

360° Image Perception with MLLMs: A Comprehensive Benchmark and a Training-Free Method

Huyen T. T. Tran¹, Van-Quang Nguyen¹, Farros Alferro¹, Kang-Jun Liu¹, Takayuki Okatani^{1,2}

¹Graduate School of Information Sciences, Tohoku University · ²RIKEN Center for AIP



Why 360° Breaks MLLMs



Equirectangular projection (**ERP**)



Cubemap projection (**CMP**)

Question: What is the relative position of **the Toys store** and **the Yorozyu store**?

- A. They are on the same side of the street but separated by a side street.
- B. They are right next to each other.
- C. They are positioned at a right angle to each other.

D. They are across from each other.

Answer:

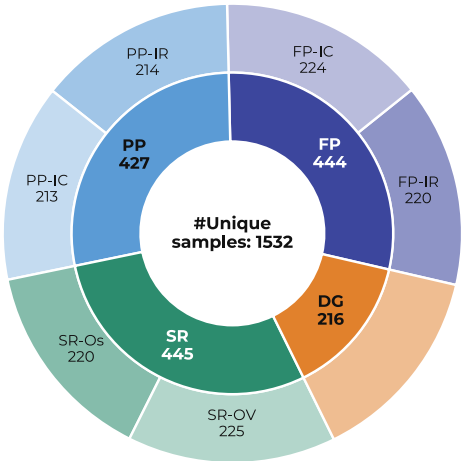
GPT-4o: C

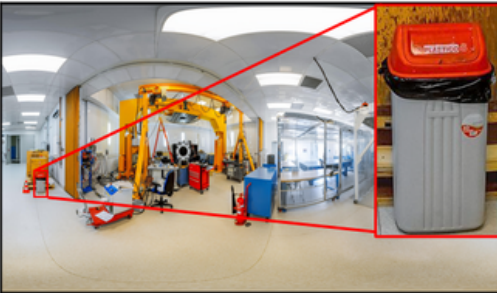
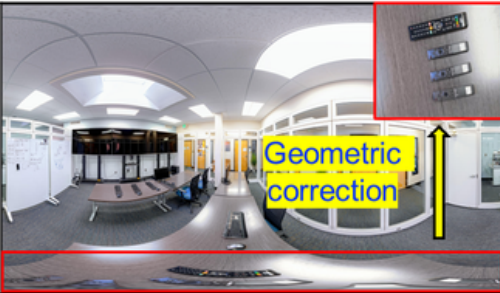
Ours: D

- ERP keeps everything in one frame, but warps geometry
- CMP stays sharp, but shatters the scene across six faces

360Bench

- 643 images @ 7K · 1,532 human-curated MCQs
- 7 subtasks: perception, distortion, spatial reasoning, direction
- **Humans 86.3% vs best model (Gemini Pro 2.5) 46.5%**



