



## 远距离和遮挡下三维目标检测算法研究

陆军, 李杨, 鲁林超

引用本文:

陆军, 李杨, 鲁林超. 远距离和遮挡下三维目标检测算法研究[J]. 智能系统学报, 2024, 19(2): 259–266.

LU Jun, LI Yang, LU Linchao. Long-distance and occluded 3D target detection algorithm[J]. *CAAI Transactions on Intelligent Systems*, 2024, 19(2): 259–266.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202301001>

## 您可能感兴趣的其他文章

### 基于改进的Faster RCNN面部表情检测算法

Facial expression recognition based on improved Faster RCNN

智能系统学报. 2021, 16(2): 210–217 <https://dx.doi.org/10.11992/tis.201910020>

### 双向特征融合与注意力机制结合的目标检测

Target detection based on bidirectional feature fusion and an attention mechanism

智能系统学报. 2021, 16(6): 1098–1105 <https://dx.doi.org/10.11992/tis.202012029>

### 基于Faster R-CNN的多任务增强裂缝图像检测方法

Multi-task enhanced dam crack image detection based on Faster R-CNN

智能系统学报. 2021, 16(2): 286–293 <https://dx.doi.org/10.11992/tis.201910004>

### 嵌入遮挡关系模块的SSD模型的输电线路图像金具检测

Fittings detection in transmission line images with SSD model embedded occlusion relation module

智能系统学报. 2020, 15(4): 656–662 <https://dx.doi.org/10.11992/tis.202001008>

### 注意力机制和Faster RCNN相结合的绝缘子识别

Insulator recognition based on attention mechanism and Faster RCNN

智能系统学报. 2020, 15(1): 92–98 <https://dx.doi.org/10.11992/tis.201907023>

### 基于跳跃连接金字塔模型的小目标检测

Skip feature pyramid network with a global receptive field for small object detection

智能系统学报. 2019, 14(6): 1144–1151 <https://dx.doi.org/10.11992/tis.201905041>

DOI: 10.11992/tis.202301001

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20231122.1201.002>

# 远距离和遮挡下三维目标检测算法研究

陆军, 李杨, 鲁林超

(哈尔滨工程大学智能科学与工程学院, 黑龙江哈尔滨 150001)

**摘要:** 针对现有三维目标检测算法对存在遮挡及距离较远目标检测效果差的问题, 以基于点云的三维目标检测算法 (3D object proposal generation and detection from point cloud, PointRCNN) 为基础, 对网络进行改进, 提高三维目标检测精度。对区域生成网络 (region proposal network, RPN) 获取的提议区域 (region of interest, ROI) 体素化处理, 同时构建不同尺度的区域金字塔来捕获更加广泛的兴趣点; 加入点云 Transformer 模块来增强对网格中心点局部特征的学习; 在网络中加入球查询半径预测模块, 使得模型可以根据点云密度自适应调整球查询的范围。最后, 对所提算法的有效性进行了试验验证, 在 KITTI 数据集下对模型的性能进行评估测试, 同时设计相应的消融试验验证模型中各模块的有效性。

**关键词:** 目标检测; 深度学习; 激光雷达点云; 远距离目标; 遮挡下目标; 自动驾驶; 区域金字塔; 特征提取

**中图分类号:** TP391.41 **文献标志码:** A **文章编号:** 1673-4785(2024)02-0259-08

中文引用格式: 陆军, 李杨, 鲁林超. 远距离和遮挡下三维目标检测算法研究 [J]. 智能系统学报, 2024, 19(2): 259-266.

英文引用格式: LU Jun, LI Yang, LU Linchao. Long-distance and occluded 3D target detection algorithm[J]. CAAI transactions on intelligent systems, 2024, 19(2): 259-266.

## Long-distance and occluded 3D target detection algorithm

LU Jun, LI Yang, LU Linchao

(College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China)

**Abstract:** To address the limitations of existing 3D target detection algorithms, particularly their poor detection performance for occluded and long-distance objects, we have implemented an enhancement to the PointRCNN network, a 3D object detection algorithm based on point cloud. We began by voxelizing the region of interest obtained from the region proposal network and constructing region pyramids of different scales to capture a wider range of points of interest. Simultaneously, we introduced a point cloud transformer module to enhance the learning of the local features of grid center points. Moreover, we incorporated a sphere query radius prediction module into the network. This addition allows the model to adaptively adjust the sphere query range according to the density of the point cloud. Finally, the effectiveness of the proposed algorithm was validated through rigorous experimental testing. We evaluated the performance of the model using the KITTI data set and designed corresponding ablation experiments to verify the effectiveness of each module in the model.

