温国强, 李承坤, 胡顺堂,等. 基于语义引导的点云行人目标检测算法研究[J]. 智能计算机与应用,2025,15(9):176-184.
DOI:10.20169/j.issn.2095-2163.25042701

基于语义引导的点云行人目标检测算法研究

温国强[1], 李承坤[1], 胡顺堂[1], 常文爽[1], 郭跃武[2], 李洪艳[3], 苏 威[4]
(1 天津中德应用技术大学 汽车与轨道交通学院, 天津 300350; 2 沂普光电(天津)有限公司, 天津 300385;
3 天津圣威科技有限公司, 天津 300074; 4 天津科技大学 电子信息与自动化学院, 天津 300222)

摘 要: 针对自动驾驶场景中激光雷达点云数据稀疏性、无序性及非均匀分布特性导致的特征提取瓶颈,以及现有基于点的检测算法在行人目标检测和建模方面的不足,本文改进了基于点的单阶段 3D 目标检测算法 IA-SSD,提出了一种基于语义引导的行人目标检测算法:SGP-SSD。该算法在下采样过程中引入语义引导的下采样算法,综合考虑点的语义信息和距离信息,同时进行了模型的结构优化,添加了反向 MLP 模块以反向瓶颈结构和可分离的 MLPs 丰富特征提取,并提出了多尺度特征聚合模块,使投票得到的质心可以同时聚合小半径和大半径的特征信息以及针对小目标检测优化的投票 loss 机制。通过对比和消融实验以及可视化验证,相对于 IA-SSD 算法,本文算法在 KITTI 验证集行人类别中,简单/中等/困难难度级别的检测精度分别提升 4.81、3.43、3.56;骑行者类别提升了 3.50、2.35、0.31。在 Waymo 验证集上,行人目标的 Level1/Level2 难度 AP/APH 指标提升 2.01/3.17 和 3.43/1.97,骑行者目标 Level1/Level2 的 AP/APH 指标提升 2.05/2.66 和 1.78/2.11。

关键词: 自动驾驶; 行人目标检测; 3D 目标检测; 语义引导; 多尺度特征聚合
中图分类号: TP391 **文献标志码**: A **文章编号**: 2095-2163(2025)09-0176-09

Research on semantic-guided point cloud-based pedestrian target detection algorithm

WEN Guoqiang[1], LI Chengkun[1], HU Shuntang[1], CHANG Wenshuang[1], GUO Yuewu[2], LI Hongyan[3], SU Wei[4]
(1 Automobile and Rail Transportation School, Tianjin Sino-German University of Applied Sciences, Tianjin 300350, China; 2 Yipu Optoelectronics (Tianjin) Co., Ltd., Tianjin 300385, China; 3 Tianjin Shengwei Technology Co., Ltd., Tianjin 300074, China; 4 College of Electronic Information and Automation, Tianjin University of Science and Technology, Tianjin 300222, China)

Abstract: Aiming at the feature extraction bottlenecks caused by the sparsity, disorderliness, and non-uniform distribution characteristics of LiDAR point cloud data in autonomous driving scenarios, as well as the limitations of existing point-based detection algorithms in pedestrian target detection and modeling, this paper improves the point-based single-stage 3D object detection algorithm IA-SSD and proposes a semantic-guided pedestrian target detection algorithm: SGP-SSD. The algorithm introduces a semantic-guided downsampling strategy during the sampling process, comprehensively considering both the semantic and distance information of points. Structural optimizations are incorporated, including a reverse MLP module with inverted bottleneck structures and separable MLPs to enhance feature extraction. Additionally, a multi-scale feature aggregation module is proposed to enable the centroids generated by voting to simultaneously aggregate features from small and large radii, along with a voting loss mechanism optimized for small target detection. Through comparative experiments, ablation studies, and visual validation, the proposed algorithm achieves significant improvements over IA-SSD. On the KITTI validation set for the pedestrian category, detection accuracy increases by 4.81, 3.43, and 3.56 for easy/moderate/hard difficulty levels, respectively, and for the cyclist category, improvements of 3.50, 2.35, and 0.31 are observed. On the Waymo validation set, the Level1/Level2 AP/APH metrics for pedestrian targets improve by 2.01/3.17 and 3.43/1.97, while cyclist targets achieve AP/APH gains of 2.05/2.66 and 1.78/2.11 at Level1/Level2.

