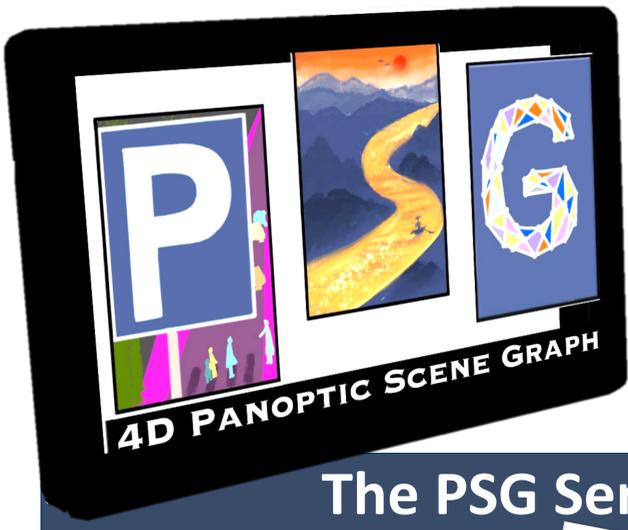
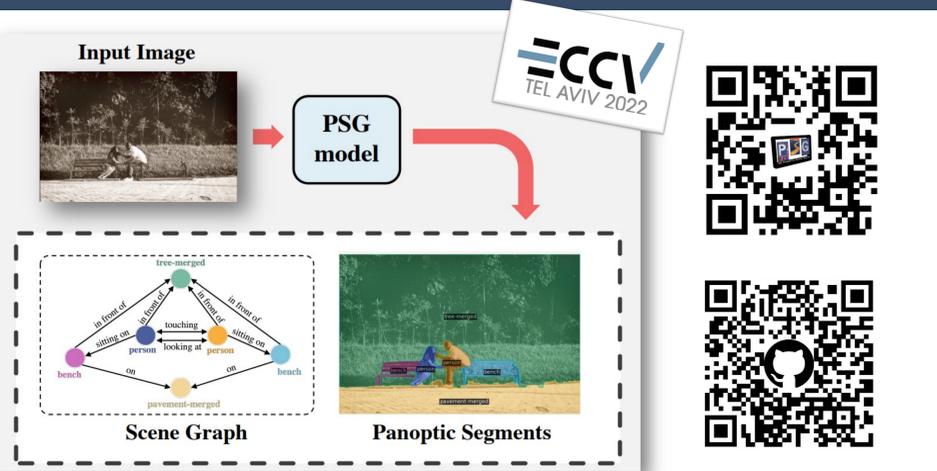


4D Panoptic Scene Graph Generation

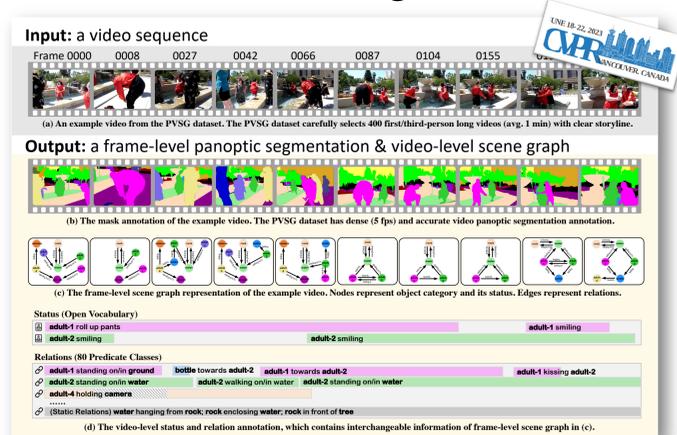
Jingkang Yang, Jun Cen, Wenxuan Peng, Shuai Liu,
Fangzhou Hong, Xiangtai Li, Kaiyang Zhou, Qifeng Chen, Ziwei Liu



