Test-Time Adaptation to Distribution Shifts by Confidence Maximization and Model Augmentation

Anonymous authors

A Appendix

A.1 Input transformation module

Note that we define our adaptable model as $g = f \circ d$, where d is a trainable network prepended to a pretrained neural network f (e.g., pretrained ResNet50). We choose $d(x) = \gamma \cdot [\tau x + (1-\tau) r_\psi(x)] + \beta$, where $\tau \in \mathbb{R}$, $(\beta, \gamma) \in \mathbb{R}^{n_{in}}$ with n_{in} being the number of input channels, r_ψ being a network with identical input and output shape, and $\cdot$ denoting elementwise multiplication. Here, β and γ implement a channel-wise affine transformation and τ implements a convex combination of unchanged input and the transformed input $r_\psi(x)$. We set $\tau = 1$, $\gamma = \mathbf{1}$, and $\beta = \mathbf{0}$, to ensure that $d(x) = x$ and thus $g = f$ at initialization. In principle, r_ψ can be chosen arbitrarily. In this work, we choose r_ψ as a simple stack of 3×3 convolutions with stride 1 and padding 1, group normalization, and ReLUs without any upsampling/downsampling layers. Specifically, the structure of g is illustrated in Figure A1.

A.2 Frozen layers in different networks

As discussed in Section 3.2.2, we freeze all trainable parameters in the top layers of the networks to prohibit “logit explosion”. That implies, we do not optimize the channel-wise affine transformations of the top layers but normalization statistics are still estimated. Similar to the hyperparameters of test time adaptation settings, the choice of these layers are made using ImageNet-C validation data. We mention the frozen layers of each architecuture below. Note that the naming convention of these layers are based on the model definition in torchvision:

- DenseNet121 - *features.denseblock4, features.norm5*.
- MobileNetV2 - *features.16, features.17, features.18*.
- ResNeXt50, ResNet50 and ResNet50 (DeepAugment+Augmix) - *layer4*.

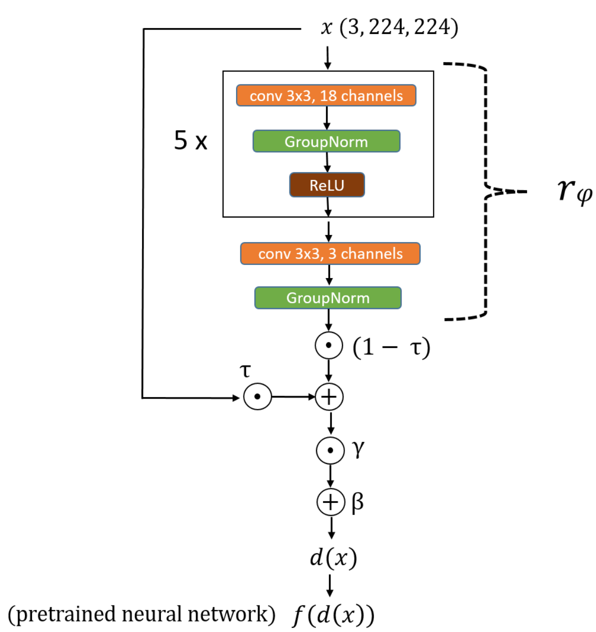
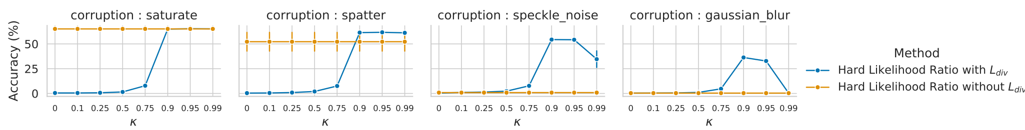


Figure A1: Structure of our adaptable model g, that comprises of r_ψ.

(a)



(b)

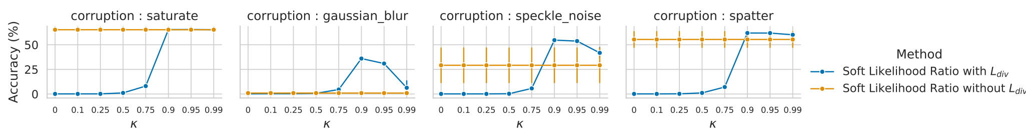


Figure A2: Effect of different κ on both (a) HLR and (b) SLR

A.3 Effect of κ

Note that the running estimate of L_{div} prevents model collapsed to trivial solutions i.e., model predicts only a single or a set of classes as outputs regardless of the input samples. L_{div} encourages model to match it's empirical distribution of predictions to class distribution of target data (uniform distribution in our experiments). Such diversity regularization is crucial as there is no direct supervision attributing to different classes and thus aids to avoid collapsed trivial solutions. In Figure A2, we investigate different values of κ on validation corruptions of ImageNet-C to study its effectiveness on our approach. It can be observed that both the HLR and SLR without L_{div} leads to collapsed solutions (e.g., accuracy drops to 0%) on some of the corruptions and the performance gains are not consistent across all the corruptions. On the other hand, L_{div} with $\kappa = 0.9$ remain consistent and improve the performance across all the corruptions.

