PrimitivePose: 3D Bounding Box Prediction of Unseen Objects via Synthetic Geometric Primitives

A. Kriegler[1,2], C. Beleznai[1], M. Murschitz[1], K. Göbel[1] and M. Gelautz[1,2]

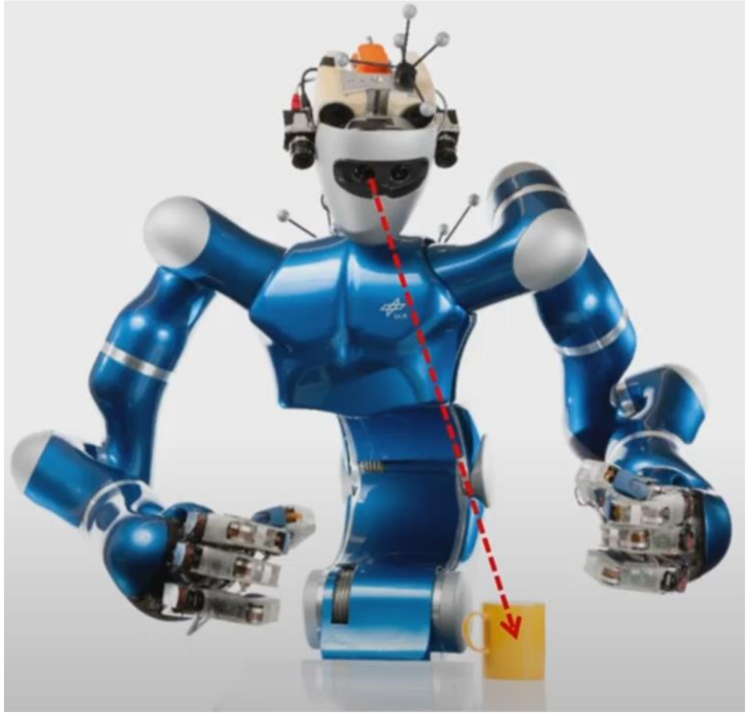
AIT Austrian Institute of Technology[1]:
Center for Vision, Automation and Control
Assistive and Autonomous Systems

TU Wien[2]:
Visual Computing and Human-Centered Technology
Computer Vision

AGENDA

- Task description
- 3D detection pipeline
  - Synthetic data generation
  - Depth estimation + surface normals
  - Learning 3D detection
- Results

3D OBJECT DETECTION FROM STEREO



No 3D models required

| Left stereo | Right stereo |

Only stereo images

$$R_{cam2obj} \in \mathbb{R}^{3x3}$$
$$t_{cam2obj} \in \mathbb{R}^{3x1}$$
$$d_{obj} \in \mathbb{R}^{3x1}$$

?

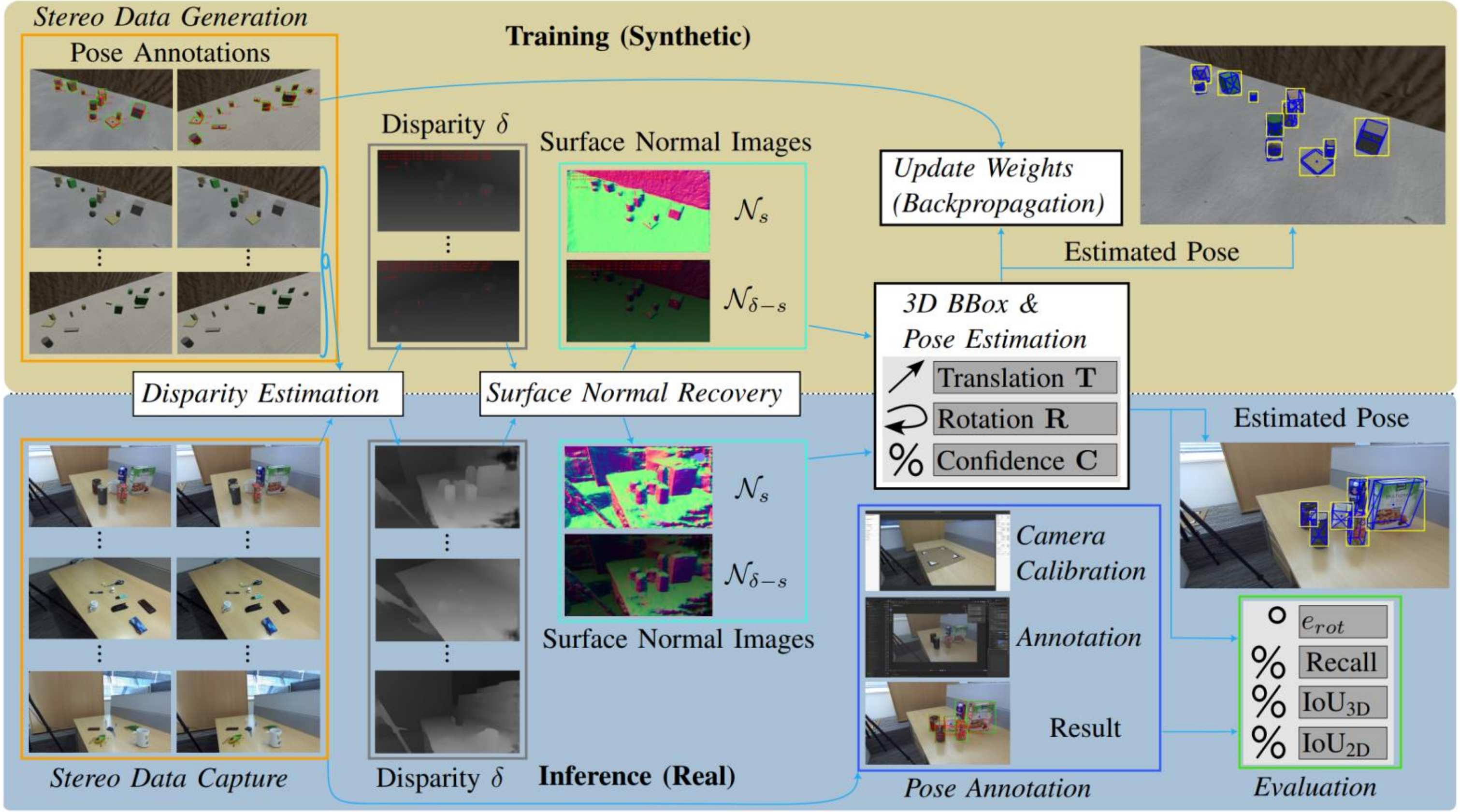
CHALLENGES

- Unseen objects
- Appearance variations
- Symmetries
- Textureless objects
- **Pose annotations for training data**
- **Synth-to-Real gap**

