

基于实例分割的目标三维位置估计方法

刘长吉, 郝志成, 杨锦程, 朱明, 聂海涛

引用本文:

刘长吉, 郝志成, 杨锦程, 等. 基于实例分割的目标三维位置估计方法[J]. 液晶与显示, 2021, 36(11): 1535–1544.

LIU Chang–ji, HAO Zhi–cheng, YANG Jin–cheng, et al. Object 3D position estimation based on instance segmentation[J]. *Chinese Journal of Liquid Crystals and Displays*, 2021, 36(11): 1535–1544.

在线阅读 View online: <https://doi.org/10.37188/CJLCD.2021–0069>

您可能感兴趣的其他文章

Articles you may be interested in

结构光3D点云的PIN针针尖提取

PIN tip extraction from 3 D point cloud of structured light

液晶与显示. 2021, 36(9): 1331–1340 <https://doi.org/10.37188/CJLCD.2020–0321>

基于笔划三维深度特征的签名识别

Signatures recognition based on strokes 3D depth feature

液晶与显示. 2019, 34(10): 1013–1020 <https://doi.org/10.3788/YJYXS20193410.1013>

时空域深度学习火灾烟雾检测

Spatio–temporal deep learning fire smoke detection

液晶与显示. 2021, 36(8): 1186–1195 <https://doi.org/10.37188/CJLCD.2020–0239>

基于ROS与三维点云图像的室内物体精准定位

Precise location of indoor objects based on ROS and point cloud images

液晶与显示. 2019, 34(6): 598–604 <https://doi.org/10.3788/YJYXS20193406.0598>

无人飞行器自主降落区识别方法研究

Method for identifying the landing area of unmanned aerial vehicle

液晶与显示. 2018, 33(3): 238–244 <https://doi.org/10.3788/YJYXS20183303.0238>

文章编号:1007-2780(2021)11-1535-10

基于实例分割的目标三维位置估计方法

刘长吉^{1,2}, 郝志成^{1,2*}, 杨锦程³, 朱 明^{1,2}, 聂海涛¹

(1. 中国科学院 长春光学精密机械与物理研究所, 吉林 长春 130033;

2. 中国科学院大学, 北京 100049;

3. 重庆嘉陵华光电科技有限公司, 重庆 400000)

摘要: 三维目标检测在实际工程应用中的难点在于深度传感器价格高、点云质量差、缺少丰富的纹理信息、三维数据训练集制作困难。对此本文提出一种基于实例分割的三维目标位置估计方法, 可以用在多种传感器中, 如相机-雷达, 双目相机等。首先在二维图像下对目标进行实例分割, 根据目标的分割掩码提取出目标的深度图像与 RGB 图像融合转化为粗略点云, 最终进行异常噪声点去除, 得到精细的目标点云。在 KITTI 数据集上进行了测试, 平均精度值(AP)可以达到 50%, 表明该方法可以准确地估计到目标位置信息。本文提出的方法无需三维数据训练集, 可以快速准确地进行三维物体点云的提取, 仅使用二维检测器就可以达到三维物体检测的目的。

关键词: 点云分割; 三维目标检测; 实例分割; 异常检测; 位置估计; 深度学习

中图分类号: TP391.4; TH691.9 **文献标识码:** A **doi:** 10.37188/CJLCD.2021-0069

Object 3D position estimation based on instance segmentation

LIU Chang-ji^{1,2}, HAO Zhi-cheng^{1,2*}, YANG Jin-cheng³, ZHU Ming^{1,2}, NIE Hai-tao¹(1. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences,
Changchun 130033, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China;

3. ChongQing JiaLing HuaGuang Photoelectric Technology Co., Ltd.,
Chongqing 400000, China)

Abstract: The difficulty of 3D target detection in practical engineering applications lies in the high price of depth perception equipment, poor point cloud quality, lack of rich texture information, difficulty in creating 3D data training sets. This paper proposes a three-dimensional target position estimation method based on instance segmentation. It can be used in a variety of sensors, such as camera-radar, binocular camera, etc. Firstly, the target segmentation is performed under the two-dimensional image, the target's depth image are extracted and RGB image according to the target segmentation mask, and it is converted into a rough point cloud. Finally, the abnormal noise points is removed to obtain a fine target point cloud. Tested on the KITTI data set, the AP can reach 50%. The results

收稿日期: 2021-03-12; **修订日期:** 2021-03-23.

基金项目: 广东省重点领域研发计划项目 (No.2019B010155003); 吉林省教育厅“十三五”科学技术项目 (No. JJKH20200780KJ)

Supported by Guangdong Key Areas R&D Program Projects (No.2019B010155003); Thirteenth Five-Year Plan Science and Technology Project of Jilin Provincial Education Department (No. JJKH20200780KJ)

* 通信联系人, E-mail: hzc972513@sina.com

show that this method can accurately estimate the target location information. The method proposed in this paper does not need 3D data training set can quickly and accurately extract the point cloud of three-dimensional objects, and only use a two-dimensional detector to achieve the purpose of three-dimensional object detection.

