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

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

Nanjing University, Nanjing, China

Learning And Mining from DatA

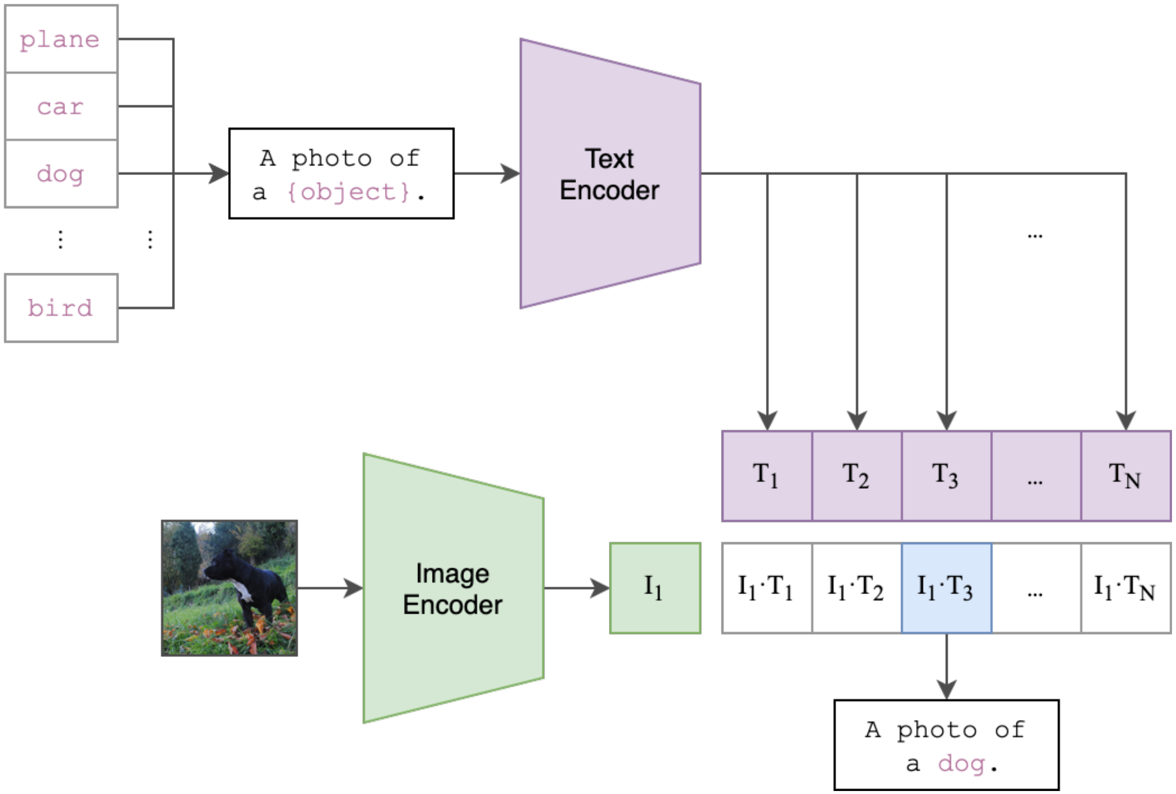
† Corresponding author

Outline

- **OPT, Open-world Prompt Tuning Problem Setting**
- DePt, Decomposed Prompt Tuning Framework
- DeCoOp, Decomposed Context Optimization Approach
- Conclusion

