

文章编号:1671-1637(2026)03-0075-14

基于多源数据与大模型的无人机 巡航风险识别技术

马涛¹, 吴俊¹, 唐樊龙^{*2}, 范剑伟², 王宁³

(1. 东南大学 交通学院, 江苏 南京 211189; 2. 金陵科技学院 网络与通信工程学院,
江苏 南京 211169; 3. 南京林业大学 土木工程学院, 江苏 南京 210037)

摘要:针对无人机在巡航过程中面临的复杂风险事件识别问题,探究了无人机巡航风险的基本要素,明确了提示词中应包含的风险特征参数;分析了多模态大模型的实现方式、架构和典型模型,提出了提示词生成模型融合多源数据的方案;结合环境感知和检测识别追踪方法建立了集成宏观场景描述、动态场景补充、突发风险检测三大模块的提示词生成模型;将提取到的特征参数集成到提示词中,通过 DeepSeek 综合分析,完成无人机巡航风险的识别与判断。研究表明:三大模块能够快速完成无人机巡航风险的识别,并获得完整的提示词;基于 Owl-ViT 模型的静态场景描述能有效识别飞行中的静态障碍物,置信度超过 80%;基于 ByteTrack 算法的动态物体抓取,可快速获取飞鸟、其他无人机等动态物体的距离、速度、坐标等动态信息;基于点云的突发风险识别可以捕捉点云障碍物信息,包括目标的距离、尺寸、体积、纵横比等,能够快速检测突发进入安全区域的障碍物;通过提示词生成的 DeepSeek 输出结果可详细展示巡航过程中的风险内容、等级并给出安全建议;开发的无人机巡航风险识别系统,可将感知识别数据可视化,并明确执行任务的设备和任务信息,进一步辅助 DeepSeek 进行风险判断。研究结果能够为无人机巡航过程中进行风险识别以及安全高效飞行提供有效技术支持。

关键词:低空交通;无人机;大模型;巡航风险识别技术;提示词生成模型;多源数据

中图分类号:U8 **文献标志码:**A **DOI:**10.19818/j.cnki.1671-1637.2026.036

Unmanned aerial vehicle cruise risk identification technology based on multi-source data and large models

MA Tao¹, WU Jun¹, TANG Fan-long^{*2}, FAN Jian-wei², WANG Ning³

(1. School of Transportation, Southeast University, Nanjing 211189, Jiangsu, China; 2. School of Network and Communication Engineering, Jinling Institute of Technology, Nanjing 211169, Jiangsu, China; 3. College of Civil Engineering, Nanjing Forestry University, Nanjing 210037, Jiangsu, China)

Abstract: To identify complex risk events during the cruise of unmanned aerial vehicles (UAVs), the basic elements of UAV cruise risks were explored, and characteristic parameters required for prompt were specified. The implementation methods, architectures, and typical models of multimodal large models were analyzed, and a scheme for integrating multi-source data in the

出版历程:2025-07-03 收稿,2025-08-19 修回,2025-09-26 录用

基金项目:国家重点研发计划(2020YFB1600102);国家自然科学基金项目(52378445);西藏自治区科技计划项目(XZ202501JX0006);金陵科技学院高层次人才科研启动项目(jit-b-202401)

作者简介:马涛(1981-),男,江苏徐州人,教授,博士生导师,工学博士,E-mail:matao@seu.edu.cn.

***通信作者:**唐樊龙(1988-),男,湖北宜昌人,讲师,工学博士,E-mail:tangfanlong@jit.edu.cn.

引用格式:马涛,吴俊,唐樊龙,等.基于多源数据与大模型的无人机巡航风险识别技术[J].交通运输工程学报,2026,26(3):75-88.

Citation:MA Tao, WU Jun, TANG Fan-long, et al. Unmanned aerial vehicle cruise risk identification technology based on multi-source data and large models[J]. Journal of Traffic and Transportation Engineering, 2026, 26(3): 75-88.

prompt generation model was proposed. By combining environmental perception, detection, identification and tracking methods, a prompt generation model integrating with macroscopic scene description, dynamic scene supplementation, and sudden risk detection was established. The extracted feature parameters were then integrated into the prompt. The UAV cruise risk identification and judgment were completed through DeepSeek's comprehensive analysis. Research results show that the three modules can quickly complete the identification of UAV cruise risks and obtain complete prompts. The static scene description based on the Owl-ViT model can effectively identify static obstacles during flight, with confidence exceeding 80%. The dynamic object capture based on the ByteTrack algorithm can quickly obtain dynamic information such as the distance, speed, and coordinates of flying birds and other UAVs. The sudden risk identification based on point clouds can capture point cloud obstacle information, including the distance, size, volume, and aspect ratio of the target, and can quickly detect obstacles that suddenly enter the safe area. The output results of DeepSeek generated by the prompt can detail the risk content and level during the cruise, and provide safety suggestions. The developed UAV cruise risk identification system can visualize the perception and identification data and determine the device and task information for the tasks, further assisting DeepSeek in risk judgment. The research results can provide effective technical support for risk identification during UAV cruise as well as safe and efficient flight.

Keywords: low-altitude traffic; UAV; large model; cruise risk identification technology; prompt word generation model; multi-source data

Publication history: Received 2025-07-03; Received in revised form 2025-08-19; Accepted 2025-09-26

Funding: National Key R&D Program of China (2020YFB1600102); National Natural Science Foundation of China (52378445); Xizang Autonomous Region Science and Technology Funding (XZ202501JX0006); Foundation of Jinling Institute of Technology (jit-b-202401)

* **Corresponding author:** TANG Fan-long, lecturer, PhD, E-mail: tangfanlong@jit.edu.cn.

0 引 言

近年来,以低空领域发展为核心的城市空中交通构想提上了日程,基于城市空中交通发展低空经济,有望成为新的经济增长引擎^[1-3]。然而,低空经济的发展还面临诸多严峻的考验,其中最重要的是保障飞行器的安全运行。2024年12月12日,一架固定翼测试无人机撞击湖北荆州奥体中心,造成一名人员严重创伤,事故引起了公众的高度关注和担忧。飞行器巡航过程中的突发状况有着严重的危险性和不可控性,且成因复杂,预测难度高^[4]。此外,飞行场景从最初的野外空旷地带到如今的城市上空、山区、水域等复杂环境,也意味着面临更加复杂和多变的风险挑战。因此,为保障飞行器巡航过程中的安全性和稳定性,有效地识别风险显得尤为重要。

初期低空经济的主要活动区域是W类空域,飞行器以微、轻、小型无人机为主^[5],因此研究首先关

注这类飞行器。针对无人机巡航过程中的风险识别,首先需要解决感知和识别问题^[6-7],再通过量化判断是否构成风险。针对感知问题,王家亮等^[8]综述了小型无人机感知环境的方法,表明视觉感知可以获取丰富的纹理和语义信息,帮助理解复杂场景,但是依赖光照条件;非视觉感知可以获取空间信息,但是信息密度低、环境适应性弱;多传感器融合则可以结合两者的优势,弥补单一传感器的缺陷,不过文章未深入探讨如何充分利用多源数据。针对识别问题,Sun等^[9]提出了光流法,将光流与其他区域不同的部分视为障碍物;Mao等^[10]提出特征识别法,用图像中的特征点定位障碍物的大致轮廓。但是这些方法在动态、非结构化场景中的泛化能力不足,促使研究者转向机器学习,利用CNN、RNN、Transformer以及混合模型等技术进行检测识别^[11-16],通过数据驱动的方式自动学习特征和规律。这些方法虽然减少了手工特征设计的依赖,但仍然不能充分融合多源数据,而且关注的风险要素也比较单一。

