

Deep Domain Adaptation

DMQA Open Seminar

2022.08.05

김태연

발표자 소개



❖ 김태연

- 고려대학교 산업경영공학과
- Data Mining & Quality Analytics Lab. (김성범 교수님)
- M.S. Student (2021.03 ~)

❖ Research Interest

- Machin Learning & Deep Learning Algorithms
- Explainable Artificial Intelligence(XAI)
- Multivariate Time-series Data

❖ Contact

- E-mail : wind0971@korea.ac.kr

Contents

1. Introduction
2. Deep Domain Adaptation
 - 2-1. DANN (Domain-Adversarial Neural Networks)
 - 2-2. CCSA-loss (Unified Deep Supervised Domain Adaptation and Generalization)
 - 2-3. SagNet (Reducing Domain Gap by Reducing Style Bias)
3. Conclusions

1. Introduction

Introduction

Background

❖ 딥러닝 분류 모델

- 원하는 task에 적합한 학습 데이터 수집



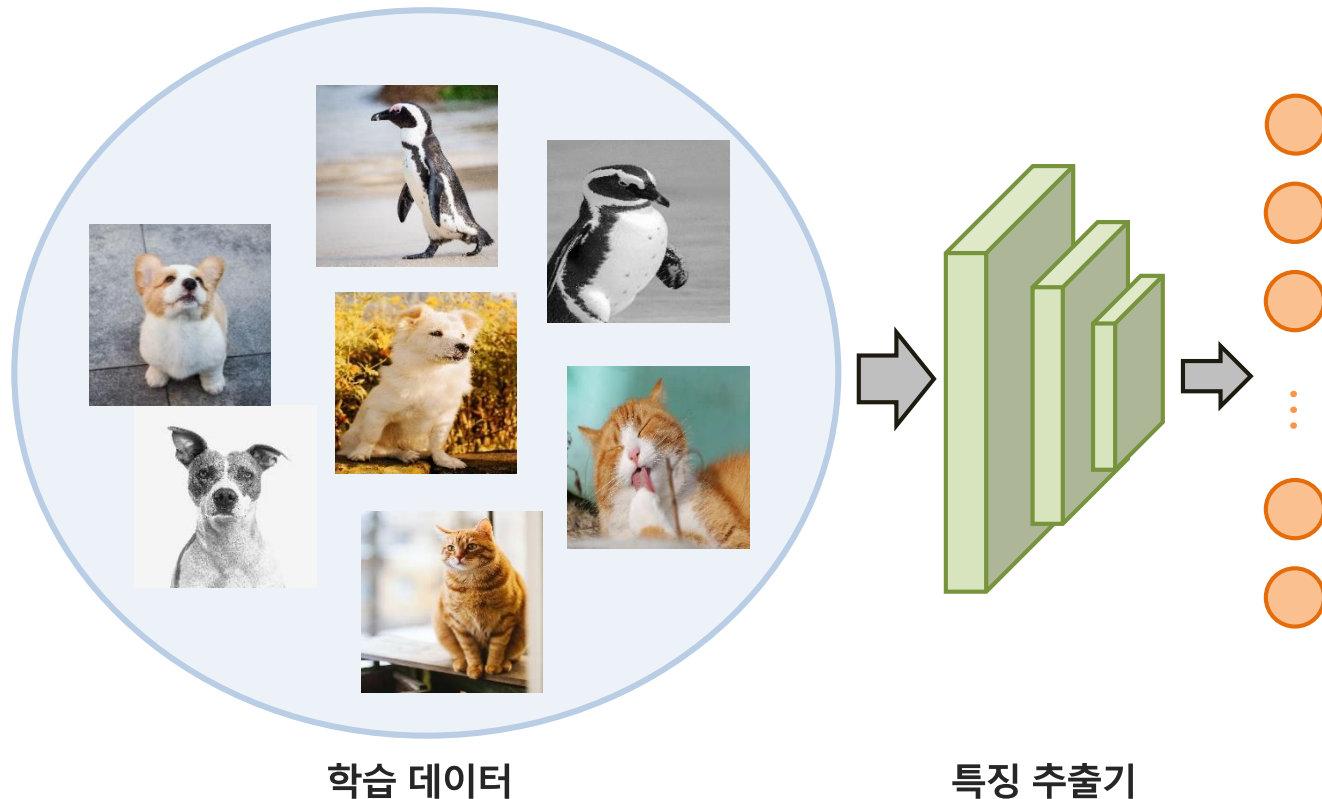
데이터 수집

Introduction

Background

❖ 딥러닝 분류 모델

- 수집된 학습 데이터를 바탕으로 각 데이터들의 특징을 파악하는 특징 추출기 학습

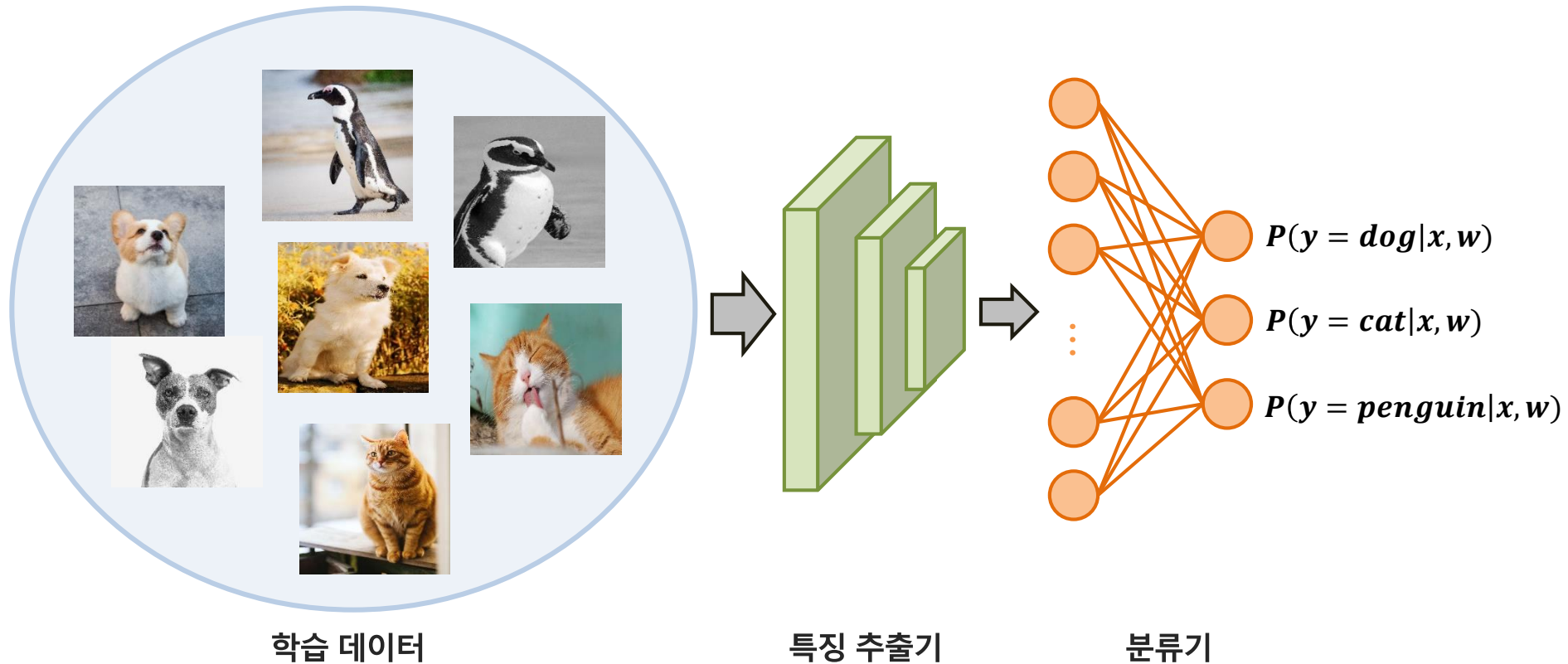


Introduction

Background

❖ 딥러닝 분류 모델

- 추출된 특징들로 데이터의 label을 예측하는 분류기 학습



Introduction

Background

❖ 딥러닝 분류 모델

- 학습된 모델을 바탕으로 테스트 데이터에서의 예측 성능 확인

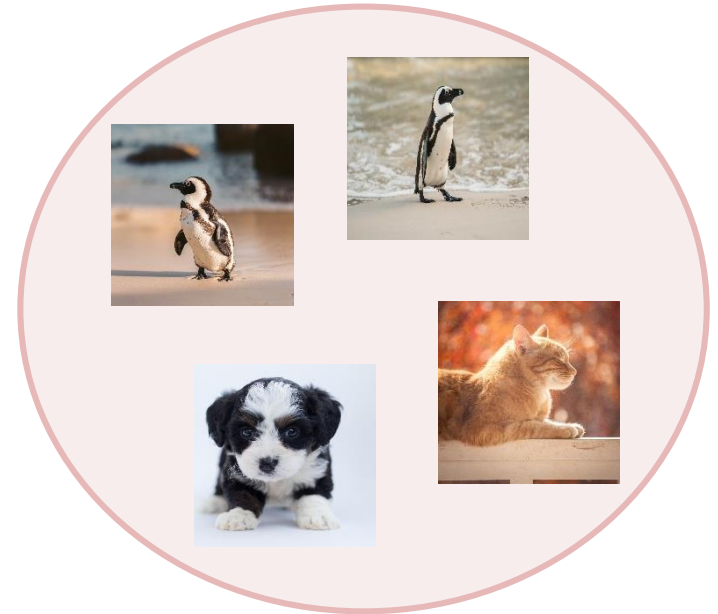


Introduction

Background



학습 데이터



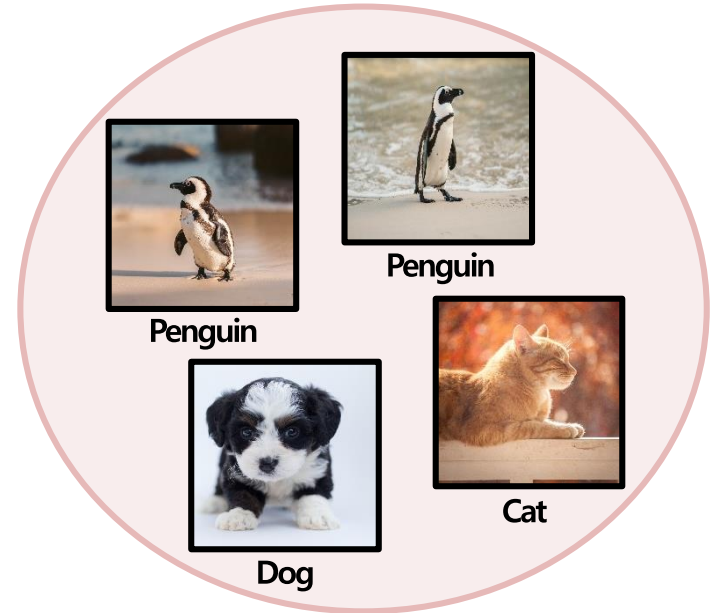
테스트 데이터

Introduction

Background



학습 데이터



테스트 데이터

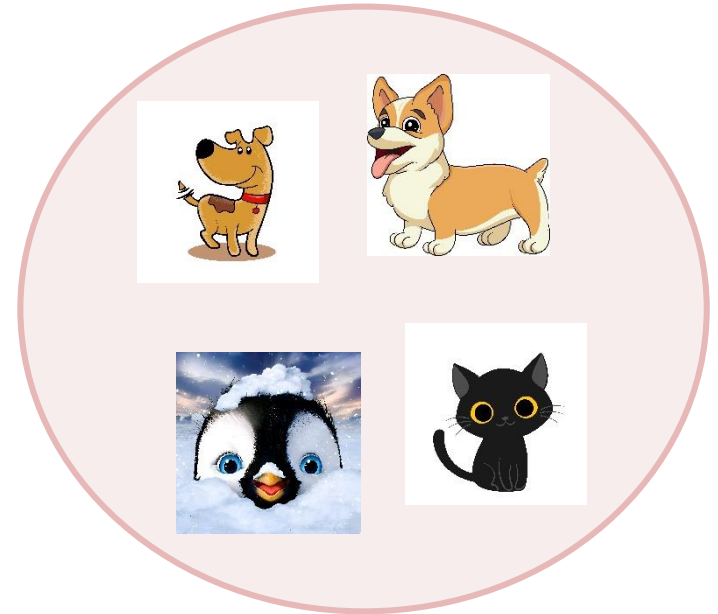
테스트 데이터에 대해 정확한 예측 가능

Introduction

Background



학습 데이터



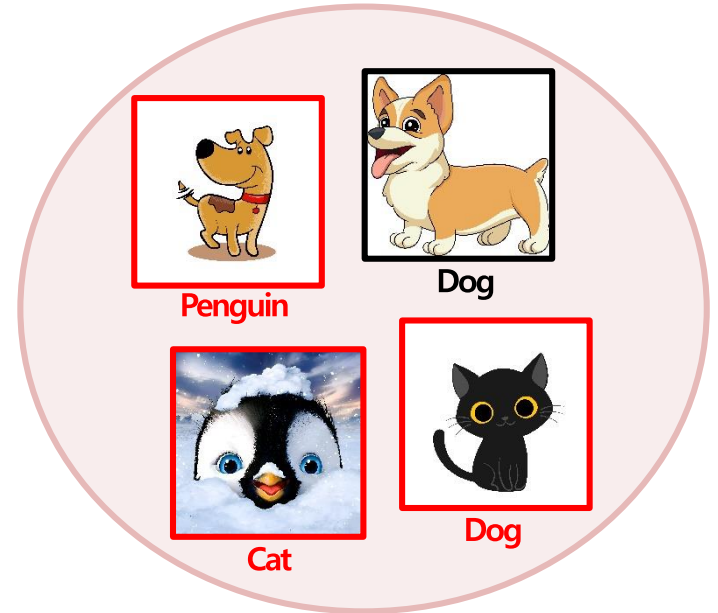
테스트 데이터

Introduction

Background



학습 데이터



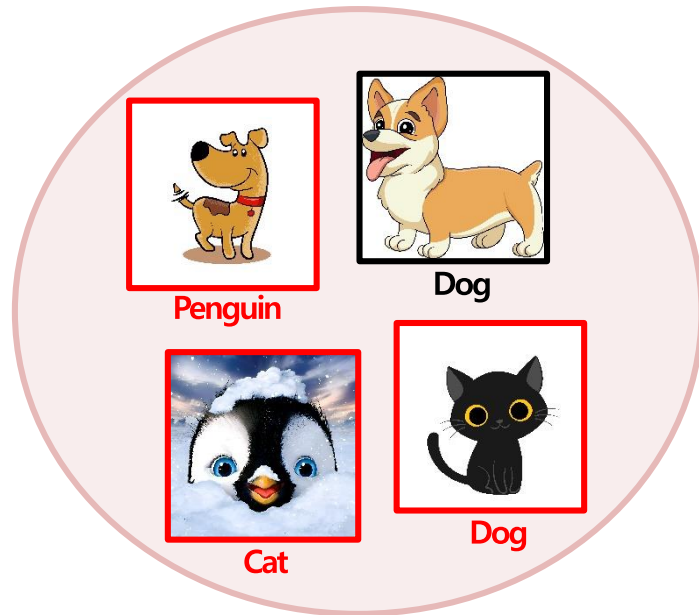
테스트 데이터

데이터간의 **도메인 차이**로 정확한 예측 불가능
(Domain Shift)

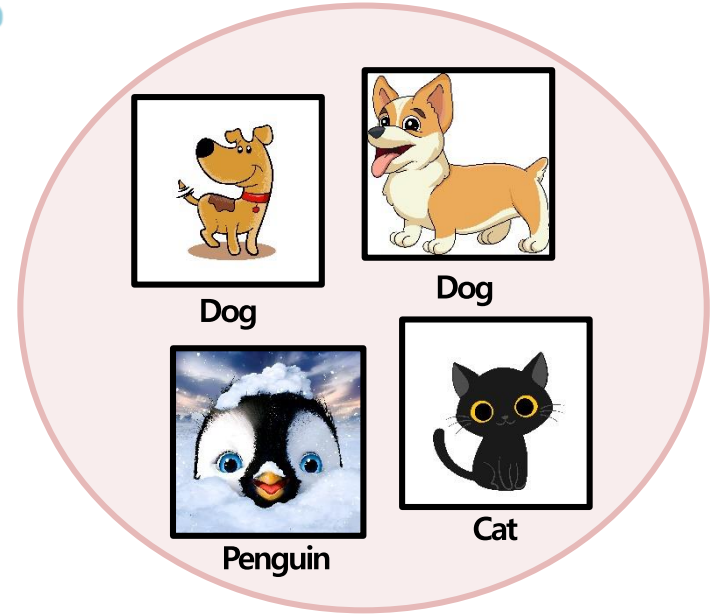
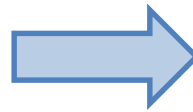
Introduction

Background

How?



