Category-Level 6D Pose and Size Estimation

Ran Ding

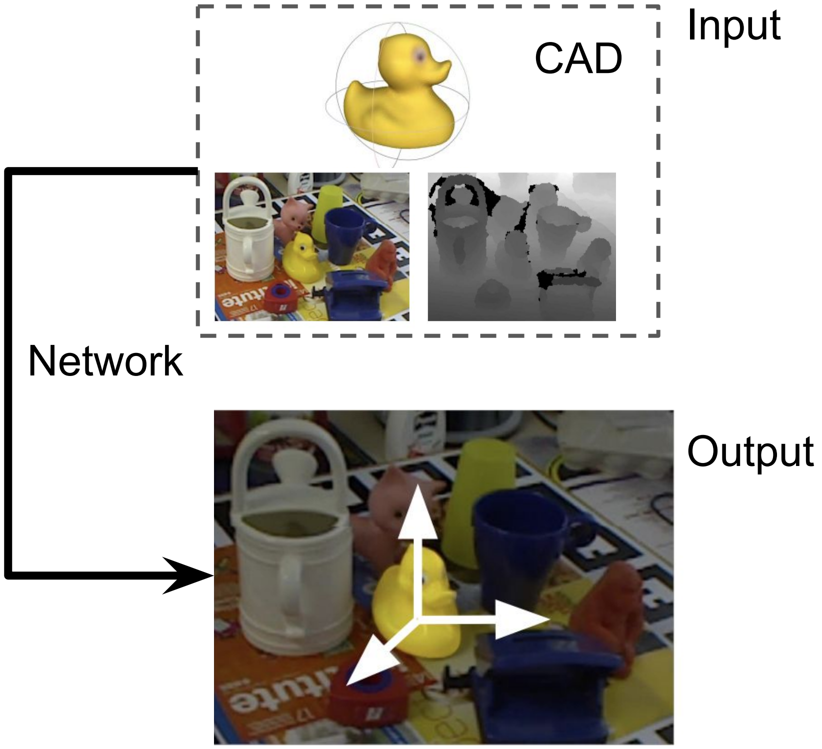
Advisor: Jiaxin Wei

Overview

1. Instance-Level 6D Pose and Size Estimation

- Task
  - Input: single RGB/RGB-D image and CAD model
  - Output: 6D pose and 3D size of the known object

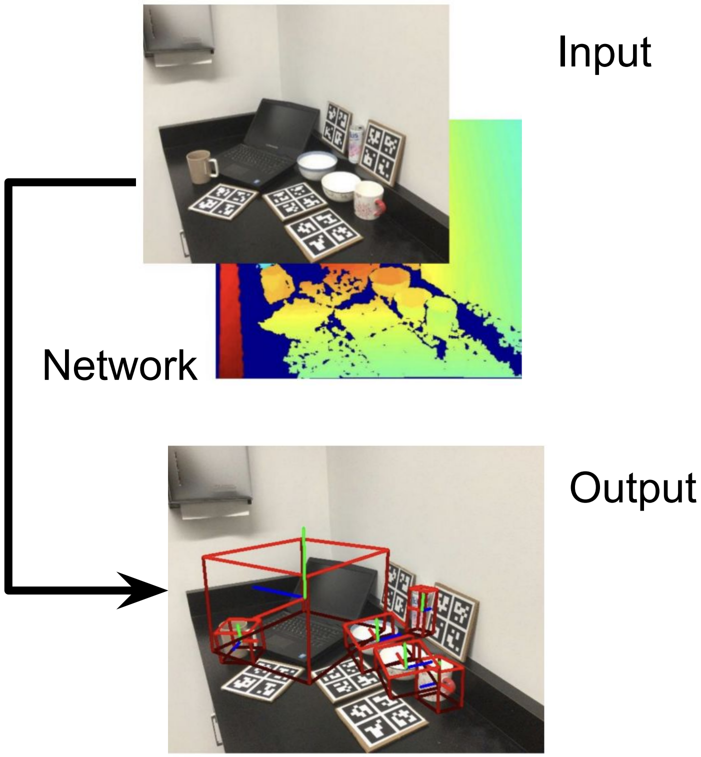
- Limitation
  - CAD model needed
  - Can’t generalize to **unseen** objects



