NEURAL INFORMATION PROCESSING SYSTEMS

NeurIPS 2024

The Thirty-Eighth Annual Conference on Neural Information Processing Systems

Exploring Structured Semantic Priors Underlying Diffusion Score for Test-time Adaptation

Mingjia Li[1] Shuang Li[2,*] Tongrui Su[1] Longhui Yuan[1] Jian Liang[3] Wei Li[4,*]

[1] Beijing Institute of Technology, China [2] Beihang University, China

[3] Kuaishou Technology, China [4] Inceptio Technology, China

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY 北京航空航天大學 BEIHANG UNIVERSITY KUAISHOU INCEPTIO TECHNOLOGY

Background

- Test-time Adaptation (TTA)
  - A task model f_θ pre-trained on labeled source data $\mathcal{D}_S = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$.
  - Adapt to unlabeled test-time target data $\mathcal{D}_T = \{\mathbf{x}_j\}_{j=1}^{M}$ on the fly.
  - Under distribution shifts: $\mathbf{x}_i \sim P_S(\mathbf{x}), \mathbf{x}_j \sim P_T(\mathbf{x}), P_S(\mathbf{x}) \neq P_T(\mathbf{x})$.
- Diffusion Models
  - A family of generative models excel at modeling data distribution
  - Learning to restore the gradually destroyed data structure
  - Discriminativeness revealed in conditional diffusion models

Motivation

- Generative Modeling
  - Captures the underlying structure of data
  - Faster adaptation to unseen data (vs. discriminative modeling)
  - Potential in facilitating discriminative tasks (e.g., JEM)
- Existing Art Diffusion-TTA
  - Employs diffusion models to achieve competitive TTA performance
  - Relies on computationally demanding Monte-Carlo method
  - Knowledge from low-dimensional conditioning space, limited versatility

Prabhudesai, Mihir, et al. "Test-time adaptation of discriminative models via diffusion generative feedback." *Advances in Neural Information Processing Systems* 36 (2024).

Method: DUSA

- Score Function: $\nabla_{\mathbf{x}} \log p(\mathbf{x})$
- Semantic Structure between Score Functions

$$\underbrace{\nabla_{\mathbf{x}} \log p(\mathbf{x})}_{\text{score function}} = \sum_{y} \underbrace{p(y \mid \mathbf{x})}_{\text{implicit priors}} \underbrace{\nabla_{\mathbf{x}} \log p(\mathbf{x} \mid y)}_{\text{cond. score functions}}$$

- Score-Noise Connection (Tweedie's Formula)

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t) = -\frac{\boldsymbol{\epsilon}}{\sqrt{1-\bar{\alpha}_t}}$$

- Conditional Score Estimation

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t \mid y) = -\frac{\boldsymbol{\epsilon}_\phi(\mathbf{x}_t, t, \mathbf{c}_y)}{\sqrt{1-\bar{\alpha}_t}}$$

Method: DUSA

- Structured Semantic Priors in Diffusion Score

$$\underbrace{\boldsymbol{\epsilon}}_{\text{real noise}} = \sum_{y} \underbrace{p(y \mid \mathbf{x}_t)}_{\text{implicit priors}} \underbrace{\boldsymbol{\epsilon}_\phi(\mathbf{x}_t, t, \mathbf{c}_y)}_{\text{cond. noise estimations}}$$

- Embed task model f_θ to extract knowledge from diffusion model $\boldsymbol{\epsilon}_\phi$

$$\mathcal{L}_{DUSA}(\theta, \phi) = \mathbb{E}_{\epsilon}\left[\left\|\boldsymbol{\epsilon} - \sum_{y} p_\theta(y \mid \mathbf{x}_0)\boldsymbol{\epsilon}_\phi(\mathbf{x}_t, t, \mathbf{c}_y)\right\|_2^2\right]$$

