Intrinsic Gradient Suppression for Label-Noise Prompt Tuning in Vision–Language Models

Jiayu Li* Jiaxin Qi* Sheng Zhou Jianqiang Huang Xiansheng Hua * Equal contribution

ICML International Conference On Machine Learning

Paper link

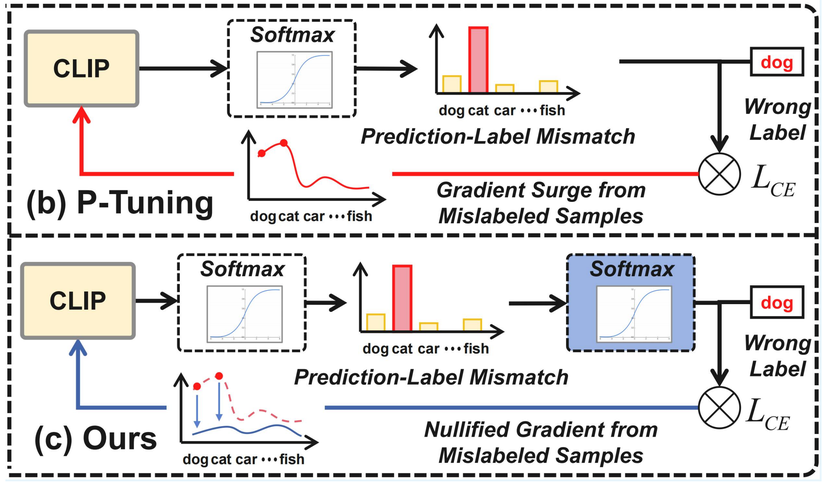
1. Challenge & Motivation

- The presence of **noisy label** disrupts the adaptation of the **prompt tuning** process of vision-language models.
- Because CLIP already provides a **near-optimal initialization**, adaptation should be inherently conservative, particularly against the extreme gradient updates common in noisy settings.

2. Our Contributions

- We provide a **systematic analysis** of CLIP's gradient dynamics under label noise. We reveal a gradient surge paradox, where CLIP possesses a strong pre-trained prior and will produce disproportionately large gradients for "incorrect" noisy samples, which aggressively overrule the original knowledge.
- We propose **Double-Softmax Prompt Tuning (DSPT)**, a hyperparameter-free framework that re-purposes "gradient vanishing" as a principled noise-filtering shield. DSPT acts as an intrinsic gradient gatekeeper that zeros out disruptive surges from noise while selectively permitting informative updates from reliable samples.
- We provide **theoretical evidence** and conduct **extensive experiments** across diverse benchmarks.

3. Our Method: DSPT



Our method is appealingly simple: adding an additional softmax layer before the cross-entropy loss, which surppresses noisy gradients!

4. Theoretical Analysis

- **Loss Value boundedness:** In a C class classification problem, for any classification model, its double-softmax cross-entropy loss is **bounded by:**

$$\log(1+\frac{C-1}{e}) \leq \mathcal{L}_{ours} \leq \log(e+C-1)$$

- **Gradient surppresion for incoherent mislabeled samples:** For high confidence single-softmax prediction $\hat{y}$ inconsistent with noisy label $\tilde{y}$, its **total gradient quantity** under double-softmax cross-entropy loss with respect to logit z will **converge to 0**:

$$\lim_{\boldsymbol{p}_{\tilde{y}}\to 1,\hat{y}\neq\tilde{y}} \sum_i \left|\frac{\partial \mathcal{L}_{ours}}{\partial z_i}\right| = 0$$

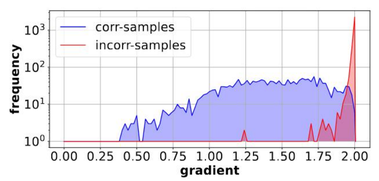
- **Additional analysis can be found in our main paper!**

5. Experiments

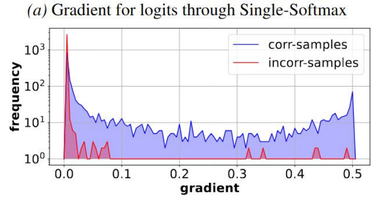
- **Comparative experiment** on nine benchmarks with multiple baselines:

Method	Dataset	Noise Type: sym 40%	60%	80%	Noise Type: pair 20%	30%	40%	Dataset	Noise Type: sym 40%	60%	80%	Noise Type: pair 20%	30%	40%
ZeroShot	Caltech 101	81.43						FGVC Aircraft	24.37					
CoOp		91.06	89.37	80.00	86.14	76.38	63.21		34.67	25.84	12.27	37.83	34.03	29.72
Smoothing		92.82	90.54	81.29	90.39	83.77	70.47		33.33	26.02	11.80	36.98	33.27	29.28
LogitNorm		83.06	82.14	80.21	86.44	85.27	82.71		26.73	23.07	17.56	29.01	26.41	25.26
NLPrompt		**95.36**	94.42	91.85	95.56	95.21	93.63		34.74	33.38	14.86	36.71	34.45	33.60
DSPT		95.07	**95.31**	**94.01**	**96.06**	**95.85**	**93.71**		**36.46**	**33.53**	**29.12**	**38.46**	**35.93**	**33.60**
ZeroShot	Stanford Cars	65.33						DTD	42.73					
CoOp		71.56	61.76	40.48	69.38	59.99	49.47		64.55	56.73	36.99	64.73	56.69	47.53
Smoothing		70.58	61.45	40.90	71.61	65.21	52.51		67.77	57.15	39.95	68.60	59.98	49.94
LogitNorm		57.05	56.97	54.05	57.96	56.49	51.48		66.84	61.81	49.07	69.43	66.50	59.30
NLPrompt		**79.70**	**77.21**	**71.98**	80.90	78.48	75.01		65.97	59.17	41.25	70.17	**70.07**	65.70
DSPT		79.64	76.78	71.50	**81.14**	**78.82**	**75.15**		**68.84**	**63.85**	**55.50**	**71.03**	69.93	**66.74**
ZeroShot	Oxford Pets	88.19						EuroSAT	42.91					
CoOp		89.17	83.05	60.43	83.22	72.96	60.76		93.09	90.75	76.03	89.75	82.43	69.77
Smoothing		89.38	85.85	62.76	87.97	79.41	65.58		92.03	91.00	77.41	92.56	86.99	73.65
LogitNorm		90.75	84.18	79.02	89.73	83.46	72.27		92.30	91.33	81.00	92.86	90.91	83.31
NLPrompt		92.41	88.97	84.18	**93.63**	**92.85**	91.27		92.04	88.83	72.01	93.93	93.19	82.00
DSPT		**93.00**	**92.60**	**91.25**	93.19	92.83	**91.74**		**93.39**	**92.64**	**82.06**	**94.42**	**93.81**	**92.38**
ZeroShot	Flowers 102	66.19						UCF101	65.19					
CoOp		91.03	84.54	66.40	86.27	74.99	59.39		79.88	75.33	62.89	73.20	65.53	54.78
Smoothing		91.89	83.89	66.91	**91.09**	81.44	65.28		79.97	76.31	63.57	78.99	72.26	59.06
LogitNorm		73.11	68.96	60.82	76.01	72.89	65.15		75.73	72.43	64.89	76.31	73.90	69.30
NLPrompt		**93.29**	84.80	80.08	90.37	**91.47**	88.39		83.00	**82.04**	**77.40**	**83.75**	82.72	**81.55**
DSPT		91.12	**88.15**	**80.35**	91.01	86.93	**90.07**		**83.56**	81.19	76.48	83.37	**83.01**	81.13
ZeroShot	Food101	85.46						AVG	62.50					
CoOp		87.76	86.63	84.45	79.58	70.71	58.60		78.09	72.67	57.77	74.46	65.97	54.80
Smoothing		87.71	86.76	84.66	84.01	77.77	64.24		78.39	73.22	58.81	78.02	71.12	58.89
LogitNorm		85.32	85.61	84.73	84.71	83.86	79.05		72.32	69.61	63.49	73.60	71.07	65.31
NLPrompt		88.50	88.08	85.78	**89.05**	88.65	88.41		80.56	77.43	68.82	81.56	**80.79**	77.73
DSPT		**88.86**	**88.62**	**87.73**	89.03	**88.96**	**88.77**		**81.10**	**79.19**	**74.22**	**81.97**	80.67	**79.25**

- **Overall gradient** of all samples in the first training epoch undel noise:

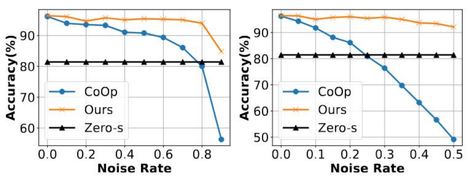


(a) Gradient for logits through Single-Softmax

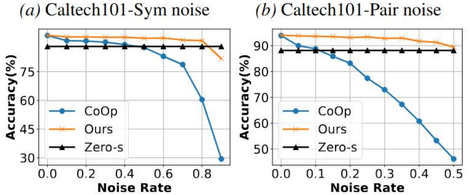


(b) Gradient for logits through Double-Softmax

- **Accuracy curve** under different noise and datasets:



(a) Caltech101-Sym noise *(b)* Caltech101-Pair noise



(c) OxfordPets-Sym noise *(d)* OxfordPets-Pair noise