

DOI: 10.19650/j.cnki.cjsi.J2310994

# 基于对称形状生成的三维目标检测网络\*

涂新奎<sup>1</sup>, 郑少武<sup>1</sup>, 于善虎<sup>2</sup>, 李巍华<sup>1</sup>

(1. 华南理工大学机械与汽车工程学院 广州 510641; 2. 广州华工机动车检测技术有限公司 广州 510640)

**摘要:**基于点云的三维目标检测在机器人、自动驾驶等领域起着至关重要的作用,激光点云能为场景理解提供精确的几何信息。然而,由于点云的稀疏性和物体间的遮挡关系,激光点云通常只能描述物体的部分形状,导致目标结构信息不完整,从而降低检测精度。针对这个问题,提出基于对称形状生成的三维目标检测网络(SSG-RCNN),一种双阶段目标检测器。考虑到感兴趣目标形状的对称性,SSG-RCNN在一阶段生成候选框的同时为每个前景点预测镜像对称点,从而恢复目标的整体形状。二阶段中,使用自注意力池化层从原始点和对称点中聚合特征用于候选框修正,完成三维框预测。KITTI数据集上的实验表明SSG-RCNN取得了卓越的检测性能,在测试集上对困难目标的检测精度达到77.64%,高于所有对比方法。

**关键词:**三维目标检测;激光雷达点云;自动驾驶

**中图分类号:** TP391.4 TH744 **文献标识码:** A **国家标准学科分类代码:** 520.20

## 3D object detection network based on symmetric shape generation

Tu Xinkui<sup>1</sup>, Zheng Shaowu<sup>1</sup>, Yu Shanhu<sup>2</sup>, Li Weihua<sup>1</sup>

(1. School of Mechanical & Automotive Engineering, South China University of Technology, Guangzhou 510641, China;  
2. Guangzhou Huagong Motor Vehicle Inspection Technology Co., Ltd., Guangzhou 510640, China)

**Abstract:** 3D object detection based on point cloud is essential in many applications, such as robotics, autonomous driving. LiDAR point clouds contain reliable geometric information for 3D scene understanding. However, due to sparsity and occlusion, point clouds depict only partial surfaces of objects, which severely degrades the detection performance. To handle this challenge, we propose a novel two-stage detector based on symmetric shape generation (SSG-RCNN). The shapes of 3D interested objects are roughly symmetric. In the first stage, SSG-RCNN predicts a symmetric point for each foreground point to complete objects shapes while generating 3D proposals. In the second stage, SSG-RCNN utilizes self-attention pooling module to aggregate proposal-wise features from raw points and symmetric points. Finally, proposal-wise features are used to refine 3D proposals. Extensive experiments on KITTI benchmark show that SSG-RCNN has remarkable detection performance. Especially for hard difficulty level objects, SSG-RCNN achieves 77.64% AP on the KITTI test set, which is better than previous state-of-the-art methods.

**Keywords:** 3D object detection; LiDAR point cloud; autonomous driving

## 0 引 言

三维目标检测是实现自动驾驶中关键的一环。不同于二维目标检测负责输出目标的2D边界框,三维目标检测需要额外推理目标的深度信息,判断目标的3D位姿。得益于激光雷达点云携带的精准深度信息,基于激光点云的三维目标检测被应用到多个与环境感知

有关的领域,如自动驾驶汽车、机器人等<sup>[1-3]</sup>。与二维图像相比,激光雷达获取的点云具有丰富的三维结构信息,且受天气等因素的影响较小。然而,由于遮挡效应的存在,激光点云无法描述目标的完整形状,导致点云中物体的结构信息不完整,这给三维目标检测带来了很大的挑战。本文探究如何在三维目标检测任务中消除遮挡效应的影响,充分利用目标形状信息实现精准的三维框预测。

收稿日期:2023-01-18 Received Date: 2023-01-18

\* 基金项目:广州市重点领域研发计划项目(202206030005)资助

目前基于点云的三维目标检测算法研究可以分为两个方向,基于体素<sup>[4-7]</sup>的方法和基于点<sup>[8-12]</sup>的方法。基于体素的方法将无序的点云转化为体素栅格并进行体素特征编码,使用高效的3D卷积模块提取特征并完成目标检测。这种方法的检测效率高,但是在点云体素化的过程中会丢失部分结构信息。基于点的方法使用PointNet<sup>[13-14]</sup>等模型直接从原始点云中提取特征,保证了特征的精细度。但是如果为所有点提取高维特征,需要庞大的计算量,因此基于点的方法通常与其他方法结合使用。一些工作<sup>[15-18]</sup>将点级特征嵌入到体素特征用于弥补体素化中丢失的信息。上述的方法均是在三维真值框的监督下完成网络的学习,没有显式地考虑目标形状信息。

面对物体结构信息缺失的问题,目标检测网络需要充分理解三维目标形状才能从局部点云推理出精准的3D边界框。一些工作通过强调点在目标上的位置来学习目标形状。SA-SSD<sup>[16]</sup>提出一项附加任务用于学习前景点与目标中心点之间的坐标偏移,在附加任务的监督下,网络提取的特征能够携带丰富的结构信息。LiDAR-RCNN<sup>[19]</sup>提出一种通用的二阶段模块用于修正候选三维框,在二阶段编码点级特征时,LiDAR-RCNN引入点与候选框边界之间的坐标偏移来表征结构信息。这些方法能够利用到部分形状信息,但是忽略了目标的被遮挡区域。为了减轻遮挡效应的影响,一些工作尝试预测目标的被遮挡区域来显式地恢复目标的完整形状。SIENet<sup>[20]</sup>在二阶段中使用对抗生成模型将候选框附近的残缺点云恢复成完整的目标形状,但是需要使用额外的数据集单独训练生成式模型。SPG<sup>[21]</sup>是一种通用的前景点恢复网络,通过在空体素中生成新的点来补充目标点云。由于无法得知目标的真实形状,SPG将三维框内的所有区域都视为前景区域,这意味生成点不仅位于目标表面,也会位于目标内部。这种数据形式与真实点云具有较大的差异,会导致网络的训练困难。

观察到交通场景下的感兴趣三维目标大致可以视为对称形状(如汽车),即对于三维目标上的一个点,始终存在着另一个镜像对称的点。如果生成点云中所有前景点的镜像对称点,可以增加大约一倍的前景点数,有效地恢复目标形状信息。这种方式生成的对称点分布在目标的表面,与真实点云的数据形式保持一致。基于上述思想,本文提出一种双阶段检测算法,基于对称形状生成的三维目标检测网络(symmetric shape generation, SSG-RCNN)。在一阶段中区域提案网络(region proposal network, RPN)生成候选框的同时,设计的对称点生成模块提取点级特征,预测前景点的对称点。将前景点沿着三维真值框的中轴面镜像翻折,可以得到用于监督学习的真实对称点标签。在二阶段中,设计的自注意力池化

层同时采样候选框附近的原始点和对称点,使用自注意力机制建模候选框内的全局语义信息,生成聚合特征用于候选框修正。一阶段预测的对称点完善了目标形状信息,二阶段引入的自注意力机制能够建模点与点之间的关联性,二者的结合使得网络可以更加准确地优化候选框位姿,提高三维目标检测的精度。在KITTI<sup>[22]</sup>数据集上的实验证明所提网络取得了可观的检测效果。综上所述,本文的主要贡献如下:

