

# 基于掩码对比学习的序列点云模型局部特征增强方法

王晓荣, 黄华

(北京交通大学计算机科学与技术学院)

**摘要:** 近年来, 基于序列的点云模型在点云处理领域逐渐展现出巨大的潜力。相比于传统的基于 kNN 的邻域特征聚合方法, 序列点云模型通过将点云数据映射为空间上邻近的点云序列, 在提升计算效率的同时, 还扩展了感受野。然而, 感受野的扩大通常伴随着对局部精细特征捕捉能力的下降, 导致模型在学习局部细节信息方面存在不足。为此, 本文提出了一种基于掩码对比学习的序列点云模型局部特征增强方法, 以解决序列点云模型对局部信息学习能力不足的问题。具体而言, 本方法设计了一种融合掩码对比学习与有监督点云分割的模型, 通过引入对比学习深入挖掘点云数据中的局部特征。为实现自监督对比学习与有监督分割任务之间的有效协同, 模型采用了一种正则化策略, 在保持特征一致性的同时平衡两种任务的学习目标。在 ShapeNet Part 和 S3DIS 数据集上的实验结果表明, 本方法有效的改进了模型在复杂场景的的分割性能。在多个评价指标上, 本方法均提升了现有模型, 充分验证了所提出方法的有效性。

**关键词:** 计算机视觉; 点云分割; 对比学习; 局部特征增强

**中图分类号:** TP391

## Contrastive Learning of Local Features for Sequential Point Cloud Models

wang xiao rong, Huang hua

(School of Computer Science & Technology, Beijing Jiaotong University)

**Abstract:** In recent years, sequential point cloud models have demonstrated substantial potential in the field of point cloud processing. Compared with traditional KNN-based neighborhood feature aggregation methods, these sequential models enhance computational efficiency while expanding the receptive field by mapping point cloud data into spatially adjacent sequences. However, the expansion of the receptive field typically compromises the model's ability to capture fine-grained local features, resulting in insufficient learning of detailed local information. To address this limitation, this paper proposes a masked contrastive learning-based approach for enhancing local features in sequential point cloud models. Specifically, we design a unified framework integrating masked contrastive learning with supervised point cloud segmentation, where contrastive learning tasks are employed to intensively explore local characteristics within point cloud data. To achieve effective collaboration between self-supervised contrastive learning and supervised segmentation tasks, the model incorporates a regularization strategy that maintains feature consistency while balancing the learning objectives of both tasks. Experimental evaluations on the ShapeNet Part and S3DIS datasets demonstrate that the proposed method significantly improves segmentation performance in complex scenarios. Across multiple evaluation metrics, our approach outperforms existing models, conclusively validating the effectiveness of the proposed methodology.

**Key words:** Computer Vision; Point Cloud Segmentation; Contrastive Learning; Local Feature Enhancement

**作者简介:** 王晓荣, 男, 硕士研究生, 3D 点云

**通信联系人:** 黄华, 男, 副教授, 硕导, 计算机视觉、机器学习. E-mail: hhua@bjtu.edu.cn

## 0 引言

45 随着 3D 传感技术的飞速发展, 3D 点云分割<sup>[1-3]</sup>已成为计算机视觉、机器人技术以及多个领域的核心技术之一。从自动驾驶车辆的环境感知, 到工业机器人在物体识别中的应用, 再到虚拟现实复杂场景的理解, 3D 点云分割的应用场景愈加广泛。如何高效、精准地对点云数据进行分割, 已成为学术界和工业界的共同关注热点。

传统的点云处理方法在处理复杂几何形状和表面细节时, 通常依赖最远点采样和 K-近邻算法来构建局部邻域。然而, 在大规模点云数据中, 这些方法往往导致较高的计算时间复杂度。此外, 现有的点云分割模型通过注意力机制来提升分割性能, 进一步增加了计算开销。近期一些研究<sup>[4-5]</sup>观察到, 注意力机制对点云局部窗口的实际形状并不敏感。因此, 新的研究趋势逐步转向将点云数据进行序列化处理。这些方法通过某种编码形式将点云数据转化为一维序列, 从而在保持点云空间局部性的同时, 显著扩大了模型的感受野。这种  
55 处理方式使得模型能够更加灵活地处理不同尺度的信息, 并通过序列化特性提升计算效率。然而, 现有的序列化方法仍面临着一个重要挑战, 即如何在扩大感受野的同时, 保持对精细局部特征的捕捉能力。

针对这一挑战, 本文提出了一种简单而有效的解决方案——通过掩码对比学习增强序列点云模型的局部特征学习能力。本方法的目标是在不牺牲模型感受野的情况下, 提高对局部细节的敏感度, 从而在点云分割任务中获得更好的性能。具体而言, 在训练阶段, 通过引入一个辅助任务, 来增强模型对局部特征的学习能力。借鉴掩码自编码器<sup>[6]</sup> (Masked Autoencoder, MAE) 的思想, 在训练过程中随机选择一部分点, 并基于这些点的邻域点构建掩码邻域。模型首先对掩码后的点云进行编码, 将其映射到高维的语义特征空间。为了进一步提升局部特征学习的效果, 本方法将对对比学习<sup>[7]</sup>框架融入到模型训练中。掩码前后的点云视作同一对象的不同视图, 辅助任务的目标是对齐掩码前后的点云视图特征。通过  
65 最小化掩码前后特征的差异, 模型能够更加专注于学习对局部细节特征。最后, 为避免模型在训练前期陷入分割精度低的困境, 本方法设计了一种正则化策略。该策略通过在损失函数中平衡对比学习与主任务 (即点云分割任务) 之间的权重, 使得模型在训练过程中既能够学习全局信息, 又能深入挖掘局部细节。在训练初期, 赋予掩码对比学习任务较大的  
70 权重, 以促进模型对局部特征的学习; 随着训练的进展, 逐步增加点云分割任务的权重, 以确保两者之间的平衡。在推理阶段, 模型不再执行掩码自编码器和对比学习操作, 而是根据输入的点云数据直接预测每个点的分割标签。通过在训练过程中增强局部特征的学习, 模型在推理时能够更准确地识别和分割不同的物体或部分, 从而提高点云分割的精度和效率。

