激光与光电子学进展 先进成像

融合语义和激光点云空间可见性特征的3D行人检测

熊璐，邓振文，田炜*，王之昂
同济大学汽车学院，上海 201804

摘要 车载激光雷达为智能驾驶汽车提供精确的周围空间几何信息而成为车载主流传感器。为克服单传感器对目标检测的局限性，对激光点云的几何特征、空间可见性特征和图像语义信息在合理设计的网络框架中进行融合，进一步提升3D行人检测精度。首先采用高效的三维空间光线投射算法形成空间可见性特征编码；其次融合图像语义类别信息，增强点云特征；最后定量和定性分析各附加信息和相关超参数对检测结果的影响。实验结果表明：相比单帧点云，结合历史前10帧点云后3D行人检测精度提升32.63个百分点；进一步融合图像语义和点云空间可见性信息，相比基准方法，所提方法的检测精度提升2.42个百分点，且超过部分主流方法，更加适用于交通场景的3D行人检测。

关键词 目标检测；图像与点云融合；点云空间可见性；智能驾驶环境感知
中图分类号 TP391.4 **文献标志码** A **DOI**：10.3788/LOP220712

Three-Dimensional Pedestrian Detection by Fusing Image Semantics and Point Cloud Spatial Visibility Features

Xiong Lu, Deng Zhenwen, Tian Wei*, Wang Zhiang
School of Automotive Studies, Tongji University, Shanghai 201804, *China*

Abstract Vehicular light detection and ranging (LiDAR) has become a standard sensor in automotive by offering accurate geometric information of the surrounding region for intelligent driving vehicles. In order to overcome the limited performance of a single sensor for object detection, the geometric and spatial visibility features of LiDAR point clouds are fused with image semantic information in a network framework to achieve accurate three dimensional (3D) pedestrian detection. First, an effective 3D ray-casting algorithm is introduced to produce spatial visibility feature encodings. Second, the image semantic information is incorporated to improve point cloud features. Finally, the impact of added information and related hyperparameters on detection findings are quantitatively and qualitatively examined. Experimental findings demonstrate that compared with the single frame point cloud, the 3D pedestrian detection accuracy is enhanced by 32.63 percentage points after aggregating the last 10 frames of the point cloud in history. By further fusing image semantics and point cloud spatial visibility information, the proposed method's detection accuracy is enhanced by 2.42 percentage points compared with the benchmark approach, and exceeds some standard approaches. Our enhanced approach is more suitable for 3D pedestrian detection in a traffic environment.

Key words object detection; image and point cloud fusion; point cloud spatial visibility; intelligent driving and environmental perception

1 引 言

车载激光雷达利用激光束在物体表面的反射探测远距离障碍物的单点精确位置，再通过多激光束旋转或光学相控阵技术生成周围环境的稀疏激光点云。作为有源传感器，车载激光雷达在曝光或低光照的环境下能够实现更可靠的检测而成为自动驾驶研究的主流传感器。但与相机相比，激光雷达容易受雨雾和灰尘影响，且制造成本随激光束增加呈指数增长，导致目前部署的自动驾驶汽车通常配备不超过64线的激光雷达。激光点云的稀疏性给目标检测识别任务，尤其是被遮挡或体积较小的行人检测[1-2]带来诸多挑战。因此，研究多传感器融合的三维行人检测对有效提升感知系统的准确性和稳定性具有重要意义。

收稿日期：2022-02-14；**修回日期**：2022-02-24；**录用日期**：2022-03-14；**网络首发日期**：2022-03-26
基金项目：国家自然科学基金青年项目(52002285)、上海市浦江人才计划(2020PJD075)、上海市科技计划(21ZR1467400)
通信作者：*tian_wei@tongji.edu.cn

图像包含丰富的语义信息，常与激光点云组成3D目标检测的数据融合方案。AVOD[3]提取图像和投影点云的感兴趣区域，然后将区域转换至鸟瞰图或前视图进行特征融合，但点云在投影过程中将丢失部分空间信息。F-PointNet[4]先完成图像2D检测，再将框中的3D点云输入至PointNet[5]完成3D位置和形状回归，但图像2D检测失效将直接导致3D检测无法实现。PointPainting[6]利用图像语义分割类别信息作为激光点云的额外特征，再通过点云处理框架VoxelNet[7]或PointPillars[8]确定3D框的类别、位置和尺寸。虽近期提出较多先进框架（如VPFNet[9]和SPG[10]）进一步提升目标检测精度，但所需运行条件难以满足车载计算平台实时性的要求。

