Semi-Supervised Domain Adaptation with Source Label Adaptation

Yu-Chu Yu Hsuan-Tien Lin

National Taiwan University

國立臺灣大學 National Taiwan University

Thu-PM-334

Semi-Supervised Domain Adaptation with Source Label Adaptation

Yu-Chu Yu   Hsuan-Tien Lin

National Taiwan University

國立臺灣大學 National Taiwan University

Thu-PM-334

Semi-Supervised Domain Adaptation

A bunch of source data

A few labeled target data

$\xRightarrow{\text{Adaptation}}$

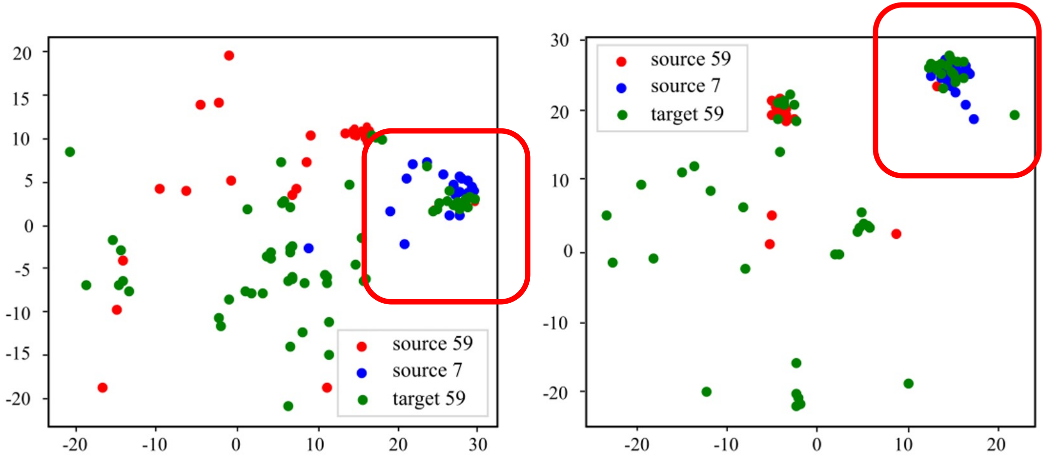
Many unlabeled target data

Goal:

- Extract invariant features across both domains
- Transfer knowledge from a source domain to another target domain

Challenge

- Domain shift
    - There is a misalignment between the 7th class of the source data and the 59th class of the target data.

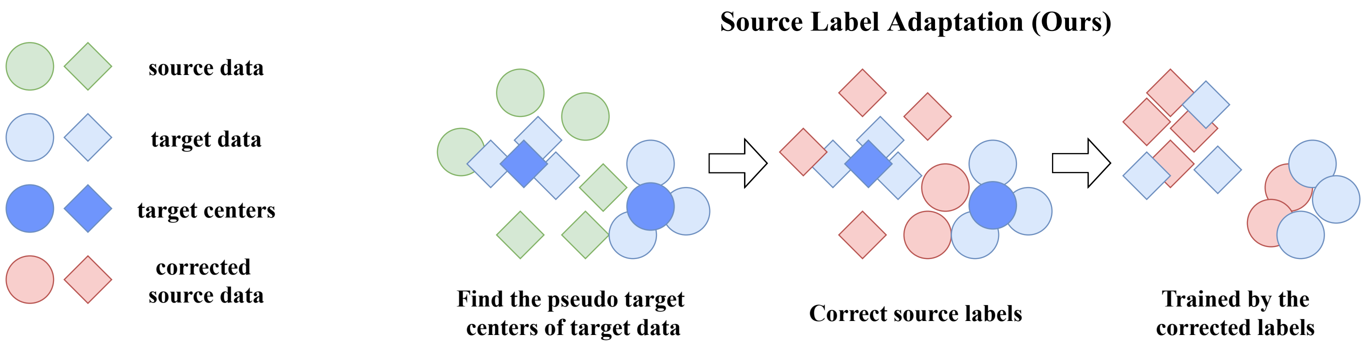


0-th iteration   5000-th iteration

Proposed Framework

Source Label Adaptation (SLA)

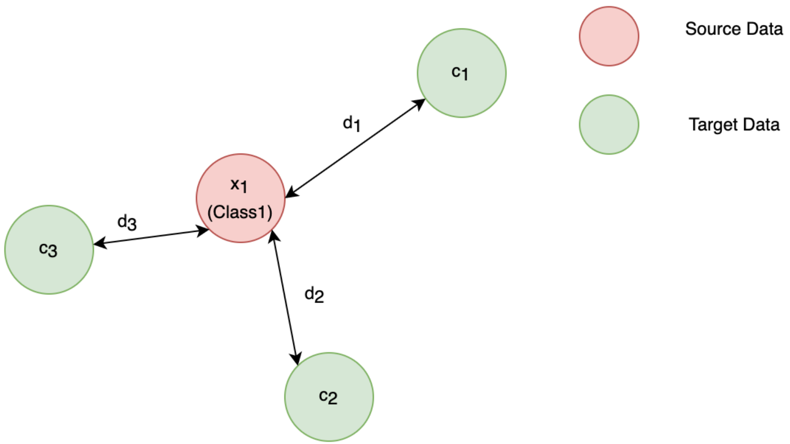
- A novel source-adaptive paradigm for Semi-Supervised Domain Adaptation.



