

DOI:10.19651/j.cnki.emt.2211269

基于骨架重建的点云补全网络^{*}杨小平^{1,2} 赵晓^{1,2}

(1. 桂林理工大学信息科学与工程学院 桂林 541004; 2. 桂林理工大学广西嵌入式技术与智能系统重点实验室 桂林 541004)

摘要: 随着计算机三维视觉的快速发展,包含空间几何信息的点云数据广泛应用于机器人、自动驾驶等场景中,然而由于遮挡、角度受限等原因经常会造成几何语义信息的缺失。为了解决这一问题,提出了SRC-Net,首先利用融合动态图卷积网络编码器和折叠网络解码器的骨架重建网络从残缺点云中重建出几何骨架,接着使用自编码器结构建立几何骨架到均匀完整点云的映射。最后在MVP数据集上的补全结果表明,SRC-Net较现有补全网络可以生成更均匀且光滑分布的高质量完整点云,并能达到更为细节的补全效果,为点云深度学习补全提供了一种新的思路和方法,具有一定指导意义。

关键词: 三维图像处理;点云;补全;自编码器;图神经网络

中图分类号: TN958.98;TN249 **文献标识码:** A **国家标准学科分类代码:** 520.2060

Point cloud completion network based on skeleton reconstruction

Yang Xiaoping^{1,2} Zhao Xiao^{1,2}(1. College of Information Science and Engineering, Guilin University of Technology, Guilin 541004, China;
2. Guangxi Key Laboratory of Embedded Technology and Intelligent System, Guilin University of Technology, Guilin 541004, China)

Abstract: With the rapid development of computer 3D vision, point cloud data containing spatial geometric information is widely used in robots, autonomous driving and other scenes. However, due to occlusion, angle limitation and other reasons, geometric information is often missing. In order to solve this problem, SRC-Net is proposed. Firstly, a geometric skeleton is reconstructed from the incomplete point cloud using a designed skeleton reconstruction network that fuses a dynamic graph convolutional network encoder and a folding network decoder, and then the auto-encoder structure is used to establish the mapping from the geometric skeleton to the uniform and complete point cloud. Finally, the completion results on the MVP dataset show that SRC-Net can generate high-quality complete point clouds that are more evenly and smoothly distributed than existing completion networks, and can achieve more detailed completion effects. It provides a new idea and method for point cloud deep learning completion, and has certain guiding significance.

Keywords: 3D image processing; point cloud; completion; auto-encoder; graph neural network

0 引言

随着3D传感器、计算机存储和图形计算技术的发展,三维数据推动了虚拟现实、自动驾驶、工业加工^[1]等方面的蓬勃发展^[2]。然而由于有限的传感器分辨率和外物遮挡、观察角度、镜面反射、回波吸收等因素^[3],所采集到的点云通常是不完整的,通过建立点云补全的上游任务,可以大大提高三维分割^[4]、分类^[5]等下游任务的精度。

传统点云补全方法分为几何补全和检索匹配两种,其中几何补全的方法通过点^[6]或网格^[7]的空间拓扑关系进行

索引,识别空洞后进行表面插值拟合^[8],这种方法并不适合直接运用于存在较大缺失的点云数据;检索匹配的方法^[9-11]通过将残缺的形状与数据库中的相似形状进行匹配来得到完整形状,其依赖于复杂的特征标注和大量的数据库。近年来,深度学习在众多领域有突破性进展,在点云方向的发展总体上是从二维到三维的一个过程。早期是将点云栅格化为二维图像进行模型预训练,文献[12]率先提出了多视图卷积神经网络(convolutional neural networks, CNN),由于投影角度问题,该方法某种程度上采集的物体空间信息不完备。在2017年文献[13]创造性地提出了

收稿日期:2022-09-05

^{*} 基金项目:863计划地球观测与导航技术领域项目(2013AA12210504)资助

PointNet 网络解决了点云无序性问题,直接对无结构的点数据应用深度学习网络,解决了点云的稀疏性和转换不变性问题。得益于 PointNet,衍生出了基于生成对抗网络(generative adversarial networks, GAN)^[14-15]、自编码器(auto-encoder, AE)^[16]等一些列基于点数据的点云补全网络,其中自编码器相对于 GAN,可以快速收敛,且具有可逆性。点云补全中的自编码器包括一个从点获取全局特征的编码器和从特征还原至完整点的解码器;对于 PointNet 编码器中每个点的特征都是独立学习的,因此无法捕获点之间的局部细节结构信息,为此 PointNet++^[17] 提供了多尺度分组(multi-scale grouping, MSG)多分辨率分组(multiresolution grouping, MRG)两种分组方法对点云进行局部分组,但不管是 MSG 还是 MRG 并没有摆脱逐点学习的实质,动态图卷积网络(dynamic graph CNN, DGCNN)^[18] 采用了图神经网络的方法,设计了边缘卷积(EdgeConv)特征提取模块,基于图的自编码器^[19-20]等一些列点云补全网络再一次得到了发展;对于解码器,首个点云补全网络 PCN 采用折叠网络(FoldingNet)^[21] 的思想设计了折叠编码器,将全局特征加上二维网格约束后,投射回三维空间,大大提高了完整点云的均匀性和平滑性。

然而受限于 GPU 的算力和内存,视深度学习网络模型大小不同,这些补全网络一般在 1 024~2 048 个输入点数据上进行学习,所补全出来的完整点云依然存在细节不够、

丢失输入点云信息等问题。在 GPU 算力的发展背景下和有限点数量学习、自编码器优化受限的前提下,为了达到较好的细节补全效果,本文从自编码器结构的本质入手,提出了基于骨架重建的点云补全网络(point cloud completion network based on skeleton reconstruction, SRC-Net),该模型首先基于一个骨架重建网络从输入残缺点云中恢复出的几何骨架,然后采用基于 DGCNN 编码器和金字塔结构^[22] 结合 FoldingNet++^[23] 的解码器结构从几何骨架推理出完整点云,以期用骨架到完整的分步补全方法提高补全精细度、以及折叠的特征解码方法提高补全的均匀完整性。通过设计和一些现有优秀点云补全网络的对比实验,证明了 SRC-Net 相对于现有部分优秀点云补全网络的优越性。这种补全模型将在自动驾驶,机器人路径规划、三维面部识别等具体应用方面具有重大意义。

