DeCoOp: Robust Prompt Tuning with Out-of-Distribution Detection

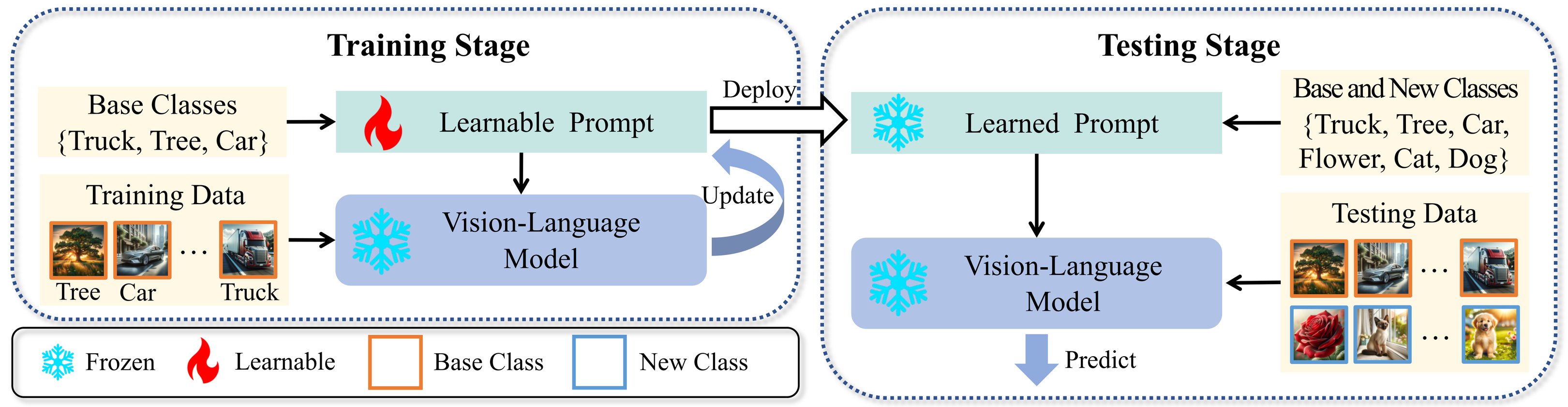
Zhi Zhou, Ming Yang, Jiang-Xin Shi, Lan-Zhe Guo, Yu-Feng Li

TL; DR We investigate a new problem setting OPT and propose DeCoOp to explore integrating out-of-distribution detection into the prompt tuning paradigm.

OPT Problem Setting

Definition

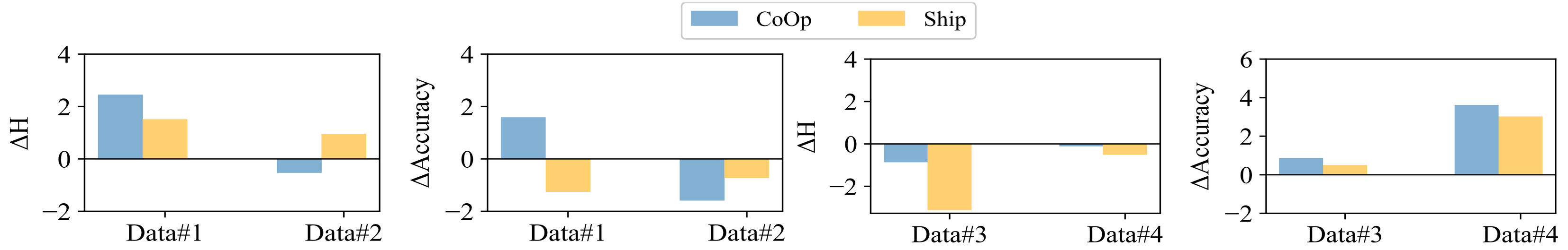
As illustrated in Figure 1, the **O**pen-world **P**rompt **T**uning (**OPT**) problem involves *tuning with only base class samples available, yet requiring classification of both base class and new class samples during testing*, with performance evaluated using accuracy metric.