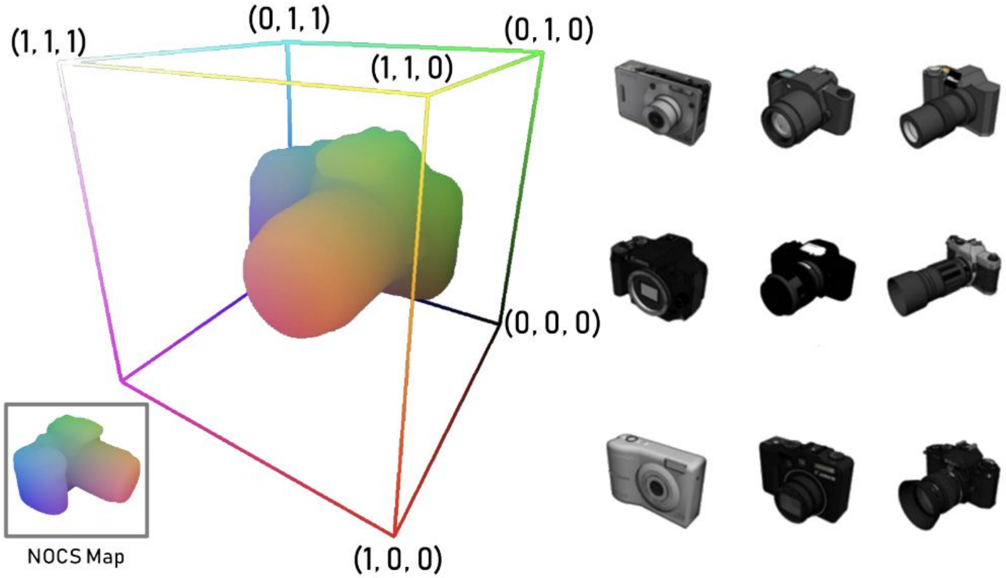
1. Category-Level 6D Pose and Size Estimation

- Task
  - Input: single RGB/RGB-D image
  - Output: 6D pose and 3D size of unknown object of the known category
  - Wide Applications in robotics and VR/AR

- Challenges
  - Intra-class shape variation
  - Lack large scale dataset



2. Normalized Object Coordinate Space (NOCS)



Rotation and Translation Recovery:

$$\hat{P} = P - \bar{P}, \hat{Q} = Q - \bar{Q}$$

$$H = \hat{P}\hat{Q}^T, H = U\Sigma V^T$$

$$R = VU^T$$

$$T = \bar{Q} - R\bar{P}$$

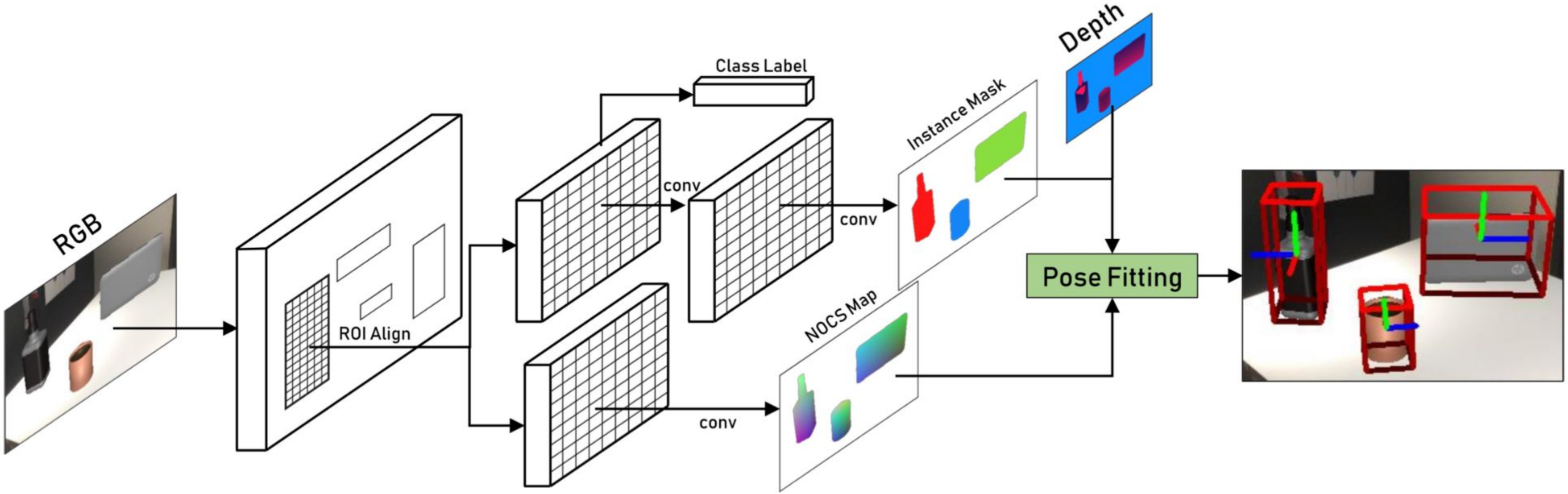
NOCS Regression Loss:

