

# 基于多编码器和 Residual-Transformer 的点云补全网络

高辉, 杨志景\*, 凌永权, 曹江中, 李为杰

广东工业大学信息工程学院, 广东 广州 510006

**摘要** 点云数据具有无序和稀疏的特点。通过不完整点云数据恢复丢失的三维几何形状的 3D 点云补全任务是 3D 视觉技术中一个具有挑战性的问题。现有的 3D 点云补全网络一般都通过编码器-解码器(Encoder-Decoder)模型直接从部分点云预测完整的点云形状, 这会干扰原始部分点云, 引入噪声, 导致几何位移损失。因此提出一个端到端的网络模型, 集中生成平滑和分布均匀的点云对象。所提网络模型主要包含三部分: 缺失点云预测、点云融合和点云平滑。第一个模块主要通过多编码器从残缺的点云对象提取局部和全局信息, 预测缺失几何部分。第二个模块通过采样算法融合点云。第三个模块基于 Residual-Transformer (RT) 预测点位移, 在避免破坏原始输入点云的空间结构下, 可以使点分布得更加均匀。在基准数据集 Shapenet-Part 上, 大量的实验结果表明, 所提网络在 3D 形状补全方面取得了更好的量化结果和更好的视觉效果。

**关键词** 点云; 3D 点云补全; 自注意力机制; 多编码器; 残差网络

中图分类号 TP242

文献标志码 A

DOI: 10.3788/LOP213314

## Point Cloud Completion Network Based on Multiencoders and Residual-Transformer

Gao Hui, Yang Zhijing\*, Ling Wing-Kuen, Cao Jiangzhong, Li Weijie

School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, Guangdong, China

**Abstract** Point cloud data has the characteristics of disorder and sparsity. The three-dimensional (3D) point cloud completion task of recovering the missing 3D geometric shapes through incomplete point cloud data is a challenging issue in 3D vision technology. The existing 3D point cloud completion network predicts the complete point cloud shape directly from a subset of the point cloud using the Encoder-Decoder model, which interferes with the original part of the point cloud, resulting in noise and geometric displacement loss. In this study, an end-to-end network model is proposed, which focuses on generating a smooth and uniformly distributed point cloud object. The proposed network model mainly consists of the following three parts: missing point cloud prediction, point cloud fusion, and point cloud smoothing. The first module mainly uses multiencoders to extract local information and global information from incomplete point cloud objects to predict the missing geometric parts. The second module merges point cloud objects by sampling algorithm. The third module is based on a Residual-Transformer (RT) to predict the displacement of the points, which can make the point distribution more uniform without destroying the spatial structure of the original input point cloud. On the benchmark dataset, Shapenet-Part, several experimental results indicate that the proposed network has achieved better quantitative results and visual effects in 3D shape completion.

**Key words** point cloud; three-dimensional point cloud completion; self-attention mechanism; multiencoders; residual network

收稿日期: 2021-12-22; 修回日期: 2022-02-09; 录用日期: 2022-03-14; 网络首发日期: 2022-03-24

基金项目: 国家自然科学基金(U1701266)、广东省自然科学基金(2021A1515011341)、广州市科技计划(202002030386)

通信作者: \*yzhj@gdut.edu.cn

# 1 引言

3D 数据应用于许多不同领域,包括自动驾驶、机器人、遥感等。然而,要想从现实世界的扫描中获得高保真度的 3D 模型是非常困难的,这不仅依赖传感器的能力,还需要良好的扫描视角。因此,近年来,基于点云表示的 3D 数据已广泛应用于计算机视觉<sup>[1-2]</sup>、机器视觉<sup>[3-4]</sup>、遥感与传感器<sup>[5-6]</sup>等领域。点云是在同一空间参考系下表达目标空间分布和表面特征的海量点的集合。点云具有非常统一的结构,可以避免组合的不规则性和复杂性。本文提出一种点云补全细化网络,通过将预测的缺失部分和原始输入融合再进行细化而不是重新预测整个 3D 点云来完成 3D 点云形状补全。

目前,3D 形状补全表示主要包括 4 种类型:深度图<sup>[6-7]</sup>、网格<sup>[8]</sup>、体素<sup>[9-10]</sup>和点云<sup>[11-12]</sup>。其中,深度图不足以表示和重构原始输入信息,网格表示过于复杂,基于体素的表示计算复杂度高。为了弥补这些表示方法的不足,研究者们提出了许多基于 3D 点云的形状补全,如 point cloud completion (PCN)<sup>[13]</sup>和 morphing and sampling network (MSN)<sup>[14]</sup>,都是利用全局特征进行点云补全的。然而,这些方法都专注于预测整个 3D 对象,破坏了输入点云的结构特征。点云的无序性要求网络对点云输入的顺序不敏感,使用卷积神经网络 (CNN) 直接处理点云存在一定的挑战。Transformer 是一种基于自注意力 (SA) 机制的深度神经网络,已被用于自然语言处理 (NLP)<sup>[15-16]</sup>、目标检测<sup>[17]</sup>、图像去噪<sup>[18]</sup>等领域。作为 Transformer 的核心部分,注意力机制本身就是一个算子,其排列不变,不依赖于点之间的连接关系,非常适合点云的学习。文献<sup>[19]</sup>构建了一个应用于点云的自注意力网络,以实现对象分类和语

义场景分割。文献<sup>[20]</sup>提出了一种点云转换器 (PCT) 网络,它可以更好地提取点云的全局和局部特征,并在分类和分割方面表现出良好的性能。

综上,以往的大部分工作通常更注重通过全局特征预测整个完整的 3D 点云对象。因此,它们可能会改变初始点云的相对位置并造成几何位移损失,忽略了细节信息。为了解决上述问题,本文提出一种与之前的网络直接通过全局结构信息来补全 3D 形状不同的模型,采用多编码器结构,分别提取输入点云对象的局部特征和全局特征来预测缺失部分,保留更多的细节信息。由于直接将预测的缺失部分和原始残缺输入点云相融合会导致点云分布不均,受 PCT<sup>[20]</sup>中使用的基于 Transformer 的编码器的启发,本文设计了一个基于 Residual-Transformer 的点云细化网络,通过 Residual-Transformer 提取重要特征,决定哪些点需要改变以提高补全的质量,解决融合部分存在的间隙问题。

