Domain Generalization :
How to improve the generalization ability of deep learning models?

DMQA Open Seminar (2023.07.21)

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
- Semi-supervised Learning

❖ **Contact**

- jihyun_k@korea.ac.kr

Contents

230721 DMQA Open Seminar:
Domain Generalization : How to improve the generalization ability of deep learning models?

1. Introduction: Domain Generalization

Introduction

Background of Domain Generalization

일반화(Generalization) 성능이 좋은 모델을 만들어보자!

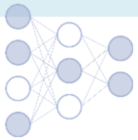
Train

Data Augmentation

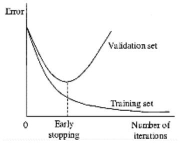
Batch Normalization

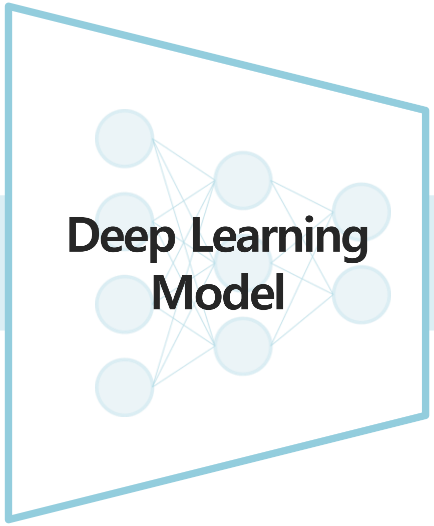
$$\frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}$$

Dropout



Early-stopping



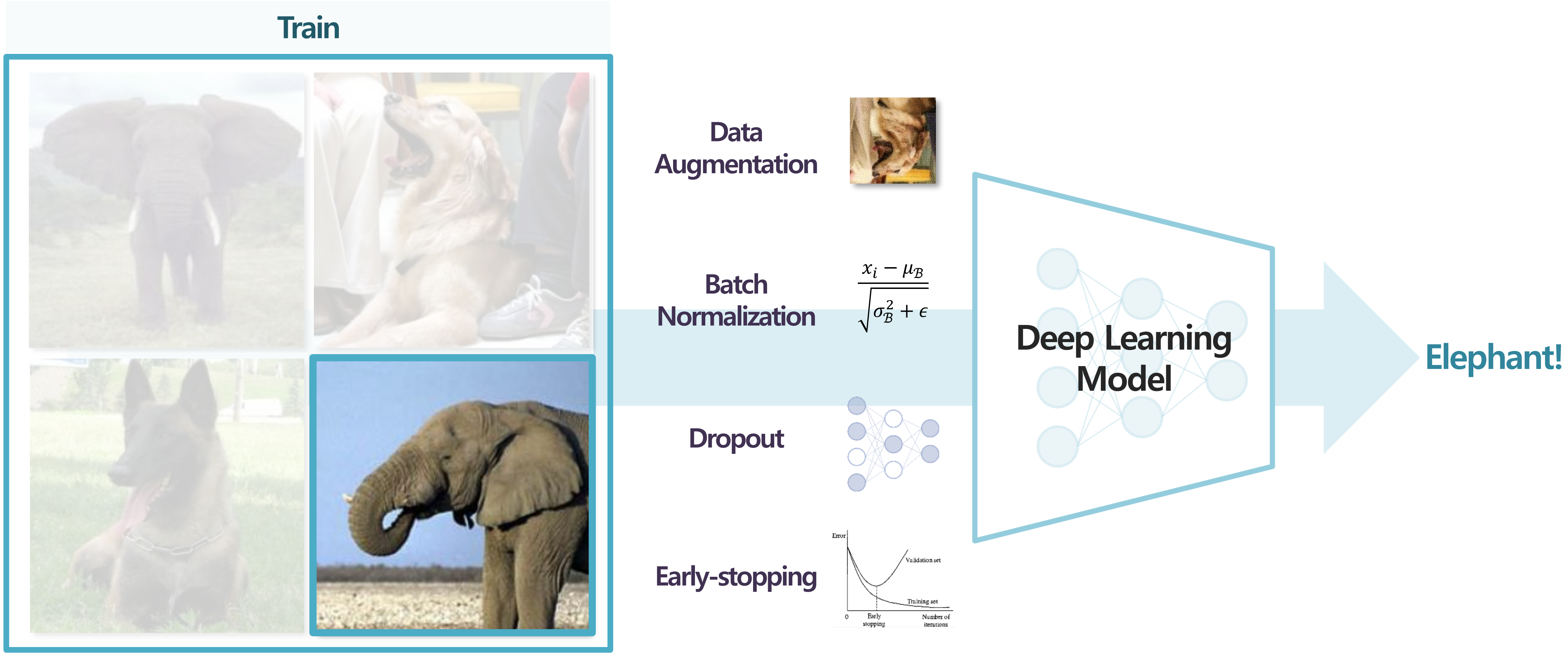


Elephant!

Introduction

Background of Domain Generalization

일반화(Generalization) 성능이 좋은 모델을 만들어보자!

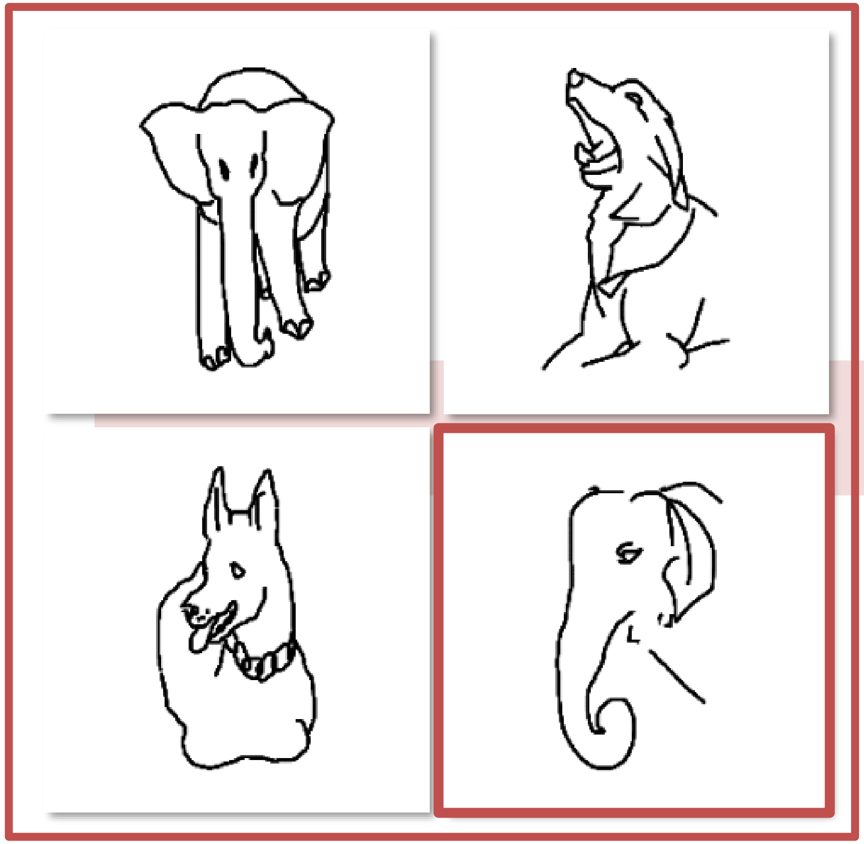


Introduction

Background of Domain Generalization

일반화(Generalization) 성능이 좋은 모델이 만들어졌을까?

Test

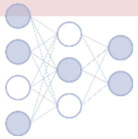


Data Augmentation

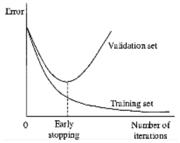
Batch Normalization

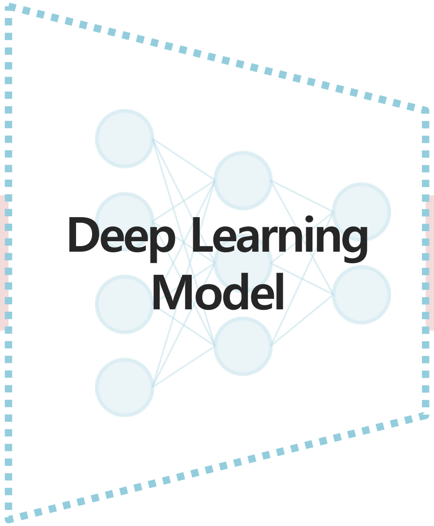
$$\frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}$$

Dropout



Early-stopping



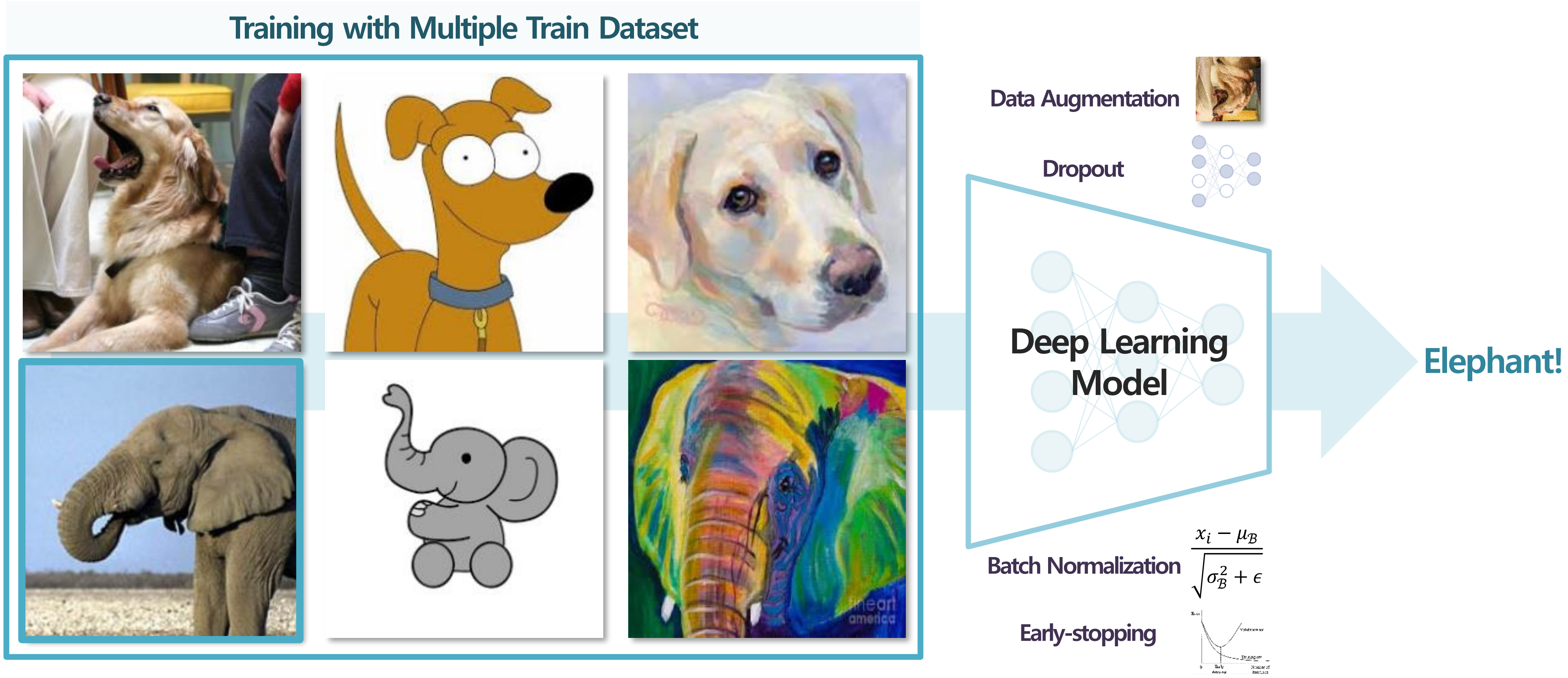


Elephant...?

Introduction

Background of Domain Generalization

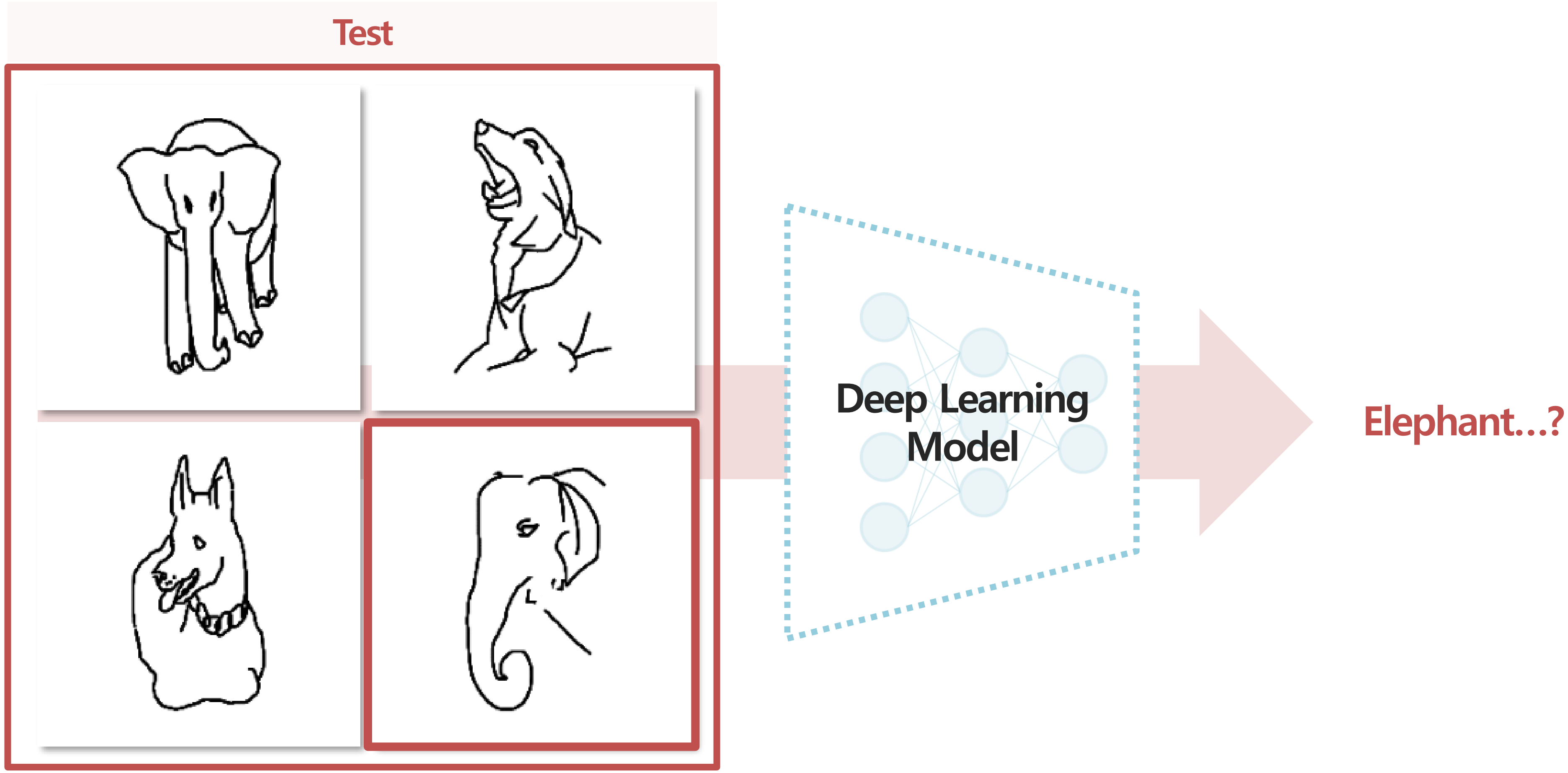
다시 한 번 더 일반화(Generalization) 성능이 좋은 모델을 만들어보자!



Introduction

Background of Domain Generalization

다시 한 번 더 일반화(Generalization) 성능이 좋은 모델을 만들어보자!



Introduction

Background of Domain Generalization

학습 데이터와 테스트 데이터의 분포가 다른 문제

Train

Test

Introduction

Background of Domain Generalization

학습 데이터와 테스트 데이터의 분포가 다른 문제

Violate i.i.d assumption, i.e.,
Training and Testing data are NOT
identically and independently distributed.

같은 분포로부터 수집되는 데이터

Domain shift!

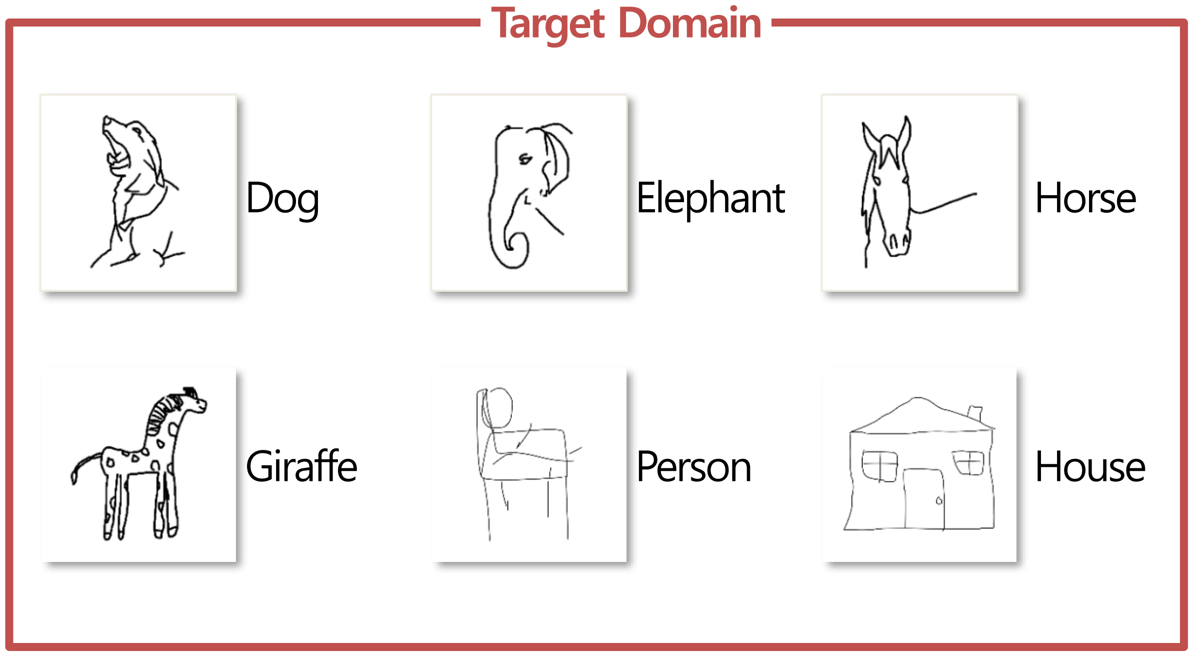
$\neq$

Source Domain Target Domain

Introduction

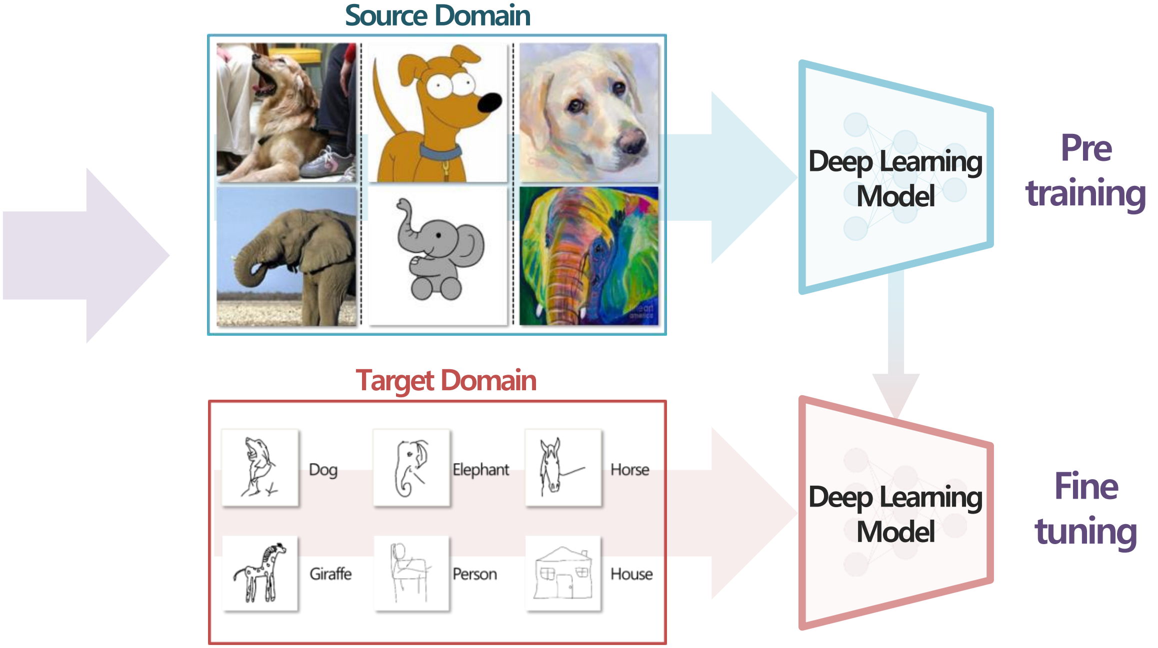
Background of Domain Generalization

❖ Generalization-related Research Topics



$\{X_t, Y_t\}$

Pretraining and Fine-tuning

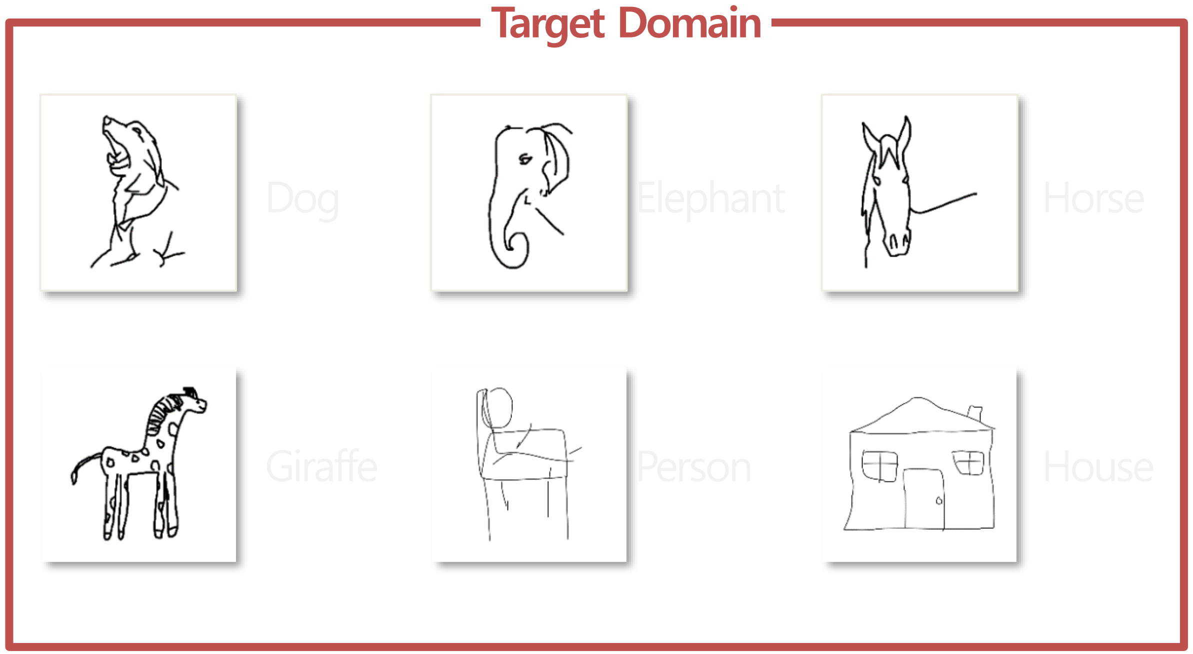


Limitation	Large-scale data collection and annotation for every new target domain is prohibitively expensive and time-consuming

Introduction

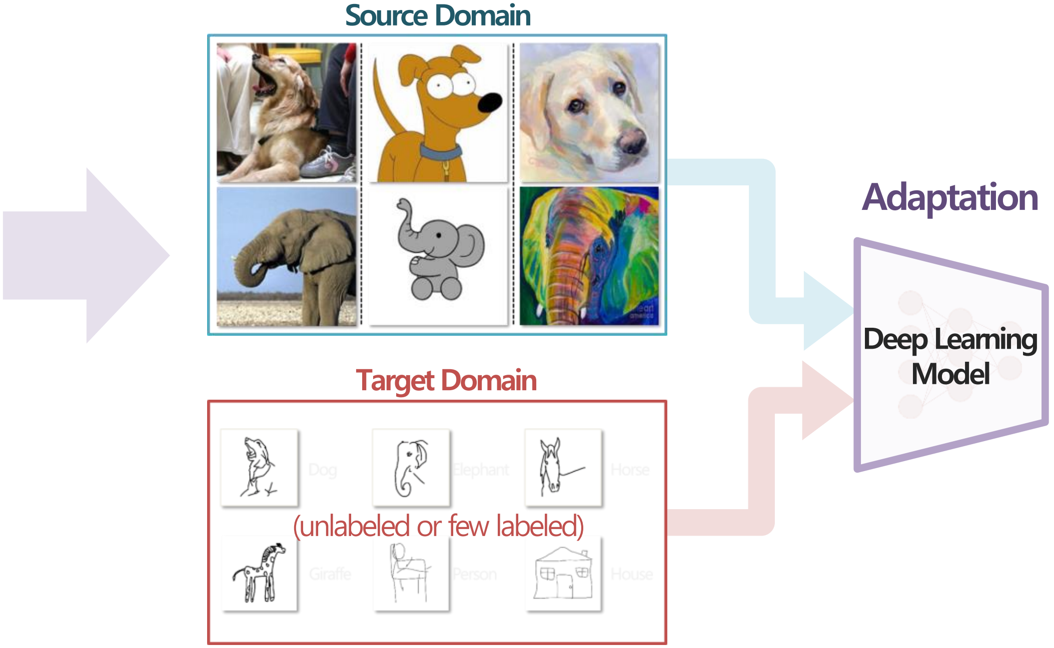
Background of Domain Generalization

❖ Generalization-related Research Topics



$\{X_t, Y_t\}$

(Unsupervised) Domain Adaptation



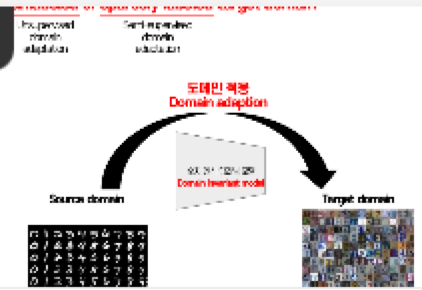
Limitation	Still requires a data collection step followed by a model adaptation step for each new domain, which might hinder its applicability.

Introduction

Background of Domain Generalization

❖ Generalization-related Research Topics

종료



Introduction to unsupervised domain ad

발표자: 이민정

2022년 9월 16일

오후 1시 ~

온라인 비디오 시청 (YouTube)

세미나 정보 보기 ⟶

종료

How to Transfer Knowledge Across Domains by Deep Neural Network?

2022. 10. 28

How to Transfer Knowledge Across Doma

발표자: 김지현

2022년 10월 28일

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 ~

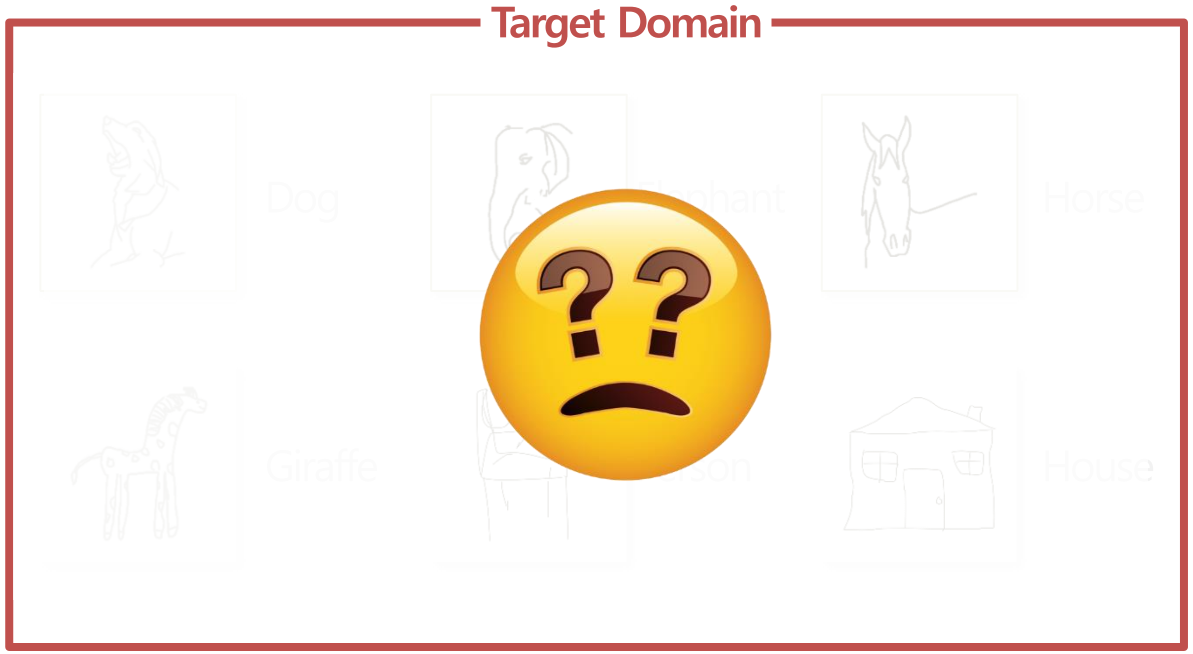
온라인 비디오 시청 (YouTube)

세미나 정보 보기 ⟶

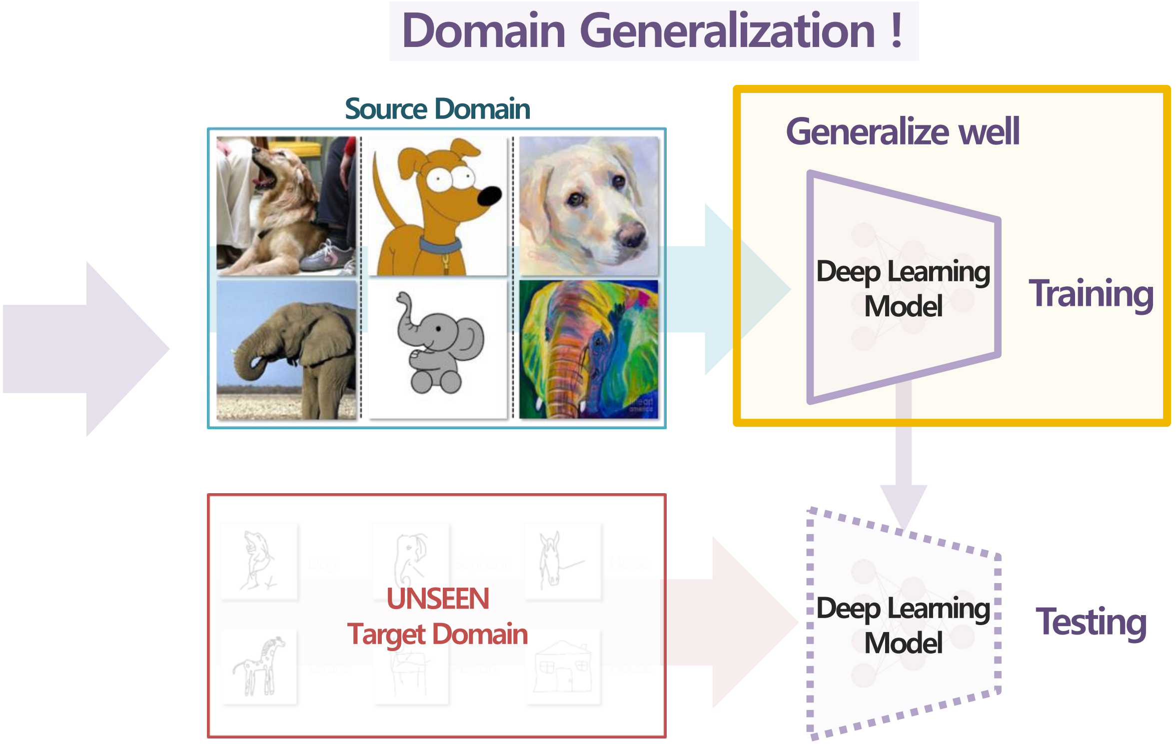
Introduction

Background of Domain Generalization

❖ Generalization-related Research Topics



None



Train a model using data from source domains and deploy model to an arbitrary unseen target domain!

