DOI: 10.19650/j.cnki.cjsi.J2514151

# 基于全局体素特征交互的 3D 目标检测算法*

刘明杰,魏 宇,陈俊生,刘 平,朴昌浩
(重庆邮电大学自动化学院 重庆 400065)

**摘 要**:针对当前多数基于激光雷达的 3D 目标检测方法中因局部感受野限制无法建模特征远距离依赖,以及对点云数据的窗口划分策略导致的拓扑结构破坏等问题,提出了一种基于全局体素特征交互的 3D 点云目标检测网络。首先,设计基于希尔伯特空间曲线和 Mamba 的长距离上下文特征提取模块,通过对体素空间进行希尔伯特曲线序列化并保持体素间的空间局部性,利用 Mamba 处理长序列的优势提取具有长距离依赖的点云上下文特征,显著提升算法对长程依赖的建模能力。其次,设计基于特征图响应强度的自适应体素扩散模块,进行体素之间大规模的长程特征交互,通过动态生成扩散体素对目标中心体素的语义表达能力进行增强。此外,提出了一种空间特征恢复算子,通过子流形卷积的局部结构保持能力和 Mamba 的全局建模特性,对局部和全局特征表达进一步进行协同优化,用于补充序列化和体素聚合过程引入的信息损失。在 KITTI 数据集进行了实验,结果表明,方法达到了先进的 3D 目标检测性能,在汽车、行人和骑行者这 3 种类别的中等检测难度下精度分别达到了 82.36%、61.96%、66.05%,同时推理速度达到 19 fps,相比于基准模型,该方法较好地保持了精度和效率间的平衡。同时,在实际道路场景中进行了可视化对比分析,该方法表现出较强的泛化能力和实际应用潜力。

**关键词**: 目标检测;激光雷达;体素特征交互;Mamba 模型;全局上下文
**中图分类号**: TP391.4 TH865 **文献标识码**: A **国家标准学科分类代码**: 520.20

# Global voxel feature interaction-based 3D object detection

Liu Mingjie, Wei Yu, Chen Junsheng, Liu Ping, Piao Changhao
(*School of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China*)

**Abstract**: To address the inability to model the long-distance dependence of features due to the limitation of local receptive fields, and the destruction of topological structure caused by the window division strategy for point cloud data in most 3D object detection, this article proposes a global voxel feature interaction-based 3D object detection method. First, a long-range context feature extraction module based on the Hilbert space-filling curves and Mamba is designed. It employs Hilbert curve ordering to serialize the voxel space while preserving spatial locality among voxels, and leverages the capability of Mamba in processing long sequences to capture point cloud context features with long-range dependencies, significantly enhancing the ability to model global contextual relationships. Secondly, an adaptive voxel diffusion module based on feature map intensity is introduced, which facilitates large-scale long-range feature interactions between voxels by dynamically generating diffused voxels to enhance the semantic representation capacity of target center voxels. Furthermore, a spatial feature recovery operator is proposed to compensate for information loss during serialization and aggregation, leveraging the local structure preservation of submanifold convolution and the global modeling capability of Mamba to further synergistically optimize both local and global feature representations. Experiments on the KITTI dataset show that the method achieves state-of-the-art performance, with 82.36%, 61.96%, and 66.05% accuracy on the car, pedestrian, and cyclist classes at moderate difficulty, while maintaining a high inference speed of 19 frames per second (FPS). The proposed method represents a superior balance between accuracy and efficiency. In addition, by comparing our method with others in real road scenes intuitively. It demonstrates that the proposed method has strong generalization ability and practical application potential.

**Keywords**: object detection; LiDAR; voxel feature interaction; Mamba model; global context

收稿日期:2025-06-12 Received Date: 2025-06-12
*基金项目:国家重点研发计划项目(2022YFE0101000)、重庆市技术创新与应用发展专项重大项目(CSTB2023TIAD-STX0035)、重庆市教育委员会科学技术研究项目(KJQN202200630)资助

## 0 引　　言

基于激光雷达的点云目标检测方法能在各种复杂环境下捕获丰富的三维空间信息,因此在自动驾驶领域受到广泛关注。然而,点云稀疏、无序且不规则的特性[1-2]使得高效的 3D 点云目标检测成为挑战。

3D 点云目标检测方法主要包括两大类:基于点的处理方法和基于体素的处理方法。基于体素化的方法则能够将非结构化的点云体素化为规则的栅格数据,因此更适用于实时高效的大规模自动驾驶场景。其中,Zhou 等[3]提出的 VoxelNet 在三维空间中将输入点云划分为一个个体素,对划分的每个非空体素利用体素编码器进行局部特征提取。Yan 等[4]提出的 SECOND 网络针对 VoxelNet 使用的 3D 卷积算子处理稀疏体素效率过低的问题,设计了稀疏卷积和子流形卷积高效地提取局部特征。Zhang 等[5]利用层级化的编码一解码网络提取不同尺度的对象特征,进而增强 3D 目标检测的性能。但是,此类基于卷积神经网络(convolutional neural network, CNN)提取特征的方法感受野受限于卷积核尺寸大小,无法捕获全局上下文信息,因此对存在遮挡情况的点云数据无法准确有效地识别。基于 Transformer[6]的方法将点云体素进行序列化之后利用注意力机制捕获全局体素信息。其中,Mao 等[7]直接在稀疏体素空间部署 Transformer,通过局部注意力和稀疏全局注意力交替操作,在降低计算复杂度的同时保留空间结构信息。Li 等[8]进一步提出混合窗口注意力机制,将规则的体素划分为非重叠窗口,在窗口内执行局部注意力并跨窗口融合全局信息。Liu 等[9]将等窗口排序策略改进为等尺寸分组并灵活改变排序轴,借助移动窗口在不同分组之间交换特征,在分组内部使用自注意力机制从不同方向高效地聚合特征信息。Transformer 虽能在理论上建模点云的全局上下文关系,但注意力机制的二次计算复杂度导致模型计算成本随点云规模急剧增长。为了降低模型的计算复杂度,现有方法往往通过窗口划分或分组的策略将点云分割为局部子集进行处理。然而,这种窗口的划分割裂了稀疏点云的空间连续性,导致全局特征的获取不彻底,即使动态调整分组轴和移动窗口的策略也仅优化了局部注意力范围,难以建立跨分组的全局关联。此类方法本质上是通过牺牲全局性来缓解计算复杂度问题,基于窗口或分组内的自注意力仅能建模局部上下文,而跨窗口的间接交互无法充分捕捉原始点云中潜在的远距离依赖关系。综上所述,基于 Transformer 的注意力机制在点云场景下面临两难选择:若直接处理原始长序列点云,则二次计算复杂度难以承受;若采用分组策略,则全局特征建模能力被显著削弱。近期 Gu 等[10]提出的 Mamba 模型凭借其线性计算复杂度特性,在保持高效计算的同时突破了 Transformer 对长序列建模的局限性,其创新的选择性状态空间模型显著提升了全局依赖关系的捕捉能力,目前已在多个领域展现出强大的全局建模能力。

基于此,本文结合点云数据量大且离散的特点以及 Mamba 在长序列建模中的优势,在 3D 点云目标检测领域中引入 Mamba 模型,提出了一种基于全局体素特征交互的 3D 点云目标检测网络(global voxel feature interaction-based 3D object detection network, GVINet)。首先,对原始点云进行动态体素编码,以提高点云处理效率;接着,构建基于希尔伯特曲线和 Mamba 的长距离上下文特征提取模块(long-distance context feature extraction module based on Hilbert curve and Mamba, LCF-HM)提取点云的拓扑信息和长距离上下文特征,设计基于希尔伯特曲线的体素重排序算法(Hilbert-based 3D reordering, H3R),通过引入变体遍历路径生成策略,提升模块对跨区域长程关联特征的捕获能力,然后采用 Mamba 提取点云的几何信息和拓扑结构;进一步地,设计基于特征图强度响应的自适应体素扩散模块,用以丰富目标中心的体素特征信息;最后,利用检测头得到包含目标类别、位置和朝向的 3D 检测框。

本文的主要贡献可以概括为:

1) 设计了一个端到端的基于全局体素特征交互的 3D 点云目标检测网络,实现跨区域长程关联特征的捕获,并以线性时间复杂度建模点云的全局上下文特征。

2) 设计了一个面向点云的长距离上下文特征提取模块,利用希尔伯特曲线对体素进行重排序,建立体素间的长距离依赖关系,再通过 Mamba 捕获其几何关系和拓扑结构,提取具有长距离依赖的点云上下文特征。同时,设计了空间特征恢复算子用于缓解点云体素化和聚合过程中的信息损失。