Key Ideas

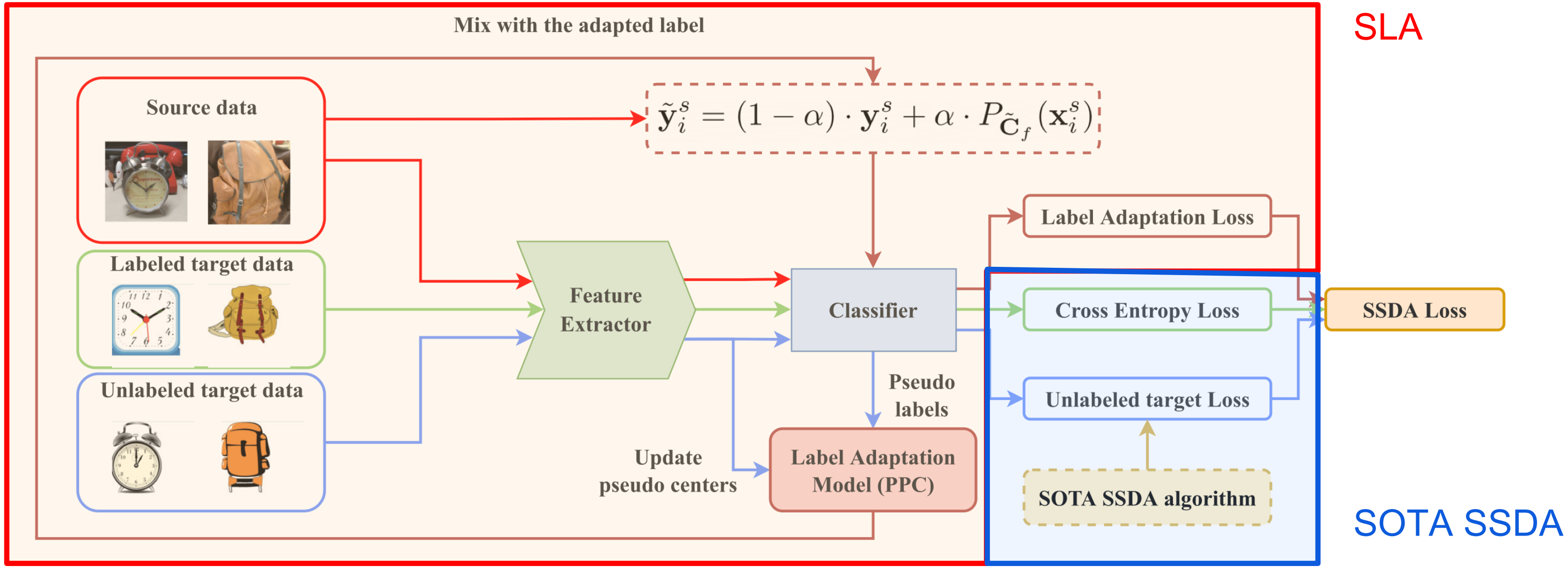
- View the source data as a version of the target data with noisy labels

- Correct the source labels with the estimated target centers in the current feature space.

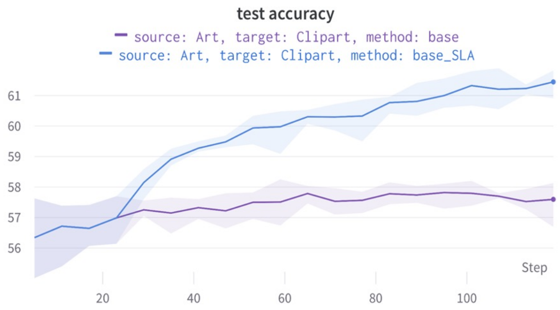


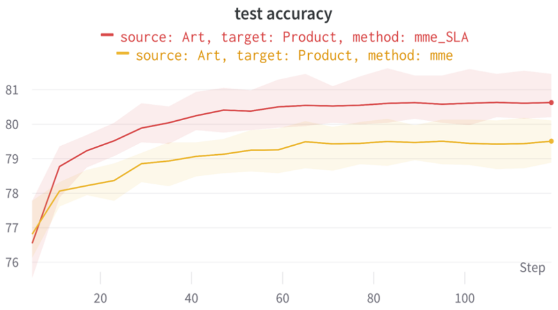
Source Label Adaptation (SLA)