75 本文提出了一种局部掩码对比学习(Local Mask Contrastive Learning, LMCL)框架, 用于增强模型对局部特征的分割能力, 如图 1 所示。在多个标准数据集上进行了广泛的实验验证, 结果表明, LMCL 不仅显著提升了模型的分割精度, 还在处理细节丰富的数据时表现

出了更强的鲁棒性。本文的主要贡献包括：

80 (1) 提出了一种新的局部特征学习框架，结合掩码自编码器与对比学习的思想，增强了模型在捕捉点云局部细节方面的能力。在不牺牲感受野大小的前提下，显著提升了模型对局部特征的理解，使其在处理复杂几何结构时具备更强的优势。

(2) 提出了一种正则化方法，旨在动态调整对比学习与点云分割任务之间损失函数的权重分配，有效避免了模型在训练初期过度倾向于点云分割任务的学习，确保了局部特征能够得到充分建模，从而提高了模型在精细分割任务中的表现。

85 (3) 通过在 ShapeNet Part<sup>[8]</sup> 和 S3DIS<sup>[9]</sup> 数据集上的实验，验证了本方法的有效性。实验结果表明，LMCL 显著提升了现有序列点云模型的分割效果，尤其在复杂场景下展示了较强的泛化能力和鲁棒性。

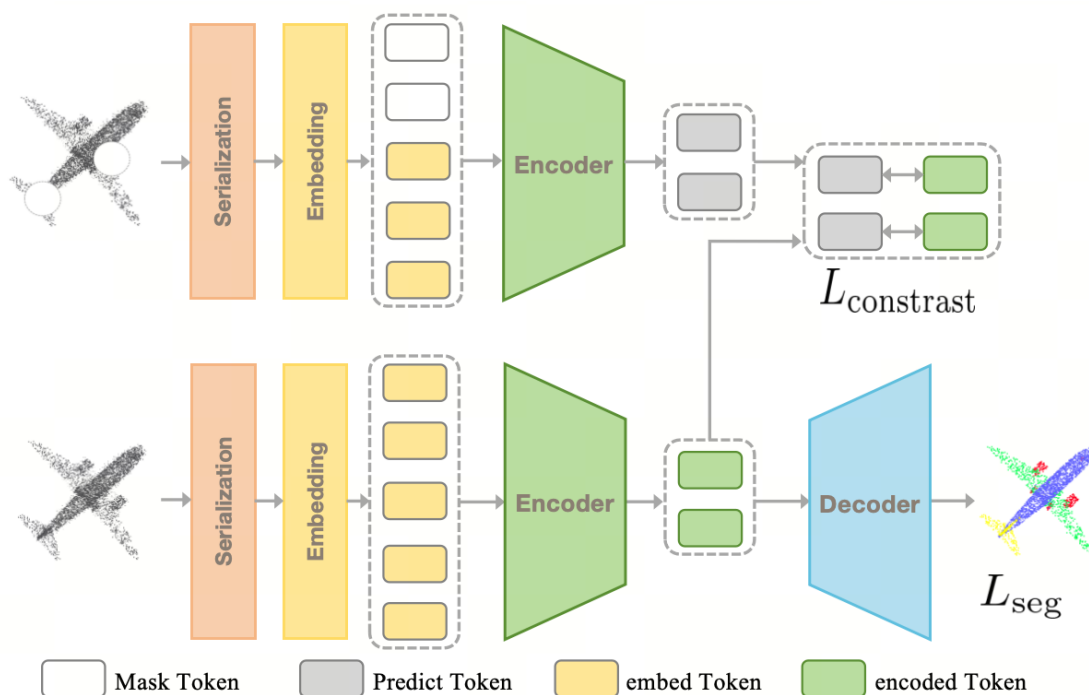


图 1 LMCL 框架示意图

Fig. 1 LMCL Framework Schematic Diagram

90

## 1 相关工作

本节主要介绍点云分割的骨干模型、序列点云模型以及自监督的研究现状。

### 1.1 点云分割方法

95 当前的点云分割模型通常依据其骨干网络的不同，主要可以分为四类：基于多层感知机（MLP-based）、基于卷积（Convolution-based）、基于图（Graph-based）以及基于 Transformer 的方法。早期基于多层感知机的方法（如 PointNet 系列<sup>[10-13]</sup>）直接处理点云坐标实现基础特征提取，其结构简洁性和计算效率推动了领域发展，但受限于局部特征建模能力，在复杂场景的细粒度识别中存在明显局限性。为增强局部特征提取，同时期一些研究者将卷积运算引入三维空间，开发出 KPConv<sup>[14,17]</sup>等序列方法，这类工作通过三维卷积核聚合邻域信息，但

100 固定卷积核带来的归纳偏差限制了其对非结构化点云的表征能力。针对点云的非规则特性，  
基于图神经网络的方法（如 DGCNN 等<sup>[18,20]</sup>）通过构建图结构有效建模局部拓扑关系，但图  
结构的动态计算显著增加了模型复杂度。近年来，基于 Transformer<sup>[21]</sup>架构的方法（如 Point  
Transformer 等<sup>[22-28]</sup>）利用注意力机制在特征表示学习方面取得突破，在多个基准数据集达  
105 到最先进性能，但其依赖 KNN 构建局部邻域的策略导致感受野受限，难以有效捕获全局上  
下文信息。

尽管上述方法在局部特征建模和上下文信息感知方面各有优势，但它们在这两者之间仍  
然面临不可避免的权衡。为此，本文采用序列化机制，在保持点云局部特征建模能力的同时，  
有效避免了传统点云局部邻域方法所带来的较高计算复杂度。