1 SRC-Net 网络模型设计

由深度传感器采集出的三维物体点云一般由三维坐标表示,如图 1 所示,输入是一个 $n \times 3$ 的残缺点云,通过骨架重建并融合输入之后,又经历了离群点去除和随机采样,作为一组由编码器和解码器组成的 AE 结构的输出。其中编码器使用改进提出的 CAEs(channel-attentive EdgeConv s),解码器融合 3 层金字塔结构和 FoldingNet++, 网络中各结构的原理和方法将从以下子节中详细阐述。

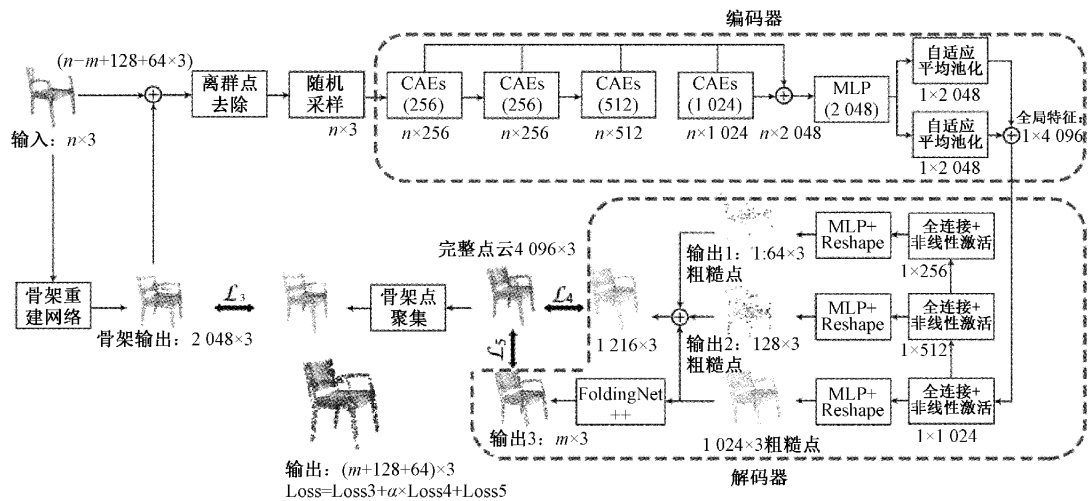


图 1 SRC-Net 模型结构

1.1 骨架重建网络

骨架重建网络用于从残缺点云粗略的重建出完整点云骨架,如图 2 所示,结构上可以分为残缺编码模块和骨架解码模块两个模块。为了提取残缺点云的特征,残缺编码模块直接使用了 DGCNN 中的 EdgeConv;骨架解码模块则直接使用 FoldingNet 的折叠操作,从全局特征中折叠出骨架输出。

EdgeConv 假设输入云中有 $x_1 \dots x_n$ 共 n 个点,每个点包含坐标信息 $x_i = (x_i, y_i, z_i)$, 用一个有向图 $G = (V, e)$

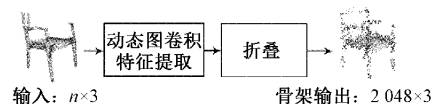


图 2 骨架重建网络结构

表示局部点云结构, $V = \{1, \dots, n\}$ 表示定点, $\epsilon \subset V \times V$ 表示边。将每个点看做定点,使用 K 近邻(K-nearest neighbor, KNN)构建图,根据每个 K 邻域中心 x_i 坐标来捕获全局形状信息, $x_j - x_i$ 捕获局部邻域信息,因此一个

K 邻域内的所有特征:

$$h_{\Theta}(x_i, x_j) = \overline{h_{\Theta}(x_i, x_j - x_i)} \quad (1)$$

其中, h_{Θ} 是具有一组可学习参数的特征提取函数, 比如多层感知机 (multilayer perceptron, MLP)。动态图卷积特征提取由 4 个串联的 EdgeConv 进行, 进而实现了从三维空间到其他 3 个 EdgeConv 特征空间的图更新, 最终将三维空间投射到全局特征向量 v 中。

折叠操作, 实际上是一个 2D 到 3D 的映射。具体地说, 对于 m 点中的每一个点 m_i , 通过变形 $u \times u$ 栅格生成 $t = u^2$ 个点的方块, 然后组织它们的坐标为 $t \times 2$ 的矩阵 G , 最后使用中心点坐标 m_i 和全局特征向量 v 的坐标拼接 G 的每一行, 并且通过 MLP 传递结果矩阵生成 $t \times 3$ 的矩阵 Q , 即以 m_i 为中心的局部块。MLP 可以被解释为一个非线性变换, 用于将 2D 栅格变换为 3D 空间中的平滑的 2D 流形。

1.2 骨架点聚集

为了给骨架输出提供监督信号, 本文设计了一种三维点云的骨架点聚集算法, 用于在完整点云上生成骨架结构突出的点云。骨架点聚集算法如图 3 所示。

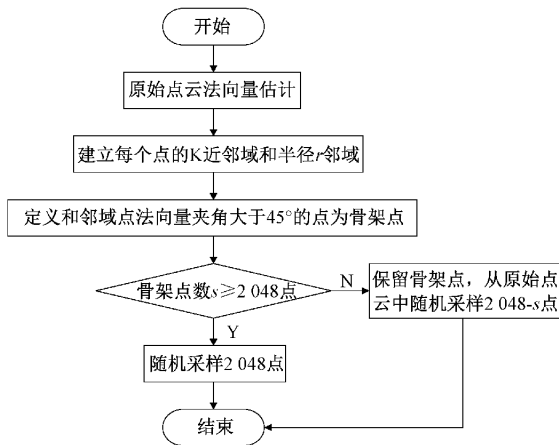


图 3 骨架点聚集算法流程

如图 4(a) 和 (b) 所示, 在完整的原始点云表面上一点处的法向量的估计即求解此表面相切的一个平面的法向量, 因此对于点云中的每个点需要用 KNN 或由半径确定一个邻域, 并求解邻域协方差矩阵的特征值和特征向量。

