基于近端策略优化算法的自动化集装箱码头自动导引车防冲突路径规划

肖世昌[1] 林 轩[1] 郑 鹏[1*] 王金凤[2]
1.上海海事大学物流工程学院,上海,201306
2.上海海事大学中国(上海)自贸区供应链研究院,上海,201306

摘要:为了提高自动化集装箱码头的运营效率和智能决策能力,针对水平运输区域中双向行驶的多自动导引车(AGV)防冲突路径规划问题展开研究。首先针对码头水平运输区域特征构建网格地图,并将AGV防冲突路径规划问题建模为以最小化所有任务最大完工时间为目标的数学规划模型。随后设计了近端策略优化(PPO)算法,搭建了面向双向引导车道的工作环境模型,设计了多AGV作业的动作空间和状态空间,通过引入防绕远策略提高算法搜索能力。最后将所提算法与商业求解器Gurobi、A*算法及遗传算法进行性能对比,仿真实验结果表明,所提算法针对大规模问题具有更强的性能稳定性和收敛能力。

关键词:自动化集装箱码头;自动导引车路径规划;近端策略优化算法;防绕远策略
中图分类号:TP242
DOI:10.3969/j.issn.1004-132X.2026.08.021 **开放科学(资源服务)标识码(OSID)**:

Anti-conflict Path Planning for AGVs in the Automated Container Terminals Based on Proximal Policy Optimization Algorithm

XIAO Shichang[1] LIN Xuan[1] ZHENG Peng[1*] WANG Jinfeng[2]
1.Logistics Engineering College, Shanghai Maritime University, Shanghai, 201306
2.China Institute of FTZ Supply Chain, Shanghai Maritime University, Shanghai, 201306

Abstract: To enhance the operational efficiency and intelligent decision making capability of automated container terminals, this study addresses the collision free path planning problem for multiple AGVs operating bidirectionally in the horizontal transport area. Considering the layout characteristics of the terminal's horizontal transport zone, a grid-based map is constructed, and the AGV collision-free path planning problem is formulated as a mathematical programming model with the objective of minimizing the makespan (i.e., the maximum completion time) of all tasks. A PPO algorithm is then designed. A simulation environment tailored for bidirectional guideway systems is developed, incorporating specifically designed action and state spaces for multi-AGV operations. An anti-detour heuristic is introduced to improve the search efficiency of the algorithm. The proposed algorithm is benchmarked against the commercial solver Gurobi, the A* algorithm, and a genetic algorithm. Simulation results demonstrate that the proposed algorithm exhibits superior solution stability and stronger convergence capability, particularly for large-scale problem instances.

Key words: automated container terminal(ACT); automated guided vehicle(AGV) path planning; proximal policy optimization(PPO) algorithm; detour prevention strategy

0 引言

近年来,自动化集装箱码头(automated container terminals, ACT)已在全球范围内得到广泛应用。作为ACT水平运输区域中的核心搬运设备,自动导引车(automated guided vehicle, AGV)的作业效率对码头整体运营性能有直接影响。特别是随着船舶装载容量的不断扩大和港口智能化发展的不断推进,AGV的作业效率不仅受到码头环境的影响,还受到设备调度和路径规划等因素的制约。在实际作业过程中,设备等待、路径冲突及死锁等问题严重影响了AGV的运行效率和系统的整体协调性,因此,多AGV防冲突路径规划问题已成为制约码头运作效率提高及实现ACT从自动化向智能化转型的关键瓶颈。

典型的ACT布局见附录一(见篇首处二维码),码头由若干堆场、岸桥及AGV双向行驶车道构成。以出口集装箱的运输为例,AGV通过场

收稿日期:2025-08-07

桥装载集装箱，随后沿指定路径将集装箱运输至对应岸桥，岸桥卸载集装箱并将其装载到船上；进口集装箱则沿相反路线进行运输。因此，合理规划AGV的行驶路径并有效解决路径冲突问题，不仅有助于降低运输成本，还能显著提高自动化集装箱码头的作业效率。

AGV之间发生碰撞或死锁等问题会降低水平运输区域效率，现有研究针对以上AGV路径规划的痛点问题已取得一定研究成果，主要集中在单向路径网络场景下的建模和求解[1-4]。虽然单向路径避免了AGV在同一路径上的冲突，但AGV的灵活性低。KIM等[5]通过仿真证明与单向路径场景对比，双向路径场景下AGV具有更短的行驶距离。现有研究表明，双向路径结构使AGV运输更加灵活，增加了AGV的可选路径，通过优化行驶路径以减少其行驶距离，可以更好地保证高效的水平运输操作[6-7]。

双向路径的灵活性虽然提高了AGV运输的效率，但也增加了冲突和死锁的复杂性。自动化集装箱码头AGV路径规划作为AGV高效完成集装箱搬运的核心问题，仍然面临着如何高效求解该类复杂模型的难题。目前用于解决路径规划问题的算法主要分为三大类：经典优化算法、智能优化算法和深度强化学习算法。经典优化算法方面，SHEN等[8]为解决动态环境下的路径规划问题，结合禁忌搜索算法与粒子群优化算法，提高了解决路径规划的优化能力。梁承姬等[9]针对集装箱码头基于卸箱任务的AGV路径规划问题，设计了一种在原路径或备选路径上插入时间窗的调整方法，实验证明该方法能有效实现AGV的避碰，提高了AGV利用率及码头运作效率。HU等[10]建立了多目标优化模型以解决自动化码头AGV的调度和路径规划问题，并结合预规划算法和实时规划算法的优点，提出了一种三阶段分解算法，将A*算法与时间窗原理相结合，实验结果展现了该方法在路径搜索和时间优化方面的有效性。XU等[11]利用AGV的缓冲区最多承载两个集装箱，设计了AGV进出路径规划模型，并采用模拟退火算法进行求解，实现了AGV的双向加载。姜辰凯等[12]针对多AGV在柔性制造系统中出现的路径规划与冲突问题，提出了一种基于时间窗的改进Dijkstra算法，在最优路径下有效避免冲突与死锁，实现了多AGV的动态路径规划。

对于高维度、复杂目标函数的AGV路径规划问题，经典优化算法的求解能力有限，智能优化算法应运而生。TANG等[13]针对港口环境下AGV节点多、距离长、转弯角度大的问题，设计几何A*算法改善传统A*算法产生的不规则路径，通过筛选函数、插值算法对路径进行平滑处理，通过实验证明该方法可有效地减少AGV行驶中的转弯次数，但未考虑AGV的大小。丁一等[14]针对AGV任务分配和路径规划问题，结合自动化码头布局特点，建立两阶段多目标优化模型，设计改进模拟退火算法和基于时空网络的路径规划算法分别进行求解，结果验证了所提方法的有效性。ZHONG等[15]建立了一个基于路径规划、集合调度、冲突和死锁的MILP模型，并验证了具有模糊逻辑控制器的混合遗传算法-粒子群优化的可用性。WU等[16]针对集装箱码头水平运输中的AGV路径优化问题，基于A*算法融合动态窗口方法，提出一种AGV路径规划和实时避障模型，发现采用改进算法能够提高集装箱装卸效率，并使规划路径更平滑。CHEN等[17]针对AGV无冲突路径规划问题，基于时间窗模型，提出了一种将传统四个方向上寻找路径的A*算法改进为任意方向搜索最短路径，但程序总运行时间过长。LIU等[18]将一种改进的遗传算法应用于U型集装箱码头AGV路径规划问题，根据问题的多目标和层次特征来安排任务和解决冲突。CAO等[19]针对集装箱码头AGV装卸过程中的最优路径问题，以AGV执行任务时行驶距离最短为目标，在传统粒子群算法中引入遗传算法交叉变异的思想，并调整惯性权重，改进了粒子群算法，通过算例证明了改进粒子群算法可以有效地缩短行驶路程。LU等[20]为实现码头AGV无冲突路径规划，以最短AGV完成总任务时间为目标，建立了min-max数学规划模型，设计遗传算法广度优先搜索得到AGV最短路径，将其作为候选路径，采取FCFS策略和时间窗进行避障，得到较优路径，但算法求解效率较低。岳春擂等[21]针对求解路径规划类问题的常用算法——蚁群算法进行改进，引入了动态启发因子的概念，对每一个栅格的邻近栅格进行方向标号，一定程度上规避了蚁群算法容易陷入局部最优解的情况，提高了其算法鲁棒性。荀燕琴[22]先采用蚁群算法对AGV进行预路径规划，再通过检验交叉栅格判断是否存在交叉冲突现象，对存在冲突现象的AGV进行蚁群局部二次路径规划。

