PARIS3D: ReasonIng-based 3D PArt Segmentation Using Large Multimodal Model

Amrin Kareem, Dr. Jean Lahoud, Dr. Hisham Cholakkal

The Problem

A novel segmentation task known as reasoning part segmentation for 3D objects.
Task: Given a 3D point cloud and a complex and implicit textual query about a specific part of a 3D object, output a segmentation mask.

"A part of this Faucet is gray colored, is located on top of and in front of it and constitutes more than half of its total volume. What is it?"
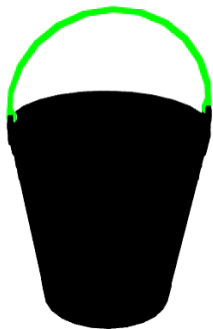
Original   Ground Truth   Ours (PARIS3D)

"Can you segment the curved semi-circular steel part of this bucket that you can hold?"
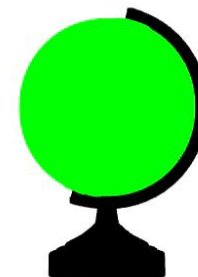
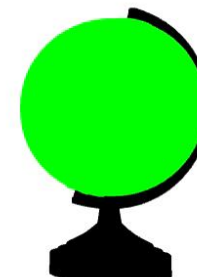Original   Ground Truth   Ours (PARIS3D)

"Segment the spherical part of this Globe."
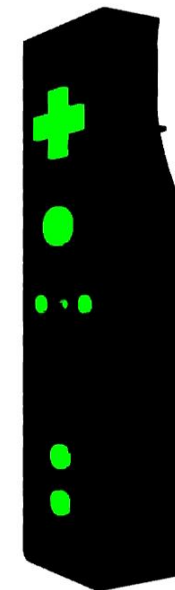
Original   Ground Truth   Ours (PARIS3D)

"Where is the part that can be pressed to change options located in this Remote?"
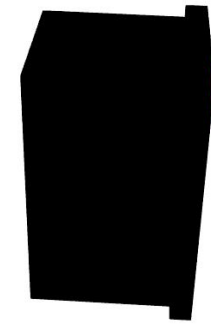
Original   Ground Truth   Ours (PARIS3D)

"Which is the thin horizontal bar-shaped part of this Dishwasher that you can hold to open and close?"

PARIS3D: [SEG]. The handle of a dishwasher is typically located on the front door of the machine and is used to open and close the door. The handle is shown in the color grey, is thin, horizontal, bar shaped and 100 per cent of the length of the Dishwasher and is located at 92 per cent of its height. It is located in front of the Dishwasher.
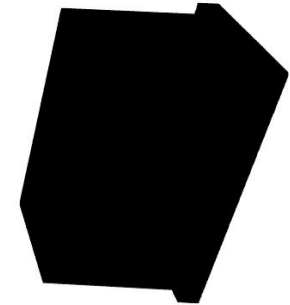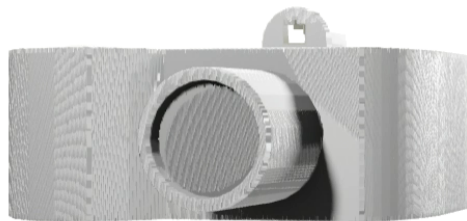

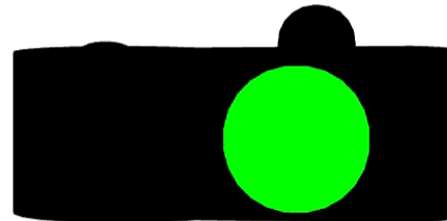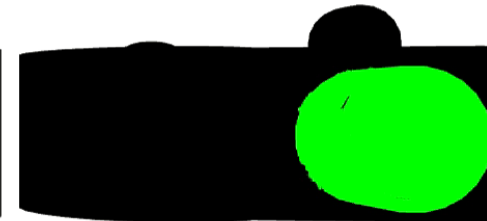Original   Ground Truth   Ours (PARIS3D)

"What is the cylindrical part in front of the Camera that focuses light and captures the image?"

