Holcombe Department of
ELECTRICAL AND
COMPUTER ENGINEERING
Clemson University

DiSa: Directional Saliency-Aware Prompt Learning for Generalizable Vision-Language Models

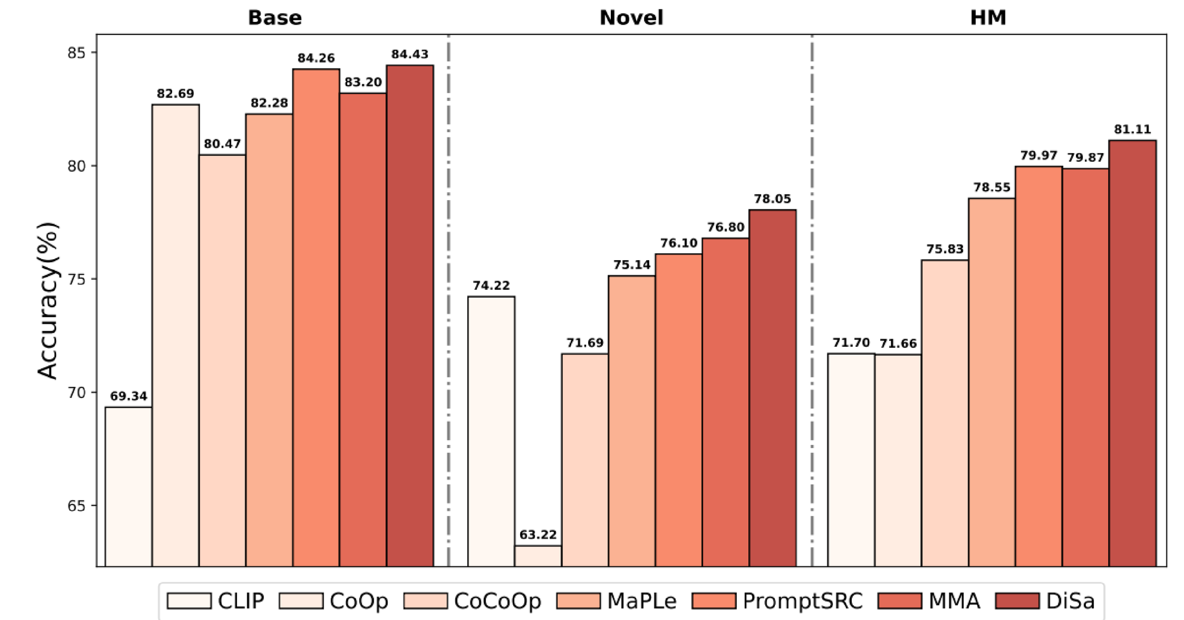
Niloufar Alipour Talemi, Hossein Kashiani, Hossein R. Nowdeh, Fatemeh Afghah
Clemson University

