# DICE: Disentangled Instance-Class knowlEdge prompt tuning via SAE for Vision-Language Models


**Daeun Lee, Seungwoo Jang, Kwangsu Kim**

Department of Computer Science and Engineering, University of Sungkyunkwan



ECCV 2026 | MALMÖ | SEPT 8–12


Paper CV

## BACKGROUND Prompt tuning for VLMs

Among VLMs, CLIP is widely used in few-shot settings. In such settings, prompt tuning is commonly employed to adapt the models LLM-based prompt tuning aims to adapt CLIP with class-level priors to obtain discriminative semantics.

## LIMITATIONS LLM-based Prompt Tuning

① Not grounded in vision

- Descriptions generated text-only, no image seen
- One description reused for every image in a class

→ captures class-average semantics only

② Entangled & class-level only

- Rich LLM semantics remain entangled
- Exploited at class level, not instance level

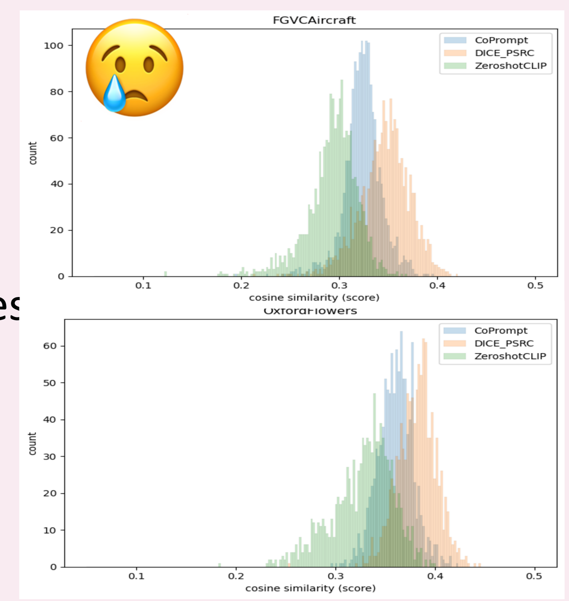
→ instance-level visual variation unused

③ No joint class–instance structure

- Class-level alone → misses fine-grained differences
- Instance-level alone → poor generalization

→ no mechanism to combine both

| FGVCAircraft DH-82 | Descriptions | Score |
|---|---|---|
| | The de Havilland DH-82 Tiger Moth is a 1930s biplane designed by Geoffrey de Havilland and manufactured by the de Havilland Aircraft Company. It was operated as a primary trainer by the Royal Air Force ( | 0.34 |
| | You can identify an aircraft DH-82 by its registration number. | 0.31 |
| | It looks like a small, single-engine airplane with two wings and a tail. | 0.27 |
| | A de Havilland Canada DHC-2 Beaver floatplane prepares to land in the waters of southeast Alaska. | **0.19** |
| | Cessna 182RG landing at McCarran International Airport | **0.11** |




## HOW Integrate class + instance knowledge

Few-shot generalization needs **both class-level structure and instance-level detail**; disentangling LLM priors into instance concepts and fusing them with class embeddings should improve generalization to unseen classes.

