A Prompt Learning Based Regulation Framework for Generalizable Point Cloud Analysis

Hongyu Sun[1,2] Qiuhong Ke[2] Yongcai Wang[1] Wang Chen[1] Kang Yang[1] Deying Li[1] Jianfei Cai[2]

[1]Department of Computer Scicence, Renmin University of China, China [2]Department of Data Science & AI, Monash University, Australia

Motivation

- **Problem:** Prompt learning on large 3D models often boosts point cloud recognition performance but harms generalization.
- **Objective:** To enhance downstream 3D tasks without compromising generalization by introducing a regulation framework for prompt learning on large 3D models.

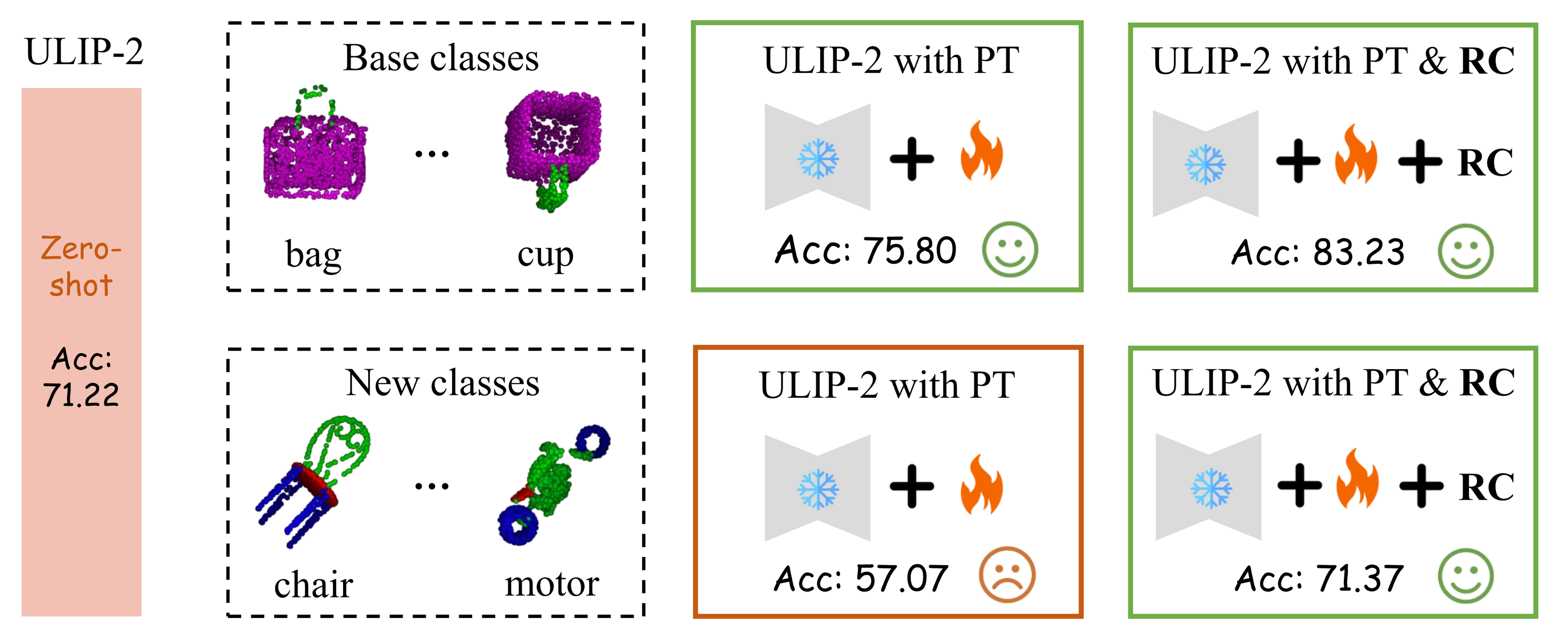

Figure 1. Motivation of our research: to promote the performances on downstream 3D tasks while maintaining good generalization of large 3D models.

Highlights

- **A Regulation Framework:** A plug-and-play framework with constraints (mutual agreement, text diversity, and model ensemble) to align prompt learning with general knowledge, improving both specific task performance and generalization.
- **Three New Benchmarks:** Created three benchmarks—base-to-new, cross-dataset, and few-shot—to test 3D domain generalization comprehensively.
- **Stunning Results:** Consistently increased accuracy across various models and datasets, showing superior generalization and robustness to corrupted data.

