

基于Mamba与循环最大池化的双流增强式点云分类网络

柴国强

上海理工大学光电信息与计算机工程学院, 上海

收稿日期: 2025年4月19日; 录用日期: 2025年5月12日; 发布日期: 2025年5月19日

摘要

点云分类作为三维视觉领域的核心任务, 面临特征表达能力有限与排列不变性处理不足的双重挑战。针对传统基于多层感知机(MLP)的网络难以有效捕捉全局特征及动态聚合局部信息的问题, 文章提出一种基于Mamba与循环最大池化的双流增强式点云分类网络。首先, 通过引入Mamba模块对原始点云进行序列化建模, 利用其长程依赖捕捉能力提取具有强区分性的全局特征; 其次, 加入循环最大池化(RMP)模块, 通过多级迭代的池化操作显式提取点云排列不变特征, 并结合循环机制实现局部特征的动态强化与上下文融合。双流架构中, 全局特征与局部特征经自适应加权后输入MLP分类头, 完成高阶语义推理。在ModelNet40与ScanObjectNN基准数据集上的实验表明, 本文方法的分类准确率分别达到93.9%与86.8%, 都高于先进的分类方法。消融实验进一步验证了Mamba的全局建模能力与RMP模块对无序点云的鲁棒性增强效果。

关键词

点云分类, 深度学习, 三维视觉, 特征提取

Dual-Stream Enhanced Point Cloud Classification Network Based on Mamba and Cyclic Maximum Pooling

Guoqiang Chai

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Apr. 19th, 2025; accepted: May 12th, 2025; published: May 19th, 2025

文章引用: 柴国强. 基于 Mamba 与循环最大池化的双流增强式点云分类网络[J]. 建模与仿真, 2025, 14(5): 503-515.
DOI: 10.12677/mos.2025.145410

Abstract

Point cloud classification, as a core task in the field of 3D vision, is faced with the dual challenges of limited feature expressiveness and insufficient handling of alignment invariance. Aiming at the problem that traditional multilayer perceptron (MLP)-based networks can hardly effectively capture global features and dynamically aggregate local information, this paper proposes a dual-stream augmented point cloud classification network based on Mamba and cyclic max-pooling. First, the original point cloud is serialized and modeled by introducing the Mamba module, and its long-range dependency capture ability is used to extract global features with strong discriminative properties; second, the Recurrent Maximum Pooling (RMP) module is added to explicitly extract the point cloud arrangement-invariant features through multi-level iterative pooling operations, and combined with the recurrent mechanism to achieve dynamic enhancement and contextual fusion of local features. In the dual-stream architecture, global features and local features are adaptively weighted and input into the MLP classification header to complete the higher-order semantic inference. Experiments on ModelNet40 and ScanObjectNN benchmark datasets show that the classification accuracy of this paper's method reaches 93.9% and 86.8%, respectively, both of which are higher than the state-of-the-art classification methods. The ablation experiments further validate the global modeling capability of Mamba with the robustness enhancement effect of the RMP module on disordered point clouds.

Keywords

Point Cloud Classification, Deep Learning, 3D Vision, Feature Extraction

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来随着激光雷达(LiDAR)和深度相机等三维传感技术的快速发展,点云数据的获取变得更加便捷和高效。点云数据以其能够精确表示三维空间中的几何结构和表面特性,广泛应用于自动驾驶、机器人导航、城市规划、文化遗产保护等领域[1][2]。然而,点云数据的稀疏性、无序性和非结构化特性为其处理和分析带来了巨大挑战。点云分类是三维计算机视觉和遥感领域中的一个重要研究热点,在点云分类任务中,如何有效地提取特征并实现高精度的分类成为了研究的核心问题。

传统的点云分类方法通常依赖于手工设计的特征,例如点云的几何特征、法向量、曲率等。这些方法虽然在特定场景下表现良好,但其泛化能力有限,难以应对复杂多变的实际环境。近年来,随着深度学习的兴起,基于神经网络的点云分类方法取得了显著进展。特别是 PointNet [3]的提出,首次直接将无序点云作为输入,通过对称函数(如最大池化)解决点云的无序性问题,并实现了端到端的分类和分割任务。此后,一系列改进方法如 PointNet++ [4]、DGCNN [5]和 PointCNN [6]等相继提出,通过引入局部特征聚合、图卷积和卷积操作等机制,进一步提升了点云分类的性能。

尽管深度学习在点云分类中取得了显著成果,但仍存在一些亟待解决的问题。首先,点云数据的稀疏性和不均匀分布导致特征提取的效率较低,尤其是在大规模场景中。其次,点云数据的噪声和遮挡问题对分类精度产生了负面影响。此外,现有方法在处理大规模点云时,由于特征提取差,往往分类结果精度低。因此,如何设计获取更加丰富的点云特征,仍然是当前研究的重要方向。本文工作的主要贡献

可以概括如下：

1) 将原始输入点云经过 Mamba 提取全局特征后得到的特征作为 MLP 的输入，MLP 通过多个全连接层对特征进行进一步处理，进而提高点云分类结果。

2) 通过引入循环最大池(RMP)模块，重复执行 max-pooling 操作以收集排列不变特征，然后循环使用这些特征来改进和提高原始模型的性能。

2. 相关研究

传统点云分类方法主要依赖于手工特征设计与传统机器学习算法，其流程包括特征提取、编码和分类器设计三阶段。虽然此类方法在特定场景下表现出有效性，但由于依赖人工设计的特征，难以捕捉复杂几何结构与深层语义信息，存在泛化能力不足的固有缺陷。深度学习技术的引入突破了传统方法的局限性，当前基于深度学习的点云分类方法可分为三类：基于变换技术的分类方法、直接处理原始点云的方法以及结合局部特征提取器的分类方法。

