Continual Test-Time Domain Adaptation

Qin Wang[1] Olga Fink[1,3*] Luc Van Gool[1,4] Dengxin Dai[2]
[1]ETH Zurich, Switzerland [2]MPI for Informatics, Germany [3]EPFL, Switzerland [4]KU Lueven, Belgium
{qin.wang,vangool,dai}@vision.ee.ethz.ch olga.fink@epfl.ch

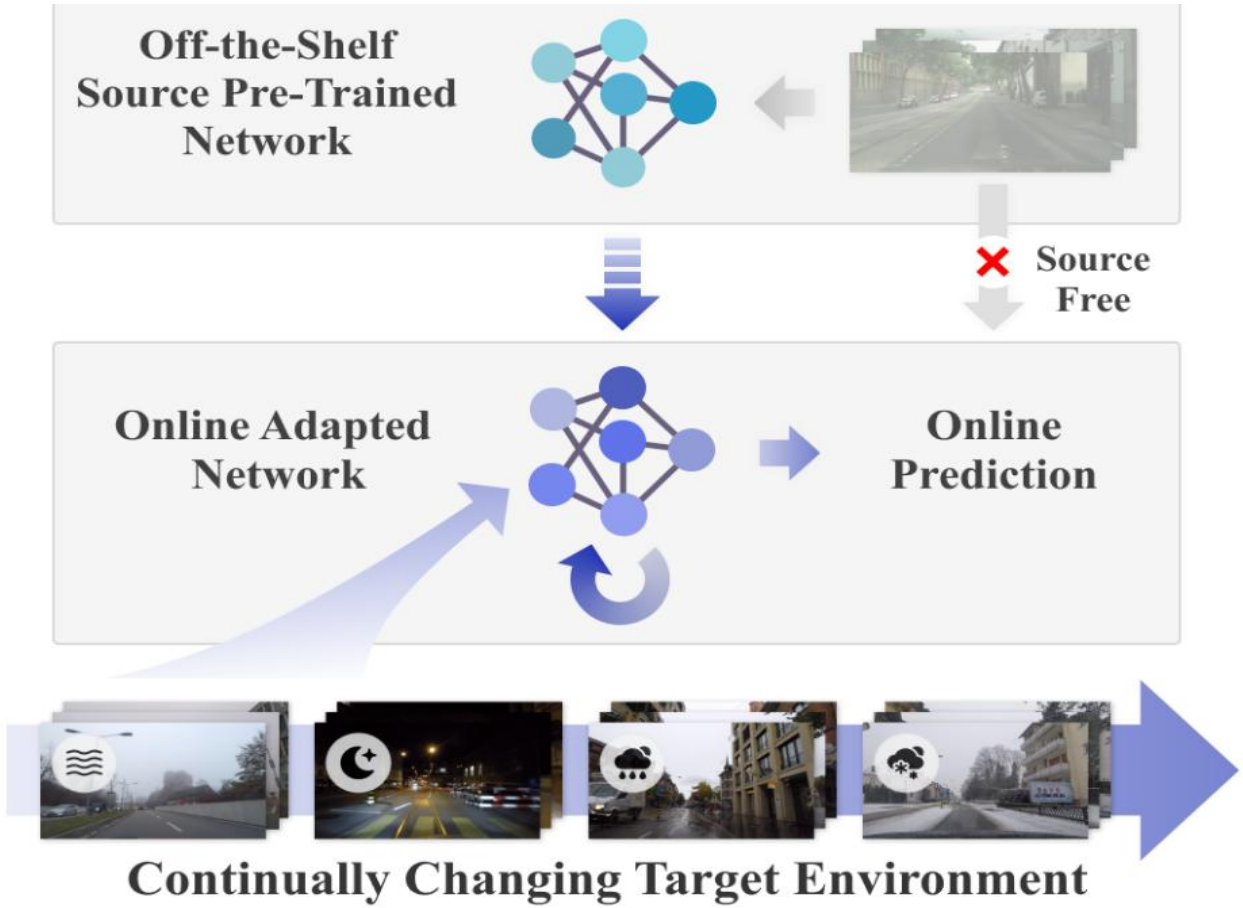
CVPR 2022

Background

Table 1. The difference between our proposed continual test-time adaptation and related adaptation settings.

