DOI:10.19651/j.cnki.emt.2417518

基于 D3QN 的目标驱动移动机器人自主导航方法*

卢赵清 王宏伟 何 丽 司盼召 陈耀华
(新疆大学机械工程学院 乌鲁木齐 830017)

摘 要:在未知或危险环境中(如应急救灾、抢险救援),传统导航方法因无法预先获得先验地图和位置信息,难以实现特定目标的导航。本文提出了一种基于竞争双深度 Q 网络(D3QN)的目标驱动移动机器人自主导航方法。该方法的跨模态融合模块对不同模态特征动态加权融合,在整合观测数据的同时充分捕捉环境信息,增强了对环境的感知能力。在此基础上,设计了一种通用的目标驱动导航方法,使用 YOLOv5 识别特定目标(如火焰、烟雾)并获取其位置,用识别出的目标位置替代深度强化学习导航中的预设位置点,实现自主导航至特定目标。仿真实验结果表明,本文方法在导航成功率等指标上具有显著优势,在简单、复杂和动态场景中,成功率分别提高了 9%、27%和 38%。此外,在简单仿真环境中训练的模型,能够直接部署在复杂的仿真环境和真实场景中,表现出良好的泛化能力。

关键词:深度强化学习;D3QN 算法;多模态融合;识别与定位;目标驱动导航
中图分类号:TP242.6;TN711 **文献标识码**:A **国家标准学科分类代码**:510.80

D3QN-based target-driven autonomous navigation for mobile robots

Lu Zhaoqing Wang Hongwei He Li Si Panzhao Chen Yaohua
(School of Mechanical Engineering, Xinjiang University, Urumqi 830017, China)

Abstract: In unknown or hazardous environments (such as emergency rescue and disaster relief), traditional navigation methods struggle to achieve specific target navigation due to the inability to obtain prior maps and location information. This paper proposes a target-driven autonomous navigation method for mobile robots based on dueling double deep Q network (D3QN). The cross-modal fusion module of this method dynamically weights and integrates features from different modalities, effectively consolidating observational data while fully capturing environmental information. This significantly enhances the capability to perceive the environment. Building on this, a general target-driven navigation approach is designed, where YOLOv5 is used to recognize specific targets (such as flames or smoke) and obtain their locations. The identified target locations are used to replace predefined waypoints in deep reinforcement learning-based navigation, enabling autonomous navigation to specific targets. Simulation results show that the proposed method has significant advantages in the navigation success rate and other indicators. In simple, complex, and dynamic scenarios, the success rates increased by 9%, 27%, and 38%, respectively. Moreover, the model trained in a simple simulation environment can be directly deployed in more complex simulation environments and real-world scenarios, exhibiting strong generalization capability.

Keywords: deep reinforcement learning; D3QN algorithm; multimodal fusion; identification and location; target-driven navigation

0 引 言

在抢险救灾等特殊作业环境中,移动机器人面临的导航环境通常是未知的,且无法提前构建地图。因此不依赖先验地图信息、能够实时规划和自主决策的无地图导航方法成为了新的研究热点[1-4]。深度强化学习(deep reinforcement learning,DRL)因其端到端的特性,可以直接从传感器感知信息(如 RGB 图像[5]、深度图像[6]和激光雷达[7-8])生成控制命令。为无地图导航提供了一种有效的解决方案,尤其适用于未知环境下的自主导航任务。

收稿日期:2024-12-02
*基金项目:新疆维吾尔自治区自然科学基金(2022D01C392)、国家自然科学基金(62063033)、新疆维吾尔自治区重点研发计划项目(2022B01050-2)资助

尽管如此，单一模态的数据在复杂环境中往往表现不足。例如，激光雷达方法在可转移性方面具有优势，但其信息无法全面描述复杂的场景；而视觉传感器不仅成本较低，还能提供更为丰富和详细的环境描述，但在动态环境中的鲁棒性较差。因此，传感器融合或多模态数据的结合成为提升机器人感知能力和导航性能的有效策略。然而，如何有效整合多模态信息仍然是一个挑战。

目前，大多数多模态融合方法采用人工设定的固定参数。例如，Huang 等[9]提出的多模态感知框架，通过拼接语义分割图和激光雷达数据，增强智能体在动态环境中的避障能力。Song 等[10]提出的基于辅助任务的多模态 DRL 方法(MDRLAT)，通过对深度图像和激光数据进行线性融合，进一步优化了避障控制策略。这些方法虽然能够有效地利用不同模态的数据，但由于参数固定，它们在动态环境中缺乏足够的灵活性和适应性，泛化能力不强。随着研究的不断深入，越来越多的工作开始采用注意力机制进行融合，例如，Han 等[11]引入了自我状态注意单元，用于融合 2D 激光雷达和单目相机数据，以实现更加灵活和精准的导航任务。Ji 等[12]提出一种多传感器信号的机器人导航主动异常检测方法(PAAD)，使用具有残差的多头注意力对相机和激光雷达融合进行特征级融合。尽管这些方法在多模态融合方面取得了进展，但仍然缺乏根据环境变化动态调整融合权重的能力，限制了其在复杂场景中的应用。

在特殊场景(如抢险救灾或复杂环境巡检)中，移动机器人需要在未知环境中探索并导航至特定目标(如火焰、烟雾)。现有目标驱动导航方法虽在自主探索方面取得了一定进展，但在实际应用中仍存在局限性。例如，Zhu 等[13]首次提出的目标驱动视觉导航方法，实现了在 3D 仿真环境中导航至不同目标，但其在真实环境中的泛化能力较弱，限制了实际应用。Zhelo 等[14]开发了一种基于 DRL 的好奇心探索导航方法，智能体在不依赖环境地图的情况下，能够自主导航至预设目标位置，但缺乏对特定目标的导航。Cimurs 等[15]提出了一种轨迹引导的目标导航方法，根据可用数据选择最佳航路点，但同样需要预先设定目标坐标，难以满足导航至特定目标的需求。

