Generated and Pseudo Content guided Prototype Refinement for Few-shot Point Cloud Segmentation

Lili Wei[1,2], Congyan Lang[1,2*], Ziyi Chen[1,2], Tao Wang[1,2], Yidong Li[1,2], Jun Liu[3]

[1]Beijing Jiaotong University

[2]Key Laboratory of Big Data & Artificial Intelligence in Transportation, Ministry of Education

[3]Lancaster University

Introduction

- **Motivation**

Challenge: In few-shot 3D point cloud semantic segmentation (FS-3DSeg), existing prototype-based methods face issues with ***low prototype quality*** due to:

- ***Semantic information constraints*** — Limited support point clouds contain only partial and incomplete object information, lacking intra-class diversity and inter-class discrimination.
- ***Class information bias*** — Intra-class object variations and feature distribution gaps between query and support point clouds.

Consequently, ***vanilla 3D prototypes lack comprehensive class information and are unsuitable for guiding query segmentation***.

Our goal: We aim to compensate for the lack of semantics in 3D support set to generate comprehensive and reliable query-specific prototypes for accurately segmenting the query point cloud.

Introduction

- **Contribution**

  - We propose ***GPCPR***, a novel end-to-end FS-3DSegframework that enhances prototype quality by simultaneously integrating ***LLM-generated content*** and ***reliable query context*** to generate query-specific prototypes. To the best of our knowledge, this is the first time leveraging LLM's capabilities to segment novel classes in FS-3DSeg.
  - We design a series of novel modules, including the ***Generated Content-guided Prototype Refinement (GCPR)*** module and the ***Pseudo Query Context-guided Prototype Refinement (PCPR)*** module, to facilitate the prototype refinement process. Additionally, we design a ***dual-distillation regularization term*** to further mutually enhance the refinement.
  - Extensive experiments demonstrate the superiority of our method, notably exceeding state-of-the-art methods by up to 12.10% and 13.75% on S3DIS and ScanNet datasets, respectively.

