基于体素化图卷积网络的三维点云目标检测方法

赵毅强，艾西丁·艾克白尔，陈　瑞，周意遥，张　琦

(天津大学 微电子学院，天津 300072)

摘　要：针对激光雷达点云的稀疏性和空间离散分布的特点，通过结合体素划分和图表示方法设计了新的图卷积特征提取模块，提出一种基于体素化图卷积神经网络的激光雷达三维点云目标检测算法。该方法通过消除传统 3D 卷积神经网络的计算冗余性，不仅提升了网络的目标检测能力，并且提高了点云拓扑信息的分析能力。文中设计的方法在 KITTI 公开数据集的车辆、行人、骑行者的 3D 目标检测和鸟瞰图目标检测任务的检测性能相比基准网络均有了有效提升，尤其在车辆 3D 目标检测任务上最高提升了 13.75%。实验表明：该方法采用图卷积特征提取模块有效提高了网络整体检测性能和数据拓扑关系的学习能力，为三维点云目标检测任务提供了新的方法。

关键词：图卷积神经网络；　激光雷达；　三维点云目标检测；　拓扑信息；　KITTI 数据集

中图分类号：TP183　　**文献标志码：**A　　**DOI：**10.3788/IRLA20200500

3D point cloud object detection method in view of voxel based on graph convolution network

Zhao Yiqiang, Arxidin· Akbar, Chen Rui, Zhou Yiyao, Zhang Qi

(School of Microelectronics, Tianjin University, Tianjin 300072, China)

Abstract: In view of the sparsity and spatial discrete distribution of lidar point cloud, a graph convolution feature extraction module was designed by combining voxel partition and graph representation, and a 3D lidar point cloud object detection algorithm in view of voxel based graph convolution neural network was proposed. By eliminating the computational redundancy of the traditional 3D convolution neural network, this method not only improved the object detection ability of the network, but also improved the analysis ability of the point cloud topology information. Compared with the baseline network, the detection performance of vehicle, pedestrian and cyclist 3D object detection and bird's eye view object detection tasks in KITTI public dataset were improved greatly, especially improved with 13.75% precision in 3D object detection task of vehicle at maximal. Experimental results show that the proposed method improves the detection performance of the network and the learning ability of data topological relationship via graph convolution feature extraction module, which provides a new method for 3D point cloud object detection task.

Key words: graph convolution neural network; lidar; 3D point cloud object detection; topological information; KITTI dataset

收稿日期：2020−12−20；　**修订日期：**2021−04−15

基金项目：国家自然科学基金 (61871284)；天津市科技重大专项研发计划新一代人工智能科技重大专项 (18ZXZNGX00320)

作者简介：赵毅强，男，教授，博士生导师，博士，主要从事集成电路设计和红外成像与感知方面的研究。

0 引 言

随着智能感知系统的快速发展,三维点云目标检测在计算机视觉领域得到了越来越多的关注。激光雷达因探测距离远,精度高,其生成的点云数据对光强和距离等环境因素的鲁棒性好,目前在自动驾驶,遥感环境监测,高精度地图等领域坐拥着重要的地位[1−2]。

激光雷达生成的三维点云由空间离散点组成,可表示目标场景的立体坐标信息。由于点云数据在三维空间的非规则排布和稀疏特性,无法直接利用传统的卷积神经网络来学习点云的局部和高级特征信息。所以通常需要借助预处理步骤来确定点云数据的局部单位,即点云的单位表示是点云数据学习的重要基础步骤。目前采用的表示方法可分为,基于体素划分表示[3−4],基于点的表示方法[5] 等。而图表示方法通过引入图概念,直接在输入点云数据上做算法处理[6]。点云数据具有丰富的拓扑信息和空间感知能力,基于图表示的方法利用数据的图结构能有效地表示空间离散分布的稀疏点云数据,同时利用图卷积操作来学习点云语义信息和邻域点之间的相关性,提高模型学习能力和泛化能力。

1 点云数据处理研究现状

基于神经网络的点云数据处理算法是目前的研究热点。基于不同的点云数据单元表示可分为基于点表示方法的神经网络方法、基于空间划分的神经网络方法和基于拓扑图表示的神经网络方法。

(1) 基于点表示的神经网络方法:无需空间划分,确定邻域搜索长度之后可以采用最临近邻域搜索或采用最远点采样来确定邻域范围内的点集,再使用多层感知机来聚合邻域信息,从而逐步生成高级点云特征。PointNet++[5] 通过最远点采样 1 024 个点,对每个采样点进行多尺度邻域聚合。并通过使用由多层感知机和局部最大值采样层构成的仿射变换矩阵,使得输入点云在特征空间上具有置换不变性。之后在全局范围内叠加使用多层感知机自下而上逐级生成最终的全局特征从而实现点级分类。该方法虽然在点云语义分割任务上得到了不错的效果,但其邻域搜索时间复杂度最高可达$O(N^2)$,在点云目标检测任务中,其时效性会使模型在点云目标的实时检测任务上的应用受到较大的阻碍。

