Autonomous Robots 稿件编号（将由编辑
插入）

关节物体的在线估计与操作

Russell Buchanan · Adrian Röfer · João Moura · Abhinav Valada ·Sethu Vijayakumar

arXiv:2601.01438v1 [cs.RO] 4 Jan 2026

收稿日期：日期 / 接受日期：日期

摘要 从冰箱到厨房抽屉，人们每天都能轻松地与关节物体交互，完成家务。为了自动化这些任务，服务机器人必须能够操作任意关节物体。近期的深度学习方法已证明，可以通过视觉预测关节物体可供性的有价值先验。相比之下，许多其他研究通过观察关节运动来估计物体关节结构，但这要求机器人已经具备操作该物体的能力。本文提出一种新颖方法，将这些方法结合起来：利用因子图在线估计关节运动，并在交互过程中融合学习得到的视觉先验与本体感知，将其转化为基于螺旋理论的关节运动解析模型。采用我们的方法，机器人系统能够在触碰物体之前，先通过视觉对关节运动进行初步预测，随后在操作过程中根据运动学和力感知快速更新估计结果。我们通过仿真和真实世界中的机器人操作实验，对该方法进行了全面评估。我们展示了若干闭环估计与操作实验，其中机器人能够打开此前未见过的抽屉。在真实硬件实验中，机器人自主打开未知关节物体的成功率达到 75%。

R. Buchanan RIPL-Lab，滑铁卢大学，加拿大 电子邮件：russell.buchanan@uwaterloo.ca A. Röfer ·A. Valada 机器人学习实验室，弗赖堡大学，德国 电子邮件：{aroefer,valada}@cs.uni-freiburg.de J. Moura · S. Vijayakumar 信息学院，爱丁堡大学，英国 电子邮件：{joao.moura,sethu.vijayakumar}@ed.ac.uk

关键词 操作感知 · 状态估计 · 深度学习

1 引言

若要让服务机器人协助人类完成烹饪和清洁等日常任务，它们必须能够与洗碗机、门和抽屉等常见关节物体互动并对其进行操作。要操作这些物体，机器人需要理解其关节运动，可以采用解析模型（例如转动、移动或螺旋关节），也可以采用通过神经网络隐式学习得到的模型。近期许多研究展示了如何利用点云测量，通过深度神经网络预测关节物体的可供性。为此，研究人员在仿真中渲染常见的家用关节物体，并将随机状态作为训练样例。由于大多数常见家用物体的关节运动都具有良好的可重复性，例如冰箱门，因此学习到的模型能够有效泛化至真实数据。然而，仅凭视觉数据预测关节运动往往并不可靠。

例如，图1 中的橱柜有四扇门，关闭时看起来完全相同。仅凭视觉，人类或机器人都不可能可靠地预测其关节运动。然而，一旦人与机器人与其进行交互，就会发现这些门的开启方式完全不同。这对于完全依赖视觉来理解关节运动、且无法在线更新其关节估计的机器人系统而言，是一项挑战。在本文中，我们提出了一种新颖的方法，通过联合优化视觉、力和运动学感知，实现对关节式物体的在线估计。

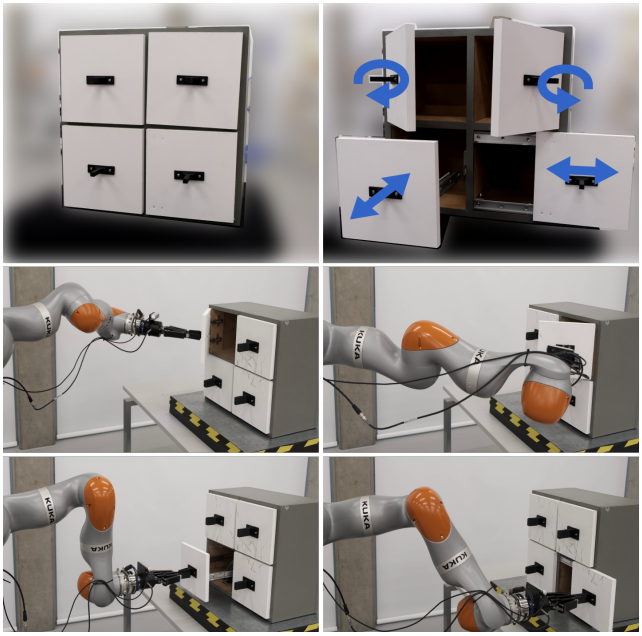

图1 顶行：一个橱柜，其柜门在视觉上完全相同。只有打开后，才能显现出它们不同的关节运动方式。仅凭视觉检查，无法预测每扇柜门的开启方式。**中间行和底行：**机器人自主打开每扇橱柜门，同时在线估计关节运动。

另一个研究分支聚焦于关节运动的概率估计。这些研究通常使用关节运动的解析模型，并通过观测物体在交互过程中的运动来估计物体的关节结构。然而，这些研究在很大程度上依赖于对关节运动的良好初始猜测，以便机器人能够开始移动物体。

在本文中，我们在 Buchanan 等人(2024年) 所介绍的前期工作基础上进行了显著改进；在该工作中，我们首次研究了关节式物体的估计问题。其中第一项改进是用于可供性预测的神经网络，该网络将预测中的不确定性纳入其中；第二项改进是采用一种全新的方法，将学习到的关节运动可供性纳入因子图，从而为关节运动提供良好的初始估计。我们还将运动学信息和力感知纳入因子图，使其能够在交互过程中在线更新估计结果。最终得到的是一个稳健的多模态关节结构估计框架。本文的贡献如下：

– 我们提出了一种在因子图框架中结合视觉与本体感知来在线估计关节参数的方法。相较于我们先前的工作，我们新增了一个能够感知不确定性的关节因子，从而提升了关节运动预测的鲁棒性。

– 此外，我们还引入了一种用于关节估计的新型力感知因子。– 我们展示了完整的系统集成，并实现了共享自主，以打开从未见过的关节式物体。– 我们通过广泛的真实世界实验验证了系统，并在闭环中运行估计，打开了视觉上难以判别关节运动方式的物体。与 Buchanan 等人(2024年) 相比，我们展示了性能提升：成功打开了图1中橱柜的所有门，而这在此前尚无法实现。

2 相关工作

本节总结关节式物体估计方面的相关工作。这是一个具有挑战性的问题，计算机视觉和机器人学领域已从许多不同角度对其展开研究。本文关注关节式物体的机器人操作。因此，在下面的章节中，我们首先介绍交互式感知方面的相关工作（Bohg 等人 (2017)）；交互式感知长期以来一直被用于关节式物体的估计与操作。随后，我们简要介绍与基于视觉的关节运动预测最相关的最新深度学习方法。最后，我们讨论一些将不同系统集成在一起、用于机器人实验的方法。

2.1 交互式感知

交互式感知的基本原理是：机器人与环境交互以收集信息时，能够显著促进机器人的感知。这一原理已广泛应用于关节物体的估计，因为机器人抓取关节物体并开始移动后，便能从多个信息来源推断其关节运动参数。如今，很少有研究利用本体感觉来估计关节运动，这主要是因为识别初始抓取点和拉动方向颇具挑战。2010年，Jain 和 Kemp(2010年)通过假设抓取姿态和初始开门力向量已知，简化了这一问题。这使得他们的方法能够仅利用力与运动学感知，自主打开橱柜、抽屉等多种日常物体。他们展示了，只要机器人与关节物体进行物理交互，并获得良好的初始运动方向，仅凭本体感觉通常就足以操控大多数关节式物体。

在近期研究中，更常见的做法是将本体感觉与视觉融合，或者仅使用视觉。Sturm 等人（2011）提出了一个概率框架，用于根据观测到的关节物体运动部件的位姿轨迹，最大化关节类型和关节

参数的概率（例如，橱柜的柜门）。为了跟踪运动部件的轨迹，他们展示了多种不同的感知方法，包括对基准标记进行视觉跟踪、基于深度图像的无标记跟踪，以及运动学感知。他们将该方法与Jain 和 Kemp （2010年）的方法集成，用于真实机器人实验，并指定初始抓取点和运动方向。

后续研究转向视觉感知，利用光束法平差在整个交互过程中跟踪视觉特征（Katz 等人 (2014年）。Martín- Martín 和 Brock(2022)还提出了一种框架，可根据视觉和触觉感知进行在线估计。与以往的工作一样，他们在机器人与物体交互时跟踪视觉特征的运动。随后将这些信息与力/力矩感知、软体机械手的触觉感知以及末端执行器位姿测量进行融合。Heppert 等人 (2022)提出了一种视觉神经网络，用于通过视觉跟踪物体各部分的运动。跟踪到的位姿通过因子图连接起来，以估计关节参数。在实验中，他们估计了未知的关节运动；然而，他们的控制器使用预先设定的运动来打开物体，从而为估计器提供了足够的信息。

所有这些交互式感知方法都需要预先确定抓取点，并对关节运动作出良好的初始估计。在我们先前的工作（Buchanan 等人(2024年）中，我们使用因子图将学习得到的视觉预测与运动学感知相结合。这使得我们的方法能够自动初步估计开启方向，并在交互过程中在线更新估计。然而，在这项早期工作中，我们只能演示打开需要拉动的物体，无法演示打开滑动门，例如图 1右下方的门。

在本研究中，与 Heppert 等人(2022)和 Buchanan 等人 (2024年)类似，我们使用因子图融合部件位姿测量值，以估计关节螺旋模型。然而，与这些先前的研究不同，我们利用基于视觉感知的基于不确定性的深度神经网络预测，为机器人提供关节运动的初始估计。随后结合力感知和运动学感知对其进行更新，从而实现任意关节运动的开启，包括滑动门。通过结合交互式感知与基于学习的预测，我们能够在开启未知关节式物体的同时进行闭环控制和估计。

2.2 基于学习的铰接预测

近年来，许多研究仅使用视觉信息结合深度学习来预测关节运动，无需与物体进行交互（Li 等人（2020年）；Jain 等人（2021年）；Jiang 等人（2022a年））。通常，这些研究使用仿真数据集，例如 Partnet-Mobility 数据集（Mo 等人（2019年）），其中包含常见关节式物体的示例。由于许多常见的家用物体具有可预测的关节运动（例如冰箱门），这些研究假设，在大多数情况下，可以通过视觉检查预测关节运动。

