



(12) 发明专利

(10) 授权公告号 CN 109523552 B

(45) 授权公告日 2021. 11. 02

(21) 申请号 201811245343.9

G06T 7/50 (2017.01)

(22) 申请日 2018.10.24

G06T 7/90 (2017.01)

(65) 同一申请的已公布的文献号

G06K 9/00 (2006.01)

申请公布号 CN 109523552 A

G06N 3/08 (2006.01)

(43) 申请公布日 2019.03.26

(56) 对比文件

(73) 专利权人 青岛智能产业技术研究院

CN 108460779 A, 2018.08.28

地址 266111 山东省青岛市高新区智力岛
路一号创业大厦B座26楼

CN 108520535 A, 2018.09.11

CN 108257139 A, 2018.07.06

(72) 发明人 沈大勇 王晓 刘胜 胡加媛

王杰 翟天亨

Charles R. Qi et al..Frustum

PointNets for 3D Object Detection from
RGB-D Data.《arXiv》.2017,第1-15页.

(74) 专利代理机构 青岛华慧泽专利代理事务所

(普通合伙) 37247

审查员 李振娟

代理人 刘娜

(51) Int. Cl.

G06T 7/10 (2017.01)

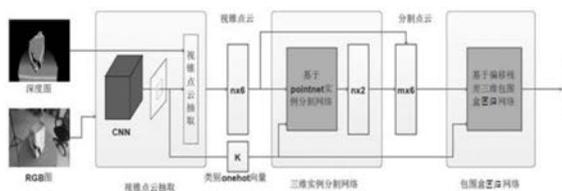
权利要求书2页 说明书8页 附图4页

(54) 发明名称

基于视锥点云的三维物体检测方法

(57) 摘要

本发明公开了一种基于视锥点云的三维物体检测方法,采集RGB图像和点云图像并进行标注,获取标注后的RGB图像数据集和点云图像数据集,将数据集作为训练样本和测试样本。主要由三个网络组成:基于二维物体检测的视锥点云抽取网络、基于Point net三维物体实例分割网络以及基于偏移残差的三维包围盒回归网络。三个网络依次训练,前一个网络的训练输出作为下一个网络的输入。设置各神经网络模型的超参数,通过Tensorflow训练三维物体检测网络模型,当代价损失函数减少到理想程度且训练达到所要求的最大迭代次数时,生成训练模型;将点云图像数据集输入至训练完毕的模型中,输出三维物体检测结果。本发明比现有三维物体检测方法更完善、难度更高。



1. 基于视锥点云的三维物体检测方法,其特征在于,包括如下步骤:

步骤一:对RGB图像和点云图像进行标注,获取标注后的RGB图像数据集和点云图像数据集,将RGB图像数据集和点云图像数据集作为基于视锥点云的三维物体检测网络模型的训练样本和测试样本;

步骤二:构建基于二维物体检测的视锥点云抽取网络,将RGB图像数据集作为基于二维物体检测的视锥点云抽取网络的输入,利用RGB信息来对物体进行检测,从而获取感兴趣物体的二维包围盒,再通过对齐和配准后RGB图像与深度图像,得到检测出来的二维物体对应的视锥点云数据;

步骤三:构建基于Point net三维物体实例分割网络,将步骤二获取的视锥点云数据作为基于Point net三维物体实例分割网络的输入,利用Point net网络对物体进行二元的实例分割,获得实例分割后的物体点云数据;

步骤四:构建基于偏移残差的三维包围盒回归网络,将步骤三获得的实例分割后的物体点云数据作为基于偏移残差的三维包围盒回归网络的输入,回归物体三维包围盒的八个顶点与其真实值之间的偏移量,通过计算即可得物体的9自由度三维包围盒对应的八个顶点,从而得到其三维包围盒;

步骤五:将RGB图像数据集和点云图像数据集输入至步骤四训练完毕的三维物体检测网络模型中,输出三维物体检测结果;

其中,基于偏移残差的三维包围盒回归网络的构建方法为:

a) 利用步骤三中得到的实例分割后物体点云数据,经过一个基于Point net的特征抽取网络A,得到每个点的局部特征和全局特征,得到全局特征后,输出有两个不同分支操作:第一个分支是进入基于偏移残差的三维包围盒回归网络B,第二个分支是进入直接回归三维包围盒的八个顶点的网络C,然后将这连个分支的输出结果进行对比;

b) 第一个分支,是将特征抽取网络A得到的全局特征和局部特征以及二维物体检测得到的先验知识one-hot向量进行连接,再经过5层MLP层操作,输出点云中第i个点预测包围盒8个顶点相对第i个点坐标的偏移值 p_{offset}^i 以及第i个点所预测的三维包围盒的得分 P_i ,选取所有点中预测的得分最高的三维包围盒作为网络的输出;然后通过真实值与偏移量计算得出预测的三维包围盒的八个顶点坐标 C_i ,其计算方法如下:

$$C_i = p_i + p_{offset}^i;$$

c) 第二个分支是直接特征抽取网络A得到的全局特征与二维检测得到的先验知识one-hot向量进行连接,经过3个MLP层操作,得到预测的三维包围盒的八个顶点坐标作为网络的输出;

基于偏移残差的三维包围盒回归网络B的损失函数为:

$$L_{reg} = \frac{1}{m} \sum_i offset_ratio_i + L_{conf};$$

其中,m是实例分割点云的数目;offset_ratio_i是第i个点预测的包围盒的偏移距离比率,偏移距离比率即预测顶点与真实顶点的偏移距离和真实包围盒对角线长度的比例,其计算方法为:

$$\text{offset_ratio} = \frac{\text{corner_distance}}{\text{diagonal_length}} ;$$

L_{conf} 是预测包围盒置信度的损失,其计算方法为:

$$L_{\text{conf}} = \frac{1}{m} \sum_i (c_{\text{pred}} - c_{\text{gt}})^2 ;$$

其中, m 是实例分割点云的数目, c_{pred} 代表预测的物体三维包围盒的置信度, c_{gt} 代表物体真实三维包围盒的置信度;

