基于视觉的机器人抓取

-从物体定位、位姿估计到抓取位姿估计

Ref: Vision-based robotic grasping from **object localization**, **object pose estimation** to **grasp estimation** for parallel grippers: a review

Co-Author: Kai Wang, Shiguo Lian, Kaiyong Zhao

arXiv: https://arxiv.org/abs/1905.06658

杜国光

2020年11月6日

https://georgedu.github.io/

主要内容

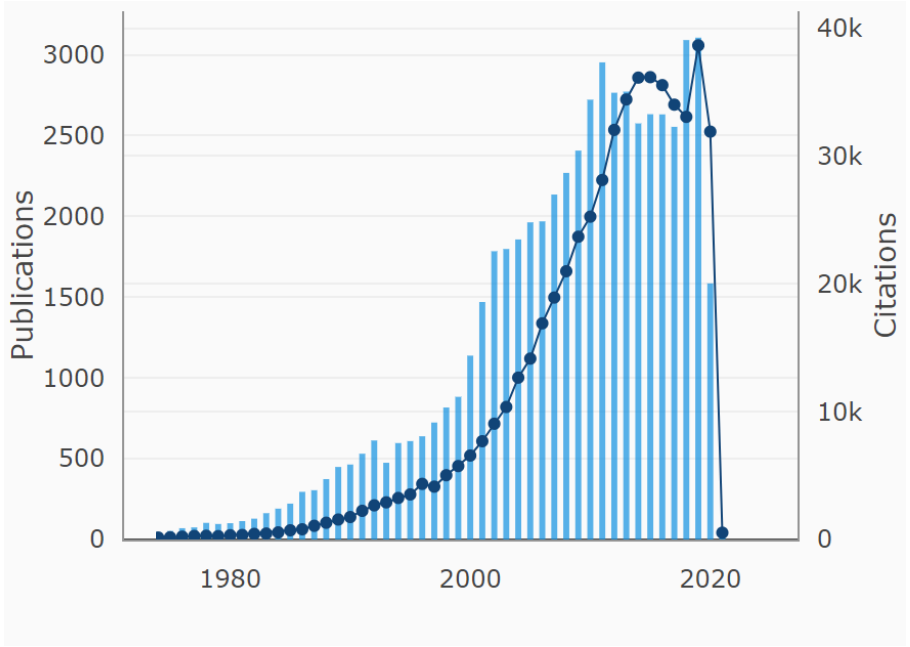
1. 基于视觉的机器人抓取流程与关键技术(Pipeline and Key Technologies)
2. 物体定位(Object localization)
3. 物体位姿估计(Object pose estimation)
4. 抓取位姿估计(Grasp pose estimation)
5. 挑战和未来的研究方向(Challenges and Future Directions)

注意：

- 总结不一定全面
- 介绍方法思想，略过算法细节

机器人抓取-学术界

- 部分高校
  - 国外：
    - MIT/CMU/Stanford/UC Berkeley/Columbia/Princeton/UW/Yale/MPI/UPenn
  - 国内：
    - THU/SJTU/CAS/BUAA/NEU
- 部分期刊
  - Robotics:
    - IJRR/RAS/TRO/AURO/RAM
  - CV/Graphics
    - IJCV/TPAMI/TOG/TVCG
- 部分国际会议
  - Robotics:
    - ICRA/IROS/CHI/HUMANOIDS/ROBIO
  - ML/CV/Graphics
    - AAAI/NeurIPS/ICML/CVPR/ECCV/ICCV/SIGGRAPH



Grasp相关论文的年发表量与引用量

Source: https://academic.microsoft.com/topic/171268870/journals?pi=1

机器人抓取-工业界

- 工业应用
  - 工件抓取
  - 物品分拣
- 服务应用
  - 端茶倒水
  - 捡垃圾

Image credit: covariant.ai

Image credit: cloudminds

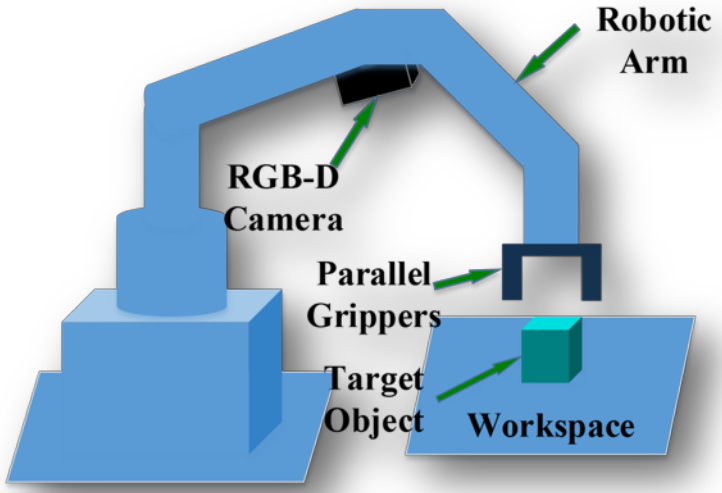
- 一些公司：
  - 国外机械臂抓取：Google X，NVIDIA，Covariant，Vicarious，Osaro等
  - 国内机械臂抓取：库柏特，梅卡曼德，星猿哲等
  - 人形抓取：优必选、达闼科技等

Part Ⅰ

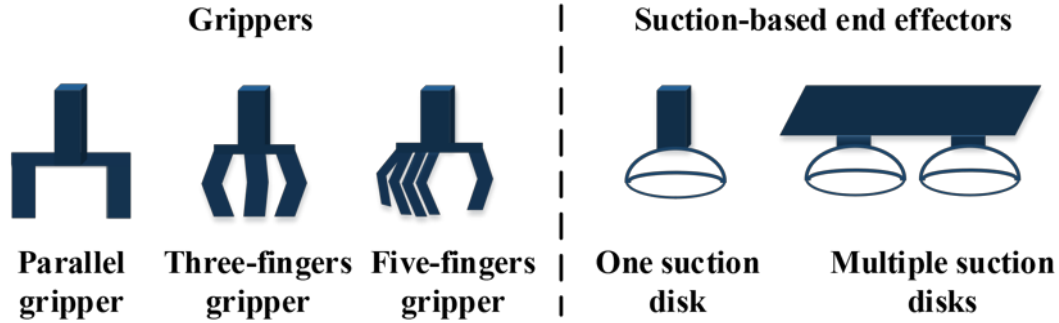
基于视觉的机器人抓取流程与关键技术

主要模块

Grasp Detection System → Path Planning System → Control System



抓取场景



末端抓取器的类型

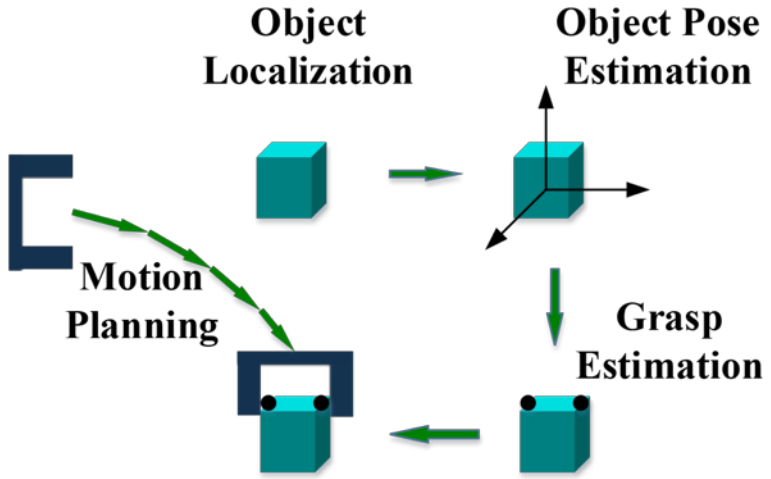
RGB-D 数据

抓取检测系统的关键技术

- 物体在哪里？- Object Localization
  - 识别出哪一个是目标物体
  - 确定目标物体在2D/3D空间的占有区域

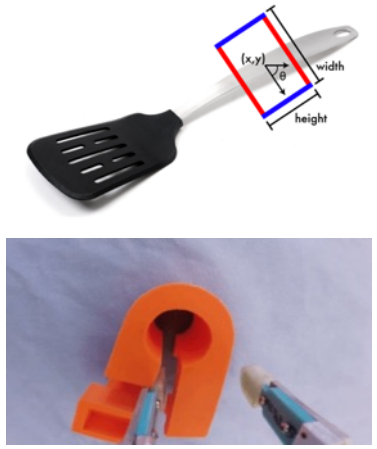
- 物体什么姿态？- Object Pose Estimation
  - 确定完整物体的3D重心和3D朝向

- 抓手怎么去抓？- Grasp Estimation
  - 抓手的最终3D位置和3D朝向
  - 存在2D平面抓取和6DoF抓取



抓取类型

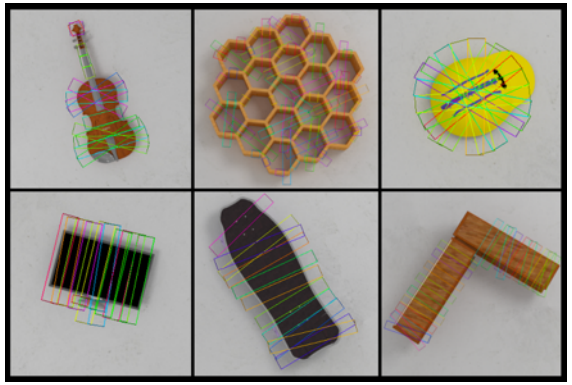
- 2D planar grasp:
  - 2D position
  - 1D angle
  - 数据集
    - Cornell dataset
    - JACQUARD



2D planar grasp

Cornell DATASET

http://pr.cs.cornell.edu/grasping/rect_data/data.php

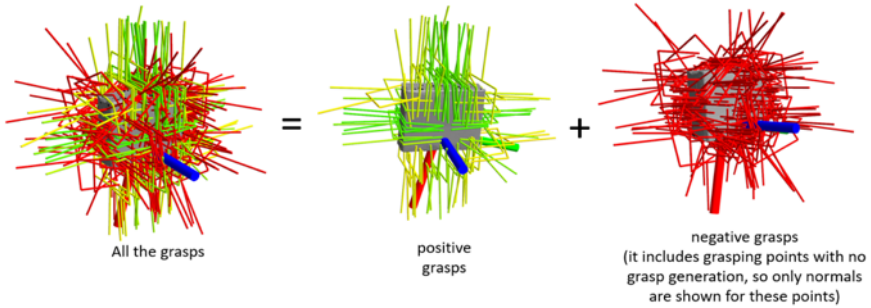


JACQUARD DATASET

https://jacquard.liris.cnrs.fr/

- 6DoF Grasp
  - 3D translation
  - 3D orientation
  - 数据集
    - YCB Video dataset
    - PointNet++ Grasping
    - EGAD!
    - GraspNet-1Billion

6-DoF GraspNet



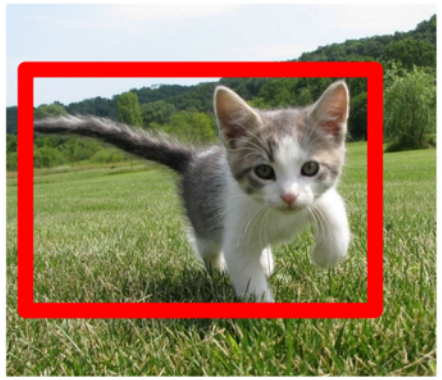
PointNet++ Grasping Dataset

Part II

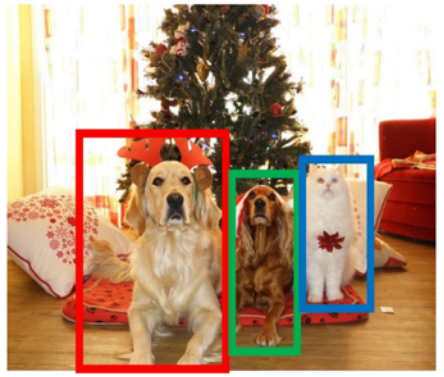
物体定位

物体定位的划分

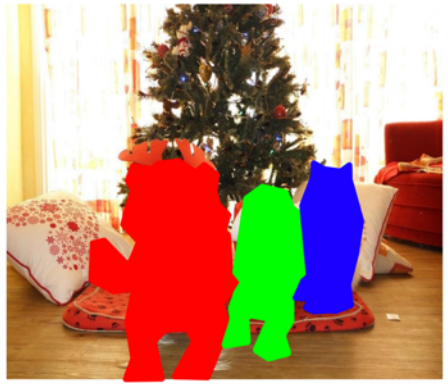
- 只定位不识别 - Object localization without classification
  - 获取目标物体的区域，但不知道目标物体的类别
- 目标检测 - Object Detection
  - 获取目标物体的区域，同时识别出物体的类别
- 目标实例分割 - Object Instance Segmentation
  - 识别物体的同时，获得像素级/点云级物体区域



只定位不识别



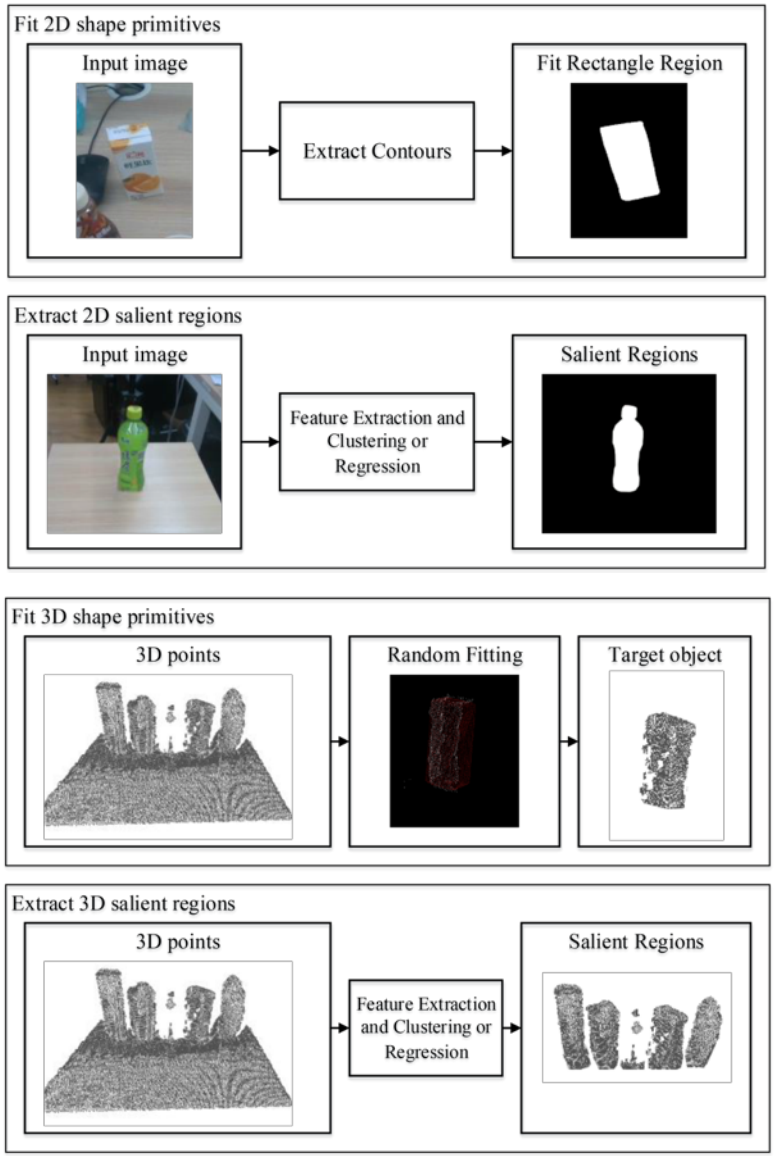
目标检测



目标实例分割

只定位不识别

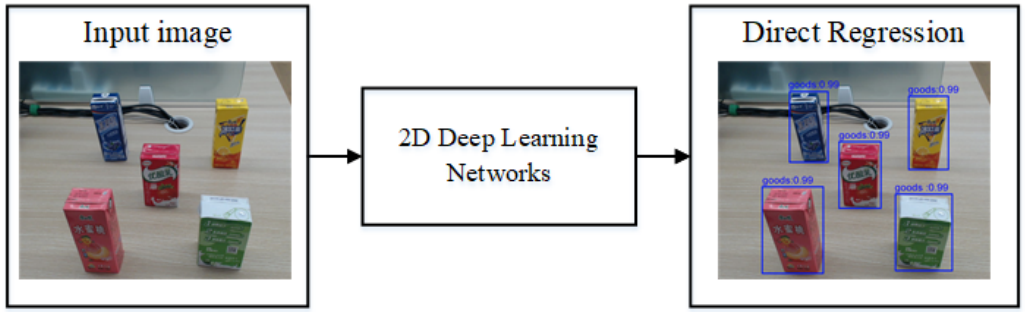
- 拟合形状基元 - Fitting Shape Primitives
  - 已知物体的几何轮廓
    - 2D：椭圆/圆形/矩形等
    - 3D：平面/球体/圆柱/圆环等
  - RANSAC/Hough Voting, etc
- 显著性物体检测 - Salient Object Detection
  - 未知物体（分割背景，提取视觉上显著性区域）
  - 基于传统方法去除背景
  - 基于DL网络提取
- 总结：
  - 处在较浅阶段，限定场景
  - Fitting Primitives
    - 只能拟合特定的几何体
    - 易受噪声影响
  - Salient Object Detection
    - 2D：基于纯色背景，绿幕布等
    - 3D：拟合桌面3D平面，或者去除已有背景的3D模型



目标检测

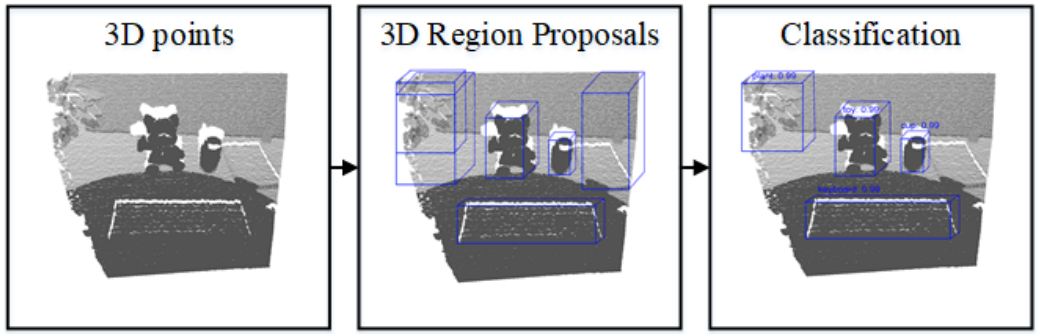
- 2D图像：两阶段法（Region proposal-based method）
  - 传统方法：Sliding window + Local feature + Classifier
  - 基于DL方法：R-CNN，Fast R-CNN，Faster R-CNN

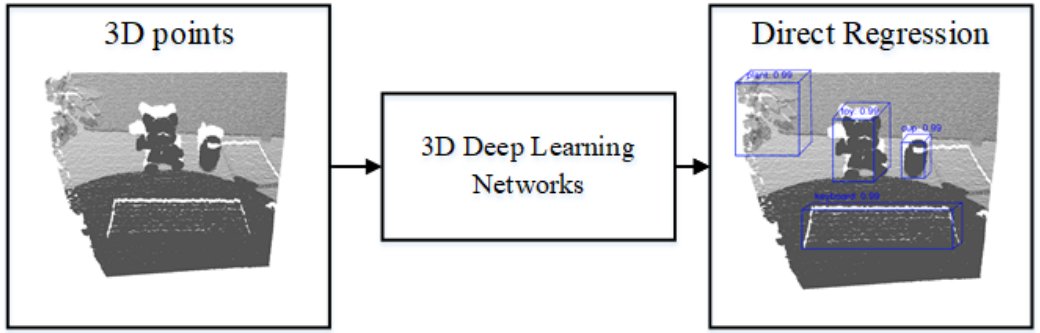
- 2D图像：单阶段法（Regression-based method）
  - 跳过候选区域生成，直接在一次估计中预测包围盒以及类别得分
  - 典型方法：
    - SSD，YOLOv1-v4
    - FCOS、CornerNet、CenterNet等
- 总结：
  - 广泛使用：
    - 先检测到候选物体，再选择要抓取的物体
    - 网络可应用于2D平面抓取估计



目标检测

- 3D检测：
  - 目标物体的完整(Amodel)3D包围盒，物体类别+3D位置+1D角度+3D尺寸
  - 通常指泛化的一类物体3D检测
- 3D点云：两阶段法
  - 传统方法：滑动窗口+局部特征+分类器
  - 基于DL方法：
    - 基于截锥体(frustum-based)
      - Frustum PointNets, FrustumConvNet等
    - 基于全局回归(global regression-based)
      - Deep Sliding Shapes, MV3D, MMF, PartA^{2}等
    - 基于局部回归(local regression-based)
      - PointRCNN，VoteNet，IMVoteNet等
- 3D点云：单阶段法
    - 通过单个网络直接预测3D包围盒及其类别
      - VoxelNet, SECOND, PointPillars, 3DSSD等
- 总结：
  - 能给出目标物体的大致位置，可用于抓取避障



