Dual Polarity Prompts with Stochastic Entropy Perturbation for Label Noise

Changhui Hu[1,†] Bhalaji Nagarajan[2,†] Ricardo Marques[3,‡] Petia Radeva[1,4,‡]

[1] Dept. de Matemàtiques i Informàtica, Universitat de Barcelona
[2] Barcelona Supercomputing Center
[3] Grup de Recerca de Tecnologies Interactives (GTI), Universitat Pompeu Fabra (UPF)
[4] Institut de Neurosciències, Universitat de Barcelona
[†] Equal Contribution. [‡] Equal Supervision.

BMVC 2025

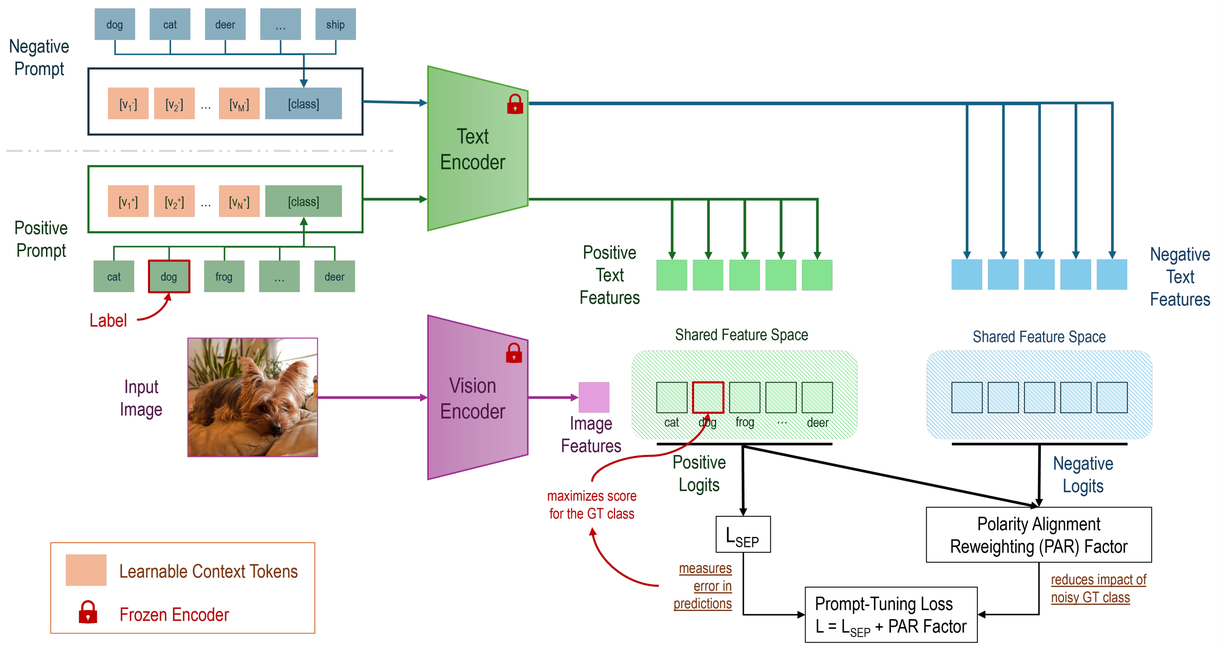
Introduction

- Pre-trained Vision-Language Models (VLMs) have been widely adopted across various applications through few-shot prompt tuning.
- However, their potential to effectively address label noise, a common challenge in real-world data, remains largely unexplored.
- While recent prompt tuning methods exhibit robustness under low to moderate noise, they struggle in handling higher noise ratios or more complex noise patterns.

Contributions

- We introduce a novel **Dual Polarity prompting** strategy along with prompt balancing functions, that enhances image-text alignment of VLMs by simultaneously maximizing similarity with *correct* descriptions and minimizing similarity with *incorrect* ones.
- We propose **Stochastic Entropy Perturbation** loss which mitigates overconfidence in noisy labels by introducing stochastic perturbations, effectively reducing the model's memorization of noisy samples.
- We perform an **extensive validation** on seven synthetic benchmarks with symmetric and instance-dependent noise and two real-world noise benchmarks, covering a various noise distributions, demonstrating the superior noise-robustness.

Overview



Overview of DP-SEP: Dual polarity prompts learn contrasting information and use various balancing functions, Stochastic perturbations mitigate the memorization of label noise.

