

## 基于对比学习和标签挖掘的点云分割算法

黄 华<sup>1,2,4</sup> 卜一凡<sup>1,2</sup> 许宏丽<sup>1,3</sup> 王晓荣<sup>1,2</sup>

<sup>1</sup>(北京交通大学计算机与信息技术学院 北京 100044)

<sup>2</sup>(交通数据分析与挖掘北京市重点实验室(北京交通大学) 北京 100044)

<sup>3</sup>(轨道工程北京市重点实验室(北京交通大学) 北京 100044)

<sup>4</sup>(智慧高铁系统前沿科学中心(北京交通大学) 北京 100044)

([hhua@bjtu.edu.cn](mailto:hhua@bjtu.edu.cn))

## Point Cloud Segmentation Algorithm Based on Contrastive Learning and Label Mining

Huang Hua<sup>1,2,4</sup>, Bu Yifan<sup>1,2</sup>, Xu Hongli<sup>1,3</sup>, and Wang Xiaorong<sup>1,2</sup>

<sup>1</sup> (*School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044*)

<sup>2</sup> (*Beijing Key Lab of Traffic Data Analysis and Mining (Beijing Jiaotong University), Beijing 100044*)

<sup>3</sup> (*Key Laboratory of Beijing for Railway Engineering (Beijing Jiaotong University), Beijing 100044*)

<sup>4</sup> (*Frontiers Science Center for Smart High-speed Railway System (Beijing Jiaotong University), Beijing 100044*)

**Abstract** Point cloud segmentation algorithm based on deep learning can effectively segment point clouds in high-dimensional space by designing complex feature extraction modules. However, the lack of feature mining for boundary point set results in suboptimal accuracy in boundary segmentation. Some studies have applied the idea of contrastive learning to point cloud segmentation to solve the problem of insufficient boundary region segmentation performance, but the disorder and sparse characteristics of point cloud have not been fully utilized, and the feature extraction is not accurate enough. To solve these problems, we propose CL2M to learn more accurate features of point clouds at different locations through the self-attention mechanism, and the contrastive learning method is introduced to improve the segmentation accuracy of point cloud boundaries. In the process of contrastive boundary learning, labels in semantic space are deeply mined and a contrastive boundary learning module based on label distribution is designed to make the label distribution of point cloud in high-dimensional space contain more semantic information. The model makes full use of the label distribution law to calculate the distance between distributions, and can accurately divide positive and negative samples, reducing the cumulative errors caused by conventional hard partition. The results on two public data sets show that CL2M is superior to the existing point cloud segmentation model on several evaluation indexes, which verifies the effectiveness of the model.

**Key words** computer vision; point cloud segmentation; contrastive learning; self-attention mechanism; boundary mining

**摘要** 基于深度学习的点云分割算法通过设计复杂的特征提取模块，可以对高维空间点云进行有效的

收稿日期：2023-06-12；修回日期：2024-02-06

基金项目：国家自然科学基金项目(51827813)；国家重点研发计划项目(2022YFB2603302)；北京市教育委员会科技重大项目(KJZD20191000402)；中央高校基本科研业务费专项资金(2022JBQY009)

This work was supported by the National Natural Science Foundation of China (51827813), the National Key Research and Development Program of China (2022YFB2603302), the Research & Development Program of Beijing Municipal Education Commission (KJZD20191000402), and the Fundamental Research Funds for the Central Universities (2022JBQY009).

分割。但由于缺乏对边界点集的特征挖掘，使得其对边界分割的精度欠佳。已有将对比学习思想用于点云分割以解决边界区域分割性能不足问题的研究中，忽略了点云无序和稀疏特性，特征提取不够准确。对此，设计了基于对比学习和标签挖掘的点云分割模型 CL2M (contrastive learning label mining)，通过自注意力机制学习不同位置处点云更为精准的特征，并引入对比学习方法，提高了点云边界处的分割精度。在对比边界学习过程中通过深入挖掘语义空间中的标签并设计了基于标签分布的对比边界学习模块，使得高维空间点云标签分布包含更多的语义信息。CL2M 充分利用标签的分布规律计算分布间的距离，可准确划分正负样本，减少了常规硬划分带来的累计错误。在 2 个公开数据集上进行的实验结果表明，CL2M 在多个评价指标上优于既有的点云分割模型，验证了模型的有效性。

关键词 计算机视觉；点云分割；对比学习；自注意力机制；边界挖掘

中图法分类号 TP391

DOI: [10.7544/issn1000-1239.202330491](https://doi.org/10.7544/issn1000-1239.202330491)

CSTR: [32373.14.issn1000-1239.202330491](https://cstr.cn/32373.14.issn1000-1239.202330491)

点云分割<sup>[1]</sup>是将点云数据集中的点按照一定的规则或标准划分为互不相交的子集，以实现对点云数据的进一步分析、处理或应用的过程。在点云分割算法中，如何学习得到合理的特征或标准对于精准分割至关重要。

点云分割的研究难点主要在于点云数据的自身特性：无序性、稀疏性和不规则性。由于点云数据中有很多个点，这些点无论以什么顺序出现，它们指代的信息并不改变，这就是无序性。稀疏性指的是采集环境、物体形状等因素可能导致点云数据的不均匀分布和缺失。点云的不规则性指的是这些点可以任意分布，不受任何规则的限制。

早期的点云分割技术主要利用传统的方法<sup>[2]</sup>，如基于边缘<sup>[3-4]</sup>、基于区域增长<sup>[5-6]</sup>、基于属性<sup>[7-9]</sup>、基于模型<sup>[10-12]</sup>等，它们通常基于手工提取的特征来进行分类或聚类。此外，基于图像处理技术的点云分割方法也比较流行，如基于 2 维图像的投影方法<sup>[13-15]</sup>，将点云数据投影到 2 维图像上进行分割。但这些方法存在一定的局限性，难以处理复杂的 3 维几何形状和表面纹理。近年来，深度学习算法和模型得到快速发展，并在各个领域<sup>[1,16]</sup>中得到广泛应用，在性能和评价指标方面有了较大的提升。深度学习模型可以自动从数据中学习到最具表征能力的特征，无需手动提取特征，具备大规模数据和并行计算的能力。因此在点云分割领域，从 PointNet<sup>[17]</sup>，PointNet++<sup>[18]</sup>到各种卷积模型如 KPConv<sup>[19]</sup>，PAConv<sup>[20]</sup>等，都设计了深度学习网络模型，能够很好地自适应提取点云特征并进行点云分割。

Transformer 是一种用于自然语言处理和其他序列任务<sup>[21-23]</sup>的神经网络架构，由于其网络核心自注意力操作原本就是一种集合操作，对于输入元素的顺

序和数量都是不敏感的，因此特别适合点云数据处理。Point Transformer<sup>[24]</sup> 将 Transformer 应用到点云分割中，通过结合自注意力机制来捕捉点云数据中的全局和局部信息。相比于传统的基于卷积操作的点云处理方法，Point Transformer 具有更强的表达能力和适应性，参数量少，计算效率更高。

尽管深度学习的引入在一定程度上提升了点云的整体分割精度，但在场景边界上的分割表现欠佳，从而导致整体分割性能下降。既有的模型鲜少提及边界的概念，且边界区域中相对非边界区域点的数量较少，因而难以提取其特征。对比学习 (contrastive learning)<sup>[25-27]</sup> 可以通过学习对象之间的相似性和差异性来提取更具判别性的特征，且只需要少量的标记数据就可以有效地学习，对于数据集较小或标记数据难以获取的任务非常适用。对比边界学习 (contrastive boundary learning, CBL) 方法<sup>[28]</sup> 将对比学习应用到点云分割任务中，将场景中不同位置的信息结合起来，提供更全面、更准确的特征表示。同时通过跨尺度的特征比较，能更好地捕捉场景的结构信息和语义关系。但该方法在处理点云数据时忽略了点云数据的 2 个特性：无序性和稀疏性，因此对于点云特征提取的精度不够准确。

本文据此提出基于对比学习和标签挖掘的点云分割模型 CL2M (contrastive learning label mining)，借鉴了 CBL 模型中利用对比学习来提高点云边界精度的思想，设计了正负样本的定义方式，使得模型更加关注边界处的分割效果；通过引入自注意力机制进行特征提取，计算每个位置与其他所有位置的特征相似度并加权聚合，可以有效应对点云数据的无序性。CL2M 引入基于语义空间的标签挖掘模块，缓解平均池化进行标签聚合导致的语义信息缺失问题；

并引入基于标签分布的对比边界学习模块,利用标签的分布规律缓解硬划分带来的错误累计问题.

