MICAS: Multi-grained In-Context Adaptive Sampling for 3D Point Cloud Processing

Supplementary Material

Summary of the Appendix

To complement the main paper, this supplementary material provides additional details and insights, structured as follows:

- Sec. A details the implementation of MICAS.
- Sec. B introduces the model variants of MICAS.
- Sec. C presents an additional ablation study on the impact of the number of candidate prompts in query-specific prompt sampling.
- Sec. D provides an ablation study to further demonstrate the robustness of our proposed MICAS.
- Sec. E offers additional qualitative analysis by visualizing sampled points.
- Sec. F discusses the limitations of our approach and its broader impacts.

A. Implementation Details

Following PIC [2], we sample $1,024$ points from each point cloud and segment them into 64 patches, each containing 32 neighboring points. PointNet [5] is used as the task encoder, point encoder, and prompt sampling module. For task-adaptive point sampling, we set the initial learning rate to $1e-4$, reducing it to $1e-6$ over 60 epochs using a Cosine Annealing Scheduler [4], with a batch size of 72 and a sampling loss hyperparameter α of 0.5. For query-specific prompt sampling, 8 candidate prompts are randomly selected per query, with a learning rate of $1e-5$, decay to $1e-6$, 30 training epochs, and a batch size of 9.

B. Model Variants

PIC [2] includes two variants: PIC-Cat and PIC-Sep, which differ in how they combine "input" and "target" point clouds. PIC-Cat concatenates the "input" and "target" point patches before feeding them into the transformer, while PIC-Sep processes the "input" and "target" point patches in parallel and merges their features after several blocks. We test our method on both variants.

C. Ablation Study: The Number of Candidates

To illustrate the impact of different K values on model performance, we conduct experiments in Table A1. As K increases, performance improves until a threshold, beyond which further increases in K result in longer inference times without improving performance. Based on this, we choose $K = 8$ for our work.

Table A1. Analysis of K in query-specific prompt sampling, where K denotes the number of candidate prompts for each query "input" point cloud. Value1/value2 denote PIC-Cat/PIC-Sep.

K	Reconstruction (CD ↓)	Denoising (CD ↓)	Registration (CD ↓)	Part seg. (mIOU ↑)	Inference time (ms)
2	4.8/4.4	4.7/5.2	12.0/3.9	87.9/86.7	22.4/20.9
4	4.8/4.4	4.6/5.1	10.7/3.7	87.9/86.8	33.6/32.8
8	4.7/4.3	4.6/5.1	9.8/3.7	87.9/86.8	47.1/45.9
12	4.7/4.3	4.6/5.1	9.6/3.7	87.9/86.8	66.0/65.0
16	4.7/4.3	4.6/5.1	9.4/3.7	87.9/86.8	79.8/78.9

D. Ablation Study: Robustness Analysis

In the task-adaptive point sampling module and query-specific prompt sampling module of our proposed MICAS, we design the task encoder, point encoder, and prompt sampling module based on PointNet [5]. To evaluate the robustness of MICAS, we conduct an additional ablation experiment by replacing PointNet with DGCNN [6], a model widely used for CNN-based high-level tasks on point clouds, such as classification and segmentation. Unlike PointNet [5], which relies on a multilayer perceptron (MLP) architecture, DGCNN [6] employs a dynamic graph CNN framework and introduces the EdgeConv operation. This operation effectively captures local geometric features of point clouds while maintaining permutation invariance.

The experimental results presented in Table A2, show that the performance trend of MICAS remains consistent across in-context learning models, including PIC-Cat [2] and PIC-Sep [2], regardless of whether PointNet [5] or DGCNN [6] is used. These findings highlight the robustness of MICAS, demonstrating its reliability across different in-context learning frameworks and point cloud models.