Method

- **Framework**

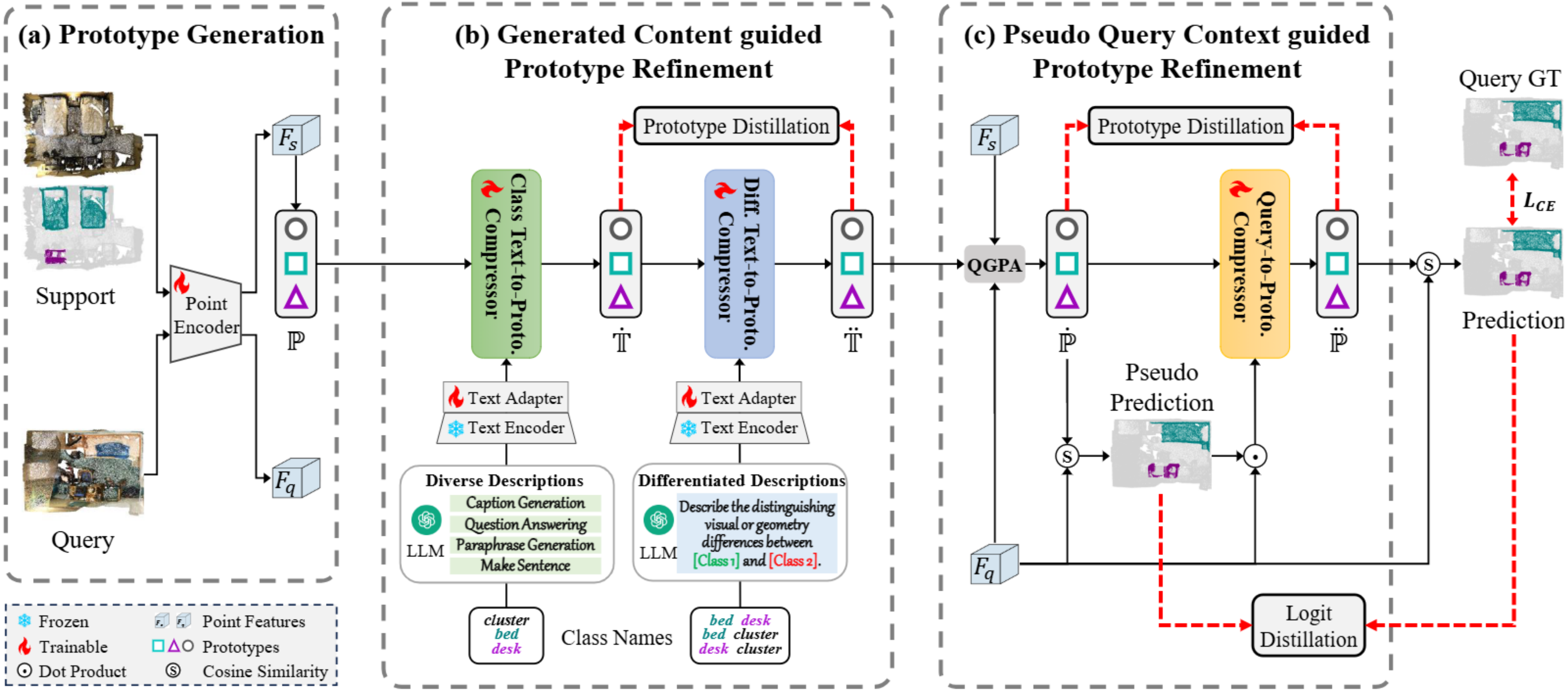


Figure 1: GPCPR framework overview.

Method

- **LLM-driven Generated Content-guided Prototype Refinement**

➢ **LLM-driven Content Generation**

① **Caption Generation:** "Describe a point cloud of a [CLASS] in one sentence."

② **Question Answering:** "What is a [CLASS] point cloud like?"

③ **Paraphrase Generation:** "Generate a synonym: A point cloud of a [CLASS]."

④ **Make Sentence:** "Make a sentence with words: point cloud, [CLASS], obscure. "

➢ **Differentiated Class Descriptions**

"Describe the distinguishing visual or geometry differences between the point clouds of [CLASS 1] and [CLASS 2] in pairs of sentences."

➢ **Text-To-Prototype Compressors**

① **Class text-to-prototype compressor:** $\dot{\mathbb{T}} = \{\dot{\mathcal{T}}^n\}_{n=0}^{N}$

