

# Feature Alignment and Uniformity for Test Time Adaptation

## Supplementary Material

Shuai Wang<sup>1</sup>   Daoan Zhang<sup>2</sup>   Zipei Yan<sup>3</sup>   Jianguo Zhang<sup>2</sup>   Rui Li<sup>1</sup>  
<sup>1</sup>Tsinghua University   <sup>2</sup>Southern University of Science and Technology  
<sup>3</sup>Hongkong Polytechnic University

s-wang20@mails.tsinghua.edu.cn, dawn\_change@126.com, zipei.yan@connect.polyu.hk  
zhangjg@sustech.edu.cn, leerui@tsinghua.edu.cn

This supplementary material provides more details about our implementation (Sec. A), additional experiments on three corruption datasets (Sec. B), some discussion (Sec. C), running time analysis (Sec. D) and detailed results (Sec. E).

### A. Other Implement Details

We run our experiments mainly on a single RTX-A6000 or RTX-2080Ti GPU, depending on the need for GPU memory. For source training, we set the batch size as 32 for each source domain and the learning rate as  $5e^{-5}$ . We set dropout probability and weight decay to zero. We train source model 5k iterations except for DomainNet. We tripled it from 5k to 15k following [2]. All images are resized to  $224 \times 224$  and data augmentation is used in source domain training, which includes randomly cropping, flipping horizontally, jittering colour, and changing the intensity.

For implementations of different test time adaptation methods, we use publicly released code of T3A [5]<sup>1</sup>, except for LAME<sup>2</sup> [1] and ETA<sup>3</sup> [10] we use source code of authors. For PL [7], we set confidence as 0.9. We resize all images to  $224 \times 224$  and no data augmentation is used during the test time adaptation process.

For different backbones, we use `torchvision` implementation<sup>4</sup> except for ViT-B/16 and MLP-mixer, we use implementation from `timm` library<sup>5</sup>. For all experiments, we use three random seeds {0,1,2} and report the average results in the main text.

### B. Experiments on Corruption Benchmark

We conduct experiments on three corruption datasets, including CIFAR-10/100-C and ImageNet-C. The results are

<sup>1</sup><https://github.com/matsuolab/T3A>

<sup>2</sup><https://github.com/fiveai/LAME>

<sup>3</sup><https://github.com/mr-eggplant/EATA>

<sup>4</sup><https://github.com/pytorch/vision>

<sup>5</sup><https://github.com/rwightman/pytorch-image-models>

listed in Table 2. From the Table 2, it is noticed that our method still achieves the state-of-the-art performance on corruption datasets.

### C. Discussion

**Performance on VLCS.** All compared methods shown poor performance on VLCS dataset. We notice that the label distribution is severely different among domains in VLCS compared to other datasets (e.g. PACS and OfficeHome), which is probably why fewer methods show performance gain in VLCS compared to other datasets. We calculated the label distribution of VLCS dataset, as shown in Table 1. Since the label distribution is not considered for existing methods, the adaptation may fail.

**Related Work.** There are some papers considering nearest neighbor information in the test time adaptation setting. TAST [6], a concurrent work published in ICLR 2023, considers nearest neighbor information to refine pseudo-labels. AdaContrast [3] uses a soft voting strategy among the nearest neighbors in the feature space to refine pseudo-labels. Both of them use nearest neighbor information to refine pseudo-labels. Different from them, our *Memorized Spatial Local Clustering* aims to cluster features with the same pseudo-label.

### D. Running Time Analysis

We provide the running time of different methods for reference. Running time is tested using RTX-A6000 GPU and AMD-EPYC-7542 32 Core Processor. The results are listed in Table 3 using ResNet50 [4] backbone.

### E. Full Results

We provide full results of different methods on ResNet-18/50, including results for each domain and error bars across three random seeds. See Table 4-11.

Table 1. Label distribution on VLCS dataset.

	0	1	2	3	4
Caltech101	237	123	118	67	870
LabelMe	80	1209	88	42	1237
SUN09	20	932	1036	30	1264
VOC2007	330	699	428	420	1499

Table 2. Accuracy on CIFAR-10/100-C and ImageNet-C. We use ResNet26 for CIFAR-10/100-C, and ResNet50 for ImageNet-C.

