第47卷 第8期 计 算 机 工 程 2021年8月
Vol.47 No.8 Computer Engineering August 2021

·图形图像处理· 文章编号：1000-3428(2021)08-0216-08 文献标志码：A 中图分类号：TP391

基于深度学习的物体点云六维位姿估计方法

李少飞，史泽林，庄春刚

（上海交通大学 机械与动力工程学院，上海 200240）

摘 要：物体位姿估计是机器人在散乱环境中实现三维物体拾取的关键技术，然而目前多数用于物体位姿估计的深度学习方法严重依赖场景的RGB信息，从而限制了其应用范围。提出基于深度学习的六维位姿估计方法，在物理仿真环境下生成针对工业零件的数据集，将三维点云映射到二维平面生成深度特征图和法线特征图，并使用特征融合网络对散乱场景中的工业零件进行六维位姿估计。在仿真数据集和真实数据集上的实验结果表明，该方法相比传统点云位姿估计方法准确率更高、计算时间更短，且对于疏密程度不一致的点云以及噪声均具有更强的鲁棒性。

关键词：点云；位姿估计；特征融合；深度学习；损失函数

开放科学（资源服务）标志码（OSID）：

中文引用格式：李少飞，史泽林，庄春刚.基于深度学习的物体点云六维位姿估计方法[J].计算机工程，2021，47(8)：216-223.

英文引用格式：LI S F，SHI Z L，ZHUANG C G.Deep learning-based 6D object pose estimation method from point clouds[J].Computer Engineering，2021，47(8)：216-223.

Deep Learning-Based 6D Object Pose Estimation Method from Point Clouds

LI Shaofei，SHI Zelin，ZHUANG Chungang

（School of Mechanical Engineering，Shanghai Jiao Tong University，Shanghai 200240，China）

【Abstract】Object pose estimation is a key technology required for enabling the robots to pick 3D objects in a cluttered environment. However，most of the existing deep learning methods for pose estimation rely heavily on the RGB information of the scene，which limits their applications. To address the problem，a deep learning-based method for 6D object pose estimation is proposed. A data set for industrial parts is generated from physical simulation，and then the 3D point cloud is mapped to the 2D plane to generate a deep feature map and normal feature map. On this basis，a feature fusion network is used for 6D pose estimation of industrial parts in cluttered environments. Experimental results on the simulation data set and the real data set show that the proposed method improves the accuracy of pose estimation and reduces time consumption compared with traditional point cloud pose estimation methods. In addition，the method displays high robustness to the point clouds with different density and noises.

【Key words】point cloud；pose estimation；feature fusion；deep learning；loss function

DOI：10. 19678/j. issn. 1000-3428. 0058768

0 概述

散乱场景中的三维物体拾取是机器人操作中的一类经典问题，利用机械臂将箱子中无序摆放、堆叠的物体取出对机器人实现自动化具有重要意义。该问题的难点在于散乱堆叠的物体之间存在大量的遮挡，这不仅影响了对物体的识别，而且使得拾取过程中的碰撞检测更加复杂。物体六维位姿识别是散乱场景中三维物体拾取的重点和难点。近年来，深度学习技术在六维位姿估计任务中得到广泛应用。文献[1-3]根据RGB数据对纹理丰富的物体实例进行六维位姿估计。文献[4]扩展二维目标检测器，提出一种基于分类离散视点的旋转位姿估计方法，但该方法仅预测真实姿态的粗略离散近似值，为达到更好的效果，还需对输出结果进行位姿细化。文献[5]先将RGB图像在2个网络中进行由粗到细的分割，再将分割结果反馈给第3个网络得到待检测目标边界框点的投影，最终利用PnP算法估计六维位姿，但

基金项目：国家自然科学基金（51775344）。

作者简介：李少飞（1995—），男，硕士研究生，主研方向为机器视觉；史泽林，硕士研究生；庄春刚，副研究员、博士生导师。

收稿日期：2020-06-28 **修回日期：**2020-08-17 **E-mail：**ee807654186@sjtu.edu.cn

该方法由于将网络分为多个阶段,因此导致运行时间非常长。文献[6]针对通过CNN检测二维关键点并利用PnP回归六维位姿的方法在遮挡和截断样本中存在的问题进行改进,对于每一个像素计算一个指向二维关键点的方向向量,并通过投票策略得到鲁棒的二维关键点,减少了物体局部缺失对位姿估计的影响。文献[7]通过训练获取输入RGB图像的六维隐变量表示,然后在数据库中查找和其最相近的位姿作为估计结果。然而,在低纹理的情况下,仅通过RGB信息估计的六维位姿准确率较低。文献[8-10]将RGB信息和深度信息相结合估计目标的六维位姿。文献[11-12]均是利用CNN学习特定的描述子进行目标检测和六维位姿估计。从RGB-D图像进行六维目标位姿估计的关键是充分利用两个互补的数据源,文献[13]提出一种新的稠密融合网络,该网络将分别处理后的两种数据源进行像素级别的特征嵌入,从而估计出物体的六维位姿。