(2) 基于空间划分的神经网络方法:是将整个目标场景划分为规则的三维立方格,从而可以采用传统的三维卷积神经网络学习编码每个体素格内的点集特征。VoxelNet[3] 通过等距划分整个空间生成规则排布的立方格,并在包含点云的体素格中使用多层感知机进行特征编码。随后采用 3D 卷积神经网络遍历整个空间提取高级特征信息,最后通过区域生成网络 (Region Proposal Network, RPN) 生成最终的目标检测结果。体素划分表示方法不但将传统卷积网络方法结合到了点云数据学习上,而且其邻域搜索时间复杂度为$O(1)$,显著提升了邻域搜索的效率,但由于体素化过程中需要规则划分点云,在相邻的目标数据上可能会产生歧义,导致发生信息的损失。

(3) 基于图表示的神经网络方法:图神经网络在点云语义分割任务中得到了不少的研究,但在点云目标检测任务上的应用并未得到足够多的研究。Point-GNN[6] 将输入点云中的每一个点作为图节点,基于节点邻域搜索的方法,建立了输入点云的图结构,从而运用多层感知机作为网络的邻域特征聚合函数,逐层进行邻域划分和特征聚合,该算法将图神经网络引入到点云目标检测任务中,但其检测速度不能完成实时检测,无法匹配激光雷达采样频率。

频谱图卷积借助图谱理论来实现在具有图结构数据上的卷积操作。Kipf 等人采用切比雪夫多项式近似基于归一化对称拉普拉斯矩阵的图卷积核,实现了对非规则排布数据的半监督节点分类[7],被广泛引用在推荐系统算法和自然语言处理等领域。其中,拉普拉斯矩阵为图结构上的拉普拉斯算子,即离散的拉普拉斯算子,可计算目标节点v_i和其邻域节点v_j之间的梯度差,学习邻域节点之间的权重。正如常规卷积中的卷积核聚合计算,逐层学习邻域间的权重,广义上和欧式空间下的卷积神经网络一致。由于点云数据同样具有非规则排布的特性,无法满足平移不变性,传统的卷积操作无法直接应用。因此,该方法对处理非规则排布的点云数据提供了相通的理论基础,实现点云数据的学习。

综上点云数据处理方法,基于点表示的网络框架虽然能对点云数据进行高细粒度分类和分割,但对于百万数量级的输入来说,其运算效率会受到一定的限

制。而基于体素划分的网络结构的时效性好,网络检测精度不俗,但其忽略了邻域之间的关系度量,而且在空间划分步骤产生一定的信息损失。图表示方法可充分利用数据的分布特性和拓扑信息,提高模型表征能力,但运算时效上具有一定局限性。

文中通过体素划分结合图表示方法,设计了一种基于图卷积神经网络的点云目标检测算法。算法采用邻域点特征拼接进行体素格特征补偿编码来弥补体素划分产生的信息损失。并采用基于拉普拉斯矩阵的图卷积模块替代基准网络 VoxelNet[3] 的 3D 卷积模块并输出图卷积编码特征,最终通过 RPN 网络模块生成三维点云目标检测框。为验证算法可行性和目标检测性能,在 KITTI 公开数据集上分别对车辆,行人和骑行者进行了三维目标检测和鸟瞰图 (Bird's Eye View, BEV) 目标检测。

2 基于体素化图卷积网络的目标检测方法

文中提出的一种基于体素化图卷积网络的目标检测算法 (Voxel-based Graph Convolution Network, VGCN),主要包括输入点云的预处理模块,体素格补偿编码层,图卷积模块以及区域生成网络模块。图卷积模块的输出特征通过区域生成网络模块由两个 2D 卷积层生成最终的检测结果,如图 1 所示。其中预处理模块基于体素划分来建立图结构,图卷积模块采用归一化对称拉普拉斯矩阵为卷积核来搭建图卷积前向传播层,不仅实现传统 3D 卷积的局部感知和卷积参数共享的特性,并且可以获取基于图结构的点云拓扑信息,有效学习非规则数据的特征信息,降低了传统卷积算法的处理点云数据时的计算冗余性,提高了目标检测精度。

