URL @KAIST
URBAN ROBOTICS LAB

CoCoA-Mix: Confusion-and-Confidence-Aware Mixture Model for Context Optimization

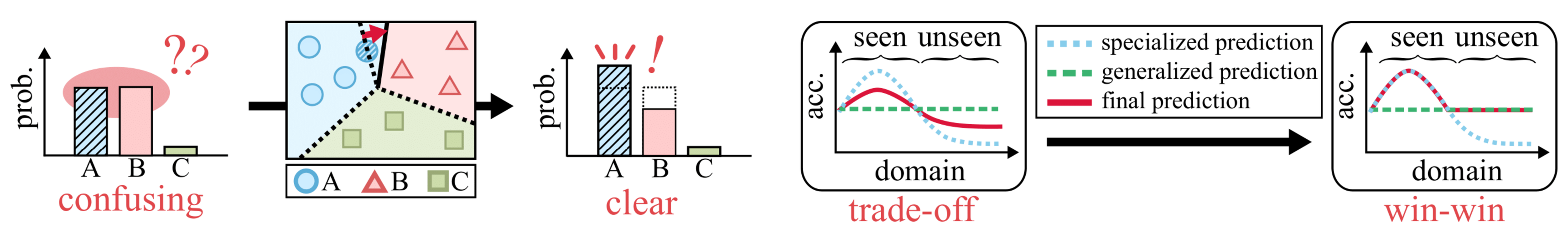
Dasol Hong[1], Wooju Lee[1], Myung Hyun[1]
[1]Urban Robotics Lab., KAIST

ICML International Conference On Machine Learning

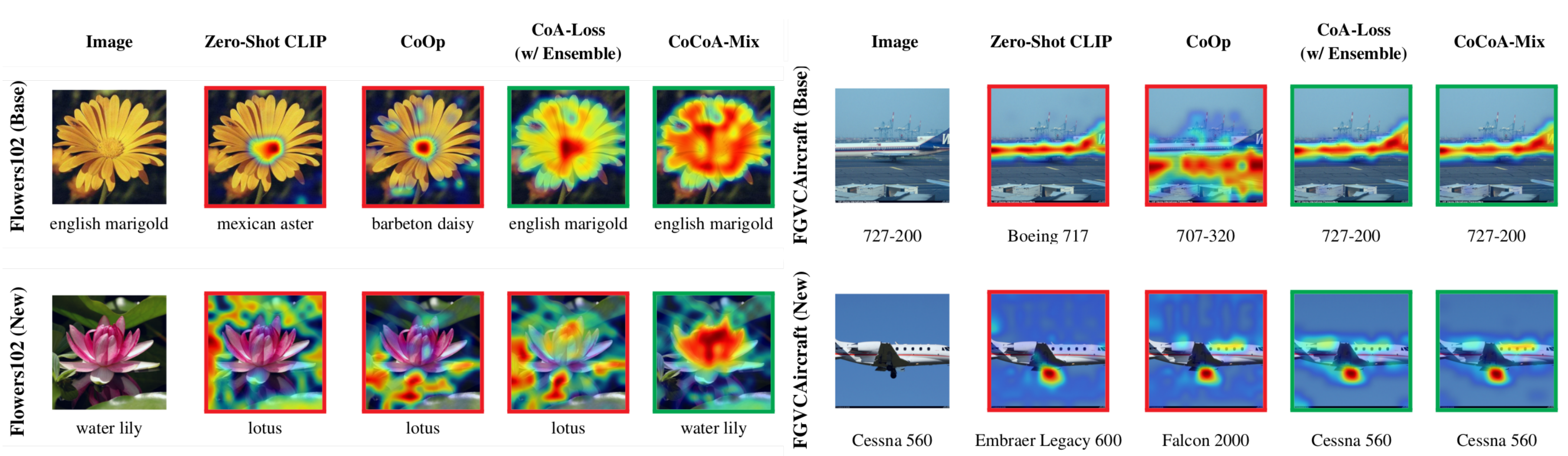
Motivation & Contribution

TL;DR

- Existing **prompt tuning** methods struggle with class confusion due to frozen encoders and often sacrifice task specialization for better generalization across unseen domains.

