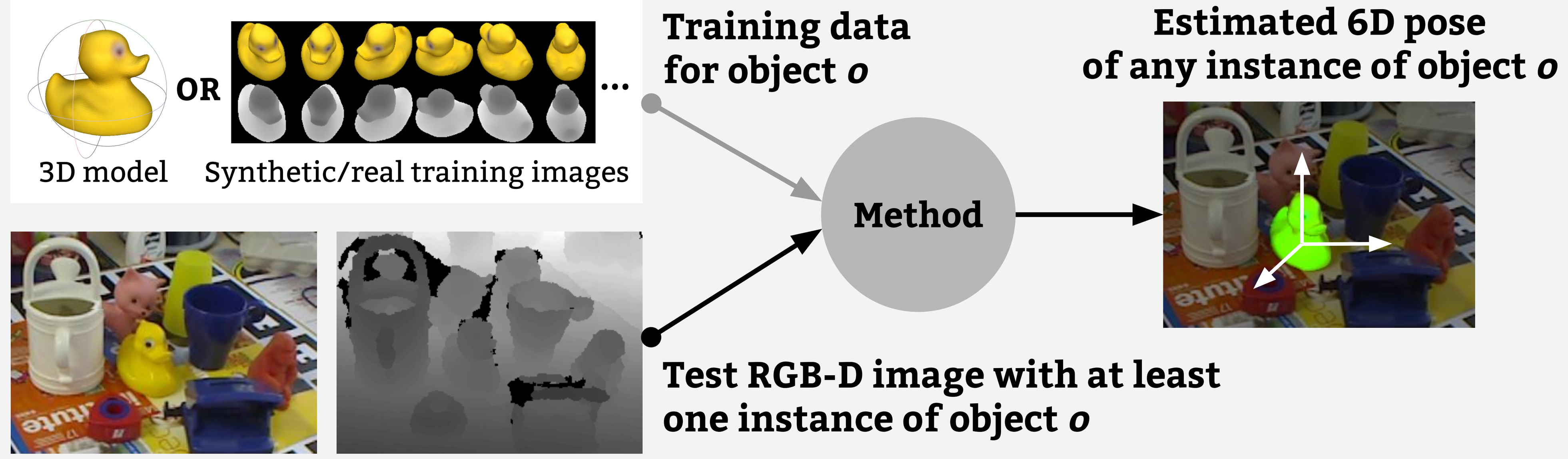
BOP: Benchmark for 6D Object Pose Estimation

Hodan[1], Michel[2], Brachmann[3], Kehl[4], Buch[5], Kraft[5], Drost[6], Vidal[7], Ihrke[2], Zabulis[8], Sahin[9], Manhardt[10], Tombari[10], Kim[9], Matas[1], Rother[3]

1 CTU Czech Technical University in Prague; 2 Technische Universität Dresden; 3 Universität Heidelberg; 4 Toyota Research Institute; 5 SDU University of Southern Denmark; 6 MVTec Software GmbH; 7 Taiwan Tech; 8 FORTH; 9 Imperial College London; 10 Technical University of Munich

Task: 6D pose estimation of a single instance of a single object

Relevant for robotics and augmented reality, addressed by all published methods



Unclear state of the art

1) No standard evaluation method, **2)** Datasets have different formats and GT quality, **3)** Methods compared with only a few competitors on a small number of datasets