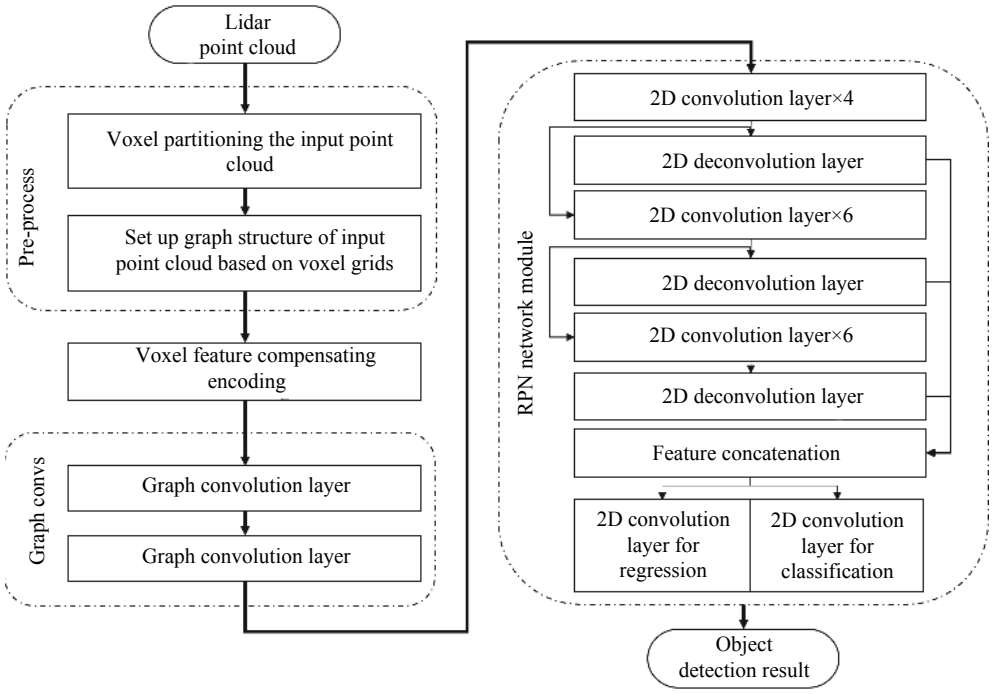


图 1 基于体素划分的图卷积三维目标检测流程

Fig.1 Pipeline of graph convolution 3D object detection based on voxelization

2.1 点云数据体素划分预处理

为方便输入点云的体素划分以及图结构的建立,对输入点云数据 $P_i \in R^3, i = 1,\cdots,N$,划分以车辆当前位置为原点,从前进方向取长度为 L,左右宽度为 W,高度为 H,单位为 m 的目标检测区域: $L \times W \times H = [0,70.4] \times [-40,40] \times [-3,1]$。并在限定的检测范围内,将整个空间划分为均匀立方格单元,生成等大体素格 $v_k, k = 1,\cdots,K$[3]。

2.2 基于体素划分的点云图结构建立

体素格尺寸的选择是影响算法的检测精度和运算效率的重要参数,当设定体素格的大小为 $v_k = (v_l, v_w, v_h) = (0.2, 0.2, 0.4)$ m 时,包含点云的非空体素格

仅占整个空间的 5%~10%, 若使用传统 3D 卷积神经网络将遍历整个空间, 即包括没有数据的体素格, 会极大地降低模型效率。而基于体素划分的图表示方法可以快速确定每个体素格的坐标, 即包含点云的有效体素格节点 V, 并建立节点之间的邻接关系 E, 从而构建输入点云的图结构 $G=(V,E)$, 如图 2 所示, 图 2 中, 左边图中的黑框为目标体素格, 红框为有效邻域体素格。

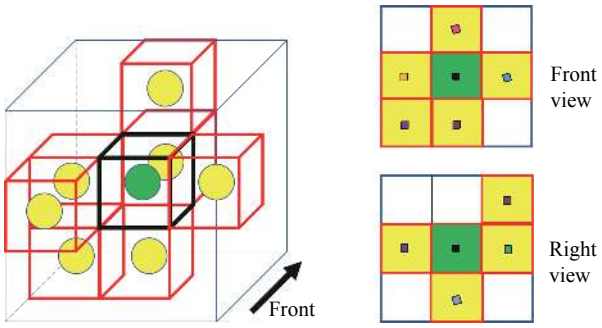


图 2 有效体素格空间邻域关系示意图

Fig.2 Schematic diagram of neighborhood of valid voxel

基于体素划分的方法不但具有较好的时效性, 而且提供了非常便捷的邻域搜索途径, 其邻域搜索时间复杂度为 $O(1)$。图卷积选取在目标节点的邻域半径 H 内的有效体素格, 建立体素格节点的邻接矩阵 A, 以及表示每个节点的邻接节点数量的度矩阵 D, 从而利用邻接矩阵和度矩阵搭建图卷积核。

2.3 体素格特征补偿编码

VoxelNet 采用的体素格特征编码 (Voxel Feature Encoder, VFE) 方法仅考虑体素格内的点云数据进行特征编码, 忽略了体素划分产生的信息损失。文中通过采用邻域点特征拼接的方法对每个包含点云的有效体素格进行体素特征补偿编码 (Voxel Feature Compensate Encoding, VFCE), 从而补偿体素划分产生的信息损失。利用先前建立的邻接矩阵 A, 可以方便的确定邻域体素格 v_j 包含的点云 p_j。从中随机采样 T 个邻域点特征 $f(p_t), t=1,2,\cdots,T$, 对目标体素格点特征 $f(p_k), k=0,1,\cdots,K$ 做特征拼接:

$$f_{VFE}=\sigma(MLP(f(p_k)\oplus f(p_t))) \tag{1}$$

式中: $f(p_k)$ 为目标体素格内的点特征; $f(p_t)$ 为随机采样的邻域点特征; $\oplus$ 为特征拼接; MLP 为多层感知机; σ 为非线性激活函数, 实验取 T=15; $VFCE$ 层的输出为有效体素格节点特征, $f_{VFE}\in R^{K\times 128}$。

