Don't Starve the Boundaries: Boundary-Constrained Label Propagation for Weakly Supervised 3D Segmentation

Shuwei Wu[1], Shuo Jin[1,2], Zhijin He[1], Siyue Yu[1]*, Eng Gee Lim[1], Qiufeng Wang[1], and Jimin Xiao[1]

[1] Xi'an Jiaotong-Liverpool University, Suzhou, China
{Shuwei.Wu24, Zhijin.He24}@student.xjtlu.edu.cn
{siyue.yu02, enggee.lim, qiufeng.wang, jimin.xiao}@xjtlu.edu.cn
[2] University of Liverpool, Liverpool, UK
shuo.jin@liverpool.ac.uk

Abstract. Although fully supervised methods have substantially advanced segmentation in boundary areas of point clouds, effective weakly supervised approaches remain scarce. This is primarily because limited supervision rarely reaches boundary regions, leaving them lacking reliable supervision. We propose a novel 2D-assisted pseudo-label propagation paradigm that does not rely on the model's own predictions or any external foundation models, yet is able to generate high-purity pseudo-labels. Compared with SAM-based 2D-3D projection, our pseudo-labels are purer and more uniformly distributed. Even under 1 pt/obj setting on S3DIS, our initial offline propagation achieves >94.4% accuracy (≈ 93k pts per scene). We decomposed the pseudo-labels generation process from the main network, and applied a divide-and-conquer strategy: supervision from interior pseudo-labels serves to stabilize the representation of core class regions, while boundary pseudo-labels are leveraged to enhance boundary robustness. This design reduces the confirmation bias inherent in classic online labeling and alleviates the lack of boundary supervision in existing weakly supervised models. Experiments show that our method outperforms existing state-of-the-art methods. Our codes will be released at: https://github.com/paul-swu/Bound3D.

Keywords: Weakly supervised · 3D segmentation · Pseudo-labels

1 Introduction

3D semantic segmentation, one of the 3D scene understanding tasks, is central to embodied AI and autonomous driving. Yet, per-point annotation for large point clouds is labor-intensive and costly, limiting the dataset scale and slowing the development of 3D foundation models. The rise of weakly supervised 3D semantic segmentation (WS3DSS) has significantly alleviated the requirement of expensive point cloud annotation, with only extremely sparse annotations (*e.g.*, 10%, 1%, 0.1%, or one labeled point per class).

* Corresponding author.

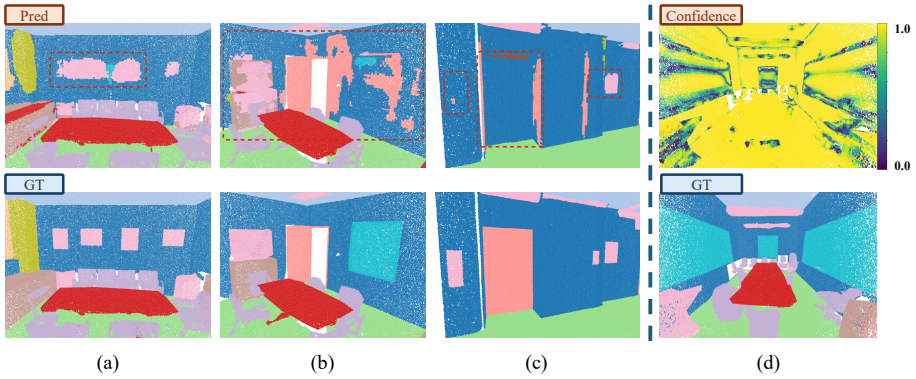

Fig. 1: (a)-(c) show systematic boundary distortion of previous method [24] across different scenes under 0.01% GT supervision. The segmentation issues are marked with red rectangle. (d) shows the predicted confidence score with its GT map.

Consistency-based training [19,24,50] is a widely used paradigm for WS3DSS. This paradigm is built upon that a point's predictions should remain invariant under different perturbations. Although these methods achieve noticeable performance improvements, they still struggle to accurately recover complete object contours, often failing to precisely delineate object boundaries. This limitation is particularly evident in complex scenes, as exemplified by several failure modes in Fig. 1 (a): boundary adhesion between adjacent clutter objects; Fig. 1 (b): the wall and its adjacent door bleed into each other, while the boundaries of a board remain largely unrecognized; and Fig. 1 (c): clutter and the door are engulfed by the dominant wall class.

To probe the key issue in boundary perception, we first try a minimal intervention experiment on the widely-adopted consistency-based training framework [47] for WS3DSS, where we explicitly identify ground-truth boundary regions and impose a higher-weighted consistency loss on these areas instead of assigning uniform loss weights to all the points. Intuitively, assigning a higher loss weight to boundary regions should encourage the model to learn more accurate boundary predictions. However, by evaluating both mIoU and B-mIoU (shown in blue and orange in Fig. 2 (a) respectively), which measure overall segmentation performance and boundary-region performance, we observe a counterintuitive result. Simply increasing the consistency loss weight on boundary points to 1.25x its original value does not yield the expected improvement. Instead, both metrics decline, with B-mIoU exhibiting a more pronounced drop than mIoU. Interestingly, when the loss weight assigned to boundary points is reduced, the model achieves better performance, particularly in terms of B-mIoU. This phenomenon suggests that the bottleneck in boundary perception is not the lack of implicit supervision intensity. But rather, the imposed consistency loss weight blends the semantic distinction and starves boundary information. We attribute this to

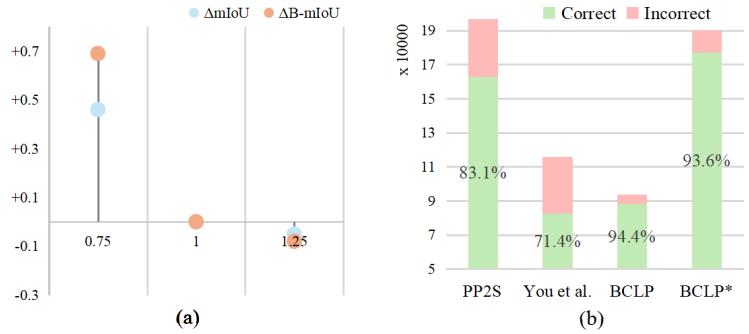

Fig. 2: (a): Vary the consistency loss weight w on boundary regions (with w=1.0 as the baseline). The y-axis reports ΔmIoU and ΔB-mIoU relative to w=1.0. B-mIoU denotes mIoU computed within boundary regions. (b): BCLP vs. SAM-based pseudo-label generation: the average number of pseudo-labels per scene (y-axis) and accuracy (percentage). * denotes a more aggressive hyperparameter setting.

boundary starvation: a lack of reliable supervision signals and explicit semantic distinctions at boundary regions.

To address the boundary starvation issue, the pseudo-label-based approach is a direct solution, as applying cross-entropy loss on boundaries avoids discarding boundary signals and encourages semantic transitions. However, pseudo-label-based self-training methods [17, 19, 25, 52] often exhibit suboptimal accuracy at boundaries. This discrepancy arises from the fact that existing methods typically evaluate the network's prediction (e.g., the maximum softmax probability) and adopt a uniform threshold to select high-confidence pseudo-labels. Yet, this operation tends to prioritize interior points and filter out boundary ones. As depicted in Fig. 1 (d), our analysis using a MinkowskiNet [4] to predict confidences on pseudo-labeled points reveals an inherent bias in this operation. It's observed that many boundary points never reach the threshold (*e.g.*, 0.90) during training, even though pseudo-labels are correct. Consequently, this operation imposes sparse supervision on boundaries and further exacerbates *boundary starvation*.