3)设计了基于特征图强度响应的自适应体素扩散模块,利用该模块进行体素之间大规模的长程特征交互,通过生成扩散体素的特征对目标中心的体素特征信息进行补充。

## 1 相关工作

### 1.1 3D 目标检测

3D 目标检测是自动驾驶中重要任务之一,目前已有很多工作对其进行了深入的研究。本文将目前存在的方法分为两类:基于点的 3D 检测方法和基于体素化的 3D 检测方法。

基于点的 3D 检测方法直接针对原始点云的无序集合进行处理,以避免数据格式转换过程出现信息丢失。

其中,PointNet++[11]是该类方法的先驱,通过多层感知机(multilayer perceptron, MLP)提取逐点特征并利用最大池化将其聚合,但缺乏对局部几何关系的显式建模。Shi 等[12]提出两阶段检测框架,第 1 阶段利用 PointNet++提取原始点云特征并生成点级前景掩码,第 2 阶段设计点云区域池化对第 1 阶段生成的前景掩码进行优化,但其性能强烈依赖耗时严重的点云采样操作。Shi 等[13]进一步将图神经网络引入 3D 目标检测,通过构建点云图结构建模局部几何关系,但图邻域搜索与动态边更新导致模型的计算复杂度随点数指数增长。综上,基于点的方法对密集点云的处理与频繁的邻域搜索导致计算效率低下,且逐点交互难以建模遮挡目标的全局上下文关系。

基于体素化的检测方法将非结构化点云转换为规则的柱状或体素网格,再利用稀疏卷积提取 3D 特征,并生成 3D 目标候选框。Wang 等[14]通过引入类感知分组,结合语义预测和类别自适应的进行体素划分,提升了模型在复杂场景下 3D 目标检测的鲁棒性。Lang 等[15]将体素简化为柱状体,并对每个柱体的高度维度进行压缩生成伪图像,再利用 2D 卷积进行目标检测,在确保检测精度的同时提升了模型处理速度。Zhang 等[16]提出了一种完全稀疏的 3D 检测网络,通过动态调整特征扩散范围的方式,在保持稀疏性的同时有效缓解了中心特征缺失问题。然而,稀疏卷积的设计难以建模不规则点云的复杂拓扑结构,固定大小的卷积核也限制了提取特征感受野的范围。

Transformer 通过自注意力机制建模全局关系,为基于体素化的点云检测提供了新的思路。该类方法对体素通过窗口划分的形式进行特征提取,以此平衡模型的计算复杂度和特征信息提取。Wang 等[17]提出动态稀疏窗口注意力机制,通过并行计算窗口内的特征大幅度提升了模型的处理速度。王溪波等[18]提出了一种融合双边特征聚合和高维空间上下文增强的网络,采用注意力机制有效提取全局特征并滤除噪声。冯凯浩等[19]通过深度可分离边缘卷积在逐通道特征提取时保留局部几何信息,显著提升了通道间的区分能力。Liu 等[20]提出了一种可变网络注意力机制,将 3D 候选框划分为均匀网络作为初始采样点,并通过预测偏移量动态调整注意力区域,增强模型对局部关键特征的捕捉能力。此类方法固有的二次方计算复杂度制约长序列处理能力,而分组策略削弱全局建模效果。

### 1.2 Mamba 模型

Mamba 模型旨在解决传统 Transformer 模型在处理长序列数据时的计算效率问题。具体地,Mamba 模型以状态空间模型(state space models, SSM)为核心架构,一方面通过引入选择性机制动态生成状态转移矩阵,实现对特征全局上下文依赖关系的捕获;另一方面通过设计基于硬件感知的高效扫描算法,实现线性时间复杂度的模型推理速度。该结构目前已在图像检测和点云分析等领域展现出强大的全局建模能力。Liu 等[21]将 Mamba 引入 2D 图像检测领域,通过交叉扫描模块解决图像数据的非因果性和方向敏感问题,在保持线性计算复杂度的同时增强了模型的全局感受野。Liang 等[22]将 Mamba 用于点云分析,提出了一种简单的排序策略用于室内点云数据处理,随后基于 Mamba 的点云分析方法被提出。Han 等[23]提出双向 SSM,通过翻转特征通道缓解点云无序性引发的顺序依赖问题;Zhang 等[24]提出一致遍历序列化方法,通过多视角序列化策略增强空间特征捕捉,将无序的点云转换为有序的序列。

借鉴上述思想,本文以 Mamba 为核心架构,通过引入希尔伯特曲线设计体素路径生成策略,提出了一种基于全局体素特征交互的 3D 点云目标检测网络 GVINet,实现高效的基于激光雷达的 3D 目标检测。

## 2 理论分析

### 2.1 网络结构

本文提出的 GVINet 网络架构如图 1 所示,其核心流程包含动态体素编码、三维特征提取、鸟瞰图(bird's eye view, BEV)特征映射与目标预测这 4 个阶段。

首先,将输入点云 $\boldsymbol{P} \in \mathrm{R}^{N\times C_0}$ 通过动态体素编码转换为规则三维体素 $\boldsymbol{V} \in \mathrm{R}^{H\times W\times D\times C_1}$;接着,将三维体素输入 3D 主干网络进行分层多尺度特征提取:1)首先通过本文设计的基于希尔伯特曲线和 Mamba 的长距离上下文特征提取模块,对三维体素进行排列重组,并提取其全局上下文信息;2)利用基于特征响应强度的自适应体素扩散模块对目标中心的特征进行补充;3)沿高度方向进行体素聚合,通过索引映射保留非空体素并压缩空间维度形成多尺度特征金字塔;4)在多尺度特征金字塔中嵌入空间特征恢复算子优化局部和全局特征表达后输出三维特征 $\boldsymbol{F}_{3\mathrm{d}} \in \mathrm{R}^{H\times W\times \frac{D}{2^N}\times C_1}$;然后,将 $\boldsymbol{F}_{3\mathrm{d}}$ 投影至 BEV 空间,通过 2D 残差网络提取 BEV 特征 $\boldsymbol{F}_{2\mathrm{d}} \in \mathrm{R}^{H\times W\times C_2}$;最终,由检测头输出目标类别、位置、尺寸及朝向参数,实现针对点云的 3D 目标检测。

### 2.2 基于希尔伯特曲线和 Mamba 的长距离上下文特征提取模块

LCF-HM 模块主要包括 H3R 和 Mamba 模块,具体结构如图 2 所示。H3R 将体素进行一维序列化并保持原有的空间邻近性;Mamba 模块则以基于 H3R 构建的一维序列为输入,获取点云的几何信息和拓扑结构。

