文章编号：2095-2163(2022)04-0130-06 中图分类号：TP391 文献标志码：A

基于改进 PointRCNN 的 3D 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 PointRCNN（3D Object Proposal Generation and Detection from Point Cloud）在面对不规则点云时很难提取出有区别特征的问题，提出了一种 Point-ANN（3D Object Proposal Generation and Aggregation Neural Network）的方法。整个框架分为 2 个阶段。第一个阶段是自下而上生成 3D 建议，第二阶段执行建议的 RoI 的感知点云汇集操作，对每个 3D 方案中的点云信息进行分组，并在坐标中改进 3D 建议。引入 RoI 感知点云汇集模块来消除点云上进行区域合并时的模糊性，从而更容易地提取出有区别的特征。通过在 KITTI 数据集上证明了改进的 Point-ANN 方法相比于其他网络在 3D 点云目标检测时精度更高。

关键词：PointRCNN；Point-ANN；RoI 感知点云模块；3D 目标检测

3D point cloud target detection based on improved PointRCNN

ZHENG Meilin，GAO Jianling
（College of Big Data and Information Engineering，Guizhou University，Guiyang 550025，China）

【Abstract】Aiming at the problem that PointRCNN（3D Object Proposal Generation and Detection from Point Cloud）is difficult to extract distinctive features in the face of irregular point cloud，a method of point ANN（3D Object Proposal Generation and Aggregation Neural Network）is proposed. The whole framework is divided into two stages. The first stage is to generate 3D suggestions from bottom to top，and the second stage is to implement the sensing point cloud collection operation of the proposed ROI，and in the process，the research groups the point cloud information in each 3D scheme，then improves the 3D suggestions in coordinates. The ROI aware point cloud aggregation module is introduced to eliminate the fuzziness of region merging on the point cloud，so as to extract different features more easily. It is proved on KITTI data set that the improved point ANN method has higher accuracy in 3D point cloud target detection than other networks.

【Key words】PointRCNN；Point-ANN；RoI perception point cloud module；3D target detection

0 引 言

当前，3D 目标检测在自动驾驶和机器人等领域获得了广泛应用，故而日益受到工业界和学术界的高度关注。激光雷达传感器被广泛地应用在自动驾驶车辆和机器人中，用于捕捉 3D 场景信息，为 3D 场景感知和理解提供了重要的线索。在本文中，提出把 RoI 感知点云模块[1]加入到 PointRCNN[2]中，从而实现高性能的三维点云目标检测。现有的大多数三维检测方法可以根据点云表示分为 2 类：基于网格的方法和基于点的方法。基于网格的方法通常将不规则的点云转换成规则的点云来表示，例如 3D 体素[3]或 2D[4]鸟瞰图，可以被 3D 或 2D 卷积神经网络（CNN）进行有效处理，以学习用于 3D 检测的点特征。基于点的方法[5]由先锋作品 PointNet[6]及其变体提供动力，直接从原始点云中提取有区别的特征用于 3D 检测。一般来说，基于网格的方法计算效率更高，但不可避免的信息丢失会降低细粒度的定位精度，而基于点的方法计算成本更高，实行难度加大。

本文提出了一种新的三维物体检测框架 Point-ANN，研究中结合了基于点和基于体素的特征学习方法的优点，提高了三维检测性能。其原理在于首先提出了一种基于自底向上点云的 3D 包围框建议生成算法，通过将点云分割成前景对象和背景，生成少量高质量的 3D 建议。从分割中学习到的点表示不仅有利于建议生成，而且有助于后期的盒子细化；引入 RoI 点云感知模块来消除汇集点的模糊性，对