Vision-Language Models

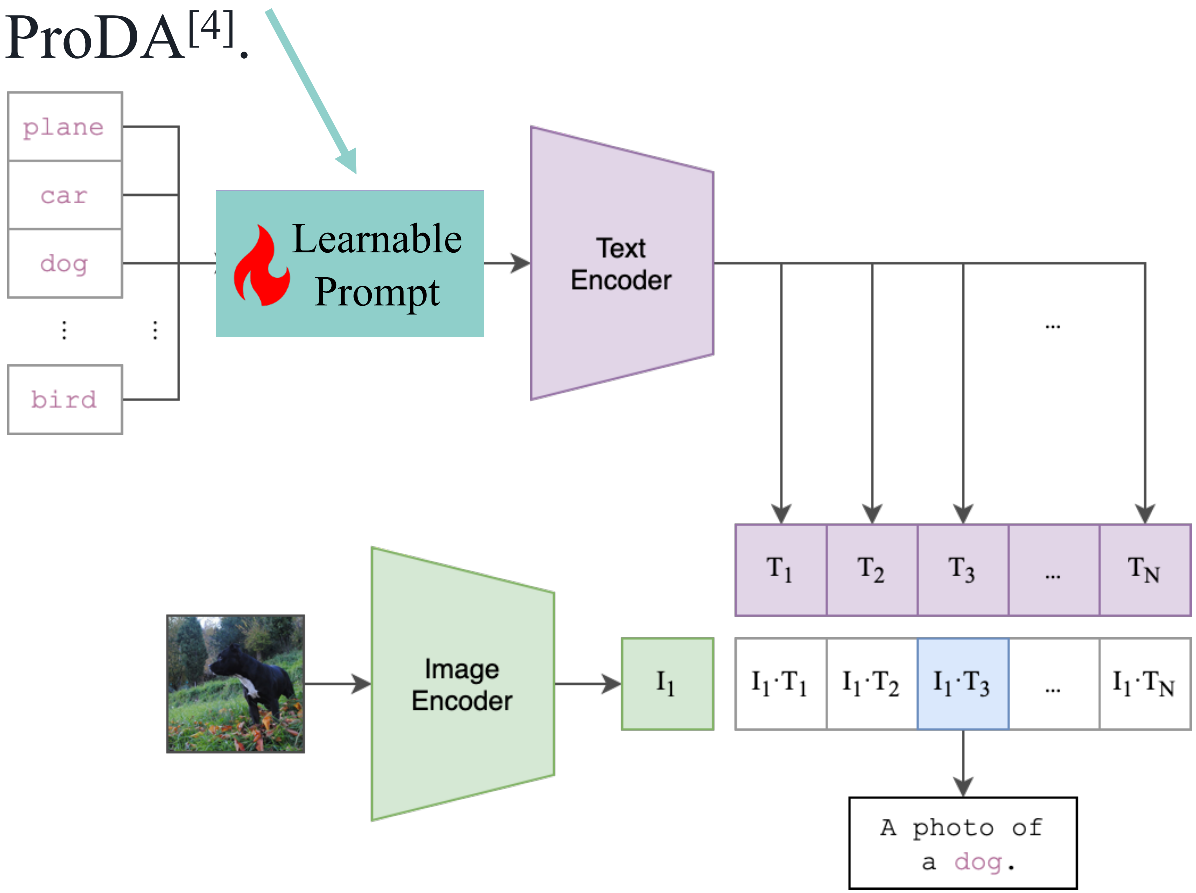
Vision-language models, like CLIP[1], demonstrate remarkable zero-shot classification capability for downstream tasks by calculating the similarity between images and class names.



[1] Learning Transferable Visual Models From Natural Language Supervision. ICML 2021.

Few-Shot Prompt Tuning

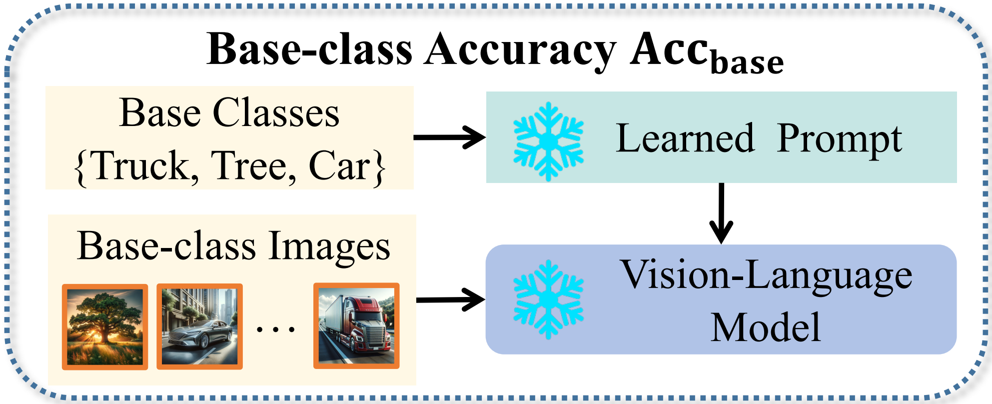
This classification capability can be further enhanced through tuning the **prompt template** using few labeled samples, e.g., CoOp[2], CoCoOp[3], ProDA[4].



