基于类别-实例分割的室内点云场景修复补全

缪永伟[1, 2)] 刘家宗[2)] 孙瑜亮[3)] 吴向阳[4)]

[1)](杭州师范大学信息科学与技术学院 杭州 311121)

[2)](浙江理工大学信息学院 杭州 310018)

[3)](浙江工业大学计算机科学与技术学院 杭州 310023)

[4)](杭州电子科技大学计算机学院 杭州 310018)

摘 要 三维室内场景修复补全是计算机图形学、数字几何处理、3D 计算机视觉中的重要问题. 针对室内场景修复补全中难以处理大规模点云数据的问题，本文提出了一种基于类别-实例分割的室内点云场景修复补全框架. 该框架包括点云场景分割模块和点云形状补全模块，前者由基于 PointNet 的类别分割网络和基于聚类的实例分割模块完成，后者由基于编码器-解码器结构的点云补全网络实现. 本文框架以缺失的室内场景点云数据为输入，首先根据“类别-实例”分割策略，采用 PointNet 对室内场景进行类别分割，并利用基于欧式距离的聚类方法进行实例分割得到室内各家具点云，然后借助点云补全网络将分割出的缺失家具点云逐一进行形状补全并融合进原始场景，最终实现室内点云场景的修复. 其中，为了实现缺失家具点云形状的补全，本文提出了一种基于编码器-解码器结构的点云补全网络，首先通过输入变换和特征变换对齐缺失的家具点云数据采样点位置与特征信息；然后借助权共享多层感知器和 PointSIFT 特征提取模块对各采样点提取形状特征和近邻点特征信息，并利用最大池化层与多层感知器编码提取出采样点的特征码字；最后将采样点特征码字加上网格坐标数据作为解码器的输入，解码器使用两个连续的三层感知器折叠操作将网格数据转变成完整的点云补全数据. 实验结果表明，本文提出的点云补全网络能够较好地补全室内场景中缺失的家具结构形状，同时基于该网络的场景修复补全框架能够有效修复大型室内点云场景.

关键词 室内场景；点云数据；类别-实例分割；编码器-解码器结构；修复补全

中图法分类号 TP391 **DOI 号** 10.11897/SP.J.1016.2021.02189

Point Cloud Completion of Indoor Scenes Based on Category-Instance Segmentation

MIAO Yong-Wei[1,2)] LIU Jia-Zong[2)] SUN Yu-Liang[3)] WU Xiang-Yang[4)]

[1)](*School of Information Science and Technology, Hangzhou Normal University, Hangzhou* 311121)

[2)](*School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou* 310018)

[3)](*College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou* 310023)

[4)](*Computer and Software School, Hangzhou Dianzi University, Hangzhou* 310018)

Abstract Point cloud completion of indoor scenes is an important issue in the literature of Computer Graphics, Digital Geometry Processing and 3D Computer Vision. Due to its difficulty for effectively processing large-scale point cloud data of the input 3d indoor scenes, a novel framework is presented in this paper which can repair the missing data of indoor scenes. Our framework consists of two modules, that is, one is the scene category-instance segmentation module and the other is the point cloud data completion module. The former is composed of PointNet based object category classification network and instance

收稿日期：2019-10-25；在线发布日期：2020-08-20. 本课题得到国家自然科学基金(No. 61972458, 61972122)资助. **缪永伟**，博士，教授，博士生导师，中国计算机学会(CCF)杰出会员，主要研究领域为计算机图形学、数字几何处理、计算机视觉、机器学习. E-mail: ywmiao2009@hotmail.com. **刘家宗**，硕士，主要研究领域为计算机图形学、计算机视觉. **孙瑜亮**，博士研究生，主要研究领域为计算机图形学、计算机视觉. **吴向阳**，博士，副教授，主要研究领域为计算机图形学、计算机视觉、可视化.

segmentation using a clustering scheme, whilst the latter is a point cloud completion network based on the encoder-decoder structure. This framework takes the incomplete point cloud data of 3d indoor scenes as input. Firstly, the furniture objects in indoor scenes can be classified by using PointNet based object category classification strategy. Then, each point cloud instance that belongs to the same object category can also be segmented effectively by using Euclidean distance based clustering scheme. The point cloud completion network can thus be adopted to repair 3d point cloud of each missing furniture, and all of the completed instances can be merged into 3d scenes for finally restoring the completed indoor scenes. Here, the data completion network is based on the encoder-decoder structure, with the missing furniture point cloud data as input. Firstly, the sampling point position and feature information of the input point clouds are aligned by input transformation and feature transformation. A weight-shared multi-layer perceptron can thus extract the local shape features for each sampling point, whilst the PointSIFT module can extract the feature information of neighbor points. Then, the feature code-words of sampling points can be extracted by using the maximum pooling layer and multi-layer perceptron. Finally, the extracted feature code-words of sampling points can be combined with 2D grid data, and the decoder will transform these grid data into the completed point cloud data by using two consecutive three-layer perceptron folding operations. Our proposed method is validated on repairing and completing both the 3d individual objects and also 3d indoor scenes. Using 3d object dataset ModelNet 40, the completion of 3d shapes has validated the effectiveness and excellent performance of our point completion network, which demonstrate that the repaired point cloud is closer to its ground truth in terms of the CD (Chamfer Distance) error and EMD (Earth Mover's Distance) error. Meanwhile, our proposed completion network is robust to the different degrees of data incompleteness, and also generalizes well for repairing 3d models not existing in the training dataset. For the Stanford 3D indoor scenes dataset S3DIS, our framework can be used to complete the missing data of the whole indoor scene effectively, and the point cloud data distribution of the completed scenes is also uniform if comparing with the original input point clouds. Experimental results demonstrate that the proposed point completion network can restore the missing 3d furniture shapes robustly, and our scene completion framework can effectively repair large-scale indoor scenes.

Keywords indoor scenes; point cloud data; category-instance segmentation; encoder-decoder structure; shape completion

1 引 言

三维室内场景修复补全是计算机图形学、数字几何处理、3D 计算机视觉领域的重要问题，其应用领域包括室内虚拟装修、室内游戏场景设计、虚拟考古博物馆、室内机器人导航等[1]. 在数字几何处理和 3D 计算机视觉中，点云数据由于其数据获取方便、无需维护拓扑连接关系、能较好地表示复杂形状等优点得到了普遍应用[1-2].

室内场景的点云数据获取通常有 2 种来源[2-3]：一是使用激光设备扫描室内场景以获取点云数据，利用传统激光扫描设备往往能扫描获取稠密且高质量的点云数据，但存在着设备价格高昂、扫描时间较长等缺陷；二是通过相机（如立体相机、RGBD 深度相机等）获取室内场景信息，并利用配准重建[3]等方法获得点云数据. 然而，通过扫描或相机拍摄获取的数据通常会受到场景中物体遮挡、深度相机的传感器距离限制、扫描设备出现误差等原因，导致获取的场景点云数据往往会产生数据缺失. 一般来说，室内场景的数据缺失主要有两类：第一类是平面结构出现孔洞区域（如墙壁、天花板、地板出现孔洞），第二类是室内家具形状结构的缺失（如桌子缺失桌腿、椅子缺失靠背等）. 针对第一类数据缺失，采用传统的点云修复方法[4]通过检测空洞边界，并利用边界点的临近点构造曲面进行孔洞修补，将能达到较好的修补效果. 而对于第二类数据缺失，则难以利用传统的点云修复方法进行补全. 一种可行的补全方法是对缺失的场景点云进行重建并

借助专业软件手工补全室内场景中家具的缺失形状. 然而，由于待修复形状复杂性高、手工修复效率低且成本高等原因，亟需提出一种高效自动的室内家具形状修复方法以实现室内点云场景的修复补全.