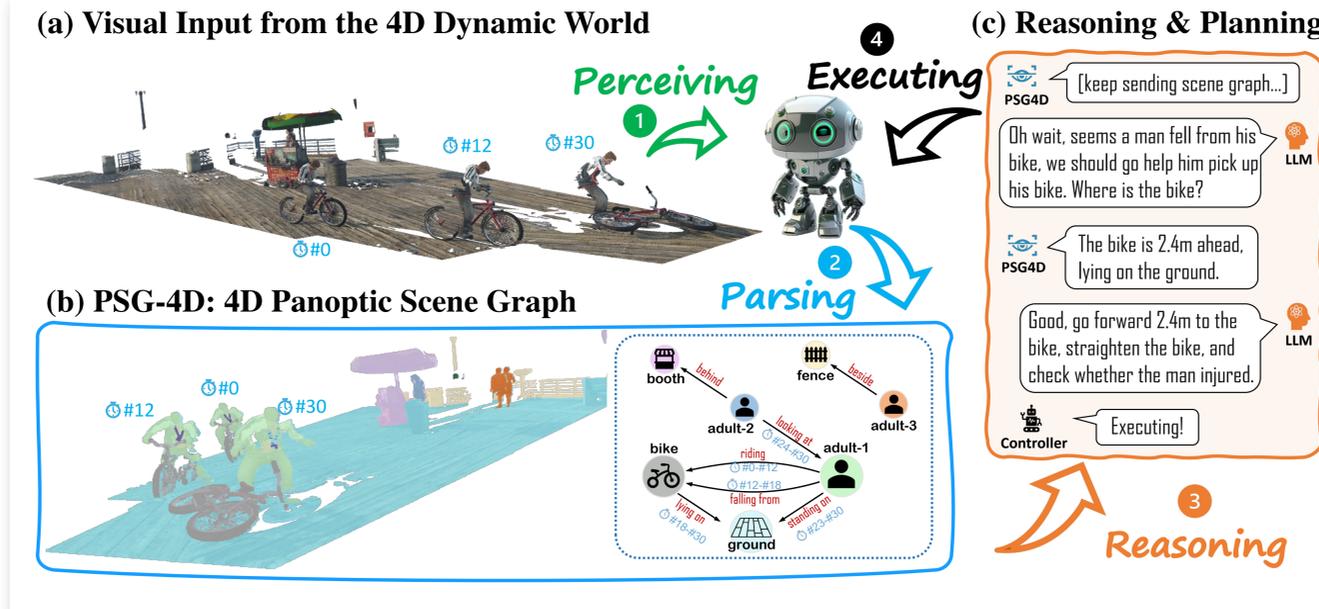
The PSG Series



Panoptic Scene Graph Generation (PSG) extends the scope of scene graph toward comprehensive scene understanding with accurate groundings.



Panoptic Video Scene Graph Generation (PVSG) represents videos using dynamic scene graph with accurate groundings on both pixels and time axis.



PSG-4D parses **RGB-D or Point Cloud Video Sequence** into a **dynamic scene graph** for later **reasoning and executing module**.

Application: Video Gaming AI, AR/VR Assistants, Robots

New Dataset: The PSG-4D Datasets

The PSG-4D Dataset covers GTA and HOI data with ~1M RGB-D video frames.

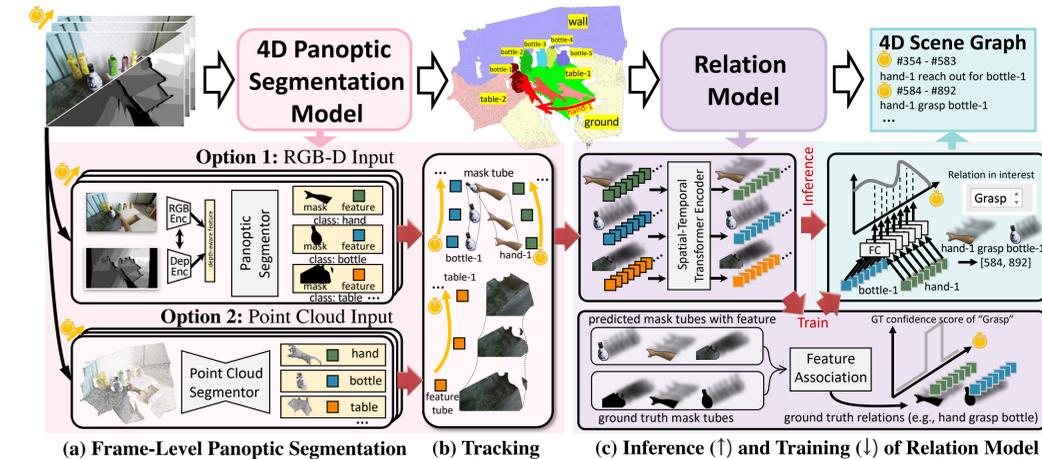


(a) PSG4D-GTA (Synthetic, Third-Person View)

(b) PSG4D-HOI (Real-World, Egocentric)

Another Real-World Egocentric PSG-4D Dataset (PSG4D-ADT) from **Aria Digital Twins** is coming soon!

New Solution: PSG4DFormer



Performance

Input Type	Method	PSG4D-GTA			PSG4D-HOI		
		R/mR@20	R/mR@50	R/mR@100	R/mR@20	R/mR@50	R/mR@100
Point Cloud Sequence	#1 3DSGG [18]	1.48 / 0.73	2.16 / 0.79	2.92 / 0.85	3.46 / 2.19	3.15 / 2.47	4.96 / 2.84
	#2 PSG4DFormer ^{/t}	2.25 / 1.03	2.67 / 1.72	3.14 / 2.05	3.26 / 2.04	3.16 / 2.35	4.18 / 2.64
	#3 PSG4DFormer	4.33 / 2.10	4.83 / 2.93	5.22 / 3.13	5.36 / 3.10	5.61 / 3.95	6.76 / 4.17
RGB-D Sequence	#4 3DSGG [18]	2.29 / 0.92	2.46 / 1.01	3.81 / 1.45	4.23 / 2.19	4.47 / 2.31	4.86 / 2.41
	#5 PSG4DFormer ^{/t}	4.43 / 1.34	4.89 / 2.42	5.26 / 2.83	4.44 / 2.37	4.83 / 2.43	5.21 / 2.84
	#6 PSG4DFormer ^{/d}	4.40 / 1.42	4.91 / 1.93	5.49 / 2.27	5.49 / 3.42	5.97 / 3.92	6.43 / 4.21
	#7 PSG4DFormer	6.68 / 3.31	7.17 / 3.85	7.22 / 4.02	5.62 / 3.65	6.16 / 4.16	6.28 / 4.97

- **Variants:** PSG4DFormer with/without temporal encoder (/t) and depth branch (/d) compared to 3DSGG baseline.
- **Input Comparison:** RGB-D outperforms point cloud, especially on PSG4D-GTA dataset.
- **Depth Importance:** Depth branch removal reduces performance, emphasizing its necessity.
- **Temporal Attention:** Essential for optimal 4D scene graph generation, as shown by variants without it.