arXiv:2302.01295v1 [cs.RO] 2 Feb 2023

Ditto in the House：通过交互式感知构建室内场景的关节模型

Cheng-Chun Hsu[1] 和 Zhenyu Jiang[1] 和 Yuke Zhu[1]

摘要——将物理世界虚拟化为虚拟模型，一直是机器人在现实世界中进行导航和规划的关键技术。为了促进日常生活中对关节物体的操作，本研究探索让机器人通过在室内场景中的目的性交互来构建室内场景的关节模型。以往的关节推理研究主要聚焦于类别有限、彼此孤立的物体。要扩展到房间尺度环境，机器人必须高效且有效地探索大规模三维空间、定位铰接物体并推断其关节运动。我们提出了一种用于此任务的交互式感知方法。我们的方法名为 *Ditto inthe House*，通过可供性预测发现可能的铰接物体，与这些物体交互以产生关节运动，并根据每次交互前后的视觉观测推断其关节属性。该方法将可供性预测与关节推理紧密结合，从而提升两项任务的表现。我们在仿真和真实场景中展示了该方法的有效性。代码和更多结果请见 https://ut-austin-rpl.github.io/HouseDitto/

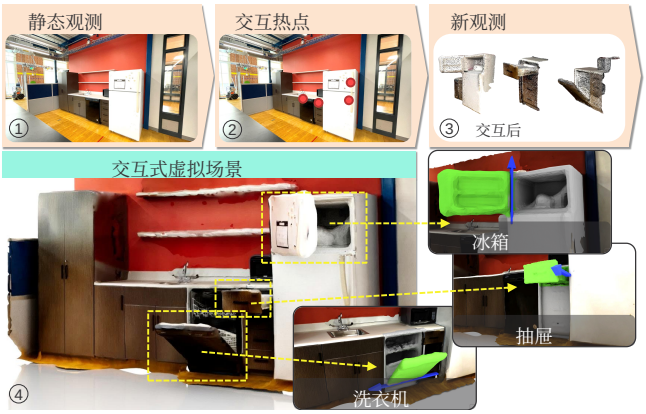

图 1：**通过交互式感知构建场景级关节模型**。给定室内场景的初始观测，我们的方法推断交互热点，引导机器人与铰接物体进行交互。随后，机器人收集交互前后的观测。基于观测到的关节运动，它为各个物体构建关节模型，并将这些模型汇总为场景级关节模型。

一、引言

将真实世界虚拟化为虚拟模型，是机器人在日常环境中执行任务的关键一步。智能机器人依赖这些模型来理解周围环境，并在非结构化场景中规划动作。结构传感器和 SLAM算法的最新进展 [9, 29,30] 为构建真实世界场景的静态复刻模型提供了方法，其逼真度达到了前所未有的水平。这些静态复刻模型有助于移动机器人进行自身定位和导航。然而，要在真实世界中进行操作，机器人就必须从重建静态场景转向揭示物体的物理属性。尤其是要与关节运动物体（如日常环境中常见的橱柜和门）进行物理交互，机器人需要理解它们的运动学和关节属性。受此目标启发，本研究旨在使机器人能够通过有目的的交互，自动构建室内场景的关节运动三维模型。

理解物体关节结构一直是计算机视觉和机器人学领域的一项长期挑战。近年来，一系列数据驱动方法尝试通过在大型三维数据集上训练模型，从静态视觉观测中推断关节运动 [22, 40]。这些模型依赖类别级先验，因此泛化能力局限于少数预设类别。另一条并行的研究路线 [13, 24] 采用交互式感知 [2]。这些方法利用物理交互生成关节运动的视觉观测，并据此估计关节属性。这些研究主要聚焦于单个物体，而要扩展到房间规模的环境，机器人就必须高效且有效地探索大规模三维空间，以开展有意义的交互。

我们提出了**Ditto in the House**，这是一种通过机器人自主探索来构建室内环境关节模型的方法。机器人会发现并与环境中的铰接物体进行物理交互。基于交互前后的视觉观测，机器人推断交互对象的关节属性。我们的方法要求机器人在大规模三维空间中识别可能发生关节运动的区域，通过操纵物体产生关节运动，并根据不完整的观测推断潜在的运动学参数。

首要挑战是确定探索环境的最有效动作，即即最有可能发现关节运动的动作。我们将这一问题表述为根据视觉观测推断场景可供性 [6, 26,28]。具体而言，我们训练一个模型，使其能够基于机器人过去与大量程序生成的仿真三维场景交互的经验来预测可供性。此外，我们还提出一种迭代细化过程，利用对关节运动的初始观测作为视觉线索，细化可供性预测。随后，机器人利用细化后的可供性指导后续交互，以获得更具信息量的观测，从而进行关节推断。