▲ Figure 1: The overall illustration of OPT problem

Motivation

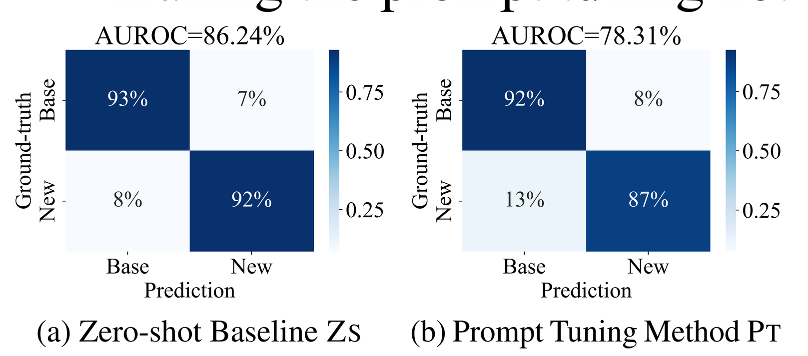
1. *The requirements to recognize new class samples emerges* in real-world applications, the , and these samples cannot be identified as a new class before testing.
2. *The performance of H and accuracy metrics are inconsistent.* Left subfigure of Figure 2 demonstrates that the improvement in the H metric corresponds to reduced accuracy, while right subfigure shows a deterioration in H is associated with increased accuracy.



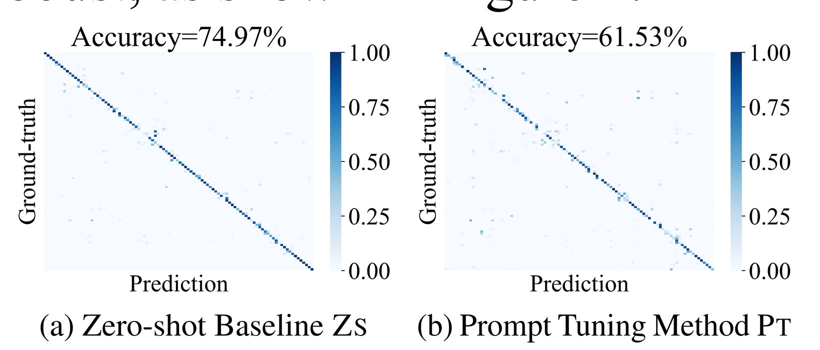
▲ Figure 2: Performance changes of different metrics

Challenges

1. *Existing methods and evaluating metrics ignore the base-to-new discriminability*, i.e., distinguishing whether a testing sample belongs to base classes and new classes. As shown in Figure 3, prompt tuning methods will degrades base-to-new discriminability.
2. *New-class discriminability degrades for prompt tuning methods*, making the prompt tuning not robust, as shown in Figure 4.



▲ Figure 3: Base-to-new discriminability



▲ Figure 4: New-class discriminability

DePt and DeCoOp Approach

DePt Framework

We propose a **De**composed **P**rompt **T**uning (**DePt**) framework, which integrates a zero-shot baseline P_{ZS}, a prompt tuning baseline P_{PT}, and an OOD detector P_{OOD} using the following formulation. The main idea is to *distinguish OOD samples and let zero-shot and prompt tuning methods handle the base classes and new classes respectively.*

$$\begin{cases} P_{\mathrm{PT}}(y|\boldsymbol{x}), & P_{\mathrm{OOD}}(y \in \mathcal{Y}_{\mathrm{b}}|\boldsymbol{x}) \geq P_{\mathrm{OOD}}(y \in \mathcal{Y}_{\mathrm{n}}|\boldsymbol{x}), \\ P_{\mathrm{ZS}}(y|\boldsymbol{x}), & P_{\mathrm{OOD}}(y \in \mathcal{Y}_{\mathrm{b}}|\boldsymbol{x}) < P_{\mathrm{OOD}}(y \in \mathcal{Y}_{\mathrm{n}}|\boldsymbol{x}). \end{cases}$$

