ECCV'20 ONLINE
23-28 AUGUST 2020

6th International Workshop on Recovering 6D Object Pose

CosyPose: Consistent multi-view multi-object 6D pose estimation

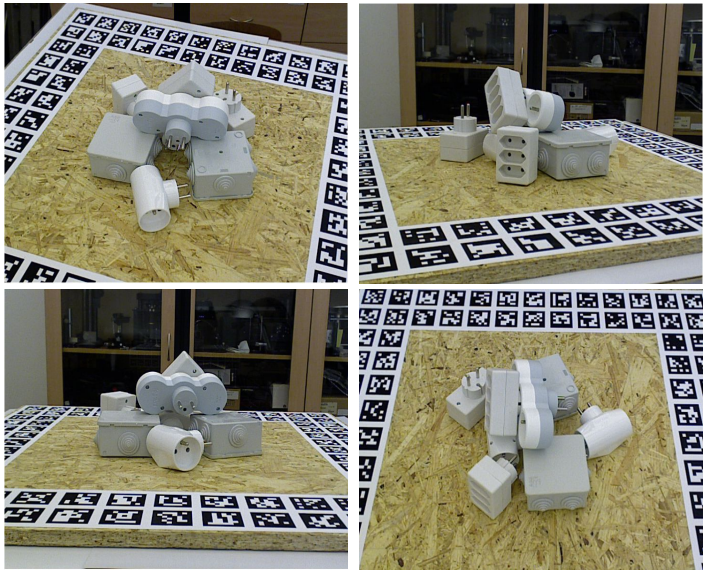
arXiv:2008.08465

Yann Labbé [1,2], Justin Carpentier [1,2], Mathieu Aubry [4], Josef Sivic [1,2,3]

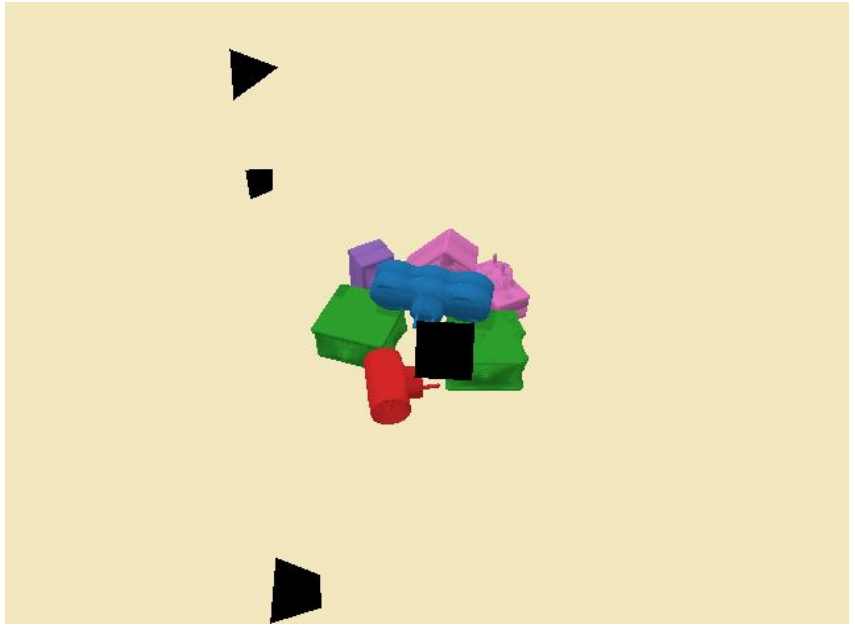
[1] Inria [2] DI ENS, PSL [3] CIIRC, CTU in Prague [4] ENPC