图1展示了我们方法的高层步骤。

[1] 计算机科学系，德克萨斯大学奥斯汀分校。联系邮箱：chengchun@utexas.edu

该方法会发现交互热点（即适合执行探测动作的位置），通过动作生成铰接运动的观测，并根据这些观测推断关节结构。为了定位交互热点，我们使用可供性模型，根据室内场景的点云观测预测可供性图。在每个交互热点处，机器人会尝试与物体交互，并收集机器人动作诱发物体运动前后的自我中心观测。基于我们先前的工作 Ditto [15], ，我们设计了一种关节结构估计模型，根据这些观测推断物体的关节结构。该模型以机器人的动作和视觉观测为输入，并预测物体的关节参数。我们通过迭代过程将可供性预测与关节结构估计紧密结合。我们根据对铰接物体可移动部件的初步估计来更新可供性预测，因为这些部件指示了可操作区域。机器人根据更新后的可供性与物体交互，从而产生更多可见的关节运动，促进关节推断。

我们使用 iGibson [21] 模拟器，在 CubiCasa5K [17] 数据集上评估了我们的方法。定量结果表明，与基线模型相比，我们的方法成功发现的部件多约 40%，且精确率更高。与基线模型相比，我们的方法还使分割交并比绝对提升了 45%。此外，消融研究证实，迭代优化过程有助于提升可供性预测和关节结构模型估计。最后，我们将该方法应用于真实厨房场景，并成功构建了该环境的关节模型。

II. 相关工作

A. 视觉可供性预测

基于视觉观测预测可供性 [10] ，是机器人规划动作、与现实世界物体交互的一项关键能力。一系列研究从人类视频 [3, 7, 12, 27]中学习视觉可供性预测。这些视频揭示了人类如何与周围环境交互的重要线索。然而，从这些视频中学到的可供性往往受人体形态限制，难以推广到机器人硬件。另一类研究则利用仿真或真实机器人获取主动训练数据，以进行可供性学习 [16, 19, 26, 28, 41]。在这些方法中，机器人探索物体和场景中的多种交互方式，并从具身经验中学习可供性。遵循这种自主探索式学习范式，我们在仿真中收集探索数据，并训练视觉可供性预测模型。该模型引导智能体发现场景中的关节物体，并与之交互以进行关节推断。

B. 关节结构模型估计

关节模型表示物体部件以及部件之间的运动学（有时还包括动力学）关系。早期工作利用概率方法 [4, 35–37]对物体部件结构及关节运动进行建模。它们通常依赖标记物或手工设计特征来跟踪可移动部件，这限制了其在自然环境中的实际应用。近年来，深度学习方法 [1, 22, 31,32, 39, 40] 被用于根据原始感知数据进行关节估计。大多数此类工作从单次观测中预测关节模型，并依赖类别级先验来估计关节参数。因此，它们依赖于类别，无法泛化到室内场景中的各种日常物体。另一类工作利用物理交互产生新的感知刺激，并根据物体状态变化推断关节模型。Katz 等 [18] 是首项引入交互式感知 [2] 以进行关节结构模型估计的工作。后续一些工作通过分层递归贝叶斯滤波器 [25], 、概率模型 [38], 和特征跟踪 [33]对其进行了扩展。近期的相关工作利用手工设计策略 [15]或学习策略 [8, 20, 23]产生的信息丰富的运动，并采用几何深度学习技术估计关节模型。上述方法均针对单个物体设计，难以直接扩展到包含多个位置未知物体的大型室内环境。相比之下，我们的方法提出了一种类别无关方法，主动探索环境并构建场景级关节模型。

III. 问题表述

我们研究如何为包含关节物体的室内场景构建关节模型。关节物体由多个部件组成，连接这些部件的关节会约束每对部件之间的相对运动。我们假设场景存在一个初始点云观测$P_s \in \mathbb{R}^{N_s \times 6}$ ，其中每个点是由其坐标x, y，z和颜色R，G，B构成的向量。N_s为点的数量。本研究旨在通过交互式感知对场景中的部件进行分割，并估计其关节模型M 。我们的方法通过主动与物体交互来估计其关节模型。

对于每个物体部件，我们识别交互热点 [27], 即机器人能够成功操纵关节物体的位置，并推断关节模型。关节模型表示物体部件之间的运动学约束。我们按照 ANCSH [22]对关节运动进行参数化。移动关节的参数包括平移轴$u^p \in \mathbb{R}^3$ 的方向和关节状态$s^p \in \mathbb{R}$。转动关节的参数包括旋转轴$u^r \in \mathbb{R}^3$的方向、旋转轴上的枢轴点$q \in \mathbb{R}^3$ 以及关节状态$s^r \in \mathbb{R}$。关节状态s^p 和s^r 分别表示交互前后关节状态的相对变化。

该过程首先从初始场景观测 $H_s \in \mathbb{R}^{N_h \times 3}$ 中寻找交互热点，其中P_s，而N_h 表示场景中的交互热点数量。根据点云观测P_s，可供性预测模型会预测一张可供性图$A_s \in \mathbb{R}^{N_s}$ ，并将其峰值位置采样为交互热点。

