

DOI:10.13232/j.cnki.jnju.2023.05.002

基于稳定性优化的三维装配补全方法

姚启皓, 王伟昊, 尤鸣宇*

(同济大学电子与信息工程学院, 上海, 201804)

摘要: 三维装配补全是一项重要的交互式装配任务, 对于一个半成品装配体, 机器人需要明确其缺失部件, 从候选部件中挑选正确部件, 计算准确的拼装位姿, 最后将半成品补全。稳定性是椅子、桌子等实际装配体设计的首要原则, 也是三维装配补全的重要目标, 现有的装配补全工作多根据部件的几何关系来优化装配补全过程, 没有考虑补全后装配体的稳定性, 导致补全结果的正确率不高, 难以满足机器人实际装配的需求。针对上述问题, 提出一种基于稳定性优化的三维装配补全方法 (Finishing the Incomplete 3D Assembly with Transformer, StableFiT), 定义了一种装配体稳定性验证方法。基于 NVIDIA Isaac Sim 仿真平台训练了一个装配体稳定性判别器, 并基于稳定性判别器提供的稳定性反馈, 优化了三维装配体的补全。在 PartNet 数据集上开展实验验证, 结果表明 StableFiT 能够有效提升补全的装配体的正确性和稳定性。

关键词: 自动装配, 三维装配补全, 稳定性验证, 稳定性优化

中图分类号: TP183

文献标志码: A

3D assembly completion with stability optimization

Yao Qihao, Wang Weihao, You Mingyu*

(College of Electronic and Information Engineering, Tongji University, Shanghai, 201804, China)

Abstract: 3D assembly completion is an essential and complex interactive assembly task. The robot must identify the true missing parts, select the correct parts from a toolkit of candidates, calculate precise assembly poses, and ultimately make the incomplete assembly complete. As the primary principle in the design of actual assemblies such as chairs and tables, stability is also an ultimate goal of 3D assembly completion. Existing works of 3D assembly completion primarily focus on geometric relationship modeling of parts, without taking into account the stability of assembly, leading to low accuracy in completion and making it a challenge to meet the actual requirements of robot assembly. To tackle this issue, we propose StableFiT (Finishing the Incomplete 3D Assembly with Transformer) for 3D assembly completion with stability optimization. We introduce a novel stability verification method for the completed assembly. By training an assembly stability discriminator using the verification results obtained from the NVIDIA Isaac Sim simulation platform, we furtherly optimize 3D assembly completion based on stability feedback from the stability discriminator. Experimental results on the PartNet dataset demonstrate that StableFiT effectively improves the correctness and stability of the completed assemblies, addressing the limitations of existing assembly completion methods.

Key words: automatic assembly, 3D assembly completion, stability verification, stability optimization

自动装配是三维计算机视觉领域一个非常重要的研究课题, 也是许多智能机器人系统的理想

功能^[1]. 将自动装配技术应用于智能机器人领域, 不仅帮助用户从头组装如积木、家具等各种装配

基金项目: 国家自然科学基金(62073244), 上海市创新行动计划(20511100500)

收稿日期: 2023-06-26

* 通讯联系人, E-mail: myyou@tongji.edu.cn

体,也可以对部件缺损的装配体进行修补,很大程度节约了人力成本,也弥补了人类能力的不足。

现有的工作已经使机器人能够从零开始将一组部件组装成一个整体,即三维部件组装^[2],但是目前对部件缺失装配体的补全以及完整组装过程中特定步骤的实现,研究还较少。例如,训练机器人修理一把缺少一条椅腿的椅子,这是人机协同以及多智能体协同装配场景下一个非常基本但重要的任务,所以,对于三维装配补全的研究将成为今后的重点。

三维装配补全任务,除了任务自身需要机器人详细理解不完整的结构并推断出候选部件及部件-结构关系带来的难度,还存在将模型推广到实际应用的问题,即模型补全的装配体应当能应用于实际场景来指导机器人在现实中正确完成相应的装配补全任务,这要求补全的装配体要兼具几何层面的准确性和物理层面的稳定性,以装配更稳定和更实用的产品。现阶段,其他装配补全工作往往忽视对装配体稳定性的考虑,不利于模型从虚拟到现实的推广,也影响装配补全本身的精度。对不稳定装配体的典型展示如图1所示。

对于装配体的稳定性研究的主要难点:(1)对各类装配体的稳定性的定义缺乏相关的评价标准;(2)因为点云中不包含部件的物理信息,所以直接利用点云进行装配体稳定性验证的难度很大,也缺乏相关的理论指导和工作支撑;(3)装配体的稳定性验证结论一般通过定性分析得出,难以用来定量优化三维装配补全模型。

针对上述问题,本文提出一种基于稳定性优化的三维装配补全方法 StableFiT (Finishing the Incomplete 3D Assembly with Transformer),指导补全出稳定性和实用性更高的装配体。

该方法的主要贡献:(1)定义了一种衡量装配体稳定性的方法,并基于 NVIDIA Isaac Sim 仿真平台实现了该方法;(2)提出一种点云格式下的装配体稳定性判别器模型,实现了点云格式下装配体的稳定性分类;(3)将提出的判别器模型引入三维装配补全模型,实现了三维装配补全的定量稳定性优化,并在定性和定量实验中取得了优异的表现。

