VLM 끄적끄적 5

Transfer Learning : Intro

VLM Introduction

1. VLM Transfer Learning의 Motivation
2. VLM Transfer Learning의 Setup
3. VLM Transfer Learning의 두 가지 연구 방향

1. VLM Transfer Learning의 Motivation

Pretraining & Downstream의 Gap 좁히기!

두 종류의 Gap이 존재한다.

- Gap 1) Image & Text distribution
  - Downstream task만의 task-specific한 image style & text format
- Gap 2) Training objectives
  - Pretraining: 주로 task-agnostic (general concept, coarse info)
  - Downstream: task-specific (coarse, fine-grained 등 다양함)

2. VLM Transfer Learning의 Setup

(1) Supervised TL

(2) Few-shot supervised TL

(3) Unsupervised TL

3. VLM Transfer Learning의 두 가지 연구 방향

(1) Prompt-tuning

(2) Feature adapter

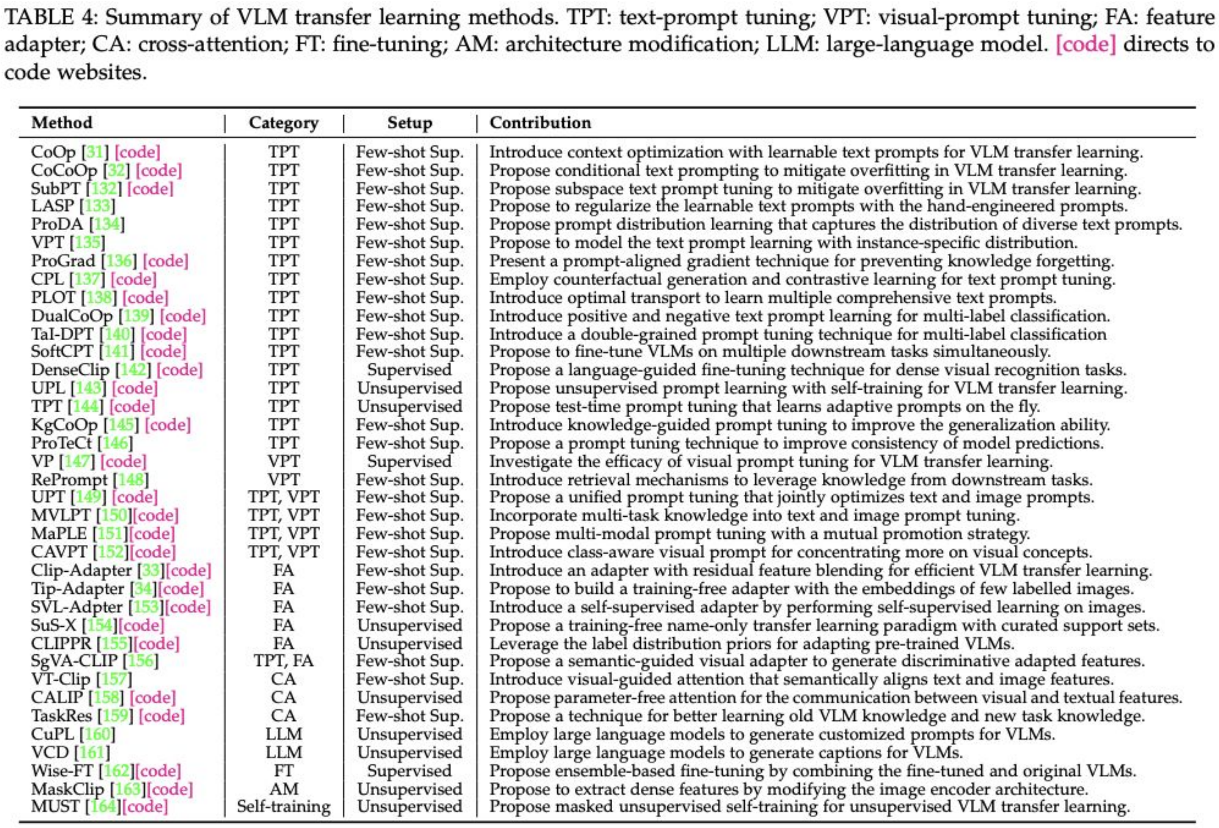

TABLE 4: Summary of VLM transfer learning methods. TPT: text-prompt tuning; VPT: visual-prompt tuning; FA: feature adapter; CA: cross-attention; FT: fine-tuning; AM: architecture modification; LLM: large-language model. [code] directs to code websites.