综上所述，尽管基于深度强化学习的移动机器人导航方法在提高自主探索能力方面取得了一定进展，但仍存在以下两个问题：1)多模态数据在融合方式上仍存在改进的空间，以更充分地捕获其互补信息；2)现有目标驱动导航方法在泛化能力弱，且在导航前需预设目标坐标，难以满足特殊场景中目标驱动的导航需求。为此，本文提出了一种目标驱动的多模态感知移动机器人自主导航方法。首先，融合了 RGB-D 相机生成的深度图和伪激光数据，利用两种模态的互补性提升对复杂环境的感知能力。其次，针对传感器视场角为 70°的限制，设计了一种多层卷积长短期记忆网络(ConvLSTM)结构，不仅扩展了视场角，还捕捉了时间维度上的环境变化，进一步增强感知能力。为了提升多模态融合的适应性，本文提出一种动态加权的跨模态融合模块，通过学习模态特征的重要性，自适应调整权重，从而提升导航性能。此外，为实现目标驱动导航，引入目标检测算法，使机器人能够实时识别特定目标(如火源或烟雾)并更新导航目标位置，有效克服了依赖预设目标坐标的局限性，在真实场景中实现目标驱动的自主导航。

1　深度强化学习与 D3QN 算法

1.1　深度强化学习

深度强化学习是将深度学习与强化学习结合的技术。通过深度神经网络处理复杂的、高维度的数据，并从中学习最优策略或值函数，使智能体在环境中通过试错自主学习，以最大化长期奖励。强化学习将任何决策者定义为一个智能体，将智能体外部一切事物定义为环境。智能体学习的目标是最大化累计奖励，通过与环境交互获得即时奖励作为训练的反馈信号。具体的流程如图 1 所示。

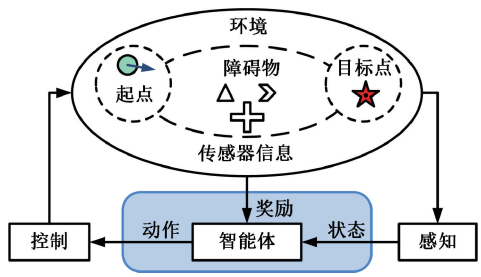


图 1　基于 DRL 的导航系统

Fig. 1　Navigation system based on DRL

通过引入折扣因子 $\gamma \in [0,1)$，可以将累计奖励表示为：

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \tag{1}$$

状态 s 的值函数为 $V_\pi(s)$，状态-动作对 (s,a) 的值函数是 $Q_\pi(s,a)$，用于评估智能体使用策略获得的预期长期回报。

$$V_\pi(s) = \mathbb{E}_\pi[R_t \mid s_t = s] \tag{2}$$

$$Q_\pi(s,a) = \mathbb{E}_\pi[R_t \mid s_t = s, a_t = a] \tag{3}$$

利用式(1)，$V_\pi(s)$ 和 $Q_\pi(s,a)$ 以递归形式建立状态 $s_t = s$ 与 $s' = s_{t+1}$ 之间的关系：

$$V_\pi(s) = \sum_a \pi(s,a) \sum_{s'} P^a_{ss'}[R^a_{ss'} + \gamma V_\pi(s')] \tag{4}$$

$$Q_\pi(s,a) = \sum_{s'} P^a_{ss'}[R^a_{ss'} + \gamma Q_\pi(s',a')] \tag{5}$$

式中：

$$P^a_{ss'} = P(s_{t+1} = s' \mid s_t = s, a_t = a) \tag{6}$$

$$R^a_{ss'} = \mathbb{E}[r_{t+1} \mid s_t = s, s_{t+1} = s', a_t = a] \tag{7}$$

式(4)和(5)被称为贝尔曼方程。通过动态规划获得贝尔曼方程近似解，从而得到当前值函数，智能体通过优化价值函数不断改进策略。

1.2　D3QN 算法

D3QN 融合了 DDQN[16]和决斗 DQN[17]的优点。

DDQN 技术由 DQN 技术发展而来,包含一个在线网络,使用参数 θ 计算在线 Q 值 $Q(s_t,a_t;\theta)$;一个目标网络,使用参数 θ^- 计算目标 Q 值 y_t:

$$y_t = r_t + \gamma \max_{a_{t+1}} Q(s_{t+1},a_{t+1};\theta^-) \tag{8}$$

为解决 Q 值的过高估计,DDQN 将式(8)中目标网络取最大 Q 值分解为动作选择和动作评价。目标 Q 值 y_t 可表示为:

$$y_t = r_t + \gamma Q(s_{t+1},\operatorname{argmax}_{a_{t+1}} Q(s_{t+1},a_{t+1};\theta);\theta^-) \tag{9}$$

D3QN 将 Q 网络划分为动作优势和状态值。Q 值可以表示为:

$$Q(s_t,a_t;\theta_C,\theta_V,\theta_A) = V(s_t;\theta_C,\theta_V) + A(s_t,a_t;\theta_C,\theta_A) - \frac{1}{N}\sum_{a'} A(s_t,a';\theta_C,\theta_A) \tag{10}$$

式中:θ_C、θ_V、θ_A 分别表示共同层、价值流和优势流的参数,N 为动作总数,$V(s_t;\theta_C,\theta_V)$ 表示状态值函数的估计,优势函数估计的计算公式为:

$$A(s_t,a_t;\theta_C,\theta_A) - \frac{1}{N}\sum_{a'} A(s_t,a';\theta_C,\theta_A) \tag{11}$$

2 目标驱动的多模态移动机器人导航

2.1 目标驱动的导航框架

针对当前视觉驱动导航在真实环境中应用所面临的挑战,本文提出了一种结合仿真训练与实际场景应用的方法。首先,基于强化学习的 D3QN 算法设计了一种跨模态注意力的位置点导航框架,并在仿真环境中训练机器人的导航能力。为进一步增强该方法在特定任务中的实用性,在位置点导航框架的基础上引入目标检测技术。通过这种方法,机器人不仅能够自主导航,还能实时识别并定位特定目标(如火焰、烟雾)位置,并根据这些目标动态调整路径。

具体导航过程如图 2 所示。在导航过程中,机器人首先根据预设的目标点进行路径规划,如未识别到特定目标,机器人将朝着既定目标点行驶。一旦检测到特定目标(如火焰、烟雾),系统将实时解算该目标的位置,并将该位置信息反馈至导航系统,随后动态调整路径以实现目标驱动的导航。

2.2 深度强化学习模块

1)参数设置

由于传感器视场角的限制,智能体在训练过程中无法完感知周围环境,因此将训练过程定义为部分可观测马尔可夫决策过程,并采取 D3QN 算法来实现深度强化学习避障策略。

观测空间:对智能体的观测,使用 70°视场角的深度图和伪激光数据。

动作空间:由线速度 υ 和角速度 ω 组成,是 10 个离散控制命令,详细数值参考文献[10]。设置机器人不能向后移动,因为传感器无法覆盖机器人的后部区域。

奖励函数:为了控制机器人在运行过程中安全运行避免碰撞并最终到达目标点,设计的奖励函数 R 如下:

$$R = \begin{cases} R_{\text{crash}}, \text{碰撞} \\ R_{reach}, \text{到达目标点} \\ (d_{t-1} - d_t)\cos(\omega_t) - ct, \quad \text{其他} \end{cases} \tag{12}$$