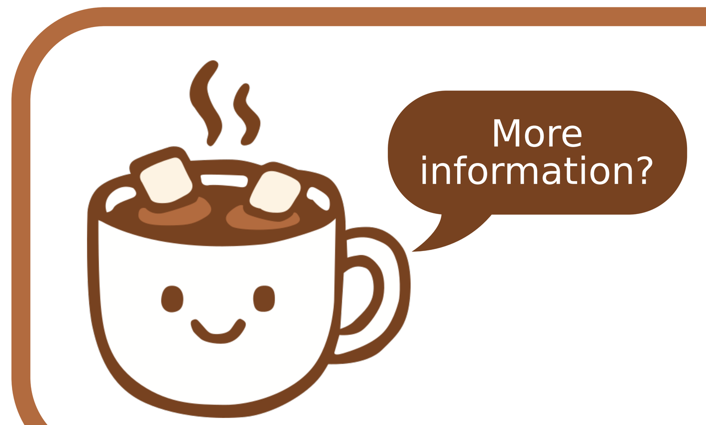


- We propose **CoCoA-Mix**, a novel framework combining a Confusion-Aware Loss **(CoA-loss)** to refine decision boundaries for confusing classes, and Confidence-Aware Weights **(CoA-weights)** to balance specialization and generalization in a mixture model.



Contribution

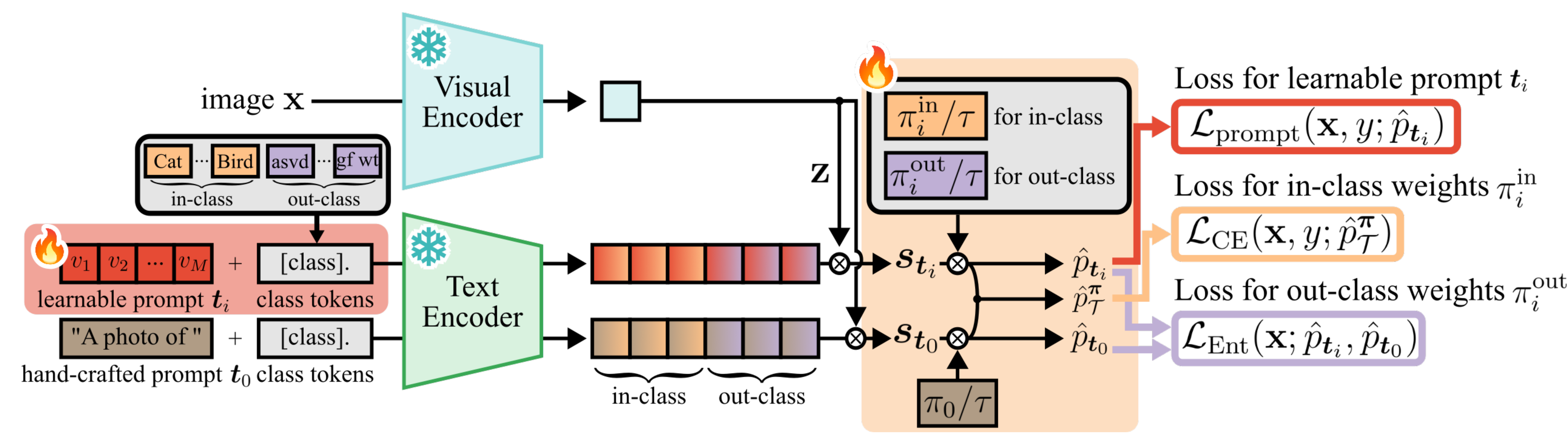
- We provide a mathematical framework demonstrating that specialization and generalization can be improved simultaneously.
- We propose a **CoCoA-Mix** framework, consisting of CoA-loss and CoA-weights. **CoA-loss** boosts specialization by improving classification for confusing cases, while **CoA-weights** improve generalization by adjusting the weights of individual prompts in the mixture model based on their confidence over class domains.
- The proposed method achieves average harmonic mean improvements of 15.28% and 3.28% over zero-shot CLIP in base-to-new generalization and cross-dataset transfer, respectively; it also improves the average accuracy in few-shot class-incremental learning by 5.6%*p*.