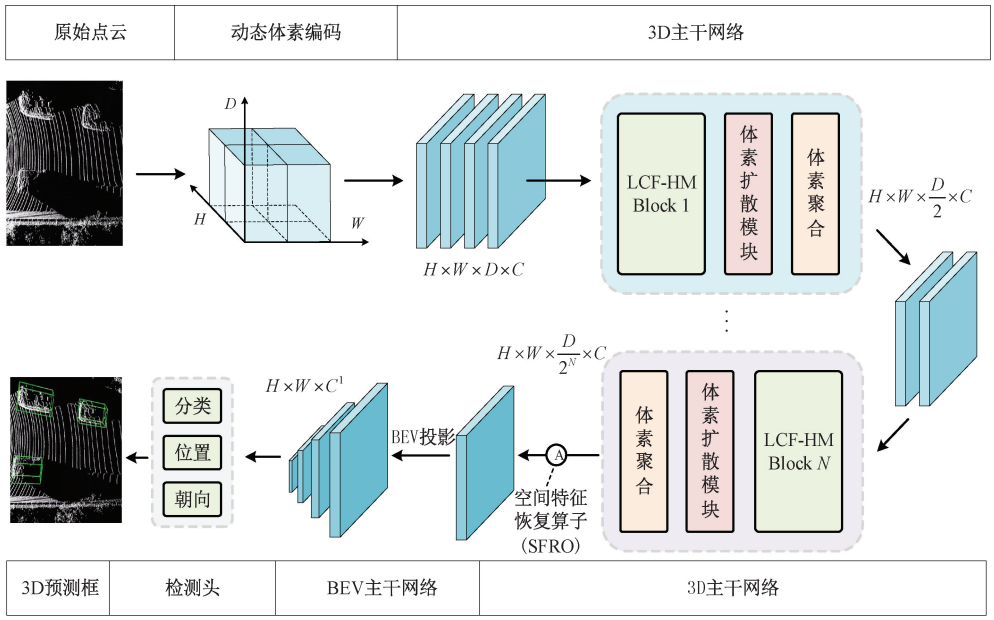



图 1　GVINet 检测网络模型整体架构

Fig. 1　Overall architecture of GVINet

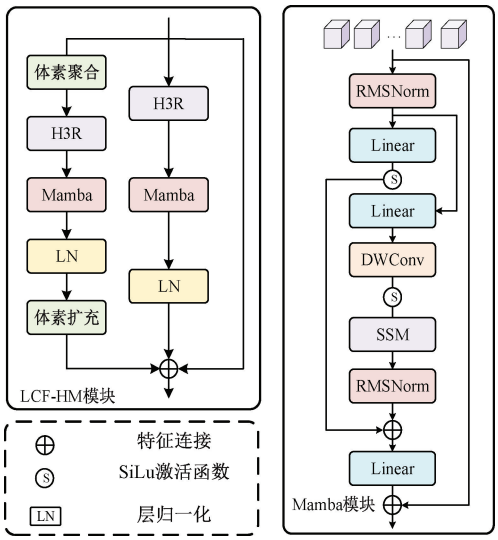



图 2　LCF-HM 模块和 Mamba 模块

Fig. 2　LCF-HM block and Mamba block

针对 Mamba 对网络输入序列化的要求,基于点云体素化的 3D 目标检测网络常采用滑动窗口分组的方式将非结构化的点云体素化为规则的栅格数据。然而,如图 3(a)所示,这种点云的分组遍历方式破坏了相邻体素的空间连续性及其局部拓扑结构,导致不同窗口之间特征交互缺失,削弱了模型对目标细节的推理能力。相比之下,如图 3(b)所示的希尔伯特(Hilbert)曲线能在遍历体素空间的同时保持相邻体素在一维序列上的邻近性。因此,本文提出基于 Hilbert 空间填充曲线的体素重排序算法 H3R,其将非空体素按照特定顺序进行重组用于输入 LCF-HM 模块,并通过引入变体遍历路径生成策略增强对跨区域长程关联特征的捕获能力。

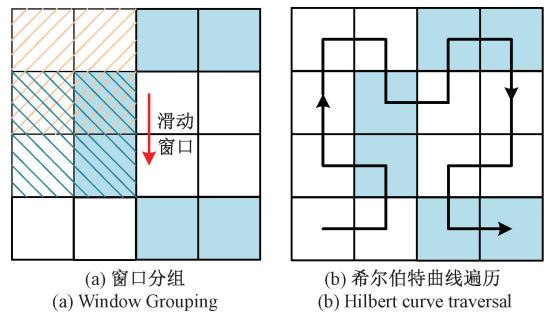



图 3　处理体素的不同方法

Fig. 3　Different methods to process voxels

Hilbert 曲线可表示为一个双射函数 $\phi:\mathrm{R}\rightarrow\mathrm{R}^3$,其中 $\mathrm{R}^3$ 表示体素坐标空间。由于 Hilbert 曲线的双射特性,存在一个逆映射 $\phi^{-1}:\mathrm{R}^3\rightarrow\mathrm{R}$,可以将三维体素坐标 $\boldsymbol{v}_i\in\mathrm{R}^3$ 映射为一个反映曲线位置信息的序列编码 $\phi^{-1}(\boldsymbol{v}_i)$。H3R 通过设计序列化的编码策略将体素的位置转换为在空间填充曲线中顺序的整数,具体地,对于体素 $\boldsymbol{V}=\{(x,y,z)\in\mathrm{R}^3\mid 0\leqslant x,y,z\leqslant 2^k-1\}$,其中 $k$ 为空间分辨率参数。首先,通过逐位模 2 运算将各维度坐标转换为 $k$ 位二进制比特串;然后,对比特串的最高有效位至最低有效位进行分层重组,前序比特为 0 时当前层执行交换操作,否则进行翻转。基于此,将所有比特位级联重组为:

$$\boldsymbol{B}=(x_k,y_k,z_k,x_{k-1},y_{k-1},z_{k-1},\cdots,x_0,y_0,z_0)\tag{1}$$

并对其执行格雷解码转换为 Hilbert 索引，即：

$$h = \sum_{j=0}^{3k-1} \left( b_j \oplus \left\lfloor \frac{1}{2} \sum_{m=j+1}^{3k-1} b_m \right\rfloor \text{bmod} 2 \right) \times 2^j \tag{2}$$

式中：⊕表示异或运算；⌊ ⌋表示向下取整。至此，基于 H3R 算法将三维体素坐标映射为 Hilbert 索引，将非空体素重组为一个特定且连续的一维序列。该序列可作为 LCF-HM 模块中 Mamba 的输入。

值得一提的是，为了提升运算效率，本文构建了离线坐标-索引映射表 $\tau: \mathrm{R}_k^3 \rightarrow \mathrm{R}_{3k}$，预先记录所有潜在体素坐标在 Hilbert 曲线上对应的索引位置，在模型推理过程中直接读取避免重复计算。

Mamba 模块以基于 H3R 构建的一维序列为输入，获取点云的几何信息和拓扑结构。具体地，首先利用均方根层归一化（root mean square normalization，RMSNorm）将输入的一维序列进行归一化处理；然后采用线性层调整其特征维度，并通过深度卷积（depthwise convolution，DWConv）提取局部特征；随后将其输入 SSM 层进行序列建模，捕捉特征间的长程依赖关系；最后再次通过 RMSNorm 与线性层输出结果。其中，SSM 作为 Mamba 的核心计算单元，其利用动态参数融合前一时刻位置 $t-1$ 累计加权的隐式全局信息 $h_{t-1}$ 前输入的局部特征 $x_t$，输出当前位置 $t$ 融合长程依赖的特征。其递归状态允许信息在图像序列块中流动，进而可以捕获任意距离的特征依赖关系，克服了传统 CNN 局部感受野的限制。在连续状态下，通常利用一组一阶微分方程实现序列 $x(t) \rightarrow y(t)$ 的映射，即：

$$\dot{h}(t) = \boldsymbol{A}h(t) + \boldsymbol{B}x(t) \tag{3}$$

$$y(t) = \boldsymbol{C}h(t) \tag{4}$$

式中：$h(t)$ 为隐式全局信息；$\dot{h}(t)$ 为 $h(t)$ 的微分；$\boldsymbol{A} \in \mathrm{R}^{N\times N}$；$\boldsymbol{B} \in \mathrm{R}^{N\times 1}$；$\boldsymbol{C} \in \mathrm{R}^{1\times N}$ 表示可学习参数。然而，针对点云序列的处理，需要将连续模型离散化。定义一个时间周期 $\Delta$，通过零阶保持器将动态参数 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 离散化为 $\bar{\boldsymbol{A}} = e^{\Delta \boldsymbol{A}}$，$\bar{\boldsymbol{B}} = (\Delta \boldsymbol{A})^{-1}(e^{\Delta \boldsymbol{A}} - \boldsymbol{I})\Delta \boldsymbol{B}$，进而得到离散 SSM，即：

$$h_t = \bar{\boldsymbol{A}}h_{t-1} + \bar{\boldsymbol{B}}x_t \tag{5}$$

$$y_t = \boldsymbol{C}h_t \tag{6}$$

基于此，可实现从输入点云序列到特征序列的映射，获取具有全局上下文感知的点云特征表示。

整体来说，LCF-HM 模块由 3 条分路构成，分路 1 直接使用 H3R 对输入体素进行一维序列化，然后将体素序列输入 Mamba 进行全局上下文信息提取，其输出特征为：

$$\boldsymbol{F}_{HM} = LN(Mamba(H3R(\boldsymbol{F}_v))) \tag{7}$$

分路 2 则是在分路 1 的基础上进行了体素聚合和体素扩充处理，其目的是为了确保在不增加模型计算复杂度的同时，实现多尺度特征的提取。其中，体素聚合的作用是降低输入数据的分辨率并获取多尺度特征；体素扩充则是用来恢复特征分辨率以便于后续多分路特征融合。具体实现如式（8）所示。

$$\boldsymbol{F}_{HM}^{low} = Up(LN(Mamba(H3R(Down(\boldsymbol{F}_v))))) \tag{8}$$

其中，*Down* 和 *Up* 分别表示体素聚合和体素扩充。由于体素数据高度稀疏的特性，不能简单地使用池化操作进行下采样和上采样，本文通过记录非空体素索引对其执行下采样实现体素聚合，对索引使用反转映射实现体素扩充。