Key words: point cloud segmentation; 3D object detection; instance segmentation; anomaly detection; position estimation; deep learning

1 引言

基于点云数据的三维物体检测的难点在于深度感知设备价格高、高质量的稠密点云获取困难且运算规模大、点云数据集制作困难。经典点云分割算法例如 LCCP^[1]、区域生长^[2]、RANSAC^[3]等,效果较差,且速度慢。近些年已经有学者提出使用深度学习进行三维物体检测,如 PointNet^[4-5]、F-PointNet^[6]等。其中 F-PointNet 提出使用基于二维检测器的三维物体检测,减少对点云的搜索空间。PointNet 直接使用全部点云信息进行分割,达到了非常不错的效果。但是基于深度学习的三维物体检测都需要使用三维点云数据作为训练样本。而在实际的工程应用当中,很难获得大量的、高质量的点云数据作为训练样本。

在图像中存在丰富的目标信息,且二维图像训练集获取容易。如今成熟的二维检测算法,例如 Yolo^[7]、Faster R-CNN^[8]等算法已经得到了广泛的应用,准确率高且速度快。经典的实例分割算法 Mask R-CNN^[9]其分割效果已经满足大部分的应用场景。因此本文提出了一种基于实例分割的三维目标位置估计方法,首先得到目标分割掩码,提取深度信息,由此获得目标粗略点云。再采用统计学异常点检测,对混入的极少量异常点进行剔除,最终得到精细的目标点云。

经过试验,该方法可以得到准确的目标物体点云,实现简单、成本低,无需任何点云数据作为训练样本。适合在目标点云提取、三维物体检测、障碍物检测等工程任务当中使用。

2 基于实例分割的目标粗略点云提取

2.1 基于 Mask R-CNN 的目标实例分割

目标检测(Object detection)需要获取图像中

目标的位置(Bounding box),还需要获取物体的类别。实例分割(Instance segmentation)在目标检测的基础上,需要对每一个像素的分类,同时获得分割掩码,如图 1 所示。

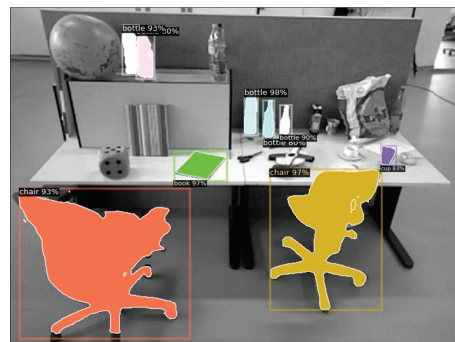


图 1 基于 Mask R-CNN 的实例分割结果

Fig.1 Instance segmentation result based on Mask R-CNN

Mask R-CNN 是一种通用图像实例分割算法,其沿用了 Faster R-CNN^[10]的思想,并且在 Faster R-CNN 的结构上增加了全卷积网络进行掩码预测^[11],并将感兴趣区域池化(ROI pooling)替换成感兴趣区域对齐(RoI Align),使特征图与原图更加准确地对准,在区域推荐的过程中,减少了量化而损失的像素偏移,使得掩码更加准确。最后通过增加全卷积网络对推荐区域进行语义分割。不管是实例分割或者语义分割,由于其最后一层的分割掩码特征图一般为 16 倍下采样,直接上采样到目标尺寸大小会造成分割边缘的不准确。在 2.2 中我们会描述分割边缘不准

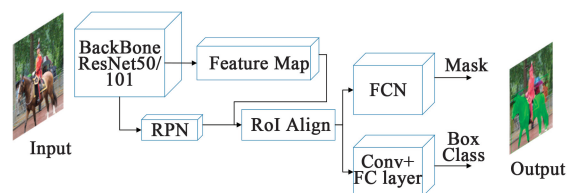


图 2 Mask R-CNN 网络结构

Fig.2 Structure of Mask R-CNN

确对点云分割造成的影响。Mask R-CNN 网络结构如图 2 所示。

2.2 RGB 图像与深度图融合

使用成熟的二维检测器在 RGB 图像中对目标进行实例分割,得到目标掩码和类别。不论是使用双目相机 TOF 相机或者是相机与雷达融合,都可以获得 RGB 图像与深度信息。如图 3 所示。



图 3 深度图像

Fig.3 RGB image and depth image

本文使用深度图来表示深度信息,深度图与 RGB 图像尺寸一致,每一个像素上储存的是与其对应 RGB 图像每一个像素的距离,使用单通道无符号 16 位整形(Uint16)来保存深度信息,单位取 mm,如图 4 所示。



图 4 使用分割掩码得到的 RGB 信息与深度信息

Fig.4 Mask,RGB and depth from instance segmentation.