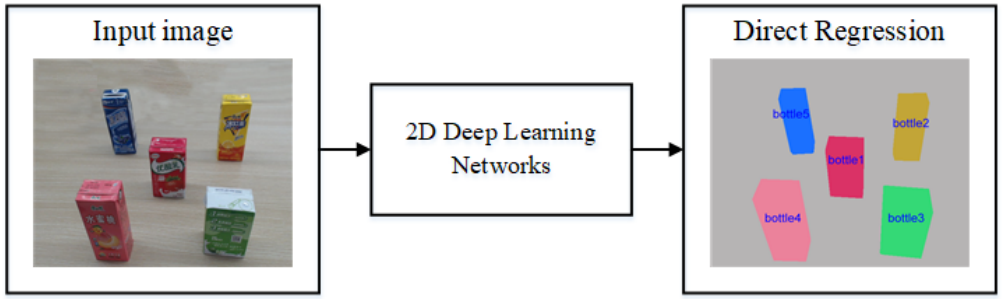


目标实例分割

- 2D图像：两阶段法
  - 成熟的2D检测器常被用来生成包围盒或者区域候选
  - 在他们内部可以进一步计算物体的mask区域
  - 代表性方法：
    - SDS, PANet, HTC, Mask R-CNN等

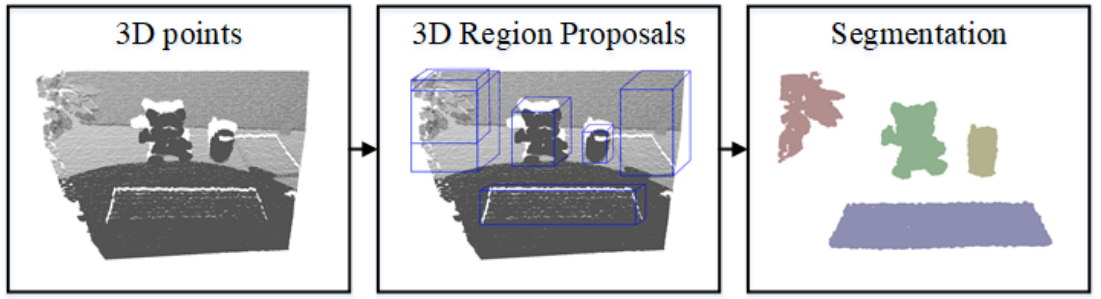
- 2D图像：单阶段法
  - 同时预测分割的mask和存在物体的得分
  - 代表性方法：
    - DeepMask, SharpMask, FCIS, TensorMask, YOLACT++，SOLO，CenterMask等

- 总结：
  - 在机器人抓取应用中被广泛使用，如果场景中同一个类别的物体只有一个实例，语义分割也可以。
  - 由于输入是RGB-D图像，有了RGB分割结果，可以快速得到对应Depth图像，得到目标物体的3D点云。

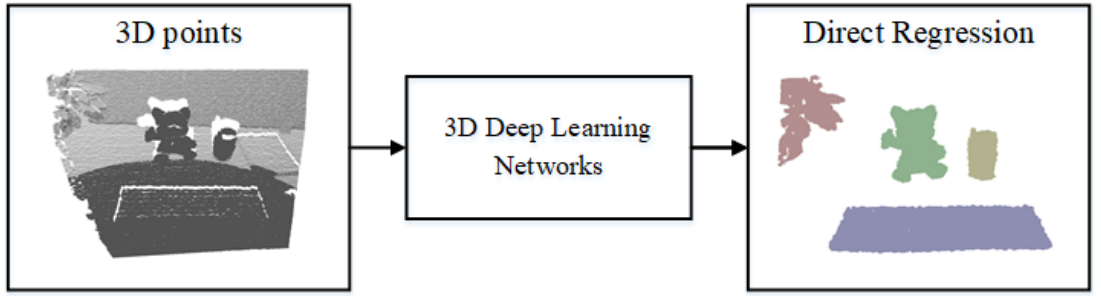


目标实例分割

- 3D点云：两阶段法
  - 一般的方法借助2D/3D检测结果，再对对应3D截锥体或者包围盒区域进行前后景分割得到目标物体的点云。
  - 代表方法：
    - GSPN，3D-SIS，OccuSeg



- 3D点云：单阶段法
  - 学习如何归类逐点的特征来完成3D实例分割
  - 代表方法：
    - SGPN, MASC, ASIS, JSIS3D，3D-BoNet等
- 应用：
  - 非常重要
  - 前景广阔



物体定位总结

- 现状：
  - 给定机器人抓取场景，目前总能找到合适的技术方案来定位目标物体的位置，进而执行后续的物体位姿估计以及机器人抓取位姿估计等任务；
  - 不过针对限定场景下特定的一些物体，当前算法已经能够得到非常好的满足落地需求的结果；

- 挑战
  - 定位但不识别的算法，要求物体在结构化的场景中或者物体与背景具有显著性差异，这些都有益于算法提取出目标物体，这就限制了应用场景；
  - 实例级目标检测算法，需要实例级目标物体的大量训练集，而且训练好的检测算法只在训练集上检测精度高；如果需要对新物体进行检测，需要再采集大量的数据进行重新训练，这个过程非常耗时耗力；
  - 通用的目标检测算法，泛化能力强，但其精度达不到实例级目标检测算法。
  - 实例分割算法也面临同样的问题，对大量训练数据集的需求以及泛化性能差，也是深度学习算法的通用性问题。

Part III

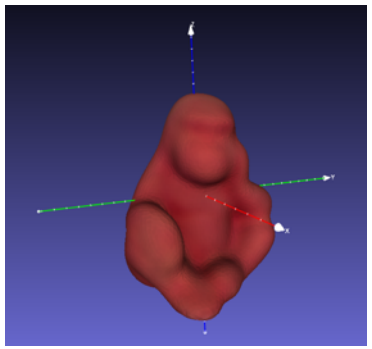
物体位姿估计

物体6D位姿的含义

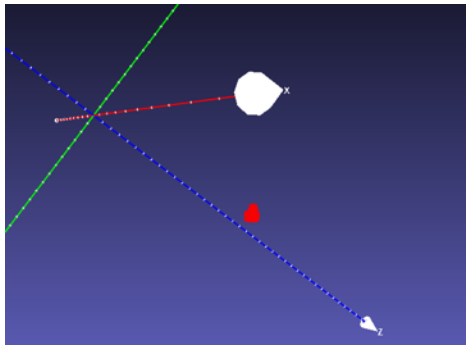
- 物体6D位姿：
  - 原始物体由所在世界系到相机系的RT变换

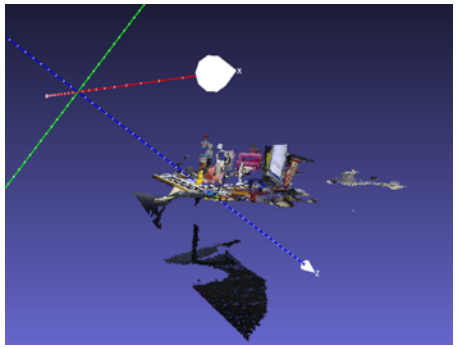
$$T_c = R_{cm} * T_m + t_{cm}$$

  - 其中：
    - T_m代表物体在世界系下的3D点，T_c代表相机系下物体的3D点
    - R_{cm}代表原始物体由所在世界系到相机系的旋转，t_{cm}代表由物体所在的世界系到相机系的平移



$[R_{cm}, t_{cm}]$





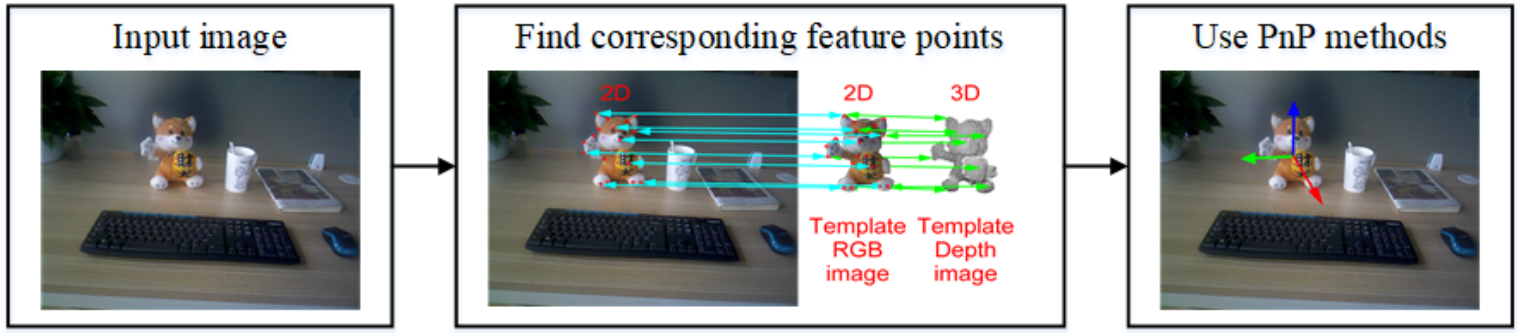
方法划分

- 根据利用特征点/利用整体/利用所有点来划分

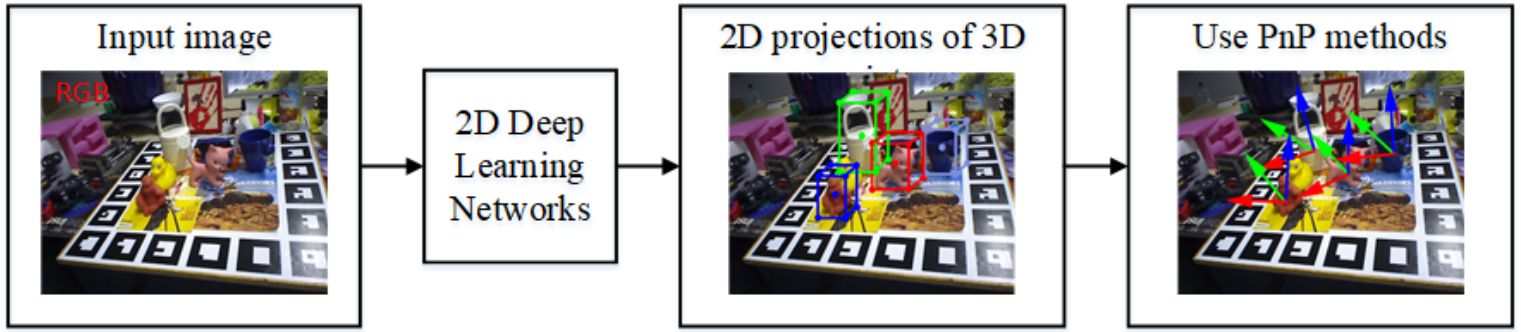
刚体6D位姿估计方法综述					
方法	简要描述	2D图像输入		3D点云输入	
基于对应点的方法（Correspondences-based method）	寻找2D-3D或3D-3D特征点的对应，包括显式的和隐式的方式	传统方法	基于DL的方法	传统方法	基于DL的方法
		SIFT，SURF，ORB等	BB8，Yolo-6D，[2018-Robust 3d]，[2018-Segmentation-driven]，DPOD，[2019-Single-stage 6d]等	Spin Images，3D Shape Context，FPFH，SHOT等	3DMatch，3DFeat-Net，StickyPillars等
基于模板的方法（Template-based method）	寻找当前输入和已有带6D位姿的模板之间的对应，包括显式的和隐式的方式	传统方法	基于DL的方法	传统方法	基于DL的方法
		LineMode，[2015-Detection and fine]等	PoseCNN，SSD6D，Deep-6DPose，[2019-CDPN]，NOCS，Latentfusion，[2020-Learning canonical]	Super 4PCS，FGR GO-ICP等	PCRNet，PointNetLK，TEASER，[2020-6d object pose]等
基于投票的方法（Voting-based method）	每个像素点或者3D点间接投票得到关键点或者直接投票得到6D位姿	间接方法	直接方法	间接方法	直接方法
		PVNet等	[2014-Learning 6d object pose]，[2014-Latent-class]等	PVN3D，6-PACK，YOLOff等	PPF，DenseFusion，[2020-Lrf-net]，[2020-6dof object pose]等

基于对应的方法-2D

- 2D-3D+PnP / Indirectly regress 2D projections of 3D points + PnP
- 显式的寻找当前图像与模板图像之间的特征点对应
  - 传统局部特征描述符如SIFT、SURF、ORB等
  - 基于深度学习的特征描述符：LIFT, GLAMpoints, LCD等

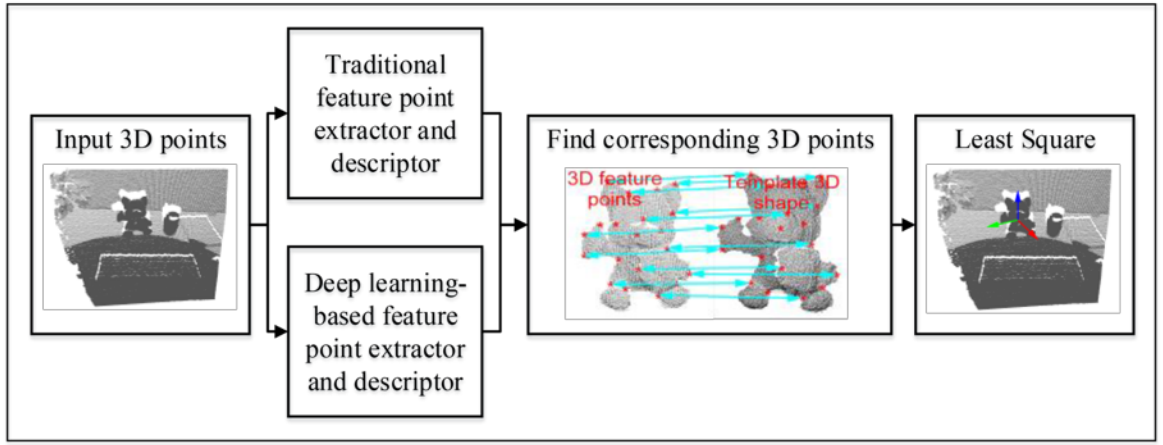


- 隐式地回归3D点在2D图像上的投影点 (外界包围盒的8个顶点，特征点或者所有点)
  - 代表方法：BB8, Yolo6D, Segmentation-driven, DPOD, EPOS等



基于对应的方法-3D

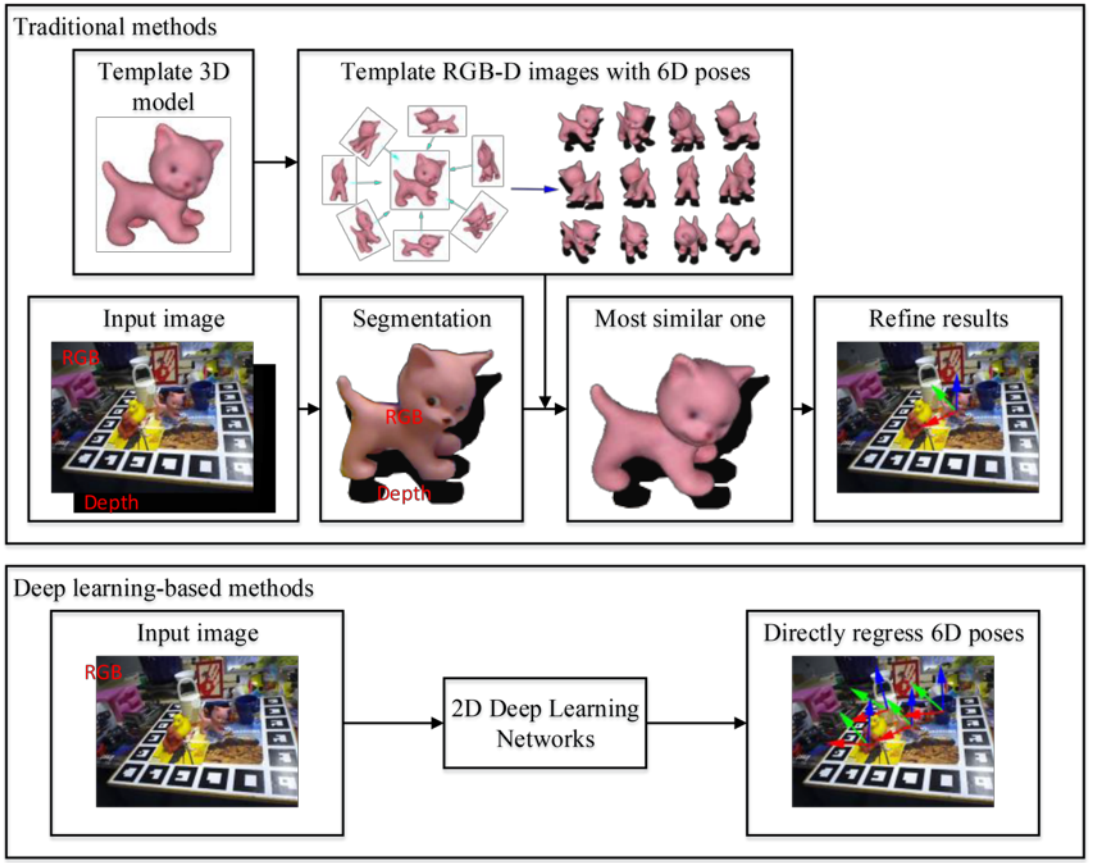
- 3D-3D + ICP / Regress 3D feature points + ICP/LeastSquares
- 提取3D descriptors，寻找3D点对应
  - 传统3D局部特征描述符，如FPFH、SHOT等
  - 基于DL的3D局部特征描述符，3DMatch, 3DFeat-Net, LCD, D3Feat等
  - 基于DL提取特征并匹配，RPM-Net, StickyPillars等



  - 基于3D DNN网络预测3D点与3D点的对应
    - 代表方法：PVN3D, 6-PACK

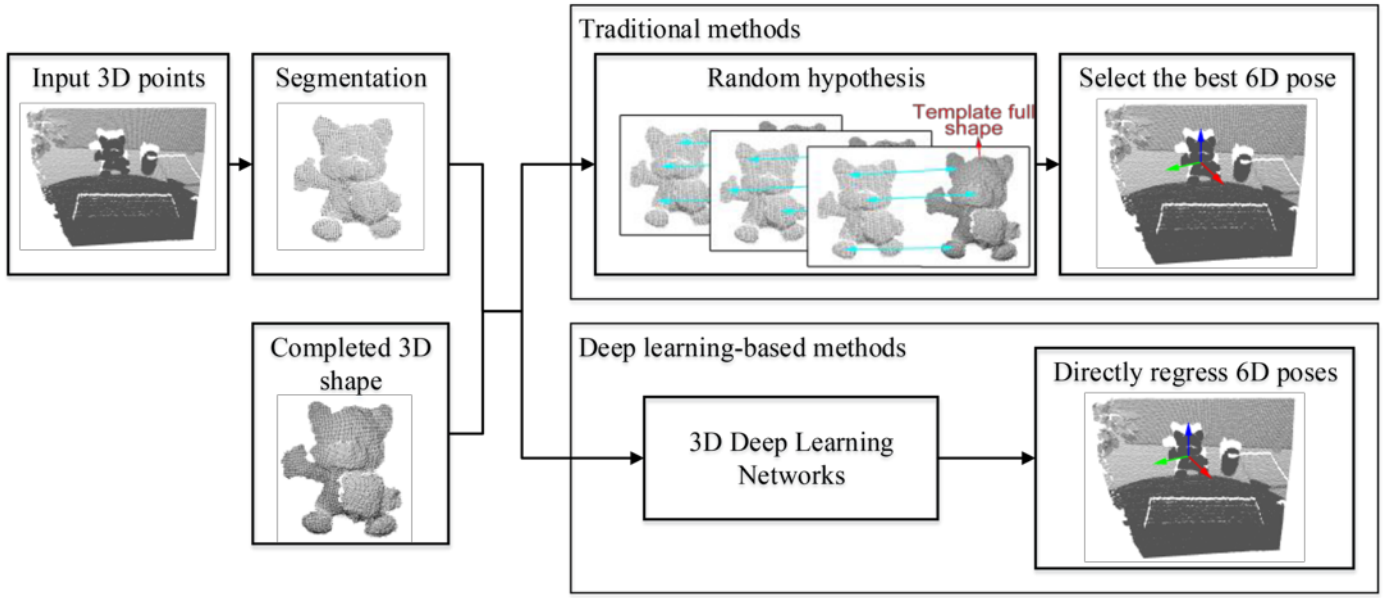
基于模板的方法-2D

- 2D image retrieval / Directly regress 6D pose
- 显式地（Explicitly）传统图像相似度比较
  - 特征丰富：使用2D特征描述符
  - 特征不丰富：使用颜色梯度(可结合深度图法向量)：2012-LineMod等
- 隐式地（Implicitly）基于深度学习的image retrieval
  - AAE[2018-Implicit 3D Orientation Learning]
- 隐式地（Implicitly）基于深度学习直接得到6D pose
  - PoseCNN，Deep-6DPose，CDPN, CosyPose等
- 隐式地（Implicitly）构建物体的潜在表示（类别级）
  - NOCS, LatentFusion，2020-Learning canonical等