**Keywords:** target detection; deep learning; Lidar point cloud; long-distance target; occluded target; autopilot; regional pyramid; feature extraction

近年来, 随着卷积神经网络<sup>[1-2]</sup>的提出及其在计算机视觉<sup>[3]</sup>和自然语言处理<sup>[4]</sup>等领域的广泛应

用, 使得深度学习在二维的图像识别<sup>[5]</sup>、语义分割<sup>[6]</sup>以及目标检测<sup>[7]</sup>等领域有了重要的突破。目前, 基于二维图像的目标检测算法已趋于成熟, 并已经被广泛地应用到我们的生活中。

尽管二维图像领域已经存在许多成熟的检测框架, 如两阶段的 R-CNN<sup>[8]</sup>、Fast R-CNN<sup>[9]</sup>、Faster

收稿日期: 2023-01-02. 网络出版日期: 2023-11-22.

基金项目: 国家自然科学基金项目 (52171332); 黑龙江省自然科学基金项目 (F201123).

通信作者: 陆军. E-mail: [lujun0260@sina.com](mailto:lujun0260@sina.com).

©《智能系统学报》编辑部版权所有

R-CNN<sup>[10]</sup> 和 Mask R-CNN<sup>[11]</sup>, 以及单阶段的 SSD<sup>[12]</sup> 和 YOLO<sup>[13]</sup> 系列等, 这些算法均展现出非常出色的效果, 但仍然难以满足一些特殊场景下的需求。随着 3D 传感器技术的迅速发展及成本的降低, 3D 目标检测<sup>[14-15]</sup> 逐渐成为当前计算机视觉领域的重要研究方向。3D 目标检测目前有着非常广泛的应用, 如在空间站的建造任务中两个航天器的交会对接就需要光学敏感器和激光雷达协同配合完成。此外, 3D 目标检测在当下还有一个热门的应用领域, 就是无人驾驶系统<sup>[16-17]</sup>。

目前 3D 目标检测算法根据对原始点云数据处理方式的不同, 主要可以分为基于体素特征 (Voxel-based) 的目标检测算法及基于点特征 (Point-based) 的目标检测算法两类。其中基于体素特征的目标检测算法将原始的点云处理成规则排列的体素或者栅格, 然后使用三维的卷积神经网络进行特征提取与目标包围框的预测, 比较具有代表性的算法是 VoxelNet 网络<sup>[18]</sup> 和 PointPillars 网络<sup>[19]</sup>, 尽管点云体素化可以提高检测的速度, 但体素化过程中会导致点云精度的丢失, 降低检测的准确率。而另一类基于点特征的目标检测算法, 如 PointRCNN<sup>[20]</sup>, 则直接在原始的点云上通过深度神经网络进行特征提取, 在原始点云上生成目标包围框, 无需对输入点云进行任何的预处理。与点云体素化相比, 尽管 PointRCNN 等基于点特征的目标检测算法提高了目标的平均检测精度, 但是对远处以及存在遮挡目标物体进行检测时, 由于点云本身的稀疏性<sup>[21]</sup>, 目标区域内包含的点云数量有限, 导致这些方法的检测精度并不高。

本研究针对现有 3D 目标检测算法对存在遮挡及距离较远目标检测效果差的问题, 在 PointRCNN 网络的基础上对其第 2 阶段网络进行改进, 提高目标检测的精度。在利用 RPN 网络<sup>[22]</sup> 获取点云数据中的 ROI 区域后, 网络通过对 ROI 体素化处理并构建不同尺度的区域金字塔来捕获更加广泛的兴趣点; 引入球查询半径预测模块, 使网络可以根据点云的密度自适应调整网格中心点的感知范围; 引入点云 Transformer 模块, 将点云与 Transformer<sup>[23]</sup> 结合来增强对网格中心点局部特征的学习; 最后在 KITTI<sup>[24]</sup> 数据集上通过消融试验对所提算法的有效性进行了验证。

## 1 ROI 区域金字塔

在现有的基于深度学习的两阶段点云目标检测算法中, 第 1 阶段 RPN 网络生成的候选框已经

有着很高的召回率, 但网络最终的检测精度仍然很低。造成这一现象的原因是一些检测目标本身所包含的点云量较少, 还有一些检测目标存在遮挡的情况, 使得 ROI 区域内包含的特征信息不足, 导致网络在第 2 阶段中对这些 ROI 区域的优化效果较差。对于这些包含点云数目较少的 ROI 区域, 除获取目标物体本身的特征信息之外, 通过扩大感受野范围获取物体周围区域的特征信息对于目标的预测也很重要。

在二维目标检测网络的 ROI Pooling 模块中, 仅使用候选区域内部的特征图就可以取得很好的检测效果, 这主要源于 2 个方面, 其一是 RGB 图像的像素点是规则排列的, 提取到的不同特征图所包含的语义信息是十分均匀且丰富的; 其二是图像特征图是经过多层卷积池化得到的, 每个像素点有着较大的感受野。而对于点云场景, 大部分候选框内部的点云稀疏且不完整, 包含的语义信息较少, 因此仅靠 ROI 内部区域的这些少量且非均匀分布的点去预测物体的形状、尺寸以及位置信息是远远不够的。