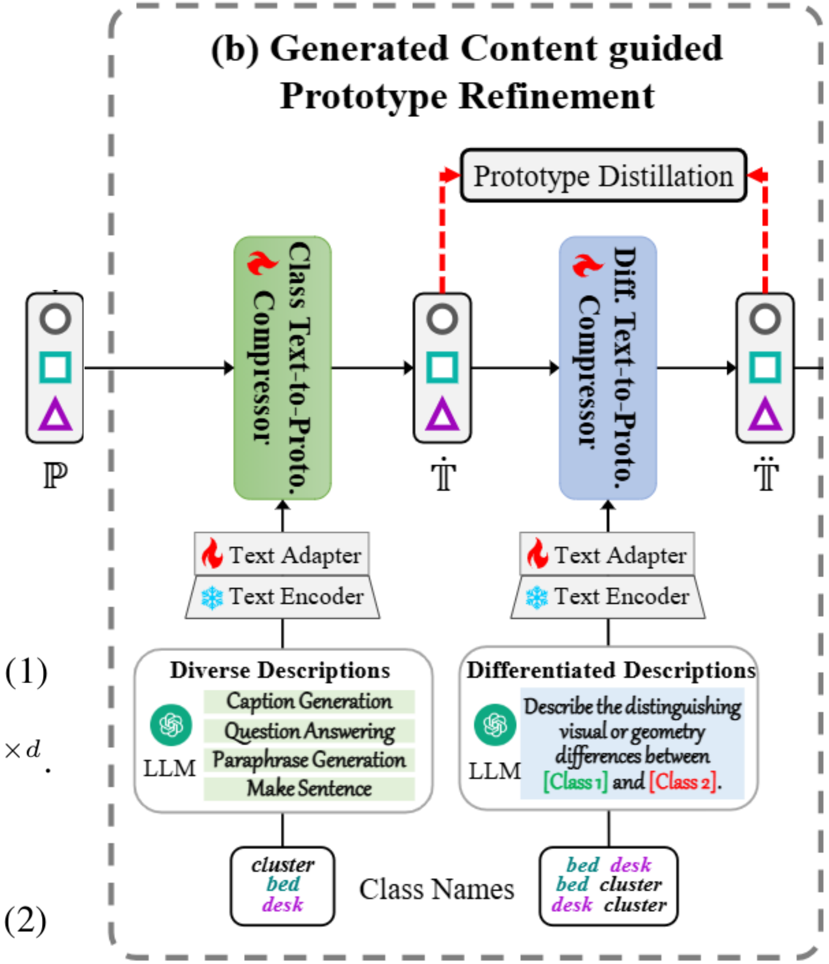
$$\dot{\mathcal{T}}^n = \mathcal{P}^n + softmax(\mathbf{Q_1}\mathbf{K_1}^\top)\mathbf{V_1}, \quad n \in \{0, ..., N\}, \tag{1}$$

where $\mathbf{Q_1} = \mathcal{P}^n W_{q1} \in \mathbb{R}^{1\times d'}$, $\mathbf{K_1} = \mathbf{E}^n W_{k1} \in \mathbb{R}^{(4\times N_{div})\times d'}$ and $\mathbf{V_1} = \mathbf{E}^n W_{v1} \in \mathbb{R}^{(4\times N_{div})\times d}$.

② **Differentiated text-to-prototype compressor:** $\ddot{\mathbb{T}} = \{\ddot{\mathcal{T}}^n\}_{n=0}^{N}$

$$\ddot{\mathcal{T}}^n = \dot{\mathcal{T}}^n + softmax(\mathbf{Q_2}\mathbf{K_2}^\top)\mathbf{V_2}, \quad n \in \{0, ..., N\}, \tag{2}$$

where $\mathbf{Q_2} = \dot{\mathcal{T}}^n W_{q2} \in \mathbb{R}^{1\times d'}$, $\mathbf{K_2} = \mathbf{E}^{n'} W_{k2} \in \mathbb{R}^{N_{diff}\times d'}$, and $\mathbf{V_2} = \mathbf{E}^{n'} W_{v2} \in \mathbb{R}^{N_{diff}\times d}$.



Method

- **Pseudo Query Context-guided Prototype Refinement**

  - **Pseudo-Query Context Generation**

    ① **QGPA:** rectify prototypes to query feature channel distribution

$$\dot{\mathbb{P}} = \{\dot{\mathbb{P}}^i\}_{i=1}^{T} \in \mathbb{R}^{T\times(N+1)\times d}$$

$$\dot{\mathcal{P}}^{i,n} = \ddot{\mathcal{T}}^n + softmax(\mathbf{Q_3}\mathbf{K_3}^\top)\mathbf{V_3}, \quad n \in \{0,...,N\}, i \in \{1,...,T\}, \tag{3}$$

where $\mathbf{Q_3} = \mathbf{F}_q^{i\top} W_{q3} \in \mathbb{R}^{d\times M'}$, $\mathbf{K_3} = \mathbf{F}_s^{n\top} W_{k3} \in \mathbb{R}^{d\times M'}$, and $\mathbf{V_3} = \ddot{\mathcal{T}}^n W_{v3} \in \mathbb{R}^{1\times d}$. $\mathbf{F}_s^n \in \mathbb{R}^{M\times d}$