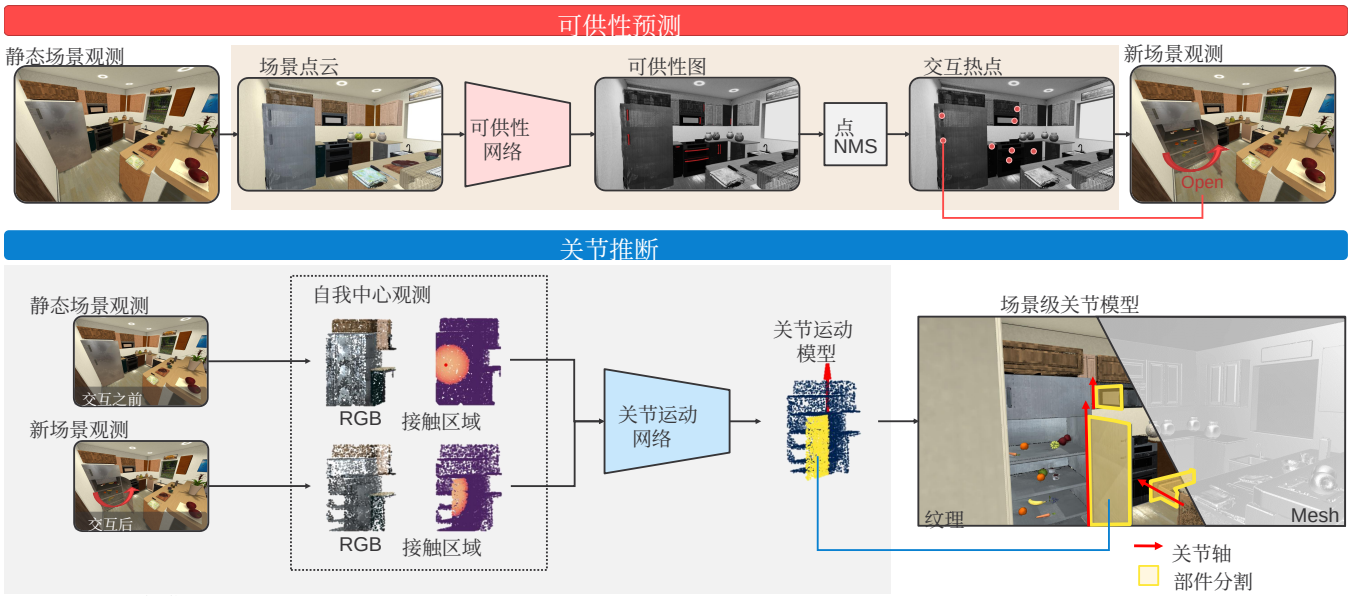

图2：**模型组件概览**。我们的方法分为两个阶段——可供性预测和关节推断。在可供性预测阶段，我们将静态场景点云输入可供性网络，预测场景级可供性图。通过应用点非极大值抑制（NMS），我们从可供性图中提取交互热点。随后，机器人根据这些接触点与物体进行交互。在关节推断阶段，我们将每次交互前后的点云观测输入关节运动模型网络，以获得关节估计。通过汇总估计得到的关节模型，我们构建整个场景的关节模型。

每次交互时，机器人都会对物体热点施加力，以产生潜在的关节运动。机器人会在交互前后分别采集两个以交互热点为中心的自中心局部点云$P \in \mathbb{R}^{N_o \times 3}$和$P' \in \mathbb{R}^{N_o' \times 3}$，其中$N_o$和$N_o'$表示物体点云中的点数。机器人还会记录相应的接触位置c，$c' \in \mathbb{R}^3$。给定这些物体点云P，P'和接触位置c，c'后，关节运动推断模型会将点云分割为静态部件和可动部件，并估计一组关节参数，即即，$\{u^p, s^p\}$（棱柱关节）或$\{u^r, q, s^r\}$（旋转关节），用于描述连接两个物体部件的关节。最后，我们将每个物体估计得到的关节模型映射至全局坐标系。这些关节模型共同构成场景级关节模型M。

IV. 方法

我们现在介绍构建室内场景关节模型的方法，如图2所示。在接下来的小节中，我们将介绍该方法的三个关键组成部分：可供性预测、关节估计，以及对可供性和关节运动进行迭代优化。

A. 可供性预测

在初始阶段，我们必须识别潜在的交互区域并发现铰接物体。我们利用 PointNet++ [34]根据场景的初始点云观测来估计场景可供性。我们将可供性预测表述为逐点二元分类问题，把场景点云输入 PointNet++ ，从而得到逐点可供性图。机器人根据该图主动探索场景，并在相应位置与物体进行交互。由于密集可供性预测可能会引入过多需要探测的潜在位置，我们采用非极大值抑制（NMS） [11] 从每个区域中提取峰值，作为交互热点。

我们的方法如下：首先，选择得分最高的点并将其加入保留集。然后，根据一定的距离阈值抑制其邻近点。重复这一过程，直到所有点都已加入保留集或被抑制。保留集中的点即为交互热点。机器人会在每个交互热点对应的位置与物体进行实体交互。这些交互会产生信息丰富的运动，以便进一步进行关节推断。

B. 关节推断

每次有效交互后，我们都会观测到物体关节结构配置的变化。机器人会收集交互前后物体的自我中心观测。我们开发了一种关节结构网络，用于根据这些观测推断关节模型。

该网络基于我们先前的工作 Ditto [15]构建。给定物体进行关节运动前后的两组视觉观测，Ditto 可同时预测物体的三维几何结构和关节模型。原版 Ditto 将观测作为输入，但并不知道观测中产生运动的动作。相比之下，我们的智能体会与物体进行物理交互，然后收集观测。因此，我们将交互信息纳入考量，并将交互的接触区域作为网络输入的一部分。对于每一对交互，机器人通过在相应的接触位置c和c'处进行交互，采集观测点云P和P'。我们分别在点云P和P'上，以c和c'为中心生成高斯热力图。这些热力图表示发生交互的区域。