本文的主要贡献有3个方面:

1)设计了基于对比学习和标签挖掘的点云分割模型CL2M,加入自注意力机制以更好地挖掘点云特征.同时,通过对比学习可以强化边界处的分割以提升整体分割性能.

2)提出了基于语义空间的标签分布挖掘算法,其包含了2个模块,分别是基于语义空间的标签挖掘和基于标签分布的对比边界学习模块.利用深度学习模型从输入的数据中提取出高层次的语义信息,并用于生成高维空间中的点云标签;利用KL散度计算标签向量之间的距离,可以更好地反映标签分布的差异来更精准地划分正负样本.

3)在S3DIS和ShapeNet Part数据集上进行了实验,结果表明CL2M在整体分割精度和边界分割精度这2个评价指标上优于对比方法,验证了模型的有效性.

## 1 相关工作

本节主要介绍对比学习和点云分割的国内外研究现状.

### 1.1 对比学习

对比学习<sup>[29]</sup>是一种无监督学习的方法,旨在学习如何比较输入中的不同部分,从而提高模型对于数据的表示能力.与有监督学习不同,对比学习不需要标注数据,因此可以应用于各种领域.对比学习的核心思想是将相同类别的样本映射到空间中的相近位置,而将不同类别的样本映射到空间中的相对较远位置,来学习更加鲁棒和具有区分性的表示.

近年来,对比学习成为了机器学习、计算机视觉和自然语言处理等领域的研究热点之一<sup>[30]</sup>,在分类、分割、预测等下游任务中均有重要应用.对比学习最常见的下游应用就是分类任务.针对图像分类任务,Hou等人<sup>[31]</sup>把对比学习融入半监督高光谱图像分类问题中,分阶段对模型分别进行基于对比学习方法的预训练,并利用有标注样本进行监督学习.除了分类任务,对比学习在分割任务中的作用也十分显著.Wang等人<sup>[32]</sup>提出密集对比学习自监督视觉预训练方法(DenseCL),通过将对比学习的思想应用于自监督学习中,训练一个神经网络来学习图像的特征表示,以此提高自监督学习的性能.在医学图像分割领域,Chaitanya等人<sup>[33]</sup>提出基于全局和局部特征之间

的对比学习,通过将不同部位的图像划分为不同的子区域来提取局部特征,并使用全局池化来获取全局特征,该方法旨在解决医学图像分割中标注数据不足的问题.康健等人<sup>[34]</sup>为了解决高分辨率SAR图像的分割问题,通过对监督对比学习进行正则化,即使用生成的建筑物足迹与真实建筑物足迹之间的差异来约束训练过程,提高了建筑物足迹的生成质量.这些论文<sup>[29-34]</sup>都在对比学习领域做出了重要的贡献,推动了该领域的发展和应用.

### 1.2 点云分割

近年来,基于深度学习的点云分割算法吸引了越来越多的学者关注<sup>[35-36]</sup>,他们在公开发布的数据集上进行研究,提出了许多效果显著的点云分割模型.基于逐点多层感知器的算法都是通过点云变换网络来消除点云无序性的影响,其优点在于可以直接处理点云数据,减少引入辅助手段带来的性能开销.2017年,Qi等人<sup>[17]</sup>提出了PointNet模型,直接操作点云数据,利用深度学习通过对每个点进行处理来学习全局特征.PointNet++<sup>[18]</sup>在PointNet的基础上提出改进,通过采样、分组,并使用PointNet模块提取特征,以此提高模型精度,但是此类方法缺乏局部特征的表述.由于卷积神经网络在2维图像上进行特征提取效果显著,而在3维点云数据上受其不规则性限制,传统卷积核无法进行卷积,因此,研究者们致力于解决并突破这一局限.2019年,Thomas等人<sup>[19]</sup>提出的KPConv模型具有可变形卷积核的特性,可以在不同的位置对卷积核进行形变,从而适应不同形状的点云.Wu等人<sup>[37]</sup>提出了PointConv,利用一组自适应权重来计算点云中每个点的特征表示,从而在点云数据上执行卷积操作,使得神经网络能够直接处理点云数据而无需转换为其他形式.

2021年,Zhao等人<sup>[24]</sup>提出了Point Transformer,将Transformer应用到点云处理上,通过使用自注意力机制在点云中计算特征表示,使其能够捕获不同点之间的复杂关系,其整体结构类似于Transformer,由编码器和解码器组成.编码器包括多层点自注意力模块,其中每个点自注意力模块处理点云中所有点之间的关系.解码器由多层传统的全连接神经网络组成,用于预测每个点的类别.该模型可以对任意数量的点进行端到端的处理,不需要任何先验网格或采样,在多个点云数据集上取得了不错的效果,并在速度上有显著优势.尽管上述模型在不同的模块中取得了这些进展,但点云分割中的边界问题却很少被探讨.2022年,Tang等人<sup>[28]</sup>设计了CBL方法,提出

边界的概念，将对比学习的思想融入到点云分割领域中，通过设计正负样本定义方式、计算距离、选择合适的损失函数，使正样本对相互靠近、负样本对逐渐远离，进一步提高了分割的准确性。尽管该方法是一种有效的点云分割方法，但是它存在一个问题：确定高维空间点云标签的过程比较简单，因此难以精确地判定正负样本，这会导致边界分割效果不够理想。

## 2 基于对比学习和标签挖掘的点云分割模型 CL2M

本节主要介绍基于对比学习和标签挖掘的点云分割模型 CL2M 的 2 个部分，分别是融合自注意力机制和对比学习的点云分割算法、基于语义空间的标签分布挖掘算法，并对边界问题给出了定义。

### 2.1 问题定义

本节阐述了边界的定义<sup>[28]</sup>，并对 CL2M 模型进行

了形式化定义，给出了相关的符号含义说明。若在一个点的邻域内存在不同标注标签的点，则定义其为边界点。用  $X = \{x_i\}$  表示点云数据，点云中第  $i$  个点记为  $x_i$ ，将  $x_i$  的局部邻域记为  $N(x_i)$ ，对应的标签记为  $\mathcal{L}_i$ ，边界点集  $B_i$  的定义如式(1)所示：

$$B_i = \{x_i \in X | \exists x_j \in N(x_i), \mathcal{L}_j \neq \mathcal{L}_i\}. \quad (1)$$