近几年,基于深度学习的六维位姿估计方法多数将RGB图和深度图作为输入。然而,一个物体处于不同的位姿却有着相似的二维图像这一现象是很常见的,这限制了基于二维图像的位姿估计的准确率。在一些工业应用中,为了获取完整场景、高精度的三维信息,通常会采用三维扫描仪获取场景点云,而有些扫描仪由于成像原理不同,不能获取RGB图和深度图。随着传感器技术的发展,获取三维点云的速度得到了大幅提升,这使得基于点云研究的实时性得到了保障。因此,基于点云的物体六维位姿估计引起了研究人员的关注。DROST等[14-15]提出基于物体点对特征(Point Pair Feature,PPF)的位姿估计算法及其变体算法,并将其成功应用于工业机器人分拣任务,然而此类算法的局限性在于:一方面,如果模板点云和场景点云的采样疏密程度不一致,将难以发现相似点对特征,从而导致匹配错误;另一方面,出现了一些先分割后配准的算法,将点云进行聚类分割后,利用点云配准的流程得到物体的位姿[16],但是此类算法计算量大,且在堆叠严重的场景中表现较差。在深度学习领域,QI等[17]基于对称函数思想,将原始点云输入网络进行训练实现分类和分割任务,并在网络中加入分层多尺度特征学习[18],该方法相比已有方法在精度上有了显著提升。之后研究人员将该方法应用于自动驾驶的目标检测提出F-PointNet[19],F-PointNet虽然在一定程度上解决了三维目标检测问题,但是激光雷达获得的点云是稀疏和不规律的,在自动驾驶场景中的物体也鲜有遮挡的情况,并且包围框的位姿也仅考虑垂直于地面的旋转,这与散乱场景中堆叠的工件有很大的差别,因此此类方法的实用性不强。

针对现有点云位姿估计方法计算量较大且在复杂场景中结果鲁棒性较差的问题,本文提出基于深度学习的物体点云六维位姿估计方法,将三维点云映射到二维平面,生成深度特征图和法线特征图,提取位姿特征。

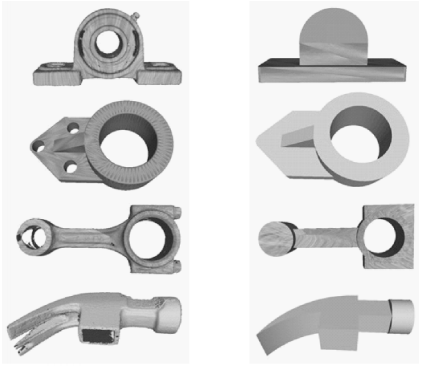
1 数据集生成

1.1 工业零件建模

现有基于深度学习的六维位姿估计方法多数是在已有的LINEMOD、OCCLUSION等数据集上进行测试。但是,由于工业零件的特殊性,在这些数据集上测试效果很好的神经网络并不能适用于一些低纹理的机械零件,因此本文提出了一种用于工业零件位姿估计的数据集生成方法。

在对数据集进行标签标注时,点云的标签标注相比二维图像标注更加困难。每训练一个新的工件,如果用真实点云生成数据集,则工作量会非常巨大,因此在仿真环境下生成数据集用于训练是很有必要的。文献[20]考虑了环境光反射的影响,利用Unity3D游戏引擎生成散乱堆叠场景的深度图数据集。文献[21]利用Blender API将提前建好的日常用品的三维模型放入仿真环境,设置模型初始位姿,并通过重力掉落以及刚体碰撞模拟真实环境。上述仿真方法均能达到较好的效果,但是所仿真模型的几何结构都是类似于圆柱体、立方体等简单的模型,而对于一些复杂的工件,首先建模精确度较低,其次仿真会出现穿模现象。

本文对文献[21]所采用的物理仿真方法进行改进,在Blender API中根据模型纹理、矩形包络、球包络等方式选择物理的碰撞类型。基于模型纹理的物理仿真方法会在模型面数较多时出现计算复杂度高的问题,从而引起穿模,而基于矩形包络、球包络等的物理仿真方法虽然可以避免模型之间产生穿模现象,但是模型形状的简化会使工件之间的堆叠不能反映真实场景中的碰撞堆叠效果。因此,本文首先利用高精度的三维扫描仪,拍摄工件多个角度的三维点云并进行配准,得到工件的完整点云;接着采用贪婪投影三角法进行曲面重建,得到复杂工业零件的精确模型,如图1(a)所示。为了尽可能减少模型面数从而减少仿真计算量,并保证物理碰撞尽可能与真实场景相似,本文对每一个特定的工业零件,实心化对物理碰撞不会产生影响的局部区域,而对于产生碰撞的区域,使用相对简单的形状进行包络拟合,如图1(b)所示。在图1中,本文采用的4种工件从上到下依次为轴承座1、轴承座2、连杆和榔头。



