PromptKD: Unsupervised Prompt Distillation for Vision-Language Models

Zheng Li, Xiang Li#, Xinyi Fu, Xin Zhang, Weiqiang Wang, Shuo Chen, Jian Yang#.

Contact: Zheng Li, **Email:** zhengli97@mail.nankai.edu.cn, **Project Page:** https://zhengli97.github.io/PromptKD

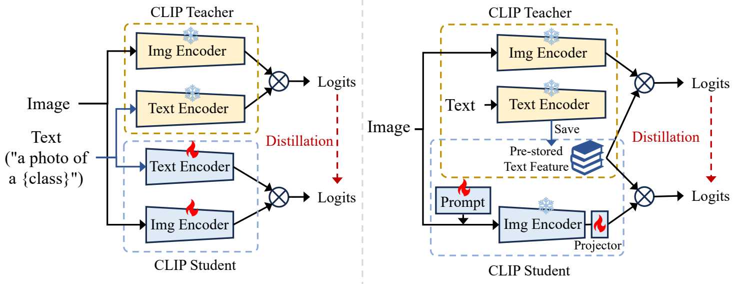
Background

Prompt learning is a technique that can efficiently adapt existing vision-language models to specific tasks through learnable textual or visual prompt.

Motivation

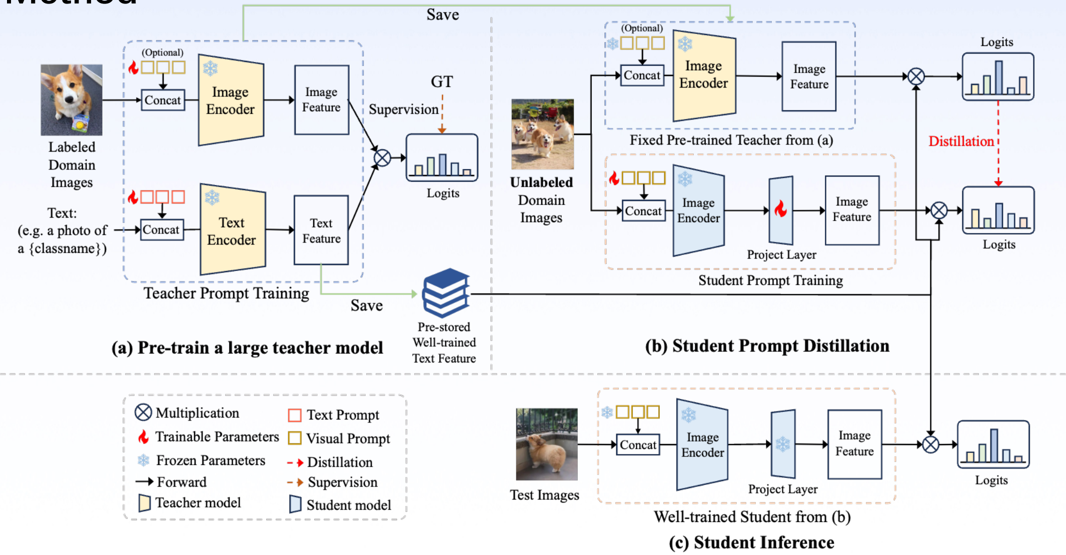
- Larger teacher models can provide better guidance for prompt learning.
- Leveraging decoupled-modality characteristics of CLIP can speed up training and inference.
- Training with extensive images leads to better prompt representations.

Overview



We propose to reuse the previously well-trained text features from the teacher pre-training stage and incorporate them into the student image encoder for both distillation and inference.

Method



- A novel two-stage unsupervised prompt distillation framework for VLMs.
- Reuse high-quality teacher text features instead of training the student's own text encoder.
- Distillation on extensive unlabeled domain images using soft labels provided by the teacher.
- PromptKD outperforms all existing methods on 11 diverse recognition datasets.

Impact of the components

Method	Base	Novel	HM
CLIP	72.43	68.14	70.22
Projector Only	78.48	72.79	75.53
Full Fine-tune	75.90	70.95	73.34
w/o Shared Text Feature	78.79	73.37	75.98
PromptKD	79.27	73.39	76.22

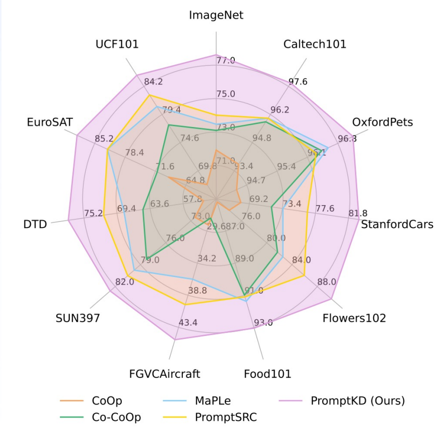
Different distillation ways.

Role (Method)	Img Backbone	Base	Novel	HM
CLIP	ViT-B/16	72.43	68.14	70.22
PromptSRC	ViT-B/16	77.60	70.73	74.01
Teacher (CLIP)	ViT-L/14	79.18	74.03	76.52
Student	ViT-B/16	76.53	72.58	74.50
Teacher (MaPLe)	ViT-L/14	82.79	76.88	79.73
Student	ViT-B/16	78.43	**73.61**	75.95
Teacher (PromptSRC)	ViT-L/14	83.24	76.83	79.91
Student	ViT-B/16	**79.27**	73.39	**76.22**

Different pre-training methods.

Experiments

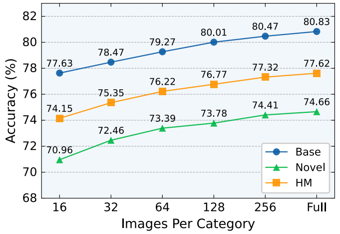
Setting: Teacher: ViT-L/14 CLIP, Student: ViT-B/16 CLIP



Base-to-novel experiments (HM) on 11 datasets.

Method	Domain Data	Base	Novel	HM
CLIP	Zero-shot	72.08	77.80	74.83
PromptSRC	Few-shot	98.07	76.50	85.95
CLIP-PR	Unlabeled	65.05	71.13	67.96
UPL		74.83	78.04	76.40
LaFTer		79.49	82.91	81.16
FPL	Few-shot + Unlabeled	97.60	78.27	86.87
IFPL		97.73	80.27	88.14
GRIP		97.83	80.87	88.54
PromptKD		99.42	82.62	90.24
Δ		+1.59	+1.75	+1.70

Comparison with existing works using unlabeled data.



Number of unlabeled images for distillation.

KD Form	Loss	Base	Novel	HM
Feature	L1	73.09	65.98	69.35
	MSE	71.89	66.17	68.91
Logit	KL	**79.27**	**73.39**	**76.22**

Different distillation forms and loss functions.