ReC-TTT: Contrastive Feature Reconstruction for Test-Time Training Supplementary Material

Marco Colussi[1]* Sergio Mascetti[1] Jose Dolz[2] Christian Desrosiers[2]

[1]Università degli studi di Milano [2]ÉTS Montréal

{marco.colussi, sergio.mascetti}@unimi.it {jose.dolz, christian.desrosiers}@etsmtl.ca

1. Extended results

For CIFAR-100C, TinyImageNet-C and VISDA our model was compared with the same state-of-the-art approaches except TTT++ where the results were not reproducible nor available: ResNet50 [3], PTBN [4], TENT [8], TIPI [5], ClusT3 [2] and NC-TTT [6]. As per previous experiments TTA methods were evaluated on the same pretrained ResNet50, while TTT approaches were trained using the same ResNet50 base architecture and the same training strategy.

1.1. VisDA

Table 1 reports the detailed results on the VisDA dataset. *ReC-TTT* outperforms most approaches on average, with a notable increase compared to the ResNet50 baseline without adaptation ($+25.81$). On $train \rightarrow val$ and $train \rightarrow test$, NC-TTT performs better than *ReC-TTT* ($\approx +1\%$ on average). Moreover, the results demonstrate that TTT methods show greater robustness on complex datasets, such as VISDA, compared to methods like Source, PTBN, and TENT, which are more competitive on the CIFAR datasets. This performance difference may be attributed to the reconstruction task's ability to capture more generalizable features, while simpler approaches struggle to detect more subtle domain shifts.

Table 1. Performance comparison with state-of-the-art on VISDA dataset (%).

	VISDA $train \rightarrow val$	VISDA $train \rightarrow test$	Average
ResNet50	35.01	36.58	35.80
PTBN	54.53	53.63	54.08
TENT	58.13	57.04	57.59
TIPI	60.22	62.26	61.24
ClusT3	60.89	61.33	61.11
NC-TTT	**62.49**	**62.57**	**62.53**
ReC-TTT	62.06	61.12	61.59

*Corresponding author.

1.2. CIFAR-100C

Table 2 shows in detail the results and the comparison with state-of-the-art approaches on all the perturbations of CIFAR-100C. *ReC-TTT* the best results, demonstrating a 30% increase in AUROC after adaptation compared to the baseline. This improvement surpasses the most recent state-of-the-art approaches as ClusT3 and NC-TTT by 3%.

1.2.1 Number of adaptation iterations

Similarly to what was identified in previous studies [2, 6, 7] and was confirmed for CIFAR-10C, also in the case of CIFAR-100C the best results are obtained after 20 adaptation iterations, while for some perturbation the same results can be obtained also with less interaction, after 20 the results tend to remain invariant for all the different perturbations. Figure 1 shows for all the corruption of CIFAR-100C the results obtained at different iterations.

1.3. TinyImagenet-C

Table 3 reports the results obtained on TinyImagenet-C, a dataset of 10.000 images with the same 15 corruptions described for CIFAR10-C and CIFAR100-C, but with 200 classes. *ReC-TTT* outperforms all the other methods also on this dataset, with a 2.46% improvement compared to NC-TTT, the second-best-performing model.

2. On the contrastive loss performances

To show the impact of our contrastive approach we implemented a TTT method based on the *SimSiam* [1] framework. This solution only compares the features at the bottleneck level and is based on a single encoder, followed by a projection head and a predictor. The model was trained with the Cross-Entropy loss and the *SimSiam* loss as auxiliary task. As reported in the paper presenting the *SimSiam* technique [1], the loss is computed as the negative cosine similarity between *i*) the features of the projector (f_E) extracted by the original image and *ii*) the features of the predictor (f_P) of the augmented version of the image with a