针对风险量化判断问题,目前研究更关注“如何检测和避开障碍物”,且采用“检测到即高风险”的简单策略,或采用既定的专家规则^[17-18]进行算法判断。这些方法很难应对真实的飞行环境及应用需求,在实际场景中则表现为无人机频繁刹车、绕行,造成不必要的能量损耗。大疆 M400 等商用无人机,凭借全向感知、超低延迟响应及 AI 驱动,已经将无人机风险识别和避障能力提升到行业顶尖水平,但是在应用于不同领域时,需要基于大量训练数据,通过迁移学习对语义分割模型及运动轨迹预测模型等 AI 模型进行场景适配,因此行业无人机多采用“按场景售卖”的策略,难以满足单一设备多场景应用的需求。

综上所述,现有的风险识别技术虽然从感知识别到量化分析环节明确,但在复杂环境准确理解、多源数据充分融合、新场景快速适配方面依然存在算法薄弱,策略简单,迁移不便等诸多问题。大模型是近年来医疗、法律、金融等领域兴起的一项新技术,通过深度优化,可以逐渐转型为具备领域知识的“专家”^[19-21],这也为无人机巡航风险识别提供了新的思路:凭借 Transformer 架构的注意力机制及参数的规模化效应表现出的强大的推理能力,能够快速理解复杂的环境^[22];通过建立统一语义空间打破模态壁垒,能够充分融合利用多源数据信息^[23-24];同时参数量巨大,通过海量语料库的自监督学习获得丰富的知识储备,能轻松实现不同场景的任务迁移,表现出强大的泛化能力^[25]。基于此,本研究采用大模型作为基座,通过构建提示词生成模型引导大模型实现风险识别和判断,为无人机在巡航过程中风险识别与飞行安全保障提供技术支持。

1 无人机巡航风险解构

构建提示词生成模型,首先需要明确大模型扮演的角色、执行的任务、分析时的注意事项和输出风险的格式;在此基础上需要明确应当包含的特征,这里的特征,严格意义上讲是特征门类,即希望通过引导大模型关注这些特征门类,从而获得完整的风险判断,输出当前场景中的风险因素、综合风险率以及最终鉴定的风险事件。因此,首先需要对无人机巡航风险进行要素解构。

无人机巡航风险事件是指无人机在巡航阶段,由于技术故障、环境干扰、人为操作失误或外部威胁等因素,导致性能下降、任务失败、财产损失或人员伤亡的潜在或实际不利事件。事件的时间范围是从无人机完成爬升进入飞行状态开始,到开始下降前

结束,空间范围涉及任务空域的限制,比如禁飞区、复杂地形、人口密集区,风险则是指超出无人机系统设计容错能力或操作规范允许范围的异常状态。主要风险事件类型见表 1。

表 1 无人机巡航风险事件类型

风险类型	主要内容
碰撞风险	与静止障碍物、其他航空器或地面物体碰撞
环境风险	极端天气(强风、雷暴)、电磁干扰、GPS 信号丢失
技术故障风险	电池、电机等动力系统失效,导航/通信中断,传感器故障
人为操作风险	操作指令错误,任务规划不合理,应急响应延迟
通信链路风险	控制信号中断,数据链延迟或被劫持
法规与合规风险	违反空域管制规定,未取得适航许可或超出操作权限
恶意攻击风险	黑客入侵、GPS 欺骗、物理劫持或电磁干扰攻击

由表 1 可以看出,无人机巡航风险事件具有动态性,飞行环境会实时变化(如突遇恶劣天气);具有隐蔽性,难以实时监测(如软件漏洞、缓慢电池衰减);具有连锁性,单一风险可能引发多次生事件(如通信中断导致时空碰撞);也可能同时面临多种类型的风险类型,受多种因素相互耦合影响,包括内部因素和外部因素,如表 2 所示。

表 2 无人机安全运行影响因素

因素	子因素	具体内容
内部因素	无人机设计可靠性	冗余系统、故障诊断能力
	能源系统的状态	电池寿命、剩余电量
	传感器与算法精度	检测识别目标物、避障、路径规划能力
外部因素	气象条件	风速、能见度、温度
	空域复杂度	障碍物密度、其他飞行器数量
	电磁环境	信号干扰源强度、频谱占用率
	人为因素	操作员经验、应急预案完善性

基于上述风险事件类型与影响因素,本研究对这些要素进行了解构,如图 1 所示。

本研究关注的空域主要是 W 类空域,该类空域管理相对自由,但环境复杂。通过分析 Aviation Safety Network 上的 190 条无人机安全事故,发现 W 类空域内由外因引起的风险事件多为碰撞,碰撞的对象有静止的建筑物、树木、电线杆、电线和运动的飞鸟和航空器等。因此,本文重点研究碰撞风险,涉及的要素主要为无人机与障碍物的相对空间信息。提示词中应包含的风险特征参数,对于静态障碍物,涉及各种障碍物的类别、位置、距离和尺寸;对于动态障碍物,除了上述要求,还应包含物体相对无

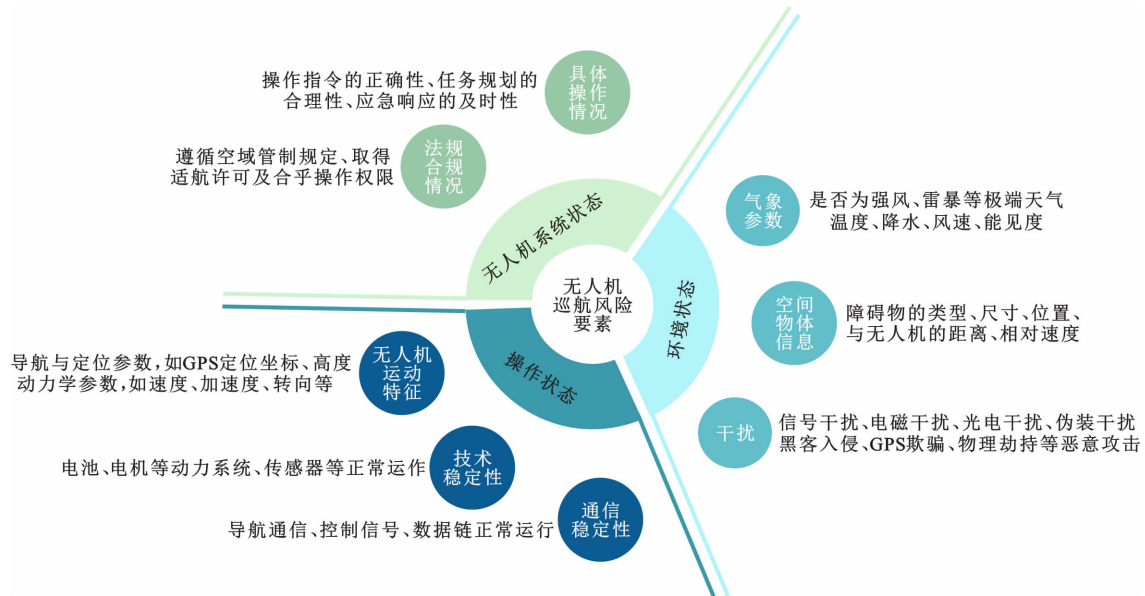


图1 巡航风险要素解构

Fig. 1 Decomposition of cruise risk factors

人机的运动速度信息。

2 多源数据融合方案研究

不同大模型对提示词有不同的要求。比较成熟的多模态大模型如 GPT4, 调用价格昂贵且目前主要擅长长文本理解、推理与生成以及图像内容的识别和分析, 无法直接处理视频、点云等数据^[26]。开源的多模态大模型, 虽然能够处理多种数据源并进行跨模态处理, 但训练数据具有局限性, 在面对差异性较大的新场景时, 泛化和理解能力会受到限制^[27]。除了多模态大模型, 一些语言类大模型在场景泛化、复杂风险理解和推理上表现良好, 如 DeepSeek^[28]。要使用大语言模型作为基座, 多模态融合的任务就需要在提示词生成模型中完成, 因此需要参考多模态大模型的实现方式, 来确定提示词生成模型融合多源数据的方案。