分路 3 不做任何操作，其直接将输入体素 $\boldsymbol{F}_v$ 作为该路径的输出特征，其目的是避免由于其他路径提取深层语义信息导致的梯度爆炸。

最后，对 3 条路径的输出特征进行加权融合。本文采用 Softmax 归一化自适应加权融合的方式对 3 种输出特征进行区分性的融合，即：

$$\boldsymbol{F}_{THM} = \alpha_1 \boldsymbol{F}_{HM} + \alpha_2 \boldsymbol{F}_{HM}^{low} + \alpha_3 \boldsymbol{F}_v \tag{9}$$

$$\alpha_i = \frac{e^{w_i}}{\sum_j e^{w_j}}, \quad i = 1,2,3;\ j = 1,2,3 \tag{10}$$

其中，$w_i$ 为初始化权重，$\alpha_i$ 为 Softmax 归一化权重，$\boldsymbol{F}_{THM}$ 为加权融合后的输出特征。对权重进行归一化可增强多输入特征动态融合的稳定性并加快网络的收敛速度。

值得注意的是，H3R 在进行体素一维序列化处理的过程中能够一定程度上保持空间邻近性，但仍不可避免会导致目标空间信息丢失。此外，在体素聚合的过程中也会产生信息的损失。为了解决该问题，本文提出简单有效的空间特征恢复算子（spatial feature restoration operator，SFRO），其能够充分利用 Mamba 捕获长距离上下文的优势对 3D 主干网络的输出特征进行增强。SFRO 如图 4 所示，首先使用 3D 子流形卷积对特征进行局部优化，然后采用 H3R 配合 Mamba 进行特征全局细化，最后进行层归一化和 GELU 非线性激活。本文将在后续的消融实验中验证该算子的有效性。

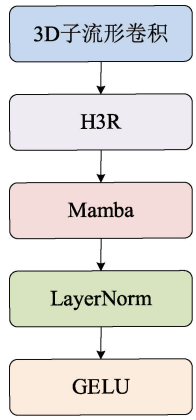



图 4 SFRO 整体流程

Fig. 4 Overall architecture of SFRO

### 2.3　基于特征图强度响应的动态体素扩散模块

激光雷达只能够采集物体表面的点云数据,这导致在 3D 目标检测过程中,目标中心及附近可能缺少有效的体素特征用于生成锚框。因此,本文提出一种基于特征图强度响应的自适应体素扩散模块用于补充目标中心的体素特征信息。受文献[25]启发,3D 主干网络输出的特征图在通道维度上的响应强度与目标区域存在强相关性。基于此,本文放弃了常用的额外监督分支和复杂前景判别机制来选择扩散区域,提出了一种基于特征图强度响应的无参数前景分割方法,具体结构如图 5 所示。

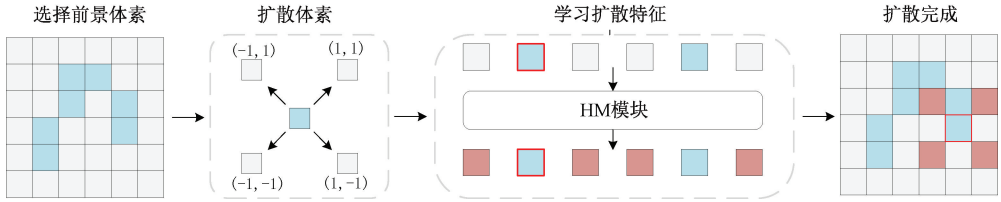



图 5　体素扩散过程

Fig. 5　Voxel diffusion process

首先,计算 LCF-HM 模块输出特征 $\boldsymbol{F}_i$ 的通道平均响应强度,即:

$$\boldsymbol{F}_i^* = \frac{1}{C}\sum_{j=1}^{C}\boldsymbol{F}^j,\quad j=1,2,\cdots,N \tag{11}$$

其中,$C$ 是特征 $\boldsymbol{F}_i$ 的通道数,$N$ 是非空体素个数。

然后,设置前景比例因子 $r\in(0,1)$,以此确定前景体素的数量,即:

$$m = r\times N \tag{12}$$

对 $\boldsymbol{F}_i^*$ 按响应值进行降序排列,取前 $m$ 个体素作为前景体素,即:

$$\boldsymbol{P}_m = Top_m(\boldsymbol{F}_i^*) \tag{13}$$

其中,$Top_m$ 表示取最大的 $m$ 个值。

确认前景体素后,需要采用一定的方法对其进行扩散,以生成目标特征。其中,K 近邻(K-nearest neighbor, KNN)算法为最常用的体素扩散方法。但是受限于局部感受野,KNN 算法将导致扩散体素的特征存在严重偏差。本文设计的 LCF-HM 模块具有捕获全局上下文信息的能力,因此利用该模块和其他体素进行大规模的长程特征交互能够有效地生成扩散体素的特征。具体地,定义 4 组空间偏移量,即:

$$\boldsymbol{\delta} = \{(-1,-1,0),(1,-1,0),(1,1,0),(-1,1,0)\} \tag{14}$$

对每个前景坐标 $\boldsymbol{p}_k$ 生成扩散坐标集合 $\boldsymbol{D}_k$,即:

$$\boldsymbol{D}_k = \{\boldsymbol{p}_k + \delta_j \mid \delta_j\in\Delta\} \tag{15}$$

将 $\boldsymbol{D}_k$ 对应的扩散特征初始化为全零矩阵 $\boldsymbol{F}_{\boldsymbol{D}_k}=\boldsymbol{0}\in \mathrm{R}^{4\times C}$,并将所有扩散特征沿体素维度聚合得到扩展特征矩阵,即:

$$\boldsymbol{F}_{\text{expand}} = \bigcup_{k=1}^{m}\boldsymbol{F}_{\boldsymbol{D}_k}\in \mathrm{R}^{4m\times C} \tag{16}$$

最后,将第 $i$ 个 LCF-HM 模块的输出特征 $\boldsymbol{F}_i$ 和扩散特征 $\boldsymbol{F}_{\text{expand}}$ 进行拼接,并输入第 $i+1$ 个 LCF-HM 模块进行扩散体素的特征学习,即:

$$\boldsymbol{F}_p = \text{Concat}(\boldsymbol{F}_i,\boldsymbol{F}_{\text{expand}})\in \mathrm{R}^{(L+4m)\times C} \tag{17}$$

$$\boldsymbol{F}_p' = \text{Block}(\boldsymbol{F}_p) \tag{18}$$

其中,Block 表示 LCF-HM 模块。

### 2.4　损失函数

目标预测主要包括目标物体分类,3D 框位置、大小及朝向,故损失函数由分类损失、位置-尺寸回归损失和角度回归损失组成。

使用 $(x_c,y_c,z_c,w,l,h,\theta)$ 定义一个 3D 预测框,$(x_c,y_c,z_c)$ 表示中心坐标,$(w,l,h)$ 表示尺寸大小,$\theta$ 表示方向角。对几何参数进行归一化回归计算,即:

$$\begin{cases}\Delta x_c = \dfrac{x_{gt}-x_{pred}}{d_{pred}}\\ \Delta w = \log\left(\dfrac{w_{gt}}{w_{pred}}\right)\\ \Delta\theta = \sin(\theta_{gt}-\theta_{pred})\end{cases} \tag{19}$$

其中,$d_{pred}=\sqrt{(l_{pred})^2+(w_{pred})^2}$ 为预测框底面对角线长度。预测框的回归损失使用 SmoothL1 进行衡量,即:

$$L_{reg} = \sum_{i\in\{x_c,y_c,z_c,w,l,h,\theta\}}\text{SmoothL1}(\Delta i) \tag{20}$$

采用 Focal Loss 作为分类损失以缓解正负样本数量不平衡问题,即:

$$L_{cls} = -\alpha(1-p_c)^{\gamma}\log(p_c) \tag{21}$$

其中,$p_c$ 为预测框类别的概率,$\alpha$ 和 $\gamma$ 分别为平衡正负样本权重的系数和调节难易样本权重的指数,本文取 $\alpha=0.5,\gamma=2$。