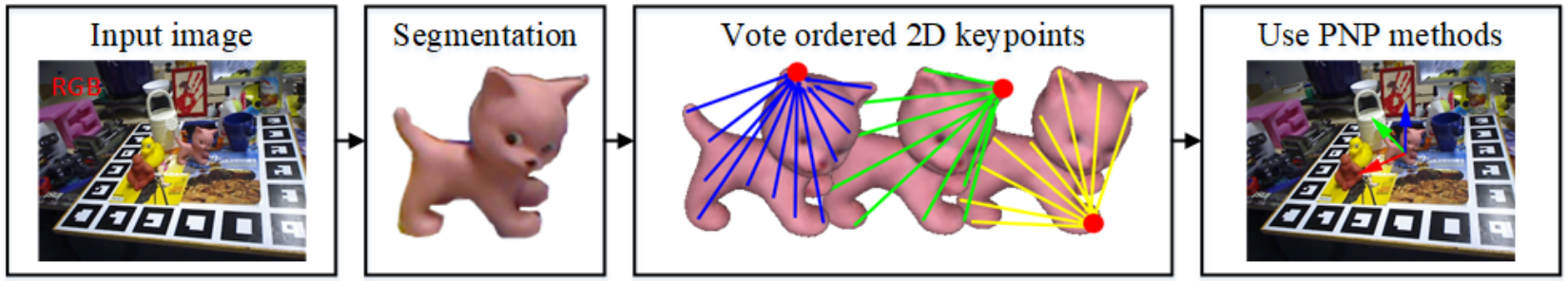
基于模板的方法-3D

- 3D partial registration / Directly regress transformations or 6D pose
- 传统partial registration方法，显式地寻找与full point cloud的变换
  - Super 4PCS，GO-ICP, FGR等
- 基于Deep learning，显式地寻找与full point cloud的变换
  - DCP，PCRNet, PointNetLK, DeepGMR, MaskNet等
- 基于Deep learning隐式地直接回归6D pose
  - [2020-6d object pose regression via supervised learning on point cloud]

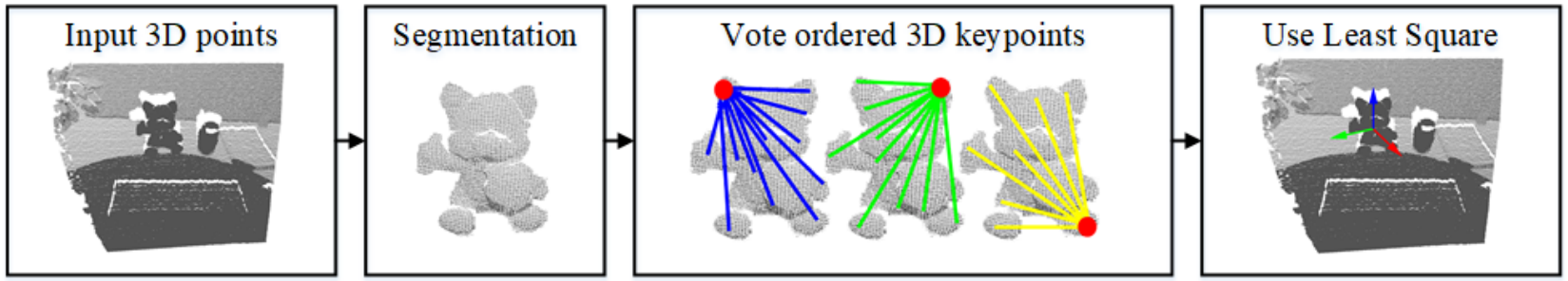


基于投票的方法

- 间接投票的方法
  - 每个2D像素或者3D点投票得到预定义的特征点，能够得到2D-3D或者3D-3D的对应
- 2D图像：每一个像素投票产生2D关键点位置+PnP
  - PVNet

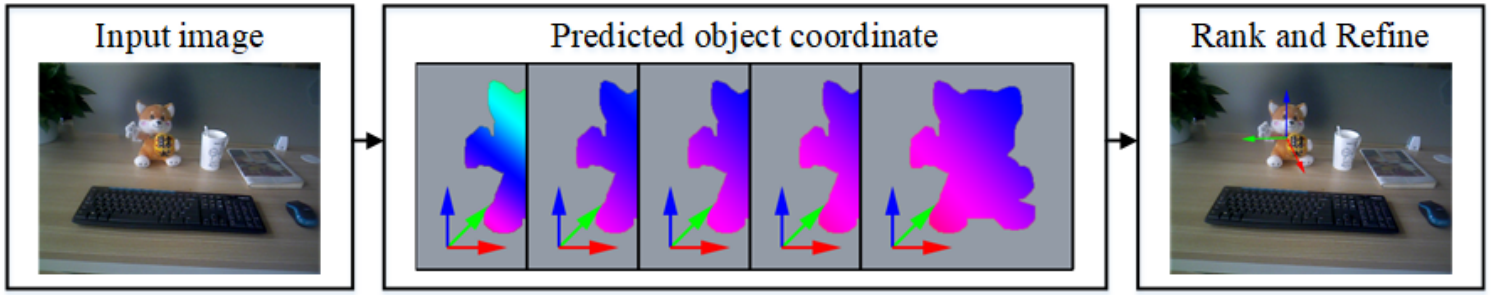


- 3D点：每一个点都预测一个3D KeyPoint + LeastSquares
  - PVN3D, 6-PACK, YOLOff等

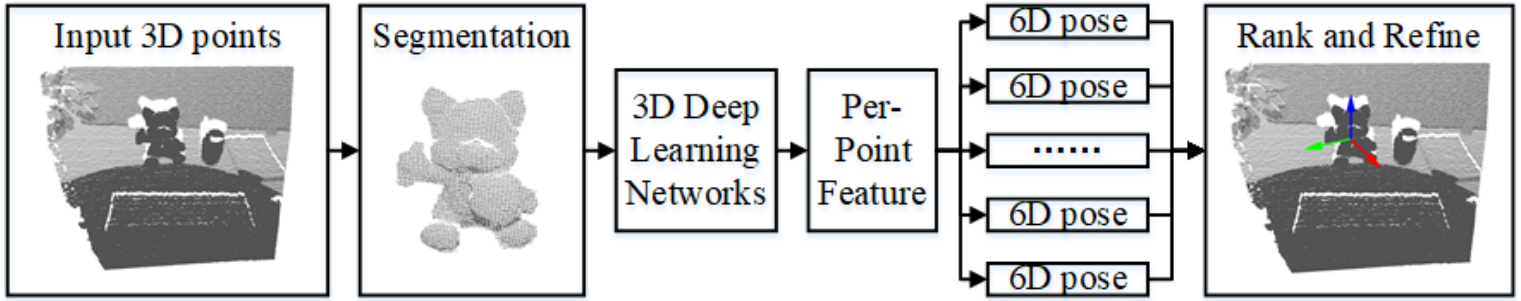


基于投票的方法

- 直接投票的方法
  - 每个像素或3D点直接投票得到一个确定的6D物体坐标系或者6D位姿
- 2D图像：每一个像素产生一个坐标架预测
    - 代表方法：2014-Learning 6D Object Pose Estimation using 3D Object Coordinates



- 3D点：每个点对应一个6D位姿
    - 代表方法：PPF，DenseFusion，[2020-Lrf-net]，[2020-6dof object pose]，MoreFusion等



物体位姿估计-方法对比

- 数据集：
  - LineMOD, Occlusion LineMOD, YCB-Video
- 度量：
  - ADD，ADD-S， AUC

$$e_{ADD} = \underset{x \in M}{avg} \left\| (Rx + T) - (\hat{R}x + \hat{T}) \right\|$$

Table 7: Accuracies of AUC and ADD-S metrics on YCB-video dataset.

Category	Method	AUC	ADD-S (<2cm)
Corre-based	Heatmaps [Oberweger *et al.*, 2018]	72.8	53.1
Template-based	PoseCNN [Xiang *et al.*, 2018]+ICP	61.0	73.8
	PoseCNN [Xiang *et al.*, 2018]+ICP	93.0	93.2
	Castro et al. [Castro *et al.*, 2020]	67.52	47.09
	PointFusion [Xu *et al.*, 2018]	83.9	74.1
	MaskedFusion [Pereira and Alexandre, 2019]	93.3	97.1
Voting-based	DenseFusion [Wang *et al.*, 2019b](per-pixel)	91.2	95.3
	DenseFusion [Wang *et al.*, 2019b](iterative)	93.1	96.8

Table 8: Accuracies of methods using ADD(-S) metric on LineMOD and Occlusion LineMOD dataset. Refine means methods such as ICP or DeepIM. IR is short for iterative refinement.

