
How to Transfer Knowledge Across Domains by Deep Neural Network?



2022. 10. 28

Data Mining & Quality Analytics Lab.

김지현

발표자 소개



❖ 김지현 (Jihyun Kim)

- 고려대학교 산업경영공학과 대학원 재학
- Data Mining & Quality Analytics Lab. (김성범 교수님)
- 석·박사 통합 과정 (2022. 03 ~ Present)

❖ Research Interest

- Deep Learning Algorithm for Multivariate Signal Analysis
- Multi Source Unsupervised Domain Adaptation

❖ Contact

- jihyun_k@korea.ac.kr

Contents

❖ Introduction

- Background
- Deep Transfer Learning

❖ Model-Based Strategy

- Sequential Training Approach

❖ Feature-Based Strategy

- Discrepancy-Based Approach
- Adversarial-Based Approach

❖ Conclusions

❖ References

1. Introduction: Deep Transfer Learning

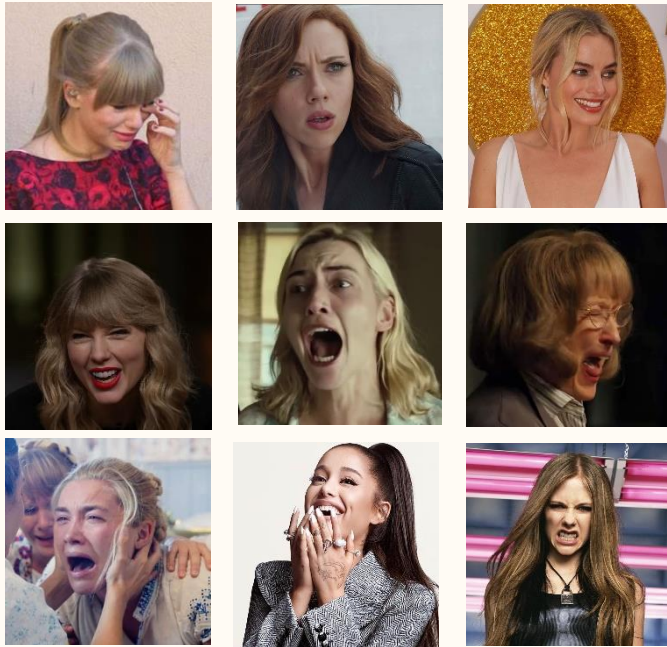
Introduction

Background

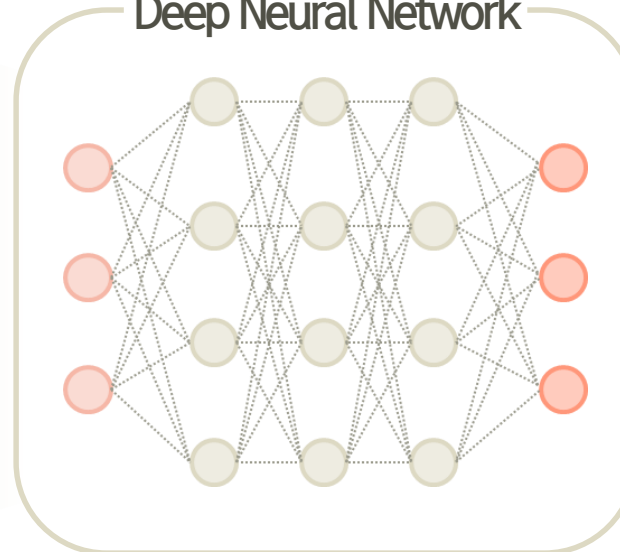
인물 사진을 이용해 감정을 분류하는 문제를 풀어보자!

서구권 여성 인물 사진 데이터

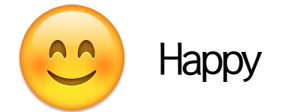
Dataset A



Deep Neural Network



Classification



Happy



Sad



Angry

Introduction

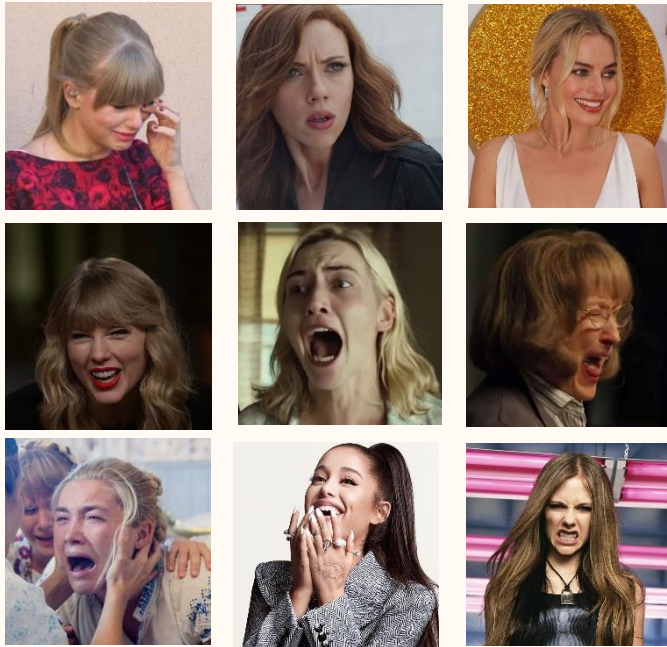
Background

인물 사진을 이용해 감정을 분류하는 문제를 풀어보자!

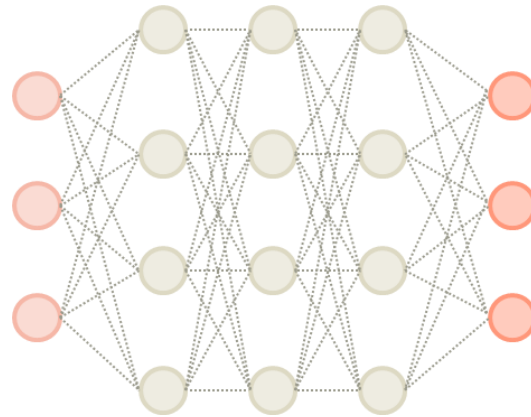
Accuracy 99%

서구권 여성 인물 사진 데이터

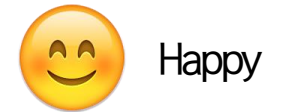
Dataset A



Deep Neural Network



Classification



Happy



Sad



Angry

Introduction

Background

인물 사진을 이용해 감정을 분류하는 문제를 풀어보자!

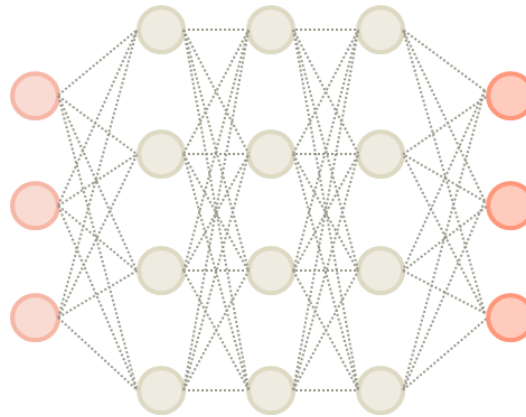
Q1. 성별이 다른 데이터셋을 이용해도 좋은 예측 성능을 보일까?

Dataset B



서구권 남성 인물 사진 데이터

Deep Neural Network



Trained on Dataset A

Classification



Happy



Sad



Angry

Introduction

Background

인물 사진을 이용해 감정을 분류하는 문제를 풀어보자!

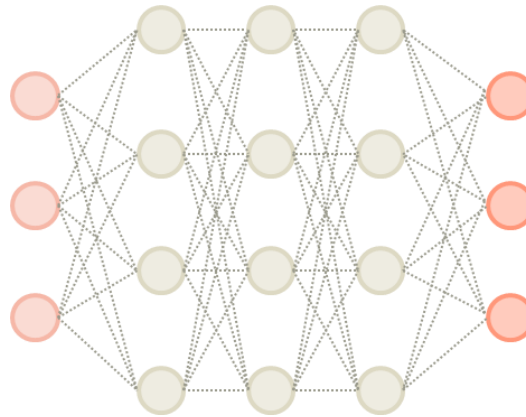
Accuracy 15%

Dataset B



서구권 남성 인물 사진 데이터

Deep Neural Network

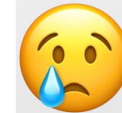


Trained on Dataset A

Classification



Happy



Sad



Angry

예측 성능 하락

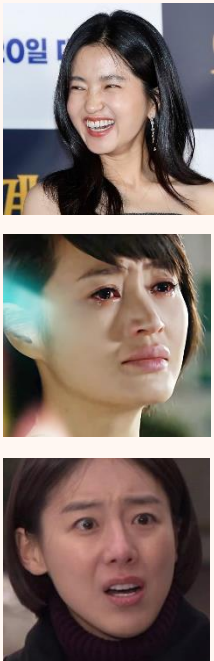
Introduction

Background

인물 사진을 이용해 감정을 분류하는 문제를 풀어보자!

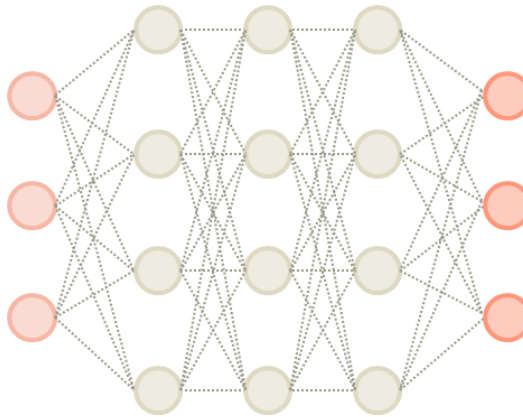
Q2. 인종이 다른 데이터셋을 이용해도 좋은 예측 성능을 보일까?

Dataset C



아시아 여성 인물 사진 데이터

Deep Neural Network



Trained on Dataset A

Classification



Happy



Sad



Angry

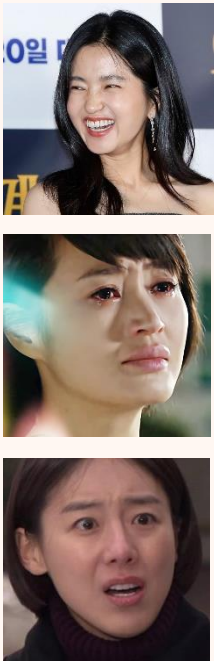
Introduction

Background

인물 사진을 이용해 감정을 분류하는 문제를 풀어보자!

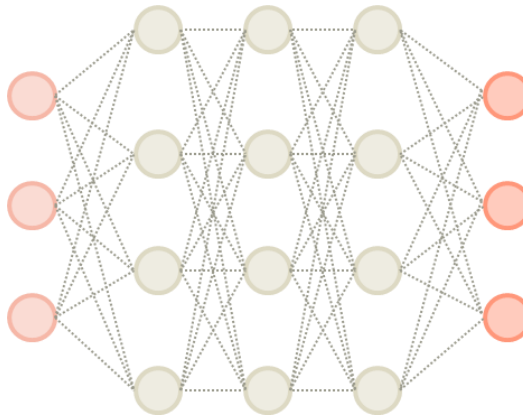
Accuracy 46%

Dataset C



아시아 여성 인물 사진 데이터

Deep Neural Network

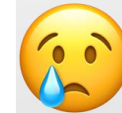


Trained on Dataset A

Classification



Happy



Sad



Angry

예측 성능 하락

Introduction

Background

인물 사진에 감정을 분류하는 문제를 풀어보자!

 Why?

Accuracy 46%

Dataset C

“**학습 데이터와 테스트 데이터가 같은 Feature Space 내에서 동일한 분포를 가져야 한다**”는 가정을 전제로 하고 있음

Domain



Trained on Dataset A

Classification

 Happy

 Sad

 Angry

예측 성능 하락

Introduction

Background

❖ Definition of Domain ($\mathcal{D}$)

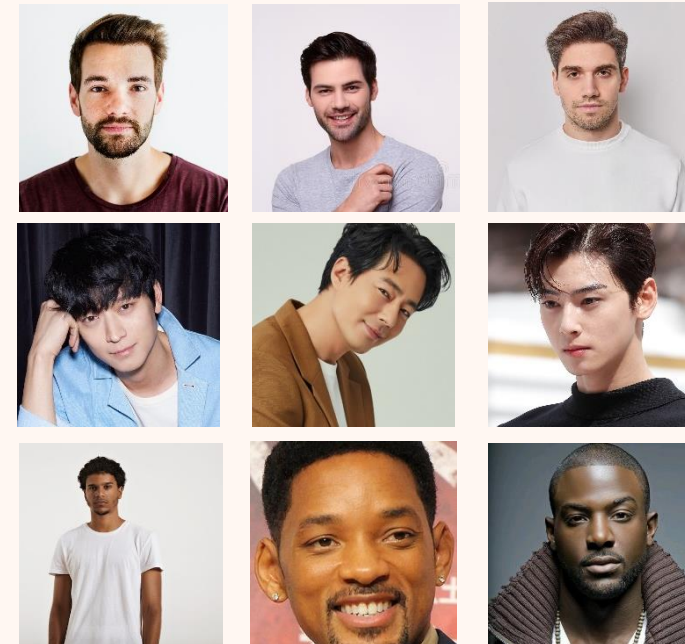
- Domain은 2가지 요소로 이루어짐: (1) Feature Space $\mathcal{X}$, (2) 결합확률분포 $P(X)$, where $X = \{x_1, x_2, \dots, x_n\} \in \mathcal{X}$
 - ✓ $\mathcal{D} = \{\mathcal{X}, P(X)\}$
- Source Domain과 Target Domain으로 구분 가능

Source Domain



≠

Target Domain



Introduction

Background

❖ Definition of Domain ($\mathcal{D}$)

- Domain은 2가지 요소로 이루어짐: (1) Feature Space $\mathcal{X}$, (2) 결합확률분포 $P(X)$, where $X = \{x | x_i \in \mathcal{X}, i = 1, \dots, n\}$
 - ✓ $\mathcal{D} = \{\mathcal{X}, P(X)\}$
- Source Domain과 Target Domain으로 구분 가능



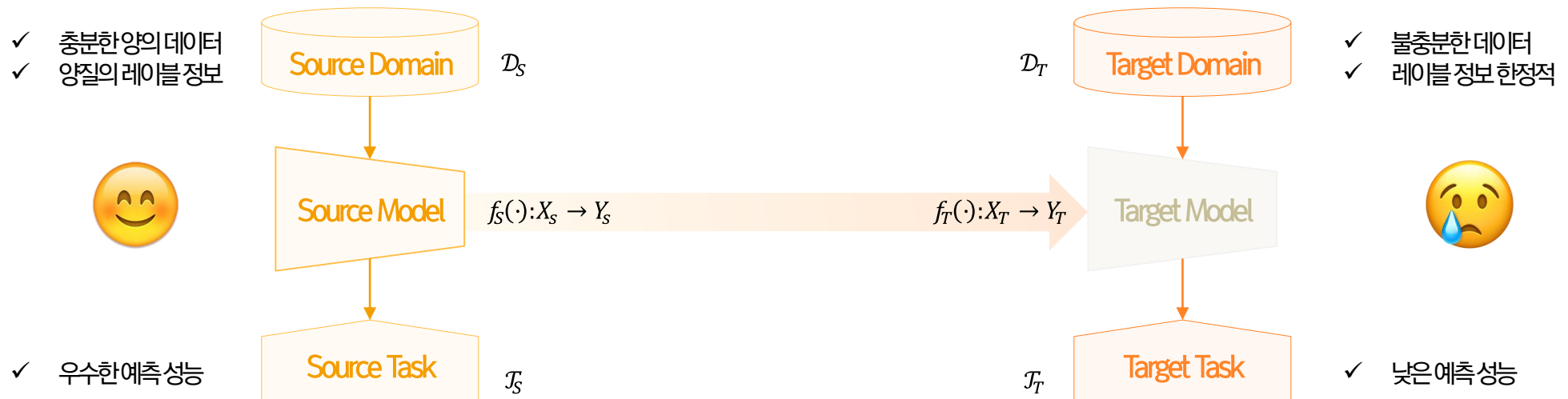
Introduction

Common Scenario of Deep Transfer Learning

❖ Why Does Transfer Learning Matter?

