# Entropy is not Enough for Test-Time Adaptation: From the Perspective of Disentangled Factors

Jonghyun Lee<sup>\*</sup>, Dahuin Jung, Saehyung Lee, Junsung Park, Juhyeon Shin, Uiwon Hwang, Sungroh Yoon

Seoul National University







#### **Test-Time Adaptation**

- Requirements
  - The pre-trained model  $\mathcal{M}_s$  trained on the source (training) data  $\mathcal{D}_s$
  - Target (test) data  $\mathcal{D}_t = \{x_t\}$
- Goal
  - Using  $\mathcal{M}_s$  and the stream of  $\mathcal{D}_t$ , obtain the best performance on target domain
- Constraints
  - No source (train) data  $\mathcal{D}_s$
  - Efficiency (memory, runtime)

## Test-Time Adaptation (cont'd)

- TTA cannot access whole target data  $\mathcal{D}_t$  before adaptation.
  - Impossible to estimate the target distribution
    - Prone to inaccurate prediction in the early stage
    - Error accumulation!
  - Need to adapt using samples that have a lower likelihood of causing error
    - Sample filtering and reweighting with confidence metric!

#### Previous work

- TENT (ICLR 2021 spotlight)
  - Entropy minimization loss
  - No filtering
- EATA (ICML 2022)
  - TENT + filtering + regularization
    - Reliable (entropy) filtering + redundant filtering
    - Fisher regularization
- SAR (ICLR 2023 oral)
  - Reliable (entropy) filtering + sharpness-aware entropy minimization loss

## Previous work (cont'd)

• Observations



- The lowest entropy interval shows the lowest accuracy
  - Unreliable!
- Samples in o~Q1
  - Correct samples: focus on bird (target) well
  - Wrong samples: relatively bad

## Entropy is not enough for TTA!

- Entropy cannot reflect 'which part' of the image assigns low entropy.
  - If birds (object)? → good!
  - If spurious features (background)? → bad...
- Let's take 'disentangled factors' notation!

- Disentangled factors
  - Disentangled latent vector  $\mathbf{v}(\mathbf{x}) = (v_0(\mathbf{x}), v_1(\mathbf{x}), \dots, v_{d_v}(\mathbf{x})) \in \mathbb{R}^{d_v}$ 
    - Each element is independent
    - **v**(**x**) can perfectly reconstruct an input image **x**
    - $v_i(\mathbf{x}) \in [0,1]$ : the *i*-th factor of  $\mathbf{x}$
  - In binary classification ( $y \in \{-1, +1\}$ )
    - Each factor  $v_i(\mathbf{x})$  has a correlation with a true label y
    - Under distribution shift, the correlation can be changed!
    - Define two correlations
      - corr<sup>train</sup><sub>*i*</sub> = corr(y<sup>train</sup>, v<sup>train</sup><sub>*i*</sub>), corr<sup>test</sup><sub>*i*</sub> = corr(y<sup>test</sup>, v<sup>test</sup><sub>*i*</sub>)
    - Then we could divide  $\mathbf{v}(\mathbf{x})$  into four partitions

$$\mathbf{v}_{pp} = \{\mathbf{v}_i | \operatorname{corr}_i^{\operatorname{train}} > 0, \operatorname{corr}_i^{\operatorname{test}} > 0\}, \quad \mathbf{v}_{pn} = \{\mathbf{v}_i | \operatorname{corr}_i^{\operatorname{train}} > 0, \operatorname{corr}_i^{\operatorname{test}} \le 0\}, \\ \mathbf{v}_{np} = \{\mathbf{v}_i | \operatorname{corr}_i^{\operatorname{train}} \le 0, \operatorname{corr}_i^{\operatorname{test}} > 0\}, \quad \mathbf{v}_{nn} = \{\mathbf{v}_i | \operatorname{corr}_i^{\operatorname{train}} \le 0, \operatorname{corr}_i^{\operatorname{test}} \le 0\}$$

• Assume the pretrained model  $\mathcal{M}_{\theta}$  as a linear classifier.

**Proposition 1.** Let us consider a pre-trained linear classifier  $\mathcal{M}_{\theta}$  that uses the latent disentangled factors  $\mathbf{v}(\mathbf{x})$  of sample  $\mathbf{x}$  as input. We define a **harmful** sample as one that reduces the difference in the mean logits between classes when used for adaptation. A sample  $\mathbf{x} \in \mathcal{X}^{\text{test}}$  is a **harmful** sample for adaptation using entropy minimization loss if it satisfies the following condition:

$$\hat{\mathbf{y}}\mathbf{v}(\mathbf{x}) \cdot (\mathbb{E}_{\mathbf{x}^{\text{test}} \sim \mathcal{X}_{+1}^{\text{test}}}[\mathbf{v}(\mathbf{x}^{\text{test}})] - \mathbb{E}_{\mathbf{x}^{\text{test}} \sim \mathcal{X}_{-1}^{\text{test}}}[\mathbf{v}(\mathbf{x}^{\text{test}})]) < 0,$$

$$where \ \mathcal{X}_{y}^{\text{test}} = \{\mathbf{x} | (\mathbf{x}, \mathbf{y}) \in \mathcal{D}^{\text{test}}, \mathbf{y} = y\}, \ and \ y \in \{1, -1\}.$$

$$(5)$$

• With Proposition 1, we can explain why the samples with low entropy can be *harmful*.

