

硕 士 研 究 生 读 书 报 告



题目《PVNet: Pixel-wise Voting Network for 6DoF Pose Estimation》读书报告

作者姓名 滕燕斌

作者学号 22151005

指导教师

学科专业 软件工程

所在学院 软件学院

提交日期 二〇 二一 年 十二 月

**摘要**

本文将会介绍《PVNet: Pixel-wise Voting Network for 6DoF Pose Estimation》一文的内容，主要包括一下几个方面：1）文章所要解决的问题；2）解决问题的算法；3）论文的创新点；3）方法的实际应用场景。

**关键字：**物体姿态估计；深度学习

1 引言

近些年AR、自动驾驶等兴欣技术广为人们所讨论，在这些技术中都需要面对一个关键的技术问题：物体六自由度位姿估计。本文所介绍的文章《PVNet: Pixel-wise Voting Network for 6DoF Pose Estimation》[1]将会从单张RGB图像去识别检测物体并估计物体的六自由度位姿。在这个技术问题中，有许多个方面具有挑战性，如其他物体的遮挡、光照的变化、背景中非目标物体的干扰等。

在过去的方法的中，有手工提取特征来估计物体位姿[2, 3, 4]，有使用深度学习的方法进行特征提取以估计物体姿态[5, 6, 7, 8]，但这些方法的泛化能力都不够好。有一些算法[9, 10, 11]，首先利用CNN去回归2D关键点，然后通过 Perspective-n-Point (PnP)计算 6D姿态参数，这些算法都取得了不错的效果，但面对存在遮挡和截断的情况时，算法泛化能力较差。

为了解决上面这些算法存在的问题，作者提出了PVNet，使用像素级投票网络的6D位姿估计。

**2 算法介绍**

对物体位姿估计的流程如图1所示，流程主要分被两大步骤，第一步，我们输入一张图片到训练好的PVNet中，PVNet输出目标物体的语义分割和指向物体关键点的向量场，随后，通过Ransac-voting，从方向向量场计算出物体的关键点。在关键点的生成过程中，PVNet同时还会生成物体关键点的概率分布，也就是关键点空间分布的mean和covariance；第二步，利用关键点的不确定性，使用改进的PnP算法完成物体姿态的估计。

因为该方法使用像素级的投票网络(PVNet)以类似搜查的方式检测2D关键点，其对于目标遮挡或者截断的鲁棒性是很强的。而基于RANSAC的投票机制，给出每个关键点的概率分布，使得我们可以用不确定驱动的PnP来估计6D的姿态

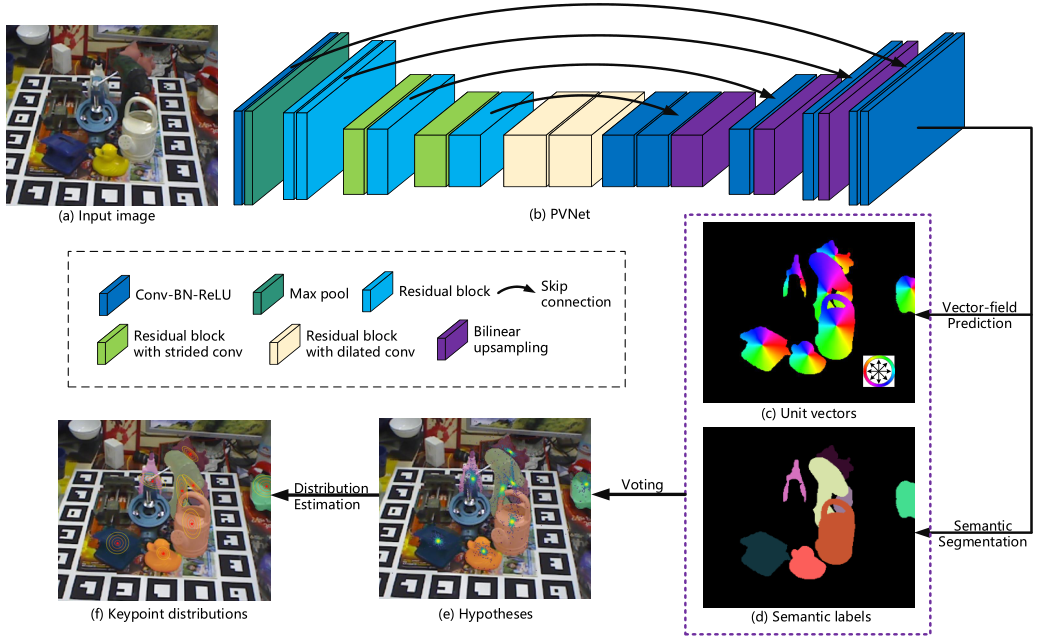


图 1

对于物体的关键点，作者发现物体关键点的定义对关键点的定位影响很大。之前的深度学习方法将物体的关键点简单地定义为物体在3D空间中的bounding box的八个角点。这8个角点在2D图片中可能离物体比较远，加大了关键点检测的难度。于是，作者使用farthest point sampling算法，生成物体表面的8个关键点。如图2所示，图中的(a)是一个三维对象模型及其三维包围盒。(b) 是PVNet对包围盒角的假设。(c) 是PVNet对物体表面选定的一个关键点产生的假设。（c）所使用方法的关键点方差较小，说明该方法比边界框角更容易定位关键点。

