pSTarC: Pseudo Source Guided Target Clustering for Fully Test-Time Adaptation

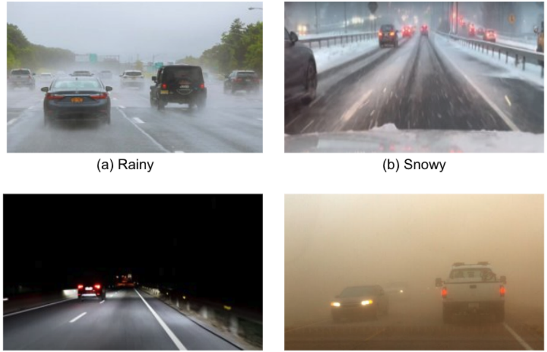
Manogna Sreenivas[†], Goirik Chakrabarty[*], Soma Biswas[†]

[†] Image Analysis and Computer Vision Lab,
Indian Institute of Science, Bengaluru, India
[*]Indian Institute of Science Education and Research, Pune, India

WACV 2024

Test Time Adaptation

Training domain



(a) Rainy (b) Snowy

(c) Night (d) Sandy

Test domains

Objective: Given a source trained model, adapt it to unseen domain shifts during test time.

pSTarC: **p**seudo-**S**ource Guided **Tar**get **C**lustering for Fully Test-Time Adaptation

DomainNet Dataset

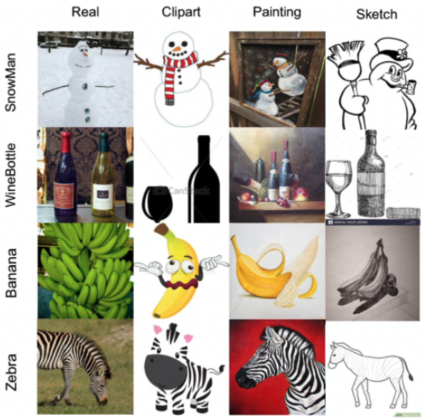


Figure: Samples from DomainNet dataset demonstrating real world domain shifts.

Domain Adaptation

- Online adaptation of models in real-time is increasingly important.
- Domain adaptation techniques aim to align distributions of training and testing data to improve model robustness.
- Four main topics in deep network robustness against distribution shifts:
  - Unsupervised Domain Adaptation (UDA)
  - Source Free Domain Adaptation (SFDA)
  - Test Time Adaptation (TTA)
  - Continuous Test Time Adaptation (CTTA)

Domain Adaptation Protocols

Table: Domain adaptation protocols

Setting	Source-free	Adaptation protocol		Target domain	
		Offline	Online	Single	Continuous
UDA		✓		✓	
SFDA	✓	✓		✓	
TTA	✓		✓	✓	
CTTA	✓		✓		✓

Test Time Adaptation

Given an off-the shelf model parameterized by θ, the objective of TTA is to adapt it using test batches $\mathbf{x}_t$ arriving in an online manner from a test domain $\mathcal{D}_{test} \neq \mathcal{D}_{train}$ by minimizing a test time objective as

$$\arg\min_{\theta} \mathcal{L}_{test}(\mathbf{x}_t; \theta) \tag{1}$$

SFDA vs TTA

- SFDA methods:
  - Leverage abundant target domain samples.
  - Employ clustering objectives.
    Attracting and Dispersing (AaD) [1]:

$$\mathcal{L}(x_i) = -\sum_{p_j \in \mathcal{N}i} p_i^T p_j + \lambda \sum_{x_m \in \mathbf{x}_t} p_i^T p_m$$

- TTA methods:
  - Classifier is fixed to preserve disciminative information learned from source.
  - Pseudo labeling, Entropy minimization[2] objectives employed to optimize a small set of network parameters.
- Can we employ SFDA objectives in TTA?