CL2M 模型网络整体采用类似 U-Net<sup>[38]</sup> 的编码器-解码器的结构完成点云分割任务。在编码器阶段，通过设计基于语义空间的标签挖掘模块，挖掘高维空间中语义信息的点云标签分布，以此指导点云的特征学习；在解码器阶段，通过设计基于标签分布的对比边界学习模块，精确划分正负样本来提升点云边界的分割精度。编码器和解码器之间通过跳跃连接，可以实现编码器和解码器的特征融合，进一步充分利用编码器中一些有用的信息。模型结构示意图如图 1 所示，骨干网络搭建如图 2 所示。其中，基于语义空间的标签分布挖掘包含基于语义空间的标签挖掘、

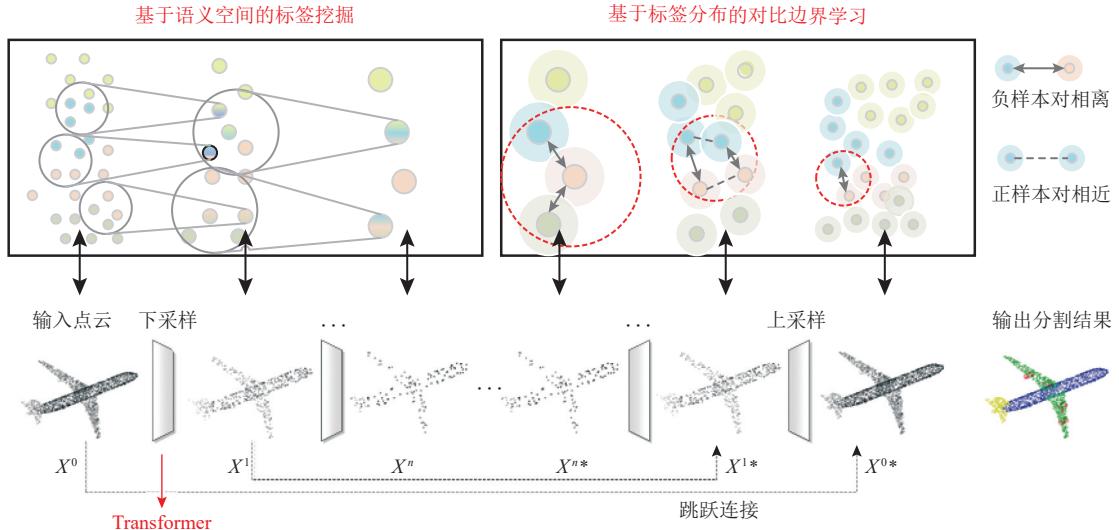


Fig. 1 Schematic diagram of CL2M model structure

图 1 CL2M 模型结构示意图

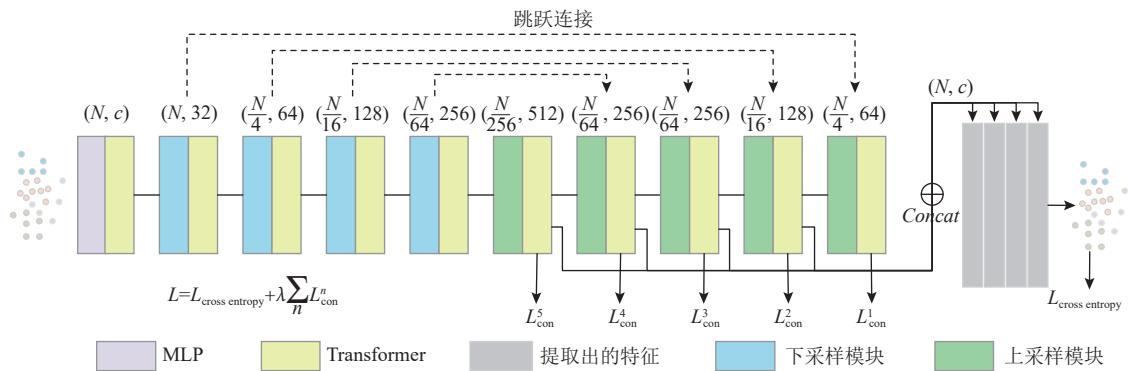


Fig. 2 CL2M backbone network

图 2 CL2M 骨干网络

基于标签分布的对比边界学习 2 个模块, 在图 2 中对应为各阶段损失的计算。此外上下采样模块依旧沿用传统方法 PointNet++中的操作。

## 2.2 融合自注意力机制和对比学习的点云分割

本节主要介绍融合自注意力机制和对比学习的点云分割, 并对其基于位置编码的特征提取模块和基于对比学习的点云分割模块进行详细阐述。

### 2.2.1 基于位置编码的特征提取

受 Point Transformer 启发<sup>[24]</sup>, 使用 Transformer 层能够适应点云的数据特性以此提取更精准的点云特征。对于进行了  $n$  次下采样阶段的点  $x_i^n$ , 将经过 Transformer 层后输出的特征  $f_i^{n+1}$  定义为:

$$f_i^{n+1} = \sum_{x_j^n \in N(x_i^n)} \rho(\gamma(\phi(f_i^n) - \psi(f_j^n) + \delta)) \odot (\alpha(f_j^n) + \delta), \quad (2)$$

其中  $f_i^n$  和  $f_j^n$  分别表示目标点  $x_i^n$  和其邻域范围内节点  $x_j^n$  的特征, 二者均为输入特征。通过使用减法运算可以求得目标点输入特征  $f_i^n$  与其邻域内某点输入特征  $f_j^n$  的差异。映射函数  $\gamma, \varphi, \psi, \alpha, \rho$  和位置编码  $\delta$ <sup>[24]</sup> 均沿用 Point Transformer 中的定义。映射函数  $\gamma$  是一个包含 2 个线性层和非线性变换 ReLU 的多层感知器, 其功能在于形成可学习的映射, 便于模型训练; 函数  $\varphi, \psi, \alpha$  对特征进行线性变换; 函数  $\rho$  采用 softmax 旨在对权重进行归一化。CL2M 通过 Transformer 层促进了特征向量之间的信息交流, 以此输出具有更准确表征的特征向量。

### 2.2.2 基于对比学习的点云分割

提升边界分割的准确度, 不仅有利于整体分割性能, 而且还能很大程度上减少稀疏数据的影响。尽管现在已经提出了各种各样的点云分割方法, 但对 3 维点云边界的关注却很少。受 CBL 启发<sup>[28]</sup>, 对比学习可以通过将点云数据分成不同的局部区域, 并将每个局部区域中的点作为正样本, 将其他区域中的点作为负样本。通过比较正样本和负样本之间的相似性, 可以学到区分边界处的点和非边界处的点的特征表示。这样就可以强化边界处的分割, 从而提升整体分割性能。CL2M 通过加入对比学习来增强跨边界的特征识别能力, 并定义正负样本对来约束损失函数。具体来说, 对于边界点  $x_i \in B_l$ , 使用 InfoNCE 损失<sup>[29]</sup> 定义边界点上的对比优化目标, 如式(3)所示:

$$L_{\text{con}} = \frac{-1}{|B_l|} \sum_{x_i \in B_l} \ln \frac{\sum_{x_j \in N(x_i) \wedge l_j = l_i} \exp(-d(f_i, f_j)/\tau)}{\sum_{x_k \in N(x_i)} \exp(-d(f_i, f_k)/\tau)}, \quad (3)$$

其中  $f_i$  是目标点  $x_i$  的特征,  $d$  是距离度量函数,  $\tau$  是温度系数。对于每个边界点  $x_i \in B_l$ , 在其局部邻域  $N(x_i)$  内采样时, 选择与  $x_i$  标签相同的点作为其正样本, 即  $\{x_j \in N(x_i) \wedge l_j = l_i\}$ , 选择与其标签不同的点作为  $x_i$  的负样本, 即  $\{x_j \in N(x_i) \wedge l_j \neq l_i\}$ 。CL2M 引入对比学习对提高边界区域的分割效果具有重要意义。