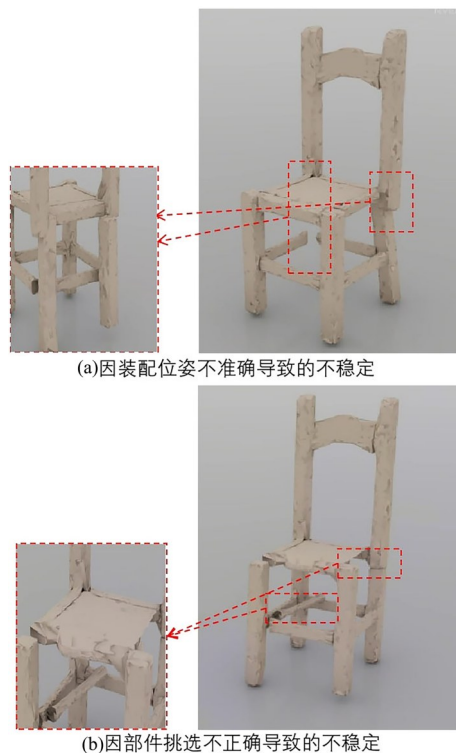


图1 不稳定装配体的典型展示

Fig. 1 A typical display of unstable assemblies

1 三维装配补全相关工作

三维部件装配是一项典型的装配任务,其目标是将一组给定的部件从头组装成完整的装配体^[3-5],目前对三维部件装配的研究工作较为广泛。Zhan et al^[2]首次正式定义三维部件装配问题,提出开创意义的DGL (Dynamic Graph Learning) 框架,基于动态图神经网络对部件之间的几何关系进行建模,预测每个部件的位姿。Harish et al^[6]提出RGL-NET (Recurrent Graph Learning),利用递归图学习范式来逐步细化部件之间的关系,能获得更准确的姿态预测。Wu et al^[7]提出一个基于部件的端到端生成式框架PQ-Net,通过三维重建对各个部件重新进行建模。Li et al^[8]提出PageNet,同样采用生成式的框架,整个框架由一个基于VAEGANs^[9-10]的部件生成器和一个部件组装器组成,能实现两阶段的三维部件组装。其他相关工作利用其他模态的信息来源作为输入,如Li et al^[11]使用图像作为输入。最新的关于三维部件装配的研究在问题设置上作出一定调整,如Wu et

al^[12]按照语义信息划分不同的部件组, Xu et al^[13]将任务目标转为无监督合成新的装配体. 但上述工作均以装配体的部件全集作为输入来预测部件的 6 自由度位姿, 三维装配补全在此基础上被提出, 两者在任务设置上存在一定差异.

三维装配补全旨在根据给定的半成品装配体, 从一系列候选部件组成的工具包中挑选合适的部件并以正确的位姿拼装, 最终将半成品补装完整. Wang et al^[14]首次提出并定义三维装配补全这一新任务, 设计了一个基于 Transformer 的框架 FiT (Finishing the Incomplete 3D Assembly with Transformer), 通过注意力机制对候选部件与半成品部件之间的几何关系进行建模, 发掘半成品真正缺失的部件并预测其位姿, 完成装配补全. 目前, 有关三维装配补全的研究仍在初步阶段.

三维点云补全是相近的一项任务, 不同于“拼装式”补全, 三维点云补全直接对点云数据中的缺失部分进行“生成式”预测和恢复, 通常采用带有编码器-解码器框架的生成式解决方案, 如 Tchapmi et al^[15]的 TopNet, Yuan et al^[16]的 PCN (Point Completion Network) 和 Wen et al^[17]的 SA-Net (Skip-Attention Network). 还有一些方法采用对抗学习的框架来提高点云补全的真实性和一致性, 如 Sarmad et al^[18]的 RL-GAN-Net (Reinforcement Learning Agent Controlled GAN Network) 和 Hu et al^[19]的 Render4Completion. 最近的一些工作将 Transformer^[20]引入三维形状补全来对点云分布进行全局建模, 如 Yan et al^[21]的 ShapeFormer 和 Mittal et al^[22]的 AutoSDF 都引入了 PQ-VAE^[23-24], 在离散空间中学习三维表征并用 Transformer 来回归补全缺失区域, Wen et al^[25]的 PMP-Net++ 和 Li et al^[26]的 ProxyFormer 同样使用 Transformer 进行三维点云补全. 总体地, 三维装

配补全侧重于部件级别的补全, 每个部件都是一组具有特定几何和语义信息的点云, 而三维点云补全则侧重于由于遮挡等造成的缺失补全.

上述装配工作仅关注部件的几何关系, 没有考虑装配体的稳定性, 这一方面会影响三维装配补全的准确性, 另一方面也不利于将三维装配补全真正应用于实际. 本文在 Wang et al^[14]的基础上, 设计了一种稳定性的验证方式与评价指标, 并将稳定性约束引入三维装配补全框架, 进一步提升补全的装配体的正确性和稳定性.

2 基于 NVIDIA Isaac Sim 仿真平台的装配体稳定性验证方法

针对上述装配体稳定性的问题, 本文创造性地提出一种基于 NVIDIA Isaac Sim 仿真平台的装配体稳定性验证方法, 如图 2 所示. 首先对数据进行从点云到 USD 的格式转换, 然后将其导入 NVIDIA Isaac Sim 仿真平台; 在物理仿真过程中, 仿真环境会实时记录装配体的状态; 仿真结束后, 系统会根据本文制定的稳定性判断规则来计算装配体的稳定性指标, 验证经过三维补全的装配体的稳定性.