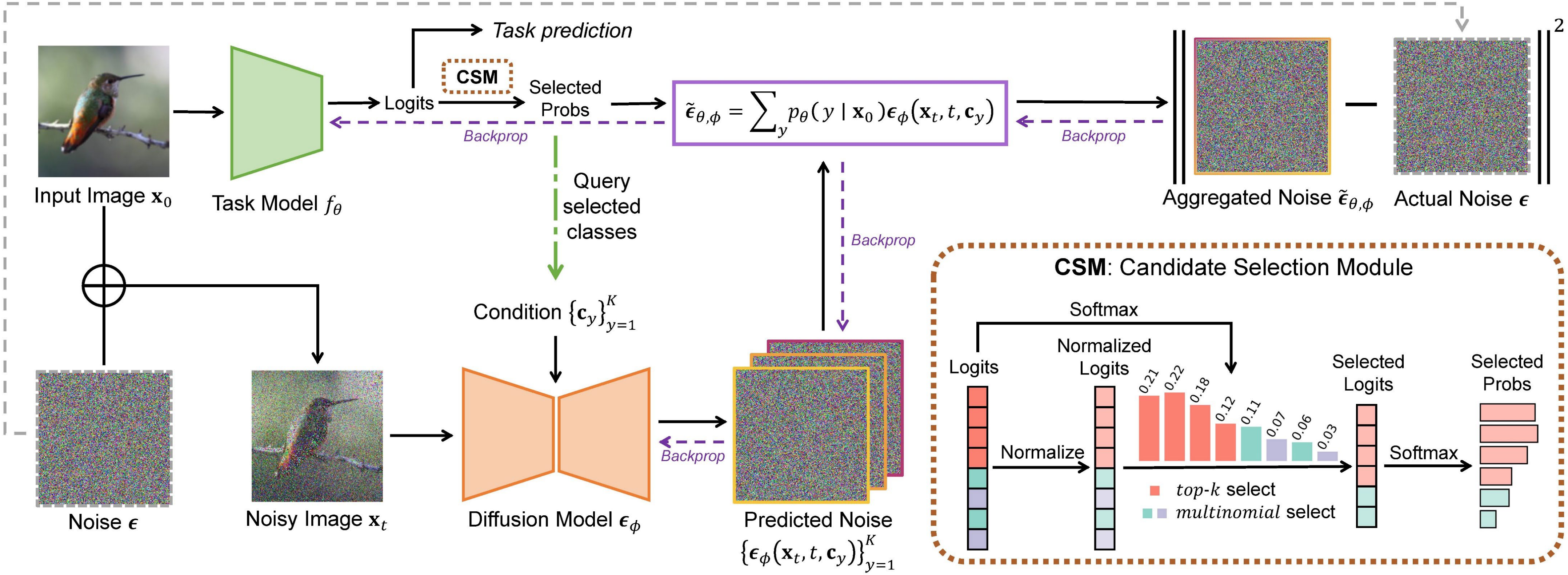
✓ Semantic priors from any single timestep

✓ Knowledge from a high-dimensional latent space (noise space)

Method: DUSA

$$\mathcal{L}_{DUSA}(\theta,\phi) = \mathbb{E}_{\epsilon}\left[\left\| \boldsymbol{\epsilon} - \sum_{y} p_{\theta}(y \mid \mathbf{x}_0)\boldsymbol{\epsilon}_{\phi}(\mathbf{x}_t, t, \mathbf{c}_y)\right\|_2^2\right]$$

- Joint update of task model f_θ and diffusion model $\boldsymbol{\epsilon}_\phi$
- A CSM to reduce computational complexity: $\mathcal{O}(K) \Rightarrow \mathcal{O}(b = k + m)$



task model-driven update

Method: DUSA-U

$$\mathcal{L}_{DUSA}(\theta,\phi) = \mathbb{E}_{\epsilon}\left[\left\|\epsilon - \sum_{y} p_{\theta}(y \mid \mathbf{x}_0)\epsilon_{\phi}(\mathbf{x}_t, t, \mathbf{c}_y)\right\|_2^2\right]$$

- Another Semantic Structure in CFG-based Diffusion Models

$$\underbrace{\epsilon_{\phi}(\mathbf{x}_t, t, \emptyset)}_{\text{uncond. noise estimation}} = \sum_{y} \underbrace{p(y \mid \mathbf{x}_t)}_{\text{implicit priors}} \underbrace{\epsilon_{\phi}(\mathbf{x}_t, t, \mathbf{c}_y)}_{\text{cond. noise estimations}}$$

implicit prior-driven update

- Separate Update of Task Model and Diffusion Model

$$\mathcal{L}_{cond}(\theta) = \mathbb{E}_{\epsilon}\left[\left\|\epsilon - \sum_{y} p_{\theta}(y \mid \mathbf{x}_0)\epsilon_{\phi}(\mathbf{x}_t, t, \mathbf{c}_y)\right\|_2^2\right], \mathcal{L}_{uncond}(\phi) = \mathbb{E}_{\epsilon}\left[\left\|\epsilon - \epsilon_{\phi}(\mathbf{x}_t, t, \emptyset)\right\|_2^2\right],$$

$$\mathcal{L}_{DUSA-U} = \mathcal{L}_{cond}(\theta) + \mathcal{L}_{uncond}(\phi)$$