A.4 Test-time adptation of pretrained models with SHOT

Following SHOT [11], we use their pseudo labeling strategy on the ImageNet pretrained ResNet50 in combination with TENT+, HLR and SLR. Note that TENT+ and pseudo labeling strategy jointly forms the method SHOT. The pseudo labeling strategy starts after the 1st epoch and thereafter computed at every epoch. The weight for the loss computed on the pseudo labels is set to 0.3, similar to [11]. Different values for this weight is explored and found 0.3 to perform best. Table A2 compares the results of the methods with and without pseudo labeling strategy. It can be observed that the results with pseudo labeling strategy perform worse than without taking this strategy into account.

We further modified the pretrained ResNet50 by following the network modifications suggested in [11], that includes adding a bottleneck layer with BatchNorm and applying weight norm on the linear classifier along with smooth label training to facilitate the pseudo labeling strategy. Table A3 shows that the pseudo labeling strategy on such network improve the results of TENT+ from epoch 1 to epoch 5. However, there are no improvements noticed in SLR. Moreover, Table A4 shows that NO pseudo labeling strategy on the same network performs better than applying the pseudo labeling strategy. Finally, the no pseduo labeling results from Table A2 and A4 shows that additional modifications to ResNet50 do not improve the performance when compared to the standard ResNet50.

A.5 Experiments on other domain adaptation datasets

We extended our experiments to VisDA-C. We followed similar network architecture from SHOT [11] and evaluated TENT+, our SLR loss function with diversity regularizer. Similar to ImageNet-C, we adapted only the channel wise affine parameters of batchnorm layers for 5 epochs with Adam optimizer with cosine decay scheduler of the learning rate with initial value $2e{-}5$. Here, the batchsize is set to 64, the weight of L_{conf} in our loss function to $\delta = 0.25$ and $\kappa = 0$ in the running estimate $p_t(y)$ of L_{div}, since the number of classes in this dataset (12 classes) is smaller than the batchsize. Setting $\kappa = 0$ enables the batch wise diversity regularizer. Table A5 shows average results from three

Table A1: Test-time adaptation of ResNet50 on ImageNet-C at highest severity level 5. Same as Table 1 with error bars.

name			Epoch 1				Epoch 5			
corruption	No adaptation	PL	TENT	TENT+	HLR	SLR	TENT	TENT+	HLR	SLR
Gauss	2.44	2.44	32.70±0.10	33.96±0.09	38.39±0.25	**39.51**±0.23	16.04±0.51	33.97±0.17	41.37±0.09	**41.52**±0.08
Shot	2.99	2.99	35.34±0.17	36.66±0.19	41.11±0.13	**42.09**±0.26	23.22±0.74	37.95±0.10	**44.04**±0.09	42.90±0.08
Impulse	1.96	1.96	35.11±0.09	35.75±0.15	40.28±0.20	**41.58**±0.04	25.85±1.01	36.93±0.09	43.68±0.06	**44.07**±0.06
Defocus	17.92	17.92	32.79±0.10	33.70±0.14	38.25±0.32	**39.35**±0.13	19.05±0.61	32.69±0.25	**41.74**±0.12	41.69±0.07
Glass	9.82	9.82	31.80±0.15	33.33±0.01	38.18±0.08	**39.02**±0.09	17.40±0.21	33.36±0.13	**41.09**±0.17	40.78±0.08
Motion	14.78	14.78	47.22±0.11	47.73±0.12	51.63±0.08	**52.67**±0.25	49.02±0.08	51.42±0.07	54.26±0.02	**54.76**±0.04
Zoom	22.50	22.50	53.02±0.06	53.22±0.07	55.55±0.06	**55.80**±0.07	52.78±0.16	54.33±0.06	56.43±0.07	**56.59**±0.05
Snow	16.89	16.89	51.82±0.05	52.16±0.09	55.45±0.11	**55.92**±0.06	52.72±0.13	54.55±0.07	57.03±0.12	**57.35**±0.03
Frost	23.31	23.31	43.42±0.30	44.79±0.20	48.96±0.07	**49.64**±0.14	34.31±0.50	45.80±0.27	50.81±0.08	**51.01**±0.02
Fog	24.43	24.43	60.44±0.08	60.62±0.05	62.19±0.03	**62.62**±0.04	61.19±0.08	62.09±0.05	63.05±0.04	**63.53**±0.08
Bright	58.93	58.93	68.82±0.02	**68.91**±0.03	68.17±0.01	68.47±0.05	68.54±0.06	**69.03**±0.06	68.29±0.09	68.72±0.10
Contrast	5.43	5.43	27.53±0.98	35.60±0.77	49.47±0.20	**50.27**±0.08	1.26±0.32	24.08±1.36	**50.98**±2.54	50.65±0.55
Elastic	16.95	16.95	58.47±0.05	58.81±0.05	60.34±0.18	**60.80**±0.08	59.26±0.06	60.36±0.02	61.15±0.04	**61.49**±0.07
Pixel	20.61	20.61	61.63±0.06	61.82±0.07	62.51±0.10	**63.01**±0.08	62.15±0.04	63.10±0.08	63.08±0.06	**63.46**±0.08
JPEG	31.65	31.65	55.98±0.09	56.23±0.05	57.42±0.13	**57.80**±0.04	56.17±0.07	57.21±0.02	58.13±0.09	**58.32**±0.05

