

DOI:10.19651/j.cnki.emt.2208783

融合多注意力机制与 PointRCNN 的 三维点云目标检测*

郑美琳 高建瓴

(贵州大学大数据与信息工程学院 贵阳 550025)

摘要: 针对三维不规则的点云格式和密度不均匀的问题,提出了一种融合多注意力机制与 PointRCNN 网络用于三维点云目标检测。本实验主要对 PointRCNN 两阶段网络分别进行改进,首先,把通道注意力与空间注意力机制串行通过调节输入到第一阶段各网络层的分布,批量归一化进一步快速识别三维特征;其次,引入交叉位置注意力机制到第二阶段网络为了避免交叉路径出现位置偏差,从而进一步精细化三维目标位置以进行特征提取。在 KITTI 数据集上实验结果表明,相比于 PointRCNN 检测网络,改进的网络在小汽车和行人测试上平均均值精度(mAP)分别提高了 1.2%、1.9%。因此改进的方法在解决了点云格式不规则和密度不均匀问题的同时还保证了检测精度。

关键词: PointRCNN;三维点云;目标检测;注意力机制;交叉位置

中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 510.4

3D point cloud target detection based on attention mechanism and pointrcnn

Zheng Meilin Gao Jianling

(College of Big Data & Information Engineering, Guizhou University, Guiyang 550025, China)

Abstract: Aiming at the problem of 3D irregular point cloud format and uneven density, a fusion of multi attention mechanism and pointrcnn network is proposed for 3D point cloud target detection. This experiment mainly improves the pointrcnn two-stage network respectively. Firstly, the channel attention and spatial attention mechanism are serially input to the distribution of each network layer in the first stage by adjusting and normalizing in batch to further quickly identify three-dimensional features; Secondly, the cross position attention mechanism is introduced into the second stage network to avoid the position deviation of the cross path, so as to further refine the three-dimensional target position for feature extraction. The experimental results on Kitti data set show that compared with pointrcnn detection network, the improved network improves the average mean accuracy (map) of car and pedestrian tests by 1.2% and 1.9% respectively. Therefore, the improved method not only solves the problems of irregular point cloud format and uneven density, but also ensures the detection accuracy.

Keywords: PointRCNN;3D point cloud;target detection;attention mechanism;cross position

0 引言

深度学习在二维计算机视觉任务上取得了显著进展,包括目标检测^[1]和实例分割^[2]等。除了对二维场景的理解之外,三维对象检测对于许多现实世界的应用来说是至关重要和不可或缺的。三维目标检测分为基于点云或者体素^[3]特征学习方法,基于点云特征网络的根基是 PointNet^[4]网络,其首次成功的解决了点云无序性的问题,

为后面的网络奠定了基础。文献[5]中将点云投影到鸟瞰地图上,并在这些地图上应用二维 CNN 进行三维检测。

Shi 等^[6]提出了点云神经网络,通过从三维标注框中获取分割标签,以自下而上的方式把原始点云直接生成三维建议,基于点云的网络如 PointRCNN^[6]和 STD^[7]网络,但在复杂环境下,网络均存在精度不高的问题;而基于体素的网络是具有三维稀疏卷积 Voxel CNN^[8],文献[9]中将点云

收稿日期:2022-01-06

* 基金项目:国家自然科学基金(62062021,61872034)、贵州省科学技术基金(黔科合基础[2020]1Y254)、贵州省自然科学基金(黔科合基础[2019]1064)项目资助

划分为三维体素,由三维 CNN 进行处理,引入三维稀疏卷积为了能够较为高效的处理三维体素,但在输入点云分布不均时难以快速把点云转化为体素从而会影响提取特征速度;由于 PC-CNN^[10], PointFusion^[11], F-PointNet^[12], Frustum ConvNet^[13] 检测器需要点云和 RGB 图像同时输入,但其突出劣势表现为整个流程对 2D 的检测效果比较依赖,且无法解决遮挡问题。因此,越来越多新提出的方法倾向于仅使用点云作为输入,例如 PointRCNN^[6]、Part-A²Net^[14]、STD^[7]、PV-RCNN^[15]、3DSSD^[16]、SASSD^[17],基于以上背景,选择其中一种(PointRCNN)研究如何处理点云密度不均匀的问题,如何在不同点云的分布差异时提高提取特征的精度。

