arXiv:2602.01092v1 [cs.RO] 1 Feb 2026

通过保守价值引导辅助实现故障感知双手遥操作

Peng Zhou*，*IEEE*会员，Zhongxuan Li*，Jinsong Wu，Jiaming Qi，Jun Hu，David Navarro-Alarcon，*IEEE*高级会员，Jia Pan，*IEEE*高级会员，Lihua Xie，*IEEE*会士，Shiyao Zhang†，*IEEE*高级会员和Zeqing Zhang†，*IEEE*会员

摘要—高精度操作的遥操作受到严格成功容差和复杂接触动力学的制约，这使得人类操作员在部分可观测条件下难以预判即将发生的失败。本文提出了一种用于双手遥操作的价值引导的故障感知框架，在持续保留人类控制权的同时提供柔顺触觉辅助。该框架完全基于异构离线遥操作数据进行训练，其中包含成功与失败的执行。任务可行性通过保守价值学习建模为保守成功评分，从而得到一种在分布偏移下仍保持可靠性的风险敏感估计。在在线操作过程中，学习得到的成功分数调节辅助程度，而学习得到的执行器提供纠正运动方向。二者通过主端的关节空间阻抗接口进行集成，从而实现持续引导，使操作员避开易导致失败的动作，同时不覆盖其意图。高接触操作任务上的实验结果表明，与传统遥操作和共享自主基线方法相比，该方法提高了任务成功率并降低了操作员工作负荷，说明保守价值学习为将失败感知嵌入双边遥操作提供了一种有效机制。实验视频见https://www.youtube.com/watch?v=XDTsvzEkDRE

索引词—精细操作；共享自主；双边遥操作；机器人学；具身人工智能。

一、引言

遥操作是收集大规模、高保真人类演示数据，以及在危险或无法进入的环境中执行机器人操作的关键基础设施。随着模仿等数据驱动方法

本研究得到国家自然科学基金委员会（NSFC）项目（编号：62403211）资助，并部分得到广东省科学技术协会青年人才支持计划资助，技术（GDSTA）资助，资助编号为 SKXRC2025092；此外，还获得广东省区域联合基础与应用基础研究基金资助（编号：2024A1515110203）。（通讯作者：*Shiyao Zhang*、*Zeqing Zhang*）

* 表示作者对本文的贡献相同。

P. Zhou、J. Hu 和 S. Zhang 任职于大湾区大学先进工程学院，地址：中国广东省（电子邮箱：pzhou@gbu.edu.cn、hujun@gbu.edu.cn、zhangshiyao@gbu.edu.cn）。

Z. Li 和 J. Pan 任职于香港大学计算与数据科学学院，地址：中国香港（电子邮箱：zxli01@connect.hku.hk、jpan@cs.hku.hk）。

J. Wu 和 D. Navarro-Alarcon 任职于香港九龙香港理工大学机械工程系（电子邮箱：jinsong.wu@connect.polyu.hk,dnavar@polyu.edu.hk）。

J. Qi 任职于中国黑龙江东北林业大学机电工程学院（电子邮箱：jiamingqi@nefu. edu.cn）。

L. Xie 和 Z. Zhang 就职于新加坡南洋理工大学电气与电子工程学院（电子邮箱：elhxie@ ntu.edu.sg，zeqing.zhang@ntu.edu.sg）。

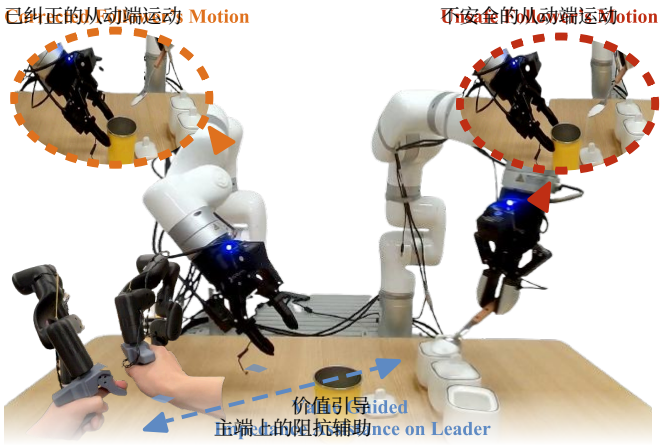


图 1. 我们的故障感知遥操作框架利用保守成功评分学习，在主臂上施加纠正力矩，引导操作员远离分布外（OOD）行为。

学习和离线强化学习日益依赖多样且精确的遥操作数据来掌握灵巧的操作技能，而这些轨迹的质量和连续性在决定策略性能方面发挥着核心作用 [1]。与此同时，对于无法进行直接人工干预的安全关键任务，遥操作仍是唯一可行的解决方案。综合而言，尽管部分可观测性、复杂的接触动力学以及通信延迟等挑战持续存在，这些需求仍要求对精细的物理交互进行精细控制。因此，亟需开发先进的遥操作系统，以便在不可逆故障发生之前预测并缓解此类故障。

大量研究 [2], [3] 致力于通过增强通信稳定性以及提供高保真力觉和位置反馈来提升遥操作性能，从而实现双边透明性。尽管这些进展提高了信号传输精度，但信息通常仅被视为被动反馈，很少用于推理任务上下文或失败风险，因此大多数系统仍依赖人类操作员来弥补不确定性。在接触密集型操作中，有效的遥操作不仅需要忠实传输信号，还需要机器人端具备运动层智能，以便结合本地物理环境解读指令。缺乏这种能力，就无法主动纠正细微的执行偏差，导致偏差不断累积，最终造成不可逆失效。此外，通信延迟和部分可观测性会持续造成感知差距，使得单纯依赖人类

进行补偿不足以实现稳健性能。

近期基于学习的范式 [4] 展现了这样的潜力：使机器人能够掌握从仿真或大规模数据中学习到的灵巧操作技能，同时允许人类在执行过程中提供高层任务指导。在这些框架中，机器人依据学习策略自主生成低层运动，而人类主要通过粗粒度的指令调节或任务级意图来影响其行为。然而，学习策略的运行在很大程度上独立于操作员的连续控制回路，因此无法向人类提供关于局部执行质量或潜在失败风险的充分反馈。因此，在遥操作过程中，机器人自主动作与操作员预期之间的差异可能不断累积，导致协调性下降，并使精细操作对接触引起的偏差更加敏感。

与此同时，共享自主框架 [5]引入了人工智能，以在任务层面提供辅助，例如进行意图推断或策略规划。然而，这些方法通常处于抽象层面，缺乏在精细、富接触交互过程中调节细微运动所需的控制分辨率。因此，现有的遥操作和共享自主系统往往要等到故障显现后才作出响应，使人机团队容易遭受本可通过更早的运动级干预加以避免的不可逆后果。

为应对这些挑战，我们提出一种框架，将失败感知与在线动作执行明确解耦。其核心思想是从离线遥操作数据中推断出对任务可行性的保守估计，用于评估给定的状态—动作对是否可能始终处于任务的成功一致区域内。我们不直接修改或覆盖人类指令，而是以尽可能少干预的方式，利用这一可行性估计调节交互。

我们的主要贡献总结如下：

- 我们提出一种从异构离线遥操作数据中学习的失败感知成功分数，为任务可行性提供保守估计，并且在操作员引起的分布偏移下仍保持可靠。
- 我们提出成功引导的辅助范式，利用学习得到的成功分数调节交互强度，同时保持人类连续控制，避免离散自主切换。
- 我们通过双边遥操作接口实现这种辅助：机器人以柔顺且具有物理依据的触觉反馈呈现其纠正意图，引导人类指令趋向成功一致的执行，同时不夺取操作员的控制权。

通过这一表述，机器人以预测性失败感知补充人类控制，从而在富接触领域实现更安全、更可靠的遥操作，并支持面向基于学习的操作的可扩展数据收集。

II. 相关工作

A. 双手遥操作系统

双手遥操作系统通过利用直接的人类控制，实现了高自由度任务中的灵巧操作，并广泛应用于接触密集型场景，在这些场景中，完全自主的策略仍不可靠。一个典型示例是 ALOHA [6], , 这是一种低成本、开源的双手遥操作平台，通过主从接口实现协调的双臂控制。多项研究致力于提升遥操作接口的易用性和反馈机制。GELLO [3] 引入了运动学匹配的操作员接口，以提升控制直观性和示范质量；而 FACTR [2] 通过反馈从端电机电流来传达接触信息，但缺乏主端的阻抗调节或运动级稳定功能，因此在接触条件快速变化时反馈有限。与此同时，近期基于学习的方法探索将遥操作用作预训练控制器的高层意图接口。例如，DexterityGen [7] 利用大规模仿真和强化学习习得灵巧操作技能，并由人工输入提供粗略引导。然而，在这些系统中，精细操作主要由机器人自主执行，任务执行期间持续的人机协调或双向反馈有限。总体而言，现有双手遥操作系统凸显了人类控制在灵巧操作和数据收集中的价值，同时也暴露出将机器人侧智能与基于物理的反馈整合到遥操作回路中的局限性。