对于直接回归三维包围盒的八个顶点的网络C的损失函数为:

$$L_{\text{reg}} = \text{offset_ratio}.$$

2.根据权利要求1所述的基于视锥点云的三维物体检测方法,其特征在于,所述步骤一中的RGB图像数据集和点云图像数据集均包括真实数据集和虚拟数据集两部分,真实数据集是利用多款深度相机在实验室场地环境下采集不同品类的RGB图和深度图;虚拟数据集是利用Blender工具人工生成不同场景、不同天气下的不同品类的RGB图和深度图。

3.根据权利要求2所述的基于视锥点云的三维物体检测方法,其特征在于,所述步骤一中,真实数据集的构建还包括:对RGB图像和深度图像进行对齐和配准,以获得需要的点云数据;利用LabelImg对RGB图像进行标注以及利用LabelFusion工具对点云数据进行标注;虚拟数据集的构建中还包括深度图的计算以及数据集自动标注过程。

4.根据权利要求1所述的基于视锥点云的三维物体检测方法,其特征在于,所述步骤二中,基于二维物体检测的视锥点云抽取网络的构建方法为:

- a) 通过特征提取网络提取图片特征,形成整张图片的Feature Map;
- b) RPN网络以整张图片Feature Map为输入,输出一系列矩形的候选区域;
- c) 使用Fast-RCNN首先对感兴趣区域进行池化操作,再通过全连接层进行分类和回归,获得最终的检测结果;
- d) 利用二维物体检测的结果,通过对齐后的RGB图像和深度图像,可以得到对应物体的视锥点云。

5.根据权利要求1所述的基于视锥点云的三维物体检测方法,其特征在于,所述步骤三中,基于Point net三维物体实例分割网络的构建方法为:

a) 使用Point net网络模型作为特征提取模型,RGB图和深度图经过配准和对齐后得到的点云图像为其输入,首先经过一个T-Net操作,以实现点云特征的对齐,经过七层MLP层得到每个点的维度为2048的全局特征,其中,在第五层MLP操作后,对所得到的维度为128的局部特征也做了一个T-Net变换,接着对 $n \times 2048$ 的特征矩阵采用最大池化的操作,得到图像的全局特征;

b) 将a)中每层MLP的结果看作为视锥点云中每个点对应的局部特征,将其与a)中得到的复制 n 份后的全局特征以及利用二维物体检测得到的先验知识one-hot向量进行连接,再经过3层MLP层得到实例分割结果,获得我们感兴趣的物体的点云数据;

基于Pointnet三维物体实例分割网络的损失函数为:

$$L_{\text{seg}} = -[y \log \hat{y} + (1 - y) \log(1 - \hat{y})] ;$$

其中 \hat{y} 表示预测为感兴趣物体的概率, $(1 - \hat{y})$ 表示预测为背景的概率。

基于视锥点云的三维物体检测方法

技术领域

[0001] 本发明涉及模式识别、机器学习及计算机视觉技术领域,特别涉及基于视锥点云的三维物体检测方法。

背景技术

[0002] 三维物体检测是模式识别和计算机视觉的一个重要研究领域,同时也是解决很多高层视觉任务的核心问题,物体检测的研究为高层视觉任务(例如:行为识别、场景理解等)的解决奠定了基础。它在人们的日常生活中以及工业生产中有着广泛的应用,如:智能视频监控、无人驾驶、智能机器人、智能交通以及增强现实等。

[0003] 在过去的几年里,随着二维物体检测方法的日趋完善以及深度相机的普及,三维物体检测技术也获得了突飞猛进的发展。三维物体检测通过识别和定位三维物体,得到物体的三维信息,输出物体的三维包围盒来表示物体在真实世界中的位置。三维包围盒由其大小 l 、 w 、 h ,中心点的坐标 c_x 、 c_y 、 c_z 以及方向角 yaw 、 $roll$ 、 $pitch$ 来决定。其中 yaw 、 $roll$ 、 $pitch$ 是相对于每个类别预定义的姿态的旋转角度。在已有的研究工作中,无人驾驶三维物体检测和室内环境的场景理解三维物体检测,其三维包围盒的自由度仅仅为7个(l 、 w 、 h 、 c_x 、 c_y 、 c_z 、 yaw),因为学者们仅仅考虑了车辆和室内环境中的物体只有航向角 yaw ,而没有俯仰角 $pitch$ 和翻滚角 $roll$ 。但在我们所研究的物流拣选和装卸领域,我们必须考虑物体精确的姿态,以达到预期的效果,因此我们需要考虑物体的俯仰角 $pitch$ 和翻滚角 $roll$,输出9个自由度的三维物体包围盒。

发明内容

[0004] 为解决上述技术问题,本发明提供了基于视锥点云的三维物体检测方法,以达到检测更完善,难度更高的目的。

[0005] 为达到上述目的,本发明的技术方案如下:

[0006] 基于视锥点云的三维物体检测方法,包括如下步骤:

[0007] 步骤一:对RGB图像和点云图像进行标注,获取标注后的RGB图像数据集和点云图像数据集,将RGB图像数据集和点云图像数据集作为基于视锥点云的三维物体检测网络模型的训练样本和测试样本;

[0008] 步骤二:构建基于二维物体检测的视锥点云抽取网络,将RGB图像数据集作为基于二维物体检测的视锥点云抽取网络的输入,利用RGB信息来对物体进行检测,从而获取感兴趣物体的二维包围盒,再通过对齐和配准后RGB图像与深度图像,得到检测出来的二维物体对应的视锥点云数据;

[0009] 步骤三:构建基于Point net三维物体实例分割网络,将步骤二获取的视锥点云数据作为基于Point net三维物体实例分割网络的输入,利用Point net网络对物体进行二元的实例分割,获得实例分割后的物体点云数据;

[0010] 步骤四:构建基于偏移残差的三维包围盒回归网络,将步骤三获得的实例分割后

的物体点云数据作为基于偏移残差的三维包围盒回归网络的输入,回归物体三维包围盒的八个顶点与其真实值之间的偏移量,通过计算即可得物体的三维包围盒对应的八个顶点,从而得到其三维包围盒;

[0011] 步骤五:将RGB图像数据集和点云图像数据集输入至步骤四训练完毕的三维物体检测网络模型中,输出三维物体检测结果。

[0012] 上述方案中,所述步骤一中的RGB图像数据集和点云图像数据集均包括真实数据集和虚拟数据集两部分,真实数据集是利用多款深度相机在实验室场地环境下采集不同品类的RGB图和深度图;虚拟数据集是利用Blender工具人工生成不同场景、不同天气下的不同品类的RGB图和深度图。

