Dual Moving Average Pseudo-Labeling for Source-Free Inductive Domain Adaptation

Carleton University

Hao Yan, Yuhong Guo

BMVC 2022

SFIDA

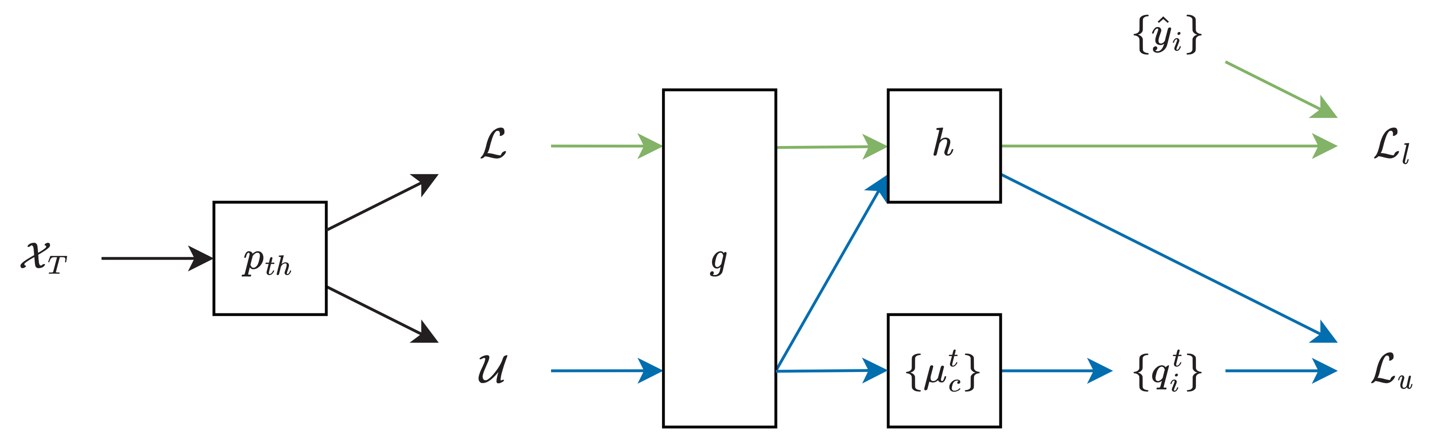
UDA vs. SFDA

- **Unsupervised Domain Adaptation (UDA)**: learn a target model given labeled source data and unlabeled target data.
- **Source-Free Domain Adaptation (SFDA)**: learn a target model given pretrained source model and unlabeled target data.
- Source-free setting preserves data privacy and avoids storing and transferring large amount of data.

Transductive vs. Inductive

- **Transductive**: target model is trained on the training set and evaluated on the training set.
- **Inductive**: target model is trained on the training set and evaluated on the testing set.
- Inductive setting evaluates methods in terms of the generalization ability on unseen test data.

Framework



- The target training data $\mathcal{X}_T$ are first split into a pseudo-labeled confident subset ($\mathcal{L}$) and a less-confident unlabeled subset ($\mathcal{U}$) based on the pre-trained source model.
- The pseudo-labeled confident subset acts as trusty supervision to prevent over adaptation.
- The unlabeled subset is gradually updated to fine-tune the prediction model ($f = h \circ g$) through the proposed dual moving average update.

Results

Table 1: Test accuracy (%) on DomainNet dataset (ResNet-101). SF means source-free.

