Feature Augmentation based Test-Time Adaptation
Supplementary Material

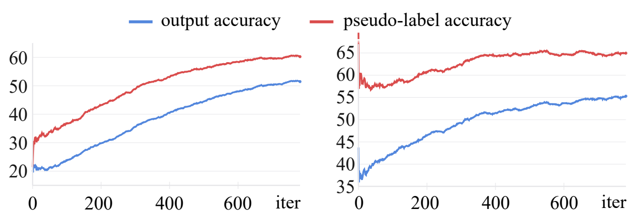
Younggeol Cho* Youngrae Kim* Junho Yoon Seunghoon Hong Dongman Lee
KAIST
{rangewing, youngrae.kim, vpdtlrdl, seunghoon.hong, dlee}@kaist.ac.kr

Due to the limited space of the main paper, we provide more experimental results and implementation details in the supplementary material.

Accuracy of pseudo-labels. Fig. A shows the accuracy of pseudo-labels as the model adapts. The pseudo-label accuracy is high compared to the output accuracy and improves over time.

Further experimental results. We present the full results of Tab. 4, Tab. 5, and Tab. 7 in the main paper, as shown in Tab. A, Tab. B, and Tab. C, respectively.

Implementation details. We provide the full data of the hyperparameters in Tab. D.



(a) elastic_transform (b) snow

Figure A. Pseudo-label accuracy (%) on ImageNet-C. We use DeYO+FATA with ResNet50-GN for this experiment. Exponential moving average ($p = 0.99$) is applied.

Method	Ar→Cl	Ar→Pr	Ar→Re	Cl→Ar	Cl→Pr	Cl→Re	Pr→Ar	Pr→Cl	Pr→Re	Re→Ar	Re→Cl	Re→Pr	Average	ΔPerf.
No Adapt	41.31	64.74	74.75	51.71	60.94	63.58	51.3	37.21	73.63	64.94	40.25	75.85	58.35	
MEMO [8]	42.18	65.65	75.03	49.94	60.26	62.82	50.31	37.25	72.44	64.81	40.37	76.73	58.15	
TENT [6]	41.4	64.74	74.75	51.71	60.89	63.64	51.3	37.16	73.65	65.02	40.21	75.87	58.36	
CoTTA [7]	41.28	64.79	71.86	51.71	60.91	63.6	51.3	37.21	72.11	64.94	40.3	70.87	57.57	
EATA [3]	41.99	64.83	74.87	51.71	60.96	63.64	51.42	37.82	73.67	65.1	41.03	75.94	58.58	
EATA+FATA	42.2	65.74	74.39	53.61	61.66	65.34	53.61	39.11	74.27	67.08	43.09	76.41	59.71	+1.13
SAR [4]	41.37	64.77	74.78	51.67	60.91	63.64	51.34	37.14	73.63	65.02	40.25	75.87	58.37	
SAR+FATA	42.02	64.63	74.73	53.32	61.59	64.82	51.63	38.72	73.88	65.51	42.47	76.82	59.18	+0.81
DeYO [2]	41.63	64.7	74.71	51.96	61.07	63.87	51.34	37.71	73.74	65.1	40.62	76.3	58.56	
DeYO+FATA	41.74	64.97	74.73	51.96	61.34	63.99	51.55	37.73	73.74	64.94	40.82	76.19	58.64	+0.08

Table A. Image classification results on Office-Home [5]. ResNet50 with GN is used for this experiment. We use the accuracy (%) as the metric. ΔPerf. is the performance gap between the original method and another version where our method is incorporated.

Method	Noise			Blur				Weather				Digital				Avg.
	Gauss.	Shot	Impul.	Defoc.	Glass	Motion	Zoom	Snow	Frost	Fog	Brit.	Contr.	Elastic	Pixel	JPEG	
No Adapt	17.98	19.84	17.88	19.75	11.35	21.42	24.92	40.43	47.30	33.59	69.28	36.27	18.61	28.40	52.28	30.62
DeYO [2]	39.46	41.90	41.03	22.27	24.11	38.48	37.87	50.51	49.59	1.43	73.17	49.95	41.54	55.96	57.82	41.67
DeYO+FATA (Layer 0)	35.58	38.56	36.51	20.32	17.61	41.07	40.53	48.89	48.40	54.22	71.25	49.74	8.08	55.29	56.76	41.52
DeYO+FATA (Layer 1)	39.84	42.11	40.47	21.84	23.68	39.93	41.17	55.03	50.82	1.09	72.86	51.13	39.57	57.26	57.97	42.32
DeYO+FATA (Layer 2)	39.72	42.77	41.03	21.44	23.04	39.74	42.76	54.51	52.04	0.84	73.57	51.67	45.92	57.89	58.16	43.01
DeYO+FATA (Layer 3)	39.71	42.49	41.37	22.29	24.46	38.90	38.31	51.23	50.02	56.31	73.19	50.03	42.10	55.99	57.79	45.61

Table B. Ablation study on the position of feature augmentation. ResNet50 with GN is used for this experiment.