尽管智能优化算法在求解AGV路径规划问题中取得了一定成效，但仍存在一定局限性，尤其在大规模动态场景下面临收敛速度慢和容易陷入局部最优的问题，求解效率有待提高。深度强化学习算法凭借其强大的环境感知及决策能力，能够适应动态场景的变化要求，近年来受到广泛关注。CRUZ等[23]为了克服传统强化学习耗时长、学

习速度慢、无法在未知环境中进行学习的缺点，使用神经网络提高强化学习拟合能力，并与核平滑技术相结合，能够完成未知环境场景下的完成路径规划问题。POPPER等[24]为解决柔性作业车间调度与AGV协调的NP-hard问题，提出一种基于多智能体强化学习的方法，实现作业调度和车辆路径规划协同工作。CETINKAYA等[25]针对动态环境中的路径规划问题，提出了一种基于深度强化学习的多智能体路径规划算法，满足了实时性要求，在仿真环境下展现了较高的效率。CHEN等[26]针对自动化集装箱码头AGV路径规划问题，提出了一种基于人工势场和孪生延迟深度确定性策略梯度框架的港口环境路径优化模型，将强化学习算法的数据输入TD3算法中，结果表明该方法在路径安全性和路径平滑性方面具有显著优势。WANG等[27]为满足高实时性要求，提出融合注意力机制的强化学习方法，用于生成多机协同路径规划。基于深度强化学习算法的路径规划方法主要采用从海量的数据中学习的模式，以实现模型的训练，可以取得很好的效果。由上可知，尽管深度强化学习算法能在一定程度上解决AGV路径规划问题，但仍存在模型训练困难、环境适应性差等问题，且当前研究主要面向生产制造系统、仓库以及一些其他场景，针对自动化集装箱码头场景下多AGV路径规划问题的相关研究报道较少。

近年来自动化集装箱码头领域的AGV路径规划技术得到了较大的发展，但仍然存在亟待解决的痛点问题：一是混合装卸箱任务场景下的AGV数量增加、运作环境空间复杂度增加，导致问题求解时间较长；二是AGV路径规划涉及多个AGV在同一运输区域协同工作，需要考虑系统的安全性和防冲突，对路径规划算法的性能及稳定性都提出了更高的要求，现有AGV路径规划方法难以应对多个AGV在同一运输区域协同工作情况下的防冲突性，同时求解路径速度慢，规划效率较低。本文针对自动化集装箱码头环境下的多AGV防冲突路径规划问题进行研究。

1　问题描述与数学模型

1.1　问题描述

在自动化集装箱码头中，AGV的行驶车道是由导引线或磁钉构成的。这些导引线或磁钉被安装在码头的地面上，通过传感器和控制系统来指引AGV的行驶方向和位置。目前，国内外对自动化集装箱码头AGV防冲突路径规划的研究大多集中在单向引导车道下的规划方法，但AGV通常可沿车道双向行驶，且双向引导车道可以提高AGV的运行效率。但为了保证行车安全和避免碰撞，码头控制系统通常会限制AGV反向行驶的条件，例如限制反向行驶的时间或指定特定的区域进行反向行驶等。图1所示为一种AGV的单、双向引导车道，箭头所指的方向为AGV被允许行驶的方向。

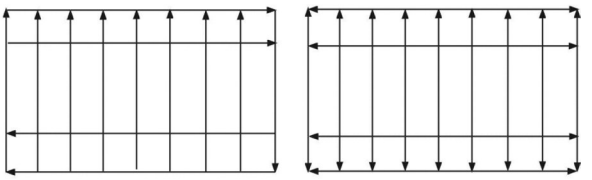


图1　AGV引导车道示意图

Fig.1　Schematic diagram of AGV guide path

在使用双向引导车道的自动化集装箱码头中，AGV在行驶过程中可能遇到以下三种冲突：①相向冲突。图2a中，当一辆AGV正在行驶时，如果另一辆AGV在同一条车道上相向而来，它们将无法同时通过，导致相向冲突。②节点冲突。图2b中，当多辆AGV同时到达同一个路口节点，会产生节点冲突。③同向冲突。图2c中，如果AGV在车道上停下来执行任务或变速、等待，可能会导致在同一车道上同向行驶的其他AGV无法通过，从而产生同向冲突。

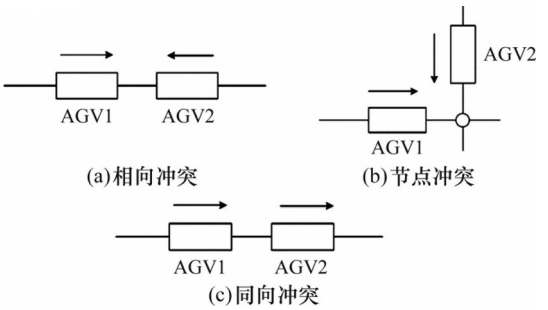


图2　AGV冲突类型

Fig.2　AGV conflict types

工作中的AGV往返于堆场与岸桥之间，在堆场的集装箱交接区装货或卸货，再通过双向行驶车道前往岸桥的集装箱交接区进行卸货或装货。码头运作时，将有多辆AGV同时工作，这导致双向行驶车道上有可能出现多辆AGV路径冲突的情况，冲突的发生会导致码头工作效率下降，进而影响港口吞吐量。因此，本文研究基于双向引导车道的自动化集装箱码头AGV防冲突路径规划方法，为各辆AGV规划合理的运行路径，以最小化所有任务最大完工时间。

1.2　模型的建立

1.2.1　假设条件

考虑到现实集装箱码头作业时的复杂性，为

了简化模型，作以下假设：①AGV的启动加速时间、刹车减速时间以及在转弯过程中所耗费的时间忽略不计；②忽略AGV在行驶过程中遇到的随机障碍物的情况；③AGV一个周期只执行一个任务，初始时刻，所有AGV均已被指派了相应的运输任务，且每个任务的起点与终点已知；④所有AGV速度一致，匀速行驶，不考虑启动时的加速过程和停车前的减速过程；⑤所有AGV不考虑中途充电的情况；⑥AGV到达对应的集装箱交接区后即视为任务完成，不考虑装卸箱时间及后续的运输任务。

1.2.2 符号定义

建模所用到的符号和变量定义如下：V为码头运输区域节点的集合，用i,j,n表示，$i,j,n\in\{0,1,\cdots,V\}$；E为引导车道线路的集合，(i,j)表示连接区域节点i和j的线路；K为AGV的集合，用k表示，$k\in\{0,1,\cdots,K\}$；T为时间集合，用t表示，$t\in\{0,1,\cdots,T\}$；T'为除去最后一个时刻的时间集合，$T'=T/t_{\max}$；C_k为连续变量，AGV k的完工时间；O_k为AGV k的起点，$k\in V$；S_k为AGV k的终点，$k\in V$；V_i为与节点i连接的节点集合，$i\in V$；x_{ijk}，如果线路(i,j)被AGV k访问，x_{ijk}取1，否则取0；y_{ijkt}，如果AGV k在时刻t占用线路(i,j)，y_{ijkt}取1，否则取0；a_{ijkt}，如果AGV k在时刻t到达线路(i,j)，a_{ijkt}取1，否则取0；d_{ijkt}，如果AGV k在时刻t离开线路(i,j)，d_{ijkt}取1，否则取0。

1.2.3 数学模型

目标函数：

$$\min\left(\max_{k\in K}C_k\right) \tag{1}$$

s.t.

