中图法分类号： **文献标识码：** **文章编号：** 1006-8961(XXXX)XX-0001-13

论文引用格式： ZHOU Min, AN Ping, HUANG Xinpeng, YANG Chao. XXXX. SMPL-Guided Human Mesh Reconstruction from Incomplete Point Clouds. Journal of Image and Graphics, XX(XX):0001-0013(周敏, 安平, 黄新彭, 杨超. XXXX. SMPL引导的残缺点云人体重建. 中国图象图形学报, XX(XX):0001-0013)[DOI:10.11834/jig.250262]

SMPL引导的残缺点云人体重建

周敏，安平*，黄新彭，杨超

上海大学 通信与信息工程学院，上海 200444

摘 要：目的 点云在人体建模和姿态估计中应用广泛，但常因传感器视角和遮挡而局部缺失，影响重建完整性和精度。点云非结构化特性使模型难以准确建模几何结构和拓扑关系，导致失真、模糊或偏差。现有方法在处理严重缺失和复杂姿态时，仍面临结构完整性和几何精度不足的挑战。**方法** 为此，本文提出一种基于蒙皮多人线性(skinned multi-person linear, SMPL)模型拓扑先验的人体重建与点云补全方法。通过引入逐点语义感知的多尺度特征提取机制，增强对局部几何结构的理解能力；在此基础上，利用SMPL的参数化网格先验构建初始人体形变骨架，有效引导缺失区域的结构推理，缓解因信息缺失带来的拓扑歧义；进一步设计顶点偏移感知优化模块，通过学习输入点云与预测网格间的空间残差，实现对人体表面细节的精细化修正。**结果** 实验结果表明，所提方法在SURREAL和AMASS数据集上均优于现有方法。在SURREAL数据集上，顶点回归误差降低4.6%，姿态估计精度达当前最优的16.2 mm；在AMASS数据集上的倒角距离指标也显著优于现有点云补全算法，进一步验证了所提出框架的良好泛化性与鲁棒性。**结论** 综上所述，将SMPL的拓扑先验有效融入点云重建流程，可显著提升重建质量，尤其在处理非结构化与不完整点云时展现出更强优势。

关键词： 点云重建；形状补全；姿态估计；多尺度特征；SMPL拓扑先验

SMPL-Guided Human Mesh Reconstruction from Incomplete Point Clouds

ZHOU Min, AN Ping, HUANG Xinpeng, YANG Chao

School of Communication and Information Engineering, Shanghai University, Shanghai, 200444, China

Abstract: Objective Point clouds data has become a fundamental modality for human body modeling and pose estimation due to their explicit and direct representation of 3D geometric structure. Unlike image-based or volumetric methods, point clouds preserve raw geometric information without discretization or projective distortion, making them especially suitable for capturing detailed surface characteristics. However, in practical acquisition scenarios—such as those using consumer-grade LiDAR sensors or depth cameras—point clouds often suffer from significant local missing regions caused by sensor viewpoint limitations, object self-occlusions, and environmental obstructions. These incompleteness issues severely degrade the completeness and accuracy of reconstruction, posing substantial challenges for downstream tasks like animation, virtual try-on, or motion analysis. Moreover, as unstructured data with irregular sampling, point clouds lack explicit connectivity information and exhibit permutation invariance. This inherent nature complicates the learning of continuous

收稿日期： 2025-06-09；**修回日期：** 2026-01-22

* **通信作者：** 安平，通信作者，女，上海大学通信与信息工程学院教授，博士，主要研究方向为图像处理、视频编码、视觉计算。E-mail: anping@shu.edu.cn；安平，通信作者，女，现任上海大学教授，主要研究方向为图像处理、视频编码、视觉计算。E-mail: anping@shu.edu.cn。

基金项目： 国家自然科学基金(No. 62471284, 62020106011, No. 62371278, No. 62371279)；上海自然科学基金(No. 22ZR1424300)

Supported by: Project supported by the National Natural Science Foundation of China (No. 62471284, 62020106011, No. 62371278, No. 62371279); Shanghai Natural Science Foundation (No. 22ZR1424300)

surface geometry and topological connectivity of the human body, which is essential for recovering plausible shapes and joint kinematics. Consequently, many existing approaches produce results with structural distortions, loss of fine-grained details, or inaccurate pose predictions—especially around articulated regions such as elbows, knees, and shoulders. Although recent deep learning models have made progress in processing point sets, they still struggle to maintain structural coherence and geometric fidelity when handling highly incomplete inputs under complex articulations. There remains a clear need for methods that can infer missing geometry with strong contextual awareness and robust priors on human body shape and mobility. **Method** To address these issues, we propose a novel method for human body reconstruction and point cloud completion by leveraging the topological prior of the SMPL (Skinned Multi-Person Linear) model. Our approach begins with a multi-scale feature extraction module that incorporates per-point semantic awareness, enabling the network to capture both fine-grained details and broader contextual information. This module employs a combination of local geometric feature learners and hierarchical aggregation layers to enhance the understanding of structural patterns at various scales. By integrating point-wise semantics during the feature learning process, it effectively encodes biological constraints into the representation, thereby generating more discriminative and context-aware point cloud features that better preserve the natural articulation and proportionality of human body shapes. Building upon this enriched feature representation, the parametric mesh topology of the SMPL model is integrated to guide the prediction of an initial deformable human skeleton. This step effectively regularizes the structural reasoning in missing or incomplete regions by introducing strong anatomical constraints, significantly alleviating topological ambiguities caused by severe data incompleteness such as occlusions or large-scale scanning artifacts. The parametric model acts as a structural template that ensures biomechanically consistent predictions, even when large portions of the input are absent. As a result, the proposed framework maintains robust performance under challenging conditions, producing plausible human body proportions and pose-dependent shape variations that align with realistic anatomical properties. Furthermore, we design a vertex offset-aware refinement module that learns spatial residuals between the input sparse point cloud and the vertices of the predicted mesh through fine-grained alignment. This module operates in a hierarchical manner, progressively correcting geometric discrepancies across multiple resolution levels. It utilizes a cascaded residual learner that first handles large-scale deformations and then gradually focuses on recovering subtle surface details—such as muscle curves, soft-tissue contours, and joint depressions—that are often lost in coarse reconstruction stages. By iteratively aligning the predicted mesh with the observed points, our method achieves precise reconstruction of detailed surface geometry with significantly improved geometric fidelity. The final output is a high-fidelity, and anatomically plausible human body model, even when reconstructed from highly incomplete or noisy input data. **Results** Experimental results demonstrate that the proposed method consistently outperforms existing approaches across both the SURREAL and AMASS datasets, highlighting its effectiveness and broad applicability. On the SURREAL dataset, the method reduces the vertex-to-vertex reconstruction error by 4.6%, setting a new state-of-the-art with a remarkable pose estimation accuracy of 16.2 mm. This improvement not only reflects enhanced precision in capturing complex human motions but also indicates stronger alignment with ground-truth data. Similarly, on the AMASS dataset, the approach achieves a significantly lower Chamfer Distance compared to current point cloud completion techniques, underscoring its ability to generate more geometrically consistent and structurally accurate reconstructions. These outcomes across diverse datasets further validate the strong generalization capacity and robustness of the proposed framework, suggesting its potential for real-world applications where reliability under varying conditions is critical. **Conclusion** In conclusion, this work demonstrates that the effective integration of SMPL's topological prior into the point cloud reconstruction pipeline leads to significant improvements in reconstruction quality. This approach proves especially advantageous when dealing with unstructured and highly incomplete input data, where traditional methods often struggle to produce plausible results. By leveraging the strong semantic and structural constraints provided by the SMPL model, our method enhances detail recovery, ensures better shape consistency, and reduces artifacts in challenging scenarios. These findings highlight the value of incorporating semantic priors in geometric reconstruction tasks and suggest promising directions for future research in robust 3D understanding and generation.

