



## 基于多模态信息融合的四足机器人避障方法

吕友豪 贾袁骏 庄圆 董琦

### Obstacle avoidance approach for quadruped robot based on multi-modal information fusion

LYU Youhao, JIA Yuanjun, ZHUANG Yuan, DONG Qi

引用本文:

吕友豪, 贾袁骏, 庄圆, 董琦. 基于多模态信息融合的四足机器人避障方法[J]. 北科大: 工程科学学报, 2024, 46(8): 1426–1433. doi: 10.13374/j.issn2095–9389.2023.07.01.002

LYU Youhao, JIA Yuanjun, ZHUANG Yuan, DONG Qi. Obstacle avoidance approach for quadruped robot based on multi-modal information fusion[J]. *Chinese Journal of Engineering*, 2024, 46(8): 1426–1433. doi: 10.13374/j.issn2095–9389.2023.07.01.002

在线阅读 View online: <https://doi.org/10.13374/j.issn2095–9389.2023.07.01.002>

## 您可能感兴趣的其他文章

### Articles you may be interested in

#### 高速公路绿篱修剪机器人手臂避障路径规划

Obstacle avoidance path planning for expressway hedgerow pruning robot manipulator

工程科学学报. 2019, 41(1): 134 <https://doi.org/10.13374/j.issn2095–9389.2019.01.015>

#### 矿区废弃地移动机器人全覆盖路径规划

Complete coverage path planning of mobile robot on abandoned mine land

工程科学学报. 2020, 42(9): 1220 <https://doi.org/10.13374/j.issn2095–9389.2019.09.09.004>

#### 基于环境语义信息的同步定位与地图构建方法综述

Survey of simultaneous localization and mapping based on environmental semantic information

工程科学学报. 2021, 43(6): 754 <https://doi.org/10.13374/j.issn2095–9389.2020.11.09.006>

#### 巡线机器人延迟容忍传感器网络数据传输策略

Data delivery scheme of delay-tolerant mobile sensor networks for high-voltage power transmission line inspection robot

工程科学学报. 2018, 40(11): 1412 <https://doi.org/10.13374/j.issn2095–9389.2018.11.015>

#### 基于BP神经网络的机器人波动摩擦力矩修正方法

Wave friction correction method for a robot based on BP neural network

工程科学学报. 2019, 41(8): 1085 <https://doi.org/10.13374/j.issn2095–9389.2019.08.014>

#### 基于深度循环神经网络的协作机器人动力学误差补偿

Error compensation of collaborative robot dynamics based on deep recurrent neural network

工程科学学报. 2021, 43(7): 995 <https://doi.org/10.13374/j.issn2095–9389.2020.04.30.003>

# 基于多模态信息融合的四足机器人避障方法

吕友豪<sup>1,2)</sup>, 贾袁骏<sup>2)</sup>, 庄 圆<sup>2)</sup>, 董 琦<sup>2)</sup>✉

1) 中国科学技术大学先进技术研究院, 合肥 230026 2) 中国电子科学研究院, 北京 100049

✉通信作者, E-mail: [dongqiouc@126.com](mailto:dongqiouc@126.com)

**摘 要** 提出了一种全新的基于多模态信息融合技术的四足机器人避障方法. 该方法将机器人的本体传感器信息与外部传感器信息相结合, 以提高机器人在复杂环境下的决策能力. 具体而言, 该方法设计了一种多模态强化学习神经网络模型, 使机器人能够从自身的传感数据和外部传感器数据中学习. 该模型采用监督和非监督学习技术相结合的方法进行训练, 以优化机器人在避障任务中的表现. 此外, 还创新地引入了 Transformer 层和注意力机制, 使机器人能够有选择地关注相关的传感信息并过滤掉无关信息, 提高在未知动态环境中的规划可靠性. 该方法在具有不同障碍物、不平坦地形等具有挑战性的模拟环境中进行了评估. 实验结果表明, 所提出的方法相较于对照组可以显著提高四足机器人的避障成功率. 此外, 由于引入了注意力机制, 所提出的算法在动态未知环境下也具有一定的可靠性, 使其在实际应用中更加实用. 本文的意义主要在于引入多模态信息融合技术和 Transformer 层, 以提高机器人在避障任务中的表现. 通过仿真环境的实验结果显示, 该学习策略能够显著改善机器人的运动控制能力, 并且多模态 Transformer 模型进一步增强了其性能使其具备优越的泛化性. 此外, 进一步的分析和可视化也表明了学习策略利用外部输入进行决策的有效性.

**关键词** 四足机器人; 路径规划; 强化学习; 信息融合; Transformer

**分类号** TP242.6

## Obstacle avoidance approach for quadruped robot based on multi-modal information fusion

LYU Youhao<sup>1,2)</sup>, JIA Yuanjun<sup>2)</sup>, ZHUANG Yuan<sup>2)</sup>, DONG Qi<sup>2)</sup>✉

1) Institute of Advanced Technology, University of Science and Technology of China, Hefei 230026, China

2) China Academy of Electronics and Information Technology, Beijing 100049, China

✉Corresponding author, E-mail: [dongqiouc@126.com](mailto:dongqiouc@126.com)

