Test-Time Domain Adaptation by Learning Domain-Aware Batch Normalization

Yanan Wu[1,*] Zhixiang Chi[2,*] Yang Wang[3] Konstantinos N. Plataniotis[2] Songhe Feng[1,†]

[1]Beijing Jiaotong University [2]University of Toronto [3]Concordia University

The 38th Annual AAAI Conference on Artificial Intelligence

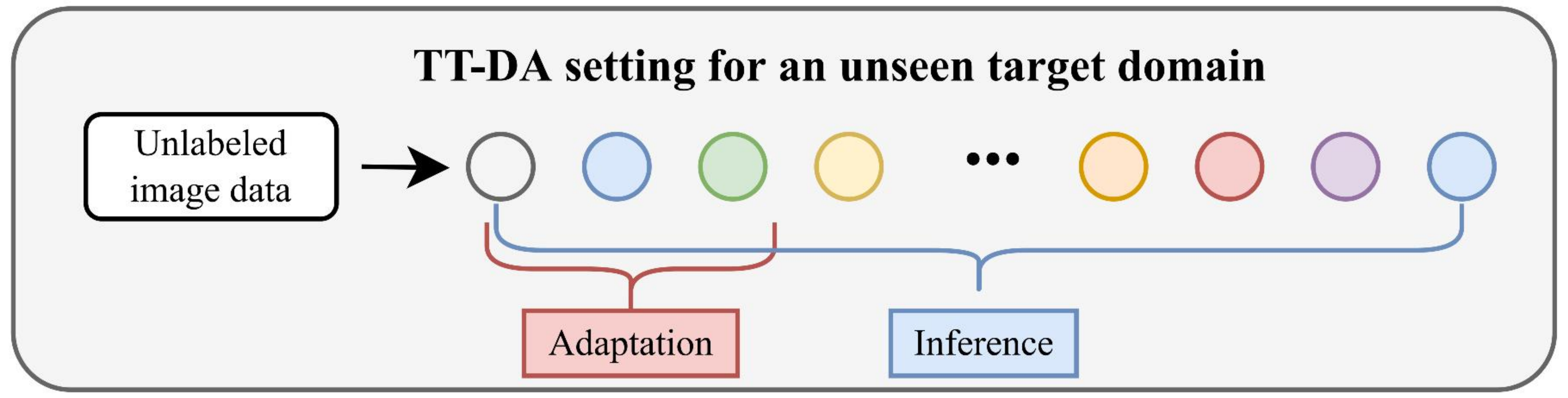
Problem and Contributions

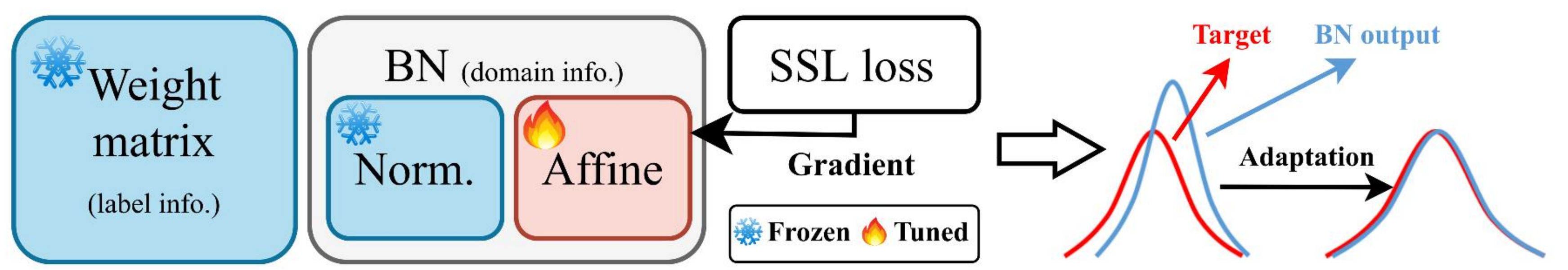
➢ A Practical Real-World Scenario

- During inference, a deployed model trained on source domain may encounter unseen domains. An intelligent system should adapt the model to unseen domain using a few unlabeled images and the adapted model is then used for inference. We refer it as Test-Time Domain Adaptation (TT-DA).