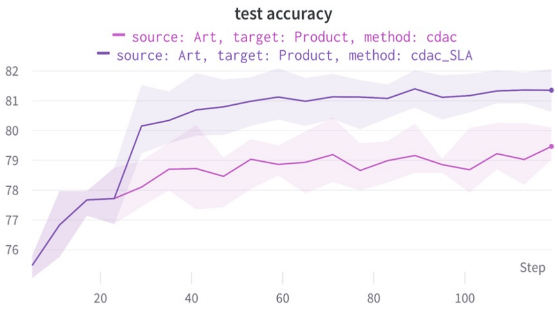
- The framework can be easily coupled with the current SOTA SSDA methods.

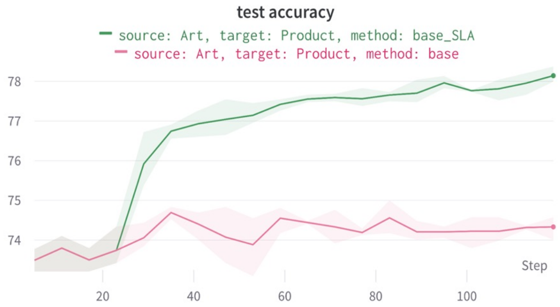


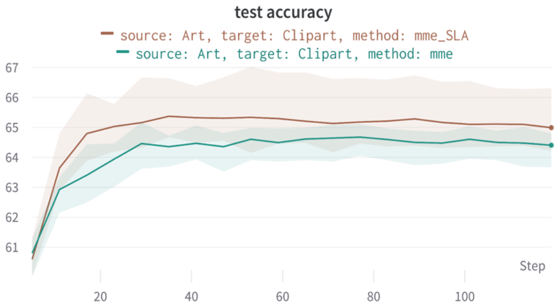
Experiment

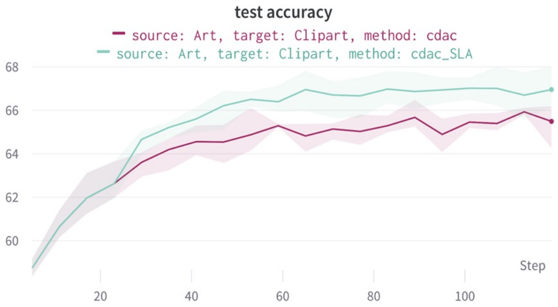

test accuracy
source: Art, target: Clipart, method: base
source: Art, target: Clipart, method: base_SLA
Step


test accuracy
source: Art, target: Product, method: mme_SLA
source: Art, target: Product, method: mme
Step


test accuracy
source: Art, target: Product, method: cdac
source: Art, target: Product, method: cdac_SLA
Step

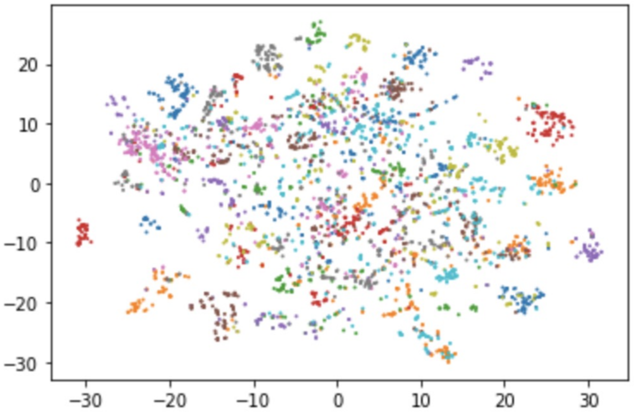

test accuracy
source: Art, target: Product, method: base_SLA
source: Art, target: Product, method: base
Step


test accuracy
source: Art, target: Clipart, method: mme_SLA
source: Art, target: Clipart, method: mme
Step

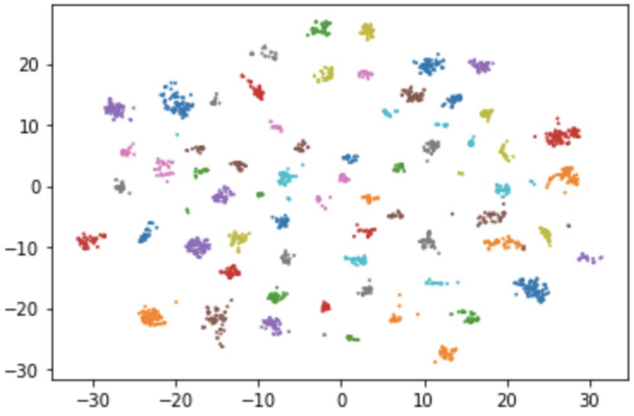

test accuracy
source: Art, target: Clipart, method: cdac
source: Art, target: Clipart, method: cdac_SLA
Step

Motivation

- Goal: Find an ideal model g^* that can minimize the target risk
- For each source data x^s, $g^*(x^s)$ is the **most suitable label** that best matches the ideal target space.



