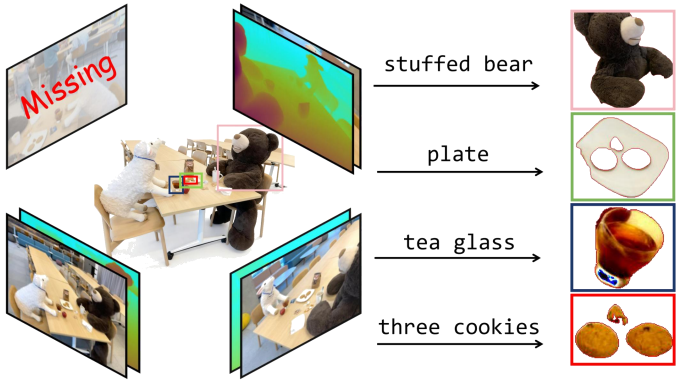
LangRef3DGS: Natural Language-Guided 3D Referential Segmentation from Partial Observations via 3D Gaussian Splatting

Xulun Ye[1]; Qin Zhang[1]; Kun Zhou[2]
[1]Ningbo University, [2]Shenzhen University

CVPR JUNE 3-7, 2026

Introduction

Language-Guided Segmentation in Partial Observability



Language-guided 3D segmentation is a fundamental task for bridging geometric perception and semantic understanding in real-world scenes. However, in practical RGB-D settings, observations are often incomplete due to occlusions, limited viewpoints, and missing frames, which makes small or partially visible objects difficult to recognize and segment reliably. Existing methods usually rely on dense annotations or assume complete visual input, so they often struggle to generalize to unseen categories and partial observations. This motivates a robust framework that can leverage incomplete RGB-D data, discover novel categories, and maintain accurate language-grounded segmentation under challenging real-world conditions.

Methods and Materials

- Dirichlet Process for Novel Class Discovery

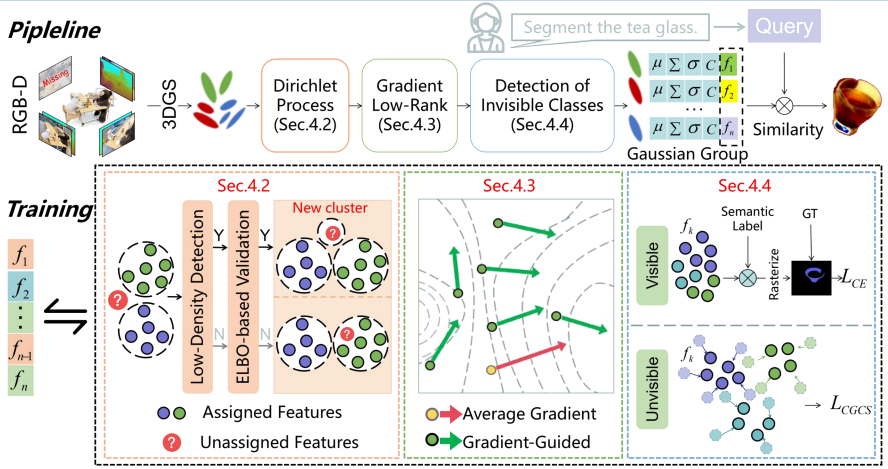
A Dirichlet Process adaptively discovers novel categories from sparse, low-density Gaussian regions under partial observations.

