Omni6D: Large-Vocabulary 3D Object Dataset for Category-Level 6D Object Pose Estimation

Mengchen Zhang[1,2], Tong Wu[3], Tai Wang[2], Tengfei Wang[2], Ziwei Liu[4], Dahua Lin[2,3]

[1] Zhejiang University, [2] Shanghai Artificial Intelligence Laboratory, [3] The Chinese University of Hong Kong, [4] Nanyang Technological University

Introduction

Problem Definition

Category-level Object 6D Pose Estimation: Given an RGBD image **I** and the category **c** of the object instance in the image, the task is to estimate the object's direction **R**, position **T**, and size **S** of the object in three-dimensional space without requiring the object's CAD model. This approach enables generalization across different objects within the same category.

Motivation

Limitations in Previous Category-level 6D Pose Estimation Datasets:

- Limited category numbers
- Lack of instance diversity within categories
- Lack of realism
- Overly simplified scenes

Omni6D vs Existing Datasets

Datasets	Mode	Realism	# Categories	# Instances	# Images
ShapeNet-SRN Cars [22]	RGB	Synthetic	1	3514	-
Sim2Real Cars [22]	RGB	Real	1	10	-
CAMERA [40]	RGBD	Synthetic	6	1085	0.3M
REAL [40]	RGBD	Real	6	42	8k
Wild6D [44]	RGBD	Real	5	1722	1M
Omni6D	RGBD	Real-Scanned	**166**	**4,688**	0.8M

Improvements of Omni6D:

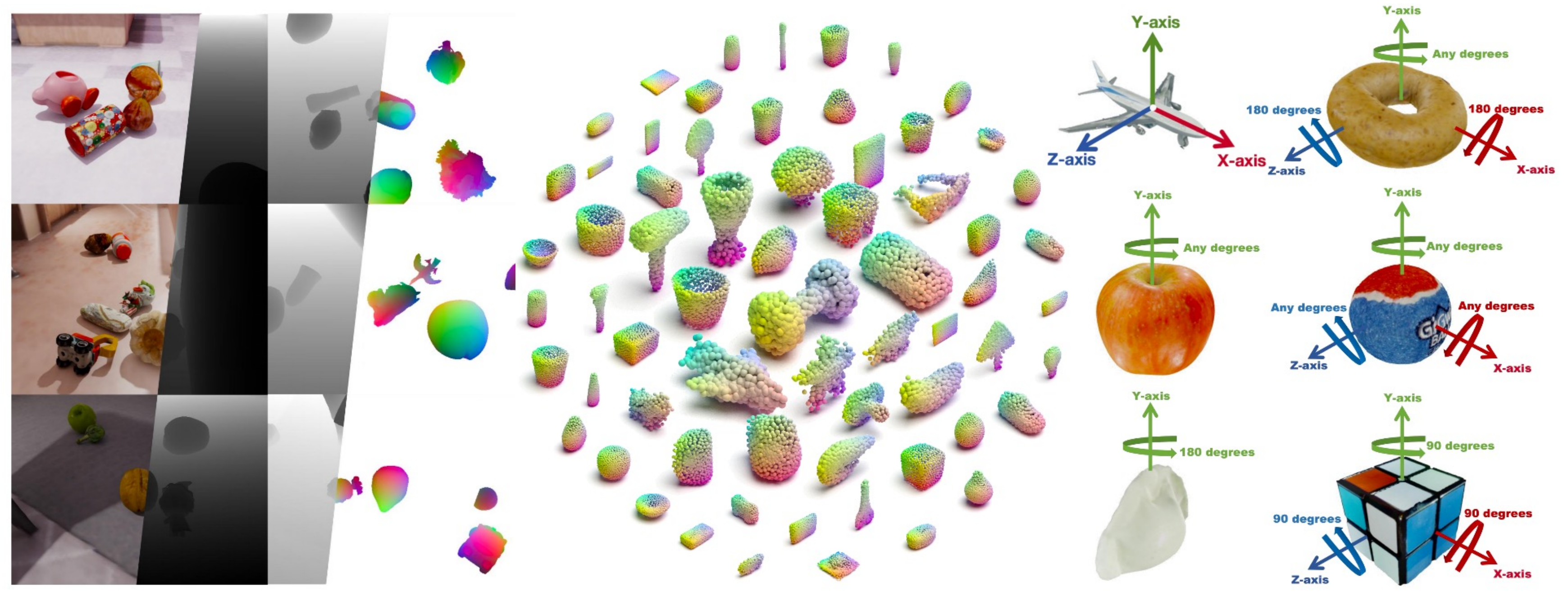
- ✓ Category and Instance Expansion: Significantly extends the range of everyday object categories and instances.
- ✓ Enhanced Realism: Utilizes real-scanned objects.
- ✓ Complex Scenes: Incorporates occlusions, changing lighting conditions, complex backgrounds, and varying viewpoints.