$g^*(x^s)$: The most suitable label for source data in the ideal target space



$g^*(x^t)$: the ideal target space

Source Label Adaptation

- We propose to adapt the original source label y^s to the ideal label $g^*(x^s)$.

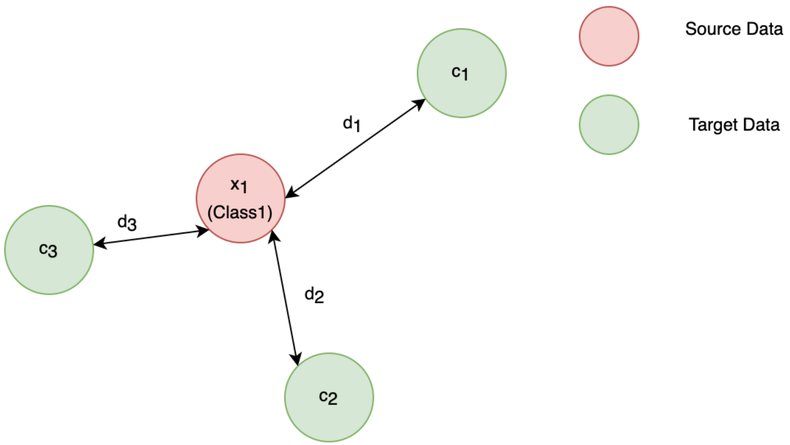
$$\underset{\text{Noisy}}{\mathbf{y}^s} \xrightarrow{\text{Label Adaptation}} \underset{\text{Clean}}{g^*(\mathbf{x}^s)}$$

- However, we are not able to access the ideal model g^*.
    - Approximate it through the current estimation of the unlabeled target data

Prototypical Network (Protonet)

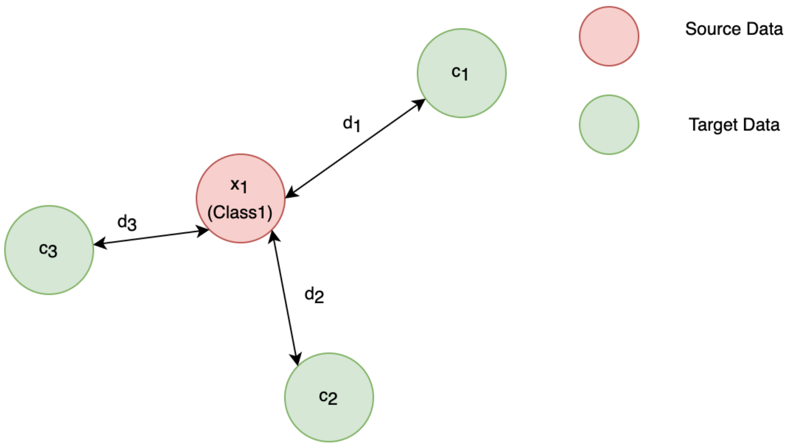
- Find the center c_k of class k over a certain feature space.
- Make predictions by the distance between the data point and each center.

$$P(\mathbf{x})_k = \frac{\exp(-d(f(\mathbf{x}), \mathbf{c}_k))}{\sum_{j=1}^{K} \exp(-d(f(\mathbf{x}), \mathbf{c}_j))}$$



Protonet with Target Centers

- We have access to a few target data.
  - Protonet with Target Centers



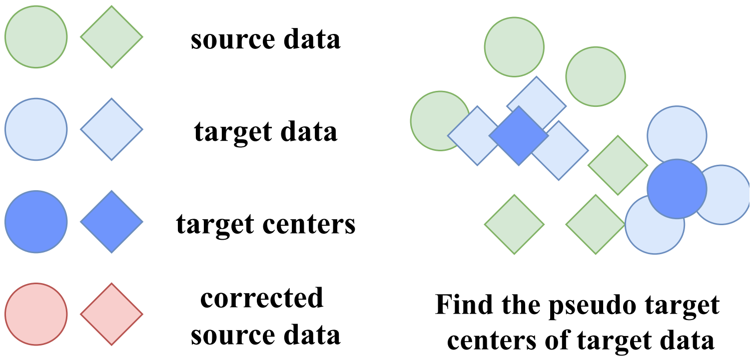
- Challenge
  - We have only 1 or 3 shot per class
  - The estimation might be inaccurate

Protonet with Pseudo Centers (PPC)

1. Determine the pseudo label for each unlabeled target data x_i^u.

$$\tilde{y}_i^u = \arg\max_k g(\mathbf{x}_i^u)_k$$

2. Find the pseudo center c_k for each class k, and construct a prototypical network P based on these centers.



Distance Comparison

From / To	labeled target centers	pseudo centers
ideal centers	10.02	4.06

Table 3. Average L2 Distance from ideal centers to labeled target centers / pseudo centers over the feature space trained by S+T (3-shot *Office-Home* A $\rightarrow$ C with ResNet34).

