

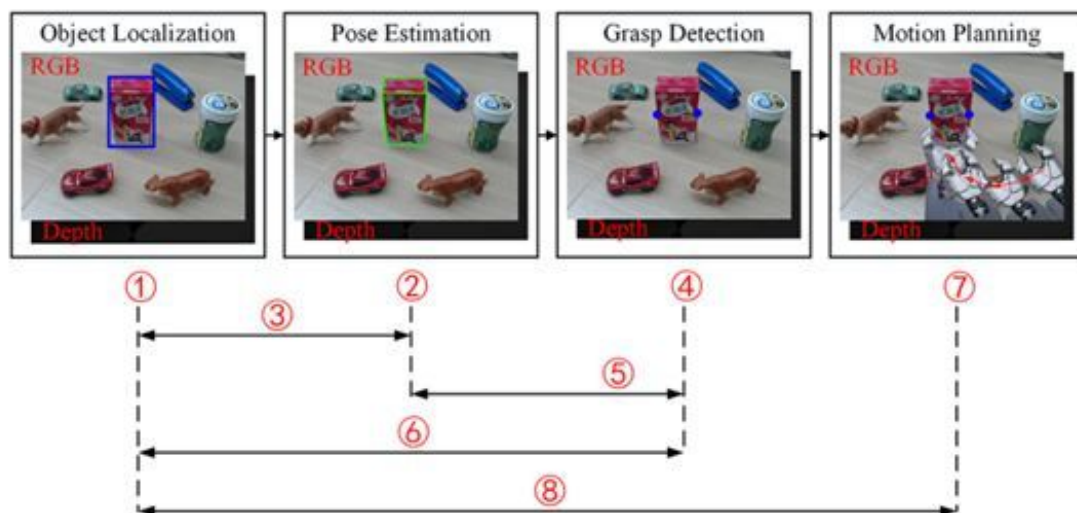
机器人抓取汇总|涉及目标检测、分割、姿态识别、抓取点检测、路径规划 - 云+社区 - 腾讯云

cloud.tencent.com/developer/article/1587723

来源：3D视觉工坊

介绍

为了完成机器人的抓取任务，机器人首先需要感知物体。随着传感器设备的不断发展，目前的机器人都配备了RGB摄像机和深度摄像机来获取丰富的环境信息。然而，原始的RGB-D图像对于机器人来说是简单的数字网格，在那里需要提取高层次的语义信息来实现基于视觉的感知。要抓取的目标对象的高层信息通常包含位置、方向和抓取位置。然后计算抓取规划以执行物理抓取。赋予机器人感知能力一直是计算机视觉和机器人学科的一个长期目标。机器人抓取不仅意义重大，而且早已被研究。机器人抓取系统由抓取检测系统、抓取规划系统和控制系统组成。其中，抓取检测系统是关键入口点，它分为三个任务：目标定位、姿态估计和抓取点检测。结合抓取规划，详细介绍四项任务。



早期的方法假设要抓取的对象被放置在一个背景简单的干净环境中，从而简化了对象定位任务，而在相对复杂的环境中，它们的能力相当有限。一些目标检测方法利用机器学习方法对基于手工二维描述符的分类器进行训练。但是，由于手工创建的描述符的限制，这些分类器的性能有限。近年来，深度学习已经开始主导图像相关的任务，如目标检测和分割。此外，从RGB图像到深度图像的训练数据，以及二维或三维输入的深度学习网络，极大地提高了目标定位的性能，促进了机器人抓取技术的发展。利用目标物的位置，可以进行抓取检测。早期的分析方法是直接分析输入数据的几何结构，根据力闭合或形状闭合来寻找合适的抓取点。然而，分析方法存在着费时、计算困难等问题。随后，随着大量三维模型的出现，可以分析数据驱动的方法，将三维模型数据库中的抓取转移到目标对象。Bohg等人介绍了数据驱动的抓取方法，并根据抓取配置是针对已知、熟悉还是未知对象计算的，将这些方法分为三组。一般来说，目标物体的6D姿态是完成这项任务的关键。基于RGB图像的方法和基于深度图像的方法都可以实现精确的姿态估计。然而，这些方法如部分配准方法易受传感器噪声或不完整数据的影响。通过深度学

