

基于多模态大模型的无人车认知决策方法研究

翁高贤 肖志烱 张钢*

海军大连舰艇学院

DOI:10.32629/pe.v4i4.21763

[摘要] 针对无人车在复杂地面环境中面临的感知信息冗余、场景语义理解不足以及规则式决策泛化能力弱等问题,本文构建了一种融合多源感知、视觉语言大模型和安全约束决策的无人车认知决策框架。系统以ROS Noetic为通信与调度基础,利用车载摄像头、激光雷达和IMU获取环境与运动状态信息,通过图像去噪、亮度均衡、边缘增强和尺寸归一化提升视觉输入质量;在认知层引入多模态大模型,通过结构化提示词将场景描述、目标状态、风险等级和建议动作输出为可解析的JSON语义结果;在决策层结合目标状态、任务规则和安全约束生成运动控制、目标跟踪、停车避障或告警等指令。基于正常通行、障碍绕行和目标跟踪三类典型无人车任务场景的180组测试结果表明,系统总体识别准确率为96.7%,目标状态判定准确率为99.4%,结构化输出合格率为100%,验证了多模态大模型赋能无人车认知决策的可行性。

[关键词] 无人车; 多模态大模型; ROS; 认知决策; 动态窗口法; 安全约束

中图分类号: TP242.6 **文献标识码:** A

Research on Cognitive Decision-Making Methods for Unmanned Ground Vehicles Based on Multimodal Large Models

Gaoxian Weng Zhihan Xiao Gang Zhang*

Dalian Naval Academy

[Abstract] To address the challenges of redundant perceptual information, insufficient scene semantic understanding, and weak generalization capability of rule-based decision-making for unmanned ground vehicles (UGVs) in complex terrestrial environments, this paper constructs a cognitive decision-making framework that integrates multi-source perception, vision-language large models, and safety-constrained decision-making. The system employs ROS Noetic as the communication and scheduling backbone, utilizing onboard cameras, LiDAR, and an IMU to acquire environmental and motion state information. Visual input quality is enhanced through image denoising, brightness equalization, edge enhancement, and size normalization. At the cognitive layer, a multimodal large model is introduced; structured prompts are designed to output scene descriptions, target states, risk levels, and suggested actions as parseable JSON semantic results. At the decision-making layer, motion control, target tracking, parking obstacle avoidance, or warning commands are generated by combining target states, mission rules, and safety constraints. Experimental results from 180 test runs across three typical UGV mission scenarios—normal traversal, obstacle detouring, and target tracking—show that the overall recognition accuracy is 96.7%, the target-state determination accuracy is 99.4%, and the structured-output qualification rate is 100%, thereby verifying the feasibility of empowering UGV cognitive decision-making with multimodal large models.

[Key words] unmanned ground vehicle; multimodal large model; ROS; cognitive decision-making; dynamic window approach; safety constraints

1 引言

地面无人车正从遥控、规则驱动模式,迭代为具备场景认知、目标研判与自主任务执行能力的智能系统,广泛应用于校园

巡检、园区安防、仓储物流、野外勘测等场景。这类平台除实现定位、导航、避障等基础功能外,还需依托视觉感知解析道路环境、目标状态与潜在风险。但传统固定规则库与单模态检测

算法,难以适配光照突变、目标遮挡、复杂背景、动态干扰等复杂工况,环境适配性与鲁棒性存在明显短板。

多模态大模型兼具图像解析与语义推理能力,为无人车从基础目标检测进阶至智能场景理解与自主决策提供了新的技术路径。本文以地面无人车为载体,聚焦三大核心工程问题:多源车载感知数据的稳定语义转化、大模型推理结果与ROS控制架构的适配融合、认知输出在安全约束下的可控运动指令生成。面向工程落地,无人车认知系统需兼顾实时性、鲁棒性与可解释性,在有限车载算力、环境扰动与模型黑箱风险的约束下,构建一体化融合架构,而非简单接入大模型模块。