近年来，随着大型三维 CAD 模型数据集[5]和大型扫描场景数据集[6]的出现，训练数据量的增加推动了深度神经网络技术在三维形状生成和修复补全方面取得了若干进展. 一般来说，三维形状作为深度神经网络的输入方式中典型的有基于多视图方式、基于体素方式等. 其中，基于多视图的输入方式[7-9]通过拍摄 3D 模型多个视角下的照片并使用处理二维图像的方式处理三维形状，然而该方式难以有效地表示三维完整模型，如复杂模型的部分结构信息往往会由于模型自身遮挡导致难以获得被遮挡部分的形状视图. 基于体素的输入方式[10-14]则将三维形状嵌入 3D 体素空间中，并将体素空间信息输入深度神经网络进行训练，然而该方式由于体素空间的三维特点导致其输入数据量和神经网络参数规模成倍增长，受制于有限的 GPU 存储空间，普通神经网络将难以处理高分辨率的体素模型. 随着基于离散点云数据输入的点云分类与分割网络 PointNet[15]的提出，使得直接将点云数据作为神经网络的输入从而实现点云形状的建模和处理成为可能. 相比于体素模型表示，直接输入点云将大大减少输入数据量和神经网络的参数规模，使得网络训练速度得到极大提高，同时也能够完整保留输入模型的全部信息. 因此，构建针对三维点云数据输入的形状修复补全网络成为三维修复补全中的一个重要问题.

三维室内场景修复通常存在两个难点：一是室内场景数据量庞大（场景点云数据通常有十几万到几百万的采样点），难以全部输入深度神经网络中直接进行训练；二是室内家具形状结构的缺失难以通过传统孔洞补全方法进行修复补全. 为了克服上述难点，本文提出了一种基于类别-实例分割的室内点云场景修复补全框架，该框架采取端到端的模式，即输入是大规模缺失点云场景，输出是大规模补全点云场景. 为了解决室内场景数据量大的问题，本文首先利用场景分割模块将大规模的室内点云场景分割成待修复的单个物体并作为补全网络的输入，从而避免了点云补全网络难以直接处理大规模场景数据的弊端. 为了能够补全室内家具形状结构的缺失，借助于端对端深度神经网络框架，同时为了在点云补全中能有效提取点云形状特征和邻域点局部特征信息，结合 PointSIFT[16]特征提取模块，本文提出了一种能够自动修复三维点云形状的编码器-解码器结构，从而最终实现室内点云场景的修复补全.

本文主要贡献在于：(1) 与以往场景补全方法大多基于体素数据表示的策略不同，提出了一个直接以场景点云数据作为输入的修复补全框架，该框架能有效实现室内场景的修复补全；(2) 基于编码器-解码器结构，结合 PointSIFT 的邻域点特征提取能力，提出了一个用于补全室内场景家具形状结构缺失的端到端深度神经网络；(3) 提出了一种用于室内场景分割的“类别-实例”分割策略，以有效分割出缺失的室内场景家具点云形状.

本文第 2 节介绍现有的三维形状修复补全方法；第 3 节详细阐述本文提出的室内点云场景修复补全框架与技术细节；第 4 节给出实验结果与讨论分析；最后一节对本文工作进行总结.

2　相关工作

现有的三维形状修复补全方法按照补全对象来分可以分为单个物体的形状修复和场景修复.

2.1　单个物体的形状修复

单个物体的形状修复方法大致可以分为基于几何、基于检索匹配和基于学习的方法.

2.1.1　基于几何的方法

基于几何的方法往往通过缺损形状中的几何线索辅助完成三维形状的修复补全. Kazhdan 等人[17]提出 Poisson 表面重建方法，该方法采取隐式拟合策略，通过求解 Poisson 方程得到点云模型所描述的形状表面隐式方程，并对该隐式方程进行抽取等值面，从而得到其形状表面模型. Pauly 等人[18]提出有效识别提取输入缺失形状的对称轴、对称结构以及重复结构的方法，并利用对称性实现对局部缺失部分的修复填充. 这些方法通常要求输入的三维形状及形状结构基本完整，从而能够利用缺失区域邻近的形状几何线索推断空洞区域的形状信息，但是其对实际获取的离散点云数据由于几何与结构线索提取的困难而并不适用.

2.1.2　基于检索匹配的方法

基于检索匹配的方法是将部分缺失模型与大型形状数据库中模型进行匹配完成三维形状的补全. Li 等人[19]通过对已有模型的对齐和缩放变形，并将变形得到的数据库中完整模型直接替代扫描缺失的模型，实现形状三维重建. Kim 等人[20]将数据库中的三维形状分割成若干部件，并生成一种基于部件的概率模板，对于缺失的待修复模型则从数据库中检索出其目标部件，并通过组装目标部件至缺失形

状上以实现模型的修复补全. Pauly 等人[21]提出先从数据库检索近似的候选模型，并通过对输入的待修复模型和候选模型进行非刚性对齐以完成三维形状的补全. 然而，该类方法的有效性通常依赖于数据库中庞大的模型数量规模和丰富的模型类型.

2.1.3 基于学习的方法

基于学习的方法通常使用深度神经网络进行形状的修复补全，该类方法直接将数据缺失的形状输入到神经网络中以修复生成三维完整形状. Wu 等人[5]提出的 3D ShapeNets 方法能从原始 CAD 模型数据中学习不同类别/姿态的三维形状分布并自动学习其形状部件的分层次组成，从而实现物体识别和形状修复，他们同时构建了 ModelNet 数据集. Thanh 等人[10]利用端到端的预测神经网络，基于马尔科夫随机场计算距离转换直接实现了扫描数据的修复. Sharma 等人[11]提出一种全卷积体积自动编码器结构，利用该结构能够通过噪声数据预测三维物体完整形状. Dai 等人[12]提出一种 3D 编码解码器，该编码解码器能完成从低分辨率、缺失的形状输入到高分辨率、完整的形状输出. Wang 等人[13]提出一种基于生成对抗网络并结合循环卷积网络的网络架构用于形状补全的目的. Han 等人[14]提出一种结合全局结构推理网络与局部几何优化网络实现对单个物体进行高分辨率形状补全的方法. 然而，上述这些神经网络都是基于三维模型的体素数据输入，往往由于 GPU 存储空间的有限，限制了其对大规模三维模型的有效处理.

2.2 场景修复

早期的场景修复方法大多不直接修复缺失的场景，而是事先修复缺失的 RGB 数据，进而生成完整的场景. Hays 等人[22]提出了一种基于 100 万张图像作为补全数据库的图像补全方法. 对于输入的缺失 RGB 图片，该方法在数据库中找到相似的图像区域来修补图像中的缺失，从而通过三维重建完成场景的修复. Li 等人[23]提出一种基于场景变换和颜色变换的图像修补算法，通过对比图像的纹理、颜色、结构信息等特征，从图像数据库中搜寻外观最相似的图像，进而完成场景修复. 近年来研究者提出的场景修复方法大多是基于深度学习的方法. Song 等人[24]提出一个预测场景体素占有率和语义标签的端到端网络，该网络以一张深度图为输入，能够同时输出完整的体素化场景和场景内物体的语义标签. Dai 等人[25]提出一种数据驱动的三维扫描场景补全方法，该方法利用全卷积自回归方法来预测完整的几何形状和三维语义分割. Hou 等人[26]提出一个以 RGBD 扫描数据为输入，将彩色图像信息和几何信息结合起来进行训练的自编码-解码结构的深度神经网络，该网络能够输出体素化的完整场景结构和语义分割信息. 然而，上述这些方法都是输出体素化补全场景，受制于 GPU 存储空间，补全的场景往往分辨率不高，难以显示复杂场景的丰富细节.

本文提出的室内场景修复补全框架直接使用场景点云数据作为输入，相较于传统体素输入的深度学习方法，点云数据往往占用相对较少的 GPU 内存，从而能有效地提高点云修复神经网络的训练速度. 同时，由于点云模型直接采用离散采样点数据表示，通常能方便准确地表达三维形状的精细结构信息并作为神经网络的输入. 考虑到室内场景点云数据的缺失往往分为平面结构的孔洞缺失以及室内家具形状结构的缺失，本文着重考虑修复补全第二类情形的场景点云数据缺失. 本文提出的框架主要包含 2 个模块，一个是场景“类别-实例”分割模块，由基于 PointNet[15]的类别分割网络和基于聚类的实例分割模块完成；另一个是点云形状补全模块，由基于编码器-解码器结构的点云补全网络实现.

3 室内点云场景修复补全

本文提出一种针对室内场景点云数据的修复补全框架，首先通过点云分割网络将室内场景进行“类别-实例”分割，将整体的场景修复补全问题转化为各个家具点云形状的修复补全任务；然后将分割出的家具点云形状作为点云补全网络的输入，并利用神经网络学习生成高质量的完整三维模型；最后，将修复补全后的家具点云模型融合到室内点云场景中，从而实现缺失场景点云数据的修复补全. 本文方法总体框架如图 1 所示.

