

2022년 8월 12일 DMQA 연구실 오픈 세미나

Deep Semi-Supervised Learning with Out-of-distribution Unlabeled Data

고려대학교 산업경영공학부 배진수



DMQA

발표자

❖ 배진수

- 건국대학교 수학과 학부 졸업
- 고려대학교 산업경영공학과 대학원 재학
- 고려대학교 DMQA 연구실 (지도교수:김성범)
- 박사과정 1년차
- wlstn215@korea.ac.kr



❖ 연구분야

- Safe Semi-Supervised Learning Using a Bayesian Neural Network
- Machine Learning Applications for COVID-19 Analysis
- Uncertainty-Aware Soft Sensor Using Bayesian Recurrent Neural Networks
- Deep Learning Model for Obscured Enemy Information Prediction in StarCraft 2

세미나 내용

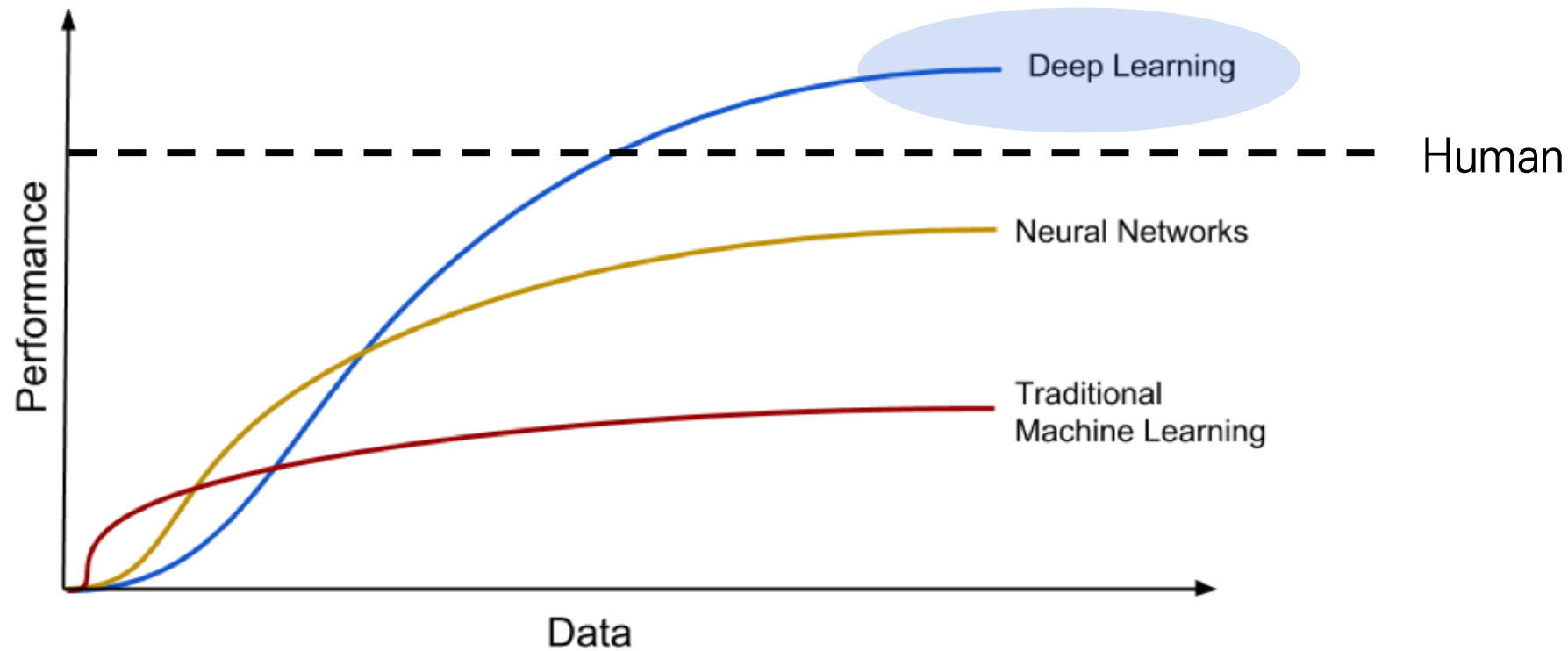
❖ Deep Semi-Supervised Learning with Out-of-distribution Unlabeled Data

