When Source-Free Domain Adaptation Meets Learning with Noisy Labels

Li Yi*, Gezheng Xu*, Pengcheng Xu, Jiaqi Li, Ruizhi Pu
Charles Ling, A. Ian McLeod, Boyu Wang

University of Western Ontario

ICLR'23 - May 2023
Presenter: Gezheng Xu

UDA and SFDA

Unsupervised Domain Adaptation (UDA)

$$D_S = \{(\boldsymbol{x}_i^S, y_i^S)\}_{i=1}^N \;+\; D_T = \{(\boldsymbol{x}_i^T)\}_{i=1}^M$$

$$\Downarrow$$

$$f_T(\boldsymbol{x})$$

Source-Free Domain Adaptation (SFDA)

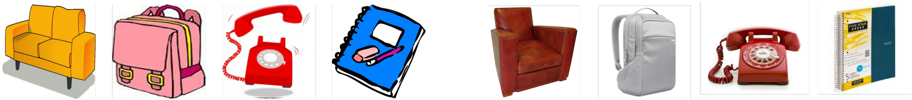
$$D_S = \{(\boldsymbol{x}_i^S, y_i^S)\}_{i=1}^N$$

$$\Downarrow \text{ Source Training}$$

$$f_S(\boldsymbol{x}) \;+\; D_T = \{(\boldsymbol{x}_i^T)\}_{i=1}^M$$

$$\Downarrow \text{ Target Adaptation}$$

$$f_T(\boldsymbol{x})$$



(a) Clipart: Clipart Images

(b) Real World: Regular Images captured with a Camera

Figure 1: Examples of Office-Home Dataset [1]: $p_S(X, Y) \neq p_T(X, Y)$

[1] Source: Venkateswara et al., Deep Hashing Network for Unsupervised Domain Adaptation. CVPR 2017.

Label Noise in SFDA

$$D_S = \{(\boldsymbol{x}_i^S, y_i^S)\}_{i=1}^N$$

$\Downarrow$ Source Training

$$f_S(\boldsymbol{x}) \; + \; D_T = \{(\boldsymbol{x}_i^T)\}_{i=1}^M$$

$\Downarrow$ Target Adaptation

$$f_T(\boldsymbol{x})$$

- **Two-Stage Training process**:
  Source Training $\Rightarrow$ **Target Adaptation**
- Key Point: Quality of the **Pseudo-Labels**
  - Domain Shift $\Rightarrow$ Severe Noise in Pseudo Labels
  - Incorrect Neighborhood/Cluster Information $\Rightarrow$ Noise Accumulation (Fig2)

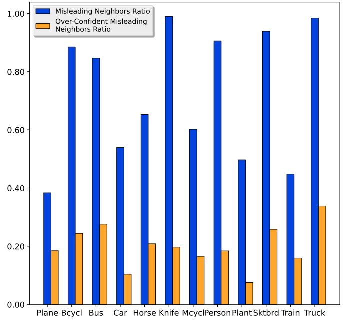


Figure 2: Neighbors Label Noise in SFDA Problem

Label Noise in SFDA

$$D_S = \{(\boldsymbol{x}_i^S, y_i^S)\}_{i=1}^N$$

$\Downarrow$ Source Training

$$f_S(\boldsymbol{x}) \;+\; D_T = \{(\boldsymbol{x}_i^T)\}_{i=1}^M$$

$\Downarrow$ Target Adaptation

$$f_T(\boldsymbol{x})$$

- Two-Stage Training process: Source Training $\Rightarrow$ Target Adaptation
- Key Point: Quality of the **Pseudo-Labels**
  - **Domain Shift $\Rightarrow$ Severe Noise in Pseudo Labels**
  - Incorrect Neighborhood/Cluster Information $\Rightarrow$ Noise Accumulation (Fig2)