2)伪激光测量

本文使用伪激光测量的目的是以低成本的方式实现高效导航,采用伪激光技术替代 2D 激光雷达,以达到相同甚至更优的导航效果。具体来说,利用 RGB-D 相机获取深图像后,借助 ROS 中的 depthimage_to_laserscan 功能包,将深度图转化为伪激光数据。

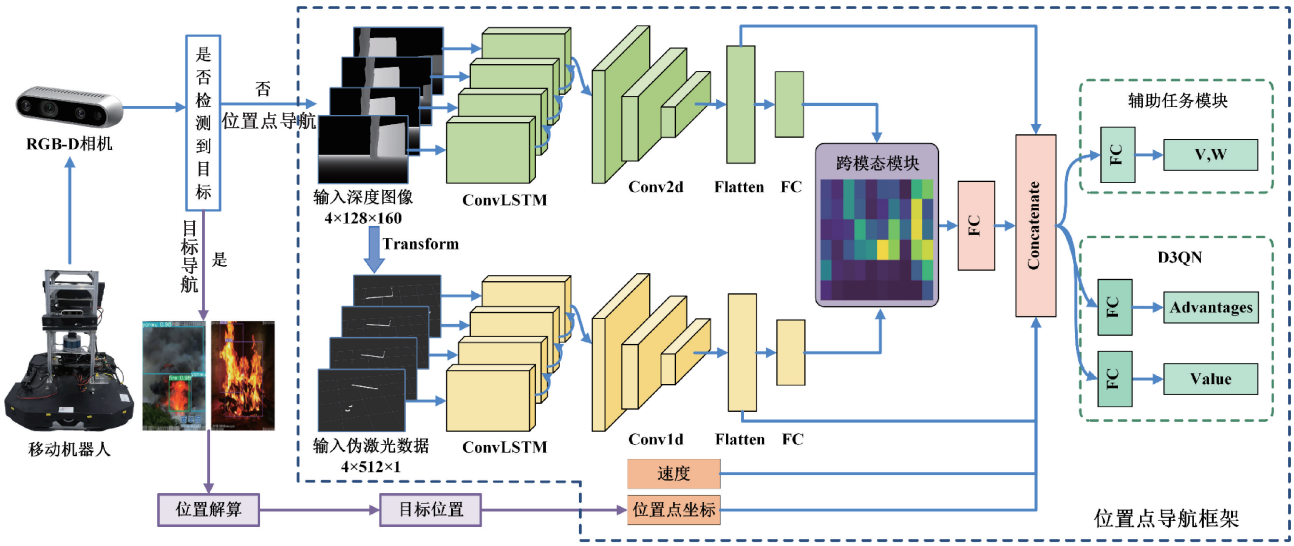


图 2 目标驱动的导航框架

Fig. 2 Goal-driven navigation framework

仿真环境下的 2D 激光雷达数据与伪激光雷达数据如图 3 所示，图 3(a)是当前时刻 RDB-D 相机拍摄的 RGB 图像，图 3(b)为此刻的深度图像，图 3(c)中的红色表示 2D 激光雷达数据，白色表示伪激光雷达数据。从图 3 中可以看出，由于伪激光数据是由深度图生成，其在检测范围和精度方面与 2D 激光雷达数据存在一定差距。但伪激光数据与深度图像更加匹配和一致，避免了时间戳不同步和坐标系转换等问题。此外，由于在机器人视角中障碍物通常出现在相机的中间范围，深度图与伪激光信息的融合增强了对中间障碍物的感知能力。

(a) RGB图像　(b) 深度图　(c) 2D激光雷达和伪激光对比

(a) RGB image　(b) Depth imahe　(c) Comparison of 2D LiDAR and pseudo-LiDAR

图 3　激光雷达数据和伪激光雷达数据对比

Fig. 3　Comparison of LiDAR data and pseudo-LiDAR data

3)网络结构

如图 2 所示，在网络特征提取部分设计了一种多层 ConvLSTM 结构，以缓解有限视场角和丰富特征提取的问题。ConvLSTM 是一种专门为处理时空数据而设计的网络，它能够同时提取空间特征并建模时空依赖关系，特别适用于需要处理时间和空间信息的任务。在本方法中，连续四帧的深度图像和伪激光数据被输入多层 ConvLSTM 模块，用于提取空间和时间特征，并整合时间序列中的动态变化和时序关系。这有助于智能体根据当前状态和历史状态做出更合理的决策。具体来说，将深度图和伪激光数据输入两个独立的多层 ConvLSTM 模块，然后通过卷积层进行特征提取，最后将智能体获得的多模态信息展平，所有卷积层的激活函数都是 ReLu 激活函数。

4)跨模态融合模块

为了充分捕捉传感器不同模态的互补信息，采用特征级深度图 f_{depth} 和伪激光雷达 $f_{pseudo\text{-}LiDAR}$ 融合作为输入，设计了跨模态注意力，将提取的特征信息进行融合，具体如图 4 所示。

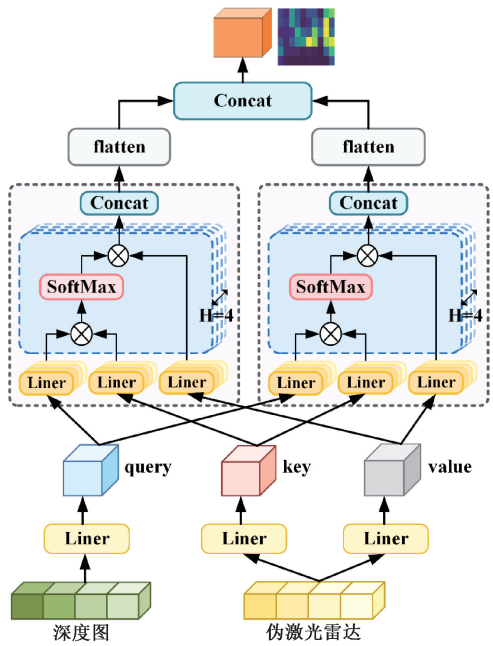


图 4　跨模态融合模块

Fig. 4　Cross-modal fusion module

跨模态注意力的核心思想是通过计算 Query(查询向量)、Key(键向量)和 Value(值向量)之间的动态相关性权重，自适应地调整模态特征的重要性。具体而言，首先将输入特征通过线性变换映射到注意力子空间：

$$\begin{cases} \boldsymbol{Q}_i = f_{depth}\boldsymbol{W}_i^{Q} \\ \boldsymbol{K}_i = f_{pseudo-LiDAR}\boldsymbol{W}_i^{K} \\ \boldsymbol{V}_i = f_{pseudo-LiDAR}\boldsymbol{W}_i^{V} \end{cases} \tag{13}$$

