

# Self-ensembling for visual domain adaptation

System Intelligence Lab

연세대학교 산업공학과  
System Intelligence Lab  
김민식 통합과정

# Table of Contents

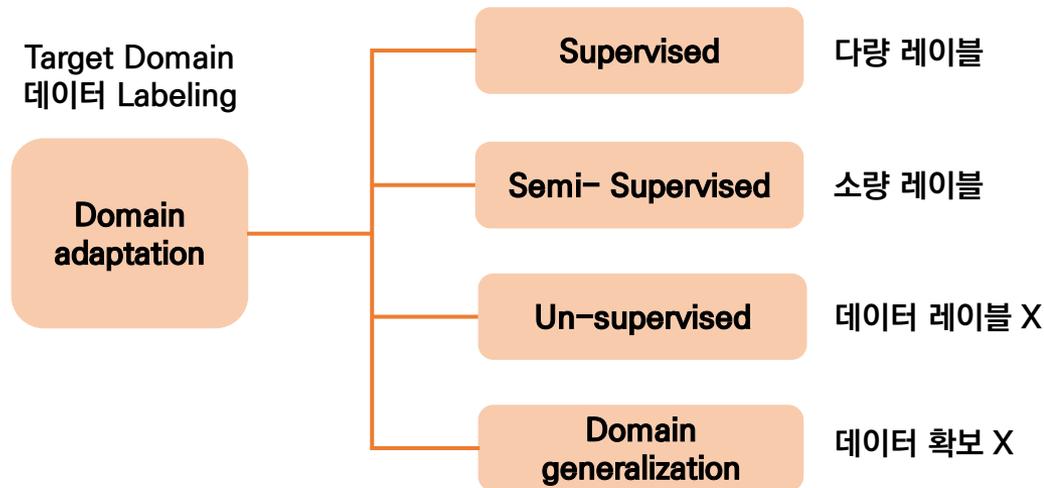
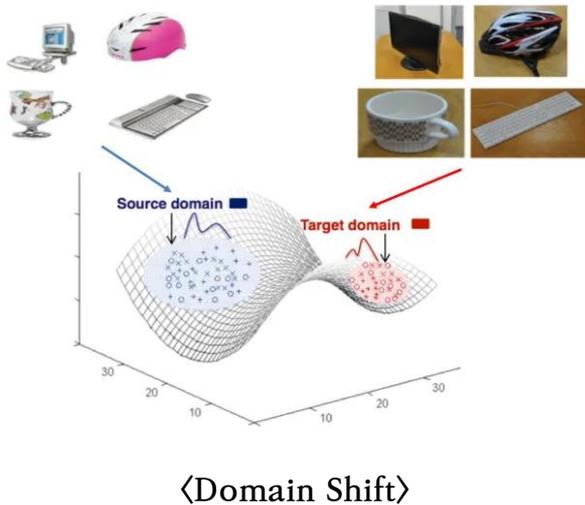
---

- **Background**
- **Introduction**
- **Proposed Method**
- **Experiment**
- **Conclusion**