- Gradient Low-Rank Regularization

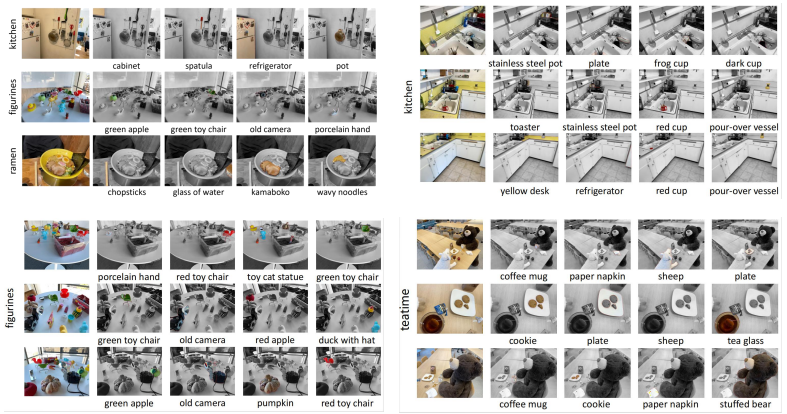
A low-rank constraint reduces feature redundancy and improves class separability.

- Invisible Class Detection and Graph Reasoning

Graph-based reasoning separates and merges visible and invisible classes for better structural consistency.



Experiments & Visualization Results

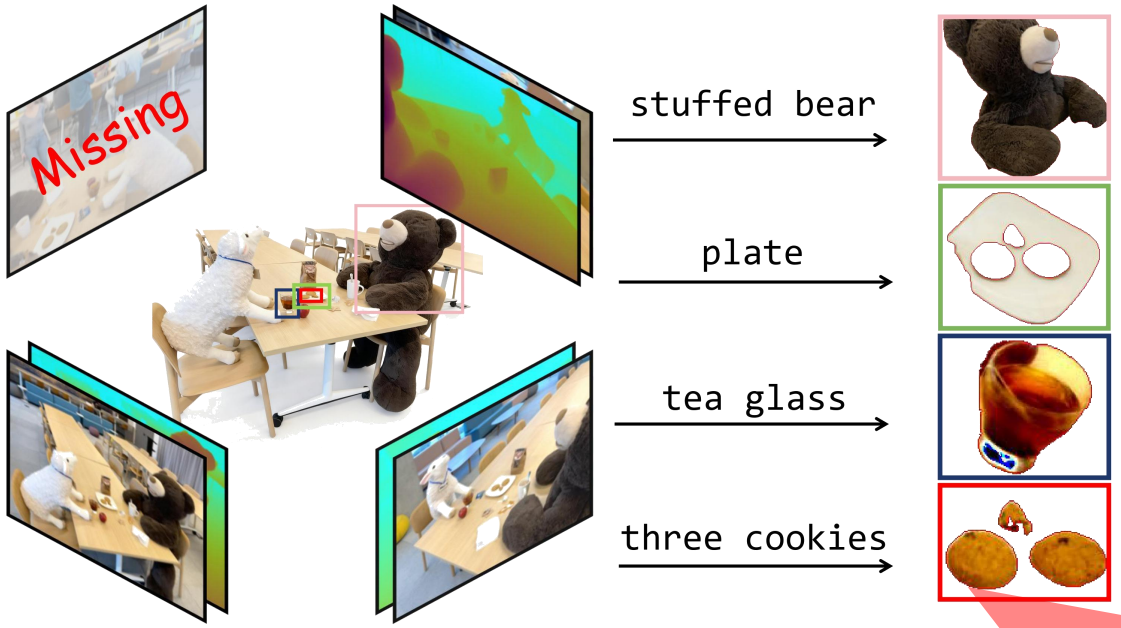


Contributions & Conclusions

- A unified 3DGS-based framework for language-guided 3D referential segmentation
- Dirichlet Process enables adaptive discovery of novel categories
- Gradient Low-Rank regularization improves feature compactness and class separability
- Strong performance in both complete-view and incomplete-view settings
- Robust to occlusion, missing observations, and partially visible objects

Introduction

Language-Guided Segmentation in Partial Observability



- Language-guided 3D segmentation links perception and semantics
- Real RGB-D scenes are often incomplete
- Occlusion, missing views, and partial visibility hurt segmentation
- Existing methods struggle in open-vocabulary partial-view settings