Table A2: Test-time adaptation of ResNet50 on ImageNet-C at highest severity level 5 with and without the pseudo labeling strategy [11].

name		No pseudo labeling: Epoch 5			Pseudo labeling: Epoch 5		
corruption	No adaptation	TENT+	HLR	SLR	TENT+	HLR	SLR
Gauss	2.44	33.97±0.17	41.37±0.09	41.52±0.08	34.08±0.11	34.88±0.35	35.58±0.06
Shot	2.99	37.95±0.10	44.04±0.09	42.90±0.08	36.74±0.26	37.61±0.49	37.98±0.19
Impulse	1.96	36.93±0.09	43.68±0.06	44.07±0.06	36.69±0.04	37.24±0.22	37.77±0.05
Defocus	17.92	32.69±0.25	41.74±0.12	41.69±0.07	33.99±0.28	34.76±0.11	35.11±0.10
Glass	9.82	33.36±0.13	41.09±0.17	40.78±0.08	34.06±0.12	34.51±0.30	34.81±0.27
Motion	14.78	51.42±0.07	54.26±0.02	54.76±0.04	50.91±0.09	48.96±0.39	49.46±0.20
Zoom	22.50	54.33±0.06	56.43±0.07	56.59±0.05	54.10±0.10	52.49±0.02	52.50±0.23
Snow	16.89	54.55±0.07	57.03±0.12	57.35±0.03	54.06±0.08	52.49±0.19	52.95±0.07
Frost	23.31	45.80±0.27	50.81±0.08	51.01±0.02	44.44±0.07	45.47±0.26	46.06±0.20
Fog	24.43	62.09±0.05	63.05±0.04	63.53±0.08	61.91±0.08	59.66±0.14	59.98±0.12
Bright	58.93	69.03±0.06	68.29±0.09	68.72±0.10	68.98±0.02	65.59±0.06	66.00±0.03
Contrast	5.43	24.08±1.36	50.98±2.54	50.65±0.55	29.37±0.95	44.58±0.38	45.64±0.47
Elastic	16.95	60.36±0.02	61.15±0.04	61.49±0.07	60.23±0.05	57.48±0.14	57.87±0.04
Pixel	20.61	63.10±0.08	63.08±0.06	63.46±0.08	62.98±0.04	59.72±0.02	60.05±0.14
JPEG	31.65	57.21±0.02	58.13±0.09	58.32±0.05	57.09±0.04	54.72±0.09	54.88±0.07

different random seeds and also shows that SLR outperforms TENT+ on this dataset. Similarly, we show the results on Office-Home dataset in Table A6.

Table A3: Test-time adaptation of modified ResNet50 (following [11]) on ImageNet-C at highest severity level 5 with pseudo labeling strategy at epoch 1 and epoch 5.

