arXiv:2602.01153v1 [cs.RO] 1 Feb 2026

UniForce：面向多种触觉传感器机器人操作的统一潜在力模型

Zhuo Chen*、Fei Ni†、Kaiyao Luo*、Zhiyuan Wu*、Xuyang Zhang*、Emmanouil Spyrakos-Papastavridis*、Lorenzo Jamone†、Nathan F. Lepora‡、Jiankang Deng† 和 Shan Luo*
*伦敦国王学院 †帝国理工学院 ‡伦敦大学学院 §布里斯托大学

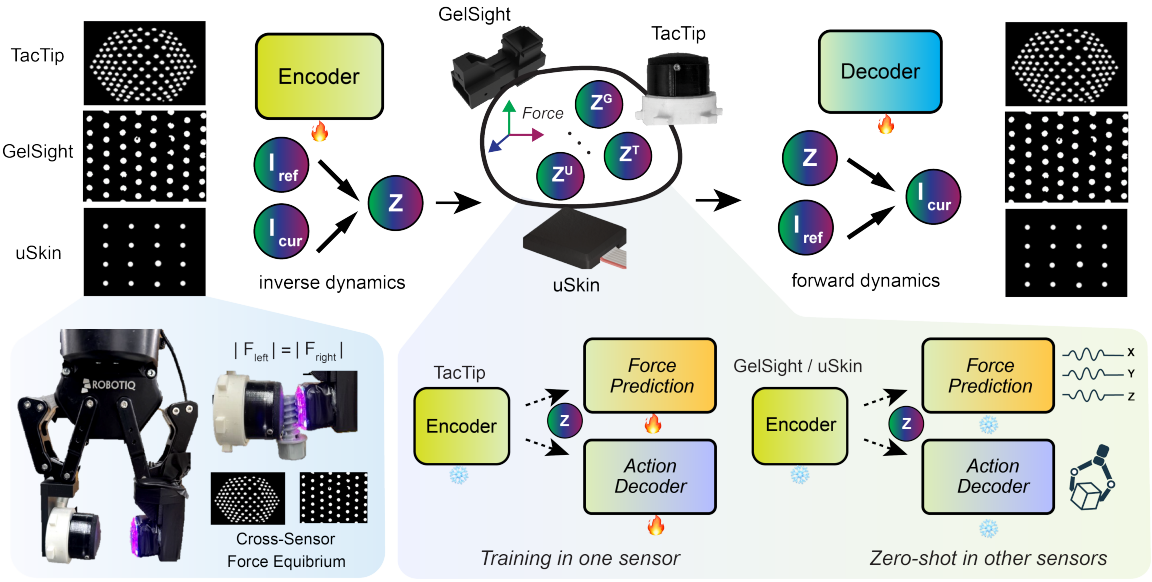

图 1：我们提出了**UniForce**，这是一个框架，能够学习适用于异构触觉传感器的机器人操作共享潜在力模型。UniForce 利用来自力平衡抓取的配对数据，预训练通用编码器，从统一的标记图像表示中提取具有物理意义的潜在力，而无需明确的力标签。随后，可通过使用单个传感器的数据训练专用预测头，将预训练编码器接入下游任务，从而实现**零样本**迁移，无需微调即可进行跨传感器力预测和力感知策略学习。

摘要——**力感知对于灵巧的机器人操作至关重要，但触觉传感器的异质性阻碍了力感知策略学习的规模化。传感原理（例如光学与磁学）、外形尺寸和材料的差异，通常要求针对每种传感器单独采集数据、进行校准和模型训练，从而限制了泛化能力。我们提出 UniForce，这是一种新颖的统一触觉表征学习框架，可在多种触觉传感器之间学习**共享潜在力空间。**UniForce通过联合建模逆动力学（图像到力）和正向动力学（力到图像），并以力平衡和图像重建损失作为约束，生成以力为基础的表征，从而减少跨传感器域偏移。为避免依赖昂贵的外部力/扭矩（F/T）传感器，我们利用静态平衡，通过传感器–物体–传感器直接交互采集力配对数据，从而实现基于接触力的跨传感器对齐。由此得到的通用触觉编码器可直接用于下游力感知机器人操作任务，实现零样本迁移，无需重新训练或微调。我们在包括 GelSight、TacTip 和 uSkin 在内的异构触觉传感器上进行了大量实验，结果表明，与以往方法相比，力估计性能均得到提升；此外，该方法还使视觉-触觉-语言-动作（VTLA）模型能够在机器人擦拭任务中实现有效的跨传感器协同。代码和数据集将公开发布。**

I. 引言

具有人类水平的机器人操作能够显著受益于触觉反馈 [13]，，尤其是力感知 [11]。然而，学习能够在不同触觉传感器之间迁移的力感知模型仍然充满挑战。关键挑战在于传感器的异质性：触觉传感器在感知原理（例如光学 [20, 16]，、磁学 [15]，和电容式 [2]）、信号模态（例如图像与多通道信号）、外形规格（例如平面与曲面）以及机械特性（例如软质与硬质弹性体）方面各不相同。因此，在一种传感器（例如 TacTip [16]）上训练的力感知策略和力预测模型，部署到 GelSight [20] 和 uSkin [15]等其他传感器上时，性能可能会显著下降。此外，为每种新传感器采集力标签并重新训练/微调成本高昂，在大规模应用中通常并不切实际，这促使人们寻求一种能够在异质传感器之间实现泛化的力感知触觉表示。

我们认为，要弥合这一差距，必须将传感器无关的接触力与传感器特定的触觉信号解耦。机器人自身的具身性为实现这种解耦提供了一种可扩展的监督信号：**跨传感器力**

平衡。在使用配备异构触觉传感器的两根手指进行准静态双边抓取时，物理规律要求相对的接触力大小相等，尽管不同传感器产生的触觉测量可能存在显著差异。基于这一先验，我们提出**UniForce**，一种能够学习由多种触觉传感器共享的统一潜在力模型的框架。不同于先前工作 [3, 4, 7, 14] 依赖位置配对数据和跨传感器图像到图像转换、却未将表征明确置于接触动力学基础之上，UniForce 学习：(i) 一个结合逆动力学的编码器，将触觉图像映射为力；(ii) 一个结合正向动力学的解码器，根据力重建触觉图像（以参考图像为条件），从而能够基于力配对的触觉图像进行自监督学习。这一以物理为基础的目标促使学习到的潜在力空间对传感器硬件保持不变，而不是像先前方法 [6, 17, 23]所研究的那样，将力与传感器特定的外观、像素或语义纠缠在一起。

据我们所知，UniForce 是首个能够学习涵盖基于视觉和非视觉式触觉传感器的潜在力空间的框架。近期的一项工作 GenForce [4],通过将力标签从一个传感器迁移到其他传感器，实现了跨触觉传感器的可迁移力感知，但需要针对每个目标传感器进行基于扩散的图像转换、力模型训练和材料补偿。相比之下，UniForce 以端到端方式学习统一的潜在力空间和通用编码器，可直接用于力感知策略学习和力预测，并实现零样本迁移，无需针对特定传感器重新训练或微调。本工作的贡献总结如下：