- 충분한 양의 레이블 된 Source Domain과 Source Task를 풀기 위해 학습된 모델만 있는 상황
- Target Domain은 Source Domain보다 관측치 수가 부족하고, 레이블이 없거나 한정적으로 있는 상황
- Domain이 다르거나 Task가 서로 달라서 Source Model을 직접적으로 활용하기 어렵고, Target Model의 예측 성능이 낮은 상황

Common Scenario of Deep Transfer Learning



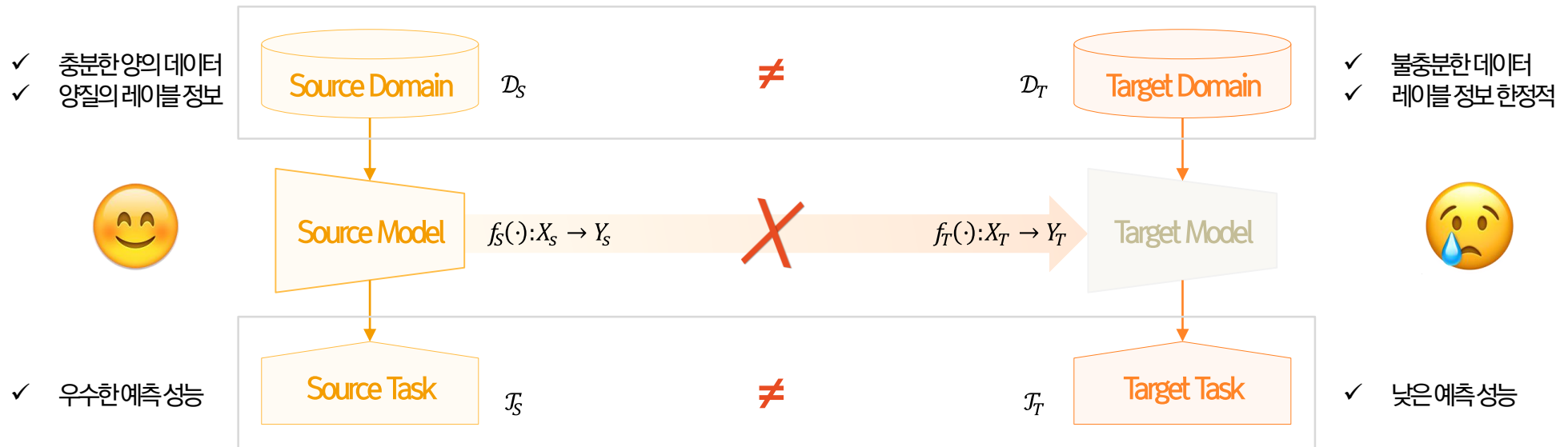
Introduction

Common Scenario of Deep Transfer Learning

❖ Why Does Transfer Learning Matter?

- 충분한 양의 레이블 된 Source Domain과 Source Task를 풀기 위해 학습된 모델만 있는 상황
- Target Domain은 Source Domain보다 관측치 수가 부족하고, 레이블이 없거나 한정적으로 있는 상황
- Domain이 다르거나 Task가 서로 달라서 Source Model을 직접적으로 활용하기 어렵고, Target Model의 예측 성능이 낮은 상황

Common Scenario of Deep Transfer Learning

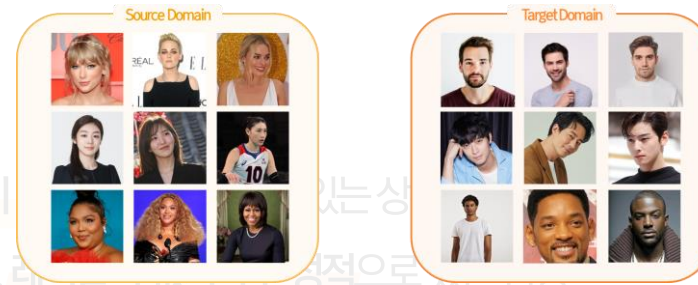


Introduction

Common Scenario of Deep Transfer Learning

❖ Why Does Transfer Learning Matter?

- 충분한 양의 레이블 된 Source Domain과 Source Task를 풀기
- Target Domain은 Source Domain보다 관측치 수가 부족하고, 레이블 정보가 한정적
- Source Domain과 Target Domain이 서로 달라서 Source Model을 직접적으로 활용하기 어렵고 Target Model의 예측 성능이 낮은 상황



전제

Source Domain과 Target Domain은 다르지만 유사

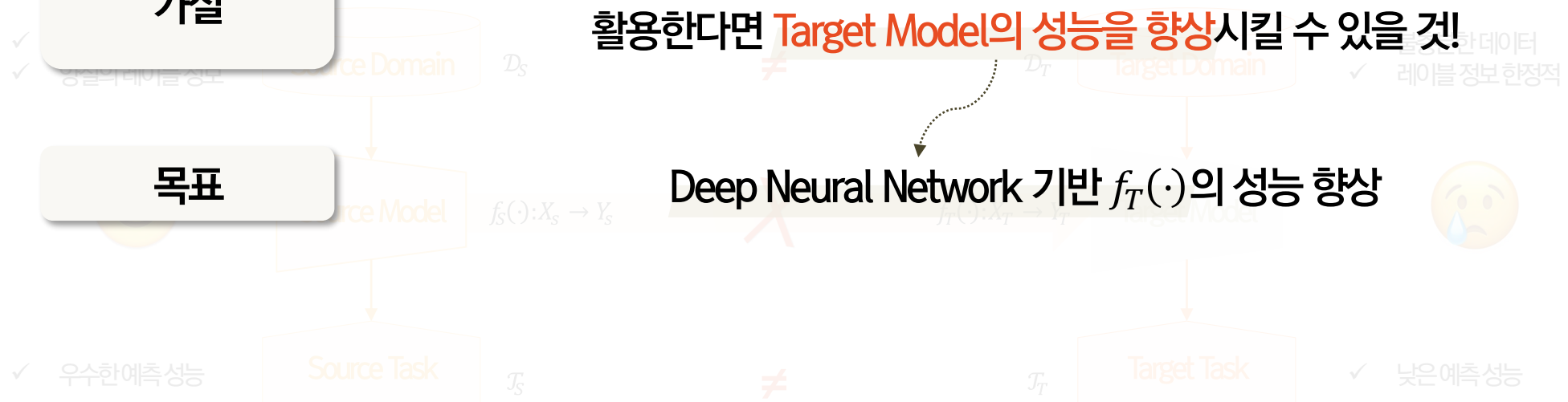
Common Scenario of Deep Transfer Learning

가설

Source Model의 지식을 일부 전이(Transfer)하여
활용한다면 Target Model의 성능을 향상시킬 수 있을 것!

목표

Deep Neural Network 기반 $f_T(\cdot)$ 의 성능 향상

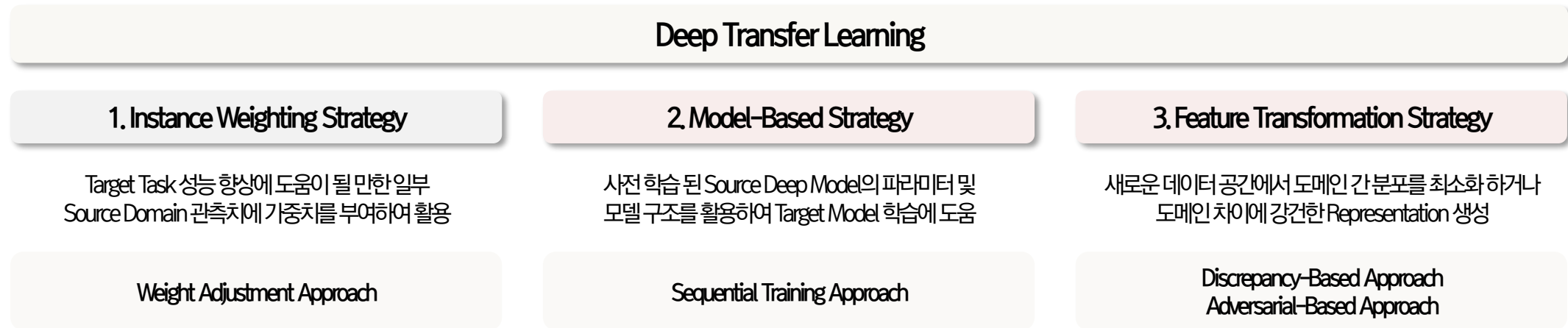


Introduction

Deep Transfer Learning

❖ How to Transfer Knowledge Across Domains by Deep Neural Network?

- Deep Transfer Learning은 방법론에 따라서 아래와 같이 3가지의 카테고리로 구분 가능



Introduction

Deep Transfer Learning

❖ How to Transfer Knowledge Across Domains by Deep Neural Network?

- Deep Transfer Learning은 방법론에 따라서 아래와 같이 3가지의 카테고리로 구분 가능 → 2가지 방법론 소개

Deep Transfer Learning

1. Instance Weighting Strategy

Target Task 성능 향상에 도움이 될 만한 일부 Source Domain 관측치에 가중치를 부여하여 활용

Weight Adjustment Approach

2. Model-Based Strategy

사전 학습된 Source Deep Model의 파라미터 및 모델 구조를 활용하여 Target Model 학습에 도움

Sequential Training Approach

3. Feature Transformation Strategy

새로운 데이터 공간에서 도메인 간 분포를 최소화하거나 도메인 차이에 강건한 Representation 생성

Discrepancy-Based Approach
Adversarial-Based Approach



2. Model-Based Strategy

Model-Based Strategy

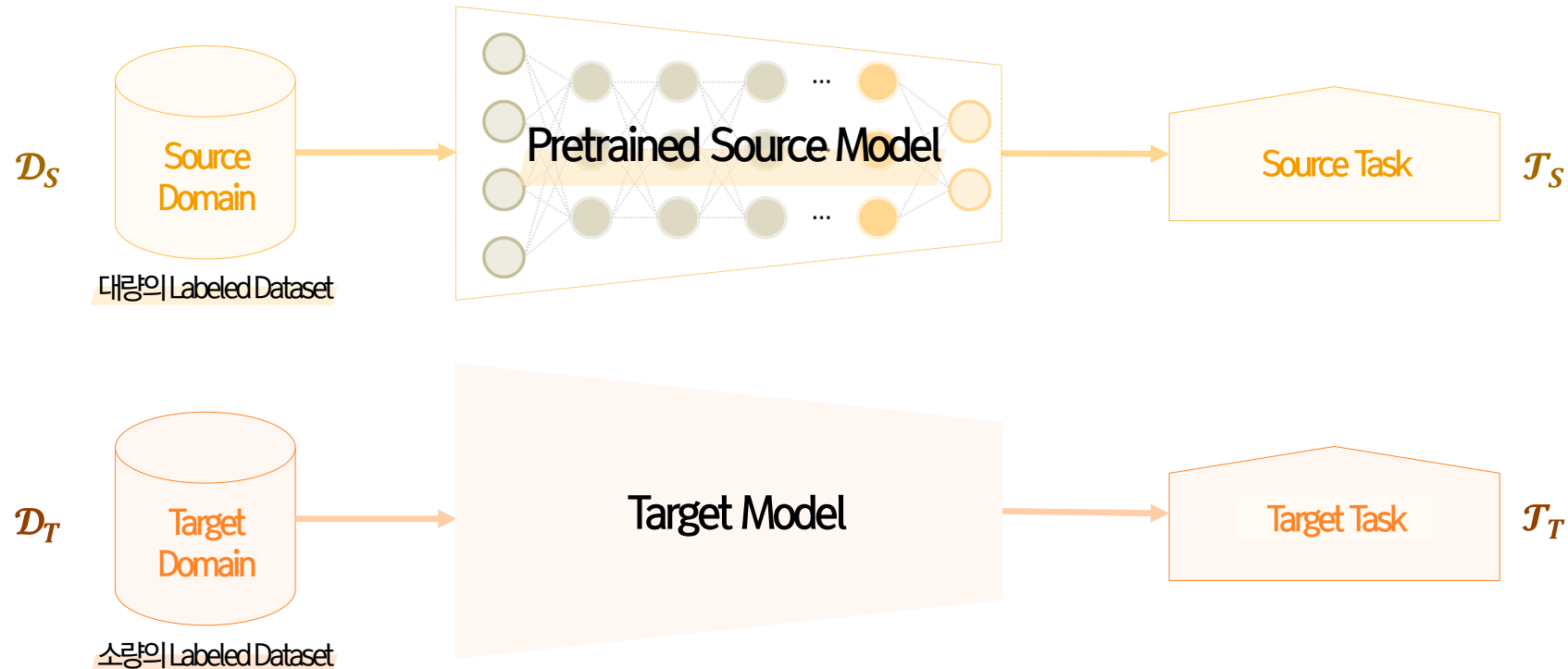
Sequential Training Approach

❖ Model-Based Strategy: Sequential Training Approach (1/2)

- 가설: Domain으로부터 얻은 지식은 모델 파라미터에 반영되며, 사전학습 시킨 Source Model의 구조와 파라미터를 활용하여 학습시키면 Target Model의 예측 성능이 향상될 것이다

