LabelAny3D: Label Any Object 3D in the Wild
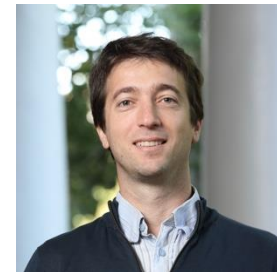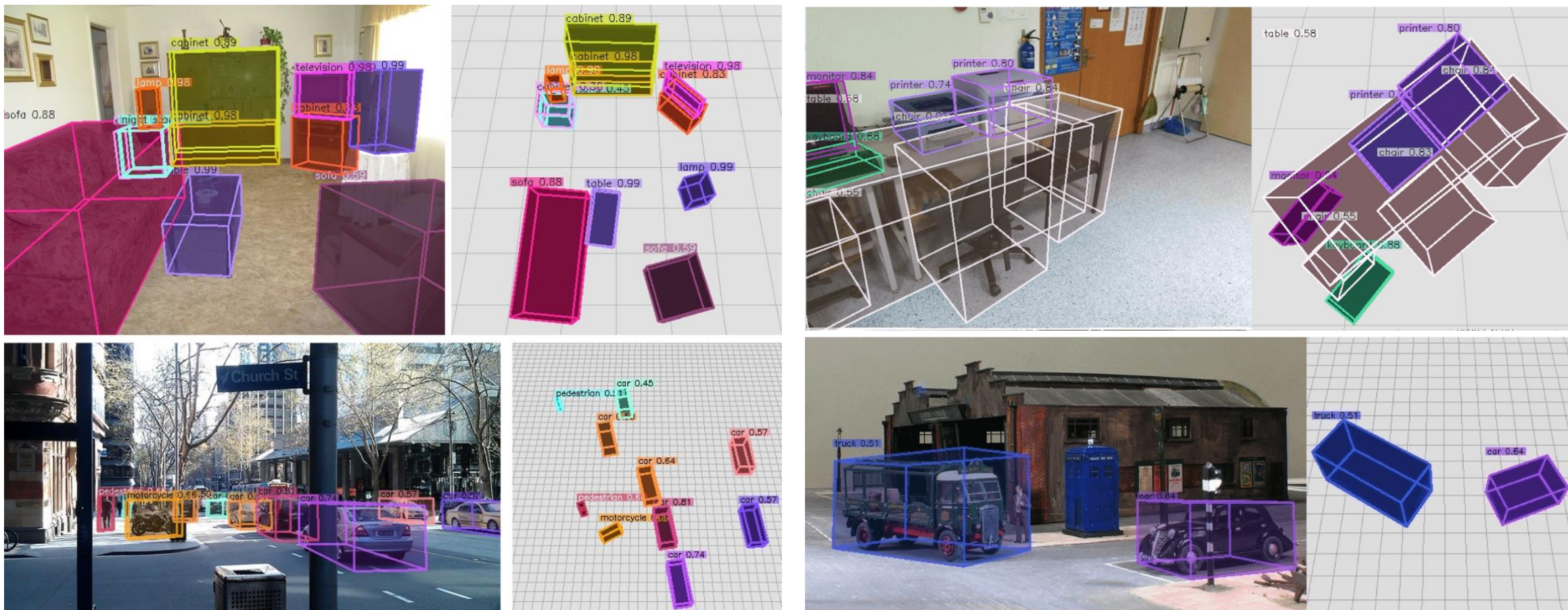
Jin Yao*, Radowan Mahmud Redoy*,
Sebastian Elbaum, Matthew B. Dwyer, Zezhou Cheng