테스트 데이터



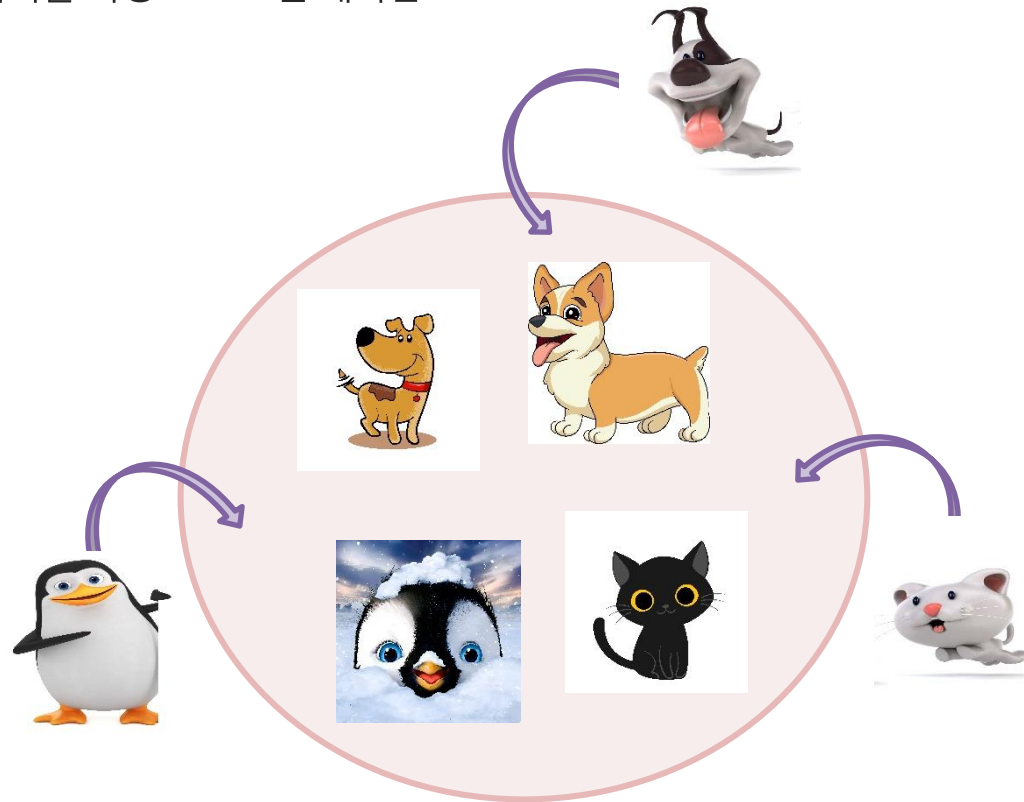
테스트 데이터

분포가 다른 데이터에서의 정확한 예측 방법에 대한 연구

Introduction

Background

- ❖ 도메인 시프트(Domain Shift) 해결 방안
 - 해당 도메인의 데이터 수집 진행
 - 수집된 데이터를 바탕으로 모델 재학습



분류하고자 하는 데이터셋

Introduction

Background

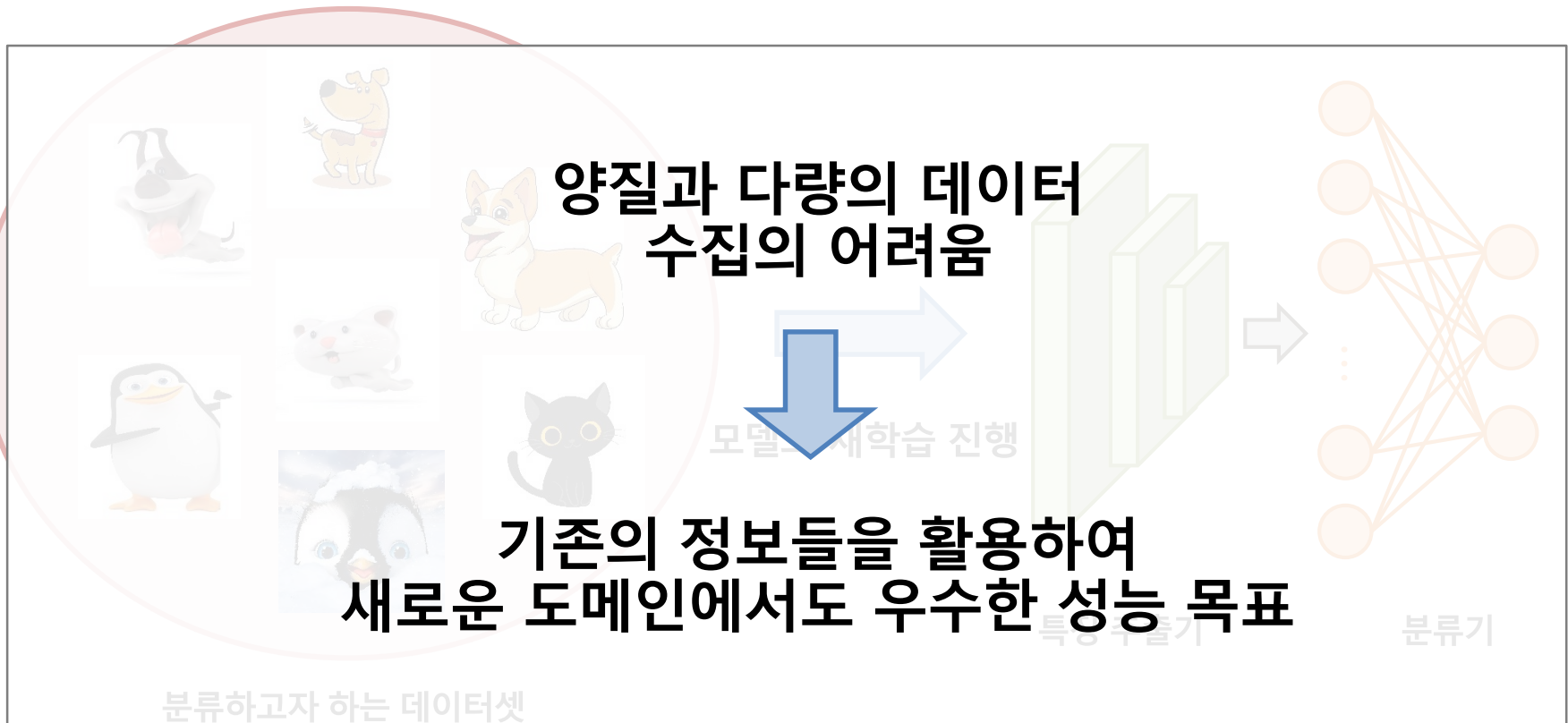
- ❖ 도메인 시프트(Domain Shift) 해결 방안
 - 해당 도메인의 데이터 수집 진행
 - 수집된 데이터를 바탕으로 모델 재학습



Introduction

Background

- ❖ **도메인 시프트(Domain Shift) 해결 방안**
 - 해당 도메인의 데이터 수집 진행
 - 수집된 데이터를 바탕으로 모델 재학습



Introduction

Background

- ❖ **도메인 시프트(Domain Shift) 해결 방안**
 - 해당 도메인의 데이터 수집 진행
 - 수집된 데이터를 바탕으로 모델 재학습

양질과 다량의 데이터
수집의 어려움

Transfer Learning

기존의 정보들을 활용하여
새로운 도메인에서도 우수한 성능 목표

Introduction

Background

❖ Task / Domain

- 태스크(Task): 주어진 데이터로 수행할 작업의 종류
 - Recognition / Detection ...
- 도메인(Domain): 데이터의 feature space \mathbb{X} 와 해당 데이터의 label space \mathbb{Y} 로 구성된 집합
 - Source / Target

Introduction

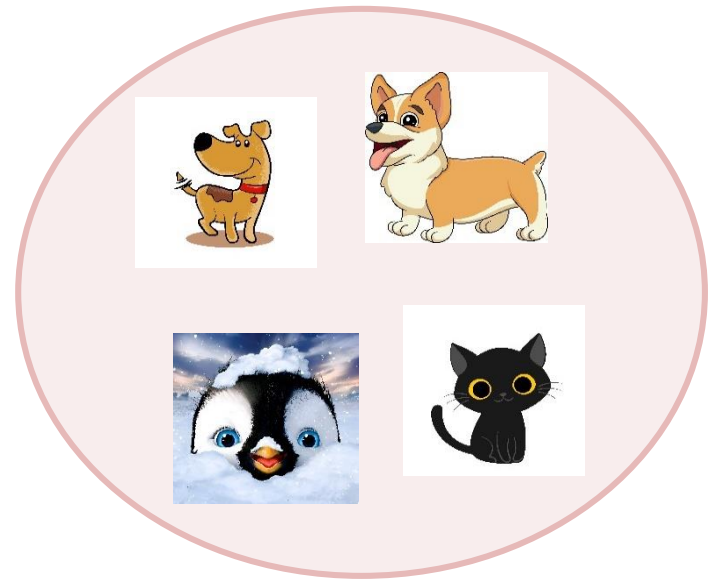
Background

❖ Task / Domain

- 태스크(Task): 주어진 데이터로 수행할 작업의 종류
 - Recognition / Detection ...
- 도메인(Domain): 데이터의 feature space \mathbb{X} 와 해당 데이터의 label space \mathbb{Y} 로 구성된 집합
 - Source / Target



