PolyTouch：一种用于接触丰富操作的鲁棒多模态触觉传感器，采用触觉扩散策略

Jialiang Zhao[1], Naveen Kuppuswamy[2], Siyuan Feng[2], Benjamin Burchfiel[2], Edward Adelson[1]

本论文荣获 ICRA 2025 最佳论文奖提名。

(a) PolyTouch: tactile, acoustic, peripheral vision

(b) Bimanual Tactile-diffusion Tasks: egg serving, fruit sorting, egg cracking, wrench insertion

图 1：(a) PolyTouch 是一种结合了触觉、声学和周边视觉传感的机器人手指。(b) 我们设计了 4 个常见的双手机器人任务来评估触觉扩散策略：递鸡蛋、分拣水果、敲鸡蛋和安装扳手。

Abstract在非结构化的家庭环境中实现鲁棒的灵巧操作仍然是机器人领域的一项重大挑战。即使采用了最先进的机器人学习方法，触觉无关的控制策略（即仅依赖外部视觉和/或本体感觉的策略）也常常因遮挡、视觉复杂性以及需要精确的接触交互控制而表现不佳。为了解决这些局限性，我们引入了 PolyTouch，这是一种新型机器人手指，它将基于相机的触觉传感、声学传感和周边视觉传感集成到一个紧凑耐用的单一设计中。PolyTouch 在多个时间尺度上提供高分辨率的触觉反馈，这对于高效学习复杂的操纵任务至关重要。实验表明，其寿命比商用触觉传感器提高了至少 20 倍，并且设计易于制造且可扩展。然后，我们将这种多模态触觉反馈与视觉本体感知观测相结合，从人类演示中合成了一个触觉扩散策略；由此产生的接触感知控制策略在多项接触感知操纵策略中显著优于触觉无关策略。本文强调了有效集成多模态接触传感如何加速有效的接触感知操纵策略的开发，为更可靠、更多功能的家用机器人铺平道路。更多信息可在 https://polytouch.alanz.info/ 获取。

I. 引言

在现实世界中实现鲁棒且可靠的灵巧操纵是一个艰巨的开放性挑战，尤其是在非结构化的家庭环境中。随着人们期望家庭机器人能够掌握各种技能，这一挑战变得更加严峻。我们越来越多地寻求机器学习方法来提供可行的解决方案。特别是，通过人类演示的监督行为学习进行策略合成正日益成为一种一种强大的替代传统方法的方法，可实现必要技能的多样性[1-3]。挑战的关键在于应对不确定的接触交互。大多数技能本质上是接触丰富的，即它们通常在整个技能过程中涉及多次接触交互。

一个关键的考虑因素是，触觉无关的控制策略，即仅使用外部视觉和/或本体感觉作为反馈的策略，可能在根本上不适合处理固有的复杂性。在各种技能中经常会遇到挑战性因素，例如视觉上复杂的场景（杂乱）、外部/自身遮挡、难以看到的物体或环境（透明、反光）或难以操作的物体（关节式、可变形）。此外，成功的技能执行可能需要在应对外部干扰的同时精确调节接触力或阻抗。

在此背景下，将触觉反馈纳入控制策略框架是缓解挑战的重要且可行的方法；触觉传感器提供了一种直接感知接触状态的方式，因此可作为推断操纵器和环境状态的附加信息来源 [4–6]。然而，有两个关键问题需要回答：(a) 从概念上讲，哪种传感器模式是正确的，以及所需的信号处理架构是什么？以及 (b) 从实践上讲，为大规模数据驱动策略合成设计紧凑、鲁棒且易于构建的传感器，其最佳设计是什么？

最近传感器设计的一个发展是基于摄像头的触觉传感，即结合摄像头和可变形的反射膜来捕捉接触交互[7]。它们为基于精细纹理的触觉传感提供了一种高分辨率的解决方案，最近的文献表明，这对于需要精确控制和增强形状或物体感知能力的操纵任务至关重要。

[1] MIT CSAIL,{alanzhao, adelson}@csail.mit.edu
[2]Toyota Research Institute, firstname.lastname@tri.global

特性[8-10]。基于接触式麦克风的振动传感是另一种触觉传感解决方案，它捕捉与物体接触时产生的声波。对感知到的振动的分析可以揭示纹理、硬度和接触事件等属性[11, 12]。此外，通过周边视觉进行的接近传感也被证明对杂乱环境中的灵巧操作很有用[13]。受其中一些结果的启发，我们寻求一种结合所有这些模态的设计，利用振动传感的高频能力、基于相机的触觉传感的高空间分辨率以及周边传感在杂乱环境中的实用性。