[1] S. Yang et al., "Attracting and dispersing: A simple approach for source-free domain adaptation", NeurIPS 2022

[2] D. Wang et al., "Tent: Fully test-time adaptation by entropy minimization", ICLR 2021

Proposed Method

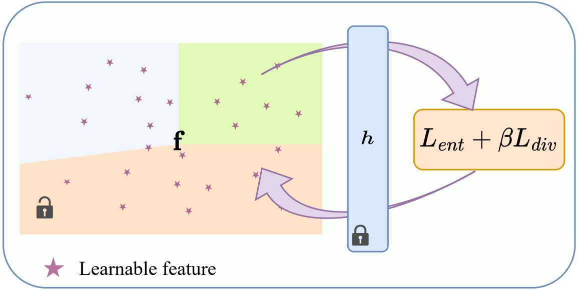
pSTarC: **p**seudo-**S**ource Guided **Tar**get **C**lustering for Fully Test-Time Adaptation

1. Pseudo-source Feature Generation:
    - The source trained classifier defined the decision boundaries.
    - As the classifier is fixed during adaptation, can we leverage this to synthesize pseudo source features?
2. Pseudo-source guided Target Clustering:
    - Leverage the generated pseudo source samples to effectively cluster the test data.

Pseudo Source Feature Generation

- Randomly initialize a feature bank $\mathbf{f} \in \mathcal{R}^{N \times d}$, $N = C \times n_c$.
- Optimize $\mathbf{f}$ using entropy minimization and diversity maximization loss.

$$\mathcal{L}_{ent}(\mathbf{f}; h) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{C} p_k \log p_k; \qquad \mathcal{L}_{div}(\mathbf{f}; h) = \sum_{k=1}^{C} \hat{p}_k \log \hat{p}_k$$



Pseudo Source features

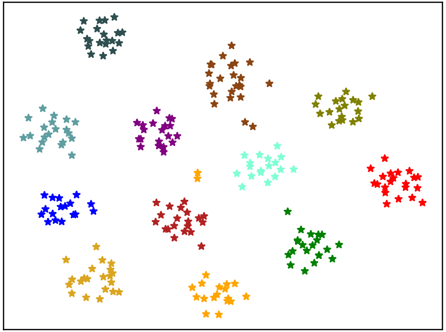
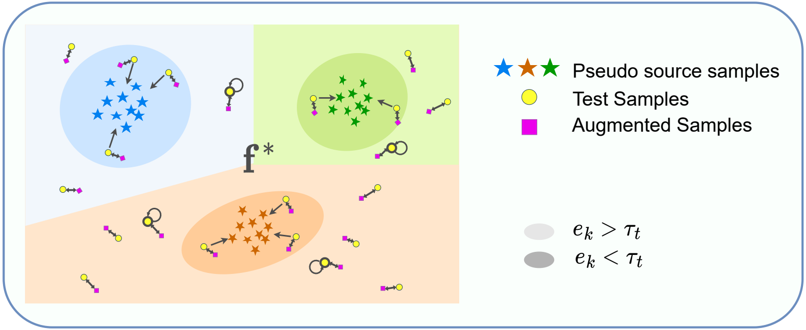


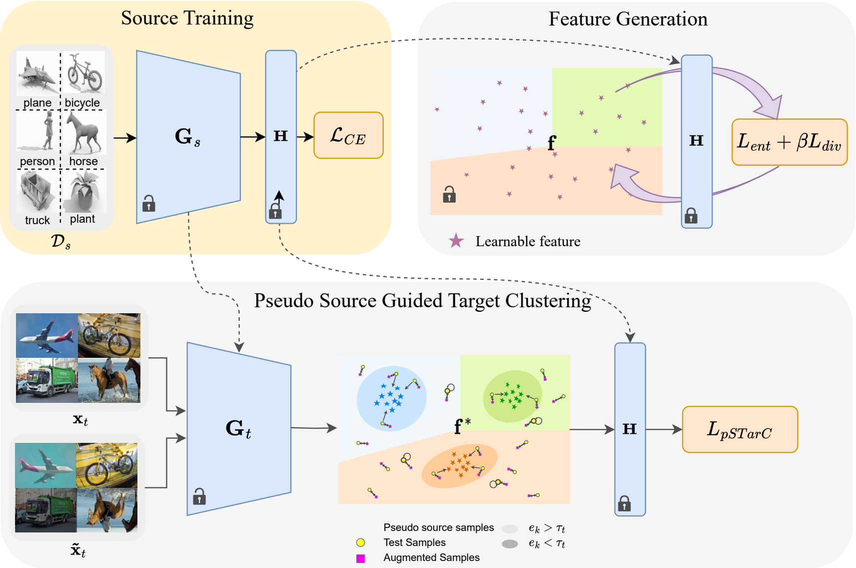
Figure: Features Generated setting C to 12 and n_c to 20 for VisDA dataset.

