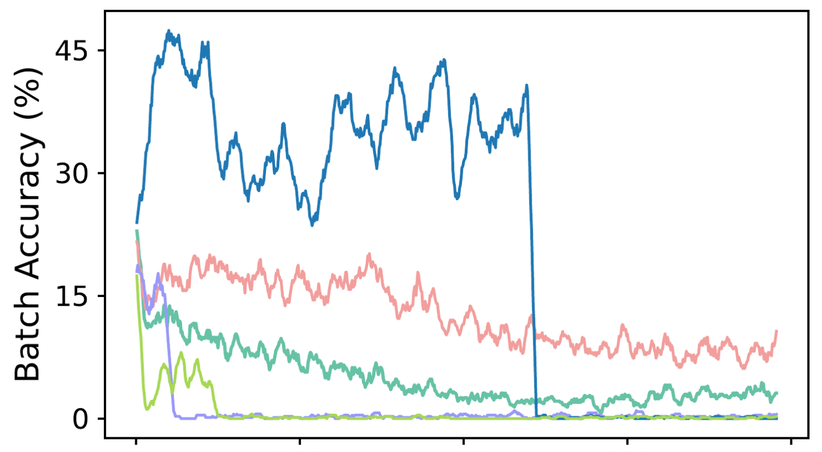
Adapting in the Dark: Efficient and Stable Test-Time Adaptation for Black-Box Models

Yunbei Zhang[1] Shuaicheng Niu[2] Chengyi Cai[3] Feng Liu[3] Jihun Hamm[1]
[1]Tulane University [2]Nanyang Technological University [3]University of Melbourne

Paper

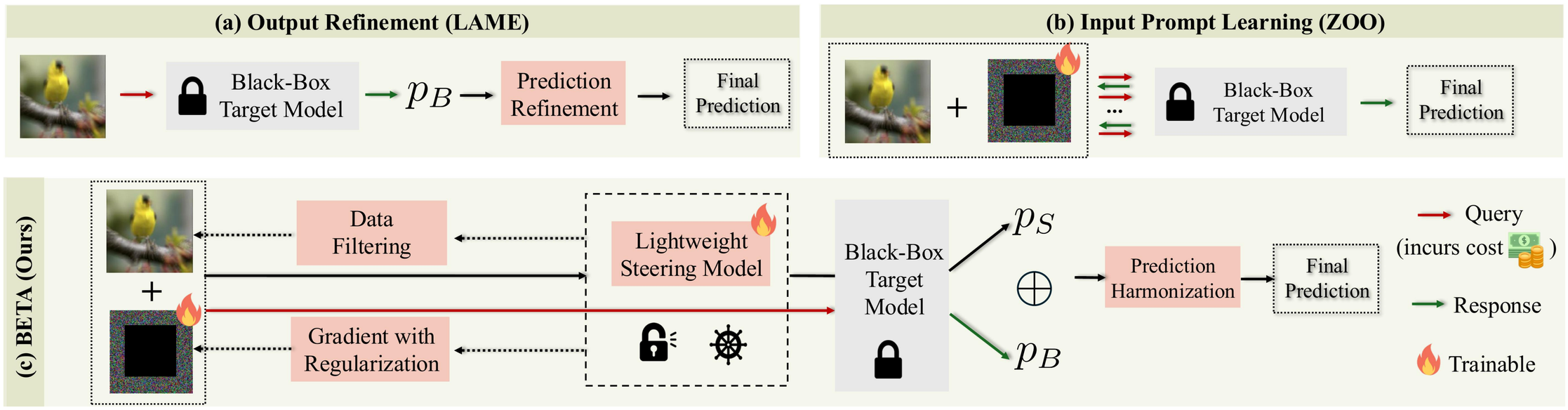
Motivation & Challenge

- **Powerful models** (e.g., GPT-5) are deployed as **black-box APIs**, but existing TTA methods require white-box/gray-box access — inapplicable.
- ZOO-based black-box methods suffer from:
  1. **Unstable optimization** under unsupervised objectives → collapse (32.6% → 4.1%)
  2. **High cost**: 16+ API calls per sample → prohibitive latency & expense
- **Our goal**: an **efficient** and **stable** TTA framework for black-box models.



