



# Learning from Unknown for Open-Set Test-Time Adaptation

Taki Hasan Rafi, Amit Agarwal, Hitesh Patel, Dong-Kyu Chae



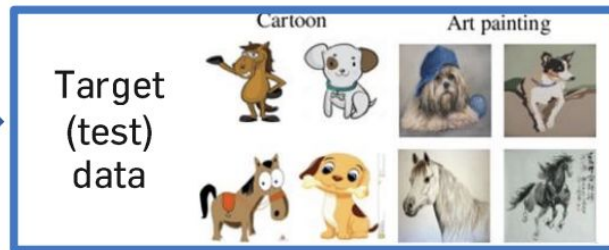
WACV 2026 (Oral)



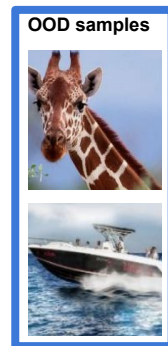
# Test-Time Adaptation



# Open-Set Test-Time Adaptation



+



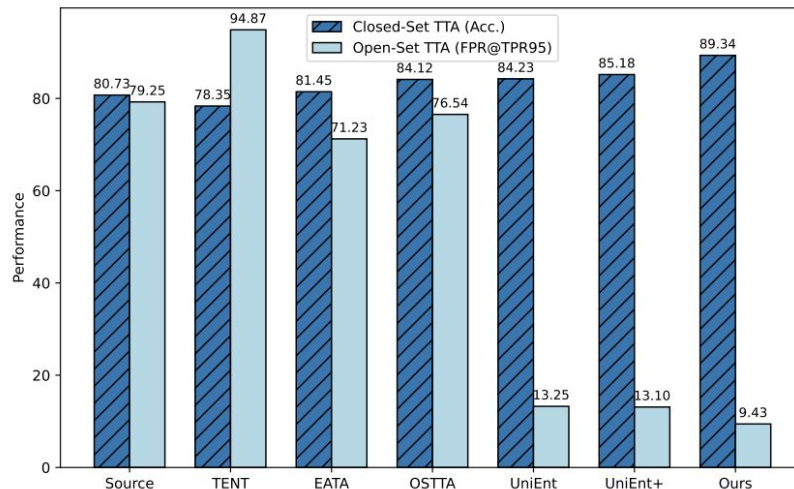
## Key Differences of TTA vs OSTTA

**Test-Time Adaptation (TTA)**: Adopts a pre-trained source model to unlabeled target domain during inference. Standard assumption: Closed-Set scenario where source and target share the same labels space.

**Open-Set Test-Time Adaptation (OSTTA)**: More realistic and challenging. Target domain contains unknown classes not seen during training. The goal is to adapt to known classes while identifying and rejecting/handling unknown samples (OOD).