➢ Challenges

- What type of knowledge is most efficacious for adapting to an unseen domain?
- How can one procure adequate supervision to guide the model's update towards that domain?

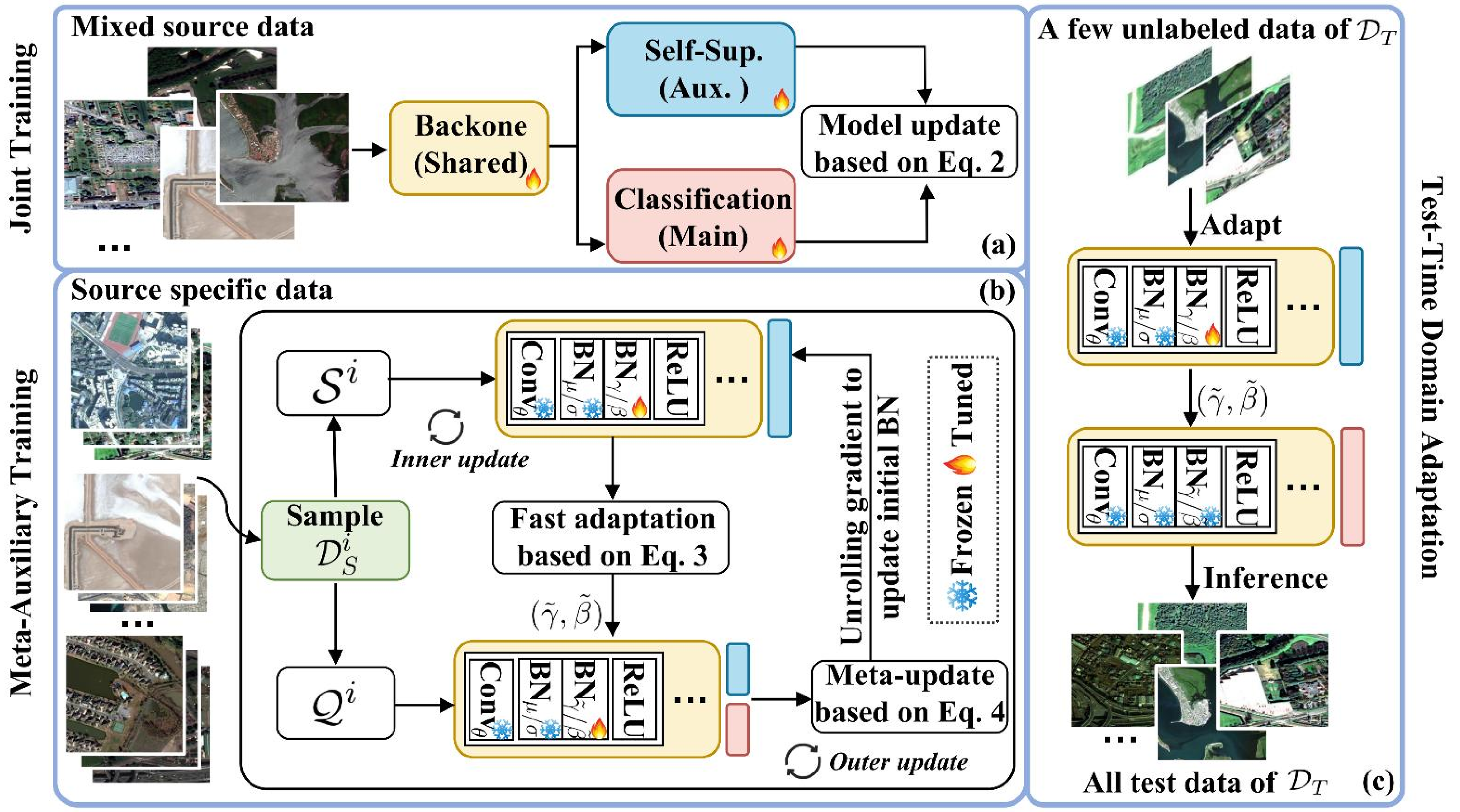




➢ Contributions

- We propose an unsupervised adaptive method that is tailored for TT-DA. We adapt only the affine parameters via a self-supervised loss to each target domain to elevate the domain knowledge learning.
- We employ a bi-level optimization to align the learning objective with the evaluation protocol to yield affine parameters that are capable of adapting domain knowledge.
- We conduct extensive experiments to show that our method is more effective in learning the domain knowledge. Thus, our domain-adapted model can be seamlessly integrated with the entropy-based TTA method(e.g. TNET) where the optimization is more toward label knowledge.
- We establish strong baselines and show that our method achieves superior performance with less hand-engineered design on six domain shift datasets.

