

DOI: 10.19650/j.cnki.cjsi.J2514002

PUDet: 基于生成上采样网络的3D目标检测方法*

许丽梅, 周治国, 周学华

(北京理工大学集成电路与电子学院 北京 100081)

摘要: 基于LiDAR的3D目标检测在性能上表现出色,但前景目标点云分布不均匀,往往削弱了其几何特征的表达。同时,远距离目标通常点数稀少,进一步影响检测效果。为此,提出了一种创新框架PUDet,将生成模型融入判别式检测器中。利用带有先验知识的点云上采样网络,增强前景目标的几何细节,从而帮助检测器实现更精确的预测。PUDet包含2个关键模块:针对近距离目标的LDEM,通过优化点云分布来提升检测效果并降低计算成本;针对远距离目标的DDAM,通过增加点密度更清晰地勾勒物体轮廓。为了验证几何轮廓的优化效果,在增强前后分别对近距离和远距离目标的均匀损失进行了实验对比,证明了LDEM和DDAM的有效性。本研究还通过目标点云的注意力图展示了模型对关键区域的关注程度,从而进一步分析了精度提升的内在机制。在KITTI测试集上的实验结果表明,PUDet将基线模型CT3D的mAP提升了1.84个百分点。本研究为3D目标检测领域提供了一种新的方法,并为自动驾驶等应用场景中的精确目标识别和处理提供了更准确、可靠的支持。

关键词: 3D目标检测;自动驾驶;点云上采样;激光雷达(LiDAR)点云

中图分类号: TH744 TP391 文献标识码: A 国家标准学科分类代码: 520.20

PUDet: Advancing 3D object detection with generative upsampling networks

Xu Limei, Zhou Zhiguo, Zhou Xuehua

(School of Integrated Circuits and Electronics, Beijing Institute of Technology, Beijing 100081, China)

Abstract: LiDAR-based 3D object detection achieves superior performance. However, the unevenly distributed point clouds on foreground objects can weaken their geometric representation. Meanwhile, far-away objects typically have very few points, which further impairs detection performance. In this article, a novel framework PUDet is presented, which integrates generative models into discriminative detectors. A point cloud upsampling network is leveraged with prior knowledge to enhance the geometric details of foreground objects, aiding the detector in achieving more accurate prediction. PUDet incorporates two key modules: LDEM for nearby objects, which optimizes point distribution while minimizing computational costs, and DDAM for distant objects, which increases point density to better delineate object contours. To evaluate the optimization of geometric contours, the uniform loss of close and long-distance targets before and after enhancement is experimentally compared, showing the efficacy of LDEM and DDAM. This article also displays the attention maps on object point clouds, explaining the observed accuracy gains. Experimental results on the KITTI testing set show that the proposed framework improves the baseline CT3D by 1.84 mAP, confirming the effectiveness of PUDet. This work introduces a novel approach to 3D object detection, enhancing precision and reliability in object recognition for applications like autonomous driving.

Keywords: 3D object detection; autonomous driving; point cloud upsampling; light detection and ranging (LiDAR) point clouds

0 引言

3D目标检测算法通过实现精确的物体识别与检测,提升了智能城市和城市区域中自动驾驶车辆的安全

性^[1-4]。当前基于LiDAR的3D检测器通过提取采集到的点云特征,并通过检测头对边界框进行分类和回归^[5-7]。LiDAR通过激光脉冲测距生成线性或带状点云,覆盖近距离前景目标的表面。然而,这些点云的几何分布通常不均匀,导致形状信息表达不足。此外,点云密度会随着距离

的增加显著降低, 远处目标通常以极为稀疏的点云形式呈现, 有时甚至仅由寥寥数点构成^[8-9]。针对这些形状信息不足且稀疏的点云数据, 将其转化为均匀分布且密集化的格式, 不仅能够提供更全面的表面覆盖, 还能显著提升特征提取的质量, 从而直观地增强检测器的精度。

生成模型在从受损数据中重建精细结构方面表现出非凡的能力, 这在图像超分辨率任务中尤为显著^[10-13]。通过将低分辨率、带有噪声的图像转换为视觉上更加清晰、细腻的高分辨率图像, 这些模型揭示出原本隐藏或模糊的细微细节, 能够赋予数据新的生命。同样地, LiDAR 点云数据在远距离或被遮挡的物体上常常表现得稀疏且不完整, 这种特性限制了后续感知检测器的有效性。受图像超分辨率任务的启发, 生成模型可以通过对点云数据的“上采样”提供一种解决方案^[14-16]。这些模型不仅能够重建缺失或遮挡区域, 还可以通过增强细节和提升精度来丰富数据集, 类似于将低分辨率图像转换为高清晰度的输出。增强后的点云数据蕴含更丰富的几何细节, 能够帮助判别模型更精确地解析和理解复杂的空间结构, 从而显著提升其在复杂环境中的性能表现。

本文提出了点云上采样 3D 目标检测框架点云上采样 3D 检测器 (point cloud upsampling 3D detector, PUDet), 旨在提升自动驾驶环境中 3D LiDAR 点云的分辨率。由于场景中点云数据量巨大, 直接对整个数据集进行上采样在计算成本上往往是难以承受的。因此, PUDet 通过区域提议网络 (region proposal network, RPN) 识别出感兴趣区域 (region of interest, ROI), 并采用选择性上采样策略来增强这些区域内点云的表示效果。为优化计算资源, PUDet 在近距离前景物体上使用最远点采样方法 (farthest point sampling, FPSampling), 以确保增强后的点数保持在合理范围内, 并均匀分布于物体表面, 从而提升几何表达的准确性。此外, 该方法还能够有效增强远距离稀疏点云的分布密度。最终, PUDet 将优化后的点云输入检测器, 用于后续的边界框分类与精细化回归。

本文的主要贡献包括:

1) 无需依赖图像数据的高效密度增强方法; 不同于以往利用 2D 图像检测结果生成虚拟点以增强目标的方法^[8], 或使用深度补全网络对整个场景进行密集化处理^[17], PUDet 完全独立于图像数据。无需额外的预处理, 直接利用精心训练的点云上采样网络, 专注于强化前景目标的潜在细节表达。

2) 通过生成模型优化数据分布: 提出了一种简单而高效的方法, 将点云生成模型与判别式 3D 目标检测相结合。对于近距离目标, 通过减少冗余点云, 优化了空间分布, 在更少的点云中精确捕捉关键几何特征; 而对于远距离目标, 则通过增加稀疏区域的点云密度, 丰富关键结构信息, 显著缓解了因距离差异导致的点云密度失衡问题,

增强了对远距离目标的几何表达能力。

3) 适用于现有当前最优算法 (state of the art, SOTA) 检测器的即插即用模块: 本文的框架能够作为现有 3D 检测器的即插即用模块。本文算法在 KITTI 数据集^[18]上进行了广泛实验 (详见第 4 章), 验证了 PUDet 算法各组件的有效性。增强后的点云分布显著提升了现有 SOTA 模型 CT3D^[19] 检测器的性能, 在 KITTI 数据集上带来了 1.84 平均精度 (mean average precision, mAP) 的提升, 并在 nuScenes 数据集^[20]上实现了 2.6 mAP 的提升。

1 相关工作

1.1 基于 LiDAR 的 3D 目标检测

在 3D 目标检测任务中, 稀疏点云通常作为主要输入用于生成边界框预测^[21-28]。单阶段模型^[29-32]通过直接从点云中提取全局和局部几何特征, 快速完成目标的分类与定位。Pillarnext^[33]通过优化数据聚合来提升检测精度, 而 Voxelnext^[34]则利用纯稀疏体素特征, 避免了因密集转换带来的高计算成本。两阶段检测方法^[7, 35-38]在第 1 阶段充分利用原始点云及其提取的初步特征生成候选区域, 并在第 2 阶段对这些候选区域进行精细化处理, 从而显著提升检测的精度和鲁棒性。在本文中, 算法基于先进且性能鲁棒的两阶段检测器 CT3D^[19], 通过增强点云分布, 展现了显著的性能提升。

1.2 点云增强方法

点云因形状缺失、稀疏性及分布不均, 削弱了其对 3D 场景目标的几何表征, 进而影响了检测器对其特征的解析。一些跨模态交互方法^[39-41]将稀疏点云与预训练图像分割网络生成的语义特征相结合。然而, 这类方法受限于分割模型的稳定性, 并依赖复杂的初始数据预处理, 限制了其实用性。本文方法在训练过程中直接对点云进行上采样, 无需依赖图像数据。另一种方法是将三原色光 (red-green-blue, RGB) 信息映射到 3D 空间, 通过深度补全生成虚拟点^[8, 17, 42], 部分研究通过优化设计减少虚拟点过度密集的问题并过滤噪声^[4]。相比之下, 本文方法不仅能够精确控制目标区域的点云数量, 还利用具有先验知识的生成模型, 有效避免了噪声点的引入。

与本文方法最接近的是一些基于 LiDAR 的检测器, 这些方法通过手工设计模块生成语义点云, 以缓解稀疏性和遮挡问题^[6, 43-46]。然而, 这类方法高度依赖训练过程的质量和稳定性, 对复杂场景的适应性有限。相较而言, 本文引入在通用数据集上预训练的生成网络, 并在训练过程中固定其权重。在加快了模型收敛速度的同时, 还提升了网络对真实世界前景物体多样性和复杂性的泛化能力。

1.3 可学习的点云上采样

近年来的研究重点集中于从低分辨率输入中提取细粒度的语义特征,生成更密集的点集以重建物体的基础表面^[14-16,47-48]。然而,这类模型在训练完成后通常只能支持固定的上采样率。相比之下,Grad-PU^[49]通过将上采样过程分为中点插值和位置细化,成功实现了对任意上采样率的适配。在本文中,模型采用这一上采样模型来改善点云分布,并填补前景区域中的小空隙,有效提升了3D检测器的性能。