为了消除角度周期性导致的方向预测错误,使用方向分类损失辅助角度回归,即:

$$L_{dir} = -[y_k\log(p_k)+(1-y_k)\log(1-p_k)] \tag{22}$$

其中,$y_k\in\{0,1\}$ 表示第 $k$ 个锚框的方向标签,$p_k$ 表示该锚框属于正方向的概率。

总损失函数为上述损失的加权和,即:

$$L_{\text{total}} = \beta_1 L_{cls}+\beta_2 L_{reg}+\beta_3 L_{dir} \tag{23}$$

为避免方向预测干扰主回归任务,分别取 $\beta_1 = 1.0$, $\beta_2 = 2.0$,$\beta_3 = 0.2$。

## 3 实验验证

### 3.1 数据集和实验环境

本文使用 KITTI 公开数据集对模型进行训练和验证评估,该数据集为室外道路真实采集的点云和图片数据,包含 7 481 个训练样本和 7 518 个测试样本。数据集主要的检测目标类别为汽车(Car)、行人(Pedestrian)和骑行者(Cyclist),并将目标按照不同的遮挡、截断情况分为简单(Easy)、中等(Moderate)和困难(Hard)这 3 个检测难度。其中,汽车、行人和骑行者类别的占比情况分别为 82.99%、12.76%和 4.25%,行人和骑行者类别的数据较少,检测难度较大。

实验的研究平台使用 Ubuntu 20.04 系统,硬件环境为英特尔 i7-13700K @ 3.4 GHz CPU 和 NVIDIA RTX 4080 GPU,算法实现基于 PyTorch 深度学习框架。本文预先对输入点云的空间范围进行归一化处理,设置 $(x,y,z)$ 这 3 个维度的有效区域分别为(0, 70.4)、(−40,40)和(−3,1),对每帧点云进行最远点采样保留 16 000 个点,体素大小设置为(0.2,0.2,0.125)。本文使用 2 张显卡进行训练,每张显卡批处理大小为 2,采用真值采样、随机翻转等方法进行数据增强,利用 AdamW 优化器进行模型训练优化,初始学习率和权重衰减设置为 $5.625\times10^{-4}$ 和 0.1,epoch 大小为 80,模型中 LCF-HM 模块的个数为 6。

### 3.2 评价指标

KITTI 数据集的评估体系中采用平均精度均值(mean average precision, mAP)作为三维目标检测性能的综合评价指标。该值由目标检测的精确率和召回率共同计算获得,其计算公式为:

$$Precision = \frac{TP}{TP + FP} \tag{24}$$

$$Recall = \frac{TP}{TP + FN} \tag{25}$$

其中,$TP$ 表示将正样本分类为正样本的数量,$FP$ 表示将负样本分类为正样本的数量,$FN$ 表示将正样本分类为负样本的数量。通过设置不同的类别置信度可以可以构建目标检测精度与召回率关系(precision-recall, PR)曲线,该曲线与坐标轴包围区域的面积可得到目标检测的平均精度。

KITTI 数据集主要针对汽车、行人和骑行者进行检测,其交并比阈值分别为 0.7、0.5 和 0.5。在计算目标检测的 mAP 时,采用 KITTI 官方度量标准 R40 插值法进行计算。该算法在 PR 曲线上均匀选取 40 个召回率位置,并在每个位置取对应召回区间内的最大精度值进行积分从而消除 PR 曲线的局部波动影响,最后将检测结果在 3D 基准上和其他模型进行对比。

### 3.3 对比实验分析

为了验证 GVINet 的有效性,本文在 KITTI 数据集上与其他基于点云的 3D 目标检测算法进行性能对比实验,具体结果如表 1 所示。相较于其他先进的 3D 目标检测算法,GVINet 在 KITTI 数据集上对汽车、行人和骑行者的检测结果表现出较好的性能,以每类目标的中等难度为例,本文方法的检测精度分别达到了 82.36%、61.96%、66.05%。

**表 1 不同方法在 KITTI 验证集上的 3D 检测精度对比**

**Table 1 Comparison of 3D detection accuracy of different methods on the KITTI validation set**

| 方法 | 汽车/% | | | 行人/% | | | 骑行者/% | | | 计算量/G | 参数量/M |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 简单 | 中等 | 困难 | 简单 | 中等 | 困难 | 简单 | 中等 | 困难 | | |
| SECOND[3] | 89.55 | 79.67 | 74.34 | 59.60 | 52.13 | 46.60 | 80.36 | 63.57 | 59.65 | 76.8 | 4.6 |
| PointPillar[14] | 88.57 | 79.45 | 76.69 | 55.54 | 49.84 | 45.99 | 82.38 | 62.91 | 58.73 | 63.4 | 4.8 |
| 改进 PointPillars[26] | 89.23 | 81.82 | 77.46 | 58.25 | 52.66 | 48.45 | 85.37 | 65.25 | 61.01 | | |
| 改进 PointRCNN[27] | 88.87 | 78.61 | 77.73 | 63.33 | 56.24 | 51.24 | | | | | |
| DSVT-Pillar[16] | 87.30 | 77.40 | 76.20 | 61.40 | 56.80 | 51.80 | 82.30 | 67.10 | 63.70 | 221.8 | 6.0 |
| DSVT-Voxel[16] | 87.80 | 77.80 | 76.80 | 66.10 | 59.70 | 55.20 | 83.20 | 66.70 | 63.20 | 303.8 | 6.1 |
| Pv-rcnn[28] | 90.25 | 81.43 | 76.82 | 52.17 | 43.29 | 40.29 | 78.60 | 63.71 | 57.65 | 89.2 | 12.4 |
| PL++[29] | 86.60 | 75.23 | 70.34 | 41.53 | 33.89 | 31.42 | 62.80 | 48.97 | 42.80 | | |
| Part-A2[30] | 85.94 | 77.86 | 72.00 | 54.49 | 44.50 | 42.36 | 78.58 | 62.73 | 57.74 | 76.0 | 23.0 |
| SeSame[31] | 85.25 | 76.83 | 71.60 | 42.29 | 35.34 | 33.02 | 69.55 | 54.56 | 48.34 | | |
| VPFNet[32] | 88.15 | 80.97 | 76.74 | 54.65 | 48.36 | 44.98 | 77.64 | 64.10 | 58.00 | | |
| LION-Mamba[26] | 88.60 | 78.30 | 77.20 | 67.20 | 60.20 | 55.60 | 83.00 | **68.60** | **63.90** | 291.4 | 4.5 |
| GVINet | **90.84** | **82.36** | **79.83** | **69.78** | **61.96** | **56.54** | **85.48** | 66.05 | 61.85 | 217.9 | 4.7 |

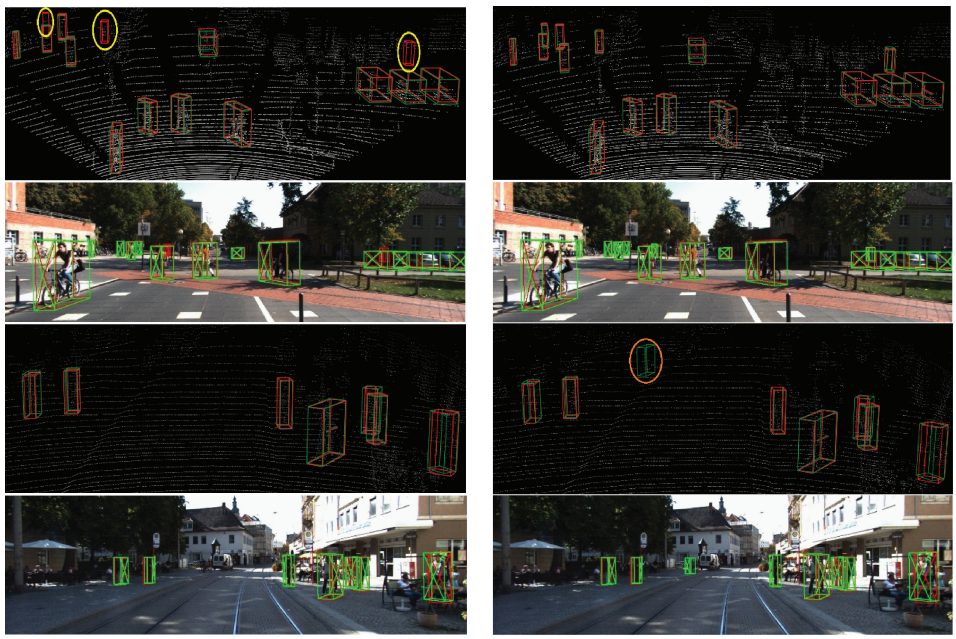
在汽车类别的检测任务中,本文方法 GVINet 在简单、中等和困难这 3 个不同难度的检测等级中,平均精度分别达到了 90.84%、82.36% 和 79.83%,较当前最新的基于 Mamba 的 3D 目标检测模型 LION-Mamba 分别提升 2.24%、4.06% 和 2.63%。本文分析认为 LION-Mamba 采用的窗口划分策略割裂了检测目标的连续空间特征,导致对遮挡目标的敏感性增加。而本文提出的 H3R 采取 Hilbert 曲线序列化策略保留了体素间的局部上下文关系,有效增强了车辆这类大尺寸目标的整体结构特征捕捉能力。此外,本文设计的体素扩散模块通过向 4 个方向扩充前景体素,缓解了激光雷达对车辆中心区域的点云稀疏性问题,尤其是在遮挡场景下的表现更为突出。