• The partitions of optimal parameters  $\boldsymbol{\theta}^*$  for the training data

$$\mathbf{v}_{pp} = \{\mathbf{v}_i | \operatorname{corr}_i^{\operatorname{train}} > 0, \operatorname{corr}_i^{\operatorname{test}} > 0\}, \quad \mathbf{v}_{pn} = \{\mathbf{v}_i | \operatorname{corr}_i^{\operatorname{train}} > 0, \operatorname{corr}_i^{\operatorname{test}} \le 0\}, \quad \boldsymbol{\theta}_{pp}^*, \boldsymbol{\theta}_{pn}^* > 0$$
$$\mathbf{v}_{np} = \{\mathbf{v}_i | \operatorname{corr}_i^{\operatorname{train}} \le 0, \operatorname{corr}_i^{\operatorname{test}} > 0\}, \quad \mathbf{v}_{nn} = \{\mathbf{v}_i | \operatorname{corr}_i^{\operatorname{train}} \le 0, \operatorname{corr}_i^{\operatorname{test}} \le 0\}. \quad \boldsymbol{\theta}_{np}^*, \boldsymbol{\theta}_{nn}^* \le 0$$

• In the early stages of adaptation, **x** with a high-confidence pseudo-label of  $\hat{y} = +1$  satisfies

$$\mathbf{a}_{\boldsymbol{\theta}}(\mathbf{x}) = \boldsymbol{\theta}_{pp} \cdot \mathbf{v}_{pp} + \boldsymbol{\theta}_{pn} \cdot \mathbf{v}_{pn} + \boldsymbol{\theta}_{np} \cdot \mathbf{v}_{np} + \boldsymbol{\theta}_{nn} \cdot \mathbf{v}_{nn} \gg 0,$$
$$|\boldsymbol{\theta}_{pp} \cdot \mathbf{v}_{pp} + \boldsymbol{\theta}_{pn} \cdot \mathbf{v}_{pn}| \gg |\boldsymbol{\theta}_{np} \cdot \mathbf{v}_{np} + \boldsymbol{\theta}_{nn} \cdot \mathbf{v}_{nn}|.$$

- Two dominant factors
  - $\mathbf{v}_{pp}$ : Commonly Positively-coRrelated with label (CPR) factors
  - **v**<sub>pn</sub>: TRAin-time only Positively-correlated with label (TRAP) factors

• By definition, the expectations of CPR factors and TRAP factors are as follows:

$$\mathbb{E}_{\mathbf{x}^{\text{test}} \sim \mathcal{X}_{+1}^{\text{test}}}[\mathbf{v}_{pp}(\mathbf{x}^{\text{test}})] > \mathbb{E}_{\mathbf{x}^{\text{test}} \sim \mathcal{X}_{-1}^{\text{test}}}[\mathbf{v}_{pp}(\mathbf{x}^{\text{test}})],$$
$$\mathbb{E}_{\mathbf{x}^{\text{test}} \sim \mathcal{X}_{+1}^{\text{test}}}[\mathbf{v}_{pn}(\mathbf{x}^{\text{test}})] \leq \mathbb{E}_{\mathbf{x}^{\text{test}} \sim \mathcal{X}_{-1}^{\text{test}}}[\mathbf{v}_{pn}(\mathbf{x}^{\text{test}})].$$

• Reformulation of Eq. (5) of Proposition 1.

$$\begin{split} \hat{\mathbf{y}} \mathbf{v}(\mathbf{x}) \cdot (\mathbb{E}_{\mathbf{x}^{\text{test}} \sim \mathcal{X}_{+1}^{\text{test}}} [\mathbf{v}(\mathbf{x}^{\text{test}})] - \mathbb{E}_{\mathbf{x}^{\text{test}} \sim \mathcal{X}_{-1}^{\text{test}}} [\mathbf{v}(\mathbf{x}^{\text{test}})]) \\ \approx \underbrace{\mathbf{v}_{pp}(\mathbf{x}) \cdot (\mathbb{E}_{\mathbf{x}^{\text{test}} \sim \mathcal{X}_{+1}^{\text{test}}} [\mathbf{v}_{pp}(\mathbf{x}^{\text{test}})] - \mathbb{E}_{\mathbf{x}^{\text{test}} \sim \mathcal{X}_{-1}^{\text{test}}} [\mathbf{v}_{pp}(\mathbf{x}^{\text{test}})]))}_{(7.a)} \\ + \underbrace{\mathbf{v}_{pn}(\mathbf{x}) \cdot (\mathbb{E}_{\mathbf{x}^{\text{test}} \sim \mathcal{X}_{+1}^{\text{test}}} [\mathbf{v}_{pn}(\mathbf{x}^{\text{test}})] - \mathbb{E}_{\mathbf{x}^{\text{test}} \sim \mathcal{X}_{-1}^{\text{test}}} [\mathbf{v}_{pn}(\mathbf{x}^{\text{test}})]))}_{(7.b)} < 0 \end{split}$$