Methods	SF	c→p	c→r	c→s	p→c	p→r	p→s	r→c	r→p	r→s	s→c	s→p	s→r	Avg.
ResNet-101	-	37.9	53.4	44.2	44.1	57.0	38.6	50.9	48.8	37.7	52.8	37.3	47.6	45.9
AdaMatch	✗	45.3	56.0	60.2	35.3	47.6	42.9	46.5	48.1	49.1	46.5	41.0	42.4	46.7
MCC	✗	37.7	55.7	42.6	45.4	59.8	39.9	54.4	53.1	37.0	58.1	46.3	56.2	48.9
CDAN	✗	40.4	56.8	46.1	45.1	58.4	40.5	55.6	53.6	43.0	57.2	46.4	55.7	49.9
CDAN+SDAT	✗	41.5	57.5	47.2	47.5	58.0	41.8	56.7	53.6	43.9	58.7	48.1	57.1	51.0
SHOT	✓	45.6	63.4	49.1	35.1	64.1	21.0	57.1	51.1	44.0	61.2	47.6	62.0	48.4
SSFT-SSD	✓	41.9	57.5	46.5	47.6	59.6	42.6	55.4	51.9	42.0	58.4	45.2	55.7	50.4
DMAPL (Ours)	✓	**46.0**	**63.7**	**49.1**	**53.2**	**64.2**	**46.0**	**61.6**	**55.4**	**47.8**	**64.1**	**50.3**	**63.5**	**55.4**
Oracle	-	71.1	83.4	70.0	78.4	83.4	70.0	78.4	71.1	70.0	78.4	71.1	83.4	75.7

Table 2: Test accuracy (%) on VisDA2017Split dataset (ResNet-101). SF means source-free.

Methods	SF	plane	bcycl	bus	car	horse	knife	mcycl	person	plant	sktbrd	train	truck	Macro	Micro
ResNet-101	-	76.7	23.9	48.1	68.0	67.8	6.5	86.0	20.6	71.8	23.9	85.0	8.4	48.9	54.1
CDAN	✗	92.7	73.5	80.0	46.4	90.2	93.2	86.1	78.4	83.8	87.3	83.2	38.3	77.8	73.7
MCC	✗	92.2	79.4	79.0	71.7	92.1	93.0	89.9	79.0	88.2	91.0	82.1	50.8	82.4	80.0
SHOT	✓	77.7	**85.8**	80.2	54.2	90.2	63.4	**82.1**	73.5	88.9	80.5	83.1	54.8	76.2	73.8
SSFT-SSD	✓	94.5	84.9	**80.9**	49.9	91.2	66.8	77.0	75.4	81.3	86.2	**89.4**	50.4	77.3	73.6
DMAPL (Ours)	✓	**95.6**	84.5	78.9	**58.7**	**92.4**	**96.6**	80.8	**82.5**	**90.3**	**88.6**	87.8	**59.1**	**83.0**	**79.1**
Oracle	-	98.2	94.7	89.5	88.0	98.7	96.4	93.6	92.8	98.0	96.5	93.4	72.6	92.7	91.5

Target Data Splitting

Target data are firstly split into confident subset $\mathcal{L}$ and less-confident subset $\mathcal{U}$ based on the pre-trained source model.

$$\mathcal{L} = \left\{ \mathbf{x}_i \in \mathcal{X}_T \,|\, \max_y p(y|\mathbf{x}_i;\theta_{fS}) \geq p_{th} \right\}, \quad \mathcal{U} = \mathcal{X}_T \backslash \mathcal{L}. \tag{1}$$

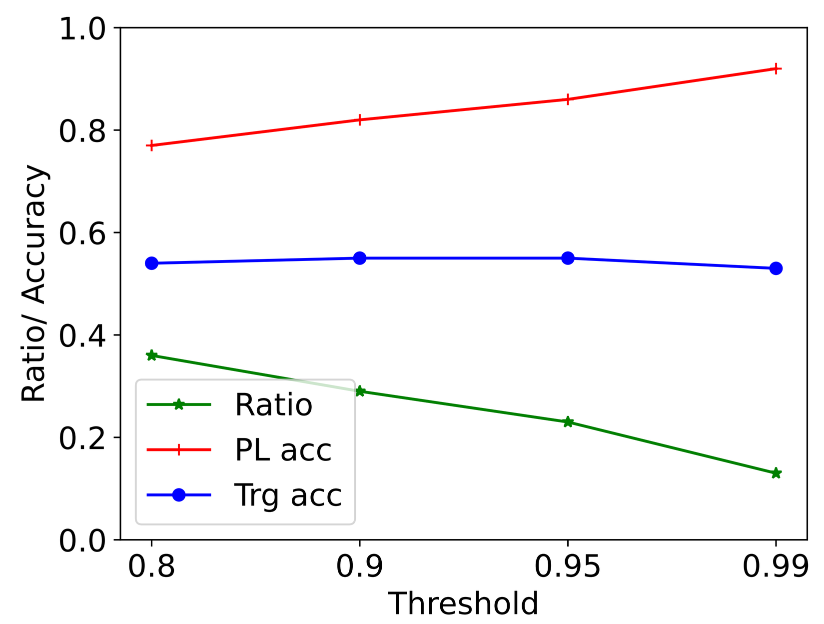
Pseudo-labels are assigned to the confident instances.