Key words: point clouds reconstruction; shape completion; pose estimation; Multi-scale features; SMPL topological prior

公式章 1 节 1 **中图法分类号**：TP391 **文献标识码**：A **文章编号**：1006-8961(年)

0 引 言

人体重建技术旨在从传感器捕获的数据中生成三维数字人体，具有广泛的应用前景，对多个行业的发展具有深远影响。该技术在娱乐行业中发挥着重要作用，尤其在运动捕捉、电影制作、增强现实、虚拟现实和混合现实等领域；另一方面，它在机器人领域也具有关键作用，支持行为理解，广泛应用于社交机器人、辅助机器人和人机交互等场景（Ren 等，2024）。近年来，基于2D图像估计与恢复参数化人体模型，如SMPL（Loper等，2023）的技术取得了显著进展。然而，2D图片缺乏深度信息，与3D输入方法相比重建效果存在较大差距，并由于拍摄实际画面可能引发隐私泄露问题。为此，随着深度传感器的不断进步和大规模数据集的应用，基于3D点云的人体重建技术逐渐受到关注。该技术能够克服2D图像作为输入的局限性和潜在的隐私风险，提供更为安全、精确的三维人体表示。

然而，通过激光扫描、深度相机等技术获取的人体点云数据，常存在不完整性、噪声和稀疏性等问题，制约了基于点云的人体重建技术的精度与发展。实际应用中，受传感器限制、人体自遮挡及环境遮挡影响，点云数据往往残缺，增加了重建难度。传统网格重建方法如泊松重建和德劳内三角化对完整点云效果较好，但难以处理残缺点云。尽管深度学习方法在点云处理领域进展显著，但在复杂人体姿态和细节重建上仍有局限。因此，研究基于残缺点云的人体网格重建技术具有重要的理论与应用价值。

近年来，基于点云输入的人体网格重建技术取得了显著进展。由于SMPL模型以其紧凑的参数化表达和良好的几何先验被广泛采用，大多数现有方法以SMPL为输出表示形式。然而，SMPL本身作为一种通用人体模板，其拓扑固定、关节驱动机制标准化，若仅将其作为最终输出目标而不改变建模方式，则容易导致方法创新性受限，难以体现针对特定输入模态（如残缺点云）的深层适配能力。因此，如何在利用SMPL拓扑先验的基础上，设计更具表现力和鲁棒性的重建架构，成为当前研究的关键挑战。目前的主流方法主要分为两类：一类是基于SMPL参数的回归方法，即通过神经网络预测姿态（pose）和形状（shape）参数，再经SMPL模型前向渲染生成三维网格；另一类则是直接顶点回归方法，跳过参数空间，直接输出SMPL网格的顶点坐标。尽管两者都使用SMPL的拓扑结构作为输出网格的连接关系，但其建模思路存在本质差异。

在点云重建领域，现有方法大多依赖于参数回归方式。例如，D-Pose（Vasilikopoulos等，2024）提出了一种关节感知的网络结构，能够从完整点云中估计SMPL参数，但在面对遮挡或截断点云时性能显著下降。VoteHMR（Jiang等，2019）引入了遮挡感知的投票网络，尝试从单帧残缺点云中估计完整人体的SMPL参数，然而该方法未能充分利用点云的空间结构信息，导致顶点精度不足；PointHPS（Cai等，2023）采用多阶段级联特征迭代优化策略，在稠密点云上取得了较好的重建效果，但其性能严重依赖高质量的输入点云，对噪声和稀疏性敏感。LiveHPS（Ren等，2024）以多帧点云作为输入，通过融合多帧时序信息并引入自适应蒸馏机制提升稳定性，但仍沿用传统的参数回归框架，推理过程需经过SMPL重参化，限制了效率与端到端优化潜力。此外，姚砺等人（姚砺等，2021）提出引入关节轴角先验作为损失约束，以抑制非生理性旋转、提高姿态合理性，虽在一定程度上缓解了异常动作问题，但仍建立在低维参数回归的基础之上，无法避免由全局参数化表示带来的几何细节丢失。上述方法普遍存在一个共性问题：它们将高维点云编码为低维SMPL参数，本质上是一种“压缩-解码”模式。这种间接建模方式割裂了输入点云与输出顶点之间的空间对应关系，使得网络难以建立细粒度的几何映射，且对特征抽象能力和全局上下文建模要求较高。此外，SMPL反向蒙皮过程计算开销大，不利于实时应用。

相比之下，顶点回归策略在图像输入任务中已有较多探索，并展现出良好潜力。例如，METRO（Lin等，2021）首次将Transformer用于人体网格重建，直接从图像特征回归SMPL顶点坐标，实现了无需显式参数估计的端到端预测。PostoMETRO（Yang等，2025）进一步引入姿势标记机制，将2D特征转化为紧凑的姿势序列输入，提升了模型对姿态语义的理解能力。PointHMR（Kim等，2023）则通过对图像特征进行关键点采用，并结合渐进式注意力掩码策略，在遮挡条件下实现了更精确的顶点定位。这类

方法的核心优势在于：直接建立输入观测与输出顶点间的映射路径，增强了模型的可解释性和几何保真度。

受上述图像输入下顶点回归方法的启发，本文提出一种面向残缺点云的端到端顶点回归框架，旨在突破传统参数回归范式的局限。与已有方法不同，我们不通过多层感知机（Multilayer Perceptron，MLP）回归SMPL参数，而是直接预测SMPL网格的顶点坐标，充分利用点云与目标网格在三维空间中的结构一致性，构建更加紧密的输入-输出关联。该策略不仅避免了复杂的后处理流程（如SMPL蒙皮），还显著提升了对遮挡、稀疏和非均匀采样点云的鲁棒性。更重要的是，我们的方法并非简单套用SMPL模板，而是在其拓扑约束下实现几何驱动的端到端重建，从而在保持标准输出格式的同时，增强了模型对局部细节的刻画能力。

针对现有方法在处理残缺点云时存在的结构感知能力弱与参数估计过程中的信息丢失问题，本文从模型结构设计与先验知识融合两个角度出发，提出三项关键技术改进：首先，引入逐点语义标签增强特征提取过程中对人体局部几何与整体姿态的理解能力；其次，结合SMPL模型的顶点拓扑关系，构建非局部的顶点-关节与顶点-顶点交互关系，实现从残缺点云到完整人体顶点的映射；最后，设计坐标细化模块，通过反馈机制对初步预测的顶点进行优化以提升重建精度。

本文方法整体采用编码器-解码器架构。其中，编码器基于多尺度特征提取机制，将输入残缺人体点云映射至高维特征空间，并融合语义信息以增强结构感知能力；解码器则依托SMPL拓扑先验，逐步回归人体网格顶点与关节坐标。解码器中，多层解码网络通过在Transformer结构中引入SMPL的邻接拓扑关系作为注意力掩码，逐渐构建由粗到细的顶点交互关系。最终通过偏移感知优化模块，对初步回归结果进行精细化校正，有效缓解输入不完整带来的几何失真问题，显著提升重建质量。