- Eq. (7.a) (related to CPR factors) becomes positive, and Eq. (7.b) (related to TRAP factors) becomes negative
  - $|(7.a)| \ll |(7.b)| \rightarrow a harmful sample$ 
    - When TRAP factors affect more than CPR factors, the sample becomes *harmful*

- If  $|(7.a)| \ll |(7.b)|$  and  $|\boldsymbol{\theta}_{pp} \cdot \mathbf{v}_{pp} + \boldsymbol{\theta}_{pn} \cdot \mathbf{v}_{pn}| \gg |\boldsymbol{\theta}_{np} \cdot \mathbf{v}_{np} + \boldsymbol{\theta}_{nn} \cdot \mathbf{v}_{nn}|$   $\rightarrow \mathbf{x}$  becomes a *harmful* sample with low entropy (high confidence)
  - Low entropy filtering cannot discern good & bad samples
- Then how to adapt?
  - Utilize the CPR factors and avoid the TRAP factors!



All sources are created by DALL-E

# Methodology

- Destroy Your Object (DeYO)
  - Utilize the factor that aligns with g.t. class (CPR factor:  $\mathbf{v}_{pp}$ ) under **any** test distribution
    - Classification task: the shape of object!
  - We can simply apply patch shuffling to destroy the shape information, preserving the patch-level local features.
  - If a prediction becomes uncertain when the shape of object is destroyed,
     → The model considers the shape of the object as the dominant factor when classifying the sample.
  - Pseudo-Label Probability Difference (PLPD)
    - Measures the extent to which the probability of pseudo-label decreases after applying the patch shuffling
  - Utilize samples with low entropy & high PLPD!

 $S_{\theta}(\mathbf{x}) = \{\mathbf{x} | \text{Ent}_{\theta}(\mathbf{x}) < \tau_{\text{Ent}}, \text{ PLPD}_{\theta}(\mathbf{x}, \mathbf{x}') > \tau_{\text{PLPD}} \}, \text{ where}$  $PLPD_{\theta}(\mathbf{x}, \mathbf{x}') = (\mathbf{p}_{\theta}(\mathbf{x}) - \mathbf{p}_{\theta}(\mathbf{x}'))_{\hat{y}},$ 

## Methodology (cont'd)

• Overall procedure of DeYO



$$S_{\theta}(\mathbf{x}) = \{\mathbf{x} | \text{Ent}_{\theta}(\mathbf{x}) < \tau_{\text{Ent}}, \text{PLPD}_{\theta}(\mathbf{x}, \mathbf{x}') > \tau_{\text{PLPD}} \}, \text{ where}$$
$$PLPD_{\theta}(\mathbf{x}, \mathbf{x}') = (\mathbf{p}_{\theta}(\mathbf{x}) - \mathbf{p}_{\theta}(\mathbf{x}'))_{\hat{y}},$$
$$\alpha_{\theta}(\mathbf{x}) = \frac{1}{\exp\{(\text{Ent}_{\theta}(\mathbf{x}) - \text{Ent}_{0})\}} + \frac{1}{\exp\{-\text{PLPD}_{\theta}(\mathbf{x}, \mathbf{x}')\}},$$

 $\mathcal{L}_{\text{DeYO}}(\mathbf{x}; \boldsymbol{\theta}) = \alpha_{\boldsymbol{\theta}}(\mathbf{x}) \cdot \mathbb{I}_{\{\mathbf{x} \in S_{\boldsymbol{\theta}}(\mathbf{x})\}} \text{Ent}_{\boldsymbol{\theta}}(\mathbf{x}),$ 

#### Experiments

• TTA on an i.i.d. sampling (mild) scenario (ResNet-50-BN)

Table 1: Comparisons with baselines on ImageNet-C at severity level 5 under a mild scenario regarding accuracy (%). The **bold** value signifies the top-performing result.

	Noise Blur						Wea	ther		Digital						
Mild	Gauss.	Shot	Impul.	Defoc.	Glass	Motion	Zoom	Snow	Frost	Fog	Brit.	Contr.	Elastic	Pixel	JPEG	Avg.
ResNet-50-BN	2.2	2.9	1.8	17.9	9.8	14.8	22.5	16.9	23.3	24.4	58.9	5.4	16.9	20.7	31.7	18.0
<ul> <li>MEMO</li> </ul>	7.5	8.8	8.9	19.8	13.0	20.7	27.7	25.3	28.7	32.2	61.0	11.0	23.8	33.0	37.6	23.9
<ul> <li>Tent</li> </ul>	29.2	31.2	30.1	28.1	27.7	41.4	49.4	47.2	41.5	57.7	67.4	29.2	54.8	58.5	52.4	43.1
• EATA	34.9	37.1	35.8	33.4	33.0	47.1	52.7	51.6	45.7	60.0	68.1	44.4	57.9	60.6	55.1	47.8
• SAR	30.6	30.6	31.3	28.5	28.5	41.9	49.4	47.1	42.2	57.5	67.3	37.8	54.6	58.4	52.1	43.9
<ul> <li>DeYO (ours)</li> </ul>	35.6±0.2	$37.9_{\pm 0.1}$	$37.1_{\pm 0.1}$	$33.8_{\pm 0.2}$	$34.1_{\pm 0.2}$	$48.5_{\pm 0.1}$	$52.8_{\pm 0.1}$	$52.7_{\pm 0.0}$	$46.4_{\pm 0.1}$	<b>60.6</b> ±0.0	$68.0_{\pm0.1}$	$46.1_{\pm 0.1}$	$58.4_{\pm 0.1}$	$61.5_{\pm 0.1}$	$55.7_{\pm 0.1}$	$\textbf{48.6}_{\pm 0.0}$

• TTA on a spurious correlations shift (biased) scenario (ResNet-18/50-BN)

Table 2:	Comparisons	with	baselines	on	Col-
oredMNI	ST regarding a	accura	acy (%).		

