[文章编号] 1007－7405(2024)04－0329－07 DOI:10.19715/j.jmuzr.2024.04.05

集装箱码头桥吊作业的强化学习优化

解静修，吴一亮

（集美大学海洋信息工程学院，福建 厦门 361021）

[摘要] 对码头桥吊装卸操作进行强化学习优化，通过Unity建立码头桥吊的模型及训练环境，用PPO算法对码头桥吊进行装卸操作训练，最终使桥吊学习在海浪及大风对集装箱船和吊具造成晃动影响时的平稳下落行为，提高桥吊抓取准确性并优化桥吊路径，实现更高效智能的码头装卸作业。

[关键词] 码头桥吊；Unity；强化学习；集装箱装卸作业

[中图分类号] TP 302.1

Reinforcement Learning Optimization of Bridge Craner Opearations at Container Terminals

XIE Jingxiu, WU Yiliang

(School of Ocean Information Egineering, Jimei University, Xiamen 361021, China)

Abstract: In this paper, reinforcement learning was used to optimize the loading and unloading operations of quay bridge cranes, a model and training environment was built for cranes through Unity, and PPO algorithm was used to train the loading and unloading operations of cranes, so that the cranes can learn to lower smoothly when waves and high winds affect the container ship and lifting gears. Eventually this can improve the accuracy of crane grasping and optimize the path of cranes to achieve more efficient and intelligent loading and unloading operations at terminals.

Keywords: crane; Unity; reinforcement learning; loading and unloading operation

0 引言

目前，尽管几乎所有类型的轨道式集装箱起重机都可以在远程控制下进行无人操作，但实际上这一转变仅是将操作员从现场转移到了远程，操作员仍需每天进行数小时的视觉操作。随着长时间的高负荷劳动，装载效率和精确度会严重下降。更重要的是桥吊会受天气、海浪等不可预见因素的影响，在操作过程中会出现吊具和集装箱摇晃的情况，这也会影响集装箱装载效率。因此，开发一种使吊具平稳下落的码头桥吊平台自适应调节方法，对于提升海运行业的效率性具有重要意义。

关于提高起重机的安全性和港口的装载效率研究，已取得一些成果。如：Kim等[1－3]提出了在模拟波浪形海洋环境下的动态定位控制系统以及用于自动定位吊具的算法；Liu等[4]通过分析集装箱顶部的红外线，拟议的系统能够计算出吊具和集装箱之间的相对位置；Mi等[5]用视觉非接触测量方法，实现了对集装箱三维姿态的实时定位；梁晓波等[6]提出了一种集装箱起重机自动装卸系统，通过解

[收稿日期] 2023－06－29

[基金项目] 福建省科技厅重大专项“集装箱码头装备群体智能协同作业关键技术研发及产业化”（2022HZ022019）；福建省自然科学基金项目“沉浸式视频流的低时延视频编码技术研究”（2021J01868）

[作者简介] 通信作者：吴一亮（1979—），博士，讲师，主要从事视频处理系统及VLSI设计方向研究。

析系统功能需求，给出了系统逻辑结构。但目前所有的研究方案仍然依赖人工操作吊具下降，缺乏对突发气候的应对能力。

强化学习算法是机器学习算法的分支，主要研究如何使智能体（agent）在与环境交互的过程中，通过不断的试验和反馈，学习到最优的行为策略。这种方法涵盖了动态规划、蒙特卡洛方法和时序差分学习等多种技术，可以用于解决各种问题，如游戏、机器人控制、自然语言处理等。相比于有监督学习中的“模型”，强化学习中的“智能体”强调的是，机器不但可以感知周围的环境信息，还可以通过做决策来直接改变这个环境，而不只是给出一些预测信号[8]。强化学习算法可以在不需要人为干预的情况下，通过观察集装箱的位置、形状等，学习自主调整自身动作，最终实现准确、高效的吊取操作。ML-Agents 是 Unity 开源的强化学习工具包，它的应用领域非常广泛，包括游戏开发、机器人控制、虚拟现实等领域。开发人员可以使用 ML-Agents 工具包来训练智能体解决各种复杂的任务。

在港口码头向智能化发展的背景下，本文采用了基于 Unity 的 ML-Agents 强化学习框架，通过构建智能体与环境之间的交互模型，以期实现码头桥吊装卸操作的自主学习和优化。