E. More Qualitative Analysis

To demonstrate the effectiveness of our proposed MICAS in central point sampling and prediction, we present a visual comparison between our task-adaptive point sampling method and Farthest Point Sampling (FPS) used in PIC-Cat [2] and PIC-Sep [2]. As shown in Figures A1 and A2, our proposed MICAS consistently selects higher-quality central points, delivering superior outcomes and overcoming the limitations of FPS. For instance, in the denoising task, FPS often prioritizes outliers, frequently selecting noisy points as central points. In contrast, MICAS effectively avoids these noisy points, focusing on more meaningful and valuable selections. In the reconstruction and

Table A2. Robustness studies on the ShapeNet In-Context Dataset [2]. ICL Model: in-context learning model. FPS: farthest point sampling. Point: task-adaptive point sampling. Prompt: query-specific prompt sampling. Introduced Model: the network model used by the task encoder, point encoder, and prompt sampling module in our proposed MICAS.

ICL Model	FPS	Point	Prompt	Reconstruction CD ↓						Denoising CD ↓						Registration CD ↓						Part Seg.	Introduced
				L1	L2	L3	L4	L5	Avg.	L1	L2	L3	L4	L5	Avg.	L1	L2	L3	L4	L5	Avg.	mIOU↑	Model
PIC-Cat [2]	√			4.9	**4.1**	**4.5**	**4.7**	6.3	4.9	**4.2**	5.1	5.9	6.8	7.8	6.0	6.5	7.8	13.6	20.4	24.5	14.5	79.9	-
		√		4.8	4.2	**4.5**	4.8	5.8	4.8	4.3	4.5	4.7	**4.9**	5.2	4.7	6.5	7.5	11.1	16.2	20.2	12.3	87.6	PointNet [5]
		√	√	**4.6**	4.2	**4.5**	4.8	**5.7**	**4.7**	**4.2**	**4.4**	**4.6**	**4.9**	**5.1**	**4.6**	**5.7**	**6.5**	**9.1**	**12.5**	**15.4**	**9.8**	**87.9**	PointNet [5]
	√			4.9	**4.1**	**4.5**	**4.7**	6.3	**4.9**	4.2	5.1	5.9	6.8	7.8	6.0	6.5	7.8	13.6	20.4	24.5	14.5	79.9	-
		√		4.9	4.2	4.6	4.9	5.9	**4.9**	4.1	4.3	4.6	**4.8**	5.0	4.6	6.6	7.5	11.5	16.7	20.5	12.6	**85.5**	DGCNN [6]
		√	√	**4.8**	4.2	4.6	4.9	**5.8**	**4.9**	**4.0**	**4.3**	**4.5**	**4.8**	**4.9**	**4.5**	**5.8**	**6.7**	**9.5**	**13.0**	**15.9**	**10.2**	85.4	DGCNN [6]
PIC-Sep [2]	√			3.9	**3.9**	**3.9**	**4.3**	6.2	4.4	6.2	7.2	7.7	8.2	8.3	7.5	7.6	7.8	8.4	9.0	10.0	8.6	78.7	-
		√		4.2	4.1	4.2	4.6	6.1	4.6	4.9	5.4	5.6	6.0	6.3	5.6	7.6	7.4	7.8	9.2	10.7	8.5	86.6	PointNet [5]
		√	√	**3.8**	**3.9**	4.0	4.4	**5.6**	**4.3**	**4.4**	**4.9**	**5.2**	**5.5**	**5.7**	**5.1**	**3.4**	**3.6**	**3.7**	**3.8**	**4.0**	**3.7**	**86.8**	PointNet [5]
	√			**3.9**	**3.9**	**3.9**	**4.3**	**6.2**	**4.4**	6.2	7.2	7.7	8.2	8.3	7.5	7.6	7.8	8.4	9.0	10.0	8.6	78.7	-
		√		4.4	4.2	4.3	4.9	6.7	4.9	4.9	5.4	5.7	6.0	6.3	5.7	8.0	8.0	8.6	9.3	9.8	8.7	83.9	DGCNN [6]
		√	√	4.0	4.0	4.2	4.6	**6.2**	4.6	**4.3**	**4.8**	**5.1**	**5.5**	**5.8**	**5.1**	**3.6**	**3.8**	**3.8**	**3.9**	**4.1**	**3.9**	**84.0**	DGCNN [6]

