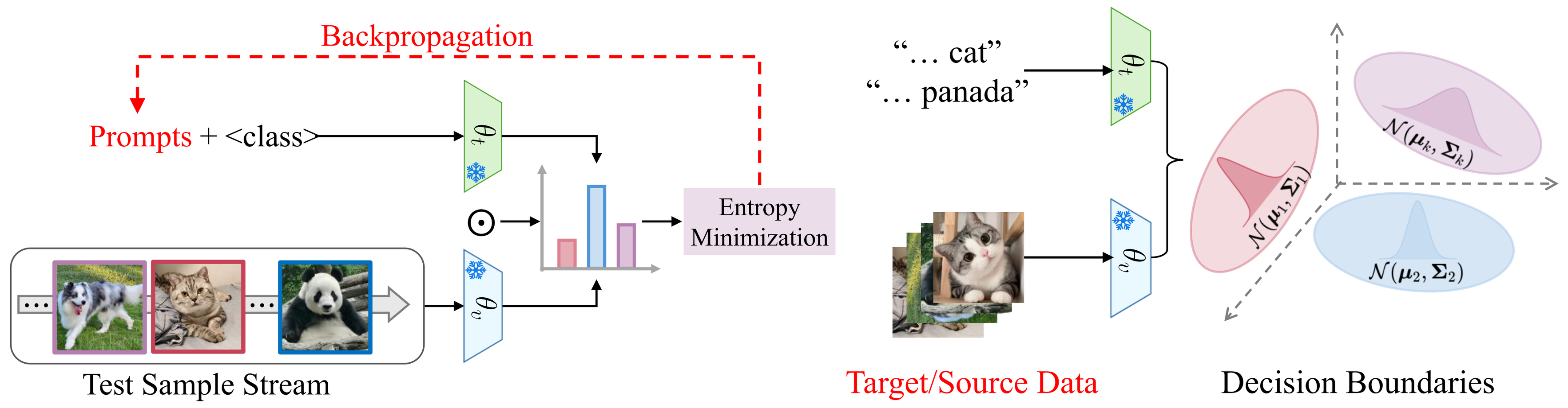
Backpropagation-Free Test-Time Adaptation via Probabilistic Gaussian Alignment

Youjia Zhang[1] Youngeun Kim[2] Young-Geun Choi[1] Hongyeob Kim[1] Huiling Liu[1] Sungeun Hong[1†]

[1]Sungkyunkwan University [2]Amazon †Corresponding author

Project Page

Introduction



(a) Backpropagation-required Online TTA (b) GDA-based Transductive TTA

- **Test-Time Adaptation (TTA):** an effective way to improve zero-shot robustness under *distribution shifts* by adapting to unlabeled test data during inference
  - **Limitations of existing Online TTA**
    - High computational cost
    - Lack explicit class distribution modeling
- **Gaussian Discriminant Analysis (GDA)**: a classical probabilistic framework that models class-conditional feature distributions and assigns labels based on likelihood estimation
  - **Limitations of GDA for TTA:**
    - Need full target/source access → not feasible for online settings
- **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)	✓	✓	✓	✓

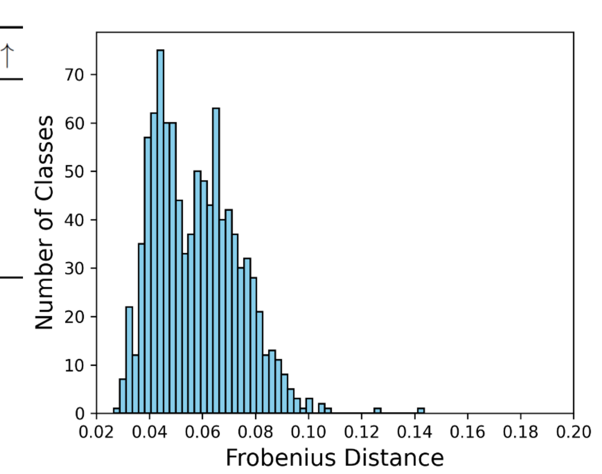
Comparison with existing TTA methods

Motivation

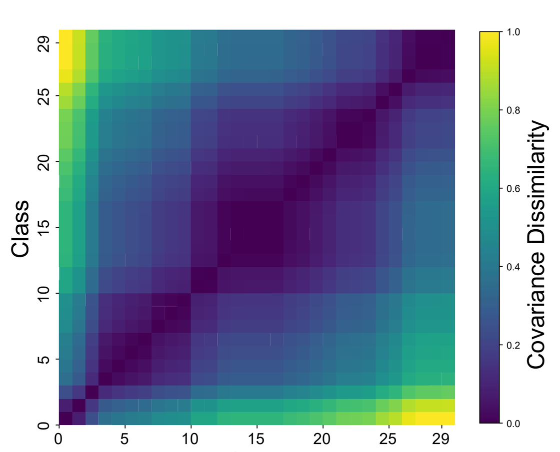
Observation 1: Gaussianity of class conditional features
Observation 2: Strong alignment of class-wise Σ_k and shared covariance Σ

