arXiv:2301.00452v2 [cs.RO] 6 Jun 2023

用于手术机器人学习的交互式仿真环境与人在回路具身智能

Yonghao Long、Wang Wei、Tao Huang、Yuehao Wang 和
Qi Dou 香港中文大学

摘要—过去十年来，手术机器人自动化引起了越来越多的研究关注，人们期待它能为外科医生、护士和患者带来潜在益处。近年来，具身智能的学习范式展现出学习各种复杂任务的优良控制策略的巨大潜力，而具身人工智能仿真器在推动相关研究方面发挥着关键作用。然而，现有的手术机器人开源仿真器仍无法充分支持通过物理输入设备进行人机交互，这进一步限制了人们对人类示范如何影响策略学习开展有效研究。在本研究中，我们借助一个用于手术机器人学习的新型交互式仿真平台，研究人在回路具身智能。具体而言，我们基于此前发布的 SurRoL模拟器搭建了该平台，并共同开发了若干新功能，以便通过输入设备实现高质量的人机交互。我们展示了这些新功能如何改进仿真环境，并以人类示范与强化学习为代表性示例，验证了将人为因素纳入具身智能的有效性。结果表明，学习效率有所提升。最后，我们开发并发布了五项新的手术机器人训练任务，希望借此为未来手术具身智能研究铺平道路。我们的学习平台现已公开发布，并将在以下网站持续更新：
https://med-air.github.io/SurRoL。

一、引言

手术机器人技术在过去十年中发展迅速，并在实践中改变了微创手术 [1]。近年来，手术任务自动化 [2] 受到研究人员越来越多的关注，因为它有望减轻外科医生的负担并提高操作效率 [3]。然而，由于复杂的手术场景和高难度的手术操作所带来的挑战，实现这一目标仍然遥不可及。传统解决方案通常依赖启发式规划方法，但难以产生可扩展的控制策略。迄今为止，成功

手术任务自动化的相关研究仍处于起步阶段 [4]。

具身智能 [5]提出基于机器人与环境之间的交互，直接学习各种技能，并在机器人任务自动化方面展现出显著能力 [6]。在此背景下，模拟器 [7–9] 是提供物理世界数字

本项目部分由香港研究资助局 TRS 项目（编号：T42-409/18-R）资助，部分由香港创新及科技委员会项目（编号：PRP/026/22FX）资助，部分由 InnoHK多尺度医疗机器人中心的拨款（编号：8312051）资助，另有部分由 Cornerstone Robotics有限公司的研究基金资助。
Y. Long、W. Wei、T. Huang、Y. Wang 和 Q. Dou 任职于香港中文大学计算机科学与工程学系。Q. Dou 还任职于香港中文大学 T Stone机器人研究所。通讯作者：*Qi Dou*（*qidou@cuhk.edu.hk*）。

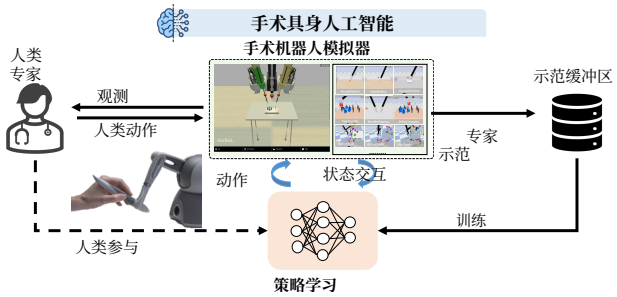

图 1. 人类参与示范的手术具身智能概念示意图，用于机器人学习。

孪生体 [10, 11] 的关键基础设施，可用于促进交互数据的收集，以及智能体的训练和测试。强化学习（RL）[12] 通常与具身人工智能结合使用，可将任务执行建模为马尔可夫决策过程，并通过机器人与环境在模拟器中的交互来优化策略学习。尽管人们已为具身智能的普遍发展付出了诸多努力 [7, 8, 13],，仍需进一步探索手术具身智能；这需要由量身定制的领域专用仿真环境提供支持。

具体而言，早期研究 [14, 15] 主要聚焦于手术机器人及其相关任务（如棋钉转移）的仿真，着重创建逼真的虚拟环境，而非像如今这样服务于人工智能。近年来，模拟器越来越注重将具身智能，尤其是强化学习，引入手术机器人领域，以开发先进的任务自动化方法。Richter 等人提出了 dVRL [16], ，这是首个面向手术机器人技术的开源强化学习环境，并展示了学习到的策略能够部署到 dVRK 平台。然而，该环境尚未提供物理输入设备接口，因此无法通过直观交互支持手动输入。这进一步限制了其收集人类示范数据以研究机器学习算法的功能。Tagliabue 等人开发了 UnityFlexML [17] ，用于在涉及软组织操作的任务中有效进行强化学习策略学习。目前，该平台仅包含少量片状虚拟资产，这可能限制可供研究的手术任务或场景的多样性。最近，Munawar 等人 发布了一个全面的仿真平台 AMBF [18] ，该平台支持交互式人类输入，因此能够收集数据，用于训练和测试手术机器人的控制机制。然而，AMBF 对人工智能算法库的支持尚不充分，难以帮助人们探索强化学习

方法，并利用大规模人类示范进行学习。

与模拟器当前存在的缺陷并行，手术机器人学习尚未解决的一项关键技术难题是如何实现更高水平的自主性 [19, 20]。解决这一问题需要注意到，大多数现有手术机器人仍采用以人为中心的方式（即遥操作） [2, 21]。换言之，人工智能的发展应将人为因素纳入其中，以利用专家的知识，提升学习复杂手术技能的认知能力 [22]。然而，实现人在回路中的手术具身智能面临诸多挑战。首先，如何在人类输入设备与手术机器人之间建立准确、直观的动作映射机制，并使其标准化且兼容机器学习。其次，如何构建逼真的物理仿真，并配合高保真场景可视化，以便在虚拟环境中根据人机交互提供沉浸式反馈。第三，在手术具身人工智能的背景下，如何理解人类示范对任务自动化策略学习的影响。

为应对上述挑战，我们提出研究人在回路中的具身智能，并开发一款专用于手术机器人学习的交互式模拟器，其概念如图所示。1。人类（通常是专家外科医生）可以通过与物理输入设备交互，操控虚拟机器人臂来操作虚拟物体，并在模拟器中感知这些交互的视觉反馈。人类示范可以保存到数据库中，形式为末端执行器动作（位置、方向和夹爪角度的变化）与环境状态构成的动作-状态对，随后用于通过强化学习进行控制策略学习。此外，在典型的具身人工智能范式中，机器人也可以通过试错（即动作-状态对）自行探索环境。