GitHub Project Page ArXiv

Methodology

CoCoA-Mix
(Confusion-and-Confidence-Aware Mixture Model)



- **Mixture Model Framework**: A theoretical mixture model that aggregates predictions from multiple prompts, enabling a unified analysis of both specialization (task-specific accuracy) and generalization (performance on unseen domains).

The expected error of the mixture model can be bounded as follows:

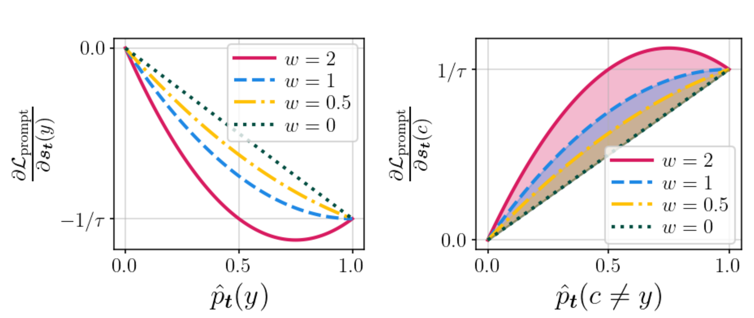
$$\epsilon_T(\hat{p}_{\mathcal{T}}^{\boldsymbol{\pi}}) \le \sum_{i=0}^{K} \lambda_i \left(\pi_i^{in} \underbrace{\epsilon_{T_i}(\hat{p}_{t_i})}_{\substack{\text{specialization}\\ \text{error}}} + \underbrace{\sum_{\substack{j=0\\ j\neq i}}^{K} \pi_j^{out} \epsilon_{T_i}(\hat{p}_{t_j})}_{\substack{\text{generalization}\\ \text{error}}} \right)$$

- $\hat{p}_{\boldsymbol{t}}$: prediction with a text prompt $\boldsymbol{t}$
- $\mathcal{T} = \{\boldsymbol{t}_0, \boldsymbol{t}_1, \cdots, \boldsymbol{t}_K\}$: K+1 different prompts
- $\boldsymbol{\pi} = \{(\pi_0^{in}, \pi_0^{out}), \cdots, (\pi_K^{in}, \pi_K^{out})\}$: a set of non-negative weights satisfying $\pi_i^{in} + \sum_{j=0, j\neq i}^{K} \pi_j^{out} = 1$
- The class set of the target domain $\mathcal{D}_T$ be partitioned into K+1 disjoint subsets, with corresponding sub-domains $\mathcal{D}_{T_0}, \mathcal{D}_{T_1}, \cdots, \mathcal{D}_{T_K}$, such that $\mathcal{D}_T = \bigsqcup_{i=0}^{K} \mathcal{D}_{T_i}$
- $\lambda_i = \Pr_{(\mathbf{x},y)\sim\mathcal{D}_T}[(\mathbf{x}, y) \in \mathcal{D}_{T_i}]$: the probability that a sample from $\mathcal{D}_T$ belongs to $\mathcal{D}_{T_i}$

- **Confusion-Aware Loss (CoA-loss)**: To enhance specialization, CoA-loss assigns higher gradients to confusing examples—samples with similar probabilities across multiple classes—thereby refining decision boundaries in prompt tuning.

$$\mathcal{L}_{prompt} = \mathcal{L}_{CE} + w\mathcal{L}_{CoA}$$

$$\mathcal{L}_{CoA}(\mathbf{x}, y; \hat{p}_{\boldsymbol{t}}) = 1 - \hat{p}_{\boldsymbol{t}}(y)$$



- **Confidence-Aware Weights (CoA-weights)**: To improve generalization without compromising specialization, the model adjusts mixture weights based on each prompt's confidence in the class domain, assigning more weight to confident prompts and less to uncertain prompts.