习方法，可以直接或间接地从输入数据中估计姿态，从而获得对传感器噪声或不完整数据的抵抗力。还存在基于深度学习的方法，不需要6D姿势来进行抓取检测。抓取结构可以通过深卷积网络直接或间接回归。Caldera等人回顾了基于深度学习的机器人抓取检测方法。他们讨论了深度学习方法的每个元素如何提高机器人抓取检测的整体性能。此外，监督学习法、强化学习法也被用来直接完成特定的任务，如玩具装配等与抓取密切相关。

四项基本任务

下面将从目标定位、姿态估计、抓取点检测以及抓取规划四个方面详细介绍机器人抓取系统的主要内容。

目标定位 (Object localization)

大多数抓取方法首先需要计算目标在输入图像数据中的位置。这涉及到目标检测和分割技术。目标检测提供目标对象的矩形包围盒，目标分割提供目标对象的精确边界。后者提供了更精确的对象区域描述，而其计算则更耗时。

针对2D目标检测：主要包括SIFT、SURF、Faster RCNN、YOLO、SSD等常见目标检测算法。

针对3D目标检测：主要包括FPFH、SHOT等传统算法以及PointRCNN、VoxelNet等深度学习方式。

针对2D目标分割：主要包括FCN、UNet、DeepLab、Mask RCNN等方式。

针对3D目标分割：主要包括PointNet、PointNet++、PointCNN等方式。

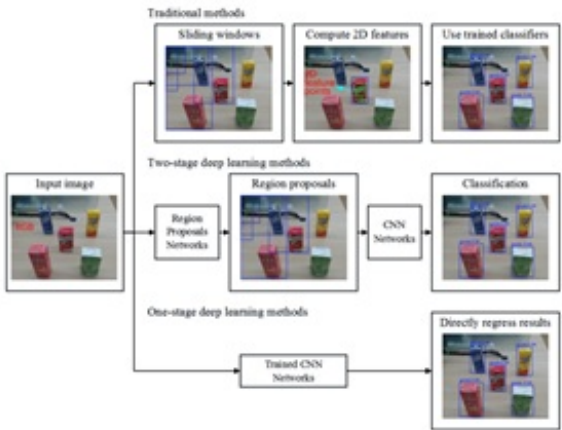
2D检测：

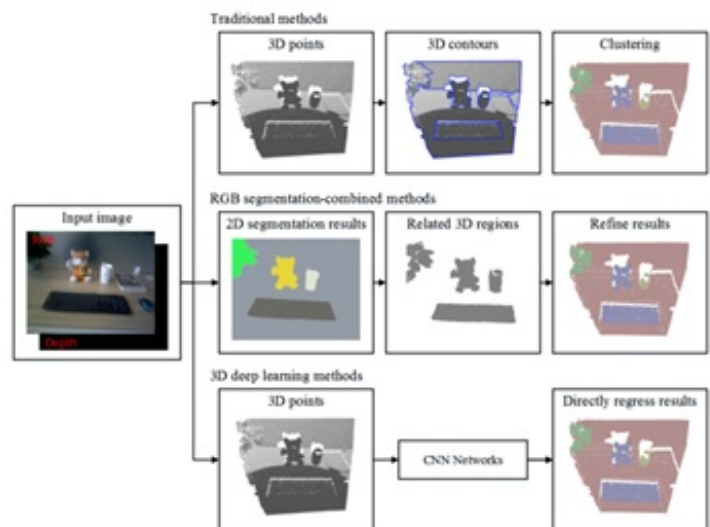
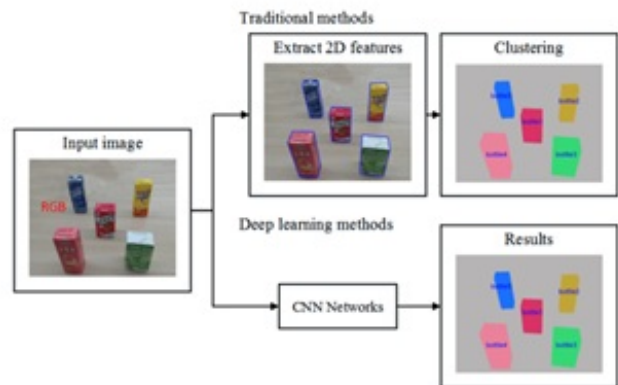
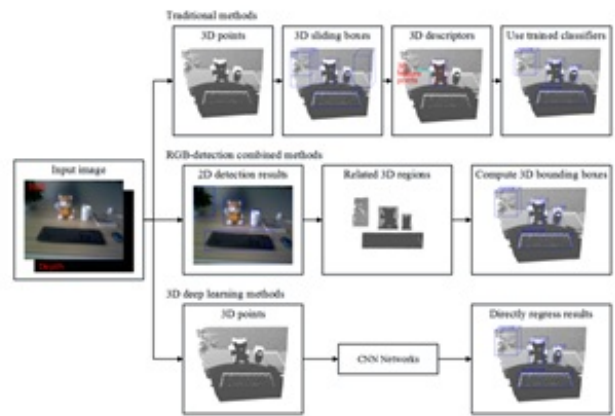
3D检测：