多模态大模型的主要架构包括模态编码器、输入映射器、语言模型骨干、输出映射器和模态生成器。模型训练阶段包括预训练阶段和微调阶段, 通常预训练阶段使用 X-Text 数据集(图像/视频/音频-文本对或交错语料)训练输入映射层, 优化跨模态对齐损失。指令微调阶段则通过指令微调格式的数据集对模型进行监督微调或基于人类反馈的强化学习, 增强模型对人类意图的遵循能力, 提升零样本泛化性^[29]。

根据上述多模态大模型的模型架构和训练方式, 可提出 3 种在提示词生成模型中融合多源数据

的方案, 如图 2 所示。3 种方案的对比见表 3、4。

表 3 数据融合方案指标对比

Table 3 Comparison of indicators for data fusion schemes

方案	训练数据量要求	算力要求 (10^9 FLOPS)	泛化能力
1	低(采用单模态预训练模型)	0.8	中等(依赖模块设计)
2	高(跨模态对齐数据)	2.5	强(模态间相互学习)
3	极高(海量多模态数据)	5.0+	极强(端到端自适应)

表 4 数据融合方案适用场景与优势分析

Table 4 Analysis of applicable scenarios and advantages of data fusion solutions

方案	适用场景	主要优势
1	适用于实时性要求高的动态场景	计算高效, 易部署, 支持增量更新
2	多模态强交互任务	跨模态理解深度优, 支持复杂语义推理
3	离线高精度分析	全模态联合优化, 小样本泛化能力突出

方案 1: 通过不同的模块独立处理各种模态, 将提取到的特征通过投票、加权、拼接等方式融合在一起。

方案 2: 通过不同的编码器独立处理各种模态, 输出的特征向量映射到统一语言向量空间中, 再由文本 Transformer 交叉注意学习其中的关系。

方案 3: 将多源数据映射到高维统一向量空间, 由自研的多模态 Transformer 交叉注意, 学习各种信号之间的关系, 实现跨模态的交互和理解, 比如 OFA、Emu。

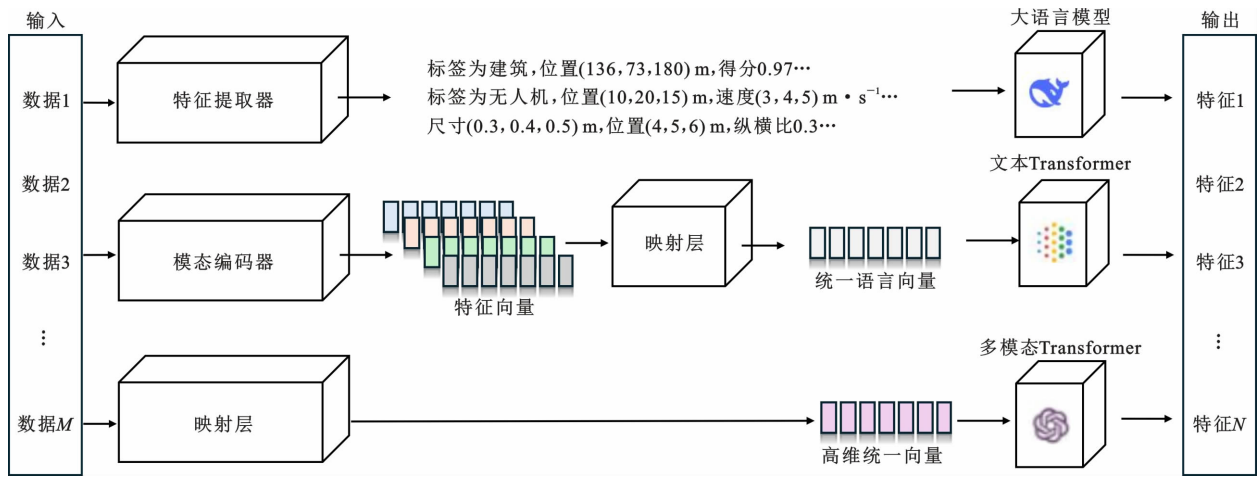


图 2 提示词生成模型数据融合方案

Fig. 2 Data fusion schemes for the prompt word generation model

3 巡航风险提示词生成模型构建

3.1 方案设计

要实现在复杂运行场景中的风险识别,一方面,需要部署尽可能多的感知设备,以感知尽可能多的环境信息,包括静态障碍物、动态障碍物、环境威胁,保证环境感知具备一定的安全余度;另一方面,需要利用大模型实现障碍物检测与识别。从硬件角度考虑,部署传感器需要考虑经济性和计算负担;从软件角度考虑,RGB图像、深度图像、点云等形式数据的处理算法更为成熟,易于在不同场景间迁移。综合考虑,宜采用单目相机、深度相机、激光雷达 3 种传感器,通过分析图像、深度图像、视频和点云数据实现风险识别。

根据表 3、4,第 2 节中提到的 3 种多源数据融合方案,第 2 种需要获取新增映射层和语言统一向量 Transformer 的参数,第 3 种需要获取映射层和高维统一向量 Transformer 的参数,都需要依赖大量的数据和算力进行训练。因此本文采用第一种方案,即通过不同的模块获取数据特征,将它们进行拼接实现数据融合,具体步骤如下。

步骤 1:由单目相机拍摄场景图片,检测目标获取各个目标物检测框的位置、长、宽等信息,结合深度图获知各个物体的实际距离,综合目标物的信息,让大模型对整个场景有一个全方位的把控和理解。

步骤 2:对于动态障碍物,同样采用上述方式,获得其每一时刻在图像中的位置信息,然后对目标进行追踪,获得它在每一个时间戳上相对于无人机的距离、运动速度和方向信息,补充动态信息。

步骤 3:由于视觉感知存在视野局限、易受光照

等条件影响等问题,可能存在突发进入安全区域但无法视觉识别的物体,可以采用点云数据对这一类障碍物进行感知,强调捕捉进入安全区域且未被动态模块持续追踪的障碍物,比如飞鸟在首次出现时属于突发风险,进入追踪后归类为动态障碍物。

3.2 提示词生成模型构建

3.2.1 基于 Owl-ViT 模型的静态场景描述

对静态场景进行描述,需要提取图像中障碍物的位置、距离等特征。目前主流目标检测模型的特性以及对比情况,如表 5 所示。GLIP 与 Owl-ViT 模型通过对比学习方式训练,能够实现开放目标检测,在场景迁移时不需要基于大量的数据进行预训练,与大模型结合泛化性较好,因此考虑采用这 2 种模型,通过在不同维度上对两者进行对比,决定采用哪一种模型,具体见表 6。

由表 6 中 Owl-ViT 和 GLIP 模型的维度对比推断,在资源环境受限,需要轻量级开放目标检测,且对实时性要求较高的情况下,Owl-ViT 将是更好的选择,因为模型检测结果后续会交给大模型进行理解,且设计短语进行匹配要比词汇复杂很多。基于此,本文采用 Owl-ViT 模型进行静态场景中的目标检测。Owl-ViT 生成图像相关提示词时,通过循环逐个读取每个目标物的类别、距离、位置以及得分信息。实际部署时也可以根据实际需求,选择得分排名靠前的一部分目标来简化计算。

3.2.2 基于 ByteTrack 算法的动态物体抓取

得到物体的类别、距离、位置、得分信息后,在此基础上对比 SORT、DeepSORT、FairMOT、OC-SORT、ByteTrack、BoT-SORT 等算法的特性,并选取最优算法实现目标追踪。算法的核心思想及数据