交互过程中接触区域的位置揭示了有关可动部件区域及其底层运动学约束的重要线索。因此，我们将点云观测和接触区域热图一并输入网络，以进行关节推断。

最后，网络会输出每个铰接物体的关节参数、关节状态和部件分割。为了构建最终的关节模型，我们将估计得到的物体级关节结构模型汇总为场景级模型。

C. 可供性与关节运动的迭代优化

关节模型的推断主要依赖交互过程中采集的视觉观测。如果观测涵盖了显著的关节运动，并且完整呈现物体内部，则估计得到的关节模型会更加准确。然而，由于动作无效（可能是可供性预测不准确或操作失败所致），这些观测可能会受到部分遮挡。部分受遮挡的观测会导致不准确的关节结构预测。根据经验，我们发现，与关节半开时相比，对于完全打开的旋转关节，例如，$>30°$，其关节估计的角度误差更小。

在实践中，我们发现，即使是粗略估计的关节模型，也能揭示有关部件级可供性的有用线索。关节模型提供了估计得到的关节轴位置和部件分割信息，我们可以据此预测物体部件可能发生的运动学运动。为了使物体部件产生更大的运动，我们会根据关节模型预测的运动来更新可供性。因此，我们设计了一种迭代流程：与部分打开的关节进行交互，并细化关节结构预测。

如图 3 所示，初始阶段估计的可供性未考虑关节模型，而是在可移动部件表面均匀分布。提取出的交互热点可能靠近旋转轴。该位置与人工设计的动作原语会在旋转轴上产生较小的扭矩。因此，机器人无法完全打开转动关节。在我们的实验中，初次交互时仅有 10% 的转动关节被打开至 $>30°$，这阻碍了关节估计的性能。

为了改进首次推断出的关节模型阶段，我们利用先前关节模型中的潜在运动信息，并提取物体级可供性。我们通过选择一对位置和动作，使关节运动最为显著，从而优化可供性预测。给定关节轴和部件分段后，我们选择预测的可移动部件上距离关节轴最远的点，作为下一个交互热点。我们将力方向设为轴的力矩，即关节轴方向与从交互热点到关节轴的投影之间的叉积。最后，我们根据更新后的交互热点和力方向与物体交互，并收集新的观测。新的

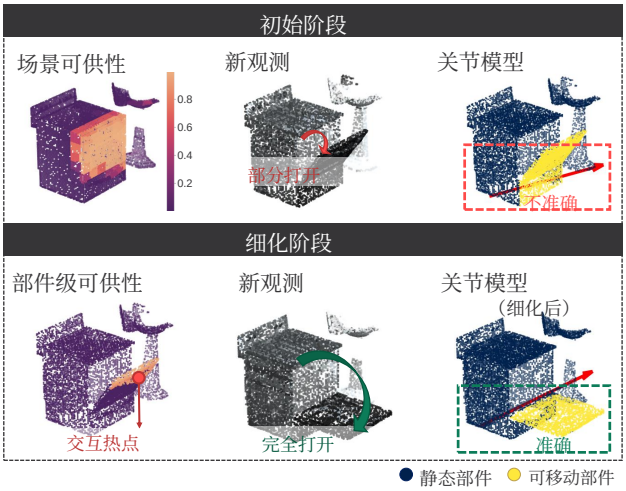

图 3：**可供性与关节运动的迭代优化**。在初始阶段，由于可供性预测不够精确，物体仅部分打开，导致关节估计不准确。在优化阶段，我们根据先前的关节估计细化物体的可供性。随后进行的新一轮交互将物体完全打开并露出内部表面，从而提高关节估计的准确性。

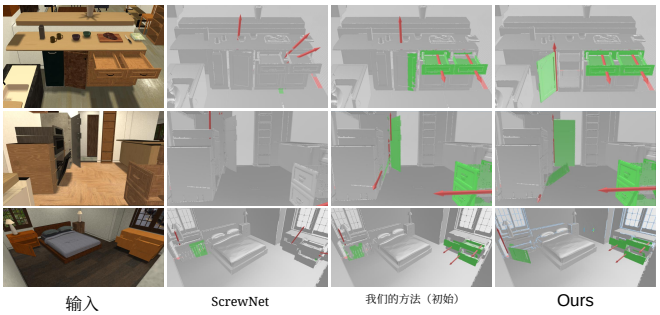

图 4：**虚拟场景上的定性结果**。静态部件以灰色表示，可动部件以绿色表示。估计关节以红色箭头表示。

观测受到的遮挡较少，从而能够准确地估计关节。

V. 实验

A. 实验设置

数据集。我们使用在仿真环境中收集的数据训练模型。具体而言，我们在 iGibson [21] 模拟器中，基于 CubiCasa5K [17] 数据集提供的场景开展实验。CubiCasa5K 是一个大规模平面图数据集，涵盖 5,000 多个场景和 80 种物体类别，例如冰箱和柜子。我们从该数据集中抽取 2,500 个场景，并将其划分为训练集、验证集和测试集，分别包含 1,500、500 和 500 个场景。由于碰撞或其他仿真问题，我们排除了可移动部件无法打开的物体。清理后，我们的测试集涵盖 1,030 个物体和 1,736 个可动部件。