现有无人车方案多采用感知、规划、控制拆分架构,仅适用于简单封闭场景。传统激光避障仅能获取距离信息,无法判别障碍物类型与运动趋势;通用检测模型仅输出目标基础信息,难以生成高层行为策略,导致底层感知与高层决策存在语义断层。针对该问题,本文构建基于多模态大模型的无人车认知决策体系,主要研究内容包括:构建图像、激光雷达与位姿信息融合的统一决策数据流;设计结构化提示词规范模型输出,保障系统稳定解析;融合大模型推理结果、DWA规划算法与安全阈值,规避模型直控风险;通过通行、绕障、目标跟踪实景实验,验证整套系统的工程实用性。

2 相关技术与研究基础

标准无人车自主系统分为感知、定位、规划、控制、任务管理五层,依托ROS分布式架构实现模块化开发维护。感知层集成相机、激光雷达、IMU、编码器采集环境与车体信息;定位层融合里程计、惯性数据、地图估算车体位姿;规划层分全局路径规划与DWA局部动态避障;控制层将速度指令转化为底盘电机输出量。

动态窗口法(DWA)由于计算效率高、约束比较直观并且适配ROS导航框架,成为了轮式机器人当前主流的局部规划算法。此方法在速度空间中分别采样线速度、角速度,并且预测下一步的运动轨迹,结合当前目标朝向、障碍物距离、路径的贴合程度和速度带来的收益来进行轨迹打分,可以纳入底盘运动约束和安全距离约束;但是该算法仅仅依靠打分程度来规划路径,无法识别目标的语义,判断任务优先级。

传统视觉算法比如目标检测、语义分割只能输出固定类别的数据,泛化能力有限,只可以回答“画面中有什么”;多模态大模型运用图文联合的方式,获得开放词汇识别和场景深度解读的能力,可以判断道路通行性、目标可跟踪性、障碍物风险等级,弥补了传统视觉算法语义缺失的短板。

当前大模型融合无人车主要存在三大缺陷:第一,文本输出的随机性比较强,难以被程序解析;第二,图像质量、提示词与上下文是影响模型误判和错判的重要因素;第三,由于认知过程是在云端进行推理,会有网络延迟和数据泄露的风险。所以本文将大模型定义为语义认知模块,搭配上结构化的输出与多层安全约束,与传统机器人路径规划算法互补协同。

3 系统架构与问题建模

本文系统构建了四大核心模块:多源感知预处理、多模态大模型认知推理、安全约束决策与运动,各模块基于ROS系统完成数据交互,构建了“感知-认知-决策-执行-反馈”的完整闭环控制系统。

多源感知预处理模块通过各类传感器采集图像、激光雷达、IMU及里程计数据,并且严格完成时间同步与标准化预处理。对图像数据进行去噪、亮度均衡与边缘增强处理,雷达与位姿数据提取障碍物距离、车速、姿态角、位置等关键状态参数,保障后续决策。

多模态大模型认知推理模块是高层语义的核心模块,通过识别障碍物、道路遮挡与目标跟踪状态输出JSON格式,使ROS节点能够稳定解析。安全约束决策模块结合当前任务、车辆状态与障碍信息约束无人车。当各项信息一致满足时无人车将执行指令进行通行、绕行、跟踪;当部分信息出现异常冲突,模块会优先采取减速、停车与告警的动作,有效抑制了大模型输出的不确定性的风险。

无人车认知决策闭环架构



认知链路: 环境感知 → 场景理解 → 风险判定 → 运动控制与任务反馈

图1 无人车认知决策闭环架构

$$X_t = \{I_t, L_t, P_t\}, Y_t = \{d_t, s_t, r_t, a_t\}, u_t = f(Y_t, C_t) \quad (1)$$

其中, I_t 、 L_t 和 P_t 分别表示图像、激光雷达和位姿信息; Y_t 表示大模型输出的描述、目标状态、风险等级和建议动作; C_t 表示安全约束集合; u_t 为无人车控制指令。