(a)精确模型 (b)简化模型

图1 精确模型与简化模型

Fig.1 Exact model and simplified model

1.2 基于物理仿真的数据集生成

本文数据集生成的步骤如下:1)将多个简化的工件模型预设置随机位姿并置于环境上方;2)工件依靠重力下落,基于模型纹理产生碰撞散乱堆叠在相机视野下,然后渲染得到每个工件的掩码与之后生成的深度图对应得到点云的类别标签;3)在获取堆叠工件位姿后,在Bullet中用重建的精确点云模型代替简化模型,渲染得到深度图,进而获得散乱场景的点云,如图2所示。这样就可以使得仿真生成的散乱堆叠工件的点云以及工件之间的碰撞效果和真实场景尽可能相似,防止由于模型面数过多造成穿模问题。由于Blender中的工件在世界坐标系下的坐标变换为$\boldsymbol{T}_{\text{W-parts}}$,因此需要将其转换到相机坐标系下,已知相机在世界坐标系下的坐标变换为$\boldsymbol{T}_{\text{W-Cam}}$,则工件在相机坐标系下的六维位姿为:

$$\boldsymbol{T}_{\text{C-parts}}=\boldsymbol{T}_{\text{W-parts}}\boldsymbol{T}_{\text{W-cam}} \tag{1}$$

(a)精确模型的点云仿真 (b)简化模型的点云仿真

图2 散乱场景的点云仿真

Fig.2 Simulation of point clouds in scattered scene

2 基于深度学习的点云位姿估计方法

直接将学习得到的原始点云特征输入全连接层进行训练可以达到很好的分类效果[17-18],但对于六维位姿估计效果并不理想,因为训练得到的全局特征和每个点的局部特征更多的是表现该工件的类别特征,而用于估计六维位姿的局部表面特征和几何特征并未进行有效提取,仅依靠神经网络本身参数的调整和训练效果较差。另外,神经网络的数据输入维度需要保持一致,而从场景分割得到的单个点云的点数是不确定的,为了使其能够输入网络,需要采样成固定点数,这会使得工件点云变得稀疏,从而损失一定的特征。近年来研究人员提出了许多成熟的处理二维图像的深度学习方法,因此本文将三维点云映射到二维平面,生成深度特征图和法线特征图并提取位姿特征,不仅保证了网络输入维度一致,而且大幅提高了基于点云的位姿估计准确率。

2.1 点云二维深度特征生成

在位姿估计前,本文利用ASIS方法[22]对散乱场景的点云进行分割预处理。对于每一个分割后的单个工件点云,计算其xyz坐标的平均值x_{m}、y_{m}、z_{m},记为点云的中心,并将点云中心移动到相机坐标系原点,如图3中A所示,记为$\boldsymbol{t}_{\text{O}}=(-x_{\text{m}},-y_{\text{m}},-z_{\text{m}})^{\text{T}}$并得到:

$$\boldsymbol{P}_{\text{O}}=\boldsymbol{T}_{\text{S-O}}\boldsymbol{P}_{\text{S}} \tag{2}$$

其中:$\boldsymbol{T}_{\text{S-O}}=\begin{pmatrix}\boldsymbol{I} & \boldsymbol{t}_{\text{O}}\\ \boldsymbol{0} & 1\end{pmatrix}$;$\boldsymbol{P}_{\text{S}}$为移动前的齐次点;$\boldsymbol{P}_{\text{O}}$为移动后的点。当点云平移到相机坐标系原点附近时,将点云向Z轴正方向投影,在XY平面生成二维图像,如图3中B和图3中C所示。同类工件所有样本的所有点中的最大最小x、y坐标之差H、W定义为图像的宽和高,z的最大值记为$z_{\max}$,其最大最小坐标的差值V对应于灰度值的最大值255。

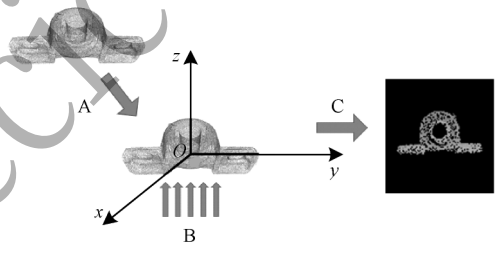


图3 点云二维特征生成

Fig.3 2D feature generation of point clouds

将点云平移到坐标原点附近可以有效减小图像尺寸,使样本点所占二维图像的比例尽可能大,增加图像特征的显著度。点云到二维图像的具体映射方法为:1)设定分辨率及宽度方向的像素个数,按照图像宽高尺寸的比例设定高度方向的像素个数;2)将点投影到图像中时,会出现一个像素中存在多个点的情况,此时仅保留z值最小的点,该点离观测视野最近,识别度最高;3)由于二维图像是单通道的灰度图,因此得到点像素的灰度值为:

