Parameter-free Online Test-time Adaptation

Malik Boudiaf
ÉTS Montreal *

Romain Mueller
FiveAI

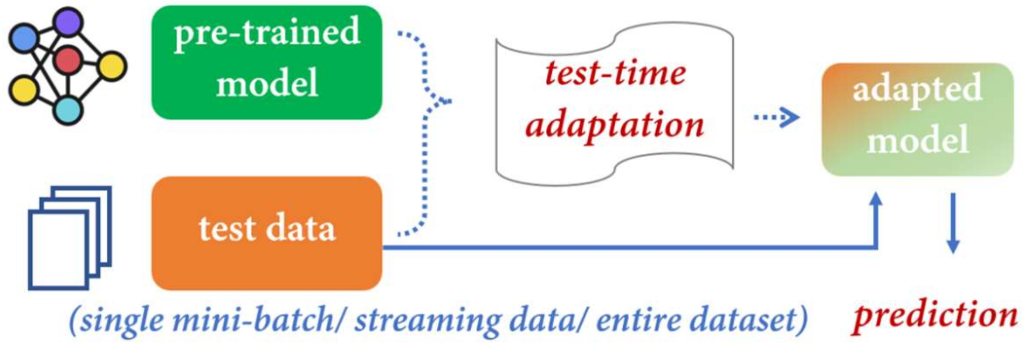
Ismail Ben Ayed
ÉTS Montreal

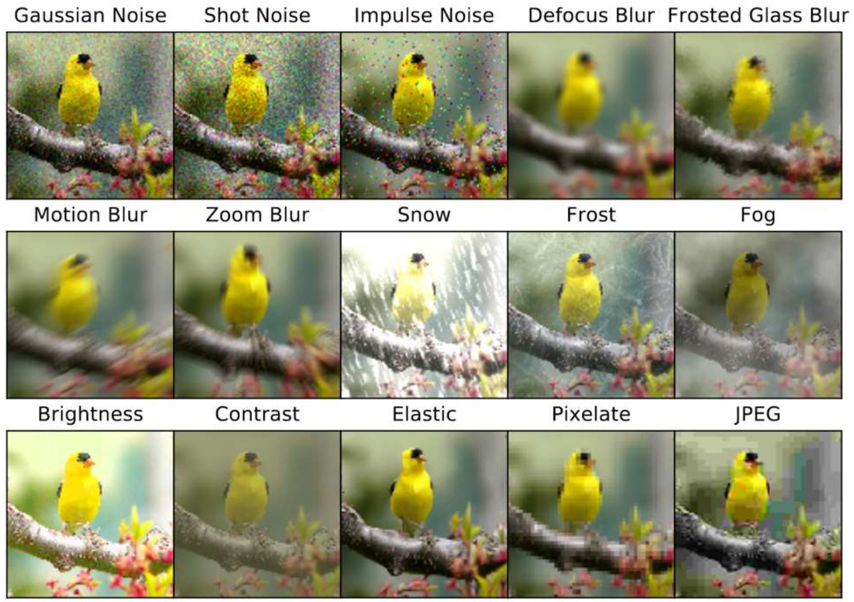
Luca Bertinetto
FiveAI

CPVR 2022

Background

Test Time Adaptation





Typical Settings Comparison

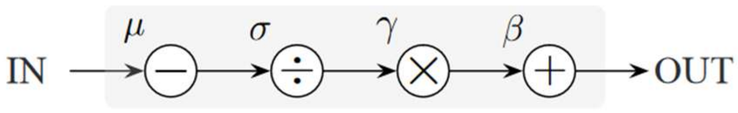
setting	source data	target data	train loss	test loss
fine-tuning	-	x^t, y^t	$L(x^t, y^t)$	-
domain adaptation	x^s, y^s	x^t	$L(x^s, y^s) + L(x^s, x^t)$	-
test-time training	x^s, y^s	x^t	$L(x^s, y^s) + L(x^s)$	$L(x^t)$
fully test-time adaptation	-	x^t	-	$L(x^t)$

