NC-TTT: A Noise Constrastive Approach for Test-Time Training - Supplementary Material

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

ÉTS Montréal, Canada

A. Deriving the posterior of Equation (6)

We start with the definition of the posterior in Eq. (5):

$$p(y_s = 1 \,|\, \widetilde{\mathbf{z}}_{i,m}) \;=\; \frac{\sigma_s^{-D} \exp\!\left(-\frac{1}{2\sigma_s^2}\|\boldsymbol{\epsilon}\|^2\right)}{\sigma_s^{-D} \exp\!\left(-\frac{1}{2\sigma_s^2}\|\boldsymbol{\epsilon}\|^2\right) + \sigma_o^{-D} \exp\!\left(-\frac{1}{2\sigma_o^2}\|\boldsymbol{\epsilon}\|^2\right)},$$

where $\boldsymbol{\epsilon} = \widetilde{\mathbf{z}}_i - \mathbf{z}_i$. Multiplying the numerator and denominator by $\sigma_s^D \exp\big(\frac{1}{2\sigma_s^2}\|\boldsymbol{\epsilon}\|^2\big)$ gives

$$p(y_s = 1 \,|\, \widetilde{\mathbf{z}}_{i,m}) \;=\; \frac{1}{1 \,+\, \big(\sigma_s/\sigma_o\big)^D \exp\!\big(-\tfrac{1}{2}(1/\sigma_s^2 - 1/\sigma_o^2)\|\boldsymbol{\epsilon}\|^2\big)}$$

We then use the following equality

$$\begin{aligned}\big(\sigma_s/\sigma_o\big)^D \;&=\; \exp\Big(\log\big(\sigma_s/\sigma_o\big)^D\Big)\\ &=\; \exp\big(D\log\big(\sigma_s/\sigma_o\big)\big)\end{aligned}$$

to obtain

$$p(y_s = 1 \,|\, \widetilde{\mathbf{z}}_{i,m}) \;=\; \frac{1}{1 \,+\, \exp\!\big(-\tfrac{1}{2}(1/\sigma_s^2 - 1/\sigma_o^2)\|\boldsymbol{\epsilon}\|^2 \,+\, D\log\big(\sigma_s/\sigma_o\big)\big)}$$

Last, since $\log\big(\sigma_s/\sigma_o\big) \;=\; -\log\big(\sigma_o/\sigma_s\big)$, we finally get $p(y_s = 1 \,|\, \widetilde{\mathbf{z}}) = 1/\big(1+\exp(-u)\big)$ with

$$u \;=\; \frac{1}{2}\left(\frac{1}{\sigma_o^2} - \frac{1}{\sigma_s^2}\right)\|\boldsymbol{\epsilon}\|^2 \,+\, D\log\left(\frac{\sigma_o}{\sigma_s}\right) \qquad \square$$

B. Results on different levels of CIFAR-10-C corruptions

We evaluate NC-TTT on the remaining severity levels of CIFAR-10-C (see Tables 1-4). Accuracy decreases on all methods as the severity augments, but NC-TTT outperforms the closest competitors from the *state-of-the-art* on the majority of the corruptions and in average for the whole dataset. To be more precise, when considering NC-TTT at a lower severity level (refer to Table 4), it demonstrates superior performance across all corruptions, with the exception of *Gaussian Noise* and *JPEG compression*. This is noteworthy, given that Gaussian noise is introduced to features during training.

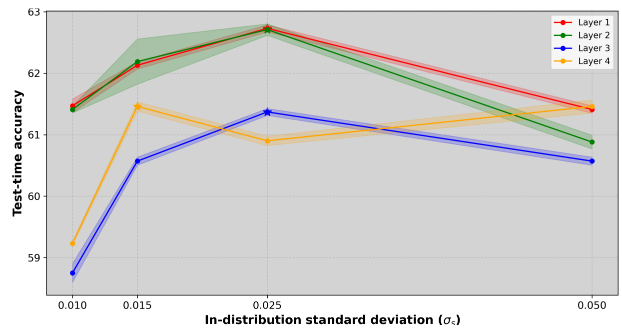

Figure 1. Test-time accuracy on different layer blocks with different in-distribution standard deviation.