2 PUDet 网络

本文提出了一种名为点云上采样3D目标检测

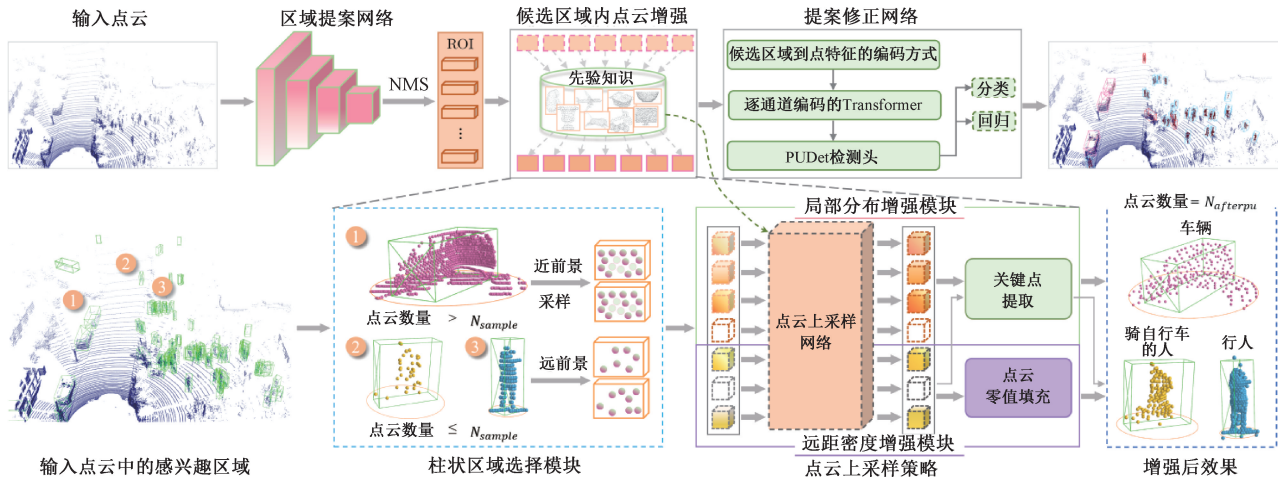


图1 PUDet 框架

Fig.1 PUDet framework

2.1 柱状区域选择模块

由于自动驾驶场景中激光雷达点云数据覆盖了大范围的空间,其中包括大量背景信息,若对整个场景的点云实施无差别的上采样,将不可避免地引入大量冗余信息与计算开销,降低算法的效率与精度。一种更有效的策略是利用两阶段检测框架中区域提案网络(RPN)的输出结果,初步筛选出可能含有前景目标的候选区域,并集中增强其中的点云。尽管RPN生成的候选区域通常存在覆盖不全或定位误差等问题,但它们仍提供了较为可靠的信息,使后续增强过程能够更加精准地聚焦于目标区域内。本文基于此设计了柱状区域选择模块,通过在感兴趣区域上构建垂直延伸的柱状体,以更完整地捕获潜在的目标点云,并根据区域内点云密度自适应地进行目标分类与采样,为后续差异化的上采样处理提供高质量的前景点云。

具体来说,柱状区域选择模块包括柱状区域提取与自适应分类采样2个关键步骤:

1) 柱状区域提取:考虑到RPN生成的初始3D候选

(PUDet)框架,具有简洁且高效的设计,旨在提升两阶段3D目标检测架构的性能。使用在通用数据集上预训练好的生成式点云上采样网络,PUDet对候选区域内的数据进行选择性增强,以提升网络对目标的分类与定位能力。如图1所示,算法的候选区域增强过程从柱状区域选择模块开始。柱状区域选择模块根据候选框中的点云,判断每个框内是否包含近前景或远前景目标,以此调整相应的采样策略。随后,局部分布增强模块(local distribution enhancement module, LDEM)用于优化近距离目标的点云分布,强化关键细节以获取更加精确的几何轮廓;远距密度增强模块(distant density augmentation module, DDAM)针对远距离目标的稀疏点云进行密度提升,补充更多有效的几何信息。

框在经过非极大值抑制(non maximum suppression, NMS)处理后,虽然能够有效去除冗余的提案框,但产生的感兴趣区域(RoI)往往难以准确捕捉完整的前景目标,尤其是在目标轮廓附近的点云容易丢失。为了增强候选区域的完整性,避免边界截断导致的前景信息损失,本文采用柱状体对RoI区域进行了扩展。具体来说,对于每个候选区域 roi_box_i ,其中心坐标为 (x_i^b, y_i^b, z_i^b) ,尺寸为 (l_i^b, w_i^b, h_i^b) ,定义一个无限高度的圆柱体,其半径为:

$$r_i = \alpha \sqrt{\left(\frac{l_i^b}{2}\right)^2 + \left(\frac{w_i^b}{2}\right)^2} \quad (1)$$

式中: α 为超参数,表示柱状区域的系数。

对于场景中的每个点云数据 $p = (x, y, z)$,若满足式(2),则该点被归入 roi_box_i 内的点云集合 \mathcal{N}_i 。

$$\sqrt{(x - x_i^b)^2 + (y - y_i^b)^2} \leq r_i \quad (2)$$

使用柱状体扩展候选区域进行点云提取具备明显优势:首先,通过适度扩展候选区域,有效减少了因RPN预

测误差导致的目标边缘点云截断问题,提升了点云的完整性;其次,柱状体在水平方向上覆盖更广,能够更好地适应目标姿态变化,使点云增强过程更具鲁棒性;此外,柱状体在高度方向上不受限制,避免了传统 3D 边界框因目标类别差异导致的高度截断问题,确保所有潜在前景点云都被完整保留,为后续的点云增强提供更加丰富的输入数据。

2) 自适应分类采样: 由于激光雷达点云的密度随目标距离的增加而急剧下降,不同目标实例内的点云数量存在显著差异。近前景目标通常包含大量点云,而远前景目标则极为稀疏。这种不均衡性若不经处理,将导致上采样时对高密度目标的冗余增强,以及对低密度目标信息的不充分补充,背离算法的初衷。因此,本模块进一步对提取的柱状区域点云进行自适应分类采样,以实现针对性的点云增强策略。

具体而言,设定点云数量的阈值为 N_{sample} , roi_box_i 可以分为两类:

(1) 若 roi_box_i 内的点云数量超过 N_{sample} (包含等于的情况), 则视作近前景区域, 即 $roi_box_i^{near}$ 。由于近前景区域点云较为密集, 直接使用全部点云可能导致冗余信息过多, 并增加计算负担, 因此采用最远点采样算法 FPSampling, 从 \mathcal{N}_i 中采样 N_{sample} 个点, 得到点集 \mathcal{N}_i^{near} 。相比于随机采样, FPSampling 在点云分布的均匀性上具有显著优势, 能够有效覆盖目标表面, 避免信息集中于局部区域, 从而确保降采样后的点云仍然能够完整表达目标的几何结构。

(2) 若 roi_box_i 内的点云数量低于 N_{sample} , 则视作远前景区域, 即 $roi_box_i^{far}$ 。由于远前景区域点云本身极为稀疏, 进一步采样会导致关键信息丢失, 因此直接将原始点云集合 \mathcal{N}_i 作为最终点集 \mathcal{N}_i^{far} 。

自适应分类采样过程不仅提高了计算效率, 同时使近前景区域在降采样后仍能保留完整的几何信息, 远前景区域的稀疏点云得以充分利用, 避免关键结构信息的丢失。接下来将基于此进一步优化点云分布, 使不同的前景目标具备更加完整的几何特征, 从而提升检测的精确度与稳定性。

2.2 点云上采样策略

如 1.2 节分析, 以往的研究通常依赖多模态输入, 如结合点云和图像的特征, 通过图像分割或深度补全生成补充信息。然而, 这类方法不仅在网络训练前需要进行大量复杂的数据预处理, 还高度依赖于多模态特征的对齐质量和单一模态的稳定性, 限制了其在实际场景中的适用性。本文方法通过引入融合了丰富先验知识的高质量生成点云上采样网络, 以端到端的方式进行点云优化。该过程由 2 个关键模块组成, 分别针对点云的分布均匀性与稀疏性问题进行针对性增强, 从而提升点云的几何

细节表达和目标检测性能。下面, 首先简要概述点云上采样网络的关键步骤, 然后构建局部分布增强模块和远距密度增强模块, 以改善数据分布并丰富几何信息。

1) 点云上采样网络

点云上采样旨在通过学习训练数据中的几何特征分布, 将给定的稀疏点云转化为密度更高且更加精细的点云集合, 以更精细地描述物体表面的几何结构。这一问题在本质上与图像超分辨率任务类似, 都是通过已有的数据分布对稀疏或退化数据进行细节重建, 但点云数据自身的不规则性和无序性使其面临独特挑战。与图像空间中规则、连续排列的像素不同, 点云数据在空间分布上缺乏明确的拓扑结构, 难以直接应用传统基于规则网格的插值或卷积操作。此外, 为了高质量地还原三维目标的真实几何结构, 新生成的点不仅需要准确地落在物体表面, 还需要遵循原始数据中隐含的局部和全局几何特征分布, 以避免几何形态的扭曲或不一致性。因此, 点云上采样的核心难点在于如何从无序且不规则的数据中学习稳定而泛化的几何表示。

Grad-PU 算法通过一种两阶段的框架解决了上述点云上采样中的关键问题^[49]。首先, 网络将上采样过程解耦为 2 个步骤: 在欧几里得空间中通过中点插值直接对输入的低分辨率点云进行上采样, 而非在特征空间中进行扩展, 这使得特征学习网络与上采样模块分离, 从而支持任意上采样率; 其次, 将插值后的粗糙点云与高分辨率真实点云之间的差异建模为逐点距离的优化问题, 实现对任意拓扑结构的有效处理, 缓解了传统方法中因几何位置估计误差导致的噪声问题。