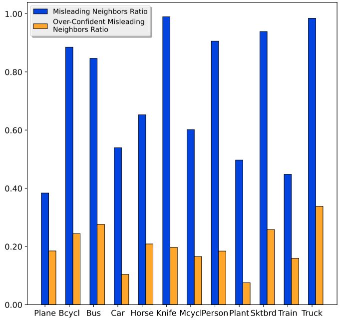


Figure 2: Neighbors Label Noise in SFDA Problem

Label Noise in SFDA

$$D_S = \{(\boldsymbol{x}_i^S, y_i^S)\}_{i=1}^N$$

$\Downarrow$ Source Training

$$f_S(\boldsymbol{x}) \;+\; D_T = \{(\boldsymbol{x}_i^T)\}_{i=1}^M$$

$\Downarrow$ Target Adaptation

$$f_T(\boldsymbol{x})$$

- Two-Stage Training process:
  Source Training $\Rightarrow$ Target Adaptation
- Key Point: Quality of the **Pseudo-Labels**
  - Domain Shift $\Rightarrow$ Severe Noise in Pseudo Labels
  - **Incorrect Neighborhood/Cluster Information** $\Rightarrow$ **Noise Accumulation** (Fig2)

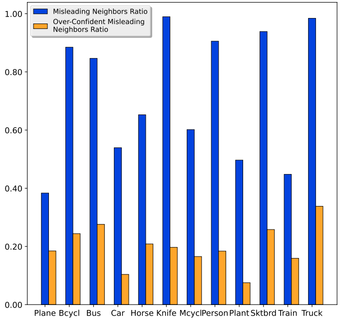


Figure 2: **Neighbors Label Noise in SFDA Problem**

Label Noise in SFDA

$$D_S = \{(\boldsymbol{x}_i^S, y_i^S)\}_{i=1}^N$$
$$\Downarrow \text{ Source Training}$$
$$\boxed{\begin{array}{c} f_S(\boldsymbol{x}) \; + \; D_T = \{(\boldsymbol{x}_i^T)\}_{i=1}^M \\ \Downarrow \text{ Target Adaptation} \\ f_T(\boldsymbol{x}) \end{array}}$$

- Two-Stage Training process:
  Source Training $\Rightarrow$ Target Adaptation
- Key Point: Quality of the **Pseudo-Labels**
  - Domain Shift $\Rightarrow$ Severe Noise in Pseudo Labels
  - Incorrect Neighborhood/Cluster Information
    $\Rightarrow$ Noise Accumulation (Fig2)

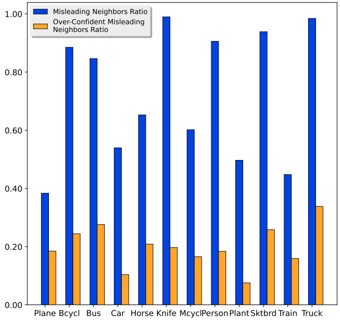


Figure 2: Neighbors Label Noise in SFDA Problem

$\Downarrow$

Incorrectly Assigned Pseudo Labels = Noisy Labels

We propose to **formulate SFDA as a Learning with Label Noise (LLN) problem.**

LLN

Learning with Label Noise (LLN)

- Given a set of **NOISY** training data $\bar{S}$
  - $\bar{S} = (\boldsymbol{x_i}, \bar{y}_i)_{i=1}^{n}$
  - $\boldsymbol{x_i}$: input data
  - $\bar{y}_i$: possibly corrupted label
  - y_i: ground-truth label
- To learn a **Noise-Robust** classifier
  $\Longrightarrow$ correctly label the new input data.

Figure 3: Example of Learning with Label Noise on Office-Home Dataset. The first row represents the ground-truth label; the second row is the possibly corrupted label.