## CONTRIBUTION Three contributions

**SAE as a learnable concept dictionary** enabling explicit and interpretable concept selection.

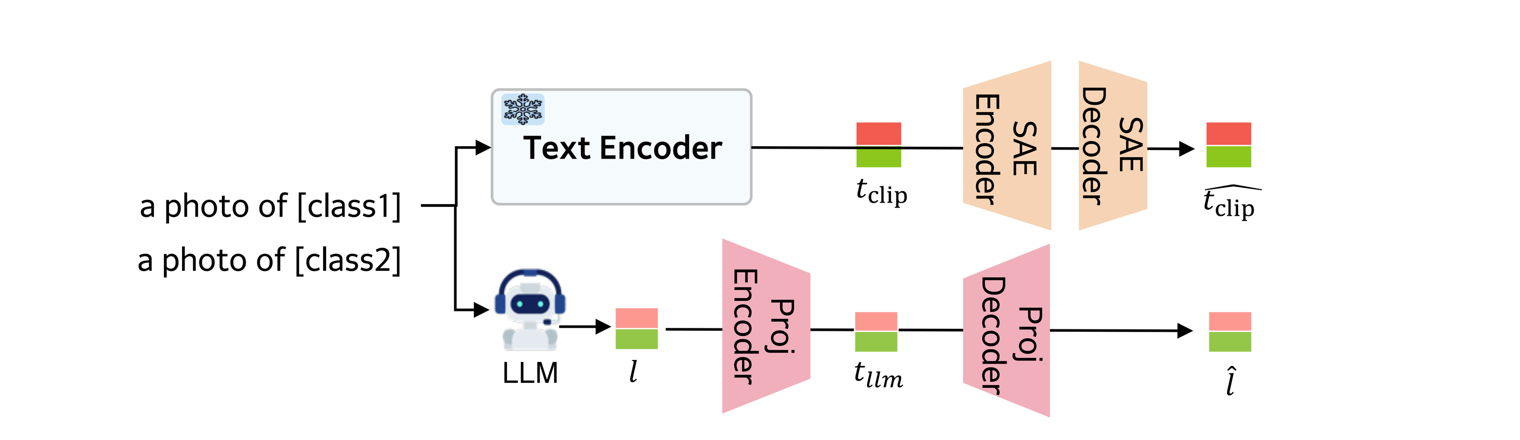
**A unified prompt** that integrates class-level priors with instance-specific visual concepts.

**Concept-based prompts** that adapt per instance, replacing predefined descriptions.

## What SAE-based disentanglement & unified context

A Sparse Autoencoder decomposes LLM-derived priors into sparse concept units; the selected instance concepts are fused with class embeddings into a unified prompt.