✓ Vastly reduced computational overhead for diffusion model update

Method: DUSA-seg

$$\mathcal{L}_{DUSA}(\theta,\phi) = \mathbb{E}_{\epsilon}\left[\left\|\epsilon - \sum_{y} p_\theta(y \mid \mathbf{x}_0)\epsilon_\phi(\mathbf{x}_t, t, \mathbf{c}_y)\right\|_2^2\right]$$

- Easily Applicable to Dense Prediction Tasks
  - Take semantic segmentation as an example
  - Correspondence between image space and latent space (LDM)
  - Per-pixel noise can be acquired by taking elements from image-level noise

$$\epsilon_\phi\left(\mathbf{x}_{t,(h,w)}, t, \mathbf{c}_k\right) \leftarrow \epsilon_\phi(\mathbf{x}_t, t, \mathbf{c}_k)_{h,w}$$

  - The objective is almost unchanged:

$$\mathcal{L}_{DUSA-seg} = \mathbb{E}_{\epsilon,(h,w)}\left[\left\|\epsilon - \sum_{k=1}^{K} p_\theta(\mathbf{y} \mid \mathbf{x}_0)_{h,w,k} \cdot \epsilon_\phi(\mathbf{x}_t, t, \mathbf{c}_k)_{h,w}\right\|_2^2\right]$$

Results: Fully TTA of ImageNet Classifiers

Table 1: *Fully test-time adaptation* of ImageNet classifiers on ImageNet-C. The best results are in bold and runner-ups are underlined. GN/LN is short for Group/Layer normalization.

