Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Reliability-aware Conflict Calibration for Generalized Few-shot Point Cloud Segmentation

He Wang
Zhejiang University
Guofeng Zhang
zhangguofeng@cad.zju.edu.cn

Zhejiang University

Research Article

Keywords: Point cloud segmentation, Generalized few-shot segmentation, 3D scene understanding, Base–novel conflict modeling

Posted Date: June 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9870233/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at The Visual Computer on October 25th, 2026. See the published version at https://doi.org/10.1007/s00371-026-04621-9.

Reliability-aware Conflict Calibration for Generalized Few-shot Point Cloud Segmentation

He Wang · Guofeng Zhang

Abstract Generalized few-shot point cloud segmentation aims to segment novel classes from a few support samples while maintaining accurate predictions for base classes, which is important for robust 3D scene understanding. Most existing methods address this problem by strengthening support representations or introducing external priors. However, the main difficulty is not just limited support information, but making reliable decisions in ambiguous regions where base and novel classes are hard to distinguish. This calls for resolving localized conflicts rather than uniformly enriching representations. Therefore, we propose a reliability-aware conflict localization and calibration framework. The method first builds hierarchical prototypes and estimates their reliability, then localizes ambiguous base–novel conflict regions, and finally calibrates only those regions to correct uncertain predictions while preserving stable base predictions elsewhere. The framework can be further extended with an optional textual prototype augmentation module to complement support-derived semantics. Experiments on multiple datasets show state-of-the-art performance under both 1-shot and 5-shot settings.

Keywords Point cloud segmentation · Generalized few-shot segmentation · 3D scene understanding · Base–novel conflict modeling

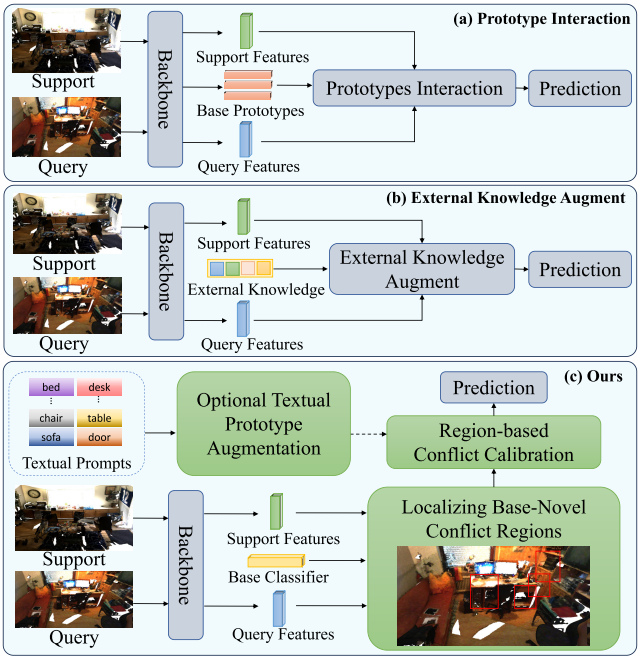


Fig. 1 Comparison of representative methods. (a) Prototype interaction and representation enhancement methods; (b) external knowledge augmentation methods; (c) our method localizes base–novel conflict regions and applies reliability-aware selective calibration, rather than uniformly enhancing representations or globally injecting external priors

1 Introduction

Fully supervised 3D segmentation has achieved strong performance on large-scale datasets [13,33,22]. However, it relies on dense point-wise labels, which are expensive to obtain for newly emerging categories. Few-shot point cloud segmentation (FS-PCS) addresses this issue by adapting to novel classes from only a few annotated support samples [24]. However, conventional FS-PCS typically treats novel classes in isolation and ignores the more practical setting in which base and novel classes coexist in the same scene. Generalized few-shot point cloud segmentation (GFS-PCS) targets this setting by requiring the model to recognize novel classes

He Wang · Guofeng Zhang
State Key Lab of CAD&CG, Zhejiang University, Hangzhou, Zhejiang, China
E-mail: zhangguofeng@cad.zju.edu.cn

from limited support while maintaining accurate predictions for base classes [31]. This capability is important for robust 3D scene understanding and content analysis.

In GFS-PCS, sparse sampling, incomplete observations, and local geometric ambiguity often make prediction errors more localized rather than uniformly distributed across the scene [1,19]. In practice, unreliable predictions tend to cluster in base–novel conflict regions, where limited support samples make base classes and novel classes hard to distinguish [26]. This suggests that merely enriching representations or introducing external priors is insufficient; instead, the key challenge is to make stable local corrections by identifying conflict regions and calibrating them while preserving reliable predictions elsewhere [8]. While existing GFS-PCS methods mainly follow two directions, prototype interaction and representation enhancement [34,35], and external knowledge augmentation [1,16], as illustrated in Fig. 1, they address base–novel conflicts only indirectly, without explicitly localizing where such conflicts arise or selectively addressing them.

Motivated by these localized conflicts in 3D point clouds, we propose a reliability-aware conflict localization and calibration framework for GFS-PCS. To enable stable local correction, the framework first estimates the reliability of hierarchical support prototypes, then localizes pair-consistent base–novel conflict regions, and finally calibrates only those regions to correct ambiguous predictions while preserving stable base predictions elsewhere. The proposed framework can be further extended with an optional textual prototype augmentation (TPA) module when sparse support alone provides insufficient semantic coverage, making it more useful for 3D scene analysis under limited supervision. Our contributions can be summarized as follows:

- We propose a reliability-aware conflict localization and calibration GFS-PCS framework, which explicitly resolves base and novel conflicts through prototype reliability estimation, conflict localization, and selective calibration, thereby improving generalized segmentation while preserving stable base predictions.
- We further extend the framework with an optional textual prototype augmentation module. Experiments show that textual priors are more effective when incorporated in a conflict-conditioned, reliability-aware manner rather than global fusion.
- Extensive experiments on multiple datasets show that the core framework already achieves strong performance by improving the base–novel trade-off, while the text-augmented version further achieves SOTA results.

2 Related Work

2.1 Generalized Few-shot Segmentation in 2D

Generalized few-shot semantic segmentation (GFSS) in 2D aims to predict both base and novel classes in a unified label space from only a few annotated support examples, making it an important setting for realistic few-shot segmentation [27]. Early works mainly focused on sparse novel-class modeling through better prototype learning and representation refinement [10,11], while more recent methods introduced external semantic priors such as textual or vision-language priors [5,20].

Beyond enriching novel-class representations, several recent studies have begun to address the base–novel competition that arises in the generalized prediction space more directly. POP [17] reduces base–novel interference by learning orthogonal or decoupled prototypes, while HBNC [18] harmonizes base and novel categories through a class-contrastive formulation. PCN [19] explicitly studies prediction calibration to mitigate biased score distributions between base and novel classes, and BCM [26] further shows that carefully calibrated decision rules can provide a strong GFSS baseline. These works show that additional representations are most effective when properly calibrated under base–novel competition, motivating our study of 3D GFS-PCS, where such calibration is especially important but cannot be directly handled by 2D methods designed for dense image features and pixel-wise prediction.

In point clouds, few-shot support samples provide sparse and uneven coverage of novel-class geometry, making support-derived prototypes less stable; meanwhile, base–novel ambiguity is spatially structured and often concentrates in local neighborhoods. Our method therefore adopts a structure-aware calibration design for 3D point clouds: reliability-aware hierarchical prototypes with slot-based local prototypes, conflict regions formed by grouping neighboring ambiguous points with consistent candidate base–novel pairs, and region-level pairwise calibration that leaves stable predictions unchanged. Thus, our difference from 2D GFSS calibration lies not only in the target modality, but also in how base–novel conflicts are represented, localized, and calibrated under sparse 3D observations.