1) 提出一种恢复三维目标形状信息的策略,通过生成前景点的对称点来恢复三维目标的被遮挡区域。该方法无需训练生成式模型,且恢复的对称点符合真实数据分布。

2) 提出一种基于对称形状生成的三维目标检测网络,在一阶段中预测候选框的同时生成对称点。并且设计了一种基于自注意力机制的候选框特征池化层,在二阶段中从原始点和对称点中提取特征用于候选框修正。

## 1 理论分析

SSG-RCNN的整体结构如图1所示。在1阶段中,首先进行点云体素化,将三维点云空间转化为体素栅格,每个体素中保留的最大点数为 $T$ ,体素特征编码为体素中所有点的平均值。堆叠3D稀疏卷积层对体素栅格进行下采样的同时扩大特征感受野,得到表征局部语义信息的三维特征体,空间分辨率为初始分辨率的 $1/8$ 。RPN模块将三维特征体降维成二维鸟瞰图特征,使用二维卷积生成候选框。对称点生成模块将三维特征体上采样到原始空间分辨率得到点级特征,利用点级特征完成前景点/背景点分类。如果一个点被判断为前景点,网络会进一步为它预测镜像对称点。RPN模块和对称点生成模块同时使用三维特征体完成不同的任务,这种特征共享模式避免了特征的重复提取,且各自的损失函数可以为对方的特征起到附加监督的效果。

在2阶段中,自注意力池化层负责聚合候选框特征,并利用这些特征完成更精细的三维框修正。一阶段生成的候选框被视为感兴趣区域,网络在感兴趣区域附近采样一定数量的点,包括原始点和对称点,使用线性层提取高维点级特征。自注意力池化层对点级特征进行注意力编码用于建模点与点之间的关联性,随后将特征解码,得到一个表征候选框全局空间语义信息的池化特征。该特征最后通过全连接层进行精准的三维框分类和回归任务。接下来会详细介绍所提网络的各个模块。

### 1.1 一阶段候选框生成

由于3D稀疏卷积<sup>[23]</sup>的高效性和高精度,实验中堆叠3D稀疏卷积层用于三维特征提取。3D稀疏卷积包含两种卷积模式,空间稀疏卷积和子流形卷积。特征提

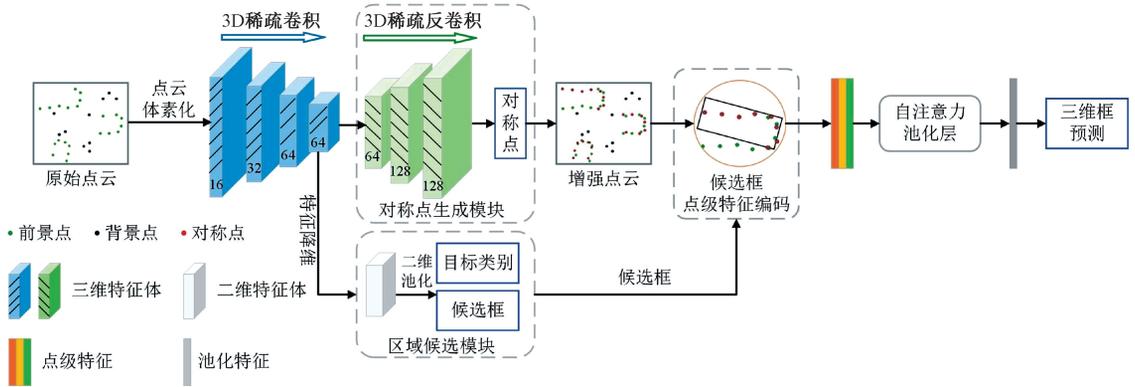


图1 SSG-RCNN网络结构图

Fig. 1 The architecture of SSG-RCNN

取过程分为一个输入层和3个下采样层,输入层由两个 $3 \times 3 \times 3$ 子流形卷积构成,用于提取初始特征。下采样层首先是一个步长为2的 $3 \times 3 \times 3$ 空间稀疏卷积,随后是连续的 $3 \times 3 \times 3$ 子流形卷积。每个卷积后面都会跟随批正则化操作和ReLU激活函数。通过特征提取器,体素栅格被转化为 $1/8$ 分辨率的三维特征体。

RPN模块负责生成一阶段的候选三维框。首先沿着Z轴拼接三维特征体,得到鸟瞰视角下的二维特征图。随后使用一系列标准 $3 \times 3$ 的2D卷积进一步聚合特征,最后利用两个线性层分别完成分类任务和三维框回归任务。采用anchor-based的方式生成三维框,在每个特征位置上设置两种朝向的anchor, $0^\circ$ 和 $90^\circ$ 。对于分类任务,网络预测anchor属于前景目标的概率;对于回归任务,网络预测三维框与anchor之间残差值。

## 1.2 对称点生成模块

如图1所示,对称点生成模块的输入同样是由特征提取器得到的下采样三维特征体。可以看出,特征提取器与对称点生成模块构成一种类似Unet的网络结构。这种下采样-上采样模式使得最后得到的点级特征具有较大的感受野,可以有效地表征每个点邻域内的点云信息。

### 1) 点级特征提取

使用3个上采样层完成点级特征提取,上采样层由一个 $3 \times 3 \times 3$ 稀疏反卷积和两个 $3 \times 3 \times 3$ 子流形卷积构成。稀疏反卷积用于恢复三维特征体的分辨率,子流形卷积负责进一步聚合深层特征信息。用 $\{f_i \in \mathbf{R}^C; i = 1, \dots, M\}$ 表示提取到点级特征, $M$ 表示体素化后的点数, $C$ 表示特征通道数。

### 2) 对称点生成

点级特征通过线性层映射形成共享特征 $\{\tilde{f}_i \in \mathbf{R}^C; i = 1, \dots, M\}$ ,此过程中保持特征通道数 $C$ 不变。使用一个线性层加sigmoid函数作为分类head,预测当前点属于前景点的概率 $\tilde{p}_i$ 。另一个线性层作为回归head,预测前

景点与对称点之间的坐标偏移量 $\tilde{s}_i, \tilde{s}_i$ 具有两个维度,代表X轴与Y轴上的坐标差值。由于三维目标检测中认为目标侧倾角为 $0^\circ$ ,可以假设对称点偏移只发生在X轴与Y轴,Z轴坐标保持不变。

### 3) 训练目标

使用 $p_i$ 表示第 $i$ 个点的类别标签。如果点位于三维标注框内部,则该点为前景点且 $p_i = 1$ ,否则该点为背景点且 $p_i = 0$ 。使用Focal Loss<sup>[24]</sup>来优化前景点/背景点分类任务,Focal Loss可以缓解点云中前景点-背景点数量不平衡的问题,分类损失如下:

$$L_{cls}^{sym} = \frac{1}{M_{pos}} \sum_i -\alpha(1 - \hat{p}_i)^\gamma \log(\hat{p}_i) \quad (1)$$

$$\hat{p}_i = \begin{cases} \tilde{p}_i, & p_i = 1 \\ 1 - \tilde{p}_i, & p_i = 0 \end{cases} \quad (2)$$