C. Hyperparameter search on VisDA-C

In order to choose the best configuration for VisDA-C, we performed a hyperparameter search considering the four different layer blocks from ResNet50 as well as four different in-distribution noise standard deviation values (i.e., 0.01, 0.015, 0.025, 0.05). The results are obtained across three executions per combination. We show in Fig.1 that the best performance can be generally obtained on the first layer of the network, consistent with previous results in the field [2, 3, 5, 6]. Furthermore, an in-distribution standard deviation $\sigma_s = 0.025$ is found to perform the best across all the different layers.

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

	ResNet50	LAME [1]	PTBN [4]	TENT [7]	TTT [6]	TTT++ [3]	ClusT3 [2]	NC-TTT (ours)
Gaussian Noise	28.02	26.22 ±0.21	61.39 ±0.10	61.19 ±0.26	70.63 ±0.04	78.70 ±4.28	**79.14 ±0.03**	78.09 ±0.02
Shot noise	38.33	37.06 ±0.17	66.57 ±0.06	66.2 ±0.18	75.18 ±0.04	80.12 ±0.12	**81.51 ±0.15**	81.33 ±0.10
Impulse Noise	46.12	45.03 ±0.17	63.56 ±0.20	62.98 ±0.19	65.91 ±0.04	70.64 ±0.53	**76.95 ±0.07**	75.52 ±0.08
Defocus blur	67.33	67.70 ±0.07	85.48 ±0.12	85.32 ±0.18	**91.95 ±0.02**	81.75 ±0.43	90.33 ±0.09	91.91 ±0.05
Glass blur	34.42	32.63 ±0.09	52.26 ±0.04	52.08 ±0.15	60.44 ±0.05	62.85 ±0.50	**71.09 ±0.17**	69.95 ±0.09
Motion blur	63.71	64.00 ±0.01	80.78 ±0.12	80.75 ±0.09	86.29 ±0.10	68.42 ±1.08	87.87 ±0.11	**89.02 ±0.10**
Zoom blur	61.27	62.12 ±0.21	83.33 ±0.11	83.28 ±0.10	89.90 ±0.04	70.74 ±2.05	88.86 ±0.04	**90.96 ±0.05**
Snow	72.15	72.18 ±0.04	73.25 ±0.16	73.17 ±0.25	81.25 ±0.02	52.43 ±0.56	84.30 ±0.07	**85.36 ±0.06**
Frost	62.27	61.72 ±0.06	73.41 ±0.22	73.54 ±0.16	83.83 ±0.04	52.80 ±2.67	87.17 ±0.07	**88.08 ±0.02**
Fog	81.86	82.07 ±0.03	83.88 ±0.06	83.81 ±0.09	90.62 ±0.05	41.75 ±0.09	90.03 ±0.02	**92.07 ±0.07**
Brightness	87.58	87.64 ±0.08	86.81 ±0.05	86.81 ±0.23	92.87 ±0.09	50.95 ±2.19	**92.99 ±0.06**	**94.41 ±0.03**
Contrast	68.62	69.02 ±0.11	84.16 ±0.09	84.23 ±0.29	90.94 ±0.07	45.28 ±0.55	**89.24 ±0.07**	**91.52 ±0.03**
Elastic transform	67.84	68.32 ±0.03	76.44 ±0.18	76.21 ±0.08	84.03 ±0.11	35.53 ±1.51	**86.74 ±0.04**	86.39 ±0.04
Pixelate	56.3	55.94 ±0.08	76.34 ±0.10	76.40 ±0.16	84.92 ±0.15	33.64 ±0.83	87.93 ±0.03	**88.06 ±0.20**
JPEG compression	70.62	70.44 ±0.18	69.64 ±0.03	69.54 ±0.05	76.46 ±0.04	28.01 ±1.75	**85.11 ±0.06**	82.73 ±0.07
Average	60.43	60.14	74.48	74.37	81.68	56.91	85.28	**85.69**

Table 1. Accuracy (%) on CIFAR-10-C dataset with Level 4 corruption for NC-TTT compared to *state-of-the-art*.