Methodology

Our framework proposes three regulation constraints: Mutual Agreement Constraint (MAC), Text Diversity Constraint (TDC), Model Ensemble Constraint (MEC)

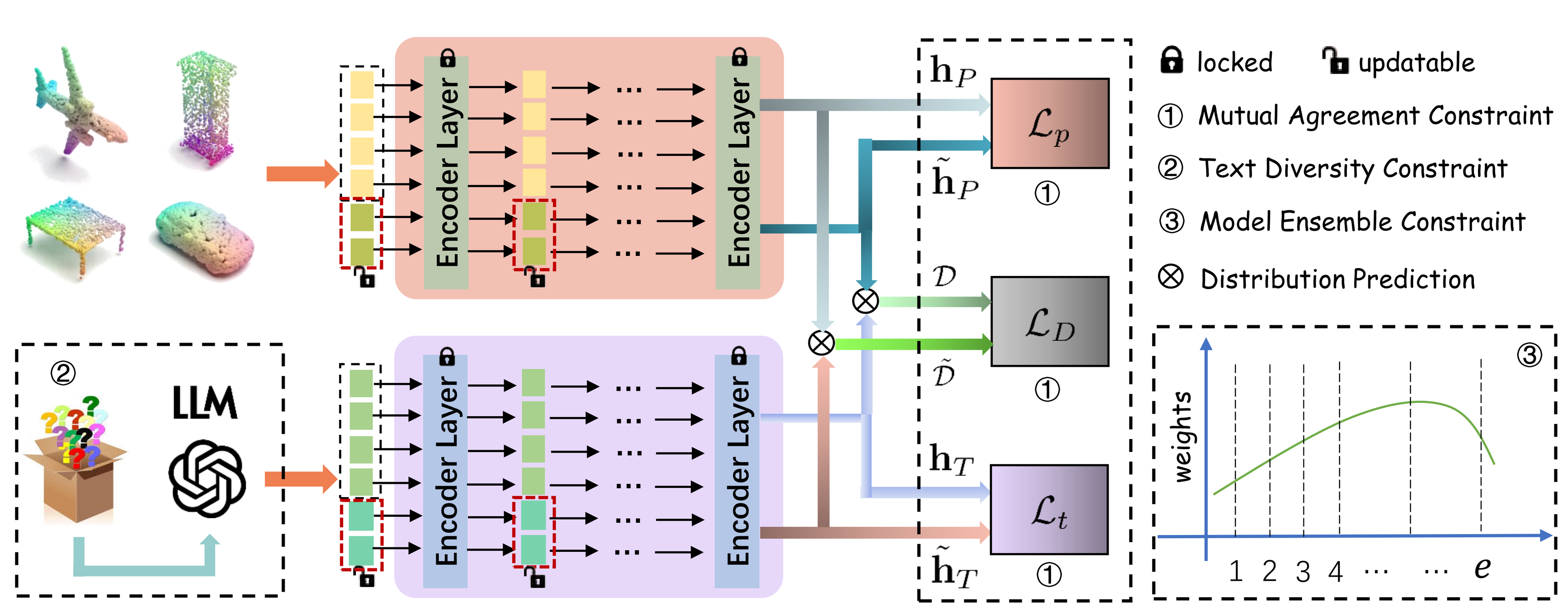

Figure 2. The overall architecture of our point regulation constraint framework, Point-PRC.

Prompt Templates to LLMs

Question Answering	Question Answering	Caption Generation	Making Sentences
What does a(n) {class} point cloud look like?	What are the identifying features of a(n) {class} point cloud?	Please describe a(n) {class} point cloud with details	Make a meaningful sentence with the following words: {class}, point cloud

Figure 3. Illustration of diverse questions to LLMs, including GPT-3.5, GPT-4 and PointLLM.

Comparison with Related Work

Methodology

- **Prior Approaches:** Focus on task-specific improvement on specific tasks for small-size point encoders but lack systematic design for generalization in large 3D models.
- **Our Contribution:** First framework to integrate regulatory constraints in prompt learning for large 3D models, offering substantial generalization gains over baseline methods in 3DDG.

Evaluation Benchmarks

- **Prior Benchmarks:** Limited scale and scope, e.g., only ∼10 classes shared between the source and the target domain in PointDA ans Sim-to-Real.
- **Our Contribution:** Designed diverse and challenging benchmarks, which contain up to 216 classes, to evaluate the generalization ability of large multi-modal 3D models

New 3DDG Benchmarks

Base-to-New Class Generalization Benchmark

- **Feature:** Tests adaptability from familiar to unseen classes within the same dataset.
- **Description:** The model is trained on a set of base classes and evaluated on unseen new classes in five point cloud datasets (e.g., S-PB_T50_RS, ShapeNetCoreV2).
- **Purpose:** Measures the ability to generalize without direct exposure to new classes during training.

Cross-Dataset Generalization Benchmark

- **Feature:** Assesses transferability across different datasets and includes out-of-distribution (OOD) generalization and robustness to data corruption.
- **Description:** The model learns from a source dataset (e.g., ShapeNetV2) and is tested on entirely different target datasets. It's also evaluated on corrupted data to test robustness.
- **Purpose:** Evaluates resilience to domain shifts, different 3D object sets, and common noise/corruptions in real-world point cloud data.