其中, $M_{pos}$ 是前景点(正样本)的总数, $\alpha$ 和 $\gamma$ 是用于调节损失函数比例的超参数,实验中设置 $\alpha = 0.25, \gamma = 2$ 。

$s_i$ 表示第 $i$ 个点与真实对称点之间的坐标偏移量标签。将前景点沿着当前目标三维标注框的中轴线镜像翻转可以得到真实对称点。使用Smooth- $l_1$ 函数来优化对称点回归任务,回归损失如下:

$$L_{reg}^{sym} = \frac{1}{M_{pos}} \sum_i \text{Smooth} - l_1(s_i, \tilde{s}_i) p_i \quad (3)$$

其中, $p_i$ 是对应的类别标签,表明只有对正样本的预测才会参与回归损失的计算。

结合分类和回归任务,对称点生成模块可以自行生成前景点的对称点从而恢复目标形状,为二阶段的候选框修正提供更完整的目标点云。2.3节的实验结果证明了对称点生成可以有效地提高目标检测精度。

### 1.3 自注意力池化层

网络第二阶段负责聚合候选框的空间特征,对候选框的位置和尺寸做进一步调整,使其更接近于目标的真实

位姿。显然,最终的预测精度与候选框特征提取的效果紧密相关。为此,本文提出自注意力池化层用于提取具有全局空间信息的候选框特征。自注意力池化层在候选

框附近同时采样原始点和对称点,获得较完整的目标形状信息,并利用自注意机制编码特征。详细过程如图2所示。

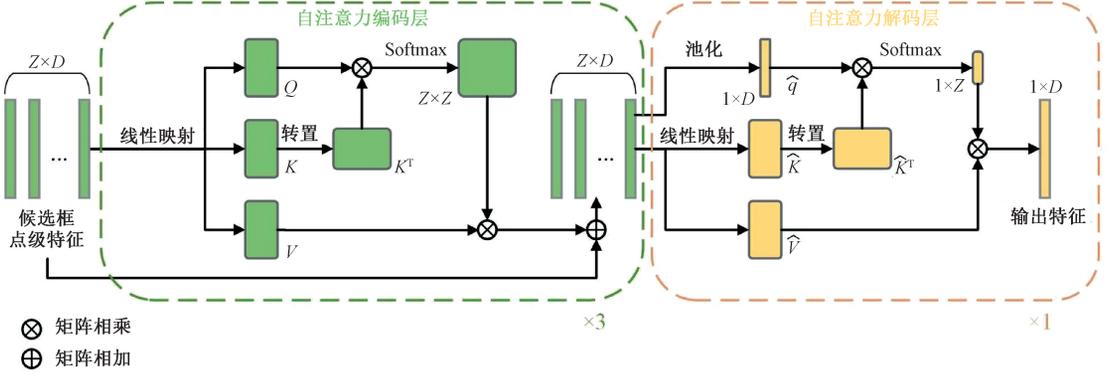


图2 自注意力池化层结构图

Fig. 2 The architecture of self-attention pooling module

### 1) 点级特征编码

首先确定点的采样范围,设置一个圆柱体区域表示候选框邻域采样范围。用参数 $(x_p, y_p, z_p, l_p, w_p, h_p, \theta_p)$ 表示候选框位姿, $(x_p, y_p, z_p)$ 表示候选框中心坐标, $(l_p, w_p, h_p)$ 分别表示长宽高, $\theta_p$ 表示朝向角。采样圆柱体的中心与候选框中心一致。圆柱的直径为 $\beta\sqrt{l_p^2 + w_p^2}$ ,与候选框底面对角线长度正相关, $\beta$ 是直径增益。圆柱的高度为 $h_p + 2\lambda$ , $\lambda$ 为高度增益。这种采样范围的设置可以大幅减少候选框偏离真实框的影响。

在圆柱体区域内随机采样 $Z$ 个点,包括原始点和对称点。对采样点进行特征编码时,使用了Lidar-RCNN<sup>[19]</sup>中的边界偏移结构,用于强化网络对候选框形状的感知能力。每个点的特征编码表示为 $(x, y, z, d_1, d_2, d_3, d_4, d_5, d_6, c)$ ,其中 $(x, y, z)$ 表示点的三维坐标, $d_1$ 到 $d_6$ 分别表示点距离候选框6个面的垂直距离。 $c$ 表示置信度,对于原始点,置信度设为1;对于生成的对称点,置信度为对称点生成模块中预测的前景概率。编码的点级特征通道数为10,通过线性层映射成通道数为 $D$ 的高维点级特征,实验中设置 $D = 256$ 。

### 2) 自注意力编码

如图2所示,本文采用的自注意力编码层结构与文献<sup>[25]</sup>保持一致。高维点级特征记为 $\mathbf{X}(\mathbf{X} \in \mathbb{R}^{Z \times D})$ ,依次通过多头自注意力模块和前馈传播网络建模点与点之间的关联性。图2中 $\mathbf{Q} = \mathbf{X}\mathbf{W}_q, \mathbf{K} = \mathbf{X}\mathbf{W}_k, \mathbf{V} = \mathbf{X}\mathbf{W}_v$ 。其中 $\mathbf{W}_q, \mathbf{W}_k, \mathbf{W}_v \in \mathbb{R}^{D \times D}$ ,均为线性映射。在多头自注意力机制中,将 $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 转化为 $\mathbf{Q} = [\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_H], \mathbf{K} = [\mathbf{K}_1, \mathbf{K}_2, \dots, \mathbf{K}_H], \mathbf{V} = [\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_H]$ ,其中 $\mathbf{Q}_h, \mathbf{K}_h, \mathbf{V}_h \in \mathbb{R}^{Z \times d}, d = \frac{D}{H}, h = 1, \dots, H, H$ 为自注意力头数。得到的多

头自注意力特征如下:

$$\mathbf{X}^{att} = \text{Cat} \left( \sigma \left( \frac{\mathbf{Q}_h \mathbf{K}_h^T}{\sqrt{d}} \right) \mathbf{V}_h \right), h = 1, \dots, H \quad (4)$$

其中,  $\text{Cat}(\cdot)$ 表示特征拼接操作,  $\sigma(\cdot)$ 表示softmax操作。添加残差连接后输出特征如下:

$$\mathbf{X}^{res} = \eta(\mathbf{X}^{att} + \mathbf{X}) \quad (5)$$

其中,  $\eta(\cdot)$ 表示正则化操作。随后通过全连接层构成的前馈传播网络,最终的编码特征为:

$$\mathbf{X}^{encode} = \eta(\tau(\mathbf{X}^{res}) + \mathbf{X}^{res}) \quad (6)$$