	ResNet50	LAME [1]	PTBN [4]	TENT [7]	TTT [6]	TTT++ [3]	ClusT3 [2]	NC-TTT (ours)
Gaussian Noise	33.99	32.37 ±0.16	64.55 ±0.13	64.67 ±0.17	74.10 ±0.09	80.29 ±0.81	**81.55 ±0.09**	81.09 ±0.09
Shot noise	46.35	45.83 ±0.14	69.82 ±0.08	70.04 ±0.14	78.43 ±0.07	82.46 ±0.37	**84.12 ±0.02**	84.03 ±0.09
Impulse Noise	59.90	59.43 ±0.13	72.08 ±0.14	71.95 ±0.33	76.32 ±0.10	79.20 ±0.38	**83.75 ±0.01**	83.73 ±0.05
Defocus blur	79.29	79.67 ±0.11	87.62 ±0.17	87.39 ±0.05	93.25 ±0.06	87.68 ±0.38	91.74 ±0.07	**93.51 ±0.08**
Glass blur	47.29	46.36 ±0.10	63.29 ±0.11	63.26 ±0.21	72.09 ±0.11	72.52 ±0.56	**79.78 ±0.02**	79.25 ±0.08
Motion blur	63.42	63.72 ±0.07	81.13 ±0.13	80.99 ±0.08	86.48 ±0.09	69.59 ±1.38	88.02 ±0.10	**89.02 ±0.02**
Zoom blur	67.86	68.23 ±0.08	84.57 ±0.11	84.34 ±0.06	91.00 ±0.02	73.23 ±2.33	89.90 ±0.07	**91.86 ±0.10**
Snow	74.93	74.78 ±0.05	75.08 ±0.14	75.14 ±0.19	83.90 ±0.07	57.96 ±1.02	86.22 ±0.07	**87.17 ±0.05**
Frost	64.54	64.16 ±0.08	74.15 ±0.04	73.98 ±0.14	84.13 ±0.10	49.94 ±3.53	**87.37 ±0.07**	88.05 ±0.05
Fog	85.73	85.98 ±0.16	86.57 ±0.09	86.38 ±0.15	92.19 ±0.08	52.89 ±4.13	91.83 ±0.01	**93.55 ±0.01**
Brightness	88.93	88.67 ±0.08	87.50 ±0.19	87.44 ±0.01	93.53 ±0.09	57.96 ±1.32	93.31 ±0.04	**94.61 ±0.01**
Contrast	79.66	79.99 ±0.05	85.63 ±0.05	85.46 ±0.08	91.85 ±0.09	53.44 ±2.37	90.83 ±0.05	**92.54 ±0.09**
Elastic transform	75.67	75.96 ±0.14	82.72 ±0.14	82.56 ±0.15	90.09 ±0.10	36.49 ±3.72	89.33 ±0.11	**90.95 ±0.03**
Pixelate	74.83	75.12 ±0.04	82.17 ±0.14	81.91 ±0.13	89.30 ±0.10	33.41 ±3.02	90.23 ±0.06	**91.44 ±0.05**
JPEG compression	73.70	73.66 ±0.16	71.54 ±0.09	71.54 ±0.15	78.95 ±0.09	28.82 ±2.74	**86.55 ±0.06**	85.10 ±0.01
Average	67.74	67.60	77.89	77.80	85.04	61.06	87.64	**88.39**

Table 2. Accuracy (%) on CIFAR-10-C dataset with Level 3 corruption for NC-TTT compared to *state-of-the-art*.