Key words: autonomous driving; pedestrian target detection; 3D object detection; semantic guidance; multi-scale feature aggregation

0 引 言

自动驾驶技术正处于飞速发展之中,是汽车行业转型升级的重要发展方向[1]。该技术旨在通过智能环境感知实现自主安全行驶[2]。环境感知系统作为核心技术,其通过雷达、摄像头等多传感器融

基金项目: 天津市教委科研项目(2020KJ086)。
作者简介: 温国强(1984—),男,博士,副教授,硕士生导师,主要研究方向:智能网联汽车技术,光电检测技术。Email:2981270182@qq.com。
收稿日期: 2025-04-27

合实现精准环境感知[3]。该系统实时感知目标物位置/速度等多维度参数，为车辆决策提供数据支撑，其通过多传感器数据融合实现对道路环境的动态解析[4]。感知数据为决策层提供关键支撑，激光雷达点云通过三维几何建模对物体进行精准定位，其空间属性解析能力超越二维检测，兼具光线鲁棒性与广域三维环境重构优势[5]。但激光雷达点云数据具有稀疏、无序和非均匀分布等特性，导致传统卷积神经网络（Convolutional Neural Networks，CNN）方法难以直接提取特征[6]。突破点云数据的三维特征提取技术仍是该领域亟待解决的关键挑战。

Shi 等学者[7]提出的 PointRCNN 率先在基于点的 3D 点云目标检测领域取得进展，该算法采用两阶段网络结构，先使用最远点采样（FPS）下采样后点云生成基于点的提案，并引入区域提议网络对提案进行优化，得到最终的边界框。由于两阶段网络结构复杂和效率低下的问题，Yang 等学者[8]提出一种高效的单阶段点云检测器 3DSSD。该算法使用一种同时利用特征和几何距离的混合采样策略，使用 PointNet 来学习点云特征，并移除了 PointNet 中的 FP 层以获得更好的性能。由于传统的最远点采样方法仅仅考虑了点云之间的距离，导致采样得到的点很多都是不重要的背景点，因此 Zhang 等学者[9]提出的单阶段算法 IA-SSD 通过 2 种可学习的、面向任务的下采样策略来选择感兴趣的前景点，并引入了上下文质心感知模块，以提高实例中心的估计精度。另一方面，Chen 等学者[10]也关注到了前景点的重要性，提出 SASA 关注于提升下采样点中前景点的比例，与 IASSD 不同的是，并不一味地选择前景点，而是综合考虑点的语义信息和距离，使用物体前景分数指导采样过程，从而更有效地捕捉到目标物体的特征。

为了避免处理不规则点云，目前的 3D 检测方法在很多方面都严重依赖基于 2D 的检测器[11]。这会牺牲几何细节，从而影响检测精度，尤其是小目标的检测精度。而基于点的方法工作在原始点之上，不需要将点转化为其他表示，相对体素网络来说能够保留更多的细节，拥有更加灵活的感受野[12]。因此本文针对点目标检测算法 IA-SSD 进行了相应改进，提出了基于语义引导的点云行人目标检测算法（SGP-SSD）。

1　IA-SSD 算法介绍

IA-SSD[9]是一个高效的单级点目标检测器，算法结构如图 1 所示。采用了轻量级编码器架构提升检测效率，通过实例感知下采样保留关键前景点降低计算成本，并利用上下文质心感知模块精准估计实例中心生成检测框。其流程依次为点云特征提取、特征下采样、实例中心定位及边界框回归。

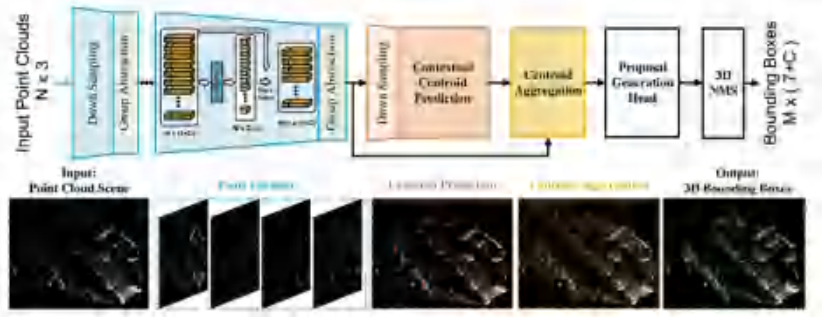


图 1　IA-SSD 算法结构图

Fig. 1　Structure diagram of IA-SSD algorithm

2　模型的结构优化

针对 IA-SSD 中 Top-K 下采样偏向大目标（如汽车）及小目标投票易受干扰的问题，本文提出语义引导单级 3D 目标检测器 SGP-SSD，改进如下：

（1）本文使用语义引导的下采样方法（S-FPS），综合考虑点的语义信息和距离信息，使得点云下采样对不同目标更加公平。

（2）此外在每次进行特征提取的 SA 模块之后，本文加入了反向 MLP（InvMLP）结构，使用反向瓶颈结构和可分离的 MLPs 作为特征提升了特征提取能力。

（3）提出多尺度特征聚合模块（MSFA），使投票得到的质心可以同时聚合小半径和大半径的特征信息。

算法结构如图 2 所示，本文提出的算法主要分为 3 个部分：语义引导的下采样模块、反向 MLP 模块和多尺度特征聚合模块。

图 2 SGP-SSD 算法结构图

Fig. 2 Structure diagram of SGP-SSD algorithm

首先将原始点云输入网络中,经过 3 次下采样和特征提取,依次为 D-FPS、D-FPS、S-FPS,结合 InvMLP 丰富 SA 模块特征提取。此外通过前景分数计算模块为 S-FPS 提供前景分数。上下文质心预测模块前先经过一次 S-FPS 下采样,得到候选点,使用 MLP 计算出坐标偏移量,然后和候选点相加得到投票出的质心点。多尺度特征聚合模块使用下采样的 SA 层以投票出的质心点为中心,同时聚合多尺度不同半径的点特征得到质心特征。最后,将质心特征输入到检测头,从而预测出回归 3D 边界框和相应的类别标签。