根据相机成像模型,物体投影在像平面上生成图像,此时从世界坐标系转换到像素坐标系,当拥有深度信息时,可从像素坐标系转化到相机坐标系。相机成像模型可简化为图 5。

空间中一点可以分别在世界坐标系 $O_w X_w Y_w Z_w$,相机坐标系 $O_c X_c Y_c Z_c$,图像坐标系 $o-x-y$,与像素坐标系 $u-v$ 中描述。世界坐标系中一点 P 转化到像素坐标系的过程如下:

从世界坐标系到相机坐标系:

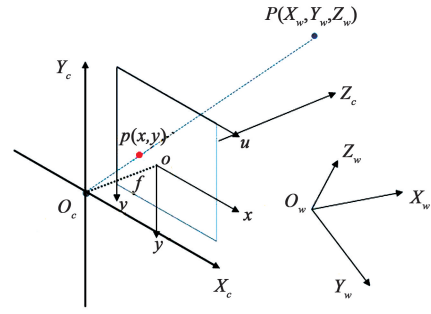


图 5 简单相机成像模型

Fig.5 Simple camera model

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix}, \quad (1)$$

其中: $\mathbf{R}$ 为旋转矩阵, $\mathbf{T}$ 为平移矩阵。

从相机坐标系到理想图像坐标系:

$$Z_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & 0 \\ 0 & f_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix}, \quad (2)$$

其中: f_x, f_y 为焦距。

假设像素坐标系原点在图像坐标系下的坐标为 (u_0, v_0) , 每个像素点在图像坐标系 x 轴、 y 轴方向的尺寸为 d_x, d_y , 且像点在实际图像坐标系下的坐标为 (x, y) , 于是可得到像点在像素坐标系下的坐标为:

从实际图像坐标系到像素坐标系:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{d_x} & 0 & u_0 \\ 0 & \frac{1}{d_y} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \quad (3)$$

式(4)被称为内参矩阵:

$$\mathbf{M} = \begin{bmatrix} \frac{f_x}{d_x} & 0 & u_0 \\ 0 & \frac{f_y}{d_y} & v_0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (4)$$

根据式(1)、(2)、(3)、(4)可以得到图像坐标系到相机坐标系的转换公式:

$$\begin{cases} X_c = \text{depth} \cdot (u - u_0) \cdot d_x / f_x \\ Y_c = \text{depth} \cdot (v - v_0) \cdot d_y / f_y \\ Z_c = \text{depth} \end{cases}, \quad (5)$$

其中 depth 为深度图中储存的深度信息,根据式

(5) 可以将 RGB 图像与深度图像融合得到点云数据, 其点云效果如图 6 所示。

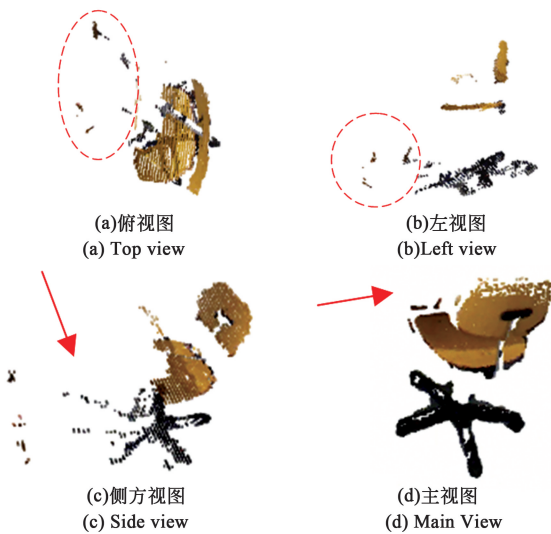


图 6 不同视角下椅子点云中混入的非目标点云^[12]

Fig.6 Outliner point cloud in different view



图 7 分割边缘不准确

Fig.7 Segmentation edge inaccuracy

由于分割边缘不准确性^[13], 如图 7 所示, 导致得到的分割掩码中混入背景, 从而造成点云中混入一定量的异常点。异常点云一般都与实际目标有一定距离, 这些非目标点云会严重影响物体位置的判断, 有必要将非目标点云进行去除。

3 点云主成分提取

3.1 自适应统计学异常点检测

由于每个目标形状差异不同, 仅使用图像分割获取到的点云区别也很大, 有些异常点云占比很大, 有些异常点云占比很少, 甚至有些规则形状的目标不存在异常点云的混入。异常点云有以下特点: (a) 异常存在性未知, 即是否存在异常点云;

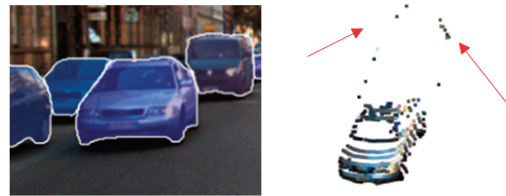
(b) 异常占比未知性, 即不同情况不同目标, 基于图像分割得到的点云中, 异常点云占比是不确定的; (c) 异常分布未知性, 即异常点云所处的位置是不确定、异常点云簇数量也不确定, 可能存在多个点云数量较少的点云簇^[14], 分布在不同的位置。

