Backpropagation-Free Test-Time Adaptation via Probabilistic Gaussian Alignment

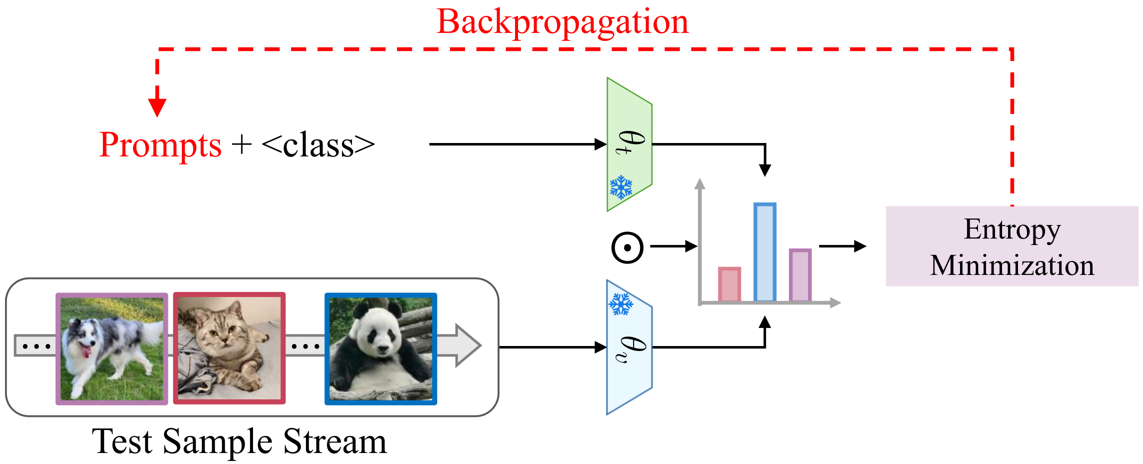





Youjia Zhang | Youngeun Kim | Young-Geun Choi | Hongyeob Kim | Huiling Liu | Sungeun Hong

Challenges

- **Test-Time Adaptation (TTA):** an effective way to improve zero-shot robustness under <u>distribution shifts</u> by adapting to unlabeled test data during inference
- **Limitations**
  - High computational cost ☹
  - Lack explicit class distribution modeling ☹



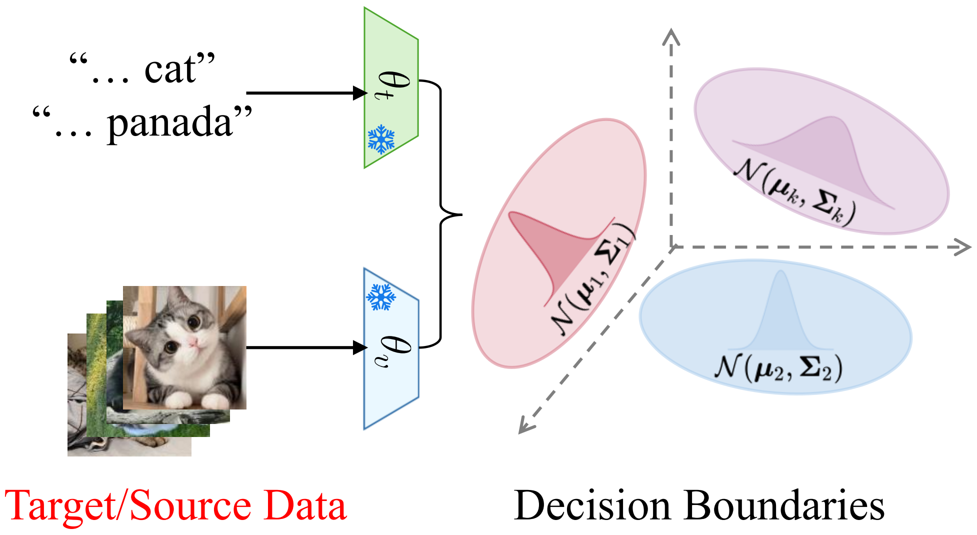
(a) Backpropagation-required TTA

TTA Method	Backpropagation-Free	Distribution-Aware	Task Setting	
			Online	Transductive
Prompt Tuning	✗	✗	✓	✗
Adapter Tuning	✗	✗	✓	✗
Similarity Score	✓	✗	✓	✗

(b) Online TTA Methods

Challenges

- **Gaussian Discriminant Analysis (GDA):** a classical probabilistic framework that models class-conditional feature distributions and assigns labels based on likelihood estimation
- **Limitations**
  - Need full target/source access → not feasible for online settings ☹



(a) GDA-based Transductive TTA

TTA Method	Backpropagation-Free	Distribution-Aware	Task Setting	
			Online	Transductive
Transductive Learning	✓	✓	✗	✓

(b) Transductive Learning in TTA

Core Idea

- Can we design a backpropagation-free and distribution-aware TTA framework that seamlessly supports both online and transductive adaptation?