Source Domain과 Target Domain은 서로 동일하지는 않아도 유사하기 때문

Sketch Map of Model-Based Strategy



Model-Based Strategy

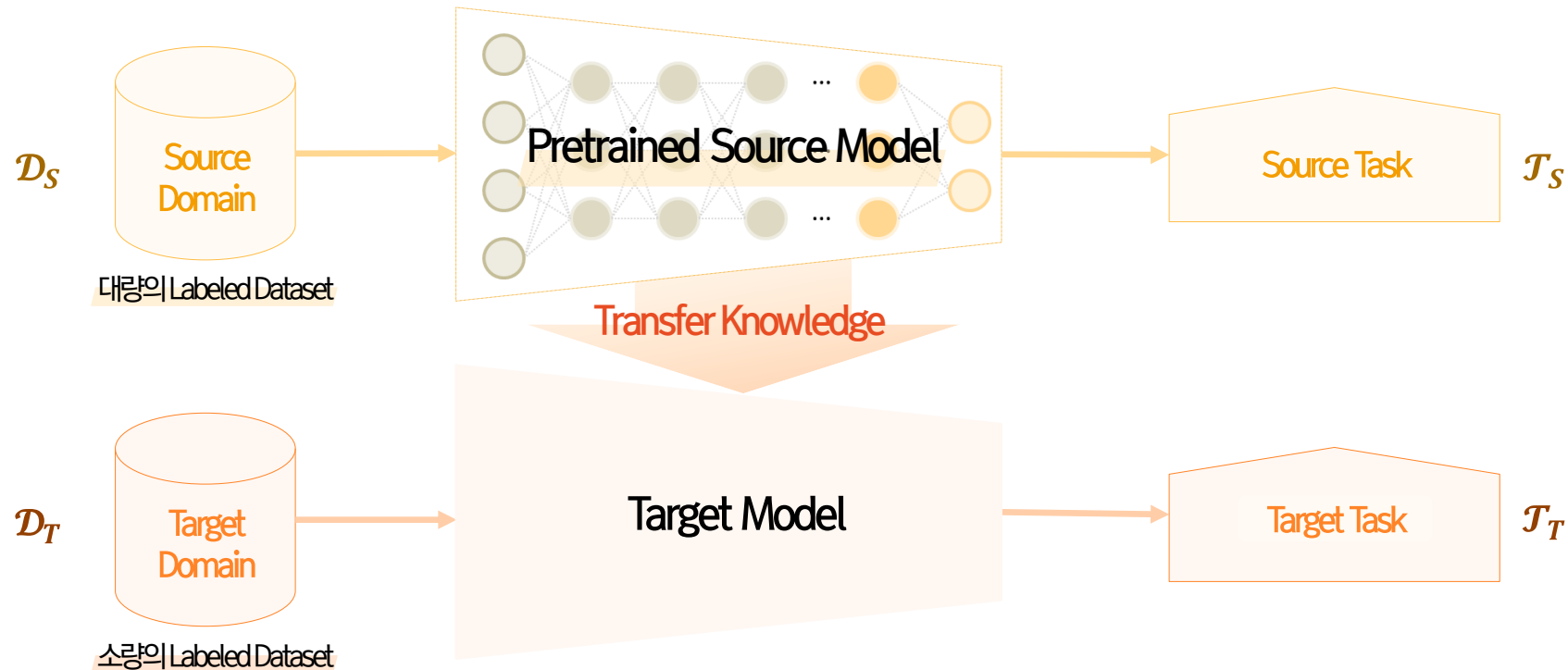
Sequential Training Approach

❖ Model-Based Strategy: Sequential Training Approach (1/2)

- 가설: Domain으로부터 얻은 지식은 모델 파라미터에 반영되며, 사전학습 시킨 Source Model의 구조와 파라미터를 활용하면 Target Model의 예측 성능이 향상될 것이다

Source Domain과 Target Domain은 서로 동일하지는 않아도 유사하기 때문

Sketch map of Model-Based Strategy



Model-Based Strategy

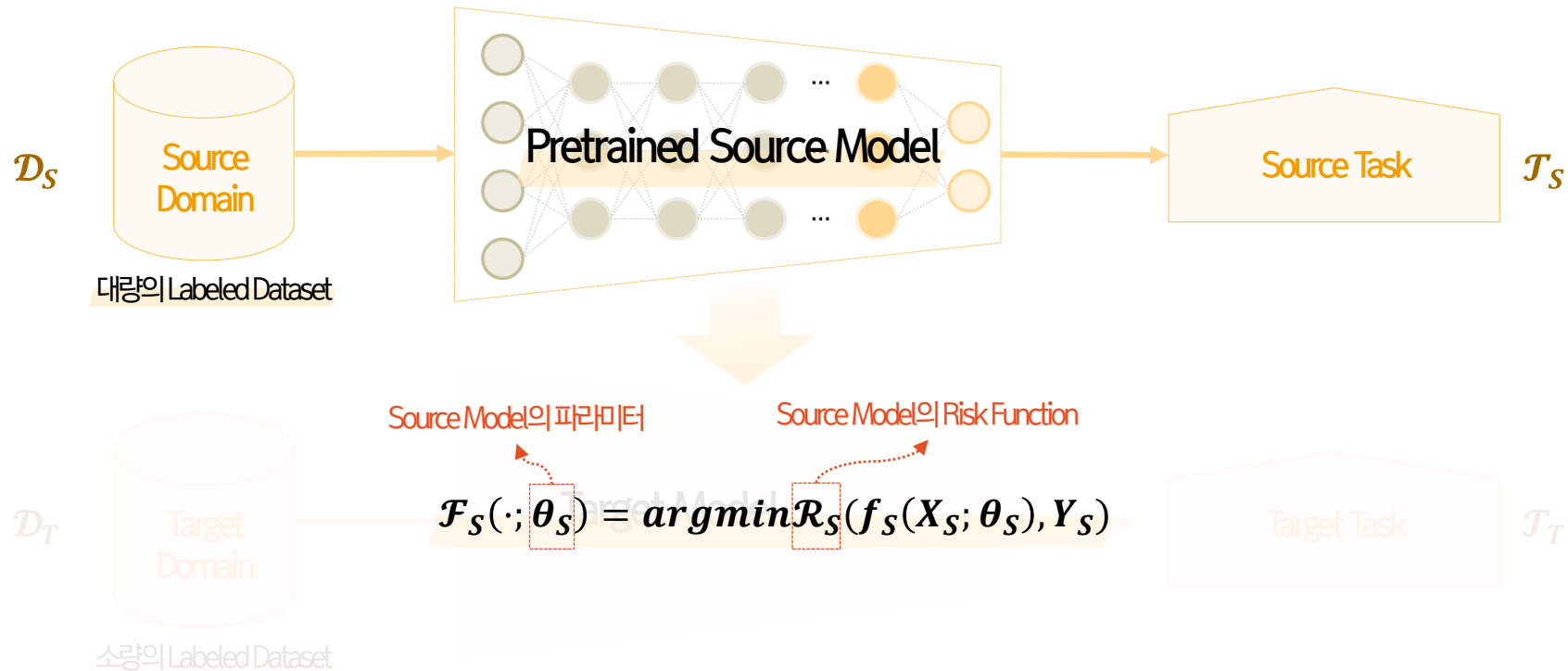
Sequential Training Approach

순차적 학습

❖ Model-Based Strategy: Sequential Training Approach (2/2)

- 방법: 1. 대량의 레이블 정보가 있는 Source Domain을 이용하여 Source Model 사전학습 수행
2. Source Model의 앞 부분 Layers는 파라미터를 Freeze하여 전이하고, 나머지 Layers는 Target Domain을 이용해 미세 조정(Fine Tuning)

Sequential Training Approach: Step 1



Model-Based Strategy

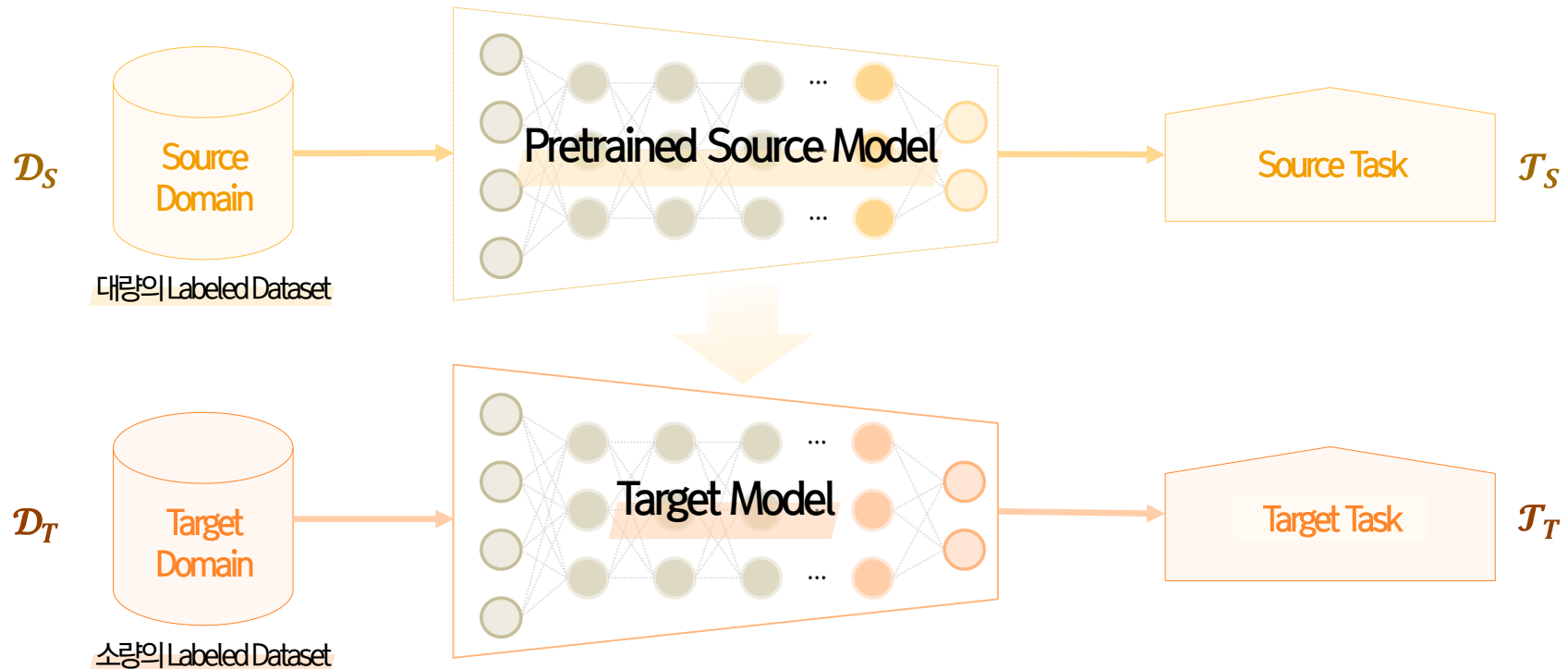
Sequential Training Approach

순차적 학습

❖ Model-Based Strategy: Sequential Training Approach (2/2)

- 방법: 1. 대량의 레이블 정보가 있는 Source Domain을 이용하여 Source Model 사전학습 수행
2. Source Model의 앞 부분 Layers는 파라미터를 Freeze하여 전이하고, 나머지 Layers는 Target Domain을 이용해 미세 조정(Fine Tuning)

Sequential Training Approach: Step 2



Model-Based Strategy

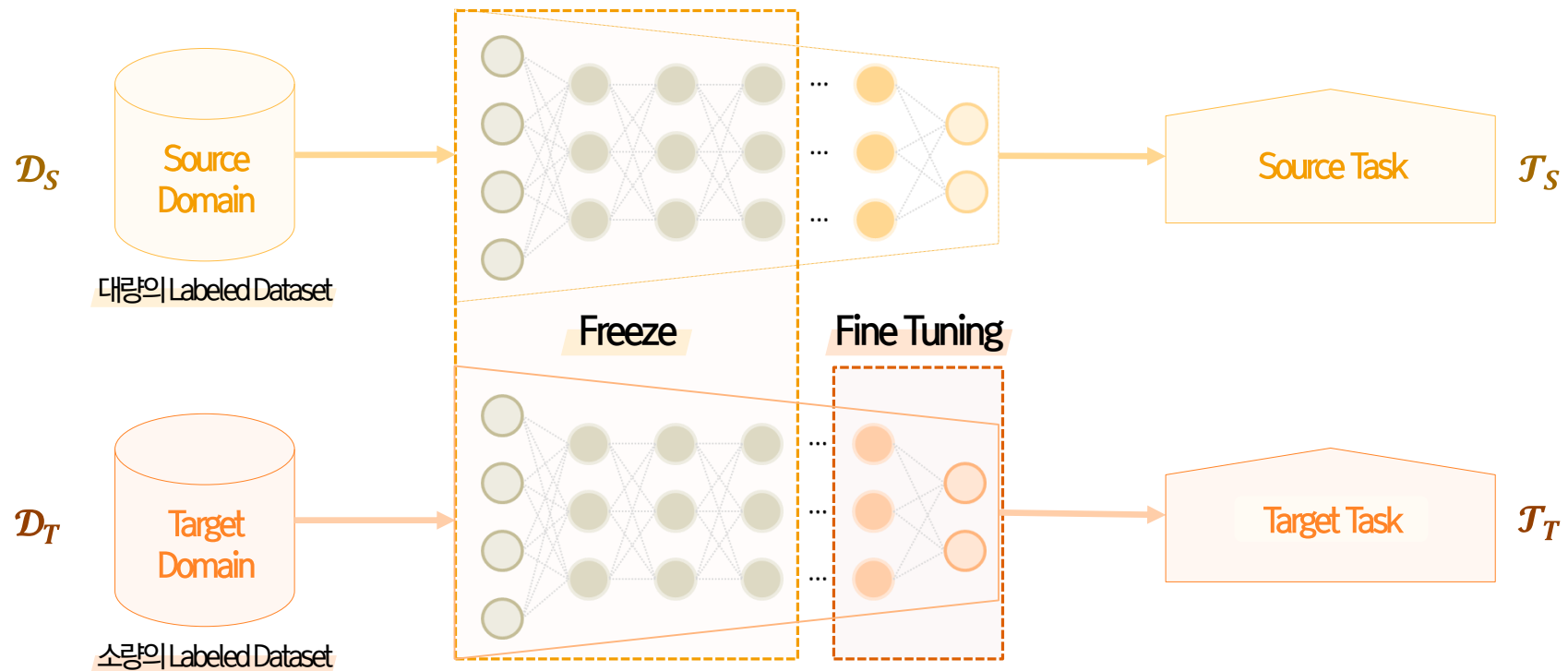
Sequential Training Approach

순차적 학습

❖ Model-Based Strategy: Sequential Training Approach (2/2)

- 방법: 1. 대량의 레이블 정보가 있는 Source Domain을 이용하여 Source Model 사전학습 수행
2. Source Model의 앞 부분 Layers는 파라미터를 Freeze하여 전이하고, 나머지 Layers는 Target Domain을 이용해 미세 조정(Fine Tuning)

Sequential Training Approach: Step 2



Model-Based Strategy

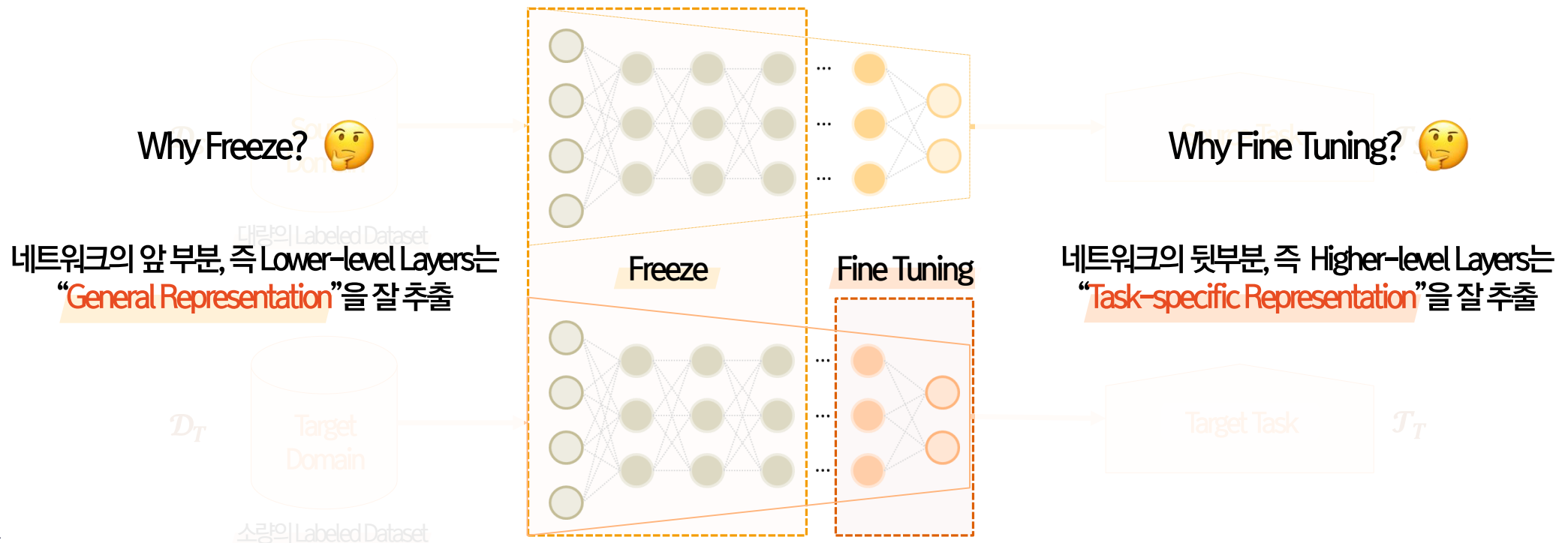
Sequential Training Approach

순차적 학습

❖ Model-Based Strategy: Sequential Training Approach (2/2)