Original   Ground Truth   Ours (PARIS3D)

Dataset – RPSeg3D

Dataset Preparation:
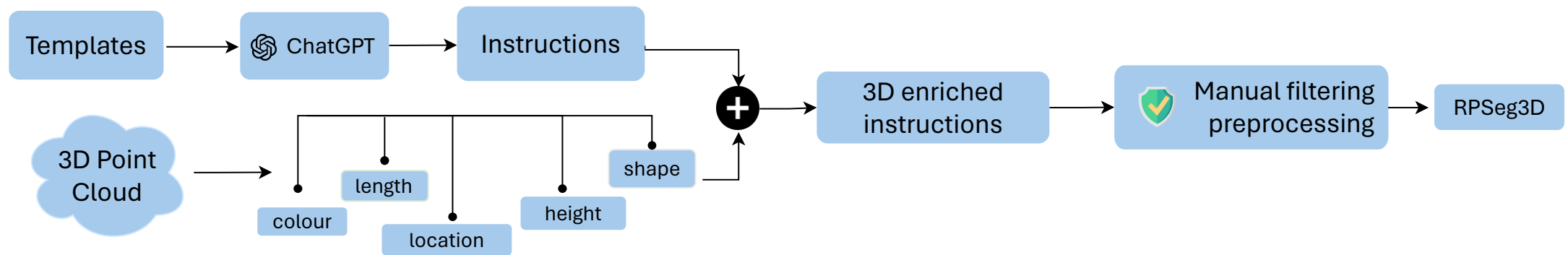
- 2624 point clouds and 60k instructions
- Train : 718 objects, Test : 1906 objects

Architecture
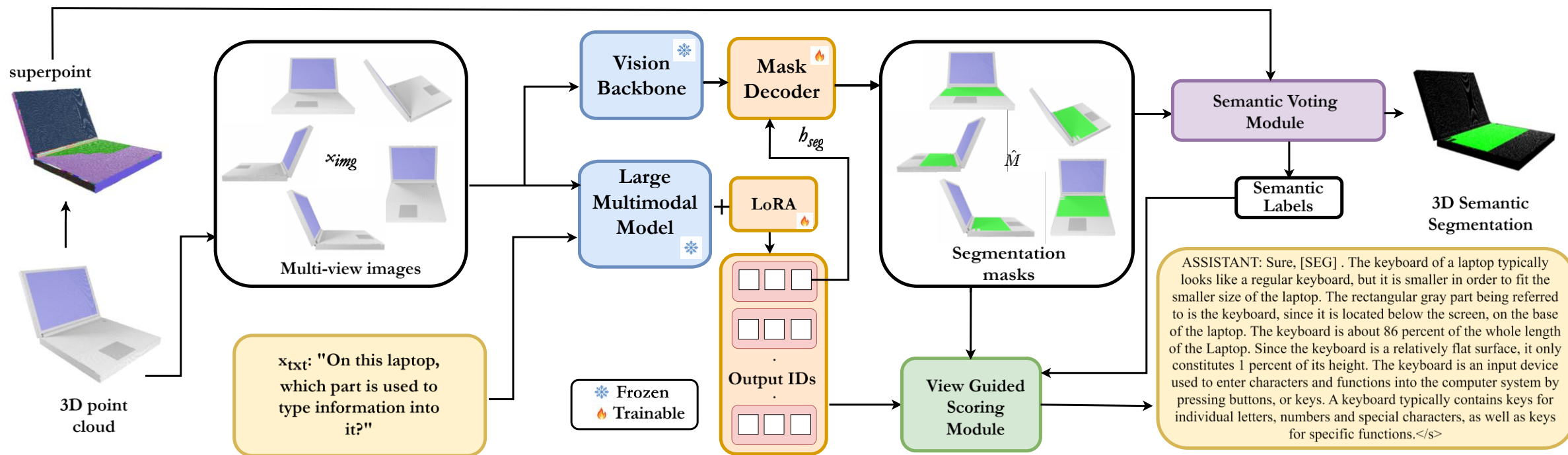
Quantitative Results

Results of reasoning part segmentation:

Method	Val			Test		
	Normal Query	3D Query	Overall	Normal Query	3D Query	Overall
LISA-MV	16.60	20.16	18.38	17.60	20.57	19.08
LISA-MV (ft)	50.43	50.28	50.35	50.75	50.81	50.78
PARIS3D	55.33	55.50	55.42	55.94	57.60	56.77

When fine-tuned with 3D information, our model performs better than the baselines for normal and 3D prompts.