**ABSTRACT** This paper proposes a multimodal information fusion neural network model that integrates visual, radar, and proprioceptive information. The model uses a spatial crossmodal attention mechanism to fuse the information, allowing the robot to focus on the most relevant information for obstacle avoidance. The attention mechanism enables the robot to selectively focus on the most relevant informative sensory inputs, which improves its ability to navigate complex terrain. The proposed method was evaluated using multiple experiments in challenging simulated environments, and the results showed a significant improvement in the obstacle avoidance success rate. The proposed method uses an actor-critic architecture and a proximal policy optimization (PPO) algorithm to train the robot in a simulated environment. The training process aims to reduce the difference between the robot's performance in simulated and real-world environments. To achieve this, we randomly adjust the simulation environment's parameters and add random noise to the robot's sensory inputs. This approach allows the robot to learn a robust planning strategy that can be deployed in real-world

收稿日期: 2023-07-01

基金项目: 网络空间安全态势感知与评估安徽省重点实验室开放课题(CSSAE-2021-003); 国家自然科学基金项目青年科学基金项目(61803353)

environments. The multimodal information fusion neural network model is designed using a transformer-based architecture. The model shares the encoding of three types of tokens and generates features for the robot's proprioceptive, visual, and point cloud inputs. The transformer encoder layers are stacked such that the token information from the three modalities can be fused at multiple levels. To balance the information from the three modalities, we first separately collect information for each modality and calculate the average value of all tokens from the same modality to obtain a single feature vector. This multimodal information fusion approach improves the robot's decision-making capabilities in complex environments. The novelty of the proposed method lies in the introduction of a spatial crossmodal attention mechanism that allows the robot to selectively attend to the most informative sensory inputs. This attention mechanism improves the robot's ability to navigate complex terrain and provides a certain degree of reliability for the quadruped robot in dynamic unknown environments. The combination of multimodal information fusion and attention mechanism enables the robot to adapt better to complex environments, thus improving its obstacle avoidance capabilities. Therefore, the proposed method provides a promising approach for improving the obstacle avoidance capabilities of quadruped robots in complex environments. The proposed method is based on the multimodal information fusion neural network model and spatial crossmodal attention mechanism. The experimental results demonstrate the effectiveness of the proposed method in improving the robot's obstacle avoidance success rate. Moreover, the potential applications of the proposed method include search and rescue missions, exploration, and surveillance in complex environments.

**KEY WORDS** quadruped robot; path planning; reinforcement learning; information fusion; transformer

四足机器人作为一类高度复杂的仿生机器人,具备先天的环境适应优势,具有广泛的应用前景和研究潜力<sup>[1-2]</sup>,提升四足机器人的运动规划能力扩大了其运动范围,进而使其能够胜任各种任务,包括野外环境探索<sup>[3-4]</sup>和工业巡检<sup>[5-6]</sup>等。所以,面向复杂地形条件下的四足机器人的运动规划研究对于其任务执行尤为关键。近年来,深度强化学习在导航和机器人操作领域取得了显著进展<sup>[7-14]</sup>,提升了四足机器人在应对现实世界复杂地形的能力。然而,当前基于强化学习的研究主要关注于无外部感知的四足机器人,忽视了视觉、雷达等外部感知的潜力。现有仿生智能研究表明<sup>[15]</sup>,视觉等外部感知输入在复杂地形运动中扮演重要角色。一些研究者<sup>[16-21]</sup>依赖于机器人周围和下方地形的高程图来规划脚步和关节角度,将深度图像融合到单个高程图中需要不同时间的相机之间的相对姿态,但这种方法带来了新的定位需求和计算负担。为此,本研究提出了一种新方法,即将本体感觉状态与视觉、雷达输入相结合,以学习四足机器人避障策略。

本文提出了基于 Transformer 模型的自主避障方法,用来融合本体感觉状态和外部传感器输入,以提高四足机器人自主避障的成功率。Transformer 模型通过将机器人状态估计、惯性测量单元读数和电机编码器等本体感知状态与来自视觉相机、激光雷达的输入相结合,实现了三种模态信息的互补和协同推理。此外,Transformer 模型引入了注

意力机制,使智能体能够有效融合多类信息,并专注于对长期和短期决策至关重要的特定区域。这种结合本体感觉状态和外部输入的方法使得避障策略更具泛化性和可解释性。

我们在模拟环境和真实环境中都进行了对比实验和消融实验,评估了算法对四足机器人运动的性能影响,实验结果表明,采用本体感觉状态和视觉、雷达输入的学习策略在模拟环境中都取得了显著的改进效果。通过仿真到现实环境的迁移实验,结果显示,该学习策略能够显著改善机器人的运动能力,并且多模态 Transformer 模型进一步增强了其性能使其具备优越的鲁棒性和泛化性。此外,进一步的分析和可视化也表明了学习策略利用外部输入进行决策的有效性。

本文工作主要意义在于:

(1)在机器人本体信息基础上,引入了视觉和雷达信息,利用端到端的强化学习训练让四足运动自主避障,穿越动态未知的复杂地形。

(2)提出了一种多模态信息融合 Transformer 架构,融合了本体感觉状态和视觉、雷达输入,以便在顺序决策中更好地进行多模态推理。

## 1 问题建模

### 1.1 马尔可夫决策过程

在本文中,四足机器人避障问题可以建模为强化学习中的马尔可夫决策问题(Part observable Markov decision process, MDP)<sup>[22]</sup>,马尔可夫过程是

指具有马尔科夫性质的随机过程, 即下一时刻状态仅由当前时刻状态决定, 而与过去状态序列无关. 马尔科夫决策过程可以由下式来描述:

$$p(s_{t+1}|s_t) = p(s_{t+1}|s_1, \dots, s_t) \quad (1)$$

其中,  $s_t$  为  $t$  时刻状态,  $p(s_{t+1}|s_t)$  为状态转移概率. 利用马尔科夫性不依赖前置状态关系这一特性, 可以将上述公式简化为:

$$a_t^* = \arg \max p(a_t|s_t) \quad (2)$$

其中,  $a_t$  表示  $t$  时刻下选择的动作,  $a_t^*$  表示在  $t$  时刻下选择的最优动作. 在环境的每次更新过程中, 下一个转变的状态仅被前一个转变所影响, 而不被更之前的转变所影响, 因此状态转变就可以用概率来表示:

$$p(s_{t+1}|s_t, a_t) \quad (3)$$

马尔科夫的决定过程包括三个方面的含义:

(1) 相邻状态是互相依存的. 因仅需要考虑前一状态, 从而极大地简化了问题.

