采用稀疏 3D 卷积的单阶段点云三维目标检测方法

李悄[1]，李垚辰[1]，张玉龙[1]，唐文能[1]，曹鲁光[2]，左良玉[1]
（1. 西安交通大学电子与信息学部，710049，西安；2. 上海机电工程研究所，201109，上海）

摘要：针对点云体素化的三维目标检测方法中点云的特征提取能力不足的问题，将三维目标检测方法 SECOND 作为基准网络，提出一种采用稀疏 3D 卷积的单阶段点云三维目标检测（Reinforced SECOND）方法。首先，改进点云分组方式形成鲁棒的抽离体素特征的体素特征编码网络；其次，为增强体素中对检测任务有显著贡献的关键特征，同时抑制不相关噪声特征，把堆叠三重注意力机制引入体素特征编码网络；然后，提出残差稀疏卷积单元，设计了残差稀疏卷积中间网络，提高了该网络层的特征提取能力，保留了更多的原始特征信息；最后，把提出的空间语义特征融合（SSFF）模块引入区域建议网络，自适应地融合低级空间特征和高级抽象语义特征，进一步提高了模型特征的表达能力。在 KITTI 开源数据集上的实验结果表明：与之前许多基于网格及基于点的方法相比，所提方法显著提高了三维目标检测性能；与基准网络相比，采用所提方法对 KITTI 测试集 car 类和 cyclist 类进行检测，中等难度级别下的 3D 检测精度分别提高了 5.85%和 8.9%，困难难度级别下的 3D 检测精度分别提高了 8.54%和 8.53%。

关键词：三维目标检测；稀疏 3D 卷积；注意力机制；点云体素化
中图分类号：TP391 **文献标志码**：A
DOI：10.7652/xjtuxb202209012 **文章编号**：0253-987X(2022)09-0112-11

OSID 码

A Single-Stage Point Cloud 3D Object Detection Method Using Sparse 3D Convolution

LI Qiao[1], LI Yaochen[1], ZHANG Yulong[1], TANG Wenneng[1], CAO Luguang[2], ZUO Liangyu[1]
(1. Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China;
2. Shanghai Institute of Mechanical and Electrical Engineering, Shanghai 201109, China)

Abstract: In view of the insufficient ability to extract features of point clouds for 3D object detection methods based on point cloud voxelization, this paper proposes a single-stage point cloud 3D object detection method (Reinforced SECOND) based on sparse 3D convolution using SECOND as the baseline. Firstly, the point cloud grouping method is improved herein to form a robust voxel feature encoding network that extracts voxel features. Secondly, to further enhance the key features in voxels that significantly contribute to the detection task while suppressing irrelevant and noisy features, this paper introduces the stacked triple attention mechanism into the voxel feature encoding network. Then, this paper proposes the residual sparse 3D convolution unit and designs the residual sparse convolution intermediate network, which improves the feature extraction ability of this network layer and retains more original feature information.

收稿日期：2022-01-18。 作者简介：李悄(1996—)，男，硕士生；李垚辰(通信作者)，男，副教授，博士生导师。 基金项目：国家自然科学基金资助项目(61803298)；陕西省重点研发计划资助项目(2022GY-080)。
网络出版时间：2022-05-29 网络出版地址：http://kns.cnki.net/kcms/detail/61.1069.T.20220527.1421.002.html

Finally, this paper introduces the proposed spatial-semantic feature fusion (SSFF) module into the region proposal network, which adaptively fuses low-level spatial features and high-level abstract semantic features, further improving the expressiveness of model features. Experimental results on the KITTI open-source dataset demonstrate the proposed method significantly improves 3D object detection performance compared to many previous grid-based and point-based methods. Compared with the benchmark network of this method, 3D car and cyclist detection in the KITTI test set are improved by 5.85% and 8.9% on the moderate mAP and by 8.54% and 8.53% on the hard mAP.

Keywords: 3D object detection; sparse 3D convolution; attention mechanism; point cloud voxelization

深度学习已经在二维目标检测的视觉任务中取得了显著的进展[1]，在人脸识别[2]、车牌识别[3]和视觉目标跟踪[4]等领域得到充分应用。除了二维场景理解，三维目标检测对于现实世界的许多应用是关键而且密不可分的，例如自动驾驶与计算机视觉。最近的三维目标检测的方法利用不同类型的数据，例如单目图像、RGB-D 图像和 3D 点云数据，最普遍使用的 3D 传感器是 LiDAR(light detection and ranging)传感器，它能够形成 3D 点云，从而来捕捉场景的三维结构[5]。然而，点云数据通常是稀疏的和无序的，如何从不规则的点中提取出独特的特征成为三维目标检测任务中的关键性挑战。

根据特征提取过程中点云的表示形式，可将基于点云的三维目标检测方法分为两类：基于点的方法(又称为直接法)和基于网格的方法(包括俯视图法与体素法)。基于点的方法[6-10]大都采用 PointNet 或者 PointNet＋＋[11-12]网络中的集合抽象(set abstraction，SA)层对输入点云进行多层次的局部特征提取。PointRCNN[7]网络和 3DSSD[8]网络等都是首先利用集合抽象层对输入点云进行下采样之后再进行下游任务的处理。此类方法在处理的过程中充分利用输入点云的几何特征，因此其能够获得更好的检测性能。由于此类方法在处理过程中需要堆叠多次下采样操作和邻域搜索操作，上述两类操作的时间复杂度分别为 $O(n^2)$ 和 $O(n)$，使得其特征提取过程需要消耗大量的时间和计算资源。基于网格的方法将输入点云转化为规则的网格，例如 3D 体素[13-16]或 2D 俯视(bird's eye view，BEV)图[17-19]，从而能够使用 3D 或者 2D CNN 提取特征。其中 PointPillar[17]网络将点云转化为一个基于俯视图的二维网格，使用 PointNet 提取每个网格的特征构成一个二维特征图，将点云压缩成 2D 数据，减少了计算规模，可以直接利用二维卷积网络进行下游任务的处理。SECOND[15]网络作为体素法，则是将点云转化为三维体素并使用稀疏三维卷积直接提取特征。相比于基于点的方法，体素法仅需利用点云坐标将其划分到对应的网格中，该过程的时间复杂度为 $O(n)$，不需进行复杂的下采样和邻域搜索。虽然体素法会对点云进行体素特征编码的预处理，但是点云是稀疏的，大部分是空体素，稀疏三维卷积的应用使得体素法仅需处理少量非空体素，此举大大提高了其检测效率。点云处理的过程中带来了一定的信息损失，使得此类方法的检测精度通常低于基于点的方法。综上可得，基于点的直接法通常具有更好的性能，基于网格的方法通常具有更高的检测效率。因此，在室外交通场景等计算能力受限的场景中，提高基于网格方法的检测性能成为近年来的研究热点。

本文以 SECOND 网络为基准网络，提出了一种基于体素的单阶段三维目标检测方法 Reinforced SECOND，该方法旨在进一步提高基于网格方法的检测精度。为了能够提高模型提取点云特征的能力，本文对基准网络中的点云处理方法和各个子网络都进行改进。

