**（一）基于视觉的从****物体定位，****姿态估计，****抓取检测到****运动规划的机器人抓取**

**Vision-based Robotic Grasping from Object Localization, Pose Estimation, Grasp**

**Detection to Motion Planning**

**----Guoguang Du , Kai Wang and Shiguo Lian**

**Cloudminds Technologies（arXiv）**

**一、科学问题**

**1.1 本文所涉及科学问题**

工业机器人、物体定位、姿态估计、抓取检测、运动规划。

**1.2 同行专家如何解决**

目前，有许多方法实现了这个工程的一小部分，比如说物体检测结合6D姿态估计，基于姿态估计的抓取识别，端到端抓取识别和端到端运动规划。

**1.3 本文所解决的问题**

本文主要解决的问题是1）如何物体定位，包括物体识别以及分割模型。2）姿态估计，包括基于RGB和RGB-D的方法。3）抓取识别，主要用深度学习的方法。4）运动规划，主要是机械臂的路径规划和运动控制，主要基于强化学习的方法。

**1.4 本文解决方案效果**

经过本次实验的研究发现本文中所用的针对各个部分的方法可以很好的实现基于视觉的机器人抓取，对于纹理不是很明显并且物体所处环境比较闭塞的情况的处理效果也很好了。

**二、研究内容**

**2.1 理论与方法介绍**

（论文主要研究内容的提出，主要技术路线、理论与方法介绍）

1）物体定位技术，主要通过物体定位和图像分割技术，主要用到了基于深度学习的算法，比如说深度卷积神经网络。2）物体分割技术，分割技术有非常多的方法，比如深度网络方法是现在用的比较多的，3D分割主要用PointNet，PointCNN。3）6D姿态估计，主要有四种类型，通信，匹配，投票和衰弱。

**2.2 验证分析与实验效果**

本文利用视频库中的大量视频来训练网络和验证实验方法，实验效果很好，让读者很容易的就能知道各种方法的区别和优缺点。

**三、论文存在问题及后续研究重点**

**3.1 论文存在问题**

传统抓取方法很难从当前抓取的物体去突破而能抓取一些以前没见过的物体。而且，不同物体之间因为闭塞而造成的信息不完全也会是一个挑战。基于深度学习的方法又比较依赖于大量的训练数据。

**3.2 后续研究重点**

未来机器抓取的突破主要靠四个方面的协助：大量训练数据的获得，语义感知能力的提升，抓取决定，数据训练。

**四、该问题相关研究成果**

**4.1 相关论文一**

**（1）题目**：Yolo3d:End-to-end real-time 3d oriented object bounding box detection from lidar point cloud. In European Conference on Computer Vision

**（2）作者介绍**：Waleed Ali, Sherif Abdelkarim, Mahmoud Zidan, Mohamed Zahran, and Ahmad El Sallab

**（3）摘要**:Object detection and classification in 3D is a key task in Automated Driving (AD). LiDAR sensors are employed to provide the 3D point cloud reconstruction of the surrounding environment, while the task of 3D object bounding box detection in real time remains a strong algorithmic challenge. In this paper, we build on the success of the one-shot regression meta-architecture in the 2D perspective image space and extend it to generate oriented 3D object bounding boxes from LiDAR point cloud. Our main contribution is in extending the loss function of YOLO v2 to include the yaw angle, the 3D box center in Cartesian coordinates and the height of the box as a direct regression problem. This formulation enables real-time performance, which is essential for automated driving. Our results are showing promising figures on KITTI benchmark, achieving real-time performance (40 fps) on Titan X GPU.

**4.2 相关论文二**

**（1）题目**：Generalization of human grasping for multi-fingered robot hands

**（2）作者介绍**：Heni Ben Amor, Oliver Kroemer, Ulrich Hillenbrand, Gerhard Neumann, and Jan Peters

**（3）摘要**：Multi-fingered robot grasping is a challenging problem that is difficult to tackle using hand-coded programs. In this paper we present an imitation learning approach for learning and generalizing grasping skills based on human demonstrations. To this end, we split the task of synthesizing a grasping motion into three parts: (1) learning efficient grasp representations from human demonstrations, (2) warping contact points onto new objects, and (3) optimizing and executing the reach-and-grasp movements. We learn low-dimensional latent grasp spaces for different grasp types, which form the basis for a novel extension to dynamic motor primitives. These latent-space dynamic motor primitives are used to synthesize entire reach-and-grasp movements. We evaluated our method on a real humanoid robot. The results of the experiment demonstrate the robustness and versatility of our approach.