近年来,注意力机制发展迅速,其广泛用于自动驾驶、机器人传感和图像编辑中。文献[18]中通过卷积运算来混合交叉通道注意力^[19]和空间注意力^[20]信息从而提取特征;对于交叉位置注意^[21]模块,融入了自我注意机制来捕获特征地图任意两个位置之间的空间依赖关系。为了完成点云分布不规则时三维场景分割并且区分一些容易混淆的类别,本文提出了把通道注意力与空间注意力模型串行引入到 PointRCNN 的第一阶段可以快速识别特征,在第二阶段加入通道注意力机制来精准检测到特征位置,该方法能在点云密度不均匀时,更快速的识别三维特征并进行特征提取,使融入多种注意力机制的网络比原有的网络性能上有了很大的提高。

1 改进 PointRCNN 模型框架

1.1 PointRCNN 框架

PointRCNN 框架(如图 1)可以直接对三维点云进行操作,并达到了较为精确的三维检测性能。由于 PointRCNN 网络是一种新的基于自底向上点云的三维包围框建议生成算法,其通过将点云分割成前景对象和背景,以生成少量高质量的三维建议。从分割中学习到的点表示不仅有利于建议生成,而且有助于后期的框细化。所提出的标准三维包围框细化利用了从第一阶段生成的高层包围框建议,并学习在标准坐标中预测框坐标的精度,基于预测框的损失具有鲁棒性。它同时可以通过仅使用点云作为输入,从而达到提出的三维检测框架点神经网络以较显著的优势优于一些方法,但同时存在提取特征速度缓慢和精度不高的问题。

1.2 改进 PointRCNN 的框架

由于点云的不规则性和不均匀密度的问题,PointRCNN 对原始点云的处理速度缓慢并且精度不高,本文提出了融入注意力机制策略的框架,所提出的框架包括两个阶段:第一阶段用自底向上的方案生成三维包围框建议,引入通道注意力与空间注意力串行的网络可以利用三维包围框生成背景分割掩膜和分割前景点,同时从分割的点中生成比原来更多的包围框建议。这种策略避免了在整个三维空间中使用大量三维锚框并且节省了计算量。在

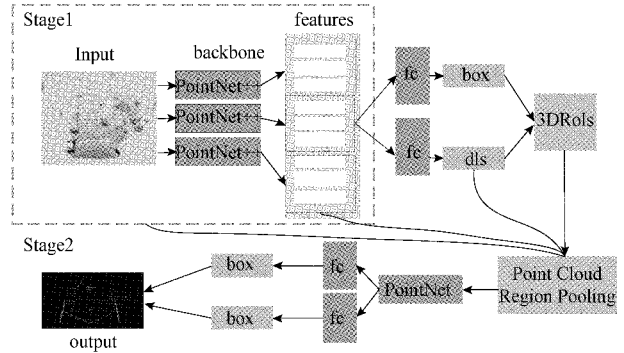


图 1 PointRCNN 的框架图

PointRCNN 的第二阶段引入交叉位置注意力机制进行了规范的三维框细化,并且在生成三维建议之后,能更精细化三维框位置以提取阶段一的特征。与直接估计全局框坐标以及把合并的三维点转换到规范坐标上不同,首先把合并的点特征以及来自第一阶段的分割掩膜相结合,然后把结合后的特征位置精细化后,再被用来对相对坐标细化。为了学习更有效的坐标细化,改进的 PointRCNN 还提出了基于全面元的三维框回归损失。此框架如图 2 所示。

