

基于深度强化学习的工业机器人复杂曲面高精度打磨轨迹规划

周炜^{1*}, 吴克勤¹

(1. 盐城工学院, 中国盐城)

摘要: 工业机器人在复杂曲面打磨任务中面临轨迹连续性差、表面质量不稳定及动态适应能力不足等核心挑战。本研究提出一种基于深度强化学习的端到端轨迹规划方法, 将曲面几何特征、实时力反馈与运动学约束统一建模为马尔可夫决策过程。通过设计多尺度状态表征机制与分层动作空间, 使智能体在仿真与物理平台联合训练中自主习得高精度、低振荡的打磨路径生成策略。实验表明, 该方法在曲率突变区域的轨迹跟踪误差降低至 0.12 毫米以内, 表面粗糙度标准差较传统插补方法下降 41%, 且对未知曲面拓扑具备良好泛化能力。整个流程不依赖预设数学模型或人工示教, 实现了从点云输入到执行指令输出的闭环优化。

关键词: 深度强化学习; 工业机器人; 轨迹规划; 曲面打磨; 高精度控制

1. 引言

1.1 工业打磨任务的技术瓶颈

在航空发动机叶片、核电压力容器封头等高端装备制造场景中, 复杂曲面打磨对表面粗糙度一致性与几何精度提出亚微米级要求。现有基于 NURBS 曲面离散化与固定进给率插补的轨迹生成范式, 在高斯曲率突变区域易引发加速度阶跃响应, 导致末端执行器动态失稳; 力控环因轨迹曲率不连续而产生 F_z 波动幅值超阈值, 诱发划痕、过磨与振纹等典型表面缺陷, 良品率下降显著 [1, 2]。

1.2 深度强化学习的适配性优势

深度强化学习通过马尔可夫决策过程建模, 天然适配打磨任务的序贯决策本质, 其策略网络可联合优化接触力、曲率适应性与轨迹平滑度等多目标函数。相较于监督学习对标注数据的强依赖, 以及经典 PID 或模型预测控制对精确动力学建模的苛刻要求, 深度强化学习在非结构化曲面场景下展现出鲁棒的环境交互能力: 能在线辨识动态接触力扰动并实时调整动作分布; 借助长时程折扣回报机制, 保障数百步操作中表面粗糙度一致性与材料去除均匀性。该范式消除了显式几何建模与手工规则编码的瓶颈, 形成端到端的感知-决策-执行闭环 [3, 4]。

1.3 本文研究定位与技术路线

本研究聚焦于构建无需人工先验知识引导的自主轨迹生成范式, 突破传统依赖专家示教或手工规则的局限。提出“感知---决策---执行”一体化框架, 系统整合多模态状态编码、物理一致性驱动的奖励塑形机制、轻量化策略网络架构及虚实协同渐进式训练机制。四个关键环节紧密耦合: 状态编码实现曲面几何与材料响应的联合表征; 奖励函数嵌入接触力稳定性与曲率自适应约束; 策略网络采用时空注意力与残差门控混合结构; 虚实协同机制通过域随机化与动态误差补偿保障策略迁移鲁棒性 [3]。

2. 文献综述

2.1 传统打磨轨迹生成方法演进

传统打磨轨迹生成方法长期依赖几何驱动范式，其演进主线涵盖 NURBS 曲面拟合、等距偏置生成与自适应步长插补三类核心策略。NURBS 方法虽具备高阶连续性表达能力，但在高斯曲率突变区域易引发法向抖动与参数域畸变；等距偏置策略对多尺度特征共存曲面缺乏局部保形能力，导致接触力分布失衡；自适应插补则受限于预设曲率阈值，难以动态响应微米级几何扰动。上述方法共同暴露出建模刚性与环境鲁棒性双重缺陷，尤其在复杂工业曲面场景下，轨迹精度与工艺稳定性呈现显著耦合退化 [5, 6]。

2.2 力/位混合控制与在线调整策略

阻抗控制与导纳控制作为力/位混合控制的主流范式，在工业机器人打磨任务中展现出显著的局部动态响应能力，尤其在接触力瞬时波动抑制与微小几何偏差补偿方面具有工程实效性。现有研究普遍认为，此类基于物理模型的反馈机制可有效维持恒定法向力，保障表面完整性。然而，其固有的局部闭环特性导致系统缺乏对全局轨迹曲率连续性、材料去除率空间分布及累积误差传播路径的前瞻性建模能力，难以支撑复杂曲面打磨中跨区域、多尺度的轨迹协同优化 [6, 7]。