## 1.2 序列点云分割模型

110 近期的点云分析研究引入了序列化技术，通过某种映射规则将非结构化的三维点云转换  
为结构化的一维序列，在保持空间局部性的同时提高了处理效率。多个方法在此方向上取得  
了进展，平衡了分割精度与计算性能。HiBertNet<sup>[29]</sup>提出了希尔伯特注意力机制，替代传统的  
3D 卷积，用注意力机制来处理点云中的网格数据，不仅降低了计算成本，还有效地近似了  
3D 卷积的表现。OctFormer<sup>[5]</sup>核心在于八叉树注意力机制，利用八叉树结构将点云划分为局  
115 部窗口，确保每个窗口中的点数相等，且窗口形状可灵活调整，从而提高点云分析的效率与  
精度。Point Transformer V3<sup>[4]</sup>通过混合序列化策略结合序列窗口注意力机制，解决了点云处  
理中准确性和效率的权衡问题。其适应性强的机制使其能够有效地聚焦于不同区域，提高整  
体性能。Mamba 系列模型<sup>[30,34]</sup>则通过定义序列化策略并在此基础上引入 Mamba 架构，提升  
了点云全局建模能力。这些模型在序列点云表示的基础上，增强了网络对局部和全局特征的  
120 捕捉能力。

然而，值得注意的是，尽管这些序列化点云模型在计算效率上表现优异，它们不可避免  
地破坏了点云数据的原始空间结构，导致局部特征的捕捉精度有所下降。如何在提高计算效  
率的同时，最大程度地保留点云的空间结构信息，是当前点云分析领域面临的一个关键挑战。

## 1.3 点云的自监督学习

125 自监督学习（Self-Supervised Learning, SSL）作为一种机器学习范式，一直受到学术界的  
广泛关注。与传统的有监督学习不同，自监督学习不依赖于显式的标签数据，而是通过设计  
一系列预训练任务来自动生成伪标签，进而挖掘数据内部的结构和规律，实现特征表示学  
习。在 3D 点云领域，自监督学习的方法取得了显著进展。PointContrast<sup>[35]</sup> 利用对比学习，  
从全局场景中学习特征表示，验证了预训练模型在不同任务中的通用性。PointBert<sup>[36]</sup> 借鉴  
130 自然语言处理的掩码语言建模策略，提出掩码点建模（MPM）。预训练目标是通过掩码点来  
学习高层次语义信息，有效提升了 Transformer 在 3D 点云任务中的性能。类似地，  
PointMAE<sup>[37]</sup> 采用掩码自监督策略，训练模型重建被掩码的点块。相比 PointBert, PointMAE  
简化了任务设计，专注于重建任务，展现了直观且有效的性能。MaskFeat3D<sup>[38]</sup>进一步发展了  
这一自监督学习的方向，其核心不再仅仅是预测掩码点的位置，而是专注于预测这些点的几  
135 何特征，如表面法线和曲率变化。通过强调局部几何属性，MaskFeat3D 在预训练阶段提升  
了模型对细粒度局部特征的学习，进而在需要深入理解点云几何信息的下游任务中，取得  
了显著的性能提升。CBL<sup>[39]</sup>（对比边界学习）提出对比边界学习框架，专注于优化边界特征表  
示。通过在多尺度场景中对比边界点，CBL 增强了边界特征的区分度，在边界检测和分割

任务中表现优异。

140 这些工作分别从不同的角度出发,验证了自监督学习,在 3D 点云任务中的有效性和泛  
化能力。本文旨在通过对比学习,作为局部特征学习与点云分割之间的桥梁,进一步提升点  
云分割的效果与精度。

## 2 方法

145 现有的序列化点云模型在一定程度上削弱了对局部特征的精细捕捉能力,为了解决这一  
问题,本方法首先提出了一种掩码对比学习的策略,以增强模型对细粒度局部特征的学习和  
区分能力,同时保持其原有的推理效率。

### 2.1 点云序列化

150 现有的序列点云模型通常采用空间填充曲线,对三维空间的点云进行序列化处理,并同  
时在序列中保持三维空间的局部性。常见的空间填充曲线有希尔伯特空间填充曲线、Z-order  
空间填充曲线。

155 希尔伯特空间填充曲线 (Hilbert Space-Filling Curve) 是一种递归定义的分形曲线,能够  
在多维空间(如二维、三维等)中遍历每一个网格,并将这些点映射到一维空间。它可以  
将一个多维空间(例如 2D 或 3D 空间)中的点按照特定的组织顺序映射到一维序列,而不会  
遗漏任何点。希尔伯特曲线具有较好的空间局部性 (Spatial Locality),即在多维空间中相  
邻的点在经过曲线映射到一维空间时仍然尽量保持相邻。

Z-order 空间填充曲线 (Z-order Curve),是一种将多维空间中的点映射到一维空间的空  
间填充曲线。它通过交替取各维度的比特位来排列多维坐标,从而生成一个一维序列。与  
希尔伯特空间填充曲线相比,Z-order 空间填充曲线实现起来相对简单,但是局部性较差。

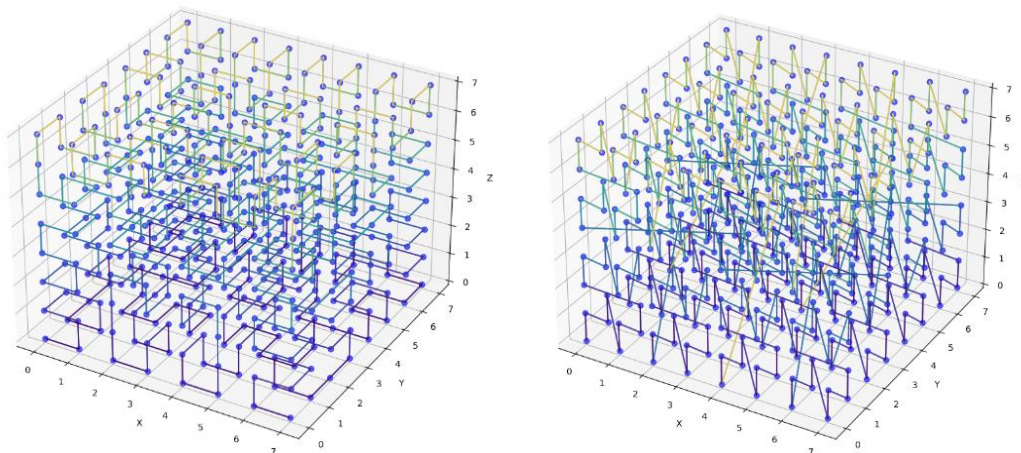


图2 空间填充曲线可视化

160 Fig. 2 The Visualization Space-filling Curve

165 在点云序列化的基础上,基于空间曲线的点云序列化方法将原始点云按空间曲线的顺序  
进行排序,然后将点云划分为多个小的分组。在每个分组内,点集在空间上是邻近的,并且  
在分组窗口内执行注意力操作。该方法通过避免邻域搜索来降低计算复杂度,序列化后的点  
云可以直接通过空间填充曲线索引来快速构建邻域。