Contribution

- We construct a dataset, **Omni6D**, which comprises an extensive spectrum of **166** categories, **4688** instances adjusted to the canonical pose, and over **0.8M** captures, significantly broadening the scope for evaluation.
- We introduce a **symmetry-aware metric** and conduct systematic **benchmarks** of existing algorithms on Omni6D, offering a thorough exploration of new challenges and insights.
- We propose an effective **fine-tuning strategy** that adapts models from previous datasets to our extensive vocabulary setting.

Omni6D Dataset

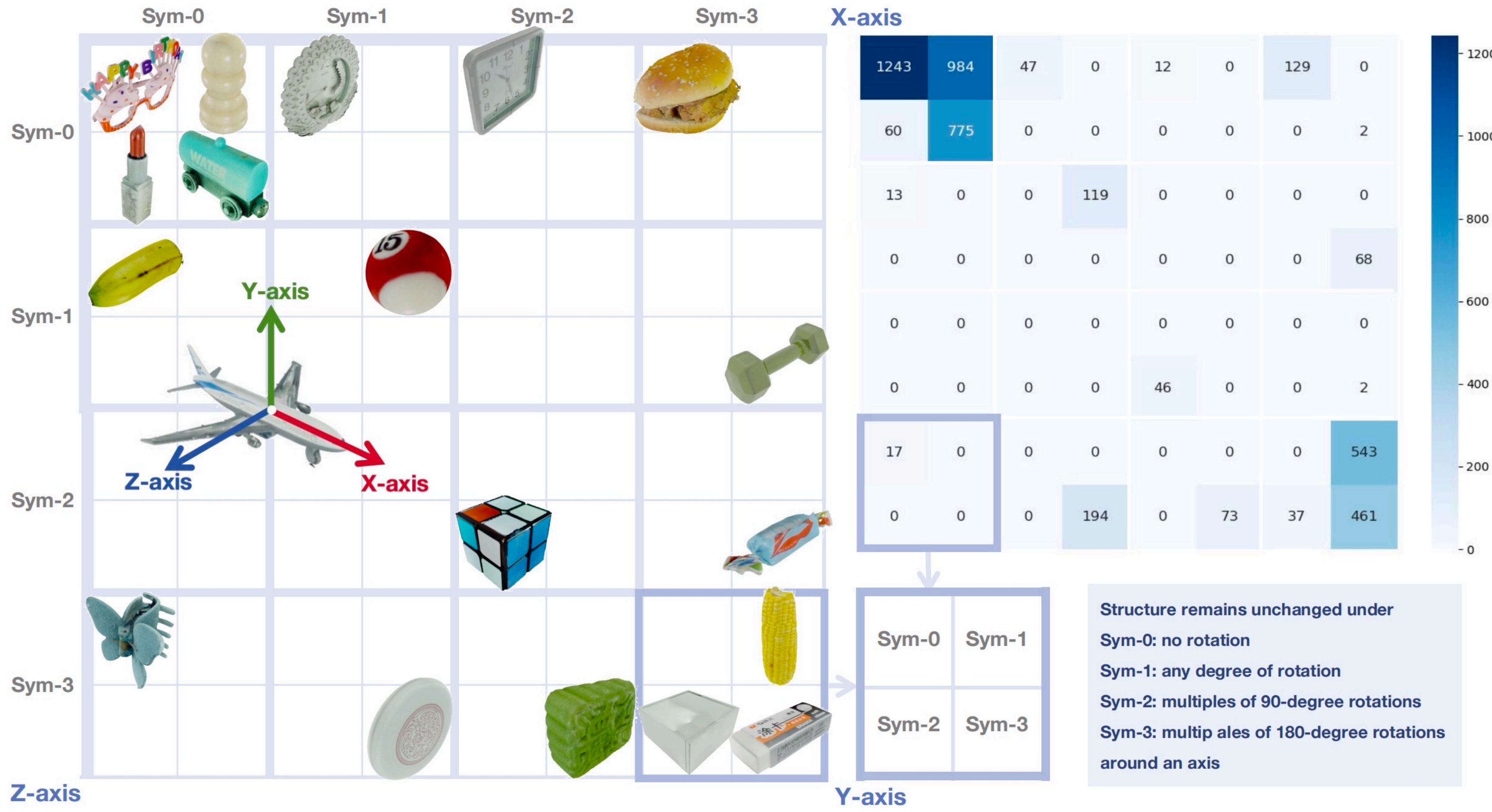
Rich Annotations



(a) Omni6D Dataset (b) Shape Priors (c) Symmetry Annotation

Each rendered output includes a rendered RGB image, instance mask, NOCS mapping, depth map, ground truth class label, as well as 6D pose and size.

Rotational Invariance



Rotational invariance implies that a symmetric object can **retain its original shape after rotation by certain angles**. As shown in the bottom-right corner of the figure, we contemplate four cases of rotational invariance where an object maintains its original shape after following rotations: in no case (**Sym-0**), any degrees (**Sym-1**), multiples of 90 degrees (**Sym-2**) and 180 degrees (**Sym-3**). The figure above illustrates all kinds of symmetry cases using object instances and quantifies their occurrence frequency.