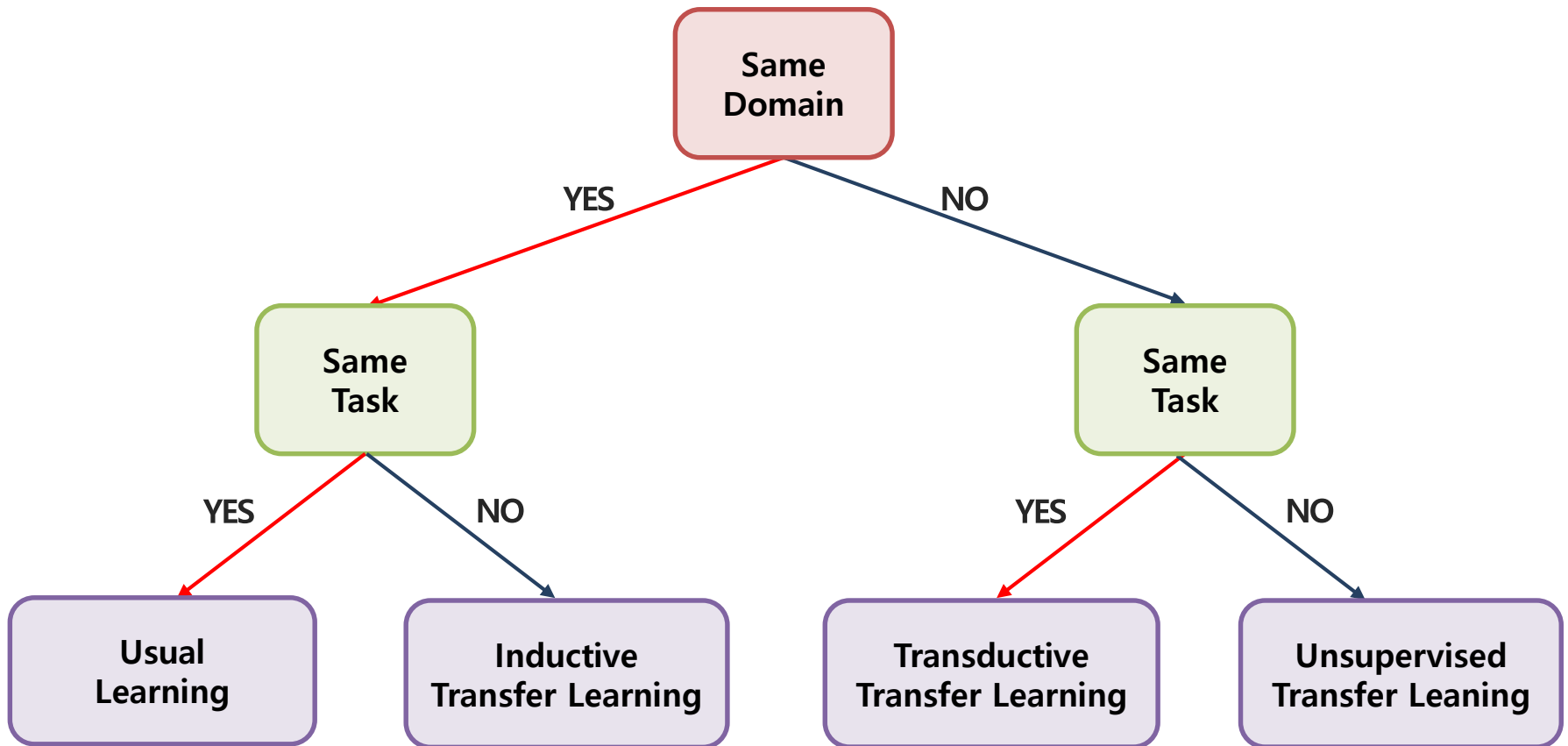
Source domain



Target domain

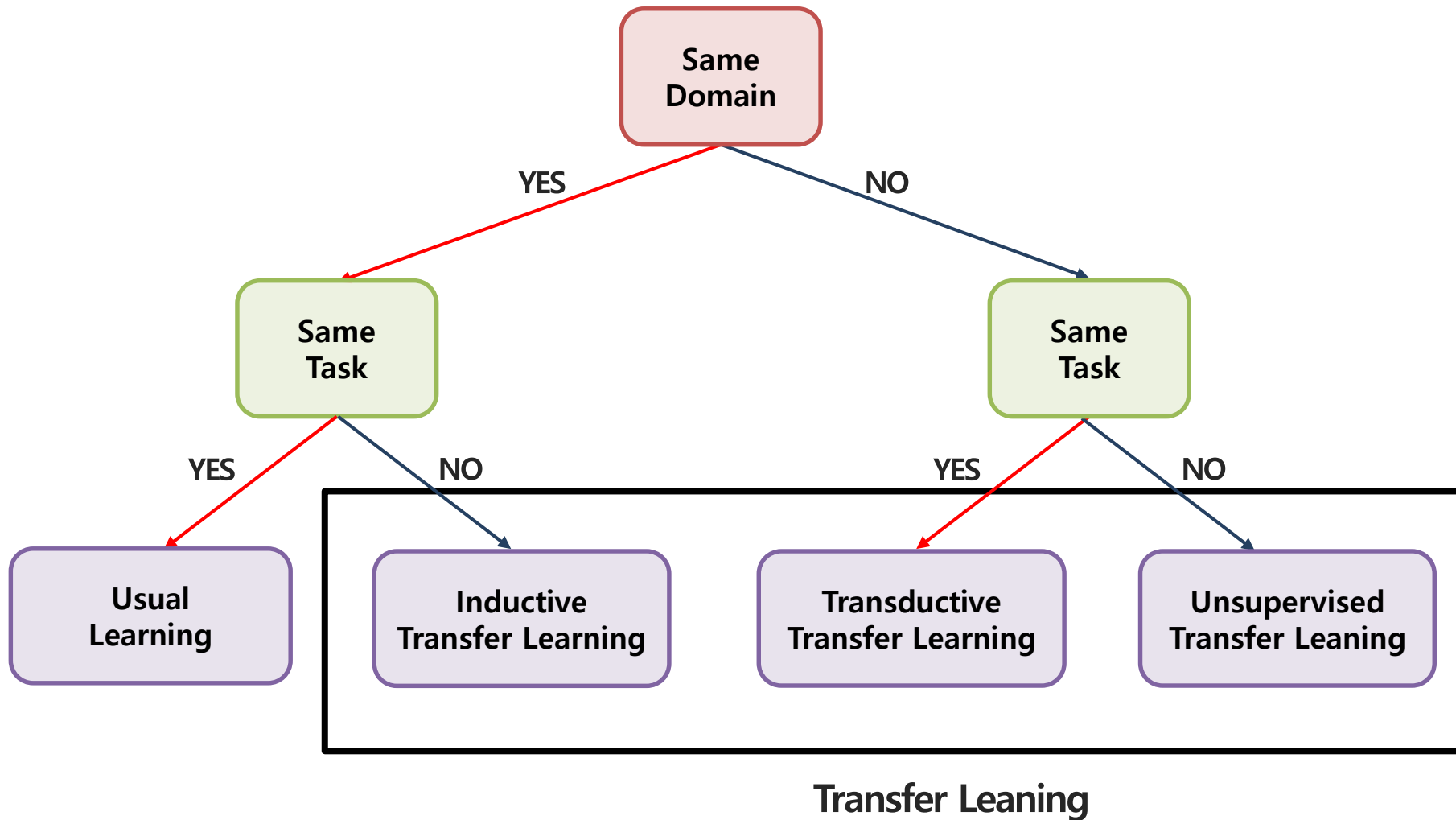
Introduction

Background



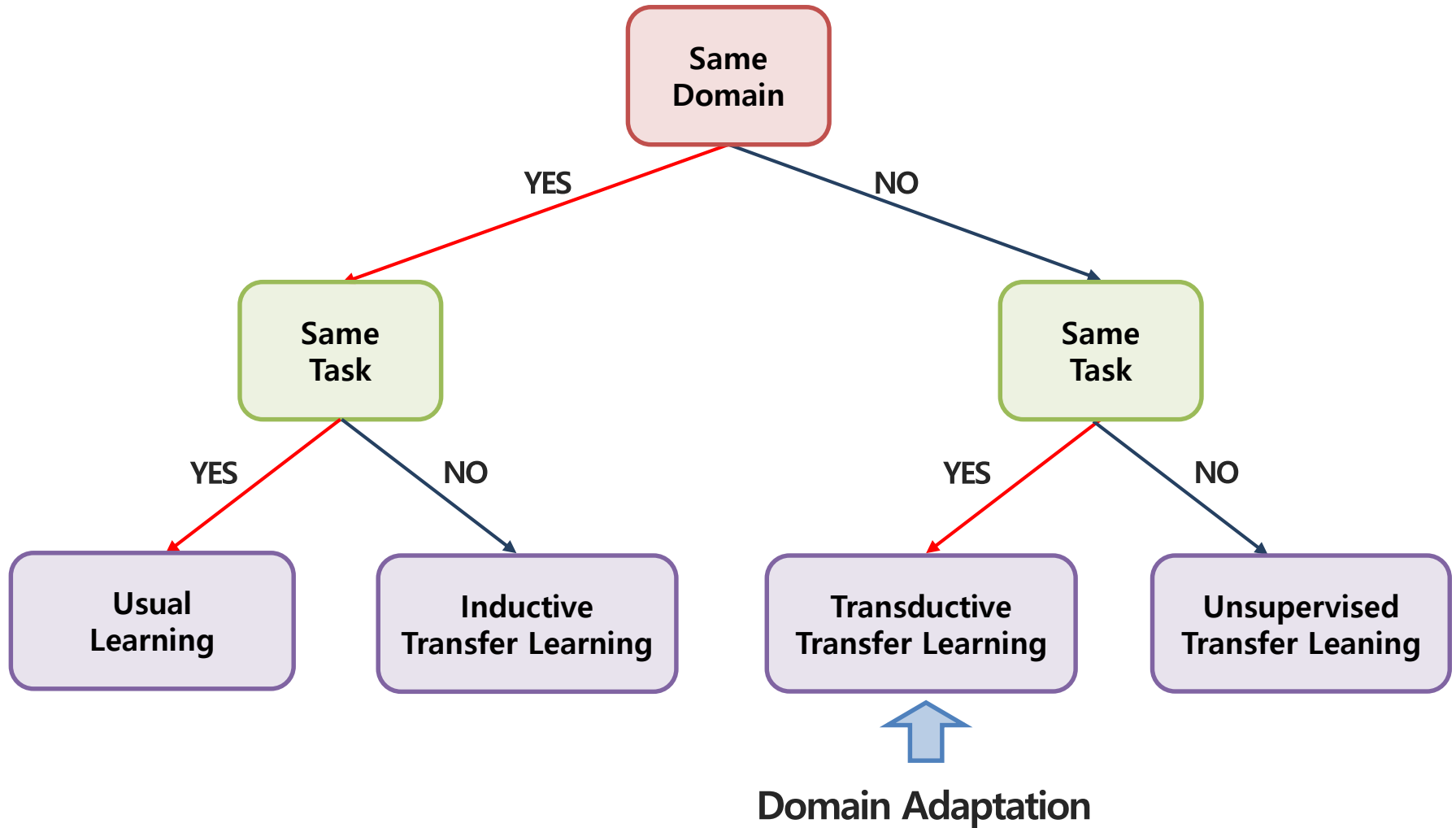
Introduction

Background



Introduction

Background



Introduction

Background

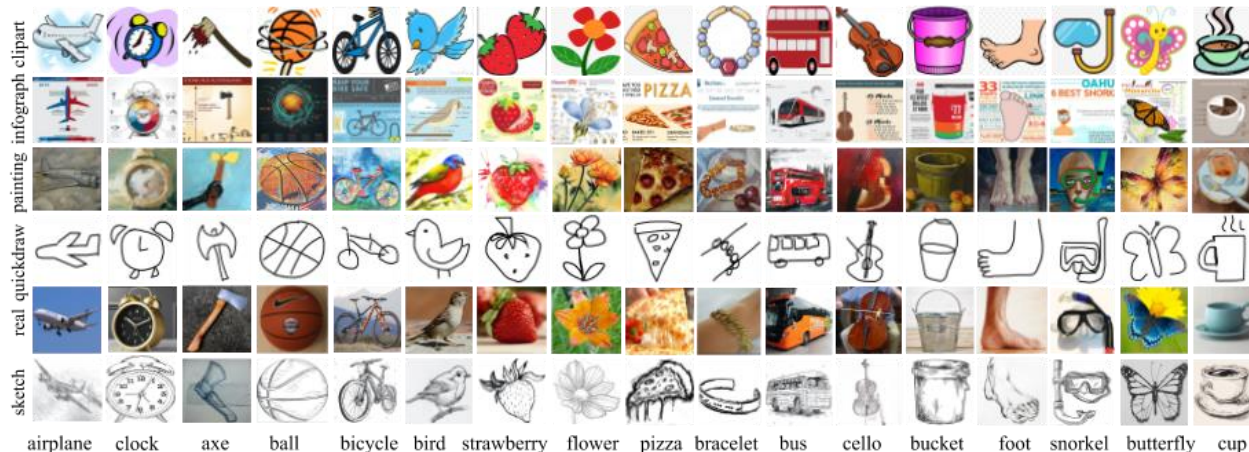
❖ Dataset for Domain Adaptation

- Domain Adaptation은 주로 computer vision과 관련하여 연구가 진행
 - Office Dataset / DomainNet Dataset

Office Dataset



DomainNet Dataset



2. Deep Domain Adaptation

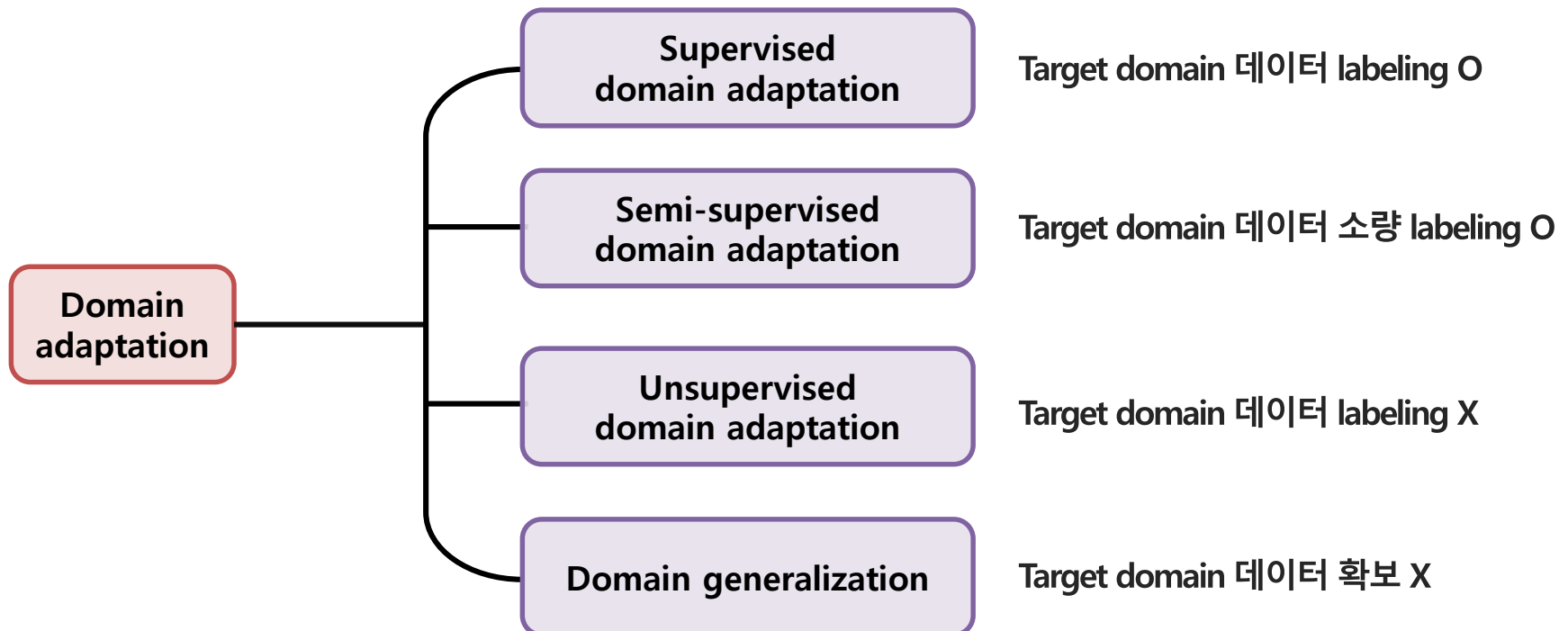
2-1. DANN (Domain-Adversarial Neural Networks)

Deep Domain Adaptation

DANN

❖ Domain-Adversarial Training of Neural Networks

- Target domain내 대량의 unlabeled data를 활용하는 unsupervised domain adaptation(UDA) 방법론
- D_S : Source domain , D_T : Target domain
- Target risk ($R_{D_T(\eta)}$)가 낮도록 분류기 $\eta : X \rightarrow Y$ 을 구축



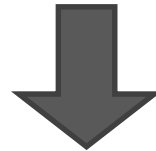
Deep Domain Adaptation

DANN

❖ Domain-Adversarial Training of Neural Networks

- Target domain내 대량의 unlabeled data를 활용하는 unsupervised domain adaptation(UDA) 방법론
- D_S : Source domain , D_T : Target domain
- Target risk ($R_{D_T(\eta)}$)가 낮도록 분류기 $\eta : X \rightarrow Y$ 을 구축

$$\text{Target risk } (R_{D_T(\eta)}) : P_{(x,y) \sim D_T}(\eta(x) \neq y)$$



$$\text{Target risk } (R_{D_T(\eta)}) \leq \text{Source risk} + \text{Domain divergence}$$

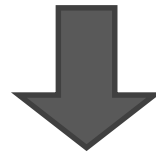
Deep Domain Adaptation

DANN

❖ Domain-Adversarial Training of Neural Networks

- Target domain내 대량의 unlabeled data를 활용하는 unsupervised domain adaptation(UDA) 방법론
- D_S : Source domain , D_T : Target domain
- Target risk ($R_{D_T(\eta)}$)가 낮도록 분류기 $\eta : X \rightarrow Y$ 을 구축

$$\text{Target risk } (R_{D_T(\eta)}) : P_{(x,y) \sim D_T}(\eta(x) \neq y)$$



$$\text{Target risk } (R_{D_T(\eta)}) \leq \text{Source risk} + \text{Domain divergence}$$

Source domain에서 잘못 예측할 위험

서로 다른 두 도메인을 구분하는 능력

Deep Domain Adaptation

DANN

❖ Domain-Adversarial Training of Neural Networks

- Target domain내 대량의 unlabeled data를 활용하는 unsupervised domain adaptation(UDA) 방법론
- D_S : Source domain , D_T : Target domain
- Target risk ($R_{D_T(\eta)}$)가 낮도록 분류기 $\eta : X \rightarrow Y$ 을 구축

$$\text{Target risk } (R_{D_T(\eta)}) \leq \text{Source risk} + \text{Domain divergence}$$

Source domain에서의 분류 성능 ↑

Source ↔ Target domain 구분 성능 ↓

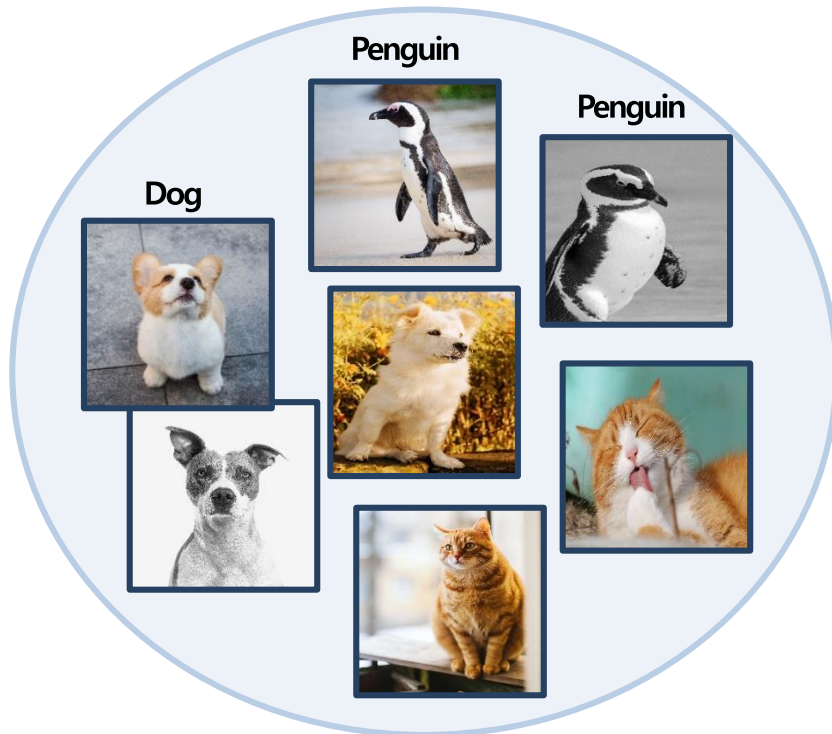
Deep Domain Adaptation

DANN

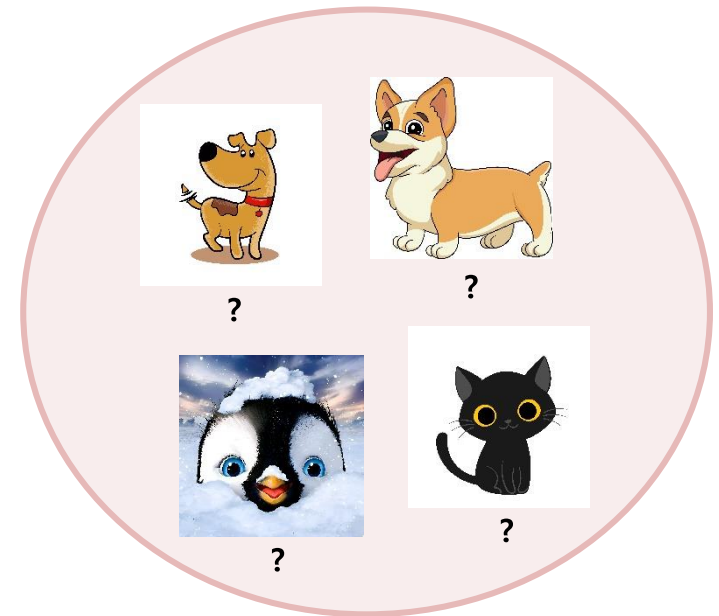
❖ Domain-Adversarial Training of Neural Networks

