南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Backpropagation-Free Test-Time Adaptation via Probabilistic Gaussian Alignment

NeurIPS 2025

Background

现有 TTA 方法存在三大核心短板：

1、传统微调类方法依赖反向传播，无法实时部署；

2、无训练方法需手动调参，泛化性差；

3、直推式学习仅适配离线场景，不支持在线流式推理。

Table 1: Comparison with existing TTA methods.

Method	BP-Free	Distribution-Aware	Task Setting	
			Online	Transductive
Prompt Tuning [33, 62, 31, 42, 54]	✗	✗	✓	✗
Adapter Tuning [11, 68, 56, 46, 61]	✗	✗	✓	✗
Similarity Score [21, 64, 65, 63, 47]	✓	✗	✓	✗
Transductive Learning [20, 29, 59, 69, 51]	✓	✓	✗	✓
ADAPT (Ours)	✓	✓	✓	✓

* *BP-Free*: No backpropagation at test time.

Background

虽然现有测试时自适应（TTA）方法在效率与数据可及性上取得进展，仍忽略了**类别条件特征分布**的结构特性：多数方法仅依赖文本原型定义决策边界，或通过任务特定缩放调整相似度，未对类别分布进行显式建模。同时，TTA 在目标域无监督运行、无法访问源数据的设定，进一步加剧了类别分布估计的难度。最终，现有方法难以捕捉类内变异与类间混淆，导致决策边界不稳定，在分布偏移场景下鲁棒性显著下降。

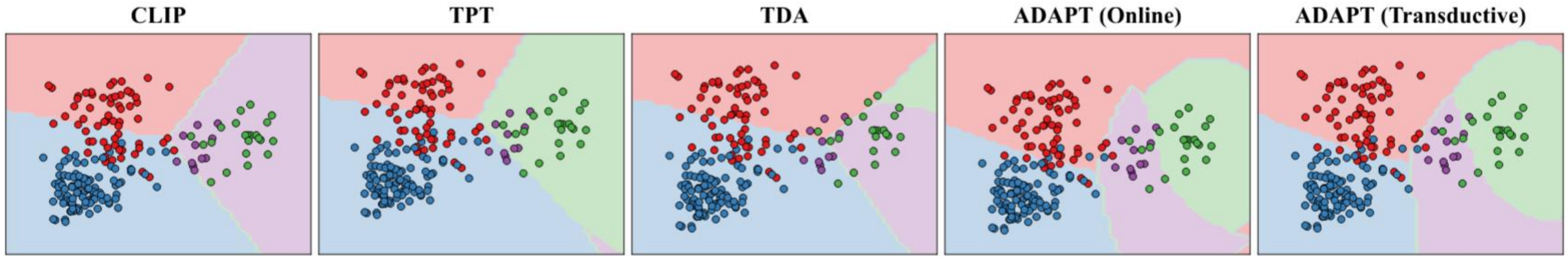


Figure 3: Visualization of decision boundaries on ImageNet-A. The colors indicate different classes.