1 强化学习训练环境搭建

图 1 展示了集装箱桥吊在风浪影响下的作业情况。连接在缆绳上的吊具受天气以及桥吊移动的影响产生摇摆，集装箱受海浪影响也产生一定幅度的摇摆。为了保证强化学习训练学习的最终效果，要尽可能地还原真实作业环境。

集装箱码头强化学习训练模型借助 Unity 的强化学习工具包 ML-Agents 搭建，并在此基础上进行物理模型搭建。

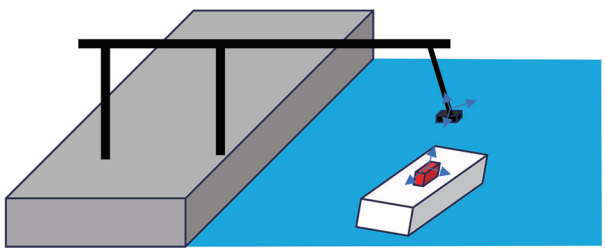


图 1 桥吊智能体作业示意图
Fig.1 Diagram of agent operation of bridge crane

1.1 强化学习环境设计

本文使用 Unity 模拟码头真实运行环境，环境包括桥吊模型、集装箱模型、港池、船舶。如图 2、3 所示。

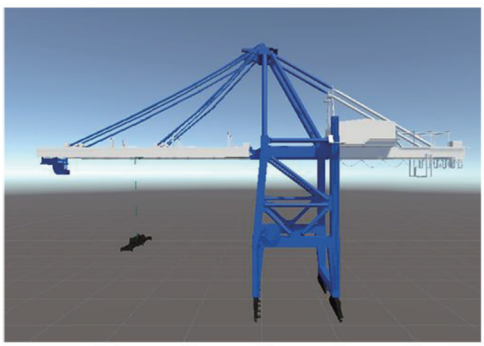


图 2 吊桥模型
Fig.2 Gantry model

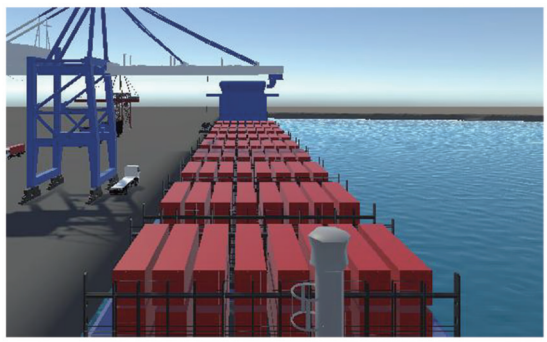


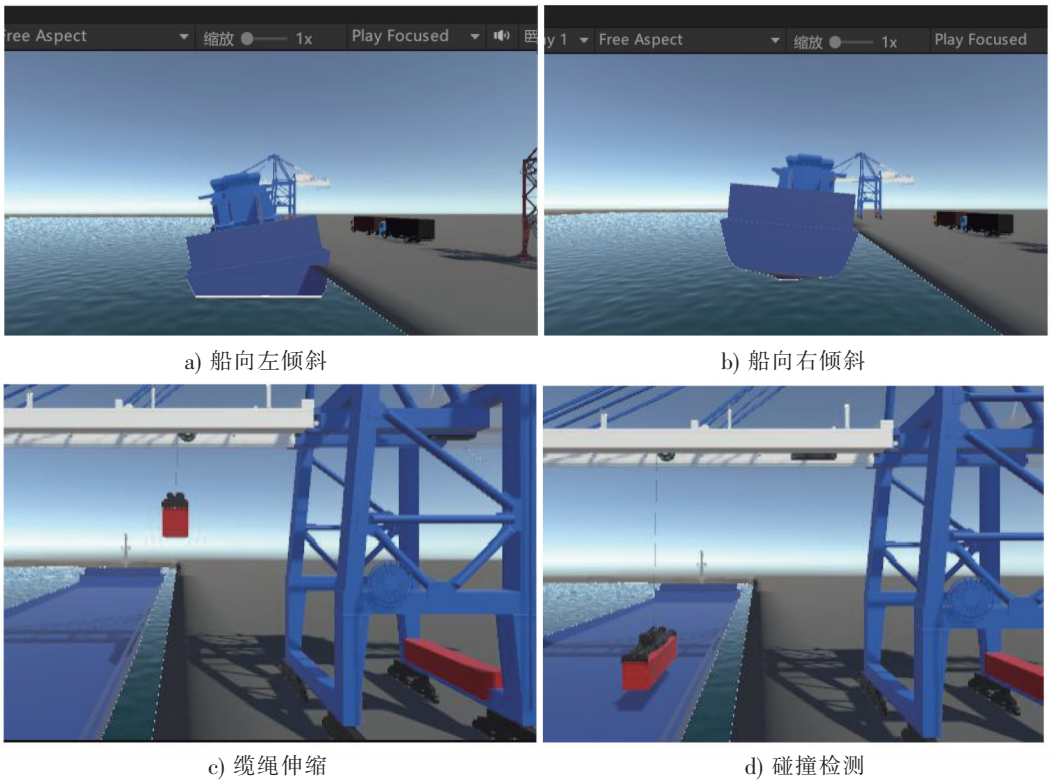
图 3 码头模型
Fig.3 Port model

在码头仿真环境中，港池是必不可少的场景因素。本次实验使用的是 Low Poly Tools Bundle 插件。它可以快速创建低面数的 3D 模型，并且可以根据具体情况调节水面反光、粒子强度、反射强度等参数。港池场景搭建过程中，还需要考虑到港池的深度和波浪强度等因素。港池模型如图 3 所示。

码头桥吊是用于港口集装箱装卸作业的机械设备，主要由三个部分构成：框架主体、可伸缩吊臂和吊具。框架主体上安装了可伸缩吊臂，可根据需要调节长度，使吊臂能够覆盖不同位置的货物。吊臂上配有小车和吊具，可以提起和移动集装箱。

在物理模型构建好以后，利用 Unity 脚本功能结合 ML-Agents 框架实现码头的功能逻辑。例如吊

桥移动功能、缆绳收放功能、海浪强度控制功能、吊具集装箱碰撞功能、任务失败成功检测功能等。详见图 4。



a) 船向左倾斜　　b) 船向右倾斜

c) 缆绳伸缩　　d) 碰撞检测

图 4　码头部分功能

Fig.4　Some functions of the terminal