University of Virginia
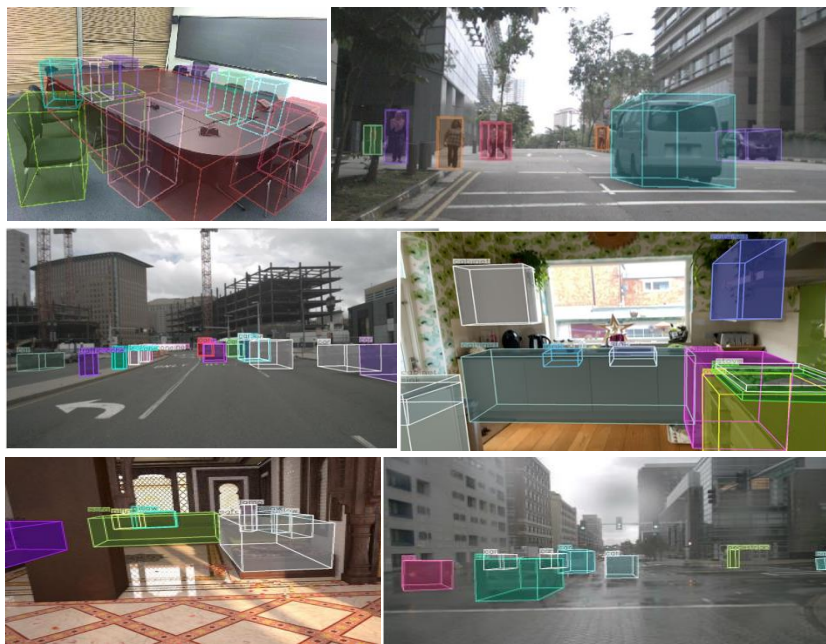
Open-Vocab Monocular 3D Object Detection

1

Bottlenecks for Open-Vocab Monocular 3D Detection

- Lack of large-scale and diverse 3D dataset for training and evaluation
  - Existing datasets focus on indoor and self-driving

- Collecting 3D data is expensive

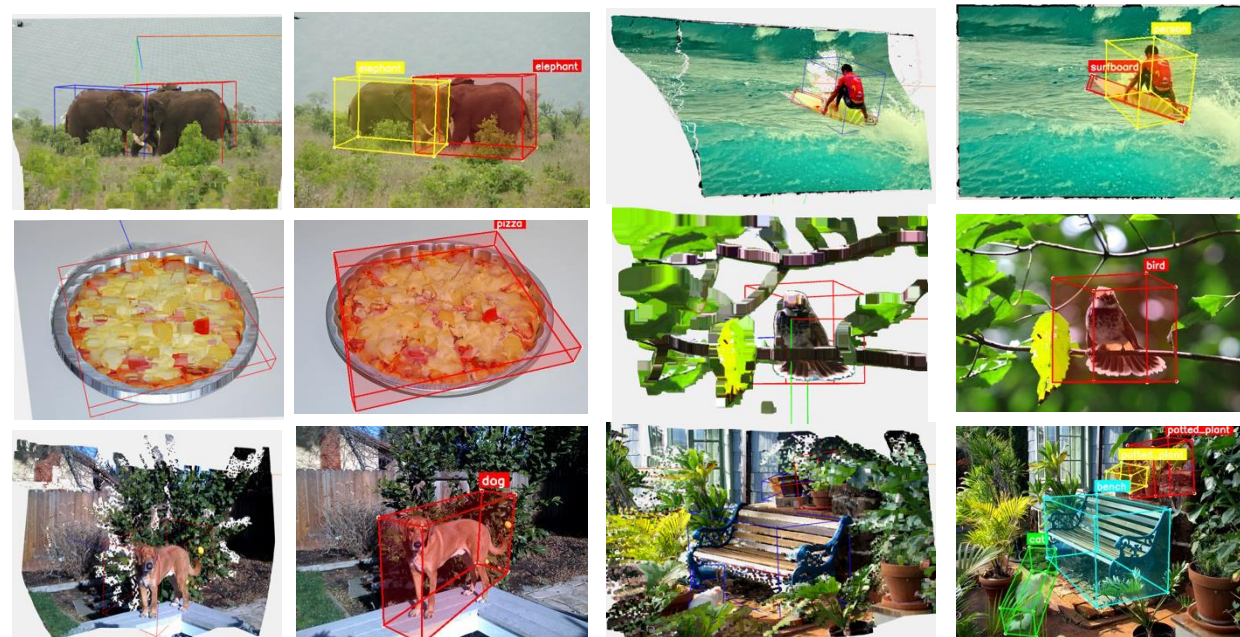- Annotating 3D data is labor-intensive

Overview

Previous Annotations

Our Pseudo Annotations

Omni3D: Only Indoor and Self-driving

LabelAny3D: Diverse Scenes and Categories

LabelAny3D automatically labels 3D box on the 2D image via 3D scene reconstruction