Biased	Avg Acc	Worst-Group Acc
ResNet-18-BN	63.40	20.05
• Tent	57.06	9.80
• MEMO	63.77	6.23
<ul> <li>SENTRY</li> </ul>	63.23	15.78
• EATA	60.81	17.98
• SAR	58.37	12.36
<ul> <li>DeYO (ours)</li> </ul>	78.24	67.39

Table 3: Comparisons with baselines on Water-Birds regarding accuracy (%).

Biased	Avg Acc	Worst-Group Acc
ResNet-50-BN	83.16	64.90
• Tent	82.95	54.14
<ul> <li>MEMO</li> </ul>	82.34	50.47
<ul> <li>SENTRY</li> </ul>	85.77	60.90
• EATA	82.38	52.38
• SAR	82.60	53.41
• DeYO (ours)	87.42	73.92

#### Experiments (cont'd)

- TTA on wild scenarios (ResNet-50-GN, ViT-16/B)
  - Temporally-correlated label shifts
  - Batch size 1

Table 5: Comparisons with baselines on ImageNet-C at severity level 5 under online imbalanced label shifts (imbalance ratio =  $\infty$ ) or under batch size 1 regarding accuracy (%).

		Noise			В	lur			We	eather			Dig	gital		
Label Shifts	Gauss.	Shot	Impul.	Defoc.	Glass	Motion	Zoom	Snow	Frost	Fog	Brit.	Contr.	Elastic	Pixel	JPEG	Avg.
ResNet-50-GN	17.9	19.9	17.9	19.7	11.3	21.3	24.9	40.4	47.4	33.6	69.3	36.3	18.7	28.4	52.2	30.6
<ul> <li>MEMO</li> </ul>	18.4	20.6	18.4	17.1	12.7	21.8	26.9	40.7	46.9	34.8	69.6	36.4	19.2	32.2	53.4	31.3
<ul> <li>Tent</li> </ul>	3.6	4.2	4.4	16.5	5.9	26.9	28.4	17.9	26.2	2.3	72.2	46.1	7.3	52.3	56.2	24.7
<ul> <li>EATA</li> </ul>	25.7	28.6	24.8	18.5	19.6	24.1	28.4	35.3	33.0	41.2	65.2	33.3	28.0	42.4	43.1	32.7
• SAR	33.7	36.9	35.3	19.3	20.3	33.8	29.8	21.9	44.7	34.9	71.9	46.7	6.6	52.3	56.2	36.3
• DeYO (ours)	$42.5_{\pm 0.5}$	<b>44.9</b> $_{\pm 0.2}$	$43.8_{\pm 0.3}$	$22.2_{\pm 0.0}$	$16.3_{\pm 10.2}$	$41.0_{\pm 0.2}$	$13.2_{\pm 9.8}$	52.2 <sub>±0.4</sub>	<b>51.5</b> ±0.5	$39.7_{\pm 27.4}$	$73.4_{\pm 0.1}$	52.6 <sub>±0.2</sub>	$46.9_{\pm 1.2}$	<b>59.3</b> $_{\pm 0.1}$	<b>59.3</b> $_{\pm 0.0}$	$43.9_{\pm 2.0}$
VitBase-LN	9.4	6.7	8.3	29.1	23.4	34.0	27.1	15.8	26.4	47.4	54.7	44.0	30.5	44.5	47.6	29.9
<ul> <li>MEMO</li> </ul>	21.6	17.4	20.6	37.1	29.6	40.6	34.4	25.0	34.8	55.2	65.0	54.9	37.4	55.5	57.7	39.1
<ul> <li>Tent</li> </ul>	33.9	1.8	27.2	54.8	52.9	58.6	54.3	12.4	11.7	69.7	76.3	66.3	59.6	69.7	66.6	47.7
<ul> <li>EATA</li> </ul>	36.2	34.7	35.5	43.4	44.3	49.3	48.5	53.2	53.5	62.3	72.7	18.8	58.0	64.7	62.8	49.2
• SAR	42.3	34.9	44.1	50.0	50.5	55.6	53.1	59.7	47.2	66.2	75.2	50.3	60.1	67.3	65.0	54.8
<ul> <li>DeYO (ours)</li> </ul>	$53.5_{\pm 0.5}$	$36.0_{\pm 25.2}$	$54.6_{\pm 0.8}$	57.6 <sub>±0.2</sub>	$58.7_{\pm 0.2}$	$63.7_{\pm 0.1}$	$46.2_{\pm 18.7}$	67.6±0.1	<b>66.0</b> ±0.1	$73.2_{\pm 0.2}$	$77.9_{\pm 0.1}$	66.7 <sub>±0.1</sub>	<b>69.0</b> ±0.1	$73.5_{\pm 0.1}$	$70.3_{\pm 0.2}$	$62.3_{\pm 1.7}$
Noise Blur																
		Noise			BI	ur			We	ather			Die	rital		
Batch Size 1	Gauss.	Noise Shot	Impul.	Defoc.	Bl Glass	ur Motion	Zoom	Snow	Wea Frost	ather Fog	Brit.	Contr.	Dig Elastic	ital Pixel	JPEG	Avg.