利用仿真平台上的稳定性验证结果, 本文训练了一个装配体稳定性判别器模型. 输入补全的装配体的点云, 模型能直接判定装配体的稳定性. 该稳定性判别器模型可以直接在点云数据格式对装配体的稳定性进行判定, 不需要在仿真平台中实现, 为将稳定性引入三维装配补全网络的优化提供了便利与可能.

2.1 装配体稳定性量化定义 参考相关国家标准^[27]中的稳定性测试方法, 本文设计了一种关于装配体稳定性的普适性验证方法与量化指标. 为了使补全的装配体具有良好的现实迁移性, 稳定性验证应最大程度地模拟装配体在现实中被使用

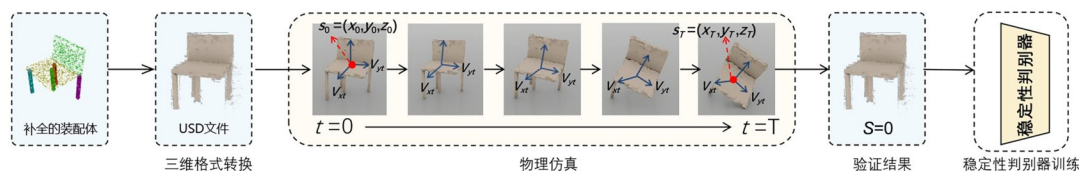


图 2 基于 Isaac Sim 仿真平台的装配体稳定性验证方法

Fig. 2 A stability verification method for the completed assembly based on Isaac Sim simulation platform

时的情况,本文的做法是将装配体静止放置于水平面上,观察其是否能保持稳定,即是否发生晃动或倾翻等现象.记初始时刻为 $t=0$,此时装配体的状态记作 $s_0=(x_0, y_0, z_0)$,其中 (x, y, z) 表示装配体重心的三维坐标.记末尾时刻为 $t=T$,此时装配体的状态记作 $s_T=(x_T, y_T, z_T)$,同时实时记录 $0\sim T$ 任意时刻装配体沿 x 和 y 方向的速度 V_{x_i} , V_{y_i} (采用右手坐标系,默认 z 方向为竖直向上, $0\leq t\leq T$).则初始时刻与末尾时刻装配体状态 s_0 和 s_T 之间的余弦相似度为:

$$\text{sim}(s_0, s_T) = \cos \theta = \frac{s_0 \cdot s_T}{|s_0| \cdot |s_T|} \quad (1)$$

记装配体稳定性为 S , S 为二值变量, $S=1$ 表示装配体稳定, $S=0$ 表示装配体不稳定,则本文提出的装配体稳定性的具体量化定义可归纳为:

$$S = \begin{cases} 0, & \max(V_{x_i}, V_{y_i}) > 0.1 \text{ or } \text{sim}(s_0, s_T) < 0.7 \\ 1, & \text{otherwise} \end{cases} \quad (2)$$

其中,对于速度的约束条件主要是确保稳定的装配体不会发生晃动,对于状态的约束条件主要是确保稳定的装配体不会发生倾翻.

式(2)是本文提出的关于装配体稳定性的一种具有普适性的定义方式,对于某些特定类型的装配体,如椅子、桌子等,其稳定性的定义以及检验方式应当在此基础上作出适当修改,比如在检验时添加适当大小的外力以更贴合其现实的使用场景.本文后续会对此进行更加深入的探索.

2.2 基于NVIDIA Isaac Sim 仿真平台的稳定性验证 NVIDIA Isaac Sim 是 NVIDIA Omniverse 平台推出的一款机器人模拟工具包, Isaac Sim 具有构建虚拟机器人世界和实验的基本功能,为研究和从业人员提供所需的工具和工作流程来创建稳定的、物理上精确的仿真环境及合成数据集.

基于 NVIDIA Isaac Sim 仿真平台,本文搭建了一个非常接近现实场景且物理精确的仿真环境来对装配体进行稳定性检验,其具体流程可以参考图2,首先添加一个具有物理属性的水平面来承载待检测的装配体,然后导入装配体.由于仿真平台只支持USD格式文件的导入,所以需要先将其点云格式的装配体转换成USD格式.导入装配体后,将其静止放置在水平面上并为其赋予物

理属性.为了使提出的稳定性检验方法具有普适性,将所有装配体统一设置为质量 $m=5\text{ kg}$,且均由相同材料制成.准备工作完成后,按下 Isaac Sim 操作界面的仿真按钮,平台内置的物理引擎将自动对物理现象进行实时的精确的模拟,同时, Isaac Sim 界面的显示面板也将实时显示装配体的位置坐标及其沿各个方向的速度,方便用户实时掌握装配体的状态.记仿真开始的时刻为0,在 T 时刻仿真自动结束.最终,仿真环境根据2.1中定义的稳定性量化指标对装配体的稳定性进行评估,并将结果返回给用户.

2.3 装配体稳定性判别器 如果按照2.2所述,每次对补全的装配体进行稳定性检验都要先转换文件格式再导入仿真环境,会大大增加实验的时间成本,并且,通过这种方式获得的装配体稳定性标签不具有梯度信息,无法用于三维装配补全网络的梯度更新与优化.

为了解决上述问题,本文设计了一个基于装配体稳定性的判别器模型,如图3所示.该模型主要由两层结构组成:第一层为特征提取层,采用 PointNet^[28] 提取部件级别的点云特征;第二层为分类层,包含两个全连接层和一个预测层,全连接层将部件级别的点云特征进行整合,提取实例级别的点云特征,最终在预测层实现从实例级点云特征到稳定性类别标签的预测.