Motivation

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

核心目标：

1、需要一种无反向传播、无需迭代优化的测试时自适应方法，适配多个部署场景；

2、现有方法普遍忽略类别条件特征分布结构，需显式建模类别分布以解决分布偏移问题；

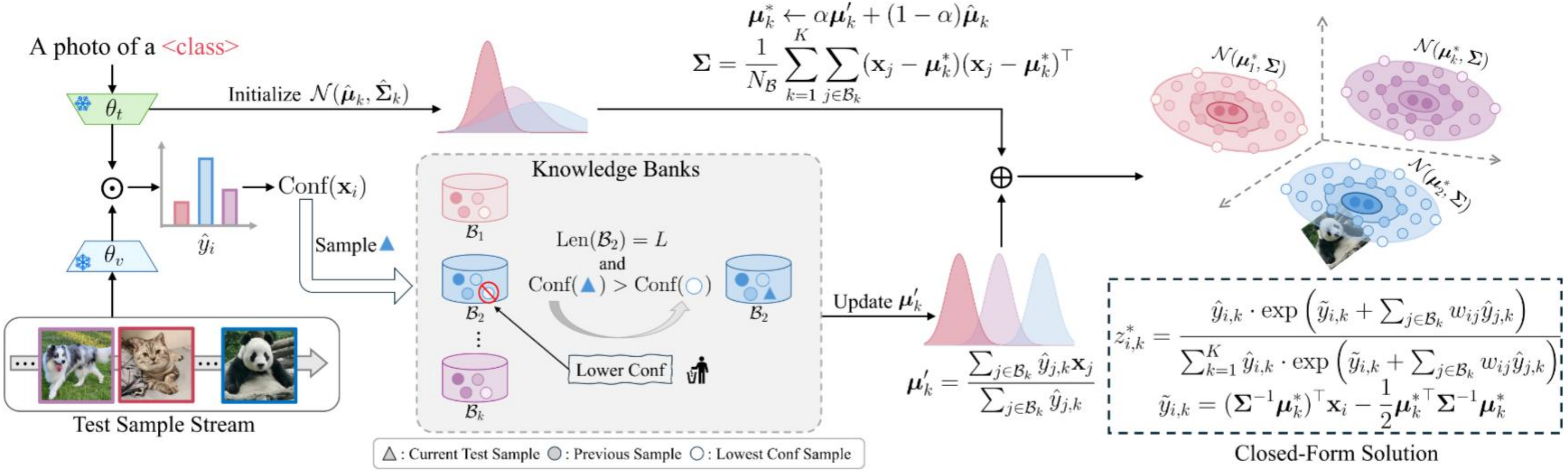
重要前提：

CLIP 特征天然呈现类条件聚类特性，
可被高斯分布良好近似，
为无梯度分布建模提供了可行性。

Table 10: Projection-based normality test results across class-conditional features.

	Low-dim	Freq of p>0.05 (%) ↑	p-value Avg. ↑
Henze–Zirkler	2	100	0.39
	4	99.90	0.32
	6	99.00	0.27
	8	96.30	0.22
	10	92.90	0.19
Shapiro-Wilk	2	100	0.31
	4	100	0.21
	6	99.50	0.16
	8	96.30	0.13
	10	92.20	0.11

Methods



A photo of a <class>
Initialize $\mathcal{N}(\hat{\boldsymbol{\mu}}_k, \hat{\boldsymbol{\Sigma}}_k)$
$\boldsymbol{\mu}_k^* \leftarrow \alpha\boldsymbol{\mu}_k' + (1-\alpha)\hat{\boldsymbol{\mu}}_k$
$\boldsymbol{\Sigma} = \frac{1}{N_{\mathcal{B}}}\sum_{k=1}^{K}\sum_{j\in\mathcal{B}_k}(\mathbf{x}_j - \boldsymbol{\mu}_k^*)(\mathbf{x}_j - \boldsymbol{\mu}_k^*)^\top$
θ_t
θ_v
$\hat{y}_i$
Conf($\mathbf{x}_i$)
Sample
Knowledge Banks
$\mathcal{B}_1$
$\mathcal{B}_2$
$\mathcal{B}_k$
Len($\mathcal{B}_2$) = L
and
Conf(▲) > Conf(○)
Lower Conf
Update $\boldsymbol{\mu}_k'$
$\boldsymbol{\mu}_k' = \frac{\sum_{j\in\mathcal{B}_k}\hat{y}_{j,k}\mathbf{x}_j}{\sum_{j\in\mathcal{B}_k}\hat{y}_{j,k}}$
$\mathcal{N}(\boldsymbol{\mu}_1^*, \boldsymbol{\Sigma})$
$\mathcal{N}(\boldsymbol{\mu}_k^*, \boldsymbol{\Sigma})$
$\mathcal{N}(\boldsymbol{\mu}_2^*, \boldsymbol{\Sigma})$
$z_{i,k}^* = \frac{\hat{y}_{i,k}\cdot\exp\left(\tilde{y}_{i,k} + \sum_{j\in\mathcal{B}_k} w_{ij}\hat{y}_{j,k}\right)}{\sum_{k=1}^{K}\hat{y}_{i,k}\cdot\exp\left(\tilde{y}_{i,k} + \sum_{j\in\mathcal{B}_k} w_{ij}\hat{y}_{j,k}\right)}$
$\tilde{y}_{i,k} = (\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_k^*)^\top\mathbf{x}_i - \frac{1}{2}\boldsymbol{\mu}_k^{*\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_k^*$
Closed-Form Solution
Test Sample Stream
△ : Current Test Sample ◯ : Previous Sample ○ : Lowest Conf Sample

