Martin Sundermeyer, DLR, TU Munich **Tomáš Hodaň**, Reality Labs at Meta **Yann Labbé**, Inria Paris **Gu Wang**, Tsinghua University **Eric Brachmann**, Niantic **Bertram Drost**, MVTec Software **Carsten Rother**, Heidelberg Uni **Jiří Matas**, CTU in Prague

7th International Workshop on Recovering 6D Object Pose ECCV 2022, October 23, Tel Aviv

bop.felk.cvut.cz

BOP: Benchmark for 6D object pose estimation

Goal: Capture and report the state of the art in estimating the 6D pose of rigid objects from RGB/RGB-D images.

BOP currently comprises of:

- **● Evaluation methodology**
- **● Online evaluation system at bop.felk.cvut.cz**
- **12 datasets in a unified format**
	- Texture-mapped 3D models of 199 objects
	- >700K training RGB-D images (mostly synthetic)
	- >100K test RGB-D images of scenes with graded complexity
	- Images are annotated with ground-truth 6D object poses

BOP publications

BOP: Benchmark for 6D Object Pose Estimation, ECCV 2018

T. Hodaň, F. Michel, E. Brachmann, W. Kehl, A. G. Buch, D. Kraft, B. Drost, J. Vidal, S. Ihrke, X. Zabulis, C. Sahin, F. Manhardt, F. Tombari, T.-K. Kim, J. Matas, C. Rother

BOP Challenge 2020 on 6D Object Localization, ECCVW 2020 T. Hodaň, M. Sundermeyer, B. Drost, Y. Labbé, E. Brachmann, F. Michel, C. Rother, J. Matas

BOP Challenge 2022 on Detection, Segmentation and Pose Estimation of Specific Rigid Object – In preparation

BOP Challenge 2022 – Tasks

- **● 2D object detection** new in BOP 2022
- **● 2D object segmentation** new in BOP 2022
- **6D object localization** as in BOP 2019 and 2020

The new tasks were introduced to address the design of many recent methods for object pose estimation, which start by detecting/segmenting objects and then estimate the poses from the predicted regions.

Evaluating the detection/segmentation stage and the pose estimation stage separately allows to better understand advances in the two stages (participants could use provided default detections/segmentations).

2D object detection task

2D object segmentation task

Evaluation of 2D object detection/segmentation

We adopt metrics from the **COCO Object Detection Challenge.**

The main metric is the **Average Precision (AP)** calculated at different Intersection over Union (IoU=.50:.05:.95) values.

A method is required to detect/segment only objects that are visible from at least 10%. If a method detects/segments also objects that are visible from less than 10%, these are ignored and not counted as false positives.

6D object localization task

BOP 2018

BOP 2019, 2020 and 2022

BOP 2019, 2020 and 2022

Why not 6D object detection, where the number of instances is unknown?

- 1. Evaluating 6D object detection is expensive as many more estimates need to be evaluated to calculate the precision/recall curve.
- 2. The scores on the simpler 6D localization task are not saturated.

Pose error functions

How good is the estimated pose?

The error of an estimated pose w.r.t. the GT pose is measured by:

- 1. **VSD: Visible Surface Discrepancy** Error calculated over the visible part \Rightarrow indistinguishable poses are equivalent.
- 2. **MSSD: Maximum Symmetry-Aware Surface Distance** Measures the surface deviation in $3D \Rightarrow$ relevant for robotic applications.
- 3. **MSPD: Maximum Symmetry-Aware Projection Distance** Measures the perceivable deviation \Rightarrow relevant for AR applications.

See bop.felk.cvut.cz for details.

Accuracy score

An estimated pose *E* is considered **correct** w.r.t. ground-truth pose *G* and pose-error function *F,* **if** *F(E, G) < θ*, where *F* is VSD, MSSD or MSPD, and *θ* is the threshold of correctness.