    ② **Extract class-specific pseudo query context**:

    predict pseudo logits and mask query features with pseudo masks

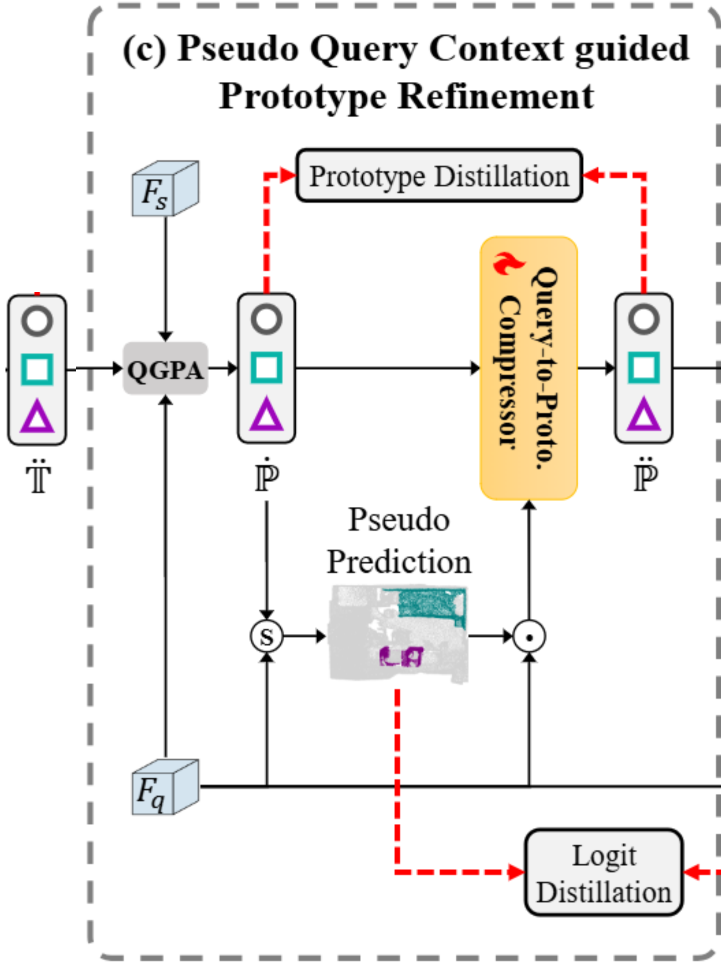
$$\dot{\mathbf{F}}_q = \{\dot{\mathbf{F}}_q^n\}_{n=0}^{N}, \dot{\mathbf{F}}_q^n \in \mathbb{R}^{(T\times M)\times d}$$

  - **Query-To-Prototype Compressor**

$$\ddot{\mathbb{P}} = \{\ddot{\mathbb{P}}^i\}_{i=1}^{T} \in \mathbb{R}^{T\times(N+1)\times d}$$

$$\ddot{\mathcal{P}}^{i,n} = \dot{\mathcal{P}}^{i,n} + softmax(\mathbf{Q_4}\mathbf{K_4}^\top)\mathbf{V_4}, \quad n \in \{1,...,N\}, \quad i \in \{1,...,T\}, \tag{4}$$

where $\mathbf{Q_4} = \dot{\mathcal{P}}^{i,n} W_{q4} \in \mathbb{R}^{1\times d'}$, $\mathbf{K_4} = \dot{\mathbf{F}}_q^n W_{k4} \in \mathbb{R}^{(T\times M)\times d'}$, $\mathbf{V_4} = \dot{\mathbf{F}}_q^n W_{v4} \in \mathbb{R}^{(T\times M)\times d}$.



