

基于扩散 Transformer 网络的激光雷达数据补全方法

李伟松, 刘佳, 张坤*

(河北科技大学 信息科学与工程学院, 石家庄 050018)

摘要: 由于设备故障和环境干扰等因素, 激光雷达在数据采集过程中常常存在数据缺失或噪声干扰的问题, 这些问题严重影响后续数据的解析和应用。为了解决这一难题, 引入了扩散 Transformer 网络(DT-Net), 将 DT-Net 用作生成器, 与自注意力单元判别器相结合。此外, 还设计了一种扩散机制用于激光雷达数据补全。实验结果表明: 相较于 PoinTr 方法, 所提出的方法在激光雷达数据补全任务方面取得了显著的改进, 平均 Chamfer 距离(CD)值降低了约 1.79%, F-Score 值提升了约 1.88%。

关键词: 扩散机制; 数据补全; 激光雷达; 实际应用

中图分类号: TN929.12 文献标志码: A 文章编号: 1002-5561(2024)03-0052-05

DOI: 10.13921/j.cnki.issn1002-5561.2024.03.010

开放科学(资源服务)标识码(OSID):



Lidar data completion method based on diffusion Transformer network

LI Weisong, LIU Jia, ZHANG Kun*

(Hebei University of Science and Technology School of Information Science and Engineering, Shijiazhuang 050018, China)

Abstract: Due to factors such as equipment failure and environmental interference, lidar often encounters problems of missing data or noise interference during the data collection process, which seriously affect the subsequent analysis and application of the data. To solve this problem, the diffusion Transformer network (DT-Net) is introduced and used as a generator in combination with a Self-Attention unit discriminator. Additionally, a diffusion mechanism is designed for lidar data completion. The experimental results show that compared to the PoinTr method, the proposed approach achieves significant improvements in lidar data completion tasks, with an average Chamfer distance(CD) value reduced by approximately 1.79% and an F-Score value increased by approximately 1.88%.

Key words: diffusion mechanism, data completion, laser radar, practical application

0 引言

激光雷达具有高分辨率的距离和速度测量能力^[1], 被广泛应用于城市规划、考古文物保护、自动驾驶、医疗成像等诸多领域。激光雷达通过发射和接收激光束, 收集目标对象表面点的三维坐标、反射率等信息,

收稿日期: 2023-11-01。

基金项目: 河北省高等学校科学技术研究项目(No.ZD2020176)资助; 河北省技术创新引导计划项目(No.21470302D)资助。

作者简介: 李伟松(1997—), 男, 硕士研究生, 现就读于河北科技大学信息科学与工程学院计算机技术专业, 研究方向为计算机视觉和三维数据处理, 主要从事激光雷达数据采集、机器视觉和补全算法研究工作。

* 通信作者: 张坤(1982—), 女, 博士研究生, 主要研究方向为机器视觉、人工智能、三维图形学。



进而快速建立目标对象的相关数据模型。然而, 在激光雷达数据采集过程中, 传感器会受到分辨率、单视点、遮挡、反射和回波等因素的影响, 这些因素会导致部分点云数据的丢失, 给后续的数据分析和处理带来了困难, 从而限制了点云数据的有效应用。因此, 如何进行激光雷达数据采集补全成为了一个重要的研究方向。

早期的一些学者^[2-5]试图通过最小二乘法将成熟的方法从二维补全任务拓展到三维点云的采集。然而, 这种方法由于存储和计算复杂度高, 以及空间不一致性, 导致结果不平滑, 尤其在语义信息的建模方面显得不足。点补全网络(PCN)^[6]是首个基于学习的点云补全架构, 专注于局部补全任务, 采用了自编码器

结构和多尺度处理来提高补全的效果和精度。例如: FoldingNet 利用共享连接层和 2 组 128×128 的二维栅格对输入点云进行补全^[7]; TopNet 引入了一个分层树形结构的解码器, 用于生成对象的形状, 显著改进了激光雷达数据采集补全任务^[8]。然而, 这些网络采用多层感知器 (MLP) 直接从不完整的点云中提取潜在特征, 容易导致局部细节的丢失。GRNet 通过独立的网格化和去网格化层, 有效地将点云转换成三维网格, 同时保留了内在的结构信息^[9]。但 GRNet 仅处理全分辨率的输入数据。

