

基于边缘卷积的三维目标识别算法

江泽宇,赵 芸

(浙江科技学院 信息与工程学院,杭州 310023)

摘 要: 三维目标检测是实现无人驾驶必不可少的技术,但很多三维检测算法采用的分割算法并不能很好地提取局部特征,导致检测精度不理想。为了改善局部特征缺失的情况,提出一种基于边缘卷积的三维目标识别算法。本算法以激光点云和 RGB(red, green, blue)图像作为输入,基于二维候选区域中的像素过滤激光点云生成视锥点云,以此提高检测速度。同时,在分割算法中,在点云的局部特征图的基础上计算目标点和相邻点之间的欧氏距离,并将其作为边缘特征赋予目标点和相邻点。此外,在卷积神经网络提取特征的过程中,每次卷积完成后都会在新的局部特征图上重新计算三维点之间的欧氏距离,为三维点构造新的边缘特征。这使得边缘特征能随着卷积神经网络的计算扩散到整个点云,从而提高局部特征的提取效果。本算法在 KITTI(Karlsruhe Institute of Technology and Toyota Technological Institute)的三维点云数据集上进行验证,分割精度达到 92.82%,相比于 F-PointNet 提高了 2.30 百分点;对不同目标的检测精度也有所提高,车辆、自行车、行人的检测精度分别达到了 85.77%,76.09%,53.08%。试验结果证明了本算法的可行性。本算法可应用于无人驾驶汽车,实现车辆、行人和自行车的定位与检测。

关键词: 边缘特征;边缘卷积;欧氏距离;视锥;KITTI

中图分类号: TP183 **文献标志码:** A **文章编号:** 1671-8798(2021)03-0214-06

3D target recognition algorithm based on edge convolution

JIANG Zeyu, ZHAO Yun

(School of Information and Electronic Engineering, Zhejiang University of
Science and Technology, Hangzhou 310023, Zhejiang, China)

Abstract: Three-dimensional (3D) target detection is an indispensable technology to realize unmanned driving, but many segmentation algorithms used cannot extract local features well, resulting in unsatisfactory detection accuracy. In order to improve the deficiency of local features, a 3D target recognition algorithm was proposed on the basis of edge convolution. This algorithm capitalized on laser point cloud and RGB(red, green, blue) images as input, and

收稿日期: 2020-07-18

基金项目: 国家自然科学基金项目(61605173);浙江省自然科学基金项目(LY16C130003)

通信作者: 赵 芸(1981—),女,浙江省杭州人,教授,博士,主要从事信息工程与人工智能研究。E-mail: zy_super0201@163.com。

filtered the laser point cloud based on pixels in the two-dimensional (2D) candidate area to generate a viewing cone point cloud to improve the detection speed. At the same time, in the segmentation algorithm, the Euclidean distance between the target point and the neighboring point was calculated on the basis of the local feature map of the point cloud, and assigned to the target point and the neighboring point as an edge feature. In addition, in the process of extracting features by the convolutional neural network, the Euclidean distance between the 3D points was recalculated on the new local feature map every time the convolution was completed, constructing new edge features for the 3D points, which enabled the edge features to be diffused to the entire point cloud with the calculation of the convolutional neural network, thereby improving the extraction effect of local features. The proposed algorithm was verified on the 3D point cloud data set of KITTI (Karlsruhe Institute of Technology and Toyota Technological Institute), with the segmentation accuracy up to 92.82%, 2.30 percentage points higher than F-PointNet and with the detection accuracy of different targets improved, the detection accuracy of vehicles, bicycles, and pedestrians reaching 85.77%, 76.09%, and 53.08% respectively. The test results prove the feasibility of the algorithm, which can be applied to driverless cars to realize the positioning and detection of vehicles, pedestrians and bicycles.