Overview of Proposed MABN



Learning label-dependent representation: train the whole model to learn tht label knowledge and normalization statistics by mixing all the source data.

$$\mathbf{F}' = \gamma \hat{\mathbf{F}} + \beta, \quad \text{where,} \quad \hat{\mathbf{F}} = \frac{\mathbf{F} - \mu}{\sqrt{\sigma^2 + \epsilon}}$$

$$\mathcal{L}_{Joint} = \mathcal{L}_{CE} + \lambda \mathcal{L}_{SSL}$$

Learning to adapt to unseen domain knowledge treat every source domain as one "task" and use few-shot unlabeled images to update the affine parameters (while keeping other model parameters fixed) via an auxiliary branch.

$$(\tilde{\gamma}, \tilde{\beta}) = (\gamma, \beta) - \alpha \nabla_{(\gamma,\beta)} \mathcal{L}_{SSL}(\mathcal{S}^i; \theta, (\gamma, \beta))$$

$$(\gamma, \beta) \leftarrow (\gamma, \beta) - \delta \nabla_{(\gamma,\beta)} \mathcal{L}_{Joint}(\mathcal{Q}^i; \theta, (\tilde{\gamma}, \tilde{\beta}))$$

To ensure that the optimization of the affine parameters is aligned with the main task, we define the meta-objective by evaluating the adapted affine parameters on a disjoint set in the task.

Experiments

➢ Quantitative Results & Ablation Studies

Methods	iWildCam Acc	iWildCam Macro F1	Camelyon17 Acc	RxRx1 Acc	FMoW WC Acc	FMoW Avg Acc	PovertyMap WC Pearson r	PovertyMap Pearson r
ERM	71.6±2.5	31.0±1.3	70.3±6.4	29.9±0.4	32.3±1.25	53.0±0.55	0.45±0.06	0.78±0.04
CORAL	73.3±4.3	32.8±0.1	59.5±7.7	28.4±0.3	31.7±1.24	50.5±0.36	0.44±0.06	0.78±0.05
Group DRO	72.7±2.1	23.9±2.0	68.4±7.3	23.0±0.3	30.8±0.81	52.1±0.5	0.39±0.06	0.75±0.07
IRM	59.8±3.7	15.1±4.9	64.2±8.1	8.2±1.1	30.0±1.37	50.8±0.13	0.43±0.07	0.77±0.05
ARM-CML	70.5±0.6	28.6±0.1	84.2±1.4	17.3±1.8	27.2±0.38	45.7±0.28	0.37±0.08	0.75±0.04
ARM-BN	70.3±2.4	23.7±2.7	87.2±0.9	31.2±0.1	24.6±0.04	42.0±0.21	0.49±0.21	**0.84±0.05**
ARM-LL	71.4±0.6	27.4±0.8	84.2±2.6	24.3±0.3	22.1±0.46	42.7±0.71	0.41±0.04	0.76±0.04
Meta-DMoE	77.2±0.3	34.0±0.6	91.4±1.5	29.8±0.4	35.4±0.58	52.5±0.18	0.51±0.04	0.80±0.03
PAIR	74.9±1.1	27.9±0.9	74.0±7.2	28.8±0.0	35.4±1.30	-	0.47±0.09	-
MABN (ours)	**78.4±0.6**	**38.3±1.2**	**92.4±1.9**	**32.7±0.2**	**36.6±0.41**	**53.2±0.52**	**0.56±0.05**	**0.84±0.04**

Method	clip	info	paint	quick	real	sketch	avg
ARM	49.7(0.3)	16.3(0.5)	40.9(1.1)	9.4(0.1)	53.4(0.4)	43.5(0.4)	35.5
Meta-DMoE	63.5(0.2)	21.4(0.3)	51.3(0.4)	14.3(0.3)	62.3(1.0)	52.4(0.2)	44.2
Ours	**64.2(0.3)**	**23.6(0.4)**	**51.5(0.2)**	**15.2(0.3)**	**64.6(0.5)**	**54.1(0.4)**	**45.5**