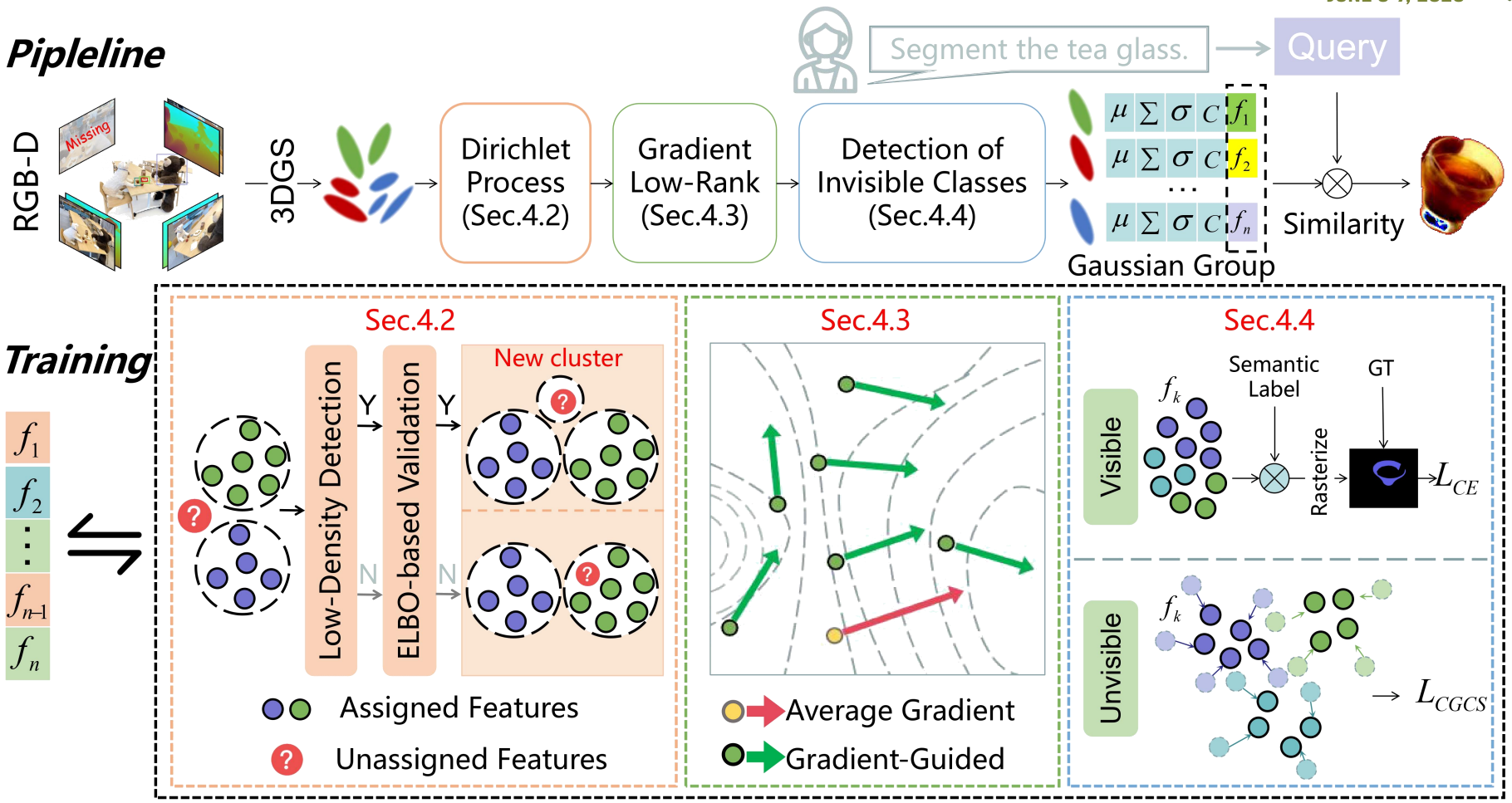
We need

Robust Segmentation From Incomplete RGB-D

Discovery of Unseen Categories

Stable Semantic Representations

Methodology



Three core modules

- Semantic Gaussian Embedding
- Dirichlet Process
- Gradient Low-Rank

Experiments:Main Results

Table 1. Performance comparison on the *LERF-Mask* dataset under the **dense-view** setting, where all RGB-D frames are available.