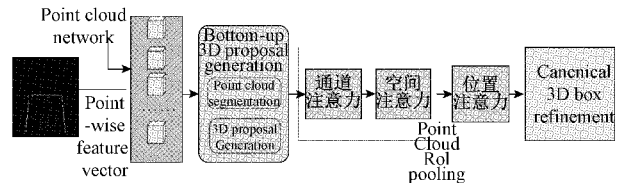


图 2 改进 PointRCNN 框架图

PointRCNN 中三维包围框的回归头被用来生成自底向上的三维建议和前景点分割。在训练过程中,只需要包围框的回归头从前景点回归三维框的位置。把三维边界框在激光雷达坐标系中表示为 $(x, y, z, h, w, l, \theta)$, 其中 (x, y, z) 是对象中心位置, (h, w, l) 是对象大小, θ 是从鸟瞰角度的对象方向。为了约束生成的三维框建议,用基于面元的回归损失来估计三维对象的边界框。为了估计物体的中心位置,将每个前景点的周围区域分成一系列沿 X 轴和 Z 轴的离散面元。具体来说,首先为当前前景点的每个 X 轴和 Z 轴设置一个搜索范围 S, 并且每个一维搜索范围被分成统一长度 δ , 用来表示 X-Z 平面上的不同对象中心 (X, Z) 。对于 X 轴和 Z 轴,使用具有交叉熵损失的基于面元的分类,而不是具有平滑 L1 损失的直接回归,使得中心定位更精确和稳健,从而更精准定位特征位置。因此,定位目标特征可以表述如下:

$$bin_x^{(p)} = \left\lceil \frac{z^p + z^{(p)} + S}{\delta} \right\rceil,$$

$$res_x^{(p)} = \frac{1}{C} \left(u^p - u^{(p)} + S - \left(bin_x^{(p)} \cdot \delta + \frac{\delta}{2} \right) \right)$$

$$res_y^{(p)} = y^p - y^{(p)}$$

(1)

2 注意力机制

2.1 通道注意力串行空间注意力机制

对于通道注意力模型,利用通道映射之间的相互依赖性,以强调相互依赖的特征映射,而空间注意力是执行特征的显式空间变换,并以端到端的学习方式来提高精度。在学习给定输入三维点云图像的特征映射之间的通道关系时,通道注意力机制中的“挤压和激励”(SE)模块通过建模通道之间的相互依赖性,自适应地重新校准通道特征响应。对于任何给定的变换 F_r , 将输入 X 映射到特征映射 U , 其中 $U \in R^{H \times W \times C}$, H, W, C 分别表示模块的高度、宽度和通道数,特征 U 首先通过一个挤压操作,该操作通过聚合其三维空间维度上的特征映射($H \times W$)来产生一个信道描述符。这个描述符用来识别信道特征响应的全局分布,允许来自网络的全局接受域的信息被网络的所有层使用。聚合之后通过一个激励操作,它采取了一个简单的自门机制的形式,以嵌入作为输入,并产生每个通道调制权值的集合。这些权重应用到特征图 U 以生成 SE 块的输出,该块可以直接输入到网络的后续层。网络将 F_r 作为卷积运算符,并使用 $V = [v_1, v_2, \dots, v_C]$ 表示已学习的过滤器内核集,其中 v_c 表示第 C 个过滤器的参数。然后我们可以把输入写为 $U = [u_1, u_2, \dots, u_C]$,

$$u_c = v_c * X = \sum_{s=1}^C V_c^s * x^s \quad (2)$$

在这里 $*$ 表示卷积, $V_c = [V_c^1, V_c^2, \dots, V_c^C]$, $X = [x^1, x^2, \dots, x^C]$ 和 $u_c \in R^{H \times W}$ 。 v_c 是一个空间核,代表 v_c 的单一通道并且作用于 X 的相应通道。

为了学习特征地图中的空间相关性,并且在三维空间地图上提高远视能力,由此引入了空间注意力机制。其中的空间转换器允许在网络中对数据进行三维空间的操作。该模块可插入现有的卷积结构中,使神经网络能够根据特征提取本身主动地对特征图进行空间变换,而无需对优化过程进行任何额外的训练监督或修改。空间转换器结构分为定位网络、电网发电机和采样器三部分:按照计算顺序,第一部分的定位网络是为了获取输入三维点云特征图,并通过多个隐藏层输出应用到特征地图的空间变换参数;然后,使用预测的变换参数创建采样网格,输入映射应在网格中采样以生成变换后的输出;最后,将特征映射和采样网格作为采样器的输入,生成从网格点处的输入采样和输出映射。这三个组件的组合构成了一个空间变换器。

