

控制与决策

Control and Decision

基于模仿学习的工业机器人控制策略

朱易凡, 刘晋飞, 黄华, 陈明

引用本文:

朱易凡, 刘晋飞, 黄华, 等. 基于模仿学习的工业机器人控制策略[J]. *控制与决策*, 2026, 41(5): 1427-1438.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2025.0749>

您可能感兴趣的其他文章

Articles you may be interested in

一种结合内在动机理论的移动机器人环境认知模型

An environment cognition model combined with intrinsic motivation for mobile robots

控制与决策. 2021, 36(9): 2211-2217 <https://doi.org/10.13195/j.kzyjc.2019.1744>

移动机器人运动规划中的深度强化学习方法

Deep reinforcement learning for motion planning of mobile robots

控制与决策. 2021, 36(6): 1281-1292 <https://doi.org/10.13195/j.kzyjc.2020.0470>

基于未知系统动态估计的机器人预设性能控制

Unknown system dynamics estimator for prescribed performance control of robotic systems

控制与决策. 2021, 36(5): 1040-1048 <https://doi.org/10.13195/j.kzyjc.2019.1219>

基于虚拟结构法的多移动机器人分布式预测控制

Distributed predictive control of multiple mobile robots based on virtual structure method

控制与决策. 2021, 36(5): 1273-1280 <https://doi.org/10.13195/j.kzyjc.2019.1136>

基于接触状态感知发育的机器人柔性装配方法

Flexible assembly method based on contact state perception development

控制与决策. 2021, 36(4): 876-884 <https://doi.org/10.13195/j.kzyjc.2019.1079>

基于模仿学习的工业机器人控制策略

朱易凡¹, 刘晋飞^{2†}, 黄华³, 陈明¹

(1. 同济大学机械与能源工程学院, 上海 201804; 2. 同济大学中德工程学院, 上海 201804;
3. 同济大学工程实践中心, 上海 201804)

摘要: 针对当前智能制造背景下复杂多变的工业应用场景对工业机器人智能化提升的需求, 提出一种基于 Transformer 的模仿学习控制策略. 首先, 设计示教机器人模型并搭建模仿学习实验平台, 降低专家演示数据采集难度, 提高效率; 其次, 提出基于夹爪状态的动作分块预测模仿学习模型 (GACT), 在其中引入面向气动夹爪状态的二元交叉熵损失函数, 并为其设计独立的交叉注意力机制; 再次, 在 MuJoCo 环境中进行仿真验证, 并通过消融实验评估模型优化效果; 最后, 在 KUKA 工业机器人平台上开展实物实验验证. 仿真结果表明, GACT 模型相较于基线及其变体具有更高的任务成功率与轨迹准确性; 实物实验结果进一步验证了该模型可有效实现工业机器人的运动控制.

关键词: 模仿学习; 工业机器人; 控制策略; 智能制造; 行为克隆; Transformer

中图分类号: TP242.2 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2025.0749

引用格式: 朱易凡, 刘晋飞, 黄华, 等. 基于模仿学习的工业机器人控制策略 [J]. 控制与决策, 2026, 41(5): 1427-1438.

Control strategy of industrial robot based on imitation learning

ZHU Yi-fan¹, LIU Jin-fei^{2†}, HUANG Hua³, CHEN Ming¹

(1. School of Mechanical Engineering, Tongji University, Shanghai 201804, China; 2. Sino-German College of Applied Sciences, Tongji University, Shanghai 201804, China; 3. Engineering Practice Center of Tongji University, Shanghai 201804, China)

Abstract: In view of the complex and dynamic industrial application scenarios in intelligent manufacturing and the growing demand for enhanced intelligence in industrial robots, this paper proposes a Transformer-based imitation learning control strategy. First, a demonstration robot model is designed and an imitation learning experimental platform is established to simplify the acquisition of expert demonstration data and improve efficiency. Second, a gripper-based action chunking with Transformers (GACT) model is proposed, which incorporates a binary cross-entropy loss function tailored to pneumatic gripper states and introduces an independent cross-attention mechanism dedicated to gripper states. Third, simulations are conducted in the MuJoCo environment, and ablation studies are performed to evaluate the effectiveness of the proposed components. Finally, real-world experiments are carried out on a KUKA industrial robot. Simulation results show that the GACT model achieves higher task success rates and trajectory accuracy than the baseline and its variants. Experimental results further confirm that the GACT can effectively realize precise motion control of industrial robots.

Keywords: imitation learning; industrial robots; control strategy; intelligent manufacturing; behavior cloning; Transformer

0 引言

随着工业机器人的大规模发展与应用, 制造业的生产效率得到了显著提高. 然而, 针对当前智能制造背景下复杂多变的工业应用场景需求, 工业机器

人的智能化水平仍有待提高. 因此, 如何提高工业机器人的智能化程度是一个重要研究方向. 其中, 机器人的目标识别与高精度抓取控制能力为其核心目标^[1]. 目前通过融合人工智能理论和机器学习技术,

收稿日期: 2025-07-16; 录用日期: 2025-12-29.

基金项目: 国家自然科学基金项目 (71601144).

责任编委: 虞文武.

†通信作者. E-mail: jinfeil@tongji.edu.cn.

本文附带电子附录文件, 可登录本刊官网该文“资源附件”区自行下载阅览.

已经出现大量机器人智能控制方法,运用机器视觉反馈结合点云与6D位姿可以让机器人精准定位目标^[2-4];运用强化学习与深度神经网络可以有效提高机器人的避障抓取能力^[5-6]。此类控制方法虽然能够有效减少工业机器人传统编程控制所需的时间,但针对不同场景需要大量调试时间,且此类基于神经网络与视觉的智能控制方法操作学习成本高,上手困难,在工业物联网环境下复杂多变的任务场景中,需要极高的部署、调试、数据传输成本与安全性。对于实际应用场景仍有局限性^[7]。

作为机器人技能学习中的一个重要分支,模仿学习近年来在机器人领域中得到了广泛的应用,是一种实现机器人智能控制的全新方式,可以大幅度减少操作学习成本与上手难度,仅需根据少量专家演示训练即可实现智能控制。模仿学习是指通过从专家示范数据中提取行为策略,使智能体在未知环境中复现专家级决策能力,尤其适用于奖励函数难以显式定义的复杂任务^[8]。在机器人领域应用模仿学习的方法包括行为克隆(BC)^[9]、逆强化学习(IRL)^[10]以及生成对抗模仿学习(GAIL)^[11]等。与基于轨迹的模仿学习研究方向不同的是,这一类以行为克隆为基础展开的模仿学习方法主要侧重解决马尔科夫决策过程(MDP)中的决策问题,其主要的特点是智能体与环境存在交互且任意时刻的交互都会影响MDP下一时刻的状态^[12]。

行为克隆无需设计复杂的奖励函数或进行试错探索,直接利用专家数据通过监督学习端到端训练策略网络,并快速提取专家策略,极大程度上减少复杂神经网络的调试部署时间,是当下端到端模仿学习的研究热点。然而,行为克隆方法存在固有挑战,其中最突出的是复合误差积累问题。在长序列决策任务中,智能体与环境的微小偏差会随时间不断累积,导致最终性能显著下降^[13]。此外,该方法对专家演示数据的质量要求较高,收集足量、高质量的专家数据集存在一定难度与时间成本。

随着越来越多的研究者致力于克服行为克隆的固有缺陷,端到端模仿学习作为一条重要的技术路线得到了显著发展。这类方法旨在直接从高维图像数据映射到动作控制指令,最大限度地减少中间表示环节。早期工作(如视觉模仿学习网络(VINN)^[14])探索了通过卷积神经网络实现从像素到动作的端到端策略学习。近年来,扩散模型等生成式AI技术也被引入该领域,例如Diffusion Policy^[15]利用扩散模型的生成能力来拟合多模态的动作分布,在策略表达的灵活性和多样性上展现出独特优势。