根据三维点云不规则区域和缺乏有序性等特性, 2021 年, GUO M H 等人^[10]设计了一种基于 Transformer 网络的用于点云学习的新框架——点云转换器 (PCT)。该转换器支持最远点采样 (FPS) 和最近邻域搜索, 能更好地捕获点云的局部上下文。同年, YU X 等人^[11]提出了一种将点云补全重新定义到集的转换方法和用于点云补全的变压器编/解码器的架构 PoinTr。该方法通过无序点组的位置嵌入来表示点云, 并利用变压器模型进行点云生成; 同时设计了一个几何感知模块, 用于显式建模局部几何关系。然而, 该方法产生参数数量过大, 并且在局部补全方面的效果有待提高。

基于此, 本文提出一种基于扩散 Transformer 网络 (DT-Net) 的激光雷达数据补全方法, 结合残差网络和自注意单元判别器网络的多尺度特征点云补全方法, 以增加系统几何感知能力。

1 基于 DT-Net 的激光雷达数据补全方法

本文提出的基于 DT-Net 的激光雷达数据补全方法模型如图 1 所示。数据补全原理如下: 首先, 将残缺的点云数据输入到 DT-Net 模型的不完整点云数据

中, 分别采用 FPS 方法、特征提取器获取中心点及中心点的周围特征, 通过残差多层感知器 (ResMLP), 使得网络更好地理解点云中的结构并更准确地估计缺失部分的形状和特征, 进行更深层次的学习; 然后, 将位置信息嵌入特征部分信息中, 并将点的代理输入到 DT-Net 中, 通过扩散几何变压编码器和扩散几何变压解码器来预测缺失部分的点代理, 而查询生成器引入一些自适应的、动态的查询信息; 最后, 自注意判别器以由粗到细的方式来补全预测点代理的点云, 并将残缺的点云数据部分补全。

1.1 ResMLP

传统的 MLP 在处理激光雷达采集数据时可能受限于模块化和特征表示能力。为了有效地提取输入点云特征, 本文引入了 ResMLP^[12], 其整体架构如图 2 所示。ResMLP 原理如下: 首先, 将 $N \times N$ 的非重叠补丁的网格作为输入 (补丁大小通常为 16×16), 这些补丁独立地通过一个线性层形成一组 N^2 个 d 维输入; 然后, 将这些 d 维数据依次通过交叉补丁子层和跨通道子层, 每个子层都与跳跃连接跨层连接 (跳跃连接是残差连接的思想, 直接将输入添加到网络层的输出中, 使得训练更加稳定和高效, 缓解梯度消失问题), 每个子层之后都有 1 个残差连接和 2 个仿射元素进行逐元素变换。与传统的 MLP 相比, ResMLP 通过 MLP 和残差连接的结合来处理点云数据, 具有更好的建模能力和表达能力, 能够更好地处理点云数据的复杂特征和关系, 有效地捕捉局部和全局的上下文信息。图中的高斯误差线性单位 (GeLU) 是一种激活函数, T 表示转置。