# 1. Background

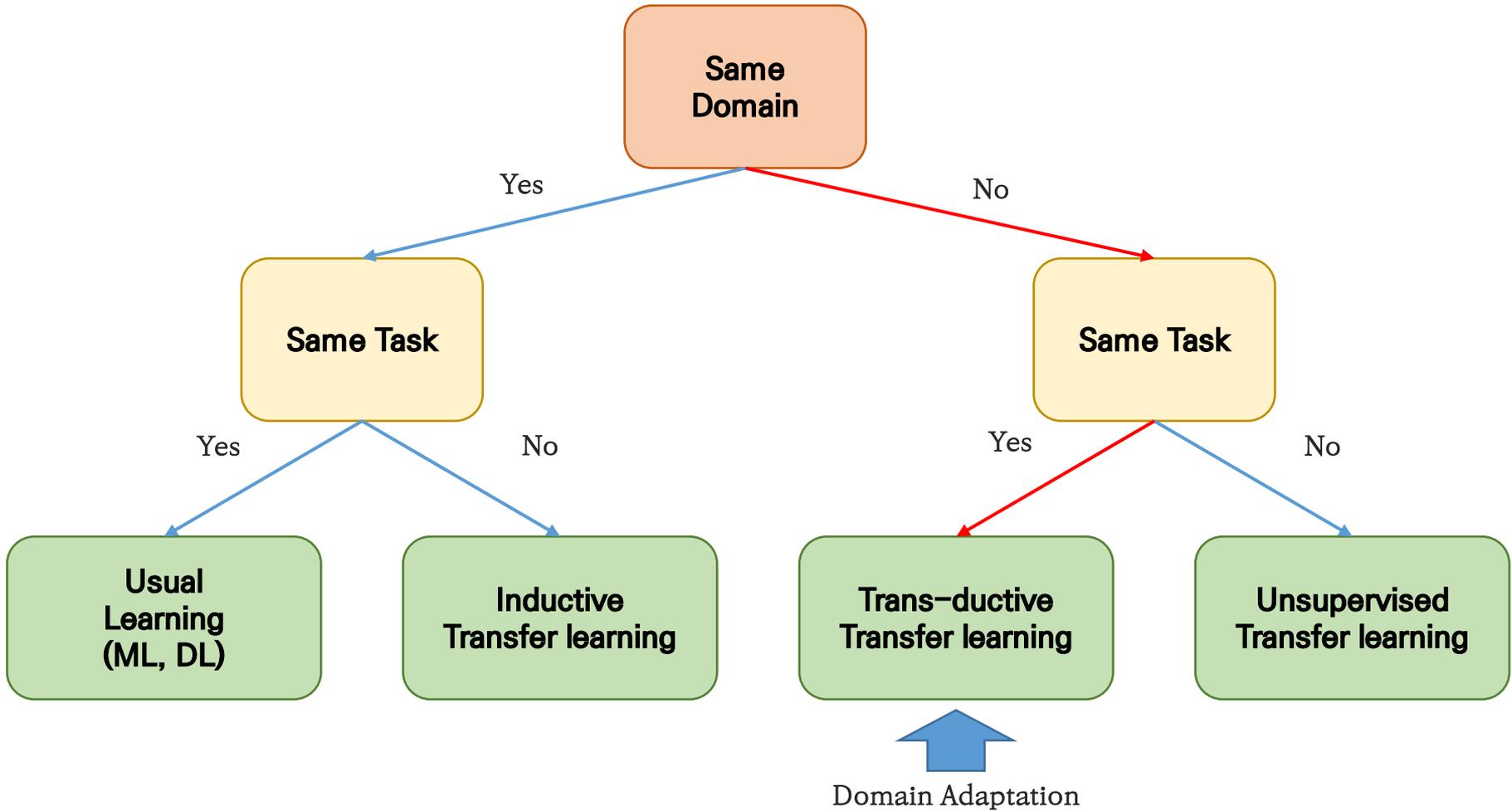
## ➤ Domain adaptation

- Machine learning과 Deep learning 학습 과정에서 중요 사안
  - 학습 데이터(Source)와 테스트 데이터(Target) 사이의 **분포(Distribution) 차이**가 발생
  - 양질의 데이터 수집 하기 어려워, 특정 도메인 데이터의 레이블이 부족함
- 도메인은 다르지만, 관련이 있는 새로운 도메인에 기존의 정보를 적용 시키는 것