2.4 基于拉普拉斯卷积核的前向传播层搭建

基于 K 个有效体素格节点的相互邻接关系建立的邻接矩阵 A 和度矩阵 D, 构建图卷积层的对称归一化拉普拉斯矩阵作为卷积核[7]:

$$Y=\widetilde{D}^{1/2}\cdot\widetilde{A}\cdot\widetilde{D}^{1/2}, Y\in R^{K\times K} \tag{2}$$

式中: $\widetilde{A}=A+I_N$ 为归一化处理的邻接矩阵, I_N 为单位矩阵, 在叠加使用多个前向传播层时可避免模型出现梯度消失或梯度爆炸等数值紊乱问题, $\widetilde{D}_{ii}=\sum_j\widetilde{A}_{ij}$ 为归一化邻接矩阵对应的度矩阵。图卷积前向传播层的表达式为:

$$O^{l+1}=\sigma\left(\widetilde{D}^{1/2}\cdot\widetilde{A}\cdot\widetilde{D}^{1/2}\cdot O^{l}\cdot W^{l}\right) \tag{3}$$

式中: $\widetilde{D}^{1/2}\cdot\widetilde{A}\cdot\widetilde{D}^{1/2}$ 为对称归一化拉普拉斯矩阵; O^l 为前一层的图卷积输出; $W^l\in R^{K\times K}$ 为第 l 层中拉普拉斯矩阵的可学习参数矩阵; σ 为非线性激活函数。图卷积层的初始输入 $O^0=f_{\mathrm{VFE}}\in R^{K\times 128}$, 为特征补偿编码后的体素格特征, 文中方法中采用两层图卷积生成图卷积模块的输出为 $f_{GCN}\in R^{K\times 128}$。

2.5 目标检测框生成网络模块

文中采用 VoxelNet [3] 的 RPN 网络模块, 该模块将图卷积模块的输出特征作为输入最终输出 3D 目标检测框。文中用 Conv2D (Cout, k, s) 表示带有非线性激活函数的 2D 卷积操作, Deconv2D (Cout, k, s) 表示带有非线性激活函数的 2D 反卷积层。其中 Cout 表示卷积输出维度, k 为卷积核尺寸, s 为卷积核移动步伐。

RPN 网络模块由三个子模块构成, 第一个子模块包括卷积核移动步伐为 2 的下采样卷积层 Conv2D (128,3,2),三个卷积层 Conv2D (128,3,1) 和反卷积层 Deconv2D (256,3,1), 第二个子模块包括下采样卷积层 Conv2D (128,3,2) 和四个卷积层 Conv2D (128,3,1) 和 Deconv2D (256,3,2), 第三个子模块包括下采样卷积层 Conv2D (256,3,2), 四个卷积层 Conv2D (256,3,1) 和 Deconv2D (256,3,4)。每个子模块的最后一层反卷积将输出射到同一尺寸进行拼接, 从而实现多尺度特征学习。RPN 网络模块最终通过 Conv2D (2,3,1) 卷积层生成正负样本分类预测 C_{pred}, Conv2D (14,3,1) 卷积层生成三维目标预测框 B_{pred}, 文中的非线性激活函数均采用 ReLU 激活函数, 网络的整体架构如图 3 所示。

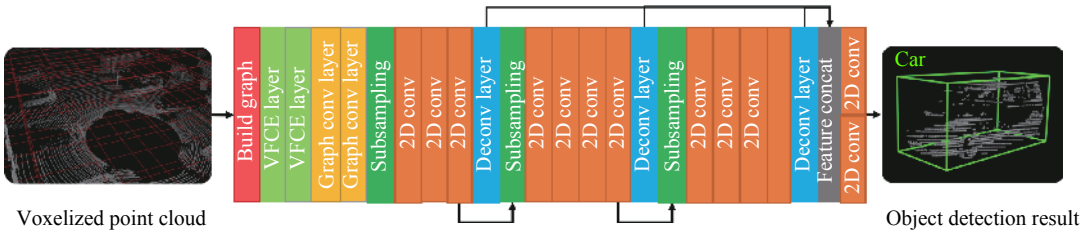


图 3 设计的三维点云目标检测网络架构图

Fig.3 Structure of designed 3D point cloud object detection network

2.6 损失函数的设计

2.6.1 目标检测框的生成

根据 KITTI 数据集的标签参数化真实检测框 $B_g=(x_g,y_g,z_g,w_g,h_g,l_g,r_g)$, 其中检测框中心点$(x_g,y_g,z_g)$, 以检测框中心点度量的长、宽、高表示为$(w_g,h_g,l_g)$, r_g表示 Z 轴为旋转轴的朝向角。