	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

(a) Projection-based normality test results



(b) Frobenius distance between Σ_k and Σ



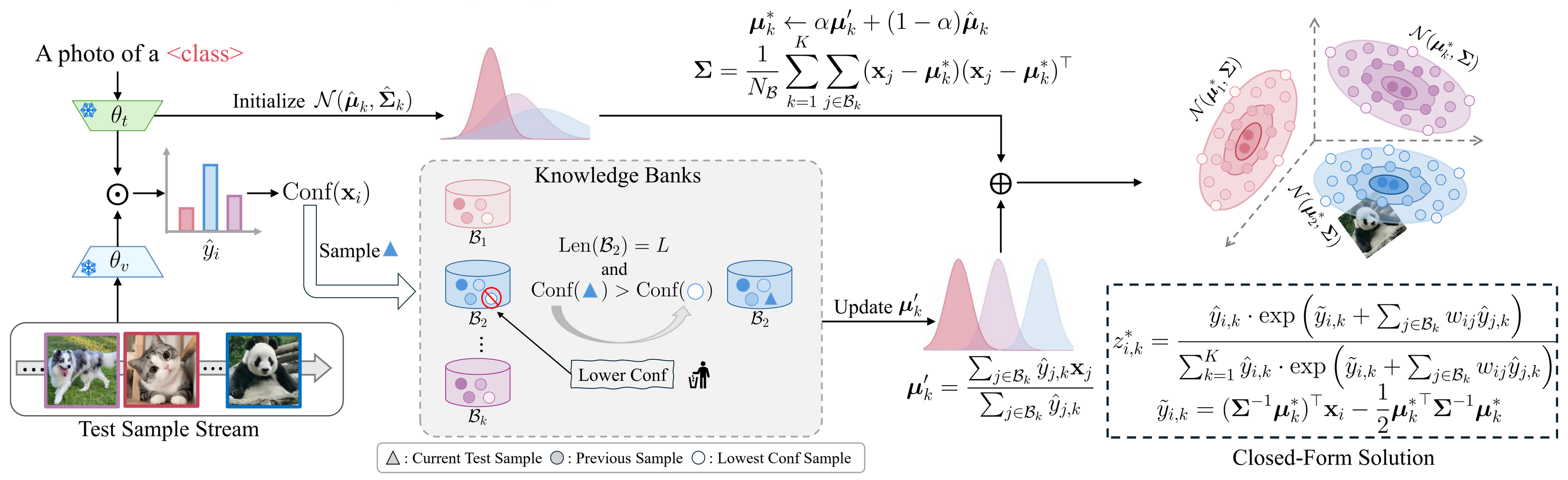
(c) class-wise covariance dissimilarity

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: Backpropagation-free and Distribution-aware TTA



- Gaussian Modeling: estimate class-conditional feature distributions.
- BP-free Adaptation: Training-free; works in both online & transductive modes.
- Closed-form Update: One-pass; efficient; no iteration or fine-tuning required

Online ADAPT

- Backpropagation-free TTA via GDA:

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

- Correcting Online Likelihood Bias via Constructed Knowledge Banks:

$$\mathcal{L}_{\text{online}}(z_i,\boldsymbol{\mu},\boldsymbol{\Sigma}) = -z_i^\top\log\mathbb{P}_i + \mathcal{R}(z_i;\hat{y}_i) + \mathcal{R}(z_i;\mathcal{B}),$$
$$\text{where } \mathcal{R}(z_i;\hat{y}_i) = \mathrm{KL}(z_i\|\hat{y}_i) + \beta\textstyle\sum_{k=1}^K \mathrm{KL}\Big(\mathcal{N}(\hat{\boldsymbol{\mu}}_k,\hat{\boldsymbol{\Sigma}}_k)\|\mathcal{N}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma})\Big),$$
$$\mathcal{R}(z_i;\mathcal{B}) = -\textstyle\sum_{j\in\mathcal{B}}\hat{y}_j^\top\log\mathbb{P}_j - \sum_{j\in\mathcal{B}} w_{ij}z_i^\top\hat{y}_j.$$

  - Online Negative Log-Likelihood $-z_i^\top\log\mathbb{P}_i$
  - CLIP Prior-based Regularization $\mathcal{R}(z_i;\hat{y}_i)$
  - Knowledge Bank-guided Consistency Regularization $\mathcal{R}(z_i;\mathcal{B})$

- One-pass distribution estimation:

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

- Closed-form Solution without Sub-iterations:

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

Transductive ADAPT

- Extend the online regularized objective to a transductive objective:

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

- Class means μ_k estimation:

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

Algorithm 1 ADAPT: Online TTA

```
Input: Test data D_u, class prototypes t and knowledge bank size L
Initialize: μ̂ ← t
for x_i ∈ D_u do
    Compute Conf(x_i) by Eq. (2)
    Update B_k with x_i if high-confidence
    Update μ* and Σ 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 D_u = {x_i}_{i=1}^N, class prototypes t and knowledge bank size L
Initialize: μ̂ ← t
Compute Conf(x) for all data by Eq. (2)
for B_k ∈ B do
    Cache Top-L confidence samples
end for
Update Σ and μ* by Eq. (10)-(67)
Compute z* = {z_i*}_{i-1}^N by Eq. (8)
return z*
```