Corruption Type	ResNet50	PTBN	TENT	TIPI	ClusT3	NC-TTT	*ReC-TTT*
Gaussian Noise	13.23	42.30	51.35	48.88	**52.79**	46.03	48.12
Shot Noise	15.46	43.30	52.63	50.61	**52.91**	47.04	50.43
Impulse Noise	7.89	37.41	45.39	43.80	**45.54**	41.53	45.29
Defocus Blur	27.36	67.46	69.44	68.72	66.66	67.00	**71.21**
Glass Blur	21.18	46.44	**51.01**	50.93	50.76	48.08	49.94
Motion Blur	38.18	64.21	67.27	66.63	62.92	64.31	**68.86**
Zoom Blur	32.81	66.68	69.33	68.84	65.42	66.24	**69.91**
Snow	44.85	55.52	**60.47**	59.51	56.65	58.70	60.21
Frost	31.56	54.76	58.35	57.90	56.91	58.55	**60.16**
Fog	32.79	56.77	**62.29**	61.12	53.95	57.73	62.22
Brightness	66.13	68.97	71.40	71.00	66.78	71.36	**73.47**
Contrast	11.87	63.47	65.63	65.17	56.46	61.53	**67.06**
Elastic Transform	48.87	57.93	60.07	59.94	59.07	60.25	**62.37**
Pixelate	26.70	59.75	**64.06**	63.56	62.26	61.17	63.61
JPEG Compression	48.88	52.45	57.84	57.79	**59.34**	55.69	57.05
Average	31.19	55.83	60.44	59.63	57.89	57.68	**60.66**

Table 2. Performance comparison with state-of-the-art on CIFAR-100C perturbations (%).

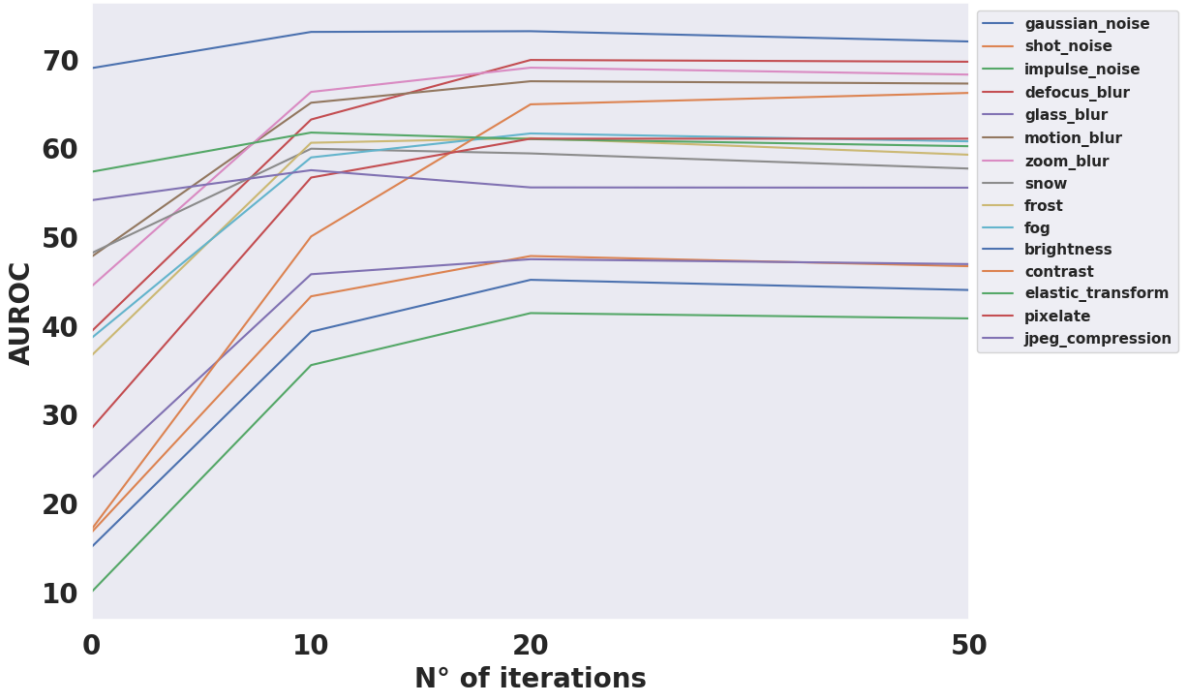

Figure 1. Performance (AUROC) reached by our method with different numbers of adaptation iterations on CIFAR-100C.

stop gradient on the predictor features. To have a fair comparison with *ReC-TTT*, we also used horizontal flip as augmentation. During the adaptation phase, we adopted the same auxiliary loss to adapt the encoder features for a total of 20 iterations.