$$C_k\geqslant t\cdot y_{iS_kkt}\quad \forall k\in K\quad i\in N_{S_k}\quad t\in T \tag{2}$$

$$x_{ijk}\leqslant\sum_{t\in T}y_{ijkt}\quad \forall k\in K\quad (i,j)\in E \tag{3}$$

$$\sum_{j\in N_{O_k}}x_{O_kjk}=1\quad \sum_{j\in N_{S_k}}x_{iS_kk}=1\quad \forall k\in K \tag{4}$$

$$\sum_{i\in N_j}x_{ijk}=\sum_{n\in N_j}y_{jnk}\quad \forall k\in K\quad j\in V(O_k\cup S_k) \tag{5}$$

$$\sum_{t\in T}\sum_{k\in K}y_{ijkt}\leqslant 1\quad \forall (i,j)\in E \tag{6}$$

$$\sum_{k\in K}\sum_{i\in N_j}y_{ijkt}\leqslant 1\quad \forall j\in V\quad t\in T' \tag{7}$$

$$\left.\begin{aligned}&a_{ijkt}\geqslant a_{ijk(t-1)}\quad d_{ijkt}\geqslant d_{ijk(t-1)}\\&\forall k\in K\quad t\in T\quad (i,j)\in E\end{aligned}\right\} \tag{8}$$

$$y_{ijkt}=a_{ijkt}-d_{ijkt}\quad \forall k\in K\quad t\in T\quad (i,j)\in E \tag{9}$$

$$x_{ijk}=a_{ijkT}\quad \forall k\in K\quad (i,j)\in E \tag{10}$$

$$\left.\begin{aligned}&\sum_{i\in N_j}d_{ijkt}=\sum_{n\in N_j}a_{jnkt}\\&\forall k\in K\quad [t\in T]\quad j\in V(O_k\cup S_k)\end{aligned}\right\} \tag{11}$$

$$d_{ijkt}\leqslant a_{ijkt}\quad \forall k\in K\quad [t\in T]\quad (i,j)\in E \tag{12}$$

$$\sum_{i\in N_{S_k}}d_{iS_kkT}\geqslant 1\sum_{i\in N_{O_k}}a_{O_kjk0}=1\quad \forall k\in K \tag{13}$$

$$\sum_{j\in N_{O_k}}x_{jO_kk}=0\quad \forall k\in K \tag{14}$$

$$\sum_{j\in N_{S_k}}x_{S_kik}=1\quad \forall k\in K \tag{15}$$

$$\left.\begin{aligned}&x_{ijk}\in\{0,1\}\quad y_{ijk}\in\{0,1\}\quad a_{ijk}\in\{0,1\}\\&d_{ijk}\in\{0,1\}\quad \forall k\in K\quad (i,j)\in E\end{aligned}\right\} \tag{16}$$

其中：式(1)为目标函数，目标是最小化所有任务最大完工时间；约束式(2)表示AGV占用通向终点的边的时间范围，定义了将占用的最大时刻作为AGV k的完工时间；约束式(3)表示当AGV k经过边(i,j)时才有占用时间，任何时刻都不占用边(i,j)时就认为其不经过边(i,j)，从而绑定x和y的关系；约束式(4)表示AGV从其起点出发、终止于对应终点；约束式(5)表示除去AGV的起始节点与终止节点，所有中间节点的流平衡约束式；约束式(6)是线路上的容量约束式，表示任何时刻每条线路上最多只能被一辆AGV占据；约束式(7)是节点容量约束式，表示一个节点任一时刻最多只能容纳一辆AGV；约束式(8)规定AGV到达和离开线路(i,j)的时间连通性；约束式(9)给出决策变量之间的关系；约束式(10)是AGV在终点的物理网络和时空网络之间的关系；约束式(11)和(12)表示AGV访问线路(i,j)时到达时间与离开时间的连续性；约束式(13)是AGV起点和终点的时间约束式，确保AGV在设定的时间移动；约束式(14)规定AGV从起点到终点的路径选择唯一，即在给定的情况下只选择一条路径到达目标节点i；约束式(15)规定从起点出发的路径唯一；约束式(16)规定4个决策变量的范围。

2 考虑防绕远策略的近端策略优化算法设计

2.1 近端策略优化算法

作为经典路径优化问题的延伸，本文研究的自动化集装箱码头环境下的多AGV防冲突路径规划问题是典型的NP难问题。在码头现实作业场景下，随着问题规模的不断增长，基于数学模型的最优化方法存在维数爆炸的问题，通常难以在可接受时间内获得最优调度方案。近端策略优化(proximal policy optimization，PPO)算法是一种基于策略梯度的强化学习算法，因其有效的策略优化过程被广泛应用于路径规划等复杂决策问题。相较于信赖域策略优化(trust region policy optimization，TRPO)算法，PPO算法在保持高效计算的同时能够显著降低算法实现的复杂度。PPO算法的核心原理是通过在目标函数中引入一个裁剪项来约束新旧策略的差异，从而限制策

略更新的步长，最小化策略更新过程中的风险。PPO算法流程如图3所示。

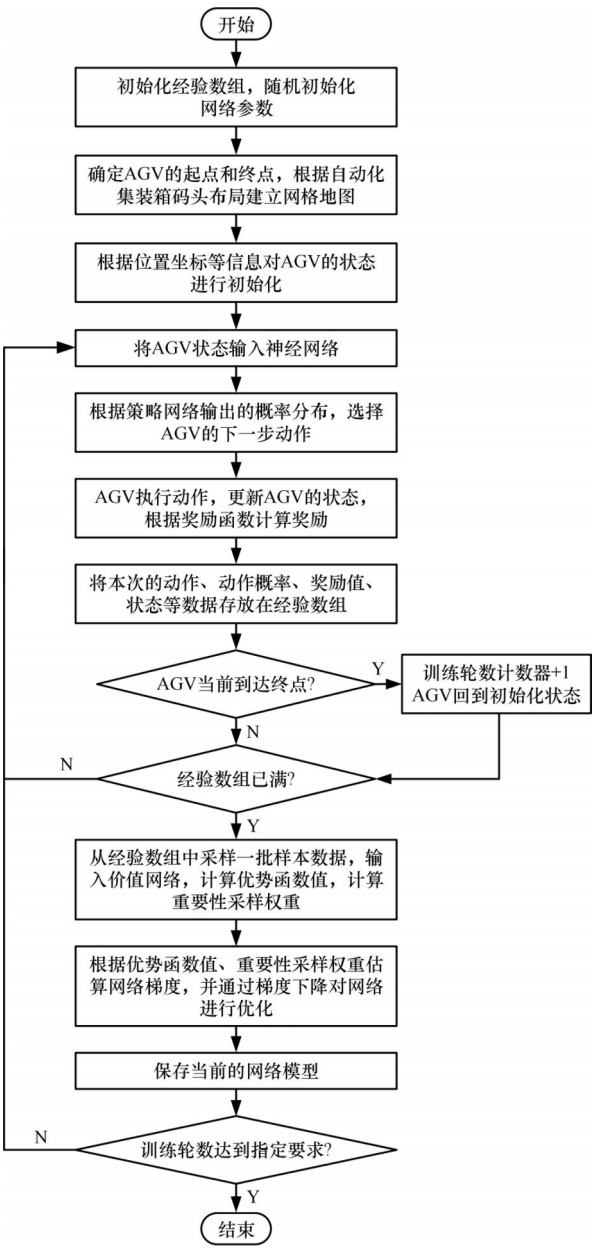


图3　PPO算法流程图

Fig.3　Flowchart of PPO algorithm

PPO算法的步骤如下：

1）初始化一个用于存储采样数据的经验数组buffer，其大小由超参数sample step决定。

2）初始化或更新概率分布dist，作为算法做出动作action决策的依据。

3）从dist中采样，获得下一步采取的动作action和该动作对应的概率prob。

4）执行action，获取奖励reward和下一个状态next_state。

5）将action、prob、reward、state、next_state等数据存储在buffer中。

6）检查buffer的大小是否达到事先设定的设定值batch size。若达到，继续进行步骤7）；若小于设定值，则跳转回步骤2）更新概率分布。

7）根据buffer中采样到的数据计算本次迭代的优势函数值。优势函数表示在状态s下，某动作a相对于平均值而言的优势，其定义为

$$A^{\pi}(s,a)=Q^{\pi}(s,a)-V^{\pi}(s) \tag{17}$$

优势函数的使用有助于减少策略梯度的方差，避免算法过拟合。PPO算法中一般会使用广义优势估计（generalized advantage estimation，GAE）来估计优势函数。GAE函数为

$$\hat{A}^{GAE(\gamma,\lambda)}=\sum_{l=0}^{T}(\gamma\lambda_{\mathrm{adv}})^{l}\delta_{t+l} \tag{18}$$