Evaluation and Analysis

Symmetry-Aware Evaluation

Algorithm 1 Compute Our Symmetry-Aware Metric L_s

procedure SYMMETRIC_METRIC(L, R, n_x, n_y, n_z)
  $\Theta_0 = \{0^\circ\}$
  $\Theta_2 = \{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$
  $\Theta_3 = \{0^\circ, 180^\circ\}$ *// Rotations around Sym-1 axis need not be considered.*
  $c = \text{count}(1 \text{ occurrences in } \{n_x, n_y, n_z\})$
  if $c \geq 2$ **then** *// The object is a sphere.*
    $L_s = L(R^*, R)$
  else if $c == 1$ **then** *// Rotations around Sym-1 axis can be disregarded.*
    Without loss of generality, assume $n_x == 1$.
    $L_s = \min_{\theta_y \in \Theta_{n_y}, \theta_z \in \Theta_{n_z}} L(R^*_{\theta_y, \theta_z}, R)$
  else if $c == 0$ **then** *// Simply enumerate all cases.*
    $L_s = \min_{\theta_x \in \Theta_{n_x}, \theta_y \in \Theta_{n_y}, \theta_z \in \Theta_{n_z}} L(R^*_{\theta_x, \theta_y, \theta_z}, R)$
  end if
  return L_s
end procedure

Omni6D has a wider range of objects with different **rotational invariances** across multiple axes. To alleviate this issue, we propose a **symmetry-aware metric**. Unlike prior works focusing solely on the y-axis, our method considers rotation symmetry around **all three axes**.