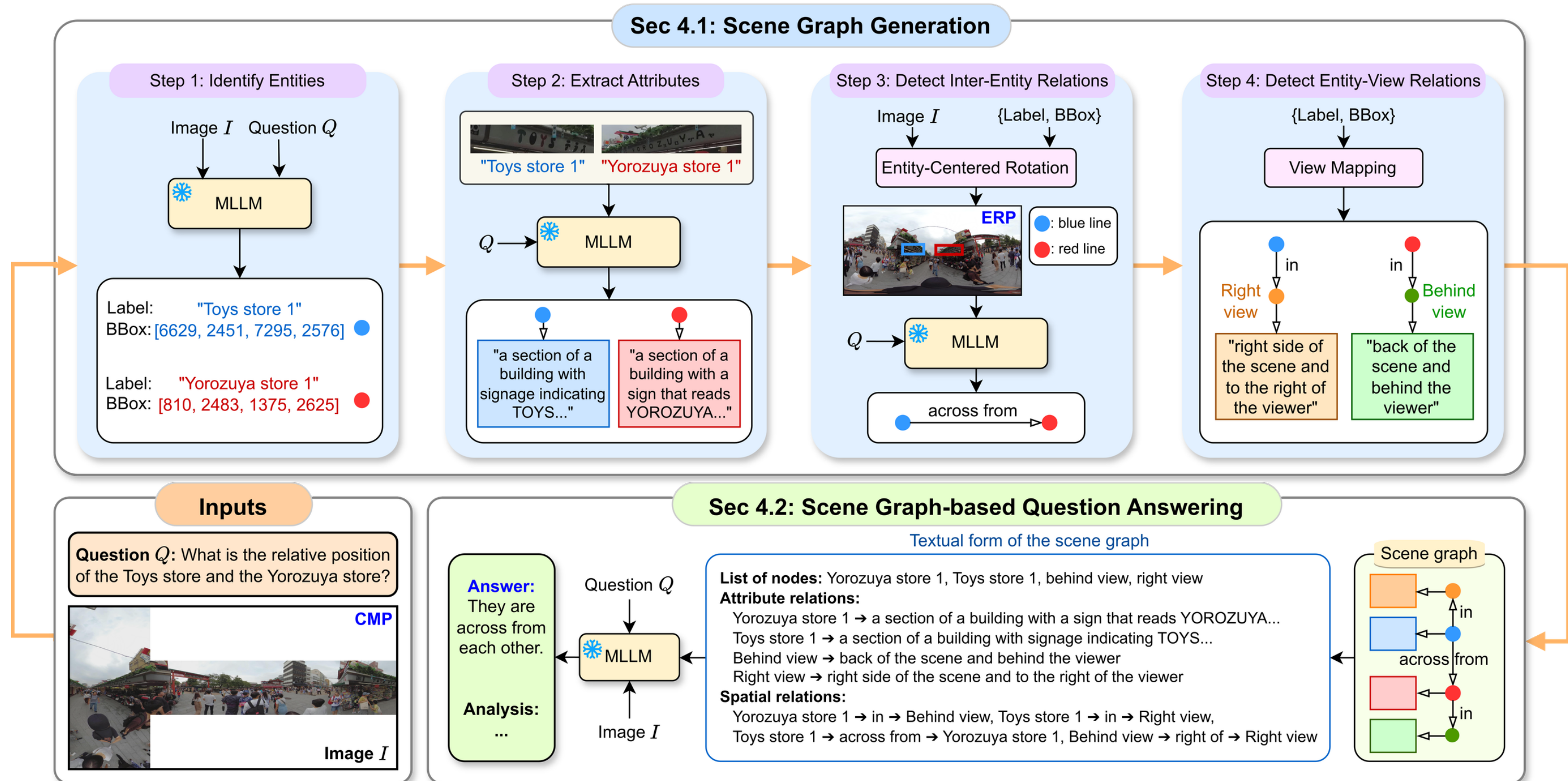
 What does the label on the trash bin say? A. Plastico B. Hazardous C. Recyclables D. Paper	 How many types of drinks are sold in the vending machine? A. 7 B. 8 C. 9 D. 10	 What type of roundabout is this? A. A five-way roundabout B. A six-way roundabout C. A three-way roundabout D. A four-way roundabout	 How many remote controls are on the table? A. 3 B. 5 C. 1 D. 4
(a) Fine-grained instance recognition (FP-IR)	(b) Fine-grained instance counting (FP-IC)	(c) Projection-distorted instance recognition (PP-IR)	(d) Projection-distorted instance counting (PP-IC)
 What is the relative position of Rooms 7 and 8? A. They are next to each other on the same side of the hallway. B. They are directly across from each other. C. They are diagonally across from each other. D. They are located on different hallways.	 Where is the fire hydrant? A. On the right B. Behind you C. In front of you D. On the left	 How can I get to Quartier Ville Vieille-Léopold? A. Cross the road, then turn right B. Cross the road, then turn left C. Turn right and go straight ahead D. Turn left and go straight ahead	
(e) Spatial reasoning between objects (SR-Os)	(f) Spatial reasoning between object and view (SR-OV)	(g) Direction-Giving (DG)	

CMP: low distortion → clean object perception
(PP ↑, up to **+14.1%**)

ERP: global layout preserved → coherent spatial reasoning (SR ↑, up to **+14.5%**)

