文章编号:1007－6735(2026)04－0513－12　DOI: 10.13255/j.cnki.jusst.20250613001

融合几何特征的多模态点云分析网络

郁森林[1]，　魏国亮[2]，　简晓富[1]
（1. 上海理工大学 光电信息与计算机工程学院，上海 200093；2. 上海理工大学 管理学院，上海 200093）

摘要： 多模态特征学习在点云领域得到广泛关注，但网络的分类分割精度仍很大程度上取决于点云分析骨干网络的设计。现有骨干网络在下采样阶段未能有侧重地压缩点云数量，在特征提取阶段难以高效捕捉多层次几何结构特征。针对上述问题，创新性地提出了一个融合多层次几何结构特征的点云分析网络，并引入多模态对齐预训练策略。在下采样阶段，将传统局部结构特征与全局注意力特征相结合，用以筛选出具有高判别力的结构关键点，从而大幅剔除对 3D 理解贡献有限的冗余点；在特征提取阶段，设计了一个轻量化模块，实现邻域中多层次几何结构特征的提取与自适应融合，在参数量与特征丰富性之间实现了良好的平衡；在预训练阶段，引入文本、图像、点云多模态特征对齐预训练策略，进一步增强模型的表现。实验结果表明，网络在 ModelNet40 和 ScanObjectNN 数据集上的分类精度分别达到了 94.2% 和 87.9%，相较于经典和近期的点云分析网络均实现显著提升，在分割任务上也表现出了良好的性能。

关键词： 深度学习；点云下采样；多模态；点云分类；部件分割
中图分类号： TP 391　　**文献标志码：** A

Multimodal point cloud analysis network integrating geometric structural features

Yu Senlin[1],　Wei Guoliang[2],　Jian Xiaofu[1]
(1. *School of Optical−Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China*; 2. *Business School, University of Shanghai for Science and Technology, Shanghai 200093, China*)

Abstract: Multimodal feature learning has received widespread attention in the point cloud domain, but the ultimate accuracy of the network in downstream tasks still depends heavily on the design of the point cloud analysis backbone network. Existing backbone networks fail to focus on compressing the number of point clouds in the down sampling phase, and it is difficult to efficiently capture multilevel geometric

收稿日期: 2025−06−13
基金项目: 国家自然科学基金资助项目(62273239)；上海市“科技创新行动计划”国内科技合作项目(20015801100)
第一作者: 郁森林(1999—)，男，硕士研究生．研究方向：点云处理、点云分割．E-mail：233370879@st.usst.edu.cn
通信作者: 魏国亮(1973—)，男，教授．研究方向：大数据分析、多智能体协同控制．E-mail：guoliang.wei@usst.edu.cn

引文格式: 郁森林，魏国亮，简晓富．融合几何特征的多模态点云分析网络[J]．上海理工大学学报，2026，48(4)：513-524.
Citation: Yu Senlin, Wei Guoliang, Jian Xiaofu. Multimodal point cloud analysis network integrating geometric structural features[J]. Journal of University of Shanghai for Science and Technology, 2026, 48(4)：513-524.

structure features in the feature extraction phase. Therefore, to address the above problems, a point cloud analysis network that incorporated multilevel geometric structure features was proposed, and a multimodal alignment pre-training strategy was introduced. In the down sampling stage, traditional local structural features were combined with global attention features to filter out structural key points with high discriminative power, thus substantially eliminating redundant points with limited contribution to 3D understanding. In the feature extraction stage, a lightweight module was designed to realize the extraction and adaptive fusion of multilevel geometric structural features in the neighboring domain, which achieved a good balance between the number of parameters and the richness of the features. In the pre-training stage, a multimodal feature alignment pre-training strategy for text, image, and point cloud was introduced to further enhance the performance of the model. The experimental results show that the network achieves classification accuracies of 94.2% and 87.9% on ModelNet40 and ScanObjectNN datasets, respectively, which is a significant improvement over both classical and recent point cloud analysis networks, and also shows good performance in segmentation tasks.

Keywords: *deep learning*; *point cloud down sampling*; *multimodal*; *point cloud classification*; *part segmentation*

近年来，随着3D传感技术的迅速发展，获取海量三维点云数据变得十分便捷。这一趋势极大地推动了以点云为核心信息的下游应用(如自动驾驶、目标检测[1]、激光SLAM(simultaneous localization and mapping)[2])的落地。因此，如何对这些原始数据进行高效的分析与解读已成为关键。本文面向的点云分类与分割任务是实现场景感知与理解的基础任务，其分析方法的性能与效率至关重要。然而，在处理日益复杂的真实场景时，现有主流点云分析网络的设计瓶颈愈发凸显，问题主要集中在下采样和局部特征提取两大核心环节。这一瓶颈不仅限制了模型精度的提高，也严重阻碍了其在自动驾驶、机器人等对实时性与算力有严格要求的场景中的实际部署与应用。一方面，下采样策略未能有效区分点的结构重要性，常导致关键几何信息丢失而保留过多冗余；另一方面，特征提取模块为追求精度而日益复杂，导致参数量与计算开销激增，在参数效率和特征丰富度之间难以取得理想的平衡。因此，设计一个能在高效筛选几何结构关键点的同时，能以更低的参数提取丰富几何特征的骨干网络，是当前点云分析领域亟待解决的关键问题。

目前，针对不规则点云的处理主要有两种范式：基于网格的间接法和基于点的直接法。基于网格的方法，如Qi等[3]、Lang等[4]将不规则点云投影到规则网络(如柱状网络或体素网络)上，然而，间接法在映射过程中常伴随信息丢失，进而影响点云表示的质量。

为了能直接对点云数据进行处理，Qi等[5]开创性地设计了能直接分析无序点云而无需预处理的PointNet算法。该网络将输入数据逐点映射到高维空间，并使用对称函数聚合所有点特征，实现了置换不变的全局特征表示。之后，Qi等[6]提出了PointNet++算法，通过考虑局部信息进一步改进了原方法，从而提高了网络的表达能力和效率。PointNet++算法提出的由下采样、局部邻域构建、局部特征提取、全局特征聚合及解码模块构成的通用框架，启发了后续众多工作的开展。其中，大部分研究集中于设计复杂的特征提取模块来替代原先使用的多层感知机(multi-layer perceptron, MLP)，早期的工作如Wang等[7]提出的DGCNN(dynamic graph convolutional neural network)、Guo等[8]提出的PCT(point cloud transformer)，以及Lu等[9]提出的3DTCN(3D temporal convolutional network)，通过引入图卷积、自注意力等机制，极大地丰富了特征表达。这一发展方向在近年的顶级会议中仍有体现，如PointMamba[10]和Mamba3D[11]算法均创新性地引入了状态空间模型来高效捕获长距离依赖，通过其线性复杂度的序列建模能力，来高效捕获长距离依赖关系，成为继Transformer算法之后一个极具潜力的新方向。这些精心设计的特征提取方式

取得了不错的性能，但复杂的特征提取器也会带来模型参数量和时间复杂度的剧增。针对这一问题，Qian 等[12] 和 Zhang 等[13] 开始放弃复杂的特征提取方式。前者通过使用残差 MLP 模块和改进后的训练策略重新挖掘出了 PointNet++算法的巨大潜能，后者则提出了一种不需要任何可学习参数即可提取点云局部特征的方法。尽管这些方法已显著降低了模型的复杂度，但上述方法在局部特征提取时仅依赖空间点的位置关系，未能充分利用局部邻域点的其他几何结构特征。