其中,  $\tau(\cdot)$ 表示前馈传播网络。实验中堆叠3层自注意力编码层构建自注意力编码器。

### 3) 自注意力解码

解码器将编码器生成的高维点级特征(记为 $\hat{\mathbf{X}}$ )解码成一个全局特征,最后通过线性层完成候选框的位姿修正。与Transformer中的原有解码器不同,预测三维框最终只需要1个特征,因此解码时也只需要1个查询特征 $\hat{\mathbf{q}}$ 。为了表征候选框空间的全局信息,对 $\hat{\mathbf{X}}$ 使用最大池化生成 $\hat{\mathbf{q}} \in \mathbb{R}^D$ :

$$\hat{\mathbf{q}} = \text{MaxPooling}(\hat{\mathbf{X}}) \quad (7)$$

$\hat{\mathbf{X}}$ 通过线性映射分别得到键特征 $\hat{\mathbf{K}} \in \mathbb{R}^{Z \times D}$ 和值特征 $\hat{\mathbf{V}} \in \mathbb{R}^{Z \times D}$ 。编码器中同样使用多头机制,将 $\hat{\mathbf{q}}, \hat{\mathbf{K}}, \hat{\mathbf{V}}$ 转化为 $\hat{\mathbf{q}} = [\hat{\mathbf{q}}_1, \hat{\mathbf{q}}_2, \dots, \hat{\mathbf{q}}_H], \hat{\mathbf{K}} = [\hat{\mathbf{K}}_1, \hat{\mathbf{K}}_2, \dots, \hat{\mathbf{K}}_H], \hat{\mathbf{V}} = [\hat{\mathbf{V}}_1, \hat{\mathbf{V}}_2, \dots, \hat{\mathbf{V}}_H]$ 。自注意力解码器的最终输出 $\hat{\mathbf{x}}^{decode}(\hat{\mathbf{x}}^{decode} \in \mathbb{R}^D)$ 由下式给出:

$$\hat{\mathbf{x}}^{att} = \text{Cat} \left( \sigma \left( \frac{\hat{\mathbf{q}}_h \hat{\mathbf{K}}_h^T}{\sqrt{d}} \right) \hat{\mathbf{V}}_h \right), h = 1, \dots, H \quad (8)$$

$$\hat{\mathbf{x}}^{res} = \eta(\hat{\mathbf{x}}^{att} + \hat{\mathbf{q}}) \quad (9)$$

$$\hat{\mathbf{x}}^{decode} = \eta(\tau(\hat{\mathbf{x}}^{res}) + \hat{\mathbf{x}}^{res}) \quad (10)$$

#### 4) 训练目标

自注意力池化层输出全局特征后,网络使用两个分支完成最终的三维框预测。对于分类任务,网络预测候选框属于正样本的概率 $\hat{y}$ ,训练标签 $y$ 由候选框和真实框之间的交并比(intersection over union, IoU)决定, $y = \min(1, \max(0, 2IoU - 0.5))$ 。分类损失如下:

$$L_{cls}^{ref} = -y \log(\hat{y}) - (1 - y) \log(1 - \hat{y}) \quad (11)$$

对于三维框回归任务,网络预测候选框与真实框之间的残差值 $\tilde{\mathbf{u}}$ ,真实标签为 $\mathbf{u}$ ,回归损失使用 Smooth- $l_1$  函数:

$$L_{reg}^{ref} = \text{Smooth} - l_1(\tilde{\mathbf{u}}, \mathbf{u}) \quad (12)$$

#### 1.4 损失函数

采用端到端的模式来训练整个目标检测网络,整体损失函数包括一阶段候选框预测损失函数 $L_{rpm}$ ,对称点生成损失函数 $L_{sym}$ ,二阶段候选框修正损失函数 $L_{ref}$ 。 $L_{rpm}$ 的构建与文献[5]保持一致,分类损失使用 Focal Loss 函数,回归损失使用 Smooth- $l_1$  函数:

$$L_{rpm} = \alpha_{rpm} L_{cls}^{rpm} + \beta_{rpm} L_{reg}^{rpm} \quad (13)$$

$L_{sym}$  包括 1.2 节中提到的损失:

$$L_{sym} = \alpha_{sym} L_{cls}^{sym} + \beta_{sym} L_{reg}^{sym} \quad (14)$$

$L_{ref}$  包括 1.3 节中提到的损失:

$$L_{ref} = \alpha_{ref} L_{cls}^{ref} + \beta_{ref} L_{reg}^{ref} \quad (15)$$

总体损失函数为:

$$L_{total} = L_{rpm} + L_{sym} + L_{ref} \quad (16)$$

实验中设置 $\alpha_{rpm} = 1.0, \beta_{rpm} = 2.0; \alpha_{sym} = 1.0, \beta_{sym} = 2.5; \alpha_{ref} = 1.0, \beta_{ref} = 1.0$ 。

## 2 实验验证

### 2.1 KITTI 三维目标检测数据集

研究中使用 KITTI 三维目标检测数据集评估所提网络的检测性能。该数据集是用于评估交通场景下环境感知技术的大型数据集,它包含了 7 481 份训练样本和 7 818 份测试样本,每份样本中包括激光点云、双目图像和目标真实标签。其中,测试数据不含真实标签,需要将检测结果提交到 KITTI 官网评估精度。每种目标类别根据大小、遮挡程度分为简单、中等和困难 3 种级别。依据 KITTI 协议,将训练数据分为训练集和验证集。训练集包含 3 712 份数据样本,验证集包含 3 769 份数据样本。在测试集上评估网络时,使用 90% 的数据样本进行训练,10% 的数据样本进行验证,最后将网络对测试数据的预测结果提交到 KITTI 官网进行评估。

### 2.2 SSG-RCNN 实现细节

目标检测考虑的点云范围沿着 X、Y 和 Z 轴分别为  $[0, 70.4]$  m、 $[-40, 40]$  m 和  $[-3, 1]$  m,点云体素化中

使用的体素尺寸为  $[0.05, 0.05, 0.1]$  m,体素内保留的最大点数  $T$  设置为 5。如图 1 所示,特征提取器中 3D 稀疏卷积层的通道数分别为 16、32、64 和 64,对称点云生成模块中用于上采样的 3D 稀疏反卷积层的通道数分别为 64、128 和 128。关于二阶段中候选框采样圆柱范围的设置,直径增益  $\beta = 1.5$ ,高度增益  $\lambda = 0.2$ ,采样点数  $Z = 384$ 。

训练过程中,一阶段生成候选框时,前景 IoU 阈值为 0.6,背景 IoU 阈值为 0.45。二阶段修正候选框时,前景 IoU 阈值为 0.75,背景 IoU 阈值为 0.25。随机采样 128 个候选框用于修正,其中正样本和负样本各占一半。只有当  $IoU > 0.55$ ,候选框的边界回归才会参与回归损失的计算。推理过程中,使用 IoU 阈值为 0.7 的非极大值抑制筛选出 100 个感兴趣候选框,完成修正后,利用 IoU 阈值为 0.1 的非极大值抑制去除冗余的检测结果。

采用端到端的模式训练整个目标检测网络,选用 ADAM 优化器,在 2 张 3090 显卡上训练网络, batch size 设置为 2,学习率设置为 0.01,迭代次数为 100 个 epochs,花费 10 h 完成网络的训练。学习率的变化采用余弦退火策略。

### 2.3 目标检测性能评估

#### 1) 评估标准

使用平均精度(average precision, AP)来评估三维目标检测的效果,平均精度综合考虑了目标检测的查准率和查全率,能够定量地评估检测网络的整体性能。