$$G=(z_{\max}-z)\times 255/V \tag{3}$$

由于设定的分辨率不同，因此每个像素包含点的数量也会发生变化，而二维图像的特征也会有所差别。图4给出了在不同分辨率下工件仅通过二维深度特征进行位姿估计的准确率。可以看出，分辨率从起始到80像素×80像素时，位姿估计的准确率提升得很快，再提高分辨率时，位姿估计准确率的提升开始减缓，并且约在100像素×100像素时达到最大，此时进一步提高分辨率，准确率开始缓慢下降。由于分辨率过大或者过小都会造成点云二维特征不够明显，因此在实验阶段，本文将特征图的分辨率设置为峰顶处的100像素×100像素。同时，本文工件的尺寸设置为10~20 cm，如果物体尺寸大于实验采用的工件尺寸，可以适当提高分辨率，反之亦然。笔者认为应谨慎降低特征图的分辨率，因为从实验结果可以看出，过大的分辨率对实验结果的影响远小于过小的分辨率。

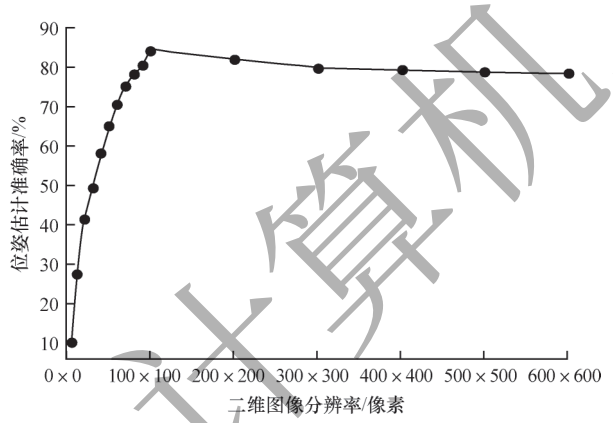


图4　不同分辨率下位姿估计的准确率

Fig.4　Accuracy of pose estimation at different resolutions

2.2　点云二维法线特征生成

点云生成的二维深度特征能够有效提取出工件的几何特征，但是一些不同的工件或者一个工件的不同局部投影到二维平面，有可能呈现类似的形状，即使深度不一致，也会影响最终的估计结果。如图5(a)、图5(b)所示，轴承座2的正反面投影到二维平面会产生上述问题。而点云法线作为点云的一种重要的几何属性，已广泛应用于特征点检测、三维重建、薄板正反面区分等场景。传统位姿估计算法的点对特征[14]就是运用两点的法线特征构建特征算子，而近年来许多基于点云分类分割的深度学习研究[17-18]也将点云的表面法线作为点云的额外信息输入网络进行训练，经过实验证明，分割准确度有了明显提升。因此，本文类比二维深度特征图的生成方式生成点云的二维法线特征图，用于增加二维特征的区分度，即使不同位姿样本的二维深度图相似，最终的位姿估计结果也不会产生误匹配的情况。

在将点云投影到二维平面生成的深度特征图前，利用Open3D库计算点云的法线，这样二维深度特征图中任意点像素都会包含这个点的深度值及其法线。将各点的三维法线特征和深度值分离，即可得到二维法线特征图。本文思想是将法线特征和深度特征分成两条支路，各自学习对应的特征，最终将网络学到的特征信息进行融合输出六维位姿。在二维法线特征图生成的过程中存在两方面的问题。一方面，通过上述方法计算出的法线并没有经过全局定向，这会极大地影响模型对工件位姿的训练。本文将所有法线的方向统一至与z轴负方向呈小于90°的夹角，解决了全局定向的问题，将二维法线特征图中计算得到的法线以及该像素缓存的三维点还原成空间点云，可以看到法线的取向是统一的，如图5(c)、图5(d)所示。另一方面，在二维法线特征计算的过程中引入了分割后的噪声，特别是在工件的边缘位置处，法线的估计会因为噪声产生很大的误差，因此本文在实验部分将噪声对位姿估计结果的影响进行实验验证。

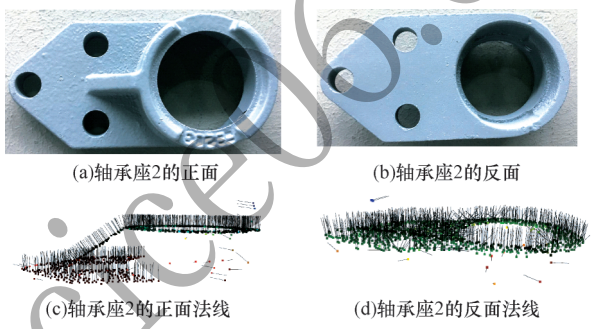


图5　轴承座2的正反面及其点云法线

Fig.5　Front and back sides and their point cloud normals of bearing pedestal 2

2.3　特征融合网络

本文提出的特征融合网络框架如图6所示。特征融合网络主要包括：1)二维深度特征提取，将点云映射为二维深度特征图，经过预处理后输入resnet50预训练模型进行预训练，每个样本得到2 048维特征，经过多个全连接层后得到256维特征；2)二维法线特征提取，投影生成二维法线特征图后，经过多个卷积层得到通道数为1 024的特征图，通过多个卷积核为2×2与5×5的卷积层得到通道数为1 024的特征图，并经过最大池化处理平铺生成1 024维的全连接层，之后分为2个支路经过全连接层分别得到256维特征，该网络采用Relu激活函数；3)将二维深度特征提取过程中得到的特征分别于二维法线特征提取过程中的两条支路进行特征

