AAAI-25 / IAAI-25 / EAAI-25
FEBRUARY 25 – MARCH 4, 2025 | PHILADELPHIA, USA

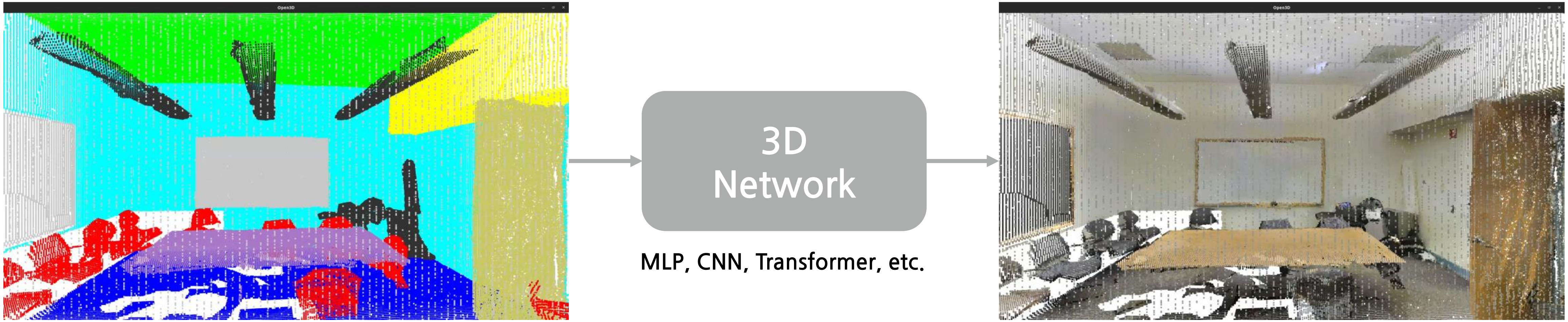
Generalized Zero-Shot Learning for Point Cloud Segmentation with Evidence-Based Dynamic Calibration

Hyeonseok Kim, Byeongkeun Kang, Yeejin Lee

SEOULTECH SEOUL NATIONAL UNIVERSITY OF SCIENCE & TECHNOLOGY

Visual Computing Laboratory

Point Cloud Segmentation

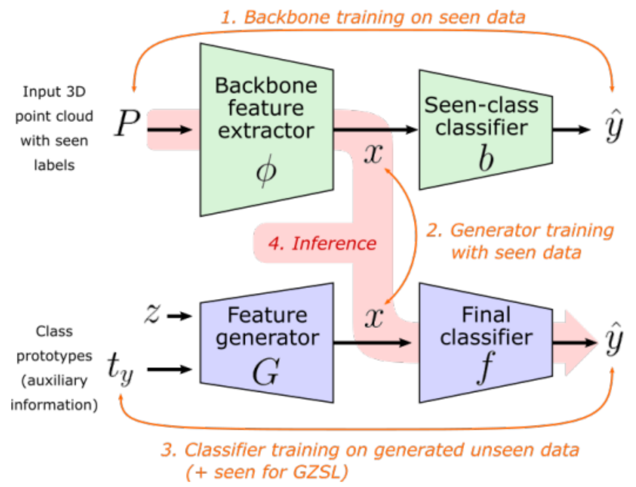


Generalized Zero-Shot Point Cloud Segmentation

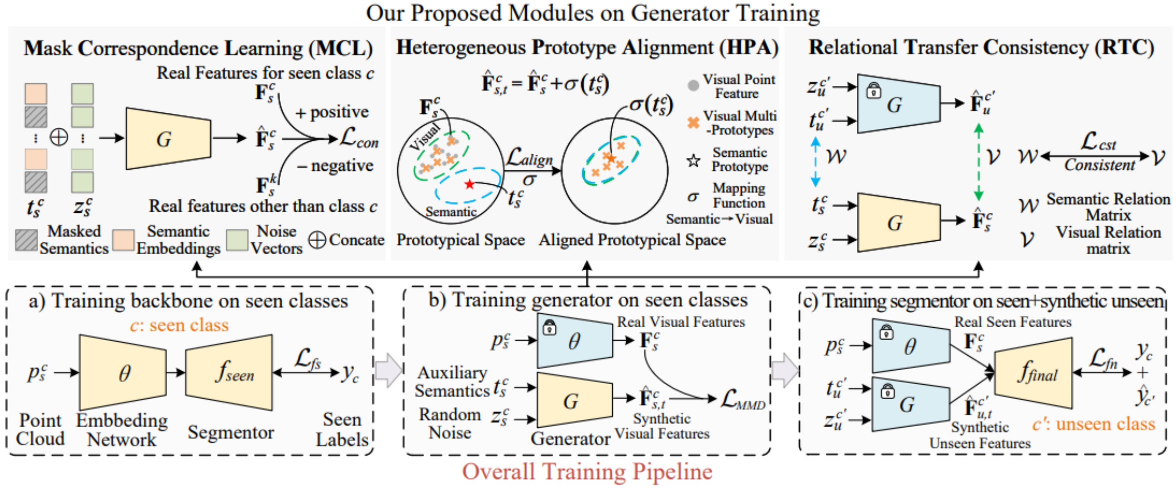
- Zero-Shot Learning(ZSL) vs Generalized Zero-Shot Learning(GZSL)