式中：$\boldsymbol{W}_i^{Q}$、$\boldsymbol{W}_i^{K}$、$\boldsymbol{W}_i^{V}$ 分别为第 i 个注意力头的权重矩阵，用于将输入特征映射到低维子空间，便于计算模态间的相似性。$\boldsymbol{Q}$ 代表深度图像特征，作为 Query；$\boldsymbol{K}$ 和 $\boldsymbol{V}$ 分别代表伪激光雷达特征，作为 Key 和 Value。

通过缩放点积注意力计算 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 之间的相似性，并使用缩放因子 $\sqrt{d_k}$ 归一化，得到未归一化的注意力权重：

$$Score_i = \frac{\boldsymbol{Q}_i\boldsymbol{K}_i^{\mathrm{T}}}{\sqrt{d_k}} \tag{14}$$

使用 softmax 将相似度分布归一化为动态权重 α：

$$\alpha_i = softmax\left(\frac{\boldsymbol{Q}_i\boldsymbol{K}_i^{\mathrm{T}}}{\sqrt{d_k}}\right) \tag{15}$$

其中，α_i 表示 Query 和 Key 之间的动态相关性权重，权重越大表明当前模态特征对融合过程的贡献越大，反之则较小。接着，将动态权重 α_i 作用于 Value $\boldsymbol{V}_i$，得到当前注意力头的输出为：

$$head_i = \alpha_i \boldsymbol{V}_i \tag{16}$$

通过这种动态权重调整机制,模型能够突出与深度图像特征相关性更强的伪激光雷达特征,同时弱化冗余或无关特征。多个注意力头在不同子空间中并行计算权重分布,以捕获深度图像与伪激光雷达特征之间的多层次、细粒度关联。所有注意力头的输出拼接起来,并通过线性变换 W^O 映射回目标特征空间,形成融合特征表示如下:

$$\begin{cases} MultiHead(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = Concat(head_1,\cdots,head_h)W^O \\ head_i = Attention(\boldsymbol{Q}\boldsymbol{W}_i^Q,\boldsymbol{K}_i^K,\boldsymbol{V}_i^V) \\ h = 4 \end{cases} \tag{17}$$

为进一步提升融合特征的表达能力,本文将两次独立的注意力输出进行拼接,形成最终的跨模态融合特征表示如下:

$$Cross-Modal\ \ Attention(f_{depth},f_{pseudo-LiDAR}) = Concat\begin{bmatrix} MultiHead(f_{depth},f_{pseudo-LiDAR}) + \\ MultiHead(f_{pseudo-LiDAR},f_{depth}) \end{bmatrix} \tag{18}$$

融合后的特征被输入到深度强化学习网络中,用于移动机器人的决策与避障任务。该机制不同于拼接或人工调参的方法,通过动态调整权重,自适应地分配模态特征的重要性,在多个子空间中捕获深层次的模态互补信息,从而提高在复杂环境中的避障能力。

2.3 识别与定位

1)基于 YOLOv5 的火焰识别

鉴于真实火焰和烟雾存在一定的危险性,本文选择采用仿真火焰及烟雾作为目标对象。通过相机获取仿真火焰数据集,数据集经数据增强后共 3 600 张图像,使用 Labelimg 工具进行标注,并按 8∶1∶1的比例随机划分为训练集、验证集和测试集,分别为 2 880 张、360 张和 360 张图像。由于 YOLOv5 是一种高效的实时目标检测模型,能够在速度与精度之间实现良好平衡,并易于部署,使用 YOLOv5 作为检测方法。识别效果如图 5 所示。

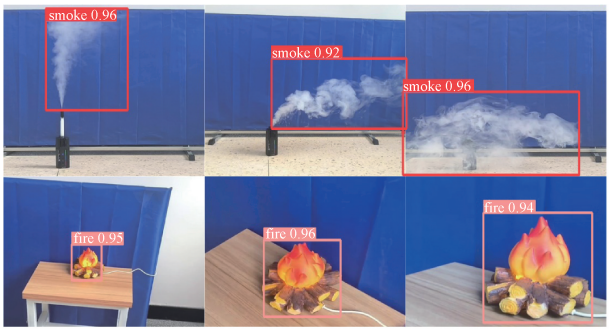


图 5 火焰及烟雾识别

Fig. 5 Flame and smoke recognition

2)基于 RGBD 相机定位

YOLOv5 模型能够实现对火焰的快速识别,输出火焰的预测框(x,y,w,h),其中(x,y)为预测框的中心坐标,w和h是预测框的宽度和高度。选择预测框的中心作为目标点,但需要获得相机到火焰的距离,通过深度相机 Realsense D435i 将 RGB 相机捕获的的彩色图像中火焰和烟雾的位置信息映射到红外相机的深度图中,如图 6 所示。

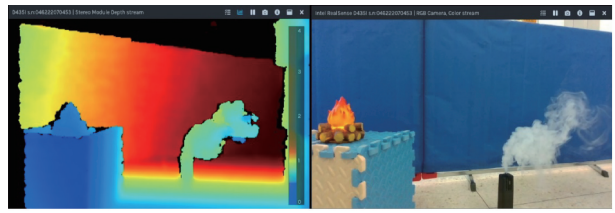


图 6 仿真火焰及烟雾深度图

Fig. 6 Simulation depth image of flame and smoke

如此得到火焰在坐标系中的真实距离坐标,测得的结果如图 7 所示。Realsense D435i 深度相机已经具备内参标定结果,保持原有的标定关系,不重新标定。

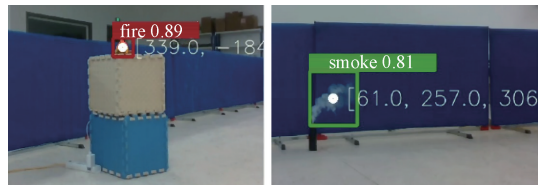


图 7 火焰及烟雾的识别定位

Fig. 7 Recognition and localization of flame and smoke

3 仿真实验

3.1 实验设置

在 Gazebo 仿真平台中构建环境以训练智能体,如图 8 所示。智能体首先在简单环境中进行训练,然后在简单、复杂和动态环境中测试其导航性能。在仿真环境中使用 Turtlebot2 智能体与环境进行交互。如果智能体成功到达目标点、与障碍物发生碰撞或运行超过 500 步,则该回合结束。训练使用 TensorFlow 作为编程环境,运行在配置有 Intel(R) Xeon(R) Gold CPU、128GB 内存和 RTX 5000 GPU(16 GB 显存)的服务器上。