- 방법: 1. 대량의 레이블 정보가 있는 Source Domain을 이용하여 Source Model 사전학습 수행
2. Source Model의 앞 부분 Layers는 파라미터를 Freeze하여 전이하고, 나머지 Layers는 Target Domain을 이용해 미세 조정(Fine Tuning)

Sequential Training Approach: Step 2



Model-Based Strategy

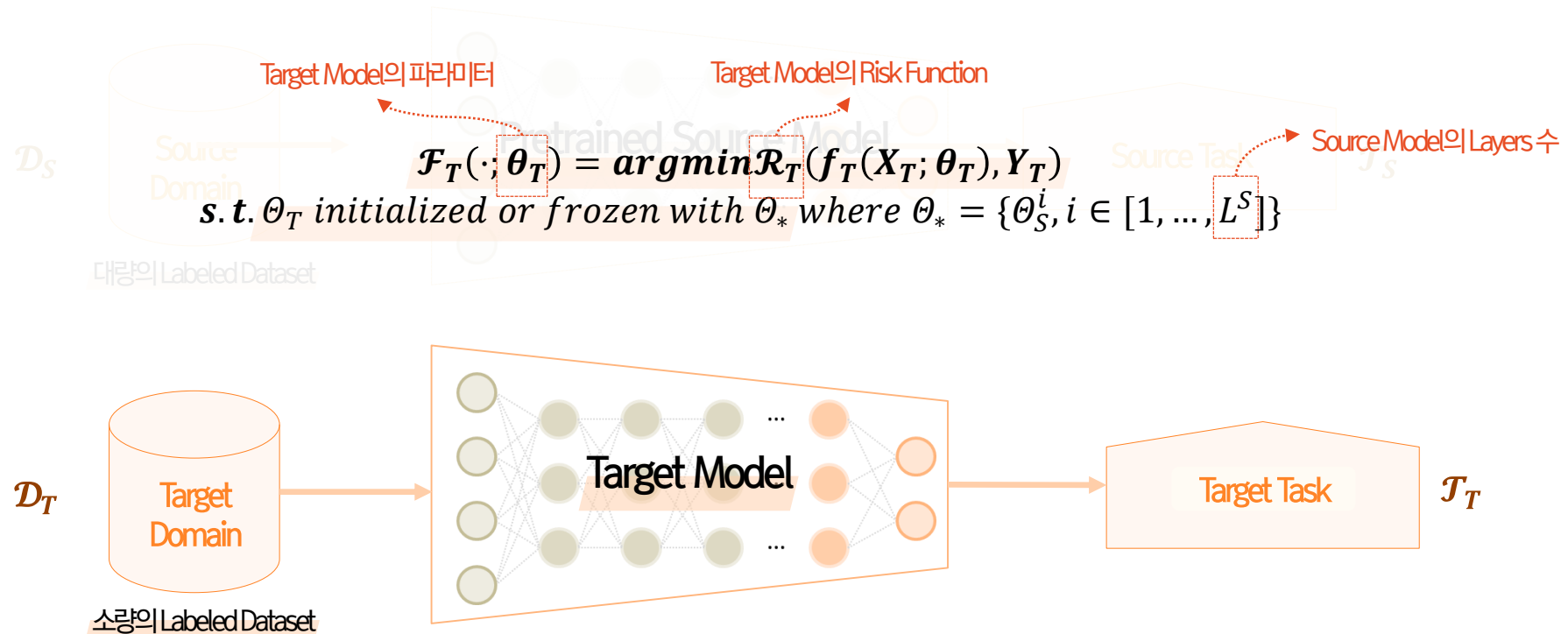
Sequential Training Approach

순차적 학습

❖ Model-Based Strategy: Sequential Training Approach (2/2)

- 방법: 1. 대량의 레이블 정보가 있는 Source Domain을 이용하여 Source Model 사전학습 수행
2. Source Model의 앞 부분 Layers는 파라미터를 Freeze하여 전이하고, 나머지 Layers는 Target Domain을 이용해 미세 조정(Fine Tuning)

Sequential Training Approach: Step 2



Model-Based Strategy

Sequential Training Approach

❖ How transferable are features in deep neural networks? (NeurIPS, 2014)

- 2022년 10월 기준 8079회 인용

Source Model의 구조 및 파라미터를 어떻게 활용해야 Target Model에서 좋은 성능이 나오는 지 분석

- 이미지 분류 문제에서 General Representation을 추출하는 Layer와 Task-specific Representation을 추출하는 Layer를 구분하기 위한 실험 진행

How transferable are features in deep neural networks?

Jason Yosinski,¹ Jeff Clune,² Yoshua Bengio,³ and Hod Lipson⁴

¹ Dept. Computer Science, Cornell University

² Dept. Computer Science, University of Wyoming

³ Dept. Computer Science & Operations Research, University of Montreal

⁴ Dept. Mechanical & Aerospace Engineering, Cornell University

Abstract

Many deep neural networks trained on natural images exhibit a curious phenomenon in common: on the first layer they learn features similar to Gabor filters and color blobs. Such first-layer features appear not to be *specific* to a particular dataset or task, but *general* in that they are applicable to many datasets and tasks. Features must eventually transition from general to specific by the last layer of the network, but this transition has not been studied extensively. In this paper we experimentally quantify the generality versus specificity of neurons in each layer of a deep convolutional neural network and report a few surprising results. Transferability is negatively affected by two distinct issues: (1) the specialization of higher layer neurons to their original task at the expense of performance on the target task, which was expected, and (2) optimization difficulties related to splitting networks between co-adapted neurons, which was not expected. In an example network trained on ImageNet, we demonstrate that either of these two issues may dominate, depending on whether features are transferred from the bottom, middle, or top of the network. We also document that the transferability of features decreases as the distance between the base task and target task increases, but that transferring features even from distant tasks can be better than using random features. A final surprising result is that initializing a network with transferred features from almost any number of layers can produce a boost to generalization that lingers even after fine-tuning to the target dataset.



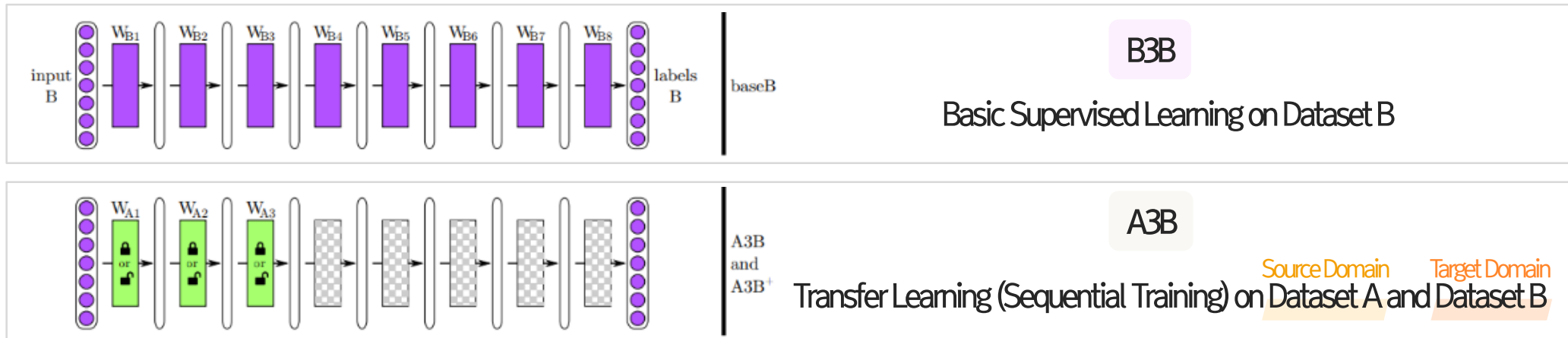
Model-Based Strategy

Sequential Training Approach

❖ How transferable are features in deep neural networks? (NeurIPS, 2014)

- 이미지 분류 문제에서 General Representation을 추출하는 Layer와 Task-specific Representation을 추출하는 Layer를 구분하기 위한 실험 진행
 - ✓ 7개의 레이어에 대해서 단계적으로 Freeze를 시켜간 결과, 첫 3개의 레이어가 General Representation을 얻을 수 있으며 다른 부분은 Fine Tuning 수행 필요
 - ✓ ImageNet과 같은 대용량 데이터셋에 대해 사전학습 된 Source Model의 General Representation을 이용한다면 Target Task에서 좋은 예측 성능을 낼 수 있음을 확인

AlexNet 구조 기반 Experiment



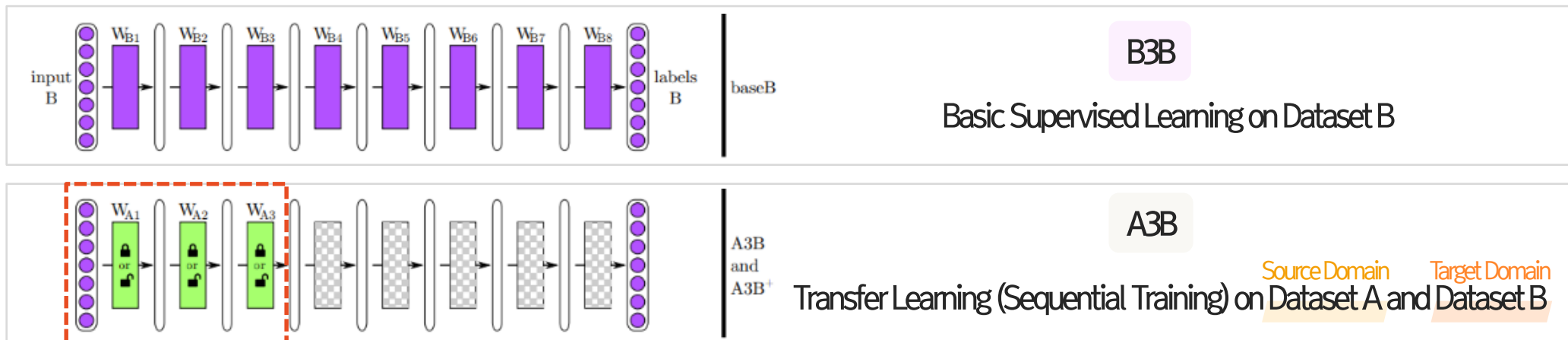
Model-Based Strategy

Sequential Training Approach

❖ How transferable are features in deep neural networks? (NeurIPS, 2014)

- 이미지 분류 문제에서 General Representation을 추출하는 Layer와 Task-specific Representation을 추출하는 Layer를 구분하기 위한 실험 진행
 - 7개의 레이어에 대해서 단계적으로 Freeze를 시켜간 결과, 첫 3개의 레이어가 General Representation을 얻을 수 있으며 다른 부분은 Fine Tuning 수행 필요
 - ImageNet과 같은 대용량 데이터셋에 대해 사전학습 된 Source Model의 General Representation을 이용한다면 Target Task에서 좋은 예측 성능을 낼 수 있음을 확인

AlexNet 구조 기반 Experiment



A3B가 B3B(Base B)만큼 성능을 낸다면 적어도 3번째 Layer까지는 B에 대해서는 General하다는 의미

3. Feature-Based Strategy

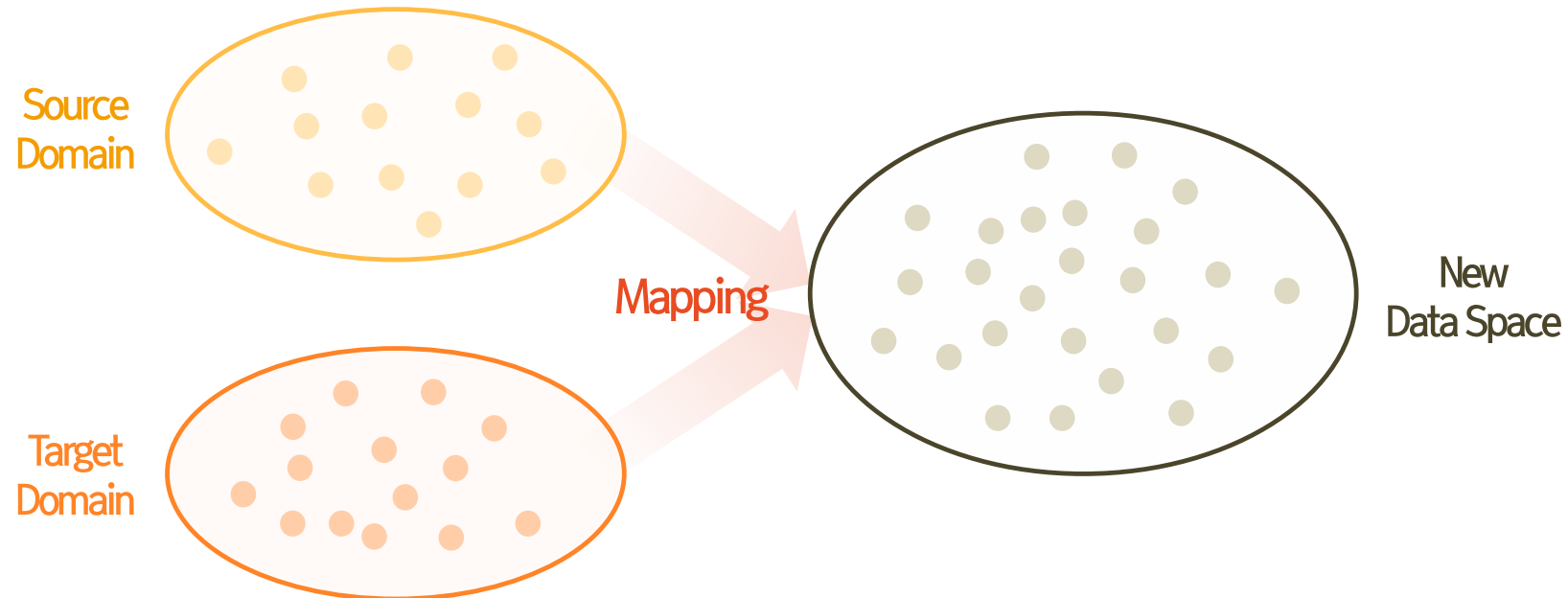
Feature-Based Strategy

Concept of Feature-Based Deep Transfer Learning

❖ Feature-Based Strategy

- 가설: Domain의 Feature들이 서로 다른 공간에 있더라도, 적절한 Mapping Function을 거쳐 나온 공간에서는 유사할 수 있다
 - ✓ Task에 대한 성능은 높이면서도 Domain 차이를 좁히는 Mapping Function 학습
- Discrepancy-Based와 Adversarial-Based로 나눌 수 있음