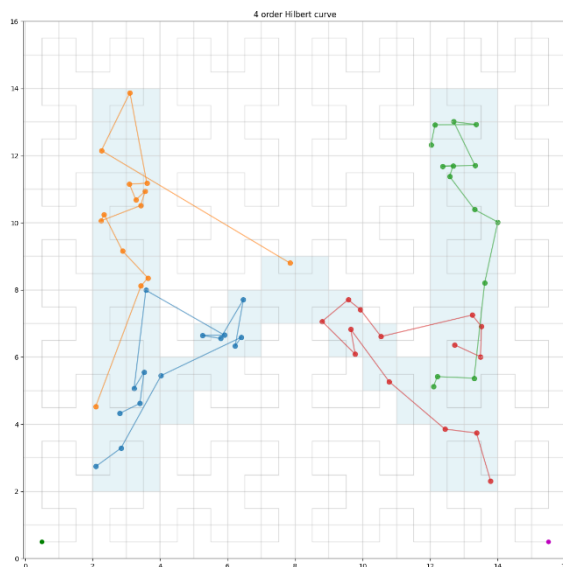


图3 点云序列化示意图

Fig. 3 Point Cloud Serialization Diagram

然而，与传统的  $K$  近邻 (KNN) 方法相比，序列化分组方法会在一定程度上牺牲邻域搜索的精度，如图 3 所示。在序列化后的点云中，点的空间位置关系不再严格保证最邻近的点会被划分到同一组。虽然这种方法能够有效降低计算复杂度，但它可能会影响邻域关系的准确性，尤其是在处理对局部几何特征敏感的任务上。此外，由于较大的感受野，模型可能难以充分学习到精细的局部特征表示。

## 2.2 局部掩码策略

为了缓解序列化注意力中的邻域关系精度损失，本方法引入了一种精确的局部掩码策略，通过局部掩码来构建精确的局部邻域，并在训练过程中强化对局部邻域的学习。

### 2.2.1 基于距离的局部掩码策略

在序列点云模型中，为了适应窗口内执行注意力的机制，本方法设计了一种针对序列窗口注意力特性的掩码策略。与之前的全局掩码方法不同，这种策略仅在同一注意力窗口内执行掩码操作，通过捕捉窗口内的几何细节，提升模型的局部特征建模能力。

基于距离的局部掩码策略利用点云中空间上邻近的点构建局部区域，掩码这些局部区域以迫使网络更好地学习局部特征表示。该策略通过以下步骤实现：

(1) 质心点采样：首先，通过最远点采样算法 (Farthest Point Sampling, FPS) 在整个点云中均匀选择质心点。FPS 算法通过最大化采样点之间的距离，确保质心点分布均匀，同时能够使每个注意力窗口内的点更具代表性。

(2) 为了便于点云模型的下采样，仅在每个注意力窗口内构建邻域。在每个注意力窗口内，采用  $K$  近邻算法 (KNN) 为每个质心点构建局部邻域。与全局掩码方法不同，本方法的邻域构建严格限制在窗口范围内。这不仅减少了跨窗口搜索的计算开销，还确保了邻域内点之间的空间连贯性和几何一致性，使其更契合序列窗口注意力机制。

(3) 在每个质心点的邻域，对其包含的点进行掩码操作。这种局部掩码方式可以更有效地利用窗口内的几何信息，避免全局掩码可能引入的噪声和冗余。此外，由于掩码是基于几何距离选择质心点，而不考虑类别信息，因此网络可以更专注于学习局部几何结构的特征表示。

掩码策略是对序列点云模型序列窗口注意力机制的进一步优化。通过仅在窗口内执行掩码操作，不仅能够提升局部特征的捕捉能力，还与模型的序列窗口注意力机制更好地结合，保证了每个窗口都有固定数量的点，同时保证其在局部区域的细节建模能力。

### 2.3 对比学习

为了解决现有序列化点云模型在局部特征捕捉方面的局限性，尤其是在建模点云中的细微几何变化时的不足，本文在局部掩码的基础上通过对比学习来增强局部特征的学习能力。本方法的核心思想是在掩码建模的基础上，引入对比学习模块，隐式增强模型对局部特征表示的学习与区分能力。

在处理输入点云数据时，首先应用随机掩码策略，遮盖掉点云中的部分点，构建精确的掩码邻域。为了更好地表征点云的局部与全局特征，本方法设计了一种双编码器架构，如图4所示。掩码编码器（Masked Encoder）将经过局部掩码处理后的点云作为输入，专注于捕捉局部邻域的特征表示；语义编码器（Semantic Encoder）则直接处理完整的点云数据，捕获完整的上下文信息。

为了进一步提升模型对局部特征的感知能力，本方法在掩码编码器输出特征之间引入了对比学习损失。具体而言，掩码编码器生成的特征作为查询，语义编码器生成的特征作为正样本，而从动态特征队列中抽取的样本作为负样本。特征对比学习模块的目标是通过最大化掩码编码器与语义编码器的特征表示之间的相似度，同时最小化与负样本的相似度，从而进一步强化局部特征的区分能力和学习效果。

在局部掩码对比学习框架下，模型能够隐式学习点云中的语义差异，从而进一步强化其在不同类别或场景下对局部特征的区分能力。特别是在复杂点云场景中的语义分割任务中。与传统方法依赖全局特征进行决策不同，掩码策略与对比学习的结合使得模型能够在全局与局部信息之间建立更为精确的映射关系，从而更好地体现局部几何差异，提高了模型对细节的敏感度。

