TEA: Test-time Energy Adaptation

Paper Code

Yige Yuan, Bingbing Xu, Liang Hou, Fei Sun, Huawei Shen, Xueqi Cheng

CAS Key Laboratory of AI Safety, Institute of Computing Technology, Chinese Academy of Sciences
University of Chinese Academy of Sciences
Kuaishou Technology

Homepage WeChat

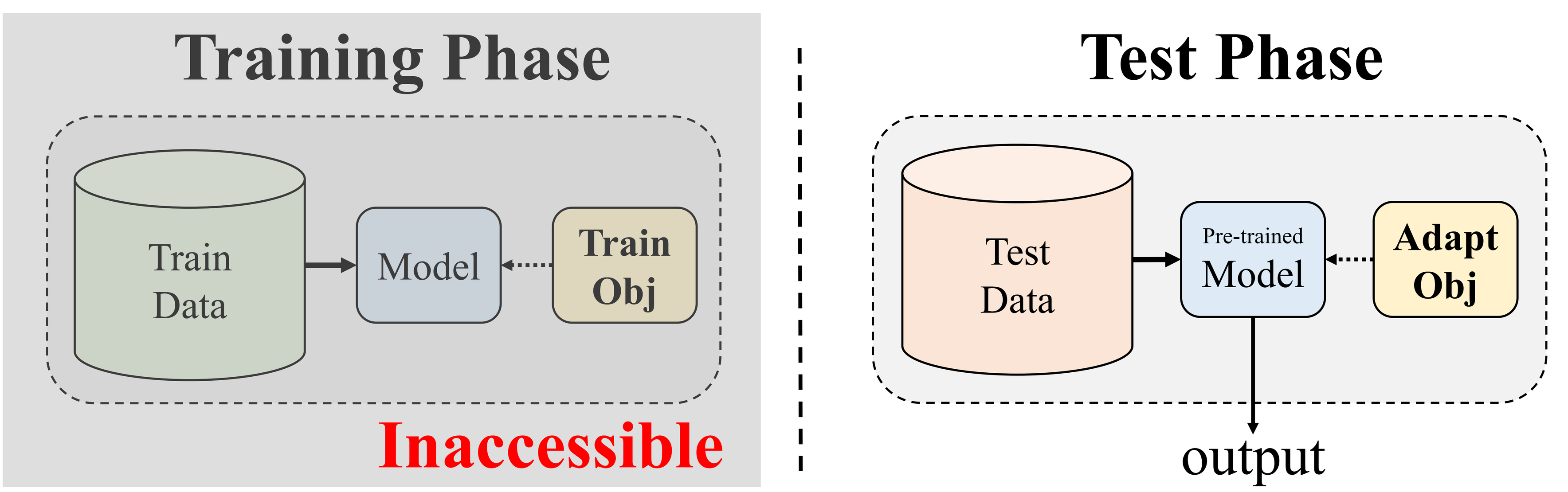
Email: yuanyige20z@ict.ac.cn

CVPR JUNE 17-21, 2024 SEATTLE, WA

INTRODUCTION

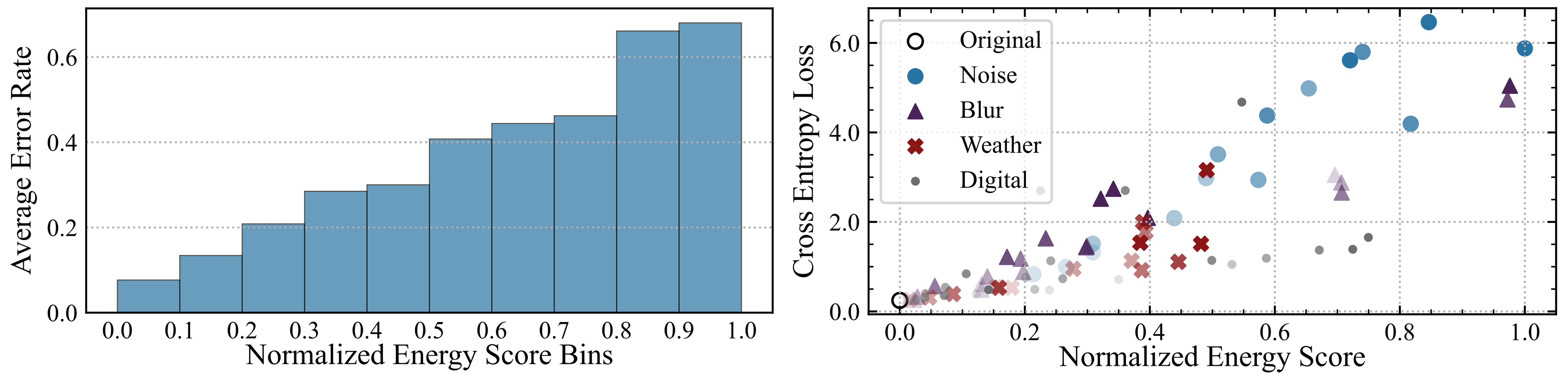
Objective: Improving model **generalizability** when test data diverges from training distribution, without requiring access to training data and processes.
Weakness of existing methods: Current TTA methods fail to address the fundamental issue: **covariate shift**, i.e., the decreased generalizability can be attributed to the model's reliance on the marginal distribution of the training data, which may impair model calibration and introduce confirmation bias.
Motivation: Transforming the trained classifier into an **energy-based** model and aligning the model's distribution with the test data's, enhancing its ability to perceive test distributions and thus improving overall generalizability.