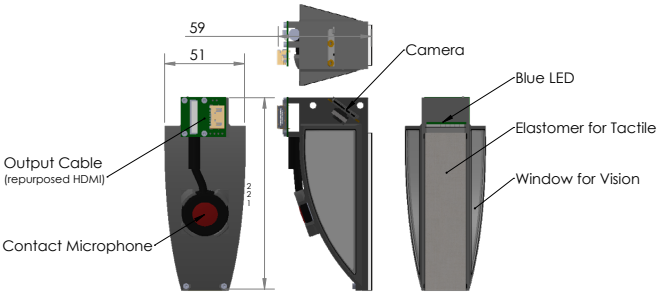

图 2：PolyTouch 的图纸和主要组件

本文介绍了PolyTouch，一种新颖的机器人手指，它将基于摄像头的纹理传感、声学传感和周边视觉传感结合到一个坚固、紧凑且易于制造的设计中。手指中集成的三种模式对于机器人高效学习复杂且接触丰富的操作策略至关重要。此外，它旨在解决当前触觉传感器在大型数据域中阻碍其可扩展性的几个关键问题，例如耐用性和可制造性差。我们的实验表明，PolyTouch的寿命比最先进的传感器长至少20倍。此外，其制造过程不需要专门的设备或专业知识，使其成为大规模机器人策略合成的理想传感器。PolyTouch的规格列于表I和图2。

本文随后利用该传感器，通过多模态传感技术合成接触感知操纵策略。我们基于扩散策略，开发了一种利用跨模态注意力的架构。然后，该触觉-扩散策略框架被用于表征和展示在富含接触的操纵中纳入触觉反馈的好处。

表 I：PolyTouch 规格

Elastomer Replacement	Gel cartridge can be quickly swapped out and replaced with new ones.
Elastomer Options	VHB tape based option (acrylic foam base material, semi-specular paint) and silicone option (silicone rubber base material, lambertian paint)
Durability	>35hrs under continuous tool using
Modalities	Texture, acoustic, and peripheral vision
Output Format	One Ethernet output (video and audio streams)
Dimension	51mm (L) x 59mm (W) x 122mm (H)
FoV for tactile	100mm x 25mm

二、相关工作

基于摄像头的纹理传感：高空间分辨率，用于精细操作。基于摄像头的触觉传感器，例如 GelSight 系列 [7]、DenseTact [14]、Bubble Grippers [15, 16]、GelSlim [17]，能够生成接触表面纹理的高分辨率视频流。它们通常包含一个柔软的接触介质（GelSight、DenseTact 和 GelSlim 中的硅胶，Bubble 中的充气空气），该介质在接触时会适应外部物体的形状，而位于另一侧的摄像头则捕捉接触表面纹理和形状的高分辨率图像。

接触。然而，尽管这些传感器提供了高度详细的信息，但存在一些众所周知的限制，阻碍了它们的广泛应用。分辨率与速度：高空间分辨率是以牺牲较低频率为代价的，这是由于相机的帧率（通常低于 100 Hz）。接触的动态信息，例如由于打滑引起的冲击和振动，需要以更高的速率进行传感。耐用性和可制造性差：柔软的弹性体承受持续的磨损和扭转。划痕、撕裂和分层是常见问题，尤其是对于硅基传感器。柔软弹性体的制造通常需要专门的设备和专业知识，这阻碍了它们的广泛应用。笨重：较大的传感表面通常需要由于光学和照明要求而将相机放置在离接触表面更远的位置，这会增加系统的笨重性。

接触式麦克风振动传感：高时间频率用于动态操作。许多研究表明，接触式麦克风捕获的较高频率振动信号可以提高需要及时冲击和碰撞检测的操作任务的性能，例如在食物切割 [18]、人机交互 [19–21] 和力水平估计 [22] 中。Liu 等人将接触式麦克风安装在手持机器人夹爪上，以收集野外数据，然后利用这些数据成功训练了机器人策略，并证明了声学数据在提高一系列动态任务的操作质量方面的实用性 [11]。

多模态感知策略学习。将多模态感知纳入机器人学习一直是活跃的研究领域。扩散策略 [3] 是一种最近提出的用于视觉运动策略学习的方法，已被证明在生成复杂的多模态动作方面非常有效和通用。该方法在机器人的动作空间上使用条件去噪扩散过程，从而形成一种策略表述，该表述可以表达任意多模态动作分布，捕获高维动作空间，并且在训练中稳定。许多工作在不同背景下选择扩散策略作为动作预测头，例如跨具身学习 [1] 和多模态感知 [23]。