$$L(\mathbf{y}, \mathbf{y}^*) = \frac{1}{n} \begin{cases} 5\,(\mathbf{y} - \mathbf{y}^*)^2, & |\mathbf{y} - \mathbf{y}^*| \leq 0.1 \\ |\mathbf{y} - \mathbf{y}^*| - 0.05, & |\mathbf{y} - \mathbf{y}^*| > 0.1 \end{cases},$$

$$\forall \mathbf{y} \in N, \mathbf{y}^* \in N_p,$$

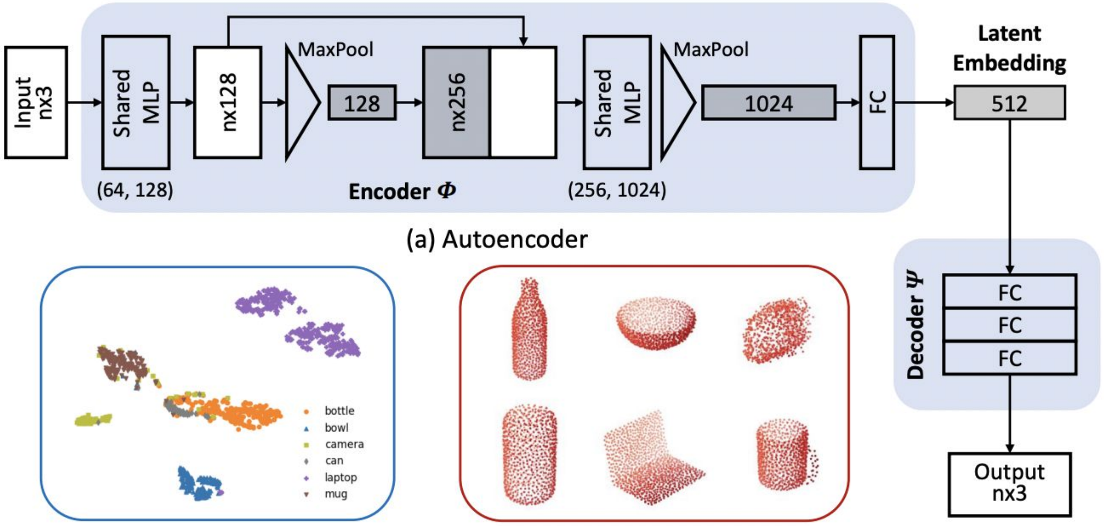
[1].Wang, He, et al. "Normalized object coordinate space for category-level 6d object pose and size estimation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.

2. Normalized Object Coordinate Space (NOCS)



[1].Wang, He, et al. "Normalized object coordinate space for category-level 6d object pose and size estimation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.