2.2 Generalized Few-shot Point Cloud Segmentation

GFS-PCS extends few-shot point cloud segmentation (FS-PCS) to a more practical setting, where the model must recognize novel classes from only a few support samples while maintaining stable predictions for base classes. This task is particularly challenging because

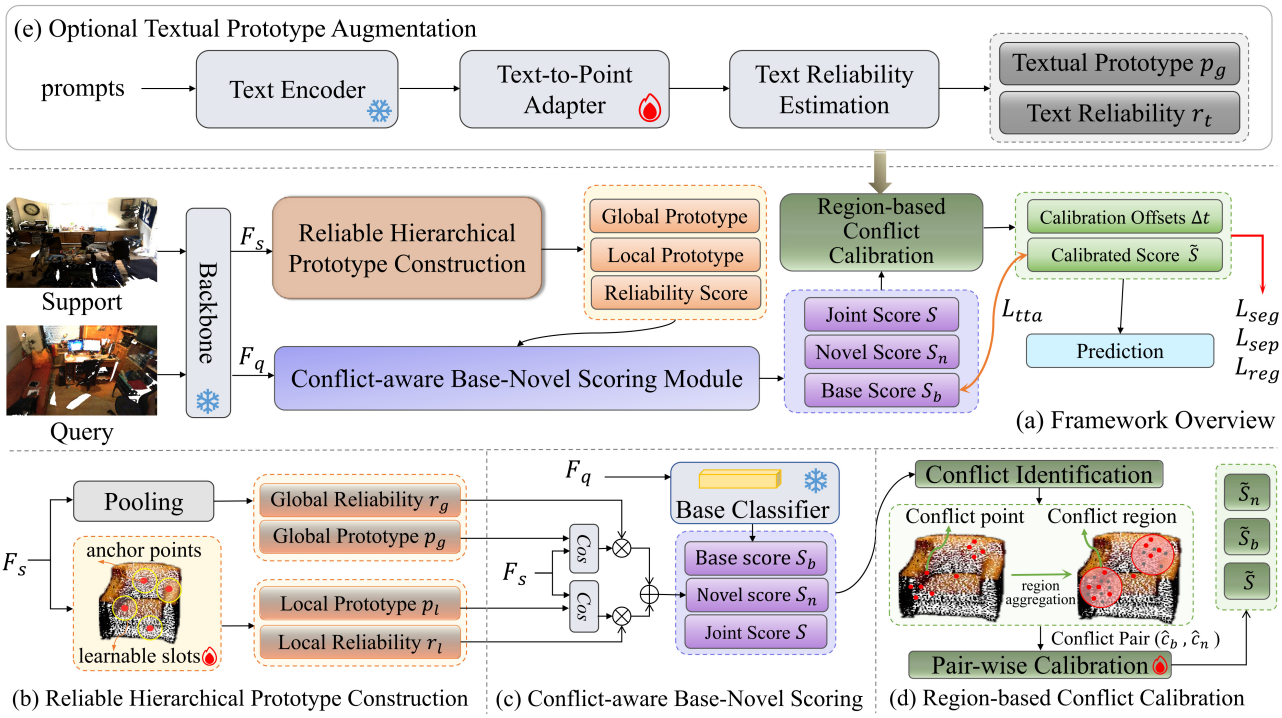


Fig. 2 Overview of the proposed support-driven reliability-aware conflict localization and calibration framework for GFS-PCS. The core framework estimates prototype reliability from the support set, localizes ambiguous base–novel conflict regions, and selectively calibrates the dominant class pair in each retained region. It can be further extended with an optional textual prototype augmentation (TPA) module that injects textual priors only into the conflict region

sparse, irregular, and incomplete point-cloud observations make support-derived representations less reliable and local ambiguities more prominent.

Existing GFS-PCS works mainly follow two directions. The first line improves support-derived representations through prototype interaction, prototype refinement, or representation enhancement. PIFS [4] introduces prototype-based incremental learning for few-shot segmentation, enabling novel classes to be added while mitigating forgetting. attMPTI [38] further represents each class with multiple prototypes to better capture the diverse feature distributions of 3D point clouds. GW [34] transfers reusable geometric words from base to novel classes. PE [28] uses pseudo-embeddings to simulate novel-class learning during episodic training, thereby improving adaptation to unseen classes. The second line introduces external priors beyond the few-shot support samples to compensate for limited novel-class information. LARM [16] enriches novel prototypes with LLM-assisted semantics and hyper-relation matching. These methods improve novel-class recognition by making novel representations stronger or semantically richer. GFS-VL [1] leverages 3D vision-language knowledge to complement sparse support annotations.

Despite these advances, existing GFS-PCS methods still primarily focus on improving novel-class representations, without explicitly modeling the localized base–novel conflicts that arise in the query point cloud. As a result, they lack an explicit mechanism to localize conflict regions and selectively calibrate them while preserving reliable base predictions.

Meanwhile, FS-PCS methods are largely built on a prototype-centric episodic pipeline, with advances in prototype construction [14,38,29], support–query interaction [21,37,39], and prototype refinement [9,30]. While these methods provide an important foundation for GFS-PCS, they are developed for conventional FS-PCS rather than the generalized setting. Our method extends this paradigm to GFS-PCS via explicit base–novel conflict localization and calibration.

3 Method

3.1 Problem Setup and Method Overview

In GFS-PCS, the label space is split into base classes $\mathcal{C}_\mathrm{b}$ and novel classes $\mathcal{C}_\mathrm{n}$, with $\mathcal{C}_\mathrm{b} \cap \mathcal{C}_\mathrm{n} = \varnothing$. During training, the model is supervised only on base classes. At test time, each episode provides a few-shot support set $\mathcal{S}$ for the novel classes together with a query point cloud $\mathcal{Q}$. The goal is to assign each query point a label from

the unified space $\mathcal{C}_{\mathrm{b}} \cup \mathcal{C}_{\mathrm{n}}$. We use a shared backbone to extract point features and obtain base–class scores using the classifier learned on base classes.

As shown in Fig. 2, our method is motivated by the observation that, in GFS-PCS, the key is not to uniformly enrich representations [5,12], but to make reliable decisions in the small number of ambiguous base–novel conflict regions where most errors concentrate under limited support [15]. We propose a reliability-aware conflict localization and calibration framework with three stages: prototype reliability estimation, conflict localization, and selective calibration. The framework first constructs hierarchical reliability-aware novel prototypes from the support set, then localizes ambiguous base–novel conflict regions in the query scene through base–novel scoring, and finally calibrates only those regions while leaving stable predictions elsewhere unchanged. It can be further extended with an optional textual prototype augmentation (TPA) module, which complements support-derived prototypes with textual priors to further improve novel-class recognition.

3.2 Prototype Reliability Estimation for Base–Novel Scoring

In 3D scenes, support prototypes for novel classes are often unreliable due to intra-class variation and viewpoint distortion. To address this issue, we introduce a prototype reliability estimation (PRE) module and a unified base–novel scoring (BNS) module. The PRE module constructs global and local support prototypes and estimates their reliability scores, while the BNS module integrates these reliability-aware novel prototypes with the base classifier to produce unified base–novel scores for query points.

For each novel class $c \in \mathcal{C}_{\mathrm{n}}$, we construct two complementary types of prototypes: a global prototype $\mathbf{p}_{\mathrm{g}}^{c}$ and local prototypes $\mathbf{p}_{\mathrm{l}}^{c,m}$, where the local prototypes are obtained using M learnable slots to aggregate support foreground patch tokens into reusable part-level patterns. The global prototype captures class-level semantics, while the local prototypes preserve finer local structure for novel-class scoring.

Since different support prototypes are not equally reliable, we estimate a reliability score for each prototype before novel-class scoring. Here, $u_{\text{cls-shot}}^{c}$ measures the consistency of the class representation across support shots, and u_{base}^{c} measures its separation from base classes. The global reliability is defined as

$$r_{\mathrm{g}}^{c} = \sigma(\alpha_1 u_{\text{cls-shot}}^{c} - \alpha_2 u_{\text{base}}^{c}), \tag{1}$$

where $\sigma(\cdot)$ is the sigmoid function, and α_1 and α_2 are learnable parameters. For each local prototype, we assess whether it captures a compact, reusable, and class-discriminative local pattern. Its reliability is estimated by jointly considering local compactness $u_{\text{comp}}^{\mathrm{l}}$, cross-shot coverage $u_{\text{cov}}^{\mathrm{l}}$, and base ambiguity $u_{\text{base}}^{\mathrm{l}}$:

$$\begin{aligned} r_{\mathrm{l}}^{c,m} = \sigma\big(&- \beta_1 u_{\text{comp}}^{\mathrm{l}}(c,m) + \beta_2 u_{\text{cov}}^{\mathrm{l}}(c,m) \\ &- \beta_3 u_{\text{base}}^{\mathrm{l}}(c,m)\big). \end{aligned} \tag{2}$$

where $\sigma(\cdot)$ is the sigmoid function, and β_1, β_2, and β_3 are learnable parameters. For each novel class c, we compute reliability-weighted global and local scores as

$$s_{\mathrm{g}}^{c}(j) = r_{\mathrm{g}}^{c} \cdot \cos\big(\mathbf{f}_j^{\mathrm{q}}, \hat{\mathbf{p}}_{\mathrm{g}}^{c}\big), \tag{3}$$

$$s_{\mathrm{l}}^{c}(j) = \max_{m=1,\dots,M} \big(r_{\mathrm{l}}^{c,m} \cdot \cos\big(\mathbf{f}_j^{\mathrm{q}}, \hat{\mathbf{p}}_{\mathrm{l}}^{c,m}\big)\big). \tag{4}$$