所提出的设计旨在结合纹理、振动、视觉感知，同时克服了可靠性

以及当代触觉传感器面临的可制造性问题。我们还提出了一个从扩散策略扩展而来的触觉扩散框架，该框架以连贯的方式结合了多模态传感，并表明额外的模态可以提高操纵性能。

三、机械设计与规格

PolyTouch 的设计目标有三点：手指需要灵敏、耐用且易于制造。

A. *Sensing*

PolyTouch 配备了三种传感模式：基于摄像头的触觉传感、声学传感和周边自然视觉传感。

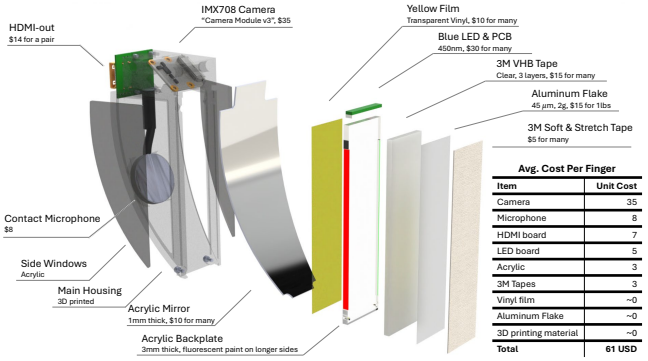

图 3：PolyTouch 的爆炸图及其估算成本。PolyTouch 的物料清单（BoM）主要由易于获取的材料组成，其构造不需要专用设备。

1) ***Camera-based tactile sensing****:* 在 PolyTouch 中，触觉感知是通过将一个 RGB 摄像头嵌入手指中，使其指向透明弹性体来实现的。我们提供两种可更换的透明弹性体选项：

- VHB 弹性体：透明 *3M VHB* 双面胶带，其外表面覆盖有 45μm 反光铝粉。
- 硅橡胶：外侧为 *Silicones Inc.* XP-565 硅橡胶，内侧为 *PRINT-ON* 灰色硅油墨。

VHB弹性体比硅酮弹性体更容易、更快地构建。然而，VHB胶带是一种粘弹性材料，在去除接触后需要一些时间才能恢复到原始形状，而硅酮弹性体响应速度要快得多。弹性体位于透明亚克力背板的顶部。亚克力的较短底部侧用蓝色LED ($\lambda = 450nm$) 照明。两条较长的侧面涂有粉红色和绿色荧光漆（Liquitex BASICS，荧光粉和荧光绿）。两种荧光漆吸收波长较短的蓝光，然后以粉红色和绿色的长波长光的形式发射能量。Adelson 在 [24] 中首次介绍了使用荧光漆作为触觉传感器照明源的方法，其优点是减小了体积和功耗。

与使用 GelSight、DIGIT 和类似传感器中的多色 LED 相比，消耗量 [7, 25]。荧光漆发出的粉色和绿色光比 LED 发出的蓝光弱。虽然人眼可以轻松区分不平衡的颜色，但蓝光很容易使大多数光学传感器饱和。为了解决这个问题，在亚克力背板内侧粘贴了一层透明的黄色乙烯基滤光片，以减少到达摄像头的蓝光量。

为了在不需要更多摄像头或增大手指尺寸的情况下实现长而大的传感区域，手指背面放置了一个曲面镜，类似于 GelSight Svelte [26, 27] 和 GelLink [28]。镜子的曲率和放置方式设计如下：(1) 亚克力背板的整个内表面都可以从摄像头的视场中看到；(2) 摄像头应该能近乎垂直地观察背板，以便所有方向的光线都能被摄像头看到，这意味着嵌入颜色中的深度信息将得到更好的保留；以及 (3) 曲面镜引入的畸变应被最小化。最终设计的反射模拟结果如图 4 所示。

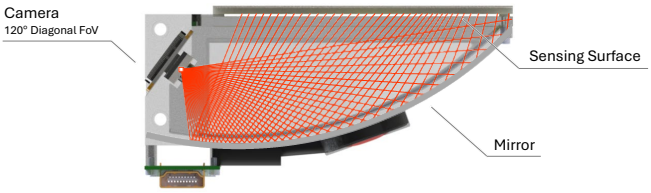

图 4：PolyTouch 的光学模拟。传感表面反射的光在到达相机之前被曲面镜重新分布。相机传感器为 4:3，对角线视场角为 120 度。

2) ***Contact microphone-based acoustic sensing****:* 将压电接触式麦克风放置在 Poly-Touch 的背面以收集声学信号。接触式麦克风以 48 kHz 的采样率进行采样，音频文件与触觉视频流同步保存或流式传输。音频和视频均由 Raspberry Pi 录制，然后 Raspberry Pi 会对信号进行转码，再将它们保存为 .mp4 文件或流式传输到机器人站。

3) ***Peripheral vision sensing****:* Poly-Touch 内的摄像头不仅收集触觉信息，还能借助两个侧窗收集周边视觉信号。侧窗与镜子一起有助于实现可见区域，包括接触表面的周围区域和传感器下方的区域。周边视觉感知的可见区域演示如图 1 所示。