1. Deep Semi-Supervised Learning (SSL)
2. Out-of-distribution (O.O.D) Unlabeled Data
3. Safe SSL Against O.O.D Unlabeled Data
4. Conclusions

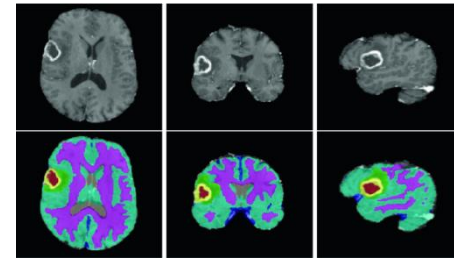
1. Deep Semi-Supervised Learning (SSL)

Deep Semi-Supervised Learning

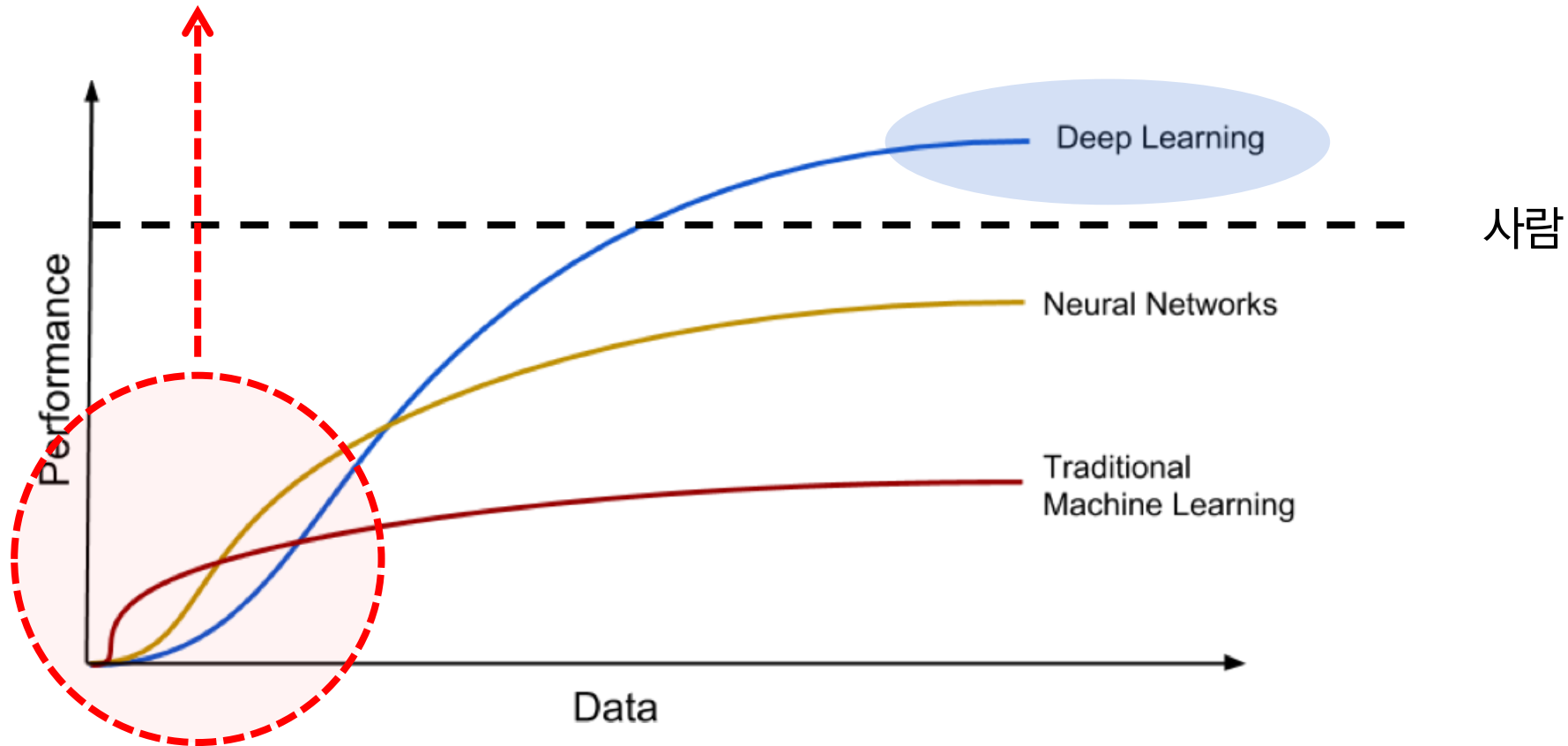
- ❖ 대규모의 데이터셋을 학습한 인공지능 모델이 사람에 버금가는 능력을 갖추기 시작하며 큰 인기를 끌게 됨 [1]