Background

Two Representative Methods

- **BN Adapt**

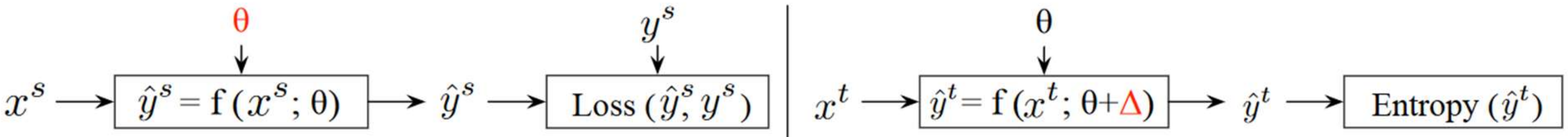
Using the BN statistics μ_t, σ_t in each test minibatch to replace the original statistics of μ_s, σ_s



$$\bar{\mu} = \frac{N}{N+n}\mu_s + \frac{n}{N+n}\mu_t, \quad \bar{\sigma}^2 = \frac{N}{N+n}\sigma_s^2 + \frac{n}{N+n}\sigma_t^2.$$

- **Tent**

Minimize the entropy of each test sample's label prediction; only update the BN affine parameters γ, β

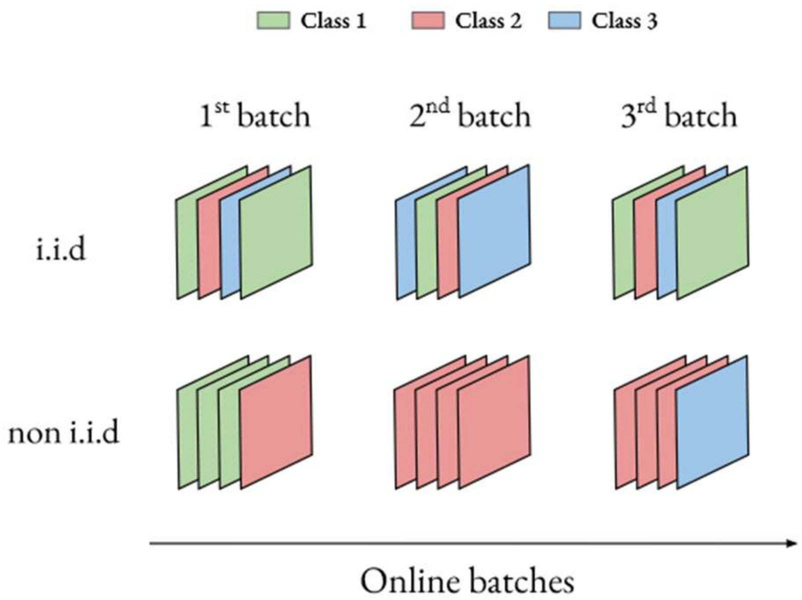


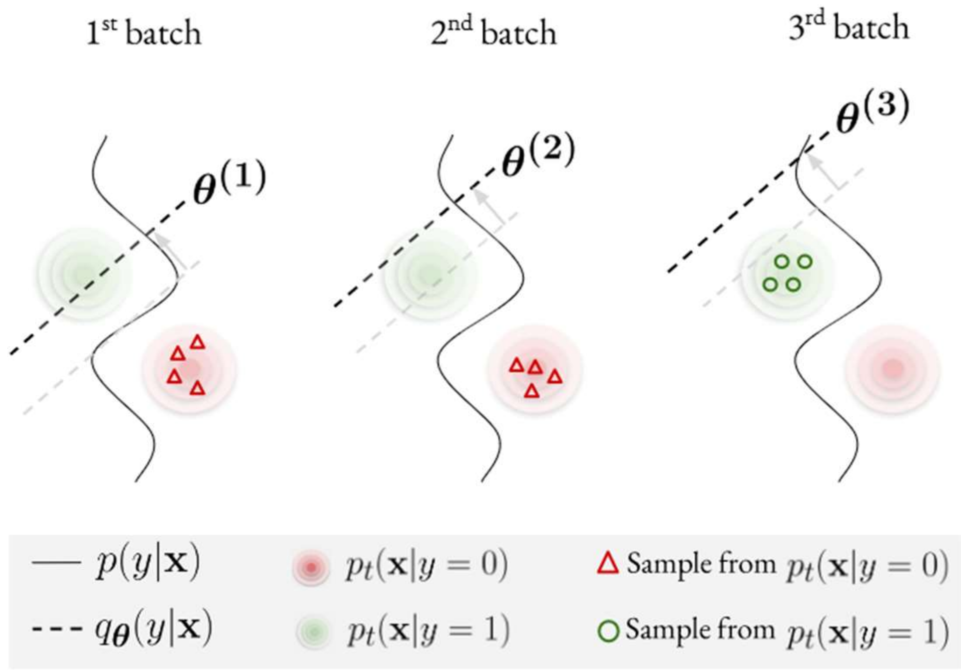
(a) training (b) fully test-time adaptation

