基于拓扑感知蒸馏与渐进式运动学推理的包容性手部姿态估计

Kanglei Zhou、Zikai Hao、Liyuan Wang、和 XiaohuiLiang、会员，*IEEE*

arXiv:2611.29163v1 [cs.CV] 22 Nov 2026

摘要——**手部姿态估计旨在根据视觉观测恢复手部关节的空间配置，是下游系统的重要感知模块，这些系统依赖完整且拓扑一致的手部表示。现有大多数方法都是针对完全铰接的手部结构开发的；当运动学结构的部分区域持续缺失时，它们往往会产生不稳定或不合理的预测，导致性能下降，甚至系统失效，尽管仍可获得部分观测。为系统地研究这一挑战，我们构建了 IncNYU，这是一个生理约束基准，通过对完整手部模型进行形态学修改来模拟结构不完整性，并保留完整区域和合成移除区域的参考位置。基于该基准，我们提出 InsHaPE，一种包容性手部姿态估计框架，旨在结构不完整时保持鲁棒性，同时维持在标准设置下对全关节手部的性能。InsHaPE 基于层次化运动学建模范式，并引入两个互补组件。首先，拓扑感知蒸馏策略从完全铰接手部表示中迁移关节级和关系先验，从而在结构缺失时提供符合解剖学一致性的初始化。其次，渐进式运动学推理模块沿运动学链执行结构化推理，使系统能够在严重结构变化下实现完整且符合解剖学一致性的姿态估计。在 NYU/IncNYU 上开展的大量实验以及在 MSRA 上进行的外部验证表明，InsHaPE 提升了结构不完整情况下的鲁棒性，同时保持了对完整手部的竞争力。此外，系统级评估进一步表明，InsHaPE 提升了需要固定拓扑手部姿态表示的下游应用的鲁棒性。我们的代码可在 https://github.com/tomhao555/InsHaPE。**

索引词—**手部姿态估计、混合现实、包容性交互，人机交互。**

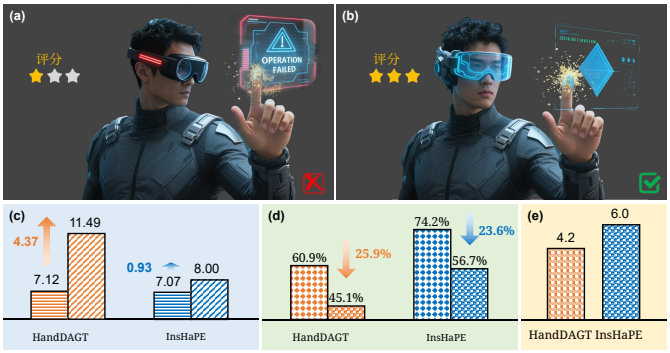


图 1：代表性基线方法 [9] 与所提出的 InsHaPE 框架在标准设置和结构不完整设置下的系统级比较。(a) 当手部结构的部分区域缺失时，基线方法会产生错误或不合理的关节预测。(b) 相比之下，在相同条件下，InsHaPE 仍能实现拓扑一致的完整姿态估计。(c) 在 NYU 和 IncNYU 上对姿态估计准确度（MPJPE）的定量比较。基线方法的性能显著下降（MPJPE 从 7.12 增至 11.49），而 InsHaPE 仅略有下降（从 7.07 增至 8.00）。(d) 完整手部与模拟的不完整观测条件下的手势识别准确率。InsHaPE 在两种设置下均优于基线方法（74.2% 对 60.9%，以及 56.7% 对 45.1%）；箭头表示准确率从完整设置到不完整设置的相对降幅。(e) 参与者在受控的系统级评估中报告的主观可用性评分，反映出 InsHaPE 在下游应用中的可靠性和兼容性有所提升

I. 引言

© 2026 IEEE。允许个人使用本材料。对于任何其他用途，包括在任何当前或未来媒体中重印或再版本材料以用于广告或宣传目的、创作新的合集作品、转售或重新分发至服务器或列表，或在其他作品中重用本作品的任何受版权保护的部分，均须事先获得 IEEE 的许可。本研究得到国家自然科学基金委员会项目 62272019、中国博士后科学基金项目 2025M781489，以及中国博士后科学基金博士后资助项目（C档）项目 GZC20260911 的资助。（通讯作者：*Xiaohui Liang*。）

K. Zhou 和 L. Wang 就职于中国北京清华大学心理与认知科学系。电子邮箱：zhoukan- glei@tsinghua.edu.cn，liyuanwang@tsinghua.edu.cn

Z. Hao 任职于中国北京的北京航空航天大学虚拟现实技术与系统国家重点实验室。电子邮箱：zy2306413@buaa.edu.cn

X. Liang 任职于中国北京的北京航空航天大学虚拟现实技术与系统国家重点实验室，同时也任职于中国北京的中关村实验室。电子邮箱：liang xiaohui@buaa.edu.cn

HAND 手部姿态估计 [1], [2], [3], [4] 旨在根据传感器数据恢复手部关节的二维/三维空间构型，是许多混合现实（MR）系统中的核心感知模块，可支持基于手势的交互、虚拟对象操作和沉浸式用户界面等应用 [5], [6]。在这些系统中，手部姿态估计充当中间表示，连接低层视觉感知与高层交互逻辑 [7]。因此，对于依赖完整且拓扑一致的手部表示的下游组件而言，准确且稳定的关节预测至关重要 [8]。

尽管近年来取得了进展，但大多数现有的手部姿态估计方法 [9], [10], [11], [12] 都是基于完全铰接的手部结构这一隐含假设开发的。永久性解剖结构缺失会使这一假设不再成立；持续的感知故障可能导致特定关节缺乏可靠证据。在这种结构不完整性下，现有方法往往会产生不稳定或不合理的预测，如图 1(c) 中显著增加的误差所示。这些

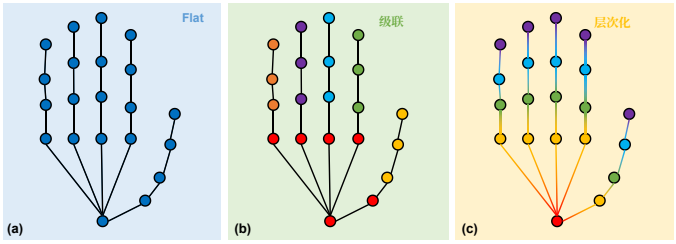


图2：三种手部姿态建模范式的比较。(a) 扁平逐关节建模，所有关节共享单一参数化，并分别进行回归。(b) 级联（逐手指）建模，在手掌和手指之间执行部件级级联预测，但缺乏明确的手指内运动学层级。(c) 我们提出的分层运动学建模，沿根节点到指尖逐步进行推理，并明确编码运动学依赖关系，从而在结构不完整的情况下实现稳健的姿态估计。

误差还会进一步传播到依赖固定拓扑姿态表示的下游混合现实系统和第三方应用，导致性能下降 [13], [14], [15], [16],，这反映在图 1(d) 所示的手势识别准确率下降上。更重要的是，由于标准设置的交互 API（例如 OpenXR）假定输入的手部姿态完整且一致，因此结构不完整情况下的不可靠估计会直接损害其与现有处理流程的兼容性。因此，在姿态估计阶段解决结构不完整性，有助于实现稳健的下游交互，而无需更改固定拓扑的应用接口。

这些失败并非偶然，而是源于现有数据、建模假设和推理策略的根本局限。首先，大多数现有基准数据集 [17] 均采集自全关节手部，并隐含地要求完整的运动学监督，导致训练后的模型倾向于预测完整的关节集合，且在学习过程中未接触过结构缺失的情况。其次，尽管已有若干方法 [18], [19], [20], [21], [22]被提出用于应对遮挡，但它们通常将观测缺失视为暂时性或噪声问题，并依赖时间平滑、空间上下文或特征幻觉。当结构持续缺失时，这些假设并不成立，因为运动学元素被永久移除，既不存在有效观测，也没有时间线索可用于可靠恢复。第三，尽管现有方法 [9], [23], [24], [25] 已探索级联细化或拓扑感知特征建模，但通常只在特征聚合阶段利用手部拓扑，而未沿解剖学运动链对最终预测过程施加明确约束，如图2(a)和2(b)所示。在结构不完整的情况下，微弱的远端关节证据会使关节预测和无约束细化变得不稳定，因此需要引入预测层面的运动学约束。

为系统研究结构不完整情况下的手部姿态估计，我们考虑手部运动学结构的部分区域持续缺失的情形。这一问题在现有基准数据集中基本未受关注。为此，我们通过扩展广泛使用的 NYU 数据集 [26]，构建了一个新颖的基准。我们采用符合生理学约束的形态学处理，生成多种结构缺失形式，包括缺失手指和缺失关节节段，同时保留真实的手部几何形状和关节活动。由此得到的基准称为 **IncNYU**，可作为一种受控且可重复的压力测试，用于评估不完整运动学结构下手部姿态估计方法的鲁棒性和泛化能力。

为应对这些挑战，我们提出**InsHaPE**，一种包容性手部姿态估计框架，旨在结构不完整的情况下保持鲁棒性，同时维持对标准设置下全关节手部的性能。这里，“包容性”表示这样一种设计目标：适应多样化的手部运动学结构，同时保持下游混合现实应用所需的一致姿态表示。InsHaPE 采用层次化运动学建模范式（见图 2(c)），用于捕获沿关节式手部结构的依赖关系。在此基础上，InsHaPE 由两个互补组件构成。首先，**拓扑感知蒸馏**策略将从全关节手部数据中学习到的结构先验迁移过来，从而在存在缺失结构的情况下提供可靠且符合解剖学一致性的初始化。在此表示的基础上，**渐进式运动学推理**模块沿运动学链推断关节，即使缺失多个运动学元素，也能实现完整且符合解剖学一致性的姿态估计。

大量实验表明，InsHaPE 在 IncNYU 上显著提升了性能，同时在 NYU 上保持了具有竞争力的准确率；在 MSRA 上的进一步验证也支持了其良好的泛化能力。此外，原型混合现实系统和用户研究证明了 InsHaPE 在结构不完整的情况下实现可靠下游混合现实交互的有效性。我们的贡献主要有三点：

- 我们针对结构不完整情况下的手部姿态估计问题进行建模，并提出 **IncNYU**，这是一个用于系统评估的生理约束基准。
- 我们提出一种**拓扑感知蒸馏**策略，可有效迁移结构先验，并提升模型在不完整运动学结构下的泛化能力。
- 我们引入**渐进式运动学推理**模块，通过层次化推断手部关节，实现稳健且拓扑一致的姿态估计。

本文其余部分安排如下。第 II节回顾手部姿态估计相关工作。第 III 节针对结构不完整情况下的手部姿态估计问题进行建模，并介绍用于系统评估的 IncNYU 基准。第 IV 节介绍所提出的 InsHaPE 框架。第 V 节提供理论分析，解释所提出的方法为何在结构不完整和标准设置下均能保持有效。第VI 节展示大量实验结果。第VII节报告系统级评估和一项用户研究。最后，第VIII节总结全文。

II. 相关工作

本节回顾手部姿态估计领域的相关工作，包括遮挡感知方法和结构感知建模策略，并讨论混合现实中面向包容性交互的现有探索，这些工作共同促成了我们的问题设定。

A. 手部姿态估计

手部姿态估计方法可根据输入模态和建模策略进行大致分类。早期方法主要依赖应用于深度图像的二维卷积神经网络 [26], [27], [28], [29], [30]。尽管计算效率较高，但这类方法难以捕捉手部完整的三维关节结构，并且容易受到视点变化的影响。随后出现的基于三维卷积神经网络的方法 [31], [32]通过在体素化深度表示上进行处理，缓解了这一局限，实现了体积推理，但代价是更高的内存占用和计算复杂度。为更好地利用几何信息，基于点云的方法（如 PointNet） [33]和 Hand-PointNet [34],直接处理三维点集以估计手部姿态。尽管这类方法能够有效捕捉空间结构，但通常依赖稀疏采样或简化表示来保持计算效率，这可能会限制其对细粒度关节配置的估计精度。近年来，人们提出了多模态方法 [35], [36] ，将二维图像线索与三维点云特征相结合，以利用互补的外观和几何信息。与此同时，人们也探索了生成式建模技术，尤其是扩散模型 [37], [38], [39],，用于手部姿态估计，以更好地建模二维到三维提升过程中的不确定性。现有大多数基于扩散的方法采用两阶段流程：首先估计二维关键点，然后细化三维姿态分布；而较新的单阶段方法则直接以三维表示为条件，从而提高效率。尽管取得了显著进展，现有方法和基准数据集主要是在手部关节完整且解剖结构正常这一假设下开发的。因此，它们对于涉及持续性结构变化的场景（例如缺失关节或异常手部形态）的鲁棒性和泛化能力仍然有限，这促使人们有必要开展结构感知建模。

B. 遮挡感知手部姿态估计

长期以来，遮挡条件下的鲁棒手部姿态估计一直是一个核心挑战，推动了多种遮挡感知建模策略的发展。早期工作（如分层混合密度网络（HMDN） [19], ）引入了概率建模方法，以利用深度图像对自遮挡关节的不确定性进行建模；当视觉证据不完整时，该方法能够实现多模态预测。后续方法融入了多视角或多模态线索，以增强鲁棒性，例如结合非模态分割与手部去遮挡模块，在严重遮挡条件下恢复完整的手部骨架 [20]。近期方法（包括 HandOccNet [22], ）采用基于Transformer的架构，将可见区域中的上下文信息注入遮挡区域，从而显著提升三维手部网格重建效果。除纯视觉线索外，研究者还探索了跨模态方法。例如，OCHID-Fi [40]将射频感知与视觉相结合，即使手部部分区域位于相机视野之外，也能估计手部姿态，凸显了互补感知模态的潜力。与此同时，掩码建模和基于注意力的框架，例如 OccRobNet [41],、HOMAE [42], 和 HOCAT [43], 利用自注意力、交叉注意力以及显式特征解耦，通过上下文感知推理来推断遮挡关节。综合性综述 [1], [44]进一步总结了遮挡感知标注策略、学习范式和数据集设计方面的进展。尽管取得了显著进展，大多数遮挡感知方法都建立在一个基本假设之上：所有手部关节在物理上均存在，只是暂时无法被观测到。因此，遮挡被视为一种暂时性或可恢复的状态，缺失信息可以通过空间上下文、时间连续性或跨模态线索进行推断。相比之下，涉及结构不完整性的场景，例如解剖学上缺失或永久变形的关节，则违背了这一假设，并且不存在有效的视觉或时间证据可用于恢复。因此，遮挡感知技术并不能直接适用于真正的结构缺失，这凸显了遮挡处理与不完整运动学结构条件下的姿态估计之间的关键空白。

