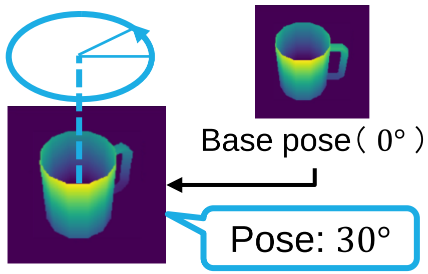
Median-shape Representation Learning for Category-level Object Pose Estimation in Cluttered Environments

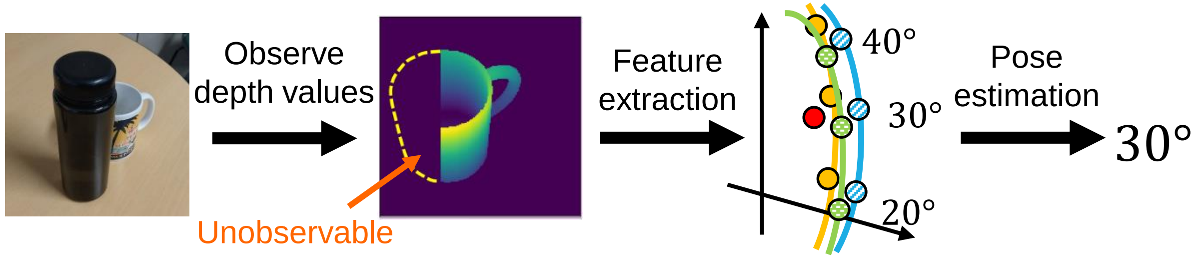
○ Hiroki Tatemichi* Yasutomo Kawanishi* Daisuke Deguchi* Ichiro Ide* Ayako Amma† Hiroshi Murase*
* Nagoya University † Toyota Motor Corporation

Goal: **Occlusion-robust category-level** pose estimation

- Input: Cropped depth image
  - Robust to variations in color and lighting conditions



