

# 基于稀疏 Transformer 的雷达点云三维目标检测

韩磊<sup>1</sup>, 高永彬<sup>1</sup>, 史志才<sup>1,2</sup>

(1. 上海工程技术大学 电子电气工程学院, 上海 201600; 2. 上海市信息安全综合管理技术研究重点实验室, 上海 200240)

**摘要:** 随着计算机视觉技术的发展, 基于点云的三维目标检测算法被广泛应用于自动驾驶、机器人控制等领域。针对点云稀疏条件下基于点云三维目标检测算法鲁棒性较差、检测精度低的问题, 提出基于稀疏 Transformer 的三维目标检测算法。在注意力矩阵生成阶段, 通过稀疏 Transformer 模块显式选择 Top- $t$  个权重元素, 以保留有利于特征提取的权重元素, 在降低环境噪声对鲁棒性影响的同时加快 Transformer 模块的运行速度。在回归阶段, 将基于空间特征粗回归模块生成的边界框作为检测头模块的初始锚框, 用于后续边界框的精细回归操作。设计基于体素的三维目标检测算法的损失函数, 以精确地衡量类别损失、位置回归损失和方向损失。在 KITTI 数据集上的实验结果表明, 相比 PointPillars 算法, 该算法的平均精度均值提高 3.46%, 能有效提高点云三维目标的检测精度且具有较优的鲁棒性。相比原始 Transformer 模块, 所提稀疏 Transformer 模块在点云图像上的平均运行速度加快了约 0.54 frame/s。

**关键词:** 机器视觉; 三维目标检测; 稀疏 Transformer; 粗回归; 损失函数

开放科学(资源服务)标志码(OSID):



**中文引用格式:** 韩磊, 高永彬, 史志才. 基于稀疏 Transformer 的雷达点云三维目标检测[J]. 计算机工程, 2022, 48(11): 104-110, 144.

**英文引用格式:** HAN L, GAO Y B, SHI Z C. Three-dimensional object detection of radar point cloud based on sparse Transformer[J]. Computer Engineering, 2022, 48(11): 104-110, 144.

## Three-Dimensional Object Detection of Radar Point Cloud Based on Sparse Transformer

HAN Lei<sup>1</sup>, GAO Yongbin<sup>1</sup>, SHI Zhicai<sup>1,2</sup>

(1. School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201600, China;

2. Shanghai Key Laboratory of Integrated Administration Technologies for Information Security, Shanghai 200240, China)

**[Abstract]** With the development of computer vision technology, three-dimensional object detection algorithms based on point cloud are widely used in automatic driving, robot control and other application scenarios. Aiming at the problems of poor robustness and low detection accuracy of three-dimensional object detection algorithms based on point cloud, under the condition of sparse point cloud, this study proposes three-dimensional object detection algorithm based on the sparse Transformer. In the attention matrix generation stage, a sparse Transformer module is used to select the Top- $t$  weight elements explicitly to retain the most favorable weight elements for feature extraction, reduce the impact of environmental noise on robustness, and accelerate the running speed of the Transformer module. In the regression stage, the bounding box generated by the coarse regression module based on spatial features is used as the initial anchor frame of the detection-head module for the subsequent fine regression operation of the bounding box. A loss function based on voxel for 3D object detection is proposed to accurately measure category, position regression, and direction losses. The experimental results on the KITTI dataset show that compared with the PointPillars algorithm, the average accuracy of the proposed algorithm is improved by 3.46%. It can effectively improve the detection accuracy of point cloud three-dimensional targets and has better robustness. Compared with the running speed of the original Transformer module, the average running speed of the proposed sparse Transformer module on point cloud image is improved by about 0.54 frame/s.

**[Key words]** machine vision; three-dimensional object detection; sparse Transformer; coarse regression; loss function

DOI: 10.19678/j.issn.1000-3428.0062440

**基金项目:** 上海市信息安全综合管理技术研究重点实验室开放项目(AGK2019004)。

**作者简介:** 韩磊(1996—), 男, 硕士研究生, 主研方向为图像与点云目标检测; 高永彬(通信作者), 副教授; 史志才, 教授。

**收稿日期:** 2021-08-23 **修回日期:** 2021-12-10 **E-mail:** hl\_sues@163.com

## 0 概述

三维目标检测广泛应用于自动驾驶<sup>[1]</sup>、增强现实<sup>[2]</sup>和机器人控制<sup>[3-4]</sup>领域中。三维目标检测算法根据输入形式的不同,分为基于图像、基于多传感器融合和基于点云的三维目标检测算法。

基于图像的三维目标检测算法根据输入RGB图像中2D/3D约束、关键点和形状,通过推理目标几何关系解决图像深度信息缺失的问题。文献[5]利用单支路网络检测三维框的多个角点以重构三维中心点,通过二分支关键点检测网络锐化目标辨识能力。文献[6]考虑到2D投影中的几何推理和未观察到深度信息的维度,通过单目RGB图像预测3D对象定位。文献[7]通过视觉深度估计方法从图像中估计像素深度,并将得到的像素深度反投影为3D点云,利用基于雷达的检测方法进行检测。文献[8]基于左右目视图的潜在关键点构建左右视图关键点一致性损失函数,以提高选取潜在关键点的位置精度,从而提高车辆的检测准确性。