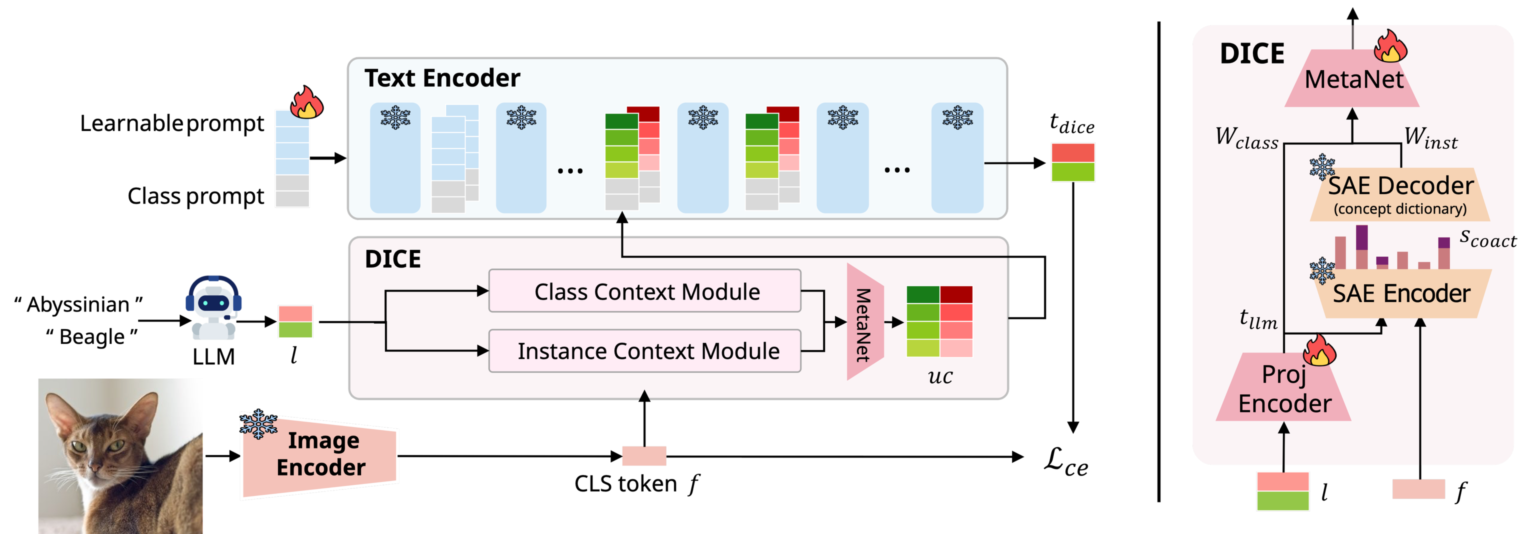
### Pre-training stage




*Training objectives*

$$\mathcal{L}_{proj} = 1 - \cos\left(t_{llm}, t_{clip}\right), \quad \mathcal{L}_{llm\_rec} = 1 - \cos(l, \hat{l}), \quad \mathcal{L}_{clip\_rec} = \left\| t_{clip} - \widehat{t_{clip}} \right\|^2$$

### Prompt-tuning stage




(a) Overall architecture (b) DICE

*1. Class context module*

$$W_{class} = E_{proj}(l), \qquad W_{class} \in \mathbb{R}^{C \times d}$$

*2. Instance context module*

*- Concept selection from the SAE decoder*

$$s_{class} = E_{sae}(t_{llm} - b_{pre}), \quad s_{inst} = E_{sae}(f - b_{pre})$$

$$s_{coact} = s_{class} \oplus s_{inst}$$

$$W_{inst} = \{ D_{sae}[i] \mid i \in \text{Top}-k\,(s_{coact}) \}, \quad W_{inst} \in \mathbb{R}^{C \times k \times d}$$

*3. Unified context injection*

$$UC = MetaNet(W_{class} + W_{inst})$$

## RESULT 1

### T1. Base-novel generalization

| Method | Type | Average Base | Average Novel | Average H | ImageNet [11] Base | ImageNet [11] Novel | ImageNet [11] H | Caltech101 [14] Base | Caltech101 [14] Novel | Caltech101 [14] H | OxfordPets [34] Base | OxfordPets [34] Novel | OxfordPets [34] H |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CoOp [52] | tp | 82.69 | 63.22 | 71.66 | 76.42 | 67.88 | 71.90 | 98.00 | 89.81 | 93.73 | 93.67 | 95.29 | 94.47 |
| CoCoOp [51] | tp | 80.47 | 71.69 | 75.83 | 75.98 | 70.43 | 73.10 | 97.96 | 93.81 | 95.84 | 95.20 | 97.69 | 96.43 |
| KART [20] | tp | 78.42 | 70.52 | 74.26 | 71.10 | 65.20 | 68.02 | 97.10 | 93.53 | 95.28 | 93.13 | 96.53 | 94.80 |
| TCP [50] | tp | 84.13 | 75.36 | 79.51 | 77.27 | 69.87 | 73.38 | 98.23 | 94.67 | 96.42 | 94.67 | 97.20 | 95.92 |
| MaPLe [22] | mp | 81.54 | 72.30 | 76.65 | 75.40 | 70.32 | 72.77 | 98.27 | 93.23 | 95.68 | 95.43 | 97.83 | 96.62 |
| PSRC [23] | mp | 84.26 | 76.10 | 79.97 | 77.60 | 70.73 | 74.01 | 98.10 | 94.03 | 96.02 | 95.33 | 97.30 | 96.30 |
| HPT [46] | mp | 84.32 | 76.86 | 80.23 | **77.95** | 70.74 | 74.17 | 98.37 | **94.98** | **96.65** | **95.78** | 97.65 | 96.71 |
| CoPrompt [39] | mp | 84.00 | 77.23 | 80.48 | 77.67 | **71.27** | **74.33** | 98.27 | 94.90 | 96.56 | 95.67 | **98.10** | **96.87** |
| DICE | tp | 84.45 | 76.77 | 80.43 | 76.87 | 70.07 | 73.31 | **98.57** | 94.33 | 96.40 | 94.80 | 97.37 | 96.07 |
| CoOp *w*/DICE | tp | 83.09 | 72.54 | 77.45 | 76.17 | 70.90 | 73.44 | 98.37 | 94.30 | 96.29 | 94.10 | 96.97 | 95.51 |
| CoCoOp *w*/DICE | tp | 82.70 | 73.19 | 77.65 | 76.30 | 69.23 | 72.59 | 98.10 | 93.87 | 95.94 | 95.27 | 97.47 | 96.36 |
| PSRC *w*/DICE | mp | **85.76** | **77.29** | **81.31** | 77.90 | 70.90 | 74.24 | 98.53 | 94.70 | 96.58 | 95.53 | 97.70 | 96.60 |