Current Limitations of LLN methods in SFDA

Different Label Noises in LLN and in SFDA Settings

- Label Noise in **LLN** (Xiao et al., 2015):
  - generated by *human annotators* or *image search engines*
  - mislabeling rate for a sample is **bounded**
  - general LLN methods: **Noise-Robust** Losses
- Label Noise in **SFDA**:
  - generated by the *source model* due to the distribution shift
  - mislabeling rate can be out of control and **unbounded**

TWO **PROBLEMS** for *correctly* applying **LLN approaches** to **SFDA**:

1. Can general noise-robust LLN methods, based on the *Bounded Noise*, be effective for SFDA problems where the label noise has different properties?
2. If NOT, what kinds of LLN methods can be helpful?

Current Limitations of LLN methods in SFDA

Different Label Noises in LLN and in SFDA Settings

- Label Noise in **LLN** (Xiao et al., 2015):
  - generated by *human annotators* or *image search engines*
  - mislabeling rate for a sample is **bounded**
  - general LLN methods: **Noise-Robust** Losses
- Label Noise in **SFDA**:
  - generated by the *source model* due to the distribution shift
  - mislabeling rate can be out of control and **unbounded**

TWO **PROBLEMS** for *correctly* applying **LLN approaches** to **SFDA**:

1. **Can general noise-robust LLN methods, based on the Bounded Noise, be effective for SFDA problems where the label noise has different properties?**
2. If NOT, what kinds of LLN methods can be helpful?

Current Limitations of LLN methods in SFDA

Different Label Noises in LLN and in SFDA Settings

- Label Noise in **LLN** (Xiao et al., 2015):
  - generated by *human annotators* or *image search engines*
  - mislabeling rate for a sample is **bounded**
  - general LLN methods: **Noise-Robust** Losses
- Label Noise in **SFDA**:
  - generated by the *source model* due to the distribution shift
  - mislabeling rate can be out of control and **unbounded**

TWO **PROBLEMS** for *correctly* applying **LLN approaches** to **SFDA**:

1. Can general noise-robust LLN methods, based on the *Bounded Noise*, be effective for SFDA problems where the label noise has different properties?
2. **If NOT, what kinds of LLN methods can be helpful?**

Unbounded Label Noise in SFDA

Theoretical Analysis 1 - Unbounded Label Noise in Source Free Domain Adaptation

Definition: Bounded and Unbounded Label Noises

With X as the input feature, Y as the ground-truth label, and $\tilde{Y}$ as the noisy label, we define the **Bounded Label Noise** scenario as:

$$\Pr[\tilde{Y} = i | Y = i, X = \boldsymbol{x}] > \Pr[\tilde{Y} = j | Y = i, X = \boldsymbol{x}], \ \forall \boldsymbol{x} \in \mathcal{X}, i \neq j$$

, and the **Unbounded Label Noise** scenario as:

$$\Pr[\tilde{Y} = j | Y = i, X = \boldsymbol{x}] \to 1, \ \exists \mathcal{S} \subset \mathcal{X}, \ \forall \boldsymbol{x} \in \mathcal{S}, i \neq j$$

- **Bounded**: A sample $\boldsymbol{x}$ has the highest probability of being in the correct class (i)
- **Unbounded**: Mislabeling rate of a sample x can be very high.

Existence of Unbounded Label Noise In SFDA (Th 3.1)

Under some mild assumptions, there exists a non-empty region $\mathbf{R} \subset \mathcal{X}$, for $(\boldsymbol{x}, y) \sim \mathcal{D}_T$, if $\boldsymbol{x} \in \mathbf{R}$, then

$$\Pr[f_S(\boldsymbol{x}) \neq y] \geq 1 - \delta,$$

where $\delta \in (0, 1)$ (i.e., $\delta = 0.01$), f_S is the optimal source classifier.

