

基于深度学习的移动机器人在非结构化物流场景中的语义障碍物识别

文豪 陈锐 李瑞雪

重庆移通学院 重庆 401420

摘要：在非结构化的物流场景中，移动机器人常常会遇到障碍物复杂、环境动态变化，使用传统几何识别方法已经不能满足高精度、智能化、智能化作业的需求。而通过深度学习实现的语义障碍物识别方法引入了图像的语义理解，机器人不仅可以识别物体的空间信息，还可以感知物体的类别及功能信息，物流场景中周转箱子、物流工人、仓储货架等。这种识别方法大大提升了机器人对复杂环境的感知及决策合理化程度，是保障机器人安全、高效、智能物流作业的关键技术，同时也是将智能物流从“会移动”向“明白在做什么”迈进的关键一步。正是由于这样的技术需求和行业的痛点，本文从深度学习下的移动机器人出发，研究非结构化物流环境下基于深度学习的移动机器人语义障碍物识别方法，希望能够给予实际项目的使用参考。

关键词：深度学习；移动机器人；非结构化物流场景；语义障碍物；识别要点

引言

随着智能制造、无人配送的发展，移动机器人的在仓储、货物分拣、物料搬运物流作业中的应用变得越来越多。而非结构化环境障碍物的复杂性表现为外形不规则、随机分布性广、动态行为不可预测，使得机器人路径规划和避障决策面临巨大挑战。几何物体的属性识别方法已不适合解决复杂环境中的不同障碍物的判定问题。在此背景下，以深度学习为核心的语义识别方法是机器人智能环境辨识技术的一个发展方向，它不仅能提高机器人的环境辨识能力，还将为其所组成的更智能、更灵活的物流自动化机械提供一定的技术基础，成为近几年学术界和工业界共同关注的热点和研究方向。

1 深度学习的定义

深度学习是机器学习的一种研究方法，模拟人脑神经网络结构和功能，构建深层的非线性变换模型，可以从巨量的数据中学习特征和完成高层次的抽象。不同于传统算法需要设计特征，深度学习可以从图像、语音、文本这些没有标记的基本信息中逐层学会从局部到整体的更加复杂的模式，学习判别数据的各种本质特征。深度学习的关键是深度神经网络。如卷积神经网络、循环神经网络、Transformer 等网络在学习过程中不断调整参数，使预测和分类的准确性进一步提升。深度学习技术目前已广泛应用

到计算机视觉、语音识别、自然语言处理等领域，其已成为人工智能的技术源泉，尤其在复杂问题中显示了深度学习的强大泛化能力和强大自学习优势^[1]。

2 移动机器人的原理

移动机器人的工作机制，以多设备协同与算法支撑为核心，通过感知、决策与执行三大核心环节的联动有序展开。它依托激光雷达、摄像头、超声波传感器等多模态感知设备实时采集环境信息，在此基础上完成环境地图构建与障碍物精准识别；结合动态更新的环境数据，通过 A* 算法、Dijkstra 算法及动态窗口法等路径规划技术开展导航决策，实现障碍主动规避与最优路径动态选择；最终由驱动系统与控制系统协同将决策指令转化为精准的机械运动，达成自主移动目标。整个运作过程需依赖嵌入式计算平台与核心算法的深度耦合，方能确保机器人在动态复杂环境中保持稳定运行状态。

3 语义识别在提升机器人理解能力中的价值

语义识别技术的核心价值，在于赋予机器人对环境中物体、场景及行为的语义级理解能力，进而显著提升机器人的认知深度与决策合理性。相较于传统障碍物识别仅聚焦物体的形状、位置等表层信息，语义识别能够进一步精准区分“箱子”与“人员”、“货架”与“临时障碍物”——这一能力使机器人不仅能完成基础的障碍规避，更能深度