Introduction

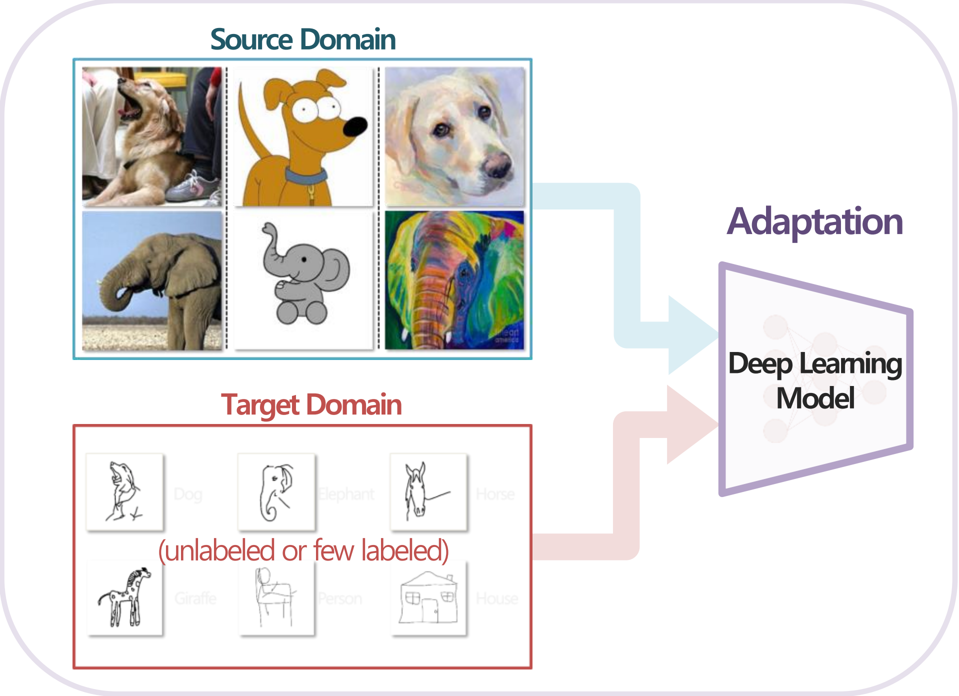
Background of Domain Generalization

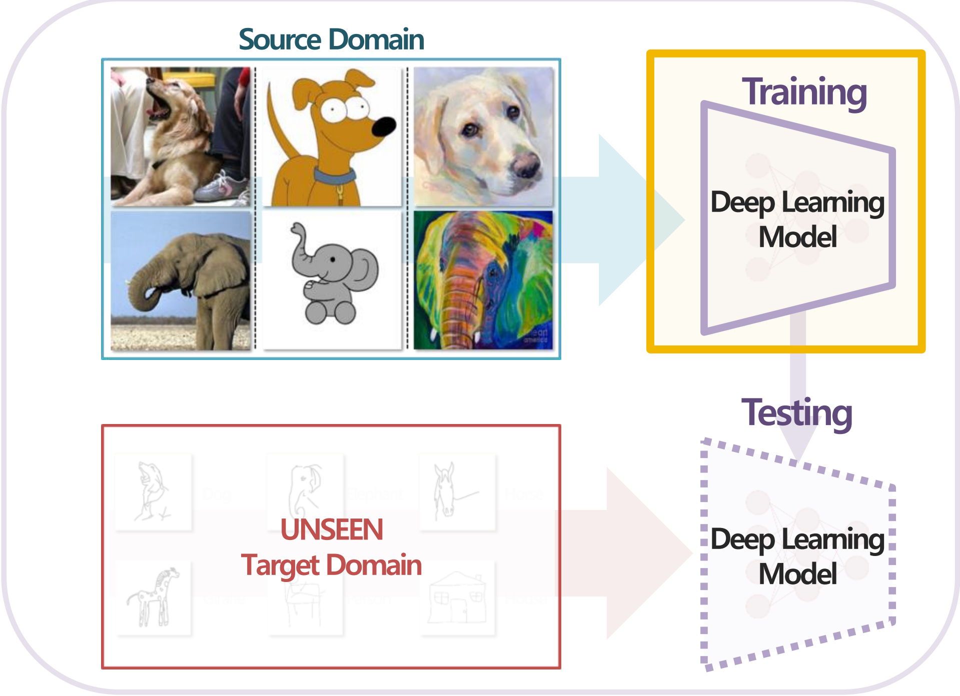
❖ Generalization-related Research Topics

Domain Adaptation

vs.

Domain Generalization

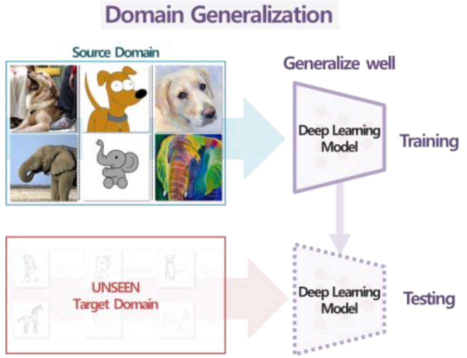




목적 : Target Domain에 대한 성능을 높이자!

Introduction

Preliminaries



❖ Problem Definition

Source Domain $\mathcal{D}_S = \{(x_i, y_i)\}_{i=1}^{N_S} \sim P_{XY}$

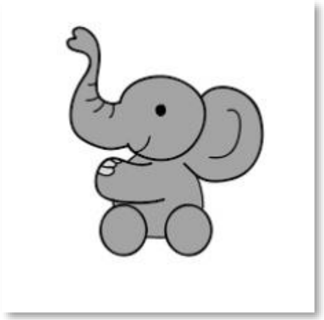
Labeled Source Domain을 상정

Dog

Elephant

Dog

Elephant



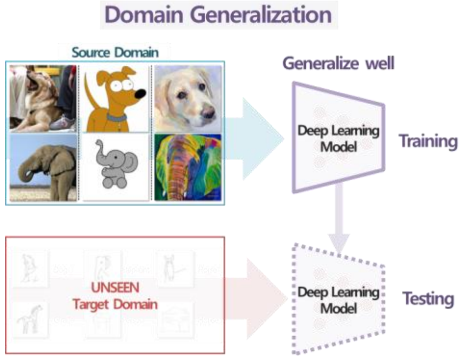
• • •

Dog

Elephant

Introduction

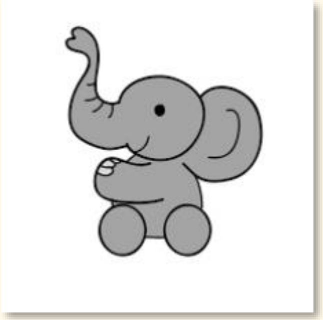
Preliminaries



❖ Problem Definition

Multi-Source Domain $\mathcal{D}_S = \{\mathcal{D}_S^k\}_{k=1}^K,\ \mathcal{D}_S^k = \{(x_i^k, y_i^k)\}_{i=1}^{N_S^k} \sim P^k{}_{XY}$

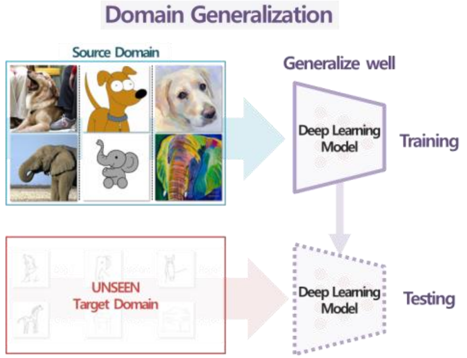
$P^1{}_{XY}$ $\neq$ $P^2{}_{XY}$ $\neq$ $P^K{}_{XY}$



…

Introduction

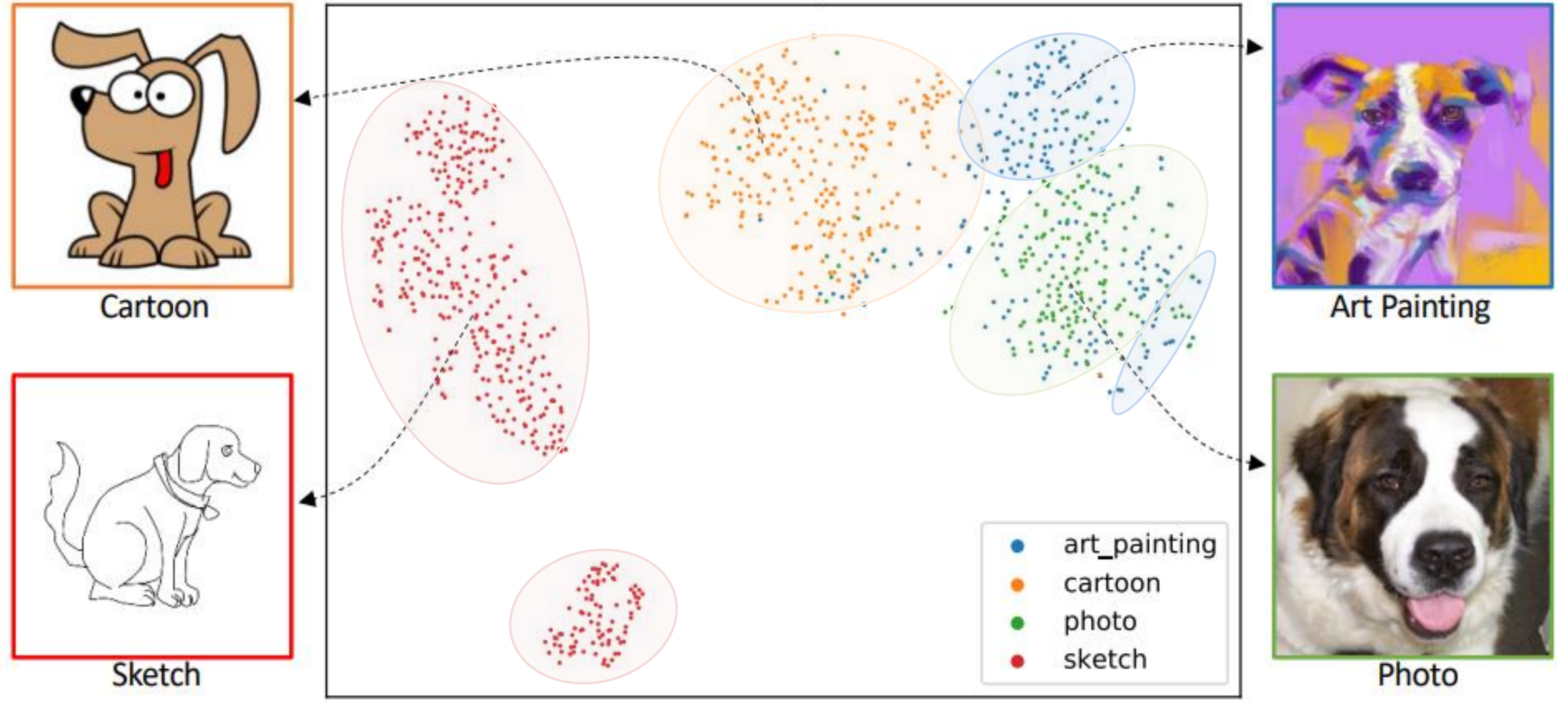
Preliminaries



❖ Problem Definition

Multi-Source Domain $\mathcal{D}_S = \{\mathcal{D}_S^k\}_{k=1}^{K},\ \mathcal{D}_S^k = \{(x_i^k, y_i^k)\}_{i=1}^{N_S^k} \sim P^k{}_{XY}$

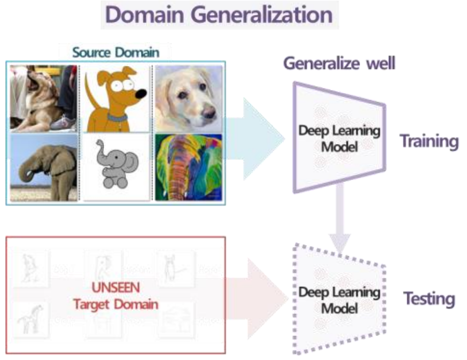
Visualization of style statistics [1]



[1] Zhou, K., Yang, Y., Qiao, Y., & Xiang, T. (2021). Domain generalization with mixstyle. arXiv preprint arXiv:2104.02008.

Introduction

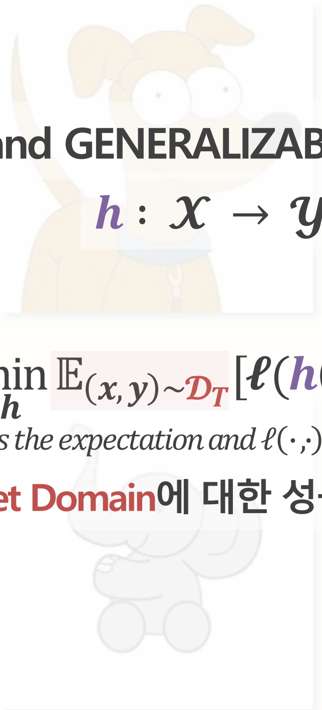
Preliminaries



❖ Problem Definition

Multi-Source Domain $\mathcal{D}_S = \{\mathcal{D}_S^k\}_{k=1}^K,\ \mathcal{D}_S^k = \{(x_i^k, y_i^k)\}_{i=1}^{N_S^k} \sim P^k{}_{XY}$

$P^1{}_{XY} \neq P^2{}_{XY} \neq \cdots P^K{}_{XY}$



Learn a ROBUST and GENERALIZABLE prediction function h

$$h : \mathcal{X} \rightarrow \mathcal{Y}$$

$$\min_{h} \mathbb{E}_{(x,y)\sim \mathcal{D}_T}[\ell(h(x), y)],$$

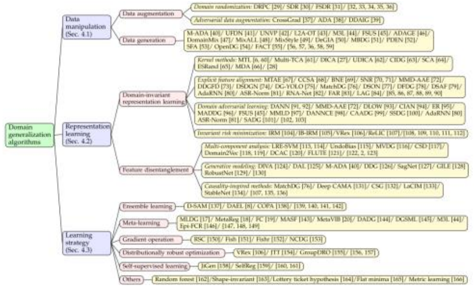
where $\mathbb{E}$ *is the expectation and* $\ell(\cdot,\cdot)$ *is the loss fuction.*

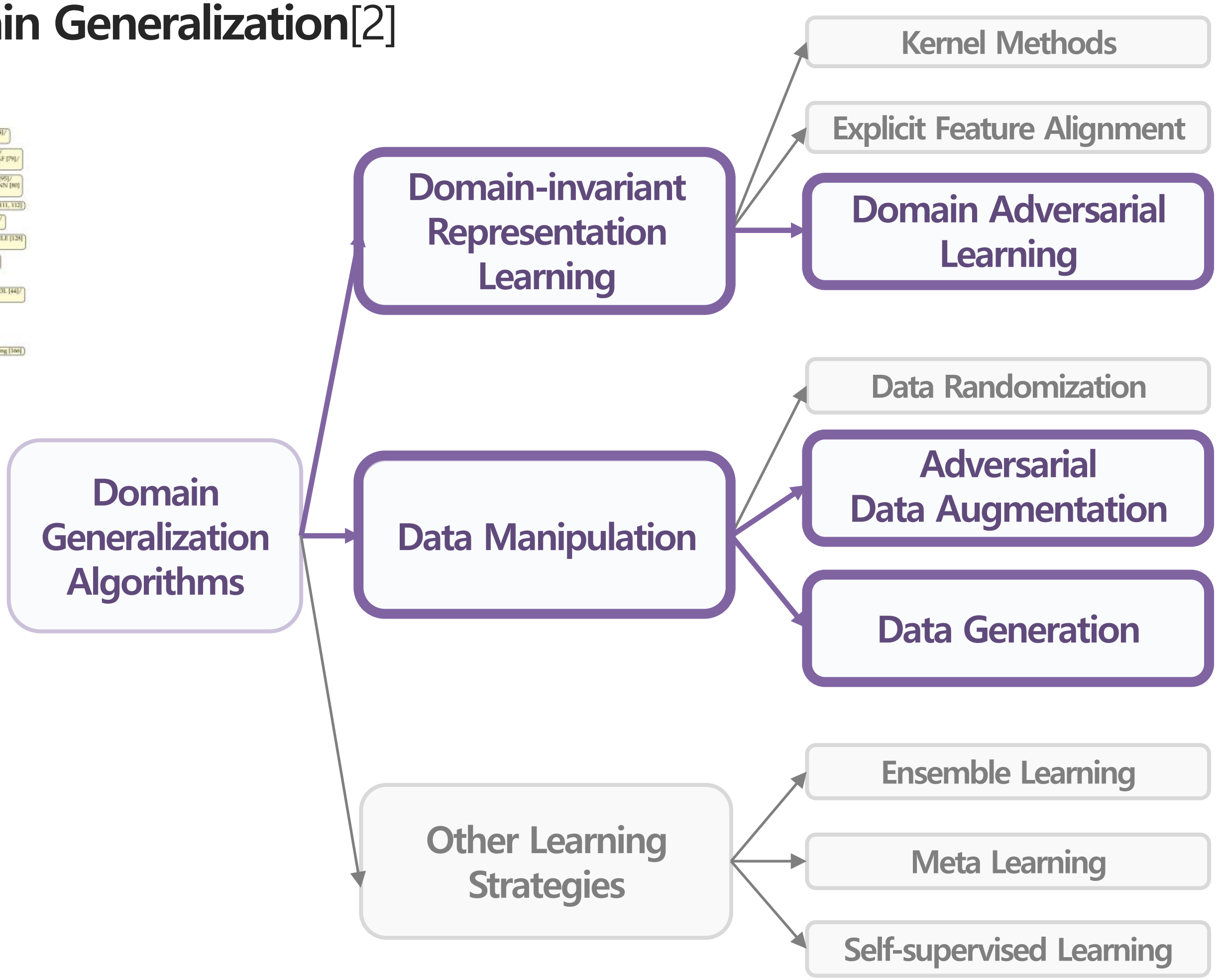
Target Domain에 대한 성능을 높이자!

Introduction

Preliminaries

❖ **Taxonomy of Domain Generalization**[2]

[2] 



[2] Wang, J., Lan, C., Liu, C., Ouyang, Y., Qin, T., Lu, W., ... & Yu, P. (2022). Generalizing to unseen domains: A survey on domain generalization. IEEE Transactions on Knowledge and Data Engineering.

230721 DMQA Open Seminar:
Domain Generalization : How to improve the generalization ability of deep learning models?

2. Methods

(1) Domain-invariant Representation Learning

Methods: Preliminaries

Domain Adversarial Learning

❖ Domain Adversarial Training of Neural Networks (DANN)[3]

- 2016년에 제안된 Domain Adaptation 방법론 (JMLR, 23년 7월 기준 7116회 인용)
- GAN[4]에서 제안된 '적대적 학습(Adversarial Learning)' 개념을 차용하여 Domain 간 차이 최소화

Domain-Adversarial Training of Neural Networks

Abstract

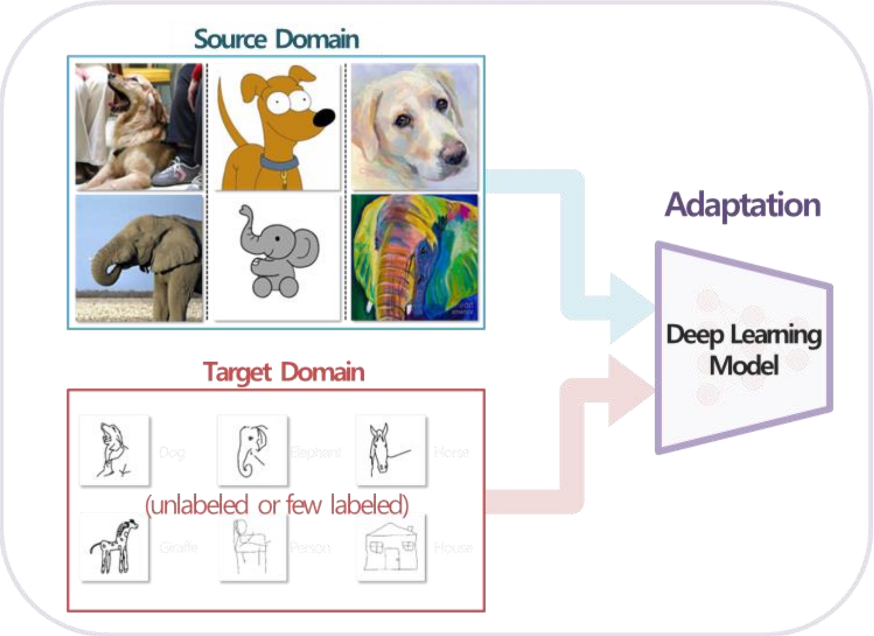
We introduce a new representation learning approach for domain adaptation, in which data at training and test time come from similar but different distributions. Our approach is directly inspired by the theory on domain adaptation suggesting that, for effective domain transfer to be achieved, predictions must be made based on features that cannot discriminate between the training (source) and test (target) domains.

The approach implements this idea in the context of neural network architectures that are trained on labeled data from the source domain and unlabeled data from the target domain (no labeled target-domain data is necessary). As the training progresses, the approach promotes the emergence of features that are (i) discriminative for the main learning task on the source domain and (ii) indiscriminate with respect to the shift between the domains. We show that this adaptation behaviour can be achieved in almost any feed-forward model by augmenting it with few standard layers and a new *gradient reversal* layer. The resulting augmented architecture can be trained using standard backpropagation and stochastic gradient descent, and can thus be implemented with little effort using any of the deep learning packages.

We demonstrate the success of our approach for two distinct classification problems (document sentiment analysis and image classification), where state-of-the-art domain adaptation performance on standard benchmarks is achieved. We also validate the approach for descriptor learning task in the context of person re-identification application.

Keywords: domain adaptation, neural network, representation learning, deep learning, synthetic data, image classification, sentiment analysis, person re-identification

Domain Adaptation



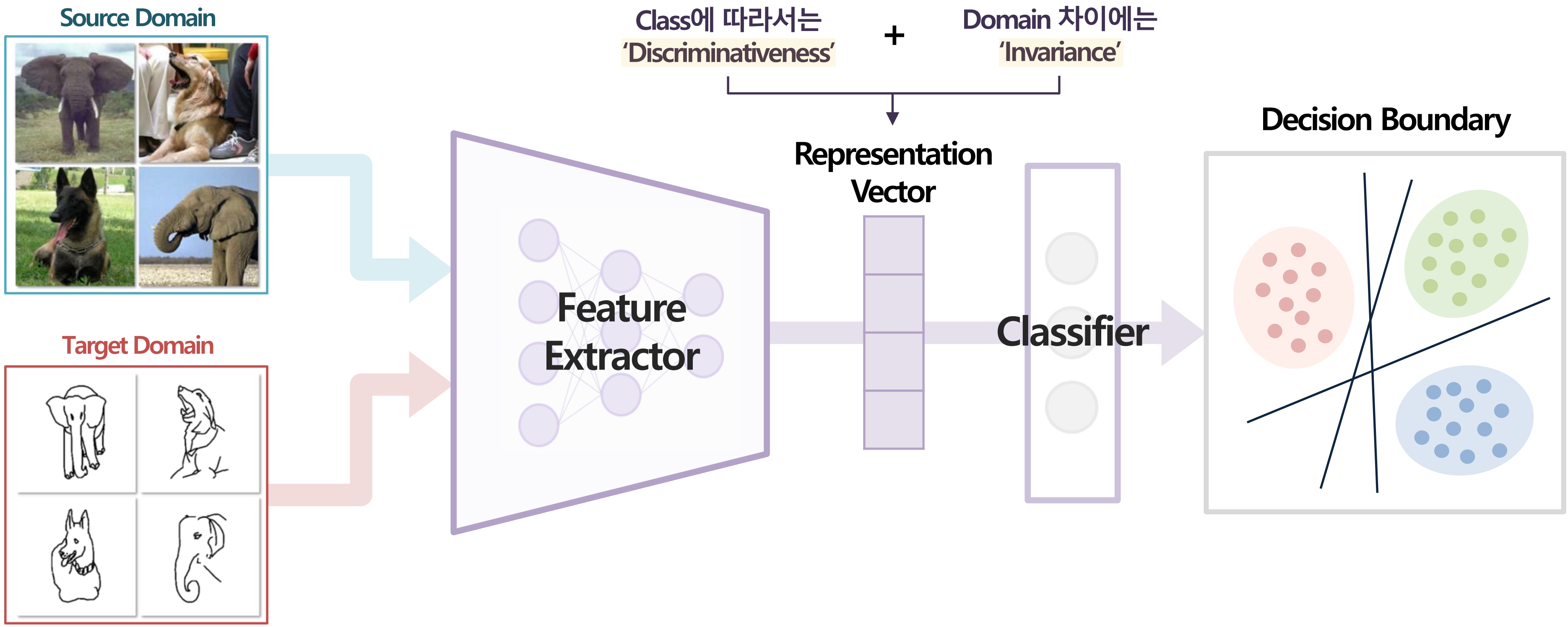
[3] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., ... & Lempitsky, V. (2016). Domain-adversarial training of neural networks. The journal of machine learning research, 17(1), 2096-2030.
[4] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.

Methods: Preliminaries

Domain Adversarial Learning

❖ Domain Adversarial Training of Neural Networks (DANN)[3]

- **Motivation** : Domain Adaptation과 Representation Learning을 하나의 학습 과정 안에서 해결하자
  - 이로써 최종적인 결정 경계(decision boundary)가 'discriminative' 하면서도 'domain-invariant'한 representation으로부터 형성 가능



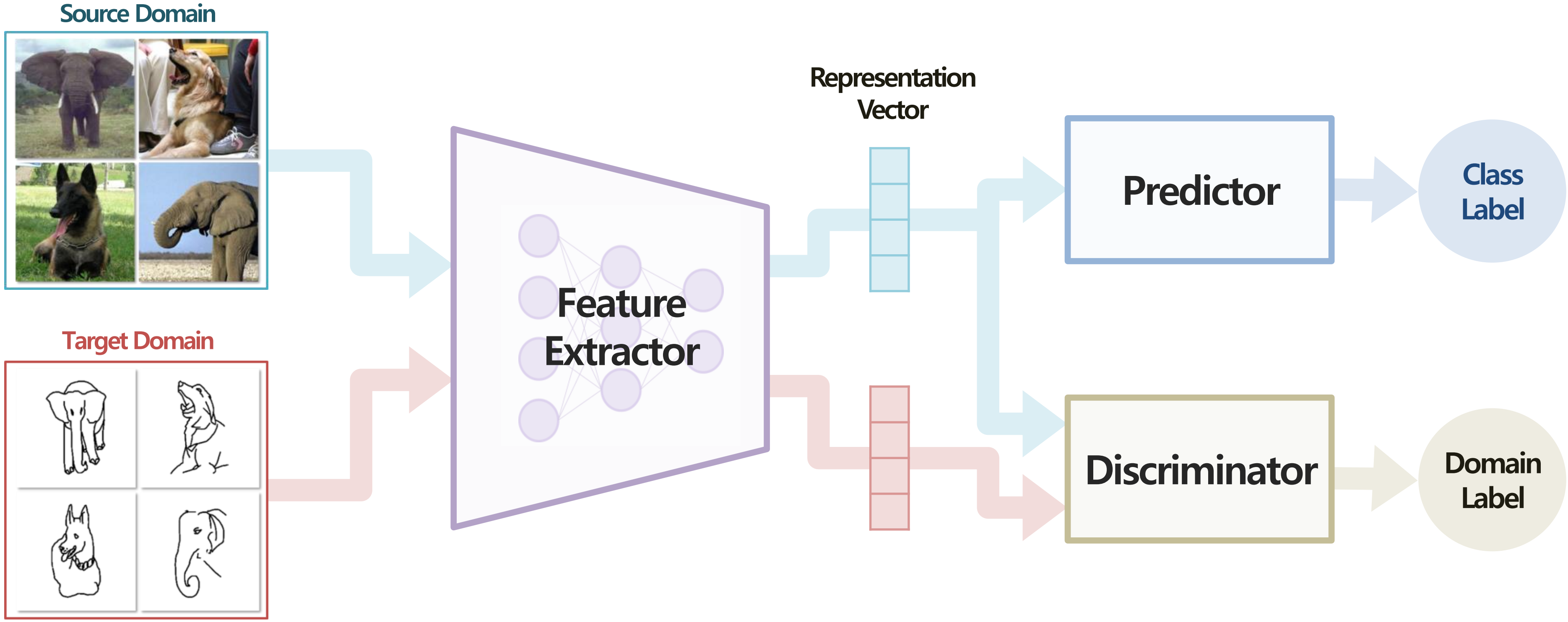
[3] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., ... & Lempitsky, V. (2016). Domain-adversarial training of neural networks. The journal of machine learning research, 17(1), 2096-2030.

Methods: Preliminaries

Domain Adversarial Learning

❖ **Domain Adversarial Training of Neural Networks (DANN)**[3]

- "**Discriminator**"라는 Classifier를 덧붙인 형태
  - Input representation이 어느 Domain에서 비롯된 것인지를 분류하는 Classifier



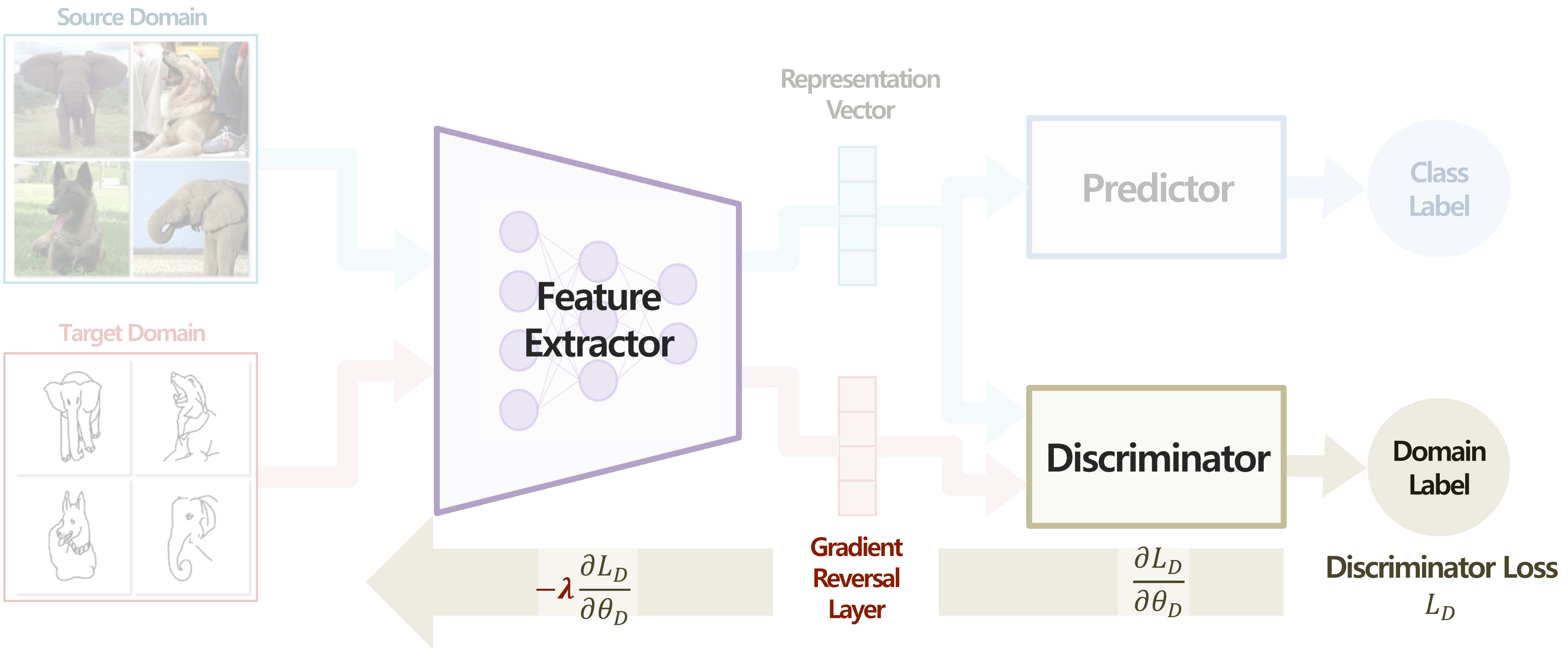
[3] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., ... & Lempitsky, V. (2016). Domain-adversarial training of neural networks. The journal of machine learning research, 17(1), 2096-2030.