在完成环境逻辑后，开始构建强化学习模型。主要分为以下 5 部分：

1）定义桥吊的强化学习任务。码头桥吊必须以最优路径将集装箱船上的集装箱准确吊起。

2）定义智能体的观察空间。包括关于桥吊当前状态的信息，例如桥吊的位置、吊臂和吊杆的位置、集装箱的位置等。

3）定义智能体的行动空间。包括启动和停止桥吊，控制桥吊、吊臂和小车的移动以及升降等。

4）编写智能体的行为决策代码。利用机器学习算法和奖惩机制，来选择智能体的下一步行动。

5）通过编写智能体的 C#脚本与 Unity 中的 ML-Agents API 进行交互，使智能体能够在桥吊环境中执行行动。

在此基础上进一步细化强化学习模型的各个关键要素。具体有：

1）环境设定—仿真码头的交互逻辑。例如桥吊移动作业，受海浪影响而摇晃的船，受桥吊移动和大风影响而晃动的吊具。

2）动作空间定义—定义智能体每个时间步所能采取的所有可能动作的集合。例如：抓取、上下移动、左右移动、前后移动。

3）状态空间定义—定义智能体所能感知到的所有可能状态的集合。例如：桥吊位置、集装箱位置、集装箱相对吊具的向量等。

4）奖励函数定义—衡量智能体行为优劣的重要指标，它会根据智能体的行为给出相应的回报。吊具做出缩短与集装箱距离的动作时，得到相应奖励；超出抓取范围时得到相应惩罚。

5）策略函数定义—策略函数是一个映射函数，它将一个状态作为输入，并给出相应的动作输出。本文选取 PPO（proximal policy optimization）算法作为策略选择。

6）训练智能体—根据训练数据和奖励函数，可以使用强化学习算法来调整智能体的参数，以便

最大化奖励。

7）优化和改进—对智能体的训练过程进行评估和改进，并根据需要调整环境设定、状态空间定义、动作空间定义、奖励函数、策略函数或训练数据的收集方法等。

1.2 强化学习算法设计

ML-Agents 工具包[10]带来了深度学习算法，如 PPO 和 SAC（soft actor critic）[11]，同时还提供了额外的算法，如模仿学习（IL）[12-13]和使用内在好奇心模块（ICM）[14]的好奇心驱动探索。所有这些算法都是通过本地运行的 Python 后台服务器实现的，可用 PyTorch 和 TensorFlow 实现这些算法，同时让开发者选择合适的框架[15]。