Experiments

Main results on fine-grained categorization:

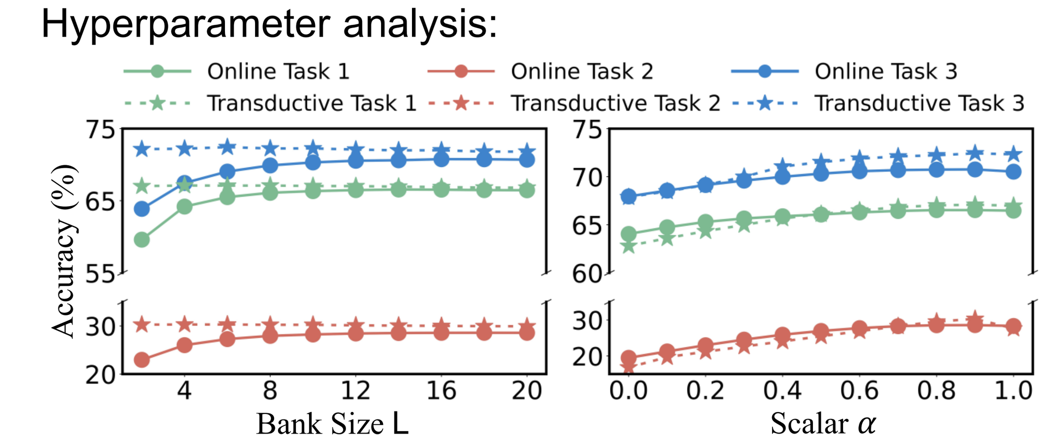
	Method	BP-free	Aircraft	Caltech	Cars	DTD	EuroSAT	Flower	Food101	Pets	Sun397	UCF101	Avg.
Online	CLIP [39]	-	23.70	92.98	65.24	44.44	41.42	67.28	83.80	87.98	62.55	65.08	63.45
	TPT [33]	✗	24.78	94.16	66.87	47.75	42.44	68.98	84.67	87.79	65.50	68.04	65.10
	DiffTPT [9]	✗	25.60	92.49	67.01	47.00	43.13	70.10	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

Ablation Studies and Further Analysis.

Ablation study:

$\mathcal{B}$	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**

Hyperparameter analysis:



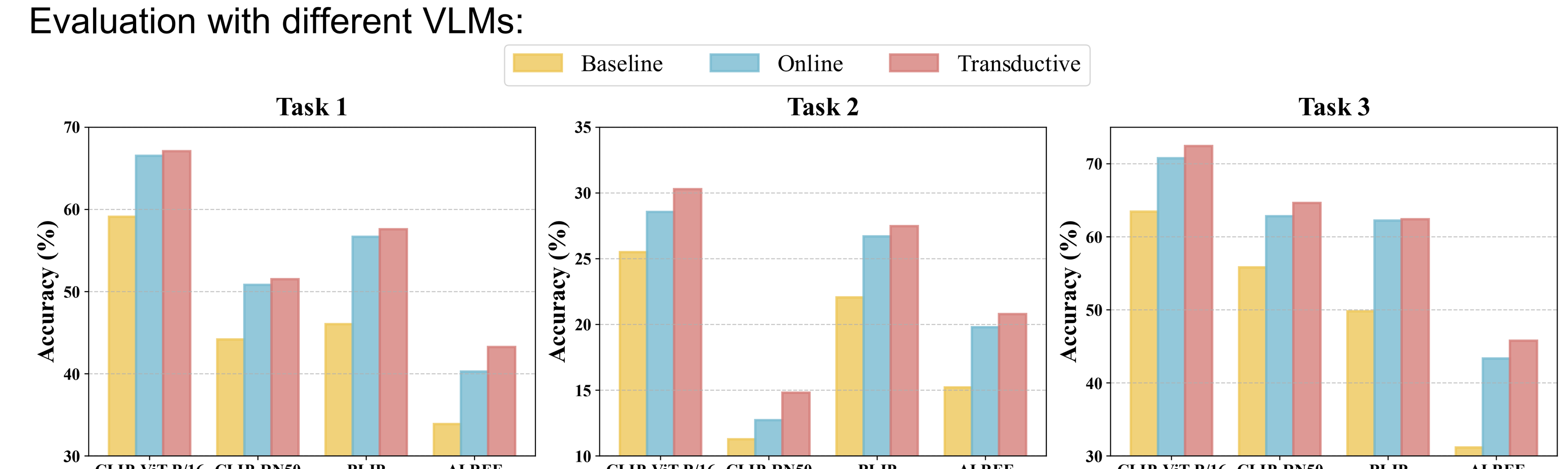
Efficiency comparison:

	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

Mean initialization comparison:

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

Evaluation with different VLMs:



Visualization of decision boundaries: