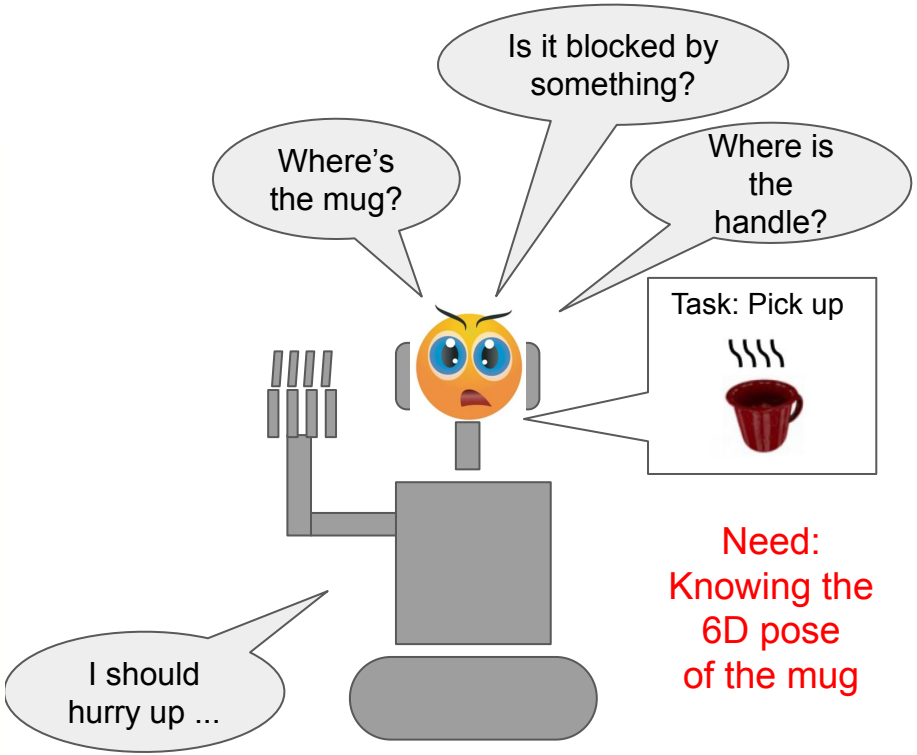
Learning 6D Object Pose on Point Clouds

Oral Defense

Ge Gao

16.06.2021, Hamburg

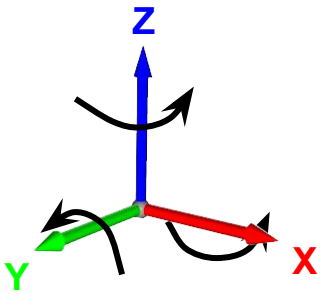
Want a morning coffee ?

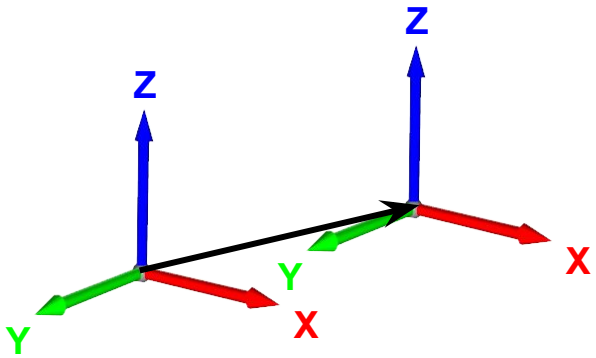


Need: Knowing the 6D pose of the mug

Not a trivial task!

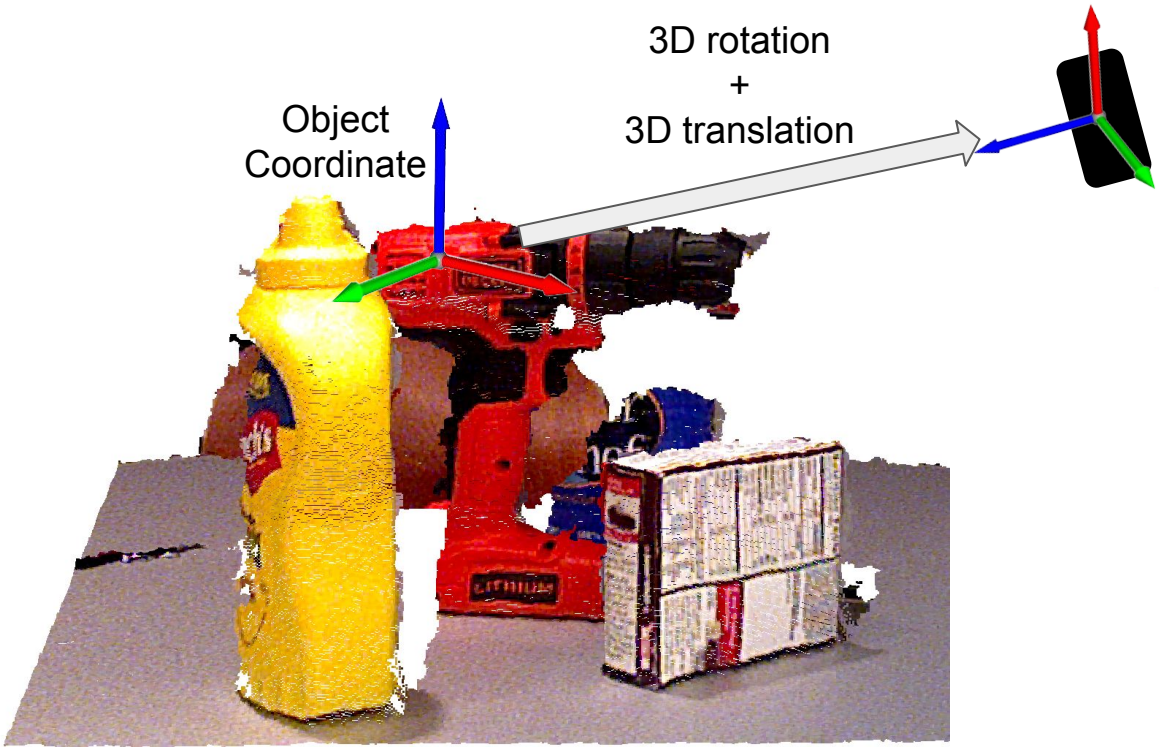
The problem: what is a 6D pose

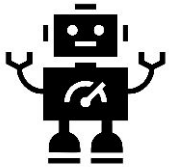




3D rotation + 3D translation = **6D Pose**

The Problem: 6D Object Pose Estimation



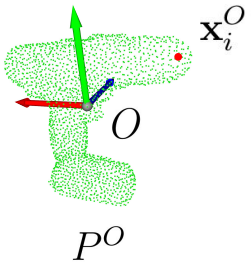


robotic grasping

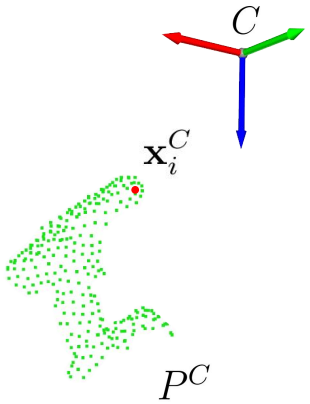
augmented reality

The Problem: 6D Object Pose Estimation for Known Objects

Known Object Model



Camera Observation



List of 3D points (point clouds)

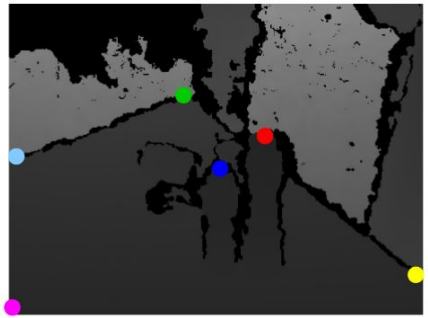
$$P^O = \{\mathbf{x}_i^O \in \mathbb{R}^3 \mid i = 1, \ldots, n\}$$

$$P^C = \{\mathbf{x}_i^C \in \mathbb{R}^3 \mid i = 1, \ldots, m\}$$

$$\mathbf{x}_i^C = R\mathbf{x}_i^O + \mathbf{t}$$

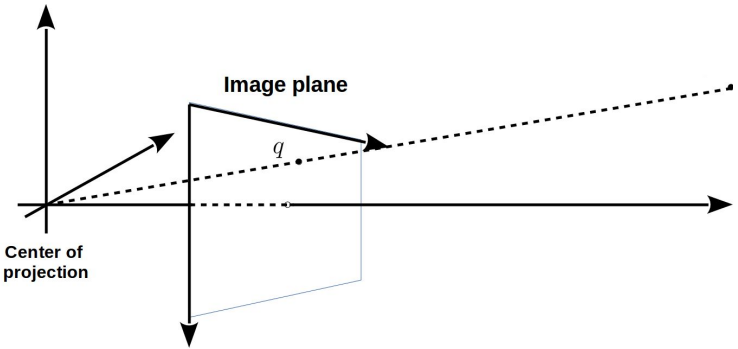
A computer vision system: possible input