早期基于学习的研究会明确对物体进行分类，以预测其关节运动；但近来，研究重点转向学习不依赖类别的关节运动可供性（Mo 等人 (2021年); Xu 等人 (2022);Eisner* 等人 (2022); Zhang 等人 (2022)），这类方法描述用户如何在不通过视觉对物体进行分类的情况下与物体交互。这通常通过一个归一化向量进行参数化，该向量描述物体关节部件上一点的运动。Bahl 等人 (2023)还使用神经网络，根据人类示范预测物体的抓取姿态和开启轨迹。一些近期的基于学习的研究也纳入了交互。 Jiang 等人 (2022b)使用了在人与目标关节式物体交互前后采集的点云数据。Nie 等人(2023)提出了一种方法，既能预测关节运动，也能提出一种交互方式，以便观察运动并更新关节估计。

所有这些基于学习的方法都有类似的局限性，阻碍了它们在真实机器人上的应用。它们只使用视觉信息，且计算需求很高，因此无法进行在线估计。因此，在使用真实数据时，它们通常只获取物体的单次“快照”，并进行单次推理。如果预测有误，就无法更新估计。此外，由于这些方法依赖于将物体与过往训练经验进行视觉相似性识别，因此在处理图1所示的物体时表现不佳，因为该物体没有任何指示其开启方式的视觉指示特征。如果预测有误，由于缺乏在线估计，这些方法只能依靠高度柔顺的控制器来弥补误差。

2.3 系统

我们的工作不仅提供了一种估计方法，还提供了一个利用共享自主开启关节物体的完整系统。因此，我们也讨论了一些开发关节式物体操作系统的相关工作。Mittal 等人 (2021年) 介绍了一种全身移动操作系统。他们使用了

该网络Li 等人 (2020年) 提出的方法。这意味着他们的方法需要事先了解机器人与之交互的物体类别。此外，在他们的方法中，他们会在交互前进行单次预测，然后依靠控制器鲁棒性来弥补预测错误。

Schiavi 等人提出了一种闭环学习估计方法 Schiavi 等人 (2023)。该方法在交互过程中多个时间步，根据视觉信息估计关节运动可供性。基于采样的控制器求解最优开启轨迹。开启物体时，由于扭矩限制，物体会停止运动；机器人便松开物体并移动到一个能够再次查看完整物体的位姿，然后基于新的视觉信息重新估计关节运动。松开物体以便重新查看会减慢开启过程，并且假设物体的门在松手后不会突然关上或自行打开。这些方法严重依赖稳健的柔顺控制器来弥补关节估计中的误差。相比之下，我们的方法可在交互过程中无缝更新关节模型估计，从而能够采用简单得多的运动生成与控制方法。

3 螺旋理论背景

螺旋理论是对扭量的几何诠释，可用于表示任意刚体运动（沙勒定理）（Murray 等人 (1994)）。螺旋运动由扭量 $\xi = (\mathbf{v}\ \boldsymbol{\omega})$，其中 $\mathbf{v}\ \boldsymbol{\omega} \in \mathbb{R}^3$。变量 $\mathbf{v}$ 表示线性运动，而 $\boldsymbol{\omega}$表示旋转。我们可以使用 $\hat{}\,\mathrm{SE}(3)$将其转换到切空间 ξ，即

$$\hat{\xi} = \begin{bmatrix} \hat{\boldsymbol{\omega}} & \mathbf{v} \\ 0 & 0 \end{bmatrix} \in \mathfrak{se}(3), \tag{1}$$

$\hat{}$ 其中帽算子 $(\cdot)$ 定义为：

$$\hat{\boldsymbol{\omega}} = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}. \tag{2}$$

在螺旋理论中，ξ 是运动方向的参数化表示，而 θ 是表示运动量的带符号标量。在纯旋转情况下，θ 的单位为弧度；在纯平移情况下，单位为米。切空间方程（1）可通过指数映射转换为齐次变换 $\hat{}\,\mathbf{T}(\xi, \theta) \in \mathrm{SE}(3)$，使用 Murray 等人（$\exp: \mathfrak{se}(3) \to \mathrm{SE}(3)$ 提出的 Murray 等人(1994）中的指数映射，其中

$$\exp(\hat{\xi}\theta) = \begin{bmatrix} \exp(\hat{\boldsymbol{\omega}}\theta) & (\mathbf{I} - \exp(\hat{\boldsymbol{\omega}}\theta))(\boldsymbol{\omega} \times \mathbf{v}) + \boldsymbol{\omega}\boldsymbol{\omega}^T\mathbf{v}\theta \\ 0 & 1 \end{bmatrix}, \tag{3}$$

而 $\exp(\hat{}\,\boldsymbol{\omega}\theta)$ 则使用 Rodriguez 公式求解：

$$\exp(\hat{\boldsymbol{\omega}}\theta) = \mathbf{I} + \hat{\boldsymbol{\omega}}\theta + \frac{(\hat{\boldsymbol{\omega}}\theta)^2}{2!} + \frac{(\hat{\boldsymbol{\omega}}\theta)^3}{3!} + \ldots \tag{4}$$

$$= \mathbf{I} + \hat{\boldsymbol{\omega}}\sin\theta + \boldsymbol{\omega}^2(1 - \cos\theta). \tag{5}$$

若定义一个固定的世界坐标系 W，则关节物体运动部件（例如柜门）所在的坐标系记为 A，它们之间的齐次变换记为 $\mathbf{T}_{\mathtt{WA}} \in \mathrm{SE}(3)$。物体的另一个非运动部件（例如柜体底座）所在的坐标系记为 B，其在 W 中的位姿定义为 $\mathbf{T}_{\mathtt{WB}} \in \mathrm{SE}(3)$。由于 $\exp(\hat{}\,\boldsymbol{\omega}\theta) \in \mathrm{SE}(3)$ 表示物体底座 B 与关节部件 A 之间的齐次变换，因此可将螺旋变换表示为：

$$\mathbf{T}_{\mathtt{BA}}(\hat{\xi}, \theta) = \exp(\hat{\xi}\theta), \tag{6}$$

其中 $\theta \in \mathbb{R}$ 为关节配置。随后，这两个物体部件通过以下方式连接：

$$\mathbf{T}_{\mathtt{WA}} = \mathbf{T}_{\mathtt{WB}}\mathbf{T}_{\mathtt{BA}}(\hat{\xi}, \theta). \tag{7}$$

4 预备知识

本研究旨在通过视觉和本体感知，在线估计单个关节的最大后验（MAP）状态。我们将时刻 t 的状态 $\boldsymbol{x}(t)$定义为

$$\boldsymbol{x}(t) \coloneqq [\xi, \theta(t)] \in \mathbb{R}^7, \tag{8}$$

其中，ξ 是我们假设始终保持不变的螺旋参数，而 $\theta(t)$ 是时间 t的关节配置。我们假设该物体仅由通过单个关节连接的两个部分组成。这涵盖了绝大多数关节式物体，因此是一个合理的简化。我们将每个部分在世界坐标系 W 中的位姿定义为 $\mathbf{T}_{\mathtt{WA}}$，$\mathbf{T}_{\mathtt{WB}} \in \mathrm{SE}(3)$，其中$\mathbf{T}_{\mathtt{WB}}$ 是静止的基座 部分，而$\mathbf{T}_{\mathtt{WA}}$是机器人抓取的关节运动部分，例如门。

我们联合估计时间索引为 k 的 K 个状态和部件位姿；因此，所有估计状态和关节部件位姿的集合可写为 $\mathcal{X} = \{\boldsymbol{x}_k, \mathbf{T}_{\mathtt{A}k}, \mathbf{T}_{\mathtt{B}k}\}_{k \in \mathtt{K}}$。为简洁起见，省略世界参考坐标系。我们融合来自三个来源的测量值：初始视觉检查得到的点云测量值、腕部安装的六轴力/力矩传感器得到的力测量值，以及机器人手臂关节编码器得到的运动学测量值。我们使用 P 个点云，每个点云都与关于 ξ 的一个预测相关联。在不失一般性的情况下，我们将 $\mathtt{P} = 1$ 设为在开始时进行一次视觉测量。在未来的工作中，

我们计划添加多个基于视觉的预测，使其能够在操作过程中异步加入因子图。力测量值在时间索引 f 处加入，最多加入 F 个测量值。我们在时间 k 使用 K 个运动学测量值，每个测量值都与一个状态估计相关联。只有当机器人与物体接触且关节部件已移动一定距离 d 后，才会选择时间 k，以避免获取过多测量值。

最后，所有测量值被归为 $\mathcal{Z} = \{\mathcal{P}, \mathcal{F}_f, \mathcal{K}_k\}_{f\in \mathsf{F} k\in \mathsf{K}}$，其中 $\mathcal{P}$是点云测量值，$\mathcal{F}$ 是力测量值，$\mathcal{K}$ 是来自机器人正向运动学的位姿测量值。

5 因子图 formulatation

给定状态历史 $\mathcal{X}$，我们最大化测量值 $\mathcal{Z}$的似然：

$$\mathcal{X}^* = \arg\max_{\mathcal{X}} p(\mathcal{X}|\mathcal{Z}) \propto p(\boldsymbol{x}_0)p(\mathcal{Z}|\mathcal{X}), \tag{9}$$

其中，$\mathcal{X}^*$ 是我们对关节运动的 MAP 估计。

我们假设测量值在条件上相互独立，且受到均值为零的高斯噪声干扰。因此，式（9）可表示为以下最小二乘最小化问题：

$$\mathcal{X}^* = \arg\min_{\mathcal{X}} \|\mathbf{r}_0\|^2_{\Sigma_0} + \|\mathbf{r}_\mathcal{P}\|^2_{\Sigma_\mathcal{P}} + \sum_{f\in \mathsf{F}} \|\mathbf{r}_{\mathcal{F}_f}\|^2_{\Sigma_\mathcal{F}} + \sum_{k\in \mathsf{K}} \left(\|\mathbf{r}_{\mathcal{A}_k}\|^2_{\Sigma_\mathcal{A}} + \|\mathbf{r}_{\mathcal{K}_k}\|^2_{\Sigma_\mathcal{K}}\right), \tag{10}$$