具体来说, 给定一个包含 N 个点的点集 $\mathcal{P} = \{p_i\}_{i=1}^N$, 点云上采样网络的目标是生成一个包含 rN 个点的密集、均匀, 且高保真的点集 $\mathcal{Q} = \{q_j\}_{j=1}^{rN}$, 其中 r 是上采样率。首先, 利用中点插值方法对输入的低密度点云 \mathcal{P} 进行空间上的密度增强。具体地, 对于每一个输入点 $p \in \mathcal{P}$, 先通过 K 近邻算法 (K -nearest neighbors, KNN) 在欧氏空间内搜索得到其邻域点集合:

$$\mathcal{P}_k(p) = \{p_1, p_2, \dots, p_k\} \quad (3)$$

随后, 对每个点 p 与其邻域点集合中的每个点 $p_l \in \mathcal{P}_k(p)$, 计算中点作为候选插值点, 即:

$$p_{interp} = \frac{p + p_l}{2}, \forall p_l \in \mathcal{P}_k(p) \quad (4)$$

在此过程中生成了大量的初始插值点, 然而这些点的分布存在冗余且不均匀的问题。因此, 通过最远点采样 (FPSampling) 进一步从这些插值点中挑选出具有代表性的 rN 个点作为中间差值点云集合 \mathcal{P}_l :

$$\mathcal{P}_l = FPS\left(\left\{\frac{p + p_l}{2}, \forall p_l \in \mathcal{P}_k(p), \forall p \in \mathcal{P}\right\}, rN\right) \quad (5)$$

经过初始插值后得到的点云 \mathcal{P}_l 仍然存在位置不准

或分布不均的问题,需要通过细化步骤对插值点的位置进行进一步修正,以更贴近高分辨率真实点云 \mathcal{P}_c 。为了实现这一目的,可以将点云的细化建模为一个迭代优化问题,即通过迭代实现 \mathcal{P}_l 与 \mathcal{P}_c 之间的差异最小化来逐步逼近理想位置。具体而言,对于中间差值点云集合 \mathcal{P}_l 中的每个点 p ,定义其与高分辨率真实点云 \mathcal{P}_c 之间的点对点距离函数,即:

$$F(p) = \min_{\hat{p} \in \mathcal{P}_c} \|p - \hat{p}\|_2 \quad (6)$$

式中: \hat{p} 表示 \mathcal{P}_c 中与 p 距离最近的点; $\|\cdot\|_2$ 表示欧几里得距离。相较于直接预测三维空间坐标, $F(p)$ 这一距离函数更容易学习,且无需显式提取目标物体表面,适用于任意拓扑结构。细化过程的目标是通过迭代优化,逐步减小 \mathcal{P}_l 中的点与 \mathcal{P}_c 之间的 $F(p)$ 。点 p 的位置更新基于

梯度下降进行迭代优化,公式为:

$$p^{t+1} = p^t - \lambda \nabla F(p^t), t = 0, \dots, N_{iter} - 1 \quad (7)$$

式中: p^0 表示初始插值点集 \mathcal{P}_l 中的点; p^t 是第 t 次迭代后的位置; λ 表示迭代步长; $\nabla F(p^t)$ 是点对点距离函数 $F(p^t)$ 的负梯度,表示插值点向高分辨率点云 \mathcal{P}_c 位置移动的最佳方向。

由于在推理阶段无法获得真实点云集合 \mathcal{P}_c ,因此需要一种能够近似点对点距离函数 $F(p)$ 的机制来指导插值点云的细化过程。如图2所示,点对点距离网络(point-to-point distance network, P2PNet)由2个主要模块组成:特征提取器和距离回归器,输入插值点云集合 \mathcal{P}_l ,最终产生每个插值点与目标真实点云表面之间的距离预测,从而引导插值点逐步逼近潜在物体的表面。

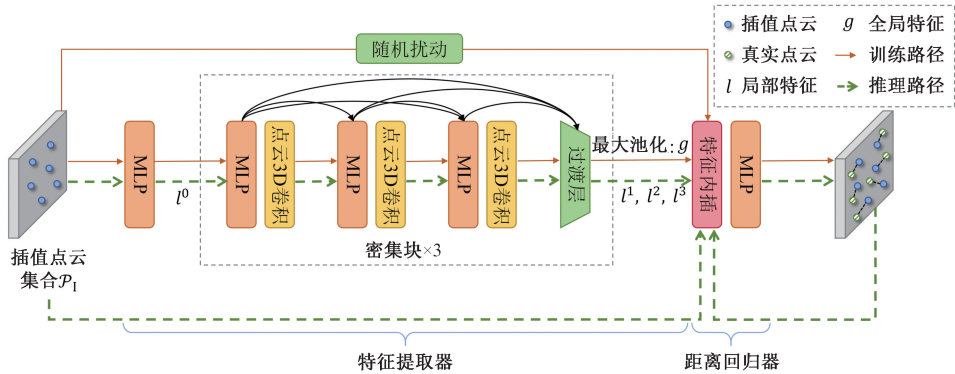


图2 点对点距离网络

Fig. 2 Point-to-point distance network

特征提取器利用层次化的特征提取机制,获取插值点 $p \in \mathcal{P}_l$ 的局部特征与全局特征,以充分表征局部区域和整体几何结构的信息。具体来说,首先通过多层感知机(multilayer perceptron, MLP)将 \mathcal{P}_l 投影到高维空间,获得局部特征 l^0 ;然后通过3个密集块分别提取出局部特征 $\{l^1, l^2, l^3\}$,其中每个密集块包含3个卷积组(MLP+点云3D卷积^[49])和1个过渡层。最终通过最大池化操作获得全局特征 g 。随后,距离回归器基于上述提取的丰富特征,通过一个4层的MLP预测每个插值点与目标真实表面之间的点对点距离 $F_N(p)$,以指导点云位置的精细化调整。为提升模型的鲁棒性和泛化性能,P2PNet在训练过程中还引入了随机扰动策略,以模拟真实数据中可能存在的采样误差与不确定性。由式(7),引入P2PNet的迭代过程可以表示为:

$$p^{t+1} = p^t - \lambda \nabla F_N(p^t), t = 0, \dots, N_{iter} - 1 \quad (8)$$

上述迭代优化过程进行 N_{iter} 次,插值点逐渐收敛到目标物体的真实表面位置,获得空间分布更加均匀、几何细节更加完整的高质量点云 $\mathcal{P}_l^{N_{iter}}$,记为 \mathcal{P}_U 。

2) 先验知识

本文使用在PU1K数据集^[50]上训练的上采样模型进行点云上采样。PU1K数据集包含1147个3D模型,涵盖了与自动驾驶场景中动态前景目标(如车辆、行人和骑行者)几何形状相似的物体。通过在该数据集上的训练,点云上采样网络掌握了丰富的几何先验知识,能够更好地捕捉和重现目标的几何特征。具体而言,该网络通过细化点云细节,优化点云的分布均匀性,有效填补因遮挡或稀疏性导致的间隙与空洞,同时保持目标的结构完整性。这种生成式上采样方法通过对目标表面细节的高保真重构,为判别式3D目标检测系统提供了更高质量的输入数据。

3) 局部分布增强模块

局部分布优化模块旨在针对近前景区域点云的分布不均及轮廓缺失等问题,通过生成式上采样网络补充细节信息,增强目标的几何一致性,恢复目标完整的结构特征,特别是在因遮挡或扫描不均匀造成的数据缺失区域。该模块主要包含2个处理步骤:点云上采样和关键点提取,以实现近前景区域的点云分布优化和计算效率平衡。

(1) 点云上采样: 为了提升点云质量, 局部分布优化模块首先对经过柱状区域选择模块筛选出的点云 \mathcal{N}^{near} 进行上采样, 该过程应该具备的特点为:

a) 局部结构一致: 新增点云应遵循原始点云的局部几何约束, 避免在局部区域产生异常密集或空洞的情况;

b) 全局形态保持: 上采样后的点云应在整体形态上与目标真实表面一致, 避免几何变形或结构扭曲;

c) 特征信息补全: 对于因遮挡或扫描限制而丢失的区域, 新增点云应能够合理补全, 使得目标的轮廓更加完整。

如 2.1 节所述, Grad-PU 网络在 PU1K 数据集上经过充分训练后, 具有良好的三维几何补全能力, 能够生成符合真实目标拓扑结构的高质量点云。本文利用该网络的生成能力, 将输入点云 \mathcal{N}^{near} 映射至更高密度的表达形式: 首先通过对 \mathcal{N}^{near} 执行中点插值(式(5)), 生成更密集的初始点云分布; 之后基于 P2PNet 的迭代优化(式(7))调整插值点的位置, 使其更均匀地分布在物体表面, 同时补全原始点云中缺失的区域。最终得到增强后的点云集合 $\mathcal{N}_U^{near} \in \mathbb{R}^{rN_{sample} \times 3}$ 。

(2) 关键点提取: 尽管上采样能够补充丢失的几何信息, 但过高密度的点云可能引入冗余信息, 不仅增加计算开销, 还可能干扰检测过程。为此, 模型进一步引入关键点提取步骤, 采用最远点采样算法(FPSampling)从 \mathcal{N}_U^{near} 中筛选出 $N_{afterpu}$ 个关键点, 形成新的点集 $\mathcal{N}_{aug}^{near} \in \mathbb{R}^{N_{afterpu} \times 3}$, 不但能够最大化点云在目标表面的覆盖范围, 同时增强目标轮廓的清晰度, 使检测模型能够更精准地提取形状特征。

4) 远距密度增强模块

远距离目标的点云稀疏性严重削弱了其在 3D 目标检测中的可辨识度, 尤其在深度学习模型高度依赖数据特征的情况下, 有限的点云信息难以提供充分的几何约束, 使得现有检测算法在远前景区域的性能大幅下降^[51]。为了提升远距离目标的几何完整性并增强点云的特征表达, 远距密度增强模块通过生成式上采样策略补全稀疏点云信息, 从而提升检测的鲁棒性。