其中，T为经验数组的长度，一般等于实验中设定的超参数batch size；γ为折扣因子，用于权衡即时奖励与未来累积奖励的权重；λ_{adv}为一个超参数，需要手动设置；δ_t为时刻t的时序差分误差，其计算公式为

$$\delta_t=r_t+\gamma V(s_{t+1})-V(s_t) \tag{19}$$

式中：r_t为AGV在时刻t获得的即时奖励。

8）计算重要性采样权重ratio。PPO算法使用了重要性采样，使得训练数据得以复用。实验中可以通过调整超参数reuse times来设定训练数据的复用次数。使用重要性采样的一个例子如下：

$$J^{\theta'}(\theta)=E_{(s_t,a_t)\sim\pi_{\theta'}}\left(\frac{p_{\theta}(a_t|s_t)}{p_{\theta'}(a_t|s_t)}A^{\theta'}(s_t,a_t)\right) \tag{20}$$

其中，$J^{\theta'}(\theta)$指要优化的目标函数，θ是训练中要更新的参数，θ'是更新前的参数，s_t和a_t是由θ'作为示范采样出来的数据。$\frac{p_{\theta}(a_t|s_t)}{p_{\theta'}(a_t|s_t)}$为重要性采样权重。在PPO算法中，通过新旧策略的概率比计算重要性采样权重，结合裁剪函数进行约束，具体运用过程为

$$J_{\mathrm{PPO}}^{\theta^k}(\theta)\approx\sum_{(s_t,a_t)}\min\left(\frac{p_{\theta}(a_t|s_t)}{p_{\theta^k}(a_t|s_t)}A^{\theta^k}(s_t,a_t),\right.$$
$$\left.clip\left(\frac{p_{\theta}(a_t|s_t)}{p_{\theta^k}(a_t|s_t)},1-\varepsilon,1+\varepsilon\right)A^{\theta^k}(s_t,a_t)\right) \tag{21}$$

其中，ε是一个超参数，需要手动设定；$clip\left(\frac{p_{\theta}(a_t|s_t)}{p_{\theta^k}(a_t|s_t)},1-\varepsilon,1+\varepsilon\right)$是一个裁剪函数，其含义为：若括号中的第一项小于第二项，则输出第二项$1-\varepsilon$；若括号中的第一项大于第三项，则输出第三项$1+\varepsilon$。在AGV路径规划中，裁剪机制可以避免AGV在移动过程中急剧变化方向，例如避免从“向上移动”突然变为“向下移动”而导致冲突或绕远。

9）使用比率ratio计算价值网络和策略网络的损失函数，用于指导策略网络和价值网络的更新。策略网络的损失函数是基于策略梯度的方法

来计算的，损失函数可以使用新旧策略之间的比率ratio来调整，以减小新策略引入的变化。

例如，假设在状态s_t下，动作a_t的旧策略概率为$p_{\theta^k}(a_t|s_t)$，新策略概率为$p_\theta(a_t|s_t)$，并且相应的优势函数为$A^{\theta^k}(s_t,a_t)$，则策略网络的损失函数为

$$L_{\text{policy}}=\min\left(\frac{p_\theta(a_t|s_t)}{p_{\theta^k}(a_t|s_t)}A^{\theta^k}(s_t,a_t),\right.$$
$$\left.clip\left(\frac{p_\theta(a_t|s_t)}{p_{\theta^k}(a_t|s_t)},1-\varepsilon,1+\varepsilon\right)A^{\theta^k}(s_t,a_t)\right) \quad (22)$$

价值网络的损失函数可以使用均方差(mean squared erro，MSE)来衡量预测的状态值和实际回报之间的差距。假设$V(s)$是价值网络对状态s的估计值，R是实际回报，价值网络的损失函数为

$$L_{\text{value}}=MSE(V(s),R) \quad (23)$$

10)反向传播更新价值网络。对价值网络的参数进行反向传播计算梯度，并使用优化算法更新价值网络的参数，以更准确地估计状态值函数。大致步骤与更新策略网络时相似。

11)重复训练步骤。重复以上步骤，不断采样、计算优势函数、计算损失函数、更新策略网络和价值网络，直到达到指定的训练轮数。

2.2 基于PPO的双向引导车道下多AGV路径规划方法

2.2.1 工作环境模型的建立

本文将AGV的运行路径抽象为一个二维的网格空间。以3辆AGV行驶在10×10的网格空间中为例，绘制AGV工作环境的抽象地图，见图4。其中，大圆点代表AGV小车；网格线表示AGV的双向引导车道，AGV只能在网格线的交叉节点处改变方向(上、下、左、右)；小圆点表示集装箱交接区，地图上的每一辆AGV都要前往对应的集装箱交接区。

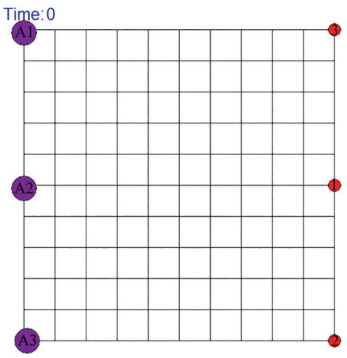


图4 AGV工作环境的表示

Fig.4 Representation of AGV working environment

2.2.2 动作空间的设计

在进行深度强化学习训练之前，需要为智能体设计合理的动作空间和状态空间。本文为每辆AGV设计了五种动作：向上、向下、向左、向右、停止，分别用0、1、2、3、4表示(式中的数字与其对应的动作如图5所示)：

$$\text{Action}\in\{0,1,2,3,4\} \quad (24)$$

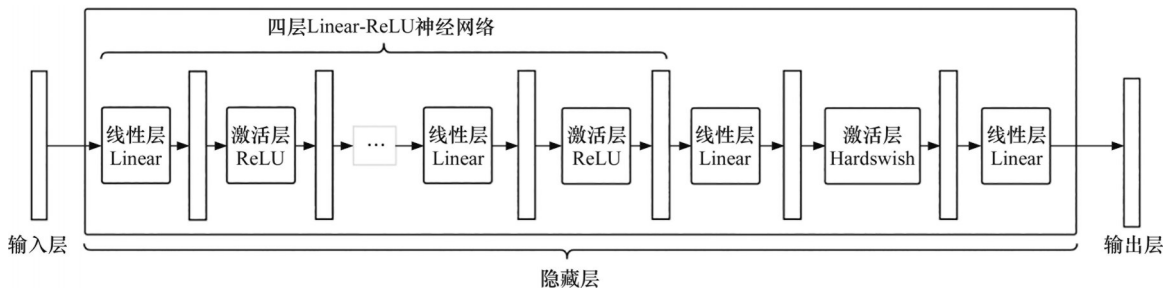


图5 神经网络结构

Fig.5 Neural network

2.2.3 状态空间与回报函数的设计

好的状态空间与回报函数相结合，可以帮助深度学习算法更快地做出更好的决策，提高学习效率，状态空间为

$$\boldsymbol{state}=(k_0,k_1,k_2,k_3,x_{\text{now}},y_{\text{now}},x_{\text{goal_now}},y_{\text{goal_now}},d_{\min}) \quad (25)$$

其中，$k_i,i\in\{0,1,2,3\}$表示当前时刻AGV k执行动作i时是否会超出地图边界，若超出，k_i为1，否则为0。假设AGV在一张有限大小的栅格地图上行驶，行驶的过程中显然不可超过地图边界，但神经网络在训练过程中有可能作出使AGV驶出边界的决策，所以在状态空间中提前告知AGV四周是否存在边界，可以有效提高训练效率。x_{now}和y_{now}分别为AGV小车当前所处位置的横坐标和纵坐标。神经网络必须掌握AGV小车的当前位置坐标，否则无法作出决策。$x_{\text{goal_now}}$和$y_{\text{goal_now}}$分别为AGV小车当前位置与目标位置的相对横坐标和相对纵坐标。对相对坐标进行预处理，可以加快神经网络的训练速度。$d_{\min}$表示当前AGV与附近最近AGV之间的距离。在多AGV共同工作的情况下，希望AGV之间能够避免冲突、死锁的情况，当$d_{\min}$过小时，可以对智能体施加一定的惩罚，以预防冲突现象的发生。动作空间的各参数及其含义见表1。