Label Adaptation Loss

- Protonet with Pseudo Centers is still an estimation of the target view.
- We introduce a hyper-parameter α to regularize the level of trust to this estimation.
- The adapted label $\tilde{y}_i^s$ is defined as follow:

$$\boxed{\tilde{\mathbf{y}}_i^s} = (1-\alpha) \cdot \boxed{\mathbf{y}_i^s} + \alpha \cdot \boxed{P(\mathbf{x}_i^s)}$$

adapted label

suggested adapted label from PPC

original label

- We propose a label adaptation loss to replace the typical source loss function.
    - H measures the cross entropy between two distributions.

$$\tilde{\mathcal{L}}_s(g|S) = \frac{1}{|S|} \sum_{i=1}^{|S|} H(g(\mathbf{x}_i^s), \tilde{\mathbf{y}}_i^s)$$

Combine with SOTA SSDA Algorithms

- Typical SSDA algorithms usually attempt to explore better use of the unlabeled target data.

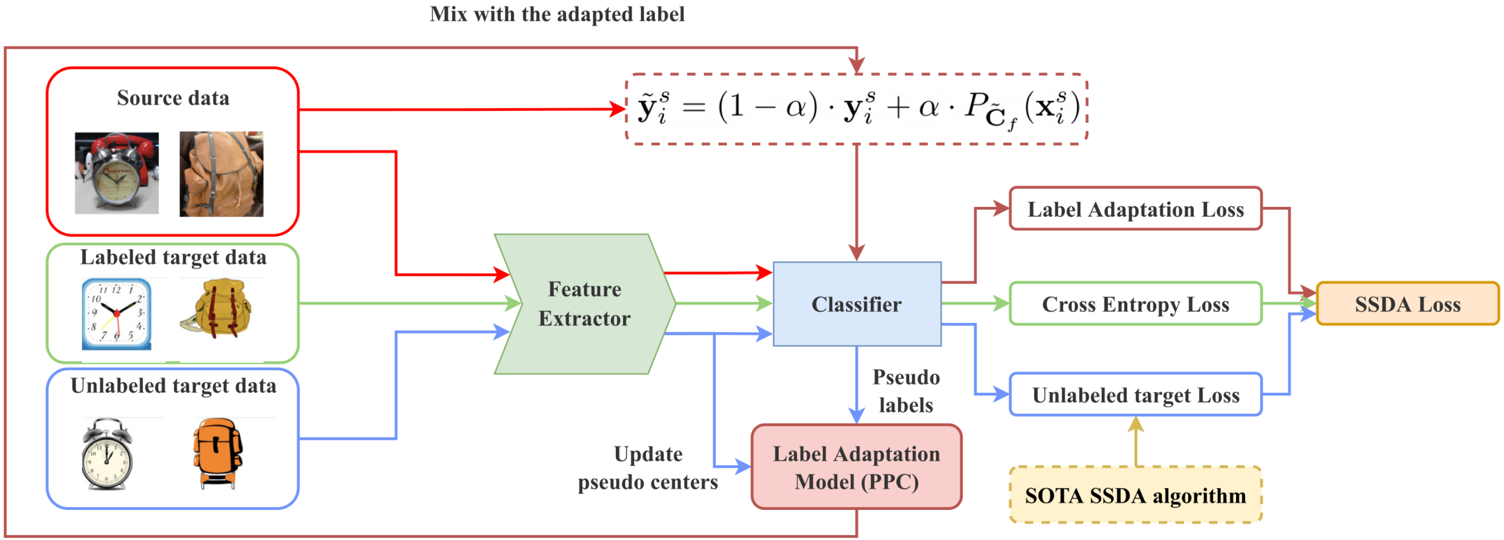
$$\mathcal{L}_{\text{SSDA}} = \mathcal{L}_s + \mathcal{L}_\ell + \boxed{\mathcal{L}_u}$$

- Our framework, on the other hand, explores the training of source data with adapted labels to better align with the ideal target space.