该模块首先对经过柱状区域选择模块筛选出的点云 $\mathcal{N}^{far} \in \mathbb{R}^{N \times 3}$ 使用 Grad-PU 网络进行上采样, 生成更高密度的增强点集 $\mathcal{N}_U^{far} \in \mathbb{R}^{rN \times 3}$, r 为上采样倍率。然而, 由于不同前景区域的初始点云数量存在显著差异, 单纯按照固定比例扩充点云可能导致部分区域仍然稀疏, 而另一些区域则生成过密点云, 影响几何一致性。这种不均匀性不仅影响目标的空间分布, 还可能在特征学习过程中引入偏差, 降低检测模型对远距离目标的识别能力; 此外, 从信号处理的角度分析, 点云可以被视为对连续目标表面进行的离散采样, 因此其重建质量依赖于采样点的密度分布。根据香农采样定理(Shannon's sampling

theorem)^[52], 若采样频率低于信号的奈奎斯特率(Nyquist rate), 则原始信号无法在不失真的情况下恢复。在点云数据中, 这意味着当某一区域的初始点云过于稀疏时, 即便进行上采样, 也难以有效重建物体的几何结构, 导致形态扭曲或局部信息缺失。因此, 仅依赖固定倍率的上采样策略难以满足不同区域的点云补全需求, 需要根据原始点云的密度自适应调整补全策略, 以保证整体点云分布的均匀性和几何保真度。

为了在不同区域保持一致的点云密度, 本文设置了一个点云数量阈值 $N_{afterpu}$ 以衡量区域的几何信息完整性。当上采样后的点云数量 $|\mathcal{N}_U^{far}|$ 超过 $N_{afterpu}$ 时, 表明该区域的几何信息较为完整, 此时采用最远点采样(FPSampling)从 \mathcal{N}_U^{far} 中筛选出 $N_{afterpu}$ 个空间分布均匀且信息丰富的点, 以优化点云的空间分布, 在保留关键几何细节的同时, 避免了不必要的计算冗余; 而当上采样后的点云数量仍然低于 $N_{afterpu}$ 时, 则说明该区域的原始信息极度缺乏, 即便经过上采样, 仍不足以进行高质量的几何补全。在这种情况下, 为确保网络输入维度的一致性, 本文采用零填充(zero-pad)标准化点数至 $N_{afterpu}$, 以便所有远前景目标区域的点云能够在统一的框架下进行后续处理。上述过程可以公式化为:

$$\mathcal{N}_{aug}^{far} = \begin{cases} FPS(\mathcal{N}_U^{far}, N_{afterpu}), & |\mathcal{N}_U^{far}| \geq N_{afterpu} \\ ZeroPad(\mathcal{N}_U^{far}, N_{afterpu}), & |\mathcal{N}_U^{far}| < N_{afterpu} \end{cases} \quad (9)$$

2.3 候选区域优化网络

在获取 RoI 区域内经过增强的点云后, 可直接引入任意现有二阶段 3D 目标检测器中的候选区域优化网络, 对点云进行特征提取, 并完成边界框的分类与回归预测任务。在本文中, 模型采用 CT3D^[19], 该方法首先通过候选区域到点特征的编码机制, 将候选区域映射为点云特征。在此过程中, 舍弃了点云的反射强度信息, 仅保留三维坐标以专注于几何结构的表达。这些包含候选区域特征的嵌入随后被输入逐通道编码的 Transformer 结构中, 该网络能够捕捉点云之间复杂而丰富的上下文依赖关系, 从而生成最终优化的检测结果。

2.4 训练损失

损失函数(loss function)用于衡量模型预测结果与真实值之间的差异, 并通过梯度反馈引导网络参数的优化方向, 以逐步提升模型性能。在 3D 目标检测任务中, 损失函数通常由分类损失和回归损失两部分组成: 分类损失用于预测目标的类别或存在性, 而回归损失用于优化目标的空间位置、尺寸及方向, 使其更精确地贴合真实目标。在训练过程中, 采用端到端的优化策略, 整体损失函数由 RPN 损失、3D 交并比(intersection over union, IoU)损失、置信度预测损失以及边界框回归损失组成。

RPN 损失 L_{RPN} 用于优化区域提案网络 (RPN), 其中前景-背景分类误差通过二元交叉熵损失 (binary cross entropy, BCE) 进行衡量, 而候选框的坐标回归误差则采用 Smooth-L1 损失进行优化。该损失设计旨在提升提案框的质量, 确保高置信度的候选区域能够准确覆盖潜在目标, 为后续检测提供可靠的基础。

3D IoU 损失 L_{IoU} 通过计算预测框与真实框之间的 3D IoU 来衡量其重叠程度, 并据此优化检测框的定位精度。相比于仅依赖坐标误差的回归损失, 该损失能够直接作用于目标边界, 促进预测框更准确地对齐真实目标, 特别是在尺度变化较大或目标密集的场景下发挥关键作用, 同时加速模型收敛。

置信度预测损失 L_{conf} 采用二元交叉熵损失 BCE 来衡量提案框的目标存在概率, 同时结合 IoU 引导的标签分配策略, 使模型能够更精准地区分前景与背景。具体来说, 该损失根据提案框与真实目标的 IoU 计算匹配度, 并动态调整正负样本的权重, 从而提升高质量 RoI 的置信度评分, 使检测网络更加关注匹配度较高的目标, 同时抑制低质量提案的干扰。

边界框回归损失 L_{reg} 采用 Smooth-L1 损失计算检测框与真实目标之间的偏差, 并仅针对 IoU 超过设定阈值的提案框进行优化。该损失约束预测框在位置、尺寸和方向上逐步逼近真实目标, 从而提升目标边界的回归精度, 确保最终检测结果能够准确对齐目标轮廓。

将上述 4 部分损失结合, 形成最终的损失函数表达式:

$$L = L_{RPN} + L_{IoU} + L_{conf} + L_{reg} \quad (10)$$

3 实验验证

3.1 实验设置

KITTI 3D 和 BEV 目标检测基准数据集^[18]是当前广泛使用的评测标准之一, 包含 7 481 个训练样本和 7 518 个测试样本。在实验中, 本文遵循 MV3D^[21]的数据划分方式, 将训练集划分为 3 712 个训练样本和 3 769 个验证样本。评估指标采用 40 个不同召回率阈值下的三维平均精度 (3D average precision, 3D AP), 并在简单 (easy)、中等 (moderate) 和困难 (hard) 3 个难度级别上进行全面评估。汽车、行人和骑行者的交并比 IoU 阈值分别设定为 0.7、0.5 和 0.5。nuScenes 数据集^[20]是一项涵盖广泛的自动驾驶评测基准, 包含 1 000 个场景, 其中 700 个用于训练, 150 个用于验证, 150 个用于测试。nuScenes 官方提供的 3D 目标检测评估指标包括鸟瞰图定位的平均精度 mAP 和 nuScenes 检测得分 (nuScenes detection score, NDS)。NDS 在结合 mAP 的基础上, 综合

考虑了物体属性 (如尺寸、朝向和位置) 的回归精度, 为检测模型性能提供了更加全面的评价标准。

本文的实现基于开源代码库 OpenPCDet^[53], 并按照 VoxelRCNN^[54], CT3D^[19], 以及 PointRCNN^[55] 中对 3D 点云输入的设置进行操作, 以确保实验的可复现性和一致性。在非极大值抑制 (NMS) 后, 训练阶段选取 128 个 RoI 区域, 而在推理阶段选择 100 个 RoI 区域。在特征提取时, 柱状区域选择模块中的超参数 α 设置为 1.2, 每个区域的采样阈值 $N_{sample} = 256$ 。对于点云上采样网络, 将上采样率 r 设置为 4, 细化迭代次数 N_{iter} 设置为 2。在训练 PUDet 时, 上采样网络的参数被冻结。roi_box 内的点云经过不同模块处理, 最终每个区域包含 $N_{afterpu} = 256$ 个点云。整个模型通过端到端的方式训练, 使用 ADAM 优化器^[56] 和 one-cycle 策略^[57], 训练 80 个 epoch。算法采用余弦退火学习率策略, 最大学习率设置为 0.003 5。训练过程中使用一块 RTX4090 GPU, 批处理大小设置为 4。

3.2 检测指标结果及可视化分析

1) KITTI 测试集

为了验证 PUDet 在不同难度级别下的性能优势, 在 KITTI 测试集上进行了全面的实验评估。如表 1 所示, 所有结果均基于 0.7 IoU 阈值和 40 个召回位置的平均精度 (3D mAP) 进行评估, 基线结果以下划线标注以便突出显示。此外, 表中输入模态中的标记, L 代表激光雷达, I 代表图像。PUDet 在中等 (moderate) 和困难 (hard) 级别上表现尤为突出, 达到了最高的 3D 平均精度, 表明其在处理稀疏点云和复杂几何结构, 尤其是远距离目标检测时, 能有效增强点云细节表达, 提升目标定位与分类精度。

具体而言, 与基线方法相比, PUDet 在所有指标上均实现了性能的全面提升, 在 3D 检测 mAP 上提高了 1.84%。然而, 在简单级别上的表现相对较弱, 这可能是由于每个候选区域内仅包含 256 个原始点, 而该级别中候选区域常包含更多点云信息, 这种限制可能影响了检测精度。尽管如此, PUDet 依然能够在复杂场景和稀疏目标下展现出强大的几何补全与特征优化能力。

本文方法在方向预测任务中也展现出了更高的稳定性, 实验结果如表 2 所示。与其他在 3D 目标检测中应用点云生成的算法, 如 SPG, BtcDet 相比, PUDet 的优势主要来源于上采样网络对几何信息的补全能力, 使得预测的目标朝向更加稳定和精确。然而, 在容易级别上, 本文算法的表现略逊于基线方法, 这可能是由于容易级别的提案框通常包含更多的原始点云, 而 PUDet 主要针对点云稀疏场景进行优化。当提案区域仅包含 256 个原始点时, 点云上采样的收益相对有限, 从而影响了在该难度级别的性能表现。