基于Transformer网络架构的方法^[16]因其在序列建模方面的核心优势而备受关注。由于Transformer的深层结构能够处理大量专家演示数据与长时间序列的复杂任务,以及使用多头注意力机制处理多模态输出的能力,融合Transformer网络与行为克隆思想的模仿学习表现出更强的预测准确性、灵活度与更高的训练效率,有效缓解了传统行为克隆中的复合误差积累问题。Shafiullah等^[17]提出行为转换器(BeT),通过离散化动作来预测多个未来动作模式;Reed等^[18]提出的Gato模型,使用单一的Transformer权重处理从文本、图像到机器人控制的多样化任务;Brohan等^[19]提出了Robotics Transformer(RT-1),是一种以端到端方式将视觉和语言指令直接映射为机器人动作的Transformer模型;Zhao等^[20]提出了基于动作分块预测模仿学习模型(ACT),仅使用50个专家演示数据便可实现对任务高达90%成功率的复现。也有大量的研究者在ACT模仿学习模型的基础上展开研究,Buamaneec等^[21]提出了Bi-ACT模型,将机械臂力学数据融入ACT模型;Li等^[22]提出了Haptic-ACT模型,改变了数据采集方式,使用虚拟现实技术作为数据采集方式;Yang等^[23]提出了MACT模型,融合记忆动作分块方法在医疗手术任务中使用模仿学习模型。这些工作表明,基于Transformer的模仿学习在未来智能制造领域中具有很强的研究价值。

尽管前景广阔,将基于Transformer的模仿学习应用于应对工业机器人复杂应用场景中的高维连续动作空间与多模态输入问题,仍是一个前沿挑战。具体而言,现有方法仍面临以下问题:在长序列场景下,复合误差积累问题仍可能导致最终任务动作预测误差失控;对专家演示的质量要求较高,收集专家数据集的难度与时间成本较高;针对离散与连续位置数据融合是一大挑战,例如工业领域的气动与液压的离散关节数据处理与电机驱动关节的连续数据融合;智能化程度有待提高,无法实现应对突发情况的实时响应。因此,现有研究对基于工业机器人采用模仿学习算法实现运动控制的关注较少,不论是控制精度,编程时间成本,离散数据处理与智能化程度,都是基于工业机器人的模仿学习模型需要解决的问题。

针对以上问题,本文提出基于工业机器人的GACT(gripper-based action chunking with transformers)模仿学习模型,该模型是以Transformer模型为基础的端到端行为克隆模式的模仿学习模型。其主要贡献为:提出了二元交叉熵损失函数,旨在解决工业领域夹爪离散状态数据与连续关节数据融合问题;使用动作分块预测方法以降低长序列场景下模仿学习

的复合误差积累, 提高控制准确性; 设计示教机器人模型, 提高专家演示数据收集效率. 具体内容包括: 首先, 搭建示教机器人模型与实验场景, 提高专家演示数据收集效率, 减少专家数据采集所需的编程时间, 为后续在工业机器人高效收集专家数据集提供参考; 其次, 设计改进 GACT 模仿学习模型, 通过使用 Resnet-18 网络作为图像处理骨干网络提升机器人对环境的感知能力, 并采用条件变分自编码器 (CVAE) 与 Transformer 解码器组合的方法, 加强模型对机器人状态空间与动作序列关系的理解能力, 学习不同专家演示动作序列数据分布, 从而提高模型鲁棒性; 最后, 通过结合动作分块预测思想减少复合误差积累问题, 并提出夹爪状态交叉注意力机制, 提高模型对夹爪状态的关注, 结合二元交叉熵损失函数, 解决模型对气动离散夹爪数据的预测问题, 使模型适配工业机器人场景. 本文通过不同场景的仿真与实际实验, 并结合消融实验, 验证 GACT 模型相较于基线模型具有更高的成功率与准确性, 并表明在工业机器人领域应用模仿学习的可行性, 有助于模仿学习方法在工业机器人领域的推广发展.

1 专家数据采集与实验场景搭建

高质量专家演示数据的获取是模仿学习应用中的关键环节. 传统的机器人轨迹模仿学习方法对数据精度要求极高, 其采集过程通常依赖专业设备的精确标定与示教, 不仅流程复杂、成本高昂, 还对操作人员的熟练度有极高要求. 尽管现有研究采用了协作机器人拖拽^[24]或手术机器人远程控制模块^[25]等方法, 但其采集高质量专家数据的难度依然居高不下.

为了提高专家演示数据的采集效率, 本文借鉴遥操作的思想^[26-27], 设计并搭建一套低成本、高效率的专家数据采集系统. 本文所提出的 GACT 模型通过条件变分自编码器学习动作序列的分布特性, 降低对专家演示数据绝对精度的依赖; 其端到端的学习范式将任务完成情况作为核心评价指标, 可进一步降低对数据采集过程的要求. 该系统的具体实现方式是构建一个示教机器人模型, 通过关节位置同步控制的方法, 实现专家对工业机器人的直观、安全遥操作, 从而高效生成专家演示数据集.

1.1 示教机器人设计

实验平台基于同济大学“工业 4.0-智能工厂实验室”的工业场景, 以 KUKA 工业机器人及 KRC4 控制柜作为研究对象. 为克服传统 KRL 编程或示教器手动控制效率低、门槛高的问题, 设计一套基于示教机器人的遥操作专家演示数据采集方案. 基于

KUKA 机器人的三维模型, 按比例缩小并保留了原有的 6 轴串联拓扑结构, 通过 3D 打印制造了示教机器人模型. 各关节集成了伺服电机、驱动器及绝对值编码器. 为了满足高精度角度传感器与直观控制器的功能, 示教模型关闭了伺服驱动器的使能信号, 使关节可被手动自由旋转; 同时为绝对值编码器独立供电, 使其能持续输出精确的关节位置信息. 示教机器人系统如图 1 所示.

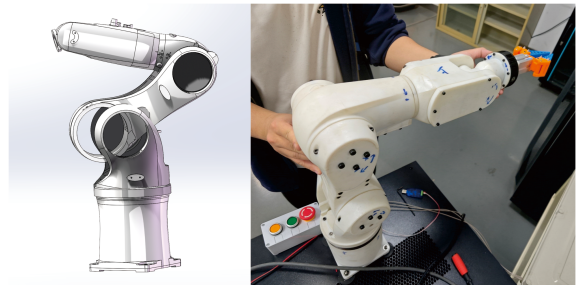


图1 示教机器人三维模型与实物图

通讯方面, 示教机器人模型的关节角度数据通过 RS-485 通讯协议实时传输至上位机. 上位机程序将关节角度数据封装成 UDP 协议数据包, 发送至 KRC4 控制柜. KUKA 机器人配置为外部自动输入模式, 接收并执行来自上位机的关节角度指令, 从而实现与示教模型的同步运动. 在此过程中, 上位机记录来自示教模型的关节角度、KUKA 机器人反馈的关节位置及夹爪状态, 并同步保存摄像头图像, 以构建专家演示数据集.

为了保证遥操作控制的准确性与实验安全性, 在实验开始前对示教模型进行关节零位对齐, 并为其关节运动范围设置物理硬限位. 同时, 在上位机程序中严格限制发送给 KUKA 机器人的关节角度范围, 在模型推理时, 对输出的动作向量进行裁剪, 防止遥操作过程中 KUKA 机器人运动失控造成安全事故.

1.2 实验场景搭建

实验场景包含 3 个摄像头, 操作平台, KUKA 机器人以及示教机器人模型. 其中 2 个摄像头分别固定于操作平台左右两侧, 作为全局场景视图, 加强机器人对操作平台环境状态的认知; 第 3 个摄像头安装在机器人第 6 轴末端, 加强机器人对自身末端位姿状态的认知. 绿色框中为目标物体出现区域, 红色标记为目标物体放置位置. 在演示数据采集过程中, 专家通过操作示教模型, 实现 KUKA 机器人关节位置与示教模型实时对应, 直观的遥操作简化了专家演示数据集的采集过程. 具体实验场景如图 2 所示.