2D分割：

3D分割：

与传统的手工描述子方法相比，基于深度学习的方法取得了较好的效果。然而，对大量训练数据的需求和训练模型的泛化能力仍然具有挑战性。





6D姿态估计

6D姿态估计在增强现实、机器人操作、自动驾驶等领域发挥着重要作用。它帮助机器人知道要抓取的物体的位置和方向。姿态估计方法大致可分为四种，分别基于对应、模板、投票和回归。

1、不基于目标检测先验信息

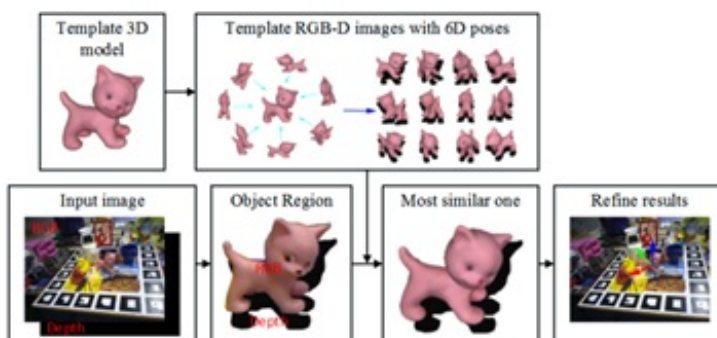
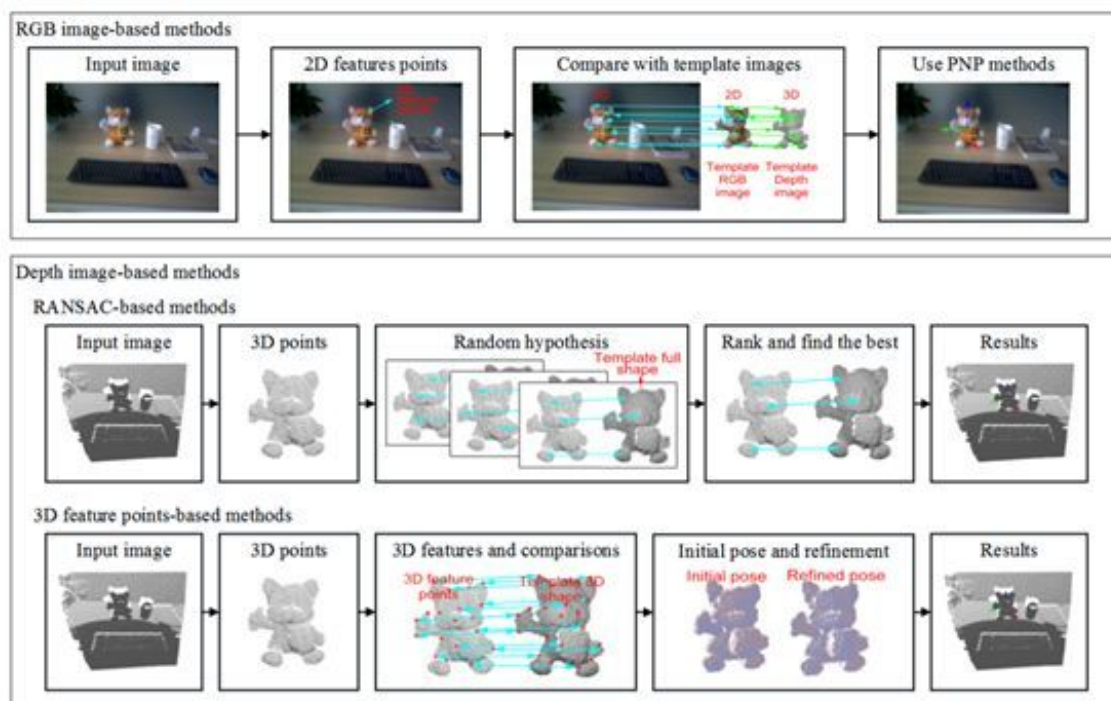
基于对应方法：