为保证内部的结构信息不变, 同时获取丰富的上下文信息, 本研究采用区域金字塔的方式对每一个 ROI 区域进行编码。

### 1.1 ROI 体素化处理

在两阶段的目标检测网络中, 第 2 阶段的网络需要经过区域池化的操作将 RPN 网络所生成 ROI 区域转换成标准的特征尺度, 以便于对包围框进行进一步的特征提取与回归。

点云目标检测网络中的区域池化的方式有两种, 第 1 种是固定采样的方式, 在每一个 ROI 区域采样相等数目的点, 如在每个区域随机采样 512 个点, 用 512 个点的特征信息来表示候选框特征。第 2 种是基于网格采样的方式, 通过将每个候选框切分成若干大小均匀的网格, 以每个网格的中心点坐标为球心, 通过球查询的方式捕获周围兴趣点的特征信息, 从而得到每个网格点的特征向量, 最后使用所有网格中心点的特征信息来表示候选框特征, 本研究采取第 2 种方式提取 ROI 特征。

设每个候选框的尺寸为  $(W, L, H)$ , 左下角顶点坐标为  $(x_0, y_0, z_0)$ , 每个 ROI 区域的格子编号为  $(i, j, k)$ , ROI 体素化后的网格中心点坐标  $p_{\text{centre}}^{ijk}$  为

$$p_{\text{centre}}^{ijk} = \left( \frac{W}{N_w}, \frac{L}{N_l}, \frac{H}{N_h} \right) \cdot (0.5 + (i, j, k)) + (x_0, y_0, z_0) \quad (1)$$

其中,  $(N_w, N_l, N_h)$  表示在宽、长、高 3 个方向上所切



分格子的数量。

## 1.2 ROI 区域金字塔构建

与二维的金字塔网络有所不同, 本研究以区域空间大小为基准构建网格金字塔, 通过构建金字塔的方式将网格的中心点扩展到候选框外部的区域。具体网格金字塔构建方式如图 1 所示。

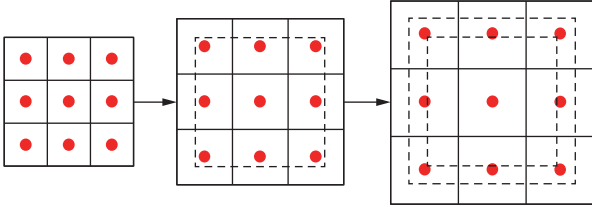


图 1 网格金字塔示意

Fig. 1 Schematic diagram of grid pyramid

由图 1 中可以看出, 网格金字塔以 ROI 的尺寸为基准向外扩张, 最左侧为对 ROI 区域标准网格化后的视角下的网格中心点分布图, 每向外扩张一个量级后, 网格中心点向外偏移。扩张的尺度越大, ROI 区域内包含的点云数目也逐渐增多。小尺度网格用于保存区域内的特征信息, 大尺度网格用于捕获区域外的特征信息。基于式 (1), 网格金字塔每个格子的中心点坐标可以重新表示为

$$p_{\text{centre}}^{ijk} = \left( \frac{p_w W}{N_w'}, \frac{p_l L}{N_l'}, \frac{p_h H}{N_h'} \right) \cdot (0.5 + (i, j, k)) + (x_o, y_o, z_o) \quad (2)$$

式中:  $(p_w, p_l, p_h)$  分别表示沿着不同坐标轴方向的放大系数, 用于控制网格金字塔的尺度大小, 当  $(p_w, p_l, p_h)$  均为 1 时, 表示标准的体素化过程, 随着  $(p_w, p_l, p_h)$  的增加, 金字塔的尺度变大, 所捕获到的上下文特征也更多。  $(N_w', N_l', N_h')$  表示不同尺度区域所切分的网格数量, 金字塔尺度越大网格尺寸也相应越大。对于所有金字塔, 将每个网格内的所有感兴趣的点进行特征聚合, 并将所有尺度金字塔的特征向量进行堆叠得到融合特征, 最终用于目标包围框的精细回归。

## 2 金字塔网格特征提取

### 2.1 Transformer 特征提取

Transformer 由于其强大的注意力机制在机器翻译<sup>[25]</sup>以及自然语言处理领域<sup>[26-28]</sup>有着非常出色的表现。自注意力机制在点云数据的处理上具有天然的优势, 因为点云数据通常不规则地嵌入在度量空间, Transformer 应用在处理点云数据上会有良好的表现, 因此本研究将该模块应用到网格点的特征提取上。

位置嵌入可以帮助网络提取点云局部的结构信息, 在序列数据或者 RGB 图像数据中, 通常依

赖手工特征提取的方式得到位置编码信息, 如通过正余弦函数或者归一化的方式。而对于三维点云数据来说, 其输入特征中本身就包含其空间位置信息, 本研究以每个网格的中心点坐标  $P_{\text{centre}}$  为球心, 通过球查询的方式获得近邻点坐标, 位置编码的公式为

$$\delta = \theta(P_{\text{centre}} - P_k) \quad (3)$$

式中:  $P_{\text{centre}}$  表示网格中心点坐标;  $P_k$  表示球查询所采样到的近邻点坐标, 其中  $k \in \Omega(r)$ ,  $r$  为球查询半径。  $\Omega(r)$  函数由多层感知机 MLP<sup>[29]</sup> 和非线性激活函数 Relu<sup>[30]</sup> 实现。