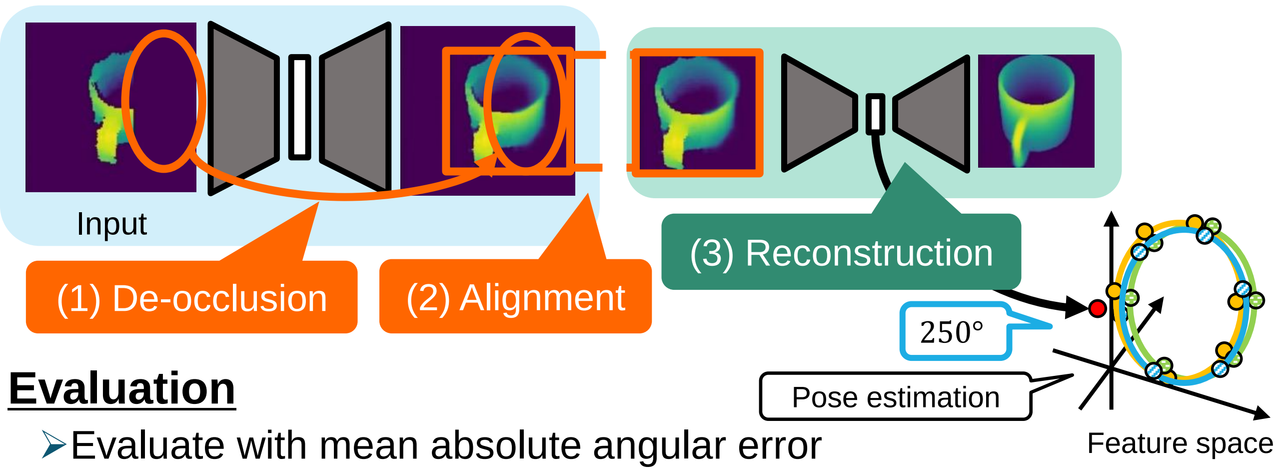
- Common approach for pose estimation



Difficulties & Approach

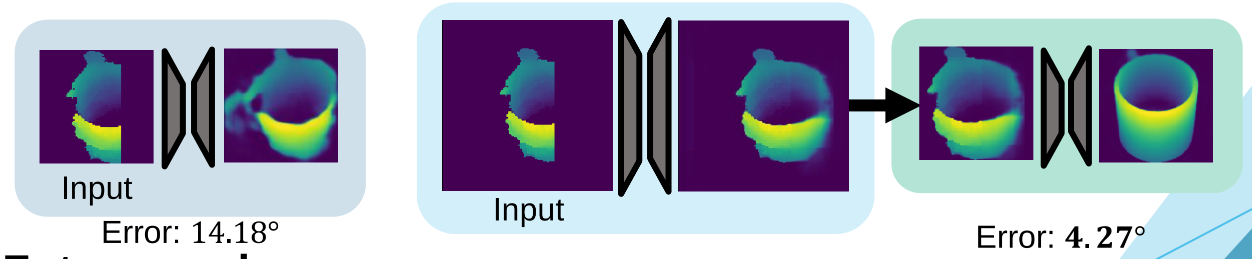
1. The occluded part of the object is unobservable
   ➡ De-occlude depth values of the occluded part
2. The true object center is shifted from the image center
   ➡ Estimate the occluded part and the offset
3. Shape variation within a category
   ➡ Recostruct the median-shaped object in the category

Proposed Method Two-stage Encoder-Decoder model



Evaluation

- Evaluate with mean absolute angular error
- Proposed method achieves the best performance (with an error of 4.27°)



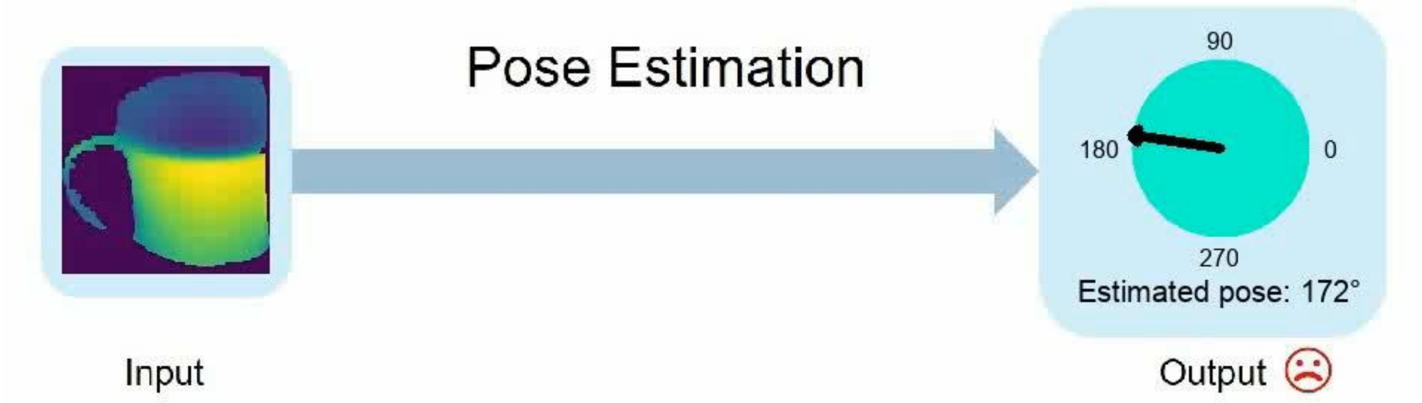
Future work

- Handle complex occlusions
- Extend to 3D axes rotations

[1] M. Sundermeyer et al., "Implicit 3D orientation learning for 6D object detection from RGB images", ECCV, 2018.

Email

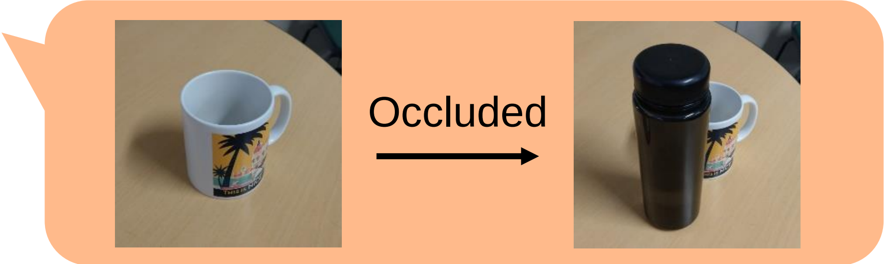
Goal: Realize **occlusion-robust** pose estimation from a depth image in the **category-level**