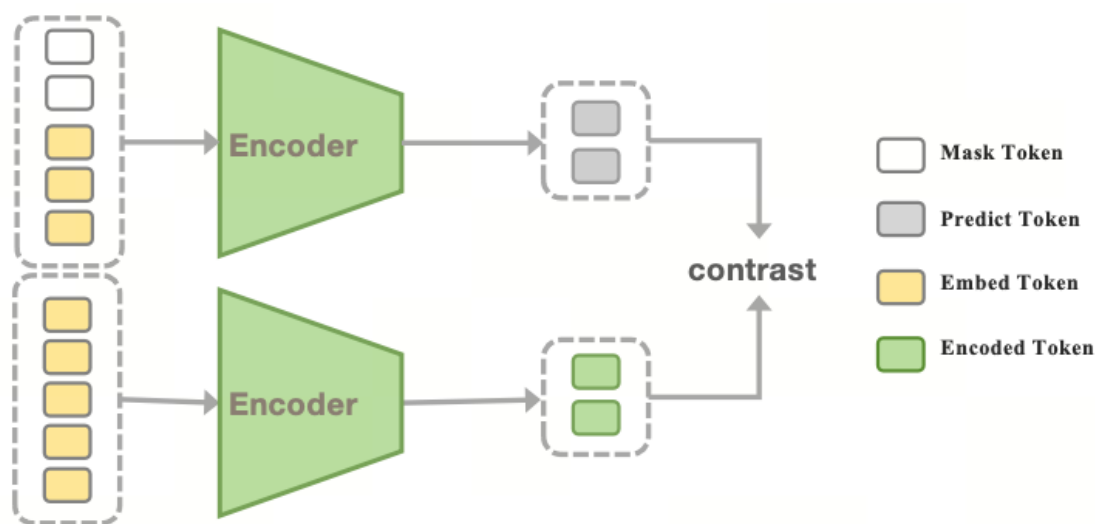


图4 双编码器对比学习示意图

Fig. 4 Dual Encoder Contrastive Learning Schematic Diagram

### 2.4 损失函数

为了更好地利用对比学习和监督学习的优势，在训练过程中需要同时优化局部掩码对比学习和点云分割两个任务。通过联合优化掩码特征对比损失和点云分割损失，模型能够在提

升局部特征学习能力的同时，保持整体分割精度。

225 对于局部特征对比学习，设掩码编码器输出的局部特征为 $q$ ，语义编码器输出的正样本特征为 $k^+$ ，负样本的特征集合为 $\{k_i^-\}$ ，对比损失定义如下：

$$\begin{aligned} & \mathcal{L}_{constast} \\ & = -\log \frac{\exp\left(\frac{\cos(q, k^+)}{\tau}\right)}{\exp\left(\frac{\cos(q, k^+)}{\tau}\right) + \sum \exp\left(\frac{\cos(q, k^-)}{\tau}\right)} \end{aligned} \quad (1)$$

对于点云分割任务，采用交叉熵损失函数

$$\begin{aligned} & \mathcal{L}_{seg} \\ 230 \quad & = \frac{1}{N} \sum_{i=1}^N CrossEntropy(y_i, \hat{y}_i) \end{aligned} \quad (2)$$

在多任务学习过程中，如果对局部特征的学习不足，模型将难以捕捉点云中的微小几何变化，而过早偏向全局特征学习则可能导致局部细节被忽视。Souhail 等<sup>[40]</sup>人表明，在训练过程中引入正则化策略有助于提高模型的泛化能力。为了更好地平衡全局和局部特征的学习，防止模型提前陷入点云分割任务的局部最优解，进一步引入了一种正则化机制。通过动态调整辅助任务和点云分割任务之间的损失权重，确保模型在不同训练阶段能够合理地聚焦于局部特征和全局语义信息的学习。因此，将损失函数中的权重动态化，最终的训练损失为：

$$\begin{aligned} & \mathcal{L} \\ & = \mathcal{L}_{seg} + \lambda(t)\mathcal{L}_{constrast} \end{aligned} \quad (3)$$

240 其中，权重调节函数  $\lambda(t)$ 采用指数衰减函数。

$$\begin{aligned} & \lambda(t) \\ & = \lambda_0 \cdot \exp(-\alpha t) \end{aligned} \quad (4)$$

$\lambda_0$ 决定了初始的权重大小， $\alpha$ 控制衰减速度， $\alpha$ 越大，权重的衰减速度越快。

### 3 实验及结果分析

245 在本节，我们在两个公开标准数据集上进行实验分析，验证 LMCL 框架的有效性，并探究了不同的参数对点云分割性能的影响。

#### 3.1 实验设置

本实验基于 Pointcept<sup>[41]</sup>代码库展开，对于训练设置，采用 AdamW 优化器优化模型参数，采用 OneCycleLR 调度器调整学习率，学习率最大为 6e-3，最小为 6e-4，总共训练 250 3000 步。为了加速模型训练，在训练过程中开启混合精度，对于 OctFormer 和 Point Transformer V3，窗口大小分别设置为 32 和 1024，其他参数遵循原始的训练设置。对于局部掩码策略，设置 K 近邻的参数 $k$ 为 16，掩码率 $r$ 为 0.5。对于训练损失，设置 $\lambda_0$ 为 1.0，衰减系数 $\alpha = 0.1$ 。

### 3.2 部件分割

#### 3.2.1 部件分割实验结果

255

ShapeNet Part 数据集是一个用于三维点云分割的基准数据集，包含 16881 个来自 16 个类别的 3D 物体。每个物体点云由 2048 个点构成，并带有精细的语义部分标签，总共包括 50 个零件标签。其类别包括飞机、椅子、桌子等，每类物体被细分为多个部件区域（如飞机的机翼、机身等）。ShapeNet Part 数据集主要用于评估模型在细粒度部件分割任务上的表现。

260

为了验证本文提出的局部掩码对比学习方法（LMCL）在点云分割任务中的有效性，在 ShapeNet Part 数据集上进行了部件分割实验，并将 LMCL 分别集成到两种现有的序列点云模型 OctFormer 和 Point Transformer V3 (PTv3) 中。实验结果以类别平均交并比 IoU (cat.mIoU) 作为衡量指标，如表 1 所示。