拼接,经过多个全连接层后,两支路分别得到三维特征和四维特征,代表工件位姿的*xyz*值以及表示旋转的四元数,将四元数转换为旋转矩阵后即可得到4×4的六维位姿矩阵。

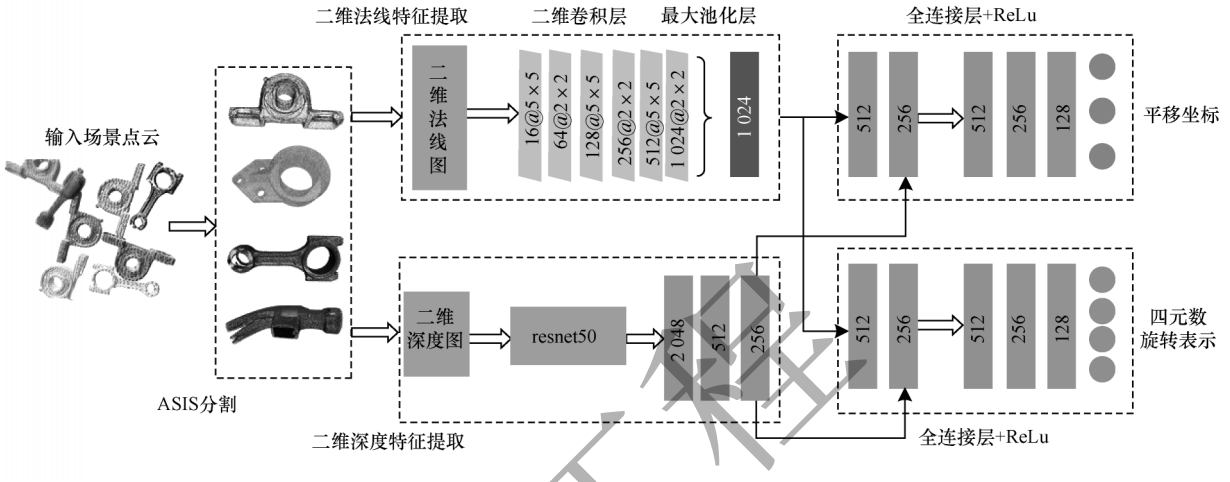


图6 特征融合网络框架

Fig.6 Framework of feature fusion network

2.4 损失函数

在基于深度学习的位姿回归中,常见的一种损失函数是计算使用真实位姿回归得到的点云和使用估计位姿回归得到的点云中对应点距离的平均值[5],记为CPLoss,计算公式如下:

$$C_{\text{CPLoss}}=\frac{1}{n}\sum_{\boldsymbol{P}\in\boldsymbol{M}}\left\|\boldsymbol{T}_{\text{g}}^{-1}\boldsymbol{P}-(\boldsymbol{T}_{\text{p}}\boldsymbol{T}_{\text{S-O}})^{-1}\boldsymbol{P}\right\| \tag{4}$$

其中:$\boldsymbol{M}$表示已事先采样的模型点云;n表示采样点个数;$\boldsymbol{T}_{\text{g}}$、$\boldsymbol{T}_{\text{p}}$分别表示标签位姿和估计位姿。需要注意的是,网络估计的位姿是分割后的局部点云到相机坐标系原点的模板点云的变换位姿,而计算损失函数使用模型点云到场景点云中的变换位姿,因此需要对变换矩阵求逆。

CPLoss损失函数可以有效地表示估计位姿回归的准确程度,但是对于一些对称物体而言,多个位姿可能对应同一个正确的姿态,从而使网络回归到另一个可代替的位姿上,造成损失函数给出不一致的训练信号。针对这一问题,本文采用类似于迭代最近点(Iterative Closest Point,ICP)算法的损失函数ICPLoss,计算估计位姿回归得到的点云中的每一个点离真实位姿回归得到点云的最近点的距离并取平均值,计算公式如下:

$$I_{\text{ICPLoss}}=\frac{1}{n}\sum_{\boldsymbol{P}_1\in\boldsymbol{M}}\min_{\boldsymbol{P}_2\in\boldsymbol{M}}\left\|\boldsymbol{T}_{\text{g}}^{-1}\boldsymbol{P}_2-(\boldsymbol{T}_{\text{p}}\boldsymbol{T}_{\text{S-O}})^{-1}\boldsymbol{P}_1\right\| \tag{5}$$

