Unsupervised Domain Adaptation with Self-Training

DMQA Open Seminar (24. 11. 29)

Data Mining & Quality Analytics Lab.

김지현

발표자 소개

❖ **김지현 (Jihyun Kim)**

- 고려대학교 산업경영공학과 대학원 재학
- Data Mining & Quality Analytics Lab. (김성범 교수님)
- Ph.D. Student (2022.03 ~ Present)

❖ **Research Interest**

- Domain Adaptation

❖ **Contact**

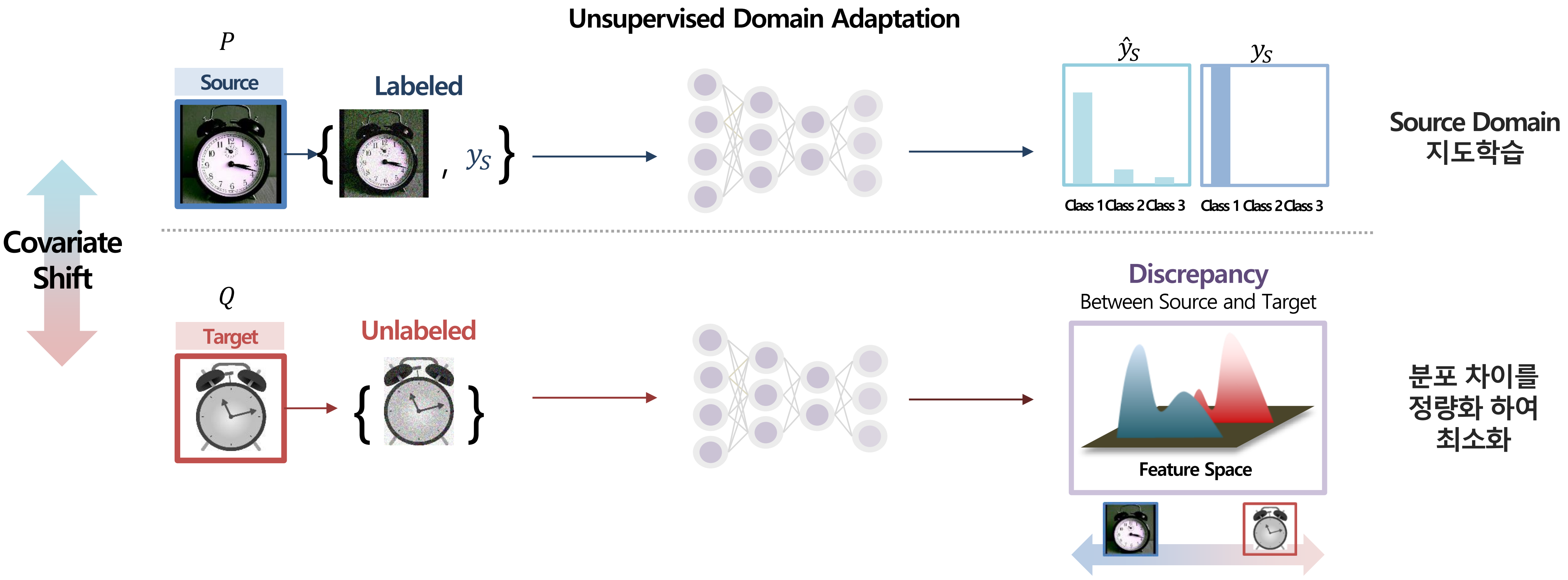
- jihyun_k@korea.ac.kr

Introduction

Background on Unsupervised Domain Adaptation

❖ **Unsupervised Domain Adaptation** with Self-Training

- **Source와 target 간 분포 차이를 줄이는 방법**으로써 self-training 기법을 이용하는 domain adaptation 연구 갈래
- Unlabeled target domain의 pseudo-labels을 기반으로 학습을 진행하여, 모델이 target domain에 점진적으로 적응하도록 함[1]



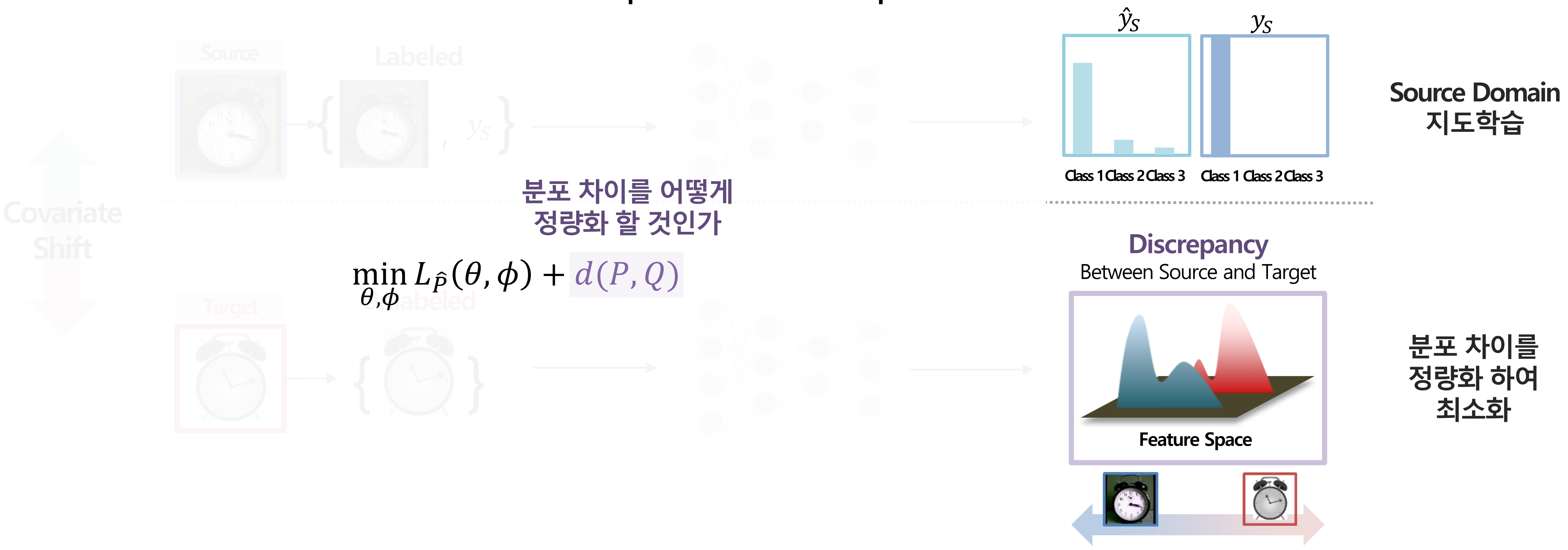
Introduction

Background on Unsupervised Domain Adaptation

❖ **Unsupervised Domain Adaptation** with Self-Training

- **Source와 target 간 분포 차이를 줄이는 방법**으로써 self-training 기법을 이용하는 domain adaptation 연구 갈래
- Unlabeled target domain의 pseudo-labels을 기반으로 학습을 진행하여, 모델이 target domain에 점진적으로 적응하도록 함[1]

Unsupervised Domain Adaptation



$$\min_{\theta,\phi} L_{\hat{P}}(\theta,\phi) + d(P,Q)$$

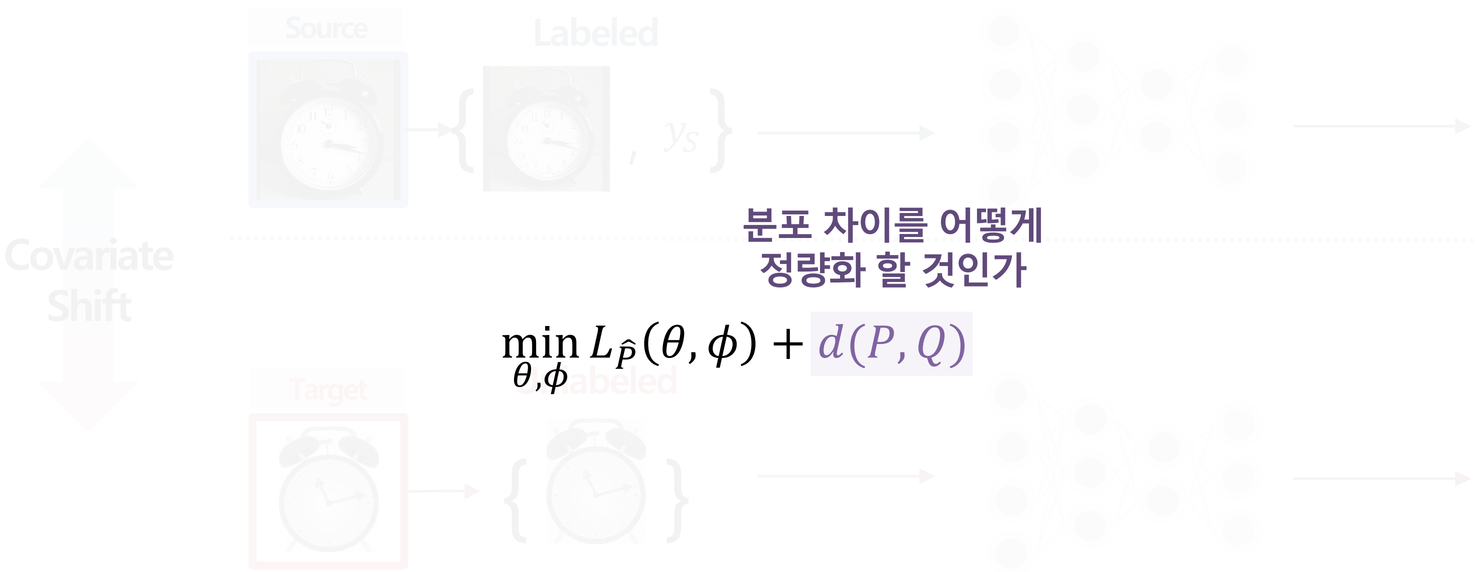
Introduction

Background on Unsupervised Domain Adaptation

❖ **Unsupervised Domain Adaptation** with Self-Training

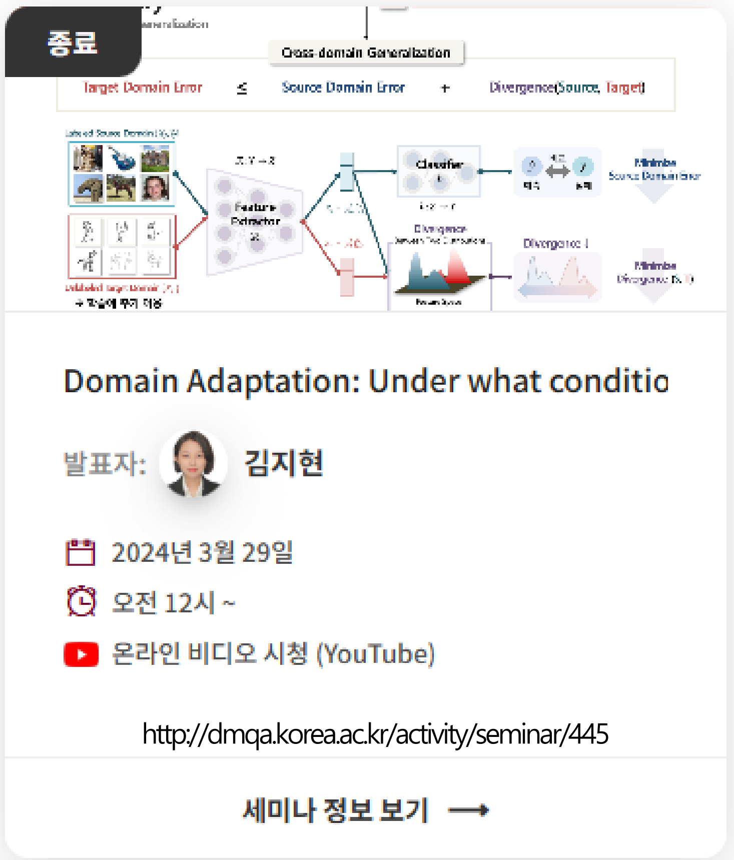
- **Source와 target 간 분포 차이를 줄이는 방법**으로써 self-training 기법을 이용하는 domain adaptation 연구 갈래
- Unlabeled target domain의 pseudo-labels을 기반으로 학습을 진행하여, 모델이 target domain에 점진적으로 적응하도록 함[1]

Unsupervised Domain Adaptation



분포 차이를 어떻게
정량화 할 것인가

$$\min_{\theta,\phi} L_{\hat{P}}(\theta,\phi) + d(P,Q)$$



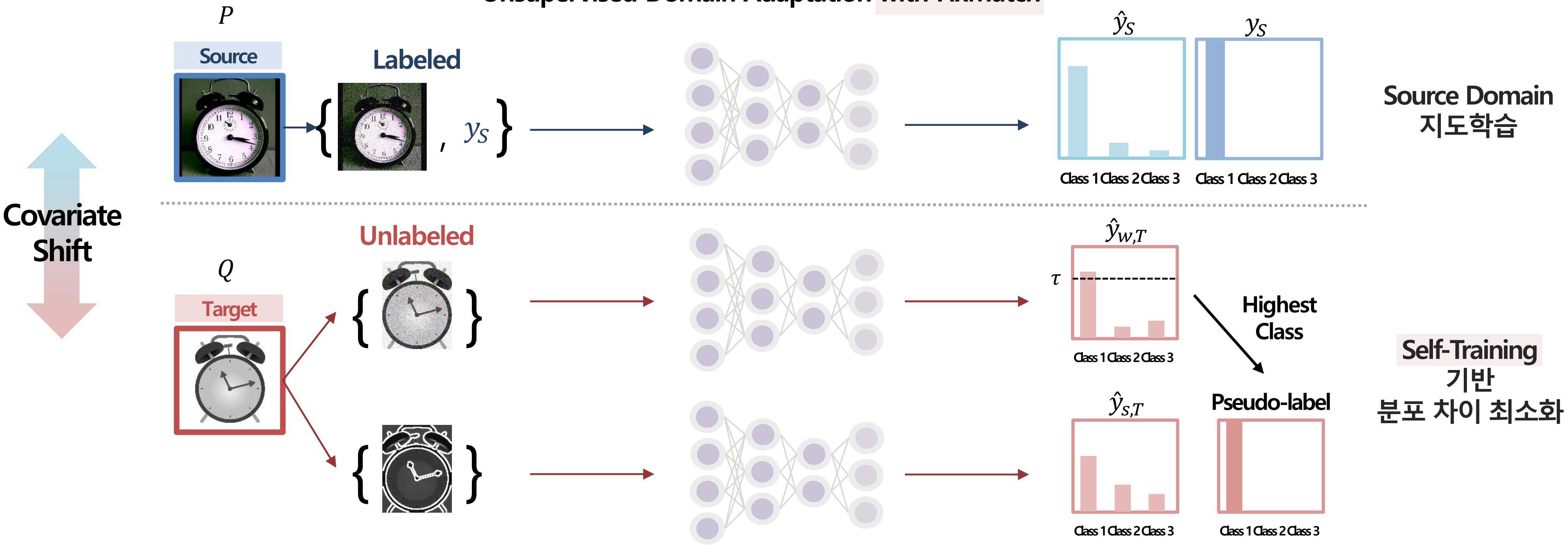
Introduction

Background on Unsupervised Domain Adaptation with Self-Training

❖ **Unsupervised Domain Adaptation with Self-Training**

- **Source와 target 간 분포 차이를 줄이는 방법**으로써 **self-training 기법을 이용**하는 domain adaptation 연구 갈래
- Unlabeled target domain의 **pseudo-labels을 기반으로 학습을 진행**하여, 모델이 target domain에 점진적으로 적응하도록 함[1]

Unsupervised Domain Adaptation with FixMatch



[1] Zou, Y., Yu, Z., Liu, X., Kumar, B. V. K., & Wang, J. (2019). Confidence regularized self-training. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 5982-5991).

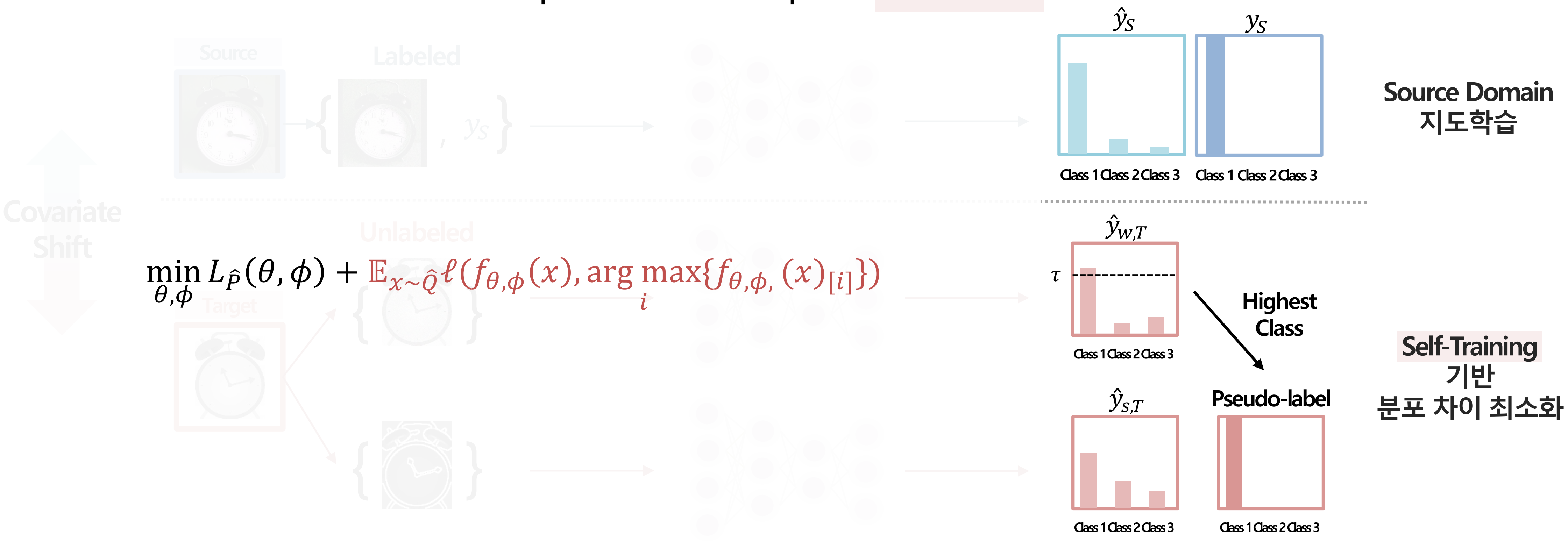
Introduction

Background on Unsupervised Domain Adaptation with Self-Training

❖ **Unsupervised Domain Adaptation with Self-Training**

- **Source와 target 간 분포 차이를 줄이는 방법**으로써 **self-training 기법을 이용**하는 domain adaptation 연구 갈래
- Unlabeled target domain의 **pseudo-labels을 기반으로 학습을 진행**하여, 모델이 target domain에 점진적으로 적응하도록 함[1]

Unsupervised Domain Adaptation with FixMatch



$$\min_{\theta,\phi} L_{\hat{P}}(\theta,\phi) + \mathbb{E}_{x\sim\hat{Q}}\ell(f_{\theta,\phi}(x), \arg\max_i\{f_{\theta,\phi,}(x)_{[i]}\})$$

[1] Zou, Y., Yu, Z., Liu, X., Kumar, B. V. K., & Wang, J. (2019). Confidence regularized self-training. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 5982-5991).

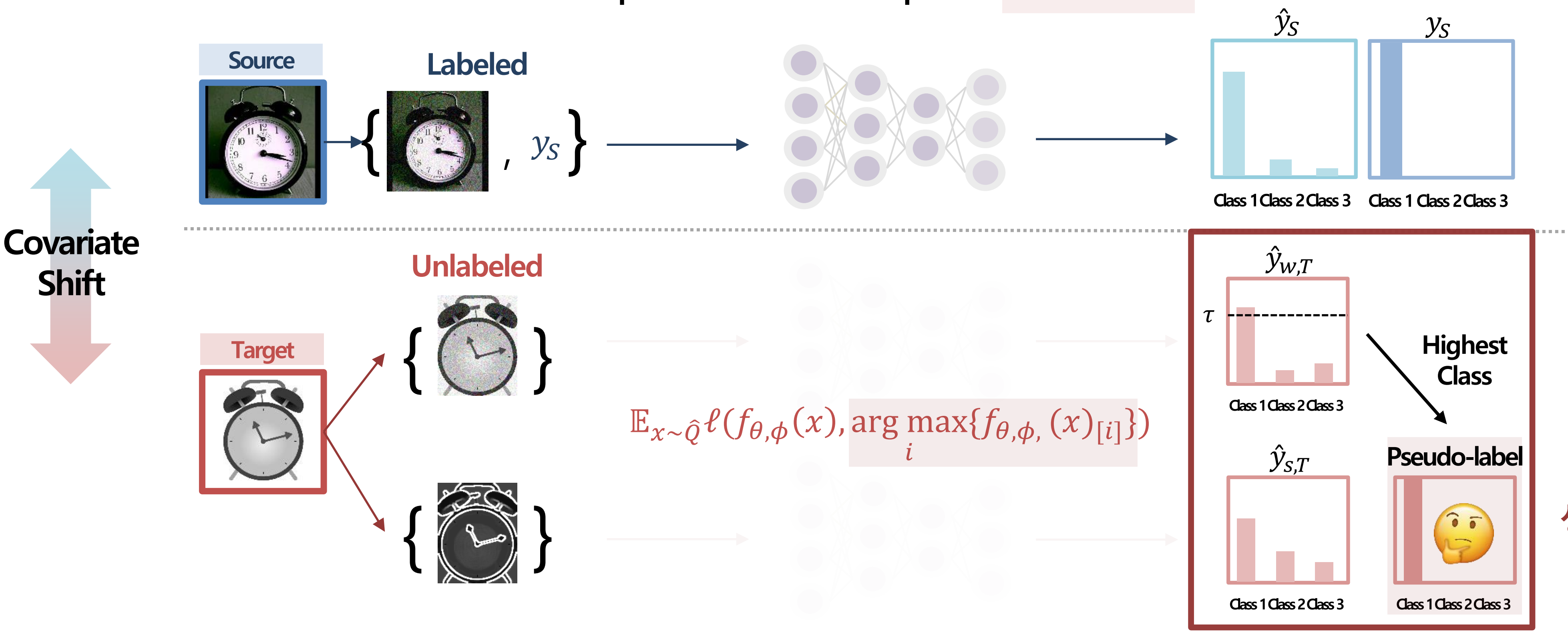
Introduction

Background on Unsupervised Domain Adaptation with Self-Training

❖ Unsupervised Domain Adaptation with Self-Training

- **Source와 target 간 분포 차이를 줄이는 방법**으로써 **self-training 기법을 이용**하는 domain adaptation 연구 갈래
- Unlabeled target domain의 **pseudo-labels을 기반으로 학습을 진행**하여, 모델이 target domain에 점진적으로 적응하도록 함[1]

Unsupervised Domain Adaptation with FixMatch



Limitation

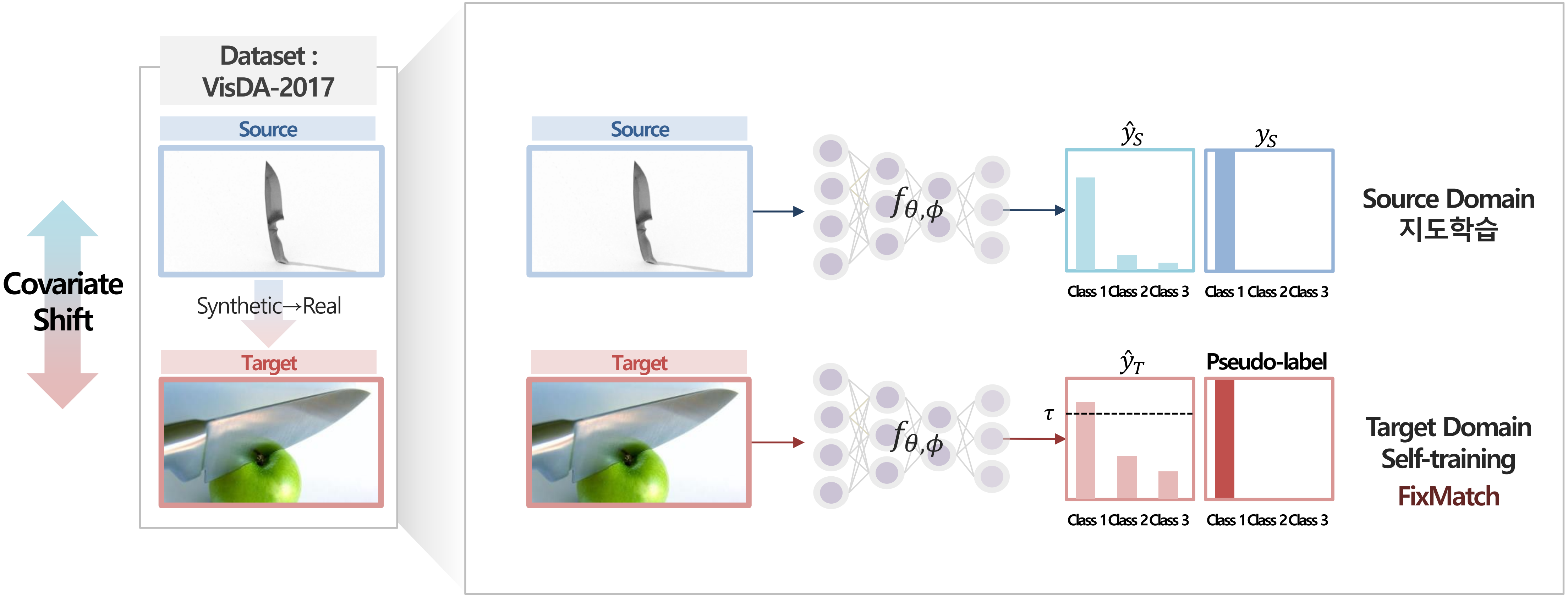
분포 차이로 인해
믿을 수 있는
Pseudo-Labels
생성/선별 어려움[2]

[1] Zou, Y., Yu, Z., Liu, X., Kumar, B. V. K., & Wang, J. (2019). Confidence regularized self-training. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 5982-5991).
[2]] Liu, H., Wang, J., & Long, M. (2021). Cycle self-training for domain adaptation. Advances in Neural Information Processing Systems, 34, 22968-22981.

Introduction

Background on Unsupervised Domain Adaptation with Self-Training

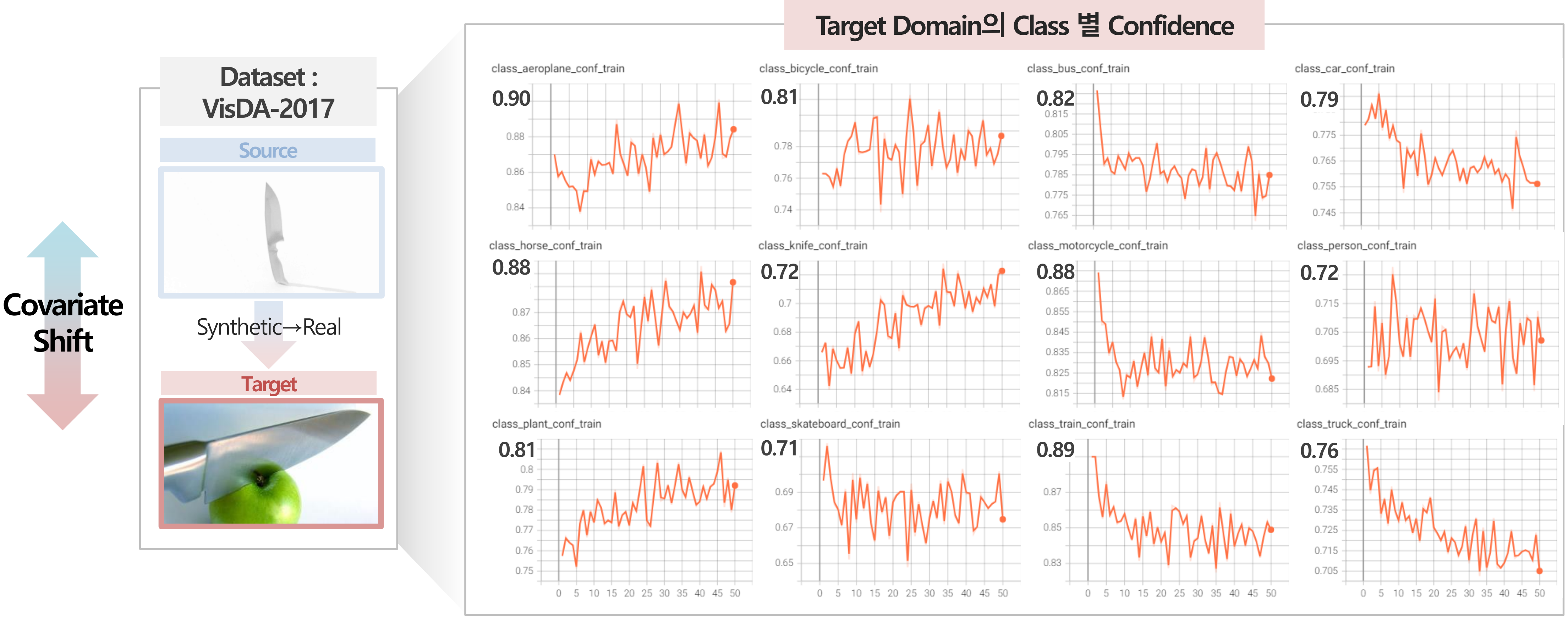
Using top-1 softmax confidence or predictive entropy and self-train on highly confident instances!



Introduction

Background on Unsupervised Domain Adaptation with Self-Training

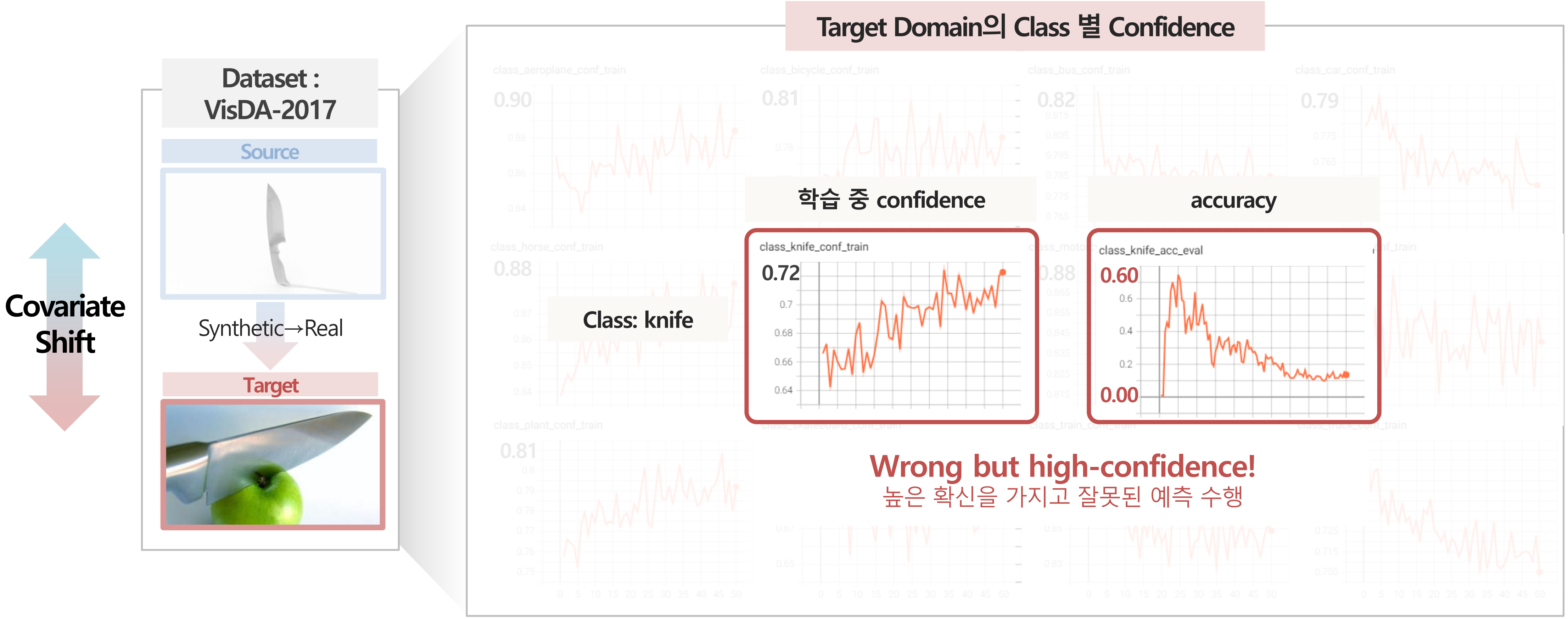
Using top-1 softmax confidence or predictive entropy and self-train on highly confident instances!
→ Such confidence measures tend to be miscalibrated and are often unreliable!