其中，每一项都是与某种测量类型相关的残差 $\mathbf{r}$，并假设其受到均值为零的高斯噪声干扰，协方差由相应的测量值决定。因子图可用于以图形方式表示式（10），如图 2所示。较大的白色圆圈表示我们希望估计的变量，较小的彩色圆圈则将残差表示为因子。下一节将详细介绍因子的实现。

6 方法

本节介绍如何在我们的因子图中融合三种测量值（点云 $\mathcal{P}$、力 $\mathcal{F}$和运动学 $\mathcal{K}$），并使用四个因子（可供性 $\mathbf{r}_\mathcal{P}$、关节运动 $\mathbf{r}_\mathcal{A}$、力 $\mathbf{r}_\mathcal{F}$ 和运动学$\mathbf{r}_\mathcal{K}$）在线估计关节运动。

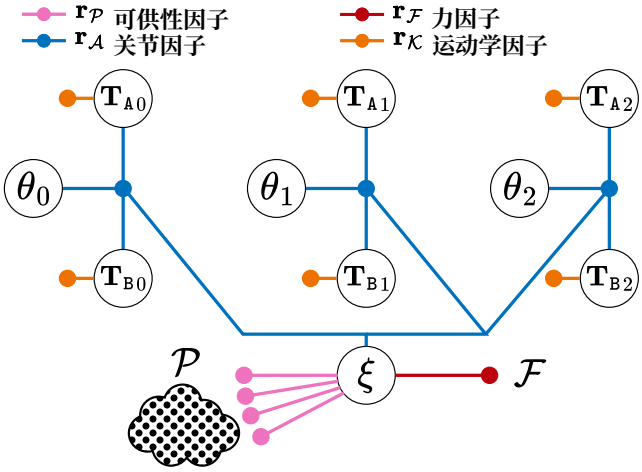

图2 因子图展示了我们正在估计的变量：$\mathbf{T}_\mathrm{A}(t)$，$\mathbf{T}_\mathrm{B}(t)$ $\theta(t)$以及 ξ，它仅存在于因子图的一个时间步中。我们展示了三个时间步，其中包括初始视觉可供性因子，它作为一元因子为 ξ 提供先验估计。

6.1 基于视觉的不确定性感知关节运动预测

我们希望利用深度神经网络，根据视觉测量预测关节运动可供性。该可供性可以用点云来表示，其中每个三维点编码了在发生少量关节运动时该点的归一化瞬时速度。如图 3所示，对于棱柱关节，所有向量都将沿运动轴指向，且大小相同。对于转动关节，所有向量都将沿圆形轨迹的切线方向，且距离旋转轴越远，向量越长。Zeng 等人(2021年) 首次提出了这种关节运动可供性表示，并将其描述为运动残差流。之后，Eisner* 等人(2022)；Zhang 等人(2022) 利用 Flowbot3D 改进了实现方式，而Buchanan 等人(2024年) 则在其框架中使用了 Flowbot3D 的网络。

我们提出了一种新的神经网络，它与Flow-bot3d一样，采用PointNet++ (Charles 等人 (2017)) 作为底层架构。我们的神经网络接收一个点云 $\mathcal{P}$，其中包含N个点，$n \in \mathsf{N}$和$\mathbf{p}_n \in \mathbb{R}^4$ 用于编码三维位置，并通过掩码指示该点属于关节部件还是固定基座部件。随后，网络预测关节部件上每个点的流$\hat{}\mathbf{f}_n \in \mathbb{R}^3$。

在我们之前的工作（Buchanan 等人 (2024年)）中，我们使用均方误差（MSE）损失函数训练网络，使其仅预测流：

$$\mathcal{L}_{\text{MSE}}(\mathbf{f}, \hat{\mathbf{f}}) = \frac{1}{p}\sum_{i=1}^{p} \left\|\mathbf{f}_i - \hat{\mathbf{f}}_i\right\|^2, \tag{11}$$

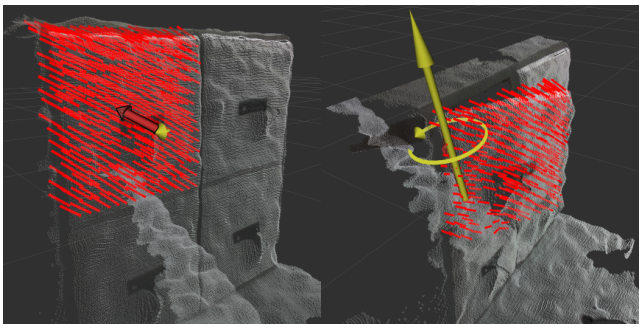

图 3 Buchanan 等人 的神经网络给出的可供性预测示例（2024年）：左侧为移动，右侧为转动。细小的红线是网络的输出，用于预测分割点上的关节运动流。粗大的红色和黄色箭头表示通过平面拟合得到的最终关节预测，方法与Buchanan 等人（2024年）中所采用的方法相同。

在这项工作中，我们希望训练神经网络，使其还能预测每个逐点流预测的偶然不确定性。为此，我们将损失函数改为 Russell 和 Reale (2022) 所示的方法，将损失函数改为以下高斯最大似然（ML）损失：

$$\mathcal{L}_{\text{ML}}(\mathbf{f}, \hat{\mathbf{\Sigma}}, \hat{\mathbf{f}}) = \frac{1}{N}\sum_{n=1}^{N} -\log\left(\frac{1}{\sqrt{8\pi^3 \det(\hat{\mathbf{\Sigma}}_n)}} e^{-\frac{1}{2}\left\|\mathbf{f}_n - \hat{\mathbf{f}}_n\right\|^2_{\hat{\mathbf{\Sigma}}_n}}\right). \tag{12}$$

这使得网络能够以监督方式根据标签$\mathbf{f}$学习预测关节运动流，并以无监督方式学习ˆ不确定性 $\mathbf{\Sigma}$ 。随后，具有训练权重 Θ 的网络函数 f 可表示为：

$$f_\Theta : (\mathcal{P}) \mapsto (\hat{\mathbf{f}}_i, \hat{\mathbf{u}}_i), \tag{13}$$

其中，ˆ$\mathbf{u}_i \in \mathbb{R}^3$可使用以下方法构造为协方差矩阵：

$$\hat{\mathbf{\Sigma}}_i(\hat{\mathbf{u}}_i) = \text{diag}(e^{2\hat{u}_x}, e^{2\hat{u}_y}, e^{2\hat{u}_z}). \tag{14}$$

这种新的关节运动预测示例如图 4 所示。在这个模拟滑动门示例中，门几乎完全打开，因此关节运动在视觉上存在歧义。网络的预测表明，门可能绕门框转动。然而，检查协方差可知， x 方向的不确定性最大，其次是 y，而 z方向的不确定性最小。稍后引入力因子时，这一点将很有用：它会以极低的不确定性将 x 方向上的运动“校正”为零，使关节估计收敛到 y 方向。

6.2 新的可供性因子

在先前利用深度学习可供性预测进行关节估计的工作中，研究者采用了一种手工设计的方法，将流动预测纳入因子图（Buchanan 等人(2024年)）。该方法将两个平面分别拟合到初始点云和表示微小关节运动的点云上。这两个平面的交线表示转动关节的测量值。如果交线距离很远，则将流动方向作为棱ˆ柱关节的测量值。这些关节运动预测 ξ 直接用于关节估计，作为一元因子，其残差如下：

$$\mathbf{r}_\mathcal{P} = \xi - \hat{\xi}. \tag{15}$$

这需要一个手动调节的不确定性（$\sigma_\mathcal{P} = 1e^{-3}$），但它无法捕捉神经网络的真实不确定性。

在我们的工作中，我们转而引入一种新的可供性因子，它直接整合预测的不确定性ˆ$\mathbf{\Sigma}$。首先，我们将 Buchanan 等人 (2024年)提出的单次关节因子改为点云中每个N个点的逐点因子之和 $\mathcal{P}$：

$$\mathbf{r}_\mathcal{P} = \sum_{i \in N} \mathbf{r}_{\mathcal{P}_i}, \tag{16}$$

如第6.1节所述，每个流向量表示运动部件上的一个点因关节角发生微小变化而产生的位置变化：$\theta = 0.05$（该 θ 增量也用于在仿真中生成训练数据）。关节运动后的新点位置可写为：

$$\hat{\mathbf{p}}_i^+ = \hat{\mathbf{f}}_i + \mathbf{p}_i. \tag{17}$$

等价地，利用关节齐次变换方程[1]，，我们可以写为：

$$\mathbf{p}_i^+ = \mathbf{T}_{\text{BA}}(\hat{\xi}, \theta)\mathbf{p}_i. \tag{18}$$

如果我们将 θ 设得非常小（≈ 0.05），那么可以预期ˆ $\mathbf{p}_i^+$ 与公式（17）中的$\mathbf{p}_i^+$ （公式 18）等价，因此我们可以逐点使用以下残差：

$$\begin{aligned}\mathbf{r}_{\mathcal{P}_i} &= \mathbf{p}_i^+ - \hat{\mathbf{p}}_i^+ \\ &= \mathbf{T}_{\text{BA}}(\hat{\xi}, \theta)\mathbf{p}_i - \hat{\mathbf{f}}_i - \mathbf{p}_i,\end{aligned} \tag{19}$$

ˆ其以神经网络在式（ $\mathbf{\Sigma}_i$**14**）中预测的协方差为条件。因此，式（$\mathbf{10}_\mathcal{P}$ ）中的点云关节残差$\mathbf{r}$被逐点残差之和（式（19）和式（16））取代，并在图2中以可视化方式表示。

[1] 在这种情况下，$\mathbf{p}_i$ 和 $\mathbf{p}_i^+$ 以齐次坐标表示。之后，我们会将 $\mathbf{p}_i$ 转换回笛卡尔坐标。