Transformer 模块的数学表达式为

$$f_{\text{grid}} = \sum_{f_k \in F(i)} \rho(\gamma(\varphi(f_i) - \psi(f_k) + \delta)) \odot (\alpha(f_i) + \delta) \quad (4)$$

式中:  $F(i)$  表示采样点的特征所组成的特征集合;  $\gamma(x)$ 、 $\varphi(x)$  以及  $\alpha(x)$  表示特征映射函数, 用于获取注意力特征, 在网络中均是由两层多层感知机 MLP 与激活函数 Relu 组成;  $\rho(x)$  由 Softmax 和归一化函数组成, 目的是将  $\gamma(x)$  所得到的注意力权重做归一化处理;  $\delta$  表示位置编码信息。

最终点云 Transformer 的结构图如图 2 所示, 在输入层中输入网格中心点以及近邻点的三维坐标  $P$ , 以及各点所对应的特征集合  $F$ 。经过该模块后可获得以网格点为中心的局部特征  $f_{\text{centre}}$ 。

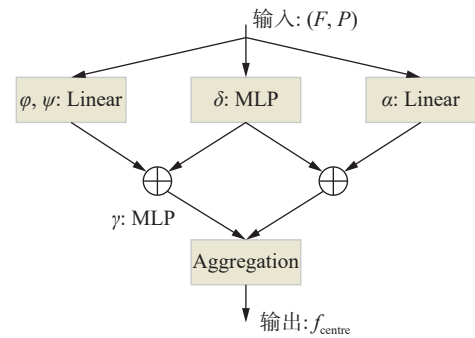


图 2 点云 Transformer 结构

Fig. 2 Structure diagram of point cloud transformer

### 2.2 球查询半径预测

在点云的特征提取中, 球查询的半径决定关键点对近邻点区间的感知范围, 由于点云的稀疏性, 固定的球查询半径会导致不同球部空间内所包含的点的数目不一致, 在 PointNet++ 网络<sup>[31]</sup>中采用多尺度半径的球查询方案, 使用该查询方案时如果半径数目设置过多会增加计算上的开销, 而半径数目设置过少可能会导致产生空球。在本节中针对这一问题采用了新的球查询方案, 在原算法的基础上加入基于密度感知的球查询半径预测模块, 设计一个可微的半径预测函数, 使得球查询半径变为一个可学习的参数, 且网络在学习过程中可以根据点云的稀疏程度调整球查询的半

径大小。

在未修改的网络中,其查询方式通过设置一个固定的半径,以网格中心点为球心,查询半径范围内球部空间内的所有点。对于未修改的球查询方式可以通过概率论的角度进行数学建模,设球查询的半径为  $r$ ,以每个网格中心点为球心,其余点到中心点的距离为

$$d_i = \|p_i - p_{\text{centre}}\|_2 \quad (5)$$

可得所有采样的点均服从 0-1 概率分布,概率分布  $p(i|r)$

$$p(i|r) = \begin{cases} 1, & d_i \leq r \\ 0, & d_i > r \end{cases} \quad (6)$$

局部区域特征的计算过程是一个求取加权平均的过程,从概率的角度出发则是在服 0-1 分布的样本中求数学期望,计算公式为

$$W^i = \rho(\gamma(\varphi(f_i) - \psi(f_k) + \delta)) \quad (7)$$

$$F^i = \alpha(f_i) + \delta \quad (8)$$

$$f_{\text{centre}} = E_{ip(i|r)}[W^i \odot F^i] \quad (9)$$

式中:  $W$  表示权重参数;  $F$  表示样本的特征向量,  $f_{\text{centre}}$  表示所提取到的网格中心点特征。由于最初的 0-1 分布是不可导的,为使得半径变为一个可学习的参数,本研究设计一个新的概率分布表达式来替代原有的 0-1 分布,新的分布函数必须同时满足两个条件,其一是本身要具备与 0-1 分布相似的性质,使大部分样本仍可以落在球部空间内,其二是可以捕获球空间以外部分点的信息,从而获得更加丰富的上下文信息。最终以 Sigmoid 函数作为基准函数重新定义概率分布表达式为

$$s(i|r) = 1 - \text{Sigmod}\left(\frac{d_i - r}{\tau}\right) \quad (10)$$

其中 Sigmoid 函数的表达式为

$$\text{Sigmod}(x) = \frac{1}{1 + e^{-x}} \quad (11)$$

式中:  $s(i|r)$  表示新的概率分布函数;  $\tau$  表示超参数,用于控制概率衰减的速度。当  $\tau$  无限小的时候,位于球内的样本输出值接近于 1,而位于球外的

样本输出值接近于 0,此时该分布近似为 0-1 分布,而且相较于 0-1 分布,该分布更加平滑且可导,随着  $\tau$  不断增加,输出概率的变换也更加均匀平缓。

将新的概率分布函数代入最初的期望表达式中,其相对于  $r$  的梯度可表示为

$$\nabla_r f_{\text{centre}} = \nabla_r E_{i \sim p(i|r)}[W^i \odot F^i] \quad (12)$$

由于无法直接计算参数化分布的梯度,因此式 (12) 仍然无法直接对半径  $r$  进行求导。针对这一问题本研究从基本分布中进行采样,将原始分布参数移动到期望函数内作为系数,最终特征表达式相对于  $r$  的梯度可重新表达为

$$\nabla_r f_{\text{centre}} = E_{i \sim U(\varepsilon)}[\nabla_r E_{i \sim p(i|r)}[W^i F^i]] \quad (13)$$

其中,基本理论分布  $U$  中的  $\varepsilon$  为 1 时表示在整个三维空间中进行采样,当  $R \gg r$  时,导致  $s(i|r)$  趋近于 0,需要将采样的范围限制在与球半径大小近似的空间内。对于新的表达式来说  $s(i|r)$  是可微的,因此可以计算特征表达式相对于  $r$  的梯度,新的特征表达式为

$$f_{\text{centre}} = \sum_{i \in U(\varepsilon)} \rho(\gamma(\varphi(f_i) - \psi(f_k) + \delta)) \odot (\alpha(f_i) + \delta) \cdot (i, r) \quad (14)$$

与原有的基于固定半径的局部特征提取方式相比,在新的特征表达式中,使用更大的基础采样范围,同时将  $s(i|r)$  添加到特征表达式中,使得半径变为一个可以通过网络训练自行学习的参数。

为网格金字塔加入球查询半径预测模块,并对不同尺度的网格金字塔设置不同的球查询基准半径  $r_0$ ,以 ROI 区域的中心点作为球心,设置 2 个固定半径大小的球空间来聚合每个 ROI 区域的上下文的特征信息,将聚合特征输入到 MLP 中预测半径的偏移量  $\Delta r$ ,由于聚合得到的 ROI 特征可以捕获该区域内点云的相关信息,如几何信息以及密度信息等,使得预测得到的  $\Delta r$  可以适应不同区域环境的变化,相比于预先定义的固定半径更加鲁棒。最终,本算法网络结构如图 3 所示。

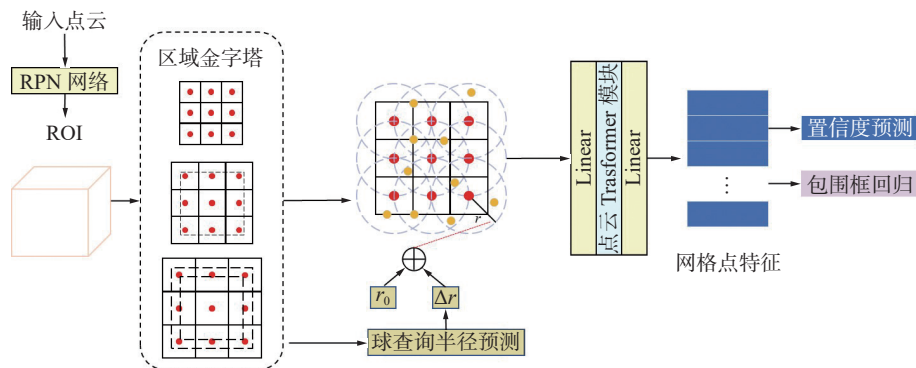


图 3 网络结构

Fig. 3 Network structure diagram

### 3 目标检测算法试验

#### 3.1 试验参数设置

##### 3.1.1 网络结构参数设置

本算法针对 PointRCNN 第 2 阶段的网络结构进行改进,其余部分的结构以及训练参数均与原算法一致,接下来对修改部分的网络结构参数进行介绍。对于第 2 阶段网络,ROI 网格化的部分设置 5 个不同尺度的网格金字塔,每个尺度对应的网格数量分别为  $6^3$ 、 $4^3$ 、 $4^3$ 、 $4^3$ 、1,对应的球查询基准半径设置为 0.2、0.4、0.6、1.2、1.6,每个尺度下的扩张系数设置为 [1, 1, 1]、[1, 1, 1]、[1.5, 1.5, 1]、[2, 2, 1]、[4, 4, 1],不同金字塔等级下,网格中心点球查询采样点的数目设置为 16、16、16、16、32。

##### 3.1.2 训练参数设置

第 1 阶段网络每个批次输入点云的数量 batch size 设置为 16,训练 500 个世代,学习率设置为 0.002,第 2 阶段网络每个批次输入点云的数量 batch size 设置为 8,训练 100 个世代,学习率设置为 0.01,学习率衰减采用余弦退火学习率策略,优化器采用 Adam,为了保证对比试验的有效性,其他训练参数保持不变。

##### 3.1.3 目标检测难度设置

试验中检测精度的评估根据 KITTI 数据标签中的目标是否遮挡、是否被截断以及目标框的边界范围将目标检测的难度划分为 3 个不同的等级,分别是简单、中等以及困难,具体的标准如表 1 所示。