Therefore, our goal is to generate high-quality pseudo-labels, especially at boundaries, to mitigate the *boundary starvation* issue. To this end, we propose a **Bound**ary-aware **3D** Label Propagation Framework (**Bound3D**) for WS3DSS. In particular, to light up the blind spots at the boundary, we introduce a Boundary Illumination Module (**BIM**) to extract boundary information first and project it into 3D space for high-quality boundary points. Building on these points, we design a Boundary-constrained Pseudo Label Generator (**BPLG**), which consists of two components: a boundary-constrained label propagation mechanism (BCLP) and an iteration trigger (BAR). BCLP leverages boundary points and geometric cues to generate reliable pseudo-labels. Meanwhile, BAR triggers iterative refinements, utilizing evolving network features to continuously correct complex boundaries. As shown in Fig.2 (b), our method achieves higher pseudo-label accuracy than existing best methods [17, 52]. Furthermore,

to avoid overemphasizing interior points, we introduce a Position-Aware Divide-and-Conquer (**PADC**) supervision strategy. Specifically, interior points are assigned to a weak perturbation view to prevent over-regularization on intra-class regions, while boundary points are supervised under a strong perturbation for robust boundary perception. Comprehensive experiments show that Bound3D achieves the new state-of-the-art (SOTA) performance for WS3DSS. Our contributions can be concluded as:

- We diagnose *boundary supervision* risks in consistency-based frameworks, and thus introduce Bound3D to strengthen the model's weakest boundary regions to boost overall performance.
- We design a Boundary-constrained Pseudo Label Generator BPLG, the core of our framework, which leverages multimodal geometric cues with an iterative refinement mechanism to generate high-quality, boundary-preserving pseudo-labels.
- We propose PADC, a divide-and-conquer supervision strategy that applies asymmetric perturbations to adaptively optimize interior and boundary regions, rather than treating all points equally.

2 Related Work

2.1 Weakly Supervised 3D Semantic Segmentation

To reduce reliance on dense 3D labels, Xu et al. [49] introduced a dual-branch framework for sparse supervision, inspiring later works [25, 27, 35, 40, 41, 53]. In parallel, perturbation consistency-based methods [18, 19, 24, 44, 47, 54] improve robustness by aligning point features across augmented views. Prototype- and metric-learning approaches further impose explicit embedding structure [20, 26, 50]. Pseudo-label–based methods [17,19,21,25] face confirmation bias, motivating efforts to improve pseudo-label quality. OTOC [25] adopts graph propagation, and a fast-rising direction leverages 2D foundation models to improve segmentation, particularly in terms of robustness [11, 48], or exploits them to generate high-quality 2D masks and 3D back-projection [17,52]. Kweon et al. [17] initiate this cross-modal paradigm via SAM-based forward–backward projection, and You et al. [52] further refine multi-view mask fusion. Related 2D segmentation advances [9,12,13,33,56] offer complementary priors for such cross-modal designs. Among self-training approaches, Q2E [6] enriches pseudo-labels via CLIP-based 2D assistance and hierarchical category optimization, while ActiveST [22] introduces active re-annotation to rectify low-quality pseudo-labels. However, both rely on either external foundation models or additional human annotation. In contrast, our method combines lightweight multi-view 2D cues with native 3D geometry alone, yielding purer pseudo-labels without relying heavily on SAM [14] or extra labels.

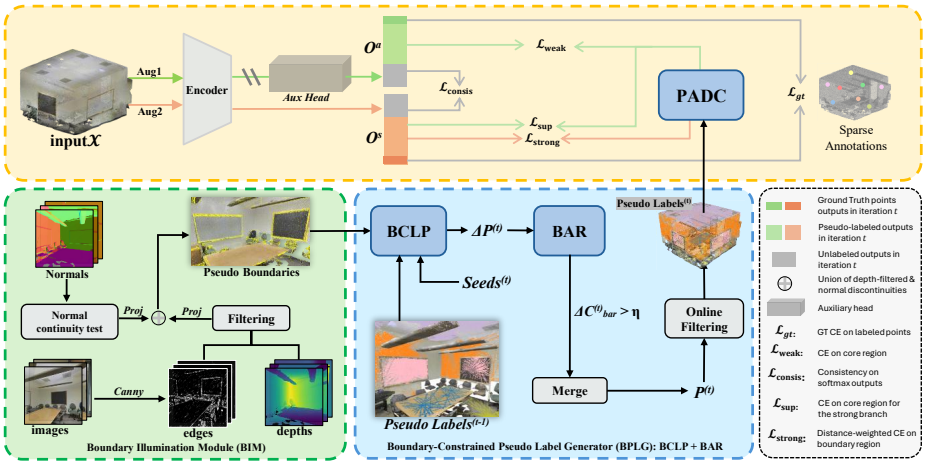

Fig. 3: Overview of our Bound3D framework. The pipeline involves a Boundary Illumination Module (BIM) for 3D pseudo-boundary generation, a Boundary-Constrained Pseudo Label Generator (BPLG) for dynamic label propagation, and a Position-Aware Divide-and-Conquer (PADC) strategy. During training, the input point cloud is augmented twice (where Aug2 denotes a stronger perturbation) and fed into a dual-branch consistency network constrained by PADC. CE denotes cross-entropy loss.

2.2 Boundary-Aware Semantic Segmentation

Incorporating boundary cues is a proven strategy to enhance segmentation robustness [3,7,15,16,29,30,36]. In 3D segmentation, ABL [39] and BECO [34] optimize boundary alignments and boundary-weighted training, respectively. Under full supervision, BAGE [8] treats boundaries as barriers during feature aggregation to block cross-class leakage, CBL [37] applies contrastive learning to boundary pairs, and BFANet [55] employs attention-based boundary fusion. However, these methods inherently rely on dense ground-truth boundaries, which are unavailable under weak supervision. WS3D [23] trains a dedicated boundary prediction network and exploits boundaries via an energy-based loss, whereas our method derives boundaries directly from multi-view geometric cues to constrain pseudo-label propagation, requiring no extra boundary network.

3 Method

3.1 Basic Settings

Given a scene point cloud $\mathbf{X} = \{x_i\}_{i=1}^{n}$ with n points, each input $x_i = [c_i; o_i]$ comprises 3D coordinates $c_i \in \mathbb{R}^3$ (XYZ) and color $o_i \in \mathbb{R}^3$. The weak labels are $\mathbf{Y} = \{y_i\}_{i \in \mathbf{L}}$, where $y_i \in \mathbf{C}$ and $|\mathbf{C}|$ is the number of classes and $\mathbf{L} \subset \mathbf{X}$. The unlabeled set is defined as $\mathbf{U} = \mathbf{X} \setminus \mathbf{L}$. For multi-view 2D observations, Let $\mathcal{V} = \{1, \ldots, V\}$ be the set of available calibrated views. For each view $j \in \mathcal{V}$,

we define an RGB image as $I_j \in \mathbb{R}^{H\times W\times 3}$, a depth map as $D_j \in \mathbb{R}^{H\times W}$, a surface-normal map as $N_j \in \mathbb{R}^{H\times W\times 3}$, the camera intrinsics K_j, and the world-to-camera transform metric $\mathcal{E}_j = \left[\, R_j \;\middle|\; \boldsymbol{\delta}_j \,\right] \in \mathbb{R}^{3\times 4}$, where H and W denote the height and width of the image for view j. During training, the model leverages the point cloud, weak labels, and 2D views, while the inference requires only the point cloud.