TTA Method	Backpropagation-Free	Distribution-Aware	Task Setting	
			Online	Transductive
Prompt Tuning	✗	✗	✓	✗
Adapter Tuning	✗	✗	✓	✗
Similarity Score	✓	✗	✓	✗
Transductive Learning	✓	✓	✗	✓
ADAPT (Ours)	✓	✓	✓	✓

Motivation

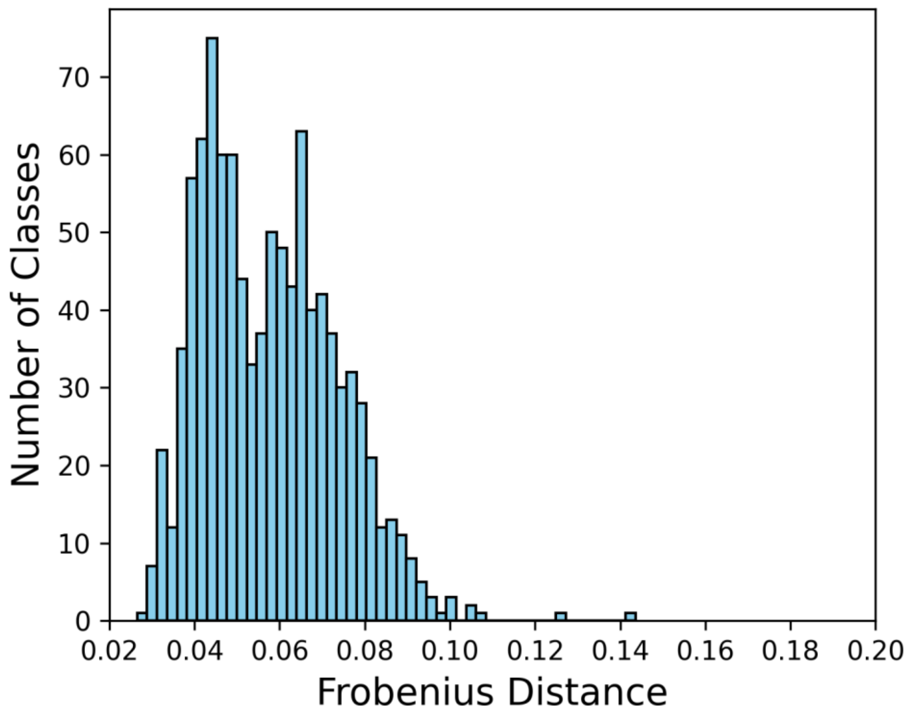
- Observation 1: Gaussianity of class conditional features
  - Class-wise features exhibit strong Gaussian patterns in CLIP space

	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

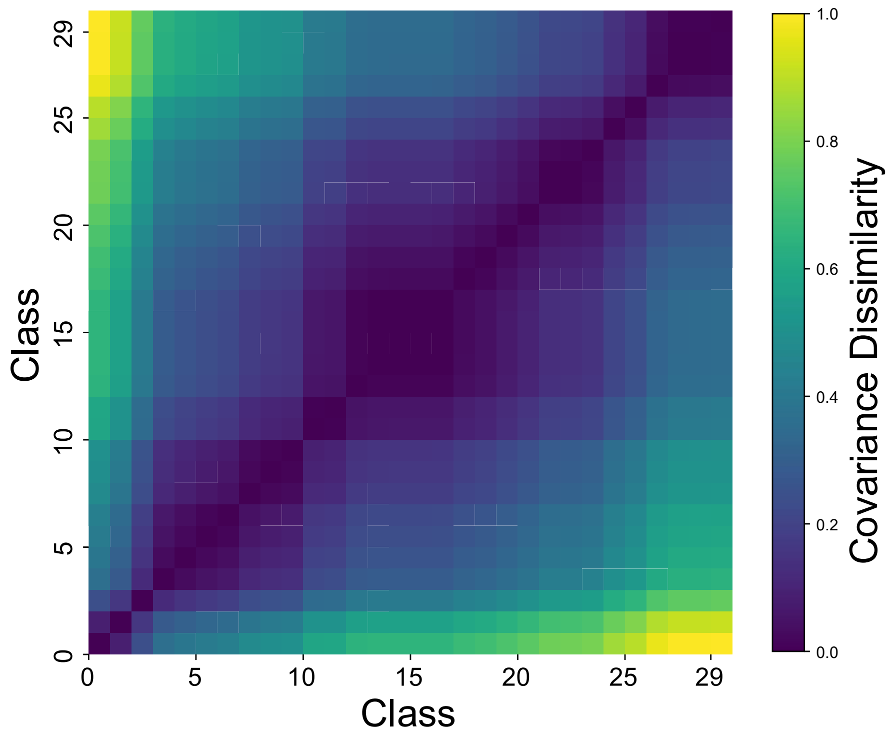
Projection-based normality test results across class-conditional features