2.1 语义引导的下采样模块

语义引导的下采样模块结构如图 3 所示。本文将类别分数预测模块替换为前景分割模块,只预测点是否为前景点,而不区分类别,得到预测分数后通过 S-FPS 对输入点采样。

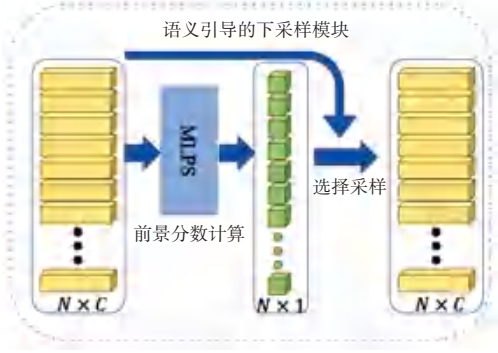


图 3 语义引导的下采样模块示意图

Fig. 3 Schematic diagram of semantic-guided downsampling module

(1)前景分割模块。在 IA-SSD 中,为了给类别感知下采样模块提高点的类别分数,使用 MLP 计算点的分数,每个点的分数为 $[x_1, x_2, x_3]$,其中 $x_i \in [0,1]$,代表每个类别的概率。

本文只把点区分为前景点或者背景点,分数计算 MLP 层是简单的 2 层 MLP,可以计算点的前景分数,其中输入特征 $\{f_1^{l_k}, f_2^{l_k}, \cdots, f_{N_k}^{l_k}\}$ 作为输入到第 k 个下采样层的 l_k 维度特征,那么第 k 个下采样层得到的每个点的类别分数为:

$$p_i^k \in [0,1] \tag{1}$$

p_i^k 由点特征 $f_i^{l_k}$ 的计算公式为:

$$p_i^k = \sigma(M_k(f_i^{l_k})) \tag{2}$$

其中,$M_k(\cdot)$ 表示第 k 层下采样层中添加的分数计算 MLP 层,$\sigma(\cdot)$ 表示 Sigmoid 函数。

为了训练每个下采样层中的点分割模块,点的前景分割标签可以自然地从框注释中得到。用交叉熵损失[13]函数计算总分割损失公式如下:

$$l_{\text{seg}} = -\sum_{k=1}^{m}\frac{\lambda_k}{N_k}\sum_{i=1}^{N_k}(p_i^k\log(\hat{p}_i^k) + (1-p_i^k)\log(1-\hat{p}_i^k)) \tag{3}$$

其中,p_i^k 和 $\hat{p}_i^k$ 分别表示预测的前景分数和真实框分割标签(1 表示来自前景,0 表示来自背景)在第 k 个下采样层的第 i 个点;N_k 和 λ_k 分别表示第 k 个 SA 层的输入点总数和分割损失权值。第 3 次和第 4 次下采样的分割损失权重分别设置为 0.01 和 0.1。

(2)语义引导的下采样方法(S-FPS)。点云中的点可划分为前景点与背景点,局部语义感知通过保留更多前景点优化检测性能。利用前文介绍的前景分数计算模块,可以获得点的语义分数。

为了有效利用点的语义分数,一种直接的方法是对分数排序,保留前景得分最高的部分点,但这种方法容易从识别的对象(汽车)中选择了太多的点,这些对象通常具有更高的前景得分。得到的关键点集不能覆盖三维场景导致大比例的小目标(行人、骑行者)被忽略,影响检测性能。

因此,本算法使用了基于语义引导的最远点采样,结合了全局场景感知和语义启发式诱导的局部目标感知,并考虑到前面分割模块产生的点向语义以及输入的点坐标。S-FPS 的主要思想是通过优先考虑前景得分较高的点来选择更多的前景点。在保持 FPS 的整体过程不变的情况下,用点前景分数来校正采样度量,得到已采样点的距离。S-FPS 算法设计描述如下。

算法　语义引导的最远点采样算法

输入　点坐标 $X = \{x_1, \cdots, x_N\} \in R^{N\times 3}$，点语义分数 $P = \{p_1, \cdots, p_N\} \in R^N$，$N$ 表示输入点的个数

输出　采样的关键点集合 $K = \{k_1, \cdots, k_M\}$。这里，M 表示算法采样输出点的个数

1. 初始化一个空采样点集 K
2. 初始化长度为 N 的距离数组 d 值为 $+\infty$
3. 初始化长度为 N 的访问数组 v 值为 $+\infty$
4. for $i = 1$ to M do
5.　if $i = 1$ then
6.　　$k_i = \text{argmax}(P)$
7.　else
8.　　$D = \{p_k^{\gamma} \cdot d_k \mid v_k = 0\}$
9.　　$k_i = \text{argmax}(D)$
10.　end if
11.　add k_i to K, $v_{k_i=1}$
12.　for $j = 1$ to N do
13.　　$d_j = \min(d_j, \|x_j - x_{k_i}\|)$
14.　end for
15. end for
16. return K