Introduction

Background on Unsupervised Domain Adaptation with Self-Training

Using top-1 softmax confidence or predictive entropy and self-train on highly confident instances!
→ Such confidence measures tend to be miscalibrated and are often unreliable!



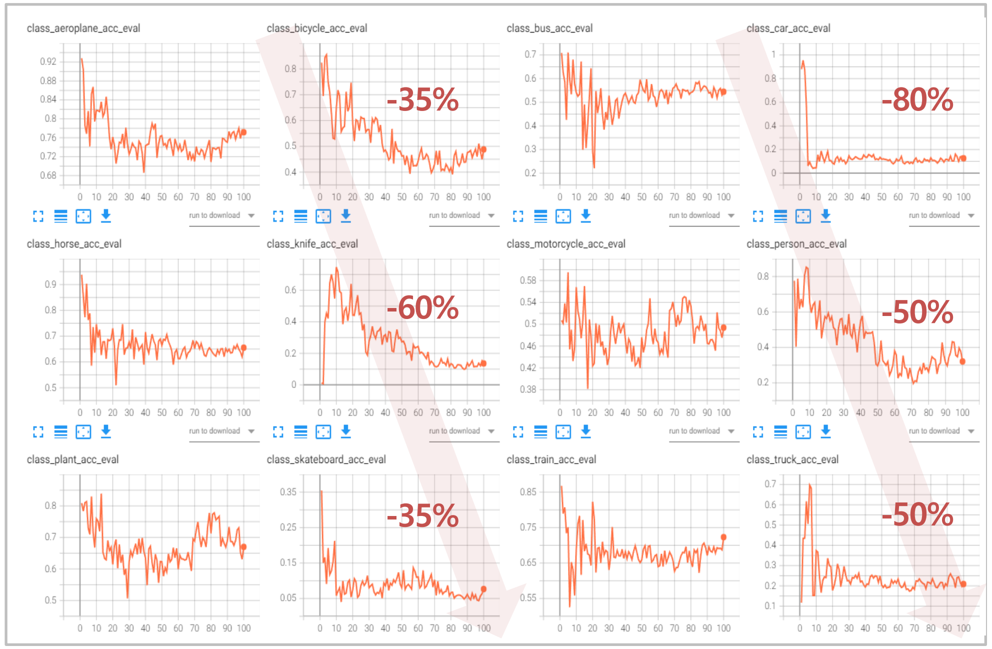
Introduction

Background on Unsupervised Domain Adaptation with Self-Training

❖ Limitations of Standard Self-Training

- Covariate shift로 인한 pseudo-labels 품질 저하
  - The distribution of pseudo-labels is significantly different from target ground-truth → Mostly misclassified into other classes!
- Note that classes 2, 7, 8 and 12 appear rarely in the target pseudo-labels in the covariate shift setting
  - Indicating that the pseudo-labels are biased towards several classes due to domain shift[2]

Target Domain의 Class 별 Accuracy



[2]] Liu, H., Wang, J., & Long, M. (2021). Cycle self-training for domain adaptation. Advances in Neural Information Processing Systems, 34, 22968-22981.

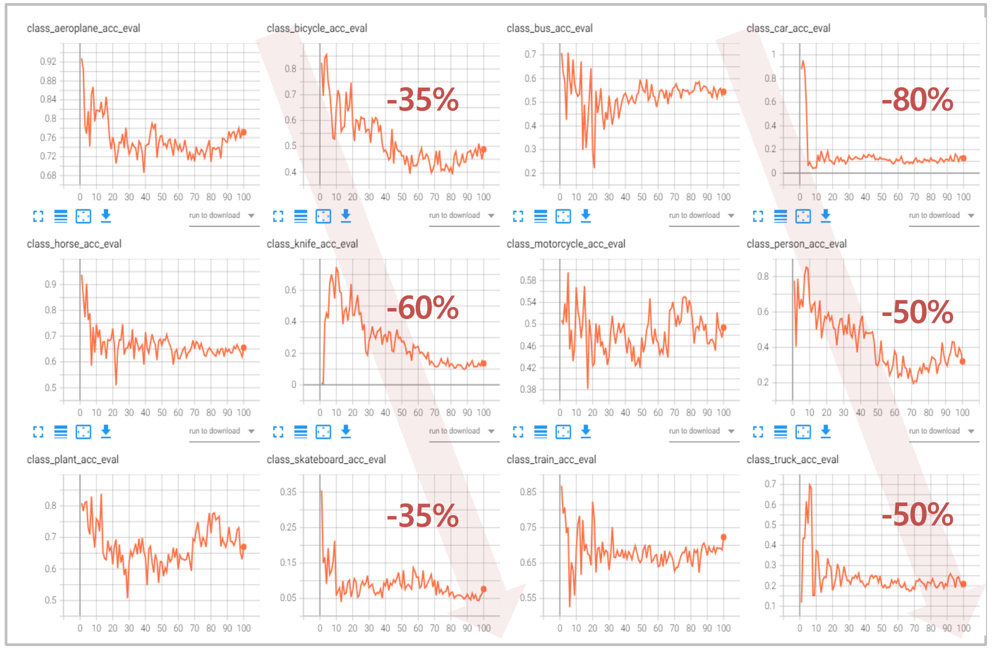
Introduction

Background on Unsupervised Domain Adaptation with Self-Training

❖ Limitations of Standard Self-Training

- Covariate shift로 인한 pseudo-labels 품질 저하
  - The distribution of pseudo-labels is significantly different from target ground-truth → Mostly misclassified into other classes!
- Note that classes 2, 7, 8 and 12 appear rarely in the target pseudo-labels in the covariate shift setting
  - Indicating that the **pseudo-labels are biased towards several classes** due to domain shift[2]

Target Domain의 Class 별 Accuracy



[2]] Liu, H., Wang, J., & Long, M. (2021). Cycle self-training for domain adaptation. Advances in Neural Information Processing Systems, 34, 22968-22981.

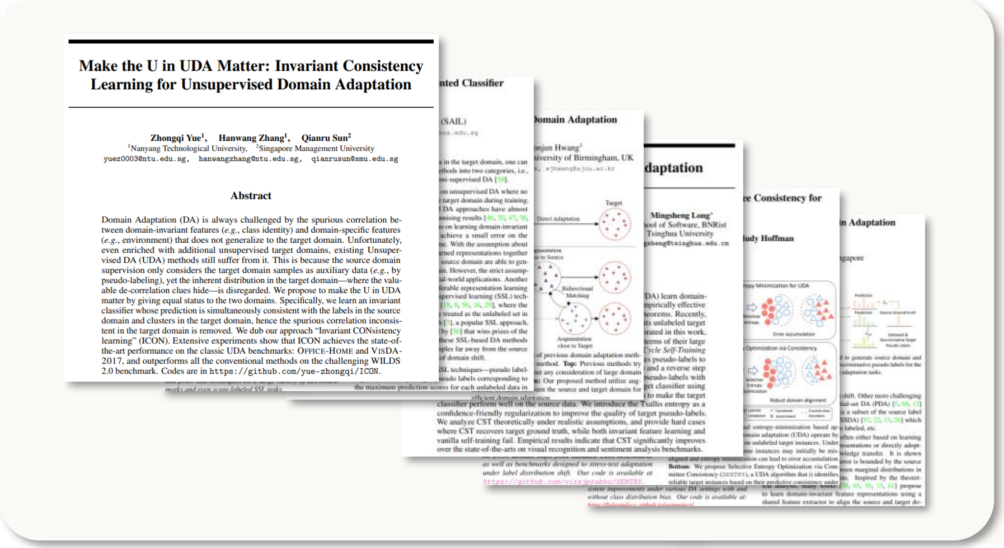
Introduction

Background on Unsupervised Domain Adaptation with Self-Training

❖ Unsupervised Domain Adaptation with Self-Training

연구 목적: Make reliable pseudo-labels under covariate shift and minimize target domain error!

Target pseudo-labels 생성을 위한 다양한 선행연구들

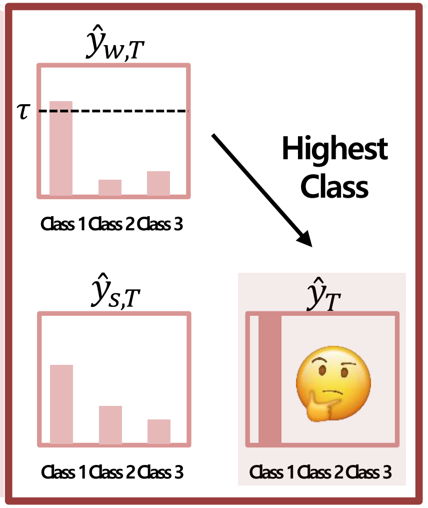


Make the U in UDA Matter: Invariant Consistency Learning for Unsupervised Domain Adaptation

Zhongqi Yue[1], Hanwang Zhang[1], Qianru Sun[2]
[1]Nanyang Technological University, [2]Singapore Management University
yuez0003@ntu.edu.sg, hanwangzhang@ntu.edu.sg, qianrusun@smu.edu.sg

Abstract

Domain Adaptation (DA) is always challenged by the spurious correlation between domain-invariant features (*e.g.*, class identity) and domain-specific features (*e.g.*, environment) that does not generalize to the target domain. Unfortunately, even enriched with additional unsupervised target domains, existing Unsupervised DA (UDA) methods still suffer from it. This is because the source domain supervision only considers the target domain samples as auxiliary data (*e.g.*, by pseudo-labeling), yet the inherent distribution in the target domain—where the valuable de-correlation clues hide—is disregarded. We propose to make the U in UDA matter by giving equal status to the two domains. Specifically, we learn an invariant classifier whose prediction is simultaneously consistent with the labels in the source domain and clusters in the target domain, hence the spurious correlation inconsistent in the target domain is removed. We dub our approach "Invariant CONsistency learning" (ICON). Extensive experiments show that ICON achieves the state-of-the-art performance on the classic UDA benchmarks: OFFICE-HOME and VISDA-2017, and outperforms all the conventional methods on the challenging WILDS 2.0 benchmark. Codes are in https://github.com/yue-zhongqi/ICON.



Limitation

분포 차이로 인해 믿을 수 있는 Pseudo-Labels 생성/선별 어려움[2]

[2]] Liu, H., Wang, J., & Long, M. (2021). Cycle self-training for domain adaptation. Advances in Neural Information Processing Systems, 34, 22968-22981.

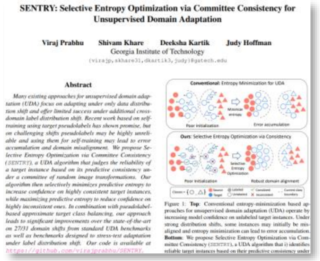
Related Works

선행연구

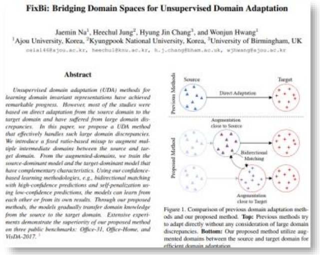
❖ 선행연구 (2021-2023)

- **2021년-2023년 중 제안된 self-training 기반 domain adaptation methods 비교 분석**
- 이 중 직접 구현까지 진행한 3가지 방법론 (SENTRY, CST, ICON)에 대해 세미나 진행

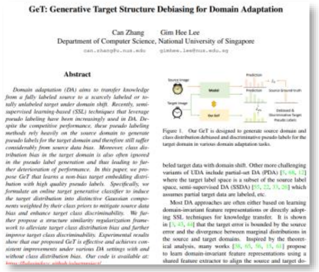
SENTRY
(ICCV 2021)


SENTRY: Selective Entropy Optimization via Committee Consistency for Unsupervised Domain Adaptation

FixBi
(CVPR 2021)


FixBi: Bridging Domain Spaces for Unsupervised Domain Adaptation

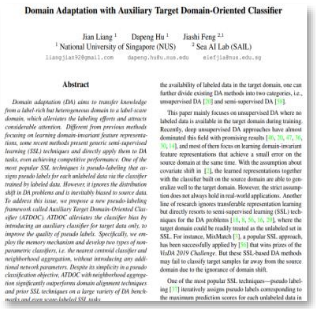
GeT
(ICCV 2023)


GeT: Generative Target Structure Debiasing for Domain Adaptation

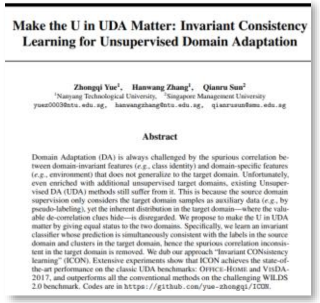
2021 → 2023


Cycle Self-Training for Domain Adaptation

CST
(NeurIPS 2021)


Domain Adaptation with Auxiliary Target Domain-Oriented Classifier

ATDOC
(CVPR 2021)


Make the U in UDA Matter: Invariant Consistency Learning for Unsupervised Domain Adaptation

ICON
(NeurIPS 2023)

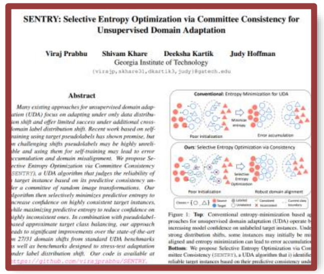
Related Works

선행연구

❖ 선행연구 (2021-2023)

- **2021년-2023년 중 제안된 self-training 기반 domain adaptation methods** 비교 분석
- 이 중 직접 구현까지 진행한 **3가지 방법론 (SENTRY, CST, ICON)**에 대해 세미나 진행

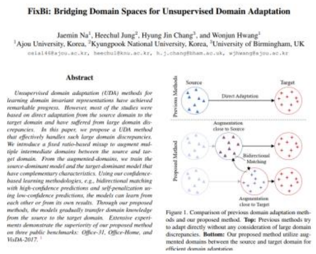
SENTRY
(ICCV 2021)


SENTRY: Selective Entropy Optimization via Committee Consistency for Unsupervised Domain Adaptation

Viraj Prabhu Shivam Khare Deeksha Kartik Judy Hoffman
Georgia Institute of Technology

Abstract

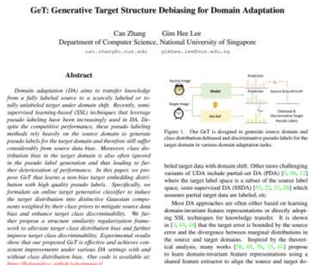
FixBi
(CVPR 2021)


FixBi: Bridging Domain Spaces for Unsupervised Domain Adaptation

Jaemin Na, Heechul Jung, Hyung Jin Chang, and Wonjun Hwang

Abstract

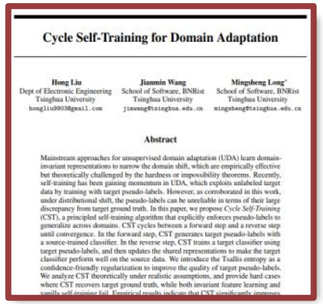
GeT
(ICCV 2023)


GeT: Generative Target Structure Debiasing for Domain Adaptation

Can Zhang Gim Hee Lee
Department of Computer Science, National University of Singapore

Abstract

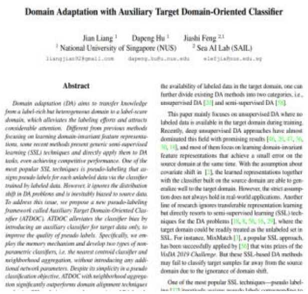
2021 → 2023


Cycle Self-Training for Domain Adaptation

Hong Liu Jianmin Wang Mingsheng Long

Abstract

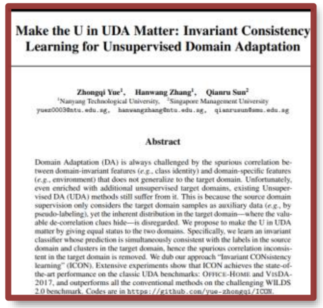
CST
(NeurIPS 2021)


Domain Adaptation with Auxiliary Target Domain-Oriented Classifier

Jian Liang Dapeng Hu Jiashi Feng

Abstract

ATDOC
(CVPR 2021)


Make the U in UDA Matter: Invariant Consistency Learning for Unsupervised Domain Adaptation

Zhongqi Yue, Hanwang Zhang, Qianru Sun

Abstract

ICON
(NeurIPS 2023)

UDA with Self-Training

[2021 ICCV]
SENTRY: Selective Entropy Optimization via Committee Consistency for Unsupervised Domain Adaptation

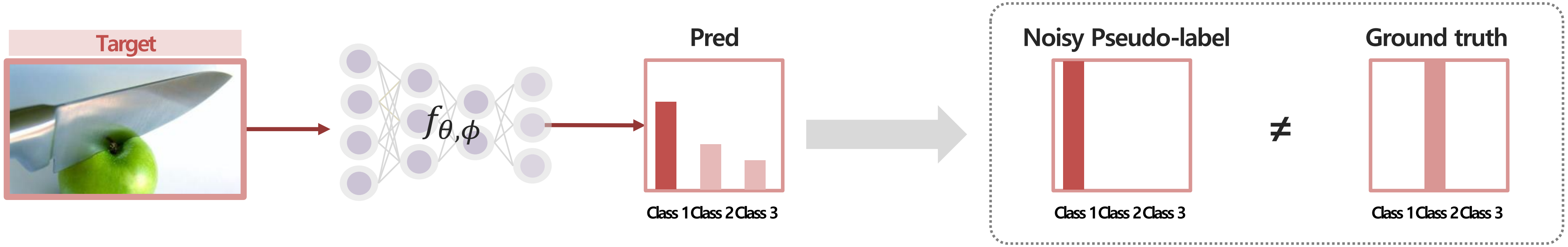
Viraj et al., Georgia Institute of Technology

SENTRY

SENTRY: Selective Entropy Optimization via Committee Consistency for Unsupervised Domain Adaptation

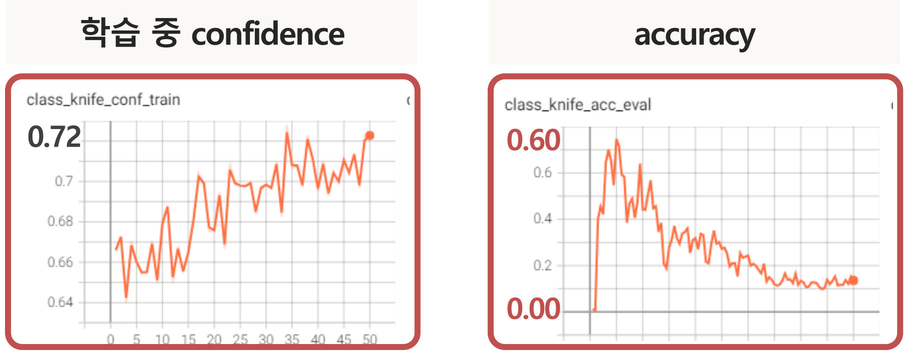
❖ **Motivation: Target pseudo-labels may be highly unreliable and using them may lead to error accumulation!**

- Previous works rely on self-training using noisy pseudo-labels or conditional entropy minimization over miscalibrated predictions



Using ***top-1 softmax confidence*** (or predictive entropy) and only self-train on highly confident instances

Class: knife



Wrong but high-confidence!
높은 확신을 가지고 잘못된 예측 수행

→ **Lead to error accumulation!**

SENTRY

SENTRY: Selective Entropy Optimization via Committee Consistency for Unsupervised Domain Adaptation

❖ **Question. How can we identify reliable target instances?**

SENTRY: Selective Entropy Optimization via Committee Consistency for Unsupervised Domain Adaptation

Viraj Prabhu **Shivam Khare** **Deeksha Kartik** **Judy Hoffman**
Georgia Institute of Technology
{virajp,skhare31,dkartik3,judy}@gatech.edu

Abstract

Many existing approaches for unsupervised domain adaptation (UDA) focus on adapting under only data distribution shift and offer limited success under additional cross-domain label distribution shift. Recent work based on self-training using target pseudolabels has shown promise, but on challenging shifts pseudolabels may be highly unreliable and using them for self-training may lead to error accumulation and domain misalignment. We propose Selective Entropy Optimization via Committee Consistency (SENTRY), a UDA algorithm that judges the reliability of a target instance based on its predictive consistency under a committee of random image transformations. Our algorithm then selectively minimizes predictive entropy to increase confidence on highly consistent target instances, while maximizing predictive entropy to reduce confidence on highly inconsistent ones. In combination with pseudolabel-based approximate target class balancing, our approach leads to significant improvements over the state-of-the-art on 27/31 domain shifts from standard UDA benchmarks as well as benchmarks designed to stress-test adaptation under label distribution shift. Our code is available at https://github.com/virajprabhu/SENTRY.

1. Introduction

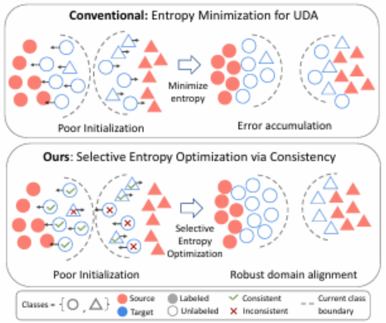


Figure 1: **Top**: Conventional entropy-minimization based approaches for unsupervised domain adaptation (UDA) operate by increasing model confidence on unlabeled target instances. Under strong distribution shifts, some instances may initially be misaligned and entropy minimization can lead to error accumulation. **Bottom**: We propose Selective Entropy Optimization via Committee Consistency (SENTRY), a UDA algorithm that i) identifies reliable target instances based on their predictive consistency under a set of random image transformations, and ii) selectively optimizes model entropy on these instances to induce domain alignment.

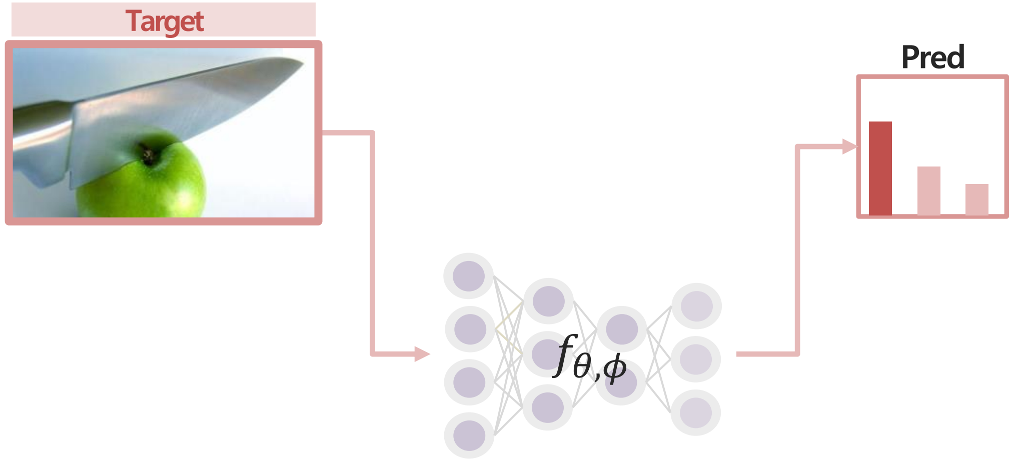
Answer. Using ***predictive consistency*** under a committee of label-preserving image transformations!

→ Selective Entropy Optimization via Committee Consistency (SENTRY)

SENTRY

SENTRY: Selective Entropy Optimization via Committee Consistency for Unsupervised Domain Adaptation

❖ **Question. How can we identify reliable target instances?**



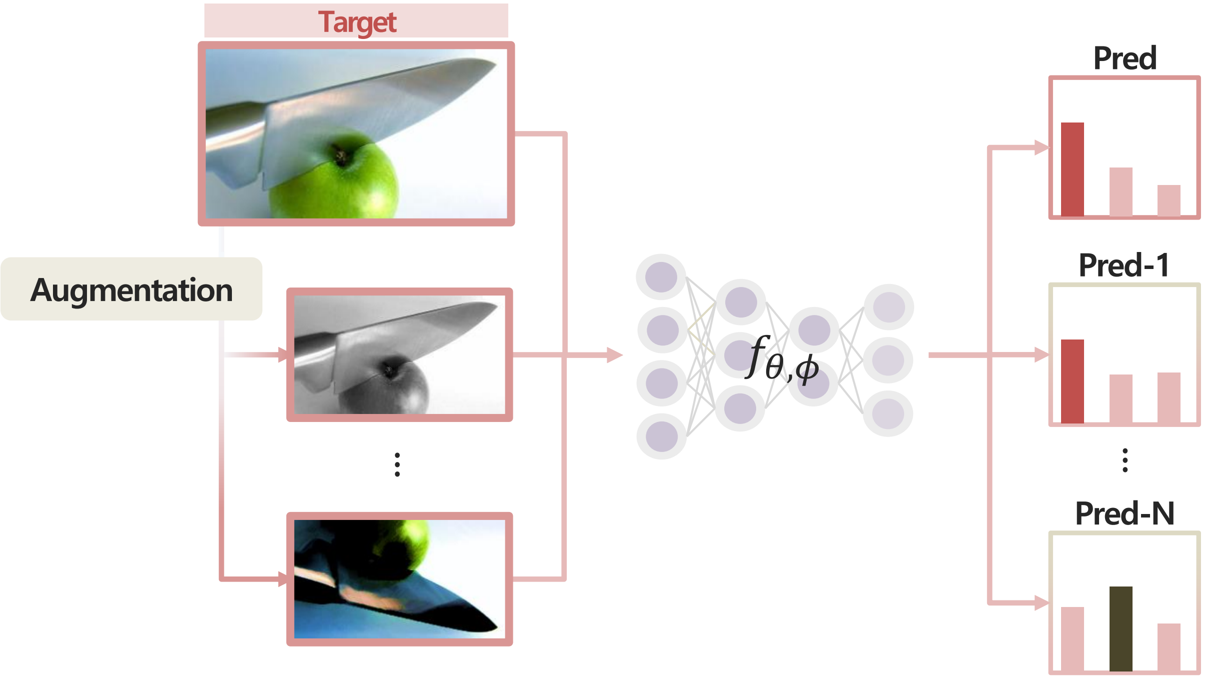
Answer. Using ***predictive consistency*** under a committee of label-preserving image transformations!

→ Selective Entropy Optimization via Committee Consistency (SENTRY)

SENTRY

SENTRY: Selective Entropy Optimization via Committee Consistency for Unsupervised Domain Adaptation

❖ **Question. How can we identify reliable target instances?**



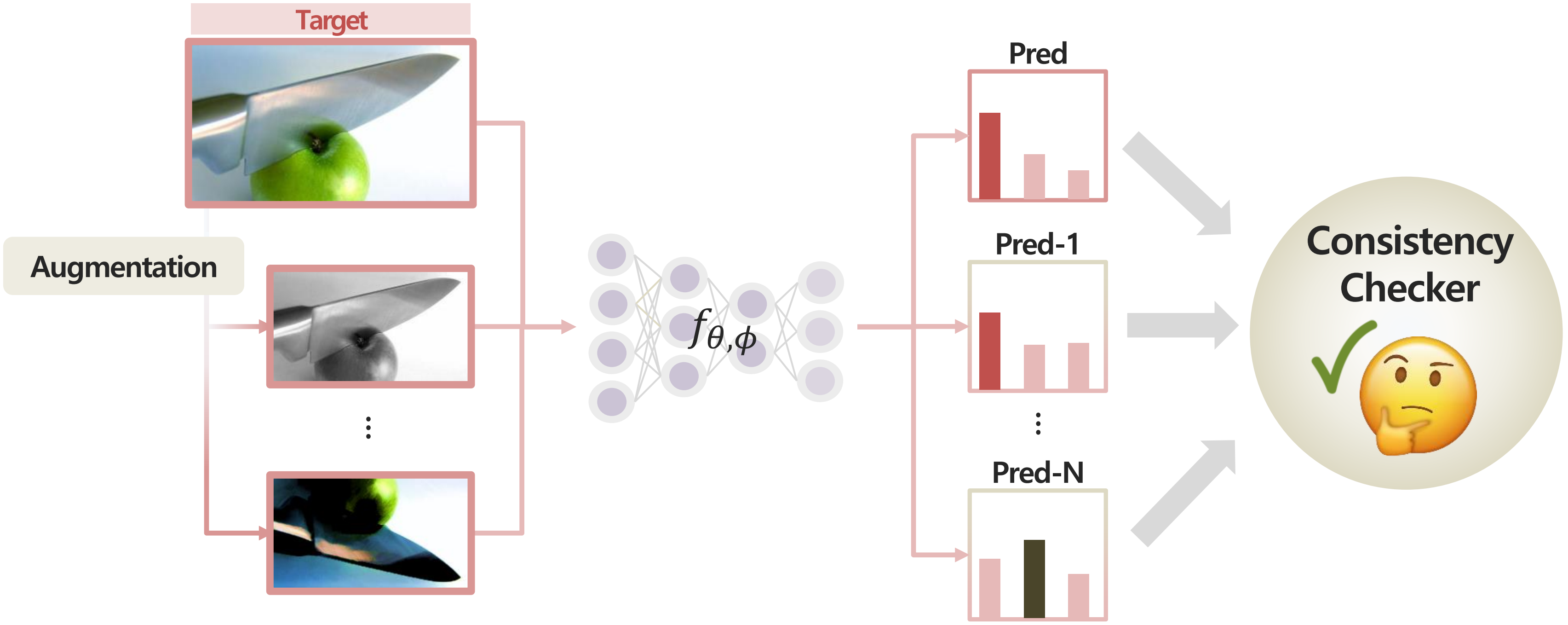
Answer. Using ***predictive consistency*** under a committee of label-preserving image transformations!

→ Selective Entropy Optimization via Committee Consistency (SENTRY)

SENTRY

SENTRY: Selective Entropy Optimization via Committee Consistency for Unsupervised Domain Adaptation

❖ **Question. How can we identify reliable target instances?**



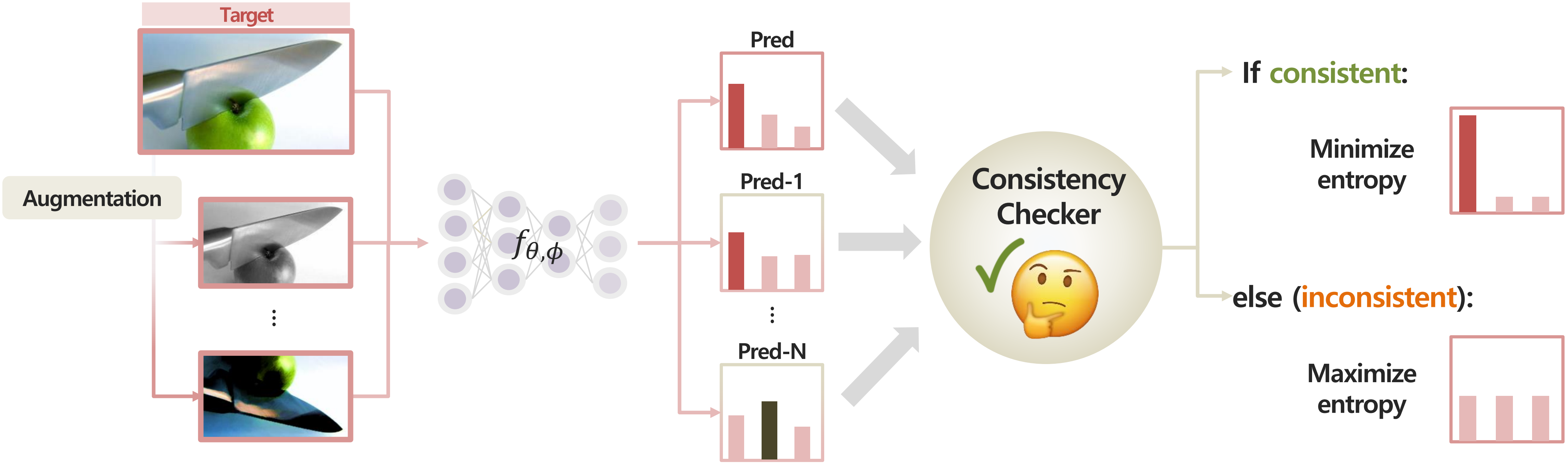
Answer. Using ***predictive consistency*** under a committee of label-preserving image transformations!

→ Selective Entropy Optimization via Committee Consistency (SENTRY)

SENTRY

SENTRY: Selective Entropy Optimization via Committee Consistency for Unsupervised Domain Adaptation

❖ **Question. How can we identify reliable target instances?**



Answer. Using ***predictive consistency*** under a committee of label-preserving image transformations!

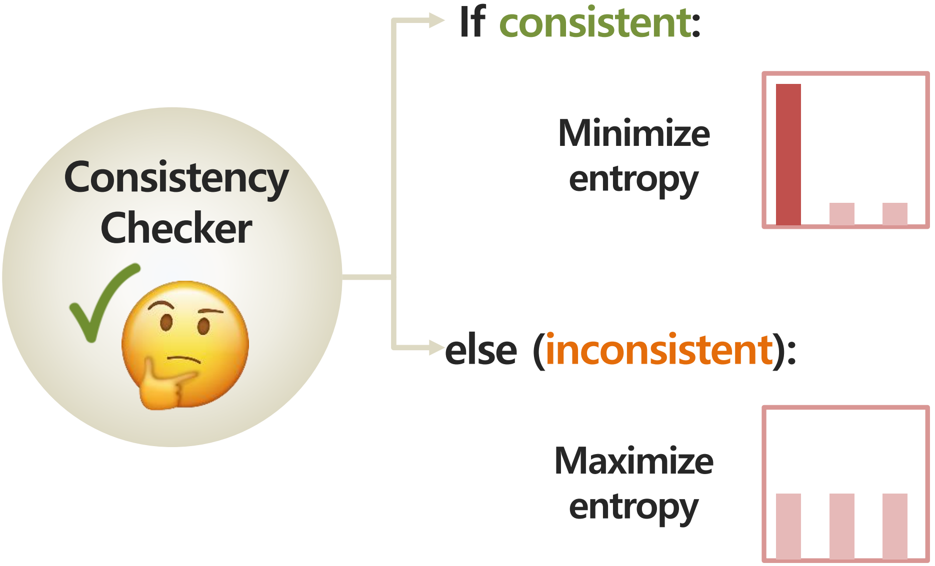
→ Selective Entropy Optimization via Committee Consistency (SENTRY)

SENTRY

SENTRY: Selective Entropy Optimization via Committee Consistency for Unsupervised Domain Adaptation

❖ Question. How can we identify reliable target instances?

Concern 1. Entropy minimization only on consistent instances might lead to the exclusion of a large percentage of target instances!



Answer. Using *predictive consistency* under a committee of label-preserving image transformations!

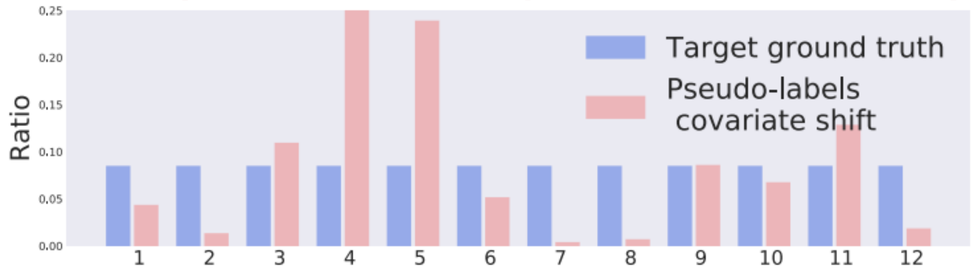
→ Selective Entropy Optimization via Committee Consistency (SENTRY)

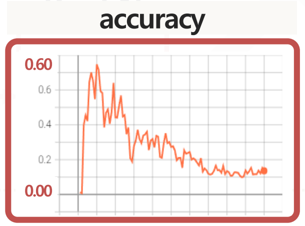
SENTRY

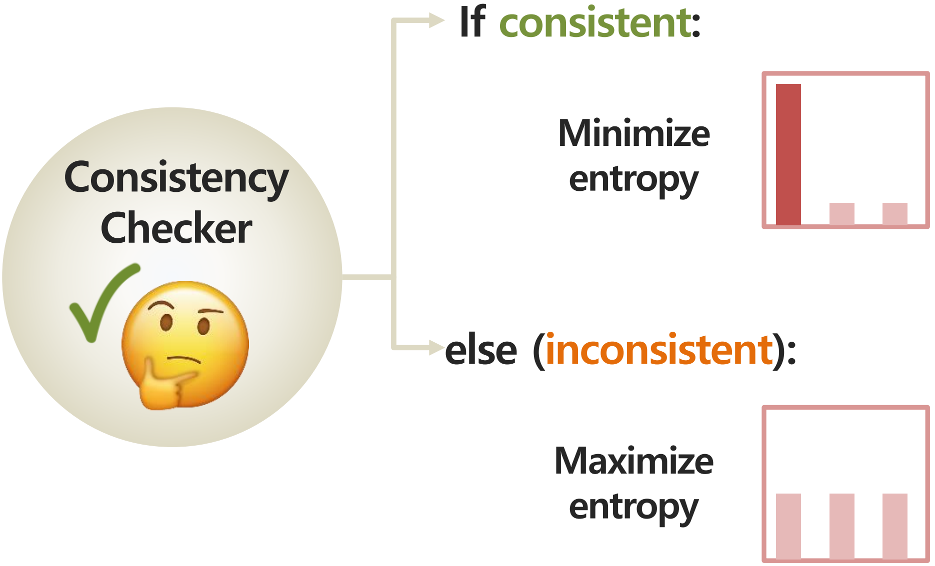
SENTRY: Selective Entropy Optimization via Committee Consistency for Unsupervised Domain Adaptation

Concern 1. Entropy minimization only on consistent instances might lead to the exclusion of a large percentage of target instances!

Concern 2. Indefinite entropy maximization on inconsistent target instances might prove detrimental to learning





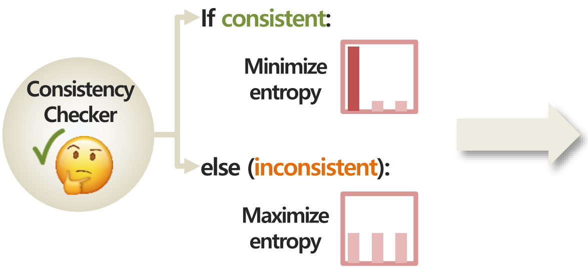


Answer. Both of these concerns are addressed via the ***adaptive selection via augmentation invariance regularization***.

SENTRY

SENTRY: Selective Entropy Optimization via Committee Consistency for Unsupervised Domain Adaptation

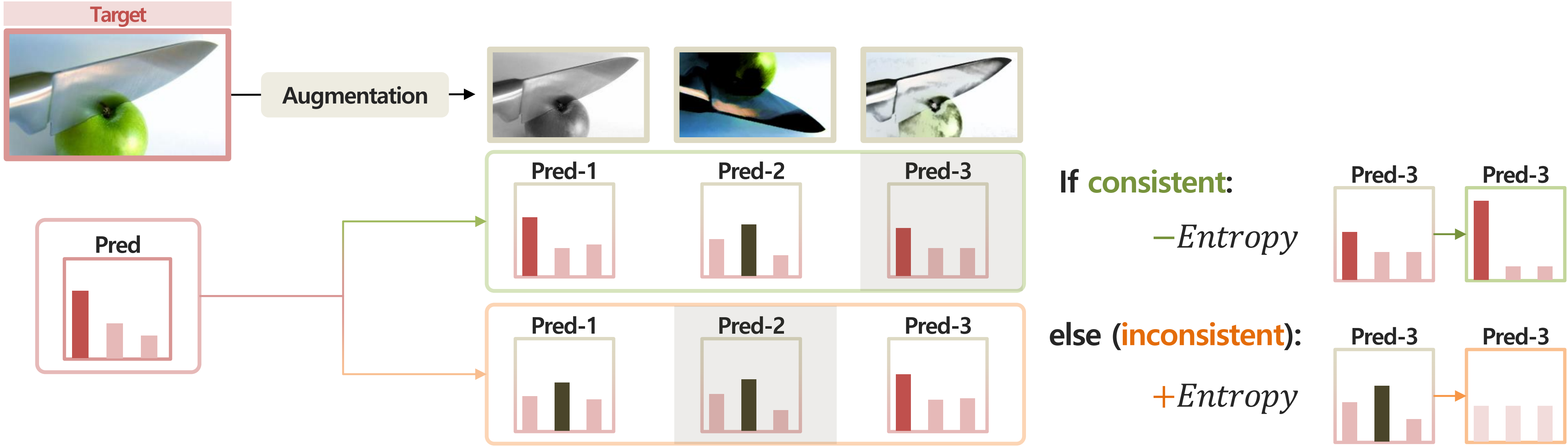
❖ **Question. What is the '*adaptive selection via augmentation invariance regularization*'?**



$$\mathcal{L}_{SENTRY}(x_T) = \begin{cases} -Entropy(y|aug_i(x_T)), & if\ consistent \\ +Entropy(y|aug_j(x_T)), & if\ inconsistent \end{cases}$$

Using LAST AUGMENTED VERSION rather than the original image itself

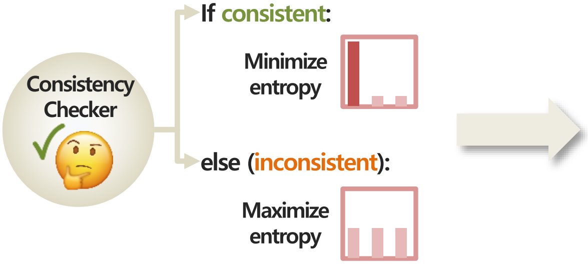
(i and j denote the index of the last consistent and inconsistent transformed version, respectively)



SENTRY

SENTRY: Selective Entropy Optimization via Committee Consistency for Unsupervised Domain Adaptation

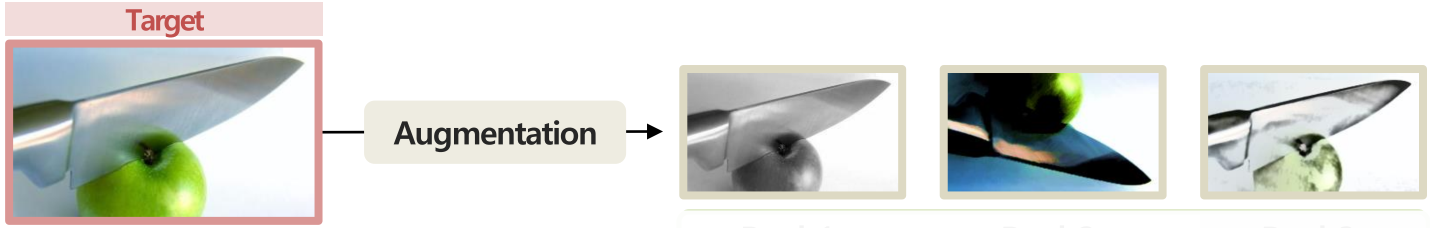
❖ **Question. What is the '*adaptive selection via augmentation invariance regularization*'?**



$$\mathcal{L}_{SENTRY}(x_T) = \begin{cases} -Entropy(y|aug_i(x_T)), & if\ consistent \\ +Entropy(y|aug_j(x_T)), & if\ inconsistent \end{cases}$$

Using LAST AUGMENTED VERSION rather than the original image itself

(i and j denote the index of the last consistent and inconsistent transformed version, respectively)



Benefit 1. Reduce Overfitting

- Low-entropy predictions across various transformations of the input image
- It helps the model become more robust to small variations in the input

Benefit 2. Augmentation Invariance

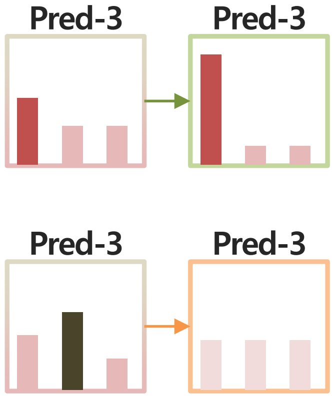
- Encouraging transformation-invariant features (more robust and generalizable features)
- More instances would be selected for entropy minimization as training progresses, making the selection process adaptive

If consistent:

$-Entropy$

else (inconsistent):

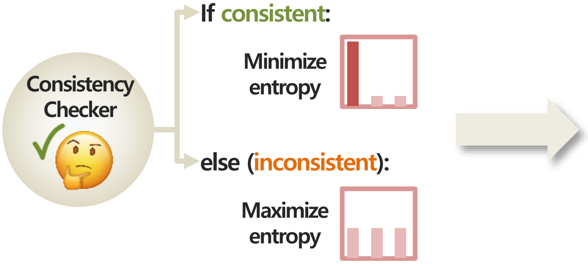
$+Entropy$



SENTRY

SENTRY: Selective Entropy Optimization via Committee Consistency for Unsupervised Domain Adaptation

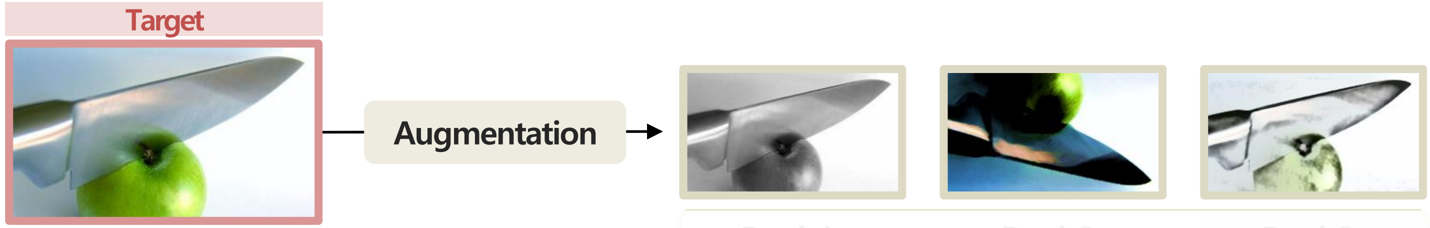
❖ **Question. What is the '*adaptive selection via augmentation invariance regularization*'?**

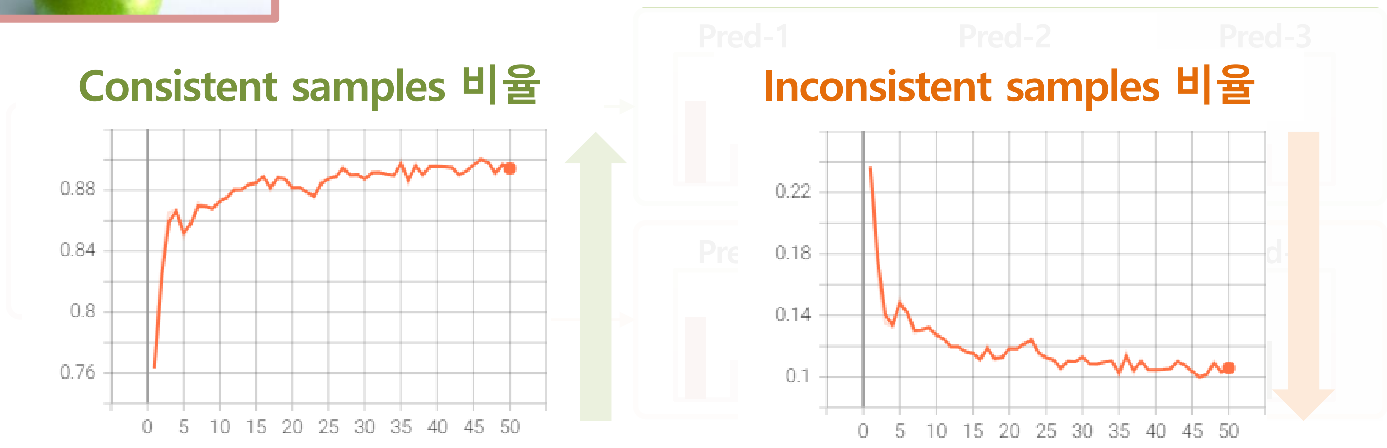


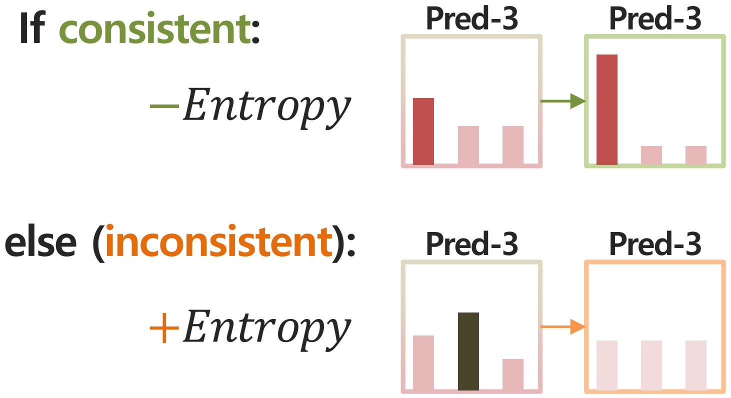
$$\mathcal{L}_{SENTRY}(x_T) = \begin{cases} -Entropy(y|aug_i(x_T)), & if\ consistent \\ +Entropy(y|aug_j(x_T)), & if\ inconsistent \end{cases}$$

Using LAST AUGMENTED VERSION rather than the original image itself

(i and j denote the index of the last consistent and inconsistent transformed version, respectively)







SENTRY

SENTRY: Selective Entropy Optimization via Committee Consistency for Unsupervised Domain Adaptation

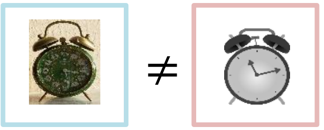
❖ **Question. How can we deal with the problem of label distribution shift (LDS)?**

I.I.D

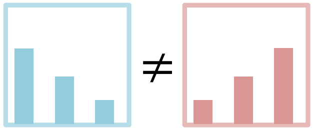
LDS

Covariate Shift, where $P_S(y|x) = P_T(y|x)$ for all x, but $P_S(x) \neq P_T(x)$;
Label Shift, where $P_S(x|y) = P_T(x|y)$ for all y, but $P_S(y) \neq P_T(y)$

$P_S(x)$ $\neq$ $P_T(x)$



$P_S(y)$ $\neq$ $P_T(y)$



Typical conditional entropy minimization method has been found to potentially encourage **trivial solutions of only predicting the majority class**

Answer ①. Pseudo Class Balancing

Source: 실제 label을 사용한 class-balanced sampling

Target: pseudo-labels을 활용한 approximate class-balanced sampling

Batch
$B = 9$

Class 1

Class 2

Class 3

SENTRY

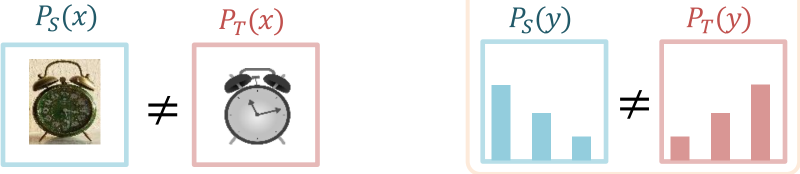
SENTRY: Selective Entropy Optimization via Committee Consistency for Unsupervised Domain Adaptation

❖ **Question. How can we deal with the problem of label distribution shift (LDS)?**

I.I.D

LDS

Covariate Shift, where $P_S(y|x) = P_T(y|x)$ for all x, but $P_S(x) \neq P_T(x)$;
Label Shift, where $P_S(x|y) = P_T(x|y)$ for all y, but $P_S(y) \neq P_T(y)$



Typical conditional entropy minimization method has been found to potentially encourage **trivial solutions of only predicting the majority class**

Answer ①. Pseudo Class Balancing

Source: 실제 label을 사용한 class-balanced sampling
Target: pseudo-labels을 활용한 approximate class-balanced sampling

Batch
$B = 9$

Class 1
Class 2
Class 3

Answer ②. Information Entropy Loss $\mathcal{L}_{IE}$

Target domain에서 모델이 다양한 예측을 하도록 장려
→ 특정 class로 예측이 치우치는 것을 방지

모델이 예측한 class c의 확률

$$\mathcal{L}_{IE} = \mathbb{E}_{x_T \sim P_T}\left[\sum_{c=1}^{K} p_\theta(y = c|x_T) \log q(\hat{y} = c)\right]$$

마지막 Q개 샘플에 대한 모델 예측 분포

$q(\hat{y} = c)$가 **uniform distribution에 가까울 수록 loss ↓**

SENTRY

SENTRY: Selective Entropy Optimization via Committee Consistency for Unsupervised Domain Adaptation

❖ **Question. How can we deal with the problem of label distribution shift (LDS)?**

$q(\hat{y}=c)$ → Uniform distribution

Q=5, 최근 5개 예측: c=1, c=1, c=2, c=2, c=3

$$q(\hat{y}=c) = \{class\ 1{:}\frac{2}{5}, class\ 2{:}\frac{2}{5}, class\ 3{:}\frac{1}{5}\}$$

$$p_\theta(y=c|x_T) = \{class\ 1{:}\ 0.7, class\ 2{:}\ 0.2, class\ 3{:}\ 0.1\}$$

$$\mathcal{L}_{IE} = 0.7 * \log\frac{2}{5} + 0.2 * \log\frac{2}{5} + 0.1 * \log\frac{1}{5}$$

$$= 0.7 * (-0.92) + 0.2 * (-0.92) + 0.1 * (-1.61) = -1.02$$

$q(\hat{y}=c)$ → Skewed distribution

Q=5, 최근 5개 예측: c=1, c=1, c=1, c=1, c=2

$$q(\hat{y}=c) = \{class\ 1{:}\ \frac{4}{5}, class\ 2{:}\frac{1}{5}, class\ 3{:}\ 0\}$$

$$p_\theta(y=c|x_T) = \{class\ 1{:}\ 0.7, class\ 2{:}\ 0.2, class\ 3{:}\ 0.1\}$$

$$\mathcal{L}_{IE} = 0.7 * \log\frac{4}{5} + 0.2 * \log\frac{1}{5} + 0.1 * \log\varepsilon$$

$$= 0.7 * (-0.22) + 0.2 * (-1.61) + 0.1 * (-\text{매우 큰 음수}) \ll -1.02$$

Answer ①. Pseudo Class Balancing

Source: 실제 label을 사용한 class-balanced sampling

Target: pseudo-labels을 활용한 approximate class-balanced sampling

Batch
$B = 9$

Class 1

Class 2

Class 3

Answer ②. Information Entropy Loss $\mathcal{L}_{IE}$

Target domain에서 모델이 다양한 예측을 하도록 장려

→ 특정 class로 예측이 치우치는 것을 방지

$$\mathcal{L}_{IE} = \mathbb{E}_{x_T \sim P_T}\left[\sum_{c=1}^{K} \underbrace{p_\theta(y=c|x_T)}_{\text{모델이 예측한 class c의 확률}} \log \underbrace{q(\hat{y}=c)}_{\text{마지막 Q개 샘플에 대한 모델 예측 분포}}\right]$$

$q(\hat{y}=c)$가 **uniform distribution에 가까울 수록 loss ↓**

SENTRY

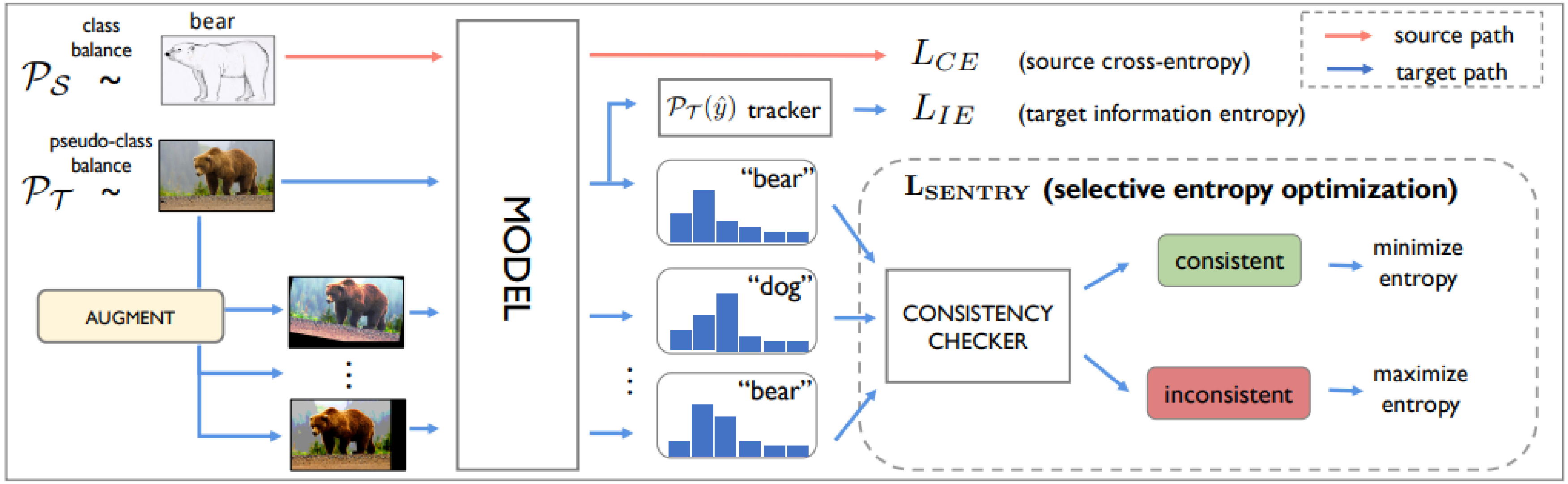
SENTRY: Selective Entropy Optimization via Committee Consistency for Unsupervised Domain Adaptation

❖ Summary

- Complete objective : $\underset{\Theta}{\text{argmin}} \; \mathbb{E}_{(\mathbf{x}_S, y_S) \overset{\text{bal}}{\sim} \mathcal{P}_S} \mathcal{L}_{CE} + \mathbb{E}_{\mathbf{x}_T \overset{\text{pbal}}{\sim} \mathcal{P}_T} \lambda_{IE} \mathcal{L}_{IE} + \lambda_{\text{SENTRY}} \mathcal{L}_{\text{SENTRY}}$

Phase 1. Pretraining with labeled source domain
Phase 2. Adaptation (SENTRY)

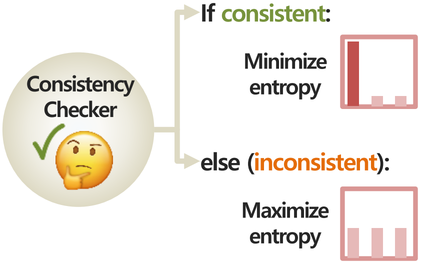
- To identify reliable target instances ($\mathcal{L}_{SENTRY}$):
  - AS-IS: using model confidence (miscalibrated) vs. TO-BE: using predictive consistency
  - Propose selective entropy optimization objective: minimize entropy if consistent else maximize entropy
- To address the problem of label distribution shift (LDS), class-balanced sampling on the source and target and $\mathcal{L}_{IE}$ are used



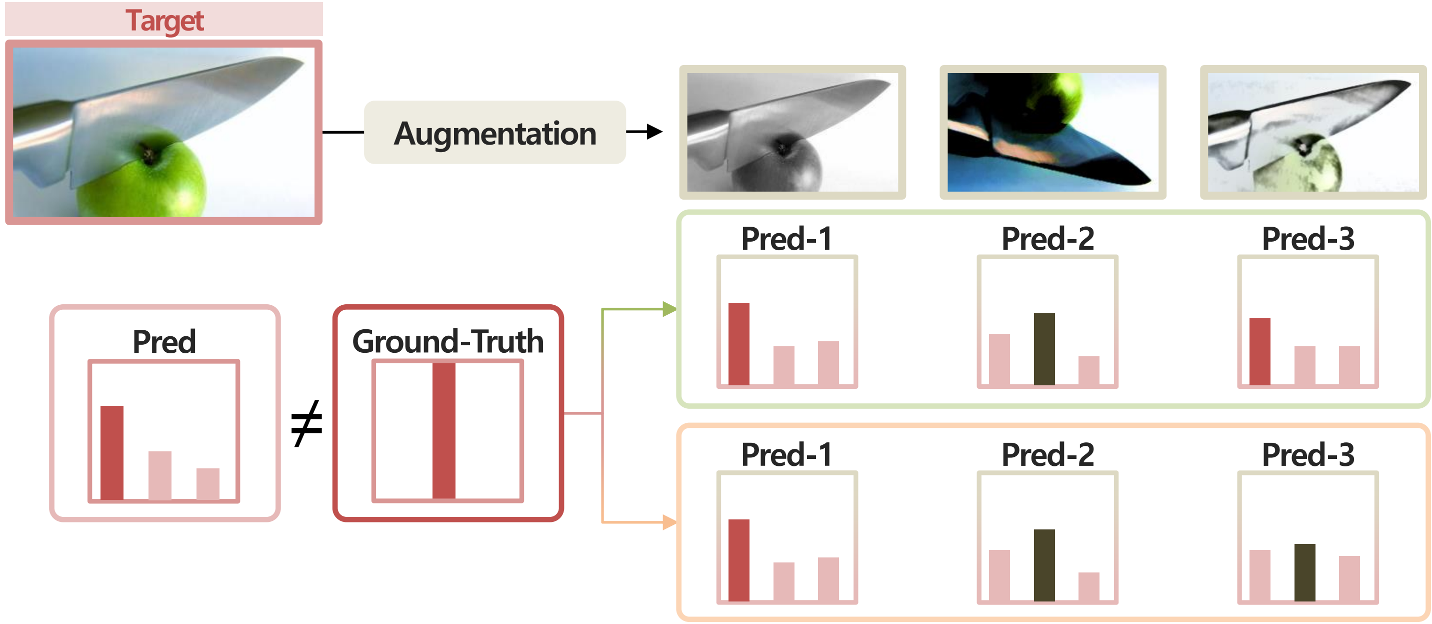
SENTRY

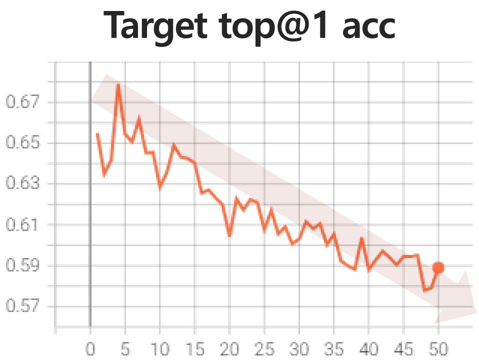
SENTRY: Selective Entropy Optimization via Committee Consistency for Unsupervised Domain Adaptation

❖ SENTRY 한계



$$\mathcal{L}_{SENTRY}(x_T) = \begin{cases} -Entropy(y|aug_i(x_T)), & if\ consistent \\ +Entropy(y|aug_j(x_T)), & if\ inconsistent \end{cases}$$





예측 일관성에 대한 규제가
예측 정확성까지 보장할까?

UDA with Self-Training

[2021 NeurIPS]
Cycle Self-Training for Domain Adaptation

Liu et al., Tsinghua University

CST

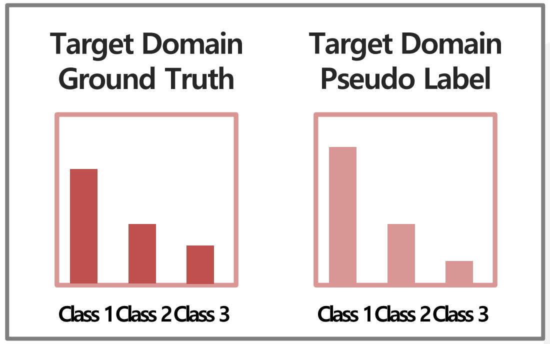
Cycle Self-Training for Domain Adaptation

❖ Limitations of Standard Self-Training

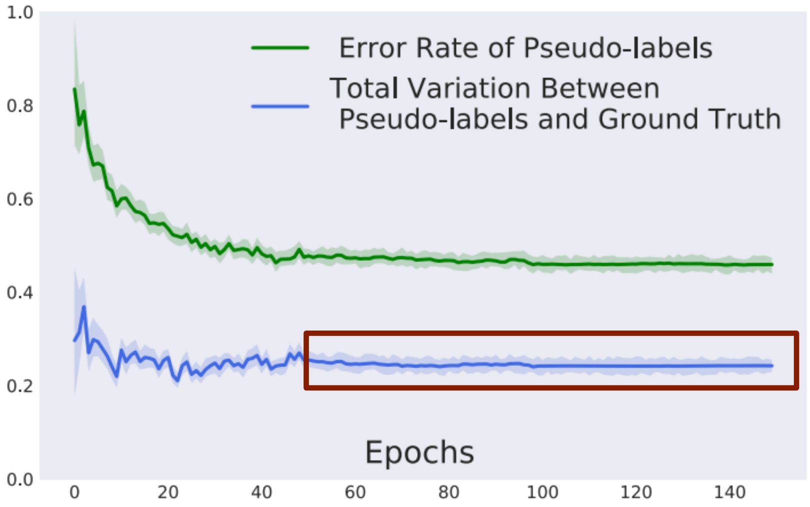
- Total variation distance (d_{TV})를 통해 실제 레이블 분포와 예측 레이블 분포 간 차이를 계산하여 biased pseudo labels 개선 필요성 강조
  - 실제 label 분포와 예측 확률 분포 간의 차이 이상으로 pseudo label이 더 잘 생성되지는 않음 (정확도의 상한)
  - 기존의 self-training 방식으로 학습하면 d_{TV}가 수렴(0.26)하고, 이에 따라 pseudo-labels은 0.74 이상의 정확도를 가질 수 없음 (품질 개선의 한계)

d_{TV}

$$d_{TV}(class\_ratio_i - pseudo\_class\_ratio_i) = \frac{1}{2}\sum_i \|class\_ratio_i - pseudo\_class\_ratio_i\|$$



Lower bound of the error rate of the pseudo-labels



d_{TV} converges to 0.26

↓

acc of pseudo-labels is then upper-bounded by 0.74

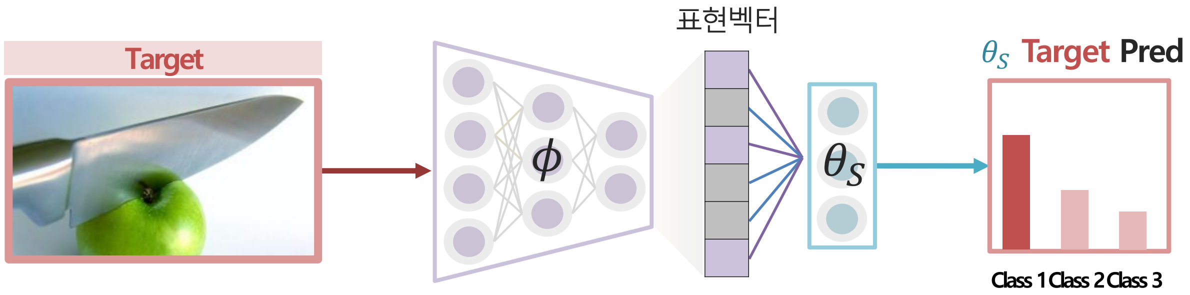
Denoising ability of pseudo-labels is needed!

CST

Cycle Self-Training for Domain Adaptation

❖ **Question. How to refine noisy pseudo-labels?**

Forward Step



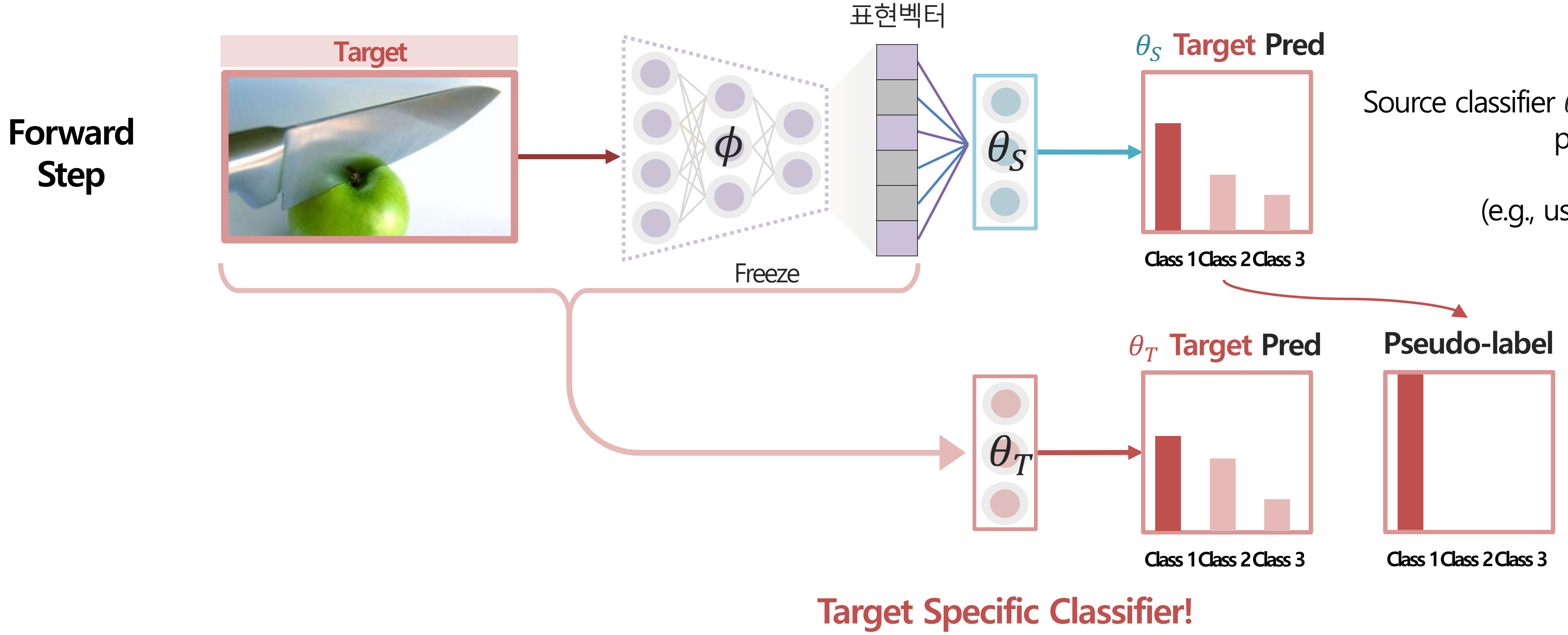
Source classifier θ_S를 이용하여 unlabeled target의 pseudo-labels 생성

(e.g., using FixMatch technique)

CST

Cycle Self-Training for Domain Adaptation

❖ **Question. How to refine noisy pseudo-labels?**



Source classifier θ_S를 이용하여 unlabeled target의 pseudo-labels 생성

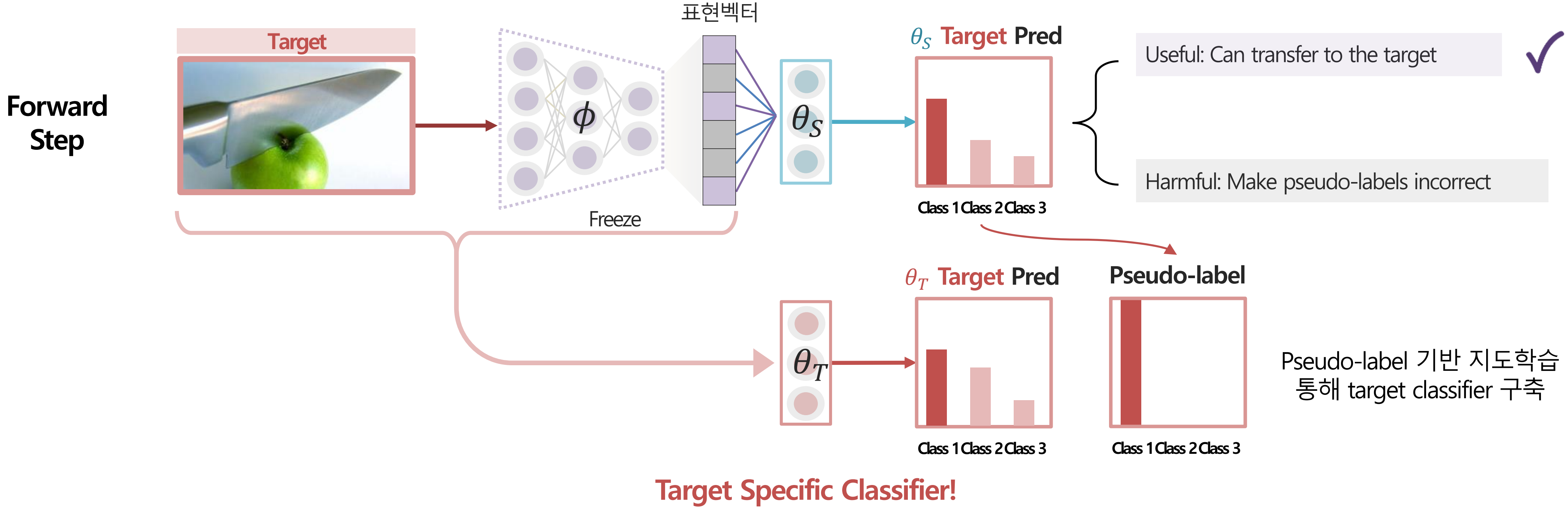
(e.g., using FixMatch technique)

Pseudo-label 기반 지도학습 통해 target classifier 구축

CST

Cycle Self-Training for Domain Adaptation

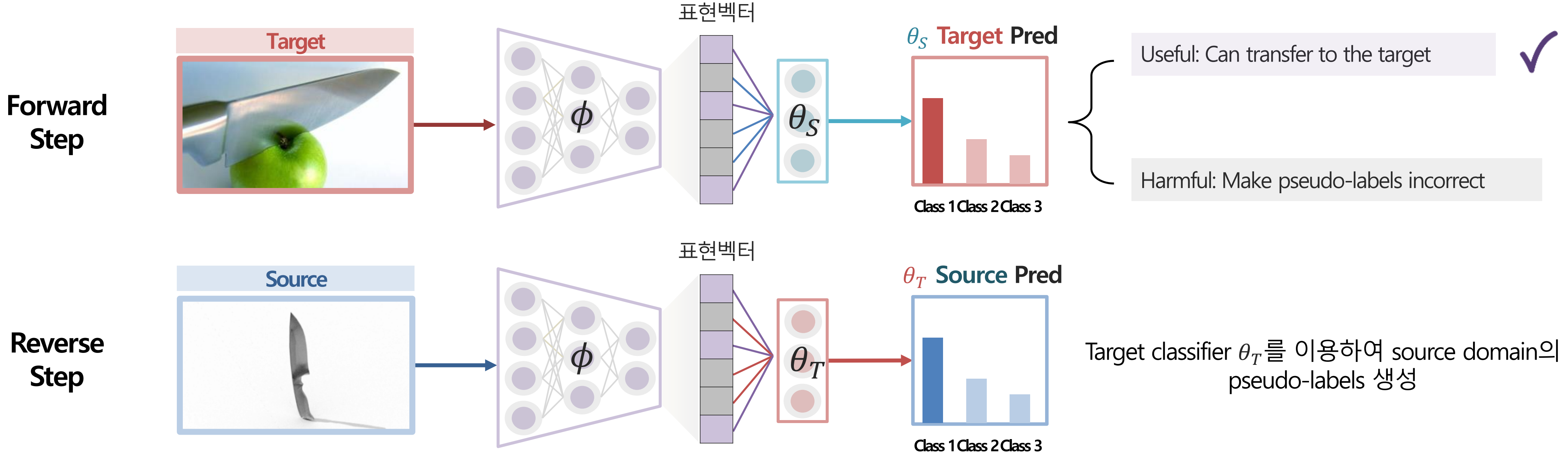
❖ **Question. How to refine noisy pseudo-labels?**



CST

Cycle Self-Training for Domain Adaptation

❖ **Question. How to refine noisy pseudo-labels?**

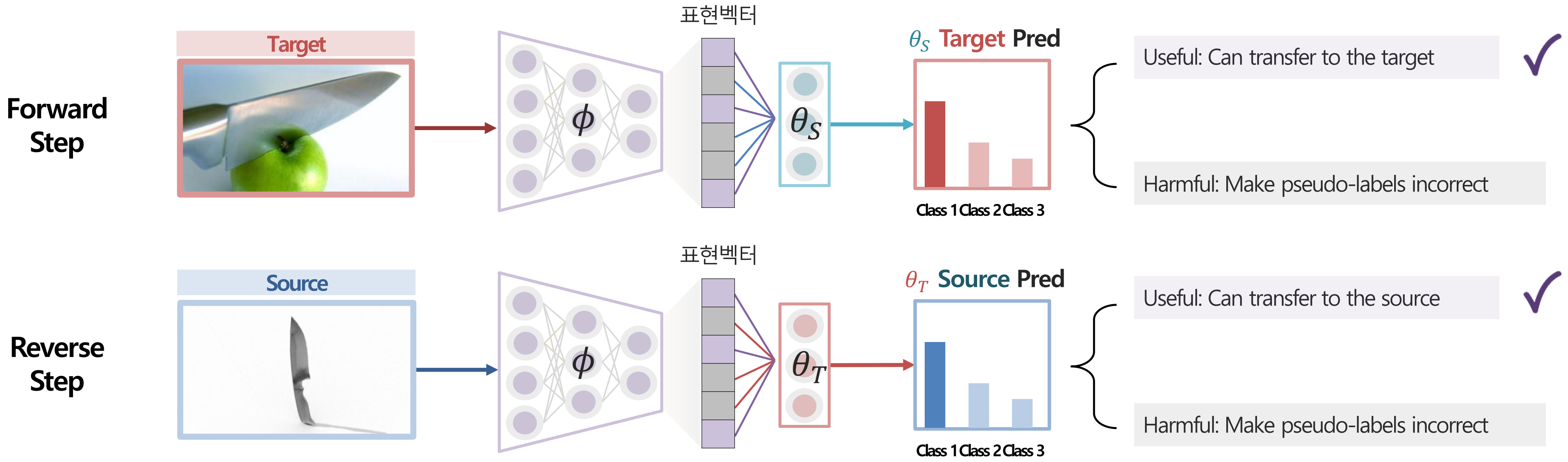


Target classifier θ_T를 이용하여 source domain의 pseudo-labels 생성

CST

Cycle Self-Training for Domain Adaptation

❖ **Question. How to refine noisy pseudo-labels?**



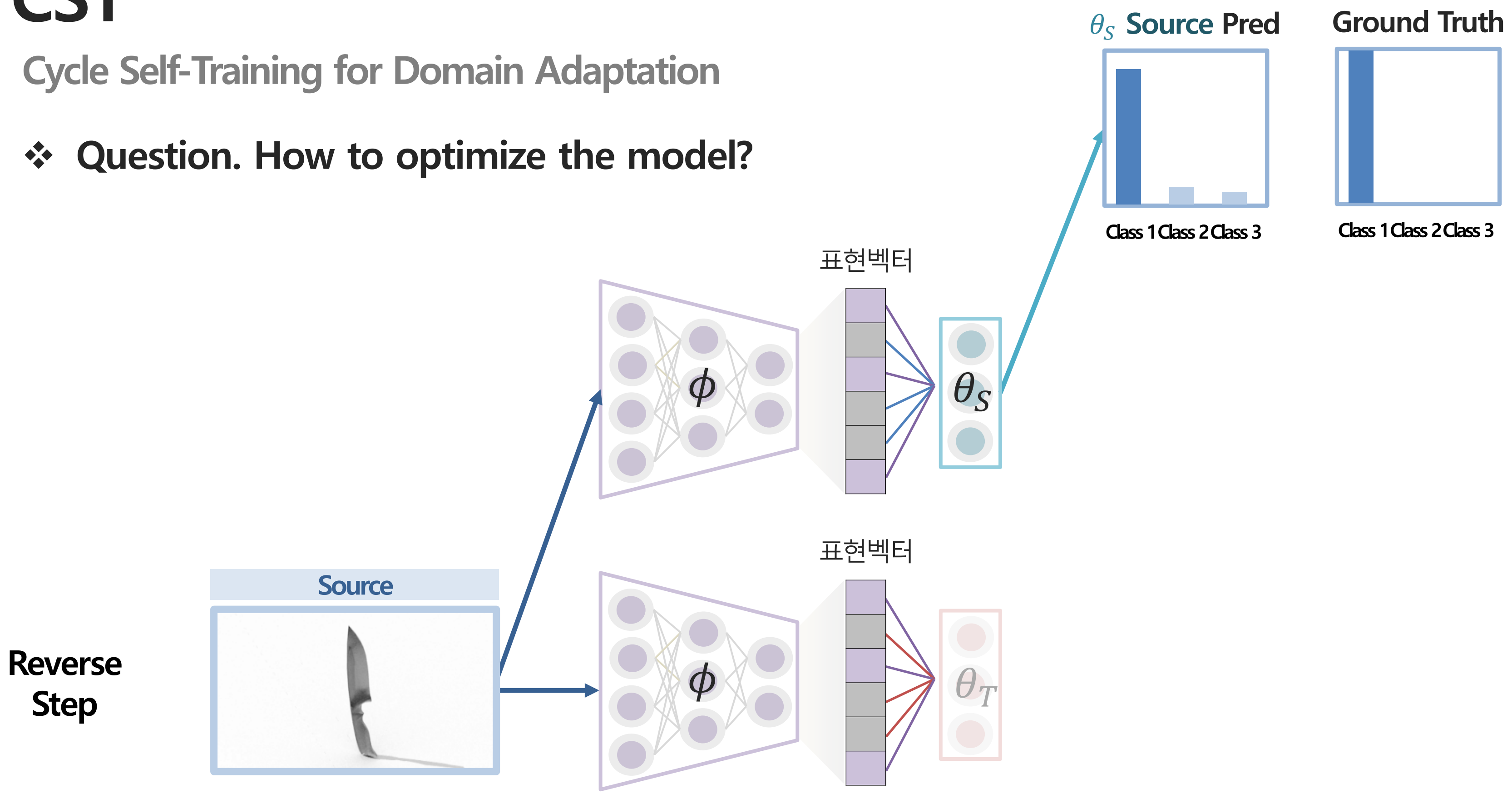
→ 우리가 source domain의 label을 기반으로 target에서도 유효한 정보를 가지고 오고 싶어하는 것처럼,
Now we can train the model to **MAKE TARGET PSUEDO-LABELS INOFRMATIVE of the SOURCE domain**!

→ So that we can **GRADUALLY REFINE noisy target pseudo-labels**!

CST

Cycle Self-Training for Domain Adaptation

❖ Question. How to optimize the model?



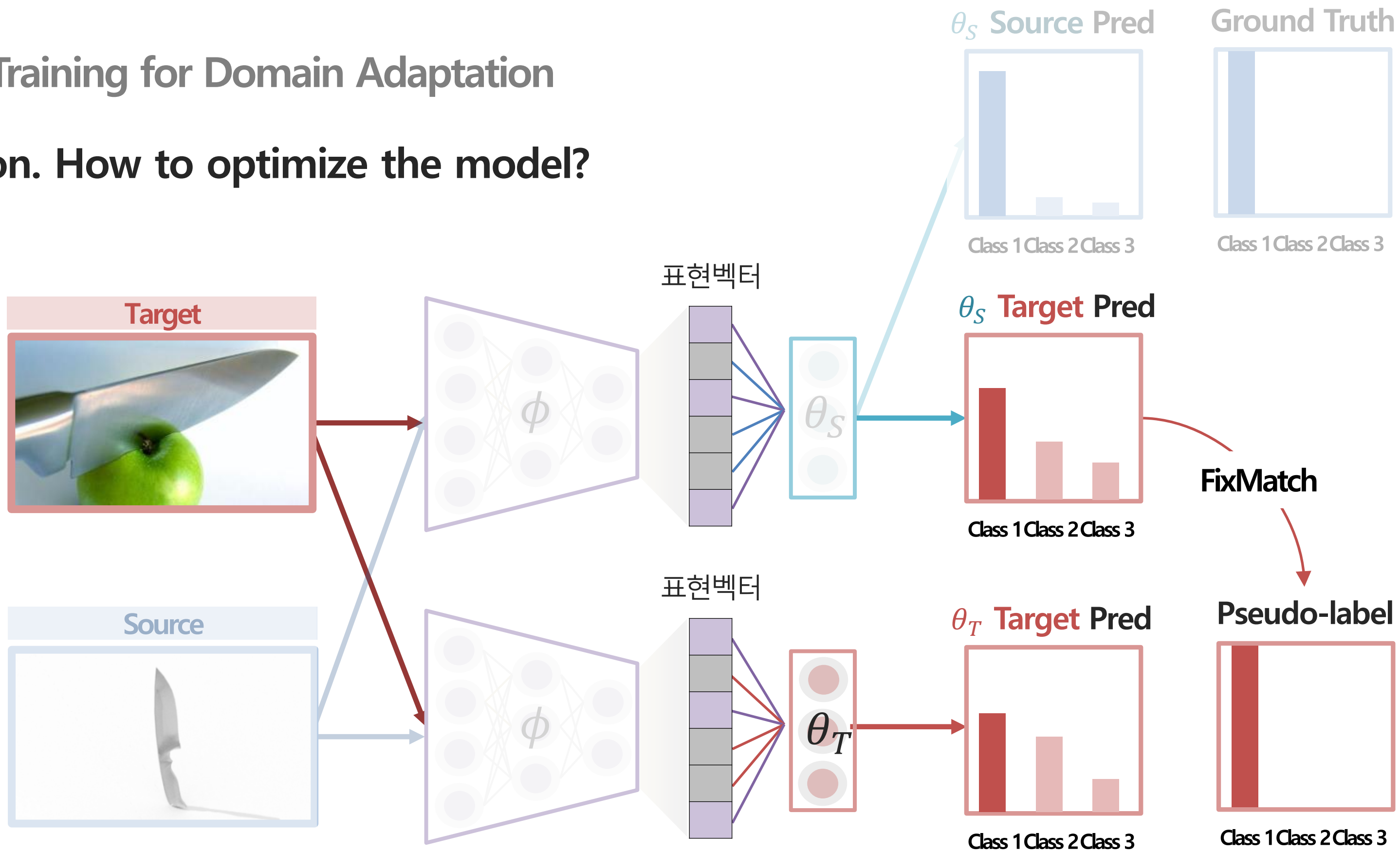
$L_{source}(\theta_S, \phi)$

Source domain 지도 학습

CST

Cycle Self-Training for Domain Adaptation

❖ Question. How to optimize the model?



$L_{source}(\theta_S, \phi)$

Source domain 지도 학습

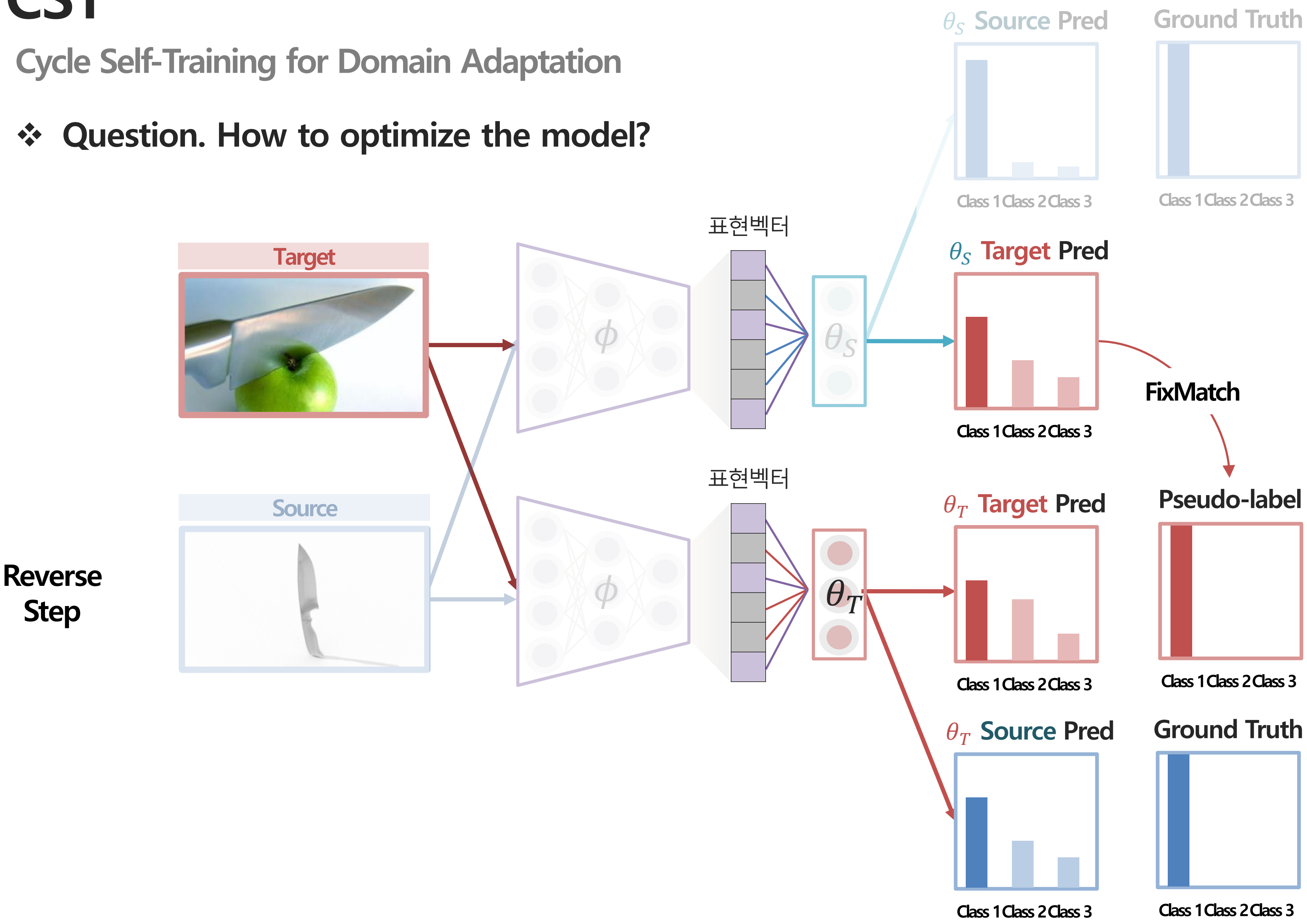
Reverse Step

$$\hat{\theta}_T(\phi) = \underset{\theta}{\operatorname{argmin}}\, \mathbb{E}_{x \sim T}\, \ell(f_{\theta,\phi}(x), y')$$

CST

Cycle Self-Training for Domain Adaptation

❖ **Question. How to optimize the model?**



$L_{source}(\theta_S, \phi)$

Source domain 지도 학습

Bi-level Optimization!

$$\hat{\theta}_T(\phi) = \underset{\theta}{\operatorname{argmin}}\, \mathbb{E}_{x\sim T}\ell(f_{\theta,\phi}(x), y')$$

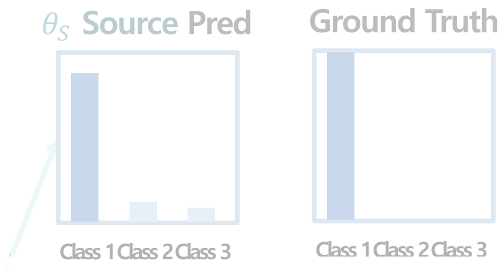
$$L_{Target}(\hat{\theta}_T(\phi), \phi)$$

Reverse self-training

CST

Cycle Self-Training for Domain Adaptation

❖ **Question. How to optimize the model?**



$L_{source}(\theta_S, \phi)$

Source domain 지도 학습

"We propose to calculate the analytical form of target classifier and directly back-propagate to the feature extractor instead of calculating the second-order derivatives as in MAML"

- **Analytical solution**
  - target classifier 최적화 문제에 대해서, "해석적으로" (수학적 공식에 의해) 직접 해를 계산할 수 있음을 의미
  - 즉, 복잡한 수치적 방법을 이용한 최적화 대신 간단한 공식을 통해 해를 구함
  - **본 논문에서는 ridge regression의 analytical solution 이용**
- **Second-order derivatives**
  - Model Agnostic Meta-Learning (MAML)과 같은 메타 학습에서는 모델 파라미터가 어떻게 변화하는지 파악하기 위해 2차 미분을 필요로 함
  - 두 단계 간의 상호작용을 파악하기 위해서 더 높은 차원의 정보를 활용 (계산 복잡도 ↑)

Bi-level Optimization!

$$\hat{\theta}_T(\phi) = \underset{\theta}{\operatorname{argmin}}\ \mathbb{E}_{x \sim T}\,\ell(f_{\theta,\phi}(x), y')$$

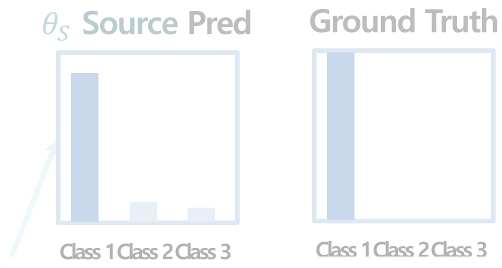
$$L_{Target}(\hat{\theta}_T(\phi), \phi)$$

Reverse self-training

CST

Cycle Self-Training for Domain Adaptation

❖ **Question. How to optimize the model?**



$L_{source}(\theta_S, \phi)$

Source domain 지도 학습

"We propose to calculate the analytical form of target classifier and directly back-propagate to the feature extractor instead of calculating the second-order derivatives as in MAML"

- **Analytical solution**
  - 본 논문에서는 ridge regression의 analytical solution 이용

Ridge

- $\min_{\theta} ||X\theta - y||^2 + \lambda||\theta||^2,$

  where X: input features, y: label, θ:weights, λ: regularization parameters
- $\theta = (X^T X + \lambda I)^{-1} X^T y$

One-hot encoded pseudo-labels

CST

- $\min_{\theta} \mathbb{E}_{x\sim T} ||\theta^T \phi(x) - y'||^2 + \lambda||\theta||^2,$

  where $\phi(x)$: X in ridge regression, y': y in ridge regression, θ: target classifier weights
- $\theta_T = (\Phi^T \Phi + \lambda I)^{-1} \Phi^T Y,$

  where Φ: feature matrix $[\Phi(x_1), \ldots, \Phi(x_n)]^T$, Y: pseudo-label matrix

Cross-entropy loss를 사용하는
일반적인 classification과 다른 접근

Bi-level Optimization!

$\hat{\theta}_T(\phi)$
$= \underset{\theta}{\operatorname{argmin}}\, \mathbb{E}_{x\sim T} \ell(f_{\theta,\phi}(x), y')$

$L_{Target}(\hat{\theta}_T(\phi), \phi)$

Reverse self-training

CST

Cycle Self-Training for Domain Adaptation

❖ Summary

- Complete objective:

Forward Step
Generate pseudo-labels on the target domain with ϕ and θ_s: $y' = \arg\max_i \{f_{\theta_s,\phi}(x)_{[i]}\}$.
Reverse Step
Train a target head $\hat{\theta}_t(\phi)$ with target pseudo-labels y' on the feature extractor ϕ:

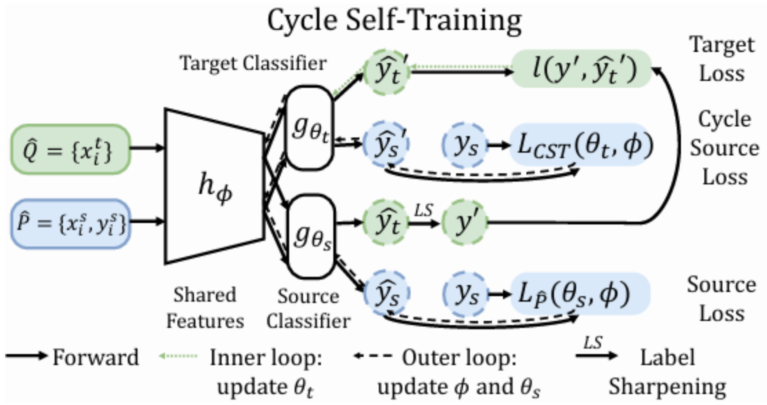
$$\hat{\theta}_t(\phi) = \arg\min_{\theta} \mathbb{E}_{x\sim\hat{Q}} \ell(f_{\theta,\phi}(x), y').$$

Update the feature extractor ϕ and the source head θ_s to make $\hat{\theta}_t(\phi)$ perform well on the source dataset and minimize the $\hat{\alpha}$-Tsallis entropy on the target dataset:

$$\phi \leftarrow \phi - \eta\nabla_\phi[L_{\hat{P}}(\theta_s,\phi) + L_{\hat{P}}(\hat{\theta}_t(\phi),\phi) + L_{\hat{Q},\mathrm{Tsallis},\hat{\alpha}}(\theta_s,\phi)]. \quad (7)$$

$$\theta_s \leftarrow \theta_s - \eta\nabla_{\theta_s}[L_{\hat{P}}(\theta_s,\phi) + L_{\hat{Q},\mathrm{Tsallis},\hat{\alpha}}(\theta_s,\phi)]. \quad (8)$$

- To refine noisy pseudo-labels (Reverse step):
  - Introduce 'cycle self-training'; train θ_T with target pseudo-labels, and make θ_T perform well on the source domain by updating the shared representations



$f_{\theta,\phi}(x_T)$

$$S_\alpha(y) = \frac{1}{\alpha - 1}\left(1 - \sum y_{[i]}^\alpha\right)$$

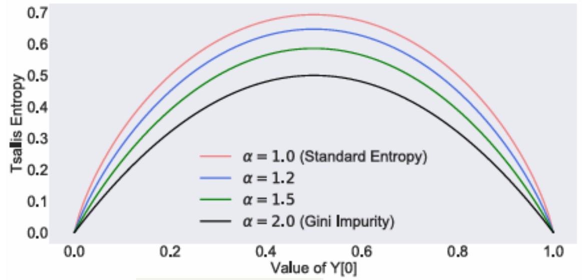


Figure 3: Tsallis entropy vs. entropic-index α.

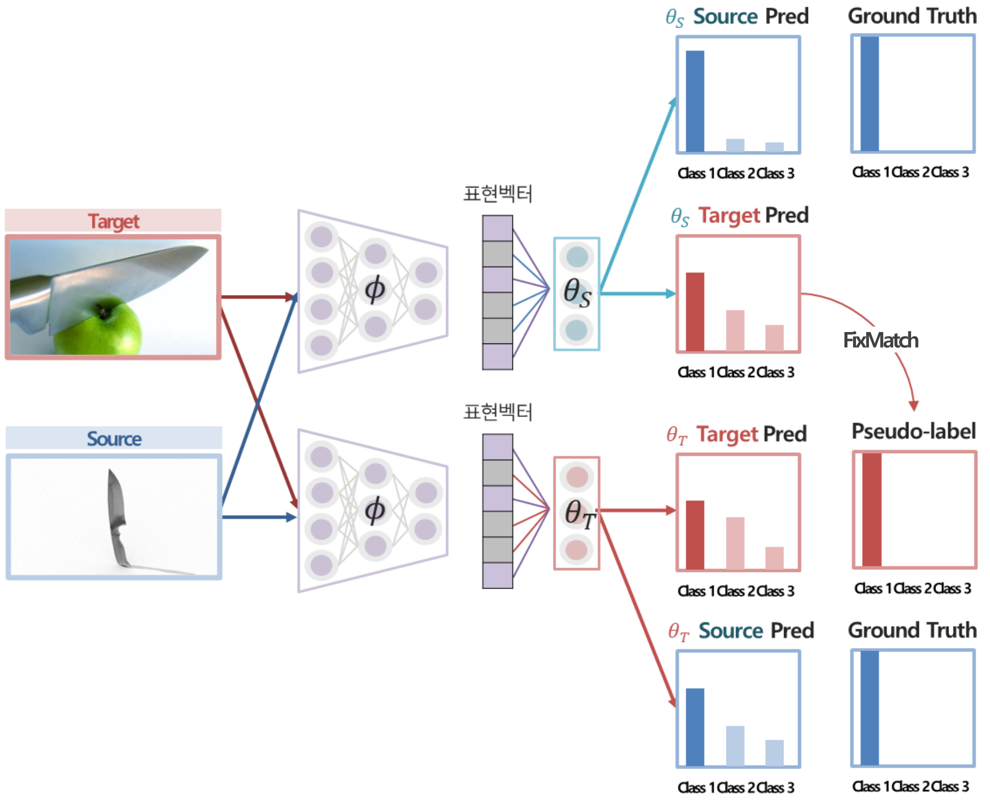
- α가 1에 가까워질 수록 Gibbs entropy로 수렴
- Gibbs는 overconfidence 문제를 강화
  (불확실성을 강하게 낮추어 예측을 확실하게 만듦)
- α에 따라 **엔트로피 민감도 조절** (flexible)

CST

Cycle Self-Training for Domain Adaptation

❖ CST 한계

- Target domain의 고유한 특성에 대한 고려 부족으로 잘못된 pseudo-labels 산출 가능성 ↑
  - θ_T 는 이미 source-specific features를 이용하여 산출된 pseudo-labels로 학습됨
    이를 다시 source data에 적합하게 학습하는 방식 (reverse step)은 오히려 악순환을 강화하는 요인으로 기능할 염려



Source classifier로부터 생성

$$\hat{\theta}_T(\phi) = \underset{\theta}{\operatorname{argmin}}\ \mathbb{E}_{x\sim T}\ell(f_{\theta,\phi}(x), y')$$

Source data 예측

$$L_{Target}(\hat{\theta}_T(\phi), \phi)$$

CST

Cycle Self-Training for Domain Adaptation

❖ CST 한계

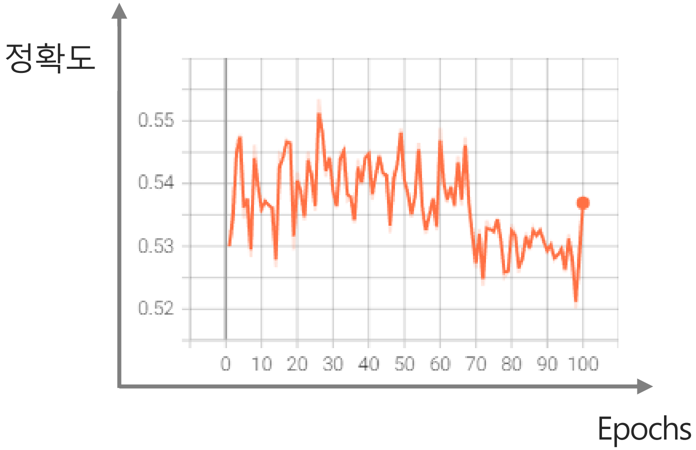
- Target domain의 고유한 특성에 대한 고려 부족으로 잘못된 pseudo-labels 산출 가능성 ↑
  - θ_T 는 이미 source-specific features를 이용하여 산출된 pseudo-labels로 학습됨
    이를 다시 source data에 적합하게 학습하는 방식 (reverse step)은 오히려 악순환을 강화하는 요인으로 기능할 염려

Self-training

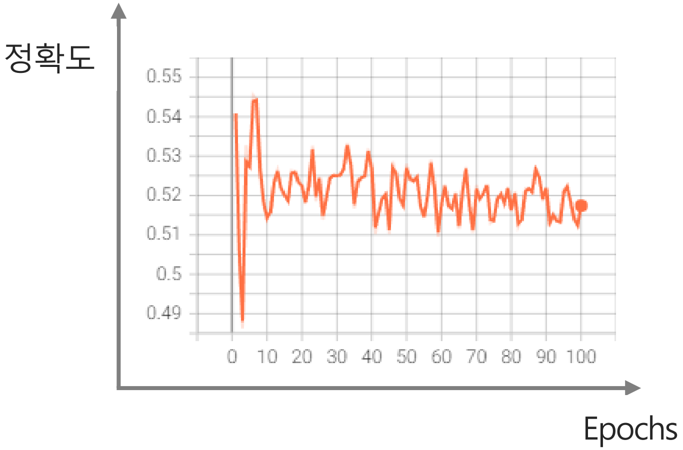
CST:
Target Domain 평균 정확도

Adversarial-training

DANN:
Target Domain 평균 정확도



CDAN:
Target Domain 평균 정확도



CST

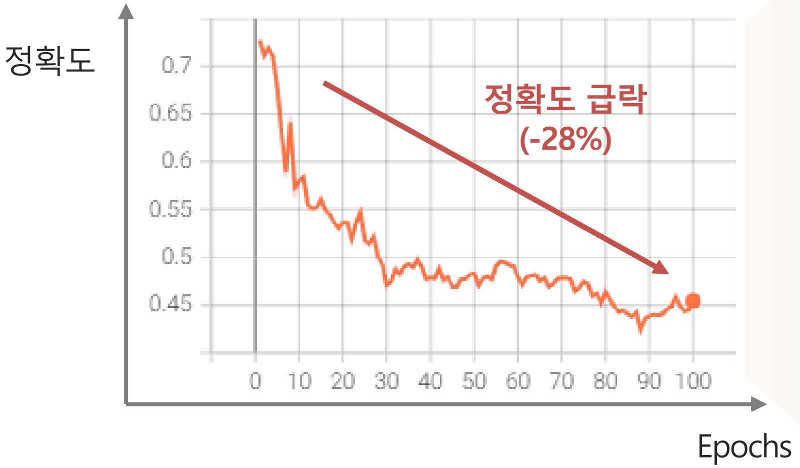
Cycle Self-Training for Domain Adaptation

❖ CST 한계

- Target domain의 고유한 특성에 대한 고려 부족으로 잘못된 pseudo-labels 산출 가능성 ↑
  - θ_T 는 이미 source-specific features를 이용하여 산출된 pseudo-labels로 학습됨
    이를 다시 source data에 적합하게 학습하는 방식 (reverse step)은 오히려 악순환을 강화하는 요인으로 기능할 염려

CST: pseudo-labels 품질 확인

CST:
Target Domain 평균 정확도



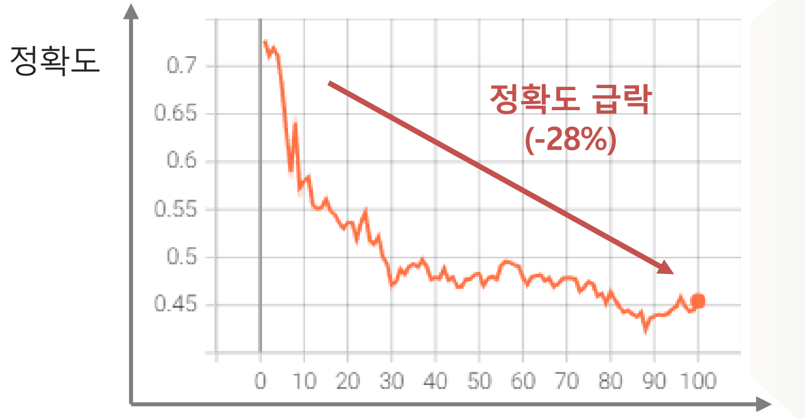
CST

Cycle Self-Training for Domain Adaptation

❖ CST 한계

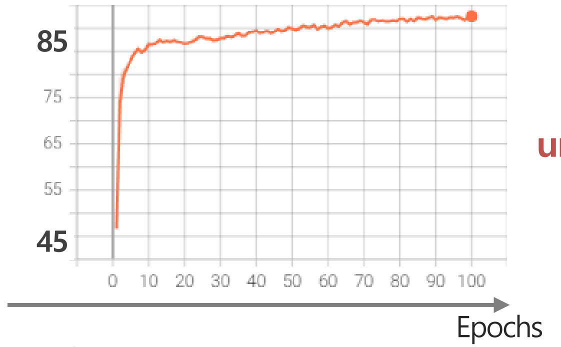
- Target domain의 고유한 특성에 대한 고려 부족으로 잘못된 pseudo-labels 산출 가능성 ↑
  - θ_T 는 이미 source-specific features를 이용하여 산출된 pseudo-labels로 학습됨
    이를 다시 source data에 적합하게 학습하는 방식 (reverse step)은 오히려 악순환을 강화하는 요인으로 기능할 염려

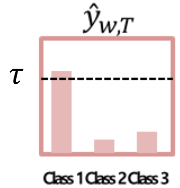
CST: Target Domain 평균 정확도



CST: pseudo-labels 품질 확인

Pseudo-labels samples 비율 (%)





(1) 높은 확신을 가지고 예측되는 target samples이 얼마나 될까?

학습이 진행될 수록 거의 모든 unlabeled target을 높은 확신을 가지고 예측

= 활용되는 pseudo-labels 개수 ↑

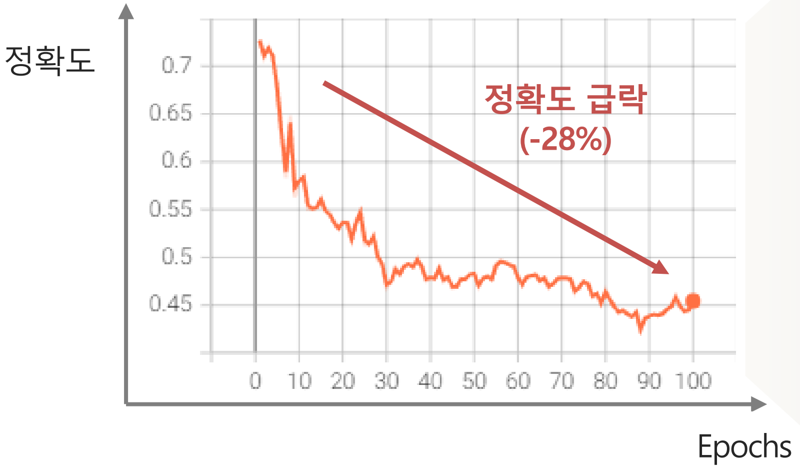
CST

Cycle Self-Training for Domain Adaptation

❖ CST 한계

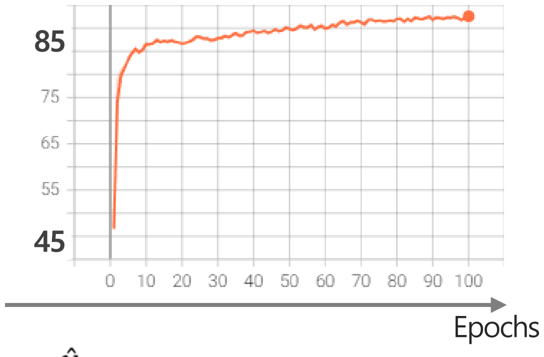
- Target domain의 고유한 특성에 대한 고려 부족으로 잘못된 pseudo-labels 산출 가능성 ↑
  - θ_T 는 이미 source-specific features를 이용하여 산출된 pseudo-labels로 학습됨
    이를 다시 source data에 적합하게 학습하는 방식 (reverse step)은 오히려 악순환을 강화하는 요인으로 기능할 염려

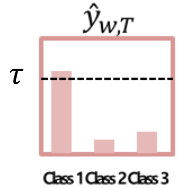
CST: Target Domain 평균 정확도



CST: pseudo-labels 품질 확인

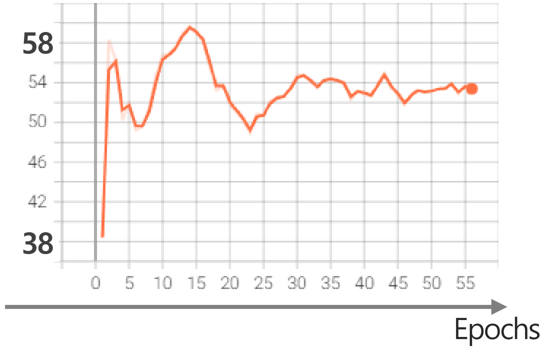
Pseudo-labels samples 비율 (%)





(1) 높은 확신을 가지고 예측되는 target samples이 얼마나 될까?

Pseudo-labels 정확도



(2) 얼마나 정확한가?

→ **잘못된 pseudo-labels을 많이 만들어내는 문제**

CST

Cycle Self-Training for Domain Adaptation

❖ CST 한계

- Target domain의 고유한 특성에 대한 고려 부족으로 잘못된 pseudo-labels 산출 가능성 ↑
  - θ_T 는 이미 source-specific features를 이용하여 산출된 pseudo-labels로 학습됨
    이를 다시 source data에 적합하게 학습하는 방식 (reverse step)은 오히려 악순환을 강화하는 요인으로 기능할 염려

CST:
Target Domain 평균 정확도

CST: components 효과 확인

CST = FixMatch + Reverse Step + Tsallis Entropy

Table 5: Ablation on VisDA-2017.

	Method	Accuracy ↑	d_{TV} ↓
FixMatch + Reverse + Gibbs	FixMatch [57]	74.5 ± 0.2	0.22
FixMatch + Reverse + Tsallis	Fixmatch+Tsallis	76.3 ± 0.8	0.15
FixMatch + Reverse + Gibbs	CST w/o Tsallis	72.0 ± 0.4	0.16
FixMatch + Reverse + Gibbs	CST+Entropy	76.2 ± 0.6	0.20
FixMatch + Reverse + Tsallis	**CST**	$\mathbf{79.9} \pm 0.5$	0.12

UDA with Self-Training

[2023 NeurIPS]
Make the U in UDA Matter: Invariant Consistency Learning for Unsupervised Domain Adaptation

Yue et al., Nanyang Technological University and Singapore Management University

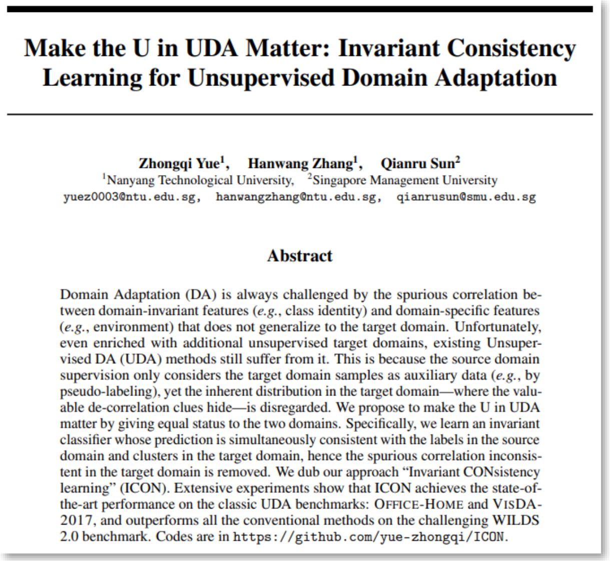
ICON

ICON: Invariant CONsistency learning

❖ Make the U in UDA Matter: Invariant Consistency Learning (ICON) for Unsupervised Domain Adaptation

- CST는 (1) transferable, (2) domain-specific 정보 중 transferable 정보에 집중할 수 있도록 학습 유도 (reverse step)
- ICON은 domain-specific 정보를 **직접적으로 제거**해야 정확한 학습이 가능함을 주장
- 학습이 진행됨에 따라 target accuracy가 하락하는 원인을 Source domain 내에 있는 'spurious correlations'로 지적

ICON (NeurIPS 2023)[3]

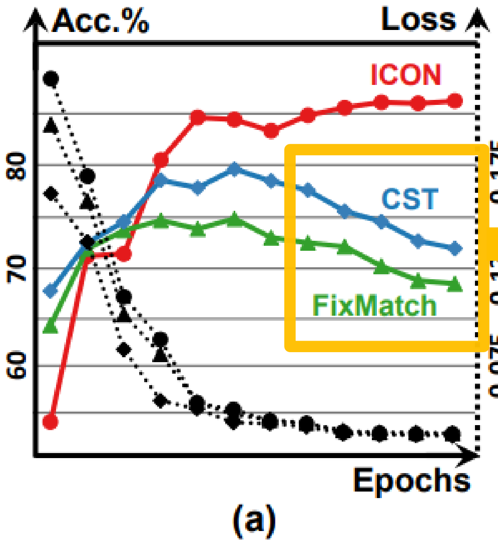


Make the U in UDA Matter: Invariant Consistency Learning for Unsupervised Domain Adaptation

Zhongqi Yue[1], **Hanwang Zhang**[1], **Qianru Sun**[2]
[1]Nanyang Technological University, [2]Singapore Management University
yuez0003@ntu.edu.sg, hanwangzhang@ntu.edu.sg, qianrusun@smu.edu.sg

Abstract

Domain Adaptation (DA) is always challenged by the spurious correlation between domain-invariant features (*e.g.*, class identity) and domain-specific features (*e.g.*, environment) that does not generalize to the target domain. Unfortunately, even enriched with additional unsupervised target domains, existing Unsupervised DA (UDA) methods still suffer from it. This is because the source domain supervision only considers the target domain samples as auxiliary data (*e.g.*, by pseudo-labeling), yet the inherent distribution in the target domain—where the valuable de-correlation clues hide—is disregarded. We propose to make the U in UDA matter by giving equal status to the two domains. Specifically, we learn an invariant classifier whose prediction is simultaneously consistent with the labels in the source domain and clusters in the target domain, hence the spurious correlation inconsistent in the target domain is removed. We dub our approach "Invariant CONsistency learning" (ICON). Extensive experiments show that ICON achieves the state-of-the-art performance on the classic UDA benchmarks: OFFICE-HOME and VISDA-2017, and outperforms all the conventional methods on the challenging WILDS 2.0 benchmark. Codes are in https://github.com/yue-zhongqi/ICON.



"CST와 FixMatch는 모두 학습이 진행됨에 따라 target domain의 정확도가 떨어지는 문제가 있다!"

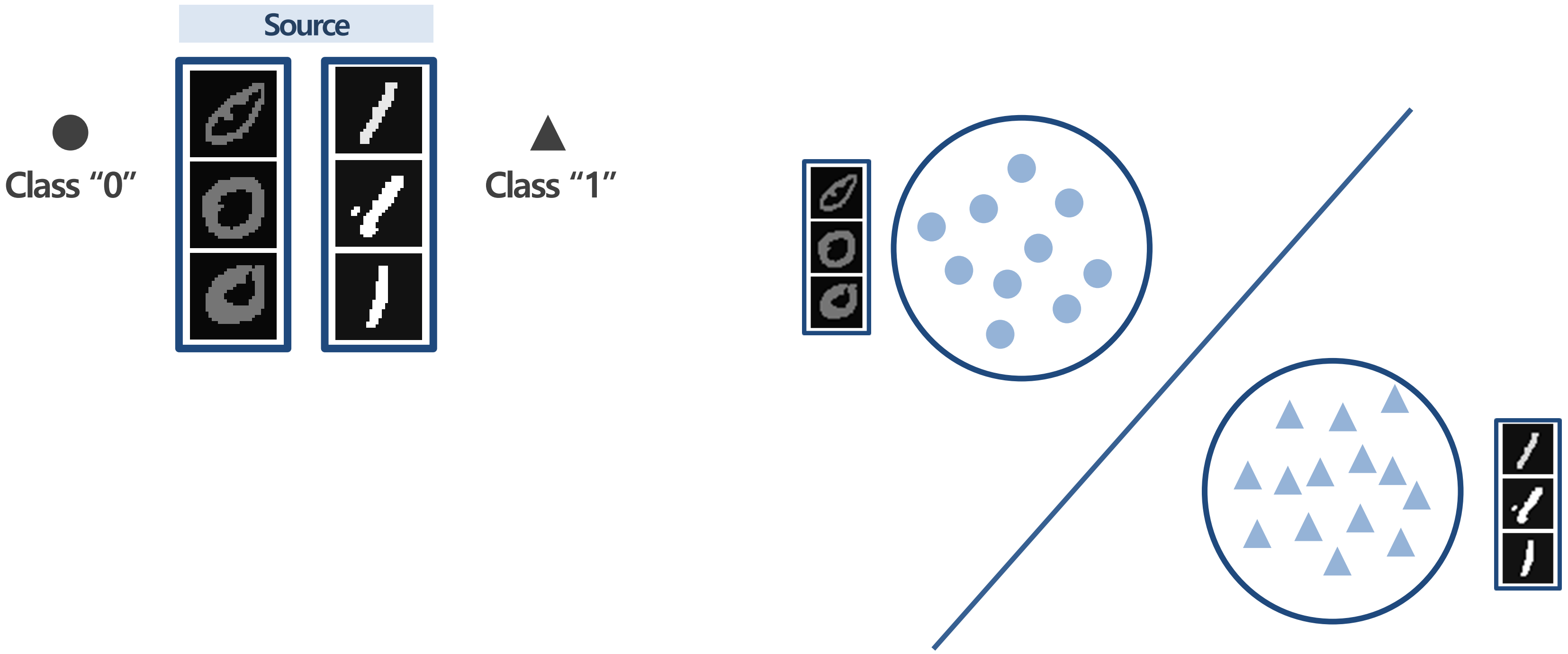
→ Source domain spurious correlations 때문!

ICON

ICON: Invariant CONsistency learning

❖ Source domain “spurious correlations”

- Label과 관련이 있는 것처럼 보이지만 실제로는 예측에 중요한 역할을 하지 않는 입력 데이터 특성[4]
  ↳ Label과의 “가짜 상관관계” (예: 어둡다/밝다)
- 오로지 source domain에만 있는 가짜 상관관계에 모델이 편향될 경우, target domain 예측을 부정확하게 만드는 요인이 됨



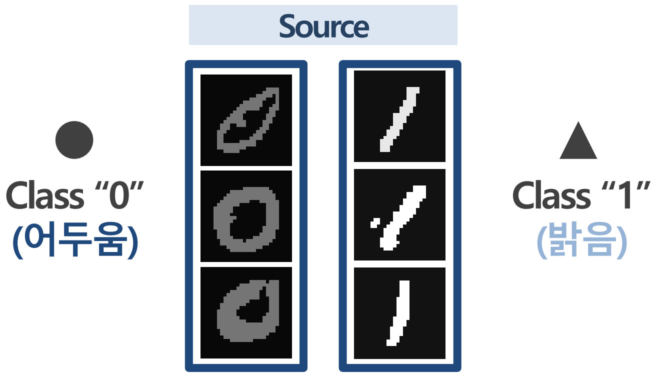
[4] Menon, A. K., Rawat, A. S., & Kumar, S. (2020). Overparameterisation and worst-case generalisation: friend or foe?. In International Conference on Learning Representations.

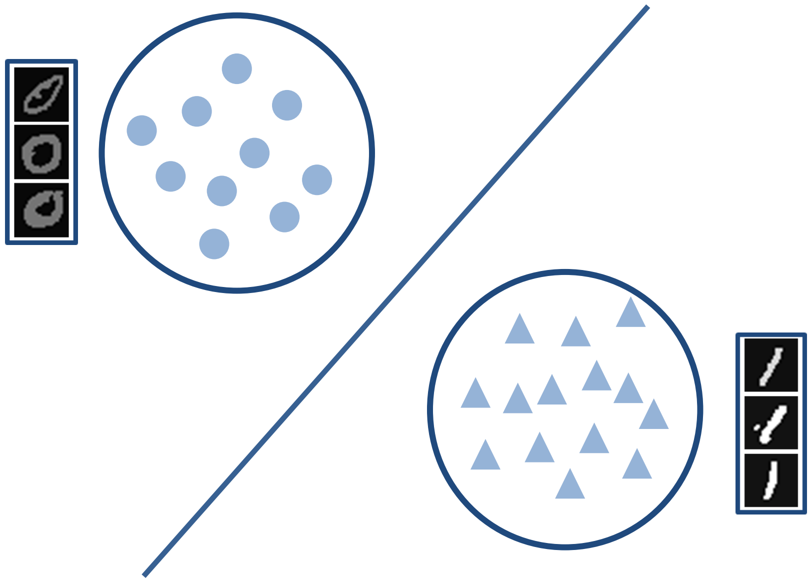
ICON

ICON: Invariant CONsistency learning

❖ **Source domain "spurious correlations"**

- Label과 관련이 있는 것처럼 보이지만 실제로는 예측에 중요한 역할을 하지 않는 입력 데이터 특성[4]
  ↳ Label과의 "가짜 상관관계" (예: 어둡다/밝다)
- 오로지 source domain에만 있는 가짜 상관관계에 모델이 편향될 경우, target domain 예측을 부정확하게 만드는 요인이 됨





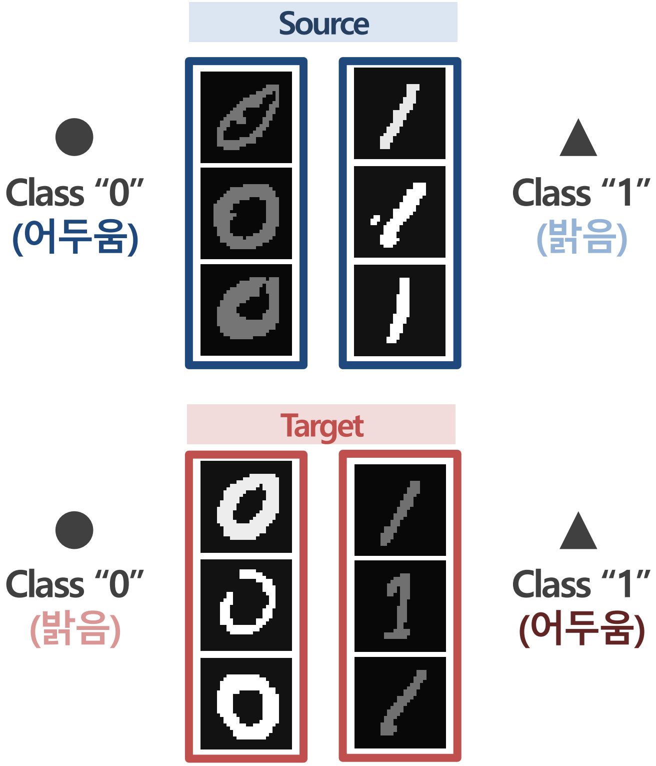
[4] Menon, A. K., Rawat, A. S., & Kumar, S. (2020). Overparameterisation and worst-case generalisation: friend or foe?. In International Conference on Learning Representations.

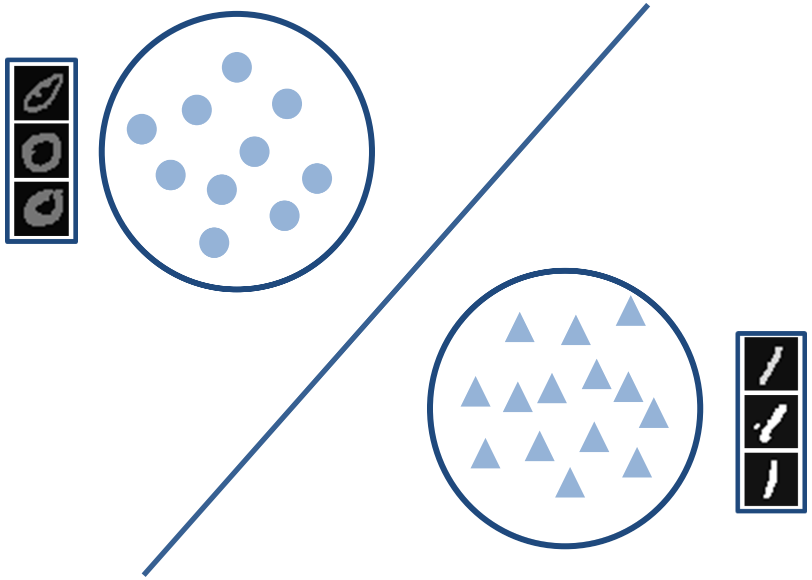
ICON

ICON: Invariant CONsistency learning

❖ Source domain "spurious correlations"

- Label과 관련이 있는 것처럼 보이지만 실제로는 예측에 중요한 역할을 하지 않는 입력 데이터 특성[4]
  ↳ Label과의 "가짜 상관관계" (예: 어둡다/밝다)
- 오로지 source domain에만 있는 가짜 상관관계에 모델이 편향될 경우, target domain 예측을 부정확하게 만드는 요인이 됨





[4] Menon, A. K., Rawat, A. S., & Kumar, S. (2020). Overparameterisation and worst-case generalisation: friend or foe?. In International Conference on Learning Representations.

ICON

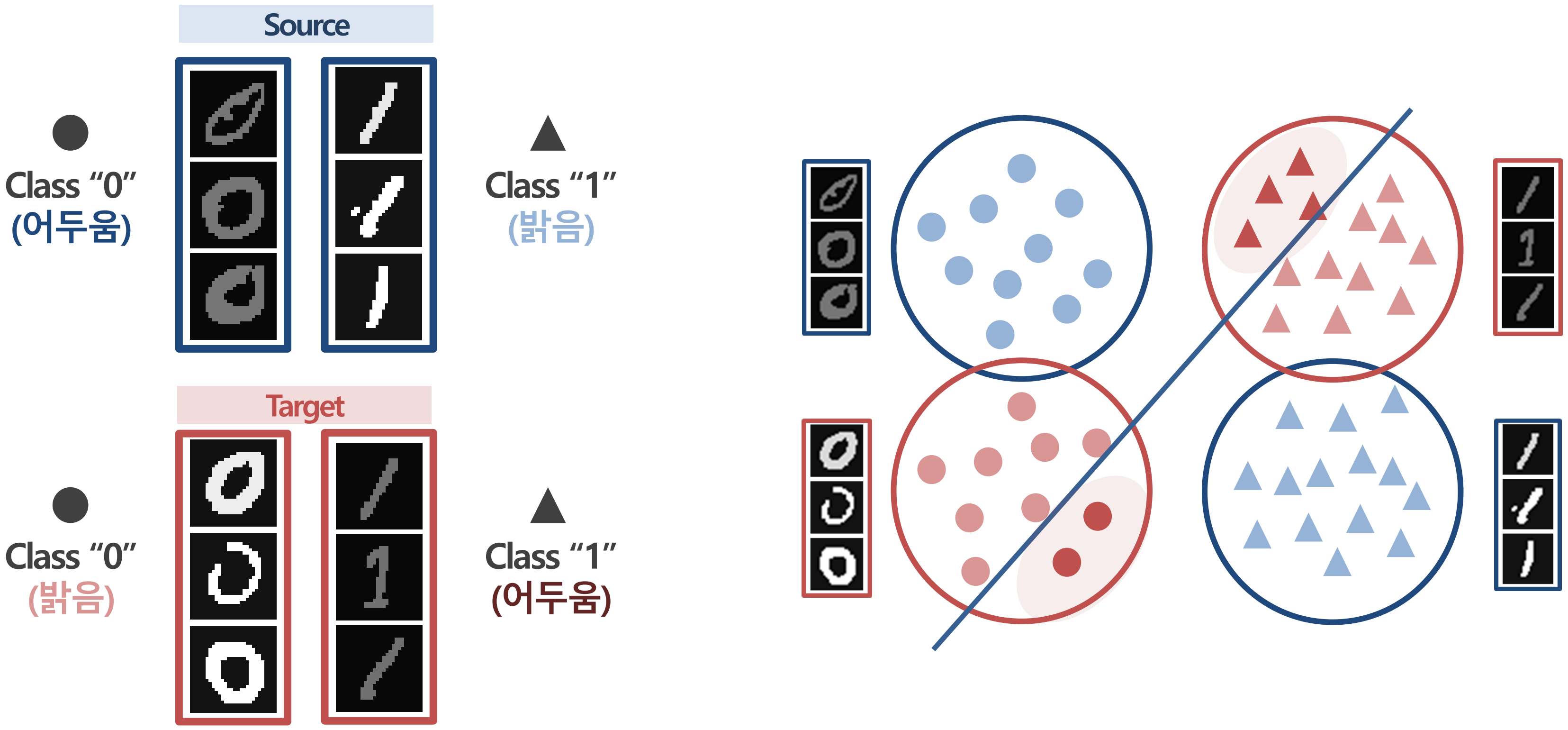
ICON: Invariant CONsistency learning

❖ Source domain "spurious correlations"

- Label과 관련이 있는 것처럼 보이지만 실제로는 예측에 중요한 역할을 하지 않는 입력 데이터 특성[4]

  ↳ Label과의 "가짜 상관관계" (예: 어둡다/밝다)

- 오로지 source domain에만 있는 가짜 상관관계에 모델이 편향될 경우, target domain 예측을 부정확하게 만드는 요인이 됨



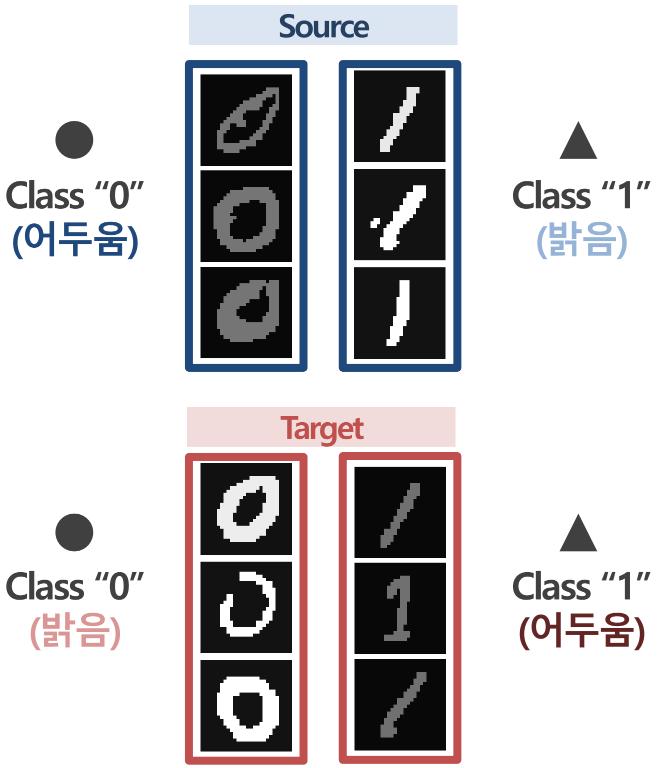
[4] Menon, A. K., Rawat, A. S., & Kumar, S. (2020). Overparameterisation and worst-case generalisation: friend or foe?. In International Conference on Learning Representations.

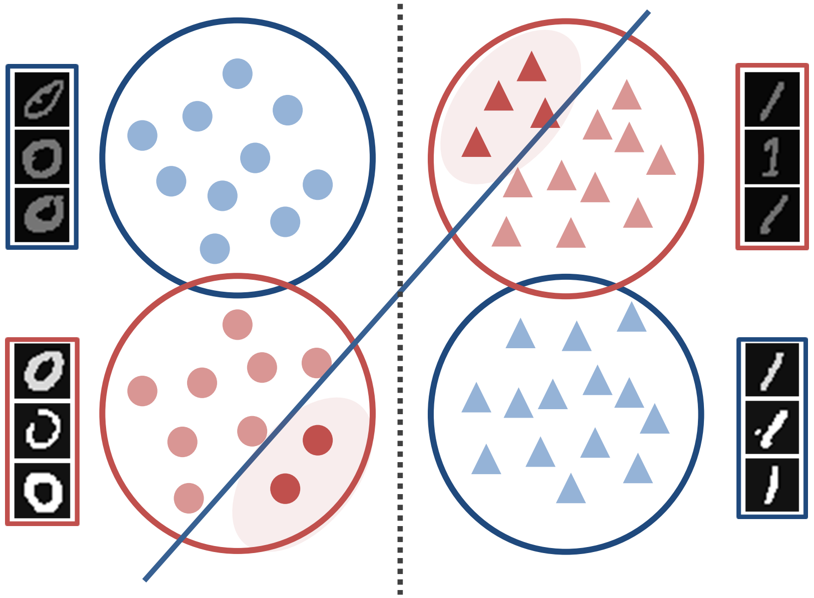
ICON

ICON: Invariant CONsistency learning

❖ **Source domain “spurious correlations”**

- Label과 관련이 있는 것처럼 보이지만 실제로는 예측에 중요한 역할을 하지 않는 입력 데이터 특성[4]
  ↳ Label과의 “가짜 상관관계” (예: 어둡다/밝다)
- 오로지 source domain에만 있는 가짜 상관관계에 모델이 편향될 경우, target domain 예측을 부정확하게 만드는 요인이 됨





이상적인 분류기

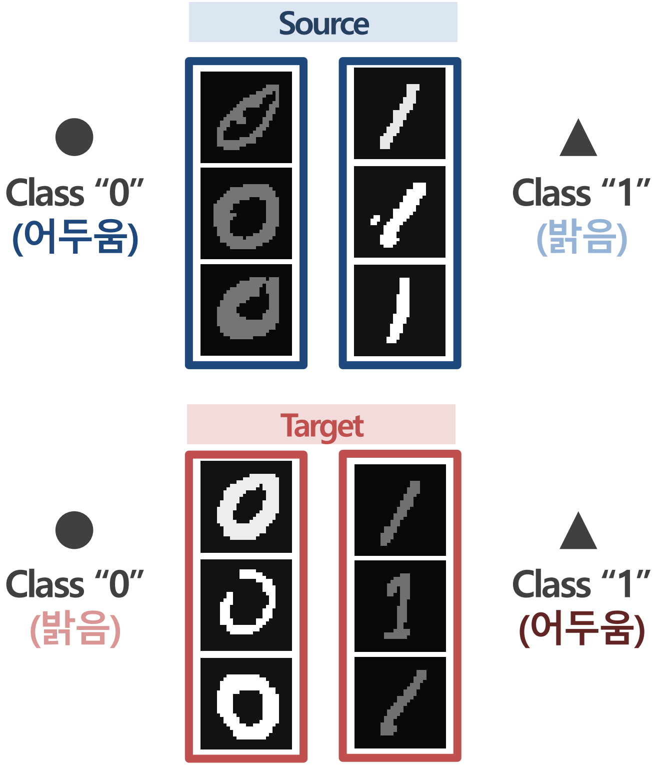
[4] Menon, A. K., Rawat, A. S., & Kumar, S. (2020). Overparameterisation and worst-case generalisation: friend or foe?. In International Conference on Learning Representations.

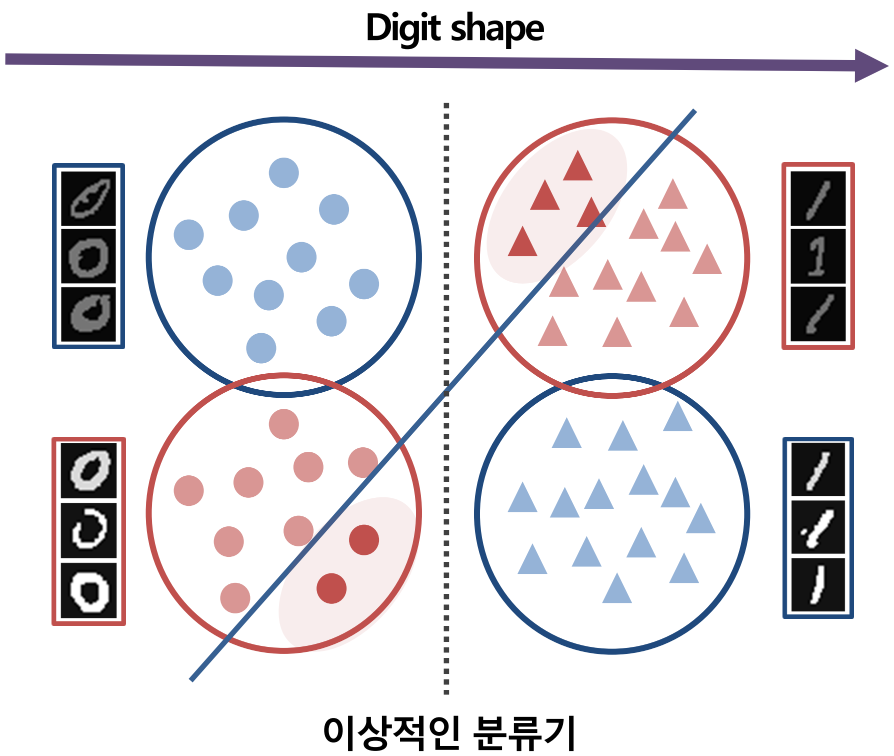
ICON

ICON: Invariant CONsistency learning

❖ **Source domain "spurious correlations"**

- Label과 관련이 있는 것처럼 보이지만 실제로는 예측에 중요한 역할을 하지 않는 입력 데이터 특성[4]
  ↳ Label과의 "가짜 상관관계" (예: 어둡다/밝다)
- 오로지 source domain에만 있는 가짜 상관관계에 모델이 편향될 경우, target domain 예측을 부정확하게 만드는 요인이 됨





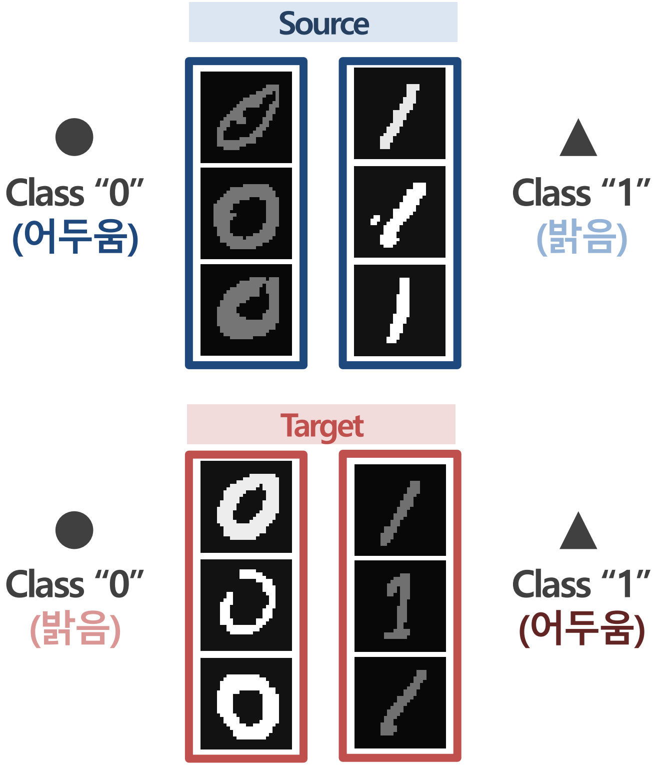
[4] Menon, A. K., Rawat, A. S., & Kumar, S. (2020). Overparameterisation and worst-case generalisation: friend or foe?. In International Conference on Learning Representations.

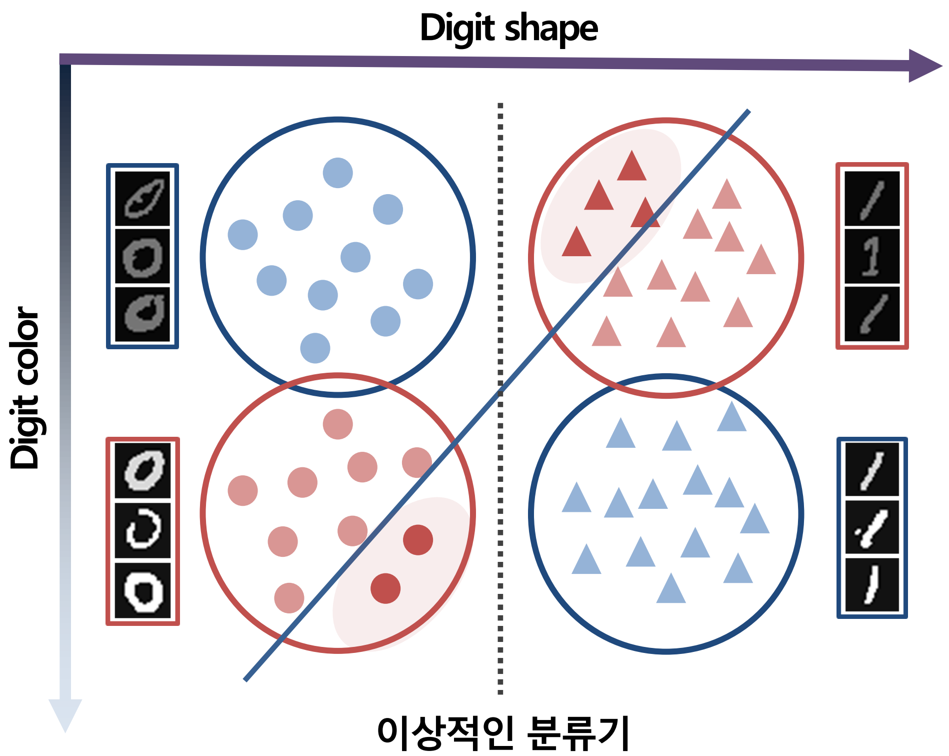
ICON

ICON: Invariant CONsistency learning

❖ Source domain "spurious correlations"

- Label과 관련이 있는 것처럼 보이지만 실제로는 예측에 중요한 역할을 하지 않는 입력 데이터 특성[4]
  ↳ Label과의 "가짜 상관관계" (예: 어둡다/밝다)
- 오로지 source domain에만 있는 가짜 상관관계에 모델이 편향될 경우, target domain 예측을 부정확하게 만드는 요인이 됨





우리의 모델은
숫자의 색깔 (어두움/밝음)에
따라 class를 구분!

[4] Menon, A. K., Rawat, A. S., & Kumar, S. (2020). Overparameterisation and worst-case generalisation: friend or foe?. In International Conference on Learning Representations.

ICON

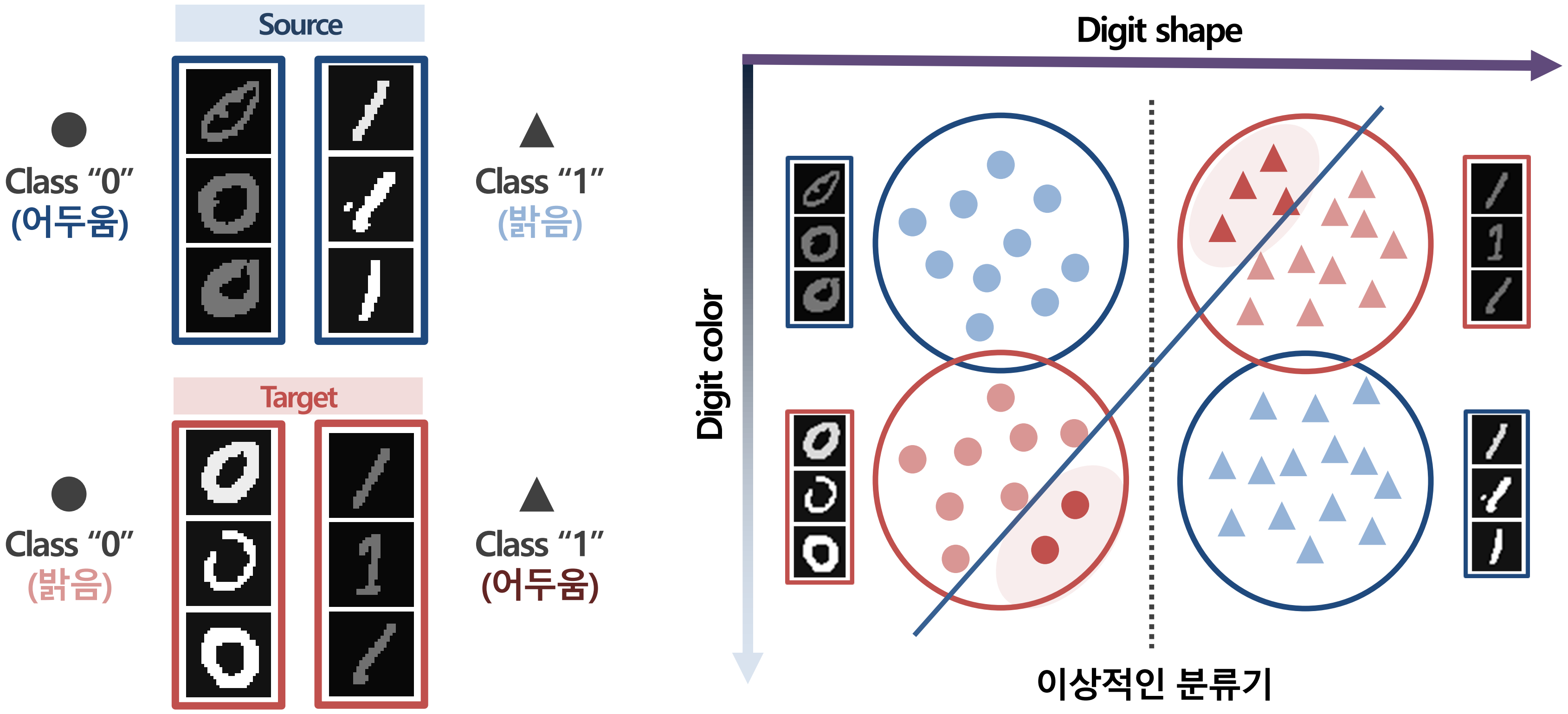
ICON: Invariant CONsistency learning

❖ Source domain “spurious correlations”

- Label과 관련이 있는 것처럼 보이지만 실제로는 예측에 중요한 역할을 하지 않는 입력 데이터 특성[4]

  ↳ Label과의 “가짜 상관관계” (예: 어둡다/밝다)

- 오로지 source domain에만 있는 가짜 상관관계에 모델이 편향될 경우, target domain 예측을 부정확하게 만드는 요인이 됨



우리의 모델은
숫자의 색깔 (어두움/밝음)에
따라 class를 구분!

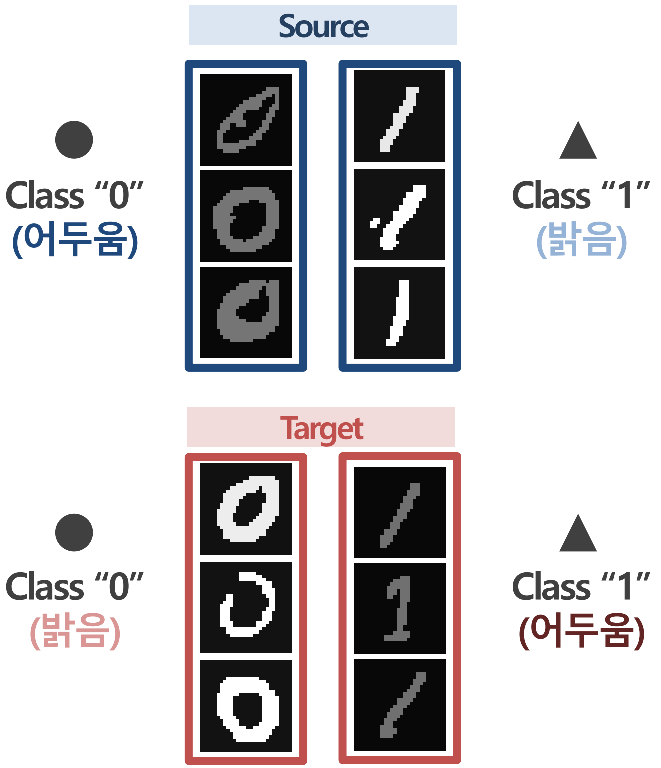
[4] Menon, A. K., Rawat, A. S., & Kumar, S. (2020). Overparameterisation and worst-case generalisation: friend or foe?. In International Conference on Learning Representations.

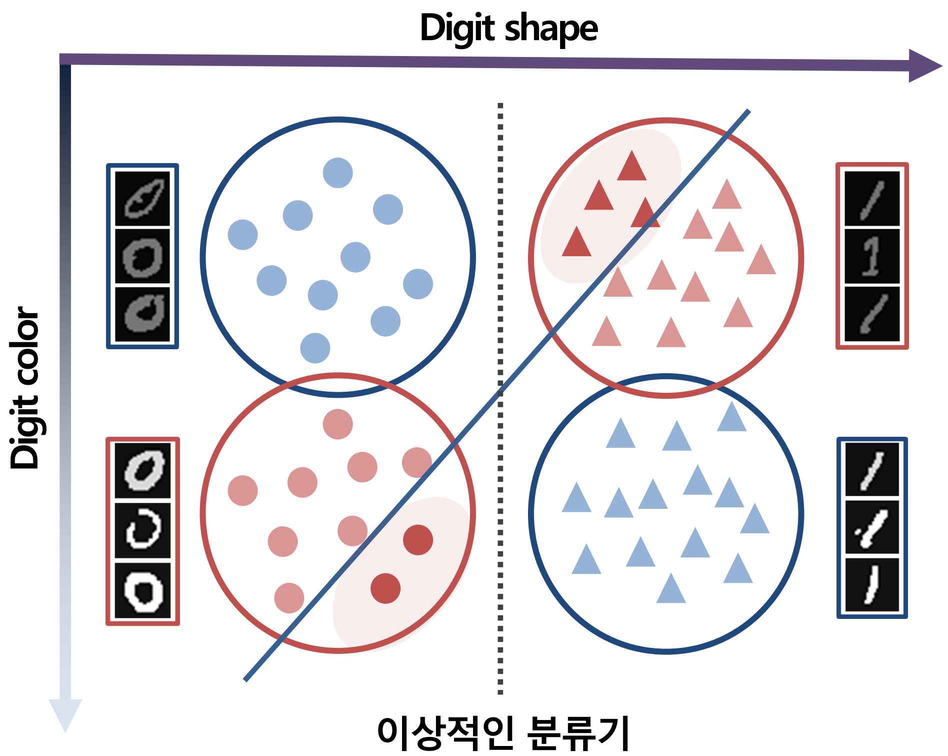
ICON

ICON: Invariant CONsistency learning

❖ Source domain “spurious correlations”

- Label과 관련이 있는 것처럼 보이지만 실제로는 예측에 중요한 역할을 하지 않는 입력 데이터 특성[4]
  ↳ Label과의 “가짜 상관관계” (예: 어둡다/밝다)
- 오로지 **source domain에만 있는 가짜 상관관계에 모델이 편향**될 경우, **target domain 예측을 부정확하게 만드는 요인**이 됨





우리의 모델은
숫자의 색깔 (어두움/밝음)에
따라 class를 구분!

[4] Menon, A. K., Rawat, A. S., & Kumar, S. (2020). Overparameterisation and worst-case generalisation: friend or foe?. In International Conference on Learning Representations.

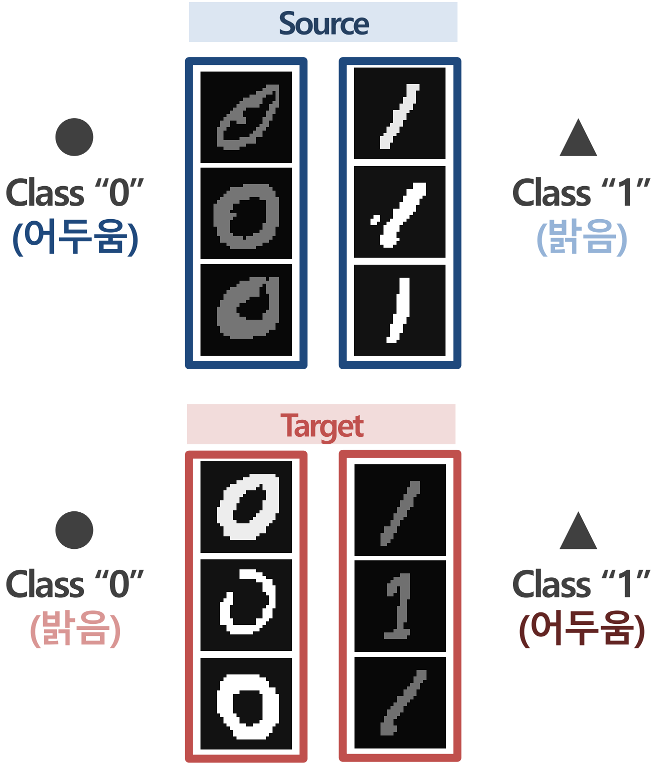
ICON

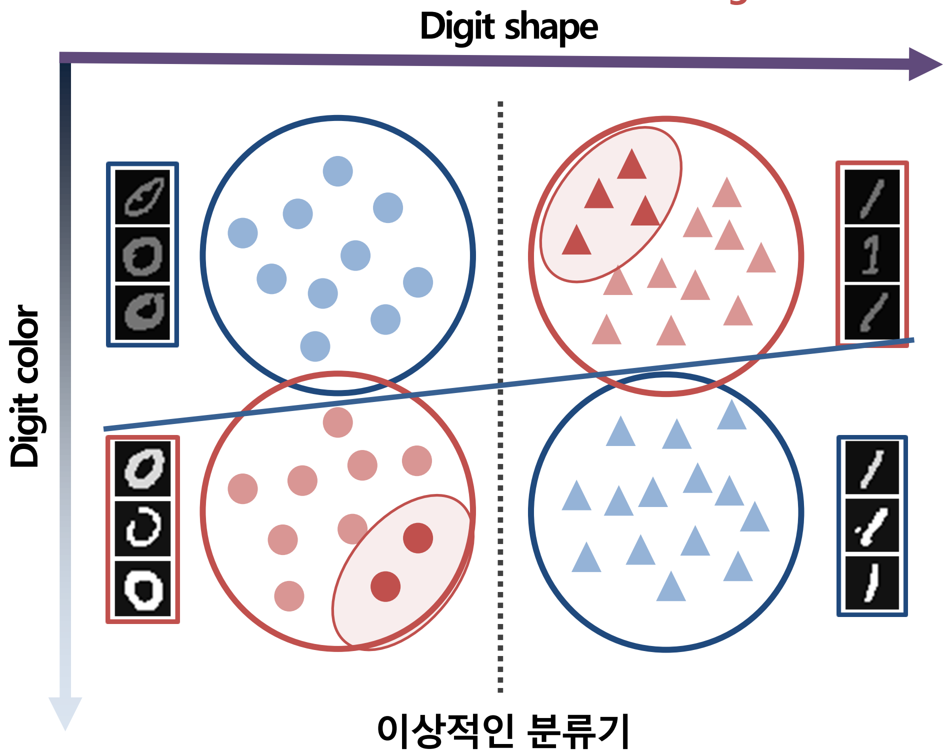
ICON: Invariant CONsistency learning

❖ **Source domain "spurious correlations"**

- Label과 관련이 있는 것처럼 보이지만 실제로는 예측에 중요한 역할을 하지 않는 입력 데이터 특성[4]
  ↳ Label과의 "가짜 상관관계" (예: 어둡다/밝다)
- 오로지 **source domain에만 있는 가짜 상관관계에 모델이 편향**될 경우, **target domain 예측을 부정확하게 만드는 요인**이 됨

Wrong but confident!





우리의 모델은
숫자의 색깔 (어두움/밝음)에
따라 class를 구분!

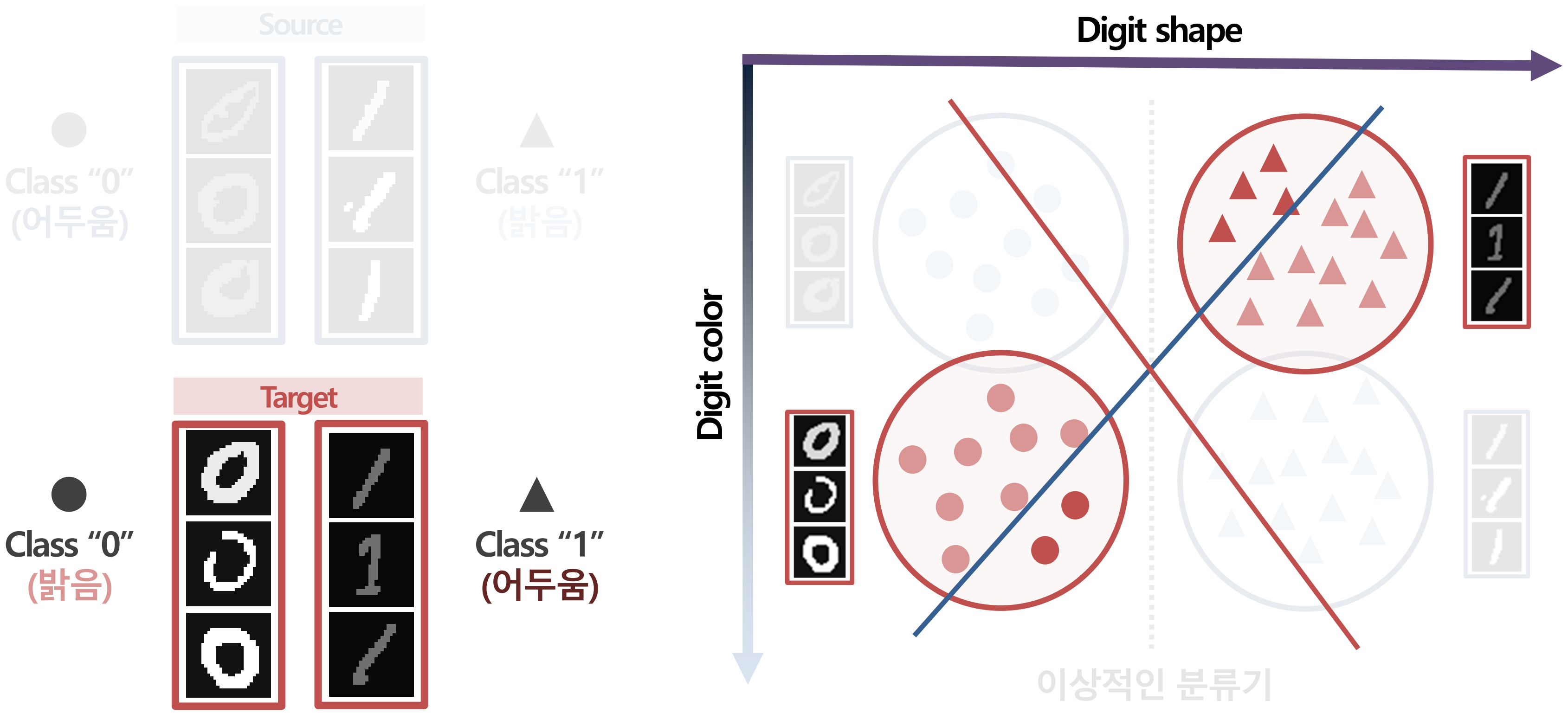
[4] Menon, A. K., Rawat, A. S., & Kumar, S. (2020). Overparameterisation and worst-case generalisation: friend or foe?. In International Conference on Learning Representations.

ICON

ICON: Invariant CONsistency learning

❖ Considering inherent distribution in target domain

- 모델은 <u>labeled</u> source domain 데이터 구조에 더 집중하여 spurious features에 과적합 되고 target domain 성능이 저하될 수 있음
- Target domain 내의 내재적 특성을 파악하여, **모델로 하여금 target domain 구조를 더 잘 반영할 수 있게끔 학습** 필요



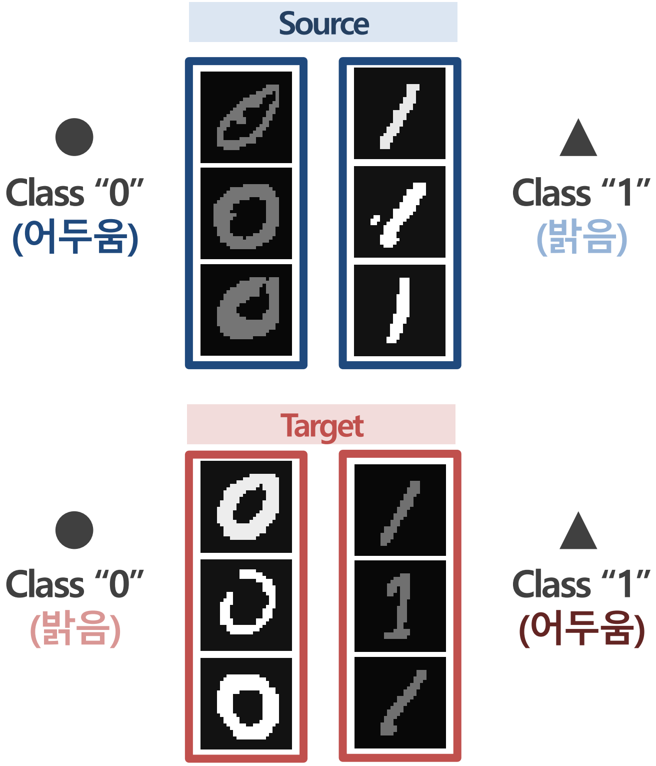
Cluster Assumption
Even though target domain is unlabeled, we can still identify the sample clusters in target.

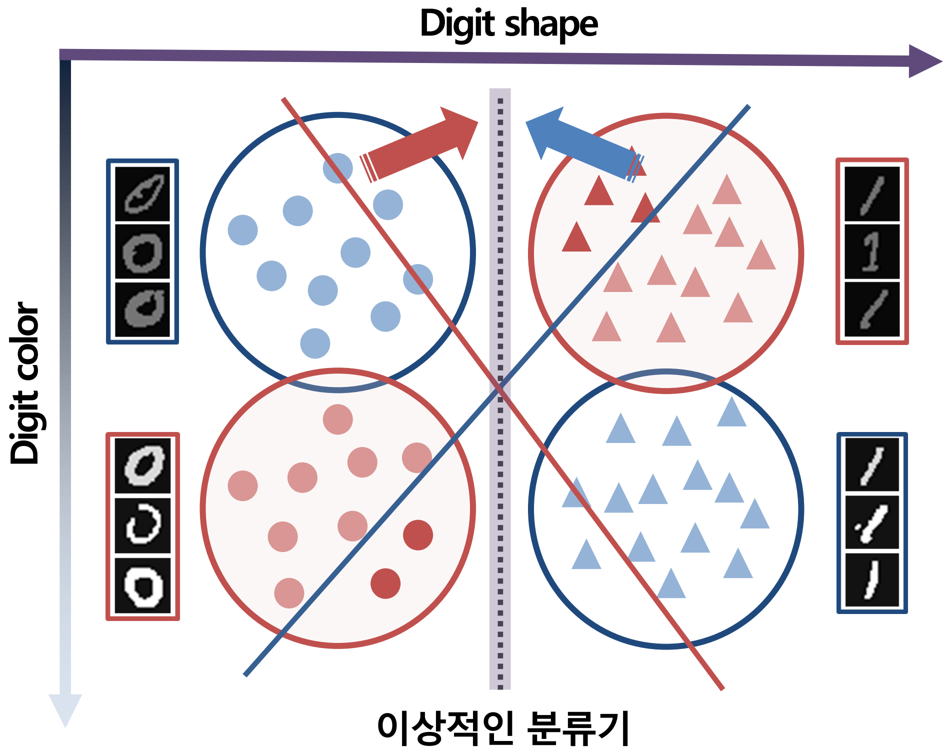
ICON

ICON: Invariant CONsistency learning

❖ Considering inherent distribution in target domain

- 모델은 labeled source domain 데이터 구조에 더 집중하여 spurious features에 과적합 되고 target domain 성능이 저하될 수 있음
- Target domain 내의 내재적 특성을 파악하여, **모델로 하여금 target domain 구조를 더 잘 반영할 수 있게끔 학습** 필요





Cluster Assumption
Even though target domain is unlabeled, we can still identify the sample clusters in target.

ICON

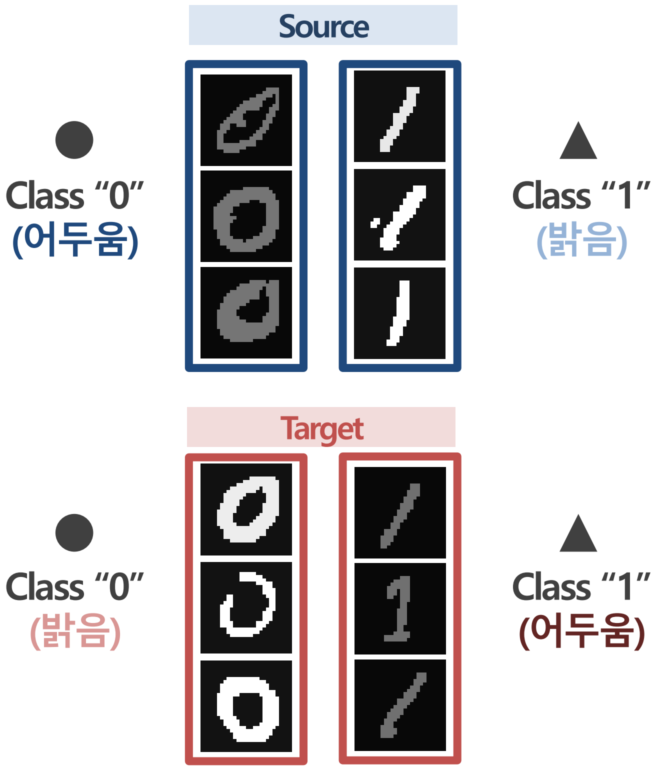
ICON: Invariant CONsistency learning

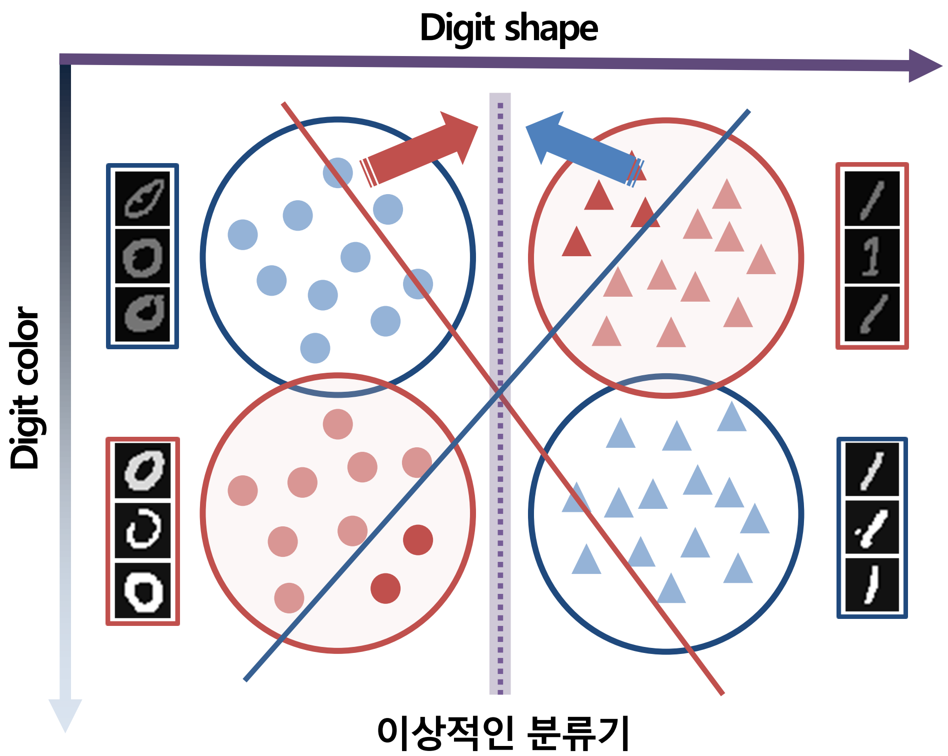
Consistency

Make a classifier whose prediction is consistent with the labels in the source and clusters in the target!

동일 class 내지는 cluster에 속하는 samples 간 유사도↑ (or vice versa)

이로써 classifier는 **두 도메인 데이터 특성을 모두 고려하여 학습**되므로 각 도메인에 특화된 **spurious correlations에 덜 의존**





Cluster Assumption

Even though target domain is unlabeled, we can still identify the sample clusters in target.

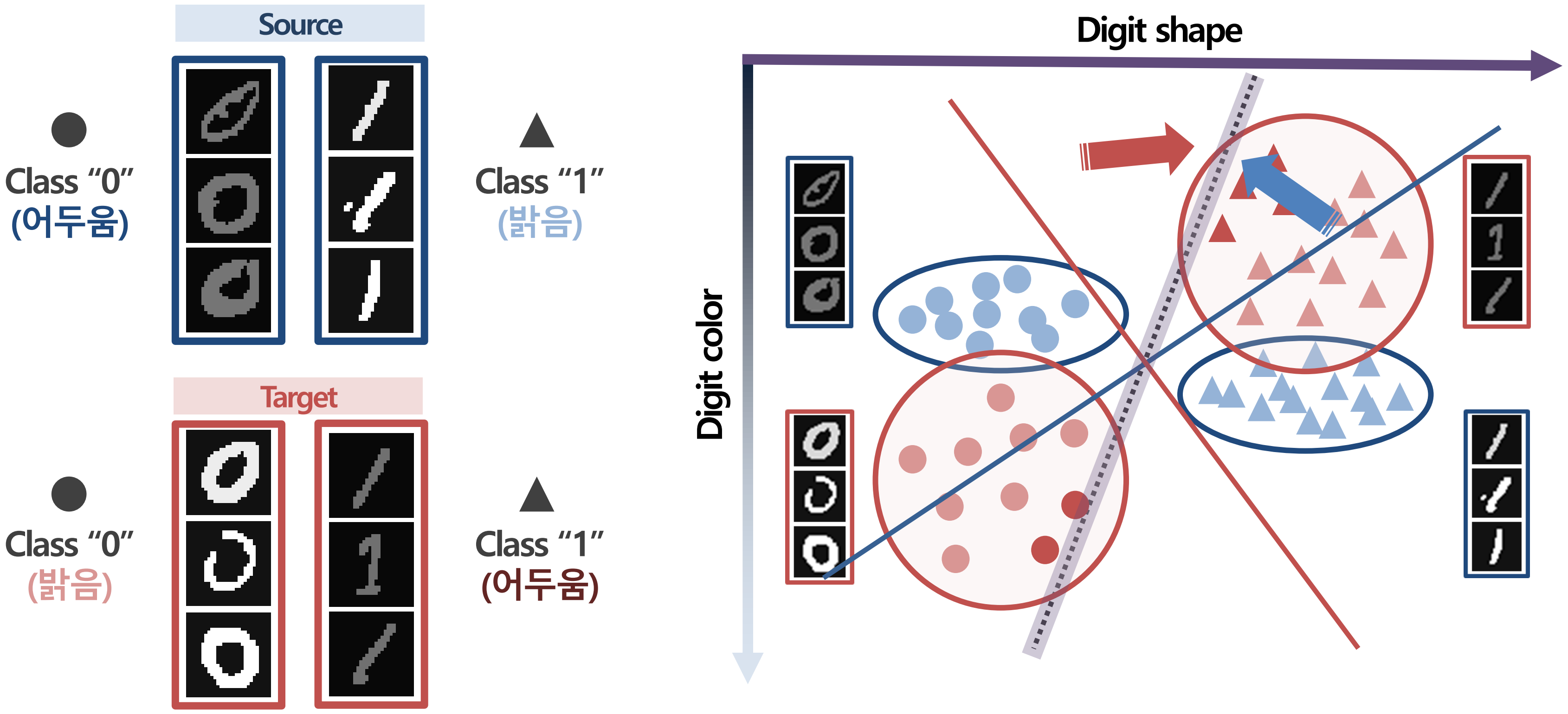
ICON

ICON: Invariant CONsistency learning

Invariance

Give equal status to the source and target domains!

어느 한 쪽 도메인에서 spurious correlation 영향력이 매우 큰 경우에는 단순히 일관된 예측 결과를 출력하는 것만으로는 부족
spurious correlation 의존성을 낮추어 일관적인 예측 결과를 출력할 뿐 아니라, **균형 잡힌 최적의 성능을 보이는 classifier** 학습



Spurious correlation이 매우 강한 경우
Spurious correlation을 규제 가능한 추가 제약 조건 필요

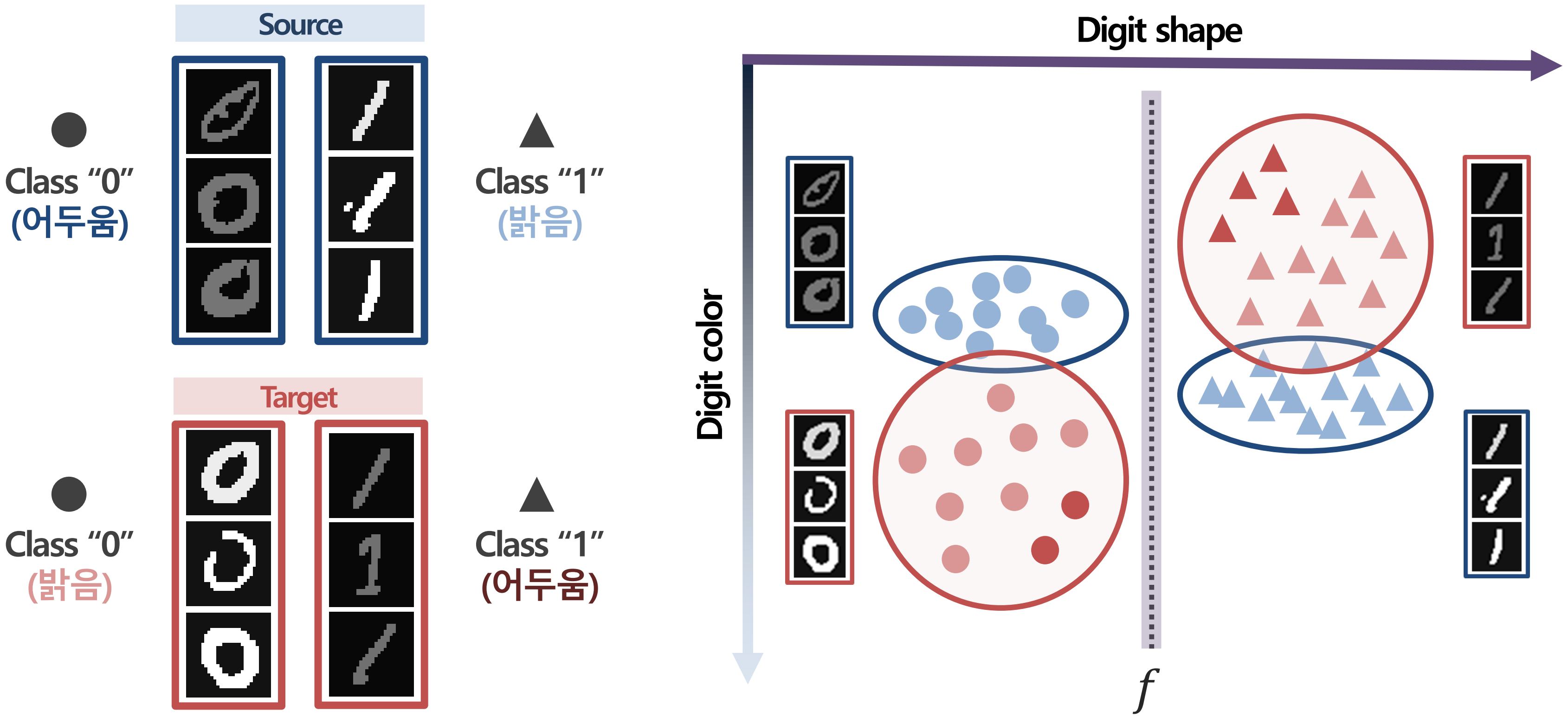
ICON

ICON: Invariant CONsistency learning

GOAL

Source와 Target 모두에서 (1) 일관적인 예측 결과를 출력하고 (CONSISTENT), (2) 최적의 성능을 보이는 (INVARIANT)

Consistent and Invariant Classifier f 구축



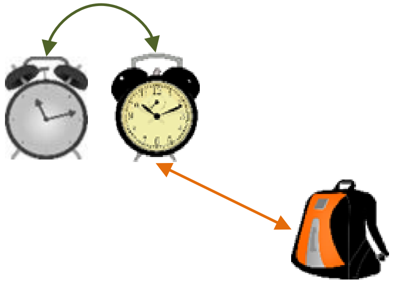
ICON

ICON: Invariant CONsistency learning

❖ **Consistency: 모든 도메인에서 예측 값의 유사성을 극대화 (=일관된 예측 결과 출력)**

Consistency with S : $BCE(S, f)$ ↓

Consistency with T : $BCE(T, f)$ ↓



Form of contrastive loss!

$$BCE(\mathcal{D}, f) = -\mathbb{E}_{(x_i, x_j)\sim\mathcal{D}}[b \log(\hat{y}) + (1-b)\log(1-\hat{y})],$$

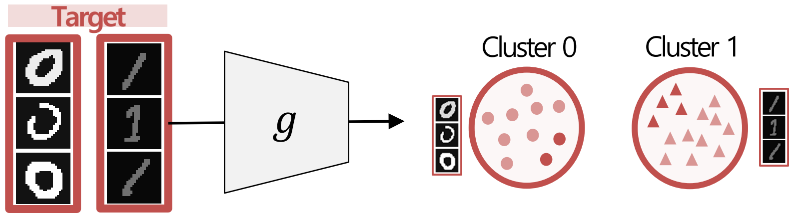
where $\hat{y} = f(x_i)^{\mathrm{T}} f(x_j)$,

and $b = \begin{cases} \mathbb{I}(y_i = y_j), \ \mathcal{D} = S \\ \mathbb{I}(\operatorname{argmax} g(x_i) = \operatorname{argmax} g(x_j)), \ \mathcal{D} = T \end{cases}$

샘플 pair가 같은 class이면 1 아니면 0

샘플 pair가 같은 cluster면 1 아니면 0

g는 rank-statistics algorithm을 통해
target samples을 clustering 하는 네트워크



ICON

GOAL: S와 T 모두에서 (1) **일관적인 예측 결과**를 출력하고, (2) **최적의 성능**을 보이는 **Consistent and Invariant Classifier** 구축

Consistency　　Invariance

ICON: Invariant CONsistency learning

❖ **Consistency: 모든 도메인에서 예측 값의 유사성을 극대화 (=일관된 예측 결과 출력)**

Consistency with S : $BCE(S, f)$ ↓　　Consistency with T : $BCE(T, f)$ ↓

Form of contrastive loss!

$$BCE(\mathcal{D}, f) = -\mathbb{E}_{(x_i, x_j) \sim \mathcal{D}}[b \log(\hat{y}) + (1-b)\log(1-\hat{y})],$$

Weaknesses:

1. The performance of ICON on Office-Home and VisDA-2017 is inferior to that of SoTA. For example, CDTrans [Xu+, ICLR2022] achieves 88.4% on VisDA-2017 and 80.5% on OfficeHome, both higher than ICON.
2. Since the assumptions underlying the theorem appear to be quite strong, it is questionable to what extent they are valid in practice. (this is discussed more or less in the limitation part in the supplementary material, though.)
3. Intuitively, the principle of ICON (i.e., bringing features within the same class/cluster closer together) seems highly similar to that of contrastive learning, which is also a major approach to unsupervised domain adaptation (e.g. [Shen+, ICML2022], [Wang+, TMM2023]). Discussing the differences will highlight the property and uniqueness of ICON.
4. While this may be outside the scope of this paper, it would be interesting to discuss the possibility of extending to more advanced domain adaptation problems, such as universal domain adaptation and source-free domain adaptation. Since ICON (perhaps implicitly) assumes that the number of classes in the source and target are the same, and the impact of the number of clusters on accuracy is significant (see Table 2). So the performance of ICON on universal domain adaptation, where the number of classes cannot be assumed to be the same, may not be as promising. Application to source-free domain adaptation is also non-trivial.

[Xu+, ICLR2022] CDTrans: Cross-domain Transformer for Unsupervised Domain Adaptation

[Shen+, ICML2022] Connect, Not Collapse: Explaining Contrastive Learning for Unsupervised Domain Adaptation

[Wang+, TMM2023] Cross-Domain Contrastive Learning for Unsupervised Domain Adaptation

GOAL: S와 T 모두에서 (1) **일관적인 예측 결과**를 출력하고, (2) **최적의 성능**을 보이는 **Consistent and Invariant Classifier** 구축

Consistency Invariance

ICON

ICON: Invariant CONsistency learning

❖ **Consistency: 모든 도메인에서 예측 값의 유사성을 극대화 (=일관된 예측 결과 출력)**

Consistency with S : $BCE(S, f)$ ↓

Consistency with T : $BCE(T, f)$ ↓

Form of contrastive loss!

$$BCE(\mathcal{D}, f) = -\mathbb{E}_{(x_i, x_j)\sim\mathcal{D}}[b \log(\hat{y}) + (1-b)\log(1-\hat{y})],$$

Weaknesses:

1. The performance of ICON on Office-Home and VisDA-2017 is inferior to that of SoTA. For example, CDTrans [Xu+, ICLR2022] achieves 88.4% on VisDA-2017 and 80.5% on OfficeHome, both higher than ICON.
2. Since the assumptions underlying the theorem appear to be quite strong, it is questionable to what extent they are valid in practice. (this is discussed more or less in the limitation part in the supplementary material, though.)
3. Intuitively, the principle of ICON (i.e., bringing features within the same class/cluster closer together) seems highly similar to that of contrastive learning, which is also a major approach to unsupervised domain adaptation (e.g. [Shen+, ICML2022], [Wang+, TMM2023]). Discussing the differences will highlight the property and uniqueness of ICON.
4. While this may be outside the scope of this paper, it would be interesting to discuss the possibility of extending to more advanced domain adaptation problems, such as universal domain adaptation and source-free domain adaptation. Since ICON (perhaps implicitly) assumes that the number of classes in the source and target are the same, and the impact of the number of clusters on accuracy is significant (see Table 2). So the performance of ICON on universal domain adaptation, where the number of classes cannot be assumed to be the same, may not be as

W3 - Differences with methods based on contrastive learning. Yes, our method can be viewed as contrastive learning. The differences with previous methods lie in ***what to contrast***. For example, [Shen+, ICML2022] contrasts augmented samples, *i.e.*, a sample under different augmented views shares similar features, and different samples have dissimilar features. [Wang+, TMM2023] contrasts ***cross-domain*** sample pairs (one from source domain S and the other from target domain T), *i.e.*, pairs from the same class share similar features (and vice versa). Unfortunately, they still generate T pseudo-labels based on S supervision like self-training methods, and hence are prone to spurious correlations (lines 52-57). Our ICON contrasts ***in-domain*** sample pairs (both samples from S or T), *i.e.*, pairs from the same class in S or cluster in T share similar features (and vice versa). In this way, our cluster labels in T only capture the inherent distribution of T, which helps remove spurious correlations (lines 66-80).