由于以上 3 点特性, 使得异常点的判断变得十分困难, 并且由于异常占比未知性, 使得异常点云的去除需要手动设置异常率的占比, 这在同一场景多目标的情况下是不可用的。图 8 所示为在复杂室外场景下, 不同目标混入的异常点云分布。



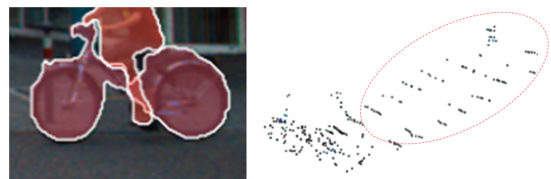
(a) 室外分割掩码与深度图

(a) Outdoor segmentation mask and depth map



(b) 汽车分割掩码与粗略点云(异常占比少)

(b) Car segmentation mask and coarse point cloud (few abnormal points)



(c) 自行车分割掩码与粗略点云(异常占比多)

(c) Bicycle segmentation mask and coarse point cloud (more anomalous point clouds)

图 8 不同目标混入的异常点云分布

Fig.8 Distribution of abnormal point clouds blended in different objects

所以我们需要一种可以不指定异常占比、依赖点云形态分布来提取目标点云的算法。通过观

察我们发现,异常点云与真实目标点云的分布存在明显不同,最明显的是数量与密度的不同。

统计学异常点移除 (Statistical Outlier detect, SOD)^[15] 是一种无监督学习方法,可以应用在数据异常数据检测、工业产品检测等领域。该算法有以下特点:(a)无需指定异常点占比;(b)无监督学习,无需训练样本。适用于连续数据的异常检测,检测分布稀疏且离密度高的群体较远的点、容易被孤立的离群点。

已有检测数据集 $X = \{x_1, x_2, x_3 \cdots x_n\}$, $\forall x_i \in X$, 且 x_i 由 m 维度的特征构成, $x_i = \{x_{i1}, x_{i2}, x_{i3} \cdots, x_{im}\}$ 。在点云数据中,只保存位置信息,则 $m=3$, 则 $x_i = \{x, y, z\}$ 。对每个点,我们计算它到所有 k 个临近点的平均距离。得到所有点的平均距离向量 $D = \{d_1, d_2, d_3 \cdots d_n\}$ 。假设得到的结果是一个高斯分布,其形状由均值 μ 和标准差 s 决定。当某一点的平均距离 $d_i > n \cdot s + \mu$ 时,可被定义为离群点并可从数据集中去除掉。算法具体流程如表 1 所示。

表 1 异常点去除算法流程图

Tab.1 Outlier removal algorithm

自适应统计学异常点检测的主成分提取

输入: X ——点云数据; t ——分块系数; n —— n 倍方差

输出:异常点云

1:计算点云总数量 a

2:取 k 个点参与平均距离计算,其中 $k = a/t$

3:构建二叉树索引

4:计算每个点到 k 临近点的平均距离

5:得到距离向量 $D = \{d_1, d_2, d_3 \cdots d_n\}$

6:遍历每一个点的平均距离 d_i :

如果 $d_i > n \cdot s + \mu$:

该点被判定为异常

否则:

该点被判定为正常

7:去除异常点,得到目标点云

3.2 k 与 n 的选取

k 值代表计算一个其最临近 k 个点,其中 k

值的大小决定了 SOD 对异常点的敏感程度。 k 越小, SOD 对局部特征越敏感。对于目标来说,异常点占比通常很小,也会出现局部聚集的情况,所以 k 不宜取过小的数。一般对于一个目标来说,假设其点云数量为 a , 令 $k = a/t$, 其中 t 代表将整个目标当做 t 个部分来考虑。我们使用 t 来控制 SOD 对整体与局部敏感度。

在计算 k 近邻平均距离后,每个点会形成新的映射关系,异常点会被映射到离散区域,如图 9 所示。SOD 去除异常点效果如图 10 所示。最终

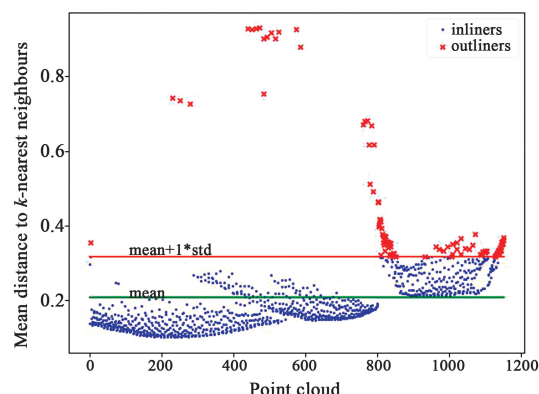


图 9 $t=3, n=1$ 时,点云 k 临近距离统计。

Fig.9 Point cloud k proximity statistics at $t=3, n=1$.