#### 2) 对称点生成有效性

首先设计实验验证网络生成的对称点可以有效地恢复目标形状信息。如图 3 所示,选取 SSG-RCNN 网络中的对称点生成线路(主要包括特征提取和对称点生成模块)作为一个独立网络。该网络可以附加到任何现有的目标检测器之前用于生成对称点,将原始点云和对称点拼接后形成增强点云,现有目标检测器消耗增强点云实现更精准的三维目标检测。

选取 3 种主流的目标检测网络, PointPillars、SECOND 和 PV-RCNN,使用 OpenPCDet<sup>[26]</sup> 框架实现,在 KITTI 验证集上对比加入独立网络前后对 Car 类别的检测性能。由表 1 可知,独立网络的加入显著地提高了各个检测器的性能,证明了对称点生成的有效性和通用性。具体而言,对于 PointPillars,加入独立网络后在 3 种级别上的精度分别提升 2.68%, 1.73%, 0.34%; 对于 SECOND,提升为 1.56%, 1.48%, 1.32%; 对于 PV-RCNN,提升为 0.37%, 0.79%, 0.28%。实验结果表明,生成对称点可以让检测网络感知到更加丰富的目标形状,进而提高目标检测的精度。

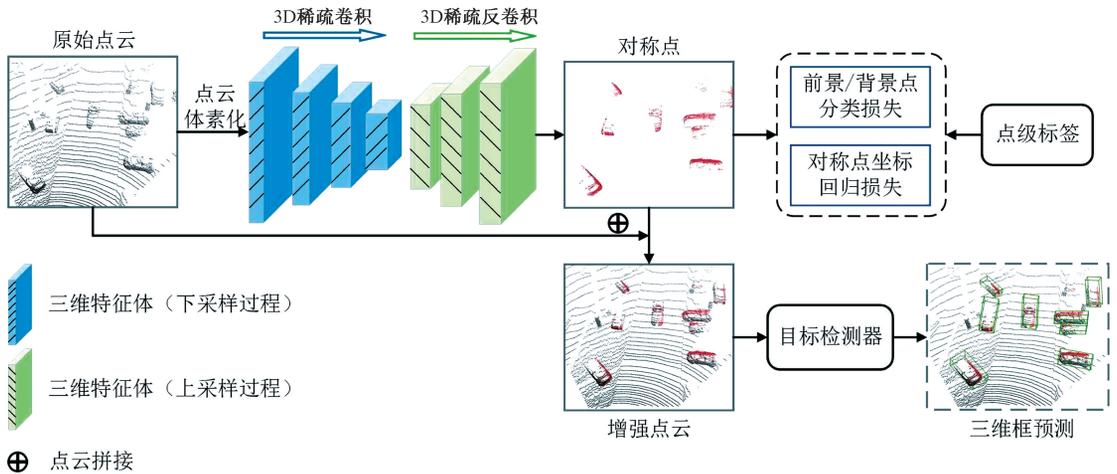


图 3 独立网络结构图

Fig. 3 The architecture of independent network

表 1 KITTI 验证集上对称点生成有效性验证

Table 1 Validity verification of symmetric point generation

| 方法                          | Car-3D detection (IoU 0.7) |              |              |
|-----------------------------|----------------------------|--------------|--------------|
|                             | 简单                         | 中等           | 困难           |
| PointPillars <sup>[6]</sup> | 87.72                      | 78.69        | 76.12        |
| PointPillars+独立网络           | <b>90.40</b>               | <b>80.42</b> | <b>76.46</b> |
| 提升                          | 2.68                       | 1.73         | 0.34         |
| SECOND <sup>[5]</sup>       | 90.22                      | 81.54        | 78.69        |
| SECOND+独立网络                 | <b>91.78</b>               | <b>83.02</b> | <b>80.01</b> |
| 提升                          | 1.56                       | 1.48         | 1.32         |
| PV-RCNN <sup>[15]</sup>     | 92.10                      | 84.36        | 82.48        |
| PV-RCNN+独立网络                | <b>92.47</b>               | <b>85.15</b> | <b>82.76</b> |
| 提升                          | 0.37                       | 0.79         | 0.28         |

3) 验证集实验结果

在 KITTI 验证集上对比 SSG-RCNN 与其他主流算法的检测效果,其他算法的精度取实验复现结果和原论文公开结果的最高值。表 2 使用的指标 AP(R11) 表示使用 11 个召回点计算的平均精度。如表 2 所示,所提网络对 Car 类别的 3 种级别的检测精度分别为 89.67%, 84.00%, 79.31%, 在简单和中等级别上均取得了较高的精度,而在困难级别上取得的精度优于所有列举的方法。困难级别意味着当前目标的截断程度较大,属于该目标的前景点数较少,而所提网络可以通过生成目标的前景点恢复形状信息,从而提高困难目标的检测精度。由此可以验证对称点生成策略可以有效地增加前景点数并恢复目标形状,减少点云遮挡的影响。此外,SSG-RCNN 对 Cyclist 和 Pedestrian 两类目标的中等级别检测精度分别

为 71.57% 和 59.77%,相较于基准模型 SECOND 提升了 3.82% 和 6.79%。

表 3 使用的指标 AP(R40) 表示使用 40 个召回点计算的平均精度,所提网络对 Car 类别的 3 种级别的检测精度分别为 92.84%, 85.27%, 83.10%, 在困难级别上依然取得了所列方法中最高检测精度。相比于最新的方法 DSPF-RCNN,所提网络在 Car 类别的 3 种级别上分别取得了 0.35%、0.02% 和 0.28% 的精度提升,证明了所提网络检测性能的先进性。对于 Cyclist 和 Pedestrian 两类目标的中等级别,所提网络在 SECOND 的基础上提升了 2.85% 和 8.08% 的检测精度。

除此以外,本文还评估了对称点生成精度,包括 2 个评估指标,前景点查全率和对称点生成的位置误差。前景点查全率是预测成功的前景点数量与所有前景点数量的比值,对称点生成的位置误差是生成对称点与真实对称点之间的绝对距离。实验发现 SSG-RCNN 中对称点生成模块与区域候选模块使用同 1 个三维特征体时,各自的损失函数会对对方的特征起到附加监督的效果,称这种训练模式为联合训练,将单独训练对称点生成模块的模式称为独立训练,同时评估联合训练和独立训练下的对称点生成精度。将不同距离的点分为不同的级别,0~20 m 为近距离,20~40 m 为中等距离,40 m 以外为远距离。对各种级别以及总体的对称点生成精度分别评估。如表 4 所示,在对称点生成的位置误差相差极小的情况下,联合训练模式可以得到更高的前景点查全率。

4) 测试集实验结果

使用 90% 的训练样本完成 SSG-RCNN 网络的训练,在 KITTI 测试集上进行实验。KITTI 官网使用的指标为 AP(R40),表示使用 40 个召回点计算的平均精度。如表 5 所示,所提网络对 Car 类别下 3 种级别的检测精

表2 KITTI 验证集上检测性能对比 (AP-R11)