Results

Dataset	Method	Symmetric 12.5%	25%	50%	65%	80%	Instance 20%	40%	60%	80%
ImageNet	PTNL	65.52	64.74	62.33	59.14	55.18	64.83	62.83	60.09	53.73
	DP-SEP	**65.63**	**65.38**	**64.59**	**63.93**	**63.02**	**65.30**	**64.58**	**63.99**	**62.95**
Caltech101	PTNL	94.29	92.84	91.22	90.13	83.29	92.99	91.42	88.95	82.63
	DP-SEP	**94.90**	**94.57**	**94.39**	**93.10**	**92.76**	**94.62**	**94.59**	**94.15**	**92.54**
OxfordPets	PTNL	83.60	81.86	74.70	56.93	50.43	81.60	79.20	70.05	29.27
	DP-SEP	**86.38**	**85.13**	**85.08**	**84.94**	**83.54**	**86.15**	**85.34**	**84.04**	**82.46**
Flowers102	PTNL	**85.75**	76.54	73.11	56.22	36.82	**84.58**	77.44	60.94	36.17
	DP-SEP	84.47	**82.93**	**78.88**	**75.86**	**70.20**	81.90	**81.30**	**76.96**	**69.73**
Food101	PTNL	77.88	77.49	74.15	69.39	61.83	77.36	75.67	72.58	61.73
	DP-SEP	**78.46**	**78.43**	**78.01**	**77.75**	**77.28**	**78.34**	**78.12**	**77.97**	**77.52**
DTD	PTNL	**61.48**	**57.70**	50.61	43.46	35.80	57.78	53.27	44.15	30.15
	DP-SEP	59.27	56.54	**52.73**	**49.52**	**43.14**	**58.87**	**55.80**	**50.72**	**41.90**
UCF101	PTNL	**73.11**	70.29	64.63	59.75	52.31	71.58	66.66	60.78	49.58
	DP-SEP	72.97	**72.25**	**70.61**	**68.12**	**64.57**	**73.18**	**71.81**	**69.53**	**65.82**

Performance Comparison under Synthetic Noise.

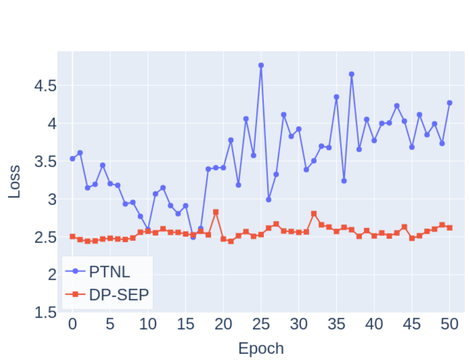
Method	WebVision Top-1	WebVision Top-5	ILSVRC12 Top-1	ILSVRC12 Top-5	Food101N
PTNL	73.55	95.95	73.29	**96.32**	61.61
DP-SEP	**75.58**	**96.17**	**75.23**	96.24	**64.51**

Performance Comparison under Real Noise.

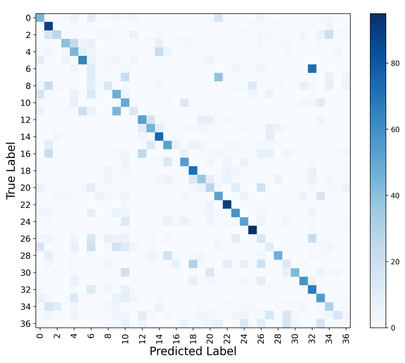
Dataset	Method	Symmetric 12.5%	25%	50%	65%	80%	Instance 20%	40%	60%	80%
OxfordPets	PTNL	79.78	79.39	79.10	77.92	74.73	79.37	**77.11**	78.22	71.93
	DP-SEP	**80.24**	**80.29**	**79.75**	**79.91**	**81.06**	**79.64**	76.48	**80.40**	**81.25**
Food101	PTNL	74.80	73.82	71.82	70.52	68.75	73.45	74.16	71.90	66.80
	DP-SEP	**75.45**	**75.07**	**75.18**	**75.71**	**76.71**	**74.20**	**74.44**	**76.13**	**76.59**
DTD	PTNL	34.84	31.01	34.15	32.07	29.79	34.57	35.43	34.79	30.27
	DP-SEP	**36.49**	**36.70**	**38.30**	**36.91**	**40.11**	**38.56**	**36.86**	**36.28**	**39.95**
UCF101	PTNL	**58.95**	58.52	**58.31**	54.37	50.91	**60.53**	**57.94**	55.17	52.26
	DP-SEP	57.63	**58.68**	57.78	**59.93**	**60.19**	58.47	57.57	**59.82**	**60.06**

Cross-Dataset Generalization
Training on ImageNet and Testing on different datasets.

