面向智能工厂管理中互联网连接的多机器人协同的量子多智能体演员-评论家神经网络

Won Joon Yun，*IEEE* 研究生会员，Jae Pyoung Kim、Soyi Jung、*IEEE* 会员、Jae-Hyun Kim、*IEEE* 会员和 Joongheon Kim，*IEEE* 高级会员

arXiv:2301.04012v1 [quant-ph] 4 Jan 2023

摘要—量子计算是学术界和工业界最新关注的领域之一，已引起广泛关注。在量子计算的诸多主题中，变分量子电路（VQC）因能够实现量子深度强化学习（QRL）而受到关注。本文验证了 QRL 的潜力，并进一步通过基于 QRL 实现量子多智能体强化学习（QMARL），尤其是在智能工厂应用中实现互联网连接的自主多机器人控制与协同，来探索这一潜力。然而，由于经典多智能体强化学习存在非平稳性，这种扩展并非易事。为应对这一问题，本文提出了适用于互联网连接场景的集中训练、分散执行（CTDE）量子多智能体强化学习框架。本文采用一个基于物联网（IoT）、包含多个智能体的智能工厂环境，以展示所提算法的有效性。仿真结果证实，所提的基于量子多智能体强化学习的自主多机器人控制与协同优于其他框架。

索引词—量子深度学习、多智能体强化学习、量子计算、机器人控制、智能工厂

I. 引言

在各种工业4.0场景中，智能工厂系统的自动化和自主化管理如今正受到广泛关注 [2]–[8]。针对工厂管理的自动化，自主多移动机器人的应用已得到广泛研究 [9]–[11]。根据Verizon报告 [12]，，工业*4.0*已在制造业全面展开。预计到*2026*年，全球市场规模将达到*2198*亿美元，而自主移动机器人正成为这一转型中的关键劳动力。要实现高效且有效的自主多机器人控制与协调，基于多智能体强化学习（MARL）的算法至关重要 [13], [14]。

近年来，随着计算硬件和深度学习算法的显著发展，分布式学习和多智能体强化学习（MARL）领域取得了革命性进展 [14]。

本文的初步版本曾发表于7月 2022 [1]在意大利博洛尼亚举行的IEEE分布式计算系统国际会议（ICDCS）。本研究由韩国国家研究基金会资助（2022R1A2C2004869，2021R1A4A1030775）。（通讯作者：*Soyi Jung*、*Jae-Hyun Kim*、*Joongheon Kim*）。Won Joon Yun、Jae Pyoung Kim和Joongheon Kim任职于大韩民国首尔02841高丽大学电气工程学院（电子邮箱：{ywjoon95,paulkim436,joongheon}@korea.ac.kr）。Soyi Jung和Jae-Hyun Kim任职于大韩民国水原16499亚洲大学电气与计算机工程系（电子邮箱：{sjung,jkim}@ajou.ac.kr）。

此外，量子计算硬件和算法的发展进一步凸显了这一趋势 [15],，推动了量子机器学习研究的发展。如今，与传统机器学习相比，量子机器学习仍处于起步阶段。例如，在量子机器学习的分类任务中，使用 MNIST 数据集时，其在量子计算机上的 Top-1 准确率仅为 32.5% [16]，而在理想量子机器上为 74.2% [17] 。然而，理论上发现的优势（即量子霸权）最近正得到实验验证 [18]–[20]。量子算法的潜力显而易见：它能够利用量子纠缠，在保持准确率的同时减少模型参数 [21]。此外， [22] 的实证结果表明，量子机器学习的表现优于经典机器学习。一个杰出的例子是变分量子电路（VQC）架构，也称为量子神经网络（QNN） [23], [24]。QNN 是一种能够复现经典深度神经网络功能的量子电路。将 QNN 与经典深度学习模型结合，可构建量子-经典混合模型，从而实现量子强化学习（QRL）。与强化学习（RL）相比，QRL 使用的模型参数更少，同时显著缩短训练和推理时间 [25], [26] ，并且消耗的计算资源也更少 [27]。因此，显而易见，利用量子计算进行量子机器学习将在不久的将来成为一大发展趋势。本文旨在将变分量子电路与经典多智能体强化学习相结合，将量子强化学习拓展至量子多智能体强化学习（QMARL）。

多智能体强化学习环境中的智能体通过合作或竞争的方式相互作用。这种交互基于物联网连接技术实现。这些交互会使每个智能体的奖励变得非平稳，从而阻碍多智能体强化学习训练的收敛。集中式训练与分布式执行（CTDE）方法用于 [28]应对多智能体强化学习模型的非平稳性。在这种场景下，通过拼接所有智能体的状态-动作对，将奖励同时分配给所有智能体。如 [1]所示，可以直接实现变分量子电路（VQC）版本的 CTDE。然而，由于通过变分量子电路进行量子强化学习时，状态-动作对由量子比特表示，这种实现会导致量子比特数量随智能体数量增加而增加。因此，量子误差也会随着量子比特数量的增加而 [29]，显著增多

影响多智能体强化学习的收敛性和可扩展性。此外，在当前含噪中等规模量子（NISQ）时代，量子纠错尚不可行。

本文旨在改进现有的多种变分量子电路、智能体策略和状态编码实现方法 [25], [30], [31]。与以往研究相比，我们所提变分量子电路有三项显著差异：参数共享、非随机变分量子电路设计和双变量密集编码。首先，参数共享是指智能体之间共享模型参数值。以往研究中的所有智能体都各自拥有不同的策略，这意味着智能体越多，所需的策略也越多，制定这些策略的过程会消耗额外的计算资源。在这一改进模型中，所有智能体将共享同一个策略，从而提高计算资源的使用效率。第二项改进是非随机变分量子电路设计。以往研究所用的变分量子电路由随机选取的量子门构成。尽管其性能出色，但由于存在随机性，难以轻易复现。由于量子门具有随机性，同一模型在另一次迭代中可能无法呈现相同的性能。本文通过设计固定模型，消除了以往变分量子电路的随机性，从而加以改进。这确保了模型的可复现性和稳定性。最后，本文所提模型采用双变量密集编码方法，而非四变量密集编码方法。原有编码方法能够降低给定数据的维度。尽管这或许适用于NISQ时代量子电路，但不可避免地会造成信息损失。所提双变量密集编码方法不会降低数据维度，同时仍兼容NISQ时代量子电路。因此，它能够避免信息损失，进而提升模型性能。

贡献。 本研究的主要贡献总结如下。

- 本文首先为智能工厂应用中的自主多机器人控制与协同提供了一种基于量子的多智能体强化学习解决方案。
- 此外，本文提出了一种改进且新颖的集中训练、分散执行量子多智能体强化学习框架，该框架在策略、变分量子电路设计和双变量密集编码方面采用了参数共享。
- 最后，通过大量实验，在智能工厂场景中进行仿真，证明所提量子多智能体强化学习框架优于经典多智能体强化学习模型。结果表明，所提模型的性能高于其他模型。

组织结构。 本文其余部分的组织结构如下。本文的预备知识将在第 II 节中介绍。面向智能制造的自主移动机器人协同将在第 III 节中介绍。第IV 节介绍所提方案的算法；所提方案的数值结果和演示将在第 V 节中展示。第 VI节总结全文并展望未来工作。需要注意的是，本文中的符号列于表I中。本文使用的大多数方程和符号均基于 狄拉克记号，具体参见 [32]。

表 I：符号列表

场景符号	
N	自主移动机器人智能体的数量
M	站点/仓库的数量
T	回合长度
o^n	第 n 个自主移动机器人智能体的观测
a^n	第 n 个自主移动机器人智能体的动作
$\mathbf{a}$	自主移动机器人智能体的动作集合，即 $\mathbf{a} = \{a_n\}_{n=1}^{N}$。
s	真实状态
$q^c_{m,t}$	时间 t第 m个仓库的负载状态
$q^e_{n,t}$	时间 t第 n个AMR智能体的负载状态
$q^e_{\max}$	仓库的承载能力
$q^e_{\max}$	AMR智能体的承载能力
量子计算符号	
$\|\psi\rangle$	纠缠量子态
$\langle O\rangle$	可观测量
Γ	Pauli-Γ 量子门，例如 $\Gamma \in \{X, Y, Z\}$
R_Γ	旋转 Γ 量子门，例如 $\Gamma \in \{X, Y, Z\}$
$(\cdot)^\dagger$	复共轭算子
$\mathcal{M}$	测量算子