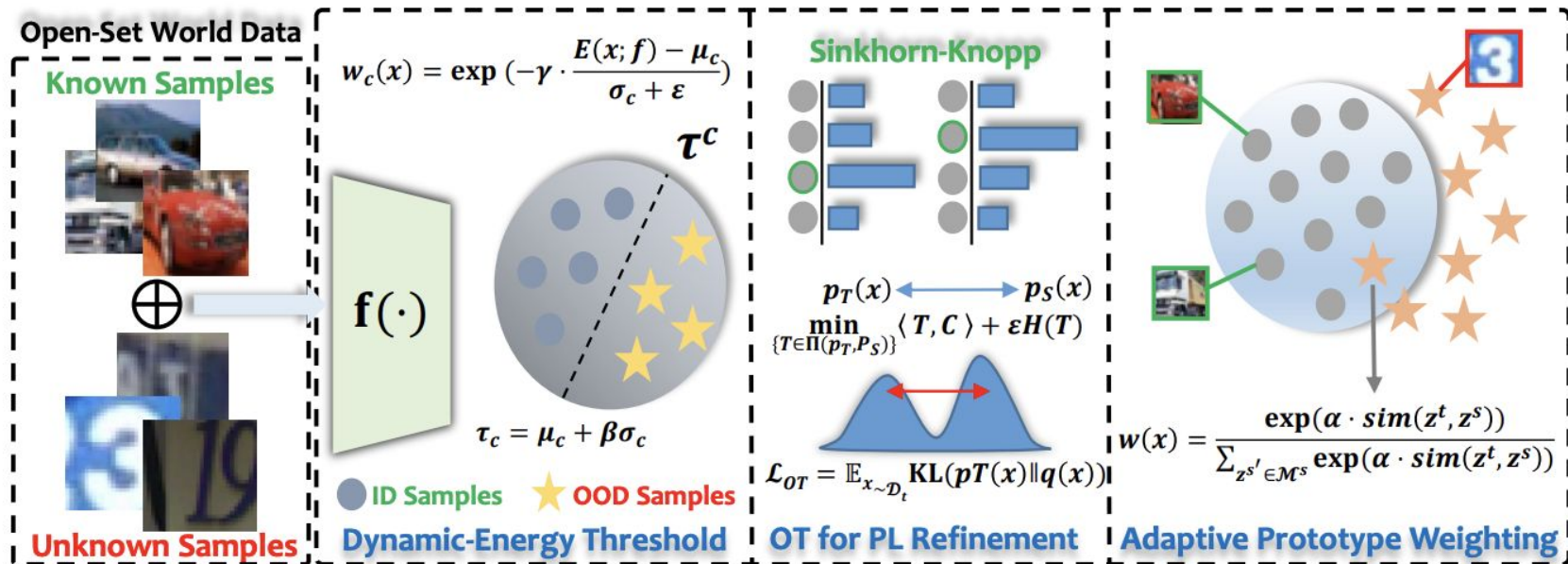
# Motivation

- **Standard TTA Limitations:** Existing TTA methods fail when unknown classes (OOD samples) appear in target domain.
- **Detect & Discard:** Current Open-Set TTA methods treat all unknown samples as harmful outliers.
- **Information Loss:** Discarding all unknowns ignores potentially useful semantic cues shared with known classes.



**Our Insight:** Not all unknowns are harmful. We can leverage semantically useful unknown samples to improve adaptation.

# Method



Overview of Learning from Unknown OSTTA (**LU-OSTTA**)

# Method: Dynamic-Energy Threshold

**Problem:** Global thresholds discard useful samples or accept harmful noise due to varying class distributions.

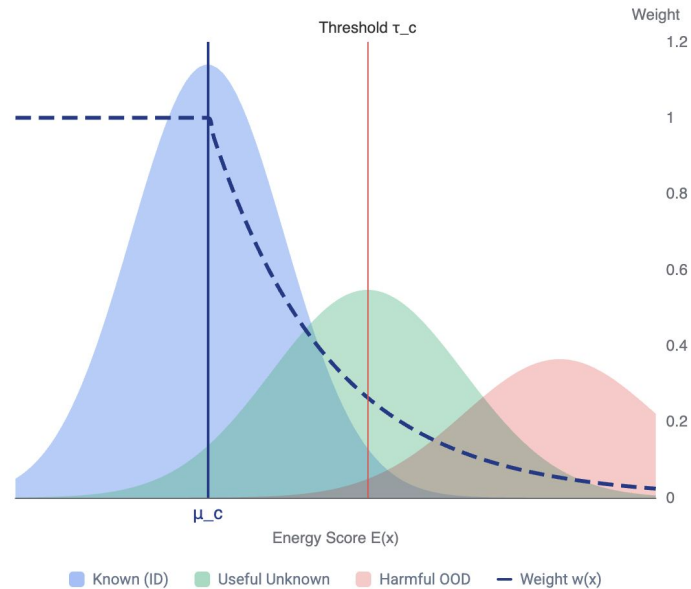
**Solution:** Compute energy statistics per-class to dynamically adjust the acceptance boundary.

