Visual + Prompt

Apr 23th 2022

Prompt

②设计灵活，可以适应不同类型的NLP任务

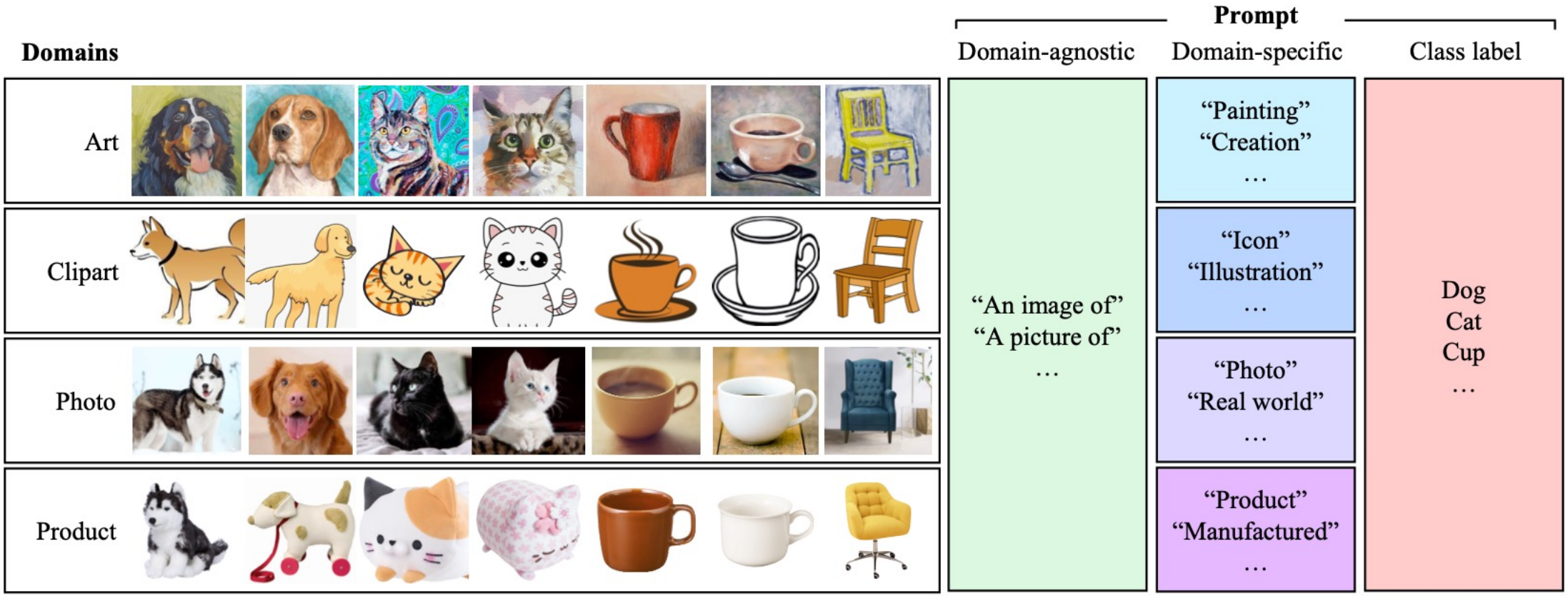
我觉得这一点是它被认为是NLP第四范式的关键

Type	Task	Input ([X])	Template	Answer ([Z])
Text CLS	Sentiment	I love this movie.	[X] The movie is [Z].	great fantastic ...
	Topics	He prompted the LM.	[X] The text is about [Z].	sports science ...
	Intention	What is taxi fare to Denver?	[X] The question is about [Z].	quantity city ...
Text-span CLS	Aspect Sentiment	Poor service but good food.	[X] What about service? [Z].	Bad Terrible ...
Text-pair CLS	NLI	[X1]: An old man with ... [X2]: A man walks ...	[X1]? [Z], [X2]	Yes No ...
Tagging	NER	[X1]: Mike went to Paris. [X2]: Paris	[X1][X2] is a [Z] entity.	organization location ...
Text Generation	Summarization	Las Vegas police ...	[X] TL;DR: [Z]	The victim ... A woman ... ...
	Translation	Je vous aime.	French: [X] English: [Z]	I love you. I fancy you. ...