LabelAny3D

(a) 2D Image Processing

Input Image

Image super-resolution

2D Instance Segmentation

Amodal Completion

(b) Monocular 3D Scene Reconstruction

3D Scene Geometry Estimation

2D-3D Alignment

Multi-Views of Aligned 3D objects

Initial 3D Reconstruction in Canonical Space

(c) 3D Annotation Generation

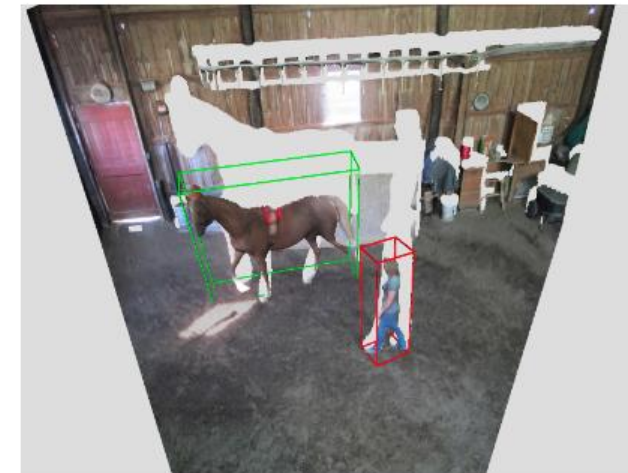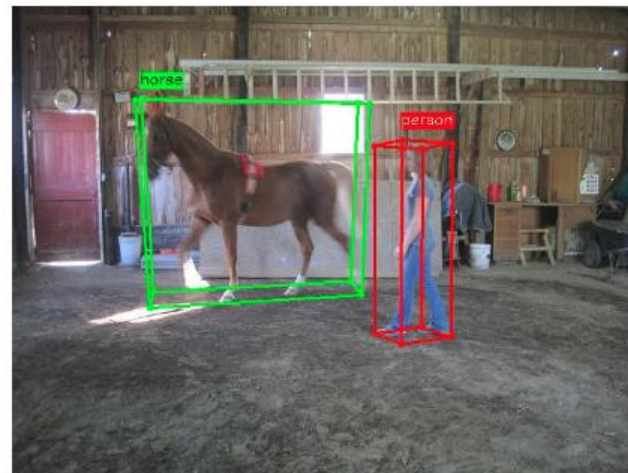*Multi-Views of 3D scene reconstruction*

LabelAny3D automatically labels 3D box on the 2D image via 3D scene reconstruction