3.1 点云场景类别-实例分割

考虑到室内场景的点云数据量过于庞大(一般为几十万到几百万规模的采样点数目)，难以一次性将室内场景直接输入到神经网络中. 本文基于“类别-实例”的分割策略对大规模室内点云场景进行有效分割. 结合场景分割，采用“分而治之”思想，将一个整体的室内场景修复补全任务分解成多个家具形状的修复补全任务，从而有效避免了直接处理庞大的点云场景数据修复补全的难题. 此外，将室内场景缺失点云数据的补全问题转化成处理采样点数目较少的室内家具点云数据，使得点云补全网络能够正常训练并高效修复. 其中，场景类别分割基于

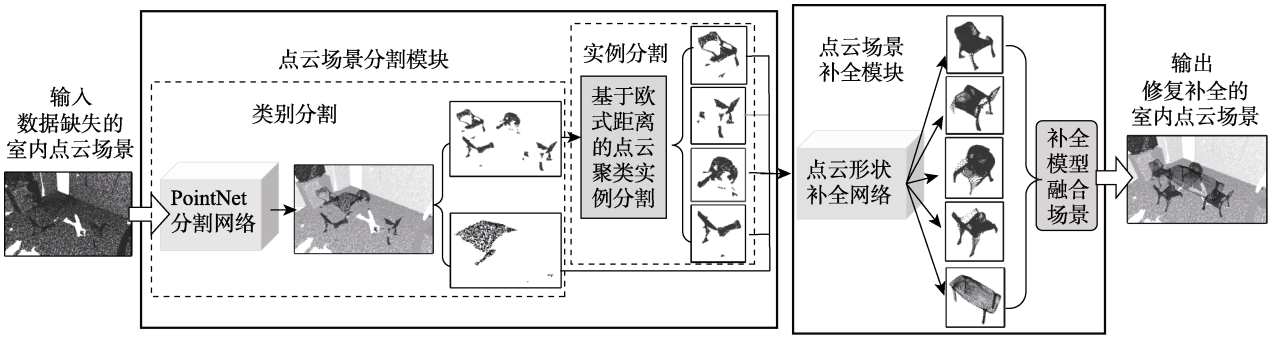


图 1　室内点云场景修复补全框架

PointNet[15]分割网络，主要目的是分割出室内场景的语义类别，如地面、墙壁、桌子、椅子等；场景实例分割则采用基于欧式距离的点云聚类方法，主要目的是将由 PointNet 分割出的语义类别组进一步分割成单个家具实例.

基于 PointNet [15]网络的室内场景物体类别分割的作用是分割出室内场景中的缺失家具类别. 相比目前的分割网络往往融合了邻域点特征信息如 PointNet++[27]等，PointNet 只提取了每个点自身的特征而没有融合邻域点的特征信息. 但由于本文的室内场景点云数据输入通常为不完整数据，并不具备缺失采样点的邻域信息. 因此，利用 PointNet 分割有缺失的室内场景点云效果将更好. 此外，由于室内场景可能有多个同类物体，如会议室场景内有多把椅子，仅采用 PointNet 网络的类别分割无法得到每把椅子实例，难以实现场景中家具形状的精确修复. 因此需要进一步实现室内场景点的实例分割.

点云场景的实例分割方法典型的有基于区域生长的分割方法[28]，但是该方法往往适用于曼哈顿结构场景，对于一般的非结构化室内场景其分割效果并不理想；而基于最小分割的点云分割方法[29]则适用于多个物体水平排列的情况，不能应用于室内场景中不同类别物体的实例分割. 本文采用基于欧式距离的点云聚类进行实例分割，该方法通常具有分割较稳定、对场景限制较少等优点. 基于欧式距离的点云聚类实例分割算法如下：

算法 1.基于欧式距离的点云聚类实例分割

输入：待分割的场景点云数据 P

输出：分割出的点云聚类 C

1.为输入点云 P 创建 Kd 树；

2.建立空列表 C 以及队列 Q；

3.遍历 P，对于每个点 $\boldsymbol{p}_i \in P$，进行如下操作：

- 将 $\boldsymbol{p}_i$ 加入当前队列 Q；
- 遍历队列 Q 中的点 $\boldsymbol{p}_i$：
  - 搜寻 $\boldsymbol{p}_i$ 在半径为 r 的球内的 K 个邻近点集 P_i^k；
  - 对邻近点集 P_i^k 内的点 P_i^k，检查点是否在 Q 内，若不在，则将其添加到 Q；
- 若 Q 被遍历完毕，将 Q 添加到聚类 C 的列表中，并将 Q 重置为空队列；

4.当点云 P 的所有点已被处理并且现在是聚类列表 C 的一部分时算法终止.

图 2 为餐厅点云场景的类别-实例分割示意图. 图 2(a)为输入的原始点云数据；图 2(b)为场景经类别分割的结果；图 2(c)和图 2(d)分别为对餐厅座椅类别、餐厅餐桌类别进行实例分割的结果. 可以看出，餐厅点云场景经过类别分割被分成墙壁、地板等平面结构和座椅、餐桌等家具类别，并将缺失的座椅和餐桌类别进行实例分割，最终分割出 8 把座椅和 2 张餐桌. 其中，点云邻域半径 r 取 0.05，场景类别分割耗时约 8 秒；实例分割耗时约 0.6 秒.

3.2　点云形状修复补全

在离散三维点云数据的修复补全过程中，由于过多的点云数据特征将会使得神经网络的训练效率降低，同时也难以分辨点云数据中含有的重要特征信息，故本文的点云形状补全网络采用编码器-解码器结构，该结构可以有效提取输入点云的形状特征信息，并将其编码成具有较小规模数据量的压缩信息，最后通过解码器重构生成点云形状的缺失部分数据.

本文基于编码器-解码器结构的点云形状补全网络是一种端对端神经网络，如图 3 所示. 其中神经网络以 N 个采样点的点云模型作为输入，经过编码器编码得到其特征码字并结合二维网格数据作为解码器的输入，解码器通过 2 次折叠操作生成 M 个点作为补全完成的点云模型. 图中各感知器下方括号中的数字表示每层感知器的输出维度. 本文方法利用含有缺失模型和完整模型的大规模数据集，训练神经网络直接从缺失点云模型预测完整点云模型，通