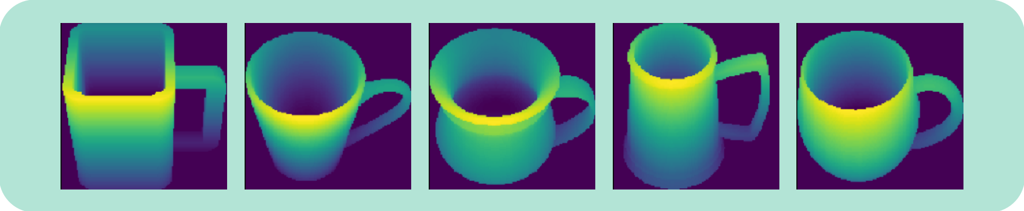
Motivation

- Grasping of objects by a robot
  - Observe an object with a depth image sensor
  - Determine the grip location of the object by pose estimation

- Difficulties in an object pose estimation task
  - Objects are often occluded in cluttered environments



  - *Instance-level* vs. *category-level*
    - *Category-level* estimation is more difficult owing to shape variations within the category

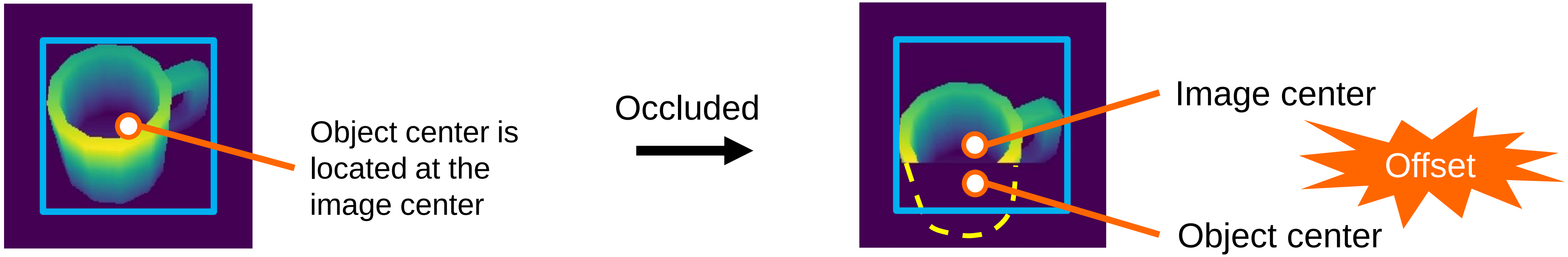


Difficulties & Approach

1. Occluded part of the object is unobservable

   ➡ De-occlude depth values of the occluded part[1]