Setting	Data		Learning	
	Source	Target	Train stage	Test stage
standard domain adaptation	Yes	stationary	Yes	No
standard test-time training [54]	Yes	stationary	Yes (aux task)	Yes
fully test-time adaptation [61]	No	stationary	No (pre-trained)	Yes
continual test-time adaptation	No	continually changing	No (pre-trained)	Yes

Background



- Existing methods, which are mostly based on self-training and entropy regularization, can suffer from these non-stationary environments.
- Due to the distribution shift over time in the target domain, pseudo-labels become unreliable.
- The noisy pseudolabels can further lead to error accumulation and catastrophic forgetting.

Methods

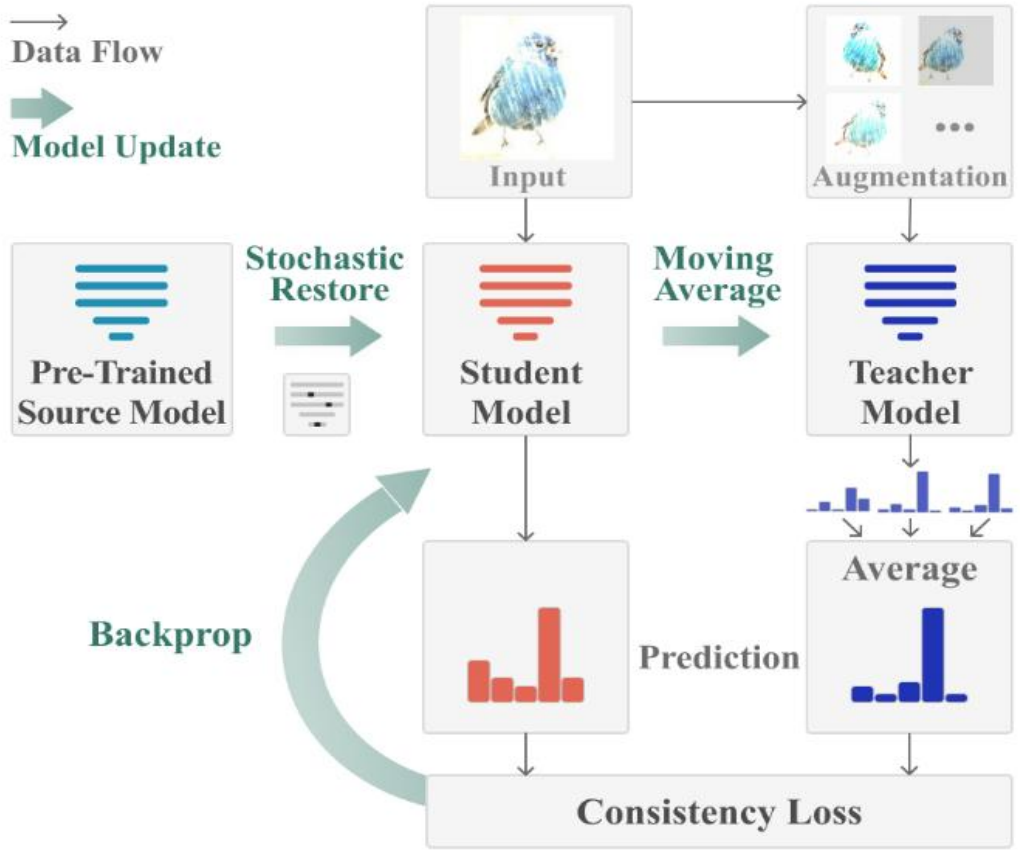


Figure 2. An overview of the proposed continual test-time adaptation (CoTTA) approach. CoTTA adapts from an off-the-shelf source pre-trained network. Error accumulation is mitigated by using a teacher model to provide weight-averaged pseudo-labels and using multiple augmentations to average the predictions. Knowledge from the source data is preserved by stochastically restoring a small number of elements of trainable weights.

Methodology

1、Weight-Averaged Pseudo-Labels

2、Augmentation-Averaged Pseudo-Labels

3、Stochastic Restoration

Experiments

Table 2. Classification error rate (%) for the standard CIFAR10-to-CIFAR10C online continual test-time adaptation task. Tesults are evaluated on WideResNet-28 with the largest corruption severity level 5. * denotes the requirement on additional domain information.