	ZSL	GZSL ★
Train	Seen	
Test	Unseen	Seen + Unseen

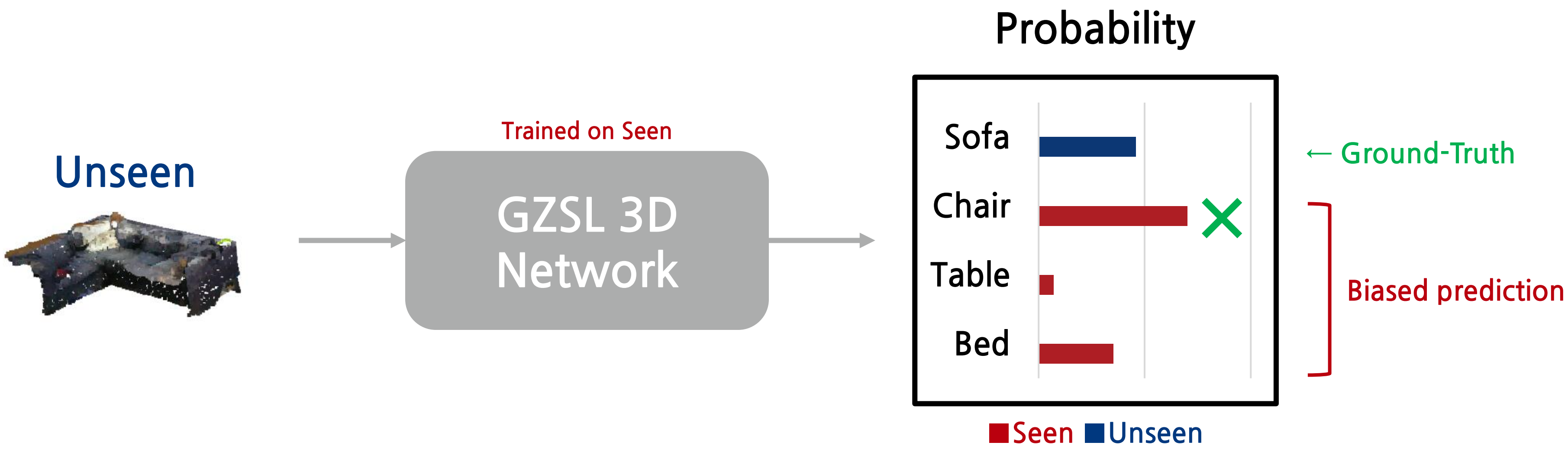
Michele et al. (3DV 2021): 3DGenZ



Yang et al. (ICCV 2023): 3DPC-GZSL



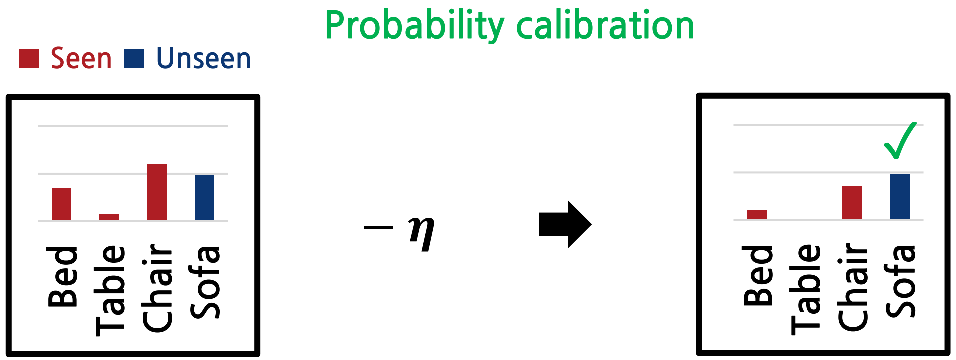
GZSL Issue: Bias of Prediction towards Seen Categories



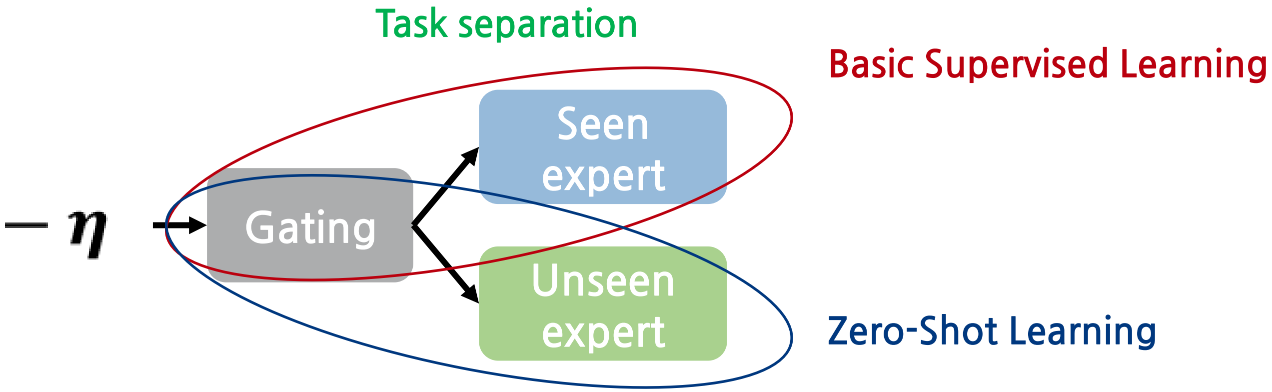
The GZSL models still show bias toward seen categories.