Batch Size 1 ResNet-50-GN	Gauss.	Noise Shot 19.8	Impul. 17.9	Defoc.	Bl Glass 11.4	ur Motion 21.4	Zoom 24.9	Snow 40.4	Wea Frost 47.3	ather Fog 33.6	Brit. 69.3	Contr. 36.3	Dig Elastic 18.6	tital Pixel 28.4	JPEG 52.3	Avg.
Batch Size 1 ResNet-50-GN • MEMO	Gauss.	Noise Shot 19.8 20.5	Impul. 17.9 18.4	Defoc. 19.8 17.1	Bl Glass 11.4 12.6	ur Motion 21.4 21.8	Zoom 24.9 26.9	Snow 40.4 40.4	Wes Frost 47.3 47.0	ather Fog 33.6 34.4	Brit. 69.3 69.5	Contr. 36.3 36.5	Dig Elastic 18.6 19.2	gital Pixel 28.4 32.1	JPEG 52.3 53.3	Avg. 30.6 31.2
Batch Size 1 ResNet-50-GN • MEMO • Tent	Gauss.	Noise Shot 19.8 20.5 4.2	Impul. 17.9 18.4 4.0	Defoc. 19.8 17.1 16.5	Bl Glass 11.4 12.6 5.3	ur Motion 21.4 21.8 27.4	Zoom 24.9 26.9 <b>30.3</b>	Snow 40.4 40.4 17.7	Wea Frost 47.3 47.0 24.9	ather Fog 33.6 34.4 2.0	Brit. 69.3 69.5 72.1	Contr. 36.3 36.5 46.2	Dig Elastic 18.6 19.2 7.8	rital Pixel 28.4 32.1 52.6	JPEG 52.3 53.3 56.3	Avg. 30.6 31.2 24.7
Batch Size 1 ResNet-50-GN • MEMO • Tent • EATA	Gauss.	Noise Shot 19.8 20.5 4.2 27.9	Impul. 17.9 18.4 4.0 25.8	Defoc. 19.8 17.1 16.5 17.9	Bl Glass 11.4 12.6 5.3 17.3	ur Motion 21.4 21.8 27.4 28.7	Zoom 24.9 26.9 <b>30.3</b> 29.3	Snow 40.4 40.4 17.7 44.7	Wes Frost 47.3 47.0 24.9 44.4	ather Fog 33.6 34.4 2.0 <b>40.2</b>	Brit. 69.3 69.5 72.1 71.0	Contr. 36.3 36.5 46.2 44.5	Dig Elastic 18.6 19.2 7.8 27.0	rital Pixel 28.4 32.1 52.6 46.8	JPEG 52.3 53.3 56.3 55.6	Avg. 30.6 31.2 24.7 36.4
Batch Size 1 ResNet-50-GN • MEMO • Tent • EATA • SAR	Gauss. 18.0 18.5 3.1 24.8 23.3	Noise Shot 19.8 20.5 4.2 27.9 26.6	Impul. 17.9 18.4 4.0 25.8 23.9	Defoc. 19.8 17.1 16.5 17.9 18.5	Bl Glass 11.4 12.6 5.3 17.3 15.2	ur <u>Motion</u> 21.4 21.8 27.4 28.7 28.6	Zoom 24.9 26.9 <b>30.3</b> 29.3 <b>30.3</b>	Snow 40.4 40.4 17.7 44.7 44.0	We: Frost 47.3 47.0 24.9 44.4 44.7	ather Fog 33.6 34.4 2.0 <b>40.2</b> 29.0	Brit. 69.3 69.5 72.1 71.0 72.3	Contr. 36.3 36.5 46.2 44.5 44.6	Dig Elastic 18.6 19.2 7.8 27.0 13.1	rital Pixel 28.4 32.1 52.6 46.8 46.8	JPEG 52.3 53.3 56.3 55.6 56.1	Avg. 30.6 31.2 24.7 36.4 34.5
Batch Size 1 ResNet-50-GN • MEMO • Tent • EATA • SAR • DeYO (ours)	Gauss. 18.0 18.5 3.1 24.8 23.3 <b>41.8</b> ±0.7	Noise Shot 19.8 20.5 4.2 27.9 26.6 <b>44.7</b> ±0.4	Impul. 17.9 18.4 4.0 25.8 23.9 <b>43.0</b> ±0.7	Defoc. 19.8 17.1 16.5 17.9 18.5 <b>22.5</b> ±0.1	Bl Glass 11.4 12.6 5.3 17.3 15.2 <b>24.7</b> ±0.3	ur Motion 21.4 21.8 27.4 28.7 28.6 <b>41.8</b> ±0.1	Zoom 24.9 26.9 <b>30.3</b> 29.3 <b>30.3</b> 24.4 <sub>±9.8</sub>	Snow 40.4 17.7 44.7 44.0 <b>54.5</b> ±0.2	We: Frost 47.3 47.0 24.9 44.4 44.7 <b>52.2</b> ±0.2	ather Fog 33.6 34.4 2.0 40.2 29.0 $20.7_{\pm 26.8}$	Brit. 69.3 69.5 72.1 71.0 72.3 <b>73.5</b> ±0.0	Contr. 36.3 36.5 46.2 44.5 44.6 <b>53.5</b> ±0.2	Dig Elastic 18.6 19.2 7.8 27.0 13.1 <b>48.5</b> ±0.3	tital Pixel 28.4 32.1 52.6 46.8 46.8 <b>60.2</b> ±0.0	JPEG 52.3 53.3 56.3 55.6 56.1 <b>59.8</b> ±0.1	Avg. 30.6 31.2 24.7 36.4 34.5 <b>44.4</b> ±1.2