Motivation

Problems:

The test data in each batch may be non i.i.d



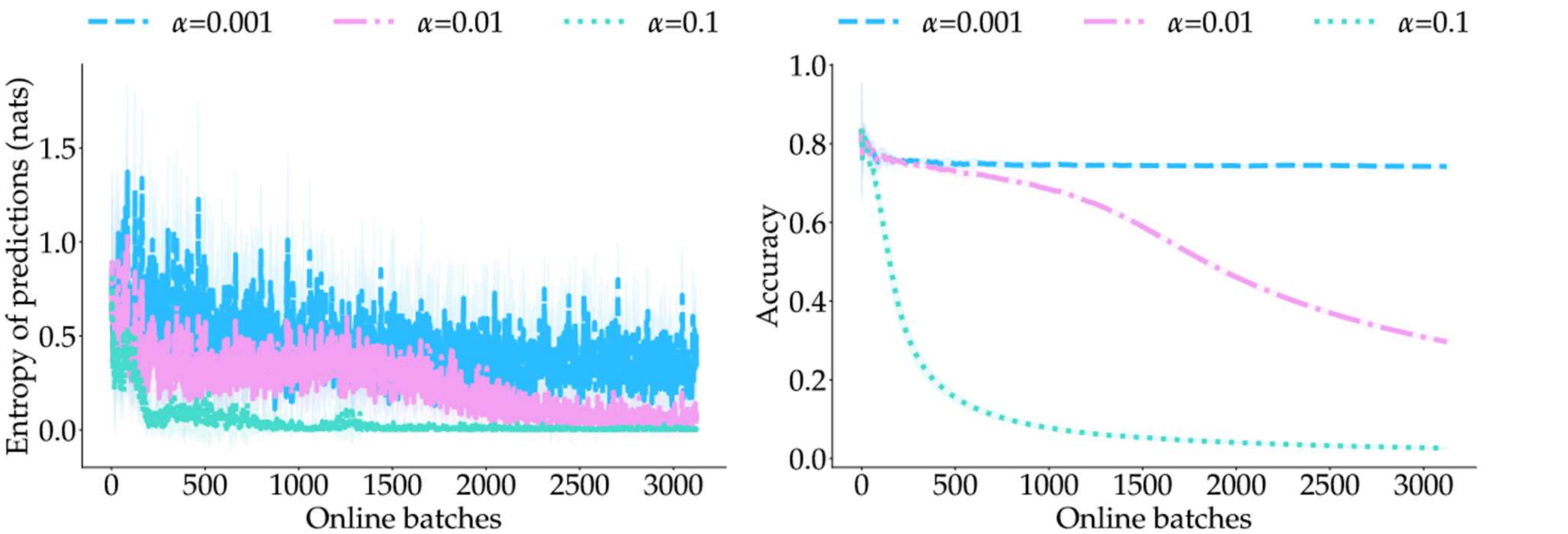


Motivation

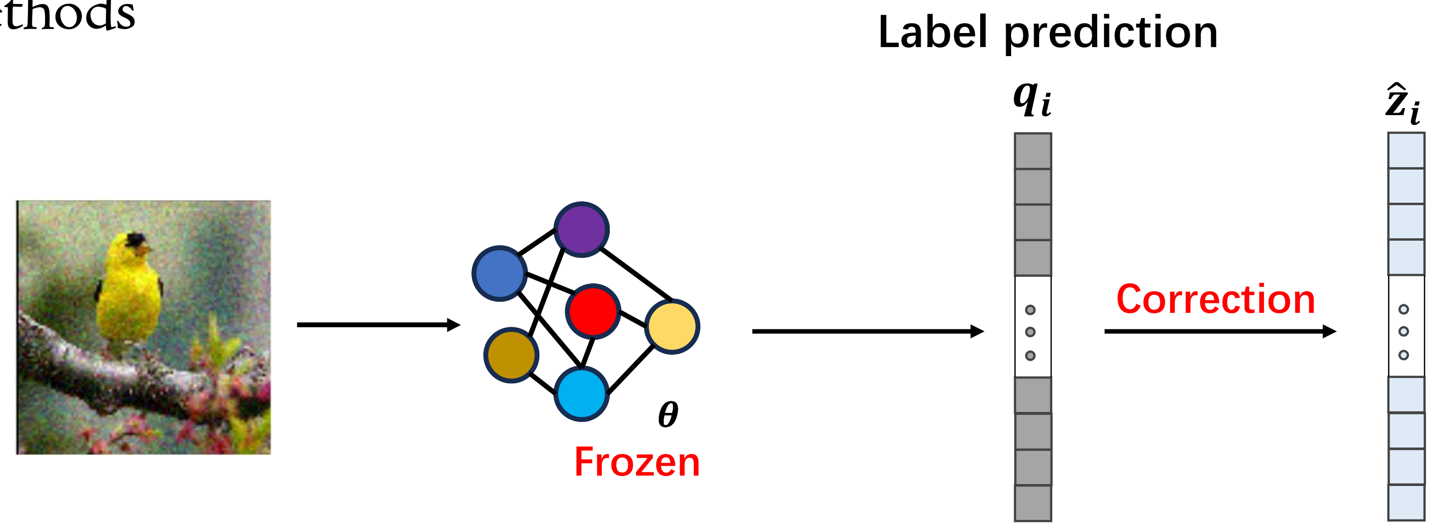
南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Problems:

TTA methods are highly sensitive to hyper-parameters



Methods



Latent assignment