当通道注意力与空间注意力模型串行时,需设定一个中间特征作为 $F \in R^{C \times H \times W}$ 输入, CBAM (convolutional block attention module)^[22] 依次推断出一个一维通道注意力 $M_c \in R^{C \times 1 \times 1}$ 和二维空间注意力 $M_s \in R^{1 \times H \times W}$ 。注意力机制的总过程可概括为:

$$\begin{aligned} F' &= M_c(F) \otimes F \\ F'' &= M_s(F') \otimes F' \end{aligned} \quad (3)$$

其中, \otimes 表示元素级乘法。 F'' 是最终的精炼输出。在 CBAM 网络中,首先通过使用平均池化和最大池化操作来收集特征地图的三维空间信息,从而生成两种不同的三维空间上下文描述符: F_{avg}^c 和 F_{max}^c , 它们分别表示平均合并的特征和最大合并的特征。然后这两个描述符被转发到由一层的多层感知器 (MLP) 组成的共享网络来生成通道注意力图 $M_c \in R^{C \times 1 \times 1}$ 。为了减少参数的数量,将隐藏的激活大小设置为 $R^{C/r \times 1 \times 1}$, 其中 r 为还原比。通道注意力串行空间注意力模型如图 3 所示,它的计算方法为:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))) \quad (4)$$

其中, σ 为 S 型函数, $W_0 \in R^{C/r \times C}$ 和 $W_1 \in R^{C \times C/r}$ 。输入的 MLP 权值 W_0 和 W_1 都是共享的, ReLU 激活函数后面是 W_0 。通过使用两个池操作聚合特征映射的通道信息来生成两个二维映射: $F_{avg}^s \in R^{1 \times H \times W}$, $F_{max}^s \in R^{1 \times H \times W}$ 。每个表示通道中的平均池特征和最大池特征。然后,通过一个标准卷积层将它们连接并卷积,生成空间注意力图。空间注意力的计算方式如下:

$$M_s(F) = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) = \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s])) \quad (5)$$

σ 表示 S 型函数, $f^{7 \times 7}$ 表示滤波器大小为 7×7 的卷积运算, \otimes 表示矩阵乘法。

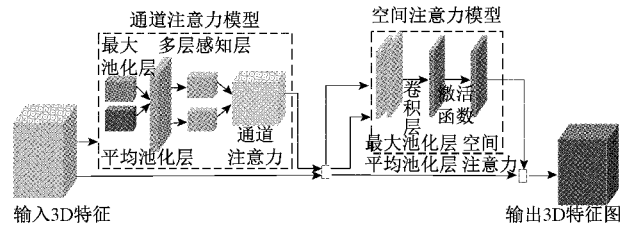


图3 空间注意力串行通道注意力

2.2 交叉位置注意力机制

为了加强目标位置的相关性从而引入了交叉位置注意力机制,其中交叉位置注意力中的交叉注意模块收集了交叉路径上所有像素的上下文信息,局部特征首先通过一个纵横交错的注意模块来收集水平和垂直方向上的上下文信息。然后,通过将生成的特征映射从第一交叉注意模块馈送到另一个交叉注意模块,从交叉路径获得的附加上下文信息最终使每个像素能够捕获完整的图像相关性,交叉位置的注意层是同一图像中所有物体的位置信号。因此,交叉位置的注意矩阵 a 如下:

$$a_{jk} = \frac{\exp(\lambda p_j^T p_k)}{\sum_k \exp(\lambda p_j^T p_k)} \quad (6)$$

其中, λ 是一个可训练的变量,而不是一个超参数的。注意权重表示位置 p_j 和位置 p_k 之间的接近性,并测量了这些位置之间的影响和相互作用。对于在 p_j 位置的项目,通过从其他位置表示的注意加权和,推导出参与项目表示

\tilde{x}_j 和参与位置表示 \tilde{p}_j :

$$\begin{aligned} \tilde{x}_j &= \sum_{k=1}^m a_{jk} x_k \\ \tilde{p}_j &= \sum_{k=1}^m a_{jk} p_k \end{aligned} \quad (7)$$