Methods: Preliminaries

Domain Adversarial Learning

❖ Domain Adversarial Training of Neural Networks (DANN)[3]

- "**Discriminator**"라는 Classifier를 덧붙인 형태
  - Input representation이 어느 Domain에서 비롯된 것인지를 분류하는 Classifier → **역전파 시 gradient에 negative scalar($-\lambda$)를 곱함**



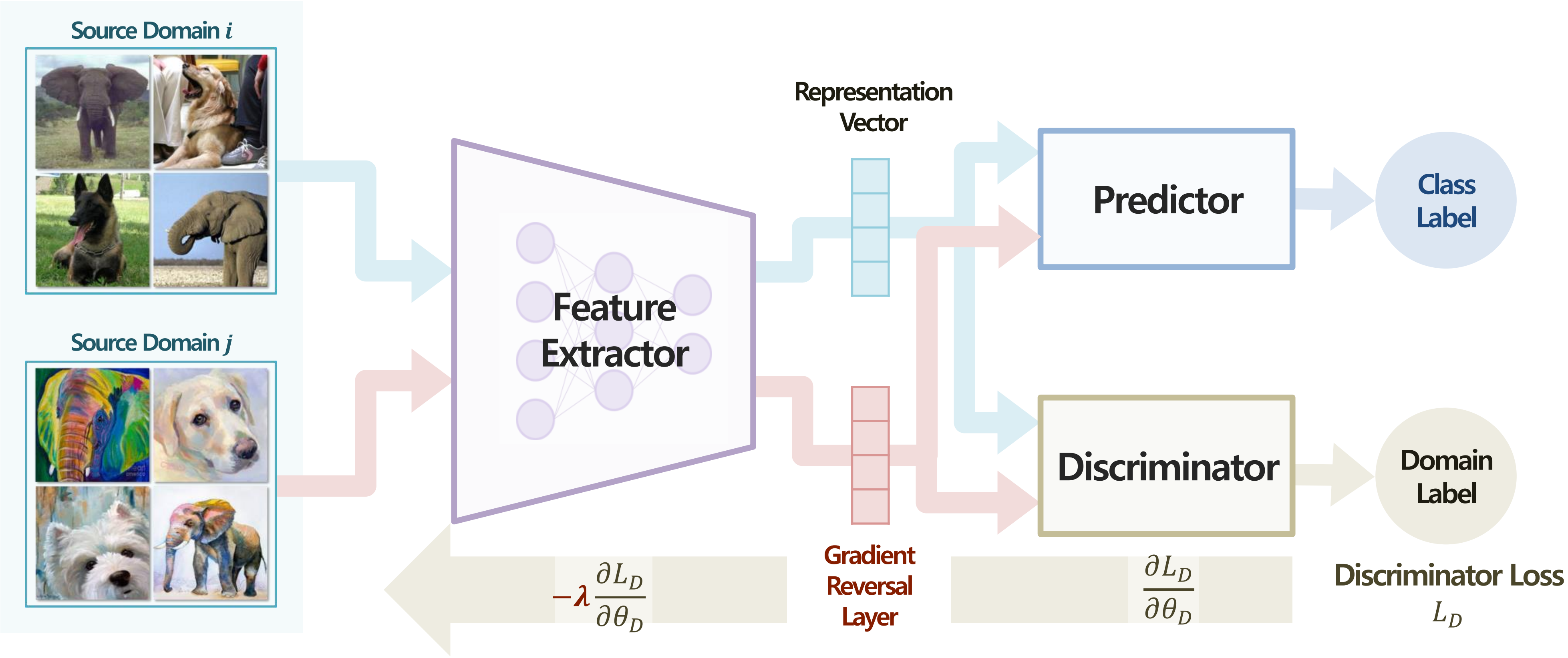
[3] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., ... & Lempitsky, V. (2016). Domain-adversarial training of neural networks. The journal of machine learning research, 17(1), 2096-2030.

Methods

1. Domain-invariant Representation Learning with Adversarial Learning

❖ DANN[3] with Domain Generalization

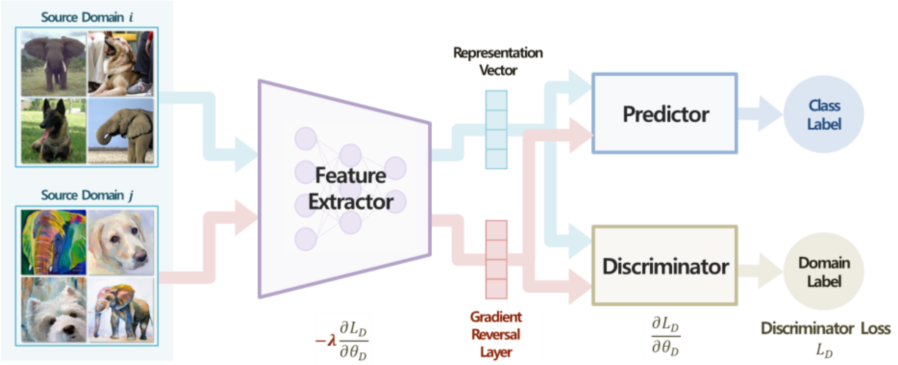
- Source Domain-invariant Representation 추출
  - **Discriminator** : Source Domain 두 쌍에 대한 Binary Classification or K개의 Source Domain에 대한 Multi-class Classification 수행



[3] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., ... & Lempitsky, V. (2016). Domain-adversarial training of neural networks. The journal of machine learning research, 17(1), 2096-2030.

Methods

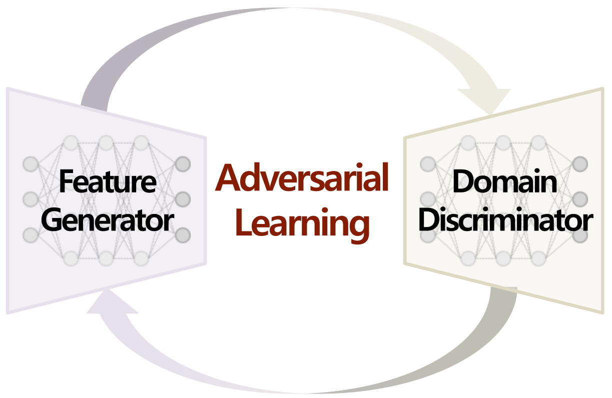
1. Domain-invariant Representation Learning with Adversarial Learning

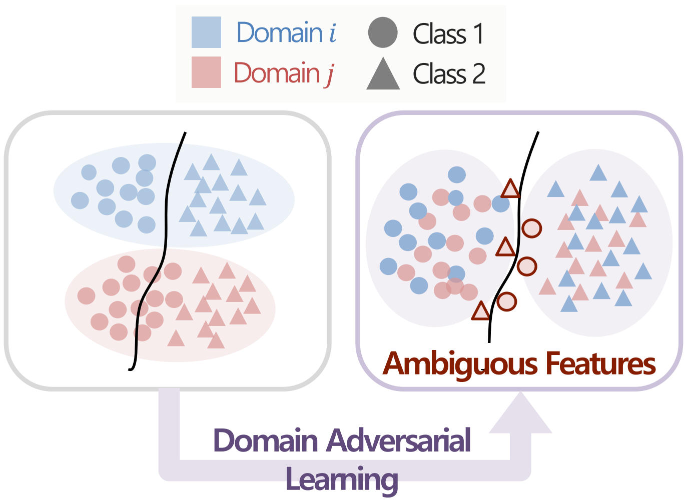


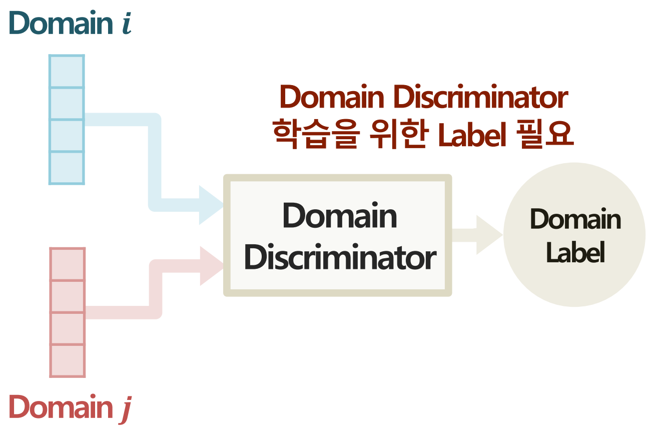
❖ Limitations of Domain Adversarial Learning

1. 적대적 학습(Adversarial Learning)의 불안정성[5]
2. 결정 경계(Decision Boundary) 근처의 Ambiguous Features를 생성하는 문제[6][7]
3. Domain Label을 필요로 하는 문제[7]

적대적 학습의 불안정성	결정 경계 근처의 features 생성	Domain Label 필요



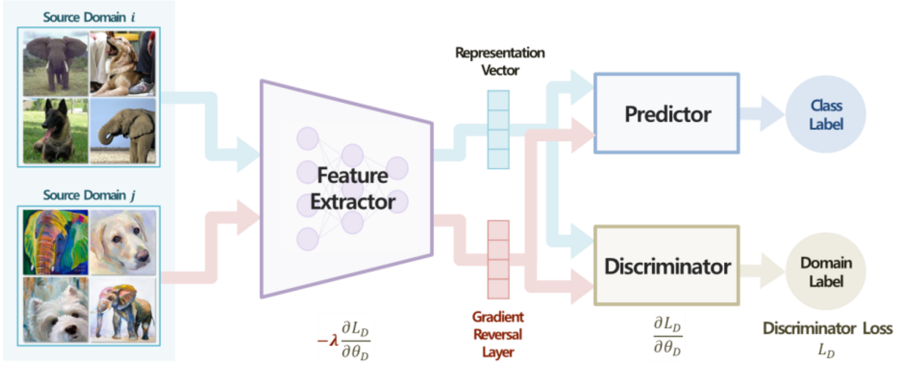




[5] Rahman, M. M., Fookes, C., Baktashmotlagh, M., & Sridharan, S. (2020). Correlation-aware adversarial domain adaptation and generalization. Pattern Recognition, 100, 107124.
[6] Saito, K., Kim, D., Sclaroff, S., Darrell, T., & Saenko, K. (2019). Semi-supervised domain adaptation via minimax entropy. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 8050-8058).
[7] Matsuura, T., & Harada, T. (2020, April). Domain generalization using a mixture of multiple latent domains. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 07, pp. 11749-11756).

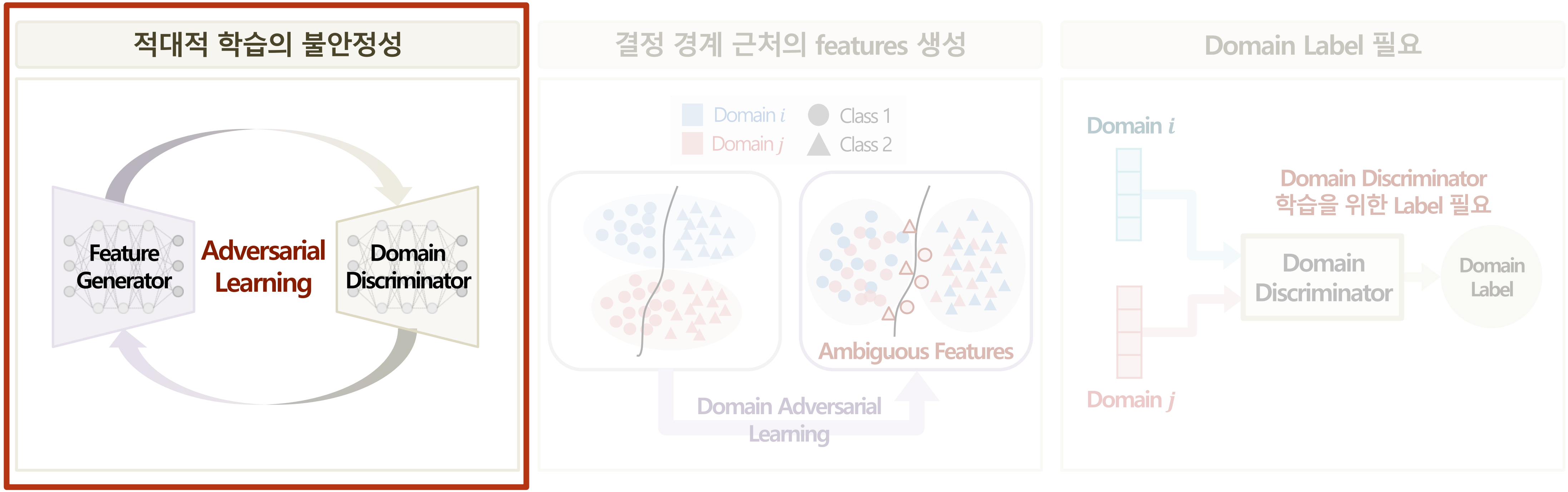
Methods

1. Domain-invariant Representation Learning with Adversarial Learning



❖ Limitations of Domain Adversarial Learning

1. 적대적 학습(Adversarial Learning)의 불안정성[5]
2. 결정 경계(Decision Boundary) 근처의 Ambiguous Features를 생성하는 문제[6][7]
3. Domain Label을 필요로 하는 문제[7]



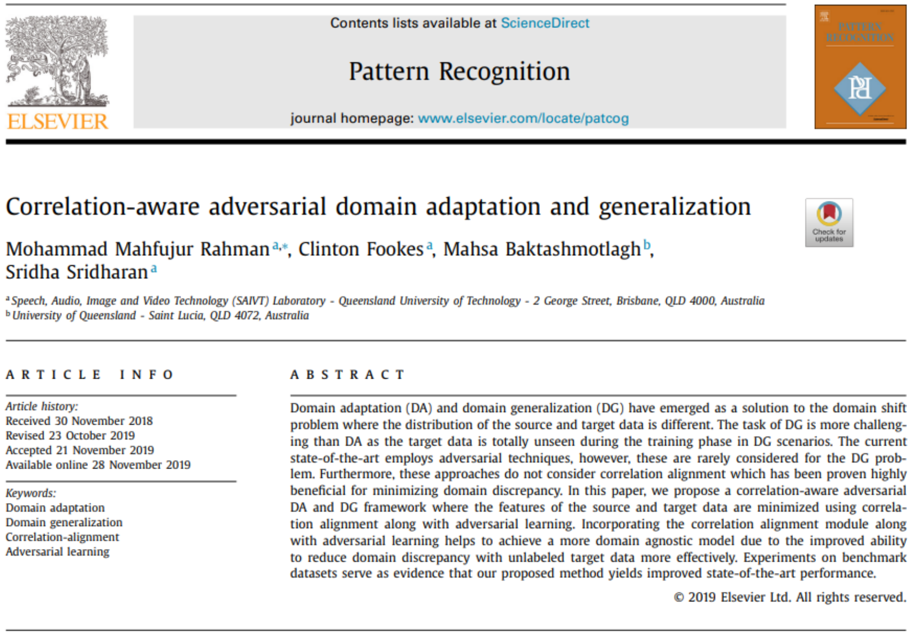
[5] Rahman, M. M., Fookes, C., Baktashmotlagh, M., & Sridharan, S. (2020). Correlation-aware adversarial domain adaptation and generalization. Pattern Recognition, 100, 107124.
[6] Saito, K., Kim, D., Sclaroff, S., Darrell, T., & Saenko, K. (2019). Semi-supervised domain adaptation via minimax entropy. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 8050-8058).
[7] Matsuura, T., & Harada, T. (2020, April). Domain generalization using a mixture of multiple latent domains. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 07, pp. 11749-11756).

Methods

1. Domain-invariant Representation Learning with Adversarial Learning

❖ Correlation-aware Adversarial Domain Adaptation and Generalization[5]

- 2020년 제안된 Domain Adaptation & Domain Generalization 방법론 (Pattern Recognition, 23년 7월 기준 89회 인용)
- Domain Adversarial Learning 구조에 CORAL loss[8]를 더하여 Domain Alignment 효과 극대화



Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Correlation-aware adversarial domain adaptation and generalization

Mohammad Mahfujur Rahman[a,*], Clinton Fookes[a], Mahsa Baktashmotlagh[b], Sridha Sridharan[a]

[a] Speech, Audio, Image and Video Technology (SAIVT) Laboratory - Queensland University of Technology - 2 George Street, Brisbane, QLD 4000, Australia
[b] University of Queensland - Saint Lucia, QLD 4072, Australia

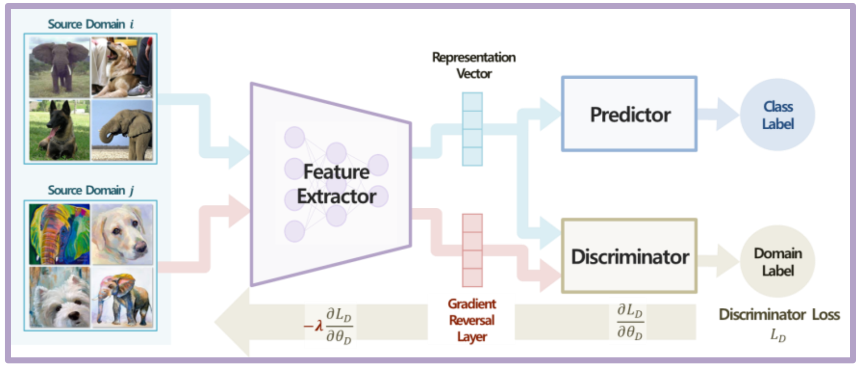
ARTICLE INFO

Article history:
Received 30 November 2018
Revised 23 October 2019
Accepted 21 November 2019
Available online 28 November 2019

Keywords:
Domain adaptation
Domain generalization
Correlation-alignment
Adversarial learning

ABSTRACT

Domain adaptation (DA) and domain generalization (DG) have emerged as a solution to the domain shift problem where the distribution of the source and target data is different. The task of DG is more challenging than DA as the target data is totally unseen during the training phase in DG scenarios. The current state-of-the-art employs adversarial techniques, however, these are rarely considered for the DG problem. Furthermore, these approaches do not consider correlation alignment which has been proven highly beneficial for minimizing domain discrepancy. In this paper, we propose a correlation-aware adversarial DA and DG framework where the features of the source and target data are minimized using correlation alignment along with adversarial learning. Incorporating the correlation alignment module along with adversarial learning helps to achieve a more domain agnostic model due to the improved ability to reduce domain discrepancy with unlabeled target data more effectively. Experiments on benchmark datasets serve as evidence that our proposed method yields improved state-of-the-art performance.



+

$$\ell_{CORAL} = \frac{1}{4d^2} ||C_S - C_T||_F^2,$$

where $||\cdot||_F^2$ *denotes the sqaured matrix Frobenius norm,*

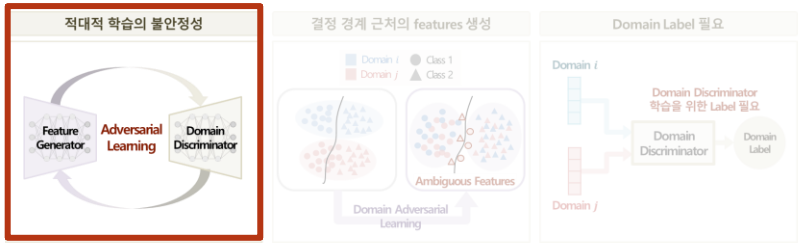
$$C_S = \frac{1}{n_S - 1}\left(\mathcal{D}_S^\top \mathcal{D}_S - \frac{1}{n_S}(1^\top \mathcal{D}_S)^\top (1^\top \mathcal{D}_S)\right),$$

$$C_T = \frac{1}{n_T - 1}\left(\mathcal{D}_T^\top \mathcal{D}_T - \frac{1}{n_T}(1^\top \mathcal{D}_T)^\top (1^\top T)\right).$$

[5] Rahman, M. M., Fookes, C., Baktashmotlagh, M., & Sridharan, S. (2020). Correlation-aware adversarial domain adaptation and generalization. Pattern Recognition, 100, 107124.
[8] Sun, B., & Saenko, K. (2016). Deep coral: Correlation alignment for deep domain adaptation. In Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14 (pp. 443-450). Springer International Publishing.

Methods

1. Domain-invariant Representation Learning with Adversarial Learning

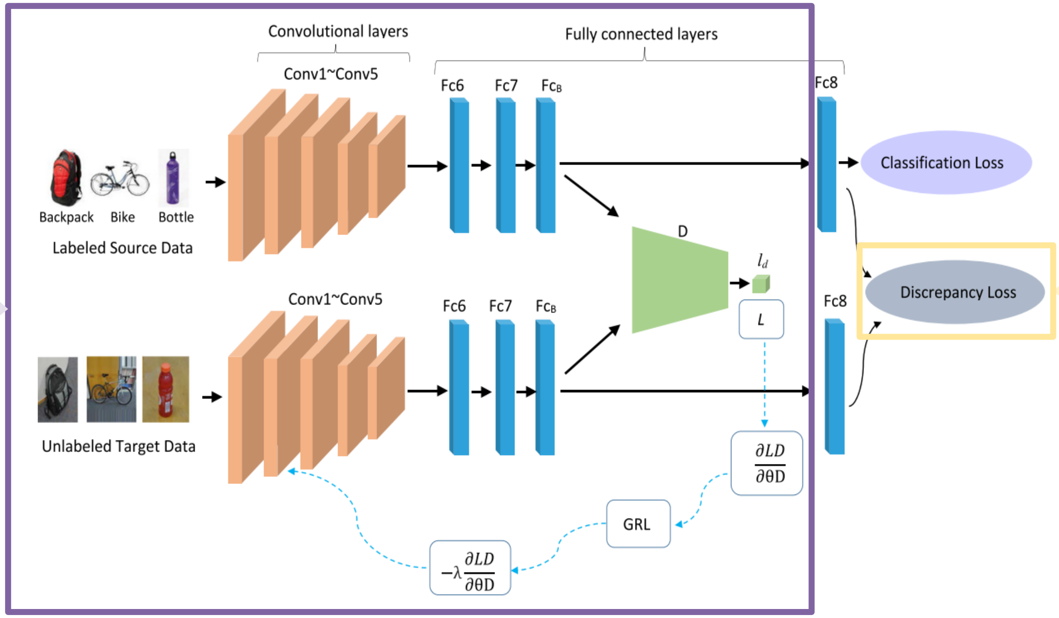


❖ Correlation-aware Adversarial Domain Adaptation and Generalization[5]

- **Motivation** : Domain 간의 차이를 좁히기 위해서, Domain의 공분산까지 정렬(**CO**Relation **AL**ignment)[8]해보자
  - Domain Adversarial Learning을 골자로 하여, Domain Alignment 효과 극대화 가능

$$\ell_{CORAL} = \frac{1}{4d^2}||C_S - C_T||_F^2,$$

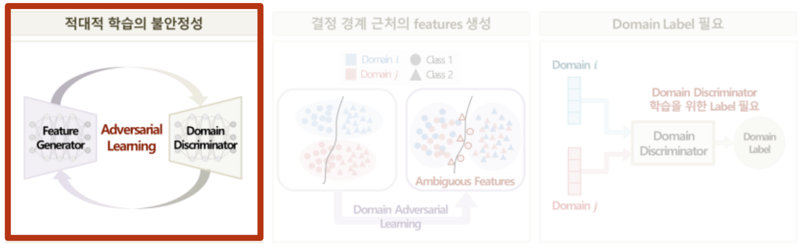
CAADA[5]



[5] Rahman, M. M., Fookes, C., Baktashmotlagh, M., & Sridharan, S. (2020). Correlation-aware adversarial domain adaptation and generalization. Pattern Recognition, 100, 107124.
[8] Sun, B., & Saenko, K. (2016). Deep coral: Correlation alignment for deep domain adaptation. In Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14 (pp. 443-450). Springer International Publishing.
[9] Sun, B., Feng, J., & Saenko, K. (2016, March). Return of frustratingly easy domain adaptation. In Proceedings of the AAAI conference on artificial intelligence (Vol. 30, No. 1).

Methods

1. Domain-invariant Representation Learning with Adversarial Learning

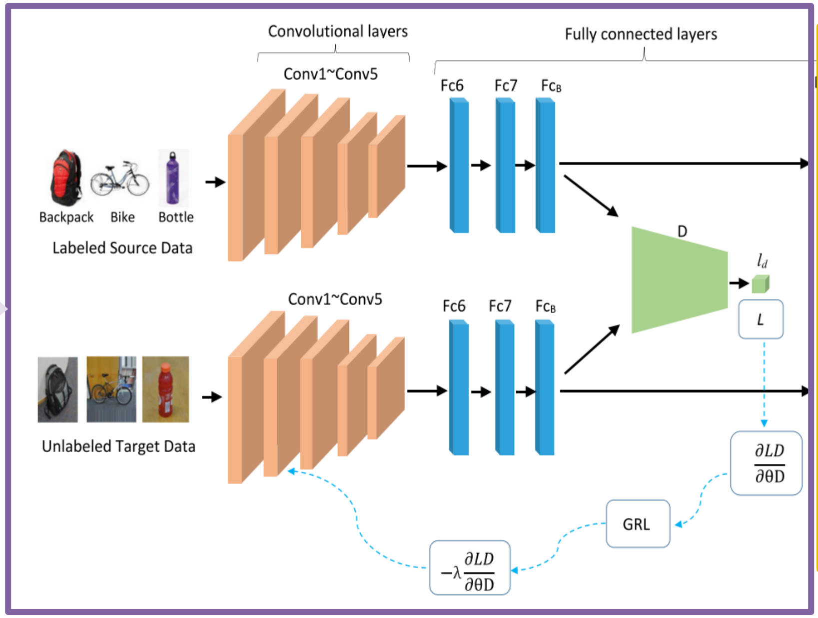


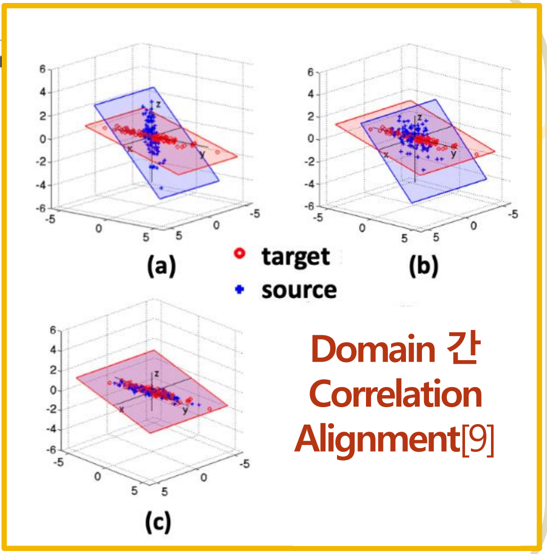
❖ Correlation-aware Adversarial Domain Adaptation and Generalization[5]

- **Motivation** : Domain 간의 차이를 좁히기 위해서, Domain의 공분산까지 정렬(**CO**Relation **AL**ignment)[8]해보자
  - Domain Adversarial Learning을 골자로 하여, Domain Alignment 효과 극대화 가능

$$\ell_{CORAL} = \frac{1}{4d^2} \|C_S - C_T\|_F^2,$$

CAADA[5]

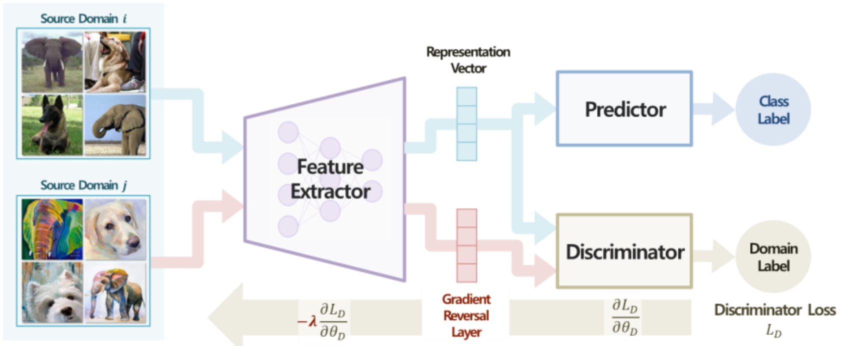




[5] Rahman, M. M., Fookes, C., Baktashmotlagh, M., & Sridharan, S. (2020). Correlation-aware adversarial domain adaptation and generalization. Pattern Recognition, 100, 107124.
[8] Sun, B., & Saenko, K. (2016). Deep coral: Correlation alignment for deep domain adaptation. In Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14 (pp. 443-450). Springer International Publishing.
[9] Sun, B., Feng, J., & Saenko, K. (2016, March). Return of frustratingly easy domain adaptation. In Proceedings of the AAAI conference on artificial intelligence (Vol. 30, No. 1).

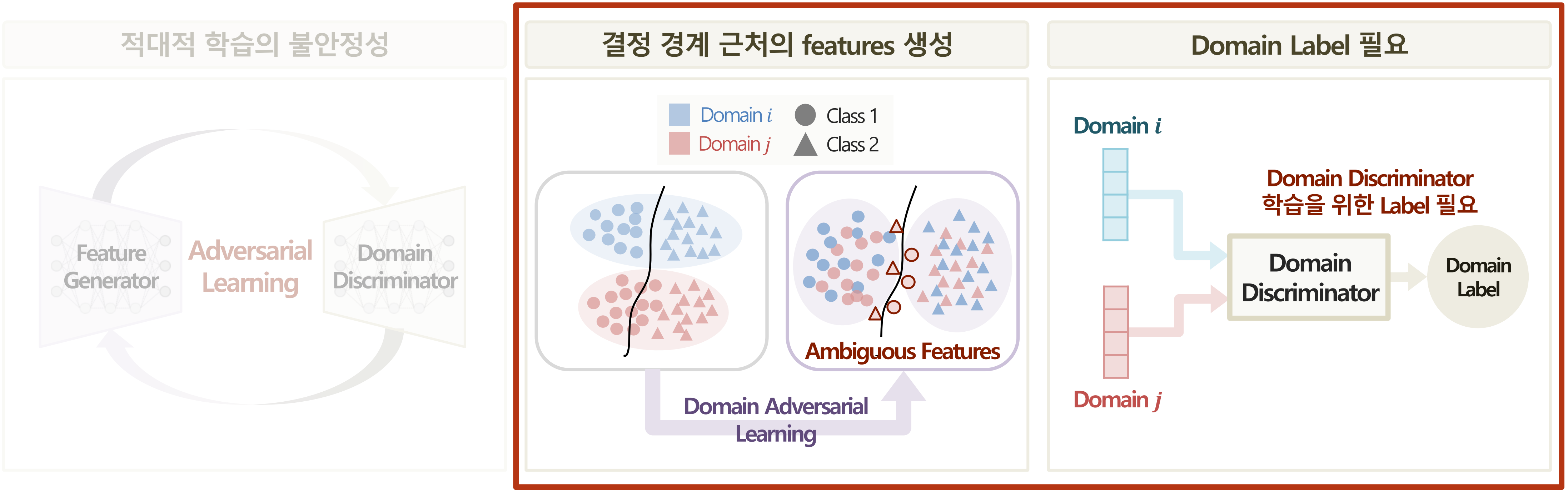
Methods

1. Domain-invariant Representation Learning with Adversarial Learning



❖ Limitations of Domain Adversarial Learning

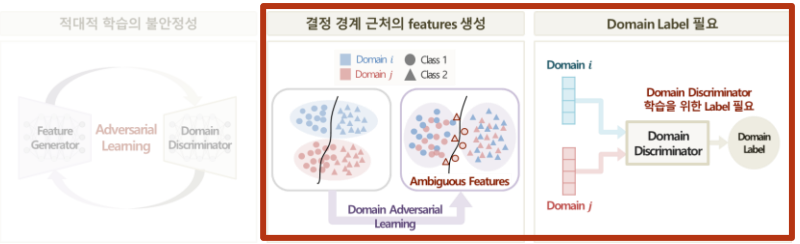
1. 적대적 학습(Adversarial Learning)의 불안정성[5]
2. 결정 경계(Decision Boundary) 근처의 ambiguous features를 생성하는 문제[6][7]
3. Domain label을 필요로 하는 문제[7]



[5] Rahman, M. M., Fookes, C., Baktashmotlagh, M., & Sridharan, S. (2020). Correlation-aware adversarial domain adaptation and generalization. Pattern Recognition, 100, 107124.
[6] Saito, K., Kim, D., Sclaroff, S., Darrell, T., & Saenko, K. (2019). Semi-supervised domain adaptation via minimax entropy. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 8050-8058).
[7] Matsuura, T., & Harada, T. (2020, April). Domain generalization using a mixture of multiple latent domains. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 07, pp. 11749-11756).