专家演示数据包括 3 个摄像头图像数据: 机器人关节状态、夹爪状态数据以及机器人动作序列. 每

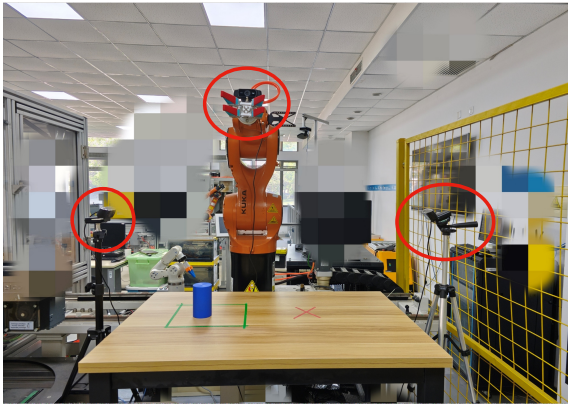


图2 KUKA 机器人实验场景

组任务采集 50 个专家演示数据样本, 每个专家演示数据样本长度固定, 在采集过程中的每个时间步, 接收示教机器人模型的各轴动作序列数据作为模型动

作空间; 接收 KUKA 机器人发送的各轴关节位置与夹爪状态数据; 同时保存摄像头图像数据, 并处理为 3 通道 RGB 数据, 最终储存为 hdf5 格式文件.

2 基于工业机器人的模仿学习模型设计

为将端到端模仿学习有效应用于工业机器人场景, 本文在 ACT 模型^[20] 基础上提出 GACT 模型. 如图 3 所示, 该模型以 Transformer 为核心架构, 融合视觉特征提取、条件变分自编码 (CVAE) 与动作分块预测机制, 并引入夹爪状态专用的交叉注意力模块与二元交叉熵损失函数, 以显式建模离散夹爪状态对动作生成的影响. GACT 模仿学习模型结构主要分为变量输入阶段、编码阶段、解码阶段以及变量输出阶段.

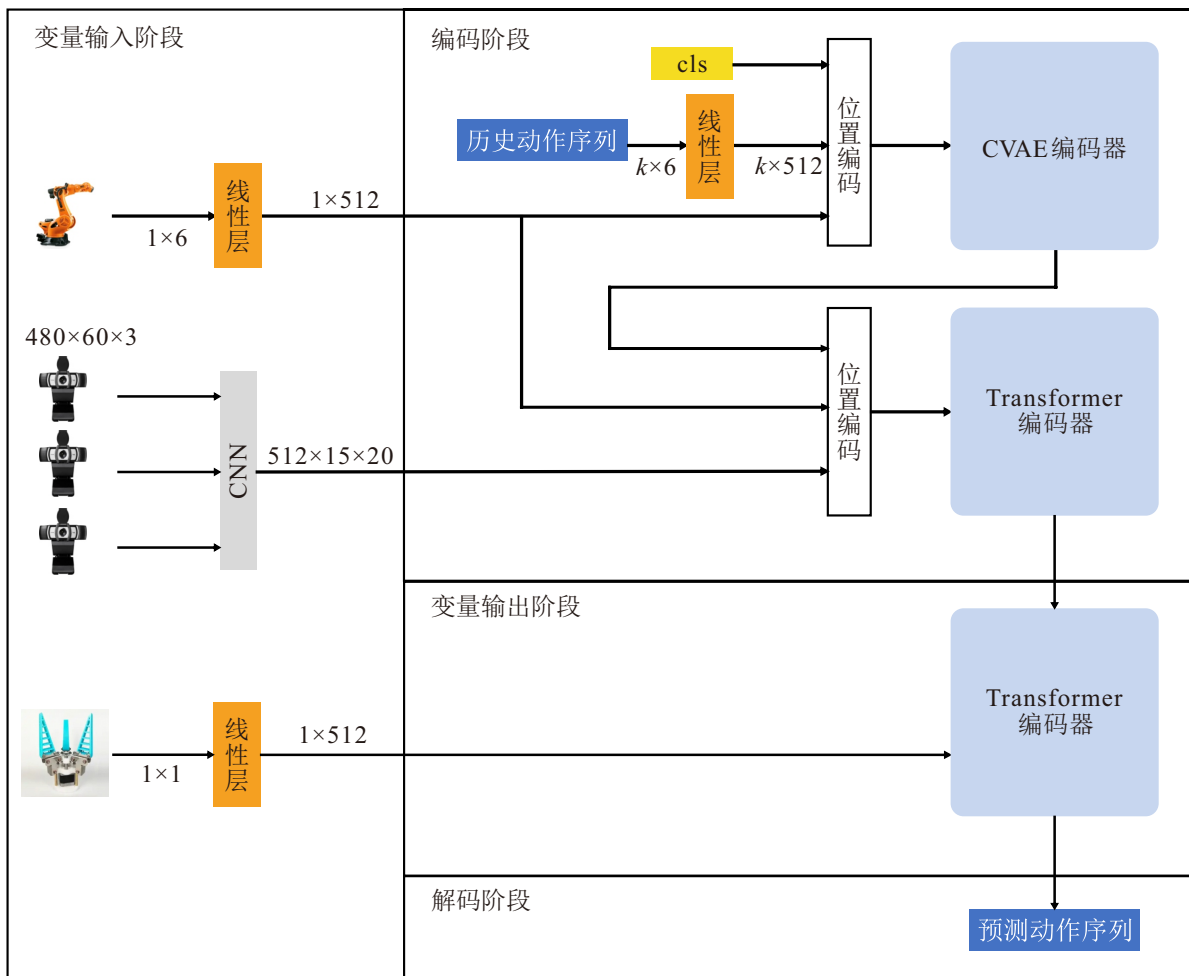


图3 GACT 模仿学习模型结构

2.1 变量输入阶段

模型的输入为多模态观测数据, 包含 3 个摄像头的图像数据、当前机器人关节位置状态及夹爪状态. 首先, 随机划分 80% 的专家演示数据为训练集, 20% 作为验证集, 并对输入数据进行预处理以便输入数据满足模型维度要求.

关节位置状态变量采用 Z-score 归一化, 夹爪状态变量保持原始形式, 通过一个线性投影层映射到高维隐藏空间. 图像通过 ResNet-18^[28] 的网络架构作为视觉特征提取骨干网络, 提取 512 维特征图, 并施加二维位置编码以保留空间信息, 最终与关节和夹爪嵌入向量拼接后输入编码器.

2.2 编码阶段

编码阶段由两个核心组件构成, 分别是条件变

分自编码器 (CVAE) 和 Transformer 编码器. 其主要结构如图 4 所示.