Sketch Map of Feature-Based Strategy

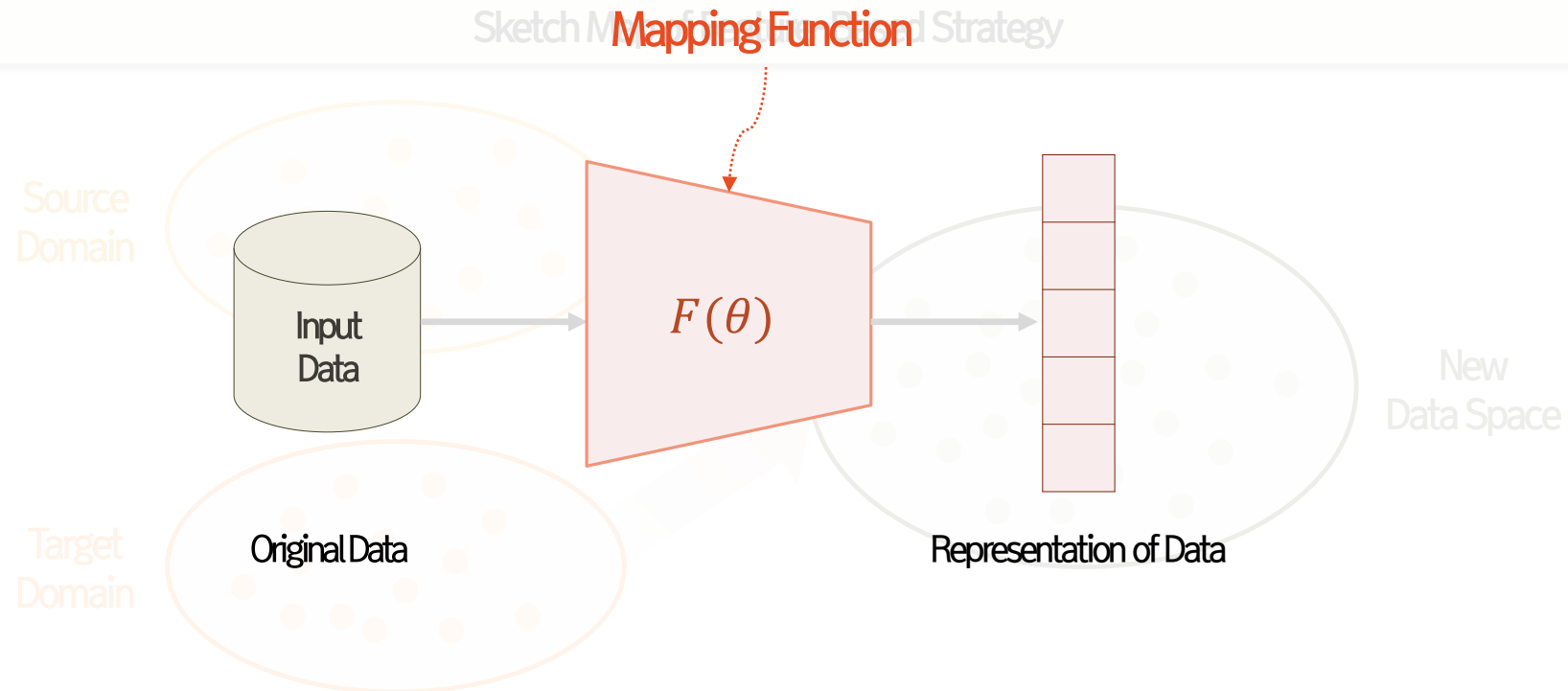


Feature-Based Strategy

Concept of Feature-Based Deep Transfer Learning

❖ Feature-Based Strategy

- 가설: Domain의 Feature들이 서로 다른 공간에 있더라도, 적절한 Mapping Function을 거쳐 나온 공간에서는 유사할 수 있다
 - ✓ Task에 대한 성능은 높이면서도 Domain 차이를 좁히는 Mapping Function 학습
- Discrepancy-Based와 Adversarial-Based로 나눌 수 있음



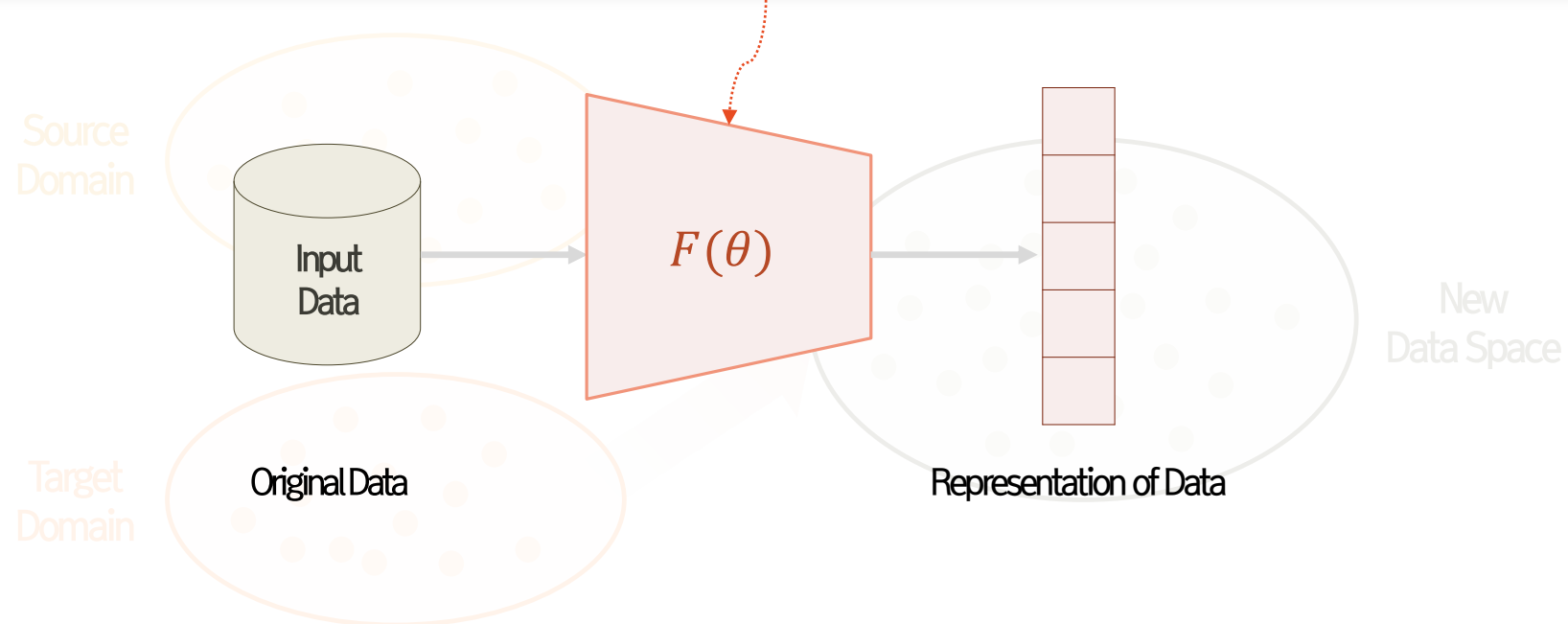
Feature-Based Strategy

Concept of Feature-Based Deep Transfer Learning

❖ Feature-Based Strategy

- 가설: Domain의 Feature들이 서로 다른 공간에 있더라도, 적절한 Mapping Function을 거쳐 나온 공간에서는 유사할 수 있다
 - ✓ Task에 대한 성능은 높이면서도 Domain 차이를 줄이는 Mapping Function 학습
- Discrepancy-Based와 Adversarial-Based로 나눌 수 있음

Sketch Mapping Function ▶ Domain Adaptation !



Feature-Based Strategy

Concept of Feature-Based Deep Transfer Learning

❖ Feature-Based Strategy

- 가설: Domain의 Feature들이 서로 다른 공간에 있더라도 적절한 mapping function을 거쳐 나온 공간에서는 유사할 수 있다

✓ Task에 대한 성능은 높으면서도 Domain 차이를 좁히는 Mapping Function 학습

- Discrepancy

참고: Domain Adaptation 관련 세미나

종료

Unsupervised domain adaptation Semi-supervised domain adaptation

도메인 적응
Domain adaptation

Source domain Target domain

2023.09.16 (Sat) 14:00 ~ 16:00

Introduction to unsupervised domain adaptation

발표자:  이민정

📅 2022년 9월 16일
🕒 오후 1시 ~
📺 온라인 비디오 시청 (YouTube)

세미나 정보 보기 →

종료 **Deep Domain Adaptation**

DMQA Open Seminar
2022.08.05

Deep Domain Adaptation

발표자:  김태연

📅 2022년 8월 5일
🕒 오후 1시 ~
📺 온라인 비디오 시청 (YouTube)

세미나 정보 보기 →

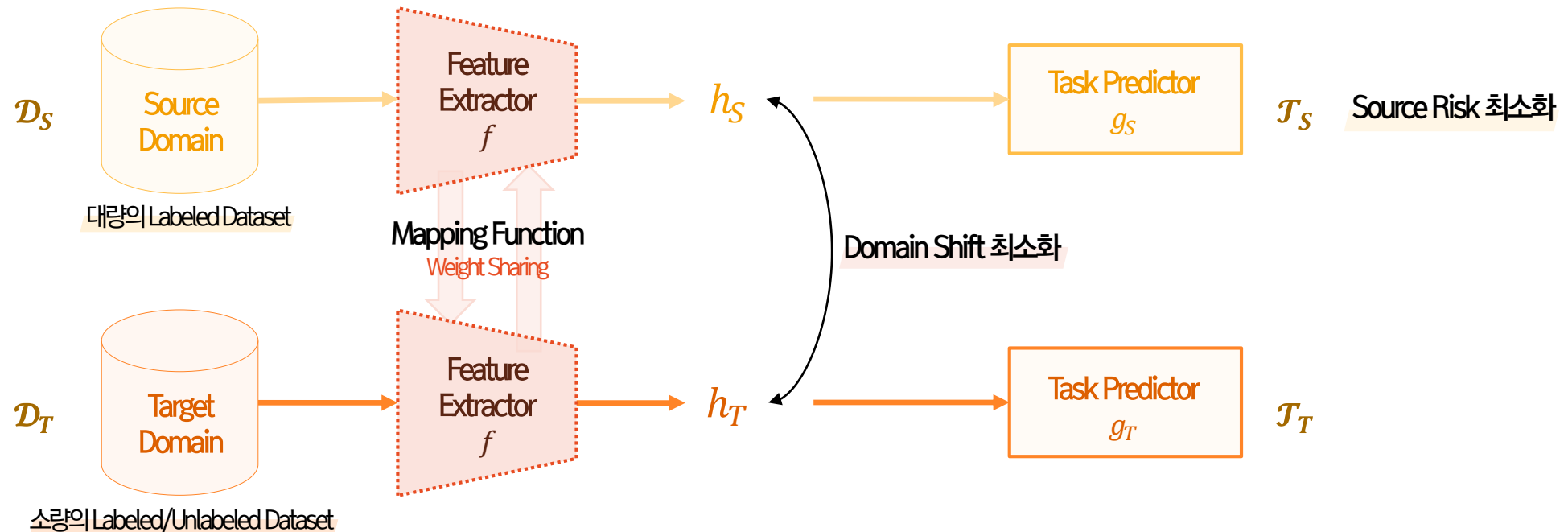
Feature-Based Strategy

Discrepancy-Based Approach

❖ Feature-Based Strategy: (1) Discrepancy-Based Approach

- 가설: Mapping Function을 통해 Task에 대한 Risk가 낮고 Domain 간 차이에 강건한 Representation을 학습할 수 있다면, Source Model의 지식을 활용하여 Target Model의 예측 성능을 향상시킬 수 있을 것이다

Sketch Map of Discrepancy-Based Approach



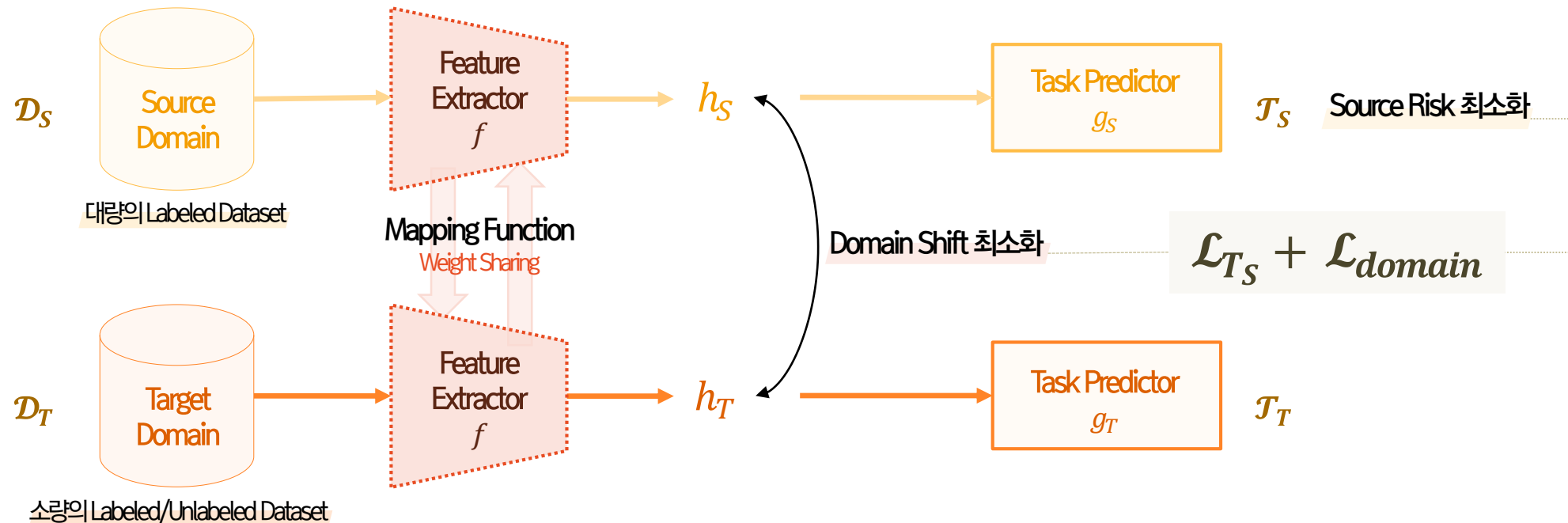
Feature-Based Strategy

Discrepancy-Based Approach

❖ Feature-Based Strategy: (1) Discrepancy-Based Approach

- 가설: Mapping Function을 통해 Task에 대한 Risk가 낮고 Domain 간 차이에 강건한 Representation을 학습할 수 있다면, Source Model의 지식을 활용하여 Target Model의 예측 성능을 향상시킬 수 있을 것이다