Table 2 Detection performance on KITTI validation set (AP-R11)

| 方法                          | 收录          | Car-3D AP (R11) |              |              | Cyc-3D AP (R11) | Ped-3D AP (R11) |
|-----------------------------|-------------|-----------------|--------------|--------------|-----------------|-----------------|
|                             |             | 简单              | 中等           | 困难           | 中等              | 中等              |
| VoxelNet <sup>[4]</sup>     | CVPR2018    | 81.97           | 65.46        | 62.85        | 47.65           | 53.42           |
| SECOND <sup>[5]</sup>       | Sensors2018 | 88.61           | 78.62        | 77.22        | 67.75           | 52.98           |
| PointPillars <sup>[6]</sup> | CVPR2019    | 86.62           | 76.06        | 68.91        | 62.68           | 52.29           |
| SA-SSD <sup>[16]</sup>      | CVPR2020    | 90.15           | 79.91        | 78.78        | -               | -               |
| 3DSSD <sup>[10]</sup>       | CVPR2020    | 89.71           | 79.45        | 78.67        | -               | -               |
| PointRCNN <sup>[8]</sup>    | CVPR2019    | 88.88           | 78.63        | 77.38        | 69.70           | <b>63.70</b>    |
| PV-RCNN <sup>[15]</sup>     | CVPR2020    | 89.35           | 83.69        | 78.70        | 70.47           | 57.90           |
| VoxelRCNN <sup>[7]</sup>    | AAAI2021    | 89.41           | <b>84.52</b> | 78.93        | -               | -               |
| Votr-TSD <sup>[27]</sup>    | ICCV2021    | 89.04           | 84.04        | 78.68        | -               | -               |
| SVGA-Net <sup>[28]</sup>    | AAAI2022    | <b>90.59</b>    | 80.23        | 79.15        | -               | -               |
| DSPF-RCNN <sup>[29]</sup>   | 中国激光 2022   | 89.59           | 84.40        | 78.99        | <b>73.90</b>    | 60.10           |
| SSG-RCNN                    | -           | 89.67           | 84.00        | <b>79.31</b> | 71.57           | 59.77           |

表3 KITTI 验证集上检测性能对比 (AP-R40)

Table 3 Detection performance on KITTI validation set (AP-R40)

| 方法                          | 收录          | Car-3D AP (R40) |              |              | Cyc-3D AP (R40) | Ped-3D AP (R40) |
|-----------------------------|-------------|-----------------|--------------|--------------|-----------------|-----------------|
|                             |             | 简单              | 中等           | 困难           | 中等              | 中等              |
| SECOND <sup>[5]</sup>       | Sensors2018 | 90.22           | 81.54        | 78.69        | 68.71           | 52.07           |
| PointPillars <sup>[6]</sup> | CVPR2019    | 87.75           | 78.39        | 75.18        | 62.94           | 51.41           |
| SA-SSD <sup>[16]</sup>      | CVPR2020    | 92.23           | 84.30        | 81.36        | -               | -               |
| PV-RCNN <sup>[15]</sup>     | CVPR2020    | 92.57           | 84.83        | 82.86        | 71.95           | 56.67           |
| VoxelRCNN <sup>[7]</sup>    | AAAI2021    | 92.38           | 85.29        | 82.86        | -               | -               |
| SPG <sup>[21]</sup>         | ICCV2021    | 92.53           | <b>85.31</b> | 82.69        | 74.35           | <b>61.80</b>    |
| DSPF-RCNN <sup>[29]</sup>   | 中国激光 2022   | 92.49           | 85.25        | 82.82        | <b>75.20</b>    | 60.07           |
| SSG-RCNN                    | -           | <b>92.84</b>    | 85.27        | <b>83.10</b> | 71.56           | 60.15           |

表4 对称点生成精度

Table 4 Accuracy of symmetric point generation

| 模式   | 前景点查全率/%     |              |              |              | 对称点生成的位置误差/m |             |             |             |
|------|--------------|--------------|--------------|--------------|--------------|-------------|-------------|-------------|
|      | 近            | 中等           | 远            | 总体           | 近            | 中等          | 远           | 总体          |
| 联合训练 | <b>89.22</b> | <b>86.97</b> | <b>64.30</b> | <b>88.28</b> | <b>0.08</b>  | <b>0.11</b> | <b>0.16</b> | 0.09        |
| 单独训练 | 85.53        | 82.42        | 61.12        | 84.42        | 0.08         | 0.11        | 0.17        | <b>0.08</b> |

度分别为 88.35%, 82.11%, 77.64%, 在简单级别上取得了较高的检测精度; 在中等和困难级别上取得了最高的检测精度, 优于所有列举的方法。实验结果充分证明了所提网络具有先进的目标检测性能。

具体而言, 相比最新的检测模型 DSPF-RCNN, 所提网络在 Car 类别的中等和困难级别上分别取得了 1.07%

和 1.10% 的精度提升。这些精度上的提升主要来源于对称点生成恢复的三维目标结构信息, 以及自注意力池化层对目标全局结构信息的建模。

#### 2.4 可视化结果

为了更直观地分析 SSG-RCNN 在三维目标检测性能上的提升, 展示部分在 KITTI 数据集上的可视化结果,

表 5 KITTI 测试集上检测性能对比 (AP-R40)

Table 5 Detection performance on KITTI testing set (AP-R40)

| 方法                          | 收录          | Car-3D AP( R40) |              |              |
|-----------------------------|-------------|-----------------|--------------|--------------|
|                             |             | 简单              | 中等           | 困难           |
| VoxelNet <sup>[4]</sup>     | CVPR2018    | 77.47           | 65.11        | 57.73        |
| SECOND <sup>[5]</sup>       | Sensors2018 | 83.34           | 72.55        | 65.82        |
| PointPillars <sup>[6]</sup> | CVPR2019    | 82.58           | 74.31        | 68.99        |
| SA-SSD <sup>[16]</sup>      | CVPR2020    | 88.75           | 79.79        | 74.16        |
| 3DSSD <sup>[10]</sup>       | CVPR2020    | 88.36           | 79.57        | 74.55        |
| PointRCNN <sup>[8]</sup>    | CVPR2019    | 86.96           | 75.64        | 70.70        |
| PV-RCNN <sup>[15]</sup>     | CVPR2020    | 90.25           | 81.43        | 76.82        |
| VoxelRCNN <sup>[7]</sup>    | AAAI2021    | <b>90.90</b>    | 81.62        | 77.06        |
| Votr-TSD <sup>[27]</sup>    | ICCV2021    | 87.83           | 81.77        | 77.16        |
| SVGA-Net <sup>[28]</sup>    | AAAI2022    | 87.33           | 80.47        | 75.91        |
| DSPF-RCNN <sup>[29]</sup>   | 中国激光 2022   | 89.90           | 81.04        | 76.54        |
| SSG-RCNN                    | -           | 88.35           | <b>82.11</b> | <b>77.64</b> |

并与 SECOND 网络检测结果进行对比。图 4 和 5 是可视化结果的两个示例,每个图片上栏是 SSG-RCNN 的检测结果,下栏是 SECOND 的检测结果,红色三维框(8 个角上带点)为真实标签,绿色三维框(8 个角上无点)为预测结果。对比图 4 中的检测结果,可以看到 SECON 没有检测出较远处的两个目标,而 SSG-RCNN 成功检测出目标,证明了 SSG-RCNN 对远距离目标的检测优势。此外,SSG-RCNN 可以利用对称点生成模块恢复三维目标结构信息,从而检测出被严重遮挡的目标。如图 5 所示,即使是 KITTI 数据集没有标注的严重遮挡目标,SSG-RCNN 也可以成功检测。



