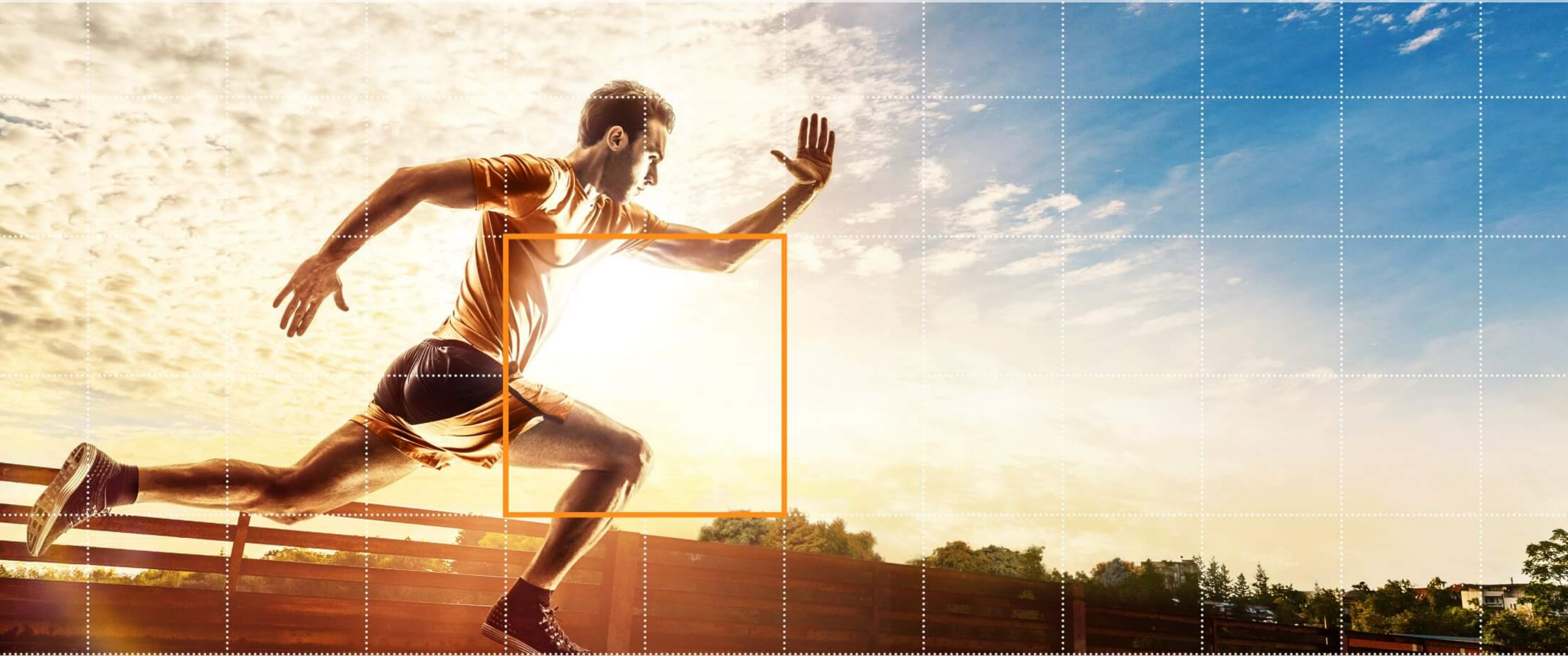


A Hybrid Approach for 6DoF Pose Estimation

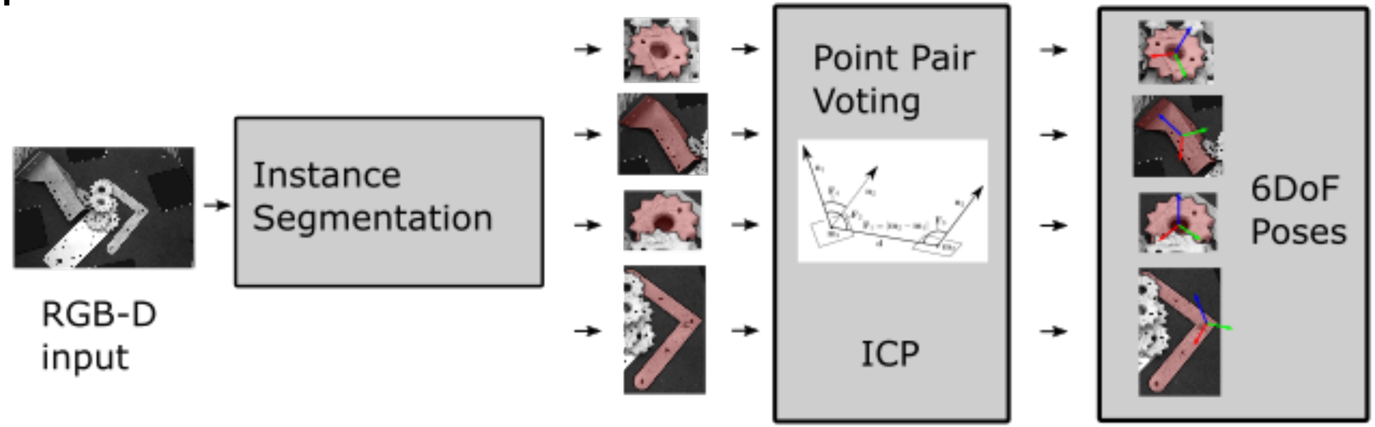
Rebecca König, Bertram Drost, MVTec Research – 6th International Workshop on Recovering 6D Object Pose – ECCV 2020