KDD2025
August 3–7, 2025

Challenges

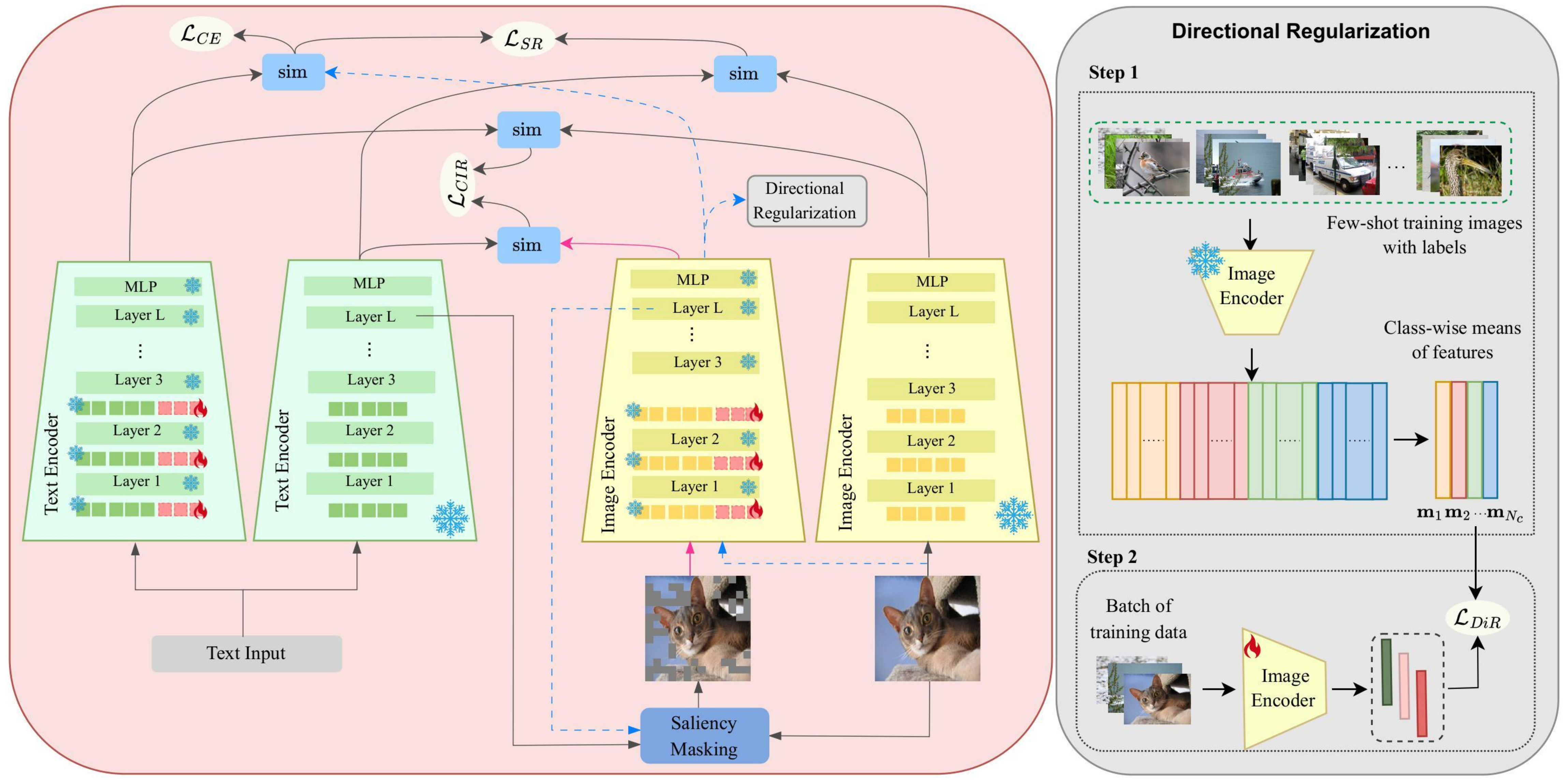
- Few-shot fine-tuning often leads to overfitting when optimizing prompts for task-specific objectives, restricting the model's ability to generalize beyond the training samples.
- This overfitting poses a significant challenge for adapting vision-language models to new domains or unseen classes within the same domain.

Contributions



- DiSa introduces CIR, a novel regularization-based prompt learning framework that promotes interaction between the modality-specific branches of prompted and frozen models.
- We introduce a novel directional regularization approach that aligns the prompted features with class-wise prototypes, represented as mean embeddings from the frozen model.
- extensive evaluations on 11 popular image classification benchmarks demonstrate the effectiveness of DiSa in all the base-to-novel generalization, cross-dataset transfer, domain generalization, and few-shot learning settings.

DiSa Framework



Experimental Results

Base-to-novel generalization evaluation.

(a) Average over 11 datasets

	Base	New	HM
CLIP [29]	69.34	74.22	71.70
CoOp [45]	82.69	63.22	71.66
CoCoOp [44]	80.47	71.69	75.83
Maple [19]	82.28	75.14	78.55
PromptSRC [20]	84.26	76.10	79.97
CoPrompt [33]	84.00	77.23	80.48
MMA [39]	83.20	76.80	79.87
APEX [40]	83.99	76.76	80.04
TCP [41]	84.13	75.36	79.51
DiSa (Ours)	**84.43**	**78.05**	**81.11**

(b) ImageNet

	Base	New	HM
CLIP[29]	72.43	68.14	70.22
CoOp[45]	76.47	67.88	71.92
CoCoOp[44]	75.98	70.43	73.10
Maple [19]	76.66	70.54	73.47
PromptSRC [20]	77.60	70.73	74.01
CoPrompt [33]	**77.67**	71.27	74.33
MMA [39]	77.31	71.00	74.02
APEX [40]	77.12	71.10	73.99
TCP [41]	77.27	69.87	73.38
DiSa (Ours)	77.56	**71.65**	**74.49**