II. 量子计算预备知识

单量子比特量子态。 QC 使用量子比特 作为基本计算单位。量子比特表示两个基态之间的量子叠加态，记为 $|0\rangle$和 $|1\rangle$。描述量子比特态有两种方式，

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \tag{1}$$

其中 $\|\alpha\|_2^2 + \|\beta\|_2^2 = 1$，以及，

$$|\psi\rangle = \cos\left(\frac{\delta}{2}\right)|0\rangle + e^{i\varphi}\sin\left(\frac{\delta}{2}\right)|1\rangle, \tag{2}$$

其中 $\delta \in [-\pi, \pi]$ 和 $\varphi \in [-\pi, \pi]$。前者基于归一化的二维复向量，后者则从几何视角基于极坐标（δ， φ）。量子比特态映射到三维单位球（布洛赫球）的表面。此外，量子门是一种酉算子，可将一个量子比特态变换为另一个量子比特态，它表示为一个含复数元素的 2×2 矩阵。单量子比特泡利量子门 X、 Y 和 Z 定义如下，

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{3}$$

还有一些常用的量子门，即 R_{x}、 R_{y} 和 R_{z}。这些旋转算子门可使单个量子比特绕布洛赫球中对应的轴旋转 δ，单量子比特操作可表示为以下方程，

$$R_X(\delta) = e^{-i\frac{\delta}{2}X}, \quad R_Y(\delta) = e^{-i\frac{\delta}{2}Y}, \quad R_Z(\delta) = e^{-i\frac{\delta}{2}Z}, \tag{4}$$

其中旋转角记为 $\delta \in \mathbb{R}[0, 2\pi]$。这些基本的泡利门和旋转门都是酉矩阵， $U^\dagger U = I$，其中I 表示单位矩阵。

多量子比特量子态。 多量子比特系统借助量子叠加实现超高速量子计算。著名的量子算法（例如肖尔算法[33] 和 Grover搜索 [34]）都以多量子比特系统为基础。含有 q 个量子比特的量子态记为 $|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_q\rangle = \sum_{n=0}^{2^q-1} \alpha_n |n\rangle$，其中 $\otimes$，且 α_n 表示叠加算符（即张量积），而 $|n\rangle$ 表示第 n - 个概率幅， n 表示 个量子比特量子态的第 - 个基 q。请注意，概率幅的模平方之和等于 1，即， $\sum_{n=0}^{2^q-1} |\alpha_n|^2 = 1$ [32]。为了实现量子叠加，可以使用作用于多个量子比特的量子门，称为受控旋转门。它们根据多个控制量子比特的信号对一个量子比特进行操作，从而在这些量子比特之间产生量子纠缠。其中，受控-X（或受控非门）量子门 $CX = \begin{bmatrix} I & 0 \\ 0 & X \end{bmatrix}$ 是广泛使用的控制门之一。这些多量子比特量子门使量子算法能够在变分量子电路上利用其特性，最终用于量子多智能体强化学习。

III. 智能制造中的自主移动机器人协同

A. 自主移动机器人系统设计

自动导引车（AGV）是一种便携式机器人，可沿着标记在地面上的线或电线行驶，也可利用无线电波、视觉摄像头、磁铁或激光进行导航。AGV广泛应用于工业领域，用于在工厂和仓库等大型工业设施内运输重型物料。因此，AGV对于智能工厂管理至关重要，这一点显而易见。自主移动机器人（AMR）与AGV不同，它配备了各种传感器，能够通过检测周围的静态和动态物体来实现自主定位识别和搜索。其路径根据静态和动态障碍物实时生成，使AMR无需预设路径即可自由行驶。尽管该系统更加灵活，实时生成路径也带来了额外挑战，车队管理系统（FMS）必须应对诸如执行运输订单配送、车辆路径规划和任务执行调度等活动。需要注意的是，自主移动机器人之间紧密耦合，导致计算复杂度较高。例如，即使自主移动机器人的数量和转运订单数量相对较少，在考虑所有可能的自主移动机器人路径时，其性能仍会受到影响。因此，集中式自主移动机器人车队管理和订单执行优化通常无法实时完成。因此，多智能体强化学习算法得到了广泛关注和研究 [35]。为进一步提升性能，还可以采用量子多智能体强化学习（QMARL），本文将对此进行讨论。

B. 配备多个自主移动机器人的自动化液晶显示器智能工厂

得益于量子计算的特性，研究表明，与经典超级计算机相比，量子计算可将能耗降低多个数量级 [36]。关于量子强化学习，近期研究显示其实现了量子霸权 [22]。本文研究一种采用基于直流的自主移动机器人的液晶显示器（LCD）智能工厂系统。如图 1所示，彩色薄膜晶体管（TFT）液晶显示面板由两块玻璃基板组成：一块薄膜晶体管阵列基板和一块彩色滤光片基板。薄膜晶体管液晶显示面板通过五道工序组合制造而成：薄膜晶体管阵列滤光工序、彩色滤光片工序、修复工序、成盒工序和模组组装工序。前两道工序（即薄膜晶体管阵列滤光工序和彩色滤光片工序）在 A站点进行，两块基板由自主移动机器人运送，其余工序则在 B站点进行。每台自主移动机器人在从提供服务到服务于灵活区域的状态转移过程中发挥作用，而这需要根据服务区域划分出的多个区域作出决策。在薄膜晶体管液晶显示器的制造过程中，可能会产生各种不良液晶显示器，这些产品不应投入使用。因此，为防止使用此类不良液晶显示器，自主移动机器人必须识别不良产品，并请求对液晶显示器进行质量验证。用于检测液晶显示器缺陷的技术已开发出来，并已在智能工厂中投入应用 [37]–[39]。通过采用上述研究提出的 精确率 参数，自主移动机器人将识别不良液晶显示器，并将其卸载到专门用于存放缺陷产品的另一个收集点。

本文假设所有自主移动机器人都配备了最优的轨迹规划器和充电调度器，例如 [40], [41]。因此，所提量子多智能体强化学习模型必须规划每台自主移动机器人的轨迹，以分拣出不良液晶显示器，同时妥善卸载正常产品。在通信方面，QDL 服务器通过有线网络连接到每个站点，而每个站点则通过无线方式与自主移动机器人连接。由于数据包大小较小，且液晶显示器智能工厂中的发射功率足够，因此我们假设丢包可以忽略不计。QDL 服务器接收自主移动机器人的观测，重新配置状态，最后向自主移动机器人发送动作决策。为了实现柔性制造，应妥善规划自主移动机器人，以装载货物、卸载货物并控制液晶显示器的质量。此外，调度和派遣这些资源的决策过程对于优化资源利用率和提升自主移动机器人的生产效率至关重要。

问题定义与表述。 在这种情况下，量子深度学习（QDL）服务器基于集中训练、分散执行的量子多智能体强化学习框架这一基本概念，支持物料搬运系统的高效调度决策过程。具体而言，QDL服务器为每台自主移动机器人做出分布式、序贯决策，以确定其货物任务（即每台自主移动机器人承载的货物数量以及是否请求质量控制），从而避免每台自主移动机器人在货物配送过程中出现溢出与欠载。

IV. 用于自主多机器人协同的量子多智能体演员-评论家网络

A. 基本马尔可夫决策过程形式化

我们所考虑的智能工厂环境中的自主多机器人协同由 M 个站点和 N 个自主移动机器人智能体组成。该智能工厂环境通过部分可观测马尔可夫决策过程进行数学建模（参见第 II节）。下文将基于第 m 个站点、第 n 个智能体以及时间步 t进行说明。



图1：系统模型：利用量子多智能体强化学习进行液晶显示面板制造

1）负载动态：每个站点都有一个仓库，其负载能力为 $c^{\mathrm{W}}_{\max}$。此外，AMR智能体的负载能力低于其最大容量 $c^{\mathrm{A}}_{\max}$。AMR智能体从其他自主移动机器人接收货物（例如液晶显示面板或薄膜晶体管）。本文将负载重量记为 $b^{\mathrm{A},n}_t$。负载重量服从均匀分布 $\forall b^{\mathrm{A},n}_t \sim \mathcal{U}(0, w_{load} \cdot b_{\max})$。仓库和AMR智能体具有装载状态 $c^{\mathrm{W},m}_t$ 和 $c^{\mathrm{A},n}_t$，表示当前装载的货物。所有AMR智能体都将货物运送至仓库。其动态如下：

$$c^{\xi,n}_{t+1} = clip(c^{\xi,n}_t - a^{\xi,n}_t + b^{\xi,n}_t, 0, c^{\xi}_{\max}), \tag{5}$$