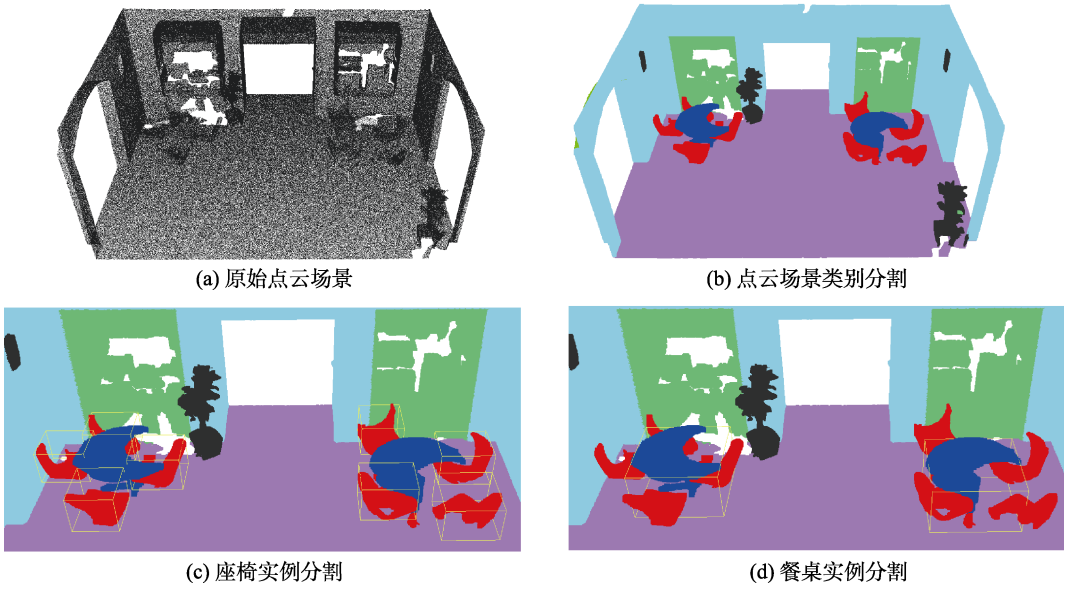


图 2 餐厅点云场景的类别-实例分割

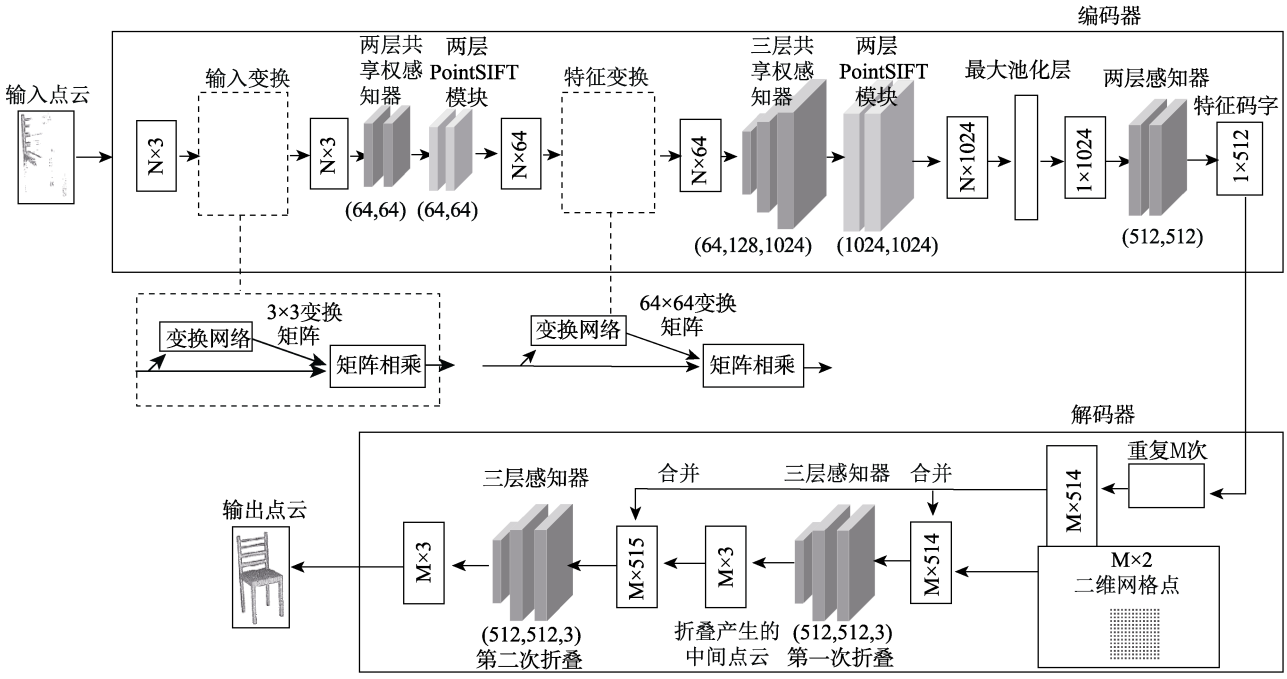


图 3 点云形状补全网络结构

过监督学习策略解决三维点云形状的修复补全问题. 该方法基于点云数据作为输入，采用基于 PointNet[15] 的编码器结构和基于 FoldingNet[30]的解码器结构，该网络训练稳定且能生成丰富的修复补全样例.

实际上，点云形状修复补全网络中的编码器采用 PointNet[15]编码结构，主要是由于该编码结构通过变换网络模块能够有效解决点云的旋转性问题，通过最大池化层解决点云的无序性问题，并且通过多重权共享感知器提取点云中每个点的对应特征. 此外，本文还对提取出的特征利用 PointSIFT[16]模块进行邻域特征加强，PointSIFT 模块可以从 8 个方向获取每个点的邻域特征信息，从而获取更加准确的特征码字. 解码器则基于折叠操作，这种折叠结构是一个通用的点云重建框架，FoldingNet[30]指出只要提供适当的特征码字，折叠操作可以将一个二维网格折叠成任意的形状，故本文采用基于折叠的解码器.

3.2.1 基于 PointNet 网络的编码器结构

如图 3 所示，点云形状补全网络的编码器输入一个 $N\times3$ 的矩阵，矩阵的每一行由采样点坐标 (x,y,z) 组成，解码器输出为 $M\times3$ 的矩阵表示修复补全完成的点云形状.

本文采用 PointNet[15]网络的编码结构作为编码器，该编码器结构能够有效解决点云数据输入的旋转性和无序性问题. 旋转性是指同一点云形状可以旋转从而产生不同的输入数据，而不同输入数据的输入不能影响其补全效果. PointNet 的变换网络能够预测仿射变换 3×3 位置矩阵和 64×64 的特征变换矩阵，并将这 2 个变换矩阵直接乘以输入矩阵和特征矩阵，从而用于配准不同输入点云数据的位置姿态和特征. 无序性问题是指由于点云数据是由一组无特定顺序的离散点组成，对输入而言其中三维离散点的顺序应不影响其在空间中对三维整体形状的表示，即不同顺序数据的输入不能影响其补全效果. PointNet[15]使用最大池化层来提取每个点对应的维度最大的特征，从而解决了无序性问题. 由于卷积层能有效地提取点云数据的特征，本文先采用多层共享权感知器提取各采样点的对应特征，再用最大池化层提取全局特征，该全局特征的各维均选取 N 个点对应维度的最大值，其大小是 1×1024 .

然而，利用 PointNet 网络中的编码器提取模型特征时通常只提取单个采样点的局部特征，而并没有考虑到点云形状中采样点与采样点之间的关联特征信息. 结合邻域采样点特征信息的神经网络（如 PointNet++[27]）其分类/分割正确率通常大于仅考虑单个采样点局部特征信息的神经网络. 本文提出的点云补全网络中编码器提取的全局特征中同时包含单个采样点及其邻域点的特征信息，能准确刻画采样点特征信息. 具体地说，在提取特征信息的共享权感知器后加入 2 层 PointSIFT[16]模块对各采样点的对应特征融合邻域点特征信息，从而加强了各采样点的特征信息，此模块的输出维度与特征的输入维度一致.

PointSIFT[16]模块能够对每个点的 K 维特征从空间中的 8 个卦限子空间寻找其对应的最近邻点的 K 维特征，并形成 8 个方向的特征向量；然后将特征向量分别在 x，y，z 坐标轴上做卷积操作，生成一个 K 维的融合邻域点特征的新特征. 值得注意的是，一般的 K 近邻方法会寻找全局的近邻点，而 PointSIFT 能够从 8 个卦限子空间寻找近邻点. 如图 4 所示，一般的寻找 K 近邻方法会选择离中心点最近的 K 个点（图中的圆圈点）. 有时，这些点会聚集在一个方向而限制了邻近特征的表达，而 PointSIFT 将选择不同方向的点（图中的灰色点），分散的近邻点能够更准确地提供采样点特征信息，模型的特征表示能力会更好.

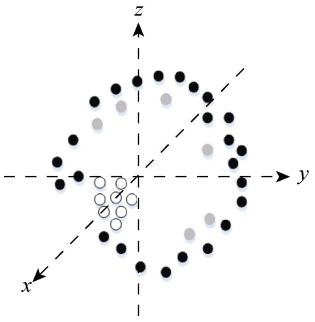


图 4　K 近邻方法与 PointSIFT[16]寻找最近邻点

3.2.2　基于折叠的解码器结构

正如 FoldingNet[30]指出，在提供适当特征码字的前提下，二维网格能够通过 2 次折叠操作生成任意的点云形状. 因此在解码器构建中，为了能补全室内家具的结构化形状缺失，本文的解码器采用 2 次折叠网格形状生成补全的点云数据. 具体来说，对输入的待修复点云数据经编码器编码得到 1×512 的特征码字矩阵（与 FoldingNet 类似,这里的“512”表示经卷积操作后，待修复点云数据在一维特征空间中表达为 512 个特征码字），再将该特征码字矩阵重复 M 次，得到 $M\times512$ 的矩阵作为解码器的输入. 然后，生成一个网格平面，此网格是以原点为中心的正方形网格，该正方形网格总共包含 M 个网格点，所有各点的坐标值形成一个 $M\times2$ 的输入矩阵，每一行为各网格点的（x，y）坐标，其坐标值分别位于[–1,1]内. 将网格点数据与特征码字矩阵合并得到 $M\times514$ 的矩阵，对该矩阵通过三层感知器进行第一次折叠，生成 $M\times3$ 的中间点云. 再将特征码字矩阵与中间点云合并得到 $M\times515$ 的矩阵，通过三层感知器中进行第二次折叠. 最后得到大小为 $M\times3$ 的重建完成的点云数据. 由于二维网格中的点是均匀分布的，因而修复补全的点云形状表面往往更加平滑，采样点分布更加均匀有序.