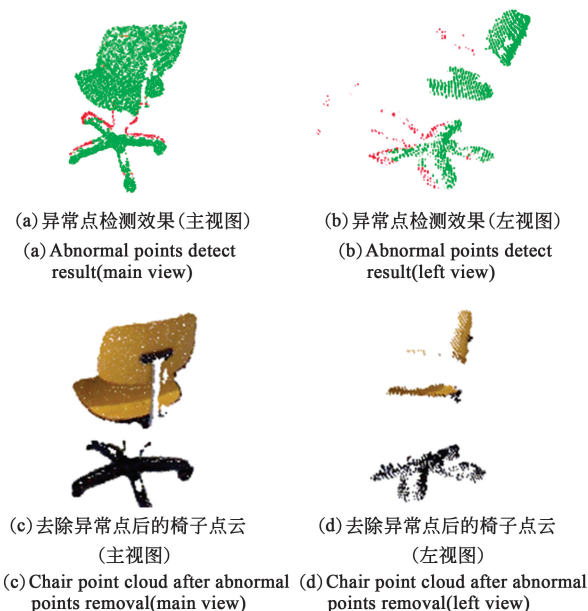


图 10 自适应 SOD 去除异常点云,红色点被判定为异常点。

Fig.10 Adaptive SOD removes the abnormal point cloud, and the red points are determined as abnormal points.

将异常点进行移除,得到目标精细点云,同时获得了目标的准确的空间位置信息。

4 实验结果与分析

实验分别使用了 TUM 数据集^[16]与 KITTI 数据集^[17-18]。TUM 数据集由尼黑工业大学的 Computer Vision Lab 制作。采集设备使用 Kinect,主要针对室内的办公室场景,于是使用飞行时间法(TOF),点云较为稠密。KITTI 数据集由德国卡尔斯鲁厄理工学院和丰田美国技术研究院联合创办,是目前国际上最大的自动驾驶场景下的计算机视觉算法评测数据集。该数据集用于评测立体图像(Stereo)、光流(Optical flow)、视觉测距(Visual odometry)、3D 物体检测(Object detection)和 3D 跟踪(Tracking)等计算机视觉技术在车载环境下的性能。3D 物体检测包含 7 481 张训练图片和其对应的点云。由于 KITTI-3D object 数据集中物体只有三维框与二维框标注,并无实例分割标注,所以我们使用了 COCO 作为训练集。COCO 中包含类别人与车,并使用 KITTI 训练集作为验证集进行性能评估。

3D 目标检测需要同时实现目标定位和目标识别两项任务。通过比较预测框与真值框之间的

交并比(Intersection over Union, IOU)来计算重叠程度。通过置信分数和 IOU 来确定检测结果的性能,最终使用单类别平均精度值(Average precision, AP)来评估单类目标检测的结果如图 11 所示。其中 2D 边界框的性能也使用 AP 来衡量,IOU 阈值使用 KITTI 标准。由于算法机制不同,我们估算物体位置使用了最小外包立方体,而远处目标的点云数量稀少且只有物体表面被激光雷达捕捉到,所以在远处物体的 3D 预测边界框会比真值 3D 边界框小很多,所以随着 IOU 增加 AP 会降低,如表 2 所示。在表 3 中,我们对比了不同算法在 KITTI 上的 AP 得分,经比较,我们的算法可以只使用 2D 数据集进行训练,在 AP 得分上,与主流的算法比较有一定差距。在测量精度上,当前主流的深度感知设备激光雷达 Velodyne HDL-64E 误差 < 2 cm,深度相机 RealsenseD435i 在 < 4 m 的距离下误差不超过 2%。

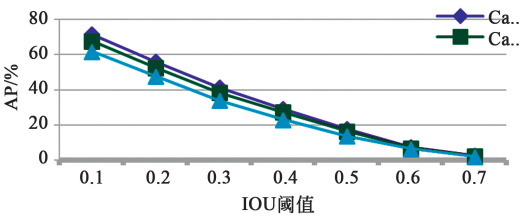


图 11 不同 IOU 阈值下的车辆 AP 值(3D)
Fig.11 Car AP at different IOU thresholds (3D)

表 2 不同 IOU 阈值下的 KITTI 数据集上的 AP
Tab.2 AP on KITTI dataset with different IOU threshold (%)

Task	IOU						
	0.1	0.2	0.3	0.4	0.5	0.6	0.7
Car Easy	71.67	56.21	41.24	29.29	17.56	7.28	2.48
Car Moderate	67.88	52.67	38.51	27.44	16.49	6.82	2.36
Car Hard	61.93	47.94	34.05	23.28	13.75	6.6	2.2
Pedestrian Easy	64.46	43.93	17.98	5.95	1.2	0.144	0.01
Pedestrian Mod	58.01	38.86	16.76	5.11	0.86	0.152	0.01
Pedestrian Hard	51.83	33.36	14.36	4.27	0.86	0.152	0.01