B. 故障感知共享自主与遥操作

共享自主（SA） [5], [8] 已得到广泛研究，旨在通过融合操作员输入与自主决策来减轻人类的认知负担。早期的共享自主框架通常将该问题建模为部分可观测马尔可夫决策过程（POMDP） [9], , 以便在感知不确定性下进行意图推断并提供目标导向辅助 [10]。这些方法虽擅长高层推理，但通常依赖简化的任务抽象，对细粒度运动调节的支持有限。随着高接触操作对安全性的要求不断提高，近期研究着重采用安全关键控制滤波器 [11], [12] , 在不完全覆盖人类控制 [13], [14]的前提下，将系统行为限制在预定义的安全集内。尽管这些方法能够提供形式化的安全保证，但它们大多是被动响应式的，且需要显式的约束建模，这对于复杂的接触动力学而言十分困难。为提升适应性，后续研究将人类干预纳入模仿学习流程。人机联合学习 [15] 及其变体 [16]–[18] 等方法允许专家在线纠正机器人的行为，从而缓解协变量偏移，但主要针对已发生的错误，而非预判失败。近期，基于基础模型的方法开始探索利用视觉语言模型（VLM）提供任务级语义引导 [19]。然而，这些方法在较粗的时间和空间分辨率下运行，无法捕捉局部接触动力学。因此，尽管已取得显著进展，大多数现有共享自主方法仍以被动响应或任务级引导为主，缺乏对运动层面的理解以及对即将发生的、往往不可逆的失败（如卡滞或错位）的预测性反馈。

C. 用于策略改进的离线强化学习

离线强化学习（RL） [20], [21] 无需在线探索，即可利用固定数据集改进策略。隐式Q学习（IQL） [22], [23] 等方法通过避免对未见动作进行显式最大化来提升离线稳定性；保守Q学习（CQL） [24] 则进一步通过抑制分布外动作的Q值来应对分布偏移，从而在富接触场景中获得更可靠的价值估计。尽管离线强化学习在自主操作方面不断取得进展，包括 $\pi^*_{0.6}$ [25]和RL100 [26], 等大规模研究，但现有方法大多针对轨迹级优化和自主执行 [27]而设计。因此，这些方法对共享自主和交互式遥操作的支持有限，因为学习策略很少能以人类操作员易于理解的形式呈现可行性或不确定性信息 [28]。这凸显了既有研究的一项关键局限：虽然离线强化学习能够从数据中捕捉任务成功结构，却很少以保留持续人类控制权 [29]的方式融入人机控制回路。要弥补这一不足，就需要将从离线数据中学习到的失败感知与直接动作执行相分离，转而通过可解释且具有物理依据的引导信号传达这些信息，以支持人类决策。

III. 问题表述

我们研究由人类操作主设备、从机器人与环境交互的双边遥操作。人类发出指令，机器人则在部分可观测性、接触不确定性和延迟的条件下执行指令。我们的目标是提供失败感知辅助，在保留人类控制权的同时减少不可逆失败。

A. 将遥操作建模为失败感知的部分可观测马尔可夫决策过程

我们将交互过程建模为一个失败感知的部分可观测马尔可夫决策过程（FA-POMDP）。

$$\mathcal{M} = \langle \mathcal{X}, \mathcal{A}, \mathcal{S}, \Omega, \mathcal{P}, \mathcal{R} \rangle, \tag{1}$$

其中， $\mathcal{X}$ 表示机器人操作系统的底层物理动力学，由其状态空间表示，包括物体位姿、接触模式和摩擦参数。 $\mathcal{A}$ 表示从机器人上的速度指令 a_{tele} 。这些观测由底层系统状态通过随机转移模型生成

$$x_{t+1} \sim \mathcal{P}(x_{t+1} \mid x_t, a_t), \tag{2}$$

其刻画了混合接触动力学以及物体-机器人交互。

鉴于遥操作具有部分可观测性，在每个时间步，人类操作员都会感知一个状态 $s_t \in \mathcal{S}$，该状态由主从关节速度 q_t 和视觉观测 o_t组成。奖励函数 R： $\mathcal{S} \times \mathcal{A} \to \mathbb{R}$ 提供了一个失败感知监督信号，用于表征任务结果。我们不依赖人工奖励工程，而是根据轨迹的终止状态，将二值信号 $r_t \in \{+1, -1\}$ 广播至整条轨迹。具体而言， $r_t = +1$表示系统仍处于与成功相一致的区域内，而 $r_t = -1$表示系统正在转向不可逆失效，或转向需要人工干预的状态。

B. 总体问题目标

本研究旨在从异构离线遥操作数据中学习一个保守的可行性估计器 Q ，并据此构建辅助策略 π_{assist} ，以主动降低失败风险，同时保留持续人类控制权。

该目标被表述为一个约束优化问题：

$$\min_{Q,\lambda} \quad \mathbb{E}_{\pi_{\text{assist}}} \left[\sum_{t=0}^{\infty} \gamma_{\text{risk}}^t \, \mathbb{I}[x_t \in \mathcal{F}_{\text{failure}}] \right], \tag{3}$$

满足最小侵入性约束，

$$\mathbb{E}_{s_t \sim d^{\pi_{\text{assist}}}} \left[\left| a_t - a_t^{\text{tele}} \right| \right] \leq \epsilon, \tag{4}$$

并满足门控透明条件，

$$\lambda_t = 0 \quad \text{if} \quad Q(s_t, a_t^{\text{tele}}) \geq \tau. \tag{5}$$

辅助策略 π_{assist} 由指令混合机制产生。在该表述中， $\lambda_t \in [0, 1]$ 表示一个随时间变化的干预权重，用于调节机器人辅助在执行指令中的影响。它并非一个自由控制变量，而是由学习得到的可行性评分 Q决定的价值门控量 λ_t ，反映了在操作员当前指令下对近期失败风险的估计。

直观而言，当评估认为操作者指令是安全的（ Q 较高且预测风险较低）时， $\lambda_t \approx 0$ 系统的行为保持透明。随着预测风险升高， λ_t会平滑增长，激活成功引导辅助，使执行偏向更安全的动作，同时不凌驾于人类意图之上。

通过引入专门的失效集 $\mathcal{F}_{\text{failure}}$，我们明确地对遥操作失效的累积性质进行建模。这类失效通常源于隐藏接触动力学下的多步状态演化，而非孤立的错误动作。进入 $\mathcal{F}_{\text{failure}}$ 意味着发生不可逆的终止事件，并触发人类操作员进行系统重置。因此，我们的目标是识别安全集 $\mathcal{F}_{\text{safe}}$ 与失效集 $\mathcal{F}_{\text{failure}}$之间的决策边界，从而及早检测迫近风险并降低系统重置的概率。

所提出的框架由三个关键组件构成：**(1)** 通过离线遥操作数据学习得到的保守成功评分估计器 ，用于刻画分布偏移下的可行性；**(2)** 成功引导执行器 ，提供与所学可行性景观中更安全区域相一致的在线纠正运动方向；以及 **(3)** 价值引导阻抗辅助模块 ，在主操作臂上以透明方式运行，提供基于物理原理的触觉反馈，引导操作员远离易导致失效的操作，同时保留人类连续控制。

IV. 方法

本节介绍成功引导遥操作辅助的整体系统架构和学习框架。我们首先介绍离线学习流程，说明如何利用异构遥操作数据训练保守可行性估计器。随后，我们描述实时推理

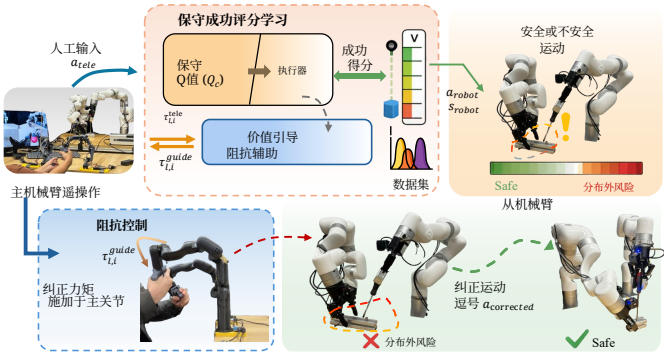


图 2. **所提出的价值引导共享自主框架概览**。上半部分：利用带有二元成功与失败标签的异构遥操作数据进行离线学习，以训练保守成功评分（评论家）。下半部分：在线部署时，成功分数决定何时需要干预，而训练得到的执行器则提供纠正性引导动作。两者均整合在共享自主框架中，引导操作员避开分布外动作和易导致失败的动作，同时保留人类连续控制。