| Method | Type | StanfordCars [24] Base | StanfordCars [24] Novel | StanfordCars [24] H | Flowers [32] Base | Flowers [32] Novel | Flowers [32] H | Food101 [4] Base | Food101 [4] Novel | Food101 [4] H | Aircraft [30] Base | Aircraft [30] Novel | Aircraft [30] H |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CoOp [52] | tp | 78.12 | 60.40 | 68.13 | 97.60 | 59.67 | 74.06 | 88.33 | 82.26 | 85.19 | 40.44 | 22.30 | 28.75 |
| CoCoOp [51] | tp | 70.49 | 73.59 | 72.01 | 94.87 | 71.75 | 81.71 | 90.70 | 91.29 | 90.99 | 33.41 | 23.71 | 27.74 |
| KART [20] | tp | 69.47 | 66.20 | 67.80 | 95.00 | 71.20 | 81.40 | 86.13 | 87.06 | 86.59 | 29.67 | 28.73 | 29.19 |
| TCP [50] | tp | 80.80 | 74.13 | 77.32 | 97.73 | 75.57 | 85.23 | 90.57 | 91.37 | 90.97 | 41.97 | 34.43 | 37.83 |
| MaPLe [22] | mp | 74.70 | 71.20 | 72.91 | 97.70 | 68.68 | 80.66 | 90.30 | 88.57 | 89.43 | 36.90 | 34.13 | 35.46 |
| PSRC [23] | mp | 78.27 | 74.97 | 76.58 | 98.07 | 76.50 | 85.95 | 90.67 | 91.53 | 91.10 | 42.73 | 37.87 | 40.15 |
| HPT [46] | mp | 76.95 | 74.23 | 75.57 | 98.17 | **78.37** | **87.16** | 90.46 | 91.57 | 91.01 | 42.68 | 38.13 | 40.28 |
| CoPrompt [39] | mp | 76.97 | 74.40 | 75.66 | 97.27 | 76.60 | 85.71 | 90.73 | **92.07** | 91.40 | 40.20 | 39.33 | 39.76 |
| DICE | tp | 80.67 | **75.57** | 78.04 | 98.07 | 77.00 | 86.27 | 90.83 | 91.63 | 91.23 | 45.77 | 38.07 | 41.56 |
| CoOp *w*/DICE | tp | 77.80 | 72.60 | 75.11 | 97.97 | 71.97 | 82.98 | 89.10 | 90.55 | 89.82 | 42.97 | 34.33 | 38.17 |
| CoCoOp *w*/DICE | tp | 74.57 | 74.00 | 74.28 | 97.27 | 73.53 | 83.75 | 90.40 | 91.20 | 90.80 | 41.77 | 33.97 | 37.47 |
| PSRC *w*/DICE | mp | **81.93** | 74.87 | **78.24** | **98.40** | 77.07 | 86.44 | **93.73** | 90.50 | **92.09** | **48.77** | **39.37** | **43.57** |