Theoretical Analysis of DePt

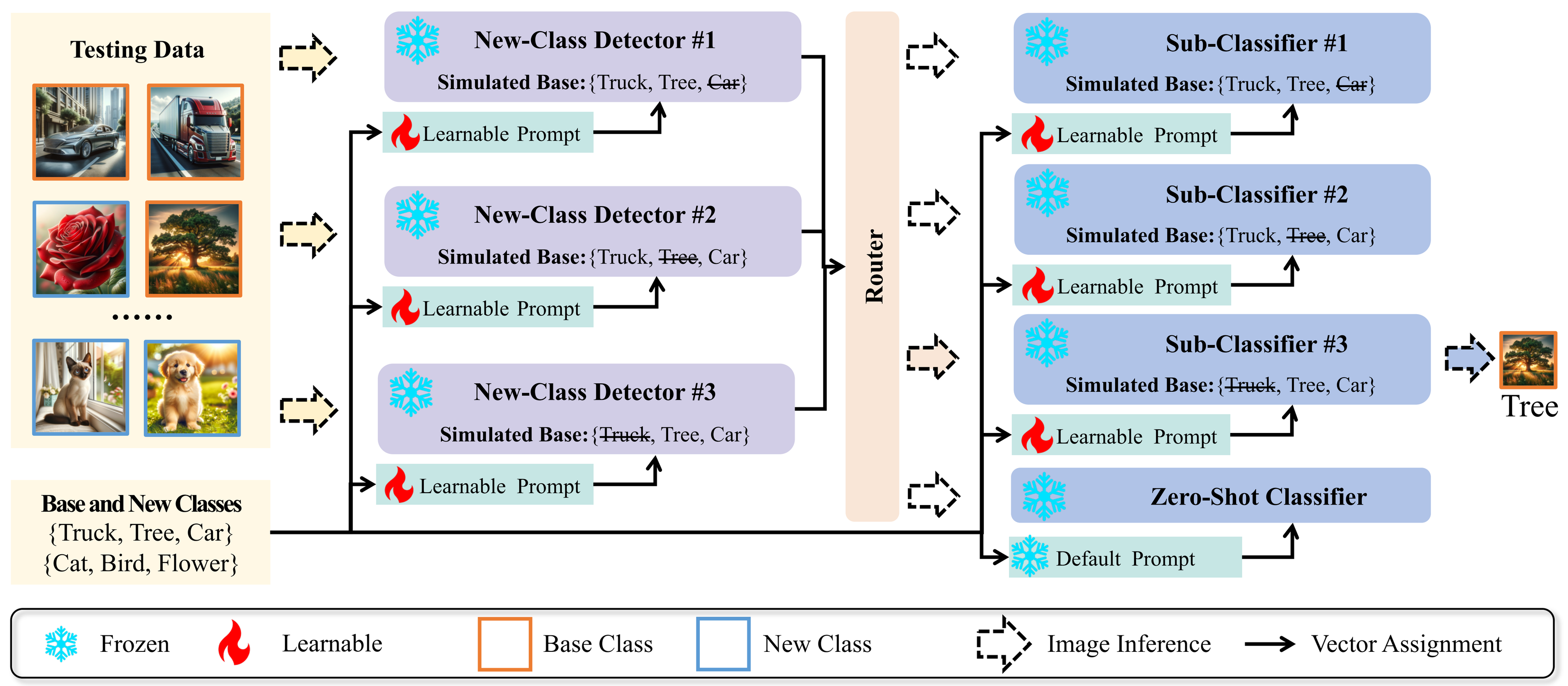
We prove that the DePt framework can achieve better performance compared to the zero-shot baseline, measuring their error using the cross-entropy metric.

Theorem 2.1. *If $\mathbb{E}_{\boldsymbol{x}}\left[H_{\mathrm{ZS}}^{\mathrm{CLS}}(\boldsymbol{x})\right] \leq \delta$ for $\boldsymbol{x}$ belonging to both base and new classes, $\mathbb{E}_{\boldsymbol{x}}\left[H_{\mathrm{PT}}^{\mathrm{CLS}}(\boldsymbol{x})\right] \leq \delta - \Delta$ for $\boldsymbol{x}$ belonging to base classes, and $\mathbb{E}_{\boldsymbol{x}}\left[H_{\mathrm{ZS}}^{\mathrm{OOD}}(\boldsymbol{x})\right] \leq \epsilon$, given a uniform mixing ratio ($\alpha : 1-\alpha$) of base classes and new classes in the testing data, we can determine that:*

$$\begin{cases} \mathbb{E}_{\boldsymbol{x}}\left[H_{\mathrm{ZS}}(\boldsymbol{x})\right] & \leq \epsilon + \delta, \\ \mathbb{E}_{\boldsymbol{x}}\left[H_{\mathrm{DePt}}(\boldsymbol{x})\right] & \leq \epsilon + \delta - \alpha \cdot \Delta. \end{cases}$$

DeCoOp Approach

Motivated by DePt framework, we propose a **De**composed **Co**ntext **Op**timization (**DeCoOp**) approach, shown in Figure 5. The main idea is to *train better OOD detector $\mathcal{M}_D$ using the leave-out strategy and train classifiers $\mathcal{M}_C$ for stronger generalization for new classes based on DePt framework.* The leave-out strategy address the challenge of lacking knowledge of new classes during training. The stronger generalization *of* $\mathcal{M}_C$ is achieved by simulating the emergence of new categories during training with the help of leave-out strategy.