和决策过程，该过程将学习到的成功信号集成到双边遥操作系统中，从而实现风险感知的指令调节和主动干预。最后，我们详细介绍控制接口和耦合机制，以确保实现无缝、透明且符合物理规律的人机交互。

A. 离线遥操作数据集收集

所有实验均使用仅通过双边遥操作收集的离线数据集，不包含任何自主辅助。数据由专家操作员和新手用户共同收集，以捕捉多样的执行风格、纠正行为以及具有代表性的失败模式。操作员通过主设备在关节速度空间中控制从机器人，并从多视角 RGB 相机配置中获取视觉反馈，其中包括外部前置、左侧和右侧相机，以及安装在两个末端执行器上的腕部相机。此配置既能捕捉全局场景上下文，也能捕捉细粒度的接触交互。

我们假设可以访问异构离线遥操作数据

$$\mathcal{D} = \mathcal{D}_{\text{succ}} \cup \mathcal{D}_{\text{fail}}, \tag{6}$$

其中， $\mathcal{D}_{\text{succ}}$ 包含成功执行的轨迹， $\mathcal{D}_{\text{fail}}$ 包含导致不可逆失败或近失败事件的轨迹。

每个遥操作回合最长为 30 秒，并在任务完成、明确指示失败或超时后终止，从而生成适用于离线学习的时间紧凑型轨迹。每条轨迹表示为

$$\tau = \{(s_t, a_t^{\text{tele}}, s_{t+1})\}_{t=0}^{T}, \tag{7}$$

其中， s_t 表示观测到的系统状态， a_t^{tele}表示人类操作员发出的关节速度指令。数据收集期间不会施加纠正动作或自主动作。

我们不对任务效率或渐进式进展进行编码，而是采用一种基于轨迹级结果的失败感知监督信号。执行结束后，每条轨迹都会被标记为成功或失败，并将相应的二元标签统一赋予所有时间步：

$$r_t = \begin{cases} +1, & \tau \in \mathcal{T}_{\text{success}}, \\ -1, & \tau \in \mathcal{T}_{\text{failure}}. \end{cases} \tag{8}$$

我们将离线数据集记为

$$\mathcal{D} = \{\tau^{(n)}\}_{n=1}^{N}, \qquad \tau^{(n)} = \{(s_t^{(n)}, a_t^{\text{tele},(n)}, s_{t+1}^{(n)})\}_{t=0}^{T_n-1}, \tag{9}$$

其中 $s_t \in \mathcal{S}$ 为观测状态， $a_t^{\text{tele}} \in \mathcal{A}$ 为人类发出的关节速度指令。所得数据集是下一节中学习任务相关价值函数的唯一监督来源。

B. 保守成功评分学习

我们学习一个保守成功评分 $Q(s, a)$，将其作为遥操作指令的近期可行性估计器。该评分定义为期望累积折扣奖励：

$$Q(s_t, a_t) = \mathbb{E}\left[\sum_{i=0}^{\infty} \gamma^i r_{t+i} \,\middle|\, s_t, a_t\right], \tag{10}$$

其中逐步监督信号 $r_t \in \{+1, -1\}$ 根据轨迹的终止结果进行赋值。

$Q(s_t, a_t^{\text{tele}})$取值较低表明，当前状态–动作对要么与离线数据集中的失败执行相关，要么偏离成功演示的支撑集，从而预示近期失效风险升高。据此，我们通过基于阈值的规则定义遥操作指令的可行性：

$$(s_t, a_t^{\text{tele}}) \in \begin{cases} \mathcal{F}_{\text{feasible}}, & \text{if } Q(s_t, a_t^{\text{tele}}) \geq \tau, \\ \mathcal{F}_{\text{failure}}, & \text{if } Q(s_t, a_t^{\text{tele}}) < \tau, \end{cases} \tag{11}$$

其中， τ 是根据成功轨迹的经验分布确定的安全阈值。这一表述能够主动识别即将发生的失败，并支持在接触丰富的遥操作场景中及时实施风险感知干预。

我们的目标是将抽象的可行性指示函数 $Q(s, a)$ 实例化为一种可学习、可靠且保守的成功评分，并从离线遥操作数据中学习得到。所学得的评分必须能够在分布偏移情况下校准地估计近期失效风险，并支持遥操作过程中的实时干预门控。

我们采用保守Q学习（CQL）作为离线强化学习框架，从固定遥操作数据集中学习这类可行性估计器。CQL 专为缓解分布外（OOD）动作上的过高估计而设计，其方法是施加悲观价值结构；这对于安全敏感的遥操作场景至关重要。

我们将成功分数参数化为一个动作价值函数

$$Q_\phi(s, a) : \mathcal{S} \times \mathcal{A} \to \mathbb{R}, \tag{12}$$

其中 ϕ表示网络参数。不同于传统的强化学习（其价值函数经过优化以最大化期望回报），此处 Q_ϕ 被解释为一种

可行性势，用于衡量在状态 s 下执行动作 a 后仍处于与成功一致的任务区域内的统计似然。较低的数值表示接近不可逆失效模式，而较高的数值则对应于成功遥操作数据所支持的动作。

评论家通过最小化保守Q学习目标进行训练

$$\mathcal{L}_{\text{critic}}(\phi) = \frac{1}{2}\mathbb{E}_{(s,a,s')\sim\mathcal{D}}\Big[\big(Q_\phi(s,a) - \hat{\mathcal{B}}Q_{\bar{\phi}}(s,a)\big)^2\Big] + \alpha\,\mathbb{E}_{s\sim\mathcal{D}}\Big[\log\sum_a \exp(Q_\phi(s,a)) - \mathbb{E}_{a\sim\mathcal{D}}Q_\phi(s,a)\Big] \tag{13}$$

其中，$\hat{\mathcal{B}}Q_{\bar{\phi}}$ 表示使用参数为 ϕ的目标网络计算得到的贝尔曼备份，而 $\alpha > 0$ 用于控制保守正则化的强度。第二项明确惩罚被赋予数据集离线支持集之外的动作的高价值，从而强制

$$Q_\phi(s, a_{\text{OOD}}) \le Q_\phi(s, a_{\text{data}}), \tag{14}$$

并由此防止分布偏移下的乐观外推。

为使学习到的决策边界在易失效区域附近更加清晰，我们为评论家增加了一个辅助头，用于预测未来 H 步内是否会发生不可逆失效。给定观测–动作对（s_t，a_t），辅助头会输出一个失效概率

$$\hat{p}_t^{(H)} = \sigma(g_\eta(s_t, a_t)), \tag{15}$$

其中，g_η 是一个与评论家骨干共享参数的轻量级预测器。相应的短时域失效标签定义为

$$y_t^{(H)} = \mathbb{I}\left[\bigcup_{k=1}^{H}\{x_{t+k} \in \mathcal{F}_{\text{failure}}\}\right]. \tag{16}$$

辅助头使用二元交叉熵损失进行训练，

$$\mathcal{L}_{\text{fail}}(\eta) = \mathbb{E}_{(s_t,a_t)\sim\mathcal{D}}\big[\text{BCE}\big(y_t^{(H)}, \hat{p}_t^{(H)}\big)\big]. \tag{17}$$

这种短时域监督与轨迹级结果相互补充，在不改变底层离线强化学习目标的情况下，提高了保守成功评分在安全–失效边界附近的敏感性。

总体而言，评论家通过最小化组合损失进行训练

$$\mathcal{L}_{\text{critic}} = \mathcal{L}_{\text{CQL}} + \lambda_{\text{fail}}\,\mathcal{L}_{\text{fail}}, \tag{18}$$

其中，$\mathcal{L}_{\text{CQL}}$ 在成功分数上强制实施保守贝尔曼一致性，$\mathcal{L}_{\text{fail}}$ 提供短时域失效监督，λ_{fail} 用于平衡这两项损失。

训练完成后，将评论家的原始输出归一化为有界的成功指标

$$\tilde{Q}(s,a) = \text{clip}\left(\frac{Q_\phi(s,a) - Q_{\min}}{Q_{\max} - Q_{\min}}, 0, 1\right), \tag{19}$$

其中，$Q_{\min}$ 和 $Q_{\max}$ 是根据训练集上 Q_ϕ 的经验分布计算得出的。实际上，$Q(s,\ a) \in [0,1]$ 提供了一个无量纲的可行性裕度，用于在遥操作控制回路中进行基于价值的干预门控和引导调制。由此得到的保守成功景观呈现出与成功演示相一致的高分盆地，并在易发生失败的区域附近具有陡峭梯度，从而能够根据部分观测可靠地评估近期风险。

C. 演员策略作为成功引导辅助

给定当前观测 s_t，执行器会输出一个机器人建议的指令

$$a_t^{\text{assist}} = \pi_\theta(s_t), \qquad \pi_\theta : \mathcal{S} \to \mathcal{A}. \tag{20}$$