RGB Image ⇨ Color



Depth Image ⇨ 2.5D(Z)

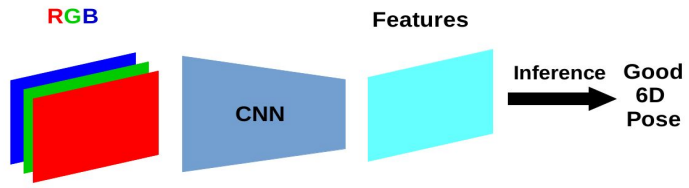
Point Cloud ⇨ 3D(XYZ)



Q=(X, Y, Z)

SOTA: Deep Learning Based Methods

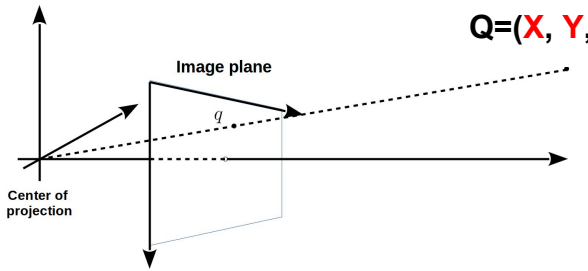
RGB Based



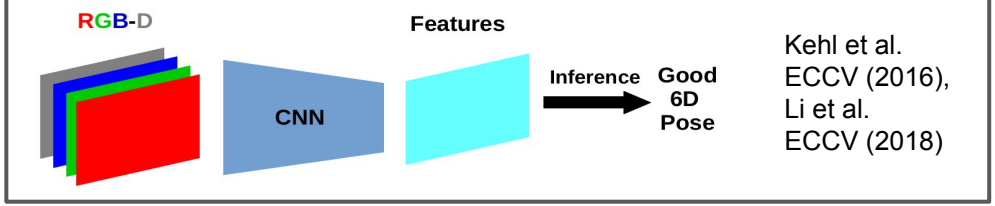
Kehl et al. ICCV (2017)
Oberweger et al. ECCV (2018)
Tremblay et al. CoRL (2018)
Park et al. ICCV (2019)

Observation: RGB is the dominant information ...

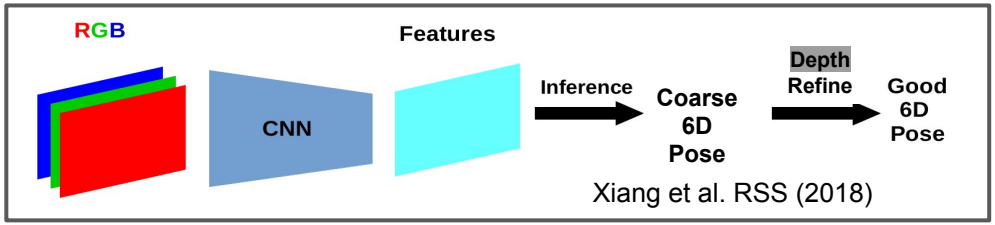
… but depth has rich geometric information!



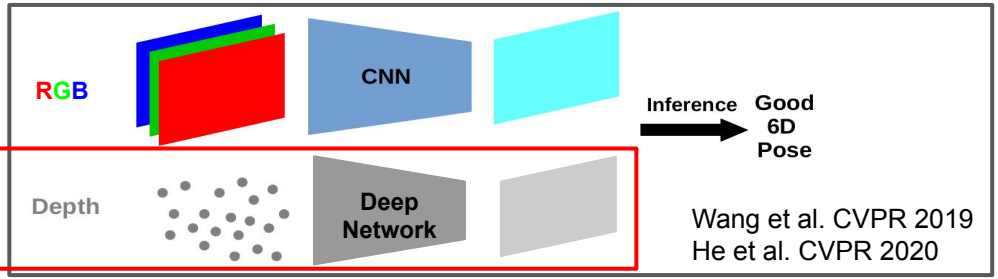
RGB-D Based



Kehl et al. ECCV (2016),
Li et al. ECCV (2018)



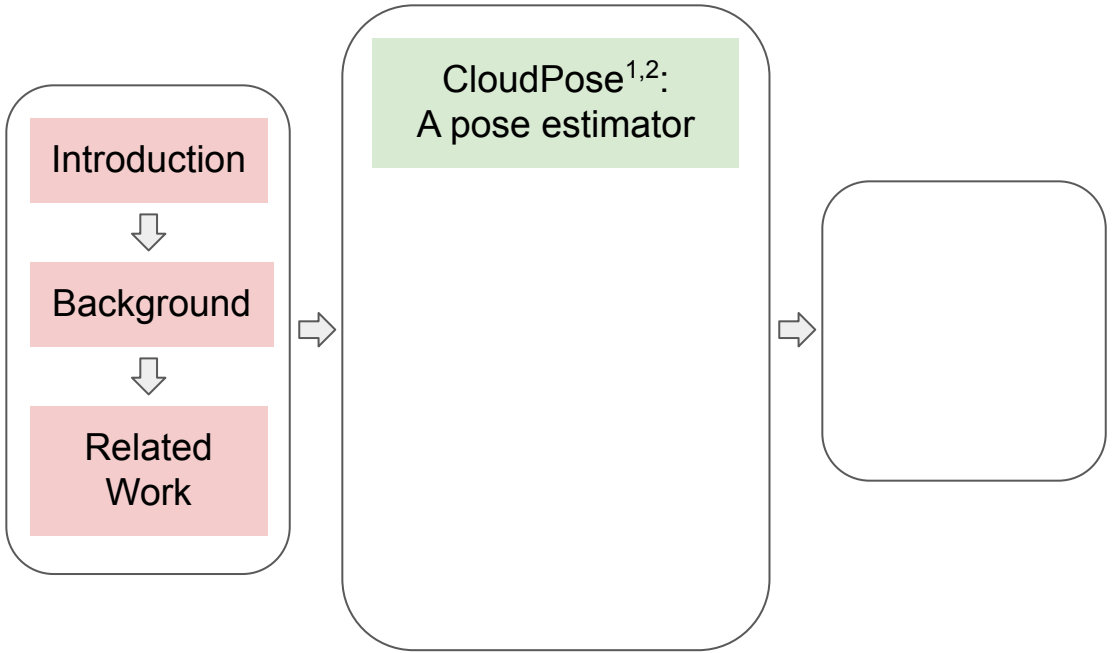
Xiang et al. RSS (2018)



This Thesis

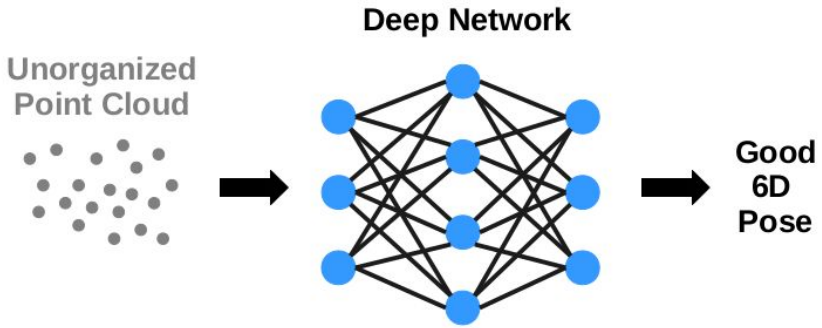
Wang et al. CVPR 2019
He et al. CVPR 2020