# 2 基本原理

网络的完整架构如图 1 所示。所提网络模型主要由 3 部分组成:缺失点云预测 (MPCP)、点云融合 (PCM) 和点云细化网络 (PRN)。提出的多编码器框架是一种端到端的三维点云补全神经网络结构。首先,通过不同编码器分别提取输入点云对象的局部特征和全局特征,更好保留了全局结构和局部结果细节信息。然后,通过一个变形解码器只预测点云残缺部分,对其与原始点云进行合并再采样,得到一个初步完整点云,防止破坏原始点云结构。最后,将初步生成的完整点云输入一个基于 Transformer 的点云优化网络,便可得到一个完整且均匀分布的点云模型。Residual-Transformer 块可以减少特征在网络传播中的损失,选

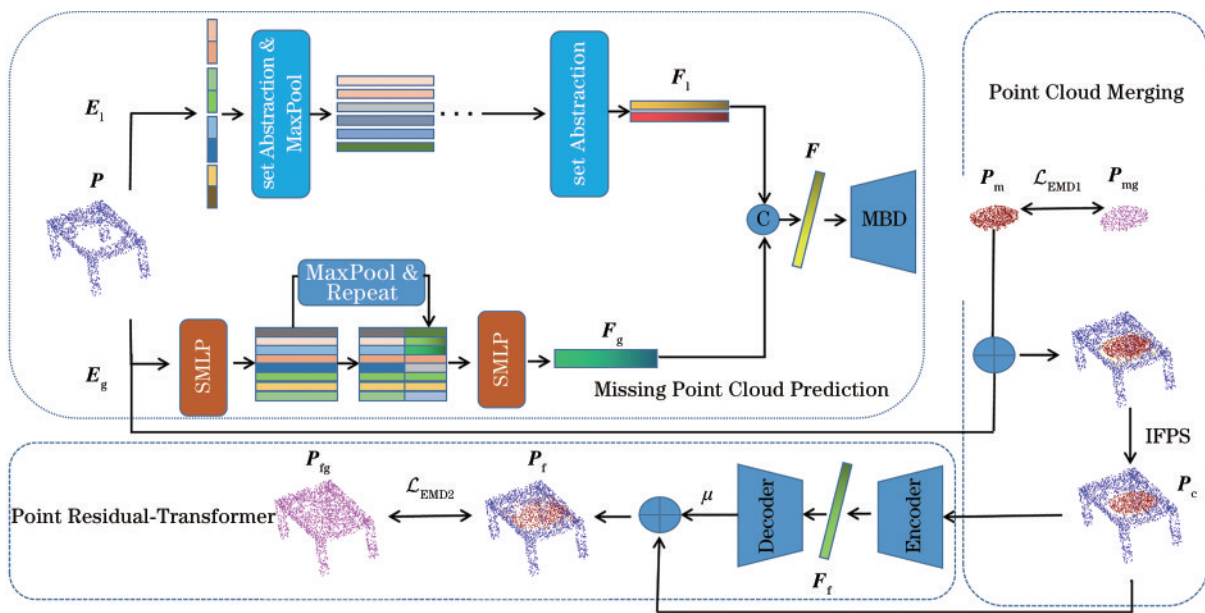


图 1 基于多编码器和 Residual-Transformer 的点云补全网络

Fig. 1 Point cloud completion network based on multi-encoders and Residual-Transformer

择性地传递局部区域的几何信息。

### 2.1 缺失点云预测

这一部分的主要目标是根据输入的残缺点云预测缺失部分的几何结构。为了充分利用输入点云的局部特征和全局特征,设计一个多编码器网络来提取这些特征,主要包含两个独立分支  $E_g$  和  $E_l$ , 结构如图 1 所示。分支  $E_g$  的结构类似于 PCN 的编码器,但是这里只用一半的网络层数。输入点云经两层共享多层感知机(SMLP)和最大池化层(MaxPool),得到一个维度为 512 的全局特征向量  $F_g$ , 计算公式为

$$F_g = \text{MaxPool}[\text{SMLP}(\mathbf{P})], \mathbf{P} \in \mathbb{R}^{N \times 3}, F_g \in \mathbb{R}^{512 \times 1}, \quad (1)$$

式中:  $\mathbf{P}$  表示输入的 3D 点云对象;  $N$  表示点的数量。

在 PointNet<sup>[21]</sup> 中,一个显著的问题是对局部特征问题的处理,在对场景进行分割时,需要得到每个点的得分,处理办法是直接点特征与全局特征结合到一块进行处理,忽略了局部特征。而 PointNet++<sup>[22]</sup> 在 PointNet 的基础上采用了分层的结构处理局部特征。因此采用 PointNet++ 的分层特征学习架构提取输入点云对象  $\mathbf{P}$  不同尺度的特征。但不同的是, PointNet++ 利用多层点集抽象(PSA)和 MaxPool 提取全局特征。本文只利用 4 层 PSA 并去除掉最后一层 PSA 的 MaxPool 来获取点的局部特征  $F_l$ , 计算公式为

$$F_l = \text{PSA}(\mathbf{P}), F_l \in \mathbb{R}^{512 \times N_l}, \quad (2)$$

式中:  $N_l$  表示在最后一层 PSA 中,通过最远点采样得到需要提取的局部特征的点的数目,  $N_l = 32$ 。

然后,将局部特征  $F_l$  和全局特征  $F_g$  融合,得到特征向量  $F$ , 计算公式为

$$F = \text{MaxPool}[\text{concat}(F_g, F_l)], F \in \mathbb{R}^{1024 \times 1}, \quad (3)$$