DAPL 2202.06687

Domain Adaptation via Prompt Learning

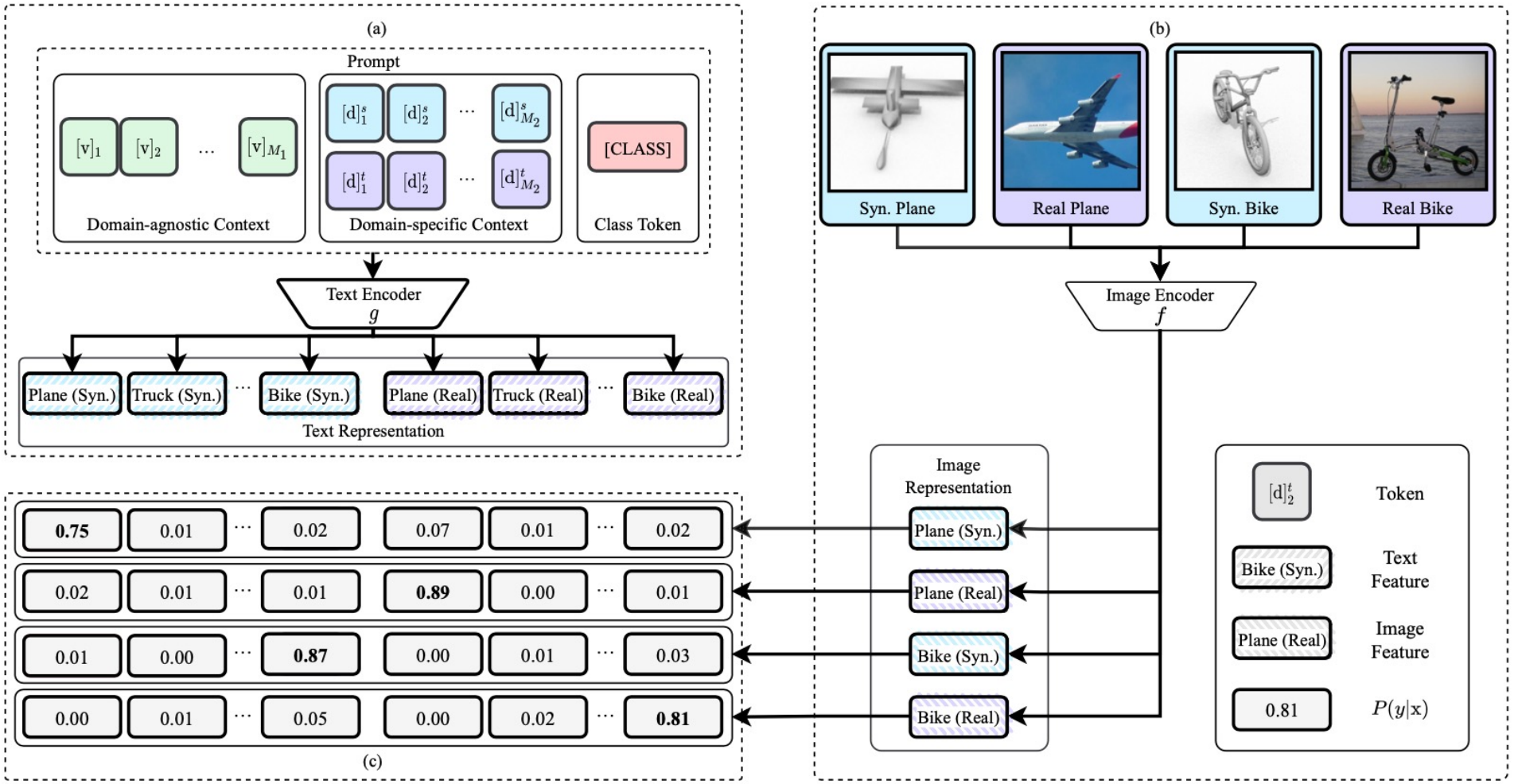
通过NLP跨模态的方式实现Domain Prompt编码



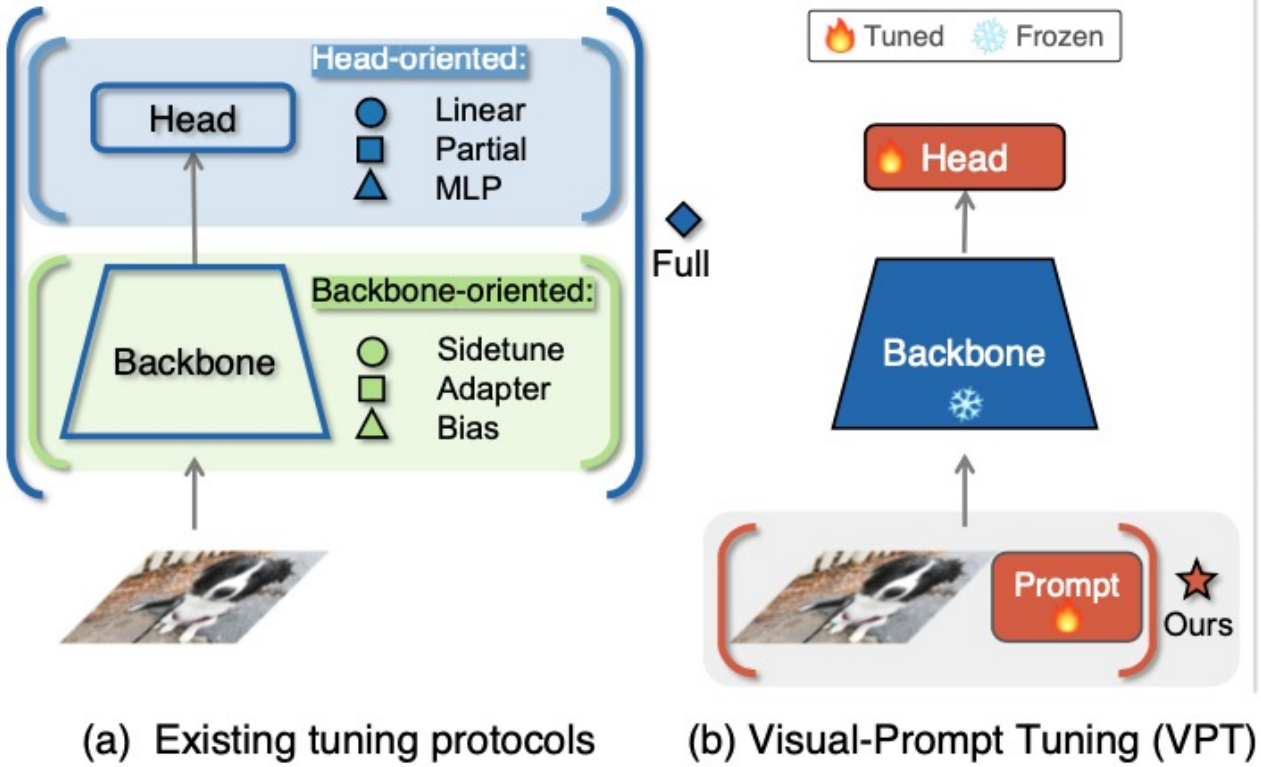
DAPL 2202.06687

Domain Adaptation via Prompt Learning

通过NLP跨模态的方式实现Domain Prompt编码

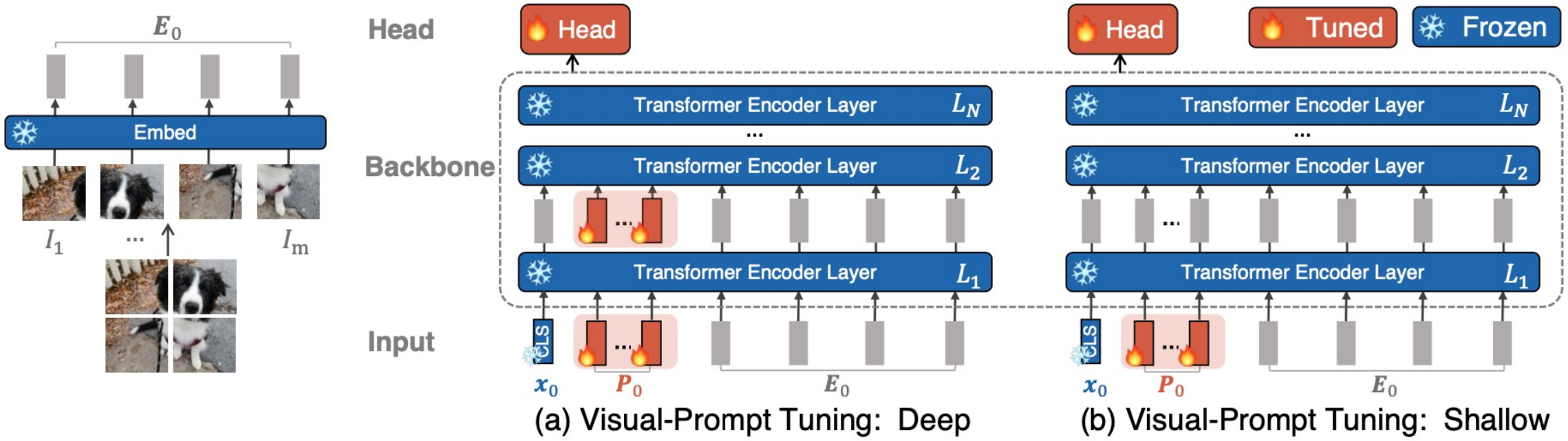


VPT visual prompt tuning



(a) Existing tuning protocols (b) Visual-Prompt Tuning (VPT) (c) Results on visual classification tasks

Prompt策略：加可学习Token在每个ViT Encoder



$$[\mathbf{x}_i, \_\_, \mathbf{E}_i] = L_i([\mathbf{x}_{i-1}, \mathbf{P}_{i-1}, \mathbf{E}_{i-1}])$$
$$\mathbf{y} = \text{Head}(\mathbf{x}_N) \quad .$$

