CZECH TECHNICAL UNIVERSITY IN PRAGUE

Pose estimation of specific rigid objects

Tomáš Hodaň

supervised by Prof. Jiří Matas

PhD defense, 7. 7. 2021

Bio

Google Scholar citations: **650**, h-index: **8**

BSc. and MSc. (with honours) in computer science (2013)
– Brno University of Technology
– Utrecht University, The Netherlands (one year)

PhD in computer vision (2021, viva today)
– Czech Technical University in Prague
– Supervisor: Prof. Jiří Matas

Intern at Microsoft Research, Redmond (2018, 3 months)
– Working with Sudipta N. Sinha, Vibhav Vineet, and Brian Guenter
– Topic: Photorealistic image synthesis for object detection

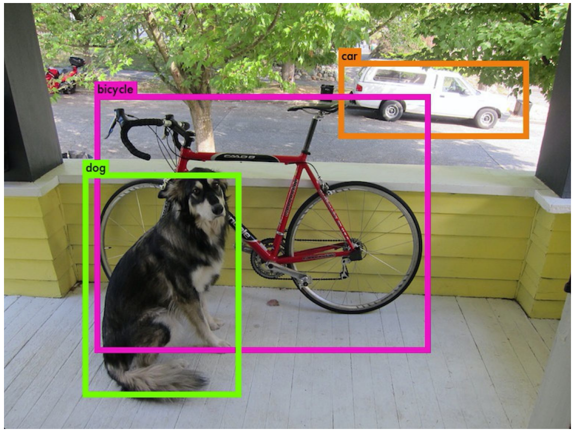
Intern at Google, Munich (2019, 6 months)
– Working with Stefan Hinterstoisser
– Topic: Fine-grained object detection

Research scientist at Facebook Reality Labs, Redmond (2020)
– Working with Cem Keskin and Robert Wang
– Topic: Hand-object pose estimation

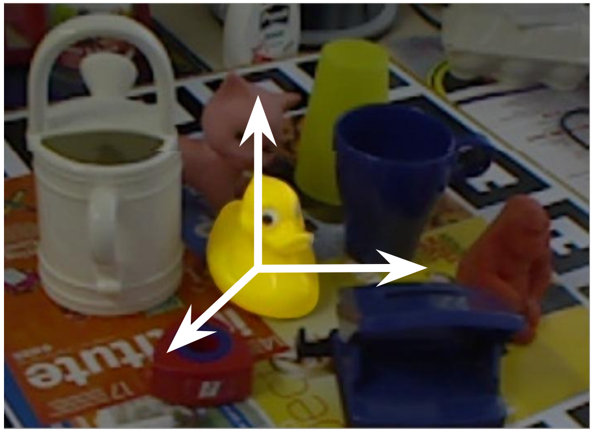
Object pose estimation

Objects in computer vision tasks:

Image classification



2D object detection

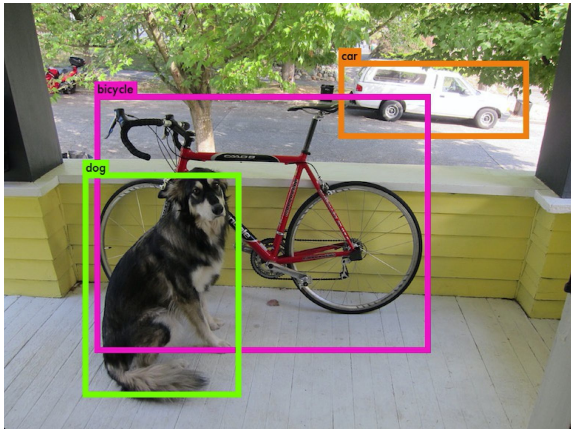


6D object pose estimation

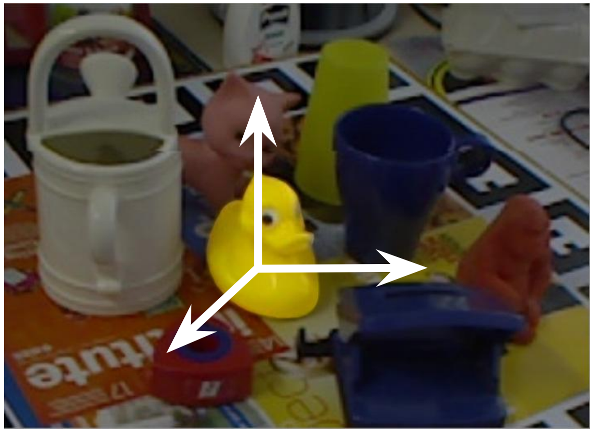
Object pose estimation

Objects in computer vision tasks:

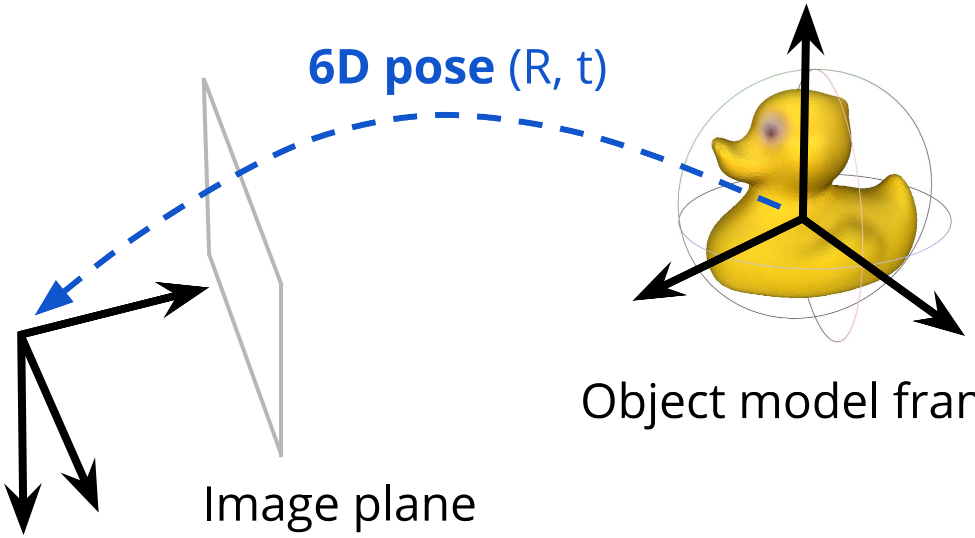
Image classification



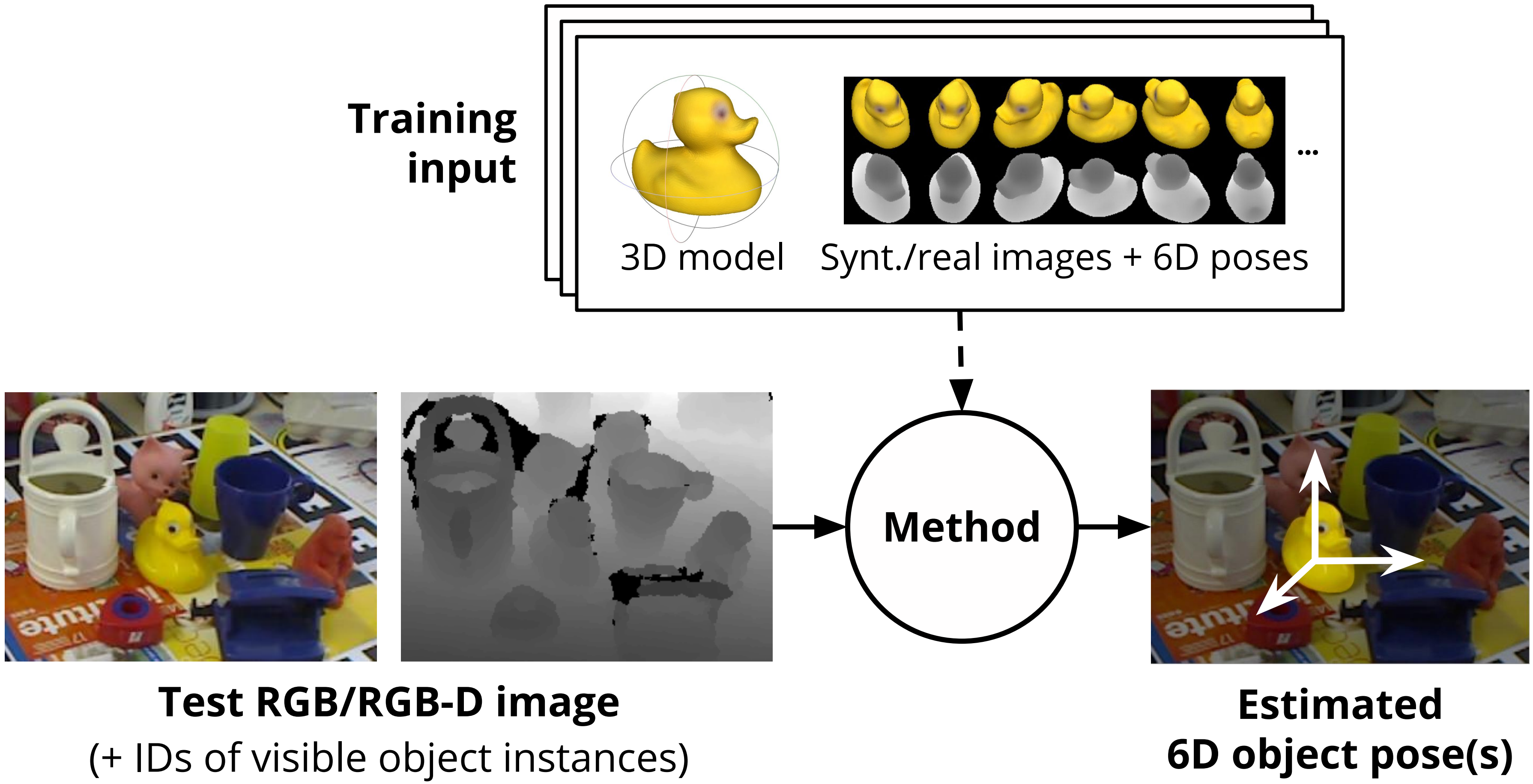
2D object detection



6D object pose estimation



Object pose estimation: input & output

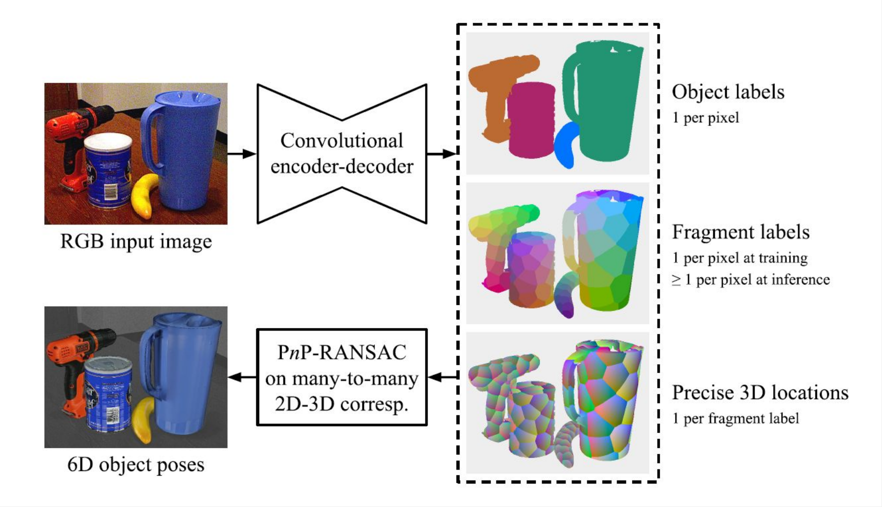


Two variants:

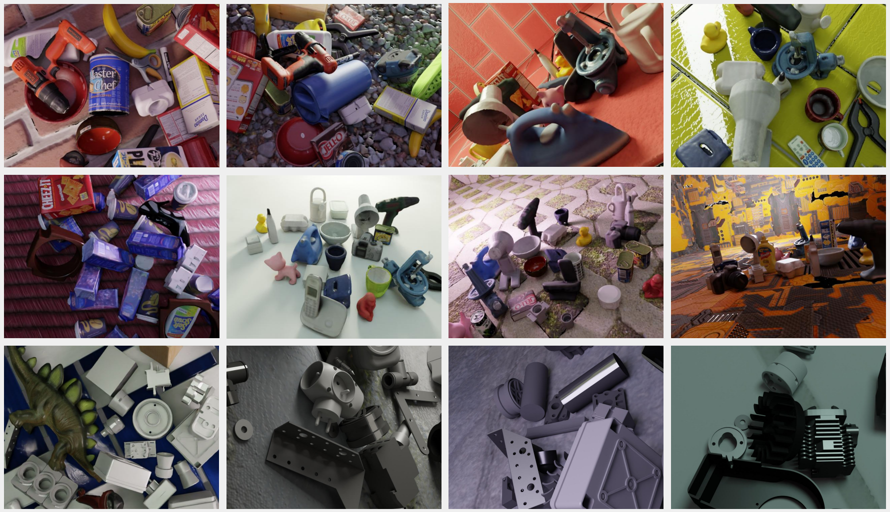
1. **6D object localization** – IDs of instances visible in the image provided (our focus).
2. **6D object detection** – No information about visible instances provided.