4
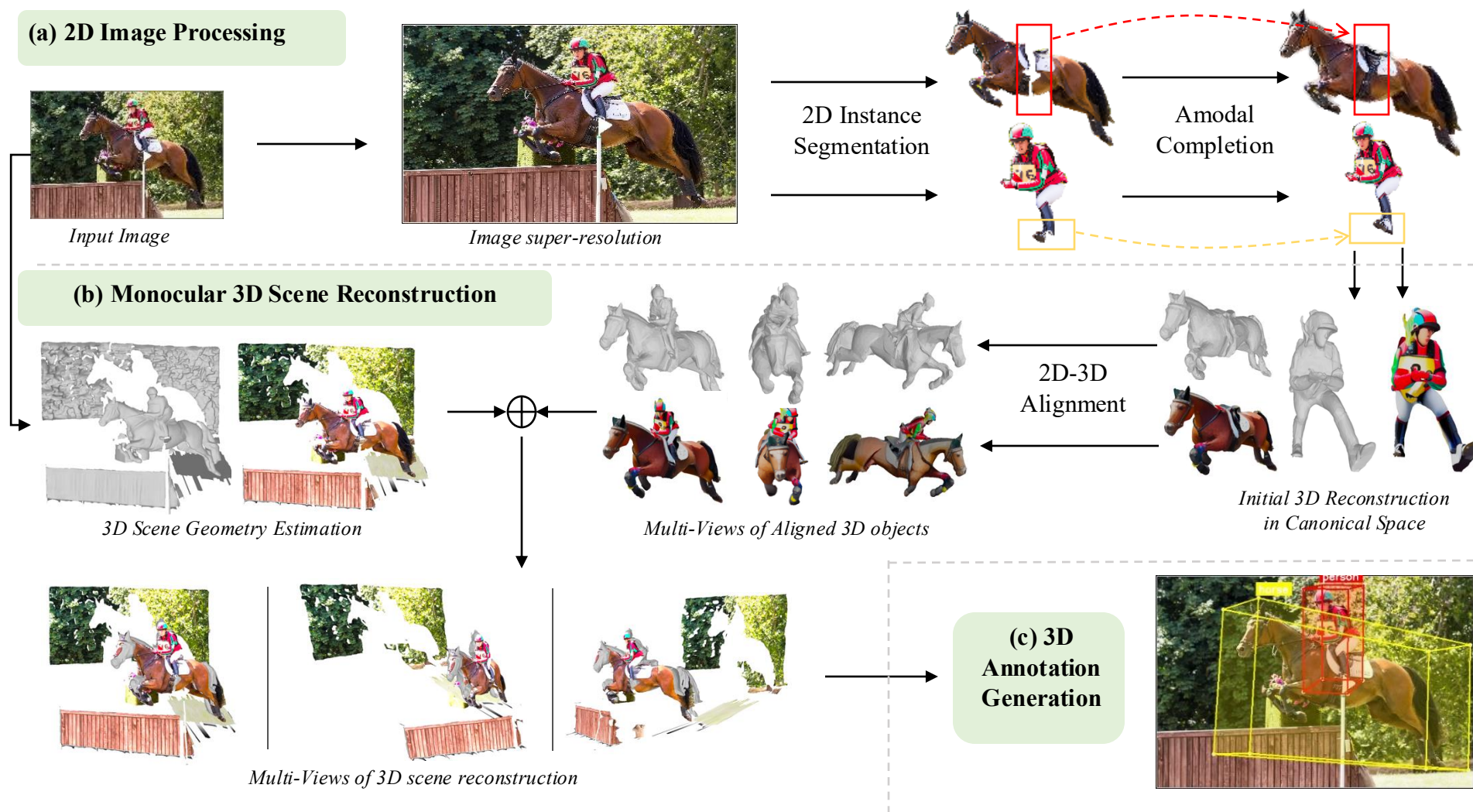
