基于深度学习的三维目标检测方法研究

王 刚[1,2] 王 沛[1]

[1](中国科学院电子学研究所 北京 100190)

[2](中国科学院大学 北京 100190)

摘 要 对于自动驾驶的感知模块,提出一种多传感器融合的改进三维目标检测算法。在 Faster RCNN 的基础上,对点云数据进行预处理,得到多通道的俯视特征图;分别对点云数据生成的俯视特征图与 RGB 图像数据进行特征提取和融合,并进行分类预测和位置回归。采用 KITTI 数据集对算法进行验证和测试,结果显示,该算法与目前两种点云与图像结合的目标检测算法比较,其 3D 平均精度(3D average precision)和俯视图平均精度(BEV average precision)都有一定的提升。

关键词 深度学习 三维目标检测 激光雷达点云

中图分类号 TP3 **文献标志码** A DOI:10.3969/j.issn.1000-386x.2020.12.026

3D OBJECT DETECTION METHOD BASED ON DEEP LEARNING

Wang Gang[1,2] Wang Pei[1]

[1](*Institute of Electrics, Chinese Academy of Sciences, Beijing 100190, China*)

[2](*University of Chinese Academy of Sciences, Beijing 100190, China*)

Abstract For the sensing module of autonomous driving, an improved 3D object detection algorithm based on multi-sensor fusion is proposed. Based on Faster RCNN, it preprocessed the point cloud data to obtain a multi-channel top view feature map, and extracted the top view feature map generated by the point cloud and RGB image data respectively. Then it fused the convolutional top view feature map and the convolutional image feature map, and the feature map was used to classify and regress the position. The KITTI dataset was used to validate and test the algorithm. The results show that the proposed algorithm compared with the current two algorithms for combining point cloud with image, the 3D average precision and the bird eye view average precision are improved.

Keywords Deep learning 3D object detection Lidar point cloud

0 引 言

近年来,随着智能化技术以及无人化技术的发展,传感器技术蓬勃发展。作为三维环境感知传感器的激光雷达自然也受到了越来越多的关注,其在无人驾驶、测绘、军事等领域都有很多运用。激光雷达的数据产品是三维点云,即三维坐标系下的点的数据集,它包含三维坐标(x,y,z)和反射强度等丰富的信息。利用激光雷达产生的点云数据,可以获得三维目标的三维信息,比图像具有更好的深度信息;而图像具有 RGB 值,具有目标的更多的细节信息。因此在自动驾驶领域,感知模块的潮流就是将激光雷达数据和二维图像相结合,进行目标检测,获得汽车的周围环境信息。

在自动驾驶的感知算法方面,主要有三种思路:

(1) 利用相机产生的二维图像进行目标检测。传统的图像目标检测算法采用方向梯度直方图(Histogram of Oriented Gradient, HOG)[1]、尺度不变特征变换(Scale-Invariant Feature Transform, SIFT)[2]等手工特征对图像进行特征提取,得到目标的边缘信息,再用支持向量机(Support Vector Machine, SVM)[3]或 AdaBoost[4]算法对目标特征进行分类检测。在神经网络

收稿日期:2019-06-15。**王刚**,硕士生,主研领域:激光雷达点云目标检测。**王沛**,副研究员。

发展之后,RCNN[5]、Fast RCNN[6]、Faster RCNN[7]等算法将目标检测提升到了一个新的高度。但由于相机图片是二维的,如果在自动驾驶场景中完全利用图像信息,很难获得三维空间目标的精确位置。

(2) 利用激光雷达的点云数据进行目标检测。激光雷达点云数据具有三维空间丰富的深度信息,利用激光雷达点云的这一特点,可以进行三维目标检测。Zhou 等[8]利用类似于图像像素的方法,将点云数据体素化,每个体素取值 0 或 1(判断体素是否含有目标),再将三维卷积神经网络运用到点云的体素网格。但是由于空体素的存在,这种方法消耗了大量的内存并且需要大量的计算量。PointNet[9-10]系列算法与 PointCNN[11]直接对点云进行处理,进行点云分类,但是这些算法只能适用于室内环境这些小场景,对于自动驾驶这种复杂场景难以适应。