多传感器融合通常将多个传感器获取的特征进行融合。文献[9]提出在两个连续的步骤中检测目标,基于摄像机图像生成区域建议,通过处理感兴趣区域中的激光雷达点以检测目标。文献[10]提出引导式图像融合模块,以基于点的方式建立原始点云数据与相机图像之间的对应关系,并自适应地估计图像语义特征的重要性。这种方式根据高分辨率的图像特征来增强点特征,同时抑制干扰图像的特征。文献[11]结合点云的深度信息与毫米波雷达输出确定目标的优势,采用量纲归一化方法对点云进行预处理,并利用处理后的点云生成特征图。文献[12]基于二维候选区域中的像素过滤激光点云,生成视锥点云,以加快检测速度。

基于点云的检测算法仅通过输入点云学习特征,

在检测网络中回归目标类别和包围框<sup>[13]</sup>。文献[14]将点云编码为体素,采用堆叠体素特征编码层来提取体素特征。文献[15]通过立柱特征网络将点云处理成伪图像,并消除耗时的3D卷积运算,使得检测速度显著提升。文献[16]将点云编码到一个固定半径的近邻图中,并设计图神经网络,以预测类别和图中每个顶点所属的对象形状。文献[17]利用三维区域生成网络,将多视图生成器模块生成的多角度点云伪图像重新校准与融合,并根据提取的语义特征进行最终的点云目标分类和最优外接包围框回归。

基于图像的三维目标检测算法无法提供可靠的三维几何信息;基于多传感器融合的三维目标检测算法输入数据较多,需要较高的算力和较复杂的特征处理与融合算法;点云数据通常极其稀疏,受噪点影响比较大,基于点云的三维目标检测算法鲁棒性较差。因此,在点云稀疏条件下提升检测精度和算法鲁棒性具有一定必要性。

本文提出基于稀疏Transformer的雷达点云三维目标检测算法。构建稀疏Transformer模块并将其应用于三维目标检测领域中,通过显式选择Top- $t$ 个权重元素,以排除对注意力干扰性较高的权重元素,从而提高检测精度。设计一种粗回归模块,将粗回归模块生成的边界框作为检测头模块的初始锚框,使检测结果生成的边界框更加精细。在此基础上,设计基于体素三维目标检测算法的损失函数,以优化检测结果。

## 1 网络模型

本文基于PointPillars在点云特征处理阶段的良好性能,延用了点云特征处理模块和2D卷积模块,并增加了稀疏Transformer模块和粗回归模块,在回归阶段使用通用的SSD<sup>[18]</sup>检测头作为检测模块。本文网络结构如图1所示。