[0013] 上述方案中,所述步骤一中,真实数据集的构建还包括:对RGB图像和深度图像进行对齐和配准,以获得需要的点云数据;利用LabelImg对RGB图像进行标注以及利用LabelFusion工具对点云数据进行标注;虚拟数据集的构建中还包括深度图的计算以及数据集自动标注过程。

[0014] 上述方案中,所述步骤二中,基于二维物体检测的视锥点云抽取网络的构建方法为:

[0015] a) 通过特征提取网络提取图片特征,形成整张图片的Feature Map;

[0016] b) RPN网络以整张图片Feature Map为输入,输出一系列矩形的候选区域;

[0017] c) 使用Fast-RCNN首先对感兴趣区域进行池化操作,再通过全连接层进行分类和回归,获得最终的检测结果;

[0018] d) 利用二维物体检测的结果,通过对齐后的RGB图像和深度图像,可以得到对应物体的视锥点云。

[0019] 上述方案中,所述步骤三中,基于Point net三维物体实例分割网络的构建方法为:

[0020] a) 使用Point net网络模型作为特征提取模型,RGB图和深度图经过配准和对齐后得到的点云图像为其输入,首先经过一个T-Net操作,以实现点云特征的对齐,经过七层MLP层得到每个点的维度为2048的全局特征,其中,在第五层MLP操作后,对所得到的维度为128的局部特征也做了一个T-Net变换,接着对 $n \times 2048$ 的特征矩阵采用最大池化的操作,得到图像的全局特征;

[0021] b) 将a)中每层MLP的结果看作为视锥点云中每个点对应的局部特征,将其与a)中得到的复制n份后的全局特征以及利用二维物体检测得到的先验知识one-hot向量进行连接,再经过3层MLP层得到实例分割结果,获得我们感兴趣的物体的点云数据;

[0022] 基于Pointnet三维物体实例分割网络的损失函数为:

$$[0023] \quad L_{seg} = -[y \log \hat{y} + (1 - y) \log(1 - \hat{y})] \quad ;$$

[0024] 其中 \hat{y} 表示预测为感兴趣物体的概率, $(1 - \hat{y})$ 表示预测为背景的概率。

[0025] 上述方案中,所述步骤四中,基于偏移残差的三维包围盒回归网络的构建方法为:

[0026] a) 利用步骤三中得到的实例分割后物体点云数据,经过一个基于Point net的特征抽取网络A,得到每个点的局部特征和全局特征,得到全局特征后,输出有两个不同分支操作:第一个分支是进入基于偏移残差的三维包围盒回归网络B,第二个分支是进入直接回

归三维包围盒的八个顶点的网络C,然后将这连个分支的输出结果进行对比;

[0027] b) 第一个分支,是将特征抽取网络A得到的全局特征和局部特征以及二维物体检测得到的先验知识one-hot向量进行连接,再经过5层MLP层操作,输出点云中第i个点预测包围盒8个顶点相对第i个点坐标的偏移值 p_{offset}^i 以及第i个点所预测的三维包围盒的得分 p_i ,选取所有点中预测的得分最高的三维包围盒作为网络的输出;然后通过真实值与偏移量计算得出预测的三维包围盒的八个顶点坐标 C_i ,其计算方法如下:

$$[0028] \quad C_i = p_i + p_{offset}^i;$$

[0029] c) 第二个分支是直接特征抽取网络A得到的全局特征与二维检测得到的先验知识one-hot向量进行连接,经过3个MLP层操作,得到预测的三维包围盒的八个顶点坐标作为网络的输出;

[0030] 基于偏移残差的三维包围盒回归网络B的损失函数为:

$$[0031] \quad L_{reg} = \frac{1}{m} \sum_i offset_ratio_i + L_{conf};$$

[0032] 其中,m是实例分割点云的数目;offset_ratio_i是第i个点预测的包围盒的偏移距离比率,偏移距离比率即预测顶点与真实顶点的偏移距离和真实包围盒对角线长度的比例,其计算方法为:

$$[0033] \quad offset_ratio = \frac{corner_distance}{diagonal_length};$$

[0034] L_{conf} 是预测包围盒置信度的损失,其计算方法为:

$$[0035] \quad L_{conf} = \frac{1}{m} \sum_i (c_{pred} - c_{gt})^2;$$

[0036] 其中,m是实例分割点云的数目, c_{pred} 代表预测的物体三维包围盒的置信度, c_{gt} 代表物体真实三维包围盒的置信度。

[0037] 对于直接回归三维包围盒的八个顶点的网络C的损失函数为:

$$[0038] \quad L_{reg} = offset_ratio。$$

[0039] 通过上述技术方案,本发明提供的基于视锥点云的三维物体检测方法采集RGB图像和点云图像并进行标注,获取标注后的RGB图像数据集和点云图像数据集,将数据集作为基于视锥点云的三维物体检测神经网络模型的训练样本和测试样本。基于视锥点云的三维物体检测网络模型主要由三个网络组成:基于二维物体检测的视锥点云抽取网络、基于Point net三维物体实例分割网络以及基于偏移残差的三维包围盒回归网络。三个网络依次训练,前一个网络的训练输出作为下一个网络的输入。设置各神经网络模型的超参数,通过Tensorflow训练三维物体检测网络模型,当代价损失函数减少到理想程度且训练达到所要求的最大迭代次数时,生成训练模型;将点云图像数据集输入至训练完毕的模型中,输出三维物体检测结果。本发明比现有三维物体检测方法更完善、难度更高。

附图说明

[0040] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍。

- [0041] 图1为基于视锥点云的三维物体检测网络结构图；
- [0042] 图2为基于二维物体检测的视锥点云抽取网络结构图；
- [0043] 图3为基于Point net三维物体实例分割网络结构图；
- [0044] 图4为基于偏移残差的三维包围盒回归网络结构图；
- [0045] 图5为本实施例中基于Point net三维物体实例分割结果展示；
- [0046] 图6为本实施例中基于视锥点云的三维物体检测方法结果展示。