问题建模上,本文将无人车认知决策过程表示为从多源观测 X_t 到结构化语义 Y_t ,再到控制指令 u_t 的映射。与传统端到端控制不同,本文显式保留语义中间层,使系统能够输出“为什么执行该动作”的解释性信息。例如,当系统选择绕行时,结构化结果中应能够体现前方障碍物、风险等级升高和建议绕行等字段;当系统选择目标跟踪时,应能够体现目标处于可见范围、风险等级较低且任务状态允许跟踪。

4 方法设计

4.1 感知预处理与数据同步

系统通过高斯去噪、亮度均衡、边缘锐化和尺寸归一化进行图像预处理,可解决地面环境中光照变化、运动模糊、背景杂波和遮挡对视觉理解的影响。预处理后的图像通过/usb_cam/image_raw发布,位姿与激光数据分别通过/odom和/scan发布,最终输入给底盘运动控制和大模型认知推理。

$$I'_t = R(E(H(G(I_t)))) \quad (2)$$

式中, G、H、E、R分别表示去噪、亮度均衡、边缘增强和尺寸归一化算子。

图像预处理是为了降低车载图像输入的不稳定性, 辅助大模型进行视觉理解。无人车运动过程中, 摄像头容易出现轻微抖动和曝光变化, 特别是在室外环境中, 逆光、阴影和地面反射会影响模型对道路边界和障碍物轮廓的判断。通过轻量级预处理, 可以在不增加计算负担的前提下提升图像可读性, 使大模型获得更稳定的视觉输入。

数据同步方面, 系统将图像帧、雷达扫描和里程计信息统一到近似一个控制周期以内。如果图像语义识别结果显示前方可通行, 而激光雷达检测到安全距离不足, 系统会优先以雷达信息作为硬约束, 而相反系统会进入保守控制状态。该设计体现出多源信息互补原则, 语义认知提供任务解释, 几何传感器提供安全边界。

4.2 DWA局部路径规划模型

运动控制层采用动态窗口法完成局部路径跟踪与避障。系统在速度空间中采样候选线速度 v 与角速度 ω , 依据差速或全向底盘运动学模型预测短时轨迹, 并综合目标趋近度、障碍物距离、路径贴合度和速度收益选择最优控制量。

$$x_t[t + \Delta t] = x_t + v \cos \theta_t \Delta t, y_t[t + \Delta t] = y_t + v \sin \theta_t \Delta t, \theta_t[t + \Delta t] = \theta_t + \omega \Delta t \quad (3)$$

$$J(v, \omega) = \alpha \cdot \text{heading}(v, \omega) + \beta \cdot \text{dist}(v, \omega) + \gamma \cdot \text{vel}(v) - \lambda \cdot \text{risk}(v, \omega) \quad (4)$$

其中, α 、 β 、 γ 、 λ 为权重系数, $\text{risk}(v, \omega)$ 用于惩罚碰撞风险、越界轨迹或不满足安全距离的候选轨迹。

在本文系统中, DWA负责底层局部运动选择, 大模型负责高层语义判断, 两者并不是替代关系。当大模型输出“绕行”建议时, DWA根据激光雷达和局部代价地图选择满足安全距离的速度组合完成任务; 当大模型输出“跟踪”建议时, DWA在保持目标方向的同时控制速度变化, 避免车辆因目标位置抖动而频繁转向; 当大模型输出“停车”或安全约束判定不通过时, 系统直接使线速度和角速度逐步降为零。

4.3 大模型结构化认知推理

认知推理模块将预处理图像编码后输入视觉语言大模型, 并通过“角色限定-字段约束-任务规则-示例引导”的提示词模板约束输出。为便于ROS节点解析, 模型输出被固定为JSON结构, 核心字段包括场景描述、目标状态、风险等级和建议动作。