表 3 在 KITTI 数据集上的性能对比

Tab.3 Performance comparison on KITTI dataset

(%)

Method	Benchmark	Easy	Moderate	Hard	训练数据集格式
Ours	Car (Detection)	69.46	64.41	51.89	2D
	Car (3D Detection@IOU:0.2)	56.21	52.63	47.94	
	Pedestrian (Detection)	65.68	59.11	52.56	
	Pedestrian(3D Detect@IOU:0.2)	43.93	38.86	33.36	
F-PointNet ^[6]	Car (Detection)	95.85	95.17	85.42	2D+3D
	Car (3D Detection)	81.20	69.79	60.59	
	Pedestrian (Detection)	89.83	80.13	75.05	
	Pedestrian (3D Detection)	51.21	44.89	40.23	
Complex-YOLO ^[19]	Car (Detection)	91.92	84.16	79.62	3D
	Car (3D Detection)	55.93	47.34	42.60	
	Pedestrian (Detection)	42.16	36.45	32.91	
	Pedestrian (3D Detection)	17.60	13.96	12.70	
Point R-CNN ^[20] (LiDAR only)	Car (3D Detection)	85.94	75.73	68.32	3D
	Pedestrian (3D Detection)	49.43	41.78	38.63	
MV3D ^[21]	Car (3D Detection)	74.97	63.63	54.00	3D

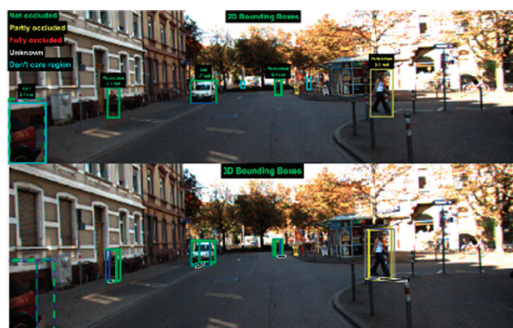
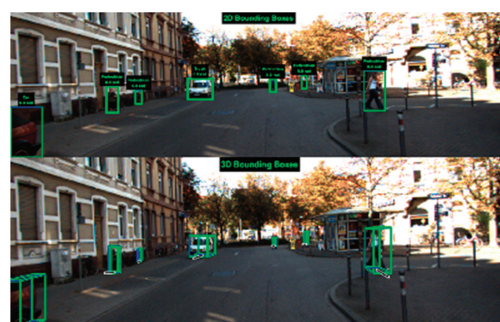
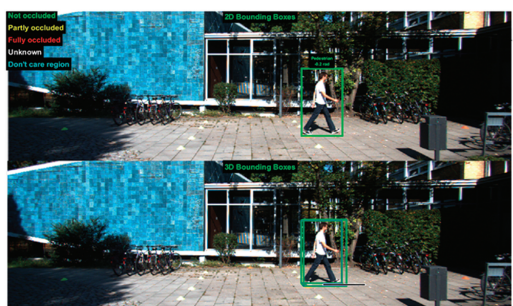
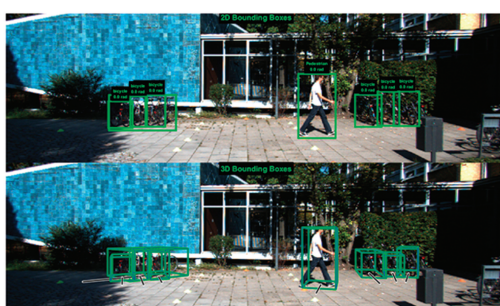
(a) 真值
(a) Ground truth(b) 检测结果
(b) Detect result(c) 真值
(c) Ground truth(d) 检测结果
(d) Detect result