Main contributions of the thesis

EPOS, HashMatch: Pose estimation methods



ObjectSynth: Photorealistic image synthesis



T-LESS: Dataset with texture-less objects

BOP: Benchmark for 6D object pose estimation

#	Method	Year	PPF	CNN	...models	Train. im.	...type	Test im.	Refine.	Avg.	LM-O	T-LESS	TUD-L	IC-BIN	ITODD	HB	YCB-V	Time
1	CosyPose-ECCV20-Synt+Real-1View-ICP	2020	No	Yes	3/dataset	RGB	Synt+real	RGB-D	RGB+ICP	0.698	0.714	0.701	0.939	0.647	0.313	0.712	0.861	13.743
2	Koenig-Hybrid-DL-PointPairs	2020	Yes	Yes	1/dataset	RGB	Synt+real	RGB-D	ICP	0.639	0.631	0.655	0.920	0.430	0.483	0.651	0.701	0.633
3	CosyPose-ECCV20-Synt+Real-1View	2020	No	Yes	3/dataset	RGB	Synt+real	RGB	RGB	0.637	0.633	0.728	0.823	0.583	0.216	0.656	0.821	0.449
4	Pix2Pose-BOP20_w/ICP-ICCV19	2020	No	Yes	1/object	RGB	Synt+real	RGB-D	ICP	0.591	0.588	0.512	0.820	0.390	0.351	0.695	0.780	4.844
5	CosyPose-ECCV20-PBR-1View	2020	No	Yes	3/dataset	RGB	PBR only	RGB	RGB	0.570	0.633	0.640	0.685	0.583	0.216	0.656	0.574	0.475
6	Vidal-Sensors18	2019	Yes	No	-	-	-	D	ICP	0.569	0.582	0.538	0.876	0.393	0.435	0.706	0.450	3.220
7	CDPNv2_BOP20 (RGB-only & ICP)	2020	No	Yes	1/object	RGB	Synt+real	RGB-D	ICP	0.568	0.630	0.464	0.913	0.450	0.186	0.712	0.619	1.462
8	Drost-CVPR10-Edges	2019	Yes	No	-	-	-	RGB-D	ICP	0.550	0.515	0.500	0.851	0.368	0.570	0.671	0.375	87.568
9	CDPNv2_BOP20 (PBR-only & ICP)	2020	No	Yes	1/object	RGB	PBR only	RGB-D	ICP	0.534	0.630	0.435	0.791	0.450	0.186	0.712	0.532	1.491
10	CDPNv2_BOP20 (RGB-only)	2020	No	Yes	1/object	RGB	Synt+real	RGB	No	0.529	0.624	0.478	0.772	0.473	0.102	0.722	0.532	0.935
11	Drost-CVPR10-3D-Edges	2019	Yes	No	-	-	-	D	ICP	0.500	0.469	0.404	0.852	0.373	0.462	0.623	0.316	80.055
12	Drost-CVPR10-3D-Only	2019	Yes	No	-	-	-	D	ICP	0.487	0.527	0.444	0.775	0.388	0.316	0.615	0.344	7.704
13	CDPN_BOP19 (RGB-only)	2020	No	Yes	1/object	RGB	Synt+real	RGB	No	0.479	0.569	0.490	0.769	0.327	0.067	0.672	0.457	0.480
14	CDPNv2_BOP20 (PBR-only&RGB-only)	2020	No	Yes	1/object	RGB	PBR only	RGB	No	0.472	0.624	0.407	0.588	0.473	0.102	0.722	0.390	0.978
15	leaping from 2D to 6D	2020	No	Yes	1/object	RGB	Synt+real	RGB	No	0.471	0.525	0.403	0.751	0.342	0.077	0.658	0.543	0.425
16	EPOS-BOP20-PBR	2020	No	Yes	1/dataset	RGB	PBR only	RGB	No	0.457	0.547	0.467	0.558	0.363	0.186	0.580	0.499	1.874
17	Drost-CVPR10-3D-Only-Faster	2019	Yes	No	-	-	-	D	ICP	0.454	0.492	0.405	0.696	0.377	0.274	0.603	0.330	1.383
18	Félix&Neves-ICRA2017-IET2019	2019	Yes	Yes	1/dataset	RGB-D	Synt+real	RGB-D	ICP	0.412	0.394	0.212	0.851	0.323	0.069	0.529	0.510	55.780
19	Sundermeyer-IJCV19+ICP	2019	No	Yes	1/object	RGB	Synt+real	RGB-D	ICP	0.398	0.237	0.487	0.614	0.281	0.158	0.506	0.505	0.865
20	Zhigang-CDPN-ICCV19	2019	No	Yes	1/object	RGB	Synt+real	RGB	No	0.353	0.374	0.124	0.757	0.257	0.070	0.470	0.422	0.513
21	PointVoteNet2	2020	No	Yes	1/object	RGB-D	PBR only	RGB-D	ICP	0.351	0.653	0.004	0.673	0.264	0.001	0.556	0.308	-
22	Pix2Pose-BOP20-ICCV19	2020	No	Yes	1/object	RGB	Synt+real	RGB	No	0.342	0.363	0.344	0.420	0.226	0.134	0.446	0.457	1.215
23	Sundermeyer-IJCV19	2019	No	Yes	1/object	RGB	Synt+real	RGB	No	0.270	0.146	0.304	0.401	0.217	0.101	0.346	0.377	0.186
24	SingleMultiPathEncoder-CVPR20	2020	No	Yes	1/all	RGB	Synt+real	RGB	No	0.241	0.217	0.310	0.334	0.175	0.067	0.293	0.289	0.186
25	Pix2Pose-BOP19-ICCV19	2019	No	Yes	1/object	RGB	Synt+real	RGB	No	0.205	0.077	0.275	0.349	0.215	0.032	0.200	0.290	0.793
26	DPOD (synthetic)	2019	No	Yes	1/scene	RGB	Synt	RGB	No	0.161	0.169	0.081	0.242	0.130	0.000	0.286	0.222	0.231

EPOS: Estimating 6D pose of objects with symmetries

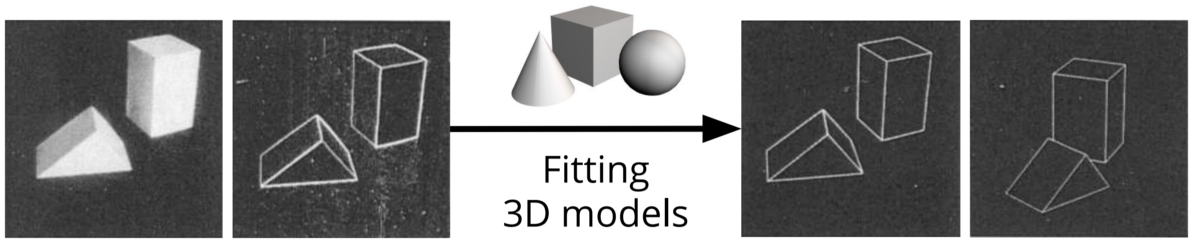
Hodaň, Baráth, Matas

CVPR 2020

Related work: Traditional methods

Fitting 3D models to edge maps *(Roberts'63, Lowe'91)*

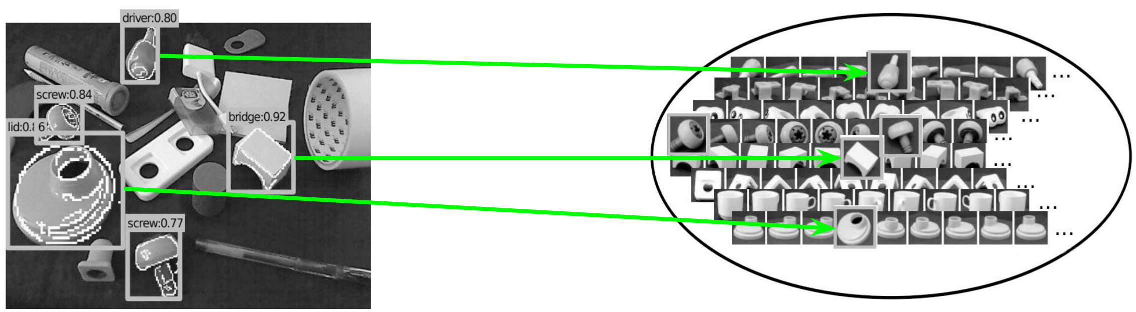
Input image,
detected edges



Two views at
a 3D polyhedral
description

Template matching *(Brunelli'09, Hinterstoisser'12, Hodan'15)*

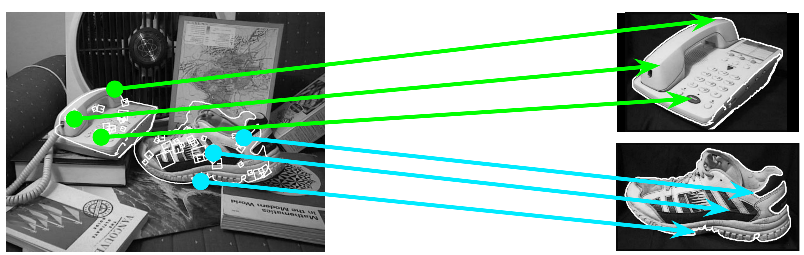
Input image
(with detections)



Templates
annotated
with 6D poses

Correspondence-based *(Lowe'99, Collet'11, Drost'10, Brachmann'14)*

Input image



Object
templates

Related work: CNN-based methods

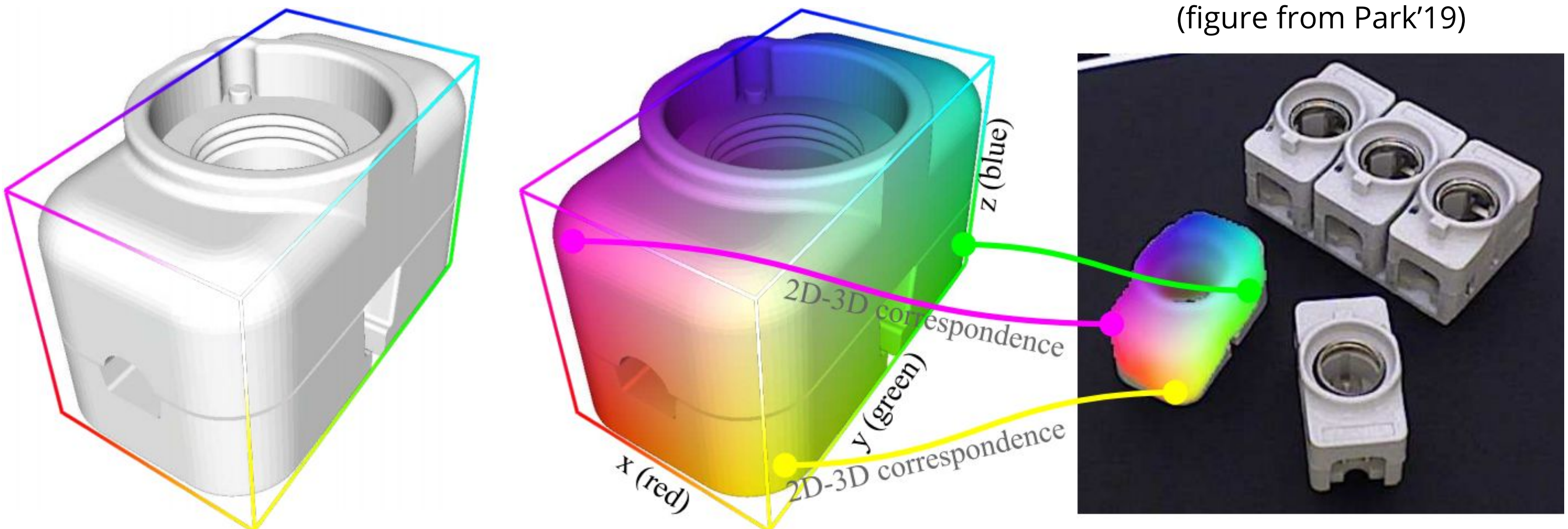
Extending 2D object detection/segmentation with pose prediction:

- Classification into discrete viewpoints *(Kehl'17, Sundermeyer'18)*
- Regression of pose/viewpoint *(Xiang'17, Li'18, Wang'19)*

Predicting 2D-3D correspondences + P*n*P-RANSAC

(Rad and Lepetit'17, Tekin'18, Peng'19, Jafari'18, Zakharov'19, Peng'19, Park'19, ...)