提示词设计遵循三个原则。第一, 任务边界明确, 要求模型只围绕无人车通行、避障、跟踪和异常状态进行分析, 避免输出与当前任务无关的开放描述。第二, 字段格式固定, 要求模型必须返回场景描述、目标状态、风险等级和建议动作四类字段, 且每个字段的取值范围受限。第三, 动作建议保守, 当图像不清晰、目标被遮挡或状态不确定时, 模型应优先输出停车、减速或请求人工确认, 而不是给出激进控制建议。

结构化输出的意义在于提升大模型与机器人系统之间的接口确定性。若模型只输出自然语言段落, 程序需要额外进行文本理解和关键词匹配, 容易受到表达差异影响。采用JSON字段

后, ROS节点可以直接读取风险等级和建议动作, 并将其映射到有限状态机中的通行、绕行、跟踪、停车和告警状态。这样既保留了大模型对复杂图像的语义理解能力, 又降低了控制系统对自由文本的依赖。

表1 大模型结构化输出字段设计

字段	取值范围	功能说明
场景描述	自然语言短句	概括道路、障碍物、目标和环境状态
目标状态	正常/遮挡/接近/异常/未知	描述目标或障碍物对无人车任务的影响
风险等级	低/中/高	供决策模块进行阈值化处理
建议动作	通行/绕行/跟踪/停车/告警	映射为 ROS 控制指令或任务状态

4.4 安全约束决策模型

决策模块并不直接执行大模型的自然语言建议, 而是将其与任务规则、速度限制、障碍物距离、通信状态和人工接管权限共同校验。只有当模型输出、传感器状态和安全约束一致时, 系统才能生成相应控制指令; 当字段缺失、置信不足或传感器信息冲突时, 系统进入停车或告警兜底状态。

$$D_t = 1(r_t \geq r_r) \cdot 1(d_obs \geq d_d) \cdot 1(C_t = \text{valid}) \quad (5)$$

$D_t=1$ 表示允许执行当前建议动作; r_t 为风险或任务置信指标, d_obs 为最近障碍物距离, C_t 表示安全约束校验结果。

安全约束决策模型主要包含三类规则。第一类是几何安全规则, 例如最近障碍物距离必须大于设定阈值, 车辆角速度变化不能超过底盘允许范围, 目标跟踪时必须保持最小安全间距。第二类是任务状态规则, 例如在巡检任务中, 只有当目标处于可见且风险较低状态时才允许跟踪; 当目标状态未知或出现遮挡时, 系统应降低速度并等待下一帧确认。第三类是异常兜底规则, 例如大模型接口超时、JSON解析失败、字段取值不在预设范围内或传感器数据长时间未更新时, 系统统一进入停车或人工接管提示状态。

该模型的核心思想是将大模型输出作为认知建议, 而不是最终控制命令。无人车的实际运动仍由安全约束和底层规划器共同决定。这样可以避免大模型偶发误判导致车辆产生不合理动作, 也使系统在工程部署中更容易通过日志追踪每一次决策的来源。

无人车认知决策方法流程



图2 无人车认知决策方法流程

5 实验设计

实验基于ABOT-M1-ATK人工智能无人车平台开展。主控单元承担图像预处理、大模型API调用、任务决策和指令下发; 感知

模块包括1080P摄像头、360° 激光雷达、九轴IMU和编码器；执行层由底盘运动控制器完成路径跟踪、避障、停车和目标跟踪等动作。

实验围绕无人车地面任务设置三类典型场景：正错通行场景，障碍物绕行场景，目标跟踪场景。正常通行场景用于验证道路和目标状态理解能力，障碍绕行场景用于验证风险识别与局部规划协同能力，目标跟踪场景用于验证大模型语义输出与运动控制指令的映射能力。设定每类场景60组样本，共计180组。

正常通行场景针对系统在无遮挡、道路边界较清晰、前方无明显障碍物条件下的稳定识别能力。系统应该输出低风险、可通行或保持当前路径等结果，并维持DWA局部规划器的正常速度指令。该场景虽然难度较低，但能够检验图像预处理、模型调用、JSON解析和控制链路是否具备基本稳定性。

