ClusT3: Information Invariant Test-Time Training – Supplementary Material

Gustavo A. Vargas Hakim* , David Osowiechi*,
Mehrdad Noori , Milad Cheraghalikhani , Ali Bahri , Ismail Ben Ayed , and Christian Desrosiers

LIVIA, ÉTS Montréal, Canada

1. Results on CIFAR-10-C [2] dataset for corruption levels 1 to 4

As shown in Tables 1, 2, 3 and 4, ClusT3 performs well on the different corruptions at different levels. It achieves a higher accuracy than ResNet50 for all corruptions, and a higher mean accuracy than all other TTA/TTT aproaches. While TTT [5] yields competitive performance, our method achieves a mean accuracy improvement of at least 2% compared to this approach, on all corruption levels.

2. Hyperparameters search on VisDA-C

We perform the hyperparameter search to find an efficient configuration for VisDA-C. We evaluate to up to 20 iterations, using all the different individual layers, as well as combinations of them. Specifically, we tested the following settings:

- A single normal projector (one 1×1 convolution) in Table 5;
- Five normal projectors in Table 6;
- Ten normal projectors in Table 7;
- A single Large projector (two $1\ 1 \times 1\ 1$ convolutions with ReLU in between) in Table 8;
- Five Large projectors in Table 9.

As observed in these results, our ClusT3 method obtains significant improvements in different settings. For this dataset, the best accuracy is achieved using a single large projector applied to the second layer.

References

[1] Malik Boudiaf, Romain Mueller, Ismail Ben Ayed, and Luca Bertinetto. Parameter-free online test-time adaptation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8344–8353, 2022.

[2] Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. 2019.

[3] Yuejiang Liu, Parth Kothari, Bastien van Delft, Baptiste Bellot-Gurlet, Taylor Mordan, and Alexandre Alahi. Ttt++: When does self-supervised test-time training fail or thrive? *Neural Information Processing Systems (NeurIPS)*, 2021.

[4] Zachary Nado, Shreyas Padhy, D. Sculley, Alexander D'Amour, Balaji Lakshminarayanan, and Jasper Snoek. Evaluating prediction-time batch normalization for robustness under covariate shift. *arXiv:2006.10963 [cs, stat]*, Jan. 2021. arXiv: 2006.10963.

[5] Yu Sun, Xiaolong Wang, Zhuang Liu, John Miller, Alexei A. Efros, and Moritz Hardt. Test-time training with self-supervision for generalization under distribution shifts. In *International Conference on Machine Learning (ICML)*, 2020.

[6] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. 2021.

*Equal contribution. Correspondence to gustavo-adolfo.vargas-hakim.1@ens.etsmtl.ca, david.osowiechi.1@ens.etsmtl.ca

	ResNet50	LAME [1]	PTBN [4]	TENT [6]	TTT [5]	TTT++ [3]	ClusT3-H15
Gaussian Noise	28.02	26.08	61.39 ±0.10	61.19 ±0.26	70.63 ±0.04	78.70 ±4.28	**79.14 ±0.03**
Shot noise	38.33	37.13	66.57 ±0.06	66.2 ±0.18	75.18 ±0.04	80.12 ±0.12	**81.51 ±0.15**
Impulse Noise	46.12	45.01	63.56 ±0.20	62.98 ±0.19	65.91 ±0.04	70.64 ±0.53	**76.95 ±0.07**
Defocus blur	67.33	67.65	85.48 ±0.12	85.32 ±0.18	**91.95 ±0.02**	81.75 ±0.43	90.33 ±0.09
Glass blur	34.42	32.73	52.26 ±0.04	52.08 ±0.15	60.44 ±0.05	62.85 ±0.50	**71.09 ±0.17**
Motion blur	63.71	64.09	80.78 ±0.12	80.75 ±0.09	86.29 ±0.10	68.42 ±1.08	**87.87 ±0.11**
Zoom blur	61.27	61.99	83.33 ±0.11	83.28 ±0.10	**89.90 ±0.04**	70.74 ±2.05	88.86 ±0.04
Snow	72.15	72.13	73.25 ±0.16	73.17 ±0.25	81.25 ±0.02	52.43 ±0.56	**84.30 ±0.07**
Frost	62.27	61.70	73.41 ±0.22	73.54 ±0.16	83.83 ±0.04	52.80 ±2.67	**87.17 ±0.07**
Fog	81.86	81.94	83.88 ±0.06	83.81 ±0.09	**90.62 ±0.05**	41.75 ±0.09	90.03 ±0.02
Brightness	87.58	87.71	86.81 ±0.05	86.81 ±0.23	92.87 ±0.09	50.95 ±2.19	**92.99 ±0.06**
Contrast	68.62	68.85	84.16 ±0.09	84.23 ±0.29	90.94 ±0.07	45.28 ±0.55	**89.24 ±0.07**
Elastic transform	67.84	68.25	76.44 ±0.18	76.21 ±0.08	84.03 ±0.11	35.53 ±1.51	**86.74 ±0.04**
Pixelate	56.3	55.83	76.34 ±0.10	76.40 ±0.16	84.92 ±0.15	33.64 ±0.83	**87.93 ±0.03**
JPEG compression	70.62	70.37	69.64 ±0.03	69.54 ±0.05	76.46 ±0.04	28.01 ±1.75	**85.11 ±0.06**
Average	60.43	60.10	74.48	74.37	81.68	56.91	**85.28**