Few-Shot Generalization Benchmark

- **Feature:** Tests model performance with limited labeled examples.
- **Description:** Models are trained with very few samples per class (e.g., 1, 2, 4, 8, or 16 shots) and tested on a full test set.
- **Purpose:** Demonstrates the capability to generalize in low-data regimes, crucial for applications with limited labeled data.

Experiments

Base-to-new generalization.

(a) **Average over 5 datasets**

Method	Base	New	HM
P-CLIP [76]	75.66	23.45	35.80
P-CLIP2 [90]	74.11	37.84	50.10
ULIP [71]	77.32	49.01	59.99
+**RC**(Ours)	**82.19**	**61.93**	**70.64**
ULIP-2 [72]	77.91	67.91	72.57
+**RC**(Ours)	**83.18**	**76.10**	**79.48**

(b) ModelNet40

Method	Base	New	HM
P-CLIP [76]	93.23	20.22	33.23
P-CLIP2 [90]	93.98	45.21	61.05
ULIP [71]	92.80	50.07	65.05
+**RC**(Ours)	**95.03**	**55.27**	**69.89**
ULIP-2 [72]	91.77	56.47	69.92
+**RC**(Ours)	**95.30**	**64.83**	**77.17**

(c) S-PB_T50_RS

Method	Base	New	HM
P-CLIP [76]	61.25	19.87	30.01
P-CLIP2 [90]	56.84	29.92	39.20
ULIP [71]	56.73	25.80	35.47
+**RC**(Ours)	**64.20**	**49.17**	**55.69**
ULIP-2 [72]	66.40	66.47	66.43
+**RC**(Ours)	**73.67**	**74.27**	**73.97**

(d) S-OBJ_BG

Method	Base	New	HM
P-CLIP [76]	72.82	23.00	34.96
P-CLIP2 [90]	70.07	35.08	46.75
ULIP [71]	73.20	47.17	57.37
+**RC**(Ours)	**79.47**	**55.20**	**65.15**
ULIP-2 [72]	77.00	83.27	80.01
+**RC**(Ours)	**80.10**	**88.93**	**84.28**

(e) S-OBJ_ONLY

Method	Base	New	HM
P-CLIP [76]	76.23	20.23	31.97
P-CLIP2 [90]	71.40	44.39	54.74
ULIP [71]	74.13	50.80	60.29
+**RC**(Ours)	**79.23**	**65.93**	**71.97**
ULIP-2 [72]	78.60	76.27	77.42
+**RC**(Ours)	**83.60**	**81.10**	**82.33**

(f) ShapeNetCoreV2

Method	Base	New	HM
P-CLIP [76]	74.78	33.92	46.61
P-CLIP2 [90]	78.27	34.58	47.97
ULIP [71]	89.73	71.20	79.40
+**RC**(Ours)	**93.03**	**84.10**	**88.34**
ULIP-2 [72]	75.80	57.07	65.38
+**RC**(Ours)	**83.23**	**71.37**	**76.85**

Figure 4. Base-to-new generalization comparison for representative large 3D models based on prompt learning.

Cross-dataset generalization.

Method	**Source**	**Target**					**Avg.**
	ShapeNetV2	ModelNet40	S-PB_T50_RS	S-OBJ_BG	S-OBJ_ONLY	Omni3D	
P-CLIP [76]	67.41(0.09)	33.20(1.86)	15.51(0.58)	18.59(1.40)	22.89(2.32)	0.48(0.17)	22.55(1.54)
P-CLIP2 [90]	68.93(1.43)	54.73(1.48)	39.53(4.22)	**34.30**(1.28)	**25.63**(1.16)	8.63(2.52)	32.56(2.13)
+**RC**(Ours)	**69.80**(2.86)	**55.37**(1.78)	**39.77**(0.45)	34.20(0.54)	24.50(1.26)	**10.20**(0.40)	**32.81**(0.89)
ULIP [71]	87.33(0.95)	56.17(1.15)	26.83(2.15)	39.43(2.17)	43.53(1.32)	6.37(0.90)	34.47(1.54)
+**RC**(Ours)	**90.43**(0.86)	**58.00**(0.57)	**28.43**(0.68)	**40.33**(0.71)	**46.33**(1.54)	**8.20**(0.50)	**36.26**(0.80)
ULIP-2 [72]	76.70(1.37)	65.27(0.66)	40.07(0.34)	53.80(1.78)	48.53(1.72)	17.27(0.54)	44.99(1.01)
+**RC**(Ours)	**76.70**(1.59)	**72.10**(0.93)	**46.77**(2.43)	**59.03**(3.02)	**56.27**(0.97)	**21.80**(0.49)	**51.19**(1.57)

Figure 5. Comparison of OOD generalization in cross-dataset benchmark.

Few-shot generalization.