Methods

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

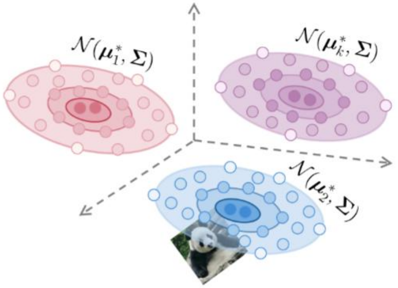
高斯假设： $$\mathbb{P}_{i,k} = \mathbb{P}(\mathbf{x}_i | y_k) = \mathcal{N}(\mathbf{x}_i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)\right)$$

From the Bayes' rule, the posterior probability of the k-th class given image $\mathbf{x}$ is $\mathbb{P}(y_k|\mathbf{x}) = \frac{\mathbb{P}(\mathbf{x}|y_k)\,\mathbb{P}(y_k)}{\mathbb{P}(\mathbf{x})} \propto \pi_k \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma})$. Assuming a uniform class prior $\pi_k = \frac{1}{K}$, then the Bayes optimal prediction label for $\mathbf{x}_i$ is known as the argmax of $\tilde{y}_{i,k}$ over $k = 1, \ldots, K$:

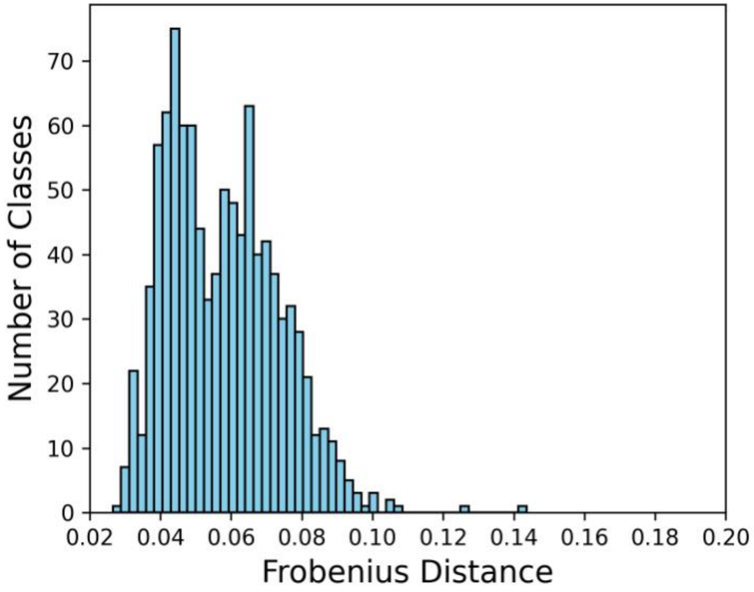
对高斯假设取对数化简： $$\tilde{y}_{i,k} = \mathrm{w}_k^\top \mathbf{x}_i + b_k, \quad \text{where } \mathrm{w}_k = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_k, b_k = -\frac{1}{2}\boldsymbol{\mu}_k^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_k.$$