Table 1. Accuracy (%) on CIFAR-10-C dataset with Level 4 corruption for ClusT3-15 compared to ResNet50, LAME, PTBN, TENT, TTT, and TTT++.

	ResNet50	LAME [1]	PTBN [4]	TENT [6]	TTT [5]	TTT++ [3]	ClusT3-H15
Gaussian Noise	33.99	32.58	64.55 ±0.13	64.67 ±0.17	74.10 ±0.09	80.29 ±0.81	**81.55 ±0.09**
Shot noise	46.35	45.88	69.82 ±0.08	70.04 ±0.14	78.43 ±0.07	82.46 ±0.37	**84.12 ±0.02**
Impulse Noise	59.90	59.61	72.08 ±0.14	71.95 ±0.33	76.32 ±0.10	79.20 ±0.38	**83.75 ±0.01**
Defocus blur	79.29	79.58	87.62 ±0.17	87.39 ±0.05	**93.25 ±0.06**	87.68 ±0.38	91.74 ±0.07
Glass blur	47.29	46.44	63.29 ±0.11	63.26 ±0.21	72.09 ±0.11	72.52 ±0.56	**79.78 ±0.02**
Motion blur	63.42	63.72	81.13 ±0.13	80.99 ±0.08	86.48 ±0.09	69.59 ±1.38	**88.02 ±0.10**
Zoom blur	67.86	68.36	84.57 ±0.11	84.34 ±0.06	**91.00 ±0.02**	73.23 ±2.33	89.90 ±0.07
Snow	74.93	74.67	75.08 ±0.14	75.14 ±0.19	83.90 ±0.07	57.96 ±1.02	**86.22 ±0.07**
Frost	64.54	64.05	74.15 ±0.04	73.98 ±0.14	84.13 ±0.10	49.94 ±3.53	**87.37 ±0.07**
Fog	85.73	85.95	86.57 ±0.09	86.38 ±0.15	**92.19 ±0.08**	52.89 ±4.13	91.83 ±0.01
Brightness	88.93	88.75	87.50 ±0.19	87.44 ±0.01	**93.53 ±0.09**	57.96 ±1.32	93.31 ±0.04
Contrast	79.66	79.83	85.63 ±0.05	85.46 ±0.08	**91.85 ±0.09**	53.44 ±2.37	90.83 ±0.05
Elastic transform	75.67	75.79	82.72 ±0.14	82.56 ±0.15	**90.09 ±0.10**	36.49 ±3.72	89.33 ±0.11
Pixelate	74.83	75.07	82.17 ±0.14	81.91 ±0.13	89.30 ±0.10	33.41 ±3.02	**90.23 ±0.06**
JPEG compression	73.70	73.51	71.54 ±0.09	71.54 ±0.15	78.95 ±0.09	28.82 ±2.74	**86.55 ±0.06**
Average	67.74	67.59	77.89	77.80	85.04	61.06	**87.64**

Table 2. Accuracy (%) on CIFAR-10-C dataset with Level 3 corruption for ClusT3-15 compared to ResNet50, LAME, PTBN, TENT, TTT, and TTT++.