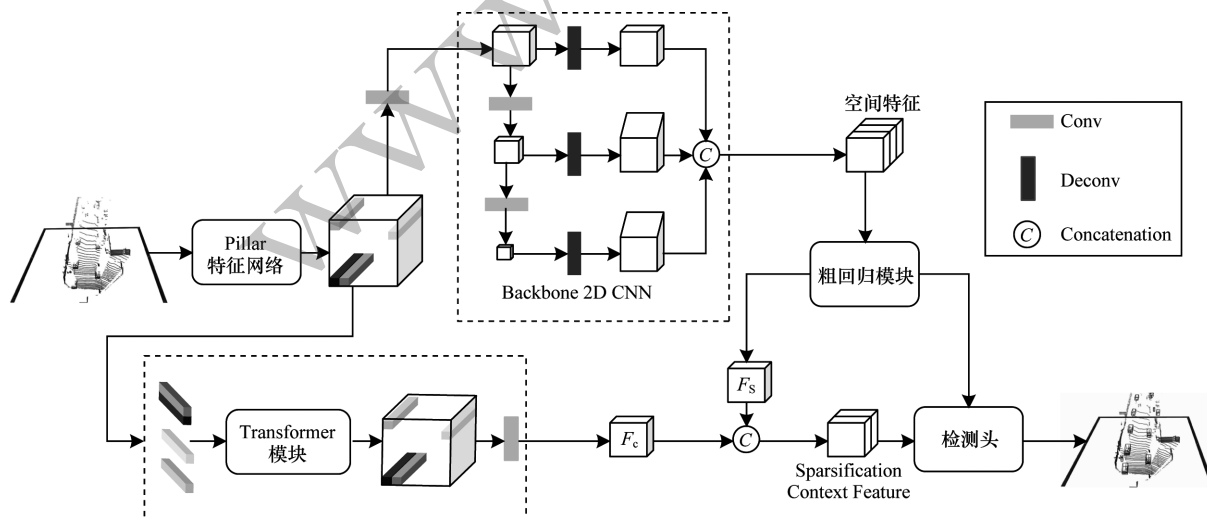


图1 本文网络结构

Fig.1 Structure of the proposed network

首先,将一帧点云图像输入到点云特征处理模块中,由该模块将点云图像划分为 $H \times W$ 的立柱,并对立柱中的点进行采样;然后,经过Pillar特征网络进行特征学习和特征展开,输出尺寸为 $(C, H, W)$ 的伪图像,将该伪图像分别送入到2D卷积模块,经过卷积操作分别产生尺寸为 $(C, H/2, W/2)$ 、 $(2C, H/4, W/4)$ 和 $(4C, H/8, W/8)$ 的特征,通过反卷积操作生成3个尺寸为 $(2C, H/2, W/2)$ 的特征后,再将这3个特征相连接,输出尺寸为 $(6C, H/2, W/2)$ 的空间特征;最后,将空间特征送入到粗回归模块,在该模块的区域建议网络(Region Proposal Network, RPN)回归粗略类别和坐标的同时,另一个分支经过卷积操作输出尺寸为 $(2C, H/2, W/2)$ 的新空间特征。与此同时,本文将Pillar特征网络中伪图像的特征展开为 $(H \times W) \times C$ 的序列形式,输入到稀疏Transformer模块,并根据原始位置对该模块输出的序列特征嵌入重新组合成伪图像特征,进行一次卷积操作,输出尺寸为 $(2C, H/2, W/2)$ 的稀疏上下文特征。本文将得到的新空间特征与稀疏上下文特征连接后输入到检测模块,在粗回归模块提供的粗略锚框坐标的辅助下更精确地回归目标物体的坐标。

### 1.1 稀疏Transformer模块

基于自注意力的Transformer<sup>[19]</sup>在一些自然语言处理和二维目标检测任务中具有较优的性能。自注意力能够模拟长期的依赖关系,但易受上下文中无关信息的影响。为解决该问题,本文引入稀疏Transformer模块<sup>[20]</sup>。稀疏Transformer模块通过显式选择最相关的片段来提高对全局上下文的关注,增强模型的鲁棒性。

稀疏Transformer模块是基于Transformer架构,通过Top- $t$ 选择将注意力退化为稀疏注意力,有助于保留注意力的成分,而去除其他无关的信息。本文提出的稀疏Transformer模块中注意力可以集中在最有贡献的元素上。这种选择方法在保留重要信息和去除噪声方面具有有效性。稀疏Transformer模块结构如图2所示。

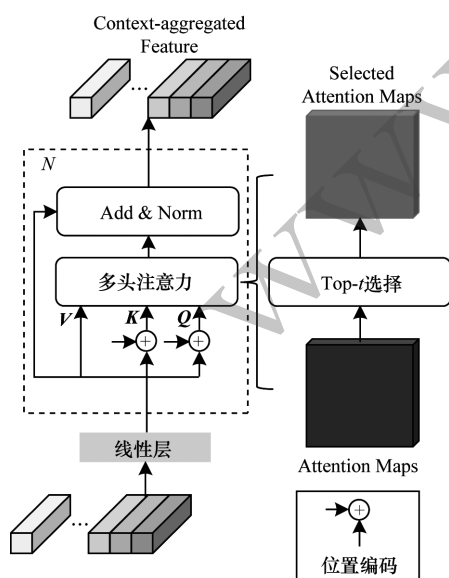


图2 稀疏Transformer模块结构

Fig.2 Structure of sparse Transformer module

对于单头自注意力,将点柱特征 $x_i$ 经线性层变换后变为值向量 $V[l_v, d]$ 、关键向量 $K[l_k, d]$ 和查询向量 $Q[l_q, d]$ 。线性变换过程如式(1)所示:

$$P = W_p x, P \in (Q, K, V) \quad (1)$$

其中: $W_p$ 为对应向量的线性变换矩阵;查询向量 $Q$ 与关键向量 $K$ 的相似性通过点乘计算。注意力得分计算如式(2)所示:

$$S = \frac{QK^T}{\sqrt{d}} \quad (2)$$

由注意力机制可知,注意力得分 $S$ 越高,特征相关性越强。因此,本文在 $S$ 上实现稀疏注意力操作,以便选择注意力矩阵中每行的前 $t$ 个有贡献的元素。本文选择 $S$ 中每行的 $t$ 个最大元素,并记录它们在矩阵中的位置 $(i, j)$ ,其中 $t$ 是一个超参数。假设第 $i$ 行的第 $t$ 个最大值是 $t_i$ ,如果第 $j$ 个分量的值大于 $t_i$ ,则记录位置 $(i, j)$ ,连接每行的阈值以形成向量 $t = [t_1, t_2, \dots, t_n]$ , $n$ 为查询向量的长度。稀疏注意力 $S_{sa}(\cdot, \cdot)$ 函数如式(3)所示:

$$S_{sa}(S, t) = \begin{cases} S_{ij}, & S_{ij} \geq t_i \\ -\infty, & S_{ij} < t_i \end{cases} \quad (3)$$

稀疏注意力模块的输出计算过程如式(4)所示:

$$A(Q, K, V) = \text{softmax}(S_{sa}(S, t))V \quad (4)$$

本文使用多头注意力机制将特征映射到不同的特征空间,以学习不同于空间的相关特征。不同的注意力头可以独立地进行特征学习,互不干扰。最后,将每个头部的结果拼接再进行一次线性变换得到的值作为多头注意力的结果。将结果与点柱特征 $x_i$ 进行残差连接,再用层归一化<sup>[21]</sup>对其进行归一化操作,层归一化操作后得到的结果即为所求。

### 1.2 粗回归模块

本文使用一个粗回归模块,该模块有两个分支,卷积分支用于调整特征尺度,RPN分支用于粗略回归目标类别和边界框,回归结果用于指导后续检测头进行精细回归操作。粗回归模块结构如图3所示。

