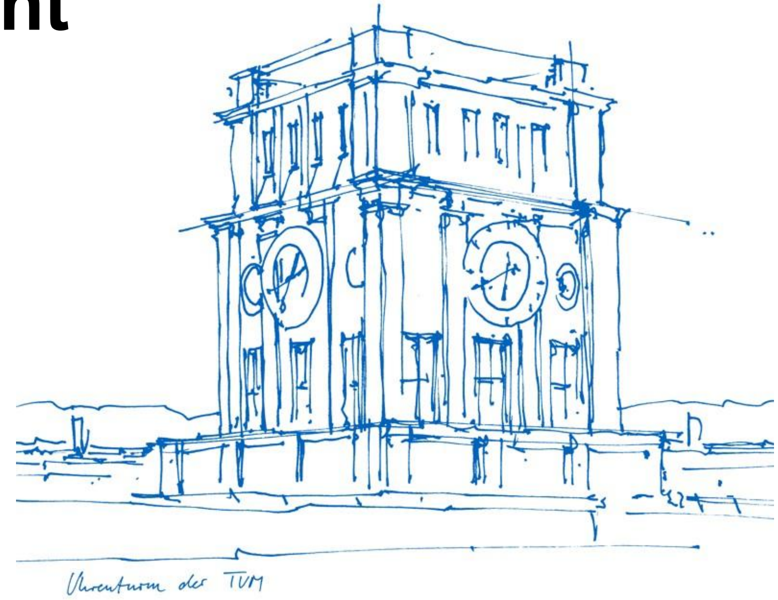
Chair for Computer Aided Medical Procedures & Augmented Reality
Department of Informatics
Technical University of Munich

# Category-level pose estimation with multi-view scale-aware refinement

**Project 5**
Zhuolun Zhou
Jaeyoung Cho

**Supervisor**
Ivan Shugurov



# Table of content

# Introduction

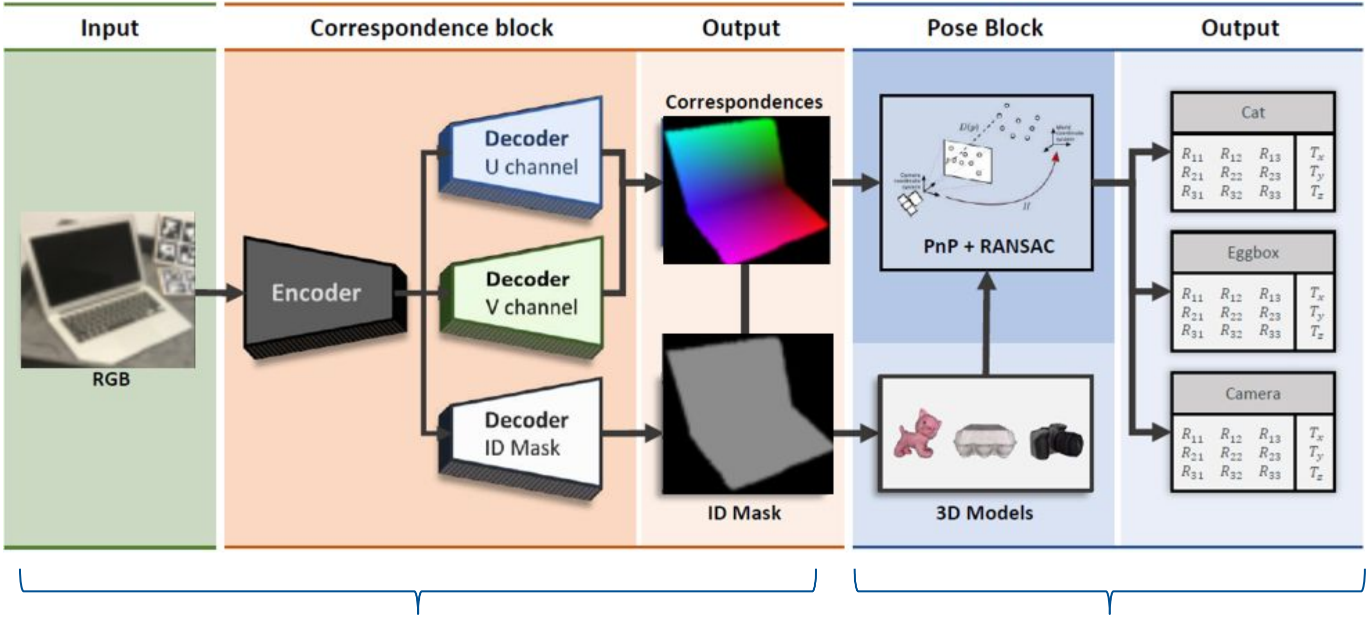
- Our task: estimate pose of an ***unseen*** object of a ***known*** class from RGB image, i.e. categorical pose estimation
- Our approach:

train

test on unseen object

pose?

# Methods: overview

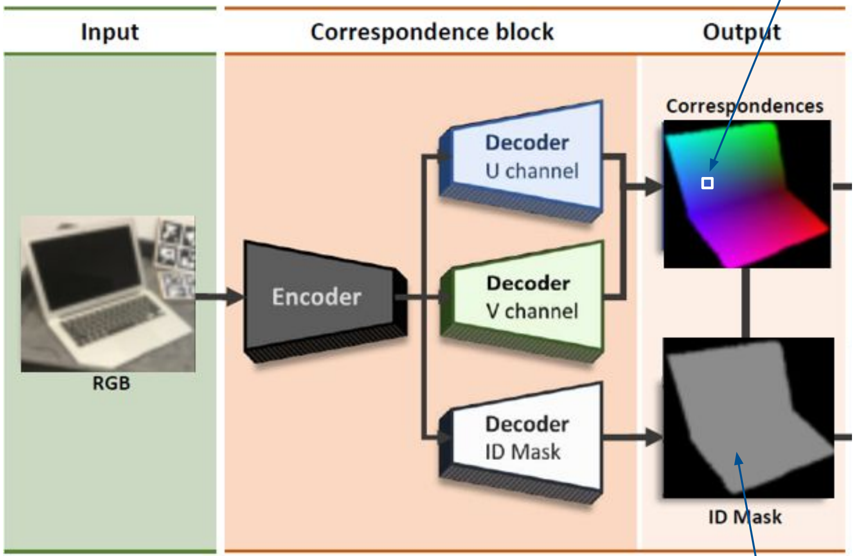



Deep learning:
predict the 2D-3D correspondences

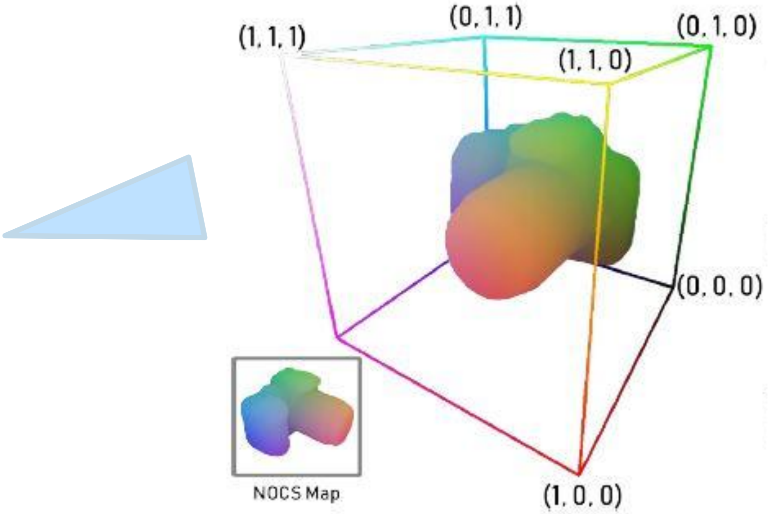
pose estimation with
PnP RANSAC

# Methods: NOCS map

(r, g, b) : normalize(x, y ,z)