在获取法向量后, 本文通过建立每个点邻域并判断该点和邻域内其他点的法向量夹角度数, 如果这个度数大于 45° , 则认为该点能够表达点云三维空间上的骨架信息。需要注意的是, 如图 4(c) 在建立每个点邻域完成子野分割的过程中, 如果仅仅采用单一的 KNN, 当 k 值选择过小 (例如 k 值小于原始点云中椅子腿部的横向点云数) 时, 具有小范围内精细结构的点云无法取出关键的骨架点, 因此本文在建立每个点邻域不仅考虑了 KNN 的方法, 还加入了半径 r 约束, 加入半径 $r = 0.001$ 约束后, 在每个子野所提取的骨架点如图 4(d) 所示, 防止了关键骨架点的采样缺失。

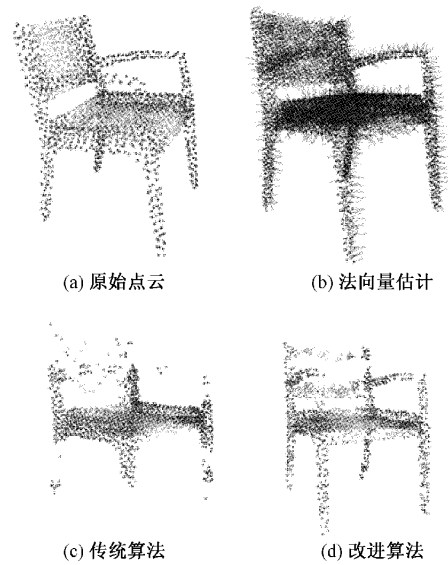


图 4 骨架点聚集算法可视化

然而由于一些具有高折叠的精细结构往往能够采样出较多的骨架点, 具有光滑结构的点云却很难有效的取出骨架, 因此当骨架点大于等于 2 048 点时在取出的骨架点中随机采样出 2 048 点, 否则在原始点云中随机采样出 2 048 点来补骨架点空缺, 以此保证不管什么样结构的点云均能取出 2 048 点数。从一个原始点云上整体来看, 本文将这一过程形象的理解为“骨架点聚集”。

1.3 离群点去除和随机采样

图 5 (a) 中的残缺点云, 进行骨架重建的结果如图 5(b) 所示, 本文将两者进行融合为图 5(c) 的融合点云。由于骨架点是通过神经网络学习而来的, 其存在一些点偏离了原始结合结构, 这些点也被称为离群点, 本文采用了统计式的剔除方式来剔除这些离群点, 这种方法也称为邻域滤波, 如图 5(d) 所示, 红色点是通过本算法剔除的离群点。同样由于 GPU 算力限制, 编码之前, 需要将输入点云个数保持在 2 048 点左右, 然而邻域滤波的结果好坏需要建立在较多的点数统计之上, 如图 5(e) 所示, 因此本文在离群点去除过程之后, 设置了随机采样将离群点去除后的点云进行下采样至 2 048 点。

其中邻域滤波采用如图 6 所示的算法流程。首先建立一组点云中每个点的 K 邻域; 然后在一个 K 邻域内通过式 (2) 计算欧氏距离的总体标准差, 其中 ρ_i 是每个点和中心点的欧氏距离由式 (3) 计算得到, $\bar{\rho}$ 是欧氏距离均值由式 (4) 计算得到, 其中欧氏距离表示三维空间中两点的绝对距离; 最后, 将偏离总体标准差 3 倍的点进行剔除。

$$\tau = \sqrt{\frac{\sum_{i=1}^k (\rho_i - \bar{\rho})^2}{k}} \quad (2)$$

$$\rho_i = s_i - s_2 = \sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2} \quad (3)$$

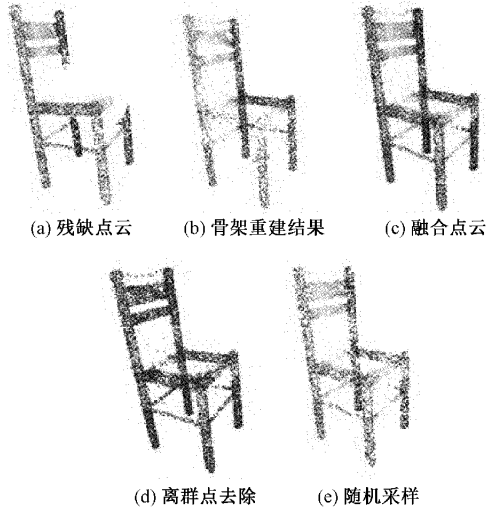


图5 离群点去除和随机采样可视化

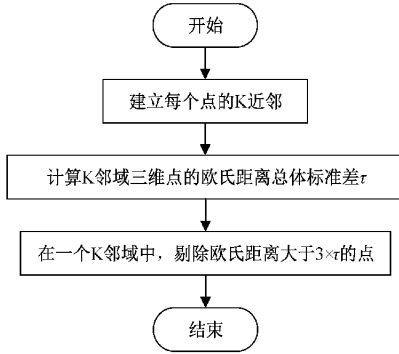


图6 邻域滤波算法流程

$$\bar{\rho} = \frac{\sum_{i=1}^k s_i - s_2}{k} = \frac{\sum_{i=1}^k \sqrt{(x_i - x_2)^2 + (y_i - y_2)^2 + (z_i - z_2)^2}}{k} \quad (4)$$