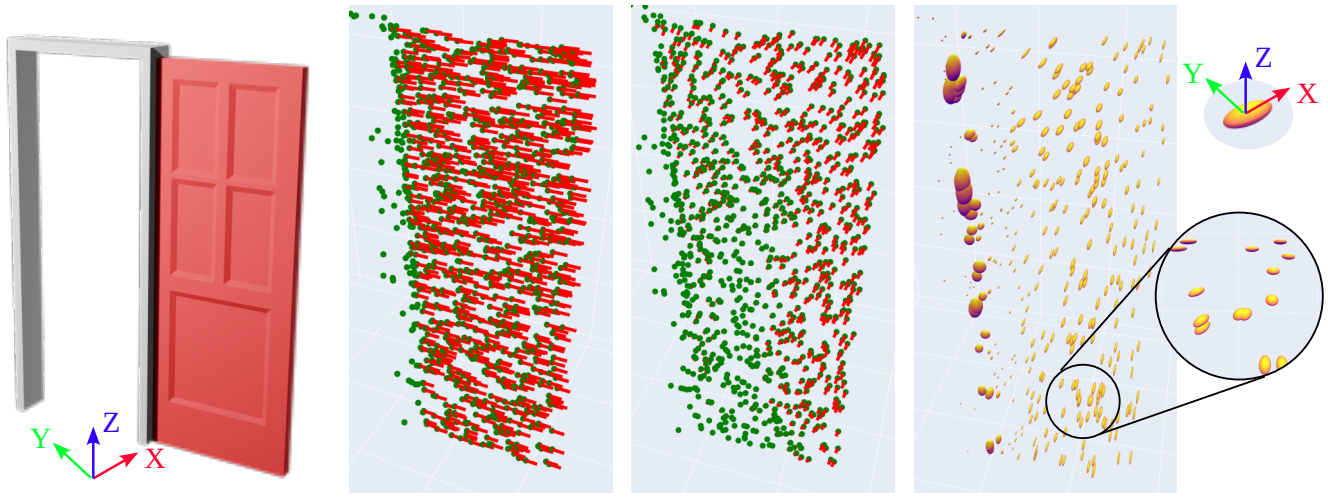

图 4 带协方差的铰接流预测示例。**左侧**：模拟滑动门的渲染图，请注意其轴。**中左**：门的绿色点云测量值，红线表示真实铰接流。**中右**：红线表示预测的铰接流。神经网络将门误判为转动式，且关节位于门框右侧。**右侧**：每个铰接流向量的协方差。x 方向的协方差最大，表明对该关节运动存在很高的不确定性。其次是 y 方向的不确定性，然后是 z 方向。

6.3 力因子

该领域先前研究的另一项局限是，要求初始开启方向必须是将门拉离物体的方向。这意味着可以估计转动门或棱柱抽屉，却无法估计移动滑动门，例如图 1所示橱柜右下方的门。这是因为神经网络会预测一个类似抽屉的移动关节，机器人便会向后拉门。然而，由于门不会移动，因此无法采集运动学测量值，也无法更新关节估计。

作为本文提出的解决方案，我们建议使用腕部安装的力传感器测得的力来推断关节运动。我们在交互开始时、抓取关节部件之后但尚未发生任何运动之前，使用力测量值。如果尝试开门时测得的反作用力达到给定阈值，那么因子图在求解关节运动的新估计值时，会将该 ˆ力 **F** 作为额外因子纳入其中。虽然可以利用力和扭矩引导机器人控制器在开门时尽量减小扭矩，但要用这些测量值推断铰接参数则不那么直接。这是因为反作用力/扭矩测量值只能表明某个特定方向不是有效运动方向，而无法指出哪种替代运动才是正确的。

然而，我们仍然可以利用这些关于无效运动的信息来排除可能的关节运动。如果机器人尝试打开一个关节物体，而没有

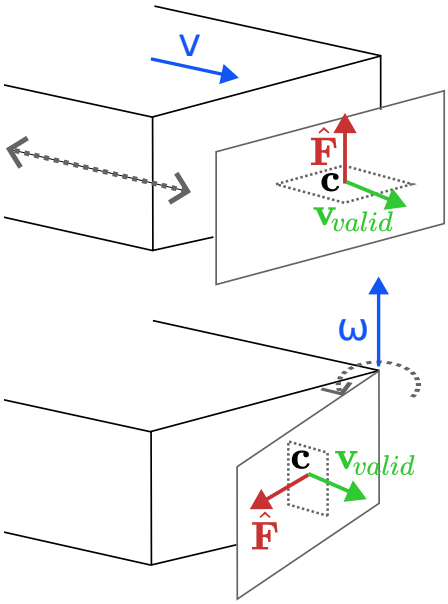

图5 施加的运动方向（灰色）、测得的反作用力与有效运动方向之间关系的示例。**上图**：向移动副施加向下的力，产生向上的反作用力 ˆ**F**。该力与一个平面（灰色虚线平面）正交归一，而我们知道有效运动 $\mathbf{v}_{valid}$ 必定位于该平面上。**下图**：朝向转动关节的铰链施加一个力，产生的反作用力垂直于运动方向，并与 $\mathbf{v}_{valid}$ 所在的平面正交归一。

运动，那么力的方向必须与有效运动方向正交。这可以看作是对Martín-Martín 和 Brock (2022所采用方法的简化版本，因为我们不需要粒子滤波器。

力的方向可以表示为：

$$\mathbf{v}_{valid} \cdot \hat{\mathbf{F}} = 0, \tag{20}$$

其中 $\hat{}\mathbf{F} \in \mathbb{R}^3$ 是测得的反作用力测量值，而 $\mathbf{v}_{valid} = \mathbf{v} + \boldsymbol{\omega} \times \mathbf{c}$ 定义了关节部件上点 $\mathbf{c}$ 的真实有效瞬时运动方向。图5展示了移动关节和转动关节中的这种关系。

直观地说，这意味着，如果机器人尝试向后拉滑动门，就可以推断出门只能沿跨越垂直平面的方向打开（即向左/右或向上/下）。要使这一点成立，我们需要作出以下几项假设：

– 如果关节运动是转动关节，抓取点不在铰链上。我们可以作出这一假设，因为是人类操作员向机器人提供抓取点。– 施力时，关节物体不会移动。

如果这两个假设成立，我们可以利用公式（20）校正估计的运动方向 $\mathbf{v}_{est}$，使其$\hat{}$位于一个法向量为$\mathbf{F}$的平面上。要执行旋转，$\hat{}$我们首先求出一个同时垂直于 $\mathbf{v}_{est}$ 和 $\mathbf{F}$的向量：

$$\mathbf{t} = \mathbf{v}_{est} \times \hat{\mathbf{F}}. \tag{21}$$

如果$\mathbf{t}$ 是零向量（即$\mathbf{v}_{est}$ 与$\hat{}\mathbf{F}$平行或反平行），我们通过将$\mathbf{F}$ 与一个标准基向量取叉积，选取另一个垂直向量$\hat{}$，同时确保：

$$\hat{\mathbf{t}} = \frac{\mathbf{t}}{\|\mathbf{t}\|}. \tag{22}$$

$\hat{}$这可确保$\mathbf{t}$ 位于该平面内，且与$\mathbf{v}_{est}$ 和$\mathbf{F}$均正交。然后，使用叉积将该向量旋转 $90°$：

$$\mathbf{v}_{\text{rot}} = \hat{t} \times \mathbf{v}_{est} \tag{23}$$

因此，我们可以定义如下残差：

$$\mathbf{r}_{\mathcal{F}_f} = \mathbf{v}_{est} - \mathbf{v}_{\text{rot}} \tag{24}$$

这会将对$\mathbf{v}_{est} = \mathbf{v} + \boldsymbol{\omega} \times \mathbf{c}$ 的估计推向存在可能运动方向的平面。当与第6.2节所述的可供性因子结合使用时，这会将优化结果推向下一个最可能的关节运动，其运动位于该平面上。对于力因子，我们将不确定性手动调得非常低（$\Sigma_{\mathcal{F}} = 1e^{-6}$），以便单个因子能够校正可供性因子。

6.4 运动学因子

我们同时对关节状态 $\boldsymbol{x}$ 和部件位姿$\mathbf{T}_{\mathtt{A}}$ 以及$\mathbf{T}_{\mathtt{B}}$进行优化。在时间 k，我们使用机器人的正向运动学计算末端执行器位姿，并假设其刚性抓取了物体的关节运动部件。这为交互过程中$\mathbf{T}_{\mathtt{A}}$ 提供了测量值。此外，我们假设$\mathbf{T}_{\mathtt{B}}$不会移动，因此在每个时间 k都重新使用初始抓取姿态。为考虑少量滑移，我们为这些测量值关联一个不确定性，该值经过手动调节 $\sigma_{\mathcal{K}} = 1e^{-3}$。残差$\mathbf{r}_{\mathcal{K}_k}$ 是 GTSAM 中的默认SE(3) 一元因子（Dellaert 和 GTSAM贡献者（2022））。

6.5 关节因子

我们使用的第四个也是最后一个因子，与先前工作中采用的关节因子等效（Buchanan 等人(2024年)）。该因子使用第 $\boldsymbol{x}$、$\mathbf{T}_{\mathtt{A}}$和$\mathbf{T}_{\mathtt{B}}$，在因子图中连接这些变量；其依据的是第3节所述的关节螺旋模型。我们将估计的部件位姿与预期的关节模型进行比较（第7节）。与Buchanan 等人(2024年)一样，将这些内容结合起来即可得到关节残差：

$$\mathbf{r}_{\mathcal{A}_k} = \mathbf{T}_{\mathtt{BA}}(\hat{\xi}, \theta_k) \boxminus \mathbf{T}_{\mathtt{B}k}^{-1} \mathbf{T}_{\mathtt{A}k}, \tag{25}$$

其中 $\boxminus$ 表示使用对数映射在流形上进行的位姿差分：

$$\mathbf{T}_{\mathtt{A}} \boxminus \mathbf{T}_{\mathtt{B}} = \operatorname{Log}(\mathbf{T}_{\mathtt{B}k}{}^{-1} \mathbf{T}_{\mathtt{A}k}) \in \mathfrak{so}(3). \tag{26}$$