2. Shape Prior Deformation (SPD)

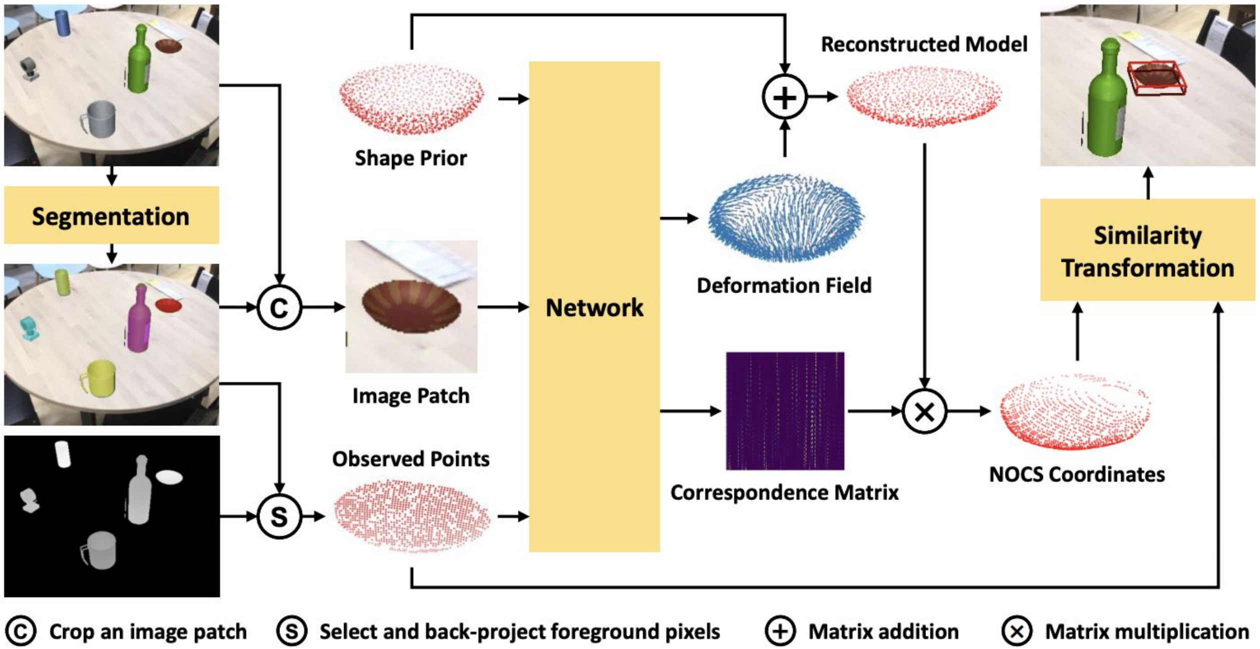


(b) Latent Embeddings
(c) Shape Priors

$$d_{\mathrm{CD}}(M_c^i, \hat{M}_c^i) = \sum_{x \in M_c^i} \min_{y \in \hat{M}_c^i} \|x - y\|_2^2 + \sum_{y \in \hat{M}_c^i} \min_{x \in M_c^i} \|x - y\|_2^2. \qquad (2)$$

