

融合 Transformer 与 SAC 算法的 无人车视觉导航模型

张益敏, 刘磊[†]

(河海大学数学学院, 江苏 南京 211100)

摘 要:为应对真实环境中自主导航任务面临的高感知力和安全性需求,提出了一种融合 Transformer 与 SAC 算法的无人车视觉导航模型。该模型利用 Vision Transformer(ViT)网络提取图像特征,作为强化学习算法的状态输入,以此来提高无人车在动态环境中的感知能力。为应对视觉传感器故障导致的安全性问题,预训练雷达策略作为决策过程中的低计算成本应急备选方案。在 Gazebo 中进行仿真实验,结果表明,融合 Vision Transformer 训练的视觉导航策略收敛速度更快且成功率更高;在传感器突发故障时,各模式策略能够协调配合,保证导航的鲁棒性;所训练的模型能在不同环境中展现出良好的适应性。

关键词:深度强化学习; Vision Transformer; 自主导航; Gazebo 仿真

中图分类号:TP18

文献标识码:A

Visual Navigation Model for Unmanned Ground Vehicles Integrating Transformer and SAC

ZHANG Yimin, LIU Lei[†]

(School of Mathematics, Hohai University, Nanjing, Jiangsu 211100, China)

Abstract: To address the high perception and safety requirements of autonomous navigation tasks in real-world environments, this paper proposes an unmanned ground vehicle (UGV) visual navigation model that integrates Transformer and SAC (Soft Actor-Critic) algorithms. This model utilizes a Vision Transformer (ViT) network to extract image features, which serve as state inputs for the reinforcement learning algorithm, thereby enhancing the UGV's perception capabilities. To mitigate safety issues arising from visual sensor failures, a pre-trained radar policy is employed as a low computational cost emergency fallback during the decision-making process. Simulation experiments conducted in Gazebo demonstrate that the visual navigation policy trained with Vision Transformer not only converges faster and achieves higher success rates but also ensures navigation stability through effective coordination of multi-modal strategies during sensor failures. Furthermore, the trained model exhibits excellent adaptability and robustness across different environments.

Key words: deep reinforcement learning; Vision Transformer; autonomous navigation; Gazebo simulation

随着人工智能的不断发展,自主导航技术成为交通领域的重要创新方向之一。自主导航有望提升出行效率,助力构建更安全、更智能的交通系统^[1]。在自主导航任务中,常采用 A* 算法^[2]、

RRT^[3]、蚁群算法^[4]、人工势场法^[5]等方法进行路径规划和避障,但这些经典的方法通常依赖于预定义规则和已知地图,缺乏对环境动态变化的适应能力,在面对复杂和不确定的环境时效果有限^[6]。强

化学习作为一种模拟智能体自主学习并优化决策的技术,逐渐成为研究自主导航的重要方法。

强化学习通过与环境交互学习策略,不依赖先验地图,具备更强的自适应性和探索能力^[7]。其中 SAC 算法因其较高的探索效率和对环境变化的适应性,在自主导航领域得到了广泛应用。例如,文献[8]采用 SAC 算法,通过结合自注意力机制和门控循环单元,实现了自主导航决策与规划的协同,大幅提升了学习速度和车速稳定性。

现有的用于自主导航的强化学习算法多采用雷达数据作为环境状态输入。因为雷达在距离测量方面具备较高的准确性,且计算复杂度较低。例如,文献[9]提出了一种基于激光雷达数据的 SAC 算法,结合人工势场法增强泛化能力,实现了自主导航中的高效避障。然而,在复杂的动态环境下,激光雷达扫描无法提供足够的信息来描述环境^[10],因此学者们开始关注基于视觉传感器的方法,通过视觉图像获取更丰富的环境信息。文献[11]提出了一种基于强化学习的视觉室内导航方法,通过卷积神经网络提取图像特征,扩展可观测性,实现了 UGV 在未知场景下的导航。但现有文献中多采用离散化的动作空间,限制了 UGV 的灵活性,且卷积神经网络在处理全局信息和长程依赖方面存在局限性。

近年来,Transformer 模型在自然语言处理和计算机视觉领域中广泛应用,尤其是 Vision Transformer(ViT)模型,它在处理图像输入时表现出色,通过自注意力机制捕捉全局依赖关系,能够直接学习到图像中的全局上下文信息^[12-14]。文献[15]采用 ViT 编码器和 Causal Transformer 解码器,通过大规模端到端强化学习训练,仅使用 RGB 信息就完成了具身机器人的室内导航任务。

基于以上启发,本文在有动态障碍物的室内环境中,采用 Vision Transformer 网络提取图像特征作为 SAC 算法的状态输入,并结合优化的奖励函数,训练了基于视觉的导航模型。考虑到 RGB 摄像头易受光线影响从而导致视觉信息缺失,本文进一步训练了不受光照条件影响的基于深度图像的策略。又预训练雷达策略作为视觉策略的对照策略,并在决策过程中作为低计算成本应急备选方案,来应对视觉传感器突发故障导致的安全性问题。通过在 Gazebo 仿真环境中的测试,验证了该方法在自主导航任务中的优越性,以及应对突发故障时的有效性。