在目标场景范围内, 以每个有效体素格中心点为锚点, 生成 r=0 和 r=90 方向的一对候选检测框 $B_a=(x_a,y_a,z_a,w_a,h_a,l_a,r_a)$, 同样用中心位置$(x_a,y_a,z_a)$, 长宽高$(w_a,h_a,l_a)$以及朝向角$r_a$来表示候选框。

C_{pos}代表B_a与B_g的重叠率大于 0.6 的正样本检测框, C_{neg}代表重叠率小于 0.4 的负样本检测框, 其余候选框视为对网络学习无贡献。网络学习包括由正负样本检测框的目标检测框B_{target}表示为:

$$\begin{aligned} &x_{target}=\frac{x_g-x_a}{d_a},y_{target}=\frac{y_g-y_a}{d_a},\\ &z_{target}=\frac{z_g-z_a}{h_a},w_{target}=\log\left(\frac{w_g}{w_a}\right),\\ &h_{target}=\log\left(\frac{h_g}{h_a}\right),l_{target}=\log\left(\frac{l_g}{l_a}\right),\\ &r_{target}=r_g-r_a \end{aligned} \tag{4}$$

式中: $d_a=\sqrt{w_a{}^2+l_a{}^2}$为检测框的地面对角线, 用来归一化$x_{target}$, y_{target}的误差。

2.6.2 网络预测回归框的总损失

网络输出的预测回归框通过计算回归损失函数, 方向损失函数及正负样本判别损失函数来输出置信度最高的检测框, 其总表达式为:

$$L_{total}=\alpha\cdot L_{cls}+\beta\cdot L_{reg}+\gamma\cdot L_{dir} \tag{5}$$

式中: $\alpha=1$, $\beta=10$, $\gamma=0.2$为损失函数权重分配, 是人为设定的超参数; L_{cls}为网络正负样本判别损失; L_{reg}为目标检测框的回归损失; L_{dir}为方向判别损失。

检测框回归损失函数: 网络中采用 Huber loss 来计算 RPN 网络生成的预测检测框B_{pred}与目标检测框B_{target}的检测误差, 具体可表示为:

$$L_\delta=\begin{cases}\dfrac{1}{2}(L_{reg})^2\delta, |L_{reg}|\leqslant\delta\\ \delta\left(|L_{reg}|-\dfrac{1}{2}\delta\right), \text{otherwise}\end{cases} \tag{6}$$

式中: L_{reg}为计算的回归误差损失; δ 为超参数, 网络中取值 δ=0.1。Huber loss 的优势在于可基于网络前向计算损失的大小调整回归损失的计算方式, 从而避免网络梯度紊乱。检测框的回归损失函数L_{reg}表示为:

$$L_{reg}=H(B_{pred\_cube}-B_{target\_cube}) \tag{7}$$

式中: $B_{pred_cube}=(x_p,y_p,z_p,w_p,h_p,l_p)$为预测检测框, $B_{target_cube}=(x_g,y_g,z_g,w_g,h_g,l_g)$为目标检测框; H 为 Huber Loss。由于场景目标丰富, 对不同目标, 网络采用不同尺寸的检测框预设, 对车辆: w=1.6 m, h=1.56 m, l=3.6 m, 对行人: w=0.6 m, h=1.73 m, l=0.8 m; 对骑行者: w=0.6 m, h=1.73 m, l=1.76 m。

检测框方向判别损失: 当 RPN 网络模块生成的检测框学习B_{target}时, 在遇到大小相等, 方向相反的检测框会无法辨别其方向, 产生较大的方向损失。对此采用正弦函数来计算方向损失, 并添加一个方向判别器来解决目标方向识别问题。若数据标签中的朝向角$r_g>0$将其设为 1, 反之设为−1[5]。目标检测框朝向损失函数L_{dir}, 具体可表示为:

$$L_{dir}=H(\sin(r_{pred\_r}-r_{target\_r})) \tag{8}$$

式中: r_{pred_r}为预测值; r_{target_r}为真实值; sin 为正弦函数。

检测框正负样本判别损失函数: 由于输入点云中的正负样本比例严重失衡, 因此网络采用全局随机旋转和比例缩放等数据增强方法[8], 并采用 Focal loss 分类损失函数来缓和样本比例失衡带来的梯度数值紊乱问题, 具体可表示为:

$$L_{cls}=\begin{cases}-\lambda(1-y')^{\rho}\log(y'), y=1\\ -(1-\lambda)y'^{\rho}\log(1-y'), y=1\end{cases} \tag{9}$$