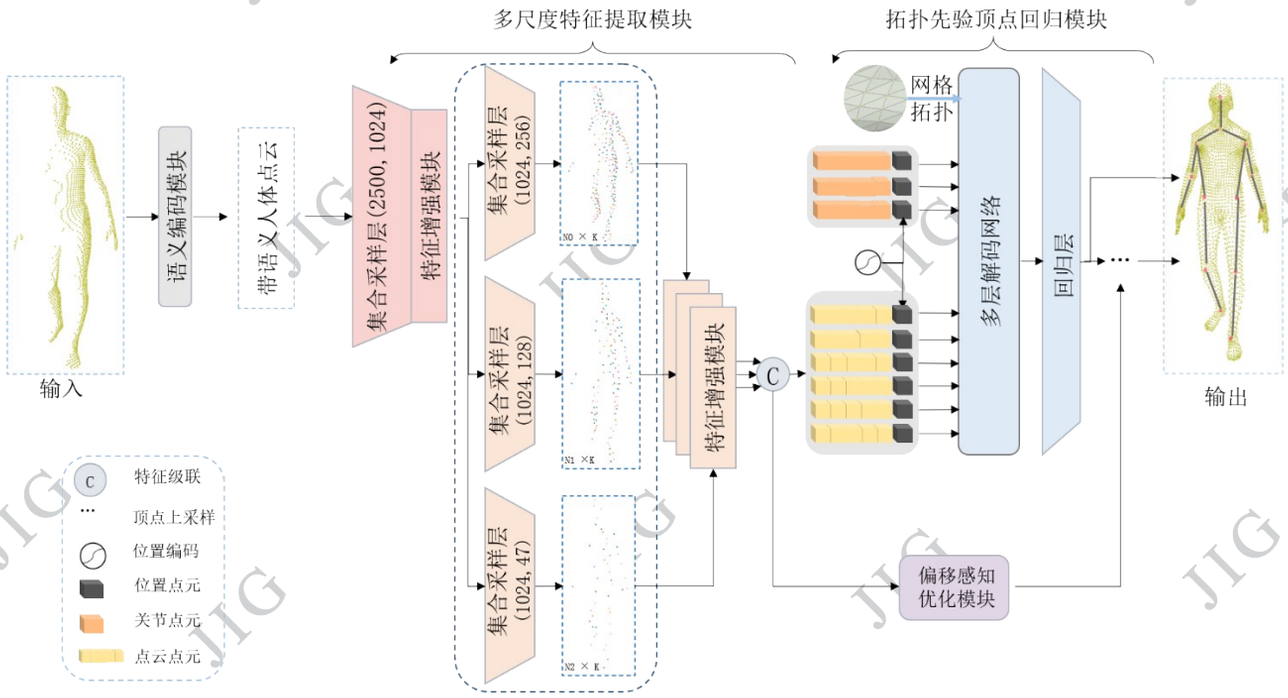


图1　基于SMPL拓扑先验的点云人体重建算法流程图

Figure 1　Flowchart of Point Clouds Human Reconstruction Algorithm with SMPL Topology

1 基于SMPL拓扑先验的人体重建

1.1 整体结构

为了从复杂姿态下的残缺人体点云数据中获得完整的人体点云表示，本文构建了一种基于SMPL拓扑先验的人体网格重建算法。该方法以残缺点云为输入，结合SMPL网格顶点的拓扑关系，构建从无序点云到有序人体网格的映射流程。

如图1所示，整体框架由多尺度特征提取模块和顶点回归模块组成。具体而言，在多尺度特征提取模块中本文方法首先在输入点云中融合逐点语义标签信息，提升网络对人体局部与全局结构的理解能力。随后，点云输入多尺度编码器，依次经过集合采样层（set abstraction，SA）和特征增强模块（feature enhance module，FEM），提取初始逐点特征，并通过下采样操作构建多尺度特征金字塔。其中，低分辨率层级用于捕获全局姿态信息，高分辨率层则保留局部几何细节，确保在遮挡或稀疏输入条件下仍能维持良好的结构感知能力。所提取的多尺度逐点特征作为表面点元，与关节点元及位置点元拼接后，共同输入至顶点回归模块。顶点回归模块基于Transformer架构，利用顶点-顶点邻接矩阵建立非局部映射关系，逐步回归SMPL模型的顶点坐标，实现从无序点云到有序网格的转换。最后，为进一步提升重建精度，引入了偏移感知优化模块（offset-aware refine module，ORM），通过反馈机制动态学习预测顶点与理想位置之间的偏差分布，并迭代修正关节与顶点坐标，从而有效缓解因输入不完整造成的几何失真问题。

1.2 多尺度特征提取模块

从输入的残缺点云中提取具有丰富几何结构信息的特征表示，对于后续顶点回归过程中实现完整人体重建和准确姿态估计至关重要，如参数化人体重建方法（李子清，2024）利用姿态估计模型生成的伪标签来区分人体的可见部位与被遮挡区域提升模型性能。对于点云任务，点云语义标签更能描述人体的结构信息。为了增强模型对人体局部细节与全局结构的理解能力，本文在点云编码之前引入了语义编码模块。通过融合逐点语义标签提升特征表达的准确性。具体而言，初始语义标签 $\boldsymbol{F}_i^0 \in \mathbf{R}^N$ 由预训练的语义分割模块（Jiang等，2019）生成，该模块将输入点云按照人体几何结构划分为24个语义类别，并为每个点分配一个对应的类别标签，标签值即为所属语义类别索引。为了实现语义信息与几何坐标的高效融合，本文采用独热编码（one-hot encoding）策略，将初始语义标签转换为稀疏二值矩阵 $\boldsymbol{F}_i \in \mathbf{R}^{N \times 24}$。在此基础上，进一步提取每个点的首个非零索引值作为其语义标识，并将其映射为低维嵌入特征，用于后续卷积层中与点云坐标进行有效的特征交互。

多尺度的特征提取可以有效感知不同尺度下的人体结构，在单图像人体表明重建方法（邵明正，2023）中，作者通过多尺度点云特征的预测补全了二维图像输入的结构缺失。基于点云输入的人体网格重建任务更需要从不同尺度对整体动作、躯干末端动作进行特征提取，因此本文提出了多尺度特征编码器。如图2所示，多尺度特征编码器由SA层和FEM组成，以残缺输入 $\boldsymbol{P}_i \in \mathbf{R}^{N \times 3}$ 和语义标识 $\boldsymbol{F}_i$ 作为输入，输出多尺度逐点特征 $\boldsymbol{F}_s \in \mathbf{R}^{N_s \times d_s}$ 以及下采样点云 $\boldsymbol{P}_s \in \mathbf{R}^{N_s \times 3}$。首先，SA层通过最远点采样（farthest point sampling，FPS）算法对输入点云 $\boldsymbol{P}_i$ 进行下采样，并借助K近邻算法（K-nearest neighbors，KNN）聚合邻域点，获得第0尺度下的点云块及其特征。随后，第0尺度下的点云块特征与语义标签 $\boldsymbol{F}_i$ 级联后通过二维卷积层进行编码，并由最大池化进行特征汇聚获得第0尺度下的逐点特征 $\boldsymbol{F}_0 \in \mathbf{R}^{N_0 \times d_0}$。受PointNet++（Qi等，2017）启发，我们用四个SA层依次处理输入点云，对其分组聚合为多尺度的点云块，并将局部结构信息编码为特征表示 $\{\boldsymbol{F}_0, \boldsymbol{F}_1, \boldsymbol{F}_2, \boldsymbol{F}_3\}$。随着网络层次深入，分辨率逐渐降低，特征维度逐渐增加，捕获多尺度信息。在每一个SA层之后，FEM均对提取的特征进行非线性变换，进一步优化特征表达，为后续解码网络输入更加有效的点云局部和全局上下文信息。

如图3所示，FEM基于点云局部自注意力机制设计，旨在增强SA层所提取的点云块特征在空间位置上的相关性，从而提升模型对局部几何结构的感知能力。具体而言，FEM首先对逐点特征 $\{\boldsymbol{F}_i\}_{i=0}^{3}$ 通过一维卷积操作生成特征矩阵 $\boldsymbol{K}$；随后，利用KNN算法在局部邻域内聚合信息，进一步生成特征矩阵 $\boldsymbol{Q}$ 和 $\boldsymbol{V}$。为了引入空间位置信息，FEM将邻域点相对于中心点的坐标差进行线性变换，得到位置编码 δ，