1 任务描述和理论基础

1.1 任务描述

本文研究融合 Transformer 和 SAC 算法的无人车视觉导航模型,来应对真实环境中自主导航任务面临的高感知力和安全性需求。要求 UGV 能够根据视觉传感器获取的环境信息做出导航决策,从随机初始位置顺利导航至固定目标点,并在导航过程中避免碰撞。同时,在视觉传感器发生故障的情况下,UGV 能切换至备选雷达模态方案,保证导航的稳定性和安全性。

1.2 强化学习

自主导航任务可以看成由五元组 $\langle S, A, P, R, \gamma \rangle$ 构成的马尔可夫决策过程,其中 S 是状态空间,表示机器人及其周围环境的可能状态; A 是动作空间,表示可能的控制命令,如速度和角速度; $P(s_{t+1} | s_t, a_t)$ 是状态转移概率,表示执行动作 a_t 后从状态 s_t 进入下一个状态 s_{t+1} 的概率; $R(s_t, a_t)$ 是奖励函数,表示在状态 s_t 执行动作 a_t 后获得的奖励; γ 是折扣因子,用于权衡当前奖励和未来奖励。

强化学习的目标是通过策略 $\pi(a_t | s_t)$ 选择动作,以最大化累积的期望奖励,即值函数:

$$V_{\pi}(s) = E_{s_t \sim P} \left[\sum_{t=0}^T \gamma^t R(s_t, a_t) \right] \quad (1)$$

同时,策略的状态行为值函数定义为:

$$Q_{\pi}(s_t, a_t) = r_t + \gamma \cdot E_{s_{t+1} \sim P} [V_{\pi}(s_{t+1})] = r_t + \gamma \cdot E_{s_{t+1} \sim P, a_{t+1} \sim \pi} [Q_{\pi}(s_{t+1}, a_{t+1})] \quad (2)$$

在本文中,采用柔性动作——评价算法(SAC)进行策略优化,该算法通过最大化熵增强智能体对环境的探索能力,其目标函数为:

$$\max_{\pi} E_{s_t \sim P, a_t \sim \pi} \left[\sum_{t=0}^T \gamma^t (r_t + \alpha H(\pi(\cdot | s_t))) \right] \quad (3)$$

其中, α 是熵的奖励系数, $H(\pi(\cdot | s_t))$ 表示策略的熵。

1.3 目标引导的 Vision Transformer

Vision Transformer (ViT) 是一种将 Transformer 应用于图像领域的模型,其主要思想是将图像拆分为若干小块,并将它们映射为线性嵌入,与标准 Transformer 架构在自然语言处理(NLP)中处理“tokens”方式类似。

基于目标引导的 ViT 模型则是将物理目标状态作为场景感知网络的输入,通过定义一个特殊的

目标 token,达到将物理目标与视觉信息结合的目的。首先,通过多层感知机将目标状态 s_{goal} 映射为向量 $G \in R^{1 \times D}$ 。具体公式为:

$$G = \text{MLP}(s_{\text{goal}}) \quad (4)$$

从而嵌入向量可以表示为:

$$z_0 = [G; \text{LP}(x_1); \text{LP}(x_2); \dots; \text{LP}(x_n)] + E_{\text{pos}} \quad (5)$$

其中 $x_1, x_2, \dots, x_n$ 为图像重塑成的一系列展平的图像块, $E_{\text{pos}} \in R^{(N+1) \times D}$ 是位置嵌入。

随后,将嵌入向量输入 Transformer 编码器中:

$$\begin{aligned} z'_l &= \text{MSA}(\text{LN}(z_{l-1}) + z_{l-1}) \\ z_l &= \text{FC}(\text{LN}(\text{MLP}(z'_l) + z'_l)) \end{aligned} \quad (6)$$