Overview



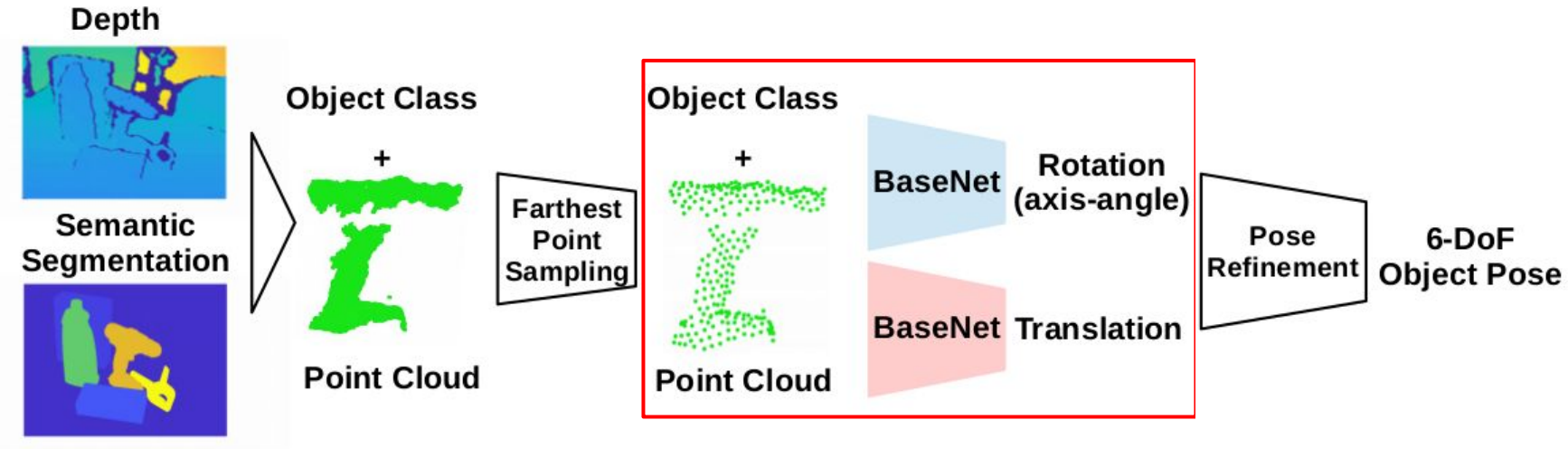
[1]Gao et al., Occlusion Resistant Object Rotation Regression from Point Cloud Segments, ECCVW, 2018
[2]Gao et al., 6D Object Pose Regression via Supervised Learning on Point Clouds, ICRA, 2020

CloudPose: 6D Object Pose Regression on Point Clouds



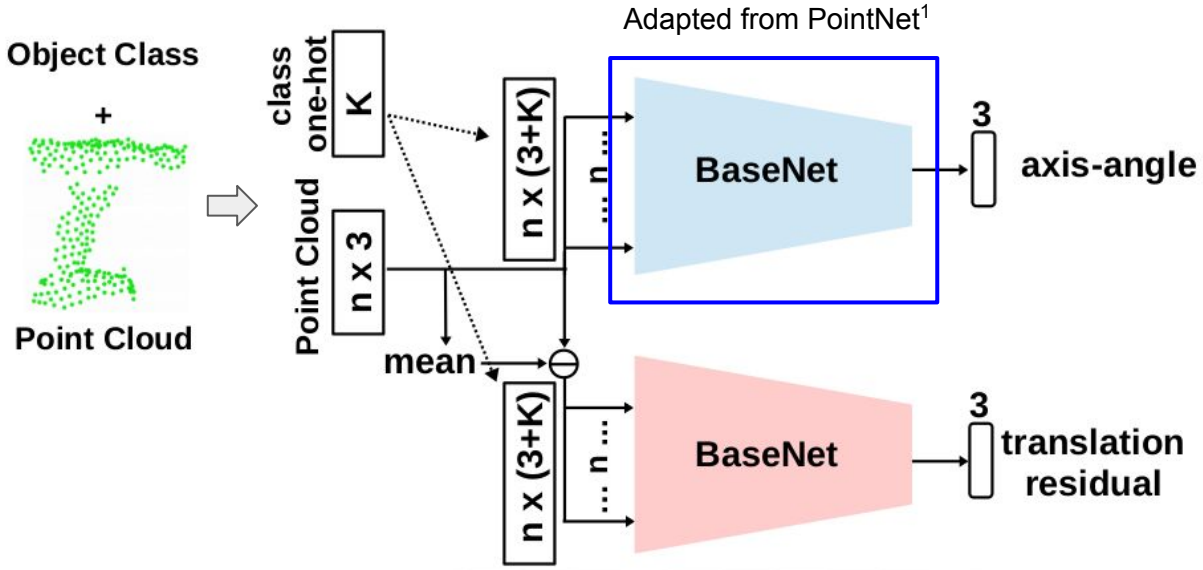
- Is it possible to achieve state-of-the-art performance when inferring 6D object pose with depth information? And how?
- Should we estimate rotation and translation with separate networks?
  - 1 radian vs. 1 meter
- What is a suitable rotation representation in a supervised learning framework?
  - rotation matrix, euler angles, quaternion, axis angle ...

CloudPose: System overview



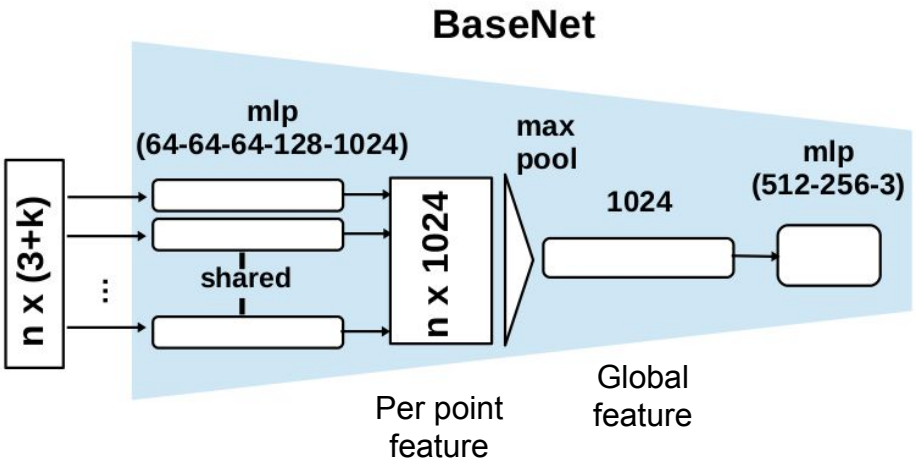
***Ge Gao**, Mikko Lauri, Yulong Wang, Xiaolin Hu, Jianwei Zhang, Simone Frintrop: 6D Object Pose Regression via Supervised Learning on Point Clouds, ICRA, 2020*

CloudPose: Details



[1]C. R. Qi and H. Su and K. Mo and L. J. Guibas. PointNet: Deep learning on point sets for 3D classification and segmentation, in CVPR, 2017.

CloudPose: Details



CloudPose: Rotation representation and loss functions

- Rotation representation
  - Axis-angle
  - Constraint free
- Loss function
  - Rotation loss l_r: geodesic distance
  - Translation loss l_t : L2
  - Total loss: $l = \alpha l_t + l_r$

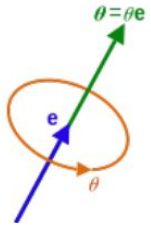
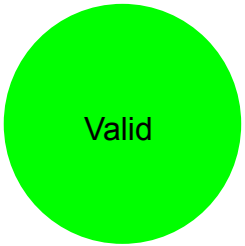
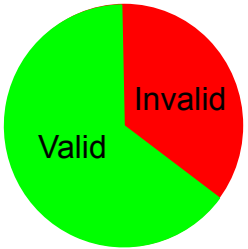


Image credit: Maschen

Network output



w/o constraint



with constraint

CloudPose: Results

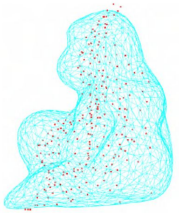
YCB Video (2018)

	RGB	Depth	**Accuracy**
PoseCNN[3]	✔	✔	93.6
DenseFusion[2]	✔	✔	93.2
CloudPose[1]		✔	**94.7**

LineMOD (2012)

	RGB	Depth	**Accuracy**
PoseCNN[3]	✔		62.7
DenseFusion[2]	✔	✔	**94.3**
CloudPose[1]		✔	58.3

Depth data quality is poor !



- **CloudPose Contribution**
  - **First system learns 6D object pose from depth**
  - **Accurate 6D pose can be estimated from depth information**
- **To be improved**
  - **Be robust against noisy depth data**

[1]*Gao et al., 6D Object Pose Regression via Supervised Learning on Point Clouds, ICRA, 2020*
[2]*Wang et al., DenseFusion: 6D object pose estimation by iterative dense fusion, CVPR, 2019*
[3]*Xiang et al., a convolutional neural network for 6D object pose estimation in cluttered scenes, RSS, 2018.*

