Biologically-Inspired Evolutionary Domain Symbiosis for Few-shot and Zero-shot Point Cloud Semantic Segmentation

Changshuo Wang[1], Zhijian Hu[2], Xiang Fang[3*], Zai Yang Yu[4, 8], Yibin Wu[5], Mingkun Xu[6], Yusong Wang[6], Xingyu Gao[7, 8], Prayag Tiwari[9]

[1]Department of Computer Science, University College London, London, United Kingdom
[2]LAAS-CNRS, University of Toulouse, CNRS, Toulouse, France
[3]Interdisciplinary Graduate Programme, Nanyang Technological University, Singapore
[4]Institute of Semiconductors, Chinese Academy of Sciences, Beijing, China
[5] Institute of Geodesy and Geoinformation, University of Bonn, Bonn, Germany
[6] Guangdong Institute of Intelligence Science and Technology, Zhuhai, China
[7] Institute of Microelectronics, Chinese Academy of Sciences, Beijing, China
[8] University of Chinese Academy of Sciences, Beijing, China
[9]School of Information Technology, Halmstad University, Sweden
wangchangshuo1@gmail.com, huzhijian1991@gmail.com, xfang9508@gmail.com, yuzaiyang@semi.ac.cn, yibin.wu@igg.uni-bonn.de, xumingkun@gdiist.cn, wangyi@lr.pi.titech.ac.jp, gxy9910@gmail.com, prayag.tiwari@ieee.org

Abstract

Few-shot and zero-shot point cloud semantic segmentation aim to accurately segment novel categories using limited or no labeled samples, respectively. However, existing methods face significant challenges including domain shifts between support and query sets and the inability to handle both few-shot and zero-shot scenarios within a unified framework. To address these issues, we propose a biologically-inspired **E**volutionary **D**omain **S**ymbiosis **Net**work **EDS-Net** for unified few-shot and zero-shot point cloud semantic segmentation. Specifically, inspired by natural symbiotic evolution, we propose a **S**ymbiotic **E**volution **M**odule (**SEM**) that models co-adaptation between support and query features through self-correlation and cross-correlation mechanisms. Second, motivated by genetic crossover mechanisms, we introduce a **V**ision-**S**emantic **B**ridging **M**odule (**VSBM**) that treats visual prototypes and semantic prototypes as two "parent" individuals, creating fused offspring prototypes through adaptive crossover operations and mutation strategies for zero-shot scenarios. Third, we develop a multi-generational evolutionary optimization framework employing an adaptive gating network to learn optimal fusion weights across different evolutionary stages. Extensive experiments demonstrate that EDS-Net with biological interpretability achieves state-of-the-art performance on both few-shot and zero-shot tasks.

1 Introduction

Point cloud semantic segmentation (Zhang et al. 2023b; Wang et al. 2025a) serves as a fundamental task in 3D computer vision, with critical applications ranging from autonomous driving (Zhao et al. 2023; Chib and Singh 2023) to robotics (Soori, Arezoo, and Dastres 2023; Goel and Gupta 2020) and augmented reality (Devagiri et al. 2022; Sereno et al. 2020). Despite significant advances in fully-supervised methods (Wang et al. 2019; Zhao et al. 2021), the acquisition of large-scale annotated 3D data remains prohibitively expensive and time-consuming. This fundamental limitation has driven researchers to explore few-shot and zero-shot learning paradigms, where models must generalize to novel categories with minimal or no labeled samples.

Table 1: Comparison of point cloud semantic segmentation methods. Green checkmarks (✓) indicate supported or used capabilities, while red crosses (✗) denote unsupported or unused features.

Method	Pre-training	Few-shot	Zero-shot
AttMPTI (Zhao, Chua, and Lee 2021)	✓	✓	✗
PAP3D (He et al. 2023)	✓	✓	✓
SegPN (Zhu et al. 2024)	✗	✓	✗
TaylorSeg (Wang et al. 2025b)	✗	✓	✗
DyPolySeg (Wang, Fang, and Tiwari 2025)	✗	✓	✗
EDS-Net (Ours)	✗	✓	✓

Few-shot point cloud semantic segmentation leverages a small set of labeled support samples to segment query point clouds of the same categories, while zero-shot segmentation aims to identify completely unseen categories without any visual examples during training. Current methods (Zhao, Chua, and Lee 2021; He et al. 2023; Zhu et al. 2024) face three critical challenges: First, existing approaches typically address either few-shot or zero-shot scenarios independently, lacking a unified framework that can handle both tasks seamlessly. Second, domain shifts between support and query sets, as well as between visual and semantic modalities, significantly degrade segmentation performance. Third, limited support samples in few-shot scenarios and the complete absence of visual examples in zero-shot settings make it difficult to learn discriminative prototypes.

Biological research (Wierstra et al. 2014; Bonner 1988) reveals that natural evolution follows sophisticated adaptive mechanisms where species co-evolve through symbiotic re-

*Corresponding Authors.

lationships, refining their characteristics across generations. Inspired by these biological principles, we observe striking parallels between natural evolution and point cloud segmentation challenges. In symbiotic evolution, two species benefit from mutual interaction and adapt to each other's presence over evolutionary time. Similarly, support and query features in few-shot segmentation can be viewed as two "species" that should co-evolve to bridge domain gaps. Furthermore, genetic crossover mechanisms in reproduction, where genetic material is exchanged between parent individuals to create superior offspring, can be leveraged to fuse visual and semantic prototypes for zero-shot scenarios.

As shown in Table 1, existing methods suffer from significant limitations. Pre-training-based approaches (Zhao, Chua, and Lee 2021; He et al. 2023) require substantial computational resources and may introduce domain biases from pre-training datasets. Non-pre-training methods (Zhu et al. 2024; Wang et al. 2025b; Wang, Fang, and Tiwari 2025) avoid these issues but are limited to few-shot scenarios only. Currently, only PAP3D (He et al. 2023) addresses both few-shot and zero-shot tasks, but it still relies on pre-training and lacks biological interpretability in its design.