C. 混合现实中的包容性设计与可访问性

近年来，提升混合现实系统的可访问性与包容性日益受到关注。以往研究考察了包容性设计原则和系统层面的框架，例如“包容式沉浸”；这类框架综合学术研究与行业实践，以识别沉浸式环境中的关键可访问性挑战 [45]。此外，相关综述还深入分析了不同障碍用户所面临的阻碍，包括视觉、听觉和运动功能障碍，并通过与领域专家及残障人士合作，对软件、硬件和伦理挑战进行分类 [46], [47]。在技术方面，现有研究探索了设备层面的包容性、自适应系统架构和以用户为中心的交互设计，以提升异质用户群体的可用性 [48],[49], [50]。其他工作则研究了医疗保健和大规模虚拟环境等特定应用领域中的可访问性，并提出了整合辅助技术、政策考量和互操作标准的高层框架 [51], [52]。尽管取得了这些进展，以往关于混合现实包容性的研究大多侧重于交互设计、界面适配或系统层面的可访问性，而对支撑众多混合现实应用的底层感知模块的鲁棒性关注相对较少。特别是，手部姿态估计作为用户动作与系统交互之间的关键接口，通常假定输入为全关节手部。这一研究空白凸显了对感知方法的需求：即使面对多样的手部运动学结构，这些方法仍能保持可靠。

III. 问题表述与基准 ARK

本节对结构不完整情况下的手部姿态估计进行问题表述，并介绍 IncNYU 基准。

A. 问题表述

我们研究结构不完整性情况下的手部姿态估计，其中手部运动学结构的部分区域是解剖学上缺失的，而非暂时被遮挡。此外，我们还将持续存在的传感器导致的观测缺失用作

一种可控的系统级压力测试，同时将这种情况与解剖学不完整性区分开来。在这两种情况下，都无法从当前视觉输入或短期时间线索中恢复缺失的信息。给定输入深度图像 $\mathbf{I} \in \mathbb{R}^{H\times W}$，目标是估计预定义的一组 K 手部关节的三维位置，

$$\hat{\mathcal{P}} = \{\hat{\boldsymbol{p}}_1, \hat{\boldsymbol{p}}_2, \ldots, \hat{\boldsymbol{p}}_K\}, \quad \hat{\boldsymbol{p}}_k \in \mathbb{R}^3, \tag{1}$$

ˆ其中 $\mathcal{P}$ 具有混合现实系统所要求的固定拓扑。

与传统的手部姿态估计不同 [9], [11], [53]，我们考虑这样一些情况：由于永久性结构变异（例如手指截肢或部分关节缺失），部分关节在解剖结构上缺失。我们将完整的关节记为 $\mathcal{K}_{\text{intact}} \subseteq \{1, \ldots, K\}$，将结构性缺失的关节记为 $\mathcal{K}_{\text{absent}} = \{1, \ldots, K\} \setminus \mathcal{K}_{\text{intact}}$。对于每个 $k \in \mathcal{K}_{\text{absent}}$，在$\mathbf{I}$中既没有有效的视觉观测，也不存在相应的物理关节活动。尽管如此，估计器仍必须输出涵盖所有 K 关节的完整姿态，因为许多下游混合现实应用 [5], [8], [15] 依赖固定拓扑姿态表示，无法适应可变关节集合。因此，我们的任务是推断一种拓扑完整且解剖学上协调的姿态表示，使其与现有交互流程保持兼容，而无需修改其固定拓扑接口。

形式上，我们学习以下映射：

$$f_\theta : \mathbf{I} \to \hat{\mathcal{P}}, \tag{2}$$

该映射同时定义在完整关节索引和结构性缺失关节索引上。对于完整关节，预测应与观测几何一致；对于结构性缺失的手部区域，则应提供运动学上合理的占位表示。

这种设定不同于能够处理遮挡的手部姿态估计；在后者中，关节被假定为存在，只是暂时不可见。相比之下，结构不完整性意味着运动学元素被永久移除，因此无法直接观测，也无法通过时间信息恢复。接下来，我们将介绍生理约束 IncNYU 基准。

B. IncNYU 基准

目前尚无公开可用的基准能够为手指截肢的手提供配对的深度观测和关节标注，而且现实案例十分罕见，难以大规模收集数据。一种直接的替代方案是通过改造面向遮挡的数据增强策略，合成类似截肢的样本。然而，

常见遮挡增强 [54], [55], [56]通常会叠加随机掩码，且不遵循解剖约束。相比之下，符合生理规律的截肢遵循结构规则，例如，只有在远节指骨也缺失时，中节指骨才可能缺失。忽略这些约束会导致不真实的样本。

为了解决这一局限，我们基于 NYU 数据集 [26]，在三维领域构建了一个符合生理特征的合成截肢基准。我们聚焦于手指截肢，因为这类情况在手部损伤中较为常见 [57]。生成截肢样本时，每根手指被选中的概率为 $1/2$。对于选中的手指截肢，

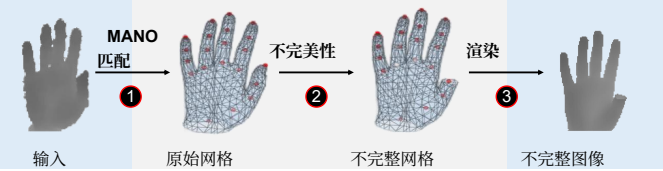


图 3：IncNYU 构建流程。与基于掩码的遮挡增强不同，我们通过根据解剖规则移除 MANO 网格片段，在三维空间中模拟手指截肢，并将生成的不完整手部渲染为深度图像。

2/3 被设定为指尖缺失，其余则随机分布在其他节段。

构建流程如图3所示。首先，我们将 MANO 网格拟合到 NYU 样本上，并手动标注手指区域，以建立区域标识符与网格顶点之间的映射。对于每个样本，根据上述概率选择截肢区域。随后，我们移除相应的网格顶点并更新网格面，按照从远端到近端的顺序递归删除节段，以确保符合生理规律，最终得到三维截肢手部网格。最后，我们使用原始 NYU 相机参数渲染修改后的网格。生成的深度图像构成了我们提出的**IncNYU**基准。

对于 IncNYU，训练集保留了原始 NYU 数据集的 10%（7,275 张深度图），其中 20% 为截肢手部，80% 为完整手部。测试集包含 8,252 张截肢深度图，规模与 NYU 测试集相同。在 14 个关键点评估子集的关节中，21% 被标记为结构上缺失。遵循先前的工作 [9]，模型预测 23 个关键点；所有 NYU 和 IncNYU 指标均使用标准 14 个关键点子集。此设计保留了真实几何形状，并支持在结构不完整的情况下进行评估。

IV. INSHAPE：包容性手部姿态估计

本节介绍所提出的 InsHaPE 框架。我们首先介绍整体架构和设计动机，随后详细说明其两个核心组件：拓扑感知蒸馏和渐进式运动学推理。最后，我们描述完整的训练流程。

A. 设计动机与框架

研究动机。 现有的大多数手部姿态估计方法 [9], [10], [11] 都是基于完整手部结构这一假设开发的，因此当运动学结构的某些部分持续缺失时，往往难以应对。因此，依赖固定拓扑手部姿态表示的混合现实应用常会出现交互不稳定、可访问性降低等问题。为支持包容性交互，即使在结构不完整的情况下，也必须估计出完整且符合解剖学的手部关键点集合。受这些挑战的启发，我们提出了**InsHaPE**，这是一种包容性手部姿态估计框架，专为处理结构不完整性而设计，同时兼容固定拓扑手部姿态表示。

核心思想。InsHaPE 的核心思想是，首先从完整手部结构中提炼拓扑感知先验，并

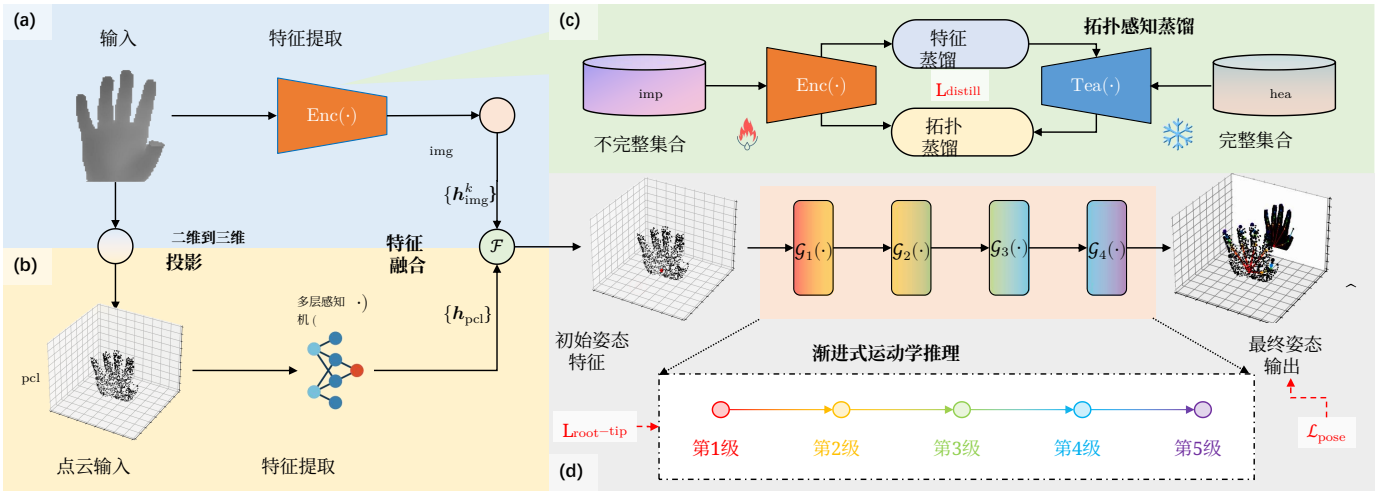


图 4：框架概览。(a) 从输入深度图像中提取外观特征。(b) 通过二维到三维反投影，将有效深度像素转换为几何点云特征。(c) 拓扑感知蒸馏从完整手部结构中迁移结构先验。(d) 渐进式运动学推理通过从根部到末端的层次化推理，实现完整姿态估计。分别表示拓扑感知蒸馏过程中的可训练骨干网络和冻结骨干网络。

在学习针对结构不完整输入的层次化运动学推理时保留这些先验。通过在推理过程中显式编码运动学层次结构，这种设计摆脱了逐关节回归或结构较弱的回归方法，即使手部结构的某些部分永久缺失，也能实现稳健且符合解剖学一致性的姿态估计。

框架概览。如图 4 所示，InsHaPE 是一个统一框架，旨在结构不完整的情况下实现稳健且包容性强的手部姿态估计。该框架旨在联合利用视觉外观和显式结构，同时明确纳入运动学拓扑，以确保与下游应用兼容。

给定输入深度图像$\mathbf{I}$，图像编码器 Enc(·) 首先提取稠密视觉特征

$$\boldsymbol{h}_{\text{img}} = \text{Enc}(\mathbf{I}), \tag{3}$$

这些特征同时编码全局上下文和局部手部部件语义。与此同时，二维到三维投影模块 $\phi(\cdot)$ [9], [10] 将有效深度像素反投影为三维点云

$$\boldsymbol{h}_{\text{pcl}} = \phi(\mathbf{I}), \tag{4}$$

提供明确的几何线索，以自然适应遮挡、形变和结构缺失。

为获得稳健且具备拓扑感知能力的表示，编码器 Enc(·) 通过以下方式进行预训练：**拓扑感知蒸馏**，如图 4(c)所示。使用具有完整手部结构 的数据训练的教师网络提供可靠的关节级关系先验，并将其蒸馏到学生编码器中。由此得到的编码器提供拓扑感知初始化，随后与下游姿态模块一同进行微调，同时保留蒸馏目标函数。

为以结构感知的方式融合外观与几何信息，首先通过关节映射函数 $\psi(\cdot)$ 将图像特征对齐到手部运动学拓扑，

$$\psi(\boldsymbol{h}_{\text{img}}) = \left\{\boldsymbol{h}^k_{\text{img}}\right\}_{k=1}^{K}, \tag{5}$$

而点云特征则通过多层感知机 $\rho(\cdot)$ 进行处理，以获得逐关节几何描述符

$$\rho(\boldsymbol{h}_{\text{pcl}}) = \left\{\boldsymbol{h}^k_{\text{pcl}}\right\}_{k=1}^{K}. \tag{6}$$

按照 [9], [10]，每个点都会附加从其投影像素处采样得到的图像特征；$\rho(\cdot)$通过共享多层感知机对距离当前关节 k 估计位置最近的 $S = 64$ 个点及其相对偏移进行编码，再经最大池化得到 $\boldsymbol{h}^k_{\text{pcl}}$；每个堆叠层都会根据精炼后的估计位置重新确定这些邻域的中心。我们将这些与关节对齐的特征融合到

$$\boldsymbol{z} = \mathcal{F}\left(\{\boldsymbol{h}^k_{\text{img}}\}_{k=1}^{K}, \{\boldsymbol{h}^k_{\text{pcl}}\}_{k=1}^{K}\right), \tag{7}$$

它对外观和运动学结构进行联合编码。

融合表示 $\boldsymbol{z}$ 随后由**渐进式运动学推理**处理，如图 4(d)所示，以推断完整的三维手部关节坐标集：

$$\hat{\mathcal{P}} = \mathcal{G}(\boldsymbol{z}), \tag{8}$$

其中 $\mathcal{G}(\cdot)$ 表示从根部到末端的层次化推理流程。通过显式建模沿关节式手部结构的运动学依赖关系，该模块能够在配置不完整的情况下实现稳健且拓扑一致的姿态估计。详情见第IV-C节。

训练遵循第IV-D节中的两阶段流程。首先仅使用 $\mathcal{L}_{\text{distill}}$ 对学生编码器进行预训练，随后通过下游微调最小化

$$\mathcal{L} = \mathcal{L}_{\text{pose}} + \lambda_1 \mathcal{L}_{\text{distill}} + \lambda_2 \mathcal{L}_{\text{root-tip}}. \tag{9}$$

此处，$\mathcal{L}_{\text{distill}}$ $K^{-1}\sum_{k=1}^{K}\|\boldsymbol{p}_k - \boldsymbol{p}^*_k\|_2^2$ $\mathcal{L}_{\text{pose}}$ 保留教师指导，用于监督关节回归，而$\mathcal{L}_{\text{root-tip}}$ 则提供链式平衡监督。

B. 拓扑感知蒸馏

结构不完整情况下手部姿态估计面临的一项关键挑战，是不完整运动学的训练数据有限

结构。主要使用完整手部训练的模型缺乏针对永久缺失关节的结构先验，因此泛化能力较差。拓扑感知蒸馏（见图4(c)）通过将关节级先验和关系先验从完整手部迁移到结构不完整表示，弥补了这一不足。

设计原理。 尽管某些运动学元素可能缺失，但不完整的手仍遵循相同的底层手部拓扑和功能约束。我们利用这一观察，将拓扑感知先验蒸馏到特征提取器中，即使缺少直接的视觉证据，也能获得符合解剖学的表示。这通过教师–学生蒸馏框架实现，该框架对齐关节级特征和关节间关系。

教师–学生特征蒸馏。 蒸馏使用成对的完整手部样本和不完整手部样本，分别记为$\mathbf{I}_c$ 和 $\mathbf{I}_i$，并通过相同的骨干网络架构进行处理。完整手部教师网络 Tea(·) 编码解剖学先验，学生编码器 Enc(·) 则将其迁移到结构不完整输入；教师网络在两个训练阶段均保持固定。

