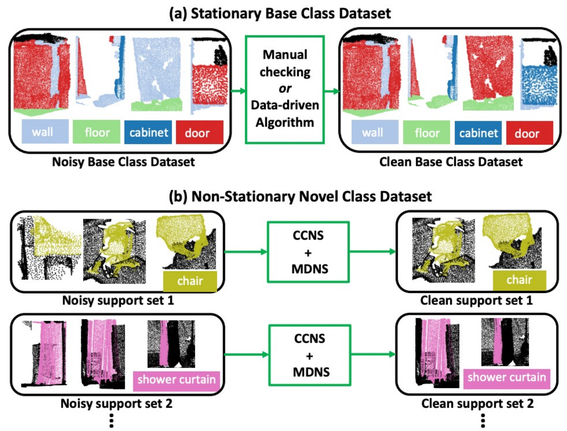
Towards Robust Few-shot Point Cloud Semantic Segmentation

Yating Xu[1], Na Zhao[2], Gim Hee Lee[1]
[1]National University of Singapore , [2]Singapore University of Technology and Design

BMVC 2023

1. Task Motivation



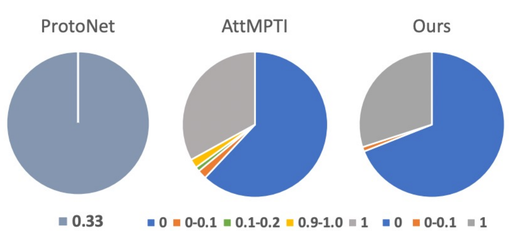
Comparison between noisy base and novel class dataset of few-shot point cloud semantic segmentation.

- Base class dataset: We can clean the label noise by either manual checking or data-driven algorithm given enough time and budget.
- Novel class dataset: It is impossible to manually clean the label noise for the *infinite novel classes*. Neither can we adopt data-driven algorithms since it would *overfit to the small size of the support set*. An example where a sofa and a curtain are wrongly annotated in support set 1 and 2, respectively.

2. Technique Contributions

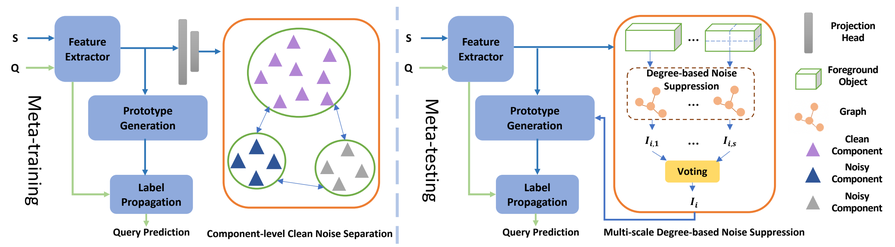
- We propose a **component-level clean noise separation** method for representation learning to enhance the class-level discrimination in the embedding space.
- We propose a **multi-scale degree-based noise suppression** scheme that is able to effectively remove noisy samples from the small support set for each new class during testing.

3. Why Choose AttMPTI[1]



- Comparison of prototype cleanness of different methods on a noisy support set. '1' means the prototype only containing clean-labeled points, and '0' means the prototype only containing points that are incorrectly labeled as the target class. Values in between 0-1 represent the portion of clean labeled points in the prototype.
- The robustness of AttMPTI is subjected to discriminativity of the feature embeddings.

4. Overall Framework



- **Component-level Clean Noise separation (CCNS)**: CCNS conducts feature component-level contrastive learning for each episode class. z_k^i is the anchor and is the *i*-th feature component of *k*-th support point cloud. z_g^j is a positive sample with the same semantic class as z_k^i.

$$\mathcal{L}_{\text{CCNS}} = \frac{1}{KR}\sum_{k=1}^{K}\sum_{i=1}^{R}\left(\frac{-1}{|A(z_k^i)|}\sum_{z_g^j \in A(z_k^i)}\log\frac{\exp\left(z_k^i \cdot z_g^j/\tau\right)}{\sum_{h,b\setminus(k,i)}\exp\left(z_k^i \cdot z_h^b/\tau\right)}\right)$$

- **Multi-scale Degree-based Noise Suppression (MDNS):**
  - We build a fully connected graph G on the K support shots for each way. The weight W_{ij} of the edge encodes the affinity between the two end nodes *i* and *j* as follow:

$$W_{ij} := \begin{cases} [x_i^\top x_j]_+^\gamma, & \text{if } i \neq j \\ 0, & \text{otherwise} \end{cases}$$

  - Degree reflects nodes connection in the graph:

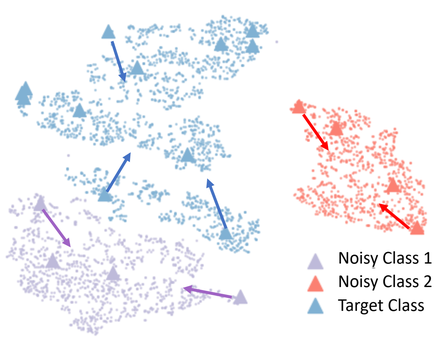
$$d_i = \sum_j W_{ij}$$

  - Compute clean indicator I_i to indicate whether *i*-th shot is clean:

$$I_i := \begin{cases} 1 & \text{if } d_i > thr \\ 0, & \text{otherwise} \end{cases}$$

  - Multi-scale evaluation: to better evaluate point cloud with complex distribution.

5. Visualization of CCNS



Visualization on a 5-shot support set with 2 noisy shots. Each dot represents a point in the feature space and each triangle represents a feature component. Different colors represent different classes with blue indicating the target class. The arrow shows the direction to pull the feature components.

6. Experiments

model	0%	In-episode Noise				Out-episode Noise			
		20%		40%		40%		60%	
	2-way	3-way	2-way	3-way	2-way	3-way	2-way	3-way	2-way	3-way
PNAL [R]	13.67	8.12	8.94	5.45	5.95	3.13	8.08	4.28	4.77	2.87
Tra-NFS [R]	44.98	31.67	43.44	30.68	37.27	27.39	41.72	28.43	35.67	23.20
ProtoNet [R]	57.02	46.78	54.21	43.57	42.57	36.71	50.01	39.31	44.96	36.08
AttMPTI [R]	65.90	51.71	60.01	47.96	38.81	37.56	58.60	44.76	51.18	40.32
Ours	**68.21**	**54.79**	**66.02**	**52.91**	**58.01**	**48.72**	**66.09**	**50.71**	**58.84**	**46.19**

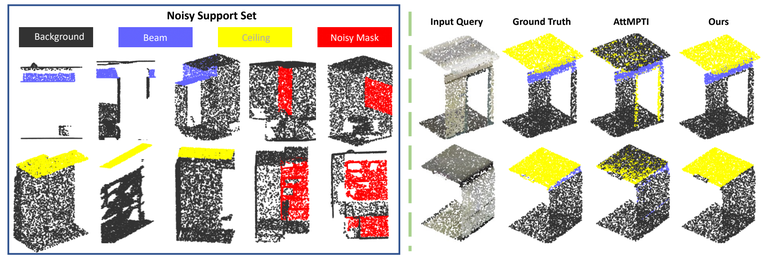
Table 1: Results on the S3DIS using mIoU metric on 2-way 5-shot and 3-way 5-shot.

model	0%	In-episode Noise				Out-episode Noise			
		20%		40%		40%		60%	
	2-way	3-way	2-way	3-way	2-way	3-way	2-way	3-way	2-way	3-way
Tra-NFS [R]	41.89	31.56	39.72	29.20	34.25	25.07	38.42	27.29	34.68	23.78
ProtoNet [R]	47.55	38.97	44.19	36.46	34.57	30.23	42.47	33.88	36.64	28.55
AttMPTI [R]	**54.16**	**44.52**	46.63	38.83	31.57	27.62	43.31	34.33	36.45	26.79
Ours	53.50	43.84	**49.78**	**41.01**	**38.70**	**34.03**	**47.90**	**38.93**	**38.42**	**28.81**

Table 2: Results on the ScanNet using mIoU metric on 2-way 5-shot and 3-way 5-shot.

- **In-episode noise** samples noisy shots from other N-1 classes of the current episode.
- **Out-episode noise** samples noisy shots from outside of the N classes in the C_{novel}.

7. Qualitative Results



Qualitative comparison of a 2-way 5-shot point cloud segmentation with 40% out-episode noise on S3DIS.

8. Ablation Study

model	0%	In-episode Noise		Out-episode Noise	
		20%	40%	40%	60%
AttMPTI	65.90	60.01	38.81	58.60	51.11
AttMPTI+CCNS	**68.50**	63.10	41.75	63.77	56.79
AttMPTI+MDNS	64.80	63.03	52.78	61.73	52.98
Ours	68.21	**66.02**	**58.01**	**66.09**	**58.84**

References

[1] Zhao, Na, Tat-Seng Chua, and Gim Hee Lee. "Few-shot 3d point cloud semantic segmentation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021