To address these limitations, we propose EDS-Net, a novel Evolutionary Domain Symbiosis Network that unifies few-shot and zero-shot point cloud semantic segmentation through biologically-inspired mechanisms. Specifically, our Symbiotic Evolution Module (SEM) simulates mutualistic relationships between support and query features, where both "species" benefit from their interaction and co-adapt to reduce domain gaps. The Vision-Semantic Bridging Module (VSBM) implements genetic crossover mechanisms, treating visual prototypes and text-generated pseudo prototypes as two "parent" individuals that exchange genetic material to produce superior "offspring" prototypes for zero-shot scenarios. A multi-generational evolutionary optimization framework employs adaptive gating networks to learn optimal fusion weights across different evolutionary stages, enabling progressive prototype refinement.

The main contributions of this work are summarized as follows:

- We propose EDS-Net, the first unified framework for few-shot and zero-shot point cloud semantic segmentation that leverages biologically-inspired evolutionary mechanisms to progressively refine prototypes through multi-generational optimization.
- We introduce a novel Symbiotic Evolution Module (SEM) that models co-adaptation between support and query features through self-correlation and cross-correlation mechanisms, simulating mutualistic relationships in biological evolution.
- We design a Vision-Semantic Bridging Module (VSBM) inspired by genetic crossover mechanisms that treats visual and semantic prototypes as "parent" individuals, generating superior "offspring" prototypes through adaptive crossover operations and mutation strategies.
- Extensive experiments on S3DIS and ScanNet datasets demonstrate that our biologically-interpretable EDS-Net achieves state-of-the-art performance on both few-shot and zero-shot settings while requiring no pre-training.

2 Related Works

2.1 3D Point Cloud Semantic Segmentation

3D point cloud semantic segmentation aims to assign semantic labels to each point in a point cloud. Early methods relied on handcrafted features (Weinmann et al. 2015), but deep learning has revolutionized this field. PointNet (Qi et al. 2017a) pioneered direct point cloud processing using shared MLPs, while PointNet++ (Qi et al. 2017b) introduced hierarchical feature learning. DGCNN (Wang et al. 2019) proposed EdgeConv operations to capture local geometric structures through dynamic graphs. Recent advances include Transformer-based methods (Zhao et al. 2021; Lai et al. 2022; Park et al. 2022; Wang et al. 2024) that leverage self-attention for long-range dependencies, and improved convolution operations (Thomas et al. 2019; Wu et al. 2022; Wang et al. 2025a) for better local feature modeling. Self-supervised approaches (Chen et al. 2023; Pang et al. 2022; Liang et al. 2025) have emerged to learn robust representations without extensive annotations. However, these fully-supervised methods require large-scale labeled datasets and struggle with novel categories.

2.2 Few-shot and Zero-shot Point Cloud Semantic Segmentation

Few-shot point cloud segmentation addresses learning with limited labeled samples. AttMPTI (Zhao, Chua, and Lee 2021) pioneered this field with episodic training and prototype-based learning. Subsequent works improved feature representations (Mao et al. 2022; Zhang et al. 2023a), optimized prototypes (He et al. 2023; Xu, Zhao, and Lee 2023), and explored novel architectures (Zhu et al. 2024; Wang et al. 2025b; Wang, Fang, and Tiwari 2025). SegPN (Zhu et al. 2024) eliminated pre-training requirements, while TaylorSeg (Wang et al. 2025b) used Taylor series-inspired convolutions for local structure modeling. DyPolySeg (Wang, Fang, and Tiwari 2025) used dynamic polynomial convolution (DyPolyConv) for local geometry and Mamba for global context, with a prototype completion module (PCM) to bridge query-support gaps. Zero-shot segmentation is more challenging, requiring identification of unseen categories without visual examples. PAP3D (He et al. 2023) represents the most relevant work, providing a unified framework for both few-shot and zero-shot scenarios by aligning visual and semantic prototypes. However, it requires extensive pre-training and lacks interpretable design principles.

3 Method

In this section, we first introduce the problem definition of point cloud few-shot and zero-shot semantic segmentation and the theory of natural evolution. Then we introduce the various modules of EDS-Net. Finally, we introduce the training process of EDS-Net, as illustrated in Fig. 1.

