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

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

Hyeonseok Kim, Byeongkeun Kang, Yeejin Lee*
Seoul National University of Science and Technology, Republic of Korea
{khslab, byeongkeun.kang, yeejinlee}@seoultech.ac.kr

SEOULTECH SEOUL NATIONAL UNIVERSITY OF SCIENCE & TECHNOLOGY

Concepts

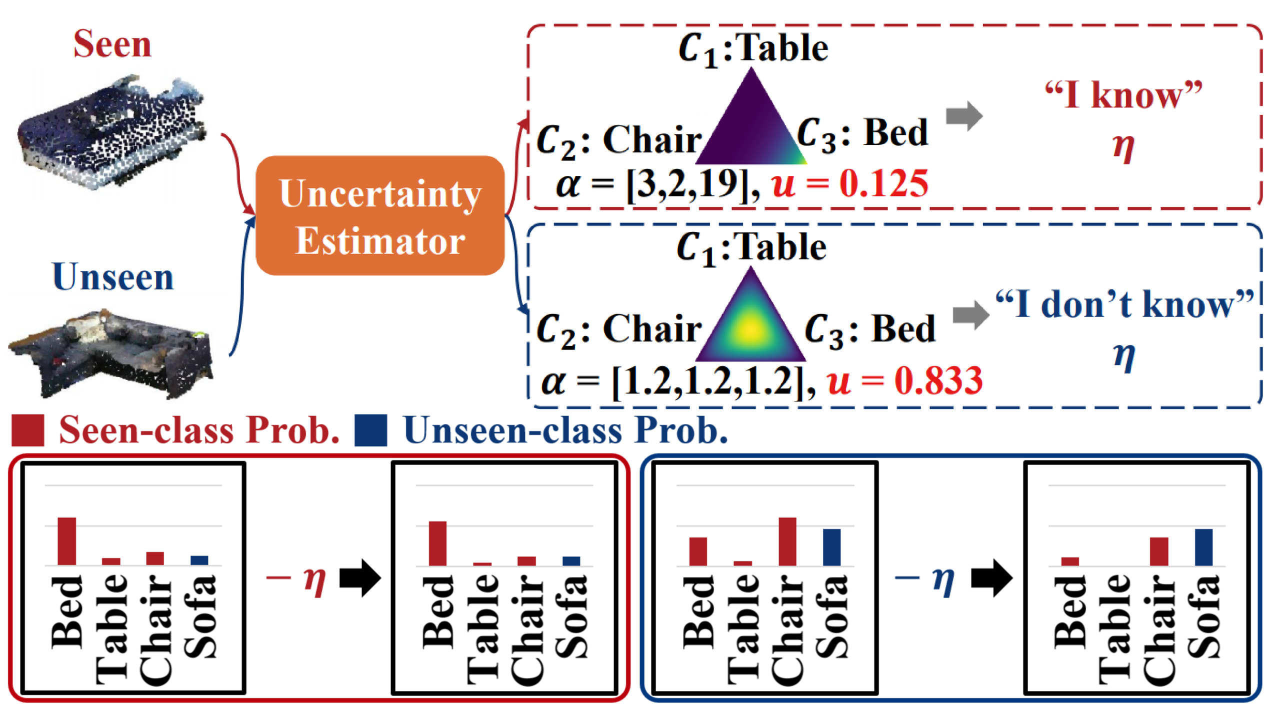

Figure 2: An illustration of E3DPC-GZSL

- ZSL models are biased toward seen categories due to training on their samples.
- Calibrated stacking reduces seen probabilities to enhance the relative probabilities of unseen categories.

$$p'_k = p_k - \eta \cdot \mathbb{1}_{\mathcal{Y}^s}(c_k)$$

- We dynamically determine the calibration factor(η) based on the uncertainty of the input data.