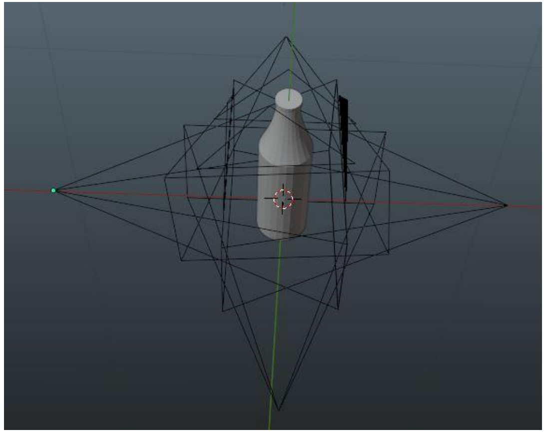

class ID



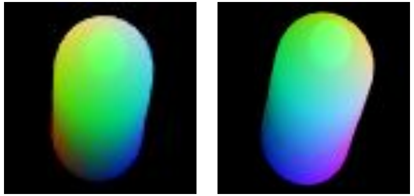

NOCS map: a RGB image encoding
2D-3D correspondences

# Methods: remap pose for symmetric objects

For symmetric objects, we map ground truth poses to a fixed pose around the symmetric axis. [4]



before remap



after remap

# Methods: MSPD metric

- Use MSPD metric to evaluate the quality of pose estimation
- MSPD (Multiple Symmetry-Aware Projection distance):

standard reprojection error:

$$\max_{x \in V_M} \left\| proj(\hat{P}x) - proj(Px) \right\|$$

$$e_{mspd} = (\hat{P}, P, S_M, V_M) = \min_{S \in S_M} \max_{x \in V_M} \left\| proj(\hat{P}x) - proj(PSx) \right\|$$

where

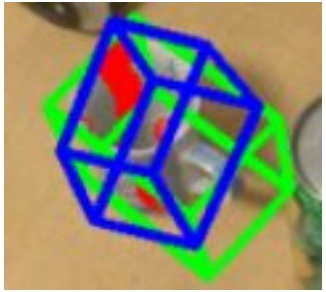
$P$ is the ground truth pose

$\hat{P}$ is the estimated pose

$S_M$ is a set of global symmetry transformations

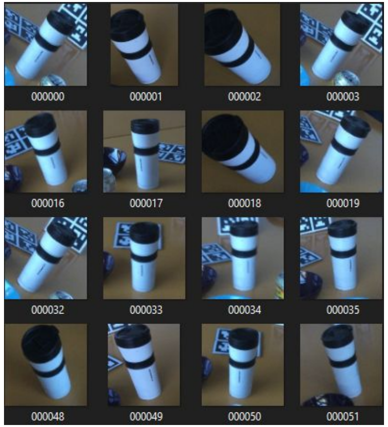
$V_M$ is a set of mesh vertices of object model