式中:  $\text{concat}(\cdot)$  是串联器。将全局特征向量  $F_g$  复制  $N_l$  次,得到  $F_g \in \mathbb{R}^{512 \times N_l}$ , 再与局部特征向量  $F_l$  串联,即可得到  $F$ 。

特征向量  $F$  通过解码器(Decoder)即可生成缺失部分的点云,本文采用一种基于变形的解码器(MBD)<sup>[14]</sup>。将生成的缺失点云用  $\mathbf{P}_m$  表示,维度大小为  $M \times 3$  ( $M = N/2$ )。对于基于变形的解码器,使用文献[14]中描述的架构,该解码器由  $K$  个(实验中为 16 个)变形网络组成,每个网络将单位正方形  $[0, 1]^2$  采样得到的 2D 表面映射到一个 3D 表面。在每次前向传输中,在单位正方形中随机采样  $M/K$  个点,在通过  $K$  个变形网络之前,首先将特征向量  $F$  连接到每个 2D 点坐标;然

后将每个采样的 2D 点映射成  $K$  个不同的散落在  $K$  个 3D 表面的点,产生  $M$  个点来描述预测的形状。每个变形网络有 4 个块:前 3 个块由卷积、批量归一化(BN)和 ReLU 激活函数组成;最后 1 个块由 1 个卷积层和 1 个 Tanh 激活函数组成。

### 2.2 点云融合

使用 MBD, 网络可以预测缺失部分的几何结构,因此将预测的缺失点云和网络输入合并,便可得到一个完整的点云对象,然而由于网络功能有限,可能会忽略某些结构。其次,一致性等问题会使最终的合并结构可能变得不切实际。因此合并操作存在 2 个问题:生成的缺失点云  $\mathbf{P}_m$  中点的密度是高于输入点云  $\mathbf{P}$  中点的密度的,因为预测的点云  $\mathbf{P}_m$  只是不完整点云  $\mathbf{P}$  所缺失的一个很小的区域,但是  $\mathbf{P}_m$  却拥有  $\mathbf{P}$  一半的点;拼接的点云的大小与真实点云模型的大小不匹配。为了解决这些问题,使用采样的方法来调整 3D 点云对象中点的分布,合并点云的大小。由于两个点云对象的密度不同且可能重叠,如果直接将其合并得到的是一个分布不均的点云对象,因此希望从合并的点云中采样具有均匀分布的点云子集。点云采样算法有很多,例如最远点采样(IFPS)和泊松盘采样(PDS)。这里使用 IFPS, 因为这种方法已经有效地应用于 PointNet++, 合并和采样架构如图 1 所示。合并和采样的过程可表示为

$$\mathbf{P}_c = \text{IFPS}(\mathbf{P}_m \cup \mathbf{P}), \mathbf{P}_c \in \mathbb{R}^{N \times 3}. \quad (4)$$

通过合并,得到一个大小为  $\frac{3}{2} \times N \times 3$  的点云,再

通过 IFPS 算法得到一个均匀分布的完整点云对象  $\mathbf{P}_c$ 。

### 2.3 点云细化网络

直接将非完整的输入点云  $\mathbf{P}$  与预测的缺失点云  $\mathbf{P}_m$  合并,可能会导致得到的完整点云对象在合并处存在缝隙。因此,所提点云细化网络的目标是通过平滑两个合并点云之间的间隙来解决这个问题。此外,细化网络还可以增强点云中的点在对象中的最终分布。PRN 的架构如图 2 所示。它主要包含一个残差转换器(RT)模块和一些卷积层。受文献[20]启发,引入其中的偏移注意力(OA)作为所提点云细化网络的 Transformer 块。相较于原始 SA, OA 的工作原理是用 SA 模块的输入和注意力特征之间的偏移量替换注意力特征,这使得同一个物体的相对坐标通常更加稳健。

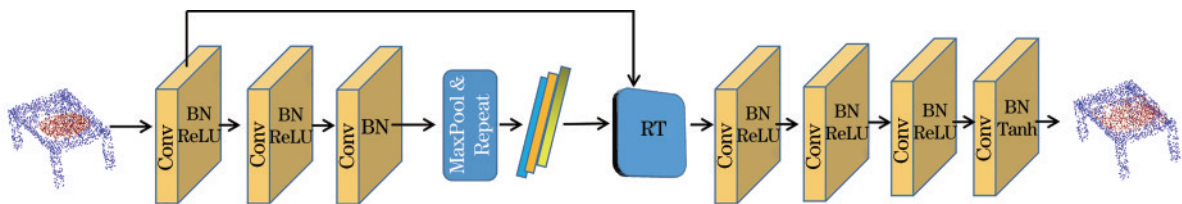


图 2 PRN 结构

Fig. 2 Structure of PRN



PRN 模块的输入是来自点云融合模块的  $P_c$  点云对象。PRN 由 Encoder、Decoder 和一个 Residual-Transformer 块组成。其中 Encoder 主要由 3 层卷积层组成:第 1、2 层由一维卷积(1D-Conv)、BN 和 ReLU 激活函数组成;第 3 层由 1D-Conv、BN 和 MaxPool 组成。经过 Encoder,可以得到一个特征通道为 1024 的全局特征向量  $F_f$ ,表达式为

$$F_f = \text{MaxPool}[\text{Conv}(P_c)], F_f \in \mathbb{R}^{1024 \times 1}, \quad (5)$$