265

表 1 模型在 ShapeNet Part 数据集上的对比结果

Tab. 1 Comparative Results of the Models on the ShapeNet Part Dataset

模型	cat.mIoU	Airplane	Bag	Cap	Car	Chair	Earphone	Motorbike	Table
OctFormer	83.97	85.18	83.14	85.33	82.68	93.94	80.15	74.08	87.29
OctFormer+LMCL	85.34	86.09	85.27	86.91	85.12	<b>94.81</b>	81.33	75.36	87.88
PTv3	85.14	87.43	84.85	86.82	83.18	93.13	79.34	77.4	89.03
PTv3+LMCL	<b>86.76</b>	<b>88.06</b>	<b>86.32</b>	<b>87.15</b>	<b>85.33</b>	94.02	<b>81.6</b>	<b>79.71</b>	<b>90.35</b>

(1) 从整体来看，OctFormer 的 cat.mIoU 从 83.93 提升至 85.33，PTv3 的 cat.mIoU 从 85.14 提高至 86.76，通过引入局部掩码对比学习，有效增强了序列点云模型对点云数据中复杂几何特征的学习和分割能力。

270

(2) 从类别来看，LMCL 对几何结构复杂的类别（如 Earphone 和 Motorbike）提升较为明显，例如 Earphone 在 Point Transformer V3 中 cat.mIoU 从 79.34 提升至 81.6，而 Motorbike 从 77.4 提升至 79.71。同时，LMCL 也展现了较强的稳健性，确保分割性能在多个类别间的均衡提升。

#### 3.2.2 显著性检验

275

为验证 LMCL 方法的鲁棒性及其对复杂细节的感知能力，采用了独立样本 t 检验，分析 LMCL 在处理具有复杂几何形状类别时性能提升的统计显著性。实验选择 ShapeNet Part 数据集中具有较为复杂结构的类别（如 Earphone 和 Motorbike）进行测试，并选择简单类别（如 Airplane 和 Table）进行对比。对于实验的每个类别，分别进行 10 次独立检验，实验结果如表 2 所示：

280

(1) 对于复杂结构类别（如 Car 和 Motorbike），LMCL 方法通过增强局部几何特征的感知能力，在这些类别上表现出了统计显著的性能提升 ( $p < 0.05$ )，表明 LMCL 在捕捉细节信息和提升复杂几何结构的分割精度方面具有显著的优势。

(2) 对于简单几何类别（如 Airplane 和 Table），虽然 LMCL 方法有一定的性能提升，但由于这些类别的几何结构相对简单，性能提升未达到统计显著性水平 ( $p \geq 0.05$ )，

285 表明 LMCL 方法对于简单类别，其改进效果相对较小。

表 2 LMCL t-检验实验结果

Tab. 2 T-test results of LMCL

类别	PTv3	PTv3+LMCL	p 值	显著性
Earphone	79.26	81.25	0.018	显著
Motorbike	77.42	79.38	0.013	显著
Airplane	87.38	88.12	0.057	非显著
Table	88.93	89.76	0.062	非显著

### 3.3 语义分割

#### 3.3.1 语义分割实验结果

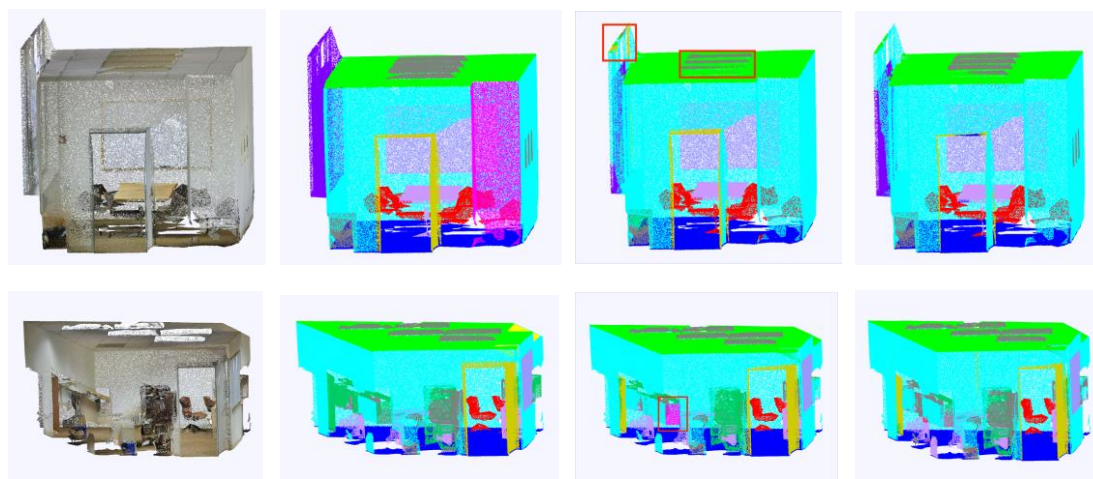
290 S3DIS 数据集专注于室内场景的点云语义分割，包含 6 个建筑区域的 3D 扫描数据，总计超过 2.7 亿个点。评估指标包括逐点分类准确率和平均交并比 (mIoU)，适用于研究室内环境下的场景理解。S3DIS 数据集广泛用于验证模型在复杂室内场景中的泛化能力。

表 3 模型在 S3DIS 数据集上的对比结果

Tab. 3 Comparative Results of the Models on the S3DIS Dataset

模型	mIoU	Ceiling	Floor	Wall	Window	Door	Table	Chair	Sofa
OctFormer	82.00	92.72	98.07	85	60.38	77.83	84.71	90.65	66.67
OctFormer+LMCL	83.03	93.31	<b>98.65</b>	86.32	60.18	<b>79.33</b>	85.07	91.8	69.63
PTV3	83.96	93.61	98.08	86.23	65.21	72.12	84.13	92.33	80.03
PTV3+LMCL	<b>85.48</b>	<b>94.31</b>	98.33	<b>87.77</b>	<b>66.07</b>	75.8	<b>85.65</b>	<b>93.19</b>	<b>82.72</b>

295 如表 3 所示，OctFormer 的 mIoU 从 82.00 提升至 83.03，PTv3 的 mIoU 则从 83.96 提升至 85.48。LMCL 的改进效果体现在多种类别上，尤其是在大多数主流类别（如 ceiling、Window 和 Sofa）的分割性能上均有稳定提升。在图 5 中可以看到 LMCL 在局部细节上分割性能提升，这表明 LMCL 的局部对比学习机制能够有效结合局部细节和全局上下文信息，使模型对点云中复杂几何结构的表征更具鲁棒性。