Multi-view 6D pose estimation

Input images



Output 3D scene

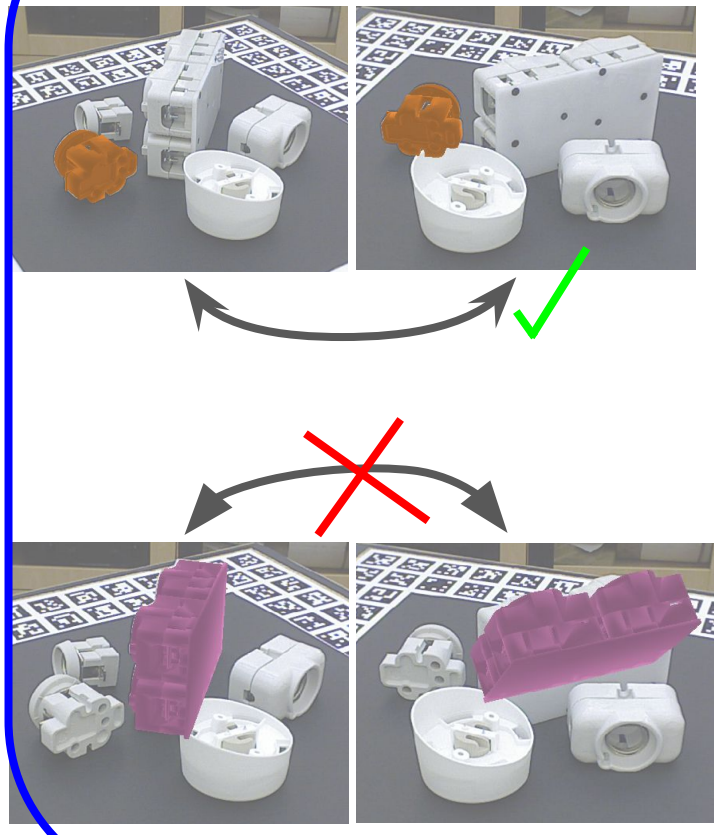


CosyPose: Approach overview

Single-view 6D pose estimation

BOP 20 Challenge

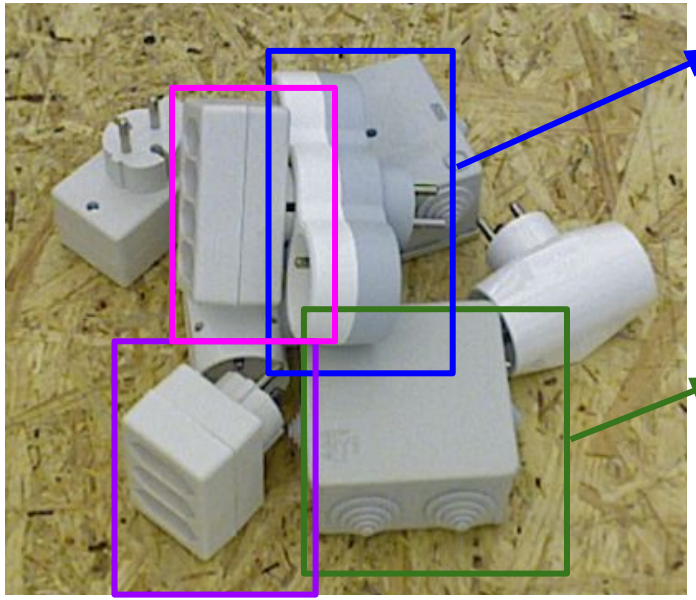
Robust multi-view multi-object reconstruction



...

Single-view CosyPose

Mask-RCNN 2D detections



Input RGB image

(only 3 networks trained per dataset)