其中，$\xi \in \{\mathrm{W}, \mathrm{A}\}$ 表示仓库和一个AMR智能体。术语 $a^{\xi,k}_t$ 和 $b^{\xi,n}_t$ 分别表示第 m个仓库或第n个AMR智能体的累计交付货物重量和接收货物重量。请注意，$a^{A,n}_t$ 是第 n个AMR智能体的动作。此外，裁剪函数定义为 $clip(x, x_{\min}, x_{\max}) \triangleq \min(x_{\max}, \max(x, x_{\min}))$。

2）质量控制：我们假设负载已由先前的缺陷检测过程完成分类。在缺陷检测过程中，负载被分为四种类型（即真正例、假正例、假负例和真负例），据此计算统计数据（即精确率、召回率和F分数）。本文将负载状态视为质量统计数据（例如精确率）。其中，质量统计数据会提供给自主移动机器人智能体。自主移动机器人智能体可以据此作出动作决策，重新请求对负载进行质量控制。如果负载被请求进行质量控制，质量工程师将对其进行质量验证。我们假设质量工程师能够完美地检测出负载上的所有缺陷。然而，由质量工程师执行的质量再确认流程还会额外产生τ_{qual} 时间延迟。

3) 效用设计：我们针对质量、时间延迟和负载均衡设计效用。首先，自主移动机器人智能体接收类型为真正例（*TP*）和假正例（*FP*）的货物。该指标，即精确率，可表示阳性预测值的比率，写作如下：

$$u^{q,n}_t = \frac{TP^n_t}{TP^n_t + FP^n_t}, \tag{6}$$

其中，$TP^n_t = \sum_{\in l^n_t} \mathbb{1}_{\text{负载}} = \sum_{=} \mathbb{1}_{\text{负载}\in l^n_t}$（负载 *TP*）和*FP*（负载 *FP*）分别表示第 n个自主移动机器人的真正例和假正例。请注意，l^n_t 表示第 n个自主移动机器人的整体负载缺陷状态。对于延迟，我们测量处理时间。因此，第 n个AMR智能体的延迟效用表示如下：

$$u^{d,n}_t = -\left(1 + \tau_{qual} \cdot q^n_t\right), \tag{7}$$

其中，q^n_t 表示质量控制动作。如果 $q^n_t = 1$，则将负载传送给质量工程师；否则，将负载传送到其他站点。最后，负载均衡旨在尽可能减少溢出负载总量以及负载为空的情况。因此，负载均衡的效用如下所示：

$$u^{b,n}_t = \mathbb{1}_{(c^{A,n}_{t+1}=0)} \cdot \tilde{c}^{A,n}_t + \mathbb{1}_{(c^{A,n}_{t+1}=c^{A}_{\max})} \cdot \hat{c}^{W,n}_t \tag{8}$$

$$u^{W,m}_t = \mathbb{1}_{(c^{W,m}_{t+1}=0)} \cdot \tilde{c}^{W,m}_t + \mathbb{1}_{(c^{W,m}_{t+1}=c^{W}_{\max})} \cdot \hat{c}^{W,m}_t \tag{9}$$

其中，$\tilde{c}^{W,m}_t = |c^{W,n}_t - a^{W,n}_t + b^{W,n}_t|$和 $\hat{c}^{W,n}_t = |c_{\max} - \tilde{c}^{W,n}_t|$。请注意，由于本文将异常装载状态（例如负载溢出或下溢）的发生视为负效用，因此 $r(s_t, \mathbf{a}_t) \in [-\infty, 0]$（负）。目标是最大化总精确率，并最小化总延迟以及溢出或下溢事件。

B. 部分可观测马尔可夫决策过程设置

本小节介绍 POMDP 的形式化定义，即智能体间存在不确定性的

随机决策模型 [42]；我们提出的 QMARL 基于 POMDP 这一基本概念进行数学建模。注意，POMDP 定义为一个元组 $\langle \mathcal{N}, \mathcal{S}, \mathcal{A}, P, r, \mathcal{Z}, O, \rho, \gamma, T \rangle$。状态集和观测集分别表示为 $\mathcal{S}$ 和 $\mathcal{Z}$。 $\mathcal{N} := \{1, \cdots, N\}$ 和 $s \in \mathcal{S}$ 分别表示 N 智能体集合以及环境的当前状态。初始状态 $s_0 \sim \rho$ 服从分布 ρ。第 n 个智能体 $a^n \in \mathcal{A}$ 的动作可以是离散动作或连续动作，联合动作用 $\mathbf{a} := \{a^n\}_{n=1}^N$ 表示。状态转移由概率函数 $P(s'|s, \mathbf{a}) : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathcal{S}$ 决定，其中 s' 表示下一个状态。所有智能体都获得共享奖励 $r_t = r(s_t, \mathbf{a}_t) : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$。在 Dec-POMDP 中，真实状态 s 不会直接提供给智能体。每个智能体 $n \in \mathcal{N}$ 通过观测函数 $O(s, a) : \mathcal{S} \times \mathcal{A} \to \mathcal{Z}$ 获得观测 $z^n \in \mathcal{Z}$。我们假设所有智能体采用参数共享策略，记为 π_θ。因此，策略 π_θ 接收第 n 个智能体的观测 $z_t^n \in \mathcal{Z}$，并确定第 n 个智能体的动作，表示为 $\pi_\theta(a|z^n) : \mathcal{Z} \times \mathcal{A} \to [0, 1]$。POMDP 的目标是获得最优策略 $\pi_\theta^* = \arg\max_{\pi_\theta} \mathbb{E}_{\mathbf{a} \sim \pi_\theta}[\sum_{t=1}^T \gamma^{t-1} \cdot r_t]$，其中 $\gamma \in \mathbb{R}[0, 1]$ 和 T 分别表示折扣因子和有限时间。基于此定义，我们将 POMDP 设计如下：

1) 观测：每个自主移动机器人智能体只能部分获取其观测。由于采用参数共享策略 π_θ，观测中包含二值指示向量 $\mathbf{n}_2$。此外，第 n 个自主移动机器人智能体根据其装载状态 $c^{A,n}_t$ 以及仓库当前的装载状态 $\{c_t^{W,m}\}_{m=1}^M$ 来决定动作。综上，第 n 个智能体的观测定义为 $z_t^n \triangleq \{\mathbf{n}_2, c^{A,n}_t\} \cup \{c_t^{W,m}\}_{m=1}^M$。

2) 状态：设计了一个状态信息变量，用于包含所有自主移动机器人智能体和仓库的装载状态信息。时间 t 的状态变量为 $s_t = \{c_t^{A,n}, u_t^{d,n}\}_{n=1}^N \cup \{c_t^{W,m}\}_{m=1}^M$。请注意，状态信息用作量子评论家网络的输入。

3) 动作：假设自主移动机器人智能体可以选择将货物运送至哪个仓库，其中目的地空间定义为 $\mathcal{I} \triangleq \{1, \cdots, M\}$。此外，自主移动机器人智能体可以确定运送至仓库的数量。运送数量和质量控制空间分别定义为 $\mathcal{P} \triangleq \{p_{\min}, \cdots, p_{\max}\}$ 和 $\mathcal{Q} \triangleq \{0, 1\}$。最后，第 n 个自主移动机器人智能体的动作及其动作空间定义为 $a_t^n := (i_t^n, p_t^n, q_t^n) \in \mathcal{A} \equiv \mathcal{I} \times \mathcal{P} \times \mathcal{Q}$。

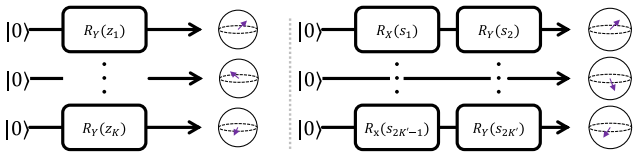
4) 奖励：部分可观测马尔可夫决策过程的目标是尽量减少溢出负载总量以及负载发生.....的事件
为空。因此，奖励 $r(s_t, \mathbf{a_t})$ 定义如下，

$$r(s_t, \mathbf{a}_t) = \sum_{n=1}^{N} (u_t^{q,t} + w_d \cdot u_t^{d,n} + w_b \cdot u_t^{b,n}) + w_W \cdot \sum_{m=1}^{M} u_t^{W,m}, \tag{10}$$