3.2 Overview

Fig. 3 illustrates the pipeline of our Bound3D, which consists of three modules: i) a boundary illumination module; ii) a boundary-constrained pseudo-label generator, which includes a boundary-constrained label propagation mechanism and a boundary arrival rate-triggered Iterator; iii) a position-aware divide-and-conquer learning strategy. The overall pipeline is as follows:
1) We first feed the given point cloud together with its multi-view RGB images, depth maps, and normal maps into BIM for boundary detection, where a 2D boundary is extracted by combining depth-filtered Canny edges [2] with normal-discontinuity masks and then projected into the corresponding 3D scene, yielding high-quality 3D boundary points for our BPLG.
2) After that, the detected boundary points guide our BPLG to generate comprehensive pseudo-labels; within BPLG, the BCLP mechanism produces the labels via boundary-constrained propagation, while boundary arrival rate is computed to control online iterations and supplies new seed points for each BCLP iteration.
3) Finally, PADC dispatches the pseudo-labels to a dual-branch consistency network in a divide-and-conquer manner. After training, the inference process requires only the raw point cloud for final segmentation results. Note here that no 2D information is needed.

3.3 Boundary Illumination Module

Detecting boundaries solely from 3D point clouds is challenging due to the limited semantic information. Thus, we propose a Boundary Illumination Module (BIM) leveraging 2D observations. For each view j, we first extract initial binary edges $E_j = \mathcal{C}(I_j)$ using the Canny detector $\mathcal{C}(\cdot)$ on the RGB image I_j. To suppress intra-object texture noise, we construct a depth-discontinuity mask M_j^{dep} using depth gradients:

$$M_j^{\text{dep}} = \mathbf{1}\left[\sqrt{(\partial_u \log D_j)^2 + (\partial_v \log D_j)^2} > \tau_d\right] \cdot \mathbf{1}[\text{valid}(D_j)], \tag{1}$$

where $\mathbf{1}[\cdot]$ is the indicator function, ∂_u, ∂_v are spatial derivatives, $\tau_d > 0$ is a threshold, and $\text{valid}(D_j(u))$ ensures valid depth. The depth-filtered edges are $E'_j(u) = E_j(u) \cap M_j^{\text{dep}}(u)$.

Since depth is view-angle sensitive and degrades at far ranges, we complement it by detecting orientation jumps in the normal map N_j, yielding a normal-

discontinuity mask:

$$M_j^{\text{norm}}(u) = \mathbf{1}\left[\min_{u' \in \mathcal{N}_{r_n}(u)} \left| \, n_j(u)^\top n_j(u') \, \right| < \cos\theta\right], \tag{2}$$

where $n_j(u)$ is the unit normal at pixel u, $^\top$ denotes the transpose operator. $\mathcal{N}_{r_n}(u) = \{\, u' : \|u' - u\|_2 \le r_n \,\}$ is a local window of radius r_n, and θ is the angle threshold. We then fuse boundary pixels by union $B_j(u) = E'_j(u) \cup M_j^{\text{norm}}(u)$ and back-project B_j using intrinsics K_j, extrinsics $\mathcal{E}_j = [R_j \,|\, \boldsymbol{\delta}_j]$, and D_j to obtain the coarse 3D pseudo-boundary set $\boldsymbol{\Phi}$.

3.4 Boundary-constrained Pseudo Label Generator

To propagate weak annotations towards the generated boundaries, our proposed BPLG employs Boundary-Constrained Label Propagation (BCLP) and a Boundary Arrival Rate (BAR) iterator.

Boundary-Constrained Label Propagation We observe that point-cloud scenes are compositions of smooth object surfaces, modeled as connected 2D manifolds ($\mathcal{M} \subset \mathbb{R}^3$). On each surface patch, semantic labels are locally constant and undergo abrupt transitions only at boundaries $\boldsymbol{\Phi}$. Following this boundary-before consistency principle, BCLP propagates pseudo-labels from a labeled seed only along paths that are certified to stay on the same manifold and not cross a boundary (Fig. 4). This is achieved via: (i) a manifold constraint proxy verifying that p and q are co-manifold, and (ii) a farthest angle sampling strategy to maximize pseudo-label diversity.

Point Cloud Manifold Constraint. Directly verifying the co-manifold condition ($p, q \in \mathcal{M}_p$) is computationally intractable. As an efficient proxy, we assume points on the same manifold are densely sampled; thus, a valid propagation path must maintain a local point density ρ above a threshold τ. This ensures that the path exclusively traverses through dense point cloud regions representing the manifold, effectively approximating the co-manifold condition and preventing labels from leaking across voids or propagating to unrelated object surfaces.

$$\mathcal{B}(x_\ell) = \left\{\, x_b \in \boldsymbol{\Phi} \,\middle|\, \|x_b - x_\ell\|_2 \le R_b \,\right\}, \tag{3}$$

where $\mathcal{B}(x_\ell)$ is the boundary-candidate set around x_ℓ. x_b denotes a single pseudo-boundary point in Φ. R_b is the neighborhood radius. For a pair (x_ℓ, x_b), define the Euclidean path $Path$ as the line segment connecting them:

$$\text{Path}(x_\ell, x_b) = \left\{\, x_\ell + \alpha\,(x_b - x_\ell) \;\middle|\; \alpha \in [0,1] \,\right\} \subset \mathbb{R}^3, \tag{4}$$

To avoid exhaustive density checks, we sample a finite set $\mathcal{G} \subset \text{Path}$ at a fixed spacing $\Delta > 0$ to form a finite set $\mathcal{G} \subset \text{Path}$ as:

$$\mathcal{G} = \left\{\, g_m = x_\ell + \alpha_m(x_b - x_\ell) \mid \alpha_m = \frac{m\Delta}{d},\ m = 1, \ldots, M-1 \,\right\}, \tag{5}$$

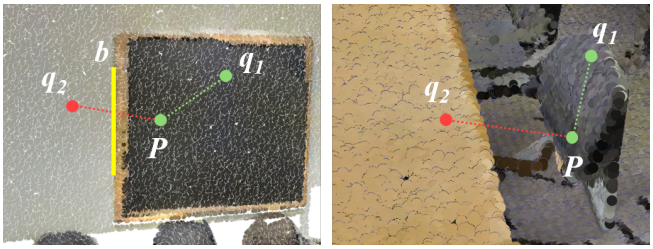

Fig. 4: Visualization of valid (green) and invalid (red) BCLP propagation paths. p denotes a labeled point, q_1, q_2 denote unlabeled candidate points. Left: p, q_1, and q_2 lie on the same manifold, but the path from p to q_2 crosses a boundary, so it is invalid, while the path from p to q_1 is accepted. Right: p and q_2 lie on different manifolds, so the path from p to q_2 is invalid.