Mitigating Bias of Prediction towards Seen Categories

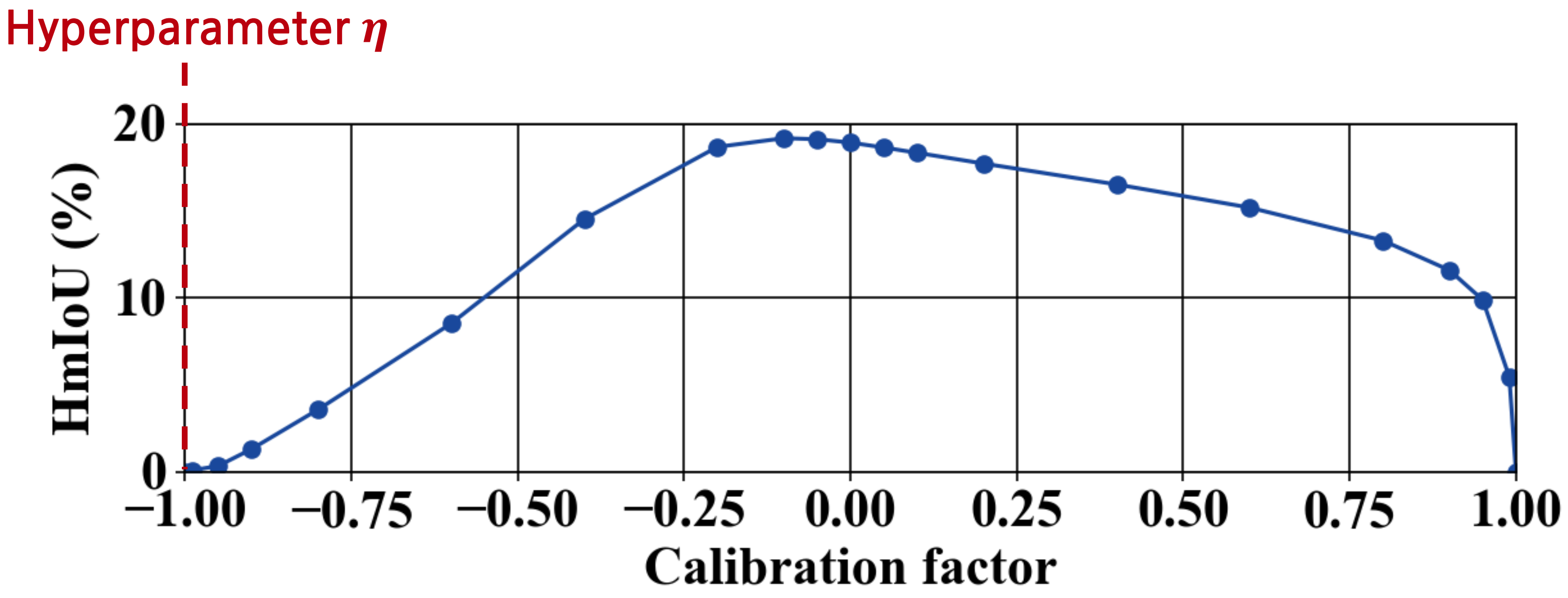
- Calibrated Stacking



- Gating methods



Preliminary Experiments: Calibrated stacking



Strongly dependent on hyperparameter

Comparison of Calibration Methods

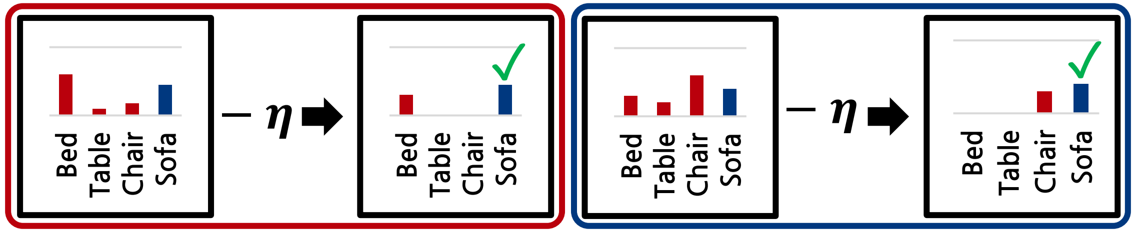
■ Seen-class Probability ■ Unseen-class Probability

Static Calibration

Hyperparameter
η

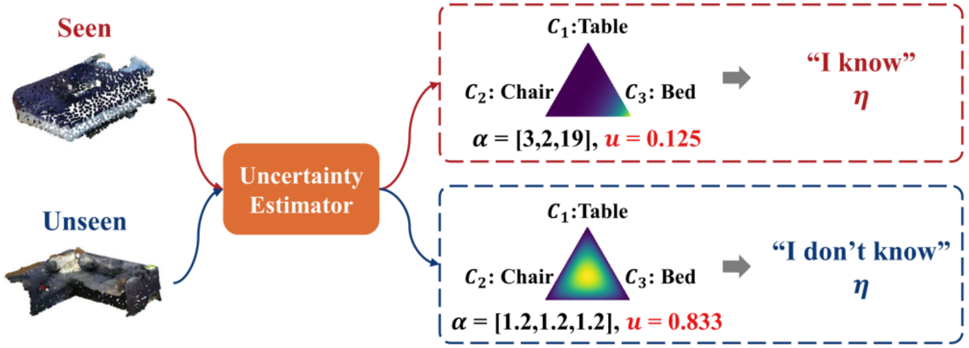
Input: Seen

Input: Unseen



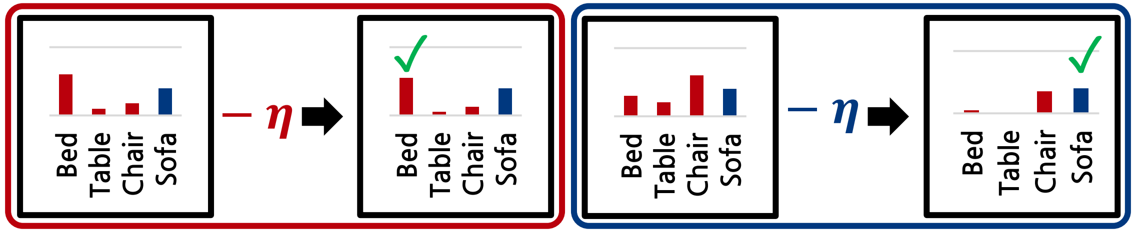
Single calibration factor

Dynamic Calibration(ours)