▲ Figure 5: The overall illustration of DeCoOp approach

Experiments

Research Question #1

Can the empirical results of the DePt framework on real-world datasets conform to our theoretical analysis?

Method	ViT-B/16		ViT-B/32	
	New Acc.	Accuracy	New Acc.	Accuracy
Zs	65.49	63.92	63.95	60.36
Pt	57.73	65.57	53.01	61.03
DePt	**68.15**	**68.03**	**65.45**	**62.92**

▲ Table 1:Performance of DePt framework

Research Question #2

Can the DeCoOp method surpass existing baseline and SOTA methods, thereby demonstrating its robustness?

	Average		ImageNet		Caltech101		OxfordPets	
	H	Acc.	H	Acc.	H	Acc.	H	Acc.
Clip	70.84	63.92	70.20 ± 0.00	66.73 ± 0.00	95.41 ± 0.00	92.90 ± 0.00	92.93 ± 0.00	88.03 ± 0.00
Prompt Ens.	71.65	65.39	72.00 ± 0.00	68.48 ± 0.00	96.20 ± 0.00	94.08 ± 0.00	92.42 ± 0.00	86.37 ± 0.00
CoOp	72.14	65.57	64.95 ± 1.11	61.79 ± 1.09	95.96 ± 0.39	93.24 ± 0.68	95.38 ± 0.33	89.61 ± 0.34
CoCoOp	74.72	67.67	72.71 ± 0.33	69.41 ± 0.36	95.55 ± 0.24	93.43 ± 0.37	**95.71 ± 0.76**	**90.24 ± 1.32**
Ship	72.26	64.51	67.29 ± 0.38	63.65 ± 0.32	95.83 ± 0.23	92.93 ± 0.37	94.44 ± 0.54	86.78 ± 1.32
DeCoOp(Ours)	**76.13**	**69.69**	**72.98 ± 0.04**	**69.62 ± 0.08**	**96.52 ± 0.09**	**94.50 ± 0.22**	95.27 ± 0.08	88.87 ± 0.28

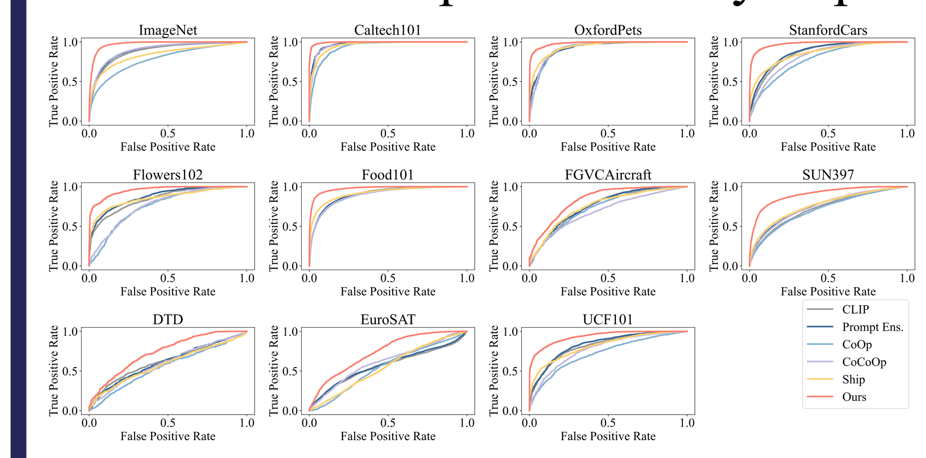
	StandfordCars		Flowers102		Food101		FGVCAircraft	
	H	Acc.	H	Acc.	H	Acc.	H	Acc.
Clip	68.75 ± 0.00	65.39 ± 0.00	72.74 ± 0.00	67.28 ± 0.00	90.18 ± 0.00	85.40 ± 0.00	30.25 ± 0.00	23.94 ± 0.00
Prompt Ens.	69.36 ± 0.00	65.95 ± 0.00	72.14 ± 0.00	67.03 ± 0.00	90.32 ± 0.00	85.54 ± 0.00	29.42 ± 0.00	23.31 ± 0.00
CoOp	68.22 ± 0.49	63.81 ± 0.44	78.33 ± 2.26	72.11 ± 2.36	86.65 ± 1.38	80.84 ± 1.50	29.38 ± 1.78	24.80 ± 1.23
CoCoOp	71.49 ± 0.62	67.75 ± 0.68	80.04 ± 1.46	71.95 ± 1.24	90.41 ± 0.24	85.61 ± 0.43	27.87 ± 11.36	21.46 ± 7.42
Ship	69.71 ± 0.43	64.67 ± 0.55	76.85 ± 2.18	70.40 ± 2.01	86.84 ± 1.49	77.39 ± 2.19	27.13 ± 1.10	24.44 ± 0.96
DeCoOp(Ours)	**73.24 ± 0.15**	**69.64 ± 0.19**	**84.16 ± 0.27**	**78.61 ± 0.59**	**90.68 ± 0.09**	**85.83 ± 0.07**	**31.44 ± 0.39**	**25.15 ± 0.31**