2.1. 点云变换技术

三维点云数据以 (x, y, z) 坐标为基础表征几何信息，常通过添加额外维度扩展特征表达。传统处理方法依赖手工设计的统计特征，其可分为对特定变换具有不变性的内在特征和外在特征两类。由于点云的无序性特征，研究者提出注意力机制实现有序向量转换，但存在几何信息保留不足的问题。典型处理方法是将不规则点云转换为规则三维体素网格后应用 CNN [7]进行特征提取。该方法面临计算复杂度高(体素分辨率立方增长)和数据稀疏性等核心挑战，虽通过分层处理优化计算效率，但难以应对大规模点云处理需求。另一方法 Multiview CNN [8]是将三维点云投影为多视角 2D 图像后使用 2D CNN 处理。该方法在形状分类任务中表现优异，但受限于投影过程中的几何信息损失，在点分类等精细任务中效果受限。此外，特征向量转换方法也可以将点云转换为特征向量后输入 DNN [9]进行分类。虽然简化了数据处理流程，但受限于特征工程的表达能力，难以适应复杂场景需求。

2.2. 原始点云处理方法

突破性工作：PointNet [3]首次实现了原始点云的端到端处理，其核心架构包含三个创新模块：对称函数(Maxpooling)保证置换不变性、仿射变换网络提升特征空间对齐、多层感知机实现特征抽象。后续改进的 PointNet++ [4]通过分层采样和 k 近邻算法构建局部感受野，递归应用 PointNet [3]实现多层次特征学习。最新进展：PointMLP [10]在保持架构轻量化的基础上，引入残差 MLP 模块和几何仿射变换，在推理速度与特征表达能力间取得显著平衡。此类方法避免了数据转换带来的信息损失，但局部特征交互效率仍有待提升。

2.3. 局部特征提取技术

当前研究重点聚焦于高效局部特征提取器的设计，现有方法主要围绕卷积运算、图结构建模和注意力机制三个方向展开，PointConv [11]通过多层感知机(MLP)动态生成连续卷积核，并建立密度函数实现点云的非规则空间卷积，显著提升了局部特征建模能力。BSC-Net [12]提出二进制稀疏卷积网络优化方案，通过筛选有效卷积子集降低量化误差，在不增加计算复杂度的情况下缩小了二值化网络与实值网络的性能差距。基于图的方法通过构建点云拓扑图实现邻域特征聚合。Wang 等人[13]提出基于球坐标系的图卷积网络(GCN)，通过姿态估计辅助网络预测旋转角度变化，有效提升了模型对刚性变换的鲁棒性。该方法将点云局部邻域映射至球坐标系，简化了复杂空间关系的计算过程。注意力机制通过动态权重分配强化关键特征提取，PCT [14]采用最远点采样策略构建局部区域，通过变换网络捕获多层次几何特征。LGR-Net [15]设计双分支 MLP 注意力模块，融合全局拓扑旋转不变特征与局部几何旋转不变特征，实现互补性特征增强。此

外 Point Transformer [16]通过自注意力机制兼顾局部结构与全局语义建模，具备强大的特征表达能力与对无序点云的适应性。其主要劣势在于计算开销大、结构复杂、对训练资源和点云密度分布较为敏感。

DMR 相较现有方法具有三大优势：一是引入 Mamba 实现全局序列建模，显著增强全局特征表达能力；二是循环最大池化(RMP)有效提取点云的排列不变特征，提升对无序输入的鲁棒性；三是双流架构融合全局与局部特征，通过自适应加权提升分类准确率与泛化能力。

3. 实验方法

3.1. 模型框架

DMR 模型框架如图 1 所示，是一个由两个子网组成的并行网络：密集网络和稀疏网络。DenseNet 是一个计算效率高的网络，并且可以从大量的点云数据中捕获关于输入对象的高级抽象。SparseNet 是一个稀疏网络，它从少量的点捕获对象的详细特征。输入的点云数据经过 Mamba 模块，利用状态空间模型(SSM)进行全局特征提取，捕捉点云中的长程依赖关系。随后，这些全局特征被传递到 PointMLP 模块，进行逐层的局部特征提取和深度处理。PointMLP 模块通过残差连接的多层感知机(MLP)结构，对每个点的特征进行非线性变换，捕捉特征之间的复杂关系。最终，经过多层级特征聚合输出全局表示，并通过分类器执行点云的分类任务。侧向连接分为中间连接和最终连接，中间连接可确保将从密集网络中大量点获取的信息传递到稀疏网络，最终连接使两个网络的输出维度是相同的，乘法和加法作为横向连接。该架构有效地结合了 Mamba 的全局信息捕获能力和 MLP 的局部特征提取优势，提升了点云分类的准确性和计算效率。