(2) 智能体确定策略. 该系统能够通过自身的行为来实现对系统的动态响应, 该动态响应是系统的动态响应与系统的随机响应共同决定的.

(3) 时间依赖. 智能体与环境之间的相互作用是按时序进行的, 智能体实施动作, 并在此过程中得到反馈, 从而使环境生成新的状态.

## 1.2 强化学习

强化学习是智能体通过动作奖励学习个体行为与状态的映射, 并将其称为策略. 强化学习包含智能体 (Agent) 和环境 (Environment) 这两个实体组成部分. 强化学习的目标是学习一个可以使最终累计期望收益最大化的最优控制策略, 智能体通过在与仿真环境交互的过程中不断学习, 如图 1 所示.

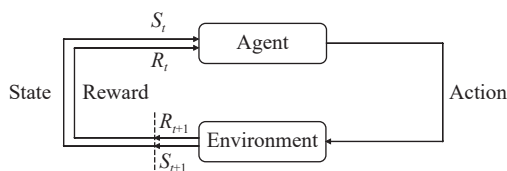


图 1 强化学习流程示意图

Fig.1 Illustration of reinforcement learning

近年来深度强化学习算法发展迅速, 诞生了一系列经典算法包括基于值函数和基于策略的算法. 我们选择在连续动作空间中处理复杂强化学习任务表现优异的近端策略优化算法 (Proximal policy optimization, PPO), 并且在训练中利用其分

布式版本以加速收敛过程.

## 2 基于多模态强化学习的四足机器人运动规划方法

### 2.1 基于 Transformer 的多模态信息融合神经网络

四足机器人具有复杂构型机械结构, 本体配备了丰富的传感器满足基本的稳定运动控制需要, 因此在进行模型设计时, 为了更好的通过多模态的异构传感器数据输入帮助智能体进行自主决策和规划等行为, 需要充分考虑对数据类型的选择. 在智能体所处环境中, 智能体利用本体感受状态和范围传感器观测进行决策. 本体感受状态和范围传感器观察是截然不同的模态: 本体感受输入是一个向量, 雷达、视觉观测信息需要额外进行编码处理, 本体信息用以估计智能体本体状态, 视觉图像提供语义信息和外界环境信息, 同时点云信息用以获取更稳定、准确的结构化信息. 为了进一步提取三种模态的有效信息并进行融合, 本文分别使用三个单独的领域特定编码器来处理本体感知、视觉、点云数据, 并在潜在空间中统一对齐、表示. 后面介绍每个编码器的架构设计, 以及如何将各模态信息转换为 Transformer 编码器的 tokens.

在本体信息上, 由于四足机器人有 12 个自由度 (DoF), 控制算法使用对电机的位置控制来实现机器人的动作切换. 所以具体而言, 本体感知输入包含关节角度 (12 维向量记录每个关节的角度)、惯性度量单元 (Inertial measurement unit, IMU) 信息 (4 维向量记录方向和角速度)、本体位置 (三维向量记录机器人的空间位置)、历史动作序列 (12 维向量记录最后一步的角度变化). 完整的本体感觉向量由最后三个步骤中的所有这些向量组成, 以保留历史状态信息进行更好地状态估计.

因此本文提出的神经网络架构如图 2 所示, 包括: 在神经网络输入端, 分别对本体状态信息、视觉信息及点云信息三种模态信息进行编码以获取特定领域的模态特征. 采用多层感知机对本体输入进行特征编码, 采用 ConvNet 编码器<sup>[23]</sup> 对第一视角图像进行特征编码, 除此之外, 采用 PointNet 编码器<sup>[24]</sup> 对点云进行特征编码, 这一步操作能使机器人更好的进行状态估计. 同时能够在隐藏空间中对三种模态特征统一表示, 将特征转化成用于 Transformer 编码的 tokens.

$$t^{\text{prop}} = W^{\text{prop}} (E^{\text{prop}}) + b^{\text{prop}}, t^{\text{prop}} \in \mathbf{R}^C \quad (4)$$