由于模型在下采样后缺失高维空间点的标签, 难以得到边界点集  $B_l$  的实际定义, 进而无法充分将对比学习的思想应用到点云分割算法中。因此, 模型需要对每个高维空间中的点进行标签标注<sup>[28]</sup>。具体来说, 定义下采样第  $n$  阶段的点云为  $X^n$ , 输入点云即为  $X^0$ 。由于下采样过程中每个点  $x_i^n$  都是由上一阶段点云  $X^{n-1}$  中的一组点经过最远点采样 (farthest point sampling, FPS)<sup>[18]</sup> 产生的, 因此可以在下采样时通过迭代的方式来确定某点在高维空间中的标签。对于  $x_i^0 \in X^0$ , 其标签  $l_i^0$  为 one-hot 类型的向量, 与 Ground-Truth 中的类别相对应。因此, 定义目标点  $x_i^n$  的标签向量  $L_i^n$  如式(4)所示:

$$L_i^n = \text{AVG}(\{L_j^{n-1} | x_j^{n-1} \in N(x_i^{n-1})\}), \quad (4)$$

其中 AVG 表示平均池化,  $N(x_i^{n-1})$  表示目标点  $x_i^n$  在第  $n-1$  阶段邻域范围内的点的集合。对于  $x_i^n \in X^n$ , 其标签向量可以表示为  $L_i^n = (l_{i,0}^n, l_{i,1}^n, \dots, l_{i,k}^n, \dots)$ ,  $L_i^n$  向量中第  $k$  个位置的  $l_{i,k}^n$  表示输入点云  $X^n$  中的点  $x_i^n$  属于第  $k$  个类别的概率。式(4)表示了通过计算  $x_i^n$  在第  $n-1$  阶段邻域范围内属于不同类别的点所占的比例可以确定标签向量  $L_i^n$ 。接着, CBL<sup>[28]</sup> 通过  $\arg \max L_i^n$  确定第  $n$  阶段节点  $x_i^n$  的类别标签, 再根据式(1)边界点集的定义可以得到当前下采样阶段的边界点。如图 2 所示, 在此基础上结合下采样点的特征, 根据 InfoNCE 损失函数将对比学习应用到各个下采样阶段就可以达到优化边界的目的, 最终的损失函数定义为:

$$L = L_{\text{cross entropy}} + \lambda \sum_n L_{\text{con}}^n, \quad (5)$$

其中  $L_{\text{con}}^n$  表示第  $n$  阶段利用函数 InfoNCE 求得的对比损失,  $\lambda$  为损失权重,  $L_{\text{cross entropy}}$  表示交叉熵损失。

## 2.3 基于语义空间的标签分布挖掘

本节主要介绍基于语义空间的标签分布挖掘算法, 并对其中的基于语义空间的标签挖掘模块和基于标签分布的对比边界学习模块进行详细阐述。

### 2.3.1 基于语义空间的标签挖掘

由于 CBL<sup>[28]</sup> 使用平均池化来确定高维空间点云标签的方法比较简单, 忽略了空间中的语义信息, 这会为精确判定正负样本造成困难, 因而导致边界分

割效果不够理想。图 3 展示了使用平均池化造成标签分布不合理现象。图 3 中用不同颜色区分标签类别，即蓝色和黄色表示不同的标签；用不同形状区分特征，即三角形和圆形表示点与点之间具有不同的特征，具有相同形状的 2 个点表示它们的特征更相似。对于目标点  $x_i^{n-1}$ ，在其邻域范围内只有 3 个点  $x_j^{n-1}$ ， $x_k^{n-1}$ ， $x_q^{n-1}$  的标签与  $x_i^{n-1}$  的标签相同，其余 6 个点与  $x_i^{n-1}$  的标签不同，但是实际上  $x_i^{n-1}$  邻域范围内某一关键点  $x_k^{n-1}$  与  $x_i^{n-1}$  的特征相似度更高，因此  $L_k^{n-1}$  对于  $x_i^n$  的标签  $L_i^n$  贡献程度更大，也就是说目标点  $x_i^{n-1}$  在经过下采样后在第  $n$  阶段的标签  $L_i^n$  应该更趋向于目标点在第  $n-1$  阶段邻域范围  $N(x_i^{n-1})$  内与其特征更相似的点的标签，即点  $x_j^{n-1}$ ， $x_k^{n-1}$ ， $x_q^{n-1}$  的标签，而非简单依靠计算目标点第  $n-1$  阶段邻域范围内属于不同类别的点所占的比例来确定其高维空间中的标签。因此，如果使用平均池化来确定高维空间点的标签会导致边界分割效果不够理想。

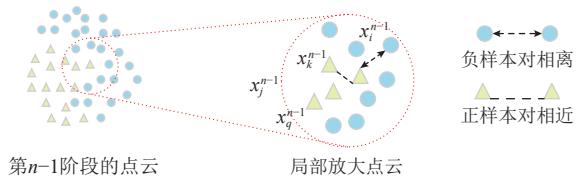


Fig. 3 Demonstration of unreasonable label distribution due to the use of average pooling

图 3 使用平均池化导致的标签分布不合理现象的展示

基于此，本文提出基于语义空间的标签挖掘方法，利用自注意力机制的思想将目标点在第  $n-1$  阶段的邻域内每一个点  $x_j^{n-1} \in N(x_i^{n-1})$  的标签  $L_j^{n-1}$  聚合生成目标点  $x_i$  在第  $n$  阶段的标签  $L_i^n$ ，核心公式如式(6)所示：

$$L_i^n = \sum_{x_j^{n-1} \in N(x_i^{n-1})} \rho(\gamma(\beta(\varphi(f_i^{n-1}), \psi(f_j^{n-1})))) \odot L_j^{n-1}, \quad (6)$$

其中  $f_i^{n-1}$  表示目标点  $x_i$  在第  $n-1$  阶段的特征， $L_j^{n-1}$  表示节点  $x_j$  在第  $n-1$  阶段的标签。 $\varphi$  和  $\psi$  是可学习的变换，负责将  $f_i^{n-1}$  和  $f_j^{n-1}$  进行线性映射，并采用点乘的方法作为函数  $\beta$  进行相似度计算。映射函数  $\gamma$  是包含 2 个线性层和非线性变换 ReLU 的多层感知器。函数  $\rho$  采用 softmax 函数旨在对权重进行归一化。总的来说，CL2M 模型根据第  $n-1$  阶段目标点  $x_i$  的特征  $f_i^{n-1}$  与邻域内节点  $x_j$  的特征  $f_j^{n-1}$  计算注意力系数，为与目标点特征相似度更高的点  $x_k^{n-1}$  的标签赋予更高的权值，更精确地获取目标点第  $n$  阶段的标签分布  $L_i^n$ 。通过上述方法，能够有效解决使用平均池化导致的标签分