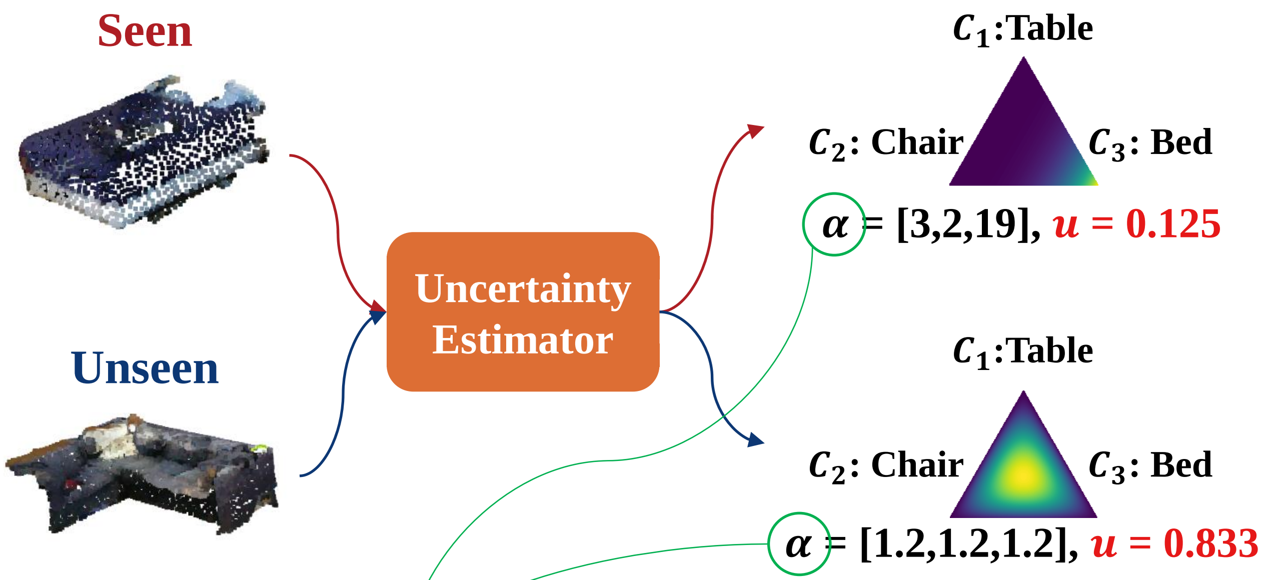
Input: Seen

Input: Unseen



Per-point calibration factor

Uncertainty Estimation



$$\mathrm{Dir}(\mathbf{p}|\boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})}\prod_{i=1}^{K} p_i^{\alpha_i - 1}$$

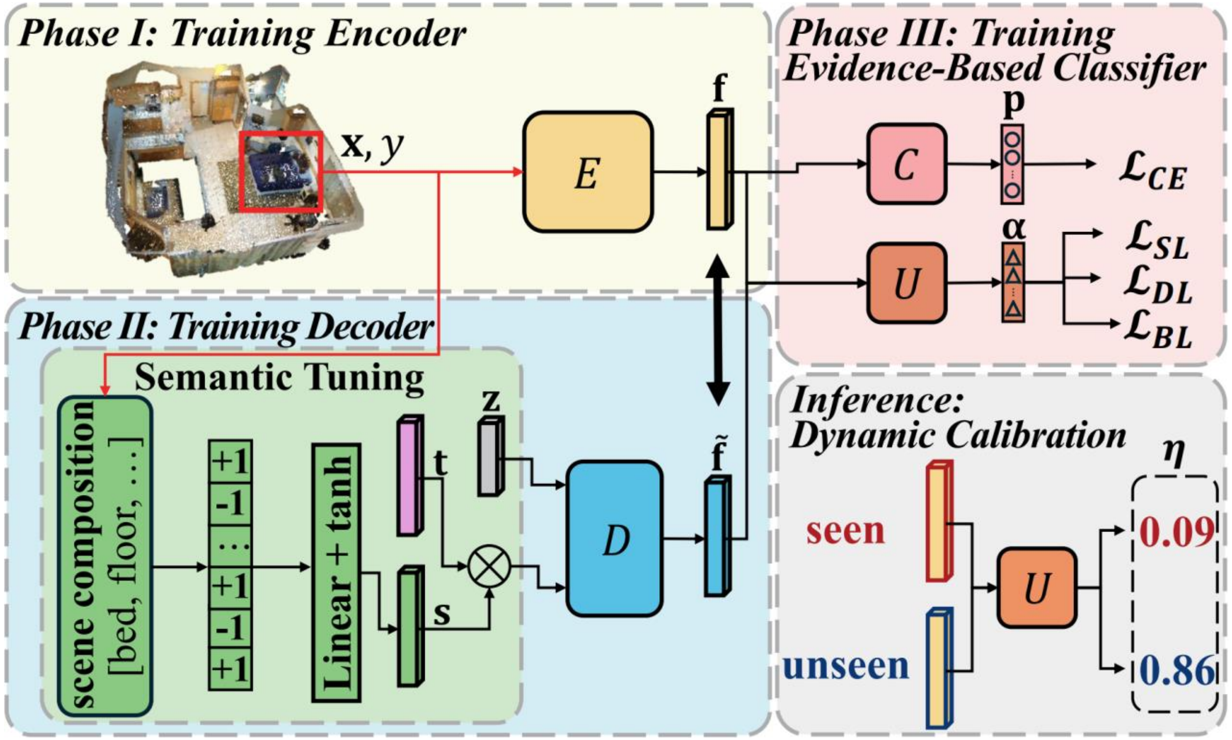
Dirichlet Distribution

- This uncertainty (evidence) can be modeled using evidence theory (Sensoy et al. 2018).