考虑车载计算平台条件的限制，本文选用具有高效点云特征编码的PointPillars[8]作为基础网络架构。采用高效的三维空间光线投射算法完成点云的空间可见性的特征编码；同时融合激光点云空间可见性特征和图像语义特征，提升3D行人检测性能并量化各因素影响程度；分析在数据融合过程中相关超参数选择对检测性能的影响。实验结果表明，图像语义和点云空间可见性信息融合在合适的超参数条件下能够提升3D行人检测精度，并超过部分主流方法，为其他网络架构的设计和优化提供重要参考。

2　点云空间可见性特征编码

基于单模态激光雷达的目标检测算法通常利用深度学习方法提取点云的几何位置特征来进行目标检测和识别，而忽略点云空间“可见性”信息，即激光束发射原点与被测物体之间为无遮挡的空闲空间，被测物体后方仍可能存在其他障碍物。这种空间可见性提供三维空间的占据状态和空闲空间分布等额外信息，可用于进一步提高对障碍物的检测精度。

本文采用一种基于Raycasting[11]的高效三维空间光线投射算法，该算法可快速计算每条激光射线在空间中经过的体素，适用于实时性要求较高的车载计算平台。首先介绍高效三维空间光线投射算法在二维平面的应用。

给定平面网格，对于任意激光射线，需列出所有遍历的体素，如图1所示，即a，b，c，d，e，f。设x'和y'为激光束沿X和Y方向的速度分量，t为传播时间，则射线传播方程为

$$\left[x,y\right]^{\mathrm{T}}=\left[x',y'\right]^{\mathrm{T}}\cdot t+\left[x_0,y_0\right]^{\mathrm{T}},\tag{1}$$

式中：$t\geqslant 0$；$(x'/y')=(x_{\mathrm{p}}/y_{\mathrm{p}})$，$(x_{\mathrm{p}},y_{\mathrm{p}})$为激光点的位置坐标；$(x_0,y_0)$为激光束起点位置坐标。计算激光束起点位置至起始网格X、Y边界的时间t_X和t_Y，以及激光束掠过单个网格水平和垂直方向距离所需时间$t_{\Delta X}$和$t_{\Delta Y}$，如图1所示。确定初始网格坐标(s_X,s_Y)且用1或−1赋值遍历步长$s_{\Delta X}$和$s_{\Delta Y}$。在迭代求遍历体素坐标的过程中，如图2所示，当$t_X<t_Y$，射线在水平方向上优先到达下一个体素，即当前网格和水平邻近网格均为遍历体素，反之，竖直方向为遍历体素。随后更新参数，进行下一次迭代，直至到达终点。

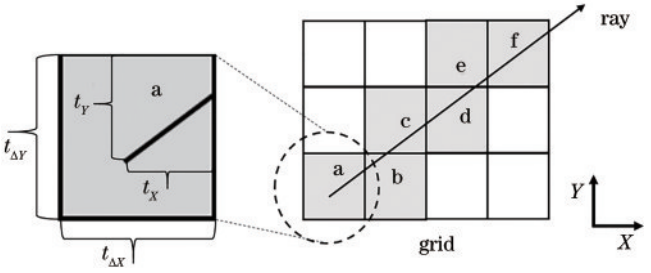

图1　射线遍历网格的示意图

Fig. 1　Schematic of ray traversing grid

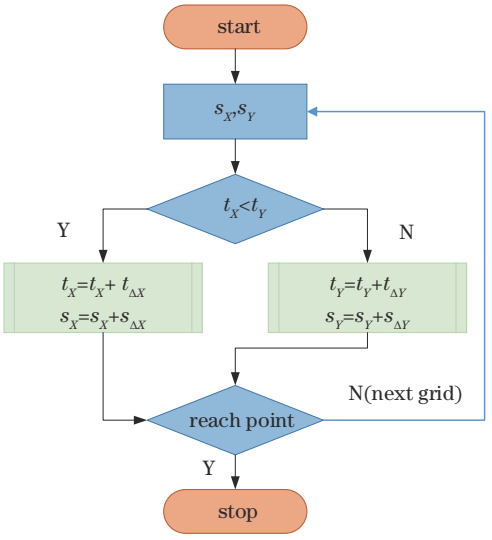

图2　二维Raycasting算法的逻辑图

Fig. 2　Logic diagram of 2D Raycasting algorithm

扩展至三维空间，在Z维度添加相应变量并在迭代过程中对比t_X、t_Y、t_Z的大小，类似找出整个三维空间中激光束所遍历的体素。算法在初始化阶段需要约33次浮点运算，每次迭代仅需2次浮点比较、1次浮点加法、2次整数比较和1次整数加法。单帧点云的空间可见性特征编码时间约为0.017 s，因而该编码方式适用于车载平台。

