NEURAL INFORMATION PROCESSING SYSTEMS

山东大学 人工智能研究中心

When Catastrophic Inheritance Meets Forgetting in Continual Adaptation of Foundation Models

NeurIPS 2026

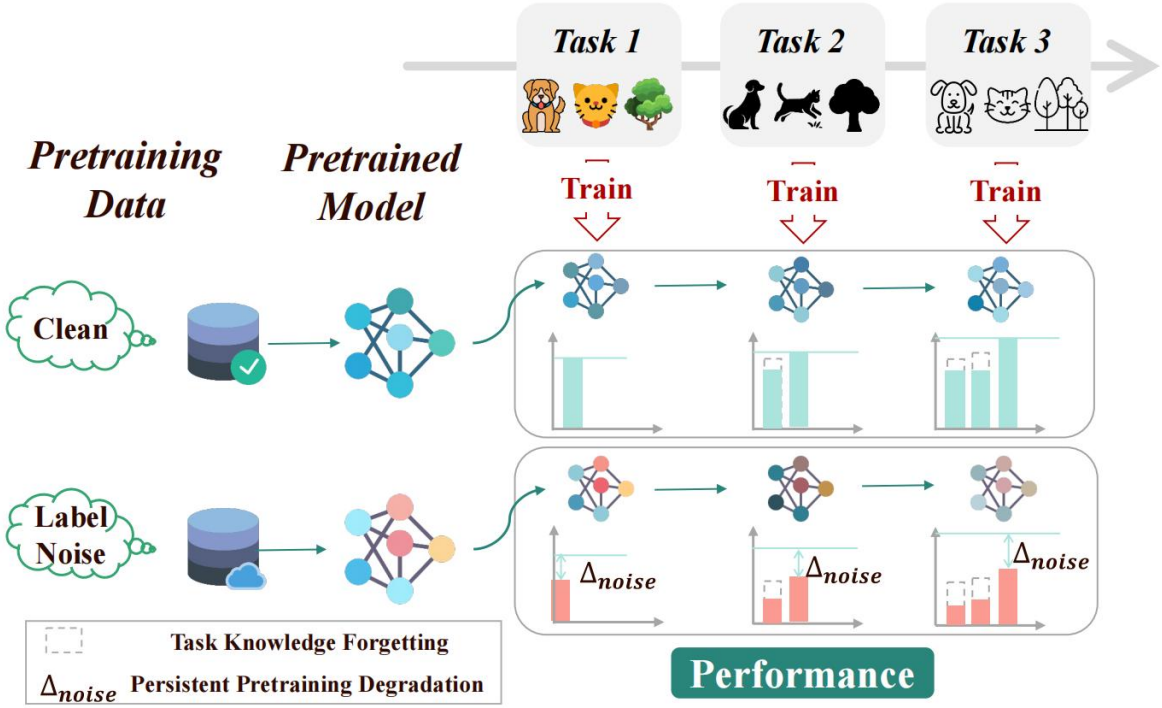
Quanyu Zhang, Zhongyi Han*, Zhenxue Chen, Xiao-Long Yin*, Henggui Zhang, Shuo Li

*corresponding author

学无止境 气有洪邪

Motivation

What happens when models pretrained on noisy data continually adapt to new tasks?