这些参与位置表示聚集了来自其他位置的信号,并根据它们与位置 p_j 的距离进行加权,它们同时也反映了相邻位置的相关性。

该模型的时间和空间复杂度均为 $(H \times W) \times (H \times W)$, 其中 $H \times W$ 表示输入三维特征图的空间维度。使用两个连续的交叉点模块,其中每个模块仅具有稀疏连接 $((H+W)-1)$ 对应于要素地图中的每个位置。“交叉注意”模块在水平和垂直方向上聚合上下文信息。通过连续堆叠两个纵横交错的注意力模块,它可以从所有像素中收集上下文信息。上述分解策略大大降低了时间和空间上的复杂度,复杂度从 $((H \times W) \times (H \times W))$ 减到 $((H \times W) \times (H \times W) - 1)$ 。

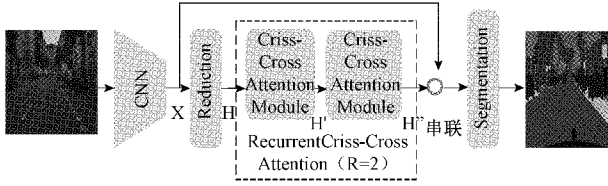


图 4 交叉位置注意力

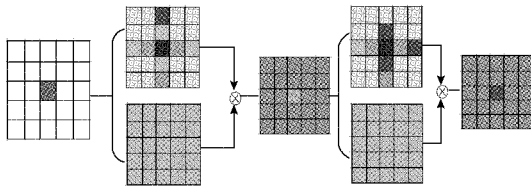


图 5 交叉注意力模块

2.3 损失函数

对于三维建议的生成和细化,使用平滑损失 L1 来回归一化框的参数,把具有不同训练损失项的总三维包围框回归损失函数 L_{reg} 表示如下:

$$\begin{aligned} L_{bin}^{(p)} &= \sum_{u \in \{x, z, \theta\}} (F_{cls}(\widehat{bin}_u^{(p)}, bin_u^{(p)}) + F_{reg}(\widehat{res}_u^{(p)}, res_u^{(p)})), \\ L_{res}^{(p)} &= \sum_{v \in \{y, h, w, l\}} F_{reg}(\widehat{res}_v^{(p)}, res_v^{(p)}), \\ L_{reg} &= \frac{1}{N_{pos}} \sum_{p \in pos} (L_{bin}^{(p)} + L_{res}^{(p)}) \end{aligned} \quad (8)$$

N_{pos} 表示前景点的数量, $\widehat{bin}_u^{(p)}$ 和 $\widehat{res}_u^{(p)}$ 表示预测 bin 的分配和前景点 p 的残差, $bin_u^{(p)}$ 和 $res_u^{(p)}$ 是(8)式中的地面真实目标, F_{cls} 表示交叉熵分类损失, F_{reg} 表示平滑 L1 损失。为了移除多余的特征,把基于鸟瞰的定向 IoU 进行非最大抑制(NMS),以生成少量高质量的特征。对于大型的室外场景,前景点的数量通常比背景点的数量少得多,因此在第一阶段中用焦点损失 L_{focal} 来处理前景点和背景点不平衡的问题,焦点损失的表达式如下:

$$L_{focal}(p_i) = -\alpha_i(1-p_i)^{\gamma} \log(p_i) \quad (9)$$

$$p_i = \begin{cases} p, & \text{前景点} \\ 1-p, & \text{其他} \end{cases}$$

第二阶段中的交叉熵损失用于分类,基于 bin 的损失用于回归,因此第二阶段子网络的总损耗可表述为式(10):

$$L_{refine} = \frac{1}{\|B\|} \sum_{i \in B} F_{cls}(prob_i, label_i) + \frac{1}{\|B_{pos}\|} \sum_{i \in B_{pos}} (\tilde{L}_{bin}^{(i)} + \tilde{L}_{res}^{(i)}) \quad (10)$$