对于成对样本 $(\mathbf{I}_c,\mathbf{I}_i)$，提取特征图如下：

$$\boldsymbol{h}_c^{\mathrm{Tea}} = \mathrm{Tea}(\mathbf{I}_c), \qquad \boldsymbol{h}_i^{\mathrm{Enc}} = \mathrm{Enc}(\mathbf{I}_i). \tag{10}$$

将 $\psi(\cdot)$应用于这两个特征图，可得到用于结构对齐的关节级表示：

$$\psi(\boldsymbol{h}_c^{\mathrm{Tea}}) = \{\boldsymbol{f}_c^k\}_{k=1}^K, \qquad \psi(\boldsymbol{h}_i^{\mathrm{Enc}}) = \{\boldsymbol{f}_i^k\}_{k=1}^K, \tag{11}$$

其中，$\boldsymbol{f}^k{}_c$和 $\boldsymbol{f}_i^k$分别是教师网络和学生网络在关节k处的特征；K 是关节数。

拓扑关系对齐。 关系知识蒸馏通过保留数据样本之间的关系来传递结构信息 [59]。遵循这一关系迁移原则，我们对成对的完整手部与不完整手部样本中的关节间特征差异进行对齐，以保留手部特有的结构关系。对于任意一对关节（i，j），我们将关系差异定义为 $\Delta f_q^{ij} = f_q^i - f_q^j$，其中 $q \in \{c, i\}$。通过最小化以下目标函数，我们对齐教师和学生的关系表示

$$\mathcal{L}_{\mathrm{topo}} = \frac{1}{K^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\left\|\Delta\boldsymbol{f}_c^{ij} - \Delta\boldsymbol{f}_i^{ij}\right\|_2^2. \tag{12}$$

其中 $\boldsymbol{r}_k = \boldsymbol{f}_c^k - \boldsymbol{f}^k{}_i$和 $\bar{\boldsymbol{r}}$分别表示其在所有关节上的均值，该目标函数等价于 $\frac{2}{K}\sum_{k=1}^{K}\|\boldsymbol{r}_k - \bar{\boldsymbol{r}}\|_2^2 = 2\mathcal{L}_{\mathrm{joint}} - 2\|\bar{\boldsymbol{r}}\|_2^2$，其中 $\mathcal{L}_{\mathrm{joint}}$见式 (13)；也就是说，这是一个居中后的特征匹配目标函数，它消除了所有关节共有的偏移，且不包含显式的骨骼边或层级权重。因此，该目标函数迁移了从完整手部中学到的关节特征相对组织方式；它所传递的解剖结构隐含在教师特征中，而显式的运动学结构则通过渐进式运动学推理和 $\mathcal{L}_{\text{root-tip}}$引入。

蒸馏目标函数。 整体蒸馏目标函数整合了全局特征对齐、关节级一致性和拓扑关系保持：

$$\mathcal{L}_{\mathrm{distill}} = \mathcal{L}_{\mathrm{global}} + \alpha\mathcal{L}_{\mathrm{joint}} + \beta\mathcal{L}_{\mathrm{topo}}, \tag{13}$$

其中 $\mathcal{L}_{\mathrm{global}} = \|\boldsymbol{h}_c^{\mathrm{Tea}} - \boldsymbol{h}_i^{\mathrm{Enc}}\|_2^2$、$\mathcal{L}_{\mathrm{joint}} = \frac{1}{K}\sum_{k=1}^{K}\|\boldsymbol{f}^k{}_c - \boldsymbol{f}_i^k\|_2^2$ $\mathcal{L}_{\mathrm{topo}}$，并确保关系一致性。在预训练期间，仅优化学生编码器 Enc(·)。预训练参数用于初始化第二阶段，在该阶段中，学生编码器继续在 $\mathcal{L}_{\mathrm{distill}}$ 以及姿态学习目标的共同作用下进行优化。

C. 渐进式运动学推理

在结构不完整的情况下准确估计完整手部姿态，需要对关节间的运动学依赖关系进行显式建模。现有基于图的方法主要在特征建模阶段利用手部拓扑，其中关节特征以高度并行的方式进行聚合与细化。这种特征层面的推理有助于捕捉关节间相关性，但并未对最终关节位置的推断顺序施加显式约束。在结构不完整的情况下，这一区别尤为重要：远端关节通常缺少局部观测或观测较弱，而其近端祖先关节通常更易观测，能够提供可靠的解剖学锚点。因此，并行预测所有关节可能迫使模型在缺乏充分运动学指导的情况下推断证据有限的远端关节，这一点也反映在我们的消融分析中（见表V）。为解决这一局限，我们引入渐进式运动学推理模块，沿解剖学根到末端链组织姿态预测，并依据父关节预测和关节特定特征来估计远端关节。

设计原理。 人手构成一棵树状运动学图，其中手腕作为根节点，每个指尖对应一个叶节点。每个关节的空间构型受其沿运动学链的祖先节点约束。基于这一结构，我们将手部姿态估计表述为一个从根节点逐步推向远端关节的渐进式推理过程。这种层次结构为固定拓扑骨架补全提供了显式的运动学先验，并与特征级图推理相辅相成。

层次化姿态推理。 设 K 表示手部关节总数，$\boldsymbol{z}$ 表示上一节中描述的融合特征嵌入。预测的关节坐标表示为 $\hat{\mathcal{P}} = \{\hat{\boldsymbol{p}}_1, \ldots, \hat{\boldsymbol{p}}_K\}$，其中 $\hat{\boldsymbol{p}}_k \in \mathbb{R}^3$。推理从根关节（手腕，即 $k = 1$）开始，其位置根据全局嵌入直接估计：

$$\hat{\boldsymbol{p}}_1 = \mathbf{W}_1\boldsymbol{z} + \boldsymbol{b}_1, \tag{14}$$

其中$\mathbf{W}_1$ 和 $\boldsymbol{b}_1$ 是可学习参数。

对于每个后续关节 $k > 1$，其位置均根据其父关节 $\pi(k)$ 以及融合全局和局部信息的关节特定表示进行推断。具体而言，我们定义一个全局—局部交互函数

$$\boldsymbol{u}_k = \mathrm{GLI}_{\ell_k}(\boldsymbol{h}^k, \boldsymbol{z}), \tag{15}$$

其中，$\boldsymbol{h}^k$ 是关节对齐的局部特征，$\boldsymbol{z}$ 提供全局结构上下文，ℓ_k 表示关节k的运动学层级。在我们的实现中，$\mathrm{GLI}_{\ell_k}(\cdot)$是一种轻量级的拼接–多层感知机融合函数。它将关节对齐的局部特征 $\boldsymbol{h}^k$ 与全局结构特征 $\boldsymbol{z}$拼接起来，并通过双层多层感知机将拼接后的特征映射到 $\boldsymbol{u}_k$。随后，融合表示 $\boldsymbol{u}_k$ 通过层级特定推理函数与父节点预测相结合：

$$\hat{\boldsymbol{p}}_k = \mathcal{G}_{\ell_k}(\hat{\boldsymbol{p}}_{\pi(k)}, \boldsymbol{u}_k), \tag{16}$$

其中，$\mathcal{G}_{\ell_k}(\cdot)$是由同一运动学层级的所有关节共享的多层感知机。这种按层级共享参数的方式体现了解剖对称性，并促进手部功能角色相似的关节之间进行一致的推理。

训练目标。 渐进式运动学推理模块采用从根部到指尖的姿态损失进行监督，该损失根据手部骨架的层次化结构组织。不同于对所有关节进行均匀聚合，该损失按照手指运动学链对关节误差进行分组，并从手腕到远端关节提供均衡的监督。令 $\mathcal{C}$ 表示与各个手指对应的从根部到指尖的运动学链集合，其中每条链 $c \in \mathcal{C}$ 都是从根节点到指尖的有序关节序列。对所有从根部到指尖的运动学链上的关节误差取均值可得

$$\mathcal{L}_{\text{root-tip}} = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \frac{1}{|c|} \sum_{k \in c} \|\hat{\boldsymbol{p}}_k - \boldsymbol{p}_k^*\|_2^2 , \tag{17}$$

其中，$\boldsymbol{p}_k^*$ 表示第 k个关节的真实位置。

通过沿从根部到指尖的运动学链组织监督，该辅助损失使关节误差的分组与渐进式推理顺序保持一致。式中的父节点条件更新(16) 提供了明确的运动学依赖关系，而该损失确保每条手指链上的近端预测和远端预测都能获得监督。

讨论。 通过显式编码运动学层次结构，并使每个关节的预测均以其父节点和全局上下文为条件，所提出的推理模块避免了平坦的关节回归，并施加了符合解剖学意义的约束。这种渐进式方法能够在严重结构变化下实现完整且符合解剖学一致性的姿态估计，并通过在姿态层面执行结构一致性推理， complement 拓扑感知蒸馏模块。

D. 训练流程

InsHaPE遵循算法 1所示的两阶段策略：阶段1学习结构特征，阶段2将这些特征与姿态模块联合调整，以适应不完整输入。

阶段1：拓扑感知蒸馏预训练。第一阶段中，教师网络在完整手部NYU训练集上进行训练，以捕获全面的解剖学和运动学表示。随后，使用配对的完整与不完整手部图像优化学生编码器，通过最小化蒸馏损失，使其关节级特征和拓扑关系与教师网络对齐。此阶段仅更新学生编码器。所得参数用于初始化下游微调阶段。

阶段2：下游微调。 第二阶段中，学生编码器、特征融合模块和渐进式推理模块将进行联合微调。配对的完整/不完整输入在微调期间继续进行拓扑感知蒸馏，同时也使用不完整输入进行姿态回归和从根部到指尖的监督。这样既保留了教师网络提供的结构指导，也使下游模型能够适应结构不完整的姿态。

算法1： InsHaPE 训练流程

输入： 成对的完整/不完整手部数据 $\mathcal{D}_{\text{c}}$， $\mathcal{D}_{\text{i}}$
输出： 训练后的 InsHaPE 模型
// 阶段1：蒸馏预训练
在 $\mathcal{D}_{\text{c}}$ 上训练 Tea(·)，用于完整手部姿态回归;
对于每个成对的小批量 $(\mathbf{I}_{\text{c}},\mathbf{I}_{\text{i}})$ **执行**
  $\boldsymbol{h}_{\text{c}}$ =茶（$\mathbf{I}_{\text{c}}$），$\boldsymbol{h}_{\text{i}}$ =编码器（$\mathbf{I}_{\text{i}}$）;
  将特征映射到关节：$\{\boldsymbol{f}^k{}_{\text{c}}\}$, $\{\boldsymbol{f}^k_{\text{i}}\}$;
  计算 $\mathcal{L}_{\text{distill}}$ 并更新 Enc(·);
结束
// 阶段2：下游微调
对于每个成对的小批量 $(\mathbf{I}_{\text{c}},\mathbf{I}_{\text{i}})$ **执行**
  $\boldsymbol{h}_{\text{c}}$ =茶（$\mathbf{I}_{\text{c}}$），$\boldsymbol{h}_{\text{i}}$ =编码器（$\mathbf{I}_{\text{i}}$）;
  将特征映射到关节：$\{\boldsymbol{f}^k{}_{\text{c}}\}$, $\{\boldsymbol{f}^k_{\text{i}}\}$;
  提取关节级点云特征 $\{\boldsymbol{h}^k_{\text{pcl}}\}$;
  融合特征：$\boldsymbol{z} = \mathcal{F}(\{\boldsymbol{f}^k{}_{\text{i}}\}, \{\boldsymbol{h}^k_{\text{pcl}}\})$;
  预测关节坐标：$\{\hat{\boldsymbol{p}}_k\} = \mathcal{G}(\boldsymbol{z})$;
  计算微调目标函数 $\mathcal{L}$（见式(9)）;
  更新 Enc(·)、融合模块和推理模块;
结束

V. 理论分析

本节提供的是分析性直觉，而非无条件的性能保证。本节刻画了为何结构不完整性 会加剧关系歧义与误差传播，并指出所提出的结构先验在何种充分条件下能够缓解这些影响。因此，我们对 InsHaPE 引入的结构归纳偏置进行了条件刻画。回顾一下，拓扑感知蒸馏首先初始化学生表示，随后在式 $\widehat{\ }$（(9)）的微调目标中持续发挥结构正则 $\mathcal{R}$ 化项的作用。令 f（$=\frac{1}{n}\sum_{i=1}^{n}\ell\big(f\ (\mathbf{I}_i)\ ,\ \mathbf{y}_i\big)$ 表示经验风险，其中 ℓ 是 L-Lipschitz 且一致有界于 B。令 f_λ 表示采用两个结构目标学习得到的模型，其中 λ统称其权重；令 f_0 表示未施加这些约束时学习得到的对应模型。

定理*1*（结构不完整条件下的归纳偏置）：对于任意 $\delta \in (0,1)$，以至少 $1-\delta$的概率，

$$\mathcal{R}(f_\lambda) \le \widehat{\mathcal{R}}(f_\lambda) + \Gamma + \Delta + \kappa + C_0, \tag{18}$$

$$\mathcal{R}(f_0) \le \widehat{\mathcal{R}}(f_0) + \Gamma_0 + D_0 + \kappa_0 + C_0', \tag{19}$$

其中，$\mathcal{R}(\cdot)$ 表示期望风险，$C_0, C_0' \ge 0$ 汇总了其余近似项和噪声项。在给定的正则性条件和局部收缩条件下，以下结论成立:

- $\Gamma = 2L\,\mathfrak{R}_n(\mathcal{H}(\rho_1,\rho_2)) + 3B\sqrt{\log(2/\delta)/(2n)}$ 和 $\Gamma_0 = 2L\,\mathfrak{R}_n(\mathcal{H}_0) + 3B\sqrt{\log(2/\delta)/(2n)}$，其中 $\mathcal{H}(\rho_1,\rho_2) = \{f \in \mathcal{H}_0 :\ \mathcal{L}_{\text{蒸馏}}(f) \le \rho_1,\ \mathcal{L}_{\text{根节点-尖端}}(f) \le \rho_2\}$；因此，$\Gamma \le \Gamma_0$。
- $\Delta = C_1\sqrt{\rho_1}$ 在关系投影器为 Lipschitz 且其输出有界的假设下，对结构不完整性引起的关系差异给出上界，而 D_0 表示相应的无约束差异项。
- $\kappa = C_2\gamma^{d_{\min}}$ 刻画了运动学链上的传播扰动，前提是学习到的局部更新满足充分的压缩条件 $\gamma \in (0,1)$，而 κ_0 则是其无约束对应形式。

因此，该定理指出了一个充分条件范围，在此范围内，所提出的约束能够降低假设复杂度、关系差异和运动学误差传播。

证明 *1*：由于约束假设类 $\mathcal{H}(\rho_1,\ \rho_2)$ 满足 $\mathcal{H}(\rho_1,\ \rho_2) \subseteq \mathcal{H}_0$，其经验 Rademacher 复杂度遵循 $\mathfrak{R}_n(\mathcal{H}(\rho_1,\ \rho_2)) \leq \mathfrak{R}_n(\mathcal{H}_0)$。根据针对 L-Lipschitz 且 B-有界损失的标准泛化界，以至少 $1-\delta$ 的概率，假设类 $\mathcal{F}$ 中的任意 f 均满足 $\mathcal{R}(f) \leq \widehat{\mathcal{R}}(f) + 2L\,\mathfrak{R}_n(\mathcal{F}) + 3B\sqrt{\log(2/\delta)/(2n)}$。将该界应用于 $\mathcal{F} = \mathcal{H}(\rho_1,\ \rho_2)$ 可得到项 Γ，而将其应用于 $\mathcal{F} = \mathcal{H}_0$ 可得到 Γ_0，因此有 $\Gamma \leq \Gamma_0$。需要指出的是，这一复杂度关系刻画了约束假设类所引入的结构归纳偏置，但其本身并不能保证较低的期望风险。