The global score captures class-level semantics, while the local score reflects the strongest match to reliable local patterns. The final score for novel class c is

$$z_{\mathrm{n}}^{c}(j) = \lambda_{\mathrm{g}} s_{\mathrm{g}}^{c}(j) + \lambda_{\mathrm{l}} s_{\mathrm{l}}^{c}(j), \quad \lambda_{\mathrm{g}} + \lambda_{\mathrm{l}} = 1. \tag{5}$$

The final novel-class score and the corresponding label are then defined as

$$s_{\mathrm{n}}(j) = \max_{c \in \mathcal{C}_{\mathrm{n}}} z_{\mathrm{n}}^{c}(j), \quad \hat{c}_{\mathrm{n}}(j) = \arg\max_{c \in \mathcal{C}_{\mathrm{n}}} z_{\mathrm{n}}^{c}(j). \tag{6}$$

For each point j, the frozen base classifier yields the highest base-class score $s_{\mathrm{b}}(j)$ and the corresponding label $\hat{c}_{\mathrm{b}}(j)$. Based on the unified base–novel scores, we identify the most competitive base–novel conflict pair for each point, denoted as $\hat{p}(j) = (\hat{c}_{\mathrm{b}}(j), \hat{c}_{\mathrm{n}}(j))$. These scores and pairs are used in the next stage to localize base–novel conflicts and determine which regions require selective calibration.

3.3 Conflict Localization and Selective Calibration

In the selective calibration (SC) module, we first localize ambiguous base–novel conflict points based on the unified base–novel scores. For each point j, we define the base–novel margin as $m(j) = s_{\mathrm{b}}(j) - s_{\mathrm{n}}(j)$, where a small $|m(j)|$ indicates that the base and novel scores are close, implying strong competition between the two classes. We further compute the entropy $H(j)$ over the unified label space to measure prediction uncertainty. Since ambiguous conflict points typically exhibit both a small base–novel margin and high entropy, the conflict score is defined as

$$\phi(j) = \exp(-|m(j)|/\tau_{\mathrm{m}}) \cdot \frac{H(j)}{\log(|\mathcal{C}_{\mathrm{b}}| + |\mathcal{C}_{\mathrm{n}}|)}, \tag{7}$$

where τ_{m} is a temperature parameter. Point-wise calibration can be unreliable because conflict points are

often sparse and scattered. We thus group spatially adjacent points with the same candidate pair into connected components as candidate conflict regions. For each region R_t, we determine its dominant conflict pair p_t by conflict-score-weighted voting and define the pair consistency score as

$$\gamma_t = \frac{\sum_{j\in R_t} \mathbb{I}[\hat{p}(j) = p_t]\phi(j)}{\sum_{j\in R_t} \phi(j)}, \tag{8}$$

where $\mathbb{I}[\cdot]$ is the indicator function. A high γ_t indicates strong support for the dominant pair within R_t, while a low γ_t indicates weak agreement on the dominant pair. We retain only regions that satisfy a minimum size threshold $n_{\min}$ and a pair-consistency threshold τ_{pair}, thereby restricting calibration to coherent and reliable local conflicts.

For each retained conflict region R_t, we aggregate its features into a compact region representation $\mathbf{h}_t$ and feed it to a lightweight calibration head to predict a scalar offset Δ_t for the dominant conflict pair p_t. The offset is applied only to the base and novel classes in p_t within R_t, leaving other classes and non-conflict regions unchanged. Thus, SC performs a local relative-margin correction for the dominant pair rather than a full multi-class update. Let $\tilde{z}_{\text{b}}^{\hat{c}_{\text{b}}^t}(j)$ and $\tilde{z}_{\text{n}}^{\hat{c}_{\text{n}}^t}(j)$ denote the post-calibrated scores of point j for the dominant base and novel classes, and define the signed post-calibration margin as $d_t(j) = \tilde{z}_n^{\hat{c}_n^t}(j) - \tilde{z}_b^{\hat{c}_b^t}(j)$, where positive values favor the dominant novel class. For each R_t, we define g_t to supervise calibration: $g_t = 1$ if the majority of points in R_t belong to the dominant novel class, and $g_t = -1$ otherwise. We then optimize

$$\mathcal{L}_{\text{sep}} = \frac{1}{|\mathcal{R}|} \sum_{R_t\in\mathcal{R}} \frac{1}{|R_t|} \sum_{j\in R_t} \log(1 + \exp(-g_t d_t(j))), \tag{9}$$

$$\mathcal{L}_{\text{reg}} = \frac{1}{|\mathcal{R}|} \sum_{R_t\in\mathcal{R}} \Delta_t^2. \tag{10}$$

Here, $\mathcal{L}_{\text{sep}}$ enforces the desired sign of the local margin, while $\mathcal{L}_{\text{reg}}$ penalizes large offsets to keep the calibration conservative. The overall training objective is

$$\mathcal{L}_{\text{train}} = \mathcal{L}_{\text{seg}} + \lambda_{\text{sep}}\mathcal{L}_{\text{sep}} + \lambda_{\text{reg}}\mathcal{L}_{\text{reg}}. \tag{11}$$

Furthermore, we perform test-time transductive adaptation (TTA) to refine the calibration module for the current episode while preserving reliable base predictions. We freeze the backbone and base classifier, and update only the calibration-related parameters. For this purpose, we define the stable base-anchor set

$$\Omega_{\text{base}} = \big\{ j \;\big|\; \hat{y}_0(j) \in \mathcal{C}_{\text{b}},\ j \notin \Omega_{\text{conf}},\ \max \mathbf{p}_0^{\text{b}}(j) > \tau_{\text{b}} \big\}, \tag{12}$$

where $\hat{y}_0(j)$ is the initial prediction before refinement, $\mathbf{p}_0^{\text{b}}(j)$ is the corresponding posterior over the base-class subspace, Ω_{conf} is the conflict point set, and τ_{b} is a base confidence threshold. By construction, Ω_{base} contains only confident base predictions outside detected conflict regions, and therefore serves as a stable anchor set that is less affected by base–novel ambiguity. Let $\mathbf{p}^{\text{b}}(j)$ denote the updated base-class posterior. We minimize $\mathcal{L}_{\text{tta}}$ only on these anchors to preserve reliable non-conflict base responses while adapting the calibration module to the current episode:

$$\mathcal{L}_{\text{tta}} = \frac{1}{|\Omega_{\text{base}}|} \sum_{j\in\Omega_{\text{base}}} \text{KL}\big(\mathbf{p}_0^{\text{b}}(j) \,\|\, \mathbf{p}^{\text{b}}(j)\big). \tag{13}$$

3.4 Optional Textual Prototype Augmentation

To complement support-derived prototypes for novel classes, we introduce an optional textual prototype augmentation (TPA) module. TPA provides an additional class-level prior for each novel class without modifying the support-derived prototypes. For each novel class c, we use a frozen CLIP text encoder [23] to encode a small set of fixed prompts derived from its canonical class name. The resulting text features are mapped into the point feature space with a lightweight adapter and then averaged to obtain a textual prototype $\mathbf{p}_{\text{t}}^c$. Since textual priors are not equally reliable across classes and episodes, we estimate a reliability score for each class:

$$r_{\text{t}}^c = \sigma\big(\gamma_1 u_{\text{pc}}^c + \gamma_2 u_{\text{sa}}^c - \gamma_3 u_{\text{ba}}^c\big), \tag{14}$$

where u_{pc}^c, u_{sa}^c, and u_{ba}^c denote prompt consistency, support agreement, and base ambiguity, indicating whether different prompts yield similar embeddings, whether the textual prototype agrees with the support-derived prototype, and whether it lies too close to base-class prototypes, respectively. For each retained region, we feed the textual prototype of its dominant novel class, together with its reliability score and a region-level text prior score, into the pairwise calibration head. As a result, textual information is incorporated only in conflict region calibration, rather than uniformly across the scene. This design preserves the support-driven nature of the framework and avoids indiscriminate fusion of textual cues. More details are provided in the Online Resource 1.

