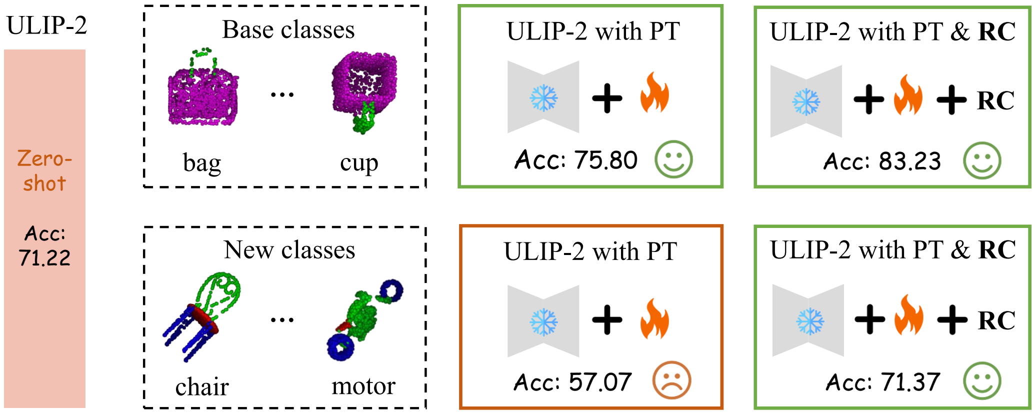


Figure 1: **Motivation of our research: to promote the performances on downstream 3D tasks while maintaining good generalization of large 3D models.** The experiments are conducted on ShapeNetCoreV2. ULIP-2 can reach 71.22% zero-shot recognition accuracy on this dataset. Recent works built on ULIP-2 introduce lightweight prompt tuning (PT) to further boost target tasks (75.80% accuracy). However, we observe the improvements come at the expenses of a severe drop in 3D domain generalization (e.g., 57.07% accuracy on new classes, much behind 71.22%), and develop a systematic regulation constraint (**RC**) framework to address this challenge, and construct three more comprehensive benchmarks to evaluate the 3DDG ability of large 3D models.

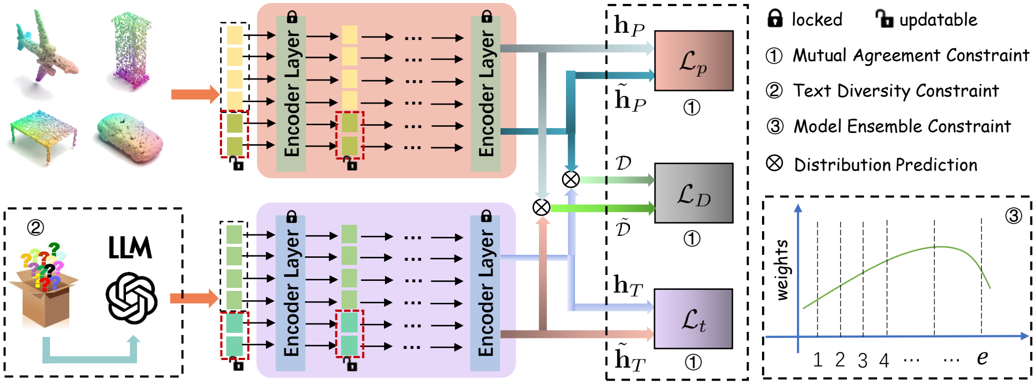
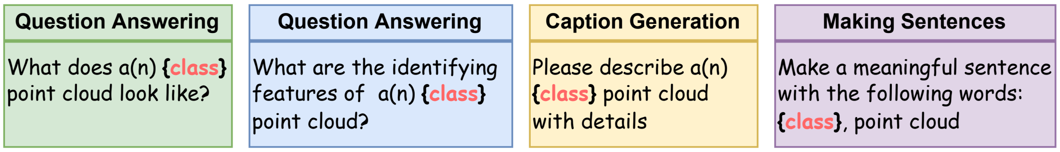


Figure 2: **The overall pipeline of our regulation constraint framework.** The large multi-modal 3D models consist of different branches that handle the point cloud (upper) and text (bottom) inputs. The backbones of different branches are frozen and the rectangles with red dashed lines represents newly introduced small number of learnable prompts. We devise a systematic framework composite of three explicit constraints (boxes with black dashed lines) to regulate the learning trajectory to enhance the task-specific performances and task-agnostic generalization.



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 the prompt templates to LLM.