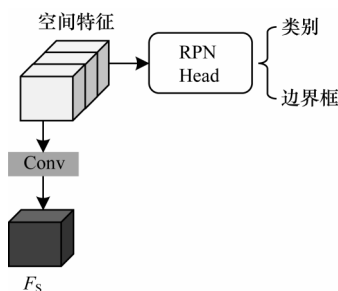


图3 粗回归模块结构

Fig.3 Structure of coarse regression module

从图3可以看出,空间特征是骨干网络进行多尺度特征串联后得到的特征,特征大小为 $(6C, H/2, W/2)$ ,该骨干网络与PointPillars算法中的骨干网络



相同。在卷积分支中主要进行 $1 \times 1$ 的卷积操作,将骨干网络输出的空间特征降维,降维后的特征大小为 $(2C, H/2, W/2)$ ,将该特征与Transformer模块生成的全局上下文特征串联。与此同时,将空间特征送入RPN分支,输出类别和边界框。RPN分支的回归结果为检测头模块提供粗略的锚框,用于后续边界框的精确回归操作。

### 1.3 损失函数

本文在SECOND<sup>[22]</sup>损失函数的基础上提出一种新的损失函数,以更好地优化粗回归和检测头模块。真值框和锚框由 $(x, y, z, w, l, h, \theta)$ 表示,其中 $(x, y, z)$ 表示框的中心点坐标, $(w, h, l)$ 表示框的宽、高、长, $\theta$ 表示框的方向角。边界框的偏移由真值框和锚框计算,如式(5)所示:

$$\begin{cases} \Delta x = \frac{x_{gt} - x_a}{d_a}, \Delta y = \frac{y_{gt} - y_a}{d_a}, \Delta z = \frac{z_{gt} - z_a}{h_a} \\ \Delta w = \log_a \frac{w_{gt}}{w_a}, \Delta l = \log_a \frac{l_{gt}}{l_a}, \Delta h = \log_a \frac{h_{gt}}{h_a} \\ \Delta \theta = \theta_{gt} - \theta_a, d_a = \sqrt{w_a^2 + l_a^2} \end{cases} \quad (5)$$

其中:gt表示真值框;a表示锚框。位置回归损失函数如式(6)所示:

$$L_{loc} = \sum_{p \in (x, y, z, w, l, h)} \text{SmoothL1}(\Delta_p^{pre} - \Delta_p^{gt}) + \text{SmoothL1}(\sin(\Delta_\theta^{pre} - \Delta_\theta^{gt})) \quad (6)$$

其中:pre表示预测值。对于角度回归,这种减少角度损失的方法解决了 $0$ 和 $\pi$ 方向的冲突问题。为解决该损失函数将方向相反的边界框视为相同的问题,本文在离散方向上使用交叉熵损失函数,使网络能够区分目标的正反方向。方向分类损失函数定义为 $L_{dir}$ 。本文使用Focal Loss定义物体分类损失,如式(7)所示:

$$L_{cls} = -\alpha(1 - p^a)^\gamma \log_a p^a \quad (7)$$

其中: $p^a$ 表示模型预测的锚框类别概率; $\alpha$ 和 $\gamma$ 表示Focal Loss的参数。

该检测网络总的损失函数如式(8)所示:

$$L_{total} = \beta_{cls} L_{cls}^C + \frac{\beta_{loc}}{N_{pos}^C} L_{loc}^C + \beta_{dir} L_{dir}^C + \lambda \left( \beta_{cls} L_{cls}^D + \frac{\beta_{loc}}{N_{pos}^D} L_{loc}^D + \beta_{dir} L_{dir}^D \right) \quad (8)$$

其中:上标C和D分别表示粗回归模块和检测头模块; $N_{pos}^C$ 表示粗回归框的正锚框数目; $N_{pos}^D$ 表示细回归框的正锚框数目; $\beta_{cls}$ 、 $\beta_{loc}$ 和 $\beta_{dir}$ 表示用于平衡类别损失、位置回归损失和方向损失的权重参数; $\lambda$ 表示用于平衡粗回归模块和检测头模块的权重。

## 2 实验结果与分析

### 2.1 实验数据集

本文在KITTI数据集上进行实验,该自动驾驶数据集是目前在三维目标检测和分割领域中使用最广泛

的数据集。该数据集包含7 481个训练样本,本文按大约1:1的比例将训练样本分为训练集和测试集,其中训练集包含3 712个样本数据,测试集包含3 769个样本数据。本文在测试集上对模型训练的汽车、行人和骑行者这3个类别进行评估。对于每个类别,本文根据3D对象的大小和遮挡程度分为简单、中等、困难3个级别。平均精度均值(mean Average Precision, mAP)作为实验结果的评估度量。本文采用官方评估建议,将汽车的交并比(Intersection Over Union, IOU)阈值设置为0.7,将行人和骑行者的IOU阈值设置为0.5。

### 2.2 实验环境与对比实验

本文实验的模型训练部分选用的设备信息:一台运行系统为Ubuntu18.04、显卡为NVIDIA RTX 8000的服务器,算法由python3.7和pytorch1.4框架实现,使用Adam优化器训练100轮,批尺寸设置为6,学习率设置为0.003。