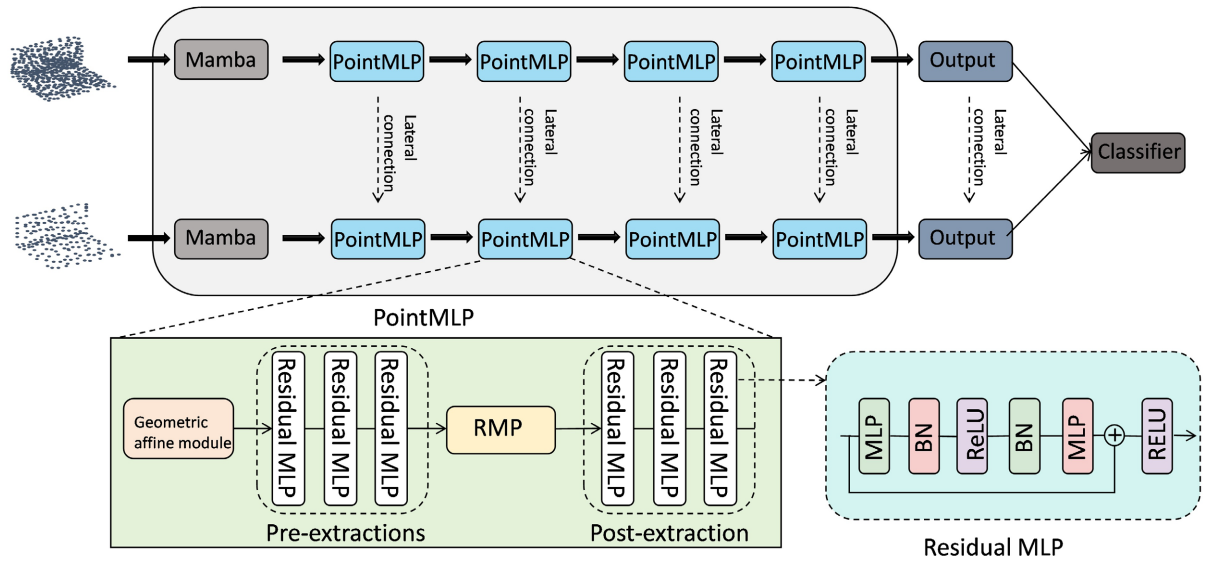


Figure 1. Overview of the DMR model
图 1. DMR 模型概览

3.2. 特征提取——Mamba 模块

在点云分类任务中，模型的性能高度依赖于特征提取和处理的有效性。提出了一种结合 Mamba 模块和多层感知机(MLP)的架构，以提升点云分类的准确性。Mamba [17]是一种基于状态空间模型(SSM)的架构，具有线性时间复杂度和强大的全局建模能力。与传统的基于点的方法相比，Mamba 在处理长程依赖关系时表现出色。其结构简洁、计算开销低，能够有效捕捉点云数据中的全局特征。在数据预处理后，输入首先经过 Layer Normalization (LN)进行归一化，然后进入线性变换层(Linear)。随后，通过深度可分离卷积

(DWConv)提取局部特征，并经过选择性状态空间模块(SSM)进一步提炼全局信息。在 SSM 处理后的特征通过乘法运算与旁路传输的特征融合，随后经过另一线性变换层，最终通过残差连接进行信息融合。此外，旁路路径包含一个额外的线性层，以确保输入和输出的特征维度匹配。在点云分类中，MLP 能够对 Mamba 提取的全局特征进行进一步处理，捕捉特征之间的复杂关系。通过共享权重的卷积实现，MLP 能够有效地对每个点进行相同的特征转换和高维映射。因此，将 Mamba 提取的全局特征输入到 MLP 中，能够充分发挥两者的优势。Mamba 负责捕捉点云数据的全局信息，而 MLP 则对这些信息进行深度处理，提取更具判别性的特征。这种组合使得模型能够更好地理解点云数据的复杂性，从而提高分类性能。