表5 主流目标检测模型的特性

Table 5 Characteristics of mainstream object detection models

特性	Fast R-CNN	YOLO	DETR	GLIP	Owl-ViT
架构基础	CNN	CNN	Transformer	Transformer	Transformer
训练方式	监督学习	监督学习	监督学习	对比学习	对比学习
检测方式	封闭词汇	封闭词汇	封闭词汇	开放词汇	开放词汇
零样本能力	无	无	有限	优秀	优秀
多模态支持	否	否	否	是	是
推理速度	较慢	极快	中等	较慢	中等
主要优势	成熟稳定	实时性	端到端检测	强零样本能力	开放世界适应性
主要局限	固定类别	固定类别	训练复杂	计算资源需求高	小目标检测弱

表6 Owl-ViT与GLIP模型不同维度对比

Table 6 Comparison of dimensions between Owl-ViT and GLIP models

对比维度	Owl-ViT	GLIP
设计理念	开放世界的视觉定位	检测与定位的统一框架
模型架构	ViT+文本Transformer	Swin/BERT+深度融合模块
预训练数据	400 M 图像-文本对	27 M 检测数据+3 M 人工标注
任务形式	图像-文本对比学习	区域-短语匹配任务
零样本迁移	通过文本提示实现	通过语言描述生成动态检测头
检测头设计	固定结构	动态生成
典型应用场景	开放词汇检索、定位	细粒度视觉语义理解
推理速度	中等(约 15 帧·s ⁻¹)	较慢(约 8 帧·s ⁻¹)
模型规模	基础版约 300 M 参数	大型版可达 1 B 参数
小目标检测	相对较弱	表现更好(Swin 层次化特征)
语言理解深度	侧重关键词匹配	支持复杂语义关系理解

关联策略见表7。

这些追踪算法的主要区别在于几个要素。

(1)是否关心低分检测框,忽略低分检测框,在目标物被遮挡或者模糊的情况下极易丢失ID,造成ID频繁切换,考虑低分检测框则可能引入噪声。

(2)运动模型,线性运动模型预测边界框位置,计算简单,但是不能适应目标的快速变向,容易导致追踪失效;非线性运动模型引入轨迹插值,即使目标运动突然变化(如急转弯、遮挡后重现),也可以继续追踪,减少ID切换,但算法复杂度高,实时性差。

(3)是否结合ReID(外观)特征,不结合外观特征,计算效率高;结合外观特征,可以区分外观相似、位置接近的不同目标,并且即使目标被短暂遮挡,重现后仍然可以通过外观特征匹配回原来的ID,而且,不依赖运动模型,对快速变向目标也有效,但是,会增加计算开销,而且需要预先训练ReID模型,结

果高度依赖数据质量。算法相关特性见表8。

结合各追踪算法的优缺点对比,为了实现追踪速度和精度的最优平衡,本文采用ByteTrack算法追踪目标。通过2次匹配(高分+低分检测框)和简单运动模型实现关联,减少遮挡导致的ID丢失,同时保证较高的实时性。同时Owl-ViT检测结果中蕴含的语义信息,可以弥补ReID模型的缺失,因此ByteTrack不需要再结合ReID模型并进行额外的外观特征模型的预训练,与检测器直接结合即可实现追踪。检测的结果是目标在每一个时间戳上的像素坐标,结合相机参数计算其在相机坐标系中的位置,再结合云台的姿态计算其在无人机坐标系中的位置,然后通过关联连续多帧,可以计算得到目标相对无人机运动速度的大小和方向,最终输出障碍物类别及在每一时刻相对无人机的运动信息,实现追踪。

3.2.3 基于点云的突发风险识别

当前还不存在点云的开放检测模型,因此暂不能实现点云与语义信息的完全对齐,但点云数据主要用于检测突发进入安全区域的障碍物,得到障碍物的尺寸、距离、形变、大致的运动趋势等信息已经足够。其中,形变在一定程度上可以反映障碍物可能对无人机造成的威胁的程度,大致可以认为,如果物体在连续帧中形状特征变化比较明显,它可能比较柔软,造成的威胁可能相对较小;反之,物体可能比较坚硬,造成的威胁可能相对较大。具体实现步骤如下。

步骤1:通过约束点云区域、去噪与下采样对点云数据进行预处理,清除噪声,减少计算量。以无人机前进方向±60°的区域建立锥形主区域,主区域和次区域采用不同的体素降采样值进行下采样,针对性地减少计算量。然后计算点云中的每个点到最近k个邻域点的平均距离,用阈值过滤,其中阈值等于

表 7 追踪算法核心思想与数据关联策略对比

Table 7 Comparison of core concepts of the tracking algorithms and the data association strategies

算法	核心思想	数据关联策略
SORT	卡尔曼滤波+匈牙利算法,仅用高分检测框	IoU 匹配
DeepSORT	结合 ReID(外观)特征,减少 ID 切换	IoU+外观相似度
FairMOT	联合检测+跟踪(JDE),端到端训练,共享特征提取器	基于中心点的关联
OC-SORT	改进运动模型,引入轨迹补偿机制预测目标位置,对非线性运动更鲁棒	运动一致性+观测平滑
ByteTrack	2 次匹配(高分+低分检测框),减少漏检,卡尔曼滤波预测边界框位置,依赖检测质量而非复杂运动模型	基于 IoU 和外观相似度(可选),但对遮挡目标更鲁棒
BoT-SORT	结合 ByteTrack 的低分策略和 DeepSORT 的 ReID 模块,平衡速度与精度	IoU+外观相似度+运动补偿

表 8 追踪算法适用场景与特点对比

Table 8 Comparison of application scenarios and characteristics of tracking algorithms

算法	适用场景	优点	缺点
SORT	实时性要求高、目标遮挡少的场景 (如简单交通监控)	计算量小,速度快,可达 $60 \text{ 帧} \cdot \text{s}^{-1}$	丢弃低分框,遮挡时易丢失目标,导致 ID 切换频繁,依赖检测质量
DeepSORT	对 ID 一致性要求高且算力充足的场景 (如体育赛事分析)	ID 稳定性高,长期跟踪效果好	ReID 模块计算量大($40 \text{ 帧} \cdot \text{s}^{-1}$),丢弃低分检测框
FairMOT	行人跟踪,多摄像头 ReID (如商场监控)	端到端,高效,检测与跟踪联合优化	ReID 与检测任务冲突,遮挡时性能下降,计算量大
OC-SORT	适合复杂运动轨迹(如无人机 视角跟踪、快速变向目标)	适合复杂运动轨迹,减少 ID 切换	算法复杂度较高,实时性稍差($30 \text{ 帧} \cdot \text{s}^{-1}$)
ByteTrack	遮挡频繁、需要实时性的场景 (如实时监控、交通流量分析)	利用低分检测框进行关联匹配,减少漏检,速度快($50 \text{ 帧} \cdot \text{s}^{-1}$),遮挡鲁棒	无 ReID 模块,相似目标区分能力较弱
BoT-SORT	需高精度+实时性的场景 (如自动驾驶、高精度监控)	结合 ByteTrack、ReID 优势, ID 稳定,速度较快	需要预先训练 ReID 模型,配置较复杂