3 实验验证

在进行位姿估计前,需要对获取的场景点云进行实例分割。本文采用ASIS[21]实例分割算法,根据同类实例点的特征向量相近、不同类实例点的特征向量相差较远的原则进行实例分割。因此,工件在无遮挡堆叠的情况下,分割效果是非常理想的,而由于本文在抓取过程中每次仅对场景中的一个实例进行位姿估计,对于遮挡堆叠严重的场景点云,将最上层实例分割分数最为理想的工件作为待抓取工件,可以避免遮挡堆叠带来的分割误差。图7(a)是真实场景的散乱堆叠工件,图7(b)、图7(c)是真实场景点云的两个分割实例。图8是针对图7(a)的真实场景位姿估计实例,通过网络估计工件位姿并利用ICP进行位姿细化得到可抓取工件的精确六维位姿,接着通过机器人进行工件的抓取,重复以上过程即是一次完整的散乱工件抓取的流程。图8(a)~图8(h)显示了将模型点云基于估计得到的精确六维位姿变换回场景中,可以看出模型点云和场景中的目标点云基本重合。

(a)真实场景的散乱堆叠工件

(b)无堆叠的点云分割实例 (c)堆叠严重的点云分割实例

图7 真实场景的点云分割实例

Fig.7 Examples of point clouds segmentation of real scene

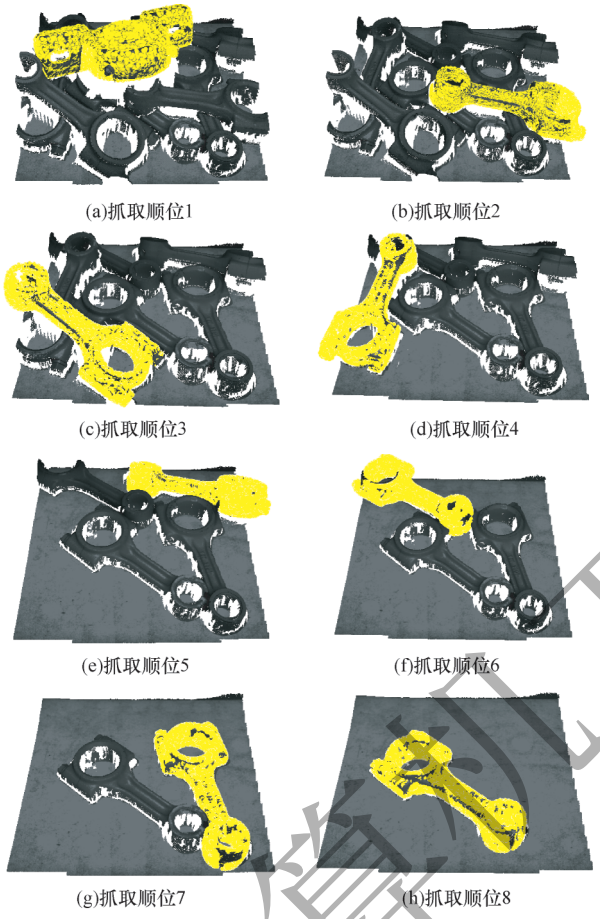


图8　真实场景的位姿估计实例

Fig.8　Examples of pose estimation of real scene

3.1　实验参数设置

本文针对4种不同的工业零件进行六维位姿估计实验。在数据集中，每类工件都有8 000个分割后的点云样本作为训练集，2 000个样本作为测试集，每个样本包含2 048个采样点。对于非对称工件，由于本文采用的工件尺寸为10~20 cm，因此将CPLoss小于工件尺寸最大直径的1/10视为位姿估计正确。对于对称工件，判别标准是ICPLoss的大小，经过实验评估，轴承座1和连杆的回归损失ICPLoss分别小于2 mm和1.4 mm时，可视为位姿回归正确。如果训练的工件尺寸和本文相差很大，则需重新选定合理的阈值。

3.2　与传统位姿估计方法的性能对比

将本文方法与粗配准+ICP、PPF、深度+ICP方法进行对比，如表1所示，其中最优指标值用加粗字体标示。可以看出，使用深度特征和法线特征相融合的位姿估计方法比仅使用深度特征的位姿估计方法具有更高的估计准确率。对于对称工件而言，即轴承座1和连杆，PPF和本文方法均能达到很高的估计准确率，而粗配准+ICP方法效果较差；对于非对称工件而言，即轴承座2和榔头，本文方法在准确率上远超粗配准+ICP和PPF方法。

表1　工业零件在不同方法下的位姿估计准确率

Table 1　Accuracy of pose estimation of industrial parts with different methods　%

工业零件	位姿估计准确率			
	粗配准+ICP方法	PPF方法	深度+ICP方法	本文方法
轴承座1	51.5	98.7	86.5	**99.5**
轴承座2	51.4	67.4	82.7	**95.7**
连杆	76.6	**98.8**	84.4	98.6
榔头	68.4	59.4	90.7	**98.8**