4 Experiments

4.1 Implementation Details

Experiments are conducted on four indoor benchmarks: S3DIS, ScanNet, ScanNet++, and ScanNet200 [3,7,36,

Table 1 Generalized few-shot point cloud segmentation results on **ScanNet**.

Prior Type	Method	5-shot				1-shot			
		mIoU-B	mIoU-N	mIoU-A	HM	mIoU-B	mIoU-N	mIoU-A	HM
–	Fully Sup.	78.80	61.22	73.25	68.91	78.80	61.22	73.25	68.91
Support only	attMPTI [38]	16.31	3.12	12.35	5.21	12.97	1.62	9.57	2.88
	PIFS [4]	35.14	3.21	25.56	5.88	35.80	2.54	25.82	4.75
	GW [34]	40.18	18.58	33.70	25.39	40.06	14.78	32.47	21.55
	PE [28]	40.42	23.34	35.30	29.55	40.47	15.57	33.00	22.47
	Ours-S	**69.49**$_{\pm 0.06}$	**41.94**$_{\pm 0.07}$	**60.79**$_{\pm 0.02}$	**52.31**$_{\pm 0.04}$	**67.34**$_{\pm 0.11}$	**35.35**$_{\pm 0.08}$	**57.24**$_{\pm 0.07}$	**46.37**$_{\pm 0.06}$
External prior	LARM [16]	42.37	28.61	38.21	34.14	42.71	21.50	36.43	28.58
	GFS-VL [1]	78.30	51.22	69.75	61.91	78.56	49.72	69.45	60.88
	Ours-ST	**78.54**$_{\pm 0.08}$	**52.17**$_{\pm 0.08}$	**70.21**$_{\pm 0.03}$	**62.69**$_{\pm 0.04}$	**78.66**$_{\pm 0.20}$	**50.62**$_{\pm 0.22}$	**69.80**$_{\pm 0.07}$	**61.60**$_{\pm 0.11}$

Table 2 Generalized few-shot point cloud segmentation results on **S3DIS**.

Prior Type	Method	5-shot				1-shot			
		mIoU-B	mIoU-N	mIoU-A	HM	mIoU-B	mIoU-N	mIoU-A	HM
–	Fully Sup.	78.62	63.50	73.57	70.26	78.62	63.50	73.57	70.26
Support only	attMPTI [38]	34.90	16.08	26.21	21.99	21.89	11.39	17.05	14.95
	PIFS [4]	56.99	19.66	39.76	29.23	57.85	14.59	37.88	23.31
	GW [34]	73.61	43.26	59.60	54.42	74.10	29.66	53.58	41.92
	PE [28]	74.77	50.23	63.44	60.06	74.54	39.78	58.50	51.34
	Ours-S	**75.81**$_{\pm 0.17}$	**54.29**$_{\pm 0.36}$	**65.88**$_{\pm 0.19}$	**63.27**$_{\pm 0.25}$	**75.75**$_{\pm 0.09}$	**46.95**$_{\pm 0.10}$	**62.46**$_{\pm 0.04}$	**57.97**$_{\pm 0.07}$
External prior	LARM [16]	75.64	54.08	65.68	63.06	75.68	44.52	61.29	56.04
	Ours-ST	**76.09**$_{\pm 0.12}$	**57.39**$_{\pm 0.16}$	**67.92**$_{\pm 0.02}$	**66.07**$_{\pm 0.06}$	**75.76**$_{\pm 0.12}$	**49.11**$_{\pm 0.24}$	**63.46**$_{\pm 0.17}$	**59.59**$_{\pm 0.21}$

Table 3 Generalized few-shot point cloud segmentation results on **ScanNet++**.

Prior Type	Method	5-shot				1-shot			
		mIoU-B	mIoU-N	mIoU-A	HM	mIoU-B	mIoU-N	mIoU-A	HM
–	Fully Sup.	64.89	37.48	49.04	48.31	64.89	37.48	49.04	48.31
Support only	PIFS [4]	39.98	5.74	19.44	10.03	36.66	4.95	17.63	8.71
	attMPTI [38]	55.89	4.17	24.87	7.78	53.16	3.55	23.40	6.66
	COSeg [2]	59.34	6.96	27.91	12.45	58.49	6.24	27.14	11.26
	GW [34]	51.35	11.03	27.16	18.15	46.71	6.63	22.66	11.59
	Ours-S	**59.51**$_{\pm 0.29}$	**18.67**$_{\pm 0.53}$	**35.01**$_{\pm 0.21}$	**28.42**$_{\pm 0.58}$	**58.50**$_{\pm 0.13}$	**13.91**$_{\pm 0.37}$	**31.75**$_{\pm 0.26}$	**22.48**$_{\pm 0.49}$
External prior	GFS-VL [1]	60.05	21.66	37.02	31.82	61.39	19.42	36.21	29.47
	Ours-ST	**61.27**$_{\pm 0.10}$	**24.45**$_{\pm 0.18}$	**39.18**$_{\pm 0.07}$	**34.95**$_{\pm 0.17}$	**62.38**$_{\pm 0.12}$	**21.77**$_{\pm 0.34}$	**38.01**$_{\pm 0.18}$	**32.27**$_{\pm 0.36}$

Table 4 Generalized few-shot point cloud segmentation results on **ScanNet200**.

Prior Type	Method	5-shot				1-shot			
		mIoU-B	mIoU-N	mIoU-A	HM	mIoU-B	mIoU-N	mIoU-A	HM
–	Fully Sup.	68.87	39.25	45.49	50.00	68.87	39.25	45.49	50.00
Support only	PIFS [4]	28.78	3.82	9.07	6.71	17.84	2.87	6.02	7.88
	attMPTI [38]	37.13	4.99	11.76	8.79	54.84	3.27	14.14	6.17
	COSeg [2]	57.67	5.21	16.25	9.54	47.03	4.03	13.09	7.42
	GW [34]	59.28	8.30	19.03	14.55	55.23	6.47	16.74	11.56
	Ours-S	**62.29**$_{\pm 0.24}$	**19.95**$_{\pm 0.19}$	**28.86**$_{\pm 0.10}$	**30.22**$_{\pm 0.18}$	**61.39**$_{\pm 0.14}$	**16.53**$_{\pm 0.06}$	**25.97**$_{\pm 0.03}$	**26.04**$_{\pm 0.06}$
External prior	GFS-VL [1]	67.57	31.67	39.23	43.12	68.48	29.18	37.45	40.92
	Ours-ST	**68.26**$_{\pm 0.10}$	**33.44**$_{\pm 0.21}$	**40.77**$_{\pm 0.19}$	**44.89**$_{\pm 0.21}$	**68.43**$_{\pm 0.08}$	**31.37**$_{\pm 0.15}$	**39.17**$_{\pm 0.12}$	**43.02**$_{\pm 0.14}$

25]. Following GFS-VL [1], each dataset is split into disjoint base and novel class sets, resulting in 7/6, 13/6, 12/18, and 12/45 base/novel classes for S3DIS, ScanNet, ScanNet++, and ScanNet200, respectively.

We chose PTV3 [32] as our default backbone for all main experiments. Ours-S denotes the support-only version of the proposed framework, while Ours-ST further incorporates the optional textual prototype augmentation module. Unless otherwise specified, we use $M = 4$ local slots, $\lambda_{\text{sep}} = 0.3$, $\lambda_{\text{reg}} = 0.05$, $\tau_{\text{b}} = 0.9$, $\tau_{\text{pair}} = 0.75$, and $n_{\min} = 12$. For conflict-region construction, the neighborhood graph is built with k-nearest neighbors using $k = 16$. During test-time transductive adaptation (TTA), only the calibration-related parameters are updated for 5 steps, while the backbone and base classifier remain frozen. All experiments are implemented on two NVIDIA L40S GPUs. More details are provided in the Online Resource 1.