智能体做出一个动作，会导致环境的状态发

表1　动作空间参数

Tab.1　Action space parameters

参数	含义
k_0	AGV从当前位置向上运动是否会超过边界
k_1	AGV从当前位置向下运动是否会超过边界
k_2	AGV从当前位置向左运动是否会超过边界
k_3	AGV从当前位置向右运动是否会超过边界
x_{now}	AGV当前所处位置的横坐标
y_{now}	AGV当前所处位置的纵坐标
$x_{\text{goal_now}}$	AGV当前位置与目标位置的相对横坐标
$y_{\text{goal_now}}$	AGV当前位置与目标位置的相对纵坐标
$d_{\min}$	当前AGV与附近最近AGV之间的距离

生改变，算法需要根据当前环境的状态来给出一个奖励值，以引导智能体做出更好的动作。奖励值通过回报函数得出，回报函数的设计会直接影响算法的收敛速度和性能。在深度强化学习中，需要额外注意避免稀疏回报[28]问题，稀疏回报会导致算法难以收敛，所以在设计回报函数时，最好的情况是智能体的每一步都能得到回报函数的反馈，并且各项奖励值的相对大小要足够合理。

为使路径规划中AGV尽快驶向目标点，应为PPO算法设计防绕远机制，具体而言，在回报函数中引入回报项：其中，$d_{\text{goal_now}}$为AGV当前位置与目标位置的直线距离，$d_{\text{goal_next}}$为AGV下一步位置与目标位置的直线距离。该回报项表明，如果AGV在行驶过程中的某一步做出了远离终点的动作，则会被施加惩罚。惩罚值与AGV当前位置和终点位置的距离有关，AGV距离终点越远，惩罚值越大。即

$$R=-kd^2_{\text{goal}_{\text{now}}}\quad d_{\text{goal_now}}<d_{\text{goal_next}} \tag{26}$$

为了选取合适的参数，对上式中的参数值进行对比仿真实验。以10×10的双向引导车道下5辆AGV的路径规划问题为例进行仿真实验，k值分别取从2.5到5，步长为0.5的6个数值分别为

$$k\in\{2.5,\ 3,\ 3.5,\ 4,\ 4.5,\ 5\} \tag{27}$$

使用每种取值情况下的算法对模型训练100轮，并重复仿真实验5次，对5次训练后所得的训练结果求均值。进而采用多样本Friedman分析方法对不同k值下算法表现的稳定性进行检验。Friedman检验用于研究多个配对定量数据是否存在显著性差异。当p值小于0.01或0.05时，说明各项数据中存在显著性差异，即重复实验的结果存在显著性差异。k取3或4.5时，$p>0.05$，说明算法表现较为稳定，为了引导防止绕远，取$k=4.5$。回报函数设计为

$$R=\begin{cases}-1000 & \text{发生相向或节点冲突}\\ -800 & \text{AGV的下一步会越界}\\ 200\,000 & \text{AGV到达目标点}\\ -\dfrac{50}{d_{\min}} & 0<d_{\min}\leqslant 2\\ -4.5d^2_{\text{goal_now}} & d_{\text{goal_now}}<d_{\text{goal_next}}\\ -d^2_{\text{goal_now}} & \text{AGV为其他合法状态}\end{cases} \tag{28}$$

回报项数值的相对大小参照蔡泽等[29]、SARTORETTI等[30]在相关实验中的设置。式(28)说明，如果算法作出的决策会使当前AGV与其他AGV发生相向冲突或节点冲突，则得到的奖励值为−1000；如果算法作出的决策会使AGV越界，则得到的奖励值为−800；如果算法作出的决策会使AGV到达目标位置，则得到的奖励值为200 000，目的是增强到达目标点的吸引力；如果当前AGV附近有距离很近的其他车辆，则存在发生同向冲突的可能，则得到的奖励值为一个与距离成反比的负数；如果AGV下一时刻与目标点的距离大于当前时刻与目标点的距离，则说明AGV有远离终点的趋势，此时对AGV施加惩罚，AGV当前位置与目标位置的直线距离越远，惩罚越大，这一项可以避免AGV反向行驶或绕远；如果AGV为其他合法状态，则得到的奖励值为一个负数，即进行一个惩罚，AGV当前位置与目标位置的直线距离越远，惩罚越大，这一项惩罚可以引导AGV更快地到达目标位置。

2.2.4　神经网络的搭建

为拟合深度强化学习中的价值函数和策略函数，搭建一个由五层全连接层组成的全连接神经网络。该神经网络中，使用了线性函数Linear：

$$y=xA^{\mathrm{T}}+b \tag{29}$$

前4层选用激活函数ReLU，第5层选用激活函数Hardswish，分别为

$$\mathrm{ReLU}(x)=(x)^{+}=\max(0,x) \tag{30}$$

$$\mathrm{Hardswish}(x)=\begin{cases}0 & x\leqslant -3\\ x & x\geqslant +3\\ x(x+3)/6 & \text{其他}\end{cases} \tag{31}$$

2.2.5　算法步骤

以PPO算法作为深度强化学习的策略学习方法，对所构建的神经网络进行参数更新。基于PPO算法的AGV路径规划算法步骤见表2。当环境中的所有AGV都到达目标位置后，一轮训练结束，可获取每辆AGV完成任务所花的时间，以及环境中所有AGV任务的最大完工时间。

3　仿真实验结果分析

3.1　仿真实验设置

使用计算机进行仿真实验，PC操作系统为

表2 基于PPO算法的AGV路径规划算法步骤

Tab.2 Implementation steps of AGV path planning algorithm based on PPO

PPO算法步骤
Input：x_{now}，y_{now}，x_{goal}，y_{goal}，buffer，γ，ε，批次大小B，数据复用次数N
Output：最大完工时间C_{max}
$C_{max}=0$ //初始化最大完工时间
$state=[0,0,0,0,x_{now},y_{now},x_{goal_now},y_{goal_now},-1]$ //初始化AGV状态
$t_{total_k}=0$ //初始化工作时间
repeat
根据$state$和γ，使用PPO算法决策下一步坐标$[x_{next},y_{next}]$，将各个动作的概率记为数组$prob$
$[x_{now},y_{now}]=[x_{next},y_{next}]$
$[x_{goal_now},y_{goal_now}]=[x_{next_now},y_{next_now}]$
根据$[x_{now},y_{now}]$计算k_0，k_1，k_2，k_3，d_{min}
$state=[k_0,k_1,k_2,k_3,x_{now},y_{now},x_{goal_now},y_{goal_now},d_{min}]$ //更新AGV状态
根据$state$计算奖励$reward$
$list=[state,reward,prob]$
$buffer=buffer+list$ //将$list$记录到经验数组$buffer$中
if $buffer_length$ mod 2^9 == 0 then
根据$buffer$，B，R，使用PPO算法更新神经网络参数
$t_{total_k}=t_{total_k}+1$
until $[x_{now},y_{now}]=[x_{goal},y_{goal}]$
$C_k=t_{total_k}$
$C_{max}=\max(C_{max},C_k)$ //更新最大完工时间
return C_{max}

Windows 10，GPU为NVIDIA GeForce MX250，程序开发环境为Python 3.6，Pytorch 1.5，开发平台为PyCharm 2021.1.1。

训练前，PPO算法设置的相关超参数值见表3。有关超参数参考SCHULMAN等[31]提出的PPO算法框架，并根据经典强化学习算法的经验值以及实际场景的需求进行设置，取值均在强化学习中常见的合理范围内。其中，学习率α和批次大小(batch size)通过多次仿真实验调整，以兼顾模型收敛速度与稳定性；折扣因子γ根据AGV路径规划任务的冲突避免需求进行设置；裁剪参数ε同样参考文献[31]提出的PPO算法标准设置；其余参数采用强化学习常见值。以上参数设置经过多轮仿真实验验证，能够实现良好的收敛效果。