Table 4 shows that the *SimSiam* contrastive learning approach, although achieving some good adaptation performances, does not achieve the same results as *ReC-TTT*. A possible reason for this result is that *SimSiam* cannot fully capture the domain shift, which is hidden in the whole representation and not only at the bottleneck level. This is the

Corruption Type	ResNet50	PTBN	TENT	TIPI	ClusT3	NC-TTT	*ReC-TTT*
Gaussian Noise	13.20	30.46	31.03	32.22	32.65	31.92	**34.87**
Shot Noise	16.28	32.26	33.07	34.27	34.72	34.47	**36.60**
Impulse Noise	7.49	20.80	21.87	23.04	22.78	22.78	**26.09**
Defocus Blur	16.71	33.09	**34.20**	31.98	29.08	25.28	31.09
Glass Blur	7.42	15.97	16.88	17.60	16.26	15.67	**19.59**
Motion Blur	27.71	43.09	44.40	43.54	43.92	43.39	**45.55**
Zoom Blur	20.98	39.76	40.89	40.01	41.17	40.46	**42.53**
Snow	31.00	36.94	37.39	38.18	42.97	**43.46**	40.33
Frost	36.28	39.29	40.21	41.43	45.32	**45.51**	44.59
Fog	16.40	31.51	32.52	32.82	**37.85**	37.68	33.08
Brightness	36.48	44.70	45.09	46.39	**51.19**	50.62	48.53
Contrast	2.59	12.22	**12.91**	10.71	2.27	2.27	8.32
Elastic Transform	28.93	39.42	39.83	40.68	41.60	41.47	**44.91**
Pixelate	37.00	47.78	48.50	48.95	37.00	39.31	**52.96**
JPEG Compression	47.04	47.78	40.88	50.21	50.57	50.91	**53.32**
Average	23.03	34.47	35.15	35.47	35.32	35.03	**37.49**

Table 3. Performance comparison with state-of-the-art on TinyImageNet-C perturbations (%).

	Impulse Noise	Brightness	Pixelate	Average
SimSiam	56.40	82.92	68.69	69.77
ReC-TTT	69.28	94.03	82.13	**82.82**

Table 4. **On the contrastive loss.** Qualitative results using SimSiam contrastive approach on CIFAR-10C (%).

main difference with *ReC-TTT* that instead compares features at different layers.

References

[1] Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 15750–15758, 2021. 1

[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. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6136–6145, 2023. 1

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1

[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 preprint arXiv:2006.10963*, 2020. 1

[5] A Tuan Nguyen, Thanh Nguyen-Tang, Ser-Nam Lim, and Philip HS Torr. Tipi: Test time adaptation with transformation invariance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24162–24171, 2023. 1

[6] David Osowiechi, Gustavo A Vargas Hakim, Mehrdad Noori, Milad Cheraghalikhani, Ali Bahri, Moslem Yazdanpanah, Ismail Ben Ayed, and Christian Desrosiers. Nc-ttt: A noise constrastive approach for test-time training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6078–6086, 2024. 1

[7] David Osowiechi, Gustavo A Vargas Hakim, Mehrdad Noori, Milad Cheraghalikhani, Ismail Ben Ayed, and Christian Desrosiers. Tttflow: Unsupervised test-time training with normalizing flow. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2126–2134, 2023. 1

[8] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. *arXiv preprint arXiv:2006.10726*, 2020. 1