其中，w_d、w_b 和 w_W 分别表示时间延迟、负载均衡以及自主移动机器人智能体和站点的奖励系数。

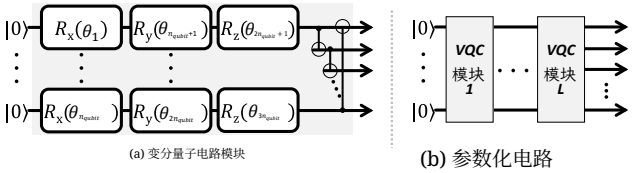
C. 量子多智能体演员-评论家网络设计

1) 状态编码电路：状态编码电路用于对状态输入进行前馈。图 2 展示了



(a) 量子行动者中的状态编码器　(b) 量子评论家中的状态编码器

图 2：状态编码器示意图。



(a) 变分量子电路模块　(b) 参数化电路

图 3：参数化电路示意图。

状态编码器的两种方案。图 2(a)/(b) 分别需要每个量子比特使用一个或两个量子门。尽管当量子比特数等于输入变量数时，编码系统的性能最佳，但强化学习中的输入变量数（即状态）必须大于量子比特数 [26]。因此，本文考虑环境因素，采用以下两种状态编码器：

$$U_{enc}^a(z) = \left[\otimes_{k=1}^K (R_Y(x_k^z)) \right] \cdot |0\rangle^{\otimes q_{actor}}, \tag{11}$$

$$U_{enc}^c(s) = \left[\otimes_{k'=1}^{K'} (R_Y(x_{2k'}^s) \cdot R_X(x_{2k'-1}^s)) \right] \cdot |0\rangle^{\otimes q_{critic}}, \tag{12}$$

其中，x_k^z 和 $x_{k'}^s$ 分别表示观测 z 的第 k 项和状态 s 的第 k' 项。注意，$U_{enc}^a(z)$ 和 $U^c{}_{enc}(s)$ 分别表示行动者观测编码器和评论家状态编码器。行动者观测编码器和评论家状态编码器分别运行于 K 量子比特和 K' 量子比特系统中。

2) 参数化电路和量子测量：参数化电路是一种使用可学习参数执行估计、优化、近似和分类等数值任务的量子电路。如图3（a）所示，变分量子电路模块由方向不同的旋转门和受控-Z 量子门组成，即 R_X、R_Y、R_Z 以及 $CZ = \begin{bmatrix} I & 0 \\ 0 & Z \end{bmatrix}$。注意，$CZ$ 用于纠缠量子比特。为提高电路性能，本文采用包含多个变分量子电路模块的参数化电路，这需要额外的可训练参数 θ，如图3所示。为获得理想的输出，采用测量 $\mathcal{M}$，该测量基于其计算基计算叠加量子态的期望值。总之，可观测量（即期望值）可写为：

$$\langle O \rangle_{x,\theta} = \prod_{M \in \mathcal{M}} \langle 0 | U_{enc}^\dagger(x) U_{VQC}^\dagger(\theta) M U_{VQC}(\theta) U_{enc}(x) | 0 \rangle, \tag{13}$$

其中，$\langle O \rangle_{x,\theta}$ 是以 x 为输入、以 θ 为电路参数的VQC输出；$\mathcal{M}$ 是VQC中的量子测量基集合，其中 $|\mathcal{M}| \leq n_{qubit}$。

3）量子行动者-评论家的实现：本文针对智能工厂提出的QMARL采用去中心化架构，以实现可扩展性。QMARL中的每个自主移动机器人智能体均采用基于VQC的策略，即智能体之间无需通信。演员/评论家的可观测量如下：

$$\langle O_a\rangle_{o,\theta}=\left\{\langle 0|U^{a\dagger}_{enc}(o)U^{a\dagger}_{VQC}(\theta)MU^{a}_{VQC}(\theta)U^{a}_{enc}(o)|0\rangle\right\}_{M\in\mathcal{M}_a}, \tag{14}$$

$$\langle O_c\rangle_{s,\phi}=\left\{\langle 0|U^{c\dagger}_{enc}(s)U^{c\dagger}_{VQC}(\phi)MU^{c}_{VQC}(\phi)U^{c}_{enc}(s)|0\rangle\right\}_{M\in\mathcal{M}_c}. \tag{15}$$

量子行动者。对于量子行动者，使用（14）的可观测量来计算每个自主移动机器人智能体的动作概率。随后，通过其可观测量的 softmax函数写出量子策略：

$$\pi_\theta(a_t|z_t)=softmax(\beta_a\langle O_a\rangle_{z_t^n,\theta}), \tag{16}$$

其中

$$softmax(\mathbf{x})\triangleq\left[\frac{e^{x_1}}{\sum_{i=1}^{N}e^{x_i}};\cdots;\frac{e^{x_N}}{\sum_{i=1}^{N}e^{x_i}}\right] \tag{17}$$

并且 β_a 分别是行动者可观测量的缩放因子。在时间 t，第 n个智能体的行动者策略根据给定的观测 o_t^n进行动作决策，记为 $\pi_\theta(a_t^n|o_t^n)$。注意， θ 表示第 n个行动者的参数。随后，动作 a_t^n 按如下方式计算：

$$a_t^n=\arg\max_a\pi_\theta(a|o_t^n), \tag{18}$$

并且请注意，所有智能体通过参数共享使用相同的策略。

量子集中式评论家。采用集中式评论家作为集中训练、分散执行中的状态价值函数。在时间 t，参数化的评论家根据 a_t估计折扣回报，如下所示：

$$V_\phi(s_t)=\beta_c\langle O_c\rangle_{s_t,\phi}\simeq\mathbb{E}[\sum_{t'=t}^{T}\gamma^{t'-t}\cdot r(s_{t'},\mathbf{u}_{t'})|s_t=s], \tag{19}$$

其中， γ、 T、$\mathbf{a}_t$、 β_c和 $r(s_{t'}, \mathbf{a}_{t'})$分别表示折扣因子 $\gamma\in[0,1)$、回合长度、所有智能体的动作、评论家可观测量的缩放因子，以及给定状态 $s_{t'}$ 和动作$\mathbf{a}'_t$ 时的奖励函数。此外， ϕ表示评论家的可训练参数。这里， s_t 是在 t时刻的真实状态。

D. 训练算法

多智能体强化学习智能体的目标是最大化折扣回报。为推导最大化目标的梯度，我们利用联合状态价值函数 V_ϕ。为训练 V_ϕ，本文采用多智能体策略梯度（多智能体策略梯度），其形式如下：

$$\nabla_\theta\mathcal{L}_{actor}=-\mathbb{E}_{\mathbf{a}\sim\pi_\theta}\left[\sum_{t=1}^{T}\sum_{n=1}^{N}y_t\nabla_\theta\log\pi_\theta(a_t^n|z_t^n)\right], \tag{20}$$

$$\nabla_\phi\mathcal{L}_{critic}=\nabla_\phi\sum_{t=1}^{T}\|y_t\|^2, \tag{21}$$

算法1： 量子多自主移动机器人智能体训练

初始化评论家网络和演员网络的重量 θ 和 ϕ 以及经验回放缓冲区 $\mathcal{D}=\{\}$;
将目标网络初始化为： $\phi^{\intercal}\leftarrow\phi$;
对于回合 = *1*, *MaxEpisode***执行**
  ▷ 初始化**智能工厂环境**;
  $t=0$;
  s_0 =初始状态;
  当 $s_t\neq terminal$ 且 $t<$回合上限 **时**
    对于每个智能体 n **执行**
      计算 $\pi_\theta(a|o_t^n)$ 并采样 a_t^n;
    end
    获取奖励 r_t 、下一个状态和观测 s_{t+1}， $\mathbf{o}_{t+1}=\{o_t^n\}_{n=1}^N$;
    $\mathcal{D}=\mathcal{D}\cup\{(s_t,\mathbf{o}_t,\mathbf{a}_t,r_t,s_{t+1},\mathbf{o}^{t+1})\}$;
    $t=t+1$;
  end
  对于每个回合中的每个时间步 t，在批次 $\mathcal{D}$ **中执行**
    获取 $V_\phi(s_t)$; $V_{\phi^{\intercal}}(s_{t+1})$;
    计算目标 y_t 使用 (22);
  end
  计算 $\nabla_\theta\mathcal{L}_{actor}$, $\nabla_\phi\mathcal{L}_{critic}$，并更新 θ, ϕ;
  如果达到目标更新周期**则**
    更新目标网络， $\phi^{\intercal}\leftarrow\phi$
  end
结束

表II：基准方案。

方案	计算方法	参数数量
所提方案	量子	≈ 110
Comp1	量子/经典	≈ 110
Comp2	经典	≈ 110
Comp3	经典	≈ 40K
对比方法	随机游走	None