Sketch Map of Discrepancy-Based Approach



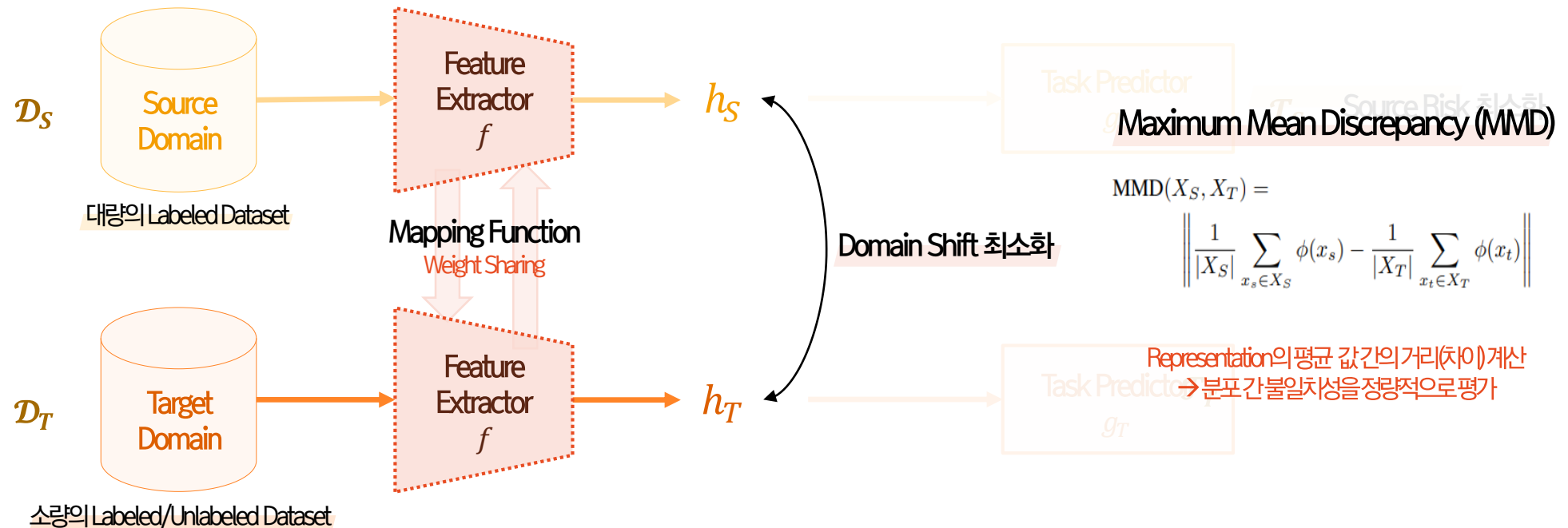
Feature-Based Strategy

Discrepancy-Based Approach

❖ Feature-Based Strategy: (1) Discrepancy-Based Approach

- Domain Shift 최소화 방법: 서로 다른 Domain 간 Representation 사이의 **거리**를 계산하여 줄여 나감
MMD, KL-Divergence, MK-MMD, Correlation Alignment, Wasserstein Distance

Sketch Map of Discrepancy-Based Approach



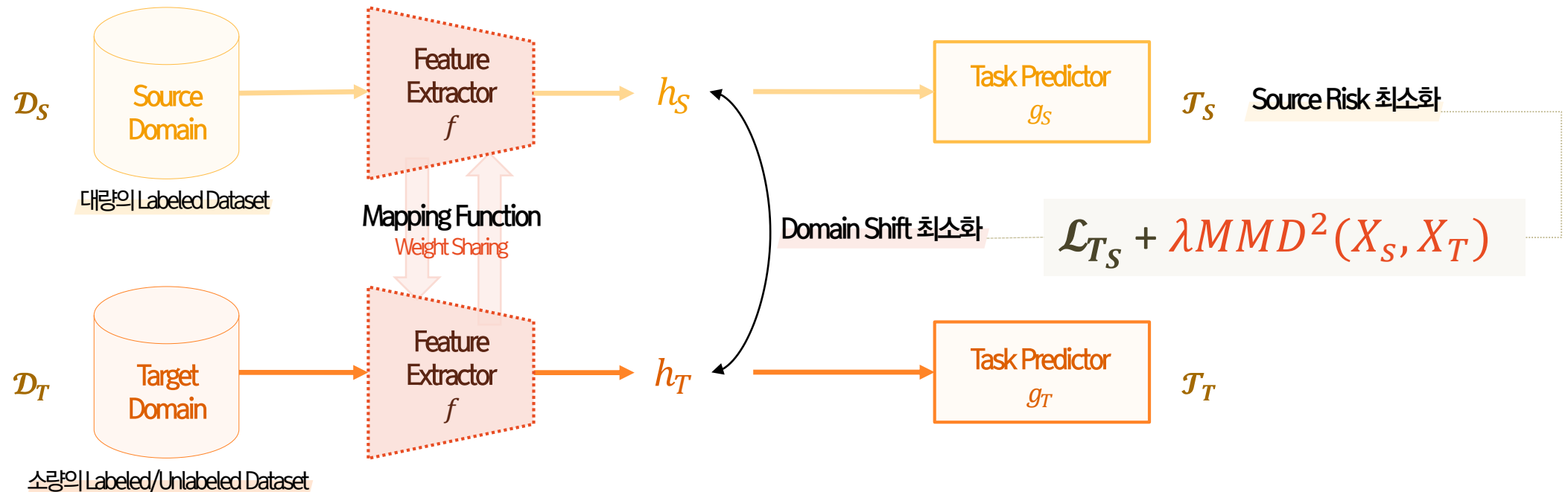
Feature-Based Strategy

Discrepancy-Based Approach

❖ Feature-Based Strategy: (1) Discrepancy-Based Approach

- Domain Shift 최소화 방법: 서로 다른 Domain 간 Representation 사이의 **거리**를 계산하여 줄여 나감
MMD, KL-Divergence, MK-MMD, Correlation Alignment, Wasserstein Distance

Sketch Map of Discrepancy-Based Approach



Feature-Based Strategy

Discrepancy-Based Approach

❖ Deep Domain Confusion: Maximizing for Domain Invariance (arXiv, 2014)

- 2022년 10월 기준 2058회 인용
- Domain 간 거리(차이)를 MMD를 통해 계산하고, Source Classifier에 대한 Risk를 최소화 하여 최종적으로 Target Task의 예측 성능을 향상
 - ✓ Mapping Function으로서 딥러닝(CNN)을 이용하며 괄목할 만한 성능을 도출

Deep Domain Confusion: Maximizing for Domain Invariance

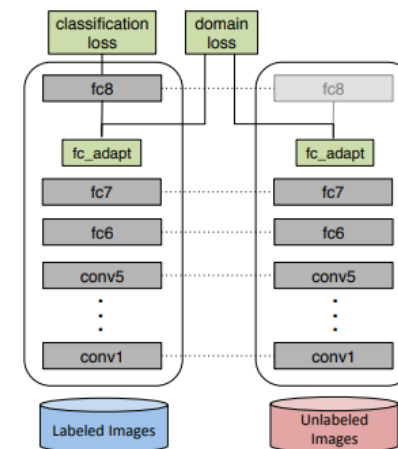
Eric Tzeng, Judy Hoffman, Ning Zhang
UC Berkeley, EECS & ICSI
{etzeng, jhoffman, nzhang}@eecs.berkeley.edu

Kate Saenko
UMass Lowell, CS
saenko@cs.uml.edu

Trevor Darrell
UC Berkeley, EECS & ICSI
trevor@eecs.berkeley.edu

Abstract

Recent reports suggest that a generic supervised deep CNN model trained on a large-scale dataset reduces, but does not remove, dataset bias on a standard benchmark. Fine-tuning deep models in a new domain can require a significant amount of data, which for many applications is simply not available. We propose a new CNN architecture which introduces an adaptation layer and an additional domain confusion loss, to learn a representation that is both semantically meaningful and domain invariant. We additionally show that a domain confusion metric can be used for model selection to determine the dimension of an adaptation layer and the best position for the layer in the CNN architecture. Our proposed adaptation method offers empirical performance which exceeds previously published results on a standard benchmark visual domain adaptation task.

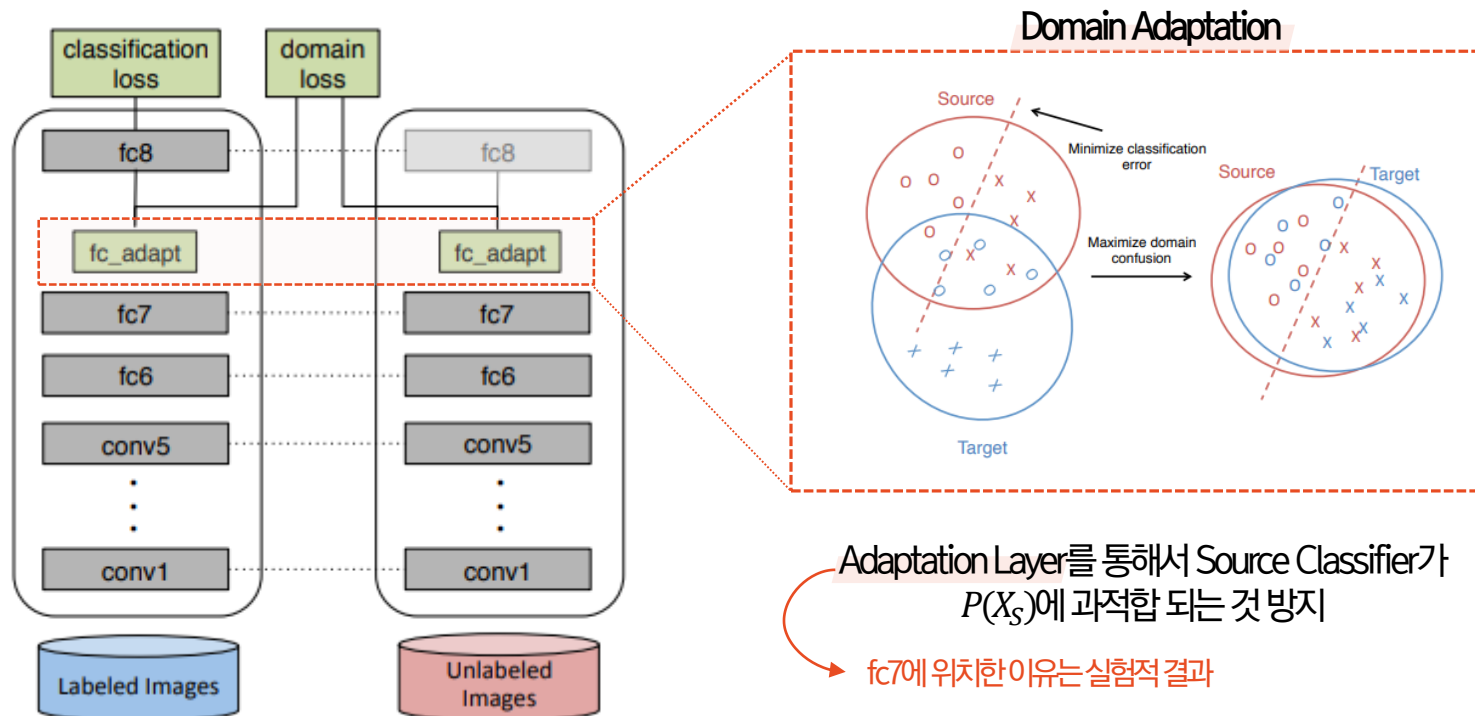


Feature-Based Strategy

Discrepancy-Based Approach

❖ Deep Domain Confusion: Maximizing for Domain Invariance (arXiv, 2014)

- 2022년 10월 기준 2058회 인용
- Domain 간 거리(차이)를 MMD를 통해 계산하고, Source Classifier에 대한 Risk를 최소화 하여 최종적으로 Target Task의 예측 성능을 향상
 - ✓ Mapping Function으로서 딥러닝(CNN)을 이용하여 괄목할 만한 성능을 도출

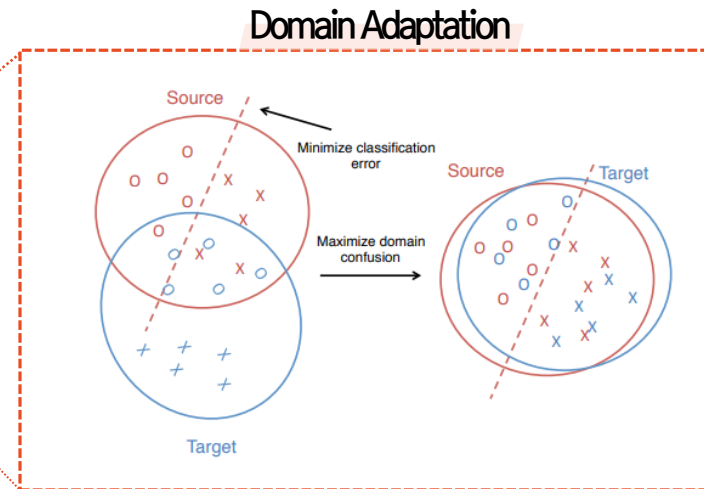
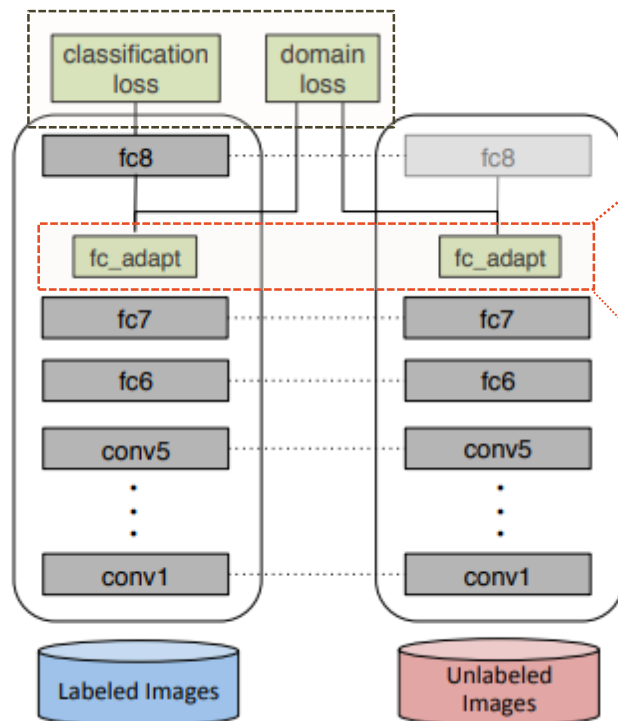


Feature-Based Strategy

Discrepancy-Based Approach

❖ Deep Domain Confusion: Maximizing for Domain Invariance (arXiv, 2014)

- 2022년 10월 기준 2058회 인용
- Domain 간 거리(차이)를 MMD를 통해 계산하고, Source Classifier에 대한 Risk를 최소화 하여 최종적으로 Target Task의 예측 성능을 향상
 - ✓ Mapping Function으로서 딥러닝(CNN)을 이용하여 괄목할 만한 성능을 도출



Adaptation Layer를 통해서 Source Classifier가 $P(X_S)$ 에 과적합 되는 것 방지

fc7에 위치한 이유는 실험적 결과

Final Loss

$$\mathcal{L} = \mathcal{L}_C(X_L, y) + \lambda \text{MMD}^2(X_S, X_T)$$
$$\text{MMD}(X_S, X_T) = \left\| \frac{1}{|X_S|} \sum_{x_s \in X_S} \phi(x_s) - \frac{1}{|X_T|} \sum_{x_t \in X_T} \phi(x_t) \right\|$$