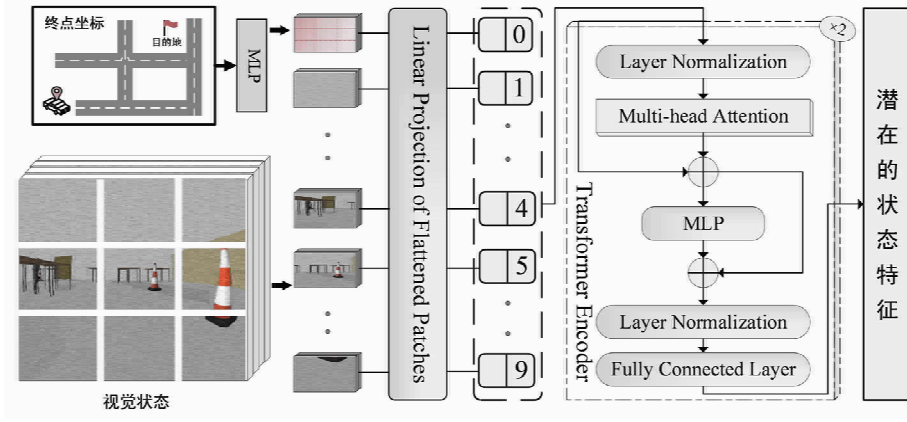


图 1 目标引导的 ViT 整体架构图

2 算法模型设计

2.1 状态空间与动作空间设计

本实验中为 UGV 配备了摄像头、深度传感器和雷达传感器,获取 RGB 图像 I_{RGB} 、深度图像 I_{Depth} 以及广角激光雷达数据 D_{Radar} ,对应传感器下的状态信息、极坐标下的目标位置信息 s_{goal} 以及 UGV 当前时刻的位置坐标 s_e 、速度 v_t 和方向角 ω_t 均包含在状态输入内。

$$S = [\alpha_{\text{RGB}} I_{\text{RGB}}, \alpha_{\text{Depth}} I_{\text{Depth}}, \alpha_{\text{Radar}} D_{\text{Radar}}, s_{\text{goal}}, s_e, v_t, \omega_t] \quad (8)$$

其中, $\alpha_{\text{RGB}}, \alpha_{\text{Depth}}, \alpha_{\text{Radar}} \in \{0, 1\}$ 为选择因子,如果某一模态被选中,则其对应的因子为 1,否则为 0。仿真过程中,若选择 RGB 图像模态,需辅以雷达模态信息用于碰撞检测。

本文为 UGV 设计了二维连续的动作空间,具体动作大小范围如表 1 所示。避障导航的 UGV 在二维空间中的运动由速度和转向联合控制。

其中, l 表示第 l 个块。

因此,潜在的特征可以从第 l 个块的输出中获得,记为:

$$h = \text{GoT}(s, G; \varphi) \quad (7)$$

其中, φ 表示 GoT 的参数。

图 1 展示了目标引导的 ViT 模型架构。如图所示,输入由两个模态组成:作为物理状态的目标信息和作为视觉状态的原始 RGB 图像。物理状态被送入 MLP 网络编码为特征块,视觉状态被分解为 9 个小图像块。Transformer 编码器由 MSA 块、MLP、FC、层归一化操作和残差连接组成。考虑到有限的计算能力和轻量级设计,本文采用两层编码器,注意力头数为 4。

UGV 必须在一个具有动态障碍物的复杂环境中避开障碍物并到达目标点,连续的动作空间可以为 UGV 提供更多的避障选择,提高导航任务的安全性和稳定性。因此,得到动作空间:

$$A = \{(v_t, \omega_t) \mid -1 \leq v_t \leq 1, -2 \leq \omega_t \leq 2\} \quad (9)$$

表 1 智能体车辆动作空间

名称	范围	单位	描述
转向	$[-2.0, 2.0]$	rad/s	负值为左转,正值为右转
速度	$[-1.0, 1.0]$	m/s	负值为后退,正值为前进

2.2 奖励函数的设置

自主导航的目标是展示一种以目标为导向的决策和无碰撞路径规划,以到达目标位置。因此,本文根据基于势能的回报塑形技术设计了奖励函数,结合连续奖励和稀疏奖励,以提高算法的收敛效率。具体来说,总奖励由以下四个独立组成部分构成:

$$r_t = r_{\text{shape}} + r_{\text{smooth}} + r_{\text{safe}} + r_{\text{target}} \quad (10)$$

基于势能的塑形回报 r_{shape} 能够激励 UGV 朝向目标位置移动,当 UGV 向目标位置靠近时就会收到正向奖励,远离目标位置时则会受到负向惩罚。因此,基于势能的塑形回报比仅针对靠近终点行为的单方面奖励具有更强的引导作用:

$$r_{\text{shape}} = \mu (\|p_{t-1} - g\|_2 - \|p_t - g\|_2) \quad (11)$$

其中, μ 为常数权重, p_t 表示 UGV 的实时位置, g 表示目标点的位置。

为了使 UGV 尽可能少地进行转向操作,平滑、快速地接近目标位置,设计了平滑奖励:

$$r_{\text{smooth}} = v_t - 2 \times |w_t| \quad (12)$$

最后,对无碰撞以及导航成功分别设计对应的回报函数:

$$r_{\text{safe}} = \begin{cases} -100, & \text{if collision} \\ k \cdot (d_{\text{obs}} - d_{\text{threshold}}), & \text{if } d_{\text{obs}} > d_{\text{threshold}} \end{cases}$$

$$r_{\text{target}} = \begin{cases} 100, & \text{if } d_t \leq 0.5 \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

其中, k 是调节系数, d_{obs} 是 UGV 与最近障碍物的距离, $d_{\text{threshold}}$ 是安全阈值, $d_t = \|p_t - g\|_2$ 。 r_{safe} 会对与障碍物保持一定安全距离的行为给予奖励,从而更好地避免碰撞发生。