理解各类物体的功能属性与潜在意图，例如精准判断物体是否具备可搬运属性、是否属于需优先避让的动态目标。此类语义理解能力在非结构化物流场景中具有突出价值，不仅能够支撑机器人实现更智能的路径规划、任务分配与人机协作，还能强化其应对复杂动态环境的适应能力，更为构建具备情境感知与逻辑推理能力的自主系统筑牢技术基础，更是推动移动机器人从“具备基础移动能力”向“深度理解作业逻辑”跨越的核心技术支柱^[2]。

4 传统障碍物识别方法的局限性

在非结构化环境的实际应用中，传统障碍物识别方法仍以几何特征提取与规则逻辑判断为核心，面对障碍物的类型多样性与状态不确定性时往往显得力不从心。这类方法对物体的形状、尺寸变化较为敏感，难以区分功能相近但形态存在差异的物体——例如物流场景中常见的纸箱与人形搬运设备。更关键的是，在光照条件波动、物体部分遮挡或动态干扰的场景下，传统方法易出现误检或漏检问题；而缺乏语义理解能力的缺陷，进一步导致机器人决策过程僵化。

5 基于深度学习的移动机器人在非结构化物流场景中的语义障碍物识别

5.1 关键技术

5.1.1 深度学习模型在视觉感知中的应用

在移动机器人的视觉感知环节，深度学习模型发挥着核心驱动作用。其中卷积神经网络（CNN）凭借多层卷积核的空间特征提取能力，展现出优异的表征性能：在ImageNet数据集上，ResNet-50以约2300万参数量达成高精度图像分类效果；YOLO系列模型则采用单阶段检测架构，如YOLOv8将目标定位与类别预测整合为统一的回归任务，并借助空间金字塔池化（SPP）结构拓展感受野，实现每秒30帧以上的实时处理速度。而Transformer架构通过自注意力机制捕捉图像中的长距离依赖关系，以DETR模型为例，其通过100个查询向量对图像全局信息进行建模，无需依赖锚框即可完成端到端目标检测。这些模型在移动机器人语义障碍物识别任务中，充分展现出强大的特征表征能力，显著提升了机器人在复杂环境下的感知鲁棒性与泛化性能。

5.1.2 语义分割与实例分割技术用于障碍物分类与定位

语义分割与实例分割技术，通过像素级精细标注为障碍物识别与定位提供了更精准的解决方案。语义分割

技术可对图像中每个像素进行预定义类别标注，例如将像素划分为“箱子”“人员”“地面”等类别，常用的DeepLabv3+模型通过融合空洞卷积（dilationrate=12,24,36）与ASPP模块，强化多尺度特征融合效果，有效提升了物体边界的识别清晰度；实例分割则进一步实现同一类别下不同个体的区分，典型方案如Mask R-CNN，其在Faster R-CNN架构基础上新增分支以预测二值掩膜，参数量约4000万，在COCO数据集上的mAP@0.5指标可达40%以上。这两项技术不仅让机器人能够准确识别障碍物类别，更能精准掌握障碍物的空间分布特征与独立个体信息，为非结构化物流场景下的安全避障决策与任务规划优化提供高精度感知保障^[3]。

5.1.3 多模态融合增强鲁棒性

在移动机器人感知系统中，多模态融合技术通过整合RGB-D图像、激光雷达点云与语义分割结果，为复杂环境下的感知鲁棒性提升提供了有效路径。其中，RGB-D数据可提供丰富纹理与颜色信息，以KinectV2为例，其 640×480 的分辨率能满足基础视觉感知需求；激光雷达则凭借 ≥ 10 万点/帧的点云密度实现高精度距离测量，进而构建环境几何结构，二者结合可有效弥补低光照条件下视觉感知失效的问题。而语义分割输出的类别概率图（如DeepLabv3+的处理结果），能进一步为障碍物赋予语义标签，形成“位置信息+类别属性”的双重约束。在融合策略上，基于注意力机制的设计（如Cross-ModalAttention模块）可通过动态调节权重参数，灵活分配各模态的贡献度——这一方案在非结构化物流场景中，能有效降低障碍物误检率，将定位准确率提升至95%以上，为机器人自主导航筑牢稳定可靠的感知基础。