Motivation and Overview

Takeaway from BOP 2019:

- **Deep Learning-based methods: Fast, good in separating clutter from data, not-so-good pose estimation (yet)**
- **Voting with Point Pairs: Locally optimal pose estimation, slow global search**
- **DL-based methods are often two-stage methods: Object detector followed by pose estimation**

- **Our approach: Use DL-based instance segmentation to localize objects, followed by PPF-Voting for pose estimation**

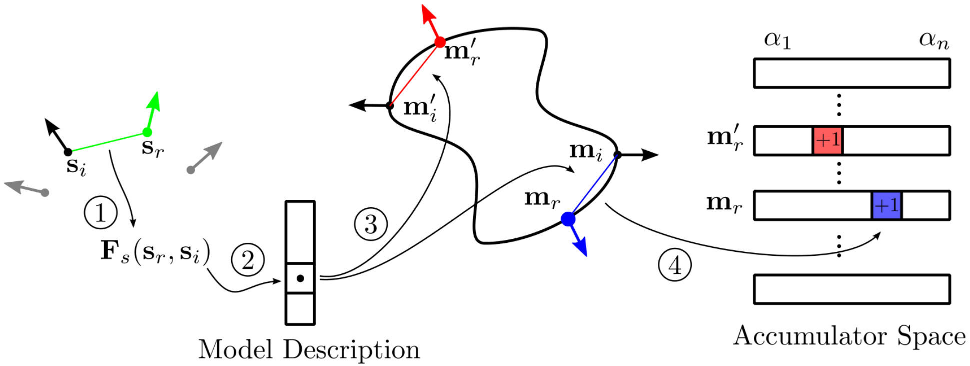


Instance Segmentation

- **High variance in datasets** (regarding training data, sensors, objects)
- **Train multiple networks, use the one with better validation error**
  - We use RetinaMask and MaskRCNN [2,3]
- **The main challenge is the training set**
  - **Partially large domain gap between training and test data for some datasets**
  - Different types of training data provided (none / CAD only, model cut-outs, synthetic images, real images)
  - PBR is a large step forward but does not fully close the domain gap
- **Our Approach**
  - Use real training images where available
  - Otherwise, augment validation / synthetic training images
    - Cut out objects, paste objects on COCO images, random scale / rotation / position
  - Use PBR images if it improves validation mAP
  - Online augmentation during training: Color variation, mirroring

Pose Estimation

- **Restrict search by using segmented instances and predicted classes**
- **Implementation of vanilla point pair voting [1]** (HALCON 20.05 progress)
  - Finds the locally best pose (largest geometric overlap)
  - Trained using CAD model only



- **Robust ICP, scoring and verification (on depth data only)**
- **Feature-point matching to resolve symmetries using texture [4]**

Results

Comparison to Baseline

Dataset	LM-O	T-LESS	TUD-L	IC-BIN	ITODD	HB	YCB-V	avg.	time
Drost et al. [1]	0.527	0.444	0.775	0.388	0.316	0.615	0.344	0.487	7.704s
Ours	0.631	0.655	0.920	0.430	0.483	0.651	0.701	0.639	0.633s

12 times faster
15% higher AR

Results

At time of submission (1 pm)…

	Date (UTC)	Method	Test image	AR_{Core}	AR_{LM-O}	AR_{T-LESS}	AR_{TUD-L}	AR_{IC-BIN}	AR_{ITODD}	AR_{HB}	AR_{YCB-V}	Time (s)
1	2020-08-19 10:19	Koenig-Hybrid-DL-PointPairs	RGB-D	0.639	0.631	0.655	0.920	0.430	0.483	0.651	0.701	0.633
2	2019-10-22 07:57	Vidal-Sensors18	D	0.569	0.582	0.538	0.876	0.393	0.435	0.706	0.450	3.220

…10 hours later

	Date (UTC)	Method	Test image	AR_{Core}	AR_{LM-O}	AR_{T-LESS}	AR_{TUD-L}	AR_{IC-BIN}	AR_{ITODD}	AR_{HB}	AR_{YCB-V}	Time (s)
1	2020-08-19	CosyPose-ECCV20-SYNT+REAL-1VIEW-ICP	RGB-D	0.698	0.714	0.701	0.939	0.647	0.313	0.712	0.861	13.743
2	2020-08-19	Koenig-Hybrid-DL-PointPairs	RGB-D	0.639	0.631	0.655	0.920	0.430	0.483	0.651	0.701	0.633
3	2020-08-18	CosyPose-ECCV20-SYNT+REAL-1VIEW	RGB	0.637	0.633	0.728	0.823	0.583	0.216	0.656	0.821	0.449

Conclusion

- **Good training data is vital**
  - **Mind the (domain) gap!**
  - **Practicability: from CAD model to training data?**

- **Automatic selection of method parameters based on validation error works**
  - and avoids dataset-specific parameters

- **Hybrid approaches that leverage advantages of learning and geometric approaches can (still?) reach state-of-the-art**

[1] Drost, B., Ulrich, M., Navab, N., Ilic, S.: Model globally, match locally: Efficient and robust 3d object recognition. In: CVPR (2010)
[2] Fu, C. Y., Shvets, M., & Berg, A. C. RetinaMask: Learning to predict masks improves state-of-the-art single-shot detection for free. arXiv:1901.03353
[3] He, K., Gkioxari, G., Dollár, P., & Girshick, R.: Mask R-CNN. ICCV 2017.
[4] Lepetit, V., Fua, P.: Keypoint recognition using randomized trees. T-PAMI 2006.