3　融合网络架构

3.1　网络总体架构

激光雷达点云为自车提供周围环境物体准确的几何位置信息，而相机可提供区分物体类别相关的语义信息。本文提出基于LiDAR和相机的数据融合方法，利用多传感器的信息互补提升3D目标检测精度。融合方法的总体架构如图3所示。首先，通过语义分割网络获取物体类别信息，并对激光点云进行特征增强；其次，使用光线投射算法重建感知空间可见性状态；然后，提取增强点云的高维特征，并对其与空间可见性特

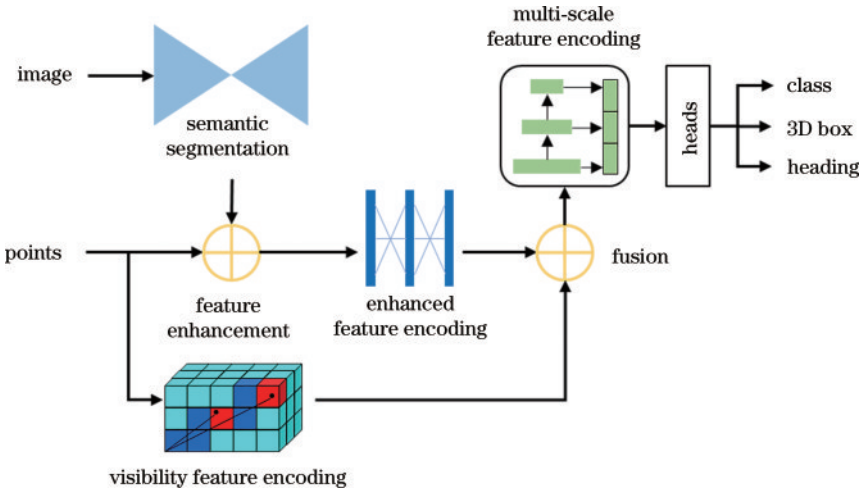

图 3　融合网络的总体架构

Fig. 3　Overall framework of fusion network

征进行融合；最后，将融合特征输入至三维目标检测框架，完成分类和回归。

3.2　语义分割及点云增强

语义分割网络从输入图像获取各像素点的背景、车辆、非机动车、行人或其他细分类别结果。根据 Siam 等[12]分析和总结的多种语义分割网络，综合考虑车载平台计算性能，本文使用 DeeplabV3+[13](ReNet18)作为语义分割网络，其单帧图像分割时间为 0.0121 s。

通过传感器外参和相机内参，将各激光点与图像像素形成数据关联，如图 4 所示，并将图像语义分割的类别结果 c_{id} 作为点云的增强特征，则点云特征由 (x, y, z, r) 增强为 (x, y, z, r, c_{id})，其中 x、y、z、r 为激光点云原始特征，分别表示点云三维空间坐标 x、y、z、反射率。

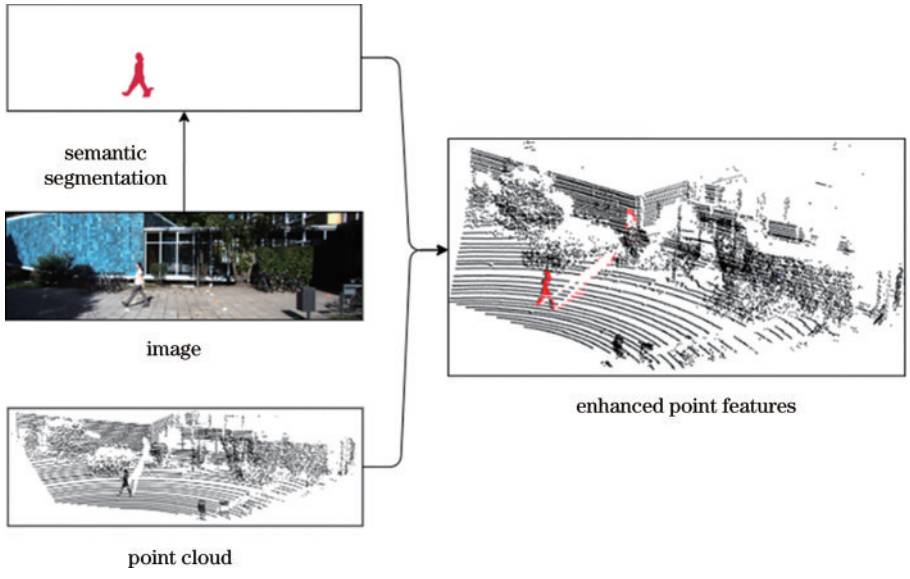

图 4　语义分割和点云特征增强

Fig. 4　Semantic segmentation and point feature enhancement

3.3　特征编码

借鉴 PointPillars[8]高效特征编码方式，特征增强后的点云回散至水平地面体柱。由多层感知机(MLP)提取各体柱内点云的高维特征，并沿数量维度进行最大池化操作，此时每个体柱表示为固定长度为 D 的特征向量，从而转换为伪图像的混合特征图，如图 5 所示。