In-Class Weights

$$\pi_i^{in} = \arg\min_{\pi_i^{in}} \mathbb{E}_{(\mathbf{x},y)\sim\mathcal{D}_{S_i}}[\mathcal{L}_{CE}(\mathbf{x}, y; \hat{p}_{\mathcal{T}}^{\boldsymbol{t}})]$$

In-Class Weights

$$\pi_i^{out} = \arg\min_{\pi_i^{out}} \mathbb{E}_{(\mathbf{x},y)\sim\mathcal{D}_{S_i}}[\mathcal{L}_{Ent}(\mathbf{x}; \hat{p}_{\boldsymbol{t}_i}, \hat{p}_{\boldsymbol{t}_0})]$$

$$\mathcal{L}_{Ent} = \max(0, H(\hat{p}_{\boldsymbol{t}_0}) - H(\hat{p}_{\boldsymbol{t}_i}) + d)$$

Experiments & Analysis

Comparative Study

- CoCoA-Mix is evaluated on 11 datasets across three tasks—base-to-new generalization, few-shot class-incremental learning (FSCIL), and cross-dataset transfer—demonstrating its broad applicability.

Base-to-New Generalization

Method	Average Base	Average New	Average H	ImageNet Base	ImageNet New	ImageNet H	Caltech101 Base	Caltech101 New	Caltech101 H	Food101 Base	Food101 New	Food101 H
CLIP	65.14	68.78	66.82	64.43	60.04	62.16	90.64	91.16	90.90	83.58	84.95	84.26
CoOp	77.23	68.56	71.33	73.72 ± 0.29	64.94 ± 0.87	69.05	97.16 ± 0.16	93.92 ± 0.80	95.51	89.19 ± 0.19	88.45 ± 0.89	88.81
ProGrad	78.74	72.19	75.06	74.81 ± 0.29	66.68 ± 0.26	70.51	97.50 ± 0.08	95.49 ± 0.27	96.48	89.33 ± 0.08	89.93 ± 0.58	89.63
KgCoOp	78.67	74.62	76.38	75.44 ± 0.08	69.43 ± 0.29	72.31	97.61 ± 0.33	94.80 ± 0.45	96.18	**90.26 ± 0.11**	**91.25 ± 0.15**	**90.75**
MaPLe	77.14	72.91	74.69	75.40 ± 0.29	**70.43 ± 0.12**	**72.83**	97.47 ± 0.31	93.77 ± 1.11	95.57	89.37 ± 0.54	90.77 ± 0.54	90.06
DePT	79.20	66.36	71.78	73.50 ± 0.22	70.00 ± 0.16	71.71	97.83 ± 0.05	**95.83 ± 0.25**	**96.82**	89.80 ± 0.08	88.10 ± 0.16	88.94
CoA-loss	79.12	73.66	76.15	**75.68 ± 0.00**	67.98 ± 0.31	71.62	97.94 ± 0.14	94.54 ± 0.24	96.21	90.11 ± 0.18	90.87 ± 0.42	90.49
CoCoA-Mix	**79.31**	**75.10**	**77.03**	75.47 ± 0.09	68.92 ± 0.10	72.04	**98.02 ± 0.03**	94.39 ± 0.10	96.17	90.09 ± 0.16	90.93 ± 0.09	90.50