相比特征提取模块的改进，点云分析中的下采样环节是一个研究较少的主题。目前，点云分析框架中普遍使用的仍然是基于距离的采样方式——最远点采样（farthest point sampling, FPS）[14]。这种采样方式可实现各个区域间点云数量的均匀下降，具有良好的鲁棒性。然而，FPS 在减少点云数量前，缺乏对每个点在物体结构中重要性的判断，忽略了不同区域的点对网络学习重要性存在差异，均匀下采样会导致后续网络在学习各区域特征时缺乏侧重。除了基于距离的采样方法外，随着神经网络的发展，还提出了几种基于网络的方法。其中：代表性的 SampleNet[15]、DA-Net[16] 算法利用多层感知机生成所需大小的新点云作为下采样结果；MOPS-Net[17] 算法利用学习到的下采样变化矩阵来对原始点云进行处理，从而降低点云数量。然而，这些方法都是基于生成式的，并且在采样时均未考虑将几何结构特征作为特殊特征来进行点的筛选。除此之外，近两年来将多模态预训练策略引入点云学习也成为了一大热门，如 ACT[18]、ULIP-2[19] 算法为代表的跨模态学习框架，通过融合语言、图像等多模态信息，在一定程度上提升了模型的表现。但下游性能的根本性突破仍取决于点云分析骨干网络的设计。

针对目前点云下采样和特征提取阶段出现的问题，本文提出了一个在下采样阶段和特征提取阶段均考虑几何结构特征的点云分析骨干网络 PointGeo（point geometry），并使用图像文本多模态预训练策略提高网络在下游任务上的表现。首先，在下采样阶段，设计了一个曲率−注意力协同采样模块，该模块可同时从全局和局部视角来分析点的重要性，将对 3D 理解任务重要性高的关键点保留下来，让后续网络在学习时有区域侧重。之后，在特征提取阶段，利用轻量化的特征提取模块来提取点云局部邻域不同层次的几何结构特征，并利用通道注意力实现特征自适应融合，在模型复杂度与特征表达力之间实现了平衡。通过从下采样到特征提取的优化改进，PointGeo 在点云分类和分割任务上均取得了不错的表现。

1　方法设计

1.1　网络总体框架

融合几何结构特征的点云分析网络 PointGeo 通过曲率−注意力协同采样（curvature-attention cooperative sampling，CACS）模块与多几何特征聚合（multivariate geometric aggregation，MGA）模块的共同作用，实现从关键点筛选到特征提取的优化。CACS 模块通过融合局部曲率与改进的全局注意力机制，筛选出既能体现局部几何细节，又能反映出全局语义重要性的关键点，有利于网络后续有侧重的学习不同区域的点云信息。MGA 模块则可学习邻域内多个不同层次的几何特征，并实现自适应融合，从而丰富单个空间邻域的特征表示。为了进一步提高模型表现，本文采取点云−文本−图像多模态特征对齐预训练策略，通过使用来自 3 种模态的特征进行预训练，让点云编码器学习图像、文本和 3D 点云的统一表示。

网络总体框架如图 1 所示。对于点云编码器部分（point encoder），3D 点云坐标首先进入嵌入层（embedding），映射为 32 维高维特征；接着，利用 CACS 模块和 K 近邻算法（K-nearest neighbors, KNN），实现点云数量的下降和局部点云邻域的构建；最后，利用 MGA 模块对各点云块进行多特征提取与聚合，聚合后的特征进一步送入由两个残差多层感知机（ResMLP）构成的深度提取模块以增强网络表达能力。对于一个分层的深度网络，与过往方法类似，递归的重复上述操作 4 次，最终将点云编码器的输出送到分类器或分割器以完成下游任务。

当使用预训练策略时，输入由点云（$\boldsymbol{P}_i$, $i = 1, 2, \cdots, N$）、图像（$\boldsymbol{I}_i$）、文本（$\boldsymbol{T}_i$）构成的三元组数据。图像编码器（image encoder）和文本编码器（text encoder）已预先在大规模图像−文本数据上完成训练，处于预先对齐的状态。预训练过程中，冻结文本和图像编码器参数，仅利用点云数据更新点云编码器，通过对比损失将点云编码器输出的

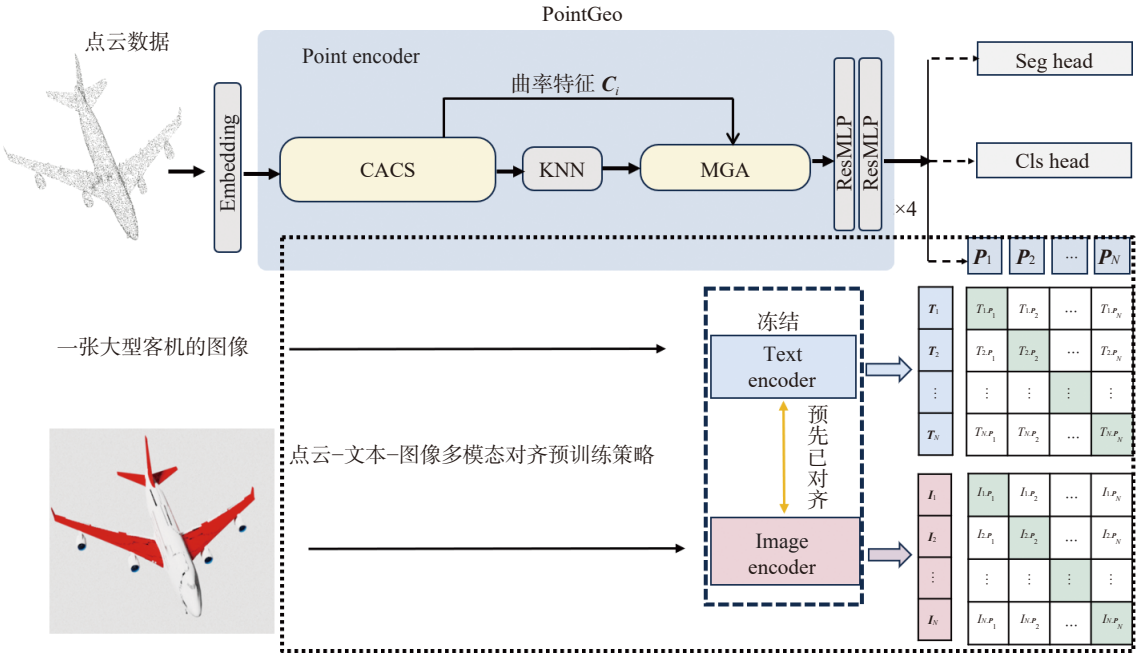

图 1 网络总体框架

Fig. 1 Overall framework of the net

3D 特征与对应的图像和文本特征逐步拉近距离，实现跨模态特征对齐，从而提高 PointGeo 算法的 3D 理解能力。

1.2 曲率−注意力协同采样模块

1.2.1 采样模块概述

下采样是点云分析框架中的关键，其核心是降低点云数量的同时保留具有判别性的关键点。然而，目前主流框架使用的最远点采样仅基于几何距离均匀降低点云数量，忽略点的结构重要性。为筛选出兼具高频几何细节表征能力与高层语义判别性的关键点，提出曲率−注意力协同采样（curvature-attention collaborative sampling, CACS）模块。CACS 模块被设计为网络层次化结构的一部分，网络的初始输入是从数据集中均匀采样获得的 1 024 个点。在后续的每个下采样阶段，CACS 模块所处理的点云数量 N 是逐层减半的。因此，CACS 模块在每一层实际处理的点云规模都远小于初始输入，确保了其计算开销在整个网络中是可控的。