name		Pseudo labeling: Epoch 1			Pseudo labeling: Epoch 5		
corruption	No adaptation	TENT+	HLR	SLR	TENT+	HLR	SLR
Gauss	2.95	31.03±0.18	34.65±0.28	37.21±0.23	35.26±0.16	35.93±0.23	37.61±0.30
Shot	3.65	33.55±0.07	38.09±0.30	40.30±0.09	37.39±0.05	38.95±0.16	40.42±0.06
Impulse	2.54	32.70±0.07	36.95±0.05	39.73±0.07	38.16±0.08	38.13±0.04	40.12±0.11
Defocus	19.36	31.66±0.15	35.08±0.05	37.18±0.15	35.95±0.17	36.72±0.13	37.96±0.25
Glass	9.72	31.06±0.06	35.46±0.12	37.62±0.10	35.98±0.04	36.84±0.11	37.90±0.02
Motion	15.66	46.96±0.12	49.95±0.12	51.87±0.14	52.24±0.02	51.90±0.12	52.76±0.09
Zoom	22.20	52.45±0.02	54.15±0.22	54.84±0.18	54.80±0.07	54.84±0.09	54.95±0.14
Snow	17.56	51.79±0.05	53.98±0.06	55.44±0.04	55.15±0.02	55.27±0.20	55.75±0.02
Frost	24.11	45.59±0.06	47.87±0.03	48.96±0.11	48.10±0.20	48.52±0.11	49.13±0.20
Fog	25.59	60.33±0.03	61.55±0.10	62.21±0.16	62.39±0.03	62.38±0.12	62.38±0.11
Bright	58.30	68.84±0.04	68.44±0.04	68.60±0.10	69.13±0.04	68.50±0.02	68.47±0.09
Contrast	6.49	42.34±0.19	47.98±0.13	50.32±0.28	42.11±0.15	49.22±0.42	50.80±0.19
Elastic	17.72	58.47±0.02	59.70±0.06	60.30±0.09	60.40±0.04	60.27±0.22	60.45±0.21
Pixel	21.29	61.39±0.06	62.10±0.07	62.71±0.10	63.04±0.02	62.71±0.07	62.81±0.07
JPEG	32.13	55.22±0.03	56.49±0.07	57.04±0.07	57.21±0.06	57.25±0.07	57.37±0.05

Table A4: Test-time adaptation of modified ResNet50 (following [11]) on ImageNet-C at highest severity level 5 with and without pseudo labeling strategy.

name		No Pseudo labeling: Epoch 5			Pseudo labeling: Epoch 5		
corruption	No adaptation	TENT+	HLR	SLR	TENT+	HLR	SLR
Gauss	2.95	34.96±0.08	38.58±0.12	39.72±0.13	35.26±0.16	35.93±0.23	37.61±0.30
Shot	3.65	37.22±0.17	41.59±0.09	42.45±0.05	37.39±0.05	38.95±0.16	40.42±0.06
Impulse	2.54	37.82±0.04	40.88±0.07	42.39±0.03	38.16±0.08	38.13±0.04	40.12±0.11
Defocus	19.36	34.46±0.12	39.22±0.15	39.78±0.09	35.95±0.17	36.72±0.13	37.96±0.25
Glass	9.72	35.12±0.05	38.83±0.13	39.37±0.07	35.98±0.04	36.84±0.11	37.90±0.02
Motion	15.66	51.91±0.09	53.23±0.05	54.00	52.24±0.02	51.90±0.12	52.76±0.09
Zoom	22.20	54.57±0.05	55.76±0.04	55.79±0.02	54.80±0.07	54.84±0.09	54.95±0.14
Snow	17.56	55.02±0.05	56.35±0.12	56.80±0.04	55.15±0.02	55.27±0.20	55.75±0.02
Frost	24.11	48.18±0.09	49.86±0.22	50.43±0.08	48.10±0.20	48.52±0.11	49.13±0.20
Fog	25.59	62.24±0.04	62.90±0.06	63.29±0.06	62.39±0.03	62.38±0.12	62.38±0.11
Bright	58.30	69.12±0.01	68.72±0.06	68.83±0.05	69.13±0.04	68.50±0.02	68.47±0.09
Contrast	6.49	33.91±0.92	52.13±0.16	53.04±0.14	42.11±0.15	49.22±0.42	50.80±0.19
Elastic	17.72	60.37±0.11	60.89±0.08	61.12±0.01	60.40±0.04	60.27±0.22	60.45±0.21
Pixel	21.29	62.97±0.02	62.95±0.05	63.21±0.05	63.04±0.02	62.71±0.07	62.81±0.07
JPEG	32.13	57.10±0.06	57.91±0.06	57.99±0.11	57.21±0.06	57.25±0.07	57.37±0.05

Table A5: Performance on VisDA-C dataset

Method	Accuracy(%)
No Adaptation	46.1
TENT+	81.83±0.16
SLR	82.32±0.16

Table A6: Accuracy (%) on Office Home

Method (Source ⟶ Target)	Art ⟶ Clipart	Art ⟶ Product	Art ⟶ RealWorld
TENT+	54.75	74.5	77.74
SLR	54.59	73.9	77.5