具体实施方式

[0047] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述。

[0048] 本发明中,RGB图指的是通过红(R)、绿(G)、蓝(B)参数值表示的图像,其通过对红(R)、绿(G)、蓝(B)三个颜色通道的变化以及它们相互之间的叠加得到各式各样的颜色;深度图是包含与视点的场景对象的表面的距离有关的信息的图像或图像通道,其类似于灰度图像,只是深度图的每个像素值是传感器距离物体的实际距离。通常RGB图像和Depth图像是配准的,因而像素点之间具有一对一的对应关系,点云图像就是指RGB图像和Depth图像通过对齐和配准之后得到的图像。

[0049] 如图1所示,基于视锥点云的三维物体检测网络模型,包括基于二维物体检测的视锥点云抽取网络、基于Point net三维物体实例分割网络、基于偏移残差三维包围盒回归网络三个子网络。具体检测方法包括:

[0050] 步骤一:对RGB图像和点云图像进行标注,获取标注后的RGB图像数据集和点云数据集,将数据集作为基于视锥点云的三维物体检测网络模型的训练样本和测试样本;

[0051] 步骤二:构建基于二维物体检测的视锥点云抽取网络,如图2所示,将RGB图像数据集作为二维物体检测的视锥点云抽取网络的输入,利用丰富的RGB信息来对物体进行检测,从而获取感兴趣物体的二维包围盒,再通过对齐和配准后的RGB图像与深度图像,得到检测出来的二维物体对应的视锥点云;

[0052] 步骤三:构建基于Point net三维物体实例分割网络,如图3所示,将步骤二获取的视锥点云数据作为基于Point net三维物体实例分割网络的输入,利用Point net网络对物体进行二元的实例分割,获得实例分割后的物体点云数据;

[0053] 步骤四:构建基于偏移残差的三维包围盒回归网络,如图4所示,将步骤三获得的实例分割后的物体点云数据作为基于偏移残差的三维包围盒回归网络的输入,回归点云中每个点预测的包围盒8个顶点相对该点坐标的偏移值以及该点所预测的三维包围盒的得分,选取得分最高的三维包围盒,再通过计算即可得物体的9自由度三维包围盒对应的八个顶点,从而得到其三维包围盒;

[0054] 步骤五:将RGB图像数据集和点云图像数据集输入至训练完毕的三维物体检测网络模型中,输出三维物体检测结果。

[0055] 基于视锥点云的三维物体检测方法的一个实施例中,步骤一中的数据集中包括真实数据集和虚拟数据集两部分。其中,真实数据集是利用奥比中光Astra Pro、英特尔SR300等多款深度相机在实验室场地环境下采集不同品类的RGB图和深度图,数据集中包括箱子、文件袋、塑包包裹、编织袋四种品类,包括地堆和货架两种分拣场景。虚拟数据集是利用

Blender工具人工生成不同分拣场景、不同天气下不同品类的RGB图和深度图,数据集中包括箱子、文件袋、塑包包裹、编织袋四种品类,分拣筐、地堆、工厂、货架四种分拣场景,包括晴天、雾天两种天气。其中,第一种获取方法采用的人工标注的方式得到标注信息,利用LabelImg对RGB图像进行标注以及利用LabelFusion工具对点云数据进行标注;第二种方法的标注信息利用编程实现。

[0056] 本实施例步骤一中的训练样本包含14151张已标注的真实数据集中RGB图像,36497张已标注的虚拟数据集中的RGB图像,测试样本包含1633张真实数据集中的RGB图像;本实施例步骤二中的训练样本包含1116张已标注的真实数据集中点云图像,9000张已标注的虚拟数据集中的点云图像,对于真实数据集和虚拟数据集按照8:1:1的比例随机的划分训练集、验证集和测试集。

[0057] 如上所述的基于视锥点云的三维物体检测方法的一个实施例中,步骤一中真实数据集中还包括对RGB图像和深度图像进行对齐和配准,以获得需要的点云数据。虚拟数据集中还包括深度图的计算以及数据集自动标注过程。

[0058] 如上所述的基于视锥点云的三维物体检测方法的一个实施例中,步骤二中基于二维物体检测的视锥点云抽取网络(如图2所示)构建方法为:

[0059] a) 通过特征提取网络如VGG16、ResNet-X等提取图片特征,形成整张图片的Feature Map;

[0060] b) RPN网络以整张图片Feature Map为输入,输出一系列矩形的候选区域;

[0061] c) 使用Fast-RCNN首先对感兴趣区域进行池化操作,再通过全连接层进行分类和回归,获得最终的检测结果;

[0062] d) 利用二维物体检测的结果,通过对齐后的RGB图像和深度图像,可以得到对应物体的视锥点云。

[0063] 如上所述的基于视锥点云的三维物体检测方法的一个实施例中,步骤三中基于Point net三维物体实例分割网络(如图3所示)构建方法为:

[0064] a) 使用Point net网络模型作为特征提取模型,RGB图和深度图经过配准和对齐后得到的点云图像为其输入,首先经过一个T-Net操作,以实现点云特征的对齐,经过七层MLP(多层感知机)层得到每个点的维度为2048的全局特征,其中,在第五层MLP操作后,对所得到的维度为128的局部特征也做了一个T-Net变换。经过七层MLP(多层感知机)层得到每个点的维度为2048的全局特征,接着对 $n \times 2048$ 的特征矩阵采用最大池化的操作,得到图像的全局特征;

[0065] b) 将a)中每层MLP的结果看作为视锥点云中每个点对应的局部特征,将其与a)中得到的复制n份后的全局特征以及利用二维检测结果的先验知识得到的one-hot向量进行连接,再经过3层MLP层得到实例分割结果,获得我们感兴趣的物体的点云数据。

[0066] 基于Point net三维物体实例分割网络的损失函数为:

$$[0067] \quad L_{seg} = -[y \log \hat{y} + (1 - y) \log(1 - \hat{y})] \quad ;$$

[0068] 其中 \hat{y} 表示预测为感兴趣物体的概率, $(1 - \hat{y})$ 表示预测为背景的概率。

[0069] 如上所述的基于视锥点云的三维物体检测方法,所述步骤四中,基于偏移残差的三维包围盒回归网络(如图4所示)构建方法为:

[0070] a) 利用步骤三中得到的实例分割后物体点云数据,经过一个基于Point net的特征抽取网络A,得到每个点的局部特征和全局特征,得到全局特征后,输出有两个不同分支操作:第一个分支是进入基于偏移残差的三维包围盒回归网络B,第二个分支是进入直接回归三维包围盒的八个顶点的网络C,我们将这连个分支的输出结果进行对比;

[0071] b) 第一个分支,是将特征抽取网络A得到的全局特征和局部特征以及二维物体检测得到的先验知识one-hot向量进行连接,再经过5层MLP层操作,输出点云中第i个点预测包围盒8个顶点相对第i个点坐标的偏移值 p_{offset}^i 以及第i个点所预测的三维包围盒的得分 p_i ,选取所有点中预测的得分最高的三维包围盒作为网络的输出;然后通过真实值与偏移量计算得出预测的三维包围盒的八个顶点坐标 C_i ,其计算方法如下:

$$[0072] \quad C_i = p_i + p_{offset}^i;$$

[0073] c) 第二个分支是直接特征抽取网络A得到的全局特征与二维检测得到的先验知识one-hot向量进行连接,经过3个MLP层操作,得到预测的三维包围盒的八个顶点坐标作为网络的输出;

[0074] 基于偏移残差的三维包围盒回归网络B的损失函数为:

$$[0075] \quad L_{reg} = \frac{1}{m} \sum_i offset_ratio_i + L_{conf};$$

[0076] 其中,m是实例分割点云的数目;offset_ratio_i是第i个点预测的包围盒的偏移距离比率,偏移距离比率即预测顶点与真实顶点的偏移距离和真实包围盒对角线长度的比例,其计算方法为:

$$[0077] \quad offset_ratio = \frac{corner_distance}{diagonal_length};$$

[0078] L_{conf} 是预测包围盒置信度的损失,其计算方法为:

$$[0079] \quad L_{conf} = \frac{1}{m} \sum_i (c_{pred} - c_{gt})^2;$$

[0080] 其中,m是实例分割点云的数目, c_{pred} 代表预测的物体三维包围盒的置信度, c_{gt} 代表物体真实三维包围盒的置信度。

[0081] 对于直接回归三维包围盒的八个顶点的网络C的损失函数为:

$$[0082] \quad L_{reg} = offset_ratio.$$

[0083] 基于深度学习的RGB-D三维物体检测方法的一个实施例中,步骤二中通过Caffe2训练二维物体检测卷积神经网络模型方法为:本文使用ImageNet数据集预训练的分类模型初始化网络模型,利用真实数据集训练网络模型。其初始化学习率设置为0.02,最大迭代次数为80k,并在30k和40k分别以0.1倍递减,batch size大小为2张图片,且权值衰减率为0.0005;

[0084] 基于深度学习的RGB-D三维物体检测方法的一个实施例中,步骤三中通过TensorFlow训练三维物体检测神经网络模型方法为:使用Point net网络预训练神经网络模型为三维物体检测神经网络进行参数初始化。

[0085] 基于深度学习的RGB-D三维物体检测方法的一个实施例中,步骤四中通过TensorFlow训练基于偏移残差的三维包围盒回归网络模型方法为:使用Point net网络预训练神经网络模型为三维物体检测神经网络进行参数初始化。

[0086] 在本实施例中步骤三、四中采用带动量参数为0.9的随机梯度下降训练模型,每一批的数据量为64条。实验采取变化的学习率进行学习,初始学习率为0.0002,针对人工生成的虚拟训练集由于其数据量比较大在7万条左右,每训练70万个数据(即大约经过10个周期(epoch))学习率下降到原来的0.7倍,针对真实采集的数据集由于其数据量大约1万条左右,每训练20万个数据(也就是大约没经过20个周期(epoch))学习率下降到原来的0.7倍,以上两种情况均是学习率小于0.00001时停止下降。实验采用了早停止的策略,训练了200个周期(epoch)。为了减少模型初始化对结果的影响,实验对每一个实验在不同初始化下训练5次,将测试的精度平均值作为模型的最终结果。

[0087] 如上所述的基于视锥点云的三维物体检测方法中,在训练阶段我们采用了三种方式来组成我们的训练数据。第一种是所有训练数据集均为真实数据集,第二种是将真实数据集和虚拟数据集混合后训练,第三种是在先在虚拟数据集上训练网络模型后,再用该网络模型训练真实数据集进行网络调优。

[0088] 如图5所示为本实施例中真实数据集测试集中三维物体实例分割结果图,其中的包围框为标注的3D包围盒,灰色的点代表感兴趣的物体的点云数据,而黑色的点代表背景物体,可以看出,基于point net的点云实例分割网络可以很好的对视锥点云数据进行分割。

[0089] 如表1所示,图中“训练集”一列中Real表示使用的训练数据全为真实数据,Syn+Real表示使用的训练数据集为真实数据集和虚拟数据集混合后得到的数据集,Syn+Real Adapted表示是网络模型先在虚拟数据集中训练得到神经网络参数后,再利用真实数据集对神经网络进行调优得到的模型,Syn表示使用的训练数据全为虚拟数据集;“测试集”一列中Real数据集代表利用真实数据集作为测试集,Syn数据集代表利用虚拟数据集作为测试集。从图中可以看出,基于Point net的点云实例分割网络在真实数据集中的平均测试精度能够达到87%左右。

[0090] 表1基于Point net三维物体实例分割网络平均精度

	训练集	测试集	平均精度 AP (%)		mAP (%)
			背景	感兴趣物体	
[0091]	Real		90.74	81.30	86.02
	Syn+Real	Real 数据集 ^a	87.23	75.42	81.32
	Syn+Real Adapted		91.19	82.89	87.04
	Syn	Syn 数据集 ^b	94.19	95.56	94.88

[0092] 如图6所示为本实施例中基于视锥点云的三维物体检测方法结果图,其中第一行表示的是基于二维物体检测的视锥点云抽取网络中的2D标注包围框;第二行是基于视锥点云的三维物体检测结果图,其中的包围框是三维物体检测出的物体的三维包围盒;第三行是基于视锥点云的三维物体检测网络中的3D标注包围盒。