BACKGROUND

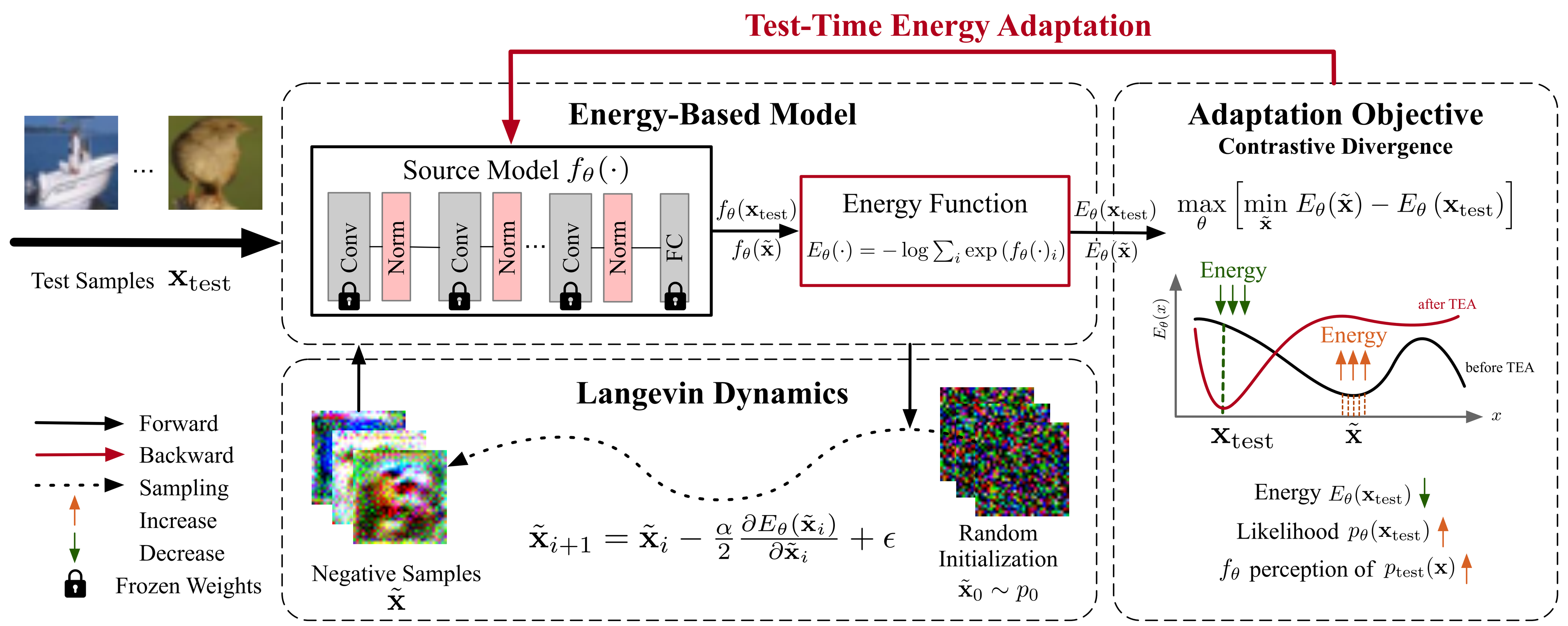


MOTIVATION

Low Energy ➡ High Probability High Performance
High Energy ➡ Low Probability Low Performance



METHOD



Treating Classifier As EBM

Pretrained Classifier

$$p_\theta(y \mid \mathbf{x}) = \frac{\exp(f_\theta(\mathbf{x})[y])}{\sum_{y'} \exp(f_\theta(\mathbf{x})[y'])}$$

introduce unknown normalizing constant Z_θ

$$p_\theta(\mathbf{x}, y) = \frac{\exp(f_\theta(\mathbf{x})[y])}{Z_\theta}$$

marginalize out y

$$p_\theta(\mathbf{x}) = \sum_y p_\theta(\mathbf{x}, y) = \frac{\sum_y \exp(f_\theta(\mathbf{x})[y])}{Z_\theta}$$

substitute E_θ

$$p_\theta(\mathbf{x}) = \frac{\exp(-E_\theta(\mathbf{x}))}{Z_\theta}$$

Energy Based Model