根据激光束的传播可以估计空间体素的可见性状态，进而可以得到空间可见性特征。首先对感知空间进行体素化，且体素的长 W、宽 H 与体柱相同，高度方向体素数量为 K。每个体素包含三种状态：未知(Unknown)、空闲(Free)和占据(Occupied)。体素初始化为未知状态，通过各激光束在体素中的扫掠更新对应体素的状态。当激光束经过体素时，该体素状态设为空闲，当激光点落在体素内，该体素状态设为占据，如图 6 所示，点云空间可见性特征进一步转换为伪图像形式的可见性特征图。

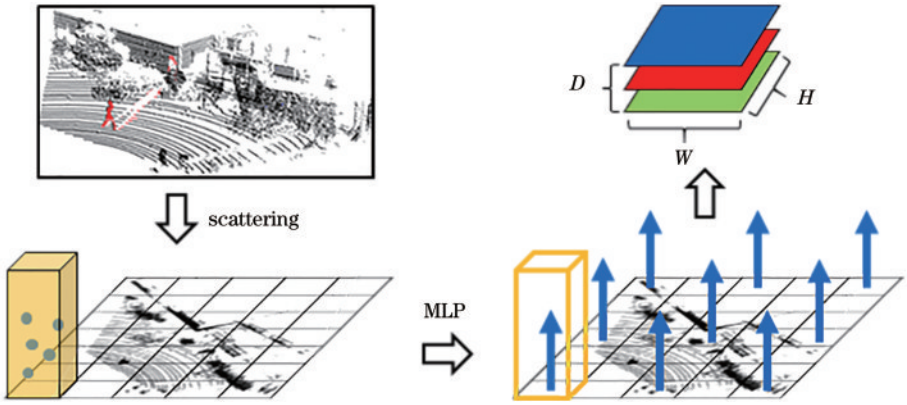

图 5　几何特征和语义特征编码

Fig. 5　Geometric feature and semantic feature encoding

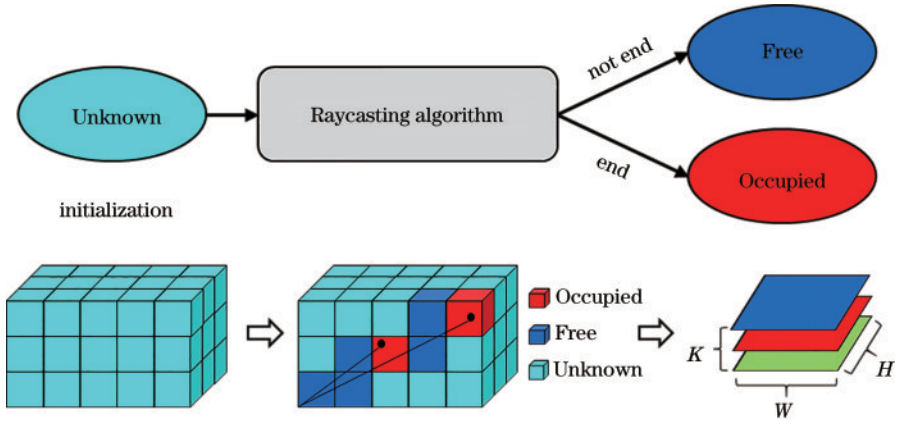

图 6　空间可见性特征编码

Fig. 6　Spatial visibility feature encoding

3.4　特征融合及检测头

在特征编码阶段，将点云特征和可见性特征转换为伪图像的特征图，从而可利用图像卷积方式完成目标检测任务。所获取的混合特征图和可见性特征图先沿通道维度堆叠，生成融合特征；再通过骨干网络从融合特征提取多尺度特征，并由反卷积模块调整为同一尺寸输出；输出特征最后通过不同卷积模块，完成 3D 目标检测的分类和回归。整个过程如图 7 所示。

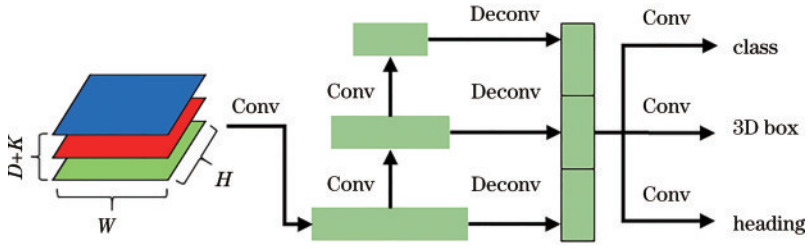

图 7　特征融合与检测头

Fig. 7　Feature fusion and detection heads

骨干网络主要包含两个子模块：top-down 特征提取模块，主要由卷积层（Conv）、批量归一化层（BN）、非线性层（ReLU）组成，通过设置步长捕获不同尺度的特征信息；second 特征处理模块，由反卷积层（Deconv）、BN、ReLU 层构成，对不同步长的特征进行上采样，并在通道维度上进行堆叠融合，将结果作为最终的输出。