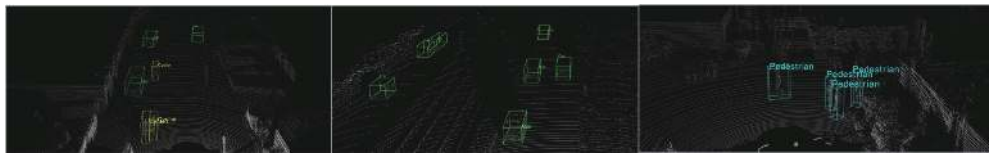
式中：λ为平衡因子；ρ为调整因子；y为分类真实值；y'为分类预测值。实验中取值λ=0.25, ρ=2。

3 实 验

3.1 数据集的准备

文中采用的激光雷达数据为 KITTI 公开数据集[9], 数据集包括 7 481 个训练数据和 7 519 个测试数据。其中点云数据采用 Velodyne HDL-64E 型线列机械旋转式激光雷达用 10 Hz 频率采集, 数据格式为点云三维坐标 (x, y, z) 及反射强度值 r, 而图像采用 CMOS 摄像机用 10 Hz 频率采集, 其水平视场角 90°, 垂直视场角 35°, 图像分辨率为 1 240 pixel×376 pixel。文中网络训练仅用激光雷达点云作为输入, 最终生成目标的 3D 检测结果和 BEV 检测结果, 并通过 KITTI 提供的传感器仿射变换矩阵将点云目标检测框映射在对应图像上, 如图 4(a) 所示,图中绿色框为车辆检测框, 蓝色框为行人检测框, 黄色框为骑行者检测框。

(a) 三维目标检测结果的二维图像映射
(a) 2D image mapping of 3D object detection results



(b) 三维点云目标检测场景示意图
(b) Scene of 3D point cloud object detection

图 4 KITTI 数据集 3D 目标实测结果

Fig.4 3D object detection result of KITTI dataset

3.2 实验环境

文中实验训练, 验证及测试使用的计算机环境为 Ubuntu 16.04 系统, Intel(R) Core™ i5-7300 CPU@ 2.50 GHz, 显卡设备为 NVIDIA 2080 Ti 型号, 整体代码在 Pycharm 编译平台基于 Google 的开源深度学习框架 Tensorflow, 采用 python 3.5 编写。

3.3 实验过程及分析

实验中网络训练将训练数据集划分为 3 712 个训练样本和 3 769 个验证样本。网络完整训练需循环遍历训练集 80 次, 且每遍历十次保存整个网络模型参数并在验证集进行检测生成模型实时检测精度, 最终的验证集检测精度如表 1、表 2 所示。

在 KITTI 公榜 (www.cvlibs.net)上的测试集检测结果的在线评估如表 3 所示, 表中每项检测任务的最高检测数值加粗给出。

实验选取基准网络 VoxelNet[3] 和基于图像和点

表 1 KITTI 验证集上的点云 3D 目标检测平均精度对比

Tab.1 Comparison of average precision of point cloud 3D object detection on KITTI validation set

Networks	Inference time/s	Sensors		Car			Pedestrian			Cyclist		
		LiDAR	Image	Moderate	Hard	Easy	Moderate	Hard	Easy	Moderate	Hard	Easy
MV3D [10]	0.36	√	√	62.35%	55.12%	71.09%	N/A	N/A	N/A	N/A	N/A	N/A
MV3D (Lidar)[10]	0.24	√	×	52.73%	51.31%	66.77%	N/A	N/A	N/A	N/A	N/A	N/A
AVOD [11]	0.08	√	√	65.78%	58.38%	73.59%	31.51%	26.98%	38.28%	44.90%	38.80%	60.11%
VoxelNet [3]	0.31	√	×	65.46%	62.85%	81.97%	53.42%	48.87%	57.86%	47.65%	45.11%	67.17%
Point-GNN[6]	0.6	√	×	79.47%	72.29%	88.33%	43.77%	40.14%	51.92%	63.48%	57.08%	78.60%
VGCN (Ours)	0.09	√	×	79.21%	78.58%	89.25%	53.28%	48.74	60.90%	71.82%	68.19%	85.89%

表 2 KITTI 验证集上的鸟瞰图目标检测平均精度对比

Tab.2 Comparison of average precision of BEV object detection on KITTI validation set

Networks	Inference time/s	Sensors		Car			Pedestrian			Cyclist		
		LiDAR	Image	Moderate	Hard	Easy	Moderate	Hard	Easy	Moderate	Hard	Easy
MV3D [10]	0.36	√	√	76.90%	68.49%	86.02%	N/A	N/A	N/A	N/A	N/A	N/A
MV3D (Lidar) [10]	0.24	√	×	77.00%	68.94%	85.82%	N/A	N/A	N/A	N/A	N/A	N/A
AVOD [11]	0.08	√	√	85.44%	77.73%	86.60%	35.24%	33.97%	42.52%	47.74%	46.55%	63.66%
VoxelNet [3]	0.31	√	×	84.81%	78.57%	89.60%	61.05%	56.98%	65.95%	52.18%	50.49%	74.41%
Point-GNN[6]	0.6	√	×	89.17%	83.90%	93.11%	43.77%	40.14%	51.92%	67.28%	59.67%	81.17%
VGCN (Ours)	0.09	√	×	87.90%	87.33%	90.23%	57.30%	52.72%	64.22%	74.94%	71.57%	87.26%