布不合理问题。

### 2.3.2 基于标签分布的对比边界学习

基于语义空间的标签挖掘方法充分利用了点云的语义特征与位置关系，获得了高维空间中点的标签分布。然而，CBL<sup>[28]</sup> 通过  $\arg \max L_i^n$  确定目标点在高维空间中的标签，可能会导致正负样本判定错误，进而影响边界的分割效果。如图 4 所示用不同颜色来区分点的标签，展示了使用硬划分导致正负样本判定不合理的现象。例如对于点  $x_p$  与点  $x_Q$ ，其标签分布分别为  $L_p = (0.4, 0.3, 0.3)$  和  $L_Q = (0.7, 0.2, 0.1)$ ，如果使用  $\arg \max L_i^n$  的方法确定该点的标签类别，则会将  $x_Q$  判定为  $x_p$  的正样本，然而实际上  $x_Q$  与  $x_p$  构成负样本对，这就造成了正负样本判定错误。因此，简单通过  $\arg \max L_i^n$  方法会导致正负样本采样不精准，模型无法学习样本之间的差异，这会影响对比学习的效果。

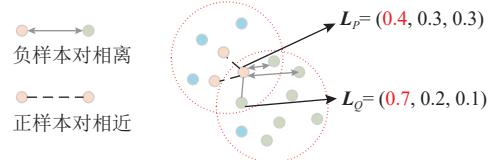


Fig. 4 Using hard division leads to the display of unreasonable positive and negative sample judgments

图 4 使用硬划分导致正负样本判定不合理的展示

为了更好地衡量正负样本之间的距离，本文提出基于标签分布的对比边界学习方法。KL 散度 (Kullback-Leibler divergence) 也称为相对熵，是一种用于衡量 2 个概率分布之间差异的指标。KL 散度是非对称的，它衡量的是从一个概率分布到另一个概率分布的信息损失。在这里，KL 散度计算标签向量分布间的距离，并据此进行负样本采样，其核心公式为：

$$F_{\text{KL}}(P \parallel Q) = \sum_x P(x) \ln \frac{P(x)}{Q(x)}, \quad (7)$$

其中  $P(x)$  表示目标点  $x_p$  的标签分布， $Q(x)$  表示目标点邻域范围内节点  $x_Q$  的标签分布， $F_{\text{KL}}(P \parallel Q)$  表示  $x_p$  与  $x_Q$  标签分布之间的距离。若  $F_{\text{KL}}(P \parallel Q) > \text{eps}$ ，则定义  $x_Q$  为  $x_p$  的负样本，反之则称  $x_Q$  为  $x_p$  的正样本，CL2M 使用这种方法能更好地划分正负样本，有效提高边界点云的挖掘精度。

## 3 结果与讨论

本节主要介绍 CL2M 的实验结果，并对其进行定量和定性分析，最后通过消融实验证各模块的有效性。

### 3.1 数据集

实验使用 2 个公开的数据集, 包括室内场景数据集 S3DIS 和部件分割 ShapeNet Part 数据集。

S3DIS(Stanford large-scale 3D indoor spaces)数据集<sup>[39]</sup>是一个用于室内 3 维场景分割的数据集。该数据集包含 6 个不同建筑物内的 11 种场景, 包括办公室、会议室、走廊、楼梯等不同类型的房间共 272 个场景, 并且标记了 13 个语义类别的标签, 如墙壁、地板、天花板、家具等。该数据集通过激光扫描设备采集点云数据以高精度捕捉室内环境, 被广泛用于 3 维场景分割的研究中, 并已成为评估算法性能和比较不同算法的标准数据集之一。

ShapeNet Part 数据集<sup>[40]</sup>是一个用于部件分割的大型 3 维数据集。该数据集包含 16 个形状类别的 16 880 个点云模型, 其中 14 006 个用于训练, 2 874 个用于测试, 每个类别标注 2~6 个部件, 共标注了 50 个不同的部件。该数据集可用于训练深度学习模型自动识别或分割 3 维模型中的不同部分, 以实现更高级别的 3 维场景理解。

### 3.2 实验设置

本节主要介绍实验的相关设置, 包含对比模型的简要介绍、点云分割算法的评价指标以及训练模型的相关设置。

#### 3.2.1 对比模型

为了验证 CL2M 的性能, 将其与 5 种方法进行对比:

1) PointNet<sup>[17]</sup>通过将点云中的每个点作为网络的输入, 并对每个点进行特征提取和聚合, 最终得到整个点云的特征表示, 这种处理方式具有可扩展性和不变性。

2) KPConv<sup>[19]</sup>通过对点云进行采样和聚类, 得到一组核点。然后, 对于每个点, KPConv 会根据它周围的核点计算出 1 组权重, 再根据这些权重将周围点的特征进行加权平均。这个过程类似于传统的卷积操作, 但是针对的是点云数据。

3) ConvNet 是 CBL 中<sup>[29]</sup>提出的一种基础网络, 通过在 3 维空间中构建的卷积神经网络来实现点云分割。

4) Point Transformer<sup>[24]</sup>通过引入 Transformer 的自注意力机制来对点云进行特征提取, 这些特征表示可以用于各种点云处理任务, 如点云分割、点云分类、点云重建等。

5) CBL<sup>[28]</sup>通过对点云边界区域进行对比学习、划分正负样本, 来提高点云边界分割的准确性。

#### 3.2.2 评价指标

为了评估 CL2M 的分割性能, 本文使用  $mIoU$ ,  $mIoU@boundary$ ,  $mIoU@inner$ <sup>[29]</sup>作为模型的性能评价

指标。

$IoU$ 是计算 2 个集合的交集和并集之比, 均交并比  $mIoU$ 是  $IoU$ 的平均值, 可以表现出每类点云的分割精度, 计算如式(8)所示:

$$IoU(X) = \frac{Gt \cap Pre}{Gt \cup Pre} \times 100\%, \quad (8)$$

其中  $Gt$  表示整片点云  $X$  中实际属于某一类标签的点的集合,  $Pre$  表示整片点云  $X$  中预测属于这一类标签的点的集合, 分子求交集表示二者标签均相同的个数, 分母求并集表示实际是此标签或者预测是此标签的个数, 最后再换化成百分比的形式。通过改变定义域的范围, 可以定义  $mIoU@boundary$  和  $mIoU@inner$ , 如式(9)(10)所示:

$$mIoU@boundary = mIoU(B_t), \quad (9)$$

$$mIoU@inner = mIoU(X - B_t), \quad (10)$$

其中  $B_t$  为边界区域点的集合,  $X - B_t$  为非边界区域点的集合。

#### 3.2.3 参数设置

CL2M 模型通过 PyTorch 实现。对于 S3DIS 数据集, 为了公平比较, 训练 CL2M 模型时将 batchsize 设为 16, 总的点云数量均为 2 048。使用 SGD 优化器训练 600 个 epoch, 权重衰减为 0.000 1, 学习率设置为 0.5。设置优化器动量为 0.9, 丢失率 dropout 为 0.5。CL2M 中对比学习相关参数设置温度系数  $\tau = 1$ , 正负样本阈值  $eps = 0.5$ , 损失权重  $\lambda = 0.1$ , 下采样过程中使用  $k$  近邻的  $k=16$ 。对于 ShapeNet Part 数据集, 除学习率设置为 0.05 外, 其余参数一致。模型使用 GeForce RTX 2080 Ti 的 GPU 完成训练。

### 3.3 定量分析

表 1 展示了 CL2M 与对比模型在 S3DIS 和 ShapeNet Part 数据集上的  $mIoU$ ,  $mIoU@boundary$ ,  $mIoU@inner$ , 表 2 和表 3 分别展示了 CL2M 与对比模型在 S3DIS 和 ShapeNet Part 数据集上的实验结果。其中, 在 S3DIS 数据集上进行的对比实验均引用 CBL<sup>[29]</sup>中展示的实验结果。通过表 1~3 得出 3 个结论:

1) 如表 1~3 所示, CL2M 对比其他方法在 S3DIS 数据集上  $mIoU$  指标最高提升了 1.3 个百分点, 在 ShapeNet Part 数据集上  $mIoU$  指标最高提升了 0.6 个百分点。CL2M 在 2 个数据集上对比其他方法  $mIoU$  指标均有所提升, 因此, 实验证明了 CL2M 模型能有效提升点云的总体分割精度。

2) 如表 1 所示, 无论是在 S3DIS 数据集还是在 ShapeNet Part 数据集上, CL2M 对比 CBL 模型  $mIoU$  的提升幅度均高于对比 Point Transformer 模型的提升

**Table 1 Experimental Results of Each Model on S3DIS and ShapeNet Part Datasets**

表 1 各模型在 S3DIS 和 ShapeNet Part 数据集上的实验结果

模型	S3DIS			ShapeNet Part			%
	<i>mIoU</i>	<i>mIoU@boundary</i>	<i>mIoU@inner</i>	<i>mIoU</i>	<i>mIoU@boundary</i>	<i>mIoU@inner</i>	
PointNet	41.1	30.2	53.4	82.9	64.0	83.5	
KPConv	67.1	50.5	70.9	86.0	65.1	86.6	
ConvNet	67.4	50.1	71.2	85.9	65.2	85.3	
Point Transformer	70.0	51.5	73.8	86.4	65.3	87.3	
CBL	69.4	52.6	73.1	86.2	66.3	86.9	
CL2M (本文)	<b>71.3</b>	<b>54.9</b>	<b>74.8</b>	<b>87.0</b>	<b>68.2</b>	<b>87.6</b>	

注：黑体数值表示最优值。

**Table 2 Experimental Results of Each Model in Some Categories on S3DIS Dataset**

表 2 各模型在 S3DIS 数据集上部分类别的实验结果

模型	<i>mIoU</i>	ceiling	floor	wall	window	door	table	chair	sofa	board	clutter	%
PointNet	41.1	88.8	97.3	69.8	46.3	10.8	59.0	52.6	5.9	26.4	33.2	
KPConv	67.1	92.8	97.3	82.4	58.0	69.0	81.5	91.0	75.4	66.7	58.9	
ConvNet	67.4	94.1	98.1	83.1	53.5	73.0	91.7	82.3	76.5	66.9	60.8	
Point Transformer	70.0	<b>95.2</b>	98.6	85.1	62.5	75.9	81.5	<b>91.0</b>	75.1	76.4	60.2	
CBL	69.4	93.9	98.4	84.2	57.7	71.9	<b>91.7</b>	81.8	<b>77.8</b>	69.1	62.9	
CL2M (本文)	<b>71.3</b>	94.7	<b>98.6</b>	<b>87.6</b>	<b>64.1</b>	<b>78.5</b>	83.2	90.5	76.6	<b>81.5</b>	<b>63.0</b>	

注：黑体数值表示最优值。

**Table 3 Experimental Results of Each Model in Some Categories on ShapeNet Part Dataset**

表 3 各模型在 ShapeNet Part 数据集上部分类别的实验结果

模型	<i>mIoU</i>	airplane	cap	car	chair	guitar	knife	lamp	laptop	mug	table	%
PointNet	82.9	82.9	80.4	74.1	87.8	90.9	85.1	79.7	95.2	91.3	80.6	
KPConv	86.0	86.1	86.5	79.0	91.8	92.5	88.5	84.3	97.1	94.0	83.3	
ConvNet	85.9	85.6	85.7	78.9	91.9	92.7	88.4	84.2	97.9	94.4	83.8	
Point Transformer	86.4	86.8	86.6	79.4	92.9	93.1	89.2	85.0	97.6	94.7	84.6	
CBL	86.2	86.0	86.7	79.5	93.0	93.2	88.9	84.8	<b>98.0</b>	<b>95.0</b>	84.7	
CL2M (本文)	<b>87.0</b>	<b>87.4</b>	<b>87.2</b>	<b>79.7</b>	<b>93.1</b>	<b>93.3</b>	<b>89.8</b>	<b>86.2</b>	97.9	94.9	<b>85.2</b>	

注：黑体数值表示最优值。

幅度。由于加入对比学习，略优于 Point Transformer，但提升性能不高，可能重在提升边界性能；由于加入自注意力机制模型分割效果远优于先前的其他模型，证明了 Transformer 结构对模型分割的优越性。

3) 如表 1 所示，CL2M 在 S3DIS 和 ShapeNet Part 数据集上对比其他方法，*mIoU@boundary* 评价指标的提升幅度高于 *mIoU@inner*。因此，实验证明了 CL2M 能够有效提升边界区域分割精度。

### 3.4 定性分析

本节展示不同模型对边界点云分割的效果。如图 5、图 6 所示，分别为各模型在 S3DIS 数据集和 ShapeNet Part 数据集上的边界效果可视化，其中：由图 5(c)(e) 和图 6(c)(e) 可知，CL2M 可以有效识别出

边界，而 Point Transformer 模型对于边界点云的识别没有明显作用。由于 Point Transformer 重在特征提取，没有专注于挖掘边界特征，因此对边界识别效果有限。由图 5(d)(e) 和图 6(d)(e) 可知，CBL 和 CL2M 都能够识别出边界，说明对比学习方法对于识别边界有效。由于对比学习通过挖掘子场景的标签可以更好地区分正负样本，因此能提升边界处的点云分割性能。并且图 5(e) 比图 5(d) 识别的边界更加精细，说明 CL2M 模型提出的标签分布学习对于识别更加精细的边界有益。综上所述，CL2M 模型对于边界点云分割效果有较大提升。