图 12 在 KITTI 上的检测结果与真值对比

Fig.12 Detection results vs. ground truth on KITTI

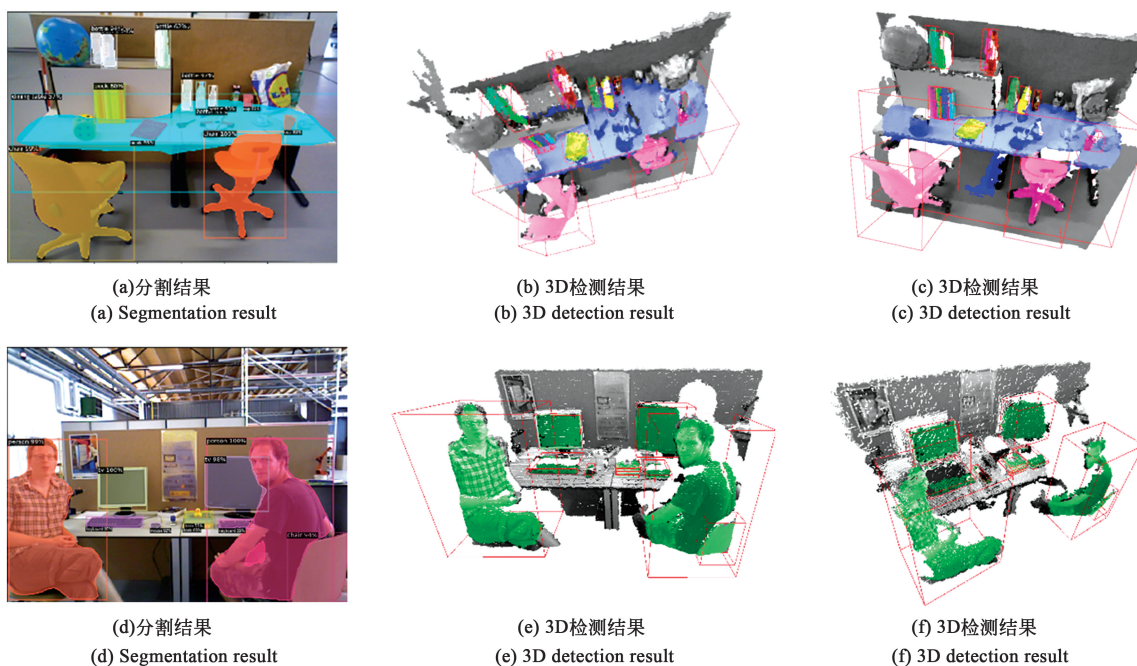


图 13 TUM 数据集上的检测结果

Fig.13 Results of TUM dataset

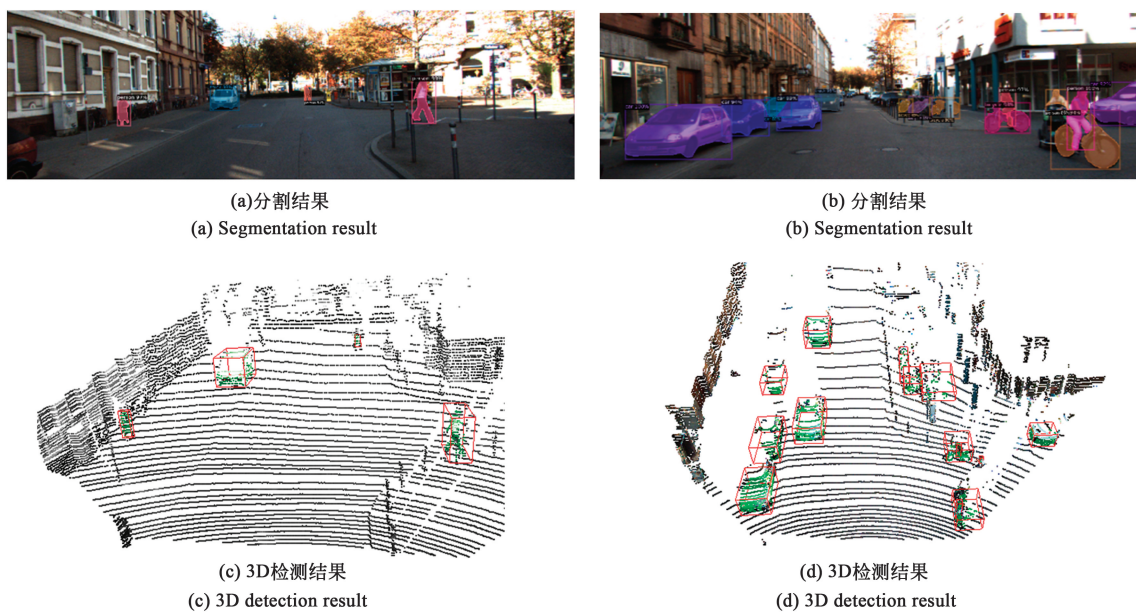


图.14 KITTI 数据集上的检测结果

Fig.14 Detection results of KITTI dataset

5 结 论

使用二维检测器驱动的三维物体检测可以得到非常好的效果,经过试验,在 KITTI 数据集上

的 AP 可以达到 50% 以上。以 Mask R-CNN 作为检测器,在 2080 Ti 上可以达到 15 FPS。

对于远距离、重叠、复杂的目标来说,点云已经无法分辨其轮廓形状,二维检测器可以提供一個强有力的前端策略,来获取目标的点云。同时,本文

使用了无监督学习来对异常点云进行检测,整个算法不需要任何的三维数据作为训练样本,轻量化,更加易于部署应用于各个环境。其中轻量化体现在硬件轻量化与训练集轻量化。硬件轻量化是指以图像为主,对点云质量要求不高,可以降低硬件成本。训练集轻量化是指仅使用了二维数据集,相比三维点云数据集,容易获取容易标注。

对于二维驱动的三维物体检测来说,其前端

可以搭载强大的二维检测器以达到更多功能的拓展。比如骨架识别、人体姿态识别、全景分割等。并且目前主流的三维目标检测方法都使用了RGB数据与点云数据的多模态融合,这也是三维物体检测的发展趋势。RGB提供丰富的语义信息,而点云数据提供准确的位置、轮廓信息。多模态的融合可以优势互补,但网络也会变得复杂,也更难应用于实时处理当中。