其中, B 是第一阶段的三维框的集合, B_{pos} 表示为存储回归的正建议,概率单位是相应标签的评估置信度, F_{cls} 是监督预测置信度的交叉熵损失,在方程中 $\tilde{L}_{bin}^{(i)}$ 和 $\tilde{L}_{res}^{(i)}$ 是类似于 $L_{bin}^{(p)}$ 和 $L_{res}^{(p)}$ 。式(10)是采用上述 \tilde{b}_i 和 \tilde{b}_i^{gt} 计算的新目标。最终应用鸟瞰 IoU 阈值为 0.01 的定向 NMS 来移除重叠的边界框,并为检测到的对象生成三维边界框。

3 相关实验

3.1 实验数据集

PointRCNN 在 KITTI 数据集具有挑战性的三维目标检测基准上进行评估^[23]。KITTI 的三维对象检测基准包含 7 481 个训练样本(train split)和 7 518 个测试样本(test split)。遵循^[24]中提到的常用的 train/val 分割,将训练样本分为 train 分割(3 712 个样本)和 val 分割(3 769 个样本),其中在列车分割时,每个范围内每个点云的平均点数如图 6 所示。在 KITTI 数据集的值分割和测试分割上,将点神经网络与一部分三维对象检测方法进行了比较。所有模型都在列车、行人分割上进行训练,并在 val 分割上进行评估。

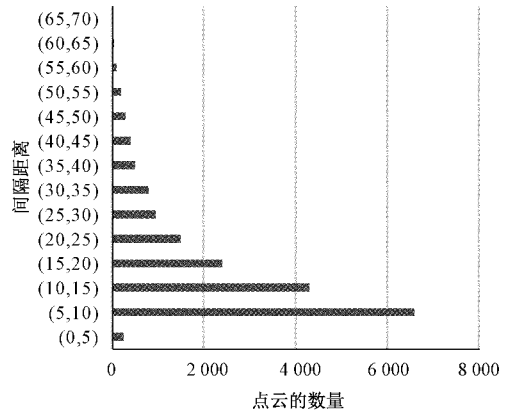


图 6 在 KITTI 数据集的列车分割上,每个范围内每个点云的平均点数

3.2 实验细节

对于训练集中的每个三维点云场景,从每个场景中抽取 16,384 个点作为输入。对于点数少于 16,384 的场景,随机重复点数,得到 16,384 点。第一阶段的子网络遵循^[25]的网络结构,其中使用具有多尺度分组的 4 个集合抽

象层将点二次抽样成大小为 4 096、1 024、256、64 的组。然后使用 4 个特征传播层来获得用于分割和建议生成的特征向量。对于精细化子网络,从每个汇集区域随机抽样 512 个点作为精细化子网络的输入。使用具有单尺度分组^[24](具有组大小 128、32、1)的三个集合抽象层来生成用于对象置信度分类和建议位置细化的单个特征向量,汽车类别的训练细节可以在 KITTI 数据集中有大多数样本,行人的超参数可以从发布的代码中找到。对于第一阶段的子网络,三维地面真值框内的所有点都被视为前景点,其他点被视为背景点。在训练过程中,忽略物体边界附近的背景点,通过在物体的每一侧将三维地面真值框放大 0.2 m 来进行鲁棒分割,因为三维地面真值框可能有小的变化。对于基于框生成的建议,超参数被设置为搜索范围 $S=3$ m,框的大小 $\delta=0.5$ m,定向框的数量 $n=12$ 。为了训练第二阶段的子网络,随机地用小的变化来增加三维框、增加框的多样性。