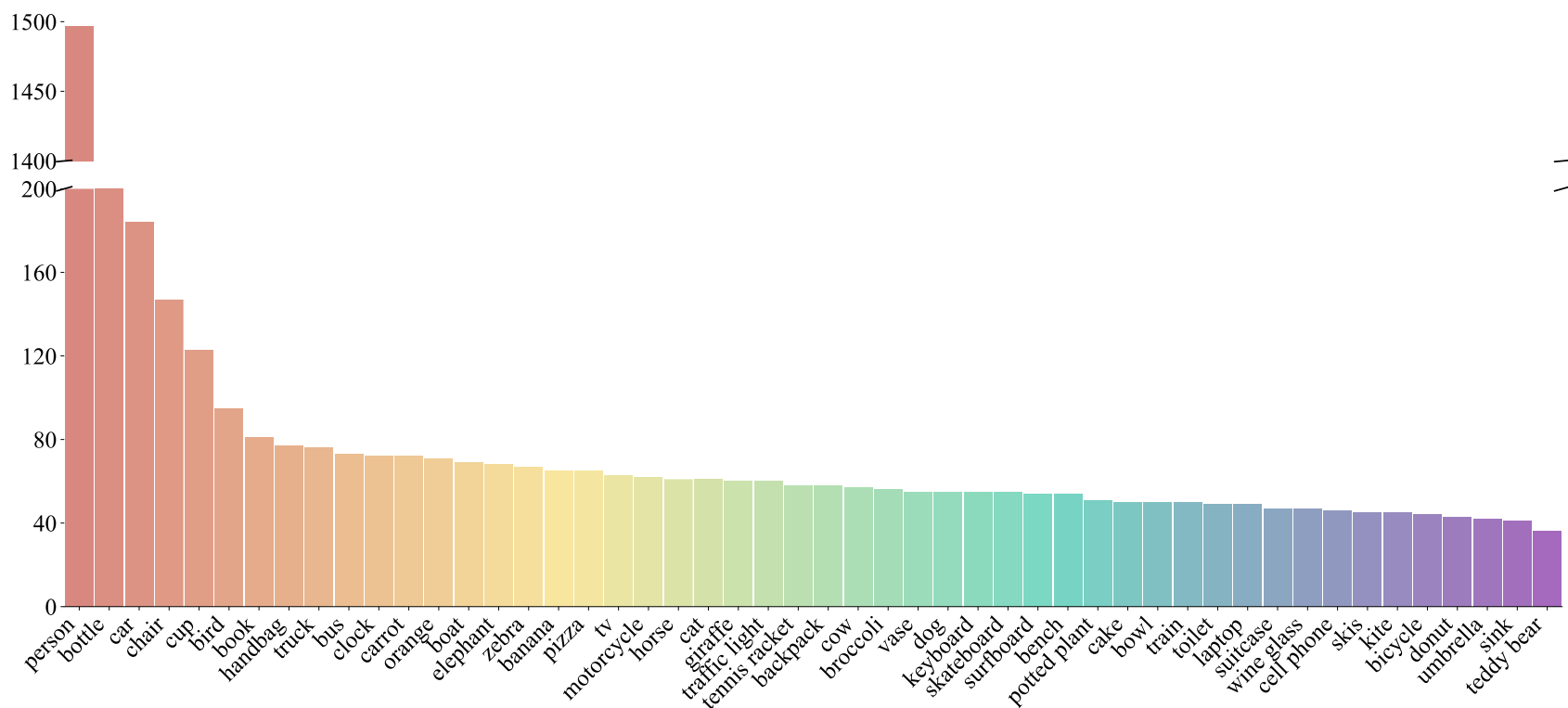# COCO3D Benchmark