where m indexes the uniformly spaced interior samples and $M = \lfloor \frac{d}{\Delta} \rfloor + 1$ is the number of segments induced by Δ and d. The path is considered to satisfy the manifold constraint if and only if the samples meet the local density condition:

$$\forall\, g \in \mathcal{G}: \quad \big|\, \mathcal{N}(g, r) \,\big| \;\geq\; \tau, \tag{6}$$

where $|\mathcal{N}(g, r)|$ is the point count within radius r of g, and τ is the minimum count threshold of points. For a validated safe path, we harvest a point set $\mathcal{S}$ from the original cloud $\mathbf{X}$ within a small search radius r_s:

$$\mathcal{S} = \{x \in \mathbf{X} \mid \min_{x_p \in \mathrm{Path}(x_\ell, x_b)} \|x - x_p\|_2 \leq r_s\}, \tag{7}$$

where $x_p \in \mathrm{Path}(x_\ell, x_b)$. Finally, all harvested points in $\mathcal{S}$ are assigned the pseudo-label y_{x_ℓ}.

Farthest Angle Sampling. Exhaustively sampling paths in $\mathcal{B}(x_\ell)$ is computationally impractical. To obtain uniform, low-correlation supervision under a fixed budget, we propose Farthest Angle Sampling (FAS). Motivated by Xu et al. [49], who show that widely distributed annotations yield more stable gradient approximations than high-variance random sampling, FAS transfers this widespread sampling to the angular domain. Concretely, let $S_k \subseteq \mathcal{B}(x_\ell)$ ($|S_k| = k$) be the selected boundary targets. For any candidate $x \in \mathcal{B}(x_\ell)$, its unit direction from x_ℓ is:

$$\mathbf{u}(x; x_\ell) = \frac{x - x_\ell}{\|x - x_\ell\|_2}. \tag{8}$$

The angle between unit vectors $\mathbf{u}$ and $\mathbf{v}$ is defined as:

$$\theta(\mathbf{u}, \mathbf{v}) = \arccos(\mathbf{u}^\top \mathbf{v}) \in [0, \pi]. \tag{9}$$

To pick endpoints maximally separated from S_k, we define the minimum angular separation of x to S_k as:

$$\theta_{\min}(x, S_k) = \min_{x_m \in S_k} \theta\big(\mathbf{u}(x; x_\ell), \mathbf{u}(x_m; x_\ell)\big), \tag{10}$$

where $x_m \in S_k$. Finally, we adopt a greedy selection to encourage angular uniformity:

$$x_{k+1} = \text{argmax}_{x \in \mathcal{B}(x_\ell) \setminus S_k} \ \theta_{\min}(x, S_k), \tag{11}$$

where x_{k+1} denotes the $(k{+}1)$-th selected candidate.

BAR Triggered Iterative Propagation While recent pseudo-label based self-training pipelines have shown potential in weakly supervised segmentation, their iterative generation mechanisms suffer from several intrinsic limitations: fixed generation schedules, confirmation bias, and error amplification. We overcome these via a dynamic, position-aware propagation scheme:

1) BAR Metric for Dynamic Triggering. To regulate iterations dynamically, we decompose the scene into a boundary region $\mathcal{R}_b$ (points within radius r_b of Φ) and a core region $\mathcal{R}_c = 1 - \mathcal{R}_b$. The Boundary Arrival Rate (BAR) coverage $\mathcal{C}_{bar}$ for pseudo-labels $P^{(t)}$ at round t is defined as:

$$\mathcal{C}_{bar} = \frac{\sum_{x \in \mathcal{P}^{(t)}} \mathbf{1}\big[x \in \mathcal{R}_b\big]}{|\mathcal{R}_b|}. \tag{12}$$

A new BCLP iteration is triggered only if the coverage increment $\Delta\mathcal{C}_{bar} = \mathcal{C}_{bar}^{(t)} - \mathcal{C}_{bar}^{(t-1)} > \eta$. Stagnant growth indicates redundant interior exploration, halting generation.

2) Position-Aware Filtering. Another primary limitation is the over-reliance on high-confidence filtering to ensure pseudo-label quality. This seemingly reasonable strategy inherently biases pseudo-label expansion towards safe interior areas. To prevent high-confidence filtering from discarding challenging boundary labels, we apply region-specific thresholds to the new labels $\Delta P^{(t)}$: τ_b for $\mathcal{R}_b$ and τ_c for $\mathcal{R}_c$. Trusting geometric cues over model predictions at boundaries, we set $\tau_b < \tau_c$. The pseudo-label set is then cumulatively updated as $P^{(t)} \leftarrow P^{(t-1)} \cup \Delta P^{(t)}$.

3) Safe Seed Selection. Existing models rely on a uniform high-confidence threshold for seed point selection in subsequent iterations. Lacking explicit position awareness, such models must resort to applying a single, extreme threshold uniformly across all points. This strategy struggles with the fact that high confidence is not infallible; a 0.8 reliability, for instance, may only correspond to 90% accuracy [17]. Selecting erroneous boundary predictions as new seeds causes severe error amplification. Thus, we explicitly exclude $\mathcal{R}_b$ from the seed pool, selecting high-confidence seeds exclusively from $\mathcal{R}_c$, and apply BCLP to recursively generate a new batch of pseudo-labels.

3.5 Position-Aware Divide-and-Conquer

To move beyond training schemes that treat all points equally and better exploit the high-purity boundary pseudo-labels in $\mathcal{R}_b$, we propose a Position-Aware Divide-and-Conquer **(PADC)** strategy. It aims to decompose the supervision signals based on position while retaining the fundamental structure of

consistency-based methods, leading to a composite loss function. Let the outputs of the weak and strong augmentation views be (F_w, O_w) and (F_s, O_s) respectively, where F denotes features from the backbone and O represents the logits from the main prediction head.

For the weak augmentation branch, we introduce a lightweight auxiliary head to prevent supervision signal conflicts. This head is tasked with stable, intra-class calibration using supervision from the core zone $\mathcal{R}_c$, and serves as a consistency anchor for transfer to the strong branch. $O_{a_x} = \text{AuxHead}(\text{sg}(F_{w_x}))$, where $\text{sg}(\cdot)$ denotes the stop-gradient operation. We denote the target pseudo-label for a point as $\tilde{y}_x$. The weak view supervision loss is defined as:

$$\mathcal{L}_{\text{w}} = \frac{1}{|\mathcal{R}_c|} \sum_{x \in \mathcal{R}_c} \text{CE}(O_{a_x}, \tilde{y}_x). \tag{13}$$

Conversely, the strong branch focuses on challenging boundary samples in $\mathcal{R}_b$. Since ambiguity increases near actual boundaries, we apply a distance-aware weight w_x to each point $x \in \mathcal{R}_b$. Based on its Euclidean distance d_x to the nearest pseudo-boundary in $\boldsymbol{\Phi}$, the weight is:

$$w_x = 1 + \lambda_d \left(1 - \frac{d_x}{r_d}\right), \tag{14}$$

where λ_d controls the weight intensity. The strong supervision loss applies this weighted cross-entropy:

$$\mathcal{L}_{\text{s}} = \frac{\sum_{x \in \mathcal{R}_b} w_x \cdot \text{CE}(O_{s_x}, \tilde{y}_x)}{\sum_{x \in \mathcal{R}_b} w_x}. \tag{15}$$