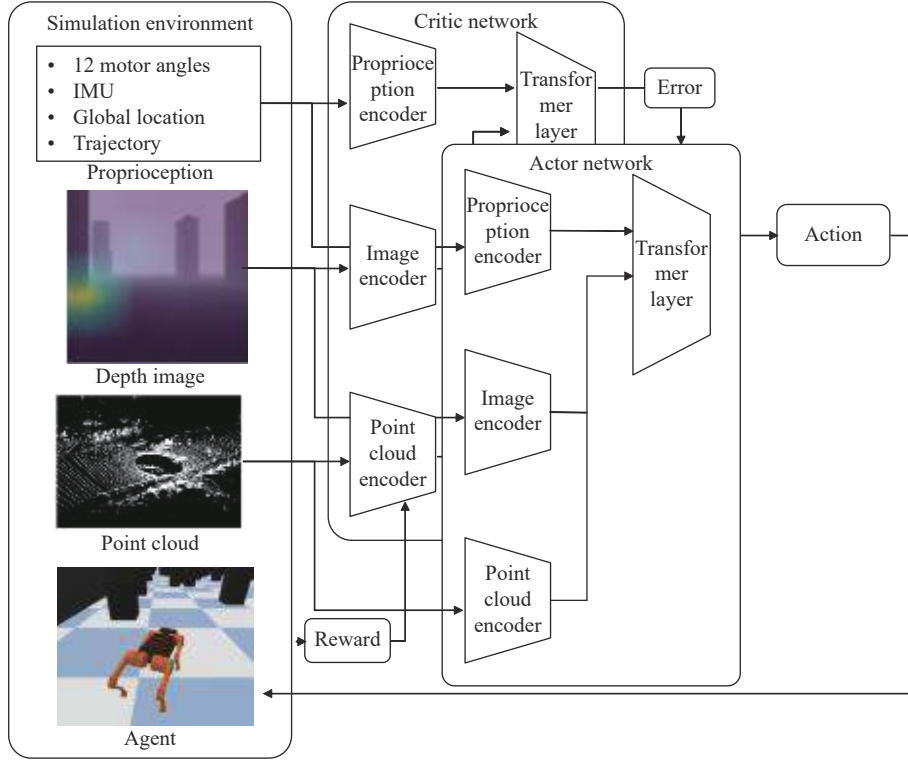


图2 本文使用的多模态信息融合神经网络模型结构及强化学习架构示意图

Fig.2 Structure of the multimodal information fusion neural network model and reinforcement learning architecture used in this paper

$$T_0 = \left[ t_{0,0}^{\text{prop}}, t_{0,0}^{\text{visual}}, t_{0,1}^{\text{visual}}, \dots, t_{N-1,N-1}^{\text{visual}}, t_{0,0}^{\text{lidar}}, t_{0,1}^{\text{lidar}}, \dots, t_{N-1,N-1}^{\text{lidar}} \right], t_{i,j}^{\text{visual}}, t_{i,j}^{\text{lidar}} \in \mathbf{R}^C \quad (5)$$

其中:  $t_i^{\text{prop}}$ ,  $t_{i,j}^{\text{visual}}$ ,  $t_{i,j}^{\text{lidar}}$  分别是本体、视觉、雷达特征;  $E^{\text{prop}}$  和  $W^{\text{prop}}$  为空间位置  $(i, j)$  处的参数矩阵;  $b^{\text{prop}}$  分别是本体感受 tokens 嵌入的线性投影的权重和偏差, 将  $T_0$  表示为输入标记序列。

然后, 采用 Transformer 模型共享编码三种 tokens 信息, 生成本体、视觉、点云特征. 将 Transformer 编码器层堆叠起来, 在多个层级上融合来自三种模态的 tokens 信息. 此外, 对于模态级融合, 在所有 tokens 上直接应用池化操作将很容易稀释本体感受信息, 因为感知信息 tokens 的数量远远超过本体感觉 tokens 的数量. 为了平衡来自三种模态的信息, 本文首先为每个模态分别汇集信息, 计算来自同一模态的所有 tokens 的平均值以获得单个特征向量。

$$T^q, T^k, T^v = TU^q, TU^k, TU^v \\ U^q, U^k, U^v \in \mathbf{R}^{C \times C} \quad (6)$$

$$W^{\text{sum}} = \text{Softmax}(T^q T^{kT} / \sqrt{D}) \\ W^{\text{sum}} \in \mathbf{R}^{(N^2+1) \times (N^2+1)} \quad (7)$$

$$SA(T) = W^{\text{sum}} T^v U^{SA} \quad U^{SA} \in \mathbf{R}^{C \times C} \quad (8)$$

其中,  $D$  是自注意层的维度,  $U$  为线性变换的参数,  $T$  为 token 序列, 自注意力机制首先在每个输入 tokens 上应用单独的线性变换以产生嵌入  $T^q, T^k, T^v$ . 然后, 本文计算输入 tokens 的加权和, 其中每个 tokens 对  $(t_i; t_j)$  的权重  $W_{i,j}^{\text{sum}}$  和计算为元素  $t_i$  和  $t_j$  的点积, 按  $1/\sqrt{D}$  缩放, 并通过 Softmax 操作进行归一化. 在权重  $W$  和值  $T^v$  之间进行矩阵乘法后, 本文将结果转发到具有参数  $U^{SA}$  的线性层, 将其表示为输出  $SA(T)$ 。

最后, 将三种模态的特征向量拼接后使用全连接网络 (Fully connected neural network, FCNN) 将串联的向量投影到最终的输出向量中. 网络结构设计如图 3 所示。

## 2.2 强化学习框架设计

在仿真环境中, 采用 Actor-Critic 架构和 PPO 算法对智能体进行避障策略训练, 如图 4 所示. 在训练过程中, 为了减小智能体在仿真环境和现实环境的差异, 实现智能体避障策略的在仿真环境的自主生成和在现实环境的低成本部署, 通过随机调整仿真环境的各项参数以及为智能体感知数据增加随机噪声, 以获得对环境鲁棒的规划策略。

模型采用了 Actor-Critic 架构<sup>[25-26]</sup>, 结合了价值评估 (Critic) 和策略改进 (Actor) 两个部分, 以实现