表3 PPO算法的相关超参数

Tab.3 Hyperparameters of PPO algorithm

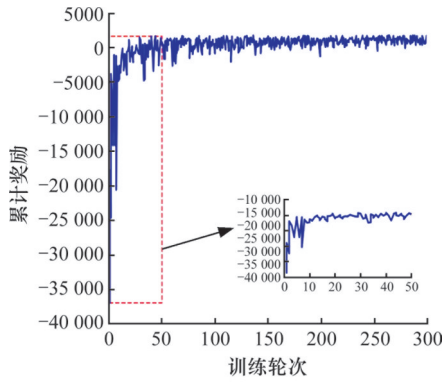
参数	取值
学习率α	0.000 007
折扣因子γ	0.88
裁剪参数ε	0.2
批次大小(batch size)	16
单轮更新的采样步数(sample step)	512
数据复用次数(reuse times)	2
限制新旧策略总体差异的系数λ_{entr}	0.025
GAE调整方差与偏差的系数λ_{adv}	0.98

为验证所提算法在不同问题复杂度下的适用性与求解性能，根据水平运输区域的网格地图大小及同时作业的AGV数量，将仿真实验划分为小、中、大三种规模。小规模算例中，有11条平行于岸线的双向引导车道和11条垂直于岸线的双向引导车道。在某一时刻t，仿真实验测试的AGV都处于工作状态，这些AGV都被要求前往对应的岸桥或堆场完成装卸任务，AGV数量分别设置3、5、10。中规模算例中，有21条平行于岸线的双向引导车道和21条垂直于岸线的双向引导车道，分别设置了5辆、10辆、15辆AGV。大规模算例中，分别设置了20辆、25辆、30辆AGV，该码头有31条平行于岸线的双向引导车道和31条垂直于岸线的双向引导车道。所有算例中，在某一时刻t，各AGV都处于工作状态，AGV根据调度指令前往对应的岸桥或堆场完成装卸集装箱任务。其中，AGV数量参考CAO等[32]在相关实验中的设置，确保了仿真实验设计的合理性与可比性。

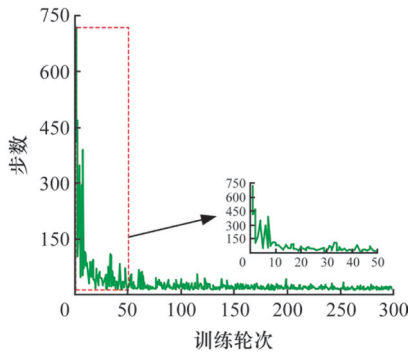
3.2 防绕远机制对训练过程的影响分析

为进一步验证所提防绕远策略的有效性，分析PPO算法回报函数的设计对训练效果的影响，采用式(26)所示的回报项。该回报项的作用是避免AGV在行驶过程中的绕远动作。使用考虑防绕远机制的PPO算法训练300轮，对算法的训练过程进行分析，训练结果如图6所示。图6a为累积奖励曲线，由结果可知，在训练初期，累积奖励值存在较大波动，这是由于智能体处于探索阶段，策略尚未稳定。随着训练轮次的增加，累积奖励曲线呈现明显的上升趋势，在第50轮左右趋于平稳，这表明PPO算法能够有效学习最优策略。图6b为每轮步数曲线，训练初期步数波动较大，随着算法收敛，每轮步数快速下降并稳定在较低水平。以上仿真实验结果说明所提PPO算法具有良好的收敛性。

将不考虑防绕远机制的PPO算法与改进后考虑防绕远机制的PPO算法分别训练300轮，对比两种算法在训练过程中的避碰等待率。避碰等待率是指智能体在训练过程中因避免碰撞而选择停车等待动作的比例，避碰等待率越小，路径规划质量越高。由图7可知，在训练过程中，不考虑防绕远机制的PPO算法的避碰等待率在训练结束后仍维持在较高水平，这说明AGV之间频繁发生冲突需要等待避让。相比之下，考虑防绕远机制的改进PPO算法的避碰等待率在训练初期快速下降，最终稳定在10%左右，表明该机制能够显著减少AGV之间的冲突等待，提高作业效率。



(a)累计奖励



(b)每轮步数

图6　改进PPO算法训练性能

Fig.6　Training performance of the improved PPO algorithm

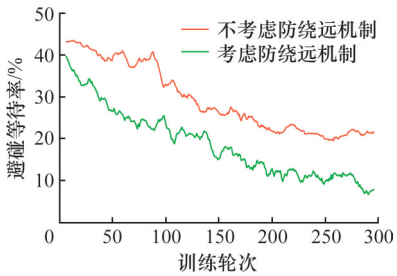
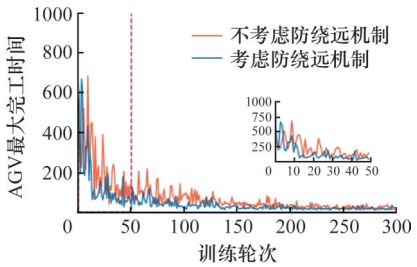


图7　两种算法的避碰等待率对比

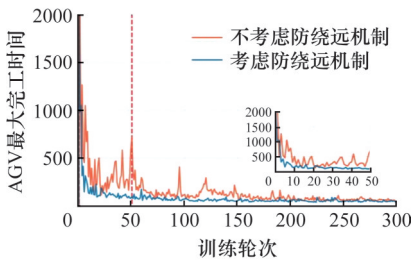
Fig.7　Comparison of collision avoidance waiting rate between two algorithms

将不考虑防绕远机制的PPO算法与改进后的考虑防绕远机制的PPO算法进行对比，分析训练效果的差异。使用不考虑防绕远机制的PPO算法和考虑防绕远机制的PPO算法训练300轮，训练结果如图8所示。

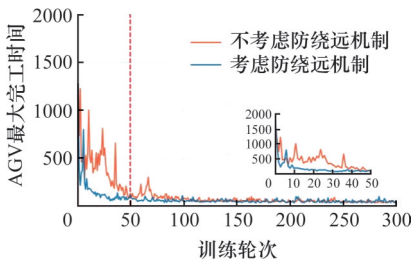
由训练结果可知，随着训练轮数的增加，两种算法均能达到收敛。图8a中的小规模算例，考虑防绕远机制的PPO算法的收敛速度较快，在第100轮左右开始收敛；不考虑防绕远机制的PPO算法的收敛速度稍显逊色，在150轮左右收敛。而图8b、图8c中的中规模和大规模算例，随着训练轮数的增加，两种算法均能基本达到收敛，但考虑防绕远机制的PPO算法在第50轮训练后的表现已趋于平稳，而不考虑防绕远机制的PPO算法在训练前期都处于剧烈波动状态。这说明考虑防绕远机制的PPO算法在更大的算例仿真实验中依然有着较好的稳定性及较快的收敛速度。



(a)小规模算例



(b)中规模算例



(c)大规模算例

图8　PPO算法训练结果

Fig.8　Training results of PPO algorithm

3.3　改进PPO算法性能对比分析

为进一步验证改进PPO算法在复杂场景中的适用性，将考虑防绕远机制的PPO算法与商业求解器Gurobi、启发式搜索算法A*和遗传算法进行对比仿真实验，重点对比考虑防绕远机制的PPO算法与其他三种算法在不同规模问题下的性能表现。仿真实验结果见表4。

由表4可以看出，在10×10小地图上求解5辆AGV的路径规划问题时，改进PPO算法所求得的最大完工时间更优，且求解时间显著短于Gurobi。当问题规模扩大，即在涉及更多数量的AGV和更复杂场景的情况下，Gurobi无法在合理时间内求解，这是由于问题的复杂度增加，Gurobi在求解过程中需要处理庞大的搜索空间，导致计算时间急剧增长，难以在有限的时间内找到全局最优解。而考虑防绕远机制的PPO算法在大规模场景下仍能快速求得最优解，这表明改进PPO算法在大规模或动态环境中能够表现出更强的适应性和计算效率。

表4 三种算法与改进PPO算法在不同规模下的性能对比

Tab.4 Performance comparison of three algorithms and the improved PPO under different scales

地图大小	AGV个数	最大完工时间(Makespan)				求解时间/s			
		Gurobi	A*	GA	改进PPO	Gurobi	A*	GA	改进PPO
10×10	3	15	15	15	15	4.33	0.01	0.50	0.37
	5	17	19	19	16	43.76	0.02	1.85	0.97
	10	18	18	18	21	478.96	0.03	3.67	1.91
20×20	5		31	31	31		0.08	4.14	0.62
	10		37	35	36		0.27	8.15	1.37
	15		34	34	34		0.34	11.61	2.06
30×30	20		47	49	49		4.39	26.38	2.11
	25		43	43	43		7.56	25.57	3.31
	30		45	45	46		13.52	36.54	5.06