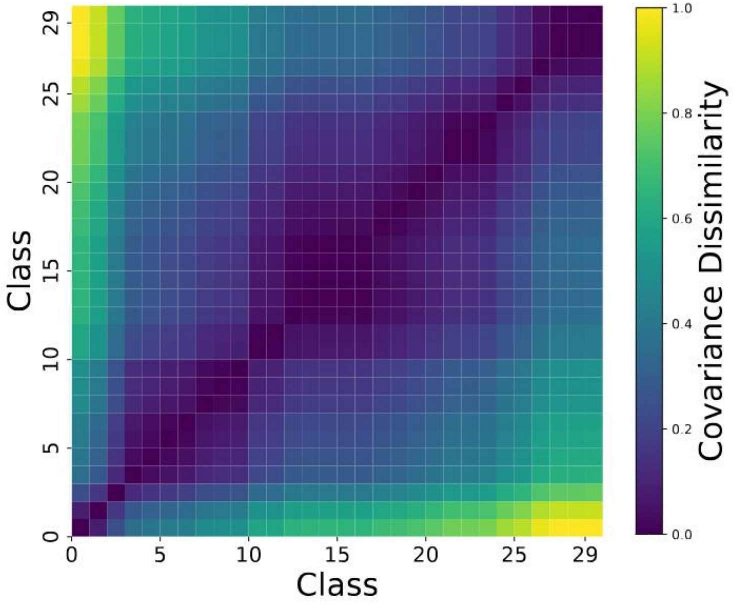
Methods

类间共享的协方差矩阵：



计算每个类别自身的协方差矩阵到全局共享协方差矩阵之间的Frobenius范数距离





随机选取多个类别，计算两两类别协方差矩阵的协方差差异度

Methods

偏差校正框架，通过最小化融合三个组件的正则化目标函数求解：

1、在线负对数似然；

2、基于 CLIP 先验的正则化；

3、知识库引导的一致性正则化

$$\mathcal{L}_{\text{online}}(z_i, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = -z_i^\top \log \mathbb{P}_i + \mathcal{R}(z_i; \hat{y}_i) + \mathcal{R}(z_i; \mathcal{B}),$$

$$\text{where } \mathcal{R}(z_i; \hat{y}_i) = \mathrm{KL}(z_i \| \hat{y}_i) + \beta \sum_{k=1}^{K} \mathrm{KL}\left(\mathcal{N}(\hat{\boldsymbol{\mu}}_k, \hat{\boldsymbol{\Sigma}}_k) \| \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma})\right),$$

$$\mathcal{R}(z_i; \mathcal{B}) = -\sum_{j \in \mathcal{B}} \hat{y}_j^\top \log \mathbb{P}_j - \sum_{j \in \mathcal{B}} w_{ij} z_i^\top \hat{y}_j.$$

得出两个闭式解：

$$z_{i,k}^* = \frac{\hat{y}_{i,k} \cdot \exp\left(\tilde{y}_{i,k} + \sum_{j \in \mathcal{B}_k} w_{ij} \hat{y}_{j,k}\right)}{\sum_{k=1}^{K} \hat{y}_{i,k} \cdot \exp\left(\tilde{y}_{i,k} + \sum_{j \in \mathcal{B}_k} w_{ij} \hat{y}_{j,k}\right)},$$

$$\boldsymbol{\mu}_k = (z_{i,k}\mathbf{x}_i + \sum_{j \in \mathcal{B}_k} \hat{y}_{j,k}\mathbf{x}_j + \beta \hat{\boldsymbol{\mu}}_k)/(z_{i,k} + \sum_{j \in \mathcal{B}_k} \hat{y}_{j,k} + \beta)$$