2.3 基于目标引导的 ViT-SAC 算法

在决策模块中,采用基于目标引导的 ViT-SAC 算法,让 SAC 算法根据 c_t 进行动作选择,并通过与环境交互不断更新策略 $\pi(a_t | c_t)$,其中 c_t 是由目标引导的 ViT 网络输出的潜在特征 h_t 结合状态空间中的其他辅助信息得到的。在 SAC 算法

中使用双 Q 网络来解决过估计问题。通过最小化平均贝尔曼平方误差 (MBSE) 损失函数来更新 Critic 网络的参数:

$$J_Q(\theta_i) = E_{h_t \sim P, a_t \sim \pi} \left[\frac{1}{2} (r_t + \gamma \cdot \hat{Q} - Q_{\theta_i}(c_t, a_t))^2 \right] \quad (14)$$

其中, $\hat{Q}$ 是由双目标 Q 网络计算得到的下一步的状态-动作值:

$$\hat{Q} = E_{h_{t+1} \sim P, a_{t+1} \sim \pi} [\min_{i=1,2} Q_{\theta_i}(c_{t+1}, a_{t+1}) - \alpha \log \pi_{\varphi}(a_{t+1} | c_{t+1})] \quad (15)$$

相应地, Actor 网络通过最大化软状态行为函数来更新其参数:

$$J_{\pi}(\varphi) = E_{h_t \sim P, a_t \sim \pi} [\min_{i=1,2} Q_{\theta_i}(c_t, a_t) - \alpha \log \pi_{\varphi}(a_t | c_t)] \quad (16)$$

融合 Transformer 与 SAC 算法的视觉导航模型如图 2 所示,模型分别从前方摄像头、深度传感器获得 RGB 图像、深度图像作为输入信息,传感器之间独立工作,互不影响。由于 RGB 图像不包含障碍物的距离信息,因此在仿真中,RGB-R 策略模型需要结合雷达数据进行碰撞检测。RGB-R 策略模型通过目标引导的 ViT 模型提取图像中关键特征,可以有效解析复杂场景,增强自主导航系统在多变环境中的视觉理解能力。Depth 策略模型可以利用物体的空间位置和形状,为自动导航系统提供重要的避障与路径规划信息,且不受室内光线强弱的影响。

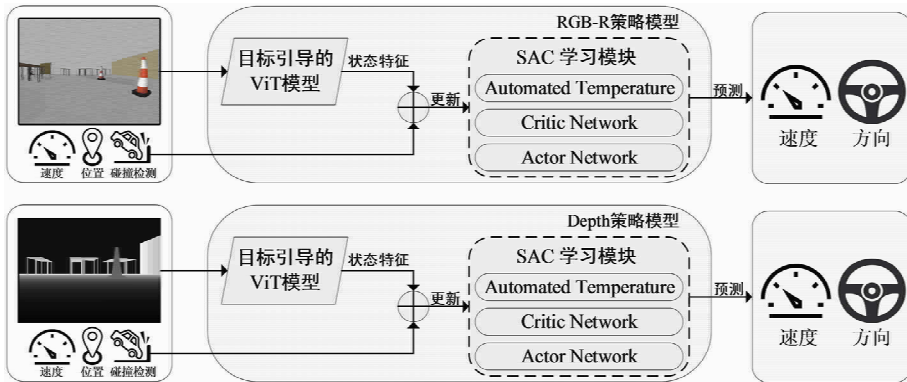


图 2 视觉模态训练

2.4 模态策略应急备选机制

在现实生活中,自主导航系统存在潜在风险,因为一旦传感器失效,UGV 将无法感知到外界信

息。在这种情况下,车辆可能做出危险的导航决策,导致碰撞或任务中断的发生。为了克服这些问题,本文提出了一种模态策略应急备选机制,该方

法能够以简单的框架保障导航的安全性和连续性。

在基于 ViT-SAC 的视觉导航模型之外,训练了一个 Radar 策略模型,如图 3 所示。与图像数据相比,雷达数据的维度和复杂性较低,使用全连接

层处理雷达数据相比于图像处理计算成本较低,可以实现更快的处理速度,且 Radar 策略模型对室内光照变化也能保持高度鲁棒性,可以作为视觉传感器突发故障时的备选策略。