折叠操作能将二维网格数据映射到三维点云数据是因为 2 次折叠操作相当于施加一个把二维网格进行变形、切割、拉伸操作并将其变成需要的点云形状，而其中特征码字则存储了折叠所需的“力”. 由于多层感知器能够有效模拟诸如变形、切割、拉伸等操作，故采用多层感知器实现二维网格的复杂折叠操作.

3.3　损失函数

本文提出的点云形状修复补全网络中损失函数的定义则用于度量修复点云和真实点云之间的差异，由于离散点云数据具有无序性，因此要求损失

函数对点云数据的顺序并不敏感. 本文采用倒角距离(Chamfer Distance，CD)[31]和推土机距离(Earth Mover's Distance，EMD)[31]用以度量修复点云与真实点云之间的差异程度. 对修复补全的点云模型 S_1 和真实点云模型 S_2 计算：

$$CD(S_1,S_2)=\frac{1}{|S_1|}\sum_{\boldsymbol{p}\in S_1,\boldsymbol{q}\in S_2} min\left\|\boldsymbol{p}-\boldsymbol{q}\right\|_2+\frac{1}{|S_2|}\sum_{\boldsymbol{q}\in S_2,\boldsymbol{p}\in S_1} min\left\|\boldsymbol{q}-\boldsymbol{p}\right\|_2$$

其中 $|S_1|$ 和 $|S_2|$ 分别为相应点云模型的采样点数目. 倒角距离 CD 误差计算修复补全点云模型 S_1 和真实点云模型 S_2 之间的平均最近点距离，其中第一项让修复补全点云靠近真实点云，第二项让真实点云覆盖修复补全点云.

$$EMD(S_1,S_2)=\min_{\phi:S_1\to S_2}\frac{1}{|S_1|}\sum_{\boldsymbol{x}\in S_1}\left\|\boldsymbol{x}-\phi(\boldsymbol{x})\right\|_2$$

推土机距离 EMD 误差则需要寻找一个双射函数 $\phi: S_1\to S_2$，使对应点之间的平均距离最小. 实际中寻找双射函数的代价过于昂贵，通常采用 Bertsekas 提出的迭代近似方案来代替双射函数[32].

在损失函数的构造中，为了能够在点云形状修复中既保持待修复物体的整体形状，又能恢复待修复物体的细节结构，结合优化 CD 距离实现物体整体形状的保持和优化 EMD 距离实现物体细节结构的恢复[33]的目的，本文提出的损失函数同时兼顾两种距离误差，将 CD 距离误差与 EMD 距离误差同时进行优化训练. 具体形式如下:

$$L(S_1,S_2)=CD(S_1,S_2)+EMD(S_1,S_2)$$

4 实验结果与分析

本文提出的点云场景修复补全框架在 CentOS 系统下得到了实现，程序的硬件环境为 CPU 处理器 Intel Xeon Gold 6148，主频 2.40GHz；GPU 为 NVIDIA Tesla V100，显存 8B. 软件平台：Python 3.5.3，Tensorflow 1.12.0.

由于本文的场景修复补全框架用于补全室内场景家具的形状缺失，其补全任务由点云形状补全网络完成，故室内家具物体补全的对比验证性实验将使用本文的点云补全网络进行有效性验证. 本文采用来自 ModelNet40 数据集[5]中的模型创建包含缺失点云模型和完整点云模型的大规模数据集. 为此从数据集的 40 种物体类别中挑选了 2048 个三维物体网格模型，通过对选取的网格模型均匀采样得到固定数量的点云数据作为完整点云数据集；而缺失点云数据集则通过对完整点云数据集的部分部位进行去除缺失得到. 在点云修复时，形状补全网络首先选取一个批次的包含若干缺失点云及其对应真实点云的数据作为输入，其中缺失点云数据输入编码器并获得用于折叠的码字信息，该码字信息将结合二维网格坐标信息用于生成补全的点云数据，即完整的三维点云形状. 该过程中由于使用基于 PointNet[15]网络的编码器，此编码结构不但解决了点云的旋转性和无序性问题，还能够提取每个点对应的特征信息，同时对提取得到的特征信息通过 PointSIFT[16]模块增加了其邻域点的特征信息. 最后得到的特征码字将含有每个采样点的独立特征信息和邻域点特征信息，因而能够更好地表示补全点云的形状特征. 解码器则使用折叠操作，将网格形状根据特征码字折叠成完整补全的点云形状，由于网格中的点分布均匀，因此修复补全的点云形状表面往往更加平滑，采样点的分布更加均匀.

4.1 数据集及预处理

为了构建实验需要的点云形状的训练数据集，本文对 ModelNet40 数据集[5]中选取的三维网格模型进行均匀采样，将模型表面采样 1024 个点作为真实点云；并对完整点云形状进行下采样得到实验用的缺损点云模型. 对每个完整点云模型进行 9 次随机部位去除处理，每个完整点云对应 9 种不同程度、不同部位的缺损点云，再将缺损点云点数量采样统一化，其中缺损点云形状的采样点数目为 540. 在缺失采样过程中，先对点云模型进行随机点的挑选，然后以该采样点为中心，以 r 为半径的球体范围进行搜索（实验中 r 取模型大小的 25%），再对该范围内的采样点进行随机缺失处理，进而得到缺失不同部位的点云形状. 其次，将每个完整点云模型再进行一次缺失处理，获得测试集. 同时，为了使修复补全网络能够快速收敛以便于训练，需要对缺失点云模型和完整点云模型的采样点坐标进行零均值归一化，即将采样点坐标值分量均归一化至[–1,1]区间内. 本文借助 Adam 优化器训练点云修复补全网络，其学习率设为 0.0001，持续 100 个周期，批次大小为 32.

4.2 点云补全网络修复效果

本文使用编码器-解码器结构提出了一种三维点云数据的修复补全网络结构. 针对 ModelNet40 数据集[5]，利用本文提出的点云补全网络验证了该神经网络结构在点云形状修复补全中的有效性和优良表现. 本文的点云补全网络在测试集数据上的修复

补全效果如图 5 所示. 同时，由于输入的点云模型缺失程度可能不同，而对于点云修复补全网络应该具备能够修复不同程度缺失的能力. 因此，需要测试本文提出的点云补全网络的鲁棒性，图 6 分别给出了对具有 25%、50%、75%缺失程度的点云模型进行修复补全. 可以看到，飞机、路锥、马桶等点云模型在不同的缺失程度下均能够较好地保持输入点云形状的结构信息. 实验表明，本文网络对不同程度的损失输入依然能较好地补全缺失部分，具有较强的鲁棒性.

此外，深度神经网络具有泛化性，所谓泛化性是指其具有能力生成不在数据集中的数据. 本文挑选了 ModelNet40 数据集[5]中的同类数据模型，按照前述的处理方式生成测试点云数据，以便对不在此数据库中的缺失数据进行补全. 图 7 给出了对具有 25%、50%、75%缺失数据的点云模型进行泛化性修复补全. 可以看到，对于不同程度缺失的模型，本文的点云形状补全网络生成了 3 种造型不同的杯子、折椅生成了靠背椅、洗手池生成了带有平面的洗手池. 这是由于测试数据不在测试集中，故可能生成特征空间内的记忆形状，而与输入不相符. 需要指出的是，利用深度神经网络进行修复补全的目的是通过网络训练能够提取并记忆不同种类形状数据的内在特征，并形成特征空间，从而使得点云修复补全网络能够在特征空间中较好地补全处于数据集涵盖的分布范围内点云数据. 由实验可知，其修复效果表明本文方法能够对不在训练数据集中的模型进行有效补全，本文网络具有良好的泛化性.