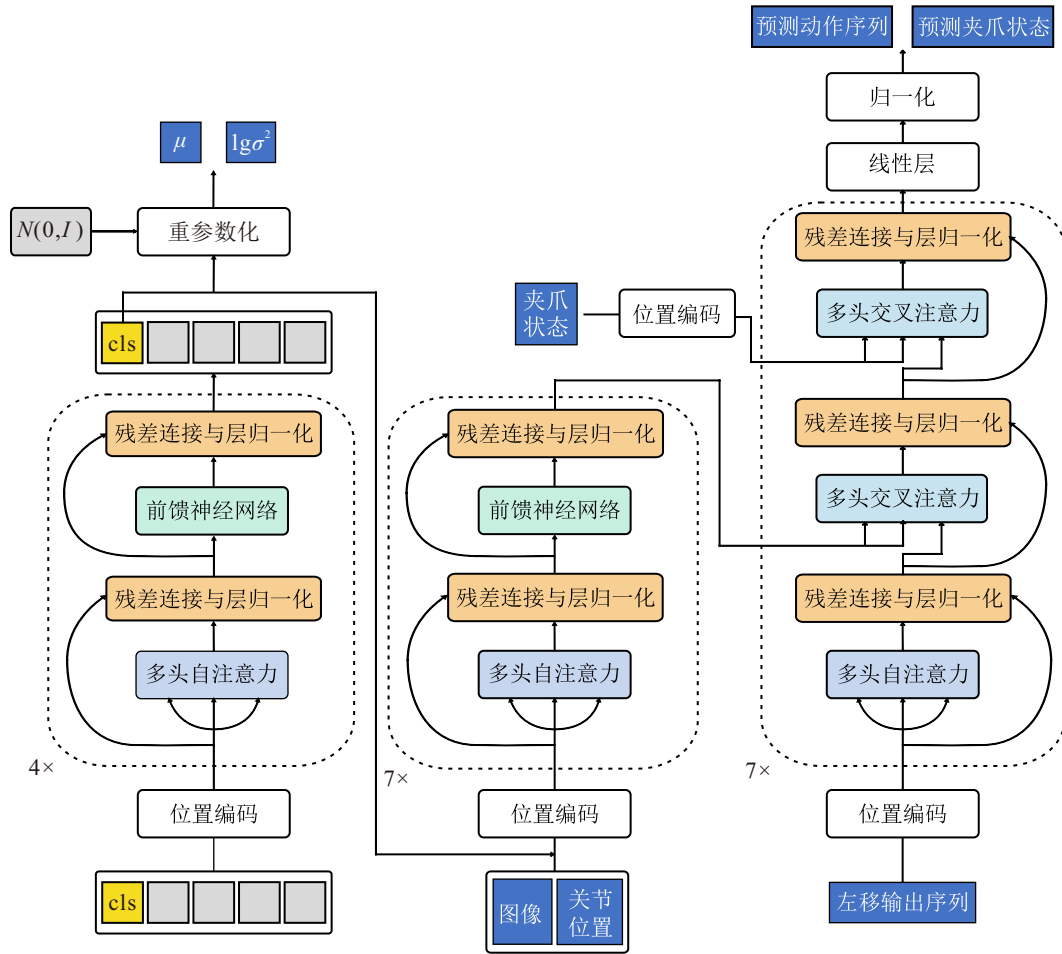


图4 编码器与解码器主要结构

2.2.1 条件变分自编码器

条件变分自编码器负责将输入数据和条件信息共同映射到潜在空间的概率分布. 与传统 CVAE 编码器不同, 本模型调用 Transformer 编码器结构代替传统 CVAE 编码器的 MLP 神经网络, 生成基于 Transformer 的条件变分自编码器. 其核心结构由 4 个 Transformer 编码层聚合而成, 以当前关节状态和未来 k 步动作为输入条件, 输出潜在变量 z 的均值与方差.

在评估阶段, 由于未来动作序列未知, 本文将 CVAE 的条件分布设置为标准正态分布, 并将潜在变量 z 的输入设为零向量, 即从标准正态分布的均值采样, 消除随机性影响, 从而满足工业控制对稳定性的要求.

2.2.2 Transformer 编码器

Transformer 编码器是由 4 个标准的 Transformer 编码层聚合而成. Transformer 编码层包含多头自注意力机制与前馈神经网络层, 并在每个层后都进行

残差连接与层归一化处理.

Transformer 编码器融合视觉特征、关节状态及潜在变量 z , 通过多头自注意力机制建模多模态输入间的全局依赖关系, 其输出传递给后续解码器作为键 (Key) 与值 (Value) 向量.

具体而言, Transformer 编码器的原理为, 利用多头自注意力机制将输入向量分别映射到多个不同的语义子空间中, 使模型能够并行地捕捉多种类型的依赖关系. 每个注意力头独立计算其对应子空间中的注意力权重, 最终将所有头的输出整合并进行线性变换, 输出为融合了全局上下文信息的深度特征表示并作为键向量和值向量提供给解码器的交叉注意力层.

2.3 解码阶段

解码阶段的核心是 Transformer 解码器, 由 7 个 Transformer 解码层堆叠而成. 每个解码层由 3 个注意力机制层组成, 分别是自注意力机制层、环境交叉注意力机制层以及夹爪交叉注意力机制层.

在每个注意力机制层后都需要残差连接与层归一化的处理,最后通过前馈网络完成特征变换得到输出.其主要结构如图4所示.

夹爪交叉注意力机制层为 GACT 模型的重点,专用于处理离散夹爪状态.其查询向量来自解码器输出,而键向量和值向量则是当前夹爪状态经过线性投影后的特征.此设计使模型能动态关注末端执行器的实时状态,将连续关节控制与离散夹爪控制分离,避免环境特征与末端执行器状态特征混淆,尤其适用于气动夹爪等离散执行器,从而根据任务的不同阶段预测动作序列,有助于提高模型的预测准确率.

最终,解码器的输出通过两个独立的线性层分别预测未来的动作序列和夹爪状态序列.

2.4 变量输出阶段

变量输出阶段作为模型训练与评估时对输出变量实现处理的重要阶段,主要由损失函数与时间整合机制组成.

2.4.1 损失函数

在 Transformer 模型中,损失函数是驱动模型优化的核心工具,利用损失函数量化模型预测结果与真实目标之间的差异,通过反向传播梯度调整自注意力机制和全连接层的参数,使模型能够提取与任务相关的关键特征,并做出符合真实状态的模型预测结果.模型总损失函数由3部分加权求和构成,以便同时优化连续动作和离散夹爪状态的预测.

本文使用 L1 损失(平均绝对误差)构建模型预测动作序列与真实动作序列之间的损失函数.L1 损失衡量的是预测与真实动作序列之间的平均绝对误差,模型中动作序列 L1 损失可以表示为

$$\mathcal{L}_{L1} = \frac{1}{n} \sum_{i=1}^n |a_{\text{pred}}^{(i)} - a_{\text{true}}^{(i)}|. \quad (1)$$

其中: n 为样本数量, $a_{\text{pred}}^{(i)}$ 为第 i 个样本真实动作序列, $a_{\text{true}}^{(i)}$ 为对应的模型预测输出动作序列.

本文使用二元交叉熵损失构建模型预测夹爪状态序列与真实夹爪状态序列之间的损失函数.二元交叉熵损失是用于衡量二分类任务中模型预测概率分布与真实标签分布之间差异的损失函数.其核心思想是通过惩罚预测概率偏离真实标签的程度,驱动模型输出逼近真实二值分布.夹爪状态序列的 BCE 损失函数可表示为

$$L_{\text{BCE}} = -\frac{1}{n} \sum_{i=1}^n [g_{\text{true}}^{(i)} \lg(g_{\text{pred}}^{(i)}) + (1 - g_{\text{true}}^{(i)}) \lg(1 - g_{\text{pred}}^{(i)})]. \quad (2)$$

其中: n 为样本数量, $g_{\text{true}}^{(i)}$ 为第 i 个样本的真实夹爪状态序列, $g_{\text{pred}}^{(i)}$ 为模型预测夹爪状态为正类的概率序列.

本文使用 KL 散度损失构建 CVAE 编码器的输出潜在变量 z 的分布与先验分布之间的损失函数.使用 KL 散度强制模型后验分布逼近先验分布,防止编码器过度拟合训练数据,确保潜在变量 z 的分布具有连续性和平滑性.由于模型中先验与后验分布均为标准高斯分布, KL 散度损失函数可以表示为

$$\mathcal{L}_{\text{KL}} = \frac{1}{2} \sum_{i=1}^K (\mu_i^2 + e^{\lg \sigma_i^2} - \lg \sigma_i^2 - 1). \quad (3)$$

其中: d 为潜在空间维度; μ_i 和 σ 为后验分布的均值和方差,是编码器网络的输出变量.