Keywords: edge features; edge convolution; Euclidean distance; viewing cone; KITTI

要实现自动驾驶必须解决道路上障碍物的检测问题,即对道路上的三维物体进行识别和位置估计。随着卷积神经网络的发展,大多数研究人员采用深度学习来处理三维物体的识别。为了能直接采用深度学习处理点云,Qi等^[1]提出了PointNet,采用最大池(max pooling)解决了点云的无序性问题。在PointNet的基础上,也出现了很多直接处理点云的分割算法和三维目标识别算法。其中,Yan等^[2]提出了一种改进的稀疏卷积算法并将其应用于三维物体检测,该算法检测速度快,但小目标检测精度不高;Wang等^[3]在PointNet的基础上提出了EdgeConv,采用局部邻域映射的方法来提取点云的局部特征并分割目标,能够捕捉到潜在的远距离相似特征,提高了分割精度;Qi等^[4]提出的F-PointNet使用RGB(red, green, blue)图像与点云的结合作为输入,并使用PointNet作为分割网络从中分离出前景点,但PointNet使用最大池化层提取点云全局特征并保证点云不变性的同时,并不能很好地提取到局部特征,因此该算法分割精度不高,从而导致某些目标的检测结果不准确。近年来有些研究者^[5-7]采用从鸟瞰图(bird's eye view, BEV)点云中分割和检测目标的方法,Chen等^[8]提出了一个采用多视角和多种数据相结合的三维目标检测网络MV3D,将前视点云和俯视点云作为输入,避免了信息的丢失又减少了卷积的计算量;王张飞等^[9]采用不同的滤波去除不必要的点,再通过深度投影的方式将点云转为鸟瞰图点云进行目标分割和检测;Wang等^[10]提出了一种多视图三维检测算法,将三维点云投影到具有特定角度的4个视图并从中提取特征,从而保留了许多低级特征,检测速度达到0.05 s/帧。但这些基于鸟瞰图点云的检测方法不能很好地检测到小物体,而且激光雷达采集到的点云转换为鸟瞰图之后所需卷积运算量也增大。

综上所述,传统的基于点云的三维检测算法由于点云的稀疏性而导致其检测精度较低。因此,笔者提出一种新的基于边缘特征的三维目标识别算法。为了提高检测速度,本算法首先基于二维检测算法生成的二维候选区域过滤点云并生成视锥点云;然后在分割网络中计算点云之间的欧氏距离并将其作为点云的边缘特征输入多层感知机(multi-layer perceptron, MLP)中,在MLP提取特征的同时重新计算点云之间的欧氏距离并更新边缘特征,以此更好地提取局部特征并实现点云的分割;最后通过检测网络输出三维边界框的参数。

1 数据集和设备

1.1 数据集

采用自动驾驶场景数据集KITTI^[11] (Karlsruhe Institute of Technology and Toyota Technological

Institute)对提出的模型进行训练和测试。该数据集具有 7 481 张分辨率为 $1\,242 \times 375$ 的 RGB 图片和与之对应的 7 481 个三维点云数据。在图片和点云中,道路上的障碍物会因其他物体的遮挡而出现不同的遮挡率,为了更好地将数据分类并测试模型,根据 KITTI 官方对数据集难度的定义及不同的遮挡率将数据分为 3 个等级:物体被遮挡率不超过 15%即为简单难度,物体被遮挡率在 15%~30%之间为中等难度,物体被遮挡率超过 30%为困难难度。为了更好地测试模型的性能,按照约 7:2:1 的比例划分数据集,其中训练集有 5 236 个数据,验证集有 1 496 个数据,测试集有 749 个数据。训练集、验证集、测试集均包含了简单、中等、困难等级的图片和点云,用训练集作为输入数据训练本研究的模型及其他三维目标检测模型,并使用划分后的测试集对所有模型的性能进行评估。

1.2 设 备

数据处理系统主要用于车辆识别,并在配置如下的计算机上进行模型训练:显卡为 NVIDIA GTX 2 080Ti,处理器为 Intel(R)Core(TM)i7,内存为 32 GB,操作系统为 Win10。软件系统为 Pycharm2018, Python2.7, Tensorflow-gpu 1.14.0, Numpy 1.16.4, Mayavi 和 OpenCV 库。