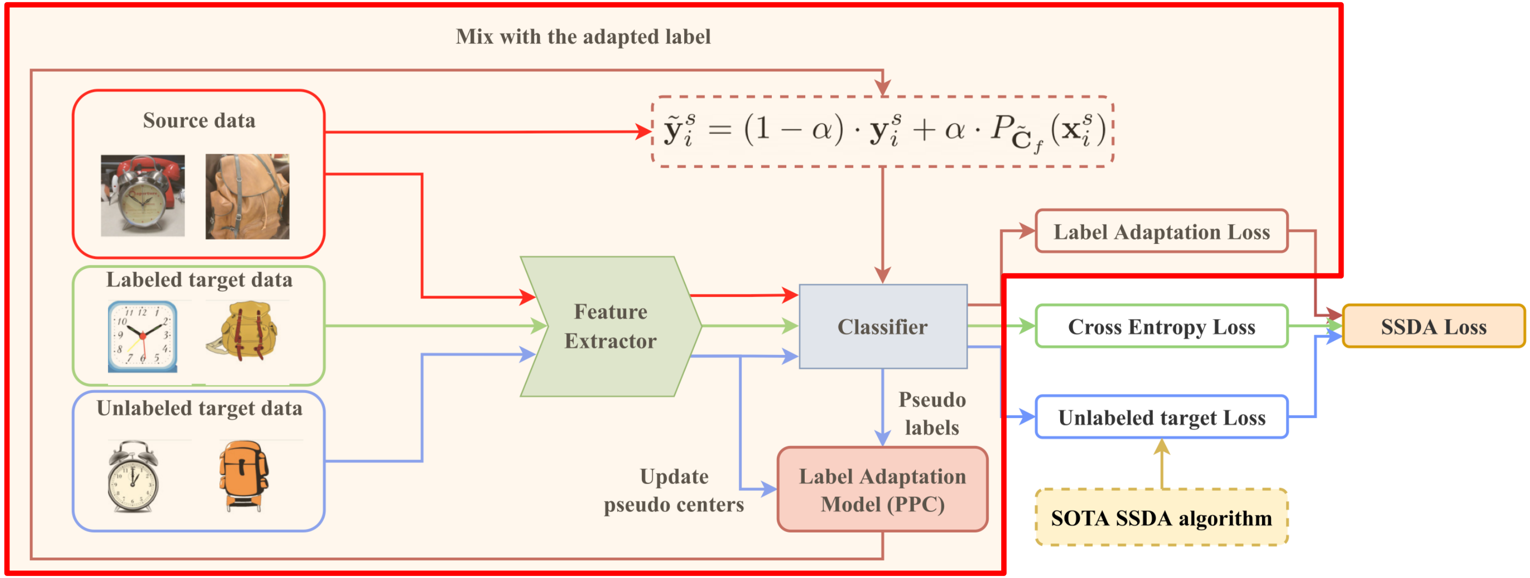
$$\mathcal{L}_{\text{SSDA w/ SLA}} = \boxed{\tilde{\mathcal{L}}_s} + \mathcal{L}_\ell + \mathcal{L}_u$$

- Thus, we can easily apply our framework to other SSDA algorithms, further boost their performance.

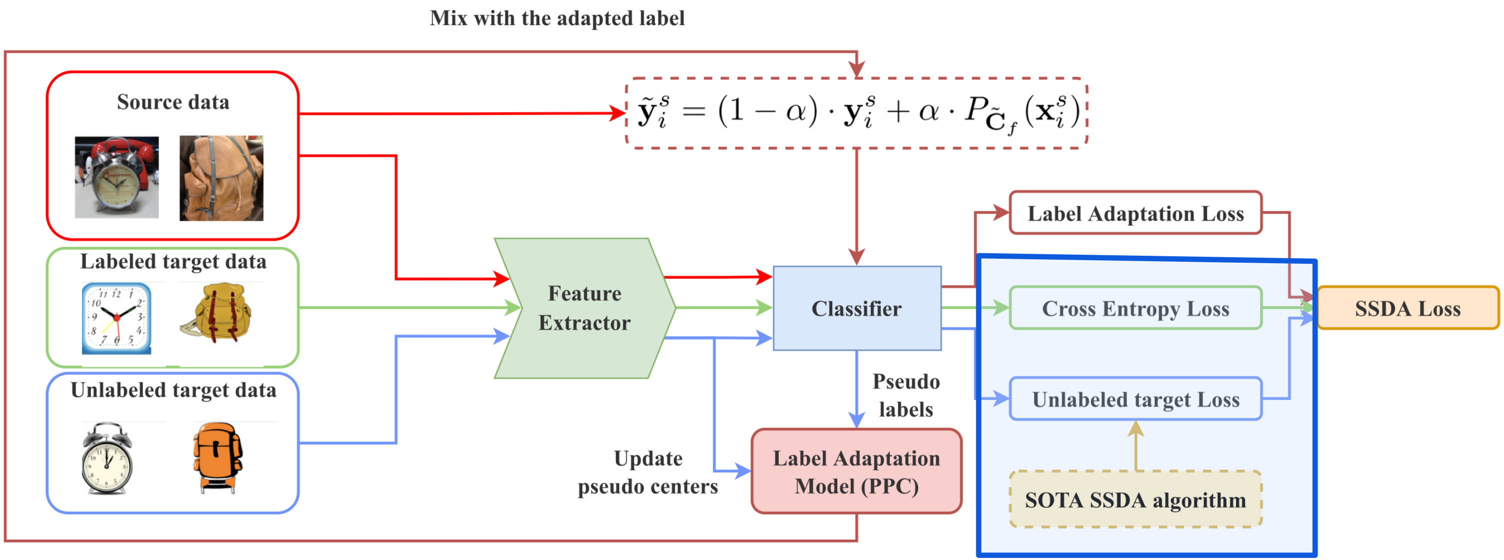
Source Label Adaptation (SLA)