Motivation

- Observation 2: Strong alignment of class-wise Σ_k and shared covariance Σ
  - Class-wise covariance matrices are nearly identical, supporting a shared-covariance assumption



(a) Frobenius distance between Σ_k and Σ



(b) class-wise covariance dissimilarity

Motivation

- Observation 1: Gaussianity of class conditional features
- Observation 2: Strong alignment of class-wise Σ_k and shared covariance Σ

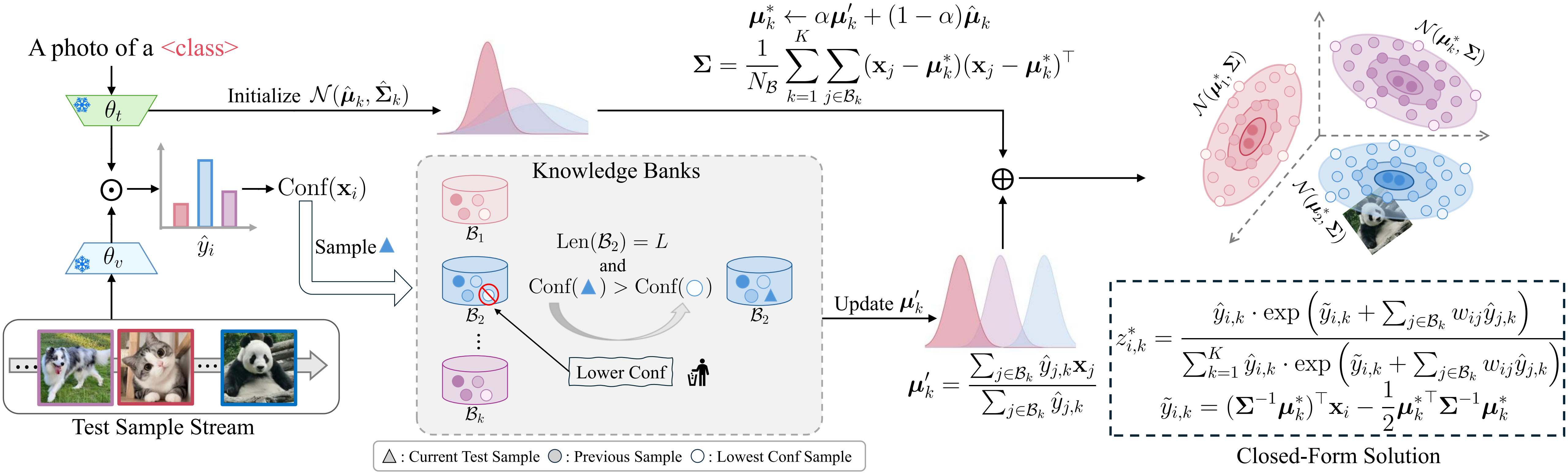
$\Downarrow$

Assumption: CLIP features conditioned on class k follow a Gaussian distribution with a shared covariance matrix:

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

Method: ADAPT

- ADAPT: Backpropagation-free and Distribution-aware Test-time Adaptation



Overview of Online ADAPT

Method: ADAPT

- Online ADAPT
  - Backpropagation-free TTA via GDA: Reframe TTA as a probabilistic inference task by modeling class-conditional likelihoods.

$$\tilde{y}_{i,k} = \mathrm{w}_k^\top \mathbf{x}_i + b_k, \quad \text{where } \mathrm{w}_k = \mathbf{\Sigma}^{-1}\boldsymbol{\mu}_k, b_k = -\frac{1}{2}\boldsymbol{\mu}_k^\top \mathbf{\Sigma}^{-1}\boldsymbol{\mu}_k.$$