首先，为每个点计算局部曲率得分 C_i 与全局注意力得分 A_i，随后相乘，并加噪得到综合评分 s_i^{score}，再用 TopK 函数选出得分最高的 M 个关键点。公式总结如下：

$$\boldsymbol{P}_{\text{sampled}} = \text{TopK}\left(\{s_i^{\text{score}}\}_{i=1}^{N}, M\right) \tag{1}$$

$$s_i^{\text{score}} = \widehat{C}_i\widehat{A}_i + \varepsilon \tag{2}$$

式中：$\widehat{C}_i$ 和 $\widehat{A}_i$ 为归一化后的局部曲率与全局注意力特征；N 为原点云数量；M 为采样后的目标点数，PointGeo 网络中 M 设置为 $N/2$；TopK 为筛选高得分函数；ε 为一个很小的随机噪声值；$\boldsymbol{P}_{\text{sampled}}$ 为被选中的关键点索引值。

通过上述操作可大幅减少对三维物体认知贡献较小的冗余点，保留关键点，引导网络有侧重地学习不同区域的特征。整体流程如图 2 所示。

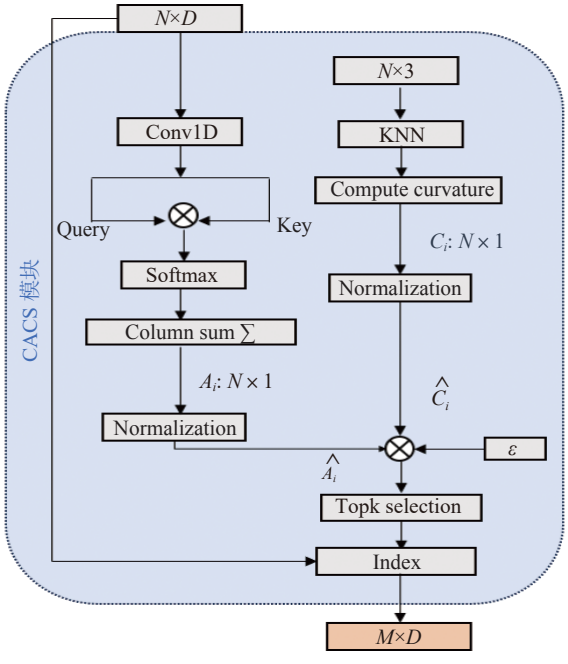

图 2 曲率−注意力协同采样流程

Fig. 2 Curvature-attention cooperative sampling pipeline

1.2.2　局部曲率得分计算

局部曲率反映了点云表面的几何突变特性（如边缘、角点等），为了量化局部曲率并将其作为筛选关键点的依据，本模块中设计了一个局部曲率得分的计算方法。具体而言，本文采用一种基于法向量空间分布的局部表面曲率估计方法。如图 3 所示，高曲率区域（如边缘点P_e）周围邻近点的法向量夹角差异显著，而低曲率区域（如平面点P_s）周围邻近点的夹角变化较小。因此，通过计算采样中心点与其 K 近邻点法向量夹角的平均值来近似表征局部曲率。对于点云中的每个点P_i，局部曲率得分C_i计算如下：

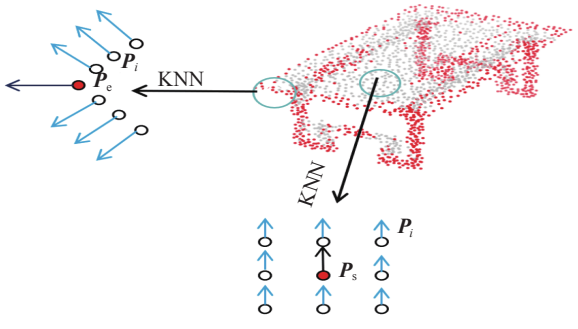

图 3　曲率与法向量的关系

Fig. 3　Relationship between curvature and normal vectors

$$C_i = \frac{1}{K}\sum_{j\in R_i}\arccos\left(\boldsymbol{n}_i\cdot\boldsymbol{n}_j\right) \tag{3}$$

式中：$\boldsymbol{n}_i$为中心点的法向量；R_i是以P_i为中心点利用 KNN 算法构建的K近邻点集；K是每个中心点周围近邻点的数量；$\boldsymbol{n}_j$为其邻近点的法向量。高曲率点的保留有助于网络捕捉物体的细节特征。

1.2.3　全局注意力得分计算与协同机制

若仅依赖曲率得分进行下采样，会出现采样点过度聚集，导致整体结构失衡。同时，曲率信息作为一个局部特征，无法构建点间的长程依赖关系，过度关注局部也会影响网络对物体宏观结构的感知能力。

为了解决上述仅使用曲率得分进行采样出现的信息局部化问题，CACS 模块中设计了一个经过视角转换的轻量级全局注意力机制，用于评估各点在整体结构中的重要性。具体流程为：舍弃传统注意力中的 Value 投影层，将已经过 embedding 层生成的所有点的高维特征$\boldsymbol{F}$通过独立卷积层生成 Query 特征$\boldsymbol{f}_q$ 和 Key 特征$\boldsymbol{f}_k$：

$$\boldsymbol{f}_q = \mathrm{Conv1D}(\boldsymbol{F}),\ \boldsymbol{f}_k = \mathrm{Conv1D}(\boldsymbol{F}) \tag{4}$$

基于此，计算所有点对间的原始注意力权重矩阵$\boldsymbol{A}$，并使用 Softmax 函数逐行进行归一化：

$$\boldsymbol{A} = \mathrm{Softmax}\left(\frac{\boldsymbol{f}_q\boldsymbol{f}_k^{\mathrm{T}}}{\sqrt{d}}\right) \tag{5}$$

式中，d为特征维度。

归一化后的$N\times N$矩阵$\boldsymbol{A}$中第h行第i列的元素记为a_{hi}：

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & & \vdots \\ a_{N1} & \cdots & a_N \end{pmatrix}_{N\times N} \tag{6}$$

与传统注意力机制通过行方向加权来更新特征不同，自注意力在此计算的目的是评估每个点在全局中的重要性。为此，本文转化分析视角，从列的方向分析矩阵$\boldsymbol{A}$，对于点i，如果它在多行中频繁出现较大的权重值，则说明该点作为关键点获得了其他点较高的注意力关注，是全局结构中值得重视的关键点。因此，计算列方向的总和，定义$A_i = \sum_{h=1}^{N} a_{hi}$，表示点$i$被其他点关注的总强度，将该值作为全局注意力得分来代表点在全局中的重要性。

每个点的C_i和A_i值分别反映局部几何重要性和全局结构重要性。为确保曲率与注意力得分的量级一致，需在融合前对二者先进行归一化处理得到$\widehat{C}_i$和$\widehat{A}_i$：

$$\widehat{C}_i = \frac{C_i - \min(\boldsymbol{C})}{\max(\boldsymbol{C}) - \min(\boldsymbol{C})},\ \widehat{A}_i = \frac{A_i - \min(\boldsymbol{A})}{\max(\boldsymbol{A}) - \min(\boldsymbol{A})} \tag{7}$$