- **提出了一种具有通用编码器的统一潜在力模型** UniForce，通过将异构触觉传感器（例如基于视觉和非视觉式触觉传感器）映射到共同的潜在力空间中进行对齐，从而学习跨传感器共享的力感知表征。
- **介绍了一种基于物理原理、无需标签的力配对数据采集流程**，该流程利用准静态力平衡，在不依赖外部力/扭矩（F/T）传感器的情况下，获取异构触觉传感器之间的力配对触觉信号。
- **面向力感知机器人操作任务的零样本迁移**由此实现：通用编码器可将知识迁移到不同触觉传感器上的下游任务，包括力估计和力感知策略学习，且无需针对特定传感器重新训练或微调。

II. 相关工作

A. 异构触觉传感器

触觉传感器的特点是具有柔软的皮肤和受人类皮肤启发的传感层 [13]。传感层可以采用不同的换能原理，包括光学 [20, 16],、磁学 [15], 、压电容式 [2], 和压阻式 [12] 设计，它们在分辨率、灵敏度和鲁棒性方面各有不同的权衡。此外，外层皮肤和外形规格通常会针对特定操作任务进行定制。例如，GelSight [20] 是一种平面视觉触觉传感器，可输出高分辨率RGB图像，常用于纹理识别和位姿估计。TacTip [16] 是一种仿生视觉触觉传感器，具有弯曲表面和内部针脚，因此对剪切力和接触力敏感。uSkin [15]是一种基于霍尔效应传感的磁性触觉传感器，可高频输出多通道信号，适用于动态交互。由于存在异质性，力感知策略通常针对特定传感器进行训练，无法直接迁移到其他传感器，甚至无法迁移到配备不同外层皮肤的同一种传感器上 [4]。我们的工作通过学习一种不受传感器硬件影响、以力为基础的表征，旨在解决这种碎片化问题。

B. 跨传感器表征学习

学习共享触觉表征并实现跨传感器迁移，近来已成为触觉机器人学领域受到关注的方向。现有工作大致可分为两类。**(1) 传感器家族内的表征学习**。多种方法为基于视觉的触觉传感器学习共享潜在空间，尤其针对 GelSight 系列。AnyTouch [6] 在静态和动态触觉观测之间学习统一表征。T3 [23] 采用共享主干网络，并搭配传感器专用编码器和任务专用解码器。UniT [17]采用类似 VQ-VAE 的目标函数学习触觉表征。这些方法在有限的硬件范围内展现出较强的迁移能力，但通常假设感知模态相似，且未明确约束潜在空间对应于某种物理变量（例如接触力），因而限制了其向非视觉传感器迁移以及在机器人技能学习中的部署能力。**(2) 通过自监督或转换实现跨传感器迁移**。另一类方法利用自监督学习或触觉域之间的图像到图像转换来实现跨传感器迁移 [7, 14]。然而，其中许多方法针对特定传感器对进行训练，随着传感器数量增加，尤其是在换能原理各异的情况下，难以扩展。GenForce [4] 与本文工作密切相关，其目标是在异构传感器间实现可迁移的力感知，但依赖力标签、逐传感器转换和训练。与之不同，UniForce 通过由跨传感器力平衡约束的共享潜在力空间对齐异构传感器，旨在支持跨模态即插即用的零样本使用。

C. 力感知机器人操作

力信号和触觉观测正日益融入策略学习，例如视觉-语言-动作（VLA）模型 [10]。一种常见策略是使用预训练的触觉编码器生成触觉词元，用于高层策略调节，例如 OmniVTLA [5] 和 TactileVLA [9],，这类方法通常基于基于视觉的触觉表征模型（例如 AnyTouch [6]）。另一种策略则将触觉信息融合到高层感知和低层动作解码中，以提升鲁棒性 [1]。然而，现有触觉编码器的训练目标是使其与语言语义建立联系，这限制了它们在以力为中心的操作任务以及

非视觉式触觉传感器中的实用性。ForceVLA [19]通过混合专家设计纳入六维腕部力/力矩传感，但腕部力/力矩测量无法直接解析滑移或分布式抓取力等接触状态变量。另一方面，直接使用基于触觉传感的力估计也面临实际挑战，其中最突出的是：每种传感器都需要昂贵的配对力-触觉数据和力模型训练 [3]。UniForce通过提供一种通用的力锚定编码器来补充这些方向，无需力标签，即可为异构触觉传感器提供共享潜在空间。

III. 方法

我们提出 **UniForce**，这是一种利用准静态力平衡，在异构触觉传感器之间学习共享潜在力表示的框架。其关键观察是，尽管不同传感器的传感原理、外形规格和皮肤材料各异，导致触觉测量差别很大，但双侧准静态抓取中的底层接触交互会施加一项物理约束：两个指尖承受的接触力近似等大反向。UniForce 将这一约束转化为隐式监督，通过逆向—正向动力学，在共同的潜在力空间中对齐异构触觉观测。

A. 问题表述

令 $\mathcal{S} = \{s_1, \ldots, s_N\}$ 表示一组异构触觉传感器。每个传感器的触觉单元布局或物理标记图案可能各不相同。遵循先前工作 [4], ，我们将每个传感器的原始信号规范化为统一的二维标记图像表示，从而能够在不同传感模态间采用单一的基于图像的架构。我们将触觉观测定义为 $\mathbf{x} = \langle I_{\text{ref}}, I_{\text{cur}} \rangle$，其中 I_{ref} 是未变形状态下的参考帧（$t = 0$）而 I_{cur} 是时间 $t = T$ 的接触帧，用于捕捉接触引起的弹性体变形。考虑一对同步传感器 $\{s_L, s_R\}$ 抓取一个形态对称的物体。尽管$\mathbf{x}_L$ 和 $\mathbf{x}_R$ 可能因域偏移（例如标记密度和材料属性）而存在显著差异，但准静态平衡会将两个指尖处的交互力约束为近似等大反向。为简单起见，我们将这一约束记为$F_L \approx F_R$ （大小相等），与图 1 一致。我们的目标是学习一个传感器不变编码器 $\mathcal{E}_\theta$ ，将特定于传感器的观测映射到共享潜在力空间。我们将接触力表示为分块潜在力图$\mathbf{z} \in \mathbb{R}^{N_p \times 6}$，其中 N_p 为空间分块的数量，且每个分块对应一个六维潜在力向量。我们将跨传感器对齐表述为匹配根据左侧和右侧观测推断出的编码器后验分布：

$$q_\theta(\mathbf{z} \mid \mathbf{x}_L) \approx q_\theta(\mathbf{z} \mid \mathbf{x}_R) \tag{1}$$

B. 架构：*UniForce* 模型

UniForce 是一种条件变分自编码器（CVAE），能够将接触物理与传感器特定的触觉信号分离（图 2）。编码器近似逆动力学（触觉观测 → 力潜变量），而解码器模型的正向动力学（力潜变量 → 接触后的触觉观测）以传感器的参考观测为条件。