在这项工作中，为了实现这一目标，我们基于现有的开源模拟器 SurRoL 开发了该平台 [23]，，即一个以强化学习为核心且兼容 dVRK 的手术机器人模拟器。新版本的重点是引入人机交互，并通过面向双手的触觉输入设备接口来实现。重要的是，我们共同开发了一系列新功能，以建立逼真的人机交互：1）标准化接口，支持通过物理输入设备进行人机交互和策略学习；2）逼真的物理仿真，对碰撞属性进行精细建模，并采用比例微分（PD）控制；3）高保真渲染，通过逼真的冯氏着色、聚光灯建模和阴影投射增强人类感知。在此基础上，我们进一步利用收集到的人类示范数据训练策略，以实现手术机器人任务自动化，并将其性能与此前使用代码生成的示范数据进行比较。结果表明，这有望提升控制策略学习的效率。最后，我们在模拟器中新增了五项外科手术训练任务（即 *PickAndPlace*、*PegBoard*、*NeedleRings*、*MatchBoard*、*MatchBoardPanel*），可用于从业人员的手术技能训练。借助这些任务，我们旨在推动未来对复杂训练任务以及人在回路手术具身智能的研究。我们的代码已作为更新分支发布于 SurRoL 仓库，地址为：https://github.com/med-air/SurRoL。

II. 相关工作

A. 借助交互式模拟器实现机器人具身人工智能

具身人工智能旨在通过与环境交互来学习复杂技能，而非直接从预先收集的视频、音频、图像和文本数据集中学习。具身人工智能的快速发展要求并推动了相关模拟器的开发 [5]。具身人工智能领域中多个具有显著贡献的模拟器都将交互作为一项重要特征。Xiang 等人提出了 SAPIEN [9]，，它支持在逼真且具有丰富物理特性的仿真环境中执行机器人交互任务。Kiana 等人提出了 ManipulaTHOR [24]，以促进使用机械臂进行视觉物体操作的研究。近年来，具身人工智能中的人机交互日益受到关注。Li 等人提出了 iGibson [8]，用于模拟多种家务任务，并支持基于虚拟现实的人机环境交互。Gan 等人提出了 ThreeDworld [25]，用于交互式多模态物理仿真，并支持人类通过虚拟现实设备进行交互。Fu等人提出了 RFUniverse [26]，，这是一个基于物理的动作中心环境，用于学习家务任务，同时也提供了供人类输入的虚拟现实接口。一些研究 [7, 27] 表明，纳入人类数据后，智能算法往往能展现出更好的性能。然而，在手术机器人学习这一研究主题中，专门针对模拟器的探索仍然很少，主要原因在于难以开发支持高质量人类参与和多种手术任务研究的开源软件基础设施。

B. 基于学习的手术任务自动化

基于学习的手术任务自动化是手术机器人技术中一个新兴领域，旨在利用机器学习技术实现手术任务自动化。该领域已研究了多种手术任务，包括缝合 [28],模式切割 [29], 组织操作 [30], 等。一种常用方法是模仿学习 [31], ，即通过监督学习技术，利用专家示范数据集训练策略。然而，这种方法无法应对分布偏移，因此泛化能力较差。另一方面，强化学习 [32, 33],允许机器人通过与环境交互来学习，展现出出色的泛化能力，能够学习不同的任务。然而，传统强化学习面临较大的探索负担，既耗时又耗费资源，且无法保证成功。近年来，人们越来越关注利用示范来提高强化学习的学习效率 [34–36]，，并取得了令人鼓舞的成果。由于真实外科手术存在潜在风险，且难以获取全面、准确的 [37] 手术数据，此类方法通常在交互式仿真环境中进行研究和测试。然而，开发工作量

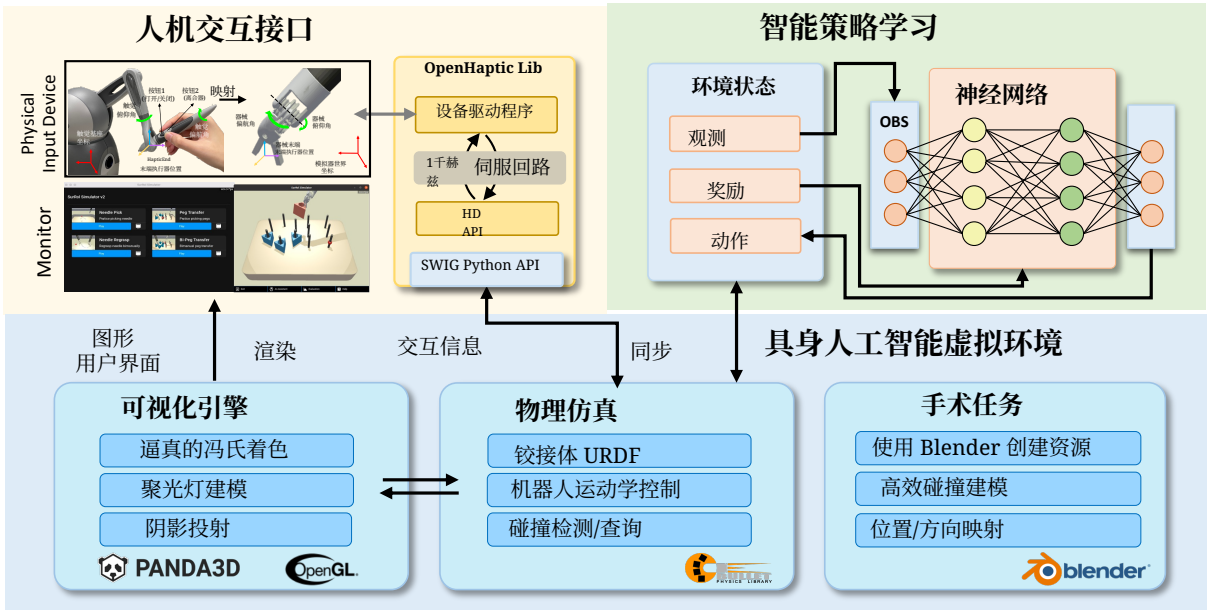

图 2。 所提出的平台 用于手术机器人学习的、配备交互式仿真环境的人在回路中的手术具身智能 g.