2D detection

6D pose estimation

Coarse network → Refiner network

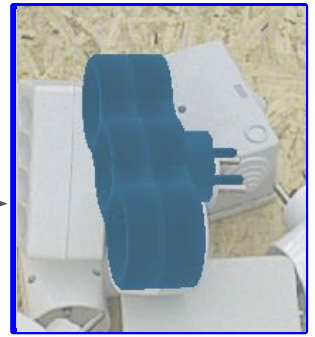
6D pose estimation

Coarse network → Refiner network

6D pose estimation

Coarse network → Refiner network

6D pose

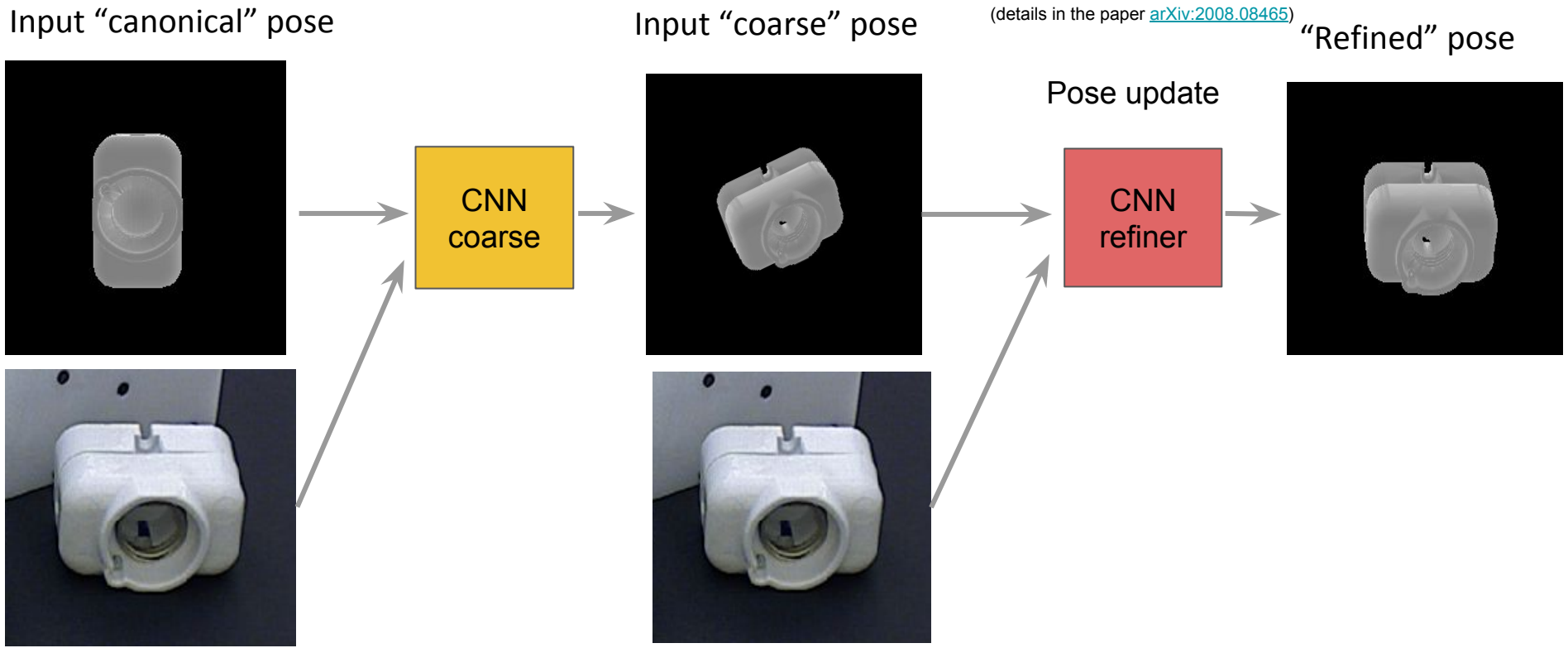


Pose estimation networks

DeepIM, Li et al, ECCV 2018

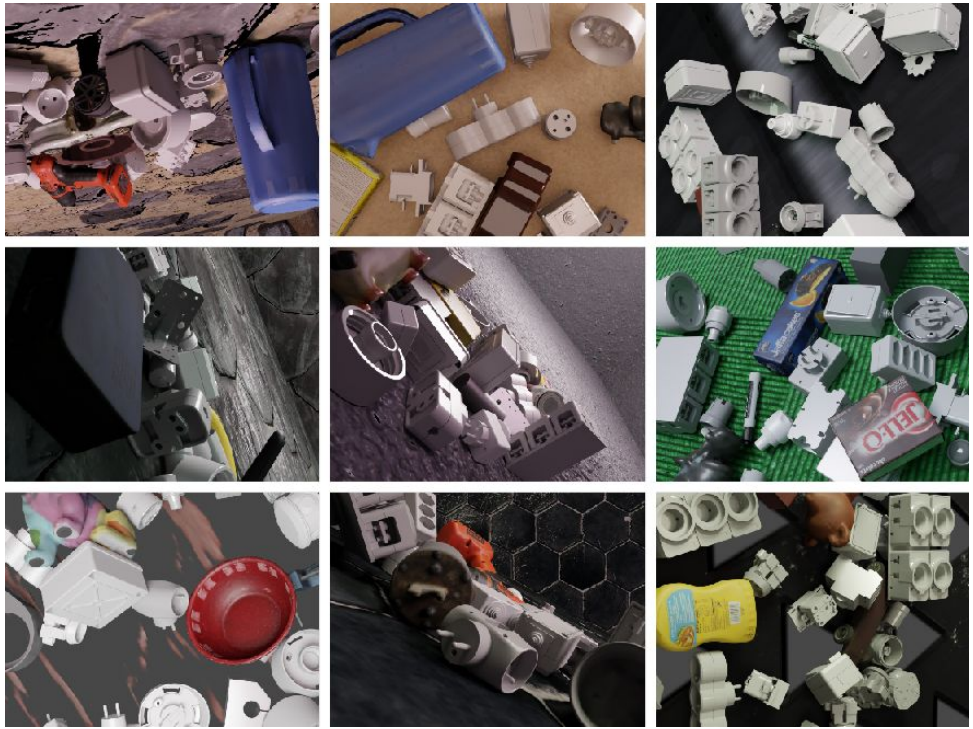
+ Network
+ Rotation parametrization
+ Loss
+ Data augmentation