**4.3 相关论文三**

**（1）题目**：Segnet: A deep convolutional encoder-decoder architecture for image segmentation

**（2）作者介绍**：Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla

**（3）摘要：**We present a novel and practical deep fully convolutional neural network architecture for semantic pixel-wise segmentation termed SegNet. This core trainable segmentation engine consists of an encoder network, a corresponding decoder network followed by a pixel-wise classification layer. The architecture of the encoder network is topologically identical to the 13 convolutional layers in the VGG16 network. The role of the decoder network is to map the low resolution encoder feature maps to full input resolution feature maps for pixel-wise classification. The novelty of SegNet lies is in the manner in which the decoder upsamples its lower resolution input feature map(s). Specifically, the decoder uses pooling indices computed in the max-pooling step of the corresponding encoder to perform non-linear upsampling. This eliminates the need for learning to upsample. The upsampled maps are sparse and are then convolved with trainable filters to produce dense feature maps. We compare our proposed architecture with the widely adopted FCN and also with the well known DeepLab-LargeFOV, DeconvNet architectures. This comparison reveals the memory versus accuracy trade

**(二)** **第四次国际工厂会议对恢复6D物体姿态估计的总结**

**A Summary of the 4th International Workshop**

**on Recovering 6D Object Pose**

**--** **Tomas Hodan, Rigas Kouskouridas, Tae-Kyun Kim, Federico Tombari，Kostas Bekris, Bertram Drost, Thibault Groueix, Krzysztof WalasVincent， Lepetit Ales Leonardis, Carsten Steger, Frank Miche， Caner Sahin, Carsten Rother, Jir Matas**

**(ECCV)**

**一、科学问题**

**1.1 本文所涉及科学问题**

6D物体姿态估计

**1.2 同行专家如何解决**

近年来，6D物体姿态估计领域受到越来越多的关注。该领域的一项重大成就是ECCV 2018颁发的最佳论文奖获奖的是Martin Sundermeyer，Zoltan Marton，Maximilian Durner，Manuel Brucker和Rudolph Triebel的论文名为“隐式3D定向学习6D”。消费者和工业级RGB-D传感器的引入，对于6D物体姿态估计的实质性改进是低的。最近发表了许多关于特定刚性物体的6D姿态估计的许多方法，但通常只与少数竞争对手进行在一小部分数据集上的比较。因此，尚不清楚在哪些情况下哪种方法表现较好。为了捕捉该领域的现状，我们在ICCV举办的第3次研讨会上组织了SIXD挑战赛，挑战的结果发布在BOP中基准论文并会在第4期研讨会上发表。

**1.3 本文所解决的问题**

对6D物体姿态的准确，快速和稳健的估计是非常重要的许多应用领域，如机器人操纵，增强现实，场景解释和自动驾驶。在机器人中，例如，6D对象

姿势有利于空间推理，并允许终结者对行动起作用对象。在增强现实场景中，可以使用对象姿势来增强通过增加具有额外信息的对象来实现对现实的感知作为装配的提示。

研讨会涉及以下主题：（a）6D物体姿态估计（a.k.a.三维物体检测）和跟踪，（b）三维物体建模和重建结构，（c）表面表示和登记，（d）对遮挡的和背景杂乱的解决，（e）多物体实例检测，（f）姿态估计 - 非刚性物体和物体类别，（g）机器人抓握和抓握ff ordances，以及（h）对象操纵和交互。

**1.4 本文解决方案效果**

该文章是第四届国际研讨会关于恢复6D物体姿态估计的总结，是ECCV 2018年举办的，有四个特邀演讲，以及研讨会论文的口头和海报展示，并介绍了用于6D物体姿态估计的BOP基准。从目前的BOP基准评分中可以明显看出，用于6D物体姿态估计的快速且稳健的方法呼声依然很高。

我们我想鼓励相关方法的作者继续向bop.felk.cvut.cz的在线评估系统提交最新研究成果。我们将很快公布下一个研讨会版本的排行榜，计划用于即将举行的重要会议。

**二、研究内容**

**2.1 理论与方法介绍**

本文首先介绍了从3D向6D的转变，我们已经有什么基础，也就是我们可以根据以前的研究怎样突破。我们要从识别简单环境下的单个物体向十分复杂环境的转变并提升精度，要促进算法的发展才能适应现在越来越多的应用和越来越复杂的物体环境。为克服上述限制，深入学习这项任务，特别强调了深度学习的两个主要研究方向且被用来改进现有技术：i）3D描述符的定义用于3D数据; ii）来自RGB（或单眼）数据的6D对象姿势估计。