Theoretical Analysis 1 - Unbounded Label Noise in Source Free Domain Adaptation

Definition: Bounded and Unbounded Label Noises

With X as the input feature, Y as the ground-truth label, and $\tilde{Y}$ as the noisy label, we define the **Bounded Label Noise** scenario as:

$$\Pr\left[\tilde{Y} = i | Y = i, X = \boldsymbol{x}\right] > \Pr\left[\tilde{Y} = j | Y = i, X = \boldsymbol{x}\right], \ \forall \boldsymbol{x} \in \mathcal{X}, i \neq j$$

, and the **Unbounded Label Noise** scenario as:

$$\Pr\left[\tilde{Y} = j | Y = i, X = \boldsymbol{x}\right] \to 1, \ \exists \mathcal{S} \subset \mathcal{X}, \ \forall \boldsymbol{x} \in \mathcal{S}, i \neq j$$

Existence of Unbounded Label Noise In SFDA (Th 3.1)

Under some mild assumptions, there exists a non-empty region $\mathbf{R} \subset \mathcal{X}$, for $(\boldsymbol{x}, y) \sim \mathcal{D}_T$, if $\boldsymbol{x} \in \mathbf{R}$, then

$$\Pr[f_S(\boldsymbol{x}) \neq y] \geq 1 - \delta,$$

where $\delta \in (0, 1)$ (i.e., $\delta = 0.01$), f_S is the optimal source classifier.

- **Theorem 3.1**: Due to the Domain Shift, **Unbounded Label Noise** exists in SFDA.

Theoretical Analysis 1 - Unbounded Label Noise in Source Free Domain Adaptation

Unsuitable LLN Losses for Unbounded Label Noise (Lemma 3.2)

Given a **bounded noise-robust loss ℓ_{LLN}** and an input sample x, we have:

$$\Pr\left[f_T^\star(\boldsymbol{x}) \neq \tilde{f}_T^\star(\boldsymbol{x})\right] \geq 1-\delta, \forall \boldsymbol{x} \in \mathbf{R}$$

where $f_T^\star$ and $\tilde{f}_T^\star$ are the global minimizers of $R(f_T)$ and $\widetilde{R}(f_T)$, the risks of the function f_T under **clean data** and **unbounded noisy data**, respectively.

- **Lemma 3.2**: many existing **Noise-Robust Loss based LLN methods**, which rely on the Bounded Label Noise assumption, are **NOT** the most suitable solutions for **SFDA**.

TWO **PROBLEMS** for *correctly* applying **LLN approaches** to **SFDA**:

1. Can general noise-robust LLN methods, based on the *Bounded Noise*, be effective for SFDA problems where the label noise has different properties? $\Rightarrow$ NO
2. If NOT, what kinds of LLN methods can be helpful?

Theoretical Analysis 1 - Unbounded Label Noise in Source Free Domain Adaptation

Unsuitable LLN Losses for Unbounded Label Noise (Lemma 3.2)

Given a **bounded noise-robust loss $\ell_{\mathbf{LLN}}$** and an input sample x, we have:

$$\Pr\left[f_T^\star(x) \neq \tilde{f}_T^\star(x)\right] \geq 1 - \delta, \forall x \in \mathbf{R}$$

where $f_T^\star$ and $\tilde{f}_T^\star$ are the global minimizers of $R(f_T)$ and $\widetilde{R}(f_T)$, the risks of the function f_T under **clean data** and **unbounded noisy data**, respectively.

- **Lemma 3.2**: many existing **Noise-Robust Loss based LLN methods**, which rely on the Bounded Label Noise assumption, are **NOT** the most suitable solutions for **SFDA**.

TWO **PROBLEMS** for *correctly* applying **LLN approaches** to **SFDA**:

1. **Can general noise-robust LLN methods, based on the Bounded Noise, be effective for SFDA problems where the label noise has different properties? $\Rightarrow$ NO**
2. If NOT, what kinds of LLN methods can be helpful?