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, 3

[2] Gustavo A. Vargas Hakim, David Osowiechi, Mehrdad Noori, Milad Cheraghalikhani, Ali Bahri, Ismail Ben Ayed, and Christian Desrosiers. ClusT3: Information Invariant Test-Time Training. pages 6136–6145, 2023. 1, 2, 3

[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. 1, 2, 3

[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]*, 2021. arXiv: 2006.10963. 2, 3

[5] David Osowiechi, Gustavo A. Vargas Hakim, Mehrdad Noori, Milad Cheraghalikhani, Ismail Ayed, and Christian Desrosiers. Tttflow: Unsupervised test-time training with normalizing flow. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 2125–2126, Los Alamitos, CA, USA, 2023. IEEE Computer Society. 1

[6] 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.

	ResNet50	LAME [1]	PTBN [4]	TENT [7]	TTT [6]	TTT++ [3]	ClusT3 [2]	NC-TTT (ours)
Gaussian Noise	50.53	50.02 ±0.24	71.31 ±0.16	71.43 ±0.08	81.18 ±0.11	85.41 ±2.26	**86.07 ±0.08**	85.37 ±0.08
Shot noise	69.27	69.47 ±0.22	78.97 ±0.19	79.02 ±0.17	87.54 ±0.10	88.79 ±0.44	89.77 ±0.04	**90.15 ±0.09**
Impulse Noise	68.57	68.68 ±0.08	77.09 ±0.13	77.03 ±0.15	82.20 ±0.13	84.27 ±0.29	86.60 ±0.03	**86.89 ±0.11**
Defocus blur	87.45	87.46 ±0.14	88.20 ±0.11	88.06 ±0.06	93.67 ±0.06	90.85 ±0.42	92.87 ±0.01	**94.32 ±0.01**
Glass blur	43.26	42.04 ±0.19	62.66 ±0.09	62.55 ±0.11	71.33 ±0.04	71.60 ±1.95	78.81 ±0.11	**79.86 ±0.06**
Motion blur	72.98	73.14 ±0.06	83.51 ±0.16	83.46 ±0.10	89.57 ±0.07	77.38 ±1.12	89.78 ±0.13	**91.23 ±0.04**
Zoom blur	74.89	75.23 ±0.18	85.81 ±0.21	85.79 ±0.05	92.05 ±0.10	80.30 ±1.45	90.82 ±0.04	**92.76 ±0.10**
Snow	71.11	70.78 ±0.12	74.73 ±0.11	74.69 ±0.22	82.96 ±0.08	68.56 ±1.36	86.30 ±0.04	**87.55 ±0.05**
Frost	76.67	76.46 ±0.02	79.54 ±0.15	79.41 ±0.27	87.67 ±0.03	63.66 ±3.39	90.27 ±0.10	**91.09 ±0.03**
Fog	88.51	88.55 ±0.08	87.62 ±0.10	87.60 ±0.17	93.23 ±0.04	64.26 ±3.37	93.07 ±0.04	**94.46 ±0.02**
Brightness	89.75	89.52 ±0.01	88.09 ±0.03	87.97 ±0.14	93.69 ±0.08	67.19 ±1.23	93.64 ±0.01	**94.97 ±0.04**
Contrast	84.58	84.87 ±0.07	86.19 ±0.17	86.41 ±0.04	92.50 ±0.12	62.90 ±1.93	92.00 ±0.01	**3.50 ±0.08**
Elastic transform	82.10	82.17 ±0.10	83.69 ±0.13	83.68 ±0.08	90.98 ±0.12	50.06 ±2.37	90.37 ±0.01	**91.60 ±0.05**
Pixelate	81.04	80.96 ±0.13	82.92 ±0.14	83.01 ±0.07	90.61 ±0.15	43.33 ±3.31	91.28 ±0.09	**92.46 ±0.06**
JPEG compression	76.06	75.92 ±0.09	73.63 ±0.02	73.56 ±0.13	81.37 ±0.11	28.26 ±2.78	**87.86 ±0.08**	86.27 ±0.03
Average	74.45	74.35	80.26	80.24	87.37	68.45	89.30	**90.17**

Table 3. Accuracy (%) on CIFAR-10-C dataset with Level 2 corruption for NC-TTT compared to *state-of-the-art*.