1) 编码器——逆动力学：给定$\mathbf{x} = \langle I_{\text{ref}}, I_{\text{cur}} \rangle \in \mathbb{R}^{2\times 3\times H\times W}$，我们将两帧划分为图像块，得到词元$\mathbf{V}_0 \in \mathbb{R}^{T\times N_p\times D}$，其中 $T = 2$, N_p 表示图像块的数量， D表示嵌入维度。我们注入正弦位置编码，以保留从参考帧到接触帧的时间顺序。为提取形变特征，我们采用因果时空变换器 [18] ，并使用因式分解注意力以提高效率并确保因果性：

- **空间注意力** 通过多头自注意力在每一帧内进行计算，以提取局部标记结构，随后应用层归一化和残差连接。
- **因果时间注意力** 沿时间轴对空间对齐的词元进行计算。因果掩码 $M_t \in \mathbb{R}^{T\times T}$允许接触帧（$t = T$）关注参考帧（$t = 0$），同时阻止参考帧关注当前接触帧。

2）概率力流形：编码器输出$\mathbf{h}_{\text{last}} \in \mathbb{R}^{N_p\times D}$，并将其投影为针对每个块的力图的对角高斯后验分布：

$$[\boldsymbol{\mu}, \log \boldsymbol{\sigma}^2] = \text{Linear}(\text{LayerNorm}(\mathbf{h}_{\text{last}})) \in \mathbb{R}^{N_p\times 12} \tag{2}$$

训练期间，我们使用 $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$采样 $\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}$。推理期间，我们使用 $\mathbf{z} = \boldsymbol{\mu}$。

3）解码器 – 物理条件正向动力学：ˆ 解码器 $\mathcal{D}_\psi$ 以参考图像 I_{ref}为条件，重建接触图像 I_{cur} ；参考图像隐式提供传感器特定外观（例如，标记/触觉单元分布）。我们通过加法注入，将潜在力与参考嵌入融合：

$$\mathbf{V}_{\text{dec}} = \text{PatchEmbed}(I_{\text{ref}}) + \text{Linear}_{\text{proj}}(\mathbf{z}) \tag{3}$$

，随后，空间变换器预测逐像素重建ˆ I_{cur} ，并以图像重建损失作为约束。

C. 跨传感器力平衡训练

我们使用配对抓取（L, R）上的三分支目标函数（图 2）训练 UniForce，无需真实力标签。设 $\mathbf{x}_i = \langle I^i_{\text{ref}}, I^i_{\text{cur}} \rangle$ 表示 $i \in \{L, R\}$, 并设 $\mathbf{z}_i \sim q_\theta(\mathbf{z} \mid \mathbf{x}_i)$。对于每个有序对 $(i, j) \in \{L, R\} \times \{L, R\}$, 我们解码

$$\hat{I}^{i\to j}_{\text{cur}} = \mathcal{D}_\psi(I^j_{\text{ref}}, \mathbf{z}_i) \tag{4}$$

其中， $i = j$ 对应自重建， $i \neq j$对应跨传感器重建。在准静态平衡条件下，ˆ$I^{i\to j}_{\text{cur}}$的监督目标是 I^j_{cur}，这促使$\mathbf{z}_i$保留与力相关的内容，同时舍弃传感器特定外观。

1）目标函数：总损失由重建、KL 正则化和平衡一致性组成：

$$\mathcal{L}_{total} = \mathcal{L}_{recon} + \lambda_{kl}\mathcal{L}_{KL} + \lambda_{eq}\mathcal{L}_{eq} \tag{5}$$

重建损失（$\mathcal{L}_{recon}$）： 我们将 ℓ_1 损失与LPIPS [21]相结合，以在四个分支中捕捉标记点形变：

$$\mathcal{L}_{recon} = \sum_{i,j\in\{L,R\}} \left(\left\| \hat{I}^{i\to j}_{\text{cur}} - I^j_{\text{cur}} \right\|_1 + \lambda_{lpips}\,\text{LPIPS}\left(\hat{I}^{i\to j}_{\text{cur}}, I^j_{\text{cur}} \right) \right) \tag{6}$$

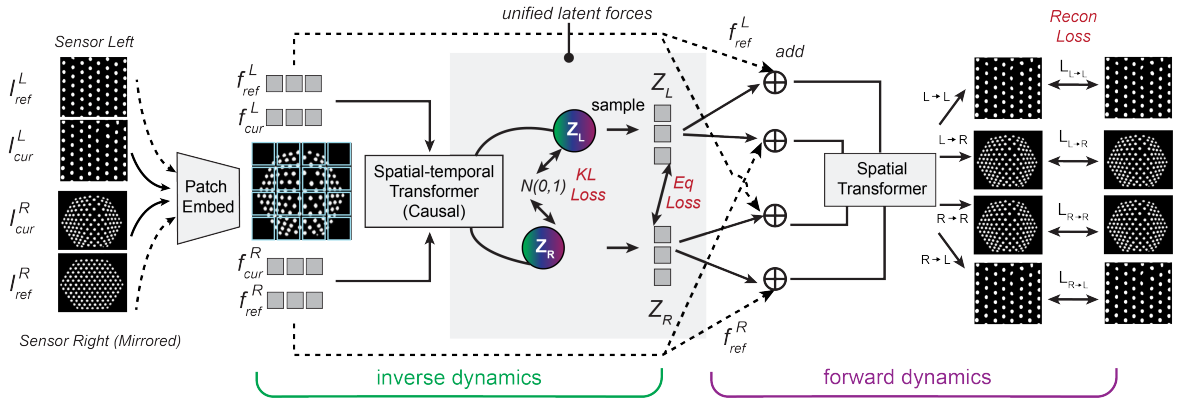

图2：**统一的力感知触觉表征学习**。在**逆动力学阶段**，编码器接收成对的参考/接触观测，并推断分块潜在力图；平衡损失对齐左/右潜在力，而KL项则对后验分布进行正则化。在**正向动力学阶段**，解码器以传感器特定参考图像和潜在力为条件重建观测，从而实现自重建和跨传感器重建。我们将右指标记图像进行镜像，以匹配左手指的坐标。

KL散度（$\mathcal{L}_{KL}$）：我们按照标准变分自编码器的做法，将 $q_\theta(\mathbf{z} \mid \mathbf{x})$正则化至$\mathcal{N}(\mathbf{0}, \mathbf{I})$。