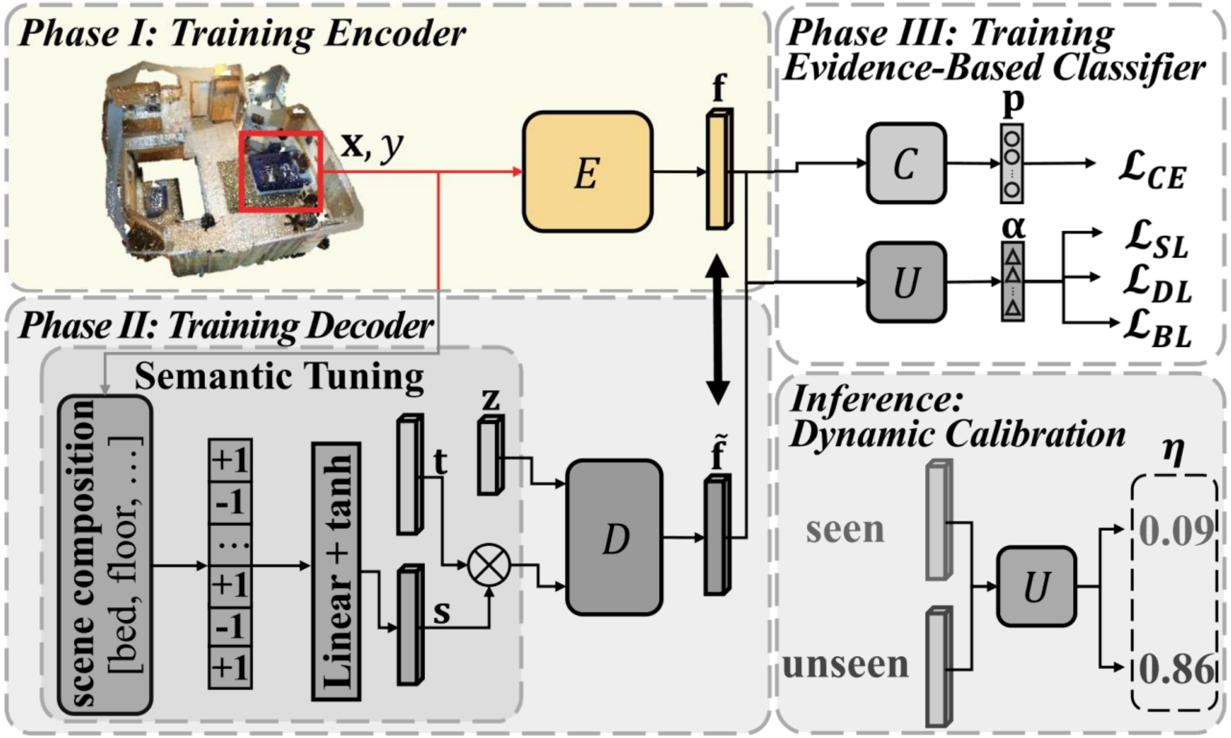
- Uncertainty: $u = \frac{K}{\sum_{k=1}^{K}\alpha_k}$
- This uncertainty is for the seen classes: $K = N_s$.

Overview

- We propose a novel method, E3DPC-GZSL.



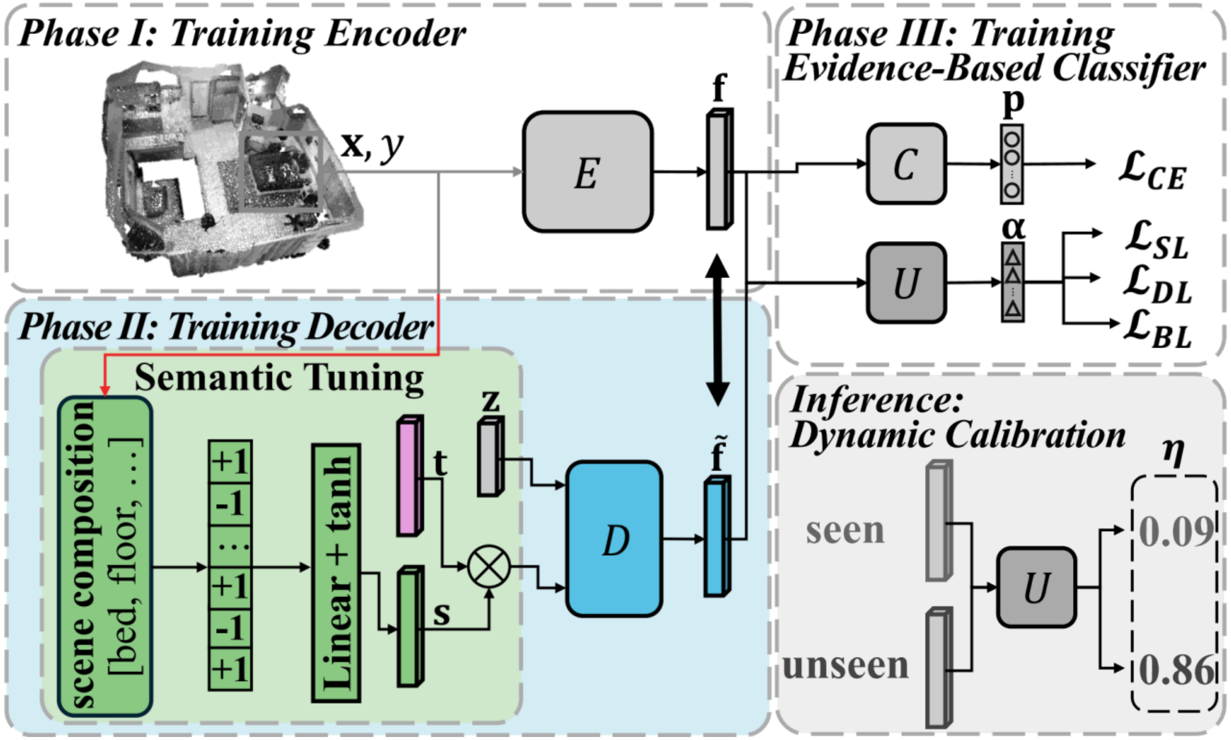
Phase I



- Training the encoder E to extract a vector f from the point cloud $\mathbf{x}$.