2 试验方法

2.1 点云预处理

在点云的预处理中,先采用二维检测算法 Fast R-CNN^[12]对输入的 RGB 图片进行检测并生成二维候选区域,将输入的点云与二维候选区域中的像素点一一对应。如果三维点能够对应二维候选区域中的像素点,那么该三维点将被保留下来,否则被过滤掉,最终保留下来的点集合到一起并生成视锥体区域。在视锥内的点一定在相机的视野内,因此只检测在视锥范围内或者与视锥视野相交的物体。点云预处理的流程如图 1 所示。

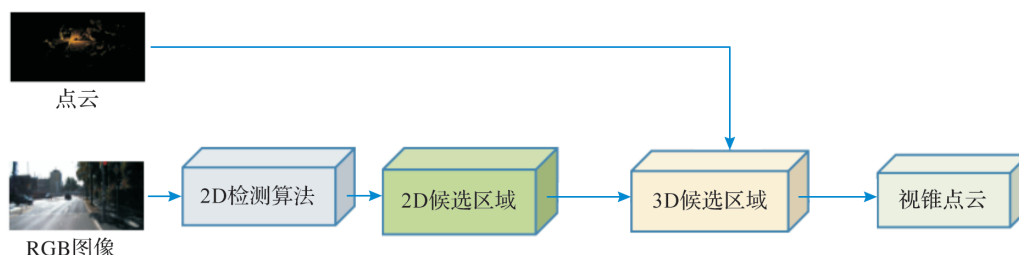


图 1 点云预处理的流程

Fig. 1 Flowchart of point cloud preprocessing

2.2 点云分割

基于边缘卷积的分割算法如图 2 所示,该分割神经网络由构造边缘特征的模块及多个 MLP 组成,同时也采用最大池化层提取全局特征。分割网络的输入数据为预处理后的视锥点云,在边缘特征中,采用 K 近邻算法确定目标点的局部特征图,并计算目标点与相邻点之间的欧氏距离,将该距离作为三维点的边缘特征并用于 MLP 的卷积计算。为了提取三维点云更多的细节特征,在网络中的较低层以更多的通道数来提取特征,如图 2 中的两层 MLP(64),卷积核大小均为 1×1 。在最大池化层之前的所有 MLP,在每次对点云进行卷积计算并提取特征后,都会重新计算新的局部特征图并构造新的边缘特征,再将新的边缘特征传入下一层 MLP,使得每个点与其他点之间的关系能在点云中扩散,以此改善 PointNet 只独立地处理点以维持置换不变性而导致的局部特征缺失的情况。点云的分割问题可以看作是对每个点的分类,本研究将分割后的点云分为两类,分别是杂乱点云和目标点云,并在分割的最终结果中输出它们对应的分类得分。欧氏距离计算如下:

$$D_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \quad (1)$$

三维空间中的欧氏距离是两点之间的实际距离。距离矩阵 (D_{ij}) 代表点云矩阵的第 i 个向量和第 j 个向量之间的欧氏距离。 x, y, z 表示三维空间中向量的坐标值。

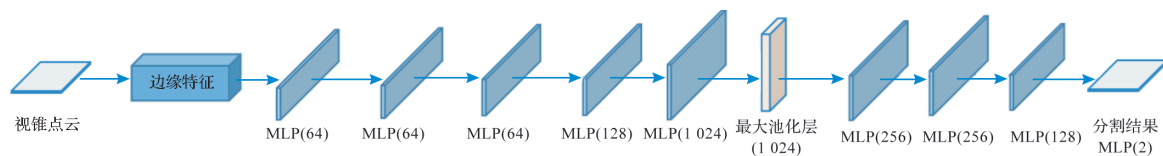


图2 基于边缘卷积的分割算法

Fig. 2 Segmentation algorithm based on edge convolution

2.3 目标检测

三维检测算法最终目的是识别并定位目标,因此采用检测网络估计目标的三维边界框并分类,最终输出三维边界框的参数,包括长度、宽度、高度、中心点的位置及边界框的方向。目标检测网络如图3所示,将分割网络得到的目标点云作为输入数据并送到检测网络,在MLP和最大池化层中进行多次卷积计算并回归后获得三维边界框的参数。本研究采用的回归方式与Faster R-CNN^[13]类似,将三维边界框回归到不同的尺寸和朝向,并以不同尺寸的三维边界框作为最终分类目标的依据。