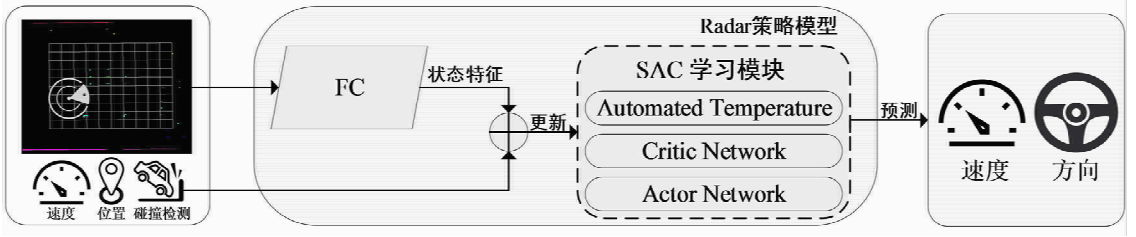


图 3 雷达模态训练

模态策略应急备选机制如图 4 所示。具体的切换条件如下:(1)正常环境:当环境稳定且光照条件良好时,系统使用 RGB-R 策略模型进行动作选择。(2)复杂光照:在复杂光照条件下(如强光、阴影或无光环境中)追求更稳定的表现,使用 Depth 策略模型进行决策。(3)视觉传感器故障:系统当前使用的视觉导航策略对应的传感器发生故障时,自动切换至另一视觉传感器。当两个视觉传感器都无法接收到有效信息时,系统将自动切换至备用的 Radar 策略模型。融合 Transformer 与 SAC 算法的无人车视觉导航模型伪代码如算法 1 所示。

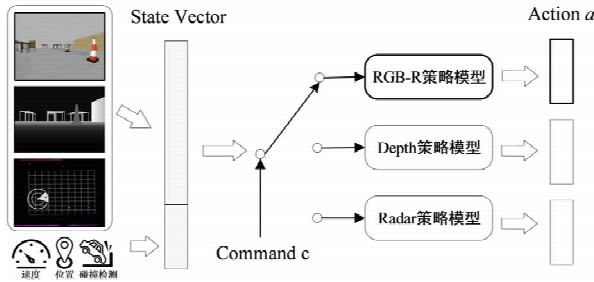


图 4 模态策略应急备选机制

算法 1:融合 Transformer 与 SAC 的无人车视觉导航算法

```

初始化基于目标引导的 ViT 的网络参数  $\varphi_{rgb}, \varphi_{depth}$ 
初始化 Actor 和 Critic 网络参数:  $\varphi_{rgb}, \varphi_{depth}, \varphi_{radar}, \theta_{rgb}, \theta_{depth}, \theta_{radar}$ 
初始化熵参数  $\alpha_{rgb}, \alpha_{depth}, \alpha_{radar}$ 
初始化批量大小  $N$ 
和经验回放池
 $D_{rgb} \leftarrow \emptyset, D_{depth} \leftarrow \emptyset, D_{radar} \leftarrow \emptyset$ 
初始化目标参数
 $\theta_{rgb}^{target} \leftarrow \theta_{rgb}, \theta_{depth}^{target} \leftarrow \theta_{depth}, \theta_{radar}^{target} \leftarrow \theta_{radar}$ 
# 训练过程
for Modal in ['rgb', 'depth', 'radar'] do:
    for episode = 1 to E do:
        初始化目标  $s_{\{goal, t\}} \sim Env$ 
        if Modal = 'rgb' or 'depth' else 初始化  $s_t \sim Env$ 
        for step = 1 to S do:
             $G_t = MLP(s_{\{goal, t\}})$ 
            if Modal != 'radar' else None

```

```

 $h_t \leftarrow GoT(G_t; \varphi_{Modal})$ 
if Modal != 'radar' else None
 $c_t \leftarrow (h_t, s_e, v_t, w_t)$ 
if Modal != 'radar' else  $s_t$ 
    依据当前策略选择动作  $a_t \leftarrow \pi_{\varphi_{Modal}}(a_t | c_t)$ 
    与环境交互  $r_t, s_{t+1} \sim Env$ 
    将  $D_{Modal} \leftarrow D_{Modal} \cup (s_t, G_t, a_t, r_t, s_{t+1})$ 
    存入经验回放池 if Modal != 'radar'
else  $D_{Modal} \leftarrow D_{Modal} \cup (s_t, a_t, r_t, s_{t+1})$ 
    基于公式 (14)(16) 更新 Critic 和 Actor 网络参数
 $\theta_{Modal}, \varphi_{Modal}$ 
软更新  $\theta_{Modal}^{target} \leftarrow \tau \cdot \theta_{Modal} + (1 - \tau) \cdot \theta_{Modal}^{target}$ 
 $s_t \leftarrow s_{t+1}$ 
end for
end for
# 根据条件选择适用的策略

```

3 实验与结果

3.1 仿真环境

本实验在 Gazebo 环境进行仿真,搭建了逼真的室内环境,设有静态障碍物、动态障碍物以及拥有动力学和多传感器仿真的 UGV 模型。RViz 仿真环境如图 5 所示, Gazebo 仿真环境如图 6 所示。在环境中由桌子、交通锥充当静态障碍物,移动的行人充当动态障碍物,蓝色部分为 UGV 上配备的雷达传感器的覆盖范围, RViz 中显示了实时的 RGB 图像和深度图像。在此场景下对算法进行训练,每回合最多训练 300 步。当 UGV 到达目标位置、发生碰撞或最大时间步耗尽时,一个训练回合结束。为了获得更真实、可迁移性更高的 DRL 的策略,本实验在每一个训练回合为 UGV 设置了随机的初始位置。实验中使用机器人操作系统 (ROS) 开源平台与 Gazebo 进行通信,通过订阅里程计消息获取目标信息。