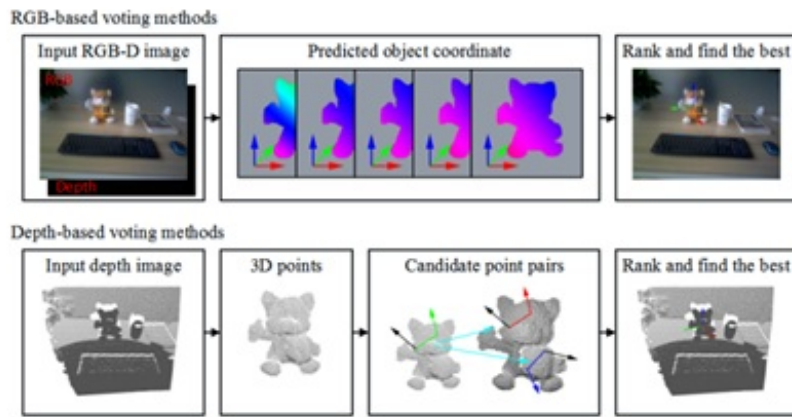
在2D点和3D点之间找到匹配，并使用PNP方法：SIFT、SURF、ORB

通过随机假设或三维描述寻找三维对应关系，并使用ICP对结果进行细化：FPFH、SHOT

基于模板方法：LineMod算法。从没有纹理的三维模型渲染图像，提取梯度信息进行匹配，并使用ICP对结果进行细化。

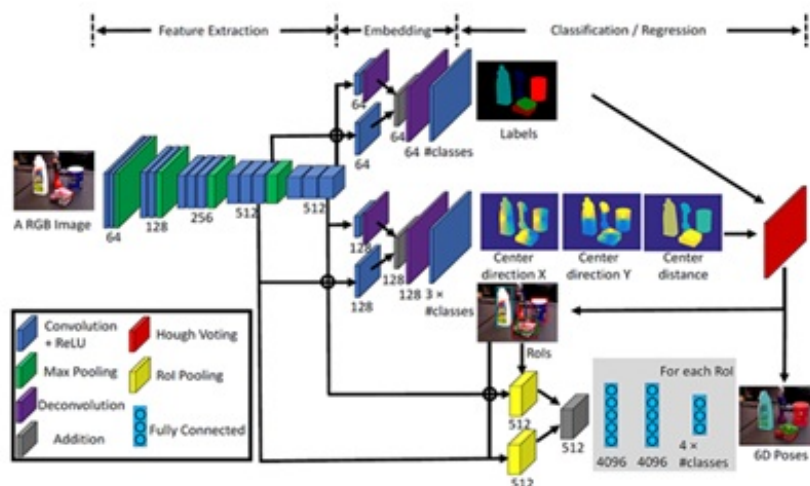
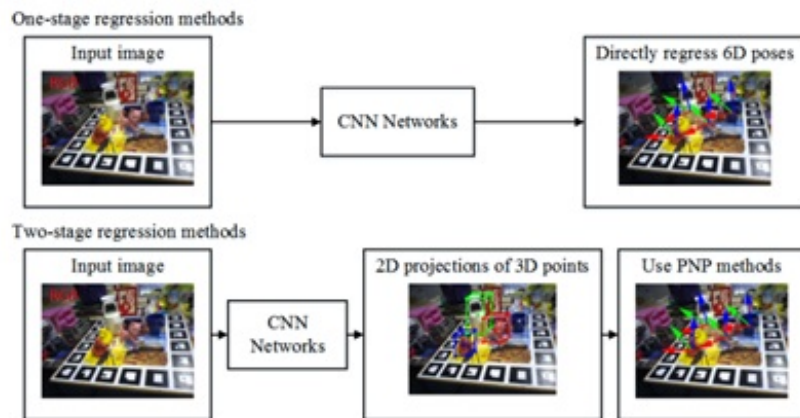
基于投票方法：PPF算法。基于三维点云或具有姿态的渲染RGB-D图像，每个局部预测一个结果，并使用RANSAC对结果进行优化。



