

Supplementary Materials: Domain-Conditioned Transformer for Fully Test-time Adaptation

Anonymous Authors

In this Supplementary Materials, we provide more details and experimental results for further understanding of the proposed Domain-Conditioned Transformer method.

A SMOOTH OPTIMIZATION FOR DOMAIN-CONDITIONED GENERATOR

We follow the baseline SAR method to perform the Sharpness-Aware Minimization (SAM) for the Domain-Conditioned Transformer to get better robustness. From the perspective of generalization and optimization, SAM not only minimizes individual points within the loss landscape criterion but also consistently reduces the loss in their surrounding neighborhoods. Contrary to the conventional approach of exclusively optimizing the model weights with low loss values, SAM seeks to identify smoother minima in the weight space. These minima are characterized by uniformly low loss \mathcal{L} in the neighborhoods ϵ of model weights θ :

$$\min_{\theta} \max_{\|\epsilon\| \leq \rho} \mathcal{L}(\theta + \epsilon; x), \quad (1)$$

where $\rho \geq 0$ is a hyper-parameter to define the scope of the neighborhoods. To address this minimax problem, SAM initially addresses the maximization problem by seeking the maximum perturbation ϵ_t at training step t . This inner maximization problem can be approximated using the first-order Taylor expansion of $\mathcal{L}(\theta + \epsilon; x)$ with respect to $\epsilon \rightarrow 0$ as follows:

$$\begin{aligned} \epsilon_t(\theta) &= \arg \max_{\|\epsilon\| \leq \rho} \mathcal{L}(\theta + \epsilon; x) \\ &= \arg \max_{\|\epsilon\| \leq \rho} \mathcal{L}(\theta; x) + \epsilon^T \nabla_{\theta} \mathcal{L}(\theta; x) + o(\epsilon) \\ &\approx \arg \max_{\|\epsilon\| \leq \rho} \epsilon^T \nabla_{\theta} \mathcal{L}(\theta; x). \end{aligned} \quad (2)$$

The value $\epsilon_t(\theta)$ that solves this approximation is given by the solution to a classical dual norm problem:

$$\epsilon_t(\theta) = \rho \cdot \text{sign}(\nabla_{\theta} \mathcal{L}(\theta; x)) \frac{|\nabla_{\theta} \mathcal{L}(\theta; x)|^{q-1}}{\|\nabla_{\theta} \mathcal{L}(\theta; x)\|_q^{q/p}}, \quad (3)$$

where $\frac{1}{p} + \frac{1}{q} = 1$. It is empirically confirmed that the optimization yields the best performance when $p = 2$, resulting in ϵ_t formulated as:

$$\epsilon_t(\theta) = \rho \frac{\nabla_{\theta} \mathcal{L}(\theta; x)}{\|\nabla_{\theta} \mathcal{L}(\theta; x)\|_2}. \quad (4)$$

Then the gradient update for the model weights θ_t is computed as:

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} \mathcal{L}(\theta; x)|_{\theta_t + \epsilon_t}. \quad (5)$$

Finally, SAM converges the model weights θ to a smooth minimum with respect to the loss function \mathcal{L} by iteratively updating equation 4 and equation 5.

B RUNNING TIME COMPARISONS

We provide a comprehensive comparison of running time costs in Table 1. Our proposed DCT model introduces a slightly higher computational overhead, approximately 13% slower than the SAR method, but much faster than MEMO and DDA. This extra overhead is mainly caused by the learning of the domain conditioner generator. In the original SAR method, it only updates the layer normalization parameters.

Table 1: Testing time cost comparison for Gaussian corruption of ImageNet-C under the single GPU NVIDIA RTX 3090.

Method	Time Cost
TENT	~ 5 min
SAR	~ 7 min
MEMO	~ 14 hours
DDA	~ 5 days
Ours	~ 8 min

C ADDITIONAL RESULTS ON IMAGENET-C WITH SEVERITY LEVEL 3

We provide additional performance comparison results for corruption severity Level 3 with the Normal, Imbalanced label shifts, and Batch size = 1 settings in Table 2, 3, and 4, respectively. The results are consistent with those in the main paper for severity level 5. We can see that our DCT method outperforms existing methods in almost all 15 corruption types.