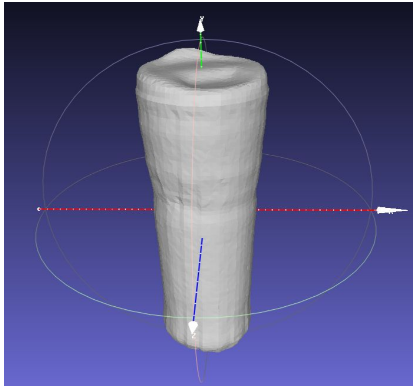
bottle, symmetric, low MSPD

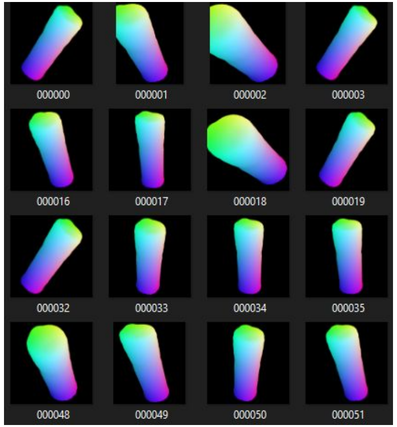


cup, non-symmetric, high MSPD

# Dataset: real train set

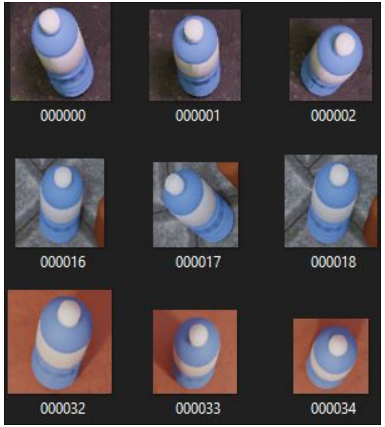


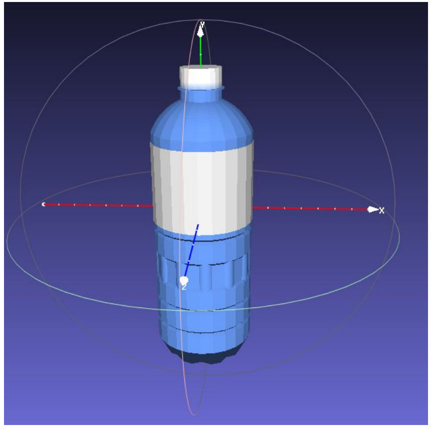






- 6 classes: bottle, bowl, can, camera, laptop, cup
- each class 3 different 3D models
- 14530 RGB images

# Dataset: synthetic train set



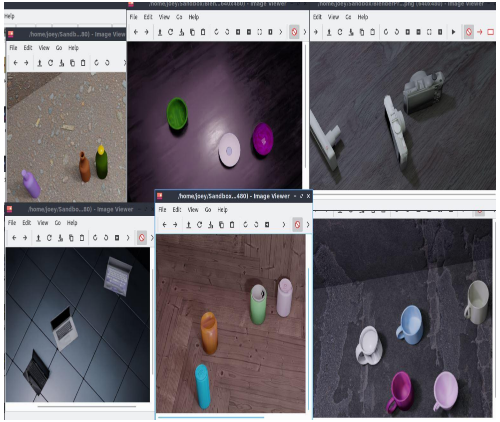



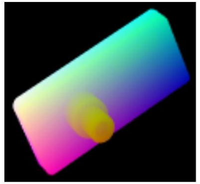

- 6 classes: bottle, bowl, can, camera, laptop, cup
- each class 6 different 3D models
- 14767 RGB images

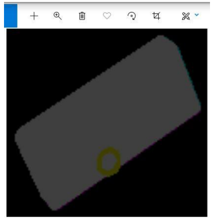
# Dataset: synthetic Data generation process

```
ply
format ascii 1.0
comment VCGLIB generated
element vertex 1504
property float x
property float y
property float z
property float nx
property float ny
property float nz
property uchar red
property uchar green
property uchar blue
property uchar alpha
element face 5604
property list uchar int vertex_indices
end_header
-0.038927 0.207289 0.145284 0 0 0 226 226 226 255
-0.075205 -0.033366 0.130258 0 0 0 226 226 226 255
```

3D model reformatting(to ply)