**Class-Conditioned Threshold:** Naively removing OOD samples can hurt performance, especially under class imbalance. To address this, we propose a dynamic energy threshold that adapts per class.

$$\tau_c = \mu_c + \beta\sigma_c$$

**Soft-Weighting:** Ensures low energy samples such as ID-samples obtain higher weights while OOD-samples are down-weighted rather than fully rejected.

$$w_c(\mathbf{x}) = \exp\left(-\gamma \cdot \frac{E(\mathbf{x}; f) - \mu_c}{\sigma_c + \varepsilon}\right)$$



# Method: Optimal Transport for Pseudo-Label Refinement

**Problem:** Teacher models produce noisy, overconfident labels on OOD data.

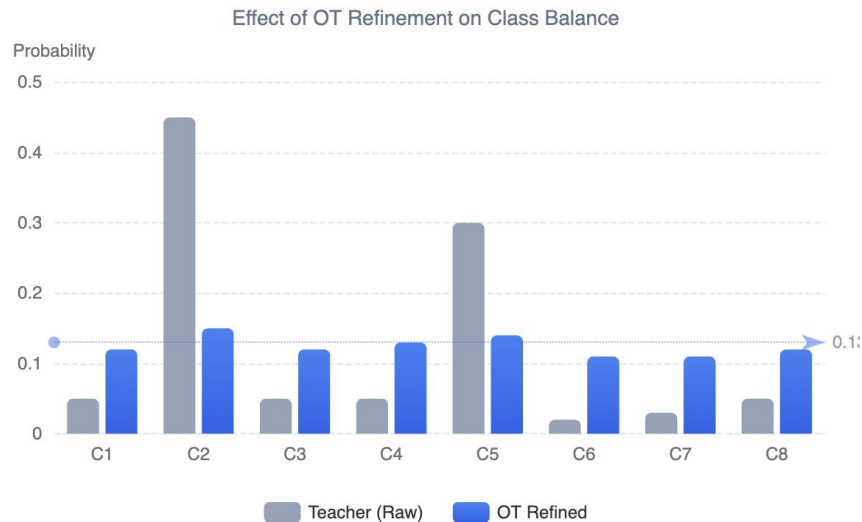
**Solution:** Use Optimal Transport (OT) to align teacher predictions with the global student distribution, enforcing class balance constraints.

**Optimal Transport Objective:** We leverage OT to align the teacher and student distributions instead of minimizing their discrepancy for each sample. Solved via Sinkhorn-Knopp Algorithm.

$$\min_{T \in \Pi(p_T, p_S)} \langle T, C \rangle + \varepsilon H(T)$$

**Refined Adaptation Loss:** Ensures the student model remains consistent with the teacher while having better and calibrated pseudo-labels.

$$\mathcal{L}_{OT} = \mathbb{E}_{\mathbf{x} \sim \mathcal{D}_t} \text{KL}(p_S(\mathbf{x}) \parallel q(\mathbf{x}))$$



# Method: Adaptive Prototype Weighting

**Goal:** Leverage semantically useful unknown samples while filtering out harmful noise.

**Method:** Maintain frozen source prototypes and updating target prototypes. Weight samples based on semantic similarity to source prototypes.

**Semantic Weight:** As target domain samples are unlabeled and far from the source distribution due to shift, it is hard to determine which target samples contain semantically useful information among OOD. To address this, we compute a semantic weight  $w(\mathbf{x})$  for each target OOD sample  $\mathbf{x}$  based on its similarity with source prototypes:

$$w(\mathbf{x}) = \frac{\exp(\alpha \cdot \text{sim}(z^t, z^s))}{\sum_{z^{s'} \in \mathcal{M}^s} \exp(\alpha \cdot \text{sim}(z^t, z^{s'}))}$$