4

满足以下约束

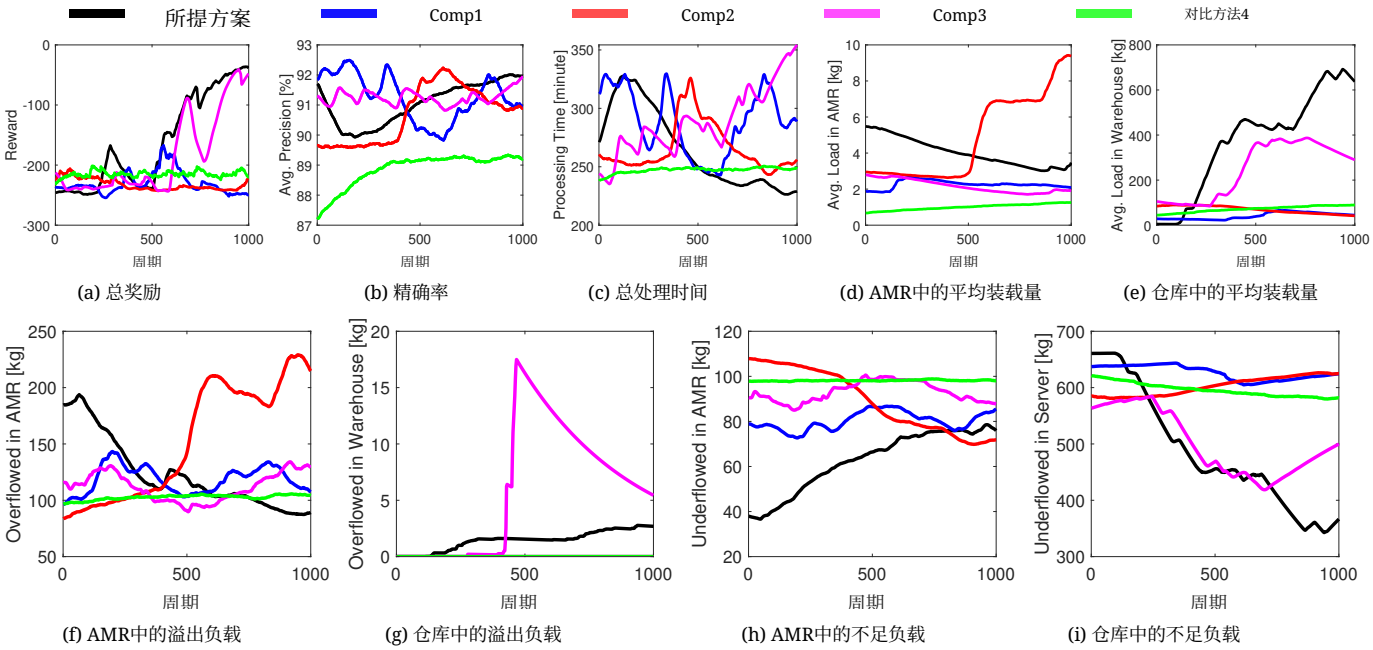
$$y_t=r(s_t,\mathbf{a}_t)+\gamma V_{\phi^{\intercal}}(s_{t+1})-V_\phi(s_t), \tag{22}$$

其中 $\phi^{\intercal}$ 是目标评论家网络的参数。请注意，（20）和（21）用于遵循参数移位规则 [43],写成如下形式：

$$\frac{\partial\mathcal{L}_{actor}}{\partial\theta_i}=\frac{\partial\mathcal{L}_{actor}}{\partial\pi_\theta}\cdot\frac{\partial\pi_\theta}{\partial\langle O\rangle_{o,\theta}}\cdot\left[\langle O\rangle_{o,\theta+\frac{\pi}{2}\mathbf{e}_i}-\langle O\rangle_{o,\theta-\frac{\pi}{2}\mathbf{e}_i}\right], \tag{23}$$

$$\frac{\partial\mathcal{L}_{critic}}{\partial\phi_j}=\frac{\partial\mathcal{L}_{critic}}{\partial V_\phi}\cdot\frac{\partial V_\phi}{\partial\langle O\rangle_{s,\phi}}\cdot\left[\langle O\rangle_{s,\phi+\frac{\pi}{2}\mathbf{e}_j}-\langle O\rangle_{o,\phi-\frac{\pi}{2}\mathbf{e}_j}\right], \tag{24}$$

其中 $\mathbf{e}_i$ 和 $\mathbf{e}j$ 分别表示参数化向量 i的第 j个和第 θ 个标准基 ϕ。需要注意的是，前两个偏导数由经典计算得出，最后一项则通过量子计算得到。详细的训练过程见**算法1**。



(a) 总奖励 (b) 精确率 (c) 总处理时间 (d) AMR中的平均装载量 (e) 仓库中的平均装载量

(f) AMR中的溢出负载 (g) 仓库中的溢出负载 (h) AMR中的不足负载 (i) 仓库中的不足负载

图4：比较不同多智能体强化学习框架的各项指标实验结果。

表III：实验参数。

参数	值
站点数量（M）	2
自主移动机器人数量（N）	6
仓库的承载能力	2，000 千克
AMR智能体的承载能力	500 千克
观测维度	6
精确率（报告值 [39]）	$\{61.9, 95.8, 97.1\}$%
薄膜晶体管液晶显示器的重量（报告值 [44]）	6 kg
动作维度	5
状态维度	8
回合长度	30 个时间步
奖励系数（w_d，w_b，w_W）	(0.1, 1, 10)
质量工程师的时间延迟（τ_{qual}）	3 个时间步
演员可观测超参数 β_a	3
评论家可观测超参数 β_c	35
优化器	Adam 优化器
U_p^a和 U_p^c中的量子门数量	54
演员的量子比特数量	8
评论家的量子比特数量	8
演员的学习率	1×10^{-2}
评论家的学习率	1×10^{-3}
权重衰减	1×10^{-5}

V. 性能评估

A. 实验设置

为验证所提量子多智能体强化学习框架（命名为 所提方案）在智能工厂管理中的有效性，将基于所提量子多智能体强化学习的算法与表 II中列出的四种对比方法进行比较。这些数值实验的目的如下：

- 为验证量子优势，本文对所提方案、Comp1 和 Comp2 进行了对比实验。为确保公平比较，各方法的参数数量设置相同。
- 本文将所提方案与 Comp3 进行比较，以验证所提方法能否取得优于最新多智能体强化学习技术的性能。
- 为验证多智能体强化学习的优越性，本文将多智能体强化学习方案与随机游走方案（即对比方法4）进行比较。
- 为研究质量控制的鲁棒性，我们在不同精确率的环境中训练基准方案。我们验证了环境中的负载质量会随时间变化这一事实，并证实了所提方案中质量控制的鲁棒性。

仿真参数设置列于表III中。由于本文使用的量子比特数量少于110，本文假设量子噪声可以忽略。Comp1 是一种采用优势演员-评论家评论家结构的量子-经典混合方法，该方法在另一项研究中提出并开发 [45]。此外，Comp2和Comp3 基于集中训练、分散执行结构。具体而言，价值分解网络（VDN） [46]。为进行公平比较，我们将神经网络构建为线性运算和激活函数（即线性层或全连接层）。我们使用 Python 软件库（`torchquantum` 和 `pytorch`）来部署变分量子电路和深度学习方法，并支持 GPU 加速 [16]。此外，所有实验均在多GPU平台上进行训练和推理/测试；该平台配备了两块 NVIDIA Titan XP GPU，主频为 1405 MHz，内存为 12 GB。

B. 训练性能

图 4 展示了与训练指标对应的数值结果。本文采用总奖励、精确率、

表 IV：不同基准方案对应的各项指标的数值结果。

指标 [(a)–(f) 的国际单位制单位：kg]	基准方案 所提方案	Comp1	Comp2	Comp3	对比方法4
(a) 自主移动机器人平均负载状态	6.0	2.9	9.3	2.9	2.6
(b) 服务器平均负载状态	511	88	87	244	87
(c) 自主移动机器人中的平均溢出负载	**81**	101	224	131	103
(d) 服务器中的平均溢出负载	**2.6**	0	0	5.1	0
(e) AMR中的平均不足负载	**77**	106	227	136	100
(f) 服务器中的平均下溢负载	**371**	628	630	493	579
(g) 负载的平均精确率 [%]	**92.1%**	90.8%	90.8%	92.2%	89.3%
（h）平均处理时间 [分钟]	**292**	294	255	371	253