始终伴随着模拟器的搭建。鉴于此，我们致力于开发一款开源手术机器人模拟器，支持高质量的人机交互和多种任务。我们希望借此加速这一快速发展的领域取得进展。

三、材料与方法

A. 仿真平台开发作为基础设施

我们的整体系统由三个主要部分组成：1）人机交互接口；2）智能策略学习；3）手术具身人工智能虚拟环境。总体框架如图2所示。首先，我们借助三维建模工具 Blender 创建手术技能训练任务的资源，并同时生成相关的碰撞模型和 URDF 描述模型，以便在仿真环境中进行物理建模。将手术任务导入模拟器后，人类可以通过手动交互设备执行手术动作，交互信息则会流式传输至虚拟环境中进行物理仿真。同时，模拟器的可视化引擎会生成视频帧，并显示在显示器上，供人类感知并进行下一步交互。策略既可以通过与虚拟环境自主交互进行学习，也可以通过记录的专家示范向人类学习。

B. 使用输入设备进行运动学映射与控制

为了在模拟器中纳入人类对手术机器人的控制，首要且关键的一步是开发一套配备物理输入设备的操作系统。本文选择*Touch* [38] (3D Systems Inc.) 作为模拟器的典型输入设备，因为它具有稳定性高、可定制性强、应用广泛 [39, 40] 且成本低等优点。具体*Touch* 设备用于模拟机器人的两条主端机械臂，以远程操控患者侧操作器（PSM）和内窥镜摄像机操作器（ECM）。

1）从输入设备到虚拟机器人的运动学映射：为了实现物理输入设备与模拟器中手术机器人之间的平滑控制，我们将输入设备的末端执行器当前动作映射到模拟器，而不是映射位姿或关节角，因为这种方式更加直观，并且与策略动作（例如强化学习策略生成的动作）兼容，也有助于记录人类示范。具体而言，在每个仿真步 k，我们首先从触觉设备（如图 2 所示的物理输入设备）获取当前步 $p(k) = \{x(k)\ y(k)\ z(k)\ rotate(k)\}$ 以及前一步 $p(k-1)$的末端执行器位置和关节角。然后，我们将当前动作计算为$p(k) - p(k-1) = \{d_x, d_y, d_z, d_{rotate}\}$，其中 d_x、d_y、d_z决定笛卡尔空间中的位置移动，d_{rotate}决定方向变化。对于 PSM，d_{rotate} 为偏航角或俯仰角，分别对应俯视或垂直空间设置，可根据具体手术需求进行调整。对于 ECM，d_{rotate} 为滚转角，使外科医生能够调整摄像机绕其纵轴的角度。按下按钮 1（如图 2 所示的物理输入设备上的按钮）时，器械钳口角度 $j(k)$每步减小一个固定值，直至闭合（$j(k) < 0$）；松开按钮后，$j(k)$则逐渐增大，直至完全张开。按下按钮 2 时，所有动作均设为零，以模拟控制的离合机制，该机制通常用于在手术过程中调整主端工作空间。所有移动动作都将乘以缩放向量（对应工具移动比例），然后加到手术器械的当前状态上，以得到目标位姿。

2) 器械末端执行器控制：完成动作映射后，我们通过逆运动学（IK）和 PD 控制器实现手术机器人的控制。具体而言，给定手术器械末端执行器的目标位姿后，根据 Denavit-Hartenberg（DH）参数（与 dVRK [2]一致）计算手术机器人的目标关节角。由于 PSM 和 ECM 分别具有

6 个自由度（DOF）和 4 个自由度（DOF），且具有远程运动中心（RCM），因此二者均存在解析解。给定器械的目标位姿和当前位姿后，我们采用位置和速度比例微分控制，实现机器人平稳、逼真的运动控制。误差定义如下：

$$e(t_k) = K_1 \cdot (p_{\text{target}}(t_k) - p_{\text{current}}(t_k)) + K_2 \cdot (v_{\text{target}}(t_k) - v_{\text{current}}(t_k)), \tag{1}$$

并且离散控制系统 [41] 被表述为：

$$u(t_k) = \left(K_p + \tfrac{K_d}{\Delta t}\right) e(t_k) - \left(K_p + \tfrac{2K_d}{\Delta t}\right) e(t_{k-1}) + \tfrac{K_d}{\Delta t} e(t_{k-2}), \tag{2}$$

其中，$u(t_k)$表示系统输出动作，K_1 和K_2 分别表示比例和积分增益，Δt 表示时间间隔。为提高触觉输入与模拟器之间高频控制通信的效率，我们利用 SWIG [42] 将 OpenHaptics 的 HD API（原始触觉软件开发工具包，使用 C/C++ 实现）直接封装到 Python 中，以便将触觉设备与基于 Python 的环境连接起来。

C. 逼真物理交互仿真

将人机交互引入模拟器也会带来一些意料之外的动作（例如突然移动和破坏性行为）。因此，我们需要进一步优化基于 SurRoL 的物理仿真和交互，以提供逼真的操作体验。

首先，为了实现包括手术器械在内的不同物体的逼真仿真，我们使用 Blender 按照真实世界的比例对它们的所有铰接体进行建模。同时，物体 URDF 文件中的接触和惯性属性（例如刚度和质量）首先根据真实世界的测量值进行初始化，然后通过反复试验仿真进行手动调整。我们反复迭代调整，直至仿真结果尽可能接近真实世界。此外，鉴于手术任务中的大多数物体资源都比单个凸包或基元（例如长方体、圆柱体和球体）[18],复杂，因此我们对网格应用 V-HACD 凸分解 [43] 和碰撞基元复合体，以生成物体 URDF 文件中的碰撞几何体，从而实现更精准的碰撞检测。然而，当人类出现某些破坏性行为（例如以极大的动作幅度推或抓物体）时，手术器械与物体之间可能会发生相互穿透问题，从而影响真实感。当资源以异常高的速度相互碰撞时，就会出现这一问题，导致碰撞求解失败，使它们相交或重叠。此外，工具无法在真实的手术机器人中从一个位置突然移动到另一个位置。因此，设计具有力和速度约束的控制器至关重要。具体而言，我们基于位置和速度比例微分控制模式，对控制 PSM 和 ECM 时关节电机的输出施加约束。在逐步仿真过程中，底层设计的控制算法会在最大电机力和速度限制下计算电机动作，以达到目标位置。此外，为了实现更稳定、逼真的抓取，我们还在夹爪与物体的交互过程中采用了关节约束。图中展示了物理仿真优化前后的对比示例。3，结果表明，我们提出的解决方案能够防止相互穿透问题，从而实现更逼真的交互。

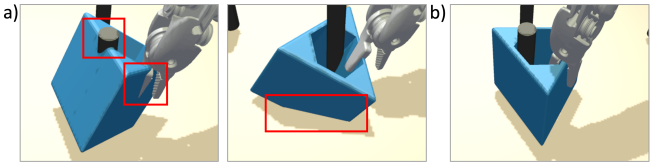

图 3. 在*PegTransfer* 任务（来自 SurRoL）中，优化物理建模前（a）和优化后（b）的示例结果。这表明，我们提出的方法能够在发生意外移动时避免相互穿透问题，从而使人类在模拟器中操作时获得更逼真的交互体验。

D. 高保真手术场景渲染

为了在模拟器中实现真实的人机交互，为渲染场景提供高度的视觉逼真度对于基于虚拟现实的手术训练中的人类感知至关重要。然而，原始 SurRoL 中的物理仿真后端（使用 PyBullet）并不注重人机交互的渲染真实感，因此仅支持用于结果可视化的简单场景渲染。为解决这一局限性，我们为 PyBullet 引擎提出了一款实用的渲染插件，支持 Panda3D [44]（一个使用 Python API 进行三维渲染的开源框架）。我们开发的插件弥合了 PyBullet 与 Panda3D 之间的差距，从而实现了场景的逼真渲染。