归一化后的得分按式 (2) 计算得到最终的综合评分s_i^{score}，不同视角下的两个重要性得分通过乘积实现协同。这种乘法机制类似于一个“逻辑与”门，充当了一个协同过滤器。它确保了只有在局部几何细节与全局结构重要性两个视角上同时表现优异的点才能获得高分，即$\widehat{C}_i$和$\widehat{A}_i$的数值均要较大。这类点通常是物体的边缘点或关键连接处（如飞机的机翼与机身连接处）的点，它们既有丰富的几何细节，又对物体的整体结构至关重要，这是设计中最希望筛选保留的点。

同时，任何一个视角的低分都会显著拉低最终得分。例如，一个孤立的噪声点虽然可能因其突兀的位置而获得极高的局部曲率得分，但由于该点与其他结构性点相距甚远，在全局注意力计算中无法与它们建立强关联，这便体现为其对整体结构缺乏贡献，其全局注意力得分会非常低，从而有效避免了噪声点被错误地采样为关键点。

这一方法同时也有效提升了模型对噪声的鲁棒性。

在筛选策略上，本文采用 TopK 函数。其现代计算库实现（如 PyTorch）无需进行完全排序，能够在近线性时间内高效完成筛选。该函数线性复杂度降低了排序这一步骤对性能的影响。此外，s_i^{score}中加入随机噪声ε是为了保证当多个点的得分相同时，通过随机扰动确保 TopK 函数返回唯一且确定的索引。

综上，CACS 模块通过协同机制，在压缩点云数量的同时尽可能地保留高判别力的关键点，为后续特征提取提供兼具几何细节与全局语义的高质量输入，从而提升模型性能。

1.3 多几何特征聚合（MGA）

MGA 为网络特征提取的核心部件，该部件旨在从每个邻域中提取并融合丰富多样的几何信息。这个提取和融合的流程分为两个主要步骤：首先，并行地进行多层次几何特征提取，包括邻域位置几何特征$\boldsymbol{f}_{\text{n}}^{\text{g}}$、邻域局部形状特征$\boldsymbol{f}_{\text{n}}^{\text{l}}$、邻域平均曲率特征$\boldsymbol{f}_{\text{n}}^{\text{c}}$；然后，将这 3 类特征输入到多尺度自适应聚合（multi-scale adaptive aggregation，MAA）模块，利用通道注意力机制实现特征间的自适应加权与聚合，每个中心点的最终综合几何特征表示为$\boldsymbol{f}^{\text{g}}$。

1.3.1 多层次几何特征提取

首先在提取邻域位置几何特征$\boldsymbol{f}_{\text{n}}^{\text{g}}$时，使用与 Point-NN[13] 相类似的方式以便降低参数量，其流程可表达如下：

$$\boldsymbol{f}_{\text{n}}^{\text{g}} = \left(\text{Concat}\left(\boldsymbol{f}_{\text{c}}, \boldsymbol{f}_{\text{n},j}\right) + \text{PosE}\left(\triangle \boldsymbol{P}_{n,j}\right)\right) \odot \text{PosE}\left(\triangle \boldsymbol{P}_{\text{n},j}\right),\ j \in N_{\text{n}} \tag{8}$$

式中：$\boldsymbol{f}_{\text{c}}$和$\boldsymbol{f}_{\text{n},j}$分别代表中心点和其邻域点的特征；$\triangle \boldsymbol{P}_{\text{n},j}$表示经过平均值和标准差归一化后的邻域点坐标；PosE(·)表示每个中心点使用三角函数实现的位置编码操作。通过式 (8) 可将邻域点的相对位置信息隐式地编码到每个中心点的特征中，而过程中不需要任何可学习参数。

除了位置几何特征，本模块还选择了平均曲率特征$\boldsymbol{f}_{\text{n}}^{\text{c}}$和局部形状特征$\boldsymbol{f}_{\text{n}}^{\text{l}}$作为辅助几何结构特征来提升单个空间邻域的特征表示。前者能反映邻域整体的弯曲程度，从而捕捉到局部区域的凸起或凹陷等细节信息。后者可有效区分邻域点分布的线性、面性或散射性的不同形态特点。这两类特征可弥补仅依赖坐标位置所无法捕捉的几何信息。

具体而言，对于平均曲率特征$\boldsymbol{f}_{\text{n}}^{\text{c}}$，首先将采样阶段计算得到每个邻域点的曲率特征$C_i$通过 MLP 进行升维，以获得相邻点的高维曲率表示；随后，对这些高维特征进行平均池化，以反映局部区域整体的平均弯曲程度:

$$\boldsymbol{f}_{\text{n}}^{\text{c}} = \text{Meanpooling}(\text{MLP}(\boldsymbol{C}_i)) \tag{9}$$

而对于一组邻近点$\boldsymbol{P}_{\text{n},j}$，其局部形状特征$f_{\text{n}}^{\text{l}}$计算过程计算如下：

首先计算其协方差矩阵，即

$$\boldsymbol{M}_{\text{n}} = \frac{1}{N_{\text{n}}} \sum_{j=1}^{N_{\text{n}}} \boldsymbol{P}_{\text{n},j} \boldsymbol{P}_{\text{n},j}^{\text{T}} - \overline{\boldsymbol{P}}_{\text{n}} \overline{\boldsymbol{P}}_{\text{n}}^{\text{T}} \tag{10}$$

式中：$\boldsymbol{M}_{\text{n}}$表示邻域点数；$N_{\text{n}}$表示邻域内点云数量；$\overline{\boldsymbol{P}}_{\text{n}}$表示邻域的质心。

$$\overline{\boldsymbol{P}}_{\text{n}} = \frac{1}{N_{\text{n}}} \sum_{j=1}^{N_{\text{n}}} \boldsymbol{P}_{\text{n},j} \tag{11}$$

接下来使用 SVD 对协方差矩阵进行分解，得到 3 个特征值λ_1、λ_2、λ_3，$\lambda_3 \leqslant \lambda_2 \leqslant \lambda_1$。利用 3 个特征值可判断邻域的几何形态。具体而言，定义了 3 个尺度不变的形状描述符：线性度L_i、面性O_i和散射度S_i。对于某点i，当一特征值远远大于其余两特征值时，即$L_i = (\lambda_1 - \lambda_2)/\lambda_1 \approx 1$时，邻域内的点分布主要沿单一方向延伸，近似线性结构；当两个特征值接近且远大于第三个特征值时，即$O_i = (\lambda_2 - \lambda_3)/\lambda_1 \approx 1$，点云分布大体呈平面状，表现出面性结构；当 3 个特征值相差较小且接近时，即$S_i = \lambda_3/\lambda_1 \approx 1$时，邻域内的点分布较为散乱，呈现出较强的散射性。然而，仅使用上述 3 个尺度不变特征时，会丢失邻域的绝对尺度信息，为此，引入了一个尺度描述符$\sum\lambda$来补充这一信息。具体而言，利用协方差矩阵的迹来作为尺度描述符补充这一信息，即$\sum\lambda = \lambda_1 + \lambda_2 + \lambda_3$，作为每个邻域内点云分布的能量指标来反映该区域的尺寸大小。至此，将每个中心点尺度不变的形状描述符与尺度描述符进行拼接，构成一个完备的四维局部几何特征向量，如图 4 所示。该向量随后被送入 MLP 网络以升维成高维局部形状特征$\boldsymbol{f}_{\text{n}}^{\text{l}}$：

$$\boldsymbol{f}_{\text{n}}^{\text{l}} = \text{MLP}\left[\text{Concat}(L_i, O_i, S_i, \sum\lambda)\right] \tag{12}$$