1.2.1 智能体

集装箱装卸作业由众多环节组成，智能体的选择显得格外重要。由于码头桥吊在实际操作中需要考虑到多个因素，如船只的大小、载重、风力等因素，因此其状态空间和动作空间非常庞大，那么智能体需要学习数据和做出的决策范围就会增加。由于吊具的状态空间和动作空间相对较少，使得算法更快速地将模型收敛，可以训练出更加精准、稳定的智能体，所以本文选择桥吊吊具作为智能体。

1.2.2 状态、动作及奖励设置

状态、动作与奖励机制是强化学习模型中的重要组成部分。

状态空间即对当前状态下外部环境信息的描述，智能体在进行决策时需要以状态空间为依据。通过对码头集装箱装卸作业的研究，状态空间的选取可以包含船舶位置信息、集装箱位置信息、吊具位置信息等方面。

动作空间的设计与集装箱的装卸作业有关，根据外部环境信息，智能体自主决定从状态空间中选择最优动作。吊具智能体动作空间包括：上下移动、左右移动、前后移动。

奖励机制的设置是智能体系统能否在不断学习中实现最优操作的关键，在对其进行设计时需要与集装箱装卸作业智能化的目标相结合，即智能实现高效可行的装卸作业计划。

1.2.3 算法设计

本文中状态、动作、奖励机制均依据吊具装卸作业的真实需求来设计，吊具智能体 PPO 决策模型算法如下所示（详见图 5）。

步骤 1）设环境信息 $S = [p^i, p^j, p^k]$，其中 p^i，p^j，p^k 分别表示桥吊位置、集装箱位置、集装箱相对吊具的向量等环境信息。将 S 输入到 actor-new 网络，得到 μ 和 σ 两个值。然后将这两个值作为正态分布的均值和方差构建正态分布 Normal0，再通过这个正态分布采样出来一个动作 a，输入到环境中得到奖励 r 和下一步的状态 $s\_$，然后存储 $[(s,a,r),\cdots]$，循环步骤 1）。直到存储了一定量的 $[(s,a,r),\cdots]$，注意这个过程中 actor-new 网络没有更新。

步骤 2）利用记忆池里存储的数据，根据总奖励 $R(\tau) = \sum_{t'=t} \gamma^{t'-t} r_{t'}$，得到 $R = [R[0], R[1], \cdots, R[t], \cdots, R[T]]$。其中：$t'$表示当前时间步，$\gamma$ 是学习因子，$r_{t'}$是当前时间步的奖励值。

步骤 3）将存储的所有 s 组合，输入到 critic-NN 网络中，得到所有状态的 $V\_$ 值，计算 $A_t = R - V\_$。

步骤 4）求 c_ loss = mean（square（策略网络）），然后反向传播更新 critic 网络。

步骤 5）将存储的所有 s 组合，输入 actor-old 和 actor-new 网络，分别得到正态分布 Normal1 和 Normal2；将存储的所有 action 组合为 actions 输入到正态分布 Normal1 和 Normal2，得到每个 actions 对应的 prob1 和 prob2，然后用 prob2 除以 prob1 得到 important weight，也就是 ratio。

步骤 6）a_loss = mean(min(ration $* A_t$, clip(ratio, $1-\xi$, $1+\xi$) $* A_t$))，然后反向传播，更新 actor-new 网络。