处理时间，以及自主移动机器人/服务器中的已加载量/溢出量/下溢量作为训练指标。如图 4(a)所示，所有训练基准方案（即 所提方案、Comp1、Comp2 和 Comp3）对于不同的总奖励均收敛至预期值。 所提方案在演员和评论家网络配置中均采用变分量子电路，可以观察到总奖励从学习开始到980 训练轮次持续增加。随后，所提方案的总奖励达到最终值 -37。 Comp1 和 Comp2共用一个由少量参数构成的状态价值网络，无法恰当地评估其取值，导致奖励处于 -205 和 -240之间。这低于 -200，即执行随机游走时总奖励的预期值。然而，由大量参数构成的经典演员和评论家($\approx$ 40K)表现与所提方案相近（例如 ± 10总奖励方面的性能差异）。

在所提方案和 Comp3 中，通过训练策略评估和改进来提高总奖励。然而，采用由少量参数构成的经典评论家网络的 Comp1 和 Comp2 虽然经过了训练（即演员损失和评论家损失均有所降低），但训练方向并未使奖励增加。换言之，策略评估和改进未能正常工作。所提方案与 Comp1 之间唯一的区别在于评论家是基于量子的还是经典的，而两者在训练性能方面存在巨大差异。此外，与 Comp3 相比，参数数量减少了 364 倍，但两者的性能几乎相当。

*C. LCD*智能工厂环境中的可行性研究

本节研究所提模型在LCD智能工厂环境中的性能。图 4 展示了训练过程中各项指标的结果，表 IV展示了训练结束后的性能。表IV所示结果为100 次迭代推理结果的平均值。图 4（b–c）分别展示了平均质量和处理时间。图 4（d–i）分别展示了仓库和自主移动机器人产生的负载量、溢出负载量和下溢负载量。在此仿真中，若溢出/下溢负载总量为 0，则表示达到目标性能。在训练过程中，可以看到这两个指标（即溢出/下溢负载量）在 所提方案 和 Comp3 中均有所降低。因此可以推断，所提方案 和 Comp3 的训练方向是正确的。另一方面，Comp1 和 Comp2中偶尔会出现溢出负载。然而，其下溢负载量最高，这证明 Comp1 和Comp2 的学习方向仅满足两个目标中的一个。此外，这种趋势也会影响仓库和自主移动机器人智能体的平均负载状态。本文提出的方案中，确认四项指标值均持续下降，直至接近 0。因此，图 4显示自主移动机器人智能体的平均负载为 3.6 千克。

D. 对状态编码方法的影响

根据 [31], [47],状态编码对量子强化学习的性能至关重要。因此，我们设计了一项实验来展示状态编码的重要性。该实验旨在将四个随机比特转换为连续标量值。输出值的计算方式为 $y = \sum_{i=1}^{4} x_i * 2^{1-i}$。在此转换过程中，$\{4, 2, 1\}$ 采用变量密集编码来比较密集编码方法。请注意，变分量子电路中的参数数量固定为 50。结果如图 6所示，由此得出结论：使用 1、2 个变量的密集编码性能较高，而使用 4 个变量的密集编码性能较低。换言之，本文使用的 2 变量编码的性能下降幅度小于 4 变量编码技术。

E. 质量控制的鲁棒性

我们设计了这项实验，以研究所提框架的鲁棒性。为评估其鲁棒性，我们构建了一个随时间变化的智能工厂环境，并将其划分为四个阶段。在阶段1中，负载的精确率从$\{61.9, 95.8, 97.1\}$%中随机选取，与训练环境相同。请注意，初始精确率服从均匀分布 $\mathcal{U}[61.9, 97.1]$%。每台自主移动机器人承载的LCD负载质量随时间变化（例如，阶段2、阶段3和阶段4分别为61.9%、95.8%和97.1%）。随后，我们测量60分钟内的平均精确率，以研究

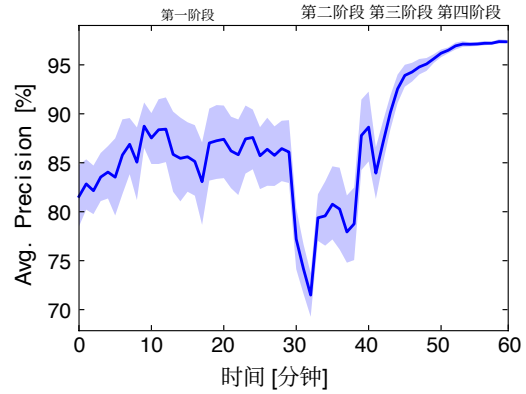


图 5：负载质量随时间变化时质量控制的鲁棒性。

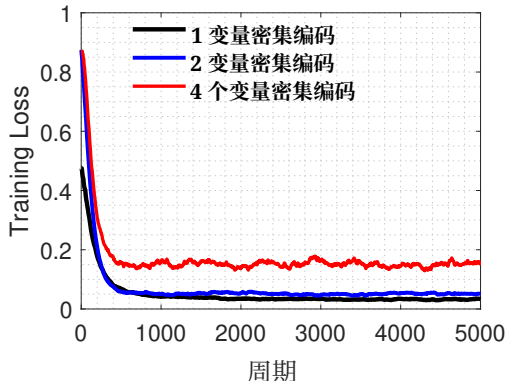


图 6：用于评估各种状态编码方法的微型实验。

质量控制的鲁棒性。图5 展示了 100 次迭代中推理的平均精确率。在阶段1，平均精确率为88.4% 。在 $t = 30$，输入负载的质量下降，即输入负载的精确率降至61.9%。因此，精确率为72.4%，是该回合中的最低精确率。针对这一结果，自主移动机器人智能体在阶段2尝试提高精确率。另一方面，在阶段3和阶段4中，由于输入负载的质量从 61.9% 提高到97.1%，自主移动机器人智能体没有采取质量控制动作。总之，通过展示自主移动机器人智能体应对不可预测的输入负载质量并妥善处理的能力，验证了所提方案中质量控制的鲁棒性。

F. 讨论

本节深入探讨所提方案为何优于其他框架。

1) 可训练参数的表达能力：的作者[48] 认为，与经典神经网络相比，变分量子电路的参数在量子神经网络中具有更强的表达能力。在强化学习/多智能体强化学习领域，可训练参数数量较少会使经典神经网络容易受到性能限制。在 [30], [31],中，研究证明，量子强化学习和量子多智能体强化学习能够达到与经典强化学习/多智能体强化学习。在本文的实验结果中，参数较少的经典神经网络性能较低，原因有二：1）对观测进行索引嵌入；2）采用参数共享策略。对智能体的观测进行索引嵌入，并采用参数共享策略，虽然会牺牲一定性能，但有助于加快收敛。此外，研究者也认为神经网络的表达能力足够，因此采用了上述两种方法 [49]。遗憾的是，无论表达能力如何，性能下降都十分显著。另一方面，即使参数数量较少，量子电路仍能成功运行。

2）与状态编码对应的维度缩减：当输入变量因维度缩减而丢失时，就会产生信息损失。在实验中，输入变量的维度设为四，而不同方案的输出变量分别设为四、二和一。在四变量密集编码中，信息损失较为严重，因为四个独立变量通过单个量子比特，分别使用四个旋转门 $R_Y(x_4)$、$R_Y(x_3)$、$R_Z(x_2)$ 和 $R_Z(x_1)$ 进行编码。对于双变量密集编码和单变量密集编码，不会发生维度缩减。通过展示布洛赫球上的编码过程，可以证明这一点。对于双变量和单变量编码，量子比特分别沿两个正交方向（例如 y 轴和 z 轴方向）旋转两次，以及沿一个方向旋转一次。因此，可以保证所得量子比特的秩。因此，四变量密集编码方法的性能最低，逊于上述其他方法。

VI. 总结

本文研究了基于变分量子电路的量子多智能体强化学习智能体的设计，并考虑部分可观测马尔可夫决策过程，以实现智能工厂管理中的自主多机器人控制与协同。将自主移动机器人用作量子多智能体强化学习智能体时，所提模型采用两变量密集编码方法，以减少量子比特的数量。此外，本文采用结合索引嵌入的参数共享策略，从而减少可训练参数的数量。通过使用上述技术，将量子策略和状态价值函数配置为量子多智能体演员-评论家。大量数值结果表明，所提基于量子多智能体强化学习的自主移动机器人控制在智能工厂管理中具有优越性。最后，与经典多智能体强化学习算法相比，在使用相同数量的参数时，所提量子多智能体强化学习具有明显的性能优势；与其他状态编码方法相比，它也不会遭受严重的数据维度缩减。