表 1 目标检测困难程度划分标准  
Table 1 Criteria for difficulty of target detection

难度	RGB 图像中 边界框范围/像素	遮挡类型	截断程度/%
简单	40	无遮挡	15
中等	25	小部分遮挡	30
困难	25	大部分被遮挡	50

#### 3.2 试验结果与分析

试验中,通过将本算法与基准算法 PointRCNN 网络对比,验证所提方法的有效性,试验结果是在 KITTI 数据集下训练得到的。本算法是针对提升对遮挡以及包含点云较少的物体的检测效果提出的,而 KITTI 数据集中行人场景数量偏少,结果不具有参考性,因此在本算法的试验中只对车辆目标的检测进行了试验。试验分别测试了车辆目标在简单、中等以及复杂 3 个等级下的检测情况。点云可视化结果中绿色的框代表预先标注好的真值框,红色的框代表算法预测的果。可视

化结果如图 4 所示。

图 4(a)、图 4(b) 分别为 PointRCNN 算法和本算法的点云检测结果,图 4(c) 为点云场景所对应的本算法 RGB 图像检测结果,在选取的 3 个场景中均存在距离雷达传感器较远的目标。在第一个场景的可视化结果中存在一辆位于右侧车道远处的车辆,PointRCNN 算法未能成功检测到该车辆,而本算法成功检测到该目标;而在第 3 个场景中本算法成功检测到存在的较小且被严重遮挡的目标,PointRCNN 算法则未能检测到,结果表明改进后的第 2 阶段网络对提高遮挡以及距离较远目标的检测精度有显著的作用。

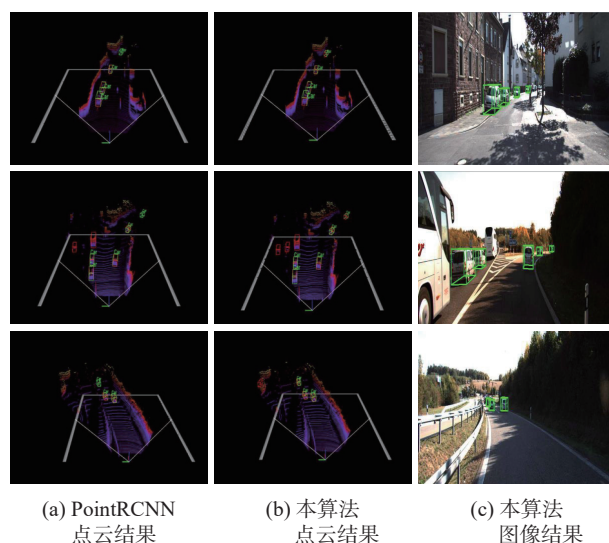


图 4 车辆检测点云可视化结果

Fig. 4 Point cloud visualization results of car detection

当 IoU 设置为 0.7 时,几种不同算法在点云场景与点云鸟瞰图场景中的精度 mAP 的检测结果如表 2 和表 3 所示。与 PointRCNN 算法相比较,本算法在对中等以及困难目标的点云场景中的检测结果中,mAP 值分别提升了 3.77% 和 1.70%,所有等级目标的平均检测精度提升了 1.75%,在鸟瞰图中平均 mAP 值提升了 1.11%,均有了十分明显的提升,而对简单目标的检测精度与改进前基本保持一致。

#### 3.3 消融试验

本算法针对 PointRCNN 算法进行改进,加入了 3 个模块:区域金字塔模块、点云 Transformer 模块以及基于密度感知的半径预测模块。为验证各个模块在模型中发挥的作用,本研究进行了消融试验。以 PointRCNN 算法作为基准网络模型,分别比较在基准模型中叠加不同模块在车辆目标中的检测结果,试验结果如表 4 所示,其中表格中 1 表示在网络中加入该模块,0 表示删除该模块。



表 2 各算法在 KITTI 验证集中车辆目标的 3D 检测结果 (mAP)

Table 2 3D detection results of each algorithm for car targets in KITTI(mAP)

%

算法	输入数据类型	简单	中等	困难	mAP
MV3D	RGB图像+点云	71.29	62.68	56.56	63.51
F-Point Net	RGB图像+点云	83.76	70.92	63.56	72.75
Voxel Net	点云	81.98	65.46	62.85	70.10
SECONED	点云	87.43	76.48	69.10	77.67
PointRCNN	点云	87.92	77.69	76.43	80.68
本算法	点云	<b>88.41</b>	<b>81.39</b>	<b>77.94</b>	<b>82.58</b>

表 3 各算法在 KITTI 验证集中车辆目标 BEV 检测结果 (mAP)

Table 3 BEV detection results of each algorithm for car targets in KITTI(mAP)