7 实现

本节介绍如图6所示的共享自主完整系统实现。该系统由三个模块组成：初始化，在开始时执行一次；估计，在线运行以估计关节运动；以及运动生成，计算机器人打开物体的轨迹。估计和运动生成均在线运行，并针对每个新的关节估计生成一条新轨迹。

7.1 初始化

对于初始化模块，我们采用关节物体深度学习领域的最新进展，并引入一个共享自主系统。首先，向用户显示物体的视频流，用户点击所需的抓取点。对于这个查询点，我们使用公开可用的分割工具 Segment Anything（SAM）（Kirillov 等人 (2023年) 对

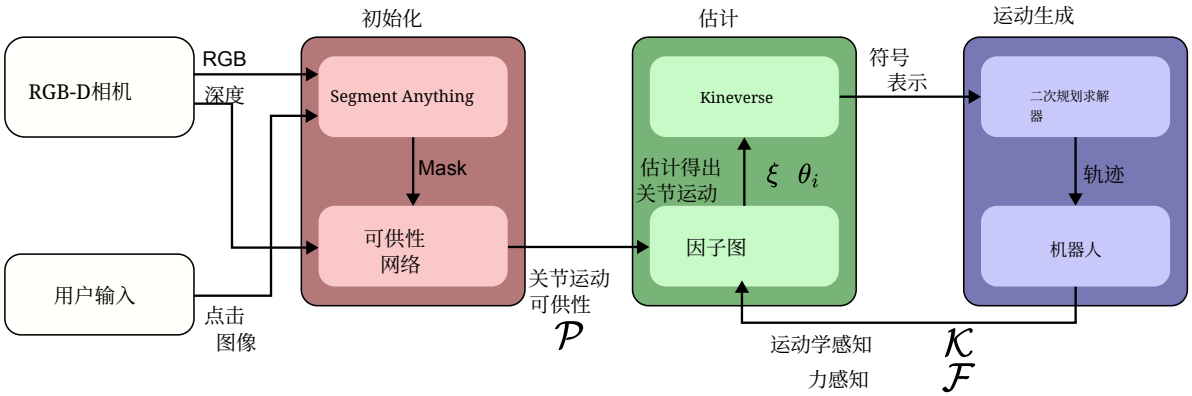

图 6 包含信息流的完整系统。RGB-D相机提供 RGB图像，并根据人类用户的点击提示对图像进行分割。由此生成关节部件的掩码，再结合相机提供的深度信息，预测初始铰接参数。随后将这些参数输入因子图；因子图还利用末端执行器的运动学测量值来估计物体关节结构。估计出的关节运动会更新机器人和物体的符号数学表示，之后将其构建为二次规划（QP）问题，以求解机器人轨迹。

非静态部分进行掩码分割。我们使用 SAM 的开源 ROS 封装，该封装最初由Buchanan 等人(2024年) 提出。随后，将图像掩码及其对应的点云传递给网络，网络会按照第 6.1 节所述，预测掩码中每个点的关节可供性。该神经网络使用 PartNet-Mobility 数据集中的关节式物体示例，在 PyBullet 仿真中进行训练（Mo 等人(2019年)）。因此，初始化步骤的输出是点云可供性 $\mathcal{P}$，以及与用户点击位置关联的 3D 点，该点将作为机器人的第一个规划目标。

7.2 估计

系统初始化后，首先使用可供性预测优化因子图 $\mathcal{P}$。我们使用 GTSAM（Dellaert 和 GTSAM贡献者 (2022))实现因子图。首次优化得出关节运动的初步预测，并立即将其发送至运动生成模块。如果机器人正在与之交互的门开始移动，则每当末端执行器移动一定距离 d，就会向因子图添加运动学测量值。我们使用 Kineverse 关节模型框架（Röfer 等人 (2022))来表示机器人和关节式物体的正向运动学及约束。Kineverse 使用 CasADi 符号数学后端（Andersson 等人 (2019年)），从而能够轻松计算任意表达式（例如关节运动）的梯度。

以往工作的一个主要局限是：如果对关节运动的初始预测与实际关节运动正交，那么交互过程中末端执行器将不会移动，因此无法根据运动学测量值更新估计。例如，在图1右下方的抽屉中，门向右滑动打开。然而，网络通常会预测一个向后拉动的移动副。机器人手臂会拉住门把手但无法使其移动，因而无法获得关于关节运动的新信息。在本工作中，我们结合第6.1 节所述的新关节因子，以及第6.3 节所述的力因子，克服了这一挑战。机器人手腕处装有力/扭矩传感器。当机器人尝试打开门时，如果反作用力大于设定阈值且门没有移动，就会触发添加力因子并执行额外的优化。由此得到新的关节运动估计，并将其传递给运动生成模块。这样，机器人会持续尝试打开门，并利用反作用力来引导下一次估计。

7.3 运动生成

给定关节运动的最新估计值 ξ，此模块会计算用于打开物体的目标机器人构型 $\mathbf{q} \in \mathbb{R}^7$。我们将机器人末端执行器坐标系定义为 E，并将机器人末端执行器的正向运动学建模为 $\mathbf{T}_{\mathtt{WE}}(\mathbf{q})$。

为计算打开关节物体所需的目标正向运动学，我们将式（6）稍作改写，得到

$$\mathbf{T}_{\mathtt{WA}}(\hat{\xi}, \theta_t) = \mathbf{T}_{\mathtt{WB}} \cdot \mathbf{T}_{\mathtt{BA}}(\hat{\xi}, \theta_t), \tag{27}$$

ˆ以$\mathbf{T}_{\mathtt{BA}}(\xi, \theta_t)$表示，$\xi$ 的最新估计值已提供，并采用目标 θ_t。位姿$\mathbf{T}_{\mathtt{WB}}$ 是一个静态变换，由用户定义的抓取位置$\mathbf{p}_{\mathtt{WB}}$ 和预先设定的抓取方向$\mathbf{R}_{\mathtt{WB}}$组成。

机器人抓取物体把手后，我们设定$\hat{\theta}_0 = 0$，从而得到$\mathbf{T}_{\mathtt{BA}}(\xi,0) = \mathbf{I}_{4\times4}$。随后，我们逐步增加期望的关节运动构型$\theta_{t+1} = \theta_t + gv\Delta t$，其中$gv$为打开/关闭关节运动的恒定速度；达到关节运动限位后，我们将gv的符号反转。对于每个θ_t，并根据对ξ的估计，我们求解逆运动学（IK）问题，并满足条件$\hat{\mathbf{T}}_{\mathtt{WE}}(\mathbf{q}_{t+1}) = \mathbf{T}_{\mathtt{WA}}(\xi,\theta_t)$。更具体地说，我们将逆运动学问题定义为一个非线性优化问题，并在其中编码以下任务空间约束

$$\begin{aligned}\left\|\mathbf{p}_{\mathtt{WE}}(\mathbf{q}_{t+1}) - \mathbf{p}_{\mathtt{WA}}(\hat{\xi},\theta_t)\right\|_F^2 = 0\\ \left\|\mathbf{R}_{\mathtt{WE}}(\mathbf{q}_{t+1}) - \mathbf{R}_{\mathtt{WA}}(\hat{\xi},\theta_t)\right\|_F^2 = 0\end{aligned} \tag{28}$$

其中$\|\cdot\|_F$表示弗罗贝尼乌斯范数。

我们利用式（28）中的约束关于$\mathbf{q}$的可微性，将问题线性化，并通过二次规划（QP）逐步求解，直至满足约束：

$$\begin{aligned}\arg\min_{\mathbf{x}} \frac{1}{2}\mathbf{x}^T\mathbf{C}\mathbf{x} \quad \text{s.t.} \quad \mathbf{lb} \le \mathbf{x} \le \mathbf{ub}\\ \mathbf{lb}_A \le \mathbf{A}\mathbf{x} \le \mathbf{ub}_A,\end{aligned} \tag{29}$$

其中，$\mathbf{x} = \langle \dot{\mathbf{q}}, \mathbf{s}\rangle$是由关节速度和松弛变量$\mathbf{s}$组成的向量，$\mathbf{A}$是任务约束及其相关松弛变量的雅可比矩阵。式（29）还对机器人关节位置和速度施加了边界限制。我们使用 Kineverse（Röfer 等人（2022））的符号表示来计算雅可比矩阵，并对式（29）中的问题进行编码和求解。

最后，我们将得到的关节位置$\mathbf{q}_{t+1}$发送给处于柔顺模式的机器人。因此，如果关节估计ξ不准确，机器人可以顺应实际的关节运动，从而使末端执行器位姿不同于$\mathbf{T}_{\mathtt{WE}}(\mathbf{q}_{t+1})$。实际末端执行器位姿$\mathbf{T}_{\mathtt{WA}}(\xi, \theta_{t+1})$作为测量值添加到图中，关联于$\mathbf{T}_{\mathtt{WA}}$。

8 实验

在以下章节中，我们将介绍为评估我们的方法而开展的实验，并报告结果。结果讨论见第9节。

8.1 仿真实验

在这些初步实验中，我们将具备不确定性感知能力的关节运动预测方法与 Flowbot3D 的可供性预测方法进行比较（Eisner 等人 (2022)）。我们在 PyBullet 中为若干个 PartNet-Mobility 数据集中的未见关节式物体模拟点云数据，然后由每个神经网络对物体的关节运动进行预测。对于每种方法，我们都按照 Flowbot3D 作者的方式选择抓取点，即选取预测流幅值最大的点。随后使用各方法预测关节运动，并在抓取点处沿关节开启方向对物体施加力进行模拟。每个物体初始时均处于关闭状态；如果物体开启至其最大开启幅度的 90%，则视为成功，这与 Eisner 等人(2022）采用的标准相同。

我们选择仿真施加力，而不是使用浮动末端执行器，因为我们发现，末端执行器常常能够通过不切实际的方式打开物体，例如穿过物体后从内部将其打开。此外，由于接触物理建模不佳，末端执行器有时会受到不切实际的巨大力，导致抓取滑脱。