registration tasks, MICAS outperforms PIC-Cat [2] and PIC-Sep [2] by producing target point clouds with clearer contours and more accurate shapes. Similarly, in the part segmentation task, MICAS achieves accurate segmentation even in areas where PIC-Cat [2] and PIC-Sep [2] encounter segmentation errors. These visualization results underscore the significance and effectiveness of our proposed MICAS in advancing point cloud in-context learning.

F. Discussion

F.1. Limitations

While our proposed MICAS represents a pioneering effort to address inter-task and intra-task sensitivity challenges in point cloud in-context learning, it has a limitation. Specifically, in the query-specific prompt sampling, we prioritize selecting the "best-performing" prompt from a sampled set of 8 candidate prompts. This process requires predicting the sampling probability for each of the 8 candidate prompts, which increases the model's inference time. As shown in Table 2 of the main paper, the query-specific prompt sampling introduces additional computation, adding approximately 25 ms to the inference time. Nonetheless, despite this slight increase in inference time, the query-specific prompt sampling achieves significant performance gains, particularly in the registration task.

In future work, we recommend addressing this limitation by making the prompt sampling module more lightweight and reducing the size of the prompt candidate pool. Specifically, a simplified prompt sampling module could be developed to streamline the prediction of sampling probabilities and enhance prediction speed. Furthermore, reducing the number of candidate prompts from 8 to 4 or even 2 would significantly lower the computational burden, thereby reducing the overall inference time.

F.2. Broader Impacts

This work highlights the limitations of existing learnable task-based sampling approaches [1, 3, 7, 8], which focus solely on inter-point cloud adaptive sampling within the same task and lack the capability to perform inter-task adaptive sampling within the same point cloud. To address this gap, we propose a novel Multi-grained In-Context Adaptive Sampling mechanism, referred to as **MICAS**, which enables adaptive sampling within the same point cloud by leveraging various prompts.

In summary, our work represents the first shift in point cloud sampling from inter-point cloud adaptive sampling within the same task to inter-task adaptive sampling within the same point cloud. Furthermore, the proposed MICAS contributes positively to the research community by advancing the field of point cloud processing and inspiring future innovations in adaptive in-context learning frameworks.

References

[1] Oren Dovrat, Itai Lang, and Shai Avidan. Learning to sample. In *CVPR*, 2019. 2

[2] Zhongbin Fang, Xiangtai Li, Xia Li, Joachim M Buhmann, Chen Change Loy, and Mengyuan Liu. Explore in-context learning for 3d point cloud understanding. In *NeurIPS*, 2024. 1, 2, 3, 4

[3] Itai Lang, Asaf Manor, and Shai Avidan. Samplenet: Differentiable point cloud sampling. In *CVPR*, 2020. 2

[4] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. In *ICLR*, 2017. 1

[5] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *CVPR*, 2017. 1, 2

[6] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *ToG*, 2019. 1, 2

Figure A1. Qualitative experimental results compared with the PIC-Cat [2]. The red and green points denote the central points selected by PIC-Cat and our proposed MICAS, respectively. (Zoom in for more details)

[7] Cheng Wen, Baosheng Yu, and Dacheng Tao. Learnable skeleton-aware 3d point cloud sampling. In *CVPR*, 2023. 2

[8] Xu Yan, Chaoda Zheng, Zhen Li, Sheng Wang, and Shuguang Cui. Pointasnl: Robust point clouds processing using nonlocal neural networks with adaptive sampling. In *CVPR*, 2020. 2