$$E_\theta(\mathbf{x}) = -\log \sum_y \exp(f_\theta(\mathbf{x})[y])$$

Inject Test Distribution Into Classifier

Contrastive Divergence

$$\frac{\partial \log p_\theta(\mathbf{x}_{\text{test}})}{\partial \theta} = \mathbb{E}_{\tilde{\mathbf{x}} \sim p_\theta}\left[\frac{\partial E_\theta(\tilde{\mathbf{x}})}{\partial \theta}\right] - \frac{\partial E_\theta(\mathbf{x}_{\text{test}})}{\partial \theta}$$

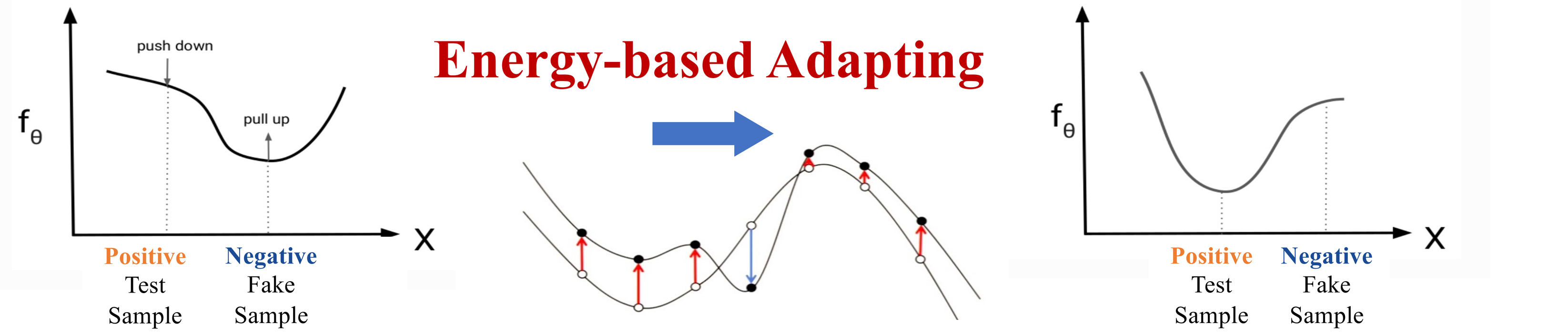
Stochastic Gradient Langevin Dynamics

$$\tilde{\mathbf{x}}_{t+1} = \mathbf{x}_t - \frac{\alpha}{2}\frac{\partial E_\theta(\tilde{\mathbf{x}}_t)}{\partial \tilde{\mathbf{x}}_t} + \sqrt{\alpha}\epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}),$$