(figure from Park'19)



3D object model

Visualization:
Color codes of locations in the 3D object model frame

Input image with ground-truth correspondences

Related work: 2D-3D correspondences

3D model

Input image
with highlighted
target objects

Approach 1: Predicting 2D projections of 3D keypoints

(Rad'17: BB8, Tekin'18: YOLO-6D, Peng'19: PVNet, ...)

Related work: 2D-3D correspondences

3D model

A fixed set of
3D keypoints

Input image
with highlighted
target objects

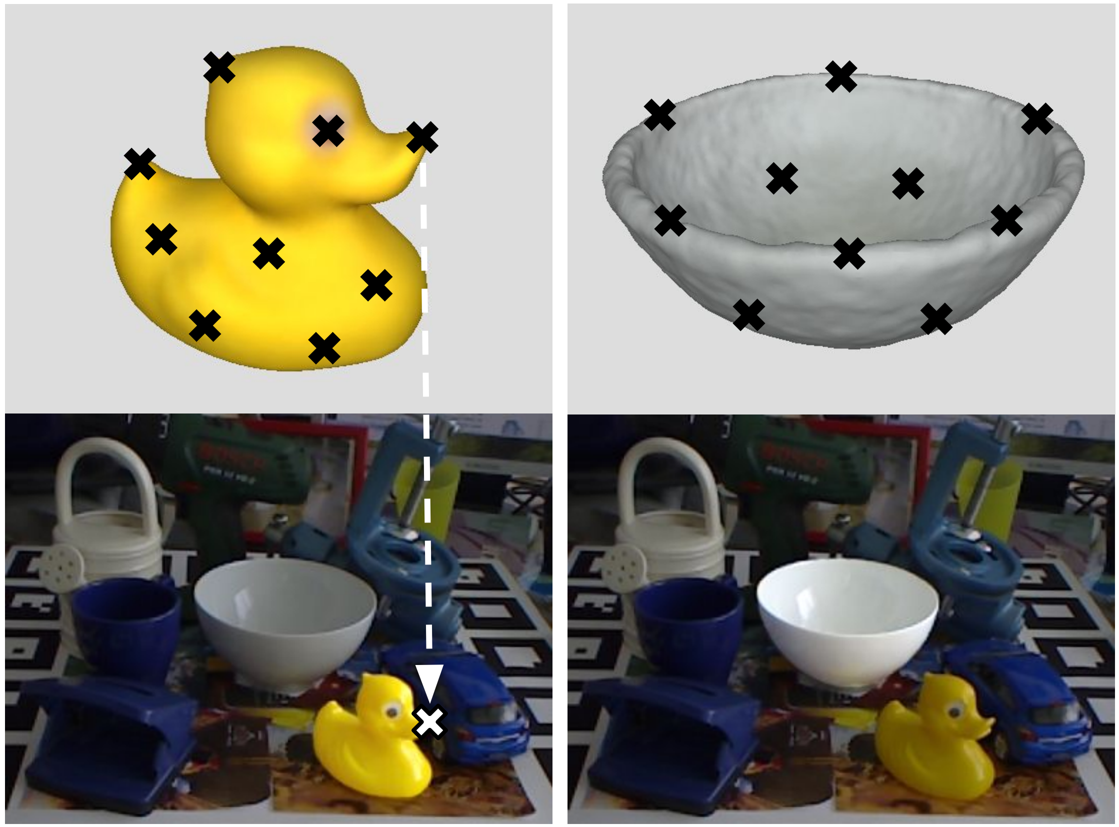
Approach 1: Predicting 2D projections of 3D keypoints

(Rad'17: BB8, Tekin'18: YOLO-6D, Peng'19: PVNet, ...)

Related work: 2D-3D correspondences

3D model

Input image
with highlighted
target objects

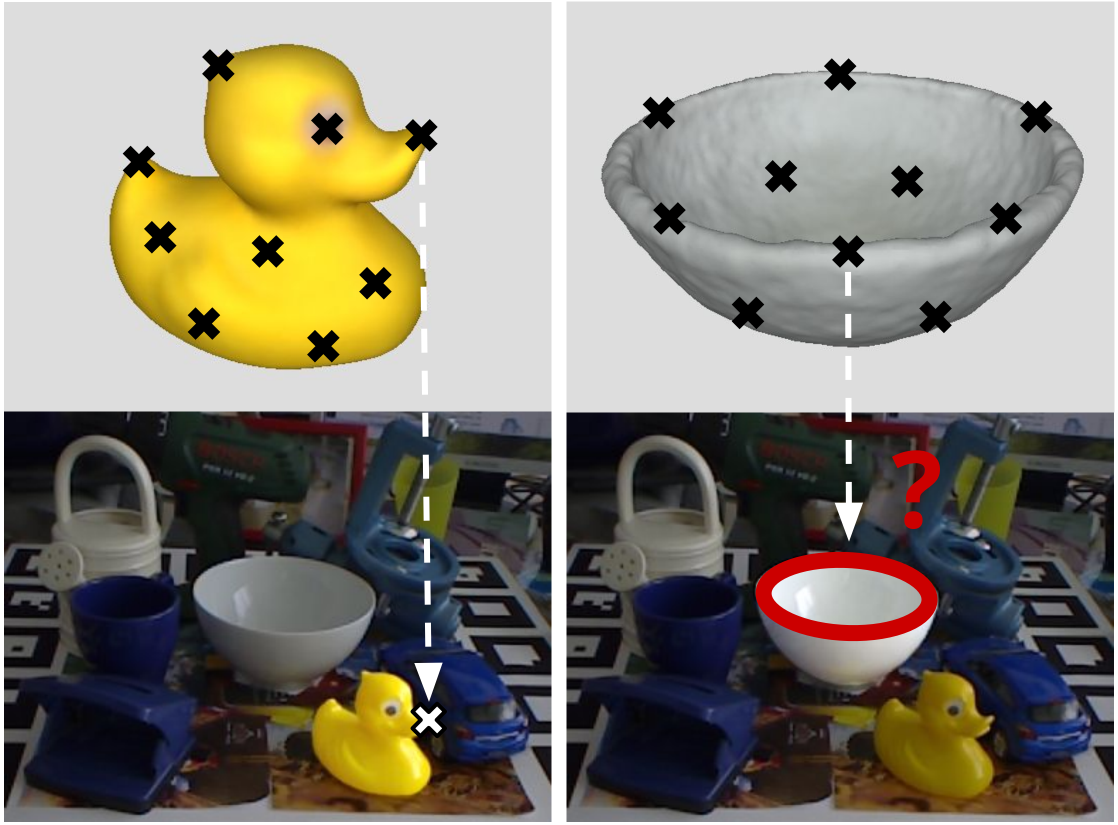


A fixed set of
3D keypoints

Approach 1: Predicting 2D projections of 3D keypoints

(Rad'17: BB8, Tekin'18: YOLO-6D, Peng'19: PVNet, ...)

Related work: 2D-3D correspondences



Approach 1: Predicting 2D projections of 3D keypoints
(Rad'17: BB8, Tekin'18: YOLO-6D, Peng'19: PVNet, ...)

In case of symmetries, methods **compromise among possible 2D locations** or consider **only the most confident one.**

Related work: 2D-3D correspondences

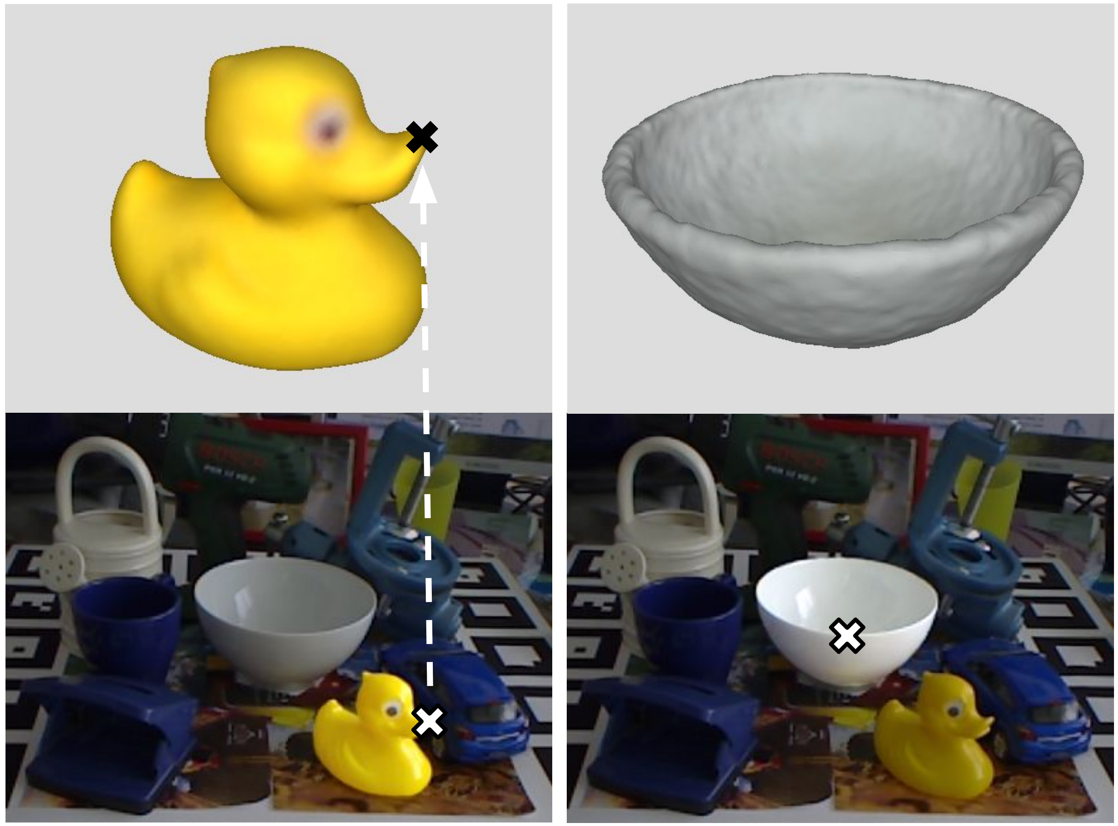
Approach 2: Predicting 3D coordinates at each pixel

(Brachmann'14, Nigam'18, Jafari'18: iPose, Zakharov'19: DPOD, ...*)*

Related work: 2D-3D correspondences

3D model

Input image
with highlighted
target objects



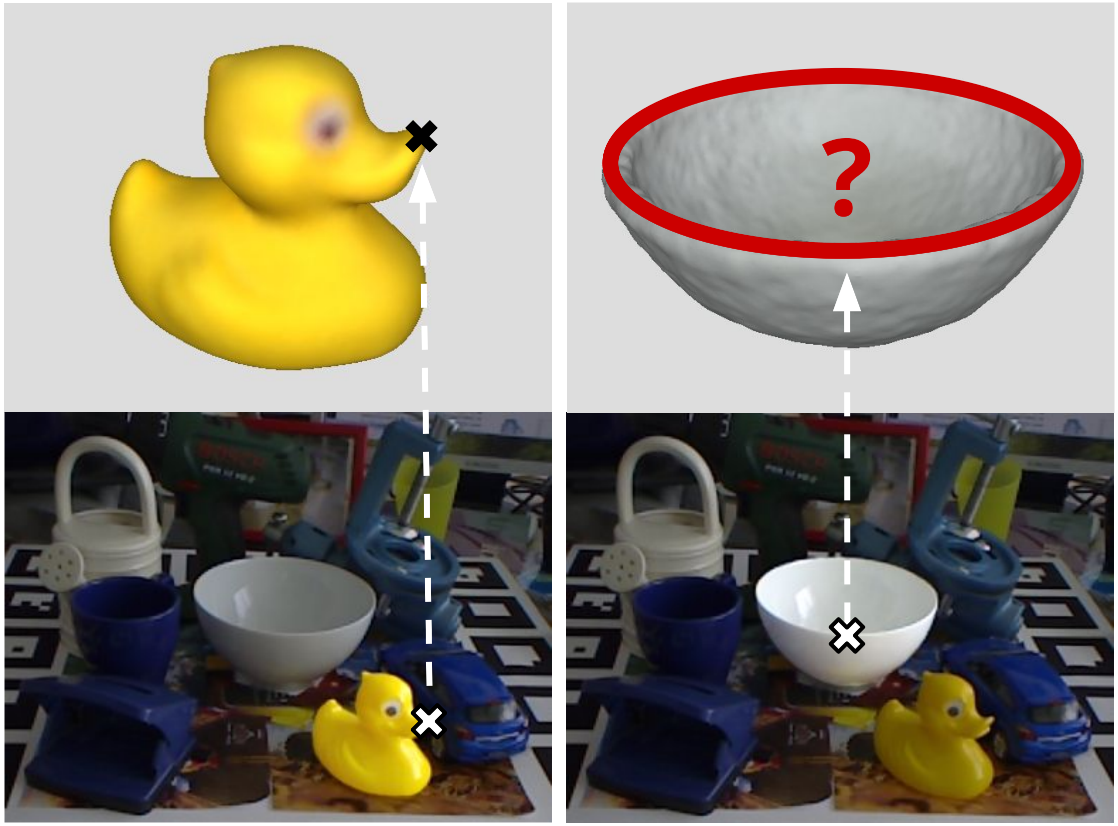
Approach 2: Predicting 3D coordinates at each pixel

(Brachmann'14, Nigam'18, Jafari'18: iPose, Zakharov'19: DPOD, ...)