2. True object center is shifted from the detected image center



   ➡ Estimate the occluded part and the offset

3. Shape variation within a category

   ➡ Recostruct a median-shaped object in the category

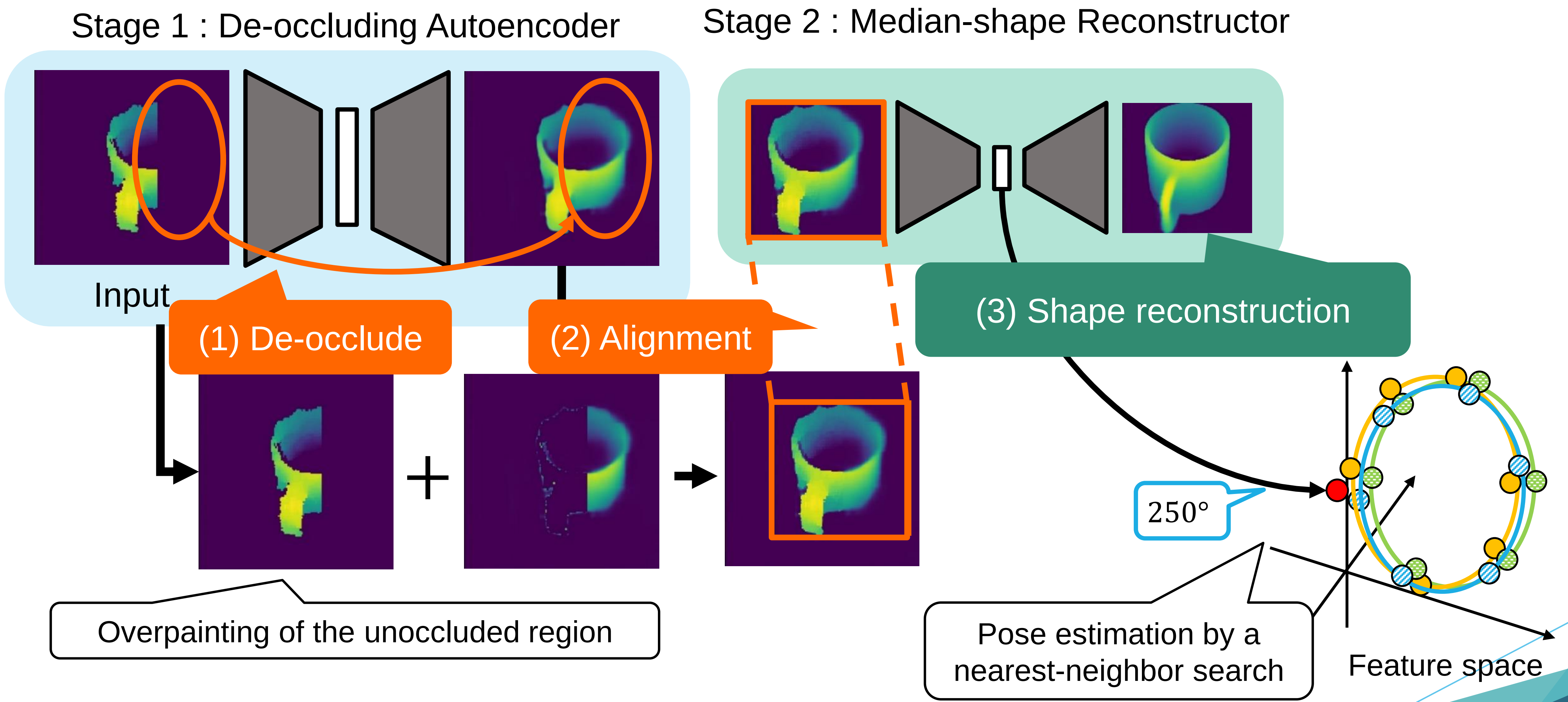
We propose a two-stage Encoder-Decoder model based on the above approaches

[1] M. Sundermeyer et al., “Implicit 3D orientation learning for 6D object detection from RGB images”, ECCV, 2018.

Proposed method (1/2)