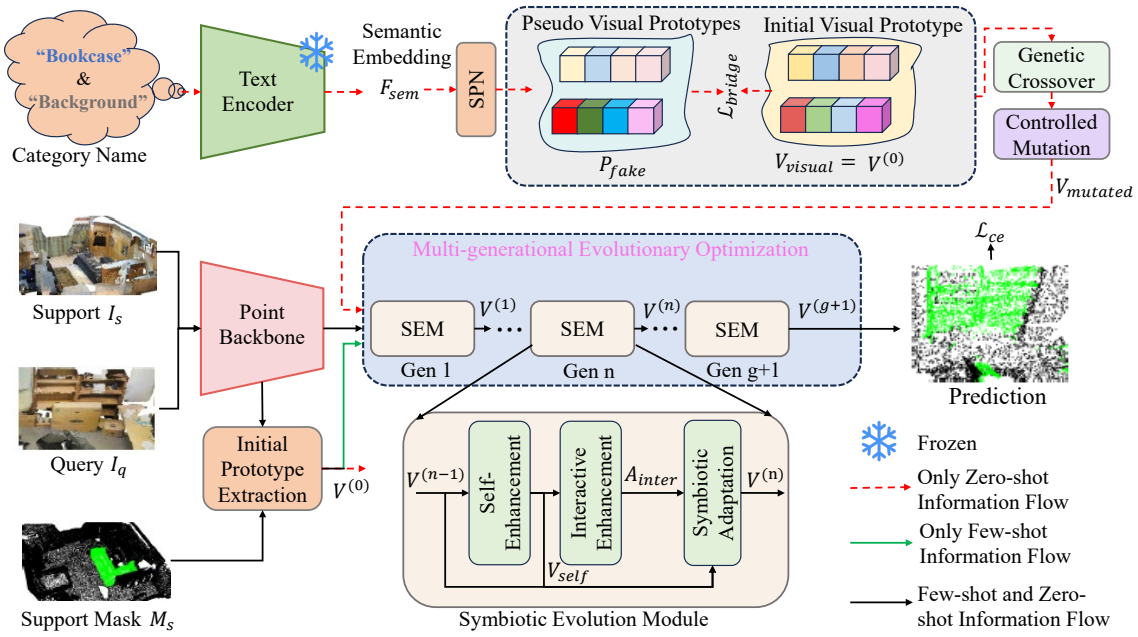

Figure 1: Overall architecture of EDS-Net for unified few-shot and zero-shot point cloud semantic segmentation. The framework employs dual pathways with biologically-inspired evolutionary mechanisms. The few-shot pathway processes support and query point clouds through Point Cloud backbone, followed by Symbiotic Evolution Module (SEM) that models co-adaptation between support and query features. The zero-shot pathway additionally incorporates semantic embeddings through Semantic Projection Network (SPN) and Vision-Semantic Bridging Module (VSBM) inspired by genetic crossover mechanisms. Both pathways utilize multi-generational evolutionary optimization. Training employs cross-entropy loss for both tasks, with additional domain bridging loss for zero-shot scenarios.

3.1 Problem Definition

Few-shot and zero-shot point cloud semantic segmentation aim to segment novel categories with limited or no visual training examples. We formulate both tasks within a unified episodic learning framework inspired by evolutionary optimization principles.

Few-shot Setting. Following standard protocols(Zhao, Chua, and Lee 2021), we adopt an N-way K-shot episodic paradigm where each episode contains a support set $\mathcal{S} = \{(I_s^{n,k}, M_s^{n,k})\}_{n=1,k=1}^{N,K}$ and a query set $\mathcal{Q} = \{I_q^i\}_{i=1}^{H}$. Here, $I_s^{n,k} \in \mathbb{R}^{T\times D}$ represents the k-th point cloud of class n with T points and D-dimensional features (coordinates plus optional attributes), while $M_s^{n,k} \in \{0,1\}^T$ denotes the corresponding binary segmentation mask. The query set contains H unlabeled point clouds to be segmented into N target classes plus background. The few-shot objective is to learn a mapping function:

$$\hat{Y}_q = f_\theta(\mathcal{S}, I_q), \tag{1}$$

where $\hat{Y}_q \in \mathbb{R}^{T\times(N+1)}$ represents predicted segmentation logits and f_θ is our EDS-Net with parameters θ.

Zero-shot Setting. Zero-shot segmentation extends the framework to handle completely unseen categories without any visual examples during training. The key difference lies in replacing visual support with semantic information. During training, we have access to seen classes $\mathcal{C}_{seen}$ with both visual point clouds and semantic embeddings $\mathcal{E}_{seen} = \{e_n \in \mathbb{R}^{D_s}\}_{n=1}^{N}$, where D_s is the semantic embedding dimension. During testing, only semantic embeddings $\mathcal{E}_{unseen}$ are available for unseen classes. The zero-shot objective becomes:

$$\hat{Y}_q = f_\theta(\mathcal{E}_{unseen}, I_q), \tag{2}$$

where the model must bridge the gap between semantic descriptions and visual point cloud features.

Unified Evolutionary Framework. Our key insight is to model both tasks as evolutionary optimization problems where prototypes undergo multi-generational refinement. We define the evolutionary process as:

$$V^{(g+1)} = \mathcal{F}_{evo}(V^{(g)}, \mathcal{S}, \mathcal{Q}; \theta^{(g)}), \tag{3}$$

where $V^{(g)} \in \mathbb{R}^{(N+1)\times C}$ represents prototype features at generation g with C feature channels, $\mathcal{F}_{evo}$ denotes the evolutionary operator (SEM for few-shot, VSBM+SEM for zero-shot), and $\theta^{(g)}$ represents generation-specific parameters.

This biological metaphor naturally unifies both scenarios: few-shot learning resembles symbiotic adaptation between closely related species (support and query from same domain), while zero-shot learning mimics cross-species genetic exchange (visual and semantic modalities). The final

segmentation prediction follows:

$$p(y_i = c|I_q) = \frac{\exp(\phi_q(p_i) \cdot V_c^{final}/\tau)}{\sum_{j=0}^{N} \exp(\phi_q(p_i) \cdot V_j^{final}/\tau)}, \quad (4)$$

where $\phi_q(p_i)$ is the feature of query point p_i, V_c^{final} is the final evolved prototype for class c, and τ is a temperature parameter controlling the sharpness of the probability distribution.

3.2 Biological Theoretical Foundation

Our method draws inspiration from biological mechanisms of natural evolution and genetic crossover, providing a robust framework for tackling domain gaps and modality fusion in point cloud segmentation.

Symbiotic Evolution Theory. Symbiotic evolution involves co-adaptive processes where species mutually shape each other's evolutionary paths through beneficial interactions (Wade 2007). In mutualistic symbiosis, species develop complementary traits that enhance survival and reproduction, as seen in the co-evolution of bees and flowering plants (Anderson 2015). In our approach, support and query features, akin to distinct "species," face domain gaps due to varying acquisition conditions or distributional shifts. Our Symbiotic Evolution Module (SEM) facilitates co-adaptation through self-enhancement (intra-species) and interactive enhancement (inter-species) mechanisms.

Genetic Crossover Mechanisms. Genetic crossover in sexual reproduction exchanges genetic material between parents to produce offspring with superior traits (Crow 1994). This process combines beneficial characteristics while preserving diversity through recombination, enhancing adaptability (Tatar et al. 2024). In zero-shot point cloud segmentation, visual prototypes from geometric features and semantic prototypes from textual descriptions act as distinct "genetic materials." Our Vision-Semantic Bridging Module (VSBM) implements adaptive crossover to combine these features based on compatibility, incorporating mutation strategies to explore novel combinations while retaining essential traits. This enables effective knowledge transfer from seen to unseen categories.

3.3 Symbiotic Evolution Module

Inspired by mutualistic symbiosis in biological systems, where species co-adapt through mutually beneficial interactions, we propose the Symbiotic Evolution Module (SEM) to model the co-adaptation between support and query features. This approach addresses the domain gap challenge that arises when support and query sets, despite sharing categorical labels, exhibit distributional differences due to varying acquisition conditions or environmental factors. SEM facilitates joint evolution of these feature representations, enabling mutual enhancement of their discriminative capabilities.