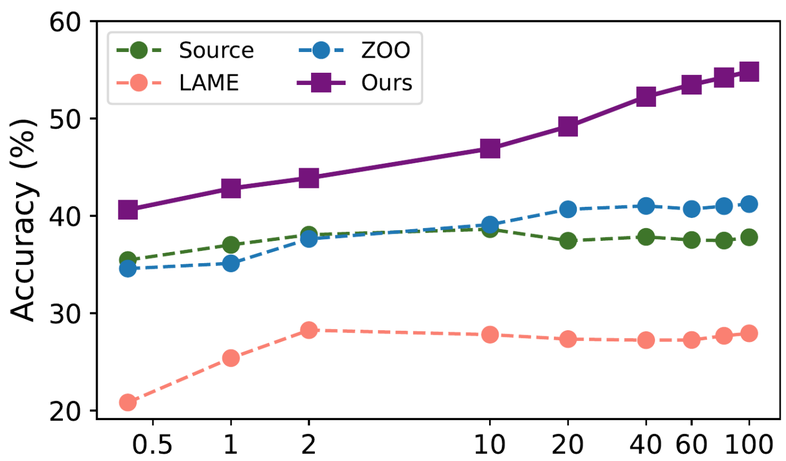
Method

Step 1: Prediction Harmonization: $p_H(x') = \alpha \cdot p_S(x') + (1-\alpha) \cdot p_B(x')$. Only **1 API query** per sample.
Step 2: Consistency Regularization KL divergence between clean & prompted predictions → **prevent collapse**.
Step 3: Data Filtering Entropy-based filtering + **confidence weighting** → stable optimization.



Results

API Efficiency with Real-world **Clarifai** API



Computation Efficiency Analysis

Method	#API /Img	Local Compute	Mem (MB)	Time (ms)	Acc (%)	Gain (%)
Source	1	✗	-	45	55.5	-
LAME	1	✓	2	46	54.1	-1.4
ZOO-SPSA-GC†	16	✓	52	450	55.1	-0.4
TT-Aug‡	64	✓	-	1,800	55.6	+0.1
DDA§	2	✓	23,427	12,722	56.9	+1.4
BETA (ViT-Tiny)	1	✓	1,292	47	58.2	+2.7
BETA (ViT-Small)	**1**	✓	2,616	48	**62.6**	+7.1

Classification Accuracy on ImageNet-C with ViT-B/16 as Black-box model

Access	Method	Noise			Blur				Weather				Digital				Avg.	Gain
		Gauss.	Shot	Impul.	Defoc.	Glass	Motion	Zoom	Snow	Frost	Fog	Bright.	Contr.	Elastic	Pixel.	JPEG		
	Source	56.8	56.8	57.5	46.9	35.6	53.1	44.8	62.2	62.5	65.7	77.7	32.6	46.0	67.0	67.6	55.5	0.0
□	TENT	60.3	61.6	61.8	59.2	56.5	63.5	59.1	54.2	64.5	2.2	79.1	67.4	61.5	72.5	70.6	59.6	+4.1
□	SAR	59.1	60.5	60.6	57.1	55.6	61.5	57.4	65.8	63.4	67.4	78.7	62.6	62.2	72.0	70.2	63.6	+8.1
□	CoTTA	63.3	63.9	64.5	55.0	51.0	63.5	56.1	68.8	69.2	71.2	78.3	9.6	64.3	73.4	71.2	61.6	+6.1
□	ETA	60.9	62.2	62.2	59.5	57.4	63.6	60.1	68.3	65.8	71.5	79.3	66.9	64.9	72.9	71.1	65.8	+10.3
▒	T3A	56.4	56.9	57.3	47.9	37.8	54.3	46.9	63.6	60.8	68.5	78.1	38.3	50.0	67.6	69.1	56.9	+1.4
▒	FOA*	57.0	58.5	57.8	51.7	35.0	37.1	27.2	20.2	11.9	72.2	76.8	0.6	39.1	66.7	67.0	44.9	-10.6
■	LAME	56.5	56.5	57.2	46.4	34.7	52.7	44.2	58.4	61.5	63.1	77.4	24.7	44.6	66.6	67.2	54.1	-1.4
■	ZOO-CMA	58.2	59.6	60.3	50.8	38.6	55.2	45.7	58.5	59.6	59.7	76.7	4.1	49.8	71.2	70.0	54.5	-1.0
■	ZOO-RGF	59.6	58.7	60.4	47.7	37.8	53.5	44.6	58.2	61.7	63.4	76.7	26.8	49.4	70.7	70.2	56.0	+0.5
■	ZOO-SPSA-GC†	59.6	58.7	60.2	47.9	38.0	53.7	44.7	58.2	61.7	63.6	76.7	12.7	49.4	70.7	70.2	55.1	-0.4
■	TT-Aug‡	55.4	54.2	55.2	43.7	48.6	48.9	45.5	57.8	63.1	60.0	76.9	49.6	41.7	65.7	67.8	55.6	+0.1
■	DDA§	**64.7**	**65.0**	**64.6**	46.3	41.3	51.7	43.7	59.1	61.3	45.0	74.9	40.6	54.4	72.2	68.4	56.9	+1.4
■	**BETA (Ours)**	60.5	60.7	61.1	**54.5**	**52.2**	**59.9**	**56.3**	**63.6**	**64.7**	**66.1**	**78.1**	**53.4**	**62.1**	**73.3**	**72.0**	**62.6**	+7.1