本文的体素特征编码网络在处理点云数据时提高了每个体素中点的信息保留，自适应地增强判别性点的特征以及抑制不稳定点。为了能够进一步解决连续稀疏卷积会丢失部分原始特征信息的问题，提出残差稀疏卷积单元，设计了残差稀疏卷积中间网络。提出的一种新颖的空间语义特征融合模块，自适应地融合低级空间特征和高级抽象语义特征，以提高区域提议网络的稳定性和鲁棒性。与基准网络相比，本文所提方法在 KITTI 测试集中的 car 类和 cyclist 类的 3D 检测精度在中等和困难等级上取得了不错的结果，这使得本文方法超越了当前的许多方法。

1 相关研究

1.1 基于点的三维目标检测

基于点云的三维目标检测方法，一般采用两种方式从不规则点云数据中提取出特征，第一种是基于点的方法。基于点的方法由 PointNet(++)及其变体提供支持，直接从原始点云中提取特征。

F-PointNet[6]使用 PointNet 在 2D 图像目标框裁剪点云完成 3D 目标检测。PointRCNN 网络借鉴 2D 检测器 Faster RCNN[20]的思想，从整个点云生成 3D 建议。3DSSD 网络最远点采样时，将欧氏度量(3DSSD 中称为 D-FPS)和特征度量(3DSSD 中称为 F-FPS)融合在一起，弥补下采样时不同前景实例内部点的损失。STD[9]网络提出从稀疏到密集的策略优化线框提议。VoteNet[10]网络采用霍夫投票进行目标特征分组。虽然通过 PointNet(++)堆叠集合抽象层为点云特征学习提供了灵活的感受域，但是三维空间中的点检索需要巨大的计算成本，本文所提模型做到了较好的实时性。

1.2 基于网格的三维目标检测

直接从不规则点云数据中提取出特征，第二种方式根据一定的分辨率将点云划分为规则的连续的网格，并使用 2D/3D CNN 网络去提取特征。PIXOR[19]网络、ComplexYOLO[18]网络和 PointPillars 网络转换点云为 2D BEV 数据，沿着 x 轴和 y 轴划分为小的像素，从而使用手工特征来代表像素特征。以上方法虽然实现了计算量的下降，但是将点云压缩成 2D 数据，不可避免的出现特征信息的丢失。另外的方法是将点云沿着 x 轴、y 轴和 z 轴均匀地划分为体素网格。早期的 VoxelNet[16]网络和 MVX-Net[14]网络将 3D CNN 应用于所有划分的体素，这导致网络性能不佳。事实上，大多数网格都是空的，对检测任务毫无用处。SECOND 网络引入稀疏卷积[21]和子流形稀疏卷积[22]，避免大量不可用的空体素对计算资源的消耗，因此具有更快的推理速度。虽然基于体素的方法在计算上是高效的，但是在离散化过程中带来了信息丢失，从而降低了细粒度的定位精度。本文从点云处理方式和子网络上进行改进，最大限度提高模型的特征提取能力。

2 采用稀疏 3D 卷积的单阶段点云三维目标检测方法

2.1 点云分组

按照 VoxelNet[16]中处理过程来获得点云数据的体素表示。点云沿 x 轴、y 轴和 z 轴被划分为相同空间大小的 3D 网格，然后对体素中的点进行随机采样。对 car 类、cyclist 类和 pedestrian 类 3 类检测目标，基于真值分布裁剪点云，统一都使用相同的点云范围，假设点云在 3D 空间沿 x 轴、y 轴和 z 轴范围为 W、H、D。使所有类别的检测任务在同一个模型当中。同时所有任务采用的体素分辨率为 v_w、v_h、v_d。相应的体素网格的尺寸是 $W'=\frac{W}{v_w}$，$H'=\frac{H}{v_h}$，$D'=\frac{D}{v_d}$。一个非空体素表示为

$$\boldsymbol{V}=\{\boldsymbol{p}_i=[x_i,y_i,z_i,r_i]^{\mathrm{T}}\in\mathbf{R}^4\},i=1,2,\cdots,t \tag{1}$$

式中：t 表示点云点数，始终满足 $t\leqslant T_{\text{points}}$，$T_{\text{points}}$ 表示每个体素中的最大点数；$\boldsymbol{p}_i$ 表示 i 个点的 x 轴、y 轴和 z 轴坐标值 x_i、y_i、z_i 和反射强度 r_i。

本文方法抛弃原来基准网络的体素特征编码层，受 PointPillars[17]的启发，用一个 10 维向量来增强表示点 p_i 的输入特征，v_x、v_y、v_z 分别表示体素所有点的 x 轴、y 轴和 z 轴坐标的算术平均值；c_x、c_y、c_z 分别表示体素中心点 x 轴、y 轴和 z 轴坐标。每个体素的输入特征集合为

$$\begin{aligned}\boldsymbol{V}_{\text{in}}=\{\hat{\boldsymbol{p}}_i=[&x_i,y_i,z_i,r_i,x_i-v_x,y_i-v_y,z_i-v_z,\\&x_i-c_x,y_i-c_y,z_i-c_z]^{\mathrm{T}}\in\mathbf{R}^{10}\},i=1,2,\cdots,t\end{aligned} \tag{2}$$

最后，点云特征被编码为 3D 向量(T_{voxels}，T_{points}，C_{in})，其中 T_{voxels}是最小批次中体素的最大数，C_{in}是点向量输入尺寸，本文为 10。

2.2 网络模型

本节主要介绍网络结构，分为 4 个子网络：①堆叠三重注意力体素特征编码网络；②残差稀疏卷积中间网络；③空间语义特征融合 2D CNN 主干网络；④多任务检测头。图 1 给出了 Reinforced SECOND 的处理过程。该模型将点云作为输入，并通过体素特征编码网络将它们编码为体素表示。残差稀疏卷积中间网络提取 3D 稀疏特征图，并将 z 轴信息压缩为 2D BEV 特征。2D CNN 主干网络在这一步实现了语义和空间特征的鲁棒提取。最后，多任务检测头生成检测结果。

2.2.1 堆叠三重注意力体素特征编码网络

本节主要介绍新设计的体素特征编码网络，称为堆叠三重注意力体素特征编码网络。体素化后的点云被编码为 3D 向量(T_{voxles}，T_{points}，C_{in})。受 PointPillars 的启发，设计了一个新的体素特征编码

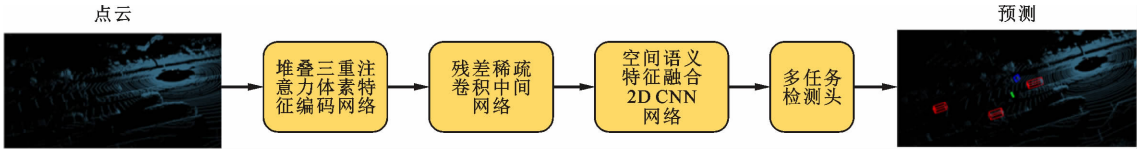


图 1　Reinforced SECOND 的处理过程

Fig. 1　The process of Reinforced SECOND

网络。但是，在提取每个体素中的特征前，为后续操作能学习到每个体素更具识别能力和鲁棒的特征表示。参照 TANet[23] 网络，对体素特征编码网络引入堆叠三重注意力，它利用了多级特征注意力。通过联合考虑 channel-wise、point-wise 和 voxel-wise 注意力，重要的 TA 模块增强了每个体素的关键信息，同时抑制不稳定的点云。图 2 给出了 TA 模块的结构，同时对 TA 模块算法的具体流程和作用进行阐述。