不同算法的三维检测结果对比如图4所示,检测的阈值均设置为0.5。在场景1中,PointPillars存在不同程度的误检,将环境中的噪点或者路灯杆检测为行人或骑行者。在场景2中,PointPillars仍存在不同程度的误检和漏检,将道闸的立柱检测为行人,把并排行走或靠近的两个人检测为一个人。

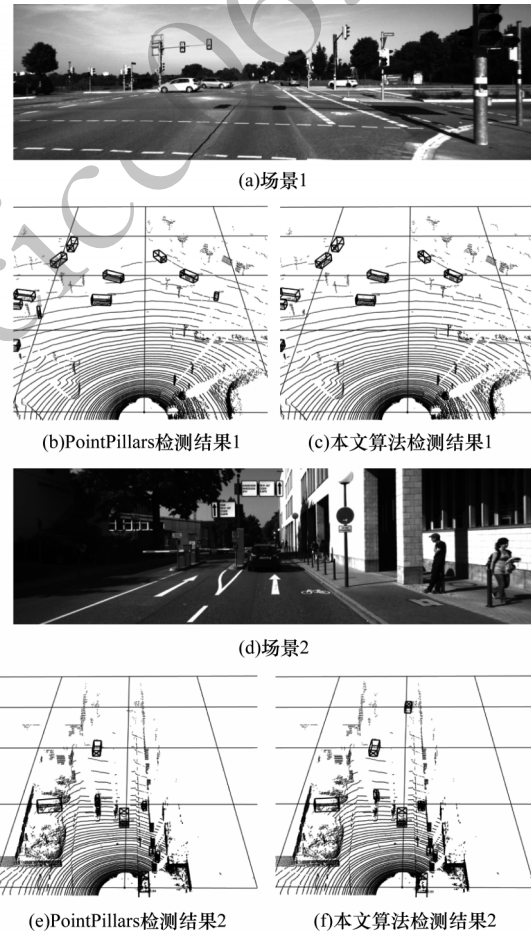


图4 不同算法的三维检测结果对比

Fig.4 Three-dimensional detection results comparison among different algorithms

鸟瞰视角下不同算法的检测结果对比如图 5 所示,图中边为白色的矩形框说明预测的边界框与实际真值框未完全重合。

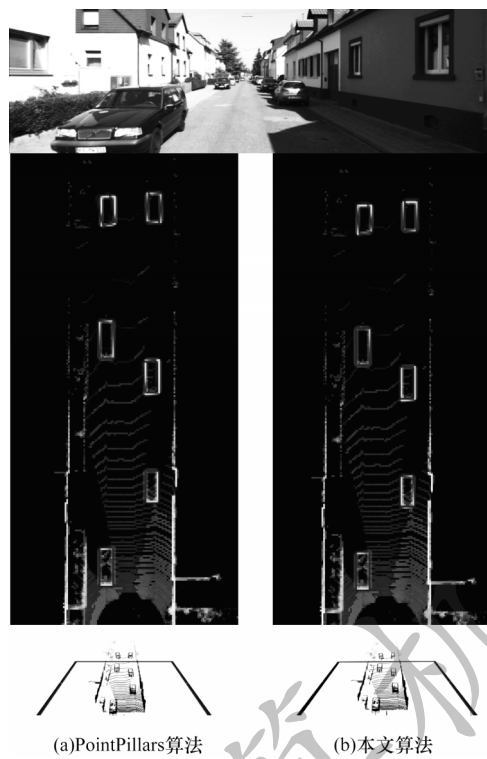


图 5 鸟瞰视角下不同算法的检测结果对比  
Fig.5 Detection results comparison among different algorithms from aerial view

本文选择 VoxelNet、SECOND、PointPillars、3D-GIoU<sup>[23]</sup>、Part-A<sup>2</sup><sup>[24]</sup>、PointRCNN<sup>[25]</sup>、Point-GNN 和 TANet<sup>[26]</sup>作为对比算法。表 1、表 2 和表 3 分别表示在 KITTI 测试集上汽车、行人和骑行者类别下本文算法与其他算法的 mAP 对比。3D mAP 是 3 种难度类别的平均精度均值。从表 1~表 3 可以看出,当检测行人和骑行者类别时,本文算法相较于其他算法具有较优的平均精度均值。

表 1 在汽车类别下不同算法的 mAP 对比				
Table 1 mAP comparison among different algorithms under car category %				
算法	mAP			3D mAP
	简单	中等	困难	
VoxelNet 算法	87.93	75.37	73.21	78.84
SECOND 算法	88.61	78.62	77.22	81.48
PointPillars 算法	87.50	77.01	74.77	79.76
3D-GIoU 算法	87.83	77.91	75.55	80.43
Part-A <sup>2</sup> 算法	89.56	79.41	78.84	82.60
PointRCNN 算法	89.01	78.77	78.10	81.96
Point-GNN 算法	89.33	79.47	78.29	82.36
TANet 算法	88.17	77.75	75.31	80.41
本文算法	88.53	79.58	78.15	82.09