[2] Learning to Prompt for Vision-Language Models. IJCV.
[3] Conditional Prompt Learning for Vision-Language Models. CVPR 2022.
[4] Prompt Distribution Learning. CVPR 2022.

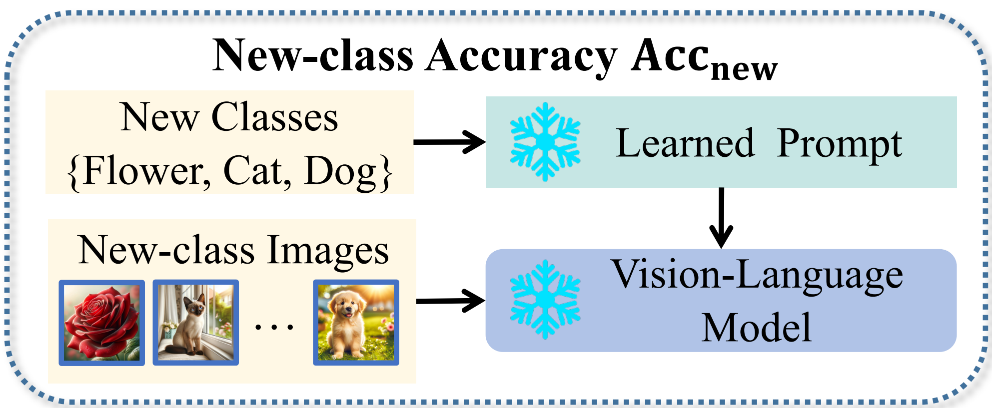
Evaluation Protocols

Existing evaluation protocols focus on separately measuring the classification capabilities of both seen classes (base classes) and unseen classes (new classes), as well as their harmonic mean.



Harmonic Mean of Accuracy

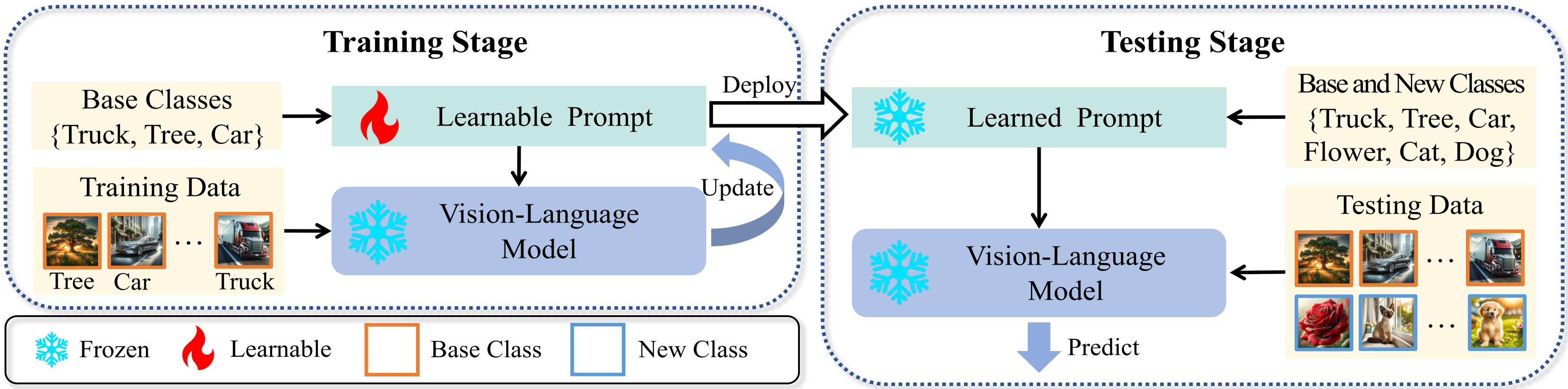
$$\frac{2 \times \mathrm{Acc}_{\mathrm{base}} \times \mathrm{Acc}_{\mathrm{new}}}{\mathrm{Acc}_{\mathrm{base}} + \mathrm{Acc}_{\mathrm{new}}}$$