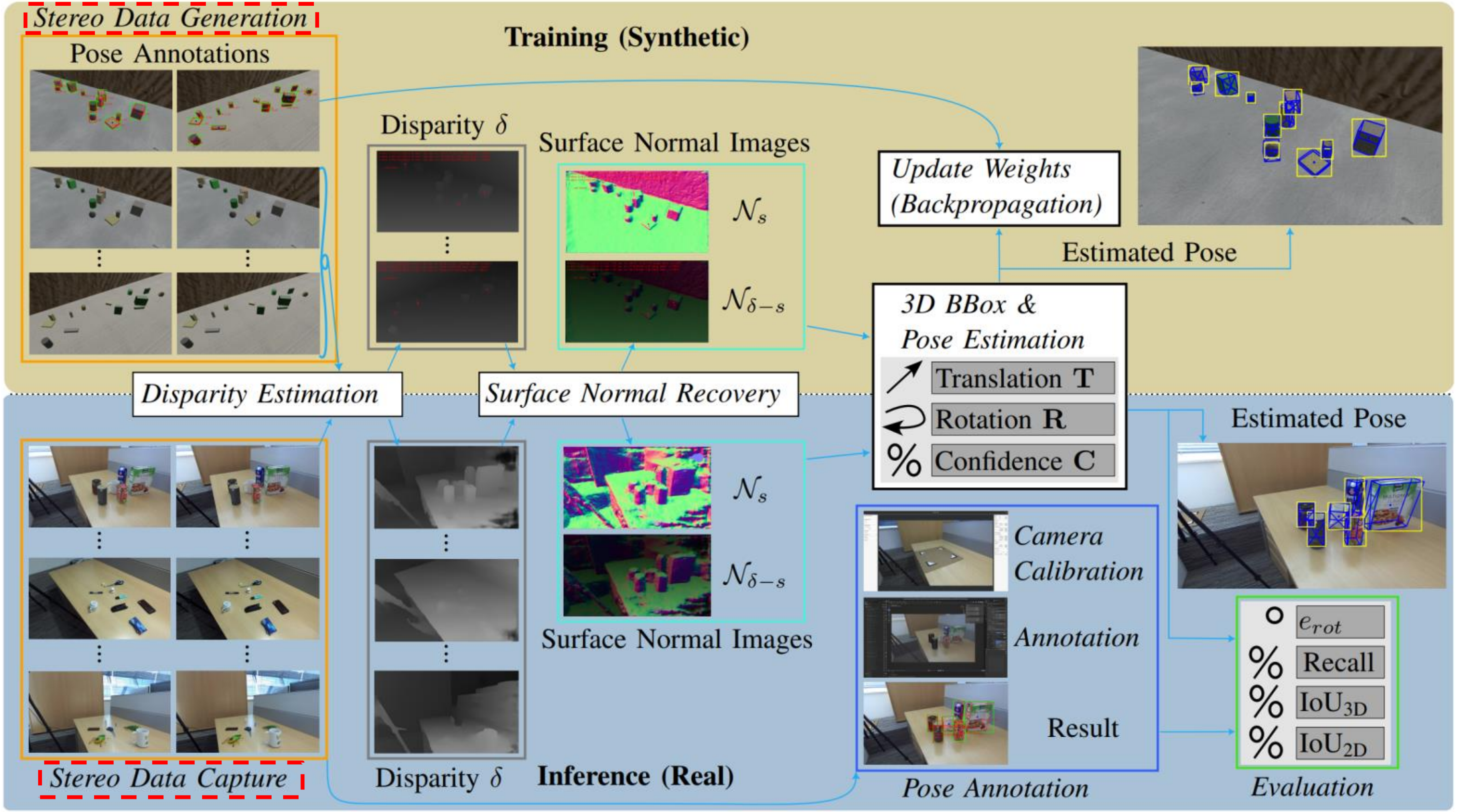
TableTop dataset [1]