表 2 在行人类别下不同算法的 mAP 对比				
Table 2 mAP comparison among different algorithms under pedestrian category %				
算法	mAP			3D mAP
	简单	中等	困难	
VoxelNet 算法	67.81	63.52	58.87	63.40
SECOND 算法	56.00	50.02	43.64	49.89
PointPillars 算法	66.73	61.06	56.50	61.43
3D-GIoU 算法	67.23	59.58	52.69	59.83
Part-A <sup>2</sup> 算法	65.69	60.05	55.45	60.40
PointRCNN 算法	62.69	55.36	51.60	56.55
Point-GNN 算法	61.92	53.77	50.14	55.28
TANet 算法	71.04	64.20	59.11	64.78
本文算法	71.27	64.70	59.26	65.08

表 3 在骑行者类别下不同算法的 mAP 对比				
Table 3 mAP comparison among different algorithms under cyclist category %				
算法	mAP			3D mAP
	简单	中等	困难	
VoxelNet 算法	77.69	58.72	51.63	62.68
SECOND 算法	80.97	63.43	56.67	67.02
PointPillars 算法	83.65	63.40	59.71	68.92
3D-GIoU 算法	83.32	64.69	63.51	70.51
Part-A <sup>2</sup> 算法	85.50	68.90	64.53	72.98
PointRCNN 算法	84.48	65.37	59.83	69.89
Point-GNN 算法	86.60	67.48	62.58	72.22
TANet 算法	85.21	65.29	61.57	70.69
本文算法	85.62	67.92	66.46	73.33

本文算法与现有执行速度表现优异算法的推理速度对比如表 4 所示。从表 4 可以看出,本文算法在提高平均精度均值的同时,推理速度平均加快了 0.535 8 frame/s。

表 4 不同算法的推理速度对比	
Table 4 Inference speed comparison among different algorithms	
算法	速度/(frame·s <sup>-1</sup> )
SECOND 算法	0.118 4
PointPillars 算法	0.062 3
PointRCNN 算法	0.178 6
本文算法(没有稀疏 Transformer 模块)	0.723 8
本文算法	0.188 0

2.3 消融实验

2.3.1 t 值的选择

由于注意力矩阵  $A$  与查询向量  $Q$ 、关键向量  $K$  有关,因此  $t$  值的大小与系数  $k$  和关键向量  $K$  的长度相关。本文 1.2 节中  $t=k \times l_k$ ,  $l_k$  是关键向量  $K$  的长度,也是注意力矩阵的列数。本文选取  $k=\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$  进行实验。实验设备

的显卡为 RTX 2080,批尺寸设置为 1。在 KITTI 测试集上汽车类别下本文算法的实验结果如图 6 所

示,以系数  $k$  为横坐标表示选择不同的  $t$  对本文算法检测精度的影响。

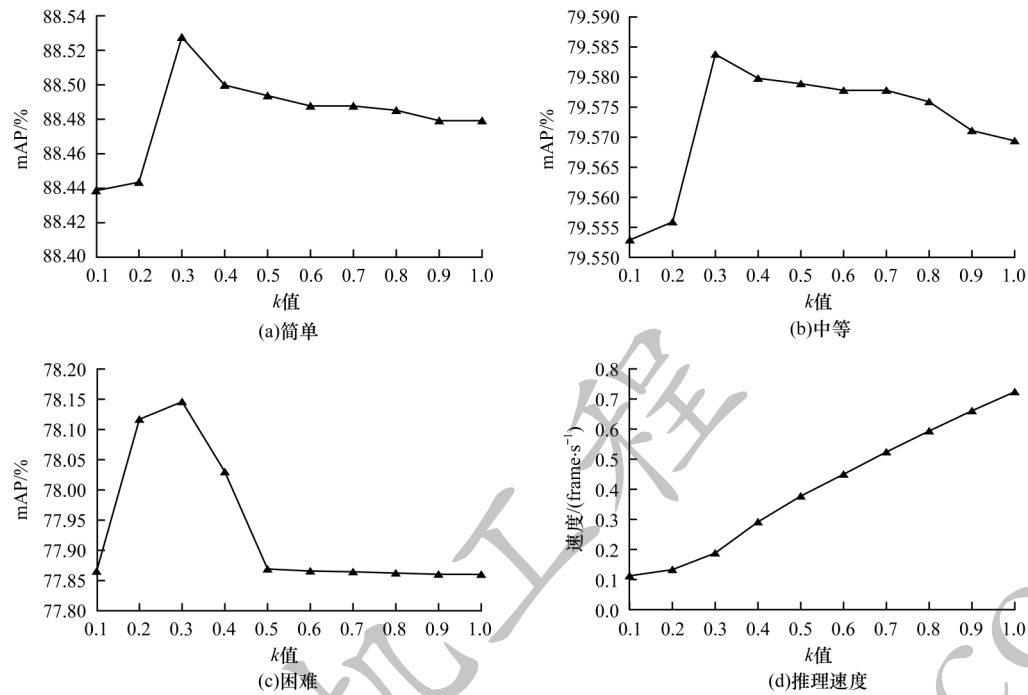


图 6 在汽车类别下本文算法的实验结果

Fig.6 Experimental results of the proposed algorithm under the car category

从图 6 可以看出,随着  $k$  值的增加,mAP 逐渐升高,当  $k=0.3$  时,mAP 达到最高,然后开始大幅度降低。其原因为本文的稀疏注意力模块对特征学习和去除噪点是有效的。在  $k=0.3$  之前,原始的注意力同样会注意到除目标以外的无关成分,对检测难度为中等和困难的目标产生的影响较大。在  $k=0.3$  之后,由于过多地过滤了目标的有用特征,因此检测精度明显下降。虽然稀疏注意力模块对检测精度的提升幅度比较微小,但是对推理速度的提升却十分显著。从图 6(d)可以看出,相较于原始的 Transformer,当  $k=0.3$  时,本文算法的平均推理速度加快了约 0.54 frame/s。