Benchmark

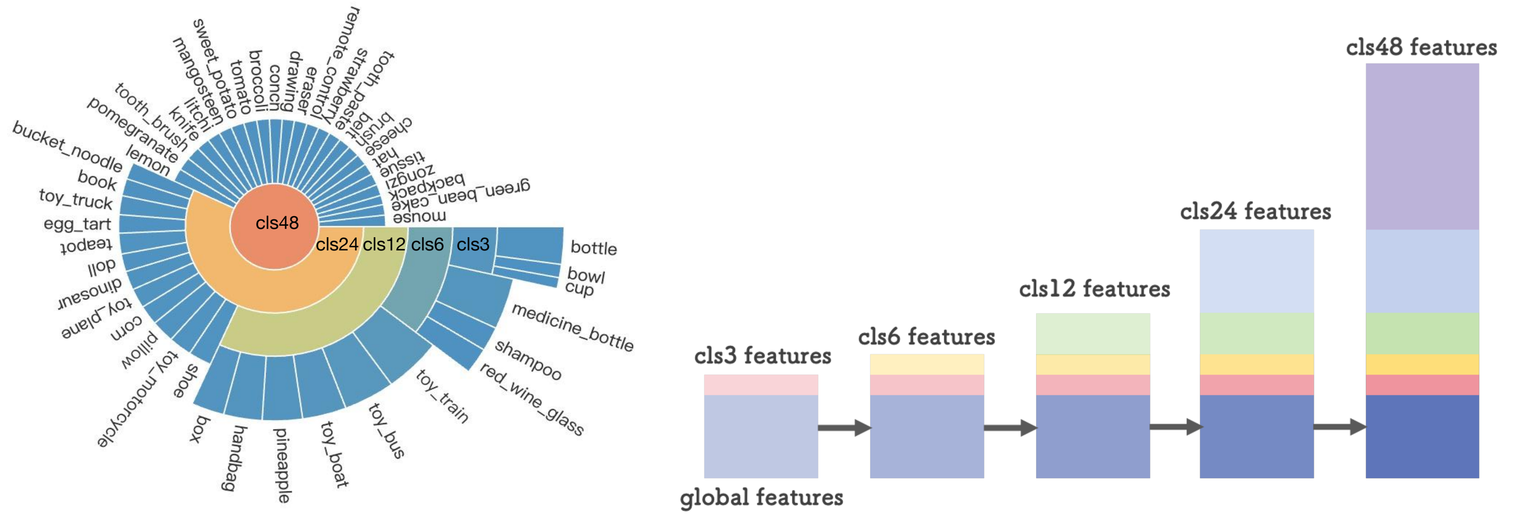
Table 2: Category-level performance on Omni6D dataset. Models are trained on Omni6D$_{train}$ and tested on Omni6D$_{test}$. Instances within each category in the test set are unseen during training, substantiating the algorithms' capacity to generalize within individual categories under large-vocabulary settings. **Bold** and underlined results indicate the best and second-best performers.

Methods	Network	IoU_{50}	IoU_{75}	$5^\circ 2cm$	$5^\circ 5cm$	$10^\circ 2cm$	$10^\circ 5cm$	5°	10°	$2cm$	$5cm$
SPD [34]	implicit	44.56	20.37	<u>7.55</u>	**9.56**	<u>14.76</u>	**19.23**	**10.68**	**21.02**	37.49	70.09
SGPA [6]	implicit	36.34	14.44	4.78	6.84	10.13	15.03	8.49	17.73	25.57	59.18
DualPoseNet [20]	hybrid	<u>58.84</u>	**25.49**	**8.28**	<u>9.30</u>	**17.26**	<u>19.05</u>	<u>9.38</u>	<u>19.18</u>	<u>73.82</u>	<u>96.37</u>
RBP-Pose [46]	hybrid	35.92	4.66	0.37	0.60	0.53	0.80	0.75	0.96	39.73	83.55
GPV-Pose [10]	explicit	15.28	0.26	0.10	0.70	0.14	0.96	2.25	2.96	5.31	33.70
HS-Pose [47]	explicit	**62.65**	<u>23.02</u>	4.26	4.85	10.49	11.61	4.96	11.75	**80.93**	**97.78**

Table 3: Category-level performance on unseen categories. Models are trained on Omni6D$_{train}$ and tested on Omni6D$_{out}$. Categories in the test set never appear in the training set, validating the algorithms' ability to generalize across categories.