B. *Durability and replaceability*

基于摄像头的触觉传感器的耐用性问题通常与传感器中的弹性体有关，这些弹性体必须与外部环境直接且重复地接触。问题有两个方面：弹性体在过大的力/扭矩下会分层，以及表面

在连续磨损下会磨损或撕裂。分层问题主要影响使用硅凝胶作为弹性体的传感器，因为硅胶难以粘合到非硅胶材料上，同时保持光学清晰度。PolyTouch 的 VHB 弹性体选项提供了一种替代方案，其本身具有高粘性，几乎完全消除了分层问题。

通过在传感器的传感表面上应用保护膜，可以提高耐磨损和抗扭曲性能。一种流行的选择是3M Tegaderm薄膜，它最初是作为伤口保护膜推出的。然而，这种薄膜很薄，在持续的磨损和扭曲下寿命有限，而且容易起皱。相反，PolyTouch选择3M Nextcare Soft & Stretch胶带作为外层保护层，它具有与人体皮肤相似的表面特性，而且不易起皱。

PolyTouch 还设计有快速弹性体更换机制，以便凝胶可以轻松快速地滑出并更换为新的。

C. Manufacturability

PolyTouch 的设计旨在最大限度地减少施工对特殊设备或专业知识的要求。PolyTouch-VHB 的弹性体制造不需要任何专用设备或专业知识：只需将胶带粘贴到激光切割的亚克力背板上，然后在胶带外侧擦拭铝粉。一个没有经验但具备基本动手能力的人制作一个 VHB 弹性体不到 5 分钟。PolyTouch-Silicone 中硅凝胶的制造过程与其他流行的软传感器相似，例如 GelSight [7]、DenseTact [14]、ReSkin [29]。

四、机器人从多模态感知中学习

我们采用了一个固定基座的双臂硬件平台，该平台由两个以直立配置安装的弗兰卡熊猫（Franka Panda）机器人手臂组成，如图1所示。每只手臂都配备了一个腕部安装的摄像头（FLIR Blackfly S）。两个固定的场景摄像头（FRAMOS D415e）被放置在能够良好地观察活动工作空间的位置。机器人本身在每个夹爪上都配备了两种手指：一种是带有凹槽的被动柔性 *fin-ray* [30]，使用TPU 3D打印而成，另一种是PolyTouch-VHB手指。这种特殊的夹爪配置能够很好地抓取任意形状的被操纵物，并能牢固地抓持工具手柄。本文分析的所有技能都在机器人设置旁边的工作台上进行演示，在数据收集过程中，机器人通过两个6自由度（6-DoF）的SpaceMouse进行遥操作。

A. Modality Encoding

时间步 t 的观测及其编码方法详述如下：

- 触觉和周边视觉 O^t_{tp} 以 RGB 图像的形式从安装在每个手臂上的 PolyTouch 获取。它们由预训练的 T3 [31] 编码器进行编码。T3 是一个一个触觉表示学习框架，该框架从大规模多传感器和多任务数据集中预先训练。
- 手腕视图 O^t_{wrist} 和场景视图 O^t_{scene} 是从双臂获取的 RGB 图像。它们由预训练的 CLIP 特征提取器 [32] 进行编码。
- 音频信号 O^t_{aud} 以声波的形式从安装在每个臂上的 PolyTouch 获得。它们首先被转换为对数梅尔频谱图，然后输入到音频频谱图 Transformer (AST, [33]) 中进行特征提取。
- 双臂本体感觉 O^t_{prop}，包括实际和期望的六维末端执行器位姿和夹爪宽度，由 MLP 编码，然后与其他模态结合。

一个 *Modality Combiner* 被设计用来组合来自每个模态的编码特征。预训练的 CLIP 视觉特征提取器和预训练的 T3 触觉特征提取器的骨干架构都是 Vision Transformers [34]，来自 O^t_{tp} 和 O^t_{scene} 的编码特征通过一个 6 块 12 头交叉注意力进行组合。然后，输出被池化（通过提取分类 token）并与来自其他模态的池化或投影特征通过一个连接和投影层进行组合。

B. Policy Learning

组合特征，表示为 O^t，被馈送到扩散策略（U-Net 变体）主干以进行动作预测（观测历史：2，动作预测范围：16，执行动作：8）。

V. 实验与讨论

一项耐久性实验旨在测试 PolyTouch 的寿命，一项操控学习实验旨在测试 PolyTouch 的附加模式的实用性。

A. Durability Testing

我们将 PolyTouch 的寿命与 GelSight Mini [35] 的寿命进行了比较，GelSight Mini 是一种流行且商业上可用的基于相机的触觉传感器。我们的实验模拟了机器人手指在使用家用工具的任务中所承受的磨损和扭转。机器人夹爪的一侧装有 GelSight Mini 传感器，另一侧装有 PolyTouch-VHB。塑料刮刀固定在工作台的固定位置。机器人以 10 至 30 N 之间的随机力抓握刮刀手柄。然后，机器人应用连续的随机 6D 运动，其中平移从 -10 至 10 毫米采样，旋转从 x、y、z 轴上的 -5 至 5 度采样。图 6 显示了故障前的时间和传感器表面的图片。