式中:Conv( $\cdot$ )表示卷积操作。

Residual-Transformer(RT)可以在网络传播中有效地利用点云的局部结构细节,选择性地从点云的局部区域传递几何信息。因此,在解码全局特征之前,将  $F_f$  复制  $N$  次,使其维度变成  $1024 \times N$ ,然后 RT 对 Decoder 的全局点特征  $F_f$  与 Encoder 的第 1 层卷积得到的特征通道为 64 的局部点特征进行融合,得到一个融合局部特征与全局特征的特征向量  $F_c$ ,公式为

$$F_c = \text{RT}[\text{concat}(F_f, F_1)], F_1 \in \mathbb{R}^{64 \times N}, F_c \in \mathbb{R}^{1088 \times N}. \quad (6)$$

$F_c$  经过由 4 层卷积组成的 Decoder,便可得到移位的 3D 点云模型  $\Delta P_c$ 。Decoder 的前 3 层由 1D-Conv、BN 和 ReLU 激活函数组成,最后一层由 1D-Conv、BN 和 Tanh 激活函数组成。

PRN 实质上就是学习一个参数函数来改进采样点集  $P_c$ 。使用 PRN 来预测位移  $P_c$ ,将  $P_c$  移动到更理想的位置,从而改善  $P_c$  内点的分布。最终输出一个平滑完整的点云  $P_f$ ,计算公式为

$$P_f = P_c + \mu \Delta P_c, \quad (7)$$

式中: $\mu \in [0, 1]$  是控制所提网络想要保持的位移量的超参数,设为 1。

## 2.4 损失函数

使用倒角距离(CD)<sup>[23]</sup>和地移距离(EMD)<sup>[24]</sup>函数作为网络训练的损失函数及实验数据的评价指标,将生成的点云与真实点云进行比较。EMD 是两个对象分布之间的度量,表示将一个分布转换为另一个分布所必需的最低成本。对于点集  $X$  和  $Y$ ,EMD 可以定义为

$$\mathcal{L}_{\text{EMD}} = \min_{\varnothing: Y \rightarrow X} \left\| \sum_{y \in Y} y - \varnothing(y) \right\|_2, \quad (8)$$

式中: $\varnothing$  是一个双射; $Y$  代表真实的 3D 点云; $X$  代表网络生成的 3D 点云。

EMD 对细节和密度分布更敏感。而 CD 通常用于测试两个点集之间的排列不变距离,并且 CD 是完全可微的,且计算量较少,表示两个点集之间最近的平方根距离,定义为

$$\mathcal{L}_{\text{CD}} = \sum_{x \in X} \min_{y \in Y} \|x - y\|_2^2 + \sum_{y \in Y} \min_{x \in X} \|x - y\|_2^2. \quad (9)$$

在第一部分网络中,虽然 MBD 可以学习从 2D 到 3D 的连续映射,在 2D 中密集采样产生较光滑的表面,

然而没有明确地防止具有相同坐标的点云的产生,表面点云之间的重叠或过度分散是不可避免的。尽管 EMD 可以引导一系列 MLP 以覆盖缺失点云的形状,点云密度分布由于过度重叠或过度分散而使点云仍然不均匀,这也使得表面元素扩展和覆盖更大的区域,2D 单位正方形难以变形并捕获局部细节。为此,使用一个延长惩罚损失(extended penalty loss)<sup>[14]</sup>作为表面元素的约束器,它使每个表面元素紧凑并集中在局部区域,公式为

$$\mathcal{L}_{\text{expansion}} = \frac{1}{KN} \sum_{1 \leq i \leq K} \sum_{(u,v) \in T_i, \rho} \{ \text{dis}(u,v) \geq \gamma l_i \} \text{dis}(u,v), \quad (10)$$

式中: $K$  表示定向最小生成树的数量; $T_i$  表示点  $i$  构建的最小生成树,描述每个 MLP 的点分布; $\text{dis}(u,v)$  是  $u$  和  $v$  之间的欧氏距离; $l_i = \left[ \sum_{(u,v) \in T_i} \text{dis}(u,v) \right] / (N-1)$  是  $T_i$  中边的平均长度; $\rho$  是长度小于  $\gamma l_i$  的滤波器边缘索引函数, $\gamma = 1.5$ 。

在所提网络框架中,主要通过计算预测缺失点云和细化网络的损失来训练模型。第一个损失函数计算了预测的残缺点云  $P_m$  和真实点云  $P_{\text{mg}}$  之间的 EMD 及表面元素的约束损失。第二个损失是生成的完整平滑点云  $P_f$  和完整真实点云  $P_{\text{fg}}$  之间的 EMD。因此,最终联合损失可以表示为

$$\mathcal{L} = \mathcal{L}_{\text{EMD2}}(P_m, P_{\text{mg}}) + \theta \mathcal{L}_{\text{expansion}} + \mathcal{L}_{\text{EMD1}}(P_f, P_{\text{fg}}), \quad (11)$$

式中: $\theta$  是延长惩罚损失的权重,这里设  $\theta = 1$ 。

## 3 分析与讨论

### 3.1 不同方法的补全效果对比

在数据集 Shapenet-Part<sup>[25]</sup>上验证所提模型,从基准数据集中选择 13 个不同的类别进行训练测试。这些类别包括总共 14473 个 CAD 模型:飞机(Airplane)、包(Bag)、帽子(Cap)、汽车(Car)、桌子(Table)、灯(Lamp)、笔记本(Laptop)、摩托车(Motorbike)、吉他(Guitar)、椅子(Chair)、杯子(Mug)、滑板(Skateboard)和手枪(Pistol)。使用 11705 个样本进行训练,使用 2768 样本进行测试。所提框架在 Pytorch 1.3.1 和 NVIDIA RTX 2080Ti GPU 中实现。在训练过程中,使用批量大小为 20 的 ADAM 优化器,初始学习率设置为 0.0005,网络的迭代训练次数为 200。图 3 为所提方法在 Shapenet-Part 数据集中对不同类型样本补全的整个过程:先生成缺失点云,然后将其与原始输入部分点云相融合再采样,最后通过细化网络生成更精细和分布均匀的完整 3D 点云。