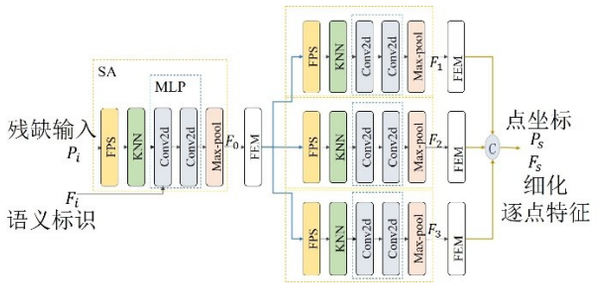


图2　多尺度编码模块结构图

Figure 2　Diagram of Multi-scale Encoder Module

其中相对坐标由KNN聚合后的局部点云块中各点与其对应中心点的坐标差计算获得。该设计使FEM能够在保留点云无序性的同时，有效建模局部区域内点与点之间的空间依赖关系，从而提升特征表达的结构敏感性和鲁棒性。$\boldsymbol{Q}$和$\boldsymbol{K}$经过卷积得到逐点相关性$\hat{h}_{ij}$。接着由最大池化和平均池化得到在通道维度上增强的注意力权重$\hat{h}_{ij}^{*}$：

$$\hat{h}_{ij} = \alpha\left(\beta(\boldsymbol{Q}) - \gamma(\boldsymbol{K})\right) + \delta, j \in \mathcal{N}(i) \tag{1}$$

$$\hat{h}_{ij}^{*} = \sigma\left(f\left(\mathrm{Cat}\left(\mathrm{Avg}\left(\hat{h}_{ij}\right), \mathrm{Max}\left(\hat{h}_{ij}\right)\right)\right) * \hat{h}_{ij}\right) \tag{2}$$

式中，$\alpha, \beta, \gamma, \sigma$为可学习的线性变换层，f表示卷积，Cat()表示特征拼接级联操作，Avg()和Max()分别表示平均池化和最大池化函数。

随后，注意力权重$\hat{h}_{ij}^{*}$与$\boldsymbol{V}$通过逐点乘积后得到细化逐点特征$\boldsymbol{F}_s$：

$$\boldsymbol{F}_s = \sum_{j \in \mathcal{N}(i)} \mathbb{N}\left(\hat{h}_{ij}^{*}\right) * \left(\psi(\boldsymbol{V}) + \delta\right) \tag{3}$$

式中：ψ均表示线性层，$\mathbb{N}(\cdot)$表示归一化处理，$*$表示逐元素乘积，$j \in \mathcal{N}(i)$表示点i周围的K近邻点j。

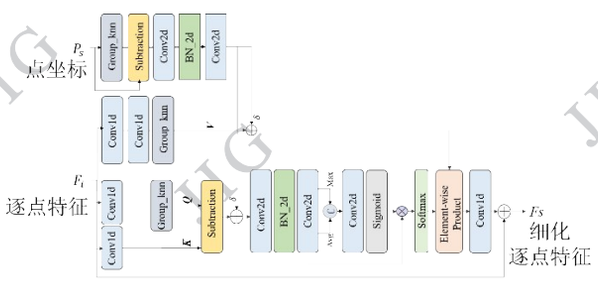


图3　特征增强模块结构图

Figure 3　Diagram of Feature Enhance Module

1.3　拓扑先验顶点回归模块

在人体重建任务中，引入结构先验知识对于提升模型对人体姿态和形状的理解能力具有重要意义。为此，本文提出一种基于SMPL模型固定拓扑结构的顶点回归模块。该模块旨在从提取的点云表征中建模顶点-顶点，顶点-关节之间的交互关系。实现在SMPL拓扑先验的引导下，将无序的逐点特征$\boldsymbol{F}_s$映射为有序的SMPL网格顶点特征的目标。该模块充分利用SMPL模型中预定义的顶点连接关系作为结构约束，指导网络从点云特征中学习符合人体拓扑结构的网格表示。如图4(a)所示，顶点回归网络由2个Transformer block组成，每个Transformer block包含两个Transformer编码层(Transformer-encoder layer)。在每一个Transformer block之后，通过线性层对嵌入特征进行降维，输出中间特征表示，并最终回归出初步的网格顶点坐标$\boldsymbol{V}_{\mathrm{coarse}}$。

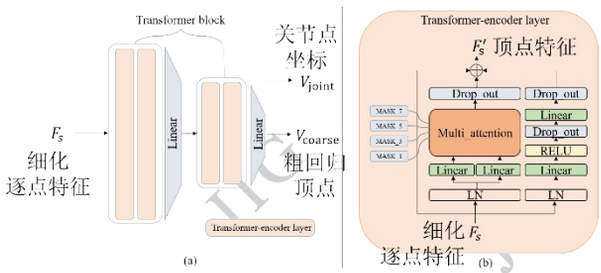


图4　拓扑先验顶点回归模块结构图

Figure 4　Diagram of Topology-Prior Vertices Regression Module

Transformer-encoder layer基于多头注意力机制实现，是顶点回归模块的核心结构，其设计如图4(b)所示，受文献(Kim等，2023)的启发，本文在每个Transformer-encoder layer中引入基于邻接矩阵的局部信息交互机制。具体而言，顶点回归模块利用SMPL模型提供的顶点邻接关系，构建一组局部邻域结构，在4个Transformer-encoder layer中引入邻接掩码$\left\{mask_{_t}\right\}^{t=1,3,5,7}$来限制注意力机制中的信息传播范围。4个掩码分别对应于邻接顶点数量为7、5、3、1的情形，在回归过程中掩码范围逐渐变小。随着网络层次加深，掩码所覆盖的邻接区域逐渐缩小，使模型的关注重点从全局结构一致性过渡到局部细节建模。这种由粗到细的掩码机制不仅提升了模型对人体复杂姿态与局部形变的适应能力，也有效增强了顶点预测结果的几何合理性。

通过多层级的Transformer结构和SMPL拓扑先验与渐进式邻接关系掩码的联合引导，本模块能够在不依赖复杂后处理的情况下，直接从残缺点云中回归出结构合理的SMPL网格顶点$\boldsymbol{V}_{\mathrm{coarse}}$及关