Category	Method	LineMOD	Occlusion
Correspondence-based methods	BB8 [Rad and Lepetit, 2017]	43.6	-
	BB8 [Rad and Lepetit, 2017]+Refine	62.7	-
	Tekin et al. [Tekin *et al.*, 2018]	55.95	6.42
	Heatmaps [Oberweger *et al.*, 2018]	-	25.8
	Heatmaps [Oberweger *et al.*, 2018]+Refine	-	30.4
	Hu et al. [Hu *et al.*, 2019]	-	26.1
	Pix2pose [Park *et al.*, 2019b]	72.4	32.0
	DPOD [Zakharov *et al.*, 2019]	82.98	32.79
	DPOD [Zakharov *et al.*, 2019]+Refine	95.15	47.25
	HybridPose [Song *et al.*, 2020]	94.5	79.2
Template-based methods	SSD-6D [Kehl *et al.*, 2017]	2.42	-
	SSD-6D [Kehl *et al.*, 2017]+Refine	76.7	27.5
	AAE [Sundermeyer *et al.*, 2018]	31.41	-
	AAE [Sundermeyer *et al.*, 2018]+Refine	64.7	-
	Castro et al. [Castro *et al.*, 2020]	59.32	-
	PoseCNN [Xiang *et al.*, 2018]	62.7	6.42
	PoseCNN [Xiang *et al.*, 2018]+Refine	88.6	78.0
	CDPN [Li *et al.*, 2019]	89.86	-
	Tian et al. [Tian *et al.*, 2020]	92.87	-
	MaskedFusion [Pereira and Alexandre, 2019]	97.3	-
Voting-based methods	Brachmann et al. [Brachmann *et al.*, 2016]	32.3	-
	Brachmann et al. [Brachmann *et al.*, 2016]+Refine	50.2	-
	PVNet [Peng *et al.*, 2019]	86.27	40.8
	DenseFusion [Wang *et al.*, 2019b](per-pixel)	86.2	
	DenseFusion [Wang *et al.*, 2019b](iterative)	94.3	
	DPVL [Yu *et al.*, 2020]	91.5	43.52
	YOLOff [Gonzalez *et al.*, 2020]	94.2	-
	YOLOff [Gonzalez *et al.*, 2020]+Refine	98.1	-
	PVN3D [He *et al.*, 2020]	95.1	-
	P^2GNet [Yu *et al.*, 2019b]	96.2	-
	P^2GNet [Yu *et al.*, 2019b]+Refine	97.4	-
	PointPoseNet [Hagelskjær and Buch, 2019]	96.3	52.6
	PointPoseNet [Hagelskjær and Buch, 2019]+Refine	-	75.1

物体位姿估计-总结

- 当前针对实例物体已能够取得较高的精度

- 不同侧重场景：
  - 物体纹理丰富 - 基于对应的方法
  - 物体纹理不丰富 - 基于模板的方法
  - 存在遮挡 - 基于投票的方法

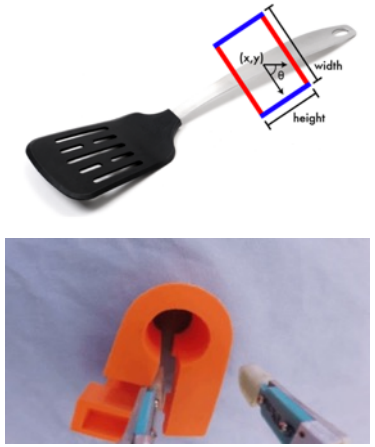
- 存在挑战：
  - 遮挡情况
  - 缺少训练数据数据集：
    - AprilTag, LabelFusion
    - BOP Toolkit
  - 实例级向类别级 (Category-Level) 扩展
    - NOCS，LatentFusion等

Part IV

抓取位姿估计

Robotic Grasping

- 2D planar estimation
  - 评估抓取接触点
  - 评估带朝向抓取四边形



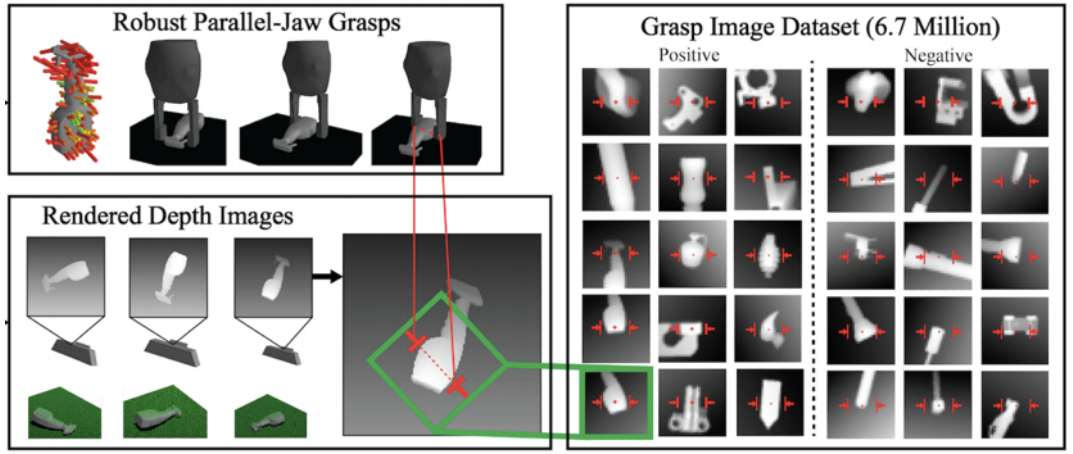
- 6DoF grasp estimation
  - 基于完整3D形状
    - 基于物体6D位姿，直接计算抓取姿态
    - 对物体进行补全，再估计抓取姿态
  - 基于单视角点云
    - 先生成候选抓取点，再评估抓取质量
    - 基于深度学习直接估计6D抓取位姿
    - 寻找点与点对应，迁移抓取点

2D planar grasp

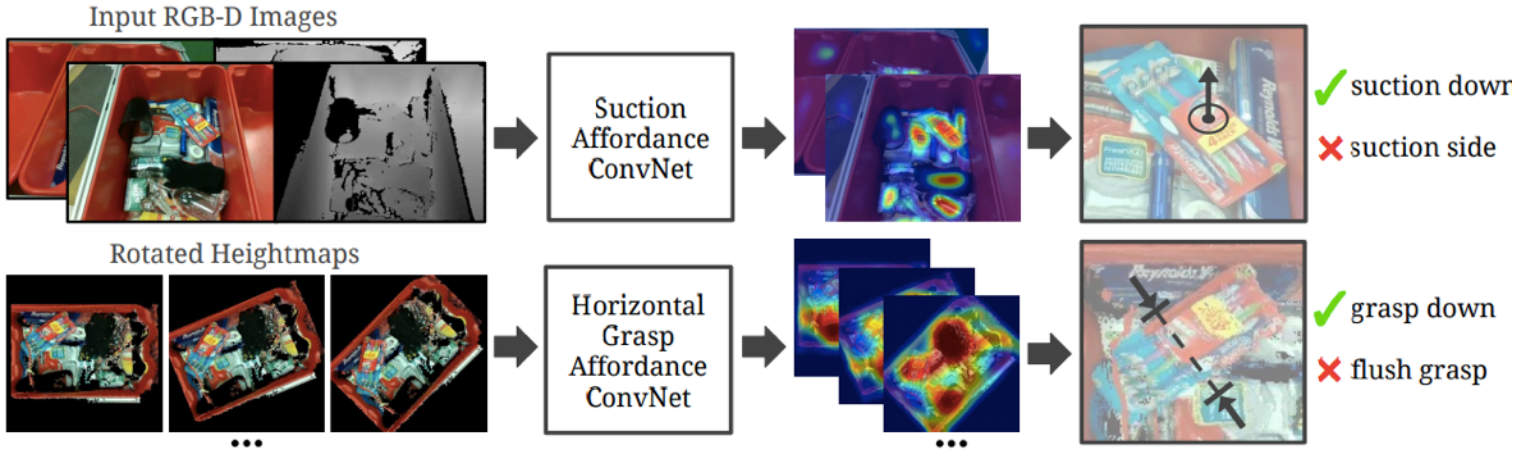
6-DoF GraspNet

2D平面抓取

- Methods of Evaluating Grasp Contact Points
  - 传统方法：
    - 2014-Fast graspability evaluation on single depth maps for bin picking with general grippers
  - 基于DL方法：
    - 评估候选质量： DextNet 2.0, GG-CNN等
    - 预测Pix-wise Affordance:
      - 2018-Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching



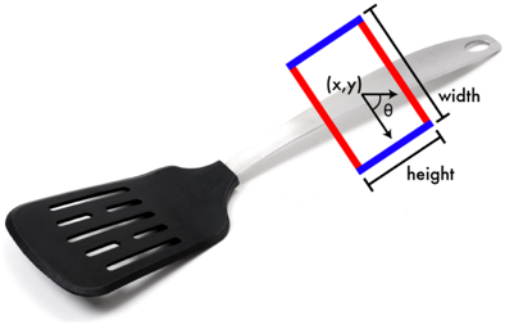
DexNet2.0



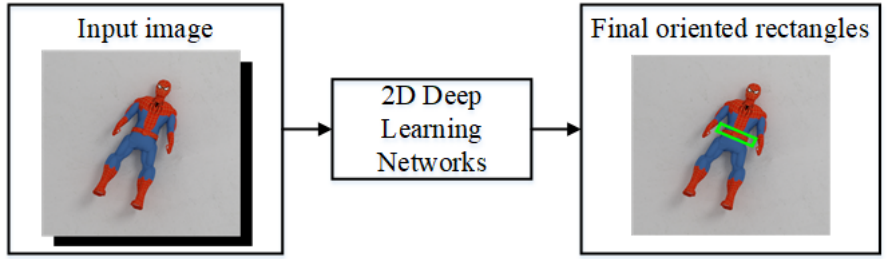
2018-Robotic pick-and-place

2D平面抓取

- Methods of evaluating oriented rectangles
  - A 5D representation for robotic grasps: (x, y, θ, h, w)
- Classification-based methods
  - Train classifiers to evaluate candidate grasps
  - 代表方法：
    - 2015-Deep learning for detecting robotic grasps
    - 2016-Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours
- Regression-based methods
  - 代表方法：
    - 2015-Real-time grasp detection using convolutional neural networks
    - 2017-Robotic grasp detection using deep convolutional neural networks
    - 2019-Antipodal robotic grasping using generative residual convolutional neural network
- Detection-based methods
  - 代表方法：
    - 2017-A hybrid deep architecture for robotic grasp detection
    - 2018-Real-world multiobject, multigrasp detection
    - 2018-Fully convolutional grasp detection network with oriented anchor box
    - 2020-Optimizing correlated graspability score and grasp regression for better grasp prediction



Grasp representation



Data from the JACQUARD dataset

2D平面抓取-总结

- Dataset
  - Cornell Grasping
    - Image-wise splitting
    - Object-wise splitting
  - Metrics
    - The grasp angle is within 30◦ of the ground truth grasp
    - And the Jaccard index J(A, B) = |A ∩ B|/|A∪B| of the predicted grasp A and the ground truth B is greater than 25%

- 总结：
  - 已经能够取得很高的精度
  - 在更复杂的数据库上测试

Table 10: Summaries of publicly available 2D planar grasp datasets.