[0093] 如表2所示,为本实施例中测试样本的三维物体检测类别和准确率,表中“训练数据”一列中Real表示使用的训练数据全为真实数据集,Syn+Real表示使用的训练数据集为

真实数据集和虚拟数据集混合后得到的数据集, Syn+Real Adapted表示是网络模型先在虚拟数据集中训练得到神经网络参数后, 再利用真实数据集对神经网络进行调优得到的模型。

[0094] 表2本实施例中测试样本的三维物体检测类别和准确率

模型	训练数据	平均精度 AP_{3D} (%)				mAP (%)
		瓦楞纸盒	信封袋	塑料包裹	编织袋	
[0095] 直接回归	Real	56.46	67.06	63.78	62.41	62.43
	Syn+Real	48.65	59.63	55.34	53.25	54.22
	Syn+Real Adapted	60.98	69.15	65.20	62.95	64.57
偏移残差回归	Real	61.76	68.76	64.17	63.57	64.57
	Syn+Real	52.15	59.43	56.65	54.32	55.64
	Syn+Real Adapted	64.62	71.45	64.21	65.09	66.59

[0096] 从表中可以看出, 在回归的平均精度方面, 除了塑料包裹这个类别直接回归的比基于偏移残差回归的精度高了大约1%, 其余的类别的回归平均精度均是基于偏移残差模型高于直接回归的模型, 总体来讲, 基于偏移残差的回归比直接回归的平均精度均值高了大约2.02%, 说明基于偏移残差回归的优于直接回归。由于直接回归包围盒的模型, 其网络输出是直接回归包围盒的8个顶点坐标, 我们所回归的物体尺寸差异性比较大, 因此直接回归具有很大的难度, 回归的效果也比较差。

[0097] 对所公开的实施例的上述说明, 使本领域专业技术人员能够实现或使用本发明。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的, 本文中所定义的一般原理可以在不脱离本发明的精神或范围的情况下, 在其它实施例中实现。因此, 本发明将不会被限制于本文所示的这些实施例, 而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