To prevent the network from developing a narrow field of view due to its primary supervision being restricted to boundary points, we introduce $\mathcal{L}_{\text{sup}}$ as a supplementary loss. This term allows a fraction of the supervision signal from easy, interior samples to leak to the main network with a minimal weight γ, ensuring the strong branch maintains supervision across the entire scene.

$$\mathcal{L}_{\text{sup}} = \frac{1}{|\mathcal{R}_c|} \sum_{x \in \mathcal{R}_c} \text{CE}(O_{s_x}, \tilde{y}_x). \tag{16}$$

Finally, we retain the standard losses from consistency-based methods to form our overall optimization objective.

$$\mathcal{L}_{\text{gt}} = \frac{1}{|\mathbf{L}|} \sum_{i \in L} \text{CE}(O_{s_i}, y_i), \tag{17}$$

$$\mathcal{L}_{\text{consis}} = \frac{1}{|\mathbf{U}|} \sum_{i \in \mathbf{U}} \text{JS}(s(O_{s_i}), s(O_{a_i})), \tag{18}$$

where $\mathbf{U}$ is the set of unlabeled indices, and $s(\cdot)$ denotes the Softmax function. The final objective is the weighted sum of these components:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{gt}} + \mathcal{L}_{\text{w}} + \lambda_1 \mathcal{L}_{\text{s}} + \lambda_2 \mathcal{L}_{\text{consis}} + \gamma \mathcal{L}_{\text{sup}}, \tag{19}$$

where λ_1 and λ_2 are balancing weights.

Fig. 5: Visualizes the pseudo-labels generated by BPLG during training on S3DIS under the 0.01% supervision setting.

Table 1: Pseudo-label quality. †: our re-implementation. Num: pts/scene.

Dataset	Method	Pub. & Year	Num	P-Acc	B-Acc
S3DIS	PP2S† [17]	CVPR 2024	278134.2	77.1	66.1
	You et al. [52]	TIP 2026	115732.7	71.6	-
	BCLP[1]	ECCV 2026	93565.4	**94.4**	**92.7**
ScanNetV2	PP2S [17]	CVPR 2024	-	-	-
	SAM3D [51]	ICCV 2023	62224.5	81.8	-
	You et al. [52]	TIP 2026	5565.5	93.0	-
	BCLP[1]	ECCV 2026	10262.6	**93.3**	**91.0**

4 Experiments

4.1 Setup

Datasets. We evaluate on S3DIS [1] and ScanNetV2 [5]. S3DIS contains 6 areas, 271 scenes, and 13 classes; we follow the standard protocol and use Area 5 for validation. ScanNetV2 contains 1,613 scenes and 20 classes with the official split of 1,201/312/100 for train/val/test. For the annotation budget, we adopt two mainstream settings: a fixed-count budget and a fixed-ratio budget. Specifically, we use 0.01% and 1 pt/obj on S3DIS, and 0.1% and 20 pts/scene on ScanNetV2.
Implementation details. Our work is built upon a consistency dual-branch baseline with a MinkowskiEngine [4] backbone and compares against self-training methods built on SAM-based and Point Transformer–based architectures [42,43].
Metrics. Alongside standard mIoU, we report 1) B-mIoU, the mIoU computed only over boundary regions; 2) P-Acc, the accuracy of pseudo-labels; 3) B-P-Acc, the accuracy of pseudo-labels within boundary regions. Following prior works [37, 55], boundary region is defined by first extracting ground-truth boundary points and then taking the union of fixed-radius neighborhoods around them. Given that boundary segmentation is particularly challenging under weak supervision, we set a slightly larger radius (0.25m).

4.2 Quality of Pseudo-Labels

Comparison with SAM-based Pseudo-Labeling. We compare the proposed BCLP against SAM-based methods [17,51,52] regarding label quantity and qual-

Table 2: Per-class label accuracy of our generated pseudo-labels on S3DIS Area 5.

Class	P-Acc	Class	P-Acc	Class	P-Acc	Class	P-Acc
Ceiling	95.22	Door	88.53	Window	91.03	Chair	91.10
Floor	99.18	Clutter	75.34	Board	82.71	Sofa	95.19
Wall	93.15	Beam	96.13	Table	92.02	**Overall**	**94.40**

Table 3: SOTA comparison on S3DIS Area5. The default backbone of our method is MinkowskiNet (MinkNet). † denotes our re-implementation. OTOC and 1 pt/obj share the same strategy for instance annotation. Bound3D strictly follows the 1 pt/obj setting with a lower annotation ratio (0.004%) compared to OTOC (0.02%).

Supervision	Method	Pub. & Year	mIoU(%)
100%	PointNet [31]	CVPR 2017	41.1
	MinkowskiNet† [4]	CVPR 2019	68.5
	PTv2 [43]	NeurIPS 2022	71.6
	PTv3 [42]	CVPR 2024	73.4
0.01%	SQN [10]	ECCV 2022	45.3
	CPCM [24]	ICCV 2023	59.3
	CPCM† [24]	ICCV 2023	59.8
	PointMatch [44]	Comput. Graph.2023	59.9
	AADNet [28]	AAAI 2025	60.8
	Bound3D-MinkNet (Ours)	ECCV 2026	**61.2**
OTOC (0.02%)	OTOC [25]	CVPR 2021	50.1
	DAT [46]	ECCV 2022	56.5
	RAC-Net [45]	IJCV 2024	58.4
	You et al.-PTv3 [52]	TIP 2026	64.4
1 pt/obj	**Bound3D-MinkNet (Ours)**	ECCV 2026	**61.9**
	AO-PTv2 [17]	CVPR 2024	62.7
	Bound3D-PTv2 (Ours)	ECCV 2026	**63.5**

ity. For fairness, we only compare the first round label initialization from ground truth. We denote the first propagation as $\mathbf{BCLP}^1$ and the subsequent BAR-triggered iterative propagation as $\mathbf{BCLP}^n$. Tab. 1 reports the results. For label budgets, BCLP^1 (ours) and AO–PP2S use 1 pt/obj, whereas You et al. [52] and SAM3D [51] adopt OTOC [25].

As shown in Tab. 1, BCLP outperforms SAM-based baselines at the generation stage. While SAM-based methods introduce cross-modal topological bias and inherited noise by projecting complex 2D masks into 3D, BCLP respects 3D geometry by treating projection noise as local and indirect. This makes the error in subsequent propagation more controllable, yielding purer and more robust pseudo-labels for 3D model training.

SAM-based methods produce denser pseudo-labels because backprojected mask points are tightly packed with few interior gaps. While BCLP^1 can be made denser (e.g., **197k**+ pts/scene) with a more aggressive sampling policy while still maintaining high accuracy (93.6%), our experiments show such redundancy is detrimental in our architecture. In fact, we prefer pseudo-labels that are uniformly distributed, broader in coverage, and of high precision. Fig. 5 visualizes the pseudo-labels generated by BPLG during training.