Method	OxfordPets Base	OxfordPets New	OxfordPets H	StanfordCars Base	StanfordCars New	StanfordCars H	Flowers102 Base	Flowers102 New	Flowers102 H	FgvcAircraft Base	FgvcAircraft New	FgvcAircraft H
CLIP	90.01	94.24	92.07	55.37	66.65	60.49	69.23	73.90	71.49	19.51	24.60	21.76
CoOp	94.10 ± 0.73	94.42 ± 4.17	94.16	69.54 ± 0.75	71.39 ± 1.28	70.44	90.60 ± 1.50	67.00 ± 1.04	77.01	26.17 ± 7.89	19.50 ± 11.94	11.46
ProGrad	95.00 ± 0.31	97.36 ± 0.42	96.16	71.45 ± 0.39	73.16 ± 0.58	72.29	91.36 ± 0.63	74.92 ± 0.90	82.32	34.21 ± 1.99	28.53 ± 2.08	30.97
KgCoOp	94.65 ± 0.15	97.59 ± 0.08	96.10	68.64 ± 0.35	74.96 ± 0.53	71.66	90.09 ± 0.63	76.31 ± 0.42	82.63	33.43 ± 0.56	32.27 ± 1.19	32.81
MaPLe	94.80 ± 0.94	97.67 ± 0.21	96.21	67.97 ± 0.29	74.40 ± 0.45	71.04	88.03 ± 1.62	73.43 ± 0.49	80.06	31.67 ± 0.66	33.13 ± 2.38	32.29
DePT	94.00 ± 0.29	88.63 ± 0.78	91.23	71.83 ± 0.52	59.27 ± 0.76	64.94	**94.53 ± 0.53**	66.30 ± 1.42	77.92	**35.93 ± 0.93**	24.33 ± 0.09	29.01
CoA-loss	94.90 ± 0.49	**97.93 ± 0.08**	**96.39**	72.70 ± 0.11	73.07 ± 1.27	72.87	88.89 ± 1.75	75.58 ± 1.31	81.67	33.91 ± 0.68	32.47 ± 0.37	33.17
CoCoA-Mix	**95.16 ± 0.38**	97.60 ± 0.09	96.36	**73.09 ± 0.25**	**74.97 ± 0.08**	**74.01**	91.04 ± 1.79	**77.37 ± 0.38**	**83.64**	33.51 ± 0.28	**34.15 ± 0.14**	**33.83**

Method	DTD Base	DTD New	DTD H	EuroSAT Base	EuroSAT New	EuroSAT H	UCF101 Base	UCF101 New	UCF101 H	SUN397 Base	SUN397 New	SUN397 H
CLIP	53.24	54.71	53.97	54.79	66.21	59.96	69.03	69.61	69.32	66.76	70.52	68.59
CoOp	71.22 ± 1.13	53.62 ± 3.45	61.03	79.93 ± 1.07	64.79 ± 6.36	71.19	80.58 ± 0.66	64.11 ± 2.84	71.32	77.37 ± 0.66	72.06 ± 1.56	74.60
ProGrad	72.07 ± 0.29	50.56 ± 2.43	59.35	81.29 ± 3.36	69.81 ± 5.56	74.80	80.97 ± 0.29	73.32 ± 1.85	76.93	**79.16 ± 0.36**	74.34 ± 0.75	76.20
KgCoOp	72.92 ± 1.05	59.14 ± 1.53	65.28	83.20 ± 0.72	**70.51 ± 9.30**	**75.61**	80.09 ± 0.24	**77.75 ± 0.40**	78.90	79.07 ± 0.24	76.78 ± 0.24	77.91
MaPLe	70.40 ± 2.57	58.40 ± 3.00	63.71	76.50 ± 3.85	55.70 ± 3.19	64.27	78.57 ± 2.11	76.60 ± 1.56	77.53	78.33 ± 0.21	**77.67 ± 0.45**	**78.00**
DePT	**74.40 ± 0.83**	53.13 ± 1.07	61.98	78.70 ± 1.56	50.53 ± 5.71	61.08	**81.57 ± 0.84**	66.53 ± 0.87	73.28	79.10 ± 0.22	67.27 ± 0.46	72.70
CoA-loss	73.23 ± 2.02	58.09 ± 0.81	64.76	83.38 ± 0.49	70.07 ± 2.49	76.09	80.83 ± 0.80	74.22 ± 0.91	77.38	78.70 ± 0.25	75.43 ± 0.72	77.03
CoCoA-Mix	72.80 ± 1.89	**64.29 ± 1.25**	**68.25**	**83.49 ± 0.66**	69.11 ± 3.10	75.54	81.28 ± 0.95	**77.75 ± 0.24**	**79.47**	78.51 ± 0.17	76.60 ± 0.24	77.54