平均值+标准差乘以标准差比率,即通过统计滤波实现去噪。为提高计算效率,对完成去噪的整体点云再进行一次下采样。

步骤 2:因为面对突发风险时难以确定物体的形状,所以采用基于密度的 DBSCAN 聚类对障碍物进行聚类分析,需要确定邻域半径和最小点数值。对于典型的邻域半径取值,在室内环境中为 $0.1 \sim 0.3 \text{ m}$,在室外环境为 $0.3 \sim 1.0 \text{ m}$ 。对于典型的最小点数值,由长距离雷达获得的稀疏点云取 $5 \sim 10$,由 RGB-D 相机获得的稠密点云取 $20 \sim 50$ 。可以动态调整这 2 个参数,如当无人机离地面比较近时,障碍物比较密集,需要设定一个较小的邻域半径区分不同物体,当无人机离地面比较远时,障碍物比较稀疏,相应的,增大邻域半径。

步骤 3:为保证检测的连续性,进行多帧关联验证,获取稳定 ID 的障碍物的尺寸、位置、形变率等特征。多帧验证在距离阈值的基础上通过运动一致性、形状相似性及短暂记忆进行辅助匹配。运动一致性是指用速度或加速度预测下一帧的位置,采用简单的线性预测。形状相似性是指匹配时检查纵横比或体积的连续性。短暂记忆是指允许短暂丢失后

重新关联,比如设定最大的跳跃帧是 3,则保存追踪 ID 三帧。

4 识别结果与识别系统构建

4.1 提示词生成模型与大模型结果

根据第 3.2 节建立完成的提示词生成模型,集成了宏观场景描述、动态场景补充、突发风险检测三大模块,将其输出全部导入到 DeepSeek 中可以实现风险判断。

将 Owl-ViT 检测阈值设置为 0.1,对得分排名前 50 位的障碍物,标注其检测分数、障碍物的距离及误差线,最后输出静态场景目标的特征。在城郊、夜间、城市复杂场景下的检测结果见图 3。检测框中包含静态场景中的房屋,树木以及栏杆等影响飞行安全的障碍物,通过映射到深度图计算出障碍物的距离。scene_context 是目标空间特征输出的具体内容,包含静态障碍物的类别、距离、得分。检测结果显示,Owl-ViT 检测模型可以准确充分地检测障碍物,生成对宏观场景的描述,且小目标障碍物的检测效果良好。

将 ByteTrack 的追踪阈值设置为 0.25,匹配阈



图 3 无人机巡航过程中的静态场景检测结果

Fig. 3 Static scene detection results during the UAV's cruising process

值设为 0.9, 关联连续多帧计算动态障碍物的特征, 如图 4 所示, 检测框中是关心的运动物体, 比如飞鸟、无人机, 检测的结果包含动态障碍物的类别、得分、在每一帧中相对无人机的位置。dynamic_context 是目标运动信息输出的具体内容, 包含在每一帧中的与无人机的距离和相对无人机的速度。追踪结果显示, Owl-ViT 检测模型与 ByteTrack 追踪算法结合可以稳定地捕捉动态障碍物, 补充动态信息。

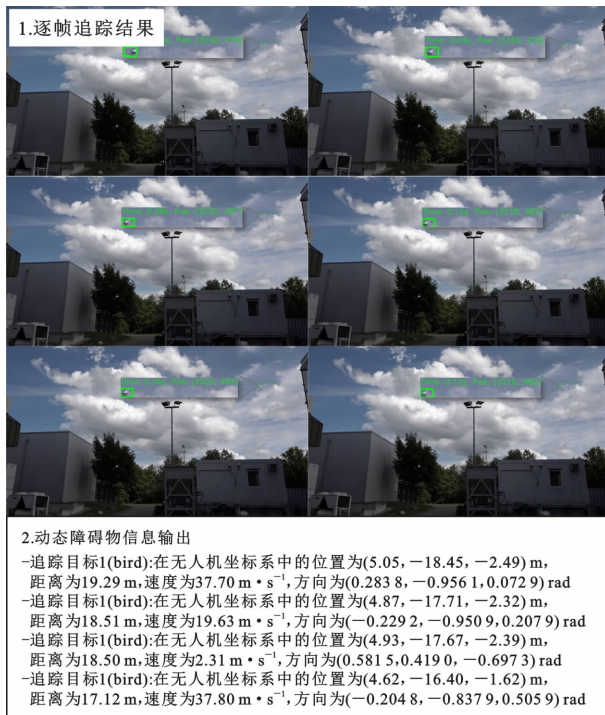


图4 无人机巡航过程中的动态障碍物追踪结果
Fig. 4 Dynamic obstacle tracking results during the UAV's cruising process

突发风险检测模块中, 将主区域角度设为 120° , 距离设为 10 m, 次区域体素降采样值设为 0.5 m。统计滤波的邻域点数为 20, 标准差比率为 1, 体素降采样值取 0.02 m。点云聚类的邻域半径是无人机运行高度和速度的函数, 小于 2.5 m, 最小点数取 5, 聚类数不超过 100。对点云进行多帧匹配时, 最大距离阈值取 2 m, 前后帧纵横比不超过 0.2, 超过 3 帧短暂记忆失效, 重新识别到后视为新目标。由连续 5 帧计算目标形变率, 体积形变率与纵横比形变率的权重分别为 0.6、0.4, 当计算得到的形变率超过 0.4 时, 认为物体较为柔软, 对无人机构成的威胁相对较小。检测结果如图 5 所示, 图中立方体区域内是空间中的障碍物。pointcloud_context 是目标聚类关联输出的具体内容, 包含目标的距离、尺寸、体积、纵横比及根据形变率判断的大致类别。检

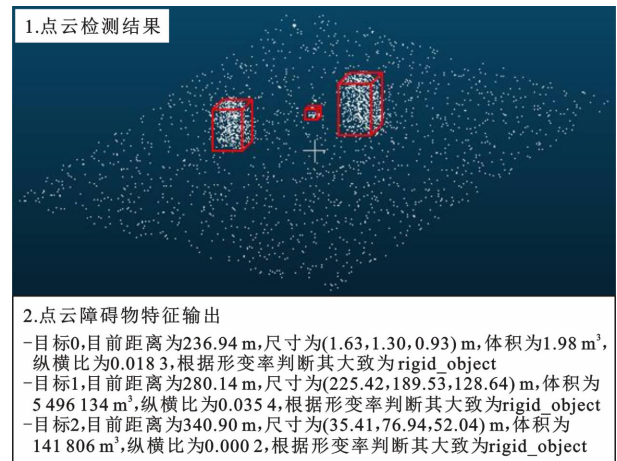


图5 无人机巡航过程中进入安全区域障碍物的检测结果
Fig. 5 Detection results of obstacles that entered the safety zone during the UAV's cruising process

测结果显示, 模型可以捕捉点云障碍物信息, 用于检测突发进入安全区域的障碍物。

将上述 3 个模块生成的 scene_context、dynamic_context、pointcloud_context 拼接输入 risk_prompt 中生成完整的提示词, DeepSeek 接收提示词后的输出结果表明, 当前无人机飞行面临的主要风险有: 距离无人机最近约 4.1 m 的高压线对无人机构成极高的碰撞和电磁干扰风险; 鸟类以高速(平均速度约 $26.8 \text{ m} \cdot \text{s}^{-1}$) 接近无人机, 最近距离约 17.1 m; 多棵树和建筑物距离无人机较近(最近约 4.5 m), 可能限制无人机的机动空间。风险等级判定为高。提出如下建议: 优先垂直爬升或水平远离高压线至安全距离(建议至少 10 m 以上); 根据鸟类运动方向[平均方向应为 (0.134, -0.124, -0.158) rad], 向相反方向(如右上方)机动避让。因此, DeepSeek 可以理解环境中的复杂风险要素, 完成风险判断和等级评定, 提出避障决策。