(3) 利用激光雷达点云数据与相机图像融合进行目标检测。百度提出的 MV3D[12]将激光雷达点云数据投影成俯视图与前视图,在点云俯视图上进行候选区域生成,再将生成的候选区域分别映射至 RGB 图像、点云俯视图和点云前视图上进行感兴趣区域(Region of Interest)的特征提取与特征融合,最后进行位置回归和目标分类。但是这种方法只利用俯视图生成候选区域,会造成分类和定位的不准确。Qi 等[13]对二维图像利用区域候选网络生成候选区域,并将生成的候选区域映射至三维点云中,运用 PointNet ++ 进行点云分类。这种方法只利用了图像信息进行候选区域生成,也会造成分类与定位的不准确。Ku 等[14]提出了 AVOD 算法,其利用深度卷积网络分别对三维点云数据的俯视图与二维图像进行特征提取,并分别将二者得到的特征图送入区域候选网络进行候选区域生成,最后进行目标分类和位置回归。

本文提出自动驾驶场景下的三维目标检测改进算法,利用点云和图像融合的方法,检测目标,并获得目标物体的三维位置信息与类别信息。本文提出的算法具有以下创新点:

(1) 对激光点云进行预处理,得到具有高度通道、点云密度通道以及反射强度的俯视特征图。

(2) 在区域候选网络之后,对点云的俯视图特征与图像特征的 ROI 使用 ROI Align 进行池化,避免了 ROI Pooling 的两次量化造成的误差。

1　整体网络结构

本文将 Faster RCNN[6]运用到点云和图像融合的三维目标检测上,其为了加强对小目标的检测准确性,在生成特征图时,引入了特征金字塔网络[15],使其生成的特征图与输入图像具有同样的尺寸,并融合了各个卷积层所提取的特征,使网络对小目标召回率提高。整体网络结构如图 1 所示。