PolyTouch-VHB 在总计 35 小时的连续摩擦中均未出现故障或图像质量下降。测试了 GelSight Inc. 提供的两种不同的标准凝胶，其中一种在凝胶完全分离前持续了 1.0 小时，而另一种在油漆脱落前持续了 3.3 小时。我们更新的含 XP-565 的硅凝胶配方（与 PolyTouch-Silicone 上的材料相同）在 GelSight Mini 上持续了 25.0 小时。

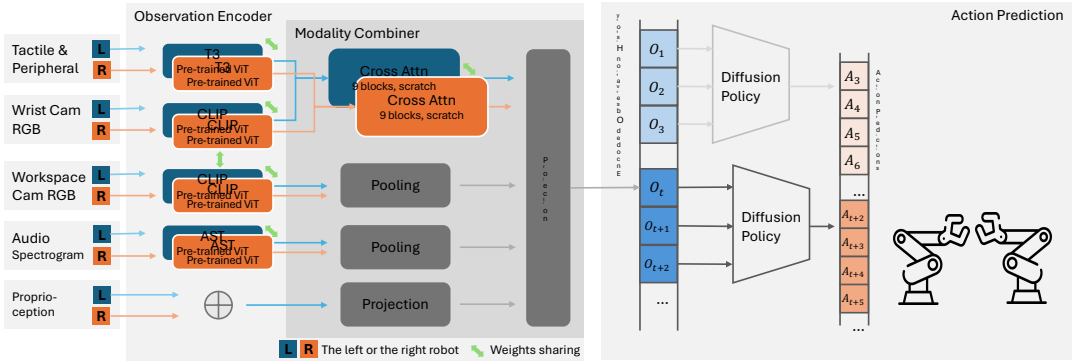

图 5：触觉扩散策略网络。来自两只手臂的感官模态均由预训练的特征提取器进行编码（本体感觉除外，它由一个从头开始的 MLP 进行编码），然后通过交叉注意力混合和连接进行组合，最后输入扩散头以预测动作。

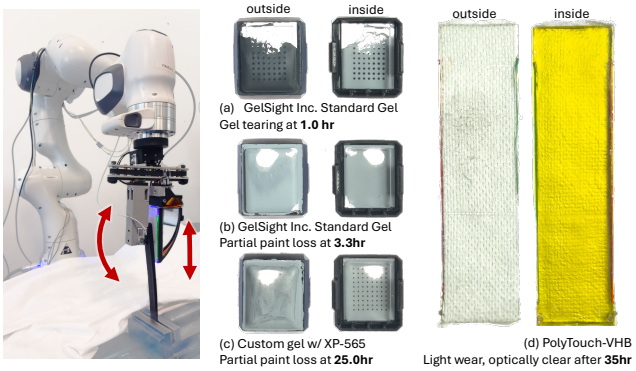

图 6：在模拟工具使用环境下的弹性体耐久性测试。Franka Panda 机器人对固定的柔性刮刀手柄进行持续的摩擦和刮擦。GelSight Inc. 生产的 GelSight Mini 和 PolyTouch -VHB 安装在相对位置。

B. Multi-modal sensing for bimanual manipulation

我们为本节设计了 4 个任务和 3 个网络变体：

BimanualInsertWrench：一个机器人拿起T型手柄的内六角扳手，将其传递给另一个机器人，然后两个机器人协同工作，瞄准并插入扳手，直到完全插入。收集约200个数据点用于训练。

BimanualSortFruit：将外观相似但表面纹理和柔软度不同的四种水果用两只手臂分拣到箱子中。收集约150个数据点用于训练。

BimanualEggCracking：一个两件式玩具蛋由一只手臂拾起，然后在另一只手臂的帮助下在平底锅上方掰开。收集约 70 个数据点用于训练。

BimanualEggServing：一个机器人借助另一只手臂拿起锅铲，然后机器人用锅铲舀起玩具煎蛋，最后将煎蛋轻轻放在一片吐司上。收集了大约150个数据点用于训练。