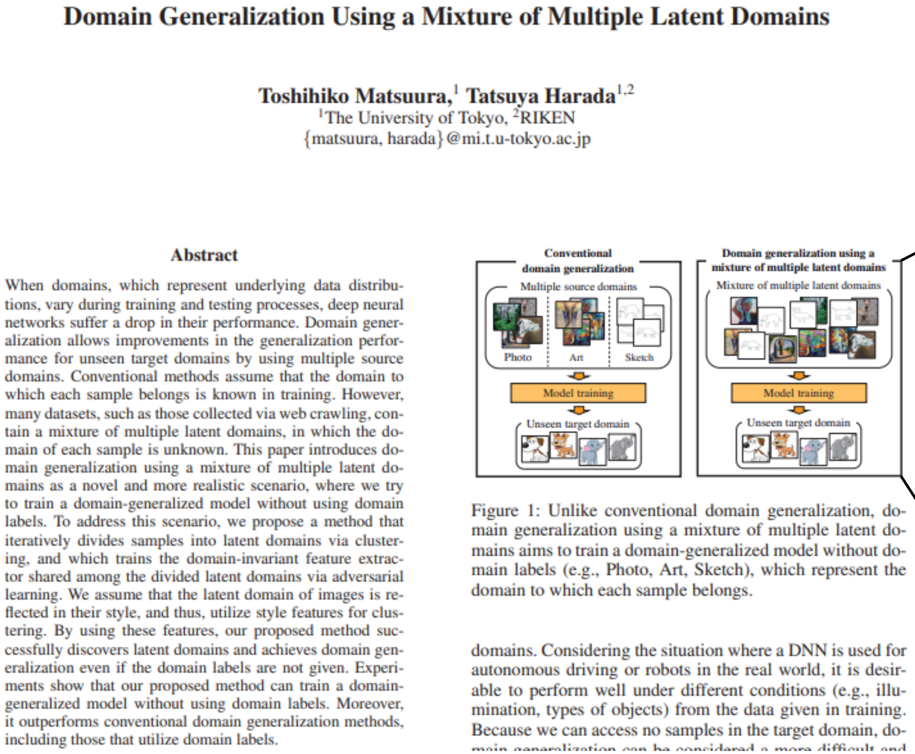
Methods

1. Domain-invariant Representation Learning with Adversarial Learning



❖ Domain Generalization using a Mixture of Multiple Latent Domains[7]

- 2020년 제안된 Domain Generalization 방법론 (AAAI, 23년 7월 기준 224회 인용)
- Domain Label이 없는 경우에도 가능한 Domain Adversarial Learning 방법론 제안



Domain Generalization Using a Mixture of Multiple Latent Domains

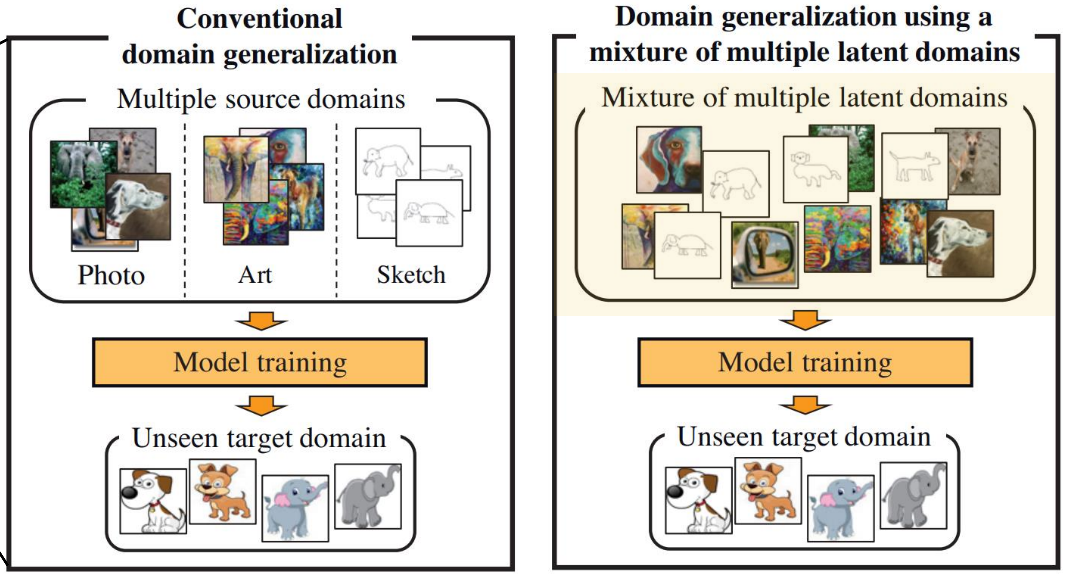
Toshihiko Matsuura,[1] Tatsuya Harada[1,2]
[1]The University of Tokyo, [2]RIKEN
{matsuura, harada}@mi.t.u-tokyo.ac.jp

Abstract

When domains, which represent underlying data distributions, vary during training and testing processes, deep neural networks suffer a drop in their performance. Domain generalization allows improvements in the generalization performance for unseen target domains by using multiple source domains. Conventional methods assume that the domain to which each sample belongs is known in training. However, many datasets, such as those collected via web crawling, contain a mixture of multiple latent domains, in which the domain of each sample is unknown. This paper introduces domain generalization using a mixture of multiple latent domains as a novel and more realistic scenario, where we try to train a domain-generalized model without using domain labels. To address this scenario, we propose a method that iteratively divides samples into latent domains via clustering, and which trains the domain-invariant feature extractor shared among the divided latent domains via adversarial learning. We assume that the latent domain of images is reflected in their style, and thus, utilize style features for clustering. By using these features, our proposed method successfully discovers latent domains and achieves domain generalization even if the domain labels are not given. Experiments show that our proposed method can train a domain-generalized model without using domain labels. Moreover, it outperforms conventional domain generalization methods, including those that utilize domain labels.

Figure 1: Unlike conventional domain generalization, domain generalization using a mixture of multiple latent domains aims to train a domain-generalized model without domain labels (e.g., Photo, Art, Sketch), which represent the domain to which each sample belongs.

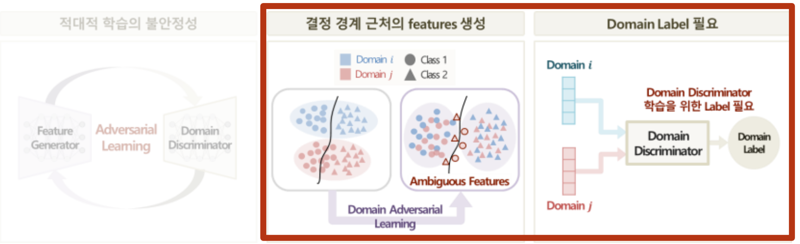
domains. Considering the situation where a DNN is used for autonomous driving or robots in the real world, it is desirable to perform well under different conditions (e.g., illumination, types of objects) from the data given in training. Because we can access no samples in the target domain, domain generalization can be considered a more difficult and



[7] Matsuura, T., & Harada, T. (2020, April). Domain generalization using a mixture of multiple latent domains. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 07, pp. 11749-11756).

Methods

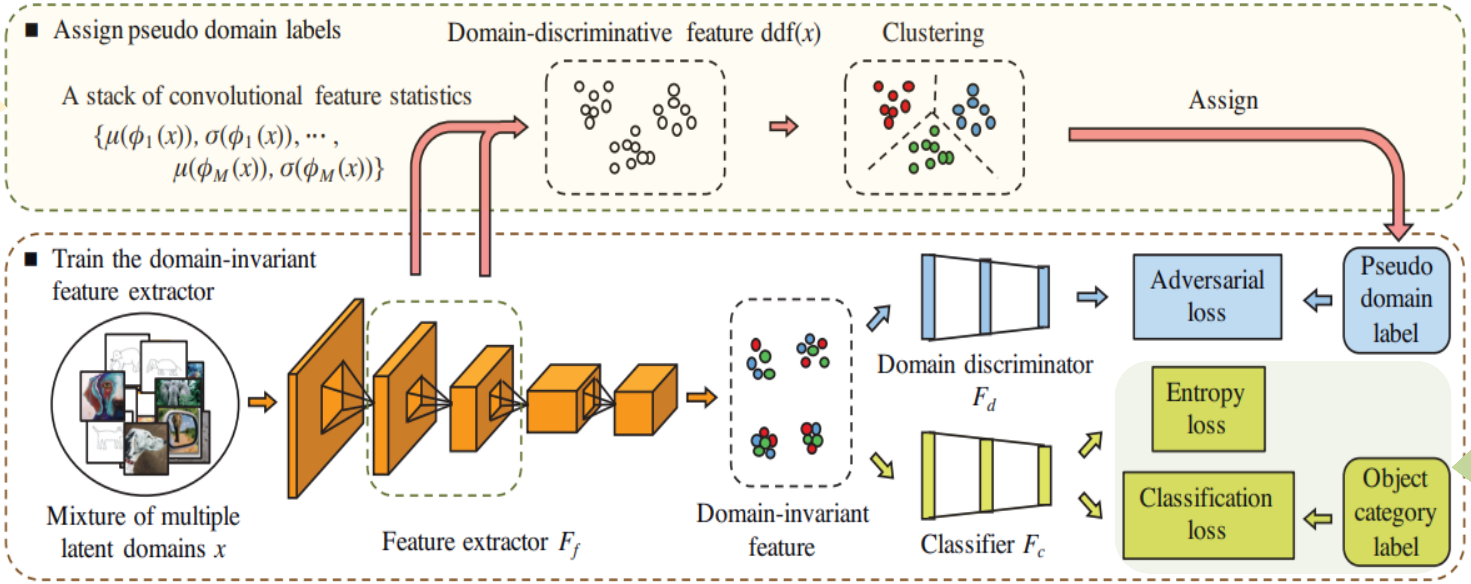
1. Domain-invariant Representation Learning with Adversarial Learning



❖ Domain Generalization using a Mixture of Multiple Latent Domains[7]

- **Motivation** : 사전에 Domain Label이 없는 경우에도 가능한 Domain Adversarial Learning 방법론이 필요
  - (1) Pseudo Domain Labels 생성을 통한 Adversarial Learning 수행, (2) Entropy Minimization을 통한 Discriminative Features 추출

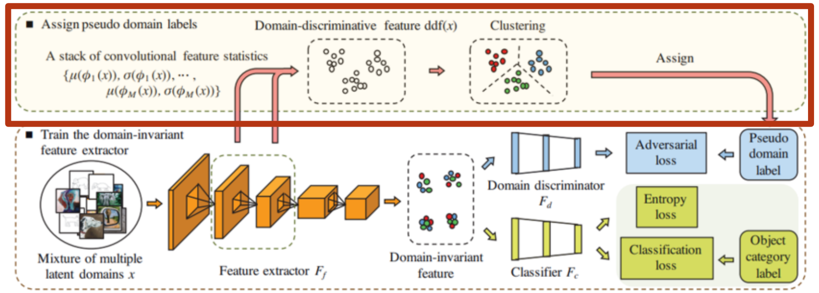
MMLD[7]



[7] Matsuura, T., & Harada, T. (2020, April). Domain generalization using a mixture of multiple latent domains. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 07, pp. 11749-11756).

Methods

1. Domain-invariant Representation Learning with Adversarial Learning

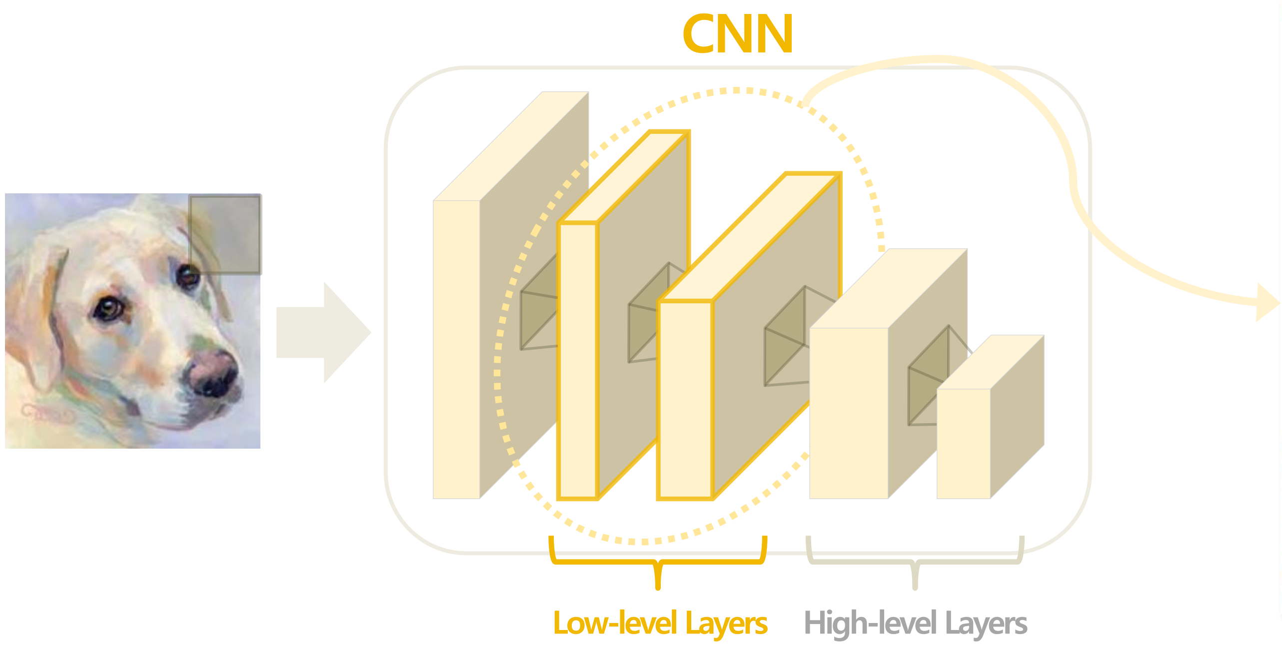


❖ Domain Generalization using a Mixture of Multiple Latent Domains[7]

- **Motivation** : 사전에 Domain Label이 없는 경우에도 가능한 Domain Adversarial Learning 방법론이 필요
  - (1) **Pseudo Domain Labels** 생성을 통한 Adversarial Learning 수행, (2) Entropy Minimization을 통한 Discriminative Features 추출

→ Domain을 구분할 수 있는 Domain-discriminative Features 추출 필요!

→ **가설:** Latent domain of images is reflected in their "style"



4.1 Different Style Representations [10]

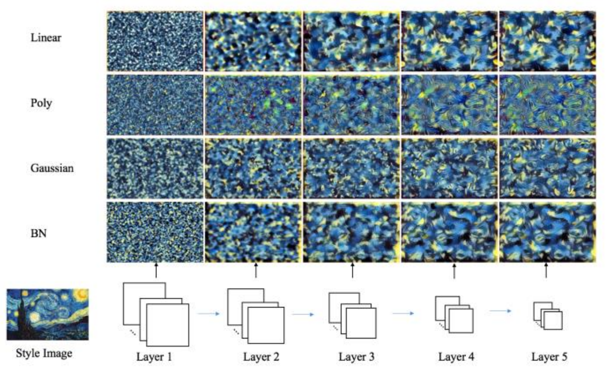


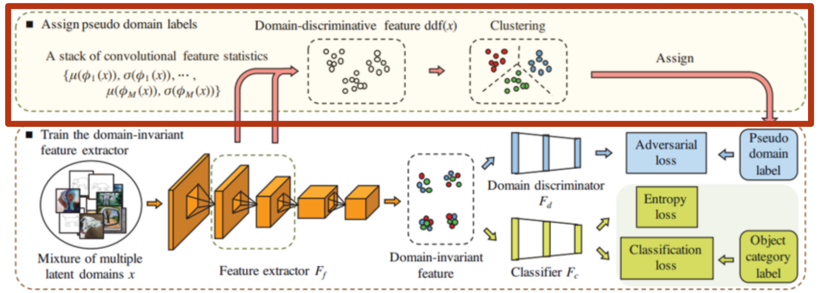
Figure 1: Style reconstructions of different methods in five layers, respectively. Each row corresponds to one method and the reconstruction results are obtained by only using the style loss $\mathcal{L}_{style}$ with $\alpha = 0$. We also reconstruct different style representations in different subsets of layers of VGG network. For example, layer 3 contains the style loss of the first 3 layers ($w_1 = w_2 = w_3 = 1.0$ and $w_4 = w_5 = 0.0$).

Layer마다 추출되는 feature map 특징이 다름

[7] Matsuura, T., & Harada, T. (2020, April). Domain generalization using a mixture of multiple latent domains. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 07, pp. 11749-11756).
[10] Li, Y., Wang, N., Liu, J., & Hou, X. (2017). Demystifying neural style transfer. arXiv preprint arXiv:1701.01036.

Methods

1. Domain-invariant Representation Learning with Adversarial Learning

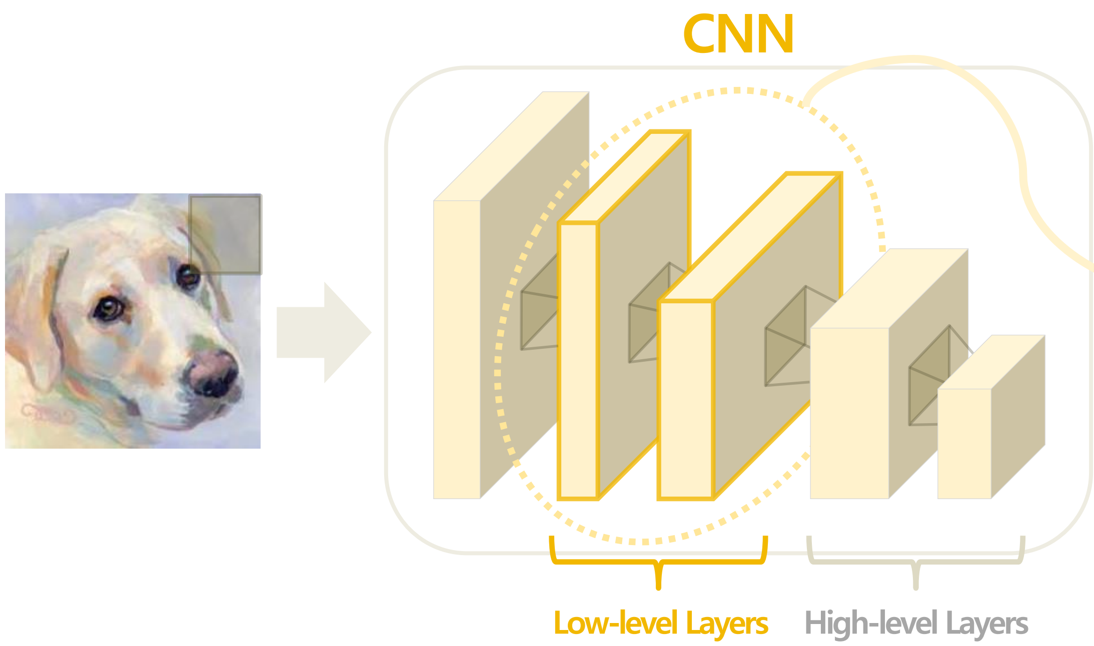


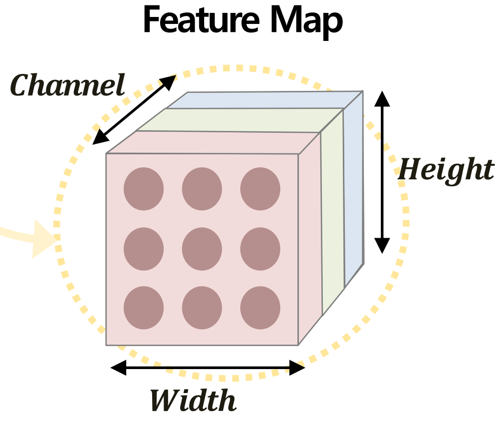
❖ Domain Generalization using a Mixture of Multiple Latent Domains[7]

- **Motivation** : 사전에 Domain Label이 없는 경우에도 가능한 Domain Adversarial Learning 방법론이 필요
  - (1) **Pseudo Domain Labels** 생성을 통한 Adversarial Learning 수행, (2) Entropy Minimization을 통한 Discriminative Features 추출

→ Domain을 구분할 수 있는 Domain-discriminative Features 추출 필요!

→ **가설:** Latent domain of images is reflected in their "style"





Feature Statistics 추출[10]

$$\mu_c(x) = \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W} x_{c,h,w}$$

$$\sigma_c(x) = \sqrt{\frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}(x_{c,h,w} - \mu_c(x))^2 + \epsilon}$$

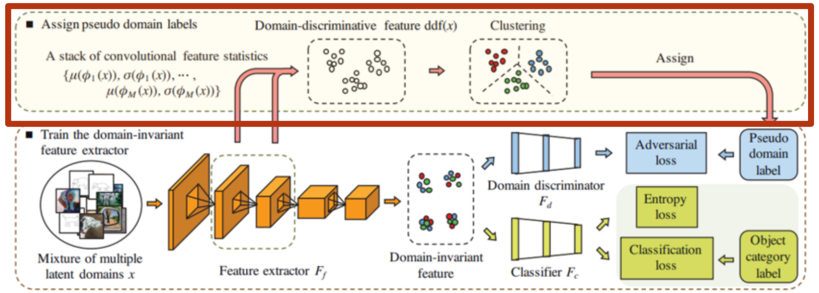
$$\boldsymbol{ddf(x)} = \{\mu(\phi_1(x)), \sigma(\phi_1(x)), \ldots, \mu(\phi_M(x)), \sigma(\phi_M(x))\}$$

Domain-discriminative Features : M개의 Layers로부터 추출된 Feature Statistics

[7] Matsuura, T., & Harada, T. (2020, April). Domain generalization using a mixture of multiple latent domains. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 07, pp. 11749-11756).
[10] Li, Y., Wang, N., Liu, J., & Hou, X. (2017). Demystifying neural style transfer. arXiv preprint arXiv:1701.01036.

Methods

1. Domain-invariant Representation Learning with Adversarial Learning



❖ Domain Generalization using a Mixture of Multiple Latent Domains[7]

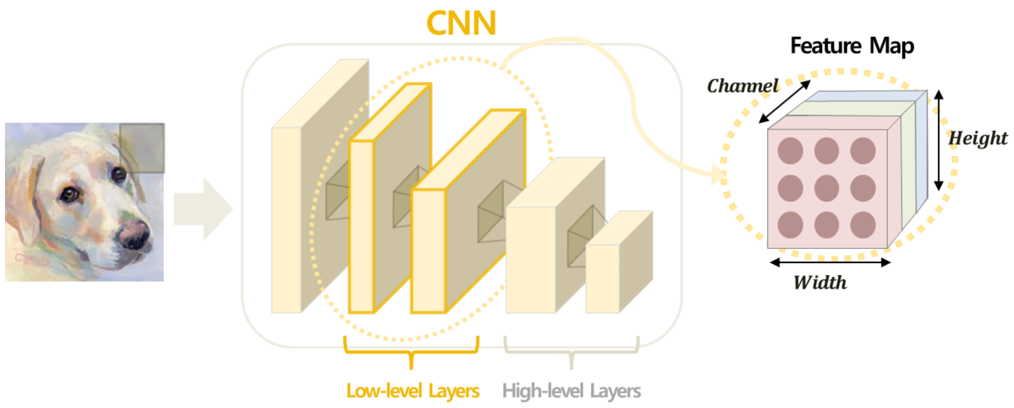
- **Motivation** : 사전에 Domain Label이 없는 경우에도 가능한 Domain Adversarial Learning 방법론이 필요
  - (1) **Pseudo Domain Labels** 생성을 통한 Adversarial Learning 수행, (2) Entropy Minimization을 통한 Discriminative Features 추출

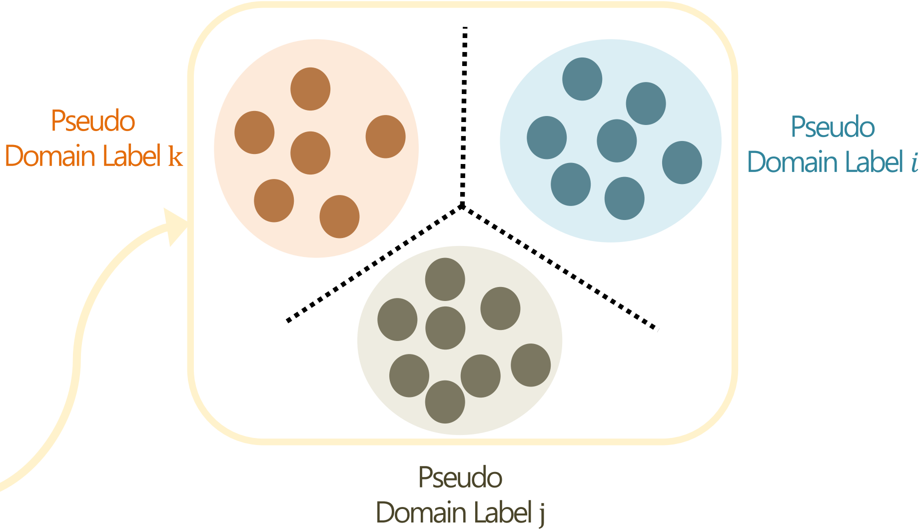
→ Domain을 구분할 수 있는 Domain-discriminative Features 추출 필요!

→ **가설:** Latent domain of images is reflected in their "style"

K-means Clustering

ddf를 Clustering 하여 pseudo domain label 부여





$$ddf(x) = \{\mu(\phi_1(x)), \sigma(\phi_1(x)), \ldots, \mu(\phi_M(x)), \sigma(\phi_M(x))\}$$

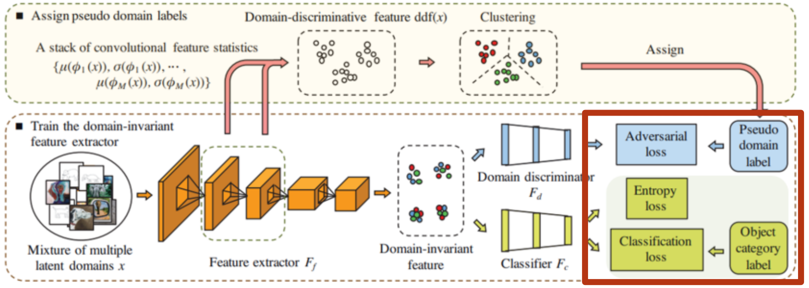
Domain-discriminative Features : M개의 Layers로부터 추출된 Feature Statistics

→ 이전 epoch의 domain label과 비교하며 반복적으로 reassign

[7] Matsuura, T., & Harada, T. (2020, April). Domain generalization using a mixture of multiple latent domains. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 07, pp. 11749-11756).

Methods

1. Domain-invariant Representation Learning with Adversarial Learning

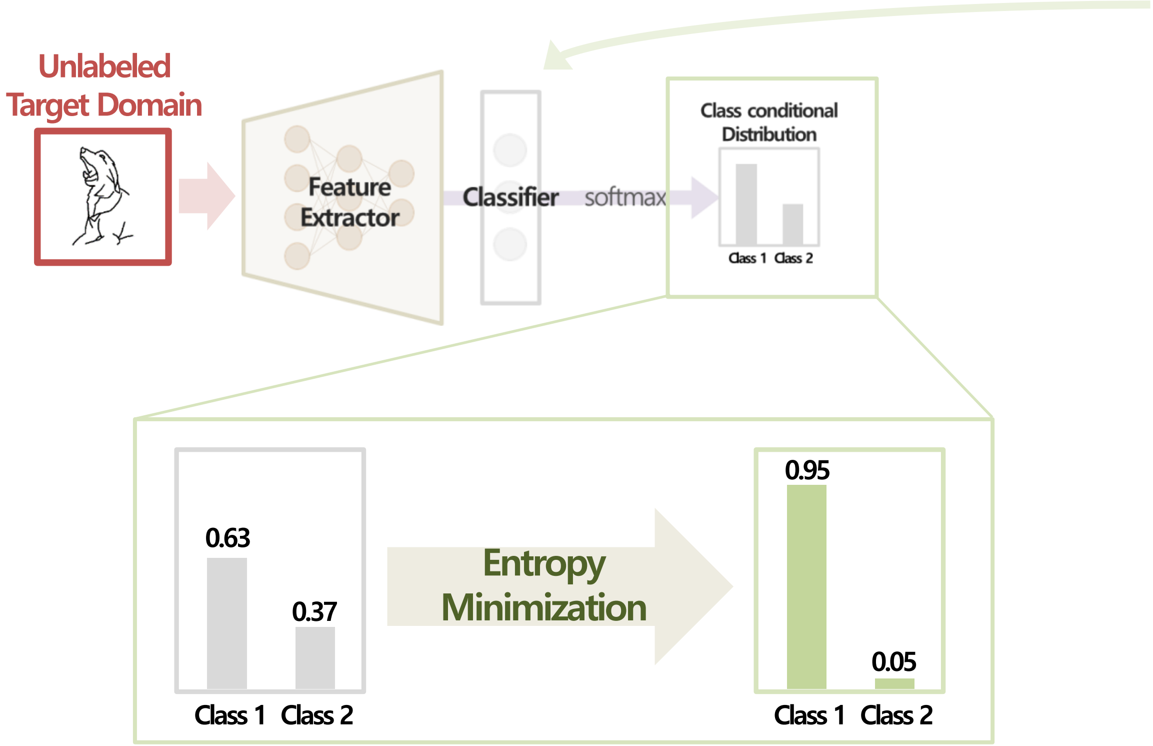


❖ Domain Generalization using a Mixture of Multiple Latent Domains[7]

- **Motivation** : 사전에 Domain Label이 없는 경우에도 가능한 Domain Adversarial Learning 방법론이 필요
  - (1) Pseudo Domain Labels 생성을 통한 Adversarial Learning 수행, (2) **Entropy Minimization**을 통한 Discriminative Features 추출

→ Class-discriminative Features 추출 필요!

→ **가설:** Minimizing class-conditional distribution entropy encourages the low-density separation between classes



Domain Adaptation[11]

$$\min_{h_t} \frac{1}{N_t} \sum_{i=1}^{N_t} H(h_t(x_i^t)),$$

$$where\ H(h_t(x_i^t)) = -\sum_{j=1}^{c} f_j^t(x_i^t) \log f_j^t(x_i^t)$$

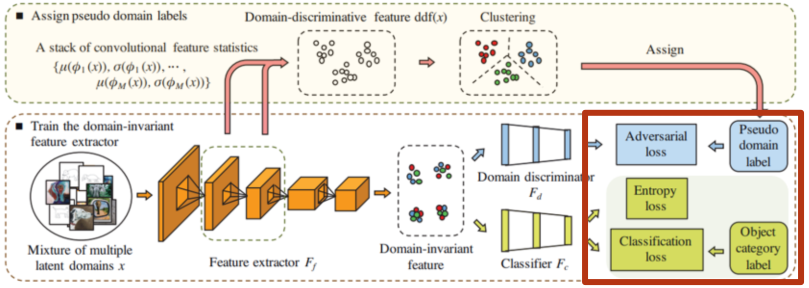
c = the number of classes

$f_j^t(x_i^t)$ = the probability of predicting point x_i^t to class j

[7] Matsuura, T., & Harada, T. (2020, April). Domain generalization using a mixture of multiple latent domains. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 07, pp. 11749-11756).
[11] Long, M., Zhu, H., Wang, J., & Jordan, M. I. (2016). Unsupervised domain adaptation with residual transfer networks. Advances in neural information processing systems, 29.

Methods

1. Domain-invariant Representation Learning with Adversarial Learning

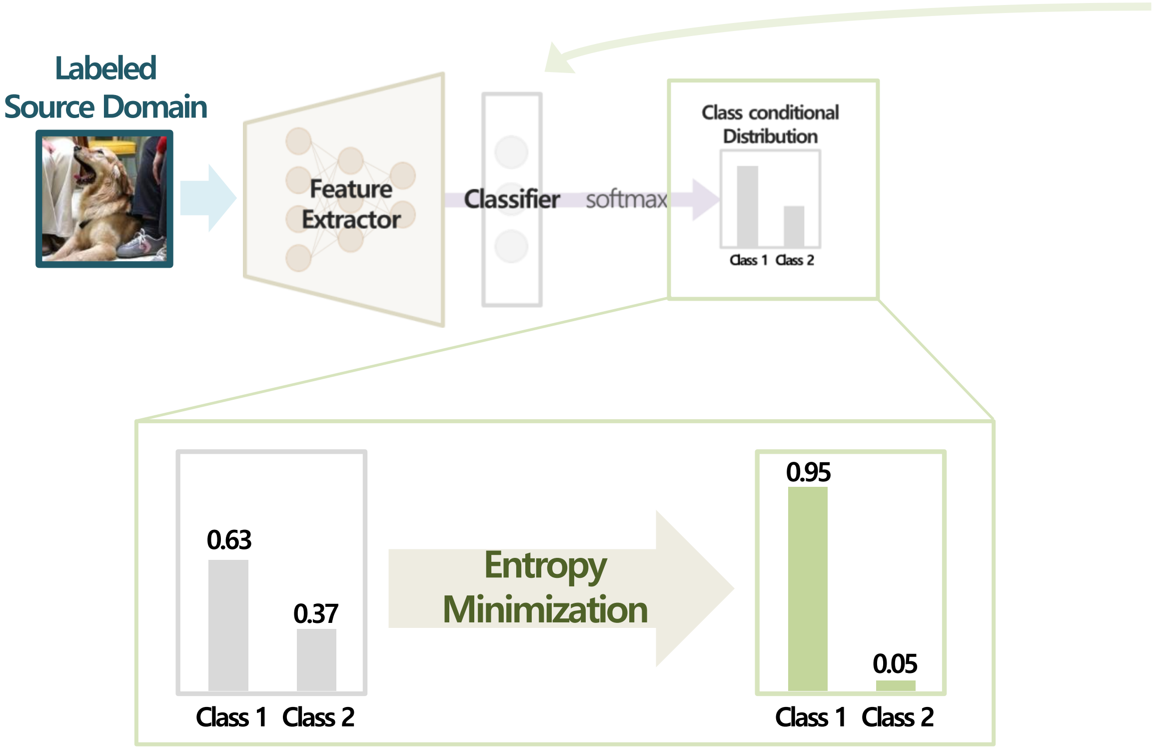


❖ Domain Generalization using a Mixture of Multiple Latent Domains[7]

- **Motivation** : 사전에 Domain Label이 없는 경우에도 가능한 Domain Adversarial Learning 방법론이 필요
  - (1) Pseudo Domain Labels 생성을 통한 Adversarial Learning 수행, (2) **Entropy Minimization**을 통한 Discriminative Features 추출

→ Class-discriminative Features 추출 필요!

→ **가설:** Minimizing class-conditional distribution entropy encourages the low-density separation between classes



Domain Generalization[7]

$$\min_{h_s} \frac{1}{N_s} \sum_{i=1}^{N_s} H(h_s(x_i^s)),$$

$$where\ H(h_s(x_i^s)) = -\sum_{j=1}^{c} f_j^s(x_i^s) \log f_j^s(x_i^s)$$

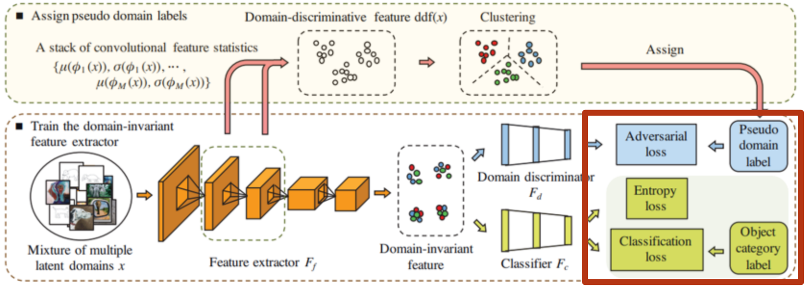
c = the number of classes

$f_j^s(x_i^s)$ = the probability of predicting point x_i^s to class j

[7] Matsuura, T., & Harada, T. (2020, April). Domain generalization using a mixture of multiple latent domains. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 07, pp. 11749-11756).

Methods

1. Domain-invariant Representation Learning with Adversarial Learning

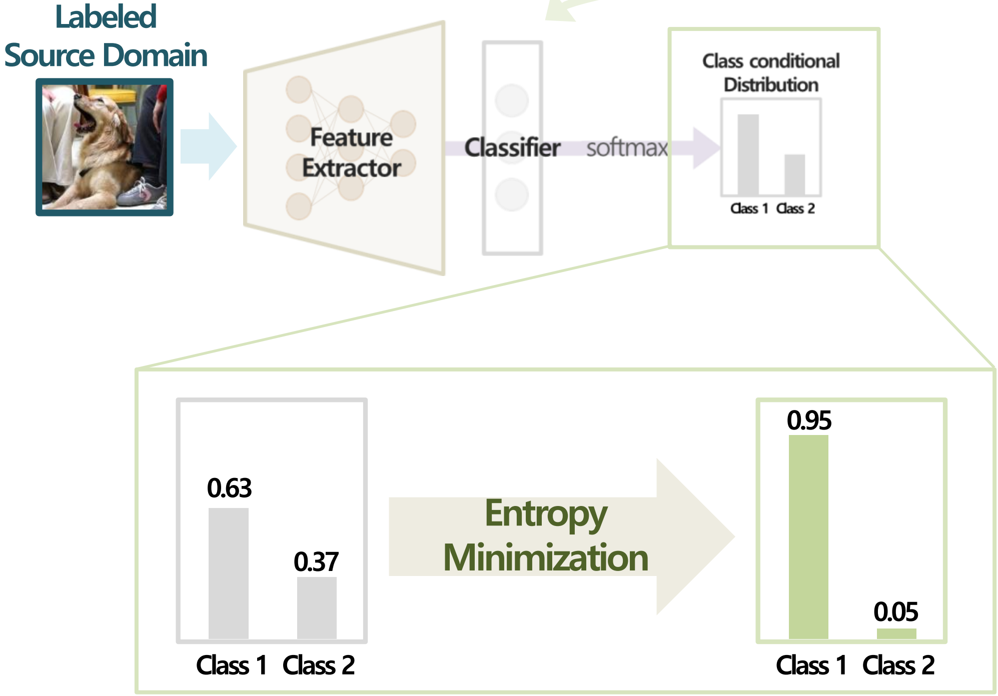


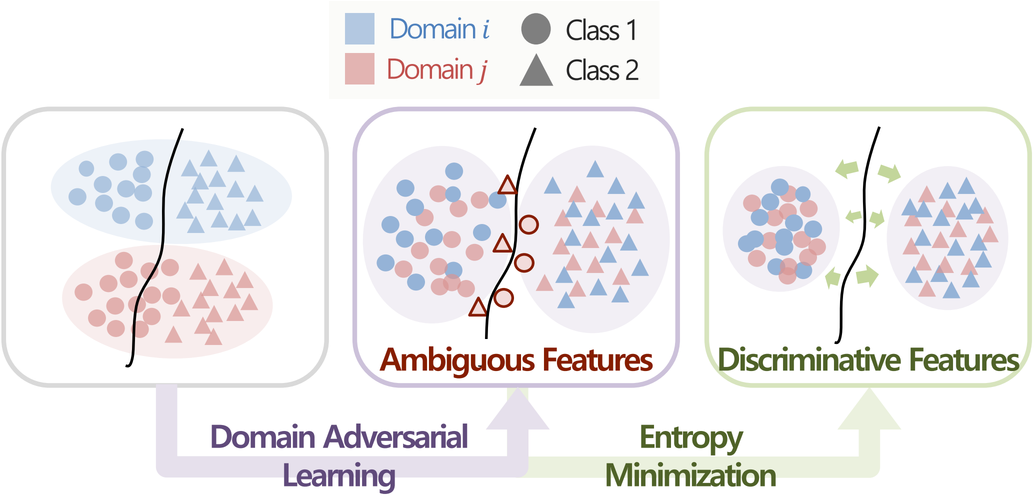
❖ Domain Generalization using a Mixture of Multiple Latent Domains[7]

- **Motivation** : 사전에 Domain Label이 없는 경우에도 가능한 Domain Adversarial Learning 방법론이 필요
  - (1) Pseudo Domain Labels 생성을 통한 Adversarial Learning 수행, (2) **Entropy Minimization**을 통한 Discriminative Features 추출

→ Class-discriminative Features 추출 필요!

→ **가설:** Minimizing class-conditional distribution entropy encourages the **low-density separation** between classes



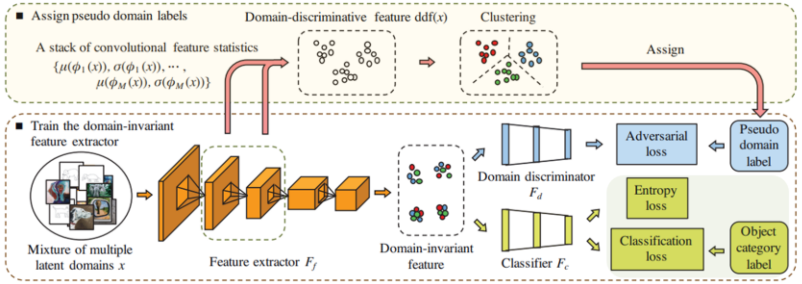


[7] Matsuura, T., & Harada, T. (2020, April). Domain generalization using a mixture of multiple latent domains. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 07, pp. 11749-11756).

Methods

1. Domain-invariant Representation Learning with Adversarial Learning

❖ Domain Generalization using a Mixture of Multiple Latent Domains[7]

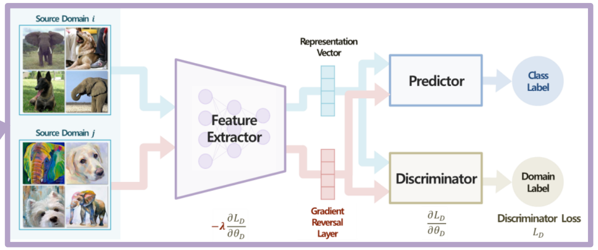


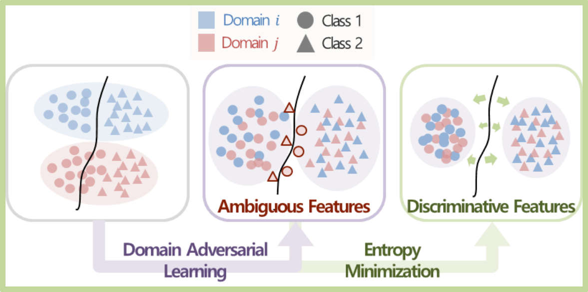
- **Ablation Study 결과**
  - Proposed Method의 Component 별 효과를 알아보기 위한 실험
  - 각 component를 제거(w/o; without)했을 때 도출된 성능을 비교하여 효과 검증

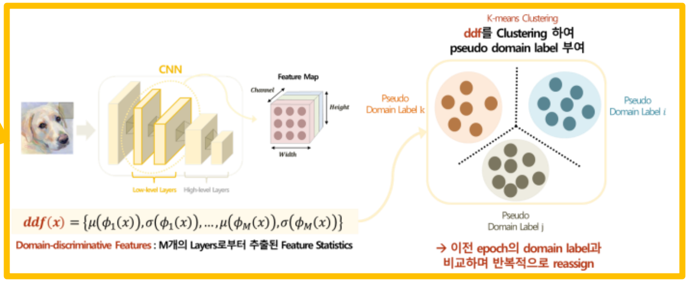
Supervised Learning
Domain Adversarial Learning
Entropy Minimization
Domain Discriminative Features
Reassigning Pseudo Domain Label
Pseudo Doman Label

PACS	Art.	Cartoon	Sketch	Photo	Avg.
AlexNet					
Deep All	68.09	70.23	61.80	88.86	72.25
Ours w/o L_{adv}	67.66	70.45	62.56	88.94	72.40
Ours w/o L_{ent}	68.31	71.13	65.26	**89.38**	73.52
Ours w/o stat.	67.37	70.22	63.12	89.20	72.48
Ours w/o iter.	69.13	70.72	65.41	89.11	73.59
Ours w/o clus.	68.49	72.24	66.31	89.27	74.08
Ours	**69.27**	**72.83**	**66.44**	88.98	**74.38**

Table 3: Results of the ablation study in the PACS dataset. For details about the meaning of columns, see Table 1.







[7] Matsuura, T., & Harada, T. (2020, April). Domain generalization using a mixture of multiple latent domains. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 07, pp. 11749-11756).

230721 DMQA Open Seminar:
Domain Generalization : How to improve the generalization ability of deep learning models?

2. Methods

(2) Data Manipulation

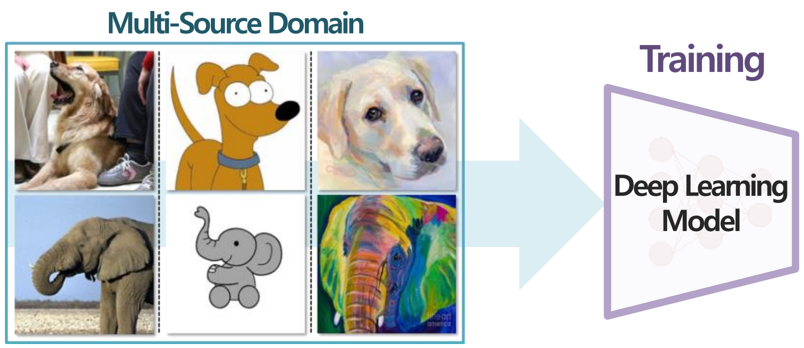
Domain Generalization Methods

2. Data Manipulation

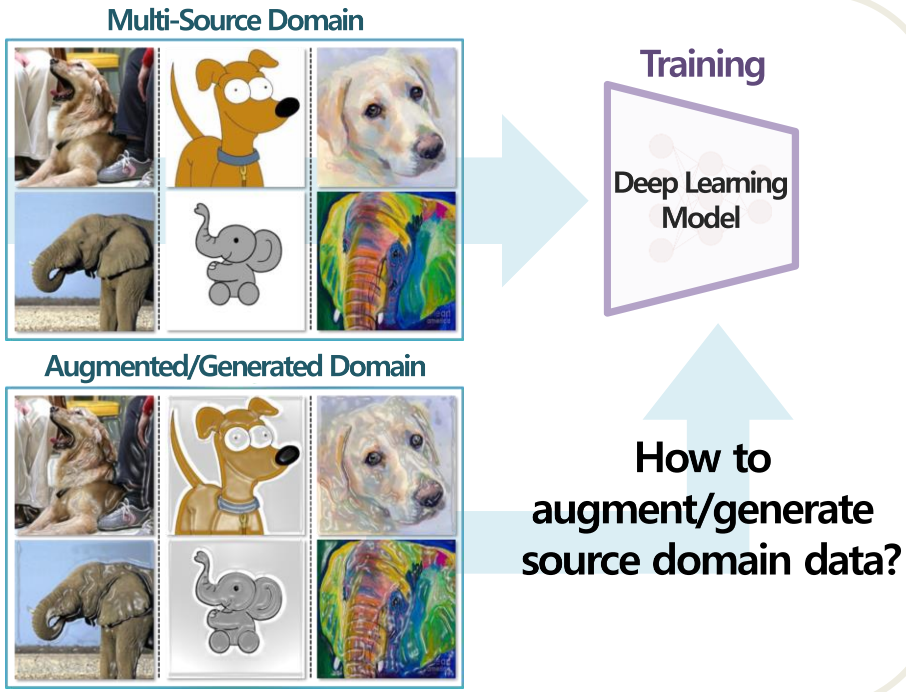
❖ Data Manipulation based Domain Generalization

- **Motivation** : 주어진 Multi-Source Domain만을 가지고 학습을 수행하면, 여전히 Source Domain에 과적합 될 확률이 높다
  - Source Domain의 domain-related properties를 바꾸어 새로운 domain의 데이터를 생성하면, 보다 일반화 된 성능 향상이 가능할 것

Representation Learning based Domain Generalization → Data Manipulation based Domain Generalization



학습 중 활용되는 Source Domain에 과적합(overfitting) 될 수 있는 한계



Domain Generalization Methods

2. Data Manipulation

❖ Data Manipulation based Domain Generalization

1. Data Augmentation
2. Data Generation

Making $x' = \mathcal{M}(x)$, where $\mathcal{M}(\cdot)$ denotes a data manipulating function

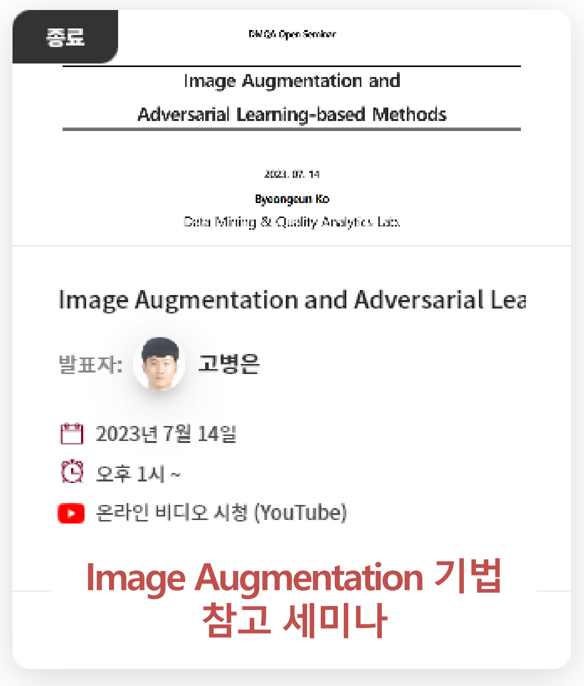


Image Augmentation 기법
참고 세미나

[12] Zhou, K., Liu, Z., Qiao, Y., Xiang, T., & Loy, C. C. (2022). Domain generalization: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Domain Generalization Methods

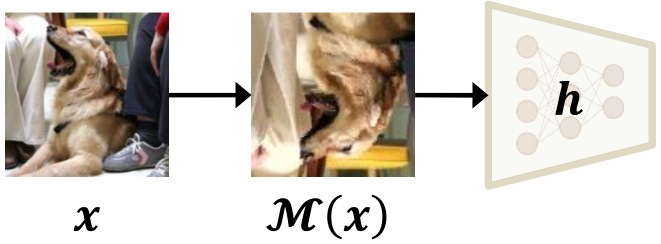
2. Data Manipulation

❖ Data Manipulation based Domain Generalization

1. Data Augmentation
2. Data Generation

Making $x' = \mathcal{M}(x)$, where $\mathcal{M}(\cdot)$ denotes a data manipulating function

Image Transformations



e.g., Random Flip, Rotation and Color augmentation

- Domain Label을 필요로 하지 않는 장점
- 어떤 증강기법을 사용할 지는 problem and data specific
  - ✓ Digit Recognition의 경우 Flip으로 인한 Label shift 발생 가능

[12]

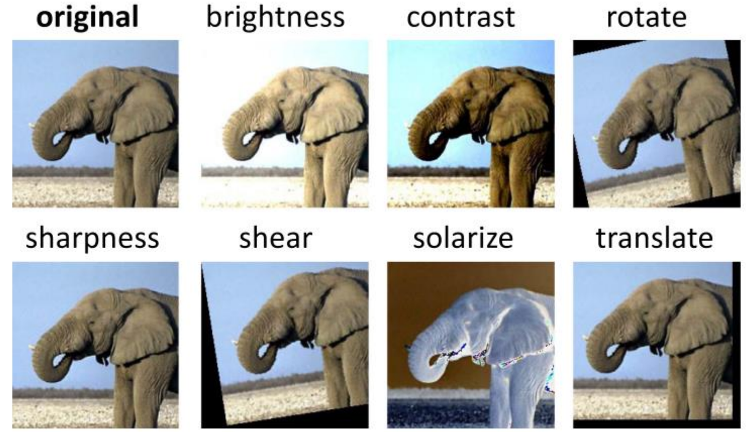


Fig. 5. Common image transformations used as data augmentation in domain generalization [39], [186], [187], [188].

[12] Zhou, K., Liu, Z., Qiao, Y., Xiang, T., & Loy, C. C. (2022). Domain generalization: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence.

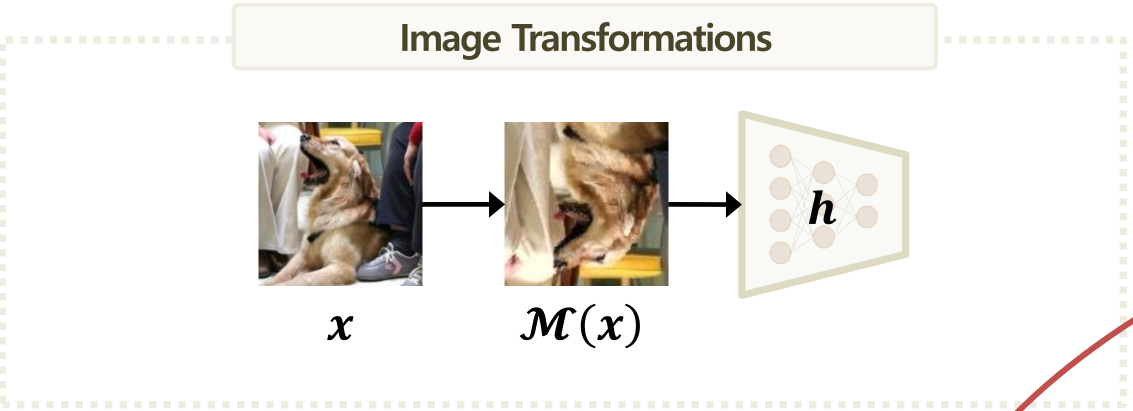
Domain Generalization Methods

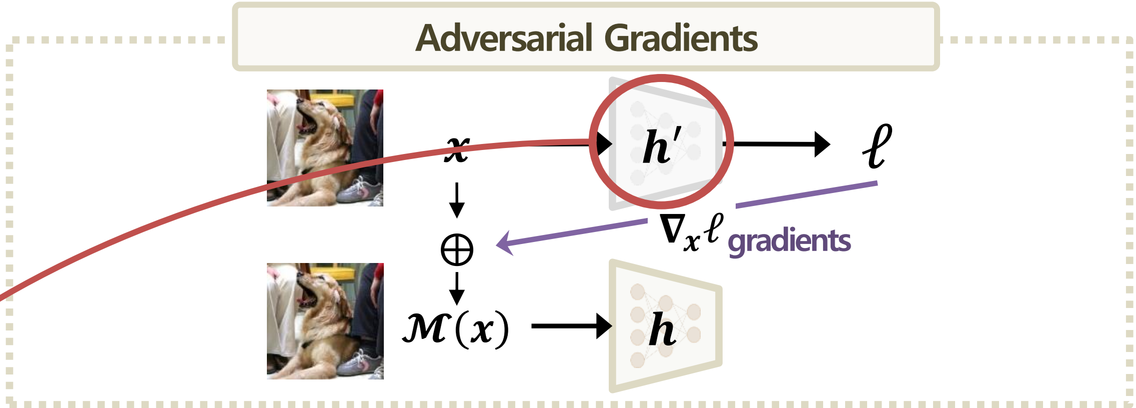
2. Data Manipulation

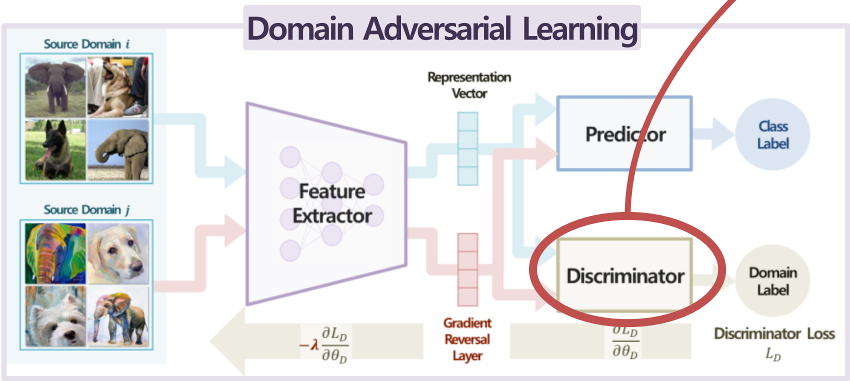
❖ Data Manipulation based Domain Generalization

1. Data Augmentation
2. Data Generation

Making $x' = \mathcal{M}(x)$, where $\mathcal{M}(\cdot)$ denotes a data manipulating function







e.g., CrossGrad[13]

Input sample x에 Domain Discriminator(h')의 gradient를 더해주는 방식
(Domain이 달라지는 방향으로 Input sample을 변형)

$$x' \leftarrow x + \epsilon \nabla L_D$$

- Only make visually imperceptible data
- Not being able to simulate real-world domain shift
- Computational cost is often doubled

[13] Shankar, S., Piratla, V., Chakrabarti, S., Chaudhuri, S., Jyothi, P., & Sarawagi, S. (2018). Generalizing across domains via cross-gradient training. arXiv preprint arXiv:1804.10745.

Domain Generalization Methods

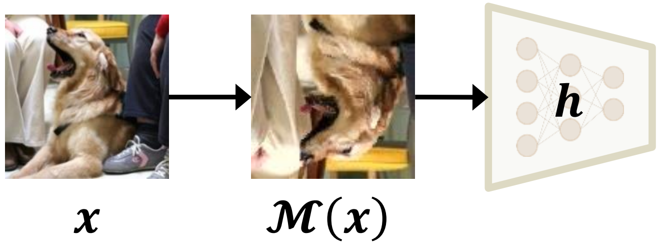
2. Data Manipulation

❖ Data Manipulation based Domain Generalization

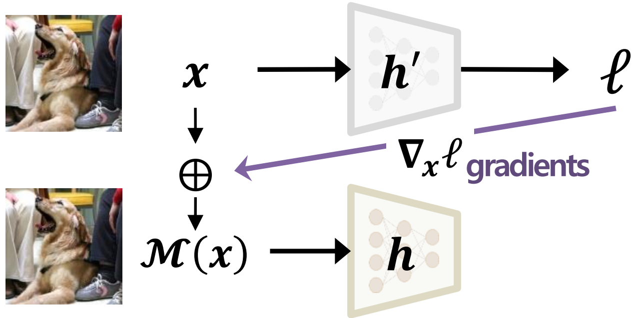
1. Data Augmentation
2. Data Generation

Making $x' = \mathcal{M}(x)$, where $\mathcal{M}(\cdot)$ denotes a data manipulating function

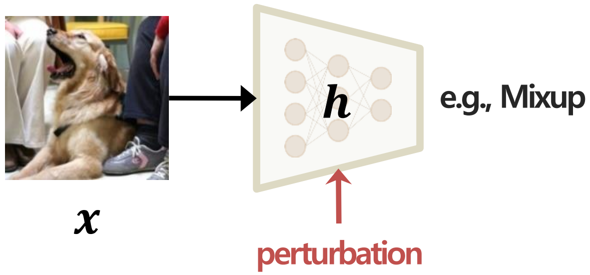
Image Transformations

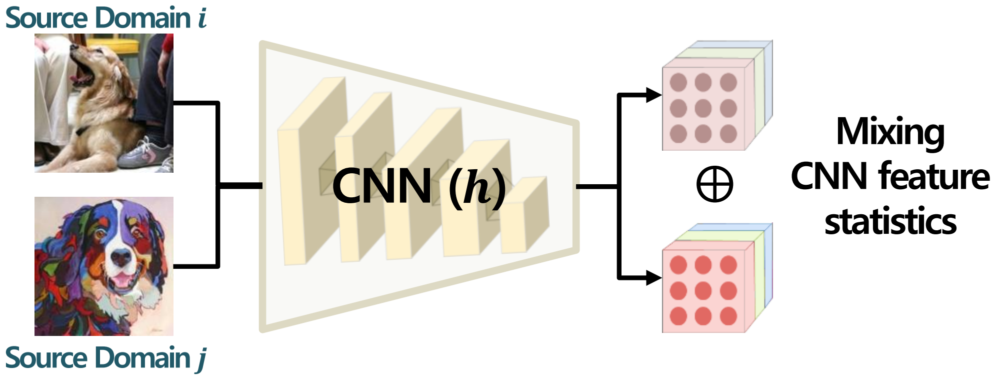


Adversarial Gradients



Feature-based Manipulation





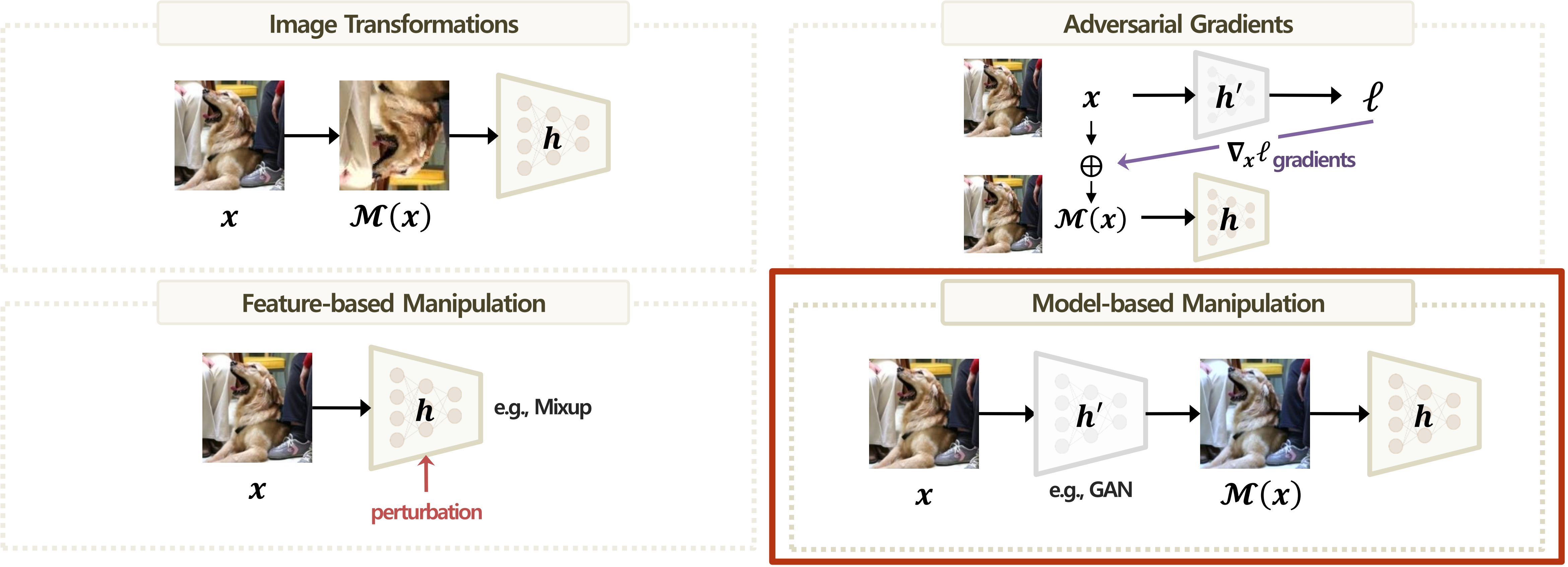
Domain Generalization Methods

2. Data Manipulation

❖ Data Manipulation based Domain Generalization

1. Data Augmentation
2. Data Generation

→ **Making $x' = \mathcal{M}(x)$, where $\mathcal{M}(\cdot)$ denotes a data manipulating function**



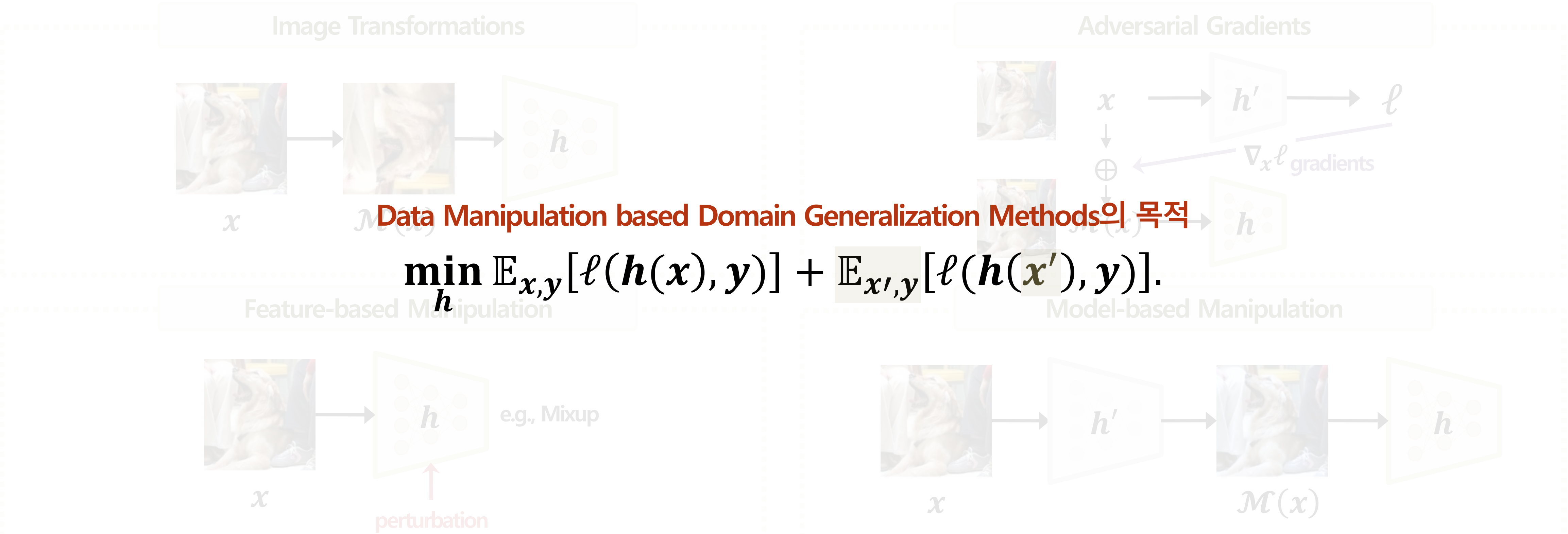
Domain Generalization Methods

2. Data Manipulation

❖ Data Manipulation based Domain Generalization

1. Data Augmentation
2. Data Generation

Making $x' = \mathcal{M}(x)$, where $\mathcal{M}(\cdot)$ denotes a data manipulating function



Data Manipulation based Domain Generalization Methods의 목적

$$\min_{h} \mathbb{E}_{x,y}[\ell(h(x), y)] + \mathbb{E}_{x',y}[\ell(h(x'), y)].$$

Domain Generalization Methods

2. Data Manipulation: (1) Data Augmentation

❖ Deep Domain-adversarial Image Generation for Domain Generalisation[14]

- 2020년 제안된 Domain Generalization 방법론 (AAAI, 23년 7월 기준 221회 인용)
- Domain Adversarial Learning을 기반으로 하는 Augmentation 기법 도입

Deep Domain-Adversarial Image Generation for Domain Generalisation

Kaiyang Zhou[1], Yongxin Yang[1], Timothy Hospedales[2,3], and Tao Xiang[1,3]
[1]University of Surrey, [2]University of Edinburgh, [3]Samsung AI Center, Cambridge
{k.zhou, yongxin.yang, t.xiang}@surrey.ac.uk, t.hospedales@ed.ac.uk

Abstract

Machine learning models typically suffer from the domain shift problem when trained on a source dataset and evaluated on a target dataset of different distribution. To overcome this problem, domain generalisation (DG) methods aim to leverage data from multiple source domains so that a trained model can generalise to unseen domains. In this paper, we propose a novel DG approach based on *Deep Domain-Adversarial Image Generation* (DDAIG). Specifically, DDAIG consists of three components, namely a label classifier, a domain classifier and a domain transformation network (DoTNet). The goal for DoTNet is to map the source training data to unseen domains. This is achieved by having a learning objective formulated to ensure that the generated data can be correctly classified by the label classifier while fooling the domain classifier. By augmenting the source training data with the generated unseen domain data, we can make the label classifier more robust to unknown domain changes. Extensive experiments on four DG datasets demonstrate the effectiveness of our approach.