表 1 各算法在 KITTI 测试集上的性能比较

Table 1 Performance comparison of various algorithms on the KITTI testing set (%)

模型	输入模态	3D 目标检测			
		简单	中等	困难	3D mAP
MV3D ^[2]	L+I	74.97	63.63	54.00	64.20
F-PointNet ^[24]	L+I	82.19	69.79	60.59	70.86
CLOCs ^[58]	L+I	89.16	82.28	77.23	82.89
MMF ^[59]	L+I	88.40	77.43	70.22	78.68
PI-RCNN ^[60]	L+I	84.37	74.82	70.03	76.41
EPNet ^[61]	L+I	89.81	79.28	74.59	81.23
EQ-PVRCNN ^[62]	L+I	90.13	82.01	77.53	83.22
SECOND ^[30]	L	83.34	72.55	65.82	73.90
PointPillars ^[29]	L	82.58	74.31	68.99	75.29
TANet ^[63]	L	84.39	75.94	68.82	76.38
PV-RCNN ^[37]	L	90.25	81.43	76.82	82.83
PartA2 ^[64]	L	87.81	78.49	73.51	79.94
CIA-SSD ^[65]	L	89.59	80.28	72.87	80.91
CT3D ^[19]	L	87.83	81.77	77.16	82.25
SPG ^[6]	L	90.50	82.13	78.90	83.84
BtcDet ^[44]	L	90.64	82.86	78.09	83.86
SASA ^[66]	L	88.76	82.16	77.16	82.69
GD-MAE ^[67]	L	88.14	79.03	73.55	80.24
PG-RCNN ^[68]	L	89.38	82.13	77.33	82.95
PUDet(本文)	L	89.89	83.02	79.35	84.09

2) KITTI 和 nuScenes 验证集

为了进一步验证 PUDet 的通用性和有效性,本文在 KITTI 和 nuScenes 验证集上进行了系统性评估。如表 3 和 4 所示。表中所有结果均基于 0.7 IoU 阈值的平均精度(average precision, AP)进行评估,其中表 4 报告了 NDS, mAP 以及各类别的 mAP,涉及的缩写类别为:施工车辆(construction vehicle, C. V.)、行人(pedestrian, Ped.)、摩托车(motorcycle, Motor.)和交通锥(traffic cone, T. C.)。实验结果显示,PUDet 在多个基准数据集上均表现出卓越的性能,与基线方法相比取得了显著提升,充分证明了其在不同场景和任务中的适用性与鲁棒性。

值得注意的是,在 nuScenes 数据集上,PUDet 在 mAP 指标上提高了 2.6%,在 NDS 指标上提升了 3.3%,表明本文方法在大规模数据集上同样具有良好的适用性。此外,在 0.5 IoU 阈值下的人行道检测和骑行者检测任务中(详见表 5),PUDet 分别将人行道和骑行者检测的

表 2 各算法在 KITTI 测试集上的方向预测准确率比较
Table 2 Comparison of direction prediction accuracy of various algorithms on the KITTI testing set (%)

模型	输入模态	方向预测			
		简单	中等	困难	mAP
MV3D ^[2]	L+I	-	-	-	-
F-PointNet ^[24]	L+I	-	-	-	-
CLOCs ^[58]	L+I	96.77	95.93	90.93	94.54
MMF ^[59]	L+I	-	-	-	-
PI-RCNN ^[60]	L+I	96.15	92.52	87.47	92.05
EPNet ^[61]	L+I	96.13	94.22	89.68	93.34
EQ-PVRCNN ^[62]	L+I	98.22	95.20	92.47	95.30
SECOND ^[30]	L	90.93	82.55	73.62	82.37
PointPillars ^[29]	L	93.84	90.70	87.47	90.67
TANet ^[63]	L	93.52	90.11	84.61	89.41
PV-RCNN ^[37]	L	98.15	94.57	91.85	94.86
PartA2 ^[64]	L	95.00	91.73	88.86	91.86
CIA-SSD ^[65]	L	96.65	93.34	85.76	91.92
CT3D ^[19]	L	<u>96.26</u>	<u>93.20</u>	<u>90.44</u>	<u>93.30</u>
SPG ^[6]	L	40.02	38.73	38.52	39.09
BtcDet ^[44]	L	39.26	38.00	36.82	38.03
SASA ^[66]	L	96.00	95.29	92.42	94.57
GD-MAE ^[67]	L	98.31	95.36	90.19	94.62
PG-RCNN ^[68]	L	96.64	95.27	90.37	94.09
PUDet(本文)	L	97.31	95.96	92.73	95.33

表 3 基线方法与 PUDet 在 KITTI 验证集上车辆检测的性能比较

Table 3 Performance comparison of vehicle detection between baseline and PUDet on KITTI validation set (%)

召回位置	模型	3D 目标检测			
		简单	中等	困难	3D mAP
40	CT3D	92.34	84.97	82.91	86.74
	PUDet	93.51	85.33	84.02	87.62
11	CT3D	89.11	85.04	78.76	84.30
	PUDet	89.74	85.34	78.58	84.61

AP 分别提升了 1.94% 和 1.22%,表明其能够提供更精确的表面细节并更好地建模上下文信息。

3) 可视化对比

如图 3 所示,在 KITTI 测试集上对检测结果进行了定性可视化分析。图中展示了 3 种不同场景的检测结果,对汽车、行人和骑行者进行了检测结果可视化。原始

表4 基线方法与 PUDet 在 nuScenes 验证集上的性能比较
Table 4 Performance comparison on the nuScenes validation set

模型	mAP	NDS	Car	Truck	C. V.	Bus	Trailer	Barrier	Motor.	Bicycle	Ped.	T. C.
CT3D	59.5	65.3	84.8	53.7	19.4	64.2	55.4	72.1	59.5	24.8	83.4	77.9
PUDet(本文)	62.1	68.6	85.4	55.1	26.1	65.7	55.3	76.3	62.6	27.1	86.4	80.5

表5 基线方法与 PUDet 在 KITTI 验证集上针对行人和骑行者检测的性能比较
Table 5 Performance comparison for pedestrian and cyclist detection on the KITTI validation set

模型	行人				骑自行车的人			
	简单	中等	困难	3D mAP	简单	中等	困难	3D mAP
CT3D	61.05	55.57	51.10	55.91	89.01	71.88	67.91	76.27
PUDet	64.32	57.33	51.91	57.85	91.62	72.31	68.53	77.49

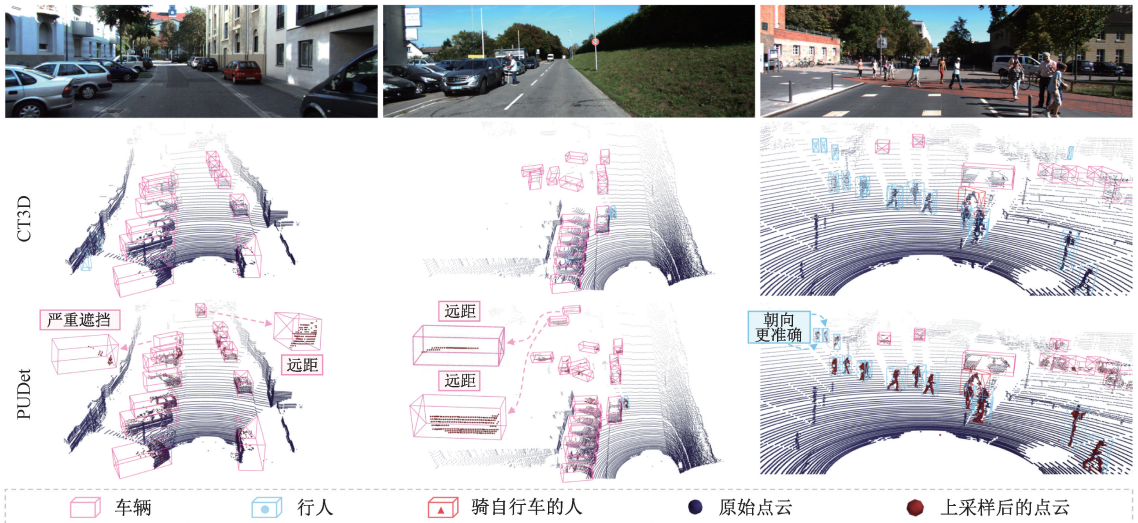


图3 KITTI 测试集的定性结果展示

Fig. 3 Qualitative results on the KITTI testing set

点云以常规样式呈现,经过 PUDet 增强的前景点云以加粗样式呈现(具体样式区分参见图3图例)。本文方法在检测严重遮挡或远距离的前景目标时显著优于基线方法,这些目标通常仅由稀疏点云表示。此外,PUDet 生成的上采样点云揭示了更丰富的上下文信息,从而有效提升了边界框方向预测的准确性。

3.3 几何轮廓优化的效果及影响

本文通过一系列实验验证了 PUDet 上采样在优化点云表现方面的显著优势,实验结果如图4所示。第2行与第4行分别展示了原始点云和经过优化后的点云结果。图4可分为2部分:左侧3列对应简单模式(点云数量接近或远超256),而右侧3列对应困难模式(点云数量显著少于256)。为进一步量化点云分布的优化效果,采用均匀性损失指标 $L_{uniform}^{[47]}$ 进行评估,其中较低的均匀性损失值表明点云分布更加合理,语义信息表达更加

完整且具有更高的几何描述能力。上采样操作在优化点云分布方面展现了显著效果,即使在点数稀少的困难模式下,均匀性损失也明显降低,相较于真实值更清晰地勾勒出目标轮廓,从而有效恢复了物体的几何结构并保持其精确形态。为直观对比,图4的第1行同时展示了与目标对应的图像区域,为点云的几何优化效果提供了进一步的视觉支撑。此外,在图4的第3、第5和第6行分别可视化了基线方法 CT3D 以及 PUDet 在训练15个和80个 epoch 后生成的注意力权重。为提供更直观的演示,展示中采用候选区域而非柱状区域。不同点云的标注样式(参见右侧图例条)对应自注意力层生成的不同等级注意力权重,直观呈现了模型对目标关键区域关注程度的差异。