4.2 Main Results

Tab. 1 – Tab. 4 report results on ScanNet, S3DIS, ScanNet++, and ScanNet200. For Ours-S and Ours-ST, we run each experiment with three independent random seeds and report the mean and standard deviation to account for run-to-run variation. Among support-only methods, Ours-S achieves the best performance under

Table 5 Core framework ablation. PRE, BNS, SC, and TTA denote reliability estimation, base–novel scoring, selective calibration, and test-time transductive adaptation.

PRE	BNS	SC	TTA	mIoU-B	mIoU-N	mIoU-A	HM
				55.8	8.9	27.7	15.4
✓				57.2	12.7	30.5	20.8
✓	✓			58.0	14.8	32.1	23.6
✓	✓	✓		59.1	17.7	34.3	27.2
✓	✓	✓	✓	**59.8**	**18.2**	**34.8**	**27.9**

Table 6 Ablation of prototype reliability estimation (PRE) on ScanNet++ (5-shot). GP, LP, and Rel. denote global prototype, local prototype, and reliability, respectively.

GP	LP	Rel.	mIoU-B	mIoU-N	mIoU-A	HM
✓			58.3	15.2	33.0	24.2
✓	✓		59.0	16.8	34.0	26.0
✓	✓	✓	**59.8**	**18.2**	**34.8**	**27.9**

both 1-shot and 5-shot settings on all datasets, showing that our framework already performs strongly without any external prior. The strong HM results further indicate that explicit base–novel conflict localization and selective calibration improve performance by enhancing novel-class recognition while maintaining a favorable balance with base classes.

Building on this strong support-only baseline, the TPA module brings further gains and yields the best results among methods using external priors across all benchmarks and shot settings. Ours-ST benefits from text as a semantic complement to support-derived prototypes, further improving novel-class prediction while keeping base-class performance largely stable.

4.3 Ablation Study

We evaluate the core framework on ScanNet++ under the 5-shot setting. Performance improves steadily as PRE, BNS, SC, and TTA are introduced, indicating that these components are complementary rather than redundant, as shown in Tab. 5. SC brings the largest gain, indicating that explicit calibration of ambiguous base–novel conflicts is the key factor in improving generalized segmentation.

We further analyze the ablation of prototype reliability estimation (PRE) on ScanNet++ (5-shot). As shown in Tab. 6, adding local prototypes already improves performance over using only a global prototype, showing the benefit of capturing finer local geometric structure. Reliability estimation provides an additional gain, suggesting that richer prototypes are most effective when their reliability is explicitly taken into account.

Tab. 7 studies the design of optional textual prototype augmentation (TPA). Applying text to all regions improves mIoU-N but slightly degrades mIoU-B, suggesting that global text fusion can interfere with already reliable base predictions. Restricting text to conflict regions is substantially more effective, and adding reliability yields a further improvement. This confirms that textual priors are most useful when introduced selectively for ambiguous regions, rather than uniformly across the scene.

Table 7 Ablation of the optional TPA design on ScanNet++ (5-shot). Region specifies where text guidance is applied, where all and conf. denote all regions and conflict regions, respectively. Rel. denotes reliability weighting.

Setting	Region	Rel.	mIoU-B	mIoU-N	mIoU-A	HM
Ours-S	–	–	59.8	18.2	34.8	27.9
Ours-S+text	all	–	58.9	20.1	35.6	30.0
Ours-S+text	conf.	–	60.6	22.7	37.9	33.0
Ours-ST	conf.	✓	**61.3**	**24.4**	**39.1**	**34.9**

Table 8 Results with Different Backbones.

Backbone	mIoU-B	mIoU-N	mIoU-A	HM
PTV3 [32] + Ours-S	59.72	18.17	34.79	27.96
PTV3 + Ours-ST	61.26	24.37	39.13	34.87
MinkowskiNet [6] + Ours-S	56.34	15.78	32.00	24.65
MinkowskiNet + Ours-ST	58.23	20.92	35.84	30.78

Table 9 Inference time and peak GPU memory cost on the ScanNet200 1-shot validation set.

Method	Inference time/scene	Total eval. time	Memory
GFS-VL	665.6 ± 4.3 ms	207.68 ± 1.36 s	4.50 GiB
Ours-S	792.8 ± 7.8 ms	247.10 ± 2.43 s	4.82 GiB
Ours-ST	832.3 ± 9.6 ms	259.68 ± 2.99 s	4.93 GiB

We further evaluate our method with different backbones in Tab. 8. Although the absolute performance varies with backbone strength, the same improvement trend is preserved. Moreover, the gains are reflected not only in mIoU-N, but also in mIoU-A and HM, suggesting that the improvement comes from better base–novel trade-off rather than simply increasing novel-class predictions. These results in Tab. 8 demonstrate that our framework is compatible with different backbone choices. In particular, Ours-ST consistently improves over Ours-S on both backbones, indicating that the benefit of optional textual prototype augmentation is not tied to a specific backbone architecture.

4.4 Computational Analysis

To further analyze the computational overhead of the proposed method, we report the inference time and peak GPU memory consumption on the ScanNet200 1-shot validation split in Tab. 9. We compare Ours-S and Ours-ST with GFS-VL under the same evaluation setting. All experiments are conducted on a single NVIDIA L40S GPU. We define inference time as the end-to-end evaluation wall-clock time, including model load-

Table 10 Hyperparameter sensitivity analysis on ScanNet++ under the 5-shot setting. The default values used in the main paper are highlighted in bold.

(a) Effect of local slots M

M	mIoU-B	mIoU-N	mIoU-A	HM
2	59.0	16.9	33.7	26.3
4	**59.8**	**18.2**	**34.8**	**27.9**
6	59.4	18.2	34.7	27.7
8	58.9	16.8	33.6	27.1

(b) Effect of τ_b

τ_b	mIoU-B	mIoU-N	mIoU-A	HM
0.7	59.3	16.7	33.7	26.1
0.8	59.6	17.6	34.4	27.2
0.9	**59.8**	**18.2**	**34.8**	**27.9**
0.95	59.2	17.9	34.4	27.5

(c) Effect of λ_{sep}

λ_{sep}	mIoU-B	mIoU-N	mIoU-A	HM
0.1	59.2	17.4	34.1	26.9
0.2	59.5	17.7	34.4	27.3
0.3	**59.8**	**18.2**	**34.8**	**27.9**
0.5	59.1	17.8	34.3	27.4

(d) Effect of λ_{reg}

λ_{reg}	mIoU-B	mIoU-N	mIoU-A	HM
0.01	59.5	17.7	34.4	27.3
0.03	59.7	17.8	34.6	27.4
0.05	**59.8**	**18.2**	**34.8**	**27.9**
0.10	59.3	17.6	34.3	27.1

(e) Effect of τ_{pair}

τ_{pair}	mIoU-B	mIoU-N	mIoU-A	HM
0.60	59.0	17.7	34.2	27.2
0.70	59.3	18.0	34.5	27.6
0.75	**59.8**	**18.2**	**34.8**	**27.9**
0.80	59.5	17.8	34.5	27.4
0.90	59.9	17.5	34.5	27.1

(f) Effect of $n_{\min}$

$n_{\min}$	mIoU-B	mIoU-N	mIoU-A	HM
6	58.9	17.5	34.1	27.0
9	59.4	17.7	34.4	27.3
12	**59.8**	**18.2**	**34.8**	**27.9**
16	59.7	18.0	34.7	27.7
20	59.3	17.8	34.4	27.4

ing, data loading, forward inference, post-processing, and metric computation. For our method, we report the mean and standard deviation over checkpoints trained with three different random seeds. For GFS-VL, we report the mean and standard deviation over the five official registration-set checkpoints.

The computational results show that our method improves generalized few-shot point cloud segmentation with a reasonable and controlled overhead compared with GFS-VL. Both variants complete end-to-end evaluation within one second per scene and keep the peak GPU memory below 5 GiB, indicating that the proposed framework maintains practical computational efficiency. In particular, the calibration head only predicts a scalar offset for the dominant base–novel pair within each retained conflict region. During test-time adaptation, only the calibration-related parameters are updated, while the backbone and base classifier remain frozen. As a result, the proposed calibration mechanism introduces only limited additional memory consumption.