- Label이 존재하는 source domain / Label 이 존재하지 않는 target domain
- Source / Target domain을 구분하기 위해 각각의 domain label S,D를 추가적으로 부여

Class Label



Source domain



Target domain

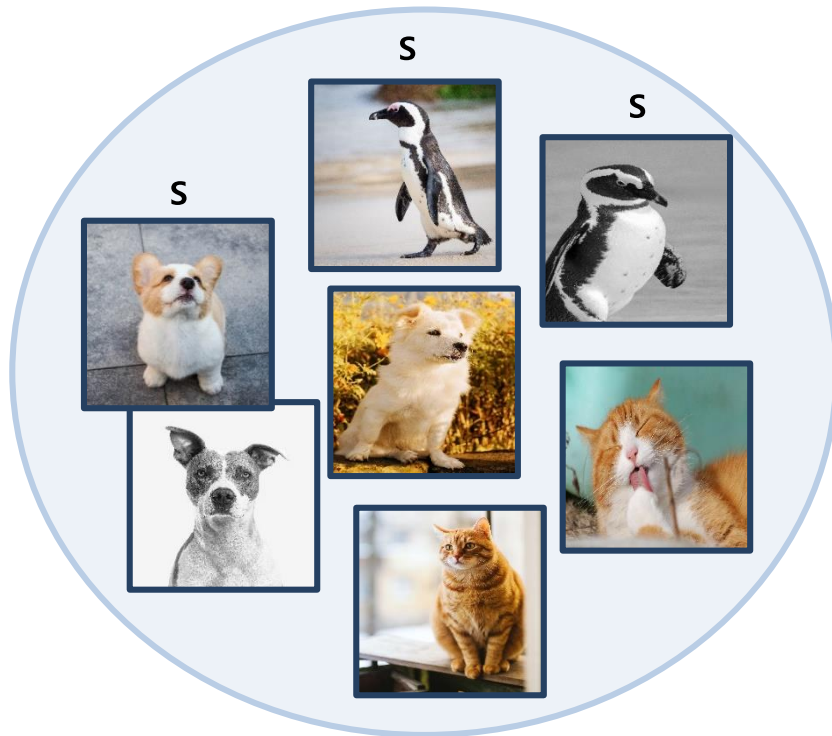
Deep Domain Adaptation

DANN

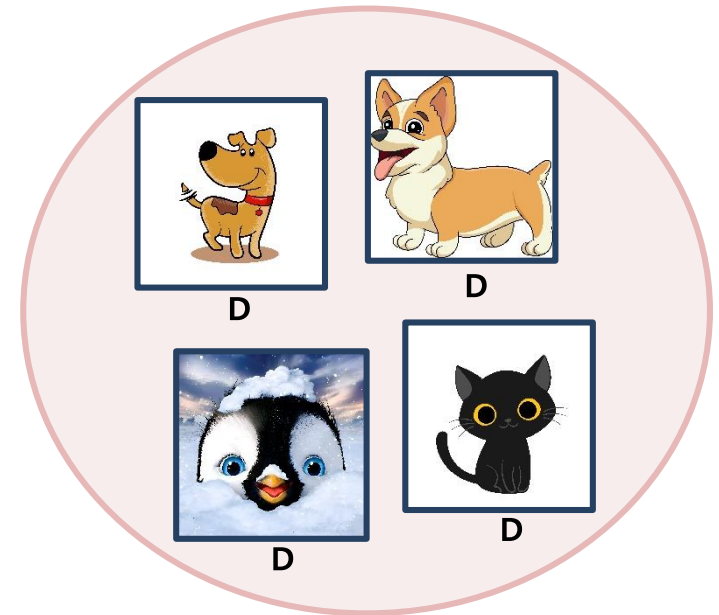
❖ Domain-Adversarial Training of Neural Networks

- Label이 존재하는 source domain / Label 이 존재하지 않는 target domain
- Source / Target domain을 구분하기 위해 각각의 domain label S,D을 추가적으로 부여

Domain Label



Source domain



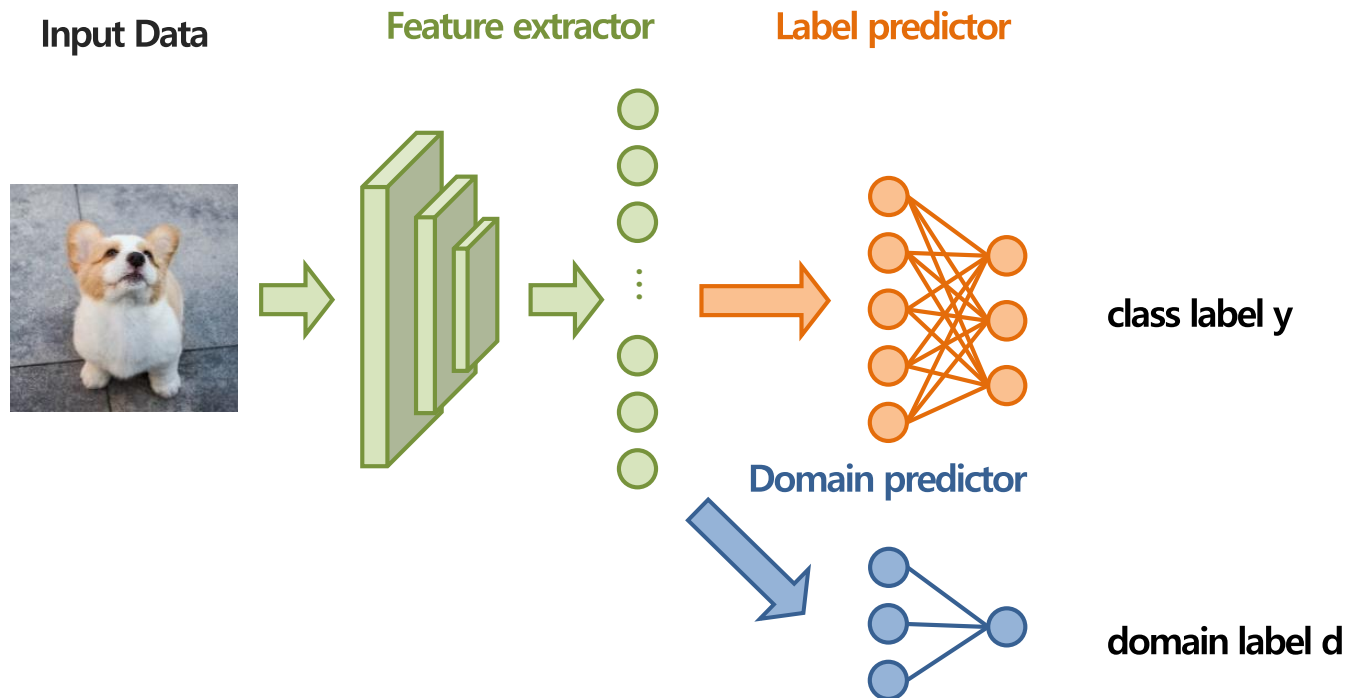
Target domain

Deep Domain Adaptation

DANN

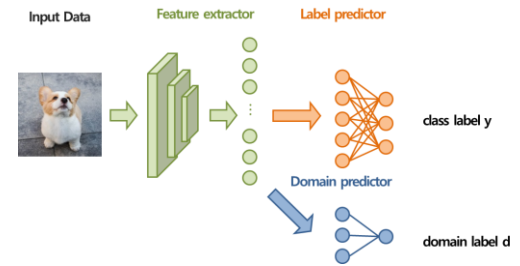
❖ Domain-Adversarial Training of Neural Networks

- 데이터의 특징 추출기(feature extractor : G_f)
- 데이터의 label 분류기(label predictor : G_y) : Source domain의 label 분류
- 데이터의 domain label 분류기 (domain predictor : G_d) : 해당 데이터가 source / target domain 인지 분류



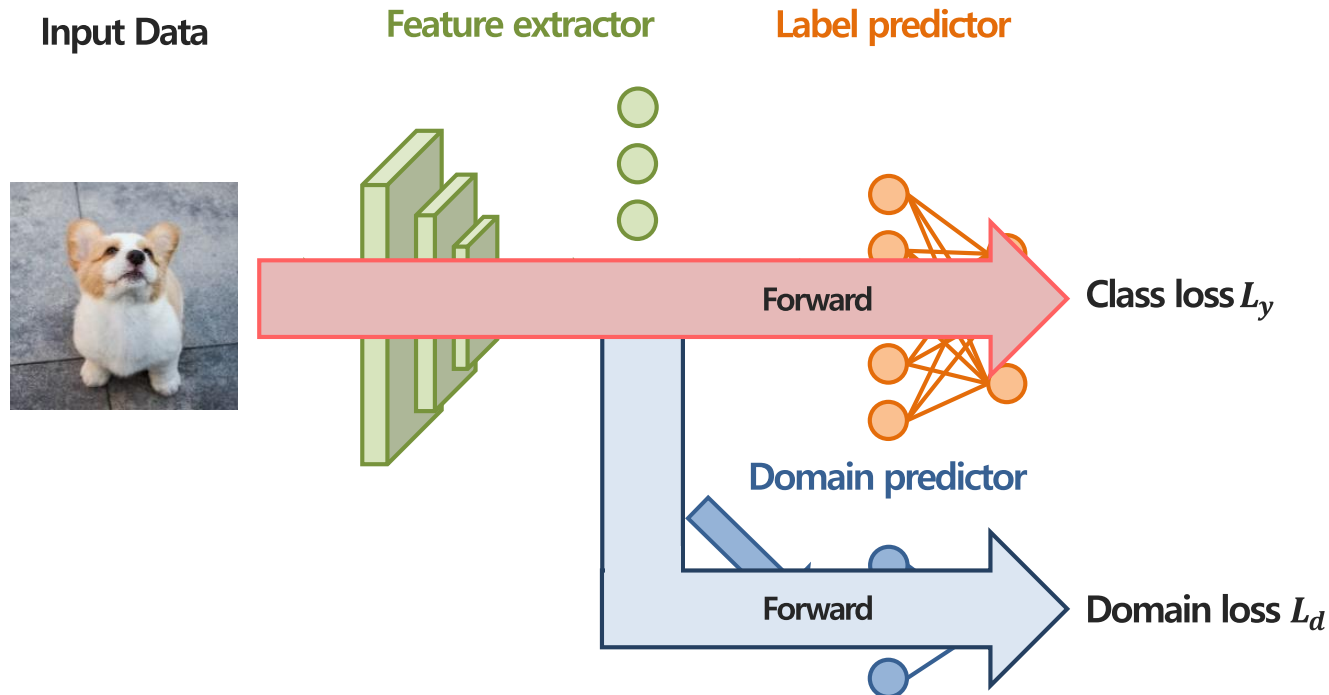
Deep Domain Adaptation

DANN



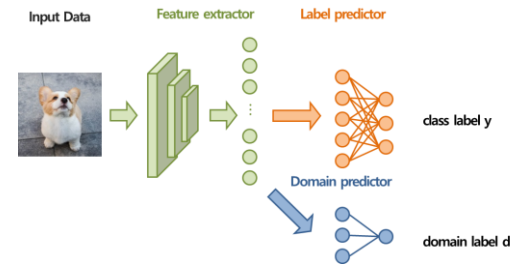
❖ Domain-Adversarial Training of Neural Networks

- 데이터의 특징 추출기(feature extractor : G_f)
- 데이터의 label 분류기(label predictor : G_y) : Source domain의 label 분류
- 데이터의 domain label 분류기 (domain predictor : G_d) : 해당 데이터가 source / target domain 인지 분류



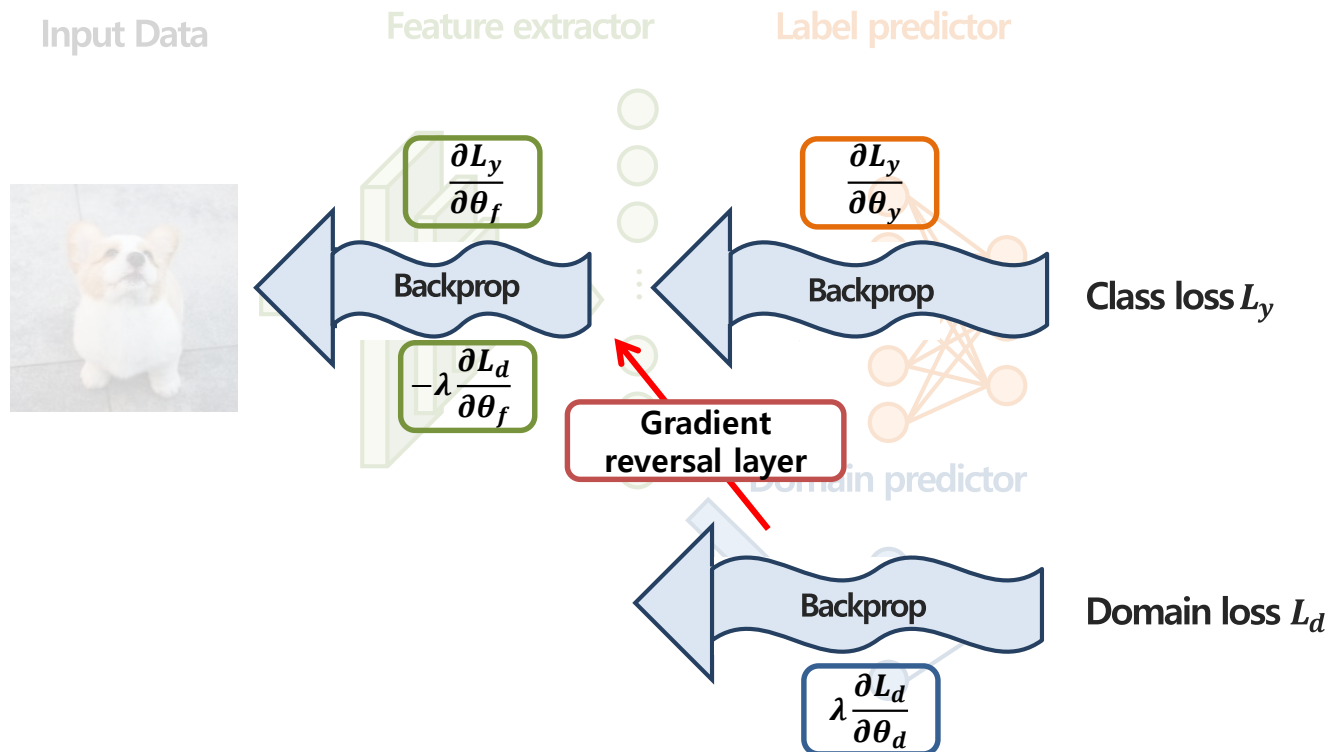
Deep Domain Adaptation

DANN



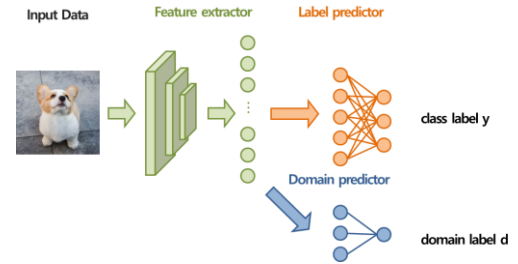
❖ Domain-Adversarial Training of Neural Networks