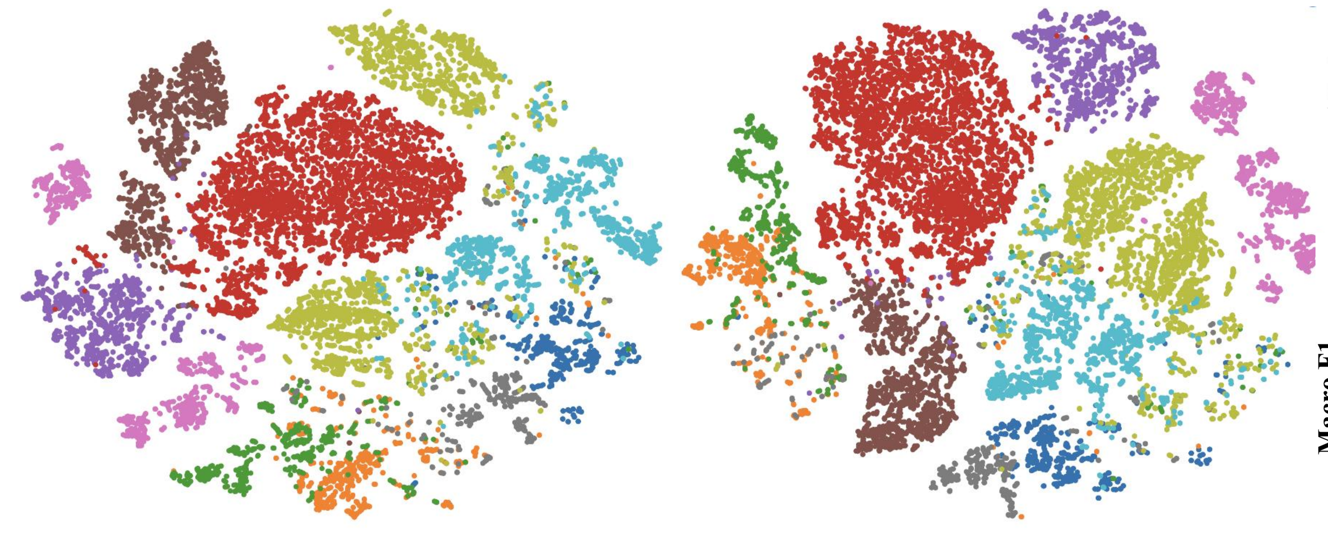
Adapted $(\tilde{\gamma}, \tilde{\beta})$	No adapt	Not-matched	Matched
Accuracy	74.69	72.39	**78.40**
Macro-F1	36.77	33.32	**38.27**

Method	Update BN Acc	Update BN Macro F1	Update Affine Acc	Update Affine Macro F1
TENT (min. entropy)	33.27	0.77	75.92	36.40
Our (min. auxiliary)	75.86	36.76	78.40	38.27
Our+TENT	75.84	31.93	79.68	38.85

Index	*SSL*	*Param.*	*TS*	*Adapt*	iWildCam Acc	iWildCam F1
1	✗	All	CE	✗	68.7	31.3
2	✓	All	Joint	✗	70.5	33.2
3	✓	BN	Joint	✓	68.2	30.5
4	✓	Aff	Joint	✓	71.1	33.9
5	✓	All	Meta	✓	72.0	29.4
6	✓	Aff	Meta	✗	74.7	36.8
7	✓	Aff	Meta	✓	78.4	38.3

Self-supervised	Backbone	Training	iWildCam Acc	iWildCam F1
None (baseline)	ResNet50	CE	68.7	31.3
Rotation (Sun et al. 2020)	ResNet50	Joint	69.2	31.5
Rotation (Sun et al. 2020)	ResNet50	Meta	72.8	33.0
MAE (He et al. 2022)	ViT-Base	Joint	71.7	33.8
MAE (He et al. 2022)	ViT-Base	Meta	74.9	35.1
Ours (BYOL)	ResNet50	Joint	70.5	33.2
Ours (BYOL)	ResNet50	Meta	78.4	38.3

➢ Visualization & Hyper-Parameter Analysis on the iWildCam Dataset



(a) Before adaptation (b) After adaptation