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

作者简介：郑美琳（1998-），女，硕士研究生，主要研究方向：智能算法、图像处理；高建瓴（1969-），女，硕士，副教授，主要研究方向：数据库系统、数据挖掘。

通讯作者：高建瓴 Email：454965711@qq.com

收稿日期：2021-11-02

每个 3D 方案中的信息进行分组，然后利用部分聚合网络根据部分特征和信息框进行评分，细化位置。本文提出了基于 PointRCNN 改进的框架 Point-ANN，把其用于 3D 目标检测中与其他框架在 Car 类和 Pedestrian 类目标中对比精度明显提高。

本文的贡献可以概括为 4 个方面：

(1)提出了 Point-ANN 框架，该框架有效地利用了基于体素和基于点的方法进行三维点云特征学习，从而在可管理的内存消耗下提高了三维对象检测的性能。

(2)提出了自上而下场景编码方案，通这些关键点特征不仅保持了准确的位置，还编码了丰富的场景上下文，显著提高了三维检测性能。

(3)提出了一个多尺度的 RoI 特征抽象层，可从场景中聚合更丰富的上下文信息，用于精确的盒子细化和置信度预测。

(4)提出的方法 Point-ANN 以显著的裕度优于所有以前的方法，在竞争激烈的 KITTI 3D 检测基准上也优于以前的方法。

1　PointRCNN 网络结构

PointRCNN 为 2 阶段检测框架，可用来从不规则点云中检测出 3D 物体，2 阶段网络首先生成 3D 建议，再进一步细化位置和建议。具体来说，先要学习逐点特征来分割原始点云，并从分割的前景点同时生成 3D 建议。基于这种自底向上的策略，避免了在 3D 空间中使用大量预定义的 3D 框，并有效限制了用于 3D 建议生成的搜索空间。通过学习分割前景点，点云网络被迫捕获上下文信息以进行精确的点方向预测，这也有利于 3D 框的生成。对于点分割，地面真值分割遮罩自然由 3D 地面真值框提供。对于大型室外场景，前景点的数量通常比背景点的数量少得多。因此，使用焦点损失[7]来处理类不平衡问题，损失函数表达式如式(1)，在训练点云分割过程中，保留默认设置 $\alpha_t = 0.25$ 和 $\gamma = 2$ 作为原始值。具体计算公式可写为：

$$L_{focal}(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t)$$
$$\text{Where } p_t = \begin{cases} p & \text{for forground point} \\ 1-p & \text{otherwise} \end{cases} \quad (1)$$

其中，PointRCNN 以 PointNet++[8]为主干网络，利用 PointNet++网络对每个前景与背景点实现分割，并且赋予一个类别信息，PointNet++框架如图 1 所示。

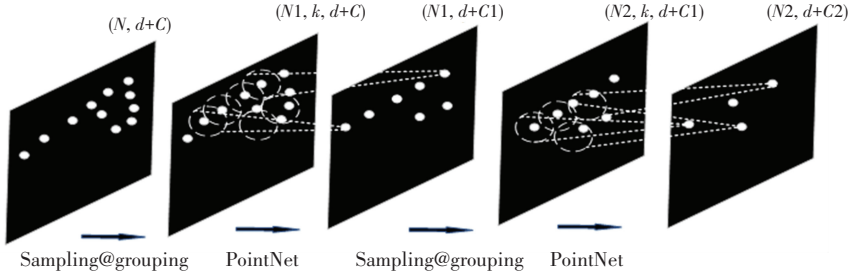

图 1　PointNet++框架图

Fig. 1　PointNet + + framework

每一组提取层的输入是 $(N,(d+C))$，其中 N 是输入点的数量，d 是坐标维度，C 是特征维度。输出是$(N',(d+C'))$，其中N' 是输出点的数量，d 是坐标维度不变，C' 是新的特征维度。