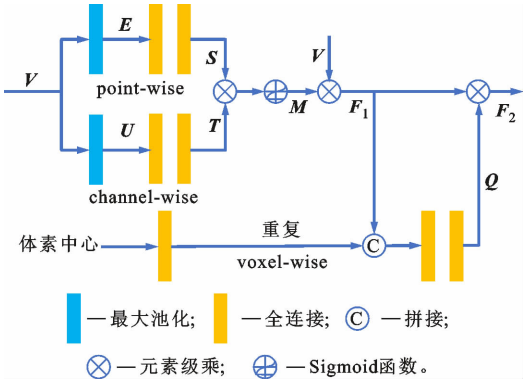


图 2　TA 模块的结构

Fig. 2　The structure of TA module

一个体素网格 $\boldsymbol{V}$ 包含 T_{voxels} 个体素，被定义成 $\boldsymbol{V}=\{\boldsymbol{V}^1,\cdots,\boldsymbol{V}^{T_{\text{voxels}}}\}$，其中 $\boldsymbol{V}^k\in\mathbf{R}^{T_{\text{points}}\times C_{\text{in}}}$ 表示体素网格 $\boldsymbol{V}$ 的第 k 个体素。

(1)point-wise 注意力。假定体素 $\boldsymbol{V}^k$，按照通道维度对体素中的点特征进行最大池化操作，得到 point-wise 特征 $\boldsymbol{E}^k\in\mathbf{R}^{T_{\text{points}}\times 1}$，通过下面的操作得到体素 $\boldsymbol{V}^k$ 的 point-wise 注意力 $\boldsymbol{S}^k\in\mathbf{R}^{T_{\text{points}}\times 1}$

$$\boldsymbol{S}^k=\boldsymbol{W}_2\delta(\boldsymbol{W}_1\boldsymbol{E}^k)\tag{3}$$

式中：$\boldsymbol{W}_1\in\mathbf{R}^{r\times T_{\text{points}}}$ 和 $\boldsymbol{W}_2\in\mathbf{R}^{T_{\text{points}}\times r}$ 表示两个全连接层权重参数；$\delta(\cdot)$ 表示 ReLU 激活函数。

(2)channel-wise 注意力。通过最大池化操作，在它们的 point-wise 维度上聚合通道特性，得到 channel-wise 特征 $\boldsymbol{U}^k\in\mathbf{R}^{1\times C_{\text{in}}}$，通过下面操作得到体素 $\boldsymbol{V}^k$ 的 channel-wise 注意力

$$\boldsymbol{T}^k=\boldsymbol{W}'_2\delta(\boldsymbol{W}'_1(\boldsymbol{U}^k)^{\mathrm{T}})\tag{4}$$

式中：$\boldsymbol{W}'_1\in\mathbf{R}^{r\times C_{\text{in}}}$ 和 $\boldsymbol{W}'_2\in\mathbf{R}^{C_{\text{in}}\times r}$ 表示特征通道对于每个体素的重要性。通过下面的操作得到体素 $\boldsymbol{V}^k$ 的注意力矩阵 $\boldsymbol{M}^k\in\mathbf{R}^{T_{\text{points}}\times C_{\text{in}}}$

$$\boldsymbol{M}^k=\sigma(\boldsymbol{S}^k\times\boldsymbol{T}^k)\tag{5}$$

式中：$\sigma(\cdot)$ 表示 Sigmoid 函数。$\boldsymbol{F}_1^k=\boldsymbol{M}^k\odot\boldsymbol{V}^k\in\mathbf{R}^{T_{\text{points}}\times C_{\text{in}}}$，有效地加权体素 $\boldsymbol{V}^k$ 内所有点在 point-wise 和 channel-wise 维度的重要性。

(3)voxel-wise 注意力。voxel-wise 注意力被用来判断体素的重要性。体素中心是每个体素内部所有点的坐标平均值。再通过全连接层转化到高维空间，并与每个体素特征 $\boldsymbol{F}_1^k$ 逐一拼接。voxel-wise 注意力权重为 $\boldsymbol{Q}=[\boldsymbol{q}^1,\cdots,\boldsymbol{q}^k,\cdots,\boldsymbol{q}^{T_{\text{voxels}}}]\in\mathbf{R}^{T_{\text{voxels}}\times 1\times 1}$。其中，$\boldsymbol{q}^k$ 是通过两个全连接层压缩体素 $\boldsymbol{V}^k$ point-wise 和 channel-wise 维度为 1 得到的。最后，通过 TA 模块对每个体素输出 $\boldsymbol{F}_2^k=\boldsymbol{q}^k\cdot\boldsymbol{F}_1^k$，得到更鲁棒、更有鉴别性的体素特征。

堆叠两个 TA 模块来提取多层特征。第一个 TA 模块直接处理点云的原始特征，第二个模块处理高维特征。对于每个 TA 模块，对其输出和输入执行拼接/元素级和操作以融合更多特征信息，然后通过全连接层获得高维特征表示。通过堆叠三重注意力模块后，体素化的点云重新编码为 3D 向量 $(T_{\text{voxels}},T_{\text{points}},C_{\text{out}})$，其中 C_{out} 是提取体素特征配置的 TA 模块的输出通道参数，设置为 64。

三重注意力模块之后，再经过全连接层后，使用通道最大池化来提取每个体素的局部聚合和对称特征，得到尺寸为 $(T_{\text{voxels}},C_{\text{out}})$ 的输出向量。图 3 给出了堆叠三重注意力体素特征编码网络的完整流程。最大池化后聚合每个体素内的点来提取尺寸为 $(T_{\text{voxels}},C_{\text{out}})$ 紧凑的特征表示，用不同深浅的红色表示不同体素的特征向量。根据体素特征在网格中的原始空间位置对其进行排列，从而得到形状为 $H\times W\times D\times C_{\text{out}}$ 的体素网格特征表示，其中每个网格包含 C_{out} 维度特征，带颜色网格与前面的紧凑排列的相同颜色的特征向量一一对应，其中空白网格表示的是空的体素。

2.2.2　残差稀疏卷积中间网络

原始的基准网络通过稀疏 3D 卷积代替传统的

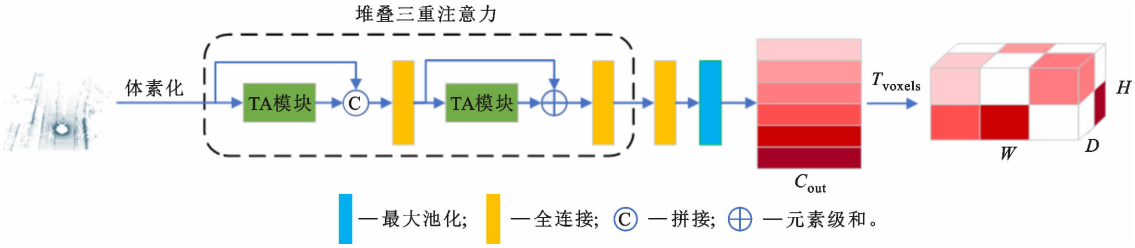


图 3 堆叠三重注意力体素特征编码网络完整流程

Fig. 3 The full pipeline of voxel encoder network with stacked triple attention

3D 卷积提取点云场景中划分的体素的特征，这样大大降低了网络的计算成本。基准网络稀疏卷积中间层网络每个块都是一个 3D 稀疏卷积或者一个 3D 子流形稀疏卷积，接着是 BatchNorm 和 ReLU 操作。

简单堆叠三维稀疏卷积会丢失大量的前期信息。参照 ResNet[24] 网络，设计了残差稀疏卷积单元。该子网络可以利用残差稀疏卷积网络结构变得更深，加快网络的收敛速度，提取到更加重要的 3D 稀疏特征。本文将这种网络命名为残差稀疏卷积中间网络。它由一系列稀疏 3D 卷积(SpConv3D)和残差稀疏卷积(ResSpConv3D)单元组成。图 4 给出了 ResSpConv3D 单元结构，主要由恒等映射和残差映射组成，其中 3×3×3 SpConv3D 和 1×1×1 SpConv3D 分别表示卷积核大小为(3,3,3)和(1,1,1)的稀疏 3D 卷积。

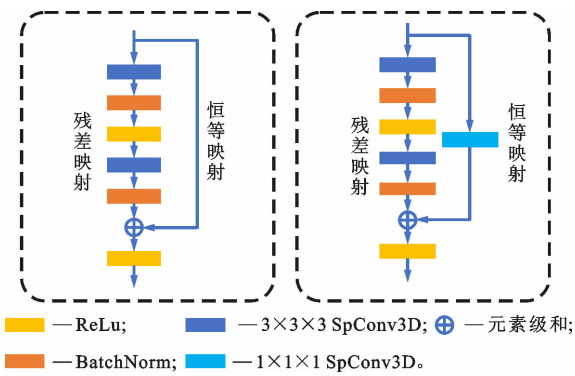


图 4 ResSpConv3D 单元结构

Fig. 4 The architecture of ResSpConv3D unit

其中一个 ResSpConv3D 单元表示为

$$h(\boldsymbol{x}_l)=\begin{cases}\boldsymbol{x}_l, & \text{如果 } \boldsymbol{x}_l \text{ 和 } \boldsymbol{x}_l+1 \text{ 是相同尺寸}\\ \boldsymbol{W}'_l\boldsymbol{x}_l, & \text{其他}\end{cases} \tag{6}$$

$$\boldsymbol{y}_l = F(\boldsymbol{x}_l)+h(\boldsymbol{x}_l) \tag{7}$$

$$\boldsymbol{x}_l+1 = \delta(\boldsymbol{y}_l) \tag{8}$$

式中：$\boldsymbol{x}_l$ 和 $\boldsymbol{x}_l+1$ 表示第 l 个残差单元的输入和输出；$F(\cdot)$ 表示残差映射，它由两个普通的 SpConv3D 组成，是图 4(a)与图 4(b)中的左分支；$h(\cdot)$表示恒等映射，分两种情况，$\boldsymbol{W}'_l$ 表示 1×1×1 卷积用于升维或降维，是图 4(a)与(b)中的右分支；+表示 $h(\boldsymbol{x}_l)$和 $F(\boldsymbol{x}_l)$直接求元素级和(图 4 中的⊕)。

残差稀疏卷积中间网络由 Block1、Block2、Block3 和 Block4 组成。将每个 Block 设计为 SpConv3D 和 ResSpConv3D 的组合，并使用一系列 SpConv3D 和 ResSpConv3D 将点云逐渐转换为 1、2、4、8 倍下采样尺寸的特征体。经过 ToDense 层将 3D 稀疏特征沿 z 轴堆叠，得到 BEV 特征图。图 5 给出了残差稀疏卷积中间网络概述。其中浅蓝色立方模块表示 3D 稀疏特征图，给出了它们的大小，同时给出 Block1、Block2、Block3 和 Block4 子模块的结构。表 1 给出了残差稀疏卷积中间网络参数信息。k、s、p 代表卷积核大小、步幅大小和填充大小。标量以简单的方式使用，例如对于 k，$k=(k,k,k)$，C_{out}代表层的输出通道数，N 代表要应用的层数。其中，ResSpConv3D 包含两个 SpConv3D，都设置为 $k=3$，$s=1$，$p=1$。

表 1 残差稀疏卷积中间网络的网络参数

Table 1 The network parameters of residual sparse convolutional intermediate networks

块编号	卷积层	k	s	p	C_{out}	N
Block1	ResSpConv3D				64	2
Block2	SpConv3D	3	2	1	64	1
Block2	ResSpConv3D				64	2
Block3	SpConv3D	3	2	1	64	1
Block3	ResSpConv3D				64	2
Block4	SpConv3D	3	2	1	128	1
Block4	ResSpConv3D				128	2

2.2.3 空间语义特征融合 2D CNN 主干网络

经过残差稀疏卷积中间网络，得到的压缩的

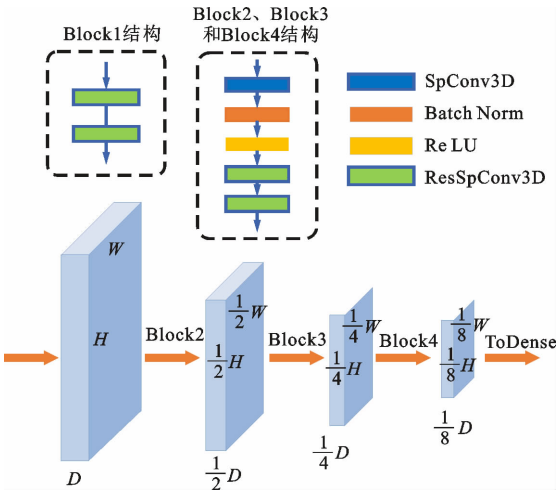


图 5　残差稀疏卷积中间网络的概述

Fig. 5　The overview of residual sparse convolutional intermediate network

BEV 特征图作为区域建议网络的输入。为了准确检测目标，必须回归目标的精确位置且分辨每个回归框作为正/负样本，因此考虑到低级空间特征和高级的抽象语义特征。当堆叠卷积层获取到高级的抽象语义特征，会导致低级空间特征在最终的特征图有所下降。因此，简单堆叠卷积层的 BEV 特征提取模块难以获得具有丰富空间信息的特征。

新设计的 2D CNN 主干网络包括两组卷积群和空间语义特征融合模块。两个卷积群分别称为空间卷积群和语义卷积群，各自的输出分别为空间特征和语义特征。图 6 为提出的区域建议网络的结构图。

空间语义特征融合模块可以自适应地融合空间特征和上采样的语义特征，以实现更鲁棒的特征提取。经过无空间语义特征融合模块 2D CNN 主干网络，空间和语义特征的大小是 $W'\times H'\times C$ 和 $\frac{W'}{2}\times\frac{H'}{2}\times 2C$，它们各自经过空间卷积群和语义卷积群，包括若干的二维卷积(Conv2D)操作，然后经过二维转置卷积(DeConv2D)，其中 DeConv2D 保证输出大小都为 $W'\times H'\times 2C$。每个 Conv2D 或者 DeConv2D 后面会有 BatchNorm 和 ReLU 操作。基准网络是将提取到的空间和语义特征进行简单的拼接操作，对于每个特征的权重无法更加动态的调整。因此在此基础上，在本子网络提出了一个自适应的特征融合模块。首先，压缩空间和语义特征的通道数为 1。接着，用 Softmax 归一化得到的两个结果，得到了 BEV 注意力图，保证了里面的权重参数都在[0,1]这个范围内，同时建立两种特征之间的依赖关系，从而实现自适应特征融合。最后，将 BEV 注意力图作为各自特征的权重，再让两个特征分别和各自 BEV 注意力特征图执行元素级乘(图 6(b)中的⊙)操作，得到新的空间和语义特征，最后再将它们进行拼接(图 6(b)中的Ⓒ)操作。

2.2.4　多任务检测头

在得到空间语义特征融合模块融合得到的特征图后，将运用 3 种卷积核大小为 1×1 二维卷积作用于得到的特征图，输出的通道数分别为 C_c、C_l 和 C_d，表示类别分类、位置回归和方向分类的输出通道数。其中图 6(c)为多任务检测头示意图。使用多个不同尺寸的锚框支持多类检测。本文使用与基准网络相同的值，并遵循 KITTI 数据集基准的交并比(intersection over union，IoU)的阈值，并采用了与基准网络相同的框编码函数。

2.3　损失函数设计

2.3.1　位置回归的 SmoothL1 函数

将位置回归分成定位回归损失 $L_{\text{reg-other}}$ 和角度回归损失 $L_{\text{reg-}\theta}$。其中，本文采用的是角度回归的正弦误差损失的角度回归方式，解决了三维回归框 0 和 π 朝向角的区分问题，自然地根据角度偏移函数对 IoU 进行建模。角度回归损失 $L_{\text{reg-}\theta}$ 的正弦误差损失定义如下

$$L_{\text{reg-}\theta}=L_{\text{smoothL1}}(\sin(\theta_{\text{p}}-\theta_{\text{t}})) \tag{9}$$

定位回归损失 $L_{\text{reg-other}}$ 定义如下

$$\begin{cases}L_{\text{reg-other}}=L_{\text{smoothL1}}(a_{\text{p}}-a_{\text{t}})\\ a\in\{x,y,z,l,w,h\}\end{cases} \tag{10}$$

式中：下标 p 表示预测值；下标 t 表示编码值；x、y 和 z 表示线框中心坐标；l、w 和 h 分别表示线框的长、宽和高；L_{smoothL1} 表示位置回归采用的是 Smooth L1 的损失函数。

2.3.2　分类的焦点损失函数

一般在 KITTI 场景的点云会预制多达 70 000 个锚框，然而只有极少的真值标注框，每个只对应 4～6 个目标框，这就导致前景框数和背景框数极不平衡。为解决此问题，引入焦点损失函数，其定义如下

$$F(p_{\text{t}})=-\alpha_{\text{t}}(1-p_{\text{t}})^{\gamma}\ln(p_{\text{t}}) \tag{11}$$

式中：p_{t} 表示样本属于真实类别概率；α_{t} 和 γ 是焦点损失函数的超参数，为了和基准网络实验对比，采用与基准网络相同的值。

2.3.3　多任务损失函数

对于每个类别，设置相同的损失函数。最终的多任务损失函数定义如下

$$L_{\text{total}}=\alpha L_{\text{cls}}+\beta(L_{\text{reg-}\theta}+L_{\text{reg-other}})+\lambda L_{\text{dir}} \tag{12}$$

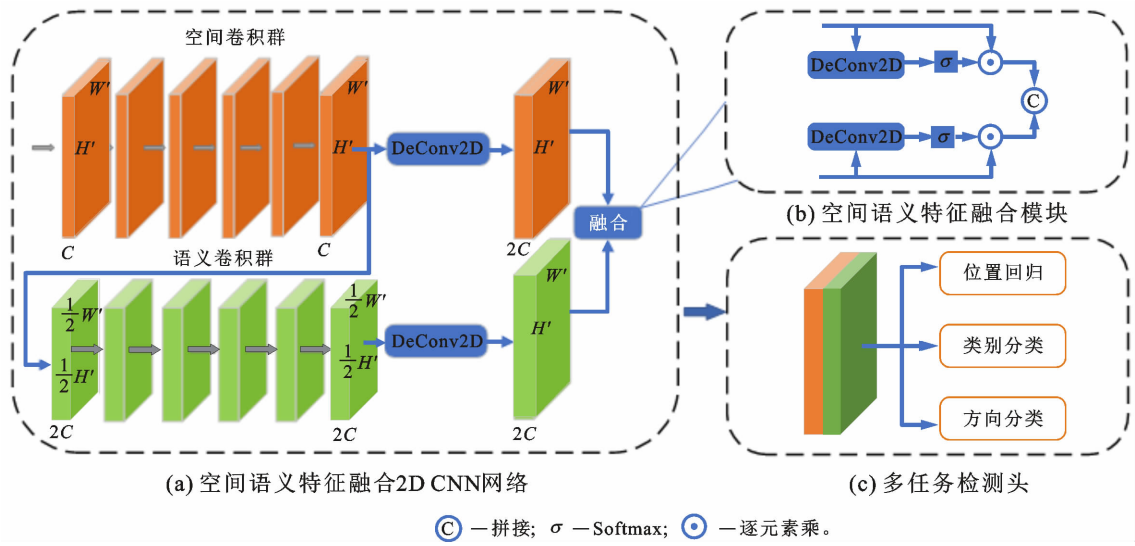


图 6 提出的区域建议网络的结构图

Fig. 6 Structure diagram of the proposed regional proposal network

式中：L_{cls}表示分类损失，使用上面提到的焦点损失函数；$L_{reg-other}$和$L_{reg-\theta}$表示定位和角度的回归损失，使用 Smooth L1 函数；L_{dir}表示方向分类损失，使用 Softmax 损失函数；α、β和λ表示不同任务的损失函数在总体损失中所占的比重。为了和基准网络对比，本文采用与基准网络相同的损失函数的常数系数，$\alpha=1.0$，$\beta=2.0$，$\lambda=0.2$。

3 实验结果及分析

3.1 实验设置

实验使用 KITTI 数据集，其中包含 7 481 个训练样本和 7 518 个测试样本。训练样本又分为训练集(3 712 个样本)和验证集(3 769 个样本)。对 car 类、cyclist 类和 pedestrian 类 3 个类进行评估。KITTI 数据集根据图像平面中边界框高度、遮挡和截断划分模型，评估难度分别为简单、中等和困难难度级别。因为对测试服务器的访问有限制，所有消融实验均使用验证集评估。按照官方 KITTI 评估指标，以平均均值精度(mean average precision，mAP)评价 3D 和 BEV 检测结果。

实验使用的点云 x 轴、y 轴、z 轴范围分别是 $W=[0\ \text{m},70.4\ \text{m}]$，$H=[-40\ \text{m},40\ \text{m}]$，$D=[-3\ \text{m},1\ \text{m}]$。选择的体素尺寸是 $v_w=0.05\ \text{m}$，$v_h=0.05\ \text{m}$，$v_d=0.1\ \text{m}$。因此，生成的体素网格大小是 $1\ 408\times1\ 600\times40$。将 T_{points}设置成 5，作为每个体素中的最大点数，同时 T_{voxles}设置成 16 000，作为最小批量中的最大非空体素数。

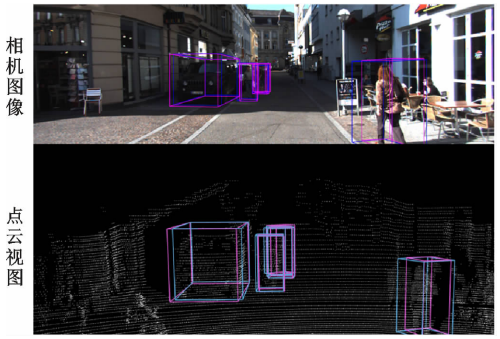
训练的整个网络设置 batch size 为 4，采用 RTX 2080 Ti GPU，设置 80 epochs。采用 Adam 优化器，初始学习率设置为 0.003，指数衰减因子为 0.8，每 15 个周期衰减一次。使用 0.01 的衰减权重，β_1为 0.9，β_2为 0.99。

在训练阶段，使用三维目标检测的数据增强策略。基准值内的点沿 z 轴方向按$[-\pi/4,\pi/4]$的均匀分布进行随机旋转，以获得基准值方位变化。此外，基准值沿 x 轴随机翻转点云。基准值使用[0.95，1.05]均匀分布的随机缩放因子进行全局缩放。这些基准值被随机采样放入原始样本中，以模拟有多个对象的场景。也采用从其他场景中随机“粘贴”一些新的基准值目标到当前的训练场景中进行基准值采样增强，模拟各种环境中的对象。

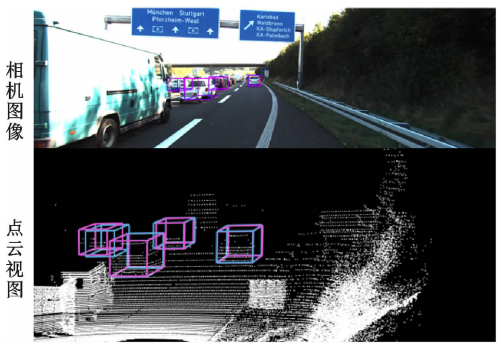
3.2 评估和对比

为了评价所提模型的性能，提供消融实验，在训练集上训练模型，并在验证集上验证结果。为了采用 KITTI 官方测试服务器对测试集进行评估，模型使用训练样本数据的 80%对模型进行训练，剩余的 20%数据用于验证。图 7 给出了 KITTI 验证集上对于 4 种场景的定性结果。通过实验结果可以看出，所提出的网络达到了意想不到的检测效果。KITTI 数据集中一些未标记的对象也可以识别；对远处的小目标、遮挡严重的目标、截断严重的目标能达到较好的识别效果。同时为了客观比较所提方法与其他方法的实时性，在本实验硬件平台上对 5 种方法在 KITTI 验证集的 3D 检测速度进行对比。

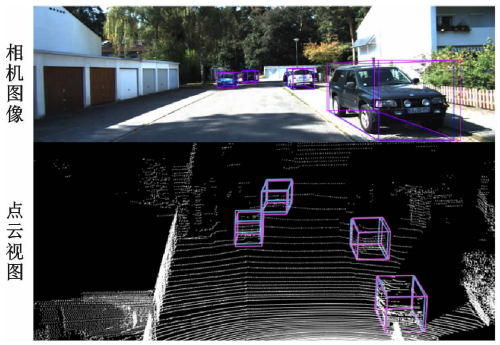
测试集的平均均值精度结果用官方 KITTI 测试服务器上的 40 个召回位置计算。在验证集的运



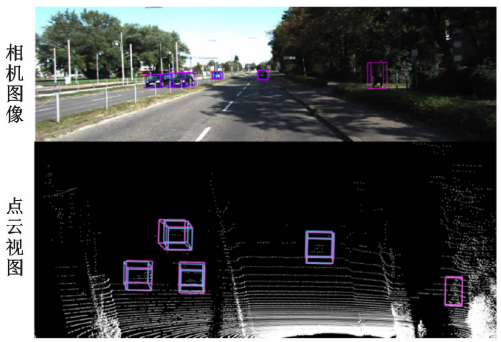
(a)城市场景



(b)高速公路场景



(c)住宅区场景



(d)乡村场景

红色边界框—预测框；蓝色边界框—真实框。

图 7　KITTI 验证集上对于 4 种场景的定性结果

Fig. 7　The qualitative results for four scenarios in the KITTI validation set

行速度，计算的是单帧检测时间/(s·帧$^{-1}$)。表 2 给出了所提方法在 KITTI 测试集上的精度性能，其性能优于之前的基准网络和许多其他方法。对于最重要的 3D 目标检测 car 类，KITTI 测试集上的 3D 检测精度在简单、中等和困难难度级别上分别比基准网络提升了 4.06%，5.85%，8.54%。而且，对于 cyclist 类来说，3D 检测精度在简单、中等和困难难度级别上分别提升了 6.95%，8.9%，8.53%。对于 car 类和 cyclist 类的 BEV 检测，本文方法在 3 个难度级别上也优于许多基于网格及基于点的方法。本文训练了一个同时用于 car 类和 cyclist 类检测的模型，而非基准网络为每个类别训练一个模型。

以上实验说明了网络在 KITTI 测试集上的有效性。在 KITTI 测试集上检测精度得到验证，表 3 给出了 5 种方法在 KITTI 验证集中 3D 检测速度的对比。由表 3 可知，本文所提出的基于体素的方法比 PointRCNN、F-PointNet 等经典的基于点的方法实时性更好，相比基准网络 SECOND，所提方法检测速度变化不大。相比于基于点的方法，在处理过程中利用 PointNet(++)的集合抽象层进行采样操作以及分组操作需要消耗大量的时间，本文方法仅需将点云划分到不同的网格中，不需进行复杂的采样和分组。稀疏三维卷积仅处理少量的非空体素，大大提升基于体素法的计算效率。

3.3　消融研究

所有模型都在训练集上进行训练，并在 KITTI 数据集的验证集上进行评估。本文使用 11 个召回位置计算平均均值精度，其中 car 类的旋转 IoU 阈值为 0.7，cyclist 类和 pedestrian 类的旋转 IoU 阈值为 0.5。表 4 和表 5 给出了 KITTI 验证集中消融实验的 3D 和 BEV 检测性能。表 4、5 中基准子网代表的是采用基准网络的子网络结构。其中消融实验的设置分别以单独、两两结合以及总体结合展示本文改进点的贡献。其中包括 3 组单独实验，2 组两两结合实验，1 组总体实验。由于残差稀疏卷积中间网络和堆叠三重注意力体素特征编码网络输出特征维度关联，因此并没有提供残差稀疏卷积中间网络改进点的单独实验。

本文以 SECOND 网络作为基准网络，尝试改进了其子网络：改进的无注意力机制的体素特征编码网络，记作 NoAtten-VFE；堆叠三重注意力的体素特征编码网络，记作 STA-VFE；残差稀疏卷积中间网络，记作 ReSpConvNet；空间语义特征融合 2D CNN 主干网络，记作 SSFF-2DCNN。

表 2 在 KITTI 测试集上本文方法与其他方法 car 类和 cyclist 类的 3D 检测和 BEV 检测平均均值精度

Table 2 The mAPs of this method and other methods for 3D and BEV detection of cars and cyclists on the KITTI test set

方法	平均均值精度/%											
	car 类-3D 检测			car 类-BEV 检测			cyclist 类-3D 检测			cyclist 类-BEV 检测		
	简单	中等	困难	简单	中等	困难	简单	中等	困难	简单	中等	困难
F-PointNet[6]	82.19	69.79	60.59	91.17	84.67	74.77	72.27	56.12	49.01	77.26	61.37	53.78
AVOD-FPN[25]	83.07	71.76	65.73	90.99	84.82	79.62	63.76	50.55	44.93	69.39	57.12	51.09
SECOND[15]	83.34	72.55	65.82	89.39	83.77	78.59	71.33	52.08	45.83	76.50	56.05	49.45
PointPillars[17]	82.58	74.31	68.99	90.07	86.56	82.81	77.10	58.65	51.92	79.90	62.73	55.58
PointRCNN[7]	86.96	75.64	70.70	92.13	87.39	82.72	74.96	58.82	52.53	82.56	67.24	60.28
TANet[23]	84.39	75.94	68.82	75.70	59.44	52.53	75.70	59.44	52.53	79.16	63.77	56.20
本文方法	86.25	78.40	74.36	91.12	88.15	84.83	78.28	60.98	54.36	81.07	64.00	57.26

表 3 5 种方法在 KITTI 验证集的 3D 检测速度的对比

Table 3 3D detection speed of five methods in the KITTI validation set

方法	SECOND[15]	F-PointNet[6]	PointRCNN[7]	AVOD-FPN[25]	本文方法
检测时间/(s·帧$^{-1}$)	0.040	0.105	0.067	0.098	0.043

表 4 在 KITTI 验证集中消融实验的 3D 检测平均均值精度

Table 4 The mAPs of 3D detection for ablation experiments in the KITTI validation set

体素特征编码网络	稀疏卷积中间网络	2D CNN 主干网络	平均均值精度/%								
			car 类-3D 检测			cyclist 类-3D 检测			pedestrian 类-3D 检测		
			容易	中等	困难	容易	中等	困难	容易	中等	困难
基准子网	基准子网	基准子网	88.14	78.20	77.04	79.11	63.68	60.26	54.17	50.82	46.12
NoAtten-VFE	基准子网	基准子网	88.09	78.34	76.93	81.78	68.61	64.42	58.74	54.13	50.27
STA-VFE	基准子网	基准子网	87.87	78.25	77.13	83.39	68.83	63.99	58.67	54.52	50.02
基准子网	基准子网	SSFF-2DCNN	88.60	78.56	77.27	81.06	67.01	62.90	57.24	53.27	49.36
STA-VFE	ReSpConvNet	基准子网	88.71	78.82	77.53	81.88	68.42	64.98	58.50	54.29	48.43
STA-VFE	基准子网	SSFF-2DCNN	87.82	78.43	77.18	83.23	68.43	64.76	58.08	55.19	50.75
STA-VFE	ReSpConvNet	SSFF-2DCNN	88.89	78.84	77.69	85.22	69.87	66.11	58.99	55.20	50.83

表 5 在 KITTI 验证集中消融实验的 BEV 检测平均均值精度

Table 5 The mAPs of BEV detection for ablation experiments in the KITTI validation set

体素特征编码网络	稀疏卷积中间网络	2D CNN 主干网络	平均均值精度/%								
			car 类-BEV 检测			cyclist 类-BEV 检测			pedestrian 类-BEV 检测		
			容易	中等	困难	容易	中等	困难	容易	中等	困难
基准子网	基准子网	基准子网	89.61	87.58	86.27	87.33	70.61	66.32	59.15	54.12	51.56
NoAtten-VFE	基准子网	基准子网	89.68	87.74	86.17	84.79	71.26	68.31	63.79	57.90	55.44
STA-VFE	基准子网	基准子网	89.70	87.56	86.26	89.57	71.61	68.49	62.93	57.36	54.45
基准子网	基准子网	SSFF-2DCNN	90.15	87.81	86.24	87.24	71.21	67.00	62.10	56.71	54.23
STA-VFE	ReSpConvNet	基准子网	89.97	88.18	86.81	85.62	72.45	69.21	64.58	59.67	55.09
STA-VFE	基准子网	SSFF-2DCNN	89.76	87.78	86.32	85.56	71.87	68.31	63.89	59.58	56.19
STA-VFE	ReSpConvNet	SSFF-2DCNN	90.23	88.20	87.09	91.04	72.64	69.23	64.59	58.34	54.95

3.3.1 无注意力机制的体素特征编码网络的效果

通过与基准网络比较来验证提出的体素特征编码网络的有效性。表 4 给出 KITTI 验证集上 3D 检测性能，在替换 NoAtten-VFE 为特征编码网络后，

模型在 car 类、cyclist 类和 pedestrian 类的中等难度级别平均均值精度分别提升了 0.14%、4.93%和 3.93%，可见对占用点云较少的小物体检测效果提升较好。因为 NoAtten-VFE 引入了 10 维向量对 point-wise 特征进行增强表示，新的结构更好地提取 voxel-wise 特征，虽然小目标点云少，但是可以提取出更多特征。

3.3.2　堆叠三重注意力的效果

为了进一步提取体素的更具辨别力和鲁棒性的特征，在体素特征编码网络引入堆叠三重注意力。同样在 KITTI 验证集进行评估，如表 4 所示，采用 STA-VFE 模型和采用 NoAtten-VFE 模型的实验结果进行对比，在中等难度级别下，cyclist 类和 pedestrian 类 3D 检测精度分别提升了 0.22%、0.39%，同时 car 类依然有轻微的下降，下降了 0.09%，但是困难难度级别下的 car 类确提升 0.21%。说明加入堆叠三重注意力增强了体素编码网络对体素的关键性特征的提取能力。