具体而言，我们重写了 PyBullet 的底层渲染接口，使其能够将资源的位姿、纹理和材质等属性从 PyBullet 数据格式转换为 Panda3D 可读格式，然后传递给 Panda3D 进行渲染。我们采用 Panda3D 的 Phong 着色 [45]模型来模拟整个虚拟环境中的光照反射与漫射。与使用纹理、法线贴图和高光贴图来控制场景真实感的传统渲染管线 [46] 相比，我们采用的方法更符合真实世界的物理规律。此外，传统的光源模拟方式是使用方向光或环境光模型，这些方式与内窥镜光源并不一致。内窥镜光源采用光纤将外部光源引导至内窥镜末端，并沿特定方向向感兴趣区域（手术场景）发射强烈的锥形光束。为模拟这种照明条件，我们采用聚光灯模型，通过将光线聚焦于特定方向和视场来复现这一效果，从而实现逼真的手术机器人内窥镜照明模拟。此外，我们启用了阴影映射，以模拟光线遮挡并呈现投射阴影。最后，之前的渲染引擎生成的图像存在明显的混叠效应，会严重影响人类的视觉体验。因此，我们在渲染管线中启用了抗锯齿，以实现更自然的可视化效果。可视化结果对比如图4所示，结果表明，所提出的方法

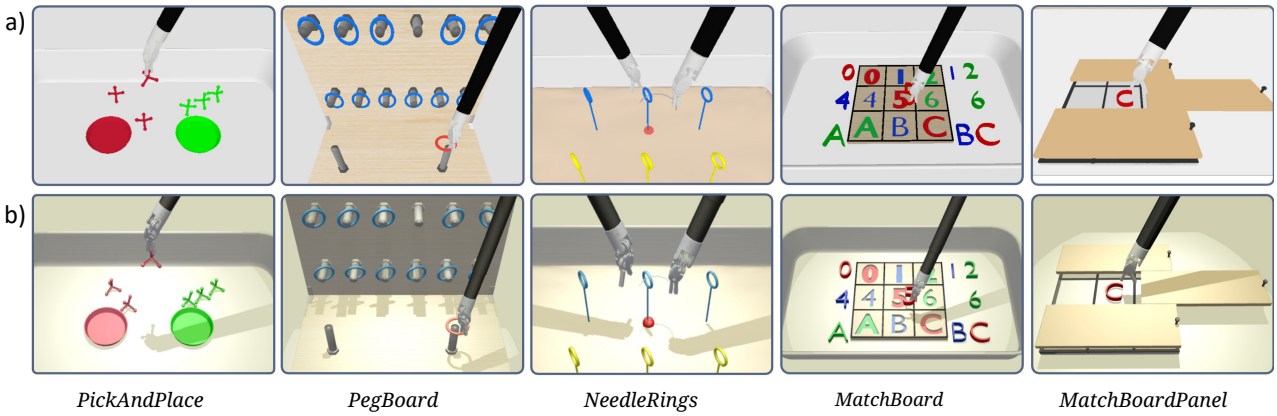

图 4. 可视化结果：（a）使用 SurRoL 的原始渲染引擎，（b）使用经过优化的真实感渲染引擎。自左至右的五个任务依次为 *PickAndPlace*、*PegBoard*、*NeedleRings*、*MatchBoard* 和*MatchBoardPanel*，相关描述见第 III-E节。

能够在照明、阴影、反射和保真度方面生成更逼真的可视化效果。最后，为了提升易用性，我们进一步利用 Panda3D 开发了一个图形用户界面（GUI）（如图2“显示器”所示）。受训者可以通过点击“播放”按钮，在面板中便捷地选择不同的外科手术训练任务；也可以通过点击“退出”按钮退出任务。该界面还可根据需要灵活扩展和定制。

值得注意的是，所提出的流程不仅可以提高人机交互中的渲染真实感，还能极大地促进基于图像的感知和仿真到现实任务的进一步研究，例如视觉强化学习 [33]，手术场景分割和动作识别，从而弥合虚拟环境与现实世界之间的差距。

E. 用于训练和自动化的手术任务

在之前版本的 SurRoL 模拟器中，我们设计了若干手术机器人任务（例如 *NeedlePick*、针重新抓取和 *EcmReach*），专门用于评估机器人学习算法。然而，这些任务不足以满足当前及未来针对手术训练的综合研究需求 [47, 48]，，其中人机交互是一个非常重要的因素。为此，我们参照机器人手术模拟训练中常见的课程任务，新增了五项任务 [49——51]，这些任务既可用于评估手术训练，也可用于评估任务自动化。具体而言，我们新增了 *PickAndPlace*、*PegBoard*、*NeedleRings*、*MatchBoard* 和 *MatchBoardPanel* 任务（见图 4）。在 *PickAndPlace*任务中，受训者需要拾取彩色插杆，并将其放入颜色相同的托盘中。只有所有彩色插杆都放入对应托盘中，任务才算成功。在 *PegBoard*任务中，受训者需要从垂直插销板上拾取圆环，再将其放到水平板的插销上。这要求受训者能够熟练控制圆环的位姿和方向，因此有助于练习其操作技能。在 *NeedleRings*任务中，受训者需要使用双机械臂传递针，使针穿过圆环。这要求受训者掌握熟练的双手针操作技能，精准控制针的位置和方向。在 *MatchBoard*任务中，受训者需要拾取数字或字母积木，并将其放置在凹槽中对应的位置，以进一步练习手眼协调。在 *MatchBoardPanel*任务中，受训者需要先抓住抽屉把手打开抽屉，然后拾取数字或字母积木，并将其放入目标网格中。这是一项包含顺序子任务的进阶技能训练任务。这些任务经过专门设计和建模，易于扩展并应用于手术训练和人在回路中的机器人学习的进一步研究。

F. 从人类示范中学习手术技能

传统的示范学习方法通常会利用机器生成示范中的数据和知识 [52, 53]，，但此类方法可能不适用于手术任务，因为这类任务需要外科医生进行精细且娴熟的手术操作，无法通过简单直接的动作来诠释。因此，在机器人手术中，示范学习主要关注专家人类示范 [36, 54]。为此，我们采用人类示范引导的强化学习作为示例，验证所提出的交互式模拟器。