$\hat{z}_i = (\hat{z}_{ik})_{1 \le k \le K} \in \Delta^{K-1}$

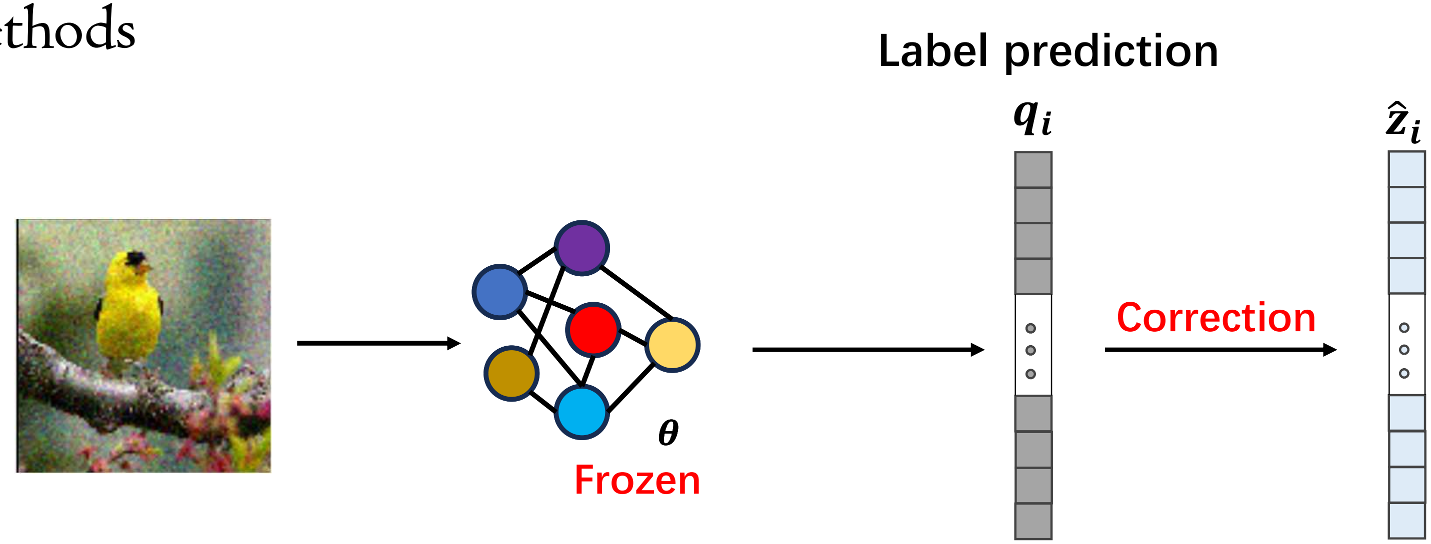
Assume true label prediction as p, maximize the log-likelihood

$$\mathcal{L}(\tilde{\mathbf{Z}}) = \log \left(\prod_{i=1}^{N} \prod_{k=1}^{K} p(\mathbf{x}_i, k)^{\tilde{z}_{ik}} \right) \stackrel{c}{=} \sum_{i=1}^{N} \tilde{\mathbf{z}}_i^T \log(\mathbf{p}_i)$$