我们训练 π_θ，使其在 Q_ϕ条件下提出具有较高保守成功评分的动作，同时保持平滑并处于驱动限制范围内。演员的目标函数定义为：

$$\mathcal{L}_\pi(\theta) = \mathbb{E}_{s\sim\mathcal{D}}\Big[-Q_\phi\big(s, \pi_\theta(s)\big) + \|\pi_\theta(s) - a^{\text{tele}}\|^2\Big], \tag{21}$$

其中 $\beta_\pi > 0$ 对幅值进行正则化（并可扩展以纳入加加速度/平滑性惩罚）。该更新将策略导向能够在局部降低预测失败风险的动作，同时在训练过程中保持与人工操作指令的接近性。

从控制理论视角来看，执行器可被解释为在学习一种沿类李雅普诺夫可行性势函数负梯度方向的反馈控制律，

$$a^{\text{assist}}(s) \;\approx\; -K\,\nabla_a Q_\phi(s,a), \tag{22}$$

其中 K 是由网络参数化所决定的隐式增益。这一视角阐明了执行器作为纠正运动生成器的作用：它使执行偏向更安全的区域，而不编码显式的任务目标。

评论家与执行器共同实例化一个学习式安全模块，该模块将保守可行性估计与有原则的、基于梯度的纠正方向相耦合，从而在富接触遥操作中实现主动且最小干预的辅助，如算法1所述

算法1： 保守成功的离线训练
评分与引导执行器

输入： 遥操作回合 $\{\tau\}$
输入： 折扣因子 γ、保守性权重 α、锚定权重 λ_{anchor}、目标周期 N_{tgt}
输入： （可选）辅助权重 λ_{fail}
输出： 评论家 $Q_\phi(s,\ a)$ 和执行器 $\pi_\theta(s)$
(1) 数据集构建。；
收集轨迹 $\tau = \{(s_t,\ a_t^{\text{tele}},\ s_{t+1})\}$ 并形成数据集 $\mathcal{D}$;
分配轨迹标签 $y(\tau) \in \{+1, -1\}$ 并设置 $r_t \leftarrow y(\tau)$，适用于所有步骤;
(2) 评论家更新（保守Q学习）。；
初始化 ϕ，$\bar{\phi} \leftarrow \phi$;
当未收敛时**执行**
  采样 $(s,\ a,\ r,\ s') \sim \mathcal{D}$;
  $y \leftarrow r + \gamma\,\mathbb{E}_{a'\sim\mathcal{D}(\cdot|s')}\,[Q_{\bar{\phi}}(s',\ a')]$;
  $\phi \leftarrow$ 使$_\phi\,(Q_\phi(s,\ a) - y)^2 + \alpha\,\mathcal{R}_{\text{cons}} + \lambda_{\text{fail}}\mathcal{L}_{\text{fail}}$最小;
  如果迭代次数模 $N_{tgt} = 0$ **则**
    $\bar{\phi} \leftarrow \phi$
(3) 演员更新（成功引导）。；
当未收敛时 **执行**
  采样 $s \sim \mathcal{D}$, $a^{\text{tele}} \sim \mathcal{D}(\cdot|s)$;
  $\theta \leftarrow$ 使$_\theta$ $Q_\phi\big(s$最小 $\pi_\theta(s)\big) + \lambda_{\text{anchor}}\|\pi_\theta(s) - a^{\text{tele}}\|^2$;
返回 Q_ϕ, π_θ;

D. 面向双手遥操作的价值引导阻抗辅助

现在，我们介绍如何通过主端机械臂上的阻抗反馈，将成功引导演员策略传达给人类操作员。系统采用双边遥操作架构，由人类操控主臂（主端）和机器人从臂（从端）组成。辅助并非直接修改操作员指令，而是通过调节主臂的机械阻抗来传达，使操作员能够通过物理交互感知引导。

如上一节所述，执行器输出成功引导辅助动作 a^{assist}，该动作根据学习到的可行性景观，编码了局部更安全的运动方向。该动作在从端空间中实例化为增量运动 $\delta\tilde{q}_{f,\ i}$。为将这一引导传达给人类操作员，从端的增量通过逆关节耦合S^{-1} 映射为主端虚拟参考，并由 $\delta q_{\max}$ 限幅，以限制辅助幅度。

a) 评论家到强度映射： 与此同时，评论家评估当前遥操作状态，并输出一个标量成功分数

$$Q_t \doteq Q_\phi(s_t, a_t^{\text{tele}}), \tag{23}$$

该值根据方程 19 进行归一化，并映射为有界的引导强度 $g_t \in [0,1]$；预测可行性越低，辅助力度越强：

$$g_t = \text{clip}\big(\sigma\big(\kappa(\tau_g - \bar{Q}_t)\big)\,,\,0,\,1\big)\,. \tag{24}$$

强度信号以伺服频率进行低通滤波，以确保触觉反馈平滑、稳定。

b) 价值引导阻抗渲染： 在每个主臂上，均以关节空间速度吸引子阻抗的形式呈现引导，

$$\tau_{\ell,i}^{\text{guide}} = K_v(g)\big(\dot{q}_{\ell,i}^{\text{des}} - \dot{q}_{\ell,i}\big) - D_0\,\dot{q}_{\ell,i}, \qquad i \in \{L,R\}, \tag{25}$$

其中 $\dot{q}_{\ell,i}^{\text{des}}$ 表示由执行器诱导的主臂侧虚拟速度参考，而 $K_v(g)$ 是由评论家导出的强度调制的半正定对角增益，

$$K_v(g) = K_{\min} + g\,(K_{\max} - K_{\min}), \quad 0 \preceq K_{\min} \preceq K_{\max}. \tag{26}$$

D_0 是一个恒定的阻尼矩阵。所有引导力矩均受到保守的幅值和变化率限值约束，以确保安全的物理人机交互。

c) 阻抗介导干预： 该公式将辅助的方向 与强度 解耦：前者由执行器通过虚拟速度参考 $\dot{q}_{\ell,i}^{\text{des}}$确定，后者则由评论家通过 g 调节。该速度在主臂侧阻抗控制器中充当虚拟运动吸引子，提供方向引导，

操作员通过阻抗反馈感受到这种作用，同时仍能直接控制运动。阻抗控制器不会施加明确的纠正力，也不会取代操作员的指令，而是让沿着容易导致失败的方向运动时受到越来越大的阻力，同时在更安全的方向上保持顺应性。

因此，人类操作员会感受到引导作用，其表现为接近不安全区域时阻力或阻尼逐渐增大。

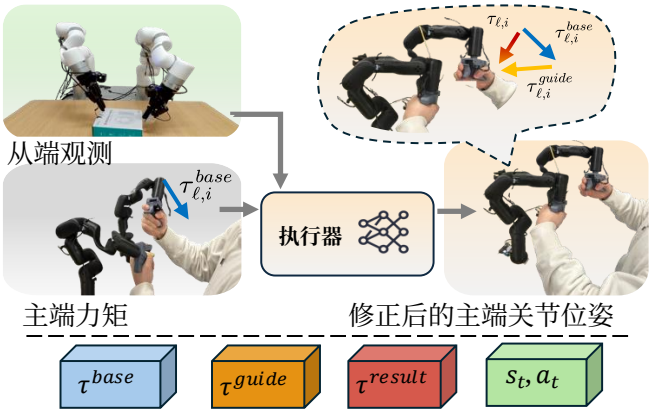


图3. **价值引导阻抗辅助**。 激活的引导会产生柔顺的主端力矩，以抑制不安全偏差，同时保持持续人类控制权。

通过双边耦合，这种阻抗整形经由物理交互影响执行过程，从而在保持连续、直观的人类控制的同时，提供主动且具备风险意识的辅助。

算法2： 价值引导阻抗辅助，用于双手遥操作

输入： 关节耦合 S，策略周期 Δt_π，伺服周期 Δt_s，执行器 π_θ，评论家 Q_ϕ

输入： 边界（$Q_{\min}$，$Q_{\max}$），Sigmoid函数（κ，τ_g），阻抗（$K_{\min}$，$K_{\max}$，D_0），限值（$\delta q_{\max}$，$\tau_{\max}$）

初始化 $g \leftarrow 0$, $\dot{q}_{\ell,i}^{\text{des}} \leftarrow \mathbf{0}$, $i \in \{L,R\}$;

当 遥操作处于活动状态**时**

  如果 策略更新可用（Δt_π） **则**

    $a^{\text{assist}} \leftarrow \pi_\theta(s_t)$;

    $Q_t \leftarrow Q_\phi(s_t, a_t^{\text{tele}})$;

    执行器输出 $\rightarrow$ 从臂增量 $\delta\tilde{q}_{f,i}$;$\delta q_\ell^{\text{ref}}$

    $\dot{q}_{\ell,\ i}^{\text{des}} \leftarrow$卫星$(S^{-1}\delta\tilde{q}_{f,\ i},\ \delta q_{\max})/\Delta t_\pi$;