将所提方法与几种最先进的方法在 Shapenet-Part 测试数据集上进行定量和定性比较。为了让所有实验有一个公平的比较,使用相同的数据分区和相同的设

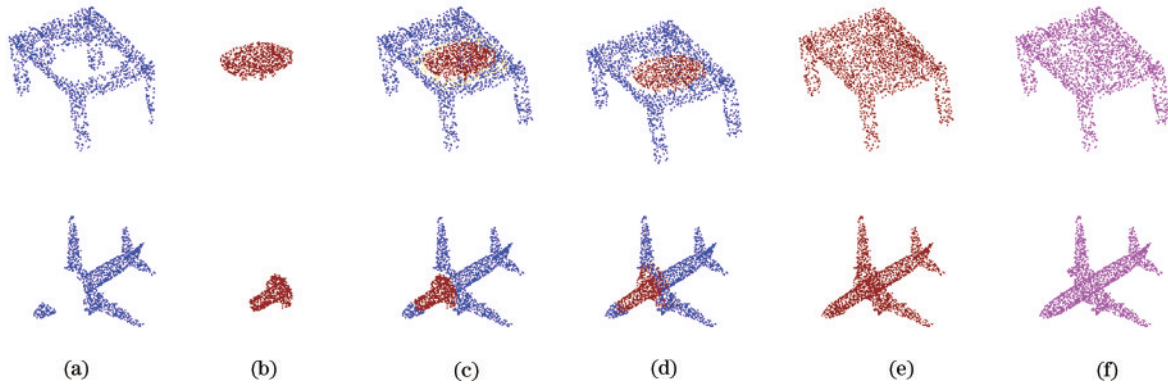


图3 所提模型的整个补全流程。(a)原始点云;(b)预测的缺失部分;(c)合并图3(a)和图3(b)的完整点云;(d)采样后的点云;(e)强化后的完整点云;(f)真实点云

Fig. 3 Entire completion process of the proposed model. (a) Original point cloud; (b) predicted missing part; (c) complete point cloud after merging Fig. 3 (a) and Fig. 3 (b); (d) point cloud after sampling; (e) complete point cloud after enhancement; (f) ground truth

置训练了所有方法,所有网络输入的非完整点云都是在完整真实点云模型上缺失 25% 的点。几种最先进的方法有:变形和采样网络(MSN)<sup>[14]</sup>,通过变形网络生成粗略的完整点云模型,使用采样算法和残差网络进一步使最终预测的点云均匀分布;点分形网络(PF-Net)<sup>[1]</sup>,通过多分辨编码器提取多尺度特征,利用金字塔式的解码器预测 3D 点云模型缺失的部分而不是点云的整体形状;全卷积自动编码器(FCAE)<sup>[26]</sup>,使用卷积层构造编码器,解码器由全连接层组成,通过构造的自动编码器直接预测完整的点云而不是缺失的部分;预测缺失部分的细化网络(RP)<sup>[26]</sup>,涉及一个包含缺失部分预测和融合细化的网络,提出的 RP-MLP 将多层感知机作为解码器来预测缺失的几何结构,再通过残差网络平滑融合的点云。

使用 PF-Net 中的评价指标:  $\text{Pred} \rightarrow \text{GT}$  误差和  $\text{Pred} \leftarrow \text{GT}$  误差。两个误差的总和即是 CD 距离。

前一个误差计算从预测点云模型中的每个点到真实点云中最近点的平均平方距离,用来衡量预测与真实情况的差异。后者用于计算从真实点云中的每个点到预测点云模型中点的最近点的平均平方距离,衡量真实点云模型与网络预测的差异程度。实验结果如表 1 所示,所有的指标都是数值越小越好,最佳结果以粗体显示。相比 MSN、PF-Net 和 FCAE,所提方法不论是  $\text{Pred} \rightarrow \text{GT}$  还是  $\text{Pred} \leftarrow \text{GT}$  指标,都取得了明显提升;RP-MLP 虽然对某些类的  $\text{Pred} \leftarrow \text{GT}$  指标略好于所提方法,但在  $\text{Pred} \rightarrow \text{GT}$  评价指标上,所提方法对所有类别都显著优于 RP-MLP,且在均值上,分别下降了 33.0% 和 3.0%。表 1 中的实验结果表明,与其他方法相比,所提方法可以实现更高精度的量化结果。选择一些样本进行可视化,结果如图 4 所示。

表 1 不同方法的  $\text{Pred} \rightarrow \text{GT}/\text{Pred} \leftarrow \text{GT}$  指标,结果都被放大 10000

Table 1  $\text{Pred} \rightarrow \text{GT}/\text{Pred} \leftarrow \text{GT}$  of different methods, results are magnified by 10000

Category	MSN <sup>[14]</sup>	PF-Net <sup>[1]</sup>	FCAE <sup>[26]</sup>	RP-MLP <sup>[26]</sup>	Proposed method
Airplane	3.113/3.500	4.604/2.997	6.980/7.047	2.892/ <b>2.212</b>	<b>2.164</b> /2.641
Bag	15.393/7.942	21.350/7.184	34.774/29.045	14.232/5.748	<b>7.127/3.340</b>
Cap	16.824/5.842	28.984/3.719	52.914/37.094	17.016/ <b>2.527</b>	<b>6.950</b> /3.065
Car	8.443/6.991	9.643/3.206	19.982/15.550	8.683/ <b>2.516</b>	<b>6.103</b> /2.592
Chair	5.934/4.434	7.803/2.774	16.408/13.733	6.049/2.051	<b>4.107/2.050</b>
Guitar	5.226/3.365	3.195/3.068	3.276/3.185	1.822/2.381	<b>1.693/2.168</b>
Lamp	19.873/9.819	24.143/10.559	39.082/21.319	15.675/ <b>9.389</b>	<b>14.506</b> /10.085
Laptop	3.543/3.910	4.879/2.262	9.252/9.623	2.832/1.430	<b>2.110/1.195</b>
Motorbike	6.802/7.178	6.473/5.028	14.094/11.592	6.197/4.162	<b>5.188/3.701</b>
Mug	8.199/6.340	10.538/3.774	25.223/22.119	7.464/ <b>3.054</b>	<b>5.481</b> /3.316
Pistol	5.092/6.349	7.613/4.944	10.898/10.115	5.399/ <b>3.952</b>	<b>3.350</b> /4.051
Skateboard	3.441/3.766	5.562/2.788	8.981/8.996	2.693/ <b>1.994</b>	<b>1.812</b> /2.101
Table	7.257/4.709	9.919/3.454	19.712/17.699	6.722/2.737	<b>4.865/2.513</b>
Average	8.395/5.704	11.131/4.289	20.121/15.932	7.514/3.396	<b>5.035/3.294</b>