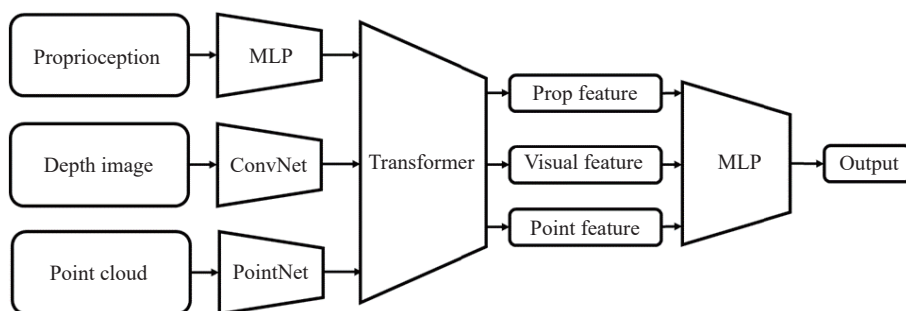


图 3 本文提出的多模态信息融合神经网络模型结构

Fig.3 Structure of the proposed multimodal information fusion neural network model

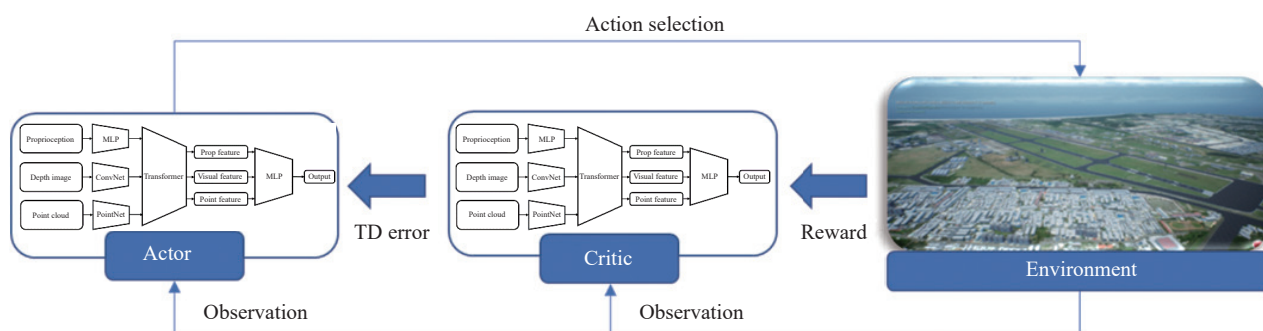


图 4 本文使用的强化学习架构

Fig.4 Actor-critic reinforcement learning architecture

对环境的学习和决策优化. Actor 网络和 Critic 网络均采用上一节中介绍的模型结构, 其中 Actor 为策略网络, 输出智能体动作, Critic 为价值网络, 输出价值估计. 通过这种设计, 使得策略优化更加高效, 增强策略的稳定性, 使得学习过程更加平滑. 为了缩小现实差距, 我们在训练阶段利用了领域随机化<sup>[26]</sup>. 在训练过程中, 对本体感觉状态在  $(-15) \sim (+15)$  内随机生成噪声向量与本体数据进行相加. 此外, 我们还对视觉输入进行域随机化. 具体来说, 在每个时间步长中, 我们在  $(64, 64)$  视觉输入中随机选择 3 到 30 个值, 并将深度读数设置为最大读数. 通过这种方法模拟了有噪声的视觉观察和本体传感器误差.

#### (1) 观测空间设计.

在所有环境中, 智能体接收本体感受状态和视觉输入如下: (i) 本体感受数据: 93 维向量包括 IMU 读数、局部关节旋转和智能体在最后三个时间步中所采取的行动; (ii) 视觉数据: 从安装在机器人头部的深度摄像机堆叠最近 4 张  $64 \times 64$  形状的密集深度图像, 为智能体提供空间和时间视觉信息; (iii) 点云数据: 包含了激光雷达连续四个时刻获取的激光数据, 每一时刻的激光数据包括 512 个距离测量值, 由于点云数据的数量庞大及盲区特性需要进一步进行处理以保证算法实时性.

#### (2) 动作空间设计.

强化学习输出指定绝对线速度和角(旋转)速度<sup>[27-29]</sup>. 线速度和角速度通过低层控制器实现对四足机器人的控制. 具体来说, 使用位置控制实现摆动动作, 扭矩控制实现站立动作. 我们使用基于有限状态机(FSM)的步态调度器来决定何时摆动每条腿以及每条腿在完整的控制周期中需要多长时间的姿势. 摆动动作由固定的脚部间隙高度决定, 并由具有目标脚部位置的 PD 控制器控制. 通过模型预测控制计算站立力(腿上每个关节的扭矩), 以跟踪所需的上层规划速度.

#### (3) 奖励函数设计.

在我们所有的实验中, 我们使用如下相同的奖励函数:

$$R = \alpha_{\text{forward}} R_{\text{forward}} + \alpha_{\text{energy}} R_{\text{energy}} + \alpha_{\text{alive}} R_{\text{alive}} + K \cdot R_{\text{sphere}} \quad (9)$$

其中,  $\alpha_{\text{forward}}$ ,  $\alpha_{\text{energy}}$ ,  $\alpha_{\text{alive}}$  分别为相应奖励的权重, 为所有任务设置  $\alpha_{\text{forward}} = 1$ ,  $\alpha_{\text{energy}} = 0.005$ ,  $\alpha_{\text{alive}} = 0.1$ .  $K$  为动态避障任务下奖励权重, 在静态避障任务下置 0,  $R_{\text{sphere}}$  为球体碰撞惩罚.  $R_{\text{forward}}$  代表前进的奖励, 在平坦环境中, 它由机器人沿  $x$  轴(默认初始方向)移动的速度定义.  $R_{\text{energy}}$  确保机器人使用最小的能量, 这已被证明可以改善机器人运动的性能, 提升机器人能耗比. 具体地说, 我们用欧

几里得范数来惩罚产生高运动力矩的动作:

$$R_{\text{energy}} = -\|\tau\|^2 \quad (10)$$

其中,  $\tau$ 为电机扭矩,  $R_{\text{alive}}$ 鼓励智能体活得更久. 它在每个时间步长上给予 1.0 的积极奖励, 直到终止. 诸如摔倒和撞上障碍物等危险行为将被判定为终止.