    $Q_t \leftarrow$剪辑$((Q_t - Q_{\min})/(Q_{\max} - Q_{\min})\,0,\ 1)$;

    $g \leftarrow$剪辑$(\sigma(\kappa(\tau_g - \bar{Q}_t))\,0, 1)$;

  （可选）在 Δt_s 处对 g 进行低通滤波;

  $K_v \leftarrow K_{\min} + g(K_{\max} - K_{\min})$;

  $\tau_{\ell,i}^{\text{引导}} \leftarrow K_v(\dot{q}_{\ell,i}^{\text{des}} - \dot{q}_{\ell,i}) - D_0\dot{q}_{\ell,\ i}$;

  $\tau_{\text{引导}\ell,\ i} \leftarrow$sat$(\tau_{\text{引导}\ell,\ i},\ \tau_{\max})$;

  将引导力矩施加到领导者控制器;

V. 实验

A. 实验设置

1) 任务设计：我们设计了十项具有代表性的日常生活机器人操作任务，以全面验证我们的方法在多种操作范式下的适应性。根据操作复杂度与交互属性，这些任务被分为两组：第1组（基础操作任务）涉及相对简单的单阶段操作，没有或仅有极少的对象间交互；第2组（交互式细粒度操作任务）需要顺序多步骤执行、对象间

交互，或对小物体/数量的精确控制。我们使用先前工作中介绍的遥操作系统Unibidex [30]。这些实验的快照分别如 图4 所示。

第1组 包含四项任务，每项均通过一个核心单步操作完成：(1) 毛巾折叠：将一条长方形棉毛巾（20cm×45 cm）折成一个10cm×10cm的小长方形，并放置在预设目标位置。此任务侧重于通过单次折叠操作进行软物体塑形。(2) 拿取钢笔：从桌面上的笔筒中取出记号笔，然后拔下笔帽，最后将记号笔放在纸上的指定位置。核心要求是刚性物体抓取和垂直提起。(3) 倒水：将100ml水从一个圆柱形杯子（容量200ml，直径8cm）倒入另一个圆柱形杯子，且不能洒出，重点是通过单次倾倒动作完成液体转移。(4) 开盒：打开一个翻盖纸板盒（15cm×10cm×8cm），将盒盖抬至> 120°的角度，然后从盒中取出卷尺，并用它测量盒口的长度。

第2组 由六项需要多步骤协调或精确交互的任务组成：(5) 杯子交接：一只机械臂从桌上的初始位置拿起一个不锈钢杯，并将其平稳地交给另一只距离初始位置超过60厘米的机械臂，后者再将不锈钢杯放置在桌上指定的位置。(6) 开碗：一只机械臂需要依次移除随机放置在桌上的三个陶瓷碗的碗盖。(7) 抓取勺子：一只机械臂从桌上的初始位置拿起一把不锈钢勺。(8)–(10) 舀取盐／米／大豆：抓取勺子的机械臂（在任务（7）中）从三个打开的陶瓷碗（来自任务（6））中舀取颗粒材料，并将其放入桌上的不锈钢杯中（来自任务（5））。三个陶瓷碗中分别装有盐、大米和大豆。

2) 平台：实验平台围绕主从式遥操作架构构建，包括一个由用户操作的主臂和两个远程控制的从端机械臂，如图2所示。该装置专门用于促进高保真操作数据的收集，如第IV-A节所述。从端机械臂是两台相同的7自由度协作机械臂（UFACTORY xArm 7），每台最大有效载荷为3.5 kg。这些机械臂为执行多样化操作任务提供了必要的灵巧性和精度。主臂是两台定制设计的3D打印操作器，其运动学结构与xArm 7一致。该设计确保操作员手部动作与从臂运动之间形成直观的一对一映射，从而在遥操作过程中最大限度地降低认知负荷。此外，数据采集桌下方放置了一个脚踏板，供操作员确认动作。为匹配两组任务的不同需求，采用了不同的末端执行器。对于第1组任务，使用双指自适应夹爪（Robotiq 2F-85）。该夹爪最大行程为85mm，抓取力范围为10-140N，适用于处理各种尺寸和刚度的物体，从柔软的毛巾到刚性的钢笔和盒子。对于第2组任务，使用灵巧型机器人手（DH-5-6）得以采用。这只五指手总共提供16个自由度，能够实现勺子操作、精准倾倒和物体交接等任务所需的复杂抓取与操作策略。其精细控制对于涉及小物体和颗粒材料的精细操作至关重要。

B. 对比方法

我们将所提方法与四种基线方法进行比较。具体而言，**人类** 表示由人直接在物理环境中用双手手动执行，不涉及任何机器人遥操作，作为具有丰富触觉反馈的人类水平表现的参考。**GELLO** [3] 是一种运动学匹配的遥操作接口，以主从控制器的形式实现，且不包含学习辅助。**虚拟现实** [31]采用基于虚拟现实的双臂遥操作接口，机器人以主从方式跟踪其指令，同样不包含学习辅助。**FACTR** [2] 是一种低成本的双向 主从遥操作系统，通过驱动主臂传递从臂的外部关节力矩来提供力反馈（并辅以重力补偿和冗余度解析），从而提升高接触任务中的遥操作表现。**Ours-LF** 是我们的方法：该方法利用从离线遥操作数据中学习得到的保守成功评分调节连续的引导强度，并由学习得到的执行器提供优选的纠正方向；二者均以柔顺的关节空间阻抗触觉引导形式呈现，在保留人类控制权的同时，引导操作员避开易导致失败的动作。

C. 实施

为确保实验结果的可靠性和统计显著性，我们严格执行以下方案：(1) 试验重复：对于五种方法（Ours-LF、Human、GELLO、VR、Naive-LF），每项任务均重复40次。每项任务与方法组合的40次重复均连续进行，以避免环境变化；不同任务与方法组合之间间隔2分钟，以防止设备过热和操作者疲劳（针对Human和VR方法）。(2) Human/VR方法的操作员选择：招募三名经验丰富的操作员（具有超过3个月的机器人遥操作经验），使用Human和VR方法执行任务。在正式实验前，所有操作员均接受1小时的培训课程，以熟悉任务要求和操作工具。(3) 数据初始化：对于Ours-LF和Naive-LF，使用第IV-A节收集并标注的遥操作数据（包括针对10项任务的500条有效操作轨迹）训练模型。训练过程在配备Intel Core i9-12900K CPU和NVIDIA RTX 5090 GPU的服务器上进行；训练100轮后固定模型参数，以确保所有试验的初始条件一致。(4) 数据记录：每次试验期间，实时记录以下数据：机械臂的关节角度、末端执行器的位置/方向、夹爪的开合程度、左/右踏板操作，以及任务的开始/结束时间。

表 I 所有十项任务中不同方法的完成时间和成功率。

方法	第1组								第2组											
	毛巾 折叠		Pen 拿取		水 倾倒		Box 打开		Cup 交接		Bowl 打开		勺子 抓取		Salt 舀取		Rice 舀取		大豆 舀取	
	时间（秒）	Succ	时间（秒）	Succ	时间（秒）	Succ	时间（秒）	Succ	时间（秒）	Succ	时间（秒）	Succ	时间（秒）	Succ	时间（秒）	Succ	时间（秒）	Succ	时间（秒）	Succ
人类	22 ± 5	40/40	7 ± 2	40/40	12 ± 3	40/40	23 ± 4	40/40	8 ± 1	40/40	14 ± 3	40/40	19 ± 3	40/40	9 ± 2	40/40	11 ± 3	39/40	12 ± 3	38/40
FACTR	198 ± 6	35/40	36 ± 4	36/40	87 ± 8	37/40	163 ± 6	34/40	77 ± 8	38/40	89 ± 9	36/40	192 ± 13	34/40	46 ± 8	37/40	48 ± 8	37/40	54 ± 9	36/40
GELLO	179 ± 12	32/40	42 ± 5	32/40	93 ± 12	35/40	176 ± 10	31/40	89 ± 9	36/40	108 ± 12	33/40	231 ± 18	31/40	58 ± 12	34/40	52 ± 9	33/40	62 ± 20	31/40
VR	234 ± 27	31/40	52 ± 9	33/40	104 ± 17	34/40	252 ± 26	30/40	98 ± 11	34/40	186 ± 17	32/40	328 ± 34	28/40	67 ± 14	31/40	74 ± 11	29/40	70 ± 12	27/40
Ours-LF	157 ± 6	39/40	31 ± 4	40/40	78 ± 8	40/40	144 ± 5	39/40	60 ± 7	40/40	78 ± 8	40/40	174 ± 13	38/40	37 ± 4	40/40	41 ± 4	39/40	44 ± 5	39/40