3.3.3　残差稀疏卷积的效果

针对体素特征编码网络引入堆叠三重注意力的改进，发现 KITTI 验证集中 car 类中等难度级别 3D 检测平均均值精度略有下降。根据 STA-VFE 网络的输出特征维度特点，设计了相应的残差稀疏卷积网络，尝试改进稀疏卷积网络来提高检测效果。如表 4 和表 5 所示，将 STA-VFE 与 ReSpConvNet 结合的模型，与只引入 STA-VFE 的模型对比，car 类在简单、中等和困难难度级别上的 3D 检测精度分别提高了 0.84%、0.57%和 0.4%。同时 BEV 检测精度在不同类的所有难度等级下都有一定提升。说明了残差稀疏卷积单元相比普通稀疏 3D 卷积对于 car 类有更好的检测提升效果。因为残差稀疏卷积的短连接结构，相当于在每个卷积又加入了上一层特征的全部信息，一定程度上保留了更多的点云原始信息。

3.3.4　空间语义特征融合模块效果

如表 4 所示，SSFF-2DCNN 在和 STA-VFE 两两结合，或者与 STA-VFE＋ReSpConvNet 总体结合的模型，都做到进一步提升了各个类不同难度级别下的在 KITTI 验证集 3D 精度。说明了本文提出的空间语义特征融合模块能够有效地融合 2D CNN 的低级空间特征和高级语义特征。

4　结　论

针对点云体素化的三维目标检测方法点云的特征提取能力不足的问题，本文提出了一种基于体素的单阶段三维目标检测（Reinforced SECOND）方法。改进的点云分组方式，对单个体素特征实现更合理的表示，并提出了一种堆叠三重注意力体素特征编码网络，该子网络增强了体素中对检测任务有着重要贡献的关键特征，同时抑制不相关噪声特征。提出残差稀疏卷积单元，设计了残差稀疏卷积中间网络，保留了 3D 稀疏特征图更丰富的信息，解决了连续卷积会丢失部分有效信息的问题。在区域建议网络中，提出了轻量级的空间语义特征融合模块，实现自适应地融合低级空间特征和高级抽象语义特征。在 KITTI 数据集的实验结果表明，与以前许多方法相比，本文方法有效提升了三维目标检测性能。

参考文献：

[1] 陈科圻，朱志亮，邓小明，等. 多尺度目标检测的深度学习研究综述 [J]. 软件学报，2021，32(4)：1201-1227.
CHEN Keqi，ZHU Zhiliang，DENG Xiaoming，et al. Deep learning for multi-scale object detection：a survey [J]. Journal of Software，2021，32(4)：1201-1227.

[2] 张帆，赵世坤，袁操，等. 人脸识别反欺诈研究进展 [J]. 软件学报，2022，33(7)：2204-2240.
ZHANG Fan，ZHAO Shikun，YUAN Cao，et al. Recent progress of face anti-spoofing [J]. Journal of Software，2022，33(7)：2204-2240.

[3] 陈晋音，沈诗婧，苏蒙蒙，等. 车牌识别系统的黑盒对抗攻击 [J]. 自动化学报，2021，47(1)：121-135.
CHEN Jinyin，SHEN Shijing，SU Mengmeng，et al. Black-box adversarial attack on license plate recognition system [J]. Acta Automatica Sinica，2021，47(1)：121-135.

[4] 孟琭，杨旭. 目标跟踪算法综述 [J]. 自动化学报，2019，45(7)：1244-1260.
MENG Lu，YANG Xu. A survey of object tracking algorithms [J]. Acta Automatica Sinica，2019，45(7)：1244-1260.

[5] 田永林，沈宇，李强，等. 平行点云：虚实互动的点云生成与三维模型进化方法 [J]. 自动化学报，2020，46(12)：2572-2582.
TIAN Yonglin，SHEN Yu，LI Qiang，et al. Parallel point clouds：point clouds generation and 3D model evolution via virtual-real interaction [J]. Acta Automatica Sinica，2020，46(12)：2572-2582.

[6] QI C R，LIU Wei，WU Chenxia，et al. Frustum PointNets for 3D object detection from RGB-D data [C]// 2018 IEEE/CVF Conference on Computer Vi-

sion and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 918-927.

[7] SHI Shaoshuai, WANG Xiaogang, LI Hongsheng. PointRCNN: 3D object proposal generation and detection from point cloud [C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2019: 770-779.

[8] YANG Zetong, SUN Yanan, LIU Shu, et al. 3DSSD: point-based 3D single stage object detector [C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 11037-11045.

[9] YANG Zetong, SUN Yanan, LIU Shu, et al. STD: sparse-to-dense 3D object detector for point cloud [C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2019: 1951-1960.

[10] QI C R, LITANY O, HE Kaiming, et al. Deep hough voting for 3D object detection in point clouds [C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2019: 9276-9285.

[11] CHARLES R Q, SU Hao, KAICHUN Mo, et al. PointNet: deep learning on point sets for 3D classification and segmentation [C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 77-85.

[12] QI C R, YI Li, SU Hao, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space [C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2017: 5105-5114.

[13] SHI Shaoshuai, WANG Zhe, WANG Xiaogang, et al. Part-A^2 net: 3D part-aware and aggregation neural network for object detection from point cloud [EB/OL]. [2021-12-09]. https://doi.org/10.48550/arXiv.1907.03670.

[14] SINDAGI V A, ZHOU Yin, TUZEL O. MVX-net: multimodal VoxelNet for 3D object detection [C]// 2019 International Conference on Robotics and Automation (ICRA). Piscataway, NJ, USA: IEEE, 2019: 7276-7282.

[15] YAN Yan, MAO Yuxing, LI Bo. SECOND: sparsely embedded convolutional detection [J]. Sensors, 2018, 18(10): 3337.

[16] 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. Piscataway, NJ, USA: IEEE, 2018: 4490-4499.

[17] 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). Piscataway, NJ, USA: IEEE, 2019: 12689-12697.

[18] SIMON M, MILZ S, AMENDE K, et al. Complex-YOLO: an Euler-region-proposal for real-time 3D object detection on point clouds [C]// Computer Vision: ECCV 2018 Workshops. Cham, Switzerland: Springer International Publishing, 2019: 197-209.

[19] YANG Bin, LUO Wenjie, URTASUN R. PIXOR: real-time 3D object detection from point clouds [C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 7652-7660.

[20] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C] // Proceedings of the 28th International Conference on Neural Information Processing Systems: Volume 1. Cambridge, MA, USA: MIT Press, 2015: 91-99.

[21] GRAHAM B. Sparse 3D convolutional neural networks [EB/OL]. [2021-12-09]. https://doi.org/10.48550/arXiv.1505.02890.

[22] GRAHAM B. VAN DER MAATEN L. Submanifold sparse convolutional networks [EB/OL]. [2021-12-09]. https://doi.org/10.48550/arXiv.1706.01307.

[23] LIU Zhe, ZHAO Xin, HUANG Tengteng, et al. TANet: robust 3D object detection from point clouds with triple attention [C]// Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2020: 11677-11684.

[24] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 770-778.

[25] 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). Piscataway, NJ, USA: IEEE, 2018: 1-8.

（编辑 武红江 刘杨）