图9给出了PPF匹配错误的两种情况，可以看出榔头正反面是两个类似的平面，而当分割后的输入点云是类似于图中这样的局部平面时，PPF或者粗配准+ICP方法很可能会将其匹配到工件的一个类似平面上，方向和位置完全错误。由此得出，传统方法是通过计算特征点对的方式进行匹配的，它们没有获取输入点云的局部外形特征和几何特征，在有相似特征的情况下很容易匹配错误，而本文方法没有出现这方面的问题。

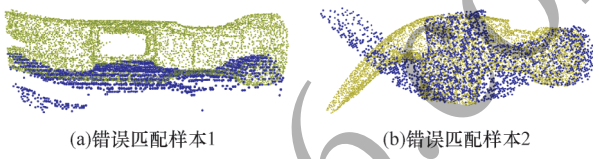


图9　PPF错误匹配样本

Fig.9　PPF error matching samples

表2给出了3种方法的平均位姿估计时间对比结果。本文所有涉及ICP位姿细化的地方，均将终止条件定为两次迭代的结果之差小于10^{-6} m。可以看出，本文方法非常高效，估计一次的时间远少于粗配准+ICP方法的时间，也略快于PPF方法。同时，对于增加的法线特征支路，其浮点运算量为2.1×10^{8}，而深度特征支路resnet50的浮点运算量为3.8×10^{9}，约为前者的1.8倍。可见，特征融合网络相比单特征网络运算复杂度和位姿估计时间并未明显增加，这是因为整个网络的运算复杂度主要由深度特征支路以及之后的全连接层决定。

表2　3种方法的平均位姿估计时间对比

Table 2　Comparison of the average pose estimation time of three methods　s

估计方法	平均位姿估计时间
粗配准+ICP方法	4.360
PPF方法	0.165
本文方法	0.154

3.3　其他因素对估计结果的影响

由于真实场景中分割得到的单个点云的点个数是不确定的，而本文训练采用的数据集中每个样本都是2 048个点，因此本文将各种采样点数的点云分

别输入训练好的模型进行位姿预测并统计各种方法在不同采样点数下的位姿估计准确率。图10给出了采用3种方法的榔头工件位姿估计准确率对比结果。可以看出,在不同采样点数下,本文方法在估计准确率上未有明显变化,说明本文训练的模型可以针对不同点数的点云进行位姿估计,而其他两种方法在点数变少时准确率出现递减的情况。

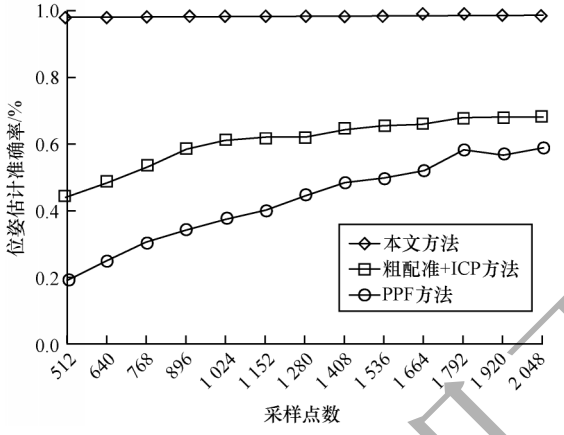


图10 不同采样点数下位姿估计准确率的对比

Fig.10 Comparison of accuracy of pose estimation under different sampling points

针对噪声对法线特征图的影响,本文对测试数据的每一个点加入随机噪声Δ:

$$\Delta=\text{random}(-1,1)\times s_{\text{物体尺寸}}\times\beta \tag{6}$$

其中:β是比例系数,为使噪声的影响更加显著,本文将其设定为0.05。表3给出了本文方法在无噪声的测试集、添加噪声样本的测试集以及添加噪声样本的训练集上进行训练后得到的测试结果。可以看出,噪声对位姿估计准确率的影响较小,并且将一些带有噪声的样本加入训练集后可以避免该影响。因此,经过实验证实,本文方法对噪声的鲁棒性较强。

表3 噪声对本文方法位姿估计准确率的影响

Table 3 The effect of noise on the accuracy of the proposed pose estimation method %

工业零件	估计准确率		
	无噪声的测试集	添加噪声的测试集	添加噪声的训练集
轴承座1	99.5	98.2	99.4
轴承座2	95.7	94.8	96.2
连杆	98.6	96.9	98.8
榔头	98.8	97.8	98.6

4 结束语

本文提出一种基于深度学习的点云位姿估计方法,将分割后的单个点云投影到二维平面,生成深度特征图和法线特征图,用于提取点云的局部表面特征和几何特征,从而估计出准确的六维位姿。在仿真数据集和真实数据集上的实验结果验证了该方法的有效性,并表明其在一定程度上解决了传统位姿估计方法计算量大且鲁棒性差的问题。但由于本文方法是基于点云的实例分割,位姿估计的准确率依赖于实例分割的准确率,因此下一步将对分割和位姿估计进行有效结合形成端到端模型,在保证点云语义实例分割准确率的前提下进一步提升算法实时性。

参考文献

[1] TEKIN B, SINHA S N, FUA P. Real-time seamless single shot 6D object pose prediction[C]//Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 292-301.