(c) Caltech101

	Base	New	HM
CLIP[29]	96.84	94.00	95.40
CoOp[45]	98.00	89.81	93.73
CoCoOp [44]	97.96	93.81	95.84
Maple [19]	97.74	94.36	96.02
PromptSRC [20]	98.10	94.03	96.02
CoPrompt [33]	98.27	94.90	96.55
MMA [39]	**98.40**	94.00	96.15
APEX [40]	98.18	95.06	96.59
TCP [41]	98.23	94.67	96.42
DiSa (Ours)	98.29	**95.41**	**96.83**

(d) OxfordPets

	Base	New	HM
CLIP[29]	91.17	97.26	94.12
CoOp[45]	93.67	95.29	94.47
CoCoOp[44]	95.20	97.69	96.43
Maple [19]	95.43	97.76	96.58
PromptSRC [20]	95.33	97.30	96.30
CoPrompt [33]	**95.67**	98.10	96.87
MMA [39]	95.40	98.07	96.72
APEX [40]	95.11	97.27	96.18
TCP [41]	94.67	97.20	95.92
DiSa (Ours)	95.48	**98.67**	**97.05**

(e) StanfordCars

	Base	New	HM
CLIP[29]	63.37	74.89	68.65
CoOp[45]	78.12	60.40	68.13
CoCoOp[44]	70.49	73.59	72.01
Maple [19]	72.94	74.00	73.47
PromptSRC [20]	78.27	74.97	76.58
CoPrompt [33]	76.97	74.40	75.66
MMA [39]	78.50	73.10	75.70
APEX [40]	80.53	**75.08**	**77.71**
TCP [41]	**80.80**	74.13	77.32
DiSa (Ours)	78.54	75.07	76.77

(f) Flowers102

	Base	New	HM
CLIP[29]	72.08	77.80	74.83
CoOp[45]	97.60	59.67	74.06
CoCoOp[44]	94.87	71.75	81.71
Maple [19]	95.92	72.46	82.56
PromptSRC [20]	98.07	76.50	85.95
CoPrompt [33]	97.27	76.60	85.71
MMA [39]	97.77	75.93	85.48
APEX [40]	97.47	**77.58**	**86.40**
TCP [41]	97.73	75.57	85.23
DiSa (Ours)	**98.14**	76.77	86.15

(g) Food101

	Base	New	HM
CLIP[29]	92.43	91.22	90.66
CoOp[45]	88.33	82.26	85.19
CoCoOp[44]	90.70	91.29	90.99
Maple [19]	90.71	92.05	91.38
PromptSRC [20]	90.67	91.53	91.10
CoPrompt [33]	90.73	92.07	91.40
MMA [39]	90.13	91.30	90.71
APEX [40]	89.60	92.06	90.81
TCP [41]	90.57	91.37	90.97
DiSa (Ours)	**90.81**	**92.32**	**91.56**

(h) FGVCAircraft

	Base	New	HM
CLIP[29]	27.19	36.29	31.09
CoOp[45]	40.44	22.30	28.75
CoCoOp[44]	33.41	23.71	27.74
Maple [19]	37.44	35.61	36.50
PromptSRC [20]	**42.73**	37.87	40.15
CoPrompt [33]	40.20	39.33	39.76
MMA [39]	40.57	36.33	38.33
APEX [40]	42.69	35.21	38.59
TCP [41]	41.97	34.43	37.83
DiSa (Ours)	42.65	**39.38**	**40.95**

(i) SUN397

	Base	New	HM
CLIP[29]	69.36	75.35	72.23
CoOp [45]	80.60	65.89	72.51
CoCoOp[44]	79.74	76.86	78.27
Maple [19]	80.82	78.70	79.75
PromptSRC [20]	82.67	78.47	80.52
CoPrompt [33]	82.63	80.03	81.31
MMA [39]	82.27	78.57	80.38
APEX [40]	81.17	78.98	80.06
TCP [41]	82.63	78.20	80.35
DiSa (Ours)	**82.69**	**80.53**	**81.60**

(j) DTD

	Base	New	HM
CLIP[29]	53.24	59.90	56.37
CoOp[45]	79.44	41.18	54.24
CoCoOp[44]	77.01	56.00	64.85
Maple [19]	80.36	59.18	68.16
PromptSRC [20]	**83.37**	62.97	71.75
CoPrompt [33]	83.13	64.73	72.79
MMA [39]	83.20	65.63	73.38
APEX [40]	82.45	63.80	71.94
TCP [41]	82.77	58.07	68.25
DiSa (Ours)	83.33	**65.71**	**73.49**