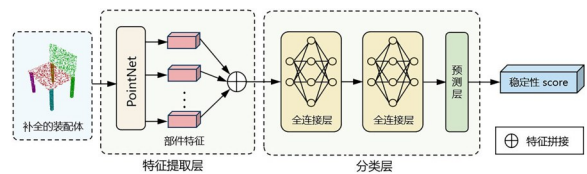


图3 装配体稳定性判别器

Fig. 3 Assembly stability discriminator

稳定性判别器模型总体实现的效果为:对模型输入装配体所有部件的点云数据,输出关于该装配体稳定性的预测评分,评分越接近1,表示装配体的稳定性越好.输出的评分还具有梯度信息,为后续对3D装配补全网络进行优化提供可能.

利用基于 Isaac Sim 仿真平台的稳定性验证方法,本文得到了一批带有稳定性标签的装配体数据,并基于此训练出装配体稳定性判别器模型,具体的训练细节参考4.3. 稳定性判别器模型训

练过程的损失函数定义如下:

$$\ell_d = \frac{1}{n} \sum_{i=1}^n |score_i - score_i^*| \quad (3)$$

其中, $score_i^*$ 表示装配体稳定性的真实标签.

3 基于稳定性优化的三维装配补全方法

首先对三维装配补全任务进行简单介绍.

给定一个具有 $N-j$ 个部件的半成品装配体 $I = \{P_i\}_{i=1}^{N-j}$, 其中, $P_i \in R^{1000 \times 3}$ 表示特定部件的一组三维点云, N 为完整装配体的部件总数. 任务目标是从候选部件库 $D = \{P_i\}_{i=1}^M$ 中挑选半成品装配体 I 缺失的部件 $M = \{P_i\}_{i=1}^j$, 同时预测这 j 个部件的 6 自由度位姿 $\{(R_i, t_i)\}_{i=1}^j$, 最终将所选部件装配到半成品装配体上. 其中, $R_i \in R^4$ 表示旋转四元数, $t_i \in R^3$ 表示平移变换, T_i 表示装配过程中基于位姿 (R_i, t_i) 的联合变换.

将本课题组已有的 FiT 框架作为三维装配补全网络的主体架构, 在其基础上引入 2.3 中的装配体稳定性判别器模型, 首创性地提出了一种基于稳定性优化的三维装配补全方法 StableFiT, 整体架构如图 4 所示. 其中, FiT 是一个基于 Transformer 的三维装配补全框架, 主要由 PointNet、编码器、解码器和预测器四部分组成. PointNet 提取半成品装配体部件以及候选部件库中各候选部件的点云特征, 编码器将半成品装配体的部件特征编码为记忆向量, 解码器将部件库中的候选部件特征编码为查询向量, 查询向量通过注意力机制进一步与记忆向量交互. 最后一部分的预测器包含两个结构相同的网络, 一个挑选可能性最大的缺失部件, 另一个预测其最佳的装配补全位姿, 最终将半成品补装完整. 稳定性判别器采用 2.3 所述预训练好的装配体稳定性判别器, 负责对补全的装配体进行稳定性检验, 检验结果以损失函数的形式反馈给三维装配补全网络 FiT, 实现基于稳定性的三维装配补全网络性能优化, 使补全的装配体更符合现实任务的需求.

StableFiT 的训练过程实际是一个多任务学习过程, 包括候选部件分类、位姿预测及稳定性优

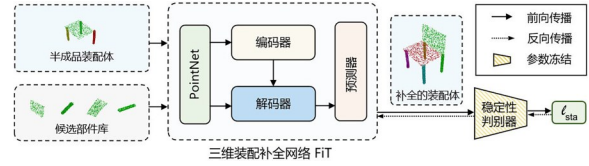


图 4 基于稳定性优化的三维装配补全方法 StableFiT
Fig. 4 StableFiT for 3D assembly completion with stability optimization

化. 本文分别为候选部件分类任务设计了分类损失 ℓ_c , 为位姿预测任务设计了平移损失 ℓ_t 、旋转损失 ℓ_r 和点云倒角距离损失 ℓ_s , 为稳定性优化任务设计了稳定性损失 ℓ_{sta} . 关于损失 ℓ_c , ℓ_t , ℓ_r 和 ℓ_s 的具体定义以及计算方式参见文献[14], 本文不再赘述.

本文设计的稳定性损失定义如下:

$$\ell_{sta} = \frac{1}{n} \sum_{i=1}^n |score_i - 1| \quad (4)$$

StableFiT 训练过程的总体损失定义如下:

$$\ell = \lambda_t \ell_t + \lambda_r \ell_r + \lambda_s \ell_s + \lambda_c \ell_c + \lambda_{sta} \ell_{sta} \quad (5)$$

其中, λ 表示损失权重, 各损失权重的设置参考 4.3.

4 实验分析

4.1 实验数据 使用 PartNet 数据集^[29]对提出的方法进行评估. PartNet 是一个大规模的三维模型数据集, 附有部件级别的实例分割标注. 选用其中的椅子和桌子类别的数据, 共包含 6323 把椅子和 8218 张桌子, 选取部件总数不超过 20 的椅子和桌子进行实验. 对每个部件使用最远点采样 (Furthest Point Sampling, FPS) 1000 个点, 并使用 PCA (Principal Component Analysis)^[30]将部件点云转换到规范空间.