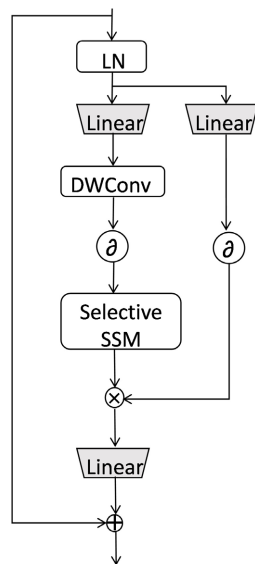


Figure 2. Mamba module diagram

图 2. Mamba 模块图

3.3. 特征加强——回收最大池化模块

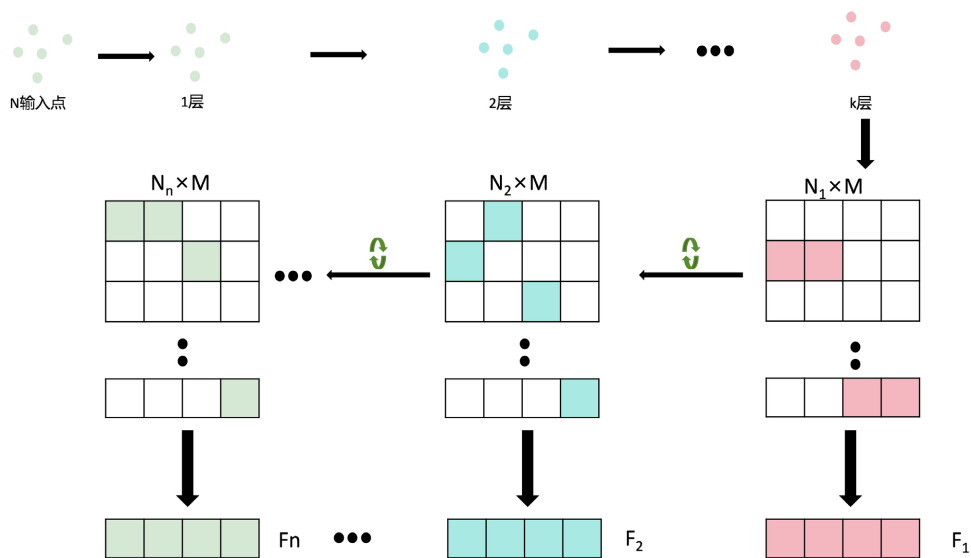


Figure 3. RMP module diagram

图 3. RMP 模块图

在点云分类任务中,最大池化(Max Pooling)操作的主要作用是通过减少数据的空间维度来提取特征,同时降低计算复杂度。最大池化通过从局部区域中选择最大值,帮助模型聚焦于最显著的,并提取出对分类更有意义的特征。在点云数据中,最大池化有助于减少信息量的同时保留重要的几何和结构特征。但传统最大池化可能丢弃一些有价值的信息(即局部区域的次最大特征),这些信息在某些情况下可能对最终分类结果有帮助。如图 3 所示,为了回收丢弃一些有价值的信息,采用 RMP [18]模块将点云分成 N 层, M 为维度。每层都进行特征获取,丢失的点进行回收再利用,从而获取更多的细节特征信息,为后续层提供了更丰富的特征,从而提升分类精度。RMP 模块通过设计一个分层损失函数来对丢失的点进行恢复,进而提高了点云特征的利用率。该损失函数可以表示为:

$$L_{RMP} = \sum_{i=1}^N \|F_i - F_{ji}\|_2^2 \quad (1)$$

其中 F_i 表示第 i 层特征信息, F_{ji} 表示回收的特征信息。

通过优化该损失函数,模型能够有效地恢复丢失的信息,并提高特征表达能力。传统最大池化的操作可能过于简化特征,忽视了不同特征之间的关系。通过回收丢失的点,RMP 能够从更广泛的角度捕捉点云的多样性,从而增强模型的泛化能力。由于 RMP 模块保留了更多的信息和细节,模型可以基于更丰富的特征进行点云分类,从而提高分类的准确度和鲁棒性。

3.4 损失函数

为了实现点云的准确分类,损失函数由交叉熵损失和 L2 正则化项联合构成,总损失函数 L_{total} 定义如下:

$$L_{total} = L_{CE} + \lambda \cdot L_{L2} + \alpha L_{RMP} \quad (2)$$

其中超参数 λ 控制正则化强度,其选取遵循以下原则:当 $\lambda \rightarrow 0$ 时,模型退化为纯交叉熵优化,易在训练集上过拟合,当 $\lambda \rightarrow \infty$ 时,权重过度收缩导致欠拟合,模型失去判别能力。 α 是 L_{RMP} 损失的权重超参数,用于控制 L_{RMP} 损失在总损失中的相对重要性。

交叉熵损失通过衡量预测概率分布与真实标签分布之间的差异,驱动模型学习判别性特征。其定义为:

$$L_{seg} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(p_{i,c}) \quad (3)$$

其中: N 为批次样本数量(Batch Size), C 表示类别总数(ModelNet40 中 $C = 40$, ShapeNetCore 中 $C = 55$)。 $y_{i,c} \in \{0, 1\}$ 为样本 i 在类别 c 上的真实标签, $p_{i,c} = \text{Softmax}(z_{i,c}) = \frac{e^{z_{i,c}}}{\sum_{k=1}^C e^{z_{i,k}}}$ 表示模型

输出的归一化概率, $e^{z_{i,c}}$ 为 MLP 最后一层的 logits 值。通过最大化正确类别的对数概率,迫使模型在特征空间中不同类别样本分离,相较于均方误差(MSE),交叉熵对标签噪声具有更强的容忍度,因其惩罚项与错误概率呈对数关系。

为防止模型在训练过程中对特定特征产生过强依赖(过拟合),引入 L2 正则化项对权重参数施加约束:

$$L_{L2} = \sum_{k=1}^K \|W_k\|_2^2 = \sum_{k=1}^K \sum_{i=1}^{d_{in}} \sum_{j=1}^{d_{out}} w_{i,j}^2 \quad (4)$$

其中, K 表示模型中可训练参数的总层数, $W_k \in R^{d_{in} \times d_{out}}$ 为第 k 层的权重矩阵, $\|\cdot\|_2$ 表示 Frobenius 范数。

4. 实验结果与分析

4.1. 数据集和评估指标

ModelNet40 [19]包含属于 40 个类别的 9843 个训练模型和 2468 个测试网格 CAD 模型。按照社区的标准实践，报告了测试集上的类准确率(Acc)。

ScanObjectNN [20]是一个基于真实世界扫描数据的点云分类数据集。该数据集包含约 15,000 个实物扫描对象，分为 15 个类别，涵盖 2902 个独特实例。数据集的特点是真实性和复杂性，包含了背景干扰、部分遮挡等真实场景中的挑战。此外，它还提供了全局和局部坐标、法线、颜色属性以及语义标签。