Overall Objective

$$\max p_\theta(\mathbf{x}_{\text{test}}) = \max_\theta \left[\min_{\tilde{\mathbf{x}}} E_\theta(\tilde{\mathbf{x}}) - E_\theta(\mathbf{x}_{\text{test}})\right]$$

Energy-based Adapting

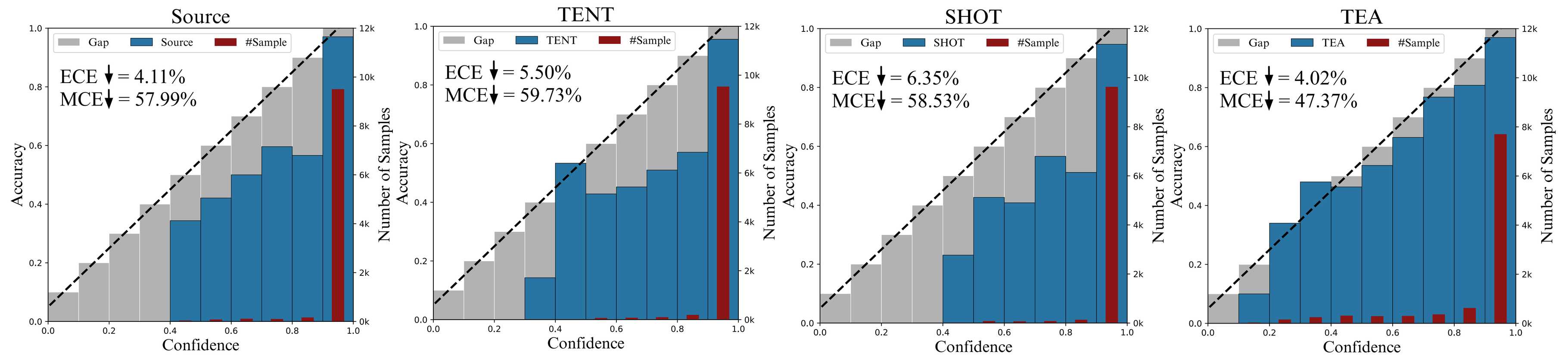


EXPERIMENTS

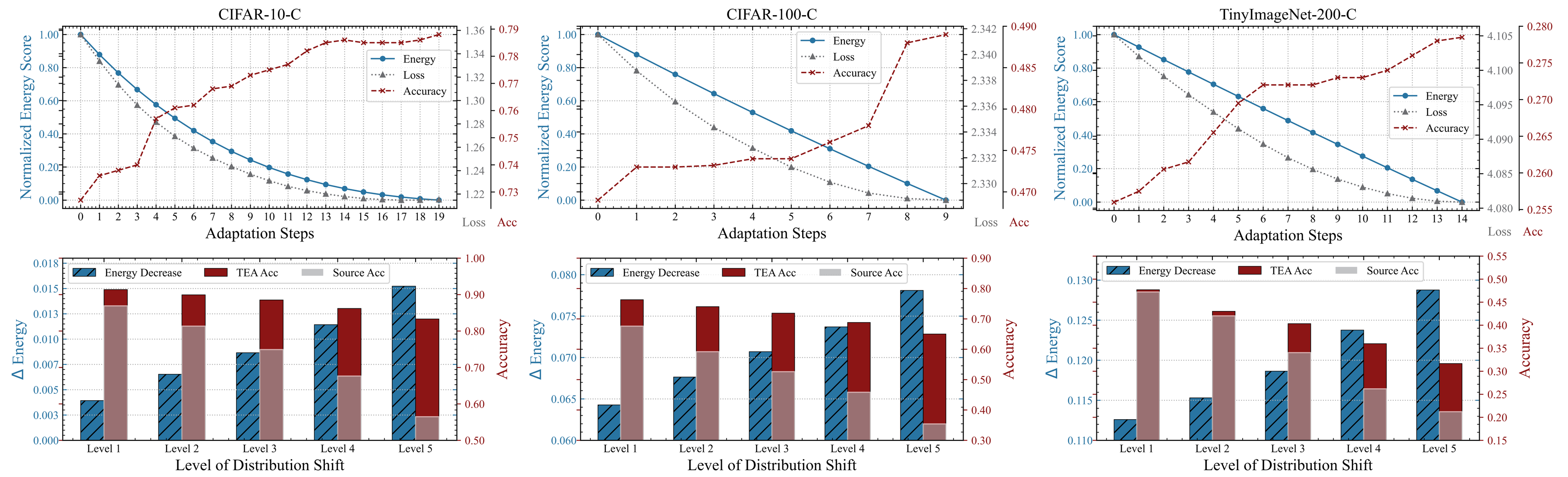
TEA's Adaptation Performance

WRN-28-10 BatchNorm		CIFAR-10(C) Clean	CIFAR-10(C) Corr Severity 5		CIFAR-10(C) Corr Severity 1-5		CIFAR-100(C) Clean	CIFAR-100(C) Corr Severity 5		CIFAR-100(C) Corr Severity 1-5		Tiny-ImageNet(C) Clean	Tiny-ImageNet(C) Corr Severity 5		Tiny-ImageNet(C) Corr Severity 1-5	
		Acc (↑)	Acc (↑)	mCE (↓)	Acc (↑)	mCE (↓)	Acc (↑)	Acc (↑)	mCE (↓)	Acc (↑)	mCE (↓)	Acc (↑)	Acc (↑)	mCE (↓)	Acc (↑)	mCE (↓)
Source		94.77	56.47	100.00	73.45	100.00	81.79	35.39	100.00	52.12	100.00	63.19	21.21	100.00	34.13	100.00
Norm	BN [52]	93.97	79.56	52.65	85.63	60.00	80.83	60.06	63.54	68.11	69.42	45.04	27.74	93.42	34.27	100.96
	DUA* [41]	-	80.10	50.78	-	-	-	-	-	-	-	-	-	-	-	-