LabelAny3D +
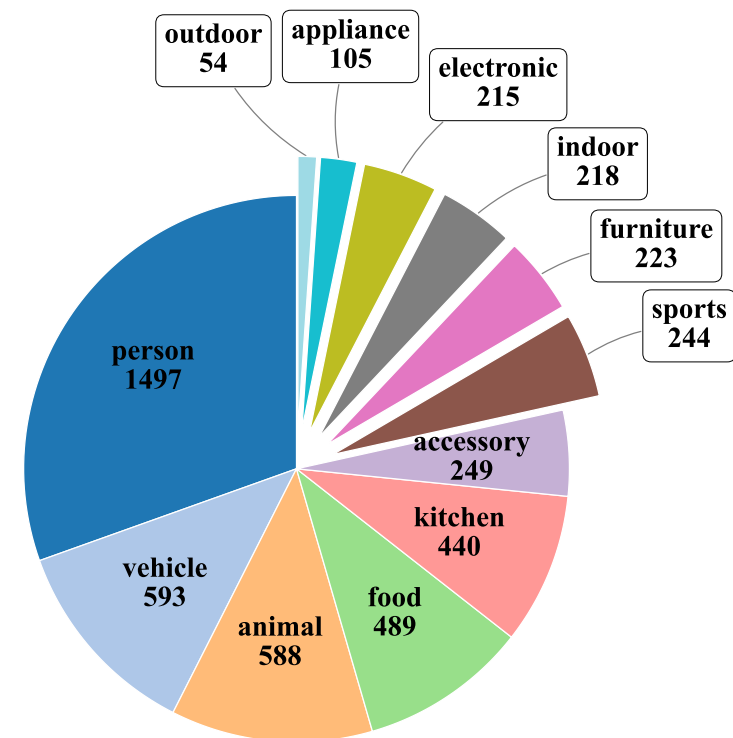Human Refinement

Images from MS-COCO Validation Set   3D Annotations

COCO3D Benchmark

Top 50 category distribution

Super category distribution
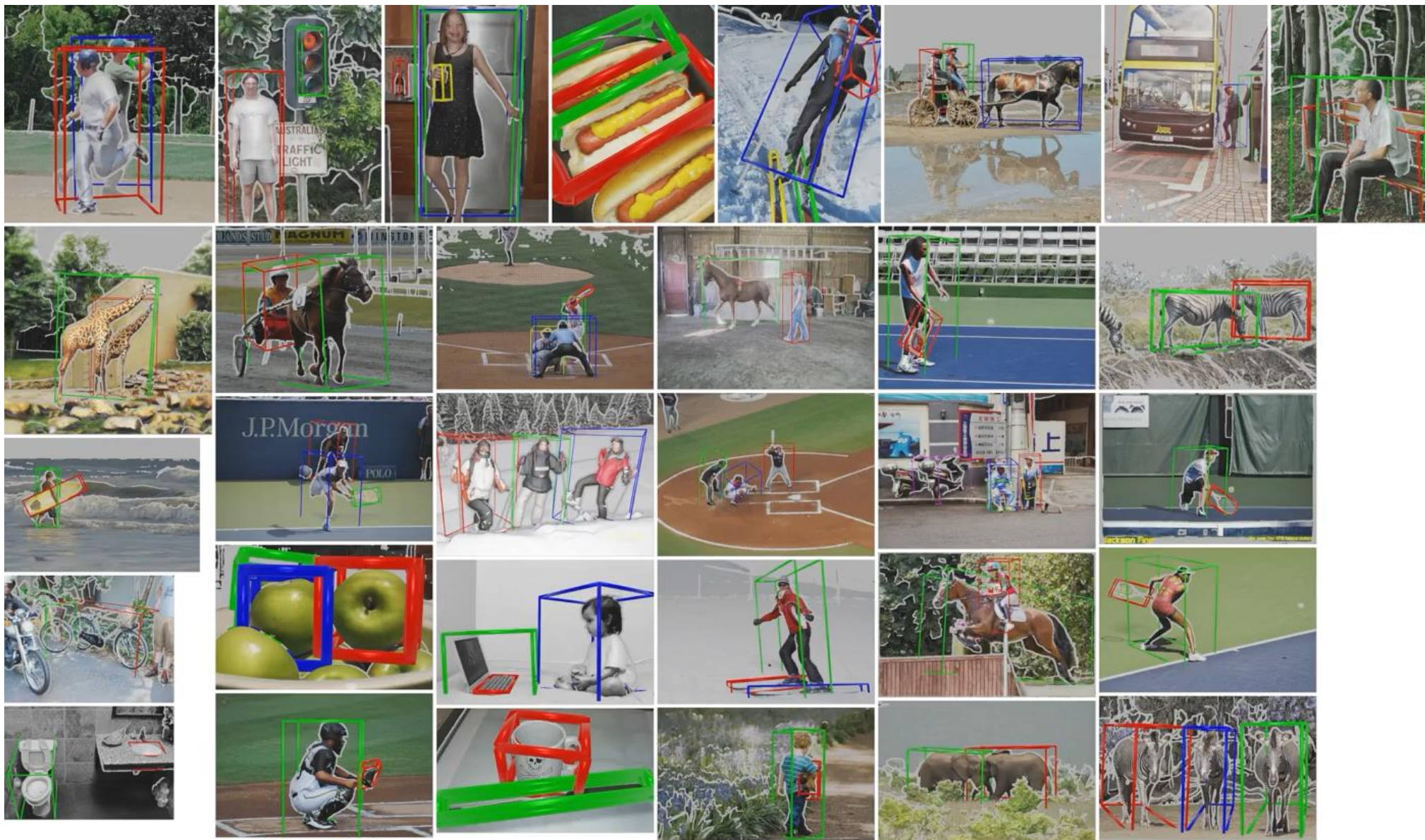
COCO3D Benchmark

Samples from our proposed **COCO3D** benchmark.

COCO3D Benchmark

Train Monocular 3D Detector with LabelAny3D

- Training Data
  - 16k MS-COCO training set images pseudo-labeled by LabelAny3D

- Evaluation on COCO3D:
  - New Metric: Relative Layout AP3D
  - Assesses the consistency of the relative spatial layout between predicted and ground-truth bounding boxes