4.2. 实现细节

训练设置：在 ScanObjectNN 数据集上训练模型。在训练过程中，batch size 为 24，epoch 为 300，初始学习率为 0.001。在 ModelNet40 数据集上的训练，batch size 为 24，epoch 为 300，初始学习率为 0.001。

实验软件环境包括 UBUNTU 18.04, CUDA 11.3, pytorch 1.8.1 和 python 3.8。硬件环境包括 NVIDIA GeForce GTX 3060 和一个 Intel i7 CPU。

4.3. ScanObjectNN 的实验结果

在本节中，将与 ScanObjectNN 数据集上先进的方法进行比较。对比结果如表 1 所示，通过与经典分类模型进行对比。分类精度较高的模型通过增强特征交互(如 MLP 多层特征融合)来提升分类性能，如 PointMLP、DensePoint 等能有效提取邻域信息；而早期方法无法学习复杂局部形状，所以分类效果差。

DMR 分类准确率都优于先前模型。在 ScanObjectNN 数据集上进行评估，本文方法的 Accuracy 为 86.7%，高于之前的先进方法。对比经典模型 PointNet++，模型获取到了更多的特征细节，从而提高了后续的分类结果。PointMLP 和 DualMLP 的 Accuracy 分别为 $85.4\% \pm 0.3\%$ 和 86.4%。PointMLP 方法得益于 MLP 架构的高效性、多尺度特征融合能力及对点云本质特性(无序、稀疏)的适配，在减少计算冗余的同时提升了分类精度和推理速度，所以有较好的分类结果。DualMLP 能够全面且高效地处理点云数据中的复杂信息，显著提升分类精度并增强模型的泛化能力和鲁棒性，因此有较好的分类结果。

Table 1. Classification results for the ScanObjectNN dataset

表 1. ScanObjectNN 数据集的分类结果

Method	Parameter	Accuracy (%)
PointNet [3]	-	63.0
PointNet++ [4]	1.4 M	77.9
SpiderCNN [21]	-	68.2
PointCNN [6]	-	78.1
DGCNN [5]	-	78.1
DRNet [22]	-	80.3
PRANet [23]	-	82.1
MVTN [24]	-	82.8
PointMLP [10]	12.6 M	85.4 ± 0.3
DualMLP [25]	14.3 M	86.4
Our	15.1 M	86.7

在表 2 中分析了 DMR 与 DualMLP 在不同数量 SparseNet 输入点下的性能，并将 DenseNet 的点数固定为 1024。可以看到，对于 128 和 256，DMR 相对于 DualMLP 的改进分别为 0.20%和 0.37%。512 个输入点实现的改善量为 0.25%。因此，可以得出的结论是，DMR 在点云输入点数通常情况下，分类实验结果比 DualMLP 更好。

Table 2. Comparison of DMR accuracy (%) with existing works
表 2. DMR 与现有作品准确率(%)的比较

Points	DualMLP	Our	Improvement (%)
128	86.36	86.56	0.20
256	86.36	86.73	0.37
512	85.67	85.92	0.25

Table 3. Results of SparseNet and DenseNet on ScanObjectNN dataset with different number of points
表 3. 不同点数的 SparseNet 与 DenseNet 在 ScanObjectNN 数据集的结果

SparseNet	DenseNet	Accuracy (%)
64	512	82.71
64	1024	84.32
64	2056	85.12
128	512	82.78
128	1024	85.39
128	2056	86.73
256	512	83.57
256	1024	84.93
256	2056	86.71

在 ScanObjectNN 数据集上，通过设置不同点数的 SparseNet 与 DenseNet，得出结果如表 3 所示，可以看出 DenseNet 固定 2056 时，随着 SparseNet 点数的增加，分类精确率也有所上升。

4.4. ModelNet40 试验结果

为了验证模型的泛化能力和提供基准性能评估，在本节中，将与 ModelNet40 数据集上先进的方法进行比较。表 4 总结了不同输入 Input 在该数据集上的分类精度 Acc 和 Parameter 的实验结果。数据集可视化如图 4 所示。

表 4 通过与经典分类模型进行对比，发现分类精度较高的方法普遍具有更强的局部结构建模能力，如 DensePoint、DGCNN 等能有效提取边缘特征；而早期方法如 PointNet 只使用全局最大池化，无法学习复杂局部形状。PointFeatNet 在分类精度和总体点精度都优于先前模型。在 ModelNet40 数据集上进行评估得出，方法的 Accuracy 为 93.9%，高于之前的先进方法。DualMLP 的 Accuracy 为 93.7%，其得益于多层次特征提取、强大的特征融合能力和高效的非线性建模能力，所以有较好的分类结果。此外相比基于 Transformer 的方法，DMR 网络在保持全局建模能力的同时，结构更轻量、计算更高效。

对比结果如表 5 所示，在表 5 中分析了 DMR 与 DualMLP 在不同数量 SparseNet 输入点下的性能，并将 DenseNet 的点数固定为 1024。可以看到，对于 128 和 256，DMR 相对于 DualMLP 的改进分别为 0.36%和 0.31%。512 个输入点实现的改善量最低，为 0.19%。因此，可以得出的结论是，DMR 在点云输