4.2 评价指标及基准模型 采用匹配精度、点云倒角距离、部件精度、连接精度和不稳定装配体比例五种评价指标对实验结果进行评估.

匹配精度 (Match Accuracy, MA) 评估候选部件挑选的准确性:

$$MA = \frac{c}{j} \quad (6)$$

其中, j 为缺失部件数量, c 为匹配正确的部件数量.

点云倒角距离 (Shape Chamfer Distance, SCD) 和部件精度 (Part Accuracy, PA) 衡量装配补全完成的整体质量.

两个点云之间的倒角距离:

$$d_c(X, Y) = \sum_{x \in X} \min_{y \in Y} \|x - y\|_2 + \sum_{y \in Y} \min_{x \in X} \|x - y\|_2 \quad (7)$$

部件精度,即与真实值的点云倒角距离在一定阈值内的部件比例:

$$PA = \frac{1}{m} \sum_{i=1}^m (d_c(T_i(p_i), T_i^*(p_i^*)) < \tau_p) \quad (8)$$

其中,阈值 τ_p 设置为 0.01.

连接精度^[2](Connectivity Accuracy, CA)衡量补全的装配体中部件的连接情况:

$$CA = \frac{1}{|C|} \sum_{\{c_{ij}, c_{ji} \in C\}} (\|T_i(c_{ij}) - T_j(c_{ji})\|_2^2 < \tau_c) \quad (9)$$

其中, c_{ij} 为部件点云 p_i 中与部件点云 p_j 距离最近的点,即接触点; C 为半成品装配体与候选部件之间接触点对的集合; 阈值 τ_c 设置为 0.01.

稳定装配体比例 (Proportion of Stable Assembly, PSA) 评估补全的装配体的稳定性分布情况:

$$PSA = \frac{s}{a} \quad (10)$$

其中, s 代表稳定装配体的数量, a 代表所有补全的装配体的数量.

采用多种基准模型与提出的 StableFiT 方法进行对比,包括 Sung et al^[31] 的 Complement, Li et al^[11] 的 Single Image 和 Wang et al^[14] 的 FiT.

Complement 方法通过有无部件的对比学习方法来学习部分形状在特征空间中的表示. 本研究保留了其中基于 PointNet 的特征提取网络和位姿解码器,并通过在候选部件库中采样错误部件的对比学习方法,训练解码器完成部件的挑选和 6 自由度拼装位姿的预测.

Single Image 方法主要采用单张图片指导完成三维组装. 本研究删除了模型中的图像编码器部分,采用 PointNet 将缺失部件的装配体建模为全局特征,将候选部件建模为局部特征,并将二者结合以实现三维装配补全.

FiT 方法在模型训练过程中提出一种对称性损失,目的是使模型补全的装配体具有更好的对称性来近似模拟稳定性,但其不具有通用性,在现实场景中的适用范围受限,故本研究在对比实验中没有对基准模型采用该损失.

4.3 实验方法

4.3.1 稳定性判别器模型训练 利用 4.1 所述数据对基准模型 FiT^[14] 进行训练,训练轮次分别为 20, 40, 80, 160, 320 和最佳轮次的模型补全的装配体,将其如 2.2 所述输入 NVIDIA Isaac Sim 仿真环境中进行稳定性验证,得到带有稳定性标签的半成品装配体数据集. 对该数据集进行一定的处理,将其中稳定装配体与不稳定装配体的比例调整至 7:3,再利用处理后的数据集训练 2.3 提出的稳定性判别器模型. 训练过程中,设置初始学习率为 0.001, batch size 为 64, 训练轮次为 200. 对于椅子和桌子数据集分别按上述步骤训练得到对应的稳定性判别器模型,最终得到的椅子稳定性判别器模型的分类准确率可达 80%, 桌子稳定性判别器模型的分类准确率可达 76%.

以上是单部件缺失时的实验方法,对于多部件缺失的情况,处理步骤与上述基本一致,只是训练数据集的构建稍有不同. 这里选择训练轮次为 40, 80, 160 和最佳轮次的模型补全的装配体进行 NVIDIA Isaac Sim 仿真环境下的稳定性验证,得到带有稳定性标签的半成品装配体数据集,后续对于稳定性判别器模型训练参数的设置与上述一致. 最终训练得到的椅子稳定性判别器模型的分类准确率可达 84%, 桌子稳定性判别器模型的分类准确率可达 77%.

4.3.2 StableFiT 模型训练 采用 AdamW 优化器对 StableFiT 模型进行训练,设置初始学习率为 1.5×10^{-4} , batch size 为 32, 训练轮次为 500, 使用四张 NVIDIA V100 显卡进行训练. 单部件缺失和多部件缺失情况下的损失函数权重设置稍有不同,具体如表 1 所示. StableFiT 框架中的稳定性判别器模型直接使用 4.3.1 中训练好的装配体稳定性判别器模型,在训练过程中其参数被冻结,不参与此处的训练.

表 1 StableFiT 训练过程中损失函数权重的设置

Table 1 Weights settings for loss function during StableFiT training process

j	λ_r	λ_p	λ_s	λ_c	λ_{sta}
1	1.0	10.0	10.0	1.5	0.1
3	1.0	10.0	5.0	1.5	0.1

4.4 实验结果与分析 为了验证提出的基于稳定性优化的三维装配补全方法 StableFiT 的有效性,从两个角度设计实验并对实验结果进行分析:(1)验证提出的方法是否可以提升现有的三维装配补全网络的精度;(2)验证提出的方法是否能有针对性地提升补全的装配体的稳定性.