# Train Monocular 3D Detector with LabelAny3D
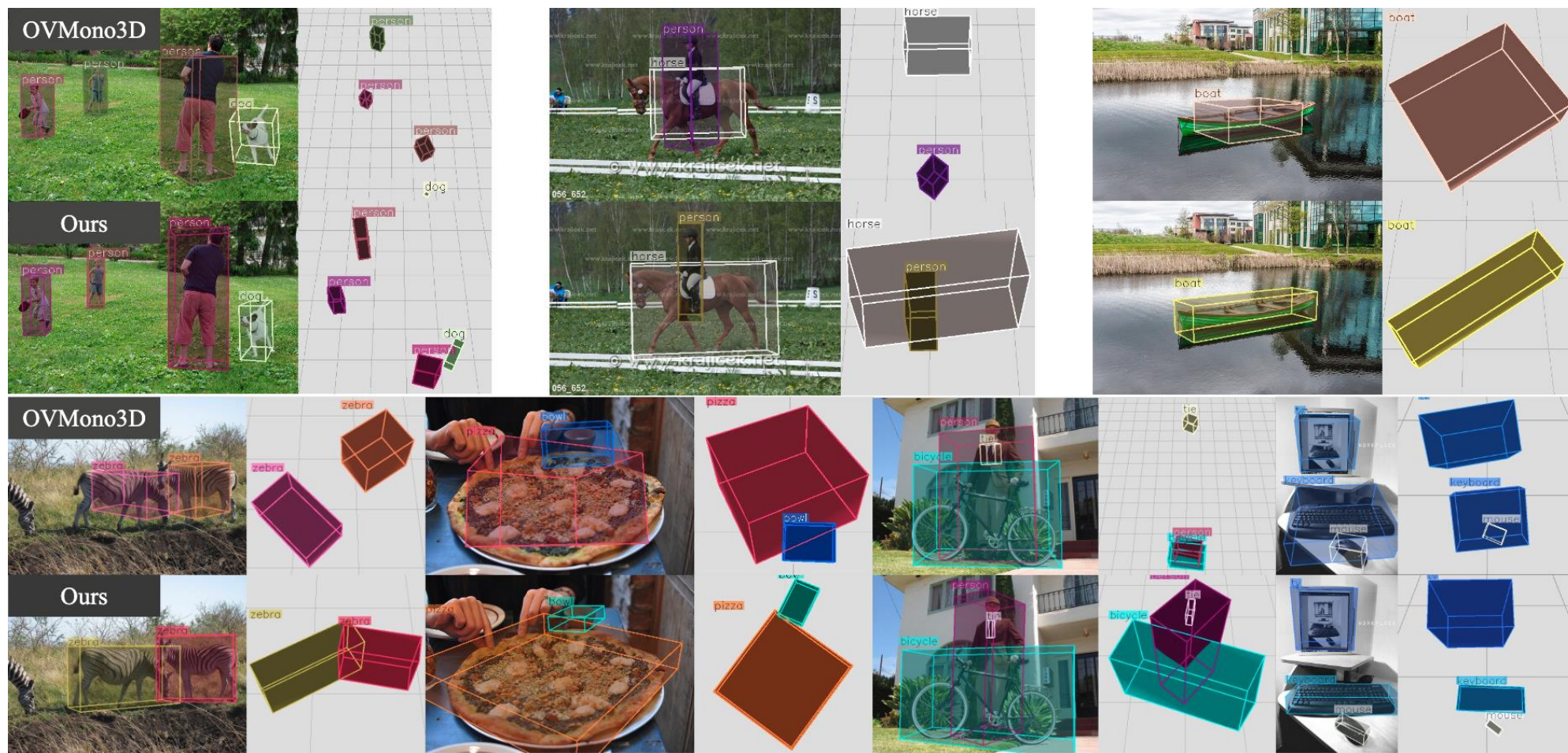
Model trained with LabelAny3D's pseudo labels shows better spatial layout prediction

Train Monocular 3D Detector with LabelAny3D

Training dataset	COCO3D				Omni3D Novel		Omni3D Base	
	$AP_{3D}\uparrow$	$AR_{3D}\uparrow$	$AP_{3D}^{Rel}\uparrow$	$AR_{3D}^{Rel}\uparrow$	$AP_{3D}\uparrow$	$AR_{3D}\uparrow$	$AP_{3D}\uparrow$	$AR_{3D}\uparrow$
Baseline: Omni3D [7, 75]	5.87	10.51	20.86	30.06	<u>16.05</u>	36.85	**24.77**	**47.28**
OVM3D-Det* [27]	2.69	5.25	7.98	12.25	5.30	15.71	7.32	26.34
Omni3D [7] + OVM3D-Det [27]	6.82	11.94	20.76	27.69	15.55	**37.18**	22.35	<u>42.68</u>
LabelAny3D	<u>7.78</u>	<u>15.41</u>	<u>24.66</u>	<u>34.54</u>	8.47	23.34	3.92	19.66
Omni3D [7] + LabelAny3D	**10.92**	**20.10**	**32.02**	**43.82**	**16.98**	<u>36.96</u>	<u>22.74</u>	42.46

Model trained with LabelAny3D's pseudo labels shows better spatial layout prediction