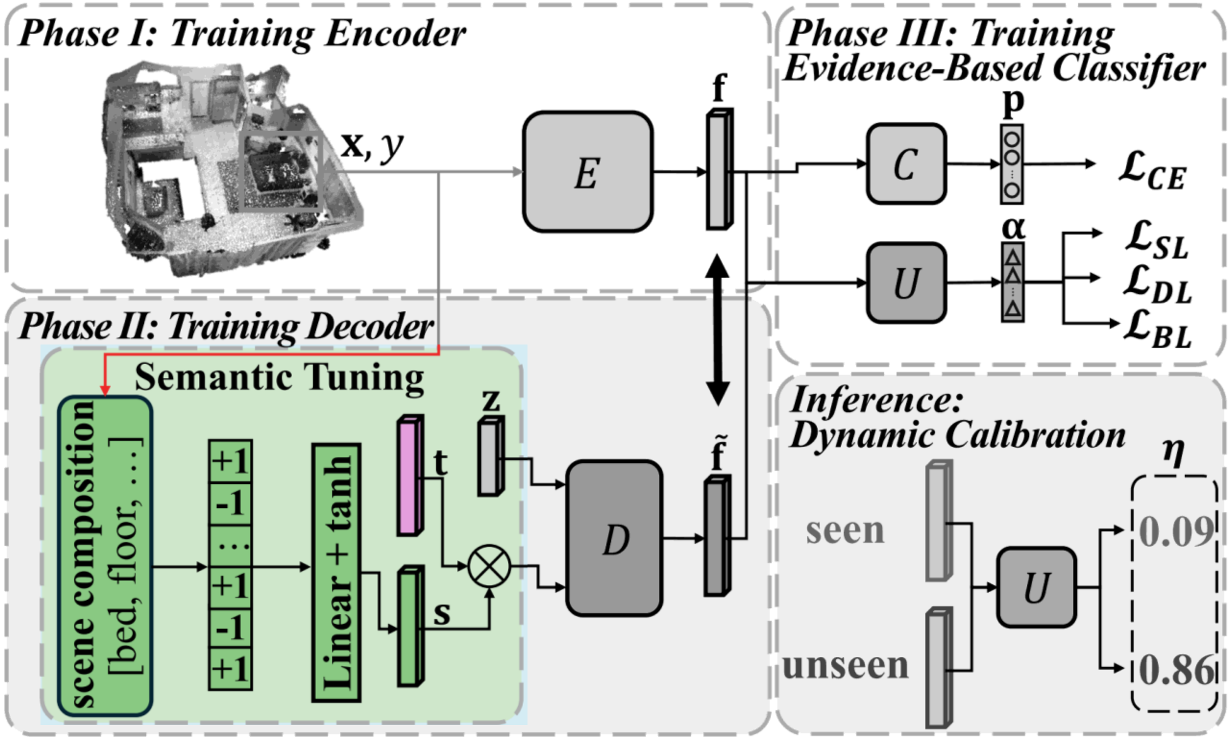
- Learning from scratch using seen data.
- Optimizing the cross-entropy.

Phase II



- The decoder D is trained to generate the feature vectors $\tilde{\mathbf{f}}$ from a random vector $\mathbf{z} \in \mathbb{R}^{N_z}$.
- Condition: text embedding $\mathbf{t} \in \mathbb{R}^{N_t}$.

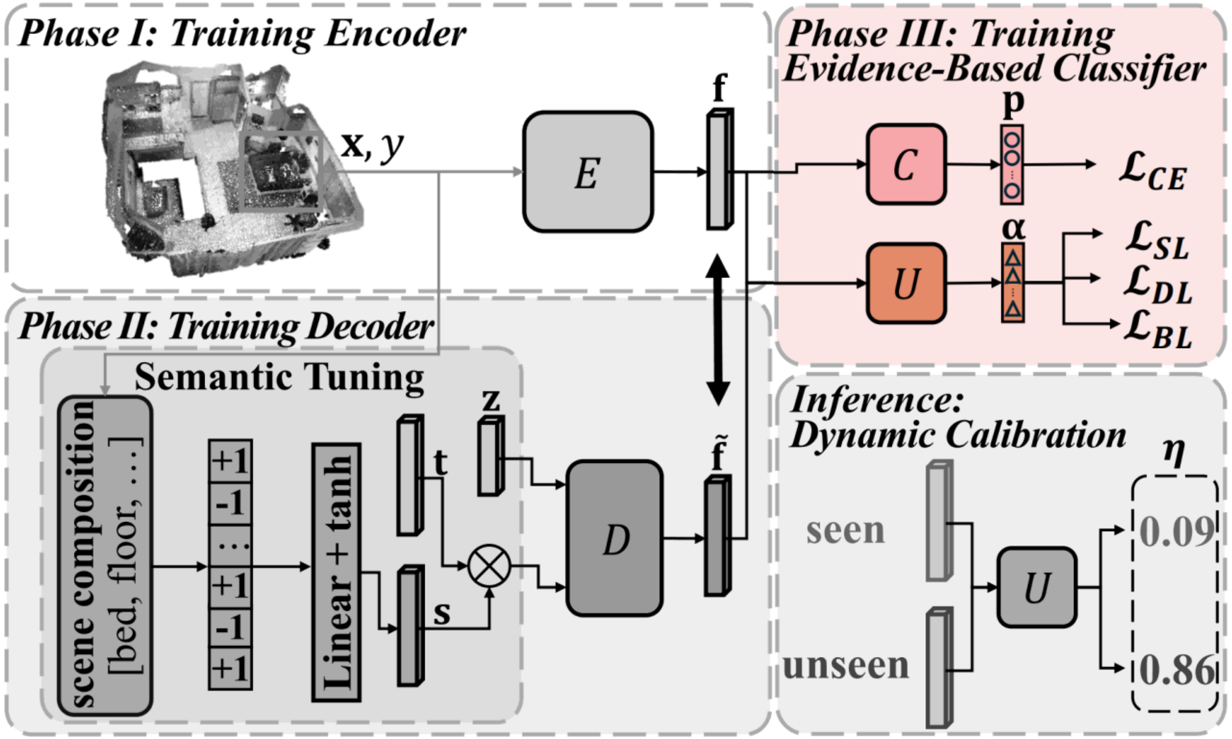
Phase II



- Semantic tuning improves synthesized features by using global scene and local point conditions.