### 3.5 消融实验

本节主要验证 CL2M 仅使用单一模块对模型性

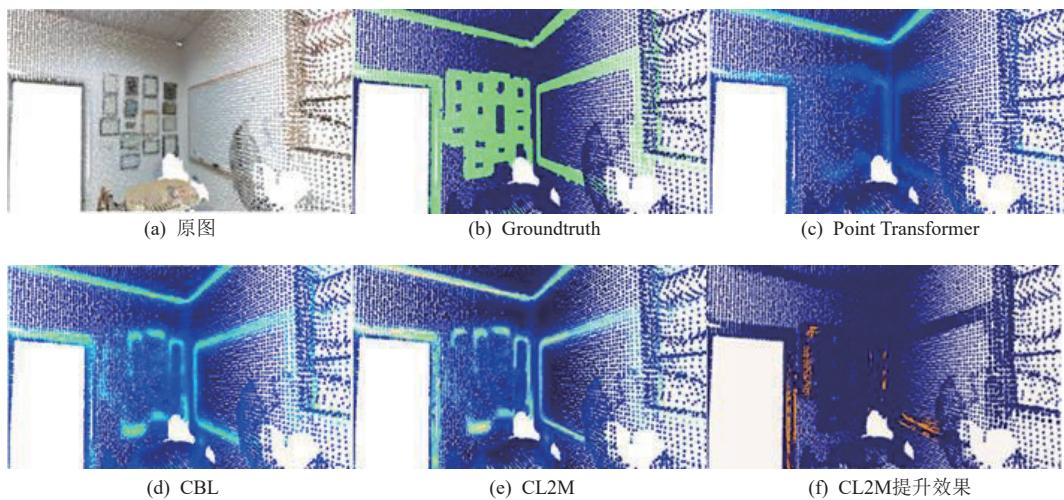


Fig. 5 Boundary performance comparison of different models on S3DIS dataset

图 5 不同模型在 S3DIS 数据集的边界效果展示

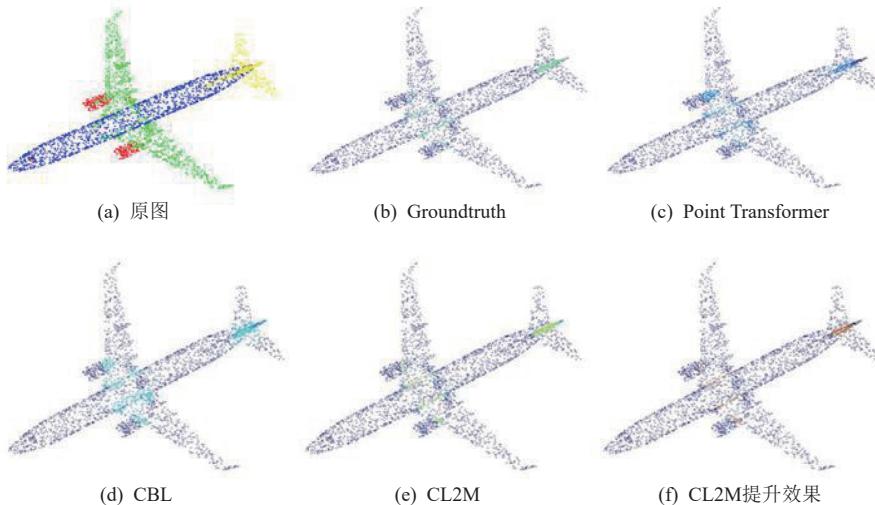


Fig. 6 Boundary performance comparison of different models on ShapeNet Part dataset

图 6 不同模型在 ShapeNet Part 数据集的边界效果展示

能的影响, 图 7 展示了不同模块对模型分割精度 *mIoU* 指标的影响。本节将这些模型与 CBL 基准模型进行比较, 并在其基础上引入自注意力机制。通过分别应用标签语义挖掘模块和标签分布对比模块, 我们证明了 CL2M 模型在性能上优于仅使用标签分布学习结构和仅使用 Transformer 结构的模型。实验结果验证了 CL2M 的 2 个结构对性能的提升效果, Transformer 结构对提升整体分割准确度的效果明显; 标签分布学习结构对模型整体分割性能提升不大, 其主要提升的是边界分割性能。

因此, CL2M 模型在语义理解和标签分布学习方面具有较高的优势。通过引入自注意力机制, CL2M 模型进一步增强了模型的表达能力和泛化能力。而标签语义挖掘模块和标签分布对比模块能够更好地

捕捉点云边界的关键信息, 从而提高了模型的性能。

### 3.6 参数灵敏度分析

为验证不同参数对模型性能的影响, 本节对温度系数和正负样本划分阈值进行参数灵敏度分析。表 4 展示了设置不同的温度系数对模型的影响, 实验结果表明, 当温度系数  $\tau = 1$  时, CL2M 效果较好。表 5 展示了不同阈值  $eps$  对模型的影响, 实验结果表明, 当  $eps = 0.5$  时 CL2M 模型效果较好。

## 4 结 论

本文提出了基于对比学习和标签挖掘的点云分割模型 CL2M, 加入自注意力机制构建 Transformer 层和对比学习分别用于学习每个位置点云更为精准的

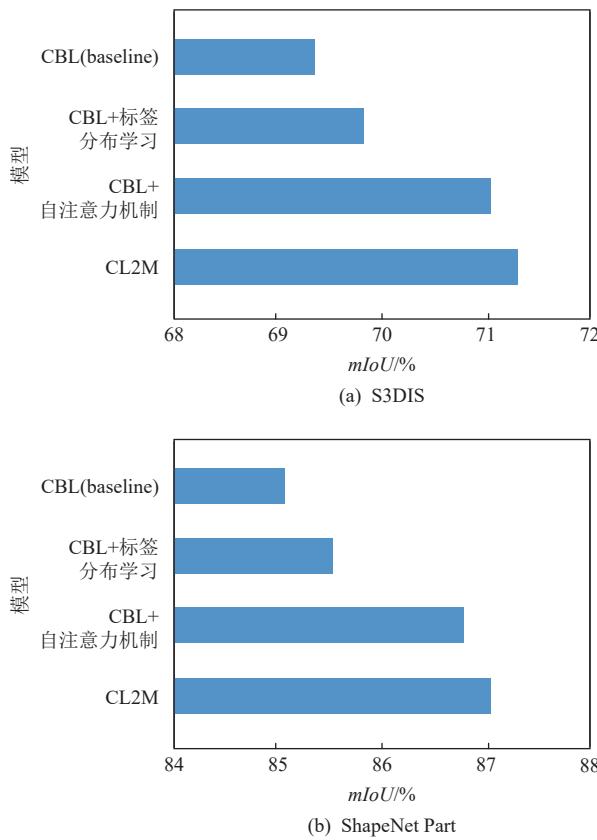


Fig. 7 Impact of different modules on the model

图 7 不同模块对模型的影响

**Table 4 Effect of Temperature Coefficient on Point Cloud Segmentation Performance**

表 4 温度系数对点云分割性能的影响

温度系数 $\tau$	mIoU/%	
	S3DIS	ShapeNet Part
0.3	70.67	86.35
0.5	70.98	86.65
1	71.33	87.03
2	70.73	86.40
5	70.03	85.71

**Table 5 Effect of Positive and Negative Sample Segmentation Thresholds on Point Cloud Segmentation Performance**

表 5 正负样本划分阈值对点云分割性能的影响

eps	mIoU/%	
	S3DIS	ShapeNet Part
0.1	70.05	86.23
0.3	70.87	86.69
0.5	71.33	87.03
0.7	70.74	86.65
0.9	70.02	86.21

特征和提高点云边界处的分割精度。通过引入基于语义空间的标签挖掘和基于标签分布的对比边界学习，使高维空间中的点云标签分布包含更多的语义信息，同时充分利用标签的分布规律计算标签向量分布间的距离，能够更加准确地划分正负样本，减少了通过硬划分带来的累计错误。实验结果表明，CL2M 对比其他点云分割算法在各个数据集上性能均有所提升，可以有效提升边界和整体的分割效果。

**作者贡献声明：**黄华提出研究思路、工作方案并修改论文；卜一凡开展研究实验并撰写论文；许宏丽审阅并提出修改意见；王晓荣参与实验。

## 参 考 文 献

- [1] Gong Zheng, Sun Junhua, Zhang Guangjun. Dynamic structured-light measurement for wheel diameter based on the cycloid constraint [J]. *Applied Optics*, 2016, 55(1): 198–207
- [2] Qin Caijie, Guan Qiang. Research status of 3D point cloud data segmentation [J]. *Journal of Yibin University*, 2017, 17(6): 30–35 (in Chinese)  
(秦彩杰, 管强. 三维点云数据分割研究现状 [J]. *宜宾学院学报*, 2017, 17(6): 30–35)
- [3] Bhanu B, Lee S, Ho C C, et al. Range data processing: Representation of surfaces by edges [C] //Proc of the 8th Int Conf on Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 1986: 236–238
- [4] Jiang Xiaoyi, Bunke H. Edge detection in range images based on scan line approximation [J]. *Computer Vision and Image Understanding*, 1999, 73(2): 183–99
- [5] Wang Zhe, Hong Liu, Yueliang Qian, et al. Real-time plane segmentation and obstacle detection of 3D point clouds for indoor scenes [C] //Proc of the Computer Vision–ECCV. Berlin: Springer, 2012: 22–31
- [6] Papon J, Alexey A, Markus S, et al. Voxel cloud connectivity segmentation-supervoxels for point clouds [C] //Proc of the 31st IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2013: 2027–2034
- [7] Zhan Qingming, Liang Yu, Liang Yubing. A point cloud segmentation method based on vector estimation and color clustering [C] //Proc of the 2nd Int Conf on Information Science and Engineering. Piscataway, NJ: IEEE, 2010: 3463–3466
- [8] Zhan Qingming, Liang Yu. Segmentation of LiDAR point cloud based on similarity measures in multi-dimension Euclidean space [C] //Proc of the 2nd Advances in Computer Science and Engineering. Berlin: Springer, 2012: 349–357
- [9] Holz D, Holzer S, Rusu R B, et al. Real-Time plane segmentation using RGB-D cameras [C] //Proc of the 15th RoboCup. Berlin: Springer, 2011: 306–317