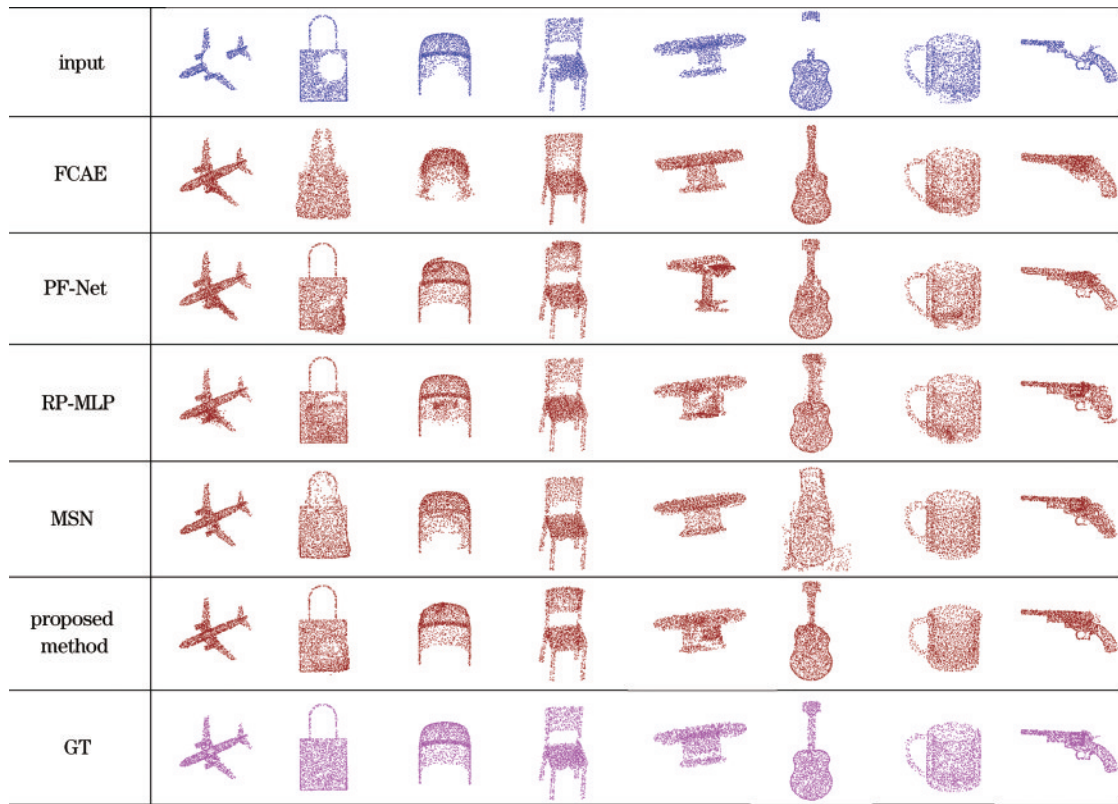


图 4 不同方法在 Shapenet-Part 数据集上的补全可视化效果对比

Fig. 4 Comparison of completion visualization effects of different methods on Shapenet-Part dataset

FCAE 只是用一个简单的编码器-解码器提取特征和补全点云,且随着网络层数的加深,会造成特征损失,因此它在预测完整的点云时会丢失几何结构,例如 Bag 和 Cap;虽然 PF-Net 提出的多分辨率模型只预测缺失部分,保留了原始输入点云的空间结构,但没有对融合部分进行平滑处理,因此在预测的可视化结果中可能存在间隙,例如 Bag;MSN 虽先生成粗糙的点云模型,后期再用残差网络来优化,但前期只提取了输入残缺点云的全局特征,因此会丢失一些细节信息,如 Airplane、Guitar 和 Table;RP-MLP 通过一个自编码器预测缺失部分的点云,再通过残差网络平滑融合的点云模型,一定程度上可以使生成的点云均匀分布,但忽视了局部特征的重要性,导致一些局部信息丢失,如

Chair 和 Pistol;所提多编码器框架分别提取输入点云的全局和局部特征信息,通过 Residual-Transformer 选择性地保留局部细节信息,可以生成平滑完整和分布均匀的点云模型。

### 3.2 消融实验

在 Shapenet-Part 数据集上进行消融实验,通过引入和不引入 Residual-Transformer 模块来分析对点云补全算法性能的影响。如表 2 所示,当网络去掉 RT 模块时,不论是  $\text{Pred} \rightarrow \text{GT}$  还是  $\text{Pred} \leftarrow \text{GT}$  评价指标,实验结果都出现明显下滑,均值分别下降 31.9% 和 32.6%。同时,综合图 5 的可视化结果,表明去掉 RT 模块后会使得补全的 3D 点云对象丢失一些局部细节信息,比如椅子的扶手、飞机的机翼等。

表 2 消融实验的  $\text{Pred} \rightarrow \text{GT}/\text{Pred} \leftarrow \text{GT}$  对比,结果都被放大 10000Table 2  $\text{Pred} \rightarrow \text{GT}/\text{Pred} \leftarrow \text{GT}$  comparison of ablation study, results are magnified by 10000