(k) EuroSAT

	Base	New	HM
CLIP[29]	56.48	64.05	60.03
CoOp[45]	92.19	54.74	68.69
CoCoOp [44]	87.49	60.04	71.21
Maple [19]	94.07	73.23	82.35
PromptSRC [20]	92.90	73.90	82.32
CoPrompt [33]	**94.60**	78.57	85.84
MMA [39]	85.46	82.34	83.87
APEX [40]	92.83	79.89	85.85
TCP [41]	91.63	74.73	82.32
DiSa (Ours)	94.10	**82.69**	**88.03**

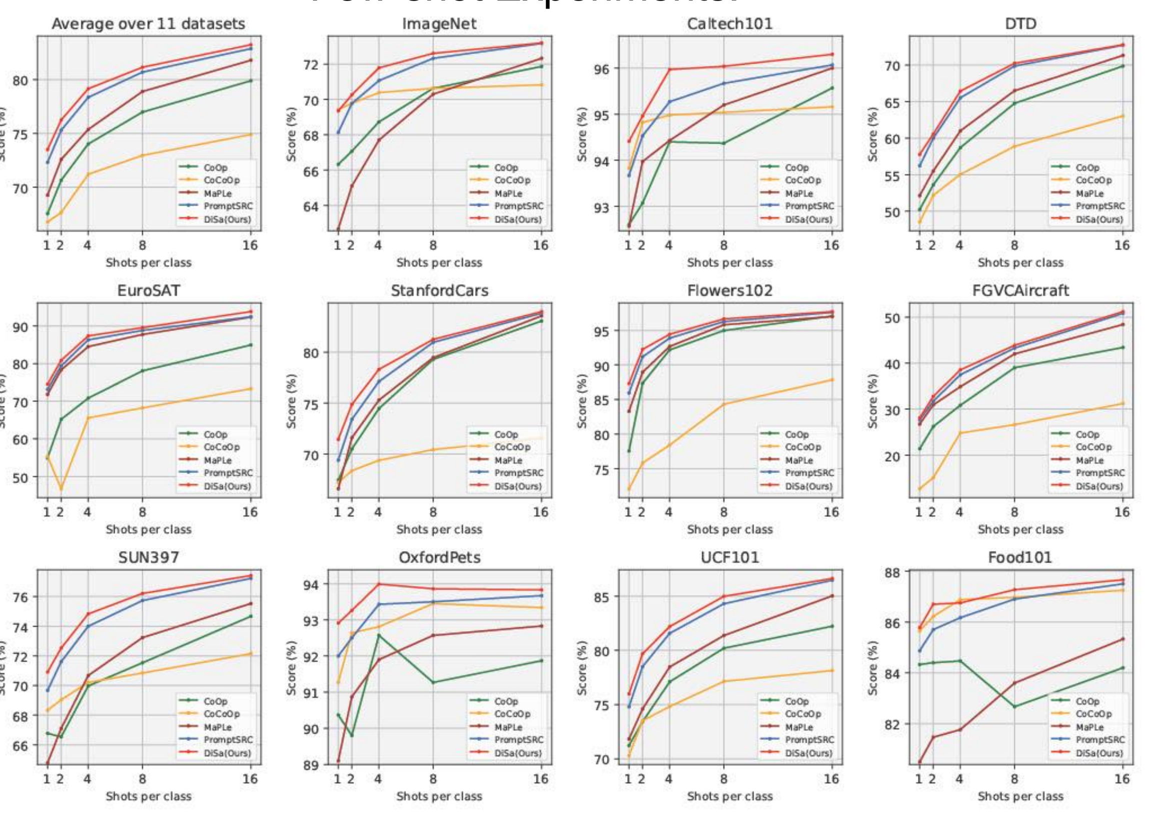
(l) UCF101

	Base	New	HM
CLIP [29]	70.53	77.50	73.85
CoOp [45]	84.69	56.05	67.46
CoCoOp[44]	82.33	73.45	77.64
Maple [19]	83.00	78.66	80.77
PromptSRC [20]	87.10	78.80	82.74
CoPrompt [33]	86.90	79.57	83.07
MMA [39]	86.23	80.03	82.20
APEX [40]	86.74	78.37	82.34
TCP [41]	87.13	**80.77**	**83.83**
DiSa (Ours)	**87.16**	80.45	83.67

Cross-dataset evaluation.