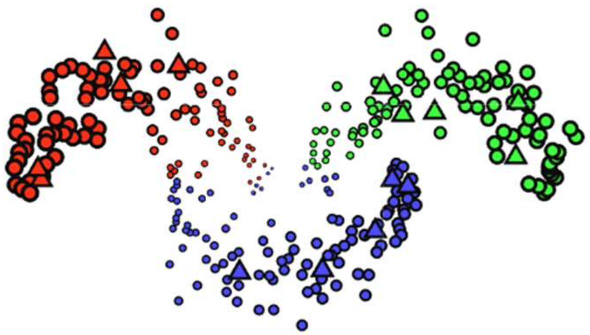
To prevent over-confident assignments; Consider a negative-entropy regularization

$$-\sum_{i=1}^{N} \tilde{\mathbf{z}}_i^T \log(\mathbf{p}_i) + \sum_{i=1}^{N} \tilde{\mathbf{z}}_i^T \log(\tilde{\mathbf{z}}_i) = \sum_{i=1}^{N} \mathrm{KL}(\tilde{\mathbf{z}}_i || \mathbf{p}_i) \Rightarrow \sum_{i} \mathrm{KL}(\tilde{\mathbf{z}}_i || \mathbf{q}_i)$$

Methods



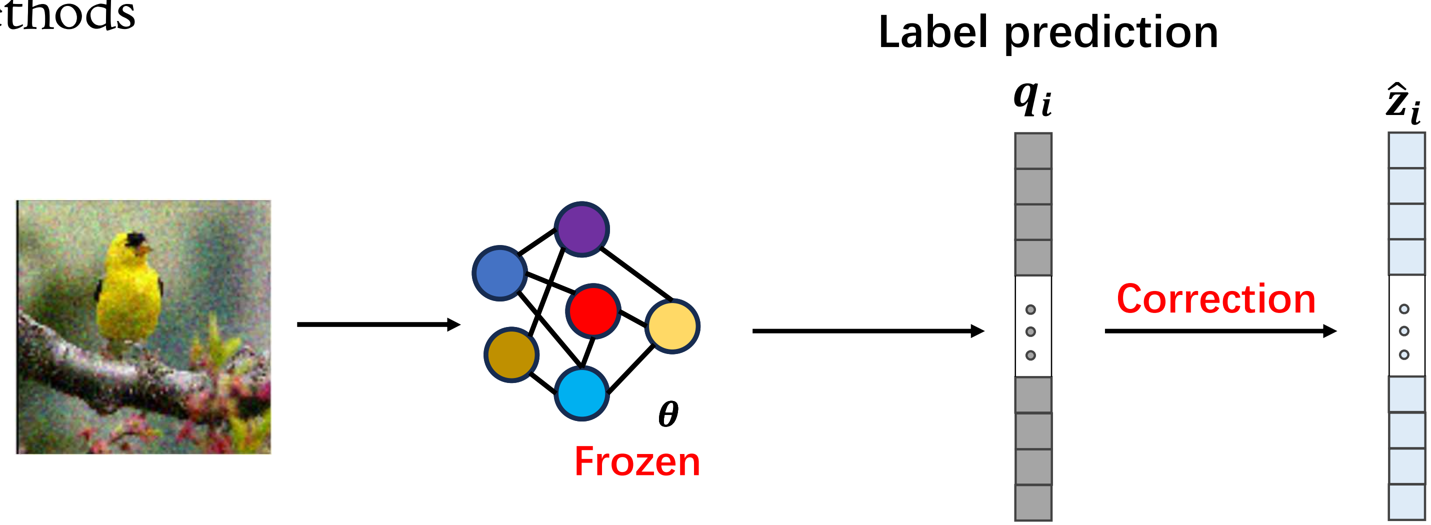
$$\mathcal{L}^{\text{LAME}}(\tilde{\mathbf{Z}}) = \sum_i \text{KL}(\tilde{\mathbf{z}}_i || \mathbf{q}_i) - \sum_{i,j} w_{ij} \tilde{\mathbf{z}}_i^T \tilde{\mathbf{z}}_j$$



$$w_{ij} = w(\phi(\mathbf{x}_i), \phi(\mathbf{x}_j))$$

$$w(\phi(\mathbf{x}_i), \phi(\mathbf{x}_j)) = \begin{cases} 1 & \text{if } \phi(\mathbf{x}_i) \in \text{kNN}(\phi(\mathbf{x}_j)) \\ 0 & \text{otherwise} \end{cases}$$