表 3 KITTI 测试集上的 3D 和鸟瞰图目标测试结果

Tab.3 3D and BEV object detection on KITTI test set

Benchmarks	Moderate	Easy	Hard
Car(3D)	77.65%	84.47%	73.36%
Car(BEV)	87.16%	90.67%	82.98%
Cyclist(3D)	62.36%	78.47%	55.88%
Cyclist(BEV)	67.04%	81.50%	59.45%
Pedestrian(3D)	37.60%	45.28%	34.96%
Pedestrian(BEV)	42.33%	50.02%	40.05%

云数据融合的网络 MV3D[10], AVOD[11] 以及基于图神经网络的 Point-GNN[6] 作为对比网络。在 KITTI 验证集和测试集的上进行 3D 目标检测和 BEV 目标检测。其中数据集分为: 简易, 中等和困难三个等级。检测目标分别为: 车辆, 行人和骑行者。点云数据中的目标在空间离散分布且稀疏, 激光雷达扫描到的目标点云为目标表面点云, 多数为残缺, 如图 4(b) 所示。且相对整副场景占比太小, 其占比约 0.3%~0.8%, 需要有效利用点云拓扑以及语义信息。因此, 文中通过 VFCE 编码以及图卷积神经网络的引入, 有效利用了点云的空间信息和拓扑连接关系。

3.3.1 改变卷积模块的层数进行消融实验

为验证图卷积模块对基准网络的检测性能带来的提升, 本节进行图卷积层的消融实验, 实验选取 3D 车辆目标检测来做验证。通过将基准网络 VoxelNet[3] 中的 3D 卷积模块用文中所提的图卷积模块替代, 改动图卷积层数来验证不同数量的图卷积层的检测精度, 如表 4 所示。实验表明: 当采用两层图卷积层时, 网络检测精度达到了最佳, 提升基准网络的检测精度高达 13.75%。

表 4 图卷积层数对车辆目标检测的影响

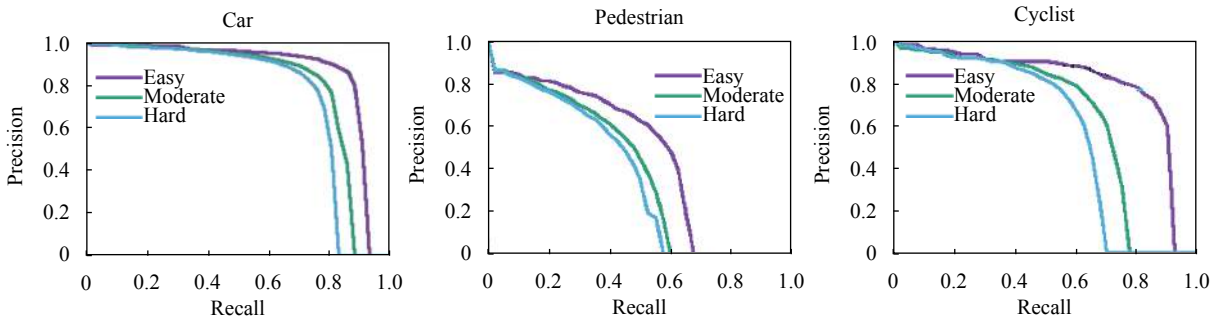
Tab.4 Effect of the number of graph convolutional layers on vehicle object detection

Networks	Moderate	Hard	Easy
VoxelNet [3]	65.46%	62.85%	81.97%
1-Graph conv	76.32%	75.18%	85.85%
2-Graph convs	79.21%	78.58%	89.25%
4-Graph convs	80.06%	75.72%	86.73%