	Noise			Blur				Weather				Digital				
Method	Gauss.	Shot	Impul.	Defoc.	Glass	Motion	Zoom	Snow	Frost	Fog	Brit.	Contr.	Elastic	Pixel	JPEG	Avg.
ResNet-50 (GN)	22.1	23.0	22.0	19.8	11.4	21.5	25.0	40.3	47.0	34.0	68.8	36.3	18.5	29.3	52.6	31.4
• Tent	25.3	29.1	24.5	14.9	9.9	21.6	22.3	27.5	32.1	3.5	69.9	42.0	10.3	48.6	54.6	29.1
• CoTTA	22.1	23.0	22.0	19.8	11.4	21.5	25.1	40.3	47.0	34.0	68.8	36.4	18.5	29.3	52.6	31.5
• EATA	38.6	40.9	39.7	27.3	26.7	36.5	38.6	50.8	49.1	55.6	72.0	49.9	40.5	55.7	58.2	45.3
• SAR	39.6	42.4	41.0	19.8	22.9	37.1	38.7	27.3	47.4	55.1	72.4	48.8	7.2	54.9	57.4	40.8
• RoTTA	22.8	23.8	22.5	19.7	12.0	21.8	25.2	41.3	47.5	34.6	69.2	36.8	19.2	29.9	52.9	31.9
• Diffusion-TTA	42.0	44.6	42.4	**38.3**	**39.5**	46.9	48.2	56.5	**56.3**	60.0	72.6	45.6	**57.9**	61.4	58.0	51.3
• DUSA (Ours)	**45.2**$_{\pm0.0}$	**47.3**$_{\pm0.0}$	**46.3**$_{\pm0.1}$	37.3$_{\pm0.1}$	37.6$_{\pm0.2}$	**48.4**$_{\pm0.0}$	**50.3**$_{\pm0.3}$	**59.1**$_{\pm0.1}$	55.6$_{\pm0.0}$	**63.3**$_{\pm0.3}$	**73.3**$_{\pm0.0}$	**55.1**$_{\pm0.0}$	56.5$_{\pm0.3}$	**63.2**$_{\pm0.1}$	**60.9**$_{\pm0.2}$	**53.3**
• DUSA-U (Ours)	45.0$_{\pm0.1}$	47.1$_{\pm0.1}$	46.1$_{\pm0.0}$	36.8$_{\pm0.2}$	37.7$_{\pm0.1}$	47.9$_{\pm0.1}$	49.5$_{\pm0.3}$	59.0$_{\pm0.1}$	55.4$_{\pm0.1}$	63.0$_{\pm0.2}$	73.1$_{\pm0.1}$	54.3$_{\pm0.0}$	56.4$_{\pm0.2}$	62.9$_{\pm0.1}$	60.5$_{\pm0.3}$	53.0
ViT-B/16 (LN)	38.3	35.4	38.1	29.5	24.2	32.8	30.5	36.4	45.0	50.4	68.3	22.5	39.4	52.7	53.5	39.8
• Tent	53.9	54.5	54.1	44.4	47.2	53.8	6.7	4.6	61.9	65.4	72.9	54.9	58.0	65.1	64.1	50.8
• CoTTA	38.3	35.4	38.1	29.5	24.2	32.8	30.5	36.4	45.0	50.4	68.3	22.5	39.4	52.7	53.5	39.8
• EATA	55.4	56.3	55.3	48.9	53.4	58.6	58.2	63.5	64.1	67.5	74.3	56.5	65.7	68.5	**66.6**	60.9
• SAR	53.9	54.3	54.1	46.0	47.8	54.2	49.4	28.2	61.4	64.3	72.8	54.3	59.2	64.8	63.5	55.2
• RoTTA	42.6	39.9	42.9	30.6	26.4	34.8	31.7	39.2	47.8	52.4	68.8	23.3	42.0	55.0	54.0	42.1
• Diffusion-TTA	52.1	54.5	53.5	49.3	52.9	56.9	55.6	60.6	63.0	64.2	72.6	47.4	66.4	67.6	62.5	58.6
• DUSA (Ours)	**56.6**$_{\pm0.2}$	**57.9**$_{\pm0.2}$	**57.0**$_{\pm0.0}$	**53.3**$_{\pm0.1}$	**56.7**$_{\pm0.3}$	**62.4**$_{\pm0.1}$	**61.6**$_{\pm0.1}$	**65.9**$_{\pm0.1}$	**65.7**$_{\pm0.1}$	**70.1**$_{\pm0.1}$	**75.3**$_{\pm0.1}$	**60.2**$_{\pm0.3}$	**67.9**$_{\pm0.1}$	**69.7**$_{\pm0.1}$	65.8$_{\pm0.1}$	**63.1**
• DUSA-U (Ours)	56.3$_{\pm0.1}$	57.6$_{\pm0.1}$	56.7$_{\pm0.1}$	52.5$_{\pm0.1}$	56.4$_{\pm0.3}$	61.9$_{\pm0.1}$	60.4$_{\pm0.2}$	65.8$_{\pm0.2}$	65.4$_{\pm0.2}$	70.0$_{\pm0.1}$	75.3$_{\pm0.0}$	58.7$_{\pm0.2}$	67.8$_{\pm0.1}$	69.4$_{\pm0.0}$	64.3$_{\pm0.1}$	62.6
ConvNeXt-L (LN)	56.7	56.2	58.3	35.1	20.7	47.6	43.5	58.9	59.8	48.0	76.6	55.7	34.0	42.3	63.3	50.5
• Tent	57.4	57.8	58.9	35.7	24.3	51.3	46.3	59.8	58.4	11.0	77.1	61.2	35.1	50.0	64.4	49.9
• CoTTA	56.7	56.2	58.3	35.1	20.7	47.6	43.5	59.0	59.9	48.0	76.6	55.7	34.0	42.3	63.3	50.5
• EATA	57.5	58.0	59.0	38.7	27.1	51.6	47.0	60.7	58.5	49.3	77.2	61.3	40.2	50.3	64.5	53.4
• SAR	57.0	56.7	58.8	37.4	26.6	50.9	46.3	60.1	57.6	12.4	77.0	61.9	37.1	51.4	64.1	50.4
• RoTTA	57.0	56.7	58.7	35.1	21.3	48.0	44.0	59.5	60.0	48.9	76.6	56.8	34.6	43.1	63.4	50.9
• Diffusion-TTA	58.7	59.6	58.3	50.3	48.8	57.6	54.8	63.3	64.8	68.6	77.4	60.9	62.0	65.6	65.5	61.1
• DUSA (Ours)	**64.2**$_{\pm0.1}$	**65.5**$_{\pm0.1}$	**65.6**$_{\pm0.1}$	**54.7**$_{\pm0.1}$	**53.6**$_{\pm0.2}$	**63.8**$_{\pm0.1}$	**61.9**$_{\pm0.1}$	**70.1**$_{\pm0.1}$	**66.6**$_{\pm0.2}$	**72.7**$_{\pm0.3}$	**79.7**$_{\pm0.0}$	**68.9**$_{\pm0.0}$	**66.1**$_{\pm0.2}$	**70.7**$_{\pm0.2}$	**69.3**$_{\pm0.1}$	**66.2**
• DUSA-U (Ours)	63.8$_{\pm0.1}$	65.2$_{\pm0.0}$	65.2$_{\pm0.1}$	54.0$_{\pm0.1}$	53.3$_{\pm0.2}$	63.3$_{\pm0.1}$	60.6$_{\pm0.1}$	69.9$_{\pm0.1}$	66.4$_{\pm0.1}$	72.5$_{\pm0.2}$	79.6$_{\pm0.0}$	68.1$_{\pm0.0}$	65.9$_{\pm0.2}$	70.3$_{\pm0.2}$	68.7$_{\pm0.1}$	65.8