至于第一个方面，我们简要介绍了3D描述符的开发对于无组织的3D表示，从手工制作的表示开始直到最近的自我学习。关于深度学习的影响在这个领域的一个重要方面的是某些3D表示，如经常用于6D刚性姿态估计任务的点集和3D网格因其内在的无组织性而不适合卷积。因此，学习3D功能的一个重要步骤是引入PointNet 等方法，相反，如3D Match ，可以直接在点云或网格上操作需要对数据进行体素化或直方图化。最近的全球和全面用于点云处理和场理解的卷积体系结构。

至于第二个研究方向,根据单目信息估计物体的6D姿势的直觉依赖于一个事实,即人类通常可以从单目线索对周围3D空间中物体的姿势有大致的想象,前提是他们熟悉对象的形状。区分这些方法的一个独特特征是网络被训练推断的输出类型:从投影边界框的 8 个角到 6D 姿势的回归或视点分类和平面内旋转。建议使用CNN从对RGB补丁中学习6D姿势细化。单目姿势估计可以通过深度学习有希望地进行,尽管还不如深度传感器准确,但两种跟踪方法之间的实际序列的"定性比较"也显示了这一点。最后,部署单目6D姿势与单目语义SLAM(如CNN-SLAM)联合部署,后者还利用深度学习获得密集的语义重构,看起来是一个有趣的方向,可以向全眼(即语义几何)场景从单目数据理解。

**2.2 验证分析与实验效果**

经过研讨会的讨论以及总结，我们可以通过几位发言人所分享的方法促进基于深度学习的6D物体姿态估计的发展，以及能够共同投入到这方面的研究中，并且能够解决将来可能会发生的一系列问题。

**三、论文存在问题及后续研究重点**

**3.1 论文存在问题**

虽然将评估系统扩展到其他任务很容易,但扩展方法却不简单。这个研讨会的另一个话题错误姿势机制，但这个机制用到的VSD有一个局限性，那就是它不能识别颜色。

**3.2 后续研究重点**

未来的计划是一步一步地发展,能够在不久的将来将更复杂的变体添加到在线评估系统,以及 6D 检测任务,其中未提供有关对象存在的先前信息。

**四、该问题相关研究成果**

**4.1 相关论文一**

**（1）题目**：6D Pose Estimation using an Improved Method based on Point Pair Features.

**（2）作者介绍**：Vidal Joel, Lin Chuiyue, Marti Robert.

**（3）摘要**: The Point Pair Feature (Drost et al. 2010) has been one of the most successful 6D pose estimation method among model-based approaches as an efficient, integrated and compromise alternative to the traditional local and global pipelines. During the last years, several variations of the algorithm have been proposed. Among these extensions, the solution introduced by Hinterstoisser et al. (2016) is a major contribution. This work presents a variation of this PPF method applied to the SIXD Challenge datasets presented at the 3rd International Workshop on Recovering 6D Object Pose held at the ICCV 2017. We report an average recall of 0.77 for all datasets and overall recall of 0.82, 0.67, 0.85, 0.37, 0.97 and 0.96 for hinterstoisser, tless, tudlight, rutgers, tejani and doumanoglou datasets, respectively.

**4.2 相关论文二**

**（1）题目**：A Probabilistic Framework for Stereo-Vision Based 3D Object Search with 6D Pose Estimation

**（2）作者介绍**：Jeremy Ma, Joel W.Burdic

**（3）摘要**：This paper presents a method whereby an autonomous mobile robot can search for a 3-dimensional (3D) object using an on-board stereo camera sensor mounted on a pan-tilt head. Search efficiency is realized by the combination of a coarse-scale global search coupled with a fine-scale local search. A grid-based probability map is initially generated using the coarse search, which is based on the color histogram of the desired object. Peaks in the probability map are visited in sequence, where a local (refined) search method based on 3D SIFT features is applied to establish or reject the existence of the desired object, and to update the probability map using Bayesian recursion methods. Once found, the 6D object pose is also estimated. Obstacle avoidance during search can be naturally integrated into the method. Experimental results obtained from the use of this method on a mobile robot are presented to illustrate and validate the approach, confirming that the search strategy can be carried out with modest computation.

**4.3 相关论文三**

**（1）题目**：Learning Compact Geometric Features

**（2）作者介绍**：Marc Khoury, Qian-Yi Zhou, Vladlen Koltun

**（3）摘要**：We present an approach to learning features that represent the local geometry around a point in an unstructured point cloud. Such features play a central role in geometric registration, which supports diverse applications in robotics and 3D vision. Current state-of-the-art local features for unstructured point clouds have been manually crafted and none combines the desirable properties of precision, compactness, and robustness. We show that features with these properties can be learned from data, by optimizing deep networks that map high-dimensional histograms into low-dimensional Euclidean spaces. The presented approach yields a family of features, parameterized by dimension, that are both more compact and more accurate than existing descriptors.