在结构不完整性下，多个完整姿态可能对应于同一个部分观测，从而导致缺失关节的潜在表示存在歧义。使用第 IV-B 节中定义的关节特征，令 $\Delta \boldsymbol{f}^{ij}{}_{\mathrm{c}} = \boldsymbol{f}^{i}{}_{\mathrm{c}} - \boldsymbol{f}^{j}{}_{\mathrm{c}}$ 和 $\Delta \boldsymbol{f}^{ij}{}_{\mathrm{i}} = \boldsymbol{f}^{i}{}_{\mathrm{i}} - \boldsymbol{f}^{j}{}_{\mathrm{i}}$。假设关节特征投影器 ψ 满足 Lipschitz 条件且其输出有界，则完整输入与不完整输入的关系分布之间的差异可以由平均关系对齐误差的某个常数倍进行上界约束，例如可通过线性核最大均值差异论证得到。由于 $\beta\mathcal{L}_{\mathrm{topo}} \leq \mathcal{L}_{\mathrm{distill}}$ 对于 $\beta > 0$成立，因此有 $\mathrm{Disc}_{\mathrm{rel}} \leq \widetilde{C}_1\sqrt{\mathcal{L}_{\mathrm{distill}}(f)}$，其中 $\widetilde{C}_1$ 吸收因子 $1/\sqrt{\beta}$。为简化符号，定理将该常数记为 C_1。因此，对于满足 $\mathcal{L}_{\mathrm{distill}}(f_\lambda) \leq \rho_1$ 的约束解，我们得到 $\mathrm{Disc}_{\mathrm{rel}}(f_\lambda) \leq C_1\sqrt{\rho_1} = \Delta$。

考虑沿运动学树进行渐进式从根部到末端的推理。对于每条边 $e = (u \to v)$，设局部更新映射为 g_e，使得 $\hat{\boldsymbol{p}}_v = g_e(\hat{\boldsymbol{p}}_u,\ \boldsymbol{u}_v)$。如果学习到的局部更新映射相对于父节点状态是压缩的，即 $\|\partial g_e/\partial \hat{\boldsymbol{p}}_u\|_2 \leq \gamma < 1$，则沿任意长度为 d 的根节点到关节路径展开该递推，可得到传播扰动的几何衰减，因此上游误差对远端关节的贡献有界于 $C_2\gamma^d$。对结构性缺失的远端关节取最小深度 $d_{\min}$，可得 $\kappa = C_2\gamma^{d_{\min}}$。需要强调的是，此处引入 $\gamma < 1$ 是作为一个充分的局部收缩条件，而非保证每个模型或输入都具备的性质。在该条件下，渐进式的父节点条件推理为沿运动学层次结构衰减上游扰动提供了一种机制。

结合这些结果，可得到所述的条件表征，其中 Γ、Δ、κ 对应于 f_λ，而 Γ_0、D_0、κ_0 对应于相应的无约束表述。因此，该结果描述了在所述充分条件成立时，所提出的结构先验如何降低有效假设复杂度、约束关系歧义并衰减误差传播。证明至此结束。

讨论。该分析将三种与结构不完整性相关的影响区分开来。拓扑感知蒸馏通过对齐从完整手部结构中学习到的关节关系，引入结构归纳偏置并约束关系歧义；而渐进式运动学推理则沿解剖层级引入显式的父节点条件依赖。在满足充分的局部收缩条件时，这种分层建模形式能够减弱近端扰动向证据受限的远端关节传播。

我们强调，这些结果为 InsHaPE 的鲁棒性提供的是有条件的分析性解释，而非对性能提升的无条件保证。

六、实验

我们首先介绍实验设置。随后，我们在标准和结构不完整基准数据集上将 InsHaPE 与最先进方法进行比较，并开展跨设置泛化研究。最后，我们提供详细的消融研究和补充分析，以考察各组件的作用，并评估 InsHaPE 的计算效率和鲁棒性。

A. 实验设置

外部数据集。除第III节所述的 NYU 和 IncNYU 外，我们还使用 MSRA数据集 [23]，并构建其结构不完整对应数据集 IncMSRA。遵循 IncNYU 构建流程，我们选取符合解剖结构的区域，从远端到近端移除相应的手部片段，并在原始采集配置下渲染修改后的手部。每个不完整样本都保留其完整对应样本和关节标注。遵循标准 MSRA协议，我们对其九名被试进行留一被试交叉验证，取九折的平均结果，并遵循其输出和评估关节协议。

评估指标。我们使用三种互补指标评估手部姿态估计的性能，指标为逐关节位置误差均值（MPJPE），单位为毫米。对于包含 N 个样本的测试集，令 $\mathcal{K}_{\mathrm{eval}}$ 表示基准特定的评估关节集，其中 $K_{\mathrm{eval}} = |\mathcal{K}_{\mathrm{eval}}|$。对于每个样本 n，令 $\mathcal{S}^{(n)}_{\mathrm{com}}$ 和 $\mathcal{S}^{(n)}$ 不完整 分别表示与完整区域和结构不完整区域相关联的关节。我们定义 $\mathcal{N}_q = \{n:$ 对于 $q \in \{,\ |\mathcal{S}^{(n)}_q| > 0\}$ com，inc$\}$。我们在这些集合上计算三种评估指标，具体如下：

$$E_{\mathrm{com}} = \frac{1}{|\mathcal{N}_{\mathrm{com}}|}\sum_{n\in\mathcal{N}_{\mathrm{com}}}\frac{1}{|\mathcal{S}^{(n)}_{\mathrm{com}}|}\sum_{k\in\mathcal{S}^{(n)}_{\mathrm{com}}}\left\|\hat{\boldsymbol{p}}^{(n)}_k - \boldsymbol{p}^{*(n)}_k\right\|_2, \quad (20)$$

$$E_{\mathrm{inc}} = \frac{1}{|\mathcal{N}_{\mathrm{inc}}|}\sum_{n\in\mathcal{N}_{\mathrm{inc}}}\frac{1}{|\mathcal{S}^{(n)}_{\mathrm{inc}}|}\sum_{k\in\mathcal{S}^{(n)}_{\mathrm{inc}}}\left\|\hat{\boldsymbol{p}}^{(n)}_k - \boldsymbol{p}^{*(n)}_k\right\|_2, \quad (21)$$

$$E_{\mathrm{avg}} = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{K_{\mathrm{eval}}}\sum_{k\in\mathcal{K}_{\mathrm{eval}}}\left\|\hat{\boldsymbol{p}}^{(n)}_k - \boldsymbol{p}^{*(n)}_k\right\|_2, \quad (22)$$

其中，$\hat{\boldsymbol{p}}^{(n)}_k$ 和 $\boldsymbol{p}^{*(n)}_k$ 分别表示第 n个样本中第 k个关节的预测位置和参考位置。仅当相应的样本集$\mathcal{N}_q$ 非空时，才报告特定区域的指标。具体而言，E_{com} 衡量解剖结构完整区域上的 MPJPE，反映模型对正常观测到的关节的预测准确度。E_{inc} 衡量与结构不完整区域相关的关节上的 MPJPE。对于根据完整手部标注构建的合成不完整手部基准数据集，隐藏的完整手部关节提供了受控参考，用于评估方法能否为缺失区域推断出符合解剖学合理性且兼容固定拓扑的占位关节。因此，E_{inc} 应被视为对持续缺失情况下结构化骨架补全能力的代理指标，而非对不存在的解剖结构进行精确恢复

在真实的不完整手部中。 E_{avg} 补充了 E_{com} 和 E_{inc}通过报告 $\mathcal{K}_{\text{eval}}$中所有关节的 MPJPE，全面评估固定拓扑手部姿态估计。对于真实的不完整手部数据，无法获得隐藏的完整手部参考；因此，我们评估下游交互，而非虚拟关节误差（见第七节）。

实现细节。 按照手部姿态估计中的常见做法，在所有实验中，我们均采用 ResNet-50 作为教师编码器和学生编码器的骨干网络。编码器输出通过一个包含两个隐藏层的轻量级多层感知机映射为逐关节特征，两个隐藏层分别包含 256 和 128 个单元。点云分支采用三层多层感知机，隐藏层维度分别为 128、128 和 64。特征融合函数 $\mathcal{F}$ 对每个关节的图像特征和点云特征进行拼接，随后进行线性投影。渐进式运动学推理模块由一组多层感知机构成，在每个运动学层级共享一个应用于所有关节的多层感知机。每个多层感知机包含两个隐藏层，均采用 128 个单元和 ReLU 激活函数。对于基于 NYU 的实验，模型预测固定的 $K = 23$ 关节表示，而报告的 NYU 指标和 IncNYU 指标采用标准的 14 关节评估子集。对于 MSRA，按照标准评估协议，在全部 21 个标注关节上计算 MPJPE [23]。我们将阶段二蒸馏权重设为 $\lambda_1 = 0.01$，将阶段二链式损失权重设为 $\lambda_2 = 0.05$，并将内部蒸馏权重设为 $\alpha = 0.01$和 $\beta = 0.01$，具体设置遵循表X中的敏感性分析。模型采用 Adam优化器进行训练，学习率为 0.01，批量大小为 64，权重衰减为 0.01。学生编码器首先使用配对的完整和不完整手部图像进行 50 个周期的预训练。随后，在保留拓扑感知蒸馏的情况下，将其与融合模块和推理模块共同进行 100 个周期的微调。所有实验均在单块 NVIDIA RTX 4090 GPU 上使用 PyTorch 完成。我们通过随机旋转（$\pm 30°$）、缩放（$\pm 10\%$）和遮挡对数据进行增强。深度图像被调整为128×128 像素，每只手采样 1024 个点。

B. 与最先进方法的比较

表 I 和 II中的结果表明，InsHaPE 在 NYU 和 IncNYU 上取得了最低的总体误差，证明了其对完整手部和结构不完整手部的有效性。

在 NYU 上的表现。我们首先在标准 NYU 数据集上评估 InsHaPE，以验证引入面向包容性的设计不会损害传统完整手部设置下的性能。如表I所示，InsHaPE 的 MPJPE 为 7.07 毫米，与近期表现最强的基线方法 HandDAGT（7.12 毫米）相当。这一结果表明，所提出的设计在保持全关节手部竞争力精度的同时，主要提升了结构不完整情况下的鲁棒性。NYU 上的性能差异较小，这在预期之中，因为该数据集仅包含完整手部结构，所有关节均具有充分的几何和外观线索。在这种情况下，大多数近期方法，尤其是采用多模态输入或基于扩散的建模的方法，其性能已接近上限。重要的是，InsHaPE 并未牺牲

表 I：与最先进方法在NYU 数据集上的性能比较青色单元格标示最佳结果。输入模态包括二维深度图像 、 、 、体素（V）和三维点云（P）。

方法	会议	输入	MPJPE
Ren-9x6x6 [60]	JVCI'18	D	12.69
DeepPrior++ [61]	ICCV'17	D	12.24
Pose-Ren [62]	NEUCOM'20	D	11.81
DenseReg [2]	CVPR'18	D	10.20
CrossInfoNet [27]	CVPR'19	D	10.08
A2J [63]	ICCV'19	D	8.61
AWR [64]	AAAI'20	D	7.48
TriHorn-Net [56]	ESWA'23	D	7.68
JGR-P2O [28]	ECCV'20	D	8.29
SSRN [65]	NEUCOM'21	D	7.37
PHG [30]	TOC'21	D	7.39
HandPointNet [34]	CVPR'18	P	10.54
Hand-Transformer [66]	ECCV'20	P	9.80
点到点 [67]	ECCV'18	P	9.10
V2V [32]	CVPR'18	V	8.42
手部折叠 [68]	ICCV'21	P	8.58
手部R2N2 [53]	ICCV'23	P	7.27
IPNet [10]	AAAI'23	D+P	7.17
HandDiff [11]	CVPR'24	D+P	7.38
HandDAGT [9]	ECCV'24	D+P	7.12
InsHaPE（我们的方法）	–	D+P	7.07

在完整手部上的精度，尽管其设计明确针对结构不完整性，这证实了其归纳偏置与标准姿态估计相容。

在 IncNYU 上的表现。在 IncNYU 基准测试中，InsHaPE 的优势更加显著；该基准测试专门评估模型在结构不完整情况下的鲁棒性。当所有方法都在 IncNYU 上重新训练时（表II下半部分），InsHaPE 确立了明确的新标杆，平均误差 E_{avg} 为 8.00。与表现最强的基线 HandDAGT 相比，InsHaPE 将 E_{com}、 E_{inc}和 E_{avg} 分别降低了 32.1%、29.9% 和 30.4%。此次比较中的基线训练时未使用成对的完整手部观测；在相同的成对监督条件下，HandDAGT 的 E_{avg} 达到 8.18（表V中的 ID 3），而 InsHaPE 为 8.00。值得注意的是，虽然大多数基线在完整关节（E_{com}）与不完整关节（E_{inc}）上的表现差距显著，但 InsHaPE 大幅缩小了这一差距。这表明，该模型不仅提升了整体准确率，还改善了结构性缺失区域的固定拓扑骨架补全能力——这正是本文始终关注的核心挑战。

从 NYU 到 IncNYU 的泛化。我们进一步评估直接泛化：将基于 NYU 训练的模型直接应用于 IncNYU，而不进行重新训练（表 II上半部分）。在此设置下，所有模型（包括 InsHaPE）都仅使用完整的 NYU 数据进行训练，且在任何训练阶段都不会接触合成的不完整观测；因此不应用 TAD。此设置尤其具有挑战性，因为模型必须从完整手部数据外推到结构不完整的构型。在这一评估方案下，基线的性能显著下降，尤其是在不完整关节上，E_{inc} 通常超过 30 毫米。InsHaPE 的整体误差最低（$E_{\text{avg}} = 17.44$），较 HandDAGT 的 18.47 提升了 5.58%，而其不完整区域误差（E_{inc} =33。40）仍高于 HandR2N2 所取得的最佳不完整区域结果

表 II 性能与当前最先进方法-水-平-方法在**IncNYU** 数据集上的比较。青色单元格表示各列中的最佳结果。输入模态：深度图像（D）、三维体素（V）和三维点云（P）。