Method	Category	Setup	Contribution
CoOp [31] [code]	TPT	Few-shot Sup.	Introduce context optimization with learnable text prompts for VLM transfer learning.
CoCoOp [32] [code]	TPT	Few-shot Sup.	Propose conditional text prompting to mitigate overfitting in VLM transfer learning.
SubPT [132] [code]	TPT	Few-shot Sup.	Propose subspace text prompt tuning to mitigate overfitting in VLM transfer learning.
LASP [133]	TPT	Few-shot Sup.	Propose to regularize the learnable text prompts with the hand-engineered prompts.
ProDA [134]	TPT	Few-shot Sup.	Propose prompt distribution learning that captures the distribution of diverse text prompts.
VPT [135]	TPT	Few-shot Sup.	Propose to model the text prompt learning with instance-specific distribution.
ProGrad [136] [code]	TPT	Few-shot Sup.	Present a prompt-aligned gradient technique for preventing knowledge forgetting.
CPL [137] [code]	TPT	Few-shot Sup.	Employ counterfactual generation and contrastive learning for text prompt tuning.
PLOT [138] [code]	TPT	Few-shot Sup.	Introduce optimal transport to learn multiple comprehensive text prompts.
DualCoOp [139] [code]	TPT	Few-shot Sup.	Introduce positive and negative text prompt learning for multi-label classification.
TaI-DPT [140] [code]	TPT	Few-shot Sup.	Introduce a double-grained prompt tuning technique for multi-label classification
SoftCPT [141] [code]	TPT	Few-shot Sup.	Propose to fine-tune VLMs on multiple downstream tasks simultaneously.
DenseClip [142] [code]	TPT	Supervised	Propose a language-guided fine-tuning technique for dense visual recognition tasks.
UPL [143] [code]	TPT	Unsupervised	Propose unsupervised prompt learning with self-training for VLM transfer learning.
TPT [144] [code]	TPT	Unsupervised	Propose test-time prompt tuning that learns adaptive prompts on the fly.
KgCoOp [145] [code]	TPT	Few-shot Sup.	Introduce knowledge-guided prompt tuning to improve the generalization ability.
ProTeCt [146]	TPT	Few-shot Sup.	Propose a prompt tuning technique to improve consistency of model predictions.
VP [147] [code]	VPT	Supervised	Investigate the efficacy of visual prompt tuning for VLM transfer learning.
RePrompt [148]	VPT	Few-shot Sup.	Introduce retrieval mechanisms to leverage knowledge from downstream tasks.
UPT [149] [code]	TPT, VPT	Few-shot Sup.	Propose a unified prompt tuning that jointly optimizes text and image prompts.
MVLPT [150][code]	TPT, VPT	Few-shot Sup.	Incorporate multi-task knowledge into text and image prompt tuning.
MaPLE [151][code]	TPT, VPT	Few-shot Sup.	Propose multi-modal prompt tuning with a mutual promotion strategy.
CAVPT [152][code]	TPT, VPT	Few-shot Sup.	Introduce class-aware visual prompt for concentrating more on visual concepts.
Clip-Adapter [33][code]	FA	Few-shot Sup.	Introduce an adapter with residual feature blending for efficient VLM transfer learning.
Tip-Adapter [34][code]	FA	Few-shot Sup.	Propose to build a training-free adapter with the embeddings of few labelled images.
SVL-Adpter [153][code]	FA	Few-shot Sup.	Introduce a self-supervised adapter by performing self-supervised learning on images.
SuS-X [154][code]	FA	Unsupervised	Propose a training-free name-only transfer learning paradigm with curated support sets.
CLIPPR [155][code]	FA	Unsupervised	Leverage the label distribution priors for adapting pre-trained VLMs.
SgVA-CLIP [156]	TPT, FA	Few-shot Sup.	Propose a semantic-guided visual adapter to generate discriminative adapted features.
VT-Clip [157]	CA	Few-shot Sup.	Introduce visual-guided attention that semantically aligns text and image features.
CALIP [158] [code]	CA	Unsupervised	Propose parameter-free attention for the communication between visual and textual features.
TaskRes [159] [code]	CA	Few-shot Sup.	Propose a technique for better learning old VLM knowledge and new task knowledge.
CuPL [160]	LLM	Unsupervised	Employ large language models to generate customized prompts for VLMs.
VCD [161]	LLM	Unsupervised	Employ large language models to generate captions for VLMs.
Wise-FT [162][code]	FT	Supervised	Propose ensemble-based fine-tuning by combining the fine-tuned and original VLMs.
MaskClip [163][code]	AM	Unsupervised	Propose to extract dense features by modifying the image encoder architecture.
MUST [164][code]	Self-training	Unsupervised	Propose masked unsupervised self-training for unsupervised VLM transfer learning.