(details in the paper arXiv:2008.08465)

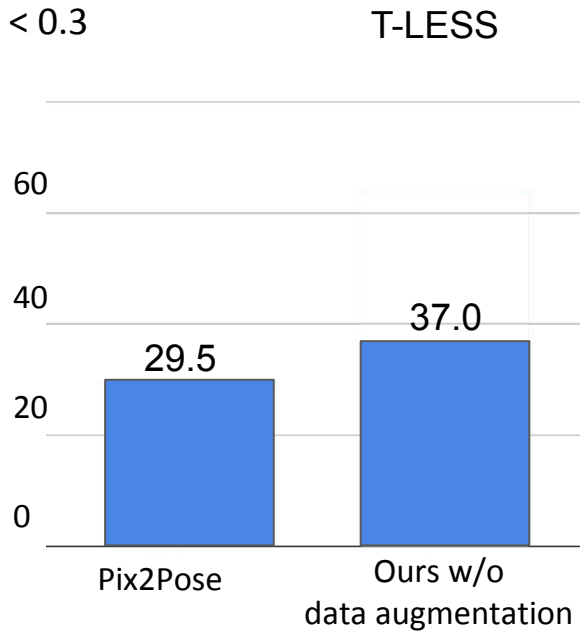


Key ingredients

Without data augmentation



$e_{vsd} < 0.3$

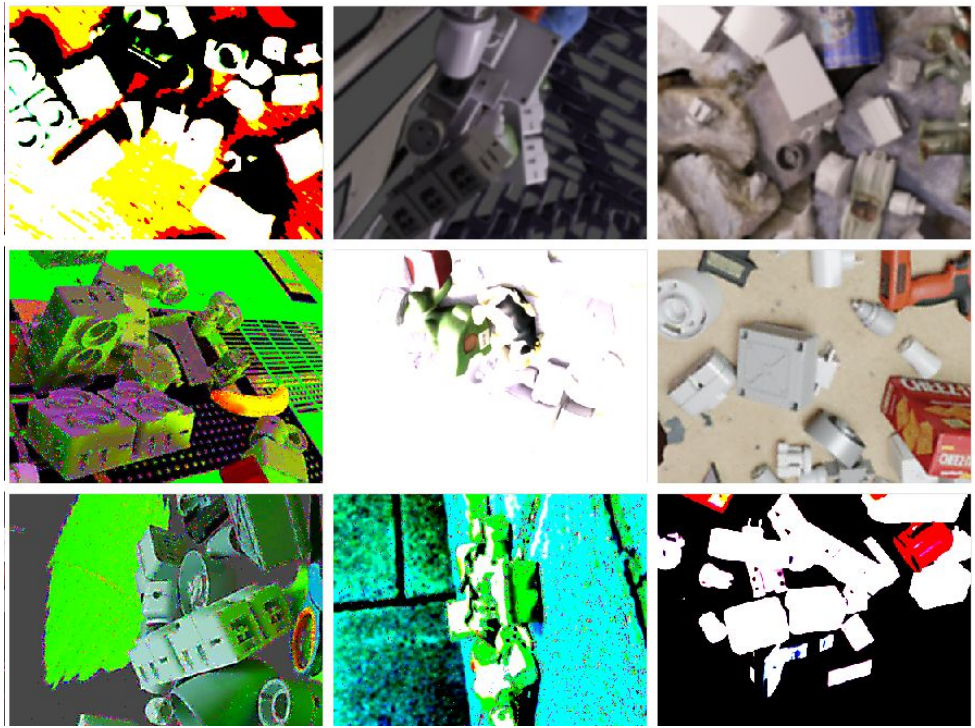


Pix2Pose, Park et al, ICCV 2019

(more ablations in the paper, Sec 3 Table 1b)