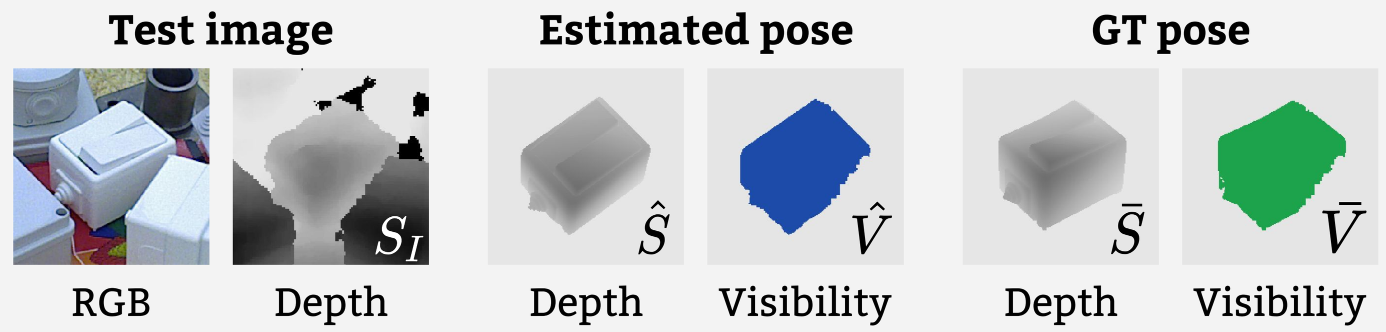
BOP includes 8 datasets in a unified format with quality GT

- **Texture-mapped 3D models of 89 diverse objects**
- **277K training RGB-D images** showing isolated objects (mostly synthetic)
- **62K test RGB-D images** of scenes with graded complexity
- **High-quality ground-truth 6D object poses** for all images
- **Six publicly available datasets**, some reduced and re-annotated
- **Two new datasets** focusing on varying lighting conditions

Test images cover different application scenarios

LM/LM-O	IC-MI	IC-BIN	T-LESS	RU-APC	TUD-L - new	TYO-L - new
Hinterstoisser et al.	Tejani et al.	Doumanoglou et al.	Hodan et al.	Rennie et al.	TU Dresden Light	Toyota Light

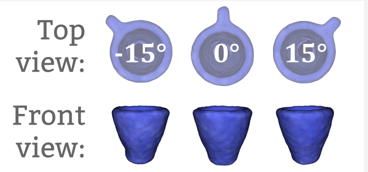
Pose error measured by Visible Surface Discrepancy (VSD)



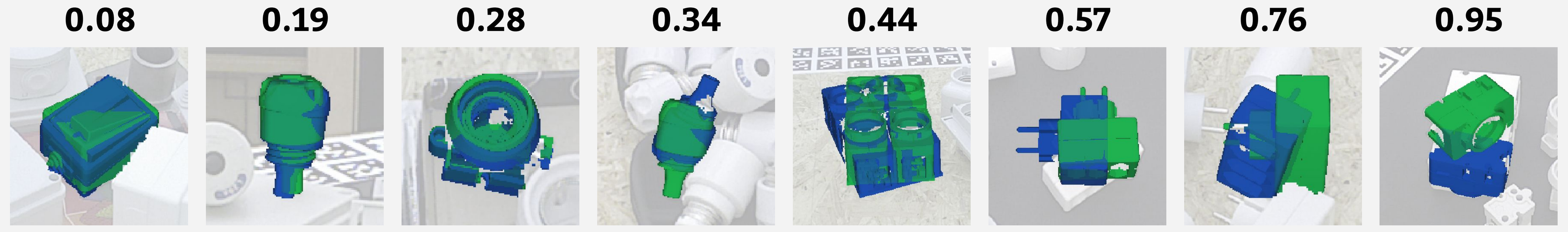
Visibility masks are obtained by comparing $\hat{S}$ and $\bar{S}$ with S_I

$$e_{\mathrm{VSD}}(\hat{S}, \bar{S}, S_I, \hat{V}, \bar{V}, \tau) = \operatorname*{avg}_{p \in \hat{V} \cup \bar{V}} \begin{cases} 0 & \text{if } p \in \hat{V} \cap \bar{V} \wedge |\hat{S}(p) - \bar{S}(p)| < \tau \\ 1 & \text{otherwise} \end{cases}$$

- Estimated pose **considered correct if** $e_{\mathrm{VSD}} < \theta$
- Pose error is calculated only over the visible part of the surface
  $\Rightarrow$ **Indistinguishable poses are treated as equivalent**