	CIFAR-10-C	CIFAR-100-C	ImageNet-C
ERM [14]	70.7	41.4	18.0
BN [12]	77.4	48.0	33.5
Tent [16]	80.7	51.7	42.7
PL [7]	80.2	49.8	38.4
SHOT-IM [9]	80.7	52.1	43.1
T3A [5]	77.4	44.6	36.5
ETA [10]	80.6	52.4	<b>48.1</b>
LAME [1]	79.4	50.6	47.6
Ours	<b>81.7</b>	<b>52.6</b>	48.0

## References

- [1] Malik Boudiaf, Romain Müller, Ismail Ben Ayed, and Luca Bertinetto. Parameter-free online test-time adaptation. In *CVPR*, 2022. 1, 2, 3
- [2] Junbum Cha, Sanghyuk Chun, Kyungjae Lee, Han-Cheol Cho, Seunghyun Park, Yunsung Lee, and Sungrae Park. SWAD: domain generalization by seeking flat minima. In *NeurIPS*, 2021. 1
- [3] Dian Chen, Dequan Wang, Trevor Darrell, and Sayna Ebrahimi. Contrastive test-time adaptation. In *CVPR*, 2022. 1
- [4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 1
- [5] Yusuke Iwasawa and Yutaka Matsuo. Test-time classifier adjustment module for model-agnostic domain generalization. In *NeurIPS*, 2021. 1, 2, 3
- [6] Minguk Jang, Sae-Young Chung, and Hye Won Chung. Test-time adaptation via self-training with nearest neighbor information. In *ICLR*, 2023. 1
- [7] Dong-Hyun Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, 2013. 1, 2, 3
- [8] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Deeper, broader and artier domain generalization. In *ICCV*, 2017. 3
- [9] Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *ICML*, 2020. 2, 3
- [10] Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Yaofo Chen, Shijian Zheng, Peilin Zhao, and Mingkui Tan. Efficient test-time model adaptation without forgetting. In *ICML*, 2022. 1, 2, 3
- [11] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *ICCV*, 2019. 3
- [12] Steffen Schneider, Evgenia Rusak, Luisa Eck, Oliver Bringmann, Wieland Brendel, and Matthias Bethge. Improving robustness against common corruptions by covariate shift adaptation. In *NeurIPS*, 2020. 2, 3
- [13] Antonio Torralba and Alexei A Efros. Unbiased look at dataset bias. In *CVPR*, 2011. 3
- [14] Vladimir Vapnik. *Statistical learning theory*. Wiley, 1998. 2, 3
- [15] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *CVPR*, 2017. 3
- [16] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *ICLR*, 2021. 2, 3

Table 3. Running time analysis of different methods on four datasets. The units used in the table are seconds. “GD” denotes whether the method requires gradient-based optimization.

	GD	PACS [8]	VLCS [13]	OfficeHome [15]	DomainNet [11]
ERM [14]	✗	17	17	56	736
BN [12]	✗	18	18	56	745
Tent [16]	✓	32	33	67	1094
PL [7]	✓	33	34	69	1339
SHOT-IM [9]	✓	37	39	72	1339
T3A [5]	✗	19	19	57	829
ETA [10]	✓	21	22	66	1094
LAME [1]	✗	18	19	56	750
Ours	✓	35	37	71	1644

Table 4. Full results on PACS with ResNet18.

	A	C	P	S	Avg
ERM	80.56±0.45	77.36±0.85	93.01±0.17	77.35±2.90	82.07±0.49
BN	81.60±0.16	82.00±0.51	92.85±0.24	74.86±1.10	82.82±0.34
Tent	83.43±0.53	83.02±0.74	93.88±0.38	79.35±1.15	84.92±0.32
PL	84.93±1.32	83.27±1.78	92.62±1.37	77.72±3.66	84.64±1.13
SHOT-IM	84.61±1.06	82.36±1.88	93.60±0.38	69.64±3.40	82.55±1.07
T3A	83.00±0.76	79.56±0.44	94.48±0.34	76.95±2.94	83.50±0.67
ETA	81.33±0.30	81.89±0.55	92.82±0.53	74.77±0.91	82.70±0.31
LAME	83.05±0.53	83.06±0.51	94.30±0.29	77.91±0.80	84.58±0.23
Ours	86.50±0.75	86.38±0.82	94.57±0.32	81.84±0.94	87.32±0.39

Table 5. Full results on OfficeHome with ResNet18.