- 데이터의 특징 추출기(feature extractor : G_f)
- 데이터의 label 분류기(label predictor : G_y) : Source domain의 label 분류
- 데이터의 domain label 분류기 (domain predictor : G_d) : 해당 데이터가 source / target domain 인지 분류



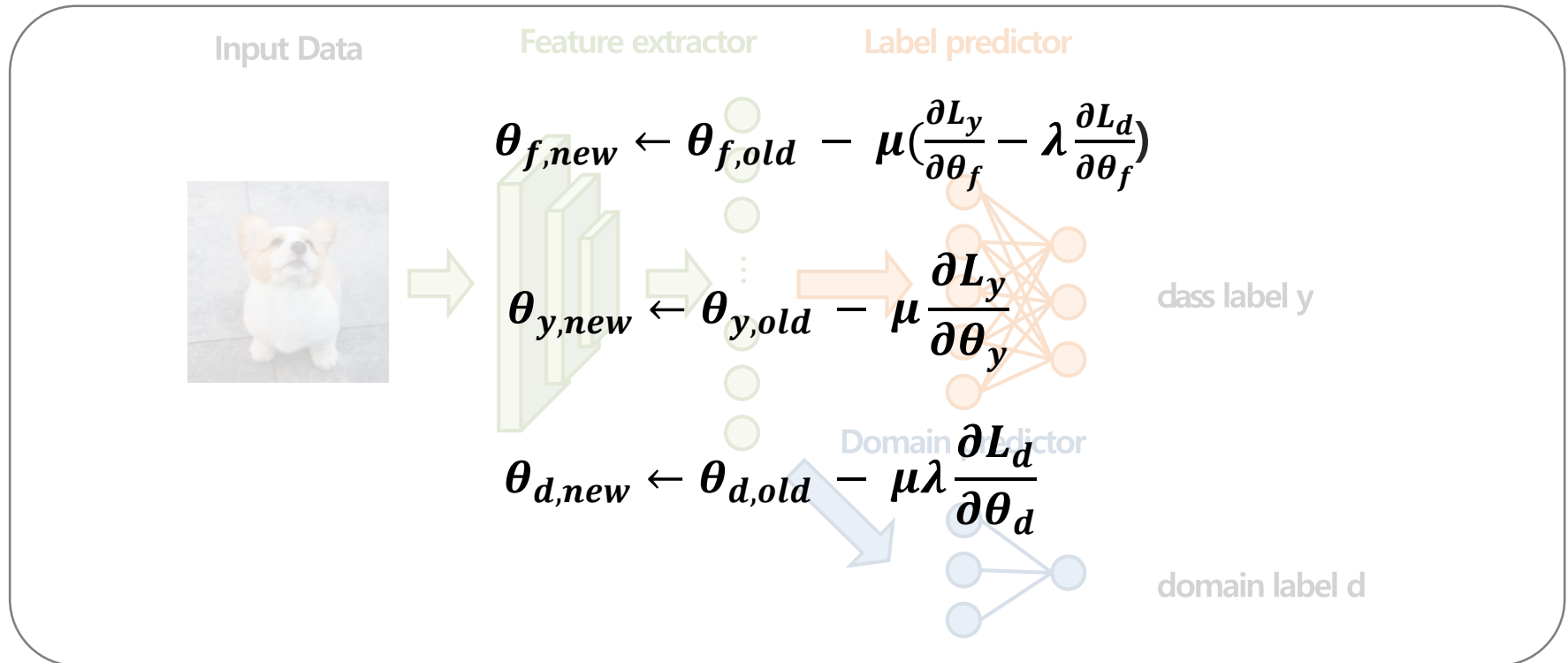
Deep Domain Adaptation

DANN



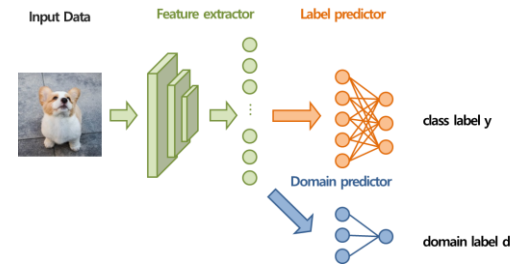
❖ Domain-Adversarial Training of Neural Networks

- 데이터의 특징 추출기(feature extractor : G_f)
- 데이터의 label 분류기(label predictor : G_y) : Source domain의 label 분류
- 데이터의 domain label 분류기 (domain predictor : G_d) : 해당 데이터가 source / target domain 인지 분류



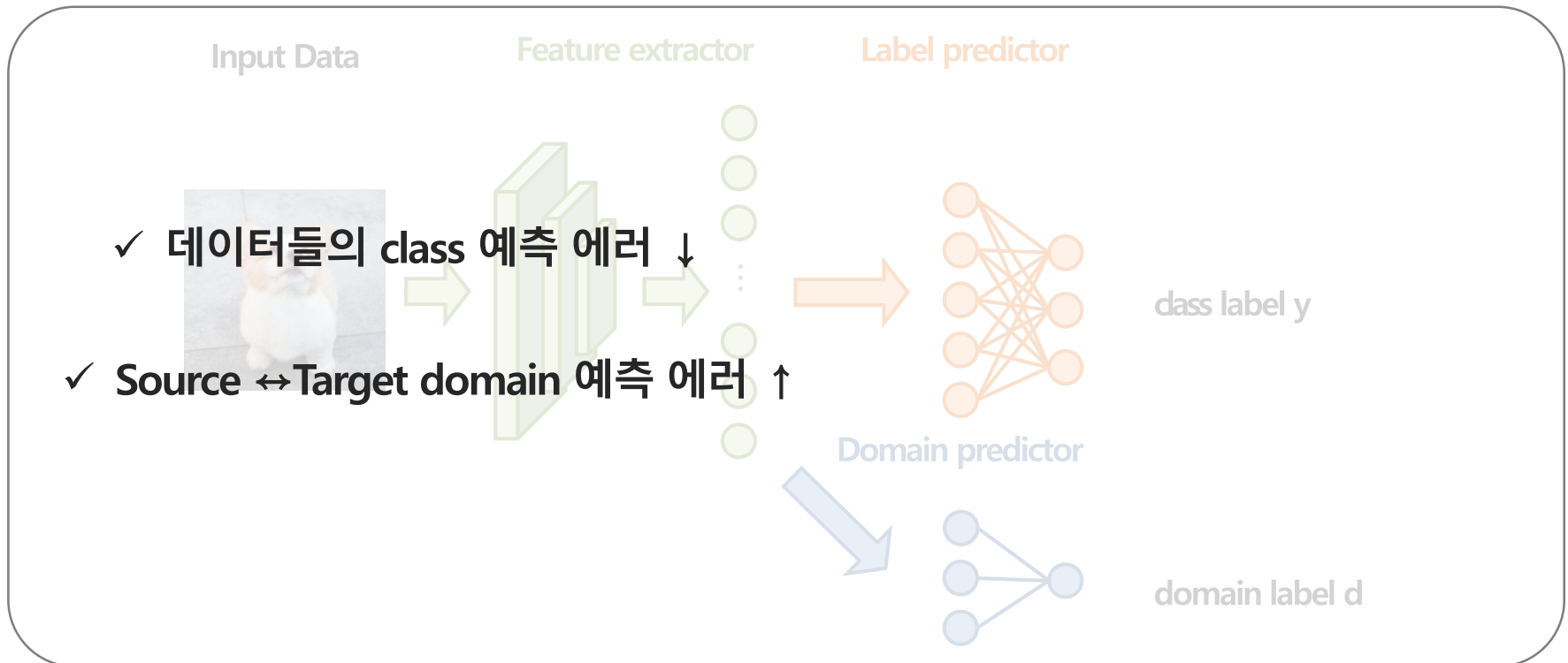
Deep Domain Adaptation

DANN



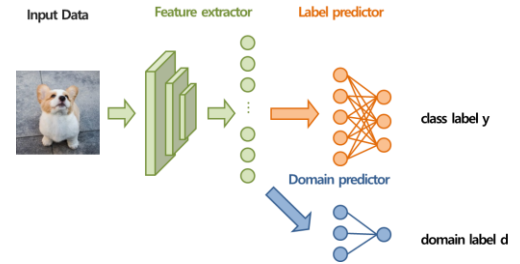
❖ Domain-Adversarial Training of Neural Networks

- 데이터의 특징 추출기(feature extractor : G_f)
- 데이터의 label 분류기(label predictor : G_y) : Source domain의 label 분류
- 데이터의 domain label 분류기 (domain predictor : G_d) : 해당 데이터가 source / target domain 인지 분류



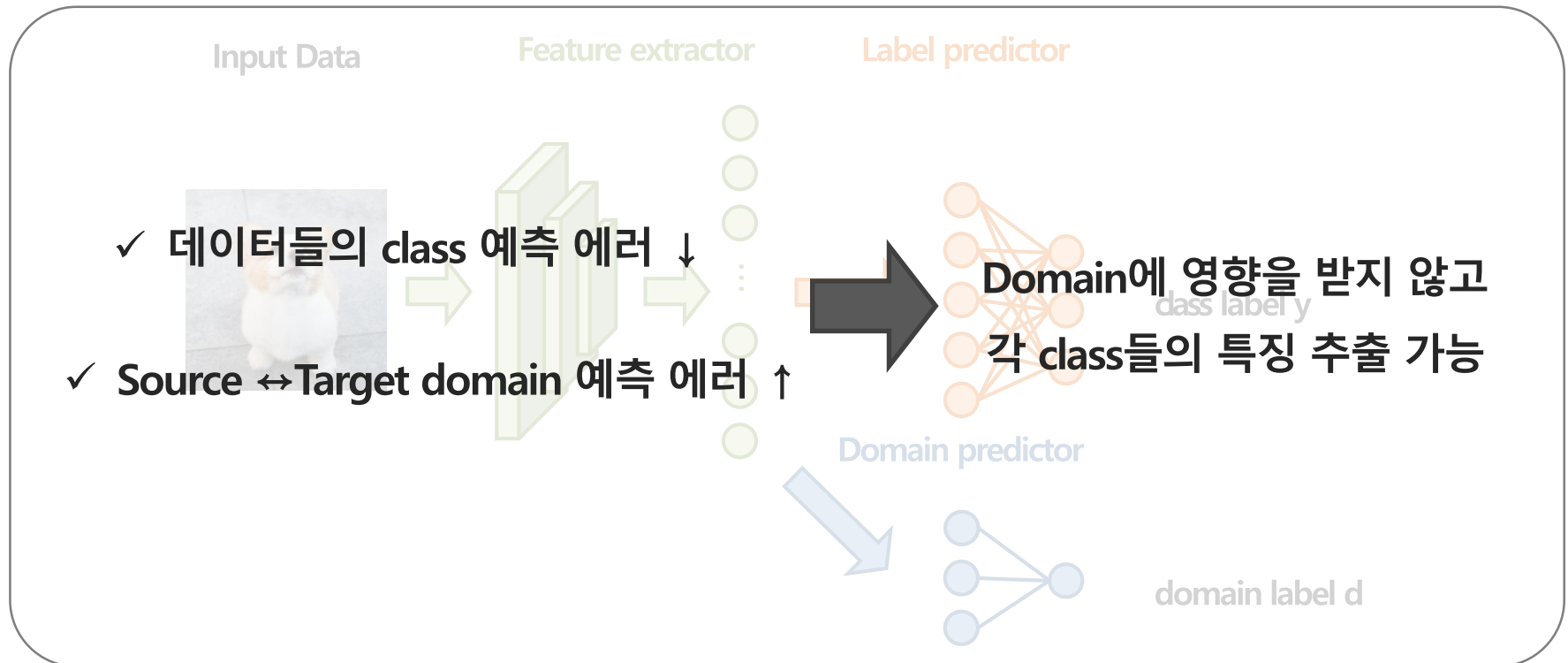
Deep Domain Adaptation

DANN



❖ Domain-Adversarial Training of Neural Networks

- 데이터의 특징 추출기(feature extractor : G_f)
- 데이터의 label 분류기(label predictor : G_y) : Source domain의 label 분류
- 데이터의 domain label 분류기 (domain predictor : G_d) : 해당 데이터가 source / target domain 인지 분류



2-2. CCSA-loss (Unified Deep Supervised Domain Adaptation and Generalization)

Deep Domain Adaptation

CCSA-loss

❖ Unified Deep Supervised Domain Adaptation and Generalization

- Target domain내 소수의 labeled data가 존재하는 경우 해당 labeled data를 활용하여 성능을 향상시키는 supervised domain adaptation(SDA) 방법론
- 서로 다른 두 도메인을 구분하는 능력인 domain divergence 를 세분화 하여 정의

Target domain 데이터들의
labeling이 필요 없는
unsupervised domain adaptation

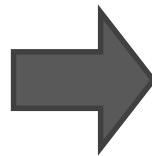
Deep Domain Adaptation

CCSA-loss

❖ Unified Deep Supervised Domain Adaptation and Generalization

- Target domain내 소수의 labeled data가 존재하는 경우 해당 labeled data를 활용하여 성능을 향상시키는 supervised domain adaptation(SDA) 방법론
- 서로 다른 두 도메인을 구분하는 능력인 domain divergence 를 세분화 하여 정의

Target domain 데이터들의
labeling이 필요 없는
unsupervised domain adaptation



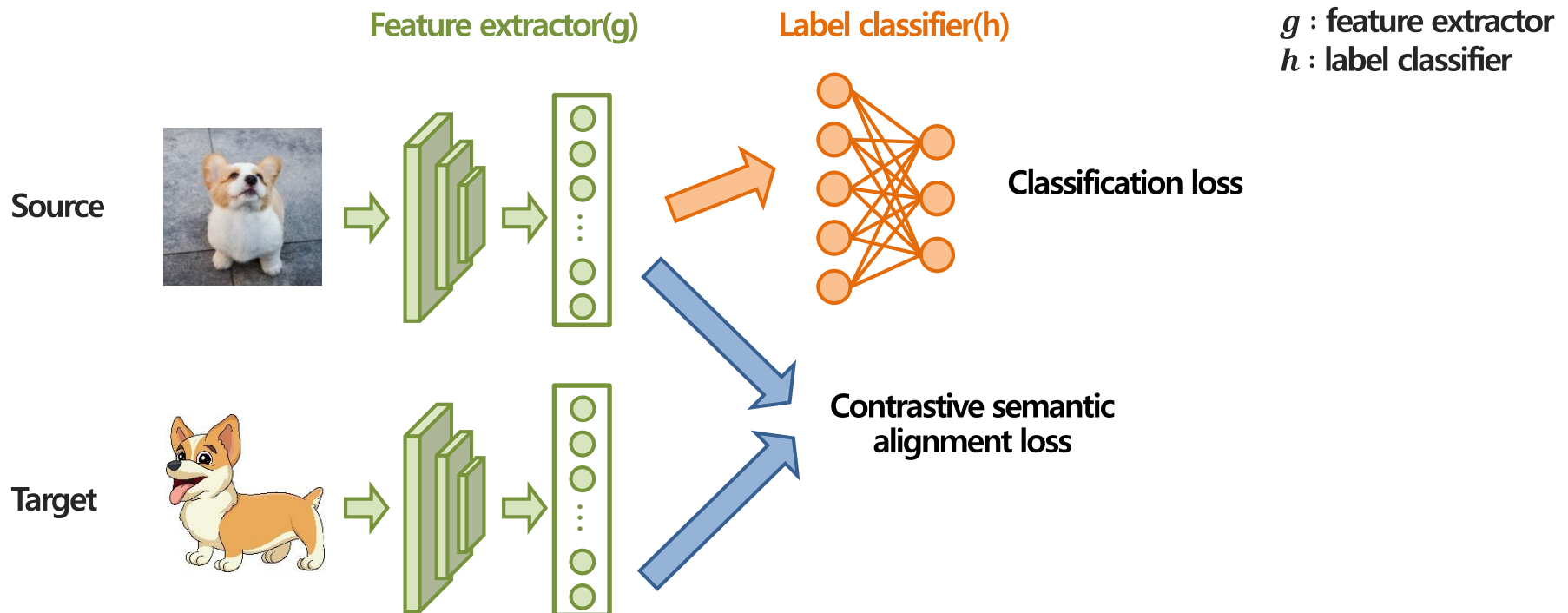
- ✓ 대량의 Target data 필요
- ✓ 동일한 양의 데이터라면
SDA가 일반적으로 우수한 성능