Methods



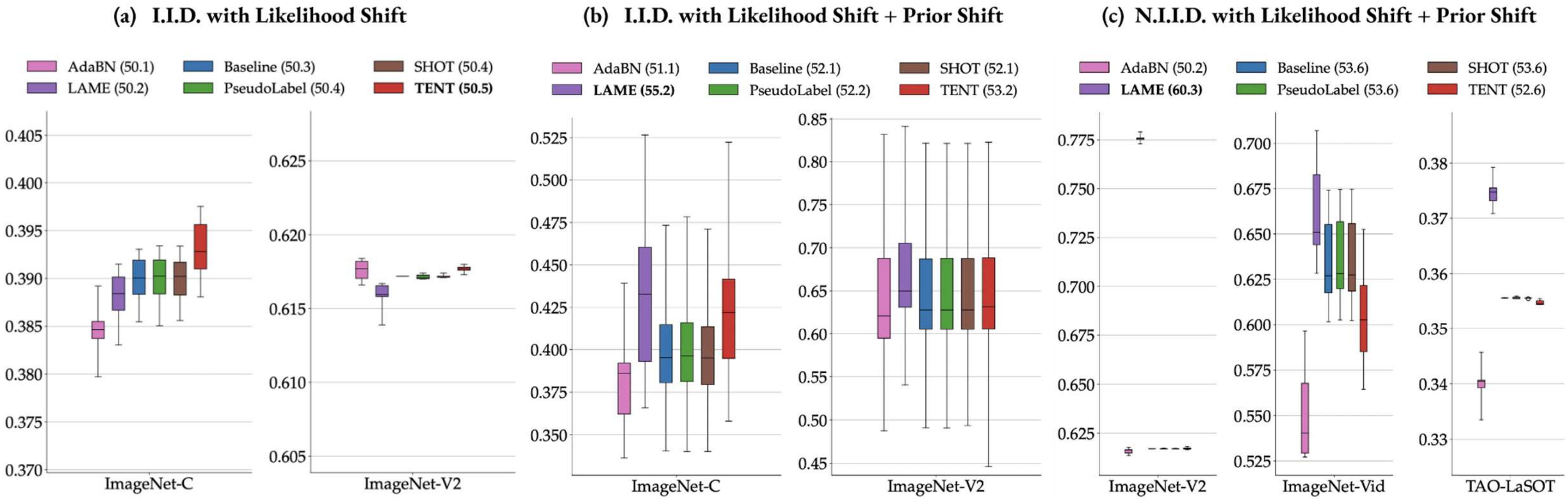
$$\mathcal{L}^{\text{LAME}}(\tilde{\mathbf{Z}}) = \sum_i \text{KL}(\tilde{\mathbf{z}}_i || \mathbf{q}_i) - \sum_{i,j} w_{ij} \tilde{\mathbf{z}}_i^T \tilde{\mathbf{z}}_j$$

Concave-Convex Procedure (CCCP)

$$\mathcal{L}^{\text{LAME}}(\tilde{\mathbf{Z}}) \overset{\text{c}}{\leq} \sum_i \text{KL}(\tilde{\mathbf{z}}_i || \mathbf{q}_i) - ((W \otimes I)\tilde{\mathbf{Z}}^{(n)})^T \tilde{\mathbf{Z}}$$

$$\tilde{z}_{ik}^{(n+1)} = \frac{q_{\boldsymbol{\theta}}(k|\mathbf{x}_i) \exp(\sum_j w_{ij} \tilde{z}_{jk}^{(n)})}{\sum_{k'} q_{\boldsymbol{\theta}}(k'|\mathbf{x}_i) \exp(\sum_j w_{ij} \tilde{z}_{jk'}^{(n)})}$$