平衡损失（$\mathcal{L}_{eq}$）：我们对左、右潜在力图之间的分块一致性进行约束：

$$\mathcal{L}_{eq} = \frac{1}{N_p} \|\mathbf{z}_L - \mathbf{z}_R\|_2^2 \tag{7}$$

IV. 力平衡数据采集

A. 设置

1）硬件：如图3所示，我们的设置包括一台配备Franka Hand的Franka FR3机械臂、对称压头，以及安装在两根手指上的触觉传感器。

我们使用三种触觉传感器：

- **GelSight 2017（基于视觉）[20]**。我们使用 Logitech C270 网络摄像头（720p，30fps），在内部制作了一款 GelSight 传感器：摄像头装在带有 RGB LED 的 3D 打印外壳内。硅橡胶弹性体上印有方形标记图案（80 个标记点）。感测区域为 $25 \times 25 \times 4$mm，最大压入深度为 1.5 mm。
- **TacTip（基于视觉）[22]**。TacTip 配有一台鱼眼相机（ELP USB 相机，1080p，30fps），安装在带有白色 LED 的 3D 打印外壳内。外层皮肤采用 Ag- ilus30，背衬为一种异常柔软的透明凝胶。该传感器含有 127 个内部标记，曲率半径为 41.5 mm，直径为 40 mm，压入深度最高可达 5 mm。
- **uSkin（基于磁学）[15]**。我们使用 XELA Robotics 的 uSkin（贴片版本），配有 4×4 触觉单元阵列（16 个触觉单元）。每个触觉单元均可提供三轴形变信号。传感器尺寸为 $24.6 \times 22.6 \times 5$mm，并以 500 Hz 输出多通道信号。

我们设计了一套具有八种形状的对称双面压头（例如正方形、三角形、曲面和球形基本体）。压头尺寸经过设计，确保其位于每个传感器的有效感测区域内。压头夹具安装在底座上，底座则刚性固定于金属光学平台。

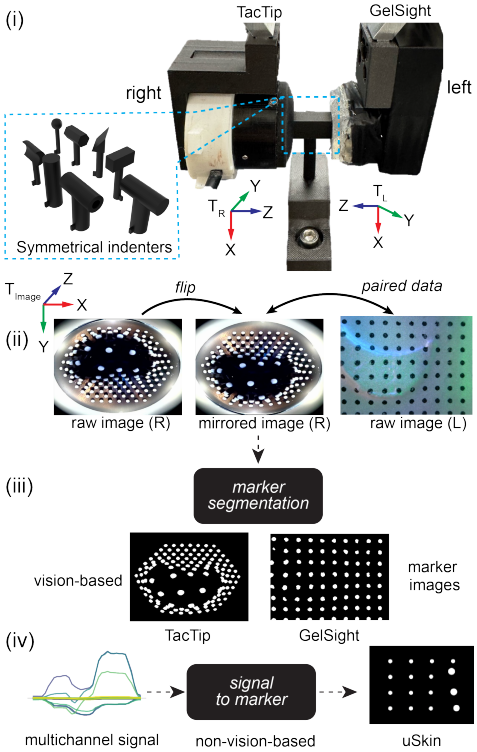

图3：**通过准静态力平衡采集配对数据**。(i) 数据采集设置；(ii) 通过镜像进行坐标对齐。(iii–iv) 参照 [4] 对异构原始信号进行规范化。

2）数据采集协议：我们将 TacTip 作为固定在右手指上的锚点传感器，同时将左手指传感器在 GelSight 和 uSkin 之间切换。由此得到两个配对数据集：⟨GelSight，TacTip⟩和 ⟨uSkin，TacTip⟩，我们将其称为*UniForce-pair* 数据集。

机器人采用笛卡尔空间阻抗控制。我们沿夹爪闭合方向设置较高的刚度，以减少接触期间意外的侧向运动，并从

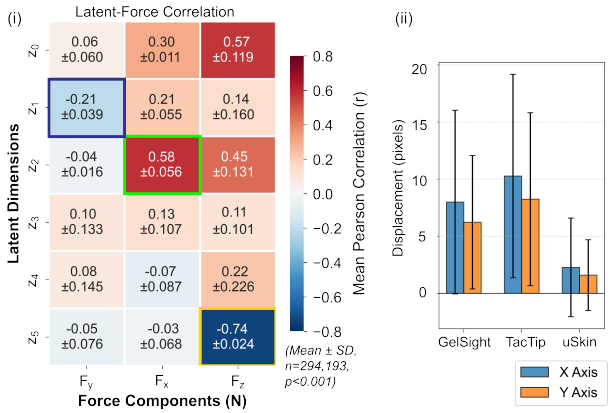

图4：（i）潜在维度 z_0–z_5 与力分量F_y、F_x、F_z（N）之间皮尔逊相关系数均值 r（±标准差）的热力图。三个传感器共计图像 $n = 294$，193张；（ii）*UniForce-pair* 数据集中标记点在x轴和y轴上的平均绝对位移。

轻柔的初始抓取开始，此时两个传感器均观测到轻微的压入。随后，人工操作员缓慢拖动或旋转机器人手，同时逐渐增加压入深度，以引发切向变形。这些操作均以低速进行，以近似准静态状态，在此状态下，预计两个指尖之间满足力平衡。对于每个压头，我们记录 1,000 帧，每个传感器对约产生 8,000 帧数据。每个传感器对的数据采集时间不到 30 分钟。

3）数据预处理：**1）通过镜像进行坐标对齐**。以图像坐标系作为参考坐标系时，左、右指尖的y轴呈镜像关系（图3）。为对齐坐标系，在构建力平衡配对之前，我们将右手指的触觉图像进行水平翻转，从而确保不同手指所表示的接触几何形状和力的方向保持一致。**2）异构原始信号的规范化**。原始RGB触觉图像可能存在光照/背景变化，而uSkin输出多通道时间序列信号。遵循GenForce [4], ，我们通过图像分割（针对视觉式传感器）或信号到标记点的渲染过程（针对uSkin），将视觉式触觉图像和非视觉触觉信号转换为统一的二值标记图像表示。在我们的实现中，这一预处理过程的运行频率超过29.6 Hz。所有标记图像均调整为 640×480，用于训练。

V. 实验与结果

我们通过回答以下四个问题来评估 UniForce：**Q1** 学习到的潜在空间是否在逆动力学编码器中编码接触力？**Q2** 解码器能否利用潜在力重建异构传感器上的形变触觉观测（正向动力学）？**Q3** 与先前的触觉表征学习方法相比，UniForce在零样本力估计方面表现如何？

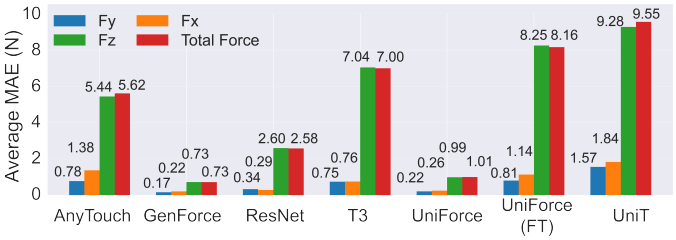

图 5：**零样本力预测误差的平均值（平均绝对误差）**。我们比较了六种异构传感器对的平均绝对误差。FT 表示全量训练。

Q4UniForce能否集成到技能学习（例如视觉-语言-动作模型）中，以借助异构触觉感知提升力感知操作?