步骤 7）循环步骤 5）~6）一定次数后，循环结束，用 actor-new 网络权重来更新 actor-old 网络。

步骤 8）循环步骤 1）~7）。

重复上述步骤 1）～8），一直重复到训练效果达到预期目标时停止，训练结束后价值网络就是最终的模型结果。

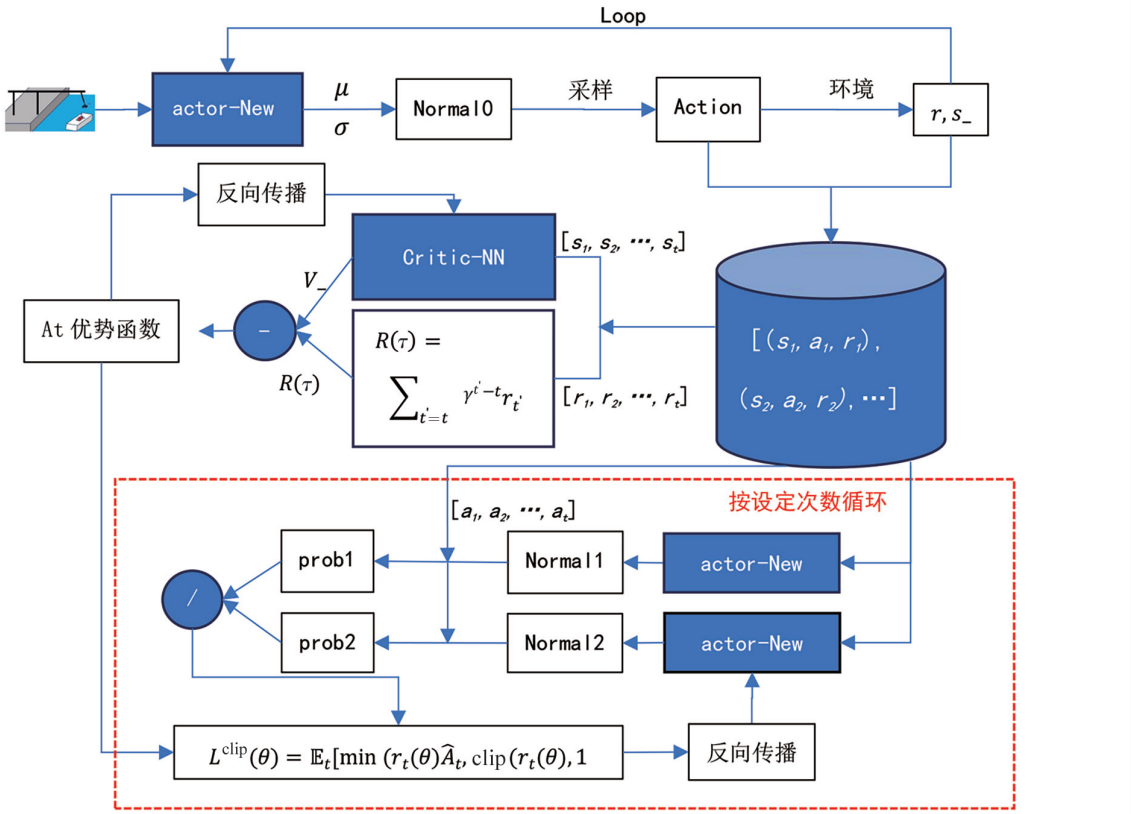


图 5　PPO 算法流程图

Fig.5　PPO algorithm flow chart

2　实验及分析

2.1　实验环境

本实验采用 Python3.6.0 开源编程语言以及 C#语言实现，并且所有实验是在带有 Core i5 3.20 GHz 的 CPU、16GB 内存、500GB 硬盘、Windows10 操作系统的计算机上进行，模拟器为 Unity3D 2021.3.18f1c1 版本。算法实现软件环境有 Python3.6.0、Mlagents-19release 强化学习框架。

模拟环境的详细参数设置如下：

1）场景设置。模拟环境中包含一个桥吊和多个集装箱。桥吊的位置和朝向可以根据实际情况灵活设置。集装箱的数量和位置也可以根据需要进行调整。

2）物理参数。桥吊和集装箱的物理参数包括质量、摩擦系数等。集装箱质量以及吊具质量分别为 1 和 0.5。摩擦系数为默认值。这些参数可以根据实际情况进行调整，以实现尽可能真实的物理模拟。

3）控制参数。桥吊的控制参数包括起升速度、移动速度等。船体晃动利用正弦函数进行模拟。桥吊横移速度设置为 0.3 m/s，吊具上升下降速度设置为 0.1 m/s，这些参数的取值可以根据实际的桥吊操作进行调整。

4）算法参数。实验采用 PPO 算法进行训练，学习率对应于每个梯度下降更新时的强度。Epsilon 对应于梯度下降更新期间新旧策略之间的可接受的差异阈值。batch_ size 对应梯度下降每次迭代中的经验数，这些数据是从经验回放池中抽取的。详细参数如表 1 所示。