Own tabletop evaluation images

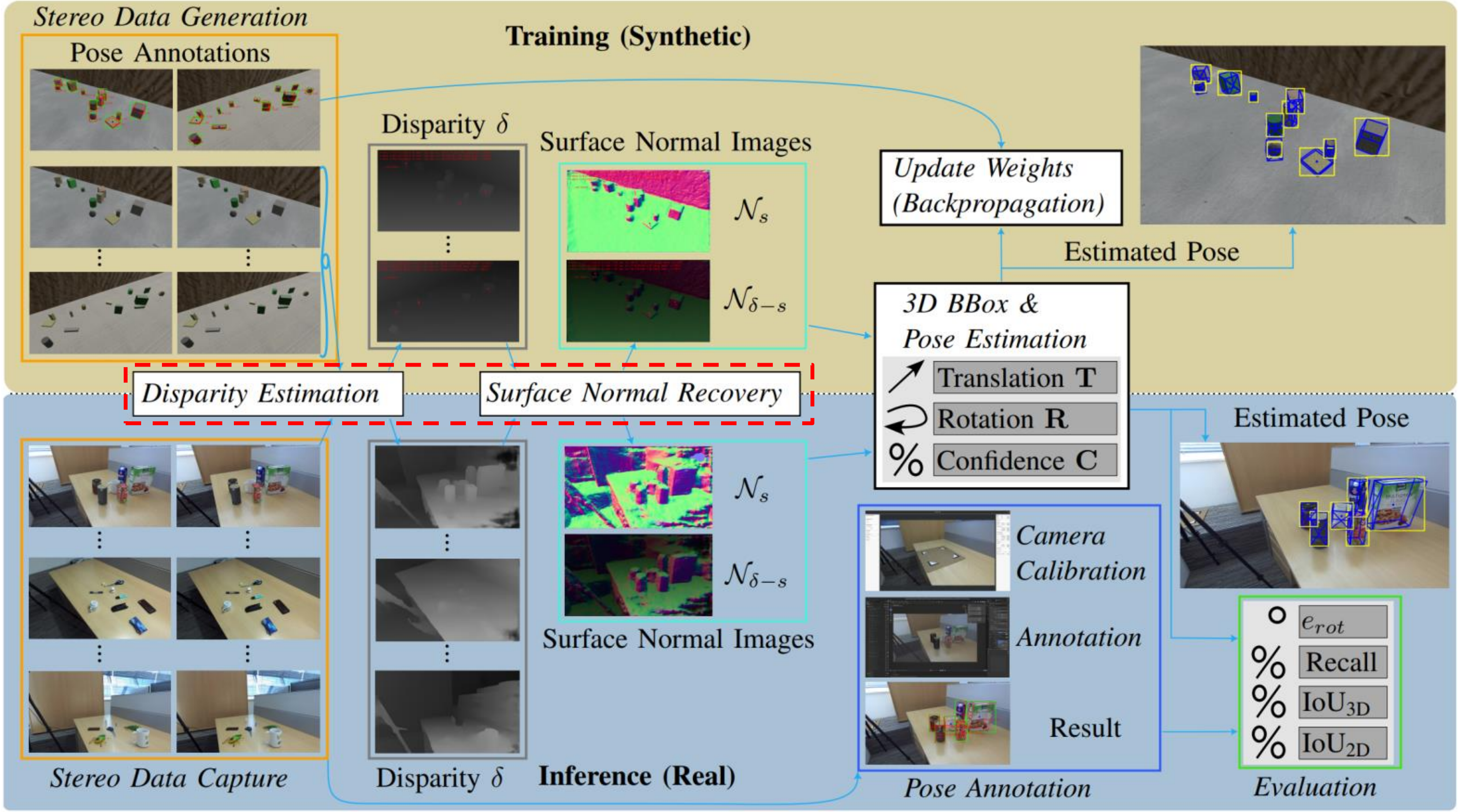
3D BOUNDING BOX DETECTION PIPELINE



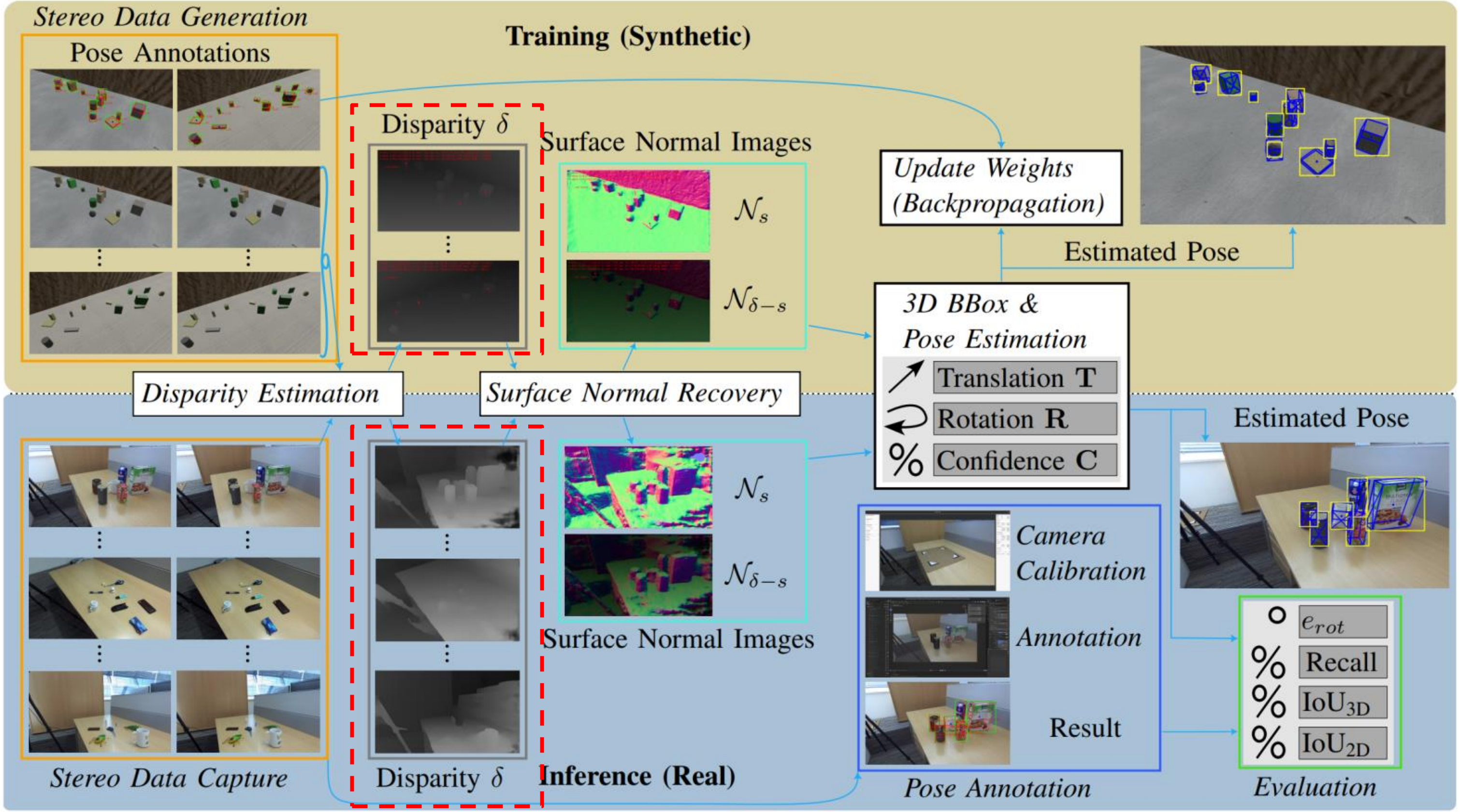
3D BOUNDING BOX DETECTION PIPELINE



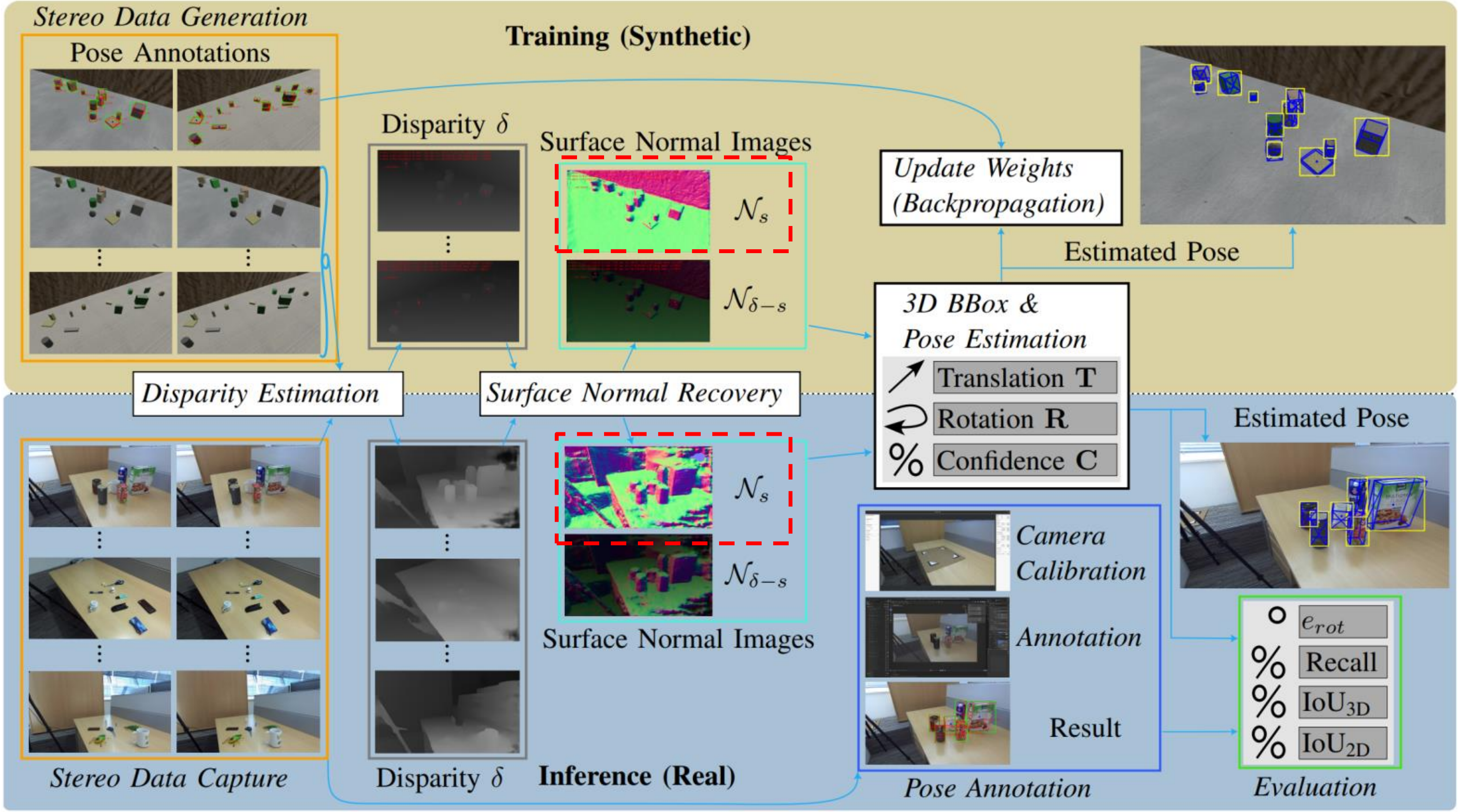
3D BOUNDING BOX DETECTION PIPELINE