λ 는 하이퍼파라미터로, Domain 조정 정도를 반영

Feature-Based Strategy

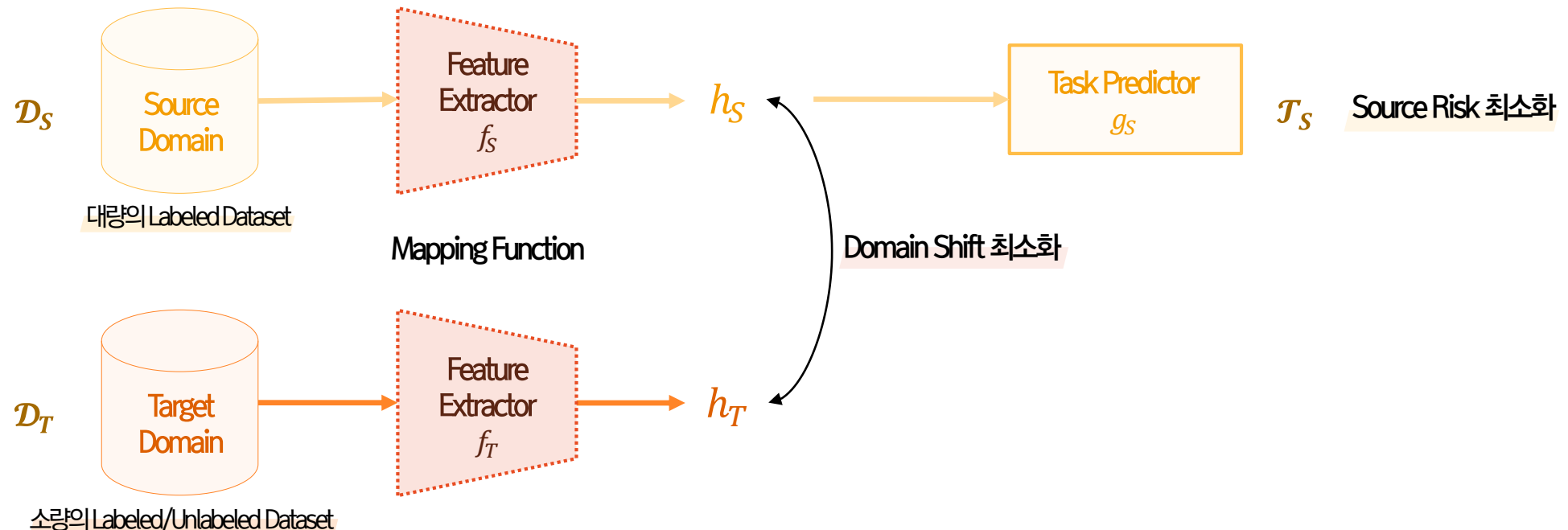
Adversarial-Based Approach

네트워크들이 서로 반대의 목적을 가지고 학습

❖ Feature-Based Strategy: (2) Adversarial-Based Approach

- 가설: Feature Extractor와 Domain 구분자(Discriminator)가 서로 적대적으로 학습한다면, Domain 차이를 구분하지 못하면서도 Target Task에 대한 예측 성능을 높이는 Representation을 산출할 수 있을 것이다
- Generative Adversarial Networks(GANs) 방법론에 기반한 알고리즘

Sketch Map of Adversarial-Based Approach



Feature-Based Strategy

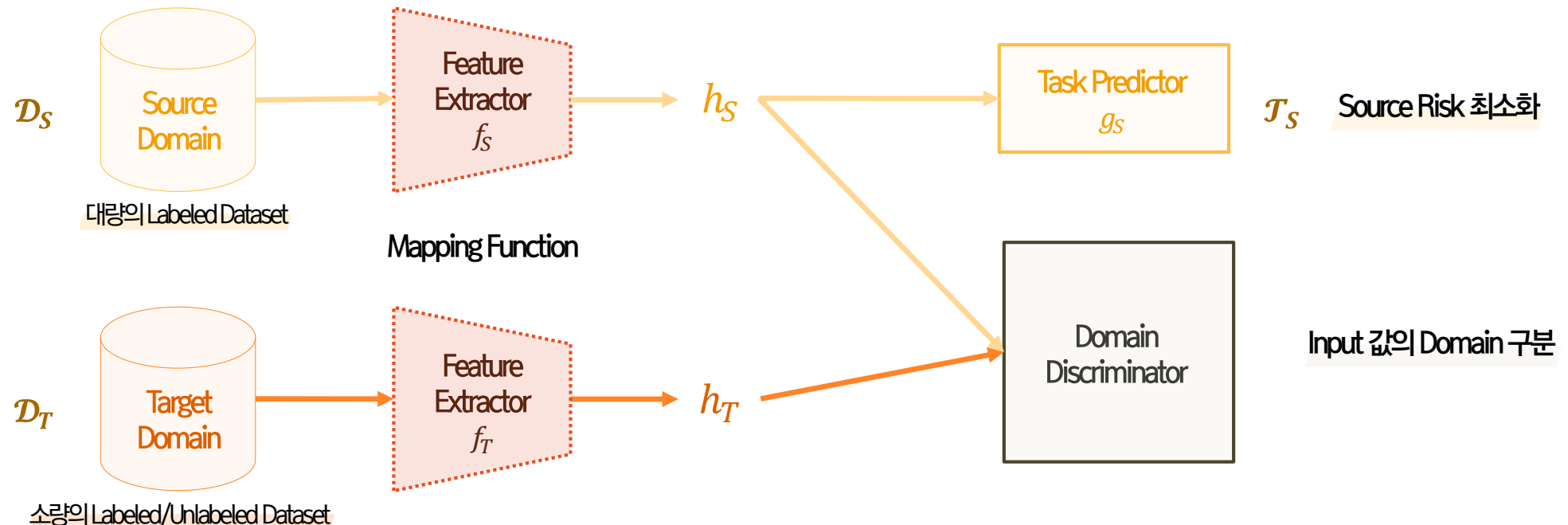
Adversarial-Based Approach

네트워크들이 서로 반대의 목적을 가지고 학습

❖ Feature-Based Strategy: (2) Adversarial-Based Approach

- 가설: Feature Extractor와 Domain 구분자(Discriminator)가 서로 적대적으로 학습한다면, Domain 차이를 구분하지 못하면서도 Target Task에 대한 예측 성능을 높이는 Representation을 산출할 수 있을 것이다
- Generative Adversarial Networks(GANs) 방법론에 기반한 알고리즘

Sketch Map of Adversarial-Based Approach



Feature-Based Strategy

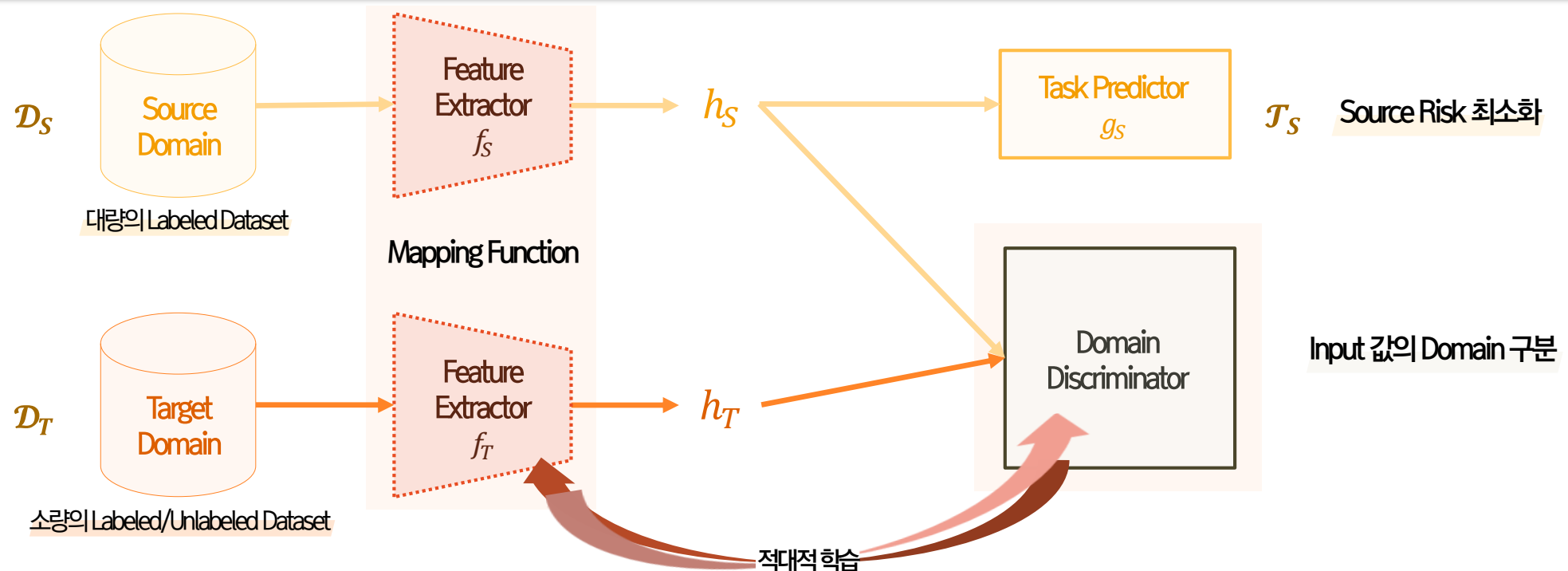
Adversarial-Based Approach

네트워크들이 서로 반대의 목적을 가지고 학습

❖ Feature-Based Strategy: (2) Adversarial-Based Approach

- 가설: Feature Extractor와 Domain 구분자(Discriminator)가 서로 적대적으로 학습한다면, Domain 차이를 구분하지 못하면서도 Target Task에 대한 예측 성능을 높이는 Representation을 산출할 수 있을 것이다
- Generative Adversarial Networks(GANs) 방법론에 기반한 알고리즘

Sketch Map of Adversarial-Based Approach



Feature-Based Strategy

Adversarial-Based Approach

네트워크들이 서로 반대의 목적을 가지고 학습

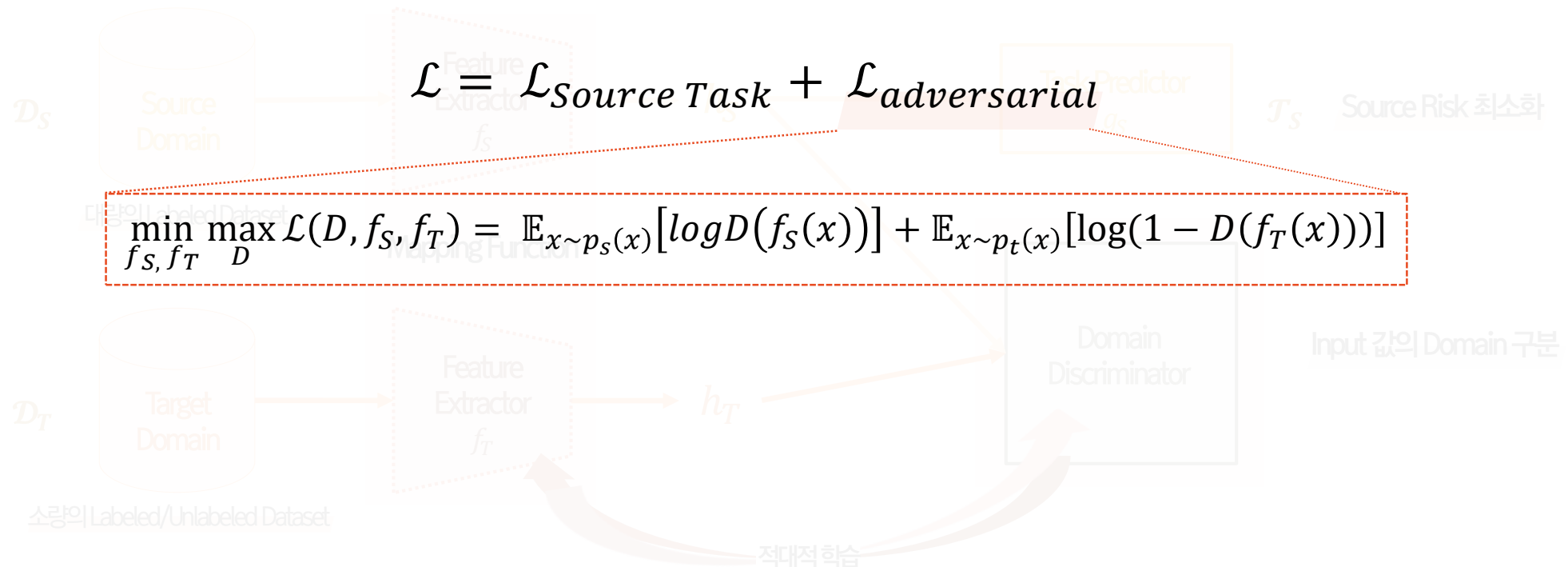
❖ Feature-Based Strategy: (2) Adversarial-Based Approach

- 가설: Feature Extractor와 Domain 구분자(Discriminator)가 서로 적대적으로 학습한다면, Domain 차이를 구분하지 못하면서도 Target Task에 대한 예측 성능을 높이는 Representation을 산출할 수 있을 것이다
- Generative Adversarial Networks(GANs) 방법론에 기반한 알고리즘

Loss Function of Adversarial-Based Approach

$$\mathcal{L} = \mathcal{L}_{Source\ Task} + \mathcal{L}_{adversarial}$$

$$\min_{f_S, f_T} \max_D \mathcal{L}(D, f_S, f_T) = \mathbb{E}_{x \sim p_S(x)} [\log D(f_S(x))] + \mathbb{E}_{x \sim p_T(x)} [\log(1 - D(f_T(x)))]$$



Feature-Based Strategy

Adversarial-Based Approach

네트워크들이 서로 반대의 목적을 가지고 학습

❖ Feature-Based Strategy: (2) Adversarial-Based Approach

- 가설: Feature Extractor와 Domain 구분자(Discriminator)가 서로 적대적으로 학습한다면, Domain 차이를 구분하지 못하면서도 Target Task에 대한 예측 성능을 높이는 Representation을 산출할 수 있을 것이다
- Generative Adversarial Networks(GANs) 방법론에 기반한 알고리즘