Method

- **Dual-Distillation Regularization**

  - **Prototype Distillation**

    Goal: achieving mutually beneficial and bi-directional optimization of multi-stage prototypes

    ① **Student:** early-stage prototypes

    ② **Teacher:** deep-stage prototypes

$$\mathcal{L}_{TP} = KL(\dot{\mathbb{T}}||\ddot{\mathbb{T}}), \quad \mathcal{L}_{QP} = KL(\dot{\mathbb{P}}||\ddot{\mathbb{P}}), \tag{5}$$

  - **Pseudo Prediction Distillation**

    Goal: improving the accuracy and quality of pseudo masks

    ① **Student:** pseudo logits

    ② **Teacher:** final predicted logits

$$\mathcal{L}_{QM} = KL(\dot{\mathbf{L}}_q||\hat{\mathbf{L}}_q). \tag{6}$$

- **Objective**

$$\mathcal{L}_{SEG} = \mathcal{L}_{CE}(\dot{\mathbf{M}}_q, \hat{\mathbf{M}}_q). \tag{7}$$

$$\mathcal{L}_{total} = \mathcal{L}_{SEG} + \lambda \times (\mathcal{L}_{TP} + \mathcal{L}_{QP} + \mathcal{L}_{QM}). \tag{8}$$

Experiments

Comparison With State-of-the-Art Methods

Method	2-way						3-way					
	1-shot			5-shot			1-shot			5-shot		
	S^0	S^1	**mean**	S^0	S^1	**mean**	S^0	S^1	**mean**	S^0	S^1	**mean**
ProtoNet [39]	48.39	49.98	49.19	57.34	63.22	60.28	40.81	45.07	42.94	49.05	53.42	51.24
AttMPTI [39]	53.77	55.94	54.86	61.67	67.02	64.35	45.18	49.27	47.23	54.92	56.79	55.86
BFG [17]	55.60	55.98	55.79	63.71	66.62	65.17	46.18	48.36	47.27	55.05	57.80	56.43
SCAT [36]	54.92	56.74	55.83	64.24	69.03	66.63	-	-	-	-	-	-
QGPNet [10]	56.30	57.62	56.96	65.34	69.01	67.17	47.00	50.12	48.56	55.80	58.54	57.17
2CBR [41]	55.89	61.99	58.94	63.55	67.51	65.53	46.51	53.91	50.21	55.51	58.07	56.79
QGE [20]	58.85	60.29	59.57	66.56	79.46	73.01	-	-	-	-	-	-
QGPA [9]	59.45	66.08	62.76	65.40	70.30	67.85	48.99	56.57	52.78	61.27	60.81	61.04
DPA [16]	66.08	74.30	70.19	71.10	77.03	74.07	50.67	59.53	55.10	64.52	63.34	63.93
Ours	**74.04**	**77.44**	**75.74**	**76.65**	**78.22**	**77.44**	**62.77**	**70.57**	**66.67**	**67.49**	**74.68**	**71.09**

Table 1: Performance on S3DIS dataset using mean-IoU metric (%).

Method	2-way						3-way					
	1-shot			5-shot			1-shot			5-shot		
	S^0	S^1	**mean**	S^0	S^1	**mean**	S^0	S^1	**mean**	S^0	S^1	**mean**
ProtoNet [39]	33.92	30.95	32.44	45.34	42.01	43.68	28.47	26.13	27.30	37.36	34.98	36.17
AttMPTI [39]	42.55	40.83	41.69	54.00	50.32	52.16	35.23	30.72	32.98	46.74	40.80	43.77
BFG [17]	42.15	40.52	41.34	51.23	49.39	50.31	34.12	31.98	33.05	46.25	41.38	43.82
SCAT [36]	45.24	45.90	45.57	55.38	57.11	56.24	-	-	-	-	-	-
QGPNet [10]	44.63	42.18	43.40	54.75	51.81	53.28	37.86	34.50	36.18	47.45	42.74	45.09
2CBR [41]	50.73	47.66	49.20	52.35	47.14	49.75	47.00	46.36	46.68	45.06	39.47	42.27
QGE [20]	43.10	46.79	44.95	51.91	57.21	54.56	-	-	-	-	-	-
QGPA [9]	57.08	55.94	56.51	64.55	59.64	62.10	55.27	55.60	55.44	59.02	53.16	56.09
DPA [16]	62.75	63.04	62.90	67.19	64.62	65.91	61.97	61.72	61.85	66.13	64.67	65.40
Ours	**75.94**	**71.92**	**73.93**	**78.42**	**78.37**	**78.40**	**70.00**	**66.61**	**68.31**	**76.73**	**68.63**	**72.68**