探索驱动的数据采集。我们通过机器人交互来收集可供性监督。首先，在关节部件和非关节部件的表面上均匀采样位置，然后让机器人与这些位置进行交互。如果机器人成功移动了物体的任一关节部件，我们就将相应位置标记为正向可供性，否则标记为负向可供性。为模拟基于夹爪的交互，我们会在每个位置进行碰撞检查

表 I：可供性预测的定量结果。

方法	精确率	覆盖率	
		棱柱关节	旋转关节 (> 15°, > 30°)
3D AffordanceNet [5]	0.56	0.23	(0。26，0。03)
我们的方法（无细化）	0.66	0.60	(0.72, 0.10)
Ours	-	0.60	(0。72，0。55)

以确保有足够的空间放置夹爪。之后，我们在物体部件与机器人之间生成一个伪链接，然后让机器人朝某个方向拉动，即左、右或后方。为简化操作，我们采用一种简化的虚拟机器人，它能在仿真环境中执行与真实机器人类似的交互。该虚拟机器人是一个浮动球，可进行单点接触，并可瞬移至场景中的任意位置。这种简化使我们能够专注于感知和交互，而无需考虑运动规划或导航。对于每次成功的交互，我们都会收集交互前后机器人的自我中心观测，以及物体的关节模型，作为训练监督。

B. 训练细节

可供性网络。 我们通过上述探索驱动方法收集可供性数据。给定场景点云后，我们随机采样落在物体上的点，将其作为潜在交互热点。通过与这些点进行物理交互，我们根据交互结果是成功 还是失败，将潜在交互热点标记为正向 或负向 可供性。训练时，尚未进行交互的其余点将被忽略。网络以点云为输入，并使用可供性标签进行监督。由于负样本占比很大，数据分布存在不平衡。为缓解这一问题，我们采用交叉熵损失和 Dice 损失的组合来优化网络，与先前工作 [5]中的做法一致。

关节结构网络。 为收集交互前后的物体中心观测，只要不会发生碰撞，机器人便会在固定距离处从物体的正面、右侧和左侧采集多视角观测。随后，在假设已知真实相机位姿的前提下，我们将这些观测整合为部分点云。网络以部分点云为输入，并使用模拟器提供的真实关节参数进行监督。我们参照 Ditto [15], 设置损失函数及其他训练细节，但不使用占用解码器。

C. 评估指标

可供性。 为了评估可供性网络，我们报告了两个指标——覆盖率和精确率。覆盖率是成功交互的部件数占可交互部件总数的比例，而精确率是智能体尝试的交互中成功交互所占的比例。对于旋转关节的覆盖率，我们进一步将特定的打开程度设为交互成功的阈值。请注意，我们没有为棱柱关节设定额外阈值，因为大多数棱柱关节无需细化即可完全打开。

表II：关节推断的定量结果。

方法	几何	关节		
	移动性 分割交并比↑	棱柱关节	旋转	
		角度误差↓	角度误差↓	位置误差↓
ScrewNet [14]	0.34	0.28	46.19	0.13
我们的方法（无细化）	0.78	0.04	49.0	0.06
我们的方法（无规则性约束）	0.78	0.05	31.0	0.05
Ours	0.81	0.04	25.2	0.04

关节模型。 为了评估估计得到的关节模型即棱柱关节/旋转关节参数和部件分割，我们采用与 Ditto [15] 相同的指标。对于这两类关节，我们测量可移动部件的逐点分割交并比（移动分割交并比）和关节轴方向误差（角度误差）。对于旋转关节，关节轴的位置同样重要，因此我们通过预测旋转轴与真值旋转轴之间的距离来测量位置误差（位置误差）。

D. 基线方法

3D AffordanceNet [5]。 为了验证探索驱动的数据采集的有效性，我们比较了从机器人第一手的具身体验中收集的可供性标签与人工可供性标注。3D AffordanceNet 是一个用于视觉物体可供性理解的基准，提供由人工标注的三维可供性图。我们使用来自 3D AffordanceNet 的真实可供性图训练了一个与我们方法架构相同的可供性模型，并在我们的测试场景上对其进行评估。

ScrewNet [14]。 为了验证我们在关节参数化方面的选择，我们修改了模型的输出，并采用 ScrewNet [14] 中基于螺旋的关节参数化。ScrewNet 使用螺旋理论统一表示不同的关节类型，并执行类别无关的关节模型估计。我们遵循 Ditto [15] ，逐点预测稠密关节，而 ScrewNet 预测一个全局关节。

我们的方法（无细化）。我们测试了所提方法的一个消融版本，该版本仅使用初始阶段的结果。与完整方法相比，我们的方法（无细化）不会根据观测到的关节运动细化可供性，也不会进一步细化关节模型。此项对比有助于验证可供性和关节运动的迭代优化。