Method	Airplane	Bag	Cap	Car	Chair	Guitar	Lamp
Proposed method	<b>2.164/</b>	<b>7.127/</b>	<b>6.950/</b>	<b>6.103/</b>	<b>4.107/</b>	<b>1.693/</b>	<b>14.506/</b>
	<b>2.461</b>	<b>3.340</b>	<b>3.065</b>	<b>2.592</b>	<b>2.050</b>	<b>2.168</b>	<b>10.085</b>
Proposed method without RT	3.208/	9.793/	11.605/	8.464/	6.407/	1.853/	18.528/
	3.356	6.232	4.262	3.730	3.413	3.214	12.221
Method	Laptop	Motorbike	Mug	Pistol	Skateboard	Table	Average
Proposed method	<b>2.110/</b>	<b>5.188/</b>	<b>5.481/</b>	<b>3.350/</b>	<b>1.812/</b>	<b>4.865/</b>	<b>5.035/</b>
	<b>1.195</b>	<b>3.701</b>	<b>3.316</b>	<b>4.051</b>	<b>2.101</b>	<b>2.513</b>	<b>3.294</b>
Proposed method without RT	2.975/	5.748/	7.922/	5.621/	7.391/	6.560/	7.390/
	2.013	5.525	4.639	6.031	4.809	4.103	4.888



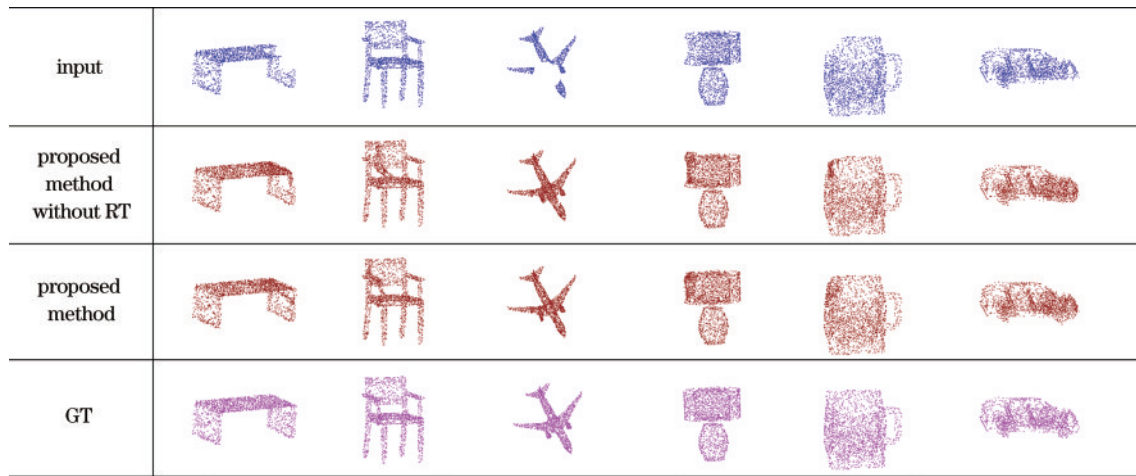


图 5 消融实验在 Shapenet-Part 数据集上的补全可视化效果对比

Fig. 5 Comparison of completion visualization effects of ablation study on Shapenet-Part dataset

为了证明所提多编码器的有效性,通过去除提取局部特征的编码器分支来分析对点云补全性能的影响。像许多先前的工作一样,利用 PCN 提出的提取全局特征的编码器来补全点云,保证网络模型其他部分不变,用 w/o Lo 表示。结果如表 3 所示,当网络只注

重提取全局特征而忽略局部特征时,不论是  $\text{Pred} \rightarrow \text{GT}$  还是  $\text{Pred} \leftarrow \text{GT}$  评价指标,实验结果都出现明显下滑,均值分别下降了 39.6% 和 44.2%。实验结果表明局部特征在点云补全任务中不容忽视,证明了所提多编码器的有效性和重要性。

表 3 消融实验的  $\text{Pred} \rightarrow \text{GT}/\text{Pred} \leftarrow \text{GT}$  对比,结果都被放大 10000

Table 3  $\text{Pred} \rightarrow \text{GT}/\text{Pred} \leftarrow \text{GT}$  comparison of ablation study, results are magnified by 10000

Method	Airplane	Bag	Cap	Car	Chair	Guitar	Lamp
Proposed method	<b>2.164/</b>	<b>7.127/</b>	<b>6.950/</b>	<b>6.103/</b>	<b>4.107/</b>	<b>1.693/</b>	<b>14.506/</b>
	<b>2.461</b>	<b>3.340</b>	3.065	<b>2.592</b>	<b>2.050</b>	<b>2.168</b>	<b>10.085</b>
Proposed method w/o Lo	5.055/	18.436/	9.441/	7.561/	12.959/	4.534/	15.442/
	7.812	4.450	<b>2.961</b>	3.938	4.838	3.363	14.081
Method	Laptop	Motorbike	Mug	Pistol	Skateboard	Table	Average
Proposed method	<b>2.110/</b>	5.188/	<b>5.481/</b>	<b>3.350/</b>	<b>1.812/</b>	<b>4.865/</b>	<b>5.035/</b>
	<b>1.195</b>	<b>3.701</b>	<b>3.316</b>	<b>4.051</b>	<b>2.101</b>	<b>2.513</b>	<b>3.294</b>
Proposed method w/o Lo	3.198/	<b>3.229/</b>	10.524/	4.440/	5.669/	7.862/	8.335/
	4.156	6.181	4.145	10.144	6.232	4.432	5.903