2.2　实验结果

在场景设置中，模拟了不同位置的集装箱船和桥吊。桥吊向当前需要卸载的集装箱移动，随着桥

吊和小车不断移动，集装箱的位置和吊具的位置也在变化，同时集装箱的大小也随之变化。

在测试环境中，利用随机种子使集装箱位置以及桥吊位置随机初始化，这样可以充分模拟集装箱船靠港位置的随机性，以及桥吊作业位置的随机性。

在任务刚开始时，集装箱桥吊模型对码头环境的了解较少，没有明确的策略去进行抓取操作，具体表现为桥吊以及小车在小范围抖动。因此在初始阶段模型的性能表现较差，无法成功抓取集装箱。在执行总步数达到 20 万步以后，桥吊模型经过一段时间的探索，逐渐积累经验并更新策略，此时集装箱桥吊模型的性能开始逐渐提升，通过不断与环境交互优化行为策略参数，此时模型开始学会如何正确地装载集装箱，并有效地完成任务。在执行总步数达到 50 万步以后，集装箱桥吊模型已经学习到一个相对稳定的最优策略，此时模型的性能达到一个稳定状态，并且不再有明显的改进。在收敛阶段，集装箱桥吊模型能够高效地完成抓取任务，达到较高的抓取成功率和操作效率。

经过强化学习算法的训练，得到了一组控制策略，可以对桥吊的状态进行判断，并在不同的状态下执行相应的动作。集装箱码头不同程度的训练模型在与人工操作进行对比的实验（见表 2）中可以明显看到，随着训练步数的增加，模型逐渐学会有效的作业行为，并逐渐超越人类的抓取效率与成功率。实验结果表明，这组策略的效果较好，能够在受风浪影响的环境下实现桥吊的高效操作。

表 1　PPO 算法参数表

Tab. 1　PPO algorithm parameters table

参数	说明
学习率	0.0003
Epsilon	0.2
batch_size	1024
隐藏层单元数	256
隐藏层数	2
max_steps	5000

表 2　智能体与人工操作成功抓取 20 个集装箱的数据对比

Tab. 2　Comparison of intelligent body and manual operation successfully grabbing 20 containers

性能	人工操作	智能体			
		10 万步	20 万步	30 万步	100 万步
平均耗时/s	67.7	240	177	67.5	5.05
总耗时/s	1354	4800	3542	1350	101
成功率/%	40	12	20	54	100

算法训练过程中的奖励变化如图 6 所示，横坐标为训练步数，纵坐标为智能体每段训练获得的奖励，可以看出，随着训练步数的增加，累积奖励值逐步增加，在大约 50 万步时，逐渐收敛到 10。这表明智能体已经学习到使累计奖励最大化的行为。

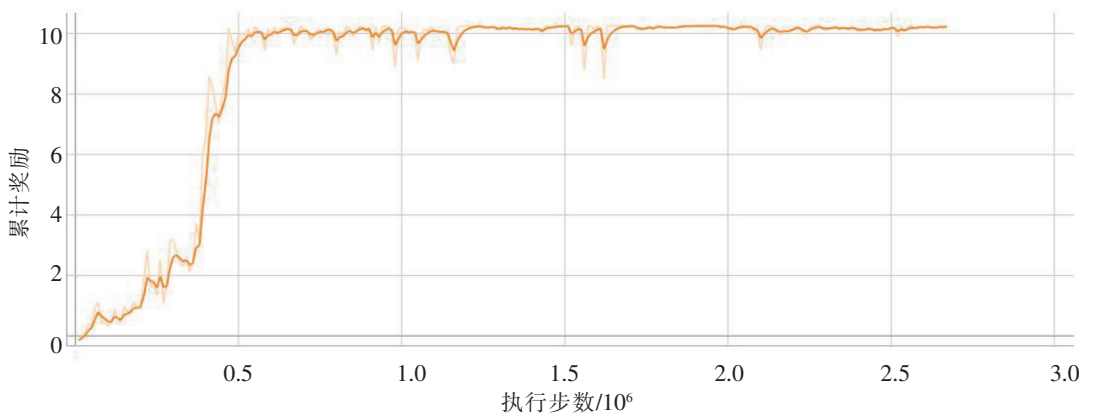


图 6　累计奖励

Fig.6　Cumulative Reward

3　结论

本文构建了码头桥吊装卸作业训练模型及环境，模拟了影响集装箱装卸效率的极端天气情况，通过对强化学习训练，在海浪和大风使集装箱船产生倾斜时，桥吊模型相较于人工作业可提升 60% 的