[2] ZAKHAROV S, SHUGUROV I, ILIC S. DPOD: 6D pose object detector and refiner[C]//Proceedings of 2019 IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2019: 1941-1950.

[3] LI Z, WANG G, JI X. CDPN: coordinates-based disentangled pose network for real-time RGB-based 6-DoF object pose estimation[C]//Proceedings of 2019 IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2019: 7678-7687.

[4] KEHL W, MANHARDT F, TOMBARI F, et al. SSD-6D: making RGB-based 3D detection and 6D pose estimation great again[C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 1521-1529.

[5] RAD M, LEPETIT V. BB8: a scalable, accurate, robust to partial occlusion method for predicting the 3D poses of challenging objects without using depth[C]// Proceedings of 2017 IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 3828-3836.

[6] PENG S, LIU Y, HUANG Q, et al. PVNet: pixel-wise voting network for 6DoF pose estimation[C]// Proceedings of 2019 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2019: 4561-4570.

[7] SUNDERMEYER M, MARTON Z C, DURNER M, et al. Implicit 3D orientation learning for 6D object detection from RGB images[C]//Proceedings of 2018 European Conference on Computer Vision. New York, USA: ACM Press, 2018: 699-715.

[8] 朱建新,沈东羽,吴钪. 基于激光点云的智能挖掘机目标识别[J]. 计算机工程,2017,43(1): 297-302.
ZHU J X, SHEN D Y, WU K. Target recognition for intelligent excavator based on laser point cloud[J]. Computer Engineering, 2017, 43(1): 297-302. (in Chinese)

[9] KRULL A, BRACHMANN E, MICHEL F, et al. Learning analysis-by-synthesis for 6D pose estimation in RGB-D images[C]//Proceedings of 2015 IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2015: 954-962.

[10] MICHEL F, KIRILLOV A, BRACHMANN E, et al. Global hypothesis generation for 6D object pose estimation[C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 462-471.

[11] WOHLHART P, LEPETIT V. Learning descriptors for object recognition and 3D pose estimation[C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2015: 3109-3118.

[12] KEHL W, MILLETARI F, TOMBARI F, et al. Deep learning of local RGB-D patches for 3D object detection and 6D pose estimation [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 205-220.

[13] WANG C, XU D, ZHU Y, et al. Dense fusion: 6D object pose estimation by iterative dense fusion[C]//Proceedings of 2019 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2019: 3343-3352.

[14] DROST B, ULRICH M, NAVAB N, et al. Model globally, match locally: efficient and robust 3D object recognition [C]//Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2010: 998-1005.

[15] JOEL V, CHYI-YEU L, XAVIER L, et al. A method for 6D pose estimation of free-form rigid objects using point pair features on range data[J]. Sensors, 2018, 18(8): 2678-2682.

[16] 龚学健. 基于RealSense的散乱零件三维目标识别[D]. 哈尔滨：哈尔滨工业大学，2018.
GONG X J. 3D object recognition of scattered parts based on realsense [D]. Harbin: Harbin Institute of Technology, 2018. (in Chinese)

[17] QI C R, SU H, MO K, et al. PointNet: deep learning on point sets for 3D classification and segmentation [C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 652-660.

[18] 罗元，王薄宇，陈旭. 基于深度学习的目标检测技术的研究综述[J]. 半导体光电，2020，41(1)：1-10.
LUO Y, WANG B Y, CHEN X. Research progresses of target detection technology based on deep learning[J]. Semiconductor Optoelectronics, 2020, 41(1): 1-10. (in Chinese)

[19] QI C R, LIU W, WU C, et al. Frustum PointNets for 3D object detection from RGB-D data[C]//Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 918-927.

[20] DYRSTAD J S, BAKKEN M, GROTLI E I, et al. Bin picking of reflective steel parts using a dual-resolution convolutional neural network trained in a simulated environment[C]//Proceedings of 2018 IEEE International Conference on Robotics and Biomimetics. Washington D. C. , USA: IEEE Press, 2018: 530-537.

[21] MITASH C, BEKRIS K E, BOULARIAS A. A self-supervised learning system for object detection using physics simulation and multi-view pose estimation[C]// Proceedings of 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems. Washington D. C. , USA: IEEE Press, 2017: 545-551.

[22] WANG X, LIU S, SHEN X, et al. Associatively segmenting instances and semantics in point clouds[C]//Proceedings of 2019 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2019: 4096-4105.

编辑　陆燕菲