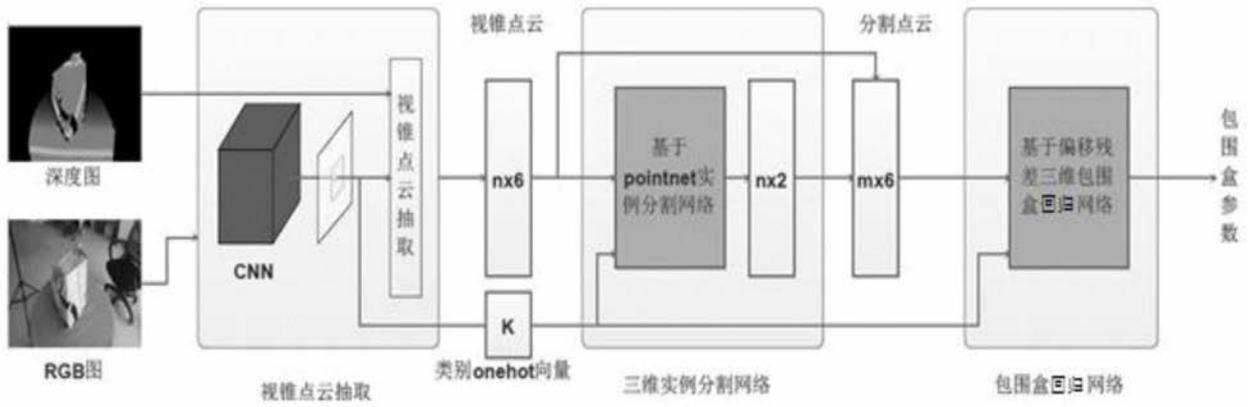


图1

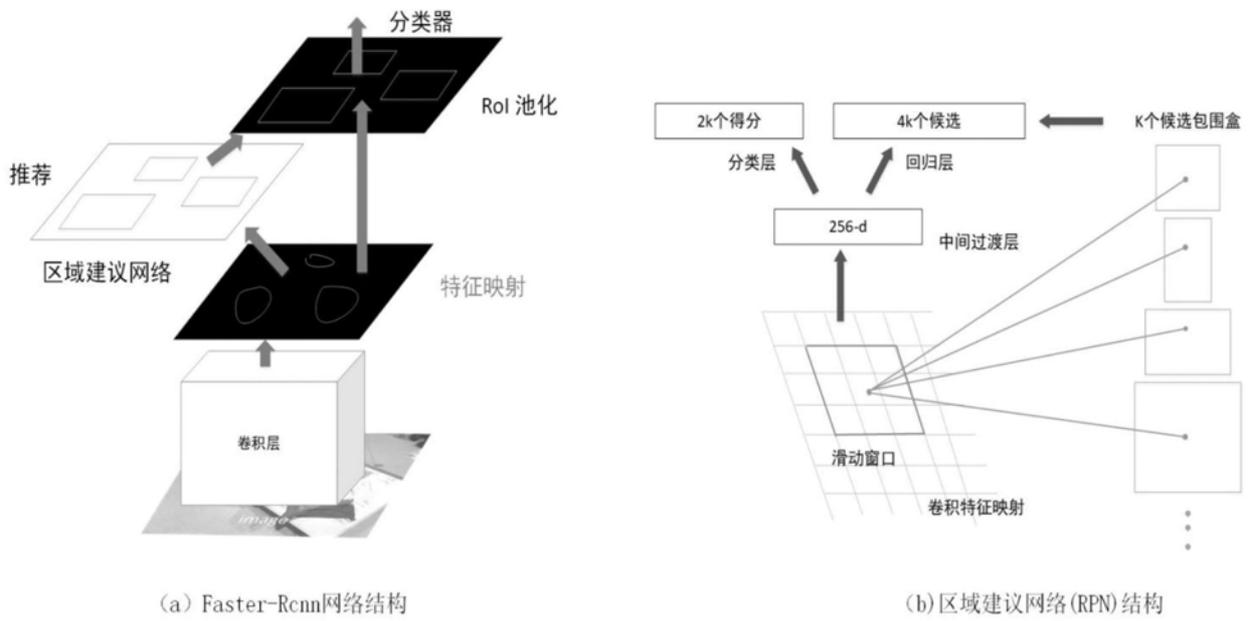


图2

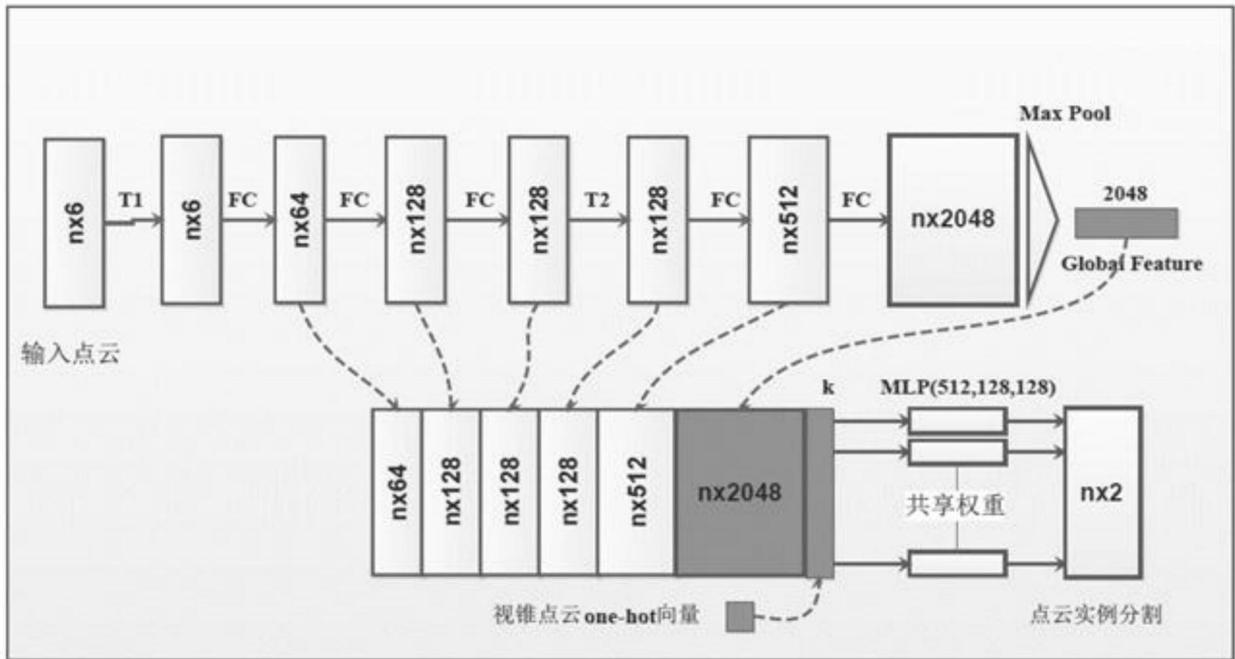


图3