1.4 编码器

1) CAEs

在2.3节中,随机采样后的点云,仍然存在部分稀疏、部分均匀的特性,稀疏点的局部特征往往对还原全局特征具有消极的作用,为了抑制不好的局部特征,启发于SpareNet^[20]提出的CAE(channel-attentive EdgeConv)编码方式,本文改进提出了CAEs,在点特征提取时,通过融合局部和全局特征,用全局特征调制局部特征,达到促进完整的局部特征,抑制不好的局部特征的目的。

CAE是一种基于信道的EdgeConv,不仅考虑了K邻域中的局部信息,而且通过聚合全局特征并相应地加权每个点的特征信道关注度,明智地利用全局上下文。边缘特征与全局特征的融合丰富了学习表征,因此网络更有力地表征了输入的精细结构。

如图7所示,让 P_m 作为CAE块的输入,假设它有 N

个点,具有特征维数 C_m 。对于每个点 $p_i \in P_m$,首先在建立 p_i 的 k 近邻子野,将一个子野内除中心点 p_i 的其他点记为 $\{q_j^i, 1 \leq j \leq k\}$, p_i 上有 k 个方向的边特征,每个边表示为 $(p_i, q_j^i - p_i)$ 。为了更好的捕捉几何轮廓,本文选用余弦相似度来表示每个边: $(p_i, \text{similarity}(q_j^i - p_i))$,并将这种方法命名为CAEs,然后利用MLP(记为 F_1)从每条边计算出一个新的特征:

$$e_i^j = F_1(p_i, \text{similarity}(q_j^i - p_i)) \quad (5)$$

$$\text{similarity}(x_i, x_j) = \cos\theta = \frac{x_i \cdot x_j}{|x_i| |x_j|} =$$

$$\frac{\sum_{n=1}^3 x_i \times x_j}{\sqrt{\sum_{n=1}^3 (x_i)^2} \times \sqrt{\sum_{n=1}^3 (x_j)^2}} \quad (6)$$

则一组点云中的所有边缘特征为: $\{e_i^j, 1 \leq i \leq N, 1 \leq j \leq k\}$ 。

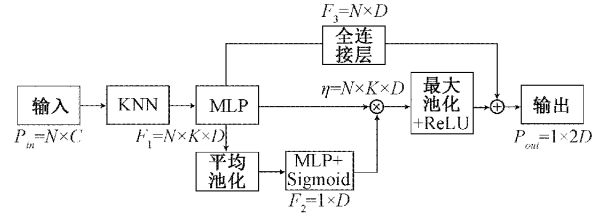


图7 CAEs结构

为了利用KNN图的全局上下文信息,CAEs将边缘特征所有全局平均值输入第2个MLP(记为 F_2)来计算门控向量 η :

$$\eta = \sigma \cdot F_2 \left[\frac{1}{kN} \times \sum_{i,j}^{N,k} e_i^j \right] \quad (7)$$

其中, σ 表示sigmoid非线性激活。CAEs将每一个边特征 e_i^j 乘以 η 重新校准,以便同时利用局部和全局上下文进行特征激活。最后,对于每个点,通过最大池化和ReLU激活将其 k 个边特征还原为新的点特征。

CAEs的输出还增加了残差,残差由跳过输出与输入连接的线性全连接层 F_3 获得。

2) 编码器结构

SRC-Net编码器由4个串联的CAEs模块进行动态图更新,并将不同感受野的CAEs模块所提取出的特征进行了整合,然后通过MLP压缩特征维度。为了不丢失特征信息和抑制过拟合,最后在每个维度上进行了自适应最大池化和平均池化,最终获得 1×4096 维度的全局特征 g 。

1.5 解码器

在图1中, SRC-Net的解码器采用3层全连接+非线性激活的金字塔结构来保留各种维度语义信息,其中高维特征包含最丰富的语义信息,也是能够还原出表达三维结构最关键点的特征,因此由 1×256 维高维特征生成最关键的64个稀疏点, 1×512 维生成128个中间点, 1×1024 维生成1024个点。并对三层粗糙点输出整合为 1216×3

后做一次监督,对于 1 024 维全局特征则采用 FoldingNet++ 折叠出 $m \times 3$ 的精细点云。最终输出整合了经过监督的 128、64、 m 点三维数据,并由完整点云进行监督。FoldingNet++ 模型结构如图 8 所示,FoldingNet++ 将折

叠操作的输出作为了一个中间输出,然后将生成的 $m \times 2$ 网格增加了一个图拓扑模块,并将中间输出和经过推理的 $m \times m$ 拓扑结构整合起来,最后经过图滤波模块折叠出最终输出 3,接下来将对这两个模块进行详细说明。

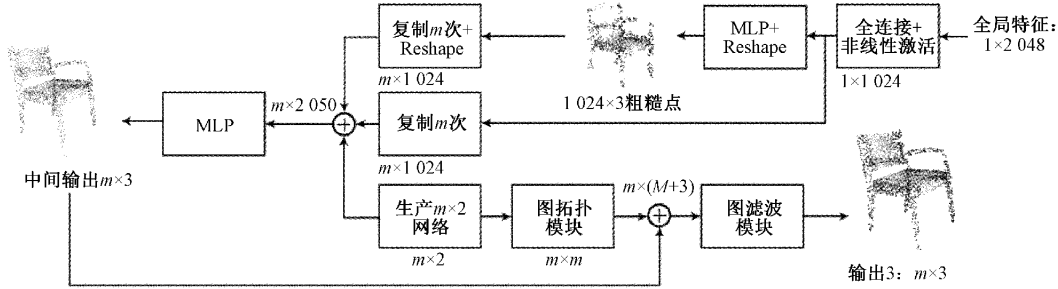


图 8 FoldingNet++ 解码器模型结构

1) 图拓扑推理模块

FoldingNet 中,2D 网格是默认是均匀的与每个点连接,然而这种方式不能够捕获非规则的模式,3 位空间的几何特征常常和 2D 平面有不同的拓扑结构。为了解决这个问题,图拓扑推理模块学习图来获取不规则的点间关系。