Theoretical Analysis 1 - Unbounded Label Noise in Source Free Domain Adaptation

Unsuitable LLN Losses for Unbounded Label Noise (Lemma 3.2)

Given a **bounded noise-robust loss** $\ell_{\mathbf{LLN}}$ and an input sample x, we have:

$$\Pr\left[f_T^{\star}(x) \neq \tilde{f}_T^{\star}(x)\right] \geq 1 - \delta, \forall x \in \mathbf{R}$$

where $f_T^{\star}$ and $\tilde{f}_T^{\star}$ are the global minimizers of $R(f_T)$ and $\widetilde{R}(f_T)$, the risks of the function f_T under **clean data** and **unbounded noisy data**, respectively.

- **Lemma 3.2**: many existing **Noise-Robust Loss based LLN methods**, which rely on the Bounded Label Noise assumption, are **NOT** the most suitable solutions for **SFDA**.

TWO **PROBLEMS** for *correctly* applying **LLN approaches** to **SFDA**:

1. Can general noise-robust LLN methods, based on the *Bounded Noise*, be effective for SFDA problems where the label noise has different properties? $\Longrightarrow$ NO
2. **If NOT, what kinds of LLN methods can be helpful?**

Early-Time Training Phenomenon in SFDA

Theoretical Analysis 2 - Early-time Training Phenomenon (ETP) exists in Unbounded Label Noise Scenario

ETP - Early-time Training Phenomenon

The **Early-time Training Phenomenon** describes the training dynamics of the classifier that preferentially **fits the clean samples** and therefore has **higher prediction accuracy for mislabeled samples** *during the early-training stage.* (Liu et al., 2020)

Existence of ETP in SFDA (Th 4.1)

In the **Unbounded Label Noise** scenario, given a set of mislabeled samples, $B = \{(\boldsymbol{x}, \tilde{y})\}$, and a classifier θ, there exists a proper time T, and a constant c_0 such that for any $0 < \sigma < c_0$, the **prediction accuracy** $\kappa(B; \theta_T)$ can satisfy the following inequality with probability $1 - o_p(1)$:

$$\kappa(B; \theta_T) \geq 1 - \exp\{-\frac{1}{200} g(\sigma)^2\},$$

where $g(\sigma)$ is a monotone decreasing function with $g(\sigma) \rightarrow \infty$ $(\sigma \rightarrow 0)$, and σ is the cluster variance.

$\Longrightarrow$ In **SFDA**,

- the **Early Adaptation Phase** is critical;
- the **Early-Time Predictions** for some easily mislabeled data could be **more promising**.

Method

Early Learning Regularization (ELR) Term (Liu et al., 2020)

$$\mathcal{L}_{\text{ELR}}(\theta_t) = \log\left(1 - \bar{y}_t^\top f(\boldsymbol{x}; \theta_t)\right)$$

where $f(\boldsymbol{x}; \theta_t)$ is the probabilistic output for the sample $\boldsymbol{x}$, and $\bar{y}_t = \beta \bar{y}_{t-1} + (1-\beta) f(\boldsymbol{x}; \theta_t)$ is the moving average prediction for $\boldsymbol{x}$.

Final Method Proposed in SFDA

Given *any* SFDA objective function $\mathcal{L}_{\text{SFDA}}$, the overall objective function is given by:

$$\mathcal{L} = \mathcal{L}_{\text{SFDA}} + \lambda \mathcal{L}_{\text{ELR}},$$

Gradient Analysis in SFDA

$$\frac{\mathrm{d}\mathcal{L}_{\text{ELR}}(\theta_t)}{\mathrm{d}f(\mathrm{x};\theta_t)} = -\frac{\bar{y}_t}{1-\bar{y}_t^\top f(\boldsymbol{x};\theta_t)}$$

- $\mathcal{L}_{\text{ELR}} \downarrow \Rightarrow \left|\frac{\mathrm{d}\mathcal{L}_{\text{ELR}}(\theta_t)}{\mathrm{d}f(\mathrm{x};\theta_t)}\right| \uparrow \Rightarrow \mathcal{L}_{\text{ELR}}$ dominates param updating
  $\Rightarrow$ Enforce the alignment of $f(\mathrm{x};\theta_t)$ with $\bar{y}_t$ rather than **noisy labels**