值得注意的是,即使整体点数相近,PUDet 在简单模式下的前景目标均匀性损失仍小于基线方法;在困难

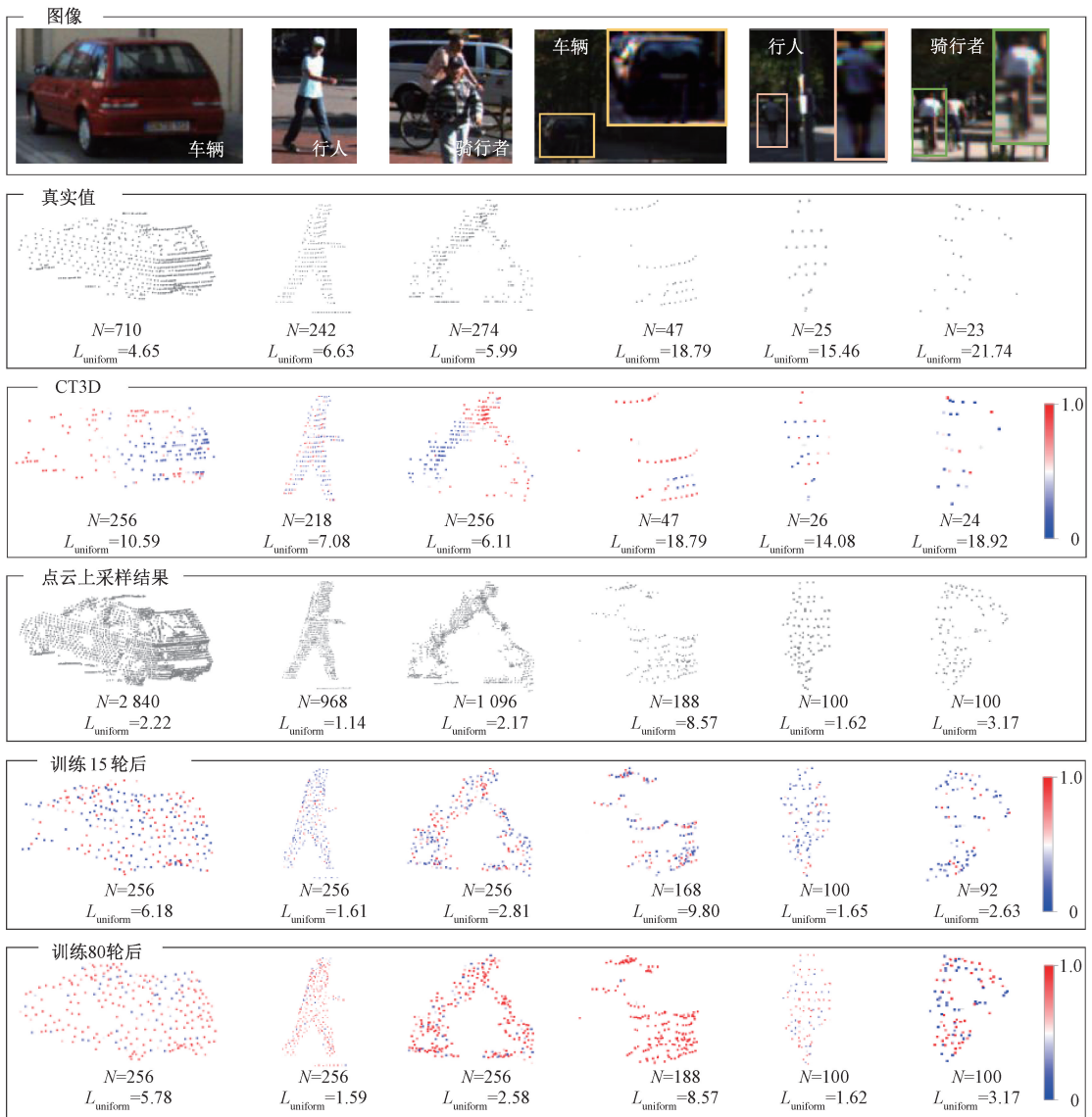


图 4 几何轮廓优化的可视化示例

Fig. 4 Visualization example of geometric profile optimization

模式下,填补了稀疏区域并均衡点间密度,均匀性损失显著降低,展现了对几乎无法辨识形状的前景目标进行高效几何恢复的能力。此外,相比基线方法,PUDet 分配高注意力权重的点数量明显更多,随着训练的逐步进行,这些高权重点逐渐覆盖整个物体。这表明网络能够有效感知并突出前景目标的整体形状,揭示了生成式上采样网络集成对性能提升的关键贡献。

3.4 消融实验

1) 柱状区域选择

为进一步验证本文算法中各部分模块的有效性,进行了一系列消融实验,并使用 KITTI 验证集基于 40 个召回位置计算的 3D 检测平均精度 (AP) 作为评估指标。

为了全面评估柱状区域选择模块的有效性,针对其不同配置进行了实验分析,并量化其对推理速度的影响,结果如表 6 所示。与此同时,实验还统计了帧率 (frame per second, FPS),用于衡量该模块在优化点云选择的同时对整体计算效率的影响,以确保在提升检测精度的同时保持高效的推理性能。实验表明,随着 N_{sample} 的增加,性能呈现出明显的提升。然而,当 N_{sample} 过大时,性能增益趋于递减,同时计算成本显著增加。例如,当性能仅提升 0.33% 时,推理时间却从 0.29 s 增加至 0.41 s。值得注意的是,过小的 N_{sample} 值也会对性能产生负面影响,例如,当 $N_{\text{sample}} = 128$ 时,性能反而低于未使用柱状区域选择模块的配置。因此,选择 $N_{\text{sample}} = 256$ 能够在性能与计算效率之间实现良好的平衡。

表6 柱状区域选择模块的消融实验结果

Table 6 Ablation results of columnar region selection module

柱状区域 提取模块	N_{sample}	3D 目标检测/%			FPS
		简单	中等	困难	
×		91.23	83.34	82.65	0.21
✓	128	89.21	81.69	81.02	0.23
✓	256	93.51	85.33	84.02	0.29
✓	512	93.97	86.48	84.08	0.41

2) 点云增强模块

为了验证点云增强各组件的有效性,设计了相关实验,结果如表7所示。第1~4行的结果显示,LDEM和DDAM模块均对PUDet的性能提升起到了关键作用。具体而言,移除LDEM会对简单级别的检测精度产生较大影响,而移除DDAM则显著削弱了困难级别的表现。这一现象充分证明了2个模块在优化近距离和远距离目标检测方面的重要性与互补性。

表7 点云增强的消融实验结果

Table 7 Ablation experiment results of point cloud enhancement

序号	LDEM	DDAM	N_{iter}	r	3D 目标检测/%		
					简单	中等	困难
1	×	×			92.34	84.97	82.91
2	✓	×	2	4	92.76	85.12	83.04
3	×	✓	2	4	92.56	85.41	83.86
4	✓	✓	2	4	93.51	85.33	84.02
5	✓	✓	5	4	93.82	85.89	84.67
6	✓	✓	10	4	94.09	86.13	84.98
7	✓	✓	2	6	93.50	85.47	84.76
8	✓	✓	2	8	93.67	85.51	85.32

上采样网络中的迭代次数 N_{iter} 对点云增强能力具有重要影响,如表7的第5~6行所示。当 $N_{iter} = 10$ 时,模型性能达到最优,这表明通过增加迭代次数以生成更精确的点云几何结构,有助于检测器更准确地捕获目标细节。然而,考虑到较高的计算成本,选取 $N_{iter} = 2$ 作为算法的参数设置,从而在性能提升与计算效率之间实现更合理的平衡;提高上采样率 r 也有助于增强困难级别目标的检测性能,但对其他级别的性能提升相对有限,同时会对实时性造成一定程度的影响。

3) ROI 区域内点云数量

在表8中,本文详细分析了上采样后点数对检测性能的影响。将 $N_{afterpu}$ 减少至128时,性能明显下降,主要由于点云数量不足以完整表达目标的关键几何特征,导致前景特征提取能力受限;而将 $N_{afterpu}$ 增加至512,则显

著提升了简单和中等级别的检测性能,这表明更高的点云密度能够增强目标几何的细节表达。然而,在困难级别下,由于点云本身的稀疏性和遮挡问题,额外的点云密度并未带来显著收益。综合考虑计算效率与性能之间的权衡,最终选择将 $N_{afterpu}$ 设置为256,以在特征表达能力与实时性之间实现更优的平衡。

表8 超参数 $N_{afterpu}$ 消融实验结果Table 8 Hyperparameter $N_{afterpu}$ ablation experiment results

序号	$N_{afterpu}$	3D 目标检测/%				FPS
		简单	中等	困难	3D mAP	
1	128	90.77	83.21	82.07	85.35	0.22
2	256	93.51	85.33	84.02	87.62	0.29
3	512	94.39	86.81	84.07	88.42	0.37

4) 对不同检测器的影响

本文首先在SECOND和PointPillar RPN中对PUDet的性能进行了验证,结果如表9所示。与基线RPN方法相比,PUDet在中等难度上性能分别提升了3.57%和5.38%,表明其在不同网络架构中的适用性和有效性。为进一步验证PUDet的通用性,将其集成到多种基于LiDAR的两阶段检测器中进行了验证。在实验中,采用一致的特征提取网络,从3D RoI内的原始点云中提取点特征,并将其与RoI网格池化特征拼接,以实现检测器的进一步优化。作为基础检测器,选择了PV-RCNN^[37]和CenterPoint-RCNN^[7]进行实验评估。实验结果如表10所示,PUDet在多种检测框架中均显著提升了检测性能,进一步验证了其在复杂场景下的鲁棒性和普适性。