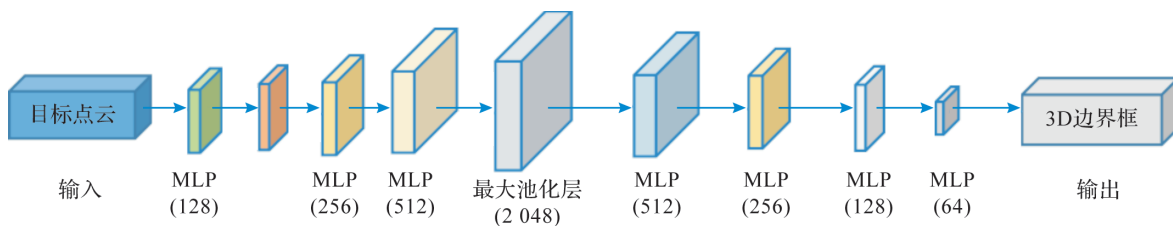


图3 目标检测网络

Fig. 3 Target detection network

3 结果与分析

3.1 点云预处理结果

点云预处理过程的可视化如图4所示。图4(a)是初始点云即未经任何处理的输入点云。图4(b)是输入的初始RGB图片。图4(c)是使用二维检测算法Fast R-CNN对RGB图片进行检测并生成对应的二维边界框的检测结果图,从图中可以看出有3辆车被检测出来。图4(d)是点云投影到图片上,并且点云中的三维点与二维候选区域中的像素点一一对应的可视化结果,在图中可以看到距离二维边界框较近的点颜色偏红;反之,距离二维边界框较远的点颜色偏绿,图中还出现了未被三维点覆盖的像素,这是由于激光雷达的扫描高度有限,因此无法覆盖所有像素。图4(e)是点云预处理后的最终结果的可视化图,从图中可以看出,预处理后的点云数量远远少于图4(a)中的点;图中的五边形边界框是为了凸显视锥区域的范围,使得视锥的可视化效果更明显。



图4 点云预处理过程的可视化

Fig. 4 Visualization of point cloud preprocessing process

3.2 试验细节参数

本试验设置的网络参数包括:输入的点云总点数为2 048,训练周期为200次,优化器为Adam,动量设置为0.9。为了训练模型,每次会取一小部分样本作为输入数据,每次取32个点云样本,即batch size为32。此外,在计算边缘特征的过程中,需要用K近邻算法确定相邻点的个数,试验验证后参数 k 的值设为6。在分割网络的全连接层使用了dropout层来防止模型过度拟合,一般取值为0.4~0.6,在本试验中该值为0.5。

3.3 分割结果

本研究使用 K 近邻算法计算欧氏距离并构造边缘特征。K 近邻算法的参数取值会影响边缘特征的构造,进而影响特征的提取效果和点云分割。为了准确选取一个合适的 k 值,对不同的 k 值进行了试验,试验后的结果见表 1。

在得到最终的检测结果之前,将本文方法与其他分割算法进行比较(表 2)。mesh 是以网格的形式表示点云, volume 是三维空间分割上的最小单位,类似像素,以立方体的形式来表示点云,而以 point 的形式处理点云是当前的主流形式,可直接将三维点云输入到卷积神经网络进行训练,本文的算法采用这种形式处理点云。

由表 2 可知,本文算法的分割精度比 multi-view segmentation 算法提高了 25.48 百分点,比 O-CNN 算法提高了 6.92 百分点,可见采用 point 形式作为输入能更好地处理点云。同时,本文算法的分割精度比 F-PointNet 提高了 2.30 百分点,测试结果证明本研究提出的构造边缘特征的方法能改善 PointNet 的局部特征缺失情况,实现更好的分割。