Source Label Adaptation (SLA)



Source Label Adaptation (SLA)



Implementation Details

- Warmup Stage
  - Our label adaptation framework relies on the quality of the predicted pseudo labels.
  - The prediction from the initial model can be noisy.
  - We introduce a warmup stage W to obtain more stable pseudo labels.

$$\tilde{\mathbf{y}}_i^s = \begin{cases} \mathbf{y}_i^s & \text{if } e \leq W \\ (1-\alpha)\cdot \mathbf{y}_i^s + \alpha \cdot P_{\tilde{\mathbf{C}}_f}(\mathbf{x}_i^s) & \text{otherwise} \end{cases}$$

Implementation Details

- Dynamic Update
  - During training phase, the feature space keeps changing for every iteration.
  - Without updating centers, the quality of the estimated pseudo centers would progressively deteriorate.
  - At certain intervals, we re-estimate the pseudo labels and centers over current feature space.

Experiments on Major SSDA Datasets

Method	R → C		R → P		P → C		C → S		S → P		R → S		P → R		**Mean**	
	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	**1-shot**	**3-shot**
S+T	55.6	60.0	60.6	62.2	56.8	59.4	50.8	55.0	56.0	59.5	46.3	50.1	71.8	73.9	56.9	60.0
DANN [5]	58.2	59.8	61.4	62.8	56.3	59.6	52.8	55.4	57.4	59.9	52.2	54.9	70.3	72.2	58.4	60.7
ENT [6]	65.2	71.0	65.9	69.2	65.4	71.1	54.6	60.0	59.7	62.1	52.1	61.1	75.0	78.6	62.6	67.6
APE [10]	70.4	76.6	70.8	72.1	72.9	76.7	56.7	63.1	64.5	66.1	63.0	67.8	76.6	79.4	67.6	71.7
DECOTA [31]	79.1	80.4	74.9	75.2	76.9	78.7	65.1	68.6	72.0	72.7	69.7	71.9	79.6	81.5	73.9	75.6
MCL [30]	77.4	79.4	74.6	**76.3**	75.5	78.8	66.4	70.9	**74.0**	**74.7**	70.7	72.3	**82.0**	**83.3**	74.4	76.5
MME [21]	70.0	72.2	67.7	69.7	69.0	71.7	56.3	61.8	64.8	66.8	61.0	61.9	76.1	78.5	66.4	68.9
MME + SLA (ours)	71.8	73.3	68.2	70.1	70.4	72.7	59.3	63.4	64.9	67.3	61.8	63.9	77.2	79.6	68.8	70.0
CDAC [12]	77.4	79.6	74.2	75.1	75.5	79.3	67.6	69.9	71.0	73.4	69.2	72.5	80.4	81.9	73.6	76.0
CDAC + SLA (ours)	**79.8**	**81.6**	**75.6**	76.0	**77.4**	**80.3**	**68.1**	**71.3**	71.7	73.5	**71.7**	**73.5**	80.4	82.5	**75.0**	**76.9**

Table 4. Accuracy (%) on *DomainNet* for 1-shot and 3-shot Semi-Supervised Domain Adaptation (ResNet34).