节点 V_{joint}。

1.4 偏移感知优化模块

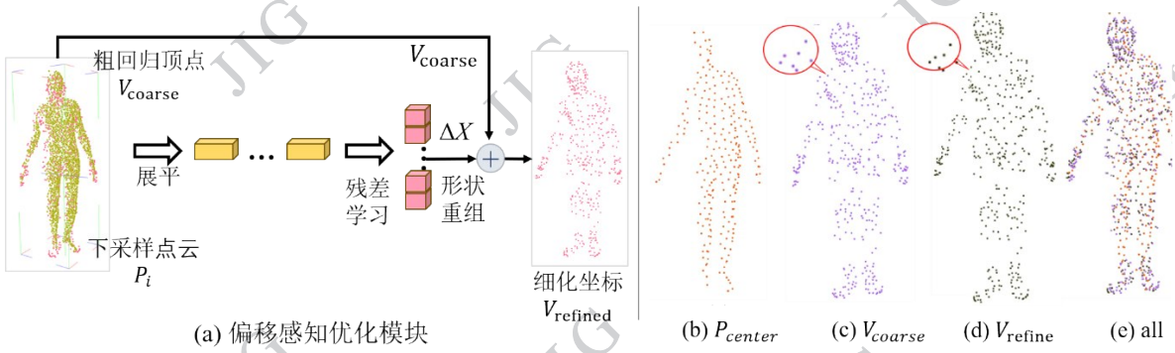


图5 偏移感知优化模块结构图

Figure 5 Diagram of Offset-aware Refine Module

如图5(a)所示，偏移感知优化模块采用反馈增强的机制，旨在通过学习粗回归结果 $\boldsymbol{V}_{coarse}$ 与输入点云之间的误差分布，进一步提升顶点预测精度。该模块的核心思想是：利用MLP从粗回归顶点 $\boldsymbol{V}_{coarse}$ 和下采样点云 $\boldsymbol{P}_{center}$ 中提取偏移误差 $\Delta\boldsymbol{X}$，并将其加到粗回归结果上，从而获得细化回归结果 $\boldsymbol{V}_{refined}$。具体而言，如图5(b)所示，输入点云经过下采样操作得到一组具有代表性的中心点云 $\boldsymbol{P}_{center}$；同时，如图5(c)所示，当前阶段输出的粗回归顶点 $\boldsymbol{V}_{coarse}$ 也作为输入之一。两者拼接后送入线性层 μ，用于学习偏移误差 $\Delta\boldsymbol{X}$，其计算方式如下：

$$\Delta\boldsymbol{X} = \mu\left(\mathrm{Cat}(\boldsymbol{V}_{coarse}, \boldsymbol{P}_{center}\right) \tag{4}$$

$$\boldsymbol{V}_{refined} = \boldsymbol{V}_{coarse} + \Delta\boldsymbol{X}. \tag{5}$$

如图5(d)所示，优化后的顶点有着更加清晰干净的点云边缘。图5(e)展示了输入点云块中心、粗回归结果与细化结果在同一坐标系下的可视化对比，直观体现了偏移感知优化的有效性。

为进一步获得完整的人体网格表示，本文参考文献(Ranjan等，2018)的上采样策略，将稀疏顶点 $\boldsymbol{V}_{coarse} \in \mathbf{R}^{N' \times 3}$ 通过上采样层扩展为稠密顶点 $\boldsymbol{V} \in \mathbf{R}^{N \times 3}$，即为SMPL模型的标准顶点数量。具体表达如下：

$$\boldsymbol{V} = \boldsymbol{U} \otimes \boldsymbol{V}_{coarse} \tag{6}$$

式中，$\boldsymbol{U}$ 是文献(Ranjan等，2018)预定义的上采样矩阵，N′为431，N为6890，即为SMPL模型的网格顶点。⊗表示矩阵乘法。

5 损失函数

损失函数 $\mathcal{L}$ 用于指导模型生成更完整的人体形状点云和准确的人体姿态。具体而言，损失函数由两部分组成：顶点重建损失 $\mathcal{L}_v$ 和关节点回归损失 $\mathcal{L}_j$。顶点重建损失 $\mathcal{L}_v$ 用于优化人体形状点云的完整性，而关节点 $\mathcal{L}_j$ 损失则确保人体姿态的准确性。通过联合优化这两部分损失，模型能够同时提升形状重建和姿态估计的性能：

$$\mathcal{L} = \lambda_v \mathcal{L}_v + \lambda_j \mathcal{L}_j \tag{7}$$

其中，λ_v=100和 λ_j = 1000分别控制顶点损失和关节点损失的权重，以平衡对人体形状和姿态建模的学习目标。

表面点损失 $\mathcal{L}_v$ 用于优化重建顶点的空间位置，使其尽可能接近真实SMPL网格顶点，从而提升人体形状的完整性与几何细节还原能力。考虑到顶点回归模块通过多分辨率策略逐步从稀疏到稠密地重建顶点，本文采用多尺度L1损失，分别对应三个不同分辨率下的输出：431、1723和6890个顶点。具体而言，该损失由三组预测顶点与对应真实顶点之间的L1距离加权组成，分别对应于不同阶段的重建结果。权重系数 λ_{v0}、λ_{v1}、λ_{v2} 用于调节各阶段损失对最终优化的影响强度。这种多分辨率损失设计使模型能够在粗粒度上快速逼近整体形状，在细粒度上进一步优化局部几何细节，从而实现高质量的人体形状重建：

$$\mathcal{L}_v = \lambda_{v0}\frac{1}{N_0}\sum_{i=1}^{N_0}\| \hat{V}_i^0 - V_i^{0,gt}\|_1 + \lambda_{v1}\frac{1}{N_1}\sum_{i=1}^{N_1}\| \hat{V}_i^1 - V_i^{1,gt}\|_1 + \lambda_{v2}\frac{1}{N_2}\sum_{i=1}^{N_2}\| \hat{V}_i^2 - V_i^{2,gt}\|_1 \quad (8)$$

式中，N_0=431、N_1=1723、N_2=6890分别表示不同分辨率下的人体顶点数量。$\hat{V}_i^k$表示第k阶段网络输出的预测顶点。$V_i^{k,gt}$表示对应的真值顶点坐标。λ_{v0}、λ_{v1}、λ_{v2}为各阶段损失的权重系数。

关节点损失$\mathcal{L}_j$用于提升模型对人体姿态的建模能力，确保重建结果符合合理的人体运动学约束。该损失由直接回归的关节点损失和SMPL拟合损失两部分构成：

$$\mathcal{L}_j = \frac{1}{J}\sum_{j=1}^{J} \| \hat{J}_j - J_{gt}\|^2 + \frac{1}{J}\sum_{j=1}^{J} \| J_j^k - J_{gt}\|^2 \quad (9)$$

式中，J为关节点数量。$\hat{J}_j$表示第j个预测关节点坐标。J_j^k是基于预测顶点经SMPL模型反推获得的关节点位置。J_{gt}表示对应的真实关节点坐标。

2 实验及结果

2.1 数据集

由于目前缺乏公开直接可用的人体点云数据集，本文基于两个广泛使用的公开数据集SURREAL和AMASS，构建用于人体点云补全任务的数据集。

SURREAL数据集基于SMPL参数化人体模型，通过物理仿真技术生成了约100万帧高质量的合成数据，包含行走、坐立等12类常见日常活动的动作序列及其对应的深度图。参考VoteHMR（Liu等，2021），本文随机选取了10万张深度图用于构建点云数据集，并按照7:1:2的比例划分为训练集、测试集和验证集。具体而言，每张深度图通过重投影技术转换为残缺稀疏点云，保留了单视角采集所导致的典型表面缺失模式；而完整点云则由SMPL模型参数直接生成，作为重建任务的目标。

为进一步评估模型在真实动作分布下的泛化能力，本文还在AMASS数据集上进行了跨数据集的性能测试。由于AMASS数据集并未包含语义信息，因此，我们在该数据集上进行的实验重建结果为无序点云，且无法获得SMPL网格，即点云补全。AMASS是一个大规模真人运动捕捉数据集，包含超过40小时、总计超过11000个动作序列，涵盖了行走、跑步、舞蹈等300多种不同类型的复杂动作。相较于SURREAL，AMASS在姿态多样性、动作复杂性和现实性方面更具优势。针对AMASS数据集，本文首先高效地生成高质量的3D人体网格模型，随后模拟深度相机对三维模型进行多角度拍摄，获得单视角的深度图像。再将这些深度图转化为点云数据，最终构建出适用于点云补全任务的AMASS-Point数据集。

最终，SURREAL点云数据集中共包含10万个样本，其中训练集7万个，测试集1万个，验证集2万个；AMASS-Point数据集中共包含4万个样本，测试集与验证集各5000个。该数据集设计不仅支持模型在合成环境中的充分训练与验证，也为跨数据集泛化能力提供了有力支撑。