为了完成第一个验证,将 StableFiT 方法与基准模型 Complement, Single Image 和 FiT 在单部件缺失($j=1$)的情况下进行对比实验,同时与基准模型 FiT 在多部件缺失($j=3$)的情况下进行对比实验,实验结果如表 2 所示,表中箭头的方向表示更优的结果方向,黑体字表示结果最优.

表 2 本文方法与基准方法的性能的比较

Table 2 Performance of our method and the baseline methods

类别	j	方法	MA \uparrow	SCD \downarrow	PA \uparrow	CA \uparrow	PSA \uparrow
椅子	1	Complement	77.69	0.0332	14.87	15.21	—
		Single Image	81.75	0.0141	34.72	29.80	—
		FiT	87.71	0.0089	63.56	49.94	90.10
		StableFiT	88.47	0.0084	66.45	51.19	90.95
	3	FiT	70.77	0.0166	44.04	38.18	75.97
		StableFiT	72.61	0.0157	48.28	37.52	77.47
桌子	1	Complement	80.13	0.0136	28.89	40.72	—
		Single Image	82.39	0.0093	47.05	54.21	—
		FiT	92.56	0.0060	75.35	68.91	91.79
		StableFiT	91.83	0.0052	76.97	69.83	92.19
	3	FiT	62.91	0.0084	49.56	43.46	82.78
		StableFiT	68.06	0.0089	50.73	52.26	83.98

由表可见,单部件缺失时,StableFiT 的性能显著优于基准模型 Complement 和 Single Image,大多数指标,尤其是部件精度 PA 和连接精度 CA 的优势更明显.这说明 Single Image 在不进行基于部件分割的特征增强时,其性能会显著下降;Complement 的部件精度 PA 表现较差,是几何相似的部件对对比学习的干扰造成的.无论是在单部件缺失还是在多部件缺失的情况下,StableFiT 的性能都基本优于基准模型 FiT,验证了 StableFiT 提升现有三维装配补全网络精度的效果.其中,和 FiT 相比,StableFiT 对部件精度 PA 的提升最显著,在单部件和多部件缺失的情况下分别提

升 2.89% 和 4.24%.这是因为装配体的稳定性直接受部件装配质量的影响,稳定的装配体要求部件以准确的位姿完成补全,所以稳定性优化能直接促进部件精度的提升.部件是否被正确选择也会影响装配体的稳定性,所以引入稳定性优化后模型的匹配精度 MA 也得到了较大的提升.这些都再次证明了本文 StableFiT 方法的有效性.

实验结果还显示,各个方法在桌子类别数据上的性能普遍优于在椅子类别数据上的性能,这是因为桌子的结构相对简单,对称性更明显,且部件构成相对固定,不同种类桌子之间的结构化差异比椅子小,因而对桌子进行三维装配补全的难度低于椅子.虽然如此,StableFiT 对椅子和桌子两种类别的数据均能很好地适应,在这两种类别数据上的实验结果均得到了一定的提升.

另外还发现,引入稳定性优化后,模型在多部件缺失场景下的性能提升优于单部件缺失场景.这是因为多部件补全任务的难度相对较大,基准模型补全的不稳定装配体的比例会相对较高,所以 StableFiT 方法发挥作用的空间相对更大.

为了完成第二个验证,分别利用训练好的 StableFiT 模型与基准模型 FiT 对同一批半成品装配体进行补全,统计装配补全结果中稳定装配体所占的比例,结果如表 2 的 PSA 一列所示.分析实验结果可知,单部件缺失时,StableFiT 装配补全结果中稳定装配体所占的比例略高于 FiT;多部件缺失时,StableFiT 装配补全结果中稳定装配体所占的比例和 FiT 相比,提升比较明显,证明了 StableFiT 对提升装配体稳定性的效果.

为了直观说明 StableFiT 对装配体稳定性提升的效果,将 StableFiT 与基准模型 FiT 补全的装配体在 NVIDIA Isaac Sim 仿真平台上进行验证,图 5 展示了引入稳定性优化前后 FiT 模型补全的部分装配体的对比情况.图 5a 和图 5b 是单部件缺失时两种方法补全的装配体,图 5c 和图 5d 是多部件缺失时两种方法补全的装配体.每幅子图中左侧的装配体是由基准模型 FiT 补全得到的,右侧的装配体是利用 StableFiT 补全得到的.图中还通过放大局部细节来更好地对比展示装配体的稳定性情况,红色框展示的是装配体不稳定的细节,蓝色框展示引入稳定性优化后的改进情况.