A. 学习到的潜在力空间（*Q1*）

为回答 **Q1**，我们冻结 UniForce 编码器，并分析推断出的潜在维度（z_0–z_5，其中 $\mathbf{z} \in \mathbb{R}^6$）是否与真实力标签相关。

数据集。我们使用 **GenForce-Hetero** 数据集 [4], ，其中包含与力标签配对的触觉序列，涵盖 TacTip（108,153 张图像）、GelSight（106,278 张图像）和 uSkin（79,762 张图像）。我们使用 18 个压头施加作用力，法向力范围为 $[0, 12]$ N，剪切力范围为 $[-3, 3]$ N。请注意，GenForce-Hetero 中的 GelSight 使用类菱形标记图案（GelSight-D），与我们的自研 GelSight 传感器不同；TacTip 和 uSkin 则与我们的硬件一致。

指标。我们计算每个潜在维度 z_i 与各个力分量（F_x、F_y、F_z）之间的皮尔逊相关系数，并报告均值 ± 标准差（所有 $p < 0.001$）。我们通过对 $N_p = 196$ 个空间分块取平均得到 z_i 。我们总共分析了三个传感器的$n = 294$、193 张图像。

法向力呈现为稳定的潜在因子。图4i显示，F_z 在不同传感器上均与单一潜在维度 z_5 密切相关（$r = -0.74 \pm 0.024$），对应于仅根据 z_5 得出的线性可预测性 $R^2 = 0.552$（TacTip：$r = -0.768$、$R^2 = 0.590$; GelSight：$r = -0.748$、$R^2 = 0.559$; uSkin：$r = -0.711$、$R^2 = 0.506$）。**剪切力得到中等程度的编码**。对于横向剪切力，F_x 与 z_2 的相关性最高（$r = 0.58 \pm 0.056$、$R^2 = 0.342$）。相比之下，F_y 的编码较弱（均值 $|r| \approx 0.21$、$R^2 = 0.047$），且相关性最高的维度因传感器而异。我们认为这主要归因于数据集构成：形变主要由抓取引起的法向力（F_z）以及手部重量导致的向下切向分量（F_x）造成，而 F_y 受到的激励较弱（图4ii），因此更难从触觉观测中推断出来。总体而言，这些相关性表明，UniForce 通过自监督训练恢复了具有物理意义的潜在因子。

B. 跨传感器图像重建（*Q2*）

UniForce 经过训练，能够根据参考观测 I_{ref} 和

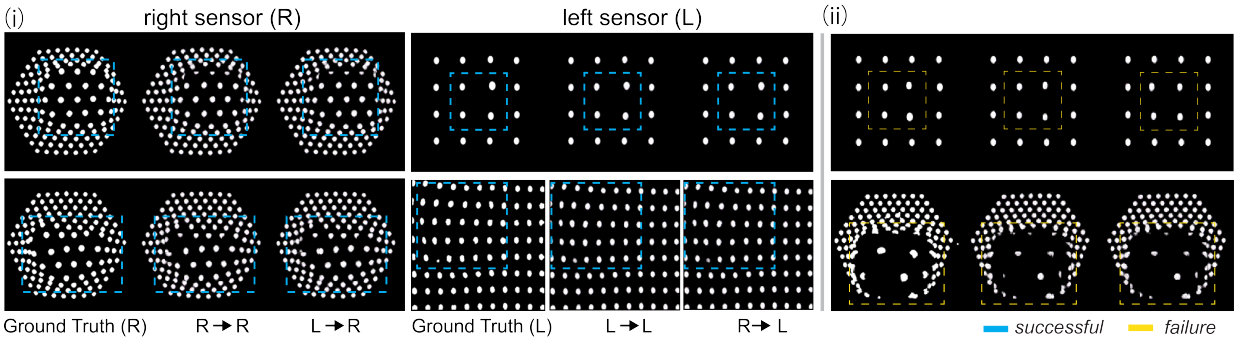

图 6：**跨传感器触觉图像重建（GelSight–TacTip 和 uSkin–TacTip）**。右手指上的 TacTip 用作锚定传感器。推断出的潜在力通过前向动力学解码器驱动四个重建分支（自重建和跨传感器重建）。一个 UniForce 模型在全部三种传感器的数据上联合训练。

表 I：使用 R^2 对**未见过的**物体进行零样本力预测的比较。缩写：**G**：GelSight，**T**：TacTip，**U**：uSkin。G→T 表示从 G 到 T 的迁移。数值越高越好（↑）。

模型	G → T			G → U			T → G			T → U			U → G			U → T		
	F_x ↑	F_y ↑	F_z ↑	F_x ↑	F_y ↑	F_z ↑	F_x ↑	F_y ↑	F_z ↑	F_x ↑	F_y ↑	F_z ↑	F_x ↑	F_y ↑	F_z ↑	F_x ↑	F_y ↑	F_z ↑
ResNet	-0.08	-1.66	0.01	-0.01	-0.71	0.04	0.02	-0.22	-0.26	0.0	-0.01	-0.09	0.03	-0.14	-2.42	-0.25	-0.12	-9.34
AnyTouch	-4.11	-0.31	-5.41	-1.97	-0.48	-12.23	-0.94	0.23	-0.07	-0.91	-3.91	-6.19	-38.84	-13.3	0.23	-43.04	-34.82	-44.27
T3	-0.2	-0.23	-3.49	-0.03	-0.2	0.17	0.3	0.2	0.1	-0.21	-0.38	-0.5	-15.93	-46.56	-59.85	-22.86	-8.41	-130.17
UniT	-1.47	-1.18	-33.74	-12.53	-0.64	-0.06	0.13	0.13	0.31	-0.53	-0.5	-6.11	-57.73	-122.61	-79.34	-105.75	-123.13	-157.72
UniForce（全量训练）	-0.48	-0.99	-14.52	-1.43	-0.04	-4.65	-0.49	-2.6	-0.85	-5.72	-0.77	-1.69	-5.9	-21.68	-37.22	-61.84	-33.62	-155.04
UniForce	**0.1**	**0.17**	**0.59**	**0.4**	**0.29**	**0.47**	**0.77**	**0.62**	**0.83**	**0.35**	**0.3**	**0.39**	**-0.23**	**-0.51**	**0.17**	**-2.36**	**-0.55**	**-1.51**
GenForce*	0.38	0.3	0.63	0.59	0.69	0.79	0.48	0.49	0.67	0.26	0.41	0.55	0.58	0.62	0.72	0.3	0.34	0.61

*GenForce transfers and tr ains 力预测模型 per 目标传感器（with additional per-sensor processing），and is 非零样本方法 .