Experimental Results

Observation of Performance Drop

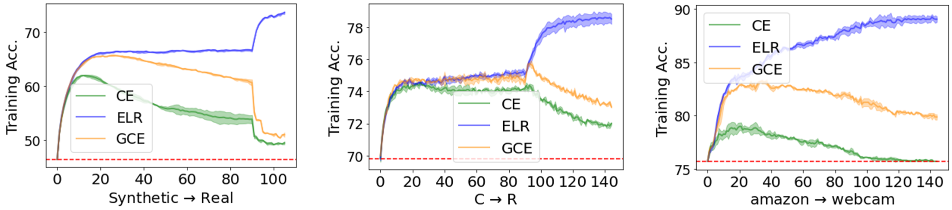


Figure 4: Performance Drop of LLN methods in Adaptation process (VisDA, DomainNet, Office-31)

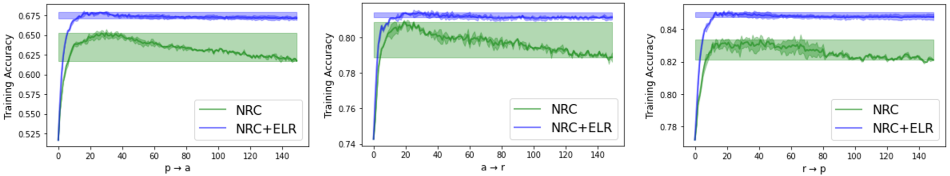


Figure 5: Performance Drop of SFDA methods in Adaptation process (Office-Home)

Main Experimental Results

- Office-Home
- VisDA-2017
- DomainNet

Method	SF	Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Cl→Pr	Cl→Rw	Pr→Ar	Pr→Cl	Pr→Rw	Rw→Ar	Rw→Cl	Rw→Pr	**Avg**
MCD (Saito et al., 2018b)	✗	48.9	68.3	74.6	61.3	67.6	68.8	57.0	47.1	75.1	69.1	52.2	79.6	64.1
CDAN (Long et al., 2018)	✗	50.7	70.6	76.0	57.6	70.0	70.0	57.4	50.9	77.3	70.9	56.7	81.6	65.8
SAFN (Xu et al., 2019a)	✗	52.0	71.7	76.3	64.2	69.9	71.9	63.7	51.4	77.1	70.9	57.1	81.5	67.3
Symnets (Zhang et al., 2019a)	✗	47.7	72.9	78.5	64.2	71.3	74.2	64.2	48.8	79.5	74.5	52.6	82.7	67.6
MDD (Zhang et al., 2019b)	✗	54.9	73.7	77.8	60.0	71.4	71.8	61.2	53.6	78.1	72.5	**60.2**	82.3	68.1
TADA (Wang et al., 2019a)	✗	53.1	72.3	77.2	59.1	71.2	72.1	59.7	53.1	78.4	72.4	60.0	82.9	67.6
BNM (Cui et al., 2020)	✗	52.3	73.9	80.0	63.3	72.9	74.9	61.7	49.5	79.7	70.5	53.6	82.2	67.9
BDG (Yang et al., 2020)	✗	51.5	73.4	78.7	65.3	71.5	73.7	65.1	49.7	81.1	74.6	55.1	84.8	68.7
SRDC (Tang et al., 2020)	✗	52.3	76.3	81.0	**69.5**	76.2	78.0	**68.7**	53.8	81.7	**76.3**	57.1	85.0	71.3
RSDA-MSTN (Gu et al., 2020)	✗	53.2	77.7	81.3	66.4	74.0	76.5	67.9	53.0	82.0	75.8	57.8	**85.4**	70.9
Source Only	✓	44.6	67.3	74.8	52.7	62.7	64.8	53.0	40.6	73.2	65.3	45.4	78.0	60.2
+ELR	✓	52.4	73.5	77.3	62.5	70.6	71.0	61.1	50.8	78.9	71.7	56.7	81.6	67.3
SHOT (Liang et al., 2020)	✓	57.1	78.1	81.5	68.0	78.2	78.1	67.4	54.9	82.2	73.3	58.8	84.3	71.8
+ELR	✓	**58.7**	**78.9**	**82.1**	68.5	79.0	77.5	68.2	57.1	81.9	74.2	59.5	84.9	**72.6**
G-SFDA (Yang et al., 2021b)	✓	55.8	77.1	80.5	66.4	74.9	77.3	66.5	53.9	80.8	72.4	59.7	83.2	70.7
+ELR	✓	56.4	77.6	81.1	67.1	75.2	77.9	65.9	55.0	81.2	72.1	60.0	83.6	71.1
NRC (Yang et al., 2021a)	✓	56.3	77.6	81.0	65.3	78.3	77.5	64.5	56.0	82.4	70.0	57.1	82.9	70.8
+ELR	✓	58.4	78.7	81.5	69.2	**79.5**	**79.3**	66.3	**58.0**	**82.6**	73.4	59.8	85.1	**72.6**