5.2 系统架构设计

5.2.1 数据采集与标注流程

高质量语义障碍物识别模型的构建，离不开科学规范的数据采集与标注流程。实践中需在真实物流场景中部署多传感器系统，该系统包含RGB-D相机、360°激光雷达与IMU，可同步采集 $\geq 1280 \times 720$ 分辨率的高清晰图像、 ≥ 50 万点/帧的稠密点云，并完成时间戳同步。采集过程需重点覆盖光照变化、天气波动、人流密度差异等典型工况，以确保数据的场景代表性与多样性。进入标注阶段后，借助LabelMe、CVAT等半自动标注工具开展像素级语义标注与实例分割标注，每帧图像平均需标注10–20个目标实例，

且标注质量以 IoU 交并比 ≥ 0.7 为阈值进行严格把控。这一完整流程不仅能保障训练数据的代表性与一致性，更为后续模型在复杂非结构化环境中实现鲁棒识别提供了可靠的数据支撑。

5.2.2 基于深度神经网络的语义障碍物检测模块

针对语义障碍物检测的深度神经网络模块，通常以轻量化骨干网络为基础构建，如选用 MobileNetV3、EfficientNet-B0，并搭配 SEBlock、CBAM 等注意力机制以强化特征提取能力。具体流程中，输入图像经预处理后送入特征提取器，生成 C3 - C5 等多尺度特征图，对应通道数依次为 256、512、1024；随后通过 FPN 特征金字塔网络融合高层语义信息与低层细节特征，提升特征表征的完整性。检测头采用 Anchor-free 设计（典型如 YOLOv8 的 PANet 结构），直接输出障碍物的边界框坐标与类别概率，且模型参数量严格控制在 1000 万以内，以适配机器人边缘计算设备。训练阶段通过交叉熵损失与 Focal Loss 协同平衡正负样本，实现高精度语义分类与定位——在非结构化物流场景测试中，其 mAP@0.5 指标稳定达到 $\geq 85\%$ ，为移动机器人提供了实时、可靠的障碍物感知能力^[4]。

5.2.3 实时推理优化与边缘计算部署策略

实时推理优化与边缘计算部署的核心目标，是提升语义障碍物检测模块在移动机器人上的运行效率与响应时效性。实践中通过结构化模型剪枝保留 80% 以上通道，搭配 INT8 精度转换可使推理延迟降低 40%；同时引入知识蒸馏技术，以 ResNet-101 作为教师模型、MobileNetV3 作为学生模型，进而将模型体积压缩至 50MB 以下，完美适配 NVIDIA Jetson AGX Orin 等嵌入式平台——该平台 275TOPS 的算力可充分支撑模型高效运行。部署阶段借助 TensorRT 或 ONNX Runtime 完成算子融合与内存优化，实现每秒 30 帧以上的推理速率，确保延迟控制在 33ms 以内；在此过程中，结合多线程调度与“CPU+GPU+NPU”异构计算模式提升硬件资源利用率。这一系列优化策略，为复杂物流场景下低延迟、高可靠的语义感知提供了保障，充分满足机器人自主导航的实时性需求。

5.3 实验验证与性能评估

5.3.1 在模拟与真实非结构化物流环境中测试识别准确率与响应速度

在验证语义障碍物识别模型的准确率与响应速度时，

研究以某智能仓储机器人在电商仓库的实际部署为场景，先在 Sim4Real 仿真平台完成模型预训练，再将其迁移至京东亚洲一号仓等真实环境，通过部署 YOLOv8-seg 语义检测模块，针对动态货架、移动叉车、穿行人员共存的复杂背景开展多轮测试。测试结果显示，模型平均识别准确率（mAP@0.5）达 87.3%，误检率控制在 5% 以下，响应延迟稳定在 28 - 35ms 之间，完全满足实时避障的技术需求。对比静态环境测试数据，真实场景中因光照突变、物体遮挡及噪声干扰，准确率出现约 6% 的下降；但通过引入自适应归一化层与多帧融合策略，模型鲁棒性得到显著提升。

