Do-Prompt: Causal Interventions Meet Variational Prompt Bottlenecks

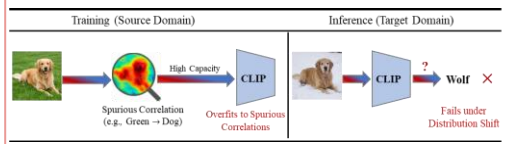
Xueting Chen, Jun-Jie Huang, Yan Yan, Long Lan, Yuhua Tang, Wenjing Yang
[1] National University of Defense Technology [2] Xiamen University

ICML International Conference On Machine Learning

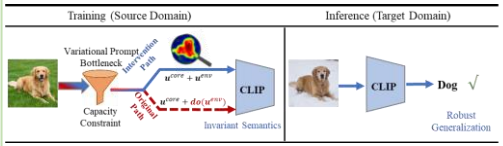
1.Motivation

Why standard prompt tuning fails under shift?

(a) Standard Prompt Tuning (Baseline)



(b) Do-Prompt (Ours)



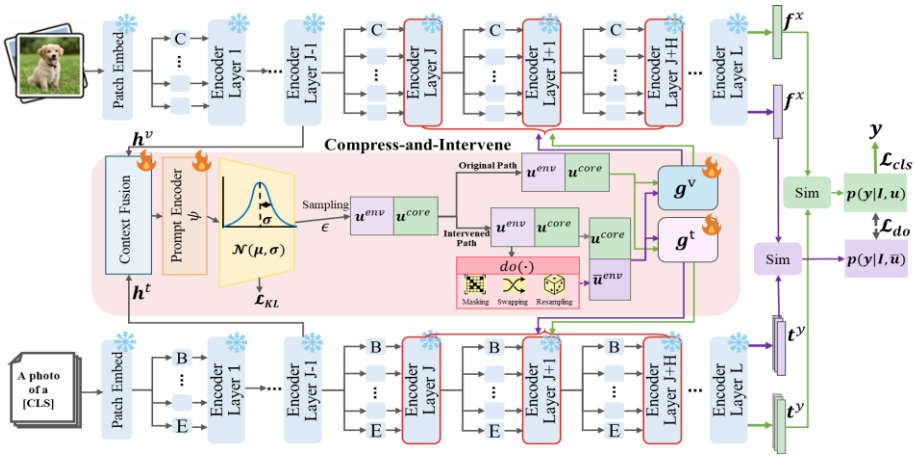
Why Do-Prompt?

- Prompts can become a high-capacity shortcut pathway that encodes environment-specific cues.
- This hurts transfer under distribution shift, especially in multi-modal prompt learning.
- Do-Prompt constrains prompt capacity and intervenes on environment-related prompt content.

Our Contributions

- Variational prompt bottleneck for stochastic, information-constrained prompts.
- Prompt-level causal interventions on the environment branch (mask / swap / resample).
- Plug-and-play integration with existing multi-modal prompt tuning pipelines.

2.Method: Do-Prompt Framework



Key Formulations

$$q_\phi(\mathbf{u}_i \mid \mathbf{h}) = \mathcal{N}\left(\boldsymbol{\mu}_i(\mathbf{h}), \mathrm{diag}\left(\boldsymbol{\sigma}_i^2(\mathbf{h})\right)\right)$$

$$\mathbf{u}_i = [\mathbf{u}_i^{\mathrm{core}}; \mathbf{u}_i^{\mathrm{env}}]$$

$$\mathcal{L}_{\mathrm{KL}} = \sum_{J}^{J+H-1} \mathrm{KL}\left(q_\phi(\mathbf{u}_i \mid \mathbf{h}) \,\|\, p(\mathbf{u}_i)\right)$$

$$\mathcal{L}_{\mathrm{do}} = \mathbb{E}_k\left[\mathrm{KL}\left(p_\theta(y \mid I; \{\mathbf{u}_i\}) \,\|\, p_\theta\left(y \mid I; \{\bar{\mathbf{u}}_i^{(k)}\}\right)\right)\right]$$

What Do-Prompt Does

- VIB limits channel capacity.
- *do*-interventions perturb only environment-related prompt components.
- Consistency regularization encourages invariant semantics.