Batch Size 1 ResNet-50-GN • MEMO • Tent • EATA • SAR • DeYO (ours) VitBase-LN	Gauss. 18.0 18.5 3.1 24.8 23.3 <b>41.8</b> ±0.7 9.5	Noise Shot 19.8 20.5 4.2 27.9 26.6 <b>44.7</b> ±0.4 6.8	Impul. 17.9 18.4 4.0 25.8 23.9 <b>43.0</b> ±0.7 8.2	Defoc. 19.8 17.1 16.5 17.9 18.5 <b>22.5</b> ±0.1 29.0	Bl Glass 11.4 12.6 5.3 17.3 15.2 <b>24.7</b> ±0.3 23.5	ur Motion 21.4 21.8 27.4 28.7 28.6 <b>41.8</b> ±0.1 33.9	Zoom 24.9 26.9 <b>30.3</b> 29.3 <b>30.3</b> 24.4 <sub>±9.8</sub> 27.1	Snow 40.4 40.4 17.7 44.7 44.0 <b>54.5</b> ±0.2 15.9	We: Frost 47.3 47.0 24.9 44.4 44.7 <b>52.2</b> ±0.2 26.5	$ \begin{array}{r} \text{Ather} \\ \hline Fog \\ 33.6 \\ 34.4 \\ 2.0 \\ 40.2 \\ 29.0 \\ 20.7_{\pm 26.8} \\ 47.2 \end{array} $	Brit. 69.3 69.5 72.1 71.0 72.3 <b>73.5</b> ±0.0 54.7	Contr. $36.3$ $36.5$ $46.2$ $44.5$ $44.6$ $53.5_{\pm 0.2}$ $44.1$	Dig Elastic 18.6 19.2 7.8 27.0 13.1 <b>48.5</b> ±0.3 30.5	tital Pixel 28.4 32.1 52.6 46.8 46.8 <b>60.2</b> ±0.0 44.5	JPEG 52.3 53.3 56.3 55.6 56.1 <b>59.8</b> ±0.1 47.8	Avg. 30.6 31.2 24.7 36.4 34.5 <b>44.4</b> ±1.2 29.9
Batch Size 1 ResNet-50-GN • MEMO • Tent • EATA • SAR • DeYO (ours) VitBase-LN • MEMO	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	Noise Shot 19.8 20.5 4.2 27.9 26.6 $44.7_{\pm 0.4}$ 6.8 17.3	Impul. 17.9 18.4 4.0 25.8 23.9 <b>43.0</b> ±0.7 8.2 20.6	Defoc. 19.8 17.1 16.5 17.9 18.5 <b>22.5</b> ±0.1 29.0 37.1	Bl Glass 11.4 12.6 5.3 17.3 15.2 <b>24.7</b> ±0.3 23.5 29.6	ur <u>Motion</u> 21.4 21.8 27.4 28.7 28.6 <b>41.8</b> ±0.1 33.9 40.4	Zoom 24.9 26.9 <b>30.3</b> 29.3 <b>30.3</b> 24.4 <sub>±9.8</sub> 27.1 34.4	Snow 40.4 40.4 17.7 44.7 44.0 <b>54.5</b> ±0.2 15.9 24.9	We: Frost 47.3 47.0 24.9 44.4 44.7 <b>52.2</b> ±0.2 26.5 34.7	$\begin{array}{r} \text{ather} \\ \hline Fog \\ 33.6 \\ 34.4 \\ 2.0 \\ 40.2 \\ 29.0 \\ 20.7_{\pm 26.8} \\ 47.2 \\ 55.1 \\ \end{array}$	Brit. 69.3 69.5 72.1 71.0 72.3 <b>73.5</b> ±0.0 54.7 64.8	Contr.           36.3           36.5           46.2           44.5           44.6 $53.5 \pm 0.2$ 44.1           54.9	Dig Elastic 18.6 19.2 7.8 27.0 13.1 $48.5_{\pm 0.3}$ 30.5 37.4	tital Pixel 28.4 32.1 52.6 46.8 46.8 <b>60.2</b> ±0.0 44.5 55.4	JPEG 52.3 53.3 55.6 56.1 <b>59.8</b> ±0.1 47.8 57.6	Avg. 30.6 31.2 24.7 36.4 34.5 <b>44.4</b> ±1.2 29.9 39.1
Batch Size 1 ResNet-50-GN • MEMO • Tent • EATA • SAR • DeYO (ours) VitBase-LN • MEMO • Tent	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	$\begin{array}{r} \text{Noise} \\ \text{Shot} \\ 19.8 \\ 20.5 \\ 4.2 \\ 27.9 \\ 26.6 \\ \textbf{44.7}_{\pm 0.4} \\ \textbf{6.8} \\ 17.3 \\ 1.6 \end{array}$	Impul. 17.9 18.4 4.0 25.8 23.9 <b>43.0</b> ±0.7 8.2 20.6 43.9	Defoc. 