我们比较了两种情况：一种是在开始时进行单次关节运动预测（单次），另一种是仿真连续的点云测量，并在交互过程中持续进行关节运动预测（多次）。在多次实验中，每次新的关节运动预测都会更新拉动方向（但不更新抓取点）。这样，如果首次预测足以使物体稍微打开、但未能完全打开，后续预测便可利用这道细小的开口，做出更准确的预测。

我们还比较了使用不同数量关节运动因子的方法。我们将点云下采样至 200、500 和 1000 个点。每个点都会在因子图中增加一个因子，从而延长优化时间。

8.1.1 结果

结果汇总于表 1，部分实验示例如图 7 所示。显然，多次推理明显优于单次推理，因为它能够在交互过程中更新预测。然而，出于若干原因，我们认为多次推理不适合部署在机器人硬件上。首先，如果摄像头安装在机器人末端执行器上，就无法在交互过程中观察关节物体，因此机器人必须放开物体并重新观察场景，如 Schiavi 等人（2023）所述。另一方面，如果摄像头安装在外部或机器人其他部位（例如人形机器人的头部），仍会受到交互中的机器人手臂造成的遮挡，而且在移动关节部件时，需要持续更新分割掩膜。

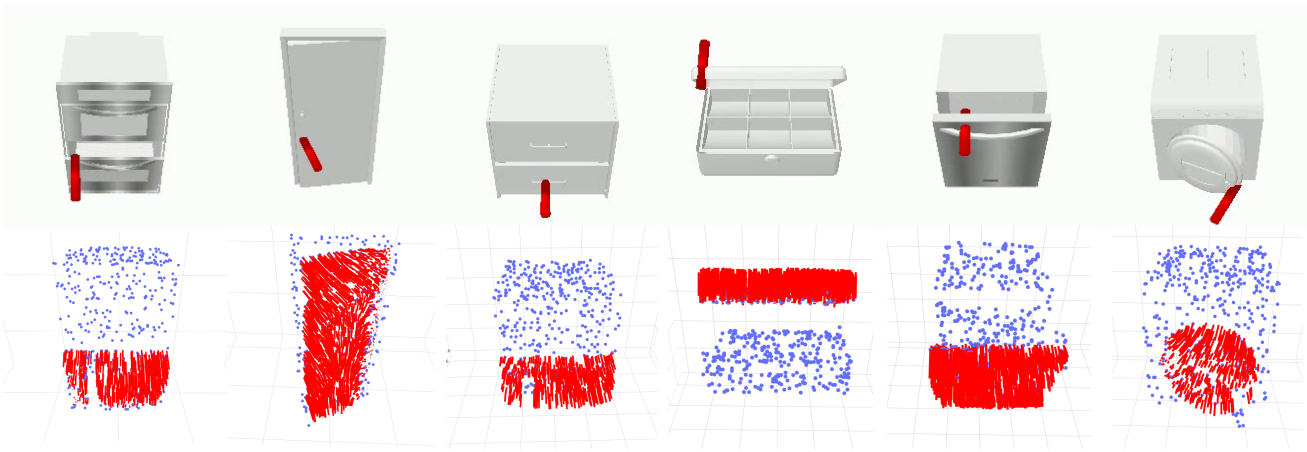

图 7 仿真实验的输出示例。**上排：**在 PyBullet 中渲染 PartNet Mobility 数据集中的物体。红线表示根据关节估计得出的拉动方向。**下排：**输入点云以蓝色显示，神经网络预测的输出关节运动流以红线叠加显示。

表 1 仿真结果以成功率百分比和每次推理所需时间（秒）表示。

方法	单次	多次	平均	最差
Flowbot 3D (Eisner* 等 (2022))	53.67%	61.47%	0.01 秒	0.02 秒
艺术因子 200	51.74%	61.49%	0.03 秒	0.35 秒
艺术因子 500	56.60%	66.81%	0.11 s	0.82 s
艺术因子 1000	57.71%	68.37%	0.22 秒	1.55 秒

相反，我们认为，将单次推理方法与所提出的运动学和基于力的感知相结合，更适合实际应用。此外，比较表 1 中的推理时间可以发现，随着因子图中添加的因子增多，延迟显著增加。由于我们计划仅在交互开始时进行一次推理，因此这种延迟增加是可以接受的。最后，我们注意到，采用 500 个和 1000 个因子的方法分别比 Flowbot3D 高出 8.7% 和 11.2%。这是因为我们的方法整合了大量关节点，以优化关节运动的整体解。相比之下，Flowbot3D 只选择最大的单个点。虽然这使其方法的延迟极低，但会增加预测的变异性，从而导致整体成功率较低。在我们的方法中，增加关节运动因子的数量能够提升性能；但我们发现，超过 1000 个点后性能不再提升，因此在后续真实机器人实验中采用了这一数值。

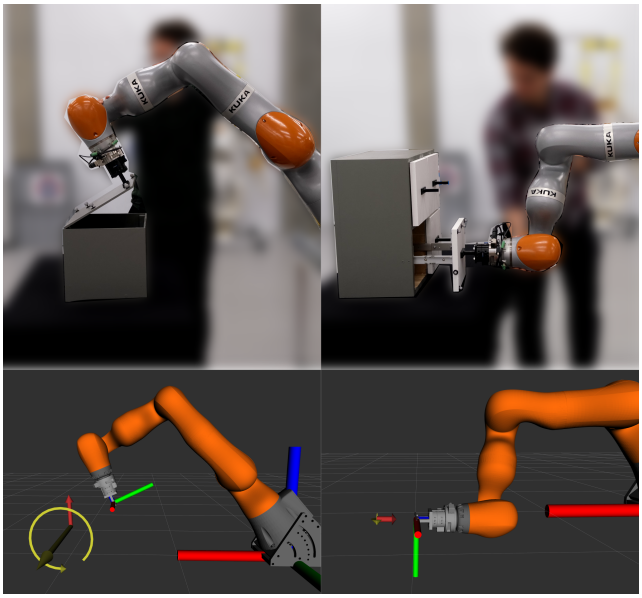

图8 上图：转动关节（左）和移动关节（右）的手动引导实验。运动捕捉标记仅用于提供真实值参考。**下图**：由关节编码器传感器得出的估计关节运动。黄色箭头表示$\boldsymbol{\omega}$，红色箭头表示$\mathbf{v}$。大坐标轴是机器人的基坐标系，用于 `W`；小坐标轴是估计位姿$\mathbf{T}_{\mathtt{A}}$。

8.2 手动引导实验

在这些实验中，我们研究了基于运动学的关节运动估计的准确性。我们将我们的方法与Heppert 等人(2022进行了比较。该方法同样使用因子图来估计螺旋参数化。作者欣然允许我们访问其代码，以便进行直接比较。

这些实验是在真实机器人硬件上进行的。我们使用了具有柔顺性的 KUKA LBR iiwa 机器人，并将机器人的末端执行器实体连接到箱盖上。随后，我们在重力补偿模式下用手引导机器人运动，以打开和关闭箱子。在这项实验中，我们记录了编码器测量的机器人关节位置，以及通过 Vicon 动作捕捉追踪的相应箱盖位姿，如图 8 所示。与 Heppert 等人(2022)类似，我们使用

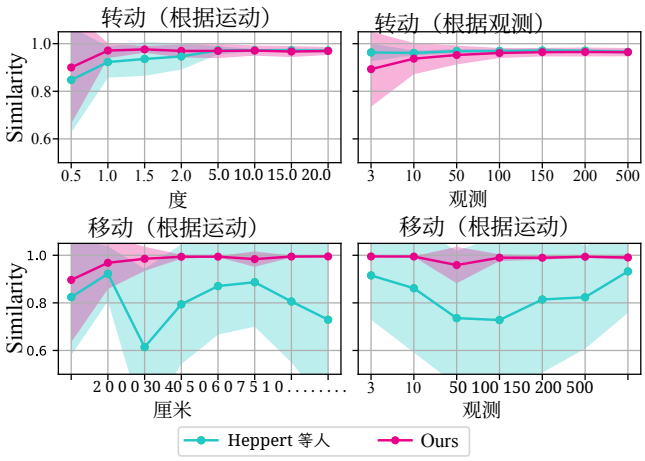

图9 手动引导实验的切向相似度。实线表示平均误差，阴影区域表示标准差。

切向相似度指标：

$$J(\mathbf{v}_{gt}, \mathbf{v}_{est}) = \frac{1}{\theta_{max} - \theta_{min}} \int_{\theta_{min}}^{\theta_{max}} \frac{\mathbf{v}_{gt}}{\|\mathbf{v}_{gt}\|} \cdot \frac{\mathbf{v}_{est}}{\|\mathbf{v}_{est}\|}, \quad (30)$$

其中，$\mathbf{v}^{gt}$ 是通过 Vicon 测得的抓取点局部线速度，$\mathbf{v}_{est}$ 是根据关节模型估计的局部速度。我们可以使用以下公式，根据 ξ 计算$\mathbf{v}_{est}$：$\mathbf{v}_{est} = \mathbf{v} + \boldsymbol{\omega} \times \mathbf{c}$ 其中，$\mathbf{c}$ 是根据运动学得出的接触点。由于$\mathbf{v}^{gt}$ 和$\mathbf{v}_{est}$ 均已归一化，因此它们表示运动方向；因此，当二者方向相同时，切向相似度为 1；当二者垂直时，切向相似度为 0。

我们记录了两次手动引导实验，一次针对转动关节，另一次针对移动关节。首先，我们按固定增量进行优化，例如，每 1° 旋转或1cm 平移进行一次优化。接着，我们测试了在整个构型范围内等间距选取固定数量测量值的方案，完整结果见图9。在因子图中，我们不区分移动副和转动关节。估计移动关节时，$\boldsymbol{\omega}$ 往往会趋近于极小值。在输出时，如果 $\|\boldsymbol{\omega}\| < 0.01$，我们将 $\boldsymbol{\omega} = 0_{3\times1}$ 设为$\mathbf{v}$并进行归一化。