2　ROI 点云感知模块

Shi 等人[2]提出了点云区域合并操作，以合并 3D 方案中的点状特征，从而在第二阶段细化特征。但该操作容易丢失 3D 建议信息，因为点在建议中没有规则地分布，并且存在从汇集点恢复 3D 框的模糊性。RoI 感知点云特征池示意见图 2。由图 2 可知，不同的提议将导致相同的汇集点，这给细化网络带来负面影响。因此，提出了 RoI 感知的点云汇集模块，以将 3D 建议均匀地划分为具有固定空间形状 $(H \times W \times L)$ 的规则空间，其中 H、W、L 是每个维度中汇集分辨率的高度、宽度和长度超参数(例如，在本文的框架中采用 14×14×14)，并且独立于 3D 建议大小。通过聚集(例如，最大池化或平均池化)该空间内的点特征来计算每个空间特征，这里空白空间的特征被设置为零并被标记为空。提议的 RoI 感知池模块是可区分的，这使得整个框架是端到端可训练的。本文提出的 RoI 感知点云汇集模块将不同的 3D 方案标准化为相同的局部空间坐标，其中每个空间编码 3D 方案中对应的是固定网格的特征。

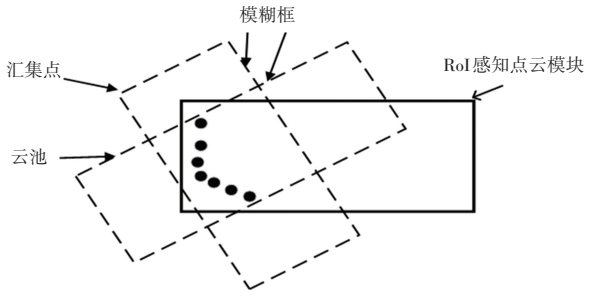

图 2 RoI 感知点云特征池示意图

Fig. 2 Schematic diagram of RoI-aware point cloud feature pool

3 改进的 Point-ANN 网络

3.1 通过点云分割自底向上生成 3D 建议

现有的 2D 物体检测方法可分为一阶段和两阶段方法。通常情况下,一阶段方法[7]更快,但是直接估计对象边界框而不进行细化,而两阶段方法[9]首先生成建议,并在第二阶段进一步细化建议和置信度。然而,两阶段方法的直接扩展是基于三维空间和点云的不规则格式。AVOD [10]在 3D 空间中放置 80~100 千个锚框,并在多个视图中为每个锚集中要素,以生成建议。FPointNet[11]从 2D 图像生成 2D 建议,再基于从 2D 区域裁剪的 3D 点来估计 3D 框,如此就可能会错过只能从 3D 空间清晰观察到的困难对象。提出了一种精确、鲁棒的三维方案生成算法,作为基于全场景点云分割的第一阶段子网络。观察到三维场景中的物体是自然分离的,相互之间没有重叠。所有三维对象的分割模板都可以通过其三维包围盒标注直接获得,即三维包围盒内的三维点被视为前景点。因此,建议以自下而上的方式生成 3D 提案。具体来说,学习逐点特征来分割原始点云,并从分割的前景点同时生成 3D 建议。基于这种自底向上的策略方法避免了在 3D 空间中使用大量预定义的 3D 框,这样一来就有效限制了用于 3D 建议生成的搜索空间。