Related work: 2D-3D correspondences

3D model

Input image
with highlighted
target objects



Approach 2: Predicting 3D coordinates at each pixel
(Brachmann'14, Nigam'18, Jafari'18: iPose, Zakharov'19: DPOD, ...)

In case of symmetries, methods **compromise among possible 3D locations** or consider **only the most confident one.**

EPOS: Object represented by surface fragments

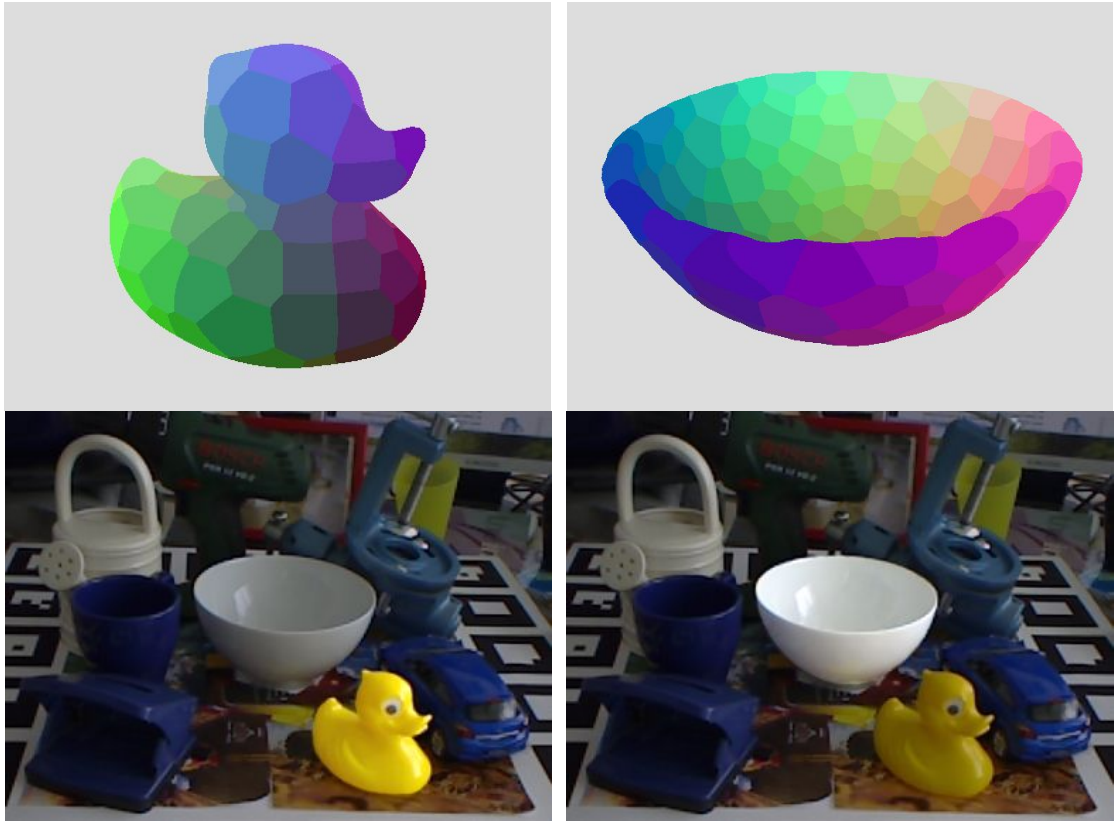
3D model

Input image
with highlighted
target objects

EPOS: Object represented by surface fragments

Surface fragments

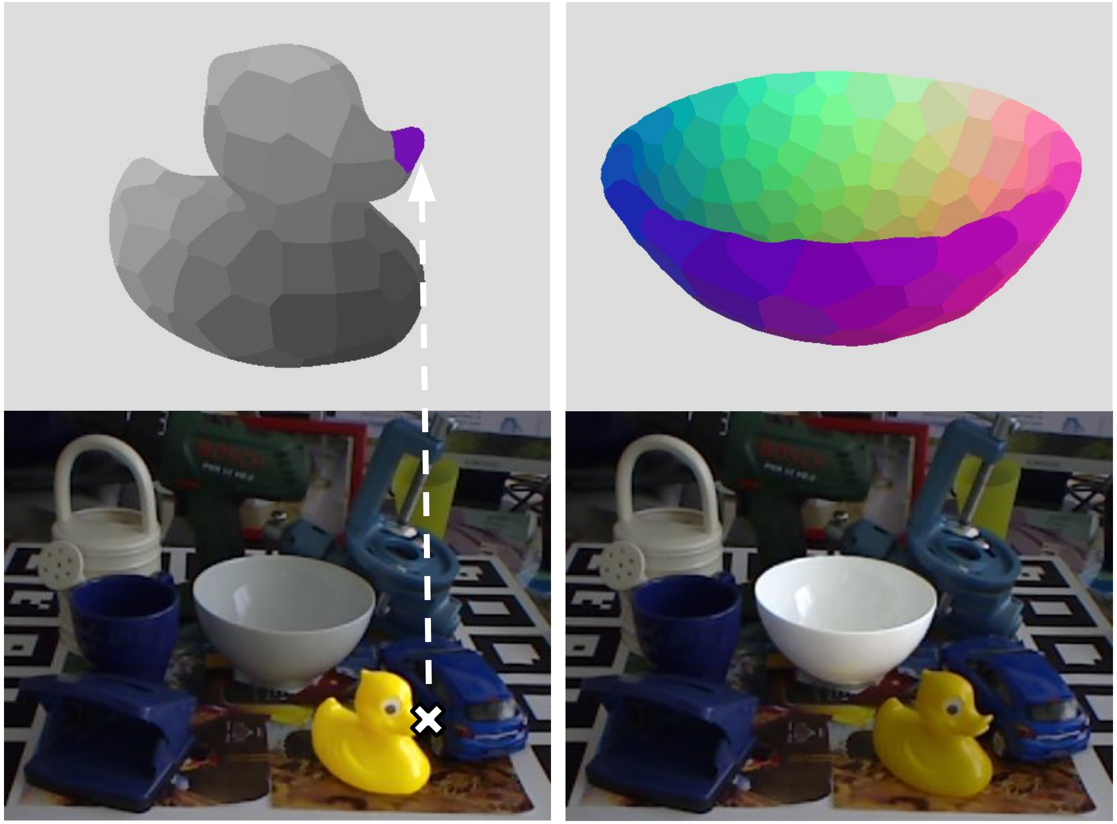
Input image with highlighted target objects



EPOS: Multiple potential 2D-3D correspondences per pixel

Surface fragments

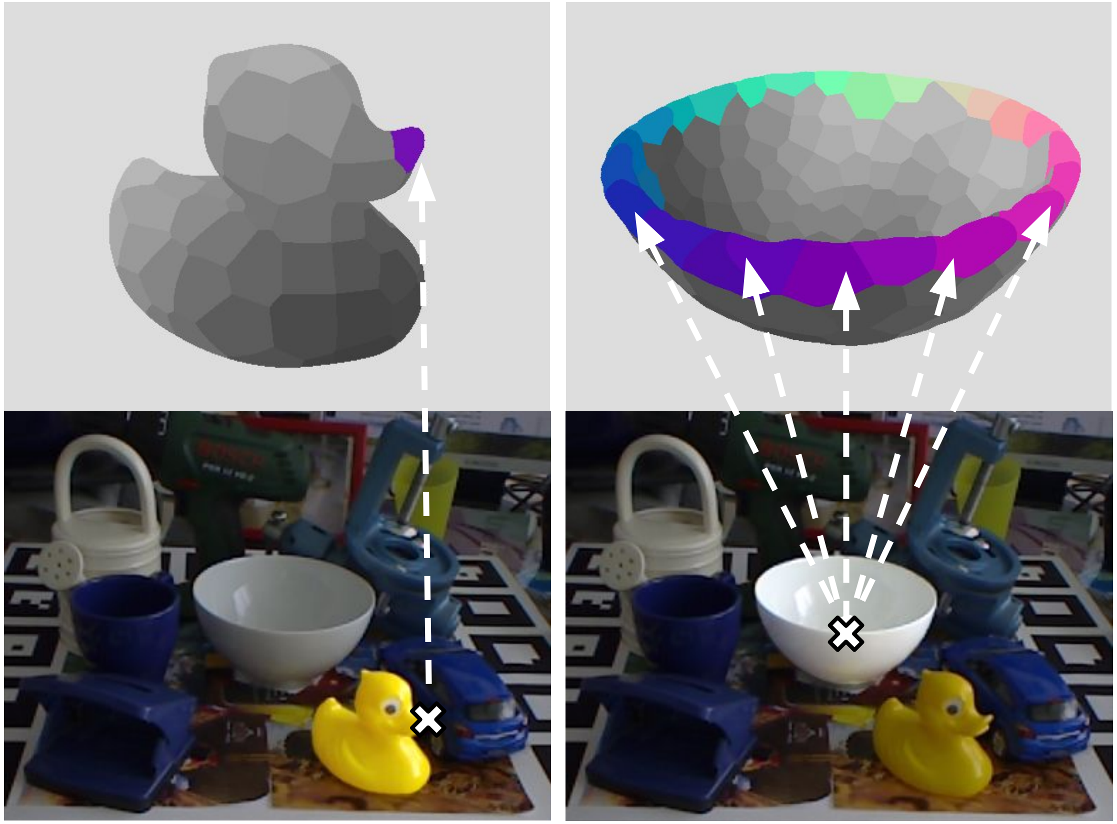
Input image with highlighted target objects



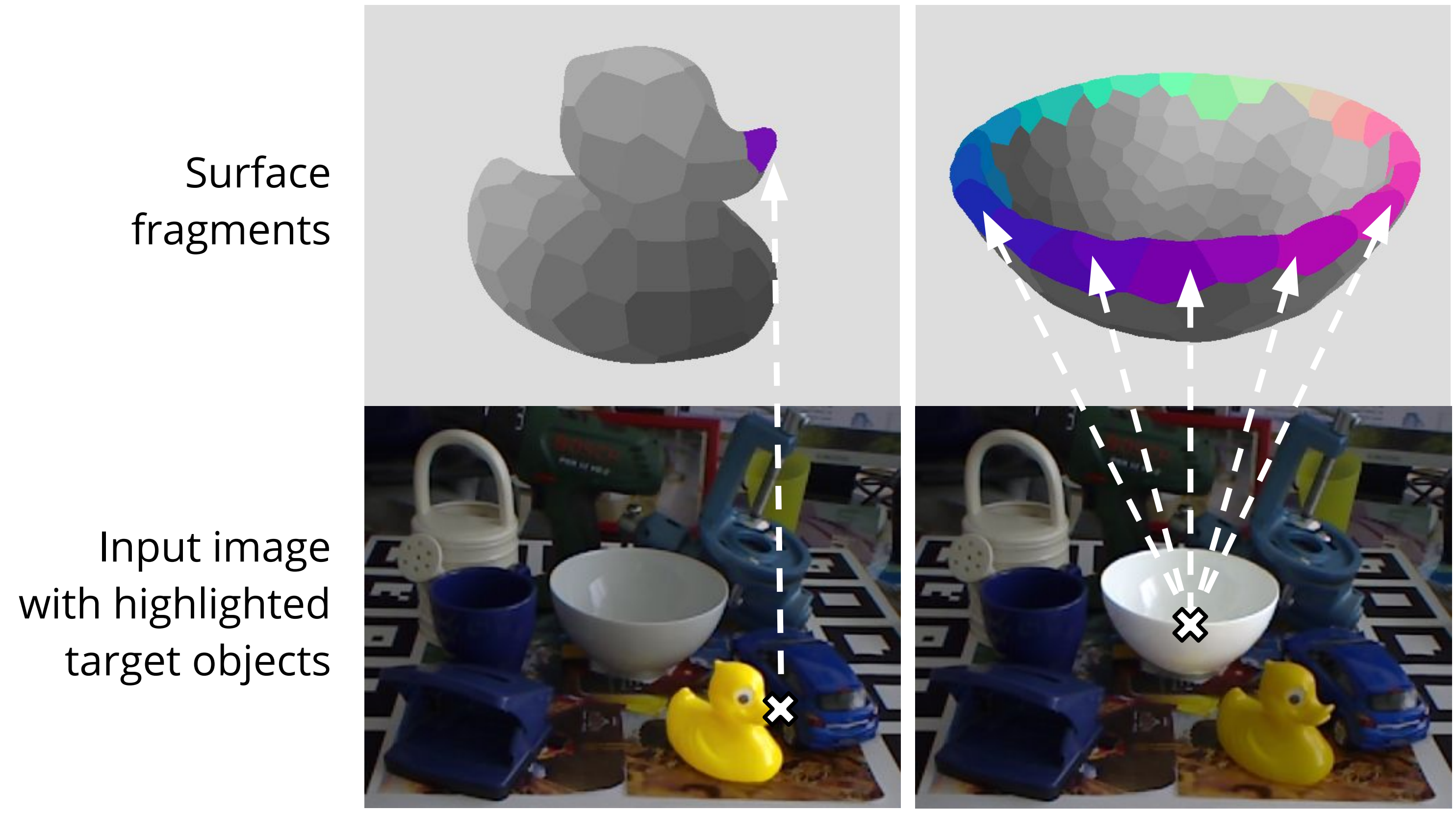
EPOS: Multiple potential 2D-3D correspondences per pixel