Initial Prototype Extraction. The evolutionary process begins with prototype initialization from the support set. For each class n, we compute the mean support feature vector using the provided masks:

$$V_n^{(0)} = \frac{1}{|\mathcal{M}n|} \sum (i,j) \in \mathcal{M}_n F_s^{i,j}, \quad (5)$$

where $\mathcal{M}_n = (i,j)|M_s^{i,j} = 1, \text{class}(i) = n$ represents the set of foreground point indices for class n, and $F_s^{i,j}$ denotes the feature vector of point j in support sample i. The background prototype $V_0^{(0)}$ is computed analogously using background points. This initialization yields the prototype set $V^{(0)} = [V_0^{(0)}, V_1^{(0)}, ..., V_N^{(0)}] \in \mathbb{R}^{(N+1)\times C}$, where C is the feature dimension.

Self-Enhancement Mechanism. The self-enhancement mechanism facilitates intra-modality adaptation by improving feature coherence within each modality. We first extract global representations through max-pooling:

$$F_q^{\text{global}} = \text{MaxPool}(F_q), \quad F_s^{\text{global}} = \text{MaxPool}(F_s), \quad (6)$$

where $F_q, F_s \in \mathbb{R}^{N_p \times C}$ represent query and support features, and $F_q^{\text{global}}, F_s^{\text{global}} \in \mathbb{R}^{D\times C}$ capture their global patterns.

Intra-modality correlations are then modeled through auto-correlation matrices:

$$G_q = (F_q^{\text{global}})^T F_q^{\text{global}}, \quad G_s = (F_s^{\text{global}})^T F_s^{\text{global}}, \quad (7)$$

where $G_q, G_s \in \mathbb{R}^{C\times C}$ encode channel-wise relationships. These are transformed into attention weights:

$$A_q = \sigma(U_q G_q), \quad A_s = \sigma(U_s G_s), \quad (8)$$

with learnable parameters $U_q, U_s \in \mathbb{R}^{C\times C}$ and sigmoid activation σ. The self-enhanced prototypes emerge as:

$$V_{\text{self}} = \phi_1(A_q \odot V^{(0)}) + \phi_1(A_s \odot V^{(0)}), \quad (9)$$

where ϕ_1 denotes a linear transformation and $\odot$ represents element-wise multiplication.

Interactive Enhancement Mechanism. The interactive enhancement mechanism implements cross-modality cooperation by modeling inter-set relationships. We compute normalized cross-correlation:

$$\mathcal{C}_{cross} = (F_q^{global})^T F_s^{global}/\sqrt{D_{feat}}, \quad (10)$$

where $\mathcal{C}_{cross} \in \mathbb{R}^{D_{feat}\times D_{feat}}$ captures cross-modal correlations with normalization factor $\sqrt{D_{feat}}$, and D_{feat} represents the feature dimension. The interactive attention weights are then computed as:

$$A_{\text{inter}} = \phi_2(\text{softmax}(C) \cdot V^{(0)}), \quad (11)$$

where ϕ_2 represents linear transformation.

Symbiotic Adaptation. The complete symbiotic adaptation combines both mechanisms through residual connections:

$$V^{(g+1)} = V_{\text{self}} + A_{\text{inter}} + V^{(g)}, \tag{12}$$

where $V^{(g+1)}$ represents the evolved prototypes at generation g+1. This formulation embodies the core symbiotic principle: both self-enhancement and interactive enhancement contribute to prototype evolution while preserving essential characteristics through residual connections. Iterative application of SEM across multiple generations enables progressive refinement.

3.4 Vision-Semantic Bridging Module

Inspired by genetic recombination in biological systems (Tatar et al. 2024), we propose the Vision-Semantic Bridging Module (VSBM) to address the modality gap in zero-shot 3D segmentation. The module establishes a genetic analogy where visual and semantic prototypes serve as parental genomes, with crossover operations producing hybrid prototypes that combine geometric and semantic understanding for unseen category recognition.

Semantic Projection Network. We first generate pseudo visual prototypes from semantic embeddings using a semantic projection network:

$$P_{fake} = SPN(F_{sem}, z), \tag{13}$$

where F_{sem} represents semantic embeddings obtained by CLIP, z is a random noise vector, and SPN denotes the semantic projection network inspired by PAP3D (He et al. 2023).

Genetic Crossover Operation. Following biological crossover mechanisms, we implement adaptive crossover between visual prototypes V_{visual} and semantic prototypes P_{fake}:

$$W_{cross} = Att([V_{visual}; P_{fake}]), \tag{14}$$

where the Att() generates crossover weights based on the concatenated features and V_{visual} is $V^{(0)}$.

The crossover operation produces offspring prototypes:

$$V_{crossed} = W_{cross} \odot V_{visual} + (1 - W_{cross}) \odot P_{fake}, \tag{15}$$

Controlled Mutation Strategy. To introduce beneficial variations, we apply mutation operations:

$$M_{mask} = rand(V_{crossed}) < \mu), \tag{16}$$

$$M_{noise} = \mathcal{N}(0, \sigma^2) \cdot scale, \tag{17}$$

$$V_{mutated} = V_{crossed} + M_{mask} \odot M_{noise}, \tag{18}$$

where μ is the mutation rate, σ controls noise magnitude, and $scale$ is a learnable parameter.

Then, $V_{mutated}$ is input to SEM instead of $V^{(0)}$.

Domain Bridging Loss. To constrain the crossover process and ensure meaningful visual-semantic alignment, we introduce a domain bridging loss inspired by PAP-FZS3D:

$$\mathcal{L}_{bridge} = \mathcal{L}_{GMMN}(V_{visual}, P_{fake}), \tag{19}$$

where $\mathcal{L}_{GMMN}$ is the Gaussian Mixture Model Network loss that minimizes distribution discrepancy between visual and semantic prototypes:

$$\mathcal{L}_{GMMN} = ||\frac{1}{n_v}\sum_{i=1}^{n_v}\phi(v_i) - \frac{1}{n_s}\sum_{j=1}^{n_s}\phi(s_j)||^2, \tag{20}$$

where ϕ represents the feature mapping function, v_i and s_j denote visual and semantic features respectively.