BOP Challenge 2019 – YCB-Video

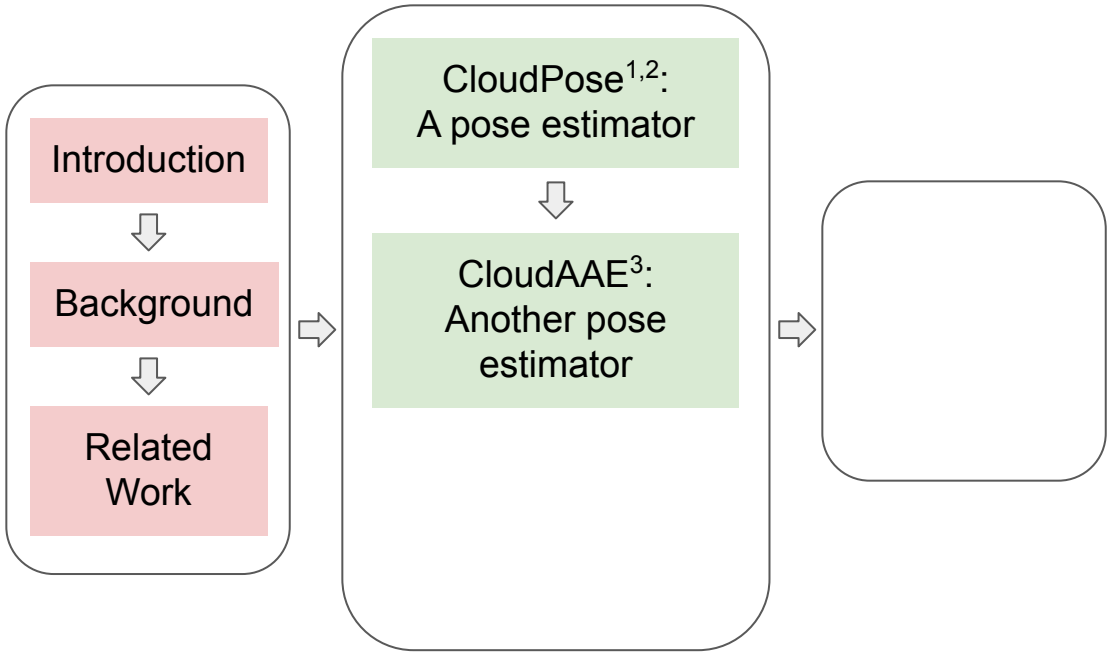
The performance scores are defined in the challenge description. The reported time is the average estimation time per image.

Show 50 entries

Search:

	Date (UTC)	Submission	Test image	AR	AR_{VSD}	AR_{MSSD}	AR_{MSPD}	Time (s)
1	2019-10-18 20:16	Pix2Pose-BOP_w/ICP-ICCV19	RGB-D	0.668	0.693	0.693	0.617	2.106
2	2019-10-05 18:03	gao-cloudpose19	D	0.569	0.540	0.659	0.508	-1.000
3	2019-10-22 09:09	Sundermeyer-IJCV19+ICP	RGB-D	0.498	0.460	0.573	0.462	1.581
4	2019-10-14 14:36	Félix&Neves-ICRA2017-IET2019	RGB-D	0.498	0.708	0.423	0.362	54.509
5	2019-10-22 07:57	Vidal-Sensors18	D	0.435	0.623	0.361	0.322	3.719
6	2019-10-22 02:34	Zhigang-CDPN-ICCV19 (Zhigang-CDPN-ICCV19)	RGB	0.418	0.319	0.432	0.503	0.295
7	2019-10-22 06:37	Sundermeyer-IJCV19	RGB	0.371	0.307	0.407	0.398	0.179
8	2019-08-22 06:12	Drost-CVPR10-Edges	RGB-D	0.368	0.537	0.303	0.263	37.878
9	2019-08-20 09:54	Drost-CVPR10-3D-Only	D	0.332	0.474	0.281	0.241	6.270
10	2019-10-17 07:05	Drost-CVPR10-3D-Only-Faster	D	0.318	0.469	0.263	0.220	1.282
11	2019-08-21 06:09	Drost-CVPR10-3D-Edges	D	0.309	0.454	0.253	0.220	37.479
12	2019-10-14 21:44	Pix2Pose-BOP-ICCV19 (Basic)	RGB	0.284	0.211	0.245	0.398	0.944
13	2019-10-10 14:22	DPOD (synthetic)	RGB	0.222	0.196	0.216	0.254	0.341

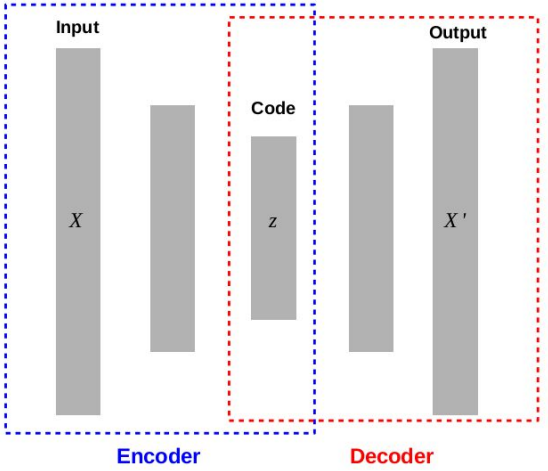
Overview

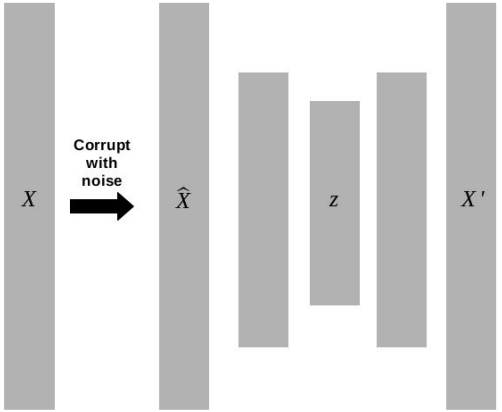


[1]Gao et al., Occlusion Resistant Object Rotation Regression from Point Cloud Segments, ECCVW, 2018
[2]Gao et al., 6D Object Pose Regression via Supervised Learning on Point Clouds, ICRA, 2020
[3]Gao et al., CloudAAE: Learning 6D Object Pose Regression with On-line Data Synthesis on Point Clouds, ICRA 2021.

CloudAAE: Motivation

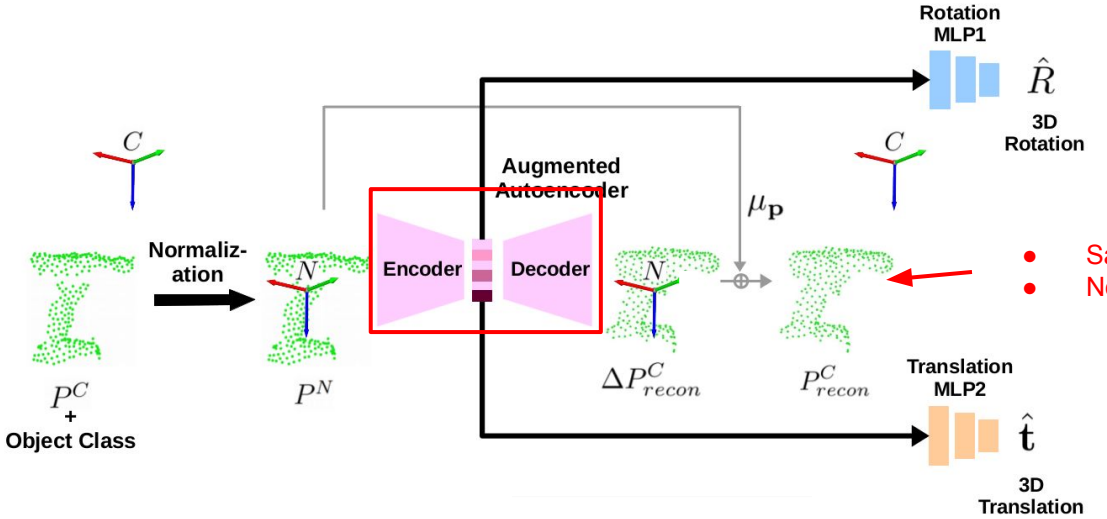
- Sparsity of information
  - Input dimension e.g. 256 × 3 = 768
  - Output dimension 3 + 3 = 6
  - Susceptible to input noise
- Augmented autoencoder[1]
  - a variant of denoising autoencoder
  - Invariant to noise corruption





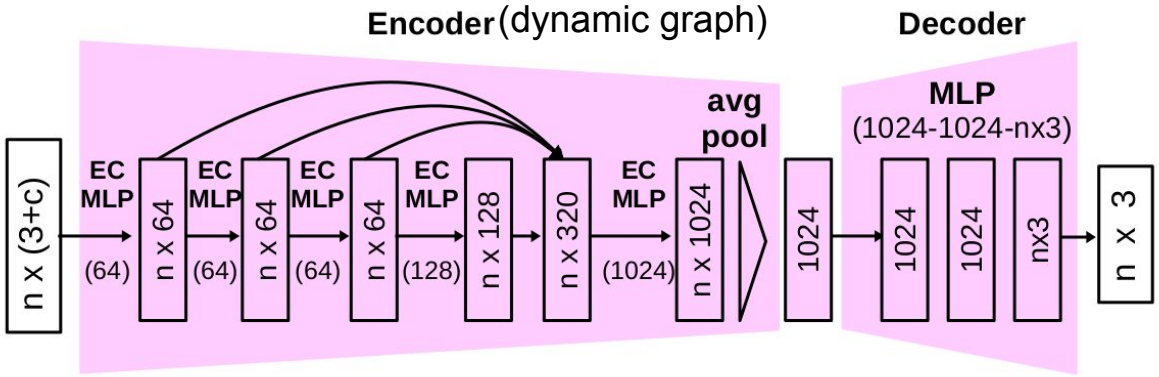
[1] *Sundermeyer et al., Implicit 3D orientation learning for 6D object detection from RGB images, ECCV, 2018.*