推断出的潜在力来重建接触后的触觉观测，充当前向动力学的替代模型。一项独特能力是跨传感器重建：当抓取处于准静态平衡状态时，利用从一个传感器推断出的潜在力，重建另一个传感器的接触观测。

如图 6 所示，UniForce 能够重建传感器内和跨传感器的触觉响应。例如，从 TacTip 推断出的潜在力可以驱动 GelSight 和 uSkin 的重建（以各自的参考观测为条件），而从 GelSight/uSkin 推断出的潜在力则可以驱动 TacTip 的重建。

尽管相同的接触力会因接触位置和压头几何形状不同而产生不同的标记点位移，但借助逐图像块的潜在设计（例如 16×16 图像块），重建结果仍能保留与真实值一致的接触位置和几何形状。我们也观察到一些失败案例。例如，uSkin 的跨传感器重建可能会出现不一致的标记点位移；而以 GelSight 为条件的 TacTip 重建在大变形情况下可能不完整。这些案例凸显了在信号稀疏或变形范围较大的情况下，跨传感器准确呈现细粒度接触模式的难度。尽管如此，成功的重建表明，解码器已学会一种与预期前向模型一致的参考条件映射：$I_{\text{cur}} \approx \mathcal{D}(I_{\text{ref}}, \mathbf{z})$。

C. 零样本力估计（*Q3*）

为回答 **Q3**，我们冻结预训练 UniForce 编码器，并连接一个可训练的力预测头。我们使用 **GenForce-Hetero** 数据集中的一种源传感器训练该预测头 [4]（GelSight-D (G)、TacTip (T) 或 uSkin (U)），随后在其他传感器上进行评估，不进行任何额外微调。在训练阶段，我们使用数据集中的已见物体（12 种）；在测试阶段，则使用未见物体（6 种）。

基线方法。我们将预训练的触觉表征模型（在可用时使用官方检查点冻结）与端到端训练的标准骨干网络进行比较。为确保公平比较，我们为每个冻结的编码器接入相同的力预测头（ConvGRU+MLP），并根据触觉序列预测力，这与先前研究一致；先前研究表明，相较于单帧预测，这种方法效果更好 [3]。

- **AnyTouch [6]:** 面向 GelSight 系列的表征学习模型，使用静态和动态数据进行训练。
- **T3 [23]:** 采用共享主干网络，并为 GelSight 系列配备传感器专用编码器和任务专用解码器。
- **UniT [17]:** 采用 VQ-VAE 风格的触觉表征，使用带标记点的 GelSight Mini 进行训练。
- **ResNet [8]:** 用于力预测的常见端到端骨干网络（完整训练）。

- **UniForce（全量训练）：** 完全训练的力模型，训练时不冻结 UniForce 编码器。

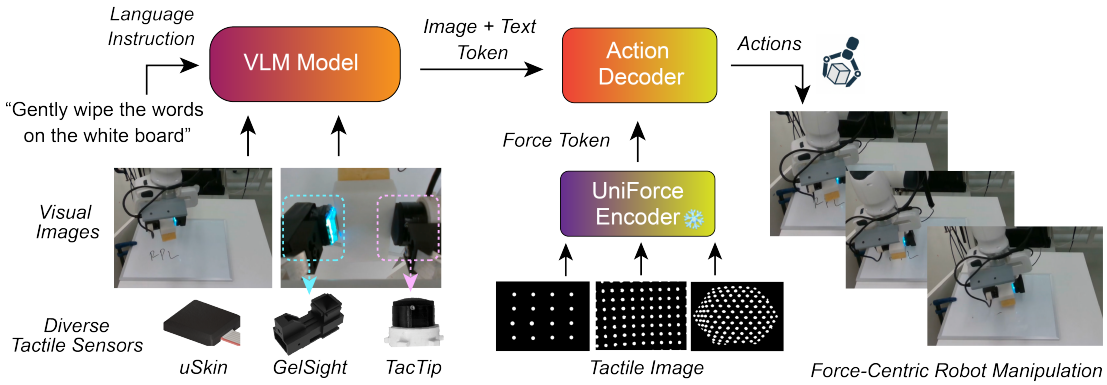

图 7：**将 UniForce 接入视觉-语言-动作模型**。通过 UniForce 将异构触觉输入统一为力感知词元，用于力感知机器人操作任务，即白板擦拭。

- **GenForce [4]:** 通过逐传感器转换和逐传感器训练实现可迁移的力感知；我们将其作为**非零样本参考**进行报告。

指标。我们报告预测力与真实力之间的决定系数（R^2）和平均绝对误差（平均绝对误差）。请注意，负的 R^2 表示其表现比预测平均力更差。

表 I 和图 5 显示，在所有传感器配对设置中，UniForce 的零样本迁移效果均优于基线方法。即使与非零样本方法相比，在从 TacTip 迁移到 GelSight（T→G）时，UniForce 在（F_x，F_y，F_z）上达到 $R^2 = \{0.77, 0.62, 0.83\}$，表现优于 GenForce 的 $\{0.48, 0.49, 0.67\}$。此外，在六种迁移设置中的平均 MAE 接近 GenForce（UniForce：$F_x/F_y/F_z$ 时分别为 0.22 N/0.26 N/0.99 N；GenForce：0.17 N/0.22 N/0.73 N）。这可能归因于 UniForce 在力感知潜在空间中进行端到端训练；与 GenForce 由图像到图像转换、力预测模型训练和材料补偿组成的三阶段流程相比，这种方式减少了信息损失和图像转换噪声。我们还观察到，以 uSkin 为源传感器时（U→G 和 U→T），迁移效果较弱；这与从低分辨率触觉信号迁移到更密集的标记图像时存在的信息瓶颈相符，GenForce 也有类似情况。尽管如此，与基线方法相比，力误差和 R^2 值显著改善，验证了 UniForce 在异构传感器间进行力估计零样本迁移的有效性。

D. 零样本力感知策略迁移（*Q4*）

为回答**Q4**，我们将预训练的 UniForce 编码器集成到 $\pi_{0.5}$ [10]（一种视觉-语言-动作模型）中，用于力感知技能学习，即白板擦拭（图 7）。UniForce 编码器将来自异构传感器的基于标记点的触觉图像映射为以力为基础的触觉词元，并将其与视觉和语言词元拼接，用于动作解码。整个过程中，UniForce 编码器始终保持冻结。

数据。我们将 TacTip 固定在右手指上，并在 GelSight 和 uSkin 之间更换左手指传感器（图 8）。我们收集了 50 条包含不同词语和初始条件的任务轨迹。**对比组**。

- **V:**仅视觉（$\pi_{0.5}$ 不含触觉词元）。
- **V+T:** 视觉 + TacTip。（$\pi_{0.5}$ 含触觉词元）
- **V+G:** 视觉 + GelSight（从 V+T 进行零样本迁移）。
- **V+U:** 视觉 +uSkin（从 V+T 进行零样本迁移）。
- **V+T+G:** 视觉 + TacTip + GelSight（使用两种传感器对策略进行微调；UniForce 冻结）。
- **V+T+U:** 视觉 +TacTip + uSkin（使用两种传感器对策略进行微调；UniForce 冻结）。

指标。我们报告每种设置下 20 个回合的成功率（SR）。若一个回合完成所有任务步骤，且未发生抓取失败，并且残留书写痕迹低于阈值（残留面积 $\leq 10\%$），则视为成功。