1.2 扩散几何变压编码器

将 Transformer 网络应用于视觉任务时, 一个主要

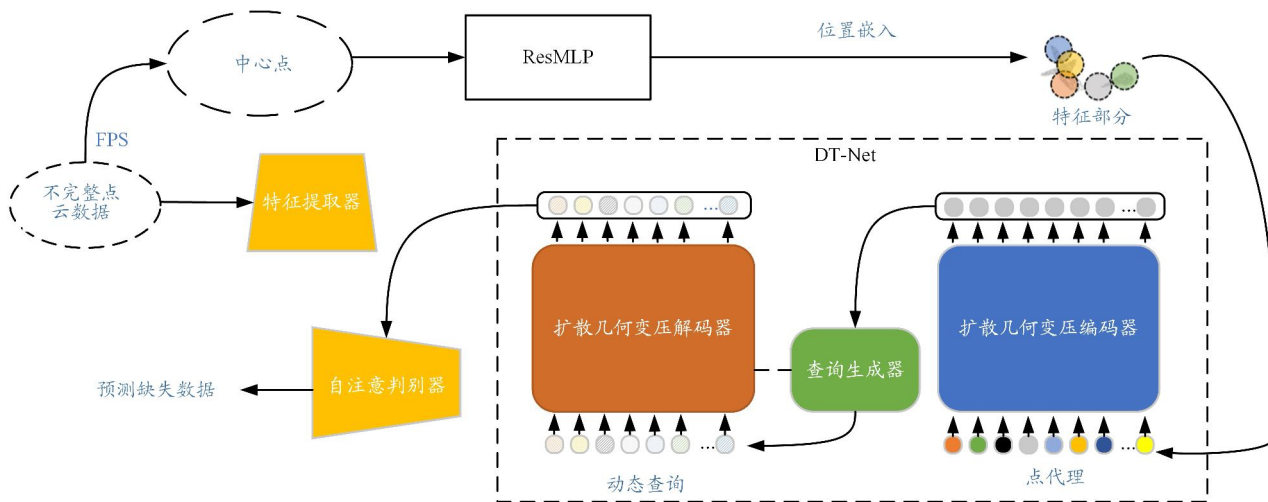


图 1 基于 DT-Net 的激光雷达数据补全方法模型

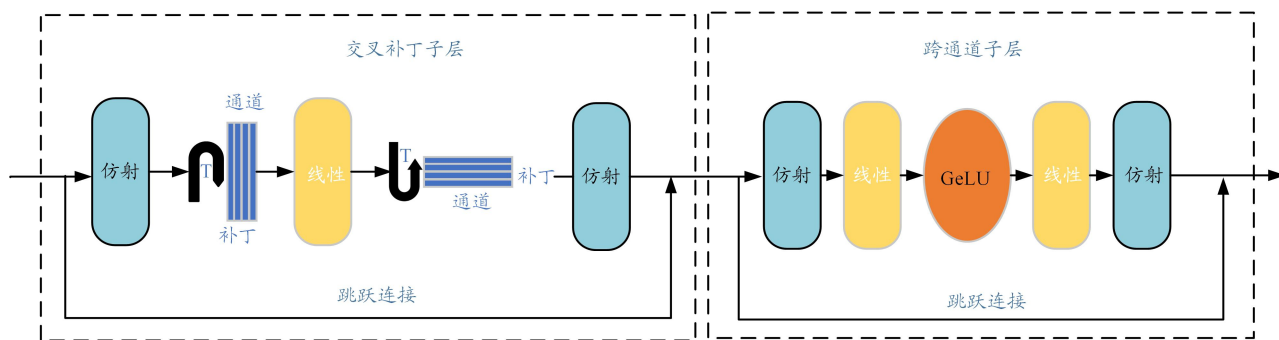


图2 ResMLP 的整体架构

挑战是缺乏传统视觉模型(例如卷积神经网络和点云网络)中固有的归纳偏置。传统视觉模型通过明确建模视觉数据的结构,比如在图像中的空间局部性或者点云中的几何结构,从而实现较好的补全效果。然而,Transformer 网络中的自注意机制并没有为这些结构提供特定的偏置。因此,为了让 Transformer 网络更好地利用点云的几何结构的归纳偏置,本文引入了可插拔的几何感知模块,可以与任何 Transformer 网络中的注意力模块结合使用。这使得 Transformer 网络能够同时捕捉点云数据的几何和语义关系。

本文采用了 K 最近邻(KNN)模型方法来捕捉点云中的几何关系。KNN 模型方法结合了近似算法和索引结构,在保证一定准确性的前提下提高 KNN 算法的效率。具体操作如下:首先,使用关键坐标来查询最近的关键点,并获取它们的特征;然后,遵循动态图卷积神经网络(DGCNN)^[13]的方法,通过线性层的特征聚合和最大池化操作来学习局部几何结构;最后,将几何特征和语义特征串联在一起,并映射回原始维度,形成最终的完整补全结果。

1.3 扩散几何变解压码器

扩散几何变解压码器需要对输入的信息进行处理和推断,然后产生相应的补全结果输出。然而,输入信息中通常包含了一定的噪音和缺失部分。将扩散模型作为一个额外的模块添加到现有的架构中,利用输

入信息的临近关系、相似性或相关性来进行信息的传播和扩散,这样解码器将不仅利用几何感知 Transformer 网络生成初步的补全结果,还会通过扩散模型理解输入信息的结构、语义和内在联系,进一步提高推断和补全的准确性。