# 1. Background

## ➤ Domain adaptation



# 1. Background

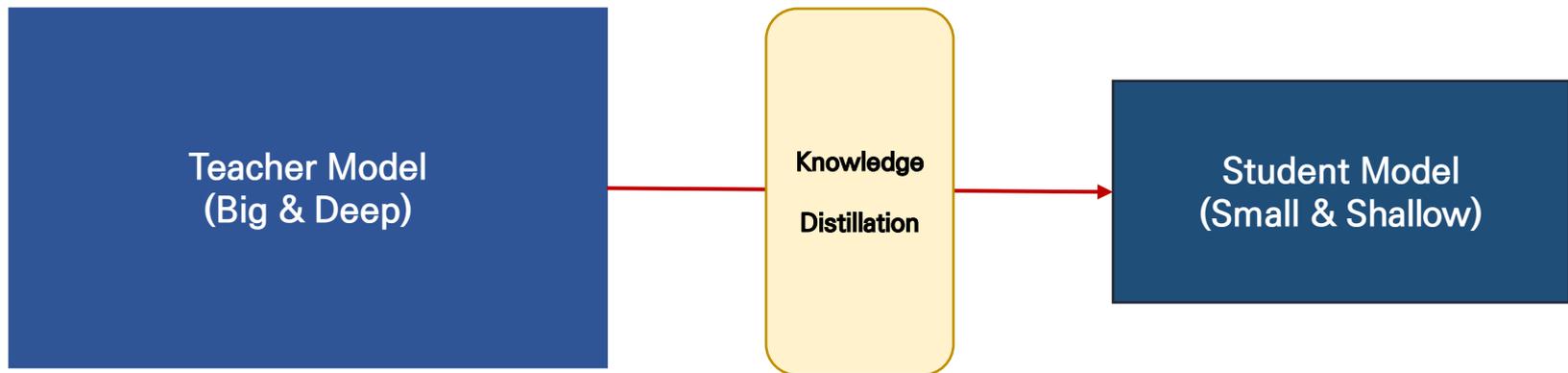
## ➤ Self-ensembling

- 신경망 모델 자체에서 데이터를 여러 번 예측한 이후 앙상블(Ensemble)하는 방식을 의미
  - 데이터 증강(Data Augmentation)을 통해 새로운 훈련 데이터 생성
  - 훈련 과정에서 Dropout, Noise condition 등을 사용해 모델이 다른 조건을 가짐
- 모델의 일반화 성능을 향상, 과적합(Over-fitting)를 줄일 수 있음
- 훈련 데이터의 레이블이 부족한 상황(Domain Adaptation 등)의 효과적

# 1. Background

## ➤ Knowledge Distillation: Teacher–Student learning

- 미리 잘 학습된 큰 네트워크(Teacher Network)의 출력(지식)을 실제로 사용하고자 하는 작은 네트워크(Student Network)에게 모방하여 학습(전달하는 것)
- 상대적으로 적은 파라미터를 가지고 모델의 높은 성능을 낼 수 있음(알고리즘 경량화)



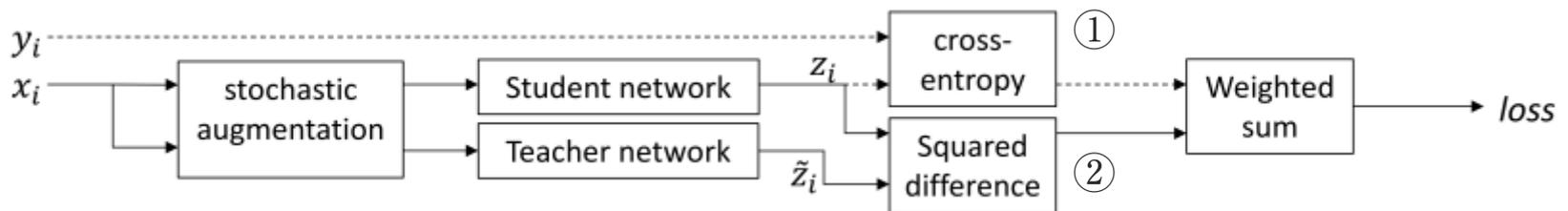
- Teacher: 높은 정확도를 가진 모델
- Teacher 모델의 **어떠한 지식(what)**을

- Student: 지식을 받는 단순한 모델
- Student 모델의 **어떻게 전달(How)**을

# 1. Background

## ➤ Mean-teacher

- 지도 학습(레이블 있는 샘플)와 비지도 학습(레이블 없는 샘플) 모두 활용하는 준(Semi) 지도 학습을 진행
  - Student Network: 경사 하강(Gradient descent) / Teacher Network: 지수 평활(EWA)
- ① 지도 학습 방법(Supervising factor)
  - 레이블이 있는 샘플의 경우, 예측된 클래스의 확률과 정답 클래스의 Cross entropy Loss 계산
  - 레이블이 없는 샘플의 경우, 손실은 0으로 masking
- ② 비지도 학습 방법
  - 동일한 Sample에 대해, Student & Teacher의 예측 클래스의 확률 차이에 대한 패널티(Loss)



〈Network Structures〉

# 2. Introduction

## ➤ Unsupervised domain adaptation

- Computer vision에서 Deep learning은 좋은 성능을 보이지만 **레이블이 있는 큰 데이터셋 필요**
- 데이터 중 일부만 레이블만 가지고 학습(Semi-supervised learning)하는 등 다양한 연구가 진행 중
- 최근에는 비지도 학습(Un-supervised learning)의 한 분야로, domain adaptation 적용
  - 레이블이 있는 Source 데이터에서 얻은 지식을 Target 데이터에 전달(Knowledge Distillation)
  - 합성 데이터와 실제 데이터를 사용해 모델을 훈련시킬 수 있는 잠재력 제공