Deep Domain Adaptation

CCSA-loss

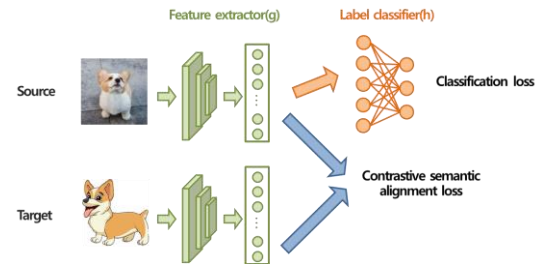
❖ Unified Deep Supervised Domain Adaptation and Generalization

- Target domain내 소수의 labeled data가 존재하는 경우 해당 labeled data를 활용하여 성능을 향상시키는 supervised domain adaptation(SDA) 방법론
- 서로 다른 두 도메인을 구분하는 능력인 domain divergence 를 세분화 하여 정의



Deep Domain Adaptation

CCSA-loss

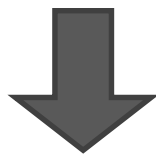


❖ Unified Deep Supervised Domain Adaptation and Generalization

- Target domain내 소수의 labeled data가 존재하는 경우 해당 labeled data를 활용하여 성능을 향상시키는 supervised domain adaptation(SDA) 방법론
- 서로 다른 두 도메인을 구분하는 능력인 domain divergence 를 세분화 하여 정의

g : feature extractor
 h : label classifier

CCSA-loss = Classification and Contrastive Semantic Alignment loss



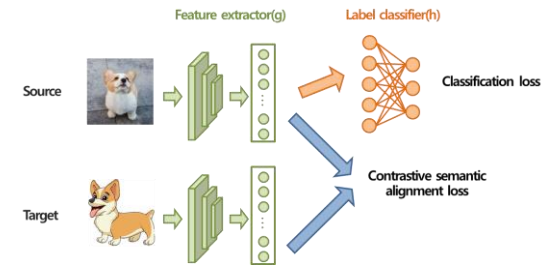
$$L_{CCSA}(h \circ g) = \underbrace{L_C(h \circ g)} + \underbrace{L_{SA}(g) + L_S(g)}$$

Source domain 에서의 분류 성능 ↑

Source ↔ Target domain 구분 성능 ↓

Deep Domain Adaptation

CCSA-loss



❖ Unified Deep Supervised Domain Adaptation and Generalization

- Target domain내 소수의 labeled data가 존재하는 경우 해당 labeled data를 활용하여 성능을 향상시키는 supervised domain adaptation(SDA) 방법론
- 서로 다른 두 도메인을 구분하는 능력인 domain divergence 를 세분화 하여 정의

g : feature extractor
 h : label classifier

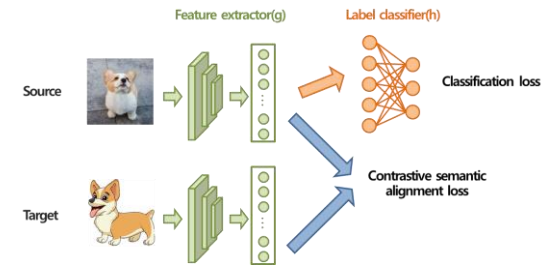
Semantic alignment loss : $L_{SA}(g) = \sum_{a=1}^C d(p(g(X_a^s)), p(g(X_a^t)))$

→ Class가 동일한 경우, source / target domain 사이 feature vector 간의 거리 ↓

$$d(g(x_i^s), g(x_j^t)) = \frac{1}{2} \|g(x_i^s) - g(x_j^t)\|^2$$

Deep Domain Adaptation

CCSA-loss



❖ Unified Deep Supervised Domain Adaptation and Generalization

- Target domain내 소수의 labeled data가 존재하는 경우 해당 labeled data를 활용하여 성능을 향상시키는 supervised domain adaptation(SDA) 방법론
- 서로 다른 두 도메인을 구분하는 능력인 domain divergence 를 세분화 하여 정의

g : feature extractor
 h : label classifier

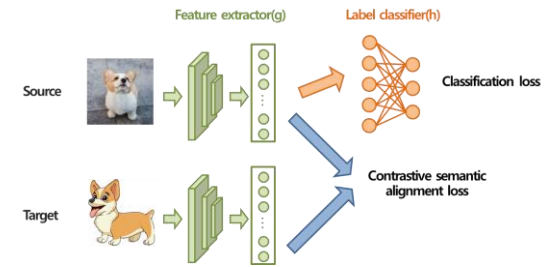
Separation loss : $L_S(g) = \sum_{a,b|a \neq b} k(p(g(X_a^s)), p(g(X_b^t)))$

→ Class가 다른 경우, source / target domain 사이 feature vector 간의 유사도 ↓ (\approx 거리 ↑)

$$k(g(x_i^s), g(x_j^t)) = \frac{1}{2} \max(0, m - \|g(x_i^s) - g(x_j^t)\|)^2$$

Deep Domain Adaptation

CCSA-loss



❖ Unified Deep Supervised Domain Adaptation and Generalization

- Target domain내 소수의 labeled data가 존재하는 경우 해당 labeled data를 활용하여 성능을 향상시키는 supervised domain adaptation(SDA) 방법론
- 서로 다른 두 도메인을 구분하는 능력인 domain divergence 를 세분화 하여 정의

g : feature extractor
 h : label classifier

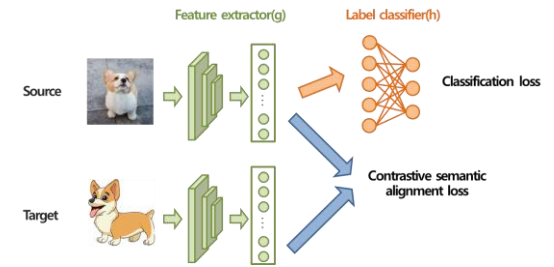
CCSA-loss = Classification and Contrastive Semantic Alignment loss

→ Class가 다른 경우, source / target domain 사이 feature vector 간의 유사도 ↓ (\approx 거리 ↑)

- ✓ Labeled source 데이터에 대한 분류 성능 ↑
- ✓ 동일 class 인 경우, domain 간의 분류 성능 ↓
- ✓ 다른 class 인 경우, domain 간의 분류 성능 ↑

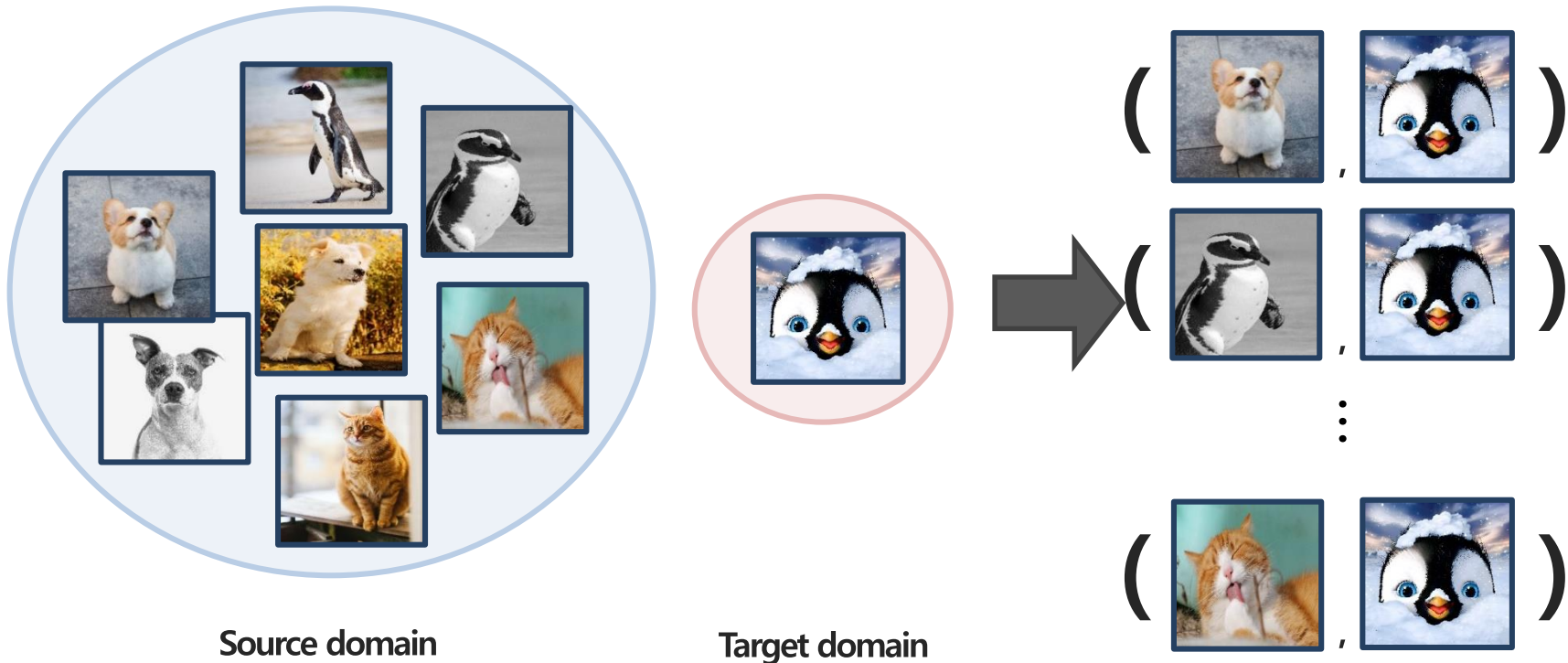
Deep Domain Adaptation

CCSA-loss



❖ Unified Deep Supervised Domain Adaptation and Generalization

- Supervised domain adaptation의 특성상 target domain내 labeled 데이터의 수가 적음
- 적은 수의 labeled target 데이터를 효율적으로 활용하기 위해, CCSA-loss 연산 과정에서 target 데이터를 모든 source 데이터와 쌍을 이루어서 학습



2-3. SagNet (Reducing Domain Gap by Reducing Style Bias)

Deep Domain Adaptation

SagNet

❖ Reducing Domain Gap by Reducing Style Bias

- CNN을 통한 특징 추출 과정에서 모양과 같은 content보다 style에 집중하는 bias가 존재하여 domain shift가 발생하는 경우 예측 성능의 하락
- Domain shift에 큰 영향을 미치는 style bias를 줄이는 방향으로 모델 개발



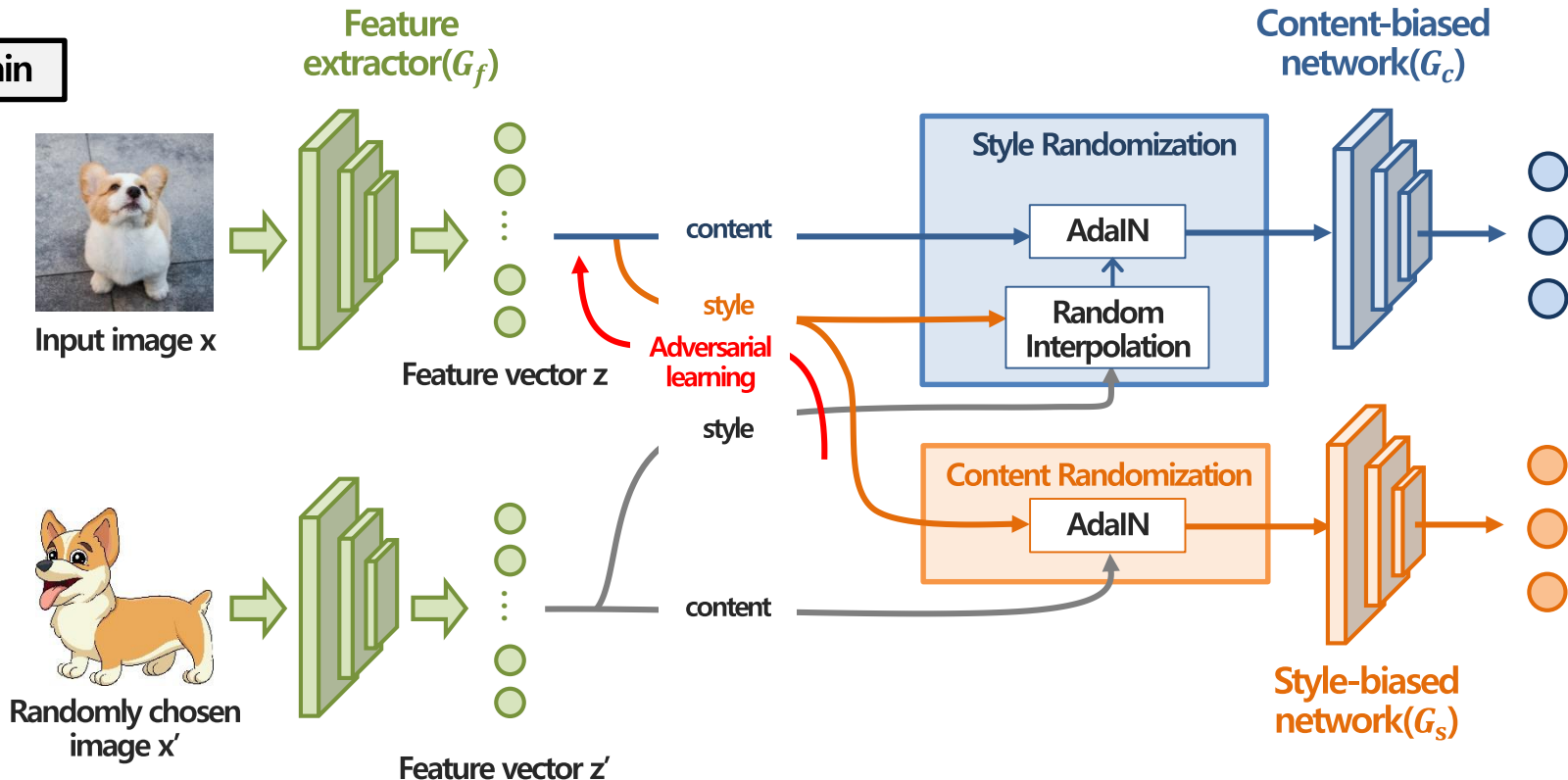
Deep Domain Adaptation

SagNet

❖ Reducing Domain Gap by Reducing Style Bias

- Feature extractor (G_f): 데이터의 content와 style을 포함하는 feature vector 추출
- Content-biased network (G_c): feature vector에 대하여 style을 랜덤하게 하여 **content에 집중한 예측**
- Style-biased network (G_s): feature vector에 대하여 content를 랜덤하게 하여 **style에 집중한 예측**