在行人类别的检测任务中,GVINet 在简单、中等和困难 3 个不同难度的检测等级上的平均精度分别达到了 69.78%、61.96% 和 56.54%,较 LION-Mamba 提升 2.58%、1.76% 和 0.94%。本文认为这得益于空间特征恢复算子(SFRO)的多尺度特征优化机制,SFRO 通过子流形卷积对局部特征进行精细化处理并结合 Mamba 模块的全局上下文建模能力,有效恢复了因体素聚合丢失的行人细节特征。此外,LCF-HM 模块的多分支动态融合策略通过自适应加权不同的分辨率特征,增强了对行人姿态多样性的鲁棒性。

在骑行者类别的中等和困难两个难度下,GVINet 的平均精度分别为 66.05% 和 61.85%,略低于 LION-Mamba 的 68.60% 和 63.90%。本文分析认为这可能源于数据类别分布不均的影响,KITTI 数据集中骑行者样本占比仅为 4.25%,并且测试集包含大量模糊和密集遮挡的场景,导致模型对复杂姿态的骑行目标泛化能力不足;此外,为了保证模型简洁性,体素扩散模块仅对固定的水平方向进行偏移扩散,导致无法完全覆盖骑行目标垂直方向的姿态变化,使得部分关键体素未被有效补充。

在参数量和计算量方面,GVINet 与基于窗口划分的 LION-Mamba 相比,参数量增加了 0.2 M,但是通过采用希尔伯特曲线遍历体素代替 LION-Mamba 的窗口分组遍历方法并设计离线坐标-索引映射表对计算复杂度进行优化,将计算量由 291.4 G 降低至 217.9 G,减少了 25.2%,并且在汽车和行人类别的检测精度上全面优于 LION-Mamba。与推理时间最短的方法 PointPillar 比较,GVINet 的参数量为 4.7 M,较 PointPillar 的 4.8 M 降低了 0.1 M,但 GVINet 的计算量是 PointPillar 的 3 倍左右。究其原因,相较于本文方法将点云转化为体素结构,PointPillar 则是将点云划分为更简单的柱状结构。这种结构很大程度上减少了模型的计算量,但是也引入了高度方向信息的损失。本文提出的 GVINet 在检测精度和模型复杂度之间进行权衡,将点云划分为携带信息更为丰富的体素结构,并引入能够捕获超长距离上下文特征的 Mamba 构建 LCF-HM 模块,通过增加部分计算量显著提高了模型的检测精度。

图 6 为 GVINet 和 LION-Mamba 在复杂行人场景下 3D 目标检测的可视化效果对比。本文利用 Open3D 在三维空间中可视化激光雷达采集的原始点云,分别标注目标物体的真实位置和预测位置的三维包围框。



(a) LION-Mamba可视化
(a) LION-Mamba visualization

(b) GVINet可视化
(b) GVINet visualization

图 6　GVINet 和 LION-Mamba 在复杂行人场景下检测效果对比

Fig. 6　Comparison of detection effects of GVINet and LION-Mamba in complex pedestrian scenes

如图 6(a)中实线椭圆标注所示,LION-Mamba 在复杂场景中对于远距离行人存在漏检现象,该现象反映了窗口划分策略对长程空间连续性建模的局限性。相比之下,GVINet 在图 6(b)中不仅成功检测到全部标注目标,还精准捕捉到点云对应图像中未标注的骑行者(虚线椭圆标注区域),该目标因标注规范限制未被纳入真值,但其空间特征与运动姿态均被 GVINet 有效识别。

图 7 为本文方法 GVINet 和 LION-Mamba 在汽车遮挡场景下 3D 目标检测的可视化效果对比。如图 7(a)中实线椭圆形标注所示,LION-Mamba 在遮挡场景中对背景干扰物敏感,将静态路灯误判为行人,同时长程建模能力不足,对远处仅部分可见的遮挡车辆出现漏检。GVINet 在图 7(b)中展现出显著优势,通过 H3R 保持空间邻近性,结合双分支 LCF-HM 模块的多尺度特征融合机制,有效区分行人特征和背景噪声;同时,体素扩散模块通过特征响应强度引导的四方向偏移扩充,成功捕获遮挡车辆体素的拓扑结构。

(a) LION-Mamba可视化
(a) LION-Mamba visualization

(b) GVINet可视化
(b) GVINet visualization

图 7 GVINet 和 LION-Mamba 在汽车遮挡场景下检测效果对比

Fig. 7 Comparison of detection effects of GVINet and LION-Mamba in car occlusion scenes

图 8 对比了 GVINet 和其他先进算法在汽车类别的中等难度下的精度及推理速度。GVINet 以 82.36% 的检测精度显著领先于现有方法,较 LION-Mamba 和 DSVT-Voxel 分别提升了 4.06% 和 4.56%。在推理速度方面,虽然 PointPillar 具有一定的优势,但其精度明显低于 GVINet。GVINet 的推理速度达到了 19 fps,较 DSVT-Voxel 和 Pv-rcnn[28] 分别提升 72.7% 和 90%,本文提出的 GVINet 在精度与效率间实现了有效平衡。

### 3.4 消融实验

为了验证 GVINet 中各个子模块的有效性,本文在基线基础上逐步引入核心模块进行消融实验。如表 2 所示,仅加入 LCF-HM 模块时,各类别的检测精度较基线均有明显提升,其中汽车中等难度检测精度提升 1.17%,行人困难难度检测精度提升 1.58%,验证了 LCF-HM 通过 Mamba 进行长序列上下文建模的基础优势。加入体素扩散模块后,行人类别的简单、中等和困难检测难度的精度

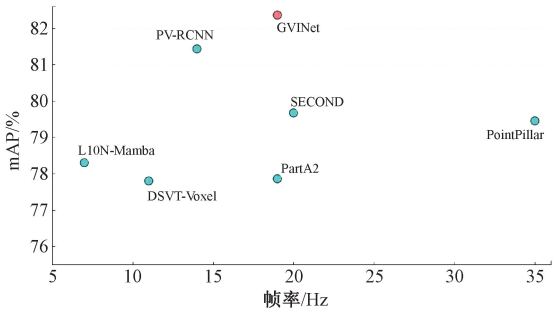



图 8　不同方法在 KITTI 数据集上检测精度和速度的比对

Fig. 8　Comparison of detection accuracy and speed of different methods on the KITTI dataset

分别提升 4.66%、3.09% 和 2.36%,表明体素扩散模块能够通过增强点云空间连续性缓解稀疏点云所导致的特征缺失问题。在 LCF-HM 基础上单独引入 SFRO 模块,全类别的精度均有不同程度提升,验证了 SFRO 对丢失空间信息的恢复能力。最终,各个模块的融合应用使各类别的检测精度达到最优,汽车、行人和骑行者的中等检测难度精度较基线分别提升 4.22%、4.26% 和 2.57%。

本文还将 SFRO 和其他算子进行了对比实验,结果如表 3 所示。SFRO 在汽车类别的简单、中等和困难 3 个不同难度的检测等级中的 AP 分别为 90.84%、82.36% 和 79.83%,相较次优的 Conv 分别提升了 0.61%、0.87% 和 1.03%。SFRO 首先通过子流形卷积对局部特征进行精细化处理,随后利用 Mamba 模块捕捉各个特征间的长程依赖关系,实现了局部细节与全局上下文的协同增强,有效提升了目标的检测精度。然而,在骑行者类别的中等难度下,SFRO 的 AP 较 SubConv 降低了 0.42%,表明全局细化过程可能对小目标的高频特征敏感性不足,未来可通过引入多粒度特征交互策略进一步优化。

