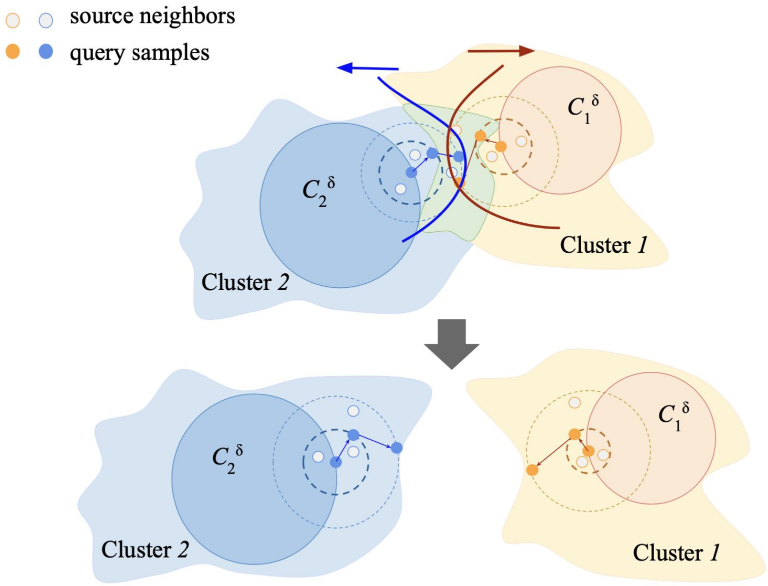
What Has Been Overlooked in Contrastive Source-Free Domain Adaptation: Leveraging Source-Informed Latent Augmentation within Neighborhood Context



Jing Wang, Wonho Bae, Jiahong Chen, Kuangen Zhang, Leonid Sigal, and Clarence W. de Silva

The University of British Columbia (UBC)

Outline

- **Motivation**
- **Contrastive Source-Free Domain Adaptation**
- **Literature Review**
- **Theoretical Analysis**
- **The Proposed SiLAN**
- **Results**
- **Conclusions**

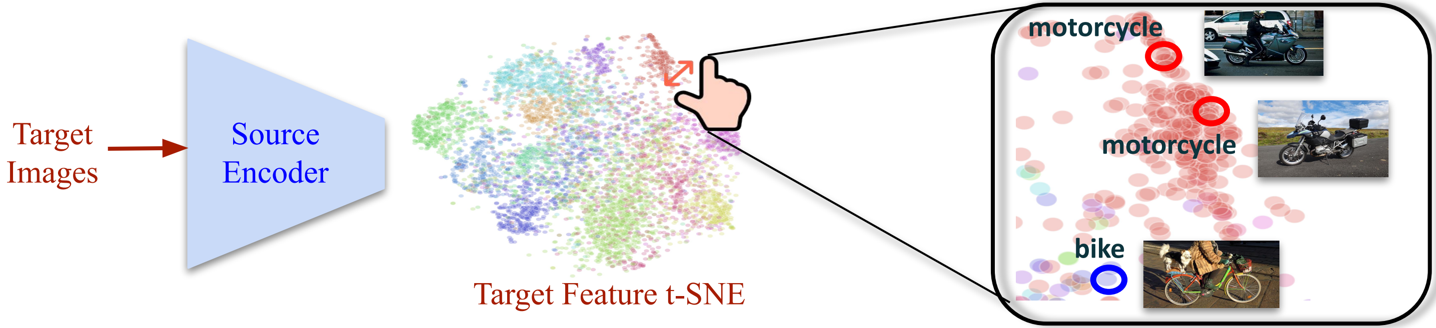
Motivation

Objectives

- We aim to investigate how variations in the design of positive keys in contrastive clustering affect the source-free domain adaptation performance.
- We aim to explore how the insights gained from the analysis can be used to improve contrastive SFDA frameworks.

Observations

- A domain shift induces a significant dispersion of target features.
- However, neighboring target features still tend to belong to the same class.