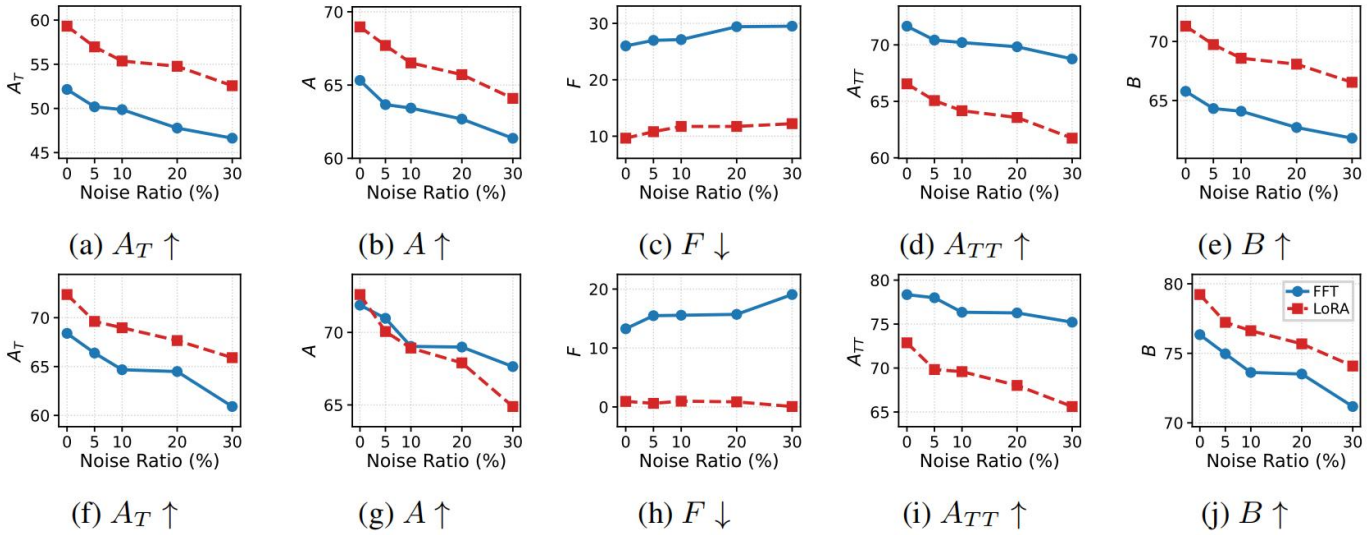
Influence of Pretraining Noise

Persistent degradation in new-task learning

Pretraining noise persistently degrades performance on new tasks.

Increased forgetting in some settings

Pretraining noise may further increase forgetting of previously adapted tasks.



(a) $A_T \uparrow$ (b) $A \uparrow$ (c) $F \downarrow$ (d) $A_{TT} \uparrow$ (e) $B \uparrow$

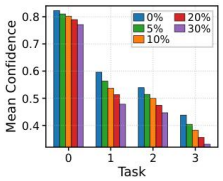
(f) $A_T \uparrow$ (g) $A \uparrow$ (h) $F \downarrow$ (i) $A_{TT} \uparrow$ (j) $B \uparrow$

ViT-B/16 on ImageNet-1K with 0–30% asymmetric label noise.
FFT and LoRA on sequential DomainNet and OfficeHome tasks.

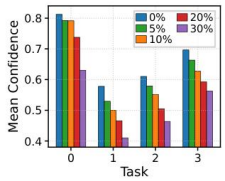
Analysis: Inherited Output Contraction

Pretraining noise compresses the logit gap between the true class and competing classes.

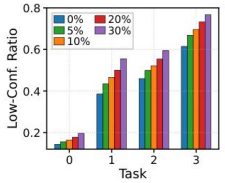
$$g_i(\theta) = z_{i,y_i} - \max_{k \neq y_i} z_{i,k}$$