$t \longrightarrow$

Method	*Weight-avg.*	*Aug-avg.*	*Stochastic Restore*	*Gaussian*	*shot*	*impulse*	*defocus*	*glass*	*motion*	*zoom*	*snow*	*frost*	*fog*	*brightness*	*contrast*	*elastic_trans*	*pixelate*	*jpeg*	Mean
Source				72.3	65.7	72.9	46.9	54.3	34.8	42.0	25.1	41.3	26.0	9.3	46.7	26.6	58.5	30.3	43.5
BN Stats Adapt				28.1	26.1	36.3	12.8	35.3	14.2	12.1	17.3	17.4	15.3	8.4	12.6	23.8	19.7	27.3	20.4
Pseudo-label				26.7	22.1	32.0	13.8	32.2	15.3	12.7	17.3	17.3	16.5	10.1	13.4	22.4	18.9	25.9	19.8
TENT-online* [61]				24.8	23.5	33.0	12.0	31.8	13.7	10.8	15.9	16.2	13.7	7.9	12.1	22.0	17.3	24.2	18.6
TENT-continual [61]				24.8	**20.6**	28.6	14.4	31.1	16.5	14.1	19.1	18.6	18.6	12.2	20.3	25.7	20.8	24.9	20.7
CoTTA (Ours)	✓			27.2	22.8	30.8	12.1	30.1	13.9	11.9	17.2	16.0	14.3	9.4	13.1	19.9	15.4	19.9	18.3
CoTTA (Ours)	✓	✓		24.5	21.0	**26.0**	12.3	27.9	13.9	12.0	16.6	15.9	14.7	9.4	13.6	19.8	14.7	18.7	17.4
CoTTA (Ours)	✓	✓	✓	**24.3**	21.3	26.6	**11.6**	**27.6**	**12.2**	**10.3**	**14.8**	**14.1**	**12.4**	**7.5**	**10.6**	**18.3**	**13.4**	**17.3**	**16.2** (0.1)

Experiments

南京航空航天大學 NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Table 4. Classification error rate (%) for the standard CIFAR100-to-CIFAR100C online continual test-time adaptation task. All results are evaluated on the ResNeXt-29 architecture with the largest corruption severity level 5.

Time	t ⟶															
Method	*Gaussian*	*shot*	*impulse*	*defocus*	*glass*	*motion*	*zoom*	*snow*	*frost*	*fog*	*brightness*	*contrast*	*elastic_trans*	*pixelate*	*jpeg*	Mean
Source	73.0	68.0	39.4	29.3	54.1	30.8	28.8	39.5	45.8	50.3	29.5	55.1	37.2	74.7	41.2	46.4
BN Stats Adapt	42.1	40.7	42.7	27.6	41.9	29.7	27.9	34.9	35.0	41.5	26.5	30.3	35.7	32.9	41.2	35.4
Pseudo-label	38.1	36.1	40.7	33.2	45.9	38.3	36.4	44.0	45.6	52.8	45.2	53.5	60.1	58.1	64.5	46.2
TENT-continual [61]	**37.2**	**35.8**	41.7	37.9	51.2	48.3	48.5	58.4	63.7	71.1	70.4	82.3	88.0	88.5	90.4	60.9
CoTTA (Proposed)	40.1	37.7	**39.7**	**26.9**	**38.0**	**27.9**	**26.4**	**32.8**	**31.8**	**40.3**	**24.7**	**26.9**	**32.5**	**28.3**	**33.5**	**32.5**

Experiments

Table 5. Semantic segmentation results (mIoU in %) on the Cityscapes-to-ACDC online continual test-time adaptation task. We evaluate the four test conditions continually for ten times to evaluate the long-term adaptation performance. To save space, we only show the continual adaptation results in the first, fourth, seventh, and last round. Full results can be found in the supplementary material. All results are evaluated based on the Segformer-B5 architecture.

Time	$t \longrightarrow$																
Round	1				4				7				10				All
Condition	Fog	Night	rain	snow	Fog	Night	rain	snow	Fog	Night	rain	snow	Fog	Night	rain	snow	Mean
Source	69.1	40.3	59.7	57.8	69.1	40.3	59.7	57.8	69.1	40.3	59.7	57.8	69.1	40.3	59.7	57.8	56.7
BN Stats Adapt	62.3	38.0	54.6	53.0	62.3	38.0	54.6	53.0	62.3	38.0	54.6	53.0	62.3	38.0	54.6	53.0	52.0
TENT-continual [61]	69.0	40.2	60.1	57.3	66.5	36.3	58.7	54.0	64.2	32.8	55.3	50.9	61.8	29.8	51.9	47.8	52.3
CoTTA (Proposed)	**70.9**	**41.2**	**62.4**	**59.7**	**70.9**	**41.0**	**62.7**	**59.7**	**70.9**	**41.0**	**62.8**	**59.7**	**70.8**	**41.0**	**62.8**	**59.7**	**58.6**

Thanks