8.2.1 结果

我们的结果表明，即使测量值数量很少，也能达到很高的准确度。旋转仅达到 0.5° 后，我们的估计器的平均切向相似度为 0.90；达到 1.0° 后，这一数值提升至 0.97。这使得在神经网络预测错误时仍能进行在线关节估计，因为机器人部件只需带动关节部件移动少量，便可更新估计结果。此外，我们还表明，在整个构型范围内均匀分布测量值时，仅需 3 个测量值就能准确估计关节。与 Heppert等人的方法相比，两种方法在转动关节上的表现相近，而我们的方法更善于区分移动关节。我们认为，这是因为我们会检查是否存在移动关节运动，而他们的方法往往会将移动副与幅度很大的转动关节运动混淆。

8.3 共享自主机器人实验

在共享自主机器人实验中，我们使用了相同的 KUKA iiwa机器人；在感知和抓取方面，我们使用了 Intel RealSense D435相机、ATI Delta力/扭矩传感器和 Robotiq 140两指夹爪。在这些实验中，我们按照第7节所述测试了完整流程，实验方案如下：人类用户查看机器人相机的画面，该画面拍摄的是与图1中相同的橱柜，然后点击图像中要抓取的位置。机器人随后移动至抓取目标并闭合夹爪。接着，机器人根据$\theta = 0$ 所学习的关节运动预测，移动至指定的上限。如果达到力阈值而夹爪没有移动，就会触发向因子图中添加一个力因子，然后对因子图进行优化，以求得下一个最大后验关节状态。机器人手臂会再次尝试打开橱柜；如果成功则任务完成，否则会继续添加力因子，直至收敛到一个门开始打开的方向。

随着估计在线进行，一旦末端执行器开始打开门，哪怕只打开一点，运动学感知就会被添加到因子图中，并更新模型。更新后的模型会以闭环方式反馈给控制器。最终，机械臂的运动使门能够打开得更多，从而产生更多运动学测量值，估计结果也会收敛到关节的正确估计值，而控制器则继续开关门。我们每获得 20 个新的运动学测量值就进行一次新的优化，并使用 $d = 2$mm 或 $d = 0.5^\circ$ 的距离阈值，触发向因子图中添加新的运动学测量值。

8.3.1 结果

图中展示了四次在线估计实验。10。与先前工作中展示的情况类似（Buchanan 等人(2024年)），在缺少关节运动视觉线索时，基于可供性的神经网络往往会预测移动关节。在本研究中，由于采用了关节运动预测因子并引入了力平面因子，即使面对右下角的滑动门案例，机器人也能够更新估计结果。我们在不同的门上重复进行了 20 次完整流程实验，并成功

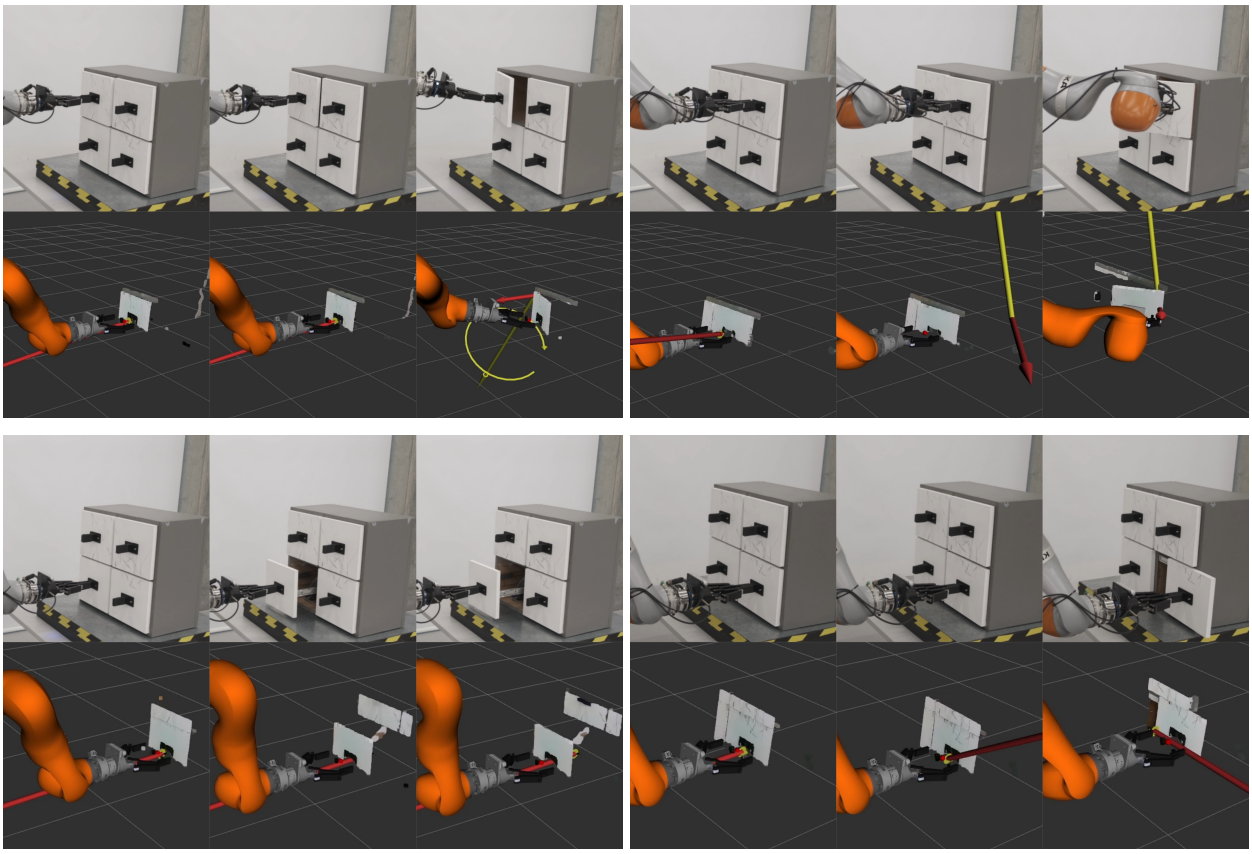

图 10 机器人实验：分别打开四扇橱柜门。**左上：**网络最初预测为移动副，当机械臂向后拉动柜门时，少量的移动使因子图能够求解出正确的转动关节。**右上：**与左上方的柜门类似，网络最初预测为移动副，但该估计通过运动学测量值进行了在线更新。**左下：**网络正确预测为移动副，机器人轻松地打开了抽屉。**右下：**网络预测为移动副，机械臂开始向后拉动。当力阈值被超过后，加入了一个力平面因子，从而再次预测为移动副。这使机器人能够使柜门略微移动，产生运动学测量值，并最终收敛到正确的估计结果。

打开了门 15 次。图11 展示了在线实验期间估计的ξ 参数。每次优化都会计算边缘协方差，其3σ 范围在图11 中以估计均值周围的阴影区域表示。

9 讨论

我们的方法包含多项改进，使机器人能够打开关节运动未知且视觉上难以区分的物体。首先，通过将神经网络改为提供不确定性预测，并修改关节运动因子，我们使关节运动的初始预测能够包含学习得到的不确定性分布，而不再像之前那样使用硬编码分布。图4展示了滑动门的一个网络输出示例，其中协方差沿 x 方向最大，随后是 y。

将力感知加入因子图后，便可利用不确定性来更新估计，从而确定下一个最可能的关节运动。例如图 4所示，网络给出的主导预测是转动门，但从不确定性分布可以明显看出，网络对横向运动的把握较低。将力平面因子加入因子图后，估计会沿着 y 方向收敛，从而正确地将估计更新为移动门。这一过程使机器人得以打开图 10右下方的滑动门。在尝试的 20 次完整系统试验中，有 4 次因夹爪打滑而失败。由于我们假设接触是刚性的，且仅发生少量滑动，因此无法区分明显打滑与有意移动以打开门的情况。未来的工作中，可利用夹爪指尖上的传感器检测这种情况。