Surface fragments

Input image with highlighted target objects

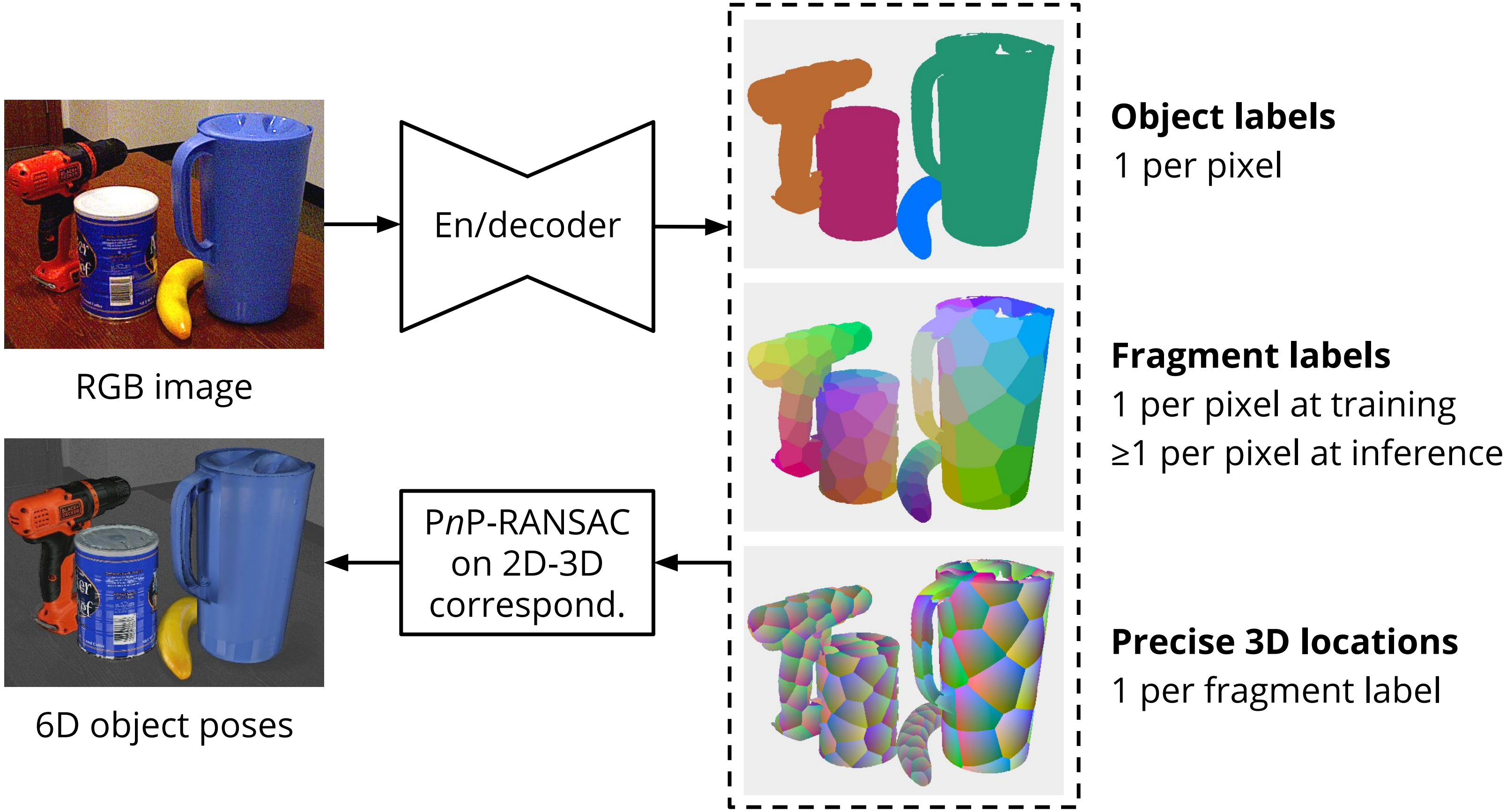


EPOS: Multiple potential 2D-3D correspondences per pixel



The distribution of corresponding fragments is predicted at each pixel, and the pixel is linked to **possibly multiple** high-confidence fragments.

EPOS: Dense prediction of 2D-3D correspondences



Potential 2D-3D correspondences are established by linking each pixel with the predicted 3D locations on possibly multiple fragments.

A custom variant of the P*n*P-RANSAC algorithm (aware of the one-to-many 2D-to-3D relationship) estimates poses from the potential correspondences.

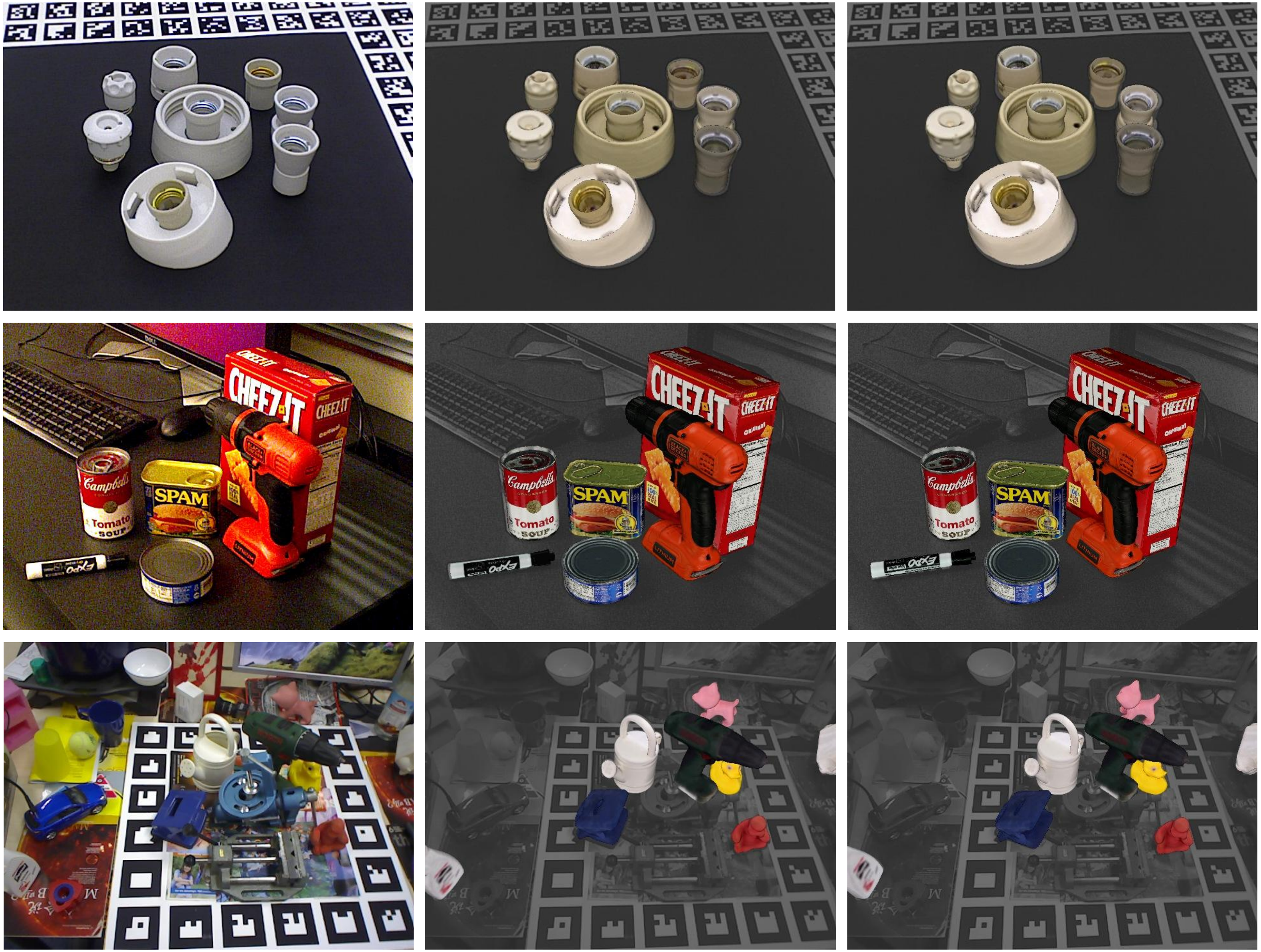
EPOS: Qualitative evaluation (1/2)

Input RGB image | Ground-truth poses | Estimated poses

T-LESS

YCB-V

LM-O



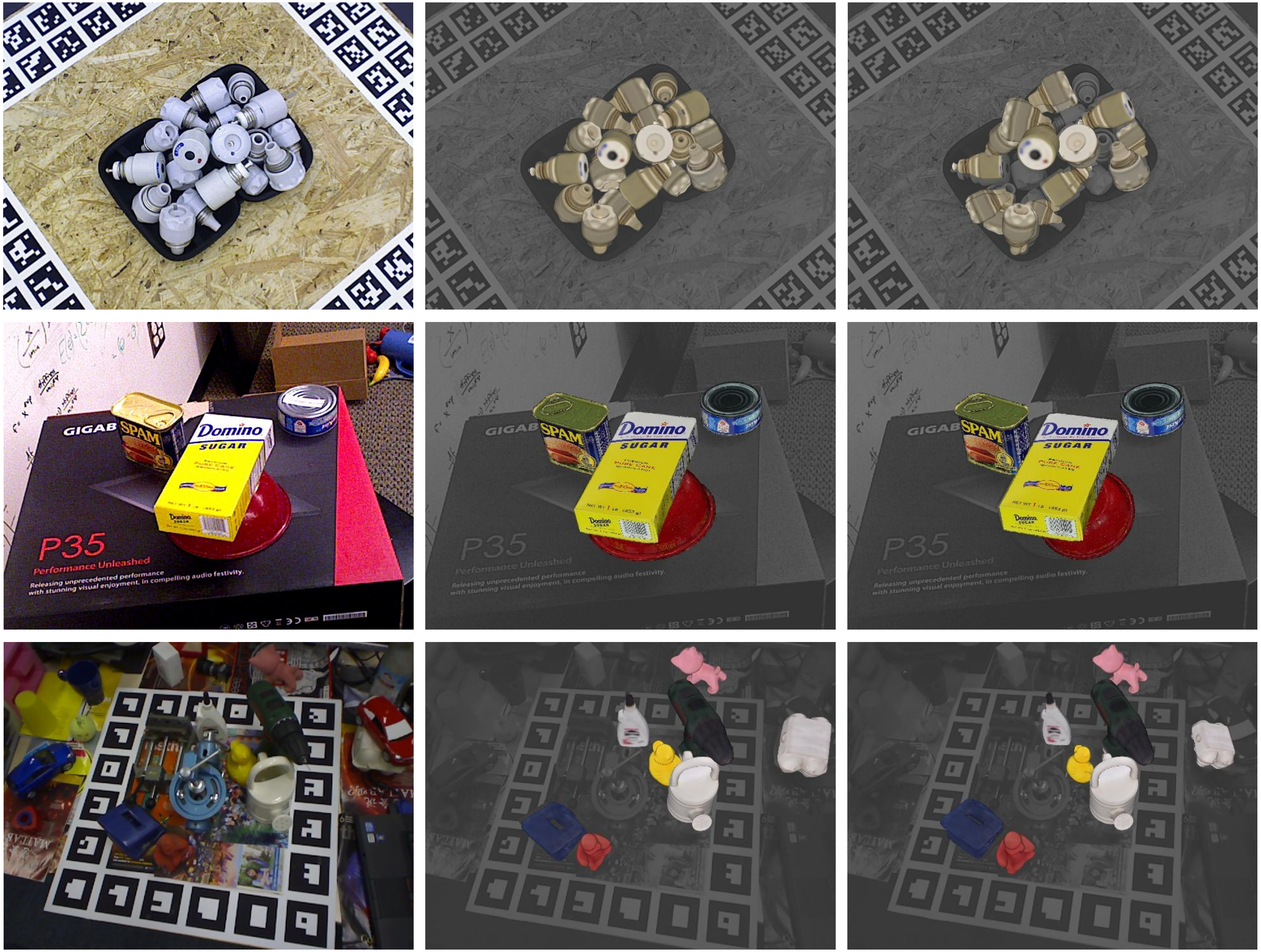
EPOS: Qualitative evaluation (2/2)