入点数通常情况下，分类实验结果比 DualMLP 更好。

Table 4. Classification results for the ModelNet40 dataset

表 4. ModelNet40 数据集的分类结果

Method	Parameter	Input	Acc (%)
PointNet [3]	–	1 K	89.2
PointNet++ [4]	1.4 M	1 K	90.7
Improved PNet [26]	–	1 K	92.2
PointCNN [6]	–	1 K	92.5
PointConv [9]	18.6 M	1 K	92.5
KPConv [27]	15.2 M	7 K	92.9
DGCNN [5]	–	1 K	92.9
RS-CNN [28]	–	1 K	92.9
DensePoint [29]	–	1 K	93.2
PointASNL [30]	–	1 K	92.9
MLMSPT [31]	–	1 K	92.9
Point Transformer [16]	–	1 K	90.6
LGR-Net [15]	–	–	91.1
PointMixer [32]	3.9 M	–	91.4
Atten.EdgeConv [33]	–	1 k	91.5
PointMLP [10]	12.6 M	1 K	94.5
PointMLP (Re.Imp.)	12.6 M	1 K	92.9
DualMLP [25]	14.3 M	0.5 k/1 k	93.7
Ours	15.1 M	1 K	93.9

Table 5. Comparison of DMR accuracy (%) with existing works

表 5. DMR 与现有作品准确率(%)的比较

Points	DualMLP	Our	Improvement (%)
128	93.31	93.67	0.36
256	93.31	93.62	0.31
512	93.72	93.91	0.19

在 ModelNet40 数据集上，通过设置不同点数的 SparseNet 与 DenseNet，得出结果如表 6 所示，可以看出 DenseNet 固定 2056 时，随着 SparseNet 点数的增加，分类精确率也有所上升。

Table 6. Results of SparseNet and DenseNet on ModelNet40 dataset with different number of points

表 6. 不同点数的 SparseNet 与 DenseNet 在 ModelNet40 数据集的结果

SparseNet	DenseNet	Accuracy (%)
64	512	82.71
64	1024	84.32
64	2056	85.12
128	512	82.78

续表

128	1024	85.39
128	2056	86.73
256	512	83.57
256	1024	84.93
256	2056	86.71

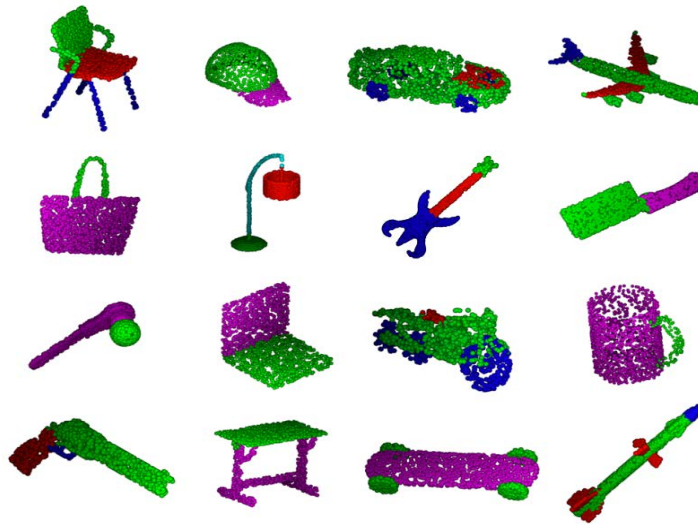


Figure 4. ModelNet40 dataset visualization
图 4. ModelNet40 数据集可视化

4.5. 消融研究

对 DMR 进行了一系列的消融实验。所有模型都是在数据集上进行训练，并在验证集上进行验证。不同配置模型的结果如表 7 所示。在表 7 中，方法 a 为 DualMLP 的分类结果，方法 b 是加入 Mamba 模块算法的结果，方法 c 是加入 RMP 模块的结果，方法 d 是结合 Mamba 模块与 RMP 模块共同的结果。

Table 7. Experimental results of different models
表 7. 不同模型实验结果

Method	Mamba 模块	RMP	ScanObjectNN 的 Acc (%)	ModelNet40 的 Acc (%)
a			86.36	93.72
b	√		86.52	93.84
c		√	86.57	93.85
d	√	√	86.71	93.92

不同模型配置的实验结果如图 5 所示，方法 b 增强了获取全局特征提取信息的能力，在两个数据集上分类精确率都有提升。方法 c 通过 RMP 模块利用循环迭代的池化策略，从而使模型获取更多的不同细节特征信息。方法 d 实验效果最好，相比方法 a，在 ScanObjectNN 数据集和 ModelNet40 数据集的分类准确率分别提升 0.35%和 0.2%。