1.4 自注意判别器

为了在点云生成任务中生成不同分辨率或不同密度的点云数据,本文提出了一个多尺度点云生成框架,即将 DT-Net 预测丢失部分的点代理作为生成器,与自注意判别器相结合来补全点云。自注意判别器原理框图如图 3 所示。其原理如下:首先,点云数据通过 MLP 映射为一个 $r_n \times C_d$ 的矩阵(r_n 表示行数, C_d 是 MLP 输出的特征维度),接着通过最大池化,对点云中的每个特征维度提取最显著的信息,从而减少维度并保留重要特征,输出的矩阵依然是 $r_n \times C_d$ (输出 $1 \times C_d$, 并重复 r_n 次),将 $r_n \times C_d$ 与 $r_n \times 2C_d$ 连接到一起,融合不同特征信息,提供多样性信息并增强判别性;然后,为了改进特征学习,在特征拼接后添加了一个自注意单元,有助于增强特征整合并提高后续特征提取的能力;最后,应用一组 MLP 和最大池化来生成全局特征,并通过全连接层进一步回归最后的置信度值。如果这个置信度值接近 1, 判别器则预测输入很可能来自具有高置信度的目标分布,否则来自生成器。

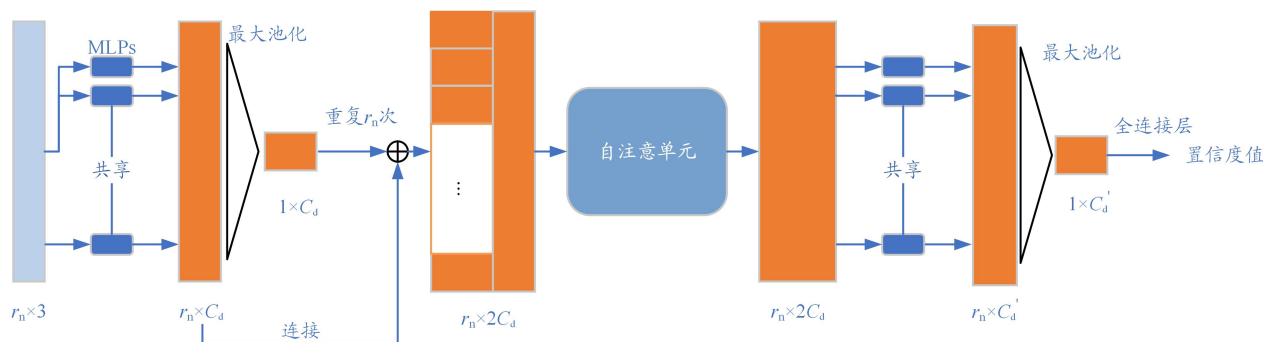


图3 自注意判别器原理框图

2 实验评估与分析

实验在 Linux 系统环境下进行。计算机 GPU 为 NVIDIA GeForce RTX 3080 Ti, 内存为 12 GB; 软件环境为 Python 3.7、cuda 11.3 和 PyTorch 1.13.1。

2.1 数据集及预处理

实验中采用的 PCN 数据集是基于 ShapeNet 中人工搭建的计算机辅助设计(CAD)模型收集的, 包含了飞机、橱柜、汽车、椅子、灯具、沙发、桌子、船舶共 8 个类别。其中, 每个完整的形状包含 16 384 个点, 因此本文首先将点数大于 2 048 的形状下采样为 2 048, 点数小于 2 048 的形状上采样为 2 048。然后, 继续在具有 2 048 个点的稀疏形状上训练本文的网络, 并通过在 PCN 数据集上预测具有 16 384 个点的密集完整形状来展示其泛化能力。实验借助自适应矩估计优化器训练三维激光雷达数据补全网络, 神经网络的学习率设置为 0.000 1, 权重衰减设置为 0.000 5, batch_size 设置为 32, epoch 设置为 600。