具体来说，给定三维坐标 $X = \{x_1, x_2, \cdots, x_N\}$ 和前景分数 $P = \{p_1, p_2 \cdots, p_N\}$ 为输入点，距离阵列 $\{d_1, d_2, \cdots, d_N\}$ 保持从第 i 点到已选关键点的最短距离。在每轮选择中，算法将语义加权距离 $\hat{d}_i$ 最高的点加入关键点集，计算方法可表示为：

$$\hat{d}_i = p_i^{\gamma} \cdot d_i \tag{4}$$

其中，γ 表示控制语义信息重要性的平衡因子。值得注意的是，当 $\gamma = 0$ 时，S-FPS 可以减少到普通 FPS，并且如果 γ 变得非常大，也可以近似于前面提到的 Top-K 选择方法。

2.2　反向 MLP 模块

受到 PointNex[14] 的启发，本算法在网络中引入了一个新的反向 MLP（InvMLP）模块。并将其添加到每个阶段的第一个 SA 块之后，就能实现高效和有效的模型缩放。InvMLP 模块的设计如图 4 所示。由于算法的深度并不大，所以没有采用 PointNext 中的残差连接。

图 4　InvMLP 模块结构

Fig. 4　Structure of InvMLP module

为了减少计算量并增强点特征的提取能力，InvMLP 模块采用可分离 MLP 替代原 SA 模块中的常规 MLP 结构（原设计通过卷积层、批归一化及 ReLU 实现邻域特征的非线性变换）。而在 InvMLP 模块中，将 MLP 分成 2 个部分：一个基于邻域特征计算的单层（位于分组和池化层之间），以及 2 个基于点特征计算的层（在池化层之后）。此外，InvMLP 模块采用了倒瓶颈设计，将第 2 个 MLP 的输出通道扩展了 4 倍。这样的设计可以增加模块的表达能力，提取更具信息量的点云特征。

2.3　多尺度特征融合模块

本文提出了一种点云多尺度特征聚合模块，旨在改进 IASSD 中的特征聚合方法。模块结构如图 5 所示。在原有的特征聚合模块中，以获得的质心为中心，聚合第 3 次下采样后质心周围的点特征。然而，由于聚合的半径较大，更适合处理大目标、如车辆，对于行人等小目标来说，聚合的范围包含了过多的信息，其中可能包含了许多相邻目标的信息。

图 5　多尺度特征聚合模块

Fig. 5　Multi-scale feature aggregation module

为了解决这一问题,本文设计了一个多尺度特征聚合模块,添加一个小半径质心特征聚合分支。以投票得到的质心使用不同半径来聚合点特征,能够更好地处理不同尺度的目标。具体而言,较小半径下的聚合感受野更小,可捕捉到更细节的信息,有助于提升对小目标的感知能力。较大半径下的聚合感受野更大,则可以更好地处理大目标。最后,将不同尺度的点特征相加,得到多尺度特征,可以综合利用不同尺度的信息,提升目标检测或识别的性能。

2.4 损失函数

算法采用多任务损失进行联合优化。总损失(L_{total})由前景分割损失(L_{seg})、投票质心预测损失(L_{vote})、分类损失(L_{cls})和盒生成损失(L_{box})组成。数学定义公式如下:

$$L_{\text{total}} = L_{\text{seg}} + L_{\text{vote}} + L_{\text{cls}} + L_{\text{box}} \tag{5}$$

其中,盒生成损失可进一步分解为位置(Loc)、尺寸(Size)、角度分区(Angle-bin)、角度分辨率(Angle-res)、角点(Corner),由此推得:

$$L_{\text{box}} = L_{\text{loc}} + L_{\text{size}} + L_{\text{angle-bin}} + L_{\text{angle-res}} + L_{\text{corner}} \tag{6}$$

3 实验与结果分析

实验在流行的 KITTI[15] 和 Waymo[16] 数据集上验证本文提出的 SGP-SSD 方法。

KITTI 数据集专注于基于激光雷达的三维目标检测,含 7 481 个训练样本和 7 518 个测试样本(常将训练集拆分为验证(3 769)/训练(3 712))。数据采集使用 64 线激光雷达(10 Hz)及多摄像头/GPS,但本研究仅用点云数据。标注涵盖目标类别、遮挡等级(0~3)、截断状态、3D 框坐标/尺寸、方向角等 16 项属性,适用于自动驾驶场景的物体检测研究。

Waymo 数据集聚焦自动驾驶多模态感知,含 1 150 段 20 s 真实道路场景(城市/郊区),提供同步校准的 LiDAR(75 m 内 3D 标注)、高清相机等多传感器数据。划分 798/202/150 场景作为训练/验证/测试集,覆盖复杂环境与天气,含精确 3D 框标注,支持鲁棒性算法研发。

3.1 实验细节

本节将具体阐述实验的 KITTI 和 Waymo 数据集的参数设置并介绍目标检测的衡量指标。

(1)KITTI 数据集设置。对于 KITTI,原始点云首先被裁剪为(0,70.4) m、(-40,40) m、(-3,1) m,X、Y、Z 轴范围的 Pillar 大小(柱体空间尺寸)为(0.16,0.16,4.00) m。经过 Pillar 特征编码器输出的特征图尺寸为(432,496,64)。在 2d 主干网络部分,3 次下采样的步长为(2,2,2),上采样的步长为(1,2,4),MobileBlock 的堆叠次数为(3,5,5),通道变换为(64,128,256)。Swin Transformer 的通道数为 256,堆叠 4 次。

(2)Waymo 数据集设置。对于 Waymo,原始点云首先被裁剪为(-74.24,74.24) m、(-74.24,74.24) m、(-2,4) m,X、Y、Z 轴范围的 Pillar 大小为(0.32,0.32,0.10) m。经过 Pillar 特征编码器输出的特征图尺寸为(464,464,64)。其他与 KITTI 设置相同。

(3)目标检测衡量指标。目标检测模型输出的结果,包括目标检测框的位置、尺寸和类别等。对于如何评价检测结果的好坏,有各种评价指标,如交并比(IoU)、精确度(Precision)、平均精度(AP)、均值平均精度(mAP)等[17] 指标。此外还介绍了 Waymo 数据集中的航向精度加权平均精度(APH)指标。对应标准公式见表 1。

表 1 衡量指标及其公式

Table 1 Evaluation metrics and the related formulas

衡量指标	公式
交并比(IoU)	$\text{IoU} = \frac{V_{\text{Pred}} \cap V_{\text{True}}}{V_{\text{Pred}} \cup V_{\text{True}}}$
精确度(Precision)	$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}$
平均精度(AP)	$\text{AP}_{40} = \frac{1}{40} \sum_{r \in \{0, 0.025, \cdots, 1\}} P_{\text{interp}}(r)$
均值平均精度(mAP)	$\text{mAP} = \frac{\sum_{i=1}^{N} \text{AP}_i}{N}$
加权平均精度(APH)	$\text{APH} = \int h(r)\,\text{d}r$

3.2 对比实验

本节将对 SGP-SSD 和多种验证集进行对比,特别是基线网络 IA-SSD,同样,本文也以同样的实验环境对其他的 3D 目标检测网络在 2 个数据集上进行了重新训练和测试,并将结果与本文的算法进行对比。但由于 Waymo 数据集的大规模数据处理的成本较高,而 KITTI 数据集规模较小,视觉任务更依赖 KITTI 数据集的标注,因此本文更加注重于基于 KITTI 数据集的评估。

(1)基于 KITTI 数据集的评估。根据常用的设置[18],算法将所有训练样例划分为训练集(3 712 个样本)和验证集(3 769 个样本),所有的实验模型在训练集上进行训练,在验证集上进行测试。在 KITTI 基准中,汽车、行人和骑自行车的对象根据难度被分为 3 个子集(“容易”、“中等”和“困难”)。

“中等”的结果通常作为最终排名的主要指标。

本文借鉴了 Mapillary 团队[19]的研究，使用 40 个召回位置。这使得结果的比较更加公平，评估结果见表 2。表 2 中，加粗数字代表最佳精度。

表 2　KITTI 数据集验证集中各类别评估结果

Table 2　Evaluation results of each category in the KITTI dataset validation set

Method	Type	Car_{R40}			$Pedestrian_{R40}$			$Cyclist_{R40}$		
		Easy	Mod	Hard	Easy	Mod	Hard	Easy	Mod	Hard
PointPillars	One	88.58	78.33	75.38	55.36	48.86	44.70	80.47	62.70	58.65
Center-Point	One	87.19	80.15	78.36	56.87	52.76	48.46	85.02	67.71	64.29
3DSSD	One	**91.43**	82.23	77.81	62.88	56.93	52.09	88.74	71.34	67.01
DBQ-SSD	One	90.43	82.55	79.66	56.78	52.35	47.17	92.49	70.27	66.40
SECOND	One	89.98	81.07	78.12	56.51	51.92	46.67	81.97	66.83	63.11
DSVT	One	88.83	80.55	78.53	57.58	52.47	48.46	90.28	71.45	67.04
PointRCNN	Two	89.21	80.38	77.98	64.87	56.35	49.31	88.33	68.38	64.00
IA-SSD	One	91.34	**83.70**	**79.89**	60.64	55.90	50.30	89.47	71.41	67.18
本文	One	90.65	83.06	79.88	**65.45**	**59.33**	**53.86**	**92.97**	**72.76**	**68.19**