每个任务收集的数据点数量是根据任务难度凭经验选择的。每个任务的说明如图1所示。我们训练和评估了三个每个任务的网络配置：

visuo-proprio 是其中触觉-周边视觉模态 O^t_{tp} 和音频信号 O^t_{aud} 被掩盖（替换为零），并且不使用交叉注意力（编码输出直接池化和投影）的变体；

multi-concate 使用所有模态，并直接使用池化和投影代替交叉注意力；

multi-crossatn 如图 5 所示，所提出的方法将所有模态与交叉注意力模块一起使用。

除是否使用交叉注意力以及是否屏蔽触觉输入外，所有网络变体都共享相同的架构和超参数。对于双臂鸡蛋服务任务，我们增加了两个实验，分别使用总数据的 1/3 和 2/3 进行训练。所有实验均在 AWS G5.48xLarge 节点（每个节点配备 8× 块 Nvidia A10G）上以 10 的批次大小训练了 500 个 epoch。在训练和评估期间，都对物体/工具的初始姿态进行了随机化处理。我们报告了两个指标的评估结果：*average task progress* 和 *average task success*。每个任务被分为 3-7 个阶段。*average task progress* 指的是每次评估运行成功完成的阶段的平均进度，而 *average task success* 代表每次评估运行的平均二元成功率。

我们强调 (1) ***visuo-proprio*** 策略是一种最先进的视觉运动策略，并结合了大量的视频馈送（总共 4 个）；(2) 评估的四个任务是标准任务，不像许多之前的触觉操纵研究中常见的任务那样，它们不是专门为触觉传感而精心设计的。我们的实验表明：

1) Tactile-inclusive policies are more robust than SOTA visuomotor policy: 使用触觉传感模态作为输入的策略变体始终优于 ***visuo- proprio*** 策略，如表 II 所示。

2) Some failure modes are unique to tactile-oblivious policies: 这些失效模式在任一包含触觉的策略中均未出现或很少出现。

在两种情况下都观察到了过度或不足的力，包括插入时施加的力过大

Task	Avg. Task Progress				Avg. Task Success			
	visuo-proprio	***multi-concate***	***multi-crossatn***	Absolute improvement	***visuo-proprio***	***multi-concate***	***multi-crossatn***	Absolute improvement
Insert Wrench	47%	60%	59%	+12%	0%	20%	18%	+18%
Sort Fruit	53%	63%	70%	+17%	33%	46%	46%	+13%
Crack Egg	70%	71%	72%	+1%	50%	53%	54%	+3%
Serve Egg (1/3 data)	10%	-	17%	+7%	0%	-	0%	-
Serve Egg (2/3 data)	53%	-	56%	+3%	13%	-	33%	+20%
Serve Egg (all data)	81%	88%	100%	+19%	66%	73%	100%	+34%

表 II：各任务和网络变体的评估性能。最佳性能以绿色标记，最差性能以红色标记。 在红色。绝对改进是***multi-crossatn*** (提出的方法)与***visuo-proprio*** (基线)之间的差值。

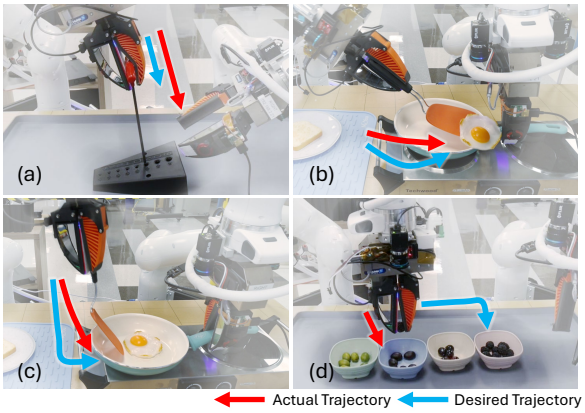

图 7：***visuo-proprio*** 策略特有的故障模式。(a) 拧紧器插入时施加的力过大。(b) 铲煎蛋时角度不足。(c) 铲动时向下压抹刀的力不足。(d) 视觉差异细微的水果分错。

扳手和舀鸡蛋时铲子按压得不够用力。这两种问题在使用包含触觉的策略时都没有发生。我们认为触觉感知，特别是基于纹理的感知模式，为策略学习者提供了信息，使其能够更好地调节力。

在抓取精度方面，***visuo-proprio***策略（评估期间有11例）比包含触觉的策略（平均3例）观察到更低的精度，包括抓取工具手柄和过于靠近边缘的水果。我们假设周边视觉模式通过在抓取过程之前和期间提供近距离视图，有助于减少此类精度问题。

在水果分拣实验中，基于纹理的分类，特别是黑莓与蓝莓的分拣任务，采用包含触觉的策略（***visuo-proprio***为20%时，包含触觉策略成功率为80%）显示出更高的成功率。这表明触觉在区分表面纹理相似的物体方面可能起着关键作用。

3) Training robotic policies with more modalities might require more data: 当我们用更少的数据训练策略时，我们注意到绝对性能提升有所下降。具体来说，平均任务进度提高了 19%