D ADDITIONAL RESULTS ON IMAGENET-C WITH ViT-L/16 BACKBONE

We extend our experimentation to encompass a larger ViT-L/16 backbone, operating within the contexts of both Normal and Batch size = 1 settings. For the Normal setting, the batch size is set to 32 due to limited memory. The learning rate of the domain-conditioner generator is set to 0.001 and 0.0001 respectively. The results, as illustrated in Table 5 and 6, consistently showcase the superiority of our proposed DCT method over the baseline SAR in both configurations. This robust performance demonstrates the efficiency of our proposed DCT method across diverse transformer backbones.

E ADDITIONAL RESULTS ON DOMAINNET-126 DATASET

Additionally, we conducted experiments on DomainNet-126, a widely used dataset in domain adaptation and domain generalization tasks. Utilizing the same ViT-B/16 pre-trained model provided by *timm*, we replace the classifier head and fine-tuned it for 100 epochs in each domain as the source model. The results are summarized in Table 7. Notably, our method surpasses the second-best approach by 0.7% on average.

Table 2: Classification Accuracy (%) for each corruption in ImageNet-C under Normal at the severity Level 3. The best result is shown in bold.

Method	gaus	shot	impul	defcs	gls	mtn	zm	snw	frst	fg	brt	cnt	els	px	jpg	Avg.
Source	72.1	71.5	71.3	62.3	51.1	69.1	59.1	69.0	60.0	70.1	80.1	74.0	75.1	77.6	75.2	69.2
DDA	62.6	63.1	62.3	50.2	54.2	50.9	43.3	41.0	41.9	14.8	60.6	26.0	62.2	62.6	62.8	50.6
TENT	74.3	73.9	73.6	70.8	66.6	73.7	66.9	73.2	68.7	76.0	81.6	78.9	78.5	79.7	77.1	74.2
SAR	74.3	73.9	73.7	70.9	66.5	73.8	66.9	73.1	68.7	75.8	81.8	78.9	78.5	79.8	77.1	74.2
Ours	74.7	74.6	74.4	71.3	69.6	74.4	70.1	74.8	71.6	78.1	81.9	79.6	79.3	80.1	78.7	75.5
	±0.1	±0.0	±0.1	±0.0	±0.1	±0.1	±0.2	±0.0	±0.1	±0.1	±0.1	±0.2	±0.1	±0.0	±0.2	±0.0

Table 3: Classification Accuracy (%) for each corruption in ImageNet-C under Imbalanced label shifts at the severity Level 3.

Method	gaus	shot	impul	defcs	gls	mtn	zm	snw	frst	fg	brt	cnt	els	px	jpg	Avg.
Source	51.5	46.8	50.4	48.7	37.1	54.7	41.6	35.1	33.3	68.0	69.3	74.9	65.9	66.0	63.6	53.8
DDA	62.6	63.1	62.3	50.2	54.2	50.9	43.3	41.0	41.9	14.8	60.6	26.0	62.2	62.6	62.8	50.6
TENT	74.6	74.3	74.0	71.4	68.4	74.6	68.3	73.7	69.8	77.0	81.9	79.2	79.2	80.0	78.1	75.0
SAR	74.7	74.4	74.2	71.5	68.7	74.8	68.6	74.0	70.2	77.0	82.1	79.4	79.3	80.2	78.2	75.1
Ours	74.6	74.5	74.3	71.5	69.5	74.7	69.8	75.0	71.4	77.9	81.9	79.6	79.2	80.2	78.9	75.5
	±0.2	±0.1	±0.1	±0.1	±0.2	±0.0	±0.2	±0.1	±0.1	±0.3	±0.1	±0.1	±0.2	±0.1	±0.1	±0.1

Table 4: Classification Accuracy (%) for each corruption in ImageNet-C under Batch size=1 at the severity Level 3.