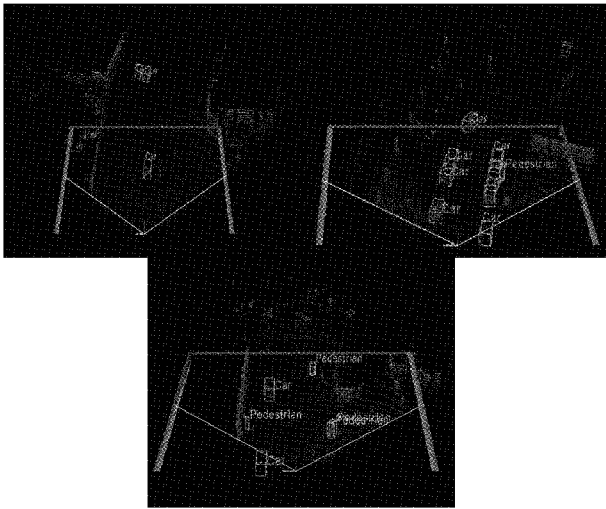


图 7 改进的网络可视化图

3.3 实验结果

在 KITTI 测试服务器的三维检测基准上评估了改进的方法,结果显示在表 1 中。对于汽车和行人的三维点云目标检测,改进的方法优于以前的 PointRCNN 网络。在小汽车测试集上,与以往的 PointRCNN 网络相比,平均均值精度(mAP)从 77.17 增加到 76.26,提高了 1.2%;在行人检测方面,与以往的方法相比,改进的方法取得了更好的结果,提高了 1.9%。然而,它的性能比具有多个传感器的方法稍差。这是由于改进的方法仅使用稀疏点云作为输入,但是行人具有较小的尺寸,并且图像可以捕捉比点云更多的行人细节,有助于三维检测。对于最重要的汽车类别,如表 2 所示,改进的方法优于以前的方法,在 val 分割集上有很大的优势。分别在难度易、难、适中的情况下,改进的方法比以前的 PointRCNN 网络分别提高了 0.81%,1.12%,1.94%,证明了所改进的网络的有效性。

表 1 KITTI 测试集上的性能比较,评价指标是平均精度(AP)

检测网络	Car(IoU=0.7)			Pedestrian(IoU=0.5)		
	Easy	Moderate	Hard	Easy	Moderate	Hard
VoxelNet ^[23]	77.21	64.89	57.38	39.06	33.25	31.13
Second ^[11]	83.06	73.25	66.03	50.63	42.07	36.77
PointRCNN ^[6]	85.45	75.17	68.16	49.11	41.42	38.19
Ours	86.33	76.12	69.07	49.89	42.21	39.06

表 2 在 Kitti val 分割集的汽车类别上性能对比

检测网络	AP(IoU=0.7)		
	Easy	Moderate	Hard
MV3D	70.75	62.34	56.18
VoxelNet	81.26	65.05	62.36
Second	86.96	76.04	68.76
PointRCNN	87.46	76.63	76.04
Ours	88.17	78.12	76.89

3.4 消融实验结果及分析

通过消融实验的结果,分析了所提出的模型的性能。首先,对不同注意力机制在 PointRCNN 的有效性进行了验证,由于在空间注意力机制串行通道注意力时参数量大且空间利用率低,而交叉位置注意力机制搜集信息的速度缓慢,故分别融入 PointRCNN 中精度不高,从而把多种注意力机制融入到 PointRCNN 得到最终结果,结果表明三种不同的融入方式 mAP 分别为 80.71,80.73,81.06,相比两种单独融入方式分别高出 0.43%,0.41%,由此证明了加入多种注意力机制模型的有效性。

表 3 PointRCNN 与不同的注意力机制性能比较

检测网络	AP(IoU=0.7)		
	Easy	Moderate	Hard
PointRCNN+channel&spatial ^[22]	87.32	77.85	76.95
PointRCNN+cross-position ^[21]	87.44	77.96	76.78
PointRCNN+channel&spatial+cross	88.17	78.12	76.89

4 结 论

本文提出了融合多注意力机制与 PointRCNN 网络的三维点云目标检测方法,针对不规则点云的无序性和检测精度不高的问题,把空间注意力机制与通道注意力机制串行用于 PointRCNN 网络第一阶段来提升特征识别的速度;引入位置注意力机制在第二阶段来精准定位以及提取三维特征。采用 KITTI 数据集进行验证,改进的网络在检测精度和速度上比 PointRCNN 网络有很大的优势。

参考文献

- [1] GIRSHICK R. Fast r-cnn[C]. IEEE International Conference on Computer Vision, 2015: 1440-1448.