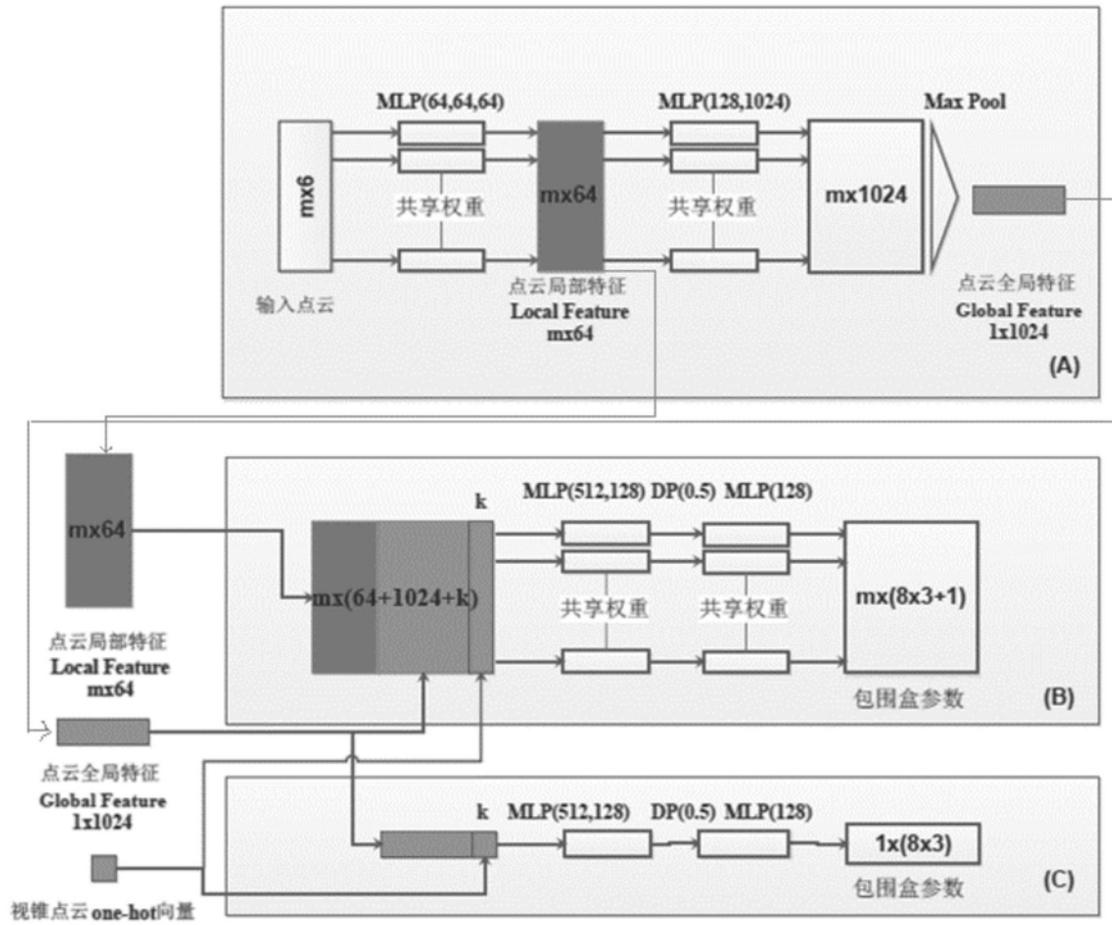


图4

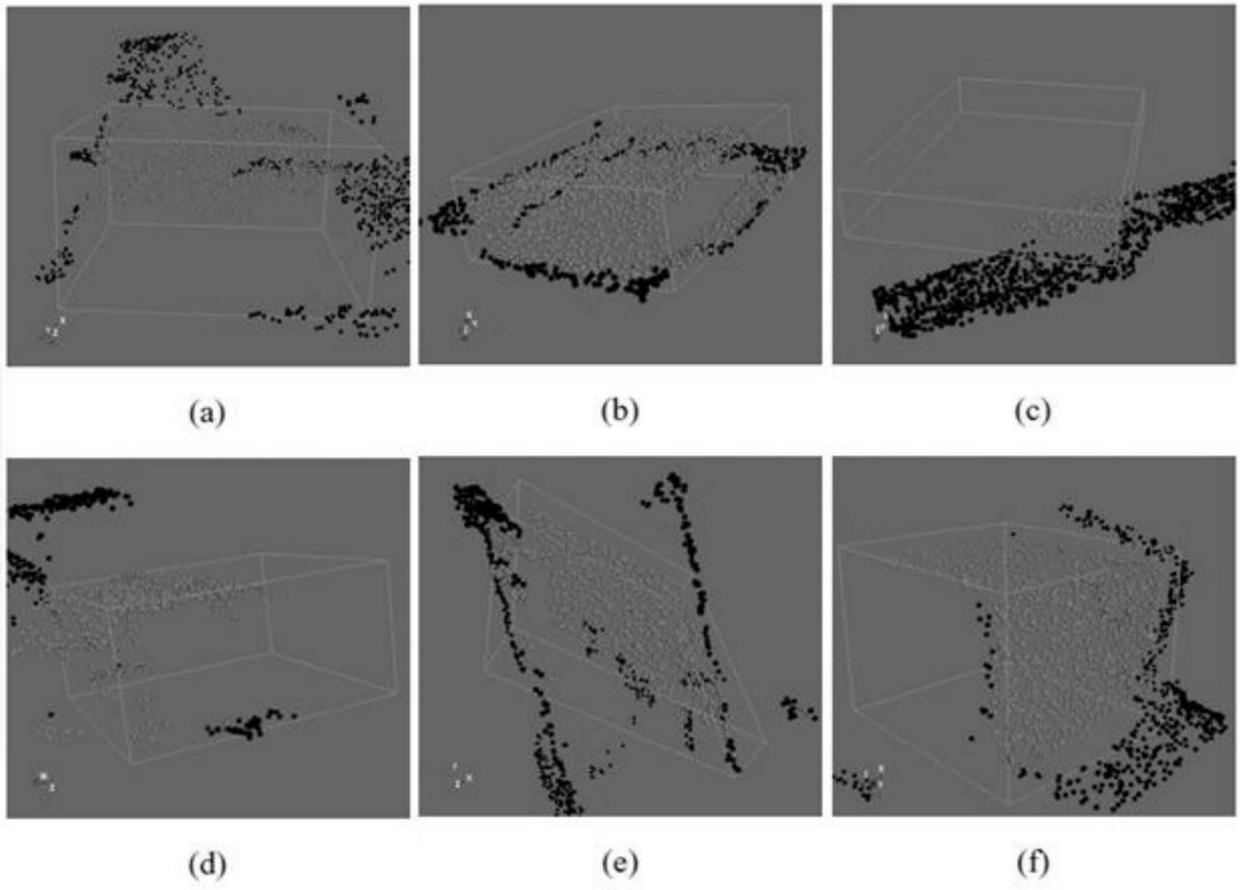


图5

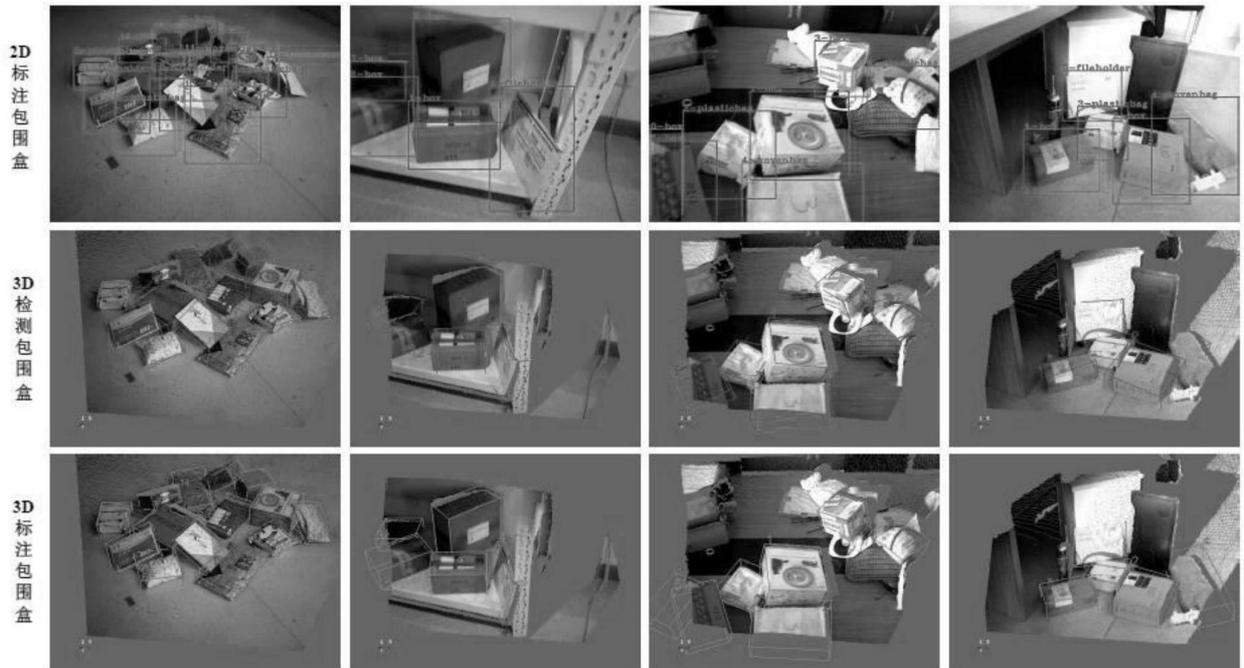


图6