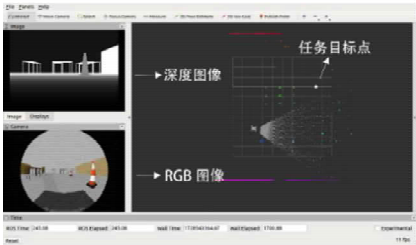


图 5 RViz 仿真环境

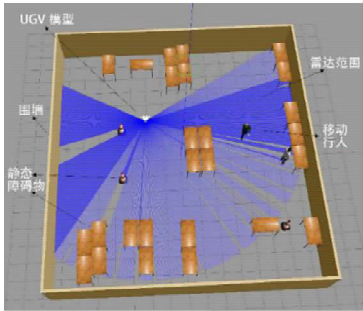


图 6 Gazebo 仿真环境

本实验在虚拟机 VMware Workstation 17Pro 内的 Ubuntu18.04 系统中进行,处理器为 13th Gen Intel(R) Core(TM) i7-13700H 2.40 GHz。无人车仿真使用 python 编程,adam 优化器进行网络训练优化。算法的网络训练参数设置如表 2 所示。

表 2 超参数设置

超参数	取值
批大小	32
每回合最大步数	300
记忆池容量	20000
折扣率	0.99
学习率	0.001
编码器层	2
注意力头	4

3.2 实验结果分析

实验中使用 120×160 的原始 RGB 图像 I_{RGB} 、 128×64 的深度图像 I_{Depth} 。图 7、图 8 分别展示了 RGB-R 策略模型和 Depth 策略模型中基于目标引导的 ViT 网络 and 传统卷积神经网络的学习曲线比较。红色实线表示基于 Transformer 网络的方法,而黑色实线表示传统卷积网络的方法。从图 7 中可以看到,基于 Transformer 网络的方法的平均奖励逐渐提高并保持稳定,而传统方法的曲线波动较大,平均奖励较低。这表明,RGB 图像提供了丰富的视觉信息,Transformer 网络更好地利

用了这些信息,相比之下,传统卷积网络则未能充分挖掘这些潜力。

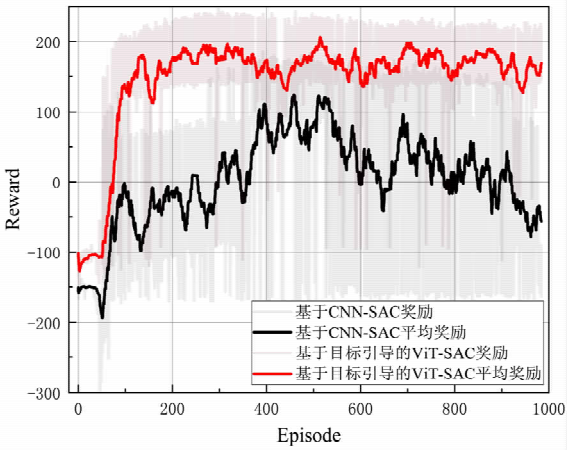


图 7 RGB-R 策略模型收敛曲线对比

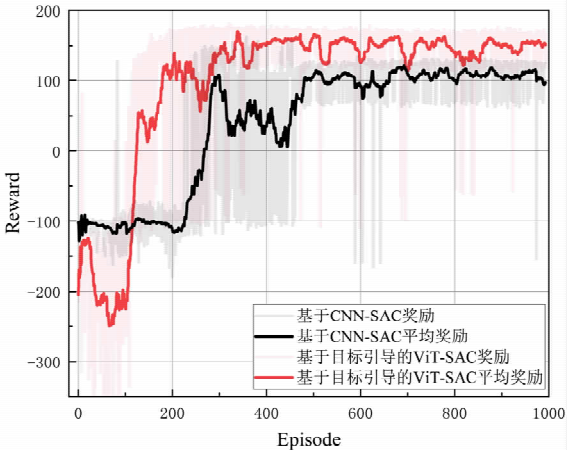


图 8 Depth 策略模型收敛曲线对比

图 8 中基于目标引导的 ViT 网络表现出更高的平均奖励和更快的收敛速度。这表明,Transformer 网络在学习和利用深度信息方面表现良好,尤其适用于处理复杂的空间和距离数据。总体来说,无论是 RGB 还是深度图像的策略模型,基于 Transformer 网络的方法都比传统卷积网络方法获得更高的奖励和更快的收敛速度,说明目标引导的场景表示和 Transformer 网络在此任务中具有显著优势。