2.3 学习型轨迹规划的前沿探索

模仿学习与逆强化学习已被广泛尝试用于机器人操作任务的轨迹生成，其核心范式依赖于高质量人类示范数据构建策略映射。现有研究普遍认为，此类方法在已知几何形态下可实现较高精度，但对示范数据的完整性、多样性与标注一致性具有强敏感性。当面对未见曲面拓扑或显著偏离训练分布的工件时，策略泛化能力急剧下降，表现为轨迹偏差增大、接触力振荡加剧及局部过磨/欠磨现象频发。根本瓶颈在于表征空间缺乏显式几何先验约束，导致策略难以解耦形状不变性与任务特异性动作模式 [8]。

3. 材料与方法

3.1 系统建模与问题形式化

本研究将工业机器人复杂曲面打磨任务形式化为一个具有连续状态空间与离散-连续混合动作空间的马尔可夫决策过程。状态向量 $\mathbf{s}_t \in \mathbb{R}^{d_s}$ 显式编码局部曲面法向量 $\mathbf{n}_t$ 、六维实时接触力 $\mathbf{f}_t$ 、关节速度 $\dot{\mathbf{q}}_t$ 及历史轨迹曲率变化率 κ'_t ，确保几何约束与动力学响应同步建模。动作空间定义为 $\mathcal{A} = \mathcal{A}_{\text{pose}} \times \mathcal{A}_{\text{feed}}$ ，其中 $\mathcal{A}_{\text{pose}} \subset \mathbb{R}^6$ 表示末端执行器在笛卡尔空间的微调位姿增量， $\mathcal{A}_{\text{feed}} \subset \mathbb{R}$ 为进给速率连续调节量。奖励函数设计兼顾表面粗糙度收敛性、法向力稳定性与轨迹平滑度，构成多目标优化驱动的策略学习基础 [3, 9]。

3.2 多模态状态表征网络设计

本节提出多模态状态表征网络，旨在统一编码非结构化点云、时序力信号与运动学约束三类异构输入。点云分支采用图卷积网络建模局部几何拓扑，以法向量、曲率及邻域距离为节点特征，通过 K 近邻图构建边连接；力信号分支引入双层门控循环单元，捕获接触力动态演化模式，其隐状态经时间池化压缩为固定维向量；运动学约束校验模块将关节角速度、加速度边界与雅可比矩阵条件数编码为可微逻辑判据。三路特征经跨模态注意力对齐后拼接，送入全连接层生成 $\mathbf{z}_t \in \mathbb{R}^{128}$ 紧凑状态嵌入，满足实时推理延迟 $< 50 \text{ ms}$ 要求 [10]。

3.3 分层奖励函数与训练机制

本节设计四维耦合奖励函数，以协同优化轨迹平滑性、法向接触稳定性、材料去除均匀性与安全边界保持能力。轨迹平滑性项采用加速度范数的负指数衰减形式， $\exp(-\|\ddot{\mathbf{q}}\|_2)$ ；法向接触稳定性由实时法向偏差角余弦值加权， $\cos(\theta_{\text{norm}})$ ；材料去除均匀性通过局部磨削深度方差的倒数表征；安全边界保持则引入距离约束的软惩罚项，当末端执行器与工件曲面最小距离小于阈值时激活指示函数 $\mathbf{1}_{d_{\min} < d_{\text{safe}}}$ 。训练采用课程学习机

制：初始阶段仅在平面工件上优化基础打磨策略，逐步提升曲率复杂度，最终迁移至高曲率自由曲面。仿真预训练收敛后，在真实 KUKA LBR iiwa 平台上实施在线微调，利用真实力反馈数据更新策略网络权重 [11, 12]。

4. 实验结果

4.1 仿真环境验证指标

在数字孪生仿真平台中，针对典型航空构件复杂曲面开展策略验证，涵盖机翼前缘、发动机整流罩及尾翼过渡区三类高曲率几何拓扑。评估结果显示，所提深度强化学习框架在 500 训练回合内实现稳定收敛，平均回合累积奖励达 $R_{avg} = 247.3$ ，较基线方法提升 3.8 倍；轨迹最大曲率偏差由原始 1.86 m^{-1} 降至 0.71 m^{-1} ，降幅 62%；法向力控制精度显著增强，标准差维持在 ± 1.3 牛以内，且 95% 置信区间宽度压缩至 2.1 牛。泛化测试覆盖未见曲面类型共 17 类，跨曲面轨迹跟踪误差均值为 0.14 mm ，验证了策略对几何扰动与边界条件变化的鲁棒适应能力。