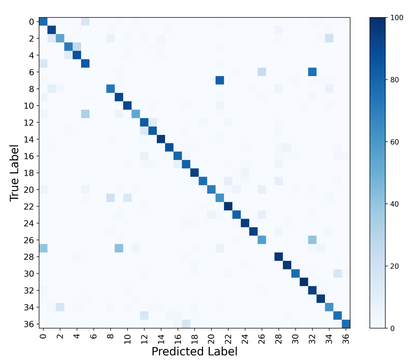
Analysis



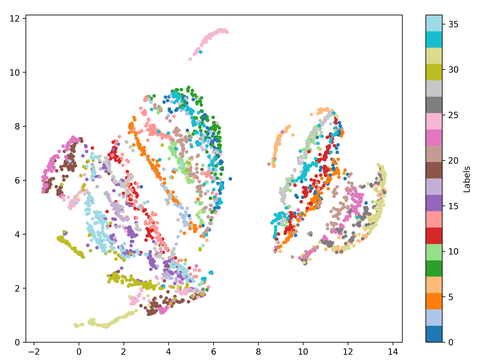
(a) Loss Evolution-OxfordPets



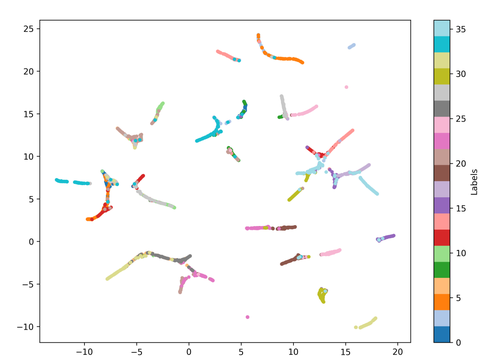
(b) Confusions-OxfordPets-PTNL



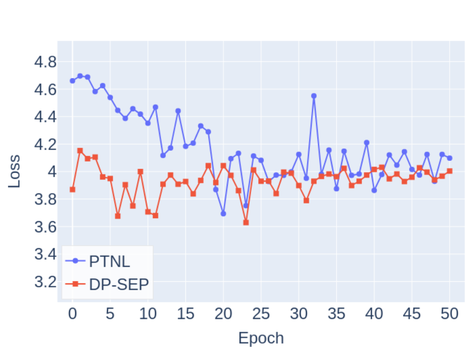
(c) Confusions-OxfordPets-DP-SEP



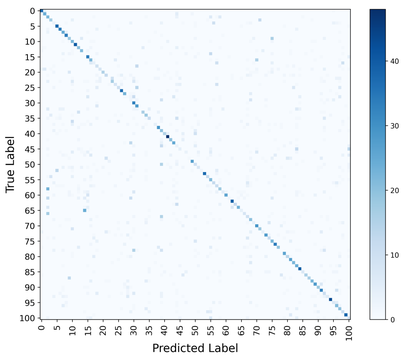
(d) Features-OxfordPets-PTNL



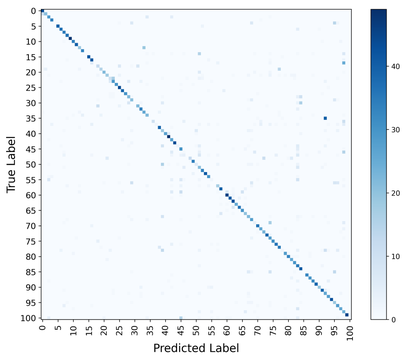
(e) Features-OxfordPets-DP-SEP



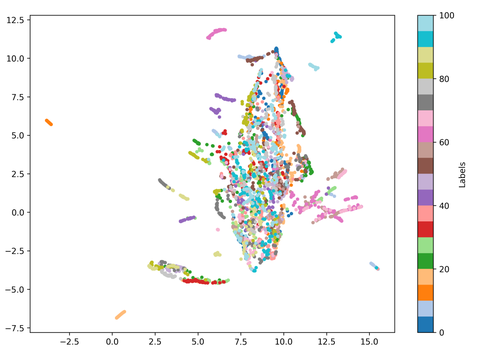
(f) Loss Evolution-UCF101



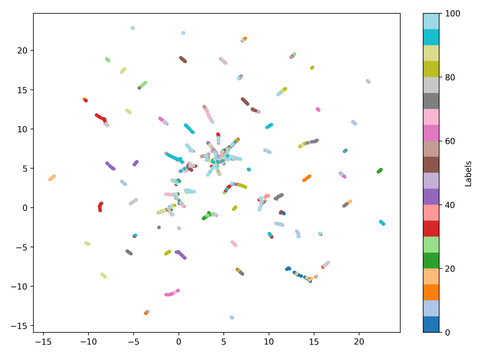
(g) Confusions-UCF101-PTNL



(h) Confusions-UCF101-DP-SEP



(i) Features-UCF101-PTNL



(j) Features-UCF101-DP-SEP

Conclusion

- As VLMs are increasingly becoming central to several applications, their sensitivity to label noise poses a significant barrier. In this regard, we focus on addressing the degradation of VLMs under high and complex noise conditions. We propose DP-SEP, a framework revolving around two core components.
- Dual Polarity Prompt constructs the discriminative relationship between *correct description* and *wrong description* in the semantic space, limiting the model's over-reliance on labels (possibly wrong).
- Stochastic Entropy Perturbation uses randomness to break the network's possibility of absolute confidence in incorrect samples.

Acknowledgments: This work was partially funded by the EU project MUSAE (No. 01070421), 2021-SGR-01094 (AGAUR), Icrea Academia'2022 (Generalitat de Catalunya), Robo STEAM (2022-1-BG01-KA220-VET-000089434, Erasmus+ EU), DeepSense (ACE053/22/000029, ACCIÓ), DeepFoodVol (AEI-MICINN, PDC2022-133642-I00) and IDEATE (AEI-MICINN, PID2022-141566NB-I00). C. Hu acknowledges China Scholarship Council (CSC) Fellowship (Ministry of Education, China, No. 202208410100). B. Nagarajan acknowledges AI4S fellowship within the "Generación D" initiative by Red.es, Ministerio para la Transformación Digital y de la Función Pública, for talent attraction (C005/24-ED CV1), funded by NextGenerationEU through PRTR.