(a) 输入模型　　(b) 输出模型　　(c) 真实模型

图 5　点云补全网络的修复效果

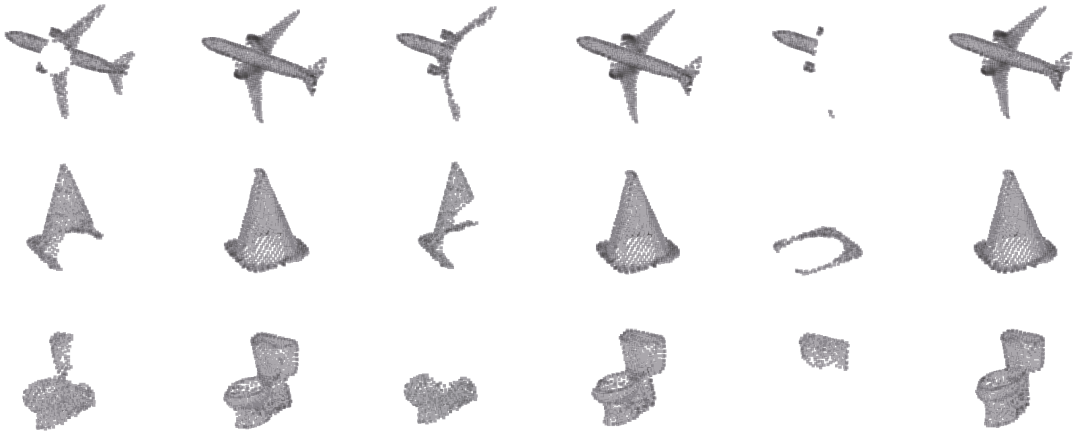
4.3　与不同的点云数据修复方法的比较

为了与本文提出的点云补全网络的修复效果进行比较，将采用相同的数据集分别在以下的网络中进行补全效果比较：

1) 不含 PointSIFT[16]模块的补全网络：该网络的结构与本文提出的点云修复补全网络的结构一致，不同的是，该网络不含有 PointSIFT 模块. 因此该网络提取的特征码字不含有邻域点信息.

2) 不含网格数据的补全网络：该网络编码结构与本文提出的点云补全网络的编码结构一致，但解码器不采用二维网格数据用于折叠，而是采用符合高斯分布的噪声点采样数据替代规则的二维网格数据并用于折叠操作.

3) 点云修复 PCN 网络[33]：该网络采用编码器-解码器结构，其编码器通过多层权值共享感知器和



(a) 缺失25%的输入 (b) 缺失25%的补全 (c) 缺失50%的输入 (d) 缺失50%的补全 (e) 缺失75%的输入 (f) 缺失75%的补全

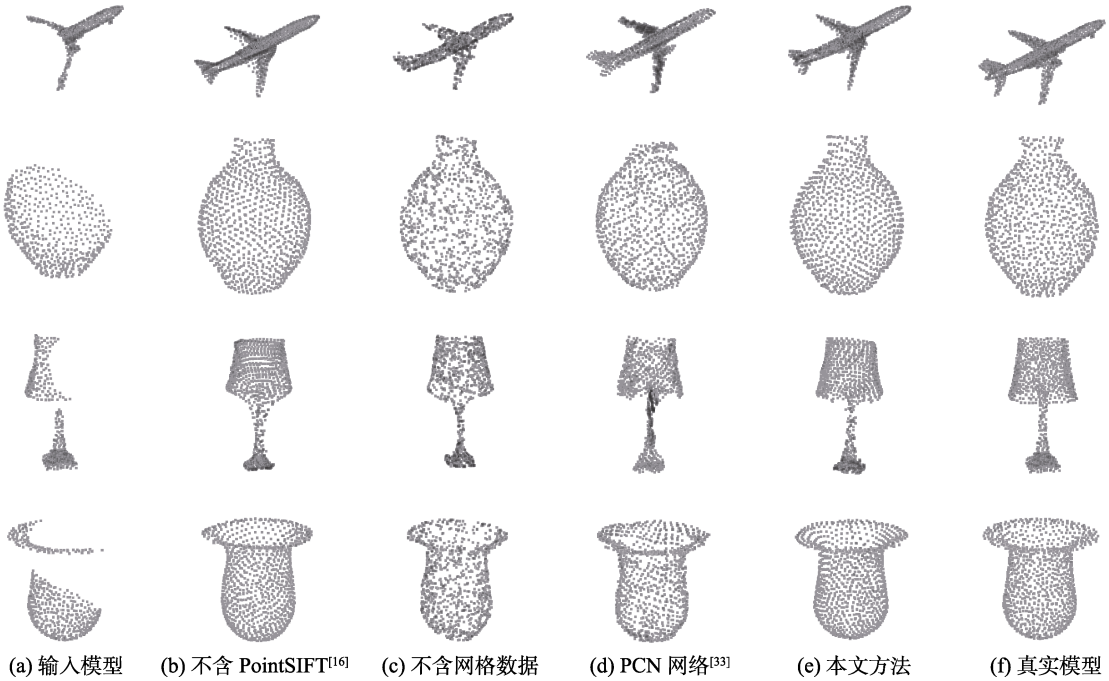
图 6　针对不同缺失程度的点云模型修复补全效果

(a) 缺失25%的输入 (b) 缺失25%的补全 (c) 缺失50%的输入 (d) 缺失50%的补全 (e) 缺失75%的输入 (f) 缺失75%的补全

图 7 针对不在训练数据集中的模型补全的泛化性实验

2 次最大池化层提取形状的全局特征. 解码器通过多层感知器生成一个点数较少的粗略点云模型，并使用添加正方形形状的点云补丁对粗略点云模型进行采样点数目扩充，最终得到完整的修复点云模型.

本文使用相同的训练集和测试集，分别利用本文方法与上述 3 个网络进行训练与模型修复，网络训练中采用相同的训练周期、学习率和批次大小. 图 8 给出了在测试集上的修复效果比较，其中输入的点云模型具有 50%的缺失数据. 可以看到，本文网络相较于不含 PointSIFT[16]的网络，补全效果两者大体一致，但是不含 PointSIFT 的网络产生的游离采样点数量比本文网络较多，如飞机机翼附近、台灯灯罩边的游离采样点等. 相较于不含网格数据的补全网络，本文网络的补全模型由于含有排列规则的网格数据，整个模型的补全效果点云分布较为均匀，而不含网格数据的补全网络其补全得到的模型点的分布较为嘈杂，补全质量不高. 利用点云修复 PCN 网络[33]的补全模型也产生了点云分布不均匀，出现点云聚集的情况，如飞机机翼、台灯灯把等. 实验表明，本文的点云补全网络较好地补全了缺失的点云数据，本文方法的修复补全效果表现优于其它 3 种网络.



(a) 输入模型 (b) 不含 PointSIFT[16] (c) 不含网格数据 (d) PCN 网络[33] (e) 本文方法 (f) 真实模型

图 8 针对不同网络的修复补全效果比较

表 1 给出点云修复模型与真实模型的倒角距离 CD 误差值和推土机距离 EMD 误差值统计(其中输入的点云模型具有 50%的缺失数据)，对于 2 个距离误差而言，误差值越接近于零表明其形状越接近对应真实点云模型的形状. 从表 1 中可知，本文方法的 CD 误差和 EMD 误差均比不含 PointSIFT[16]网络的对应误差小，表明使用 PointSIFT 模块融合近邻点特征信息的编码器能够更好地表示模型的特征，并生成接近真实的点云模型. 本文方法的 CD 距离误差和 EMD 距离误差均比不含网格数据的网络的对应误差小，表明了网格折叠操作的有效性，能够更好地逼近真实点云. 表 1 数据表明，本文网络在 CD 距离误差和 EMD 距离误差通常都取得了较小的值，表明生成的点云模型更接近真实模型.

在损失函数的选择比较方面,相较于单独使用倒角 CD 距离训练本文提出的点云形状修复补全网络与单独使用推土机 EMD 距离训练网络中，在相同的网络框架下分别考虑仅使用 CD 距离、仅使用 EMD 距离和本文 CD+EMD 距离作为损失函数并用于神经网络训练. 表 2 统计了实验中不同点云模型在利用不同损失函数定义下的修复补全网络进行模型补全后修复模型与真实模型之间的 CD 距离误差、EMD 距离误差和 CD+EMD 距离误差. 表 2 数据表明，本文采用的 CD+EMD 距离损失函数相较于仅使用 CD 距离的损失函数时得到的不同点云模型修复结果均取得了较小的误差，说明使用本文的损失函数比单独使用 CD 距离作为损失函数往往更能使生成的补全模型接近真实模型. 然而，本文采用的 CD+EMD 距离损失函数相较于仅使用 EMD 距离的损失函数时得到的不同点云模型修复结果的误差比较中，两者从 CD 距离误差、EMD 距离误差和 CD+EMD 距离误差来看总体差别不大，说明使用本文的损失函数与单独使用 EMD 距离损失函数均能够得到相近的效果. 在神经网络的训练时间比较方面，分别采用上述 3 种不同损失函数定义下的点云形状修复补全网络在一个周期(也就是将所有训练集数据均训练一遍)内所耗费的训练时间均需要约 4 分钟，因而采用不同的损失函数所耗费的网络训练时间差别不大.