3.5 EDS-Net Overview

As shown in Fig.1, EDS-Net integrates the biologically-inspired SEM and VSBM modules within a unified multi-generational evolutionary optimization framework. The architecture seamlessly handles both few-shot and zero-shot scenarios through adaptive pathway selection.

Multi-generational Evolutionary Optimization. The core innovation lies in modeling prototype refinement as a multi-generational evolutionary process where each generation builds upon the improvements of its predecessors. For generation g, the evolutionary update follows:

$$V^{(g+1)} = \begin{cases} \mathcal{F}_{SEM}(V^{(g)}, \mathcal{S}, \mathcal{Q}) & \text{few-shot} \\ \mathcal{F}_{SEM}(\mathcal{F}_{VSBM}(V^{(g)}, E), \mathcal{S}, \mathcal{Q}) & \text{zero-shot,} \end{cases} \tag{21}$$

where $\mathcal{F}_{SEM}$ and $\mathcal{F}_{VSBM}$ represent the SEM and VSBM operators respectively, and E denotes semantic embeddings for zero-shot scenarios.

To optimally combine information across generations, we employ an adaptive gating network that learns generation-specific importance weights based on query features:

$$\alpha^{(g)} = \text{softmax}(\text{FC}(\text{GAP}(F_q))), \tag{22}$$

where $\text{GAP}(\cdot)$ denotes global average pooling, $\text{FC}(\cdot)$ is a fully connected layer, and $\alpha^{(g)} \in \mathbb{R}^G$ represents the weight distribution over G generations. The final refined prototypes are obtained through weighted aggregation:

$$V_{\text{final}} = \sum_{g=1}^{G} \alpha^{(g)} \cdot V^{(g)}. \tag{23}$$

Training Process. The training procedure differs between few-shot and zero-shot settings to accommodate their distinct requirements and data availability.

Few-shot Training: Episodes are constructed by sampling N classes from seen categories, with K support samples and multiple query samples per class. The model optimizes prototypes using only the SEM module with cross-entropy loss:

$$\mathcal{L}_{\text{few}} = \mathcal{L}_{\text{CE}}(V_{\text{final}}, Y_{\text{query}}), \tag{24}$$

where Y_{query} represents ground-truth query labels.

Table 2: **Few-shot Results (%) on S3DIS.** S_i denotes the split i is used for testing. *Avg* is their average mIoU.

Methods	Param.	2-way						3-way					
		1-shot			5-shot			1-shot			5-shot		
		S_0	S_1	*Avg*	S_0	S_1	*Avg*	S_0	S_1	*Avg*	S_0	S_1	*Avg*
DGCNN (Wang et al. 2019)	0.62 M	36.34	38.79	37.57	56.49	56.99	56.74	30.05	32.19	31.12	46.88	47.57	47.23
ProtoNet (Snell, Swersky, and Zemel 2017)	0.27 M	48.39	49.98	49.19	57.34	63.22	60.28	40.81	45.07	42.94	49.05	53.42	51.24
MPTI (Zhao, Chua, and Lee 2021)	0.29 M	52.27	51.48	51.88	58.93	60.56	59.75	44.27	46.92	45.60	51.74	48.57	50.16
AttMPTI (Zhao, Chua, and Lee 2021)	0.37 M	53.77	55.94	54.86	61.67	67.02	64.35	45.18	49.27	47.23	54.92	56.79	55.86
BFG (Mao et al. 2022)	-	55.60	55.98	55.79	63.71	66.62	65.17	46.18	48.36	47.27	55.05	57.80	56.43
2CBR (Zhu et al. 2023)	0.35 M	55.89	61.99	58.94	63.55	67.51	65.53	46.51	53.91	50.21	55.51	58.07	56.79
PAP3D (He et al. 2023)	2.45 M	59.45	66.08	62.76	65.40	70.30	67.85	48.99	56.57	52.78	61.27	60.81	61.04
Seg-PN(Zhu et al. 2024)	0.24 M	64.84	67.98	66.41	67.63	71.48	69.56	59.11	60.42	59.77	59.48	64.72	62.10
TaylorSeg-PN (Wang et al. 2025b)	0.27 M	67.12	71.11	69.12	70.44	72.23	71.34	60.28	65.70	63.00	61.59	67.06	64.33
DyPolySeg (Wang, Fang, and Tiwari 2025)	2.64 M	72.02	73.82	72.92	75.99	75.32	75.66	64.54	67.93	66.24	65.61	70.22	67.92
EDS-Net	2.82 M	**73.32**	**74.67**	**74.00**	**76.89**	**76.55**	**76.72**	**65.67**	**68.94**	**67.31**	**66.82**	**71.36**	**69.09**
Improvement	-	+1.30	+0.85	+1.08	+0.90	+1.23	+1.06	+1.13	+1.01	+1.07	+1.21	+1.14	+1.17

Table 3: Few-shot Results (%) on ScanNet. S_i denotes the split i is used for testing. *Avg* is their average mIoU.