CloudAAE: System Architecture

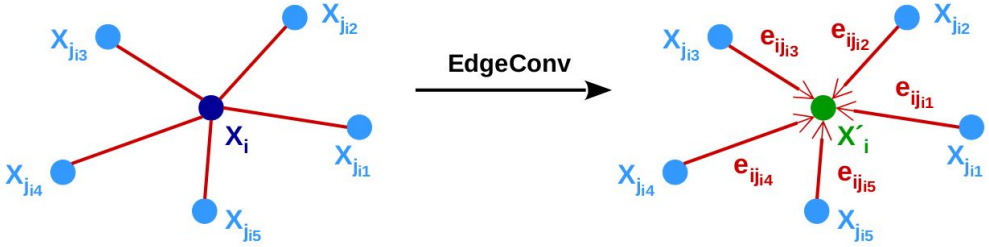


***Ge Gao**, Mikko Lauri, Xiaolin Hu, Jianwei Zhang, Simone Frintrop: CloudAAE: Learning 6D Object Pose Regression with On-line Data Synthesis on Point Clouds, ICRA, 2021*

CloudAAE: Details



EC = EdgeConv[1]



[1] *Wang et al., Dynamic Graph CNN for Learning on Point Clouds, TOG, 2019*

CloudAAE: Results

YCB Video (2018)

	RGB	Depth	Accuracy
DenseFusion[1]	✔	✔	93.2
PVN3D[2]	✔	✔	**96.1**
CloudPose[3]		✔	94.7
CloudAAE[4]		✔	94.0

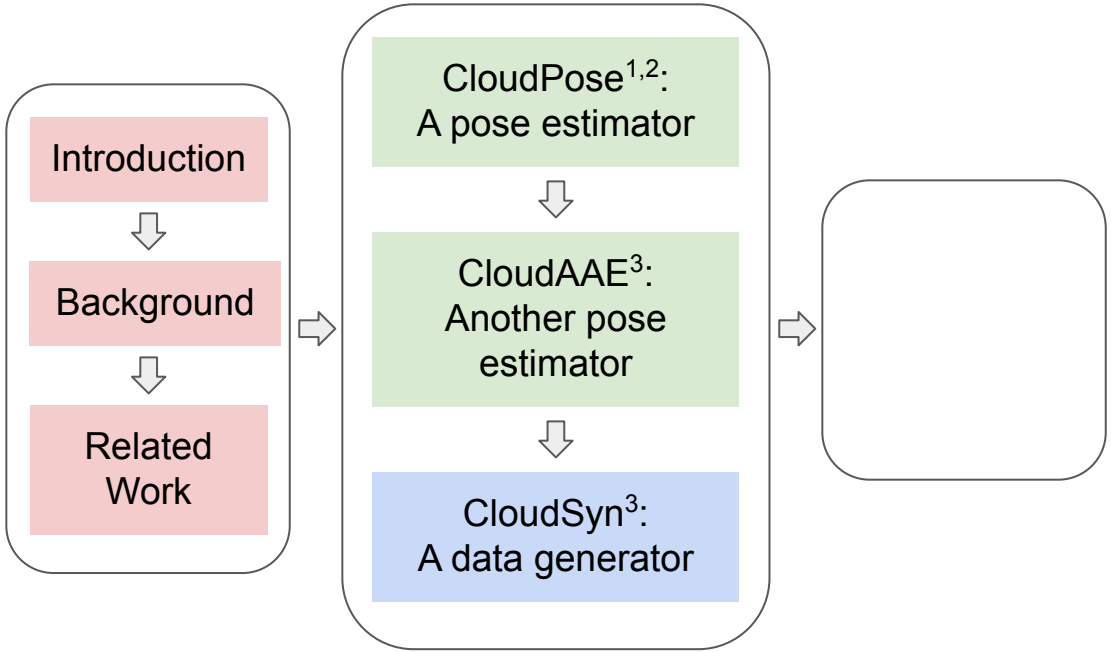
LineMOD (15% / 85%, train / test data, 2012)

	RGB	Depth	Accuracy
PoseCNN[5]	✔		62.7
DenseFusion[1]	✔	✔	**94.3**
CloudPose[3]		✔	58.3
CloudAAE[4]		✔	86.7

- **CloudAAE contribution**
  - **Point cloud-based AAE helps to increase robustness against noise**
- **To be improved**
  - **Getting more training data (easily)**

[1]*Wang et al., DenseFusion: 6D object pose estimation by iterative dense fusion, CVPR, 2019*
[2]*He et al., PVN3D: A deep point-wise 3D keypoints voting network for 6DoF pose estimation, CVPR, 2020.*
[3]*Gao et al., 6D Object Pose Regression via Supervised Learning on Point Clouds, ICRA, 2020*
[4]*Gao et al., CloudAAE: Learning 6D Object Pose Regression with On-line Data Synthesis on Point Clouds, ICRA, 2021*
[5]*Xiang et al., PoseCNN: a convolutional neural network for 6D object pose estimation in cluttered scenes, RSS, 2018*

Overview



[1]*Gao et al., Occlusion Resistant Object Rotation Regression from Point Cloud Segments, ECCVW, 2018*
[2]*Gao et al., 6D Object Pose Regression via Supervised Learning on Point Clouds, ICRA, 2020*
[3]*Gao et al., CloudAAE: Learning 6D Object Pose Regression with On-line Data Synthesis on Point Clouds, ICRA 2021.*

CloudSyn: Motivation

Background

Real data -> Expensive

$ $$$

Synthetic data -> Less expensive

Denninger et al.,
BlenderProc, arXiv 2019.

Existing Methods

Textured 3D model

\+

Physically-Based Renderer

$ (time, hardware storage, off-line)

illumination, shadow, reflection,

physically plausible pose

Synthetic RGB-(D) Images for training

CloudSyn: Motivation

Existing Methods

Textured 3D model

\+

Physically-Based Renderer

$ (time, hardware storage, off-line)

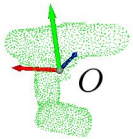
illumination, shadow, reflection,

physically plausible pose

Source of cost: color information

Synthetic RGB-(D) Images for training

Hypothesis



Texture-less 3D model

\+

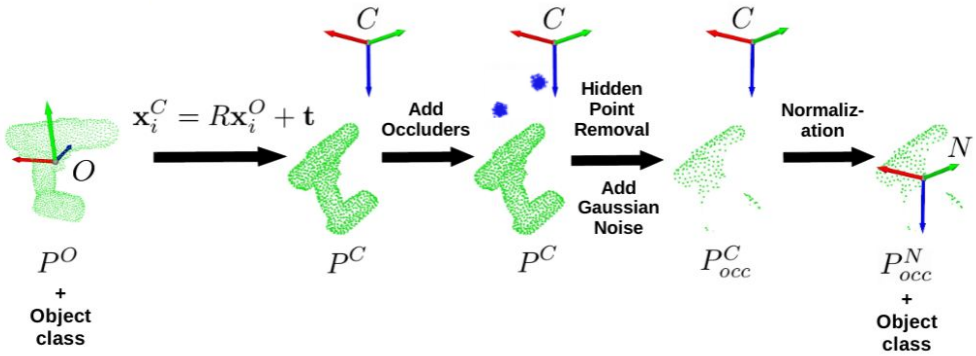
Desired Pose Space

R $\mathbf{t}$

~ 0 cost (time, hardware storage, on-line)

Synthetic Point Clouds for training

CloudSyn: Data Synthesis Pipeline



Real Segment
Synthetic Segment

Speed: < 30 ms for 128 samples

***Ge Gao**, Mikko Lauri, Xiaolin Hu, Jianwei Zhang, Simone Frintrop: CloudAAE: Learning 6D Object Pose Regression with On-line Data Synthesis on Point Clouds, ICRA, 2021*

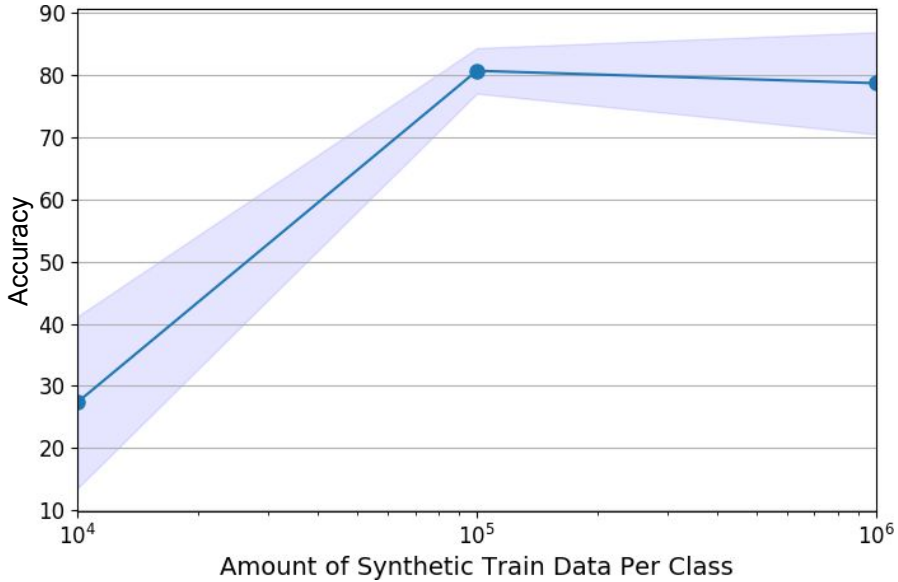
Train CloudAAE with CloudSyn

LineMOD (Synthetic training)

