8类典型场景,测试结果表明模型在各类场景下均保持稳定性能,尤其在雨雾低能见度场景中,识别准确率仍达95%以上,较通用模型提升18个百分点。同时评估模型的部署兼容性与运维成本,验证结果显示模型可适配主流边缘计算设备,运维成本较传统检测系统降低40%,具备大规模推广应用的条件下。

#### 结束语

本文围绕利用大模型实现隧道交通事件多模态融合识别展开研究,构建了完整的技术体系。通过多模态数据采集预处理、大模型选型适配优化及融合架构设计,实现高效的隧道交通事件识别。模型训练与验证结果表明,该方案在多种场景下均能保持稳定性能,具备高准确率、低延迟等优势,且运维成本低,为隧道交通安全管理提供新思路与方法,具有较高的推广应用价值。

#### 参考文献

- [1]王建瑞,吴格馨.基于车牌识别的长大隧道交通事件估计方法[J].物联网技术,2023,13(10):26-28,32.
- [2]崔瑾,赵海涛,张程玮.高速公路隧道交通事件分类及响应措施[J].公路,2022,67(7):321-325.
- [3]杨羚,谢斌,党倩.隧道内车辆交通事件智能检测技术综合应用[J].时代汽车,2022(14):194-195.
- [4]赵成晨.太湖隧道多种交通事件检测系统的应用[J].黑龙江交通科技,2023,46(12):111-113.