Figure 6: Accuracies (%) on Office-Home for ResNet50-based methods

Main Experimental Results

- Office-Home
- **VisDA-2017**
- DomainNet

Method	SF	plane	bcycl	bus	car	horse	knife	mcycl	person	plant	sktbrd	train	truck	**Per-class**
DANN (Ganin et al., 2016)	✗	81.9	77.7	82.8	44.3	81.2	29.5	65.1	28.6	51.9	54.6	82.8	7.8	57.4
DAN (Long et al., 2015)	✗	87.1	63.0	76.5	42.0	90.3	42.9	85.9	53.1	49.7	36.3	85.8	20.7	61.1
ADR (Saito et al., 2018a)	✗	94.2	48.5	84.0	72.9	90.1	74.2	92.6	72.5	80.8	61.8	82.2	28.8	73.5
CDAN (Long et al., 2018)	✗	85.2	66.9	83.0	50.8	84.2	74.9	88.1	74.5	83.4	76.0	81.9	38.0	73.9
SAFN (Xu et al., 2019a)	✗	93.6	61.3	84.1	70.6	94.1	79.0	91.8	79.6	89.9	55.6	89.0	24.4	76.1
SWD (Lee et al., 2019)	✗	90.8	82.5	81.7	70.5	91.7	69.5	86.3	77.5	87.4	63.6	85.6	29.2	76.4
MDD (Zhang et al., 2019b)	✗	-	-	-	-	-	-	-	-	-	-	-	-	74.6
MCC (Jin et al., 2020)	✗	88.7	80.3	80.5	71.5	90.1	93.2	85.0	71.6	89.4	73.8	85.0	36.9	78.8
STAR (Lu et al., 2020)	✗	95.0	84.0	84.6	73.0	91.6	91.8	85.9	78.4	94.4	84.7	87.0	42.2	82.7
RWOT (Xu et al., 2020)	✗	95.1	80.3	83.7	**90.0**	92.4	68.0	92.5	82.2	87.9	78.4	**90.4**	**68.2**	84.0
Source Only	✓	60.9	21.6	50.9	67.6	65.8	6.3	82.2	23.2	57.3	30.6	84.6	8.0	46.6
+ELR	✓	95.4	45.7	89.7	69.8	94.1	97.1	**92.9**	80.1	89.7	52.8	83.3	4.3	74.6
SHOT (Liang et al., 2020)	✓	94.3	88.5	80.1	57.3	93.1	94.9	80.7	80.3	91.5	89.1	86.3	58.2	82.9
+ELR	✓	95.8	84.1	83.3	67.9	93.9	**97.6**	89.2	80.1	90.6	90.4	87.2	48.2	84.1
G-SFDA (Yang et al., 2021b)	✓	96.0	87.6	85.3	72.8	95.9	94.7	88.4	79.0	92.7	93.9	87.2	43.7	84.8
+ELR	✓	**97.3**	89.1	**89.8**	79.2	**96.9**	97.5	92.2	**82.5**	**95.8**	**94.5**	87.3	34.5	**86.4**
NRC (Yang et al., 2021a)	✓	96.9	89.7	84.0	59.8	95.9	96.6	86.5	80.9	92.8	92.6	90.2	60.2	85.4
+ELR	✓	97.1	**89.7**	82.7	62.0	96.2	97.0	87.6	81.2	93.7	94.1	90.2	58.6	85.8