5.3.2 对比传统方法与语义识别方案的差异

以某物流园区 AGV 在分拣区的避障任务为研究对象，传统方案依赖激光雷达点云聚类与几何特征匹配，将欧式距离阈值设为 0.5m，仅能判断障碍物是否存在，无法区分“纸箱”与“人员”等不同目标，导致 AGV 频繁出现误停现象，甚至存在碰撞风险。反观采用语义识别方案的系统，其搭载基于 DeepLabv3+ 的 RGB-D 融合模型，在相同场景中实现像素级分类与实例分割，可准确识别人员、托盘、作业设备等目标类别。测试数据显示，传统方法在分拣区货物转运的复杂交叉场景下，误报率高达 22%，且响应延迟波动较大，平均达 45ms；而语义识别方案将误报率降至 6%，响应延迟稳定在 32ms 以内，更能依据语义信息判断目标属性。如对人员采取绕行策略，而非简单强制停止。这一对比案例表明，语义识别不仅提升了环境感知精度，更赋予机器人情境理解能力，推动其决策逻辑从“被动避障”升级为“智能避让”，显著增强了 AGV 在真实非结构化环境中的作业安全性与效率。

5.3.3 分析复杂遮挡、光照变化等干扰因素下的稳定性

某港口自动化堆场的实测中，移动机器人需在集装箱高密度堆放、阳光直射与阴影交替的环境下运行。传统方法因依赖单一 RGB 图像特征，强光下易出现过曝、弱光环境下纹理信息模糊，导致障碍物漏检率高达 35%；语义识别方案则通过融合 RGB-D 与激光雷达数据，搭配 CBAM 等注意力机制强化关键区域感知权重，即便遮挡率超 40%（如部分箱体被其他货物遮挡），仍能维持 82% 以上的识别准确率。例如，当叉车从侧面突然驶入视野并遮挡部分货架时，模型借助多模态信息互补与上下文推理，仍能准确判断其为“可通行障碍”而非“静态障碍”，有效规避误停。

该案例印证,语义识别在复杂干扰场景下具备更优的鲁棒性与情境适应能力,堪称提升机器人在动态不确定物流环境中稳定运行的关键技术支撑^[5]。

6 结束语

综上,语义障碍物识别作为衔接感知与决策的核心环节,正逐步重塑移动机器人在非结构化环境中的行为逻辑。未来需聚焦模型轻量化、跨场景泛化能力及多源信息融合策略,进一步提升识别效率与鲁棒性;与此同时,还需重点关注实际部署中的实时性、安全性与可解释性问题,助力技术从实验室研究迈向规模化工程应用。

参考文献:

[1] 漆亮,杨逸飞,黄阳.基于激光跟踪仪的移动机器人精度测试 [J].装备制造技术,2024,(06):65–68.

[2] 王春茂.机器视觉技术在物流领域的应用与创新概述 [J].物流技术与应用,2024,29(06):96–101.

[3] 康丽丽.智能仓储物流移动机器人任务协同调度方法研究 [J].现代计算机,2024,30(11):40–44.

[4] 汪传雷,杨依婉,秦琴.面向新质生产力的物流机器人技术融合研究 [J].重庆工商大学学报(社会科学版),2025,42(03):96–111.

[5] 刘元华,李超群,郭乙运.仓储物流机器人多启发蚁群路径规划与避障 [J].机械设计与制造,2024,(11):296–300.

基金项目: 2024 年重庆市教育委员会科学技术研究项目,项目名称: 复杂物流环境下移动机器人障碍物识别及导航研究 (项目编号: KJQN202402405)