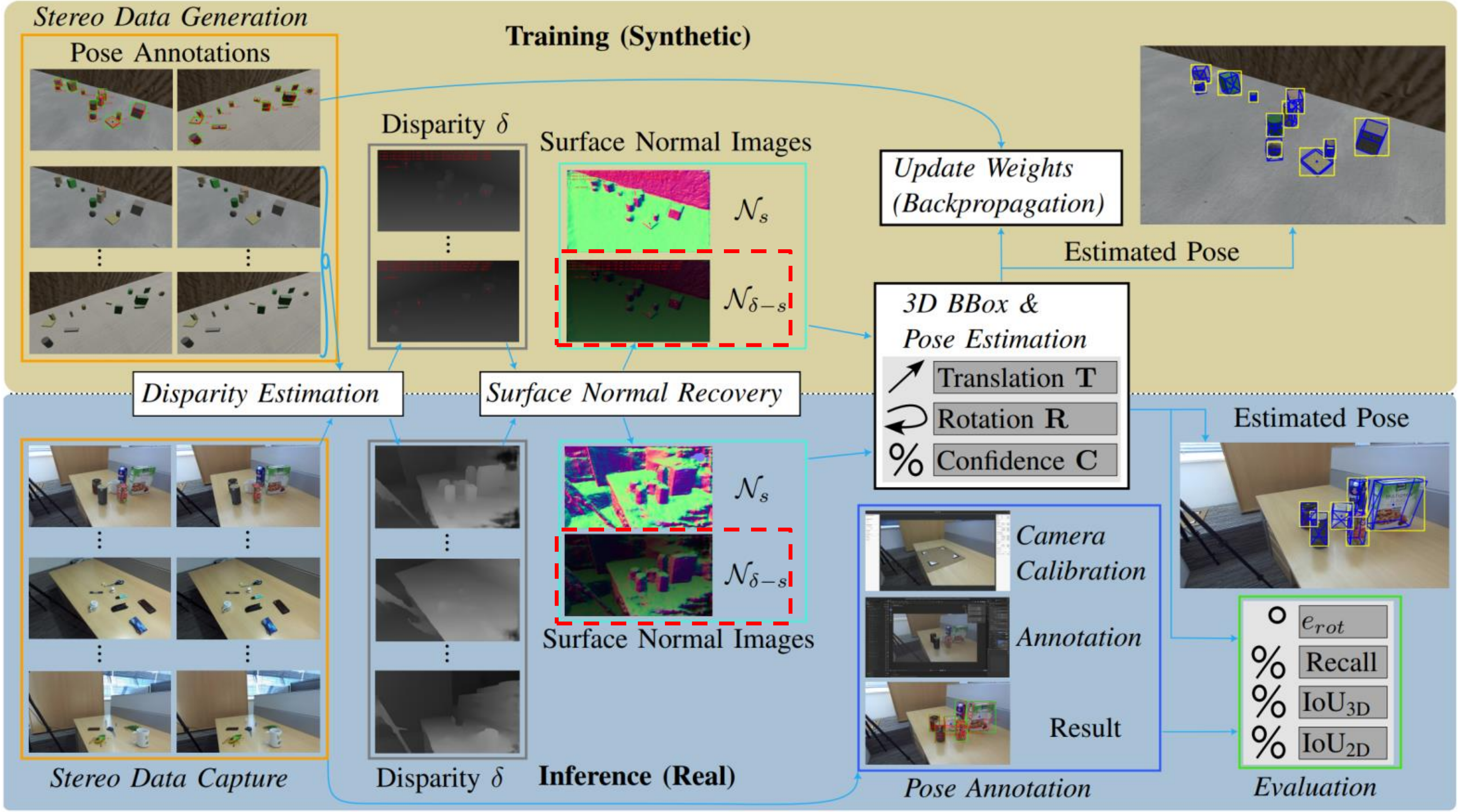
3D BOUNDING BOX DETECTION PIPELINE



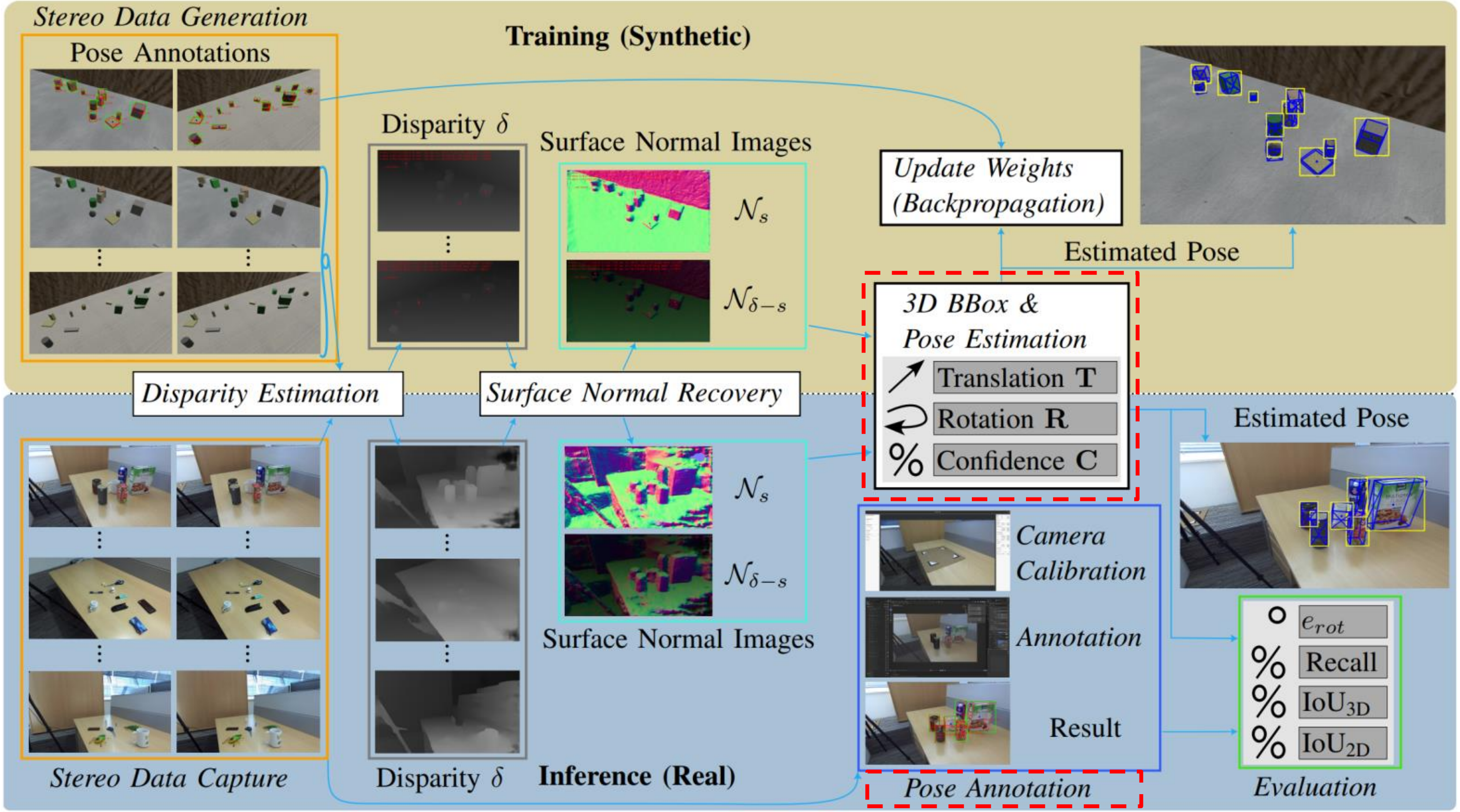
3D BOUNDING BOX DETECTION PIPELINE



3D BOUNDING BOX DETECTION PIPELINE

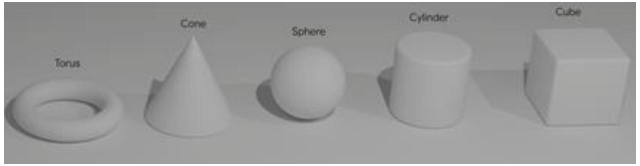


3D BOUNDING BOX DETECTION PIPELINE