## 4 结 论

在传统的 Encoder-Decoder 网络基础上,提出一个基于多编码器的端到端的网络模型用于 3D 点云补全。提出的多编码器网络结构既可以学习点云的全局特征,又可以提取点云的局部特征信息,防止特征丢失,更好地保留了模型的细节信息。此外,针对生成的点云存在密度分布不均匀的情况,提出了一个基于 Residual-Transformer 的细化网络,该网络不仅可防止局部区域结构信息的丢失,同时能预测出点云的位移量,使得生成的点云模型中的点均匀分布。实验结果表明:与 MSN、PF-Net 和 RP-MLP 等最新算法相比较,在相同实验条件下,所提方法的量化评价指标均取得了较好的结果,补全的可视化点云对象也更加平滑,保留了更多的局部信息。

## 参 考 文 献

- [1] Huang Z T, Yu Y K, Xu J W, et al. PF-net: point fractal network for 3D point cloud completion[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 7659-7667.
- [2] Tchapmi L P, Kosaraju V, Rezatofighi H, et al. TopNet: structural point cloud decoder[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 383-392.
- [3] 李晓波, 袁希平, 甘淑, 等. 一种基于对偶四元素描述的线面特征约束的点云配准方法[J]. 光学学报, 2022, 42(2): 0214003.  
Li R B, Yuan X P, Gan S, et al. Point cloud registration method based on dual quaternion description of line-planar feature constraints[J]. Acta Optica Sinica, 2022,

- 42(2): 0214003.
- [4] 汪文琪, 李宗春, 付永健, 等. 基于改进多规则区域生长的点云多要素分割[J]. 光学学报, 2021, 41(5): 0528001.  
Wang W Q, Li Z C, Fu Y J, et al. Multi-factor segmentation of point cloud based on improved multi-rule region growing[J]. Acta Optica Sinica, 2021, 41(5): 0528001.
- [5] 李维刚, 樊响, 梅洋, 等. 基于车载激光点云的道路标线提取方法[J]. 中国激光, 2022, 49(7): 0710001.  
Li W G, Fan X, Mei Y, et al. Road marking extraction method based on vehicle laser point cloud[J]. Chinese Journal of Lasers, 2022, 49(7): 0710001.
- [6] 张磊, 徐孝彬, 曹晨飞, 等. 基于动态特征剔除的图像与点云融合的机器人位姿估计方法[J]. 中国激光, 2022, 49(6): 0610001.  
Zhang L, Xu X B, Cao C F, et al. Robot pose estimation method based on image and point cloud fusion with dynamic feature elimination[J]. Chinese Journal of Lasers, 2022, 49(6): 0610001.
- [7] Soltani A A, Huang H B, Wu J J, et al. Synthesizing 3D shapes via modeling multi-view depth maps and silhouettes with deep generative networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2511-2519.
- [8] Eigen D, Puhrsch C, Fergus R. Depth map prediction from a single image using a multi-scale deep network[C]//Advances in Neural Information Processing Systems 27, December 8-13, 2014, Montreal, Canada. New York: Curran Associates, 2014: 2366-2374.
- [9] Wang N, Zhang Y, Li Z, et al. Pixel2mesh: generating 3D mesh models from single RGB images[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 111215: 55-71.
- [10] Qi C R, Su H, Nießner M, et al. Volumetric and multi-view CNNs for object classification on 3D data[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 5648-5656.
- [11] Liu Z J, Tang H T, Lin Y J, et al. Point-voxel CNN for efficient 3D deep learning[EB/OL]. (2019-07-08)[2021-02-05]. <https://arxiv.org/abs/1907.03739>.
- [12] Wang X G, Ang M H, Lee G H. Cascaded refinement network for point cloud completion[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 787-796.
- [13] Yuan W T, Khot T, Held D, et al. PCN: point completion network[C]//2018 International Conference on 3D Vision (3DV), September 5-8, 2018, Verona, Italy. New York: IEEE Press, 2018: 728-737.
- [14] Liu M H, Sheng L, Yang S, et al. Morphing and sampling network for dense point cloud completion[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 11596-11603.
- [15] Lin Z H, Feng M W, Santos C N D, et al. A structured self-attentive sentence embedding[EB/OL]. (2017-03-09)[2021-05-06]. <https://arxiv.org/abs/1703.03130>.
- [16] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[EB/OL]. (2017-06-12)[2021-05-06]. <https://arxiv.org/abs/1706.03762>.
- [17] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[EB/OL]. (2020-05-26)[2021-04-05]. <https://arxiv.org/abs/2005.12872>.
- [18] Liu R J, Yuan Z J, Liu T, et al. End-to-end lane shape prediction with transformers[C]//2021 IEEE Winter Conference on Applications of Computer Vision, January 3-8, 2021, Waikoloa, HI, USA. New York: IEEE Press, 2021: 3693-3701.
- [19] Engel N, Belagiannis V, Dietmayer K. Point transformer [J]. IEEE Access, 2021, 9: 134826-134840.
- [20] Guo M H, Cai J X, Liu Z N, et al. PCT: point cloud transformer[J]. Computational Visual Media, 2021, 7(2): 187-199.
- [21] Charles R Q, Hao S, Mo K C, et al. PointNet: deep learning on point sets for 3D classification and segmentation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 77-85.
- [22] Qi C R, Yi L, Su H, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space [EB/OL]. (2017-06-07)[2021-05-08]. <https://arxiv.org/abs/1706.02413>.
- [23] Rubner Y, Tomasi C, Guibas L J. The earth mover's distance as a metric for image retrieval[J]. International Journal of Computer Vision, 2000, 40(2): 99-121.
- [24] Borgefors G. Hierarchical chamfer matching: a parametric edge matching algorithm[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1988, 10(6): 849-865.
- [25] Wu Z R, Song S R, Khosla A, et al. 3D ShapeNets: a deep representation for volumetric shapes[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 1912-1920.
- [26] Mendoza A, Apaza A, Sipiran I, et al. Refinement of predicted missing parts enhance point cloud completion [EB/OL]. (2020-10-08)[2021-07-08]. <https://arxiv.org/abs/2010.04278>.