1) 问题表述：具体而言，我们考虑一个与模拟器交互的强化学习智能体，并将其建模为马尔可夫决策过程。在每个时间步 k，智能体根据其策略 $\pi : \mathcal{S} \rightarrow \mathcal{A}$，在状态 $s_k \in \mathcal{S}$下采取动作 $a_k \in \mathcal{A}$。随后，它接收奖励信号 r_k 并转移到后继状态 s_{k+1}，重复执行策略直至回合结束；每次经验（s_k，a_k，r_k，s_{k+1}）都会存入经验回放缓冲区 $\mathcal{D}_A$。与此同时，智能体还维护另一个缓冲区 $\mathcal{D}_E$，其中包含预先提供的专家示范。智能体的目标是找到最优策略$\pi^\star$，使折扣因子 $\gamma \in$ 下折扣累积奖励（又称回报）的期望最大化（$0, 1]$。许多强化学习方法通过估计Q值函数Q 来实现这一目标：$\mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$，该函数给出状态 s_k下动作 a_k 的预期回报。本文采用深度确定性策略梯度 DDPG [55]，——一种广泛用于手术机器人学习的深度强化学习方法 [16, 30]，——通过最小化平方贝尔曼误差 [55]来学习参数为 θ 的Q值函数 Q_θ：

$$L_Q(\theta) = \mathbb{E}_{\mathcal{D}_A}\left[\left(r_k + \gamma Q_\theta(s_{k+1}, a_{k+1}) - Q_\theta(s_k, a_k)\right)^2\right]. \quad (3)$$

为利用示范中的知识，我们遵循 [56]中利用状态-动作对的方法

，从而鼓励智能体与专家之间的行为相似性。该方法首先使用行为克隆损失对由π_ϕ 参数化的 ϕ 策略进行预训练， [57]，然后在在线阶段最小化以下目标函数：

$$\mathcal{L}_\pi(\phi) = \mathbb{E}_{s\in\mathcal{D}_A}\left[-Q_\theta(s_k, \pi_\phi(s_k))\right] + \mathbb{E}_{s,a\in\mathcal{D}_E}\left[\|\pi_\phi(s_k) - a_k\|_2^2\right]. \tag{4}$$

我们进一步采用后见经验回放 HER [58]作为两个缓冲区的采样策略，以解决目标条件环境中的稀疏奖励问题。

2）实验设置：我们选择 *NeedlePick* 作为具有代表性的任务进行验证。该任务是一项重要的外科手术训练任务，用户需要操纵机械臂到达托盘上的针并将其拾起，然后将其移动到目标位置。该任务包含与虚拟环境的多次交互，因此需要具备相关的手术技能才能成功完成。奖励函数参照先前的 SurRoL 定义为：$r(s,a) = -\mathbb{I}_{(\|o_g - o_c\|_2 \geq \epsilon)}$，其中 o_c 和 o_g 分别表示针中心的当前位置和目标位置，而 ϵ 表示两者之间的容许距离。实验过程中，我们使用所提出的模拟器收集成功的人类专家和非专家示范。同时，我们还开发了一个线性路径规划脚本，可生成机器人当前位置与目标位置之间的一系列航点，以收集脚本示范。随后，我们比较使用人类专家示范、非专家示范以及脚本示范所取得的结果。为了评估学习到的策略，我们采用成功率、完成步数（时间成本）和动作经济性（轨迹距离）作为评估指标。这些指标因能有效反映手术技能水平，已被广泛用于评估外科医生在机器人手术模拟训练中的技能 [15, 50, 51]。

3）实现细节：按照 SurRoL 中的设置，工作空间设为 $10.0cm^2$，目标与当前状态之间的容差距离为 $0.5cm$。每个训练轮次包含 40 个回合，每个回合设为 150 个时间步。模拟时间步设为 0.2 秒。如果用户在规定的时间步内未能完成任务，环境将重置。每次环境重置时，都会随机采样初始条件和目标条件。策略采用具有 ReLU 激活函数的四层多层感知机建模，每层包含 256 个隐藏维度。我们基于 DEX [59] 框架实现了该方法，该框架构建于 OpenAI Baselines [60] 之上，并采用其超参数设置，以确保比较公平。在收集示范时，我们从四名医学生（每人 25 个）那里录制了 100 个专家人类示范；他们接受过机器人手术训练，并能熟练使用我们的模拟器。此外，我们从四名工程专业学生（每人 25 个）那里录制了 100 个非专家人类示范；他们有使用我们模拟器的实践经验。同时，我们通过专门设计的路径规划脚本生成了另外 100 个脚本示范，以供比较。为保证公平，示范的平均完成步数在数值上接近：人类专家约为 113 步，非专家为 117 步，脚本为 109 步。类似地，人类专家、非专家和脚本示范的平均轨迹距离分别为 $12.1cm$ 、 $12.5cm$ 和 $11.7cm$ 。

IV. 结果

当前通过基于熟练度的渐进训练（PBP） [61, 62] 评估手术模拟训练的方案，强调学习效率对于评估手术技能水平的重要性，这与先前关于强化学习的研究 [63, 64] 中的理念相呼应：这些研究主张在训练早期阶段评估所提出算法的样本效率。鉴于此，为了分析学习过程和策略性能，我们分析了策略的训练曲线，以及在 50 个训练轮次（早期训练阶段）和 100 个训练轮次（所有结果均不再呈现上升趋势时）的结果。

为了展示策略的学习效率，我们使用四个随机种子对其进行训练，并绘制平均成功率，结果如图 5所示。我们观察到，在整个训练过程中，从人类专家示范中学习的成功率始终高于从非专家示范和脚本示范中学习的成功率。具体而言，基于人类专家示范的强化学习策略的学习曲线偏差小于非专家示范和脚本示范，表明策略学习的稳定性有所提升。经过最初 50 个训练轮次后，*NeedlePick* 在从人类专家示范中学习时，成功率达到 $89.7\% \pm 6.4\%$ ，优于从非专家示范中学习（$66.3\% \pm 12.7\%$ 成功率）23.4%，也优于脚本示范（$43.2\% \pm 21.9\%$ 成功率）46.5% （如表 I）。由专家示范学到的策略也能实现更快的完成步数（39.2 ± 14.5），优于非专家（59.5 ± 17.1）和脚本（101.8 ± 27.1）。此外，我们计算了平均动作经济性，发现使用专家示范时运动轨迹更短， $12.2cm \pm 1.3cm$；使用非专家示范时为 $14.2cm \pm 3.1cm$，使用脚本时为 $15.0cm \pm 2.5cm$ 。训练 100 轮后，基于人类示范的学习结果比非专家和脚本更稳定，方差也更小。同时，基于专家示范学到的策略始终优于

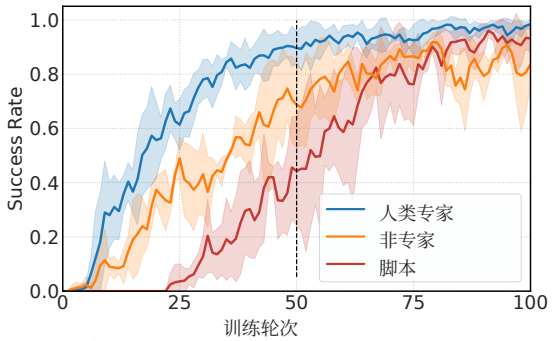

图5. 使用滑动窗口平滑处理的*NeedlePick* 成功率学习曲线。结果表明，通过专家示范学习的策略始终优于通过非专家示范或脚本示范学习的策略。阴影区域表示四个随机种子下的标准差。