19.8 17.1 16.5 17.9 18.5 <b>22.5</b> $\pm$ 0.1 29.0 37.1 52.8	Bl Glass 11.4 12.6 5.3 17.3 15.2 <b>24.7</b> ±0.3 23.5 29.6 48.8	ur Motion 21.4 21.8 27.4 28.7 28.6 <b>41.8</b> ±0.1 33.9 40.4 55.9	Zoom 24.9 26.9 <b>30.3</b> 29.3 <b>30.3</b> 24.4 <sub>±9.8</sub> 27.1 34.4 51.3	Snow 40.4 40.4 17.7 44.7 44.0 <b>54.5</b> $\pm 0.2$ 15.9 24.9 22.9	We Frost 47.3 47.0 24.9 44.4 44.7 <b>52.2</b> $_{\pm 0.2}$ 26.5 34.7 21.1	$\begin{array}{r} \text{ather} \\ \hline \text{Fog} \\ \hline 33.6 \\ 34.4 \\ 2.0 \\ \textbf{40.2} \\ 29.0 \\ \hline 20.7_{\pm 26.8} \\ 47.2 \\ 55.1 \\ 66.9 \\ \end{array}$	Brit. 69.3 69.5 72.1 71.0 72.3 <b>73.5</b> ±0.0 54.7 64.8 75.1	Contr.           36.3           36.5           46.2           44.5           44.6 $53.5 \pm 0.2$ 44.1           54.9           65.0	$\begin{array}{c} \text{Dig}\\ \hline \text{Elastic}\\ 18.6\\ 19.2\\ 7.8\\ 27.0\\ 13.1\\ \textbf{48.5}_{\pm 0.3}\\ 30.5\\ 37.4\\ 54.0\\ \end{array}$	tital Pixel 28.4 32.1 52.6 46.8 46.8 <b>60.2</b> ±0.0 44.5 55.4 67.0	JPEG 52.3 53.3 55.6 56.1 <b>59.8</b> ±0.1 47.8 57.6 64.3	Avg. 30.6 31.2 24.7 36.4 34.5 <b>44.4</b> ±1.2 29.9 39.1 48.9
Batch Size 1 ResNet-50-GN • MEMO • Tent • EATA • SAR • DeYO (ours) VitBase-LN • MEMO • Tent • EATA	$\begin{tabular}{ c c c c c c c c c c c c c c c c c c c$	$\begin{array}{r} \text{Noise} \\ \text{Shot} \\ 19.8 \\ 20.5 \\ 4.2 \\ 27.9 \\ 26.6 \\ \textbf{44.7}_{\pm 0.4} \\ \textbf{6.8} \\ 17.3 \\ 1.6 \\ 26.7 \end{array}$	Impul. 17.9 18.4 4.0 25.8 23.9 <b>43.0</b> ±0.7 8.2 20.6 43.9 30.3	$\begin{array}{c} \text{Defoc.} \\ 19.8 \\ 17.1 \\ 16.5 \\ 17.9 \\ 18.5 \\ \textbf{22.5}_{\pm 0.1} \\ 29.0 \\ 37.1 \\ 52.8 \\ 43.8 \\ \end{array}$	$\begin{array}{c} \text{Bl} \\ \hline \text{Glass} \\ 11.4 \\ 12.6 \\ 5.3 \\ 17.3 \\ 15.2 \\ \hline \textbf{24.7}_{\pm 0.3} \\ 23.5 \\ 29.6 \\ 48.8 \\ 40.1 \\ \end{array}$	ur Motion 21.4 21.8 27.4 28.7 28.6 <b>41.8</b> ±0.1 33.9 40.4 55.9 47.7	Zoom 24.9 26.9 <b>30.3</b> 29.3 <b>30.3</b> 24.4 <sub>±9.8</sub> 27.1 34.4 51.3 42.6	Snow           40.4           40.7           44.7           44.7           44.0 <b>54.5</b> ±0.2           15.9           24.9           35.7	Wea           Frost           47.3           47.0           24.9           44.4           44.7 <b>52.2</b> $\pm 0.2$ 26.5           34.7           21.1           43.4	$\begin{array}{r} \text{ather} \\ \hline \text{Fog} \\ \hline 33.6 \\ 34.4 \\ 2.0 \\ \textbf{40.2} \\ 29.0 \\ \hline 20.7_{\pm 26.8} \\ \hline 47.2 \\ 55.1 \\ 66.9 \\ 60.8 \\ \end{array}$	Brit. 69.3 69.5 72.1 71.0 72.3 <b>73.5</b> $\pm 0.0$ 54.7 64.8 75.1 65.6	Contr. $36.3$ $36.5$ $46.2$ $44.5$ $44.6$ $53.5_{\pm 0.2}$ $44.1$ $54.9$ $65.0$ $61.1$	Dig Elastic 18.6 19.2 7.8 27.0 13.1 48.5±0.3 30.5 37.4 54.0 46.5	$\begin{array}{r} \text{pixel} \\ \hline 28.4 \\ 32.1 \\ 52.6 \\ 46.8 \\ 46.8 \\ \hline 60.2 \pm 0.0 \\ 44.5 \\ 55.4 \\ 67.0 \\ 60.5 \end{array}$	JPEG 52.3 53.3 56.3 55.6 56.1 <b>59.8</b> ±0.1 <b>47.8</b> 57.6 64.3 58.2	Avg. 30.6 31.2 24.7 36.4 34.5 <b>44.4</b> ±1.2 29.9 39.1 48.9 46.3