参 考 文 献:

- [1] STEIN S C, SCHOELER M, PAPON J, *et al.* Object partitioning using local convexity [C]//*Proceedings of the 2014 IEEE Computer Vision and Pattern Recognition*. Columbus, OH, USA: IEEE, 2014.
- [2] RABBANI T, VAN DEN HEUVEL F, VOSELMAN G. Segmentation of point clouds using smoothness constraint [J]. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2006, 36(5): 248-253.
- [3] FISCHLER M A, BOLLES R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography [J]. *Communications of the ACM*, 1981, 24(6): 381-395.
- [4] QI CHARLES R, SU H, KAICHUN M O, *et al.* PointNet: deep learning on point sets for 3D classification and segmentation [C]//*Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA: IEEE, 2017.
- [5] QI C R, YI L, SU H, *et al.* PointNet++: deep hierarchical feature learning on point sets in a metric space [J]. *arXiv*:1706.02413, 2017.
- [6] QI C R, LIU W, WU C X, *et al.* Frustum PointNets for 3D object detection from RGB-D data [C]//*2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA: IEEE, 2018: 918-927.
- [7] 任彬,王宇庆,聂海涛,等. 基于 MPSOC 的航空图像目标检测系统设计 [J]. *液晶与显示*, 2020; 1-12, doi:10.37188/CJLCD.2020-0310.
REN B, WANG Y Q, NIE H T, *et al.* Design of aerial image target detection system based on MPSOC [J]. *Chinese Journal of Liquid Crystals and Displays*, 2020; 1-12, doi: 10.37188/CJLCD.2020-0310. (in Chinese)
- [8] 李娜,姜志,王军,等. 基于 Faster R-CNN 的仪表识别方法 [J]. *液晶与显示*, 2020, 35(12): 1291-1298.
LI N, JIANG Z, WANG J, *et al.* Instrument recognition method based on Faster R-CNN [J]. *Chinese Journal of Liquid Crystals and Displays*, 2020, 35(12): 1291-1298. (in Chinese)
- [9] HE K M, GKIOXARI G, DOLLÁR P, *et al.* Mask R-CNN [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(2): 386-397.
- [10] REN S Q, HE K M, 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): 1137-1149.
- [11] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 39(4): 640-651.
- [12] ZHOU Q Y, PARK J, KOLTUN V. Open3D: A modern library for 3D data processing [J]. *arXiv*:1801.09847, 2018.
- [13] KIRILLOV A, WU Y X, HE K M, *et al.* PointRend: image segmentation as rendering [C]//*Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE, 2020.
- [14] 马鑫,魏仲慧,何昕,等. 三维枪弹痕点云数据处理及特征提取研究 [J]. *液晶与显示*, 2016, 31(9): 889-896.
MA X, WEI Z H, HE X, *et al.* Processing and feature extraction for three-dimensional bullet point cloud data [J]. *Chinese Journal of Liquid Crystals and Displays*, 2016, 31(9): 889-896. (in Chinese)

- [15] CHHIKARA R S, HORTON C L, MENDLOWITZ M A, *et al.* Statistical outlier detection (SOD): A computer program for detecting outliers in data [R]. NTRS. NASNA, 1980.
- [16] STURM J, ENGELHARD N, ENDRES F, *et al.* A benchmark for the evaluation of RGB-D SLAM systems [C]//*Proceedings of 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Vilamoura-Algarve, Portugal: IEEE, 2012.
- [17] GEIGER A, LENZ P, STILLER C, *et al.* Vision meets robotics: The KITTI dataset [J]. *The International Journal of Robotics Research*, 2013, 32(11): 1231-1237.
- [18] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? The KITTI vision benchmark suite [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Providence, RI, USA: IEEE, 2012.
- [19] SIMON M, MILZ S, AMENDE K, *et al.* Complex-YOLO: real-time 3D object detection on point clouds [J]. *arXiv*:1803.06199, 2018.
- [20] SHI S S, WANG X G, LI H S. PointRCNN: 3D object proposal generation and detection from point cloud [C]//*2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, CA, USA: IEEE, 2019: 770-779.
- [21] CHEN X, MA H, WAN J, *et al.* Multi-view 3D object detection network for autonomous driving [C]//*Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA: IEEE, 2017.

作者简介:



刘长吉(1990—),男,辽宁营口人,硕士研究生,2017年于辽宁工业大学获得学士学位,主要从事三维物体检测方面的研究。E-mail: liuchangji18@mails.ucas.ac.cn



郝志成(1978—)男,辽宁营口人,博士,研究员,2007年于中国科学院长春光学精密机械与物理研究所获得博士学位,主要从事数字图像处理、自动目标识别、图像跟踪与测量方面的研究。E-mail: hzc972513@sina.com