为证明所提方法的有效性,使用不同方法进行对比。

1)multi[9]:深度图和 2D 激光雷达特征通过拼接的方式输入到 D3QN 模型。

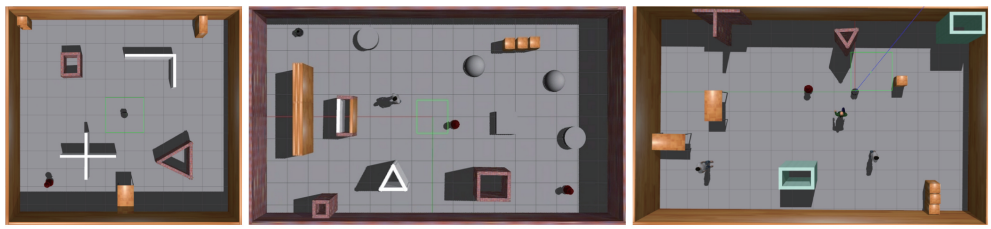
2)MDRLAT[10]:深度图和 2D 激光雷达特征通过线性融合的方式输入到 D3QN 模型。

3)PAAD[12]:深度图和 2D 激光雷达特征通过多头注意力融合的方式输入到 D3QN 模型。

4)本文:深度图和伪激光特征通过跨模态融合方式输入到 D3QN 模型。

使用以下指标评估导航算法的性能:

1)成功率:定义为 $SR = (1/M)\sum_i^M S_i$,其中,S_i 是第 i 回合成功与否的二进制表示,M 是回合数,定义 500 步内导航到目标位置为一个回合成功。



（a）简单环境　（b）复杂环境　（c）动态行人环境
（a）Simple environment　（b）Complex environment　（c）Dynamic pedestrian environment

图 8　仿真实验中的训练和测试环境

Fig. 8　Training and testing environments in simulation experiments

2)碰撞率:计算方法与成功率相同。

3)超时率:计算方法与成功率相同。

4)平均回合奖励:五个回合的平均累积奖励。智能体训练过程中,每 2 000 步评估一次。

5)平均步数:在多次重复导航实验中,导航到达目标点的步数与实验次数之比,代表导航中动作策略的效率。

6)平均时间:在多次重复导航实验中,成功到达目标点时间与时间次数之比,代表导航的效率。

3.2　训练实验

所有的方法在图 8 的简单环境中训练 40 万步,智能体训练过程中的学习曲线如图 9 所示。批量参数 N_B 和折扣因子 γ 分别设置为 32 和 0.99。此外,动作选择依赖于 ε-greedy 策略。ε 的初始值被设置为 0.1,并在前 20k 步中现行衰减到 0.000 1。