障碍绕行场景针对系统对前方静态障碍物或临时阻挡物的识别能力。实验中，障碍物可能出现在路径中央或偏侧区域，系统需要结合图像语义和激光雷达距离判断其障碍物是否影响通行。当风险等级达到阈值时，决策模块不直接依据图像给出固定转向，而是将“需要绕行”的任务状态交由DWA局部规划器处理，生成满足安全距离的速度指令。

目标跟踪场景针对系统对指定目标状态的连续理解能力。目标可能处于静止、缓慢移动或短时遮挡状态，系统要根据图像结果判断目标是否还处于可跟踪范围内，并结合底盘运动状态生成跟踪、减速或停车指令。该场景能够体现大模型语义理解与传统运动控制之间的协同关系。

$$Acc = N_{correct}/N_{total} \times 100\%, \quad P_{json} = N_{valid-json}/N_{total} \times 100\% \quad (6)$$

Acc表示识别准确率， P_{json} 表示结构化JSON输出合格率。



图3 ABOT-M1-ATK人工智能无人车平台

表2 无人车典型任务场景识别与决策结果

测试场景	样本数量	正确识别数	识别准确率	状态判定率	指令执行率
正常通行	60	59	98.3%	100.0%	96.7%
障碍绕行	60	57	95.0%	98.3%	98.3%
目标跟踪	60	58	96.7%	100.0%	100.0%
总计/平均	180	174	96.7%	99.4%	98.3%

6 结果与分析

实验结果表明，系统在180组测试样本中正确识别了174组，整体识别准确率达96.7%，目标状态判定准确率为99.4%，所有大模型输出均严格符合JSON结构化格式，无字段缺失与格式异常问题。分场景测试显示，系统可在正常通行场景准确识别可通行状态并稳定维持规划路径；在障碍绕行场景融合视觉语义与激光雷达距离信息完成安全避障；在目标跟踪场景依据目标状态输出跟踪或停车控制指令。

其中，正常通行场景识别精度最高，在道路清晰、图像质量良好的工况下环境感知稳定性优异。该场景少量识别误差源于光照变化与地面纹理干扰，但激光雷达的测距约束可通过安全决策模块规避误判，保障控制指令稳定。障碍绕行场景准确率相对偏低，主要受障碍物形态多样、背景色彩混淆影响，导致模型边界与风险判定存在小幅波动，但多源传感器融合约束有效弥补了单一视觉语义的不确定性，整体指令执行精度较高，模型输出风险异常时系统可主动执行绕行、停车等保守策略。

目标跟踪场景可在目标可见时稳定输出跟踪控制指令，在目标短时遮挡、脱离视野时及时更新目标状态，触发减速等待策略，可适配简单动态目标工况，但在长时间遮挡、多目标交叉、目标快速移动等复杂场景下，时序建模能力仍有待优化。相较于传统固定规则决策，本方法依托大模型实现了高层场景语义解析与结构化输出；相较于大模型直接决策，本方法通过安全约束与有限状态机抑制了输出不确定性，提升了控制链路稳定性，可作为传统无人车系统的认知增强模块，有效补足传统算法的语义理解短板。

表3 不同决策策略的定性对比

方法/配置	结构化输出	安全约束	泛化能力	适用特点
固定规则决策	否	中	弱	实现简单，但难以适应复杂地面场景
大模型直接决策	部分	弱	中	具备语义理解能力，但控制安全性不足
本文方法	是	强	较强	结构化认知与安全约束控制协同

7 结论与展望

本文面向无人车复杂地面任务，提出了一种融合多模态大模型、ROS分布式通信和安全约束决策的认知决策方法。研究表明，结构化提示词能够提升大模型输出的可解析性和稳定性，安全约束决策模型能够避免自然语言建议直接进入控制链路所带来的不确定风险，使无人车在通行、避障和目标跟踪等任务中形成较为完整的认知决策闭环。