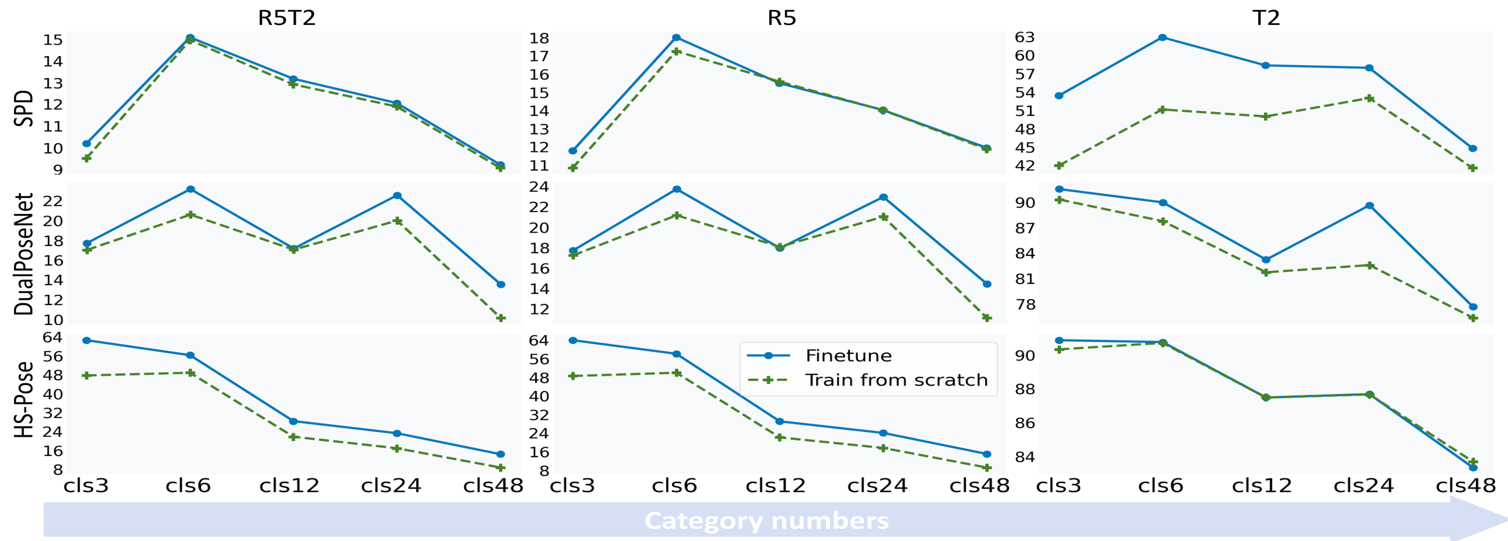
Methods	Network	IoU_{50}	IoU_{75}	$5^\circ 2cm$	$5^\circ 5cm$	$10^\circ 2cm$	$10^\circ 5cm$	5°	10°	$2cm$	$5cm$
SPD [34]	implicit	7.56	0.95	0.18	0.40	0.80	1.65	0.65	2.36	8.88	40.59
SGPA [6]	implicit	7.05	0.60	0.07	0.28	0.19	0.82	0.53	1.69	3.87	28.28
DualPoseNet [20]	hybrid	**36.85**	**12.06**	**3.24**	**3.37**	**8.04**	**8.51**	**3.39**	**8.64**	<u>78.00</u>	<u>98.60</u>
RBP-Pose [46]	hybrid	26.18	1.95	0.01	0.02	0.02	0.03	0.02	0.03	16.74	43.06
GPV-Pose [10]	explicit	10.97	0.14	0.03	0.18	0.12	0.57	0.30	1.07	7.14	41.30
HS-Pose [47]	explicit	<u>36.75</u>	<u>8.92</u>	<u>1.54</u>	<u>1.66</u>	<u>4.67</u>	<u>5.16</u>	<u>1.75</u>	<u>5.38</u>	**79.95**	**98.27**