图 4 可视化结果实例 1

Fig. 4 Visualizations of detection example 1

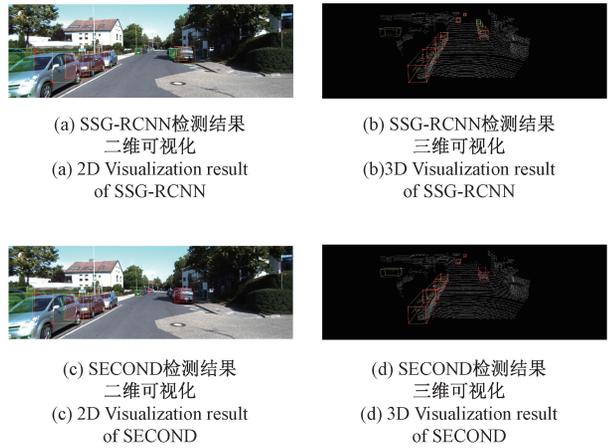


图 5 可视化结果实例 2

Fig. 5 Visualizations of detection example 2

### 2.5 消融实验

建立消融实验验证网络中各个模块的有效性,包括对称点生成、自注意力池化层。关于对称点生成的验证,直接在对比模型中去掉相关部分。关于自注意力池化层的验证,在对比模型中使用 PointNet 池化层代替。消融实验结果如表 6 所示。

表 6 KITTI 验证集上消融实验

Table 6 Ablation study on KITTI validation set

| SECOND | 模块           |         |       | Car-3D detection (IoU 0.7) |              |              |
|--------|--------------|---------|-------|----------------------------|--------------|--------------|
|        | PointNet 池化层 | 自注意力池化层 | 对称点生成 | 简单                         | 中等           | 困难           |
| ✓      |              |         |       | 90.22                      | 81.54        | 78.69        |
| ✓      | ✓            |         |       | 92.36                      | 84.57        | 82.36        |
| ✓      |              | ✓       |       | 92.37                      | 84.81        | 82.74        |
| ✓      | ✓            |         | ✓     | <b>93.02</b>               | 84.96        | 82.72        |
| ✓      |              | ✓       | ✓     | 92.84                      | <b>85.27</b> | <b>83.10</b> |

对比表 6 的第 2 行和第 4 行,以及对比表 6 的第 3 行和第 5 行,可以看到无论是使用 PointNet 池化层还是自注意力池化层,对称点生成模块都可以有效地提高网络的三维目标检测精度。具体而言,两组对比数据下,对称点生成模块在中等级别上带来的精度提升分别为 0.39% 和 0.46%。对称点生成模块利用交通场景下三维目标形状的镜像对称性,生成点云中前景点的对称点,有效恢复目标的整体形状信息,从而提高最终的检测精度。图 6 和 7 分别展示了原始点云和对称点生成点云的可视化结果。由两图对比可以看出,经过对称点生成后,点云中的三维目标结构得到了明显的增强,可以帮助网络实现更精准的三维目标检测。

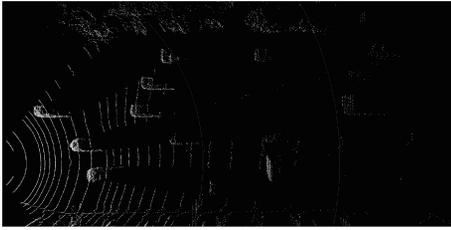


图6 原始点云

Fig. 6 Raw point cloud



图7 对称点生成后的点云

Fig. 7 Point cloud with symmetric points

对比表6的第4行和第5行,以及对比表6的第2行和第3行,相比常规的 PointNet 池化层,所设计的自注意力池化层在总体上能够有效地提升三维目标检测精度。具体而言,两组对比数据下,自注意力池化层在中等级别上带来的精度提升分别为0.24%和0.31%。自注意力池化层在聚合候选框特征时,利用 Transformer 结构对候选框附近的点建模全局语义信息,最后提取的候选框特征能够表征更丰富的目标整体信息,从而提高二阶段修正的精度。

2.6 检测速度对比

将 SSG-RCNN 网络的运行速度与其他的双阶段检测算法进行对比,所提网络的运行速度基于 3090 显卡进行计算,其他对比方法的检测速度取自论文公开结果。

结果如表7所示,所提方法的检测速度为 89 ms,频率为 11 Hz,可以满足激光雷达 10 Hz 的点云采集频率。相比于其他的双阶段检测网络,SSG-RCNN 的运行速度居于中位水平,快于部分双阶段检测网络。

表7 检测速度对比

Table 7 Detection speed comparison

| 方法    | Point RCNN | PVRCNN | STD | SVGA-Net | DSPF-RCNN | SSG-RCNN |
|-------|------------|--------|-----|----------|-----------|----------|
| 耗时/ms | 100        | 91     | 80  | 62       | 64        | 89       |

2.7 算法在本地数据集验证

为了进一步验证所提算法的工程实用性,在园区路段采集本地三维点云数据进行算法的训练和验证。如图8和9所示,在园区内搭建路侧激光雷达设备采集三维点云数据。



图8 实验路段示意图

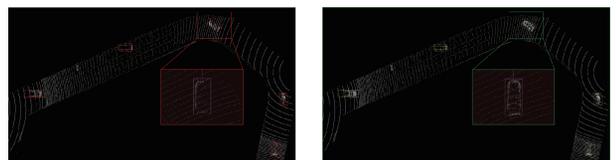
Fig. 8 Experimental road section



图9 路侧设备

Fig. 9 Road side unit

所提算法在本地数据集上的实验结果实例如图10和11所示,每张图片左侧为原始点云和真值三维框,右侧为对称点生成后的点云和预测三维框。由图10和11可知,面对本地实验数据,所提算法可以正确地预测三维目标框,且所提的对称形状生成策略可以有效地增强前景点数并恢复三维目标的形状信息,证明所提算法具有较好的工程实用性。



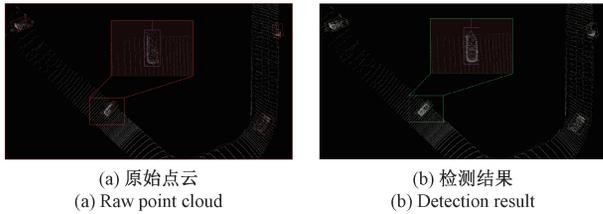
(a) 原始点云  
(a) Raw point cloud

(b) 检测结果  
(b) Detection result

图10 本地数据上的检测结果示例1

Fig. 10 Detection result of SSG-RCNN on local data.

Example 1



(a) 原始点云  
(a) Raw point cloud

(b) 检测结果  
(b) Detection result

图 11 本地数据上的检测结果示例 2

Fig. 11 Detection result of SSG-RCNN on local data.