	ResNet50	LAME [1]	PTBN [4]	TENT [7]	TTT [6]	TTT++ [3]	ClusT3 [2]	NC-TTT (ours)
Gaussian Noise	71.38	71.35 ±0.05	79.22 ±0.13	79.52 ±0.12	88.38 ±0.12	90.14 ±1.05	**90.35 ±0.05**	90.29 ±0.01
Shot noise	80.39	80.32 ±0.07	82.21 ±0.05	82.18 ±0.15	90.43 ±0.02	90.89 ±0.29	91.42 ±0.02	**92.25 ±0.02**
Impulse Noise	80.04	79.98 ±0.09	82.39 ±0.08	82.48 ±0.15	88.23 ±0.02	87.76 ±0.06	90.51 ±0.06	**91.07 ±0.05**
Defocus blur	90.17	89.9 ±0.06	88.28 ±0.04	88.26 ±0.15	93.89 ±0.04	91.51 ±0.48	93.72 ±0.09	**95.12 ±0.02**
Glass blur	40.96	39.87 ±0.16	63.19 ±0.05	63.22 ±0.15	71.12 ±0.07	72.12 ±2.13	79.01 ±0.21	**79.78 ±0.05**
Motion blur	82.78	82.81 ±0.11	85.99 ±0.09	85.89 ±0.08	91.97 ±0.05	84.11 ±0.91	91.50 ±0.13	**93.15 ±0.07**
Zoom blur	78.58	79.03 ±0.06	86.19 ±0.06	86.23 ±0.04	92.21 ±0.08	81.76 ±1.38	90.87 ±0.04	**92.60 ±0.07**
Snow	83.45	83.32 ±0.11	82.94 ±0.13	82.84 ±0.35	88.90 ±0.04	75.89 ±0.75	90.33 ±0.02	**91.57 ±0.03**
Frost	84.84	84.44 ±0.10	83.88 ±0.15	83.71 ±0.24	91.17 ±0.03	71.54 ±3.13	92.19 ±0.06	**93.16 ±0.08**
Fog	90.15	90.05 ±0.05	88.31 ±0.13	88.05 ±0.06	93.71 ±0.09	70.58 ±1.29	93.64 ±0.01	**95.11 ±0.03**
Brightness	90.35	90.24 ±0.06	88.28 ±0.09	88.35 ±0.25	93.90 ±0.06	64.40 ±2.69	93.83 ±0.05	**95.28 ±0.02**
Contrast	89.52	89.57 ±0.07	87.98 ±0.09	87.93 ±0.08	93.61 ±0.05	53.60 ±3.80	93.61 ±0.03	**94.95 ±0.06**
Elastic transform	82.46	82.72 ±0.06	83.29 ±0.17	83.28 ±0.27	90.55 ±0.09	39.92 ±1.52	90.33 ±0.06	**91.62 ±0.07**
Pixelate	87.27	87.18 ±0.08	85.79 ±0.12	85.81 ±0.17	92.24 ±0.01	36.04 ±3.47	92.74 ±0.04	**93.84 ±0.03**
JPEG compression	82.03	81.66 ±0.07	79.72 ±0.10	79.82 ±0.14	86.86 ±0.08	30.90 ±1.18	**90.90 ±0.01**	90.18 ±0.05
Average	80.96	80.83	83.17	83.17	89.81	69.41	91.00	**92.00**

Table 4. Accuracy (%) on CIFAR-10-C dataset with Level 1 corruption for NC-TTT compared to *state-of-the-art*.

1, 2, 3

[7] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. 2021. 2, 3