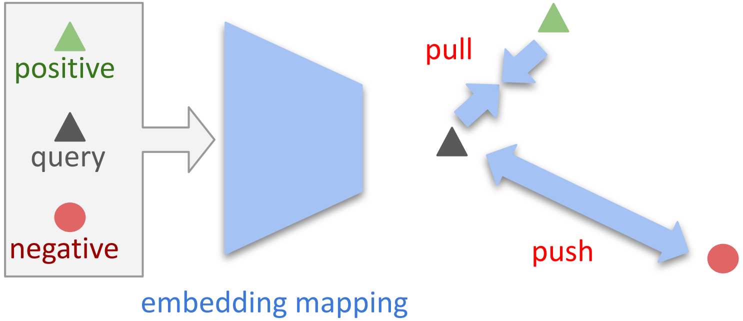
Contrastive Source-Free Domain Adaptation

Source-Free Domain Adaptation (SFDA)

- Adapting a model initially trained on a source domain to perform effectively on a target domain, without access to any source domain data during the adaptation process.
- In this scenario, the target adaptation becomes an unsupervised learning challenge, involving:
    - a pre-trained model
    - completely unlabeled data

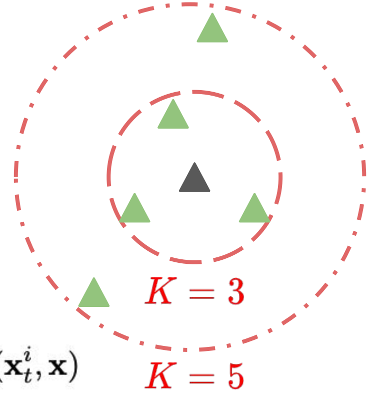
Contrastive SFDA

- Resolving SFDA problems by contrastive clustering akin to unsupervised learning.
    - The effectiveness of contrastive clustering heavily depends on how positive keys are defined.
        - augmentation-based
        - neighborhood-searching-based



$$d(\mathbf{x}_t^i, \mathbf{x}_t^j) := \frac{G(\mathbf{x}_t^i) \cdot \mathbb{B}_{T,j}}{||G(\mathbf{x}_t^i)|| \cdot ||\mathbb{B}_{T,j}||}$$

$$\mathbb{N}_K(\mathbf{x}_t^i) := \arg max_{\mathcal{S} \subset \mathbb{B}_T, |\mathcal{S}|=K} \sum_{\mathbf{x} \in \mathcal{S}} d(\mathbf{x}_t^i, \mathbf{x})$$



Theoretical Analysis

Target Adaptation with Contrastive Clustering

- **Given an InfoNCE [1]**

$$\mathcal{L}_{cont} = -\sum_{i=1}^{m} log \frac{e^{f_t^\top(\mathbf{x}_i) f_t(\mathbf{x}_i^+)/\tau}}{\sum_{j \neq i} e^{f_t^\top(\mathbf{x}_i) f_t(\mathbf{x}_j^+)/\tau}}.$$

Proposition 1. *When formulated in the output logit space, the InfoNCE-based contrastive loss, denoted as $\mathcal{L}_{cont}$, serves as an upper bound for the misalignment associated with two distinct alignment errors in predictions.*

$$\sum_{i=1}^{m}\left(\log(m-1) + \frac{\|f_t(\mathbf{x}_i) - f_t(\mathbf{x}_i^+)\|_2^2}{2\tau} - \frac{1}{m-1}\sum_{j \neq i}\frac{\|f_t(\mathbf{x}_i) - f_t(\mathbf{x}_j^+)\|_2^2}{2\tau}\right) \leq \mathcal{L}_{cont}.$$

[1] Oord, Aaron van den, Yazhe Li, and Oriol Vinyals. "Representation learning with contrastive predictive coding." arXiv preprint arXiv:1807.03748 (2018).

Theoretical Analysis

Unveiling the Uniqueness: Exploring Contrastive Clustering in SFDA

- **Preliminary**
    - The samples predicted to be class ***z*** can be enclosed within a ball of $\mathcal{C}_z^\delta$ diameter δ
        - a subset of positive keys are defined within the ball $\mathcal{S}_z^+ = \{\mathbf{x}^+ \in \mathcal{C}_z : \forall \mathbf{x} \in \mathcal{C}_z^\delta, ||f_t(\mathbf{x}^+) - f_t(\mathbf{x})||_2^2 \leq \delta\}$,
        - a subset of negative keys are defined outside the ball $\mathcal{S}_z^- = \{\mathbf{x}^- \in \mathcal{C}_z : \forall \mathbf{x} \in \mathcal{C}_z^\delta, ||f_t(\mathbf{x}^-) - f_t(\mathbf{x})||_2^2 > \delta\}$.
        - reformulate the target classification error *w.r.t.* alignment errors

$$\epsilon_{\mathcal{D}_T} = \sum_{z=1}^{Z} (\mathbb{P}[f_t(\mathbf{x}_t) \neq z, \forall \mathbf{x}_t \in \mathcal{C}_z] + \mathbb{P}[z \neq \mathbf{y}_t, \forall \mathbf{x}_t \in \mathcal{C}_z]),$$

- **A New Upper Bound for Target Classification Error**

Lemma 2. *If $\mathcal{C}_z^\delta \cap \mathcal{C}_l^\delta = \emptyset$ holds for any $l \neq z$, then the error $\epsilon_{\mathcal{D}_T}$ defined on the groups of logits is upper bounded by:*

$$\epsilon_{\mathcal{D}_T} \leq R_\delta + \sum_{z=1}^{Z} \Big(\mathbb{P}[f_t(\mathbf{x}^+) \neq \mathbf{y}_t, \forall \mathbf{x}^+ \in \mathcal{S}_z^+] + \mathbb{P}[f_t(\mathbf{x}^-) = \mathbf{y}_t, \forall \mathbf{x}^- \in \mathcal{S}_z^-] \Big),$$

where $R_\delta = \frac{\cup_{z=1}^{Z}(\mathcal{C}_z - \mathcal{C}_z^\delta)}{\cup_{z=1}^{Z}\mathcal{C}_z}$ and $(\mathcal{C}_z - \mathcal{C}_z^\delta)$ may overlap with $(\mathcal{C}_l - \mathcal{C}_l^\delta)$ for any $l \neq z$.

Theoretical Analysis (continued)

Unveiling the Uniqueness: Exploring Contrastive Clustering in SFDA

- **A New Upper Bound for Target Classification Error**

Lemma 2. *If $\mathcal{C}_z^\delta \cap \mathcal{C}_l^\delta = \emptyset$ holds for any $l \neq z$, then the error $\epsilon_{\mathcal{D}_T}$ defined on the groups of logits is upper bounded by:*

$$\epsilon_{\mathcal{D}_T} \leq R_\delta + \sum_{z=1}^{Z} \Big(\mathbb{P}[f_t(\mathbf{x}^+) \neq \mathbf{y}_t, \forall \mathbf{x}^+ \in \mathcal{S}_z^+] + \mathbb{P}[f_t(\mathbf{x}^-) = \mathbf{y}_t, \forall \mathbf{x}^- \in \mathcal{S}_z^-] \Big),$$

where $R_\delta = \frac{\cup_{z=1}^{Z}(\mathcal{C}_z - \mathcal{C}_z^\delta)}{\cup_{z=1}^{Z}\mathcal{C}_z}$ and $(\mathcal{C}_z - \mathcal{C}_z^\delta)$ may overlap with $(\mathcal{C}_l - \mathcal{C}_l^\delta)$ for any $l \neq z$.

Beyond the Blind Spots: Revisiting Three Oversights in Existing Contrastive SFDA Methods

- Data augmentation lacking domain-specific insight can misguide the model under the domain shift.
- Expanding neighborhood search range leads to smoother predictions but also heightens the risk of overlap.
- Enhancing the source pre-trained model's utility for positive keys can decrease transformation misclassification errors.

The Proposed SiLAN

Addressing Three Oversights Leads to a Simplified SFDA Framework

Source-Informed Latent Augmented Neighborhood (*SiLAN*)

Adding target-centric and source-scattered Gaussian noise to target latent features

$$\xi \sim \mathcal{N}(\mathbf{0}, {\sigma^{s}_{K}}^{2}(\mathbf{x}^{i}_{\mathbf{t}}))$$

$$\hat{\mathbf{h}} := G_t(\mu^{t}_{K}(\mathbf{x}^{i}_{\mathbf{t}})) + \xi$$

$\mathbf{x}^{i}_{t}$	target input data (**query**)
$G_t(\cdot)$	target encoder
$F_t(\cdot)$	target classifier
$\mu^{t}_{K}(\cdot)$	mean of the query's KNNs
$\sigma^{t}_{K}(\cdot)$	Standard deviation of the query's KNNs

Our SFDA Framework

InfoNCE loss with SiLAN augmentation

$$\mathcal{L}_{SiLAN} = -\sum_{i=1}^{m} log \frac{e^{[F_t(G_t(\mathbf{x}^{i}_{\mathbf{t}}))]^{\top} F_t(\hat{\mathbf{h}}_i)/\tau}}{\sum_{j \neq i} e^{[F_t(G_t(\mathbf{x}^{i}_{\mathbf{t}})]^{\top} F_t(\hat{\mathbf{h}}_j)/\tau}}$$

The Proposed SiLAN (continued)

The Reasoning Behind SiLAN

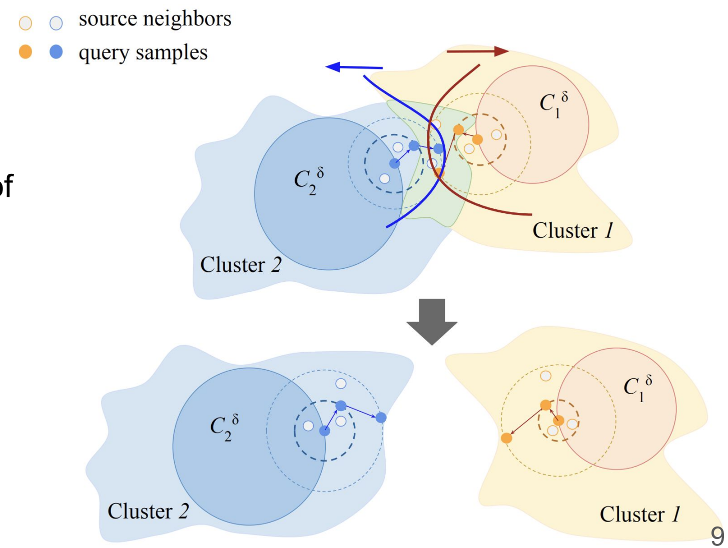
Let us examine a general scenario in which $\xi_0 \sim \mathcal{N}(\mathbf{0}, \sigma^2)$ is introduced to the queries' latent neighborhood centroid for generating positive keys.

$$\hat{\mathbf{h}}_0 := G_t(\mu_K^t(\mathbf{x}_t^i)) + \xi_0$$

When aligning the predictions of queries and their augmented latent representations using contrastive loss, the samples will cluster in their latent space due to the pull (positive) and push (negative) effects.

Observations

- The radius of the spherical Gaussian profiles is determined by σ.
  - Larger radius leads to wider traversal
- Contrastive loss pulls the features of positive keys closer to the features of queries while pushing away those of negative keys.
- The pull and push effects will widen the gap between clusters as their augmentations traverse along the Gaussian profile.



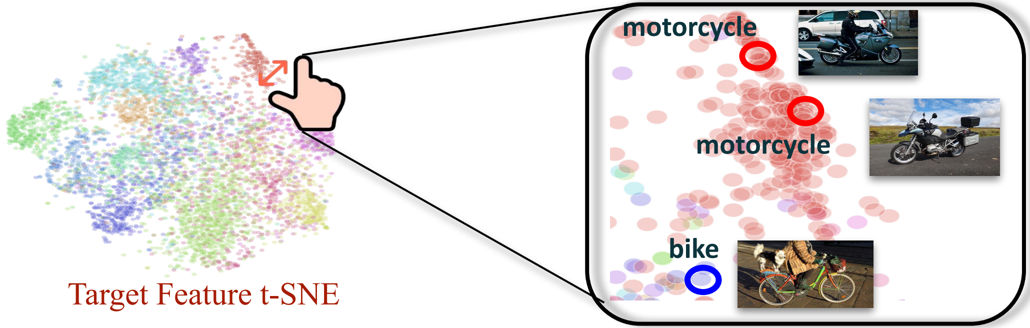
Conclusions

- A larger σ promotes a wider gap between clusters but also increases the risk of incorporating irrelevant features into the clusters

The Proposed SiLAN (continued)

How Should We Determine σ?