Figure 6: Accuracies (%) on VisDA-C (Synthesis → Real) for ResNet101-based methods

Main Experimental Results

- Office-Home
- VisDA-2017
- **DomainNet**

Method	SF	R→C	R→P	R→S	C→R	C→P	C→S	P→R	P→C	P→S	S→R	S→C	S→P	**Avg**
MCD (Saito et al., 2018b)	✗	61.9	69.3	56.2	79.7	56.6	53.6	83.3	58.3	60.9	81.7	56.2	66.7	65.4
DANN (Ganin et al., 2016)	✗	63.4	73.6	72.6	86.5	65.7	70.6	86.9	73.2	70.2	85.7	75.2	70.0	74.5
DAN (Long et al., 2015)	✗	64.3	70.6	58.4	79.4	56.7	60.0	84.5	61.6	62.2	79.7	65.0	62.0	67.0
COAL (Tan et al., 2020)	✗	73.9	75.4	70.5	89.6	70.0	71.3	89.8	68.0	70.5	88.0	73.2	70.5	75.9
MDD (Zhang et al., 2019b)	✗	77.6	75.7	**74.2**	89.5	74.2	**75.6**	90.2	76.0	**74.6**	86.7	72.9	73.2	78.4
Source Only	✓	53.7	71.6	52.9	70.8	49.5	58.3	85.2	59.6	59.1	30.6	74.8	65.7	61.0
+ELR	✓	70.2	81.7	61.7	79.9	63.8	67.0	90.0	72.1	66.8	85.1	78.5	68.8	73.8
SHOT (Liang et al., 2020)	✓	73.3	80.1	65.8	**91.4**	74.3	69.2	91.9	77.0	66.2	87.4	81.3	75.0	77.7
+ELR	✓	**78.0**	81.9	67.4	91.1	**75.9**	71.0	92.6	79.3	68.0	88.7	84.8	77.0	**79.7**
G-SFDA (Yang et al., 2021b)	✓	65.8	78.9	60.2	80.5	64.7	64.6	89.3	69.9	63.6	86.4	78.8	71.1	72.8
+ELR	✓	69.4	80.9	60.6	81.3	67.2	66.4	90.2	73.2	64.9	87.6	82.1	71.0	74.6
NRC (Yang et al., 2021a)	✓	69.8	81.1	62.9	83.4	74.4	66.3	90.3	73.4	65.2	88.2	82.2	75.8	76.4
+ELR	✓	75.6	**82.2**	65.7	91.2	77.2	68.5	**92.7**	**79.8**	67.5	**89.3**	**85.1**	**77.6**	79.4

Figure 6: Accuracies (%) on DomainNet for ResNet50-based methods

Summary

In this work, we

1. *Distinguish* Label Noises in **SFDA** from **Traditional LLN Settings**;
2. *Justify* the existence of **ETP** in **Unbound Label Noise**;
3. *Identify* **effective** LLN methods for **SFDA**;
4. *Introduce* the **ELR** term to enhance SFDA performance.

We hope this work can **INSPIRE** more research on

- Exploring the Training Dynamic of Early-Time Adaptation
- and Utilizing the Early-Time Training Phenomenon in Unbounded Label Noise.

Thanks!

Poster Session 4: May 2, 16:30 - 18:30 #144

Project Code