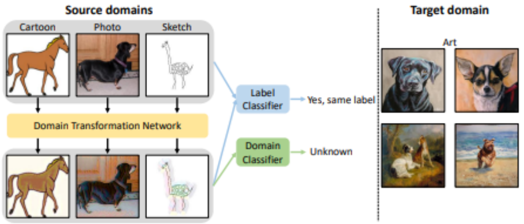


Figure 1: Given multiple source domains, e.g. Cartoon, Photo and Sketch, we learn a domain transformation network (DoTNet) to transform images to unseen domains, which maintain the class labels but change domain-related properties. Both original and transformed images are used to train a label classifier, which is applied to an unseen target domain, e.g. Art.

get data and perform supervised model fine-tuning. However, large-scale data collection and annotation for every new target domain is prohibitively expensive and time-

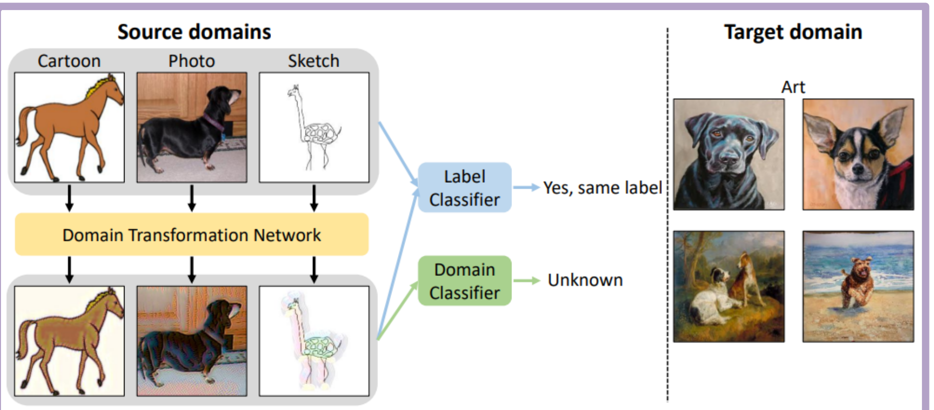


Figure 1: Given multiple source domains, e.g. Cartoon, Photo and Sketch, we learn a domain transformation network (DoTNet) to transform images to unseen domains, which maintain the class labels but change domain-related properties. Both original and transformed images are used to train a label classifier, which is applied to an unseen target domain, e.g. Art.

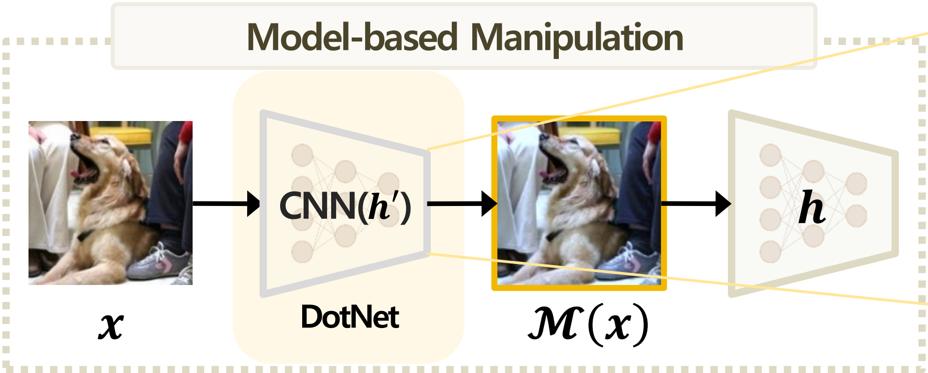
[14] Zhou, K., Yang, Y., Hospedales, T., & Xiang, T. (2020, April). Deep domain-adversarial image generation for domain generalisation. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 07, pp. 13025-13032).

Domain Generalization Methods

2. Data Manipulation: (1) Data Augmentation

❖ Deep Domain-adversarial Image Generation for Domain Generalisation[14]

- **Motivation** : Domain이 기존과 크게 달라지면서도 Semantic 정보는 유지하도록 데이터를 생성하자
  - Domain Discriminator를 속일 수 있을 만한 이미지를 생성하는 동시에 Label Predictor는 Domain 차이에 강건해지도록 학습



Can makes more sophisticated and more overt perturbations!

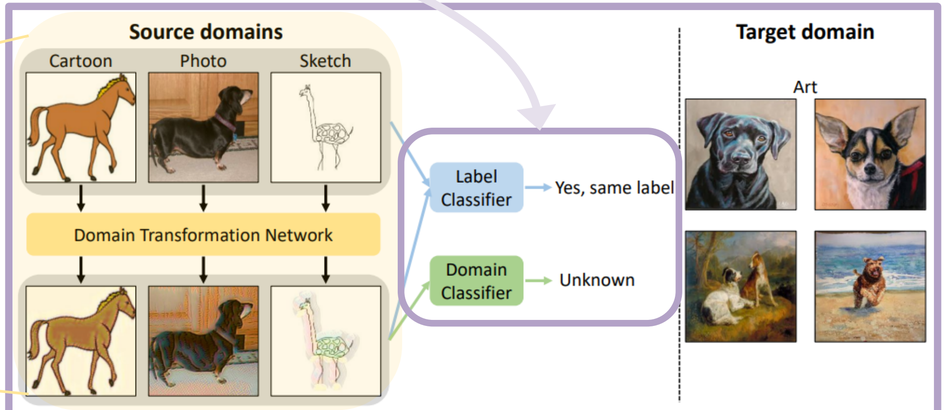


Figure 1: Given multiple source domains, e.g. Cartoon, Photo and Sketch, we learn a domain transformation network (DoTNet) to transform images to unseen domains, which maintain the class labels but change domain-related properties. Both original and transformed images are used to train a label classifier, which is applied to an unseen target domain, e.g. Art.

[14] Zhou, K., Yang, Y., Hospedales, T., & Xiang, T. (2020, April). Deep domain-adversarial image generation for domain generalisation. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 07, pp. 13025-13032).

Domain Generalization Methods

2. Data Manipulation: (1) Data Augmentation

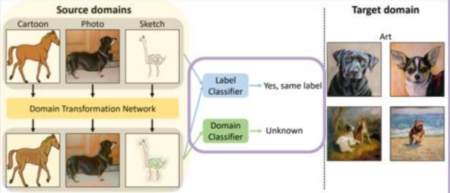


Figure 1: Given multiple source domains, e.g. Cartoon, Photo and Sketch, we learn a domain transformation network (DoTNet) to transform images to unseen domains, which maintain the class labels but change domain-related properties. Both original and transformed images are used to train a label classifier, which is applied to an unseen target domain, e.g. Art.

❖ Deep Domain-adversarial Image Generation for Domain Generalisation[14]

- **DDAIG** : (1) Domain Transformation Network (DoTNet), (2) Label Classifier, (3) Domain Classifier
  - 총 3가지 모델로 구성

DDAIG[14]

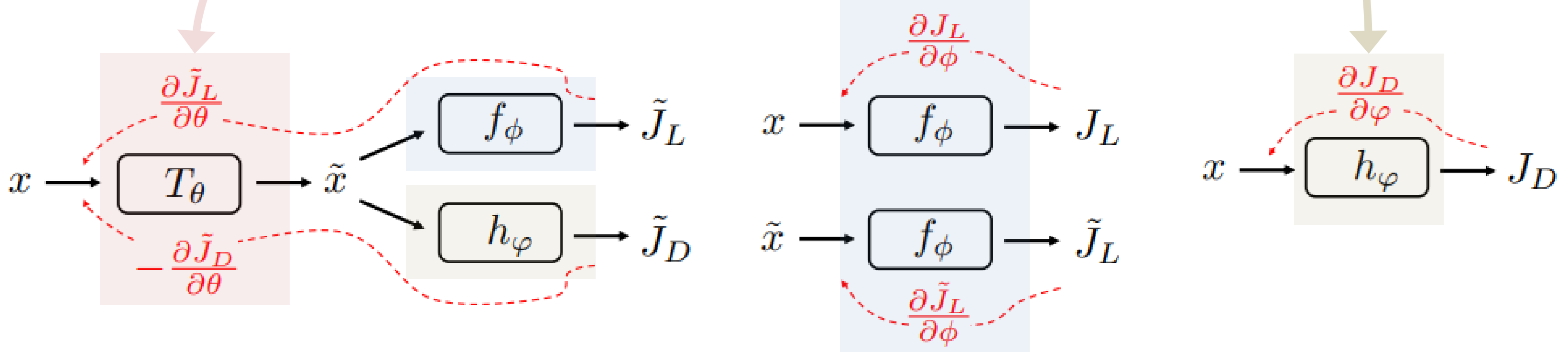


Figure 2: Overview of our framework. A domain transformation network T_θ is trained by minimising the label classification loss $\tilde{J}_L$ while maximising the domain classification loss $\tilde{J}_D$ on the transformed data $\tilde{x}$. The label classifier f_ϕ is learned by minimising the label classification loss given both original and transformed data. The domain classifier h_φ is trained to classify each instance into one of source domains. The red dashed arrows represent the gradient flow.

[14] Zhou, K., Yang, Y., Hospedales, T., & Xiang, T. (2020, April). Deep domain-adversarial image generation for domain generalisation. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 07, pp. 13025-13032).

Domain Generalization Methods

2. Data Manipulation: (1) Data Augmentation

DDAIG[14]

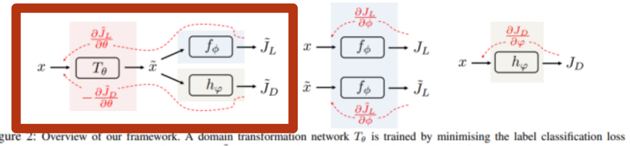
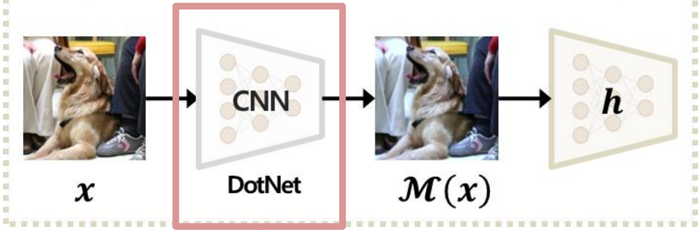


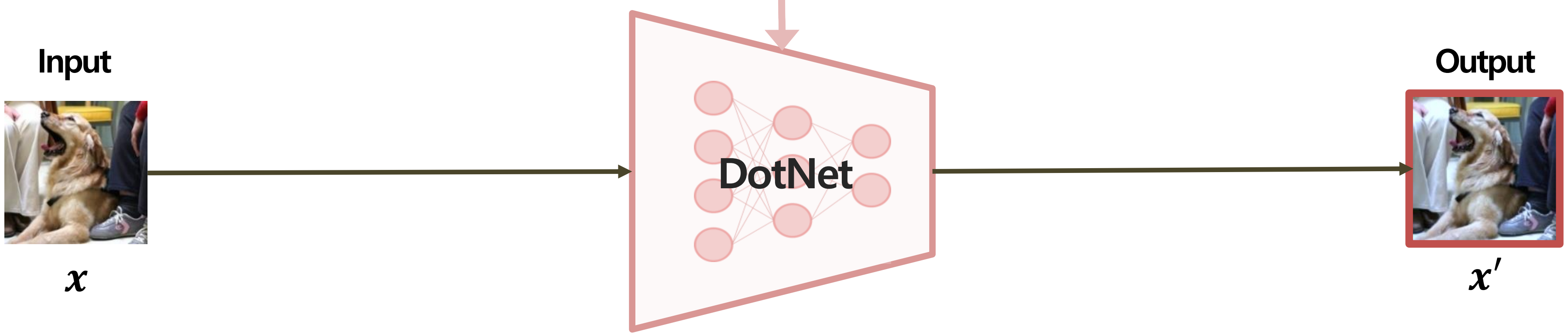
Figure 2: Overview of our framework. A domain transformation network T_θ is trained by minimising the label classification loss $\tilde{J}_L$ while maximising the domain classification loss $\tilde{J}_D$ on the transformed data $\tilde{x}$. The label classifier f_ϕ is learned by minimising the label classification loss given both original and transformed data. The domain classifier h_φ is trained to classify each instance into one of source domains. The red dashed arrows represent the gradient flow.

❖ Deep Domain-adversarial Image Generation for Domain Generalisation[14]

- **DDAIG** : (1) Domain Transformation Network (DoTNet), (2) Label Classifier, (3) Domain Classifier
  - DotNet은 perturbed image를 생성하는 CNN 모듈

Model-based Manipulation





[14] Zhou, K., Yang, Y., Hospedales, T., & Xiang, T. (2020, April). Deep domain-adversarial image generation for domain generalisation. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 07, pp. 13025-13032).

Domain Generalization Methods

2. Data Manipulation: (1) Data Augmentation

DDAIG[14]

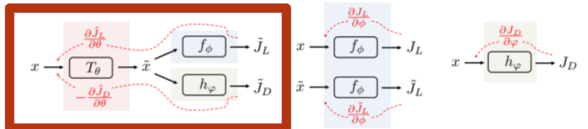
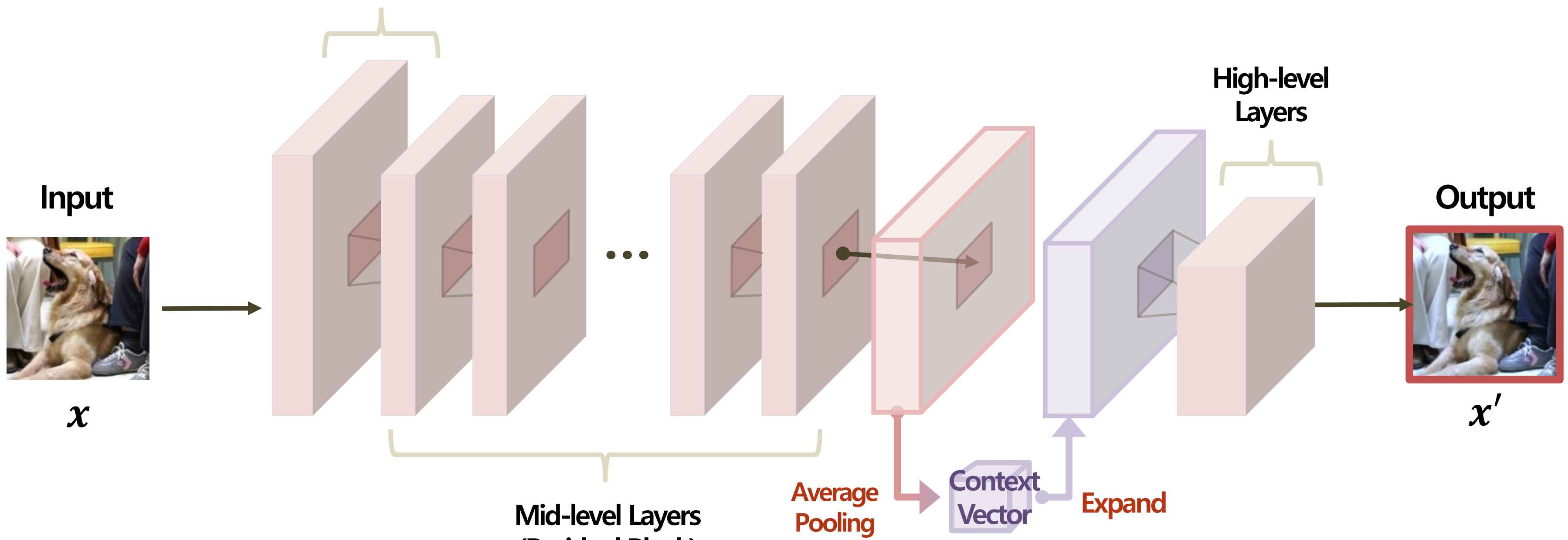


Figure 2: Overview of our framework. A domain transformation network T_θ is trained by minimising the label classification loss $\tilde{J}_L$ while maximising the domain classification loss $\tilde{J}_D$ on the transformed data $\tilde{x}$. The label classifier f_ϕ is learned by minimising the label classification loss given both original and transformed data. The domain classifier h_φ is trained to classify each instance into one of source domains. The red dashed arrows represent the gradient flow.

❖ Deep Domain-adversarial Image Generation for Domain Generalisation[14]

- **DDAIG** : (1) Domain Transformation Network (DoTNet), (2) Label Classifier, (3) Domain Classifier
  - DotNet은 perturbed image를 생성하는 CNN 모듈



[14] Zhou, K., Yang, Y., Hospedales, T., & Xiang, T. (2020, April). Deep domain-adversarial image generation for domain generalisation. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 07, pp. 13025-13032).

Domain Generalization Methods

2. Data Manipulation: (1) Data Augmentation

DDAIG[14]

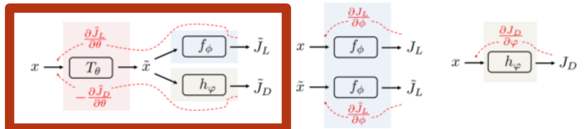
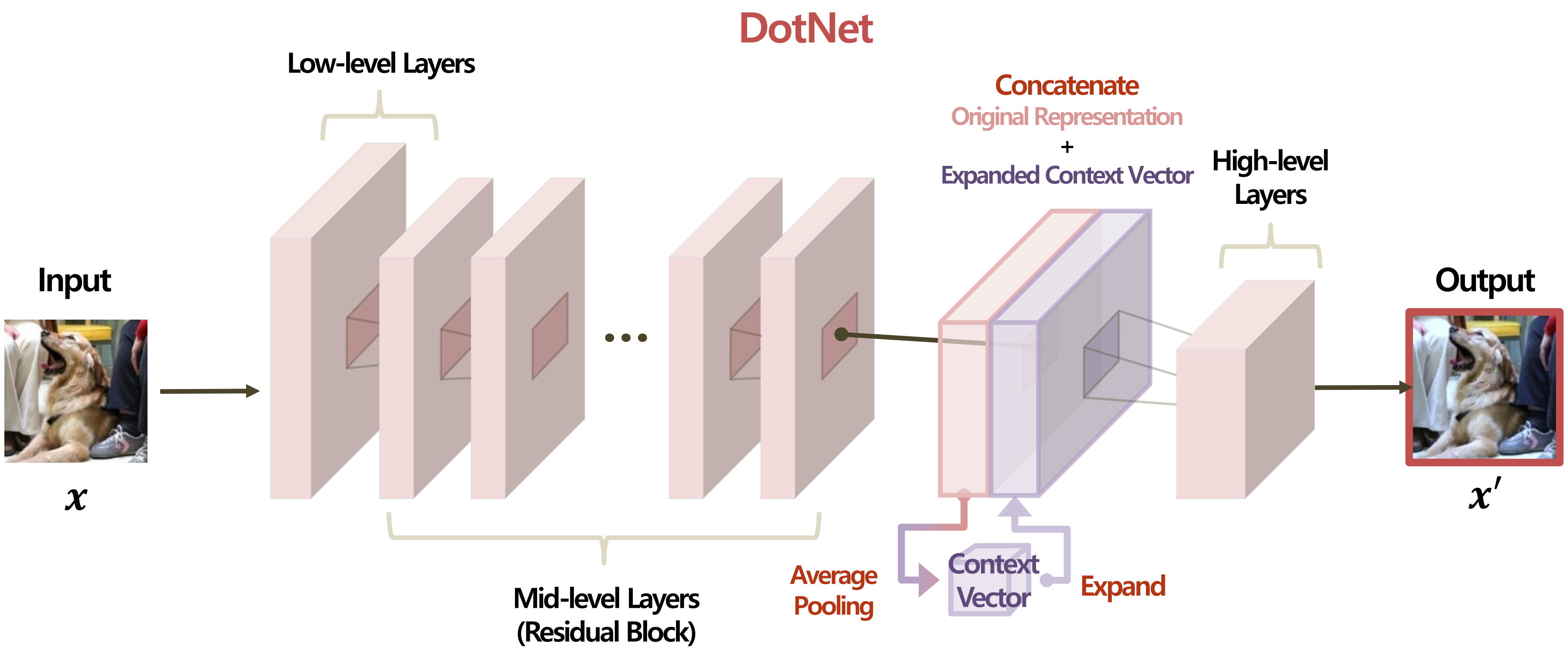


Figure 2: Overview of our framework. A domain transformation network T_θ is trained by minimising the label classification loss $\tilde{J}_L$ while maximising the domain classification loss $\tilde{J}_D$ on the transformed data $\tilde{x}$. The label classifier f_ϕ is learned by minimising the label classification loss given both original and transformed data. The domain classifier h_φ is trained to classify each instance into one of source domains. The red dashed arrows represent the gradient flow.

❖ Deep Domain-adversarial Image Generation for Domain Generalisation[14]

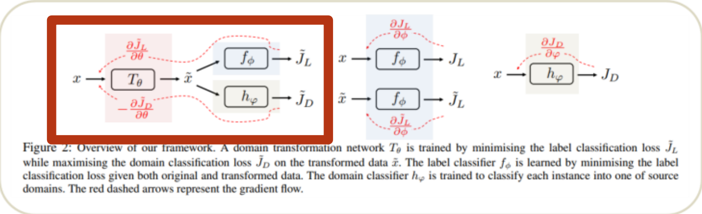
- **DDAIG** : (1) Domain Transformation Network (DoTNet), (2) Label Classifier, (3) Domain Classifier
  - DotNet은 perturbed image를 생성하는 CNN 모듈



[14] Zhou, K., Yang, Y., Hospedales, T., & Xiang, T. (2020, April). Deep domain-adversarial image generation for domain generalisation. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 07, pp. 13025-13032).

Domain Generalization Methods

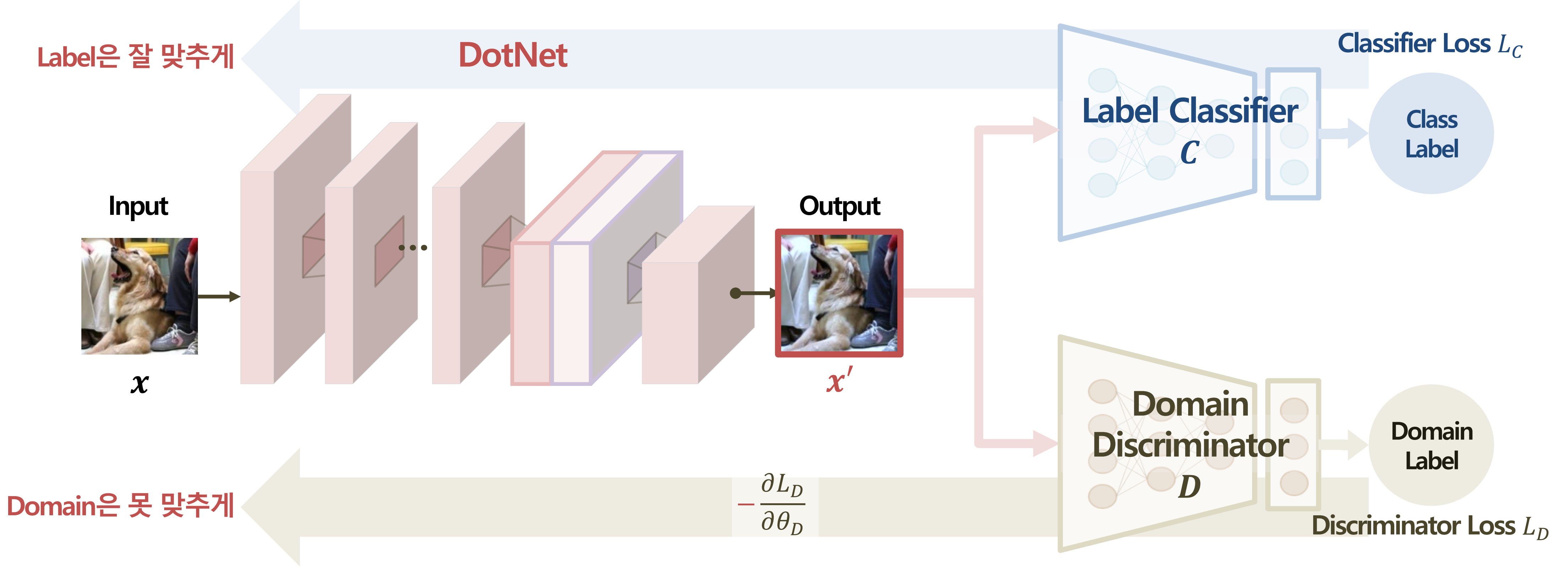
2. Data Manipulation: (1) Data Augmentation



'Classifier'는 Feature Extractor와 Classification Layer가 합해진 형태를 의미

❖ Deep Domain-adversarial Image Generation for Domain Generalisation[14]

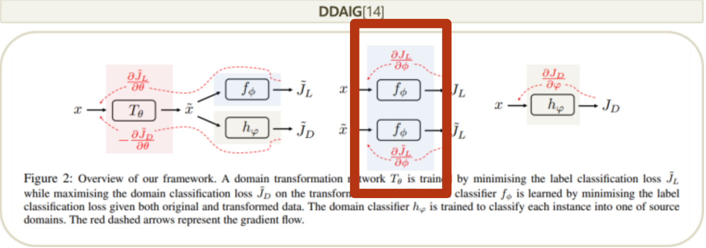
- **DDAIG** : (1) Domain Transformation Network (DoTNet), (2) Label Classifier, (3) Domain Classifier
  - DotNet은 perturbed image를 생성하는 CNN 모듈 → **x'가 입력 값일 때 L_C - L_D가 최소화 되는 방향으로 학습**



[14] Zhou, K., Yang, Y., Hospedales, T., & Xiang, T. (2020, April). Deep domain-adversarial image generation for domain generalisation. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 07, pp. 13025-13032).

Domain Generalization Methods

2. Data Manipulation: (1) Data Augmentation



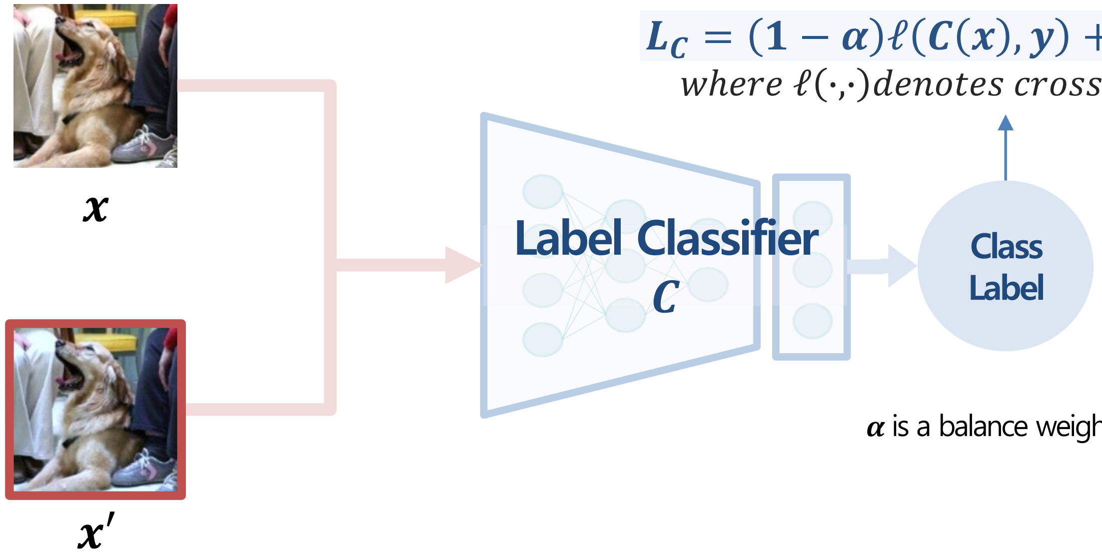
'Classifier'는 Feature Extractor와 Classification Layer가 합해진 형태를 의미

❖ Deep Domain-adversarial Image Generation for Domain Generalisation[14]

- **DDAIG** : (1) Domain Transformation Network (DoTNet), (2) Label Classifier, (3) Domain Classifier
  - Label Classifier는 Task를 잘 수행하기 위한 Predictor → **x와 x'를 모두 입력으로 받아 L_C가 최소화 되는 방향으로 학습**

Classifier Loss L_C

$$L_C = (1-\alpha)\ell(C(x), y) + \alpha\ell(C(x'), y),$$
$$where\ \ell(\cdot,\cdot) denotes\ cross\ entropy\ loss$$



α is a balance weight (0.5)

[14] Zhou, K., Yang, Y., Hospedales, T., & Xiang, T. (2020, April). Deep domain-adversarial image generation for domain generalisation. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 07, pp. 13025-13032).

Domain Generalization Methods

2. Data Manipulation: (1) Data Augmentation

DDAIG[14]

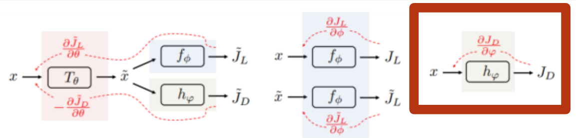


Figure 2: Overview of our framework. A domain transformation network T_θ is trained by minimising the label classification loss $\tilde{J}_L$ while maximising the domain classification loss $\tilde{J}_D$ on the transformed data $\tilde{x}$. The label classifier f_ϕ is learned by minimising the label classification loss given both original and transformed data. The domain classifier h_φ is trained to classify each instance into one of source domains. The red dashed arrows represent the gradient flow.

'Classifier'는 Feature Extractor와 Classification Layer가 합해진 형태를 의미

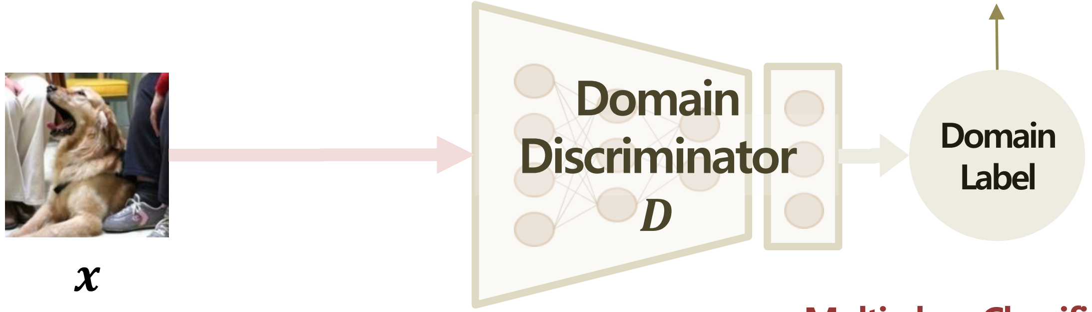
❖ **Deep Domain-adversarial Image Generation for Domain Generalisation**[14]

- **DDAIG** : (1) Domain Transformation Network (DoTNet), (2) Label Classifier, (3) Domain Classifier
  - Domain Classifier는 Domain을 잘 구분하기 위한 Predictor → **x를 입력으로 받아 L_D가 최소화 되는 방향으로 학습**

Discriminator Loss L_D

$$L_D = \ell(D(x), d),$$

where $\ell(\cdot,\cdot)$ denotes cross entropy loss and d is the domain label



Multi-class Classification!
(다른 Source Domain에 가까운 이미지를 생성하는 것을 방지)

[14] Zhou, K., Yang, Y., Hospedales, T., & Xiang, T. (2020, April). Deep domain-adversarial image generation for domain generalisation. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 07, pp. 13025-13032).