然而，在 ***multi-crossatn*** 策略优于 ***visuo-proprio*** 策略的情况下，使用 2/3 和 1/3 的训练数据，改进幅度仅分别为 3 % 和 7%。平均成功率也观察到相同的趋势，使用全部数据或 2/3 数据训练时，性能分别提高了 34% 和 20%。尽管多模态策略的性能始终优于 ***visuo-proprio*** 策略，但这表明可能需要更多数据才能充分发挥多模态作为输入的潜力。这个问题可能是由于模型尺寸较大，并且可以通过使用在其他领域数据（例如在不同具身或不同任务上收集的数据）预训练的基础策略来缓解。

六、局限性与未来工作

使用VHB胶带作为相机式触觉传感器的弹性体是一种新颖的方法，它对制造设备和专业知识的要求最低，旨在吸引更广泛的机器人社区成员在其流程中采用触觉传感器。尽管其易于制造且灵敏度高，但VHB胶带中使用的丙烯酸泡沫基材的粘度引起的滞后可能会导致响应变慢或在动态操作任务中产生歧义。这个问题可以通过软件方面来缓解，通过获取时间差图像并交叉检查周边视觉信号。寻找一种易于制造且粘度较低的新材料可能是未来潜在的改进方向。

VII. 结论

在这项工作中，我们提出了 PolyTouch，这是一种坚固且易于制造的机器人手指，它将触觉、声学和外围传感模式结合在一个紧凑的尺寸中。我们介绍了两种弹性体配方，包括一种新颖的基于 VHB 胶带的配方，以实现终极的制造简易性，以及支持它们的制造工具。与最先进的替代品相比，我们的传感器更加耐用，其紧凑的尺寸和多模态特性使其自然成为大规模机器人策略合成研究的理想选择。然后，我们提出了一个触觉扩散策略框架，该框架将多模态触觉信息与视觉和本体感觉相结合。我们的实验表明，在接触丰富的操作中，通过多模态触觉传感训练的策略始终优于最先进的视觉运动策略。

参考文献

[1] O. M. Team, D. Ghosh, H. Walke, K. Pertsch, K. Black, O. Mees, S. Dasari, J. Hejna, T. Kreiman, C. Xu, *et al.*, "Octo: An open-source generalist robot policy," *arXiv preprint arXiv:2405.12213*, 2024. [2] L. Wang, X. Chen, J. Zhao, and K. He, "Scaling proprioceptive-visual learning with heterogeneous pre-trained transformers," in *Arxiv*, 2024. [3] C. Chi, S. Feng, Y. Du, Z. Xu, E. Cousineau, B. Burchfiel, and S. Song, "Diffusion policy: Visuomotor policy learning via action diffusion," *arXiv preprint arXiv:2303.04137*, 2023. [4] A. Bronars, S. Kim, P. Patre, and A. Rodriguez, "Texterity: Tactile extrinsic dexterity," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 7976–7983. [5] J. Zhao, M. Bauza, and E. H. Adelson, "Fingerslam: Closed-loop unknown object localization and reconstruction from visuo-tactile feedback," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 8033–8039. [6] S. Suresh, H. Qi, T. Wu, T. Fan, L. Pineda, M. Lambeta, J. Malik, M. Kalakrishnan, R. Calandra, M. Kaess, *et al.*, "Neural feels with neural fields: Visuo-tactile perception for in-hand manipulation," *arXiv preprint arXiv:2312.13469*, 2023. [7] W. Yuan, S. Dong, and E. H. Adelson, "Gelsight: High-resolution robot tactile sensors for estimating geometry and force," *Sensors*, vol. 17, no. 12, p. 2762, 2017. [8] M. Burgess, "Learning object compliance via young's modulus from single grasps with camera-based tactile sensors," *arXiv preprint arXiv:2406.15304*, 2024. [9] W. Yuan, R. Li, M. A. Srinivasan, and E. H. Adelson, "Measurement of shear and slip with a gelsight tactile sensor," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 304–311. [10] S. Dong, W. Yuan, and E. H. Adelson, "Improved gelsight tactile sen- sor for measuring geometry and slip," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 137–144. [11] Z. Liu, C. Chi, E. Cousineau, N. Kuppuswamy, B. Burchfiel, and S. Song, "Maniwav: Learning robot manipulation from in-the-wild audio-visual data," *arXiv preprint arXiv:2406.19464*, 2024. [12] S. Jin, H. Liu, B. Wang, and F. Sun, "Open-environment robotic acoustic perception for object recognition," *Frontiers in neurorobotics*, vol. 13, p. 96, 2019. [13] A. Yamaguchi and C. G. Atkeson, "Combining finger vision and optical tactile sensing: Reducing and handling errors while cutting vegetables," in *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2016, pp. 1045–1051. [14] W. K. Do and M. Kennedy, "Densetact: Optical tactile sensor for dense shape reconstruction," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 6188–6194. [15] N. Kuppuswamy, A. Alspach, A. Uttamchandani, S. Creasy, T. Ikeda, and R. Tedrake, "Soft-bubble grippers for robust and perceptive manipulation," *International Conference on Intelligent Robots and Systems (IROS)*, 2020. [16] A. Alspach, K. Hashimoto, N. Kuppuswarny, and R. Tedrake, "Soft-bubble: A highly compliant dense geometry tactile sensor for robot manipulation," in *RoboSoft*, 04 2019, pp. 597–604. [17] E. Donlon, S. Dong, M. Liu, J. Li, E. Adelson, and A. Rodriguez, "Gelslim: A high-resolution, compact, robust, and calibrated tactile- sensing finger," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 1927–1934. [18] K. Zhang, M. Sharma, M. Veloso, and O. Kroemer, "Leveraging multimodal haptic sensory data for robust cutting," in *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2019, pp. 409–416. [19] X. Fan, D. Lee, Y. Chen, C. Prepscius, V. Isler, L. Jackel, H. S. Seung, and D. Lee, "Acoustic collision detection and localization for robot manipulators," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 9529–9536. [20] X. Fan, R. Simmons-Edler, D. Lee, L. Jackel, R. Howard, and D. Lee, "Aurasense: Robot collision avoidance by full surface proximity de- tection," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 1763–1770. [21] J. J. Gamboa-Montero, F. Alonso-Martin, J. C. Castillo, M. Malfaz, and M. A. Salichs, "Detecting, locating and recognising human touches in social robots with contact microphones," *Engineering Applications of Artificial Intelligence*, vol. 92, p. 103670, 2020.