### 3 实验

#### 3.1 实验设置

在模拟环境中评估本文提出的方法, 仿真环境的搭建利用开源的 Pybullet 软件, Pybullet 可以提供包括四足机器人在内多种物理模型导入和自定义环境的摩擦系数、重力、碰撞系数等物理属性配置, 四足机器人的视觉、激光雷达、惯性测量单元等传感器通过插件形式加装在模型上, 同时搭建了一组具有挑战性和多样性的环境进行四足机器人实验.

本实验在安装 Ubuntu 操作系统的计算机上运行, 神经网络在 NVIDIA 的 3090 显卡上进行训练. 训练场景如图 5 所示, 其中, 黑色柱体为静态障碍物, 红色球体为随机产生的动态障碍物. 每个训练周期开始时环境都会随机生成障碍物的位置以及目标点的位置, 智能体在运行 1000 个时间步后仍然没有达到目标点, 或者智能体在某个时间步发生碰撞都会开始新的训练回合.

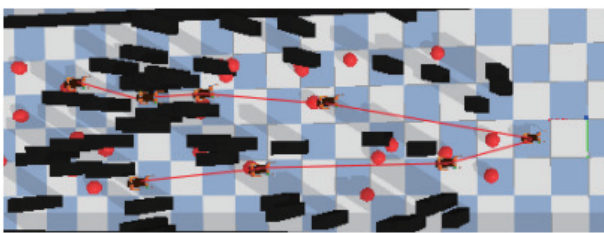


图 5 搭建的仿真训练环境

Fig.5 Built simulation training environment

#### 3.2 实验结果

为了证明外部感知信息在复杂环境中运动的重要性, 以及利用注意力机制进行信息融合的有效性, 本文将提出的方法与仅使用本体感觉状态的实验组、仅使用视觉观察的实验组、仅使用雷达电源的实验组以及使用本体感觉状态和视觉、雷达特征拼接(未用注意力机制)的实验组进行对比. 状态-图像实验组使用线性投影将视觉特征投影到与本体感觉特征具有相同维度的特征向量中. 然后, 状态-图像实验组将两个向量拼接起来, 并将其输入到价值和策略网络中. 使用 PPO 算法训

练所有实验组中的智能体, 并为价值和策略网络共享相同的本体感知和外部信息编码器.

本文通过平均回报率和两个特定领域的评估指标来评估策略: (i) 智能体沿着其目标方向移动的距离(以米为单位); 以及(ii) 每回合与障碍物的碰撞次数(长度为 1000 步). 在每个时间步骤检查碰撞, 并且只计算机器人经过至少一个障碍物时的碰撞. 本文用 5 种不同的随机种子训练 15M 样本的所有方法, 并报告最终策略的平均值和标准差.

根据图 6 训练曲线图, 可以发现以下情况: 单一模态信息相较于多模态信息更容易收敛, 且本体信息的震荡明显更小, 这可能是因为由 IMU 和电机数据组成的本体信息更有利于进行更准确的状态估计和预测, 但由于缺少感知信息, 这并不能为训练过程带来更高的奖励. 通过比较两种多模态方法可以看出注意力机制能够更快收敛且获得更高的奖励, 我们推测注意力机制可以更有效地对齐多模态特征, 隐式获取了机器人与外界环境的空间关系, 具有更好的空间感知和决策能力.

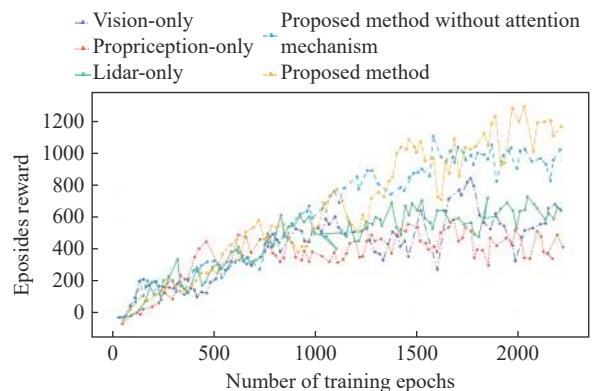


图 6 本文提出的方法与 baseline 的性能比较

Fig.6 Performance comparison of the proposed algorithm and baseline algorithm

根据结果推测, 当环境外观相对简单时, 仅图像实验组可以在不使用本体感觉状态的情况下学习合理的策略, 智能体可以从策略学习的视觉观察中推断出本体感受状态的一部分. 这一现象表明, 建模不同模态之间的相关性和更好的融合技术对于一个好的策略训练至关重要. 除此之外还观察到, 更简单的实验组在简单环境中的性能与本文的模型未有明显差距, 这是因为区分障碍物和平坦地形不是一项感知上复杂的任务, 因此, 简单的拼接就足以进行策略学习.