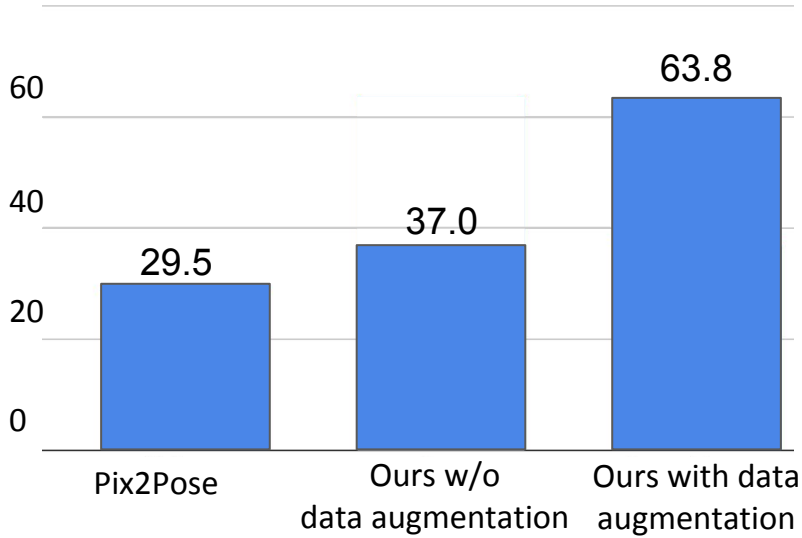
Key ingredients

With data augmentation



$e_{vsd} < 0.3$

T-LESS

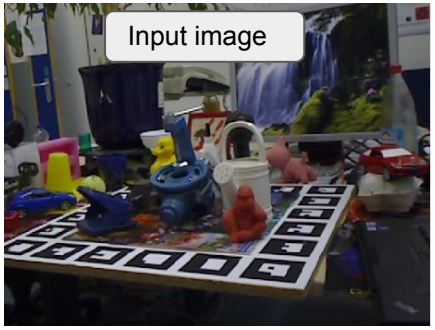


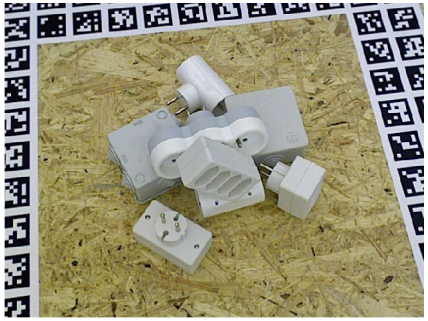
Pix2Pose, Park et al, ICCV 2019

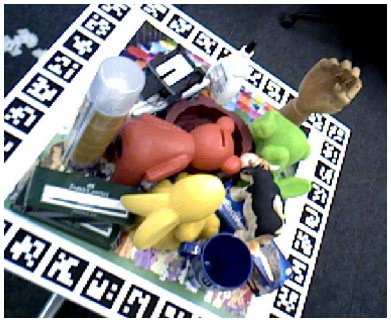
(more ablations in the paper, Sec 3 Table 1b)

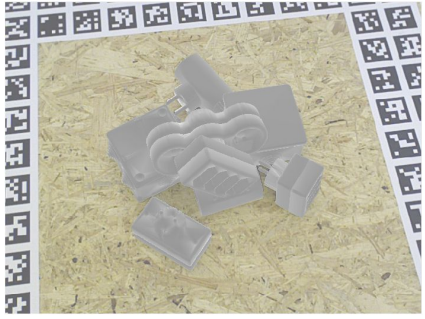
+ Access to a GPU cluster*
training 1 pose network: **~10 hours on 32 GPUs**