For Ours-ST, textual prototype augmentation introduces only a small extra cost over Ours-S. The per-scene time increases by 39.5 ms, and the peak memory increases by 0.11 GiB. This is because textual prototypes are class-level semantic priors and are incorporated in a conflict-conditioned manner, instead of being densely fused into all query points. Moreover, the textual branch does not require an additional dense vision-language forward pass for each scene during evaluation.

4.5 Hyperparameter Sensitivity Analysis

We further analyze the sensitivity of several key hyperparameters used in the proposed framework on ScanNet++ under the 5-shot setting. As shown in Tab. 10, we study several key hyperparameters, including the number of local slots M, the stable base-anchor threshold τ_{b}, the separation loss weight λ_{sep}, the regularization weight λ_{reg}, the pair-consistency threshold τ_{pair}, and the minimum region size $n_{\min}$. In our method, M controls how support patch tokens are aggregated into local prototypes, τ_{b} determines which base predictions are treated as stable anchors during refinement, and λ_{sep} and λ_{reg} balance the pairwise separation and conservative regularization terms, respectively. As shown in Tab. 10, the method remains stable within a reasonable range of settings, and the best performance is consistently achieved near the final configuration used in the main paper, i.e., $M = 4$, $\tau_{\text{b}} = 0.9$, $\lambda_{\text{sep}} = 0.3$, and $\lambda_{\text{reg}} = 0.05$.

As shown in Tab. 10(a), the performance first improves and then declines as M increases, and the best result is achieved at $M = 4$. When M is too small, the local prototype bank cannot adequately capture diverse part-level structures, which weakens novel-class discrimination. In contrast, using too many slots tends to fragment support information into overly fine local patterns, reducing prototype stability and slightly degrading both mIoU-N and HM. Therefore, $M = 4$ provides the best balance between representational diversity and reliability.

Tab. 10(b) shows the effect of τ_{b}, which determines which base predictions are selected as stable anchors during test-time refinement. A lower threshold introduces less reliable base anchors into transductive refinement, which weakens the base-preserving constraint and leads to less stable generalized adaptation. On the other hand, an overly strict threshold such as 0.95 reduces anchor coverage and limits the effectiveness of refinement.

Tab. 10(c) and (d) analyze the two loss weights in region-based conflict calibration. With λ_{reg} fixed, performance peaks at $\lambda_{\text{sep}} = 0.3$. Smaller values do not

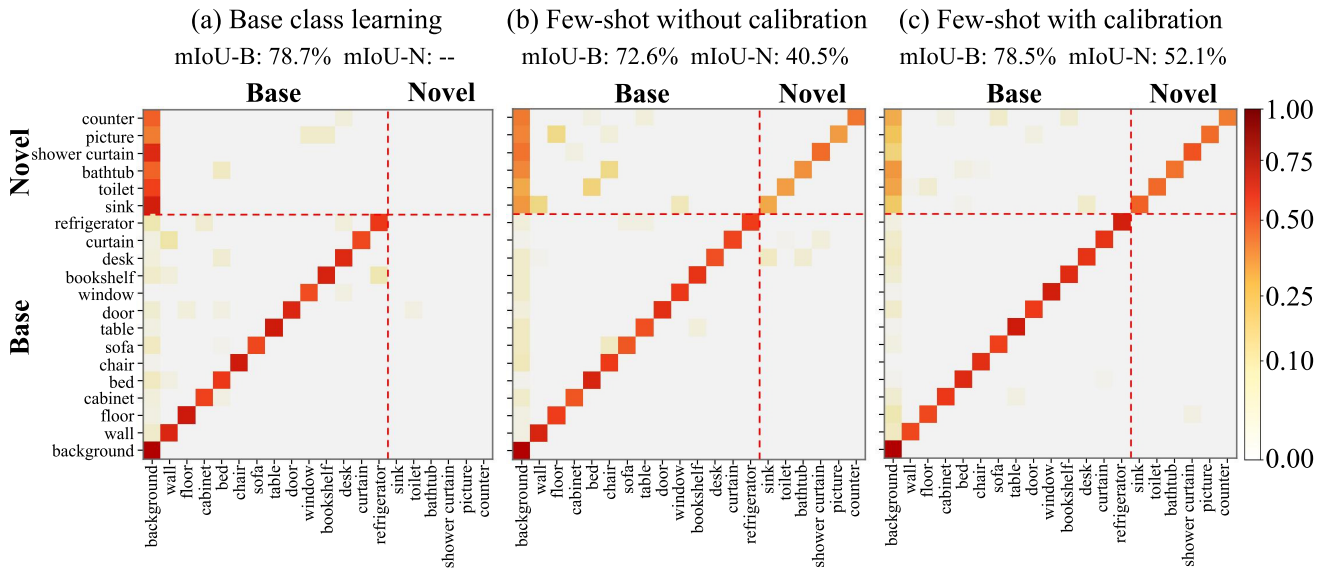


Fig. 3 Visualization of class-level confusion matrices on ScanNet (5-shot)

provide sufficient supervision for separating ambiguous base–novel conflict pairs, while a larger value overemphasizes local separation and slightly harms the global base–novel balance. Similarly, with λ_{sep} fixed, the best result is achieved at $\lambda_{\text{reg}} = 0.05$. If λ_{reg} is too small, the calibration module becomes less conservative; if it is too large, the predicted offsets are overly suppressed, weakening the corrective effect in true conflict regions. In general, these results confirm that the default setting $M = 4$, $\tau_{\text{b}} = 0.9$, $\lambda_{\text{sep}} = 0.3$, and $\lambda_{\text{reg}} = 0.05$ achieves the best overall trade-off.

Tab. 10(e) analyzes the effect of the pair-consistency threshold τ_{pair}. A smaller threshold retains more candidate regions but may introduce less coherent base–novel conflicts, while a larger threshold is more conservative and may discard useful conflict regions. The performance is stable around the default setting, and $\tau_{\text{pair}} = 0.75$ gives the best results, achieving 34.8 mIoU-A and 27.9 HM. Therefore, we use $\tau_{\text{pair}} = 0.75$ in all experiments.

Tab. 10(f) analyzes the effect of the minimum region size n_{min}. A smaller value may keep noisy small components, whereas a larger value may remove valid but fragmented conflict regions. The best performance is obtained when $n_{\text{min}} = 12$, with 34.8 mIoU-A and 27.9 HM. This setting provides a good balance between suppressing noisy conflict proposals and preserving meaningful local conflicts.

4.6 Visualization

We further provide visual analysis to better understand the effect and limitations of the proposed calibration framework. These visualizations provide complementary evidence from three perspectives: whether calibration improves base–novel discrimination at the class level, whether it corrects localized conflicts at the scene level, and where the current framework may still fail.

Fig. 3 compares the confusion matrices before and after calibration. Before calibration, ambiguous points near base–novel decision boundaries are often assigned to visually similar competing categories, producing evident off-diagonal entries between base and novel classes. After calibration, the diagonal entries become clearer for both base and novel classes, while base–novel confusion is reduced. This shows that calibration does not simply increase novel-class confidence; instead, it improves the relative decision boundary between competing base and novel classes. The clearer diagonal structure supports our motivation that generalized few-shot point cloud segmentation benefits from explicitly resolving base–novel conflicts rather than uniformly enhancing novel representations.