Train



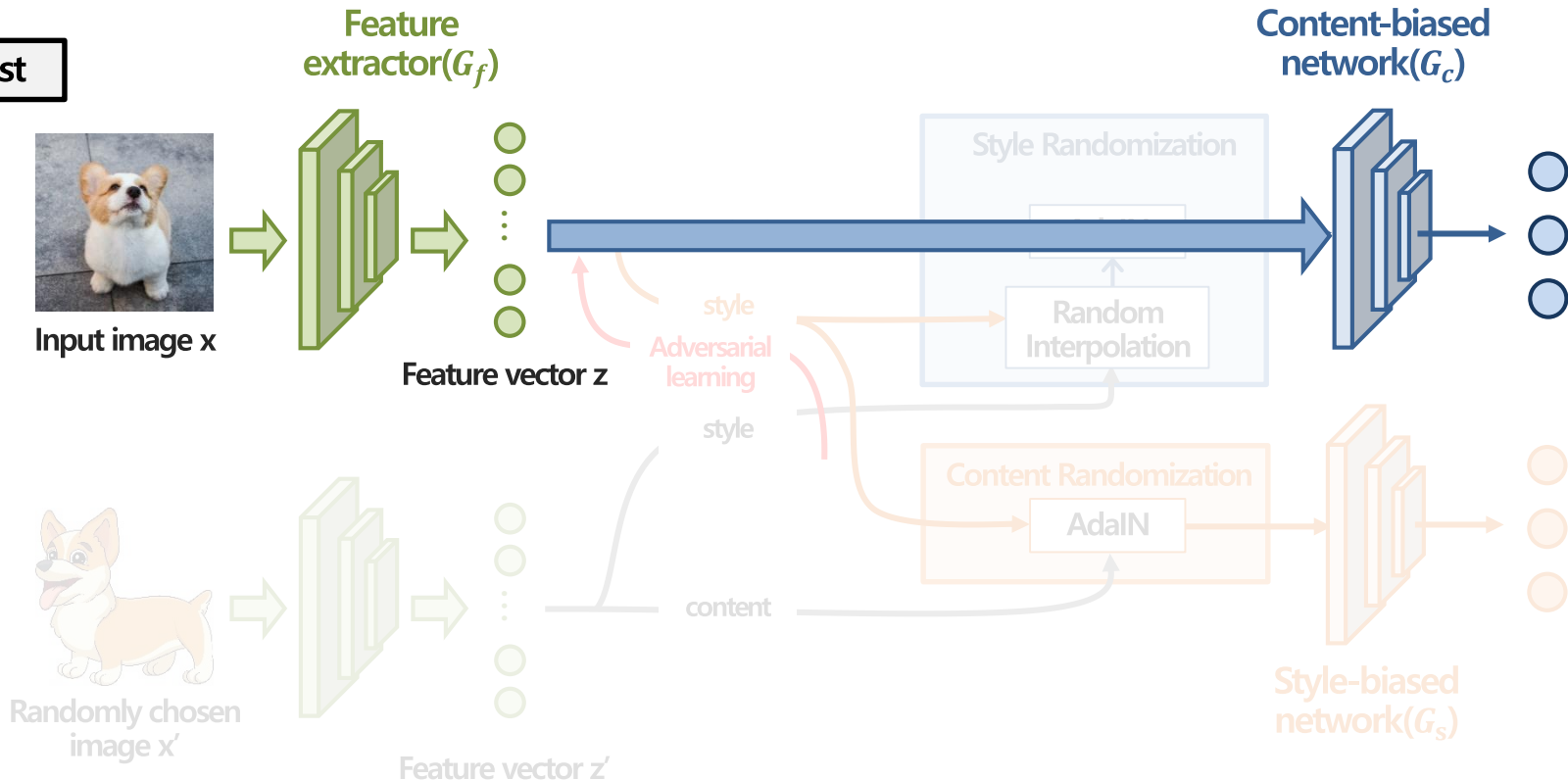
Deep Domain Adaptation

SagNet

❖ Reducing Domain Gap by Reducing Style Bias

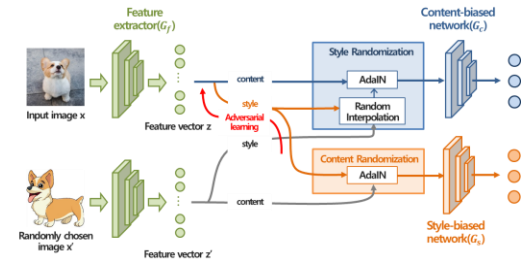
- Feature extractor (G_f): 데이터의 content와 style을 포함하는 feature vector 추출
- Content-biased network (G_c): feature vector에 대하여 style을 랜덤하게 하여 **content에 집중한 예측**
- Style-biased network (G_s): feature vector에 대하여 content를 랜덤하게 하여 **style에 집중한 예측**

Test



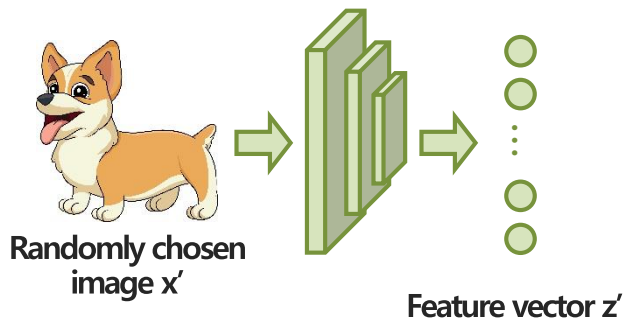
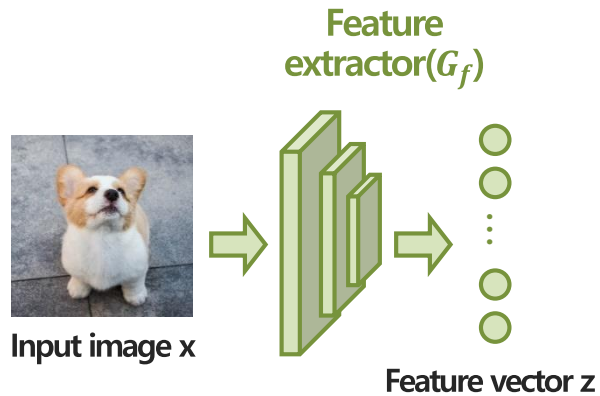
Deep Domain Adaptation

SagNet



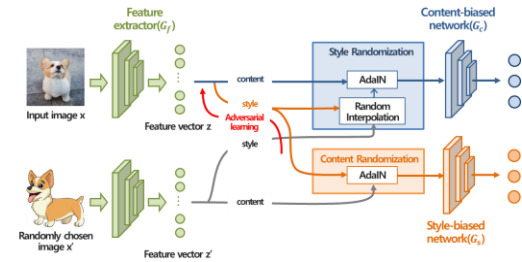
❖ Content-Biased Learning

- Feature extractor를 통하여 input image x 와 랜덤하게 선택된 x' 의 feature vector z, z' 을 추출
- Feature vector에서 style을 표현하는 평균(μ)과 표준편차(σ)를 channel 별로 계산



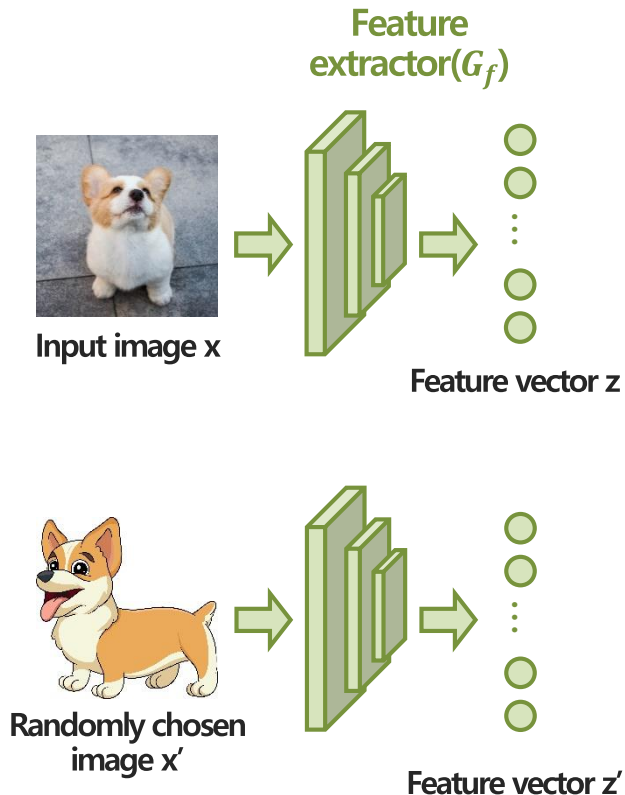
Deep Domain Adaptation

SagNet



❖ Content-Biased Learning

- Feature extractor를 통하여 input image x 와 랜덤하게 선택된 x' 의 feature vector z, z' 을 추출
- Feature vector에서 style을 표현하는 평균(μ)과 표준편차(σ)를 channel 별로 계산

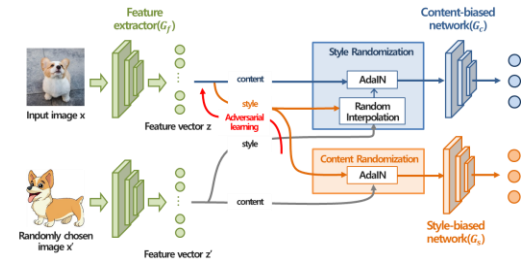


$$\mu(z) = \frac{1}{HW} \sum_{h=1}^H \sum_{w=1}^W z_{hw}$$

$$\sigma(z) = \sqrt{\frac{1}{HW} \sum_{h=1}^H \sum_{w=1}^W (z_{hw} - \mu(z))^2 + \varepsilon}$$

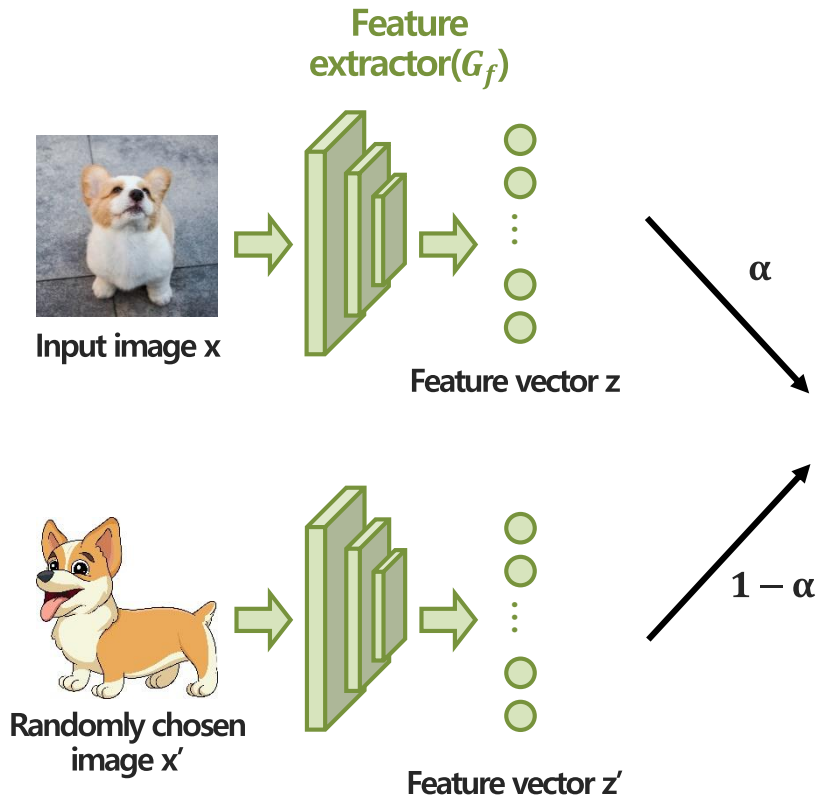
Deep Domain Adaptation

SagNet



❖ Content-Biased Learning

- Feature vector z, z' 의 style $\mu(z), \sigma(z)$ 와 $\mu(z'), \sigma(z')$ 을 선형 결합하여 random style $\hat{\mu}, \hat{\sigma}$ 을 계산
- Adaptive instance normalization(AdaIN)을 활용하여 Input image의 content는 유지하며, style을 random style로 변경

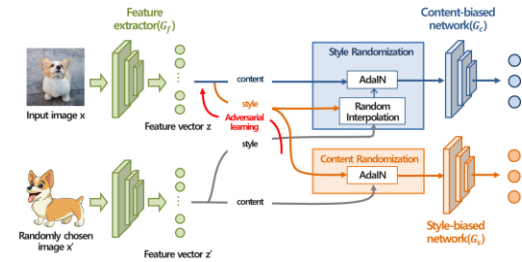


$$\text{Random style } \hat{\mu} = \alpha \cdot \mu(z) + (1 - \alpha) \cdot \mu(z')$$

$$\text{Random style } \hat{\sigma} = \alpha \cdot \sigma(z) + (1 - \alpha) \cdot \sigma(z')$$

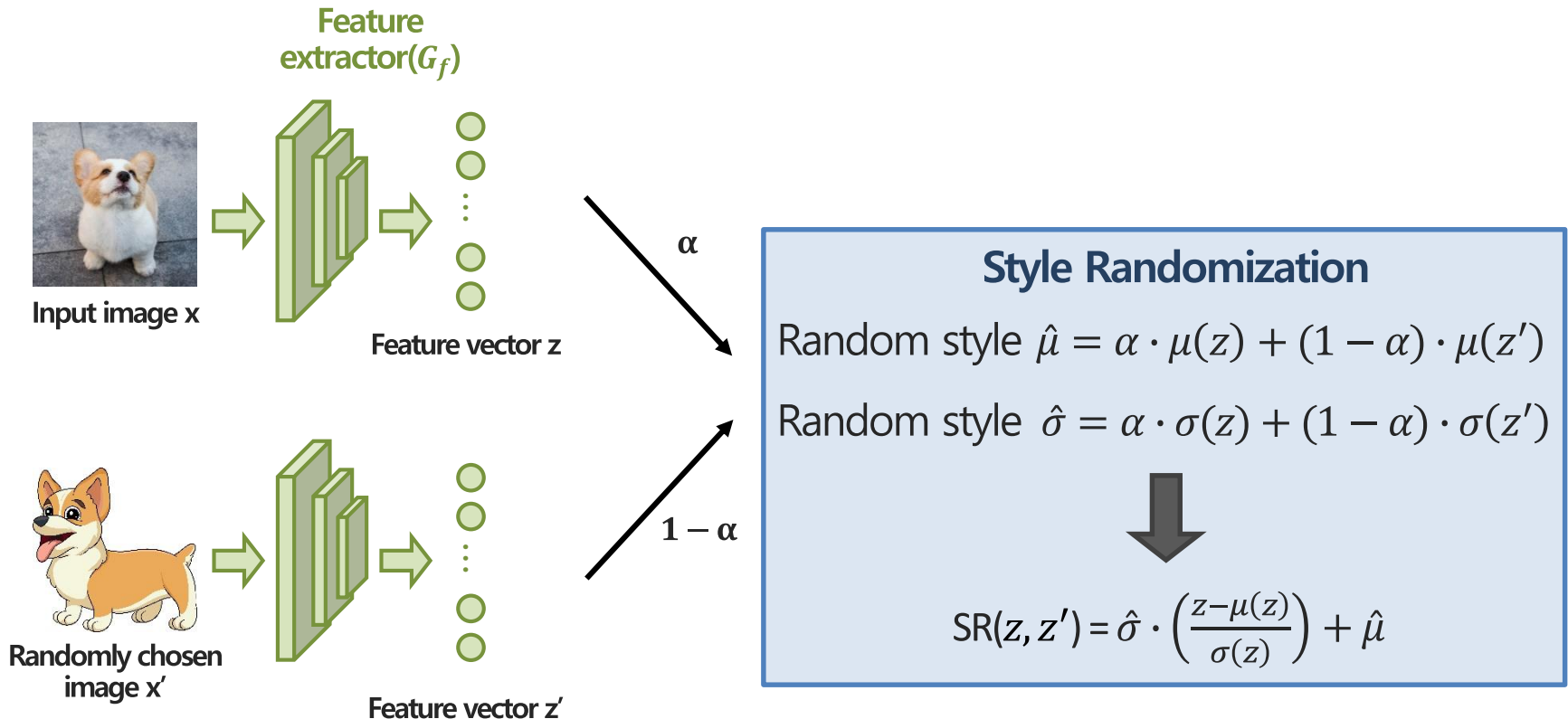
Deep Domain Adaptation

SagNet



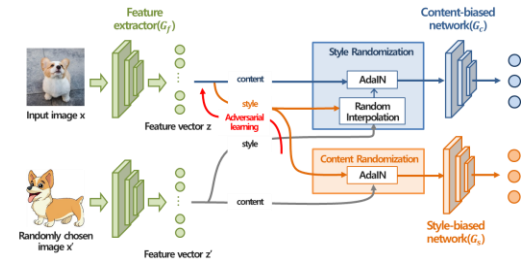
❖ Content-Biased Learning

- Feature vector z, z' 의 style $\mu(z), \sigma(z)$ 와 $\mu(z'), \sigma(z')$ 을 선형 결합하여 random style $\hat{\mu}, \hat{\sigma}$ 을 계산
- Adaptive instance normalization(AdaIN)을 활용하여 Input image의 content는 유지하며, style을 random style로 변경