检测头在特征图上的每个锚点位置处预设朝向为 0°和 90°两个方向的锚框，则检测头的三个卷积模块输出的通道数量分别为锚框数乘以类别数量得到的结果、锚框数乘以 3D 框参数得到的结果、锚框数乘以方向数量得到的结果。损失函数及优化器等参数均与基础网络架构相同。

4　实验与分析

在单模态激光点云的 3D 目标检测架构

PointPillars[8]基础上，增加网络分支并优化网络结构，重点验证图像语义信息和空间可见性信息对3D行人检测的影响，并在主流数据集KITTI[14]和nuScenes[15]同时完成对比实验、定量分析实验、定性分析实验。KITTI数据集为安装有前视双目立体相机和64线激光雷达的自动驾驶汽车在20多个不同场景下采集的真实道路原始数据，并已公开语义分割和目标检测的标注信息。数据集中的每个目标按照标注检测框的最小边界框高度和遮挡程度分为三个难度等级（简单、中等、困难）。而nuScenes数据集为6个相机和1个32线激光雷达采集1000个场景的道路数据，在360°感知范围内标注交通目标的3D检测框，额外提供4个距离范围（0.5，1，2，4 m）的目标检测平均精度，专注于目标检测任务。实验基于先进的Pytorch深度学习框架，在Ubuntu 18.04系统的单个RTX 2080 Ti GPU上进行网络训练和测试。由于数据集存在部分差异，部分实验仅适用于单个数据集。

4.1 对比实验

在点云空间可见性特征编码过程中，需采用3个不同的数值描述体素空间可见性的三种状态。实验对比两种描述方式，即[Unknown为0，Occupied为1，Free为－1]和[Unknown为0.5，Occupied为0.7，Free为0.4][16]。实验结果如表1所示，使用[0.5，0.7，0.4]描述体素空间可见性状态，相比[0，1，－1]，平均精度均值（mAP）提升2.06个百分点。

表1 空间可见性描述方式在KITTI数据集的性能对比

Table 1 Performance comparison of different visibility description methods on KITTI dataset

Visibility code [U, O, F]	3D pedestrian detection AP /%			mAP /%
	Easy	Moderate	Hard	
[0, 1, −1]	68.76	62.49	57.74	63.00
[0.5, 0.7, 0.4]	70.84	64.57	59.77	65.06

点云空间可见性的特征编码需将感知空间划分为稠密体素，进而转换为可见性特征图。实验分别对比体素高度方向不同情况下稠密程度对目标检测的影响，即体素数量分别为1（体柱形式）和32[17]。不同距离区间范围的行人检测结果如表2所示。相比体柱形式的单通道特征，沿感知范围的高度方向划分为32个体素后mAP提升7.37个百分点。

表2 在nuScenes数据集上，体素稠密程度（高度方向）性能对比

Table 2 Performance comparison of different density along height direction on nuScenes dataset

Number of channels	3D pedestrian detection AP /%				mAP /%
	0.5 m	1.0 m	2.0 m	4.0 m	
1	62.24	64.36	66.36	68.78	65.44
32	69.68	71.87	73.73	75.97	72.81

基于单帧激光点云的目标检测，由于激光点云的稀疏性，较远距离行人的表面覆盖激光点较少，从而降低对远距离小目标的检测精度。若将相邻历史帧点云叠加至当前帧点云进行稠密化，即可增加行人表面点云数量。实验对比单帧点云和结合历史前10帧点云的行人检测结果，如表3所示。数据表明：结合历史多帧点云形成的稠密输入可获得更精确的三维目标检测；相比单帧点云，结合历史前10帧点云数据后平均检测精度提升32.63个百分点。实际算法在移植至实车应用平台时需对历史帧点云数据进行位置补偿。

表3 在nuScenes数据集上，不同帧数的性能对比

Table 3 Performance comparison of different number of frames on nuScenes dataset

Nunber of frames	3D pedestrian detection AP /%				mAP /%
	0.5 m	1.0 m	2.0 m	4.0 m	
1	37.46	38.22	39.29	40.40	38.84
10	68.36	70.63	72.30	74.59	71.47