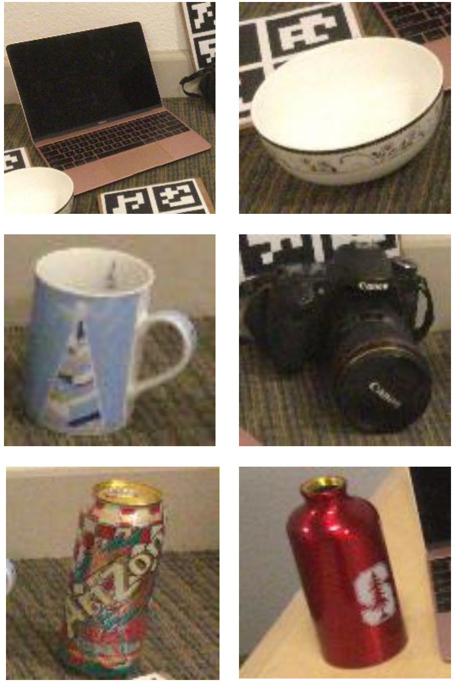


Blender ShapeNet → BenderProc → BOP format Dataset → NOCS Map, ID Mask

- models
- rgb
- scene_camera.json
- scene_gt.json

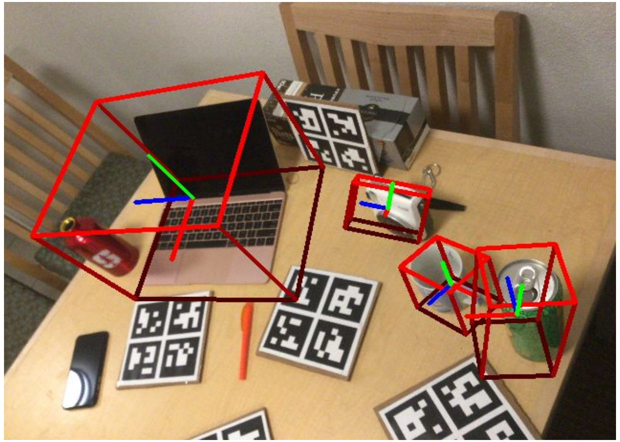
scene_camera.json: Intrinsic
scene_gt.json: Extrinsic

# Dataset: real testset

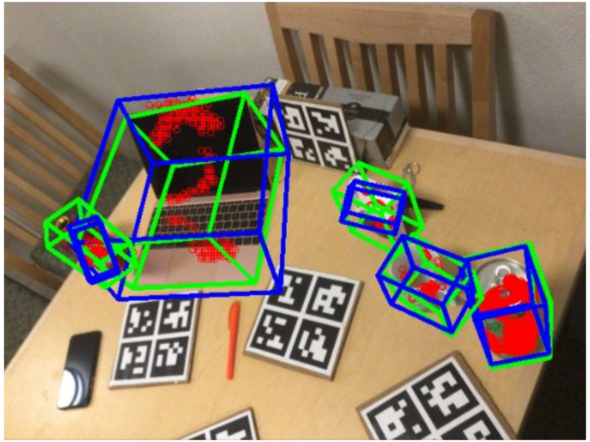


- 6 classes: bottle, bowl, can, camera, laptop, cup
- each class with ***unseen*** 3D models

# Qualitative results



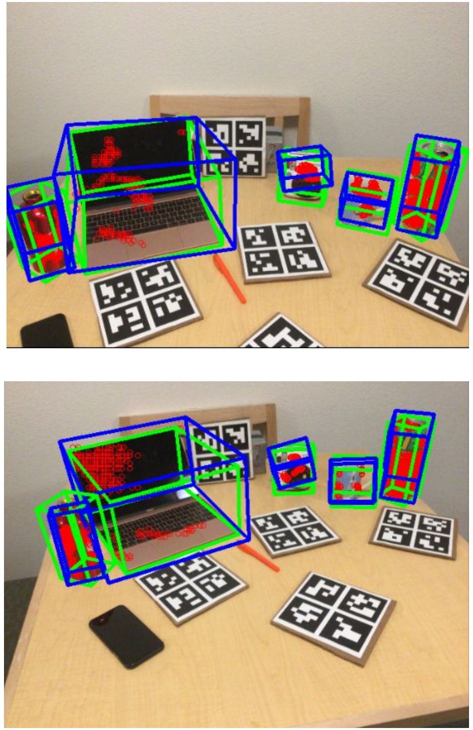
NOCS[2]