2.2 实验配置

所有实验均在配备NVIDIA RTX4090GPU的高性能计算平台上进行。训练过程中，批大小设置为32，模型训练总轮数为400。优化器采用Adam，其初始学习率设置为0.001，并配合warmup学习率预热调整策略，预热阶段持续20个轮次，以稳定训练初期的梯度更新。学习率衰减采用的权重衰减系数为0.98。

在网络架构设计中，特征维度、关节点表示维度以及位置编码维度均设置为256，以确保模型在特征提取和空间信息编码上的表达能力。此外，偏移感知优化模块中的嵌入层维度设置为2048。

损失函数的设计中，各损失项的权重配置如下：顶点重建损失（λ_v）权重为100，关节点回归损失（λ_j）权重为1000。为了进一步提升模型在多尺度下的几何一致性，多尺度顶点重建损失的权重分配为λ_{v0}=0.33、λ_{v1}=0.33、λ_{v2}=0.33。

本文所提出模型的总参数量为22.0M，在单张NVIDIA RTX 4090 GPU上的平均单帧推理时间为19.6 ms（约51 FPS），所有实验均基于PyTorch框架实现，未采用模型剪枝、量化或TensorRT等加速技术。该推理效率满足虚拟试衣、动作捕捉等准实时应用场景的需求，体现了方法在精度与效率之间的良好平衡。

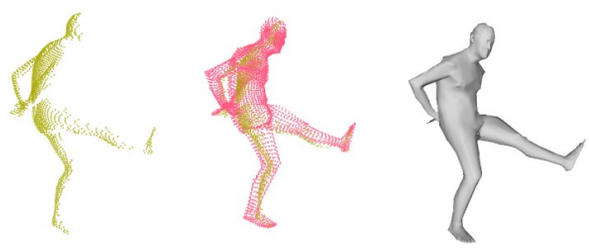
2.3 SURREAL数据集重建结果

本文的核心目标是对输入的残缺点云进行高精度补全，并生成符合SMPL参数化人体模型拓扑结构的网格顶点。与传统的点云补全任务主要关注几何完整性不同，本文不仅追求形状的视觉合理性，更

强调补全结果与SMPL模型之间的拓扑一致性。这种一致性意味着补全后的点云在结构上具备明确的人体语义，能够直接映射为具有标准连接关系的SMPL网格模型，而无需后续复杂的配准或变形操作。

为此，本文在补全过程引入SMPL模型预定义的拓扑先验信息（顶点邻接关系），作为对人体结构的强先验约束。该先验不仅指导模型合理恢复因单视角采集导致的表面缺失区域，还有效避免了非人体结构的不合理补全，显著提升了重建的准确性和结构鲁棒性。通过将补全过程约束在SMPL流形空间内，所提方法能够输出既具几何完整性又满足拓扑一致性的点云结果，便于下游任务（如姿态估计、动作识别等）直接使用。

实验表明，SMPL拓扑先验的引入为残缺点云的补全提供了强有力的结构引导，使模型能够在缺乏多视角信息的情况下仍保持对人体结构的高度还原能力。最终生成的完整点云图6(b)可直接转换为具有语义一致性的SMPL网格模型图6(c)，验证了本方法在结构保持与语义表达方面的有效性。



(a)残缺点云 (b)重建结果 (c)SMPL网格

图6　在SURREAL测试集SMPL网格重建结果

Figure 6　SMPL mesh reconstruction results on SURREAL

此外，本文在多种人体姿态和点云缺失模式下对所提方法进行了全面测试，以验证其鲁棒性与泛化能力。图7展示了6组具有代表性的重建结果，涵盖了不同的人体姿态（如站立、行走、下蹲、跳跃、背手等）、多样化的体型特征，以及不同程度和位置的点云缺失情况。每组数据从左到右依次展示输入的残缺点云、重建得到的SMPL网格顶点以及对应的真值，其中中间列进一步呈现了重建结果（红色）与原始输入点云（绿色）在同一坐标系下的叠加效果，直观体现了模型在几何对齐与结构恢复方面的表现。

从图7可以看出，尽管输入点云存在显著的缺失，重建结果仍能够较好地保留原始姿态与整体形状，并在缺失区域生成符合人体结构语义的补全内容。具体而言，在输入存在大面积缺失的情况下（如数据(a)-(c)），本文方法依然能够基于SMPL拓扑先验和上下文信息合理预测出完整的人体结构，即使输入仅包含躯干或部分肢体的关键点，也能准确恢复其余部位的姿态与连接关系。对于复杂姿态（如下蹲、跳跃、抬手等，对应数据(d)-(f)），这些动作通常伴随着肢体交叉、大角度弯曲等挑战性结构变化，而模型在输入点云严重遮挡的情况下仍能保持良好的姿态一致性与结构完整性。

此外，对图7进一步分析，尽管由SMPL引入的先验拓扑关系相对固定，但对于不同体型的人体，例如点云(c)和点云(f)，两组点云的手部动作类似，但体型存在明显区别，而我们的重建结果能够对这两组点云进行准确重建，包括缺失点云的补全，最重要的是重建的结果也与真值保持一致。

综上所述，本文通过引入SMPL参数化模型的顶点拓扑关系作为约束，使得模型能够在有限信息下依然生成结构合理、语义一致的重建结果。图7充分展示了本文方法在面对多样化的人体姿态、体型以及复杂缺失条件时，具备较强的适应能力和稳定的重建性能。

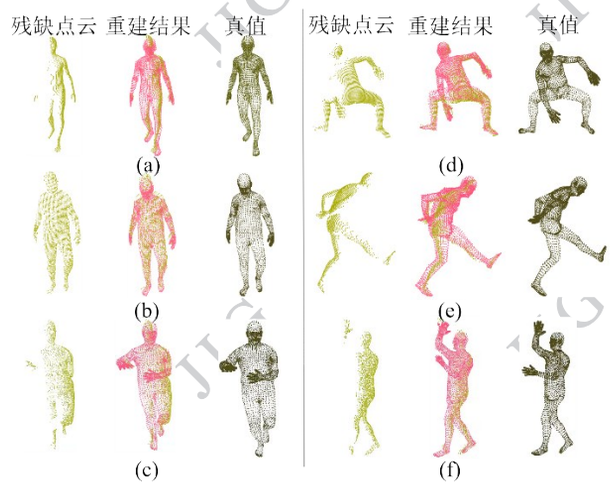


图7　在SURREAL测试集的重建结果

Figure 7　Reconstruction results on SURREAL

2.4　量化比较

为更加全面且客观地评估所提算法的性能，本文在SURREAL数据集上采用多个权威评价指标，与经典方法及当前最新的人体重建方法进行了对比

实验。具体而言，本文选用平均顶点误差（mean per vertex error，MPVE）、平均逐关节位置误差（mean per joint position error，MPJPE）以及对齐后的平均逐关节位置误差（procrustes-aligned MPJPE，PA-MPJPE）三个指标，分别从几何形状精度、关节点定位准确性和姿态一致性三个方面对人体重建质量和姿态估计能力进行量化评估。通过与现有方法在上述指标上的系统对比，能够有效验证本文方法在保持人体结构完整性、提升重建精度以及增强姿态估计鲁棒性方面的优势。

根据输入数据的来源，本文将现有基于点云的三维人体重建方法分为两类：基于深度图点云的方法和基于雷达点云的方法。如表1所示，本文将所提出算法与这两类方法进行了系统的量化对比，以全面评估其在不同感知条件下的性能表现。在与深度图点云方法的比较中，我们沿用VoteHMR（Liu等，2021）的评估策略，在SMPL模型输出的1723个顶点上计算MPVE指标。本文方法在SURREAL数据集上的MPVE达到了18.7mm；进一步将重建网格上采样至标准的6890个顶点后，顶点误差为19.5mm，分别优于PointHPS（Cai等，2023）和VoteHMR（Li等，2021）4.6%（从19.6mm降至18.7mm）和7.4%（从20.2mm降至18.7mm），表明本文方法在整体重建精度上的显著优势。此外，在姿态估计估计方面，本文方法在MPJPE和PA-MPJPE指标上分别取得了16.2mm和12.0mm的性能，较VoteHMR分别提升了8.5%（从17.7mm降至16.2mm）和1.6%（从12.2降至12.0mm），展现出在姿态估计中的领先能力。