- [10] Li Yangyan, Wu Xiaokun, Chrysathou Y, et al. Globfit: Consistently fitting primitives by discovering global relations [C/OL] //Proc of the 11th ACM SIGGRAPH. New York: ACM, 2011[2023-01-20]. <https://dl.acm.org/doi/abs/10.1145/1964921.1964947>
- [11] Awadallah M, Lynn A, Sherin G. Segmentation of sparse noisy point clouds using active contour models [C]//Proc of the 21st IEEE Int Conf on Image Processing (ICIP). Piscataway, NJ: IEEE, 2014: 6061–6065
- [12] Wang Yanmin, Shi Hongbin. A segmentation method for point cloud based on local sample and statistic inference [C] //Proc of the 2nd Int Conf on Geo-Informatics in Resource Management and Sustainable Ecosystem(GRMSE). Berlin: Springer, 2015: 274–282
- [13] Boulch A, Guerry J, Le Saux B, et al. SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks[J]. Computers & Graphics, 2018, 71: 189–198
- [14] Wang Wei, Xu Yuan, Ren Yingchao, et al. Parsing of urban facades from 3D point clouds based on a novel multi-view domain[J]. Photogrammetric Engineering & Remote Sensing, 2021, 87(4): 283–293
- [15] Kundu A, Yin X, Fathi A, et al. Virtual multi-view fusion for 3D semantic segmentation [C] //Proc of the 16th European Conf on Computer Vision (ECCV). Berlin: Springer, 2020: 518–535
- [16] Chen Xiaozhi, Ma Huimin, Wan Ji, et al. Multi-view 3D object detection network for autonomous driving [C] //Proc of the 35th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 1907–1915
- [17] 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
- [18] 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. [2024-01-20]. [https://proceedings.neurips.cc/paper\\_files/paper/2017/hash/d8bf84be3800d12f74d8b05e9b89836f-Abstract.html](https://proceedings.neurips.cc/paper_files/paper/2017/hash/d8bf84be3800d12f74d8b05e9b89836f-Abstract.html)
- [19] 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
- [20] 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
- [21] Wu Felix, Fan A, Baevski A, et al. Pay less attention with lightweight and dynamic convolutions [J]. arXiv preprint, arXiv: 1901.10430, 2019
- [22] Yang Zhilin, Dai Zihang, Yang Yiming, et al. XLNet: Generalized autoregressive pretraining for language understanding[J]. Neural Information Processing Systems, 2019, 33: 5753–5763
- [23] Devlin J, Chang Ming-Wei, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [J]. arXiv preprint, arXiv: 1810.04805, 2018
- [24] 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
- [25] Chen Ting, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations [C] //Proc of the 37th Int Conf on Machine Learning. Washington, DC: ICLR, 2020: 1597–1607
- [26] Khosla P, Teterwak P, Wang Chen, et al. Supervised contrastive learning[J]. Neural Information Processing Systems, 2020, 33: 18661–18673
- [27] Xu Yufei, Zhang Qiming, Zhang Jing, et al. RegionCL: Can simple region swapping contribute to contrastive learning[J]. arXiv preprint, arXiv: 2111.12309, 2021
- [28] Tang Liyao, Zhan Yibing, Chen Zhe, et al. Contrastive boundary learning for point cloud segmentation[C]//Proc of the 40th IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 8489–8499
- [29] Oord A, Li Yazhe, Vinyals O. Representation learning with contrastive predictive coding[J]. arXiv preprint, arXiv: 1807.03748, 2018
- [30] Zhang Chongsheng, Chen Jie, Li Qilong, et al. A review of deep contrast learning[J]. Acta Automatica Sinica, 2023, 49(1): 15–39 (in Chinese)  
(张重生, 陈杰, 李岐龙, 等. 深度对比学习综述[J]. 自动化学报, 2023, 49(1): 15–39)
- [31] Hou Sikang, Shi Hongye, Cao Xianghai, et al. Hyperspectral imagery classification based on contrastive learning[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1–13
- [32] Wang Xinlong, Zhang Rufeng, Shen Chunhua, et al. Dense contrastive learning for self-supervised visual pre-training[C] //Proc of the 39th IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 3024–3033
- [33] Chaitanya K, Erdil E, Karani N, et al. Contrastive learning of global and local features for medical image segmentation with limited annotations[J]. Advances in Neural Information Processing Systems, 2020, 33: 12546–12558
- [34] Kang Jian, Wang Zhirui, Zhu Ruoxin, et al. High-resolution SAR image building extraction method based on supervised contrast learning regularization[J]. Journal of Radar, 2022, 11(1): 157–167 (in Chinese)  
(康健, 王若鑫, 祝若鑫, 等. 基于监督对比学习正则化的高分辨率 SAR 图像建筑物提取方法[J]. 雷达学报, 2022, 11(1): 157–167)
- [35] Hou Linlin. Research on 3D point cloud segmentation algorithm based on geometric features [D]. Beijing: Beijing Jiaotong University, 2019 (in Chinese)  
(侯琳琳. 基于几何特征的三维点云分割算法研究[D]. 北京: 北京交通大学, 2019)
- [36] Li Jiaojiao, Sun Hongyan, Dong Yu, et al. Survey of 3-dimensional point cloud processing based on deep learning[J]. Journal of Computer Research and Development, 2022, 59(5): 1160–1179 (in Chinese)  
(李娇娇, 孙红岩, 董雨, 等. 基于深度学习的 3 维点云处理综述[J]. 计算机研究与发展, 2022, 59(5): 1160–1179)
- [37] 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

- [38] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation[C] //Proc of the 18th Medical Image Computing and Computer-Assisted Intervention (MICCAI). Berlin: Springer, 2015: 234–241
- [39] 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
- [40] 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



**Huang Hua**, born in 1977. PhD, associate professor. Member of CCF. His main research interests include computer vision, machine learning, and the applications in the railway field.

黄 华, 1977 年生. 博士, 副教授. CCF 会员. 主要研究方向为计算机视觉、机器学习及其在铁路领域中的应用.



**Bu Yifan**, born in 1998. Master. Her main research interests include 3D vision and point cloud segmentation.

卜一凡, 1998 年生. 硕士. 主要研究方向为 3D 视觉、点云分割.



**Xu Hongli**, born in 1963. PhD, professor. Her main research interests include machine learning and machine vision.

许宏丽, 1963 年生. 博士, 教授. 主要研究方向为机器学习、机器视觉.



**Wang Xiaorong**, born in 1999. Master candidate. Student member of CCF. His main research interests include 3D vision and point cloud segmentation.

王晓荣, 1999 年生. 硕士研究生. CCF 学生会员. 主要研究方向为 3D 视觉、点云分割.