◆Two-stage Encoder-Decoder model

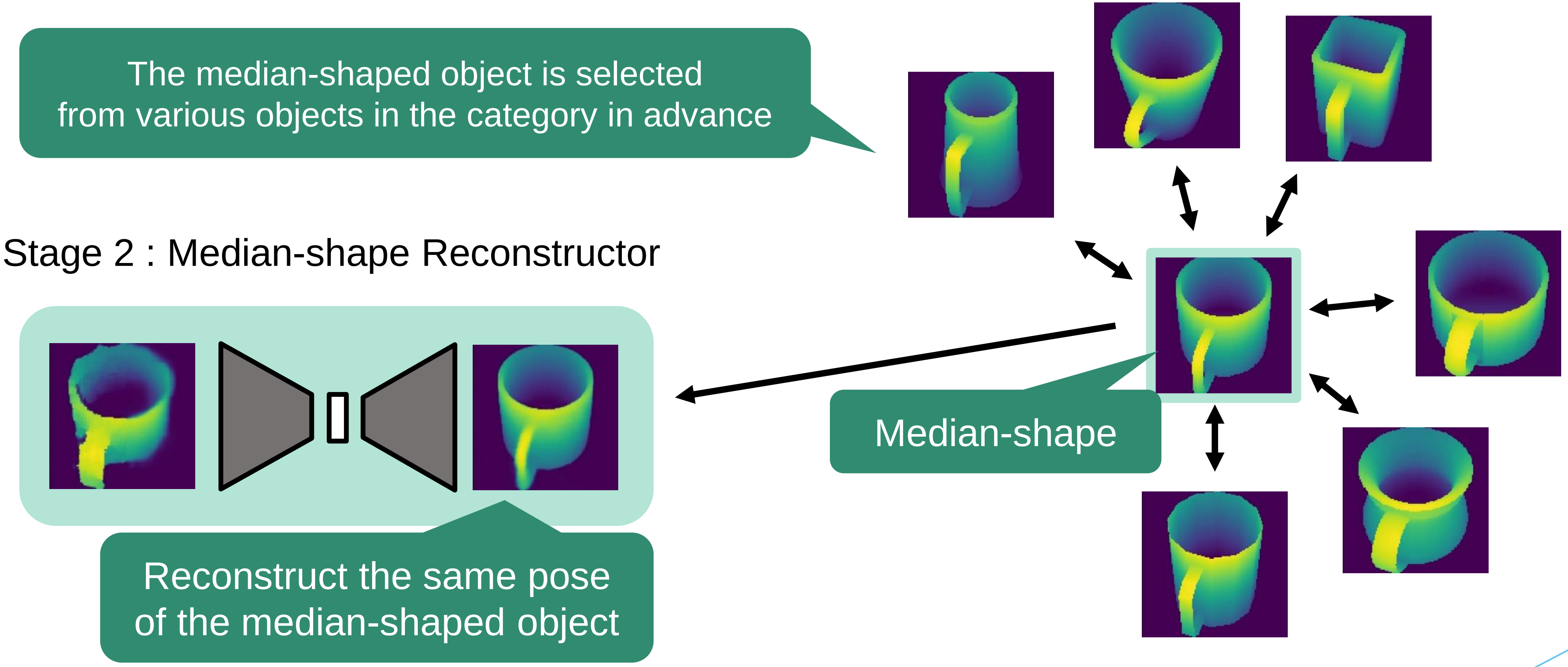
➢Extract features of a de-occluded object whose center is aligned to the image center



Proposed method (2/2)