4.2 定量分析实验

在对比实验基础上选取最优方案，同时融合图像语义信息，在主流数据集上与基准方法（Baseline）进行定量分析实验，在KITTI验证集上的定量实验结果如表4所示。表4中列出参考基准和3种不同优化配置方法的实验结果，其中PP是PointPillars缩写，Vis.是Visibility即可见性缩写，Img.是Image缩写。

表4 不同优化配置方法在KITTI验证集的性能对比

Table 4 Performance comparison of different optimized methods on KITTI validation set

Method	3D pedestrian detection AP /%			mAP /%
	Easy	Moderate	Hard	
PP（Baseline）	70.16	63.40	57.49	63.68
PP＋Vis.	71.86	64.49	58.72	65.02
PP＋Img.	72.08	65.56	60.34	65.99
PP＋Vis.＋Img.	71.84	65.65	60.81	66.10

表4实验结果表明：补充点云空间可见性信息后，行人检测性能在3D检测评估标准下的三个难度等级上都获得了明显的提升，其中在简单级别获得最大提升，为1.70个百分点，主要因为KITTI上定义的简单模式代表目标被遮挡面积较少，可观性与可见性从原理上表征的含义相同；增加图像语义信息后，三个难度等级下的行人检测性能同样都得到了显著提升，相比参考基准，困难级别下的检测精度提升最高，为2.85个百分点。与可见性信息的作用不同，图像信息更有助于困难模式检测目标，主要因为困难模式遮挡比较严重且目标表面点云更加稀疏，但通过出色的语义分割结果为点云提供类别信息，有效地弥补遮挡引起的点云缺失、稀疏而造成的检测性能降低的不足。补充图像信息后对行人检测精度的提升略大于补充可见性信息后的行人检测精度。同时补充图像和可见性信

息，行人检测发挥最佳性能，mAP 相比参考基准提升 2.42 个百分点。数据表明，同时利用图像和可见性信息能更有效地改善三维行人检测的综合性能。

在 KITTI 测试集上，对所提方法与目前先进的三维目标检测方法进行行人检测性能对比，结果如表 5 所示。

表 5　不同方法在 KITTI 测试集的性能比较

Table 5　Performance comparison of different methods on KITTI test set

Method	3D pedestrian detection AP /%			Speed / Hz
	Easy	Moderate	Hard	
VoxelNet[7]	39.48	33.69	31.51	4.4
AVOD[3]	36.10	27.86	25.76	10
SECOND[18]	51.07	42.56	37.29	20
F-PointNet[4]	50.53	42.15	38.08	5.9
PointPainting[6]	50.32	40.97	37.87	2.5
PointRCNN[19]	47.98	39.37	36.01	10
Proposed method	51.07	41.36	37.83	30

表 5 结果表明：相比其他三维目标检测方法，所提方法在简单模式的 3D 行人检测平均精度超过其他方法，而在其余两个难度级别上获得前三表现；虽近期提出较多其他先进框架，但所提方法的单帧数据处理速度为 30 Hz，从而更适用于自动驾驶车辆的车载计算平台。

4.3　定性分析实验

首先对点云空间可见性特征进行可视化，以更直观地理解可见性信息和真实环境的内在联系。其次从训练数据中选取几组样本进行目标检测结果的可视化，对所提方法和参考基准进行对比，从直观的角度对 3D 行人检测结果进行定性分析。

图 8(a)为单帧激光点云俯视图，图 8(b)为同帧点云数据空间可见性特征的中间层通道可视化示意图，为更清晰呈现可见性，将未知状态的体素设为 0.5，占据状态的体素设为 1，空闲状态的体素设为 0。图 8 结果表明：可见性特征图与真实环境具有较好的一致性，从定性的角度验证了空间可见性特征的准确性，为更精确的目标检测提供支撑。

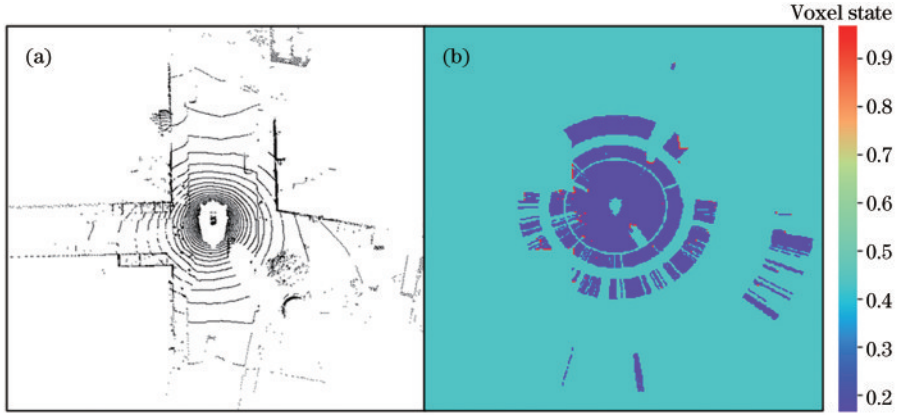

图 8　可见性特征可视化。(a)点云俯视图；(b)单层特征图

Fig. 8　Visibility feature visualization. (a) BEV of point cloud; (b) single layer feature