Methods



Likelihood Shift $p_s(\boldsymbol{x}|y) \neq p_t(\boldsymbol{x}|y)$

Prior Shift $p_s(y) \neq p_t(y)$

Experiments

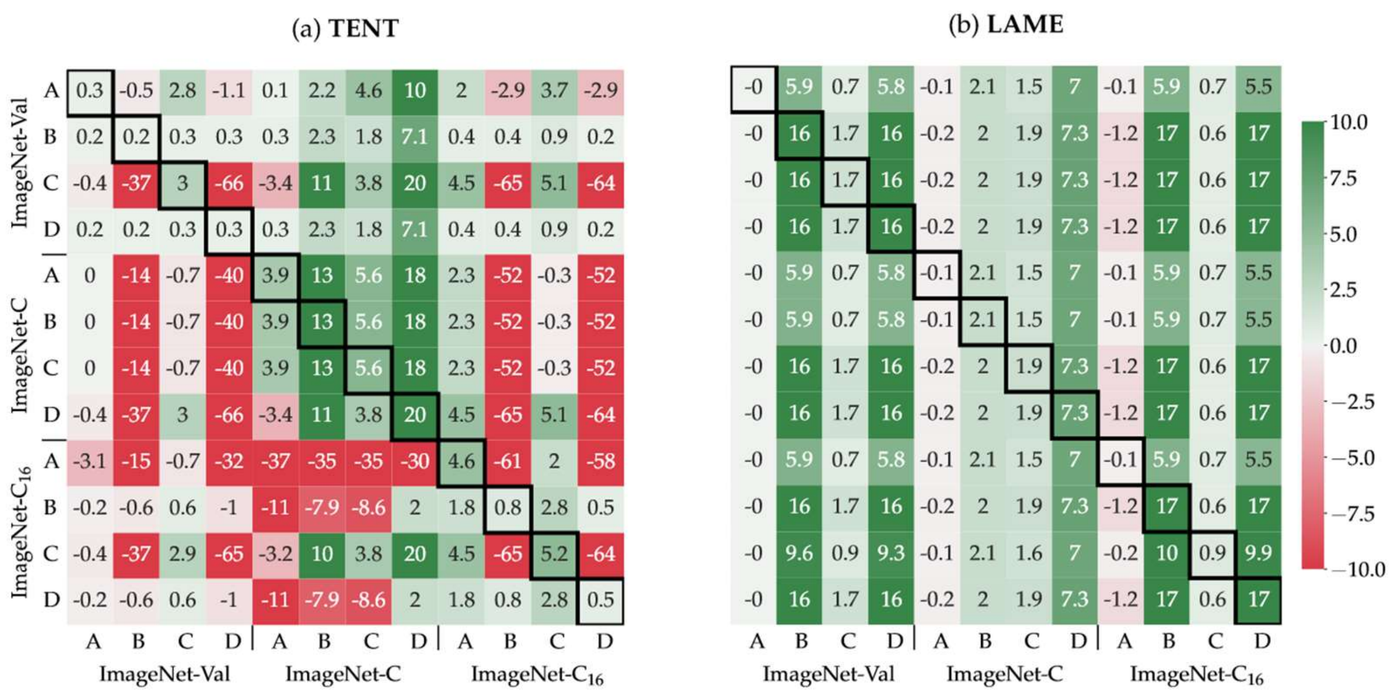


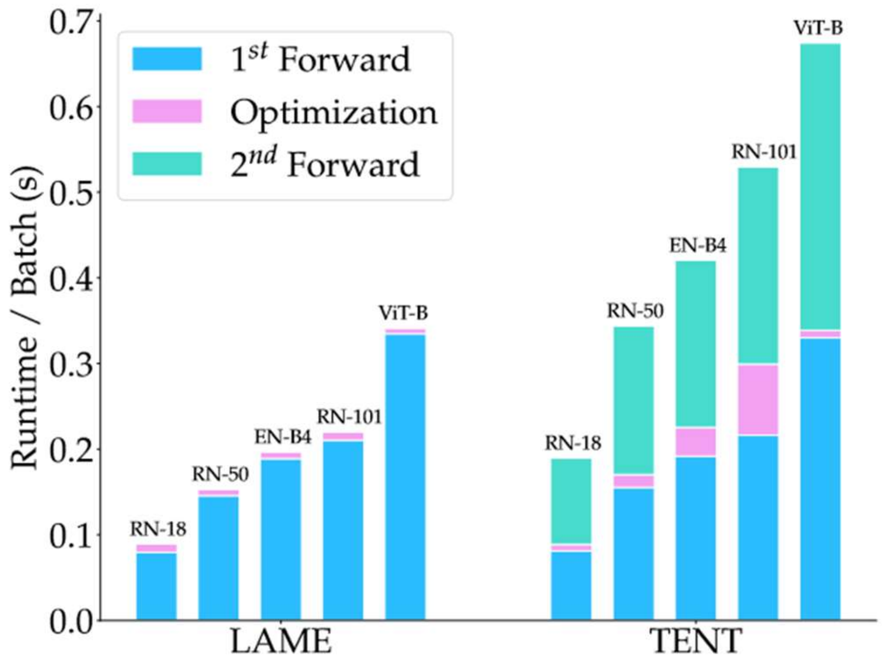
Figure 3. *Cross-shift* validation for TENT [57] (left) and our proposed LAME (right). A cell at position (i, j) shows the absolute improvement (or degradation) of the current method w.r.t. to the baseline when using the optimal hyperparameters for scenario i, but evaluating in scenario j. Legend: A = i.i.d., B = non i.i.d., C = i.i.d. + prior shift, D = non i.i.d. + prior shift. More details on the scenarios in Sec. 6