图拓扑推理模块在 FoldingNet 初始化的 2D 网格基础上加入了边信息构成一个图结构,其中边根据训练过程进行更新。图的初始邻接矩阵定义为:

$$A_{ij}^0 = \begin{cases} \frac{1}{z_i} \exp(-\frac{\|z_i - z_j\|_2^2}{2\sigma^2}), & z_i \in N_i \\ 0, & \text{其他} \end{cases} \quad (8)$$

其中, z_i 是 2D 网络中的第 i 个结点, σ 反应了衰减率, N_i 表示 z_i 的 k 个邻居点,归一化项为 $\frac{1}{z_i} \exp(-\frac{\|z_i - z_j\|_2^2}{2\sigma^2})$, 确保了 $\sum_j A_{ij}^0 = 1$ 。

由于通过编码器产生的全局特征保留了原始点云的一些信息,图拓扑推理模块将全局特征与 A^0 的每一行拼接起来,然后使用 MLP 来实现图拓扑结构推理模块,最终推理出 $m \times m$ 的图拓扑结构。

图邻接矩阵提供了 2D 网格各种点之间的关联性和全局特征提供的机制来适应图邻接矩阵的边权值,也能够反应点对关系。同时,在训练过程中,图拓扑推理模块将全局特征保留的空间信息来指引邻接矩阵的进化。

2) 图滤波模块

图滤波模块将中间输出和图拓扑推理模块合并作为输入,优化粗略重构的效果。FoldingNet++ 中认为点云中的每个点沿表面的曲率引起了不连续性,但是,3D 坐标却总是沿表面平滑的。也就是说,如果一个图完美的反映出表面,3D 坐标在图域里面是平滑的。因此图滤波模块根据图邻接矩阵设计低通滤波器来获取改进的重构。滤波器使得每个点可以从其周边邻居获取一些信息。同时图滤波让网络学习图拓扑结构,用来保留 3D 点在图中的平滑性。

FoldingNet++ 提供了基于图邻接矩阵的滤波器和基于拉普拉斯矩阵的滤波器,由于邻接矩阵到拉普拉斯矩阵需要计算出度矩阵来进行转换,因此本文考虑使用基于图邻接矩阵的滤波器:

$$h(A) = \sum_{l=0}^{L-1} h_l A^l \in \mathbb{R}^{m \times m} \quad (9)$$

其中, h_l 为滤波参数, L 是滤波器顺序。大的 L 表明了更大的感受野;然而大 L 也会减速训练。简单起见,本文将 L 设置为 1,且 $h_0 = h_1 = 0.5$ 。最终重构为:

$$\hat{X} = \frac{1}{2}(I + A)X' \in \mathbb{R}^{m \times 3} \quad (10)$$

其中, X' 是中间输出,相应的 \hat{X} 的第 i 行为:

$$\hat{X}_i = (X'_i + \sum_{j \in N_i} A_{ij} X'_j) / 2 \quad (11)$$

将式(11)视为低通 Haar 滤波器^[24]。在传统处理过程中,Haar 滤波器用来平滑时间或图像中的噪声,图滤波模块使用图 Haar 滤波器来平滑粗略的 3D 点云局部来完整点云。

1.6 损失函数

在神经网络中损失函数(loss function)是不可缺少的组成部分,其也被称作代价函数,主要是用于评估模型的真实值与预测值之间的误差,并反向传播优化网络模型参数该结果。虽然损失函数能看到模型性能的优劣,但选取合适与否的损失函数也会影响模型本身性能。对于评价点云的回归损失,Liu 等^[25]提出了两个置换不变的度量来比较 s_1, s_2 两组无序点云,它们是倒角距离(chamfer distance, CD)和地球移动距离(earth move distance, EMD)。

$$d_{CD}(s_1, s_2) = \frac{1}{s_1} \sum_{x \in s_1} \min_{y \in s_2} \|x - y\|_2 + \frac{1}{s_2} \sum_{y \in s_2} \min_{x \in s_1} \|y - x\|_2 \quad (12)$$

$$d_{EMD}(s_1, s_2) = \min_{\phi: s_1 \rightarrow s_2} \sum_{x \in s_1} \|x - \phi(x)\|_2, s_1 \rightarrow s_2 \text{ 为一个双射} \quad (13)$$

EMD 的计算往往十分耗时,且需要 s_1, s_2 点数相同才能一一映射。由于 2.5 中解码器输出为 $m+64+128$ 点,故很难保证模型输出和原始点云骨架点聚集后的点云双射,因此本文选用倒角距离作为 loss 来进行梯度优化,倒角距离是计算生成点云和理想点云之间平均的最短点距离。式(12)的前一部分保证生成点云与理想完整点云点云之间的距离最小,后一部分保证理想完整点云点云在生成点云中的覆盖率。

本文中,假设训练集的完整点云 U 通过骨架聚集形成一个子集 R ,金字塔结构的 1 216 点粗糙输出为 V , FoldingNet++ 的 m 点输出 3 为 W 。对于 V 和 W 需要利用 U 来进行,监督骨架输出 X 则利用 R 来评估,因此 SRC-Net 的整体损失函数为:

$$Loss = d_{cd}(X, R) + \alpha \cdot d_{cd}(V, U) + d_{cd}(W, U) = L_3 + \alpha \cdot L_4 + L_5 \quad (14)$$

本文对于金字塔结构整合后的 1 216 点粗糙输出 V 在做监督时增加了权重 α , 来降低匹配所有点的限制,提高了对点云分布差异的容忍度,其中 α 为 0~1 之间的权重值,并设置在训练中自动更新。