Average Recall w.r.t. function *F***:** AR_F = the average of recall rates calculated for multiple settings of threshold *θ* (and tolerance *τ* for VSD). Recall rate = the fraction of objects for which a correct pose is estimated.

Average Recall on dataset *D***:** AR_D = (AR_{VSD}+ AR_{MSSD}+ AR_{MSPD}) / 3

The overall accuracy (AR) = the average of per-dataset AR_{*D*} scores. ⇒ Each dataset is treated as a separate sub-challenge which avoids the overall score being dominated by larger datasets.

See **bop.felk.cvut.cz** for details.

Rules

1. **For training, a method** *can***...**

- a. use the provided 3D object models and training images.
- b. render extra training images.
- c. use the range (not a probability distribution) of all GT poses in the test images (e.g. objects are from 20 to 100 cm from the camera).

2. **For training, a method** *cannot***...**

- a. use a single pixel of test images.
- b. use the individual ground-truth poses from test images.
- 3. **A fixed set of hyper-parameters** required for all objects/datasets.

BOP Toolkit

Scripts for reading the standard dataset format, evaluation etc.

Classical pre-DNN (RGB-D and D) methods on the **SiSo task.**

Pose error measured with only **Visible Surface Discrepancy (VSD).**

Methods based on Point Pair Features Template matching methods, Learning-based methods Methods based on 3D local features

Classical pre-DNN (RGB-D and D) methods on the **SiSo task.**

Pose error measured with only **Visible Surface Discrepancy (VSD).**

Methods based on Point Pair Features Template matching methods, Learning-based methods Methods based on 3D local features

Classical and DNN (RGB, RGB-D and D) methods on the **ViVo task.**

Evaluation methodology as in BOP 2020 and 2022.

Classical and DNN (RGB, RGB-D and D) methods on the **ViVo task.**

Evaluation methodology as in BOP 2020 and 2022.

Classical and DNN (RGB, RGB-D and D) methods on the **ViVo task.**

Evaluation methodology as in BOP 2020 and 2022.

Classical methods outperformed DNN methods, because of:

1. **Insufficient number of real training images** annotated with 6D object poses – annotation is expensive!

2. **Large domain gap** between real test images and the commonly used synthetic training images (objects rendered on random background).

- **BlenderProc4BOP** an open-source photorealistic (PBR) renderer.
- **350K pre-rendered training images** provided to the participants.

DNN-based methods finally caught up with PPF-based methods!

Most methods used both synthetic and real training images, but...

Competitive results can be achieved with PBR training images only. (For LM-O, IC-BIN, ITODD and HB, only synthetic training images are provided.)

PBR training images yield a noticeable improvement over "naively" synthesized images (objects rendered on random backgrounds). Similarly to CDPN, EPOS jumped from 0.44 to 0.55 on LM-O with PBR images.

BOP Challenge 2022 – Submissions

Submission system: bop.felk.cvut.cz, deadline: October 16, 2022.

49 pose estimation methods (23 since BOP 2020) evaluated on all 7 core datasets: LM-O, T-LESS, TUD-L, IC-BIN, ITODD, HB, YCB-V.

The submission form stays open! Coming soon: All raw predictions available on the BOP website.

Signed in as **admin:** My methods My submissions Edit account Sign out

HOME CHALLENGES DATASETS LEADERBOARDS SUBMIT RESULTS

 $Task:$ Pose estimation (BOP 2019-2022) Detection (BOP 2022) Segmentation (BOP 2022)

Dataset: Core datasets LM LM-O T-LESS ITODD HB HOPE YCB-V RU-APC IC-BIN IC-MI TUD-L TYO-L

Pose estimation (BOP 2019-2022) - Core datasets

This leaderbord shows the overall ranking on the core datasets (LM-O, T-LESS, TUD-L, IC-BIN, ITODD, HB, YCB-V). For each method, the date of the latest considered submission is reported. If more submissions of a method are available for a dataset, the submission with the highest AR_{Core} score is considered. The performance scores are defined in the BOP Challenge 2019 description. The reported time is the average image processing time averaged over the core datasets.