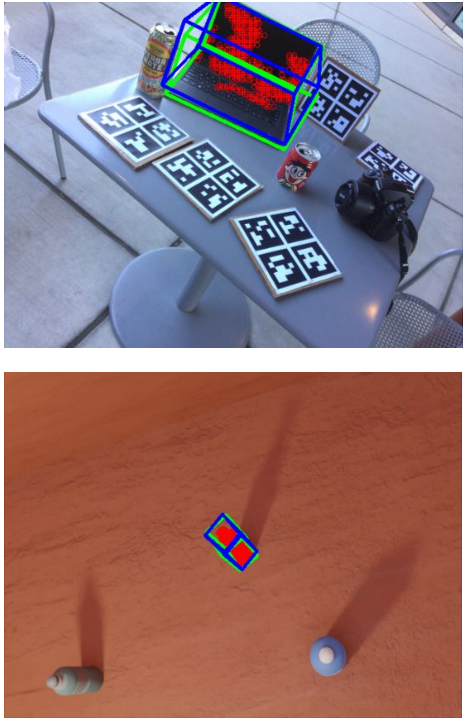
our method (green: ground truth pose, blue: estimated pose)

nocs map predicted by our method

# Qualitative results:

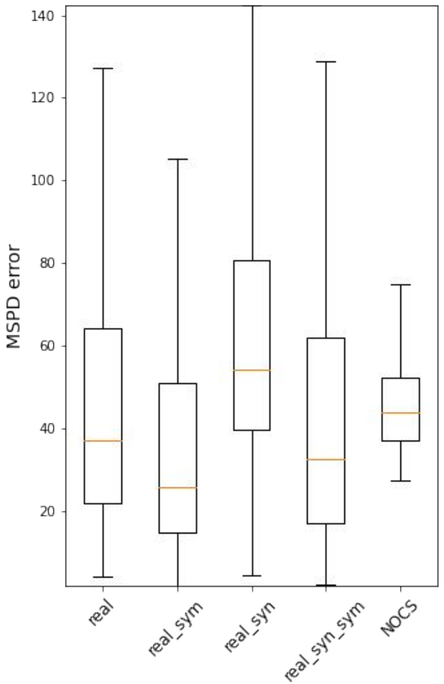


degraded performance on unseen objects

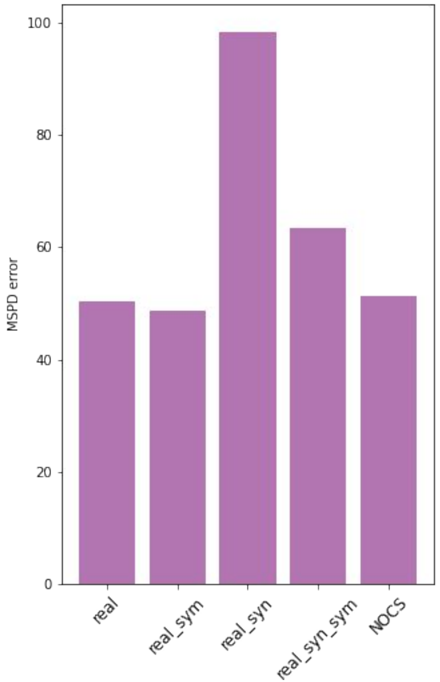


good performance on known objects

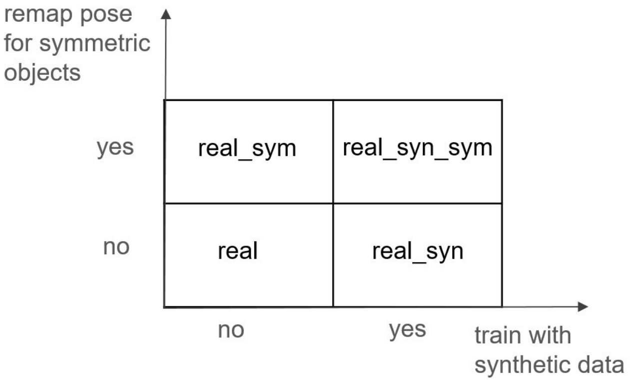
# Quantitative Results: average MSPD error