1.3.2 多尺度自适应聚合（MAA）模块

为了聚合目前得到的 3 个不同层次特征，与

简单的特征相加（Add）或者与之前 Zhang 等[20]使用多个 MLP 聚合多个特征的方式不同，本文设计的 MAA 模块以轻量化的方式，按通道地学习不同特征的重要性。该模块的核心是引入了一个高效的通道注意力机制，为每一个通道动态生成一个 0～1 的权重，从而自适应关注信息丰富的通道，抑制噪声与冗余信息。该注意力机制通过一个参数量极小的瓶颈层 MLP 实现，在显著提升性能的同时，几乎不增加额外的模型计算负担，完美地平衡了性能与效率。

如图 4 所示，MAA 模块具体工作流程如下：首先将输入的 3 个几何特征沿通道维度进行拼接，形成一个聚合所有信息的特征描述符；其次，为了提取更全面的上下文，采用并行池化策略，同时对拼接后的特征进行全局平均池化和全局最大池化；最后，两个池化后的特征描述符被送入一个共享权重的瓶颈层 MLP 中。该 MLP 由两个全连接层构成，负责学习通道间的非线性依赖关系：第一个全连接层将特征维度先压缩到原先的 1/2，用于提取核心通道信息并降低计算复杂度；第二个全连接层再将维度恢复至原始大小，以便为每个通道生成对应的注意力权重。这种瓶颈设计在显著减少参数量的同时还有助于防止过拟合。

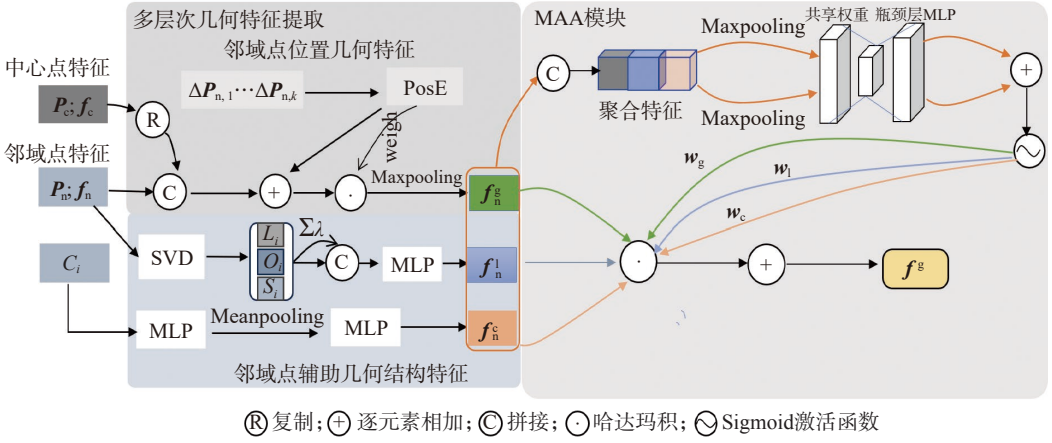

图 4　多几何特征聚合

Fig. 4　Aggregation of multi-geometric features

最终瓶颈层 MLP 的输出经逐元素相加和 Sigmoid 激活后，生成最终的注意力权重向量$\boldsymbol{w}$，将$\boldsymbol{w}$切分成$\boldsymbol{w}_{\mathrm{g}}$、$\boldsymbol{w}_{\mathrm{l}}$、$\boldsymbol{w}_{\mathrm{c}}$这 3 个部分，通过逐元素乘积的方式分别对原始的 3 个特征进行加权，得到融合后的综合几何特征$\boldsymbol{f}^{\mathrm{g}}$：

$$\boldsymbol{f}^{\mathrm{g}}=(\boldsymbol{w}_{\mathrm{g}}\odot\boldsymbol{f}_{\mathrm{n}}^{\mathrm{g}})+(\boldsymbol{w}_{\mathrm{l}}\odot\boldsymbol{f}_{\mathrm{n}}^{\mathrm{l}})+(\boldsymbol{w}_{\mathrm{c}}\odot\boldsymbol{f}_{\mathrm{n}}^{\mathrm{c}}) \tag{13}$$

如此，MAA 模块便完成了对多源几何信息的自适应加权与聚合。整个流程如图 4 所示。

1.4　点云−文本−图像多模态特征对齐预训练

为了进一步提升模型的表现能力，可引入多模态对齐预训练策略对 PointGeo 模型进行预训练，预训练时的输入数据为点云、图像、文本构成的三元组，通过同时对齐图像、文本和 3D 点云的特征，将 3 种模态映射到统一的特征空间，从而充分利用跨模态之间的互补信息，使骨干网络 PointGeo 学习到一种统一的表示。

具体而言，预训练的目标是将 PointGeo 点云编码器生成的特征，与已在大规模图文数据上预训练好、并被证明含有丰富通用语义的文本编码器和图像编码器所生成的特征进行对齐。借助这两个编码器中已对齐的语义信息，来有效提高点云编码器的 3D 理解和泛化能力。

在预训练过程中，本文采取冻结文本和图像编码器参数，仅更新点云编码器的方式。使用较为刚性的冻结策略是为了避免出现灾难性遗忘的现象。图像与文本编码器已在海量数据上完成预训练，具备了强大且泛化的语义表示能力，而用于 3D 预训练的数据集规模远小于前者。在该情况下，若微调图文编码器，多方向的进行优化，会导致图文编码器迅速遗忘已学到的通用知识，转而过拟合到小规模的 3D 数据上，这会损害其作为“教师模型”指导点云编码器学习的能力。

在上述冻结策略下，图像与文本间的对比损失可被忽略，即$L_{(I,T)}=0$，仅需要利用点云数据更新点云编码器的参数，通过最小化点云模态与图

像及文本模态之间的总对比损失L_t来实现对齐。总对比损失度L_t计算方法如下：

$$L_t = L_{(\boldsymbol{I},\boldsymbol{P})} + L_{(\boldsymbol{P},\boldsymbol{T})} \tag{14}$$

式中：$L_{(\boldsymbol{I},\boldsymbol{P})}$表示图像模态$\boldsymbol{I}$与点云模态$\boldsymbol{P}$之间的双向对比损失；$L_{(\boldsymbol{P},\boldsymbol{T})}$表示点云模态$\boldsymbol{P}$与文本模态$\boldsymbol{T}$之间的双向对比损失。其具体计算方法如下：

$$L_{(\mathrm{M}_1,\mathrm{M}_2)} = -\frac{1}{2}\sum_{(i,j)}\left(\log\frac{\exp\left(\boldsymbol{f}_i^{\mathrm{M}_1}\boldsymbol{f}_j^{\mathrm{M}_2}/\tau\right)}{\sum_k \exp\left(\boldsymbol{f}_i^{\mathrm{M}_1}\boldsymbol{f}_k^{\mathrm{M}_2}/\tau\right)} + \log\frac{\exp\left(\boldsymbol{f}_i^{\mathrm{M}_1}\boldsymbol{f}_j^{\mathrm{M}_2}/\tau\right)}{\sum_k \exp\left(\boldsymbol{f}_k^{\mathrm{M}_1}\boldsymbol{f}_j^{\mathrm{M}_2}/\tau\right)}\right) \tag{15}$$