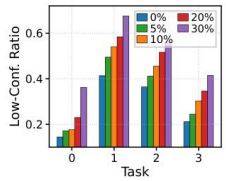
(a) DomainNet, Mean



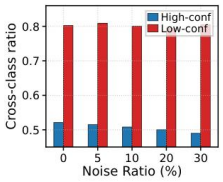
(b) OfficeHome, Mean



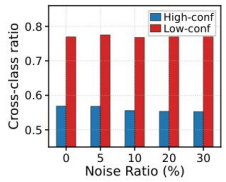
(c) DomainNet, Low



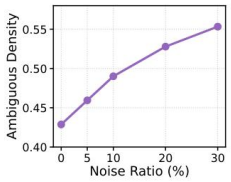
(d) OfficeHome, Low



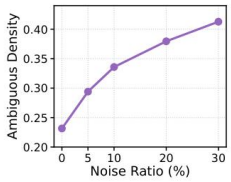
DomainNet
(a) Cross-class ratio



OfficeHome
(b) Cross-class ratio



DomainNet
(c) Ambiguous density



OfficeHome
(d) Ambiguous density

Lower confidence

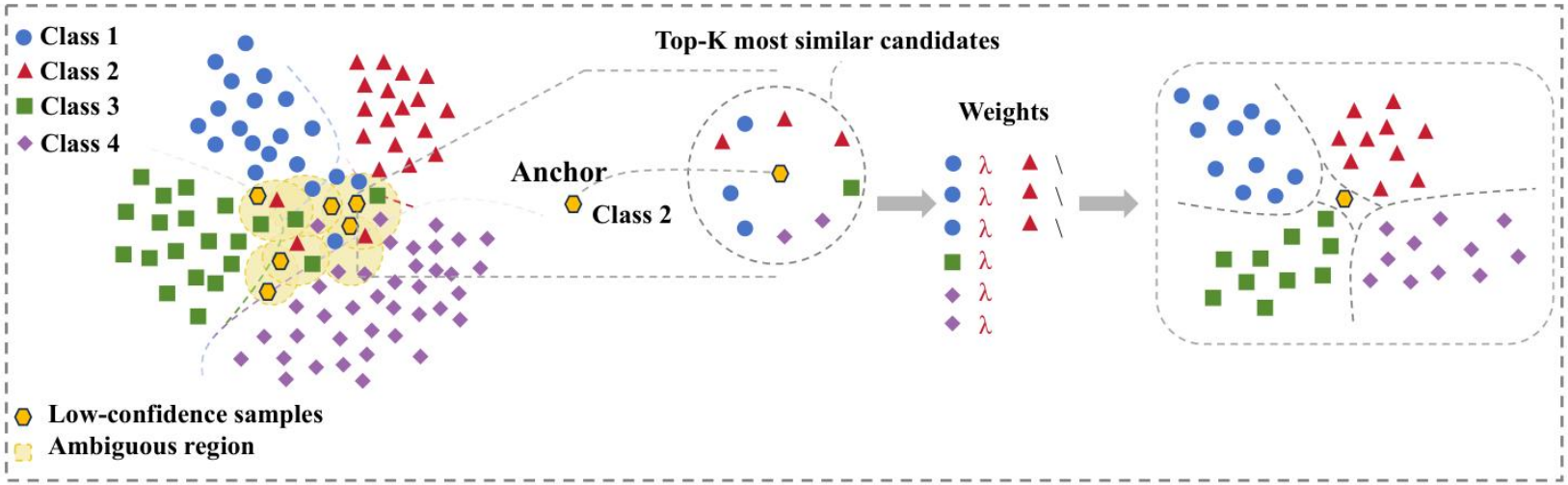
Lower mean confidence and a higher low-confidence ratio.

Cross-class ambiguity

Higher cross-class neighborhood overlap.

Method: Ambiguous Boundary Correction

Anchor-guided neighborhood reweighting



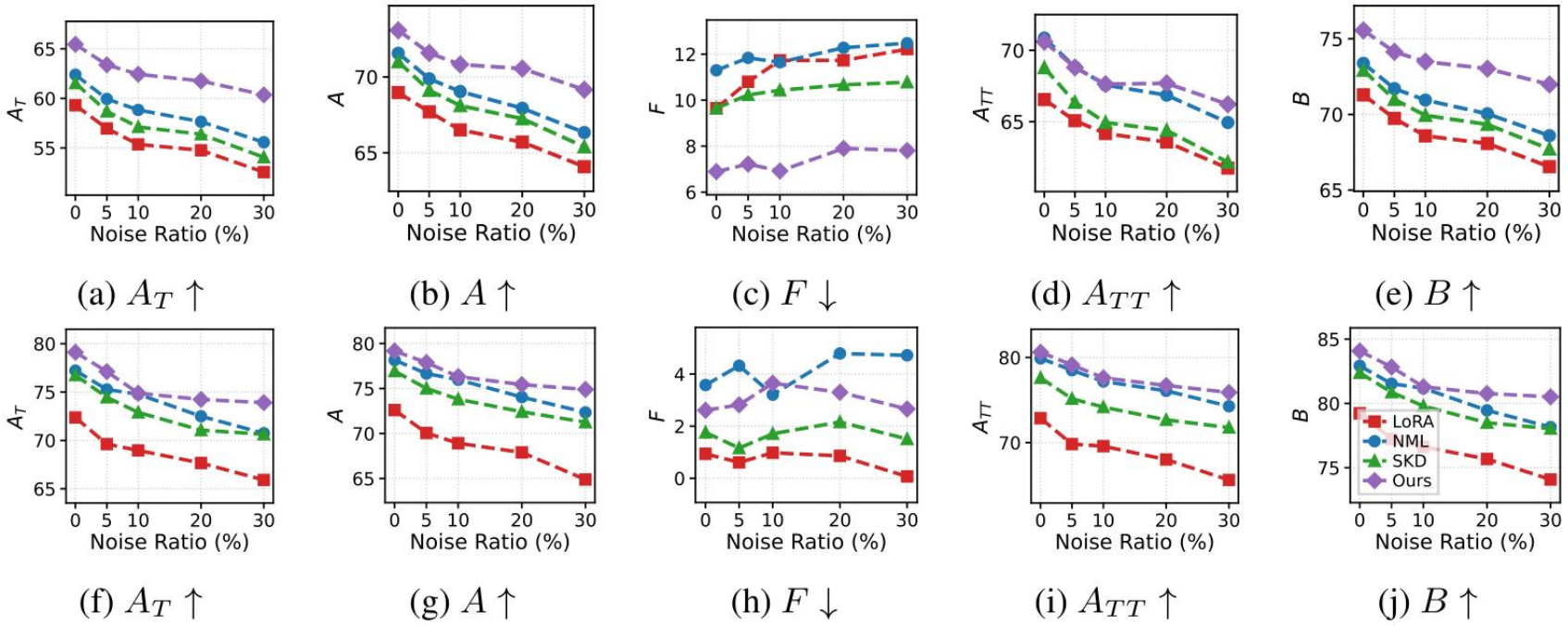
1. Low-confidence anchors $c_i = \max_k p_\theta(k|x_i), \quad \mathcal{A} = \{i \mid c_i < \tau\}$