BOP Challenge 2022: 6D object localization

37

וטע

Most top BOP methods use 3 stages:

- 1. Detection / segmentation
- 2. Pose estimation
- 3. Pose refinement

Most top BOP methods use 3 stages:

- 1. Detection / segmentation
- 2. Pose estimation
- 3. Pose refinement
- Therefore, BOP 2022:
- Measures COCO metrics (AP) of detection / segmentation

Most top BOP methods use 3 stages:

- 1. Detection / segmentation
- 2. Pose estimation
- 3. Pose refinement

Therefore, BOP 2022:

- Measures COCO metrics (AP) of detection / segmentation
- Provides the best detections from BOP 2020 (CosyPose)

Most top BOP methods use 3 stages:

- 1. Detection / segmentation
- 2. Pose estimation
- 3. Pose refinement

Therefore, BOP 2022:

- Measures COCO metrics (AP) of detection / segmentation
- Provides the best detections from BOP 2020 (CosyPose)

YOLOX from GDRNPP gains +16.8AP over MaskRCNN from Cosypose!

ءه ا

 T_{r2}

and the state

BOP Challenge 2022: 2D object segmentation

ZebraPose refines masks from CosyPose detections: **+18.2 AP!**

As for detection, segmentation methods still use RGB only.

Object poses estimated by GDRNPP [1]

[1] Image courtesy Xingyu Liu 54

Object poses estimated by GDRNPP [1]

Sponsors of BOP 2022 Awards

Donated \$4000 (each \$2000)

Meta VNIANTIC

The Overall Best Detection Method GDRNPPDet_PBRReal

Xingyu Liu, Ruida Zhang, Chenyangguang Zhang, Bowen Fu, Jiwen Tang, Xiquan Liang, Jingyi Tang, Xiaotian Cheng, Yukang Zhang, Gu Wang, Xiangyang Ji

The Best BlenderProc-Trained Detection Method

GDRNPPDet_PBR

Xingyu Liu, Ruida Zhang, Chenyangguang Zhang, Bowen Fu, Jiwen Tang, Xiquan Liang, Jingyi Tang, Xiaotian Cheng, Yukang Zhang, Gu Wang, Xiangyang Ji

The Overall Best Segmentation Method

ZebraPoseSAT-EffnetB4 (DefaultDetection)

Yongzhi Su, Praveen Nathan, Torben Fetzer, Jason Rambach, Didier Stricker, Mahdi Saleh, Yan Di, Nassir Navab, Benjamin Busam, Federico Tombari, Yongliang Lin, Yu Zhang

The Best BlenderProc-Trained Segmentation Method

ZebraPoseSAT-EffnetB4 (DefaultDet+PBR_Only)

Yongzhi Su, Praveen Nathan, Torben Fetzer, Jason Rambach, Didier Stricker, Mahdi Saleh, Yan Di, Nassir Navab, Benjamin Busam, Federico Tombari, Yongliang Lin, Yu Zhang

The Best Methods on Individual Datasets

T-LESS, ITODD, YCB-V, HB: GDRNPP-PBRReal-RGBD-MModel

Xingyu Liu, Ruida Zhang, Chenyangguang Zhang, Bowen Fu, Jiwen Tang, Xiquan Liang, Jingyi Tang, Xiaotian Cheng, Yukang Zhang, Gu Wang, Xiangyang Ji

IC-BIN: RCVPose 3D_SingleModel_VIVO_PBR

Yangzheng Wu, Alireza Javaheri, Mohsen Zand, Michael Greenspan

LM-O: RADet+PFA-MixPBR-RGBD

Yang Hai, Rui Song, Zhiqiang Liu, Jiaojiao Li, Mathieu Salzmann, Pascal Fua, Yinlin Hu

TUD-L: Coupled Iterative Refinement

Lahav Lipson, Zachary Teed, Ankit Goyal, Jia Deng

Award money: \$100 per dataset