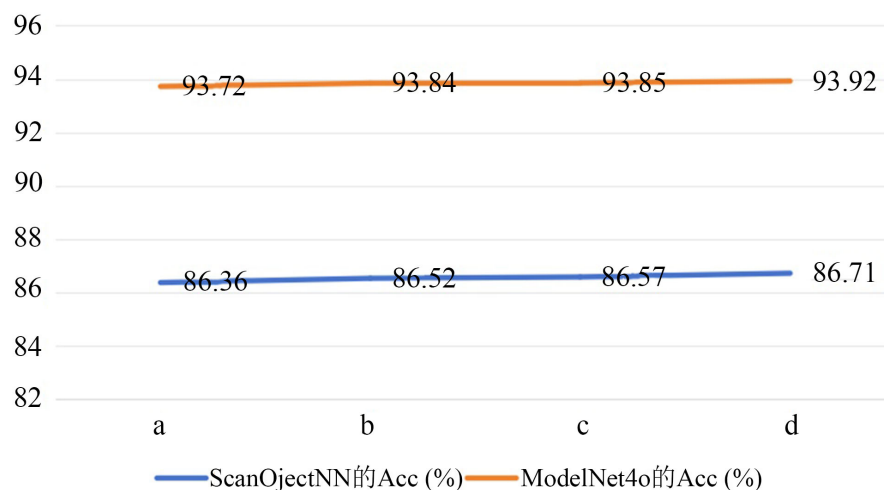


Figure 5. Experimental results for different model configurations

图 5. 不同模型配置的实验结果

5. 结论

本文提出了 DMR 网络，并提升了其在点云分类任务中的性能。通过加入了 Mamba 模块，以便于点云数据通过 Mamba 模块的状态空间模型提取全局特征并捕获长程依赖关系。将 PRM 模块取代 PointMLP 中的最大池化，其通过循环迭代的最大池化操作显式强化点云排列不变性，并基于动态上下文融合机制提升局部特征表征能力，从而显著增强模型对无序点云数据的分类鲁棒性。在 ScanObjectNN 数据集和 ModelNet40 上取得了令人满意的性能，验证了所提出方法在点云数据分类任务中的有效性。

DMR 模型在点云分类中展现出较强性能，但仍存在以下局限：细粒度特征区分能力不足，缺乏对点云中微小结构差异的精准捕捉，影响复杂场景下的分类精度；计算效率较低，动态模块的引入导致推理速度受限，难以满足实时任务需求。未来研究可以引入层级化注意力机制，通过通道与空间注意力联合优化，强化模型对关键局部特征的聚焦能力，提升细粒度分类性能；轻量化模型设计，结合动态稀疏激活、模型剪枝与量化技术，降低计算复杂度，同时采用知识蒸馏压缩模型规模，实现精度与效率的平衡。未来工作可基于上述优化，推动 DMR 在自动驾驶等实时场景中的实际应用。

参考文献

- [1] You, Y. (2023) Enhancing 3D Perception with Unlabeled Repeated Historical Data for Autonomous Vehicles. Ph.D. Thesis, Cornell University.
- [2] Kahn, G., Abbeel, P. and Levine, S. (2021) BADGR: An Autonomous Self-Supervised Learning-Based Navigation System. *IEEE Robotics and Automation Letters*, **6**, 1312-1319. <https://doi.org/10.1109/lra.2021.3057023>
- [3] Qi, C.R., Su, H., Mo, K. and Guibas, L.J. (2017) PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 77-85.
- [4] Qi, C.R., Yi, L., Su, H. and Guibas, L.J. (2017) PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. arXiv: 1706.02413.
- [5] Wu, B., Liu, Y., Lang, B. and Huang, L. (2018) DGCNN: Disordered Graph Convolutional Neural Network Based on the Gaussian Mixture Model. *Neurocomputing*, **321**, 346-356. <https://doi.org/10.1016/j.neucom.2018.09.008>
- [6] Li, Y., Bu, R., Sun, M., Wu, W., Di, X. and Chen, B. (2018) PointCNN: Convolution on X-Transformed Points. *Neural Information Processing Systems*. Curran Associates Inc.
- [7] O'Shea, K. and Nash, R. (2015) An Introduction to Convolutional Neural Networks. arXiv: 1511.08458.
- [8] Elhoseiny, M., Elgaaly, T., Bakry, A. and Elgammal, A. (2016) A Comparative Analysis and Study of Multiview CNN