参考文献

[1] W. J. Yun、Y. Kwak、J. P. Kim、H. Cho、S. Jung、J. Park 和 J. Kim，《通过变分量子电路设计实现量子多智能体强化学习》，载于 *IEEE*国际分布式计算系统会议（*ICDCS*）论文集，意大利博洛尼亚，2022年7月。[2] W. Na、Y. Lee、N.-N. Dao、D. N. Vu、A. Masood 和 S. Cho，《面向智能制造系统实时数据处理的定向链路调度》，*IEEE*《物联网期刊》，第5卷，第5期，第3661–3671页，2018年。

[3] S. Jeong、W. Na、J. Kim 和 S. Cho，《智能制造系统中的物联网：资源分配中的信任问题》，IEEE《物联网期刊》，第5卷，第6期，第4418–4427页，2018年12月。[4] W. Yu、Y. Liu、T. Dillon、W. Rahayu 和 F. Mostafa，《在智能工厂中采用物联网和大数据技术的健康状态监测集成框架》，IEEE《物联网期刊》，第9卷，第3期，第2443–2454页，2022年。[5] S. Savazzi、V. Rampa 和 U. Spagnolini，《物联网工厂的无线云网络：连接性建模与布局设计》，IEEE《物联网期刊》，第1卷，第2期，第180–195页，2014年。[6] Y. Jin、B. Huang、Y. Yan、Y. Huan、J. Xu、S. Li、P. Gope、L. Xu、Z. Zou 和 L. Zheng，《面向人工智能物联网的基于边缘的协同训练系统》，IEEE工业信息学汇刊，第1–1页，2022年。[7] Z. Nie 和 K.-C. Chen，《智能工厂中的超图实时多机器人任务分配》，IEEE工业信息学汇刊，第1–1页，2021年。[8] G. Fortino、F. Messina、D. Rosaci、G. M. L. Sarn´e 和 C. Savaglio，《面向智能工厂移动智能的基于信任的团队组建框架》，IEEE工业信息学汇刊，第16卷，第9期，第6133–6142页，2020年9月。[9] S. Jeong、W. Na、J. Kim 和 S. Cho，《智能制造系统中的物联网：资源分配中的信任问题》，IEEE《物联网期刊》，第5卷，第6期，第4418–4427页，2018年。[10] E. A. Oyekanlu、A. C. Smith、W. P. Thomas、G. Mulroy、D. Hitesh、M. Ramsey、D. J. Kuhn、J. D. Mcghinnis、S. C. Buonavita、N. A. Looper、M. Ng、A. Ng'oma、W. Liu、P. G. Mcbride、M. G. Shultz、C. Cerasi 和 D. Sun，《自动导引车技术的最新进展综述：基于5G的智能制造应用的集成挑战与研究方向》，IEEE Access，第8卷，第202 312–202 353页，2020年。[11] W. Xia、J. Goh、C. A. Cortes、Y. Lu 和 X. Xu，《工业4.0背景下柔性工厂自动化的自主式自动导引车分布式协同》，载于IEEE自动化科学与工程国际会议（CASE）论文集，2020年，第488–493页。[12] P. Apte，《什么是自主移动机器人？它们如何改变制造业？》，Verizon商业资源与行业洞察文章，2021年10月。 [在线]。网址：https: //
w
w
w.verizon.com/business/resources/articles/s/autonomous-mobile-robots-can-drive-smart-manufacturing/ [13] J. Hu、H. Niu、J. Carrasco、B. Lennox 和 F. Arvin，《通过深度强化学习在未知环境中进行基于Voronoi图的多机器人自主探索》，IEEE车辆技术汇刊，第69卷，第12期，第14 413–14 423页，2020年12月。[14] W. J. Yun、S. Park、J. Kim、M. Shin、S. Jung、A. Mohaisen 和 J.-H. Kim，《通过自主多无人机控制实现可靠监视的协作式多智能体深度强化学习》，IEEE工业信息学汇刊，第1–1页，2022年。[15] M. Schuld 和 N. Killoran，《量子优势是量子机器学习的正确目标吗？》，CoRR，第abs:2203.01340卷，2022年。[16] H. Wang、Y. Ding、J. Gu、Z. Li、Y. Lin、D. Z. Pan、F. T. Chong 和 S. Han，《QuantumNAS：面向稳健量子电路的噪声自适应搜索》，载于IEEE高性能计算机体系结构国际研讨会（HPCA）论文集，2022年4月。[17] W. J. Yun、H. Baek 和 J. Kim，《用于多类别分类的基于投影值测量的量子机器学习》，CoRR，第abs/2210.16731卷，2022年。[18] A. Cornelissen、Y. Hamoudi 和 S. Jerbi，《用于多变量均值估计的近乎最优的量子算法》，载于ACM SIGACT计算理论研讨会（STOC）论文集，美国纽约州纽约，2022年，第33–43页。[19] S. Wiedemann、D. Hein、S. Udluft 和 C. Mendl，《通过振幅估计和Grover搜索实现量子策略迭代——迈向强化学习的量子优势》，arXiv预印本 arXiv:2206 04741，2022年。.[20] F. Arute、K. Arya、R. Babbush、D. Bacon、J. C. Bardin、R. Barends、R. Biswas、S. Boixo、F. G. Brandao、D. A. Buell等，《使用可编程超导处理器实现量子霸权》，《自然》，第574卷，第7779期，第505–510页，2019年。[21] S. Oh、J. Choi 和 J. Kim，《量子卷积神经网络（QCNN）教程》，载于IEEE ICT融合国际会议（ICTC）论文集，2020年10月。[22] S. Jerbi、C. Gyurik、S. Marshall、H. J. Briegel 和 V. Dunjko，《用于强化学习的变分量子策略》，载于神经信息处理系统（NeurIPS）论文集，2021年12月。[23] Z. Hong、J. Wang、X. Qu、X. Zhu、J. Liu 和 J. Xiao，"用于蛋白质距离预测的量子卷积神经网络"，载于 IEEE 国际神经网络联合会议（IJCNN）论文集，2021年7月。[24] Y. Kwak、W. J. Yun、S. Jung 和 J. Kim，"量子神经网络：概念、应用与挑战"，载于 IEEE 泛在网络与未来网络国际会议（ICUFN）论文集，2021年8月。[25] S. Y.-C. Chen、C.-H. H. Yang、J. Qi、P.-Y. Chen、X. Ma 和 H.-S. Goan，"用于深度强化学习的变分量子电路"，IEEE Access，第8卷，页码141 007–141 024，2020年。[26] Y. Kwak、W. J. Yun、S. Jung、J.-K. Kim 和 J. Kim，"量子强化学习导论：理论与基于 PennyLane 的实现 "，载于 IEEE ICT 融合国际会议（ ICTC ）论文集，2021年10月。[27] G. Carleo、I. Cirac、K. Cranmer、L. Daudet、M. Schuld、N. Tishby、L. Vogt-Maranto 和 L. Zdeborov´a，"机器学习与物理科学"，《现代物理评论》，第91卷，第4期，文章编号045002，2019年。[28] P. Sunehag、G. Lever、A. Gruslys、W. M. Czarnecki、V. Zambaldi、M. Jaderberg、M. Lanctot、N. Sonnerat、J. Z. Leibo、K. Tuyls 等，"用于协作式多智能体学习的价值分解网络"，CoRR，2017年6月。[29] P. W. Shor，"降低量子计算机存储器退相干的方案"，《物理评论A辑》，第52卷，第4期，页码R2493，1995年。[30] W. J. Yun、Y. Kwak、J. P. Kim、H. Cho、S. Jung、J. Park 和 J. Kim，"通过变分量子电路设计实现量子多智能体强化学习"，arXiv预印本 arXiv:2203.10443，2022年。[31] O. Lockwood 和 M. Si，"基于量子变分电路的强化学习"，载于 AAAI人工智能与交互式数字娱乐会议（AIIDE）论文集，2020年10月。[32] O. Simeone，"面向工程师的量子机器学习导论"，CoRR，第abs/2205.09510卷，2022年6月。[33] P. W. Shor，"量子计算算法：离散对数与因数分解"，载于 IEEE 计算机科学基础研讨会（FOCS）论文集，美国新墨西哥州圣菲，1994年11月，页码124–134。[34] L. K. Grover，"一种快速的量子力学数据库搜索算法"，载于 ACM计算理论研讨会（STOC）论文集，系列：STOC '96，美国纽约州纽约，1996年，页码212–219。[35] A. Bolu 和 Korc¸ak，"智能仓库中多台移动机器人的路径规划"，载于 IEEE控制、机电一体化与自动化国际会议（ICCMA）论文集，2019年，页码144–150。[36] B. Villalonga、D. Lyakh、S. Boixo、H. Neven、T. S. Humble、R. Biswas、E. G. Rieffel、A. Ho 和 S. Mandr`a，"通过281 pflop/s仿真确立量子霸权前沿"，《量子科学与技术》，第5卷，第3期，文章编号034003，2020年。[37] X. Bai、Y. Fang、W. Lin、L. Wang 和 B.-F. Ju，"利用相位谱对工业图像进行基于显著性的缺陷检测"，IEEE工业信息学汇刊，第10卷，页码2145，2014 4，2135 no. pp. – 。[38] J.-Y. Lee、T.-W. Kim 和 H. J. Pahk，"非周期性薄膜晶体管液晶显示器焊盘区域的稳健缺陷检测方法"，《精密工程与制造国际期刊》，第18卷，第8期，页码1093–1102，2017年。[39] Y. Xia、C. Luo、Y. Zhou 和 L. Jia，"结合频域与空间域技术的薄膜晶体管液晶显示器电路缺陷检测混合方法"，《IEEE半导体制造汇刊》，页码1–1，2022年。[40] T. Xue、R. Li、M. Tokgo、J. Ri 和 G. Han，"基于混合改进量子粒子群优化算法的自主移动机器人轨迹规划"，《软计算》，第21卷，第9期，页码2421–2437，2017年。[41] S. Jung、W. J. Yun、M. Shin、J. Kim 和 J.-H. Kim，"面向云辅助多无人机充电系统的协同调度与多智能体深度强化学习"，IEEE车辆技术汇刊，第70卷，第6期，页码5362–5377，2021年6月。[42] F. A. Oliehoek 和 C. Amato，《去中心化部分可观测马尔可夫决策过程简明导论》。Springer Publishing Company, Incorporated，2016年。[43] G. E. Crooks，"利用参数移位规则和量子门分解计算参数化量子门的梯度"，CoRR，第abs/1905.13311卷，2019年5月。[44] D. Company，"Dell E系列E2311H显示器"，戴尔公司，2022年。[在线]。网址：http://www1.la.dell.com/content/products/productdetails.aspx/monitor-dell-e2311h?c=sr&l=en&s=corp&~tab=specstab [45] M. Schenk、E. F. Combarro、M. Grossi、V. Kain、K. S. B. Li 和 S. Popa，Mircea-Marian aFnd Vallecorsa，"用于 CERN束流线量子强化学习的混合演员-评论家算法"，CoRR，第abs/2209.11044卷，2022年9月。[46] P. Sunehag、G. Lever、A. Gruslys、W. M. Czarnecki、V. Zambaldi、M. Jaderberg、M. Lanctot、N. Sonnerat、J. Z. Leibo、K. Tuyls 等，"用于协作式多智能体学习的价值分解网络"，CoRR，卷 abs/1706.05296 ，2017年6月。