表 1　不同网络修复点云模型的倒角距离 CD 误差和推土机距离 EMD 误差统计

距离类型	点云模型	距离误差			
		不含 PointSIFT[16]	不含网格数据	PCN 网络[33]	本文网络
CD 距离	飞机	0.04528	0.05033	0.04336	0.04194
	坛子	0.06760	0.07931	0.07101	0.06328
	台灯	0.06575	0.06960	0.05704	0.05719
	阔口瓶	0.06573	0.07748	0.07169	0.06469
EMD 距离	飞机	0.04966	0.06582	0.11016	0.04928
	坛子	0.06075	0.08816	0.07773	0.05838
	台灯	0.09072	0.09781	0.15522	0.05531
	阔口瓶	0.08042	0.10059	0.09913	0.06563

表 2　不同损失函数定义下修复点云模型的误差统计

点云模型	仅使用 CD 距离训练			仅使用 EMD 距离训练			本文损失函数训练		
	CD 距离	EMD 距离	CD+EMD	CD 距离	EMD 距离	CD+EMD	CD 距离	EMD 距离	CD+EMD
飞机	0.03921	0.07542	0.11463	0.04049	0.05109	0.09158	0.04194	0.04928	0.09122
坛子	0.06274	0.06601	0.12875	0.07099	0.06274	0.13373	0.06328	0.05838	0.12166
台灯	0.06320	0.30325	0.36645	0.07674	0.09170	0.16844	0.05719	0.05531	0.11250
阔口瓶	0.06049	0.09044	0.15093	0.06399	0.05588	0.11987	0.06469	0.06563	0.13032

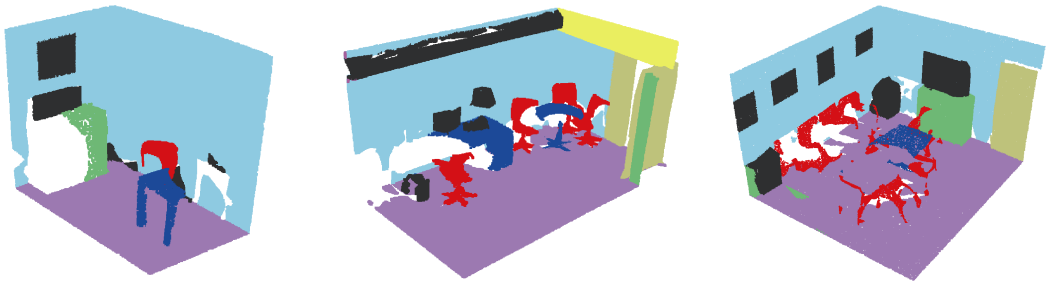
4.4　室内点云场景修复补全

为了验证本文提出的修复补全框架在室内点云场景修复中的有效性，本文使用 Stanford 大型三维室内场景数据集 S3DIS (Stanford 3D Indoor Semantics Dataset)[6]作为提供训练与验证模型的数据集. S3DIS 中有 6 个区域，每个区域有类似会议室、办公室、走廊、仓库、厕所等几十种点云场景，每个点云场景中含有桌子、椅子、沙发等常见点云家具模型. 本文挑选数据集中的家具点云模型按前述方式生成训练模型数据库(真实点云模型的点数量取 4096，缺损的点云采样点数目为 1792)，将其放入点云补全网络中训练. 测试集的生成由随机 1 个

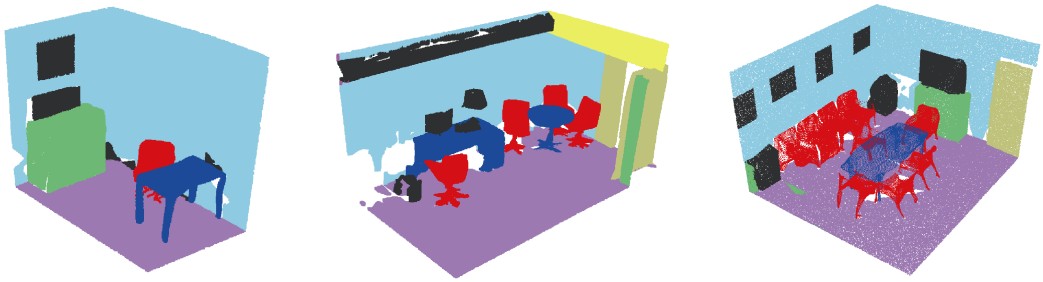
区域中的室内场景做缺失处理，形成缺失的点云场景. 测试时将缺失的点云场景输入点云场景分割模块得到分割完成的点云模型，再将每个模型输入点云补全网络中生成完整的补全点云模型，最终将补全的点云形状融合进场景中完成室内场景的修复补全. 如图 9 所示，工作室场景中有柜子、四脚桌等的数据缺失，办公室场景中有电脑桌桌面、电脑椅座面、圆桌桌腿等的数据缺失，会议室场景中有桌子桌面、桌子腿、椅子靠背及座面、椅子腿等的数据缺失，这些均得到了有效补全. 同时，可以看到相对于场景的其他原始点云，经修复补全的家具模型其点云数据分布较均匀. 实验表明，本文提出的框架能够补全室内场景家具结构缺失；表 3 分别统计了图 9 中室内点云场景修复补全前后的采样点数目、场景分割出的实例及其个数、分割模块和补全模块的效率.

表 3　室内点云场景修复补全数据统计

点云场景	修复补全前/后采样点数目	分割类别/实例个数	分割/补全时间(s)
工作室	919926/945270	柜子/1 个 工作椅/1 个 四脚桌/1 个	8.427/7.523
办公室	1516065/1529889	圆桌/1 个 电脑桌/1 个 电脑椅/4 个	13.612/8.242
会议室	907253/914165	会议桌/1 个 四脚椅/9 个	8.281/9.563



(a) 缺失数据的室内点云场景



(b) 经修复补全的室内点云场景

图 9　室内点云场景的修复补全（从左至右分别是工作室、办公室和会议室场景）

目前典型的场景修复方法大多以单张、多张图片或体素化缺失场景为输入并且输出体素化的场景补全工作. 如 Dai 等人[25]和 Hou 等人[26]提出的方法均以缺失的体素化场景为输入，经过编码解码器结构输出预测的体素完成场景修复. 然而，体素化的室内场景表达往往丢失了室内场景的精细细节，若要输入高分辨率的体素场景，有限的 GPU 内存又难以训练过于庞大的体素空间数据. 相反地，本文提出的点云场景补全框架则直接处理大规模的点云场景数据，既保持了场景的精细细节结构又使得深度神经网络能够进行有效训练，其具体做法是利用“分而治之”思想将一个补全大规模的缺失场景数据任务转换成补全场景中分割出的一个个实例模型任务. 补全功能的实现则由本文提出的点云修复补全网络完成，由于采用编码器-解码器结构，需要将点云形状特征进行压缩表达，特征压缩的质量往往决定了补全的点云形状质量. 因此，点云数据的特征提取往往十分重要. 基于体素化的场景补全通常直接使用 3D 卷积操作对输入的体素空间直接卷积并提取场景特征，但是点云数据由于其旋转性和无序性等问题，不能够直接进行卷积提取特征. 本文利用编码器-解码器结构压缩点云形状的编码表达以

有效修复点云. 本文借鉴用于点云分类/分割的 PointNet[15]网络架构,该网络克服了点云的旋转性和无序性问题，并使用卷积操作提取每个点的特征. 同时，为了加强点云特征信息的自动提取，本文网络中引入 PointSIFT[16]增强了对各采样点的邻域点特征信息的提取，从而提升了三维离散点云形状的压缩表达.