Pseudo Source Guided Target Clustering



$$\mathcal{L}_{\text{pSTarC}}(x_k) = \underbrace{-\, p_k^T \tilde{p}_k}_{L_{aug}} \underbrace{-\sum_{p_j^+ \in \mathbf{p}^+} p_k^T p_j^+}_{L_{attr}} + \lambda \underbrace{\sum_{x_j \in \mathbf{x}_t} p_k^T p_j}_{L_{disp}}$$

pSTarC



Source Training
plane
bicycle
person
horse
truck
plant
$\mathcal{D}_s$
$\mathbf{G}_s$
H
$\mathcal{L}_{CE}$
Feature Generation
f
H
$L_{ent} + \beta L_{div}$
Learnable feature
Pseudo Source Guided Target Clustering
$\mathbf{x}_t$
$\tilde{\mathbf{x}}_t$
$\mathbf{G}_t$
$\mathbf{f}^*$
H
L_{pSTarC}
Pseudo source samples
$e_k > \tau_t$
Test Samples
$e_k < \tau_t$
Augmented Samples

Experimental Results

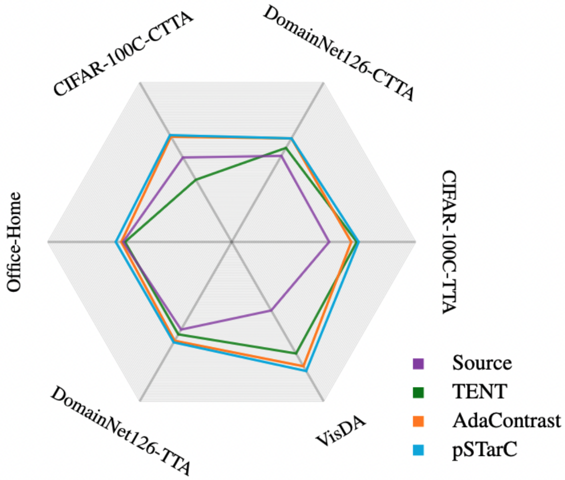


Figure: Comparison of pSTarC with prior fully TTA methods.

Ablation Study

L_{aug}	L_{attr}	L_{disp}	VisDA	DomainNet-126
✓	✓		68.8	58.8
✓		✓	78.2	59.7
	✓	✓	80.0	63.0
✓	✓	✓	**81.9**	**63.7**

Table: Ablation study on importance of each loss term.

Performance on varying batch sizes

Method	Batch size 8	16	32	64	128	Average
TENT	38.8	55.4	58.6	59.1	58.9	54.2
AdaContrast	50.1	57.9	60.8	62.4	62.4	58.7
pSTarC	54.1	59.2	61.3	63.8	63.7	**60.4**

Table: Performance on varying batch size on DomainNet-126 dataset

Complexity Analysis

Method	AdaContrast	Source-Proxy-TTA	C-SFDA	pSTarC
#Parameters	86M	43M	86M	43M
Memory	4.67M	3.76M	-	0.03M
#Forward	3	3	13	2
#Backward	1	1	1	1

Table: Complexity Analysis of TTA methods on VisDA

Conclusion

- In pSTarC framework, we propose a simple and efficient way to leverage fixed source classifier to generate pseudo source samples.
- Pseudo source samples generated, acting like a proxy for the labeled training data, can be effectively used to aid clustering the test samples during TTA.
- Experimental evidence on diverse datasets and setting including TTA and CTTA justify the effectiveness of the proposed pSTarC framework.

Thank You!