Methods	Param.	2-way						3-way					
		1-shot			5-shot			1-shot			5-shot		
		S_0	S_1	*Avg*	S_0	S_1	*Avg*	S_0	S_1	*Avg*	S_0	S_1	*Avg*
DGCNN (Wang et al. 2019)	1.43 M	31.55	28.94	30.25	42.71	37.24	39.98	23.99	19.10	21.55	34.93	28.10	31.52
ProtoNet (Snell, Swersky, and Zemel 2017)	0.27 M	33.92	30.95	32.44	45.34	42.01	43.68	28.47	26.13	27.30	37.36	34.98	36.17
MPTI (Zhao, Chua, and Lee 2021)	0.29 M	39.27	36.14	37.71	46.90	43.59	45.25	29.96	27.26	28.61	38.14	34.36	36.25
AttMPTI (Zhao, Chua, and Lee 2021)	0.37 M	42.55	40.83	41.69	54.00	50.32	52.16	35.23	30.72	32.98	46.74	40.80	43.77
BFG (Mao et al. 2022)	-	42.15	40.52	41.34	51.23	49.39	50.31	34.12	31.98	33.05	46.25	41.38	43.82
2CBR (Zhu et al. 2023)	0.35 M	50.73	47.66	49.20	52.35	47.14	49.75	47.00	46.36	46.68	45.06	39.47	42.27
PAP3D (He et al. 2023)	2.45 M	57.08	55.94	56.51	64.55	59.64	62.10	55.27	55.60	55.44	59.02	53.16	56.09
Seg-PN (Zhu et al. 2024)	0.24 M	63.15	64.32	63.74	67.08	69.05	68.07	61.80	65.34	63.57	62.94	68.26	65.60
TaylorSeg-PN (Wang et al. 2025b)	0.27 M	67.52	70.75	69.14	68.39	71.55	69.97	63.60	67.55	65.58	66.98	69.78	68.38
DyPolySeg (Wang, Fang, and Tiwari 2025)	2.64 M	71.05	72.73	71.89	71.25	73.66	72.46	67.65	71.24	69.45	68.73	69.62	69.18
EDS-Net (Ours)	2.82 M	**72.15**	**73.54**	**72.85**	**72.37**	**74.51**	**73.44**	**68.70**	**72.45**	**70.58**	**69.52**	**70.12**	**69.82**
Improvement	-	+1.10	+0.81	+0.96	+1.12	+0.85	+0.98	+1.05	+1.21	+1.13	+0.79	+0.50	+0.64

Zero-shot Training: Training uses seen classes with both visual point clouds and semantic embeddings. The VSBM module first generates crossed prototypes, which are then refined through SEM. The total loss combines cross-entropy and domain bridging components:

$$\mathcal{L}_{\text{zero}} = \mathcal{L}_{\text{CE}}(V_{\text{final}}, Y_{\text{query}}) + \lambda \mathcal{L}_{\text{bridge}}, \tag{25}$$

where λ is a hyperparameter balancing the two loss terms.

4 Experiments

4.1 Datasets and Evaluation Metrics

Datasets: We used two datasets: S3DIS (Armeni et al. 2016) is a dataset of 3D RGB point clouds collected from 272 rooms across 6 indoor environments. Each point is annotated with one of 13 semantic labels (12 semantic classes plus clutter). The ScanNet (Dai et al. 2017) contains a total of 1513 scanned scenes. All points, except for unannotated spaces, are annotated with 20 semantic classes.

Evaluation Metric: We choose mIoU (Mean Intersection over Union), which is widely used in point cloud segmentation, as the performance evaluation metric.

4.2 Implementation Details

We implement EDS-Net using PyTorch framework on an NVIDIA GeForce RTX 4090 GPU. Our network is built upon the DyPolySeg backbone for geometric feature extraction. For the Symbiotic Evolution Module (SEM), we set the local neighborhood size to 16 points using K-NN. The Vision-Semantic Bridging Module (VSBM) employs CLIP text embeddings with dimension 512. We set the mutation rate $\mu = 0.1$ and noise variance $\sigma^2 = 0.01$. During training, we use the AdamW optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$) with an initial learning rate of 0.001, which is halved every 7,000 iterations. The domain bridging loss weight λ is set to 0.5. For multi-generational evolutionary optimization, we set the number of generations $G = 4$. In episodic training, each batch contains 1 episode with one support set and one query set following the N-way K-shot paradigm.

4.3 Comparison with State-of-the-Art Methods

Few-shot Results on S3DIS Dataset. As shown in Table 2, EDS-Net outperforms all baseline methods across different few-shot settings. Compared to the previous state-of-the-art DyPolySeg, our method achieves notable improvements of +1.08% and +1.06% mIoU in 2-way 1-shot and 5-shot scenarios, respectively. In the more challenging 3-way settings, EDS-Net demonstrates robust performance with +1.07% and +1.17% improvements over DyPolySeg. These consistent gains validate the effectiveness of our biologically-inspired evolutionary mechanisms in capturing discriminative features and reducing domain gaps between support and query sets. Notably, EDS-Net achieves these improvements with only a modest parameter increase (2.82M vs 2.64M), demonstrating its efficiency in leveraging evolutionary optimization for few-shot point cloud segmentation.

Few-shot Results on ScanNet Dataset. Table 3 demonstrates the superior performance of EDS-Net on the ScanNet dataset across all few-shot configurations. Compared to the previous state-of-the-art DyPolySeg, our method achieves consistent improvements of +0.96% and +0.98% mIoU in

Table 4: Comparison of zero-shot 3D semantic segmentation results on S3DIS and ScanNet dataset using mean-IoU (%).

Method	S3DIS				ScanNet			
	2-way		3-way		2-way		3-way	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
3DGenZ (Michele et al. 2021)	34.93	36.12	23.08	27.52	29.07	31.65	28.13	28.01
PAP3D (word2vec) (He et al. 2023)	59.98	63.54	48.91	55.62	54.72	58.94	53.78	53.50
PAP3D (CLIP) (He et al. 2023)	61.09	64.91	50.18	59.10	56.12	60.65	55.24	55.04
EDS-Net (Ours)	**73.38**	**72.86**	**66.10**	**63.94**	**69.63**	**68.98**	**69.11**	**65.42**
Improvement	+12.29	+7.95	+15.92	+4.84	+13.51	+8.33	+13.87	+10.38

Table 5: Effect of different modules on S3DIS under 2-way-1-shot settings on the S_0 and S_1 split.

SEM	VSBM	Multi-Gen	S_0	S_1	Avg
✗	✗	✗	72.02	73.82	72.92
✓	✗	✗	72.85	74.31	73.58
✗	✓	✗	72.47	74.12	73.30
✗	✗	✓	72.34	74.05	73.20
✓	✓	✗	73.18	74.55	73.87
✓	✗	✓	73.05	74.42	73.74
✗	✓	✓	72.89	74.28	73.59
✓	✓	✓	**73.32**	**74.67**	**74.00**

2-way 1-shot and 5-shot scenarios, respectively. In the more challenging 3-way settings, EDS-Net shows particularly strong performance in 1-shot scenarios with +1.13% improvement, while maintaining competitive results in 5-shot settings (+0.64% improvement). The consistent gains across both datasets validate the generalization capability of our biologically-inspired evolutionary framework.