Quantitative Results

Comparison to previous 3D part segmentation methods:

| 3D Data	Method	Overlapping Categories									Non-overlapping Categories									Overall (45)
		Bottle	Chair	Display	Door	Knife	Lamp	Storage-Furniture	Table	Overall (17)	Camera	Cart	Dis-Penser	Kettle	Kitchen Pot	Oven	Suit-case	Toaster	Overall (28)	
Extra data (45x8+28k)	PointNet++ [10]	48.8	84.7	78.4	45.7	35.4	68.0	46.9	**63.7**	55.6	6.5	6.4	12.1	20.9	15.8	34.3	40.6	14.7	25.4	36.8
	PointNeXt [11]	68.4	**91.8**	**89.4**	43.8	58.7	64.9	**68.5**	52.1	**58.5**	33.2	36.3	26.0	45.1	57.0	37.8	13.5	8.3	45.1	50.2
	SoftGroup [12]	41.4	88.3	62.1	53.1	31.3	**82.2**	60.2	54.8	50.2	23.6	23.9	18.9	57.4	45.5	13.6	18.3	26.4	30.7	38.1
Few-shot (45x8)	PartSLIP* [9]	83.4	85.3	84.8	40.8	65.2	66.0	53.6	42.4	56.3	58.3	88.1	73.7	77.0	69.6	73.5	70.4	60.0	61.3	59.4
	PointNet++ [10]	27.0	42.2	30.2	20.5	22.2	10.5	8.4	7.3	18.1	9.7	11.6	7.0	28.6	31.7	19.4	3.3	0.0	21.8	20.4
	PointNeXt [11]	67.6	65.1	53.7	46.3	59.7	55.4	20.6	22.1	39.2	26.0	47.7	22.6	60.5	66.0	36.8	14.5	0.0	41.5	40.6
	SoftGroup [12]	20.8	80.5	39.7	16.3	38.3	38.3	18.9	24.9	32.8	28.6	40.8	42.9	60.7	54.8	35.6	29.8	14.8	41.1	38.0
	ACD [8]	22.4	39.0	29.2	18.9	39.6	13.7	7.6	13.5	19.2	10.1	31.5	19.4	40.2	51.8	8.9	13.2	0.0	25.6	23.2
	Prototype [13]	60.1	70.8	67.3	33.4	50.4	38.2	30.2	25.7	41.1	32.0	36.8	53.4	62.7	63.3	36.5	35.5	10.1	46.3	44.3
	PartSLIP+	64.8	69.5	59.5	24.5	34.5	37.1	32.0	40.1	35.3	25.5	75.7	15.6	30.5	58.4	31.1	49.4	6.6	26.7	29.9
	Ours	**84.0**	81.0	70.1	**68.4**	47.2	61.2	39.4	45.1	55.1	29.3	71.7	40.1	59.3	**78.8**	59.1	61.6	24.9	59.1	57.6

More Qualitative Results

" Which is the thin, long part that extends behind the ears on either side of the body of the eyeglasses?"
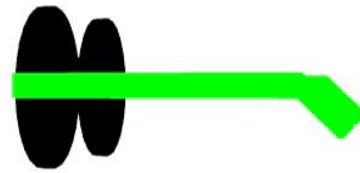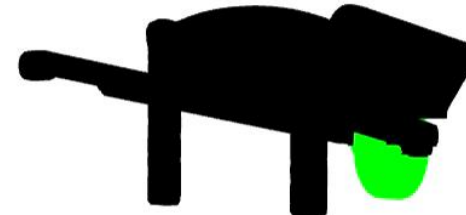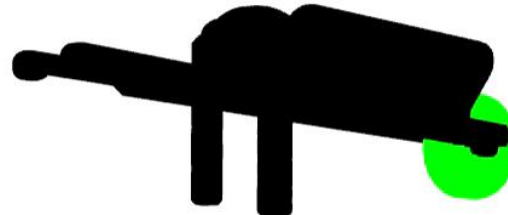
Original	Ground Truth	Ours (PARIS3D)

"Which is the circular part located at the bottom of an in front of this cart that rolls on the ground? Please output a segmentation mask."

Conclusion

We presented:

- Novel challenge within 3D segmentation

- A pipeline that integrates 3D segmentation capabilities into Large Multimodal Models (LMMs)

- Ability to identify part concepts, reason about them, and complement them with world knowledge.