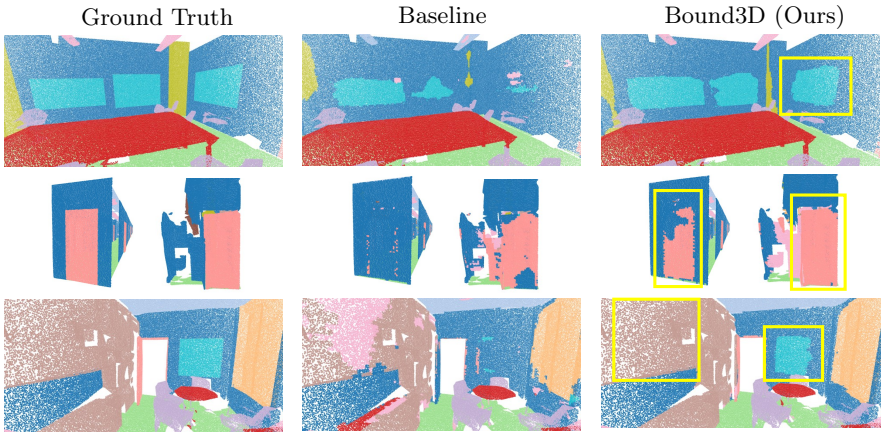

Fig. 6: Visualization results on the S3DIS test set. From left to right are Ground Truth, baseline, and Bound3D. Yellow boxes highlight the superior performance of Bound3D in boundary regions, effectively identifying clear object boundaries.

To verify the high P-Acc is not inflated by dominant planar classes, we report per-class P-Acc in Tab. 2. BCLP sustains high purity across minor and medium-sized classes: structured classes such as board (82.7%) and door (88.5%) stay well above 80%. This indicates the overall 94.4% reflects genuine pseudo-label quality rather than dominance by large regions. However, our boundary cues are derived from 2D edges and geometric discontinuities, which can be insufficient for thin or small objects whose boundaries are weakly observed across views (as reflected in the lower clutter accuracy in Tab. 2). Distance-based shape modeling could better capture such fine structures; we regard this, together with extension to outdoor LiDAR-dominated settings that exhibit sparser viewpoints and noisier geometry, as a promising direction for extending Bound3D.

4.3 Comparison with State-of-the-Art Methods

Tab. 3 compares Bound3D with SOTA methods on S3DIS. Under 0.01% supervision, Bound3D outperforms PointMatch by 1.3% mIoU points and CPCM by 1.4%, and it nearly matches SQN trained with 0.1% labels. Moreover, under the same setting as AO (1 pt/obj), despite using neither SAM nor a Point Transformer backbone, Bound3D is only 0.8% mIoU behind, yet surpasses it by 0.8% under the same PTv2 backbone. Even in this extremely sparse setting, Bound3D still outperforms most mainstream models under the OTOC (0.02%) setup. Beyond quantitative metrics, the qualitative visualizations in Fig. 6 further demonstrate superior performance of Bound3D. Tab. 4 reports results on the ScanNetV2 test set. With 0.1% labels, Bound3D outperforms SQN by 6.3% mIoU. Under the 20 pts/scene setting, it outperforms MIL, DAT, and OTOC by 8.0%, 7.2%, and 3.0% mIoU, respectively. These results indicate that Bound3D effectively exploits extremely sparse per-instance annotations, even

Table 4: SOTA comparison on ScanNetV2 (Test) under varying supervision budgets.

Supervision	Method	Pub. & Year	mIoU(%)
100%	PointNet++ [32]	NeurIPS 2017	55.7
	KPConv [38]	ICCV 2019	68.4
	MinkowskiNet [4]	CVPR 2019	73.6
0.1%	SQN [10]	ECCV 2022	56.9
	Bound3D-MinkNet (Ours)	ECCV 2026	**63.2**
20 pts	MIL [50]	CVPR 2022	54.4
	DAT [46]	ECCV 2022	55.2
	OTOC [25]	CVPR 2021	59.4
	Bound3D-MinkNet (Ours)	ECCV 2026	**62.4**

Table 5: Ablation study of Bound3D on S3DIS Area 5 (0.01%). **Upper:** Incremental addition of core modules to the baseline. **Lower:** Component-level leave-one-out analysis within the iterative BCLP^n generator. The full configuration is in bold.

Method	mIoU (%)	B-mIoU (%)
Module-level		
MinkowskiNet	49.3	44.4
Baseline	55.8	49.0
Baseline + BCLP^1	56.7	49.4
Baseline + BCLP^n	60.7	52.9
Baseline + BCLP^n + PADC (Bound3D)	**61.2**	**54.5**
Component-level (leave-one-out within BCLP^n)		
w/o FAS	59.3	51.6
w/o distinct thresholds	58.9	51.2
w/o distance weighting	60.0	52.8

though they account for only a very small proportion of all scene points. Additionally, Bound3D avoids the extensive model inferences and repeated projection operations required by SAM-based pipelines.

4.4 Ablation on core components

As shown in Tab. 5, we conduct a comprehensive ablation study to validate the effectiveness of each key component in Bound3D. We start with a baseline model (Mink [4]) trained only on sparse annotations; lacking a mechanism to handle sparse labels, this model exhibits poor performance. Building upon this, we introduce a perturbation consistency framework (partially adopting strategies from DAT [46]), which serves as our main baseline. This baseline achieves a significant performance gain by effectively utilizing unlabeled data. The third row validates the efficacy of the BCLP strategy, which lifts mIoU by 0.9% and B-mIoU by 0.4% over the baseline, showing that our label propagation already yields useful supervision beyond standard consistency. The fourth row show that BPLG adds a further 4.0% mIoU / 3.5% B-mIoU, indicating that its pseudo-label filtering and iteration mechanism provides a significant boost. Finally, PADC further sharpens boundary learning, yielding 0.5% mIoU and 1.6% B-mIoU. This demonstrates that PADC successfully enhances boundary robustness without compromising the stability of interior regions.

Table 6: Sensitivity analysis of key hyperparameters in Bound3D on S3DIS Area 5 under 0.01% supervision. (a) Boundary-illumination and label-propagation hyperparameters. (b) Additional hyperparameters of the propagation, triggering, and supervision stages. Default values are highlighted in gray.