Methods

最终均值更新的滑动平均公式：

$$\boldsymbol{\mu}_k^* \leftarrow \alpha \boldsymbol{\mu}_k' + (1-\alpha)\hat{\boldsymbol{\mu}}_k, \quad \text{where } \boldsymbol{\mu}_k' = \frac{\sum_{j\in\mathcal{B}_k} \hat{y}_{j,k}\mathbf{x}_j}{\sum_{j\in\mathcal{B}_k} \hat{y}_{j,k}}, \alpha = \frac{\sum_{j\in\mathcal{B}_k} \hat{y}_{j,k}}{\sum_{j\in\mathcal{B}_k} \hat{y}_{j,k} + \beta}.$$

共享的协方差矩阵计算公式：

$$\boldsymbol{\Sigma} = \frac{1}{N_{\mathcal{B}}} \sum_{k=1}^{K} \sum_{j\in\mathcal{B}_k} (\mathbf{x}_j - \boldsymbol{\mu}_k^*)(\mathbf{x}_j - \boldsymbol{\mu}_k^*)^\top, \quad \boldsymbol{\Sigma}^{-1} = d\left((N_{\mathcal{B}}-1)\boldsymbol{\Sigma} + \operatorname{tr}(\boldsymbol{\Sigma})I_d\right)^{-1}$$

直推式损失计算公式：

$$\mathcal{L}_{\text{trans}}(z, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = -\sum_{i=1}^{N} z_i^\top \log \mathbb{P}_i + \sum_{i=1}^{N} \mathcal{R}(z_i; \hat{y}_i) + \sum_{i=1}^{N} \mathcal{R}(z_i; \mathcal{B}).$$

直推式均值计算原始公式：

$$\boldsymbol{\mu}_k = \left(\sum_{i=1}^{N} z_{i,k}\mathbf{x}_i + \sum_{j\in\mathcal{B}_k} \hat{y}_{j,k}\mathbf{x}_j + \beta\hat{\boldsymbol{\mu}}_k\right) / \left(\sum_{i=1}^{N} z_{i,k} + \sum_{j\in\mathcal{B}_k} \hat{y}_{j,k} + \beta\right)$$

直推式均值更新计算简化版公式：

$$\boldsymbol{\mu}_k^* \leftarrow \alpha \boldsymbol{\mu}_k' + (1-\alpha)\hat{\boldsymbol{\mu}}_k, \boldsymbol{\mu}_k' = \frac{\sum_{i=1}^{N} \hat{y}_{i,k}\mathbf{x}_i + \sum_{j\in\mathcal{B}_k} \hat{y}_{j,k}\mathbf{x}_j}{\sum_{i=1}^{N} \hat{y}_{i,k} + \sum_{j\in\mathcal{B}_k} \hat{y}_{j,k}}, \alpha = \frac{\sum_{i=1}^{N} \hat{y}_{i,k} + \sum_{j\in\mathcal{B}_k} \hat{y}_{j,k}}{\sum_{i=1}^{N} \hat{y}_{i,k} + \sum_{j\in\mathcal{B}_k} \hat{y}_{j,k} + \beta}.$$

Experiments

数据集构成

1、**基于自然分布偏移：** ImageNet， ImageNet-A， ImageNet-R， ImageNet-V， ImageNet-Sketch

2、**10个细粒度识别数据集：** Aircraft， Caltech101， Cars， DTD， EuroSAT， Flower102， Food101， Pets， SUN397 和 UCF101。