	Source	Target										
	ImageNet	Caltech101	OxfordPets	StanfordCars	Flowers102	Food101	Aircraft	SUN397	DTD	EuroSAT	UCF101	Average
CoOp [45]	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 [44]	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 [19]	70.72	93.53	90.49	65.57	72.23	86.20	24.74	67.01	46.49	48.06	68.69	66.30
PromtSCR [20]	71.27	93.60	90.25	65.70	70.25	86.15	23.90	67.10	46.87	45.50	68.75	65.81
CoPrompt [33]	70.80	94.50	90.73	65.67	72.30	86.43	24.00	67.57	47.07	**51.90**	69.73	67.00
MMA [39]	71.00	93.80	90.30	66.13	72.07	86.12	**25.33**	68.17	46.57	49.24	68.32	66.61
APEX [40]	**72.00**	94.46	90.06	65.46	71.58	86.44	24.44	67.20	45.70	47.58	68.80	66.16
TCP [41]	71.40	93.97	**91.25**	64.69	71.21	**86.69**	23.45	67.15	44.35	51.45	68.73	66.29
DiSa	71.21	**94.62**	90.94	**66.22**	**72.51**	86.64	25.26	**68.32**	**47.23**	50.84	**69.93**	**67.25**

Few-shot Experiments.



Methodology

The DiSa employs two complementary regularization approaches: saliency-aware cross-interactive regularization and directional regularization.

Cross-Interactive Regularization Loss:

$$\mathcal{L}_{CIR} = \mathcal{D}_{\mathrm{KL}}(q^{f_p g_o}, q^{f_o g_p}),$$

$$q^{f_p g_o} = \mathrm{sim}(\mathbf{f}_p, \mathbf{g}_o),\; q^{f_o g_p} = \mathrm{sim}(\mathbf{f}_o, \mathbf{g}_p),$$

Directional Regularization Loss:

$$\mathcal{L}_{DiR} = \left|1 - \cos(\mathbf{f}_p, \mathbf{m}_i)\right|, \quad \mathbf{m}_i = \frac{1}{|I_j|}\sum_{j \in I_j} \mathbf{f}_{o_j},$$

Score-based Loss:

$$\mathcal{L}_{SR} = \mathcal{D}_{\mathrm{KL}}(q^{f_p g_p}, q^{f_o g_o}),$$

$$q^{f_p g_p} = \mathrm{sim}(\mathbf{f}_p, \mathbf{g}_p),\; q^{f_o g_o} = \mathrm{sim}(\mathbf{f}_o, \mathbf{g}_o).$$

Total Loss:

$$\mathcal{L}_{total} = \mathcal{L}_{CE} + \mathcal{L}_{SR} + \mathcal{L}_{CIR} + \lambda\mathcal{L}_{DiR},$$

Experiments

- We validate our method across four different settings: generalization from base-to-novel, classes, cross-dataset evaluation, domain generalization, and few-shot learning.
- For base-to-novel, cross-dataset and few-shot experiments: Generic-object datasets (ImageNet and Caltech101), Fine-grained datasets (Oxford Pets, Stanford Cars, Flowers102, Food101, and FGVC Aircraft), remote sensing classification dataset (EuroSAT), scene recognition dataset (SUN397), Action recognition dataset (UCF101), Texture dataset (DTD). For domain generalization experiments: ImageNetV2, ImageNet Sketch, ImageNet-A, ImageNet-R.
- Ablations studies proves that components complement each other to mitigate overfitting in vision-language model adaptation, leading to improved generalization performance.

Analysis of the effectiveness of each component in DiSa.

Approach					Accuracy		
$\mathcal{L}_{CIR}$	*Masking*	$\mathcal{L}_{SR}$	$\mathcal{L}_{DiR}$ Sample	$\mathcal{L}_{DiR}$ Prototype	Base	Novel	HM
					84.21	71.79	77.51
✓					84.35	74.23	78.97
✓	✓				84.31	74.86	79.30
✓	✓	✓			84.27	76.53	80.21
✓	✓	✓	✓		84.25	77.09	80.51
✓	✓	✓		✓	**84.43**	**78.05**	**81.11**

Comparison of feature alignment strategies.