在我们的实验中，我们发现机器人手臂的柔顺性会显著影响成功率。例如，如果机器人手臂在某个特定方向上过于柔顺，就可能无法克服关节处的摩擦力来打开门。另一方面，如果

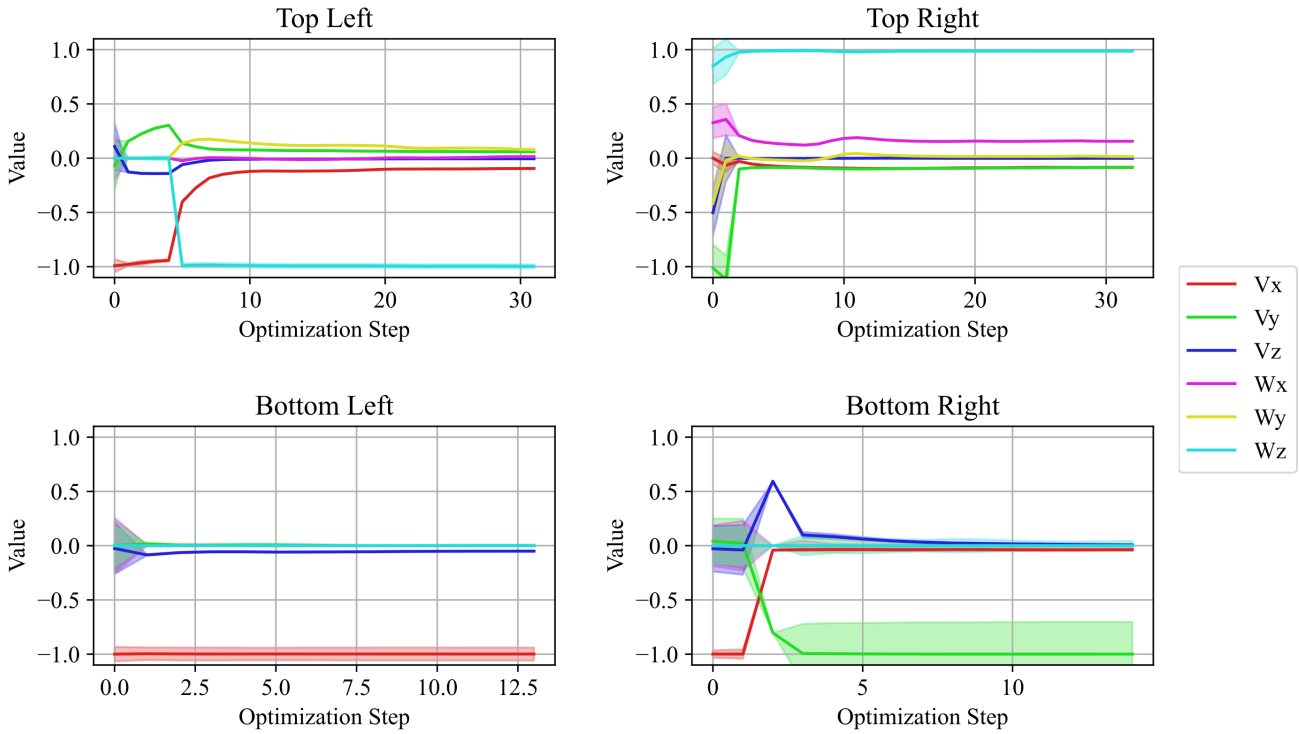

图11 e 的曲线图　　估计的 ξ 参数（对应图10 中的实验）及相关的 3σ 计算　　d 由边　　ginal 协变　　量。

机械臂过于僵硬，则可能会损坏门。这为未来的研究提供了一个方向：根据对关节运动的估计，在线调整刚度参数。基于模型或基于学习的方法都可以应用于这一问题。基于学习的方法也为学习安全交互力的先验提供了途径。目前，我们系统施加的最大力是一个需要针对具体设置进行调整的超参数。虽然柜子大小的物体所需的交互力都非常相近，但我们当然希望系统不仅能操作重型门和小型首饰盒，还无需人工干预。我们认为，可以基于稳定的语言对齐的视觉特征学习此类先验（Rad- ford 等人(2022)），既可以通过增量学习，也可以从人类示范中学习（Li 等人 (2022)）.

最后，我们仅在交互开始时使用力感知来估计关节运动；不过，在未来的工作中，我们打算探索基于学习的方法，从力感知中估计关节运动，类似于学习姿态估计器（Del Aguila Ferran- dis 等人(2024年)）。尽管接触物理存在仿真到现实差距，我们相信，利用仿真训练（Aoyama 等人(2024年)）能够显著提高机器人与关节式物体进行交互时在真实世界中的训练成功率。

10 结论

在这项工作中，我们提出了一种用于在线估计并打开未知关节式物体的新方法。我们的方法能够使机器人打开各种各样的关节式物体，包括常见的家居用品，以及关节运动在视觉上并不明显的物体。我们的方法融合了视觉、力与运动学感知，这些信息既来自神经网络的学习预测，也来自基于物理的建模，并利用螺旋理论对关节运动进行建模。该方法的核心估计框架基于因子图，并与共享自主框架相结合：用户只需点击要打开的位置，机器人便会打开门。

这项工作在我们的前期成果基础上取得了多项重大进展。我们对神经网络进行了修改，使其能够预测不确定性，并引入了一种新的关节因子，以便将这种不确定性纳入因子图。我们还新增了一种感知模态：力感知。这些改进相结合，使我们的系统能够更有效地打开不同类型的关节结构，包括视觉上不明显的关节结构。我们在真实机器人上实现了该方法，并与一个视觉上模糊的关节运动物体进行交互，取得了很高的成功率。

在交互任务中，尽管仍可进一步研究如何将力感知融入该框架，但本文展示了将本体感知与学习得到的视觉先验融合应用于物体操作的显著优势。

参考文献

Andersson JAE, Gillis J, Horn G, Rawlings JB, Diehl M（2019年）CasADi——一种用于非线性优化与最优控制的软件框架。Mathematical Programming Computation 11(1):1–36 Aoyama MY, Moura J, Saito N, Vijayakumar S（2024年）通过预训练触觉表征，从示范中进行基于力的运动少样本学习。载于：IEEE机器人与自动化国际会议（ICRA） Bahl S, Mendonca R, Chen L, Jain U, Pathak D（2023）将人类视频中的可供性作为机器人学的通用表征。载于：IEEE/CVF计算机视觉与模式识别会议（CVPR）Bohg J, Hausman K, Sankaran B, Brock O, Kragic D, Schaal S, Sukhatme G（2017）交互式感知：在感知中利用动作，在动作中利用感知。IEEE机器人学汇刊 33(6):1273–1291 Buchanan R, Röfer A, Moura J, Valada A, Vijayakumar S（2024年）结合视觉与本体感知，对关节物体进行因子图在线估计。载于：IEEE机器人与自动化国际会议（ICRA）Charles RQ, Su H, Kaichun M, Guibas LJ（2017）Point-net：用于三维分类与分割的点集深度学习。载于：IEEE计算机视觉与模式识别会议（CVPR），第77–85页，DOI 10.1109/CVPR.2017.16 Del Aguila Ferrandis J, Pousa De Moura J, Vijayakumar S（2024年）学习视触觉估计与控制，以在遮挡情况下进行非抓取操作。载于：机器人学习会议（CoRL） Dellaert F, GTSAM贡献者（2022）borglab/gtsam。DOI 10.5281/zenodo.5794541, URLhttps://github.com/borglab/gtsam)Eisner* B, Zhang* H, Held D（2022）Flowbot3d：学习三维关节运动流以操作关节式物体。载于：机器人学：科学与系统（RSS）Heppert N, Migimatsu T, Yi B, Chen C, Bohg J（2022）基于因子图的类别无关关节物体跟踪。载于：IEEE/RSJ智能机器人与系统国际会议（IROS），第3800–3807页 Jain A, Kemp CC（2010年）拉开门和抽屉：协调全向底盘与柔顺机械臂的平衡点控制。载于：IEEE国际机器人与自动化会议（ICRA），第1807–1814页。Jain A、Lioutikov R、Chuck C、Niekum S（2021）ScrewNet：利用螺旋理论和深度图像进行类别无关的关节模型估计。载于：IEEE机器人与自动化国际会议（ICRA），第13670–13677页，DOI: 10.1109/ICRA48506.2021.9561132。Jiang H、Mao Y、Savva M、Chang AX（2022a）Opd：单视图三维可开启部件检测。载于：欧洲计算机视觉会议（ECCV），第410–426页。Jiang Z、Hsu CC、Zhu Y（2022b）Ditto：通过交互构建关节物体的数字孪生。载于：IEEE/CVF计算机视觉与模式识别会议（CVPR），第5606–5616页。Katz D、Orthey A、Brock O（2014）关节物体的交互式感知。载于：实验机器人学：第12届实验机器人学国际研讨会，《Springer先进机器人学系列》，Springer，柏林、海德堡，第301–315页。Kirillov A、Mintun E、Ravi N、Mao H、Rolland C、Gustafson L、Xiao T、Whitehead S、Berg AC、Lo WY、Dollár P、Girshick R（2023）Segment anything。arXiv预印本 arXiv:230402643。Li X、Wang H、Yi L、Guibas LJ、Abbott AL、Song S（2020）类别级关节物体位姿估计。载于：IEEE/CVF计算机视觉与模式识别会议（CVPR），第3703–3712页。Li Z、Sedlar J、Carpentier J、Laptev I、Mansard N、Sivic J（2022）从互联网视频中估计人类与物体交互的三维运动和作用力。《国际计算机视觉杂志》130(2)：363–383。Martín-Martín R、Brock O（2022）用于关节物体在线交互式感知的耦合递归估计。《国际机器人研究杂志》41：741–777。Mittal M、Hoeller D、Farshidian F、Hutter M、Garg A（2021）在未知场景中通过全身移动操作实现关节物体交互。arXiv预印本 arXiv:210310534。Mo K、Zhu S、Chang AX、Yi L、Tripathi S、Guibas LJ、Su H（2019）Partnet：用于细粒度分层部件级三维物体理解的大规模基准。载于：IEEE/CVF计算机视觉与模式识别会议（CVPR），第909–918页。Mo K、Guibas L、Mukadam M、Gupta A、Tulsiani S（2021）Where2act：从像素到关节式三维物体的动作。载于：IEEE/CVF国际计算机视觉会议（ICCV），第6793–6803页。Murray RM、Li Z、Sastry S（1994）《机器人操作数学导论》，第1版。CRC出版社。

Nie N、Gadre SY、Ehsani K、Song S（2023）Structure from Action：学习交互以发现三维铰接物体结构。载于：IEEE/RSJ智能机器人与系统国际会议（IROS），第1222–1229页。Radford A、Kim JW、Hallacy C、Ramesh A、Goh G、Agarwal S、Sastry G、Askell A、Mishkin P、Clark J等（2021）从自然语言监督中学习可迁移视觉模型。载于：机器学习国际会议，第8748–8763页。Russell RL、Reale C（2022）深度学习中的多元不确定性。《IEEE神经网络与学习系统汇刊》33(12)：7937–7943。Röfer A、Bartels G、Burgard W、Valada A、Beetz M（2022）Kineverse：用于模型无关移动操作的符号铰接模型框架。《IEEE机器人与自动化快报》7(2)：3372–3379，DOI：10.1109/LRA.2022.3146515。Schiavi G、Wulkop P、Rizzi G、Ott L、Siegwart R、Chung JJ（2023）学习智能体感知可供性，以实现与关节物体的闭环交互。载于：IEEE机器人与自动化国际会议（ICRA），第5916–5922页。Sturm J、Stachniss C、Burgard W（2011）用于学习关节物体运动学模型的概率框架。《人工智能研究杂志》41(2)：477–526。Xu Z、He Z、Song S（2022）用于关节物体的通用操作策略网络。《IEEE机器人与自动化快报》7(2)：2447–2454。Zeng V、Lee TE、Liang J、Kroemer O（2021）关节物体部件的视觉识别。载于：IEEE/RSJ智能机器人与系统国际会议（IROS），第2443–2450页。Zhang H、Eisner B、Held D（2022）Flowbot++：通过关节投影学习广义铰接物体操作。载于：机器人学习会议（CoRL）。