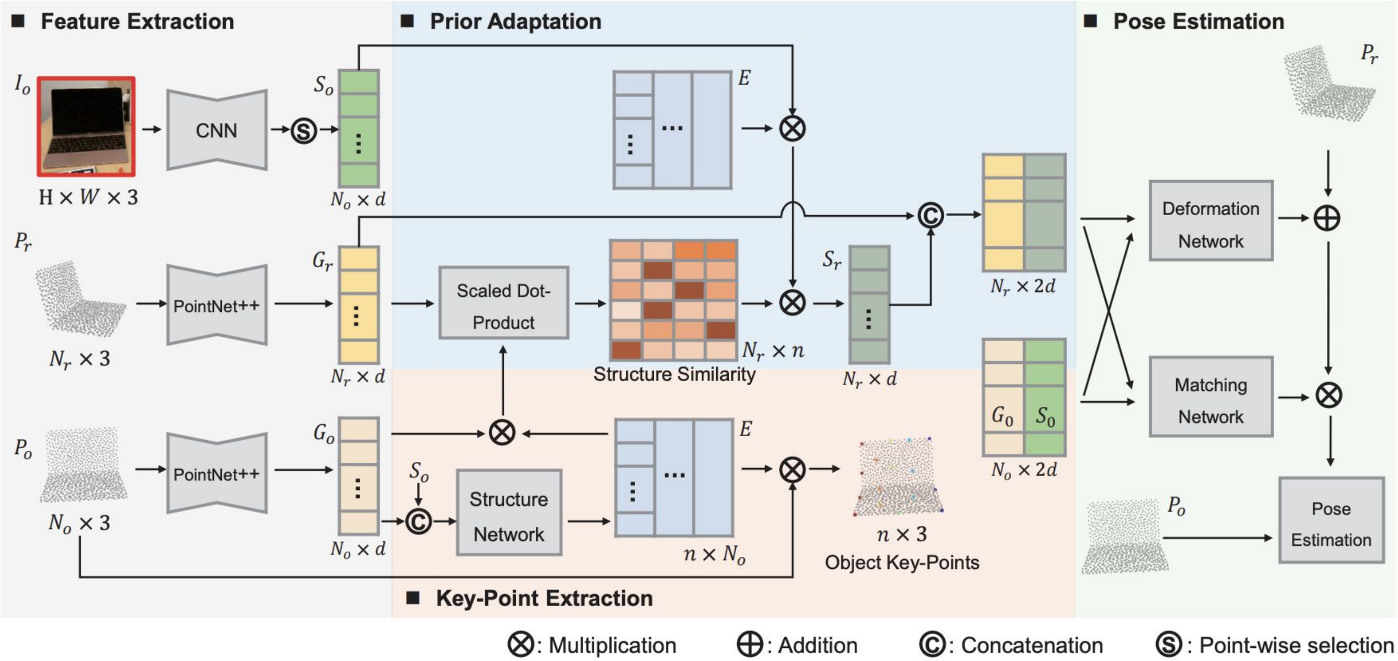
[2]. Tian, Meng, Marcelo H. Ang, and Gim Hee Lee. "Shape prior deformation for categorical 6d object pose and size estimation." Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020,

2. Shape Prior Deformation (SPD)



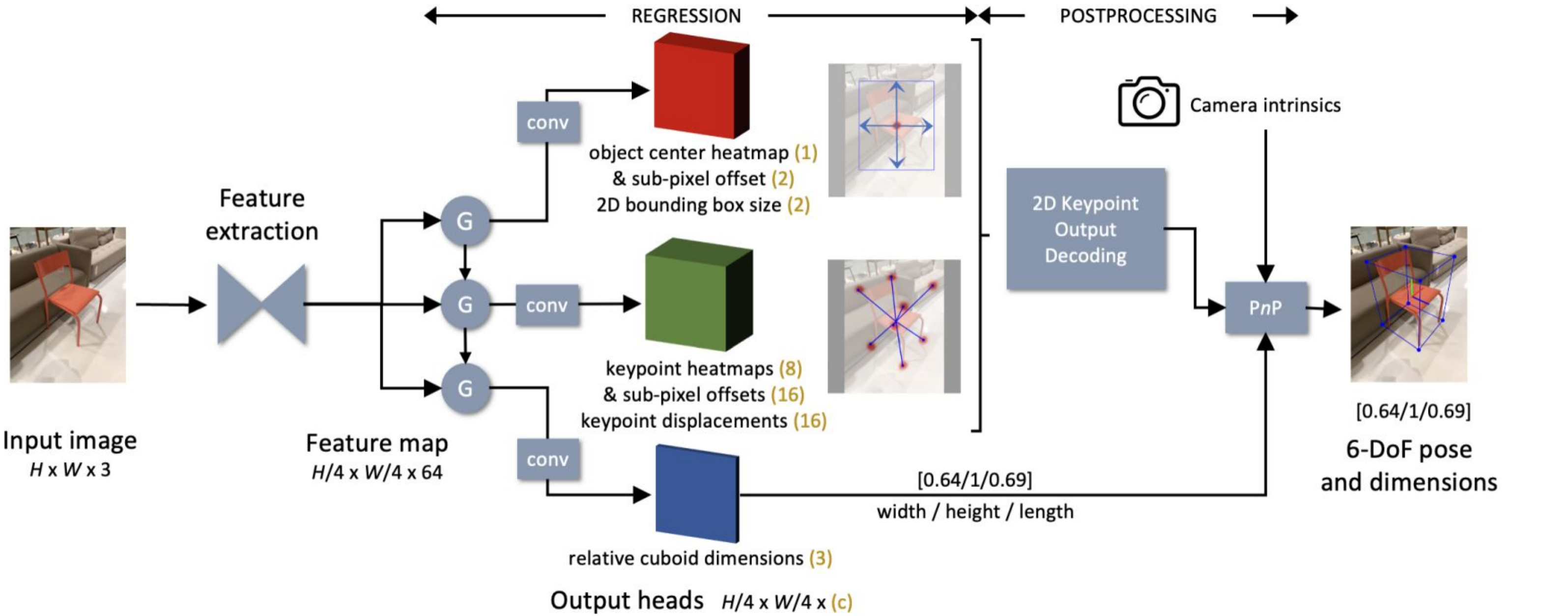
[2]. Tian, Meng, Marcelo H. Ang, and Gim Hee Lee. "Shape prior deformation for categorical 6d object pose and size estimation." Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020,