Method	Venue	figurines		teatime		ramen		mean	
		mIoU	mBIoU	mIoU	mBIoU	mIoU	mBIoU	mIoU	mBIoU
LERF [20]	CVPR2023	33.5	30.6	49.7	42.6	28.3	14.7	37.2	29.3
LangSplat [34]	CVPR2024	52.8	50.5	69.5	65.6	50.4	44.7	57.6	53.6
Gaussian Grouping [48]	ECCV2024	69.7	67.9	71.7	66.1	77.0	68.7	72.8	67.6
CGC [40]	ICPR2024	91.6	88.8	80.5	78.9	68.7	63.1	80.3	76.9
Feature-3DGS [52]	CVPR2024	58.8	52.5	40.5	36.8	43.7	38.3	47.7	45.7
ILGS [16]	ICCV2025	75.9	73.8	81.8	78.8	84.3	75.5	80.5	76.0
OpenSplat3D [31]	CVPR2025	92.3	**89.4**	83.7	78.8	75.9	68.2	84.0	78.8
LangRef3DGS	Ours	**92.8**	88.7	**84.3**	**79.1**	**84.3**	**75.5**	**84.9**	**79.1**

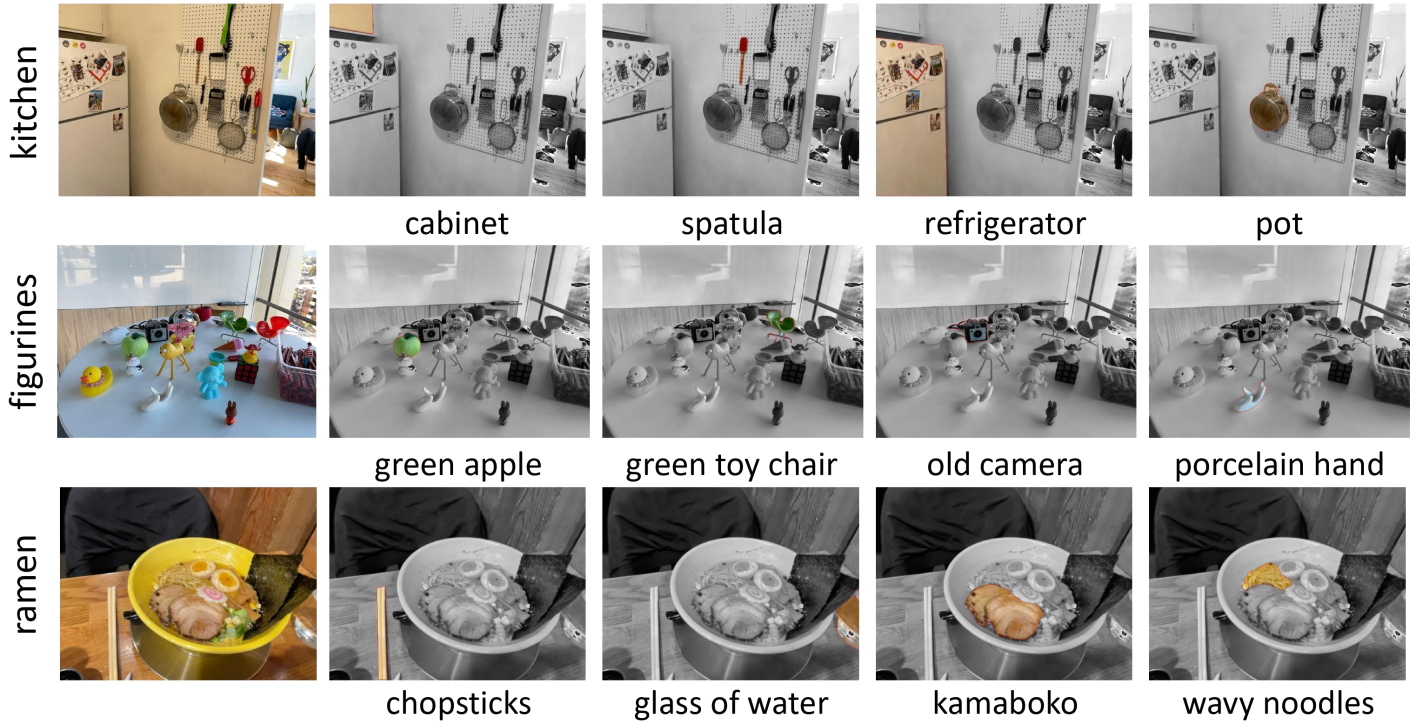
Table 2. Performance comparison on the *LERF-OVS* dataset under the **dense-view** setting, where all RGB-D frames are available.