- Models for Joint Object Categorization and Pose Estimation. *Proceedings of the 33rd International Conference on Machine Learning*, New York, 20-22 June 2016, 888-897.
- [9] Gong, M., Zhao, J., Liu, J., Miao, Q. and Jiao, L. (2016) Change Detection in Synthetic Aperture Radar Images Based on Deep Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, **27**, 125-138. <https://doi.org/10.1109/tnnls.2015.2435783>
 - [10] Ma, X., Qin, C., You, H., Ran, H. and Fu, Y. (2022) Rethinking Network Design and Local Geometry in Point Cloud: A Simple Residual MLP Framework. arXiv: 2202.07123.
 - [11] Wu, W., Qi, Z. and Fuxin, L. (2019) PointConv: Deep Convolutional Networks on 3D Point Clouds. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 9613-9622. <https://doi.org/10.1109/cvpr.2019.00985>
 - [12] Li, Y., Niu, Z., Sun, Q., Xiao, H. and Li, H. (2022) BSC-Net: Background Suppression Algorithm for Stray Lights in Star Images. *Remote Sensing*, **14**, Article 4852. <https://doi.org/10.3390/rs14194852>
 - [13] Zhang, S., Tong, H., Xu, J. and Maciejewski, R. (2019) Graph Convolutional Networks: A Comprehensive Review. *Computational Social Networks*, **6**, Article No. 11. <https://doi.org/10.1186/s40649-019-0069-y>
 - [14] Guo, M., Cai, J., Liu, Z., Mu, T., Martin, R.R. and Hu, S. (2021) PCT: Point Cloud Transformer. *Computational Visual Media*, **7**, 187-199. <https://doi.org/10.1007/s41095-021-0229-5>
 - [15] Xu, H., Yang, Y., Aviles-Rivero, A.I., Yang, G., Qin, J. and Zhu, L. (2024) LGRNet: Local-Global Reciprocal Network for Uterine Fibroid Segmentation in Ultrasound Videos. In: Linguraru, M.G., et al., Eds., *Medical Image Computing and Computer Assisted Intervention—MICCAI 2024*, Springer, 667-677. https://doi.org/10.1007/978-3-031-72083-3_62
 - [16] Zhao, H., Jiang, L., Jia, J., Torr, P. and Koltun, V. (2021) Point Transformer. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 16239-16248. <https://doi.org/10.1109/iccv48922.2021.01595>
 - [17] Gu, A. and Dao, T. (2023) Mamba: Linear-Time Sequence Modeling with Selective State Spaces. arXiv: 2312.00752.
 - [18] Chen, J., Kakillioglu, B., Ren, H. and Velipasalar, S. (2022) Why Discard If You Can Recycle? A Recycling Max Pooling Module for 3D Point Cloud Analysis. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 549-557. <https://doi.org/10.1109/cvpr52688.2022.00064>
 - [19] Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., & Tang, X., et al. (2015) 3D ShapeNets: A Deep Representation for Volumetric Shapes. arXiv: 1406.5670.
 - [20] Uy, M.A., Pham, Q., Hua, B., Nguyen, T. and Yeung, S. (2019) Revisiting Point Cloud Classification: A New Benchmark Dataset and Classification Model on Real-World Data. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27 October-2 November 2019, 1588-1597. <https://doi.org/10.1109/iccv.2019.00167>
 - [21] Xu, Y., Fan, T., Xu, M., Zeng, L. and Qiao, Y. (2018) SpiderCNN: Deep Learning on Point Sets with Parameterized Convolutional Filters. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision—ECCV 2018*, Springer, 90-105. https://doi.org/10.1007/978-3-030-01237-3_6
 - [22] Chang, Z., Gao, X., Li, N., Zhou, H. and Lu, Y. (2024) DRNet: Disentanglement and Recombination Network for Few-Shot Semantic Segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, **34**, 5560-5574. <https://doi.org/10.1109/tcsvt.2024.3358679>
 - [23] Cheng, S., Chen, X., He, X., Liu, Z. and Bai, X. (2021) PRA-Net: Point Relation-Aware Network for 3D Point Cloud Analysis. *IEEE Transactions on Image Processing*, **30**, 4436-4448. <https://doi.org/10.1109/tip.2021.3072214>
 - [24] Garg, M., Ghosh, D. and Pradhan, P.M. (2024) GestFormer: Multiscale Wavelet Pooling Transformer Network for Dynamic Hand Gesture Recognition. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, 17-18 June 2024, 2473-2483. <https://doi.org/10.1109/cvprw63382.2024.00254>
 - [25] Paul, S., Patterson, Z. and Bouguila, N. (2023) DualMLP: A Two-Stream Fusion Model for 3D Point Cloud Classification. *The Visual Computer*, **40**, 5435-5449. <https://doi.org/10.1007/s00371-023-03114-3>
 - [26] Paul, S., Patterson, Z. and Bouguila, N. (2022) Improved Training for 3D Point Cloud Classification. In: Krzyzak, A., Suen, C.Y., Torsello, A. and Nobile, N., Eds., *Structural, Syntactic, and Statistical Pattern Recognition*, Springer, 253-263. https://doi.org/10.1007/978-3-031-23028-8_26
 - [27] Thomas, H., Qi, C.R., Deschaud, J., Marcotegui, B., Goulette, F. and Guibas, L. (2019) KPConv: Flexible and Deformable Convolution for Point Clouds. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27 October-2 November 2019, 6410-6419. <https://doi.org/10.1109/iccv.2019.00651>
 - [28] Hu, L., Qin, M., Zhang, F., Du, Z. and Liu, R. (2020) RSCNN: A CNN-Based Method to Enhance Low-Light Remote-Sensing Images. *Remote Sensing*, **13**, Article 62. <https://doi.org/10.3390/rs13010062>
 - [29] Liu, Y., Fan, B., Meng, G., Lu, J., Xiang, S. and Pan, C. (2019) DensePoint: Learning Densely Contextual Representation for Efficient Point Cloud Processing. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27

-
- October-2 November 2019, 5238-5247. <https://doi.org/10.1109/iccv.2019.00534>
- [30] Yan, X., Zheng, C., Li, Z., Wang, S. and Cui, S. (2020) PointASNL: Robust Point Clouds Processing Using Nonlocal Neural Networks with Adaptive Sampling. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 5588-5597. <https://doi.org/10.1109/cvpr42600.2020.00563>
- [31] Han, X.F., Kuang, Y.J. and Xiao, G Q. (2021) Point Cloud Learning with Transformer. arXiv: 2104.13636.
- [32] Choe, J., Park, C., Rameau, F., Park, J. and Kweon, I.S. (2022) PointMixer: Mlp-Mixer for Point Cloud Understanding. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M. and Hassner, T., Eds., *Computer Vision—ECCV 2022*, Springer, 620-640. https://doi.org/10.1007/978-3-031-19812-0_36
- [33] Cui, Y., Liu, X., Liu, H., Zhang, J., Zare, A. and Fan, B. (2021) Geometric Attentional Dynamic Graph Convolutional Neural Networks for Point Cloud Analysis. *Neurocomputing*, **432**, 300-310. <https://doi.org/10.1016/j.neucom.2020.12.067>