从表 2 中结果来看，与 IA-SSD 及其他方法相比，在大多数指标中，本文提出的算法在 KITTI 数据集上取得了更好的性能。具体而言，相较于 IA-SSD 而言，在行人三个难度中分别提升了 4.81、3.43、3.56，在骑行者三个难度中分别提升了 3.50、2.35、0.31。

值得注意的是，在汽车类别的简单和中等难度上比 IA-SSD 下降了 0.69、0.64。在困难难度上几乎没有下降，这是由于本文提出的各个模块主要针对小目标做出改进，对车辆目标的特征有所减弱。而困难难度的车辆也像行人一样需要细粒度的信息，下采样方法一定程度上应该提高了困难汽车类别的采样点数，但是其他 2 个模块（InvMLP、MSFA）一定程度上对特征有削弱。

行人和骑行者类别上的巨大提升进一步说明本文提出的方法有利于提升网络在小目标上的检测精度。此外，KITTI 数据集验证集 mAP 对比结果见表 3。表 3 中，将 IA-SSD 和 SGP-SSD 在简单、中等和困难 3 种难度上的平均精度对比发现，本文提出的方法在 3 种难度上的 mAP 值均高于 IA-SSD。

表 3　KITTI 数据集验证集 mAP 对比结果

Table 3　Experimental results of mAP in KITTI dataset validation set

方法	Easy	Mod	Hard
IA-SSD	80.48	70.34	65.79
本文	83.02	71.72	67.31

（2）基于 Waymo 数据集的评估。Waymo 数据集验证集中各类别评结果见表 4。表 4 中，本文提出的算法在行人目标 Level_1、Level_2（Level_1：如果点数大于 5 并且在发布的数据中未标记为 Level_2。Level_2：如果点数大于等于是 1 且小于等于 5，或者在发布的数据中标记为 Level_2。）难度级别上的 AP/APH 指标提升了 2.01/3.17、3.43/1.97，在骑行者目标 Level_1、Level_2 难度级别上的 AP/APH 指标提升了 2.05/2.66、1.78/2.11。而对于 Level_2 的 AP/APH 相比 IA-SSD 却上升了 0.92/0.32，因为困难难度的车辆也像行人一样需要细粒度的信息，所以网络针对小目标的改进也提升了困难的车辆检测效果。

表 4　Waymo 数据集验证集中各类别评估结果

Table 4　Evaluation results of each category in the Waymo dataset validation set

Methods	Vehicle		Pedestrian		Cyclist	
	Level_1 3D AP/APH	Level_2 3D AP/APH	Level_1 3D AP/APH	Level_2 3D AP/APH	Level_1 3D AP/APH	Level_2 3D AP/APH
PointPillars	70.44/69.53	61.91/61.35	67.36/49.26	59.30/43.26	59.96/55.29	55.77/53.20
SECOND	**70.96/70.34**	**62.58/62.02**	65.23/54.24	57.22/47.49	57.13/55.62	54.97/53.53
IASSD	70.13/69.37	60.96/60.53	69.08/57.32	59.84/50.26	66.63/64.73	64.45/62.39
SGP-SSD	70.06/69.11	61.88/60.85	**71.09/60.49**	**63.27/52.23**	**68.68/67.39**	**66.23/64.56**

综合在 2 个数据集上的结果而言,本文提出的算法 SGP-SSD 相对于 IA-SSD 表现出更高的检测精度。这些结果证明了算法的有效性和鲁棒性,在不同难度级别下都能取得较好的性能。

3.3 消融实验

在本小节中对提出的语义引导的下采样模块、特征增强模块和投票增强模块皆以代表性的中等难度精度在 KITTI 数据集上做了各种消融实验,并分析了各个模块对模型的影响。各个模块的增量实验见表 5。

表 5 KITTI 验证集各个模块的增量实验结果

Table 5 Incremental experimental results for each module of the KITTI validation set

S-FPS	INS	MSFA	Car	Pedestrian Moderate	Cyclist
			83.70	55.90	70.41
√			83.28	57.37	71.27
	√		83.22	57.12	71.03
		√	83.50	57.48	70.20
√	√		83.36	58.26	71.63
√	√	√	83.06	**59.33**	**72.76**

(1)语义引导的下采样模块。由表 5 可知,本文使用的 S-FPS 下采样方法提升了基线网络在行人和骑行者类别上的检测精度,尤其是行人这样的小目标。S-FPS 可以使得采样的点不集中于易于识别的对象(汽车),从而从不容易识别的目标(行人,骑车者)中采集到了更多的点,而且汽车对象的精度有所下降,是因为少采集了一部分前景点,造成特征的减少。

语义引导的下采样模块参数对比实验结果见表 6。在表 6 中,相对于基线网络,本文只改变了下采样方法,并设置不同的语义平衡因子 γ 值,来探究最有效的设置。结果表明,过大或过小的 γ 都不能适当地提高检测精度。如前所述,如果 γ 变得非常大,S-FPS 将近似于前景分数的 Top-K 选择。采样的关键点可能会挤在少数容易识别的实例中,而无法覆盖遥远或被遮挡的实例。另一方面,当 γ 接近于 0 时,S-FPS 会退化为普通 FPS,改进有限。适当的 γ 能显著提高其性能。当 $\gamma = 1$ 时,3 个难度等级同时达到较为满意的性能。