Labeling Performance
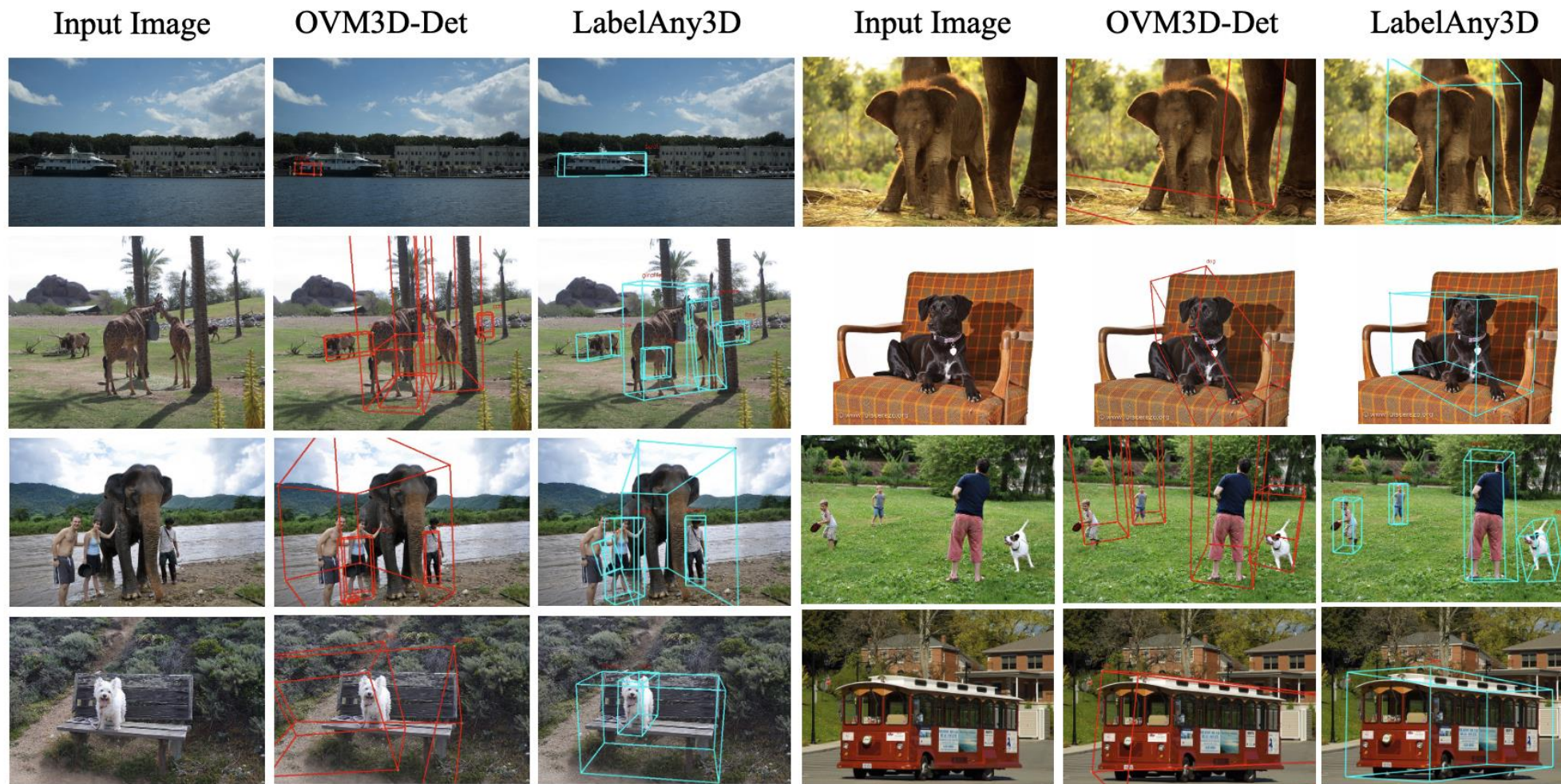
| Input Image | OVM3D-Det | LabelAny3D | Input Image | OVM3D-Det | LabelAny3D |

LabelAny3D achieves better pseudo annotations on in-the-wild images

12

Labeling Performance

Methods	$AP_{3D}\uparrow$	$AP_{3D}^{15}\uparrow$	$AP_{3D}^{25}\uparrow$	$AP_{3D}^{50}\uparrow$	$AR_{3D}\uparrow$	$AP_{3D}^{Rel}\uparrow$	$AR_{3D}^{Rel}\uparrow$
OVM3D-Det* [27]	10.03	16.88	9.03	1.44	17.82	10.04	17.84
LabelAny3D	**64.17**	**82.11**	**74.47**	**57.34**	**73.57**	**64.17**	**73.57**

LabelAny3D achieves better pseudo annotations on in-the-wild images

Conclusion and Limitations

- We present:
  - LabelAny3D, an automatic 3D box labeling tool works on in-the-wild images
  - COCO3D, a 3D detection benchmark with diverse scenes and categories

- Limitations
  - Fails on scenarios such as heavy occlusion and textureless objects
  - Occasional reconstruction errors due to RGB-to-3D ambiguity
  - Lacks robust training strategies for noisy pseudo-labels
  - Not exhaustively annotated benchmark to ensure dataset quality

Thanks!