3.2 用于 3D 盒细化的零件位置聚合

通过考虑建议中所有 3D 点的预测对象内零件位置的空间分布,用聚集预测零件位置来评估该建议的质量是合理的。实际上,可以将其公式转化为一个优化问题,同时通过拟合相应方案中所有点的预测零件位置来直接求解 3D 包围盒的参数。然而,研究发现这种基于优化的方法对异常值和预测零件位置的质量很敏感。为了解决这一问题,提出了一种基于学习的方法来鲁棒地聚合零件位置信息,用于盒评分和位置细化。对于每个 3D 建议,将建议的 RoI 感知点云汇集操作分别应用于来自第一阶段的预测的点状零件位置(平均汇集)和点状特征(最大汇集),这导致大小为(14×14×14×4)和(14×14×14× C) 的 2 个特征图,其中预测的零件位置图是 4 维的;3 个用于零件位置 (x,y,z 维度),1 个用于前景分割分数,而 C 是由第一阶段转换的点状特征的特征维度在汇集操作之后,以分层方式实现零件聚集网络,以从预测的对象内部零件位置的空间分布中学习。具体来说,首先使用核大小为 3×3×3 的稀疏卷积层将 2 个合并的特征映射转换为相同的特征维数。在连接这 2 个特征映射后,堆叠了 4 个核大小为 3×3×3 的稀疏卷积层,以随着接收域的增加逐渐聚集部分信息。这里,还在第二个卷积层之后利用一个具有 2×2×2 的内核大小和 2×2×2 步长的稀疏最大池来将特征映射下采样到 7×7×7,以节省计算和参数。此后将其矢量化为一个特征向量,并添加 2 个分支,用于最终的盒子评分和位置细化。与将合并后的三维特征图直接矢量化为特征向量的简单方法相比,提出的零件聚集策略可以通过将特征从局部尺度聚集到全局尺度来有效地学习预测零件位置的空间分布。

3.3 改进的 Point-ANN 网络框架

Point-ANN 以自下而上的方式从原始点云生成 3D 建议,然后用第二个零件聚合阶段执行建议的 RoI 感知点云汇集操作,对每个 3D 方案中的零件信息进行分组,使用零件聚合网络对盒子进行评分,再根据零件特征和信息进行位置优化。引入 RoI 点云感知模块可以通过编码盒形,并用稀疏卷积加以有效处理。研究得到的改进的 Point-ANN 框架如图 3 所示。

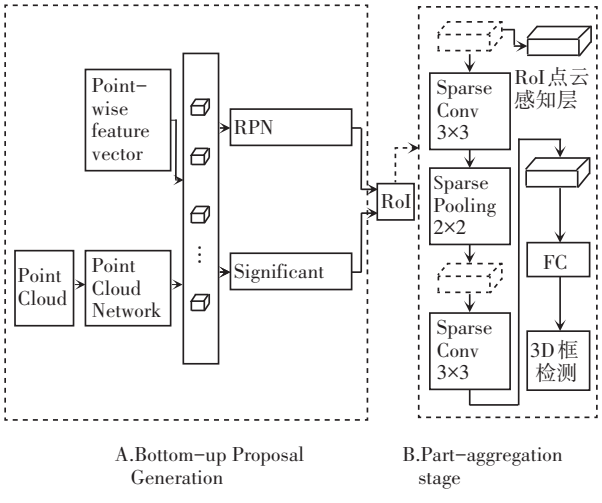

图 3 改进的 Point-ANN 框架

Fig. 3 Improved Point-ANN framework

3.4 损失函数

将 2 个分支附加到从预测零件信息聚合的矢量化特征向量。对于盒子评分分支，使用 3D 建议书与其对应的地面真值盒之间的 3D IoU 作为建议书质量评估的软标签，这也是通过二元交叉熵损失作为 Eq 来学习的。对于三维建议的生成和细化，采用平滑 L_1 损失来回归归一化盒参数，如下所示：

$$\Delta x = \frac{x^g - x^a}{d^a},\ \Delta y = \frac{y^g - y^a}{h^a},\ \Delta z = \frac{z^g - z^a}{d^a},$$

$$\Delta l = \log\left(\frac{l^g}{l^a}\right),\Delta h = \log\left(\frac{h^g}{h^a}\right),\Delta w = \log\left(\frac{w^g}{w^a}\right) \quad (2)$$

$$\Delta\theta = \theta^g - \theta^a, d^a = \sqrt{(l^a)^2 + (w^a)^2}$$

其中，$d^a = \sqrt{(l^a)^2 + (w^a)^2}$ 归一化鸟瞰图中的中心偏移；$(x^a, y^a, z^a, h^a, w^a, l^a, \theta^a)$ 是 3D 锚/方案的参数；$(x^g, y^g, z^g, h^g, w^g, l^g, \theta^g)$ 表示其对应的地面真值框。