抓取成功率，作业的平均操作时间、准确度均得到了显著改善，大大缩短了成功抓取集装箱的时间。这表明本文提出的方法可以有效地使桥吊学习在海浪及大风对集装箱船和吊具造成晃动影响时的平稳下落行为，提高桥吊抓取准确性并优化桥吊路径，从而提高整个装箱流程的效率和安全性。

本文的实验结果可用于辅助集装箱码头装卸作业。未来随着集装箱码头数字孪生模型数据的广度及精度的不断提高，有望实现集装箱码头的完全自主作业。

[参 考 文 献]

[1]KIM E H,KWAK K W,KIM Y K,et al. Auto-positioning of sliding planes based on virtual force[J]. Int J Control Autom Syst,2013,11:798-804. DOI:10.1007/s12555-012-0300-1.

[2]KIM D,PARK Y. Tracking control in x-y plane of an offshore containercrane[J]. Journal of Vibration and Control,2017,23(3):469-483. DOI:10.1177/1077546315581091.

[3]JUNG Y,JANG I G,KWAK B M,et al. Advanced sensing system of crane spreader motion(for Mobile Harbor)[J]. 2012 IEEE International Systems Conference. Vancouver,BC,Canada:IEEE,2012:1-5. DOI:10.1109/SysCon.2012.6189443.

[4]LIU Y,WANG Y,LV J,et al. Automatic spreader-container alignment system using infrared structured lights[J]. Applied optics,2012,51(16):3205-3213.

[5]MI C,HUANG S,ZHANG Y,et al. Design and implementation of 3-D measurement method for container handling target[J]. Journal of Marine Science and Engineering,2022,10(12):1961. DOI:10.3390/jmse10121961.

[6]梁晓波,程文明,郭鹏. 集装箱起重机自动装卸系统的研究与设计[J]. 计算机应用,2015,35(S1):229-231,251.

[7]肖智清. 强化学习原理与 Python 实现[M]. 北京:机械工业出版社,2019:139-170.

[8]张伟楠,沈键,俞勇. 动手学强化学习[M]. 北京:人民邮电出版社,2022:3.

[9]VINCENT G,ANUPAM B. ML-agents toolkit overview[EB/OL]. (2018-02-07)[2023-06-09]. https://github.com/Unity-Technologies/ml-agents/blob/release_19/docs/ML-Agents-Overview.md.

[10]JULIANI A,BERGES V,VCKAYE,et al. Unity:a general platform for intelligent agents[J]. ArXiv,abs/1809.02627.

[11]HAARNOJA T,ZHOU A,HARTIKAINEN K,et al. Soft actor-critic algorithms and applications[J]. ArXiv,vol. abs/1812.05905,2018. [Online]. Available:https://arxiv.org/abs/1812.05905.

[12]TORABI F ,WARNEL G,STONE P. Behavioral cloning from observation[C]//Proceedings of the 28th International Joint Conference on Artificial Intelligence,2018:4950-4957. DOI:10.24963/ijcai.2018/687. corpus ID:23206414.

[13]HO J,ERMON S. Generative adversarial imitation learning[C]//Advances in Neural Information Processing Systems. Curran Associates Inc. [2022-01-15]. https://proceedings.neurips.cc/paper/2016/file/cc7e2b878868cbae992d1fb743995d8f-Paper.pdf.

[14]PATHAK D,AGRAWAL P,EFROS A A,et al. Curiosity-driven exploration by self-supervised prediction[J]. JMLR ,2017,70:2778-2787.

[15]ROBERT RAUCH,STEFAN KORECKO,JURAJ GAZDA. Evaluation of proximal policy optimization with extensions in virtual environments of various complexity[C]//32nd International Coference of Radioelektronika(RADIOELEKTRONIKA). Kosice,Slovakia:IEEE,2022:1-5. DOI:10.1109/RADIOELEKTRONIKA54537.2022.9764924.

(责任编辑　朱雪莲　英文审校　周云龙)