Domain Generalization Methods

2. Data Manipulation: (2) Data Generation

❖ Learning to Generate Novel Domains for Domain Generalization[15]

- 2020년 제안된 Domain Generalization 방법론 (ECCV, 23년 7월 기준 259회 인용)
- Conditional Generator를 기반으로 데이터를 생성하여 학습 중 활용 가능한 Domain의 다양성 확보

Learning to Generate Novel Domains for Domain Generalization

Kaiyang Zhou[1], Yongxin Yang[1], Timothy Hospedales[2,3], and Tao Xiang[1,3]

[1] University of Surrey
{k.zhou, yongxin.yang, t.xiang}@surrey.ac.uk
[2] University of Edinburgh
t.hospedales@ed.ac.uk
[3] Samsung AI Center, Cambridge

Abstract. This paper focuses on domain generalization (DG), the task of learning from multiple source domains a model that generalizes well to unseen domains. A main challenge for DG is that the available source domains often exhibit limited diversity, hampering the model's ability to learn to generalize. We therefore employ a data generator to synthesize data from pseudo-novel domains to augment the source domains. This explicitly increases the diversity of available training domains and leads to a more generalizable model. To train the generator, we model the distribution divergence between source and synthesized pseudo-novel domains using optimal transport, and maximize the divergence. To ensure that semantics are preserved in the synthesized data, we further impose cycle-consistency and classification losses on the generator. Our method, L2A-OT (Learning to Augment by Optimal Transport) outperforms current state-of-the-art DG methods on four benchmark datasets.

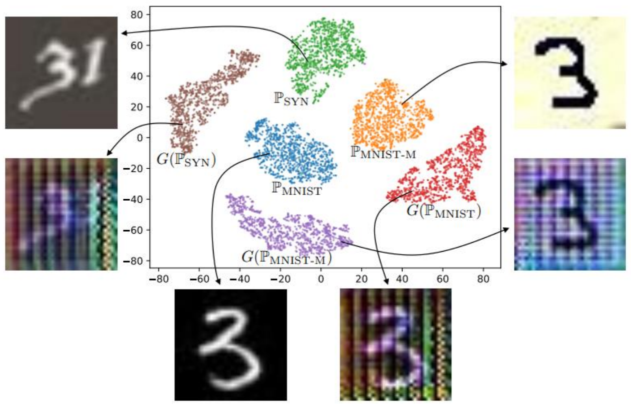


Fig. 1. Motivation of our approach. We improve generalization by increasing the diversity of training domains by learning a generator network G to map images of a source distribution, e.g., $\mathbb{P}_{MNIST}$, to a novel distribution, i.e. $G(\mathbb{P}_{MNIST})$. We then combine both source and novel domains for model learning.

"Increasing the diversity of available source domains (to improve model generalization) "

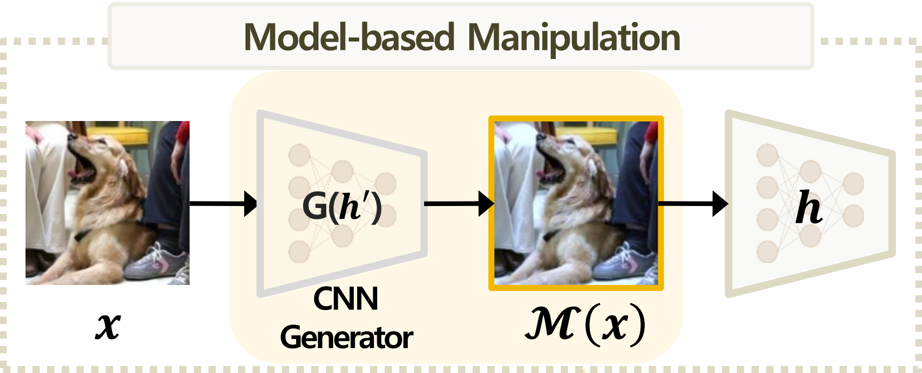
[15] Zhou, K., Yang, Y., Hospedales, T., & Xiang, T. (2020). Learning to generate novel domains for domain generalization. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16 (pp. 561-578). Springer International Publishing.

Domain Generalization Methods

2. Data Manipulation: (2) Data Generation

❖ Learning to Generate Novel Domains for Domain Generalization[15]

- **Motivation** : 기존의 Source Domains 만으로 학습을 수행했을 때, Source Domain에 과적합 될 수 있는 여지를 줄이자
  - Conditional Generator를 통해 K개의 Source Domains 각각에 mapping되는 K개의 Pseudo Novel Domain 생성



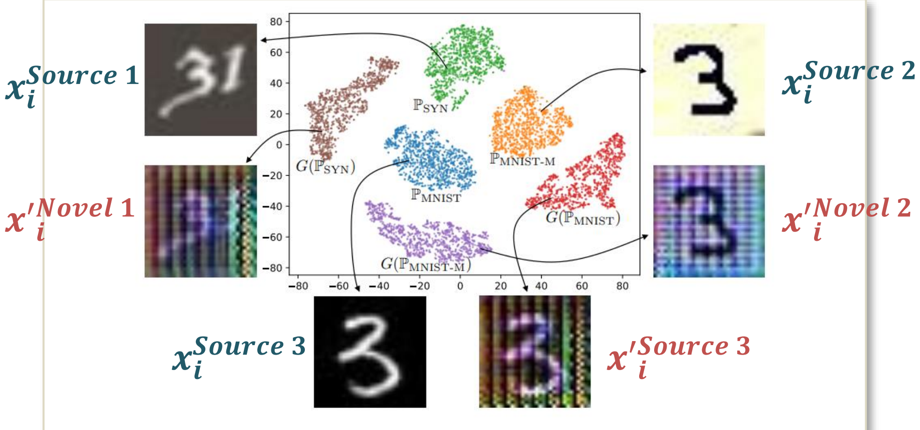


Fig. 1. Motivation of our approach. We improve generalization by increasing the diversity of training domains by learning a generator network G to map images of a source distribution, e.g., $\mathbb{P}_{MNIST}$, to a novel distribution, i.e. $G(\mathbb{P}_{MNIST})$. We then combine both source and novel domains for model learning.

[15] Zhou, K., Yang, Y., Hospedales, T., & Xiang, T. (2020). Learning to generate novel domains for domain generalization. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16 (pp. 561-578). Springer International Publishing.

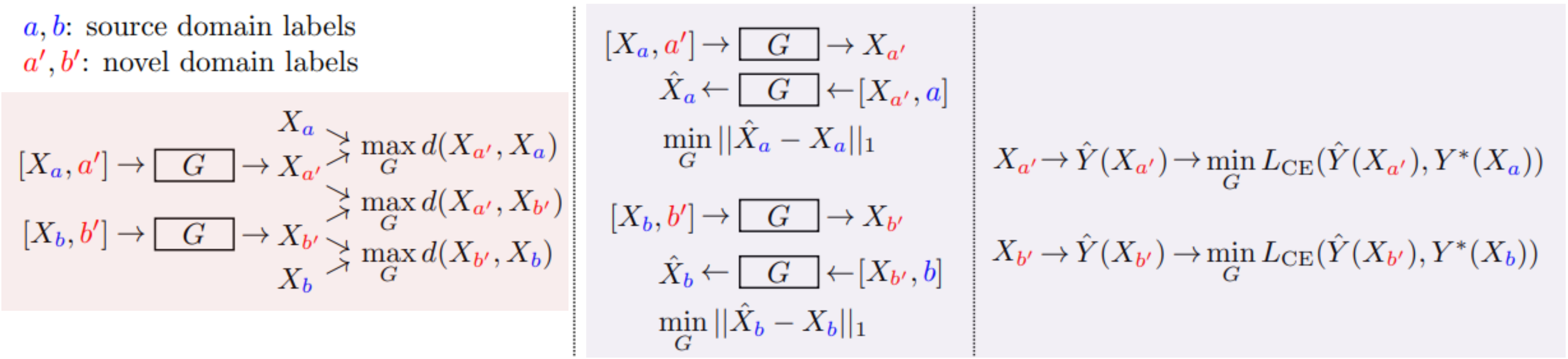
Domain Generalization Methods

2. Data Manipulation: (2) Data Generation

❖ Learning to Generate Novel Domains for Domain Generalization[15]

- **L2A-OT**(Learning to Augment by Optimal Transport) : (1) Conditional Generator G, (2) Classifier $\hat{Y}$
  - 총 2가지 네트워크로 구성되어 있으며, 3단계를 거쳐 학습

L2A-OT[15]



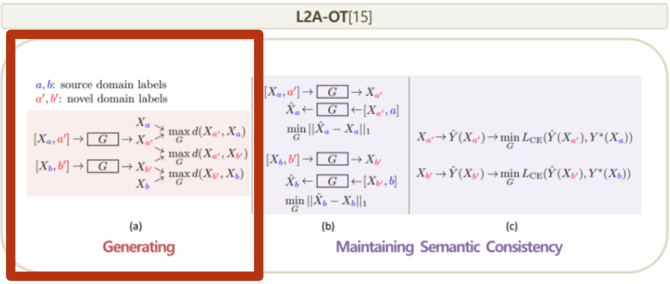
(a) (b) (c)

Generating — Maintaining Semantic Consistency

[15] Zhou, K., Yang, Y., Hospedales, T., & Xiang, T. (2020). Learning to generate novel domains for domain generalization. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16 (pp. 561-578). Springer International Publishing.

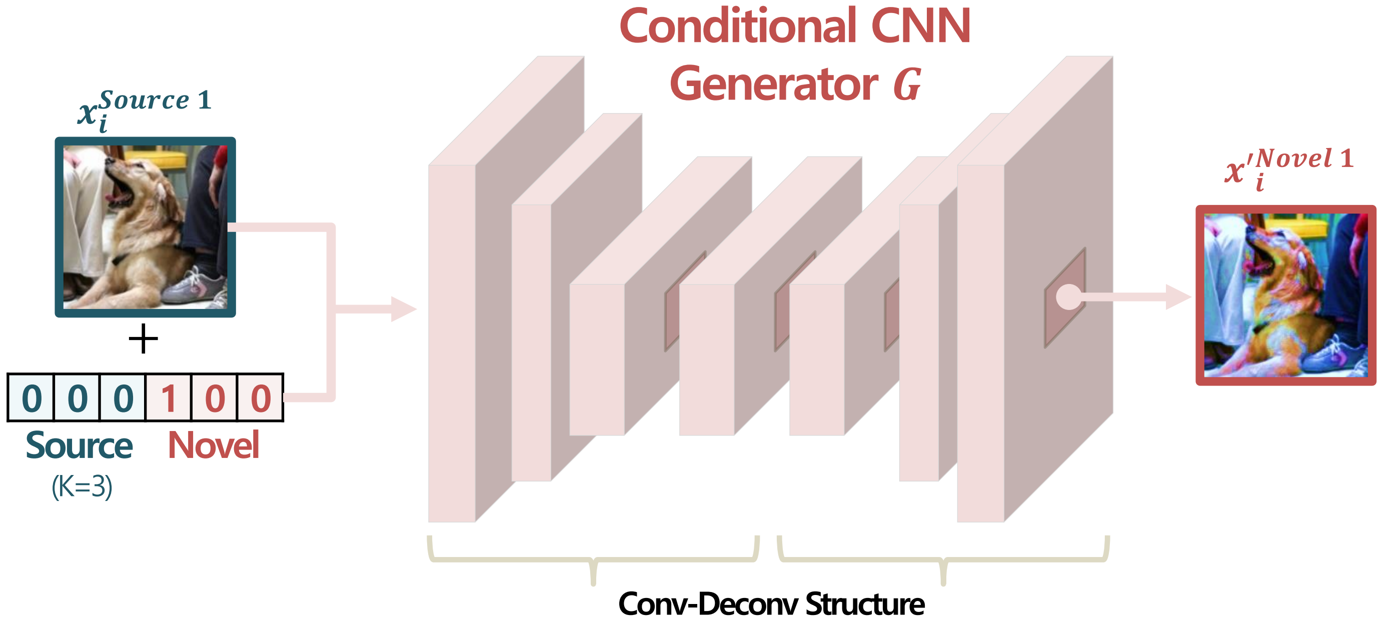
Domain Generalization Methods

2. Data Manipulation: (2) Data Generation



❖ Learning to Generate Novel Domains for Domain Generalization[15]

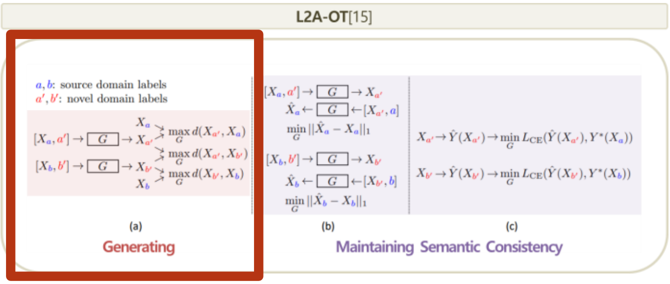
- **L2A-OT**(Learning to Augment by Optimal Transport) : (1) Conditional Generator G, (2) Classifier $\hat{Y}$
  - **(1)-1. Generating** → Domain 간의 Wasserstein distance[16]가 커지는 방향으로 학습하여 다양한 도메인의 데이터 생성



[15] Zhou, K., Yang, Y., Hospedales, T., & Xiang, T. (2020). Learning to generate novel domains for domain generalization. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16 (pp. 561-578). Springer International Publishing.
[16] Arjovsky, M., Chintala, S., & Bottou, L. (2017, July). Wasserstein generative adversarial networks. In International conference on machine learning (pp. 214-223). PMLR.

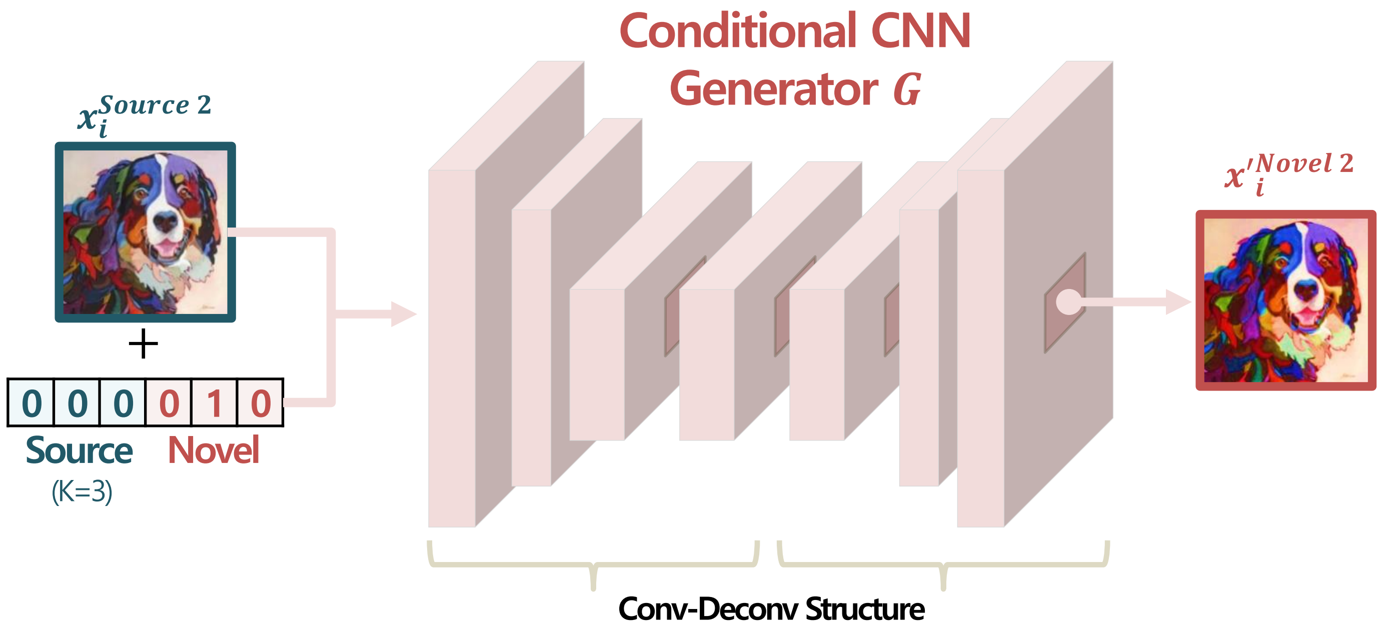
Domain Generalization Methods

2. Data Manipulation: (2) Data Generation



❖ Learning to Generate Novel Domains for Domain Generalization[15]

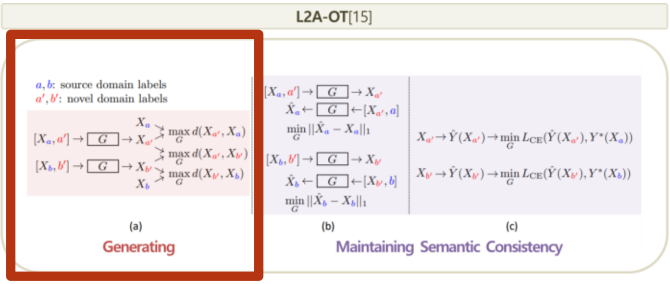
- **L2A-OT**(Learning to Augment by Optimal Transport) : (1) Conditional Generator G, (2) Classifier $\hat{Y}$
  - **(1)-1. Generating** → Domain 간의 Wasserstein distance[16]가 커지는 방향으로 학습하여 다양한 도메인의 데이터 생성



[15] Zhou, K., Yang, Y., Hospedales, T., & Xiang, T. (2020). Learning to generate novel domains for domain generalization. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16 (pp. 561-578). Springer International Publishing.
[16] Arjovsky, M., Chintala, S., & Bottou, L. (2017, July). Wasserstein generative adversarial networks. In International conference on machine learning (pp. 214-223). PMLR.

Domain Generalization Methods

2. Data Manipulation: (2) Data Generation



❖ Learning to Generate Novel Domains for Domain Generalization[15]

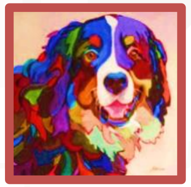
- **L2A-OT**(Learning to Augment by Optimal Transport) : (1) Conditional Generator G, (2) Classifier $\hat{Y}$
  - **(1)-1. Generating** → Domain 간의 Wasserstein distance[16]가 커지는 방향으로 학습하여 다양한 도메인의 데이터 생성

Objective Functions for Generating Novel Domain

$$\max_{G} \ distance\left(x_i^{Source\,1}, \; x_i'^{Novel\,1} \right)$$


Source와 다른 Domain 생성

$$\max_{G} \ distance\left(x_i'^{Novel\,1}, \; x_i'^{Novel\,2} \right)$$

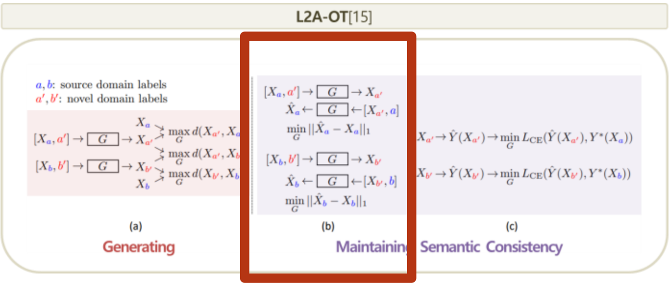
K개의 서로 다른 Domain 생성

Wasserstein Distance[16]

[15] Zhou, K., Yang, Y., Hospedales, T., & Xiang, T. (2020). Learning to generate novel domains for domain generalization. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16 (pp. 561-578). Springer International Publishing.
[16] Arjovsky, M., Chintala, S., & Bottou, L. (2017, July). Wasserstein generative adversarial networks. In International conference on machine learning (pp. 214-223). PMLR.

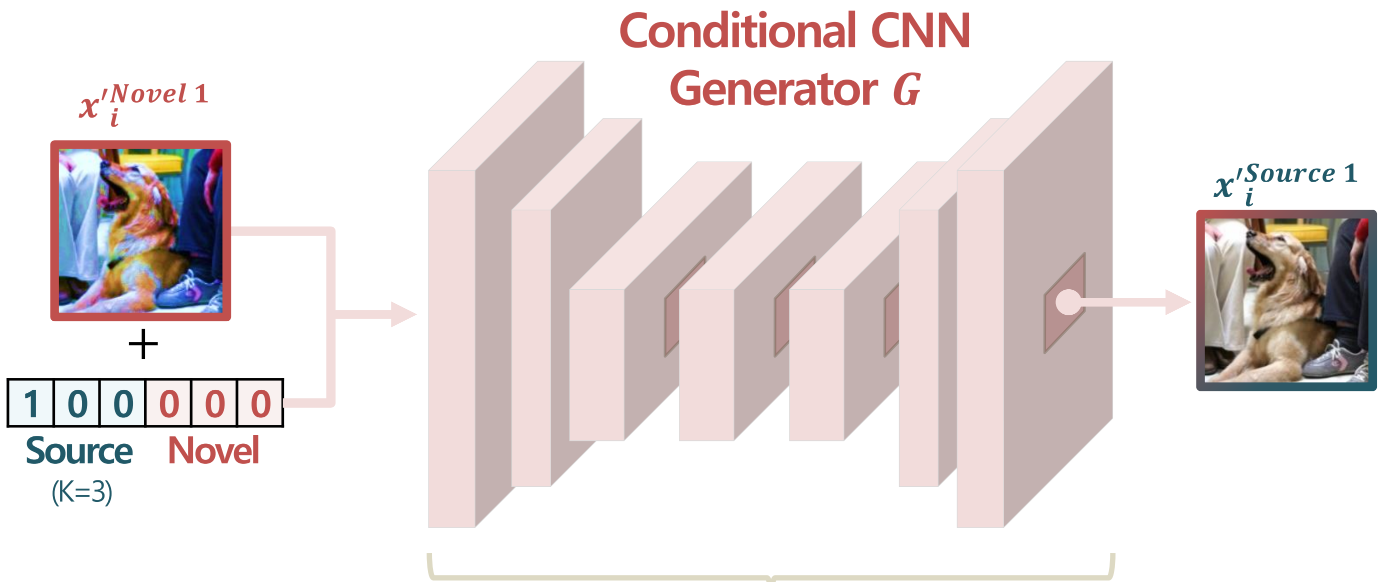
Domain Generalization Methods

2. Data Manipulation: (2) Data Generation



❖ Learning to Generate Novel Domains for Domain Generalization[15]

- **L2A-OT**(Learning to Augment by Optimal Transport) : (1) Conditional Generator G, (2) Classifier $\hat{Y}$
  - **(1)-2. Maintaining Semantic Consistency** → Novel Domain이 원본 이미지의 Semantic Content를 유지하도록 하기 위한 전략



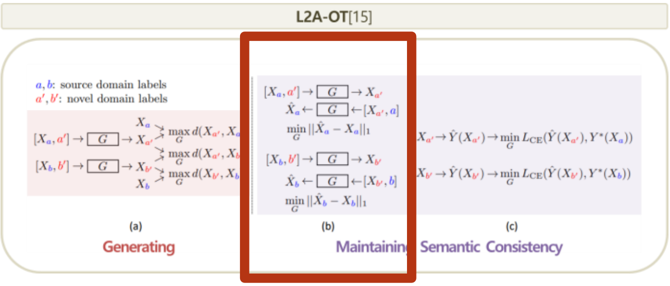
원본 Source Domain과
같은 Domain의 이미지 생성

StarGAN[17] 구조

[15] Zhou, K., Yang, Y., Hospedales, T., & Xiang, T. (2020). Learning to generate novel domains for domain generalization. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16 (pp. 561-578). Springer International Publishing.
[17] Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8789-8797).

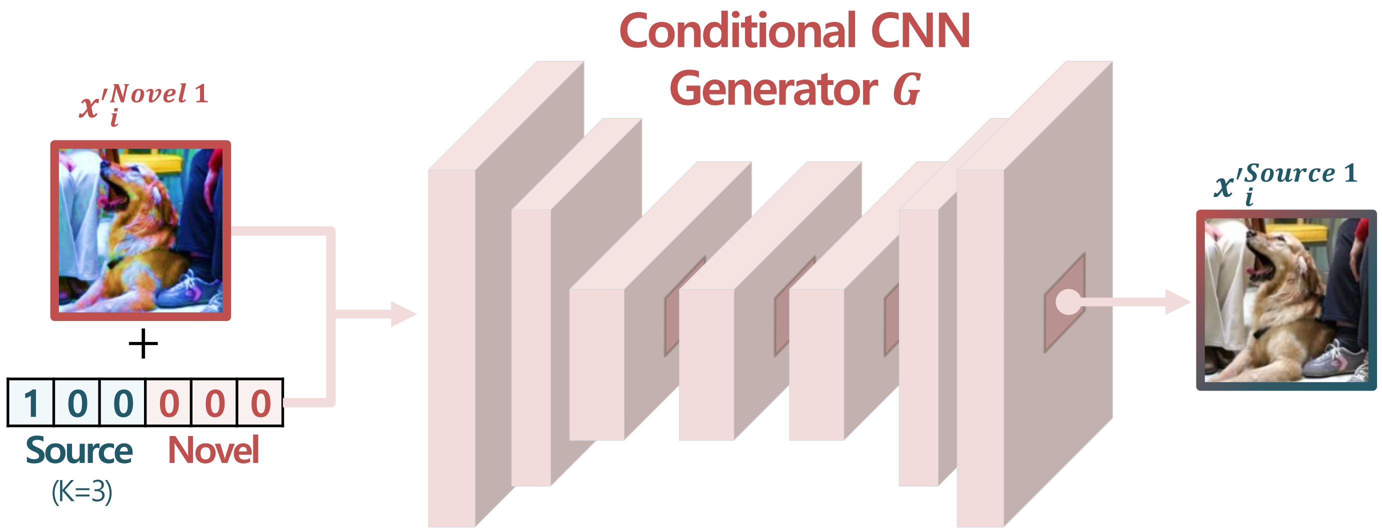
Domain Generalization Methods

2. Data Manipulation: (2) Data Generation



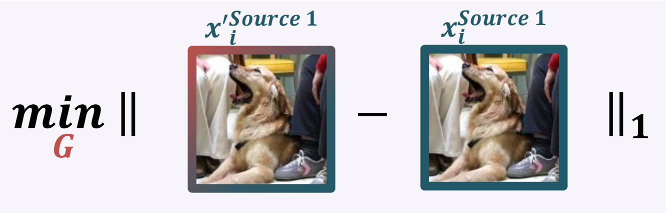
❖ Learning to Generate Novel Domains for Domain Generalization[15]

- **L2A-OT**(Learning to Augment by Optimal Transport) : (1) Conditional Generator G, (2) Classifier $\hat{Y}$
  - **(1)-2. Maintaining Semantic Consistency** → Novel Domain이 원본 이미지의 Semantic Content를 유지하도록 하기 위한 전략



원본 Source Domain과 같은 Domain의 이미지 생성

Cycle-consistency Loss[18]

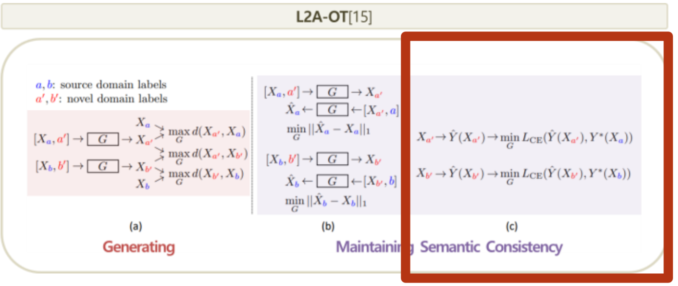


To guarantee structural consistency

[15] Zhou, K., Yang, Y., Hospedales, T., & Xiang, T. (2020). Learning to generate novel domains for domain generalization. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16 (pp. 561-578). Springer International Publishing.
[18] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).

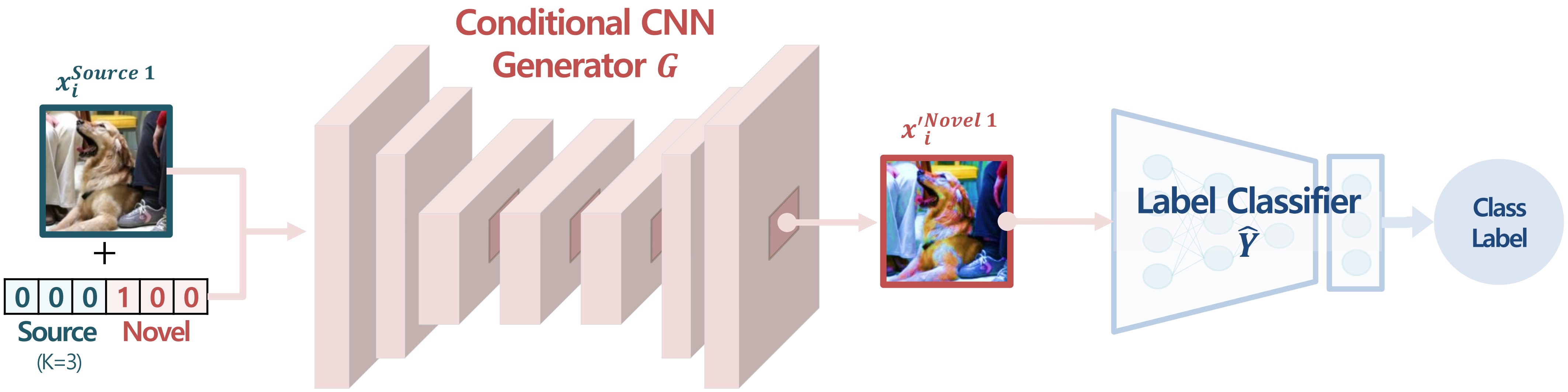
Domain Generalization Methods

2. Data Manipulation: (2) Data Generation



❖ Learning to Generate Novel Domains for Domain Generalization[15]

- **L2A-OT**(Learning to Augment by Optimal Transport) : (1) Conditional Generator G, (2) Classifier $\hat{Y}$
  - **(1)-2. Maintaining Semantic Consistency** → Novel Domain이 원본 이미지의 Semantic Content를 유지하도록 하기 위한 전략



Cross-entropy Loss $\min_{G} \ell($ Class Label , Dog $)$

$x_i^{\prime Novel\ 1}$ Class 예측 값

$x_i^{Source\ 1}$ Class 실제 값

To maintain the category label and thus enforce semantic consistency

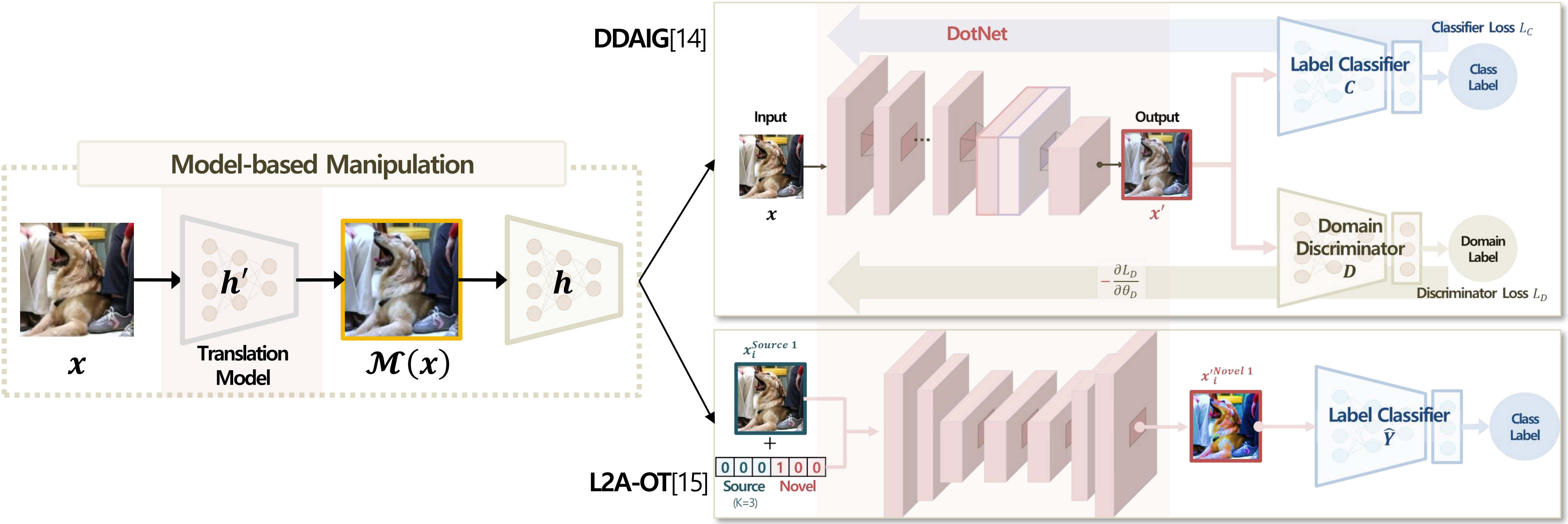
[15] Zhou, K., Yang, Y., Hospedales, T., & Xiang, T. (2020). Learning to generate novel domains for domain generalization. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16 (pp. 561-578). Springer International Publishing.