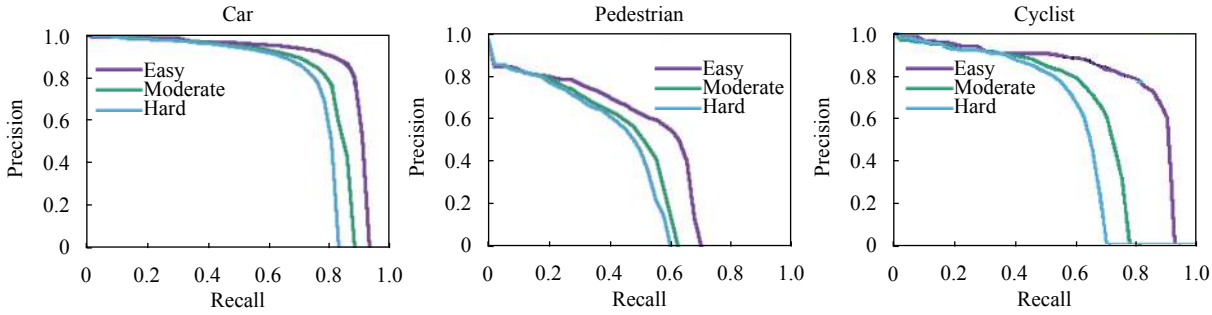
3.3.2 体素格补偿编码的检测性能对比

体素格补偿编码不但能补偿体素划分产生的信息损失, 同时使该方法的目标检测回召率有了一定提升。回召率 (Recall) 是样本中正例被预测正确的比例, 在回召率一定的情况下, 检测精度 (Precision) 越高表示算法检测能力越好, 一般采用 P-R 曲线来评估算法的检测性能。通过实验对比观察, 相比 VoxelNet[3] 的 VFE 编码, VFCE 编码方法提升了算法在车辆, 行人和骑行者目标检测的回召率和检测精度, 见图 5。

如表 1、表 2 所示, 文中提出的目标检测算法 VGCN 在 KITTI 验证集上的测试结果, 结果显示相比基准网络 VoxelNet[3], 基于数据融合方法的 MV3D[10], AVOD[11] 以及基于图神经网络的 Point-GNN[6] 方法, 该方法在三个目标检测任务中的检测速度和精度上均有了一定的提升。网络在检测识别行人, 骑行者等小尺寸目标的精度最高提升了 24.17%, 车辆 3D 检测和鸟瞰图检测任务上最高提升了 13.75% 检测精度。如表 3 所示的 KITTI 测试集测试结果表明: 该方法在各项检测任务上相比对比网络均有了一定的提升, 验证了基于图卷积神经网络的三维目标检测算法的点云目标检测性能。



(a) 使用 VFE 编码层
(a) Using VFE encoding layer



(b) 使用 VFCE 编码层
(b) Using VFCE encoding layer

图 5 3D 目标检测 P-R 曲线对比

Fig.5 Comparison of P-R curve of 3D object detection

4 结 论

文中研究了基于体素划分的图卷积神经网络在点云目标检测中的应用，提出一种基于体素化图卷积网络的三维点云目标检测算法。针对三维点云数据的非规则排布和稀疏性，引入了基于体素划分的图表示方法，有效提高了目标检测网络的数据表征和学习能力。通过设计体素格补偿编码层以及采用图卷积模块替代传统的 3D 卷积层，消除了网络的计算冗余性，提高了体素格节点之间的相关性学习，从而提升了网络算法的目标检测性能。

在公开数据集 KITTI 的检测实验表明，与已有的目标检测算法相比，通过引入基于拉普拉斯矩阵的图卷积模块有效提升了网络的综合检测精度，在 3D 目标检测和鸟瞰图目标检测任务上进一步优化了基准目标检测算法的性能。但相比图像数据，点云数据稀疏，低分辨率特性使其在捕捉行人，骑行者等扫描到的点数少，尺寸小的目标物时会产生漏检情况。因此，后续工作将进一步提升网络对点云语义信息的学习能力，提升目标场景的全覆盖检测能力和检测精度。

参考文献：

[1] Zhang Nan, Sun Jianfeng, Jiang Peng, et al. Pose estimation algorithms for lidar scene based on point normal vector [J]. ***Infrared and Laser Engineering***, 2020, 49(1): 0105004. (in Chinese)

[2] Xia Xianzhao, Zhu Shixian, Zhou Yiyao, et al. LiDAR K-means clustering algorithm based on threshold [J]. ***Journal of Beijing University of Aeronautics and Astronautics***, 2020, 46(1): 115-121. (in Chinese)

[3] Zhou Y, Tuzel O. Voxelnet: End-to-end learning for point cloud based3d object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4490-4499.

[4] Yan Yan, Mao Yuxing, Bo Li. Second: Sparsely embedded convolutional detection [J]. ***Sensors***, 2018, 18 (10) : 3337.

[5] Qi C R, Yi L, Su H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space[C]// Advances in Neural Information Processing Systems, 2017: 5099–5108.

[6] Shi Weijing, Rajkumar Raj. Point-gnn: Graph neural network for 3d object detection in a point cloud[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern

Recognition, 2020: 1711-1719.

[7] Kipe Thomas N, Welling Max. Semi-supervised classification with graph convolutional networks [J]. ***arXiv preprint arXiv***, 2016, 1609: 02907.

[8] Xue Shan, Zhang Zhen, Lu Qiongying, et al. Image recognition method of anti UAV system based on convolutional neural network [J]. ***Infrared and Laser Engineering***, 2020, 49(7): 20200154. (in Chinese)

[9] 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, 2012: 3354–3361.

[10] Chen Xiaozhi, Ma Huimin, Wan Ji, et al. Multi-view 3d object detection network for autonomous driving[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1907-1915.

[11] Yan Yan, Mao Yuxing, Bo Li. Second: Sparsely embedded convolutional detection [J]. ***Sensors***, 2018, 18(10): 3337.