	SUN397		DTD		EuroSAT		UCF101	
	H	Acc.	H	Acc.	H	Acc.	H	Acc.
Clip	72.26 ± 0.00	62.57 ± 0.00	57.32 ± 0.00	44.56 ± 0.00	58.16 ± 0.00	41.40 ± 0.00	71.00 ± 0.00	64.97 ± 0.00
Prompt Ens.	75.04 ± 0.00	65.97 ± 0.00	59.63 ± 0.00	46.28 ± 0.00	58.45 ± 0.00	48.91 ± 0.00	73.17 ± 0.00	67.33 ± 0.00
CoOp	71.37 ± 1.21	61.82 ± 1.11	57.22 ± 2.37	48.18 ± 1.78	74.33 ± 4.35	59.65 ± 5.07	71.68 ± 2.84	65.41 ± 2.18
CoCoOp	77.17 ± 0.27	68.17 ± 0.33	60.59 ± 1.51	47.90 ± 1.43	73.77 ± 3.58	58.08 ± 1.49	76.59 ± 0.79	70.39 ± 1.25
Ship	72.57 ± 0.38	60.42 ± 0.48	56.82 ± 2.18	47.58 ± 1.62	73.29 ± 2.67	54.11 ± 1.73	74.09 ± 2.09	67.24 ± 1.94
DeCoOp(Ours)	**78.11 ± 0.09**	**69.33 ± 0.05**	**62.72 ± 1.23**	**51.44 ± 1.04**	**74.61 ± 3.82**	**61.90 ± 3.72**	**77.67 ± 0.50**	**71.71 ± 0.79**

▲ Table 2: Performance of DeCoOp approach

Research Question #3

Does the DeCoOp successfully improve the base-to-new discriminability?



▲ Figure 6: AUROC of OOD detection

Dataset	Clip	CoCoOp	Ship	DeCoOp(Ours)
ImageNet	88.34	88.05	84.71	**97.48**
Caltech101	97.03	95.71	96.94	**99.58**
OxfordPets	92.66	91.15	93.30	**98.12**
StanfordCars	86.24	83.00	87.23	**97.63**
Flowers102	84.92	79.63	84.84	**95.75**
Food101	89.88	88.19	89.92	**97.59**
FGVCAircraft	75.08	69.00	75.78	**84.06**
SUN397	72.46	73.75	74.78	**90.21**
DTD	62.29	60.65	60.66	**75.47**
EuroSAT	56.40	57.74	59.32	**77.78**
UCF101	82.03	79.03	80.35	**93.56**
Average	80.67	78.72	80.71	**91.57**

▲ Table 3: AUROC of OOD detection

- ✓ If you are interested in this paper, please feel free to contact Zhi Zhou (zhouz@lamda.nju.edu.cn) or visit our project homepage for more details (https://wnjxyk.github.io/DeCoOp) .
- ✓ This research was supported by National Science and Technology Major Project (2022ZD0114803) and the National Science Foundation of China (62306133, 62176118).