表 I模拟器中*NeedlePick* 的评估结果。

演示类型	50 个训练轮次			100 个训练轮次		
	脚本	非专家	人类专家	脚本	非专家	人类专家
成功率 / % (↑)	43.2 (±21.9)	66.3 (±12.7)	**89.7** (±6.4)	93.2 (±6.9)	86.7 (±2.4)	**99.3** (±0.1)
完成步数 (↓)	101.8 (±27.1)	59.5 (±17.1)	**39.2** (±14.5)	24.0 (±10.4)	21.4 (±5.9)	**15.0** (±5.5)
动作经济性 / 厘米 (↓)	15.0 (±2.5)	14.2 (±3.1)	**12.2** (±1.3)	13.5 (±1.6)	12.6 (±2.9)	**11.6** (±1.1)

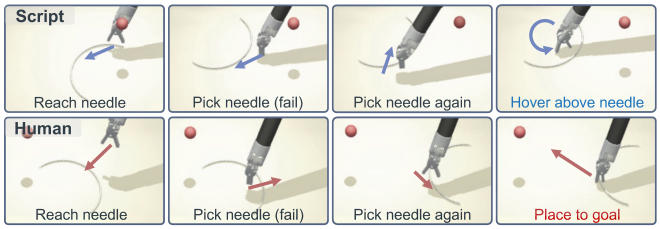

图6. 一个示例案例表明，在早期训练阶段，从人类示范中学习的策略能够掌握处理失败尝试的技能，并随后完成任务；而从脚本示范中学习的策略则无法做到这一点。

基于非专家或脚本学到的策略在所有指标上的表现。结果表明，专家人类示范提升了策略学习效率。

除了上述定量结果外，我们还进一步分析了 50 个训练轮次时的定性结果。我们以视频形式展示所有案例，并比较从人类（包括专家和非专家）以及脚本中学习到的策略所产生的动作。我们发现了一个显著现象，如图 6 所示。第一行展示了一个从脚本中学习到的策略的失败案例：器械首先接近针并尝试将其拾起，但未能抓住针。第一次尝试失败后，该策略在随后的步骤中仍未能再次抓住针。相比之下，从人类示范中学习时，策略在首次抓取失败后能够再次抓住针并完成任务。结果表明，在早期学习阶段，使用人类示范有助于策略更快掌握处理失败情况的技能，这证明了以人类示范的形式将人为因素纳入策略学习的有效性。

V. 结论与讨论

本文提出了一个基于现有 SurRoL 模拟器、用于人在回路具身智能的交互平台。我们新增了多项功能以增强人机交互，包括触觉交互界面、物理仿真，以及逼真度更高的场景渲染。我们还共同开发了五项具有代表性的新外科训练任务，为未来开展人在回路的外科手术任务学习研究做准备。我们首先进行了基础实验，以人类示范的形式验证将人类纳入具身智能的构想。令人鼓舞的结果证明了我们所提方法的有效性，并为相关研究方向奠定了基础。未来的潜在研究方向包括：在基于人类示范进行策略学习时考虑手术安全 [65, 66] ，研究从人类反馈中学习的有效性，以及人机协作 [22]。我们期待未来开展广泛研究，探索人为因素如何发挥作用，并借助我们为手术机器人提供的开源具身人工智能平台，推动手术机器人学习的发展。

参考文献

[1] R. H. Taylor、N. Simaan、A. Menciassi 和 G.-Z. Yang，《手术机器人技术与计算机集成介入医学》，Proceedings of the IEEE，第 110 卷，第 7 期，第 823–834 页，2022 年。[2] C. D'Ettorre、A. Mariani、A. Stilli、P. Valdastri、A. Deguet、P. Kazanzides、R. H. Taylor、G. S. Fischer、S. P. DiMaio、A. Menciassi 等人，《加速手术机器人研究：回顾 dVRK 十年来的研究》，arXiv 预印本 arXiv:2104.09869，2021 年。[3] G.-Z. Yang、J. Bellingham 等人，《Science Robotics 面临的重大挑战》，Science Robotics，第 3 卷，第 14 期，文章编号 eaar7650，2018 年。[4] L. Maier-Hein、M. Eisenmann、D. Sarikaya、K. M¨arz、T. Collins、A. Malpani、J. Fallert、H. Feussner、S. Giannarou、P. Mascagni 等人，《手术数据科学——从概念走向临床转化》，Medical Image Analysis，第 76 卷，文章编号 102306，2022 年。[5] J. Duan、S. Yu、H. L. Tan、H. Zhu 和 C. Tan，《具身人工智能综述：从模拟器到研究任务》，IEEE Transactions on Emerging Topics in Computational Intelligence，2022 年。[6] M. Savva、A. Kadian、O. Maksymets、Y. Zhao、E. Wijmans、B. Jain、J. Straub、J. Liu、V. Koltun、J. Malik 等人，《Habitat：具身人工智能研究平台》，载于 IEEE/CVF 国际计算机视觉会议论文集，2019 年，第 9339–9347 页。[7] R. Ramrakhya、E. Undersander、D. Batra 和 A. Das，《Habitat-web：从大规模人类示范中学习具身物体搜索策略》，载于 IEEE/CVF 计算机视觉与模式识别会议论文集，2022 年，第 5173–5183 页。[8] C. Li、F. Xia、R. Mart´ın-Mart´ın、M. Lingelbach、S. Srivastava、B. Shen、K. Vainio、C. Gokmen、G. Dharan、T. Jain 等人，《iGibson 2.0：面向机器人日常家务任务学习的以物体为中心的仿真》，arXiv 预印本 arXiv:2108.03272，2021 年。[9] F. Xiang、Y. Qin、K. Mo、Y. Xia、H. Zhu、F. Liu、M. Liu、H. Jiang、Y. Yuan、H. Wang、L. Yi、A. X. Chang、L. J. Guibas 和 H. Su，《SAPIEN：基于仿真部件的交互环境》，载于 IEEE 计算机视觉与模式识别会议，2020 年 6 月。[10] S. Bonne、W. Panitch、K. Dharmarajan、K. Srinivas、J.-L. Kincade、T. Low、B. Knoth、C. Cowan、D. Fer、B. Thananjeyan 等人，《面向网络服务质量变化情况下远程手术的数字孪生框架》，载于 CASE。IEEE，2022 年，第 1325–1332 页。[11] H. Shu、R. Liang、Z. Li、A. Goodridge、X. Zhang、H. Ding、N. Nagu- ruru、M. Sahu、F. X. Creighton、R. H. Taylor 等人，《Twin-S：颅底手术的数字孪生》，arXiv 预印本 arXiv:2211.11863，2022 年。[12] J. Ibarz、J. Tan、C. Finn、M. Kalakrishnan、P. Pastor 和 S. Levine，《如何通过深度强化学习训练机器人：我们的经验教训》，The International Journal of Robotics Research，第 40 卷，第 4–5 期，第 698–721 页，2021 年。[13] E. Kolve、R. Mottaghi、W. Han、E. VanderBilt、L. Weihs、A. Herrasti、D. Gordon、Y. Zhu、A. Gupta 和 A. Farhadi，《AI2-THOR：面向视觉人工智能的交互式三维环境》，arXiv，2017 年。[14] P. A. Kenney、M. F. Wszolek 等人，《DV-Trainer（一种新型机器人手术虚拟现实模拟器）的表面效度、内容效度和结构效度》，Urology，第 73 卷，第 6 期，第 1288–1292 页，2009 年。[15] G. Whittaker、A. Aydin、N. Raison、F. Kum、B. Challacombe、M. S. Khan 等人，《Robotix Mentor 机器人手术模拟器的验证》，《腔内泌尿学杂志》，第 30 卷，第 3 期，第 338–346 页，2016 年。[16] F. Richter 等人，《面向手术机器人的开源强化学习环境》，arXiv 预印本 arXiv:1903.02090，2019 年。[17] E. Tagliabue、A. Pore、D. Dall'Alba、E. Magnabosco、M. Piccinelli 和 P. Fiorini，《用于自主机器人手术操作任务学习的软组织仿真环境》，载于智能机器人与系统国际会议（IROS）。IEEE，2020 年，第 3261–3266 页。[18] A. Munawar、J. Y. Wu 等人，《用于机器人辅助外科缝合学习与练习的开放式仿真环境》，IEEE RAL，第 7 卷，第 2 期，第 3843–3850 页，2022 年。[19] G.-Z. Yang、J. Cambias、K. Cleary、E. Daimler、J. Drake、P. E. Dupont、N. Hata、P. Kazanzides、S. Martel、R. V. Patel 等人，《医疗机器人——提升自主性水平所涉及的监管、伦理和法律考量》，文章编号 eaam8638，2017 年。