  - Correcting Online Likelihood Bias via Constructed Knowledge Banks

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

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

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

  - Closed-form Solution without Sub-iterations

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

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

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

Method: ADAPT

- Transductive ADAPT
  - Extend the online regularized objective to a transductive objective

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

  - Closed-form Solution without Sub-iterations

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

  - One-pass estimate for class means

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

Method: ADAPT

Algorithm 1 ADAPT: Online TTA

Input: Test data $\mathcal{D}_u$, class prototypes $\mathbf{t}$ and knowledge bank size L
Initialize: $\hat{\boldsymbol{\mu}} \leftarrow \mathbf{t}$
for $\mathbf{x}_i \in \mathcal{D}_u$ **do**
    Compute Conf$(\mathbf{x}_i)$ by Eq. (2)
    Update $\mathcal{B}_k$ with $\mathbf{x}_i$ if high-confidence
    Update $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}$ by Eq. (9)-(10)
    Compute z_i^* by Eq. (8)
end for
return $\{z_i^*\}_{i-1}^N$

Algorithm 2 ADAPT: Transductive TTA

Input: Test data $\mathcal{D}_u = \{\mathbf{x}_i\}_{i=1}^N$, class prototypes $\mathbf{t}$ and knowledge bank size L
Initialize: $\hat{\boldsymbol{\mu}} \leftarrow \mathbf{t}$
Compute Conf$(\mathbf{x})$ for all data by Eq. (2)
for $\mathcal{B}_k \in \mathcal{B}$ **do**
    Cache Top-L confidence samples
end for
Update $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}^*$ by Eq. (10)-(67)
Compute $z^* = \{z_i^*\}_{i-1}^N$ by Eq. (8)
return z^*

Experiments

- Task 1: 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**

Experiments

- Task 2: 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

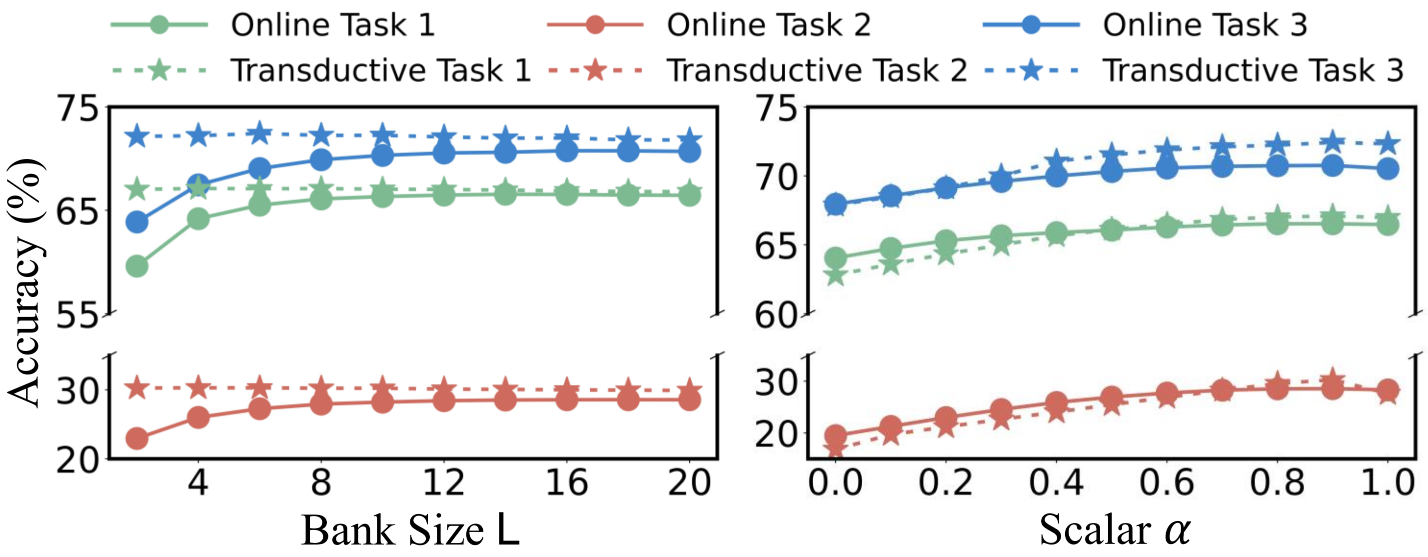
- Task 3: Fine-Grained Categorization

	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

Experiments

- Ablation Studies and Hyperparameter analysis

$\mathcal{B}$	Update μ	Update Σ	Task 1	Task 2	Task 3
✗	✗	✗	59.11	25.50	63.45
✗	✗	✓	49.64	9.58	60.02
✗	✓	✗	61.54	25.42	67.03
✗	✓	✓	49.65	9.58	60.04
✓	✗	✗	64.89	25.08	67.06
✓	✗	✓	64.05	19.49	67.95
✓	✓	✗	65.27	25.67	67.43
✓	✓	✓	**66.53**	**28.56**	**70.76**