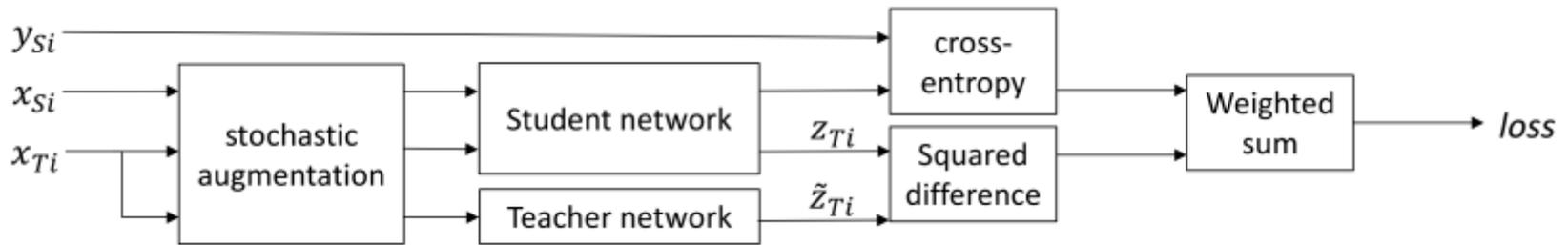
## ➤ Contribution

- 본 논문은 Domain adaptation 문제에 Self-ensembling과 Teacher-Network 적용한 방법론 제시
  - VISDA-2017 visual domain adaptation challenge 우승
  - (이미지의 크기가 작은) 여러 데이터셋에서 지도 학습 방식과 거의 유사한 정확도를 달성

# 3. Proposed method

## ➤ Adapting to domain adaptation

- 공통점: Mean-teacher와 동일한 Loss 구성
- 차이점
  - (Semi-supervised) 기존에는 하나의 데이터 분포(동일한 기저 분포)를 가진 데이터셋
  - (Unsupervised-domain adaptation) 현재는 서로 다른 기저 분포를 가진 두 개 데이터셋
    - 레이블이 지정된 Source 데이터셋과 레이블이 지정되지 않는 Target 데이터셋
    - Source & Target 데이터셋에 대해서 **별도로 Batch Normalization** 동시 처리



〈Network Structures(Adapting domain adaptation)〉

# 3. Proposed method

## ➤ 추가적인 수정사항

### ① Confidence thresholding

- Target 데이터셋에 대한 Teacher Net 예측 확률이 임계 값 기준으로 필터링 작업
- Confidence 값이 임계 값(0.968) 아래일 경우 해당 샘플을 학습하지 않음(Loss 0으로 Mask)

### ② Data Augmentation

- 이미지 크기가 작은 데이터셋에 대한 실험으로 3가지 데이터 증강 기법을 사용
- 1) Gaussian Noise(0.1), 2) 수평 전환 및 [-2, 2] 정규화, 3) random affine 변환

### ③ Class Balance Loss

- 일부 데이터셋(MNIST → SVNH)의 경우 클래스 불균형적으로 예측을 진행
- 클래스 불균형적으로 예측을 할 때 패널티를 주는 Loss 값 추가
- Target의 각 클래스별로 예측된 평균 확률과 uniform probability 사이의 Binary cross entropy

# 4. Experiment

## ➤ 데이터셋과 실험

- 데이터셋 별로 Domain Adaptation하기 위한 추가 전처리 진행
  - 해상도, 입력 이미지 포맷(채널), 분류 Target 등 Matching
- 성능 지표는 정확도를 사용하며, 제안 방법론에 대한 구성요소별 실험 진행
  - MT(Mean teacher), CT(confidence thresholding), TF(데이터 증강), TFA(데이터 증강 추가)
  - 제안 방법 이외의 Classification에서 baseline(train on source)과 최고 성능(train on target)

(a) MNIST ↔ USPS



(b) CIFAR-10 ↔ STL



(c) Syn-digits → SVHN



(d) Syn-signs → GTSRB



(e) SVHN → MNIST



(f) MNIST (specific augmentation) → SVHN



〈Small image example Images〉

Dataset	Target	Classes	Resolution	Channel
USPS	Digits	10	16 * 16	Mono
MNIST	Digits	10	28 * 28	Mono
SVHN	Digits	10	32 * 32	RGB
Syn-Digits	Digits	10	32 * 32	RGB
Syn-Signs	Traffic signs	43	40 * 40	RGB
GTSRB	Traffic signs	43	Varies	RGB
CIFAR-10	Object ID	10	32 * 32	RGB
STL	Object ID	10	96 * 96	RGB