式中：M_1和M_2表示参与对齐的任意两个模态；$\boldsymbol{f}_i^{\mathrm{M1}}$、$\boldsymbol{f}_j^{\mathrm{M2}}$表示来自两个模态的特征向量；$\tau$是一个可学习的温度参数，用于缩放相似度值。通过该预训练过程完成对齐后，得到的点云编码器PointGeo便可针对下游应用场景的不同进行微调，以释放其3D分析潜力。

2 实 验

2.1 实验数据集

对于3D分类任务，本文在ModelNet40和ScanObjectNN两个主流基准数据集上对PointGeo的性能进行了测试。ModelNet40是一个由干净CAD模型构成的标准数据集，共包含40个常见对象类别的12 311个模型。实验遵循其官方划分，使用9 843个模型进行训练，并在2 468个模型构成的测试集上进行评估。此外，为了评估模型在真实复杂场景下的鲁棒性，本文还在ScanObjectNN数据集中最具挑战性的变体PB_T50_RS上进行了测试。该数据集包含从真实场景中扫描得到的对象实例，分为15个类别，并带有严重的背景噪声、遮挡和旋转扰动。根据其官方划分，使用2 309个样本进行训练，581个样本进行测试。

对于部件分割任务，本文则采用了ShapeNetPart数据集进行评估。该数据集由16个大类的16 881个形状构成，并提供了50个部件标签；根据其标准划分，训练集包含14 007个形状，测试集则包含了2 874个形状。

2.2 实验细节

仅测试PointGeo算法的性能时，网络的模型训练和测试均在GeForce RTX 4090 GPU上完成。实验时，使用初始学习率为10^{-4}的AdamW优化器，并通过余弦退火策略将学习率衰减至10^{-8}，批次大小设置为32。对ModelNet40数据集训练300轮，而对于ScanObjectNN和ShapeNetPart数据集训练轮数为200轮。

验证多模态预训练策略的模型表现时，为了减少预训练时间，本文使用较小版本的ULIP-ShapeNet Triplets数据集来代替原先超过1 TB的完整数据集进行预先训练，训练时使用10^{-3}作为学习率，AdamW作为优化器。之后，训练好的点云编码器在分类任务上的微调实验环境配置与单独测试PointGeo时的设置相同。为了后文便于区分使用预训练策略模型的表现，使用预训练策略后的网络表现记为PointGeo+ULIP（point geometry+unified representation of language, images, and point clouds）。

2.3 ModelNet40数据集上的分类实验

本文采用类平均精度(mean accuracy, mAcc)与总体精度(overall accuracy, OA)作为分类效果评价指标，此外还比较了使用单模态特征的点云分析网络的可学习参数量。如表1所示，PointGeo在ModelNet40上达到了91.1%的mAcc和94.0%的OA，显著优于经典的PointNet和DGCNN模型。在与最新的研究方法比较中，PointGeo不仅在参数效率和总体精度上优于PointMamba、Mamba3D等先进的状态空间算法，并且与最新方法ResPOL和CrossAlignNet相比，也能在关键指标上达到与之

表1 ModelNet40数据集上的分类结果

Tab. 1 Classification results on ModelNet40

算法	mAcc%	OA%	参数量
PointNet[5]	86.2	89.2	3.5×10^4
DGCNN[7]	90.2	92.9	1.7×10^4
PointMamba[10]	—	93.6	12.3×10^4
Mamba3D[11]	91.8	93.4	16.9×10^4
ResPOL[21]	91.1	93.9	1.5×10^4
CrossAlignNet[22]	—	93.2	1.2×10^4
PointGeo	91.1	94.0	1.8×10^4
ACT[18]	—	93.6	—
PointNet++ + ULIP[23]	91.2	93.4	1.5×10^4
PointBERT + ULIP[23]	—	94.1	21.9×10^4
PointGeo+ULIP	92.2	94.2	1.8×10^4

相当的水平。值得注意的是，在仅使用单模态训练时，PointGeo 表现甚至优于使用多模态交叉学习的ACT 和PointNet++ULIP 算法，略逊于PointBERT+ULIP 算法。

进一步，当使用 ULIP-ShapeNet Triplets 数据集进行多模态预训练后，PointGeo+ULIP 在 OA 上虽仅提升了 0.2%，但在 mAcc 上获得了超过 1% 的显著提升，达到了表中各方法的最优性能。此外，图像与文本编码器仅在预训练阶段提供语义监督，在推理阶段均被移除，仅保留 PointGeo 骨干网独立工作。因此，模型实际部署的参数量仍然保持不变。为了更直观地展示 CACS 的效果，本文对 ModelNet40 上分类实验中前两轮下采样结果进行了可视化，如图 5 所示：灰色表示被舍弃的点，绿色则为被保留的关键点。相较于 FPS 这一传统的均匀下采样方式，CACS 更倾向于丢弃平面区域中对物体认知贡献较小的冗余点，优先保留那些具有几何细节（如边缘轮廓）和结构语义（如连接处）的关键点，从而为后续特征提取提供更具判别性的输入。

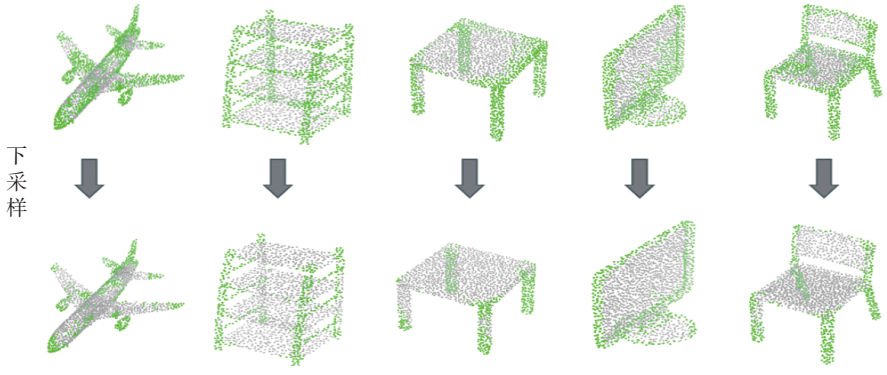

图 5　曲率–注意力协同下采样可视化

Fig. 5　Visualization of curvature-attention cooperative sampling

2.4　ScanObjectNN 数据集上的分类实验

尽管 PointGeo 已经在合成数据集 Modelnet40 上取得了优越表现，但 ModelNet40 是由干净 CAD 模型构成的合成数据集，不包含真实世界的噪声。为了更全面地评估本文方法在复杂真实场景下的鲁棒性，本文进一步在真实世界数据集 ScanObjectNN 上进行了测试，并与最近的方法进行了比较，结果如表 2 所示。在不使用预训练的情况下，PointGeo 的性能已优于其他单模态方法，从实验层面说明了 CACS 模块协同机制内在的噪声抑制能力，展现了其骨干网络设计的有效性。

当使用多模态进行特征对齐时，PointGeo 的性能得到了巨大提升，其OA（87.9%）和mAcc（86.6%）超越了同样使用预训练的 PointBERT+ULIP。同时，这一结果还表明，多模态预训练对 PointGeo 在 ScanObjectNN 上的性能增益（OA +2.5%, mAcc +2.8%）远大于在 ModelNet40 上的增益。这一差异主要来源于 ModelNet40 作为合成数据集，其理想 CAD 模型几何信息本身已足够丰富，因此，预训练带来的效益相对较小。而 ScanObjectNN 的真实世界数据常伴有噪声、遮挡与信息缺失，其不完美的几何结构使得模型更依赖于 ULIP 所提供的跨模态高级语义先验来进行补充。