Zero-shot Results. Table 4 demonstrates the exceptional performance of EDS-Net in zero-shot task across both datasets. Compared to the previous best method PAP3D (CLIP), our approach achieves substantial improvements ranging from +4.84% to +15.92% mIoU across different settings. Particularly notable are the significant gains in challenging scenarios: +15.92% in S3DIS 3-way 1-shot and +13.87% in ScanNet 3-way 1-shot settings. These remarkable improvements validate the effectiveness of our Vision-Semantic Bridging Module in handling the modality gap between visual features and semantic embeddings.

4.4 Ablation Studies

Effect of Different Modules Table 5 demonstrates the critical contributions of each module in our EDS-Net framework. Starting with the baseline (72.92% mIoU), each proposed module brings consistent improvements: SEM contributes +0.66%, VSBM adds +0.38%, and Multi-Gen optimization provides +0.28%. The combination of SEM and VSBM achieves 73.87% mIoU, validating their complementary effects in handling support-query co-adaptation and visual-semantic bridging respectively. Notably, incorporating all modules together yields the optimal performance of 74.00% mIoU, representing a +1.08% improvement over the baseline and confirming the synergistic effect of our complete biologically-inspired evolutionary architecture.

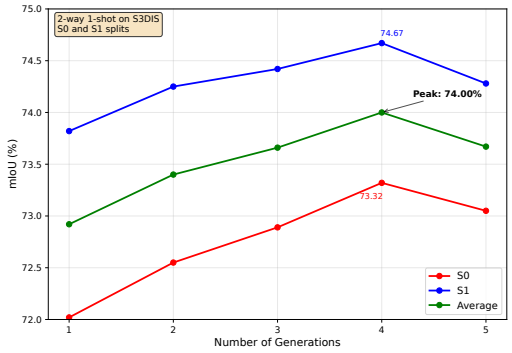

Figure 2: Impact of different generation numbers on S3DIS under 2-way 1-shot settings across S0 and S1 splits.

Impact of Generation Numbers Fig. 2 illustrates the impact of generation numbers in our multi-generational evolutionary optimization framework on S3DIS dataset under 2-way 1-shot settings. Both S0 and S1 splits demonstrate consistent improvement trends, with performance progressively increasing from 1 generation (average: 72.92%) to the optimal point at 4 generations (average: 74.00%). Beyond this point, performance slightly decreases at 5 generations (average: 73.67%), indicating potential overfitting. The consistent pattern across both dataset splits validates the robustness of our approach. This analysis confirms that 4 generations provide the optimal balance between prototype refinement and computational efficiency.

5 Conclusion

In this paper, we propose a novel EDS-Net for few-shot and zero-shot point cloud semantic segmentation. Specifically, Our SEM enables mutual enhancement between support and query features, effectively addressing domain gaps through self-enhancement and interactive enhancement mechanisms. For zero-shot scenarios, our VSBM implements genetic crossover operations that treat visual prototypes and semantic embeddings as "parent" individuals, generating superior "offspring" prototypes through adaptive crossover and controlled mutation strategies. In the future, we will explore additional evolutionary operators inspired by other biological mechanisms and investigate extending the framework to incremental learning scenarios.

6 Acknowledgments

This work was supported in part by the European Union's Horizon 2024 Research and Innovation Programme for the Marie Skłodowska-Curie Actions under Grant No. 101211118. This work was also supported by the UKRI Future Leaders Fellowship [MR/V025333/1] (RoboHike). Shuting He was sponsored by Shanghai Pujiang Programme 24PJD030 and Natural Science Foundation of Shanghai 25ZR1402138.

References

Anderson, B. 2015. Coevolution in mutualisms. *Mutualism*, 107–130.

Armeni, I.; Sener, O.; Zamir, A. R.; Jiang, H.; Brilakis, I.; Fischer, M.; and Savarese, S. 2016. 3d semantic parsing of large-scale indoor spaces. In *CVPR*, 1534–1543.

Bonner, J. T. 1988. *The evolution of complexity by means of natural selection*. Princeton University Press.

Chen, G.; Wang, M.; Yang, Y.; Yu, K.; Yuan, L.; and Yue, Y. 2023. Pointgpt: Auto-regressively generative pre-training from point clouds. *Advances in Neural Information Processing Systems*, 36: 29667–29679.

Chib, P. S.; and Singh, P. 2023. Recent advancements in end-to-end autonomous driving using deep learning: A survey. *IEEE Transactions on Intelligent Vehicles*.

Crow, J. F. 1994. Advantages of sexual reproduction. *Developmental genetics*, 15(3): 205–213.

Dai, A.; Chang, A. X.; Savva, M.; Halber, M.; Funkhouser, T.; and Nießner, M. 2017. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *CVPR*, 5828–5839.

Devagiri, J. S.; Paheding, S.; Niyaz, Q.; Yang, X.; and Smith, S. 2022. Augmented Reality and Artificial Intelligence in industry: Trends, tools, and future challenges. *Expert Systems with Applications*, 207: 118002.

Goel, R.; and Gupta, P. 2020. Robotics and industry 4.0. *A Roadmap to Industry 4.0: Smart Production, Sharp Business and Sustainable Development*, 157–169.

He, S.; Jiang, X.; Jiang, W.; and Ding, H. 2023. Prototype adaption and projection for few-and zero-shot 3d point cloud semantic segmentation. *TIP*, 32: 3199–3211.

Lai, X.; Liu, J.; Jiang, L.; Wang, L.; Zhao, H.; Liu, S.; Qi, X.; and Jia, J. 2022. Stratified transformer for 3d point cloud segmentation. In *CVPR*, 8500–8509.

Liang, D.; Zhou, X.; Xu, W.; Zhu, X.; Zou, Z.; Ye, X.; Tan, X.; and Bai, X. 2025. Pointmamba: A simple state space model for point cloud analysis. *NeurIPS*, 37: 32653–32677.