对 3 个模块及 DeepSeek 进行测算, 结果见表 9。其中宏观场景描述由 Owl-ViT-Large 模型实现, 在 RTX 4090 上平均延迟为 60~80 ms, 识别准确率达 75%; 动态场景补充由 Owl-ViT-Large 检测和 ByteTrack 追踪获得障碍物运动信息, 延迟时间与识别准确率取决于目标检测模型, 跟踪算法延迟仅增加 5~10 ms, 语义信息的补充保证了目标的稳定追踪; 突发风险检测通过“点云处理-聚类分析-多帧关联-障碍物属性计算”流程实现, 平均延迟 25~35 ms, 识别准确率达 85%; 最后, 巡航风险判断由 DeepSeek 实现, 平均延迟 40~50 ms, 识别准确率达 80%。整个流程中, 动态场景补充和突发风险检测并行, 延迟 70~85 ms, 再由 DeepSeek 推断, 增加

延迟 40~50 ms,因此风险判断总延迟 105~140 ms,通过模型量化与边缘部署可缩减至 50~80 ms。

由表 10 避障全链路延迟分解,总延迟为 70~145 ms,满足无人机低速巡航可容忍延迟要求。

表 9 三大检测模块及 DeepSeek 测算结果

Table 9 Three major detection modules and DeepSeek measurement results

模块	功能实现模型	平均延迟时间	识别准确率
宏观场景描述	Owl-ViT-Large	60~80 ms	75%
动态场景补充	Owl-ViT-Large+ByteTrack	跟踪算法延迟仅增加 5~10 ms,延迟主要在检测	依赖检测模型,准确率 75%,ID 稳定
突发风险检测	点云处理-聚类分析-多帧关联-障碍物属性计算	25~35 ms	85%
巡航风险判断	DeepSeek	40~50 ms	80%

表 10 避障全链路延迟分解

Table 10 Obstacle avoidance full-link delay decomposition

环节	子任务	延迟范围/ms	影响因素及优化方向
传感器数据采集	摄像头图像传输激光雷达点云生成	5~20	传感器硬件性能
风险识别	宏观场景检测(Owl-ViT)突发障碍检测(点云)	50~80	模型量化、硬件加速(TensorRT)
决策规划	路径重规划动态障碍预测	10~30	算法复杂度
控制执行	电机响应、姿态调整	5~15	飞控板性能
总延迟		70~145	

4.2 无人机巡航风险识别系统的开发

为实现全流程中的无人机设备管理、飞行任务配置、实时监测结果传递、以及 DeepSeek 风险判断结果等多维度数据控制的协同统一,在上述研究的基础上开发了无人机巡航风险识别系统。系统包括实时监控、设备管理、任务执行以及用户配置 4 个界面分别如图 6~9 所示。

其中实时监控界面展示检测追踪结果,输出对应的 scene_context、dynamic_context、pointcloud_context,并且结合巡航设备信息、任务情况、气象信息以及设备状态作为提示词,引导 DeepSeek 生成风险判断。设备管理界面用于管理监测的设备,涉及的参数有设备的类型、尺寸、质量、续航、最小转弯半径、最大飞行高度、最大飞行速度。任务执行界面