在评估过程中, 本文固定了 8 个视点, 为了方便起见, 设置点数为 2 048 个。

2.2 评估指标

在激光雷达数据补全任务中, 由于点云的无序性, 许多传统的损失函数并不适用。因此, 需要一种用于定量测量输出点云质量的方法。为了解决这个问题, 本文使用 Chamfer 距离(CD)^[14]和 F-Score 作为评估指标, 计算网络补全后的点云 S_1 与真实点云 S_2 之间的误差。其中, CD 是生成点云与真实点云的平均最近平方距离, 其值越小, 表示用于激光雷达数据补全的神经网络效果越好; F-Score 是用来评价模型补全质量的指标, 它基于精确度、召回率这两个指标的调和平均数计算得出。F-Score 越高, 表明模型补全的质量越好, 激光雷达数据补全的准确率也相应越高。

2.3 实验结果分析

本文所提方法与 5 种激光雷达数据补全方法(FoldingNet、PCN、TopNet、GRNet、PoinTr)在 PCN 数据集上的实验对比结果(蓝色字体为最优解)如表 1 所示。可以看出, 本文所提方法除了在灯具类和沙发类的 CD 值略高于 PoinTr、GRNet 外, 其它各类别的 CD 值均优于对比方法, 并且平均 CD 值相比最优的 PoinTr 方法还降低了约 1.79%; 与 PoinTr 方法相比, 本文所提方法的 F-Score 值提升了约 1.88%。这种优势是由以下几种原因导致的:

1) 本文通过特征点采样方法将输入转换为多分辨率, 并且采用 ResMLP 提取全局特征, 以此保证残缺点云的特征全面性与完整性。此外, 利用全连接层生成粗糙点云, 以更好地预测输出点云的全局几何结构。因此, 本文方法在输入不完整点云的情况下, 能够有效地补全激光雷达采集的数据, 提高网络的泛化性。

2) 本文将传统的单分辨率特征提取方式优化为多特征提取方式, 这样可以捕获丰富的信息, 帮助模型更全面地理解点云的结构和属性。同时, 不同的特征通道可以捕获不同层次和角度的信息, 可以更好地理解点云中的局部细节和全局结构, 为后续的编码和恢复完整点云提供了很好的帮助。

3) 编码器中添加了扩散模型, 利用输入信息的临近关系进行信息的传播和扩散, 进一步提高补全的准确性, 使得网络补全性能达到最优。

本文使用觅道 Mid-40/Mid-100 激光雷达, 激光波长为 905 nm, 量程(@100 klx)分别为 90 m@10%反射率、130 m@20%反射率、260 m@80%反射率; 视场(FOV)分别为 38.4°圆形(Mid-40)、98.4°(水平)×38.4°(竖直)(Mid-100); 距离精度(1σ @20m)为 2 cm; 角度精度小于 0.1°, 虚警率(@100 klx³)小于 0.01%; 光束

表 1 PCN 数据集上不同网络的点云补全对比

数据补全 方法	各类 CD/a.u.								平均 CD/a.u.	F-Score/%
	飞机	橱柜	汽车	椅子	灯具	沙发	桌子	船舶		
FoldingNet	9.49	15.80	12.61	15.55	16.41	15.97	13.65	14.99	14.31	0.322
PCN	5.50	22.70	10.63	8.70	11.00	11.34	11.68	8.59	9.64	0.695
TopNet	7.61	13.31	10.90	13.82	14.44	14.78	11.22	11.12	12.15	—
GRNet	6.45	10.37	9.45	9.41	7.96	10.51	8.44	8.04	8.83	0.708
PoinTr	4.75	10.47	8.68	9.39	7.75	10.93	7.78	7.29	8.38	0.745
本文方法	4.69	10.06	8.24	9.21	7.93	11.14	7.48	7.11	8.23	0.759

李伟松, 刘佳, 张坤: 基于扩散 Transformer 网络的激光雷达数据补全方法

发散度为 0.28° (垂直) $\times 0.03^\circ$ (水平); 数据率分别为 100 000 points/s (Mid-40)、300 000 points/s (Mid-100); 视场覆盖率分别为 $20\% @ 0.1 \text{ s}^2$ 、 $68\% @ 0.5 \text{ s}^2$ 、 $93\% @ 1 \text{ s}^2$ 。本文对椅子进行了数据采集, 并将采集到的点云输入到 DT-Net 模型中, 可以看到椅子在全局效果较好, 但是在局部细节不清晰或缺失, 后续会对细节补全部分做进一步优化。