Input RGB image | Ground-truth poses | Estimated poses

T-LESS

YCB-V

LM-O



EPOS: Evaluation on BOP Challenge 2019 (bop.felk.cvut.cz)

Method	Image	T-LESS (AR)	YCB-V (AR)	LM-O (AR)	Time (s)
EPOS	RGB	**0.40**	**0.68**	**0.39**	0.63
Zhigang-CDPN-ICCV19	RGB	0.09	0.42	0.37	0.51
Sundermeyer-IJCV19	RGB	0.25	0.37	0.15	0.19
Pix2Pose-BOP-ICCV19	RGB	0.23	0.28	0.08	0.79
DPOD (synthetic)	RGB	0.07	0.22	0.17	0.23

Pix2Pose-BOP-ICCV19	RGB-D	-	**0.67**	-	
Drost-CVPR10-Edges	RGB-D	**0.44**	0.37	**0.52**	87.57
Félix&Neves-ICRA2017-IET2019	RGB-D	0.19	0.50	0.39	55.78
Sundermeyer-IJCV19+ICP	RGB-D	0.41	0.50	0.24	0.87

Vidal-Sensors18	D	**0.47**	**0.44**	**0.58**	3.22
Drost-CVPR10-3D-Edges	D	0.35	0.31	0.47	80.06
Drost-CVPR10-3D-Only	D	0.38	0.33	0.53	7.70
Drost-CVPR10-3D-Only-Faster	D	0.35	0.32	0.49	1.38

Accuracy: **EPOS outperformed all RGB methods and most RGB-D/D methods.**

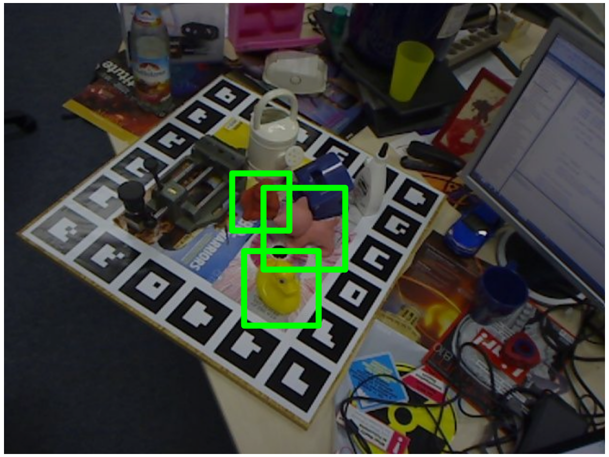
Speed: **~1.5 FPS** (non-optimized implementation) = noticeably faster than traditional methods and comparable to other CNN-based methods.

HashMatch: Hashing for Efficient Template Matching

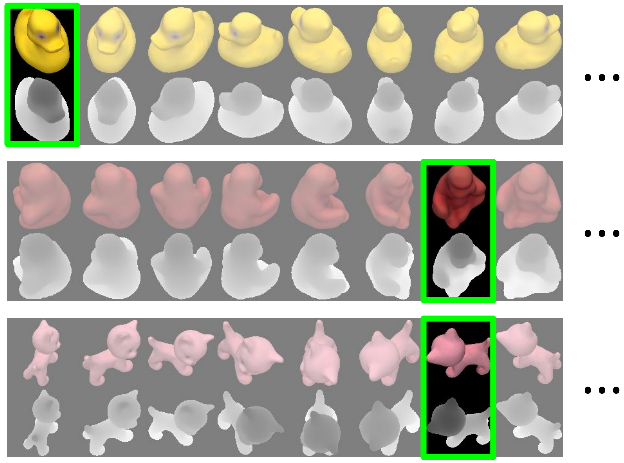
Hodaň, Haluza, Obdržálek, Matas, Lourakis, Zabulis

IROS 2015

HashMatch: The proposed method

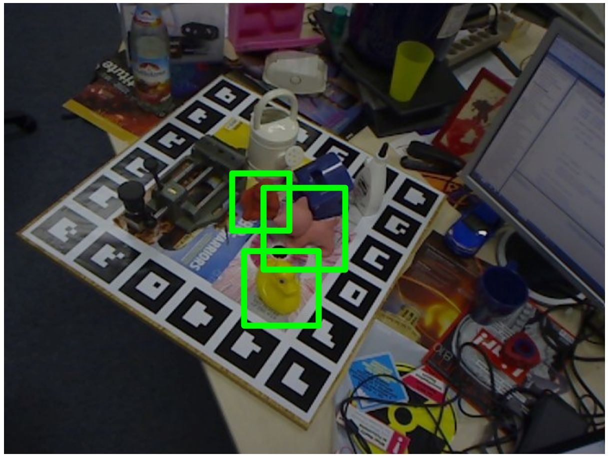


Sliding window
over test RGB-D image

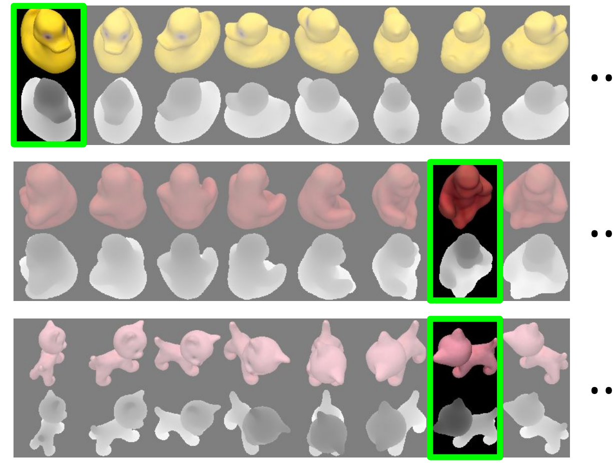


RGB-D templates
annotated with 6D poses

HashMatch: The proposed method



Sliding window
over test RGB-D image



RGB-D templates
annotated with 6D poses

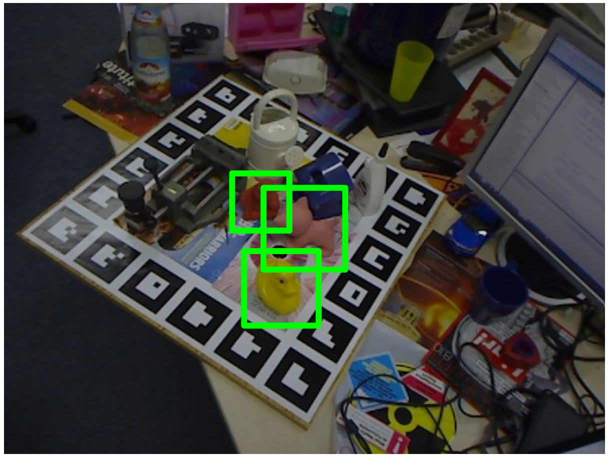
Window locations → 10^9 → Template matching → 10^2 → Non-max suppress. → 10^0 → 6D pose refinement → 10^0 → Fine 6D poses

LT = the number of window-template comparisons

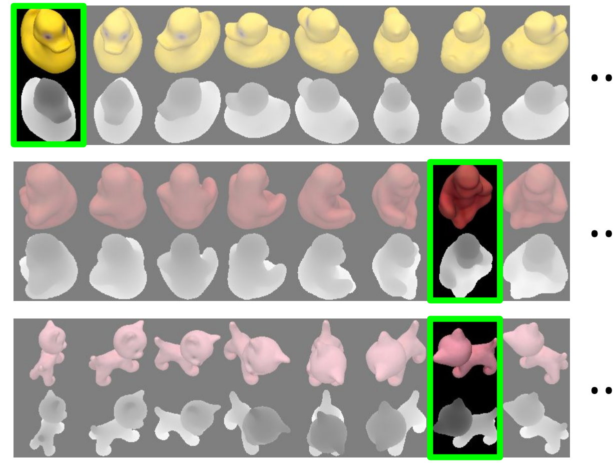
L = # of sliding window locations

T = # of templates

HashMatch: The proposed method



Sliding window
over test RGB-D image



RGB-D templates
annotated with 6D poses

Window locations → 10^9 → Objectness filter → 10^8 → Hashing → 10^5 → Template matching → 10^2 → Non-max suppress. → 10^0 → 6D pose refinement → 10^0 → Fine 6D poses

Objectness filter: reducing ***L***

Hashing: reducing ***T***

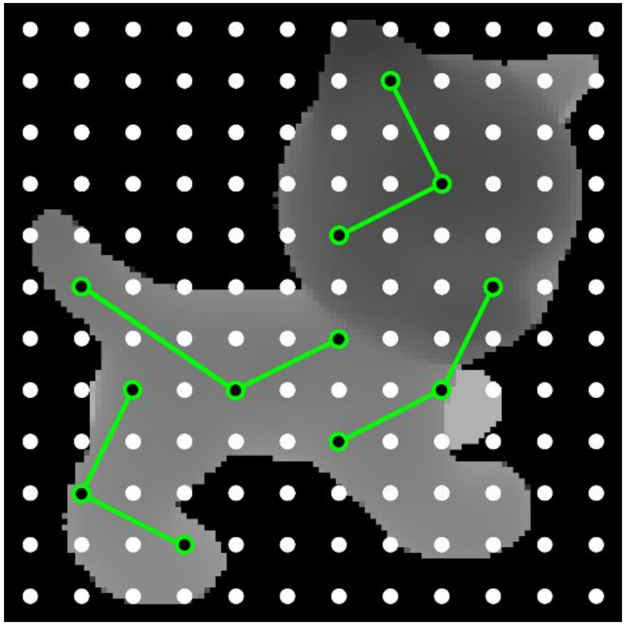
LT = the number of window-template comparisons

L = # of sliding window locations

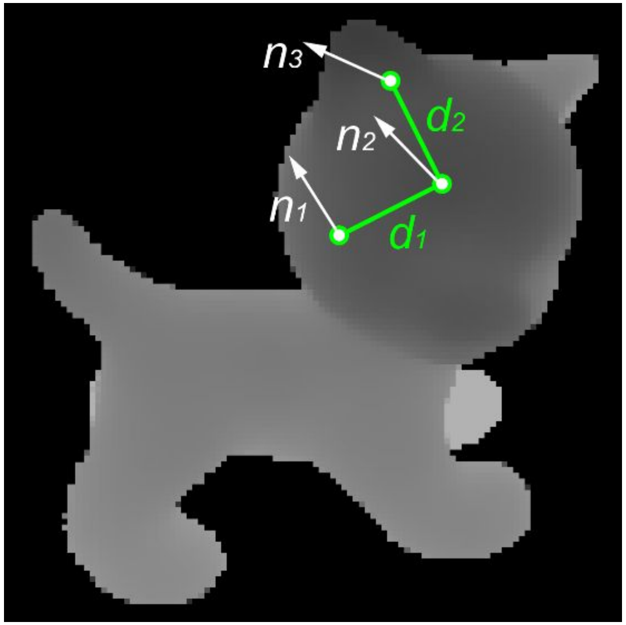
T = # of templates

HashMatch: Hashing

1. **A grid or reference points** is attached to the sliding window.
2. A triplet of points is described by **surface normals and depth differences.**
3. The descriptor is **quantized and used to retrieve identifiers of templates** with the same quantized descriptor.
4. The retrieved identifiers **vote for potentially matching templates.**
5. **A small set of templates with most votes** is passed to the next stage.