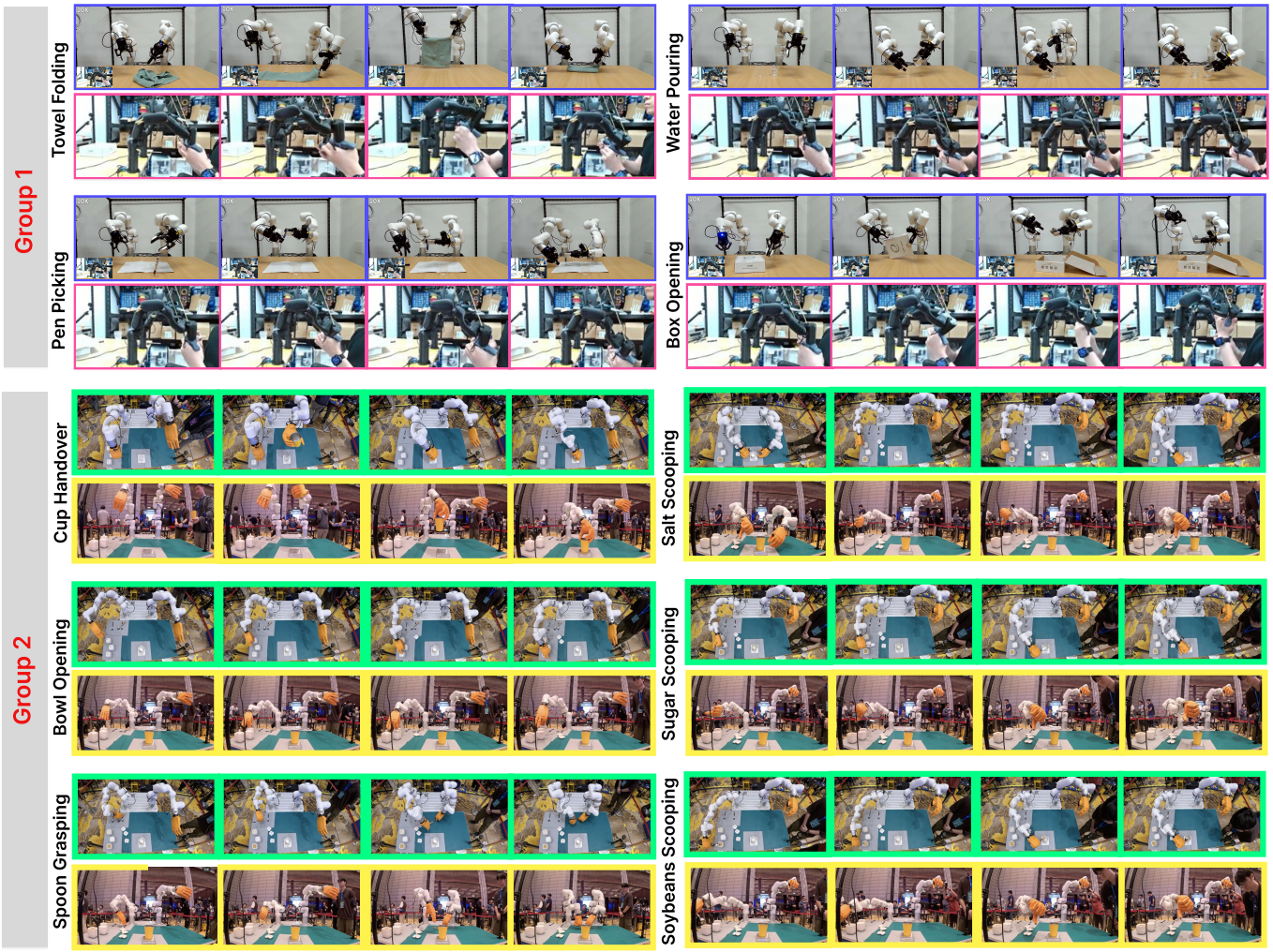


图4. 十项实验的所有快照，包括第1组（基础操作任务）中的四项任务，其跟随者视角（紫色框内）和领导者视角（粉色框内），以及第2组（交互式精细操作任务）中的六项任务，其跟随者视角（黄色框内）和俯视视角（绿色框内）。

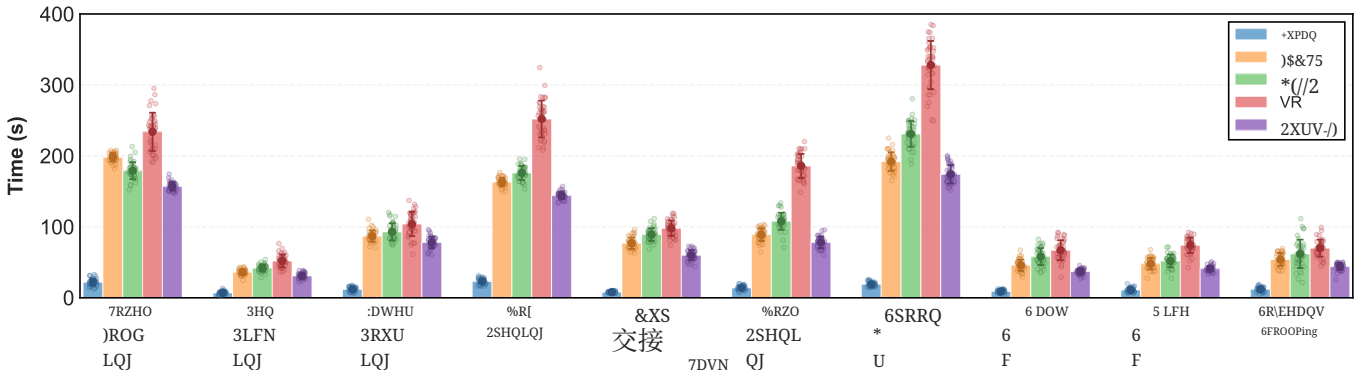


图5. 比较　　基线方法与我们的方法在所有任务中的完成时间　　。

D. 实验结果

采用两个关键指标评估全部十项任务的实验结果。**完成时间**：从任务开始到成功完成所用的时间。失败情况不计入完成时间。**成功率**：在规定时限内成功完成试验的百分比。五种对比方法下10项任务的完成时间和成功率统计结果分别列于表 I中。具体而言，完成时间在图5中进行了进一步展示，直观反映了五种方法在任务执行效率方面的差异。

对于第1组任务，人类操作员作为性能上限，在全部四项任务中均取得最短的完成时间（例如，拿取钢笔为 7 ± 2 秒）和100%的成功率（40/40）。在遥操作方法中，Ours-LF展现出更高的效率：对于毛巾折叠，其完成时间（157 ± 6秒）明显短于FACTR（198 ± 6秒）、GELLO（179 ± 12秒）和VR（234 ± 27秒），同时保持近乎完美的成功率（39/40）。对于拿取钢笔、倒水和开盒，Ours-LF在完成时间（分别为31 ± 4 秒、 78 ± 8秒和 144 ± 5 秒）方面优于其他遥操作方法，同时保持较高的成功率（拿取钢笔和倒水为40/40，开盒为39/40），因此其性能与人类基线相当。

对于第2组任务，人类操作员可作为性能基准，在大多数任务中展现出最短的完成时间（例如，抓取勺子仅需 19 ± 3 秒），并且成功率接近完美（四项任务均为 40/40，舀米为 39/40，舀大豆为 38/40）。在遥操作方法中，Ours-LF 展现出更高的效率和可靠性。在杯子交接任务中，其完成时间（60 ± 7 秒）短于 FACTR（77 ± 8 秒）、GELLO（89 ± 9 秒）和虚拟现实（98 ± 11 秒），且成功率达到 100%（40/40）。在开碗任务中，Ours-LF 将完成时间缩短至 78 ± 8 秒（其他基线方法均超过 89 ± 9 秒），同时实现了 100% 的成功率。在舀盐等精细操作任务中，Ours-LF 的完成时间最短（37 ± 4 秒），成功率达到 100 %，优于 FACTR（46 ± 8 秒）、Gello（58 ± 12 秒）和虚拟现实（67 ± 14 秒）。在抓取勺子、舀米和舀大豆任务中，Ours-LF 的完成时间更短，成功率（$\geqslant$38/40）也高于其他遥操作方法。

E. 结果分析

从上述实验结果可以看出，人类基线的平均成功率为 ~100%，且在所有任务中的完成时间最短。Ours-LF 的平均成功率达到 $\geqslant$98%（FACTR/GELLO/VR 为 ~80%），平均完成时间比其他遥操作方法短 ~25%，如图5所示。Ours-LF 的卓越表现源于所提出的遥操作框架中的两项核心创新，它们解决了现有方法的关键局限性。