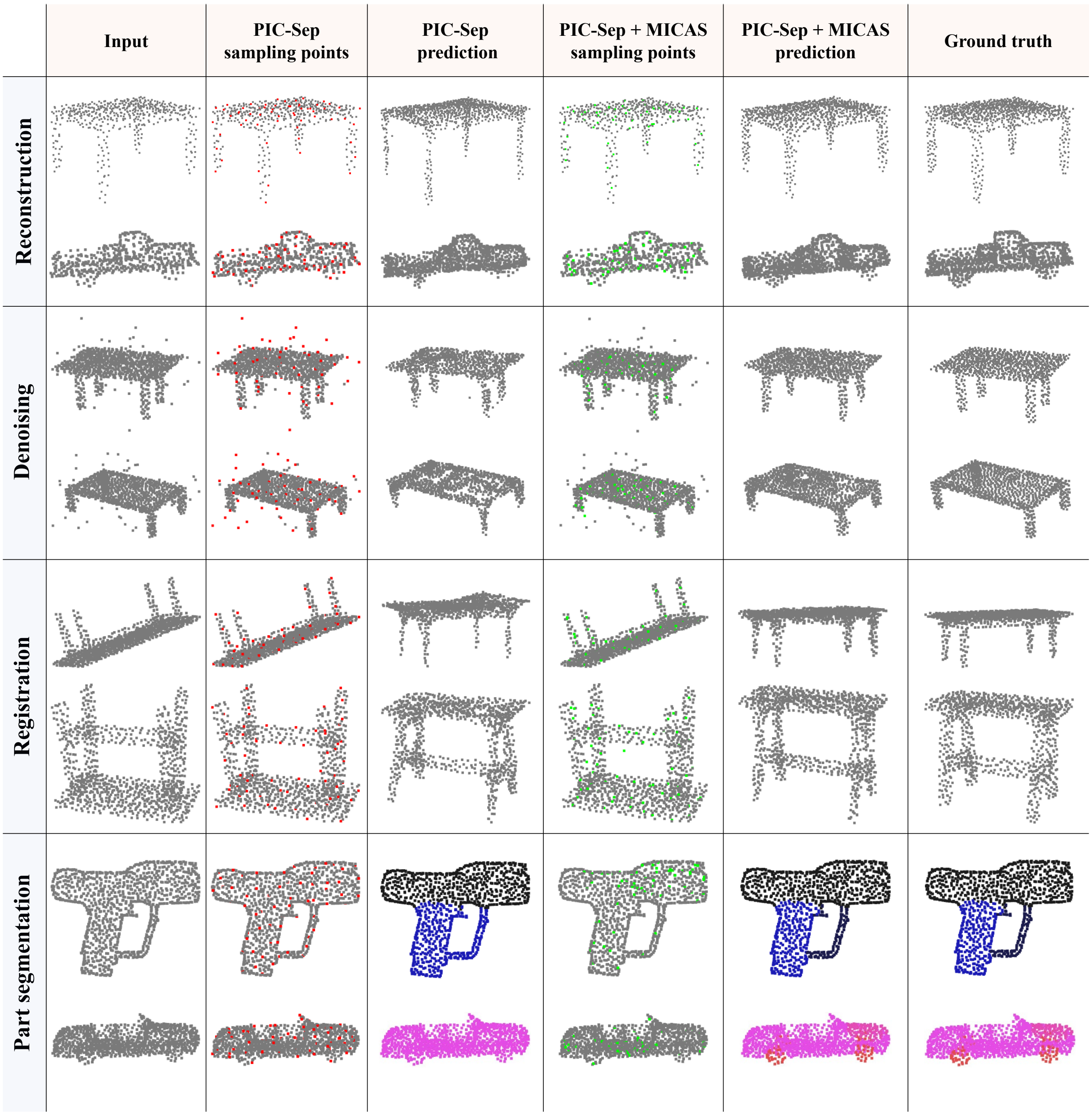

Figure A2. Qualitative experimental results compared with the PIC-Sep [2]. The red and green points denote the central points selected by PIC-Sep and our proposed MICAS, respectively. (Zoom in for more details)