Fine-Tuning Strategy



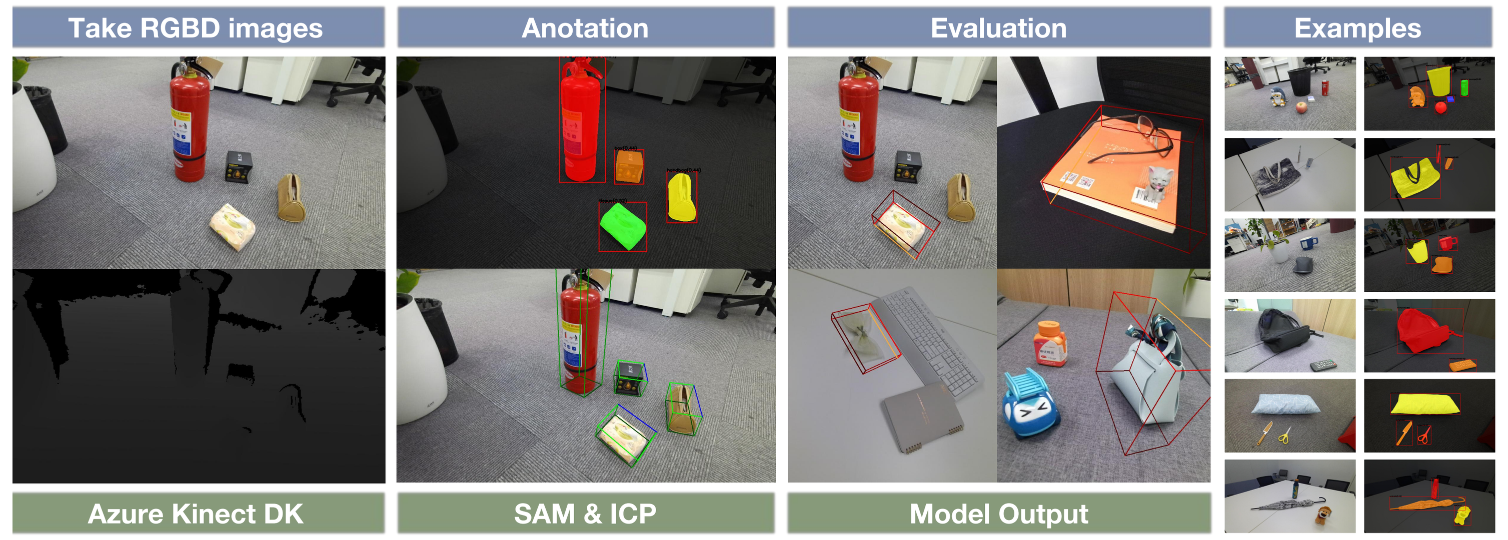
(a) Category inventory of clsn (b) Our finetune strategy

We propose a fine-tuning strategy to extend methods from a limited set of categories to a large vocabulary. This involves **an iterative fine-tuning process** on a progressively expanded category dataset until the desired number of categories is reached.



Even with an exponential increase in the number of categories, pre-trained models remain pivotal in our fine-tuning strategy.

Sim2real Capability



To further validate the sim2real capability of models trained with Omni6D, we constructed a real-world dataset, **Omni6D-Real**, comprising 30 scenes, 39 categories, 73 instances, and 1k RGBD images captured using the Azure Kinect.