3.4 检测结果

在测试结果中随机选取 3 张点云图,从数据集中选取它们对应的初始点云图,并将这些点云图可视化,如图 5 所示。由图 5 可知,图 5(a)检测出了 4 辆汽车,图 5(b)检测出了 5 辆汽车和 1 个行人,这两个测试结果都正确地检测出了所有目标,图 5(c)中只检测出了 1 辆汽车和 4 个行人,有 1 个行人未被检测到,这可能是由于该行人的三维点云不够密集而导致的漏检。

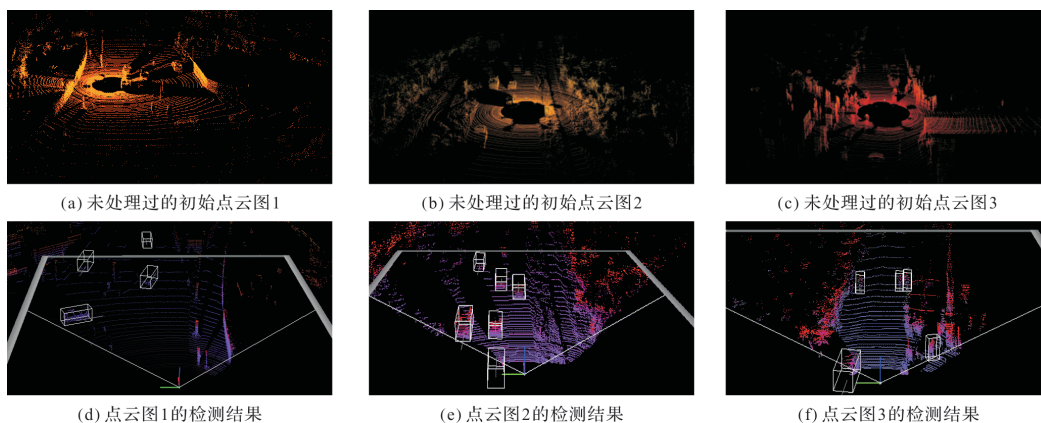


图 5 检测结果的可视化图

Fig. 5 Visualization of detection results

为了更好地对比本文算法与其他算法的性能,采取 KITTI 数据集的三维点云测试集对本文算法进行了验证,对汽车、自行车和行人的检测精度对比结果见表 3。检测的对象分别是汽车、自行车和行人,由于每种目标都存在着不同遮挡率,因此把数据划分为了简单、中等和困难 3 个等级并一一进行验证。由表 3 可知,相比于基于迭代算法的三维目标检测算法^[16],采用鸟瞰视角的 MV3D 算法及采用 PointNet 作为分割网络的 F-PointNet 算法,本研究提出的基于边缘卷积的三维目标识别算法在不同目标的识别精度上都有所提高。随着遮挡率的增大,目标的检测精度下降,其中行人的检测精度最低,这是因为行人点云的几何形状不明显,易与其他物体点云混淆在一起,不易过滤出来,且行人的三维点可能会由于太稀疏而出现误检或漏检等情况的发生,从而使得行人的检测精度远低于车辆和自行车。验证结果证明了本文算法实现了最佳检测性能,这说明本研究提出的边缘卷积有助于提取局部特征并提高检测精度。

表 1 不同的 k 值对分割精度的影响

Table 1 Effect of different k values on segmentation accuracy

k 的取值	分割精度/%
2	90.33
4	91.54
6	92.82
8	89.66
10	91.47

表 2 不同分割算法的分割精度对比

Table 2 Comparison of segmentation accuracy of different segmentation algorithms

算法	输入形式	分割精度/%
multi-view segmentation ^[14]	mesh	67.34
O-CNN ^[15]	volume	85.90
F-PointNet ^[4]	point	90.52
本文算法	point	92.82