2. Structure-Guided Prior Adaptation (SGPA)



[3]. Chen, Kai, and Qi Dou. "Sgpa: Structure-guided prior adaptation for category-level 6d object pose estimation." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

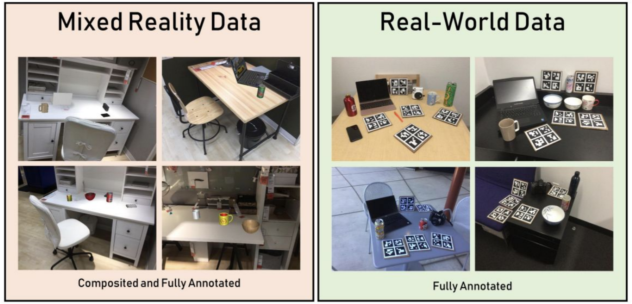
3. RGB Based Category-Level Object Pose Estimation



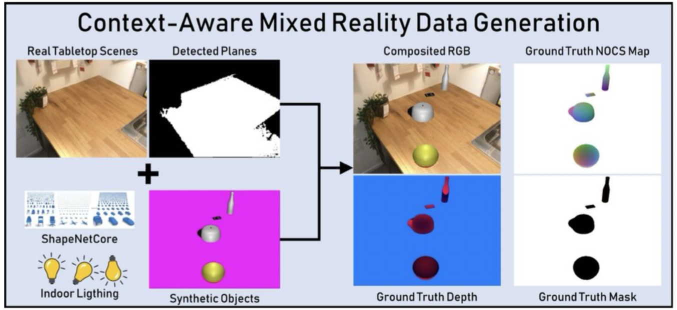
[4]. Lin, Yunzhi, et al. "Single-stage keypoint-based category-level object pose estimation from an RGB image." 2022 International Conference on Robotics and Automation (ICRA). IEEE, 2022.

4. Experiments

- Common Benchmark
  - CAMERA25 (6 Categories, 184 Instances, 300K images), REAL275 (6 Categories, 24 Instances, 6K images)
- Evaluation Metrics
  - 3D Size Estimation: IoU (25, 50, 75)
  - 6D Pose Estimation: Rotation Error, Translation Error (5 degree, 10 degree, 5cm, 10cm)