2 实验

2.1 数据集

在本论文中主要使用到的公共数据集为 MVP 数据集^[26]。MVP 是目前唯一一个专门用于点云补全的数据集, MVP 数据集由南洋理工大学、商汤科技和上海人工智能实验室从 16 个类别的 3D CAD 模型中采集。

如图 9(a) 所示,遗憾的是 MVP 数据集只提供了 2 048 点的数据,为了实现 2 048 点的骨架点聚集和数据增强,本文采用对训练集进行了正态抖动预处理。假设一组点云用 $2\ 048 \times 3$ 维数组 A 表示,随机生成服从正态分布的 $2\ 048 \times 3$ 维随机数 A_n 。

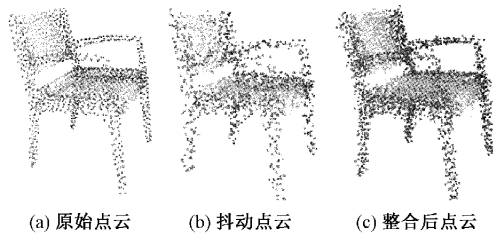


图 9 数据增强和扩充可视化

为了实现每个点云的正态抖动本文定义了式(15):

$$A_d = A + \gamma(0.01A_n) \quad (15)$$

式中: γ 函数由 numpy.clip 实现,如式(15)所示,首先对 A_n 进行 0.01 倍的缩放,然后通过 γ 函数在 A_n 中进行 ± 0.05 的筛选,超出部分以 ± 0.05 替代,最后的结果 A_d 为 A 中每个点上 ± 0.05 范围内的加和。生成的抖动点云 A_d 如图 9(b) 所示。

如图 9(c) 所示,为了对原始点云 A 进行数据扩充至

4 096 点,本文将抖动点云 A_d 和原始点云 A 进行整合为 A_l, A_l 点云即完整包含了较为平滑的原始表征信息,又实现了数据的增强处理。

2.2 实验环境和参数

本文的网络均在软件及硬件如表 1 所示的配置下部署及测试。

表 1 实验环境

系统/平台	版本
操作系统	ubuntu18.04
显卡	QuadroRTX5000 16 GB
CPU	AMDEPYC7302
CUDA	10.2
cuDNN	7.6.5
GNU 编译器	gcc-10.2.0
框架	PyTorch 1.5.0
语言	Python 3.8

在网络训练时设置训练集学习类别为 all,学习率为 0.000 1,衰减率为 0.7,使用 Adam 优化器对网络参数进行优化。每个点云点数为 4 096,使用批量训练的方式,设置批量大小为 32 以加快网络的训练速度。权重 α 设置为 0.01, 0.05, 0.1, 0.5, 1, 并在第 5、15、30、50 进行更新。

2.3 结果与分析

为了验证 SRC-Net 补全效果,本文利用了数据增强和扩充后的 MVP 数据集进行了训练,可视化采用方法 Open3D 工具箱,从红到蓝是根据深度不同进行的点云颜色渲染。如图 10 和 11 本文选择了 3 种现有的先进方法与本文提出的 SRC-Net 在补全结果的 CD 指标及视觉效果进行比较,以验证本章提出的网络模型的有效性。就像 SRC-Net 一样,这 4 种方法也是以自编码器端到端的方式训练网络模型,为了保证对比实验的有效性,本文使用相同的数据集和相同的训练相关参数设置对 3 种方法的模型进行训练,这 3 种网络分别是: PCN, PF-Net^[15], VRC-Net^[26]。

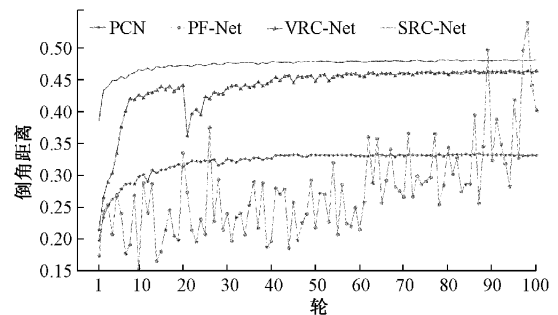


图 10 训练过程中的 CD 值

PCN 在 2019 年由卡内基梅隆大学提出,是第 1 个用于点云补全的深度学习网络,网络结构主要由两个部分

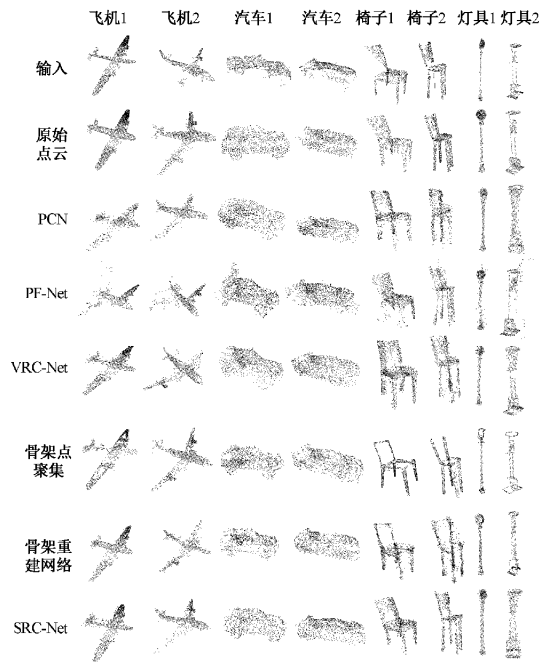


图 11 测试集补全效果可视化对比

组成,由 Pointnet++ 组成的编码器以及由 FoldingNet 组成的解码器。编码器采用输入点云并输出一个 k 维特征向量,解码器采用这个特征向量加上随机生成的二维网格通过多层感知机生成完整输出点。从图 10 中的模型学习过程中可以看出 PCN 具有稳定的收敛速度,但收敛不到位,因此在图 11 中可以看出 PCN 对汽车 1 等简单结构具有很好的补全效果,但是对于椅子 2、灯 2 等复杂模型的高折叠部分存在全覆盖的现象,灯具 1 的杆部细节也没有体现出来。总体上来看,PCN 利用折叠网络还原出了平滑的三维点云,同时带来的却是因为二维网格大小选取过小和折叠不到位而损失的细节结构,如果仅通过增大网格和多次折叠,必然会导致模型过拟合且增加 GPU 运行负担。