图 6 实时监控模块

Fig. 6 Real-time monitoring module

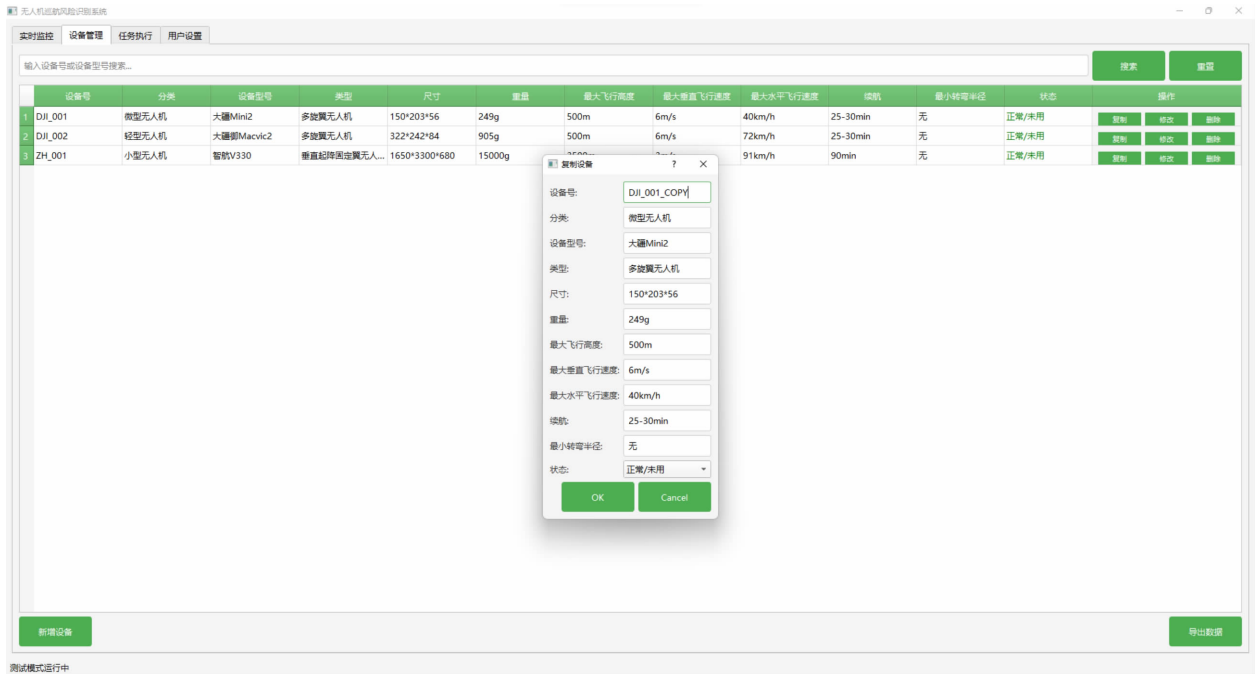


图 7 设备管理模块

Fig. 7 Equipment management module

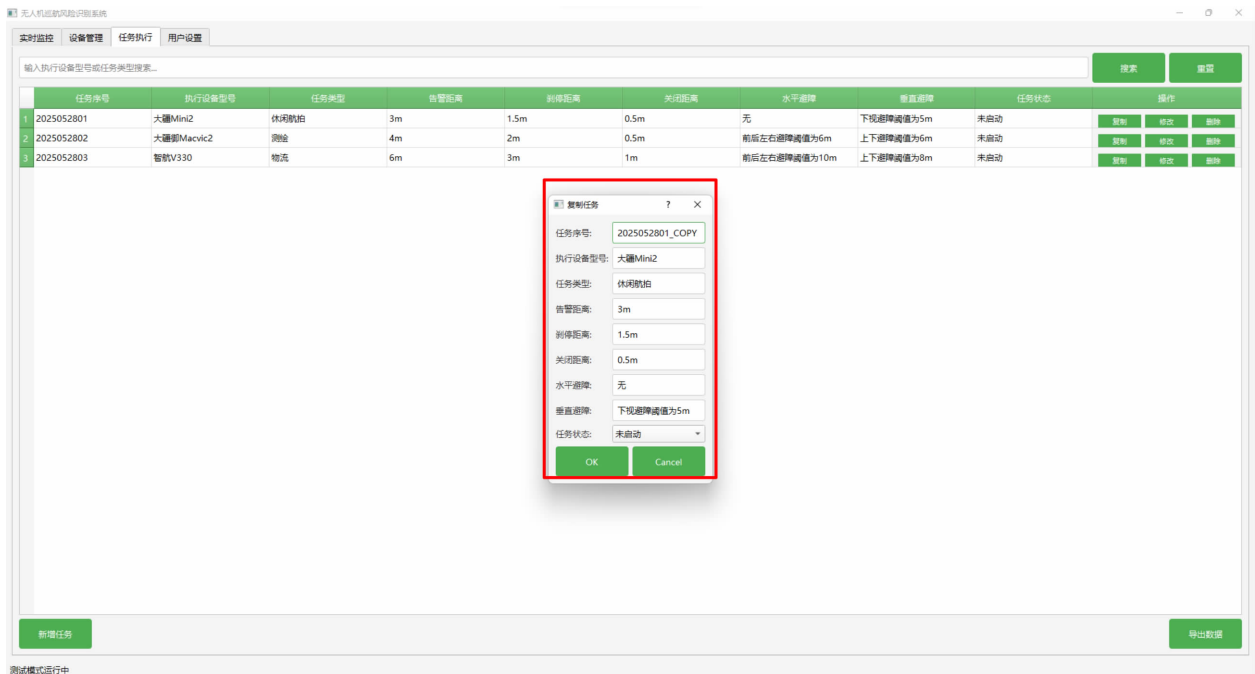


图 8 任务执行模块

Fig. 8 Task execution module

用于管理监测设备正在执行的任务,涉及的参数有任务类型、告警距离、刹停距离、关闭距离、垂直避障和水平避障的距离阈值。这 2 个界面明确执行任务的设备和任务信息,辅助 DeepSeek 进行风险判断。用户配置模块用于确定模式,获取调用 DeepSeek 的 API,以及由用户根据试验及实际场景选择参数取值。

5 结 语

(1)对无人机巡航风险事件的分析和解构发现,主要活动在 W 类空域的无人机由外因引起的风险事件多为碰撞,其包含的主要要素为无人机与障碍物的相对空间信息,且风险源可以划分为静态障碍物、动态障碍物以及突发类障碍物三大类。

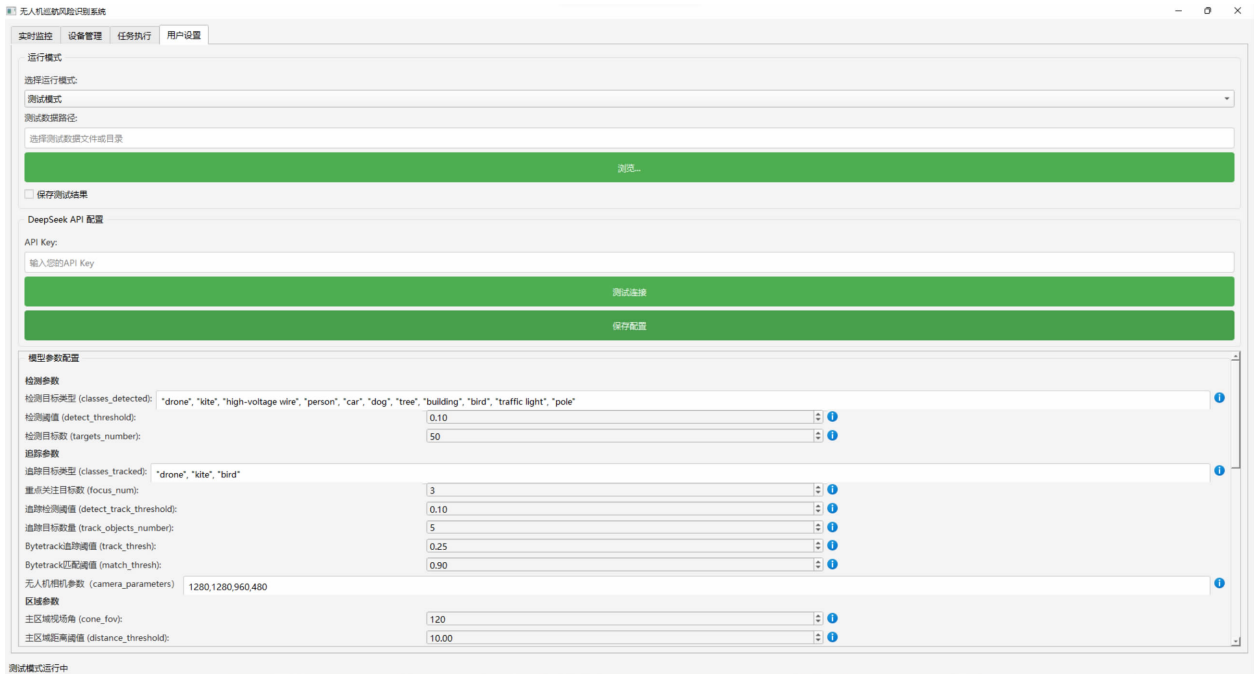


图 9 用户配置模块

Fig. 9 User configuration module

(2) 基于多模态大模型的模型架构和训练方式,建立了 3 个不同的模块来获取数据特征,实现多源数据融合。其中基于 Owl-ViT 模型的静态场景模块能有效识别飞行中的静态障碍物。基于 ByteTrack 算法的动态物体抓取模块,可快速获取飞鸟、其他无人机等动态物体的距离、速度、坐标等动态信息。基于点云的突发风险识别模块,可以捕捉点云障碍物信息,包括目标的距离、尺寸、体积、纵横比等,能够快速检测突发进入安全区域的障碍物。由 3 个不同模块生成的提示词引导 DeepSeek 输出的结果可详细展示巡航过程中的风险内容、风险等级并给出安全建议,表明通过提示词引导的 DeepSeek 模型可以理解飞行环境中的复杂风险要素。

(3) 本文提出的无人机巡航风险识别方法,可以通过多个模块及 DeepSeek 实现多源数据的充分融合,开放词汇的目标检测与 DeepSeek 结合可以帮助无人机更轻松地实现场景迁移,DeepSeek 的推理能力和知识储备有助于实现风险要素的充分理解。结果表明,所述方法在多个场景中的表现良好,且能够满足无人机巡航实时避障的需求。

(4) 开发无人机巡航风险识别系统,系统包括实时监控,设备管理,任务执行以及用户配置 4 个功能界面,可实时检测追踪结果形成风险判断,并将感知数据可视化。同时明确执行任务的设备和任务信

息,进一步辅助 DeepSeek 进行风险判断,为无人机巡航风险分析与预警功能实现提供了可实操性的平台支撑。

(5) 本文开发了基于多源数据与大模型的无人机巡航风险识别技术以及相对应的巡航风险识别系统,但研究中发现随着环境的复杂程度加剧,Owl-ViT 检测以及 DeepSeek 推理会造成延时,导致该方法的实时性不强,后续还需要对模型进行优化,提高模型运行速度。

参 考 文 献 :

References :

[1] 胡文晓,牟迪,李智,等.以关键技术问题创新引领低空经济发展对策研究[J/OL].航空学报,2024,https://doi.org/10.7527/S1000-6893.2024.31539.
HU Wen-xiao, MU Di, LI Zhi, et al. Research on countermeasures for promoting the development of low-altitude economy through innovation in key technical issues[J/OL]. Acta Aeronautica et Astronautica Sinica, 2024, https://doi.org/10.7527/S1000-6893.2024.31539.

[2] 贾永楠.低空空域无人系统交通管理方案初探[J].航空学报,2025,46(11):114-140.
JIA Yong-nan. A scheme for unmanned aerial system traffic management in low-altitude airspace[J]. Acta Aeronautica et Astronautica Sinica, 2025, 46(11): 114-140.