2.3.2 稀疏 Transformer 模块的作用

由于在 KITTI 数据集中噪点数目无法得知,因此本文在每个目标物的真值框内添加相同数量的噪点,模拟实际场景中噪点对模型的负影响,以测试该模块对模型鲁棒性和检测精度的贡献。基于 2.3.1 节的实验结果,本文将  $k$  值设置为 0.3,在不改变其他模块的情况下,在 KITTI 测试集上汽车类别下本文算法(稀疏 Transformer 模块)与 PointPillars 算法(普通 Transformer 模块)的 mAP 对比如表 5 所示。手动在每个目标物真值框内随机添加 100 个噪点,本文算法的 mAP 仅下降 1.60%,优于 PointPillars

算法。

表 5 在不同噪点数量下 PointPillars 和本文算法的 mAP 对比

Table 5 mAP comparison among PointPillars and the proposed algorithms with different number of noises %					
算法	噪点数量	mAP			3D mAP
		简单	中等	困难	
PointPillars 算法	0	87.50	77.01	74.77	79.76
	50	87.21	76.74	74.54	79.50
	100	86.62	76.06	68.91	77.20
本文算法	0	88.53	79.58	78.15	82.09
	50	88.21	79.12	77.23	81.52
	100	87.15	78.47	75.86	80.49

2.3.3 粗回归模块对结果的影响

在不改动其余模块的情况下,本文去掉粗回归模块的回归分支,以验证粗回归模块的有效性。本文算法 1 不包含粗回归模块,本文算法 2 包含粗回归模块。本文算法在 KITTI 测试集上汽车、行人和骑行者类别下的实验结果分别如表 6、表 7 和表 8 所示。从表中可以看出,相比不包含粗回归模块算法的测试结果,在不同检测难度下有粗回归模块算法的 mAP 分别提升了 0.61、1.01 和 0.95 个百分点。因此,包含粗回归模块的算法能够更精确地回归目标物体的坐标。



表6 在汽车类别下粗回归模块对检测精度的影响

Table 6 Influence of coarse regression module on detection accuracy under car category %

算法	mAP			3D mAP
	简单	中等	困难	
本文算法1	88.24	78.97	77.23	81.48
本文算法2	88.53	79.58	78.15	82.09

表7 在行人类别下粗回归模块对检测精度的影响

Table 7 Influence of coarse regression module on detection accuracy under pedestrian category %

算法	mAP			3D mAP
	简单	中等	困难	
本文算法1	70.88	63.41	57.92	64.07
本文算法2	71.27	64.70	59.26	65.08

表8 在骑行者类别下粗回归模块对检测精度的影响

Table 8 Influence of coarse regression module on detection accuracy under cyclist category %

算法	mAP			3D mAP
	简单	中等	困难	
本文算法1	85.36	66.56	65.21	72.38
本文算法2	85.62	67.92	66.46	73.33

### 3 结束语

本文提出基于稀疏Transformer的点云三维目标检测算法。通过稀疏Transformer模块显示选择与注意力相关的信息,以学习点云的全局上下文特征,从而提高模型的精确度。设计基于空间特征的粗回归模块,将其生成的初始锚框作为后续回归精确操作的边界框。在KITTI数据集上的实验结果表明,本文算法具有较优的检测精度和鲁棒性。下一步将在点云处理阶段引入点云关键点的采样信息,结合基于关键点和基于体素点云处理算法的优点,设计一种融合特征提取与体素关键点的目标检测算法,以扩大检测网络的感受野并提高定位精度。