	ResNet50	LAME [1]	PTBN [4]	TENT [6]	TTT [5]	TTT++ [3]	ClusT3-H15
Gaussian Noise	50.53	49.99	71.31 ±0.16	71.43 ±0.08	81.18 ±0.11	85.41 ±2.26	**86.07 ±0.08**
Shot noise	69.27	69.47	78.97 ±0.19	79.02 ±0.17	87.54 ±0.10	88.79 ±0.44	**89.77 ±0.04**
Impulse Noise	68.57	68.69	77.09 ±0.13	77.03 ±0.15	82.20 ±0.13	84.27 ±0.29	**86.60 ±0.03**
Defocus blur	87.45	87.47	88.20 ±0.11	88.06 ±0.06	**93.67 ±0.06**	90.85 ±0.42	92.87 ±0.01
Glass blur	43.26	42.01	62.66 ±0.09	62.55 ±0.11	71.33 ±0.04	71.60 ±1.95	**78.81 ±0.11**
Motion blur	72.98	73.11	83.51 ±0.16	83.46 ±0.10	89.57 ±0.07	77.38 ±1.12	**89.78 ±0.13**
Zoom blur	74.89	75.24	85.81 ±0.21	85.79 ±0.05	**92.05 ±0.10**	80.30 ±1.45	90.82 ±0.04
Snow	71.11	70.74	74.73 ±0.11	74.69 ±0.22	82.96 ±0.08	68.56 ±1.36	**86.30 ±0.04**
Frost	76.67	76.56	79.54 ±0.15	79.41 ±0.27	87.67 ±0.03	63.66 ±3.39	**90.27 ±0.10**
Fog	88.51	88.47	87.62 ±0.10	87.60 ±0.17	**93.23 ±0.04**	64.26 ±3.37	93.07 ±0.04
Brightness	89.75	89.57	88.09 ±0.03	87.97 ±0.14	**93.69 ±0.08**	67.19 ±1.23	93.64 ±0.01
Contrast	84.58	84.79	86.19 ±0.17	86.41 ±0.04	**92.50 ±0.12**	62.90 ±1.93	92.00 ±0.01
Elastic transform	82.10	82.26	83.69 ±0.13	83.68 ±0.08	**90.98 ±0.12**	50.06 ±2.37	90.37 ±0.01
Pixelate	81.04	80.94	82.92 ±0.14	83.01 ±0.07	90.61 ±0.15	43.33 ±3.31	**91.28 ±0.09**
JPEG compression	76.06	76.04	73.63 ±0.02	73.56 ±0.13	81.37 ±0.11	28.26 ±2.78	**87.86 ±0.08**
Average	74.45	74.36	80.26	80.24	87.37	68.45	**89.30**

Table 3. Accuracy (%) on CIFAR-10-C dataset with Level 2 corruption for ClusT3-15 compared to ResNet50, LAME, PTBN, TENT, TTT, and TTT++.

	ResNet50	LAME [1]	PTBN [4]	TENT [6]	TTT [5]	TTT++ [3]	ClusT3-H15
Gaussian Noise	71.38	71.54	79.22 ±0.13	79.52 ±0.12	88.38 ±0.12	90.14 ±1.05	**90.35 ±0.05**
Shot noise	80.39	80.44	82.21 ±0.05	82.18 ±0.15	90.43 ±0.02	90.89 ±0.29	**91.42 ±0.02**
Impulse Noise	80.04	80.05	82.39 ±0.08	82.48 ±0.15	88.23 ±0.02	87.76 ±0.06	**90.51 ±0.06**
Defocus blur	90.17	89.96	88.28 ±0.04	88.26 ±0.15	**93.89 ±0.04**	91.51 ±0.48	93.72 ±0.09
Glass blur	40.96	39.79	63.19 ±0.05	63.22 ±0.15	71.12 ±0.07	72.12 ±2.13	**790.1 ±0.21**
Motion blur	82.78	82.75	85.99 ±0.09	85.89 ±0.08	**91.97 ±0.05**	84.11 ±0.91	91.50 ±0.13
Zoom blur	78.58	78.90	86.19 ±0.06	86.23 ±0.04	**92.21 ±0.08**	81.76 ±1.38	90.87 ±0.04
Snow	83.45	83.33	82.94 ±0.13	82.84 ±0.35	88.90 ±0.04	75.89 ±0.75	**90.33 ±0.02**
Frost	84.84	84.48	83.88 ±0.15	83.71 ±0.24	91.17 ±0.03	71.54 ±3.13	**92.19 ±0.06**
Fog	90.15	90.10	88.31 ±0.13	88.05 ±0.06	**93.71 ±0.09**	70.58 ±1.29	93.64 ±0.01
Brightness	90.35	90.19	88.28 ±0.09	88.35 ±0.25	**93.90 ±0.06**	64.40 ±2.69	93.83 ±0.05
Contrast	89.52	89.33	87.98 ±0.09	87.93 ±0.08	**93.61 ±0.05**	53.60 ±3.80	**93.61 ±0.03**
Elastic transform	82.46	82.57	83.29 ±0.17	83.28 ±0.27	**90.55 ±0.09**	39.92 ±1.52	90.33 ±0.06
Pixelate	87.27	87.15	85.79 ±0.12	85.81 ±0.17	92.24 ±0.01	36.04 ±3.47	**92.74 ±0.04**
JPEG compression	82.03	81.73	79.72 ±0.10	79.82 ±0.14	86.86 ±0.08	30.90 ±1.18	**90.90 ±0.01**
Average	80.96	80.82	83.17	83.17	89.81	69.41	**91.00**