[22] M. Ono, B. Shizuki, and J. Tanaka，"使用主动声学传感感知触觉力"，载于 *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*，2015年，第355–358页。[23] L. Wang, J. Zhao, Y. Du, E. H. Adelson, and R. Tedrake，"Poco: 用于异构机器人学习的策略组合"，载于 *arXiv preprint arXiv:2402.02511*，2024年。[24] E. Adelson，"具有荧光照明的逆向图形传感器"，2024年2月22日，美国专利申请 18/267,032。[25] M. Lambeta, P.-W. Chou, S. Tian, B. Yang, B. Maloon, V. R. Most, D. Stroud, R. Santos, A. Byagowi, G. Kammerer, *et al.*，"Digit: 一种用于低成本紧凑型高分辨率触觉传感的新型设计，并应用于手内操作"，载于 *IEEE Robotics and Automation Letters*，第5卷，第3期，第3838–3845页，2020年。[26] J. Zhao and E. H. Adelson，"Gelsight svelte: 一种人手指形单摄像头触觉机器人手指，具有大传感覆盖范围和本体感觉"，载于 *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*。IEEE，2023年，第8979–8984页。[27] J. Zhao and E. Adelson，"Gelsight svelte hand: 一种三指、两自由度、触觉丰富、低成本的机器人手，用于灵巧操作"，载于 *arXiv preprint arXiv:2309.10886*，2023年。[28] Y. Ma, J. Zhao, and E. Adelson，"Gellink: 一种具有基于视觉的触觉传感和本体感觉的紧凑型多指节手指"，载于 *2024 IEEE International Conference on Robotics and Automation (ICRA)*。IEEE，2024年，第1107–1113页。[29] R. Bhirangi, T. Hellebrekers, C. Majidi, and A. Gupta，"Reskin: 通用、可更换、持久的触觉皮肤"，载于 *5th Annual Conference on Robot Learning*，2021年。[30] W. Crooks, G. Vukasin, M. O'Sullivan, W. Messner, and C. Rogers，"受鳍条®效应启发的软机器人夹持器：从Robosoft大挑战到优化"，载于 *Frontiers in Robotics and AI*，第3卷，第70页，2016年。[31] J. Zhao, Y. Ma, L. Wang, and E. H. Adelson，"可迁移的触觉Transformer，用于跨不同传感器和任务的表示学习"，载于 *arXiv preprint arXiv:2406.13640*，2024年。[32] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, *et al.*，"从自然语言监督中学习可迁移的视觉模型"，载于 *International conference on machine learning*。PMLR，2021年，第8748–8763页。[33] Y. Gong, Y.-A. Chung, and J. Glass，"Ast: 音频频谱图Transformer"，载于 *arXiv preprint arXiv:2104.01778*，2021年。[34] A. Dosovitskiy，"一张图像值16x16个词：用于大规模图像识别的Transformer"，载于 *arXiv preprint arXiv:2010.11929*，2020年。[35] G. Inc.，"GelSight Mini"，https://www.gelsight.com/gelsightmini/，2022年，[在线]。