首先，我们的框架能够稳健地识别分布外条件并进行实时校正，即故障感知校正机制。当检测到分布外状态时（例如图 6 所示，笔帽位于夹爪的可达范围之外），

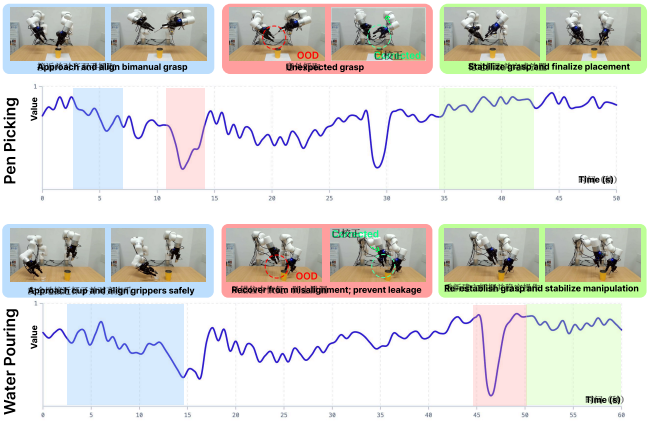


图6. 故障校正。在远程操作过程中，我们的故障感知框架能够检测分布外条件并主动介入，及时向主端机械臂施加适当动作，从而使从端机械臂恢复至正常状态（绿色虚线箭头所示）。

框架会向主端机械臂发送有针对性的电机指令，从而使从端机械臂调整其轨迹，回到正常的操作范围。该机制缩短了从偏差中恢复所需的时间，同时减轻了操作者的认知负荷，使用户能够专注于核心操作任务，而不必直接处理错误恢复。这使得 Ours-LF 的完成时间更短，成功率也接近完美。

其次，我们框架中集成的力控制反馈提升了性能，尤其是在接触密集型操作任务中。例如，在抓取勺子任务中，从端机械臂受到的环境力会迅速传递至主端机械臂，从而实现自适应动作调节。缺乏力反馈的方法（如 VR、GELLO）由于机器人与环境之间的碰撞未得到调节，常常触发紧急停止，因此在第2组的所有任务中成功率最低（见表 I）。FACTR 虽然采用了力控制，但由于缺乏故障感知校正，其完成时间更长，成功率也低于 Ours-LF。总之，故障感知校正与力反馈的协同作用，是 Ours-LF 在全部10项任务中表现卓越的关键。

F. 局限性

尽管取得了令人鼓舞的成果，所提出的框架仍存在若干局限性。首先，引导策略根据学习得到的成功分数优化纠正方向，但未明确施加接触约束，例如单边接触或摩擦模式。因此，建议的运动可能与瞬时接触状态相冲突，导致富接触交互中出现局部振荡和不一致的触觉反馈。其次，失败感知能力本质上受限于离线数据集的支持范围。尽管保守成功评分提高了系统对操作员引起的分布偏移的鲁棒性，但在任务条件发生变化时（例如末端执行器几何形状、材料），其可靠性可能下降，因此需要补充数据或进行适应，以维持一致的引导效果。

VI. 结论

本文提出了一种用于富接触操作的双手遥操作故障感知共享自主框架。利用包含成功与失败案例的异构离线遥操作数据，我们学习得到一种保守成功评分，可在分布偏移情况下提供风险敏感的可行性估计。在执行过程中，该评分用于调节辅助强度，同时由学习得到的执行器提供纠正运动方向；两者均通过关节空间阻抗接口实现，从而在不损害人类控制权的前提下，提供顺应且立足于物理交互的触觉引导。在十项日常生活操作任务上的实验表明，与传统遥操作和共享自主基线方法相比，该框架在成功率和效率方面均取得了稳定提升。总体而言，结果表明，保守价值学习为将预测性失败感知融入双边遥操作提供了有效途径，从而提升了富接触场景下的可靠性。

参考文献

[1] Y. Qin, W. Yang, B. Huang, K. Van Wyk, H. Su, X. Wang, Y.-W. Chao, and D. Fox, 《Anyteleop：一种通用的基于视觉的灵巧机器人双臂遥操作系统》，*arXiv*预印本 *arXiv:2307.04577*，2023年。[2] J. J. Liu, Y. Li, K. Shaw, T. Tao, R. Salakhutdinov, and D. Pathak, 《Factr：面向接触丰富策略学习的力关注课程训练》，*arXiv*预印本 *arXiv:2502.17432*，2025年。[3] P. Wu, Y. Shentu, Z. Yi, X. Lin, and P. Abbeel, 《Gello：一种通用、低成本且直观的机器人操作臂遥操作框架》，载于*2024*年*IEEE/RSJ*智能机器人与系统国际会议（*IROS*）。IEEE，2024年，第12 156–12 163页。[4] S. An, Z. Meng, C. Tang, Y. Zhou, T. Liu, F. Ding, S. Zhang, Y. Mu, R. Song, W. Zhang 等，《通过模仿学习实现灵巧操作：综述》，*arXiv*预印本 *arXiv:2504.03515*，2025年。[5] A. D. Dragan and S. S. Srinivasa，《一种用于共享控制的策略融合形式体系》，国际机器人研究期刊，第32卷，第7期，第790–805页，2013年。[6] Z. Fu, T. Z. Zhao, and C. Finn，《Mobile aloha：通过低成本全身遥操作学习双手移动操作》，载于机器人学习会议（*CoRL*），2024年。[7] Z.-H. Yin, C. Wang, L. Pineda, F. Hogan, K. Bodduluri, A. Sharma, P. Lancaster, I. Prasad, M. Kalakrishnan, J. Malik 等，《Dexterity-gen：实现前所未有灵巧性的基础控制器》，*arXiv*预印本 *arXiv:2502.04307*，2025年。[8] P. A. Lasota, T. Fong, J. A. Shah 等，《安全人机交互方法综述》，机器人学基础与趋势®，第5卷，第4期，第261–349页，2017年。[9] A. R. Cassandra，《POMDP应用综述》，载于*AAAI 1998*年秋季“部分可观测马尔可夫决策过程规划”研讨会工作札记，第1724卷，1998年。[10] S. Javdani, H. Admoni, S. Pellegrinelli, S. S. Srinivasa, and J. A. Bagnell，《通过事后优化实现用于遥操作和协作的共享自主》，国际机器人研究期刊，第37卷，第7期，第717–742页，2018年。[11] P. Aigner and B. McCarragher，《利用势场将人类融入机器人控制》，载于国际机器人与自动化会议论文集，第1卷。IEEE，1997年，第291–296页。[12] S. Bansal, M. Chen, S. Herbert, and C. J. Tomlin，《哈密顿–雅可比可达性：简要概述与最新进展》，*IEEE*控制系统杂志，第37卷，第6期，第28–42页，2017年。[13] C. E. Mower, J. Moura, and S. Vijayakumar，《基于技能的共享控制》，载于机器人学：科学与系统*2021*。机器人学：科学与系统基金会，2021年。[14] K. Muelling, A. Venkatraman, J.-S. Valois, J. E. Downey, J. Weiss, S. Javdani, M. Hebert, A. B. Schwartz, J. L. Collinger, and J. A. Bagnell，《融入自主能力的遥操作及其在脑机接口控制操作中的应用》，自主机器人，第41卷，第6期，第1401–1422页，2017年。[15] H. Liu, S. Nasiriany, L. Zhang, Z. Bao, and Y. Zhu，《在岗机器人学习：部署期间的人在回路自主系统与学习》，国际机器人研究期刊，第44卷，第10–11期，第1727–1742页，2025年。