300 (a)原图 (b)ground truth (c)PTv3 (d)PTv3+LMCL

图 5 不同模型在 S3DIS 数据集上的对比结果

Fig. 5 Comparative Results of Different Models on the S3DIS Dataset

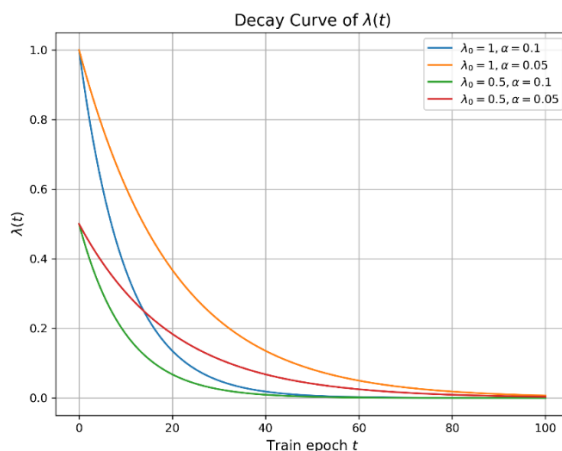
3.3.2 参数分析

305 为了进一步评估对比学习中超参数对最终分割性能的影响，本节重点分析正则化策略中  $\lambda_0$ 和 $\alpha$ 的不同组合对模型表现的具体影响。具体来说，在 S3DIS 数据集上测试了以下参数组合：

表 4 超参数组合  
Tab. 4 Hyperparameter Combination

	$\alpha=0.1$	$\alpha=0.05$
$\lambda_0=1$	(1.0, 0.1)	(1.0, 0.05)
$\lambda_0=0.5$	(0.5, 0.1)	(0.5, 0.05)

上述不同参数组合的衰减可视化结果如图 6 所示



310

图 6 不同参数组合的衰减曲线可视化

Fig. 6 Visualization of Decay Curves for Different Parameter Combinations

315 采用 Point Transformer V3 模型在 S3DIS 数据集上进行参数实验分析，为了全面评估模型的表现，选择 mIoU（mean Intersection over Union）和 OA（Overall Accuracy）作为本次实验主要的评估指标。通过结合这两个指标，分析模型在不同任务设置下的表现，并进一步探讨各个设计选择对最终性能的影响。

表 5 超参数组合对点云分割性能的影响

Tab. 5 The Impact of Hyperparameter Combinations on Point Cloud Segmentation Performance

参数组合	mIoU	OA
$\lambda_0=0.5, \alpha=0.05$	83.82	91.81
$\lambda_0=0.5, \alpha=0.1$	84.19	91.91
$\lambda_0=1.0, \alpha=0.05$	84.55	92.06
$\lambda_0=1.0, \alpha=0.1$	85.48	92.25

320

(1) 当 $\lambda_0=1.0$ 时，相较于 $\lambda_0=0.5$ ，整体性能更优，表明更高的初始正则化有助于对比学习框架中的特征区分度提升。

(2) 减小 $\alpha$ 的值（从 0.1 降低到 0.05）导致 mIoU 和 Overall Accuracy 均有所下降。这可能是因为较小的 $\alpha$ 会限制正负样本对之间的对比，进而降低特征空间的区分能力。

325 (3) 组合 (1.0, 0.1) 的参数配置在所有评估指标上均取得了最佳结果, 进一步验证了适当的正则化和衰减系数对模型性能的积极影响。实验结果表明, 在适当的参数配置下, 正则化和衰减策略能够有效提升模型的整体分割精度。

## 结论

330 本文针对序列点云模型在扩展感受野时对精细局部特征捕捉能力不足的问题, 提出了一种基于掩码对比学习的局部特征增强方法。本方法通过引入掩码对比学习任务以及正则化策略, 有效地平衡了掩码对比学习任务与点云分割任务之间的训练目标, 从而显著提升了模型对局部几何细节的表征能力。在 ShapeNet Part 和 S3DIS 数据集上的实验结果表明, 本方法在多个评价指标上均取得了优于现有方法的分割性能, 尤其在复杂几何结构和边界区域的分割精度上表现突出。本文的方法也存在一定的局限性。模型对窗口大小和掩码比例是固定的, 需要进一步优化以适应不同类型的点云数据。未来的研究可以考虑通过动态调整窗口范围和掩码比例, 进一步提升方法的适应性。

335 综上所述, 本文提出的基于掩码对比学习的局部特征增强方法, 在提升序列点云模型对局部几何特征的捕捉能力以及分割精度方面取得了一定的进展, 展现了其在复杂点云场景中的应用潜力, 希望能够为更精细化的点云分析与应用提供一些参考。