| Method | Type | SUN397 [48] Base | SUN397 [48] Novel | SUN397 [48] H | DTD [8] Base | DTD [8] Novel | DTD [8] H | EuroSAT [19] Base | EuroSAT [19] Novel | EuroSAT [19] H | UCF101 [42] Base | UCF101 [42] Novel | UCF101 [42] H |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CoOp [52] | tp | 80.60 | 65.89 | 72.51 | 79.44 | 41.18 | 54.24 | 92.19 | 54.74 | 68.69 | 84.69 | 56.05 | 67.46 |
| CoCoOp [51] | tp | 79.74 | 76.86 | 78.27 | 77.01 | 56.00 | 64.85 | 87.49 | 66.00 | 75.24 | 85.23 | 71.97 | 78.04 |
| KART [20] | tp | 79.40 | 74.33 | 76.78 | 75.97 | 58.30 | 65.97 | 84.80 | 67.57 | 75.21 | 80.83 | 67.10 | 73.33 |
| TCP [50] | tp | 82.63 | 78.20 | 80.35 | 82.77 | 58.07 | 68.25 | 91.63 | 74.73 | 82.32 | 87.13 | 80.77 | 83.83 |
| MaPLe [22] | mp | 78.47 | 76.93 | 77.69 | 80.67 | 56.48 | 66.44 | 83.90 | 66.00 | 73.88 | 85.23 | 71.97 | 78.04 |
| PSRC [23] | mp | 82.67 | 78.47 | 80.52 | 83.37 | 62.97 | 71.75 | 92.90 | 73.90 | 82.32 | 87.10 | 78.80 | 82.74 |
| HPT [46] | mp | 82.57 | **79.26** | **80.88** | **83.84** | 63.33 | 72.16 | 94.24 | 77.12 | 84.82 | 86.52 | 80.06 | 83.16 |
| CoPrompt [39] | mp | 82.63 | 80.03 | 81.31 | 83.13 | 64.73 | 72.79 | 94.60 | 78.57 | 85.84 | 86.90 | 79.57 | 83.07 |
| DICE | tp | 82.73 | 78.23 | 80.42 | 82.30 | 61.53 | 70.42 | 91.83 | 78.53 | 84.66 | 86.53 | **82.13** | **84.28** |
| CoOp *w*/DICE | tp | 80.73 | 74.00 | 77.22 | 81.37 | 55.67 | 66.11 | 91.33 | 63.07 | 74.61 | 84.33 | 73.80 | 78.71 |
| CoCoOp *w*/DICE | tp | 81.53 | 76.17 | 78.76 | 80.13 | 55.80 | 65.79 | 90.20 | 65.03 | 75.57 | 84.13 | 74.77 | 79.17 |
| PSRC *w*/DICE | mp | **82.95** | 78.77 | 80.81 | 83.63 | **65.27** | **73.32** | **94.70** | **80.57** | **87.06** | **87.30** | 80.53 | 83.78 |

### T2. Few-shot generalization

| Method | Shots 1 | Shots 2 | Shots 4 | Shots 8 | Shots 16 |
|---|---|---|---|---|---|
| CoOp | 67.56 | 64.78 | 71.61 | 76.09 | 79.89 |
| CoCoOp | 66.79 | 67.65 | 71.21 | 72.96 | 74.90 |
| PSRC | 72.32 | 75.29 | 78.06 | 80.69 | 82.87 |
| CoPrompt[1] | 68.08 | 71.33 | 74.51 | 77.75 | 79.27 |
| HPT[1] | 69.72 | 72.49 | 76.19 | 78.53 | 81.15 |
| CoOp *w*/DICE | 68.93 | 72.33 | 75.18 | 77.70 | 80.96 |
| CoCoOp *w*/DICE | 66.92 | 70.52 | 72.68 | 75.25 | 77.97 |
| PSRC *w*/DICE | **72.94** | **75.90** | **79.01** | **81.19** | **83.21** |