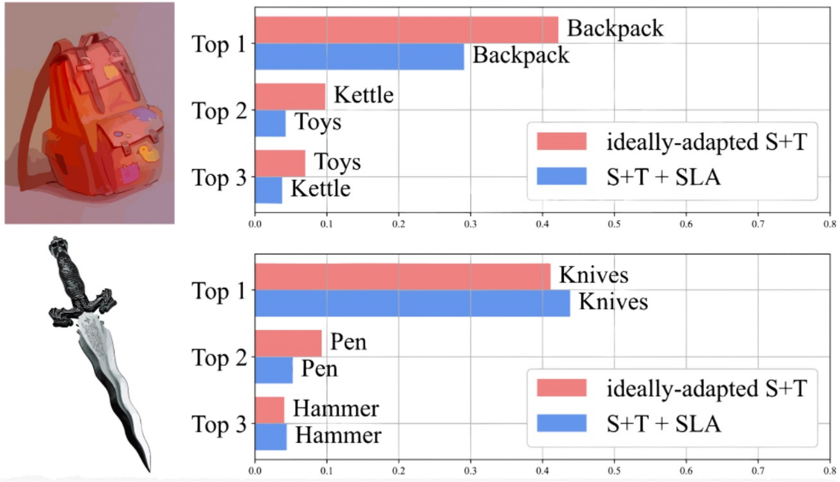
Method	A→C	A→P	A→R	C→A	C→P	C→R	P→A	P→C	P→R	R→A	R→C	R→P	**Mean**
						Three-shot							
S+T	54.0	73.1	74.2	57.6	72.3	68.3	63.5	53.8	73.1	67.8	55.7	80.8	66.2
DANN [5]	54.7	68.3	73.8	55.1	67.5	67.1	56.6	51.8	69.2	65.2	57.3	75.5	63.5
ENT [6]	61.3	79.5	79.1	64.7	79.1	76.4	63.9	60.5	79.9	70.2	62.6	85.7	71.9
APE [10]	63.9	81.1	80.2	66.6	79.9	76.8	66.1	65.2	82.0	73.4	66.4	86.2	74.0
DECOTA [31]	64.0	81.8	80.5	68.0	**83.2**	79.0	69.9	68.0	82.1	74.0	**70.4**	**87.7**	75.7
MME [21]	63.6	79.0	79.7	67.2	79.3	76.6	65.5	64.6	80.1	71.3	64.6	85.5	73.1
MME + SLA (ours)	65.9	81.1	80.5	**69.2**	81.9	79.4	69.7	67.4	81.9	**74.7**	68.4	87.4	75.6
CDAC [12]	65.9	80.3	80.6	67.4	81.4	**80.2**	67.5	67.0	81.9	72.2	67.8	85.6	74.8
CDAC + SLA (ours)	**67.3**	**82.6**	**81.4**	**69.2**	82.1	80.1	**70.1**	**69.3**	**82.5**	73.9	70.1	87.1	**76.3**

Table 5. Accuracy (%) on *Office-Home* for 1-shot and 3-shot Semi-Supervised Domain Adaptation (ResNet34).

Adapted Labels

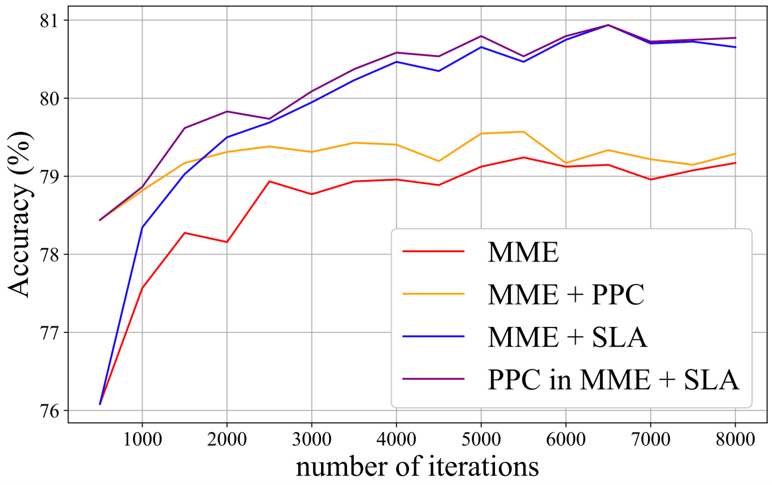
- For the backpack case, SLA suggests to adapt the label from 100% backpack to:
  - 30% Backpack
  - 5% Toys
  - 4% Kettle.

- The adapted labels are much closer to the ideally-adapted labels ($g^*(x^s)$).



The Intermediate Results in SLA

- PPC is actually a strong model that has performed well on the target domain at the early stage.
- However, without updating the source labels, it will end up converge to the same performance as the original method.
- On the other hand, in our SLA framework, the model leverages the benefits of PPC, resulting in better performance.



Conclusion

- General framework
    - Source Label Adaptation for Semi-Supervised Domain Adaptation
- Rethinking the usage of source data
    - Approach Domain Adaptation as a Noisy Label Learning problem.
- Empirical Improvement
    - Our method improve 2 representative SSDA algorithms on 2 major datasets for both 1-shot and 3-shot settings.

Visit our project page for more details!

Code is available