Results: Continual TTA of ImageNet Classifiers

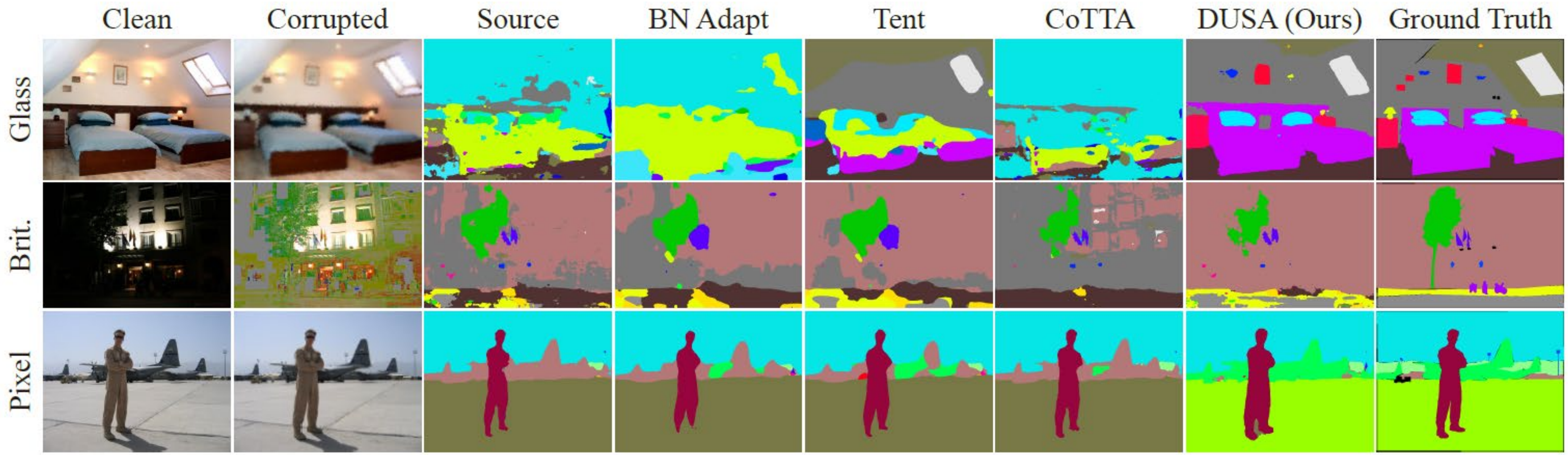
Table 2: *Continual test-time adaptation* of ImageNet pre-trained ConvNext-L on ImageNet-C. The best results are in bold and runner-ups are underlined. LN is short for Layer normalization.