	A	C	P	R	Avg
ERM	55.49±0.61	51.41±0.46	71.92±0.47	73.67±0.18	63.12±0.26
BN	54.36±1.00	51.14±0.25	71.20±0.44	72.52±0.37	62.30±0.25
Tent	55.85±0.91	53.38±0.29	72.50±0.65	73.26±0.37	63.75±0.23
PL	54.64±0.96	48.36±1.58	68.83±1.08	69.05±0.58	60.22±0.37
SHOT-IM	55.45±1.08	52.32±0.99	73.23±0.76	72.67±0.37	63.42±0.52
T3A	56.18±0.48	52.90±0.67	73.44±0.48	74.48±0.19	64.25±0.22
ETA	54.88±0.33	51.05±0.21	71.18±0.44	72.72±0.30	62.46±0.14
LAME	54.84±0.86	50.90±0.26	70.85±0.30	72.19±0.30	62.20±0.21
Ours	57.87±0.81	53.40±0.27	74.20±0.38	73.86±0.22	64.83±0.46

Table 6. Full results on VLCS with ResNet18.

	C	L	S	V	Avg
ERM	92.44±0.78	62.72±0.69	69.26±2.01	66.58±1.43	72.75±0.29
BN	76.23±1.99	57.84±0.80	58.04±0.73	65.12±0.35	64.31±0.47
Tent	82.65±1.34	60.02±1.04	60.49±0.84	66.28±0.46	67.36±0.43
PL	86.64±3.45	61.66±1.47	61.78±2.66	65.64±1.53	68.93±1.07
SHOT-IM	76.65±2.94	57.34±1.26	59.13±0.85	66.46±0.55	64.90±0.70
T3A	96.76±1.22	63.80±0.39	64.98±0.45	66.55±0.38	73.03±0.35
ETA	76.1±1.90	57.89±0.87	58.12±0.73	65.31±0.23	64.35±0.40
LAME	94.7±0.21	62.69±0.25	67.58±0.12	66.55±0.45	72.88±0.13
Ours	97.2±1.15	64.50±0.28	65.42±0.34	67.32±0.38	73.61±0.42

Table 7. Full results on DomainNet with ResNet18.

	clipart	infograph	painting	quickdraw	real	sketch	Avg
ERM	55.86±0.15	16.85±0.06	44.80±0.20	12.49±0.38	56.74±0.04	46.96±0.12	38.95±0.08
BN	55.90±0.09	12.07±0.15	43.58±0.06	11.63±0.15	56.47±0.09	47.16±0.15	37.80±0.06
Tent	56.67±0.14	13.58±0.19	45.02±0.13	11.52±0.33	57.25±0.05	48.59±0.16	38.77±0.10
PL	55.99±0.12	14.44±0.27	44.29±0.52	4.34±0.46	45.22±1.31	47.09±0.13	35.23±0.32
SHOT-IM	56.73±0.18	14.02±0.22	44.61±0.06	16.13±0.24	57.51±0.13	48.20±0.17	39.53±0.09
T3A	55.82±0.18	16.71±0.20	43.43±0.18	17.86±0.26	57.58±0.06	46.28±0.08	39.61±0.04
ETA	56.46±0.11	14.67±0.12	45.20±0.10	14.13±0.18	57.70±0.05	48.44±0.14	39.43±0.04
LAME	55.42±0.09	12.09±0.16	43.35±0.08	11.52±0.16	55.69±0.09	46.86±0.17	37.49±0.07
Ours	56.79±0.12	18.42±0.14	46.71±0.12	13.45±0.22	57.65±0.12	48.12±0.24	40.19±0.08

Table 8. Full results on PACS with ResNet50.