2. Cross-class neighbors $\mathcal{S}_i = \mathrm{TopK}_{j \neq i}\left(\cos(f_i, f_j), K\right), \quad \mathcal{N}_i = \{j \in \mathcal{S}_i \mid y_j \neq y_i\}, \quad \mathcal{N} = \bigcup_{i \in \mathcal{A}} \mathcal{N}_i$

3. Reweighted training $\mathcal{L}_{\mathrm{ABC}} = -\frac{1}{n}\sum^{n} w_i \log p_\theta(y_i|x_i), \quad w_i = \begin{cases} 1 + \lambda \frac{n}{|\mathcal{N}|}, & \text{if } i \in \mathcal{N}, \\ 1, & \text{otherwise.} \end{cases}$

Experiments: Controlled Pretraining Noise

Main Results



(a) $A_T \uparrow$ (b) $A \uparrow$ (c) $F \downarrow$ (d) $A_{TT} \uparrow$ (e) $B \uparrow$

(f) $A_T \uparrow$ (g) $A \uparrow$ (h) $F \downarrow$ (i) $A_{TT} \uparrow$ (j) $B \uparrow$

Experiments: Controlled Pretraining Noise

CLIP / ViT-L/14 under domain shifts and class shifts

Domain shifts

Dataset	Method	$A_T \uparrow$	$A \uparrow$	$F \downarrow$	$A_{TT} \uparrow$	$B \uparrow$
DomainNet	LoRA	80.93	85.60	3.45	83.52	86.65
	NML	81.11	86.58	5.17	84.43	86.73
	SKD	80.37	85.57	3.36	82.89	86.37
	Ours	**82.30**	**86.61**	**2.94**	84.50	**87.62**
	Ours (Feature Projection)	81.15	86.15	4.74	**84.70**	86.82
OfficeHome	LoRA	88.05	87.41	**0.00**	86.82	90.57
	NML	88.27	89.06	1.88	89.69	91.28
	SKD	88.32	88.08	0.42	88.49	91.12
	Ours	**90.97**	**91.19**	1.24	**91.61**	**93.13**
	Ours (Feature Projection)	89.92	90.39	1.86	91.31	92.44
CDDB	LoRA	65.67	61.63	8.64	69.33	72.00
	NML	63.82	61.41	8.94	69.63	71.48
	SKD	59.06	58.58	**0.74**	58.81	68.93
	Ours	**70.51**	**66.09**	8.03	72.45	**75.25**
	Ours (Feature Projection)	68.53	64.78	11.33	**74.22**	74.05
TerraIncognita	LoRA	62.01	70.40	**6.03**	66.28	73.16
	NML	57.56	69.47	17.43	70.63	70.06
	SKD	56.31	67.98	8.36	62.34	69.57
	Ours	65.40	73.64	10.37	72.80	75.37
	Ours (Feature Projection)	**69.26**	**76.84**	13.88	**79.67**	**77.97**

Class shifts

Dataset	Method	$A_T \uparrow$	$A \uparrow$	$F \downarrow$	$A_{TT} \uparrow$	$B \uparrow$
CUB-40	LoRA	57.22	66.16	8.64	63.80	69.64
	NML	72.74	76.99	13.32	83.40	79.95
	SKD	67.93	75.75	7.88	74.20	77.50
	Ours	**76.39**	**80.79**	**7.54**	82.42	**83.02**
	Ours (Feature Projection)	76.06	80.07	11.37	**85.16**	82.48
CUB-20	LoRA	47.08	59.71	11.32	57.17	63.16
	NML	66.22	74.23	16.98	81.51	76.25
	SKD	54.88	69.95	**9.38**	63.32	69.69
	Ours	67.19	75.29	10.77	76.84	77.14
	Ours (Feature Projection)	**68.82**	**76.67**	16.87	**84.00**	**78.16**



山东大学
人工智能研究中心

Q & A

Thanks!



学无止境　气有浩然