[3] 张学军,李诚龙,张志远,等.低空飞行系统实时风险管理能力构建:概念、挑战与技术[J].航空学报,2025,46(11):8-34.
ZHANG Xue-jun, LI Cheng-long, ZHANG Zhi-yuan, et al. Constructing in-time risk management capabilities for low-

- altitude aviation systems: Concepts, technologies, and challenges[J]. *Acta Aeronautica et Astronautica Sinica*, 2025, 46(11): 8-34.
- [4] 林嘉美霖,黄龙杨. 无人机运行风险评估及风险缓解研究[J]. *舰船电子工程*, 2024, 44(11): 156-160.
LIN Jia-mei-lin, HUANG Long-yang. Unmanned aerial system (UAS) operation risk assessment and risk mitigation research[J]. *Ship Electronic Engineering*, 2024, 44(11): 156-160.
- [5] 于立深. 低空经济有序发展的政府管制逻辑及法律保障[J]. *江西社会科学*, 2025, 45(3): 33-47, 206.
YU Li-shen. On government regulation and legal protection in the orderly development of low-altitude economy [J]. *Jiangxi Social Sciences*, 2025, 45(3): 33-47, 206.
- [6] ASWINI N, KUMAR E K, UMA S V. UAV and obstacle sensing techniques—a perspective[J]. *International Journal of Intelligent Unmanned Systems*, 2018, 6(1): 32-46.
- [7] SKARKA W, ASHFAQ R. Hybrid machine learning and reinforcement learning framework for adaptive UAV obstacle avoidance[J]. *Aerospace*, 2024, 11(11): 870.
- [8] 王家亮,董楷,顾兆军,等. 小型无人机视觉传感器避障方法综述[J]. *西安电子科技大学学报(自然科学版)*, 2025, 52(1): 60-79.
WANG Jia-liang, DONG Kai, GU Zhao-jun, et al. Review of obstacle avoidance methods for small UAVs using visual sensors[J]. *Journal of Xidian University (Natural Science)*, 2025, 52(1): 60-79.
- [9] SUN S, HAN Y Q, LI Y, et al. Analysis and research of intelligent distribution UAV control system based on optical flow sensor[C]//Springer. 2023 International Conference on Innovative Computing. Munich: Springer, 2023: 128-137.
- [10] MAO Y M, CHEN M, WEI X L, et al. Obstacle recognition and avoidance for UAVs under resource-constrained environments[J]. *IEEE Access*, 2020, 8: 169408-169422.
- [11] WANG D S, LI W, LIU X G, et al. UAV environmental perception and autonomous obstacle avoidance: A deep learning and depth camera combined solution[J]. *Computers and Electronics in Agriculture*, 2020, 175: 105523.
- [12] LIN H Y, PENG X Z. Autonomous quadrotor navigation with vision-based obstacle avoidance and path planning[J]. *IEEE Access*, 2021, 9: 102450-102459.
- [13] CHEN P H, LEE C Y. UAVNet: An efficient obstacle detection model for UAV with autonomous flight[C]//IEEE. 2018 International Conference on Intelligent Autonomous Systems (ICoIAS). New York: IEEE, 2018: 217-220.
- [14] FU Q, YANG Y H, CHEN X Y, et al. Vision-based obstacle avoidance for flapping-wing aerial vehicles[J]. *Science China Information Sciences*, 2020, 63(7): 170208.
- [15] 张午阳,章伟,宋芳,等. 基于深度学习的四旋翼无人机单目视觉避障方法[J]. *计算机应用*, 2019, 39(4): 1001-1005.
ZHANG Wu-yang, ZHANG Wei, SONG Fang, et al. Monocular vision obstacle avoidance method for quadcopter based on deep learning[J]. *Journal of Computer Applications*, 2019, 39(4): 1001-1005.
- [16] SINGLA A, PADAKANDLA S, BHATNAGAR S. Memory-based deep reinforcement learning for obstacle avoidance in UAV with limited environment knowledge[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2021, 22(1): 107-118.
- [17] 陈艺君,余莎莎,张学军. 城市低空场景下无人机运行对地风险量化评估[J]. *北京航空航天大学学报*, 2025, 51(3): 806-815.
CHEN Yi-jun, YU Sha-sha, ZHANG Xue-jun. Ground risk quantitative assessment for UAV operations in urban low-altitude scenarios[J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2025, 51(3): 806-815.
- [18] 贺洪波,徐晨晨,叶虎平. 无人机低空飞行障碍物环境风险评估方法研究:以京津新城为例[J]. *地理科学进展*, 2021, 40(9): 1503-1515.
HE Hong-bo, XU Chen-chen, YE Hu-ping. Environmental risk assessment of obstacles in low-altitude flight of unmanned aerial vehicle: Taking the Beijing-Tianjin New Town as an example[J]. *Progress in Geography*, 2021, 40(9): 1503-1515.
- [19] 张钦彤,王昱超,王鹤羲,等. 大语言模型微调技术的研究综述[J]. *计算机工程与应用*, 2024, 60(17): 17-33.
ZHANG Qin-tong, WANG Yu-chao, WANG He-xi, et al. Comprehensive review of large language model fine-tuning [J]. *Computer Engineering and Applications*, 2024, 60(17): 17-33.
- [20] 刘学博,户保田,陈科海,等. 大模型关键技术与未来发展方向:从 ChatGPT 谈起[J]. *中国科学基金*, 2023, 37(5): 758-766.
LIU Xue-bo, HU Bao-tian, CHEN Ke-hai, et al. Key technologies and future development directions of large language models: Insights from ChatGPT [J]. *Bulletin of National Natural Science Foundation of China*, 2023, 37(5): 758-766.
- [21] LIU P F, YUAN W Z, FU J L, et al. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing[J]. *ACM Computing Surveys*, 2023, 55(9): 1-35.
- [22] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[EB/OL]. arXiv, 2017, <https://arxiv.org/abs/1706.03762>.
- [23] 肖建力,邱雪,张扬,等. 交通大模型综述[J]. *交通运输工程学报*, 2025, 25(1): 8-28.
XIAO Jian-li, QIU Xue, ZHANG Yang, et al. Review on large language models in transportation[J]. *Journal of Traffic and Transportation Engineering*, 2025, 25(1): 8-28.
- [24] YIN S K, FU C Y, ZHAO S R, et al. A survey on multimodal large language models[J]. *National Science Review*, 2024, 11(12): nwae403.
- [25] BROWN T B, MANN B, RYDER N, et al. Language models

- are few-shot learners[EB/OL]. arXiv, 2005, <https://arxiv.org/abs/2005.14165>.
- [26] 严 昊, 刘禹良, 金连文, 等. 类 ChatGPT 大模型发展、应用和前景[J]. 中国图象图形学报, 2023, 28(9): 2749-2762.
YAN Hao, LIU Yu-liang, JIN Lian-wen, et al. The development, application, and future of LLM similar to ChatGPT [J]. Journal of Image and Graphics, 2023, 28(9): 2749-2762.
- [27] 刘 静, 郭龙腾. GPT-4 对多模态大模型在多模态理解、生成、交互上的启发[J]. 中国科学基金, 2023, 37(5): 793-802.
LIU Jing, GUO Long-teng. Inspiration of GPT-4 on multimodal foundation models in multimodal understanding, generation, and interaction[J]. Bulletin of National Natural Science Foundation of China, 2023, 37(5): 793-802.
- [28] 吴文峻, 廖星创, 赵金琨. DeepSeek 技术创新与通用人工智能发展趋势[J]. 科技导报, 2025, 43(6): 14-20.
WU Wen-jun, LIAO Xing-chuang, ZHAO Jin-kun. DeepSeek: Technological innovations and development trends toward artificial general intelligence[J]. Science & Technology Review, 2025, 43(6): 14-20.
- [29] 李 耕, 王梓烁, 何相腾, 等. 从 ChatGPT 到多模态大模型: 现状与未来[J]. 中国科学基金, 2023, 37(5): 724-734.
LI Geng, WANG Zi-shuo, HE Xiang-teng, et al. From ChatGPT to large multimodal model: Present and future[J]. Bulletin of National Natural Science Foundation of China, 2023, 37(5): 724-734.