Example 2

### 3 结 论

本文提出了基于对称形状生成的三维目标检测网络,SSG-RCNN。网络一阶段通过预测前景点的镜像对称点恢复三维目标的对称形状,为后续的三维框预测提供更完整的目标结构信息。网络二阶段引入自注意力机制从原始点和对称点中聚合候选框特征,建模候选框的全局空间信息,增加三维目标检测精度。在KITTI数据集上对所提网络进行评估,并与其他主流方法对比,实验表明SSG-RCNN取得了卓越的检测性能。特别是对于Car类别的困难级别,无论是验证集还是测试集,SSG-RCNN在所提方法中均取得了最高的检测精度,证明了其先进的检测性能。所提网络的推理时间为89 ms,可以满足激光雷达的点云采集速率,具有实际应用的价值。

### 参考文献

[ 1 ] 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.

[ 2 ] 郭保青,余祖俊,张楠,等. 铁路场景三维点云分割与分类识别算法[J]. 仪器仪表学报, 2017, 38(9): 2103-2111.  
GUO B Q, YU Z J, ZHANG N, et al. 3D point cloud segmentation, classification and recognition algorithm of railway scene [J]. Chinese Journal of Scientific Instrument, 2017, 38(9): 2103-2111.

[ 3 ] 陈丽,陈洋,杨艳华. 面向三维结构视觉检测的无人机覆盖路径规划[J]. 电子测量与仪器学报, 2023, 37(2): 1-10.

CHEN L, CHEN Y, YANG Y H. UAV coverage path planning for 3D structure visual inspection[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(2): 1-10.

[ 4 ] ZHOU Y, 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, 2018: 4490-4499.

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

[ 6 ] LANG A H, VORA S, CAESAR H, et al. Pointpillars: Fast encoders for object detection from point clouds[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 12697-12705.

[ 7 ] DENG J J, SHI SH SH, LI P W, et al. Voxel R-CNN: Towards high performance voxel-based 3d object detection[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(2): 1201-1209.

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

[ 9 ] CHEN Y L, LIU SH, SHEN X Y, et al. Fast point R-CNN[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9775-9784.

[ 10 ] YANG Z T, SUN Y N, LIU SH, et al. 3Dssd: Point-based 3D single stage object detector[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11040-11048.

[ 11 ] YANG Z T, SUN Y N, LIU SH, et al. Std: Sparse-to-dense 3d object detector for point cloud[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 1951-1960.

[ 12 ] 冯明驰,高小倩,汪静姝,等. 基于立体视觉与激光雷达的车辆目标外形位置融合算法研究[J]. 仪器仪表学报, 2021, 42(10): 210-220.  
FENG M CH, GAO X Q, WANG J SH, et al. Research on the fusion algorithm of vehicle object shape-position based on stereo vision and lidar[J]. Chinese Journal of Scientific Instrument, 2021, 42(10): 210-220.

[ 13 ] QI C R, SU H, MO K CH, et al. Pointnet: Deep

- learning on point sets for 3d classification and segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 652-660.
- [14] QI C R, YI L, SU H, et al. Pointnet ++: Deep hierarchical feature learning on point sets in a metric space[J]. Advances in Neural Information Processing Systems, 2017, 30.
- [15] SHI SH SH, GUO CH X, JIANG L, et al. Pv-RCNN: Point-voxel feature set abstraction for 3d object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10529-10538.
- [16] HE CH H, ZENG H, HUANG J Q, et al. Structure aware single-stage 3D object detection from point cloud[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11873-11882.
- [17] LI J L, SUN Y, LUO SH J, et al. P2V-RCNN: Point to voxel feature learning for 3D object detection from point clouds[J]. IEEE Access, 2021, 9: 98249-98260.
- [18] NOH J, LEE S, HAM B. Hvr: Hybrid voxel-point representation for single-stage 3D object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 14605-14614.
- [19] LI ZH CH, WANG F, WANG N Y. LiDAR R-CNN: An efficient and universal 3D object detector [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 7546-7555.
- [20] LI Z Y, YAO Y C, QUAN ZH B, et al. Sienet: Spatial information enhancement network for 3D object detection from point cloud [J]. Arxiv Preprint, 2021, ArXiv: 2103.15396.
- [21] XU Q G, ZHOU Y, WANG W Y, et al. Spg: Unsupervised domain adaptation for 3D object detection via semantic point generation [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 15446-15456.
- [22] 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, IEEE, 2012: 3354-3361.
- [23] GRAHAM B, MAATEN L V D. Submanifold sparse convolutional networks [J]. Arxiv Preprint, 2017, Arxiv: 1706.01307.
- [24] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [25] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. ArXiv Preprint, 2017, ArXiv:1706.03762.
- [26] TEAM O D. OpenPCDet: An open-source toolbox for 3D object detection from point clouds[Z]. Github, 2020.
- [27] MAO J G, XUE Y J, NIU M ZH, et al. Voxel transformer for 3D object detection [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 3164-3173.
- [28] HE Q D, WANG ZH N, ZENG H, et al. Svga-Net: Sparse voxel-graph attention network for 3D object detection from point clouds [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022.
- [29] 胡杰,安永鹏,徐文才,等. 基于激光点云的深度语义和位置信息融合的三维目标检测 [J]. 中国激光, 2023, 50(10): 200-210.
- HU J, AN Y P, XU W C, et al. 3D object detection based on deep semantics and position information fusion of laser point cloud [J]. Chinese Journal of Lasers, 2023, 50(10): 200-210.

### 作者简介



**涂新奎**, 2020 年于华南理工大学获得学士学位, 现为华南理工大学硕士研究生, 主要研究方向为智能汽车环境感知。

E-mail: 763149484@qq.com

**Tu Xinkui** received his B.Sc. degree from South China University of Technology in 2020. He is currently a master student at South China University of Technology. His main research interests include intelligent driving and environment perception.



**郑少武**, 2018 年于华南理工大学获得学士学位, 现为华南理工大学博士研究生, 主要研究方向为智能驾驶环境感知、多传感器融合。

E-mail: mezhengsw@mail.scut.edu.cn

**Zheng Shaowu** received his B.Sc. degree from South China University of Technology in 2018. He is currently a Ph.D.

candidate at South China University of Technology. His main research interests include intelligent driving, environment perception and multi-sensor fusion.



于善虎,1988 年于华南理工大学获得学士学位,现为广州华工机动车检测技术有限公司副总经理及高级工程师,主要研究方向为智能网联汽车、机动车检测。

E-mail: yushanhu@163.com

**Yu Shanhu** received his B.Sc. degree from South China University of Technology in 1988. He is currently the Deputy General Manager and Senior Engineer of Guangzhou Huagong Motor Vehicle Inspection Technology Co., Ltd. His main research interests include intelligent connected vehicle and vehicle inspection.



李巍华(通信作者),分别在 1995 年和 1998 年于太原理工大学获得学士学位和硕士学位,2003 年于华中科技大学获得博士学位,现为华南理工大学教授,主要研究方向为智能网联汽车的环境感知及路径规划、装备智能运维、数字孪生。

E-mail: whlee@scut.edu.cn

**Li Weihua** (Corresponding author) received his B.Sc. and M.Sc. degrees both from Taiyuan University of Technology in 1995 and 1998, respectively, and received his Ph.D. degree from Huazhong University of Science and Technology in 2003. He is currently a professor at South China University of Technology. His main research interests include environment perception & path planning for intelligent connected vehicles, intelligent maintenance and health management, and digital twins.