	A	C	P	S	Avg
ERM	82.50±1.83	80.80±0.33	94.05±0.30	80.99±1.29	84.59±0.40
BN	83.27±0.47	84.91±0.43	94.03±0.31	77.92±1.23	85.03±0.20
Tent	85.28±1.07	86.75±0.92	94.94±0.83	82.96±1.20	87.48±0.52
PL	83.96±1.63	84.15±2.91	93.82±1.74	78.99±2.64	85.23±1.70
SHOT-IM	84.31±0.63	85.74±0.56	94.04±0.67	77.91±0.94	85.50±0.31
T3A	84.07±0.68	82.37±0.92	95.02±0.27	82.72±1.06	86.04±0.24
ETA	83.27±0.47	84.91±0.43	94.03±0.31	77.92±1.24	85.04±0.20
LAME	84.97±0.77	85.50±0.55	95.04±0.23	80.97±1.09	86.62±0.22
Ours	87.68±0.84	88.78±0.63	96.17±0.37	85.01±1.52	89.41±0.51

Table 9. Full results on OfficeHome with ResNet50.

	A	C	P	R	Avg
ERM	60.71±0.88	55.74±0.79	76.18±0.65	76.83±0.41	67.37±0.06
BN	58.23±0.76	55.62±0.68	75.08±0.60	75.47±0.35	66.10±0.20
Tent	60.55±0.93	58.73±0.85	76.48±0.53	76.07±0.59	67.96±0.24
PL	59.14±0.93	57.29±0.62	76.24±0.69	75.85±0.28	67.13±0.17
SHOT-IM	59.20±0.76	57.54±0.58	76.53±0.44	76.27±0.37	67.39±0.16
T3A	61.23±1.00	56.69±1.11	77.95±0.43	77.31±0.22	68.29±0.21
ETA	58.38±0.80	55.78±0.69	75.17±0.56	75.53±0.34	66.21±0.19
LAME	58.67±0.70	55.58±0.53	75.09±0.65	75.40±0.31	66.19±0.20
Ours	62.32±0.52	57.45±0.71	77.48±0.45	77.45±0.38	68.67±0.14

Table 10. Full results on VLCS with ResNet50.

	C	L	S	V	Avg
ERM	94.91±0.32	65.20±2.35	66.52±1.65	69.41±3.38	74.01±1.32
BN	75.26±1.15	56.85±0.53	60.87±0.68	66.16±0.59	64.78±0.34
Tent	84.75±2.25	60.70±1.26	64.94±1.40	66.42±0.89	69.20±0.83
PL	86.75±4.25	61.19±2.22	63.36±3.33	62.77±2.84	68.52±1.79
SHOT-IM	76.54±1.27	55.90±1.20	61.26±0.71	67.24±0.86	65.23±0.32
T3A	97.06±0.30	63.96±0.89	67.14±0.52	67.75±0.31	73.98±0.32
ETA	75.27±1.13	56.85±0.52	60.89±0.67	66.16±0.58	64.79±0.34
LAME	96.25±0.55	61.39±0.40	70.25±0.67	67.89±0.38	73.94±0.14
Ours	97.40±0.41	64.89±0.64	68.05±0.45	68.12±0.43	74.52±0.27

Table 11. Full results on DomainNet with ResNet50.

	clipart	infograph	painting	quickdraw	real	sketch	Avg
ERM	64.76±0.06	22.11±0.11	51.77±0.18	13.84±0.14	64.66±0.21	54.04±0.27	45.20±0.09
BN	64.46±0.09	15.62±0.05	50.64±0.08	11.84±0.05	63.86±0.11	53.86±0.19	43.38±0.03
Tent	65.78±0.08	18.18±0.02	52.96±0.01	10.77±0.11	64.85±0.09	55.71±0.09	44.71±0.03
PL	64.96±0.05	19.00±0.03	50.30±0.27	4.21±0.68	54.40±0.55	54.20±0.15	41.18±0.13
SHOT-IM	65.62±0.05	18.73±0.21	52.41±0.08	19.01±0.24	66.47±0.12	55.54±0.08	46.30±0.07
T3A	64.77±0.05	22.10±0.09	50.89±0.15	19.41±0.18	65.85±0.06	53.96±0.17	46.16±0.03
ETA	65.11±0.09	19.37±0.19	52.69±0.11	18.24±0.33	65.92±0.10	55.48±0.16	46.13±0.08
LAME	64.18±0.12	15.64±0.07	50.54±0.04	11.77±0.05	63.46±0.08	53.65±0.18	43.20±0.03
Ours	66.12±0.08	24.12±0.12	52.82±0.10	18.17±0.08	68.45±0.12	56.72±0.12	47.73±0.05