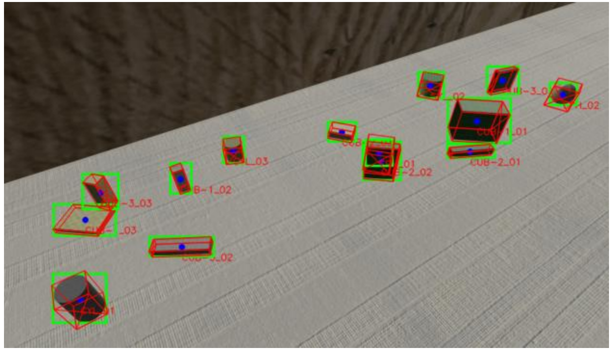


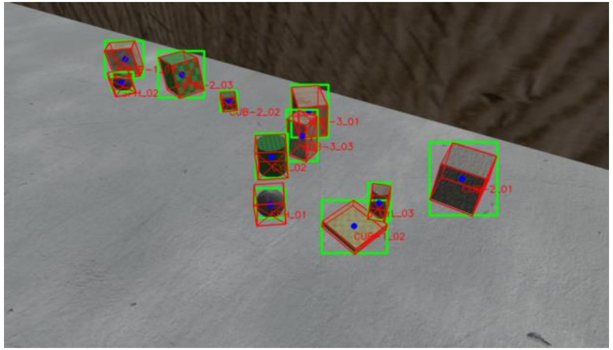
SYNTHETIC DATA GENERATION

- Blender to render synthetic stereo images
- 3D primitives



- Obtain GT 6D pose, object size, occlusion etc.
- Overcomes data sparsity but what about synth-to-real gap?