PF-Net 在 2020 年由上海交通大学提出,编码器部分通过最远点采样,对残缺点云进行多分辨率的特征学习后将全局特征整合为一维向量,特征学习的方法同样借鉴 Pointnet++。解码器采用金字塔结构+MLP 的方法生成粗糙点云,然后模仿对抗生成网络(GAN)的思想,仅仅生成缺失部分的点云并放到一个判别器中,最终来判断补全的结果的优劣性。从训练过程中的 CD 值来看,虽然设置了相同的学习率,但是 CD 值仍然波动较大,学习能力较弱,模型收敛速度较慢,因此缺失的点云没有很好的收敛到原有位置,可能需要增加训练次数,但是 PF-Net 保留残缺点云,仅仅在缺失部分进行补全的方法值得借鉴。

VRC-Net 在 2021 年由南洋理工大学 S-Lab 实验室提出,VRC-Net 由概率模型网络(PMNet)和关系增强网络(RENet)两个子网络级联构成。首先,PMNet 基于残缺点

云生成一个粗略的点云形状框架,然后再由 RENet 结合前阶段生成的粗略框架和残缺点云观测,推测具有相关性的结构,以实现最后生成点云细节特征的增强。从图 10 可以看出 VRC-Net 具有较快的收敛速度和优化效果。VRC-Net 的补全可视化结果相比较之前的点云补全方法主要有两方面的优势:1)可以生成更均匀和光滑分布的高质量完整点云;2)可以有效地捕捉到关联性结构并对缺失部分实现合理的生成和补全,但在椅子 1 中的脚撑部分出现了较多的散乱点,汽车 2 的前轮和灯具 2 细节不够,VRC-Net 在复杂结构中表现不是很好,可能是收敛速度过快导致的鲁棒性差。

本文的 SRC-Net 则包括两个串联的自编码器,一个用来学习残缺点云到骨架的映射,另一个则学习从骨架到均匀完整点云的映射。两步学习的方法大大降低了补全难度,从图 10 的训练过程中来看, SRC-Net 一开始的 CD 值就较高,这是因为其输入是整合了残缺点云和骨架重建网络的结果,因此和完整点云之间差异较小,也正是这个原因, SRC-Net 相对于其他网络 CD 指标能够收敛至 0.48,比 VRC-Net 提升了 11.6%,比 PCN 提升了 50%。从均匀、平滑性角度对比 PCN 和 SRC-Net,证明了本文增加图拓扑推理模块和图滤波模块的折叠操作相对 PCN 简单折叠操作,优越性在于图拓扑推理模块模拟 3D 点云之间的关系,全局特征保留了 3D 点云的坐标和点对的关联性,因此可以不受特定监督的推断,图滤波模块将粗重构变为精细重构结果,改善了补全均匀效果,尤其表现于椅子 1 的腿部和椅子 2 的坐垫上表面均匀度;从精细化补全角度对比 VRC-Net 和 SRC-Net,证明了本文从粗糙到完整的分步补全方法,其在复杂结构上保证了细节结构不会缺失,比如汽车 2 的前轮胎和车头细节,椅子 2、灯具 2 细节展现,但这其中也有离群点去除带来的细节贡献。总体上来看, SRC-Net 生成点云的整体均匀性却不如 VRC-Net,例如 SRC-Net 在灯具 1 和灯具 2 的补全结果中点云还较为集中于三维骨架上,其他 6 个模型却较为均匀,也侧面说明了 SRC-Net 的泛化性有待提高,因此 VRC-Net 利用关系增强网络从骨架点云中生成点云细节特征的增强对均匀性和泛化性提升有很大帮助。

为了验证 SRC-Net 在实际补全中的表现,本文绘制并打印了空客 A320 的三维模型,如图 12(a)采用通过 iPhone 12 前置红外扫描仪进行三维扫描为图 12(b)的曲面文件,从扫描后的曲面文件中泊松圆盘采样出完整点云并随机裁剪出图 12(c)的残缺点云。用训练好的 SRC-Net 对残缺点云进行的补全结果如图 12(d)所示,从均匀、平滑性看,机头机翼位置甚至超过了泊松圆盘采样结果,证明了本文 SRC-Net 模型的合理性;从精细化角度看,尾翼位置得到了高折叠的精细补全,证明了本文所提出的残缺到骨架、骨架到完整的分步补全方法可行性及优越性。

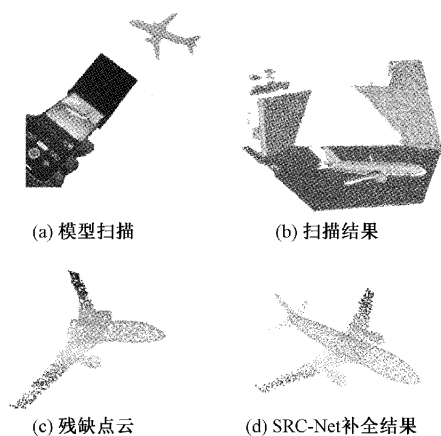


图12 SRC-Net补全实例

3 结 论

本文以重建工作中的物体三维形状补全任务为研究背景,在分析国内外点云补全方法后,对基于深度学习的三维点云补全问题进行了研究,提出了“两步走”的补全方法,即首先对从残缺点云进行骨架重建,然后从骨架点云恢复出完整点云,最终提出了 SRC-Net,其利用骨架重建网络对残缺点云重建出几何骨架,然后将两者融合并采用邻域滤波算法进行离群点去除,为建立去除离群点后的骨架到完整点云的映射,改进提出了 CAEs 编码器来同时利用局部和全局上下文进行特征激活,解码器则 FoldingNet++ 进行精细结构重构。