表 II 显示，与仅使用视觉（V：20%）相比，加入触觉传感可提高 SR。UniForce 还能将使用 TacTip 训练的策略（50%）迁移至 GelSight（45%）和 uSkin（60%），而无需针对每种传感器分别进行触觉编码器微调。使用来自两根手指的触觉信号（V+T+G：75%，V+T+U：80%）比使用单个触觉传感器获得更高的 SR，这表明 UniForce 能够利用异构触觉输入进行力感知技能学习。一种可能的解释是，双侧触觉反馈提供了互补的接触信息（例如，两侧的法向和剪切线索），有助于在擦拭过程中保持稳定接触，并减少表 II 所示由滑移或过大力导致的失败。我们还在图 9 中绘制了最短成功回合的末端执行器笛卡尔 z位置。使用 UniForce 和触觉传感时，擦拭阶段的曲线更加平坦，表明垂直接触调节更加稳定；这与不当作用力（例如滑移引起的高度变化或过度按压）发生次数更少的情况相符。

VI. 局限性与讨论

A. 准静态平衡假设

UniForce 在收集配对数据时依赖准静态力平衡先验。在实践中，快速运动、滑移、夹爪机构的柔顺性，或两根指尖与压头之间不完美的对齐/平行度，都可能违反这一假设。

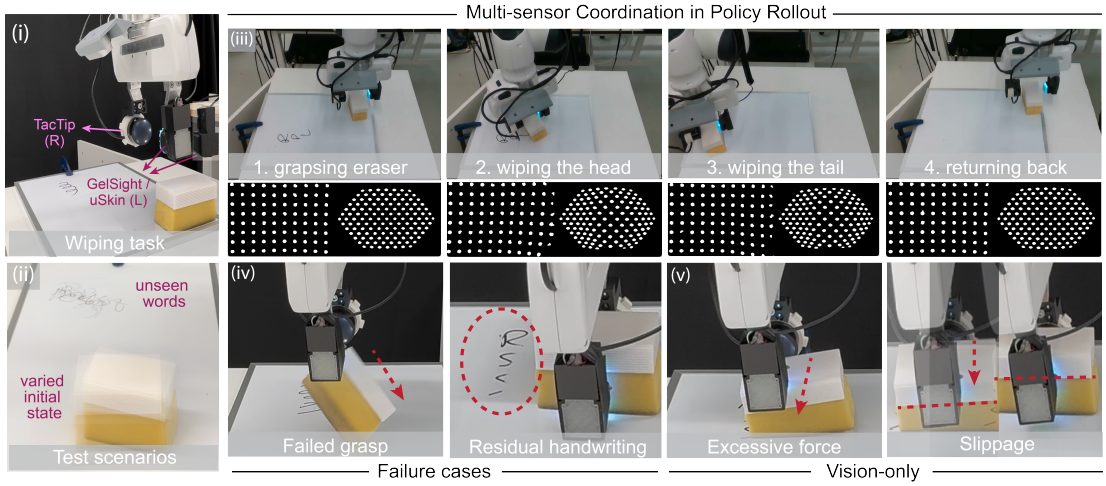

图 8：**使用异构触觉传感器进行白板擦拭。**（i）擦拭任务设置。（ii）不同的测试场景。（iii）多传感器协同下的策略执行。（iv）失败案例。（v）不当作用力会导致滑移和过度按压。

表 II：使用不同输入模态进行擦拭时的策略性能。

模态	成功 Rate	失败 Case		不当力	
		失败抓取	残差笔迹	滑移	超额力
V	4/20（20%）	12	4	3	3
V+T	10/20（50%）	8	2	2	1
V+G	9/20（45%）	9	2	2	1
V+U	12/20（60%）	7	1	1	/
V+T+G	15/20（75%）	3	2	1	/
V+T+U	16/20（80%）	2	2	1	/

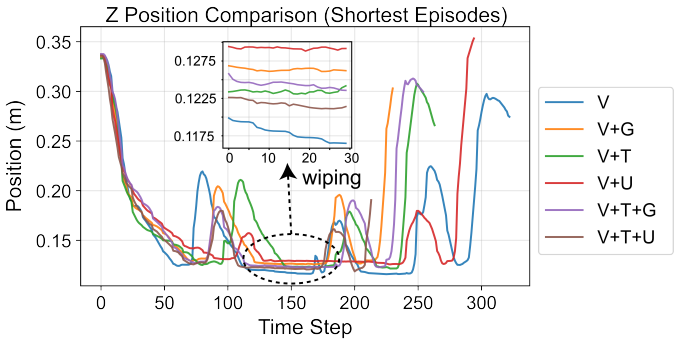

图 9：**擦拭过程中机器人手的 Z 轴位置。**

当平衡被破坏时，左右两侧的交互可能不再对应同一个接触力。一个切实可行的方向是通过闭环控制自动化数据收集，以强制实现准静态接触（例如限制速度/加速度），并主动纠正错位，从而提高力配对监督的质量。

B. 材料建模与长期动力学

UniForce 根据参考观测 I_{ref} 对解码器进行条件化，并注入潜在力以重建 I_{cur}，但并未显式建模材料属性。软弹性体可能表现出非线性且与历史状态相关的行为（例如滞后、黏弹性和延迟恢复），而标记图案相似但表面硬度不同的传感器，在相同载荷下可能会产生不同响应。一个很有前景的方向是将显式的材料/状态变量（例如硬度参数）纳入前向模型，从而提升模型对表面差异和长期材料漂移的鲁棒性。

C. 对未见传感模态的泛化

UniForce 专为原始信号可转换为标记点/触觉单元图像表示的触觉传感器而设计。如何泛化到差异显著的模态（例如基于触须的触觉感知）仍有待研究，尤其是在信号几何结构无法自然地进行二维标记图像规范化时。未来工作可能需要采用双指力平衡之外更广泛的物理约束。

D. 大型模型中的触觉集成

在 VTLA 实验中，我们将 UniForce 触觉词元与视觉与语言词元拼接，并输入动作解码器。这种简单的融合方式可能无法充分利用触觉信息进行高层推理。未来工作可探索结构化融合，例如将触觉线索集成到视觉语言模型的推理流程中，同时向低层动作模块提供以力为依据的词元，从而更好地利用不同模态之间的互补信息。

VII. 结论

我们提出了**UniForce**，这是一个面向采用异构触觉传感器的机器人操作的统一潜在力表征学习框架。UniForce利用力平衡数据学习共享的潜在力空间，而无需在表征学习过程中使用真实力标签。所学习的潜在力向量支持在不同触觉传感器之间对下游力估计和力感知策略学习进行零样本迁移。该方法朝着可扩展、以力为基础的触觉表征迈出了一步，可减少逐传感器再训练，并促进其在不同本体上的部署。