CMP to see · ERP to reason

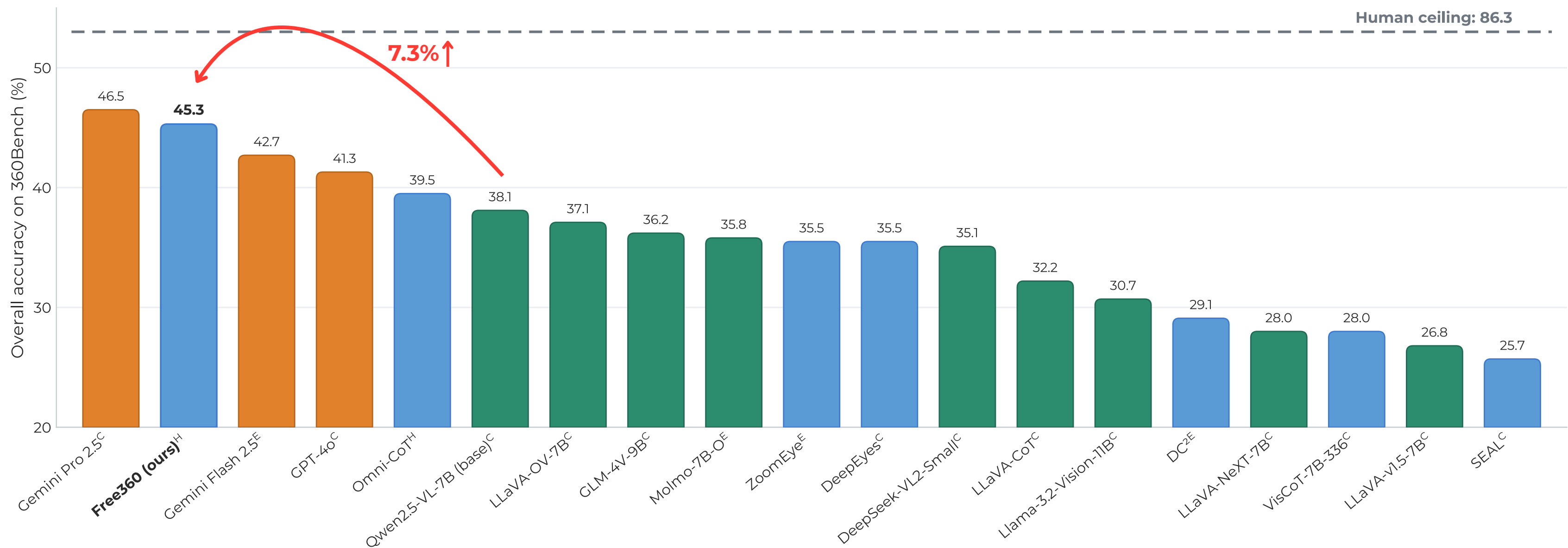
Free360: a training-free scene graph



- **Frozen** Qwen2.5-VL-7B. No fine-tuning, builds a **question-relevant scene graph**
- 4 steps: **entities** (CMP) → **attributes** → **relation** (rotate to ERP) → **viewer-centered** direction

Results

- **+7.3 overall** · up to **+22.9 on subtasks**
- All components are important
- Gains hold at 3B / 7B / 32B



Impact of scene graph components

Setting	Acc.
Full	45.3
w/o Crop	42.7
w/o Rotate	44.7
w/o EVR	42.8

Effectiveness across model scales

Model scale	Qwen2.5-VL		Free360 Hybrid	Acc. Gain	
	ERP	CMP		ERP	CMP
3B	35.2	37.2	40.8	+5.5	+3.5
7B	37.3	38.1	45.3	+8.1	+7.3
32B	37.8	39.7	45.0	+7.2	+5.3

Takeaway

- 360° perception is a real, measurable **weakness** in today's MLLMs
- The two projections are complementary: **CMP to see, ERP to reason**
- A simple, training-free scene graph closes much of the gap



Try answering 360Bench here!



Visit our project page

Come to our poster! Poster Session 6 · Sat Sep 12, 15:00–17:00 CEST · MalmöMässan
ExHall #154