Loss Function of Adversarial-Based Approach

$$\mathcal{L} = \mathcal{L}_{Source\ Task} + \mathcal{L}_{adversarial}$$

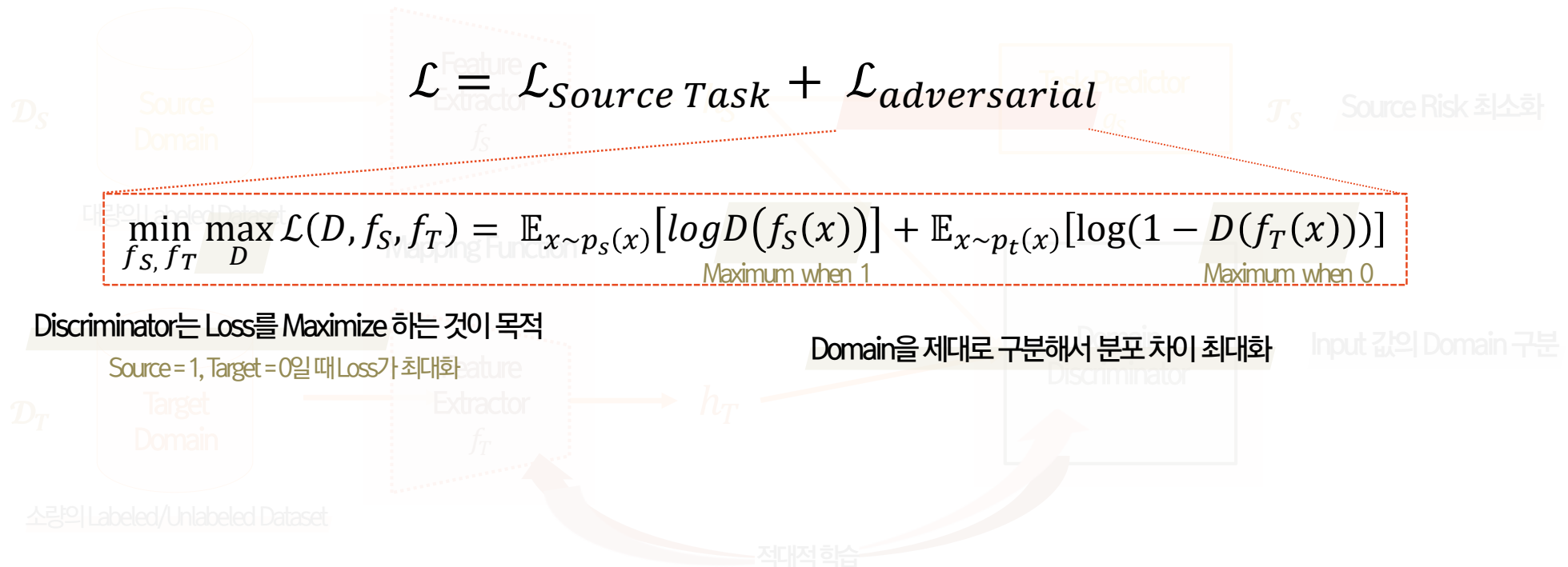
$$\min_{f_S, f_T} \max_D \mathcal{L}(D, f_S, f_T) = \mathbb{E}_{x \sim p_S(x)} [\log D(f_S(x))] + \mathbb{E}_{x \sim p_T(x)} [\log(1 - D(f_T(x)))]$$

Maximum when 1 Maximum when 0

Discriminator는 Loss를 Maximize 하는 것이 목적

Source = 1, Target = 0일 때 Loss가 최대화

Domain을 제대로 구분해서 분포 차이 최대화



Feature-Based Strategy

Adversarial-Based Approach

네트워크들이 서로 반대의 목적을 가지고 학습

❖ Feature-Based Strategy: (2) Adversarial-Based Approach

- 가설: Feature Extractor와 Domain 구분자(Discriminator)가 서로 적대적으로 학습한다면, Domain 차이를 구분하지 못하면서도 Target Task에 대한 예측 성능을 높이는 Representation을 산출할 수 있을 것이다
- Generative Adversarial Networks(GANs) 방법론에 기반한 알고리즘

Loss Function of Adversarial-Based Approach

$$\mathcal{L} = \mathcal{L}_{Source\ Task} + \mathcal{L}_{adversarial}$$

$$\min_{f_S, f_T} \max_D \mathcal{L}(D, f_S, f_T) = \mathbb{E}_{x \sim p_S(x)} [\log D(f_S(x))] + \mathbb{E}_{x \sim p_T(x)} [\log(1 - D(f_T(x)))]$$

Minimum when 1

Feature Extractor는 Loss를 Minimize 하는 것이 목적

Target = 1일 때 Loss가 최소화

Domain을 제대로 구분하지 못하게 만들어서 분포 차이 최소화

소량의 Labeled/Unlabeled Dataset

적대적 학습

Feature-Based Strategy

Adversarial-Based Approach

❖ Adversarial Discriminative Domain Adaptation (CVPR, 2017)

- 2022년 10월 기준 3666회 인용
- **적대적 학습 방식으로 Domain 간의 차이를 조정하고 Source Model의 지식을 이용해 Target Task에 대한 성능을 높임**

Adversarial Discriminative Domain Adaptation

Eric Tzeng
University of California, Berkeley
etzeng@eecs.berkeley.edu

Judy Hoffman
Stanford University
jhoffman@cs.stanford.edu

Kate Saenko
Boston University
saenko@bu.edu

Trevor Darrell
University of California, Berkeley
trevor@eecs.berkeley.edu

Abstract

Adversarial learning methods are a promising approach to training robust deep networks, and can generate complex samples across diverse domains. They can also improve recognition despite the presence of domain shift or dataset bias: recent adversarial approaches to unsupervised domain adaptation reduce the difference between the training and test domain distributions and thus improve generalization performance. However, while generative adversarial networks (GANs) show compelling visualizations, they are not optimal on discriminative tasks and can be limited to smaller shifts. On the other hand, discriminative approaches can handle larger domain shifts, but impose tied weights on the model and do not exploit a GAN-based loss. In this work, we first outline a novel generalized framework for adversarial adaptation, which subsumes recent state-of-the-art approaches as special cases, and use this generalized view to better relate prior approaches. We then propose a previously unexplored instance of our general framework which combines discriminative modeling, untied weight sharing, and a GAN loss, which we call Adversarial Discriminative Domain Adaptation (ADDA). We show that ADDA is more effective yet considerably simpler than competing domain-adversarial methods, and demonstrate the promise of our approach by exceeding state-of-the-art unsupervised adaptation results on standard domain adaptation tasks as well as a difficult cross-modality object classification task.

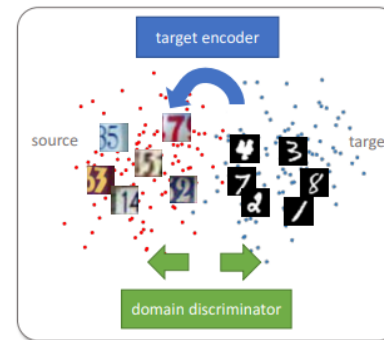


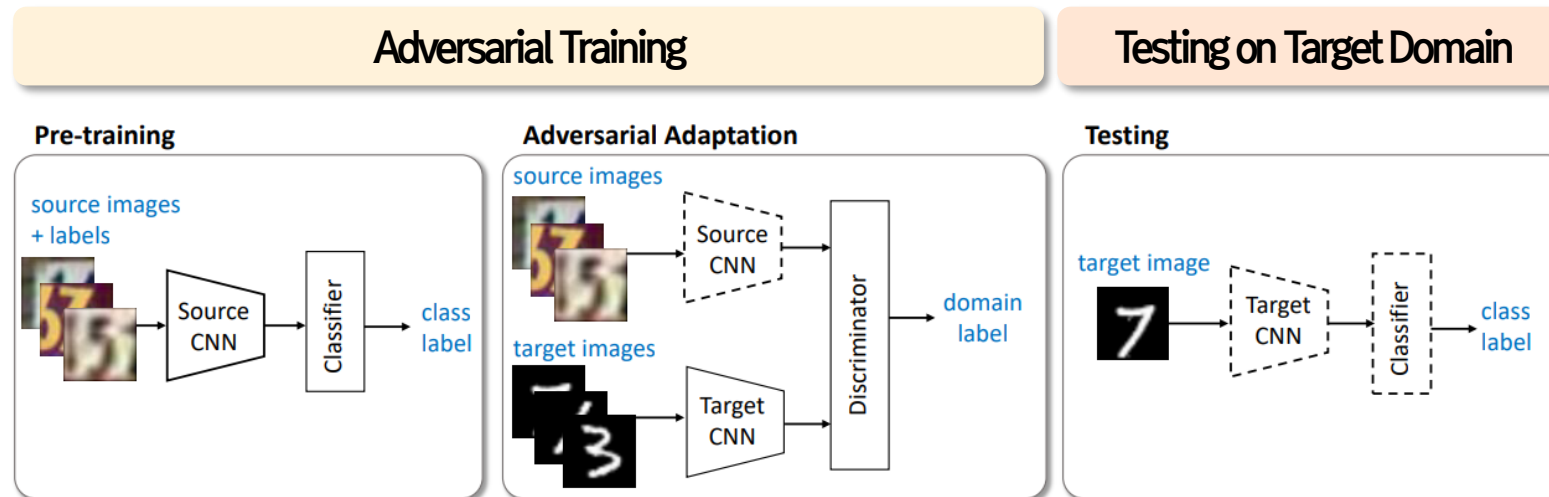
Figure 1: We propose an improved unsupervised domain adaptation method that combines adversarial learning with discriminative feature learning. Specifically, we learn a discriminative mapping of target images to the source feature space (target encoder) by fooling a domain discriminator that tries to distinguish the encoded target images from source examples.

Feature-Based Strategy

Adversarial-Based Approach

❖ Adversarial Discriminative Domain Adaptation (CVPR, 2017)

- 2022년 10월 기준 3666회 인용
- **적대적 학습 방식으로 Domain 간의 차이를 조정**하고 Source Model의 지식을 이용해 Target Task에 대한 성능을 높임



학습 2단계, Target Domain에 대한 Inference 과정

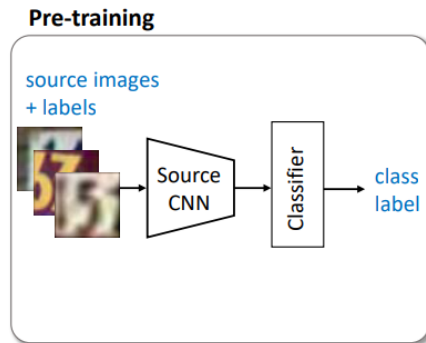
Feature-Based Strategy

Adversarial-Based Approach

❖ Adversarial Discriminative Domain Adaptation (CVPR, 2017)

- 2022년 10월 기준 3666회 인용
- **적대적 학습 방식으로 Domain 간의 차이를 조정하고 Source Model의 지식을 이용해 Target Task에 대한 성능을 높임**

① Discriminative Model (Pretraining Source Model)



$$\min_{M_s, C} \mathcal{L}_{\text{cls}}(\mathbf{X}_s, Y_t) = \mathbb{E}_{(\mathbf{x}_s, y_s) \sim (\mathbf{X}_s, Y_t)} - \sum_{k=1}^K \mathbb{1}_{[k=y_s]} \log C(M_s(\mathbf{x}_s))$$

→ 일반적인 지도학습 방식을 이용하여 Image Classification 문제에 대한 Source Risk를 최소화

Feature-Based Strategy

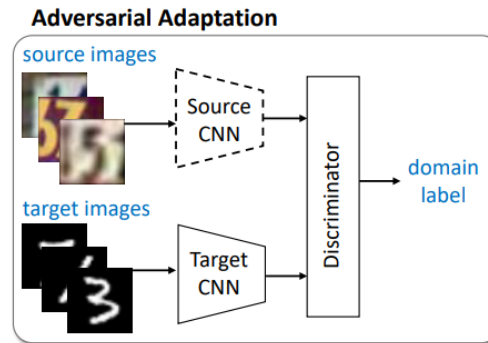
Adversarial-Based Approach

❖ Adversarial Discriminative Domain Adaptation (CVPR, 2017)

- 2022년 10월 기준 3666회 인용
- **적대적 학습 방식으로 Domain 간의 차이를 조정하고 Source Model의 지식을 이용해 Target Task에 대한 성능을 높임**

② Adversarial Adaptation

$$\min_D \mathcal{L}_{\text{adv}_D}(\mathbf{X}_s, \mathbf{X}_t, M_s, M_t) =$$
$$- \mathbb{E}_{\mathbf{x}_s \sim \mathbf{X}_s} [\log D(M_s(\mathbf{x}_s))]$$
$$- \mathbb{E}_{\mathbf{x}_t \sim \mathbf{X}_t} [\log(1 - D(M_t(\mathbf{x}_t)))]$$



$$\min_{M_t} \mathcal{L}_{\text{adv}_M}(\mathbf{X}_s, \mathbf{X}_t, D) =$$
$$- \mathbb{E}_{\mathbf{x}_t \sim \mathbf{X}_t} [\log D(M_t(\mathbf{x}_t))]$$

→ Source Feature Extractor는 고정 (파라미터 업데이트 하지 않음)

→ 독립적인 Target Feature Extractor만 업데이트 수행

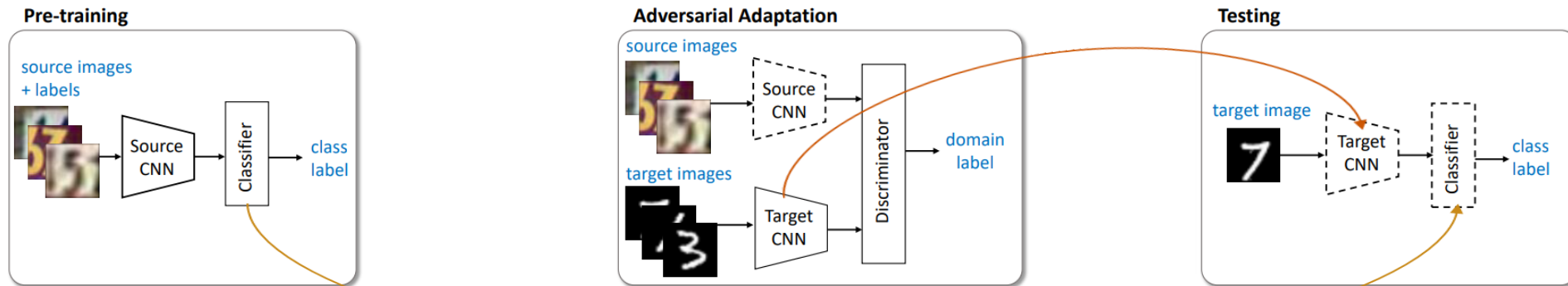
Feature-Based Strategy

Adversarial-Based Approach

❖ Adversarial Discriminative Domain Adaptation (CVPR, 2017)

- 2022년 10월 기준 3666회 인용
- **적대적 학습 방식으로 Domain 간의 차이를 조정하고 Source Model의 지식을 이용해 Target Task에 대한 성능을 높임**
- **Gradient Reversal Layer를 활용한 DANN(MLR, 2016)보다 좋은 성능**

③ Inference (Testing on Target Domain)



→ 사전 학습된 Source Classifier을 이용하여 Target Domain으로 Target Task 수행

4. Conclusion

Conclusions

Deep Transfer Learning: How to Transfer Knowledge Across Domains by Deep Neural Network?

- ❖ 대표적인 Deep Transfer Learning 방법론에 대해 살펴보고 특징 파악
 - Transfer Learning 이 필요한 시나리오 및 경우의 수가 여러가지이므로, 문제 상황에 대한 정의가 중요
- ❖ 각 방법론마다 가정과 특색이 다르기에 상황에 맞게 활용하는 것이 필요
- ❖ 최근 Deep Transfer Learning 분야는 Target Domain의 레이블 정보를 필요로 하지 않는 'Unsupervised Domain Adaptation' 과 같은 분야에서 활발히 수행되고 있음 (Source Task = Target Task)
- ❖ 모델 활용에 있어 실제 현장에서 적용 시 강건한 예측 성능을 내기 위한 방법론으로써, 향후 다양한 Application에서 적용 가능
 - 산업 현장, Bioinformation, Computer Vision 등

5. Reference

References

- [1] A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges
- [2] Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks?. *Advances in neural information processing systems*, 27.
- [3] Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018, October). A survey on deep transfer learning. In *International conference on artificial neural networks* (pp. 270–279). Springer, Cham.
- [4] Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., ... & He, Q. (2020). A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1), 43–76.
- [5] Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., & Darrell, T. (2014). Deep domain confusion: Maximizing for domain invariance. *arXiv preprint arXiv:1412.3474*.
- [6] Tzeng, E., Hoffman, J., Saenko, K., & Darrell, T. (2017). Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7167–7176).

Thank You