*Jean-zay, French national cluster managed by GENCI-IDRIS

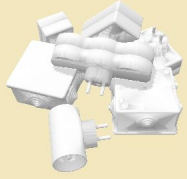




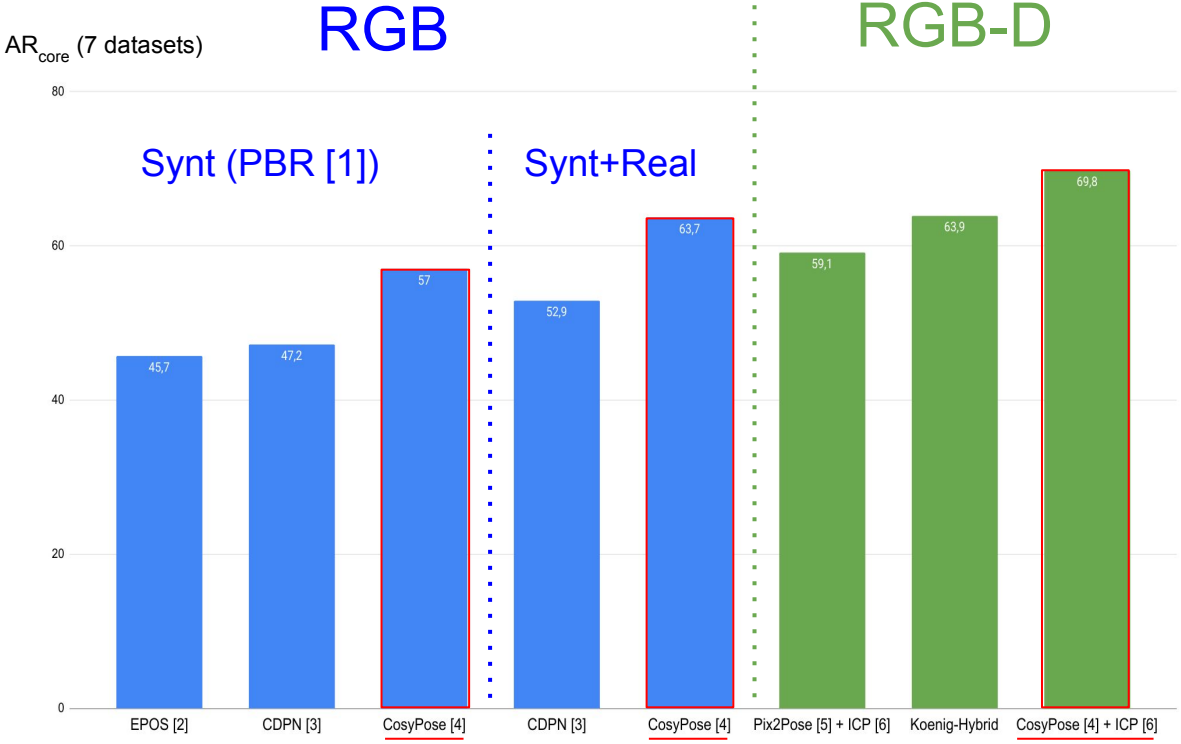




3D visualization



BOP20 results



\+ running time < 0.5s per image

[1] BlenderProc: Denninger, Sundermeyer, Winkelbauer, Olefir, Hodan, Zidan, Elbadrawy, Knauer, Katam, Lodhi in RSS workshops.

[2] EPOS, Hodan et al, CVPR 2020

[3] CDPN, Li et al, ICCV 2019

[4] CosyPose, Labbé et al, ECCV 2020

[5] Pix2Pose, Park et al, ICCV 2019

[6] https://github.com/kirumang/Pix2Pose

Code

- State-of-the-art pre-trained models for multiple datasets
- RGB single-view and multi-view modular framework
- Full training code

```python
# One or multiple RGB images
images = ...

# Define camera intrinsic parameters
intrinsics = ...

# 2D detections
detections = detector.get_detections(images)

# Single-view pose estimation
pose_predictions = pose_predictor.get_predictions(images, intrinsics, detections, n_refiner_iterations=4)

# Object-level multi-view reconstruction (if multiple images available)
scene_state = multiview_predictor.predict_scene_state(pose_predictions, intrinsics)
```

https://github.com/ylabbe/cosypose

ECCV'20 ONLINE
23-28 AUGUST 2020

6th International Workshop on Recovering 6D Object Pose

CosyPose: Consistent multi-view multi-object 6D pose estimation

arXiv:2008.08465

Yann Labbé [1,2], Justin Carpentier [1,2], Mathieu Aubry [4], Josef Sivic [1,2,3]

[1] Inria [2] DI ENS, PSL [3] CIIRC, CTU in Prague [4] ENPC

https://github.com/ylabbe/cosypose