在这里，为了细化建议，基于 3D 建议的参数把直接回归相对偏移或大小比，因为建议的 RoI 感知点云池模块已经编码了 3D 建议的完整几何信息，并将不同的 3D 建议传输到相同的归一化空间坐标系。因此，具有相等权重的部分感知阶段有 3 个损失，包括前景点分割的焦点损失、对象内部分位置回归的二元交叉熵损失和 3D 建议生成的平滑损失。对于部分聚集阶段，有 2 个损失也具有相同的损失权重，包括 IoU 回归的二元交叉熵损失和位置细化的平滑损失。

4 实验及结果分析

4.1 实验细节网络架构

对于训练集中的每个 3D 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。对于阶段一子网络，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每点特征向量。对于盒子提议细化子网络，从每个提议的汇集区域随机抽样 512 个点作为细化子网络的输入。使用具有单尺度分组（具有组大小 128、32、1）的 3 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。在这里，报告了汽车类别的训练细节，则因其在 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。对于阶段—子网络，3D 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 3D 地面真值框放大 0.2 m 来进行鲁棒分割，因为 3D 地面真值框可能有小的变化。对于基于箱的建议生成，超参数被设置为搜索范围 $S = 3$ m，箱大小 $\delta = 0.5$ m，定向箱数量 $n = 12$。对于部分聚集阶段，RoI 感知点云汇集模块的汇集分辨率为 14×14×14，经稀疏卷积和特征维数为 128 的最大汇集处理后下采样为 7×7×7。将下采样的特征映射矢量化为单个特征向量，用于最终的盒子评分和位置细化。使用自动数据管理优化器对整个网络进行端到端培训，批量为 650 个时期。使用余弦退火学习速率策略，初始学习速率为 0.001。从每个场景中随机选择 128 个建议用于第二阶段的训练，其中 50%的建议具有 3D IoU，其对应的地面真值框至少为 0.55。在训练期间进行常见的数据增强，包括随机翻转、从 $[0.95, 1.05]$ 采样的比例因子的全局缩放、从 $\left[-\frac{\pi}{4}, \frac{\pi}{4}\right]$ 采样的围绕垂直轴的角度的全局旋转。从推论来看，只有 100 个建议来自部分知晓阶段，*NMS* 阈值为 0.7，然后由下一个部分汇总阶段对其进行评分和细化。最终应用阈值为 0.01 的旋转 *NMS* 去除冗余盒，并生成最终的 3D 检测结果。

4.2 KITTI 上的 3D 对象检测

KITTI 的 3D 对象检测基准包含 7 481 个训练样本和 7 518 个测试样本（测试分割）。遵循常用的 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）。在 KITTI 数据集的值分割和测试分割上，将点神经网络与最先进的 3D 对象检测方法进行了比较。所有模型都在列车分割上进行训练，并在测试分割和价值分割上进行评估。对于 3D 对象检测的评估，在 KITTI 数据集的具有挑战性的 3D 对象检测基准上对 Point-ANN 进行评估，与 VoxelNet[12]、SECOND[13]、PointRCNN[2] 等方法在 Car 类和 Pedestrian 类目标中进行对比实验，结果见表 1，通过以下度量来评估预测的对象内零件位置：

$$AbsError_u = \frac{1}{\|G\|}\sum_{i=G}|P_u^i - Q_u^i|, u \in \{x, y, z\} \quad (3)$$

表 1 在 KITTI 3D 物体检测测试服务器上的性能评估(测试分割)

Tab. 1 Performance evaluation (test segmentation) on KITTI 3D object detection test server