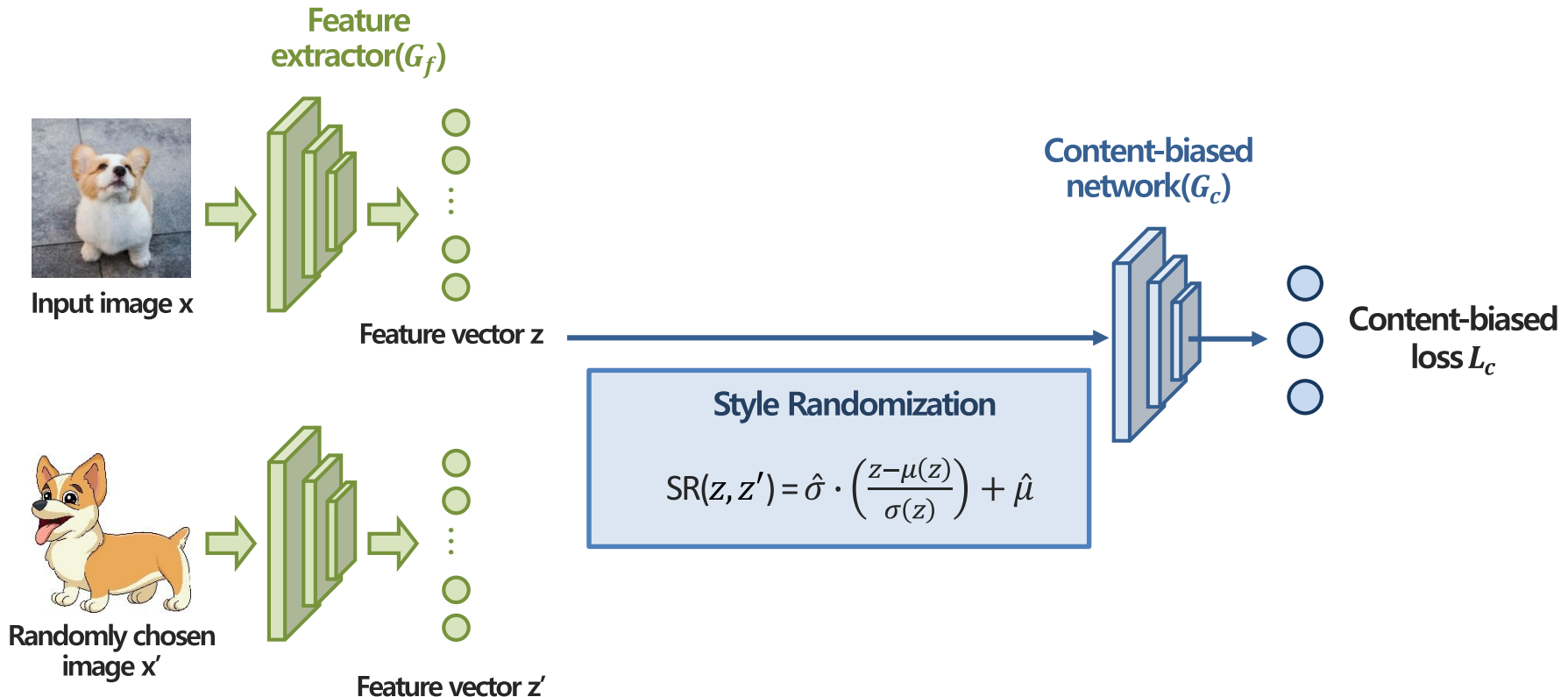
Deep Domain Adaptation

SagNet



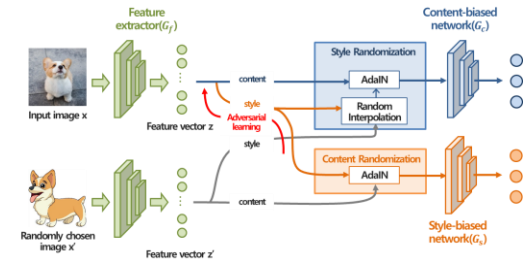
❖ Content-Biased Learning

- Content 정보는 유지되며, random style로 변경된 feature vector를 content-biased network를 통하여 content-biased prediction 진행
- 실제 label과 content-biased prediction label 사이의 content-biased loss L_c



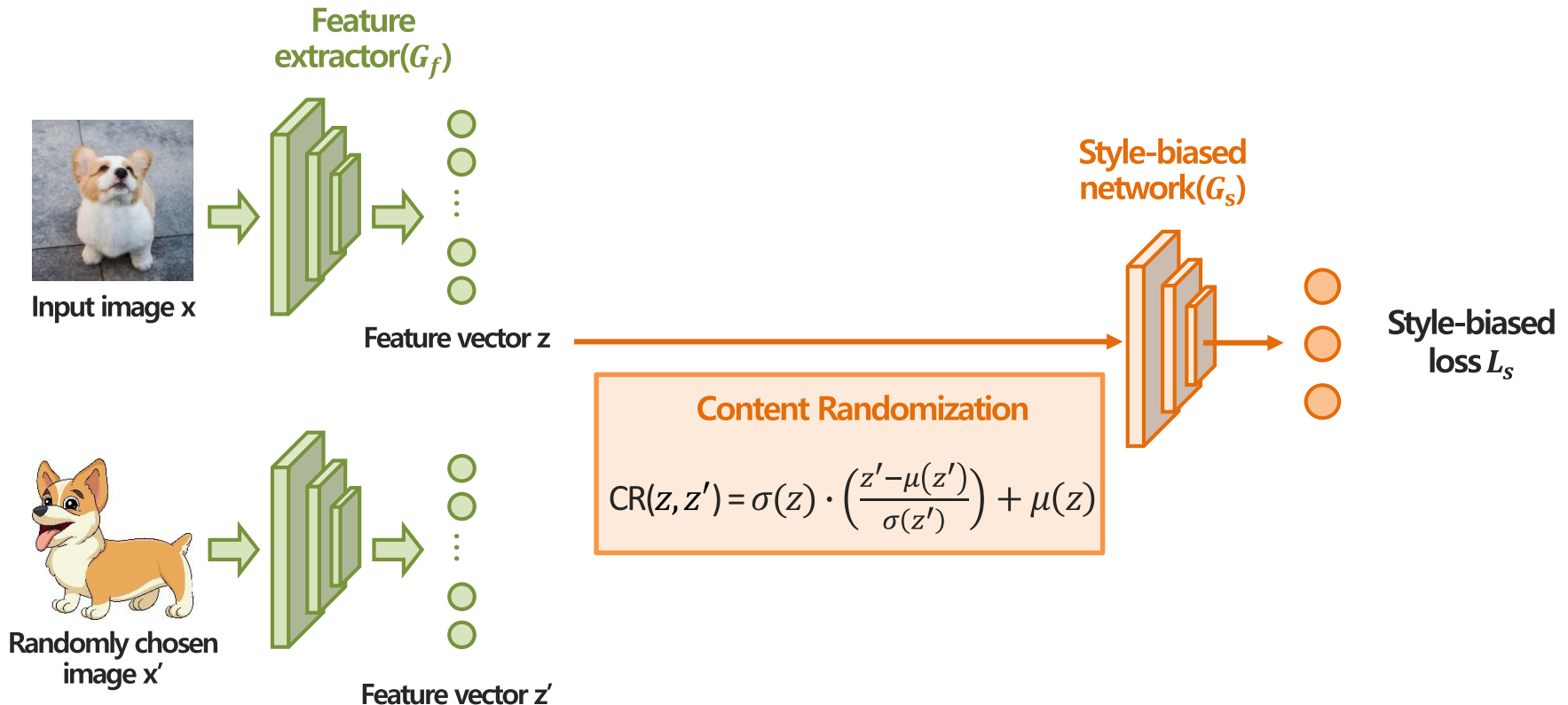
Deep Domain Adaptation

SagNet



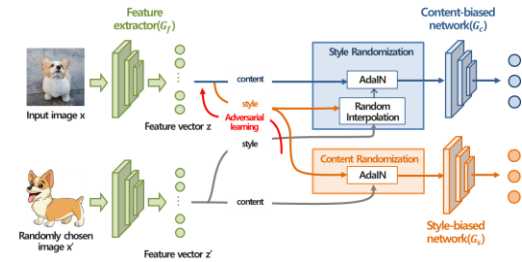
❖ Adversarial Style-Biased Learning

- Feature extractor를 통하여 input image x 와 랜덤하게 선택된 x' 의 feature vector z, z' 을 추출
- Adaptive instance normalization(AdaIN)을 활용하여 z' 의 content는 유지하며, z 의 style을 적용
- Style-biased network를 통하여 style-biased prediction 진행 및 실제 label과 style-biased loss L_s



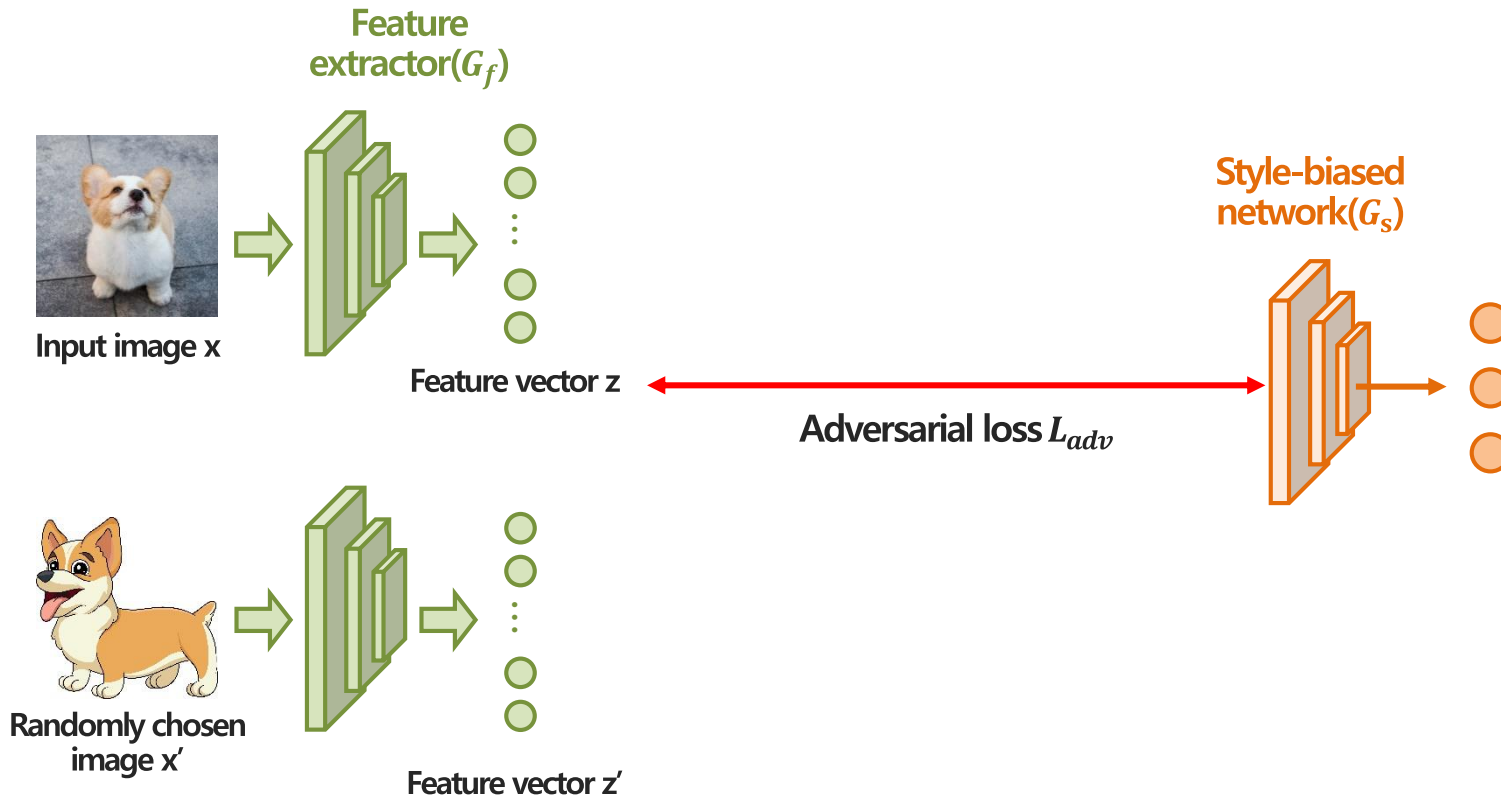
Deep Domain Adaptation

SagNet



❖ Adversarial Style-Biased Learning

- Style에 집중하지 않도록 style-bias가 적은 모델 구축이 목적
- Style에 집중하는 style-biased network를 통한 예측과 uniform 분포와의 adversarial loss L_{adv}
 - Feature extractor가 style-biased network를 속이도록 학습



3. Conclusions

Conclusions

❖ Conclusions

- Domain adaptation은 기존의 정보들을 활용하여 새로운 도메인에서도 우수한 성능을 유지하기 위한 연구 분야
- Source 데이터들의 class에 대해서는 예측 에러를 감소시키며, domain 사이에 대해서는 분류하지 못하도록 학습
- Transfer learning의 한 종류로 동일한 task에 대하여 활용 데이터의 domain이 서로 다른 경우를 해결하는 분야로 실제 세상에서 자주 일어나는 환경이기에 집중되는 연구 분야

References

1. Kouw, Wouter M., and Marco Loog. "A review of domain adaptation without target labels." IEEE transactions on pattern analysis and machine intelligence 43.3 (2019): 766-785.
2. Ben-David, Shai, et al. "A theory of learning from different domains." Machine learning 79.1 (2010): 151-175.
3. Ganin, Yaroslav, et al. "Domain-adversarial training of neural networks." The journal of machine learning research 17.1 (2016): 2096-2030.
4. Motiian, Saeid, et al. "Unified deep supervised domain adaptation and generalization." Proceedings of the IEEE international conference on computer vision. 2017.
5. Nam, Hyeonseob, et al. "Reducing domain gap by reducing style bias." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

Thank you