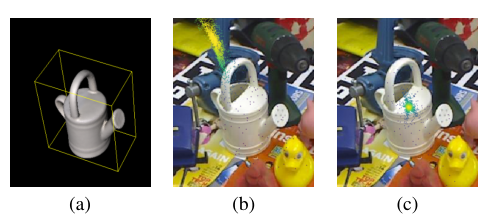


图 2

获得目标对象的2D关键位置后，可以使用 PnP 算法求解6D姿态。如之前很多的算法采用EPnP，但是他们忽略每个关键点的置信度是由差异的，存在不确定性，在使用PnP进行求解的时候，我们应该考虑到这些问题。如第3.1节所述，投票机制出来的结果是每个关键的对应的概率分布，对于每个关键点的评估标准，是计算他们的均值以及协方差（其中，k = 1, 2, 3, …, K）, 通过最小化Mahalanobis distance 计算6D姿态估算pose (R,t)。

其中代表关键点的3D坐标，是的二维映射，EPnP算法根据四个关键点对参数R和t进行初始化，其协方差矩阵轨迹最小。然后利用LevenbergMarquardt算法求解，在其他的一些工作中还通过最小化逼近值来考虑特征的不确定性Sampson错误，在文章的方法中，我们直接最小化重投影误差。

**3 新点与不足**

文章的创新之处在于提出了一种新的二维目标关键点表示和改进的PnP位姿估计算法。作者的方法使用像素级的投票网络(PVNet)以类似搜查的方式检测2D关键点，其对于目标遮挡或者截断的鲁棒性是很强的。基于RANSAC的投票机制，给出每个关键点的概率分布，使得可以用不确定驱动的PnP来估计6D的姿态。

对于部分可见的小物体，PVNet估计出的姿态会有较大的误差，如图3所示，其中蓝色是预测的姿态，绿色是真实姿态。

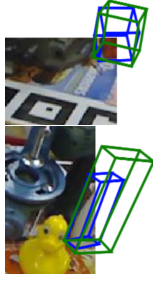


图 3

**4 应用**

在增强现实技术中，我们需要让虚拟世界的东西与真实世界互动，以实现超越现实的感官体验。比如，我想在一张桌子的某个地方摆上一个虚拟的花瓶，我就需要实时检测画面中桌子的位姿，才能让花瓶始终摆在桌子对的位置上。

虽然文章中并没有说可以估计人体姿态，但理论上也是可以应用的。只需要将人体的关节分成几个部分，看成是多物体姿态估计即可。这样我们就可以使用普通的相机来完成动作识别、运动捕捉等工作了。

参考文献

[1] Sida Peng, Yuan Liu, Qixing Huang, Xiaowei Zhou, and Hujun Bao. Pvnet: Pixel-wise voting network for 6dof pose estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4561–4570, 2019.

[2] D. G. Lowe. Object recognition from local scale-invariant features. In ICCV, 1999.

[3] V. Lepetit, P. Fua, et al. Monocular model-based 3d tracking of rigid objects: A survey. Foundations and Trends in Computer Graphics and Vision, 1(1):1–89, 2005.

[4] S. Hinterstoisser, V. Lepetit, S. Ilic, S. Holzer, G. Bradski, K. Konolige, and N. Navab. Model based training, detection and pose estimation of texture-less 3d objects in heavily cluttered scenes. In ACCV, 2012.

[5] H. Su, C. R. Qi, Y. Li, and L. J. Guibas. Render for cnn: Viewpoint estimation in images using cnns trained with rendered 3d model views. In ICCV, 2015.

[6] W. Kehl, F. Manhardt, F. Tombari, S. Ilic, and N. Navab. Ssd6d: Making rgb-based 3d detection and 6d pose estimation great again. In ICCV, 2017.

[7] Y. Xiang, T. Schmidt, V. Narayanan, and D. Fox. Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes. In Robotics: Science and Systems (RSS), 2018.

[8] M. Bui, S. Zakharov, S. Albarqouni, S. Ilic, and N. Navab. When regression meets manifold learning for object recognition and pose estimation. In ICRA, 2018.

[9] G. Pavlakos, X. Zhou, A. Chan, K. G. Derpanis, and K. Daniilidis. 6-dof object pose from semantic keypoints. In ICRA, 2017.

[10] M. Rad and V. Lepetit. Bb8: A scalable, accurate, robust to partial occlusion method for predicting the 3d poses of challenging objects without using depth. In ICCV, 2017.

[11] B. Tekin, S. N. Sinha, and P. Fua. Real-time seamless single shot 6d object pose prediction. In CVPR, 2018.