[20] H. Saeidi、J. D. Opfermann、M. Kam、S. Wei、S. L´eonard、Hsieh 等人，《用于肠吻合术的自主机器人腹腔镜手术》，Science Robotics，第 7 卷，第 62 期，文章编号 eabj2908，2022 年。[21] D. Zhang、Y. Guo、J. Chen、J. Liu 和 G.-Z. Yang，《用于机器人辅助手术显微术的手持式主控制器》，载于智能机器人与系统国际会议。IEEE，2019 年，第 394–400 页。[22] D. Zhang、Z. Wu、J. Chen、R. Zhu、A. Munawar、B. Xiao 等人，《基于上下文感知仿真到现实适应的手术机器人人机共享控制》，arXiv 预印本 arXiv:2204.11116，2022 年。[23] J. Xu、B. Li、B. Lu、Y.-H. Liu、Q. Dou 和 P.-A. Heng，《SurRoL：以强化学习为核心且兼容 dVRK 的开源手术机器人学习平台》，载于 IROS。IEEE，2021 年。[24] K. Ehsani、W. Han、A. Herrasti、E. VanderBilt、L. Weihs、E. Kolve、A. Kembhavi 和 R. Mottaghi，《ManipulaTHOR：视觉物体操作框架》，CVPR，第 4495–4504 页，2021 年。[25] C. Gan、J. Schwartz、S. Alter、M. Schrimpf、J. Traer、J. De Freitas、J. Kubilius 等人，《ThreeDWorld：交互式多模态物理仿真平台》，arXiv 预印本 arXiv:2007.04954，2020 年。[26] H. Fu、W. Xu、H. Xue、H. Yang、R. Ye、Y. Huang 等人，《RFUniverse：面向日常家务任务、基于物理且以动作为中心的交互环境》，arXiv 预印本 arXiv:2202.00199，2022 年。[27] A. Mandlekar、D. Xu、J. Wong、S. Nasiriany、C. Wang、R. Kulkarni、L. Fei-Fei、S. Savarese、Y. Zhu 和 R. Mart´ín-Mart´ín，《对于从离线人类示范中学习机器人操作而言，哪些因素至关重要》，arXiv 预印本 arXiv:2108.03298，2021 年。[28] K. L. Schwaner、I. Iturrate、J. K. Andersen、P. T. Jensen 和 T. R. Savarimuthu，《基于示范技能学习的自主双手外科缝合》，载于智能机器人与系统国际会议（IROS）。IEEE，2021 年，第 4017–4024 页。[29] B. Thananjeyan、A. Garg、S. Krishnan、C. Chen、L. Miller 和 K. Goldberg，《利用深度强化学习策略在二维正交各向异性纱布上进行多边手术图案切割与张紧》，载于 ICRA。IEEE，2017 年，第 2371–2378 页。[30] C. D'Ettorre、S. Zirino、N. N. Dei、A. Stilli、E. De Momi 和 D. Stoy- anov，《利用基于上下文的强化学习进行术中器官操作》，《国际计算机辅助放射学与外科学杂志》，第 17 卷，第 8 期，第 1419–1427 页，2022 年。[31] J. Hua、L. Zeng、G. Li 和 Z. Ju，《机器人学习：深度强化学习、模仿学习与迁移学习》，Sensors，第 21 卷，第 4 期，文章编号 1278，2021 年。[32] M. Yip 和 N. Das，《手术中的机器人自主性》，载于《医学机器人百科全书：第 1 卷 微创手术机器人学》。World Scientific，2019 年，第 281–313 页。[33] P. M. Scheikl、E. Tagliabue、B. Gyenes、M. Wagner、D. Dall'Alba、P. Fiorini 和 F. Mathis-Ullrich，《用于机器人辅助手术中可变形物体操作的视觉强化学习仿真到现实迁移》，IEEE RAL，第 8 卷，第 2 期，第 560–567 页，2022 年。[34] X. Tan、C.-B. Chng、Y. Su、K.-B. Lim 和 C.-K. Chui，《利用深度强化学习进行腹腔镜手术机器人辅助训练》，IEEE《机器人与自动化快报》，第 4 卷，第 2 期，第 485–492 页，2019 年。[35] A. Pore、E. Tagliabue、M. Piccinelli、D. Dall'Alba、A. Casals 和 P. Fiorini，《从示范中学习自主软组织牵拉》，载于 2021 ISMR。IEEE，2021 年，第 1–7 页。[36] B. Li、R. Wei、J. Xu、B. Lu 等人，《有限示范下用于机器人手术腹腔镜控制的三维感知模仿学习》，载于 ICRA。IEEE，2022 年，第 7664–7670 页。[37] F. Richter、J. Lu、R. K. Orosco 和 M. C. Yip，《部分可见运动链条件下的机器人手术器械跟踪：统一方法》，IEEE《机器人学汇刊》，第 38 卷，第 3 期，第 1653–1670 页，2021 年。[38] M. A. Arteaga、A. Guti´errez-Giles 等人，《Geomagic Touch 触觉设备》，载于《机器人操纵器控制中的局部稳定性与最终有界性》。Springer，2022 年，第 361–374 页。[39] A. Munawar、Z. Li、P. Kunjam、N. Nagururu、A. S. Ding、P. Kazanzides、T. Looi、F. X. Creighton、R. H. Taylor 和 M. Un- berath，《用于协同手术训练与数据生成的虚拟现实》，Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization，第 10 卷，第 4 期，第 366–374 页，2022 年。[40] Simulated Surgical Systems，《机器人手术模拟器（RoSS™）》，2020 年，http://simulatedsurgicals.com/projects/ross/。[41] K. H. Ang、G. Chong 和 Y. Li，《 PID 控制系统分析、设计与技术》，IEEE《控制系统技术汇刊》，第 13 卷，第 4 期，第 559–576 页，2005 年。