Batch Size 1 ResNet-50-GN • MEMO • Tent • EATA • SAR • DeYO (ours) VitBase-LN • MEMO • Tent • EATA • SAR	$\begin{tabular}{ c c c c c c c c c c c c c c c c c c c$	Noise Shot 19.8 20.5 4.2 27.9 26.6 $44.7_{\pm 0.4}$ 6.8 17.3 1.6 26.7 36.9	Impul.           17.9           18.4           4.0           25.8           23.9           43.0±0.7           8.2           20.6           43.9           30.3           41.9	Defoc.           19.8           17.1           16.5           17.9           18.5 <b>22.5</b> ±0.1           29.0           37.1           52.8           43.8           53.7	$\begin{array}{c} & \text{Bl}\\ \hline \\ Glass\\ 11.4\\ 12.6\\ 5.3\\ 17.3\\ 15.2\\ \textbf{24.7}_{\pm 0.3}\\ \textbf{23.5}\\ 29.6\\ 48.8\\ 40.1\\ 50.5\\ \end{array}$	$\begin{array}{c} \text{ur}\\ \hline \text{Motion}\\ \hline 21.4\\ 21.8\\ 27.4\\ 28.6\\ \hline 41.8_{\pm 0.1}\\ \hline 33.9\\ 40.4\\ 55.9\\ 47.7\\ 57.4\\ \end{array}$	$\begin{array}{c} \hline Zoom \\ 24.9 \\ 26.9 \\ 30.3 \\ 29.3 \\ 30.3 \\ 24.4_{\pm 9.8} \\ 27.1 \\ 34.4 \\ 51.3 \\ 42.6 \\ 52.8 \\ \end{array}$	$\begin{tabular}{c} Snow \\ \hline 40.4 \\ 40.4 \\ 17.7 \\ 44.7 \\ 44.0 \\ \hline 54.5 {\pm} 0.2 \\ 15.9 \\ 24.9 \\ 22.9 \\ 35.7 \\ 58.9 \\ \end{tabular}$	Wes           Frost           47.3           47.0           24.9           44.4           44.7           52.2 $\pm$ 0.2           26.5           34.7           21.1           43.4           52.7	$\begin{array}{r} & Fog \\ \hline 33.6 \\ 34.4 \\ 2.0 \\ \textbf{40.2} \\ 29.0 \\ \hline 20.7_{\pm 26.8} \\ \hline 47.2 \\ 55.1 \\ 66.9 \\ 60.8 \\ 68.9 \\ \end{array}$	Brit. 69.3 69.5 72.1 71.0 72.3 <b>73.5</b> ±0.0 54.7 64.8 75.1 65.6 76.0	Contr. 36.3 36.5 44.5 44.5 44.6 <b>53.5</b> ±0.2 44.1 54.9 65.0 61.1 65.8	Dig Elastic 18.6 19.2 7.8 27.0 13.1 48.5±0.3 30.5 37.4 54.0 46.5 57.9	tital Pixel 28.4 32.1 52.6 46.8 46.8 46.8 $60.2_{\pm 0.0}$ 44.5 55.4 67.0 60.5 68.9	JPEG 52.3 53.3 56.3 55.6 56.1 <b>59.8</b> ±0.1 <b>47.8</b> 57.6 64.3 58.2 65.8	Avg.           30.6           31.2           24.7           36.4           34.5           44.4±1.2           29.9           39.1           48.9           46.3           56.6

## Thank you!

- TL;DR
  - Address the limitations of relying solely on entropy as a confidence metric for TTA.
- Summary
  - Theoretically prove why entropy is not enough for TTA.
    - Entropy cannot discern the CPR and TRAP factors.
  - Introduce an effective TTA method based on the proposed novel confidence metric.
  - Achieve state-of-the-art performances in various TTA scenarios.
- More details can be found:
  - Paper: <u>https://openreview.net/forum?id=9w3iw8wDuE</u>
  - Project page: <u>https://whitesnowdrop.github.io/DeYO/</u>
  - Code: <u>https://github.com/Jhyun17/DeYO</u>
  - Poster Session: Tue 7 May 10:45 am 12:45 pm at Halle B