表9 不同RPN骨干网络消融实验结果

Table 9 Ablation experimental results of different RPN backbone networks

RPN	PUDet	3D 目标检测/%			
		简单	中等	困难	3D mAP
SECOND	×	90.31	81.76	78.88	83.65
	✓	93.51	85.33	84.02	87.62
PointPillar	×	87.75	78.39	75.18	80.44
	✓	91.61	83.77	81.59	85.66

5) 模型参数及推理速度

为了更全面地分析生成式点云上采样网络对不同两阶段基线的影响,在表11中报告了将PUDet集成到CenterPoint-RCNN、PV-RCNN和CT3D后的模型参数与推理速度变化情况。结果基于KITTI验证集,在IoU阈值为0.5的条件下报告了模型参数和推理速度等关键性能指标。

表 10 4 种不同的基线方法与 PUDet 在 KITTI 验证集的性能比较

Table 10 Performance comparison of four different baseline and PUDet on the KITTI validation set

基线方法	PUDet	召回位置	车辆/%	行人/%	骑行者/%
PV-RCNN	×	40	86.37	59.43	76.71
	✓	40	87.70	59.76	77.30
CenterPoint-RCNN	×	40	85.32	57.90	78.33
	✓	40	86.82	59.07	78.51
PVRCNN	×	11	83.79	60.12	74.63
	✓	11	85.00	60.81	74.79
CenterPoint-RCNN	×	11	82.13	58.19	76.10
	✓	11	84.45	59.02	76.14

表 11 不同基线方法及其结合 PUDet 增强后的性能比较
Table 11 Performance comparison of different baseline and their enhancements with PUDet

基线方法	PUDet	模型参数/MB	推理速度/Hz
CT3D	×	7.85	8.33
	✓	7.92	3.45 ↓ 4.88
PV-RCNN	×	13.12	9.09
	✓	21.60	2.78 ↓ 6.31
CenterPoint-RCNN	×	9.67	11.12
	✓	18.16	4.54 ↓ 6.58

3.5 PUDet 性能分析

为探究 PUDet 在不同场景中的性能表现, 对其在不同距离和目标密度下进行了详细评估。评估指标为基于 40 个召回位置计算的 3D 平均精度 (AP), 涵盖车辆、行人和骑行者 3 类目标。正如图 5 和 6 所示, PUDet 在远距离和稀疏目标上的提升效果最为显著, 充分验证了其在处理稀疏点云方面的优势。然而, 对于极远距离的目

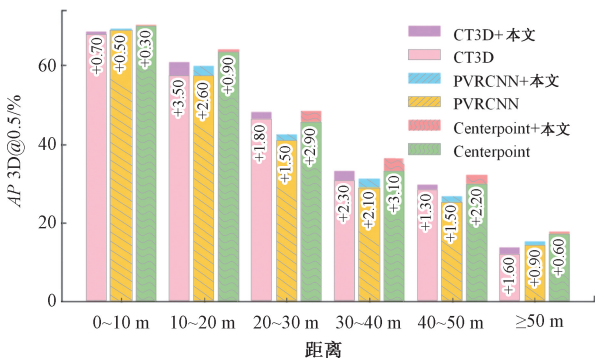


图 5 在 KITTI 验证集上针对不同距离范围的实验

Fig. 5 Experiments on the KITTI validation set for different distance ranges

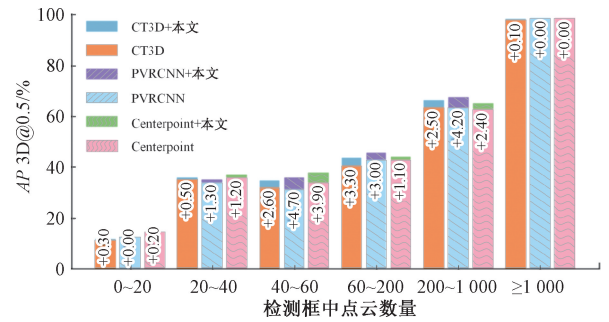


图 6 在 KITTI 验证集上针对不同目标密度的实验

Fig. 6 Experiments on the KITTI validation set for different object densities

标或仅包含极少点数的目标, 由于可恢复的细节有限, 上采样网络的增益效果相对受限。

4 结 论

本研究提出了一种创新性框架 PUDet, 将生成式点云上采样网络融入判别式 3D 目标检测任务, 充分利用生成模型的先验知识, 有效提升了点云的几何结构表达能力与检测精度。针对近距离目标, PUDet 优化了点云分布, 增强了几何细节的表达; 对于远距离目标, 通过对稀疏点云的密度增强, 有效提升了检测性能, 尤其在稀疏场景与遮挡情况下表现尤为优异。实验结果表明, PUDet 在检测精度和鲁棒性上取得了显著提升, 同时展现了生成式点云上采样网络在复杂 3D 场景中目标识别与定位任务中的巨大潜力。通过优化点云几何结构表达与分布均衡, PUDet 为构建高精度、低成本的自动驾驶感知系统奠定了坚实基础, 并为智能驾驶技术的持续发展提供了重要支持; 还为更复杂、更动态的 3D 场景下的目标检测与定位问题提供了精细化和高效的解决方案, 为下一代智能交通系统与无人驾驶技术的发展注入了新的可能性。

参考文献

[1] FAN L, WANG F, WANG N Y, et al. Fully sparse 3D object detection [J]. Advances in Neural Information Processing Systems, 2022, 35: 351-363.

[2] LIANG T T, XIE H W, YU K CH, et al. Bevfusion: A simple and robust LiDAR-camera fusion framework [J]. Advances in Neural Information Processing Systems, 2022, 35: 10421-10434.

[3] WU H, WEN CH L, LI W, et al. Transformation-equivariant 3D object detection for autonomous driving [C]. 2023 AAAI Conference on Artificial Intelligence, 2023, 37(3): 2795-2802.

- [4] WU H, WEN CH L, SHI SH SH, et al. Virtual sparse convolution for multimodal 3D object detection[C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023; 21653-21662.
- [5] SCHINAGL D, KRISPEL G, POSSEGGGER H, et al. OccAM's laser: Occlusion-based attribution maps for 3D object detectors on LiDAR data[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022; 1141-1150.
- [6] XU Q G, ZHOU Y, WANG W Y, et al. Spg: Unsupervised domain adaptation for 3D object detection via semantic point generation[C]. 2021 IEEE/CVF International Conference on Computer Vision, 2021; 15446-15456.
- [7] YIN T W, ZHOU X Y, KRAHENBUHL P. Center-based 3D object detection and tracking[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021; 11784-11793.
- [8] YIN T W, ZHOU X Y, KRÄHENBUHL P. Multimodal virtual point 3D detection[J]. Advances in Neural Information Processing Systems, 2021, 34: 16494-16507.
- [9] ZHANG H L, YANG D F, YURTSEVER E, et al. Faraway-frustum: Dealing with LiDAR sparsity for 3D object detection using fusion[C]. 2021 IEEE International Intelligent Transportation Systems Conference, 2021; 2646-2652.
- [10] LI Y W, FAN Y CH, XIANG X Y, et al. Efficient and explicit modelling of image hierarchies for image restoration[C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023; 18278-18289.
- [11] LUO ZH X, HUANG Y, LI SH, et al. Learning the degradation distribution for blind image super-resolution[J]. ArXiv preprint arXiv:2203.04962, 2022.
- [12] WANG X T, XIE L B, DONG CH, et al. Real-esrgan: Training real-world blind super-resolution with pure synthetic data[C]. 2021 IEEE/CVF International Conference on Computer Vision, 2021; 1905-1914.
- [13] ZHANG K, LIANG J Y, VAN GOOL L, et al. Designing a practical degradation model for deep blind image super-resolution[C]. 2021 IEEE/CVF International Conference on Computer Vision, 2021; 4791-4800.
- [14] LI R H, LI X ZH, HENG P A, et al. Point cloud upsampling via disentangled refinement[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021; 344-353.
- [15] LIU H, YUAN H, HOU J H, et al. PUFA-GAN: A frequency-aware generative adversarial network for 3D point cloud upsampling[J]. IEEE Transactions on Image Processing, 2022, 31: 7389-7402.
- [16] YU L Q, LI X ZH, FU C W, et al. PU-NET: Point cloud upsampling network[C]. 2018 IEEE Conference on Computer Vision and Pattern Recognition, 2018; 2790-2799.
- [17] WU X P, PENG L, YANG H H, et al. Sparse fuse dense: Towards high quality 3D detection with depth completion[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022; 5418-5427.
- [18] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]. 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012; 3354-3361.
- [19] SHENG H L, CAI S J, LIU Y, et al. Improving 3D object detection with channel-wise transformer[C]. 2021 IEEE/CVF International Conference on Computer Vision, 2021; 2743-2752.
- [20] CAESAR H, BANKITI V, LANG A H, et al. nuScenes: A multimodal dataset for autonomous driving[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020; 11621-11631.
- [21] CHEN X ZH, MA H M, WAN J, et al. Multi-view 3D object detection network for autonomous driving[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017; 1907-1915.
- [22] HUANG K C, LYU W J, YANG M H, et al. PTT: Point-trajectory Transformer for efficient temporal 3D object detection[C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024; 14938-14947.
- [23] QI C R, LITANY O, HE K M, et al. Deep hough voting for 3D object detection in point clouds[C]. 2019 IEEE/CVF International Conference on Computer Vision, 2019; 9277-9286.