表 1 显示了定量评估结果的平均值. 我们发现未利用注意力机制的多模态信息基线足以达到



具有明显优势的效果, 但与基于注意力的方法在障碍环境中将回合奖励分别提高了 69% 和 56%。我们观察到, 当障碍物环境进行评估时, 我们的方法平均移动距离更远, 并且比 vision-only、lidar-only 和 proprioception-only 三类方法减少了 183.7%、153.1% 和 464.3% 的碰撞次数。proprioception-only 与模型预测控制 (Model predictive control, MPC) 方法由于无法获取外部感知信息, 在各种环境中频繁发生碰撞, 但凭借出色的状态估计能力, 甚至能在陷入障碍物后进行逃脱。而视觉传感器对与外部的感知给机器人带来了明显增益, 不过由于视觉传感器易受噪声影响以及缺失本体信息, 四足机器人非常容易因环境障碍密集和环境可观测性下降影响, 这也在与我们的推测相符合, 也是我们设计之初增加雷达传感器的原因之一。而对于我们提出的方法, 它要么趋于稳定, 要么随着时间的推移而增加。注意力机制的引入也有效减少了实验结果的方差, 这表明我们的方法在训练过程中更有效地捕捉外部感知和本体感知信息中的基本信息, 并且不容易过度拟合训练环境。

表 1 提出算法与基线算法在仿真环境比较结果

Table 1 Comparison of the proposed and baseline algorithms

Method	Moving distance/m	Collision probability/%
Model predictive control	5.2±1.2	45.5±3.7
Vision-only	8.0±3.2	18±5.1
Proprioception-only	4.8±0.8	44.2±10.2
Lidar-only	8.3±1.8	15±2.7
Proprioception-vision-lidar (without attention)	10.0±1.3	12.3±4.2
Proprioception-vision-lidar (with attention)	12.2±2.3	9.8±2.3

本文还评估了未知环境泛化的所有方法。表 1 的显示了每种方法的碰撞概率和移动平均距离, 本文的方法对比其他所有实验组上都有很大幅度的改善。由此有以下推断: 尽管可以获得图像信息, 但单一模态的实验组并没有显示出任何明显的改善; 通过结果推测将空间视觉特征直接编码到向量中, 并通过简单的拼接来融合多模态信息, 很容易丢失视觉信息的空间结构; 尽管本体信息基线在基线之间移动得更远, 但它没有获得更高的回合奖励, 这表明本体信息基线无法利用朝向目标的视觉引导; 本文基于 Transformer 的方法可以更好地捕捉空间信息, 如地形环境的全局和局部特征, 并且比简单的连接更成功地融合空间和

本体感觉信息。

### 3.3 消融实验

在仿真环境中进一步评估了提出模型的两个参数的重要性: (1) Transformer 编码器层数; 以及 (2) tokens 的数量 (评估结果见表 2 和表 3)。

表 2 融合网络层数消融实验

Table 2 Ablation result of transformer layers

Method	Average reward	Moving distance/m	Collision probability/%
Proposed (Layer=1)	1209.7±122.4	11.6±1.3	12.3±4.2
Proposed (Layer=2)	1151.5±142.5	9.8±1.8	9.8±3.3
Proposed (Layer=3)	1123.5±96.6	11.0±2.3	13.7±4.8
Proposed (Layer=4)	1257.5±105.4	10.2±2.9	10.6±5.6

表 3 token 数量消融实验

Table 3 Ablation result of token numbers

Method	Average reward	Moving distance/m	Collision probability/%
Proposed (N=1)	1015.6±113.5	6.3±1.3	8.7±3.7
Proposed (N=2)	1085.2±152.9	7.2±2.3	11.0±4.2
Proposed (N=3)	1251.4±132.2	11.8±2.5	9.4±3.5

从表 2 中, 本文观察到本文模型的性能对 Transformer 编码器层的数量相对不敏感。为了消除 tokens 的数量对实验结果的影响, 本文通过改变视觉编码器中最后一个卷积层的卷积核大小和步长, 以获得具有不同数量视觉标记的视觉特征。从表 3 中, 可以看到, 本文提出方法的性能与 tokens 的数量呈正相关。在特征图的固定大小下, 更高数量的标记直接导致每个 tokens 的接受域更小。因为本文的方法在所有 tokens 上执行空间跨模态注意力机制, 所以本文的模型受益于更丰富的低级特征信息。这表明本文的模型有可能在更复杂的环境和复杂的任务中更多的利用到外部感知的信息。

## 4 总结

本文提出了一种将本体感觉、视觉信息和雷达信息结合的避障策略神经网络模型。通过引入视觉、雷达等外部传感器输入, 并利用注意力机制进行信息融合, 智能体可以在运动过程中自主避开障碍物通过复杂地形。除了固定训练环境之外, 本文还通过实验表明, 在未知环境中进行测试时使用多模态信息融合方法的智能体获得了更好的泛化结果。这表明本文提出的模型在本体感受和



外部感知信息之间提供了一种有效的融合机制, 以及利用多模态信息进行强化学习的新可能性。在未来的工作中, 我们将尝试引入 Visual transformer 等表征学习领域的最新研究成果, 替代当前特定模态编码网络结构, 以期减少感知信息的损失, 进一步提高策略模型的决策水平。