3.Main Results

A. Base to novel generalization results

Method	Average			ImageNet			Caltech101			OxfordPets		
	Base	Novel	HM	Base	Novel	HM	Base	Novel	HM	Base	Novel	HM
CoOp (Zhou et al., 2022a)	82.69	63.22	70.83	76.47	67.88	71.92	98.00	89.81	93.73	93.67	95.29	94.47
CoCoOp (Zhou et al., 2022b)	80.47	71.69	75.83	75.98	70.43	73.10	97.96	93.81	95.84	95.20	97.69	96.43
APP (Cho et al., 2024)	83.0	65.8	72.61	69.9	63.2	66.4	95.2	91.0	93.0	96.8	88.3	92.4
PromptSRC* (Khattak et al., 2023b)	84.93	74.49	78.61	76.77	67.8	72.01	98.07	94.03	96.01	95.27	97.23	96.24
MaPLe* (Khattak et al., 2023a)	82.03	75.03	78.37	74.96	66.97	70.74	97.83	94.87	96.33	95.20	98.13	96.64
MaPLe* + Do-Prompt (Ours)	**83.82**	**76.68**	**80.12**	76.06	68.80	72.25	98.35	94.87	96.58	**96.17**	97.77	**96.96**
Improvements Δ	+1.79	+1.65	+1.75	+1.10	+1.83	+1.51	+0.52	0.0	+0.25	+0.97	-0.36	+0.32
MMRL* (Guo & Gu, 2025)	85.54	76.52	80.59	77.55	67.43	72.14	98.93	**94.60**	**96.72**	95.27	97.23	96.24
MMRL* + Do-Prompt (Ours)	**86.38**	**79.16**	**82.56**	**77.90**	**68.83**	**73.09**	**99.10**	94.27	96.62	95.93	97.57	96.74
Improvements Δ	+0.84	+2.64	+1.97	+0.35	+1.40	+0.95	+0.17	-0.33	-0.10	+0.66	+0.34	+0.50

Method	Source	Target										
	ImageNet	Caltech101	OxfordPets	StanfordCars	Flowers102	Food101	FGVCAircraft	SUN397	DTD	EuroSAT	UCF101	Average
MaPLe* (Khattak et al., 2023a)	67.96	93.17	90.20	65.97	71.07	86.33	23.23	67.23	47.20	45.70	66.27	65.63
MaPLe* + Do-Prompt (Ours)	**68.66**	**93.80**	**90.73**	**66.30**	**74.57**	**86.40**	**25.37**	**67.70**	**48.17**	**54.90**	**70.87**	**67.19**
Improvements Δ	+0.70	+0.63	+0.53	+0.33	+3.50	+0.07	+2.14	+0.47	+0.97	+9.20	+4.60	+1.56
MMRL* (Guo & Gu, 2025)	70.13	94.30	91.00	66.20	71.53	86.20	26.07	67.50	46.93	49.83	69.13	66.87
MMRL* + Do-Prompt (Ours)	**71.00**	**94.60**	**91.97**	**66.63**	**75.57**	**86.43**	**28.33**	**67.90**	**48.27**	**58.97**	**72.50**	**68.64**
Improvements Δ	+0.87	+0.30	+0.97	+0.43	+4.04	+0.23	+2.26	+0.40	+1.34	+9.14	+3.37	+1.77

C. Base to novel generalization results

Method	Source	Target			
	ImageNet	ImageNetV2	ImageNet-S	ImageNet-A	ImageNet-R
MaPLe* (Khattak et al., 2023a)	67.96	61.57	47.70	48.80	75.33
MaPLe* + Do-Prompt (Ours)	**68.66**	**62.80**	**49.20**	**50.30**	**76.53**
Improvements Δ	+0.70	+1.23	+1.50	+1.50	+1.20
MMRL* (Guo & Gu, 2025)	70.13	62.20	47.80	48.90	75.03
MMRL* + Do-Prompt (Ours)	**71.00**	**63.80**	**49.70**	**50.80**	**76.83**
Improvements Δ	+0.87	+1.60	+1.90	+1.90	+1.80

4.Ablation & Analysis

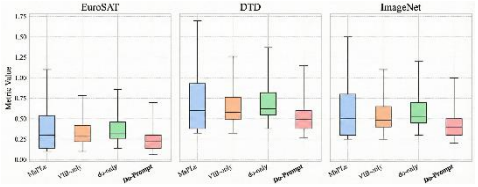
A. Ablation on key components

Variant	VIB	do-int.	HM
MaPLe* baseline			78.37
+ VIB only	✓		78.93
+ do-intervention only		✓	79.02
MaPLe* + Do-Prompt	✓	✓	**80.12**
MMRL* baseline			80.59
+ VIB only	✓		81.32
+ do-intervention only		✓	81.43
MMRL* + Do-Prompt	✓	✓	**82.56**

B. Intervention operators

Intervention	Base	Novel	Avg HM
None (MaPLe)	82.03	75.03	78.37
Mask only	83.01	75.97	79.01
Swap only	82.78	75.75	78.94
Resample only	83.06	75.82	79.15
All (Do-Prompt)	**83.82**	**76.68**	**80.12**

C. Prediction invariance under prompt interventions



D. Overhead of Do-Prompt

Method	Train Time / Iter	Inference Time / Img	Params (Δ)
MaPLe (Khattak et al., 2023a)	1.00×	1.00×	–
MaPLe + Do-Prompt	1.8×	1.02×	+4.0M
MMRL (Guo & Gu, 2025)	1.00×	1.00×	–
MMRL + Do-Prompt	1.9×	1.03×	+7.0M