(2)特征增强模块。由表 6 第 3 行可知,特征增强模块有效地提升了模型在行人和骑行者类别上的检测精度,证明该模块提供了有效的模型缩放,丰富了网络特征提取。但是在汽车类别上却出现了下降,这是因为更广泛的网络往往能够捕获更多细粒度的特征,并且更容易训练,然而极宽但较浅的网络往往难以捕获更高级的特征[20]。

表 6 语义引导的下采样模块参数对比实验

Table 6 Comparative experiment on parameters of semantically guided downsampling module

Sampling Method	Car	Pedestrian Moderate	cyclist
D-FPS	82.34	49.27	65.28
Ctr-aware	82.82	56.18	71.85
FPS	82.51	53.45	65.84
S-FPS-0.1	82.97	54.51	66.74
S-FPS-0.5	**83.52**	54.83	66.28
S-FPS-1	83.06	**59.33**	**72.76**
S-FPS-2	83.37	57.20	68.54
S-FPS-5	83.38	58.02	69.46
S-FPS-10	83.27	57.16	69.53

(3)多尺度特征聚合模块。由表 6 中数据可知,对车辆、行人和骑行者都有精度的提升,证明该模块可以聚合不同尺度、不同层次的特征,对不同目标都有特征的增强。

综前所述,本文进一步提出猜想,第二次经过 D-FPS 采样的点云更为密集,是否更适合小目标特征聚合,因此本文做了相关实验,实验结果见表 7。

表 7 多尺度特征聚合模块采样点云分析

Table 7 Analysis of sampled point clouds by multi-scale feature aggregation module

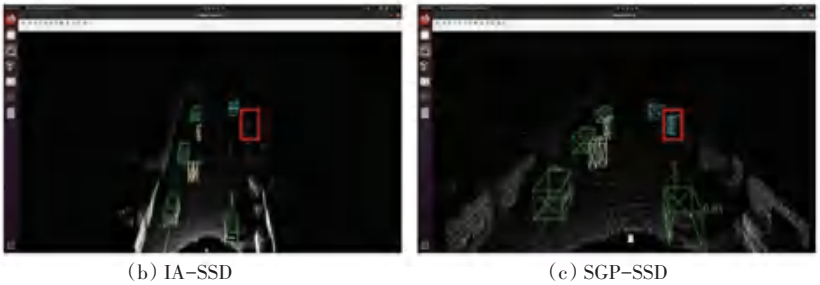
小半径采样层	Car	Pedestrian Moderate	Cyclist
2	82.72	57.24	70.48
3	83.06	59.33	72.76

由实验结果可知第二次下采样的点用于小目标检测并不理想,本文推测是因为未经过语义引导采样的点云还是存在大量背景点影响前景特征提取。

3.4 可视化分析

本文算法在 KITTI 上的可视化效果如图 6、图 7 所示。图 6、图 7 中,图 6(a)、图 7(a)是真实场景下的相机图像,图 6(b)、图 7(b)是 IA-SSD 算法的点云目标检测效果图,图 6(c)、图 7(c)是本文算法 SGP-SSD 的点云目标检测图。在点云检测图中,绿色框代表汽车类别的三维检测框,蓝色框代表行人类别的三维检测框,黄色框代表骑行者类别的三维检测框。

（a）道路场景的相机图像

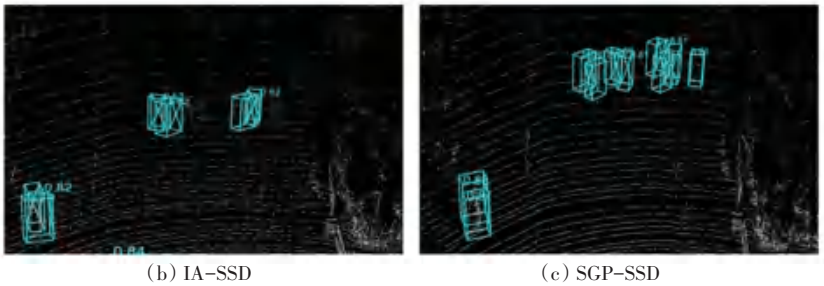


（b）IA-SSD　　（c）SGP-SSD

图 6　道路场景可视化结果对比图

Fig. 6　Comparison chart of road scene visualization results

（a）行人场景的相机图像



（b）IA-SSD　　（c）SGP-SSD

图 7　行人场景可视化结果对比图

Fig. 7　Comparison chart of pedestrian scene visualization results

在图 6 中，2 种算法都有效地检测出汽车、骑行者目标，但是 IA-SSD 没有检测出远处的行人目标，SGP-SSD 在红色框处检测出了行人目标。这证明本文提出的算法相比 IA-SSD 能够更加有效地提取远处行人目标的特征并检测出目标框。