DEPTH ESTIMATION AND SURFACE NORMALS

- Stereo model (AANet [2]) for disparity
- „Real“ stereo-matcher on synthetic images delivers close-to-real disparity maps [3]
- Obtain surface normal images from disparity
- Disparity-scaled surface normal images preserve depth information as well

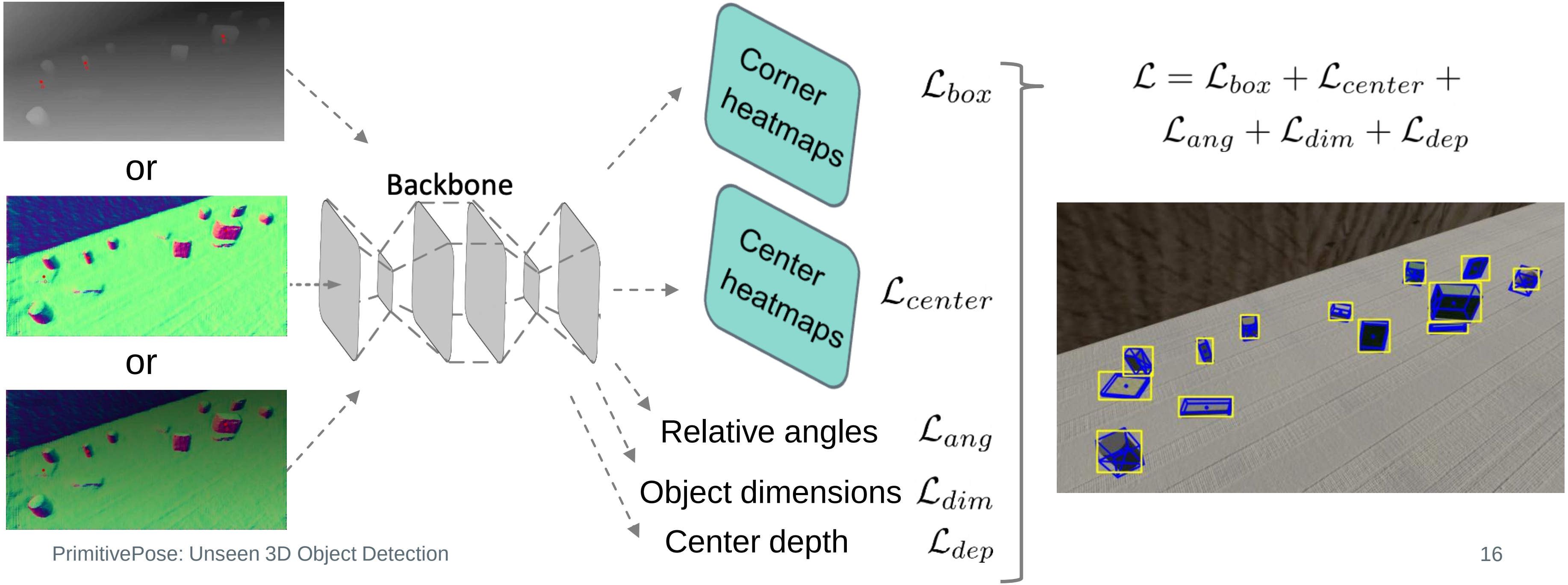
δ

$\mathcal{N}_s$

$\mathcal{N}_{\delta-s}$

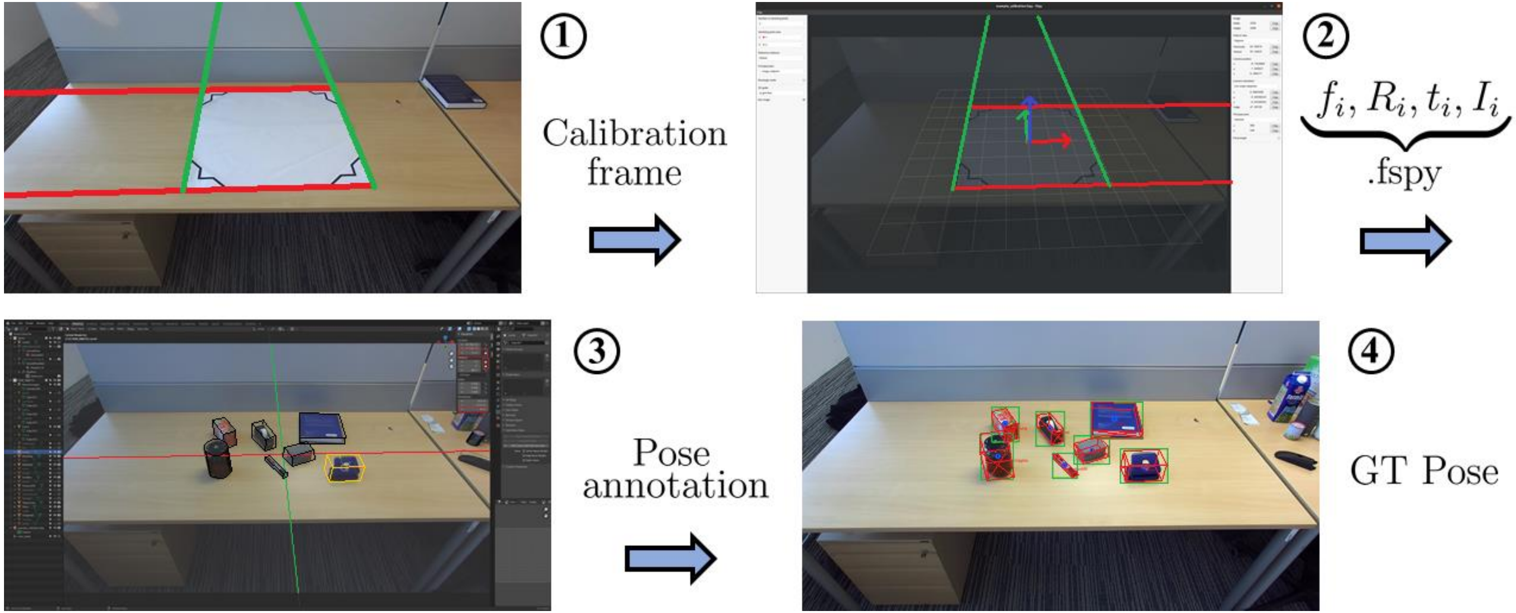
LEARNING 3D OBJECT DETECTION

- Objects as points (CenterNet [4]) extended for 3D detection



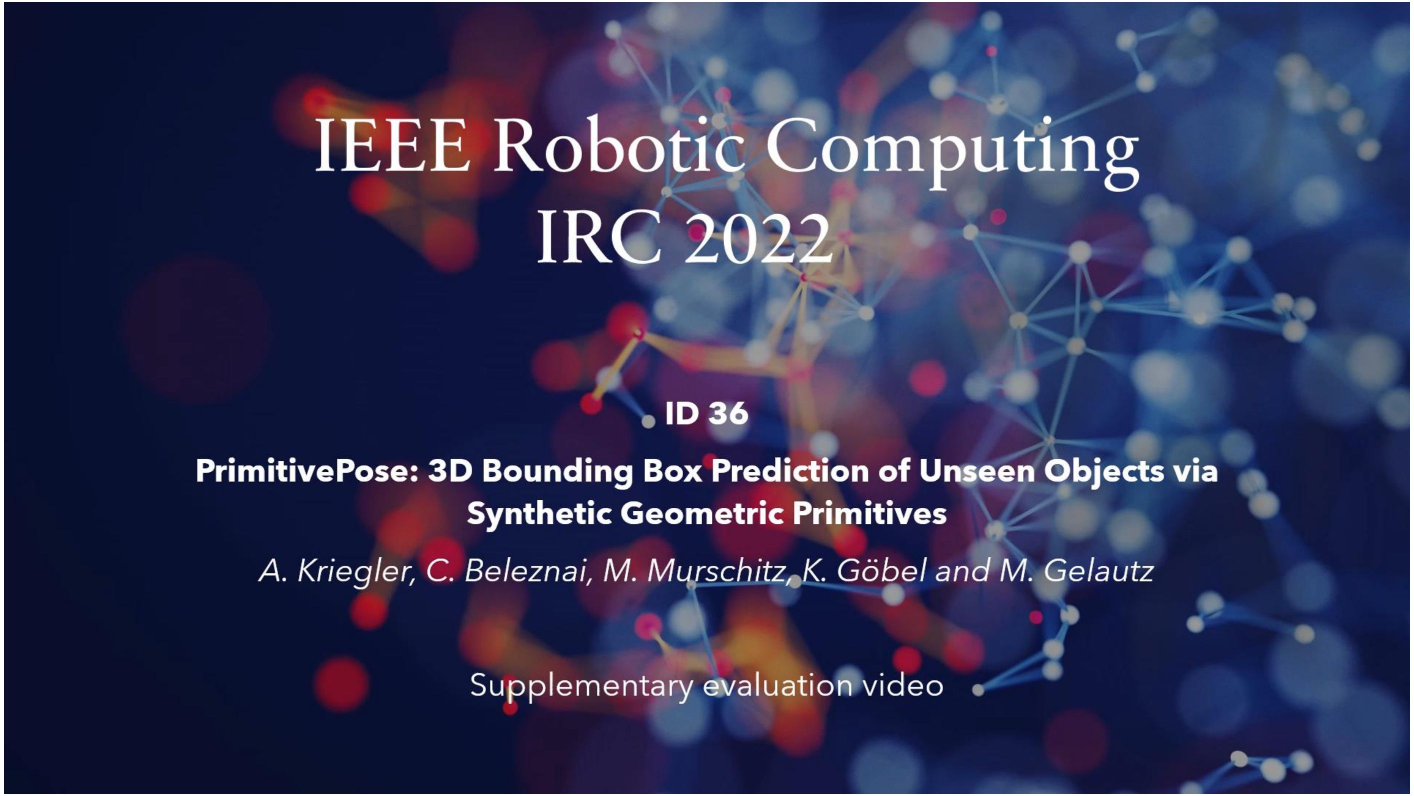
REAL 3D OBJECT POSE ANNOTATION

- Annotate 6DoF pose of arbitrary objects from RGB images (uncalibrated)



https://github.com/vindicate-git/PrimitivePose-Annotation-Tool

UNSEEN OBJECT 3D DETECTION



IEEE Robotic Computing IRC 2022

ID 36

PrimitivePose: 3D Bounding Box Prediction of Unseen Objects via Synthetic Geometric Primitives

A. Kriegler, C. Beleznai, M. Murschitz, K. Göbel and M. Gelautz

Supplementary evaluation video

RESULTS ON OUR TABLETOP DATASET

- Correct pose if rotational error e_{rot} [5] is below a certain threshold
- Compare to PoseCNN [6], very popular method trained on YCB objects
- Also report 3D IoU of 3D bounding boxes

Method	Large	Mixed	Small	Large	Mixed	Small	Large	Mixed	Small	Large	Mixed	Small	Large	Mixed	Small	Large	Mixed	Small
	Recall ($e_{rot} < 2^\circ$)			Recall ($e_{rot} < 5^\circ$)			Recall ($e_{rot} < 10^\circ$)			Recall ($e_{rot} < 15^\circ$)			Recall ($e_{rot} < 25^\circ$)			Recall ($e_{rot} < 40^\circ$)		
PoseCNN [5]	0.0	0.0	0.0	0.0	1.2	0.0	2.5	1.6	0.8	3.7	2.3	1.1	4.4	3.7	4.1	7.4	9.3	10.6
δ	**27.5**	**22.5**	**47.7**	**38.6**	32.8	**55.6**	**45.5**	39.2	**60.3**	49.9	43.6	**62.7**	53.8	47.1	**65.0**	57.1	49.9	**67.4**
$\mathcal{N}_s$	19.1	14.6	29.1	33.6	26.4	40.0	42.3	35.7	45.0	47.5	40.8	48.0	51.7	44.9	51.2	55.5	48.2	54.0
$\mathcal{N}_{\delta-s}$	22.0	19.3	29.7	37.2	**34.5**	43.4	45.3	**42.1**	50.4	**50.7**	**46.9**	54.6	**55.8**	**50.8**	57.6	**59.8**	**54.2**	60.5
	IoU_{3D} ($d_{tol} = 0\%$)			IoU_{3D} ($d_{tol} = 4\%$)			IoU_{3D} ($d_{tol} = 8\%$)			IoU_{3D} ($d_{tol} = 12\%$)			IoU_{3D} ($d_{tol} = 16\%$)			IoU_{3D} ($d_{tol} = 20\%$)		
δ	**15.3**	**12.2**	**5.0**	**20.0**	**16.7**	**8.6**	**24.2**	**20.5**	**12.4**	**27.3**	**23.0**	**15.5**	**29.5**	**24.7**	**17.6**	**31.0**	**25.9**	**19.0**
$\mathcal{N}_s$	6.8	1.8	0.2	9.1	3.0	0.6	11.7	4.7	1.5	14.4	6.6	3.0	16.8	8.5	4.7	18.9	10.2	6.1
$\mathcal{N}_{\delta-s}$	12.4	7.0	2.4	15.9	10.2	4.0	19.2	13.4	6.4	21.9	16.3	8.7	24.0	18.6	10.6	25.7	20.3	12.3

- Quantitatively, simple disparity δ is best model input
- Disparity-scaled surface normals $\mathcal{N}_{\delta-s}$ are second -> depth information