## 340 [参考文献] (References)

- [1] Xu, W., & Baolong, L. (2021). A Review of Segmentation Technology Based on 3D Point Cloud. International Journal of Advanced Network, Monitoring and Controls, 6(1), 35-40.
- [2] Lu, Dening, et al. "Transformers in 3d point clouds: A survey." arXiv preprint arXiv:2205.07417 (2022).
- 345 [3] Guo Y, Wang H, Hu Q, Liu H, Liu L, Bennamoun M. Deep learning for 3d point clouds: A survey. IEEE transactions on pattern analysis and machine intelligence. 2020 Jun 29;43(12):4338-64.
- [4] Wu, Xiaoyang, et al. "Point Transformer V3: Simpler Faster Stronger." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.
- [5] Wang, Peng-Shuai. "Octformer: Octree-based transformers for 3d point clouds." ACM Transactions on Graphics (TOG) 42.4 (2023): 1-11.
- 350 [6] He, Kaiming, et al. "Masked autoencoders are scalable vision learners." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.
- [7] He, Kaiming, et al. "Momentum contrast for unsupervised visual representation learning." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
- 355 [8] Yi Li, Shao Lin, Manolis S, et al. Large-Scale 3D shape reconstruction and segmentation from shapenet core55[J]. arXiv preprint, arXiv:1710.06104, 2017
- [9] Armeni I, Sax S, Zamir A R, et al. Joint 2D-3D-semantic data for indoor scene understanding[J]. arXiv preprint, arXiv:1702.01105, 2017
- [10] Qi C R, Su Hao, Mo K, et al. Pointnet: Deep learning on point sets for 3d classification and segmentation [C] //Proc of the 35th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 652-660
- 360 [11] Qi C R, Yi L, Su Hao, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space[J/OL]Advances in Neural Information Processing Systems, 2017, 30
- [12] Ma X, Qin C, You H, Ran H, Fu Y. Rethinking network design and local geometry in point cloud: A simple residual MLP framework[J]. , 2022.
- 365 [13] Qian G, Li Y, Peng H, Mai J, Hammoud H, Elhoseiny M, Ghanem B. PointNeXt: Revisiting PointNet++ with Improved Training and Scaling Strategies[J]. , 2022.
- [14] Wu Wenxuan, Qi Zhongang, Li Fuxi. Pointconv: Deep convolutional networks on 3D point clouds[C]//Proc of the 39th IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ:IEEE, 2019: 9621-9630
- [15] Thomas H, Qi C R, Deschaud J E, et al. Kpconv: Flexible and deformable convolution for point clouds [C] //Proc of the 37th IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 6411-6420
- 370 [16] Xu Mutian, Ding Runyu, Zhao Hengshuang, et al. Paconv: Position adaptive convolution with dynamic kernel assembling on point clouds [C] //Proc of the 39th IEEE/CVF Conf on Computer Vision and Pattern Recognition.

- Piscataway, NJ: IEEE,2021: 3173-3182
- [17] Yang F, Davoine F, Wang H, Jin Z. Continuous Conditional Random Field Convolution for Point Cloud Segmentation[J]. Pattern Recognition, 2022: 108357.
- 375 [18] Wang Y, Sun Y, Liu Z, Sarma S E, Bronstein M M, Solomon J M. Dynamic Graph CNN for Learning on Point Clouds[J]. ACM Transactions on Graphics, 2019: 1-12.
- [19] Lei H, Akhtar N, Mian A. SegGCN: Efficient 3D Point Cloud Segmentation With Fuzzy Spherical Kernel[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). , 2020.
- 380 [20] Wei M, Wei Z, Zhou H, Hu F, Si H, Chen Z, Zhu Z, Qiu J, Yan X, Guo Y, Wang J, Qin J. AGConv: Adaptive Graph Convolution on 3D Point Clouds[J]. IEEE transactions on pattern analysis and machine intelligence, 2022.
- [21] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AIDANN, Kaiser L, Polosukhin I. Attention Is All You Need[J]. Neural Information Processing Systems,Neural Information Processing Systems, 2017.
- [22] Zhao Hengshuang, Li Jiang, Jia Jiaya, et al. Point transformer [C] //Proc of the 39th IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2021: 16259-16268
- 385 [23] Wu X, Lao Y, Jiang L, Liu X, Zhao H. Point Transformer V2: Grouped Vector Attention and Partition-based Pooling[J]. , 2022.
- [24] Guo M-H, Cai J-X, Liu Z-N, Mu T-J, Martin R R, Hu S-M. PCT: Point cloud transformer[J]. Computational Visual Media, 2021: 187-199.
- 390 [25] Park C, Jeong Y, Cho M, Park J. Fast Point Transformer[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). , 2022.
- [26] Lai X, Liu J, Jiang L, Wang L, Zhao H, Liu S, Qi X, Jia J. Stratified Transformer for 3D Point Cloud Segmentation[J]. , 2022.
- [27] Yang Y-Q, Guo Y-X, Xiong J-Y, Liu Y, Pan H, Wang P-S, Tong X, Guo B. Swin3D: A Pretrained Transformer Backbone for 3D Indoor Scene Understanding[J]. , 2023.
- 395 [28] Yang, Y. Q., Guo, Y. X., & Liu, Y. Swin3D++: Effective Multi-Source Pretraining for 3D Indoor Scene Understanding. arXiv preprint arXiv:2402.14215.
- [29] Efficient Point Cloud Analysis Using Hilbert Curve.pdf[J]. .
- [30] Gu, A., & Dao, T. (2023). Mamba: Linear-time sequence modeling Gu, A., & Dao, T. (2023). Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752.
- 400 [31] Han X, Tang Y, Wang Z, Li X. Mamba3D: Enhancing Local Features for 3D Point Cloud Analysis via State Space Model[J]. .
- [32] Mamba24:8D- Enhancing Global Interaction in Point Clouds via State Space Model.pdf[J]. .
- [33] Liang D, Zhou X, Wang X, Zhu X, Xu W, Zou Z, Ye X, Bai X. PointMamba: A Simple State Space Model for Point Cloud Analysis[J]. .
- 405 [34] Zhang, Tao, et al. "Point Cloud Mamba: Point Cloud Learning via State Space Model." CoRR (2024).
- [35] Xie S, Gu J, Guo D, Qi C R, Guibas L, Litany O. PointContrast: Unsupervised Pre-training for 3D Point Cloud Understanding[G]//Computer Vision - ECCV 2020,Lecture Notes in Computer Science. , 2020: 574-591.
- [36] Yu X, Tang L, Rao Y, Huang T, Zhou J, Lu J. Point-BERT: Pre-training 3D Point Cloud Transformers with Masked Point Modeling[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). , 2022.
- 410 [37] Pang Y, Wang W, Tay F, Liu W, Tian Y, Yuan L. PointMAE Masked Autoencoders for Point Cloud Self-supervised Learning[J]. , 2022.
- [38] S. Yan et al., 3D Feature Prediction for Masked-AutoEncoder-Based Point Cloud Pretraining, in The Twelfth International Conference on Learning Representations, 2024.
- 415 [39] Tang L, Zhan Y, Chen Z, Yu B, Tao D. Contrastive Boundary Learning for Point Cloud Segmentation[J]. .
- [40] To Supervise or Not to Supervise- Understanding and Addressing the Key Challenges of Point Cloud Transfer Learning.pdf[J]. .
- [41] Pointcept Contributors.Pointcept: A Codebase for Point Cloud Perception Research. <https://github.com/Pointcept/Pointcept>
- 420