GOAL: S와 T 모두에서 (1) **일관적인 예측 결과**를 출력하고, (2) **최적의 성능**을 보이는 **Consistent and Invariant Classifier** 구축

Consistency　　Invariance

ICON

ICON: Invariant CONsistency learning

❖ **Invariant Consistency (ICON): 모든 도메인에 대해 최적의 성능을 일관되게 출력**

- **Objective**:

$$\min_{\theta,f} \overbrace{CE(S,f) + \alpha\mathcal{L}_{self-training}}^{Self\ Training} + \overbrace{BCE(S,f) + BCE(T,f)}^{Consistency}$$

$$\underbrace{s.t.\ \ f \in \operatorname*{argmin}_{\bar{f}} BCE\left(S,\bar{f}\right) \cap \operatorname*{argmin}_{\bar{f}} BCE\left(T,\bar{f}\right).}_{Invariance}$$

f는 S, T 모두에서 BCE loss를 최소화 하는 분류기 집합 $\bar{f}$ 에 속해야 함

ICON

ICON: Invariant CONsistency learning

GOAL: S와 T 모두에서 (1) **일관적인 예측 결과**를 출력하고, (2) **최적의 성능**을 보이는 **Consistent and Invariant Classifier** 구축

Consistency　　　Invariance

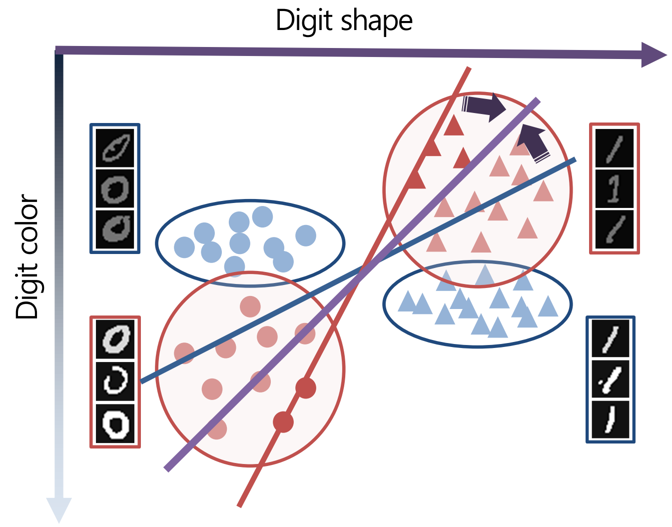
❖ **Invariant Consistency (ICON): 모든 도메인에 대해 최적의 성능을 일관되게 출력**

- **Objective**:

$$\min_{\theta,f} \overbrace{CE(S,f) + \alpha\mathcal{L}_{self-training}}^{Self\ Training} + \overbrace{BCE(S,f) + BCE(T,f)}^{Consistency}$$

$$\underbrace{s.t.\ \ f \in \operatorname*{argmin}_{\bar{f}} BCE\left(S,\bar{f}\right) \cap \operatorname*{argmin}_{\bar{f}} BCE\left(T,\bar{f}\right).}_{Invariance}$$

f는 S, T 모두에서 BCE loss를 최소화 하는 분류기 집합 $\bar{f}$ 에 속해야 함



- Source가 target을 dominate하여, **spurious correlation에 큰 가중치를 부여하는 분류기 학습 위험**
  - ✓ Target 고유의 군집 구조를 학습하지 못하고, source에서 학습된 잘못된 상관관계를 단순히 적용하는 방향으로 학습할 가능성이 있음
- Target의 군집 구조도 반영하여 최적화되도록 강제함으로써 (or vice versa), source spurious correlations이 강하게 학습되더라도 이를 완화하도록 규제

ICON

GOAL: S와 T 모두에서 (1) **일관적인 예측 결과**를 출력하고, (2) **최적의 성능**을 보이는 **Consistent and Invariant Classifier** 구축

Consistency　　Invariance

ICON: Invariant CONsistency learning

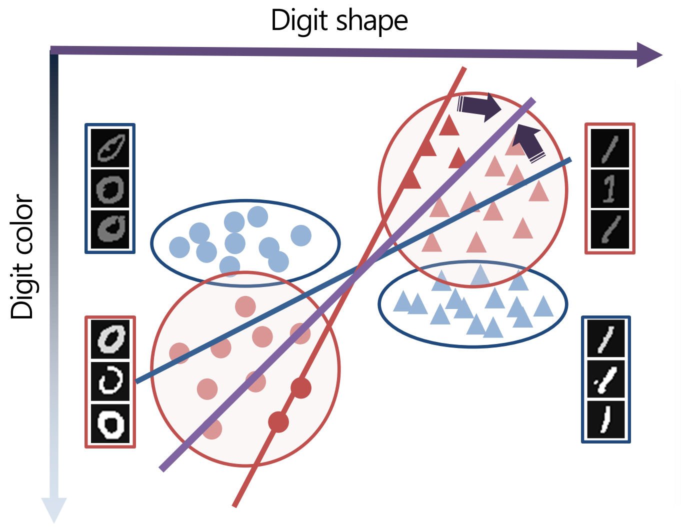
❖ Invariant Consistency (ICON): 모든 도메인에 대해 최적의 성능을 일관되게 출력

- **Objective**:

$$\min_{\theta,f} \underbrace{CE(S,f) + \alpha\mathcal{L}_{self-training}}_{Self\ Training} + \underbrace{BCE(S,f) + BCE(T,f)}_{Consistency}$$

$$\underbrace{s.t.\ \ f \in \operatorname*{argmin}_{\bar{f}} BCE\left(S,\bar{f}\right) \cap \operatorname*{argmin}_{\bar{f}} BCE\left(T,\bar{f}\right).}_{Invariance}$$

f는 S, T 모두에서 BCE loss를 최소화 하는 분류기 집합 $\bar{f}$ 에 속해야 함



- Source가 target을 dominate하여, **spurious correlation에 큰 가중치를 부여하는 분류기 학습 위험**
  - ✓ Target 고유의 군집 구조를 학습하지 못하고, source에서 학습된 잘못된 상관관계를 단순히 적용하는 방향으로 학습할 가능성이 있음
- **Target의 군집 구조도 반영하여 최적화되도록 강제함**으로써 (or vice versa), source spurious correlations이 강하게 학습되더라도 이를 완화하도록 규제

ICON

GOAL: S와 T 모두에서 (1) **일관적인 예측 결과**를 출력하고, (2) **최적의 성능**을 보이는 **Consistent and Invariant Classifier** 구축

Consistency Invariance

ICON: Invariant CONsistency learning

❖ **Invariant Consistency (ICON): 모든 도메인에 대해 최적의 성능을 일관되게 출력**

- **Objective**:

Bi-level Optimization!

$$\min_{\theta,f} \overbrace{CE(S,f) + \alpha\mathcal{L}_{self-training}}^{Self\ Training} + \overbrace{BCE(S,f) + BCE(T,f)}^{Consistency}$$

$$\underbrace{s.t.\ \ f \in \operatorname*{argmin}_{\bar{f}} BCE\left(S,\bar{f}\right) \cap \operatorname*{argmin}_{\bar{f}} BCE\left(T,\bar{f}\right).}_{Invariance}$$

f는 S, T 모두에서 BCE loss를 최소화 하는 분류기 집합 $\bar{f}$ 에 속해야 함

→ f를 찾는 과정이 θ에 의존적 (θ가 update 될 때마다 변하는 feature space에 맞춰 재탐색 필요)

→ 즉, **f와 θ를 함께 찾는 bi-level optimization problem**

Practical Implementation

$$\min_{\theta,f} CE(S,f) + \alpha\mathcal{L}_{self-training} + BCE(S,f) + BCE(T,f) + \beta\mathrm{Var}(\{BCE(S,f), BCE(T,f)\})$$

수렴이 어려운 문제 때문에 위 제약 조건을 완화한 형태의
손실함수인 **REx loss**[5]를 이용함

[5] Krueger, D., Caballero, E., Jacobsen, J. H., Zhang, A., Binas, J., Zhang, D., ... & Courville, A. (2021, July). Out-of-distribution generalization via risk extrapolation (rex). In International conference on machine learning (pp. 5815-5826). PMLR.

ICON

GOAL: S와 T 모두에서 (1) **일관적인 예측 결과**를 출력하고, (2) **최적의 성능**을 보이는 **Consistent and Invariant Classifier** 구축

Consistency　　　Invariance

ICON: Invariant CONsistency learning

❖ Invariant Consistency (ICON): 모든 도메인에 대해 최적의 성능을 일관되게 출력

- Objective:

Bi-level Optimization!

Self Training　　　Consistency

REx (Risk Extrapolation) loss:

- **도메인 간 편차를 초월 (extrapolate)**하여 모델이 입력 데이터의 원인적 특성 (causal feature)을 학습하도록 유도
  - ✓ 즉, 'digit color'와 같이 도메인마다 다른 **환경적 특성을 배제**하고, 도메인에 무관하게 공통적으로 나타나는 **원인적 특성 (e.g., digit shape)에 기반한 예측을 수행**하도록 함

✓ $L_{REx} = \max_{\mathcal{D}} R_{\mathcal{D}}(f) - \min_{\mathcal{D}} R_{\mathcal{D}}(f)$

그러나, 특정 도메인에서의 손실이 지나치게 크면 다른 도메인의 학습에 방해가 될 수 있음

VREx (Variance Risk Extrapolation) loss:

- **REx의 변형으로, 도메인 별 손실의 "분산"을 최소화 하는 방식으로 정의됨**

✓ $L_{VREx} = Var(R_{\mathcal{D}}(f))$

함

취 재탐색 필요).

Practical Implementation

$$\min_{\theta,f} CE(S,f) + \alpha\mathcal{L}_{self-training} + BCE(S,f) + BCE(T,f) + \beta \mathrm{Var}(\{BCE(S,f), BCE(T,f)\})$$

수렴이 어려운 문제 때문에 위 제약 조건을 완화한 형태의 손실함수인 **REx loss**[5]를 이용함

[5] Krueger, D., Caballero, E., Jacobsen, J. H., Zhang, A., Binas, J., Zhang, D., ... & Courville, A. (2021, July). Out-of-distribution generalization via risk extrapolation (rex). In International conference on machine learning (pp. 5815-5826). PMLR.

ICON

GOAL: S와 T 모두에서 (1) **일관적인 예측 결과**를 출력하고, (2) **최적의 성능**을 보이는 **Consistent and Invariant Classifier** 구축

Consistency Invariance

ICON: Invariant CONsistency learning

❖ Summary

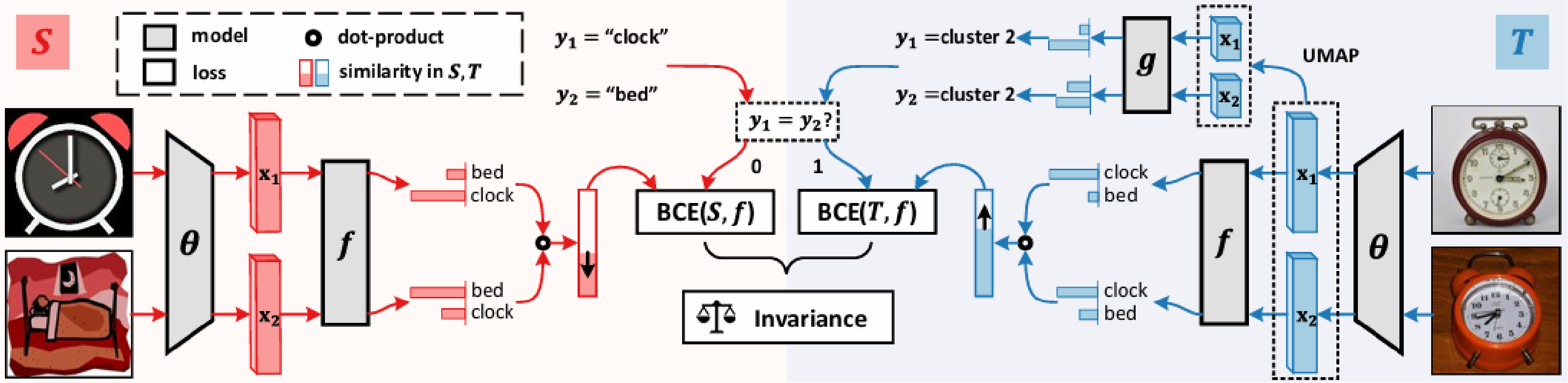
- Complete Objective: $\min_{\theta,f} CE(S,f) + \alpha\mathcal{L}_{self-training} + BCE(S,f) + BCE(T,f) + \beta\mathrm{Var}(\{BCE(S,f), BCE(T,f)\})$

  $+ \, Cluster\ loss + Cluster\ Self\ Training\ loss + \ Tsallis\ Entropy\ loss + Equivariance\ loss$[6]

  $\mathcal{L}_{EqInv} = \mathbb{E}_{x\sim S\cup T}\left\|\hat{y}(aug(x)) - aug(\hat{y}(x))\right\|^2$

  특정 변환 (transform)에 대해 동등성을 유지하도록, 즉, 모델이 입력 데이터에 적용된 변환을 반영한 표현을 학습하도록 유도

- To refine noisy pseudo-labels (Consistency + Invariance):
  - Giving equal status to the two domains; learning an invariant classifier whose prediction is simultaneously consistent with the labels in the source domain and clusters in the target domain



[5] Wad, T., Sun, Q., Pranata, S., Jayashree, K., & Zhang, H. (2022, October). Equivariance and invariance inductive bias for learning from insufficient data. In European Conference on Computer Vision (pp. 241-258). Cham: Springer Nature Switzerland.

ICON

GOAL: S와 T 모두에서 (1) **일관적인 예측 결과**를 출력하고, (2) **최적의 성능**을 보이는 **Consistent and Invariant Classifier** 구축

Consistency Invariance

ICON: Invariant CONsistency learning

❖ Summary

- Complete Objective: $\min_{\theta,f} CE(S,f) + \alpha\mathcal{L}_{self-training} + BCE(S,f) + BCE(T,f) + \beta\mathrm{Var}(\{BCE(S,f), BCE(T,f)\})$

$$+ Cluster\ loss + Cluster\ Self\ Training\ loss + Tsallis\ Entropy\ loss + Equivariance\ loss$$[6]

$$\mathcal{L}_{EqInv} = \mathbb{E}_{x\sim S\cup T}\left\|\hat{y}(aug(x)) - aug(\hat{y}(x))\right\|^2$$

특정 변환 (transform)에 대해 동등성을 유지하도록, 즉, 모델이 입력 데이터에 적용된 변환을 반영한 표현을 학습하도록 유도

Components ablation study

Method	OFFICE-HOME	VISDA-2017
FixMatch	69.1	76.6
FixMatch+CON	74.1	82.0
FixMatch+CON+INV	75.8	87.4
Cluster with 2×#classes	69.7	78.6
Cluster with 0.5×#classes	67.5	76.2
Cluster with k-NN	72.1	85.6

Table 4: Ablations on each ICON component. CON denotes the consistency loss in S and T. INV denotes the invariance constraint.

[5] Wad, T., Sun, Q., Pranata, S., Jayashree, K., & Zhang, H. (2022, October). Equivariance and invariance inductive bias for learning from insufficient data. In European Conference on Computer Vision (pp. 241-258). Cham: Springer Nature Switzerland.

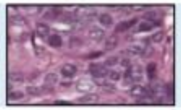
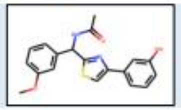
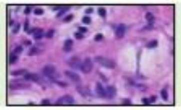
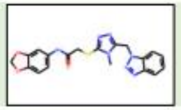
ICON

ICON: Invariant CONsistency learning

❖ Experiment Settings

- **Datasets** : **10개** – OFFICE-HOME, VISDA-2017 | WILDS 2.0 (8개 데이터, image, text, graph 등 **다양한 modalities**) (**다양한 tasks**: +reg, +detection)

  다양한 modality의 데이터셋 활용 → image 등에만 국한된 기법 사용하지 않음 (e.g. mixup)

Dataset	OFFICE-HOME	VISDA-2017	IWILDCAM	CAMELYON17	FMOW	POVERTYMAP	GLOBALWHEAT	OGB-MOLPCBA	CIVILCOMMENTS	AMAZON
Sample x	object image	object image	camera trap photo	tissue slide	satellite image	satellite image	wheat image	molecular graph	online comment	product review
Label y	65 categories	12 categories	182 species	tumor/not	62 land uses	asset wealth	wheat bbox	bioassays	toxic/not	5 review scores
Task	classification	classification	classification	classification	classification	regression	detection	classification	classification	classification
Source S	various*	synthetic images	photos from 243 traps	slides from hospital 1-3	images from 2002-2013	images in 14 countries	images in Europe	44,930 scaffold groups	online articles*	1,252 reviewers
Example S									I applaud your father. He was a good man! We need more like him.	Super easy to put together. Very well built.
#Samples S	average 3,875	152,397	129,809	302,436	76,863	~10,000	2,943	350,343	269,038	245,502
Target T	various*	real photos	photos from 3215 traps	slides from hospital 5	images from 2016-2018	images in 5 countries	images across the world	43,793 scaffold groups	online articles*	1,334 reviewers
Example T									As a Christian, I will not be patronizing any of those businesses.	I am disappointed in the quality of these.
#Samples T	average 3,875	55,388	819,120	600,030	173,208	261,396	42,445	517,048	1,551,515	268,761
Evaluation	average acc.	mean-class accuracy	macro-F1	acc.	worst-region acc.*	Pearson correlation* %	acc.	average precision	worst-group acc.*	10th percentile acc.

UMAP, EqInv를 이용한 Feature preprocessing 수행

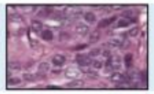
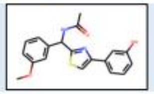
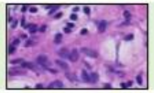
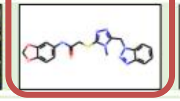
→ (highlight causal feature to improve clustering)

ICON

ICON: Invariant CONsistency learning

❖ Experiment Settings

- **Backbones** :
  - Image Classification Task : (1) Office-Home, VisDA, IWildCam - ResNet-50 / (2) FMOW – DenseNet-121 (Both backbones are pretrained on ImageNet)
  - Medical Img. Classification Task (CamelYon17) : DenseNet-121 pretrained by SwAV with unlabeled S and T
  - Object Detection Task (Global Wheat) : Faster-RCNN
  - Text Classification Task (Civil Comments, Amazon) : DistilBERT
  - Regression Task (**Poverty Map**) : **No Pretrained Model Available**, Multi-spectral ResNet-18 **trained with the labeled source domain**
  - Molecular Graph Classification Task (**OGB-MolPCBA**) : **No Pretrained Model Available**, Graph isomorphism network **trained with the labeled source domain**

	ResNet			SwAV	DenseNet	Source	FasterRCNN	Source	DistilBERT	
Dataset	**OFFICE-HOME**	**VISDA-2017**	**IWILDCAM**	**CAMELYON17**	**FMOW**	**POVERTYMAP**	**GLOBALWHEAT**	**OGB-MOLPCBA**	**CIVILCOMMENTS**	**AMAZON**
Sample x	object image	object image	camera trap photo	tissue slide	satellite image	satellite image	wheat image	molecular graph	online comment	product review
Label y	65 categories	12 categories	182 species	tumor/not	62 land uses	asset wealth	wheat bbox	bioassays	toxic/not	5 review scores
Task	classification	classification	classification	classification	classification	regression	detection	classification	classification	classification
Source S	various*	synthetic images	photos from 243 traps	slides from hospital 1-3	images from 2002-2013	images in 14 countries	images in Europe	44,930 scaffold groups	online articles*	1,252 reviewers
Example S									I applaud your father. He was a good man! We need more like him.	Super easy to put together. Very well built.
#Samples S	average 3,875	152,397	129,809	302,436	76,863	~10,000	2,943	350,343	269,038	245,502
Target T	various*	real photos	photos from 3215 traps	slides from hospital 5	images from 2016-2018	images in 5 countries	images across the world	43,793 scaffold groups	online articles*	1,334 reviewers
Example T									As a Christian, I will not be patronizing any of those businesses.	I am disappointed in the quality of these.

ICON

ICON: Invariant CONsistency learning

❖ Main Results

- **WILDS 2.0 – Failure Case**
  - **POVERTYMAP & OGB-MOLPCBA**에 대해서는 상대적으로 **성능 향상이 두드러지지 않음**
  - 이유 : (1) no pretraining-available → source domain으로만 사전학습을 하였는데, 이로써 **source domain 특성에 biased 되었을 가능성**

    (2) the ground-truth number of classes T is nor well-defined → regression + OGB는 nan 값 존재 → clustering assumption 미충족

Dataset	OFFICE-HOME	VISDA-2017	IWILDCAM	CAMELYON17	FMOW	POVERTYMAP	GLOBALWHEAT	OGB-MOLPCBA	CIVILCOMMENTS	AMAZON
Task	classification	classification	classification	classification	classification	regression	detection	classification	classification	classification
Evaluation	average acc.	mean-class accuracy	macro-F1	acc.	worst-region acc.*	Pearson correlation* %	acc.	average precision	worst-group acc.*	10th percentile acc.
Existing Methods	**GVB**		**Empirical Risk Minimization (ERM)**							
	70.4	75.3	47.0 / 32.2	90.6 / 82.0	60.6 / 34.8	65 / 48	77.8 / 51.0	- / **28.3**	89.8 / 66.6	**72.0** / 54.2
	TCM		**CORAL**							
	70.7	75.8	40.5 / 27.9	90.4 / 77.9	58.9 / 34.1	54 / 36	- / -	- / 26.6	- / -	71.7 / 53.3
	SENTRY		**DANN**							
	72.0	76.7	48.5 / 31.9	86.9 / 68.4	57.9 / 34.6	50 / 33	- / -	- / 20.4	- / -	71.7 / 53.3
	CST		**Pseudo-Label**							
	72.2	80.6	47.3 / 30.3	91.3 / 67.7	60.9 / 33.7	- / -	73.3 / 42.9	- / 19.7	90.3 / 66.9	71.6 / 52.3
	ToAlign	**MDD**	**Noisy Student**							
	72.7	77.8	47.5 / 32.1	93.2 / 86.7	61.3 / 37.8	61 / 42	78.1 / 46.8	- / 27.5	- / -	- / -
	FixBi	**MT+16augs**	**FixMatch**						**Masked LM**	
	73.0	82.8	46.3 / 31.0	91.3 / 71.0	58.6 / 32.1	54 / 30	- / -	- / -	89.4 / 65.7	71.9 / 53.9
	ATDOC	**MCC+NWD**	**SwAV**						**ERM (labelled *T*)**	
	73.2	83.7	47.3 / 29.0	92.3 / 91.4	61.8 / 36.3	60 / 45	- / -	- / -	89.9 / 69.4	73.6 / 56.4
ICON	**75.8** +2.6	**87.4** +3.7	**50.6 / 34.5** +2.3	**95.6 / 93.8** +2.4	**62.2 / 39.9** +2.1	**65 / 49** +1	**78.6 / 52.3** +1.3	**- / 28.3** +0.0	89.7 / **68.8** +1.9	71.9 / **54.7** +0.5

Source Test 성능 / Target Test 성능

Conclusion

❖ **How to make reliable target pseudo-labels?**

1. **SENTRY** (2021, ICCV) : augmented samples과의 예측 일관성 높이기

Need to refine noisy pseudo-labels

2. **CST** (2021, NeurIPS) : target classifier가 source domain에서도 잘 동작하도록 만들기 (reverse step)

Need to consider inherent distribution of target domain

3. **ICON** (2023, NeurIPS) : target domain의 구조적 정보를 잘 활용하여 예측 정확도 및 일관성 높이기

Conclusion

❖ **How to make reliable target pseudo-labels?**

1. **SENTRY** (2021, ICCV) : augmented samples과의 예측 일관성 높이기
   - Source domain으로 사전학습된 모델을 활용하여 source 정보에 과적합될 가능성
   - 예측 일관성을 높이는 것이 예측 정확도를 보장하지 않는 문제

2. **CST** (2021, NeurIPS) : target classifier가 source domain에서도 잘 동작하도록 만들기 (reverse step)
   - Target classifier 구축 시 labeled source domain으로 훈련된 source classifier의 예측 값에 의존적
   - Ablation study 확인 결과, 제안한 reverse step보다 Tsallis entropy의 효과가 두드러짐을 확인

3. **ICON** (2023, NeurIPS) : target domain의 구조적 정보를 잘 활용하여 예측 정확도 및 일관성 높이기
   - 기존에 제안된 손실함수를 적절히 종합하여, 적게는 7개, 많게는 8개 이상의 losses를 활용하여 학습 (데이터셋에 따라 상이)

Need to refine noisy pseudo-labels

Need to consider inherent distribution of target domain

Thank You

ICON

ICON: Invariant CONsistency learning

❖ Main Results (1/2)

- **Classic UDA Benchmarks (OFFICE-HOME, VisDA-2017)**
  - OFFICE-HOME : Spurious correlation을 제거하는 것이 목적인 타 UDA 방법론 CST와 ATDOC 보다 좋은 성능 → Spurious Corr. 제거의 중요성 강조
    - 하지만 두 방법론은 target domain의 pseudo label 산출을 위해 source classifier의 가중치로 초기화된 T-classifier를 이용 → S의 spurious corr. 영향력 여전
  - VisDA : 기존의 VisDa SoTA 중 하나였던 MT+16augs 성능 이김, backbone으로 ResNet-101 사용한 타 방법론보다도 좋은 성능 냈음을 강조
    - 질문. OFFICE-HOME과 VisDA의 비교 방법론 (baselines)이 같지 않음

OFFICE-HOME

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
DANN [15] (2016)	45.6	59.3	70.1	47.0	58.5	60.9	46.1	43.7	68.5	63.2	51.8	76.8	57.6
CDAN [36] (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
SymNet [71] (2019)	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 [72] (2019)	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
SHOT[30] (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
ALDA [7] (2020)	53.7	70.1	76.4	60.2	72.6	71.5	56.8	51.9	77.1	70.2	56.3	82.1	66.6
GVB [9] (2020)	57.0	74.7	79.8	64.6	74.1	74.6	65.2	55.1	81.0	74.6	59.7	84.3	70.4
TCM [69] (2021)	58.6	74.4	79.6	64.5	74.0	75.1	64.6	56.2	80.9	74.6	60.7	84.7	70.7
SENTRY [46] (2021)	61.8	77.4	80.1	66.3	71.6	74.7	66.8	**63.0**	80.9	74.0	66.3	84.1	72.2
CST [34] (2021)	59.0	79.6	83.4	68.4	77.1	76.7	68.9	56.4	83.0	75.3	62.2	85.1	73.0
ToAlign [66] (2021)	57.9	76.9	80.8	66.7	75.6	77.0	67.8	57.0	82.5	75.1	60.0	84.9	72.0
FixBi [42] (2021)	58.1	77.3	80.4	67.7	79.5	78.1	65.8	57.9	81.7	**76.4**	62.9	86.7	72.7
ATDOC [31] (2021)	60.2	77.8	82.2	68.5	78.6	77.9	68.4	58.4	83.1	74.8	61.5	87.2	73.2
SDAT [47] (2022)	58.2	77.1	82.2	66.3	77.6	76.8	63.3	57.0	82.2	74.9	64.7	86.0	72.2
MCC+NWD [6] (2022)	58.1	79.6	83.7	67.7	77.9	78.7	66.8	56.0	81.9	73.9	60.9	86.1	72.6
kSHOT* [57] (2022)	58.2	80.0	82.9	61.1	80.3	80.7	**71.3**	56.8	83.2	75.5	60.3	86.6	73.9
ICON (Ours)	**63.3**	**81.3**	**84.5**	**70.3**	**82.1**	**81.0**	70.3	61.8	**83.7**	75.6	**68.6**	**87.3**	**75.8**

Table 2: Break-down of the accuracies in each domain on OFFICE-HOME dataset [62]. *: kSHOT [57] additionally uses the prior knowledge on the percentage of samples in each class in the testing data. Published years are in the brackets after the method names.

VisDA-2017

Method	Backbone	Acc.
MT+16augs [13] (2018)	ResNet-50	82.8
MDD [72] (2019)	ResNet-50	77.8
GVB [9] (2020)	ResNet-50	75.3
TCM [69] (2021)	ResNet-50	75.8
SENTRY [46] (2021)	ResNet-50	76.7
CST [34] (2021)	ResNet-50	80.6
CAN [26] (2019)	ResNet-101	87.2
SHOT [30] (2020)	ResNet-101	82.9
FixBi [26] (2021)	ResNet-101	87.2
MCC+NWD [6] (2022)	ResNet-101	83.7
SDAC [47] (2022)	ResNet-101	84.3
kSHOT* [57] (2022)	ResNet-101	86.1
ICON (Ours)	ResNet-50	**87.4**

Table 3: Mean-class accuracy (Acc.) on VISDA-2017 Synthetic→Real task with the choice of feature backbone. *: details in Table 2 caption. Published years are in the brackets after the method names.

ICON

ICON: Invariant CONsistency learning

❖ Main Results (2/2)

- **WILDS 2.0 – Success Case**
  - IWILDCAM : long-tail distribution (y 분포) 있음에도 좋은 성능
  - CIVILCOMMENTS : "even under the SSL setting"에서 제안 방법론의 우수함을 증명했다고 하는데, "SSL setting"이 뭘 의미하는 지 아직 파악 못함
  - AMAZON: 성능 향상이 그리 크지 않지만, ERM (labeled T)와 근접함을 강조

Dataset	OFFICE-HOME	VISDA-2017	IWILDCAM	CAMELYON17	FMOW	POVERTYMAP	GLOBALWHEAT	OGB-MOLPCBA	CIVILCOMMENTS	AMAZON
Task	classification	classification	classification	classification	classification	regression	detection	classification	classification	classification
Evaluation	average acc.	mean-class accuracy	macro-F1	acc.	worst-region acc.*	Pearson correlation* %	acc.	average precision	worst-group acc.*	10th percentile acc.
Existing Methods	**GVB**		**Empirical Risk Minimization (ERM)**							
	70.4	75.3	47.0 / 32.2	90.6 / 82.0	60.6 / 34.8	65 / 48	77.8 / 51.0	- / **28.3**	89.8 / 66.6	**72.0** / 54.2
	TCM		**CORAL**							
	70.7	75.8	40.5 / 27.9	90.4 / 77.9	58.9 / 34.1	54 / 36	- / -	- / 26.6	- / -	71.7 / 53.3
	SENTRY		**DANN**							
	72.0	76.7	48.5 / 31.9	86.9 / 68.4	57.9 / 34.6	50 / 33	- / -	- / 20.4	- / -	71.7 / 53.3
	CST		**Pseudo-Label**							
	72.2	80.6	47.3 / 30.3	91.3 / 67.7	60.9 / 33.7	- / -	73.3 / 42.9	- / 19.7	90.3 / 66.9	71.6 / 52.3
	ToAlign	**MDD**	**Noisy Student**							
	72.7	77.8	47.5 / 32.1	93.2 / 86.7	61.3 / 37.8	61 / 42	78.1 / 46.8	- / 27.5	- / -	- / -
	FixBi	**MT+16augs**	**FixMatch**						**Masked LM**	
	73.0	82.8	46.3 / 31.0	91.3 / 71.0	58.6 / 32.1	54 / 30	- / -	- / -	89.4 / 65.7	71.9 / 53.9
	ATDOC	**MCC+NWD**	**SwAV**						**ERM (labelled *T*)**	
	73.2	83.7	47.3 / 29.0	92.3 / 91.4	61.8 / 36.3	60 / 45	- / -	- / -	89.9 / 69.4	73.6 / 56.4
ICON	**75.8** +2.6	**87.4** +3.7	**50.6 / 34.5** +2.3	**95.6 / 93.8** +2.4	**62.2 / 39.9** +2.1	**65 / 49** +1	**78.6 / 52.3** +1.3	**- / 28.3** +0.0	89.7 / **68.8** +1.9	71.9 / **54.7** +0.5

Source Test 성능 / Target Test 성능

모든 경우에 대해
Source Test에 대해서도
좋은 성능을 유지함을
강조
(invariance objective)