Time	$t \longrightarrow$															
Method	Gauss.	Shot	Impul.	Defoc.	Glass	Motion	Zoom	Snow	Frost	Fog	Brit.	Contr.	Elastic	Pixel	JPEG	Avg.
ConvNeXt-L (LN)	56.7	56.2	58.3	35.1	20.7	47.6	43.5	58.9	59.8	48.0	76.6	55.7	34.0	42.3	63.3	50.5
• Tent	57.4	60.0	62.9	38.7	32.8	53.7	50.0	60.3	60.2	67.4	77.5	64.9	23.4	52.3	64.6	55.1
• CoTTA	56.7	56.2	58.3	35.1	20.7	47.6	43.5	59.0	59.9	48.1	76.6	55.8	34.1	42.3	63.3	50.5
• SAR	57.0	59.6	62.6	40.9	32.5	55.1	51.1	61.1	61.2	68.3	78.0	65.4	28.4	52.1	65.2	55.9
• EATA	57.6	61.0	63.5	42.5	35.2	55.3	52.4	62.3	62.9	68.6	78.3	66.1	46.2	56.7	66.9	58.3
• RoTTA	57.0	58.2	60.9	34.2	24.5	47.9	45.3	60.9	62.5	51.7	74.9	49.8	39.3	42.6	62.5	51.5
• Diffusion-TTA	58.1	63.2	63.2	54.1	**56.6**	61.8	62.5	65.2	65.5	68.1	75.3	58.9	37.3	54.8	60.9	60.4
• DUSA (Ours)	**64.1**$_{\pm 0.1}$	**67.7**$_{\pm 0.0}$	**68.3**$_{\pm 0.1}$	**54.8**$_{\pm 0.3}$	56.2$_{\pm 0.2}$	**64.6**$_{\pm 0.0}$	**65.6**$_{\pm 0.1}$	**69.8**$_{\pm 0.0}$	**69.9**$_{\pm 0.2}$	**74.5**$_{\pm 0.1}$	**79.0**$_{\pm 0.1}$	**70.3**$_{\pm 0.0}$	**68.5**$_{\pm 0.1}$	**71.9**$_{\pm 0.1}$	**70.7**$_{\pm 0.2}$	**67.7**

Results: Fully TTA of ADE20K Segmentors

Table 3: *Test-time semantic segmentation* of ADE20K pre-trained SegFormer-B5 on ADE20K-C. The best results are in bold and runner-ups are underlined. LN/BN is short for Layer/Batch normalization.

	Noise			Blur				Weather				Digital				
Method	Gauss.	Shot	Impul.	Defoc.	Glass	Motion	Zoom	Snow	Frost	Fog	Brit.	Contr.	Elastic	Pixel	JPEG	Avg.
Segformer-B5 (LN+BN)	14.2	15.8	15.6	23.1	16.8	22.5	10.3	22.3	21.5	38.6	42.0	23.1	24.5	33.1	35.3	23.9
• BN Adapt	10.8	12.0	11.7	16.6	12.8	16.6	7.9	17.0	16.8	29.6	32.4	18.2	19.2	25.5	26.3	18.2
• Tent	11.2	13.0	12.5	17.0	13.5	16.9	7.7	17.7	17.4	29.7	32.5	18.6	20.0	25.8	26.4	18.7
• CoTTA	14.6	16.1	15.8	22.6	16.5	22.1	9.8	20.9	20.4	38.8	42.3	21.9	24.3	33.6	35.4	23.7
• DUSA (Ours)	**23.6**$_{\pm1.3}$	**24.5**$_{\pm1.0}$	**23.2**$_{\pm0.3}$	**24.7**$_{\pm0.5}$	**23.2**$_{\pm1.2}$	**24.7**$_{\pm0.6}$	**12.5**$_{\pm0.6}$	**27.3**$_{\pm1.2}$	**26.7**$_{\pm0.8}$	**39.3**$_{\pm0.2}$	**42.6**$_{\pm0.3}$	**27.1**$_{\pm1.2}$	**30.6**$_{\pm0.6}$	**35.7**$_{\pm0.7}$	**35.6**$_{\pm0.7}$	**28.1**