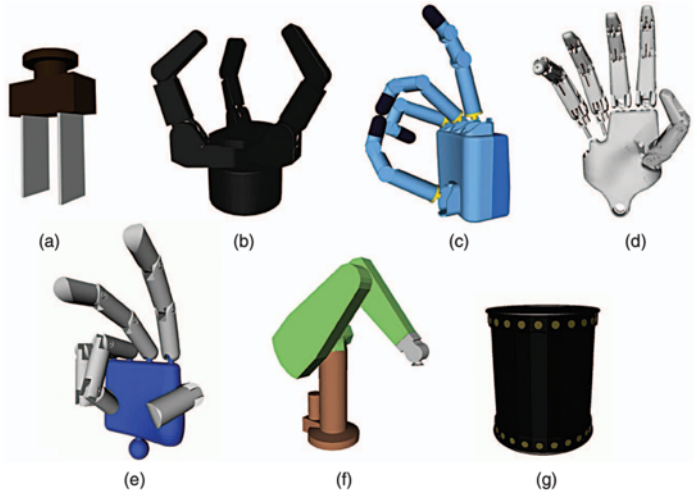
Dataset	Objects Num	Num of RGB-D images	Num of grasps
Stanford Grasping [Saxena *et al.*, 2008b; Saxena *et al.*, 2008a]	10	13747	13747
Cornell Grasping [Jiang *et al.*, 2011]	240	885	8019
CMU dataset [Pinto and Gupta, 2016]	over 150	50567	no
Dex-Net 2.0 [Mahler *et al.*, 2017]	over 150	6.7 M(Depth only)	6.7 M
JACQUARD [Depierre *et al.*, 2018]	11619	54485	1.1 M

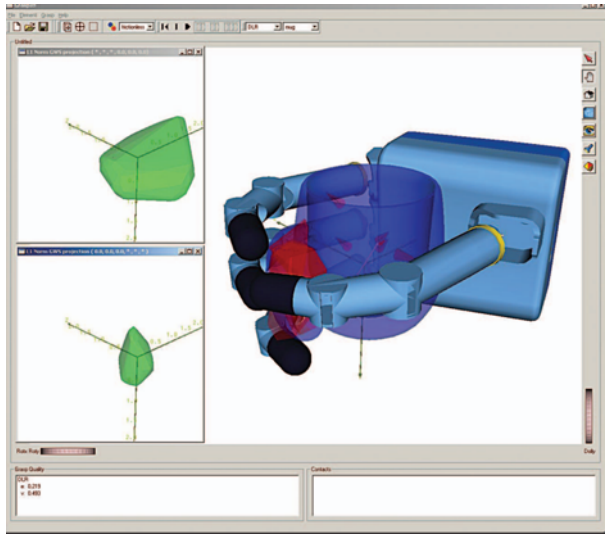
Table 11: Accuracies of grasp prediction on the Cornell Grasp dataset.

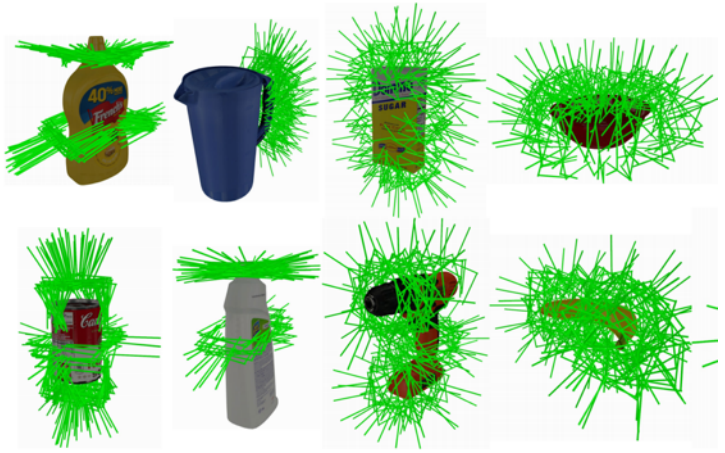
Method	Input Size	Accuracy(%)		Time
		Image Split	Object Split	
Jiang et al. [Jiang *et al.*, 2011]	227 x 227	60.50	58.30	50sec
Lenz et al. [Lenz *et al.*, 2015]	227 x 227	73.90	75.60	13.5sec
Morrison et al. [Morrison *et al.*, 2018]	300 x 300	78.56	-	7ms
Redmon et al. [Redmon and Angelova, 2015]	224 x 224	88.00	87.1	76ms
Zhang et al. [Zhang *et al.*, 2017]	224 x 224	88.90	88.20	117ms
Kumra et al. [Kumra and Kanan, 2017]	224 x 224	89.21	88.96	103ms
Chun et al. [Park and Chun, 2018]	400 x 400	89.60	-	23ms
Asif et al. [Asif *et al.*, 2018]	224 x 224	90.60	90.20	24ms
Wang et al. [Wang *et al.*, 2019d]	400 x 400	94.42	91.02	8ms
Chu et al. [Chu *et al.*, 2018]	227 x 227	96.00	96.10	120ms
Chun et al. [Park *et al.*, 2018]	360 x 360	96.60	95.40	20ms
Zhou et al. [Zhou *et al.*, 2018]	320 x 320	97.74	96.61	118ms
Park et al. [Park *et al.*, 2019a]	360 x 360	98.6	97.2	16ms

6D抓取

- 传统方法如何计算抓取位姿
  - 基于已知CAD模型的完整物体
  - 计算对点（Antipodal points）使满足力封闭（Force Closure）约束
  - 得到预先计算的物体世界系下的抓取6D抓取位姿
- 基于物体6D的位姿，直接获得抓取器在相机系下的抓取位姿
  - 绝大多数6DoF抓取采用这种方式







Examples of 100 precomputed grasps

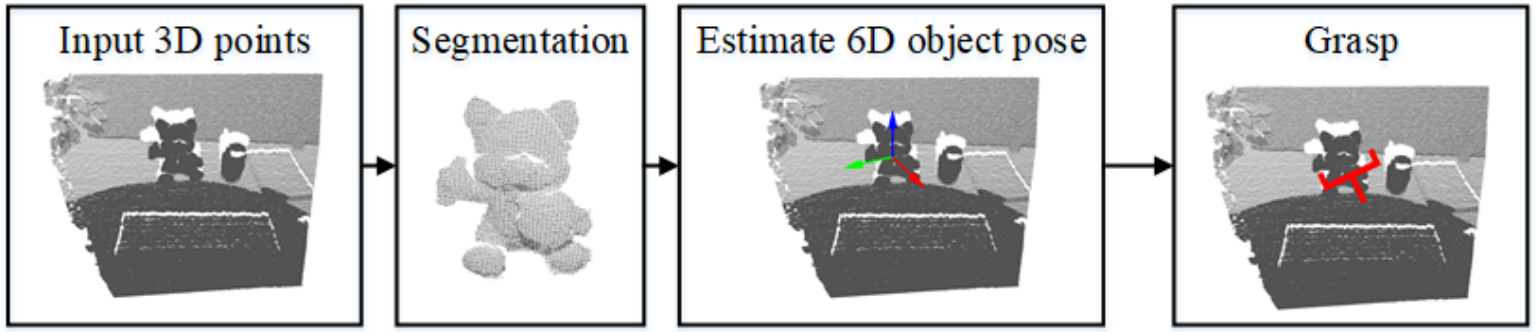
2004-Graspit!: A Versatile Simulator for Robotic Grasping

基于完整形状的抓取

- Methods of estimating the 6D object pose
  - 估计物体6D位姿估计 + 直接获取预先计算的Grasps

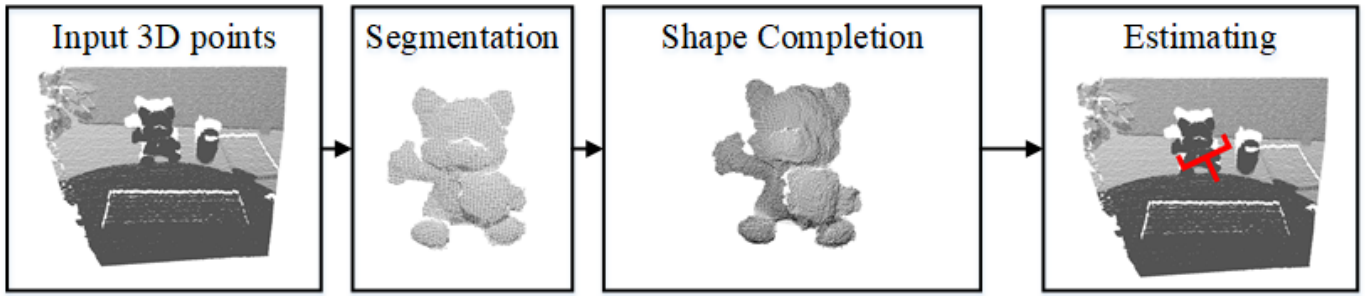
- 代表性论文
  - Amazon Picking Challenge
  - 2017- Multi-view self-supervised deep learning for 6d pose estimation in the amazon picking challenge
  - 2017- Segicp: Integrated deep semantic segmentation and pose estimation

- 优点
  - 精度高
- 缺点
  - 实例级



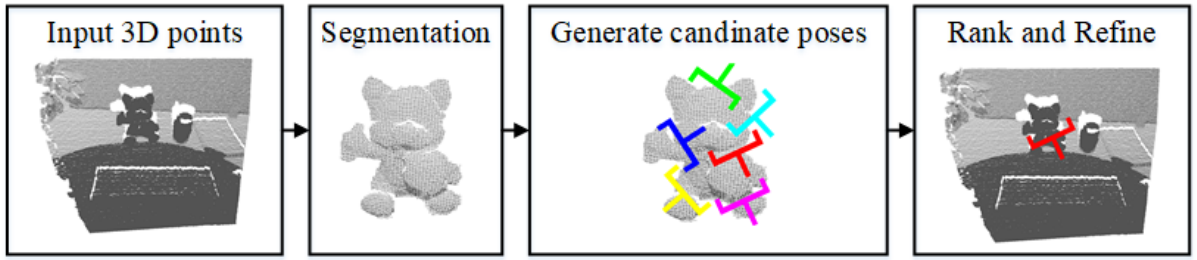
基于完整形状的抓取

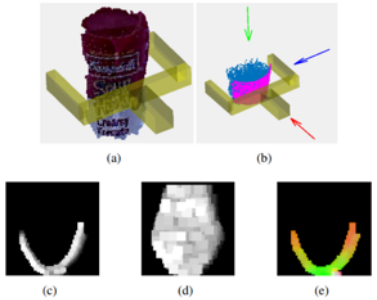
- Methods of conducting shape completion
  - 恢复完整物体的3D网格 + 估计可能的抓取位姿
  - From partial point cloud
    - 2017-Shape completion enabled robotic grasping，3DCNN+Mesh+GraspIt!
    - 2019-Robust grasp planning over uncertain shape completions
    - 2019-Learning continuous 3d reconstructions for geometrically aware grasping
  - From RGB-D images
    - 2018-Learning 6-dof grasping interaction via deep geometry-aware 3d representations
    - 2019-Data-efficient learning for sim-to-real robotic grasping using deep point cloud prediction networks
  - Combine tactile information
    - 2018-3d shape perception from monocular vision, touch, and shape priors
    - 2019-Multi-modal geometric learning for grasping and manipulation
    - 2019-kpamsc: Generalizable manipulation planning using keypoint affordance and shape completion