RESULTS ON OUR TABLETOP DATASET

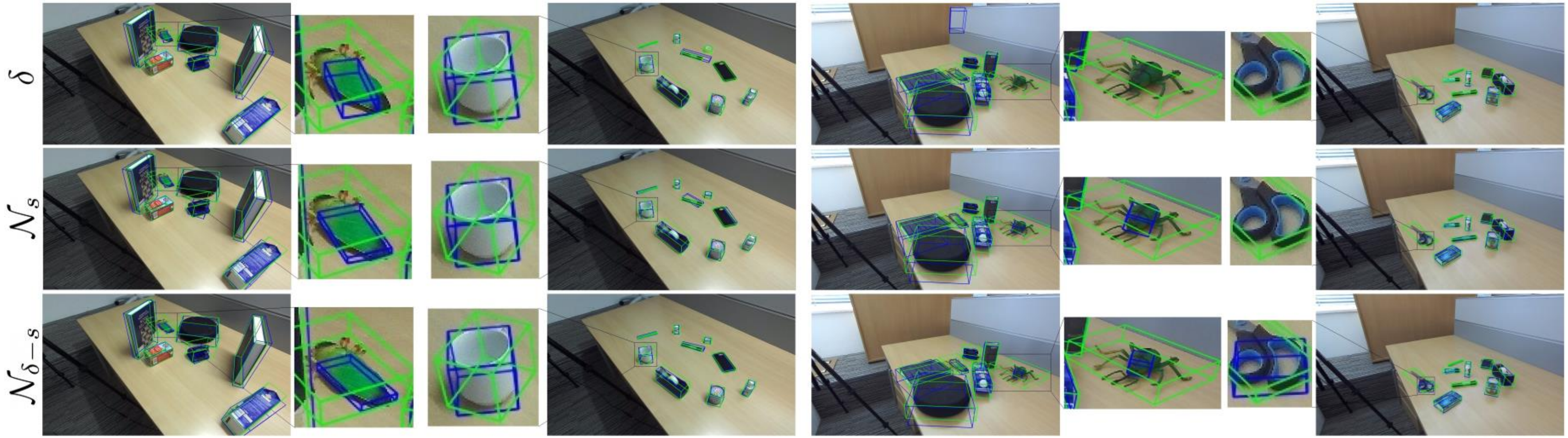
- Correct pose if rotational error e_{rot} [5] is below a certain threshold
- Compare to PoseCNN [6], very popular method trained on YCB objects
- Also report 3D IoU of 3D bounding boxes

Method	Large	Mixed	Small	Large	Mixed	Small	Large	Mixed	Small	Large	Mixed	Small	Large	Mixed	Small	Large	Mixed	Small
	Recall ($e_{rot} < 2°$)			Recall ($e_{rot} < 5°$)			Recall ($e_{rot} < 10°$)			Recall ($e_{rot} < 15°$)			Recall ($e_{rot} < 25°$)			Recall ($e_{rot} < 40°$)		
PoseCNN [5]	0.0	0.0	0.0	0.0	1.2	0.0	2.5	1.6	0.8	3.7	2.3	1.1	4.4	3.7	4.1	7.4	9.3	10.6
δ	**27.5**	**22.5**	**47.7**	**38.6**	32.8	**55.6**	**45.5**	39.2	**60.3**	49.9	43.6	**62.7**	53.8	47.1	**65.0**	57.1	49.9	**67.4**
$\mathcal{N}_s$	19.1	14.6	29.1	33.6	26.4	40.0	42.3	35.7	45.0	47.5	40.8	48.0	51.7	44.9	51.2	55.5	48.2	54.0
$\mathcal{N}_{\delta-s}$	22.0	19.3	29.7	37.2	**34.5**	43.4	45.3	**42.1**	50.4	**50.7**	**46.9**	54.6	**55.8**	**50.8**	57.6	**59.8**	**54.2**	60.5
	$\mathrm{IoU_{3D}}$ ($d_{tol} = 0\%$)			$\mathrm{IoU_{3D}}$ ($d_{tol} = 4\%$)			$\mathrm{IoU_{3D}}$ ($d_{tol} = 8\%$)			$\mathrm{IoU_{3D}}$ ($d_{tol} = 12\%$)			$\mathrm{IoU_{3D}}$ ($d_{tol} = 16\%$)			$\mathrm{IoU_{3D}}$ ($d_{tol} = 20\%$)		
δ	**15.3**	**12.2**	**5.0**	**20.0**	**16.7**	**8.6**	**24.2**	**20.5**	**12.4**	**27.3**	**23.0**	**15.5**	**29.5**	**24.7**	**17.6**	**31.0**	**25.9**	**19.0**
$\mathcal{N}_s$	6.8	1.8	0.2	9.1	3.0	0.6	11.7	4.7	1.5	14.4	6.6	3.0	16.8	8.5	4.7	18.9	10.2	6.1
$\mathcal{N}_{\delta-s}$	12.4	7.0	2.4	15.9	10.2	4.0	19.2	13.4	6.4	21.9	16.3	8.7	24.0	18.6	10.6	25.7	20.3	12.3

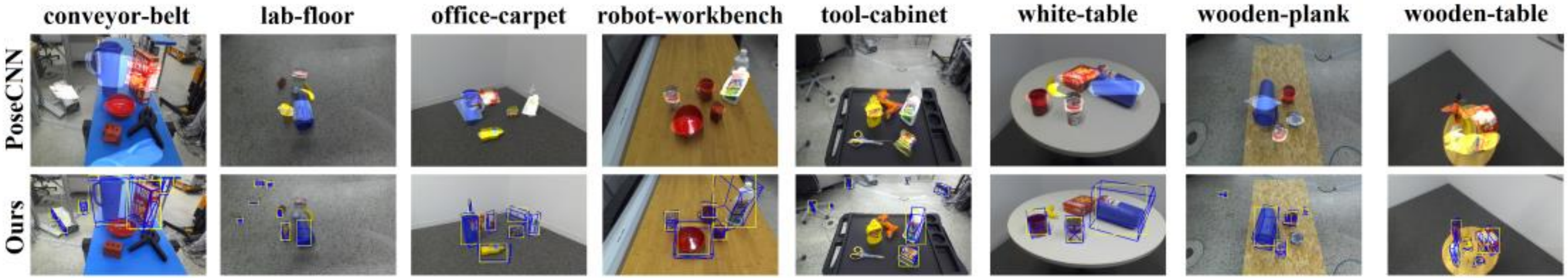
- Quantitatively, simple disparity δ is best model input
- Disparity-scaled surface normals $\mathcal{N}_{\delta-s}$ are second -> depth information

FAILURE CASES

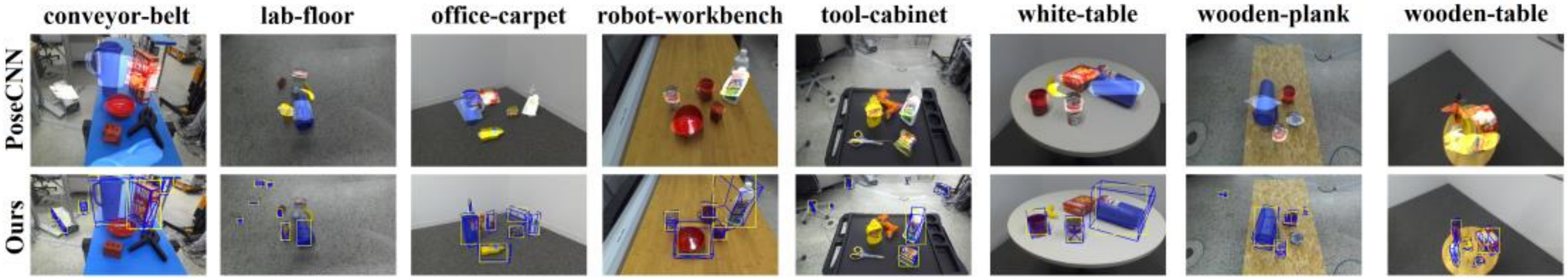
- Non-compact objects
- Symmetry due to self-occlusion
- Object cavities



RESULTS ON STIOS DATASET [7]