深度图点云通常由深度相机采集并转换生成，具有高分辨率和接近真实场景的空间分布特性。相比之下，雷达点云多来自激光雷达，分辨率较低，且受扫描模式影响显著，如机械旋转式雷达产生的点云呈“层状”稀疏分布，给处理带来额外挑战。

为更全面、公平地评估算法在低分辨率与稀疏输入下的鲁棒性，本文针对雷达点云类方法设置了更具实际意义的输入配置：将输入点云下采样至约512个点（几位Ours-512），以模拟真实部署中感知能力受限的场景。在此设定下，本文方法，本文方法在MPVE和MPJPE指标上仍分别达到了27.0mm和20.9mm，相较于雷达点云方法LiveHPS（Ren等，2024）分别提升了2.3%（从30.8mm降至27.0mm）和12.9%（从24.0mm降至16.2mm）。

综上所述，通过与基于深度图点云和雷达点云的代表性方法进行全面对比，本文方法在多种评价指标下均取得了最优性能，充分验证了其在三维人体重建任务中的有效性与优越性。

表1　在SURREAL数据集上进行三维人体重建的实验结果

Table 1　3D human reconstruction experimental results on SURREAL dataset

类别	方法	MPVE↓ (mm)	MPJPE↓ (mm)	PA-MPJPE↓ (mm)
深度图点云	Jiang et al.(Jiang等，2019)	80.5	–	–
	3DCODED(Groueix等，2018)	41.8	38.5	33.1
	Sequential (Param.)(Wang等，2020)	24.3	21.9	18.2
	Sequential (Non Param.)(Wang等，2020)	21.2	18.7	13.4
	VoteHMR(Liu等，2021)	20.2	17.7	12.2
	PointHPS(Cai等，2023)	19.6	–	–
雷达点云	LiDARCap(Li等，2022)	42.8	54.1	–
	LIP(Ren等，2023)	31.7	42.4	–
	LiveHPS(Ren等，2024)	30.8	24.0	–
	Ours-512	27.0	20.9	15.3
	Ours	**18.7(19.5)**	**16.2**	**12.0**

注：表中加粗值表示最佳性能；“–”表示结果不可用。

此外，为进一步评估所提出方法在点云补全任务中的适用性，本文在AMASS-Point数据集上开展了跨数据集测试。为确保实验对比的公平性与可重复性，本文对点云补全领域的经典算法以及前沿方法（Rong等，2024）进行复现与重训练。实验采用多维度评估体系，从全局几何一致性和局部完整性两个方面对补全效果进行量化评估。如表2所示，本文方法在Chamfer距离L1范数（chamfer distance L1-norm，CD-L1）指标上取得最优结果 6.27×10^{-3}，相较当前最优方法CRA-PCN（Rong等，2024）的 10.28×10^{-3} 提升了39.1%，表明本方法在全局结构重建方面具有更强的能力。与此同时，在F1-Score指标上，本文方法得分为0.61，显著优于所有已有方法（如Seedformer（Zhou等，2022）的0.49和CRA-PCN的0.55），说明其在保持点云完整性方面也展现出良好性能。

综上所述，通过与多种代表性点云补全方法的系统对比，本文方法在整体重建质量上展现出明显优势，验证了其在复杂人体姿态下进行点云补全的有效性与鲁棒性。

表2　在AMASS数据集上进行点云补全的实验结果

Table 2　3D human point clouds completion experimental results on AMASS dataset

方法	CD-L1↓ $\times10^{-3}$	F1-Score↑
FoldingNet(Yang等，2018)	31.16	0.11
PCN(Yuan等，2018)	23.13	0.21
PoinTr(Yu等，2021)	25.12	0.28
Pmp-net++(Wen等，2022)	18.69	0.31
Seedformer(Zhou等，2022)	16.92	0.49
CRA-PCN(Rong等，2024)	10.28	0.55
Ours	**6.27**	**0.61**

注：表中加粗值表示最佳性能。

2.5　消融实验

在表3中，本文评估了语义编码、多尺度特征提取模块和偏移感知优化模块对模型性能的影响。性能指标包括顶点误差（PVE）、关节位置误差（MPJPE）。通过不同模块的组合实验，表格清晰地展示了各模块对模型性能的贡献，为优化人体点云重建任务提供了重要的参考依据。

方法A和方法B作为基线模型，均未引入语义编码、多尺度特征提取及偏移感知模块。其中，方法A以完整的下采样点云作为输入，而方法B则采用基于深度图采样生成的残缺点云作为输入。实验结果表明，方法A在MPVE和MPJPE指标上分别达到了9.3mm和6.3mm，显著优于方法B的32.7mm和24.2mm。这一对比结果清晰地表明，在相同的网络架构下，完整点云输入的重建性能显著优于残缺点云输入。因此，残缺输入所带来的数据不完整性是当前任务的主要挑战。

方法C在基线模型中加入语义编码，将MPVE降低至23.1mm，性能提升了32.0%。通过引入语义标签，可以为人体缺失区域提供先验的语义知识，辅助模型推断缺失部分的合理结构。这种语义引导不仅增强了模型对人体点云局部特征的感知能力，还提升了其对全局结构的理解，改善了特征提取的完整性和鲁棒性，从而显著提升模型重建的性能。

方法D、E在方法C的基础上逐步引入了多尺度编码和偏移感知模块。方法D较方法E额外引入偏移感知模块，MPVE为19.5mm，提升了14.5%。偏移感知模块通过显式学习输入残缺点云与初步回归的顶点偏移误差，进一步优化了顶点回归的精度。这一结果表明，偏移感知模块能够以简单有效的网络结构校正模型回归的顶点位置，从而提升重建结果的准确性。方法D重建结果的MPVE为22.8mm，带来了1.3%的提升，说明多尺度编码器结构可以更好感知点云特征。对于基于点云输入的重建任务，网格的顶点的回归方式，避免了复杂的后处理流程（如SMPL蒙皮），充分利用点云与目标网格在三维空间中的结构一致性，构建更加紧密的输入-输出关联，显著提升了对遮挡、稀疏和非均匀采样点云的鲁棒性。更重要的是，通过偏移感知优化，可以进一步提升模型对肢体细节的感知能力。

3　结　论

本文针对当前点云人体重建方法中对点云特征利用不足以及复杂姿态下残缺严重的重建问题，提出了一种基于SMPL拓扑先验的人体重建算法，旨在从输入残缺点云重建完整人体网格表示。该算法首先提出融合语义对输入残缺点云从多个尺度上提取丰富几何结构信息。随后，在SMPL顶点邻接关

表3　语义、多尺度编码和偏移感知优化模块的消融实验
Table 3　Ablation Study of the Semantic, Multi-scale encoder, and Offset-aware refine Modules

方法	残缺	语义编码	多尺度	偏移感知	PVE ↓ (mm)	MPJPE ↓ (mm)
A	×	×	×	×	9.3	6.6
B	√	×	×	×	32.7	24.2
C	√	√	×	×	23.1	17.6
D	√	√	√	×	22.8	17.1
E	√	√	√	√	**19.5**	**16.2**