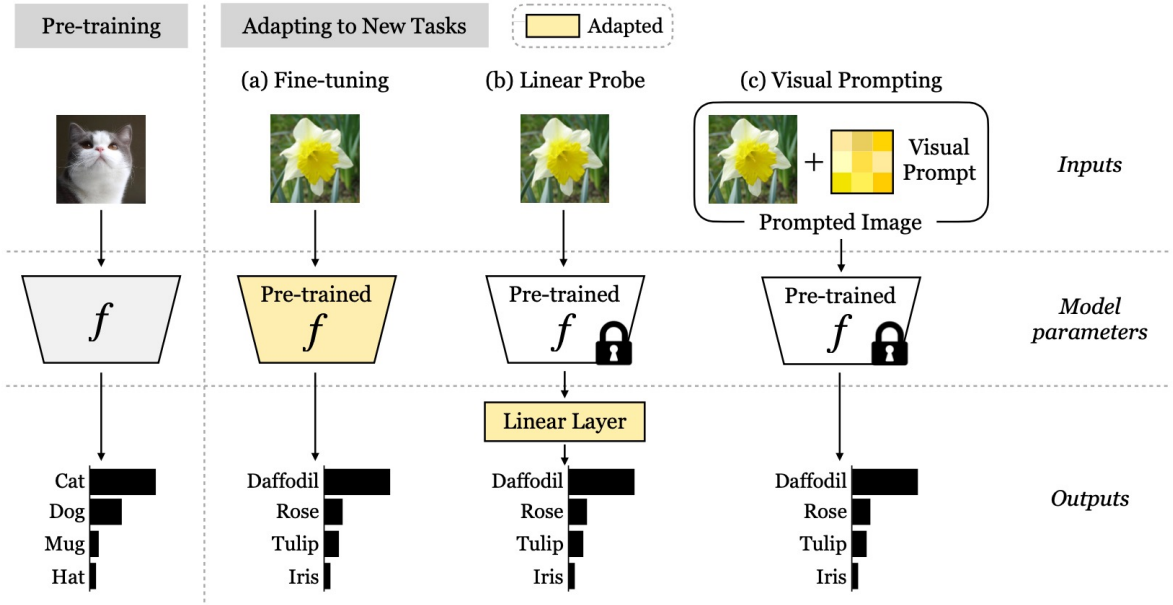
$$[\mathbf{x}_1, \mathbf{Z}_1, \mathbf{E}_1] = L_1([\mathbf{x}_0, \mathbf{P}, \mathbf{E}_0])$$
$$[\mathbf{x}_i, \mathbf{Z}_i, \mathbf{E}_i] = L_i([\mathbf{x}_{i-1}, \mathbf{Z}_{i-1}, \mathbf{E}_{i-1}])$$
$$\mathbf{y} = \text{Head}(\mathbf{x}_N) \quad ,$$

复现 https://github.com/sagizty/VPT

Visual Prompting:

Modifying Pixel Space to Adapt Pre-trained Models

通过加入Prompt Token进行训练（类似一种VPT-Shallow）



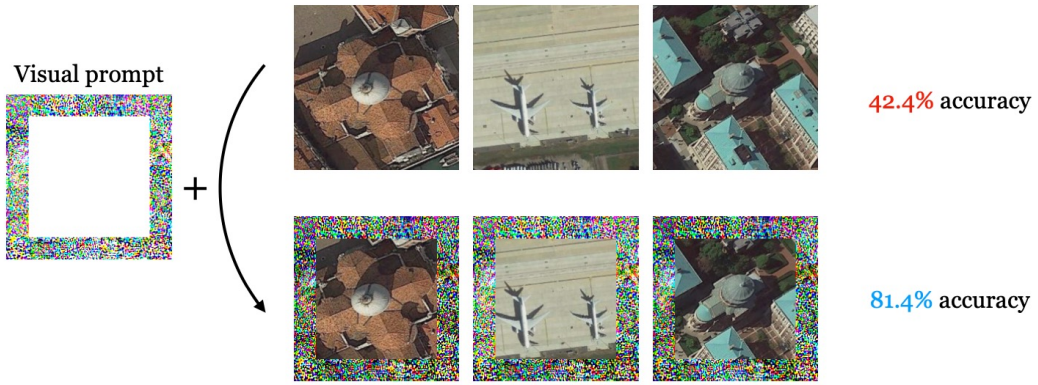


Figure 1: **All you need is a single patch.** Given a frozen, pre-trained model (CLIP [1] in this example), we wish to adapt the model to perform well on a new downstream task. Adding a *single, input-agnostic* prompt to input images improves performance across *all* object categories in a dataset.

Visual Prompting:

Modifying Pixel Space to Adapt Pre-trained Models

通过加入Prompt Token进行训练
（类似一种VPT-Shallow）