Mao, Y.; Guo, Z.; Xiaonan, L.; Yuan, Z.; and Guo, H. 2022. Bidirectional feature globalization for few-shot semantic segmentation of 3d point cloud scenes. In *2022 International Conference on 3D Vision (3DV)*, 505–514. IEEE.

Michele, B.; Boulch, A.; Puy, G.; Bucher, M.; and Marlet, R. 2021. Generative zero-shot learning for semantic segmentation of 3d point clouds. In *2021 International Conference on 3D vision (3DV)*, 992–1002. IEEE.

Pang, Y.; Wang, W.; Tay, F. E.; Liu, W.; Tian, Y.; and Yuan, L. 2022. Masked autoencoders for point cloud self-supervised learning. In *European conference on computer vision*, 604–621. Springer.

Park, C.; Jeong, Y.; Cho, M.; and Park, J. 2022. Fast point transformer. In *CVPR*, 16949–16958.

Qi, C. R.; Su, H.; Mo, K.; and Guibas, L. J. 2017a. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *CVPR*, 652–660.

Qi, C. R.; Yi, L.; Su, H.; and Guibas, L. J. 2017b. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *NIPS*, 30.

Sereno, M.; Wang, X.; Besançon, L.; Mcguffin, M. J.; and Isenberg, T. 2020. Collaborative work in augmented reality: A survey. *TVCG*, 28(6): 2530–2549.

Snell, J.; Swersky, K.; and Zemel, R. 2017. Prototypical networks for few-shot learning. *NIPS*, 30.

Soori, M.; Arezoo, B.; and Dastres, R. 2023. Artificial intelligence, machine learning and deep learning in advanced robotics, a review. *Cognitive Robotics*, 3: 54–70.

Tatar, A.; Fat, N.; Petrovan, A.; and Matei, O. 2024. A New Vision of Social Behavior on Genetic Algorithm Performance. In *International Conference on Soft Computing Models in Industrial and Environmental Applications*, 241–250. Springer.

Thomas, H.; Qi, C. R.; Deschaud, J.-E.; Marcotegui, B.; Goulette, F.; and Guibas, L. J. 2019. Kpconv: Flexible and deformable convolution for point clouds. In *ICCV*, 6411–6420.

Wade, M. J. 2007. The co-evolutionary genetics of ecological communities. *Nature Reviews Genetics*, 8(3): 185–195.

Wang, C.; Fang, X.; and Tiwari, P. 2025. DyPolySeg: Taylor Series-Inspired Dynamic Polynomial Fitting Network for Few-shot Point Cloud Semantic Segmentation. In *ICML*.

Wang, C.; He, S.; Fang, X.; Han, J.; Liu, Z.; Ning, X.; Li, W.; and Tiwari, P. 2025a. Point Clouds Meets Physics: Dynamic Acoustic Field Fitting Network for Point Cloud Understanding. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 22182–22192.

Wang, C.; He, S.; Fang, X.; Wu, M.; Lam, S.-K.; and Tiwari, P. 2025b. Taylor series-inspired local structure fitting network for few-shot point cloud semantic segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 7527–7535.

Wang, C.; Wu, M.; Lam, S.-K.; Ning, X.; Yu, S.; Wang, R.; Li, W.; and Srikanthan, T. 2024. Gpsformer: A global perception and local structure fitting-based transformer for point cloud understanding. In *European conference on computer vision*, 75–92. Springer.

Wang, Y.; Sun, Y.; Liu, Z.; Sarma, S. E.; Bronstein, M. M.; and Solomon, J. M. 2019. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (tog)*, 38(5): 1–12.

Weinmann, M.; Jutzi, B.; Hinz, S.; and Mallet, C. 2015. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS*

Journal of Photogrammetry and Remote Sensing, 105: 286–304.

Wierstra, D.; Schaul, T.; Glasmachers, T.; Sun, Y.; Peters, J.; and Schmidhuber, J. 2014. Natural evolution strategies. *The Journal of Machine Learning Research*, 15(1): 949–980.

Wu, X.; Lao, Y.; Jiang, L.; Liu, X.; and Zhao, H. 2022. Point transformer v2: Grouped vector attention and partition-based pooling. *NIPS*, 35: 33330–33342.

Xu, Y.; Zhao, N.; and Lee, G. H. 2023. Towards Robust Few-shot Point Cloud Semantic Segmentation. *arXiv preprint arXiv:2309.11228*.

Zhang, C.; Wu, Z.; Wu, X.; Zhao, Z.; and Wang, S. 2023a. Few-shot 3d point cloud semantic segmentation via stratified class-specific attention based transformer network. In *AAAI*, volume 37, 3410–3417.

Zhang, H.; Wang, C.; Tian, S.; Lu, B.; Zhang, L.; Ning, X.; and Bai, X. 2023b. Deep learning-based 3D point cloud classification: A systematic survey and outlook. *Displays*, 79: 102456.

Zhao, H.; Jiang, L.; Jia, J.; Torr, P. H.; and Koltun, V. 2021. Point transformer. In *ICCV*, 16259–16268.

Zhao, J.; Zhao, W.; Deng, B.; Wang, Z.; Zhang, F.; Zheng, W.; Cao, W.; Nan, J.; Lian, Y.; and Burke, A. F. 2023. Autonomous driving system: A comprehensive survey. *Expert Systems with Applications*, 122836.

Zhao, N.; Chua, T.-S.; and Lee, G. H. 2021. Few-shot 3d point cloud semantic segmentation. In *CVPR*, 8873–8882.

Zhu, G.; Zhou, Y.; Yao, R.; and Zhu, H. 2023. Cross-class bias rectification for point cloud few-shot segmentation. *TMM*, 25: 9175–9188.

Zhu, X.; Zhang, R.; He, B.; Guo, Z.; Liu, J.; Xiao, H.; Fu, C.; Dong, H.; and Gao, P. 2024. No Time to Train: Empowering Non-Parametric Networks for Few-shot 3D Scene Segmentation. In *CVPR*, 3838–3847.