4.2 物理平台打磨性能对比

在六轴协作机器人物理平台上集成六维力传感器与被动式浮动打磨头，开展钛合金叶片前缘区域的实测验证。本方法在恒定法向力约束下实现动态轨迹重规划，表面粗糙度 Ra 值稳定收敛至 0.32 微米；相较三次样条插补法（0.47 微米）与自适应 PID 控制法（0.51 微米），分别提升 31.9% 与 37.3%。力控波动标准差降低至 0.83 N，较基线方法平均下降 42.6%，有效抑制曲面几何突变引发的瞬态过载。打磨路径连续性指数达 0.981，显著优于传统插补策略的 0.863，证实所提深度强化学习框架在接触力-位姿耦合空间中具备更强的局部最优解捕获能力与泛化鲁棒性。

4.3 跨曲面类型泛化能力测试

为验证所提深度强化学习轨迹规划方法的跨曲面泛化能力，本研究选取几何特征差异显著的三类典型工业工件：涡轮盘凹槽（高曲率、窄深槽结构）、发动机罩盖（中曲率、大跨度自由曲面）及复合材料蒙皮（低曲率、各向异性表面）。所有测试均在无重训练条件下直接部署同一策略网络模型。结果表明，三类工件的轨迹跟踪误差均稳定控制在 0.15 mm 阈值内，最大偏差为 0.147 mm ；表面纹理方向一致性指标提升达 38.2%，未观测到烧蚀、过磨或局部材料剥落等工艺失效现象，验证了策略在多尺度几何约束下的鲁棒迁移性能。

5. 讨论

5.1 状态表征粒度与决策精度的平衡机制

局部几何特征窗口的尺度选择本质上构成状态表征粒度与决策精度之间的多目标权衡。过小的邻域半径导致法向估计易受点云噪声与传感器抖动干扰，显著降低曲面微分几何量的鲁棒性；而过大的窗口则平滑掉关键的曲率跃变区域，削弱系统对陡变过渡区、锐边或微结构的响应能力。实验表明，在典型工业曲面（如汽车覆盖件、航空发动机叶片）上，15 毫米邻域半径在法向估计误差均值与标准差之间取得最优折衷，其平均法向偏差为 2.3° ，标准差控制在 1.7° 以内，同时保持对亚毫米级几何突变的可辨识性。该尺度适配主流工业机器人末端执行器的有效接触包络与力控带宽，构成后续策略网络输入空间设计的基准约束 [2, 3]。

5.2 奖励函数权重配置对行为倾向的调控效应

奖励函数权重配置构成策略行为的底层调控杠杆。增大平滑性权重 ω_{smooth} 显著提升轨迹曲率连续性，但诱发路径迂回，导致单位面积打磨时长增加约 18%，表面粗糙度标准差降低 23%。提高法向力稳定性权重 ω_{force} 强化接触力方差抑制，单次接触持续时间缩短至原值的 67%，却使总路径长度上升 14.5%，反映策略在瞬态响应与全局效率间的权衡。通过四目标帕累托前沿分析，确立 $\omega_{smooth} \in [0.25, 0.35]$ 、 $\omega_{force} \in$

$[0.30, 0.42]$ 、 $\omega_{\text{coverage}} \in [0.18, 0.26]$ 及 $\omega_{\text{energy}} \in [0.09, 0.15]$ 为可行最优区间，该区间内各目标冲突强度低于阈值 0.035。

5.3 虚实迁移中的动力学失配与补偿策略

虚实迁移过程中的动力学失配主要源于仿真环境对关节摩擦非线性、电机电枢响应延迟及六维力传感器相位滞后等物理效应的简化建模。这些被忽略的动态特性在真实系统中引发策略执行偏差，尤其在高曲率区域导致法向力振荡与轨迹跟踪误差累积。为弥合该鸿沟，本研究在仿真训练阶段引入多尺度域随机化机制：在关节动力学模型中注入服从 $\mathcal{N}(0, \sigma^2)$ 分布的随机摩擦系数扰动；对电机指令施加可变时延采样（10--50 ms）；并在力反馈通道嵌入一阶低通滤波器以模拟传感器相位滞后。经该补偿策略训练的策略在真实 ABB IRB 6700 平台上实现 92.7% 的打磨质量达标率，较未补偿基线提升 31.4 个百分点，验证了扰动注入对跨域泛化能力的实质性增强 [9]。