Depth Thr. τ_d			**Normal Thr.** θ			**Radius** r_s			**Step Size** Δ		
Val.	mIoU	B-mIoU	Val.	mIoU	B-mIoU	Val.	mIoU	B-mIoU	Val.	mIoU	B-mIoU
0.05	58.9	52.5	10°	56.5	50.5	0.01	58.8	52.3	0.06	60.8	54.1
0.10	60.7	54.0	20°	59.4	53.3	0.03	60.7	54.1	0.08	61.2	54.4
0.15	**61.2**	**54.5**	**30°**	**61.2**	**54.5**	**0.05**	**61.2**	**54.5**	**0.10**	**61.2**	**54.5**
0.20	61.1	54.6	40°	61.0	54.3	0.07	59.5	53.0	0.15	60.1	54.5
0.25	60.6	54.1	45°	59.3	52.9	0.10	56.8	50.6	0.20	60.7	54.2

(a) Boundary illumination and label propagation

τ		R_b		r_b		η		r_d	
Val.	P-Acc	Val.	P-Acc	Val.	mIoU	Val.	mIoU	Val.	B-mIoU
2	94.29	1.0	93.80	0.15	60.91	0.06	60.78	0.10	54.13
4	94.35	1.5	94.14	0.20	61.15	0.08	60.95	0.15	54.33
5	**94.40**	**2.0**	**94.40**	**0.25**	**61.21**	**0.10**	**61.21**	**0.20**	**54.52**
10	94.69	2.5	94.37	0.30	61.06	0.12	60.82	0.25	54.47

(b) Propagation, triggering, and supervision

Within BCLP^n, a leave-one-out analysis confirms each component contributes: distinct thresholds are especially critical for boundary quality (-1.8% mIoU / -1.7% B-mIoU when removed), while FAS and distance weighting yield consistent gains. We further verify the stop-gradient in PADC's auxiliary head: removing it lets noisy boundary pseudo-labels corrupt the main head's interior semantics, dropping performance to 60.3/53.0 mIoU/B-mIoU ($-0.9/-1.5$). Furthermore, we conduct a detailed sensitivity analysis across all stages of Bound3D (Tab. 6), covering the boundary-illumination and propagation hyperparameters (τ_d, θ, r_s, Δ) as well as the propagation, triggering, and supervision hyperparameters (τ, R_b, r_b, η, r_d). Bound3D attains its best performance under the default configuration and remains robust within a reasonable range of variation.

5 Conclusion

We identify *boundary starvation* in WS3DSS, and introduce BCLP, which leverages multi-modal boundary cues to produce purer pseudo-labels without relying on costly vision models. Furthermore, our Position-aware Divide-and-Conquer strategy breaks the uniform regularization that typically suppresses boundary learning. We hope this work will motivate the community to prioritize discriminative feature learning at point cloud class boundaries, rather than solely pursuing overall mIoU metrics that are dominated by interior and dominant classes.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (No. 62301451, 62471405, 62331003), Basic Research Program of Jiangsu (BK2024 1814), Suzhou Basic Research Program (SYG202316), XJTLU REF-22-01-010.

References

1. Armeni, I., Sax, S., Zamir, A.R., Savarese, S.: Joint 2D-3D-Semantic data for indoor scene understanding. arXiv preprint arXiv:1702.01105 (2017)
2. Canny, J.: A computational approach to edge detection. IEEE TPAMI **8**(6), 679–698 (1986)
3. Cheng, B., Girshick, R., Dollar, P., Berg, A.C., Kirillov, A.: Boundary IoU: Improving object-centric image segmentation evaluation. In: CVPR. pp. 15334–15342 (2021)
4. Choy, C., Gwak, J., Savarese, S.: 4D spatio-temporal convnets: Minkowski convolutional neural networks. In: CVPR. pp. 3070–3079 (2019)
5. Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T., Nießner, M.: Scannet: Richly-annotated 3D reconstructions of indoor scenes. In: CVPR. pp. 2432–2443 (2017)
6. Deng, J., Lu, J., Zhang, T.: Quantity-quality enhanced self-training network for weakly supervised point cloud semantic segmentation. IEEE TPAMI **47**(5), 3580–3596 (2025). https://doi.org/10.1109/TPAMI.2025.3532637
7. Du, S., Ibrahimli, N., Stoter, J., Kooij, J., Nan, L.: Push-the-boundary: Boundary-aware feature propagation for semantic segmentation of 3D point clouds. In: 3DV. pp. 124–133 (2022). https://doi.org/10.1109/3DV57658.2022.00025
8. Gong, J., Xu, J., Tan, X., Zhou, J., Qu, Y., Xie, Y., Ma, L.: Boundary-aware geometric encoding for semantic segmentation of point clouds. In: AAAI. pp. 1424–1432 (2021)
9. He, Z., Jin, S., Yu, S., Wu, S., Zhang, B., Yu, L., Xiao, J.: TF-SSD: A strong pipeline via synergic mask filter for training-free co-salient object detection. In: CVPR. pp. 32216–32225 (2026)
10. Hu, Q., Yang, B., Fang, G., Guo, Y., Leonardis, A., Trigoni, N., Markham, A.: SQN: Weakly-supervised semantic segmentation of large-scale 3D point clouds. In: ECCV. pp. 600–619 (2022)
11. Jiang, L., Shi, S., Schiele, B.: Open-vocabulary 3d semantic segmentation with foundation models. In: CVPR. pp. 21284–21294 (2024)
12. Jin, S., Yu, S., Zhang, B., Sun, M., Dong, Y., Xiao, J.: Feature purification matters: Suppressing outlier propagation for training-free open-vocabulary semantic segmentation. In: ICCV. pp. 20291–20300 (2025)
13. Jin, S., Yu, S., Zhang, B., Yao, C., Liu, M., Xiao, J.: Talent: Target-aware efficient tuning for referring image segmentation. In: CVPR. pp. 7472–7482 (2026)
14. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: ICCV. pp. 4015–4026 (2023)
15. Kirillov, A., Wu, Y., He, K., Girshick, R.: PointRend: Image segmentation as rendering. In: CVPR. pp. 9799–9808 (2020)
16. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected CRFs with gaussian edge potentials. In: NeurIPS. pp. 109–117 (2011)
17. Kweon, H., Kim, J., Yoon, K.J.: Weakly supervised point cloud semantic segmentation via artificial oracle. In: CVPR. pp. 3721–3731 (2024)
18. Lei, X., Guan, H., Ma, L., Yu, Y., Dong, Z., Gao, K., Delavar, M.R., Li, J.: WSPointNet: A multi-branch weakly supervised learning network for semantic segmentation of large-scale mobile laser scanning point clouds. International Journal of Applied Earth Observation and Geoinformation **115**, 103129 (2022). https://doi.org/10.1016/j.jag.2022.103129