对于不同规模问题，A*算法和考虑防绕远机制的PPO算法求得的最大完工时间相近，这表明改进PPO算法在解的质量上能够接近传统求解算法。当求解10×10地图5辆AGV的路径规划问题时，改进PPO算法求得的最大完工时间为16，相比A*算法所求得的结果，目标值降低了15.79%。随着问题规模的增大，A*算法求解时间的增速显著大于改进PPO算法的相应值。在10×10地图3辆AGV场景下，A*算法的求解时间为0.01 s，改进PPO算法则为0.37 s；而在30×30地图30辆AGV场景下，A*算法的求解时间达到13.52 s，改进PPO算法仅需约5.06 s，由此可见，考虑防绕远机制的PPO算法在大规模算例下具有更强的稳定性、更好的路径平滑性和更低的潜在冲突风险。

遗传算法作为一种经典的元启发式方法，广泛应用于组合优化问题。由表4可以看出，GA通过选择、交叉和变异等操作进行全局搜索，随着问题场景复杂度的增加，GA的性能明显下降，在大规模场景下的求解时间远不如改进PPO算法。在30×30地图20辆AGV场景下，GA的求解时间为26.38 s，而改进PPO算法仅需2.11 s，求解效率提高了92%。相比之下，改进PPO算法在完成训练后可直接用于决策，无需重复进行迭代，且在各种场景下都能保持稳定的求解质量和较高的计算效率。

3.4 典型案例分析

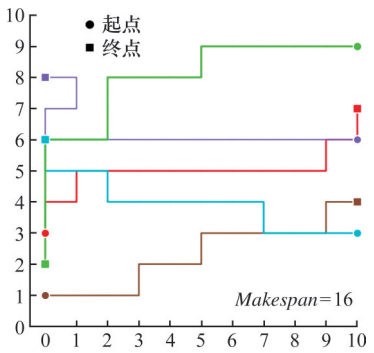
为进一步验证算法的路径规划效果，选取一个典型的小规模案例进行分析。表5给出了5辆AGV的起始位置与终点位置，AGV需要在对应的场景地图内完成装卸任务。其中，AGV1～2从堆场前往岸桥执行装箱任务，AGV3～5从岸桥返回堆场执行卸箱任务，形成典型的双向混合作业场景。

表5 5辆AGV起始位置与终点位置

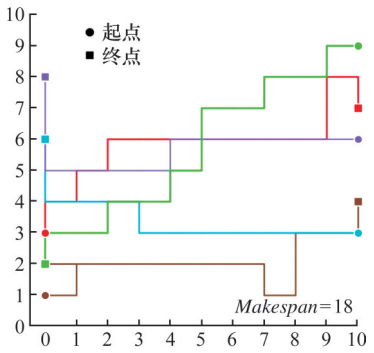
Tab.5 Starting and ending positions of 5 AGVs

AGV编号	起始位置坐标	终点位置坐标	任务类型
1	(0,1)	(10,4)	前往岸桥装箱
2	(0,3)	(10,7)	前往岸桥装箱
3	(10,3)	(0,6)	前往堆场卸箱
4	(10,6)	(0,8)	前往堆场卸箱
5	(10,9)	(0,2)	前往堆场卸箱

图9所示为两种算法在该算例下的路径规划结果。可以看出，图9a中考虑防绕远策略的PPO算法规划的路径更加简洁直接，虽然个别AGV在终点附近存在轻微的绕行以避让其他车辆，但总体来看，各AGV路径较为平直，转向次数少；



(a)考虑防绕远机制的PPO算法路径



(b)不考虑防绕远机制的PPO算法路径

图9 两种算法的路径规划结果对比

Fig.9 Comparison of path planning results between the two algorithms

而图9b中不考虑防绕远策略的PPO算法规划的部分路径存在反复移动和明显的绕远行为，如AGV1在起点附近、AGV3在运输过程中均出现了不必要的绕远。

改进后的考虑防绕远机制的PPO算法和不考虑防绕远机制的PPO算法在小规模算例下的规划结果对比见表6。对比指标包括地图大小、AGV个数、算法类型、训练轮数、最优的最大完工时间、AGV完工时间之和以及AGV路径步数之和。其中，完工时间以AGV每移动一步的时间作为单位时间进行计算，且包含AGV停车等待的时间。路径步数则表示AGV从起点到终点的移动次数。

表6　两种PPO算法求解双向车道上的AGV路径规划结果对比

Tab.6　Comparison of path planning results for AGVs on bidirectional lanes using two PPO algorithms

地图大小	AGV个数	PPO算法是否考虑防绕远机制	训练轮数	最大完工时间（Makespan）	AGV完工时间之和	AGV路径步数之和
10×10	5	否	300	18	80	79
		是	300	16	69	73
10×10	10	否	300	23	195	185
		是	300	21	168	167

由表6可以看出，考虑防绕远机制的PPO算法在两种规模算例中所求得的最大完工时间、AGV完工时间之和以及AGV路径步数之和均优于不考虑防绕远机制的PPO算法的相应值。两种算法的训练轮数均为300，这表明防绕远机制在不增加训练成本的前提下能够有效提高路径规划质量和作业效率，验证了该机制的有效性。

4　结语

本研究的主要贡献如下：①根据自动化集装箱码头的装卸搬运流程对AGV防冲突路径规划问题进行研究，建立了以最小化所有任务最大完工时间为目标的数学规划模型；②针对单向路径存在的不足，探索双向行驶路径，旨在实现道路资源的最大化利用与调度灵活性的显著提高；③提出采用深度强化学习方法进行AGV防冲突路径规划，引入防绕远机制，提高了PPO算法求解自动化集装箱码头AGV防冲突路径规划的性能及求解效率。

在与Gurobi求解器的性能对比仿真实验中，小规模算例下改进PPO算法的最大完工时间与Gurobi求解器差距较小，但是中、大规模算例下Gurobi因求解时间过长而无法完成求解，改进PPO算法仍能快速求解，求解时间均在5.06 s以内，验证了模型的可行性与改进PPO算法的有效性；在与A*算法和遗传算法的对比仿真实验中，不同规模算例下改进PPO算法的最大完工时间平均差距仅为1.10%和1.25%，随着问题规模的增大，改进PPO算法的求解时间增速显著低于A*算法和遗传算法的相应值，验证了改进PPO算法在性能上具有较大的优势；在与基础PPO算法的对比仿真实验中，不同规模算例下改进PPO算法均具有更好的稳定性及更快的收敛速度。本文研究的多AGV防冲突路径规划问题能够有效提高码头水平运输区域的效率，对自动化集装箱码头智能化决策具有重要的参考价值。

考虑到现实集装箱码头作业时的复杂性，在模型构建中忽略了一些实际场景中的动态变化或复杂约束，未来研究可将更多现实因素纳入模型中，针对更加复杂的场景（如动态障碍、多任务冲突等）进行研究，以更全面地满足实际应用需求，进一步提高码头运营效率。

参考文献：

[1] ROY D, de KOSTER R, BEKKER R. Modeling and Design of Container Terminal Operations[J]. Operations Research, 2020, 68(3): 686-715.

[2] YUAN Ruiping, DONG Tingting, LI Juntao. Research on the Collision-free Path Planning of Multi-AGVs System Based on Improved A* Algorithm[J]. American Journal of Operations Research, 2016, 6(6): 442-449.

[3] YANG Yongsheng, ZHONG Meisu, DESSOUKY Y, et al. An Integrated Scheduling Method for AGV Routing in Automated Container Terminals[J]. Computers & Industrial Engineering, 2018, 126: 482-493.

[4] LÜ Min, GAO Tong, ZHANG Nian. Research of AGV Scheduling and Path Planning of Automatic Transport System[J]. International Journal of Control and Automation, 2016, 9(4): 1-12.

[5] KIM C W, TANCHOCOJ J M A. Operational Control of a Bidirectional Automated Guided Vehicle System[J]. International Journal of Production Research, 1993, 31(9): 2123-2138.