Method	gaus	shot	impul	defcs	gls	mtn	zm	snw	frst	fg	brt	cnt	els	px	jpg	Avg.
Source	51.6	46.9	50.5	48.7	37.2	54.7	41.6	35.1	33.5	67.8	69.3	74.8	65.8	66.0	63.7	53.8
DDA	62.6	63.1	62.3	50.2	54.2	50.9	43.3	41.0	41.9	14.8	60.6	26.0	62.2	62.6	62.8	50.6
TENT	74.4	73.9	73.6	70.9	66.6	73.7	67.0	73.1	68.7	76.0	81.6	79.0	78.5	79.8	77.1	74.3
SAR	74.9	74.6	74.0	71.4	68.2	74.5	68.2	73.9	70.0	76.3	81.2	78.9	78.4	79.3	77.1	74.7
Ours	74.6	74.4	74.1	71.5	69.6	74.4	69.1	73.4	71.7	77.4	81.1	79.0	77.9	79.2	78.1	75.0
	±0.1	±0.3	±0.2	±0.2	±0.2	±0.3	±0.3	±0.2	±0.1	±0.5	±0.2	±0.2	±0.3	±0.1	±0.2	±0.0

Table 5: Classification Accuracy (%) in ImageNet-C with ViT-L/16 under Normal at the highest severity (Level 5).

Method	gaus	shot	impul	defcs	gls	mtn	zm	snw	frst	fg	brt	cnt	els	px	jpg	Avg.
Source	62.1	61.4	62.3	52.7	45.1	60.6	55.1	66.2	62.4	62.5	80.2	39.8	56.2	74.3	72.7	60.9
TENT	65.6	68.3	67.6	63.4	59.9	66.8	60.7	69.0	68.5	67.4	81.0	28.9	64.7	77.2	74.7	65.6
SAR	66.0	66.6	66.2	61.3	55.1	66.1	58.3	68.4	65.7	66.3	81.0	26.8	63.7	74.5	73.6	64.0
Ours	67.1	68.2	66.9	64.0	62.4	66.7	64.1	71.1	69.1	70.4	81.3	65.3	69.7	77.6	75.5	69.3
	±0.7	±0.4	±0.9	±1.0	±1.3	±0.5	±0.5	±0.9	±0.5	±0.7	±0.1	±0.6	±1.9	±0.1	±0.5	±0.3

Table 6: Classification Accuracy (%) in ImageNet-C with ViT-L/16 under Batch size = 1 at the highest severity (Level 5).

Method	gaus	shot	impul	defcs	gls	mtn	zm	snw	frst	fg	brt	cnt	els	px	jpg	Avg.
Source	62.1	61.4	62.3	52.7	45.1	60.6	55.1	66.2	62.4	62.5	80.2	39.8	56.2	74.3	72.7	60.9
TENT	55.2	67.8	67.8	43.5	59.6	66.9	62.8	69.9	67.4	68.3	81.4	31.4	64.7	77.4	73.0	63.8
SAR	67.9	64.5	67.7	63.0	61.6	63.7	62.4	70.4	67.6	68.7	75.7	60.2	55.5	76.4	74.7	66.7
Ours	67.5	69.2	67.9	64.6	64.5	68.5	65.7	71.0	68.8	71.6	80.6	65.0	71.2	76.5	75.3	69.9
	±0.3	±0.2	±0.0	±0.6	±0.9	±0.1	±1.6	±0.5	±0.7	±0.0	±0.1	±0.4	±1.7	±0.2	±0.7	±0.3

Table 7: Classification Accuracy (%) for test-time adaptation of all transfer tasks in DomainNet-126 dataset.

Method	C→P	C→R	C→S	P→C	P→R	P→S	R→C	R→P	R→S	S→C	S→P	S→R	Avg.
Source	71.7	84.0	69.2	72.0	85.6	61.0	60.5	65.1	48.4	72.6	72.7	82.9	70.5
TENT	71.9	83.8	69.1	72.3	85.4	52.0	60.7	65.3	47.2	74.5	69.2	82.3	69.5
SAR	71.7	83.5	69.2	72.1	85.2	60.2	60.9	65.8	49.3	74.6	72.4	82.5	70.6
Ours	72.6	82.8	71.5	74.4	85.7	49.5	70.9	72.3	57.0	72.3	65.1	81.1	71.3
	±0.6	±0.3	±0.1	±0.3	±0.7	±1.7	±0.5	±0.2	±0.5	±0.7	±1.0	±0.2	±0.3