Triplets of grid points



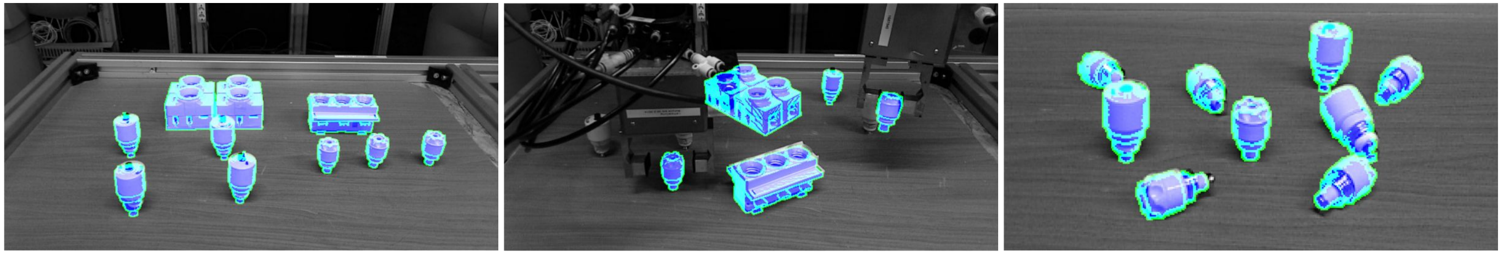
Triplet descriptor

HashMatch: Evaluation on BOP Challenge 2018

#	Method	LM	LM-O	IC-MI	IC-BIN	T-LESS	RU-APC	TUD-L	Average	Time (s)
1.	Vidal-18	87.83	59.31	95.33	96.50	66.51	36.52	80.17	74.60	4.7
2.	Drost-10-edge	79.13	54.95	94.00	92.00	67.50	27.17	87.33	71.73	21.5
3.	Drost-10	82.00	55.36	94.33	87.00	56.81	22.25	78.67	68.06	2.3
4.	Hodan-15	87.10	51.42	95.33	90.50	63.18	37.61	45.50	67.23	13.5
5.	Brachmann-16	75.33	52.04	73.33	56.50	17.84	24.35	88.67	55.44	4.4
6.	Hodan-15-nopso	69.83	34.39	84.67	76.00	62.70	32.39	27.83	55.40	12.3
7.	Buch-17-ppfh	56.60	36.96	95.00	75.00	25.10	20.80	68.67	54.02	14.2
8.	Kehl-16	58.20	33.91	65.00	44.00	24.60	25.58	7.50	36.97	1.8
9.	Buch-17-si	33.33	20.35	67.33	59.00	13.34	23.12	41.17	36.81	15.9
10.	Brachmann-14	67.60	41.52	78.67	24.00	0.25	30.22	0.00	34.61	1.4
11.	Buch-17-ecsad	13.27	9.62	40.67	59.00	7.16	6.59	24.00	22.90	5.9
12.	Buch-17-shot	5.97	1.45	43.00	38.50	3.83	0.07	16.67	15.64	6.7
13.	Tejani-14	12.10	4.50	36.33	10.00	0.13	1.52	0.00	9.23	1.4
14.	Buch-16-ppfh	8.13	2.28	20.00	2.50	7.81	8.99	0.67	7.20	47.1
15.	Buch-16-ecsad	3.70	0.97	3.67	4.00	1.24	2.90	0.17	2.38	39.1

Average image processing time (with 43740 templates of 15 objects):

- Exhaustive template matching: ~15s
- HashMatch: ~2s → **sub-linear complexity in the number of templates**



Used for robotic assembly in the DARWIN EU project

ObjectSynth: Synthesis of Photorealistic Training Images

Hodaň, Vineet, Gal, Shalev, Hanzelka,
Connell, Urbina, Sinha, Guenter

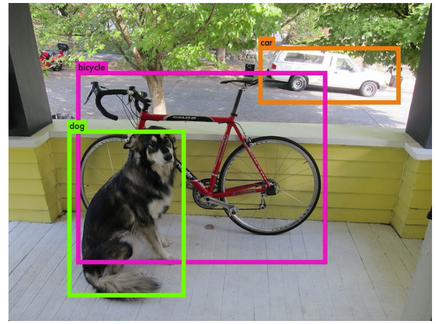
ICIP 2019

Microsoft

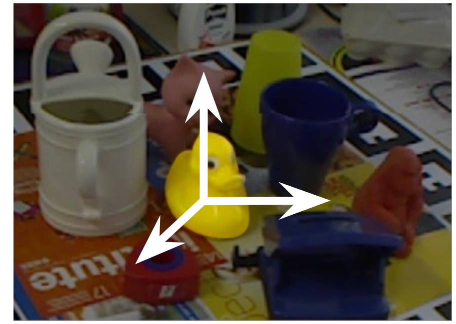
Neural networks are great, but data hungry

GT annotation of a large number of real images is **expensive.**

Image classification
$



Object detection
$$



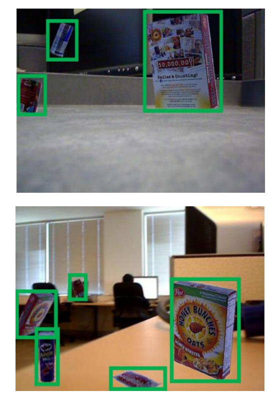
Object pose estimation
$$$

Many object pose estimation methods rely on **“cut & paste”** synthetic images:

Object segments cut from real or rendered images

+

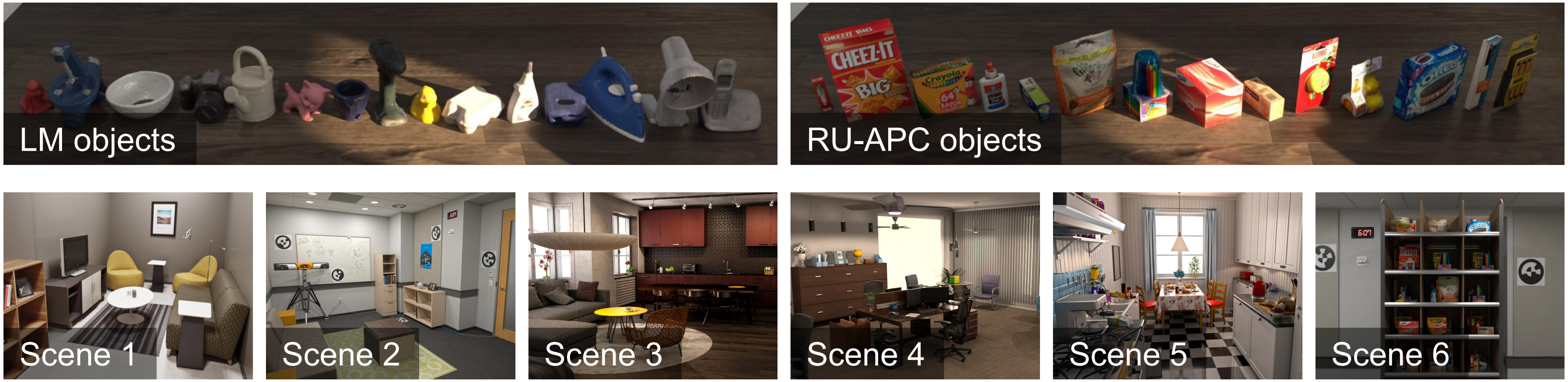
=



Lack of photorealism (inconsistent lighting, missing interreflections and shadows, unnatural object pose and context) **enlarges the synthetic-real domain gap.**

ObjectSynth: Reducing the gap with photorealistic images

3D object models rendered in 3D scene models by **ray tracing:**



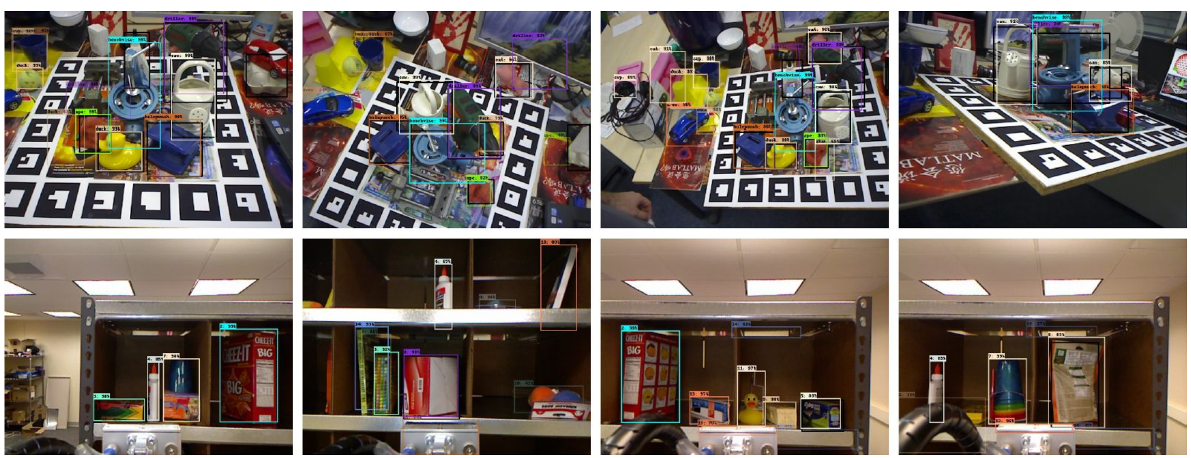
Examples of rendered images rendered with the **Arnold ray-tracer**

ObjectSynth: Evaluation

Photorealistic training images

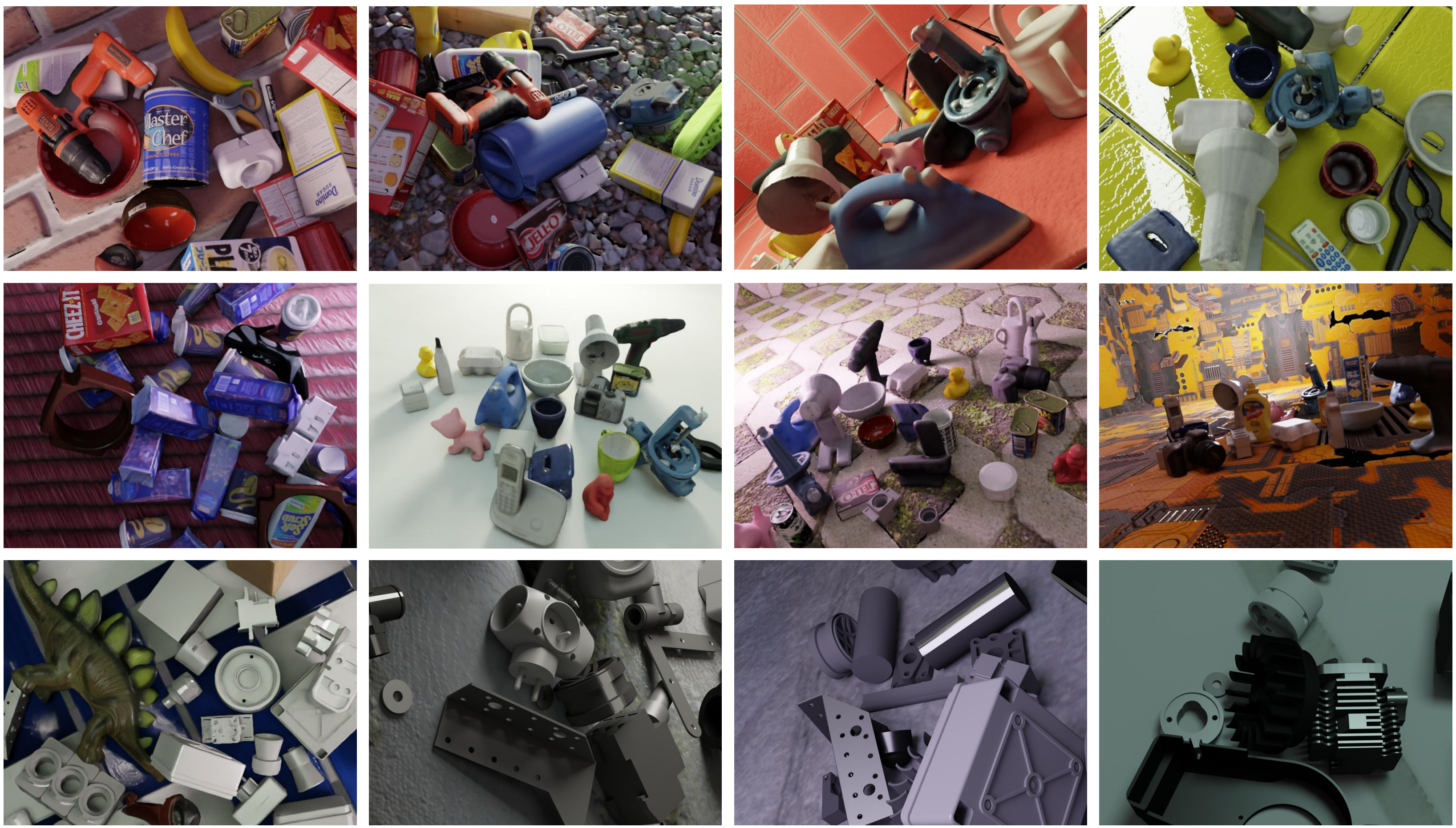
Cut & paste baseline: 3D object models on random photographs (in the same poses as in the photorealistic images)

Faster R-CNN achieves **11–24% higher mAP@.75IoU** on real test images when trained on the ray-traced images.