为了验证本文模型,对 MVP 数据集上进行了数据增强和扩充,并设计了对比实验。最后从模型训练过程和重建可视化结果表明,本文的补全方法可以生成更均匀和光滑分布的高质量完整点云,并能达到更为细节的补全效果,证明了 CAEs 编码器和 FoldingNet++ 解码器的优越性、以及“两步走”补全法的可行性。但骨架点聚集仍然依靠在子野内进行法向量比较的传统算法,可以进一步效仿 PointNet 的旋转网络(T-Net)设计方法,设计出一个轮廓提取网络,用学习的方法提取几何骨架,同时加入 Transformers 自注意力机制,为骨架和完整点云分配权重,实验验证点云补全的骨架注意力。

参考文献

- [1] 张潇丹,单彦虎,储成群.基于点云处理的小型工业零件高度精密测量[J].电子测量技术,2022,45(11):134-139.
- [2] 王春香,梁亮,王耀,等.三维点云模型孔洞边界识别的研究综述[J].现代制造工程,2019(7):157-162.
- [3] NASEER M, KHAN S H, PORIKLI F. Indoor scene understanding in 2.5/3D for autonomous agents: A survey[J]. IEEE Access, 2019, 7:1859-1887.
- [4] 任兵,陈卫国,饶银辉,等.基于移动式浮标的无人艇航

- 行监测系统设计[J].中国测试,2022,48(4):123-128.
- [5] 徐平,沙从术.产品表面缺陷检测系统设计与开发[J].中国测试,2020,46(6):34-38.
- [6] CHARLES R Q, SU H, KAICHUN M, et al. PointNet: Deep learning on point sets for 3D classification and segmentation [C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR),2017.
- [7] 赵江洪,孙铭悦,王殷瑞,等.三维点云孔洞修复方法综述[J].测绘科学,2021,46(1):114-123.
- [8] KAZHDAN M, HOPPE H. Screened poisson surface reconstruction [J]. Acm Transactions on Graphics, 2013,32(3):1-13.
- [9] KIM V G, LI W, MITRA N J, et al. Learning part-based templates from large collections of 3D shapes[J]. ACM Transactions on Graphics, 2013, DOI: 10.1145/2461912.2461933.
- [10] ROCK J, GUPTA T, THORSEN J, et al. Completing 3D object shape from one depth image[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA, 2015: 2484-2493.
- [11] YIN K, HUANG H, ZHANG H, et al. Morfit. interactive surface reconstruction from incomplete point clouds with curve-driven topology and geometry control [J]. ACM Transactions on Graphics, 2014, 33(6):202.
- [12] SU H, MAJI S, KALOGERAKIS E, et al. Multi-view convolutional neural networks for 3D shape recognition[J]. 2015 IEEE International Conference on Computer Vision (ICCV), 2015, DOI: 10.1109/ICCV.2015.114.
- [13] QI C R, SU H, MO K, et al. PointNet: Deep learning on point sets for 3D classification and segmentation[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017, DOI: 10.1109/CVPR.2017.16.
- [14] 叶荆荆,叶海良,曹飞龙.基于生成对抗自关注的点云补全网络[J].中国计量大学学报,2021,DOI:10.3969/j.issn.2096-2835.2021.03.012.
- [15] HUANG Z, YU Y, XU J, et al. PF-Net: Point fractal network for 3D point cloud completion [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2020, DOI: 10.1109/CVPR42600.2020.00768.
- [16] YUAN W, KHOT T, HELD D, et al. PCN: Point completion network[J]. IEEE,2018:728-737.
- [17] QI C R, YI L, SU H, et al. PointNet ++: Deep

- hierarchical feature learning on point sets in a metric space[J]. *Advances in Neural Information Processing Systems*, 2017, 30.
- [18] WANG Y, SUN Y, LIU Z, et al. Dynamic graph CNN for learning on point clouds[J]. *ACM Transactions on Graphics*, 2018, DOI:10.1145/3326362.
- [19] LIANG P. ECG: Edge-aware point cloud completion with graph convolution [J]. *IEEE Robotics and Automation Letters*, 2020, DOI:10.1109/LRA.2020.2994483.
- [20] XIE C, WANG C, ZHANG B, et al. Style-based point generator with adversarial rendering for point cloud completion[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 4619-4628.
- [21] YWANG Y, FENG C, SHEN Y, et al. FoldingNet: Point cloud auto-encoder via deep grid deformation [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 206-215.
- [22] LIN T Y, DOLLAER P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE Computer Society, 2017: 936-944.
- [23] CHEN S, DUAN C, YANG Y, et al. Deep unsupervised learning of 3D point clouds via graph topology inference and filtering[J]. *IEEE Transactions on Image Processing*, 2019, 29: 3183-3198.
- [24] VETTERLI M, KOVÁČEVÍČ J, GOYAL V K. *Foundations of signal processing [M]*. Cambridge: Cambridge University Press, 2014.
- [25] LIU J, YU F, FUNKHOUSER T. Interactive 3D modeling with a generative adversarial network [J]. *IEEE*, 2017, DOI:10.1109/3DV.2017.00024.
- [26] PAN L, CHEN X, CAI Z, et al. Variational relational point completion network[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 8524-8533.

作者简介

杨小平, 教授, 硕士生导师, 主要研究方向为智能感知技术、计算机视觉、图像处理。

E-mail: gutyxp@126.com

赵晓, 硕士研究生, 主要研究方向为计算机视觉、三维点云处理。

E-mail: 2120200988@glut.edu.cn