为了避免过早抑制 Transformer 模型的表达能力,模型采用 KL 余弦退火,周期性调节 KL 项的强度.权重系数 $\beta(t)$ 随训练步数 t 按以下规律变化:

$$\beta(t) = \beta_{\min} + \frac{1}{2} (\beta_{\max} - \beta_{\min}) \left(1 + \cos\left(\frac{t\pi}{T}\right)\right). \quad (4)$$

其中: T 为退火周期; β_{\min} 和 β_{\max} 为权重的最小值和最大值,模型中设置为 0 和 10.在每个周期 T 内, $\beta(t)$ 从 β_{\min} 开始,先升至 β_{\max} ,再降回 β_{\min} ,形成升降周期,避免单一方向优化导致的局部最优.

因此,模型的总损失函数由 L1 损失、二元交叉熵损失以及 KL 散度损失求和组成.总损失函数

$$\mathcal{L}_{\text{total}} = \beta(t)\mathcal{L}_{\text{KL}} + \lambda_1 \mathcal{L}_{L1} + \lambda_2 \mathcal{L}_{\text{BCE}}. \quad (5)$$

其中: $\beta(t)$ 为 KL 退火系数; λ_1 和 λ_2 为 L1 损失与 BCE 损失权重,用于调整各个损失对于模型的重要性,模型中对于 L1 损失与 BCE 损失权重均设置为 1.

2.4.2 时间整合策略

本模型的输出预测基于动作分块的方式,其核心思想是根据当前状态预测未来 k 步动作.将机器人的状态空间与动作空间分成大小相同的动作块,再分别投入训练与评估过程.使用此方法能够有效减少长时间序列可能积累的复合误差.

为了消除将完整任务分成多个动作块可能导致的运动割裂,确保动作序列平滑,本文采用时间整合(temporal aggregation)的策略,给历史预测中同一时间步的动作分配指数衰减的权重,时间越近的预测权重越高.该指数衰减函数计算公式为

$$\omega_i = \frac{e^{-ki}}{N-1}, \quad i \in [0, N-1]. \quad (6)$$

其中: $k = 0.01$ 为衰减率, N 为有效动作数量.该权重赋予近期动作更高重要性,同时保留历史动作的

平滑贡献.

基于该权重序列, 提取历史模型输出序列中当前时间步的动作预测指令, 对其进行加权求和, 生成当前时间步的最终动作指令

$$\text{raw_action} = \sum_{i=0}^{N-1} \omega_i \cdot \text{actions_for_curr_step}[i]. \quad (7)$$

该时间整合机制显著提升了连续控制任务中动作输出的稳定性和抗噪能力, 适用于需要长程状态依赖的复杂操作场景.

3 实验与验证

3.1 超参数选择

超参数的合理选择对模仿学习模型的训练效果至关重要. GACT 模型的超参数可分为训练策略参数、模型架构参数以及任务配置参数. 在训练策略参数方面, 学习率 lr 设置为 $1e-5$; 单次梯度计算的批次大小 batch_size 设置为 8; KL 散度退火策略的 KL 散度权重最大值 kl_weight 设置为 10. 时间序列分块长度 chunk_size 设置为 100, 约为完整任务示教样本

长度 400 时间步的 1/4, 旨在平衡模型的长序列预测能力与训练效率. 在模型架构参数方面, 隐藏层维度 hidden_dim 设置为 512, 前馈网络维度设置为 3 200, 编码器层数为 4, 解码器层数为 7, 旨在平衡图像特征提取效率与处理多模态信息的融合与输出的能力. 多头注意力头数为 8. 视觉骨干网络选择 Resnet-18 而非更大模型, 以避免过度高维视觉特征挤压机器人控制信号的权重. 在任务配置参数方面, 根据实际任务时间长度, 设置每个专家演示样本长度为 400 时间步, 采样周期为 0.02 s, 机器人状态维度为 6, 对应机器人 6 轴关节位置.

3.2 仿真实验场景

为了保证工业机器人实验场景的安全性并高效验证模仿学习模型性能, 本实验使用 MuJoCo 物理仿真环境构建与实物一致的“抓取-放置”仿真环境, 其实验过程如图 5 所示. 仿真环境中导入了 KUKA 机器人模型及 3 路虚拟摄像头. 目标物体随机出现在指定的红色区域内, 机器人需将其抓取并放置到黄色标记位置.

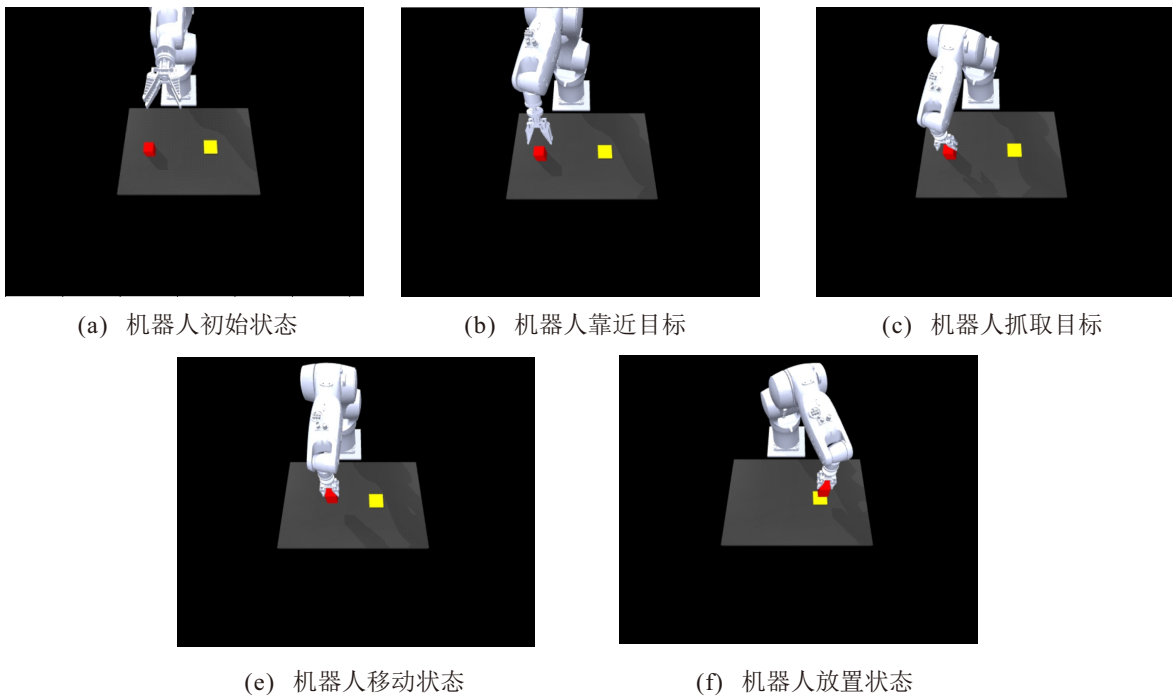


图5 “抓取-放置任务”仿真实验过程

仿真环境下的专家数据采集包含两个阶段: 路径演示记录与策略验证执行. 在路径演示记录阶段中, 通过在机器人末端添加铰链约束, 并编写拖拽路径来模拟专家示教过程. 在策略验证执行阶段中, 将拖拽过程的关节角度序列作为控制指令发送给机器人的关节执行器, 驱动机器人完整复现抓取任务, 从而生成可用于模型训练的专家演示数据.