Table 2: Performance on ScanNet dataset using mean-IoU metric (%).

Computational Complexity

Phase	S3DIS	ScanNet
Description Generation: gpt-3.5-turbo	30.23 min	67.15 min
Text Feature Extraction: CLIP rn50	10.95 s	17.79 s
Total	30.41 min	67.45 min

Table 3: Offline time cost

Methods	#Params	FLOPs (G)	FPS	Inference Time (ms)	S3DIS	ScanNet
attMPTI	357.82K	152.65	1.47	678.67	54.86	41.69
QGPA	2.79M	16.30	38.68	25.85	62.76	56.51
DPA	4.85M	15.49	32.35	30.91	70.19	62.90
Ours	4.22M	18.96	20.57	48.61	**75.74**	**73.93**

Table 4: Online time cost

Achieving Stare-of-the-Art Performance with Moderate Computational Cost.

Experiments

- **Qualitative results**

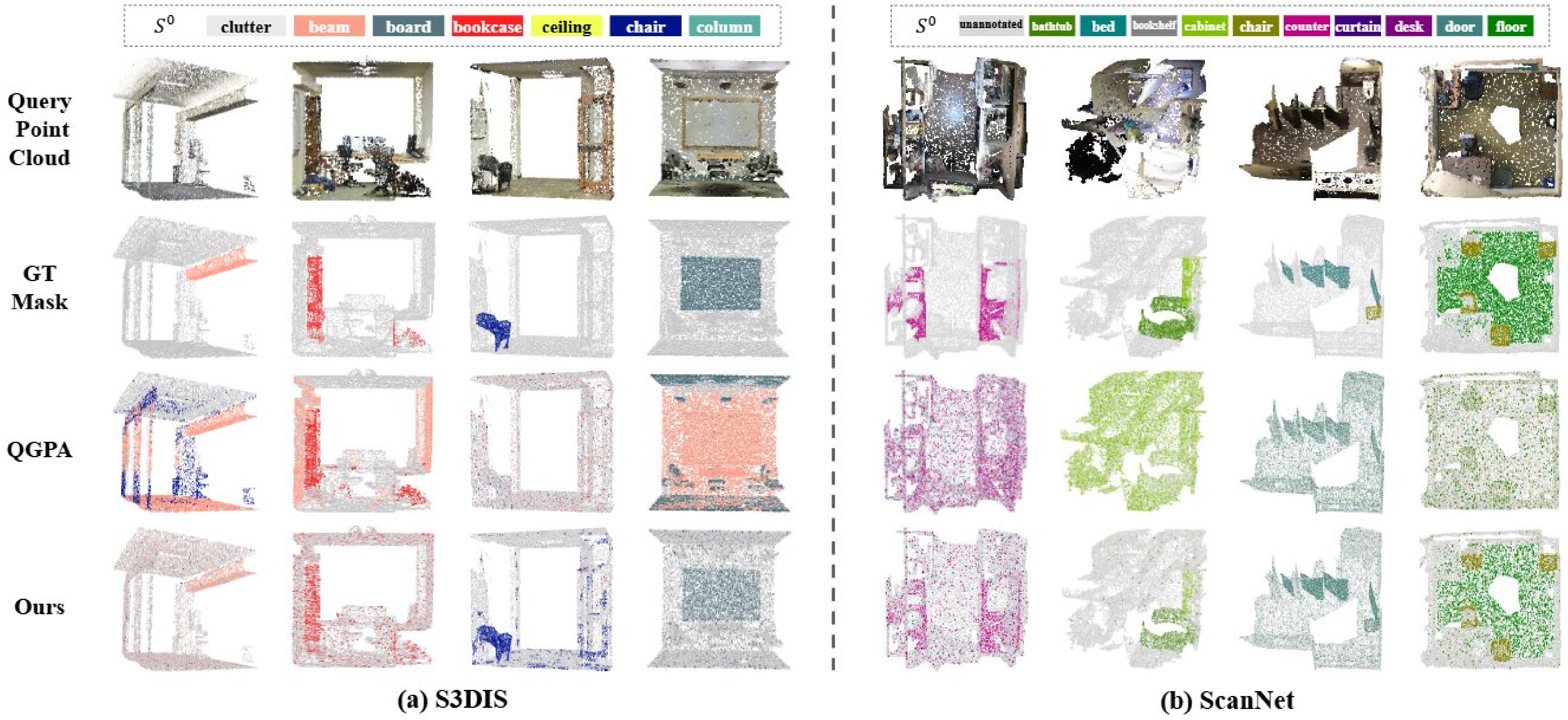


Figure 2: Qualitative results.

Experiments

- **Ablation Study**

GCPR D	GCPR D'	PCPR	DD loss $\mathcal{L}_{TP}$	DD loss $\mathcal{L}_{QP}$	DD loss $\mathcal{L}_{QM}$	2-way 1-shot S^0	2-way 1-shot S^1	2-way 1-shot mean
						58.96	63.08	61.02
		✓				65.01	74.39	69.70
✓						66.06	69.03	67.55
	✓					66.71	70.18	68.45
✓	✓					68.57	74.71	71.64
✓	✓	✓				68.68	75.73	72.21
✓	✓	✓	✓			69.36	75.85	72.61
✓	✓	✓		✓		71.09	76.09	73.59
✓	✓	✓			✓	71.07	76.47	73.77
✓	✓	✓	✓	✓	✓	**74.04**	**77.44**	**75.74**

Table 5: Ablation study of key components.