实验中使用的广角激光雷达数据 D_{Radar} 覆盖 UGV 正前方 240° 扫描范围。从图 9 中可以看出, Radar 策略模型在初始阶段表现不稳定,奖励值波动较大,成功率低,但随着训练的进行,1500 回合后其性能逐渐稳定,成功率也有了较快的提升,表明 SAC 算法已有效掌握利用雷达数据进行有效导航的策略,显示出良好的学习和适应能力。使用全

连接层处理雷达数据相比于图像处理计算成本较低,可以实现更快的处理速度,适用于紧急情况下的实时避障导航。但由于任务的复杂度和环境的不确定性,Radar 策略模型需要更多的训练回合来学习如何从这些相对较简单的数据中提取足够的信息,解析环境中的障碍物位置、距离及其可能的动态变化,以做出有效的导航决策。

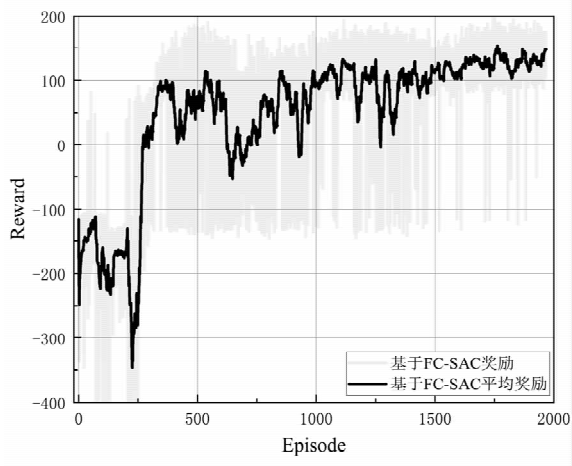


图 9 Radar 策略模型收敛曲线

模型测试中,将 100 回合平均奖励视为“驾驶分数”来评价导航性能。从表 3 中结果可以观察到,RGB-R 和 Depth 两大视觉模态策略导航在驾驶分数和导航成功率上的表现都优于雷达导航。这并不奇怪,因为 Radar 策略模型只利用了距离信息,其能力仅限于在不与周围物体发生碰撞的情况下向目的地靠拢。从表 3 可以看出 RGB-R 策略模型效果最好,这说明视觉输入对驾驶过程中的决策有更显著的影响。

表 3 模型测试及结果对比

模型	回合数	导航成功率	驾驶分数
RGB-R 策略模型	100	100%	182.95
Depth 策略模型	100	100%	173.57
Radar 策略模型	100	92%	121.84

3.3 鲁棒性测试

为测试模型的鲁棒性,本实验为视觉传感器设置了随机的故障率,本文将传感器故障定义为 UGV 完全接收不到感知信息,图像传入数据为 None,且在传感器故障仿真中,摄像头和深度传感器可能同时发生故障。

当某一传感器发生故障时,其对应的单模态策

略将无法做出动作决策。因此,使用模态策略备选机制来展示本文提出的方法应对传感器突发故障的能力。由于本文的仿真环境为简化的室内空间,未涉及光线变化,因此在传感器故障发生前,模型按照预训练中驾驶分数最高的 RGB-R 策略进行动作选择。图 10 展示了模型在不同故障率下的驾驶分数变化趋势。结果表明,在低故障率下,驾驶分数保持相对稳定,模态策略备选机制表现出较强的适应性。然而,随着传感器故障率提升,驾驶分数有所下降,这是因为不同模态的策略模型基于不同的控制逻辑和对环境的理解,频繁切换会导致机器人接收到突变的控制信号,导致运动轨迹不平滑,影响驾驶性能。但在传感器故障率小于 30% 的情况下,模型依然保持了较高的驾驶性能,自主导航的整体效果良好。

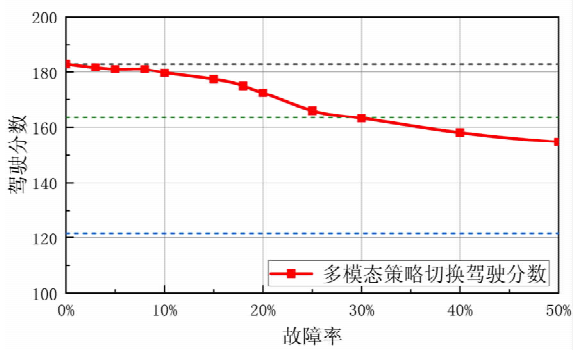


图 10 对传感器故障的鲁棒性

3.4 泛化性测试

为了进一步评估模型在不同环境下的迁移能力,将传感器故障率设置为 30%,在训练环境、测试环境 a 和测试环境 b(图 11)中进行了对比测试。相比于训练环境,测试环境中障碍物的类型、形状和布局更加复杂,包含了更多不规则和分散的结构,可用于评估模型在复杂多样的导航场景下的鲁棒性和迁移能力。测试结果表明,模型在训练环境和新环境中均有极高的成功率和驾驶分数。详细测试结果如表 4 所示。