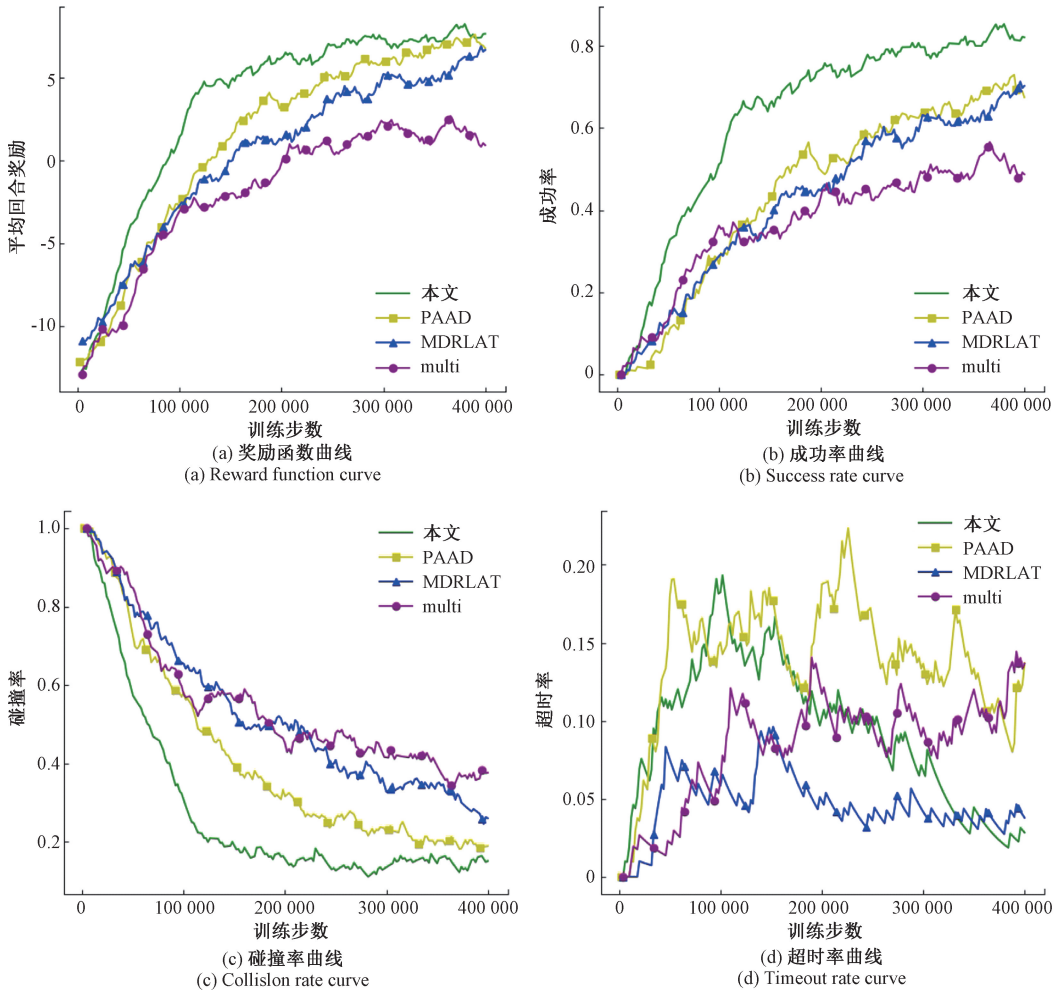


图 9　训练过程中各项指标曲线

Fig. 9　Curves of various metrics during training

从图 9 中可以看出,所提方法在奖励、成功率、碰撞率和超时率方面表现最好。具体来说,所提方法在训练初期表现出显著的优势,能够将碰撞率转化为超时率,使机器人学会避开障碍物,减少碰撞,最终成功到达目标点,从而达到了较高的成功率。在 D3QN 算法中,使用不同的视场角和融合方式进行训练,包括拼接的融合方式(multi),线性融合方式(MDRLAT),多头注意力融合方式(PAAD)及跨模态融合方式(本文方法)。本文方法虽然具有较小的视场角,但通过多层 ConvLSTM 弥补视场角,同时设计了一个跨模态融合的方式对提取到的特征进行融合,使得机器人能够有效避开障碍物,减少碰撞。相比之下,拼接和线性融合方法在感知能力上的表现较弱,这两种方法均在碰撞率上存在较高值。特别是在训练的后期,拼接融合方式仍然呈现较高的超时率,导致智能体无法成功到达目标点。此外,尽管多头注意力融合方式在奖励上取得了较好的表现,但其成功率较低。在导航任务中,导航成功率是最关键的性能指标。从图 9 中的数据可以看出,所提方法在成功率上取得了 82.38%的最佳表现,而拼接方法、线性融合方法和多头注意力融合方法分别仅达到了 48.96%、64.52%和 63.24%。这些结果表明,所提方法显著优于其他融合方式,在提升导航成功率方面表现出色。

3.3 测试评估

为验证所提方法的泛化性,在简单、复杂和动态行人环境中对所有方法训练后的模型进行了跨目标和跨场景泛化测试,每个方法重复 5 次独立实验。智能体从原点出发,导航到非障碍物区域的随机目标,每种方法每次实验生成 100 个评估结果。成功率、碰撞率和超时率如表 1 所示。

表 1 随机目标点的导航性能对比

Table 1 Navigation performance comparison for random target points %

环境	方法	成功率↑	碰撞率↓	超时率↓
简单环境	multi	49.6	40.2	10.2
	MDRLAT	79.2	18.2	**2.6**
	PAAD	80.4	15.0	4.6
	本文	**88.2**	**9.0**	2.8
复杂环境	multi	30.2	63.4	6.4
	MDRLAT	47.6	52.2	2.0
	PAAD	51.6	43.6	4.8
	本文	**74.6**	**23.6**	**1.8**
动态行人环境	multi	15.2	78.2	6.6
	MDRLAT	24.6	70.0	**5.4**
	PAAD	25.4	66.8	7.8
	本文	**61.6**	**18.6**	19.8

在简单环境中,对随机生成的 100 个目标点进行跨目标测试。所提方法在成功率上取得了 88.2%,并且其碰撞率仅为 9.0%,表现最佳。其低碰撞率表明智能体能够有效避开障碍物,展示了良好的感知能力。具体来说,所提方法通过多层 ConvLSTM 和跨模态融合模块有效弥补了视场角的限制,增强了感知能力。相比之下,MDRLAT 方法的成功率为 72.9%,碰撞率为 15.0%,虽然其超时率为 10.2%,与所提方法相当,但整体导航性能较差,表明线性融合的人工调参方式还是有一定的局限性。Multi 融合方式的表现最差,其成功率为 49.6%,碰撞率为 40.2%,超时率为 10.2%,表明数据拼接这种融合方式未能有效提升感知能力。PAAD 方法的成功率为 80.4%,碰撞率为 13.8%,超时率为 2.6%,显示出较好的性能,但仍不及所提方法。

在复杂和动态行人环境中进行的跨场景泛化实验,验证了不同方法的适应性。

在复杂环境中,所提方法成功率为 74.6%,碰撞率为 23.6%,超时率为 1.8%,显示了较强的泛化能力,能够较好应对未知复杂环境。相比之下,MDRLAT 的成功率为 47.6%,碰撞率为 52.2%,表明其感知和避障能力在复杂环境中受限;PAAD 的成功率为 51.6%,碰撞率为 43.6%,虽略优于 MDRLAT,但碰撞率依然较高,表明其多头注意力机制在复杂场景中的适应性仍不足;multi 方法则表现最差,成功率为 30.2%,碰撞率高达 63.4%。

在动态行人环境中,机器人从原点出发,需避开以半径为 1 的单位长度绕原点运动的行人,且行人速度快于机器人,这增加了导航的复杂性。在这种环境下,所提方法的表现最为优越,成功率达到 61.6%,显著高于其他方法,且碰撞率为 18.6%,超时率为 19.8%。因为环境场地较大,超时率相对较高,但所提方法依然展现出较强的任务完成能力和优异的避障效果,能够有效应对动态障碍的挑战。PAAD 方法的成功率为 25.4%,碰撞率为 66.8%,表现较差,未能有效应对动态障碍。MDRLAT 方法的成功率为 24.6%,碰撞率为 70.0%,超时率为 5.4%,同样未能有效应对动态行人。Multi 方法的表现最差,成功率为 15.2%,碰撞率高达 78.2%,显示出在应对动态障碍时的能力严重不足。说明本文方法比 PAAD、MDRLAT 和 multi 更有可能在未知环境中完成导航任务。

由于上述随机生成的目标点无法对所有方法的轨迹长度和运行时间进行比较,故在简单、复杂场景和动态行人场景中分别选取一个固定目标点进行实验。每个实验重复 5 次独立实验,并每次进行 50 次测试。实验结果如表 2 所示,图 10 展示了不同方法在这些环境中的轨迹路径。