### T3. Cross generalization

| Method | Type | ImageNet | Caltech | Pets | Cars | Flowers | Food | Aircraft | SUN397 | DTD | EuroSAT | UCF | Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CoOp | tp | 71.51 | 93.70 | 89.14 | 64.51 | 68.71 | 85.30 | 18.47 | 64.15 | 41.92 | 46.39 | 66.55 | 63.88 |
| CoCoOp | tp | 71.02 | 94.43 | 90.14 | 65.32 | 71.88 | 86.06 | 22.94 | 67.36 | 45.73 | 45.37 | 68.21 | 65.74 |
| MaPLe | mp | 70.72 | 93.53 | 90.49 | 65.57 | 72.23 | 86.20 | 24.74 | 67.01 | 46.49 | 48.06 | 68.69 | 66.30 |
| PSRC | mp | 71.27 | 93.60 | 90.25 | 65.70 | 70.25 | 86.15 | 23.90 | 67.10 | 46.87 | 45.50 | 68.75 | 65.81 |
| TCP | tp | 71.40 | 93.97 | 91.25 | 64.69 | 71.21 | 86.69 | 23.45 | 67.15 | 44.35 | 51.45 | 68.73 | 66.29 |
| HPT | mp | 71.72 | 94.20 | **92.63** | **66.33** | **74.84** | 86.21 | 25.68 | 68.75 | 50.87 | 47.36 | **70.50** | 67.73 |
| CoPrompt | mp | 70.80 | 94.50 | 90.73 | 65.67 | 72.30 | 86.43 | 24.00 | 67.57 | 47.07 | 51.90 | 69.73 | 67.00 |
| CoOp *w*/DICE | tp | **72.03** | 93.03 | 89.50 | 64.37 | 70.90 | 85.57 | 20.40 | 66.00 | 42.90 | 49.50 | 67.70 | 64.99 |
| CoCoOp *w*/DICE | tp | 71.76 | **96.13** | 90.23 | 59.57 | 70.73 | **88.93** | **25.83** | **71.25** | **50.93** | **57.60** | 70.20 | **68.14** |

## RESULT 2

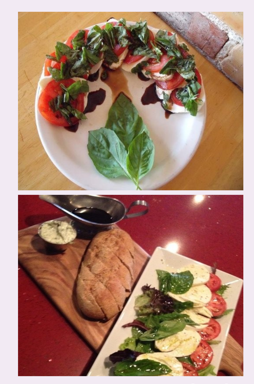
### F1. Examples of disentangled latent interpretation

freckled

| Concept | Latent | $s_{coact}$ |
|---|---|---|
| facial | 7878 | 0.20 |
| freckle | 9584 | 0.17 |
| marbling | 2862 | 0.17 |
| patchiness | 15844 | 0.15 |

2012 Chevrolet Silverado 2500HD Regular Cab

| Concept | Latent | $s_{coact}$ |
|---|---|---|
| four-wheeler | 7830 | 0.20 |
| truck bed | 14 | 0.19 |
| Truck | 16148 | 0.18 |
| hauling | 1202 | 0.18 |



caprese salad

| Concept | Latent | $s_{coact}$ |
|---|---|---|
| salad | 8684 | 0.20 |
| Ploughman’s lunch | 833 | 0.17 |
| mozzarella | 8015 | 0.17 |
| tomato | 2650 | 0.15 |

| Concept | Latent | $s_{coact}$ |
|---|---|---|
| Italian bread | 11935 | 0.16 |
| Green salad | 14713 | 0.16 |
| mozzarella | 8015 | 0.16 |
| Salad plate | 8712 | 0.15 |

### T4. Class-wise latent activation analysis.

| Dataset | Class | Latent | Act. | Ref. |
|---|---|---|---|---|
| StanfordCars | 2007 Chevrolet Corvette Ron Fellows Edition Z06 | 4910 | 0.1968 | edge |
| | | 6382 | 0.1885 | dashboard |
| | | 16351 | 0.1715 | chair |
| | | 12769 | 0.1503 | sports car |
| Caltech101 | elephant | 10579 | 0.1244 | elephant |
| | | 4544 | 0.1187 | tusker |
| | | 10851 | 0.1165 | African elephant |
| | | 724 | 0.1052 | big head |
| UCF101 | Apply Eye Makeup | 15299 | 0.1893 | whiteface |
| | | 4 | 0.1869 | eyelash |
| | | 4440 | 0.1848 | front |
| | | 11337 | 0.1785 | eyeliner |

### T5. Intra/inter-class similarity.

| Dataset | $s_{coact}$ Intra ↑ | $s_{coact}$ Inter ↓ | $s_{inst}$ Intra ↑ | $s_{inst}$ Inter ↓ |
|---|---|---|---|---|
| FGVCAircraft | 0.14 | 0.18 | 0.07 | 0.62 |
| Food101 | 0.12 | 0.09 | 0.04 | 0.41 |
| OxfordPets | 0.14 | 0.11 | 0.05 | 0.45 |

### F2. Comparison of UMAP between LLM-based prompt tunings and ours.