(a) box plot



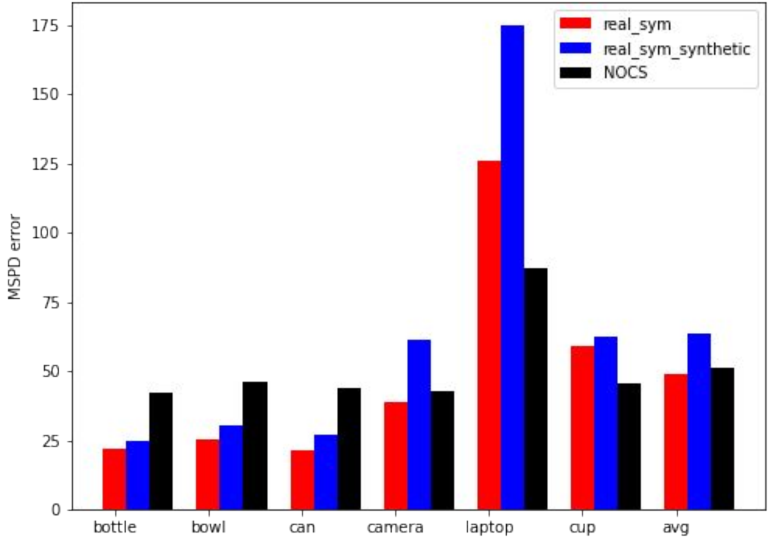

(b) bar plot



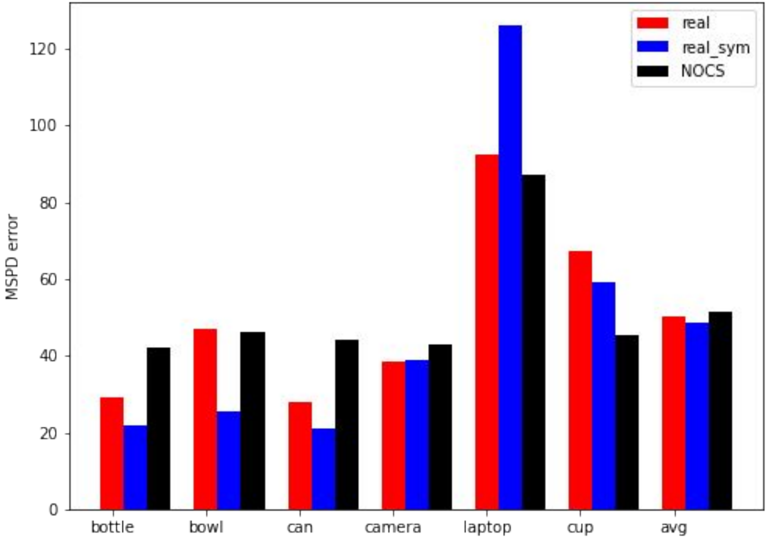

4 different train settings:
real, real_sym,
real_syn, real_syn_sym

Fig. Comparing the average MSPD error over all classes on test set in different train settings

# Quantitative Results: MSPD error per class




(a) effects of including synthetic data in trainset




(b) effects of remapping pose for symmetric objects

# Summary

- Implemented part of the DPOD[1] for categorical pose estimation
- Synthesized dataset for training
- Achieved comparable results with NOCS[2]

# Possible improvements:

- Generate more realistic data to boost performance on testdata
- Apply differential renderer in [3] to refine the pose estimation
- Use object detector (e.g. YOLO) to crop out single objects and determine the class of object

# Reference

[1] Sergey Zakharov, Ivan Shugurov, and Slobodan Ilic. DPOD: 6d pose object detector and rener. In Proceedings of the IEEE International Conference on Computer Vision, pages 1941-1950, 2019.

[2] He Wang, Srinath Sridhar, Jingwei Huang, Julien Valentin, Shuran Song, and Leonidas J Guibas. Normalized object coordinate space for category-level 6d object pose and size estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2642-2651, 2019.

[3] Ivan Shugurov, Ivan Pavlov, Sergey Zakharov and Slobodan Ilic. Multi-View Object Pose Refinement with Differentiable Renderer

[4] Pitteri, Giorgia, et al. "On object symmetries and 6d pose estimation from images." *2019 International Conference on 3D Vision (3DV)*. IEEE, 2019.

Thank you for your attention!

Questions?