表 2　在 ScanObjectNN 数据集上的分类结果

Tab. 2　Classification results on ScanObjectNN

算法	mAcc%	OA%
PointNet[5]	63.4	68.2
DGCNN[7]	—	78.1
SimpleView[24]	—	80.5±0.3
MaskPoints[25]	—	84.6
RepSurf-U[26]	—	84.3
CrossAlignNet[22]	—	84.5
PointGeo	83.8	85.4
PointBERT +ULIP[23]	—	86.4
PointGeo+ULIP	86.6	87.9

2.5　ShapeNetPart 数据集上的部件分割实验

除了点云分类任务，PointGeo 同样适用于点云分割。本文在 ShapeNetPart 数据集上进行了部件分割实验，评价指标与现有工作保持一致，包括实例的平均交并比（mean intersection over union per instance，mIoUI）和类别平均交并比（mean intersection over union per category，mIoUC），以及每个类别的 IoU 值。如表 3 所示，PointGeo 取得了

表 3 ShapeNetPart 数据集上消融实验结果

Tab. 3 Ablation study results on ShapeNetPart

模型	mIoUC	mIoUI	飞机	包	帽子	汽车	椅子	耳机	吉他	刀	灯	电脑	摩托	浴缸	手枪	火箭	滑板	桌子
PointNet[5]	80.4	83.7	83.4	78.7	82.5	74.9	89.6	73.0	91.5	85.9	80.8	95.3	63.2	93.0	81.2	57.9	72.8	80.6
DGCNN[7]	82.3	85.2	84.0	83.4	87.7	77.3	90.8	71.8	91.0	85.9	83.7	95.3	71.6	94.1	81.3	58.7	76.4	82.6
PointMLP[27]	**84.6**	86.1	83.5	83.4	87.5	**80.5**	90.3	78.2	**92.2**	88.1	82.6	**96.2**	77.5	**95.8**	85.4	64.6	83.3	84.3
PointBERT[28]	84.1	85.6	84.3	**84.8**	88.0	79.8	91.0	**81.7**	91.6	87.9	85.2	95.6	75.6	94.7	84.3	63.4	76.3	81.5
CrossAlignNet[22]	84.2	85.9	84.7	85.1	**89.0**	80.2	91.2	80.6	91.6	87.8	**85.3**	95.8	75.1	94.8	84.3	62.4	76.7	82.2
PointGeo	84.5	85.9	**84.9**	84.7	87.7	78.2	**92.1**	80.9	91.1	**88.2**	85.1	95.5	74.9	94.8	84.2	**66.5**	76.7	**85.7**

不错的表现：相较于经典网络 PointNet 和 DGCNN，其分割精度均有显著提升；与近期工作 PointBERT 相比，在 mIoUC 上提高了 0.4%， mIoUI 上提高了 0.3%；与 PointMLP 相比，PointGeo 在 mIoUC 指标上取得了相近表现。

从具体类别上看，由于 PointGeo 的核心模块专注于捕捉点云物体的几何突变，模型在飞机、椅子、刀等结构复杂、部件连接处曲率变化剧烈的类别上展现出极强的判别力，相较于 PointMLP 取得了显著提升，例如飞机的分割准确率提升了 1.4%。然而，对于汽车、吉他等主要由大面积平滑曲面构成的物体，其局部曲率特征趋于一致，导致 CACS 模块难以筛选出具有显著几何意义的关键点，同时 MGA 模块提取的辅助几何特征信息量也随之减少。这解释了为何 PointGeo 在这些平滑类别上的分割精度表现相对一般，略逊于表中最优性能。这种性能分布表明 PointGeo 更适合处理具有丰富几何细节的复杂 3D 物体。部分物体部件分割可视化效果可见图 6。

2.6 消融实验

为了验证本文提出的 CACS 和 MGA 模块的有效性，本文在 ModelNet40 数据集上进行了分类对比实验，结果如表 4 所示，实验分别评估了不同模块组合对网络分类性能的影响。首先，对比实验组 A 和 F 可以看出，在使用相同的多层次几何特征和聚合方式时，本文提出的 CACS 采样策略相较于传统的 FPS 采样取得了有效提升，证明了其在筛选高判别力关键点方面的优势。其次，对于 MGA 模块，对比实验组 B 与 C、D、F 的结果表明，平均曲率特征 $\boldsymbol{f}_{\mathrm{n}}^{\mathrm{c}}$ 和局部形状特征 $\boldsymbol{f}_{\mathrm{n}}^{\mathrm{l}}$ 作为辅助几何信息均能有效提升模型性能，且当同时引入两者时（实验组 F）分类精度最高，验证了多层次几何特征的互补性。最后，对比实验组 E 和 F，验证了 MAA 模块相较于简单的 Add 模块具有更优越的特征整合能力。

此外，针对多模态预训练阶段图像与文本编码器是否应使用过于刚性的冻结策略，本文进行了进一步的实验对比。将本文采用的冻结策略与允许参数更新的全参数微调策略（非冻结）进行了对比，记录了在 ModelNet40 数据集上预训练前 50 轮的零样本分类准确率变化情况，对比结果如图 7 所示，较为刚性的冻结策略在当前条件下展现出

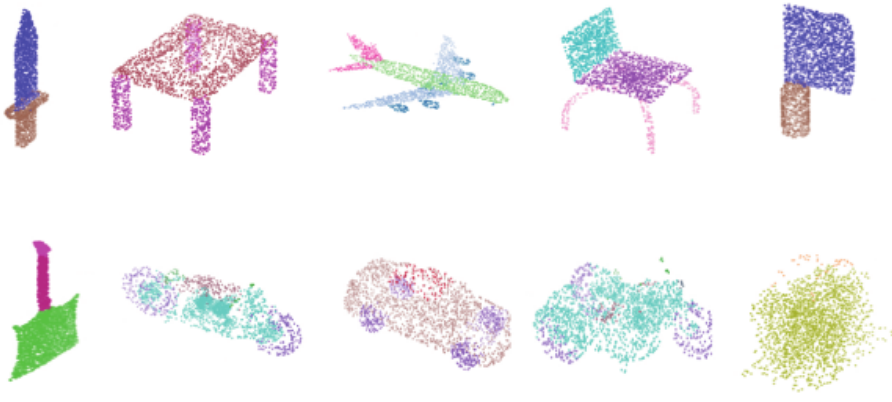

图 6 ShapeNetPart 数据集上点云分割可视化结果

Fig. 6 Visualization of part segmentation results on ShapeNetPart

表 4　ModelNet40 数据集上分类对比实验结果

Tab. 4　Classification comparison results on ModelNet40

实验组	采样方式	辅助几何特征	聚合方式	分类总体精度/%
A	FPS	$f_n^c+f_n^l$	MAA	93.65
B	CACS	None	None	93.81
C	CACS	f_n^c	MAA	93.89
D	CACS	f_n^l	MAA	93.86
E	CACS	$f_n^c+f_n^l$	Add	93.92
F(本文)	CACS	$f_n^c+f_n^l$	MAA	94.00