3.3 消融实验验证

本文设计 3 类任务用于消融实验模型评估. 第 1 类任务为专家任务复现, 用于验证模型复现专家演示任务的能力; 第 2 类任务为动态干扰任务, 用于验证模型应对动态目标位置变化的智能化水平; 第 3 类任务为长序列任务, 用于验证模型完成连续任务的能力. 每种模型进行 100 次验证实验, 使用相同的

随机种子确保各组目标物体点位保持相同。

3.3.1 实验设置与模型训练

为验证 GACT 模型中各创新组件的有效性,在“抓取-放置”任务上设计系统的消融实验,并从任务成功率与运动精度两个维度进行评估。对“抓取-放置”实验采集 50 次相同的专家演示数据并用于所有模型的训练。

在 MuJoCo 仿真环境中对模型的性能进行消融实验,将 3 种模型结构进行对比。ACT 模型为 Zhao 等^[20]提出的原始模型,该模型各关节动作信息都是连续信息,无法处理工业机器人离散气动夹爪状态,在这里作为基线模型;ACT w/BCE 模型增加一个独立的线性输出层和二元交叉熵损失函数,但缺少夹爪注意力机制层。GACT 模型为本文所提出模型,在对比模型的基础上,创新性地引入夹爪状态交叉注

意力机制,将夹爪状态作为独立模态输入解码器。

本文针对消融实验使用抓取位置平均误差来量化运动精度并衡量模仿学习任务的准确性。在本实验中利用 MuJoCo 仿真环境读取坐标的机制,记录抓取瞬间机器人末端执行器与目标物体在 XY 平面上的欧氏距离。

所有模型都基于同样的 50 次专家演示数据,在 RTX 2080 Ti 显卡上训练 16000 代。通过读取 GACT 模型解码器的注意力权重并对比,得到夹爪状态注意力权重比环境交叉注意力权重更高,初步验证了该注意力机制的有效性。

3.3.2 专家任务复现结果

针对第 1 类任务,设计了实验 1 用于比较每种模型复现“抓取-放置”任务的效果,实验结果如表 1 所示。

表1 实验 1 专家任务复现结果

模型结构	成功率/%	抓取目标总平均误差/m	抓取目标成功平均误差/m	抓取目标失败平均误差/m
ACT	91	0.0103	0.0094	0.0291
ACT w/BCE	94	0.0095	0.0082	0.0290
GACT	96	0.0096	0.0086	0.0342

由表 1 结果可知,GACT 模型具有最高的成功率。与基线 ACT 模型相比,GACT 与 ACT w/BCE 模型的抓取总平均误差均更低。这表明,在复现专家任务方面,GACT 模型表现出更优的性能,说明夹爪交叉注意力机制带来了积极效果,尽管在此相对简单的任务中提升幅度有限。

3.3.3 动态干扰任务结果

针对第 2 类任务,设计了实验 2 与实验 3 两种动态干扰方式:实验 2 为固定位置干扰,在机器人抓取过程中将目标物体移至目标区域中心的固定位置;实验 3 为随机位置干扰,在抓取过程中将目标物体移至目标区域内的随机位置,以测试模型的泛化与应变能力。

固定位置干扰任务实验结果如表 2 所示。由实

验结果可以看出,GACT 模型在该任务上同样具有最高的任务成功率,表明在应对确定性干扰时,GACT 模型凭借其对于夹爪状态的深层关注,更有效地调整策略,成功完成了再定位与抓取。针对抓取误差进行分析,基线 ACT 模型的抓取总平均误差最低,但其成功率也最低,表明其策略相对保守,即使在干扰后仍试图执行原轨迹,导致虽能接近新位置但最终抓取失败。而 GACT 模型在保持高成功率的同时,其成功抓取的平均误差与基线模型处于同一量级,表明其在成功案例中的定位精度较高。

随机位置干扰任务实验结果如表 3 所示,所有模型的成功率均出现预期中的下降,但 GACT 模型仍以 48% 的成功率显著优于基线模型。尤为关键的是,GACT 模型具有最小的抓取总平均误差。这一指

表2 实验 2 固定位置干扰结果

模型结构	成功率/%	抓取目标总平均误差/m	抓取目标成功平均误差/m	抓取目标失败平均误差/m
ACT	62	0.0093	0.0071	0.0131
ACT w/BCE	63	0.0127	0.0100	0.0175
GACT	69	0.0112	0.0094	0.0154

表3 实验 3 随机位置干扰结果

模型结构	成功率/%	抓取目标总平均误差/m	抓取目标成功平均误差/m	抓取目标失败平均误差/m
ACT	37	0.0211	0.0093	0.0284
ACT w/BCE	39	0.0228	0.0110	0.0307
GACT	48	0.0173	0.0128	0.0215

标表明,即使在任务失败的情况下,GACT模型驱动机器人到达的位置也更接近目标,其行为更加智能,体现出更强的环境认知能力。

3.3.4 长序列任务结果

针对第3类长序列任务,实验设置总时间步为单次任务的4倍,并在每次完成任务后自动重置目标物体位置,用于模拟真实工厂流水线物体抓取场景,比较在相同时间内完成抓取任务的平均抓取次数。结果如表4所示。

表4 长序列任务平均抓取次数

模型结构	平均抓取次数
ACT	2.7
ACT w/BCE	3.82
GACT	3.92

分析实验结果,在仅训练单次抓取任务的情况下,基于Transformer预测3种模型都能够通过分析环境状态的变化,实现重复的长序列抓取任务。GACT模型在单位时间内的平均抓取次数最高,接近理论最大值4次,且显著优于基线模型。这表明其能够明确预测和理解夹爪状态,可以更好地判断任务阶段并进行自主循环,从而胜任更长的序列任务。同时,表明GACT模型具有更强的对自身状态认知能力与重复长序列任务的完成能力。

3.3.5 消融实验总结

综上消融实验结果可知,引入离散夹爪预测是性能提升的关键,它直接解决了基线模型在工业场景中应用的根本缺陷;夹爪交叉注意力机制能带来进一步的智能增益,尤其在需要应对动态变化和长程规划的任务中,该机制通过增强模型对任务阶段的感知,显著提升了系统的鲁棒性和智能化水平。消融实验充分验证了GACT模型在工业机器人模仿学习中的应用潜力与价值。

3.4 前沿模型对比分析

为全面评估GACT模型的性能,将其与多种代表性模型进行对比实验,包括经典的深度学习模型(如CNN、LSTM)以及前沿的模仿学习模型(如BeT、Diffusion Policy)。所有模型均在相同的50次专家演示数据上进行训练,并针对工业场景中经常出现的“抓取-放置”与更复杂的“分拣-放置”任务进行测试。

“分拣-放置”任务模拟工业场景中普遍的分拣任务,场景包含一个红色目标物体及蓝、绿两个随机位置的干扰物体,如图6所示。任务要求机器人抓取目标物体,并将其准确放置于目标位置,对模型的感

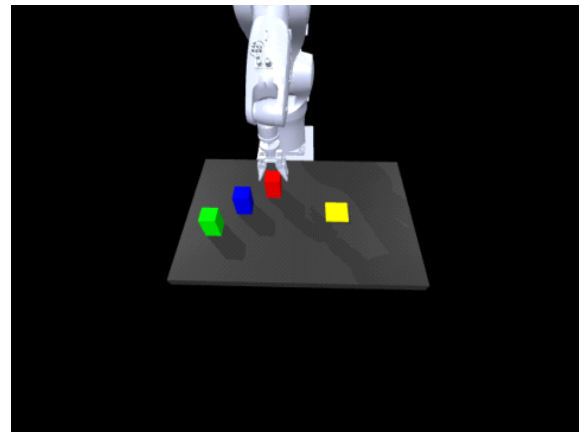


图6 “分拣-放置”任务仿真实验环境

知与泛化能力提出了更高要求。

任务成功率对比结果如表5所示。任务场景使用相同随机种子,针对“抓取-放置”任务与“分拣-放置”任务各进行100次任务评估。

表5 不同模型的任务成功率 %

模型结构	“抓取-放置”任务		“分拣-放置”任务	
	抓取成功率	放置成功率	分拣成功率	放置成功率
CNN	69	42	29	11
LSTM	51	28	32	10
BeT	78	55	56	26
DP	89	81	68	40
GACT	100	96	97	78

卷积神经网络(CNN)^[29]虽能有效处理图像空间特征,但其缺乏时序建模能力,难以捕捉动作序列的连贯性。在测试中常出现重复输出无效动作,导致系统“卡死”的现象,容易陷入局部行为模式,尤其在复杂的“分拣-放置”任务中成功率极低。

长短期记忆网络(LSTM)^[30]通过门控机制处理序列依赖,但其参数量大,在仅有50条演示数据的小样本场景下极易过拟合,无法学习到有效的泛化策略,导致性能不稳定,成功率相比CNN模型更低。