**表 2　消融实验结果**

**Table 2　Ablation experiment results**　(%)

| 基线 | LCF-HM | 体素扩散 | SFRO | 汽车 | | | 行人 | | | 骑行者 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 简单 | 中等 | 困难 | 简单 | 中等 | 困难 | 简单 | 中等 | 困难 |
| √ | | | | 87.77 | 78.14 | 75.82 | 61.86 | 57.70 | 51.69 | 82.18 | 63.48 | 59.92 |
| √ | √ | | | 89.53 | 79.31 | 77.59 | 64.02 | 58.34 | 53.47 | 83.11 | 63.53 | 59.70 |
| √ | √ | √ | | 89.71 | 80.85 | 78.23 | 68.68 | 61.43 | 55.83 | 83.29 | 65.97 | 60.03 |
| √ | √ | | √ | 90.23 | 81.11 | 78.14 | 65.33 | 59.13 | 54.91 | 83.58 | 64.16 | 59.88 |
| √ | √ | √ | √ | **90.84** | **82.36** | **79.83** | **69.78** | **61.96** | **56.54** | **85.48** | **66.05** | **61.85** |

**表 3　SFRO 有效性分析**

**Table 3　SFRO effectiveness analysis**　(%)

| 方法 | 汽车 | | | 行人 | | | 骑行者 | | |
|---|---|---|---|---|---|---|---|---|---|
| | 简单 | 中等 | 困难 | 简单 | 中等 | 困难 | 简单 | 中等 | 困难 |
| Baseline | 89.71 | 80.85 | 78.23 | 68.68 | 61.43 | 55.83 | 82.29 | 65.97 | 59.03 |
| MLP | 89.84 | 81.28 | 78.36 | **69.92** | 61.44 | 55.79 | 83.73 | 66.41 | 60.35 |
| Conv | 90.23 | 81.49 | 78.80 | 68.72 | 61.91 | 56.32 | 85.01 | **66.47** | 61.31 |
| SFRO | **90.84** | **82.36** | **79.83** | 69.78 | **61.96** | **56.54** | **85.48** | 66.05 | **61.85** |

此外,本文分析了不同的 LCF-HM 模块个数对点云目标检测性能的影响,揭示了深度与效率的平衡问题。实验以 2 个 LCF-HM 模块为一组,分别设置模块个数为 2、4、6、8 个。如表 4 所示,当模块数量从 2 增至 6 时,模型获得了更深层的全局上下文建模能力,在各个类别上的检测精度整体呈上升趋势;但当数量增至 8 时,行人和骑行者类别的检测精度呈断崖式下降,其中行人类别的简单、中等和困难下的精度分别下降 8.39%、6.88% 和 6.33%,骑行者类别分别下降 3.54%、5.37% 和 5.21%。该现象表明过深的网络结构会导致对稀疏目标的过拟合,尤其是在样本量有限的行人和骑行者类别中更为显著。因此,本文选择 6 个 LCF-HM 模块作为 GVINet 的配置,在参数量与检测精度间实现了有效平衡。

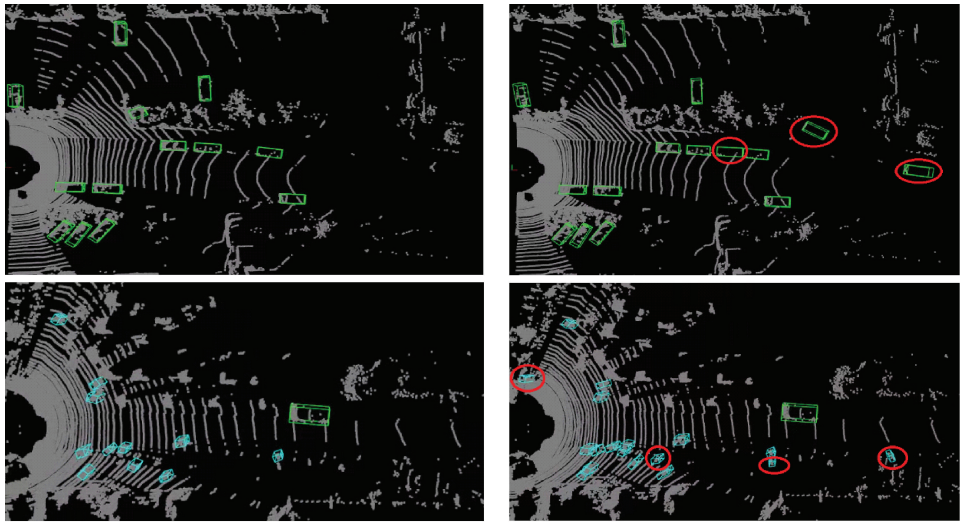
### 3.5　实际场景可视化验证

为了验证 GVINet 在实际场景中的目标检测的有效性,本文采用 AGV 小车搭载 RS-Helios-32 线激光雷达在真实道路环境下进行数据采集,并通过 NVIDIA Jetson Orin 控制器处理并保存点云数据。图 9 对本文提出的 GVINet 与同样是融入了 Mamba 模块的 LION-Mamba

表 4 LCF-HM 模块数量对模型性能的影响

Table 4 The impact of the number of LCF-HM modules on model performance (%)

| LCF-HM 数量 | 汽车/% | | | 行人/% | | | 骑行者/% | | | 参数量/M |
|---|---|---|---|---|---|---|---|---|---|---|
| | 简单 | 中等 | 困难 | 简单 | 中等 | 困难 | 简单 | 中等 | 困难 | |
| $N=2$ | 89.80 | 81.03 | 78.73 | 67.16 | 61.03 | 55.93 | **86.17** | 65.02 | 61.11 | 4.0 |
| $N=4$ | 89.21 | 81.65 | 78.73 | 67.83 | 61.04 | 55.52 | 85.41 | **66.45** | **62.53** | 4.4 |
| $N=6$ | 90.84 | **82.36** | **79.83** | **69.78** | **61.96** | **56.54** | 85.48 | 66.05 | 61.85 | 4.7 |
| $N=8$ | **91.22** | 81.86 | 78.88 | 61.39 | 55.08 | 50.21 | 81.94 | 60.68 | 56.64 | 5.1 |



(a) LION-Mamba检测结果可视化
(a) LION-Mamba visualization of detection

(b) GVINet检测结果可视化
(b) GVINet visualization of detection

图 9 GVINet 和 LION-Mamba 在实际道路上检测效果对比

Fig. 9 Comparison of detection effects of GVINet and LION-Mamba on actual roads scenes

在实际道路条件下的实际检测结果进行了可视化。可以看出,GVINet 能够有效检测到 LION-Mamba 所漏检的远处目标,体现了其获取长距离上下文信息的优势,进一步验证了该方法在实际应用中的有效性。

## 4 结 论

本文提出了一种基于状态空间模型的 3D 点云目标检测网络 GVINet,有效解决了传统卷积网络全局建模能力不足与 Transformer 方法二次计算复杂度之间的矛盾。通过 H3R 实现三维体素到一维因果序列的高效映射,在保持空间邻近性的前提下,结合双分支 LCF-HM 模块的线性复杂度全局特征交互机制,构建了多尺度上下文感知框架。针对体素聚合过程中的信息损失,提出 SFRO 通过子流形卷积与 Mamba 协同优化,实现了局部细节与全局拓扑的联合增强;提出基于特征响应强度的自适应体素扩散模块有效缓解了目标中心区域特征稀疏性问题。在 KITTI 数据集上进行的实验表明,GVINet 能够有效提高不同目标类别的检测精度,验证了状态空间模型在稀疏点云长程依赖建模中的优势。进一步地,GVINet 通过增加部分计算量获得了检测精度的大幅提升,但检测速度受到了限制,后续的工作可以通过知识蒸馏降低模型参数量以提高检测性能。

## 参考文献

[1] 陈慧娴, 吴一全, 张耀. 基于深度学习的三维点云分析 方法研究进展[J]. 仪器仪表学报, 2023, 44(11): 130-158.
CHEN H X, WU Y Q, ZHANG Y. Research progress of 3D point cloud analysis methods based on deep learning[J]. Chinese Journal of Scientific Instrument, 2023, 44(11):130-158.

[2] 陈熙源, 戈明明,姚志婷, 等. 雨雪天气下的激光雷达滤波算法研究[J]. 仪器仪表学报, 2023, 44(7): 172-181.
CHEN X Y, GE M M, YAO ZH T, et al. Research on LiDAR filtering algorithm for rainy and snowy wea-

ther[J]. Chinese Journal of Scifetific Instrument, 2023, 44(7): 172-181.

[3] ZHOU Y, TUZEL O. Voxelnet: End-to-end learning for point cloud based 3D object detection[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 4490-4499.