在 KITTI 数据集中选取几组样本，通过与基准方法进行对比，分别在图像和点云中对行人检测结果进行可视化。基准方法 PointPillars 和所提方法的检测结果如图 9 和图 10 所示，其中检测结果的三维框转换为

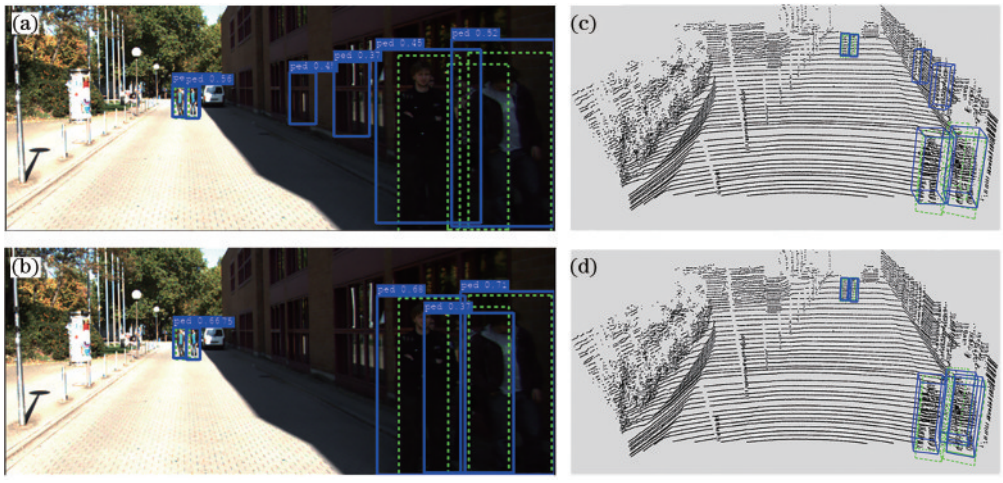

图 9　行人检测对比结果(示例一)。(a)(c)基准结果；(b)(d) 所提方法得到的结果

Fig. 9　Comparison results of pedestrian detection (example 1). (a) (c) Benchmark results; (b) (d) results obtained by proposed method

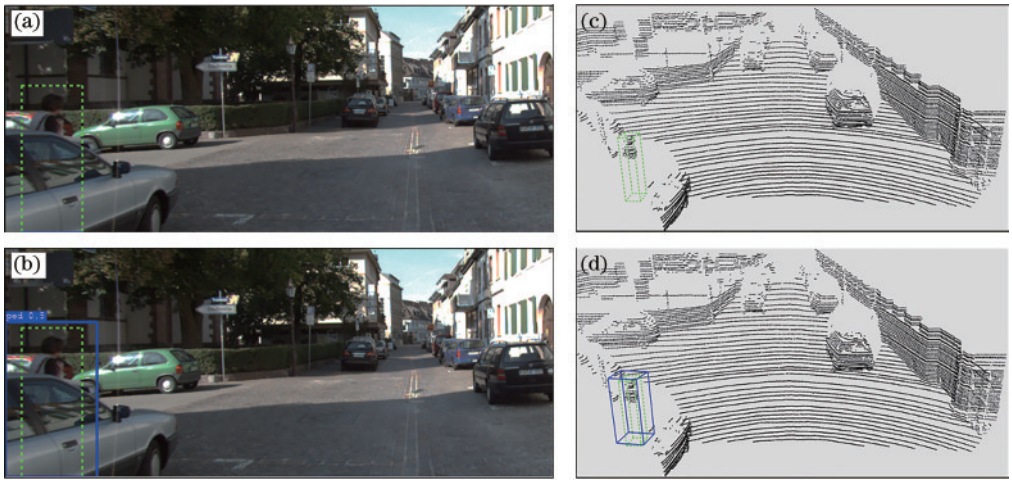

图 10　行人检测对比结果(示例二)。(a)(c)基准结果;(b)(d)所提方法得到的结果

Fig. 10　Comparison results of pedestrian detection (example 2). (a) (c) Benchmark results; (b) (d) results obtained by proposed method