**Adaptation Objective:** Contrastive loss aligns useful features, and entropy regularization prevents model collapse.

$$\mathcal{L}_{\text{con}} = -\frac{1}{|\mathcal{B}_t|} \sum_{\mathbf{x} \in \mathcal{B}_t} w(\mathbf{x}) \cdot \log \frac{\exp(\text{sim}(z^t, z^s)/\tau)}{\sum_{z^{s'} \in \mathcal{M}^s} \exp(\text{sim}(z^t, z^{s'})/\tau)} \quad \left\{ \begin{array}{l} \text{contrastive loss} \end{array} \right\}$$
$$\mathcal{L}_{\text{ent}} = -\frac{1}{|\mathcal{B}_t|} \sum_{\mathbf{x} \in \mathcal{B}_t} \sum_{c=1}^K p(c|\mathbf{x}) \log p(c|\mathbf{x}) \quad \left\{ \begin{array}{l} \text{entropy regularization} \end{array} \right\}$$

# Experiments

## Main Results: Evaluated on CIFAR-10/100-C benchmarks

Method	CIFAR-10-C				CIFAR-100-C			
	Acc.↑	AUROC↑	FPR@TPR95↓	OSCR↑	Acc.↑	AUROC↑	FPR@TPR95↓	OSCR↑
Source [85]	80.73	76.44	79.25	67.44	53.68	61.55	93.20	39.71
TENT [72]	78.35	64.79	94.87	55.92	55.71	66.23	93.89	41.34
CoTTA [74]	83.34	84.07	72.40	77.45	55.33	77.23	80.78	48.32
EATA [56]	81.45	83.76	71.23	72.56	60.78	<u>87.24</u>	94.88	42.28
OSTTA [38]	84.12	72.45	76.54	65.32	60.18	75.34	82.12	50.83
UniEnt [24]	84.23	95.45	13.25	83.45	59.45	92.21	<u>23.24</u>	57.31
UniEnt+ [24]	85.18	96.12	13.10	84.27	60.32	92.43	23.69	58.23
S-OSTTA [37]	<u>86.23</u>	<u>96.74</u>	<u>11.24</u>	<u>85.03</u>	<u>61.34</u>	<u>95.37</u>	23.75	<u>58.35</u>
<b>Ours</b>	<b>89.34</b> <sub>(+3.11)</sub>	<b>98.56</b> <sub>(+1.82)</sub>	<b>9.43</b> <sub>(-1.81)</sub>	<b>88.74</b> <sub>(+3.71)</sub>	<b>64.76</b> <sub>(+3.42)</sub>	<b>97.67</b> <sub>(+2.30)</sub>	<b>20.34</b> <sub>(-3.41)</sub>	<b>62.45</b> <sub>(+4.10)</sub>

# Experiments

Main Results: Evaluated on Tiny-ImageNet benchmark

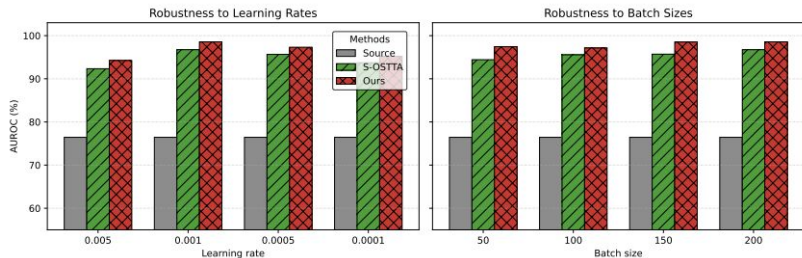
Method	Tiny-ImageNet-C			
	Acc $\uparrow$	AUROC $\uparrow$	FPR@TPR95 $\downarrow$	OSCR $\uparrow$
Source [85]	22.14	54.86	93.32	16.12
TENT [72]	28.56	48.10	95.32	19.41
CoTTA [74]	30.12	57.24	81.78	23.45
EATA [56]	32.92	58.31	83.66	25.64
OSTTA [38]	37.46	55.78	93.78	27.67
UniEnt [24]	37.23	56.02	91.25	27.98
UniEnt+ [24]	38.00	57.12	90.34	28.43
S-OSTTA [37]	<u>39.12</u>	<u>58.90</u>	<u>89.36</u>	<u>29.89</u>
<b>Ours</b>	<b>42.78</b> <sub>(+3.66)</sub>	<b>62.34</b> <sub>(+3.44)</sub>	<b>86.57</b> <sub>(-2.79)</sub>	<b>32.47</b> <sub>(+2.58)</sub>

