

Active Prompt Learning in Vision Language Models

What is the problem?

- Pre-trained Vision Language Models (VLMs), like CLIP, show impressive **zero-shot performance** across tasks such as image classification and retrieval. However, when adapting these models to new tasks:
- Task-specific knowledge** is necessary, but acquiring labeled data is expensive.
- Active learning** methods are often used to select a small number of samples for labeling. However, applying **traditional active learning** techniques to VLMs does not consistently improve performance.
- One major issue is **class imbalance** introduced by VLMs, which leads to poor sample selection, affecting overall performance.

What has been done earlier?

- Pre-trained VLMs** like CLIP have successfully been adapted for zero-shot tasks without fine-tuning, showing **impressive generalization** across unseen datasets.
- To fine-tune models for specific tasks while keeping computational costs low, methods like **Prompt Learning** (e.g., CoOp) were introduced. They allow training only a small number of parameters (prompts) rather than fine-tuning the entire model.
- Active learning** methods, such as **uncertainty-based sampling** and **diversity-based sampling**, have been used to select informative samples, improving model performance with limited labeling. Examples include techniques like **Entropy Sampling** and **BADGE**.

What are the remaining challenges? What novel solution proposed by the authors to solve the problem?

Remaining Challenges-

- **Class Imbalance:** VLMs can introduce an imbalance in labeled data selection, leading to biased models. This imbalance arises because VLMs have unequal knowledge of various classes.
- **Active Learning Issues:** Applying conventional active learning techniques does not fully address the **class imbalance problem**, leading to performance degradation in some cases. Simply relying on uncertainty-based or diversity-based sampling is not enough to rectify this imbalance.

Proposed Solution-

- **PCB** (Pseudo-Class Balance) is a novel active learning framework that addresses class imbalance in VLMs. The key idea is to use the **knowledge embedded in pre-trained VLMs** to guide the sample selection process.