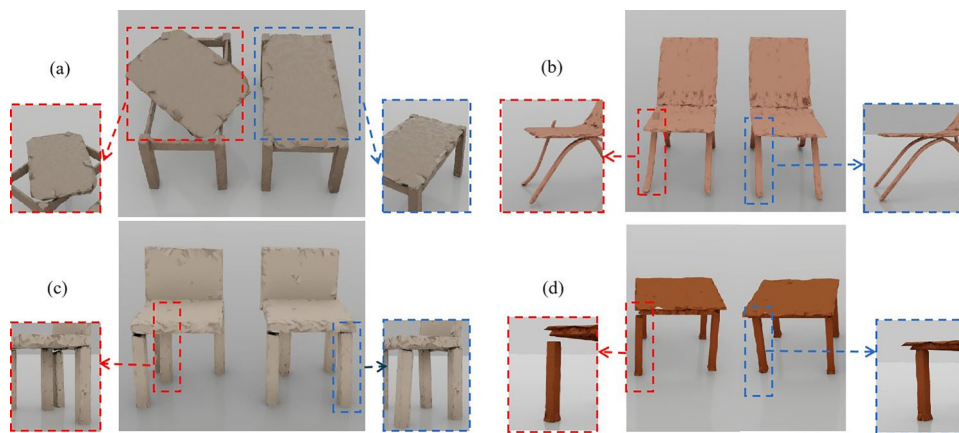


图5 StableFiT与FiT补全装配效果的对比

Fig. 5 Assembly completion results of StableFiT and FiT

图5a和图5b中由FiT模型补全的装配体都存在典型的拼装错误.如红色框中所示,图5a中装配体的桌面在装配补全时发生拼装位姿的错误,没有和桌子的边框对齐,图5b中装配体的椅腿在装配补全过程中拼装姿态发生错误,这些都会导致装配体的不稳定.引入稳定性优化后,利用StableFiT补全的装配体不存在上述问题,稳定性得到了提升,如蓝色框中所示.图5c中由FiT模型补全的装配体存在多部件补全中常见的部件选择错误,导致其在某个待补全位置没有椅腿,而在另一个待补全位置上有两条椅腿,这显然是不稳定的.引入稳定性优化后,装配体的部件被正确挑选,并以相对准确的位姿被装配到待补全位置上.图5d中由FiT模型补全的装配体存在典型的连接性错误,桌腿和桌面没有连接成整体,这也是一种不稳定现象.引入稳定性优化后,这种错误得到了针对性的解决.

综上,本文提出的基于稳定性优化的三维装配补全方法StableFiT,一方面可以提升现有三维装配补全网络的精度,补全出更高质量的装配体,另一方面,也能针对性地提升补全的装配体的稳定性,证明了稳定性优化的效果.

4.5 消融实验 为了充分探讨损失函数中各损失项在三维装配补全任务中的作用,在PartNet数据集中某一类别数据的单部件补全场景下设计了全面的消融实验,分析不同损失项对性能的影响.

通过单独删除损失函数中的每个损失项来评估其对性能的影响,实验结果如表3所示,表中箭

头的方向表示更优的结果方向,黑体字表示结果最优.由表可知,损失函数的每个损失项都有助于模型性能的提高.具体地,平移损失 ℓ_t 对部件拼装位姿的预测至关重要,删除 ℓ_t 之后部件精度和连接精度都大幅降低;旋转损失 ℓ_r 对部件正确挑选的影响更大,删除 ℓ_r 之后匹配精度比其他指标下降得更明显;点云倒角距离损失 ℓ_s 同样有助于部件以正确的位姿完成拼装,主要影响的是点云倒角距离,但删除 ℓ_s 之后匹配精度反而得到小幅的提升,综合所有指标来看, ℓ_s 对部件拼装的正向作用高于其潜在的负面作用,应在损失函数中保留.本文设计的稳定性损失 ℓ_{sta} 的添加,一方面使稳定装配体的比例上升,另一方面,也使其他的性能指标得到提升,充分证明了其有效性.

表3 各损失项对实验性能的影响

Table 3 Effect of loss components on experiment results

ℓ_t	ℓ_r	ℓ_s	ℓ_{sta}	MA \uparrow	SCD \downarrow	PA \uparrow	CA \uparrow	PSA \uparrow
\times	\checkmark	\checkmark	\checkmark	87.10	0.0106	49.35	36.96	82.31
\checkmark	\times	\checkmark	\checkmark	84.91	0.0086	64.73	49.63	90.58
\checkmark	\checkmark	\times	\checkmark	90.01	0.0100	63.00	48.68	91.20
\checkmark	\checkmark	\checkmark	\times	87.71	0.0089	63.56	49.94	90.10
\checkmark	\checkmark	\checkmark	\checkmark	88.47	0.0084	66.45	51.19	90.95

5 结论

三维装配补全任务尝试解决半成品装配体的部件补全问题,具有很高的实用价值.针对现有的三维装配补全方法较少考虑装配体稳定性的问题,本文提出一种基于稳定性优化的三维装配补

全方法 StableFiT,一方面可以提高补全后装配体的稳定性,使其更加满足机器人装配实施的需求,另一方面,也可对现有的三维装配补全过程进行性能优化,提升补全的装配体的整体质量.后续将继续完善稳定性验证方式,增强其合理性,并基于不同的装配场景设计特定的稳定性验证方式;还将把本文提出的稳定性优化方法拓展到其他装配任务中,如三维部件装配.