方法	发表 venue	输入	未使用 IncNYU 重新训练		
			E_{com}	E_{inc}	E_{avg}
A2J [63]	ICCV'19	D	82.58	91.85	87.28
AWR [64]	AAAI'20	D	67.44	94.16	73.72
TriHorn-Net [56]	ESWA'23	D	15.59	46.79	22.07
手R2N2 [53]	ICCV'23	P	16.25	26.36	19.23
IPNet [10]	AAAI'23	D+P	20.49	35.36	23.65
HandDiff [11]	CVPR'24	D+P	15.17	38.22	20.20
HandDAGT [9]	ECCV'24	D+P	13.28	34.53	18.47
InsHaPE（我们的方法）	–	D+P	13.28	33.40	17.44
			使用 IncNYU 重新训练		
A2J [63]	ICCV'19	D	15.54	31.83	20.16
AWR [64]	AAAI'20	D	15.51	33.37	19.83
TriHorn-Net [56]	ESWA'23	D	11.22	33.12	15.67
HandR2N2 [53]	ICCV'23	P	10.76	22.56	14.08
IPNet [10]	AAAI'23	D+P	11.42	15.87	12.36
HandDiff [11]	CVPR'24	D+P	13.76	25.57	15.99
HandDAGT [9]	ECCV'24	D+P	8.54	19.61	11.49
InsHaPE（我们的方法）	–	D+P	5.80	13.75	8.00

26.36。这些结果表明，模型整体上对 NYU 到 IncNYU 的域偏移具有更强的鲁棒性，但并不意味着其在每个关节子集上都更优。

方法级分析。性能提升与 InsHaPE 两个组件所发挥的互补作用一致。首先，拓扑感知蒸馏机制从完整手部结构中迁移关系先验，减少缺失关节潜在表示中的歧义，并缓解完整手部与不完整手部之间的域偏移。这一点在 IncNYU 中尤为关键，因为同一部分观测可能对应多种符合解剖学的补全结果。其次，渐进式运动学推理模块显式编码层次化运动学依赖关系，并约束误差沿手部骨架传播。通过依据近端估计来预测远端关节，该模块主要降低了受控消融实验中不完整关节的误差（表 V 和 VII）；由于对比方法与 InsHaPE 在多个组件上有所不同，其相较于 HandDAGT 等基于图的方法所取得的提升，体现的是整个框架的效果，而非该模块单独的作用。总体而言，这些设计使 InsHaPE 能够在标准基准数据集上取得更高的准确率，同时显著提升结构不完整情况下的鲁棒性和泛化能力，验证了其适用于包容性混合现实应用。

训练稳定性。为了评估 InsHaPE 对随机初始化的鲁棒性，我们进一步使用三个不同的随机种子在 IncNYU 上开展实验。模型在各次运行中表现一致，83 ± 0.04、13。55 ± 0.35 和 8。01 ± 0.10，分别对应 E_{com}，E_{inc}和 E_{avg}。平均结果与表II中的主要比较结果一致，表明所报告的性能在不同训练运行中保持稳定。

在 MSRA 和 IncMSRA 上进行外部验证。我们进一步在 MSRA 基准及其结构

表 III MSRA 和 IncMSRA 上的性能比较。我们报告标准 MSRA 设置下的 MPJPE（毫米）、从 MSRA 到 IncMSRA 的直接泛化结果（不进行重新训练），以及在 IncMSRA 上重新训练后的性能。

方法	会议	输入 MSRA		MSRA→ IncMSRA	IncMSRA
IPNet [10]	AAAI'23	D+P	6.64	7.18	7.36
HandDiff [11]	CVPR'24	D+P	6.53	6.90	5.89
HandDAGT [9]	ECCV'24	D+P	6.38	6.79	6.74
InsHaPE（我们的方法）	–	D+P	**5.52**	**6.23**	**5.77**

不完整变体 IncMSRA 上评估 InsHaPE。如表III所示，InsHaPE 在三种评估设置下均取得最低的 MPJPE。在标准 MSRA数据集上，InsHaPE 的误差为 5.52 毫米，分别较 HandDAGT、HandDiff 和 IPNet 优越 13.5%、15.5% 和 16.9%。在更具挑战性的直接泛化设置下，即从 MSRA 直接泛化到 IncMSRA 且不进行重新训练，InsHaPE 仍保持明显优势，MPJPE 为 6.23 毫米，而 HandDAGT、HandDiff 和 IPNet 分别为 6.79 毫米、6.90 毫米和 7.18 毫米，对应的相对降幅分别为 8.2%、9.7% 和 13.2%。在 IncMSRA 上重新训练后，InsHaPE 达到 5.77 毫米，较 HandDiff 的次优结果 5.89 毫米提升 2.0%。这些结果进一步表明，InsHaPE 能够在 NYU/IncNYU 之外实现泛化，同时保持对结构不完整性的鲁棒性。

C. 与最先进方法的计算比较

完整模型比较。我们从计算效率方面评估 InsHaPE（采用 3 个堆叠），并将其与近期同时使用深度图像和点云（D+P）的最先进方法进行比较。结果汇总于表 IV。尽管 InsHaPE 实现了最低的平均误差（$E_{avg} = 8.00$），但在所比较的 D+P 方法中，其测得的模型推理延迟也最低，在 NVIDIA RTX 4090 上每帧平均运行时间为 29.43 毫秒。这明显快于 HandDAGT（45.47 毫秒）和 IPNet（45.65 毫秒）。这些测量仅涉及模型推理，不包括传感器采集、数据传输、下游识别或渲染延迟。在模型复杂度方面，InsHaPE 的参数数量（108.01M）多于所比较的基线。不过，其浮点运算次数（23.47G）与 HandDAGT（23.74G）相当，且显著低于 HandDiff（46.68G）。这表明，额外的参数并未按比例增加浮点运算次数或实测模型推理延迟。重要的是，更大的模型带来了显著的性能提升：在 IncNYU 上，InsHaPE 相较于 HandDAGT 将平均误差降低了 30.4%，同时在测试硬件上的模型推理速度更快。这提供了良好的权衡，但端到端可穿戴混合现实延迟尚未经过测试。

堆叠深度与效率分析。与 HandDAGT 类似，[9]，InsHaPE 采用堆叠式渐进式运动学推理模块，迭代优化手部姿态预测并确保结构一致性。每个堆叠模块都将前一堆叠模块预测的关节坐标和关节特征作为输入，而第一个堆叠模块则以根据全局图像回归得到的初始估计为起点

表 IV INCNYU 数据集上的计算比较。

方法	参数量 (M)	浮点运算次数 (G)	推理 (ms)	E_{avg}
IPNet [10]	39.51	2.97	45.65	12.36
HandDiff [11]	67.55	46.68	59.75	15.99
HandDAGT [9]	69.48	23.74	45.47	11.49
InsHaPE，使用 1 个堆叠（我们的方法）	74.20	17.09	18.12	10.67
InsHaPE，使用 2 个堆叠（我们的方法）	91.10	20.42	23.44	8.10
InsHaPE，使用 3 个堆叠（我们的方法）	108.01	23.47	29.43	8.00

以及点云编码。虽然增加堆叠深度可以提升表征能力，但也会增加计算成本，这可能会限制其在对延迟敏感的混合现实系统中的部署。为分析这一权衡，我们评估了采用一个、两个和三个推理堆栈的 InsHaPE。如表 IV所示，随着堆栈数量增加，准确率持续提升，误差从 10.67（1 个堆叠）降至 E_{avg} 8.10（2 个堆叠），最终降至 8.00（3 个堆叠）。这一改进以更高的计算量为代价：推理时间从 18.12ms 增至 29.43ms，浮点运算次数从 17.09G 增至 23.47G，这表明超过两个堆叠后，准确率提升逐渐趋于有限。重要的是，即使只有一个推理堆栈，InsHaPE 的表现也已优于所有参与比较的最先进方法。具体而言，采用一个堆叠的 InsHaPE 平均误差为 10.67，显著低于 HandDAGT（11.49）、IPNet（12.36）和 HandDiff（15.99），同时推理延迟不到 HandDAGT 的一半。因此，即使只有一个堆叠，所提出的两个模块仍然有效。

D. 消融研究

核心模块的影响。我们首先分析 InsHaPE 中各个核心组件的贡献，结果汇总于表 V。移除全局-局部交互（GLI，ID 2）后，E_{avg} 在单次运行中从 8.00 增至 8.15，这表明显式融合关节级局部线索与全局结构上下文是有益的。这与 GLI 平衡细粒度外观线索和整体手部结构的作用相一致。对于 ID 3，我们按照常见的扁平关节预测范式，将渐进式根到指尖推理模块替换为并行预测头 [9]。该预测头以相同的融合特征作为输入，并行预测所有关节偏移量，不进行运动学级分解或父节点条件推理。由于所有配置均采用相同的三堆栈设计，且 ID 3 也禁用了 GLI，因此 ID 1 与 ID 3 之间的差异反映了这两个组件的联合效果。在 GLI 均缺失的 ID 2 和 ID 3 之间，PKR* 使 E_{inc}从 14.15 降至 13.88，但完整关节方面略有损失；鉴于这一次运行中的差异处于完整模型的随机种子变化范围内，表 VII 中的随机种子匹配对照为父节点条件化提供了主要证据。没有 TAD 时，PKR* 的作用要大得多：ID 5 和 ID 6 仅在 PKR* 上有所不同，后者使 E_{inc} 从 19.61 降至 16.31，并使 E_{avg} 从 11.49 降至 10.95。应用 TAD 后，增益较小，这与两个组件都能降低证据有限的远端关节歧义相一致。这一结果也符合渐进式运动学推理的设计动机：在结构不完整的情况下，远端关节的局部观测有限，且

表 V 核心模块消融实验结果：在NYU上的结果。所示百分比表示相对于完整模型（ID 1）的相对误差变化。**TAD**：拓扑感知蒸馏（§IV-B）；PKR*：渐进式**运动学推理**（§IV-C），不含全局-局部交互（GLI，公式（15））。

ID	配置			E_{com}	E_{inc}	E_{avg}
	TAD	PKR*	GLI			
1	✔	✔	✔	5.80	13.75	8.00
2	✔	✔	✗	5.93 $^{\uparrow 2\%}$	13.88 $^{\uparrow 1\%}$	8.15 $^{\uparrow 2\%}$
3	✔	✗	✗	5.77$^{\approx 0\%}$	14.15 $^{\uparrow 3\%}$	8.18 $^{\uparrow 2\%}$
4	✗	✔	✔	6.94 $^{\uparrow 20\%}$	13.84 $^{\uparrow 1\%}$	9.56 $^{\uparrow 20\%}$
5	✗	✔	✗	8.16 $^{\uparrow 41\%}$	16.31 $^{\uparrow 19\%}$	10.95 $^{\uparrow 37\%}$
6	✗	✗	✗	8.54 $^{\uparrow 47\%}$	19.61 $^{\uparrow 43\%}$	11.49 $^{\uparrow 44\%}$

表 VI ONINCNYU 上的蒸馏损失消融结果。报告的百分比是与 ID 1 相比的误差变化。

ID	配置			E_{com}	E_{inc}	E_{avg}
	$\mathcal{L}_{global}$	$\mathcal{L}_{joint}$	$\mathcal{L}_{topo}$			
1	✔	✔	✔	5.80	13.75	8.00
2	✔	✔	✗	6.72 $^{\uparrow 16\%}$	14.20 $^{\uparrow 3\%}$	8.81 $^{\uparrow 10\%}$
3	✔	✗	✗	7.66 $^{\uparrow 32\%}$	16.69 $^{\uparrow 21\%}$	10.27 $^{\uparrow 28\%}$
4	✗	✔	✔	5.97 $^{\uparrow 3\%}$	14.00 $^{\uparrow 2\%}$	8.27 $^{\uparrow 3\%}$
5	✗	✔	✗	6.19 $^{\uparrow 7\%}$	14.10 $^{\uparrow 3\%}$	8.41 $^{\uparrow 5\%}$
6	✗	✗	✔	6.37 $^{\uparrow 10\%}$	14.23 $^{\uparrow 3\%}$	8.62 $^{\uparrow 8\%}$
7	✗	✗	✗	6.94 $^{\uparrow 20\%}$	13.84 $^{\uparrow 1\%}$	9.56 $^{\uparrow 20\%}$

可以通过更可靠的近端线索加以约束。排除拓扑感知蒸馏（ID 4 和 ID 5）会导致 E_{avg} 明显下降 20–37%，表明它有助于缩小完整手部结构与不完整手部结构之间的域差距。值得注意的是，在保留渐进式推理的同时移除蒸馏（ID 4），已经会导致 E_{com} 大幅增加，这表明缺少迁移来的结构先验时，学习到的表征本身会变得不那么可靠。最后，移除全部三个组件（ID 6）后，平均误差大幅增加了 44%。这种严重的性能下降表明，InsHaPE 并不依赖任何单一模块；相反，稳健的性能来自拓扑感知蒸馏、GLI 和渐进式运动学推理之间的协同作用。

蒸馏损失的影响。我们进一步研究了每种蒸馏损失的作用，结果见表 VI。完整组合 $\mathcal{L}_{global}+\mathcal{L}_{joint}+\mathcal{L}_{topo}$（ID 1，见公式（13））在所有指标上均取得最低误差。移除拓扑对齐、保留全局对齐和关节级对齐（ID 2）后，E_{avg} 从 8.00 增加到 8.81。移除全局对齐、保留关节级对齐和拓扑对齐（ID 4）后，其增加到 8.27。关节级对齐的益处得到了两组受控对比的支持：在仅使用全局蒸馏的基础上加入关节级对齐，可将 E_{avg}从 10.27 降至 8.81（ID 3 和 ID 2）；而在仅使用拓扑蒸馏的基础上加入关节级对齐，可将误差从 8.62 降至 8.27（ID 6 和 ID 4）。单损失配置（ID 3、5 和 6）的总体误差均高于完整组合，而移除所有蒸馏损失（ID 7）则使 E_{avg} 增至 9.56。综上，这些结果支持全局对齐、关节级对齐和拓扑对齐所带来的互补益处。

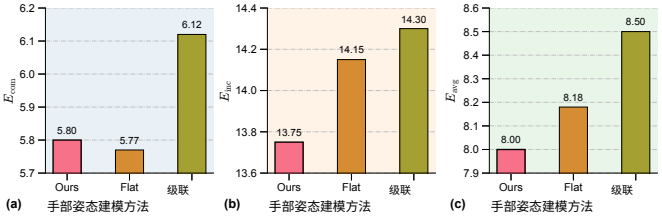


图5：不同手部姿态建模范式的消融结果（参见图 2）。我们比较了扁平逐关节建模、级联逐手指建模，以及本文提出的分层根到指尖推理，比较指标包括 (a) E_{com}、(b) E_{inc} 和 (c) E_{avg}。