二维框投影至图像进行显示,实线为预测框,虚线为真值框。分析图 9 可知:基于单模态激光点云的行人检测方法容易将杆状或柱状的窗户误检为行人,且无法准确地检测出并排行人中的单独个体,从而形成误检和漏检;所提方法将空间可见性信息和图像语义信息融入点云特征,不仅能预测所有的行人目标,而且显著提升预测边界框的置信度。

图 10 的示例样本结果表明:被严重遮挡的行人难以通过基于单模态激光点云的基准方法检测出来,主要因为该行人表面的点云较少,没有表现出行人的完整点云特征;补充了语义信息后,所提方法成功检测到该行人。结果说明在点云添加图像语义特征和点云空间可见性特征后,更利于检测严重遮挡或表面覆盖点云较少的环境目标。图 10(c)和图 10(d)为样本在点云中可视化结果对比,其中行人点云仅包含头部少量点云,几何特征不明显,补充语义信息和空间可见性特征的点云有助于解决有挑战性的(诸如遮挡)行人检测任务。

5　结　　论

旨在充分利用激光点云的空间特征,包括点云几何特征和空间可见性特征,并融合图像语义特征共同完成自动驾驶场景下的三维目标检测任务,通过实验探索融合过程中各核心因素对 3D 行人检测的影响程度。实验结果表明:结合点云空间可见性信息和图像语义信息后在目标观测性良好的情况下可进一步提升精度和置信度;而当目标存在部分遮挡时,图像语义信息可充分降低误检率和漏检率。在 KITTI 验证集中同时补充可见性特征和语义信息,所提方法在三个难度级别下的行人检测平均精度达 66.10%,相比参考基准,提升了 2.42 个百分点,且显著提升预测边界框的置信度,从而验证所提融合方法的有效性。

参　考　文　献

[1] 邹梓吟, 盖绍彦, 达飞鹏, 等. 基于注意力机制的遮挡行人检测算法[J]. 光学学报, 2021, 41(15): 1515001.
Zou Z Y, Gai S Y, Da F P, et al. Occluded pedestrian detection algorithm based on attention mechanism[J]. Acta Optica Sinica, 2021, 41(15): 1515001.

[2] 尧佼, 于凤芹. 基于候选区域定位与 HOG-CLBP 特征组合的行人检测[J]. 激光与光电子学进展, 2021, 58(2): 0210015.
Yao J, Yu F Q. Pedestrian detection based on combination of candidate region location and HOG-CLBP features[J]. Laser & Optoelectronics Progress, 2021, 58(2): 0210015.

[3] Ku J, Mozifian M, Lee J, et al. Joint 3D proposal generation and object detection from view aggregation[C]// 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), October 1-5, 2018, Madrid, Spain. New York: IEEE Press, 2018: 18392975.

[4] Charles R Q, Liu W, Wu C X, et al. Frustum PointNets for 3D object detection from RGB-D data[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 918-927.

[5] Charles R Q, Hao S, Mo K C, et al. PointNet: deep learning on point sets for 3D classification and segmentation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 77-85.

[6] 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 (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 4603-4611.

[7] 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, June 18-23, 2018, Salt Lake City, UT,

USA. New York: IEEE Press, 2018: 4490-4499.

[8] 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 (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 12689-12697.

[9] Zhu H Q, Deng J J, Zhang Y, et al. VPFNet: improving 3D object detection with virtual point based LiDAR and stereo data fusion[EB/OL]. (2021-11-29) [2022-01-08]. https://arxiv.org/abs/2111.14382.

[10] 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 (ICCV), October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2021: 15426-15436.

[11] Aleksandrov M, Zlatanova S, Heslop D J. Voxelisation algorithms and data structures: a review[J]. Sensors, 2021, 21(24): 8241.

[12] Siam M, Gamal M, Abdel-Razek M, et al. A comparative study of real-time semantic segmentation for autonomous driving[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 18-22, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 700-70010.

[13] Chen L C, Zhu Y K, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[EB/OL]. (2018-02-07)[2021-05-06]. https://arxiv.org/abs/1802.02611.

[14] 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, June 16-21, 2012, Providence, RI, USA. New York: IEEE Press, 2012: 3354-3361.

[15] 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 (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11618-11628.

[16] Hornung A, Wurm K M, Bennewitz M, et al. OctoMap: an efficient probabilistic 3D mapping framework based on octrees[J]. Autonomous Robots, 2013, 34(3): 189-206.

[17] Hu P Y, Ziglar J, Held D, et al. What You see is what You get: exploiting visibility for 3D object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 10998-11006.

[18] Yan Y, Mao Y X, Li B. SECOND: sparsely embedded convolutional detection[J]. Sensors, 2018, 18(10): 3337.

[19] Shi S S, Wang X G, Li H S. PointRCNN: 3D object proposal generation and detection from point cloud[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 770-779.