- [24] QI C R, LIU W, WU CH X, et al. Frustum PointNets for 3D object detection from RGB-D data [C]. 2018 IEEE Conference on Computer Vision and Pattern Recognition, 2018; 918-927.
- [25] YANG Z T, SUN Y N, LIU SH, et al. 3DSSD: Point-based 3D single stage object detector [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020; 11040-11048.
- [26] 涂新奎, 郑少武, 于善虎, 等. 基于对称形状生成的三维目标检测网络 [J]. 仪器仪表学报, 2023, 44(6): 252-263.
- TU X K, ZHENG SH W, YU SH H, et al. 3D object detection network based on symmetric shape generation [J]. Chinese Journal of Scientific Instrument, 2023, 44(6): 252-263.
- [27] 金宇锋, 陶重犇. 基于 Transformer 的融合信息增强 3D 目标检测算法 [J]. 仪器仪表学报, 2024, 44(12): 297-306.
- JING Y F, TAO CH B. Fusion information enhanced method based on Transformer for 3D object detection [J]. Chinese Journal of Scientific Instrument, 2024, 44(12): 297-306.
- [28] 郑少武, 李巍华, 胡坚耀. 基于激光点云与图像信息融合的交通环境车辆检测 [J]. 仪器仪表学报, 2019, 40(12): 143-151.
- ZHENG SH W, LI W H, HU J Y. Vehicle detection in the traffic environment based on the fusion of laser point cloud and image information [J]. Chinese Journal of Scientific Instrument, 2019, 40(12): 143-151.
- [29] LANG A H, VORA S, CAESAR H, et al. Pointpillars: Fast encoders for object detection from point clouds [C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019; 12697-12705.
- [30] YAN Y, MAO Y X, LI B. Second: Sparsely embedded convolutional detection [J]. Sensors, 2018, 18(10): 3337.
- [31] YANG B, LUO W J, URTASUN R. Pixor: Real-time 3D object detection from point clouds [C]. 2018 IEEE Conference on Computer Vision and Pattern Recognition, 2018; 7652-7660.
- [32] ZHOU Y, TUZEL O. Voxelnet: End-to-end learning for point cloud based 3D object detection [C]. 2018 IEEE Conference on Computer Vision and Pattern Recognition, 2018; 4490-4499.
- [33] LI J Y, LUO CH X, YANG X D. PillarNeXt: Rethinking network designs for 3D object detection in LiDAR point clouds [C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023; 17567-17576.
- [34] CHEN Y K, LIU J H, ZHANG X Y, et al. Voxelnext: Fully sparse voxelnet for 3D object detection and tracking [C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023; 21674-21683.
- [35] ERABATI G K, ARAUJO H. Li3DeTr: A LiDAR based 3D detection transformer [C]. 2023 IEEE/CVF Winter Conference on Applications of Computer Vision, 2023; 4250-4259.
- [36] 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; 3164-3173.
- [37] 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; 10529-10538.
- [38] 周治国, 马文浩. 一种多层多模态融合 3D 目标检测方法 [J]. 电子学报, 2024, 52(3): 696-708.
- ZHOU ZH G, MA W H. 3D object detection based on multilayer multimodal fusion [J]. Acta Electronica Sinica, 2024, 52(3): 696-708.
- [39] VORA S, LANG A H, HELOU B, et al. PointPainting: Sequential fusion for 3D object detection [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020; 4604-4612.
- [40] WANG CH W, MA CH, ZHU M, et al. Pointaugmenting: Cross-modal augmentation for 3D object detection [C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021; 11794-11803.
- [41] XU SH Q, ZHOU D F, FANG J, et al. FusionPainting: Multimodal fusion with adaptive attention for 3D object detection [C]. 2021 IEEE International Intelligent Transportation Systems Conference, 2021; 3047-3054.
- [42] JIAO Y, JIE Z Q, CHEN SH X, et al. Msmdfusion: Fusing LiDAR and camera at multiple scales with multi-depth seeds for 3D object detection [C]. 2023 IEEE/

- CVF Conference on Computer Vision and Pattern Recognition, 2023; 21643-21652.
- [43] SHROUT O, BEN-SHABAT Y, TAL A. Gravos: Voxel selection for 3D point-cloud detection[C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023; 21684-21693.
- [44] XU Q G, ZHONG Y Q, NEUMANN U. Behind the curtain: Learning occluded shapes for 3D object detection[C]. 2022 AAAI Conference on Artificial Intelligence, 2022, 36(3): 2893-2901.
- [45] ZHENG W, TANG W L, JIANG L, et al. SE-SSD: Self-ensembling single-stage object detector from point cloud[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021; 14494-14503.
- [46] 赵树恩, 袁亮, 赵东宇. 基于要素信息补全的自动驾驶复杂场景语义理解[J]. 仪器仪表学报, 2025, 46(4): 295-305.
- ZHAO SH EN, YUAN L, ZHAO D Y. Semantic understanding of complex scenarios in autonomouddriving based on element information completion[J]. Chinese Journal of Scientific Instrument, 2025, 46(4): 295-305.
- [47] LI R H, LI X ZH, FU C W, et al. PU-GAN: A point cloud upsampling adversarial network[C]. 2019 IEEE/CVF International Conference on Computer Vision, 2019; 7203-7212.
- [48] WANG Y F, WU SH H, HUANG H, et al. Patch-based progressive 3D point set upsampling[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019; 5958-5967.
- [49] HE Y, TANG D H, ZHANG Y D, et al. Grad-PU: Arbitrary-scale point cloud upsampling via gradient descent with learned distance functions[C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023; 5354-5363.
- [50] QIAN G CH, ABUALSHOUR A, LI G H, et al. PU-GCN: Point cloud upsampling using graph convolutional networks[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021; 11683-11692.
- [51] ENGELS G, ARANJUELO N, ARGANDA-CARRERAS I, et al. 3D object detection from LiDAR data using distance dependent feature extraction[J]. ArXiv preprint arXiv:2003.00888, 2020.
- [52] SHANNON C E. Communication in the presence of noise[J]. Proceedings of the IRE, 1949, 37(1): 10-21.
- [53] OpenPCDet Development Team. Openpcdet: An opensource toolbox for 3D object detection from point clouds[EB/OL]. 2020.
- [54] DENG J J, SHI SH SH, LI P W, et al. Voxel R-CNN: Towards high performance voxel-based 3D object detection [C]. 2021 AAAI Conference on Artificial Intelligence, 2021, 35(2): 1201-1209.
- [55] SHI SH SH, WANG X G, LI H SH. PointRCNN: 3D object proposal generation and detection from point cloud[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019; 770-779.
- [56] KINGMA D, BA J. Adam: A method for stochastic optimization[J]. Computer Science, 2014; 6980.
- [57] GUGGER S. The 1cycle policy[EB/OL]. 2018.
- [58] PANG S, MORRIS D, RADHA H. CLOCs: Camera-LiDAR object candidates fusion for 3D object detection[C]. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2020; 10386-10393.
- [59] LIANG M, YANG B, CHEN Y, et al. Multi-task multi-sensor fusion for 3D object detection[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019; 7345-7353.
- [60] XIE L, XIANG CH, YU ZH X, et al. PI-RCNN: An efficient multi-sensor 3D object detector with point-based attentive cont-conv fusion module [C]. 2020 AAAI Conference on Artificial Intelligence, 2020, 34(7): 12460-12467.
- [61] HUANG T T, LIU ZH, CHEN X W, et al. EPNet: Enhancing point features with image semantics for 3D object detection [C]. Computer Vision-ECCV 2020, 2020; 35-52.
- [62] YANG Z T, JIANG L, SUN Y N, et al. A unified query-based paradigm for point cloud understanding[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022; 8541-8551.
- [63] LIU ZH, ZHAO X, HUANG T T, et al. TANet: Robust 3D object detection from point clouds with triple attention[C]. 2020 AAAI Conference on Artificial Intelligence, 2020, 34(7): 11677-11684.
- [64] SHI SH SH, WANG ZH, WANG X G, et al. Part-A²

net: 3D part-aware and aggregation neural network for object detection from point cloud [J]. ArXiv preprint arXiv:1907.03670, 2019.

- [65] ZHENG W, TANG W L, CHEN S J, et al. CIA-SSD: Confident IoU-aware single-stage object detector from point cloud [C]. 2021 AAAI Conference on Artificial Intelligence, 2021, 35(4): 3555-3562.
- [66] CHEN CH, CHEN ZH, ZHANG J, et al. SASA: Semantics-augmented set abstraction for point-based 3D object detection [C]. 2022 AAAI Conference on Artificial Intelligence, 2022, 36(1): 221-229.
- [67] YANG H H, HE T, LIU J H, et al. GD-MAE: Generative decoder for MAE pre-training on LiDAR point clouds [C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 9403-9414.
- [68] KOO I, LEE I, KIM S H, et al. PG-RCNN: Semantic surface point generation for 3D object detection [C]. 2023 IEEE/CVF International Conference on Computer Vision, 2023: 18142-18151.

作者简介



许丽梅, 2022 年于哈尔滨工程大学获得学士学位, 现为北京理工大学研究生, 主要研究方向为深度学习、3D 目标检测和自动驾驶。

E-mail: 3220221541@bit.edu.cn

Xu Limei received her B. Sc. degree from Harbin Engineering

University in 2022. She is currently a M. Sc. candidate at Beijing Institute of Technology. Her main research areas are deep learning, 3D object detection and autonomous driving.



周治国 (通信作者), 1998 年于华中科技大学获得学士学位, 2001 年于北京理工大学获得硕士学位, 2009 年于北京理工大学获得博士学位, 现为北京理工大学副教授, 主要研究方向为具身智能无人系统。

E-mail: zhiguo Zhou@bit.edu.cn

Zhou Zhiguo (Corresponding author) received his B. Sc. degree from Huazhong University of Science and Technology in 1998, his M. Sc. degree from Beijing Institute of Technology in 2001, and his Ph. D. degree from Beijing Institute of Technology in 2009. He is currently an associate professor at Beijing Institute of Technology. His main research direction is embodied intelligent unmanned systems.



周学华, 2010 年于北京理工大学获得学士学位, 2013 年于北京理工大学获得硕士学位, 现为北京理工大学工程师, 主要研究方向为人工智能、无人平台系统、实时仿真及图像处理。

E-mail: xuehuazhou@bit.edu.cn

Zhou Xuehua received his B. Sc. degree from Beijing Institute of Technology in 2010, and his M. Sc. degree from Beijing Institute of Technology in 2013. He is currently an engineer at Beijing Institute of Technology. His main research areas are artificial intelligence, unmanned platform systems, real-time simulation and image processing.