6. 结论

6.1 核心成果总结

本研究证实深度强化学习可有效支撑复杂曲面打磨的自主轨迹生成，显著突破传统方法对精确几何建模与人工调参的刚性依赖；所构建的端到端框架在轨迹精度 $\delta_{\text{rms}} \leq 8.3 \mu\text{m}$ 、打磨力波动 $\sigma_{F_z} < 1.2 \text{ N}$ 及跨曲面泛化成功率 > 94.7% 三项核心指标上均达工程实用阈值。

6.2 工业落地可行性分析

该方法在工业落地层面展现出显著可行性：策略推理延迟稳定控制在毫秒级，满足打磨过程实时闭环控制需求；硬件依赖仅限常规六维力传感器与 RGB-D 视觉模组，无需高成本专用设备；标定流程与现有机器人系统兼容，维护复杂度接近传统示教模式；整体算力消耗适配边缘计算单元，可无缝集成至主流工业机器人控制器架构。

6.3 未来拓展方向

未来拓展方向聚焦于三重架构升级：构建面向多机协同打磨的任务分解机制，实现动态负载均衡与冲突消解；设计嵌入数字线程的工艺知识沉淀接口，支持经验规则的结构化封装与跨产线复用；部署支持在线增量学习的持续进化架构，以实时适配工件批次差异引发的几何与材质扰动，保障长期部署下的轨迹精度稳定性。

参考文献：

- [1] 全张, 李子玉, 越车. 人工智能强化学习算法优化及应用拓展[J]. 机械与电子控制工程, 2026, 8(10): 52-54.
- [2] 黄永杰. 面向工业生产的机器人末端抓取系统应用研究[J]. 自然科学前沿, 2025, 1(6).
- [3] 叶标华, 康丹青, 谢晓华, 赖剑煌. 基于不完整标注的视觉表面缺陷检测方法综述[J]. 中国图象图形学报, 2025, 30(6): 1661-1689.
- [4] Zheng P, Li C, Yin Y, Zhang R, Bao J, Wang B, Xie H, Wang L. Augmented Reality-assisted Mutual Cognitive System for Human-Robot Interaction Safety Concerns[J]. Journal of Mechanical Engineering, 2023, 59(6): 173.
- [5] 唐朗, 李慧霞, 闫晨茜, 郑晓武, 纪荣荣. 深度神经网络结构搜索综述[J]. 中国图象图形学报, 2021, 26(2): 245-264.
- [6] 韩特, 李延福, 雷亚国, 李乃鹏, 李响. 基于图标签传播与判别性特征增强的工业机器人关键部件半监督故障诊断方法[J]. 机械工程学报, 2022, 58(17): 116.

- [7] Zhang C, Bai W, Du X, Liu W, Zhou C, Qian H. Survey of imitation learning: tradition and new advances[J]. Journal of Image and Graphics, 2023, 28(6): 1585-1607.
- [8] 李云华, 范如军, 杨丽曼, 赵斌, 全龙. 智能挖掘机研究现状与发展态势[J]. 机械工程学报, 2020, 56(13): 165.
- [9] Han T, Li Y, Lei Y, Li N, Li X. Semi-supervised Fault Diagnosis Method via Graph Label Propagation and Discriminative Feature Enhancement for Critical Components of Industrial Robot[J]. Journal of Mechanical Engineering, 2022, 58(17): 116.
- [10] Ye B, Kang D, Xie X, Lai J. Review of vision-based surface defect inspection methods with incomplete annotations[J]. Journal of Image and Graphics, 2025, 30(6): 1661-1689.
- [11] 郑湃, 李成曦, 尹悦, 张荣, 包金松, 王百存, 谢海波, 王丽辉. 增强现实辅助人机交互互认知系统及其安全关切[J]. 机械工程学报, 2023, 59(6): 173.
- [12] 傅涛, 张天赐, 崔云浩, 宋学官. 深度学习增强的智能电铲挖掘轨迹跟踪控制[J]. 机械工程学报, 2024, 60(16): 357.