5　结 语

针对室内场景的修复补全，本文提出一个以场景点云数据作为输入、端到端的室内场景修复补全框架；该框架能有效实现室内家具形状结构的缺失补全并输出补全的场景点云数据. 其中，点云场景分割模块能够分割出场景内的待补全缺失家具点云，从而将一个场景的修复补全任务转换成多个家具点云的修复补全任务，并解决了神经网络难以处理大规模点云场景数据修复的问题；点云补全模块采用编码器-解码器结构的深度神经网络，其编码器能够提取点云每个点的特征信息和其邻域点的特征信息，生成一个具有丰富折叠信息的特征码字，解码器根据此特征码字结合网格数据并采用折叠操作,利用 CD 距离和 EMD 距离以指导点云补全网络的不断优化，最终实现点云形状的修复补全，并有效解决了点云场景缺失数据补全的问题. 实验表明,本文提出的点云补全网络针对 ModelNet40 数据集取得了较好的修复补全效果；该方法对于不同程度的缺失模型均取得了较好的补全效果,具有很好的鲁棒性；对不在训练集中的模型也能较好补全，方法具有泛化性. 本文提出的场景补全框架在 S3DIS 数据集上较好地实现了室内家具形状结构的缺失补全.

然而，本文点云场景缺失数据的修复补全效果通常受场景分割结果的质量影响. 如果分割质量不高，如某个模型带有其它模型的一部分则难以实现针对实例模型的有效补全. 未来的工作可以考虑在场景分割和修复补全中加入几何与语义约束，以取得更加鲁棒的场景修复补全效果.

致　谢　本项工作中缪永伟教授受杭州市钱江特聘专家计划（杭州师范大学）资助.

参　考　文　献

[1] Gross M, Pfister H. Point Based Graphics. Burlington, USA: Morgan Kaufmann Publisher, 2007

[2] Miao Yong-Wei, Xiao Chun-Xia. Geometric Processing and Shape Modeling of 3d Point-Sampled Models. Beijing: Science Press, 2014(in Chinese)
(缪永伟, 肖春霞. 三维点采样模型的几何处理和形状造型. 北京: 科学出版社, 2014)

[3] Zhang Cong-Yi, Wei Zi-Zhang, Xu Hao-Wen et al. Scale variable fast global point cloud registration. Chinese Journal of Computers, 2019, 42(9): 1939-1952(in Chinese)
(张琮毅, 魏子庄, 徐昊文 等. 尺度可变的快速全局点云配准方法. 计算机学报, 2019, 42(9): 1939-1952)

[4] Chalmovianský P, Jüttler B. Filling holes in point clouds// Proceedings of the 10th IMA International Conference on the Mathematics of Surfaces. Leeds, UK, 2003: 196-212

[5] Wu Z, Song S, Khosla A, et al. 3d shapenets: A deep representation for volumetric shapes//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, Massachusetts, USA, 2015: 1912-1920

[6] Armeni I, Sener O, Zamir A R, et al. 3d semantic parsing of large-scale indoor spaces//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 1534-1543

[7] Su H, Maji S, Kalogerakis E, et al. Multi-view convolutional neural networks for 3d shape recognition//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile, 2015: 945-953

[8] Chen X, Ma H, Wan J, et al. Multi-view 3d object detection network for autonomous driving//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2017: 1907-1915

[9] Qi C R, Su H, Nießner M, et al. Volumetric and multi-view cnns for object classification on 3d data//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 5648-5656

[10] Thanh Nguyen D, Hua B S, Tran K, et al. A field model for repairing 3d shapes//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 5676-5684

[11] Sharma A, Grau O, Fritz M. Vconv-dae: Deep volumetric shape learning without object labels//Proceedings of the European Conference on Computer Vision. Amsterdam, The Netherlands, 2016: 236-250

[12] Dai A, Ruizhongtai Qi C, Nießner M. Shape completion using 3d-encoder-predictor cnns and shape synthesis//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2017: 5868-5877

[13] Wang W, Huang Q, You S, et al. Shape inpainting using 3d generative adversarial network and recurrent convolutional networks//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 2298-2306

[14] Han X, Li Z, Huang H, et al. High-resolution shape completion using deep neural networks for global structure and local geometry inference//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 85-93

[15] Qi C R, Su H, Mo K, et al. Pointnet: Deep learning on point sets for 3d classification and segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2017: 652-660

[16] Jiang M, Wu Y, Zhao T, et al. PointSIFT: A sift-like network module for 3d point cloud semantic segmentation. arXiv preprint arXiv:1807.00652, 2018
[17] Kazhdan M, Hoppe H. Screened poisson surface reconstruction. ACM Transactions on Graphics, 2013, 32(3): Article No. 29
[18] Pauly M, Mitra N J, Wallner J, et al. Discovering structural regularity in 3D geometry. ACM Transactions on Graphics, 2008, 27(3): Article No. 43
[19] Li Y, Dai A, Guibas L, et al. Database-assisted object retrieval for realtime 3d reconstruction. Computer Graphics Forum, 2015, 34(2): 435- 446
[20] Kim V G, Li W, Mitra N J, et al. Learning part-based templates from large collections of 3D shapes. ACM Transactions on Graphics, 2013, 32(4): Article No. 70
[21] Pauly M, Mitra N J, Giesen J, et al. Example-based 3D scan completion//Proceedings of the Eurographics Symposium on Geometry Processing. Vienna, Austria, 2005: 23-32
[22] Hays J, Efros A A. Scene completion using millions of photographs. ACM Transactions on Graphics, 2007, 26(3): Article No. 4
[23] Li H, Wang S, Zhang W, et al. Image inpainting based on scene transform and color transfer. Pattern Recognition Letters, 2010, 31(7): 582-592
[24] Song S, Yu F, Zeng A, et al. Semantic scene completion from a single depth image//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2017: 1746-1754
[25] Dai A, Ritchie D, Bokeloh M, et al. Scancomplete: Large-scale scene completion and semantic segmentation for 3d scans// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 4578-4587
[26] Hou J, Dai A, Nießner M. 3D-SIC: 3D semantic instance completion for RGB-D scans. arXiv preprint arXiv:1904.12012, 2019
[27] Qi C R, Yi L, Su H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space//Proceedings of the Advances in Neural Information Processing Systems. Long Beach, USA, 2017: 5099-5108
[28] Vo A V, Truong-Hong L, Laefer D F, et al. Octree-based region growing for point cloud segmentation. ISPRS Journal of Photogrammetry and Remote Sensing, 2015, 104: 88-100
[29] Golovinskiy A, Funkhouser T. Min-cut based segmentation of point clouds//Proceedings of the IEEE International Conference on Computer Vision Workshops, ICCV Workshops. Kyoto, Japan, 2009: 39-46
[30] Yang Y, Feng C, Shen Y, et al. Foldingnet: Point cloud auto-encoder via deep grid deformation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 206-215
[31] Fan H, Su H, Guibas L J. A point set generation network for 3d object reconstruction from a single image//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2017: 605-613
[32] Bertsekas D P. A distributed asynchronous relaxation algorithm for the assignment problem//Proceedings of the 24th IEEE Conference on Decision and Control. Fort Lauderdale, USA, 1985: 1703-1704
[33] Yuan W, Khot T, Held D, et al. PCN: Point completion network//Proceedings of the International Conference on 3D Vision (3DV). Verona, Italy, 2018: 728-737

MIAO Yong-Wei, Ph.D., professor, Ph.D. supervisor. His current research interests include computer graphics, digital geometry processing, computer vision, machine learning.

LIU Jia-Zong, M.S. His current research interests include computer graphics, computer vision.

SUN Yu-Liang, Ph.D. candidate. His current research interests include computer graphics, computer vision.

WU Xiang-Yang, Ph.D., associate professor. His current research interests include computer graphics, computer vision, visualization.

Background

Three-dimensional point cloud data is one of the most important and popular data forms which can be obtained easily from the real world. However, the point cloud data acquired by 3d scanning devices or depth cameras will always be missing data or low quality due to the object occlusions in the complex scenes, the limited sensor distance of the depth camera, and the scanning errors of the scanning devices, etc. For improving the input scanning point cloud data, the point cloud completion of 3d indoor scenes or 3d shapes is an important issue in the literature of Computer Graphics, Digital Geometry Processing and 3D Computer Vision.

In generally speaking, there are two types of 3d indoor scene completion, one is repairing the missing plane structures, the second is completing the lack of interior furniture shape structures. In this paper, we focus on the latter issue and proposed a novel framework that can repair the missing point cloud data of 3d furniture shapes in indoor scenes and also complete the underlying 3d indoor scenes. The framework consists of point cloud scene segmentation module and point cloud completion module. Our proposed method is validated on repairing and completing both the 3d individual shapes and also 3d indoor scenes.

Prof. Yongwei Miao is a Qianjiang Special Expert of Hangzhou city at Hangzhou Normal University. This research is supported by the National Natural Science Foundation of China under grant Nos. 61972458, 61972122.