Indistinguishable poses



Values of e_{VSD} for example pose estimates, in blue, the GT in green

Experimental setup

- The methods were **evaluated by their authors**
- **Parameters of each method were fixed** for all objects and datasets
- **Test** defined by a pair *(I, o)*, image *I* shows at least one instance of object *o*
- The performance was measured by **recall**, i.e. the fraction of tests for which a correct object pose was estimated, with **misalignment tolerance** τ **= 20 mm** and **correctness threshold** θ **= 0.3**

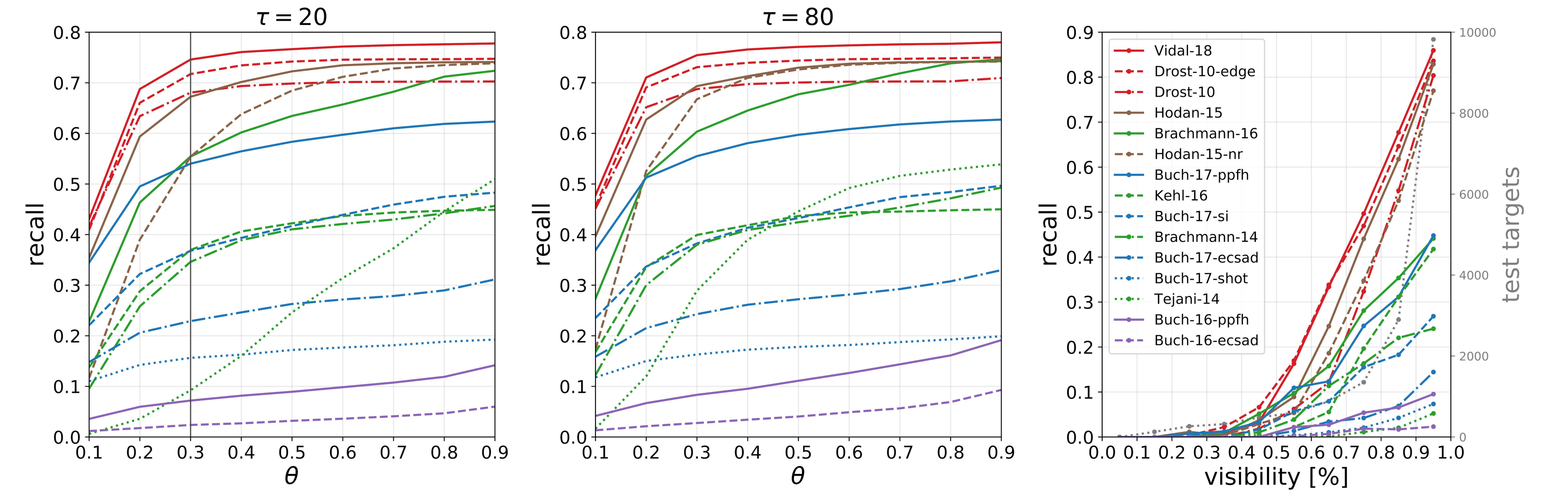
Online evaluation system: bop.felk.cvut.cz

Up-to-date leaderboards + a form for submission of new results

Evaluation of 15 recent methods

1) Methods based on point pair features, 2) Template matching methods, 3) Learning-based methods, 4) Methods based on 3D local features

# 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



- **Poses estimated by most methods are either of a high quality or totally off** – the scores increase only slightly if τ is increased from 20 to 80 mm, or if $\theta > 0.3$
- **Occlusion is a big challenge for current methods** – all methods perform on LM by at least 30% better than on LM-O, which includes the same but occluded objects
- **Object symmetries and similarities** of the T-LESS objects cause problems to methods based on 3D local features and learning-based methods
- **Varying lighting conditions** present a challenge for methods that rely on synthetic training RGB images rendered with fixed lighting
- **Noisy depth images** in RU-APC present problems to all methods
- Methods were **optimized primarily for recall**, not for speed

ECCV 2018 European Conference on Computer Vision
8 – 14 September 2018 | Munich, Germany