Domain Generalization Methods

2. Data Manipulation

❖ Data Manipulation based Domain Generalization Methods

- **Model-based Manipulation : DDAIG**[14], **L2A-OT**[15]
  - 학습 시 활용 가능한 Domain의 범위가 확장되고, Label space가 다른 경우에도 강건한 성능 도출 가능
  - 보조 생성 모델에 대한 학습 필요 (Computational Cost ↑)



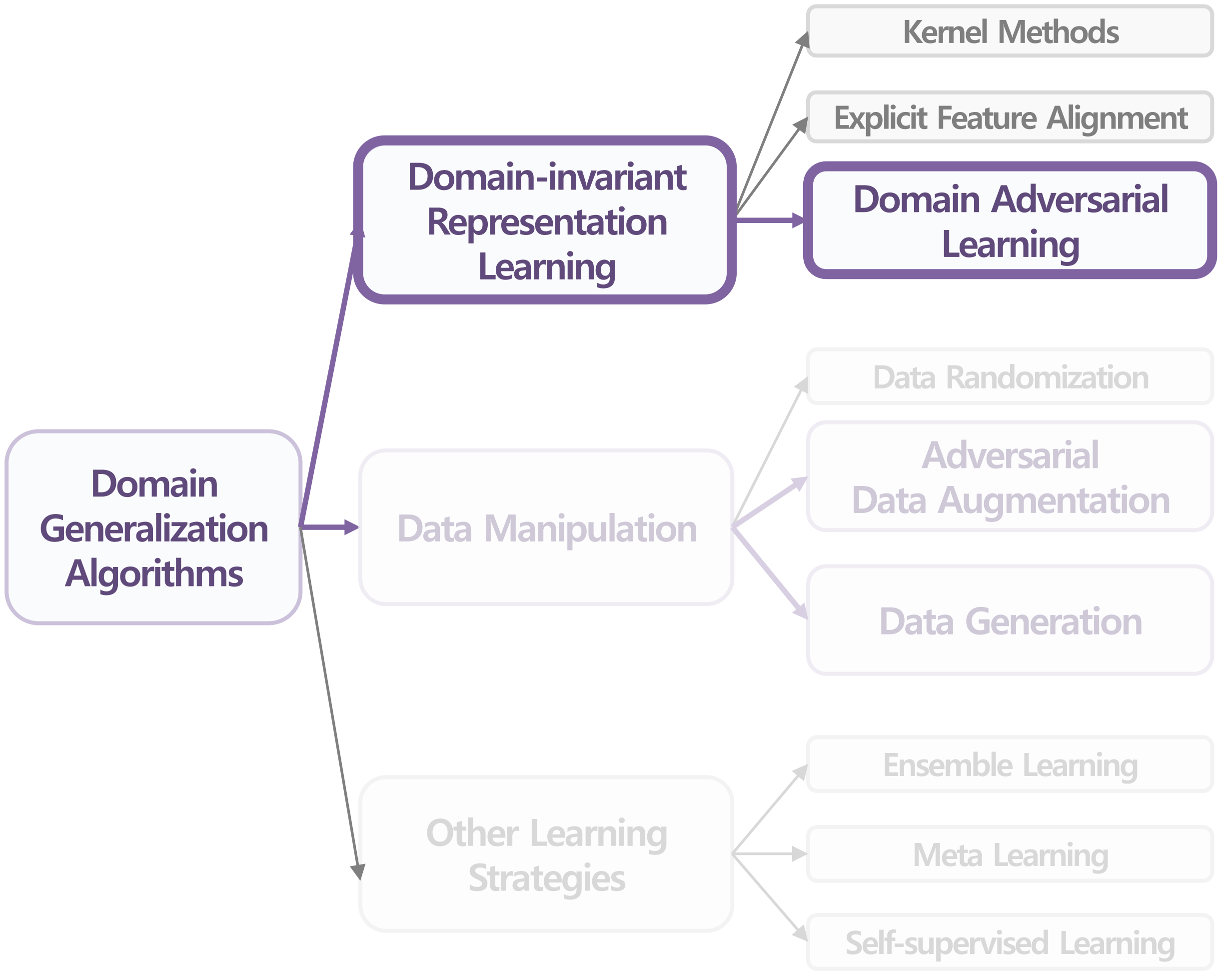
[15] Zhou, K., Yang, Y., Hospedales, T., & Xiang, T. (2020). Learning to generate novel domains for domain generalization. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16 (pp. 561-578). Springer International Publishing.

230721 DMQA Open Seminar:
Domain Generalization : How to improve the generalization ability of deep learning models?

3. Conclusions

Conclusion

Summary



DANN[3]

Domain-Adversarial Training of Neural Networks

Abstract

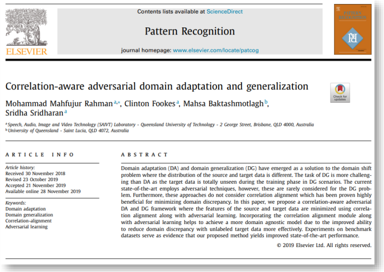
We introduce a new representation learning approach for domain adaptation, in which data at training and test time come from similar but different distributions. Our approach is directly inspired by the theory on domain adaptation suggesting that, for effective domain transfer to be achieved, predictions must be made based on features that cannot discriminate between the training (source) and test (target) domains.

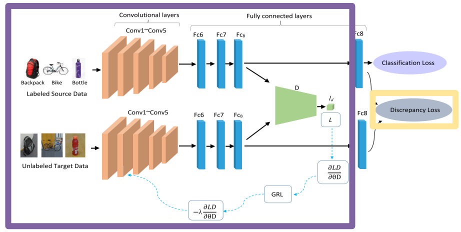
The approach implements this idea in the context of neural network architectures that are trained on labeled data from the source domain and unlabeled data from the target domain (no labeled target-domain data is necessary). As the training progresses, the approach promotes the emergence of features that are (i) discriminative for the main learning task on the source domain and (ii) indiscriminate with respect to the shift between the domains. We show that this adaptation behaviour can be achieved in almost any feed-forward model by augmenting it with few standard layers and a new *gradient reversal* layer. The resulting augmented architecture can be trained using standard backpropagation and stochastic gradient descent, and can thus be implemented with little effort using any of the deep learning packages.

We demonstrate the success of our approach for two distinct classification problems (document sentiment analysis and image classification), where state-of-the-art domain adaptation performance on standard benchmarks is achieved. We also validate the approach for descriptor learning task in the context of person re-identification application.

Keywords: domain adaptation, neural network, representation learning, deep learning, synthetic data, image classification, sentiment analysis, person re-identification

CAADA[5]

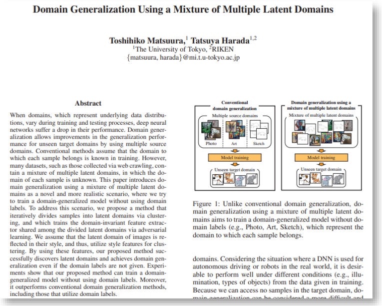


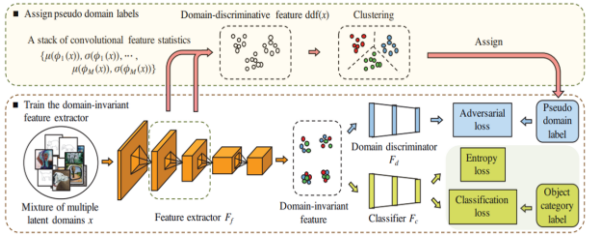


DANN +
Deep CORAL[8]

Explicit Feature Alignment based Method

MMLD[7]



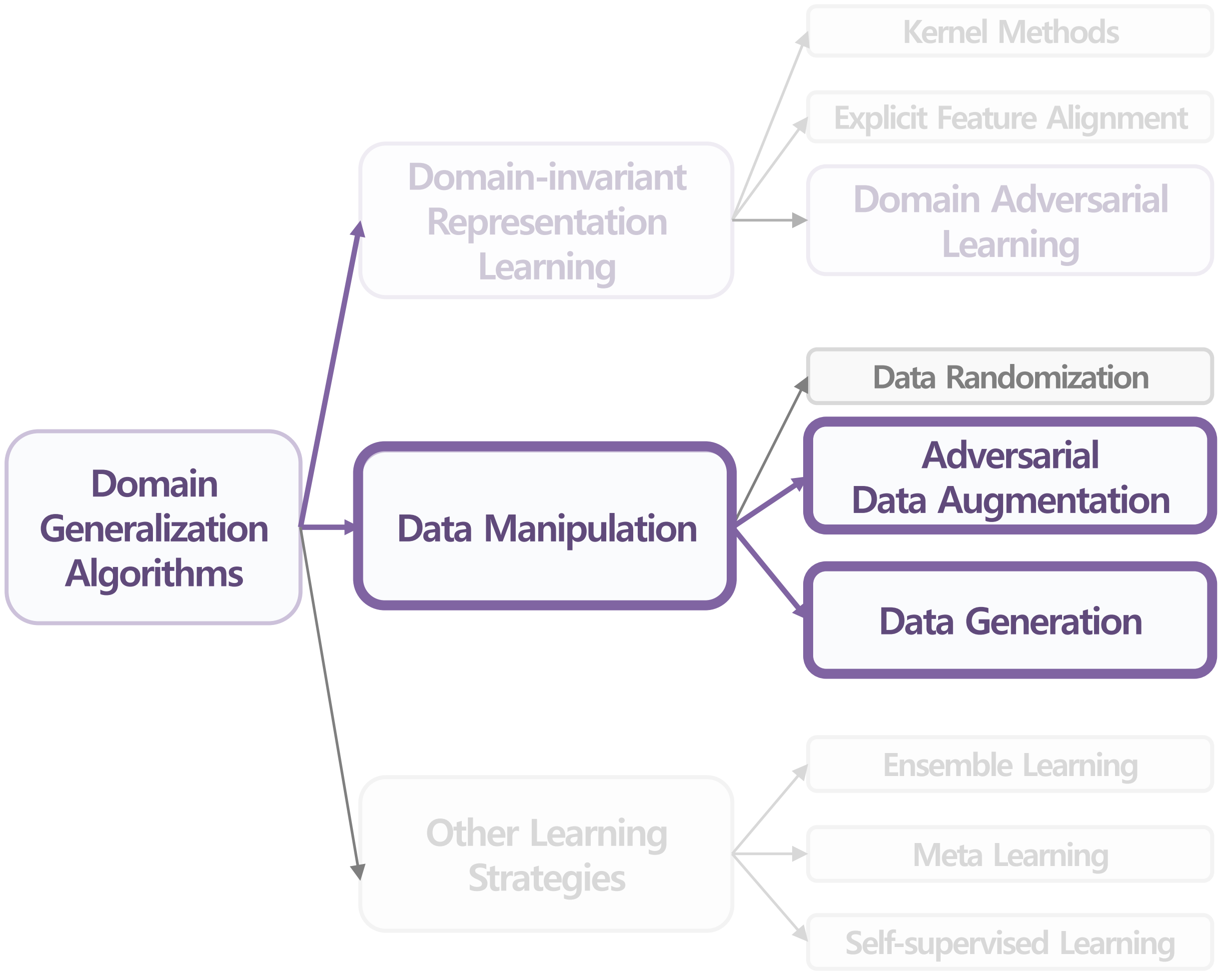


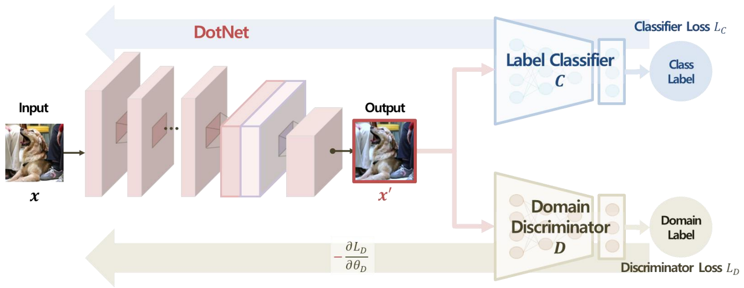
DANN +
Domain Label[10] +
Entropy Minimization[11]

Domain Adaptation Method

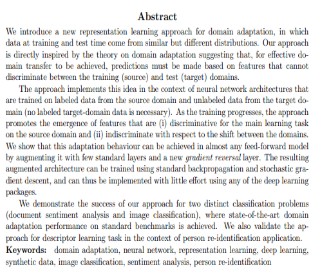
Conclusion

Summary





DANN[3]



Domain-Adversarial Training of Neural Networks

Abstract

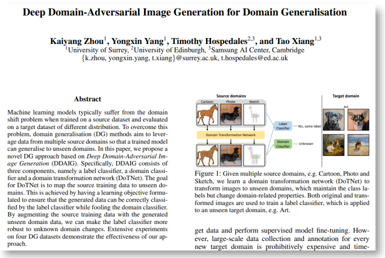
We introduce a new representation learning approach for domain adaptation, in which data at training and test time come from similar but different distributions. Our approach is directly inspired by the theory on domain adaptation suggesting that, for effective domain transfer to be achieved, predictions must be made based on features that cannot discriminate between the training (source) and test (target) domains.

The approach implements this idea in the context of neural network architectures that are trained on labeled data from the source domain and unlabeled data from the target domain (no labeled target-domain data is necessary). As the training progresses, the approach promotes the emergence of features that are (i) discriminative for the main learning task on the source domain and (ii) indiscriminate with respect to the shift between the domains. We show that this adaptation behaviour can be achieved in almost any feed-forward model by augmenting it with few standard layers and a new *gradient reversal* layer. The resulting augmented architecture can be trained using standard backpropagation and stochastic gradient descent, and can thus be implemented with little effort using any of the deep learning packages.

We demonstrate the success of our approach for two distinct classification problems (document sentiment analysis and image classification), where state-of-the-art domain adaptation performance on standard benchmarks is achieved. We also validate the approach for descriptor learning task in the context of person re-identification application.

Keywords: domain adaptation, neural network, representation learning, deep learning, synthetic data, image classification, sentiment analysis, person re-identification

DDAIG[14]



Deep Domain-Adversarial Image Generation for Domain Generalisation

Kaiyang Zhou[1], Yongxin Yang[1], Timothy Hospedales[2,3], and Tao Xiang[1,3]
[1]University of Surrey, [2]University of Edinburgh, [3]Samsung AI Center, Cambridge
{k.zhou, yongxin.yang, t.xiang}@surrey.ac.uk, t.hospedales@ed.ac.uk

Abstract

Machine learning models typically suffer from the domain shift problem when trained on a source dataset and evaluated on a target dataset of different distribution. To overcome this problem, domain generalisation (DG) methods aim to leverage data from multiple source domains so that a trained model can generalise to unseen domains. In this paper, we propose a novel DG approach based on *Deep Domain-Adversarial Image Generation* (DDAIG). Specifically, DDAIG consists of three components, namely a label classifier, a domain classifier and a domain transformation network (DoTNet). The goal for DoTNet is to map the source training data to unseen domains. This is achieved by having a learning objective formulated to ensure that the generated data can be correctly classified by the label classifier while fooling the domain classifier. By augmenting the source training data with the generated unseen domain data, we can make the label classifier more robust to unknown domain changes. Extensive experiments on four DG datasets demonstrate the effectiveness of our approach.

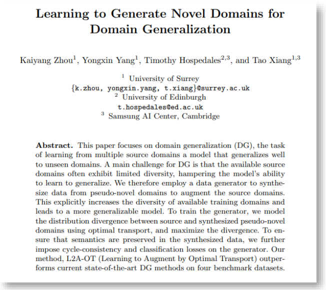
Figure 1: Given multiple source domains, e.g. Cartoon, Photo and Sketch, we learn a domain transformation network (DoTNet) to transform images to unseen domains, which maintain the class labels but change domain-related properties. Both original and transformed images are used to train a label classifier, which is applied to an unseen target domain, e.g. Art.

get data and perform supervised model fine-tuning. However, large-scale data collection and annotation for every new target domain is prohibitively expensive and time-

Model-based Manipulation

Domain Transformation Network + **DANN**

L2A-OT[15]

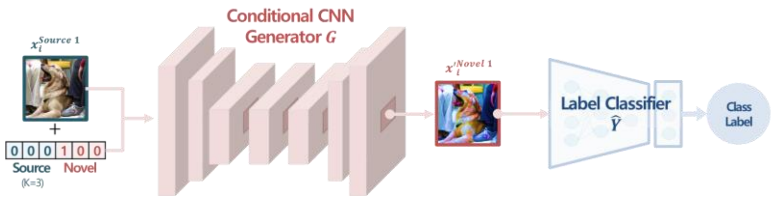


Learning to Generate Novel Domains for Domain Generalization

Kaiyang Zhou[1], Yongxin Yang[1], Timothy Hospedales[2,3], and Tao Xiang[1,3]

[1] University of Surrey
{k.zhou, yongxin.yang, t.xiang}@surrey.ac.uk
[2] University of Edinburgh
t.hospedales@ed.ac.uk
[3] Samsung AI Center, Cambridge

Abstract. This paper focuses on domain generalization (DG), the task of learning from multiple source domains a model that generalizes well to unseen domains. A main challenge for DG is that the available source domains often exhibit limited diversity, hampering the model's ability to learn to generalize. We therefore employ a data generator to synthesize data from pseudo-novel domains to augment the source domains. This explicitly increases the diversity of available training domains and leads to a more generalizable model. To train the generator, we model the distribution divergence between source and synthesized pseudo-novel domains using optimal transport, and maximize the divergence. To ensure that semantics are preserved in the synthesized data, we further impose cycle-consistency and classification losses on the generator. Our method, L2A-OT (Learning to Augment by Optimal Transport) outperforms current state-of-the-art DG methods on four benchmark datasets.



StarGAN[17] 구조 +
CycleGAN [18] consistency loss +
WGAN [16] distance metric

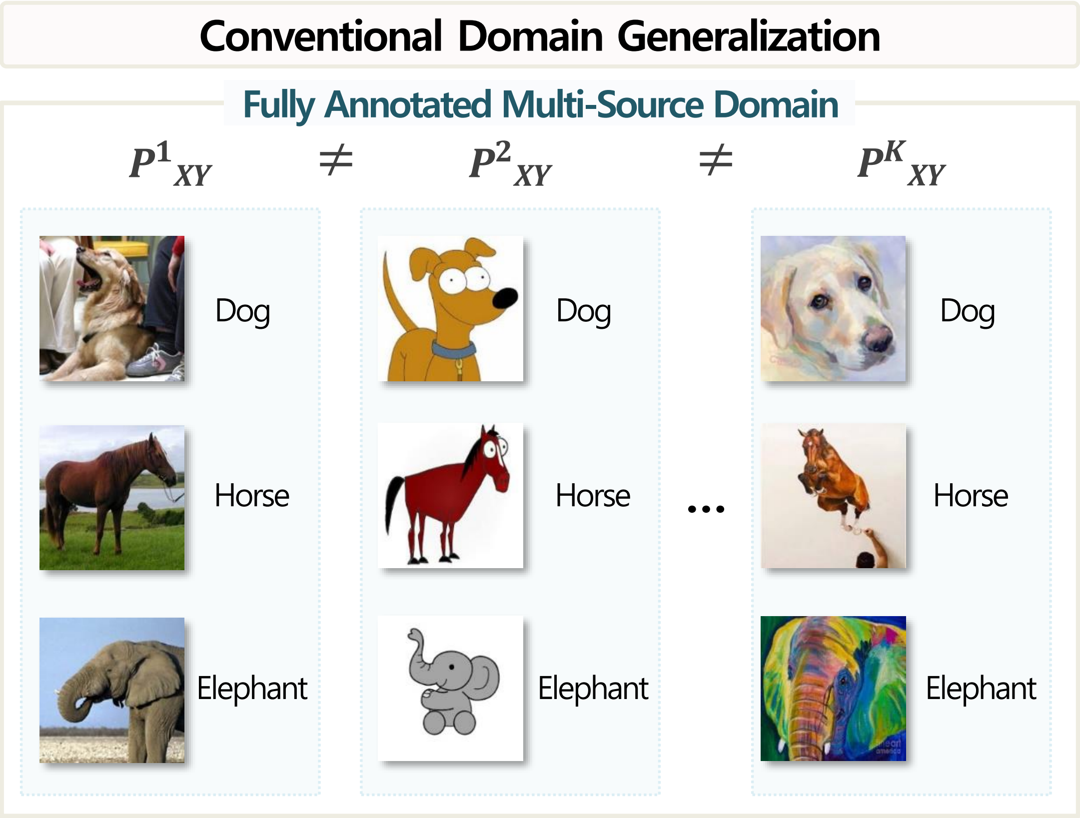
Model-based Manipulation

Conclusion

Other Domain Generalization Research Areas

❖ Future Research Directions

- **What are worth exploring to further this field? :** New scenarios to push the frontiers of Domain Generalization



Conclusion

Other Domain Generalization Research Areas

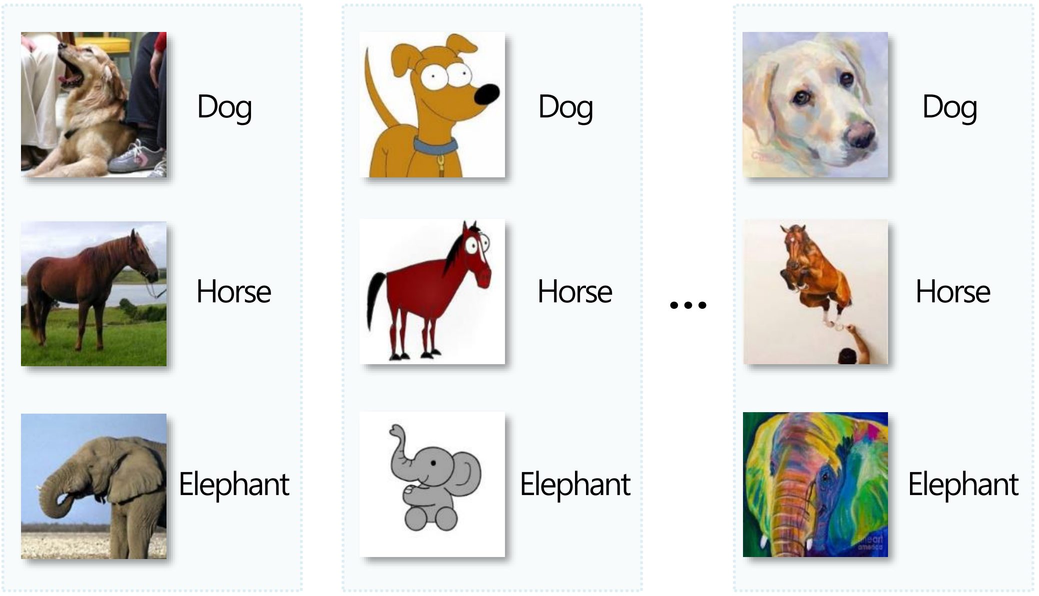
❖ Future Research Directions

- **Single-source Domain Generalization** : 'Multi-Source'를 가정할 수 없을 때

Conventional Domain Generalization → **Single-source Domain Generalization**

Fully Annotated Multi-Source Domain

$P^1_{XY} \neq P^2_{XY} \neq P^K_{XY}$



Fully Annotated Single-Source Domain

P^1_{XY}

활용 가능한 Source Domain이 하나밖에 없는 경우

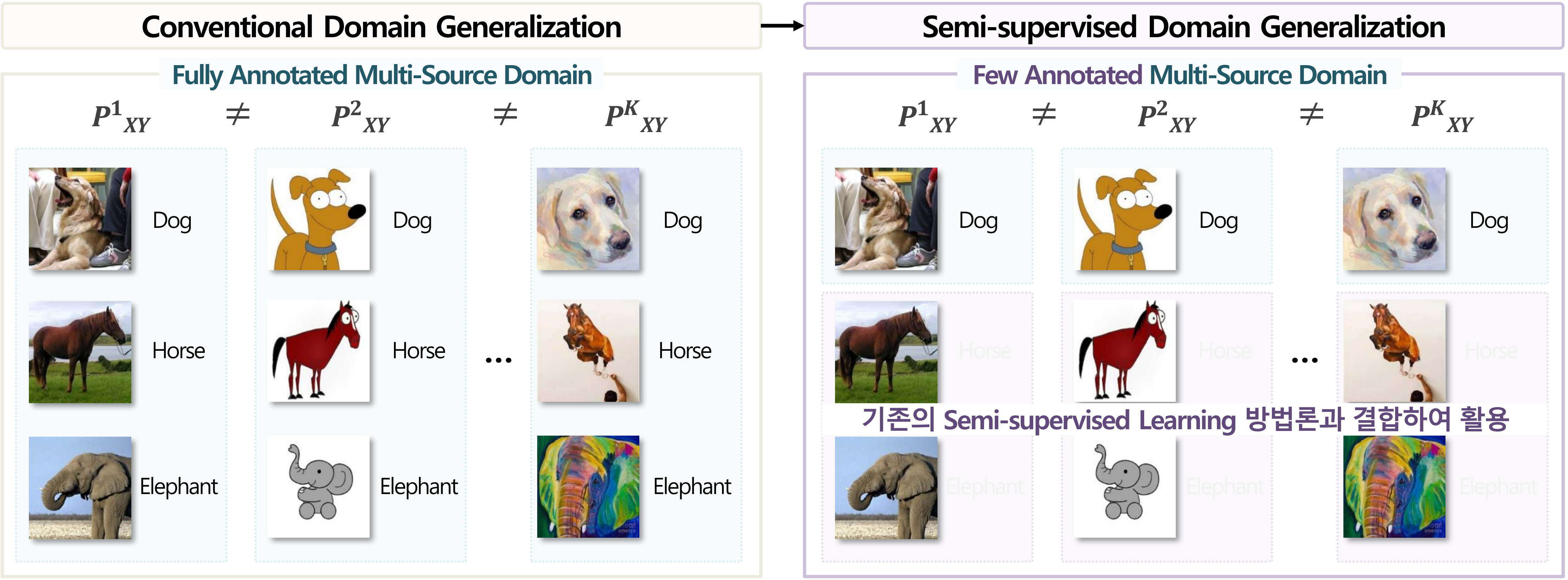
→ Data Generation based Methods를 주로 활용

Conclusion

Other Domain Generalization Research Areas

❖ Future Research Directions

- **Semi-supervised Domain Generalization** : 'Fully Annotated'를 가정할 수 없을 때

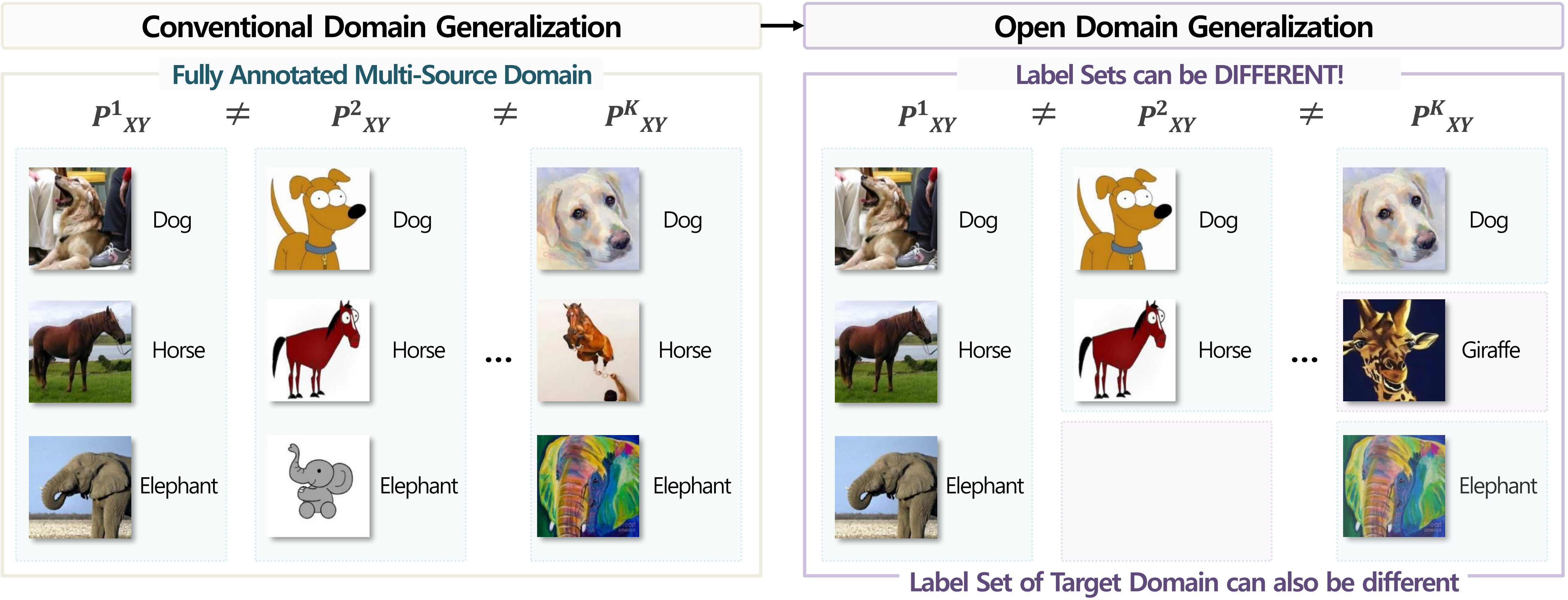


Conclusion

Other Domain Generalization Research Areas

❖ Future Research Directions

- **Open Domain Generalization** : 'Same Label Set'을 가정할 수 없을 때 → Target Domain에도 Unknown Classes (Out-of-distribution; OOD)가 있는 경우



Thank You

References

[1] Zhou, K., Yang, Y., Qiao, Y., & Xiang, T. (2021). Domain generalization with mixstyle. arXiv preprint arXiv:2104.02008.

[2] Wang, J., Lan, C., Liu, C., Ouyang, Y., Qin, T., Lu, W., ... & Yu, P. (2022). Generalizing to unseen domains: A survey on domain generalization. IEEE Transactions on Knowledge and Data Engineering.

[3] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., ... & Lempitsky, V. (2016). Domain-adversarial training of neural networks. The journal of machine learning research, 17(1), 2096-2030.

[4] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.

[5] Rahman, M. M., Fookes, C., Baktashmotlagh, M., & Sridharan, S. (2020). Correlation-aware adversarial domain adaptation and generalization. Pattern Recognition, 100, 107124..

[6] Saito, K., Kim, D., Sclaroff, S., Darrell, T., & Saenko, K. (2019). Semi-supervised domain adaptation via minimax entropy. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 8050-8058).

[7] Matsuura, T., & Harada, T. (2020, April). Domain generalization using a mixture of multiple latent domains. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 07, pp. 11749-11756).

[8] Sun, B., & Saenko, K. (2016). Deep coral: Correlation alignment for deep domain adaptation. In Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14 (pp. 443-450). Springer International Publishing.

[9] Sun, B., Feng, J., & Saenko, K. (2016, March). Return of frustratingly easy domain adaptation. In Proceedings of the AAAI conference on artificial intelligence (Vol. 30, No. 1).

[10] Li, Y., Wang, N., Liu, J., & Hou, X. (2017). Demystifying neural style transfer. arXiv preprint arXiv:1701.01036.

[11] Long, M., Zhu, H., Wang, J., & Jordan, M. I. (2016). Unsupervised domain adaptation with residual transfer networks. Advances in neural information processing systems, 29.

[12] Zhou, K., Liu, Z., Qiao, Y., Xiang, T., & Loy, C. C. (2022). Domain generalization: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[13] Shankar, S., Piratla, V., Chakrabarti, S., Chaudhuri, S., Jyothi, P., & Sarawagi, S. (2018). Generalizing across domains via cross-gradient training. arXiv preprint arXiv:1804.10745.

[14] Zhou, K., Yang, Y., Hospedales, T., & Xiang, T. (2020, April). Deep domain-adversarial image generation for domain generalisation. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 07, pp. 13025-13032).

[15] Zhou, K., Yang, Y., Hospedales, T., & Xiang, T. (2020). Learning to generate novel domains for domain generalization. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16 (pp. 561-578). Springer International Publishing.

[16] Arjovsky, M., Chintala, S., & Bottou, L. (2017, July). Wasserstein generative adversarial networks. In International conference on machine learning (pp. 214-223). PMLR.

[17] Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8789-8797).

[18] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).