Method	Venue	figurines		teatime		ramen		waldo_kitchen		mean	
		mIoU	mAcc.	mIoU	mAcc.	mIoU	mAcc.	mIoU	mAcc.	mIoU	mAcc.
LangSplat [34]	CVPR2024	10.16	8.93	11.38	20.34	7.92	11.27	9.18	9.09	9.66	12.41
LEGaussians [39]	CVPR2024	17.99	23.21	19.27	27.12	15.79	26.76	11.78	18.18	16.21	23.82
OpenGaussian [46]	NeurIPS2024	39.29	55.36	60.44	76.27	31.01	42.25	22.70	31.82	38.36	51.43
OpenInsGaussian [15]	ICCV2025	-	-	-	-	-	-	-	-	42.62	62.11
OpenSplat3D [31]	CVPR2025	60.71	85.71	73.27	**88.14**	49.20	76.06	55.63	77.27	59.70	81.79
LangRef3DGS	Ours	**62.52**	**86.65**	**73.41**	87.81	**49.73**	**77.50**	**57.11**	**77.68**	**60.69**	**82.41**

Experiments:Main Results

Table 3. Comparison on LERF-OVS under 20% view removal. Our method consistently outperforms existing approaches across both benchmarks.

Method	Venue	LERF-Mask		LERF-OVS	
		mIoU	mBIoU	mIoU	mAcc.
LERF [20]	CVPR2023	33.9	25.8	–	–
LangSplat [34]	CVPR2024	53.5	49.8	6.4	8.0
Gaussian Grouping [48]	ECCV2024	68.5	63.2	–	–
CGC [40]	ICPR2024	76.8	72.4	–	–
LEGaussians [39]	CVPR2024	–	–	12.4	19.1
OpenGaussian [46]	NeurIPS2024	–	–	34.3	48.6
Feature-3DGS [52]	CVPR2024	43.1	41.0	–	–
OpenInsGaussian [15]	ICCV2025	–	–	38.5	58.4
ILGS [16]	ICCV2025	76.3	71.8	–	–
LangRef3DGS	Ours	79.6	74.9	57.3	78.6



Effective in **both** complete-view and incomplete-view settings

Experiments：Robustness and Ablation

Removed Frames	mIoU	mAcc.
10%	60.5	81.2
20%	57.3	78.6
30%	54.0	75.5
40%	50.8	72.3

Configuration	mIoU	mAcc.
Baseline (without DP/GLR/CGSL)	48.2	69.1
+ Dirichlet Process (DP)	51.0	72.4
+ Gradient Low-Rank (DP + GLR)	54.1	75.2
+ Contrastive Graph Semantic Loss (Full model)	57.3	78.6

Conclusion

1. 3DGS-based continuous semantic field
2. DP for novel class discovery
3. GLR for compact and robust features
4. *Strong performance in complete and incomplete views*
5. *Unified framework for language-guided 3D segmentation under partial observations*