表 4 30%故障率下模型测试

测试	环境	回合数	成功率	驾驶分数
1	训练环境	100	100%	163.25
2	测试环境 a	100	100%	155.39
3	测试环境 b	100	98%	121.16

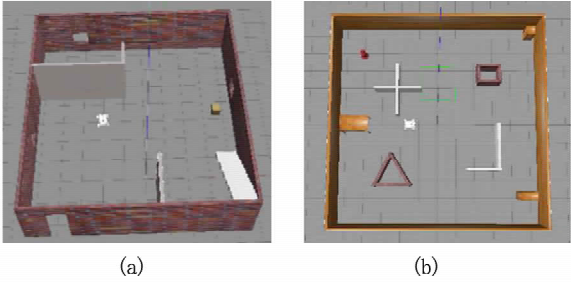


图 11 测试环境

4 结 论

本文提出了一种基于视觉的深度强化学习自主导航方法。通过使用 Vision Transformer 网络提取图像特征提高了无人车在复杂动态环境下的感知能力,将视觉特征作为深度强化学习算法的状态输入增强了模型训练效率。分别训练了适用于正常光照环境的 RGB-R 策略模型和不受光照影响的 Depth 策略模型,扩大了视觉导航的适用场景。并以模态策略应急备选机制应对导航过程中突发传感器故障导致的安全性问题,将雷达策略模型作为视觉策略模型的备选提升了导航系统在紧急情况下的鲁棒性。仿真实验表明,所提方法在具有动态障碍物的室内环境中表现出优异的导航性能,在面对传感器故障的情况下,凭借备选机制的保护能够维持较高驾驶分数和导航成功率。未来的研究将着眼于将视觉导航系统应用于更真实的驾驶场景中,以应对更加复杂的环境和多变的驾驶条件。

参考文献

[1] ALEDHARI M, RAHOUTI M, QADIR J, et al. Motion comfort optimization for autonomous vehicles: concepts, methods, and techniques[J]. IEEE Internet of Things Journal, 2024, 11(1): 378—402.

[2] LIN Z, WU K, SHEN R, et al. An efficient and accurate a-star algorithm for autonomous vehicle path planning[J]. IEEE Transactions on Vehicular Technology, 2024, 73(6): 9003—9008.

[3] GANESAN S, RAMALINGAM B, MOHAN R E. A hybrid sampling-based RRT* path planning algorithm for autonomous mobile robot navigation[J]. Expert Systems with Applications, 2024, 258: 125206.

[4] SU S, JU X, XU C, et al. Collaborative motion planning based on the improved ant colony algorithm for multiple autonomous vehicles [J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(3): 2792—2802.

[5] YANG H J, HE Y Q, XU Y, et al. Collision avoidance for autonomous vehicles based on MPC with adaptive APF[J]. IEEE Transactions on Intelligent Vehicles, 2024, 9(1): 1559—1570.

[6] CHALVATZARAS A, PRATIKAKIS I, AMANATIADIS A. A survey on map-based localization techniques for autonomous vehicles[J]. IEEE Transactions on Intelligent Vehicles, 2023, 8(2): 1574—1596.

[7] CHIB P S, SINGH P. Recent advancements in end-to-end autonomous driving using deep learning: a survey[J]. IEEE Transactions on Intelligent Vehicles, 2024, 9(1): 103—118.

[8] 唐斌,刘光耀,江浩斌,等. 基于柔性演员-评论家算法的决策规划协同研究[J]. 交通运输系统工程与信息, 2024, 24(2): 105—113.

[9] WEN S H, SHU Y L, RAD A, et al. A deep residual reinforcement learning algorithm based on Soft Actor-Critic for autonomous navigation[J]. Expert Systems with Applications, 2024, 259: 125238.

[10] GAJJAR H, SANYAL S, SHAH M. A comprehensive study on lane detecting autonomous car using computer vision[J]. Expert Systems with Applications, 2023, 233: 120929.

[11] ZHANG W R, LIU H, WU J B, et al. MVSSC: meta-reinforcement learning based visual indoor navigation using multi-view semantic spatial context[J]. Pattern Recognition Letters, 2023, 177: 75—81.

[12] VASWANI A, SHAZEER N M, PARMAR N, et al. Attention is all you need[J]. Neural Information Processing Systems, 2017, 30: 1—11.

[13] LI G F, QIU Y F, YANG Y F, et al. Lane change strategies for autonomous vehicles: a deep reinforcement learning approach based on transformer[J]. IEEE Transactions on Intelligent Vehicles, 2023, 8(3): 2197—2211.

[14] ZHANG Z M, LEI Z Y, OMURA M, et al. Dendritic learning-incorporated vision transformer for image recognition[J]. IEEE/CAA Journal of Automatica Sinica, 2024, 11(2): 539—541.

[15] ZENG K H, ZHANG Z C, EHASNI K, et al. PoliFormer: scaling on-policy RL with transformers results in masterful navigators [EB/OL]. [2024-06-28]. <https://arxiv.org/abs/2406.20083>.