我们的方法（无规律性）。 我们利用动作感知中的规律性，将交互的接触区域纳入关节结构网络。在该消融版本中，我们在训练和推理期间移除了接触区域信息。我们展示此结果，以验证这些规律性对关节推断的作用。

E. 可供性预测

可供性预测的定量结果如表 I 所示。我们的方法（无细化）取得了显著优于基于 3D AffordanceNet 训练的模型的结果。人工标注的可供性偏向于人类经验，可能无法推广到机器人操作。标注场景与部署环境之间的分布偏移也带来了额外挑战

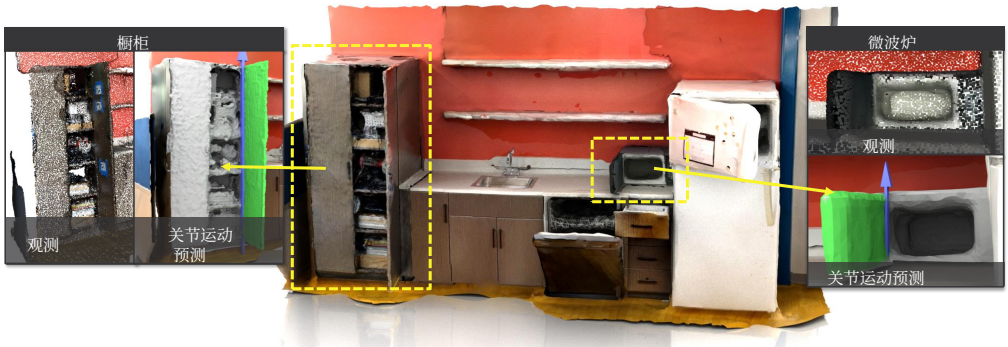

图5：**真实场景的重建关节模型**。静态部件以灰色显示，可动部件以绿色显示。估计关节以蓝色箭头表示。同一场景的更多结果见图1。

，影响泛化能力。相比之下，我们使用通过自主探索收集的数据训练可供性模型。这些数据根据机器人在环境中执行动作后的结果进行自标注。因此，训练得到的模型能够轻松适应测试场景，并表现出显著更好的交互性能。

通过对可供性和关节运动进行迭代优化，我们的最终模型进一步提升了转动关节的覆盖性能。打开超过30°的转动关节比例从 10% 增至 55%。这一提升验证了迭代优化方案。关节模型的初始估计为部件级可供性提供了重要线索。推断出的部件级可供性使关节运动更加明显，并使带有转动关节的部件打开得更充分。请注意，我们的迭代优化过程不会新增或移除已有的交互热点。因此，我们不报告优化后的精确率结果。

F. 关节推断

表II展示了关节估计结果。基于 ScrewNet 的模型 [14] 在关节估计方面表现不佳，尤其是在部件分割上。我们的方法（无规则性）未利用动作发生位置的交互信息，因此在部件分割和关节轴预测方面不如完整模型。如图4所示，基于 ScrewNet 的模型几乎无法分割出任何可动部件。该基线模型对关节轴的预测也与正确结果相去甚远。我们的方法（无细化）在带有棱柱关节的物体部件上表现好得多。然而，对于带有转动关节的橱柜门，其结果不够准确。我们的方法（无细化）分割出的门不到一半，估计的转动关节轴位置也偏离了正确位置。经过细化的模型首先将橱柜打开得更大，从而显露出更多此前被遮挡的表面。借助物体状态变化更明显的新观测，经过细化的模型能够预测出更准确的部件分割和关节参数。

G. 真实世界评估

最后，我们在一个真实的家庭场景中评估我们的方法。我们使用 iPhone 12 Pro 的激光雷达和摄像头，通过三维扫描重建该场景，而不是使用实体机器人。我们预测交互热点，并亲自动手与这些热点处的物体进行交互。随后，我们收集新观测，并运行我们的方法来构建场景级关节模型。图1和图5中的结果表明，我们的方法无需任何修改即可应用于真实场景，并能重建出准确的场景关节模型。

H. 局限性

我们使用模拟机器人夹爪和人类交互，简化虚拟与真实世界实验中的探索和物体操作过程。因此，本研究存在以下局限性：a) 探索：为了在不同地点之间移动，我们直接让机器人瞬移，未考虑导航和运动规划问题。此外，在重建静态模型时，我们假设里程计和深度估计完全准确；b) 交互：所有交互均通过在机器人与物体之间创建伪连接，或由人类完成。在物体操作过程中，我们未考虑关节约束、机器人自碰撞等问题。

VI. 结论

我们提出了一种交互式感知方法，用于构建场景级关节模型。机器人通过与关节物体进行物理交互，根据探索动作前后的视觉观测推断其关节属性。我们进一步通过在迭代过程中结合可供性预测与关节推断，提高预测质量。定量结果表明，在可供性预测和关节估计两项任务中，我们的方法都显著优于基线方法。消融研究证实，迭代优化过程能够提升这两项任务的表现。最后，我们展示了该方法能够泛化至真实世界观测，并据此构建厨房场景的关节模型。这些结果彰显了我们的方法在构建日常环境中用于机器人操作的交互式模型方面的潜力。