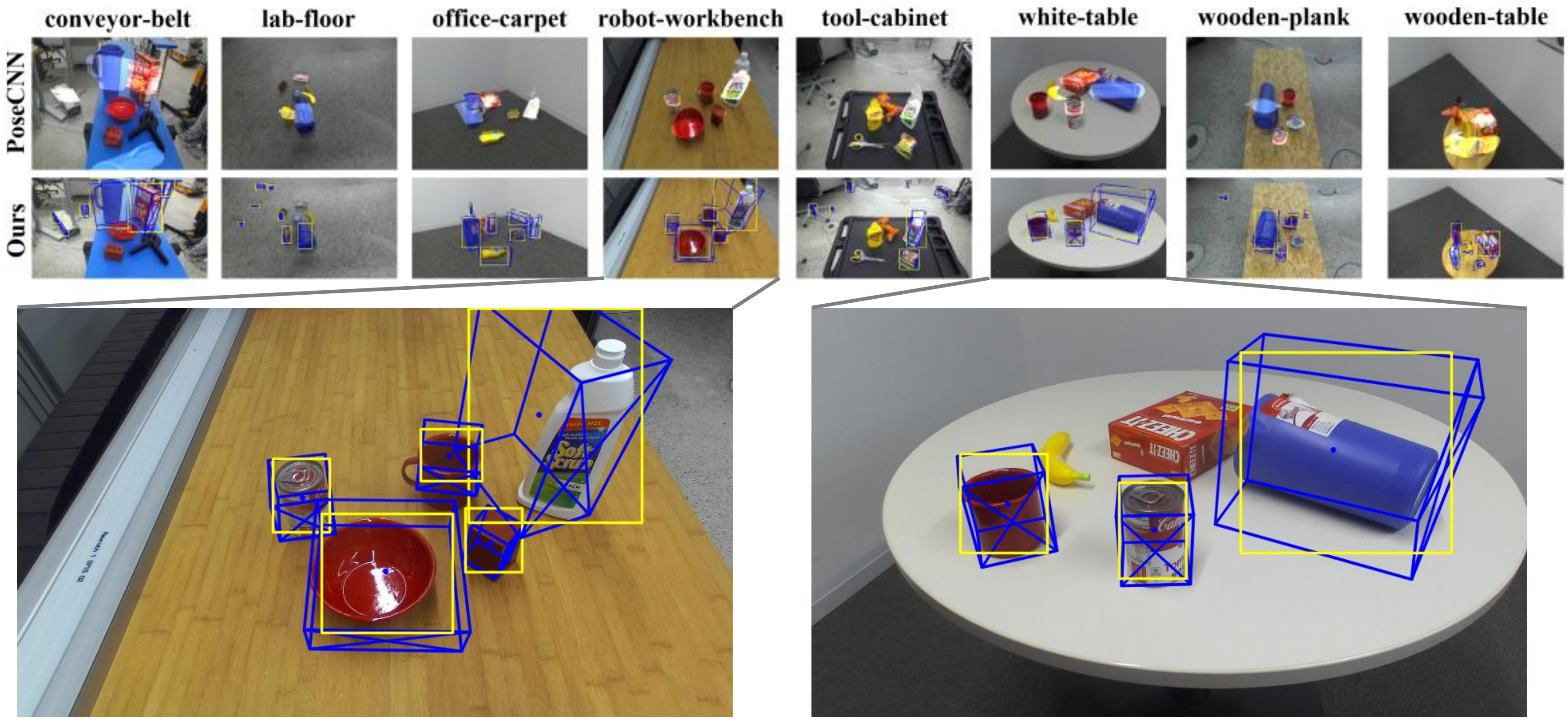


RESULTS ON STIOS DATASET [7]

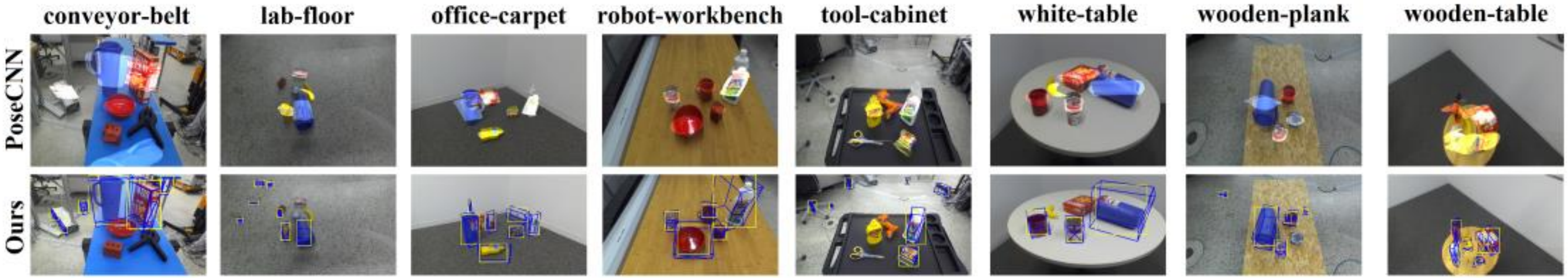


- 8 different environments with YCB objects
- Different object configurations
- Provides stereo images and semantic masks but no pose annotations

RESULTS ON STIOS DATASET [7]



RESULTS ON STIOS DATASET [7]

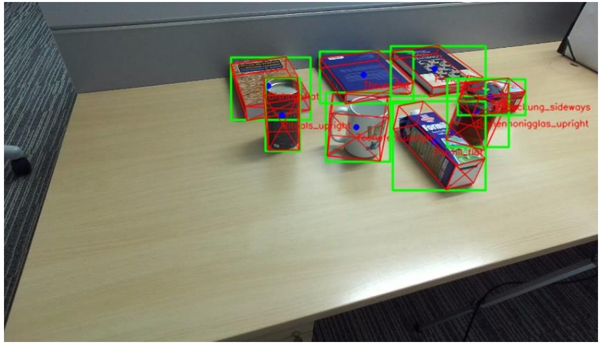


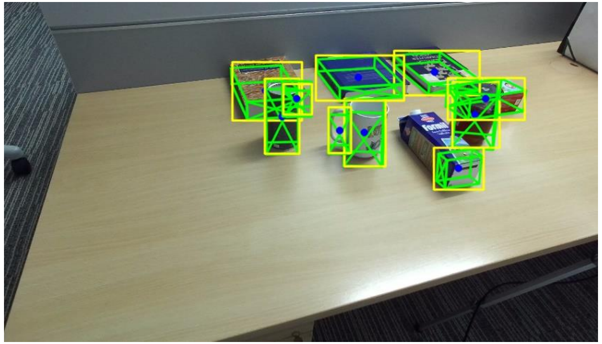
- 8 different environments with YCB objects
- Different object configurations
- Provides stereo images and semantic masks but no pose annotations

- **Problems with maximum disparity**
- **Acceptable results but dataset transfer is always hard**

MAIN CONTRIBUTIONS

- Stereo-based, real-time 3D object detection
- Learned from synthetic data, no pose-annotated images required
- Generalizes reasonably to other environments
- Toolkit for pose annotation of objects from monocular, uncalibrated camera images





REFERENCES

[1] A. Murali, A. Mousavian, C. Eppner, C. Paxton and D. Fox., "6-DOF Grasping for Target-driven Object Manipulation in Clutter," *International Conference on Robotics and Automation (ICRA)*, pp. 6232-6238, 2020.

[2] H. Xu and J. Zhang, "AANet: Adaptive Aggregation Network for Efficient Stereo Matching," *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1959-1968, 2020.

[3] A. Kriegler, C. Beleznai and M. Gelautz, "Evaluation of Monocular and Stereo Depth Data for Geometry-Assisted Learning of 3D Pose," *OAGM Workshop 2021*, pp. 1-7, 2021.

[4] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang and Q. Tian, "CenterNet: Keypoing Triplets for Object Detection," *International Conference on Computer Vision (ICCV)*, pp. 6569-6578, 2019.

[5] T. Hodan, J. Matas and S. Obdrzalek, "On Evaluation of 6D Object Pose Estimation," *European Conference on Computer Vision Workshops (ECCVW)*, pp. 609-619, 2016.

[6] Y. Xiang, T. Schmidt, V. Narayanan and D. Fox, "PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation in Cluttered Scenes," *Robotics: Science and Systems XIV (RSS)*, 2018.

[7] M. Durner and W. Boerdijk, "Stereo Instances on Surfaces (STIOS," [Online]. Available: https://zenodo.org/record/4706907

THANK YOU

Andreas Kriegler, Csaba Beleznai, Markus Murschitz, Kai Göbel,
Margrit Gelautz 2022-12-05