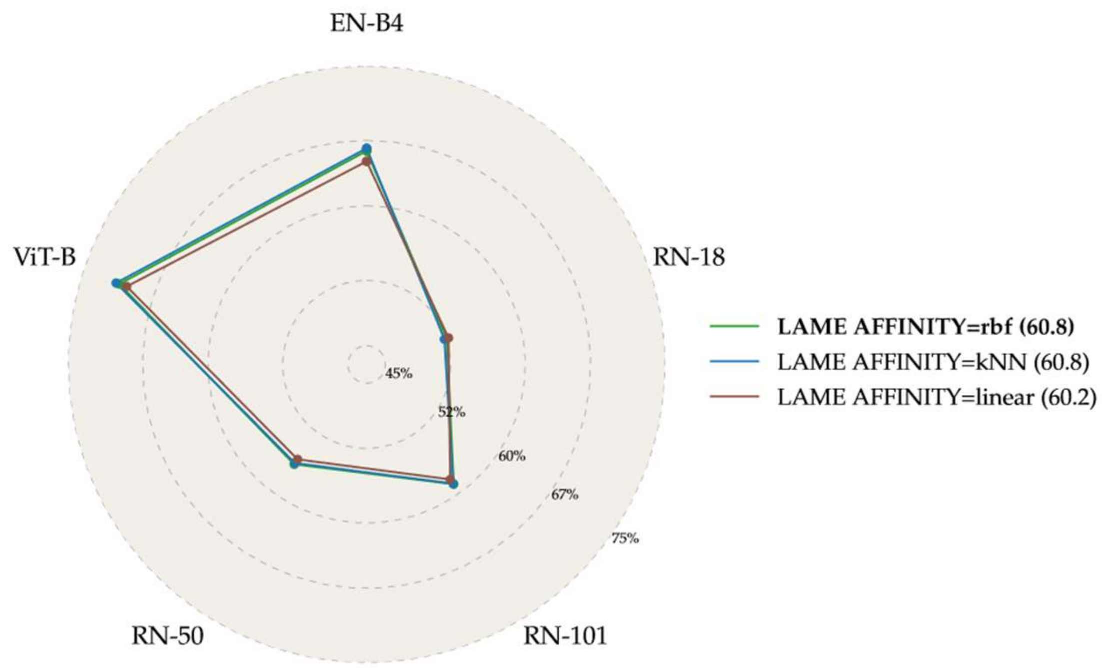
Experiments

Method	*Original*	*Torchvision*	*SimCLR*	Mean
Baseline	52.07	53.39	47.68	51.0
PseudoLabel [25]	52.17	45.53	47.67	48.5
SHOT-IM [30]	52.11	50.77	47.64	50.2
TENT [57]	49.74	28.54	37.00	38.4
AdaBN [28]	49.31	50.43	41.07	46.9
LAME	**55.70**	**57.74**	**51.46**	**55.0**

Method	RN-18	RN-50	RN-101	EN-B4	ViT-B/16	Mean
Baseline	47.28	52.07	54.93	62.67	67.69	56.9
PseudoLabel [25]	33.13	52.17	55.17	59.64	69.86	54.0
SHOT-IM [30]	42.95	52.11	54.98	59.64	65.33	55.0
TENT [57]	28.59	49.74	53.04	53.27	54.00	47.7
LAME	**51.46**	**55.70**	**58.79**	**66.17**	**71.22**	**60.7**

Experiments





Thanks