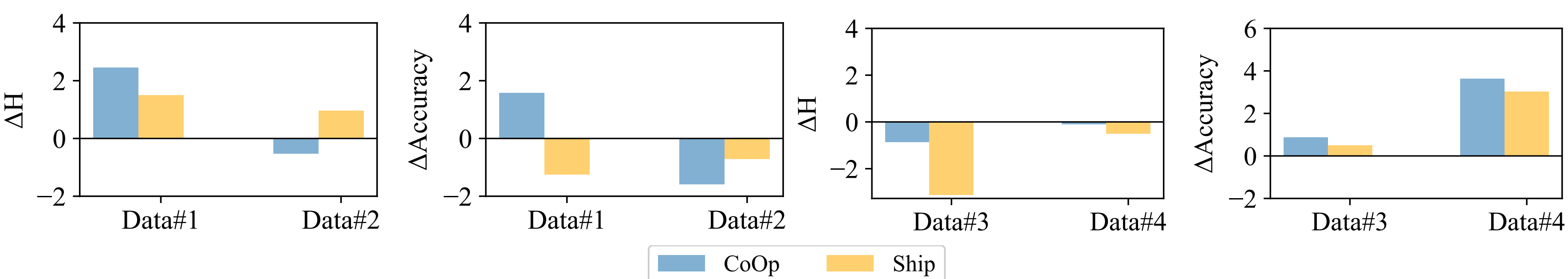


These metrics fail to measure **the base-to-new discriminability** of tuned prompt.

OPT Problem Setting

Open-world **P**rompt **T**uning (**OPT**) problem evaluates the accuracy of base-class and new-class images together, measuring ***base-class, new-class, and base-to-new discriminability*** at the same time. The inconsistency between H and our metrics highlights necessity of OPT problem setting.





Outline

Learning And Mining from DatA

- OPT, Open-world Prompt Tuning Problem Setting
- **DePt, Decomposed Prompt Tuning Framework**
- DeCoOp, Decomposed Context Optimization Approach
- Conclusion

DePt Framework

We decompose the original problem into two **classification problems** and one **OOD detection problem**.

$$P(y|\boldsymbol{x}) = \sum_{i \in \{\mathrm{b},\mathrm{n}\}} \boxed{P(y|y \in \mathcal{Y}_i, \boldsymbol{x})} \cdot \boxed{P(y \in \mathcal{Y}_i|\boldsymbol{x})}$$

We integrate 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***.

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

DePt Framework

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:*

$$\begin{cases} \mathbb{E}_{\boldsymbol{x}}\left[H_{\mathrm{ZS}}(\boldsymbol{x})\right] & \leq \epsilon + \delta, \\ \mathbb{E}_{\boldsymbol{x}}\left[H_{\mathrm{DEPT}}(\boldsymbol{x})\right] & \leq \epsilon + \delta - \alpha \cdot \Delta. \end{cases} \tag{5}$$

DePt Framework

To verify our theorem, we conducted experiments on 11 datasets using ViT-B/16 and ViT-B/32 architectures. The experimental result suggests that the DePt framework effectively mitigates performance degradation on new classes through the utilization of the OOD detector, which aligns well with 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**