〈Dataset summary〉

# 4. Experiment

## ➤ 실험 방향 및 결과

	USPS	MNIST	SVHN	MNIST	CIFAR	STL	Syn Digits	Syn Signs
	-	-	-	-	-	-	-	-
	MNIST	USPS	MNIST	SVHN	STL	CIFAR	SVHN	GTSRB
TRAIN ON SOURCE								
SupSrc*	77.55	82.03	66.5	25.44	72.84	51.88	86.86	96.95
	±0.8	±1.16	±1.93	±2.8	±0.61	±1.44	±0.86	±0.36
SupSrc+TF	77.53	95.39	68.65	24.86	75.2	59.06	87.45	97.3
	±4.63	±0.93	±1.5	±3.29	±0.28	±1.02	±0.65	±0.16
SupSrc+TFA	91.97	96.25	71.73	28.69	75.18	59.38	87.16	98.02
	±2.15	±0.54	±5.73	±1.59	±0.76	±0.58	±0.85	±0.20
Specific aug. <sup>b</sup>	-	-	-	61.99	-	-	-	-
				±3.9				
RevGrad <sup>a</sup> [1]	74.01	91.11	73.91	35.67	66.12	56.91	91.09	88.65
DCRN [2]	73.67	91.8	81.97	40.05	66.37	58.65	-	-
G2A [3]	90.8	92.5	84.70	36.4	-	-	-	-
ADDA [4]	90.1	89.4	76.00	-	-	-	-	-
ATT [5]	-	-	86.20	52.8	-	-	93.1	96.2
SBADA-GAN [6]	97.60	95.04	76.14	61.08	-	-	-	-
ADA [7]	-	-	97.6	-	-	-	91.86	97.66
OUR RESULTS								
MT+TF	98.07	<b>98.26</b>	99.18	13.96 <sup>c</sup>	80.08	18.3	15.94	98.63
	±2.82	±0.11	±0.12	±4.41	±0.25	±9.03	±0.0	±0.09
MT+CT*	92.35	88.14	93.33	33.87 <sup>c</sup>	77.53	71.65	96.01	98.53
	±8.61	±0.34	±5.88	±4.02	±0.11	±0.67	±0.08	±0.15
MT+CT+TF	97.28	98.13	98.64	34.15 <sup>c</sup>	79.73	<b>74.24</b>	96.51	98.66
	±2.74	±0.17	±0.42	±3.56	±0.45	±0.46	±0.08	±0.12
MT+CT+TFA	<b>99.54</b>	98.23	<b>99.26</b>	37.49 <sup>c</sup>	<b>80.09</b>	69.86	<b>97.11</b>	<b>99.37</b>
	±0.04	±0.13	±0.05	±2.44	±0.31	±1.97	±0.04	±0.09
Specific aug. <sup>b</sup>	-	-	-	<b>97.0<sup>c</sup></b>	-	-	-	-
				±0.06				
TRAIN ON TARGET								
SupTgt*	99.53	97.29	99.59	95.7	67.75	88.86	95.62	98.49
	±0.02	±0.2	±0.08	±0.13	±2.23	±0.38	±0.2	±0.32
SupTgt+TF	99.62	97.65	99.61	96.19	70.98	89.83	96.18	98.64
	±0.04	±0.17	±0.04	±0.1	±0.79	±0.39	±0.09	±0.09
SupTgt+TFA	99.62	97.83	99.59	96.65	70.03	90.44	96.59	99.22
	±0.03	±0.17	±0.06	±0.11	±1.13	±0.38	±0.09	±0.22
Specific aug. <sup>b</sup>	-	-	-	97.16	-	-	-	-
				±0.05				

〈Small image benchmark classification result〉

# 5. Conclusion

## ➤ Conclusion

- 지도학습과 거의 유사한 정확도를 달성한 Visual Domain Adaptation 알고리즘을 제안
  - 다양한 네트워크에서 사용할 수 있을 정도로 유연한 알고리즘
- Source와 Target 데이터셋의 집합 분포를 맞추기 위해 데이터 증강이 중요함
  - 효과적인 Domain adaptation을 하기 위해서 Self-ensembling 활용하는 것이 적합
  - 데이터 증강으로 데이터셋 간의 집합 분포 맞추기 어려운 경우, 사전 훈련된 네트워크 활용

**Thank you.**