	RGB	Depth	Accuracy
EEGP-AAE[1]	✔		89.2
SSD-6D[2]	✔	✔	90.0
CloudPose[3]		✔	75.2
CloudAAE+Syn[4]		✔	**92.5**

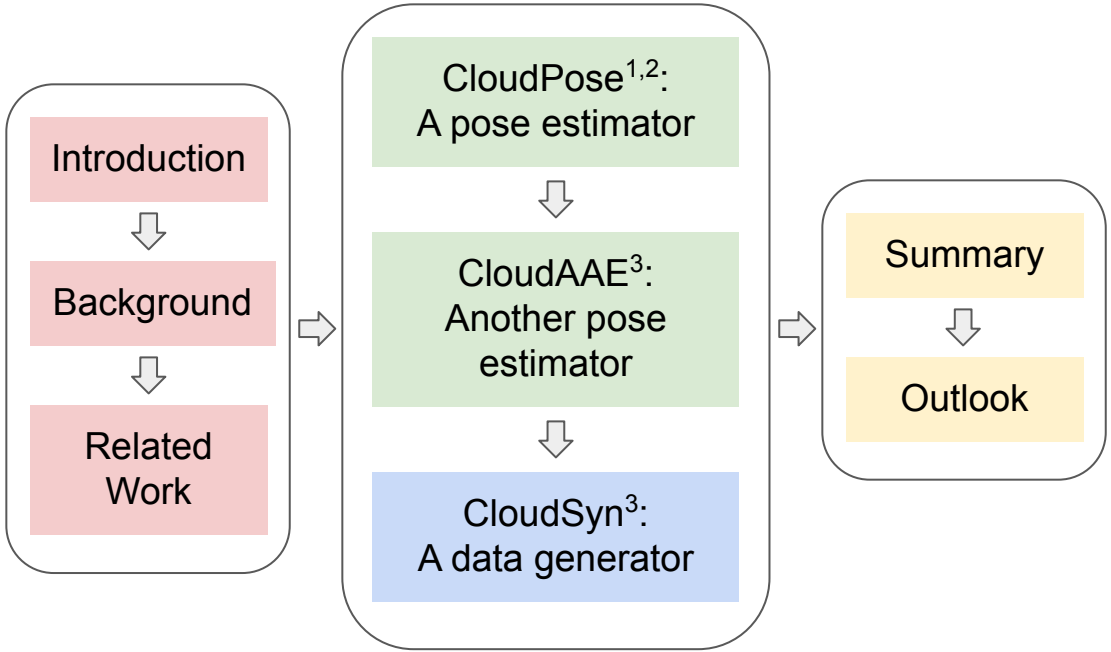
- **CloudSyn contribution**
  - **Point cloud-based synthetic training data can be useful and cheap**

How much data is needed (per class) ?



[1]*Wen et al., Edge enhanced implicit orientation learning with geometric prior for 6D pose estimation, RAL, 2020*
[2]*Kehl et al., SSD-6D: Making RGB-based 3D detection and 6D pose estimation great again, ICCV, 2017*
[3]*Gao et al., 6D Object Pose Regression via Supervised Learning on Point Clouds, ICRA, 2020*
[4]*Gao et al., CloudAAE: Learning 6D Object Pose Regression with On-line Data Synthesis on Point Clouds, ICRA, 2021*

Overview



[1]*Gao et al., Occlusion Resistant Object Rotation Regression from Point Cloud Segments, ECCVW, 2018*
[2]*Gao et al., 6D Object Pose Regression via Supervised Learning on Point Clouds, ICRA, 2020*
[3]*Gao et al., CloudAAE: Learning 6D Object Pose Regression with On-line Data Synthesis on Point Clouds, ICRA 2021.*

Summary

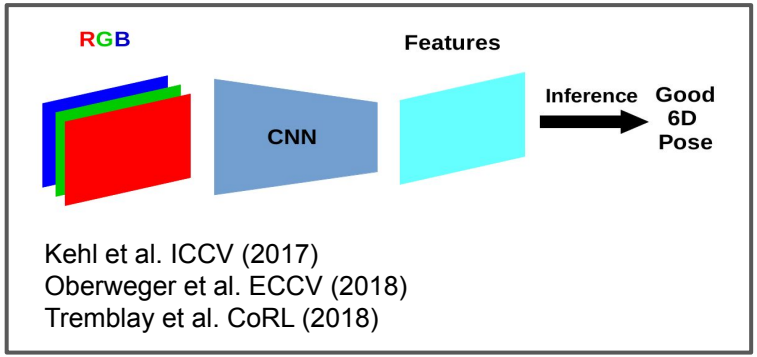
- CloudPose[1,2]
  - Accurate 6D pose can be estimated from depth information in point clouds
- CloudAAE[3]
  - Point cloud-based AAE helps to increase robustness against noise
- CloudSyn[3]
  - Point cloud-based synthetic training data can be useful and cheap

Related Publication

1. **Ge Gao**, Mikko Lauri, Jianwei Zhang, Simone Frintrop: Occlusion Resistant Object Rotation Regression from Point Cloud Segments, **ECCVW**, 2018
2. **Ge Gao**, Mikko Lauri, Yulong Wang, Xiaolin Hu, Jianwei Zhang, Simone Frintrop: 6D Object Pose Regression via Supervised Learning on Point Clouds, **ICRA**, 2020
3. **Ge Gao**, Mikko Lauri, Xiaolin Hu, Jianwei Zhang, Simone Frintrop: CloudAAE: Learning 6D Object Pose Regression with On-line Data Synthesis on Point Clouds, **ICRA**, 2021

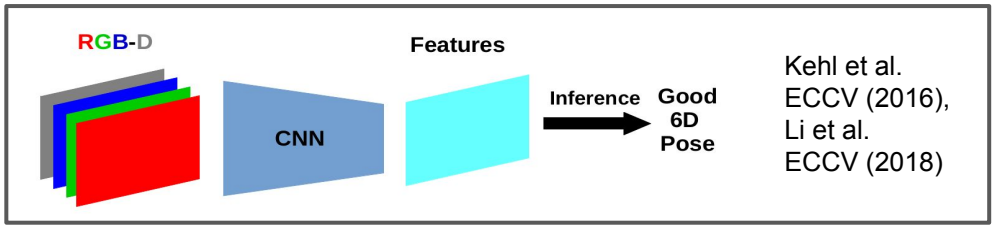
Summary

RGB Based

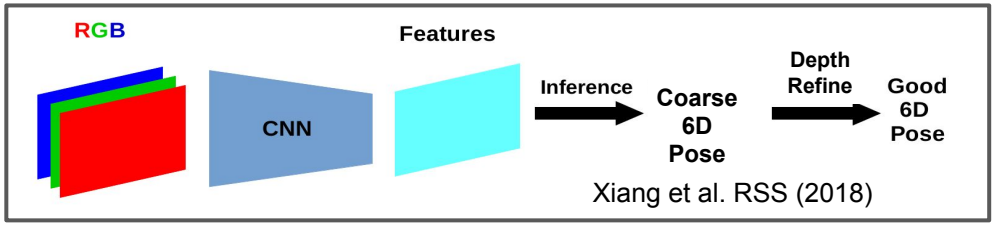


Kehl et al. ICCV (2017)
Oberweger et al. ECCV (2018)
Tremblay et al. CoRL (2018)

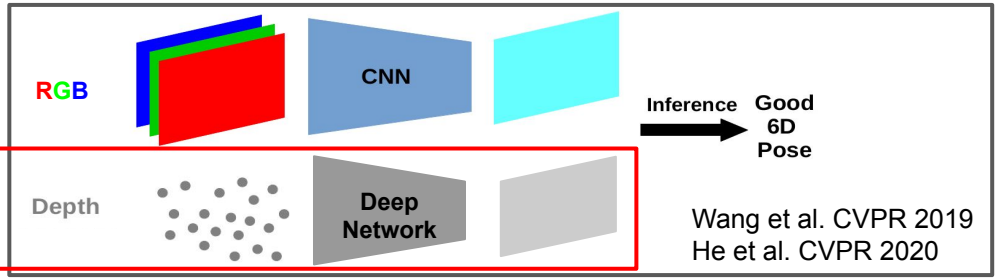
RGB-D Based



Kehl et al. ECCV (2016),
Li et al. ECCV (2018)