[4] YAN Y, MAO Y X, LI B. Second: Sparsely embedded convolutional detection[J]. Sensors, 2018, 18(10): 3337.

[5] ZHANG G, CHEN J N, GAO G H, et al. Hednet: A hierarchical encoder-decoder network for 3D object detection in point clouds[J]. ArXiv preprint arXiv. 2310.20234, 2023.

[6] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. ArXiv preprint arXiv. 1706.03762, 2017.

[7] MAO J G, XUE Y J, NIU M ZH, et al. Voxel transformer for 3D object detection[C]. 2021 IEEE/CVF International Conference on Computer Vision, 2021: 3144-3153.

[8] LI Y M, YU ZH D, CHOY C, et al. Voxformer: Sparse voxel transformer for camera-based 3D semantic scene completion[C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 9087-9098.

[9] LIU ZH J, YANG X Y, TANG H T, et al. Flatformer: Flattened window attention for efficient point cloud transformer[C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 1200-1211.

[10] GU A, DAO T. Mamba: Linear-time sequence modeling with selective state spaces[J]. ArXiv preprint arXiv: 2312.00752, 2023.

[11] QI C R, YI L, SU H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space[J]. ArXiv preprint arXiv:1706.02413,2017.

[12] SHI SH SH, WANG X G, LI H SH. Pointrcnn: 3D object proposal generation and detection from point cloud[C]. 2019 IEEE/CVP Conference on Computer Vision and Pattern Recognition, 2019: 770-779.

[13] SHI W J, RAJKUMAR R. Point-GNN: Graph neural network for 3D object detection in a point cloud[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1708-1716.

[14] WANG H Y, DING L H, DONG SH C, et al. CAGroup3D: Class-aware grouping for 3D object detection on point clouds[J]. Advances in Neural Information Processing Systems, 2022, 35: 29975-29988.

[15] LANG A H, VORA S, CAESAR H, et al. Pointpillars: Fast encoders for object detection from point clouds[C]. 2019 IEEE/CVP Conference on Computer Vision and Pattern Recognition, 2019: 12689-12697.

[16] ZHANG G, CHEN J N, GAO G H, et al. Safdnet: A simple and effective network for fully sparse 3D object detection[C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 14477-14486.

[17] WANG H Y, SHI CH, SHI SH SH, et al. Dsvt: Dynamic sparse voxel transformer with rotated sets[C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 13520-13529.

[18] 王溪波, 曹士彭, 赵怀慈, 等. 双边特征聚合与注意力机制点云语义分割[J]. 仪器仪表学报, 2021, 42(12): 175-183.
WANG X B, CAO SH P, ZHAO H C, et al. Semantic segmentation of point cloud via bilateral feature aggregation and attention mechanism[J]. Chinese Journal of Scifetific Instrument, 2021, 42(12): 175-183.

[19] 冯凯浩, 陶志勇, 李衡, 等. 基于 Transformer 的逐通道点云分析网络[J]. 电子测量与仪器学报, 2025, 39(2): 49-59.
FENG K H, TAO ZH Y, LI H, et al. Transformer-based channel-by-channel point cloud analysis network[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(2): 49-59.

[20] LIU ZH, HOU J H, YE X Q, et al. Seed: A simple and effective 3D DETR in point clouds[C]. Computer Vision-ECCV 2024, 2025: 110-126.

[21] LIU Y, TIAN Y J, ZHAO Y ZH, et al. Vmamba: Visual state space model[J]. Advances in Neural Information Processing Systems, 2024, 37: 103031-103063.

[22] LIANG D K, ZHOU X, XU W, et al. Pointmamba: A simple state space model for point cloud analysis[J]. ArXiv preprint arXiv:2402.10739, 2024.

[23] HAN X, TANG Y, WANG ZH X, et al. Mamba 3D: Enhancing local features for 3D point cloud analysis via state space model[C]. 32nd ACM International Conference on Multimedia, 2024: 4995-5004.

[24] ZHANG T, YUAN H B, QI L, et al. Point cloud mamba: Point cloud learning via state space model[J]. ArXiv preprint arXiv:2403.00762, 2024.

[25] LIU ZH, HOU J H, WANG X Y, et al. Lion: Linear group rnn for 3D object detection in point clouds[J]. ArXiv preprint arXiv:2407.18232, 2024.

[26] 汤新华, 代道文, 陈熙源, 等. 基于 PointPillars 的改进三维目标检测算法[J]. 仪器仪表学报, 2024, 45(9): 260-269.
TANG X H, DAI D W, CHEN X Y, et al. Improved three-dimensional object detection algorithm based on PointPillars[J]. Chinese Journal of Scientific Instrument, 2024, 45(9): 260-269.

[27] 王庆林, 李辉, 谢礼志, 等. 基于激光雷达点云的车辆目标检测算法改进研究[J]. 电子测量技术, 2023, 46(1): 120-126.
WANG Q L, LI H, XIE L ZH, et al. Research on improving vehicle target detection algorithm based on LiDAR point cloud[J]. Electronic Measurement Technology, 2023, 46(1): 120-126.

[28] SHI SH SH, GUO CH X, JIANG L, et al. PV-RCNN: Point-voxel feature set abstraction for 3D object detection[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10526-10535.

[29] GONG X, HUANG X, CHEN SH SH, et al. Enhancing 3D detection accuracy in autonomous driving through Pseudo-LiDAR augmentation and down sampling[C]. 2024 International Conference on Image Processing, Computer Vision and Machine Learning, 2024: 1105-1110.

[30] SHI SH SH, WANG ZH, SHI J P, et al. From points to parts: 3D object detection from point cloud with part-aware and part-aggregation network[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(8): 2647-2664.

[31] HAYEON O, YANG C, HUH K. SeSame: Simple, Easy 3D object detection with point-wise semantics[C]. Computer Vision-ACCV 2024, 2025: 211-227.

[32] WANG C H, CHEN H W, CHEN Y, et al. VoPiFNet: Voxel-pixel fusion network for multi-class 3D object detection[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(8): 8527-8537.

## 作者简介

**刘明杰**,2019 年于韩国仁荷大学获得博士学位,现为重庆邮电大学副教授,主要研究方向为智能网联汽车环境智能感知、多源信息融合、机器学习。
E-mail:liumj@ cqupt. edu. cn

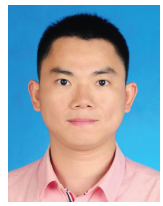
**Liu Mingjie** received his Ph. D. degree from Inha University in 2019. He is currently an associate professor at Chongqing University of Posts and Telecommunications. His main research interests include environment perception for the internet of vehicles, multi-information fusion, and machine learning.

**魏宇**,2023 年于重庆邮电大学获得学士学位,现为重庆邮电大学硕士研究生,主要研究方向为智能网联汽车环境智能感知。
E-mail:S230301059@ stu. cqupt. edu. cn

**Wei Yu** received his B. Sc. degree from Chongqing University of Posts and Telecommunications in 2023. He is currently a master′s student at Chongqing University of Posts and Telecommunications. His main research interest is environmental perception for internet of vehicles.

**陈俊生**, 2019 年于重庆大学获得博士学位,现为重庆邮电大学讲师,主要研究方向为智能网联汽车环境智能感知、能源设备状态智能诊断。
E-mail:chenjunsheng@ cqupt. edu. cn

**Chen Junsheng** received his Ph. D. degree from Chongqing University in 2019. He is currently a lecturer at Chongqing University of Posts and Telecommunications. His main research interests include environment perception for internet of vehicles, intelligent diagnosis of energy equipment status.



**刘平**,2017 年于浙江大学获得博士学位,现为重庆邮电大学副教授,主要研究方向为智能网联汽车环境智能感知及控制、轨迹优化。
E-mail:liuping_cqupt@ cqupt. edu. cn

**Liu Ping** received his Ph. D. degree from Zhejiang University in 2017. He is currently an associate professor at Chongqing University of Posts and Telecommunications. His main research interests include environment perception & control for internet of vehicles, and trajectory optimization.

**朴昌浩**(通信作者),2001 年于西安交通大学获得学士学位,2006 年于韩国仁荷大学获得博士学位,现为重庆邮电大学教授,主要研究方向为自动驾驶、智能网联汽车。
E-mail:piaoch@ cqupt. edu. cn

**Piao Changhao** (Corresponding author) received his B. Sc. degree from Xi′an Jiaotong University in 2001, and his Ph. D. degree from Inha University in 2006. He is currently a professor at Chongqing University of Posts and Telecom-munications. His main research interests include autonomous driving and internet of vehicles.