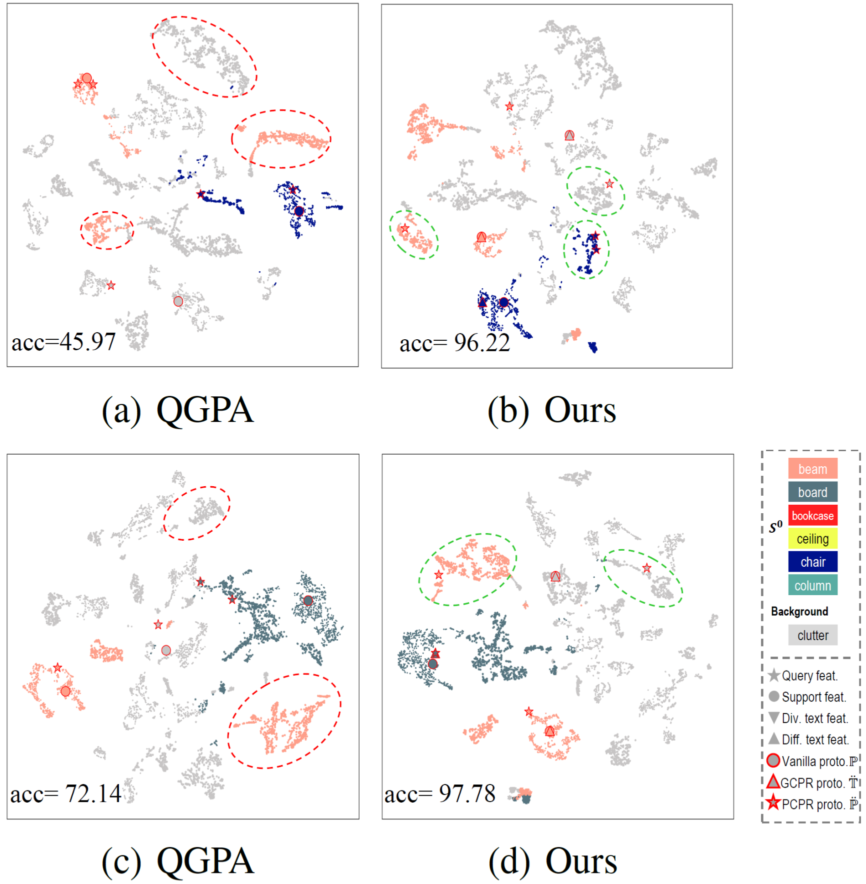
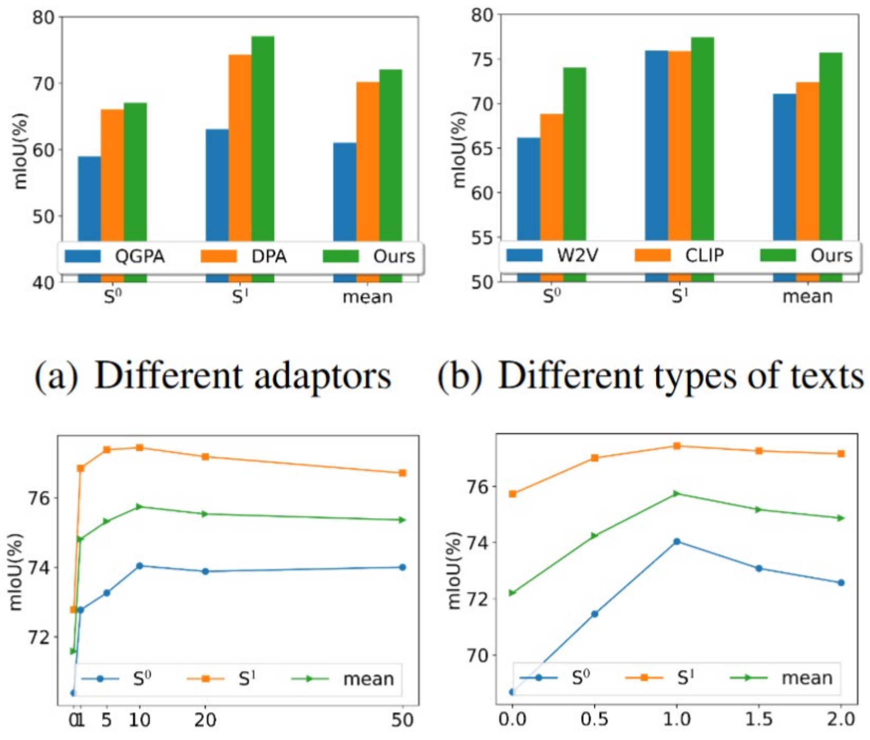


Figure 3: Visualization of feature distribution and prototype distribution.

Experiments

- **Ablation Study**



(a) Different adaptors (b) Different types of texts

(c) Text number N_{div} (d) Coefficient λ

Figure 4: Ablation study of modules and hyper-parameters.

Methods	S^0	S^1	mean
attMPTI	53.77	55.94	54.86
QGPA	59.45	66.08	62.76
DPA	66.08	74.30	70.19
Ours (gpt-4o-mini)	71.64	76.11	73.88
Ours (gpt-3.5-turbo)	**74.04**	**77.44**	**75.74**

Table 6: Effects of different LLMs.

Contact Us

Fell free to contact me:

- **Email:** *20112014@bjtu.edu.cn*
- **WeChat:** *W1248776580*

THANKS