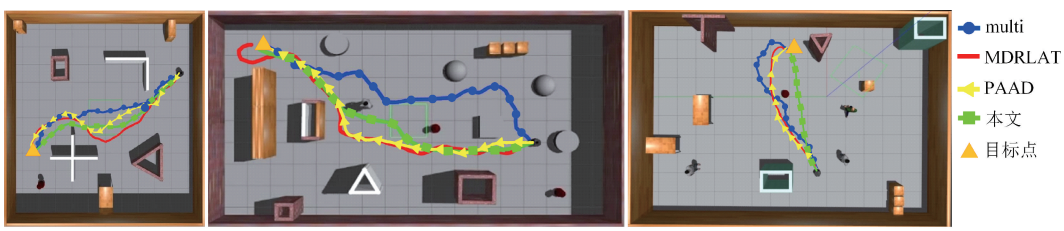
结合表 2 的评估结果和图 10 的轨迹路线,所提方法在简单环境、复杂环境和动态行人环境中表现出显著的优势。具体来说,在简单和复杂环境中,所提方法展示出较强的适应性和稳定性,能够较好地避开障碍物并持续朝目标前进。尤其在动态行人环境中,尽管环境复杂且有快速

移动的障碍物，所提方法仍能维持较高的成功率和较低的碰撞率。路径规划和决策速度相对较快，能够高效避开障碍，完成目标任务。PAAD 方法在简单环境中表现良好，但在复杂环境和动态行人环境中遇到了较大的挑战，尤其是在动态环境下，其性能较差、成功率低、碰撞率高，显示出其在应对动态障碍时的不足。MDRLAT 方法在所有环境中的表现都较差，尤其在复杂环境和动态行人环境中，感知和避障能力无法有效应对障碍物，导致导航成功率低、碰撞频繁，未能成功避免障碍物。Multi 方法在所有环境中的表现最差，尤其是在动态环境下，成功率和避障能力都远远落后于其他方法，表现出严重的局限性。由此可以看出，所提方法不仅适用于静态环境中的标准任务，而且在面对动态环境中的复杂挑战时，能够保持高效且稳定的性能，显示出其广泛的应用潜力和优越性。

表 2　固定目标点的导航性能对比

Table 2　Navigation performance comparison for fixed target points

环境	方法	成功率/% ↑	碰撞率/% ↓	超时率/% ↓	时间/s ↓	步长 ↓
简单环境	multi	22.0	71.2	6.8	51.18	255.77
	MDRLAT	53.0	43.8	3.2	44.78	233.40
	PAAD	69.2	25.4	5.4	41.56	207.38
	本文	**74.0**	**26.0**	**0**	**40.17**	**184.99**
复杂环境	multi	20.8	63.2	16.0	72.38	361.40
	MDRLAT	31.6	68.0	0.4	64.86	324.13
	PAAD	40.8	58.8	0.4	69.20	315.44
	本文	**64.4**	**35.6**	**0**	**42.37**	**211.49**
动态行人环境	multi	10.2	83.0	6.8	122.67	381.40
	MDRLAT	19.8	79.0	**1.2**	95.92	252.54
	PAAD	21.2	74.4	4.4	87.82	241.20
	本文	**56.6**	**32.6**	10.8	**70.45**	**218.15**



（a）简单环境　（b）复杂环境　（c）动态行人环境

（a）Simple environment　（b）Complex environment　（c）Dynamic pedestrian environment

图 10　不同方法和环境下的路径轨迹

Fig. 10　Path trajectory visualization under different methods and environments

4　真实场景实验

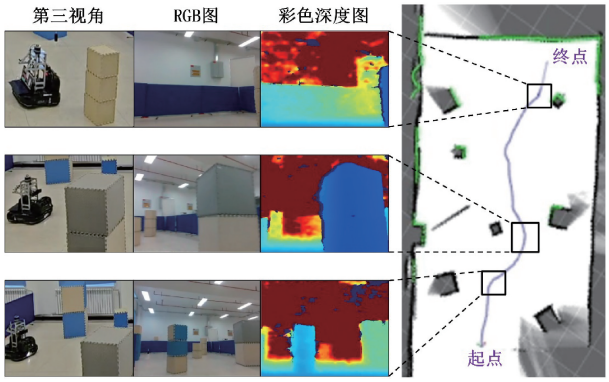
为验证本文所提目标驱动的移动机器人导航方法在真实环境下的有效性，使用松灵 Tracer 移动机器人执行导航命令，Realsense D435i 深度相机获取深度图，线速度和角速度直接由提出的模型从深度图中以端到端的方式推导出来，并将速度信息发送给机器人，实验场景如图 11 所示。

图 11　实验场景

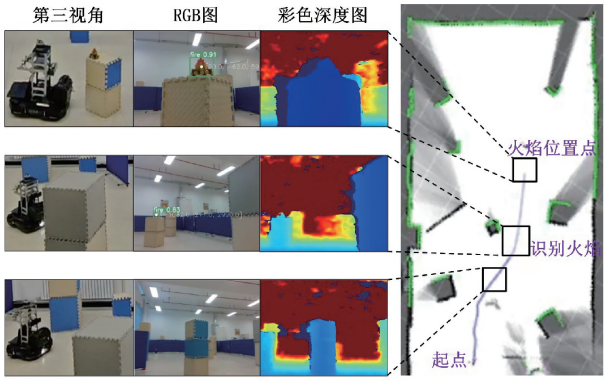
Fig. 11　Experimental scenarios

在搭建的实验场景中，使用本文方法进行了如图 12 所示的 3 组实验，每组实验在出发前都设定了图 11 所示的终点位置。导航过程如图 12 所示，对关键位置的第三视角、RGB 图像、彩色深度图和轨迹路线进行了展示。图 12(a)在导航过程中并未发现特定目标，故在避开障碍物后最终到达预设位置，实现位置点导航。而图 12(b)在导航过程中，机器人识别到火焰并对其进行定位，将火焰位置解算并代替预设位置，实现仿真火焰驱动的导航。在图 12(c)中，机

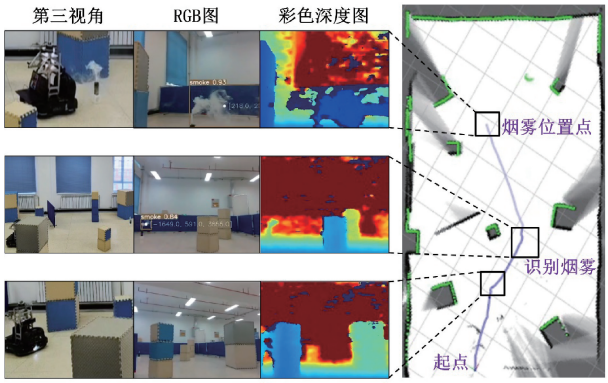
器人识别到烟雾并定位,解算出烟雾的位置并用此替代预设的目标位置,实现仿真烟雾驱动的导航。