CAMERA25 and REAL75 Data



Data Generation

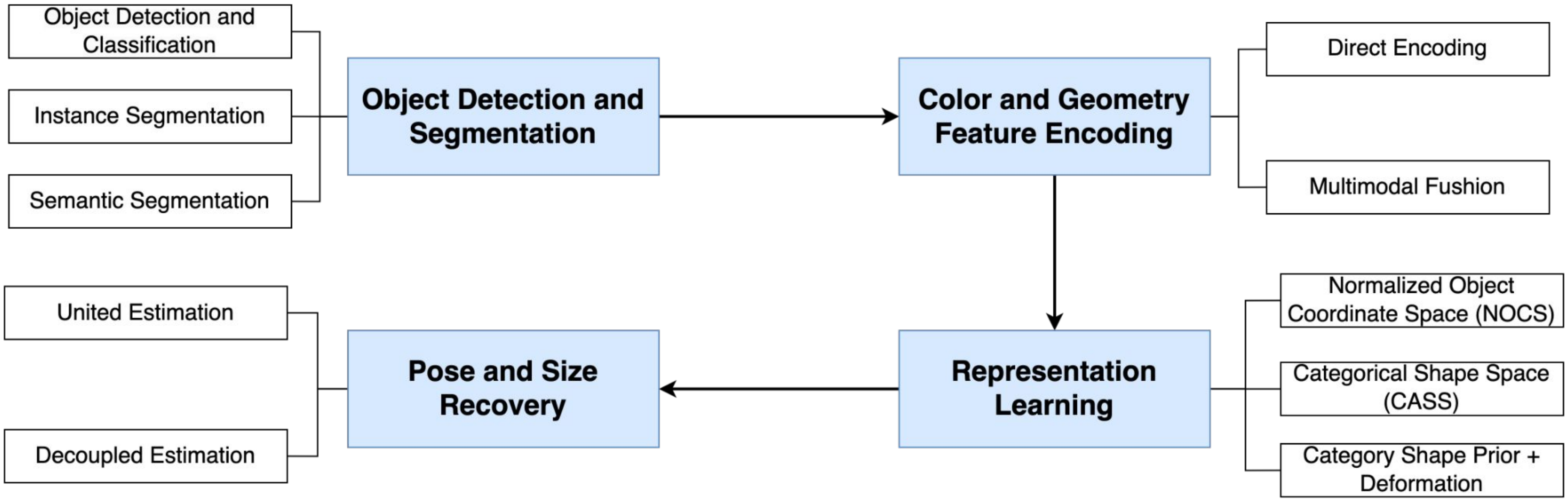
4. Evaluation results on REAL275 and CAMERA25

Methods	IOU 0.5	IOU 0.75	5deg2cm	5deg5cm	10deg2cm	10deg5cm
NOCS [1]	78 (83.9)	30.1 (69.5)	7.2 (32.3)	10 (40.9)	13.8 (48.2)	25.2 (64.6)
SPD [2]	77.3 (**93.2**)	53.2 (83.1)	19.3 (54.3)	21.4 (59)	43.2 (73.3)	54.1 (81.5)
SGPA [3]	**80.1** (**93.2**)	**61.9** (**88.1**)	**35.9** (**70.7**)	**39.6** (**74.5**)	**61.3** (**82.7**)	**70.7** (**88.4**)