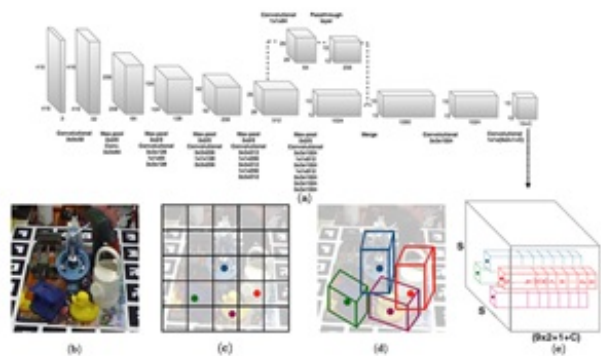


2、基于目标检测信息的6D姿态估计

这种方法也称为基于回归的方法，它同时完成目标检测和6D姿态估计。基于回归的方法：BB8、SSD6D、PoseCNN、Deep6DPose。



独立于目标检测的方法在通常发生遮挡的杂乱场景中显示出明显的局限性；基于目标检测的方法是缺乏足够的训练数据。



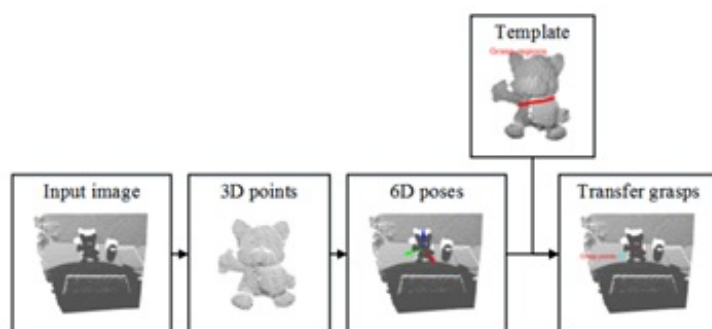
抓取点检测（Grasp Detection）

抓取检测被定义为能够识别任何给定图像中物体的抓取点或抓取姿势。抓取策略应确保对新物体的稳定性、任务兼容性和适应性，抓取质量可通过物体上接触点的位置和手的配置来测量。为了掌握一个新的对象，完成以下任务，有分析方法和经验方法。分析方法根据抓取稳定性或任务要求的运动学和动力学公式，选择手指位置和手部构型，经验方法根据具体任务和目标物体的几何结构，使用学习算法选择抓取。根据是否需要目标定位，需要确定目标的姿态，进一步将其分为三类：具有已知定位和姿势的方法、具有已知定位和无姿态的方法、无定位和无姿态的方法。

1、具有已知定位和姿势

针对已知目标的经验方法，利用姿态将已知目标的抓取点转换为局部数据。主要算法有：

- Multi-view self-supervised deep learning for 6d pose estimation in the amazon picking challenge.
- Silhonet: An RGB method for 3d object pose estimation and grasp planning.