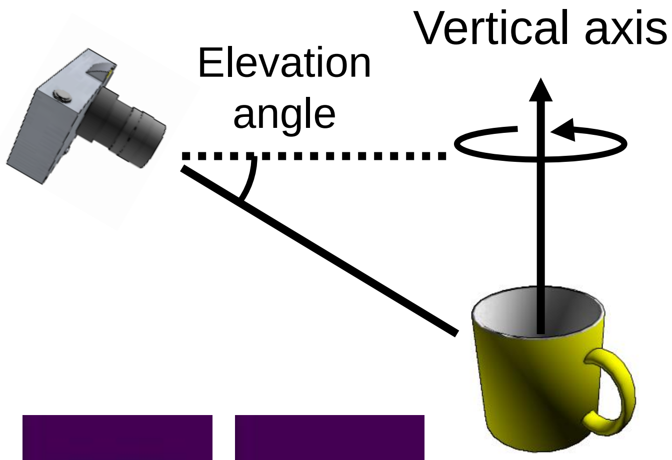
◆Median-shape Reconstructor

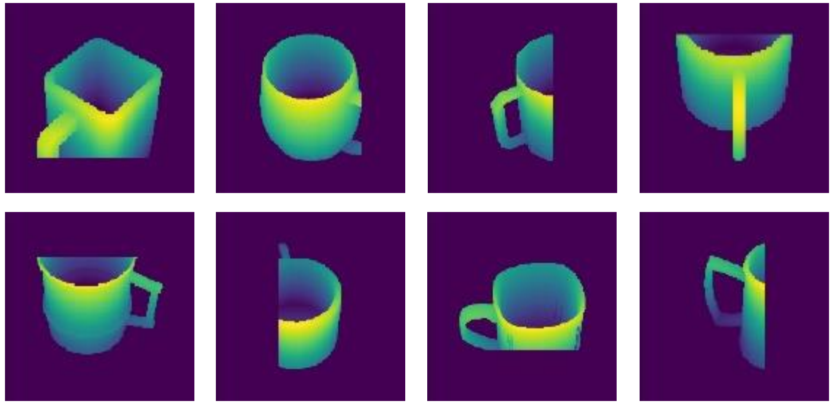
- Absorb shape variations in a category



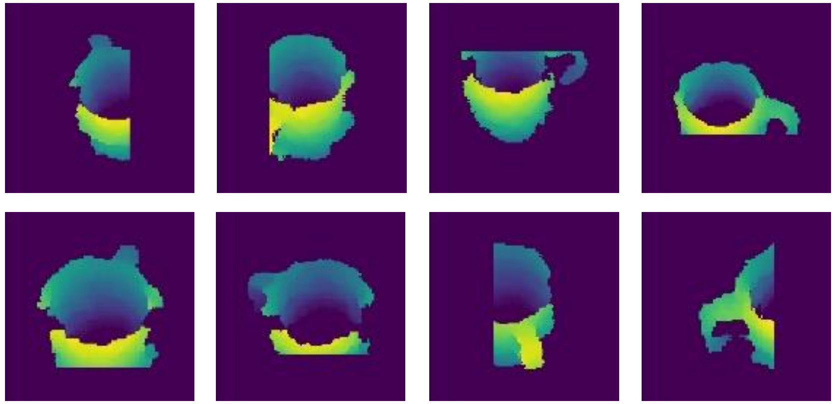
Evaluation (1/2)

- Estimated the pose around the vertical axis with an interval of 1°
- Used the below datasets we prepared
  - Training: Large-scale virtual dataset + Real dataset
  - Evaluation: Real dataset





Large-scale virtual dataset



Real dataset

Variously-shaped objects for evaluation

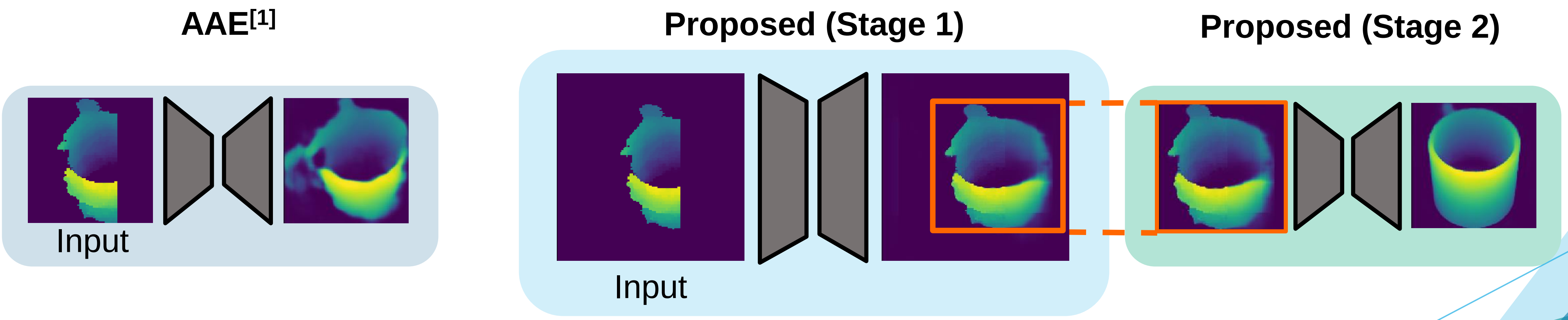
Evaluation (2/2)

◆Pose estimation results

➢The proposed method achieves the best performance

	Mean absolute angular error ↓		Ratio of absolute angle error within 5° ↑	
Elevation angle	30°	50°	30°	50°
AAE[1] (Auto Encoder which only de-occludes)	17.85°	14.18°	43.6%	61.3%
Proposed	**5.15°**	**4.27°**	**61.7%**	**74.2%**

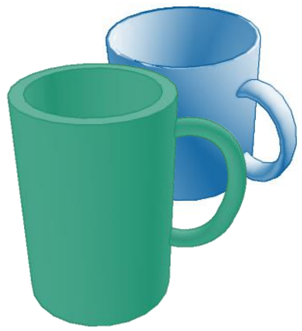
◆Visualization results



[1] M. Sundermeyer et al., "Implicit 3D orientation learning for 6D object detection from RGB images", ECCV, 2018.

Conclusion

- Proposed a *category-level* occlusion-robust pose estimation method
  - **Two-stage Encoder-Decoder model** to extract features of a de-occluded object whose center is aligned to the image center
  - **Median-shape Reconstructor** to absorb shape variations in a category
- Demonstrated the performance of the proposed method by evaluating it using a large-scale virtual dataset and a real dataset
- Future work
  - Handle complex occlusions
  - Extend the rotation to 3D axes



Complex occlusion