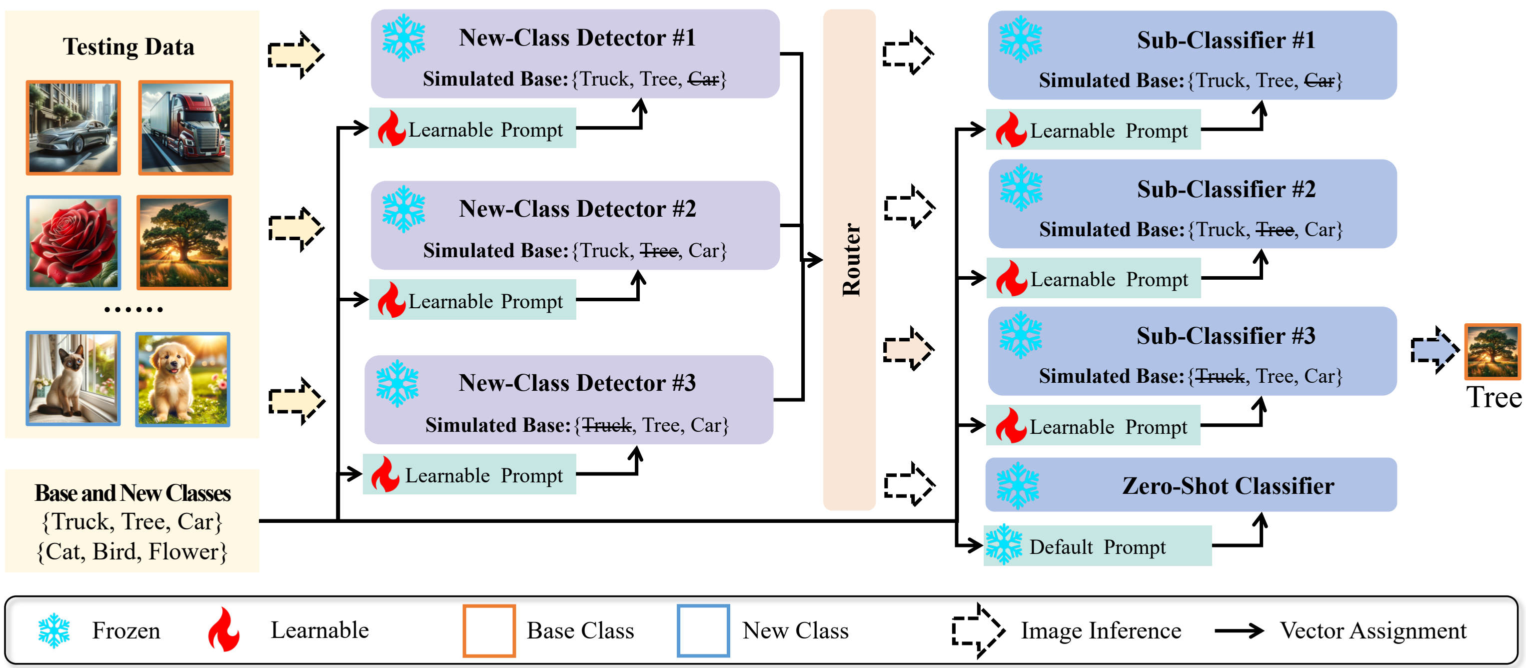
Outline

- OPT, Open-world Prompt Tuning Problem Setting
- DePt, Decomposed Prompt Tuning Framework
- **DeCoOp, Decomposed Context Optimization Approach**
- Conclusion

DeCoOp Approach

Learning And Mining from DatA

Motivated by DePt framework, we propose a **De**composed **Co**ntext **Op**timization (**DeCoOp**) approach, shown in following figure. 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.*



DeCoOp Approach

We evaluate the DeCoOp approach on 11 datasets using ViT-B/16 architecture. Our DeCoOp approach demonstrates superior average performance on both the H metric and Accuracy, showcasing 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

	StanfordCars		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**

DeCoOp Approach

Learning And Mining from DatA

We evaluate the base-to-new discriminability of our approach and selected methods using the MSP method with the ViT-B/16 architecture. The results clearly indicate that our DeCoOp approach significantly improves base-to-new discriminability, which accounts for its SOTA performance.

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**

Outline

- OPT, Open-world Prompt Tuning Problem Setting
- DePt, Decomposed Prompt Tuning Framework
- DeCoOp, Decomposed Context Optimization Approach
- **Conclusion**

Conclusions

Learning And Mining from DatA

- We investigate a novel OPT problem setting.
- We introduce the DePt framework, exploring the integration of OOD detection into prompt tuning.
- We propose the DeCoOp approach to achieve state-of-the-art performance on the OPT problem.

➢ 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).