3 结束语

本文提出了一种基于 DT-Net 的激光雷达数据补全方法, 将扩散模型与 Transformer 网络的编码器融合, 通过点与点之间的相互作用来捕获全局依赖关系, 同时在点云中传播信息以帮助识别局部之间的关联性, 既增强了全局特征信息的提取, 又有助于更好地理解局部特征信息。然后, 将 Transformer 网络作为生成器与自注意判别器融合, 有助于提高生成点云的质量, 同时自注意判别器使得局部信息进一步增强。实验结果表明: 本文所提方法除了在灯具类和沙发类的 CD 略高于 PoinTr、GRNet 外, 其它各类别的 CD 值均优于对比方法, 并且平均 CD 值相比最优的 PoinTr 方法还降低了约 1.79%; 与 PoinTr 方法相比, 本文所提方法的 F-Score 值提升了约 1.88%。

本文的方法不仅保证了点云的全局结构, 并且还对局部细节信息有更好的把控, 体现出了良好的泛化性和鲁棒性。这项研究对于提升激光雷达感知系统在自动驾驶、机器人导航等领域的应用具有重要意义。

参考文献:

- [1] 丘焕然, 刘偲嘉, 甘育娇, 等. 半导体脉冲激光器发展综述[J]. 光通信技术, 2021, 45(10): 1-6.
- [2] SCHIEBENER D, SCHMIDT A, VAHRENKAMP N, et al. Heuristic 3D object shape completion based on symmetry and scene context [C] //IEEE. Proceedings of 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Daejeon: IEEE, 2016: 74-81.
- [3] FIGUEIREDO R, MORENO P, BERNARDINO A. Automatic object shape completion from 3d point clouds for object manipulation[C]//IEEE. Proceedings of International Joint Conference on Computer Vision Theory and Applications. Porto: IEEE, 2017: 565-570.
- [4] LIU Z, TANG H, LIN Y, et al. Point-voxel cnn for efficient 3d deep learning[C]//IEEE. Proceedings of Thirty-third Conference on Neural Information Processing Systems 2019. Vancouver: IEEE, 2019: 965-975.
- [5] SHARMA A, GRAU O, FRITZ M. Vconv-dae: deep volumetric shape learning without object labels[C]//ECCV Workshops. Proceedings of European Conference on Computer Vision 2016. Amsterdam: ECCV Workshops, 2016: 236-250.
- [6] YUAN W, KHOT T, HELD D, et al. Pcn: point completion network[C] //IEEE. Proceedings of International Conference on 3D Vision (2018). Verona: IEEE, 2018: 728-737.
- [7] YANG Y, FENG C, SHEN Y, et al. FoldingNet: interpretable unsupervised learning on 3d point clouds [C]//IEEE. Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 1-14.
- [8] TCHAPMI L P, KOSARAJU V, REZATOFIGHI H, et al. TopNet: structural point cloud decoder [C]//IEEE. Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019: 383-392.
- [9] XIE H, YAO H, ZHOU S, et al. Grnet: gridding residual network for dense point cloud completion[C]//ECCV Workshops. Proceedings of European Conference on Computer Vision 2020. Glasgow, ECCV Workshops, 2020: 365-381.
- [10] GUO M H, CAI J X, LIU Z N, et al. PCT: point cloud transformer[J]. Computational Visual Media, 2021, 7(2): 187-199.
- [11] YU X, RAO Y, WANG Z, et al. PoinTr: diverse point cloud completion with geometry-aware transformers [C]//IEEE. Proceedings of CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021: 12478-12487.
- [12] TOUVRON H, BOJANOWSKI P, CARON M, et al. Resmlp: feedforward networks for image classification with data-efficient training [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(4): 5314-5321.
- [13] ZHANG K, HAO M, WANG J, et al. Linked dynamic graph cnn: Learning on point cloud via linking hierarchical features [C]//IEEE. Proceedings of International Conference on Mechatronics and Machine Vision in Practice (M2VIP). Shanghai: IEEE, 2021: 7-12.
- [14] FAN H, SU H, GUIBAS L J. A point set generation network for 3d object reconstruction from a single image [C]//IEEE. Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 2463-2471.
- [15] QI C R, YI L, SU H, et al. Pointnet++: deep hierarchical feature learning on point sets in a metric space [C]//IEEE. Proceedings of 31st international conference on neural information processing systems. Long Beach: IEEE, 2017: 5105-5114.