开集TTA+细粒度

超参数设置

模型架构：使用CLIP ViT-B/16作为骨干网络

滑动平均系数α：设为0.9

knowledge bank size L： 16 for online

6 for transductive

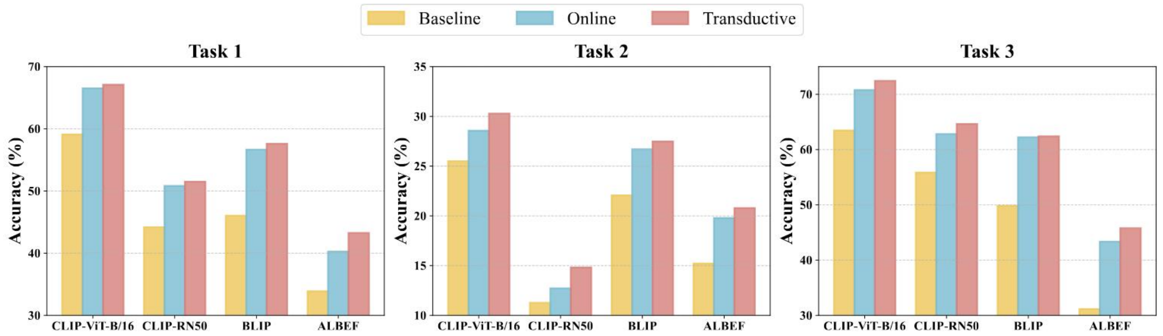


Figure 6: Performance comparison of proposed ADAPT on different VLMs.

Experiments

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Table 2: Top-1 accuracy (%) comparison on natural distribution shift.

	Method	BP-free	ImageNet	ImageNet-A	ImageNet-V	ImageNet-R	ImageNet-S	**OOD Avg.**	**Avg.**
Online	CLIP [39]	-	66.74	47.79	60.89	73.99	46.12	57.20	59.11
	Tip-Adapter [63]	✗	70.75	51.04	63.41	77.76	48.88	60.27	62.37
	TPT [33]	✗	68.98	54.77	63.45	77.06	47.97	60.81	62.45
	DiffTPT [9]	✗	70.30	55.68	65.10	75.00	46.80	60.65	62.58
	C-TPT [55]	✗	68.50	51.60	62.70	76.00	47.90	59.55	61.34
	DMN [65]	✗	72.25	58.28	65.17	78.55	**53.20**	63.80	65.49
	DPE [61]	✗	**71.91**	59.63	**65.44**	80.40	52.26	64.43	65.93
	TPS [46]	✗	70.38	59.21	63.80	77.49	49.57	62.52	64.09
	DynaPrompt [54]	✗	69.61	56.17	64.67	78.17	48.22	61.81	63.37
	B^2TPT [34]	✗	69.57	55.26	65.40	78.64	49.53	62.21	63.68
	MTA [57]	✓	70.08	58.06	64.24	78.33	49.61	62.56	64.06
	TDA [21]	✓	69.51	60.11	64.67	80.24	50.54	63.89	65.01
	ZERO [7]	✓	69.31	59.61	64.16	77.22	48.40	62.35	63.74
	AWT [70]	✓	71.32	60.33	65.15	80.64	51.60	64.43	65.81
	RA-TTA [24]	✓	70.58	59.21	64.16	79.68	50.83	63.47	64.89
	BCA [67]	✓	70.22	61.14	64.90	80.72	50.87	64.41	65.57
	TCA [52]	✓	68.88	50.13	62.10	77.11	48.95	59.57	61.43
	Dota [12]	✓	70.68	61.19	64.41	**81.17**	51.33	64.53	65.76
	ADAPT	✓	70.91	**63.32**	64.64	80.66	53.13	**65.44**	**66.53**
Trans.	GDA-CLIP [51]	✓	64.13	19.72	55.67	55.30	34.32	41.25	45.83
	TransCLIP [59]	✓	70.30	49.50	62.30	75.00	49.70	59.13	61.36
	Frolic [69]	✓	70.90	60.40	64.70	**80.70**	53.30	64.78	66.00
	TIMO [28]	✓	64.63	22.06	56.40	58.47	35.96	43.22	47.50
	ADAPT	✓	**71.56**	**63.77**	**65.59**	80.64	**53.87**	**65.97**	**67.09**

Table 4: Top-1 accuracy (%) comparison on fine-grained categorization.