%

算法	输入数据类型	简单	中等	困难	mAP
MV3D	RGB图像+点云	86.55	78.10	76.67	80.44
F-Point Net	RGB图像+点云	88.16	84.02	76.44	84.45
Voxel Net	点云	89.60	84.81	78.57	84.33
SECONED	点云	89.96	87.07	79.66	85.56
PointRCNN	点云	91.21	86.89	82.51	86.87
区域金字塔算法	点云	91.15	<b>88.05</b>	<b>84.64</b>	<b>87.95</b>

表 4 消融试验比较结果 (mAP)

Table 4 Results of ablation experiment (mAP) %

算法	区域 金字塔	点云 Transformer模块	半径 预测模块	mAP
PointRCNN	0	0	0	80.83
A	1	0	0	81.75
B	1	1	0	82.26
C	1	1	1	82.58

试验结果显示,原网络加入了网格金字塔模块后平均精度提升了 0.92%,结果表明网格金字塔可以捕获更加丰富的上下文信息,对提升目标的检测精度有显著作用;再加入点云 Transformer 模块后目标检测精度提升了 0.51%,结果表明点云 Transformer 模块的加入能有效提升点云的特征提取效果;加入半径预测模块以后目标检测精度再次提升了 0.32%,表明半径预测模块能有效扩大稀疏处点云感受野范围。

综合以上结果来看,3 个模块都能够提升网络模型的检测精度,其中网格金字塔与半径预测模块均能通过扩大特征提取时的感受野范围从而获得更丰富的上下文信息。相对而言,网格金字塔模块中感受野扩大范围是固定的,并且融合了多尺度特征,因此其能有效提高各种难度目标检测精度;而半径预测模块中感受野扩大范围与点云稀疏程度相关,在检测遮挡程度较高的中等和

困难目标时,其能在网格金字塔模块基础上进一步扩大感受野范围从而提高检测精度。

## 4 结束语

针对目前 3D 目标检测算法对远处及存在遮挡目标物体检测精度不高的问题,以 PointRCNN 网络为基础进行改进,提高目标检测的精度。通过在 PointRCNN 第 2 阶段网络构建多尺度的网格金字塔,捕获 ROI 区域以外的点云,获得更加丰富的 ROI 上下文特征,提高目标包围框的回归精度。同时将 Transformer 中的自注意力机制应用到点云的特征学习中,优化点云的特征提取效率。加入基于密度感知的球查询半径预测模块,根据点云的密度自适应的调整球查询范围,解决了应对点云分布不均匀的问题。最终在 KITTI 数据集中的试验结果表明本研究所提算法能有效提高远距离和遮挡目标的检测精度,同时在对中等以及困难程度目标的检测精度上均有较为明显的提升。

## 参考文献:

- [1] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.

- [2] 张驰, 郭媛, 黎明. 人工神经网络模型发展及应用综述[J]. 计算机工程与应用, 2021, 57(11): 57–69.  
ZHANG Chi, GUO Yuan, LI Ming. Review of development and application of artificial neural network models[J]. Computer engineering and applications, 2021, 57(11): 57–69.
- [3] 许德刚, 王露, 李凡. 深度学习的典型目标检测算法研究综述[J]. 计算机工程与应用, 2021, 57(8): 10–25.  
XU Degang, WANG Lu, LI Fan. Review of typical object detection algorithms for deep learning[J]. Computer engineering and applications, 2021, 57(8): 10–25.
- [4] GUO Menghao, XU Tianxing, LIU Jiangjiang, et al. Attention mechanisms in computer vision: A survey[J]. Computational visual media, 2022(3): 331–368.
- [5] ZHANG Suzhi, WU Yuhong, CHANG Jun. Survey of image recognition algorithms[C]//2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference. Chongqing: IEEE, 2020: 542–548.
- [6] YUAN Xiaohui, SHI Jianfang, GU Lichuan. A review of deep learning methods for semantic segmentation of remote sensing imagery[J]. *Expert systems with applications*, 2021, 169: 114417.
- [7] FANG Wei, SHEN Liang, CHEN Yupeng. Survey on image object detection algorithms based on deep learning[C]//SUN X, ZHANG X, XIA Z, et al. International Conference on Artificial Intelligence and Security. Cham: Springer, 2021: 468–480.
- [8] UIJLINGS J R R, VAN DE SANDE K E A, GEVERS T, et al. Selective search for object recognition[J]. *International journal of computer vision*, 2013, 104(2): 154–171.
- [9] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2016: 1440–1448.
- [10] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2017, 39(6): 1137–1149.
- [11] HE Kaiming, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//IEEE Transactions on Pattern Analysis and Machine Intelligence. [S. l.]: IEEE, 2018: 386–397.
- [12] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//European Conference on Computer Vision. Cham: Springer, 2016: 21–37.
- [13] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 779–788.
- [14] XIA C, WEI P, WEI W, et al. A multilevel fusion network for 3D object detection[J]. *Neurocomputing*, 2021, 437(2): 107–117.
- [15] WEN Lihua, JO K H. Three-attention mechanisms for one-stage 3-D object detection based on LiDAR and camera[J]. *IEEE transactions on industrial informatics*, 2021, 17(10): 6655–6663.
- [16] SHI Shaoshuai, GUO Chaoxu, JIANG Li, et al. PV-RCNN: point-voxel feature set abstraction for 3D object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 10526–10535.
- [17] YIN Tianwei, ZHOU Xingyi, KRÄHENBÜHL P. Center-based 3D object detection and tracking[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 11779–11788.
- [18] ZHOU Yin, 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. Salt Lake City: IEEE, 2018: 4490–4499.
- [19] 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. Long Beach: IEEE, 2020: 12689–12697.
- [20] LI Y, BU R, SUN M, et al. Point CNN: Convolution On X-Transformed Points[EB/OL]. (2018–01–23)[2023–10–15]. <https://arxiv.org/abs/1801.07791>.
- [21] LI Yusheng, TIAN Yong, TIAN Jiandong. Sparse 3D point clouds segmentation considering 2D image feature extraction with deep learning[C]//Proc SPIE 11179, Eleventh International Conference on Digital Image Processing. Guangzhou: SPIE, 2019, 11179: 62–70.
- [22] TANG Xu, ZHANG Huayu, MA Jingjing, et al. Supervised adaptive-RPN network for object detection in remote sensing images[C]//IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium. Waikoloa: IEEE, 2021: 2647–2650.
- [23] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. *Advances in neural information processing systems*, 2017, 30: 6000–6010.
- [24] GEIGER A, LENZ P, STILLER C, et al. Vision meets robotics: the KITTI dataset[J]. *The international journal of robotics research*, 2013, 32(11): 1231–1237.
- [25] 李亚超, 熊德意, 张民. 神经机器翻译综述[J]. 计算机学报, 2018, 41(12): 2734–2755.  
LI Yachao, XIONG Deyi, ZHANG Min. A survey of neural machine translation[J]. *Chinese journal of computers*, 2018, 41(12): 2734–2755.
- [26] GU Fengwei, LU Jun, CAI Chengtao, et al. EANTrack:



an efficient attention network for visual tracking[J]. IEEE transactions on automation science and engineering, 2023(99): 1–18.

- [27] GU Fengwei, LU Jun, CAI Chengtao. RPformer: a robust parallel transformer for visual tracking in complex scenes[J]. IEEE transactions on instrumentation and measurement, 2022, 71: 1–14.
- [28] 石磊, 王毅, 成颖, 等. 自然语言处理中的注意力机制研究综述 [J]. 数据分析与知识发现, 2020, 4(5): 1–14.  
SHI Lei, WANG Yi, CHENG Ying, et al. Review of attention mechanism in natural language processing[J]. Data analysis and knowledge discovery, 2020, 4(5): 1–14.
- [29] 周志华. 机器学习 [M]. 北京: 清华大学出版社, 2016: 298–300.
- [30] AGARAP A F. Deep learning using rectified linear units (ReLU)[EB/OL]. (2018–04–22)[2023–10–15]. <https://arxiv.org/abs/1803.08375>.
- [31] QI C R, YI Li, SU Hao, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space[J]. Advances in neural information processing systems, 2017, 30: 30–38.

## 作者简介:



陆军, 教授, 博士生导师, 主要研究方向为计算机视觉、智能控制技术和机械臂控制。获国防科学技术进步奖一等奖 3 项, 省部级二等奖 1 项, 省部级三等奖 3 项, 授权发明专利 10 项。发表学术论文 70 余篇。E-mail: [lujun0260@sina.com](mailto:lujun0260@sina.com)。



李杨, 硕士研究生, 主要研究方向为点云目标检测、目标跟踪, 机器视觉、图像处理。E-mail: [liyang142857@126.com](mailto:liyang142857@126.com)。



鲁林超, 硕士研究生, 主要研究方向为点云目标检测、特征融合, 图像处理。E-mail: [llczsr@163.com](mailto:llczsr@163.com)。

# 2024 IEEE“一带一路”人工智能可持续发展大会

携手共赴金色十年壮阔新征程, 合力共创智能时代可持续发展。

2024 年是“一带一路”倡议“下一个金色十年”的崭新起点, 也是高质量共建“一带一路”八项行动的重要节点。推动科技创新作为“八项行动”的关键内容, 为通用人工智能时代下世界各国现代化发展带来无限可能, 为一带一路可持续发展目标实现注入强劲动能。

为此, 中国人工智能学会联合 IEEE China Council 主办《2024 IEEE“一带一路”人工智能可持续发展大会》, 携手打造国际化交流平台, 构建全球性共创模式, 为一带一路高质量、可持续发展凝共识、集众智、聚合力、谋未来。

大会将于 2024 年 6 月 22—23 日在中国杭州举办, 邀请国内外“一带一路”相关国家政府机构代表、科研院所及企业负责人、学术研究机构、社会组织各界嘉宾共同参加, 围绕人工智能大模型应用创新, 开展大模型创新算法及关键技术研究, 探索通用人工智能新路径, 推动“一带一路”通用人工智能可持续发展。

会议网址

<https://aicon2024.casconf.cn/>