当前研究仍存在一定局限：实验规模仍以典型场景为主，尚未充分覆盖雨雾、夜间、强遮挡、多目标交互和长时间连续运行等复杂工况；云端大模型调用对网络稳定性仍有依赖；激光雷达、IMU与视觉图像尚未在大模型语义层实现深度融合。后续研究将围绕边缘部署、多模态时序融合、多目标任务规划和人机协同接管机制展开。

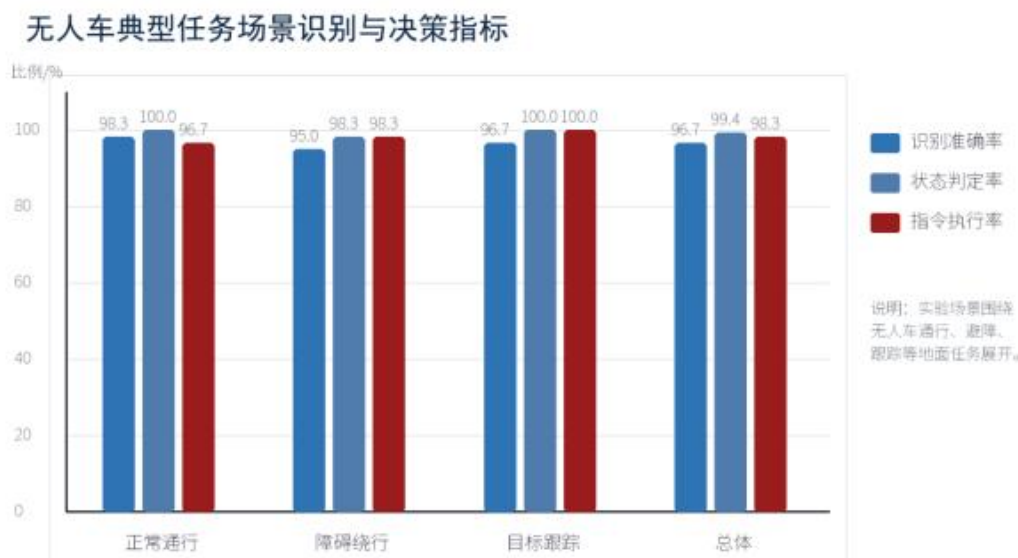


图4 无人车典型任务场景识别与决策指标

未来可从三个方面继续完善。第一,在模型部署方面,可通过量化、蒸馏和边缘推理加速等方式,将认知模型部署到车载计算平台,降低网络依赖和推理延迟。第二,在感知融合方面,可进一步研究视觉、雷达和位姿数据的时序联合表示,使模型不仅理解单帧图像,还能理解目标运动趋势和车辆自身状态变化。第三,在决策机制方面,可将安全约束模型与任务规划、行为树或强化学习策略结合,使无人车在多目标、多任务和长时间运行场景中具备更强的自适应能力。

[参考文献]

[1]仲训昱,武东杰,陈登龙,等.基于多模态感知与融合的无人车韧性导航系统[J].导航定位与授时,2021,8(06):1-12.

[2]王浩森,袁莞迈,马可,等.基于联盟博弈的无人机/无人车异构集群验证[J].工程科学学报,2024,46(07):1207-1215.

[3]王金桥,杨蓓莹.多模态大模型的发展综述及思考[J].无线电工程,2025,55(09):1727-1742.

[4]Fox D,Burgard W,Thrun S.The Dynamic Window Approach to Collision Avoidance[J].IEEE Robotics & Automation Magazine,1997,4(1):23-33.

[5]Thrun S,Burgard W,Fox D.Probabilistic Robotics[M].Cambridge:MIT Press,2005.

作者简介:

翁高贤(2005--),男,汉族,江西上饶人,本科在读。

肖志烜(2005--),男,汉族,湖北广水人,本科在读。

*通讯作者:

张钢(1992--),男,汉族,山东潍坊人,博士,讲师,研究方向为兵器科学。