DeYO	Augmentation loss	Noise			Blur				Weather				Digital				Avg.
		Gauss.	Shot	Impul.	Defoc.	Glass	Motion	Zoom	Snow	Frost	Fog	Brit.	Contr.	Elastic	Pixel	JPEG	
-	-	17.98	19.84	17.88	19.75	11.35	21.42	24.92	40.43	47.30	33.59	69.28	36.27	18.61	28.40	52.28	30.62
-	FATA loss (Ours)	37.92	39.99	38.70	**25.96**	22.21	36.06	37.77	**51.99**	**51.33**	2.85	73.11	48.83	13.10	53.78	56.73	39.35
✓	-	39.46	41.90	41.03	22.27	24.11	38.48	37.87	50.51	49.59	1.43	73.17	49.95	41.54	55.96	**57.82**	41.67
✓	Simple Aug.	38.19	41.31	39.31	21.64	23.43	36.96	34.93	48.72	46.50	1.14	72.90	49.19	37.43	55.39	57.39	40.30
✓	MSE loss	0.22	0.46	0.57	0.88	0.89	0.57	2.15	2.00	7.49	8.18	61.78	6.37	0.83	2.19	3.52	6.54
✓	FATA Loss (Ours)	**39.71**	**42.49**	**41.37**	22.29	**24.46**	**38.90**	**38.31**	51.23	50.02	**56.31**	**73.19**	**50.03**	**42.10**	**55.99**	57.79	**45.61**

Table C. Ablation study on the augmentation loss. ResNet50 with GN is used for this experiment.

Parameter	Meaning	Value
i	Layer to inject the feature augmentation	3 (ResNet50), 11 (ViT-B)
E_0	Entropy threshold for sample filtering	$0.5 \ln C$
E_ω	Normalization factor for sample weighting	$0.4 \ln C$
σ_n	Standard deviation for the noise	1.0
λ_{EMA}	Smoothing factor for exponential moving average	0.95
B	Default batch size	64
η	Learning rate	0.0005 (ResNet50, $B = 64$)
		0.001 (ViT-B, $B = 64$)
		0.00025 (ResNet50, $B = 1$)
		0.000016 (ViT-B, $B = 1$)

Table D. Hyperparameters for the experiment.

Rationale for replacing variance with standard deviation in Eq. (3). Eq. (3) is derived from Normalization Perturbation (NP) [1], which is formulated as follows:

$$\mathbf{z}' = \alpha\sigma_c \cdot \frac{\mathbf{z} - \mu_c}{\sigma_c} + \beta\mu_c = \alpha\mathbf{z} + (\beta - \alpha)\mu_c, \tag{A}$$

where α, $\beta \in \mathbb{R}^{B\times C}$ are the random noises sampled from $N(\mathbf{I}, \sigma_n\mathbf{I})$, μ_c, σ_c are the channel-wise feature mean and standard deviation, respectively. As formulated by Eq. (3), NP+ introduces the term $\delta = \mathrm{Var}(\mu_c)/\max(\mathrm{Var}(\mu_c))$ to control the magnitude of the noises α and β. We replace variance with standard deviation because the noises are related to standard deviation σ_c, not variance, as shown in Eq. (A).

References

[1] Qi Fan, Mattia Segu, Yu-Wing Tai, Fisher Yu, Chi-Keung Tang, Bernt Schiele, and Dengxin Dai. Towards robust object detection invariant to real-world domain shifts. In *The Eleventh International Conference on Learning Representations*, 2023. 2

[2] Jonghyun Lee, Dahuin Jung, Saehyung Lee, Junsung Park, Juhyeon Shin, Uiwon Hwang, and Sungroh Yoon. Entropy is not enough for test-time adaptation: From the perspective of disentangled factors. In *The Twelfth International Conference on Learning Representations*, 2024. 1

[3] Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Yaofo Chen, Shijian Zheng, Peilin Zhao, and Mingkui Tan. Efficient test-time model adaptation without forgetting. In *International conference on machine learning*, pages 16888–16905. PMLR, 2022. 1

[4] Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Zhiquan Wen, Yaofo Chen, Peilin Zhao, and Mingkui Tan. Towards stable test-time adaptation in dynamic wild world. In *The Eleventh International Conference on Learning Representations*, 2023. 1

[5] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5018–5027, 2017. 1

[6] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *International Conference on Learning Representations*, 2021. 1

[7] Qin Wang, Olga Fink, Luc Van Gool, and Dengxin Dai. Continual test-time domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7201–7211, June 2022. 1

[8] Marvin Zhang, Sergey Levine, and Chelsea Finn. Memo: Test time robustness via adaptation and augmentation. *Advances in Neural Information Processing Systems*, 35:38629–38642, 2022. 1