了更好的鲁棒性。具体而言，解冻策略的准确率始终在低位震荡，未能实现有效对齐；而冻结策略则稳步上升。造成这一差异的原因主要有两点：首先是数据量级悬殊，CLIP 的 4 亿预训练数据与 ShapeNet55 的 5 万组数据存在近 8 000 倍的规模差，小数据微调大模型极易引发灾难性遗忘；其次是计算资源的制约，解冻庞大的编码器显著增加了显存开销，限制了批次大小，进一步加剧了过拟合风险。因此，保留 CLIP 先验知识的冻结策略是当前实验条件下平衡计算效率与模型泛化性能的更优方案。

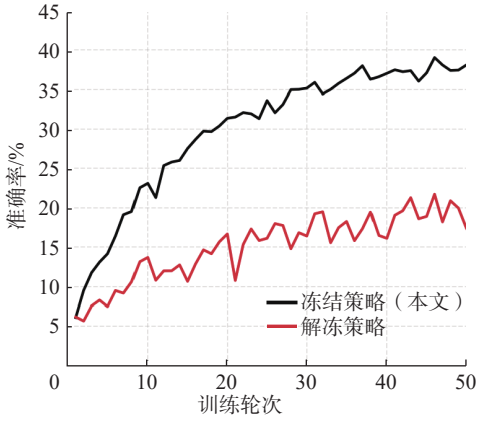

图 7　预训练阶段冻结策略与解冻策略的零样本分类准确率对比

Fig. 7　Comparison of zero-shot classification accuracy between frozen and unfrozen strategies in the pre-training stage

3　结束语

本文提出了一种支持多模态预训练策略优化的点云分析骨干网络，旨在保证几何特征表达能力的同时兼顾网络的轻量化与扩展性需求。在下采样阶段，本文提出的 CACS 模块综合考虑点的局部曲率和全局注意力响应，有效识别并保留结构关键点，显著减少了对 3D 语义理解影响较小的冗余点。在特征提取阶段，设计了一个高效的轻量级几何特征提取模块，利用通道注意力机制实现了多尺度结构特征的自适应融合，在保留细节的同时降低了模型复杂度，实现了特征表达的充分性与模型效率之间的良好平衡。在 ModelNet40、ScanObjectNN 和 ShapeNetPart 等多个数据集上的分类与分割实验均取得了良好表现，消融实验进一步验证了所提模块的有效性。未来工作将以本网络为基础，进一步挖掘点云中潜在的空间、语义与上下文结构信息。同时，也将从模型结构和学习策略角度优化网络在更大规模场景、动态点云和复杂干扰条件下的鲁棒性与泛化能力。

参考文献：

[1] 刘振威, 黄影平, 梁振明, 等. 基于点云图卷积神经网络的 3D 目标检测 [J]. 上海理工大学学报, 2024, 46(3): 320−330.

[2] 虞沈豪, 魏国亮, 张顺, 等. 融合多元信息的激光 SLAM 回环检测方法 [J]. 信息与控制, 2024, 53(5): 594−602.

[3] Qi C R, Litany O, He K M, et al. Deep hough voting for 3D object detection in point clouds[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 9277–9286.

[4] Lang A H, Vora S, Caesar H, et al. PointPillars: fast encoders for object detection from point clouds[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 12689–12697.

[5] Qi C R, Su H, Mo K, et al. PointNet: deep learning on point sets for 3D classification and segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 652–660.

[6] Qi C R, Yi L, Su H, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc. , 2017: 5105–5114.

[7] Wang Y, Sun Y B, Liu Z W, et al. Dynamic graph CNN for learning on point clouds[J]. ACM Transactions on Graphics (TOG), 2019, 38(5): 146.

[8] Guo M H, Cai J X, Liu Z N, et al. PCT: point cloud transformer[J]. Computational Visual Media, 2021, 7(2): 187−199.

[9] Lu D N, Xie Q, Gao K, et al. 3DCTN: 3D convolution-

transformer network for point cloud classification[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(12): 24854−24865.

[10] Liang D K, Zhou X, Xu W, et al. PointMamba: a simple state space model for point cloud analysis[C]//Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc. , 2024: 1026.

[11] Han X, Tang Y, Wang Z X, et al. M_{AMBA}3D: enhancing local feature for 3D point cloud analysis via state space model[C]//Proceedings of the 32nd ACM International Conference on Multimedia. Melbourne: ACM, 2024: 4995–5004.

[12] Qian G C, Li Y C, Peng H W, et al. PointNeXt: revisiting PointNet++ with improved training and scaling strategies[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans: Curran Associates Inc. , 2022: 1685.

[13] Zhang R R, Wang L H, Wang Y L, et al. Starting from non-parametric networks for 3D point cloud analysis[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 5344–5353.

[14] Moenning C, Dodgson N A. Fast marching farthest point sampling[R]. Cambridge: University of Cambridge Computer Laboratory, 2003: 1–2.

[15] Lang I, Manor A, Avidan S. SampleNet: differentiable point cloud sampling[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 7575–7585.

[16] Lin Y N, Huang Y, Zhou S H, et al. DA-Net: density-adaptive downsampling network for point cloud classification via end-to-end learning[C]//Proceedings of the 4th International Conference on Pattern Recognition and Artificial Intelligence (PRAI). Yibin: IEEE, 2021: 13–18.

[17] Qian Y, Hou J, Zhang Q, et al. MOPS-Net: a matrix optimization-driven network for task-oriented 3D point cloud downsampling[EB/OL]. (2021-04-12). https://doi.org/10.48550/arXiv.2005.00383

[18] Dong R P, Qi Z K, Zhang L F, et al. Autoencoders as cross-modal teachers: can pretrained 2D image transformers help 3D representation learning?[C]// Proceedings of the Eleventh International Conference on Learning Representations. Kigali: OpenReview. net, 2023: 1–17.

[19] Xue L, Yu N, Zhang S, et al. ULIP-2: towards scalable multimodal pre-training for 3D understanding[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024: 27081–27091.

[20] Zhang W W, Zhou H, Sun S Y, et al. Robust multi-modality multi-object tracking[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 2365–2374.

[21] Hazer A, Yildirim R. Hybrid method for point cloud classification[J]. IEEE Access, 2025, 13: 8825−8838.

[22] Wang F, Dong X Z, Wu J, et al. CrossAlignNet: a self-supervised feature learning framework for 3D point cloud understanding[J]. PeerJ Computer Science, 2025, 11: e3194.

[23] Xue L, Gao M F, Xing C, et al. ULIP: learning a unified representation of language, images, and point clouds for 3D understanding[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 1179–1189.

[24] Goyal A, Law H, Liu B W, et al. Revisiting point cloud shape classification with a simple and effective baseline[C]//Proceedings of the 38th International Conference on Machine Learning. PMLR, 2021: 3809–3820.

[25] Liu H T, Cai M, Lee Y J. Masked discrimination for self-supervised learning on point clouds[C]//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv: Springer, 2022: 657–675.

[26] Ran H X, Liu J, Wang C J. Surface representation for point clouds[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 18920–18930.

[27] Ma X, Qin C, You H X, et al. Rethinking network design and local geometry in point cloud: a simple residual MLP framework[EB/OL]. (2022-11-29). https://doi.org/10.48550/arXiv.2202.07123.

[28] Yu X M, Tang L L, Rao Y M, et al. Point-BERT: pre-training 3D point cloud transformers with masked point modeling[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 19291–19300.

（编辑：丁红艺）