- Tuning vector $\mathbf{s} \in \mathbb{R}^{N_t}$: encode the categories present in the point cloud.
- Training strategy: 3DPC-GZSL(Yang et al. 2023)

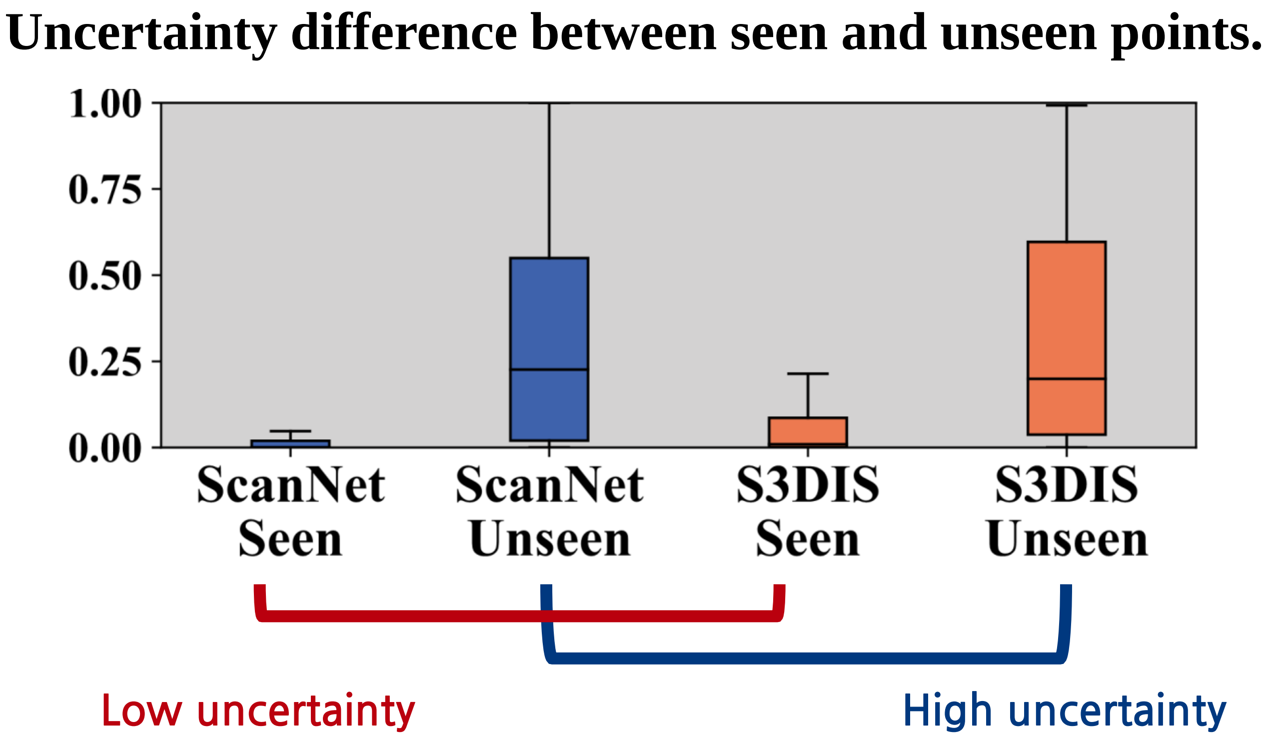
Phase III



- The object is to train the classifier C and uncertainty estimator U.

- C: assign class probabilities to points.
- U: obtain the uncertainty of a point.