Xiang et al. RSS (2018)

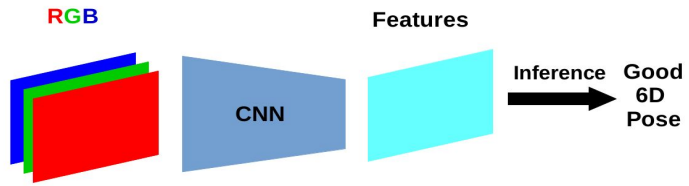


This Thesis

Wang et al. CVPR 2019
He et al. CVPR 2020

Summary

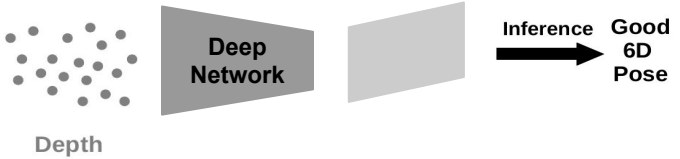
RGB Based



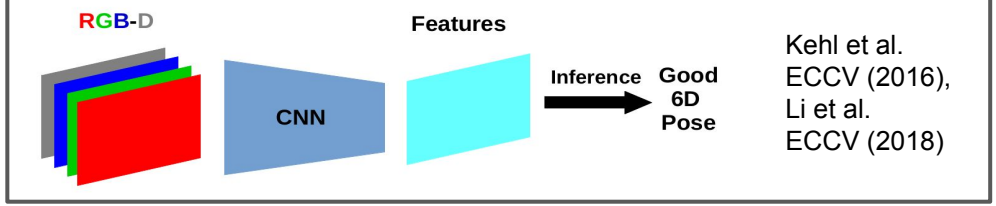
Kehl et al. ICCV (2017)
Oberweger et al. ECCV (2018)
Tremblay et al. CoRL (2018)

Depth Based

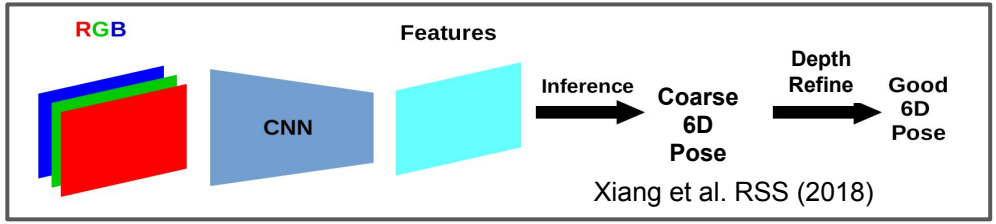
Gao et al. ECCVW (2018)
ICRA (2020, 2021)



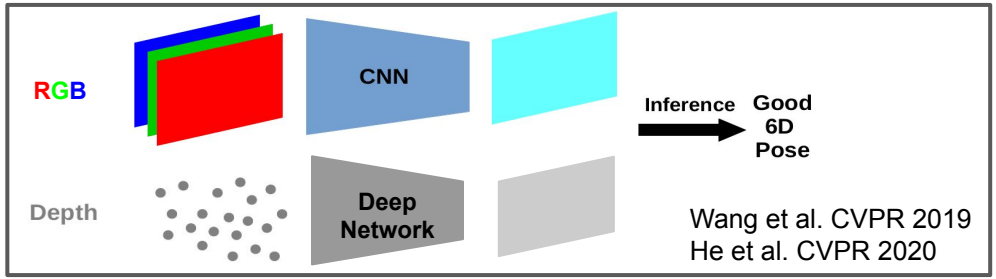
RGB-D Based



Kehl et al. ECCV (2016), Li et al. ECCV (2018)



Xiang et al. RSS (2018)

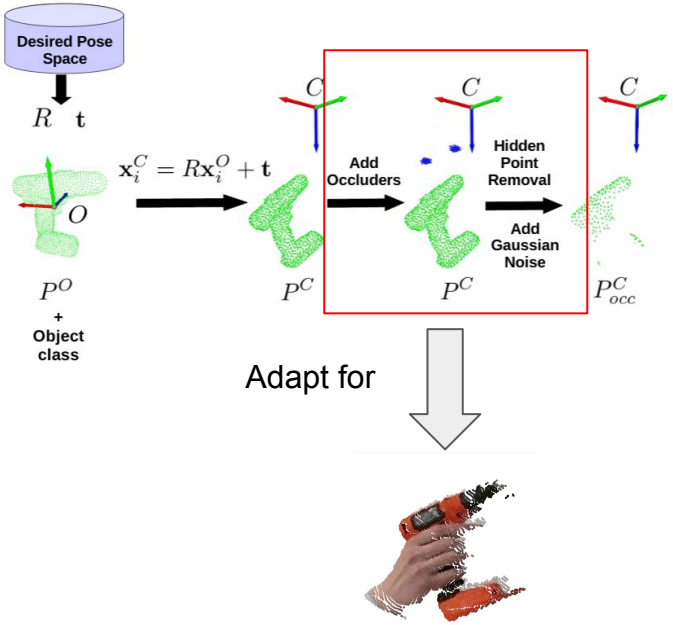


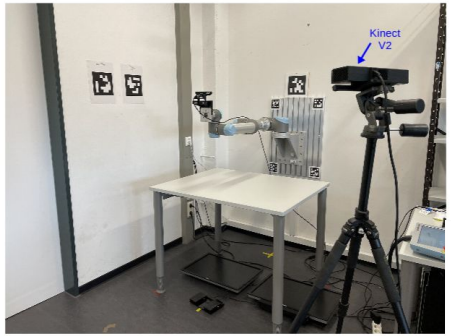
Wang et al. CVPR 2019
He et al. CVPR 2020

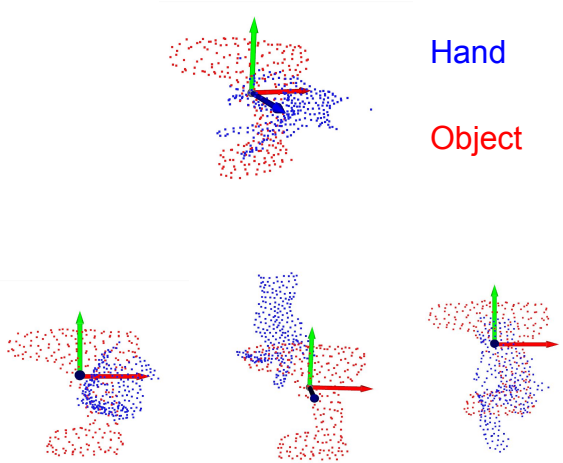
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Outlook

- Hand-over task



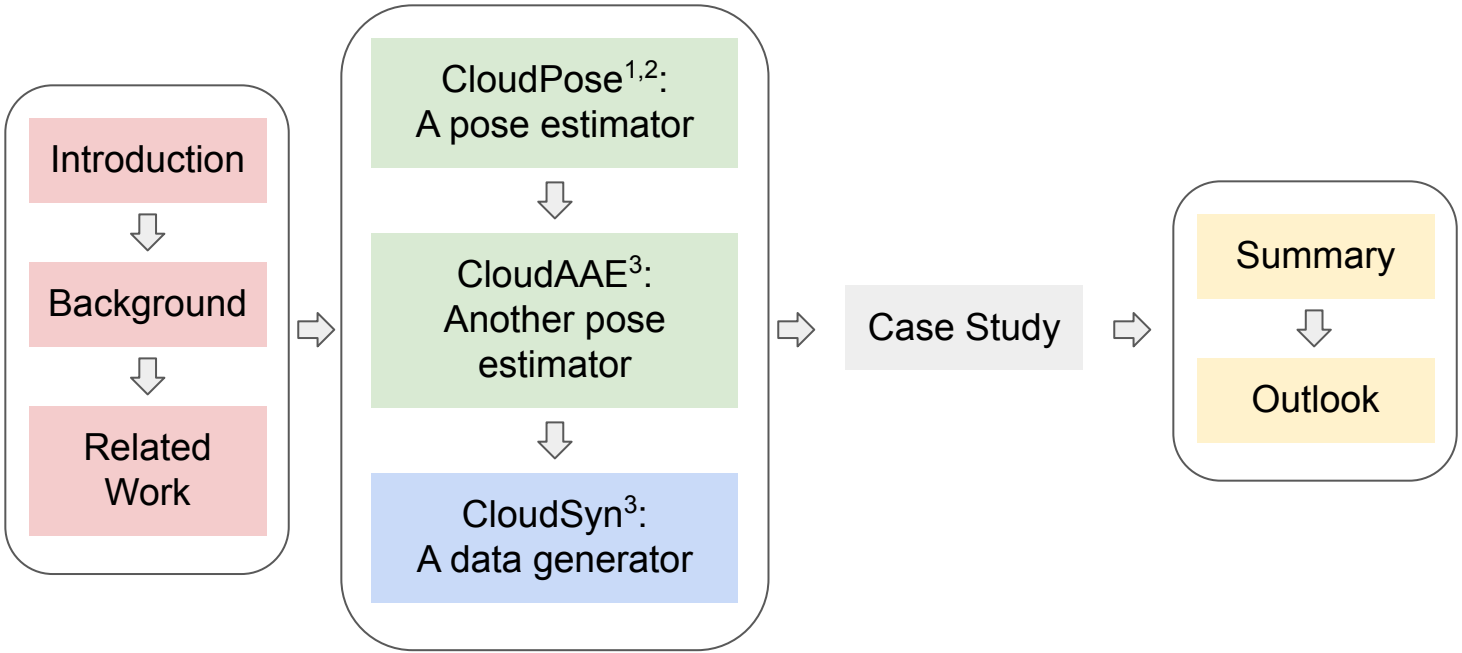




Open topic:

- Sim2Real gap for (complicated) 3D data

Overview

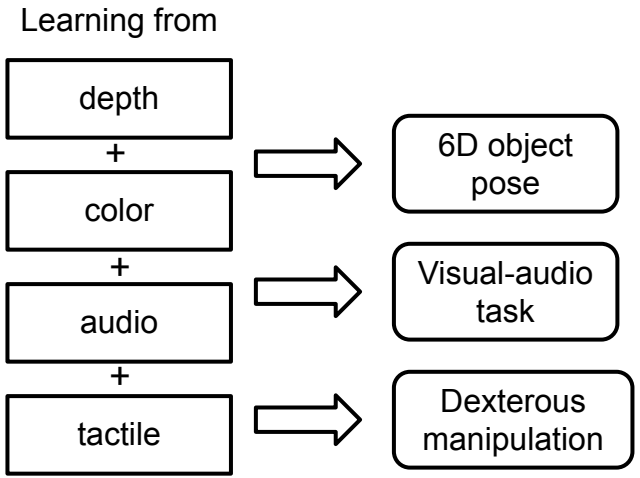


[1]Gao et al., Occlusion Resistant Object Rotation Regression from Point Cloud Segments, ECCVW, 2018
[2]Gao et al., 6D Object Pose Regression via Supervised Learning on Point Clouds, ICRA, 2020
[3]Gao et al., CloudAAE: Learning 6D Object Pose Regression with On-line Data Synthesis on Point Clouds, ICRA 2021.

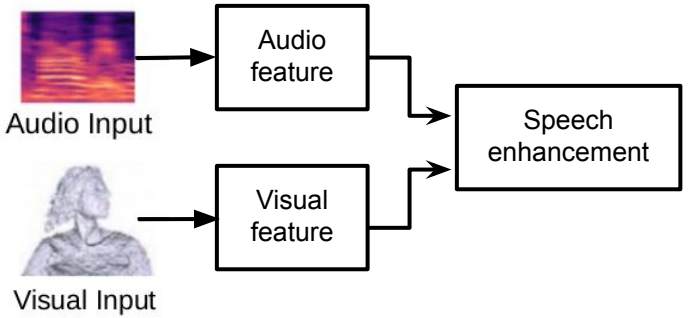
Outlook

- Multimodal data

Image credit:livejapan

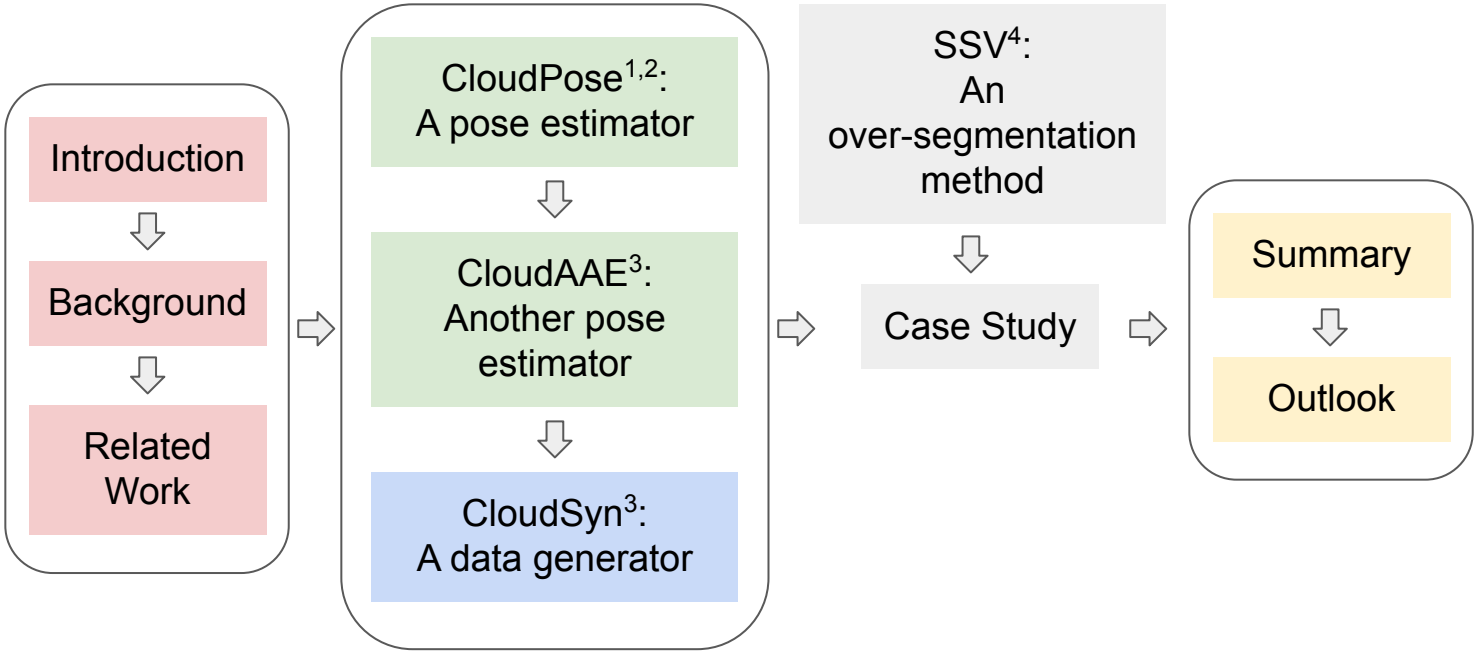


Saturated benchmarking (color+depth)
- ~96% on YCB video dataset,
- ~99% on LineMOD



Open topic:
- (fixed) Fusion strategy
- Dynamically choosing the reliable modality

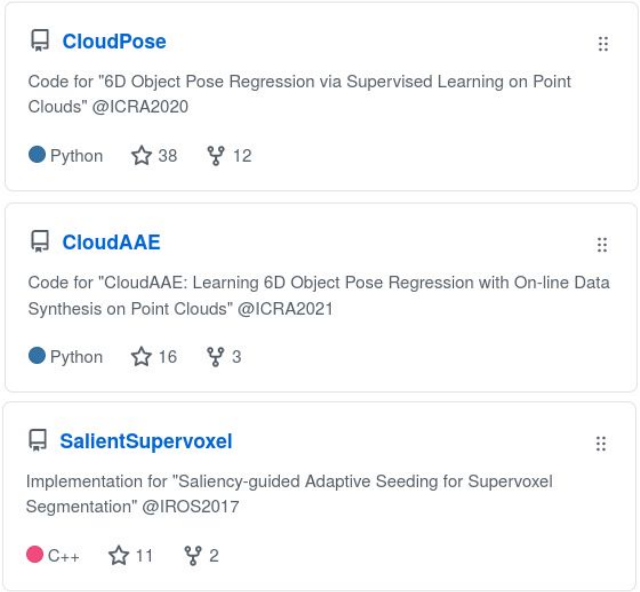
Overview



[1]Gao et al., Occlusion Resistant Object Rotation Regression from Point Cloud Segments, ***ECCVW***, 2018
[2]Gao et al., 6D Object Pose Regression via Supervised Learning on Point Clouds, ***ICRA***, 2020
[3]Gao et al., CloudAAE: Learning 6D Object Pose Regression with On-line Data Synthesis on Point Clouds, ***ICRA***, 2021.
[4]Gao et al., Saliency-guided Adaptive Seeding for Supervoxel Segmentation, ***IROS***, 2017

Thank you!

All contributions are open source:
https://github.com/GeeeG



Thesis Findings

- CloudPose
  - Accurate 6D pose can be estimated from depth information in point clouds
- CloudAAE
  - Point cloud-based AAE helps to increase robustness against noise
- CloudSyn
  - Point cloud-based synthetic training data can be useful and cheap