	Method	BP-free	Aircraft	Caltech	Cars	DTD	EuroSAT	Flower	Food101	Pets	Sun397	UCF101	**Avg.**
Online	CLIP [39]	-	23.70	92.98	65.24	44.44	41.42	67.28	83.80	87.98	62.55	65.08	63.45
	TPT [33]	✗	24.78	94.16	66.87	47.75	42.44	68.98	84.67	87.79	65.50	68.04	65.10
	DiffTPT [9]	✗	25.60	92.49	67.01	47.00	43.13	70.10	87.23	88.22	65.74	68.22	65.47
	C-TPT [55]	✗	24.00	93.60	65.80	46.00	43.20	**79.80**	83.70	88.20	64.80	65.70	64.48
	DMN [65]	✗	**30.03**	**95.38**	67.96	**55.85**	59.43	74.49	85.08	**92.04**	70.18	**72.51**	70.30
	TPS [29]	✗	26.27	94.56	67.00	53.80	42.11	71.69	84.78	87.82	68.25	71.18	66.75
	DPE [61]	✗	28.95	94.81	67.31	54.20	55.79	75.07	86.17	91.14	70.07	70.44	69.40
	HisTPT [62]	✗	26.90	94.50	69.20	48.90	49.70	71.20	**89.30**	89.10	67.20	70.10	67.61
	DynaPrompt [54]	✗	24.33	94.32	67.65	47.96	42.28	69.95	85.42	88.28	66.32	68.72	65.52
	MTA [57]	✓	25.32	94.13	66.36	45.59	38.71	68.26	84.95	88.22	64.98	68.11	64.46
	TDA [21]	✓	23.91	94.24	67.28	47.40	58.00	71.42	86.14	88.63	67.62	70.66	67.53
	ZLaP [20]	✓	25.40	93.10	65.60	48.60	55.60	73.50	86.90	87.10	67.40	71.50	67.47
	ZERO [7]	✓	25.21	93.66	68.04	46.12	34.33	67.68	86.53	87.75	65.03	67.77	64.21
	BCA [67]	✓	28.59	94.69	66.86	53.49	56.63	73.12	85.97	90.43	68.41	67.59	68.58
	OGA [10]	✓	23.20	93.60	68.10	47.90	54.20	69.20	85.60	89.40	67.90	71.40	67.05
	TCA [52]	✓	24.87	93.63	65.33	46.16	**70.43**	73.33	85.31	89.53	65.92	72.38	68.69
	Dota [12]	✓	25.59	94.32	**69.48**	47.87	57.65	74.67	87.02	91.69	69.70	72.06	69.01
	ADAPT	✓	28.95	94.48	68.19	55.20	68.19	75.56	83.81	92.01	**70.57**	70.66	**70.76**
Trans.	GDA-CLIP [51]	✓	18.69	87.53	60.78	46.81	49.92	72.65	78.25	89.90	63.60	68.70	63.68
	ZLaP [20]	✓	26.30	91.80	66.80	46.00	57.70	67.90	**87.20**	87.90	67.80	73.80	67.32
	TransCLIP [59]	✓	26.90	92.70	69.40	49.50	65.10	76.70	87.10	92.60	68.90	74.40	70.33
	Frolic [69]	✓	**31.40**	95.10	69.10	56.10	58.50	74.80	87.10	**92.90**	70.80	**75.20**	71.10
	StatA [58]	✓	24.70	94.20	68.00	48.40	**67.30**	75.20	87.10	92.40	68.70	73.50	69.95
	ADAPT	✓	30.81	**95.46**	**71.32**	**56.86**	65.93	**80.11**	85.15	92.59	**72.25**	73.86	**72.43**
	Oracle ADAPT	✓	41.88	98.26	82.89	60.87	56.51	81.93	85.74	92.61	80.04	90.14	77.09

Table 3: Top-1 accuracy (%) comparison on corruption robustness.