Preliminary Experiments

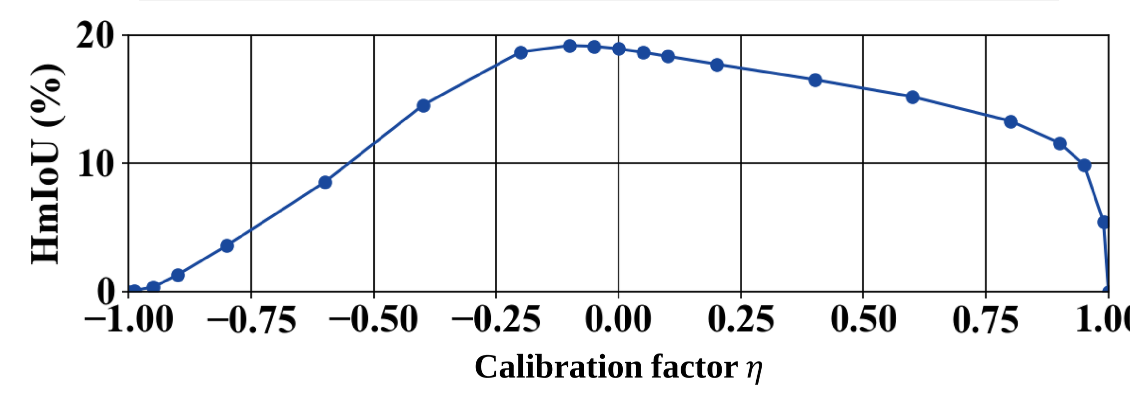

Figure 1a: Segmentation performance variation with different calibration factors

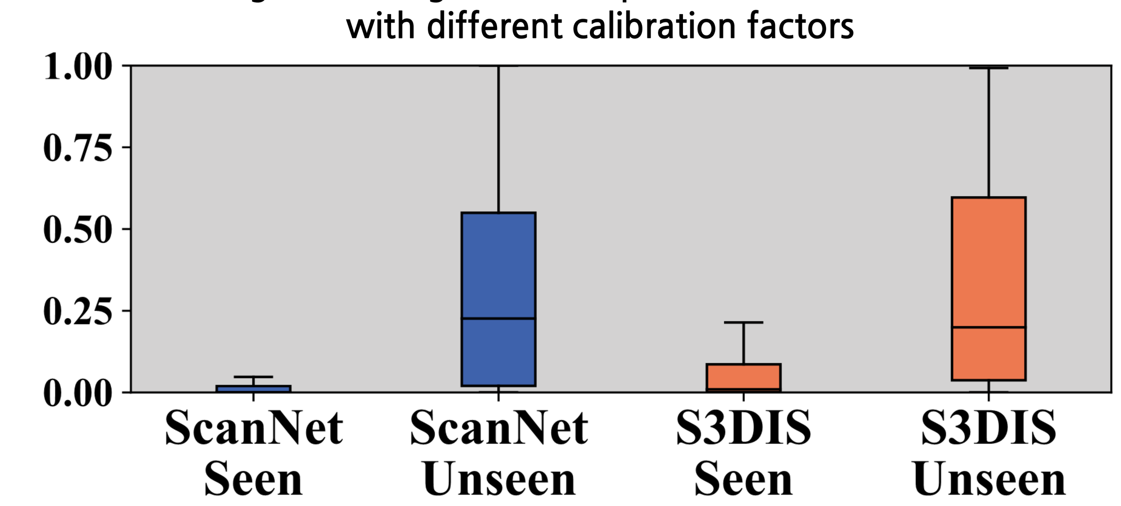

Figure 1b: Uncertainty difference between seen and unseen points

- Conventional calibration stacking employs a static calibration factor (η), heavily dependent on hyperparameters.
- GZSL models make uncertain predictions for unseen data. We exploit this aspect to dynamically obtain the calibration factor at the point level.