Fig. 4 further shows scene-level segmentation results on ScanNet200. Before calibration, errors are concentrated in local ambiguous regions where base and novel classes have similar responses, especially around object boundaries, small objects, and cluttered areas. After calibration, many of these regions are corrected, while large stable regions such as walls, floors, and confident base-class areas remain largely unchanged. This behavior matches the design of selective calibration: the model localizes conflict regions and adjusts only the dominant base–novel pair within them. Thus, the qualitative results suggest that the performance gain mainly

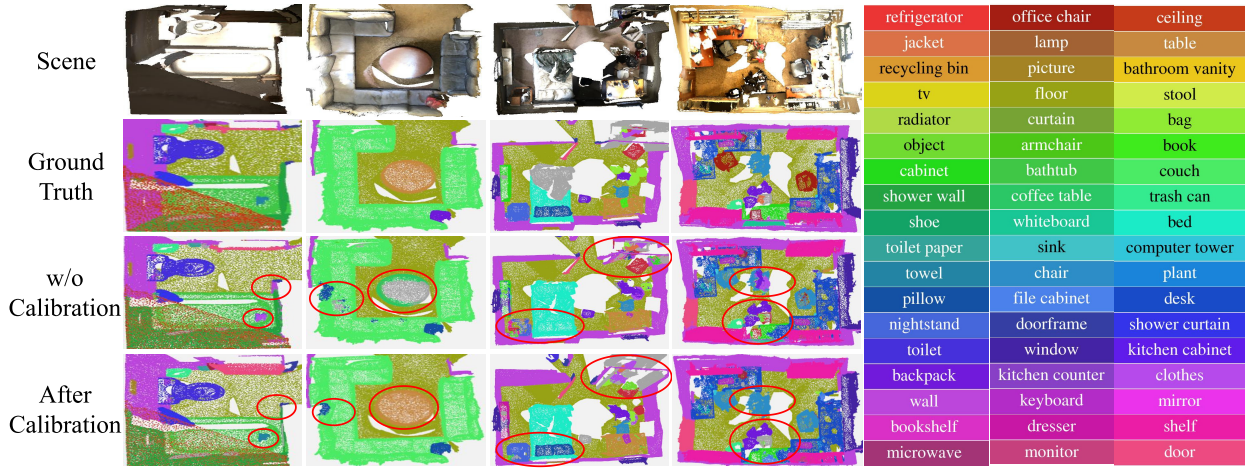


Fig. 4 Qualitative comparison on ScanNet200 before and after calibration

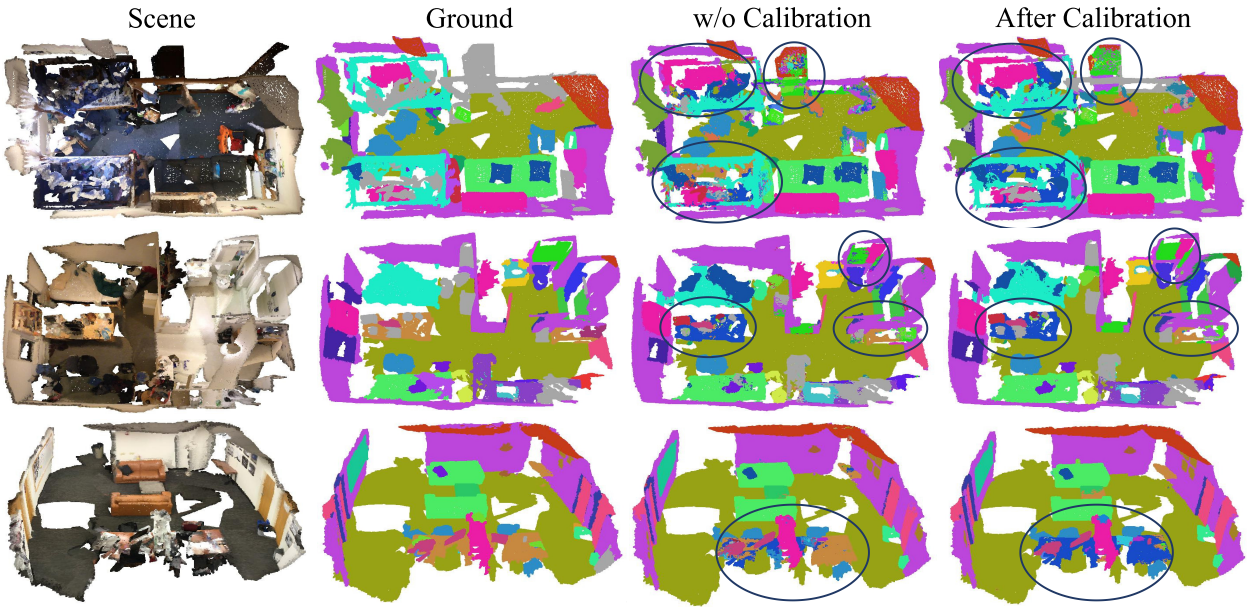


Fig. 5 Representative failure cases

comes from local conflict correction rather than globally reshaping the prediction map.

Fig. 5 visualizes representative failure cases, where the highlighted regions indicate remaining errors after calibration. These cases are consistent with the limitations discussed in the Online Resource 1. In the first row, cluttered structures and visually similar categories lead to errors, suggesting that single dominant-pair correction may be insufficient when multiple fine-grained classes compete in a compact area. In the second row, some ambiguous regions are fragmented and only partially corrected. This indicates that the final prediction depends on the quality of conflict localization and retained region construction; if the true conflict area is only partially covered, selective calibration cannot fully correct all erroneous points. In the third row, errors remain around sparse and incomplete point cloud structures, where missing observations and weak local geometry make the conflict regions less stable. These examples show that our method is most effective when ambiguous regions can be localized as coherent pair-dominant conflicts, but it may still struggle under fine-grained ambiguity, sparse observations, or fragmented conflict regions.

Overall, the visualization results support the main design principle of our framework. Calibration improves generalized segmentation by selectively correcting localized base–novel conflicts while preserving stable predictions elsewhere. At the same time, the failure cases reveal that more robust conflict localization and finer-

grained, possibly multi-pair, calibration are promising directions for future work.

5 Conclusion

We proposed a reliability-aware conflict localization and calibration framework for generalized few-shot point cloud segmentation. The key idea is to treat generalized few-shot segmentation not only as a problem of learning better novel-class prototypes, but also as a problem of making reliable decisions in local base–novel conflict regions. To this end, our method estimates the reliability of hierarchical support prototypes, identifies coherent base–novel conflicts, and calibrates only the ambiguous regions while keeping confident predictions unchanged. We further introduce an optional textual prototype augmentation module, where textual priors are incorporated as a semantic complement in a conflict-conditioned and reliability-aware manner. The results show that the major gains of our method come from explicit conflict modeling in the support-only setting, where Ours-S consistently and substantially improves over previous support-driven methods across datasets and shot settings. The ablation study further confirms that selective calibration is the most influential component of the core framework, highlighting the importance of locally resolving base–novel conflicts. With textual augmentation, Ours-ST brings more moderate but consistent improvements over strong vision-language baselines, with larger gains on more challenging benchmarks such as ScanNet++ and ScanNet200, where base–novel trade-offs are harder. These results suggest that the proposed framework is broadly applicable to GFS-PCS and provides an interpretable way to improve generalized segmentation by focusing on uncertain conflict regions. Future work may extend it toward multi-pair calibration, stronger support–text alignment, and more robust conflict localization under complex multi-class ambiguity and fine-grained novel categories.

Compliance with Ethical Standards

Funding: This work was partially supported by NSF of China (No. 62425209) and SenseCore.
Conflict of interest: The authors declare that they have no conflict of interest.

Data Availability

No new dataset was generated or collected in this study. The experiments were conducted on publicly available datasets, including S3DIS, ScanNet, ScanNet++, and ScanNet200. These datasets are available from their official project pages subject to their respective licenses and terms of use. The dataset splits and evaluation protocols follow the settings described in the cited benchmark and baseline papers. Additional experimental details are provided in Online Resource 1.

References

1. An, Z., Sun, G., Liu, Y., Li, R., Han, J., Konukoglu, E., Belongie, S.: Generalized few-shot 3D point cloud segmentation with vision-language model. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 16,997–17,007 (2025)
2. An, Z., Sun, G., Liu, Y., Liu, F., Wu, Z., Wang, D., Van Gool, L., Belongie, S.: Rethinking few-shot 3D point cloud semantic segmentation. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 3996–4006 (2024)
3. Armeni, I., Sener, O., Zamir, A.R., Jiang, H., Brilakis, I., Fischer, M., Savarese, S.: 3D semantic parsing of large-scale indoor spaces. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 1534–1543 (2016)
4. Cermelli, F., Mancini, M., Xian, Y., Akata, Z., Caputo, B.: Prototype-based incremental few-shot semantic segmentation. In: Proc. Br. Mach. Vis. Conf. (2021)
5. Chen, X., Shi, M., Zhou, Z., He, L., Tsoka, S.: Enhancing generalized few-shot semantic segmentation via effective knowledge transfer. In: Proc. AAAI Conf. Artif. Intell., vol. 39, pp. 2256–2265 (2025)
6. Choy, C., Gwak, J., Savarese, S.: 4D spatio-temporal convnets: Minkowski convolutional neural networks. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 3075–3084 (2019)
7. Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T., Nießner, M.: ScanNet: Richly-annotated 3D reconstructions of indoor scenes. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 5828–5839 (2017)
8. Hajimiri, S., Boudiaf, M., Ben Ayed, I., Dolz, J.: A strong baseline for generalized few-shot semantic segmentation. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 11,269–11,278 (2023)
9. He, S., Jiang, X., Jiang, W., Ding, H.: Prototype adaption and projection for few- and zero-shot 3D point cloud semantic segmentation. IEEE Trans. Image Process. **32**, 3199–3211 (2023). DOI https://doi.org/10.1109/TIP.2023.3279660
10. Hossain, M.R.I., Siam, M., Sigal, L., Little, J.J.: Visual prompting for generalized few-shot segmentation: A multi-scale approach. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 23,470–23,480 (2024)
11. Huang, K., Wang, F., Xi, Y., Gao, Y.: Prototypical kernel learning and open-set foreground perception for generalized few-shot semantic segmentation. In: Proc. IEEE/CVF Int. Conf. Comput. Vis., pp. 19,256–19,265 (2023)
12. Jia, Y., Li, J., Wang, Q.: Generalized few-shot semantic segmentation for remote sensing images. IEEE Trans. Geosci. Remote Sens. **63**, 1–10 (2025). DOI https://doi.org/10.1109/TGRS.2025.3531874
13. Lai, X., Liu, J., Jiang, L., Wang, L., Zhao, H., Liu, S., Qi, X., Jia, J.: Stratified transformer for 3D point cloud segmentation. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 8500–8509 (2022)