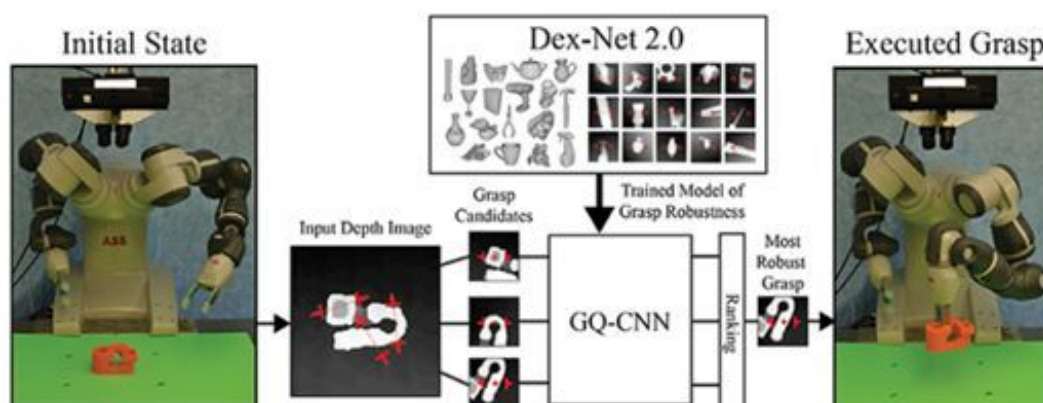


2、具有已知定位和无姿态的方法

主要方法：

- Automatic grasp planning using shape primitives.
- Part-based grasp planning for familiar objects.
- Transferring grasp configurations using active learning and local replanning.

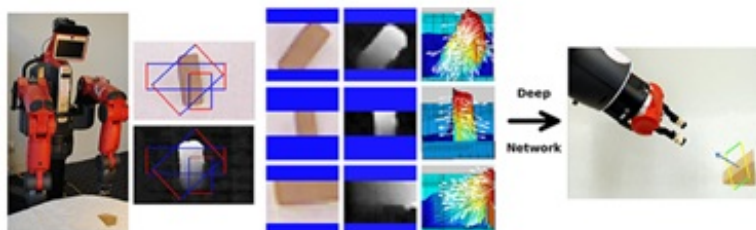
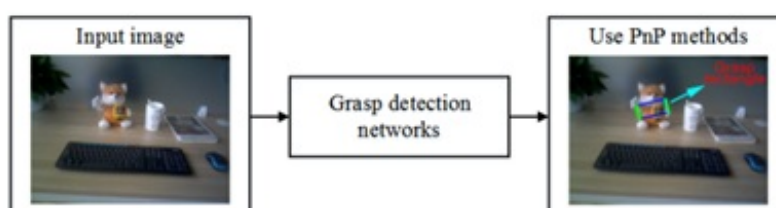
- Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics.

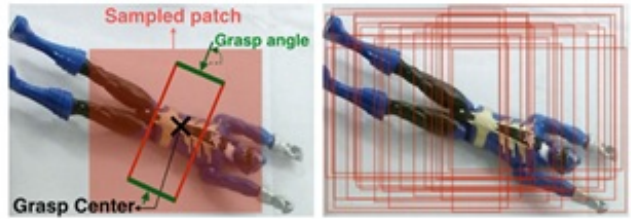


3、无定位无姿态的方法

主要基于深度学习方法，包括：

- Deep learning for detecting robotic grasps.
- Real-time grasp detection using convolutional neural networks.
- Object discovery and grasp detection with a shared convolutional neural network.
- Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours.
- Real-time, highly accurate robotic grasp detection using fully convolutional neural networks with high-resolution images.
- Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching.





运动规划

1、已有抓取点

本节介绍开环方法，其中假设抓取点已通过上述步骤检测到。这些方法设计了从机器人手到目标物体抓取点的路径。这里运动表示是关键问题。虽然存在从机器人手到目标抓取点的无限数量的轨迹，但是由于机器人臂的限制，许多区域无法到达。因此，需要对轨迹进行规划。主要有三种方法，如传统的基于DMP的方法、模仿学习的方法和基于强化学习的方法。

基于DMP的方法：主要包括DMP算法。形式化为稳定的非线性吸引子系统。

基于模仿学习的方法：Generalization of human grasping for multi-fingered robot hands.

基于强化学习的方式：Task-agnostic self-modeling machines.

2、无抓取点

主要通过强化学习直接完成对原始RGB-D图像的抓取任务。

主要有：

- Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection.
- Qt-opt: Scalable deep reinforcement learning for vision-based robotic manipulation.
- Robustness via retrying: Closed-loop robotic manipulation with selfsupervised learning.

参考：<https://arxiv.org/abs/1905.06658>