	Method	**Blur**				**Weather**				**Digital**				**Noise**			**Avg.**
		Defo.	Glas.	Moti.	Zoom	Snow	Fros.	Fog	Brig.	Cont.	Elas.	Pix.	JPEG	Gauss.	Shot	Impu.	
Online	CLIP [39]	24.25	15.71	24.46	22.60	33.08	31.06	37.61	55.62	17.11	13.43	33.04	33.70	13.25	14.16	13.48	25.50
	TPT [33]	**27.56**	15.48	26.16	**26.94**	**36.74**	34.28	39.38	60.22	16.96	15.64	**40.74**	**37.90**	10.64	11.94	10.92	27.43
	DiffTPT [9]	25.63	16.96	26.74	25.40	35.99	34.57	39.83	59.01	17.32	**17.16**	38.43	35.47	12.97	13.60	13.21	27.49
	TDA [21]	26.53	17.91	**27.35**	25.90	36.50	**34.84**	40.53	58.57	**20.16**	16.62	35.65	36.69	15.42	16.46	**16.03**	28.34
	DMN [65]	26.06	17.19	26.61	25.23	34.81	33.48	38.93	58.70	19.38	15.40	35.32	36.49	14.33	15.33	14.69	27.46
	ADAPT	26.30	**18.01**	27.31	25.54	36.19	34.67	**40.96**	**60.29**	19.95	16.09	37.44	37.22	**15.76**	**16.84**	15.90	**28.56**
Trans.	ZLaP [20]	24.88	16.13	25.77	24.36	34.43	32.63	38.56	58.42	17.53	14.21	33.72	35.52	12.83	14.03	13.27	26.42
	TransCLIP [59]	25.35	16.40	25.53	23.22	34.58	32.47	39.65	59.04	17.72	14.76	35.22	35.53	14.82	16.11	15.60	27.07
	StatA [58]	20.23	13.29	20.38	18.84	31.30	29.80	34.58	54.79	11.24	11.80	26.31	33.20	9.58	10.52	10.12	22.40
	ADAPT	**27.98**	**19.78**	**29.00**	**27.38**	**38.09**	**36.44**	**42.43**	**62.21**	**21.94**	**18.40**	**39.89**	**38.23**	**17.71**	**18.81**	**18.09**	**30.29**

Experiments

更丰富、更具描述性的初始化可提升性能

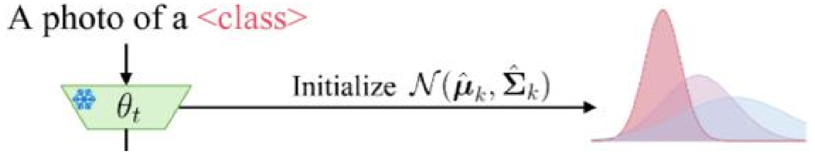


Table 6: Efficiency comparison on ImageNet.

	Method	BP-free	Acc (%) ↑	Gain (%) ↑	Time ↓	Mem.(GB) ↓
Online	CLIP [39]	✓	66.74	-	8m	0.79
	TPT [33]	✗	68.95	2.21	9h 45m	4.29
	DiffTPT [9]	✗	70.30	3.56	> 20h	4.60
	TDA [21]	✓	69.51	2.77	50m	0.84
	TPS [46]	✗	70.38	3.64	1h 19m	1.71
	ADAPT	✓	70.91	4.17	1h 11m	0.93
Trans.	GDA-CLIP [51]	✓	64.13	-2.61	1.31m	10.03
	TransCLIP [59]	✓	70.30	3.56	1.34m	16.17
	StatA [58]	✓	69.90	3.16	1.5m	20.74
	ADAPT	✓	71.56	4.82	0.73m	3.37

Table 7: Mean initialization comparison.

Mean initialization $\hat{\mu}$	Task 1	Task 2	Task 3
Vanilla [39]	64.70	27.52	67.43
Ensemble [39]	66.56	28.54	67.74
CLIP Template [39]	66.51	28.44	67.62
GPT [70]	66.53	28.56	70.76
GPT & Ensemble	66.57	28.98	69.95
GPT & CLIP Template	66.58	28.91	69.90

Thanks