* The results in parentheses come from CAMERA 25

* RePoNet-sup[4] refers to the RePoNet trained in fully supervised setting

5. Summary

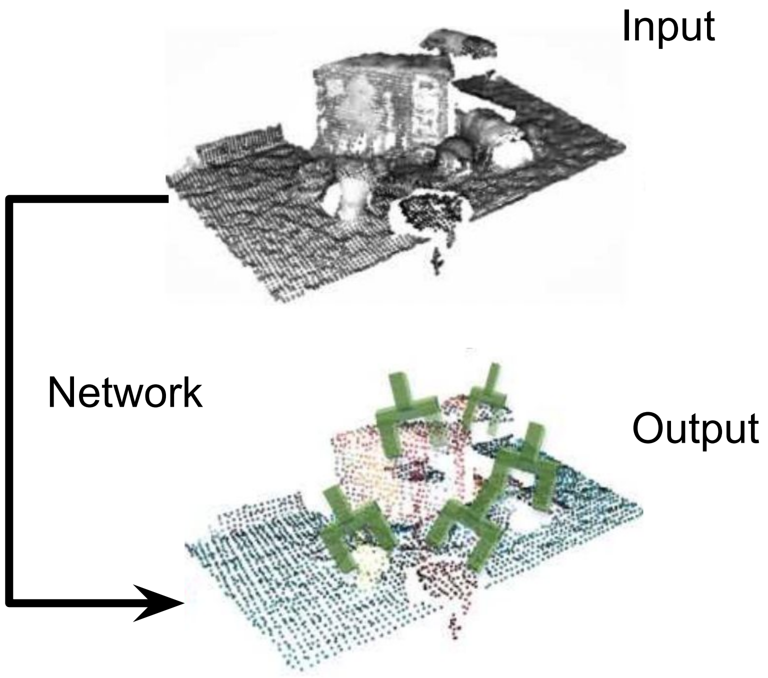


Object Detection and Classification
Instance Segmentation
Semantic Segmentation
Object Detection and Segmentation
Color and Geometry Feature Encoding
Direct Encoding
Multimodal Fushion
United Estimation
Decoupled Estimation
Pose and Size Recovery
Representation Learning
Normalized Object Coordinate Space (NOCS)
Categorical Shape Space (CASS)
Category Shape Prior + Deformation

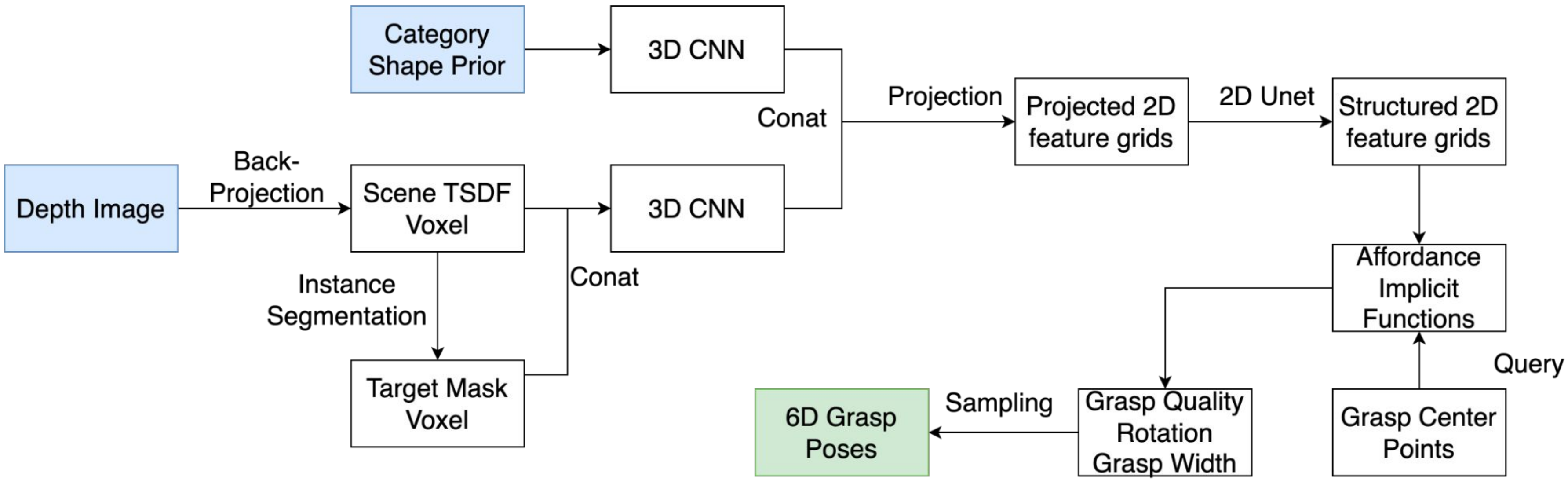
6. Future Work: Robot Grasp Pose Estimation

- Motivation
  - Core task in robotics and object level perception

- Task
  - Input: a single Depth image/RGB-D image of the cluttered scene
  - Output: 6D Pose and width of the grasps for the target object



6. Predicting Target Object Affordance from single depth image with Category Shape Prior



References

[1]. Wang, He, et al. "Normalized object coordinate space for category-level 6d object pose and size estimation." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019.

[2]. Tian, Meng, Marcelo H. Ang, and Gim Hee Lee. "Shape prior deformation for categorical 6d object pose and size estimation." *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16*. Springer International Publishing, 2020.

[3]. Chen, Kai, and Qi Dou. "Sgpa: Structure-guided prior adaptation for category-level 6d object pose estimation." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.

[4]. Lin, Yunzhi, et al. "Single-stage keypoint-based category-level object pose estimation from an RGB image." *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022.

Thank you for listening!