[42] D. M. Beazley 等人，《 SWIG：一种易于使用的工具，用于将脚本语言与 C 和 C++ 集成》，载于 Tcl/Tk研讨会，1996 年。[43] K. Mamou、E. Lengyel 和 A. Peters，《体积层次近似凸分解”，载于《Game Engine Gems 3》。AK Peters，2016年，第141–158页。[44] M. Goslin和M. R. Mine，“Panda3D图形引擎”，Computer，第37卷，第10期，第112–114页，2004年。[45] G. Bishop和D. M. Weimer，“快速Phong着色”，ACM SIGGRAPH Computer Graphics，第20卷，第4期，第103–106页，1986年。[46] H. Gouraud，“曲面的连续着色”，IEEE计算机汇刊，第100卷，第6期，第623–629页，1971年。[47] Y. Long、J. Cao、A. Deguet、R. H. Taylor和Q. Dou，“将人工智能与增强现实整合到机器人手术中：采用外科教育场景的初步dVRK研究”，载于2022年国际医疗机器人研讨会（ISMR），2022年，第1–8页。[48] R. Smith、V. Patel和R. Satava，“机器人手术基础：基于14个学会共识框架的机器人手术基本技能课程，该框架涵盖结局测量指标和课程开发”，《国际医学机器人与计算机辅助手术杂志》，第10卷，第3期，第379–384页，2014年。[49] D. Sanford、R. Ma、A. Ghoreifi等，“虚拟现实模拟与实时手术中的缝合技术技能评估分数之间的关联”，《腔内泌尿学杂志》，编号ja，2022年。[50] C. Perrenot、M. Perez、N. Tran、J.-P. Jehl、J. Felblinger、L. Bresler和J. Hubert，“虚拟现实模拟器dv-trainer®是评估机器人手术技能的有效工具”，《外科内镜》，第26卷，第9期，第2587–2593页，2012年。[51] A. Cowan、J. Chen、S. Mingo、S. S. Reddy、R. Ma等，“虚拟现实与干实验室模型：比较机器人模拟训练中的自动化性能指标和认知工作负荷”，《腔内泌尿学杂志》，第35卷，第10期，第1571–1576页，2021年。[52] S. Levine和V. Koltun，“引导式策略搜索”，载于国际机器学习会议。PMLR，2013年，第1–9页。[53] M. Vecerik、T. Hester、J. Scholz、F. Wang、O. Pietquin、B. Piot、N. Heess、T. Roth¨orl、T. Lampe和M. Riedmiller，“利用示范，通过稀疏奖励在机器人学问题上进行深度强化学习”，arXiv预印本arXiv:1707.08817，2017年。[54] H. Su、A. Mariani、S. E. Ovur、A. Menciassi、G. Ferrigno和E. De Momi，“探索机器人辅助微创手术中的示范教学”，TASE，第18卷，第2期，第484–494页，2021年。[55] T. P. Lillicrap、J. J. Hunt、A. Pritzel、N. Heess、T. Erez、Y. Tassa、D. Silver和D. Wierstra，“深度强化学习下的连续控制”，载于ICLR，2016年。[56] V. G. Goecks、G. M. Gremillion、V. J. Lawhern、J. Valasek和N. R. Waytowich，“结合行为克隆与强化学习，以提升稠密和稀疏奖励环境下的性能”，载于第19届自主智能体与多智能体系统国际会议论文集，2020年，第465–473页。[57] M. Bain和C. Sammut，“行为克隆框架”，载于Machine Intelligence 15，1995年，第103–129页。[58] M. Andrychowicz、F. Wolski、A. Ray、J. Schneider、R. Fong、P. Welinder、B. McGrew、J. Tobin、O. P. Abbeel和W. Zaremba，“后见经验回放”，载于NeurIPS，2017年。[59] T. Huang、K. Chen、B. Li、Y.-H. Liu和Q. Dou，“结合高效探索的示范引导强化学习，用于手术机器人的任务自动化”，IEEE ICRA，2023年。[60] P. Dhariwal、C. Hesse、O. Klimov、A. Nichol、M. Plappert、A. Radford、J. Schulman、S. Sidor、Y. Wu和P. Zhokhov，“OpenAI Baselines”，https://github.com/openai/baselines，2017年。[61] R. M. Satava，“外科模拟的未来”，载于《综合医疗模拟：外科与外科亚专科》，2019年。[62] A. G. Gallagher等，《外科模拟基础：原理与实践》。Springer Science & Business Media，2011年。[63] M. Laskin、A. Srinivas和P. Abbeel，“CURL：用于强化学习的对比式无监督表征”，载于国际机器学习会议。PMLR，2020年，第5639–5650页。[64] M. Laskin、K. Lee、A. Stooke、L. Pinto、P. Abbeel和A. Srinivas，“基于增强数据的强化学习”，《神经信息处理系统进展》，第33卷，第19 884–19 895页，2020年。[65] B. Thananjeyan、A. Balakrishna、U. Rosolia、F. Li、R. McAllister、J. E. Gonzalez、S. Levine、F. Borrelli和K. Goldberg，“基于示范的安全增强价值估计（SAVED）：适用于稀疏成本机器人任务的安全深度基于模型的强化学习”，IEEE《机器人与自动化快报》，第5卷，第2期，第3612–3619页，2020年。[66] A. Pore、D. Corsi、E. Marchesini、D. Dall' Alba、A. Casals、A. Farinelli和P. Fiorini，“利用形式化验证实现组织牵拉的安全强化学习，用于自主机器人辅助手术”，载于IROS。IEEE，2021年，第4025–4031页。