# Experiments

## Ablation of Model Components

DET	OTPL	APW	Acc $\uparrow$	AUROC $\uparrow$	FPR@TPR95 $\downarrow$	OSCR $\uparrow$
$\times$	$\times$	$\times$	64.23	61.33	96.31	48.42
$\checkmark$	$\times$	$\times$	82.27	79.93	39.45	68.73
$\times$	$\checkmark$	$\times$	74.34	76.14	42.59	66.56
$\checkmark$	$\checkmark$	$\times$	83.45	95.61	14.56	83.11
$\checkmark$	$\checkmark$	$\checkmark$	<b>89.34</b>	<b>98.56</b>	<b>9.43</b>	<b>88.74</b>

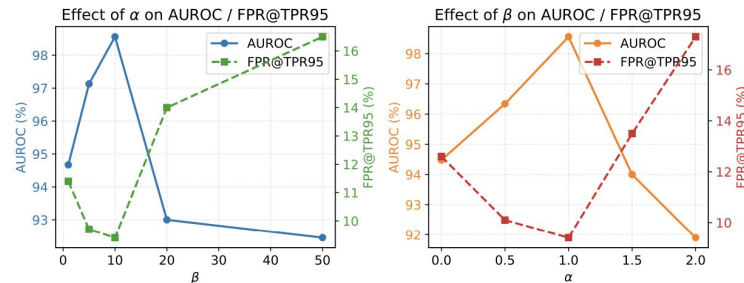
## Difference LR and Batch Sizes



## Different Confidence Thresholds

Method	CIFAR-10		
	AUROC $\uparrow$	FPR@TPR95 $\downarrow$	OSCR $\uparrow$
MSP [29]	48.08	93.34	42.90
Max Logit [27]	50.51	90.67	45.66
Energy [47]	51.25	90.89	47.76
<b>DET (Ours)</b>	<b>79.93</b>	<b>42.59</b>	<b>68.73</b>

## Different Hyperparameters



# Experiments

## Per-Batch Inference Time

Method	CIFAR-10-C						Ours
	TENT	EATA	CoTTA	OSTTA	UniEnt	S-OSTTA	
Inference Time (ms)↓	26.34	57.76	887.34	59.89	75.34	90.34	93.34
OSCR↑	55.92	72.56	77.45	65.32	83.45	85.03	<b>88.74</b>

## Difference No. of Unknown Class

Method	2	4	6	8	10
Source	72.45	69.67	69.33	69.17	68.05
TENT	49.44	49.23	50.56	48.25	50.56
UniEnt+	78.45	78.46	78.51	78.34	76.98
S-OSTTA	80.34	79.26	79.76	80.00	78.28
<b>Ours</b>	<b>84.12</b>	<b>83.44</b>	<b>83.33</b>	<b>83.67</b>	<b>83.00</b>

## Different Known and Unknown Ratios

Method	0.1	0.5	1.5	2.0
Source	40.00	40.30	38.97	38.18
TENT	47.76	44.27	44.89	41.79
UniEnt+	56.43	57.85	56.12	55.04
S-OSTTA	60.23	60.43	59.90	58.67
<b>Ours</b>	<b>63.56</b>	<b>63.30</b>	<b>62.92</b>	<b>62.76</b>

## Gradually Changing

Avg. Error (%)↓	Source	TENT [72]	CoTTA [74]	OSTTA [38]	UniEnt [24]	UniEnt+ [24]	S-OSTTA [37]	Ours
CIFAR-10-C	26.5	33.6	12.2	24.6	11.3	11.2	9.4	<b>8.1</b>

**Thank you!**