表3 不同算法对汽车、自行车和行人的检测精度对比

Table 3 Comparison of detection accuracy of different algorithms for vehicles, bicycles and pedestrians %									
算法	汽车			自行车			行人		
	简单	中等	困难	简单	中等	困难	简单	中等	困难
迭代算法	65.80	46.11	39.39	—	—	—	—	—	—
MV3D	75.84	64.73	56.21	—	—	—	—	—	—
F-PointNet	83.52	71.45	61.32	74.13	57.85	49.88	52.77	44.56	39.04
本文算法	85.77	73.21	62.13	76.09	58.16	51.39	53.08	45.22	40.81

4 结 语

本研究提出了一种基于边缘卷积的三维目标检测算法,采用由激光雷达获取到的点云和由 RGB 摄像机采集到的图像作为输入数据来识别和定位道路上的障碍物。试验结果证明了本研究提出的基于边缘卷积的分割网络的有效性,且分割精度和检测精度相比于其他算法都有所提升。验证结果也说明了该方法存在一些尚待完善的地方,其中一个可能的扩展是将 RGB 图像更好地与点云融合,并且提高 RGB 图像的检测精度也能更好地提高模型的性能。另一种可能性是完善目标点云,使得目标的三维点足够密集,减少因三维点不足而导致的误检和漏检等情况的发生。我们在未来将更多地研究 RGB 图像与点云的融合,以获得更好的检测效果。

参考文献:

- [1] 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. Honolulu: IEEE,2017:657.
- [2] YAN Y, MAO Y, LI B, et al. Second: sparsely embedded convolutional detection[J]. Sensors,2018,18(10):3337.
- [3] WANG Y, SUN Y, LIU Z, et al. Dynamic graph CNN for learning on point clouds[J]. ACM Transactions on Graphics,2019,38(5):1.
- [4] QI C R, LIU W, WU C X, et al. Frustum PointNets for 3D object detection from RGB-D data[C]//IEEE Conference Computer Vision and Pattern Recognition. Salt Lake City: IEEE,2018:923.
- [5] 郑少武,李巍华,胡坚耀. 基于激光点云与图像信息融合的交通环境车辆检测[J]. 仪器仪表学报,2019,40(12):143.
- [6] ZHANG T, KAN Y, JIA H, et al. Urban vehicle extraction from aerial laser scanning point cloud data[J]. International Journal of Remote Sensing,2020,41(17):6670.
- [7] 张耀威,卞春江,周海,等. 基于图像与点云的三维障碍物检测[J]. 计算机工程与设计,2020,41(4):1169.
- [8] CHEN X Z, MA H M, WAN J, et al. Multi-view 3D object detection network for autonomous driving[C]//IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE,2017:1910.
- [9] 王张飞,刘春阳,隋新,等. 基于深度投影的三维点云目标分割和碰撞检测[J]. 光学精密工程,2020,28(7):1600.
- [10] WANG L, FAN X Y, CHEN J H, et al. 3D object detection based on sparse convolution neural network and feature fusion for autonomous driving in smart cities[J]. Sustainable Cities and Society,2020,54:102002.
- [11] GEIGER A, LENZ P, STILLER C, et al. Vision meets robotics: the kitti dataset[J]. The International Journal of Robotics Research,2013,32(11):1231.
- [12] GIRSHICK R. Fast R-CNN[C]//IEEE International Conference on Computer Vision. Santiago: IEEE,2015:1445.
- [13] REN S Q, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,39(6):1137.
- [14] LU Y H, ZHEN M M, FANG T. Multi-view based neural network for semantic segmentation on 3D scenes[J]. Science China(Information Sciences),2019,62(12):2.
- [15] WANG P S, LIU Y, GUO Y X, et al. O-CNN: octree-based convolutional neural networks for 3D shape analysis [J]. ACM Transactions on Graphics,2017,36(4):1.
- [16] 王康如,谭锦钢,杜量,等. 基于迭代式自主学习的三维目标检测[J]. 光学学报,2020,40(9):133.