Methods

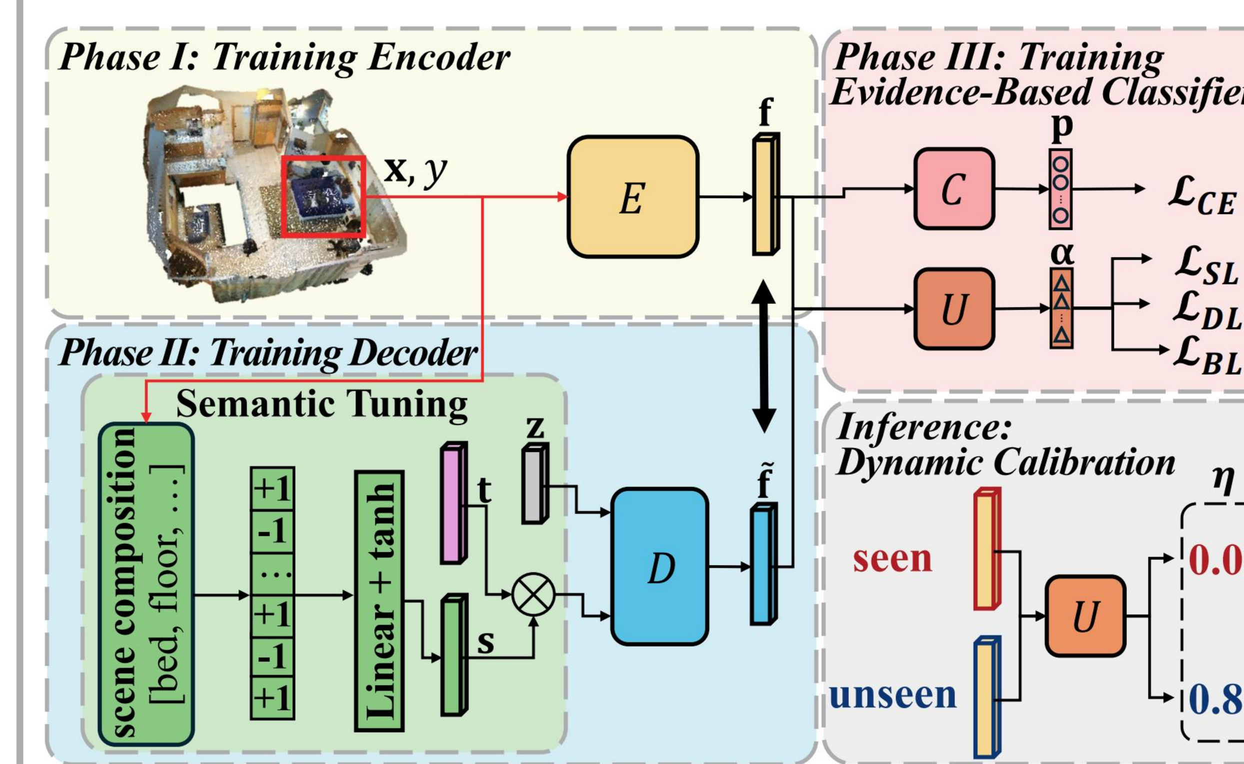

Figure 3: E3DPC-GZSL architecture

- We propose a novel method, E3DPC-GZSL, which mitigates the overconfidence of zero-shot models on seen categories by redistributing prediction probabilities using estimated uncertainty.
- We propose a new strategy for tuning semantic embeddings to overcome data scarcity in 3D zero-shot learning.

Results and Discussion

Quantitative Result

	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. 2023)	$\mathcal{Y}^s$	$\mathcal{Y}^s \cup \mathcal{Y}^{\tilde{u}}$	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***

Table 1: Performance Comparisons of 3D GZSL semantic segmentation benchmarks

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

Table 2: Analysis of the effects of each module on segmentation performance.

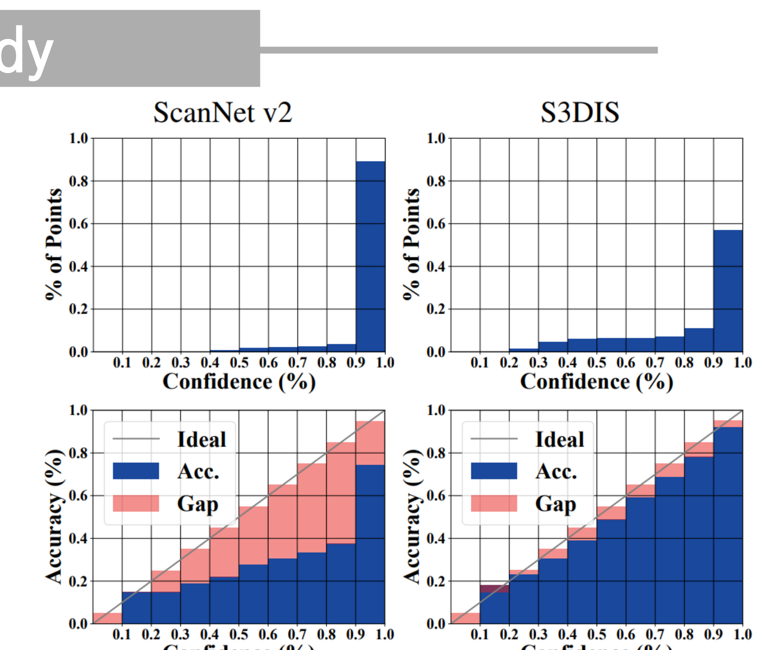

Figure 4: Analysis of model confidence

Discussion

- Our proposed method achieves improvements over the state-of-the-art in both seen and unseen mIoU metrics, demonstrating an increase of 1.4% HmIoU on ScanNet v2 and 3.7% HmIoU on S3DIS.
- The results demonstrate that dynamic calibration using module **U** effectively adjusts the module's predictions independent of semantic tuning. Furthermore, they show that applying semantic tuning enhances classifier performance by increasing the expressive power of the decoder.
- ScanNet v2 shows only marginal performance improvement with module **U**, which is attributed to the model's overconfidence issue.

Qualitative Result

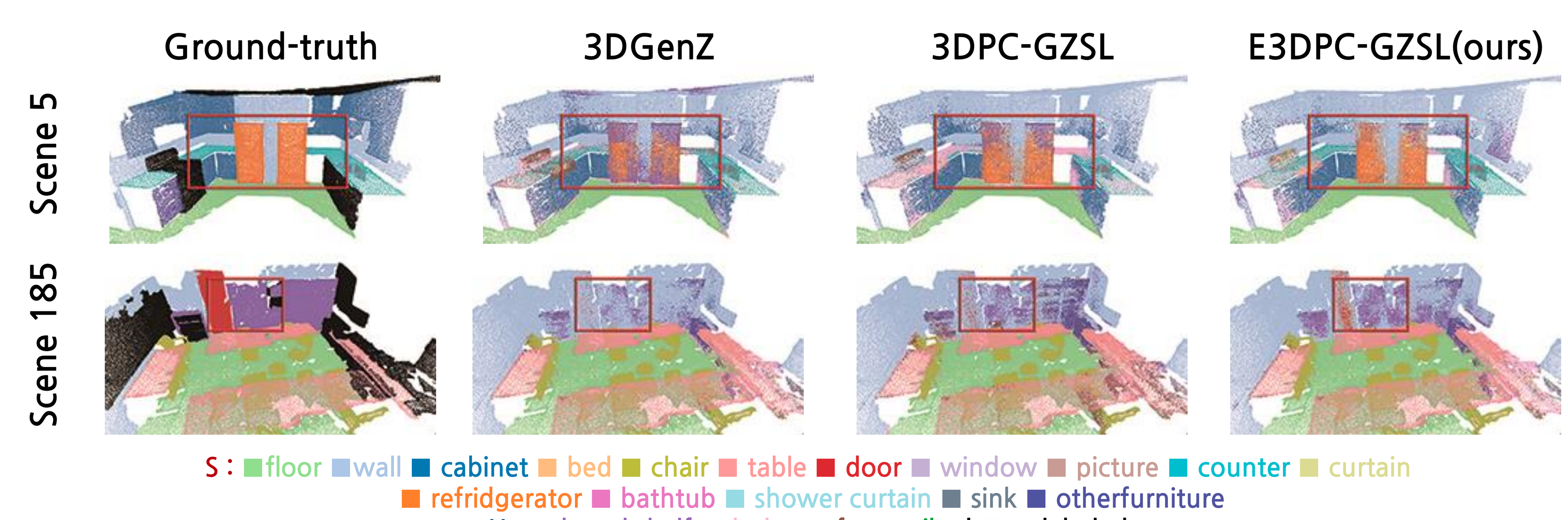

Figure 5a: Qualitative comparison on ScanNet v2 dataset

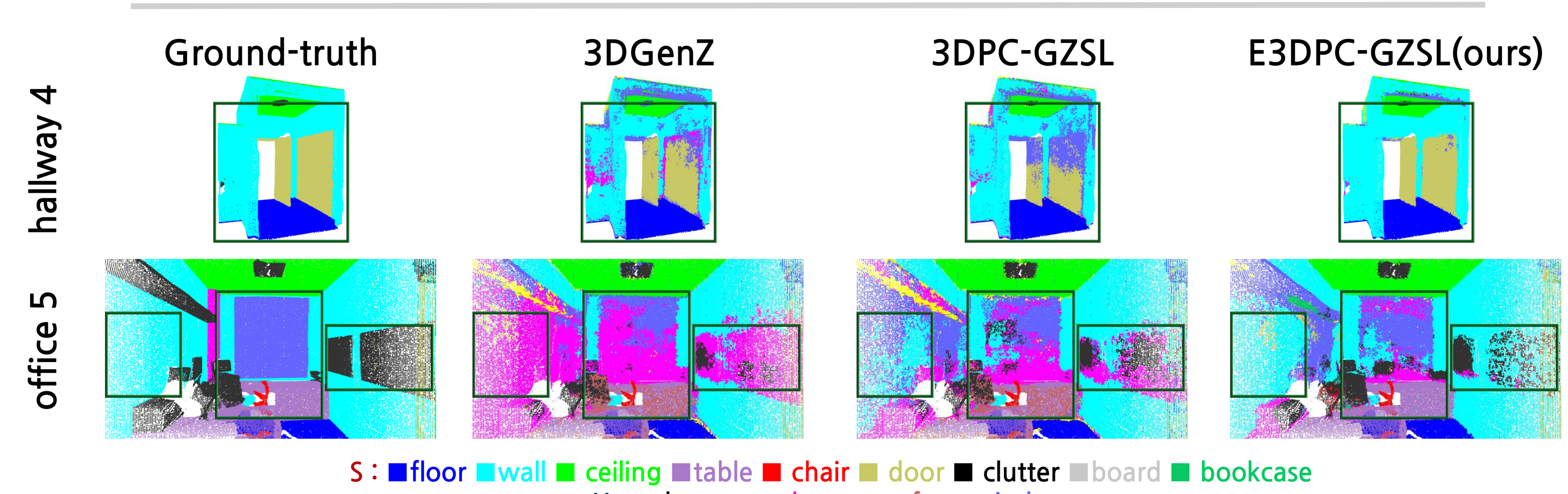

Figure 5b: Qualitative comparison on S3DIS dataset

Discussion

- For ScanNet v2, in scene 5, the proposed E3DPC-GZSL method classifies the refrigerator more accurately than the other methods.
- For S3DIS, in the Hallway 4, the proposed E3DPC-GZSL method successfully segments the door and clusters the closest points around it, while other methods struggle to distinguish the door from the wall.

Conclusion

- We proposed 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 introduced 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.