在图 7 中，对于正前方的聚集行走的 8 个行人，IA-SSD 仅仅检测出了 4 个行人，SGP-SSD 则将 8 个行人全部检测了出来。此外，对于左下角的 2 个行人，IA-SSD 虽然输出了 2 个目标框，但是其中行人的方向预测错误，本文算法 SGP-SSD 则获得了正确的判断。这更加证明本文提出的算法相比 IA-SSD 检测精度更高，效果更好。

综上所述，本文提出的算法在 KITTI 场景点云中能够有效地检测出车辆、行人和骑行者，并且相比于 IA-SSD，在行人方面表现出更好的性能。本文算法能够准确地区分出点云中密集的行人群体，避免误判或漏检的情况。这使得算法在高密度行人区域的检测任务中表现出更高的准确性和鲁棒性。

4　结束语

本文对自动驾驶点云目标检测在检测行人方面的局限，通过在基于原始点的 3D 目标检测网络 IA-

SSD 中引入语义引导的点采样方法,综合考虑点的语义信息和距离信息,使网络将关注点从更易识别的车辆大目标转移到行人这样的小目标上。同时在经典的点云特征提取结构 SA 模块中加入了反向瓶颈结构,以实现高效的模型缩放,增加模型的特征提取能力。为了改善原网络特征单一、聚合特征模块聚合半径大、对小目标不友好等状况,提出多尺度特征聚合模块,有效地融合了不同尺度和层次的特征。提出的基于语义引导的点云行人目标检测算法,对行人、骑行者目标的各个难度的检测均有一定的提升。

参考文献

[1] 李磊磊. 自动驾驶汽车产业发展研究及展望[J]. 汽车文摘, 2023(9):1-10.

[2] MAO Jiageng, SHI Shaoshuai, WANG Xiaogang, et al. 3D object detection for autonomous driving: A comprehensive survey [J]. International Journal of Computer Vision, 2023, 131(8): 1909-1963.

[3] HOSS M, SCHOLTES M, ECKSTEIN L. A review of testing object-based environment perception for safe automated driving [J]. Automotive Innovation, 2022, 5(3): 223-250.

[4] 赵建辉. 智能驾驶多模态感知方法研究[D]. 北京:清华大学, 2022.

[5] 邓迪杭. 基于激光雷达的 3D 目标检测算法研究[D]. 重庆:重庆大学,2022.

[6] 吴一全,陈慧娴,张耀. 基于深度学习的三维点云处理方法研究进展[J]. 中国激光,2024,51(5):0509001.

[7] SHI Shaoshuai, WANG Xiaogang, LI Hongsheng. PointrCNN:3D object proposal generation and detection from point cloud [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019:770-779.

[8] YANG Zetong, SUN Yanan, LIU Shu, et al. 3Dssd: Point-based 3D single stage object detector[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 11040-11048.

[9] ZHANG Yifan, HU Qingyong, XU Guoquan, et al. Not all points are equal: Learning highly efficient point-based detectors for 3D lidar point clouds [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 18953-18962.

[10] CHEN Chen, CHEN Zhe, ZHANG Jing, et al. Sasa: Semantics-augmented set abstraction for point-based 3D object detection [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022,36(1):221-229.

[11] 刘洋宏,付杨悠然,董性平. HDMapFusion:用于自动驾驶的多模态融合高清地图生成[J/OL]. 计算机工程. [2025-04-29]. https://doi.org/10.19678/j.issn.1000-3428.0070569.

[12] 霍威乐,荆涛,任爽. 面向自动驾驶的三维目标检测综述[J]. 计算机科学,2023,50(7):107-118.

[13] MAO Anqi, MOHRI M, ZHONG Yutao. Cross-entropy loss functions: Theoretical analysis and applications [C]// International Conference on Machine Learning. Cambridge, MA: PMLR, 2023: 23803-23828.

[14] QIAN Guocheng, LI Yuchen, PENG Houwen, et al. Pointnext: Revisiting PointNet++ with improved training and scaling strategies [J]. Advances in Neural Information Processing Systems, 2022, 35: 23192-23204.

[15] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? the kitti vision benchmark suite [C]// Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway,NJ:IEEE ,2012: 3354-3361.

[16] SUN Pei, KRETZSCHMAR H, DOTIWALLA X, et al. Scalability in perception for autonomous driving: Waymo open dataset [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway,NJ:IEEE, 2020:2446-2454.

[17] 徐鸿盛. 基于目标检测和图像分割的自动驾驶方法研究[D]. 合肥:合肥大学,2024.

[18] ZHOU Yin, TUZEL O. Voxelnet: End-to-end learning for point cloud based 3D object detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway,NJ:IEEE,2018:4490-4499.

[19] SIMONELLI A, BULO S R, PORZI L, et al. Disentangling monocular 3D object detection [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway, NJ:IEEE,2019: 1991-1999.

[20] KOONCE B. Efficient Net [M]//Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization. Berkeley, USA: Apress, 2021:109-123.