- [2] DAI J, HE K, SUN J. Instance-aware semantic segmentation via multi-task network cascades [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 3150-3158.
- [3] MATURANA D, SCHERER S. Voxnet: A 3d convolutional neural network for real-time object recognition [C]. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2015: 922-928.
- [4] QI C R, SU H, MO K, et al. Pointnet: Deep learning on point sets for 3d classification and segmentation [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 652-660.
- [5] YANG B, LUO W, URTASUN R. Pixor: Real-time 3d object detection from point clouds [C]. IEEE conference on Computer Vision and Pattern Recognition, 2018: 7652-7660.
- [6] SHI S, WANG X, LI H. Pointcnn: 3d object proposal generation and detection from point cloud [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 770-779.
- [7] YANG Z, SUN Y, LIU S, et al. Std: Sparse-to-dense 3d object detector for point cloud [C]. IEEE/CVF International Conference on Computer Vision, 2019: 1951-1960.
- [8] GRAHAM B, ENGELCKE M, VAN D M L. 3d semantic segmentation with submanifold sparse convolutional networks [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 9224-9232.
- [9] ZHOU Y, TUZEL O. Voxnet: End-to-end learning for point cloud based 3d object detection [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4490-4499.
- [10] JIANG R, MEI S. Polar coordinate convolutional neural network: From rotation-invariance to translation-invariance [C]. 2019 IEEE International Conference on Image Processing (ICIP), 2019: 355-359.
- [11] XU D, ANGUELOV D, JAIN A. Pointfusion: Deep sensor fusion for 3d bounding box estimation [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 244-253.
- [12] CAO P, CHEN H, ZHANG Y, et al. Multi-view frustum pointnet for object detection in autonomous driving [C]. IEEE International Conference on Image Processing (ICIP), 2019: 3896-3899.
- [13] WANG Z, JIA K. Frustum convnet: Sliding frustums to aggregate local point-wise features for amodal 3d object detection [C]. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2019: 1742-1749.
- [14] SHI S, WANG Z, WANG X, et al. Part-a² net: 3d part-aware and aggregation neural network for object detection from point cloud [J]. Conference on Computer Vision and Pattern Recognition, 2019, DOI: 10.48550/arXiv.1907.03670.
- [15] SHI S, GUO C, JIANG L, et al. Pv-rnn: Point-voxel feature set abstraction for 3d object detection [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10529-10538.
- [16] YANG Z, SUN Y, LIU S, et al. 3dssd: Point-based 3d single stage object detector [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11040-11048.
- [17] T ARMEILLE T, K PAMELA N, PIERRE M, et al. Antiplasmodial limonoids from trichilia rubescens (meliaceae) [J]. Medicinal Chemistry, 2016: 655-661.
- [18] 何新卫, 闫阿泽, 陈龙, 等. 基于改进 PointPillars 的快速准确的三维目标检测算法 [J]. 江苏大学学报(自然科学版), 2020, 41(3): 268-273.
- [19] LI H, XIONG P, AN J, et al. Pyramid attention network for semantic segmentation [J]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, DOI: 10.48550/arXiv.1805.10180.
- [20] JADERBERG M, SIMONYAN K, ZISSERMAN A. Spatial transformer networks [J]. Conference on Computer Vision and Pattern Recognition, 2015, DOI: 10.48550/arXiv.1506.02025.
- [21] HUANG Z, WANG X, HUANG L, et al. Ccnet: Criss-cross attention for semantic segmentation [C]. IEEE/CVF International Conference on Computer Vision, 2019: 603-612.
- [22] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module [C]. European Conference on Computer Vision (ECCV), 2018: 3-19.
- [23] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? the kitti vision benchmark suite [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2012: 3354-3361.
- [24] CHEN X, MA H, WAN J, et al. Multi-view 3d object detection network for autonomous driving [C]. IEEE conference on Computer Vision and Pattern Recognition, 2017: 1907-1915.
- [25] 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: 5099-5108.

作者简介

郑美琳, 硕士研究生, 主要研究方向为图像处理。

E-mail: 2496004224@qq.com

高建瓴, 硕士, 副教授, 主要研究方向为数据库系统、数据挖掘。