[16] S. Luo, Q. Peng, J. Lv, K. Hong, K. R. Driggs-Campbell, C. Lu, and Y.-L. Li，《人类-智能体联合学习，实现高效的机器人操作技能习得》，载于*2025*年*IEEE*国际机器人与自动化会议（*ICRA*）。IEEE，2025年，第1370–1377页。[17] C. Schaff and M. R. Walter，《面向共享自主的残差策略学习》，*arXiv*预印本 *arXiv:2004.05097*，2020年。[18] S. Reddy, A. D. Dragan, and S. Levine，《通过深度强化学习实现共享自主》，*arXiv*预印本 *arXiv:1802.01744*，2018年。[19] H. Liu, R. Shah, S. Liu, J. Pittenger, M. Seo, Y. Cui, Y. Bisk, R. Mart´ın-Mart´ın, and Y. Zhu，《Casper：利用视觉语言模型推断多样化意图以辅助遥操作》，*arXiv*预印本 *arXiv:2506.14727*，2025年。[20] S. Levine, A. Kumar, G. Tucker, and J. Fu，《离线强化学习：教程、综述与开放问题展望》，*arXiv*预印本 *arXiv:2005.01643*，2020年。[21] C. J. Watkins and P. Dayan，《Q学习》，机器学习，第8卷，第3期，第279–292页，1992年。[22] I. Kostrikov, A. Nair, and S. Levine，《基于隐式Q学习的离线强化学习》，*arXiv*预印本 *arXiv:2110.06169*，2021年。[23] B. D. Ziebart, A. L. Maas, J. A. Bagnell, A. K. Dey 等，《最大熵逆强化学习》。载于*Aaai*，第8卷。美国伊利诺伊州芝加哥，2008年，第1433–1438页。[24] A. Kumar, A. Zhou, G. Tucker, and S. Levine，《面向离线强化学习的保守Q学习》，神经信息处理系统进展，第33卷，第1179–1191页，2020年。[25] Physical Intelligence, A. Amin, R. Aniceto, A. Balakrishna, K. Black, K. Conley, G. Connors, J. Darpinian, K. Dhabalia, J. DiCarlo 等，《$\pi_{0.6}$：一种从经验中学习的视觉-语言-动作模型》，*arXiv*预印本 *arXiv:2511.04352*，2025年。 [在线]。网址：https://arxiv.org/abs/2511. 04352[26] K. Lei, H. Li, D. Yu, Z. Wei, L. Guo, Z. Jiang, Z. Wang, S. Liang, and H. Xu，《Rl-100：利用真实世界强化学习实现高性能机器人操作》，*arXiv*预印本 *arXiv:2510.14830*，2025年。[27] Y. Xu, Z. Liu, G. Duan, J. Zhu, X. Bai, and J. Tan，《三思而后行：结合人类干预的安全基于模型强化学习》，载于机器人学习会议。PMLR，2022年，第332–341页。[28] W. Schwarting, J. Alonso-Mora, L. Pauli, S. Karaman, and D. Rus，《自动驾驶车辆中的并行自主：以最少干预实现安全运动生成》，载于*2017*年*IEEE*国际机器人与自动化会议（*ICRA*）。IEEE，2017年，第1928–1935页。[29] Z. Zhang, W. Lu, L. Yang, W. Jing, B. Tang, and J. Pan，《盲人机器人的协作装配策略学习》，载于*2025*年*IEEE*国际机器人与仿生学会议（*ROBIO*）。IEEE，2025年，第198–205页。[30] Z. Li, Z. Guo, J. Hu, D. Navarro-Alarcon, J. Pan, H. Wu, and P. Zhou，《Unibidex：用于机器人双手灵巧操作的统一遥操作框架》，载于*2025*年*IEEE*国际机器人与仿生学会议（*ROBIO*）。IEEE，2025年。[31] R. Ding, Y. Qin, J. Zhu, C. Jia, S. Yang, R. Yang, X. Qi, and X. Wang，《Bunny-visionpro：用于模仿学习的实时双手灵巧遥操作》，载于*2025*年*IEEE/RSJ*智能机器人与系统国际会议（*IROS*）。IEEE，2025年，第12 248–12 255页。



Peng Zhou 于2022年获得香港理工大学机器人学博士学位。2021年，他曾在瑞典斯德哥尔摩瑞典皇家理工学院的机器人、感知与学习实验室担任访问研究员。2022年至2024年，他在香港大学计算机科学系担任博士后研究员。目前，他是大湾区大学（GBU）先进工程学院的助理教授。他的研究兴趣包括机器人操作、机器人学习以及任务与运动规划。

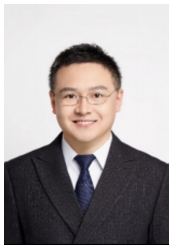
周博士目前担任IEEE机器人与自动化快报和《机器人与人工智能前沿》的副编辑。



Zhongxuan Li 获得了英国伦敦伦敦帝国理工学院电气与电子工程综合工程硕士学位（已纳入学士学位）。目前，他正在香港香港大学（HKU）计算机科学系攻读机器人学博士学位。他还是大湾区大学Email实验室的访问学生。他的研究兴趣包括机器人任务理解与规划。



Jinsong Wu 于2021年获得英国利物浦大学机电一体化与机器人系统工程学学士学位，并于2022年获得英国伦敦帝国理工学院医疗机器人与图像引导干预研究硕士学位。2022年至2023年，他曾在香港中文大学香港校区担任研究助理。自2023年起，他在香港香港理工大学攻读机械工程博士学位。他目前的研究兴趣包括医疗机器人与控制理论。



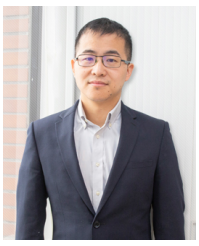
Jiaming Qi 于2023年和2018年分别获得中国哈尔滨哈尔滨工业大学控制科学与工程博士学位和集成电路工程理学硕士学位。2019年，他曾在香港中国香港特别行政区香港理工大学机器人与机器智能实验室担任访问博士生。2023年，他曾在中国香港特别行政区香港大学计算机科学系制衣生产中心有限公司担任博士后研究员。他目前是一名助理教授，任职于中国哈尔滨东北林业大学机电工程学院。他的研究兴趣包括可变形物体操作、视觉伺服、人机协作和机器人柔顺控制。



Jun Hu 于2023年获得中国合肥中国科学技术大学控制科学与工程博士学位。他目前是中国广东大湾区大学先进工程学院的博士后研究员。他的研究兴趣集中于机器人操作与感知，尤其关注灵巧操作、触觉感知、多模态感知以及机器人系统的技能学习。他近期的研究包括触觉传感器设计与校准、用于灵巧操作的遥操作手部，以及用于机器人操作中力与接触理解的基于学习的方法。



David Navarro-Alarcon （IEEE高级会员）于2014年获得香港中文大学机械与自动化工程博士学位。自2017年起，他任职于香港理工大学，目前担任机械工程系副教授。他目前的研究兴趣包括感知机器人学和控制系统。目前，他担任IEEE《机器人学汇刊》副编辑。



Jia Pan （IEEE高级会员）于2013年获得美国北卡罗来纳大学教堂山分校计算机科学博士学位。他目前是香港大学计算机科学系副教授，同时也是香港制衣生产中心有限公司的成员。他的研究兴趣包括机器人学和人工智能在自主系统中的应用，尤其是在导航和操作等具有挑战性的任务中实现有效在密集人群中移动，以及为服装自动化操纵可变形物体。



Lihua Xie （IEEE 会士）于1983年和1986年分别获得中国南京的南京理工大学电气工程学士学位和硕士学位，并于1992年获得澳大利亚新南威尔士州卡拉汉纽卡斯尔大学电气工程博士学位。自1992年以来，他一直任职于新加坡南洋理工大学电气与电子工程学院，目前担任教授，并任先进机器人技术创新中心主任。他1986年至1989年，他在南京理工大学自动控制系任教。2011年7月至2014年6月，他担任控制与仪器部门主任；2016年至2021年，他担任Delta-NTU网络物理系统企业实验室联合主任。他的研究兴趣包括鲁棒控制与状态估计、网络化控制系统、多智能体网络、定位和无人系统。Xie博士是《无人系统》的主编，也是《IEEE控制系统杂志》的副编辑。他曾任IET控制丛书编辑，并担任多家期刊的副编辑，包括《IEEE自动控制汇刊》、《Automatica》、《IEEE控制系统技术汇刊》、《IEEE网络控制系统汇刊》和《IEEE电路与系统汇刊II》。他曾任IEEE杰出演讲人（2012年1月至2014年12月）。他是新加坡工程院、IEEE、IFAC和中国自动化学会会士。



Shiyao Zhang （IEEE高级会员）现任大湾区大学先进工程学院助理教授。他于2014年获得美国印第安纳州西拉法叶普渡大学电气与计算机工程专业理学学士学位（荣誉），于2016年获得美国加利福尼亚州洛杉矶南加州大学电气工程硕士学位，并于2020年获得中国香港特别行政区香港大学博士学位。他曾以博士后研究员身份任职于高等研究院，南方科技大学，自2020年至2022年，并于2022年至2024年担任南方科技大学可信自主系统研究院研究助理教授。他的研究兴趣包括智能交通系统、具身人工智能和交通电气化。他已在顶级国际期刊和会议上发表60余篇学术论文。他是《智能与机器人》和《人工智能与自主系统》期刊的青年编辑。



Zeqing Zhang （IEEE 会员）于 2019 年获得香港中文大学哲学硕士学位，并于 2024 年获得香港大学计算机科学博士学位。2024 年至 2025 年，他担任香港大学荣誉研究助理，并担任由创新及科技基金资助的博士后研究员，任职于香港 InnoHK 变革性服装生产中心（TransGP）。目前，他是新加坡南洋理工大学的研究员。他的研究兴趣主要是利用机器人本体感知和外感知，感知并操作软体（例如服装）及颗粒介质（例如沙粒）。他主要运用视觉、触觉、力觉传感和文本模态来处理这些非线性材料。他曾于 2024 年作为访问学者访问日本大阪大学，并于 2025 年访问美国斯坦福大学。