19. Li, M., Xie, Y., Shen, Y., Ke, B., Qiao, R., Ren, B., Lin, S., Ma, L.: Hybridcr: Weakly-supervised 3d point cloud semantic segmentation via hybrid contrastive regularization. In: CVPR. pp. 14930–14939 (2022)
20. Li, R., Cao, A.Q., de Charette, R.: COARSE3D: Class-prototypes for contrastive learning in weakly-supervised 3D point cloud segmentation. In: BMVC (2022)
21. Li, X., Xu, Q., Zhang, J., Zhang, T., Yu, Q., Sheng, L., Xu, D.: Multi-modality affinity inference for weakly supervised 3D semantic segmentation. In: AAAI. pp. 3216–3224 (2024)
22. Liu, G., van Kaick, O., Huang, H., Hu, R.: Active self-training for weakly supervised 3d scene semantic segmentation. Computational Visual Media **10**(3), 425–438 (2024). https://doi.org/10.1007/s41095-022-0311-7
23. Liu, K., Zhao, Y., Nie, Q., Gao, Z., Chen, B.M.: Weakly supervised 3D scene segmentation with region-level boundary awareness and instance discrimination. In: ECCV. pp. 37–55 (2022)
24. Liu, L., Zhuang, Z., Huang, S., Xiao, X., Xiang, T., Chen, C., Wang, J., Tan, M.: CPCM: Contextual point cloud modeling for weakly-supervised point cloud semantic segmentation. In: ICCV. pp. 18413–18422 (2023)
25. Liu, Z., Qi, X., Fu, C.W.: One Thing One Click: A self-training approach for weakly supervised 3D semantic segmentation. In: CVPR. pp. 1726–1736 (2021)
26. Pan, Z., Gao, W., Liu, S., Li, G.: Distribution guidance network for weakly supervised point cloud semantic segmentation. NeurIPS **37**, 32400–32420 (2024)
27. Pan, Z., Zhang, N., Gao, W., Liu, S., Li, G.: Less is more: Label recommendation for weakly supervised point cloud semantic segmentation. In: AAAI. pp. 4397–4405 (2024)
28. Pan, Z., Zhang, N., Gao, W., Liu, S., Li, G.: Point cloud semantic segmentation with sparse and inhomogeneous annotations. In: AAAI. pp. 6354–6362 (2025). https://doi.org/10.1609/aaai.v39i6.32680
29. Peng, S., Jiang, W., Pi, H., Li, X., Bao, H., Zhou, X.: Deep snake for Real-Time instance segmentation. In: CVPR (2020)
30. Pu, M., Huang, Y., Liu, Y., Guan, Q., Ling, H.: EDTER: Edge detection with transformer. In: CVPR. pp. 1392–1412 (2022)
31. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: PointNet: Deep learning on point sets for 3D classification and segmentation. In: CVPR. pp. 77–85 (2017)
32. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: PointNet++: Deep hierarchical feature learning on point sets in a metric space. In: NeurIPS. pp. 5099–5108 (2017)
33. Qiu, X., Yu, S., Zhang, B., Zhang, Z., Tillo, T., Xiao, J.: Unleashing the power of optimal head in CLIP and DINO for weakly supervised semantic segmentation. Pattern Recognition **178**, 113454 (2026)
34. Rong, S., Tu, B., Wang, Z., Li, J.: Boundary-enhanced co-training for weakly supervised semantic segmentation. In: CVPR. pp. 19574–19584 (2023)
35. Shen, L., Cao, Y., Wei, X., Li, S.: Weakly supervised point cloud semantic segmentation using pseudo-label reliability and consistency regularization. Neurocomputing **639**, 130241 (2025). https://doi.org/10.1016/j.neucom.2025.130241
36. Su, Z., Zhang, J., Wang, L., Zhang, H., Liu, Z., Pietikäinen, M., Liu, L.: Lightweight pixel difference networks for efficient visual representation learning. IEEE TPAMI **45**(12), 14956–14974 (2023). https://doi.org/10.1109/TPAMI.2023.3300513
37. Tang, L., Zhan, Y., Chen, Z., Yu, B., Tao, D.: Contrastive boundary learning for point cloud segmentation. In: CVPR. pp. 8489–8499 (2022)
38. Thomas, H., Qi, C.R., Deschaud, J.E., Marcotegui, B., Goulette, F., Guibas, L.J.: Kpconv: Flexible and deformable convolution for point clouds. In: ICCV. pp. 6410–6419 (2019)

39. Wang, C., Zhang, Y., Cui, M., Ren, P., Yang, Y., Xie, X., Hua, X.S., Bao, H., Xu, W.: Active boundary loss for semantic segmentation. In: AAAI. pp. 2397–2405 (2022)
40. Wei, J., Lin, G., Yap, K.H., Hung, T.Y., Xie, L.: Multi-path region mining for weakly supervised 3D semantic segmentation on point clouds. In: CVPR. pp. 4384–4393 (2020)
41. Wu, T.H., Liu, Y.C., Huang, Y.K., Lee, H.Y., Su, H.T., Huang, P.C., Hsu, W.H.: ReDAL: Region-based and diversity-aware active learning for point cloud semantic segmentation. In: ICCV. pp. 15510–15519 (2021)
42. 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: CVPR. pp. 4840–4851 (2024)
43. Wu, X., Lao, Y., Jiang, L., Liu, X., Zhao, H.: Point transformer V2: Grouped vector attention and partition-based pooling. In: NeurIPS (2022)
44. Wu, Y., Yan, Z., Cai, S., Li, G., Han, X., Cui, S.: PointMatch: A consistency training framework for weakly supervised semantic segmentation of 3D point clouds. Computers & Graphics **116**, 427–436 (2023). https://doi.org/10.1016/j.cag.2023.09.006
45. Wu, Z., Wu, Y., Lin, G., Cai, J.: Reliability-adaptive consistency regularization for weakly-supervised point cloud segmentation. IJCV pp. 2276–2289 (2024). https://doi.org/10.1007/s11263-023-01975-8
46. Wu, Z., Wu, Y., Lin, G., Cai, J., Qian, C.: Dual adaptive transformations for weakly supervised point cloud segmentation. In: ECCV. pp. 78–96 (2022)
47. Xie, S., Gu, J., Guo, D., Qi, C.R., Guibas, L., Litany, O.: PointContrast: Unsupervised pre-training for 3D point cloud understanding. In: ECCV. pp. 574–591 (2020)
48. Xu, X., Yuan, Y., Li, J., Zhang, Q., Jie, Z., Ma, L., Tang, H., Sebe, N., Wang, X.: 3D weakly supervised semantic segmentation with 2D vision-language guidance. In: ECCV. pp. 87–104 (2024)
49. Xu, X., Lee, G.H.: Weakly supervised semantic point cloud segmentation: Towards 10x fewer labels. In: CVPR. pp. 13703–13712 (2020)
50. Yang, C.K., Wu, J.J., Chen, K.S., Chuang, Y.Y., Lin, Y.Y.: An MIL-derived transformer for weakly supervised point cloud segmentation. In: CVPR. pp. 11830–11839 (2022)
51. Yang, Y., Wu, X., He, T., Zhao, H., Liu, X.: Sam3D: Segment anything in 3D scenes. arXiv preprint arXiv:2306.03908 (2023)
52. You, L., Wu, Z., Liu, W., Yang, X., Cheng, J., Zhou, W., Veeravalli, B., Lin, G.: Integrating sam supervision for 3D weakly supervised point cloud segmentation. IEEE TIP **35**, 5212–5223 (2026). https://doi.org/10.1109/TIP.2026.3691686
53. Zhang, Y., Qu, Y., Xie, Y., Li, Z., Zheng, S., Li, C.: Perturbed self-distillation: Weakly supervised large-scale point cloud semantic segmentation. In: ICCV. pp. 15520–15528 (2021)
54. Zhang, Z., Girdhar, R., Joulin, A., Misra, I.: Self-supervised pretraining of 3D features on any point-cloud. In: ICCV. pp. 10232–10243 (2021)
55. Zhao, W., Zhang, R., Wang, Q., Cheng, G., Huang, K.: BFANet: Revisiting 3D semantic segmentation with boundary feature analysis. In: CVPR. pp. 29395–29405 (2025)
56. Zhu, H., Jin, S., Liao, W., Xiao, J., Zhu, Y., Yu, S., Dai, F.: VIP: Visual-guided prompt evolution for efficient dense vision-language inference. arXiv preprint arXiv:2605.12325 (2026)