[47] O. Lockwood 和 M. Si，《使用混合量子-经典强化学习玩 Atari》，载于*NeurIPS 2020*年机器学习预注册研讨会论文集，2021年12月，第 285–301 页。[48] A. Abbas、D. Sutter、C. Zoufal、A. Lucchi、A. Figalli 和 S. Woerner，《量子神经网络的力量》，《自然计算科学》，第 1 卷，第 6 期，第 403–409 页，2021年。[49] T. Rashid、M. Samvelyan、C. Schroeder、G. Farquhar、J. Foerster 和 S. Whiteson，《QMIX：用于深度多智能体强化学习的单调价值函数分解》，载于国际机器学习会议（*ICML*）论文集，瑞典斯德哥尔摩，2018年7月，第 4295–4304 页。



Won Joon Yun 自2021年3月起在大韩民国首尔的高丽大学攻读电气与计算机工程博士学位，并于该校获得电气工程理学学士学位。2022年夏季，他曾在美国加利福尼亚州圣何塞的 Cipherome Inc. 担任访问研究员；2022年冬季，他还曾在美国加利福尼亚州洛杉矶的南加州大学担任访问研究员，参与与南加州大学维特比工程学院谢明电气与计算机工程系 Andreas F. Molisch 教授合作的联合项目。

他的当前研究兴趣包括各领域的机器学习、量子机器学习和多智能体强化学习。



Jae Pyoung Kim 自2023年3月起在大韩民国首尔的高丽大学攻读电气与计算机工程理学硕士学位，并于该校获得电气工程理学学士学位。电气与计算机工程专业学生。2021年至2022年，他在高丽大学人工智能与移动性（AIM）实验室担任研究工程师。他当前的研究兴趣包括量子机器学习。



Soyi Jung 自2022年9月起担任大韩民国水原亚洲大学电气与计算机工程系助理教授。她还于2021年至2022年在美国加利福尼亚州的加利福尼亚大学欧文分校唐纳德·布伦信息与计算机科学学院担任访问学者。2021年，她在大韩民国首尔的高丽大学担任研究教授。2015年至2016年，她还在大韩民国果川的韩国测试研究院（KTR）担任研究员。

她分别于2013年、2015年和2021年在大韩民国水原亚洲大学获得电气与计算机工程理学学士、理学硕士和博士学位。

她目前的研究兴趣包括自动驾驶汽车通信的网络优化、分布式系统分析、大数据处理平台以及概率接入分析。她曾获KICS最佳论文奖（2015年）、WISET和KICS青年女性研究者奖（2015年）、IEEE首尔分会学生论文竞赛铜奖论文奖（2018年）、《电子时报》ICT论文竞赛奖（2019年）以及IEEE ICOIN最佳论文奖（2021年）。



Jae-Hyun Kim 于1991年、1993年和1996年分别获得韩国安山汉阳大学计算机科学与工程专业的理学学士、理学硕士和博士学位。在（年份缺失），他曾作为访问学者在日本东京通信研究实验室工作。1997年4月至1998年10月，他在加利福尼亚大学洛杉矶分校电气工程系担任博士后研究员。1998年11月至2003年2月，他在性能建模和QoS管理部门担任技术人员，ment部门、贝尔实验室、朗讯科技（新泽西州霍姆德尔）。自2003年起，他一直在韩国水原亚洲大学电气与计算机工程系担任教授。

他是韩国信息通信技术振兴院资助的卫星信息融合应用服务研究中心（SICAS）中心主任。自2018年起，他担任韩国5G论坛智慧城市委员会主席。自2022年起，他担任韩国通信与信息科学学会（KICS）副会长。他是IEEE、KICS、韩国电子与信息工程师学会（IEIE）以及韩国信息科学家与工程师学会（KIISE）的会员。他曾获IEEE ICOIN最佳论文奖（2021年）。



Joongheon Kim （M’06–SM’18）自2019年起任职于韩国首尔的高丽大学，目前担任电气工程学院副教授，同时兼任通信工程系（由三星电子设立/资助）和半导体工程系（由SK海力士设立/资助）的兼职教授。他于2004年和2006年分别获得韩国首尔高丽大学计算机科学与工程专业的理学学士和理学硕士学位；并获得计算机

他于2014年在美国加利福尼亚州洛杉矶的南加州大学（USC）获得理学学位。在加入高丽大学之前，他曾任LG电子（韩国首尔，2006—2009年）研究工程师、英特尔公司（美国加利福尼亚州圣克拉拉，2013—2016年）系统工程师，以及中央大学（韩国首尔，2016—2019年）计算机科学与工程助理教授。

他担任IEEE《车辆技术汇刊》、IEEE《机器学习与通信和网络汇刊》以及IEEE《通信标准杂志》的编辑。他还是*IEEE*通信学会（*ComSoc*）和*IEEE*系统委员会的杰出演讲人。

他曾获南加州大学（USC）博士录取时颁发的安嫩伯格研究生奖学金（2009年）、英特尔公司下一代与标准（NGS）部门表彰奖（2015年）、IEEE《系统期刊》最佳论文奖（2020年）、IEEE ComSoc多媒体通信技术委员会（MMTC）杰出青年研究员奖（2020年）、IEEE ComSoc MMTC最佳期刊论文奖（2021年），以及*ICT Express*（爱思唯尔）颁发的最佳专刊客座编辑奖（2022年）。他还获得了多项IEEE会议奖项，包括IEEE ICOIN最佳论文奖（2021年）、IEEE车辆技术学会（VTS）首尔分会奖（2019年、2021年和2022年），以及IEEE ICTC最佳论文奖（2022年）。