参考文献

[1]Jianxin Bi、Kevin Yuchen Ma、Ce Hao、Mike Zheng Shou 和 Harold Soh。Vla-touch：利用双层触觉反馈增强视觉-语言-动作模型。*arXiv* 预印本 *arXiv:2507.17294*，2025年。[2]Zhuo Chen、Yang Zhang、Bin Zhu、Yigen Wu、Xiao-hui Du、Liwei Lin 和 Dezhi Wu。用于宽工作范围超高灵敏度离子电子传感器的激光雕刻分层棘状结构。*ACSApplied Materials &Interfaces*，14(17)：19672–19682，2022年。[3]Zhuo Chen、Ni Ou、Xuyang Zhang 和 Shan Luo。Transforce：通过序列图像转换实现面向视觉触觉传感器的可迁移力预测。载于*2025 IEEE*机器人与自动化国际会议（*ICRA*），第237–243页。IEEE，2025年。[4]Zhuo Chen、Ni Ou、Xuyang Zhang、Zhiyuan Wu、Yongqiang Zhao、Yupeng Wang、Emmanouil Spyrakos Papastavridis、Nathan Lepora、Lorenzo Jamone、Jiankang Deng 等。训练触觉传感器相互学习力感知。*NatureCommunications*，2026年。[5] Zhengxue Cheng、Yiqian Zhang、Wenkang Zhang、Haoyu Li、Keyu Wang、Li Song 和 Hengdi Zhang。OmniVTLA：具有语义对齐触觉感知的视觉-触觉-语言-动作模型。*arXiv* 预印本 *arXiv:2508.08706*，2025年。[6]Ruoxuan Feng、Jiangyu Hu、Wenke Xia、Tianci Gao、Ao Shen、Yuhao Sun、Bin Fang 和 Di Hu。Any-touch：跨多种视觉触觉传感器学习统一静态-动态表征。*arXiv* 预印本 *arXiv:2502.12191*，2025年。[7] Francesco Grella、Alessandro Albini、Giorgio Cannata 和 Perla Maiolino。触觉到触觉转换——学习异构触觉感知技术之间的映射。*arXiv* 预印本 *arXiv:2411.02187*，2024。[8] Kaiming He、Xiangyu Zhang、Shaoqing Ren 和 Jian Sun。用于图像识别的深度残差学习。载于*IEEE*计算机视觉与模式识别会议论文集，第770–778页，2016年。[9] Jialei Huang、Shuo Wang、Fanqi Lin、Yihang Hu、Chuan Wen 和 Yang Gao。Tactile-VLA：释放视觉-语言-动作模型的物理知识，实现触觉泛化。*arXiv* 预印本 *arXiv:2507.09160*，2025年。[10] Physical Intelligence、Kevin Black、Noah Brown、James Darpinian、Karan Dhabalia、Danny Driess、Adnan Esmail、Michael Equi、Chelsea Finn、Niccolo Fusai、Manuel Y. Galliker、Dibya Ghosh、Lachy Groom、Karol Hausman、Brian Ichter、Szymon Jakubczak、Tim Jones、Liyiming Ke、Devin LeBlanc、Sergey Levine、Adrian Li-Bell、Mohith Mothukuri、Suraj Nair、Karl Pertsch、Allen Z. Ren、Lucy Xiaoyang Shi、Laura Smith、Jost Tobias Springenberg、Kyle Stachowicz、James Tanner、Quan Vuong、Homer Walke、Anna Walling、Haohuan Wang、Lili Yu 和 Ury Zhilinsky。$\pi_{0.5}$：具有开放世界泛化能力的视觉-语言-动作模型，2025年。URL https://arxiv.org/abs/2504.16054.

[11] Roland S Johansson 和 J Randall Flanagan。物体操作任务中指尖触觉信号的编码与使用。《自然·神经科学评论》，10（5）：345–359，2009年。[12]Zhi Liu、Xiaonan Hu、Renheng Bo、Youzhou Yang、Xu Cheng、Wenbo Pang、Qing Liu、Yuejiao Wang、Shuheng Wang、Shiwei Xu 等。模仿人类机械感觉的三维架构电子皮肤。《科学》，384（6699）：987–994，2024。[13]Shan Luo、Nathan F Lepora、Wenzhen Yuan、Kaspar Althoefer、Gordon Cheng 和 Ravinder Dahiya。触觉机器人学：展望。《*IEEE*机器人学汇刊》，2025年。[14] Samanta Rodriguez、Yiming Dou、Miquel Oller、Andrew Owens 和 Nima Fazeli。Touch2touch：用于物体操作的跨模态触觉生成。*arXiv*预印本 *arXiv:2409.08269*，2024。[15]Tito Pradhono Tomo、Massimo Regoli、Alexander Schmitz、Lorenzo Natale、Harris Kristanto、Sophon Somlor、Lorenzo Jamone、Giorgio Metta 和 Shigeki Sugano。用于uSkin——一种柔性、分布式、数字式三轴皮肤传感器的新型硅胶结构及其在人形机器人 iCub 上的集成。《*IEEE*机器人与自动化快报》，3（3）：2584–2591，2018年。[16] Benjamin Ward-Cherrier、Nicholas Pestell、Nathan F 等。TacTip系列：具有3D打印仿生形态的柔性光学触觉传感器。《软体机器人》，5（2）：216–227，2018年。[17] Zhengtong Xu、Raghava Uppuluri、Xinwei Zhang、Cael Fitch、Philip Glen Crandall、Wan Shou、Dongyi Wang 和 Yu She。UniT：具有未见物体泛化能力的数据高效触觉表征。《*IEEE*机器人与自动化快报》，2025年。[18] Seonghyeon Ye、Joel Jang、Byeongguk Jeon、Sejune Joo、Jianwei Yang、Baolin Peng、Ajay Mandlekar、Reuben Tan、Yu-Wei Chao、Bill Yuchen Lin 等。从视频中进行潜在动作预训练。*arXiv*预印本 *arXiv:2410.11758*，2024。[19] Jiawen Yu、Hairuo Liu、Qiaojun Yu、Jieji Ren、Ce Hao、Haitong Ding、Guangyu Huang、Guofan Huang、Yan Song、Panpan Cai 等。ForceVLA：通过力感知混合专家模型增强视觉-语言-动作模型，以实现富接触操作。*arXiv*预印本 *arXiv:2505.22159*，2025年。[20] Wenzhen Yuan、Siyuan Dong 和 Edward H Adelson。GelSight：用于估计几何形状和力的高分辨率机器人触觉传感器。传感器，17（12）：2762，2017年。[21] Richard Zhang、Phillip Isola、Alexei A Efros、Eli Shechtman 和 Oliver Wang。深层特征作为感知度量的非凡有效性。载于*IEEE*计算机视觉与模式识别会议论文集，第586–595页，2018年。[22]Xuyang Zhang、Tianqi Yang、Dandan Zhang 和 Nathan F Lepora。TacPalm：具有仿生光学触觉手掌的柔性夹爪，用于稳定、精确的抓取

。*IEEE*传感器期刊，2024。[23] Jialiang Zhao、Yuxiang Ma、Lirui Wang 和 Edward H Adelson。可迁移触觉变换器：跨多样传感器和任务的表征学习。*arXiv*预印本 *arXiv:2406.13640*，2024。