注：表中加粗值表示最佳性能。

系的掩码下，通过Transformer建立顶点-顶点，顶点-关节点之间的交互关系，并对网格顶点和关节点进行回归。最终，偏移感知优化模块将回归的网格顶点结合输入点云进一步预估偏移误差从而提升回归精度。实验结果表明，与现有的基于点云的人体重建技术相比，本文方法在MPVE和MPJPE均达到了最先进的性能水平，同时在人体点云补全场景也具有最优的性能。未来将进一步优化模型轻量化与细粒度重建能力，推动该方法在虚拟现实、智能交互与数字内容创作等领域的实用化部署及深度应用。

参考文献（References）

Cai Z, Pan L, Wei C, Yin W, Hong F, Zhang M, Loy C C, Yang LandLiu Z.2023.Pointhps: cascaded 3d human pose and shape estimation from point clouds.[EB/OL].[2023-8-28]. https://arxiv.org/pdf/2308. 14492. pdf

Groueix T, Fisher M, Kim V G, Russell B CandAubry M.2018. 3d-coded: 3d correspondences by deep deformation//Proceedings of the European Conference on Computer Vision Munich: Springer: 230-246 [DOI: 10.1007/978-3-030-01216-8_15]

Jiang H, Cai JandZheng J.2019. Skeleton-aware 3d human shape reconstruction from point clouds//Proceedings of the IEEE/CVF International Conference on Computer Vision.IEEE: 5431-5441 [DOI: 10.1109/ICCV.2019.00553]

Kim J, Gwon M-G, Park H, Kwon H, Um G-MandKim W.2023. Sampling is matter: point-guided 3d human mesh reconstruction//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.IEEE: 12880-12889 [DOI: 10.1109/CVPR52729.2023.01238]

Li Z Q, Zhang Y L, Wang C Y and Chen X J. 2024. Research on 3D human body reconstruction algorithm based on parametric models [D]. East China Normal University, Shanghai (李子清，张一林，王晨宇，陈晓军. 2024. 基于参数化模型的三维人体重建算法研究 [D]. 华东师范大学，上海) [DOI: 10.27149/d.cnki.ghdsu.2024.003820]

Shao M Z, Liu Y F, Xu J and Zhao H B. 2023. Research on single-image human surface reconstruction technology based on implicit functions [D]. Anhui University, Hefei (邵明正，刘亚飞，徐杰，赵宏博. 2023. 基于隐式函数的单幅图像人体表面重建技术研究 [D]. 安徽大学，合肥) [DOI: 10.26917/d.cnki.ganhu.2023.001442]

Li J, Zhang J, Wang Z, Shen S, Wen C, Ma Y, Xu L, Yu JandWang C.2022. Lidarcap: Long-range marker-less 3d human motion capture with lidar point clouds//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE: 20502-20512 [DOI: 10.1109/CVPR52688.2022.01985]

Lin K, Wang LandLiu Z.2021. End-to-end human pose and mesh reconstruction with transformers//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE: 1954-1963 [DOI: 10.1109/CVPR46437.2021.00199]

Liu G, Rong YandSheng L.2021. Votehmr: Occlusion-aware voting network for robust 3d human mesh recovery from partial point clouds//Proceedings of the 29th ACM International Conference on Multimedia.Chengdu: ACM: 955-964 [DOI: 10.1145/3474085.3475309]

Loper M, Mahmood N, Romero J, Pons-Moll GandBlack M J. 2023. Smpl: A skinned multi-person linear model. Seminal Graphics Papers: Pushing the Boundaries, Volume 2.851-866. [DOI: 10.1145/2816795.2818013]

Qi C R, Yi L, Su HandGuibas L J.2017.Pointnet++: Deep hierarchical feature learning on point sets in a metric space. [EB/OL]. [2017-6-7]. https://arxiv.org/pdf/1706.02413.pdf

Ranjan A, Bolkart T, Sanyal SandBlack M J.2018. Generating 3d faces using convolutional mesh autoencoders//Proceedings of the European Conference on Computer Vision. Munich: Springer: 704-720 [DOI: 10.1007/978-3-030-01219-9_43]

Ren Y, Han X, Zhao C, Wang J, Xu L, Yu JandMa Y.2024. Livehps: Lidar-based scene-level human pose and shape estimation in free environment//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Yokohama: IEEE: 1281-1291 [DOI: 10.1109/CVPR52733.2024.00128]

Yao L, Zhang Y A, Zhang M X, et al. The impact of joint axis angle prior on the results of 3D human body reconstruction [J]. Journal of Image and Graphics, 2021, 26(12): 2918 - 2930. (姚砺，张幼安，张梦雪，等. 关节轴角先验对3维人体重建结果的影响[J]. 中国图象图形学报，2021，26(12)：2918-2930). [DOI: 10.11834/jig.200348]

Ren Y, Zhao C, He Y, Cong P, Liang H, Yu J, Xu LandMa Y.2023. Lidar-aid inertial poser: Large-scale human motion capture by sparse inertial and lidar sensors. IEEE Transactions on Visualiza-

tion and Computer Graphics. 29（5）：2337-2347.［DOI：10.1109/TVCG.2023.3247088］

Rong Y，Zhou H，Yuan L，Mei C，Wang JandLu T. 2024. Cra-pcn：Point cloud completion with intra-and inter-level cross-resolution transformers//Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver：AAAI Press：4676-4685［DOI：10.1109/TVCG.2023.3247088］

Vasilikopoulos N，Drosakis DandArgyros A. 2024. D-pose：Depth as an intermediate representation for 3d human pose and shape estimation.［EB/OL］.［2024-10-7］. https://arxiv.org/pdf/2410.04889.pdf

Wang K，Xie J，Zhang G，Liu LandYang J. 2020. Sequential 3d human pose and shape estimation from point clouds//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle：IEEE：7275-7284［DOI：10.1109/CVPR42600.2020.00730］

Wen X，Xiang P，Han Z，Cao Y-P，Wan P，Zheng WandLiu Y-S. 2022. Pmp-net++：Point cloud completion by transformer-enhanced multi-step point moving paths. IEEE transactions on pattern analysis and machine intelligence. 45（1）：852-867.［DOI：10.1109/TPAMI.2022.3159003］

Yang W，Jiang Z，Zhao SandZhou S K. 2025. Postometro：Pose token enhanced mesh transformer for robust 3d human mesh recovery//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa：IEEE：［DOI：10.1109/WACV61041.2025.00465］

Yang Y，Feng C，Shen YandTian D. 2018. Foldingnet：Point cloud auto-encoder via deep grid deformation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City：IEEE：206-215［DOI：10.1109/CVPR.2018.00029］

Yu X，Rao Y，Wang Z，Liu Z，Lu JandZhou J. 2021. Pointr：Diverse point cloud completion with geometry-aware transformers//Proceedings of the IEEE/CVF International Conference on Computer Vision. Vancouver：IEEE：12498-12507［DOI：10.1109/ICCV48922.2021.01227］

Yuan W，Khot T，Held D，Mertz CandHebert M. 2018. Pcn：Point completion network//2018 International Conference on 3D Vision. Vancouver：IEEE：728-737［DOI：10.1109/3DV.2018.00088］

Zhou H，Cao Y，Chu W，Zhu J，Lu T，Tai YandWang C. 2022. Seedformer：Patch seeds based point cloud completion with upsample transformer//European Conference on Computer Vision. Tel Aviv：Springer：416-432［DOI：10.1007/978-3-031-20062-5_24］

作者简介

黄新彭，男，上海大学通信与信息工程学院副教授，博士，主要研究方向为光场数据压缩与处理。E-mail：huangxinpeng@shu.edu.cn

杨 超，男，上海大学通信与信息工程学院副教授，博士，主要研究方向为视频编码与质量评价。E-mail：yangchaoie@shu.edu.cn