$$\hat{y}_i = \arg\max_y \, p(y|\mathbf{x}_i;\theta_{fS}), \quad \mathbf{x}_i \in \mathcal{L}. \tag{2}$$

The pseudo-labeled confident subset $\mathcal{L}$ is used as trusty supervision.

$$\mathcal{L}_l = \mathbb{E}_{\mathbf{x}_i \in \mathcal{L}} \left[-\log(p(\hat{y}_i|\mathbf{x}_i;\theta_f)) \right], \tag{3}$$

Higher threshold p_{th} leads to more accurate pseudo-labels but less amount of confident instances, however the final accuracy values of the learnt target model are similar.



Dual Moving Average based Model Fine-Tuning

Denote the normalized feature vector as $\mathbf{z}_i = g(\mathbf{x}_i)/\|g(\mathbf{x}_i)\|_2$. We calculate the feature mean of the c-th class in the current iteration t as,

$$\mathbf{v}_c^t = \frac{\sum_{\mathbf{x}_i \in (X_l \cup X_u)} \mathbb{1}(\bar{y}_i = c) \cdot \mathbf{z}_i}{\sum_{\mathbf{x}_i \in (X_l \cup X_u)} \mathbb{1}(\bar{y}_i = c)}, \tag{4}$$

where

$$\bar{y}_i = \begin{cases} \hat{y}_i, & \mathbf{x}_i \in \mathcal{L} \\ \arg\max_y p(y|\mathbf{x}_i;\theta_f), & \mathbf{x}_i \in \mathcal{U} \end{cases} \tag{5}$$

We then calculate the centroid μ_c^t of the prototypical classifier for the current iteration t as the weighted average of the centroid μ_c^{t-1} from the previous iteration and the feature mean $\mathbf{v}_c^t$ in the current iteration,

$$\mu_c^t = \text{Normalize}(\alpha\mu_c^{t-1} + (1-\alpha)\mathbf{v}_c^t), \tag{6}$$

The prototypical classifier assigns a new one-hot label vector $\tilde{\mathbf{y}}_i^t$ to each unlabeled instance $\mathbf{x}_i \in X_u$ as follow,

$$(\tilde{\mathbf{y}}_i^t)_j = \begin{cases} 1, & j = \arg\max_{c \in \{1,\cdots,C\}} \mathbf{z}_i^\top \mu_c^t, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

This newly assigned pseudo-label vector is further used to update the soft-labels of the unlabeled subset in the following moving average manner.

$$\mathbf{q}_i^t = \beta \mathbf{q}_i^{t-1} + (1-\beta)\tilde{\mathbf{y}}_i^t, \tag{8}$$

The soft-label vectors for the instances in the unlabeled subset $\mathcal{U}$ are further used to fine-tune the target model f by minimizing the following *soft cross-entropy loss* in the t-th iteration:

$$\mathcal{L}_u = \mathbb{E}_{\mathbf{x}_i \in \mathcal{U}} \left[\sum_y -(\mathbf{q}_i^t)_y \log p(y|\mathbf{x}_i;\theta_f) \right] \tag{9}$$

By taking both subsets $\mathcal{L}$ and $\mathcal{U}$ into consideration, the overall loss minimization for the proposed semi-supervised fine-tuning method is shown as follows,

$$\min_{\theta_f} \mathcal{L}_u + \lambda \mathcal{L}_l, \tag{10}$$

The coefficient parameters α and β control the updating degrees for centroid and soft-label updates. Obviously slower updates are more beneficial for the proposed method, duo to better training stability.