BeT^[17]模仿学习采用Transformer架构和k-means离散化动作空间。其缺陷在于,为追求动作精度需大量聚类中心,大幅增加计算负担与训练时间,且离散-连续的混合动作生成方式导致动作不平滑,影响实时控制性能与智能性。针对复杂的“分拣-放置”任务则需要更高的模型参数量,训练成本无法负担。

Diffusion Policy(DP)^[15],作为当前性能强大的生成式模型,它通过迭代去噪过程生成动作,在多模态策略表达上具有优势,其性能显著优于BeT与其他模型。然而,其多步迭代的采样过程导致推理速度较慢,难以满足工业机器人控制对高实时性的严苛要求,且针对复杂“分拣-放置”任务场景图像变化较大,

噪声不稳定,容易干扰模型的噪声输出,预测动作序列跳动较大。

实验结果表明,GACT模型在两项任务上的成功率均显著领先于所有对比模型。首先,GACT模型具有针对性的架构设计,GACT模型集成了Transformer优异的序列建模能力和CVAE的分布学习能力,尤其适合从小样本专家数据中学习鲁棒策略;其次,GACT模型具有高效的动作生成方式,基于Transformer解码器的自回归式动作分块预测,推理速度远快于Diffusion Policy的迭代采样,更能满足工业实时控制需求;最后,GACT模型具有强大的泛化能力,即使在复杂的“分拣-放置”任务中,GACT模型仍能有效忽略干扰物,精准定位目标,其接近100%的分拣成功率验证了其出色的感知与决策泛化能力。综上所述,GACT模型在保持前沿性能的同时,兼顾了高数据效率与高推理效率,在工业机器人模仿学习应用中展现出更强的实用价值与普适性。

3.5 实物实验验证

为评估GACT模型在真实工业环境下的性能,在实际KUKA机器人实验平台上进行实物实验。实验任务包括“抓取-放置”实验与“分拣-放置”实验,旨在验证模型从仿真到实物的迁移能力及其在真实场景中的有效性与实时性。对两项任务分别利用

示教机器人模型,采集50组专家演示数据,使用GACT模型进行训练。模型总参数量为91.27 MB,训练8000代。

“抓取-放置”任务1与“分拣-放置”任务2实物实验评估场景及过程如图7所示,蓝色物体为目标物体,红色与绿色物体为干扰物体,干扰物体放置位置接近目标物体且随机放置。

对两个实验分别进行10次评估,利用程序记录每个时间步模型的推理时间以及完成任务所需总任务时间,评估结果如表6所示。GACT模型在真实环境中能够有效完成任务,“抓取-放置”任务成功率达80%，“分拣-放置”任务成功率达60%。更重要的是,模型表现出优异的实时性能,平均单步推理时间稳定在30 ms以内,能够满足近30 Hz的机器人控制频率,且推理时间标准差极低,表明GACT模型能够为工业机器人提供高频、确定性的控制指令,满足实时控制对稳定性的苛刻要求。

针对实物实验中的失败场景进行分析有助于明确改进方向。在“抓取-放置”实验中,失败原因多为抓取姿态不佳导致物体在移动中掉落。这反映了在专家采集过程中,抓取姿态的多样性可能未被模型完全学习,未来可通过提升专家演示数据的质量来改善。在“分拣-放置”任务中,失败原因多为机械臂

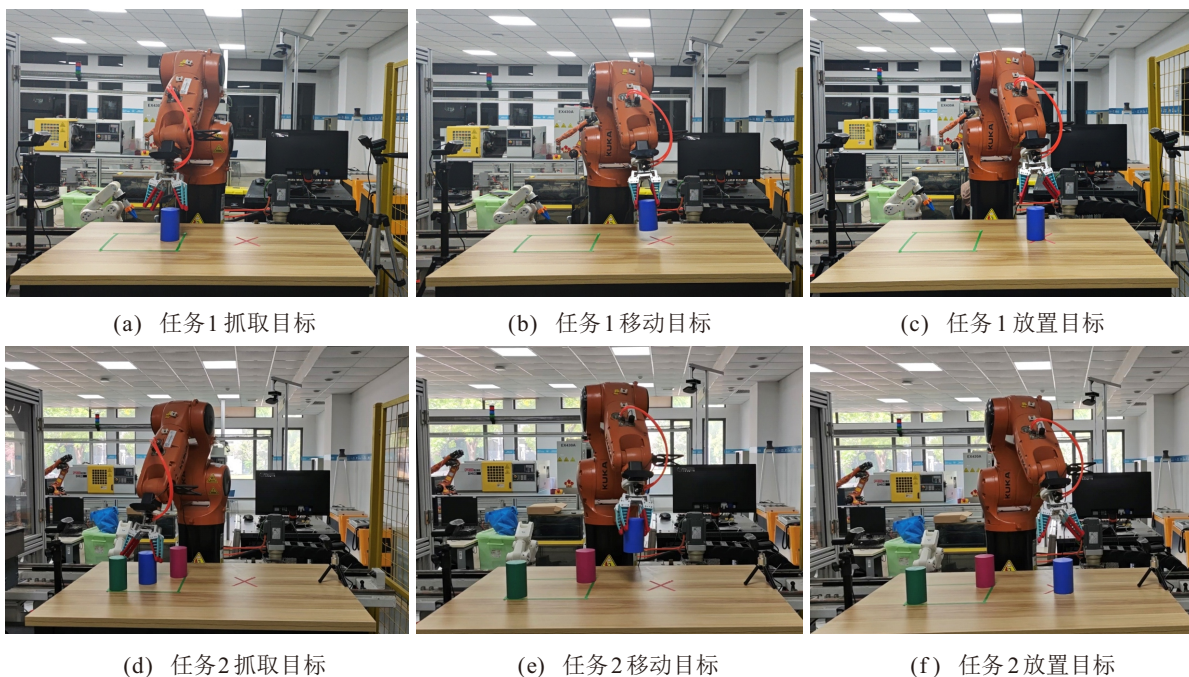


图7 KUKA 机器人实物实验过程

表6 实物实验评估结果

任务类型	成功率/%	平均推理时间/ms	推理时间标准差/ms	平均总任务时间/s
“抓取-放置”	80(8/10)	22.735	1.43	37.2
“分拣-放置”	60(6/10)	24.185	2.305	42.4

在分拣过程中与干扰物体发生冲突, 导致分拣过程失败. 这表明在小样本的复杂场景中, 模型规避动态干扰的能力仍需加强, 需要通过增加相关场景的专家数据来提升. 在验证过程中, 实验发现光照条件对模型预测结果影响较大, 这与模型高度依赖视觉输入有关, 在与数据采集过程的光照条件不同的环境中进行的实物实验误差较大, 导致整体成功率降低, 未来可以提升模型在各环境中的鲁棒性.

实物实验成功验证了 GACT 模型在真实工业机器人平台上的应用可能性. 模型在保证实时高频控制的前提下, 取得了具有说服力的任务成功率. 当前的主要挑战在于对复杂任务的泛化能力和环境鲁棒性, 这需要通过优化数据采集、增加数据质量等方法进一步改进. 结果表明, 本文提出的 GACT 模型为工业机器人的智能控制提供了一种有效的解决方案.

4 结论

本文针对工业机器人智能化控制的需求, 提出了一种基于 Transformer 的端到端模仿学习模型——GACT 模仿学习模型. 该模型设计了夹爪状态交叉注意力机制, 将离散的夹爪状态作为独立模态融入 Transformer 解码器, 有效提升了模型对任务阶段的理解和对混合模态动作空间的协同预测能力. 同时, 构建了一套基于遥操作的示教机器人系统, 显著降低了高质量专家演示数据的采集门槛与成本. 本文通过消融实验验证了 GACT 模型在面对动态干扰与长序列任务时, 表现出了更强的应变能力和决策智能, 凸显出夹爪状态交叉注意力机制在提升机器人智能化水平方面的关键作用; 通过系统的仿真实验对比, 在“抓取-放置”及更复杂的“分拣-放置”任务上, GACT 模型的任务成功率显著高于传统深度学习模型与前沿模仿学习模型, 表明了其性能的优越性; 通过实物实验验证了 GACT 模型能满足工业级实时控制要求, 且动作轨迹平滑稳定, 充分表明了模仿学习技术在真实工业场景中应用的可行性.