Method	Modality	Car (IoU = 0.7)			Pedestrian (IoU = 0.5)		
		Easy	Moderate	Hard	Easy	Moderate	Hard
MV3D	RGB+LiDAR	72.08	63.40	56.21	—	—	—
UberATG-ContFuse	RGB+LiDAR	83.65	67.32	65.14	—	—	—
AVOD-FPN	RGB+LiDAR	82.65	72.15	67.05	50.45	43.05	41.92
F-PointNet	RGB+LiDAR	82.21	71.06	63.01	51.36	44.92	42.02
VoxelNet	LiDAR	78.18	66.06	58.09	39.68	34.68	35.65
SECOND	LiDAR	84.02	74.15	67.09	52.02	43.21	38.11
PointRCNN	LiDAR	87.06	77.42	76.34	50.19	44.32	39.25
Part-A^2	LiDAR	86.42	77.65	77.51	50.02	43.66	39.11
The proposed	LiDAR	88.06	78.01	76.52	54.17	46.88	40.02

在 KITTI 测试服务器的 3D 检测基准上评估的方法结果显示在表 1 中,表 2、表 3 显示了融入的 RoI 模块与其他聚合部分相比的结果以及影响。对于汽车和行人的 3D 检测,本文的方法优于以前最先进的方法,在 Easy、Moderate、Hard 三个困难等级方面都有显著的优势。虽然以前的大多数方法都使用 RGB 图像和点云作为输入,但是本文的方法通过仅使用点云作为输入来获得更好的性能。在行人检测方面,与以往的仅使用激光雷达的方法相比,改进的方法取得了更好或相当的结果。然而,本文方法的性能比具有多个传感器的方法稍差。认为这是由于本文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 3D 检测。对于最重要的汽车类别,改进的方法优于以前的先进方法,在价值分割上有很大的利润。特别是在难度较大的情况下,改进的方法比以前最好的方法有所提高,证明了所提出的点神经网络的有效性。

表 2 RoI 感知点云区域池的影响

Tab. 2 Effects of RoI-aware point cloud region pool

Method	AP_{Easy}	$AP_{Mod.}$	AP_{Hard}
RoI fifixed-sized pool(14×14×14)& sparse conv	88.23	78.36	78.03
RoI-aware pool (14×14×14)&FCs	89.21	78.45	78.16
RoI-aware pool(14×14×14)&sparse conv	89.26	78.51	78.20

表 3 不同部分聚合网络结构的比较

Tab. 3 Comparison of different part-aggregation network structures

Method	AP_{Easy}	$AP_{Mod.}$	AP_{Hard}
RoI-aware pool 7 × 7 × 7 & FCs	89.21	80.02	78.65
RoI-aware pool 7 × 7 × 7 & sparse conv	89.35	80.32	78.72
The proposed	89.62	80.45	78.98

5 结束语

(1)本文考虑到 PointRCNN 在面对大量不规则点云时不能精确地提取出特征这一问题,利用改进的 Point-ANN 网络框架从点云中检测三维物体来弥补这一不足。

(2) 融入 RoI 点云感知模块,让每个对象的预测内部对象部分位置被池化。因此,随后的部件聚集阶段可以考虑预测的对象内部件位置的空间关系,以及对盒子进行评分并细化相应的位置。

(3) 实验表明,本文的方法在具有挑战性的 KITTI 3D 检测基准上取得了最先进的性能,证明了改进方法有效性。

参考文献

[1] SHI Shaoshuai, WANG Zhe, SHI Jianping, 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, 2020,43(8):2647-2664.

[2] SHI Shaoshuai, WANG Xiaogang, LI Xin. PointRCNN: 3d object proposal generation and detection from point cloud [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach:IEEE , 2019:770-779.

[3] CHEN Yilun, LIU Shu, SHEN Xiaoyong, et al. Fast point R-CNN [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South):IEEE, 2019:9774-9783.

[4] CHEN Xiaozhi, MA Huimin, WAN Ji, et al. Multi-view 3D object detection network for autonomous driving[C]// The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA:IEEE, 2017:6526-6534.

[5] QI C R, LIU Wei, WU Chenxia, et al. Frustum pointnets for 3d object detection from RGB-D data[C]// The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City:IEEE, 2018:1-15.

(下转第 139 页)