运动学建模范式的影响。 为分析渐进式运动学推理的作用，我们比较了三种范式（见图 2）：扁平逐关节建模、级联逐手指建模和分层根到指尖推理。三种变体均采用相同的三堆栈设计、姿态回归头、目标函数（见公式（9））和训练计划，参数量分别为 107.95M、114.57M 和 108.01M；它们的区别在于每个堆栈中对关节关系的建模方式不同。所有变体均保留 HandDAGT 的特征级图推理 [9]：扁平变体对应表 V中的 ID 3，即带有 TAD 的 HandDAGT；级联变体使用独立模块对每根手指进行建模；只有分层变体，即完整模型，使用 GLI 和父节点条件化。如图 5所示，分层建模形式取得了最佳总体性能，将E_{avg} 降至 8.00，而扁平建模和级联建模分别为 8.18 和 8.50。与级联建模相比，分层根到指尖推理分别将 E_{com}、 E_{inc} 和E_{avg} 降低了 5.23%、3.85% 和 5.88%；如下文所示，这一差距并不能仅由 GLI 解释。尽管扁平建模取得了略低的 E_{com} （5.77 对 5.80），但它没有提供显式的结构推理，因而导致更高的总体误差。这些结果表明，PKR 带来的增益不能仅仅归因于引入了额外的预测阶段。由于 GLI 仅用于分层变体，图 5对整体预测架构进行了比较。当从分层模型中同样移除 GLI（表 V中的 ID 2）后，其仍然取得三种架构中最低的 E_{inc} （13.88 对 14.15 和 14.30），并且 E_{avg}低于级联变体（8.15 对 8.50），而 E_{avg} 在这次单次运行中接近扁平变体（8.18）。

父级条件控制。 为单独考察父级条件的作用，我们使用式 $\hat{\boldsymbol{p}}_{\pi(k)}$ 中的父节点预测重新训练完整模型，(16) 将其设为零，同时保持其他所有组件和训练设置不变，并为两个模型使用相同的三个随机种子。由于该模型使用这一恒定输入重新训练后，该输入不携带任何有关父节点预测的信息，因此在保持架构不变的同时移除了父节点条件化。如表 VII所示，移除父节点输入主要影响不完整关节，使 E_{inc} 在三个种子中的每一个上均有所上升，平均值从 13.55 升至 14.08；这一差异大于任一模型的种子间标准差，而 E_{com}几乎没有变化；表 V 中的完整模型结果就是这三次运行中的一次。这与其预期作用一致

表 VII IncNYU 上的父级条件控制（在三个匹配随机种子上的均值 ± 标准差）。

配置	E_{com}	E_{inc}	E_{avg}
InsHaPE（完整模型）	5.83 ± 0.04	13.55 ± 0.35	8.01 ± 0.10
无父节点条件化	5.86 ± 0.05	14.08 ± 0.28	8.13 ± 0.08

表 VIII IncNYU 上的几何点云消融。

配置	E_{com}	E_{inc}	E_{avg}
InsHaPE（完整模型）	5.80	13.75	8.00
不含点云	9.34 $^{\uparrow 61\%}$	16.71 $^{\uparrow 22\%}$	11.35 $^{\uparrow 42\%}$

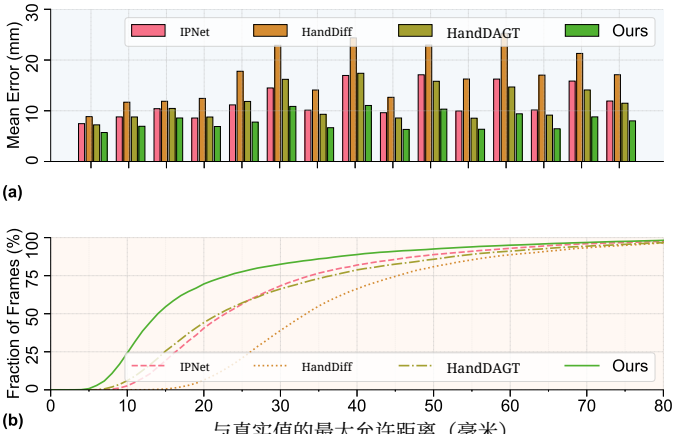


图6：IncNYU 数据集上的逐关节误差分析。（a）：平均关节位置误差（MPJPE），显示了从手掌到指尖难度逐渐增加的趋势。（b）：PCK风格曲线，展示了在给定距离阈值内与真实值（GT）的距离符合要求的帧所占比例。

：当远端关节缺乏直接的视觉观测时，近端估计最为重要。

几何点云特征的影响。为评估几何点云特征的作用，我们移除点云分支，同时保持其他组件不变。这里的点云由输入深度图像经深度到三维投影得到。如表VIII所示，移除几何表示后，总体误差从 8.00 增至 11.35，完整关节和不完整关节的表现均有所下降。这些结果表明，由深度信息得到的三维几何表示提供了深度外观特征之外的互补结构信息，这对于结构不完整情况下的稳健姿态估计至关重要。尽管性能有所下降，不含点云的变体仍可与表II中的现有仅使用深度的方法相竞争，这表明所提出的框架能够有效利用深度输入中的互补外观和几何线索。

配对完整手部监督的影响。为了考察获取成对的完整手部信息所带来的益处，我们还使用 InsHaPE 可用的相同配对完整/不完整训练数据对 HandDiff [11] 进行了适配。具体而言，我们利用对应的完整手部观测对其特征提取模块进行配对特征蒸馏，然后使用蒸馏后的特征提取器初始化 HandDiff，以进行后续训练。这使 HandDiff 能够获取相同的成对完整手部信息

表 IX IncNYU 上成对监督的比较。

配置	E_{com}	E_{inc}	E_{avg}
HandDiff	13.76	25.57	15.99
HandDiff +成对监督	9.87 $^{\downarrow 28\%}$	16.93 $^{\downarrow 34\%}$	12.29 $^{\downarrow 23\%}$
InsHaPE（我们的方法）	**5.80**	**13.75**	**8.00**

表 X INCNYU 上损失相关超参数的敏感性分析。我们报告 E_{avg} 在各超参数取不同值时的结果，同时将其他超参数保持为默认设置。每行中的最佳结果以粗体标出。

超参数	0.005	0.01	0.05	0.1
λ_1	8.14	**8.00**	8.29	8.47
λ_2	8.18	8.12	**8.00**	8.32
α	8.07	**8.00**	8.14	8.18
β	8.11	**8.00**	8.15	8.23

，同时保留其原有的姿态估计框架。如表 IX所示，该配置将 HandDiff 的 E_{avg} 从 15.99 降至 12.29，表明获取成对的完整手部信息是有益的。在所评估的设置下，InsHaPE 的总体误差进一步降至 8.00。该比较控制了对配对训练数据的获取，但两种方法利用这些信息的方式不同。与 InsHaPE 不同，该 HandDiff 变体仅将蒸馏后的特征提取器用于初始化，并未在微调期间保留蒸馏。与 InsHaPE 在这方面保持一致的 HandDAGT 对照实验见表 V中的 ID 6 和 3：ID 6 对应 HandDAGT，ID 3 则加入了在微调期间保留的 TAD，使 E_{avg} 在 IncNYU 上从 11.49 降至 8.18。

损失权重的影响。 我们分析了 InsHaPE 对损失相关超参数的敏感性，包括阶段二蒸馏权重（λ_1）、阶段二链式损失权重（λ_2）以及蒸馏目标函数中的内部平衡因子 α 和 β。如表 X所示，InsHaPE 在测试的超参数取值范围内表现稳健。最佳性能出现在 $\lambda_1 = 0.01$，取值较小或较大时性能仅略有下降，这表明拓扑感知蒸馏无需精细调参。对于 λ_2，性能同样稳健，最优值约为 $\lambda_2 = 0.05$，说明渐进式运动学推理发挥的是结构正则化项的作用，而非占主导地位的损失项。用于平衡关节级对齐与拓扑对齐的内部权重 α 和β的影响也很小：即使它们相差一个数量级，性能仍接近最优。总体而言，这些结果表明，InsHaPE 无需对损失权重进行太多调优。

E. 定量与定性结果

关节级误差统计。图6 给出了 IncNYU 上关节级误差分布的结果。如图6(a)所示，掌心处的误差最低，并向远端关节逐渐增大，尤其是指尖。这一趋势符合预期，因为远端关节具有更高的自由度和更大的变化性，因此更难准确估计。值得注意的是，InsHaPE 在所有关节上始终保持较低误差，且误差向指尖方向增长得更为平缓，表明其在结构不完整情况下具有较强的鲁棒性。如图6(b)所示，我们报告了

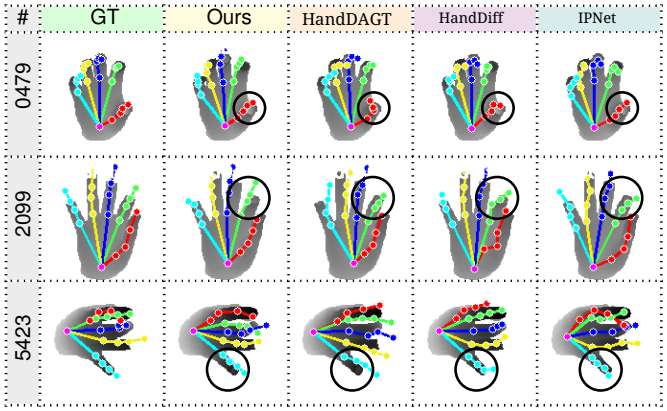


图7：InsHaPE 与 IncNYU 数据集上最先进基线方法的定性比较。圆圈标出了重点部分。

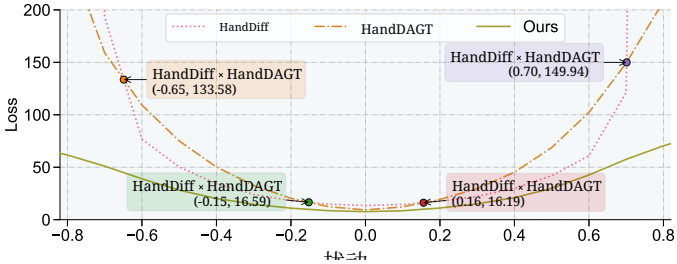


图8：IncNYU 上的一维损失景观切片。与比较方法相比，InsHaPE 沿采样的滤波器归一化扰动方向呈现出更平坦的损失曲线。

百分比正确关键点（PCK）曲线，展示了关节误差处于不同距离阈值范围内的帧所占比例。InsHaPE 获得的曲线下面积更大，表明其在广泛的容差范围内始终具有更高的准确率。这一结果说明，InsHaPE 不仅提高了均值准确率，也减少了大误差，尤其是在较难估计的远端关节上。

定性比较。 图 7展示了 InsHaPE 与 IncNYU 数据集上具有代表性的最先进方法（包括 HandDAGT、HandDiff 和 IPNet）之间的定性比较。基线方法在截肢或严重结构不完整性影响的区域中经常失败，从而导致关节缺失、手指构型扭曲或运动学上不合理的姿态。相比之下，InsHaPE 能够更加稳定地产生解剖学上协调的固定拓扑姿态表示。具体而言，即使中间关节缺失，InsHaPE 也能更好地定位指尖等远端关节，并保持手部的整体结构。这些视觉结果与定量改进相一致，凸显了拓扑感知蒸馏和渐进式运动学推理在处理结构不完整手部方面的有效性。

损失景观分析。为考察训练后模型对参数扰动的局部敏感性，我们使用 [69] 中的滤波器归一化技术，在 IncNYU 上可视化平均一维损失剖面。对于每个训练好的模型，我们分别采样 10 个随机扰动方向，并应用滤波器归一化。我们评估沿每个方向的损失，扰动系数在 $[-0.8, 0.8]$ 范围内变化，并在每个系数处对 10 个方向的损失求平均。

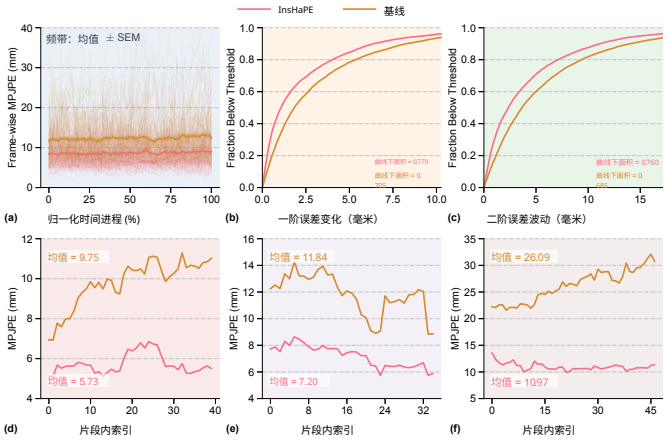


图 9：IncNYU 上的时间误差变化分析。(a) 推断出的伪连续片段中的逐帧 MPJPE，以及均值趋势和均值标准误。(b–c) 一阶误差变化和二阶误差波动的基于阈值的累积分布；曲线越靠近左上角，表示标量误差的变化越小。(d–f) 分别展示来自低、中、高误差区间的代表性伪连续片段。

处的损失。如图 8 所示，与 HandDAGT 和 HandDiff 相比，InsHaPE 的平均损失剖面更平坦，表明其对采样参数扰动的敏感性较低。尽管对多个方向求平均可以降低对任何单一方向的依赖，但这些剖面仍只是对局部参数敏感性的定性评估。它们无法保证泛化能力，也不能验证第 V 节中的压缩条件。

时间误差变化。为补充逐帧精度评估，我们分析了 IncNYU 上标量姿态误差的时间行为（见图 9）。由于 NYU手部姿态数据集 [26], 未提供明确的序列边界，我们根据真实运动推断伪连续片段：当相邻帧之间的根部位移或 14 个评估关节的平均相对根部位移超过该量在测试集上的中位数加上五倍中位数绝对偏差时，就在这两帧之间设定边界；所有时间比较均限制在这些片段内。I ncNYU 的缺失手指配置会随帧变化，因此在一个片段内并非固定；两种方法均在相同帧上评估，因此面对相同的配置变化。具体而言，我们保留了 114 个至少包含 10 帧的片段，涵盖 8,252 个测试帧中的 7,383 帧（89.47%）。如图 9(a)所示，InsHaPE 在推断出的片段中始终保持较低的逐帧 MPJPE。总体平均 MPJPE 从 HandDAGT 基线的 11.49 毫米[9] 降至 InsHaPE 的 8.00 毫米，相对降低了 30.4%。我们进一步使用一阶差分和二阶差分来刻画时间误差变化，二者分别量化相邻帧之间的误差变化和误差轨迹的高阶波动。如图 9(b)和

9(c)中，InsHaPE 的一致性曲线始终更接近左上区域。相应的归一化曲线下面积从一阶度量的 0.705 提升至 0.779，并从二阶度量的 0.685 提升至 0.760

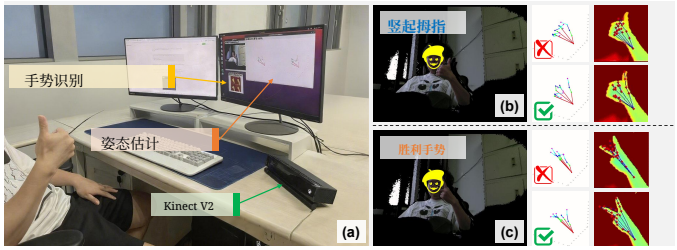


图 10：由 InsHaPE 驱动的原型混合现实系统。(a) 手部姿态估计与交互流程。(b–c) 在结构不完整的手部构型下的代表性交互场景。完整演示请参阅补充视频。

，相对提升幅度分别为 10.5% 和 11.0%。图 9(d) 至9(f) 进一步展示了低误差、中等误差和高误差区间中的代表性节段。在这三个区间内，InsHaPE 均表现出更低的逐帧误差，以及比基线更小的标量误差变化。这些结果表明，InsHaPE 的标量误差随时间变化更小。由于方向相反但幅值相等的偏移会使标量误差保持不变，因此标量误差无法衡量关节级时间稳定性。