致谢我们感谢 Huihan Liu、Mingyo Seo 和 Soroush Nasiriany 对本文提出的宝贵意见。本研究部分得到 NSF CNS-1955523 和 FRR-2145283、德克萨斯大学奥斯汀分校机器学习实验室的 MLL研究奖，以及亚马逊研究奖的资助。

参考文献

[1] H. Abdul-Rashid、M. Freeman、B. Abbatematteo、G. Konidaris 和 D. Ritchie，“学习推断新物体实例的运动学层级结构”，载于《2022年国际机器人与自动化会议（ICRA）》，2022年，第8461–8467页。[2] J. Bohg、K. Hausman、B. Sankaran、O. Brock、D. Kragic、S. Schaal 和 G. S. Sukhatme，“交互式感知：利用感知中的动作和动作中的感知”，《IEEE机器人学汇刊》，第33卷，第6期，第1273–1291页，2017年。[3] S. Brahmbhatt、C. Ham、C. C. Kemp 和 J. Hays，“ContactDB：通过热成像分析和预测抓取接触”，载于《IEEE/CVF计算机视觉与模式识别会议论文集》，2019年，第8709–8719页。[4] A. Dearden 和 Y. Demiris，“为机器人学习前向模型”，载于 IJCAI，第5卷，2005年，第1440页。[5] S. Deng、X. Xu、C. Wu、K. Chen 和 K. Jia，“3D AffordanceNet：视觉物体可供性理解基准”，载于《IEEE/CVF计算机视觉与模式识别会议论文集》，2021年，第1778–1787页。[6] T.-T. Do、A. Nguyen 和 I. Reid，“可供性网络：一种用于物体可供性检测的端到端深度学习方法”，载于《2018年IEEE国际机器人与自动化会议（ICRA）》，IEEE，2018年，第5882–5889页。[7] D. F. Fouhey、V. Delaitre、A. Gupta、A. A. Efros、I. Laptev 和 J. Sivic，“观察人群：人类动作作为单视图几何的线索”，载于《欧洲计算机视觉会议》，施普林格，2012年，第732–745页。[8] S. Y. Gadre、K. Ehsani 和 S. Song，“各司其职：学习用于铰接物体部件发现的交互策略”，载于《IEEE/CVF国际计算机视觉会议论文集》，2021年，第15 752–15 761页。[9] A. Geiger、J. Ziegler 和 C. Stiller，“StereoScan：实时稠密三维重建 ”，载于《 2011 年 IEEE智能车辆研讨会（IV）》，IEEE，2011年，第963–968页。[10] J. J. Gibson，“《视觉知觉的生态学方法：经典版》”，心理学出版社，1979年。[11] R. Girshick、J. Donahue、T. Darrell 和 J. Malik，“用于精确物体检测和语义分割的丰富特征层级结构”，载于《IEEE计算机视觉与模式识别会议论文集》，2014年，第580–587页。[12] H. Hamer、J. Gall、T. Weise 和 L. Van Gool，“从稀疏训练数据中获取依赖于物体的手部姿态先验”，载于《2010年IEEE计算机学会计算机视觉与模式识别会议》，IEEE，2010年，第671–678页。[13] K. Hausman、S. Niekum、S. Osentoski 和 G. S. Sukhatme，“通过交互式感知进行主动铰接模型估计”，载于《2015年IEEE国际机器人与自动化会议（ICRA）》，IEEE，2015年，第3305–3312页。[14] A. Jain、R. Lioutikov 和 S. Niekum，“ScrewNet：利用螺旋理论从深度图像中进行类别无关的关节模型估计”，arXiv预印本 arXiv:2008.10518，2020年。[15] Z. Jiang、C.-C. Hsu 和 Y. Zhu，“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论文集》，2022年，第5616–5626页。[16] Z. Jiang、Y. Zhu、M. Svetlik、K. Fang 和 Y. Zhu，“可供性与几何之间的协同效应：通过隐式表示进行六自由度抓取检测”，arXiv预印本 arXiv:2104.01542，2021年。[17] A. Kalervo、J. Ylioinas、M. H¨aiki¨o、A. Karhu 和 J. Kannala，“CubiCasa5K：用于户型图像分析的数据集及改进的多任务模型”，载于《斯堪的纳维亚图像分析会议》，施普林格，2019年，第28–40页。[18] D. Katz、M. Kazemi、J. A. Bagnell 和 A. Stentz，“未知三维铰接物体的交互式分割、跟踪和运动学建模”，载于 ICRA，2013年，第5003–5010页。[19] M. Khansari、D. Kappler、J. Luo、J. Bingham 和 M. Kalakrishnan，“动作图像表示：在无需真实世界数据的情况下学习可扩展的深度抓取策略”，载于《2020年IEEE国际机器人与自动化会议（ICRA）》，IEEE，2020年，第3597–3603页。[20] K. N. Kumar、I. Essa、S. Ha 和 C. K. Liu，“利用非抓取式操作估计关节物体的质量分布”，arXiv预印本 arXiv:1907.03964，2019年。