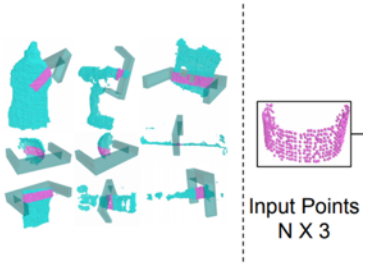
基于部分点云的抓取

- Classification-based method
  - 先生成候选抓取位姿，再评估抓取质量
  - 传统几何方法评估：
    - 2010-Learning grasping points with shape context
    - 2015-Using geometry to detect grasps in 3d point clouds
  - 基于深度学习评估：
    - Multi-view Images:
      - 2017-Grasp pose detection in point clouds（GPD）
    - Voxel grids:
      - 2019-Learning to generate 6-dof grasp poses with reachability awareness
    - Point cloud:
      - 2019-Pointnetgpd: Detecting grasp configurations from point sets
      - 2020-6-DoF Graspnet: Variational grasp generation for object manipulation

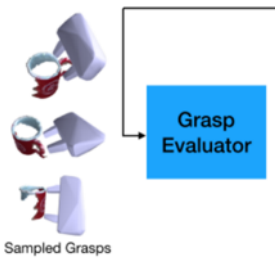




GPD



PointNetGPD



6-DoF GraspNet

基于部分点云的抓取

- Regression-based method
  - 基于深度学习直接估计6D抓取位姿
  - 代表方法：
    - 2020-S^{4}G: Amodal single-view single-shot se (3) grasp detection in cluttered scenes
    - 2020-REGNet: Region-based grasp network for single-shot grasp detection in point clouds
    - 2020-Pointnet++ grasping: Learning an end-to-end spatial grasp generation algorithm from sparse point clouds

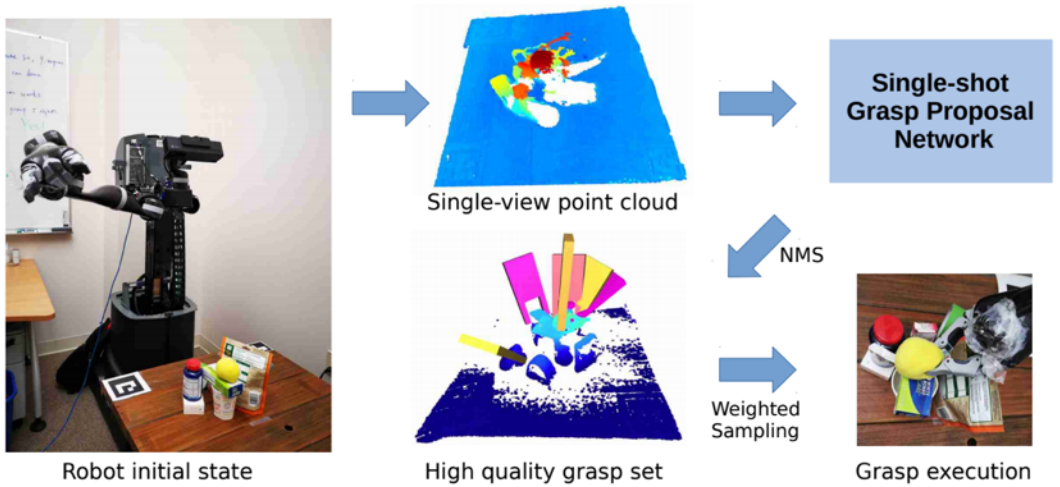


Figure 1: Illustration of the pipeline of Single-Shot SE(3) Grasp Detection (S^{4}G). Taking as input the view point cloud from the depth sensor, S^{4}G regresses the 6-DoF grasp pose directly and predicts the grasp quality for each point, which is more robust and effective.

S^{4}G

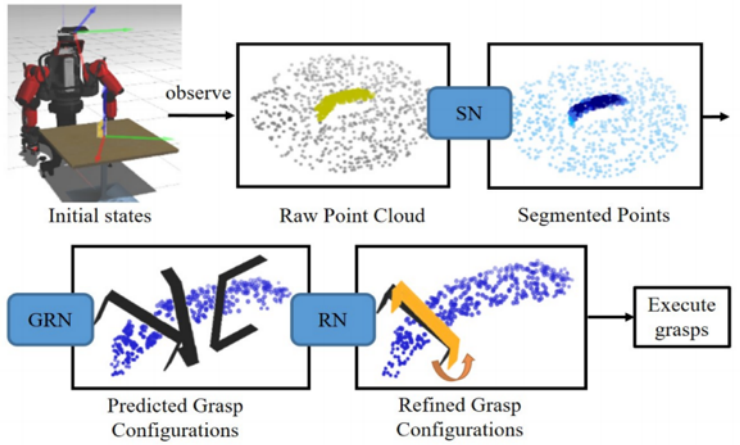
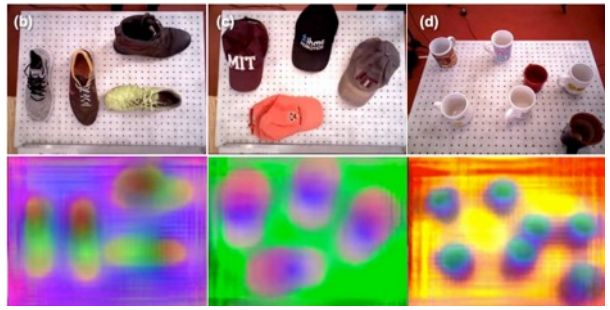


Fig. 1. Illustration of the pipeline in this paper. SN uses the point cloud as input, and outputs segmented points with grasp confidence. Then GRN predicts grasps from grasp regions. After RN refining the predicted grasps, we execute grasping.

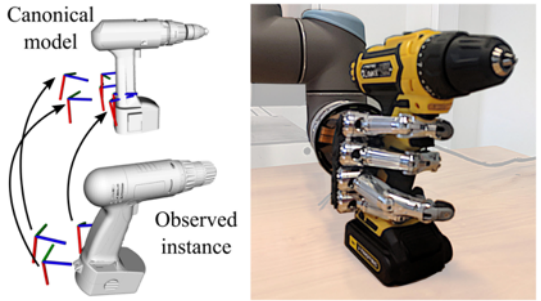
REGNet

基于部分点云的抓取

- 从已有抓取迁移
  - 针对一个类别内的物体，基于已有的抓取，迁移到Novel Objects
  - 代表方法：
    - 2015-Category-based task specific grasping
    - 2016-Part-based Grasp Planning for Familiar Objects
    - 2018-Dense object nets: Learning dense visual object descriptors by and for robotic manipulation（2D planar grasp）
    - 2019-Transferring grasp configurations using active learning and local replanning
    - 2019-Transferring Category-based Functional Grasping Skills by Latent Space Non-Rigid Registration
    - 2020-Dgcm-net: Dense geometrical correspondence matching network for incremental experience-based robotic grasping



Dense object nets



2019-Transferring Category-based Functional

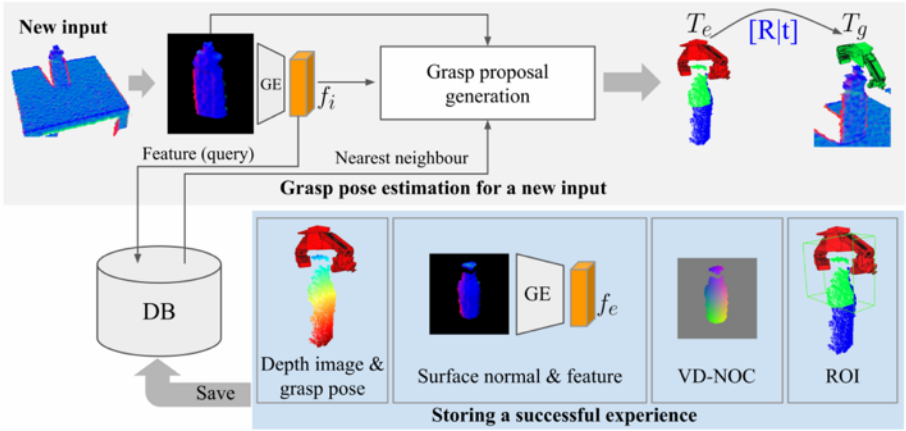


Figure 1: Overview of storing and retrieving experience with the incremental grasp learning framework.

2020-DGCM-Net

6D抓取-总结

- 基于完整形状的抓取
- 当3D模型存在
  - 可以估计物体6D位姿，进而得到抓取
  - 不足：
    - 需要实例分割
    - 需要模型完全一致
- 当3D模型不存在，可以重建得到完整物体的Mesh，再估计抓取位姿
  - 缺少对侧的信息
  - 借助多源数据

- 基于部分点云的抓取
- 没有相似抓取库
  - 间接或直接估计抓取位姿
    - 在数据集上表现很好
    - 具有一定的泛化性能
  - 不足：
    - 需要精确分割，否则影响候选位姿生成
    - 结果不固定，不能保证总能得到很好的结果
- 当有相似抓取库：
  - 迁移抓取位姿
    - 具有一定的可扩展性
  - 不足：
    - 精度不易保证

Part V

挑战和未来研究方向

挑战和未来研究方向

- Insufficient information in data acquisition
  - 只有单个角度，只能得到这个角度下质量较高的抓取
  - 需要已有3D模型配准或者三维补全
  - Multi-view data / Multi-sensor data
- Insufficient amounts of training data
  - 需要训练2D/3D目标检测/分割以及物体/抓取6D位姿估计
  - Simulation / semi-supervised / self-supervised
- Scalabilities in grasping novel objects
  - 如何扩展到新颖物体，并保证较高的成功率
  - Category-level 6D object pose estimation methods
  - Involve more semantic information
- Grasping transparent objects
  - 当前深度相机无法获取透明物体的深度
  - Hardware / Estimate 6D pose / Estimate 3D geometry
- 其他工程性问题
  - 硬件质量/算法精度速度/算法鲁棒性
- 任务导向的抓取/支持灵巧手/视觉驱动的闭环抓取等

总结

- Reference paper lists:
  - https://github.com/GeorgeDu/vision-based-robotic-grasping

个人微信号

Thanks!

公众号：3D视觉前沿