### 参考文献

- [1] 刘丹,马世霞.融合超像素3D与Appearance特征的可行驶区域检测[J].计算机工程,2017,43(7):293-297.  
LIU D, MA S X. Travelable area detection fusing superpixel 3D and apperance feature[J]. Computer Engineering, 2017, 43(7):293-297. (in Chinese)
- [2] PARK Y, LEPETIT V, WOO W. Multiple 3D object tracking for augmented reality[C]//Proceedings of the 7th International Symposium on Mixed and Augmented Reality. Washington D. C., USA: IEEE Press, 2008: 117-120.
- [3] 葛俊彦,史金龙,周志强,等.基于三维检测网络的机器人抓取方法[J].仪器仪表学报,2021,41(8):146-153.  
GE J Y, SHI J L, ZHOU Z Q, et al. A robotic grasping method based on three-dimensional detection network[J]. Chinese Journal of Scientific Instrument, 2021, 41(8): 146-153. (in Chinese)
- [4] 方海国.基于深度学习的3D目标检测与抓取研究[D].湘潭:湘潭大学,2020.  
FANG H G. Research on 3D object detection and grasping technology based on deep learning[D]. Xiangtan: Xiangtan University, 2020. (in Chinese)
- [5] 迟旭然,裴伟,朱永英,等. Fast Stereo-RCNN 三维目标检测算法[J].小型微型计算机系统,2022,43(10):2157-2167.  
CHI X R, PEI W, ZHU Y Y, et al. Fast Stereo-RCNN 3D target detection algorithm [J]. Journal of Chinese Computer Systems, 2022, 43(10): 2157-2167. (in Chinese)
- [6] QIN Z Y, WANG J L, LU Y. MonoGRNet: a geometric reasoning network for monocular 3D object localization [EB/OL]. [2021-07-20]. <https://arxiv.org/pdf/1811.10247.pdf>.
- [7] WANG Y, CHAO W L, GARG D, et al. Pseudo-LiDAR from visual depth estimation: bridging the gap in 3D object detection for autonomous driving [C]//Proceedings of Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2019: 8437-8445.
- [8] 于洁潇,张美琪,苏育挺.基于双目视觉的三维车辆检测算法[J].激光与光电子学进展,2021,58(2):301-306.  
YU J X, ZHANG M Q, SU Y T. Three-dimensional vehicle detection algorithm based on binocular vision[J]. Laser & Optoelectronics Progress, 2021, 58(2): 301-306. (in Chinese)
- [9] SHIN K, KWON Y P, TOMIZUKA M. RoarNet: a robust 3D object detection based on region approximation refinement[C]//Proceedings of Intelligent Vehicles Symposium. Washington D. C., USA: IEEE Press, 2019: 2510-2515.
- [10] HUANG T T, LIU Z, CHEN X W, et al. EPNet: enhancing point features with image semantics for 3D object detection [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2020: 35-52.
- [11] 王海,刘明亮,蔡英凤,等.基于激光雷达与毫米波雷达融合的车辆目标检测算法[J].江苏大学学报(自然科学版),2021,42(4):389-394.  
WANG H, LIU M L, CAI Y F, et al. Vehicle target detection algorithm based on fusion of lidar and millimeter wave radar[J]. Journal of Jiangsu University (Natural Science Edition), 2021, 42(4): 389-394. (in Chinese)
- [12] 江泽宇,赵芸.基于边缘卷积的三维目标识别算法[J].浙江科技学院学报,2021,33(3):214-219.  
JIANG Z Y, ZHAO Y. 3D target recognition algorithm based on edge convolution[J]. Journal of Zhejiang University of Science and Technology, 2021, 33(3): 214-219. (in Chinese)
- [13] 刘高天,段锦,范祺,等.基于改进RFBNet算法的遥感图像目标检测[J].吉林大学学报(理学版),2021,59(5):1188-1198.  
LIU G T, DUAN J, FAN Q, et al. Target detection for remote sensing image based on improved RFBNet algorithm [J]. Journal of Jilin University (Science Edition), 2021, 59(5): 1188-1198. (in Chinese)
- [14] ZHOU Y, TUZEL O. VoxelNet: end-to-end learning for point cloud based 3D object detection[C]//Proceedings of Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 4490-4499.

(下转第144页)

(上接第110页)

- [15] LANG A H, VORA S, CAESAR H, et al. PointPillars: fast encoders for object detection from point clouds [C]// Proceedings of Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2019: 12689-12697.
- [16] SHI W J, RAJKUMAR R. Point-GNN: graph neural network for 3D object detection in a point cloud [C]// Proceedings of Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2020: 1708-1716.
- [17] 杨永光. 基于点云的目标检测方法研究[D]. 南京: 南京邮电大学, 2020.
- YANG Y G. Research on point cloud based on object detection algorithm [D]. Nanjing: Nanjing University of Posts and Telecommunications, 2020. (in Chinese)
- [18] BERG A C, FU C Y, SZEGEDY C, et al. SSD: single shot multi-box detector [C]// Proceedings of the European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 21-37.
- [19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. New York, USA: ACM Press, 2017: 6000-6010.
- [20] ZHAO G X, LIN J Y, ZHANG Z Y, et al. Sparse Transformer: concentrated attention through explicit selection [EB/OL]. [2021-07-20]. <https://arxiv.org/abs/1912.11637>.
- [21] BA J L, KIROS J R, HINTON G E. Layer normalization [EB/OL]. [2021-07-20]. <https://arxiv.org/pdf/1607.06450.pdf>.
- [22] YAN Y, MAO Y, LI B. SECOND: sparsely embedded convolutional detection [J]. Sensors, 2018, 18(10): 3337.
- [23] XU J, MA Y X, HE S H, et al. 3D-GIoU: 3D generalized intersection over union for object detection in point cloud [J]. Sensors, 2019, 19(19): 4093-4108.
- [24] SHI S S, WANG Z, SHI J P, et al. From points to parts: 3D object detection from point cloud with part-aware and part-aggregation network [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(6): 836-842.
- [25] SHI S S, WANG X G, LI H S. PointRCNN: 3D object proposal generation and detection from point cloud [C]// Proceedings of Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2019: 770-779.
- [26] LIU Z, ZHAO X, HUANG T T, et al. TANet: robust 3D object detection from point clouds with triple attention [C]// Proceedings of the 24th AAAI Conference on Artificial Intelligence. New York, USA: AAAI Press, 2020: 11677-11684.

编辑 薛晋栋