[21] C. Li、F. Xia、R. Martı’n-Martı’n、M. Lingelbach、S. Srivastava、B. Shen、K. Vainio、C. Gokmen、G. Dharan、T. Jain 等，“iGibson 2.0：以物体为中心的仿真：用于机器人学习日常家务任务，” arXiv预印本 arXiv:2108.03272，2021年。[22] X. Li、H. Wang、L. Yi、L. J. Guibas、A. L. Abbott 和 S. Song，“类别级可动关节物体姿态估计”，载于《IEEE/CVF计算机视觉与模式识别会议论文集》，2020年，第3706–3715页。[23] J. Lv、Q. Yu、L. Shao、W. Liu、W. Xu 和 C. Lu，“Sagci-system：迈向样本高效、可泛化、组合式和增量式机器人学习”，载于《2022年国际机器人与自动化会议（ICRA）》，IEEE，2022年，第98–105页。[24] R. Martı’n-Martı’n 和 O. Brock，“基于任务特定先验和多层递归估计的铰接物体在线交互式感知”，载于《2014年IEEE/RSJ智能机器人与系统国际会议》，IEEE，2014年，第2494–2501页。[25] R. Martı’n-Martı’n、S. H¨ofer 和 O. Brock，“铰接物体视觉感知的集成方法”，载于《2016年IEEE国际机器人与自动化会议（ICRA）》，IEEE，2016年，第5091–5097页。[26] K. Mo、L. Guibas、M. Mukadam、A. Gupta 和 S. Tulsiani，“Where2act：从像素到可动关节三维物体的动作”，arXiv预印本 arXiv:2101.02692，2021年。[27] T. Nagarajan、C. Feichtenhofer 和 K. Grauman，“从视频中提取有依据的人物—物体交互热点”，载于《IEEE/CVF国际计算机视觉会议论文集》，2019年，第8688–8697页。[28] T. Nagarajan 和 K. Grauman，“学习可供性景观，以探索三维环境中的交互”，《神经信息处理系统进展》，第33卷，第2005–2015页，2020年。[29] R. A. Newcombe、D. Fox 和 S. M. Seitz，“Dynamicfusion：非刚性场景的实时重建与跟踪”，载于《IEEE计算机视觉与模式识别会议论文集》，2015年，第343–352页。[30] R. A. Newcombe、S. Izadi、O. Hilliges、D. Molyneaux、D. Kim、A. J. Davison、P. Kohi、J. Shotton、S. Hodges 和 A. Fitzgibbon，“Kinectfusion：实时稠密表面映射与跟踪”，载于《2011年IEEE第十届混合现实与增强现实国际研讨会》，IEEE，2011年，第127–136页。[31] A. Noguchi、X. Sun、S. Lin 和 T. Harada，“神经可动关节辐射场”，arXiv预印本 arXiv:2104.03110，2021年。[32] M. Omran、C. Lassner、G. Pons-Moll、P. Gehler 和 B. Schiele，“神经人体拟合：统一深度学习与基于模型的人体姿态和形状估计”，载于《2018年国际三维视觉会议（3DV）》，IEEE，2018年，第484–494页。[33] S. Pillai、M. R. Walter 和 S. Teller，“从视觉演示中学习关节运动”，arXiv预印本 arXiv:1502.01659，2015年。[34] C. R. Qi、L. Yi、H. Su 和 L. J. Guibas，“Pointnet++：度量空间中点集的深度层次特征学习”，arXiv预印本 arXiv:1706.02413，2017年。[35] J. Sturm、C. Plagemann 和 W. Burgard，“通过自我感知进行无监督身体结构学习”，载于《2008年IEEE国际机器人与自动化会议》，IEEE，2008年，第3328–3333页。[36] J. Sturm、C. Plagemann 和 W. Burgard，“用于稳健机器人操作的自适应身体结构模型”，载于《机器人学：科学与系统》，苏黎世，2008年。[37] J. Sturm、V. Pradeep、C. Stachniss、C. Plagemann、K. Konolige 和 W. Burgard，“学习铰接物体的运动学模型”，载于《第二十一届国际人工智能联合会议》，2009年。[38] J. Sturm、C. Stachniss 和 W. Burgard，“用于学习铰接物体运动学模型的概率框架”，《人工智能研究杂志》，第41卷，第477–526页，2011年。[39] X. Wang、B. Zhou、Y. Shi、X. Chen、Q. Zhao 和 K. Xu，“Shape2motion：对三维形状中的运动部件及其属性进行联合分析”，载于《IEEE/CVF计算机视觉与模式识别会议论文集》，2019年，第8876–8884页。[40] Y. Weng、H. Wang、Q. Zhou、Y. Qin、Y. Duan、Q. Fan、B. Chen、H. Su 和 L. J. Guibas，“Captra：基于点云跟踪刚性及铰接物体的类别级姿态”，载于《IEEE/CVF国际计算机视觉会议论文集（ICCV）》，2021年10月，第13 209–13 218页。[41] A. Zeng、S. Song、S. Welker、J. Lee、A. Rodriguez 和 T. Funkhouser，“利用自监督深度强化学习学习推动与抓取之间的协同效应”，载于《2018年IEEE/RSJ

智能机器人与系统国际会议 *(IROS)*，IEEE，2018年，第 4238–4245 页。