[6] WANG Zehao, ZENG Qingcheng. A Branch-and-bound Approach for AGV Dispatching and Routing Problems in Automated Container Terminals[J]. Computers & Industrial Engineering, 2022, 166: 107968.

[7] 高一鹭，胡志华. 基于时空网络的自动化集装箱码头自动化导引车路径规划[J]. 计算机应用，2020，40(7)：2155-2163.
GAO Yilu, HU Zhihua. Path Planning for Automated Guided Vehicles Based on Tempo-spatial Network at Automated Container Terminal[J]. Journal of Computer Applications, 2020, 40(7): 2155-2163.

[8] SHEN Lixin, WANG Yaodong, LIU Kunpeng, et al. Synergistic Path Planning of Multi-UAVs for Air Pollution Detection of Ships in Ports[J]. Transportation Research Part E: Logistics and Transportation Review, 2020, 144: 102128.

[9] 梁承姬，沈珊珊，胡文辉. 基于路段时间窗考虑备选路径

的AGV路径规划[J]. 工程设计学报，2018，25(2)：200-208.
LIANG Chengji，SHEN Shanshan，HU Wenhui. AGV Path Planning Considering Alternative Paths Based on Time Window of Road Section[J]. Chinese Journal of Engineering Design，2018，25(2)：200-208.

[10] HU Yejun，DONG Liangcai，XU Lei. Multi-AGV Dispatching and Routing Problem Based on a Three-stage Decomposition Method[J]. Mathematical Biosciences and Engineering，2020，17(5)：5150-5172.

[11] XU Yixiang，QI Liang，LUAN Wenjing，et al. Load-in-load-out AGV Route Planning in Automatic Container Terminal[J]. IEEE Access，2020，8：157081-157088.

[12] 姜辰凯，李智，盘书宝，等. 基于改进Dijkstra算法的AGVs无碰撞路径规划[J]. 计算机科学，2020，47(8)：272-277.
JIANG Chenkai，LI Zhi，PAN Shubao，et al. Collision-free Path Planning of AGVs Based on Improved Dijkstra Algorithm[J]. Computer Science，2020，47(8)：272-277.

[13] TANG Gang，TANG Congqiang，CLARAMUNT C，et al. Geometric A-star Algorithm：an Improved A-star Algorithm for AGV Path Planning in a Port Environment[J]. IEEE Access，2021，9：59196-59210.

[14] 丁一，袁浩，方怀瑾，等. 考虑冲突规避的自动化集装箱码头AGV优化调度方法[J]. 交通信息与安全，2022，40(3)：96-107.
DING Yi，YUAN Hao，FANG Huaijin，et al. An Optimal Scheduling Method of AGVs at Automated Container Terminal Considering Conflict Avoidance[J]. Journal of Transport Information and Safety，2022，40(3)：96-107.

[15] ZHONG Meisu，YANG Yongsheng，DESSOUKY Y，et al. Multi-AGV Scheduling for Conflict-free Path Planning in Automated Container Terminals[J]. Computers & Industrial Engineering，2020，142：106371.

[16] WU Maopu，GAO Jian，LI Le，et al. Control Optimisation of Automated Guided Vehicles in Container Terminal Based on Petri Network and Dynamic Path Planning[J]. Computers and Electrical Engineering，2022，104：108471.

[17] CHEN Tianjian，SUN Yuan，DAI Wei，et al. On the Shortest and Conflict-free Path Planning of Multi-AGV System Based on Dijkstra Algorithm and the Dynamic Time-window Method[J]. Advanced Materials Research，2013，645：267-271.

[18] LIU Wenqian，ZHU Xiaoning，WANG Li，et al. Multiple Equipment Scheduling and AGV Trajectory Generation in U-shaped Sea-rail Intermodal Automated Container Terminal[J]. Measurement，2023，206：112262.

[19] CAO Menglong，PENG Zhang. Research on Loading and Unloading Path Optimization for AGV at Automatic Container Terminal Based on Improved Particle Swarm Algorithm[C]//2020 4th Annual International Conference on Data Science and Business Analytics (ICDSBA). IEEE，2020：1-5.

[20] LU Tong，SUN Zhaohui，QIU Siqi，et al. Time Window Based Genetic Algorithm for Multi-AGVs Conflict-free Path Planning in Automated Container Terminals[C]// 2021 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM). IEEE，2021：603-607.

[21] 岳春擂，黄俊，邓乐乐. 改进蚁群算法在AGV路径规划上的研究[J]. 计算机工程与设计，2022，43(9)：2533-2541.
YUE Chunlei，HUANG Jun，DENG Lele. Research on Improved Ant Colony Algorithm in AGV Path Planning [J]. Computer Engineering and Design，2022，43(9)：2533-2541.

[22] 荀燕琴. 基于群体智能优化的AGV路径规划算法研究[D]. 长春：吉林大学，2017.
XUN Yanqin. Research on AGV Path Planning Algorithm Based on Group Intelligence Optimization[D]. Changchun：Jilin University，2017.

[23] CRUZ D L，YU Wen. Path Planning of Multi-agent Systems in Unknown Environment with Neural Kernel Smoothing and Reinforcement Learning [J]. Neurocomputing，2017，233：34-42.

[24] POPPER J，YFANTIS V，RUSKOWSKI M. Simultaneous Production and AGV Scheduling Using Multi-agent Deep Reinforcement Learning[J]. Procedia CIRP，2021，104：1523-1528.

[25] ÇETINKAYA M. Multi-agent Path Planning Using Deep Reinforcement Learning [J]. ArXiv Preprint ArXiv：2110.01460，2021.

[26] CHEN Xinqiang，LIU Shuhao，ZHAO Jiansen，et al. Autonomous Port Management Based AGV Path Planning and Optimization via an Ensemble Reinforcement Learning Framework [J]. Ocean & Coastal Management，2024，251：107087.

[27] WANG Tingzhong，ZHANG Binbin，ZHANG Mengyan，et al. Multi-UAV Collaborative Path Planning Method Based on Attention Mechanism [J]. Mathematical Problems in Engineering，2021，2021：6964875.

[28] RIEDMILLER M，HAFNER R，LAMPE T，et al. Learning by Playing—Solving Sparse Reward Tasks from Scratch[C]//Proceedings of the 35th International Conference on Machine Learning (ICML). Stockholmsmässan，2018：4344-4353.

[29] 蔡泽，胡耀光，闻敬谦，等. 复杂动态环境下基于深度强化学习的AGV避障方法[J]. 计算机集成制造系统，2023(1)：236-245.
CAI Ze，HU Yaoguang，WEN Jingqian，et al. Collision Avoidance for AGV Based on Deep Reinforcement Learning in Complex Dynamic Environment[J]. Computer Integrated Manufacturing Systems，2023(1)：236-245.

[30] SARTORETTI G，KERR J，SHI Yunfei，et al. PRIMAL：Pathfinding via Reinforcement and Imitation Multi-agent Learning[J]. IEEE Robotics and Automation Letters，2019，4(3)：2378-2385.

[31] SCHULMAN J，WOLSKI F，DHARIWAL P，et al. Proximal Policy Optimization Algorithms [PP/OL]. V2. arXiv(2017-08-28). https://doi.org/10.48550/arXiv.1707.06347.

[32] CAO Yu，YANG Ang，LIU Yang，et al. AGV Dispatching and Bidirectional Conflict-free Routing Problem in Automated Container Terminal[J]. Computers & Industrial Engineering，2023，184：109611.

（编辑 陈 勇）

作者简介：肖世昌，男，1987年生，副教授、博士。研究方向为港口调度优化/智能制造系统建模与调度优化。E-mail：scxiao@shmtu.edu.cn。郑 鹏*（通信作者），男，1990年生，讲师、博士。研究方向为复杂物流/制造系统的建模、仿真与优化。E-mail：zhengpeng@shmtu.edu.cn。

本文引用格式：

肖世昌，林轩，郑鹏，等. 基于近端策略优化算法的自动化集装箱码头自动导引车防冲突路径规划[J]. 中国机械工程，2026，37(8)：2017-2028.

XIAO Shichang，LIN Xuan，ZHENG Peng，et al. Anti-conflict Path Planning for AGVs in the Automated Container Terminals Based on Proximal Policy Optimization Algorithm[J]. China Mechanical Engineering，2026，37(8)：2017-2028.