Pseudo	PL [34]	93.75	51.42	106.98	72.62	99.37	80.52	53.40	72.12	64.53	75.29	47.84	28.26	91.22	39.83	91.67
	SHOT [36]	93.25	74.77	63.19	82.35	72.61	80.52	56.53	68.01	66.00	73.28	47.95	29.14	90.16	**40.01**	**91.41**
Entropy	TENT [60]	93.66	81.41	48.13	86.75	56.17	80.14	63.09	59.42	69.47	67.80	39.54	26.31	95.52	32.03	104.49
	ETA [45]	93.96	79.58	52.64	85.63	59.99	80.65	59.82	64.52	67.17	72.40	43.20	27.28	94.12	33.46	102.25
	EATA [45]	93.96	79.59	52.62	85.64	59.98	80.68	60.24	63.75	67.48	71.66	43.42	27.28	94.09	33.47	102.24
	SAR [46]	93.97	79.77	51.94	85.83	58.97	80.84	62.95	59.37	70.01	65.99	41.58	28.21	92.82	34.60	100.47
Energy	TEA	**94.09**	**83.34**	**43.69**	**87.88**	**52.00**	**80.88**	**65.10**	**56.07**	**71.22**	**63.72**	**51.65**	**31.67**	**87.99**	39.96	92.12

TEA's Improvements in Confidence Calibration



Relation between TEA's Energy Reduction and Generalizability Enhancement



TEA's Distribution Perception and Generation