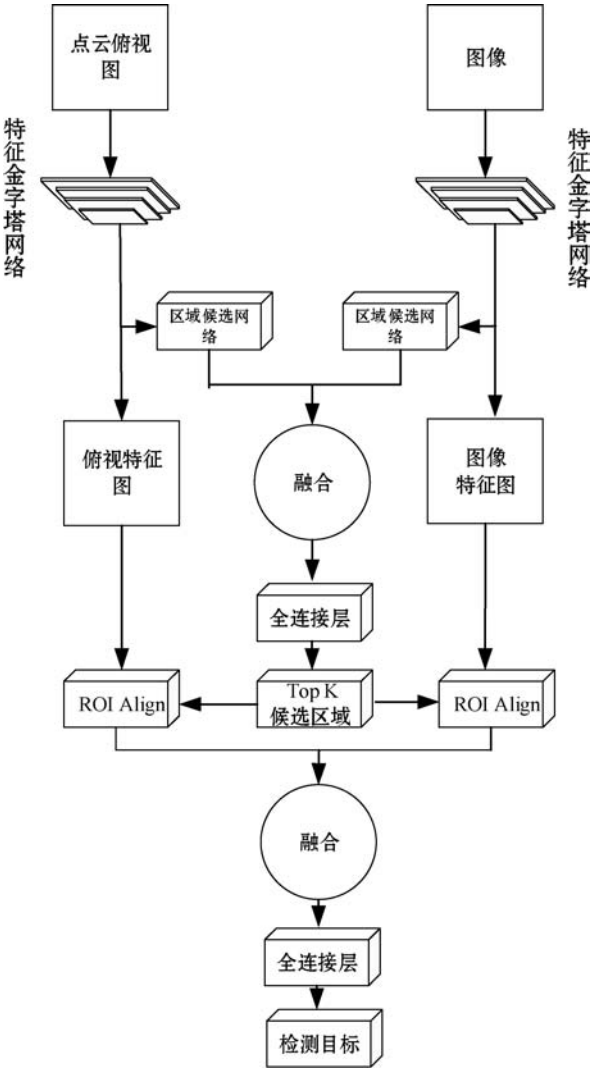


图 1　整体结构图

1.1　Faster RCNN 网络

Faster RCNN 算法是 2015 年提出的两阶段目标检测算法,是目标检测的经典框架,其提出了区域候选网络。目标检测具有两个任务:目标分类和位置回归。基于区域候选网络的方法能够很好地完成这两个任务。其主要分为两步:

Step 1　将图像作为输入,使用深度网络提取输入图像的特征图,区域候选网络对前面生成的特征图进行裁剪,使其生成一定量的 anchor,然后区域候选网络再对这些 anchor 作分类(判断是不是目标)和位置回归(粗定位),生成一定量的候选区域。区域候选网络如图 2 所示。

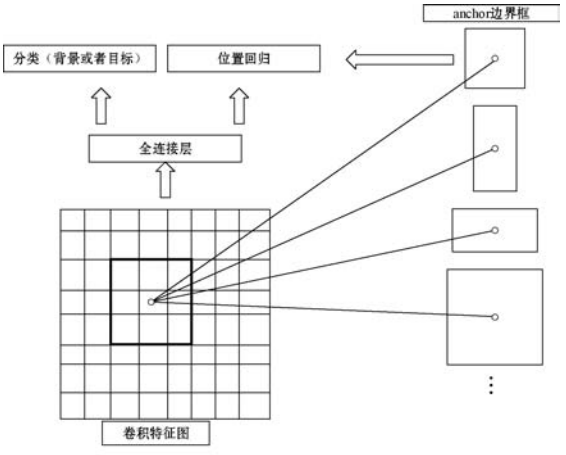


图 2 特征候选网络

Step 2 将生成的候选区域映射至输入图像生成的特征图，得到感兴趣区域，并进行 ROI Pooling，得到固定大小的候选区域特征图。最后用全连接层进行分类（具体类别）和位置回归（精确定位）。

1.2 特征金字塔网络

特征金字塔网络通过简单地改变网络的连接，在几乎不增加网络计算量的情况下，提升了网络对小目标的检测性能。在深度卷积神经网络中，网络的层数越深，得到的特征图的分辨率越低，但语义信息却越丰富。特征金字塔网络通过将卷积网络底层的高分辨率、低语义信息的特征图与网络上层的低分辨率、高语义信息的特征图进行由上到下的连接，使各个尺度下的特征图都具有较为丰富的语义信息。特征金字塔网络主要分为三个部分：自下而上的卷积特征提取网络，自上而下的上采样过程，以及同一层间特征的横向连接。自下而上的卷积特征提取网络为卷积网络的前向过程，得到语义信息逐渐增强，分辨率逐渐变小的特征图。自上而下的上采样过程将上层特征图逐渐向下一层上采样，得到与下一层相同尺寸的特征图。横向连接将上采样得到的特征图与下一层特征图进行融合，使得到的特征图既具有高层语义信息，又具有低层的定位细节信息。特征金字塔网络如图 3 所示。

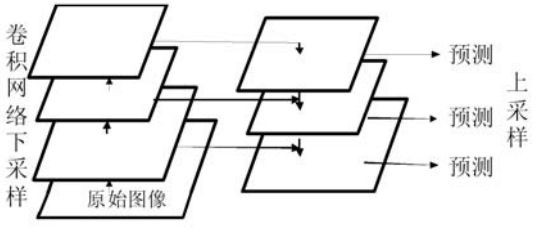


图 3 特征金字塔网络

2 点云数据预处理

KITTI 点云数据集是由 Velodyne 64 线激光雷达获取的点云数据集，通过激光照射到物体表面获取对目标的距离信息 (x,y,z) 以及反射强度值。首先对三维点云数据以 0.1 m 的分辨率进行投影，得到点云数据的俯视图。俯视图的特征通道由高度特征图、密度特征图和反射强度特征图组成。对于高度特征图，将点云沿 Z 轴平均划分为 5 个切片，将每个切片的每个像素位置的最大高度值作为这个像素值，因此具有 5 个高度特征图。对于密度特征图，取每个像素位置的占有点的数量 N 作为密度值，还需对这个密度值以 $\min\left(1.0,\frac{\log(N+1)}{\log(64)}\right)$ 进行归一化。本文加入了之前点云预处理算法中没有使用的反射强度（intensity）特征图，取每个像素位置的最大高度处的点的反射强度作为其反射强度特征。经过预处理后得到的 7 个通道的特征图如图 4 所示。

(a) 高度图 (b) 密度图 (c) 反射强度图

图 4 点云俯视图

3 ROI Align

在 Faster RCNN 中，区域候选网络生成候选区域后，网络将候选区域映射到前面卷积网络生成的特征图中，然后使用 ROI Pooling 对目标区域进行池化，但是这个步骤会有两次量化操作，如图 5 所示。由于在卷积过程中，图像进行了下采样，所以在将候选区域映射至卷积特征图的过程中，也需要将候选框下采样同样的倍数，如果对候选框下采样不能够除尽，将其量化为整数，这就出现了第一次量化操作。在池化过程中，需要将候选区域平均分割为 $k\times k$ 个单元，同样，如果不能整除，这就会产生第二次量化操作。经过这两次量化后，回归出来的目标框与量化后的目标框发生了一定的偏差，因此会导致检测精度下降。

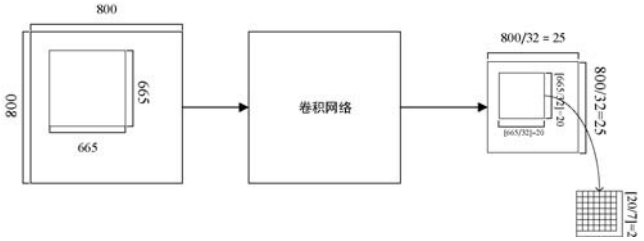


图 5 ROI Pooling

采用 ROI Align[16] 解决这个问题，如图 6 所示。

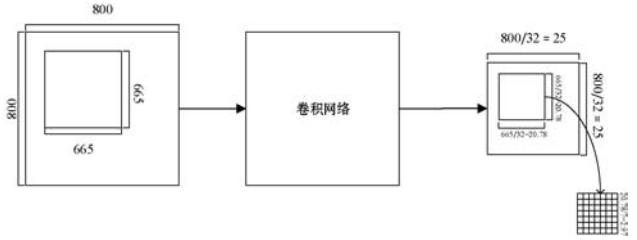


图 6　ROI Align

取消两次量化操作,使用双线性插值获得坐标为浮点数的像素点上的图像数值。具体过程如下:

(1) 将目标候选区域映射至卷积特征图上,映射过程中不做量化处理。

(2) 将候选区域划分为 $k \times k$ 个单元,每个单元的边界也不做量化处理。

(3) 对每个单元取四个固定的坐标点,用双线性插值的方法计算出这四个点的坐标,然后对其进行最大池化操作。

在 ROI Pooling 操作中,反向传播公式为:

$$\frac{\partial L}{\partial x_i} = \sum_r \sum_j [i = i^*(r,j)] \frac{\partial L}{\partial y_{rj}} \tag{1}$$

式中:x_i表示在池化操作之前卷积特征图上的像素点;y_{rj}表示 ROI Pooling 之后的第 r 个候选框的第 j 个点;$i^*(r,j)$代表点 y_{rj}池化之前的坐标点。由式(1)可以看出,只有在 ROI Pooling 之后的点的像素值在 Pooling 操作中使用了当前点 x_i的像素值(即当 $i = i^*(r,j)$)时,x_i的梯度才反向传播。

ROI Align 的反向传播公式为:

$$\frac{\partial L}{\partial x_i} = \sum_r \sum_j [d(i,i^*(r,j)) < 1](1 - \Delta h)(1 - \Delta w) \frac{\partial L}{\partial y_{rj}} \tag{2}$$

式中:$d(\cdot)$表示两点之间的距离;Δh 和 Δw 表示 x_i与 $x_i^*(r,j)$横纵坐标的差值,这里作为双线性内插的系数乘在原始的梯度上。可以看出,在 ROI Align 中,$x_i(r,j)$的坐标值是一个浮点数(卷积网络前向传播得到的坐标点),在 Pooling 操作之前的卷积特征图中,与 x_i与 $x_i^*(r,j)$横纵坐标都小于 1 的点都要接收与此对应的池化后的点 $y_{r,j}$回传的梯度。

4　实　验

4.1　实验数据集

本文实验采用 KITTI 数据集,它是自动驾驶领域最出名的数据集之一,目前自动驾驶领域的大量算法都在此数据集下进行实验。本文利用其三维点云数据集和图像数据集,包含 7 481 个三维点云文件和图像文件。点云文件被裁减到以激光雷达为原点,横纵坐标分别为[-40,40] ×[0,70] m 的范围内。

4.2　实验环境

本文是在 Ubuntu 16.04 系统下,采用 TensorFlow 1.9 深度学习框架,CPU 为 Intel(R) Core(TM) i7 - 3770,GPU 为 MSI 1080Ti,开发工具为 Pycharm + Anaconda,Python 版本为 3.6。训练大约需要 15 h。

4.3　实验结果分析

在训练集,测试集与验证集的分割与目前两种基于激光雷达点云数据和图像融合的算法分割相同的情况下,由表 1 与表 2 可以看出,在 KITTI 数据集中,本文算法在加入反射强度信息后,3D 平均精度(AP-3D)和俯视图平均精度(AP-BEV)都有一定的提升,说明反射强度信息对神经网络的特征提取具有一定的帮助。在用 ROI Align 替代 ROI Pooling 后,3D 平均精度和俯视图平均精度也有一定的提升,特别是在检测小目标方面,其中容易、中等、困难为目标检测的难度。检测结果如图 7 所示。

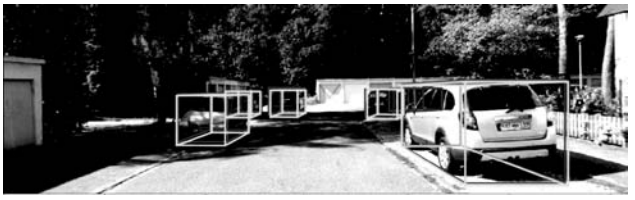
表 1　各方法 3D 平均精度(AP -3D)对比(Car)

方法	AP-3D		
	容易	中等	困难
MV3D[12]	0.710 9	0.623 5	0.551 2
AVOD[14]	0.735 9	0.657 8	0.583 8
AVOD-FPN[14]	0.819 4	0.7188	0.663 8
反射强度	0.816 7	0.727 4	0.668 5
ROI Align	**0.822 9**	**0.738 6**	**0.676 3**

表 2　各方法俯视图平均精度(AP-BEV)对比(Car)

方法	AP-BEV		
	容易	中等	困难
MV3D[12]	0.860 2	0.769 0	0.684 9
AVOD[14]	0.868 0	0.854 4	0.777 3
AVOD-FPN[14]	0.885 3	0.837 9	0.779 0
反射强度	0.892 1	0.864 1	0.795 2
ROI Align	**0.892 3**	**0.870 4**	**0.796 9**

(a) 2D 车辆检测结果



(b) 3D 车辆检测结果

图 7 检测结果

5 结 语

本文提出基于深度学习的三维目标检测改进方法,并且实现改进的检测网络。通过将激光雷达点云数据与图像相结合,使检测网络不仅能够提取激光雷达点云的深度信息,还可以提取图像的颜色细节信息,减少单一输入形式对目标检测准确率造成的影响。将反映目标材质信息的反射强度引入点云俯视图的信息通道中,对特征的提取有一定的帮助;将候选区域网络得到的候选区域映射至卷积特征图之后,采用 ROI Align,避免了原来池化过程中的两次量化操作。本文提出的对三维目标检测的改进方法,在使用激光雷达点云与图像融合的前提下,三维目标检测效果有一定的提升。

参 考 文 献

[1] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05),2005.

[2] 李晓明,郑链,胡占义. 基于 SIFT 特征的遥感影像自动配准[J]. 遥感学报,2006,10(6):885 - 892.

[3] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报,2000,26(1):32 - 42.

[4] 曹莹,苗启广,刘家辰,等. AdaBoost 算法研究进展与展望[J]. 自动化学报,2013,39(6):745 - 758.

[5] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition,2014.

[6] Girshick R. Fast R-CNN[J]. 2015 IEEE International Conference on Computer Vision (ICCV), 2015.

[7] Ren S, Girshick R, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137 - 1149.

[8] Zhou Y, 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, 2018.

[9] Charles R Q, Hao S, Mo K, et al. PointNet: Deep learning on point sets for 3D classification and segmentation[C]//2017 IEEE Conference on Computer Vision & Pattern Recognition, 2017.

[10] Qi C R, Yi L, Su H, et al. PointNet++: Deep hierarchical feature learning on point sets in a metric space[EB]. arXiv: 1706.02413,2017.

[11] Li Y, Bu R, Sun M, et al. PointCNN: Convolution on χ-transformed points[EB]//arXiv:1801.07791,2018.

[12] Chen X, Ma H, Wan J, et al. Multi-view 3D object detection network for autonomous driving[C]//2017 IEEE Conference on Computer Vision & Pattern Recognition,2017.

[13] Qi C R, Liu W, Wu C, et al. Frustum pointnets for 3D object detection from RGB-D data [EB]. arXiv: 1711.08488,2017.

[14] Ku J, Mozifian M, Lee J, et al. Joint 3D proposal generation and object detection from view aggregation[EB]. arXiv: 1712.02294,2017.

[15] Lin T Y, Dollár Piotr, Girshick R, et al. Feature pyramid networks for object detection[EB]. arXiv:1612.03144,2016.

[16] He K, Gkioxari G, Dollar P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2020, 42(2):386 - 397.

(上接第 94 页)

参 考 文 献

[1] 董煜. 基于以太网技术的无人值守地磅系统设计[D]. 济南:齐鲁工业大学,2015.

[2] 王凯. 基于 SIMPASS 技术的门禁系统设计开发[D]. 南京:南京信息工程大学,2011.

[3] Su B, Zhang B. Notice of retraction: The application of MES in electronic assembly industry based on RFID[C]//2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC),2011.

[4] 潘鑫. 超高频激光全息 RFID 标签设计与实现[D]. 武汉:华中科技大学,2013.

[5] 靳钊. 无源超高频射频识别标签设计中的关键技术研究[D]. 西安:西安电子科技大学,2011.

[6] 罗胜彬. 磁电式扭矩传感器的研究[D]. 成都:西华大学,2015.

[7] 孙丽敬. RFID 在汽车生产线应用的电磁兼容性研究[D]. 重庆:重庆大学,2010.

[8] 吕凌. 工作于 860 - 960MHz 频段的 RFID 读写器和标签一致性指标测试[J]. 电子质量,2016(8):57 - 62,65.

[9] 傅益标. 基于 EPC C1G2 协议的超高频 RFID 系统设计及仿真[D]. 哈尔滨:哈尔滨工业大学,2008.

[10] 王明星,李利民. 改进的遗传算法在机加车间排产优化的研究[J]. 制造业自动化,2014,36(15):34 - 37,48.