## 参 考 文 献

- [1] Miki T, Lee J, Hwangbo J, et al. Learning robust perceptive locomotion for quadrupedal robots in the wild. *Sci Robot*, 2022, 7(62): eabk2822
- [2] Grandia R, Jenelten F, Yang S H, et al. Perceptive locomotion through nonlinear model-predictive control. *IEEE Trans Robotics*, 2023, 39(5): 3402
- [3] Zhou S Y, Xi J Q, McDaniel M W, et al. Self-supervised learning to visually detect terrain surfaces for autonomous robots operating in forested terrain. *J Field Robotics*, 2012, 29(2): 277
- [4] Velasquez A E B, Higuti V A H, Gasparino M V, et al. Multi-sensor fusion based robust row following for compact agricultural robots [J/OL]. *arXiv preprint* (2021-01-28) [2023-07-01]. <https://arxiv.org/abs/2106.15029>
- [5] Ji T C, Sivakumar A N, Chowdhary G, et al. Proactive anomaly detection for robot navigation with multi-sensor fusion. *IEEE Robotics Autom Lett*, 2022, 7(2): 4975
- [6] Ji T C, Vuppala S T, Chowdhary G, et al. Multi-modal anomaly detection for unstructured and uncertain environments [J/OL]. *arXiv: preprint* (2020-12-15) [2023-07-01]. <https://arxiv.org/abs/2012.08637>
- [7] Weerakoon K, Sathyamoorthy A J, Patel U, et al. TERP: Reliable planning in uneven outdoor environments using deep reinforcement learning // 2022 *International Conference on Robotics and Automation (ICRA)*. Philadelphia, 2022: 9447
- [8] Kahn G, Abbeel P, Levine S. BADGR: An autonomous self-supervised learning-based navigation system. *IEEE Robotics Autom Lett*, 2021, 6(2): 1312
- [9] Kahn G, Abbeel P, Levine S. LaND: Learning to navigate from disengagements. *IEEE Robotics Autom Lett*, 2021, 6(2): 1872
- [10] Bledt G, Kim S. Extracting legged locomotion heuristics with regularized predictive control // 2020 *IEEE International Conference on Robotics and Automation (ICRA)*. Paris, 2020: 406
- [11] Carius J, Ranftl R, Koltun V, et al. Trajectory optimization for legged robots with slipping motions. *IEEE Robotics Autom Lett*, 2019, 4(3): 3013
- [12] Li T Y, Calandra R, Pathak D, et al. Planning in learned latent action spaces for generalizable legged locomotion. *IEEE Robotics Autom Lett*, 2021, 6(2): 2682
- [13] Lu G L, Chen T, Rong X W, et al. Whole-body motion planning and control of a quadruped robot for challenging terrain. *J Field Robotics*, 2023, 40(6): 1657
- [14] Jiang X Y, Chi W C, Zheng Y, et al. Locomotion generation for quadruped robots on challenging terrains via quadratic programming. *Auton Robot*, 2023, 47(1): 51
- [15] Xu S H, Zhu L J, Ho C P. Learning efficient and robust multimodal quadruped locomotion: A hierarchical approach // 2022 *International Conference on Robotics and Automation (ICRA)*. Philadelphia, 2022: 4649
- [16] Aladem M, Baek S, Rawashdeh S A. Evaluation of image enhancement techniques for vision-based navigation under low illumination. *J Robotics*, 2019, 2019: 1
- [17] Song H L, Li A, Wang T, et al. Multimodal deep reinforcement learning with auxiliary task for obstacle avoidance of indoor mobile robot. *Sensors*, 2021, 21(4): 1363
- [18] Guan T R, Kothandaraman D, Chandra R, et al. GA-nav: Efficient terrain segmentation for robot navigation in unstructured outdoor environments. *IEEE Robotics Autom Lett*, 2022, 7(3): 8138
- [19] Mansouri S S, Kanellakis C, Kominiak D, et al. Deploying MAVs for autonomous navigation in dark underground mine environments. *Robotics Auton Syst*, 2020, 126: 103472
- [20] Cai P D, Wang S K, Sun Y X, et al. Probabilistic end-to-end vehicle navigation in complex dynamic environments with multimodal sensor fusion. *IEEE Robotics Autom Lett*, 2020, 5(3): 4218
- [21] Qu Y H, Yang M H, Zhang J Q, et al. An outline of multi-sensor fusion methods for mobile agents indoor navigation. *Sensors*, 2021, 21(5): 1605
- [22] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition // 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, 2016: 770
- [23] Qi C R, Su H, Mo K C, et al. Pointnet: deep learning on point sets for 3D classification and segmentation // 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, 2017: 652
- [24] Huang X Q, Deng H, Zhang W, et al. Towards multi-modal perception-based navigation: A deep reinforcement learning method. *IEEE Robotics Autom Lett*, 2021, 6(3): 4986
- [25] Park D, Kim H, Kemp C C. Multimodal anomaly detection for assistive robots. *Auton Robot*, 2019, 43: 611
- [26] Wellhausen L, Ranftl R, Hutter M. Safe robot navigation via multimodal anomaly detection. *IEEE Robotics Autom Lett*, 2020, 5(2): 1326
- [27] Sorokin M, Tan J, Liu C K, et al. Learning to navigate sidewalks in outdoor environments. *IEEE Robotics Autom Lett*, 2022, 7(2): 3906
- [28] Rusli L, Nurhalim B, Rusyadi R. Vision-based vanishing point detection of autonomous navigation of mobile robot for outdoor applications. *J Mechatron Electr Power Veh Technol*, 2021, 12(2): 117
- [29] Li Z, Li B, Liang Q X, et al. A quadruped robot obstacle avoidance and personnel following strategy based on ultra-wideband and three-dimensional laser radar. *Int J Adv Robotic Syst*, 2022, 19(4): 17298806221114705