尽管 GACT 模型展现出良好的性能, 但也揭示了其在环境鲁棒性和复杂任务泛化能力方面的局限. 未来工作将聚焦提升模型对多任务和未知环境的泛化与适应能力; 探索数据增强与元学习等策略, 提高模型训练效率与鲁棒性; 研究模型在更复杂工业场景中的部署与优化, 推动其走向实际应用.

参考文献 (References)

- [1] 吴昊天, 王耀南, 朴玄斌, 等. 智能制造工业机器人技术应用及发展趋势[J]. *中国工程科学*, 2025, 27(3): 83-97.
- [2] Drost B, Ulrich M, Navab N, et al. Model globally, match locally: Efficient and robust 3D object recognition[C]. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, 2010: 998-1005.
- [3] 周游, 双丰, 李金科, 等. 基于视觉反馈的多机器人自重构系统研究[J]. *控制与决策*, 2022, 37(8): 2127-2133. (Zhou Y, Shuang F, Li J K, et al. Research on multi-robot self-reconfiguration system based on visual feedback[J]. *Control and Decision*, 2022, 37(8): 2127-2133.)
- [4] 陈钧, 宋薇, 周洋. 一种多模块神经网络与遗传算法相结合的单目位姿估计方法[J]. *机器人*, 2023(2): 187-196. (Chen J, Song W, Zhou Y. A monocular pose estimation method based on multi-module neural network and genetic algorithm[J]. *Robot*, 2023(2): 187-196.)
- [5] 王勇, 李邑灵, 苗夺谦, 等. 基于 Transformer-CNN 混合架构的跨模态融合抓取检测[J]. *控制与决策*, 2024, 39(11): 3607-3616. (Wang Y, Li Y L, Miao D Q, et al. Cross-modal fusion grab detection based on Transformer-CNN hybrid architecture[J]. *Control and Decision*, 2024, 39(11): 3607-3616.)
- [6] Sangiovanni B, Rendiniello A, Incremona G P, et al. Deep reinforcement learning for collision avoidance of robotic manipulators[C]. 2018 European Control Conference. Limassol, 2018: 2063-2068.
- [7] 柴浩轩, 金曦, 许驰, 等. 面向工业物联网的 5G 机器学习研究综述[J]. *信息与控制*, 2023, 52(3): 257-276. (Chai H X, Jin X, Xu C, et al. Overview of 5G machine learning for industrial Internet of Things[J]. *Information and Control*, 2023, 52(3): 257-276.)
- [8] 李晓豪, 郑海斌, 王雪柯, 等. 基于改进交叉熵的模仿学习鲁棒性增强方法[J]. *控制与决策*, 2024, 39(3): 768-776. (Li X H, Zheng H B, Wang X K, et al. Robustness enhancement method of imitation learning based on improved cross entropy[J]. *Control and Decision*, 2024, 39(3): 768-776.)
- [9] Pomerleau D A. Alvin: An autonomous land vehicle in a neural network[J]. *Advances in Neural Information Processing Systems*, 1988, 1: 305-313.
- [10] Abbeel P, Ng A Y. Apprenticeship learning via inverse reinforcement learning[C]. The 21st International Conference on Machine Learning. Banff, 2004: 1.
- [11] Ho J, Ermon S. Generative adversarial imitation learning[J/OL]. 2016: arXiv: 1606.03476.
- [12] 黄艳龙, 徐德, 谭民. 机器人运动轨迹的模仿学习综述[J]. *自动化学报*, 2022, 48(2): 315-334. (Huang Y L, Xu D, Tan M. Overview of imitation learning of robot trajectory[J]. *Acta Automatica Sinica*,

- 2022, 48(2): 315-334.)
- [13] Ross S, Gordon G J, Bagnell J A. A reduction of imitation learning and structured prediction to No-regret online learning[J/OL]. 2010, arXiv: 1011.0686.
- [14] Pari J, Shafiullah N M, Arunachalam S P, et al. The surprising effectiveness of representation learning for visual imitation[J/OL]. 2021, arXiv: 2112.01511.
- [15] Chi C, Xu Z J, Feng S Y, et al. Diffusion policy: Visuomotor policy learning via action diffusion[J/OL]. 2023, arXiv: 2303.04137.
- [16] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. *Advances in Neural Information Processing Systems*, 2017, 30: 5998-6008.
- [17] Shafiullah N M M, Cui Z J, Altanzaya A, et al. Behavior transformers: Cloning k modes with one stone[J/OL]. 2022, arXiv: 2206.11251.
- [18] Reed S, Zolna K, Parisotto E, et al. A generalist agent[J/OL]. 2022, arXiv: 2205.06175.
- [19] Brohan A, Brown N, Carbajal J, et al. RT-1: Robotics transformer for real-world control at scale[J/OL]. 2022, arXiv: 2212.06817.
- [20] Zhao T Z, Kumar V, Levine S, et al. Learning fine-grained bimanual manipulation with low-cost hardware[J/OL]. 2023, arXiv: 2304.13705.
- [21] Buamane T, Kobayashi M, Uranishi Y, et al. Bi-ACT: Bilateral control-based imitation learning via action chunking with transformer[J/OL]. 2024, arXiv: 2401.17698.
- [22] Li K L, Wagh S M, Sharma N, et al. Haptic-ACT: Bridging human intuition with compliant robotic manipulation via immersive VR[J/OL]. 2024, arXiv: 2409.11925.
- [23] Yang B C, Deng K Z, Peters C J, et al. Memorized action chunking with Transformers: Imitation learning for vision-based tissue surface scanning[J/OL]. 2024, arXiv: 2411.04050.
- [24] Rozo L, Calinon S, Caldwell D G, et al. Learning physical collaborative robot behaviors from human demonstrations[J]. *IEEE Transactions on Robotics*, 2016, 32(3): 513-527.
- [25] 李至, 潘越, 陈殿生, 等. 基于模仿学习的眼底手术行为机器人复现[J]. *机器人*, 2024, 46(3): 361-369. (Li Z, Pan Y, Chen D S, et al. Fundus surgical behavior reproduction of robot based on imitation learning[J]. *Robot*, 2024, 46(3): 361-369.)
- [26] Abu-Dakka F J, Nemeč B, Kramberger A, et al. Solving peg-in-hole tasks by human demonstration and exception strategies[J]. *Industrial Robot*, 2014, 41(6): 575-584.
- [27] 訾斌, 潘敬锋, 李元, 等. 智能喷涂机器人编程方法与遥操作技术研究综述[J]. *控制与决策*, 2025, 40(4): 1057-1078. (Zi B, Pan J F, Li Y, et al. Programming methods and teleoperation technology for intelligent spray-painting robot: A review[J]. *Control and Decision*, 2025, 40(4): 1057-1078.)
- [28] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2012, 60(6): 84-90.
- [29] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, 2016: 770-778.
- [30] Hochreiter S, Schmidhuber J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.

作者简介

朱易凡 (2001-), 男, 硕士生, 主要研究方向为机器人、智能控制, E-mail: 2330501@tongji.edu.cn;

刘晋飞 (1981-), 男, 副教授, 博士, 主要研究方向为分布式协同控制与决策、制造系统智能管理与优化, E-mail: jinfel@tongji.edu.cn;

黄华 (1992-), 男, 工程师, 硕士, 主要研究方向为智能制造、机器人控制, E-mail: huanghua0805@tongji.edu.cn;

陈明 (1964-), 男, 教授, 博士, 主要研究方向为智能制造, E-mail: chen.ming@tongji.edu.cn.