Recall the observations from examining the t-SNE visualization of target features extracted by the source pre-trained model:



Target Feature t-SNE

The source-model-informed variance of the latent features in the query's neighborhood guides us in determining the appropriate size of the Gaussian profile for effective augmentation traversal: **simply set the variance of Gaussian noise for latent augmentation to σ_K^s!**

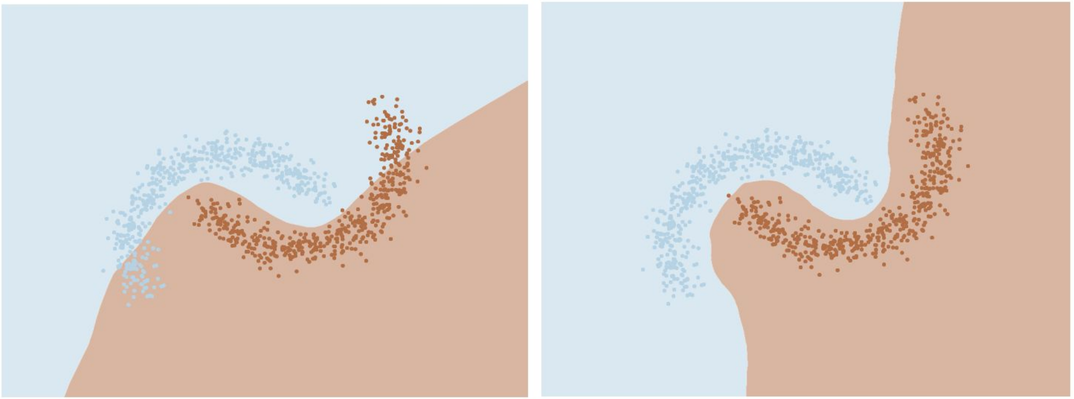
Results

Experiments on Toy Dataset

moon dataset

Simulating domain shift by rotating data sample orientations, each domain consists of an interleaving half circle with 1000 data points. The target domain replicates this structure with a 30-degree rotation around the mean. Both domains include Gaussian noise with a standard deviation of 0.1 and are generated using distinct random seeds.

Backbone: 5-layer MLP without nonlinear activation



(a) Source Only (b) SiLAN

Results

Experiments on Benchmark Datasets

Office-31 dataset

- 4,652 images of 31 object classes captured from three domains: *Amazon* (**A**), *Webcam* (**W**), and *DSLR* (**D**).

Results

Table 1: Comparison of SFDA methods using ResNet-50 on *Office-31*. The best results are highlighted.

Method	A→D	A→W	D→W	D→A	W→D	W→A	Avg.
ResNet-50 He et al. (2016)	68.9	68.4	96.7	62.5	99.3	60.7	76.1
SHOT Liang et al. (2020)	94.0	90.1	98.4	74.7	99.9	74.3	88.6
NRC Yang et al. (2021a)	96.0	90.8	99.0	75.3	100.0	75.0	89.4
3C-GAN Li et al. (2020)	92.7	93.7	98.5	75.3	99.8	**77.8**	89.6
HCL Huang et al. (2021a)	94.7	92.5	98.2	75.9	100.0	77.7	89.8
AaD Yang et al. (2022)	96.4	92.1	**99.1**	75.0	100.0	76.5	89.9
$SF(DA)^2$ Hwang et al. (2024)	95.8	92.1	99.0	75.7	99.8	76.8	89.9
SiLAN (Ours)	**97.1**	**95.8**	98.9	**76.4**	**100.0**	76.9	**90.7**

Results

Experiments on Benchmark Datasets

Office-Home dataset

- 15,500 images of 65 classes from four domains: *Artistic* (**Ar**), *Clipart* (**Cl**), *Product* (**Pr**), and *Real-World* (**Rw**).