(a) 位置点导航
(a) Position navigation



(b) 火焰驱动的导航
(b) Flame-driven navigation



(c) 烟雾驱动的导航
(c) Smoke-driven navigation

图 12 机器人的实时导航过程

Fig. 12 Real-time navigation process of the robot

将仿真环境中训练的导航模型迁移到真实环境中进行位置点导航,如图 12(a)所示。在此基础上,通过引入目标检测算法,机器人能够在导航过程中识别并定位特定目标(如火焰和烟雾),进而实现特定目标驱动的导航。此方法具有通用性,如需要对其他目标实现驱动导航,只需要对这类目标进行训练即可。具体来说,导航系统的核心部分保持不变,仅需更新目标识别模块以适应新的目标类型。这种灵活性使得系统能够快速适应不同的目标驱动任务,而无需重新训练整个深度强化学习模型。

5 结 论

本文提出了一种结合仿真训练与实际场景应用的方法:基于 D3QN 的目标驱动移动机器人自主导航。该方法集成深度图像与伪激光数据,利用单个 RGB-D 相机实现复杂环境下的导航任务。为增强对环境的感知能力,设计了一种跨模态融合模块以整合观测数据并充分捕捉环境信息。仿真实验结果表明,所提方法在导航成功率、碰撞率、超时率、路径长度和步数上都取得了一定的成效,提升了导航性能。在此基础上,利用目标检测方法对特定目标进行识别、定位并解算其目标位置,从而在真实场景中实现目标驱动的导航。未来工作将巡检与导航功能相结合,使移动机器人能够在特种环境中完成日常巡检和应急响应等任务。

参考文献

[1] 何丽,姚佳程,廖雨鑫,等. 深度强化学习求解移动机器人端到端导航问题的研究综述[J]. 计算机工程与应用,2024,60(14):1-13.
HE L,YAO J CH,LIAO Y X,et al. Research review on deep reinforcement learning for solving end-to-end navigation problems of mobile robots[J]. Computer Engineering and Applications,2024,60(14):1-13.

[2] 刘紫燕,杨模,袁浩,等. 结合拆分注意力机制和下一次预期观察的视觉导航[J]. 电子测量与仪器学报,2023,37(1):96-105.
LIU Z Y, YANG M, YUAN H, et al. Visual navigation combining split attention mechanism and next expected observation[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(1): 96-105.

[3] 邓修朋,崔建明,李敏,等. 深度强化学习在机器人路径规划中的应用[J]. 电子测量技术,2023,46(6):1-8.
DENG X P,CUI J M,LI M,et al. Application of deep reinforcement learning in robotpath planning[J]. Electronic Measurement Technology,2023,46(6):1-8.

[4] 王典,周阳,宋毅,等. 基于 Q 学习的生物启发式目标导向导航路径规划模型[J]. 电子测量与仪器学报,2023,37(6):68-76.
WANG D, ZHOU Y, SONG Y, et al. Model of path planning in biological inspired goal-oriented navigation based on Q-learning[J]. Journal of Electronic Measurement and Instrumentation,2023,37(6):68-76.

[5] DUAN Z M,CHEN Y W,YU H J,et al. RGB-fusion: Monocular 3D reconstruction with learned depth prediction[J]. Displays, 2021, 70: 102100.

[6] WU K Y, WANG H, ESFAHANI M A, et al. Learn to navigate autonomously through deep reinforcement learning [J]. IEEE Transactions on Industrial Electronics, 2021, 69(5): 5342-5352.

[7] 姜杨,曾铁文,万东东,等.基于 TS-TD3 的动态环境端到端无地图导航方法[J].机器人,2023,45(6):655-669.
JIANG Y, ZENG T W, WAN D D, et al. An end-to-end mapless navigation method based on TS-TD3 in dynamic environment [J]. Robot, 2023, 45(6): 655-669.

[8] 冷忠涛,张烈平,彭建盛,等.基于自适应探索 DDQN 的移动机器人路径规划[J].电子测量技术,2024,47(22):84-93.
LENG ZH T, ZHANG L P, PENG J SH, et al. Path planning for mobile robots based on self-adaptive exploration DDQN [J]. Electronic Measurement Technology, 2024, 47(22): 84-93.

[9] HUANG X Q, DENG H, ZHANG W, et al. Towards multi-modal perception-based navigation: A deep reinforcement learning method [J]. IEEE Robotics and Automation Letters, 2021, 6(3): 4986-4993.

[10] SONG H L, LI AO, WANG T, et al. Multimodal deep reinforcement learning with auxiliary task for obstacle avoidance of indoor mobile robot [J]. Sensors, 2021, 21(4): 1363.

[11] HAN Y H, ZHAN I H, ZHAO W, et al. Deep reinforcement learning for robot collision avoidance with self-state-attention and sensor fusion [J]. IEEE Robotics and Automation Letters, 2022, 7(3): 6886-6893.

[12] JI T C, SIVAKUMAR A N, CHOWDHARY G, et al. Proactive anomaly detection for robot navigation with multi-sensor fusion [J]. IEEE Robotics and Automation Letters, 2022, 7(2): 4975-4982.

[13] ZHU Y K, MOTTAGHI R, KOLVE E, et al. Target-driven visual navigation in indoor scenes using deep reinforcement learning [C]. 2017 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2017: 3357-3364.

[14] ZHELO O, ZHANG J W, TAI L, et al. Curiosity-driven exploration for mapless navigation with deep reinforcement learning [C]. Machine Learning in the Planning and Control of Robot Motion (MLPC), 2018.

[15] CIMURS R, SUH I H, LEE J H. Goal-driven autonomous exploration through deep rein-forcement learning [J]. IEEE Robotics and Automation Letters, 2022, 7(2): 730-737.

[16] HASSELT V H, GUZE A, SILVER D. Deep reinforcement learning with double Q-learning [C]. AAAI Conference on Artificial Intelligence, 2016.

[17] WANG Z, SCHAUL T, HESSEL M, et al. Dueling network architectures for deep reinforcement learning [C]. International Conference on Machine Learning, PMLR, 2016: 1995-2003.

作者简介

卢赵清,硕士研究生,主要研究方向为移动机器人自主导航。

E-mail: 2317027232@qq.com

王宏伟(通信作者),副教授,硕士生导师,主要研究方向为智能制造、计算机视觉、智能故障诊断与寿命预测。

E-mail: wanghongwei_xju@126.com

何丽,教授,硕士、博士生导师,主要研究方向为移动机器人共融导航、模式识别与智能控制技术、化工园区安全风险管控辅助决策技术。

E-mail: xju_heli@163.com