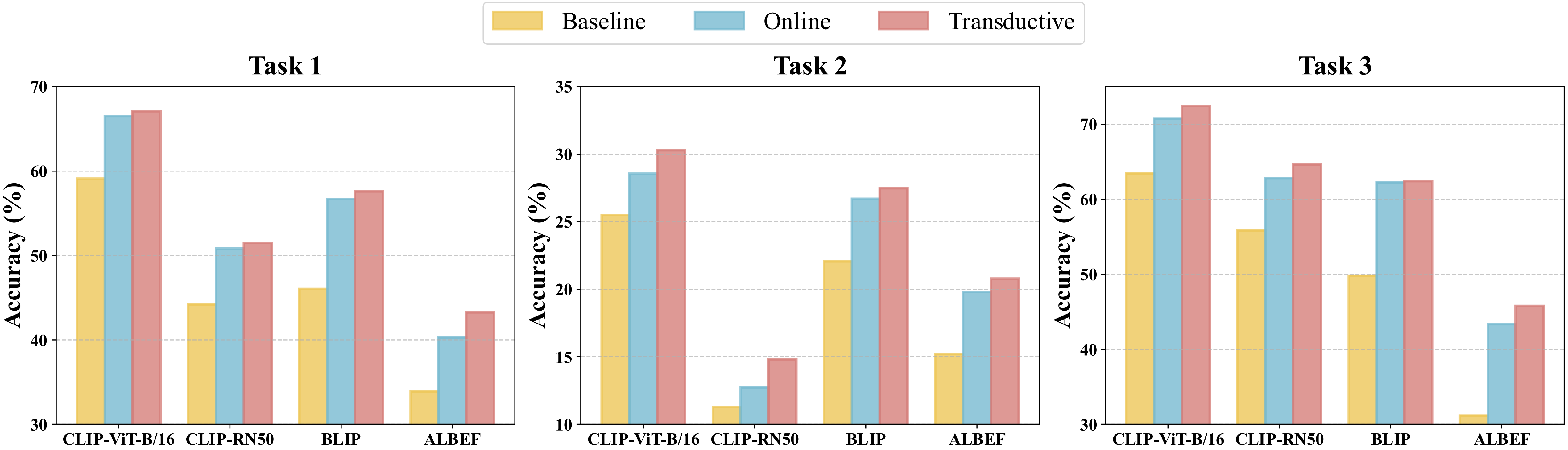
Experiments

- 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

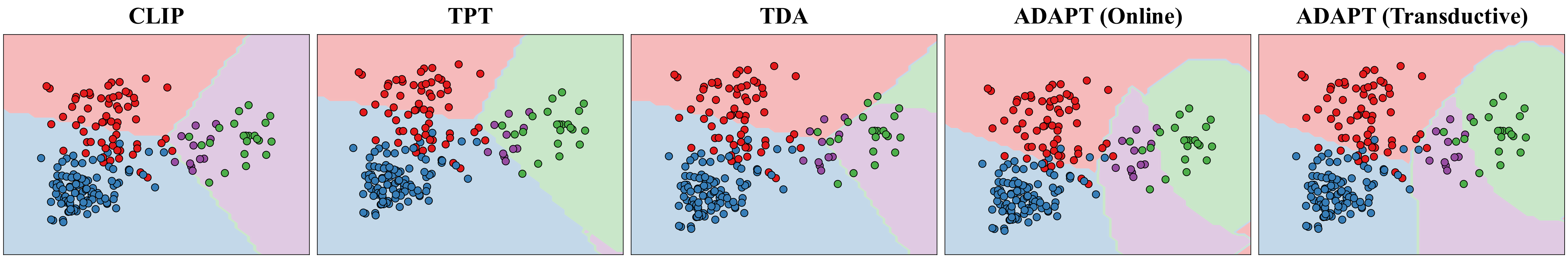
Experiments

- Evaluation with Different VLMs



Experiments

- Visualization of Decision Boundaries on ImageNet-A



Thank you

Project page is here!