VII. 原型系统与用户研究

虽然第 VI 节展示了 InsHaPE 在基准数据集上的准确性和鲁棒性，但仍有必要考察这些提升能否转化为下游混合现实交互中的实际收益。许多混合现实系统将手部姿态作为中间表示，其中不稳定或不符合解剖结构的骨架可能直接影响交互质量和下游动作理解 [8]。因此，我们将 InsHaPE 集成到一个原型混合现实系统中（见图 10），并使用我们先前混合现实系统中的基于规则的手势识别器评估其下游影响 [8],。该识别器根据手部关节之间预定义的距离对每个手势进行分类，对于两种估计器均无需训练或修改。我们将 InsHaPE 估计的骨架与当前最先进的基线 HandDAGT 生成的骨架进行比较 [9], ，同时保持下游识别框架不变。我们开展了两项互补评估：一项针对模拟持续不完整情况的受控研究，以及一项涉及真实手指不完整用户的试点研究。

A. 通过模拟不完整性进行的受控研究

参与者与流程。 共有18名大学生（男性9名，女性9名）参与了本研究，年龄为20–27岁（均值 = 23.1，标准差 =2.1）。所有参与者的视力均正常或经矫正后正常，均报告无手部相关疾病，并提供了知情同意。研究方案在参与者招募前获得了伦理批准，批准文件已随补充材料一并提交。此前无人使用过受测系统，但大多数人熟悉混合现实设备。为提供受控且可复现的环境，参与者在一根手指上佩戴轻质实体遮罩，以模拟持续缺失的观测。我们强调，该设置模拟的是健全参与者持续的视觉缺失，

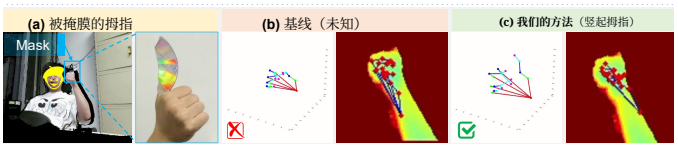


图 11：使用光敏手指掩膜对持续缺失的观测进行受控模拟。(a) 实验设置。(b) HandDAGT [9] 在模拟不完整观测条件下会产生错误且不合理的骨架预测。(c) InsHaPE 能够维持更合理且符合解剖学一致性的骨架估计。

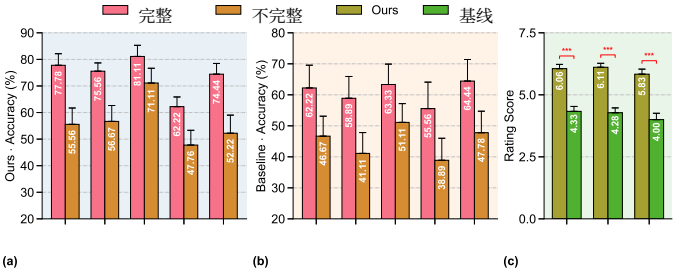


图 12：模拟持续不完整条件下受控用户研究的结果。(a–b) InsHaPE 和 HandDAGT 在完整手部条件和模拟不完整手部条件下的手势识别准确率。(c) 受试者对感知自然度、交互易用性和可靠性的主观评分，采用 7点评分量表。柱形表示均值，误差线表示均值标准误，(c) 中的 *** 表示评分配对检验中的 $p < 0.001$ 。

不应被解读为真实的解剖学不完整性。研究在安静的实验室环境中进行（见图 10(a)）。参与者首先完成了一个简短的训练环节，以熟悉系统。我们使用相同的研究方案评估 InsHaPE 和 HandDAGT。参与者在两种观测条件下完成五种预定义手势（竖起拇指、捏合、张开手掌、握拳和胜利手势）：（1）完整手部；（2）遮挡一根手指的模拟不完整手部（见图 11(a)）。两种估计器的测试顺序以及两种观测条件的顺序在参与者之间进行了平衡，各手势在每种条件下重复五次。每个估计器均在独立的环节中实时运行，因此两种方法处理的是相同手势的不同次表演，而非完全相同的录制内容。期间安排了短暂休息以减少疲劳；每次环节持续约30分钟。

评估协议。我们从两个角度评估该系统。(1)客观性能：手势识别准确率，即正确分类的手势所占百分比；我们比较不同姿态估计方法（InsHaPE 与 HandDAGT）以及不同观测条件（完整与模拟不完整）下的表现。(2)主观用户体验：参与者分别使用 7 分制对基于各估计器的系统进行评分，评价感知自然度、交互易用性和可靠性。自然度评估系统与手部运动的一致性，交互易用性反映所需付出的努力，可靠性衡量响应一致性。比较以参与者为配对单位。

结果。在完整手部条件下，InsHaPE 的识别准确率高于 HandDAGT（见图 12(a)和 12(b)），这表明，当手部被完整观测时，其结构建模有利于下游手势识别。在模拟不完整条件下，HandDAGT 的性能大幅下降，而 InsHaPE 仍能保持较高的识别准确率（见图12(a)和 12(b)）。对各类手势取均值后，InsHaPE 提升了

在完整设置和模拟不完整设置下，识别准确率分别提高了 13.3 和 11.6 个百分点。我们仅将这些差异作为系统层面的度量进行描述，不进行参与者层面的配对推断；由于每个估计器都在各自的实时会话中运行，这些差异也包含不同表演之间的变化。这表明，更具鲁棒性且符合解剖学一致性的骨架估计有助于下游手势识别，尤其是在局部观测持续缺失时。主观评分（均值 ± 标准差）显示，InsHaPE 在三个维度上均具有明显优势（见图 12(c)）：自然度（6.06 ± 0.73 vs.4.33 ± 0.84）、交互易用性（6.11 ± 0.68 vs. 4.28 ± 0.83）和可靠性（5.83 ± 0.86 vs. 4.00 ± 1.08）。配对

表 XI 具有真实手指不完整情况的三名参与者的试点评估。在与受控研究相同的下游MR流程下，我们报告手势识别准确率（%）以及关于自然度、交互便捷性和可靠性的主观评分。

参与者	识别准确率		自然度		交互便捷性		可靠性	
	基线	我们的方法	基线	我们的方法	基线	我们的方法	基线	我们的方法
P1	48.0	**64.0**	5	**6**	5	**6**	4	**6**
P2	44.0	**60.0**	4	**6**	4	**6**	4	**5**
P3	40.0	**52.0**	4	**5**	4	**5**	3	**5**
Mean	44.0	**58.7**	4.33	**5.67**	4.33	**5.67**	3.67	**5.33**

均值差异（基于取整前的显示均值计算）分别为 1.72（95% 置信区间 [1.35, 2.10]）、1.83（95%置信区间 [1.48, 2.18]）和 1.83（95%置信区间 [1.37, 2.29]）；所有比较均达到显著水平，$p < .001$ 且配对效应量较大（$d_z = 2.29$、2.59 和 1.99）。综上，这些结果表明，在受控的持续缺失观测条件下，InsHaPE 在 IncNYU 上取得的鲁棒性提升能够转化为更可靠的下游混合现实交互。

B. 真实不完整手部试点研究

参与者与流程。为补充受控模拟并直接获取结构不完整性的证据，我们在北京市海淀区残疾人联合会的协助下，招募了三名手指不完整的男性参与者。参与者年龄为28–35岁（均值 = 32.0，标准差 = 3.6），其部分手指缺失的情况各不相同。所有参与者均在研究开始前签署了书面知情同意书，同意为研究目的收集和使用实验数据。我们将这项试点研究与受控遮挡用户研究分开报告，因为真实的解剖结构缺失与模拟缺失观测有本质区别。每位参与者均使用第 VII-A 节所述的混合现实流程和识别协议执行预定义手势。

评估协议。我们遵循与受控研究相同的下游评估协议，以便进行直接比较。由于真实不完整手部中的缺失关节不存在物理上的真实位置，我们不计算幻影关节误差。相反，我们在7点评分量表上，同时评估手势识别准确率、参与者报告的自然度、交互易用性以及可靠性

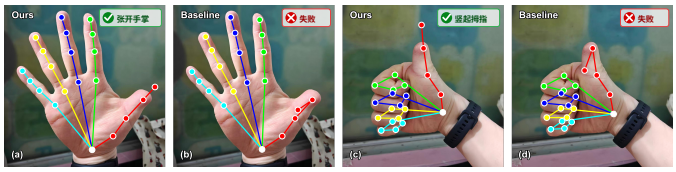


图 13：真实不完整手部试点研究中一位部分拇指截肢参与者的代表性结果。(a–b) 分别为 InsHaPE 和 HandDAGT 基线估计的张开手掌手势。(c–d) 分别为 InsHaPE 和 HandDAGT 基线估计的竖起拇指手势。在这些测试中，下游识别器使用 InsHaPE 时识别成功，使用基线时则失败。

。鉴于参与者队列规模较小，我们报告参与者级结果和描述性平均值，不进行推断性统计检验。

结果。 如表 XI所示，InsHaPE 持续提升了三位参与者的下游手势识别性能，将平均准确率从 44.0% 提高至 58.7%。主观评分也呈现出相同趋势：InsHaPE 在自然度（5.67 对 4.33）、交互易用性（5.67 对 4.33）和可靠性（5.33 对 3.67）方面均获得了更高的平均分。在图 13中，我们可视化了部分张开手掌和竖起拇指的试验结果，为参与者层面的结果提供了定性背景。这些一致的参与者层面改进为以下结论提供了初步证据：InsHaPE 生成的固定拓扑骨架在真实结构不完整性下仍可用于下游混合现实交互。总体而言，该试点研究通过具有实际手指不完整性的用户证据，为可复现的模拟实验提供了补充。

VIII. 结论与讨论

本研究重新审视了混合现实手部姿态估计系统中的一项假设：手部具有固定拓扑的完整手部结构。我们提出了 InsHaPE，这是一种包容性手部姿态估计框架，能够处理结构不完整性，同时保持标准设置下全关节手部的性能。在 NYU/IncNYU 上开展的大量实验以及在 MSRA 上进行的额外验证表明，该方法对结构不完整性具有较强的鲁棒性，在完整手部上具有具有竞争力的性能，并且能够泛化到主要基准之外。这些结果表明，对结构不完整性进行显式建模，可以提升固定拓扑手部姿态估计对缺失结构的鲁棒性。除基准评估外，原型混合现实系统和用户研究进一步表明，InsHaPE 能够为下游交互提供更加鲁棒且拓扑一致的骨架表示，在无需修改现有流程的情况下，提升手势识别性能和用户感知的交互质量。本研究强调了以手为中心的感知中的结构鲁棒性，并为在不完整手部构型下实现可靠的混合现实交互提供了切实可行的一步。

尽管取得了这些进展，当前方法仍存在若干局限。拓扑感知蒸馏在训练期间依赖成对的完整/不完整观测；这类配对数据可以为 IncNYU 生成，但通常无法为存在真实解剖学缺失的用户获取。因此，当前的真实用户试点研究仅提供下游证据，并未消除对这种监督的依赖。本研究还聚焦于结构不完整性，尚未应对运动不稳定性或长期疲劳等其他挑战因素。该模型的参数也多于轻量级基线，且缺少在边缘设备或可穿戴设备上的端到端评估。尽管 IncNYU 提供了一个受控且符合生理约束的基准，但它源自 NYU，无法涵盖真实手部解剖结构、感知条件和交互行为的全部多样性。未来工作将探讨弱监督适应、更广泛的手部变异、多样化的真实世界数据，以及在实际场景中实现高效可穿戴部署。

参考文献

[1] T. Ohkawa、R. Furuta 和 Y. Sato，“三维手部姿态估计的高效标注与学习：综述”，国际计算机视觉杂志，第 131 卷，第 12 期，第 3193–3206 页，2023年。[2] C. Wan、T. Probst、L. Van Gool 和 A. Yao，“用于手部姿态估计的稠密三维回归”，载于 *IEEE*计算机视觉与模式识别会议，第 5147–5156 页，2018 年。[3] G. Moon、H. Choi 和 K. M. Lee，“用于全身三维人体网格估计的精确三维手部姿态估计”，载于 *IEEE/CVF*计算机视觉与模式识别会议研讨会，第 2308 2317 页，2022 – 。[4] S. Salter、R. Warren、C. Schlager、A. Spurr、S. Han、R. Bhasin、Y. Cai、P. Walkington、A. Bolarinwa、R. J. Wang、等，“emg2pose：用于表面肌电手部姿态估计的大规模多样化基准”，神经信息处理系统进展，第 37 卷，第 55703– 55728 页，2024年。[5] L. Zhao、X. Lu、R. Fan、S. K. Im 和 L. Wang，“GaussianHand：用于手部虚拟形象动画的实时三维高斯渲染”，*IEEE*《可视化与计算机图形学汇刊》，第 31 卷，第 9 期，第 6484–6496 页，2025年。[6] Y. Shi、L. Zhao、X. Lu、T. Hoang 和 M. Wang，“在虚拟现实环境中使用虚拟手抓取三维物体”，载于 *ACM SIGGRAPH*虚拟现实连续体及其工业应用国际会议，第 1–8 页，2022 年。[7] K. Zhou、R. Cai、L. Wang、H. P. H. Shum 和 X. Liang，“动作质量评估的综合综述：方法与基准”，模式识别，第 179 卷，文章 113933，2026年。[8] K. Zhou、Z. Cheng、H. P. Shum、F. W. Li 和 X. Liang，“STGAE：用于手部运动去噪的时空图自编码器”，载于*IEEE*混合与增强现实国际研讨会，第 41–49 页，2021 年。[9] W. Cheng、E. Kim 和 J. H. Ko，“HandDAGT：用于三维手部姿态估计的去噪自适应图变换器”，载于 欧洲计算机视觉会议，第 15146 卷，第 35–52 页，2024年。[10] P. Ren、Y. Chen、J. Hao、H. Sun、Q. Qi、J. Wang 和 J. Liao，“两个头胜过一个头：用于基于深度的三维手部姿态估计的图像-点云网络”，*AAAI* 人工智能会议，第 37 卷，第 2 期，第 2163–2171 页，2023 年。[11] W. Cheng、H. Tang、L. Van Gool 和 J. H. Ko，“HandDiff：基于图像-点云扩散的三维手部姿态估计”，载于 *IEEE/CVF*计算机视觉与模式识别会议，第 2274– 2284 页，2024年。[12] C. Jiang、Y. Xiao、C. Wu、M. Zhang、J. Zheng、Z. Cao 和 J. T. Zhou，“A2J-Transformer：用于从单幅 RGB图像进行三维交互手部姿态估计的锚点到关节变换器网络”，载于 *IEEE/CVF*计算机视觉与模式识别会议，第 8846– 8855 页，2023 年。[13] G. B. Mo、J. J. Dudley 和 P. O. Kristensson，“Gesture knitter：面向头戴式混合现实应用的手势设计工具”，载于 *CHI*计算系统中的人为因素会议，第 1–13 页，2021 年。[14] L. Zhao、X. Lu、Q. Bao 和 M. Wang，“通过长时记忆增强网络进行原地手势分类”，载于 *IEEE*混合与增强现实国际研讨会，第 224–233 页，2022 年。[15] K. Zhou、C. Chen、Y. Ma、Z. Leng、H. P. Shum、F. W. Li 和 X. Liang，“用于模拟微重力环境中手物交互的混合现实训练系统”，载于*IEEE*混合与增强现实国际研讨会，第 167–176 页，2023 年。[16] T. Ohkawa、K. He、F. Sener、T. Hodan、L. Tran 和 C. Keskin，“AssemblyHands：通过三维手部姿态估计实现自我中心视角活动理解”，载于 *IEEE/CVF*计算机视觉与模式识别会议，第 12999–13008 页，2023 年。