Results: Ablation Study

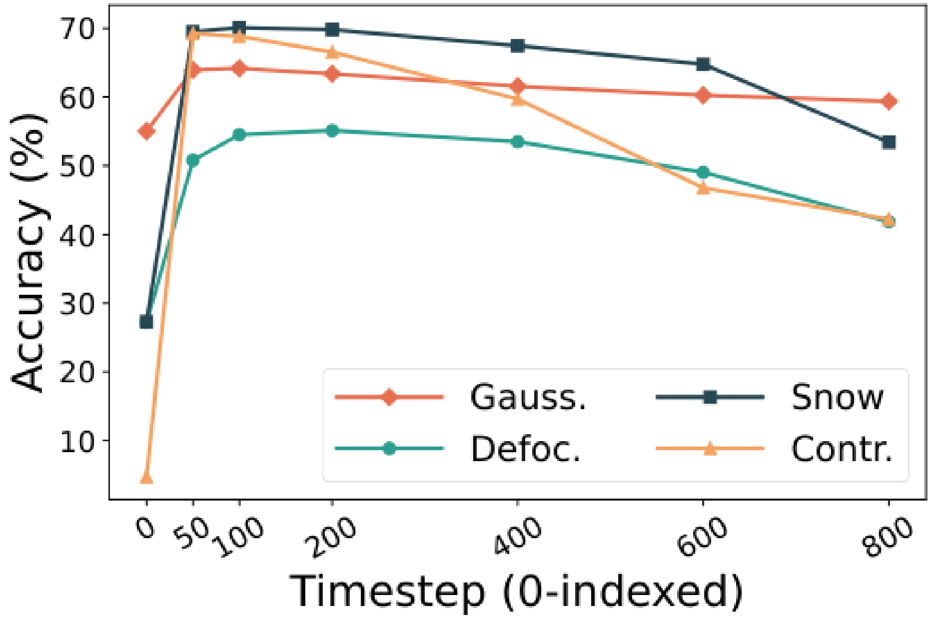


Figure 3: Accuracy of ConvNeXt-L across different selections of timestep.

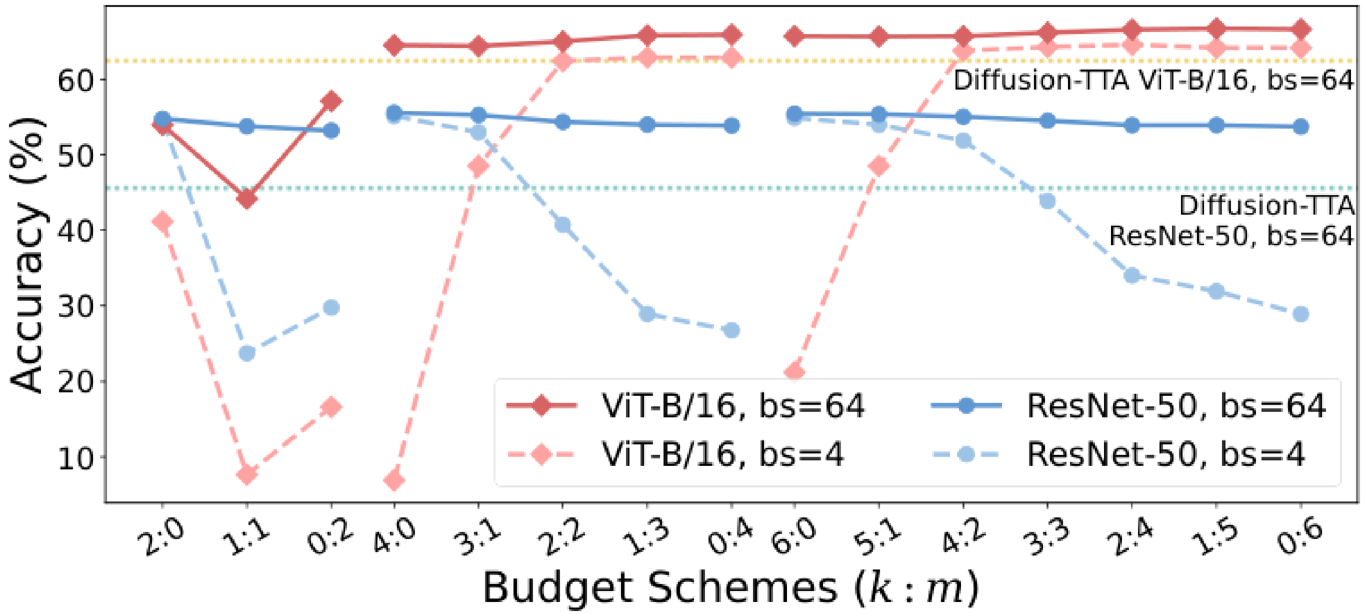


Figure 4: Accuracy of ViT-B/16 on JPEG and ResNet-50 on Contrast, across different budgets for adaptation.

Thanks for Listening!

Contact: mingjiali@bit.edu.cn

Code

https://github.com/BIT-DA/DUSA

Project Page

https://kiwixr.github.io/projects/dusa