Training images for BOP Challenge 2020

- **BlenderProc4BOP** – an open-source and light-weight physically-based renderer which implements a refined version of ObjectSynth.
- **350K pre-rendered training images** provided to the participants.
- **5th method** (out of 26) was trained only on these images (with no real).

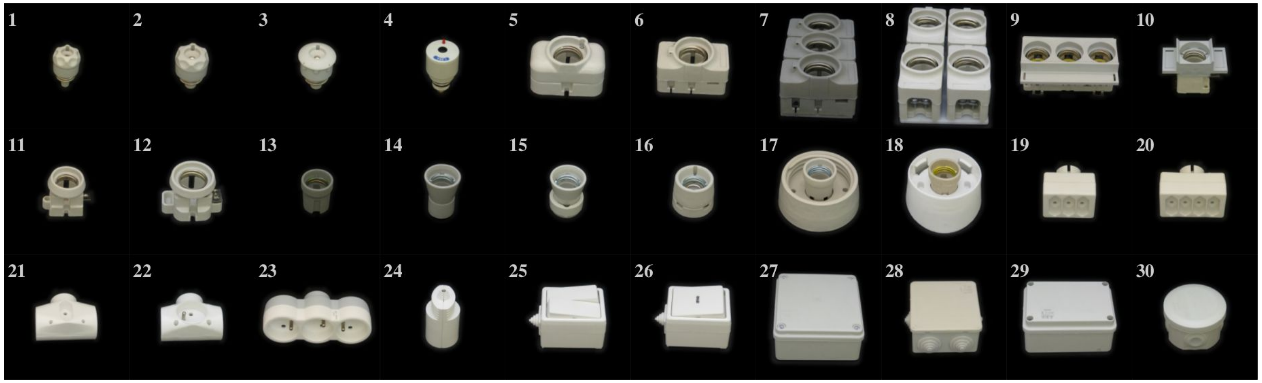


T-LESS: An RGB-D Dataset with Texture-less Objects

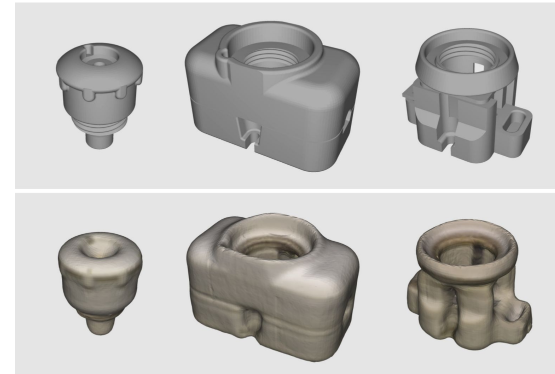
Hodaň, Haluza, Obdržálek, Matas, Lourakis, Zabulis

WACV 2017

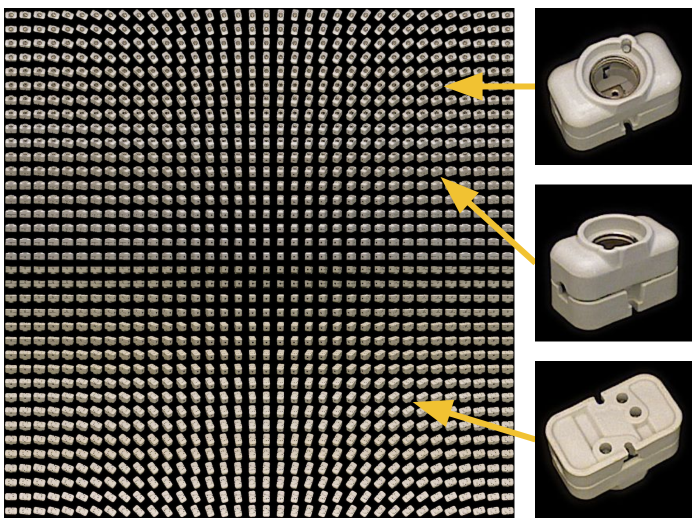
The T-LESS dataset



30 objects with **no significant texture or color**,
with symmetries and mutual similarities in shape or size



CAD and reconstructed
3D object models



38K training images

10K test images from 20 scenes
with accurate ground-truth 6D poses

Well accepted (>200 citations) and still one of the more difficult datasets.

BOP: Benchmark for 6D Object Pose Estimation

Hodaň, Sundermeyer, Michel, Labbé, Brachmann, Kehl, Buch, Kraft, Drost, Vidal, Ihrke, Zabulis, Sahin, Manhardt, Tombari, Kim, Obdržálek, Matas, Rother

ECCVW 2016, ECCV 2018, ECCVW 2020

Motivation: Unclear state of the art

SOTA unclear because:

- No standard evaluation methodology.
- New methods usually compared with only a few competitors on a few datasets.
- Scores on the most commonly used Linemod dataset have been saturated.

BOP includes:

- Evaluation methodology (task definition, new pose-error functions).
- 11 RGB-D datasets in a unified format + more are coming.
- Online evaluation system at bop.felk.cvut.cz (40K visits by 14K users since July’19).
- Public workshops and challenges at ICCV and ECCV conferences.

R6D: International workshops on recovering 6D object pose

T. Hodaň, M. Sundermeyer, E. Brachmann, R. Kouskouridas, B. Drost, T.-K. Kim, J. Matas, C. Rother, V. Lepetit, A. Leonardis, K. Walas, C. Steger, J. Sock

ECCV'20
SEC, Glasgow
23-28 AUGUST 2020

BOP Challenge 2020

#	Method	Year	PPF	CNN	...models	Train. im.	...type	Test im.	Refine.	Avg.	LM-O	T-LESS	TUD-L	IC-BIN	ITODD	HB	YCB-V	Time
1	**CosyPose-ECCV20-Synt+Real-1View-ICP**	2020	No	Yes	3/dataset	RGB	Synt+real	RGB-D	RGB+ICP	0.698	0.714	0.701	0.939	0.647	0.313	0.712	0.861	13.743
2	**Koenig-Hybrid-DL-PointPairs**	2020	Yes	Yes	1/dataset	RGB	Synt+real	RGB-D	ICP	0.639	0.631	0.655	0.920	0.430	0.483	0.651	0.701	0.633
3	**CosyPose-ECCV20-Synt+Real-1View**	2020	No	Yes	3/dataset	RGB	Synt+real	RGB	RGB	0.637	0.633	0.728	0.823	0.583	0.216	0.656	0.821	0.449
4	**Pix2Pose-BOP20_w/ICP-ICCV19**	2020	No	Yes	1/object	RGB	Synt+real	RGB-D	ICP	0.591	0.588	0.512	0.820	0.390	0.351	0.695	0.780	4.844
5	**CosyPose-ECCV20-PBR-1View**	2020	No	Yes	3/dataset	RGB	PBR only	RGB	RGB	0.570	0.633	0.640	0.685	0.583	0.216	0.656	0.574	0.475
6	**Vidal-Sensors18**	2019	Yes	No	-	-	-	D	ICP	0.569	0.582	0.538	0.876	0.393	0.435	0.706	0.450	3.220
7	**CDPNv2_BOP20 (RGB-only & ICP)**	2020	No	Yes	1/object	RGB	Synt+real	RGB-D	ICP	0.568	0.630	0.464	0.913	0.450	0.186	0.712	0.619	1.462
8	**Drost-CVPR10-Edges**	2019	Yes	No	-	-	-	RGB-D	ICP	0.550	0.515	0.500	0.851	0.368	0.570	0.671	0.375	87.568
9	**CDPNv2_BOP20 (PBR-only & ICP)**	2020	No	Yes	1/object	RGB	PBR only	RGB-D	ICP	0.534	0.630	0.435	0.791	0.450	0.186	0.712	0.532	1.491
10	**CDPNv2_BOP20 (RGB-only)**	2020	No	Yes	1/object	RGB	Synt+real	RGB	No	0.529	0.624	0.478	0.772	0.473	0.102	0.722	0.532	0.935
11	**Drost-CVPR10-3D-Edges**	2019	Yes	No	-	-	-	D	ICP	0.500	0.469	0.404	0.852	0.373	0.462	0.623	0.316	80.055
12	**Drost-CVPR10-3D-Only**	2019	Yes	No	-	-	-	D	ICP	0.487	0.527	0.444	0.775	0.388	0.316	0.615	0.344	7.704
13	**CDPN_BOP19 (RGB-only)**	2020	No	Yes	1/object	RGB	Synt+real	RGB	No	0.479	0.569	0.490	0.769	0.327	0.067	0.672	0.457	0.480
14	**CDPNv2_BOP20 (PBR-only&RGB-only)**	2020	No	Yes	1/object	RGB	PBR only	RGB	No	0.472	0.624	0.407	0.588	0.473	0.102	0.722	0.390	0.978
15	**leaping from 2D to 6D**	2020	No	Yes	1/object	RGB	Synt+real	RGB	No	0.471	0.525	0.403	0.751	0.342	0.077	0.658	0.543	0.425
16	**EPOS-BOP20-PBR**	2020	No	Yes	1/dataset	RGB	PBR only	RGB	No	0.457	0.547	0.467	0.558	0.363	0.186	0.580	0.499	1.874
17	**Drost-CVPR10-3D-Only-Faster**	2019	Yes	No	-	-	-	D	ICP	0.454	0.492	0.405	0.696	0.377	0.274	0.603	0.330	1.383
18	**Félix&Neves-ICRA2017-IET2019**	2019	Yes	Yes	1/dataset	RGB-D	Synt+real	RGB-D	ICP	0.412	0.394	0.212	0.851	0.323	0.069	0.529	0.510	55.780
19	**Sundermeyer-IJCV19+ICP**	2019	No	Yes	1/object	RGB	Synt+real	RGB-D	ICP	0.398	0.237	0.487	0.614	0.281	0.158	0.506	0.505	0.865
20	**Zhigang-CDPN-ICCV19**	2019	No	Yes	1/object	RGB	Synt+real	RGB	No	0.353	0.374	0.124	0.757	0.257	0.070	0.470	0.422	0.513
21	**PointVoteNet2**	2020	No	Yes	1/object	RGB-D	PBR only	RGB-D	ICP	0.351	0.653	0.004	0.673	0.264	0.001	0.556	0.308	-
22	**Pix2Pose-BOP20-ICCV19**	2020	No	Yes	1/object	RGB	Synt+real	RGB	No	0.342	0.363	0.344	0.420	0.226	0.134	0.446	0.457	1.215
23	**Sundermeyer-IJCV19**	2019	No	Yes	1/object	RGB	Synt+real	RGB	No	0.270	0.146	0.304	0.401	0.217	0.101	0.346	0.377	0.186
24	**SingleMultiPathEncoder-CVPR20**	2020	No	Yes	1/all	RGB	Synt+real	RGB	No	0.241	0.217	0.310	0.334	0.175	0.067	0.293	0.289	0.186
25	**Pix2Pose-BOP19-ICCV19**	2019	No	Yes	1/object	RGB	Synt+real	RGB	No	0.205	0.077	0.275	0.349	0.215	0.032	0.200	0.290	0.793
26	**DPOD (synthetic)**	2019	No	Yes	1/scene	RGB	Synt	RGB	No	0.161	0.169	0.081	0.242	0.130	0.000	0.286	0.222	0.231

A detailed analysis at: **bop.felk.cvut.cz**

Summary

EPOS (CVPR’20) – an RGB method applicable to a broad range of objects.

HashMatch (IROS’15) – efficient RGB-D template matching.

ObjectSynth (ICIP’19, RSSW’20) – synthesis of photorealistic training images.

T-LESS (WACV’17) – an RGB-D dataset with texture-less objects.

BOP (ECCVW’16, ECCV’18, ECCVW’20) – a benchmark for 6D object pose estimation.

Thank you!

Real-world demo: EPOS applied frame by frame on a video from a cell phone.