Table 4. Accuracy (%) on CIFAR-10-C dataset with Level 1 corruption for ClusT3-15 compared to ResNet50, LAME, PTBN, TENT, TTT, and TTT++.

	Layers							
Iterations	**1**	**2**	**3**	**4**	**1, 2**	**2, 3**	**3, 4**	**All**
No adaptation	45.31	45.57	45.67	47.09	45.66	44.27	38.17	42.89
1	49.07	48.73	48.96	52.96	50.13	47.85	49.35	48.16
3	52.53	54.31	53.99	56.23	52.18	54.7	55.57	54.23
10	57.67	58.19	58.27	58.79	57.78	58.74	58.31	57.56
20	56.84	**59.82**	56.61	57.34	57.41	57.63	56.31	55.87

Table 5. Accuracy (%) values on VisDA-C with 1 normal projector on different layers.

	Layers							
Iterations	**1**	**2**	**3**	**4**	**1, 2**	**2, 3**	**3, 4**	**All**
No adaptation	45.31	45.57	45.67	47.09	45.66	44.27	38.17	42.89
1	49.57	49.84	50.28	50.69	49.58	47.38	41.39	47.72
3	54.39	54.51	55.73	54.47	53.99	52.11	47.15	53.75
10	58.77	58.85	60.28	58.31	58.47	57.34	55.94	57.89
20	57.66	**61.02**	58.43	57.17	56.83	56.52	56.31	55.87

Table 6. Accuracy (%) values on VisDA-C with 5 normal projectors on different layers.

	Layers							
Iterations	**1**	**2**	**3**	**4**	**1, 2**	**2, 3**	**3, 4**	**All**
No adaptation	46.51	46.02	44.69	45.97	44.22	46.02	43.2	41.05
1	49.28	50.31	49.38	50.38	48.43	50.35	46.19	44.25
3	54.79	55.69	55.31	55.19	53.34	55.71	51.35	44.19
10	61.13	60.93	60.61	59.6	59.62	60.78	59.09	59.06
15	**61.53**	61.16	60.59	59.97	60.37	60.55	59.82	59.97
20	61.33	60.86	59.92	59.8	60.22	59.89	59.73	58.98

Table 7. Accuracy (%) values on VisDA-C with 10 normal projectors on different layers.

	Layers							
Iterations	**1**	**2**	**3**	**4**	**1, 2**	**2, 3**	**3, 4**	**All**
No adaptation	43.91	46.41	46	42.46	46.54	45.45	44.27	44.09
1	48.28	51.79	49.82	45.96	51.06	49.23	49.41	49.36
3	54.23	56.72	54.24	51.62	56.6	54.36	54.89	49.43
10	60.04	61.72	59.16	59.56	60.93	60.64	61.12	60.18
15	60.25	**61.93**	59.44	60.16	60.81	60.91	61.64	60.27
20	59.98	61.57	59.14	59.88	60.31	60.72	61.16	59.92

Table 8. Accuracy (%) values on VisDA-C with 1 Large projector on different layers.

	Layers							
Iterations	**1**	**2**	**3**	**4**	**1, 2**	**2, 3**	**3, 4**	**All**
No adaptation	46.57	44.66	46.01	43.86	45.21	46.37	46.58	46.81
1	50.67	47.68	50.16	48.61	49.70	49.46	49.46	52.34
3	56.18	52.77	54.90	54.27	54.65	52.88	53.43	52.29
10	61.45	61.03	59.87	60.51	59.57	59.95	58.12	**61.69**
15	61.48	61.54	60.40	60.96	60.16	61.44	58.83	61.51
20	60.91	60.90	59.81	60.49	59.83	59.30	58.69	60.66

Table 9. Accuracy (%) values on VisDA-C with 5 Large projectors on different layers.