Results

Table 2: Comparison of the SFDA methods on *Office-Home* (ResNet-50).

Method	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.
ResNet-50 He et al. (2016)	34.9	50.0	58.0	37.4	41.9	46.2	38.5	31.2	60.4	53.9	41.2	59.9	46.1
G-SFDA Yang et al. (2021b)	57.9	78.6	81.0	66.7	77.2	77.2	65.6	56.0	82.2	72.0	57.8	83.4	71.3
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
NRC Yang et al. (2021a)	57.7	80.3	82.0	68.1	**79.8**	78.6	65.3	56.4	**83.0**	71.0	58.6	85.6	72.2
AaD Yang et al. (2022)	59.3	79.3	82.1	68.9	**79.8**	79.5	67.2	57.4	83.1	72.1	58.5	85.4	72.7
DaC Zhang et al. (2022)	**59.5**	79.5	81.2	69.3	78.9	79.2	67.4	56.4	82.4	74.0	**61.4**	84.4	72.8
SF(DA)^2 Hwang et al. (2024)	57.8	80.2	81.5	69.5	79.2	79.4	66.5	57.2	82.1	73.3	60.2	83.8	72.6
SiLAN (Ours)	58.2	**81.2**	**82.5**	**69.8**	78.6	**80.3**	**68.4**	**58.6**	82.5	**75.6**	60.8	**86.1**	**73.6**

Results

Experiments on Benchmark Datasets

VisDA-2017-C dataset

- A large-scale dataset used for the 2017 *ICCV* visual DA challenge, with 280K images of 12 object categories. The source domain contains synthetic images generated via 3D model rendering, while the target domain consists of real images.

Results

Table 3: Comparison of the SFDA methods on *VisDA2017* (ResNet-101).

Method	plane	bcycl	bus	car	horse	knife	mcycl	person	plant	sktbrd	train	truck	Avg.
ResNet-101 He et al. (2016)	55.1	53.3	61.9	59.1	80.6	17.9	79.7	31.2	81.0	26.5	73.5	8.5	52.4
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
HCL Huang et al. (2021a)	93.3	85.4	80.7	68.5	91.0	88.1	86.0	78.6	86.6	88.8	80.0	74.7	83.5
G-SFDA Yang et al. (2021b)	96.1	88.3	85.5	74.1	97.1	95.4	89.5	79.4	95.4	92.9	89.1	42.6	85.4
NRC Yang et al. (2021a)	96.8	**91.3**	82.4	62.4	96.2	95.9	86.1	80.6	94.8	94.1	90.4	59.7	85.9
AaD Yang et al. (2022)	95.2	90.5	85.5	79.2	96.4	96.2	88.8	80.4	93.9	91.8	91.1	55.9	87.1
DaC Zhang et al. (2022)	96.6	86.8	**86.4**	78.4	96.4	96.2	**93.6**	**83.8**	**96.8**	95.1	89.6	50.0	87.3
SF(DA)^2 Hwang et al. (2024)	96.8	89.3	82.9	81.4	96.8	**95.7**	90.4	81.3	95.5	93.7	88.5	**64.7**	88.1
SiLAN (Ours)	**97.5**	90.1	85.8	**80.4**	**97.6**	95.5	92.0	82.9	96.5	**95.3**	**92.6**	53.4	**88.3**

Conclusions

Our Contributions

- We hypothesize that domain shift causes significant dispersion in target features yet nearby points still tend to share similar labels, explaining the success of contrastive clustering in SFDA.
- Our theoretical analysis of contrastive SFDA reveals that three often-overlooked factors, associated with the aforementioned hypotheses, have significant implications for target classification performance.
- To address these three issues, we introduce *SiLAN*, a simple yet effective latent augmentation technique explicitly designed to improve contrastive SFDA.
- Experimental results support our theoretical findings, demonstrating that *InfoNCE*, when augmented with *SiLAN*, achieves state-of-the-art performance in SFDA.

Thank You!