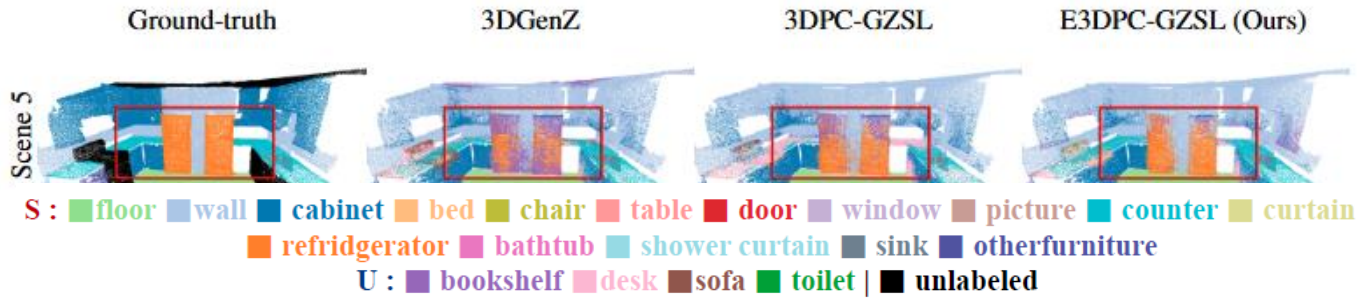
Inference: uncertainty estimation



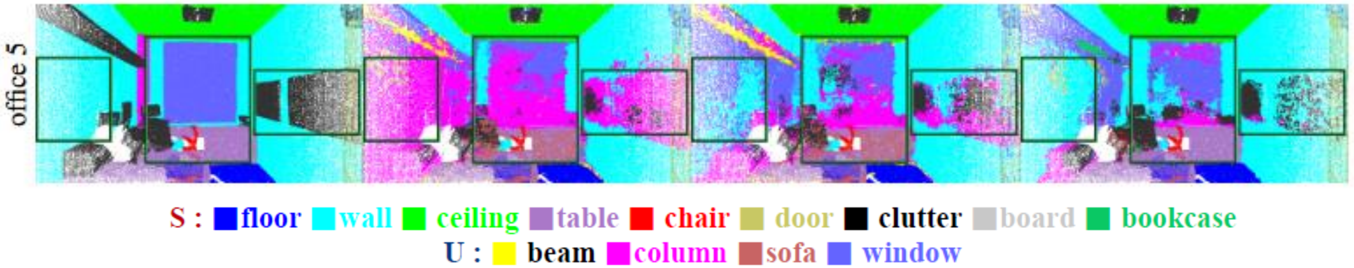
Evidence-Based Dynamic Calibration is performed based on the uncertainty of each input point.

Experiments: Results

	Training set		ScanNet v2				S3DIS			
	Encoder	Classifier	mIoU Seen	mIoU Unseen	mIoU All	HmIoU	mIoU Seen	mIoU Unseen	mIoU All	HmIoU
Full supervision	$\mathcal{Y}^s \cup \mathcal{Y}^u$	$\mathcal{Y}^s \cup \mathcal{Y}^u$	43.3	51.9	45.1	47.2	74.0	50.0	66.6	59.6
Full supervision only for classifier	$\mathcal{Y}^s$	$\mathcal{Y}^s \cup \mathcal{Y}^u$	41.5	39.2	40.3	40.3	60.9	21.5	48.7	31.8
Supervision with seen	$\mathcal{Y}^s$	$\mathcal{Y}^s$	39.0	0.0	31.3	0.0	70.2	0.0	48.6	0.0
3DGenZ (Michele et al. 2021)	$\mathcal{Y}^s$	$\mathcal{Y}^s \cup \mathcal{Y}^{\tilde{u}}$	32.8	7.7	27.8	12.5	53.1	7.3	39.0	12.9
3DPC-GZSL (Yang et al. 2023a)	$\mathcal{Y}^s$	$\mathcal{Y}^s \cup \mathcal{Y}^{\tilde{u}}$	34.5	14.3	30.4	20.2	58.9	9.7	43.8	16.7
E3DPC-GZSL (ours)	$\mathcal{Y}^s$	$\mathcal{Y}^s \cup \mathcal{Y}^{\tilde{u}}$	**36.1**	**15.4**	**32.0**	**21.6**	**67.9**	**12.0**	**50.7**	**20.4**
			+1.6	+1.1	+1.6	+1.4	+9.0	+2.3	+6.9	+3.7



ScanNet v2

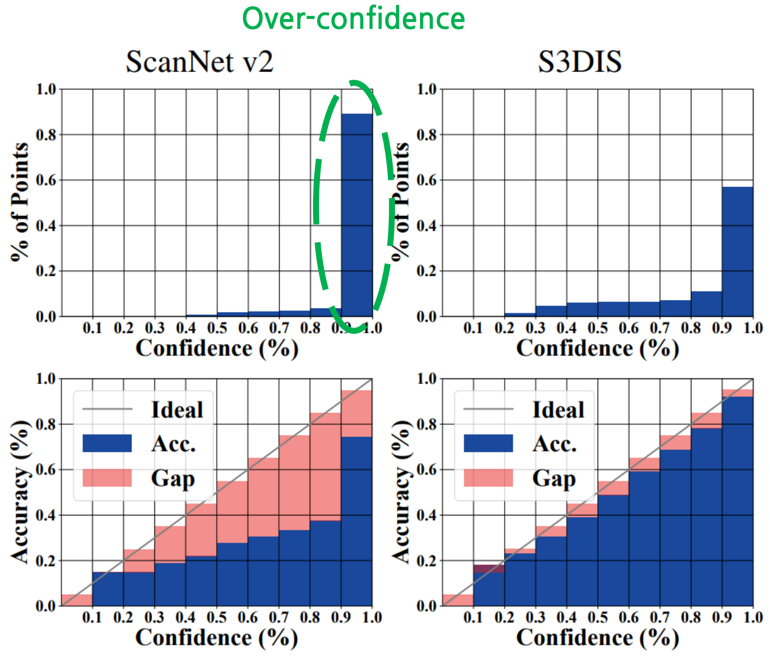


S3DIS

Experiments: Ablation Study

Dataset	B	S	U	mIoU Seen	mIoU Unseen	mIoU All	HmIoU
ScanNet v2	✓	–	–	34.78	14.80	30.79	20.77
	✓	–	✓	34.86	14.87	30.87	20.85
	✓	✓	–	36.09	15.23	31.91	21.42
	✓	✓	✓	**36.11**	**15.40**	**31.97**	**21.59**
S3DIS	✓	–	–	65.03	10.17	48.15	17.60
	✓	–	✓	66.58	11.27	49.56	19.28
	✓	✓	–	66.46	11.02	49.40	18.90
	✓	✓	✓	**67.90**	**12.01**	**50.70**	**20.42**

- B: Baseline model without calibrated stacking
- S: Semantic tuning
- U: Evidence-based dynamic calibration



- This over-confidence issue makes it difficult to calibrate the prediction probabilities.

Summary

- We propose E3DPC-GZSL, a novel approach for generalized zero-shot point cloud semantic segmentation.
- Our method exploits the uncertainty of input points to dynamically calibrate classifier predictions, which mitigates the bias of zero-shot models towards seen classes and improves generalization performance.
- To address the issue of data scarcity, we introduce a novel training strategy that refines the semantic space by applying semantic tuning to text embeddings.
- Regularizing the model's overconfidence issue could improve performance in generalized zero-shot settings.
- We consider this as a direction for future research.

Thank You!