[17] L. Li、L. Tian、X. Zhang、Q. Wang、B. Zhang、L. Bo、M. Liu 和 C. Chen，“RenderIH：用于三维交互手部姿态估计的大规模合成数据集”，载于 IEEE/CVF国际计算机视觉会议，第 20395–20405 页，2023 年。[18] K. Zhou、H. P. Shum、F. W. Li 和 X. Liang，“用于手部运动去噪的多任务时空图自编码器”，IEEE《可视化与计算机图形学汇刊》，第 30 卷，第 10 期，第 6754– 6769 页，2024年。[19] Q. Ye 和 T.-K. Kim，“使用分层混合密度网络的遮挡感知手部姿态估计”，载于 欧洲计算机视觉会议，第 11214 卷，第 817–834 页，2018 年。[20] H. Meng、S. Jin、W. Liu、C. Qian、M. Lin、W. Ouyang 和 P. Luo，“通过手部去遮挡与移除实现三维交互手部姿态估计”，载于 欧洲计算机视觉会议，第 13666 卷，397，2022 380页。– 。[21] Q. Liu、Y. Zhang、S. Bai 和 A. Yuille，“多人三维人体姿态估计的显式遮挡推理”，载于 欧洲计算机视觉会议，第 13665 卷，第 497–517 页，2022 年。[22] J. Park、Y. Oh、G. Moon、H. Choi 和 K. M. Lee，“HandOccNet：遮挡鲁棒三维手部网格估计网络”，载于 IEEE/CVF计算机视觉与模式识别会议，第 1496– 1505 页，2022 年。[23] X. Sun、Y. Wei、S. Liang、X. Tang 和 J. Sun，“级联手部姿态回归”，载于 IEEE计算机视觉与模式识别会议，第 824–832 页，2015 年。[24] X. Deng、D. Zuo、Y. Zhang、Z. Cui、J. Cheng、P. Tan、L. Chang、M. Pollefeys、S. Fanello 和 H. Wang，“使用级联姿态引导三维对齐的循环三维手部姿态估计”，IEEE《模式分析与机器智能汇刊》，第 45 卷，第 1 期，第 932–945 页，2023 年。[25] Z. Leng、J. Chen、H. P. Shum、F. W. Li 和 X. Liang，“基于图神经网络的震颤条件下稳定手部姿态估计”，载于 IEEE虚拟现实与三维用户界面会议，第 226–234 页，2021 年。[26] J. Tompson、M. Stein、Y. LeCun 和 K. Perlin，“使用卷积网络实时连续恢复人手姿态”，ACM《图形学汇刊》，第 33 卷，第 5 期，第 1–10 页，2014 年。[27] K. Du、X. Lin、Y. Sun 和 X. Ma，“CrossInfoNet：基于多任务信息共享的手部姿态估计”，载于 IEEE/CVF计算机视觉与模式识别会议，第 9896–9905 页，2019 年。[28] L. Fang、X. Liu、L. Liu、H. Xu 和 W. Kang，“JGR-P2O：基于联合图推理的像素到偏移预测网络，用于从单幅深度图像进行三维手部姿态估计”，载于 欧洲计算机视觉会议，第 12351 卷，第 120–137 页，2020 年。[29] L. Ge、H. Liang、J. Yuan 和 D. Thalmann，“ 单幅深度图像中的鲁棒三维手部姿态估计：从单视图 CNN 到多视图 CNN”，载于 IEEE计算机视觉与模式识别会议，第 3593–3601 页，2016 年。[30] P. Ren、H. Sun、J. Hao、Q. Qi、J. Wang 和 J. Liao，“姿态引导的分层图推理：从单幅深度图像进行三维手部姿态估计”，IEEE《控制论汇刊》，第 53 卷，第 1 期，第 315– 328 页，2023 年。[31] L. Ge、H. Liang、J. Yuan 和 D. Thalmann，“用于从单幅深度图像高效、鲁棒地进行手部姿态估计的三维卷积神经网络”，载于 IEEE 计算机视觉与模式识别会议，第 1991–2000 页，2017 年。[32] G. Moon、J. Y. Chang 和 K. M. Lee，“V2V-PoseNet：用于从单幅深度图进行精确三维手部与人体姿态估计的体素到体素预测网络”，载于 IEEE 计算机视觉与模式识别会议，第 5079–5088 页，2018 年。[33] C. R. Qi、H. Su、K. Mo 和 L. J. Guibas，“PointNet：基于点集的三维分类与分割深度学习”，载于 IEEE计算机视觉与模式识别会议，第 652–660 页，2017 年。[34] L. Ge、Y. Cai、J. Weng 和 J. Yuan，“Hand PointNet：使用点集进行三维手部姿态估计”，载于 IEEE计算机视觉与模式识别会议，第 8417–8426 页，2018 年。[35] Q. Lin、L. Yang 和 A. Yao，“具有双模态的跨域三维手部姿态估计”，载于 IEEE/CVF计算机视觉与模式识别会议，第 17184–17193 页，2023 年。[36] Y. Ge、C. Xu 和 L. Cheng，“用于歧义手部姿态估计的由粗到细的多假设方法”，IEEE《图像处理汇刊》，第 34 卷，第 4302–4314 页，2025年。[37] J. Ho、A. Jain 和 P. Abbeel，“去噪扩散概率模型”，神经信息处理系统进展，第 33 卷，第 6840– 6851 页，2020 年。[38] J. Sohl-Dickstein、E. Weiss、N. Maheswaranathan 和 S. Ganguli，“使用非平衡热力学的深度无监督学习”，载于 国际机器学习会议，第 37 卷，机器学习研究论文集

，第 2256–2265 页，2015 年。

[39] M. Ivashechkin、O. Mendez 和 R. Bowden，《基于图像的三维手部姿态估计中的去噪扩散》，载于 IEEE/CVF国际计算机视觉会议研讨会，第3136–3145页，2023年。[40] S. Zhang、T. Zheng、Z. Chen、J. Hu、A. Khamis、J. Liu 和 J. Luo，《OCHID-Fi：通过射频视觉实现遮挡鲁棒的三维手部姿态估计》，载于 IEEE/CVF国际计算机视觉会议，第15112–15121页，2023年。[41] M. Garg、P. M. Pradhan 和 D. Ghosh，《OccRobNet：用于准确三维交互式手物姿态估计的遮挡鲁棒网络》，载于全国通信会议，第1–6页，2025年。[42] H. Yang、W. Sun、J. Liu、J. Zheng、J. Xiao 和 A. Mian，《利用掩码自编码器进行遮挡条件下的三维手物姿态估计》，IEEE多媒体汇刊，第1–11页，2026年。[43] Y. Cai、H. Pan、J. Yang、Y. Liu、Q. Gao 和 X. Wang，《利用层次化特征解耦进行遮挡条件下的几何感知三维手物姿态估计》，《电子学》，第14卷，第5期，文章编号1029，2025年。[44] T. Woo、W. Park、W. Jeong 和 J. Park，《手物交互图像中手部姿态估计的深度学习方法与数据集综述》，《计算机与图形学》，第116卷，第474–490页，2023年。[45] J. Dudley、L. Yin、V. Garaj 和 P. O. Kristensson，《包容式沉浸：改善虚拟现实、增强现实和元宇宙可访问性的工作综述》，Virtual Reality，第27卷，2023年，第2989–3020页，第4期。[46] C. Creed、M. Al-Kalbani、A. Theil、S. Sarcar 和 I. Williams，《包容性增强现实与虚拟现实：研究议程》，《国际人机交互杂志》，第40卷，第20期，第6200–6219页，2024年。[47] C. Creed、M. Al-Kalbani、A. Theil、S. Sarcar 和 I. Williams，《包容性AR/VR：沉浸式技术的可访问性障碍》，《信息社会中的普遍访问》，第23卷，第1期，第59–73页，2024年。[48] P. Biswas、P. Orero、M. Swaminathan、K. Krishnaswamy 和 P. Robinson，《自适应可访问AR/VR系统》，载于CHI计算系统中的人为因素会议，第1–7页，2021年。[49] K. Krug、J. M´endez、W. Luo 和 R. Dachselt，《别把我排除在外：为混合现实协作设计设备包容性》，arXiv预印本 arXiv:2409.05374，2024年。[50] J. Kim、J.-H. Ahn 和 Y. Kim，《面向包容性虚拟现实导航的沉浸式交互：提升社会弱势用户的可访问性》，《电子学》，第14卷，第5期，文章编号1046，2025年。[51] A. Othman、K. Chemnad、A. E. Hassanien、A. Tlili、C. Y. Zhang、D. Al-Thani、F. Altınay、H. Chalghoumi、H. S. Al-Khalifa、M. Obeid 等，《无障碍元宇宙：元宇宙可访问性与包容性的理论框架》，《多模态技术与交互》，第8卷，第3期，文章编号21，2024年。[52] P. Radanliev、D. De Roure、P. Novitzky 和 I. Sluganovic，《元宇宙中面向残障用户和内容创作者的新型信息与通信技术的可访问性与包容性》，《残疾与康复：辅助技术》，第19卷，2024年，第1849–1863页，第5期。[53] W. Cheng 和 J. H. Ko，《HandR2N2：使用残差循环神经网络进行迭代式三维手部姿态估计》，载于 IEEE/CVF国际计算机视觉会议，第20904–20913页，2023年。[54] H. Choi、G. Moon、J. Park 和 K. M. Lee，《学习从野外拥挤场景中估计鲁棒的三维人体网格》，载于 IEEE/CVF计算机视觉与模式识别会议，第1475–1484页，2022年。[55] M. Everingham 和 J. Winn，《PASCAL视觉目标类别挑战赛2011 (VOC2011) 开发工具包》。技术报告，2011年。[56] M. Rezaei、R. Rastgoo 和 V. Athitsos，《TriHorn-Net：一种用于准确的基于深度的三维手部姿态估计的模型》，《专家系统与应用》，第223卷，文章编号119922，2023年。[57] A. Kawaiah、M. Thakur、S. Garg、S. H. Kawasmi 和 A. Hassan，《指尖损伤与截肢：文献综述》，《Cureus》，第12卷，第5期，文章编号e8291，2020年。[58] A. Armagan、G. Garcia-Hernando、S. Baek、S. Hampali、M. Rad、Z. Zhang、S. Xie、M. Chen、B. Zhang、F. Xiong 等，《评估三维手部姿态估计在手物交互场景下对未见视角、关节构型、形状和物体的泛化能力》，载于欧洲计算机视觉会议，第12368卷，第85–101页，2020年。[59] W. Park、D. Kim、Y. Lu 和 M. Cho，《关系知识蒸馏》，载于 IEEE/CVF计算机视觉与模式识别会议，第3967–3976页，2019年。[60] H. Guo、G. Wang、X. Chen、C. Zhang、F. Qiao 和 H. Yang，《区域集成网络：深度三维手部姿态估计的良好实践》，《视觉通信与图像表示杂志》，第55卷，第404–414页，2018年。

[61] M. Oberweger 和 V. Lepetit，《DeepPrior++：提升快速且准确的三维手部姿态估计》，载于 *IEEE* 国际计算机视觉会议研讨会，第585–594页，2017年。[62] X. Chen、G. Wang、H. Guo 和 C. Zhang，《用于级联手部姿态估计的姿态引导结构化区域集成网络》，神经计算，第395卷，第138–149页，2020年。[63] F. Xiong、B. Zhang、Y. Xiao、Z. Cao、T. Yu、J. T. Zhou 和 J. Yuan，《A2J：用于从单幅深度图像估计三维关节构型姿态的锚点到关节回归网络》，载于 *IEEE/CVF*国际计算机视觉会议，第793–802页，2019年。[64] W. Huang、P. Ren、J. Wang、Q. Qi 和 H. Sun，《AWR：用于三维手部姿态估计的自适应加权回归》，*AAAI* 人工智能会议，第34卷，第07期，第11061–11068页，2020年。[65] P. Ren、H. Sun、W. Huang、J. Hao、D. Cheng、Q. Qi、J. Wang 和 J. Liao，《用于实时三维手部姿态估计的空间感知堆叠回归网络》，神经计算，第437卷，第42–57页，2021年。[66] L. Huang、J. Tan、J. Liu 和 J. Yuan，《Hand-Transformer：用于三维手部姿态估计的非自回归结构化建模》，载于 欧洲计算机视觉会议，第12370卷，第17–33页，2020年。[67] L. Ge、Z. Ren 和 J. Yuan，《用于三维手部姿态估计的点对点回归 PointNet》，载于 欧洲计算机视觉会议，第11217卷，第489–505页，2018年。[68] W. Cheng、J. H. Park 和 J. H. Ko，《HandFoldingNet：一种通过二维手部骨架的多尺度特征引导折叠进行三维手部姿态估计的网络》，载于 *IEEE/CVF* 国际计算机视觉会议，第11260–11269页，2021年。[69] H. Li、Z. Xu、G. Taylor、C. Studer 和 T. Goldstein，《神经网络损失景观的可视化》，神经信息处理系统进展，第31卷，第6391–6401页，2018年。



Kanglei Zhou 现任清华大学心理与认知科学系博士后研究员。他于2025年获北京航空航天大学计算机科学与工程博士学位，并于2020年获河南师范大学工学学士学位。2024年，他曾在杜伦大学担任访问学生。他的研究兴趣包括人体运动分析和持续学习，相关论文发表于 TPAMI、国际计算机视觉期刊、IEEE 图像处理汇刊、IEEE 可视化与计算机图形学汇刊、国际机器学习会议、ICLR、CVPR、ECCV 和 IEEE 国际混合与增强现实研讨会等顶级学术会议和期刊。



Zikai Hao 获太原理工大学工学学士学位。目前，他正在北京航空航天大学同一所学校攻读软件工程硕士学位。他目前从事人体运动分析和手部姿态估计研究。



Liyuan Wang 现任清华大学心理与认知科学系助理教授。他在清华大学获得学士和博士学位，并在那里开展博士后研究。他具有神经科学和机器学习的跨学科背景。他关于持续学习的研究成果发表于相关领域的主要会议和期刊，如《自然·机器智能》、IEEE 模式分析与机器智能汇刊、TNNLS、NeurIPS、ICLR、CVPR、ICCV、ECCV 等。



Xiaohui Liang 在中国北京航空航天大学获得计算机科学与工程博士学位。目前，他是北京航空航天大学计算机科学与工程学院的教授。其主要研究兴趣包括计算机图形学与动画、可视化和虚拟现实。