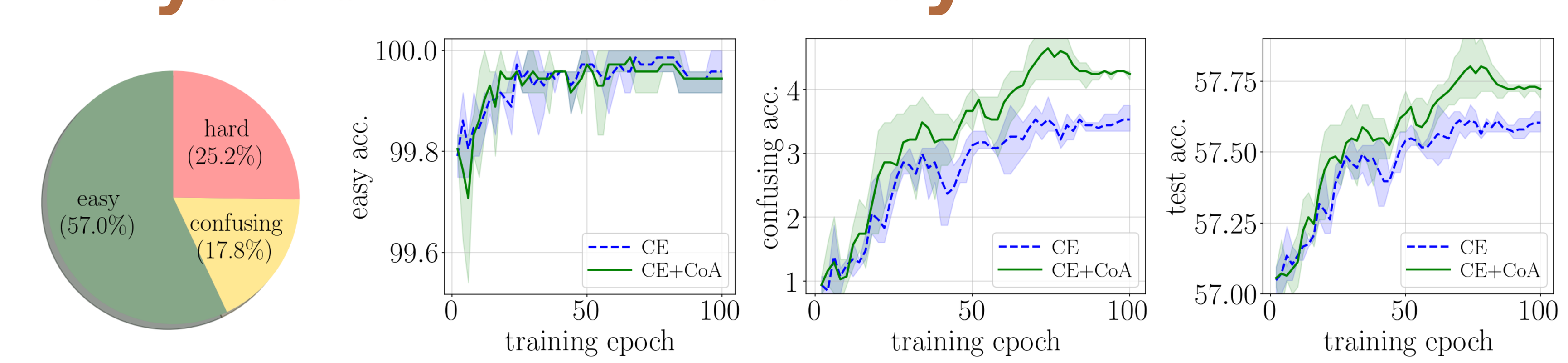
FSCIL
(Few-Shot Class-Incremental Learning)

Method	Acc(%)↑ 0	1	2	3	4	5	6	7	8	Mean↑	PD↓
L2P	**89.9**	**86.0**	81.8	80.3	80.0	74.6	73.2	72.6	65.0	78.2	24.9
CLIP-ZSL	–	–	–	–	–	–	–	–	–	77.9	–
CoOp-FSCIL	88.6	78.9	77.5	76.0	76.8	78.3	79.2	79.8	79.3	79.4	9.3
FACT w/ CLIP	87.8	84.0	81.4	78.0	77.8	76.3	75.0	72.5	71.9	78.3	15.9
FSPT-FSCIL	86.9	83.1	81.9	80.7	80.4	79.9	80.1	79.9	79.4	81.4	7.5
CoCoA-Mix (Ours)	88.2	85.6	**84.6**	**82.7**	**82.8**	**82.5**	**82.3**	**81.8**	**80.8**	**83.5**	**7.4**

Cross-Dataset Transfer

Method	Source	Target	H
CLIP	66.73	64.89	63.97
CoOp	69.06 ± 0.43	59.88	61.52
ProGrad	70.21 ± 0.16	62.36	63.58
KgCoOp	70.52 ± 0.05	64.45	65.17
MaPLe	69.53 ± 0.39	65.24	65.26
DePT	68.03 ± 0.09	65.06	64.42
CoCoA-Mix	**70.85 ± 0.09**	**65.27**	**66.07**

Analysis & Ablation Study



Backbone	Method	Loss	Ensemble	Base	*New*	H
ViT-B/16	CLIP	-	-	65.1	68.8	66.8
	CLIP (w/ Ensemble)	-	Uniform Ensemble	70.6	74.3	72.3
	CoA-loss (w/o Ensemble)	CoA-loss	-	78.6	68.5	72.9
	CoA-loss (w/ Ensemble)	CoA-loss	Uniform Ensemble	79.1	73.7	76.2
	CoCoA-Mix	CoA-loss	CoA-weights	**79.3**	**75.1**	**77.0**
RN50 + RN101 + ViT-B/16 + ViT-B/32	T_{En} + CoCoOp	CoCoOp	Sample-Aware Weight Generator	84.1	75.5	79.2
	T_{En} + CoA-Loss	CoA-loss	Sample-Aware Weight Generator	85.3	75.2	79.5
	CoCoA-Mix	CoA-loss	CoA-weights	**85.4**	**76.3**	**80.3**

- Analyses show that CoA-loss significantly boosts performance on confusing samples, while CoA-weights improve generalization across domains without sacrificing specialization.