14. Li, G., Jampani, V., Sevilla-Lara, L., Sun, D., Kim, J., Kim, J.: Adaptive prototype learning and allocation for few-shot segmentation. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 8334–8343 (2021)
15. Li, Z., Lu, F., Zou, J., Hu, L., Zhang, H.: Generalized few-shot meets remote sensing: Discovering novel classes in land cover mapping via hybrid semantic segmentation framework. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, pp. 2744–2754 (2024)
16. Li, Z., Wang, Y., Xiong, G., Li, W., Pan, Y., Zhang, T.: Generalized few-shot point cloud segmentation via LLM-assisted hyper-relation matching. In: Proc. IEEE/CVF Int. Conf. Comput. Vis., pp. 23,063–23,073 (2025)
17. Liu, S.A., Zhang, Y., Qiu, Z., Xie, H., Zhang, Y., Yao, T.: Learning orthogonal prototypes for generalized few-shot semantic segmentation. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 11,319–11,328 (2023)
18. Liu, W., Wu, Z., Zhao, Y., Fang, Y., Foo, C.S., Cheng, J., Lin, G.: Harmonizing base and novel classes: A class-contrastive approach for generalized few-shot segmentation. Int. J. Comput. Vis. **132**(4), 1277–1291 (2024). DOI https://doi.org/10.1007/s11263-023-01939-y
19. Lu, Z., He, S., Li, D., Song, Y.Z., Xiang, T.: Prediction calibration for generalized few-shot semantic segmentation. IEEE Trans. Image Process. **32**, 3311–3323 (2023). DOI https://doi.org/10.1109/TIP.2023.3282070
20. Lüddecke, T., Ecker, A.: Image segmentation using text and image prompts. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 7086–7096 (2022)
21. Min, J., Kang, D., Cho, M.: Hypercorrelation squeeze for few-shot segmentation. In: Proc. IEEE/CVF Int. Conf. Comput. Vis., pp. 6941–6952 (2021)
22. Qian, G., Li, Y., Peng, H., Mai, J., Hammoud, H., Elhoseiny, M., Ghanem, B.: Pointnext: Revisiting pointnet++ with improved training and scaling strategies. Adv. Neural Inf. Process. Syst. **35**, 23,192–23,204 (2022)
23. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: Proc. Int. Conf. Mach. Learn., pp. 8748–8763 (2021)
24. Ren, W., Tang, Y., Sun, Q., Zhao, C., Han, Q.L.: Visual semantic segmentation based on few/zero-shot learning: An overview. IEEE/CAA J. Autom. Sinica **11**(5), 1106–1126 (2024). DOI https://doi.org/10.1109/JAS.2023.123207
25. Rozenberszki, D., Litany, O., Dai, A.: Language-grounded indoor 3D semantic segmentation in the wild. In: Eur. Conf. Comput. Vis., pp. 125–141 (2022)
26. Sakai, T., Qiu, H., Katsuki, T., Kimura, D., Osogami, T., Inoue, T.: A surprisingly simple approach to generalized few-shot semantic segmentation. Adv. Neural Inf. Process. Syst. **37**, 27,005–27,023 (2024)
27. Tian, Z., Lai, X., Jiang, L., Liu, S., Shu, M., Zhao, H., Jia, J.: Generalized few-shot semantic segmentation. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 11,563–11,572 (2022)
28. Tsai, C.J., Chen, H.T., Liu, T.L.: Pseudo-embedding for generalized few-shot 3D segmentation. In: Eur. Conf. Comput. Vis., pp. 383–400 (2024)
29. Wang, K., Liew, J.H., Zou, Y., Zhou, D., Feng, J.: Panet: Few-shot image semantic segmentation with prototype alignment. In: Proc. IEEE/CVF Int. Conf. Comput. Vis., pp. 9197–9206 (2019)
30. Wei, L., Lang, C., Chen, Z., Wang, T., Li, Y., Liu, J.: Generated and pseudo content guided prototype refinement for few-shot point cloud segmentation. Adv. Neural Inf. Process. Syst. **37**, 31,103–31,123 (2024)
31. Wu, P., Mei, J., Zhao, X., Hu, Y.: Generalized few-shot semantic segmentation for LiDAR point clouds. In: Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), pp. 7622–7628 (2023). DOI https://doi.org/10.1109/IROS55552.2023.10341978
32. Wu, X., Jiang, L., Wang, P.S., Liu, Z., Liu, X., Qiao, Y., Ouyang, W., He, T., Zhao, H.: Point transformer v3: Simpler faster stronger. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 4840–4851 (2024)
33. Wu, X., Lao, Y., Jiang, L., Liu, X., Zhao, H.: Point transformer v2: Grouped vector attention and partition-based pooling. Adv. Neural Inf. Process. Syst. **35**, 33,330–33,342 (2022)
34. Xu, Y., Hu, C., Zhao, N., Lee, G.H.: Generalized few-shot point cloud segmentation via geometric words. In: Proc. IEEE/CVF Int. Conf. Comput. Vis., pp. 21,506–21,515 (2023)
35. Yang, S., Ding, H., Jiang, X.: Generalized few-shot 3D point cloud segmentation. In: IEEE Int. Symp. Circuits Syst., pp. 1–5 (2024)
36. Yeshwanth, C., Liu, Y.C., Nießner, M., Dai, A.: ScanNet++: A high-fidelity dataset of 3D indoor scenes. In: Proc. IEEE/CVF Int. Conf. Comput. Vis., pp. 12–22 (2023)
37. Zhang, C., Wu, Z., Wu, X., Zhao, Z., Wang, S.: Few-shot 3D point cloud semantic segmentation via stratified class-specific attention based transformer network. In: Proc. AAAI Conf. Artif. Intell., vol. 37, pp. 3410–3417 (2023)
38. Zhao, N., Chua, T.S., Lee, G.H.: Few-shot 3D point cloud semantic segmentation. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 8873–8882 (2021)
39. Zheng, C., Liu, L., Meng, Y., Peng, X., Wang, M.: Few-shot point cloud semantic segmentation via support-query feature interaction. IEEE Trans. Circuits Syst. Video Technol. **34**(11), 10,753–10,763 (2024). DOI https://doi.org/10.1109/TCSVT.2024.3417233



He Wang is currently a Ph.D. student at Zhejiang University. She received the B.S. and M.S. degrees from the College of Computer Science and Technology, Harbin Engineering University, in 2016 and 2019, respectively. Her research interests include computer vision, 3D scene understanding, point cloud segmentation, and augmented reality, with applications in spatial computing.



Guofeng Zhang is currently a professor at Zhejiang University. He received the B.S. and Ph.D. degrees in computer science and technology from Zhejiang University in 2003 and 2009, respectively. He received the National Excellent Doctoral Dissertation Award, the Excellent Doctoral Dissertation Award of China Computer Federation and the ISMAR 2020 Best Paper Award. His research interests include 3D vision, augmented reality and spatial intelligence.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplymentarymaterials.pdf