参考文献

- [1] Zhang R F, Kong T, Wang W H, et al. 3D part assembly generation with instance encoded transformer. *IEEE Robotics and Automation Letters*, 2022, 7(4): 9051—9058.
- [2] Zhan G Q, Fan Q N, Mo K C, et al. Generative 3D part assembly via dynamic graph learning// *Proceedings of the 34th International Conference on Neural Information Processing Systems*. 2020: 6315—6326.
- [3] Chaudhuri S, Kalogerakis E, Guibas L, et al. Probabilistic reasoning for assembly - based 3D modeling// *Proceedings of the ACM SIGGRAPH 2011*. Vancouver, Canada: ACM, 2011: Article No.35.
- [4] Jaiswal P, Huang J M, Rai R. Assembly - based conceptual 3D modeling with unlabeled components using probabilistic factor graph. *Computer - Aided Design*, 2016(74): 45—54.
- [5] Kalogerakis E, Chaudhuri S, Koller D, et al. A probabilistic model for component - based shape synthesis. *ACM Transactions on Graphics*, 2012, 31(4): 55.
- [6] Harish A N, Nagar R, Raman S. RGL - NET: A recurrent graph learning framework for progressive part assembly// *Proceedings of 2022 IEEE/CVF Winter Conference on Applications of Computer Vision*. Waikoloa, HI, USA: IEEE, 2022: 647—656.
- [7] Wu R D, Zhuang Y X, Xu K, et al. PQ - NET: A generative part Seq2Seq network for 3D shapes// *Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, WA, USA: IEEE, 2020: 829—838.
- [8] Li J, Niu C J, Xu K. Learning part generation and assembly for structure - aware shape synthesis// *Proceedings of the 34th AAAI Conference on Artificial Intelligence*. New York, NY, USA: AAAI Press, 2020: 11362—11369.
- [9] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of wasserstein GANs// *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, CA, USA: Curran Associates Inc., 2017: 5769—5779.
- [10] Adler J, Lunz S. Banach wasserstein GAN// *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. Montréal, Canada: Curran Associates Inc., 2018: 6755—6764.
- [11] Li Y C, Mo K C, Shao L, et al. Learning 3D part assembly from a single image// *Proceedings of the 16th European Conference on Computer Vision*. Springer Berlin Heidelberg, 2020: 664—682.
- [12] Wu C, Zheng J, Pfrommer J, et al. Attention - based part assembly for 3D volumetric shape modeling// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Vancouver, Canada: IEEE, 2023: 2716—2725.
- [13] Xu X H, Guerrero P, Fisher M, et al. Unsupervised 3D shape reconstruction by part retrieval and assembly// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada: IEEE, 2023: 8559—8567.
- [14] Wang W H, Zhang R F, You M Y, et al. 3D assembly completion// *Proceedings of the 37th AAAI Conference on Artificial Intelligence*. Washington DC, USA: AAAI Press, 2023: 2663—2671.
- [15] Tchapmi L P, Kosaraju V, Rezatofighi H, et al. TopNet: Structural point cloud decoder// *Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, CA, USA: IEEE, 2019: 383—392.
- [16] Yuan W T, Khot T, Held D, et al. PCN: Point completion network// *Proceedings of 2018 International Conference on 3D Vision*. Verona, Italy: IEEE, 2018: 728—737.
- [17] Wen X, Li T Y, Han Z Z, et al. Point cloud completion by skip - attention network with hierarchical folding// *Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern*

- Recognition. Seattle, WA, USA: IEEE, 2020: 1939—1948.
- [18] Sarmad M, Lee H J, Kim Y M. RL-GAN-Net: A reinforcement learning agent controlled GAN network for real-time point cloud shape completion//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019: 5898—5907.
- [19] Hu T, Han Z Z, Shrivastava A, et al. Render4 Completion: Synthesizing multi-view depth maps for 3D shape completion//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision Workshop. Seoul, Korea (South): IEEE, 2019: 4114—4122.
- [20] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA, USA: Curran Associates Inc., 2017: 6000—6010.
- [21] Yan X G, Lin L Q, Mitra N J, et al. ShapeFormer: Transformer - based shape completion via sparse representation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 6239—6249.
- [22] Mittal P, Cheng Y C, Singh M, et al. AutoSDF: Shape priors for 3D completion, reconstruction and generation//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA: IEEE, 2022: 306—315.
- [23] van den Oord A, Vinyals O, Kavukcuoglu K. Neural discrete representation learning//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA, USA: Curran Associates Inc., 2017: 6309—6318.
- [24] Esser P, Rombach R, Ommer B. Taming transformers for high - resolution image synthesis//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2021: 12873—12883.
- [25] Wen X, Xiang P, Han Z Z, et al. PMP-Net++: Point cloud completion by transformer-enhanced multi-step point moving paths. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(1): 852—867.
- [26] Li S S, Gao P, Tan X Y, et al. ProxyFormer: Proxy alignment assisted point cloud completion with missing part sensitive transformer//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE, 2023: 9466—9475.
- [27] 国家市场监督管理总局, 国家标准化管理委员会. 办公家具 桌台类 稳定性、强度和耐久性测试方法. GB/T 38607-2020. (State Administration of Market Supervision and Administration of the People's Republic of China, Standardization Administration of the People's Republic of China. Office furniture: Tables and desks. Test methods for the determination of stability, strength and durability. GB/T 38607-2020.)
- [28] Qi Charles R, Su H, Mo K C, et al. PointNet: Deep learning on point sets for 3D classification and segmentation//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 652—660.
- [29] Mo K C, Zhu S L, Chang A X, et al. PartNet: A large - scale benchmark for fine - grained and hierarchical part: Level 3D object understanding//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019: 909—918.
- [30] Pearson K. LIII. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1901, 2(11): 559—572.
- [31] Sung M, Su H, Kim V G, et al. ComplementMe: Weakly - supervised component suggestions for 3D modeling. ACM Transactions on Graphics, 2017, 36(6): 226.

(责任编辑 杨可盛)