Deep Semi-Supervised Learning



❖ 학습 데이터가 적을 때에는 좋은 성능을 발휘하기 어려움



Deep Semi-Supervised Learning

- ❖ Labeled Data는 사람이 데이터에 대한 정보를 표기해둔 데이터로, 인건비가 비쌀 때 풍부한 양의 데이터 수집이 어려움
- ❖ Unlabeled Data는 사람의 검수를 받지 않고 수집된 데이터들을 의미하여, 상대적으로 많은 양의 데이터 수집이 가능함

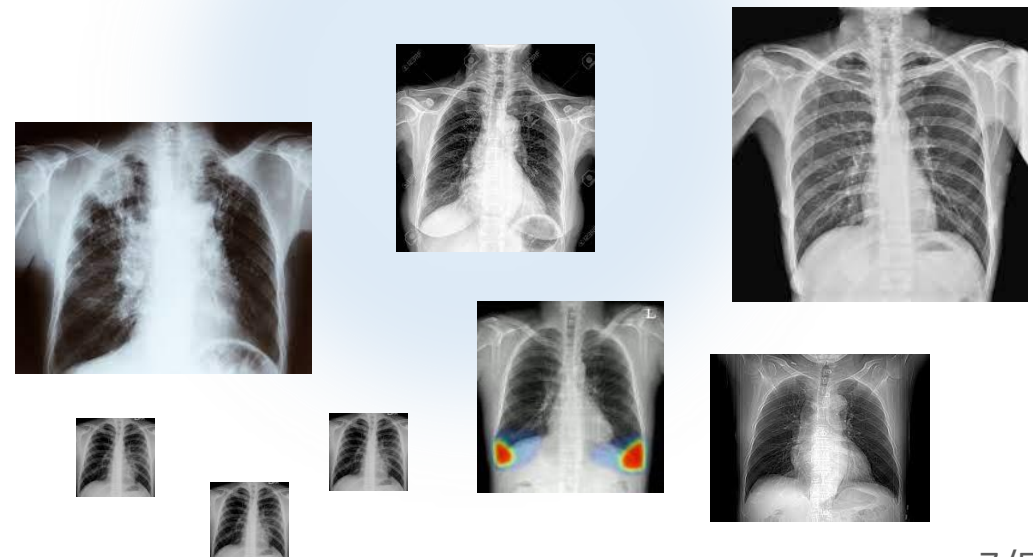
레이블링된 데이터

수집 비용 높음



레이블이 부여되지 않은 데이터

수집 비용 낮음



Deep Semi-Supervised Learning

- ❖ Labeled Data로 사람이 표기해 데이터 정보와 데이터를 함께 인공지능 모델로 학습 → Supervised Learning
- ❖ Unlabeled Data로 데이터의 특성만을 인공지능 모델로 학습 → Unsupervised Learning

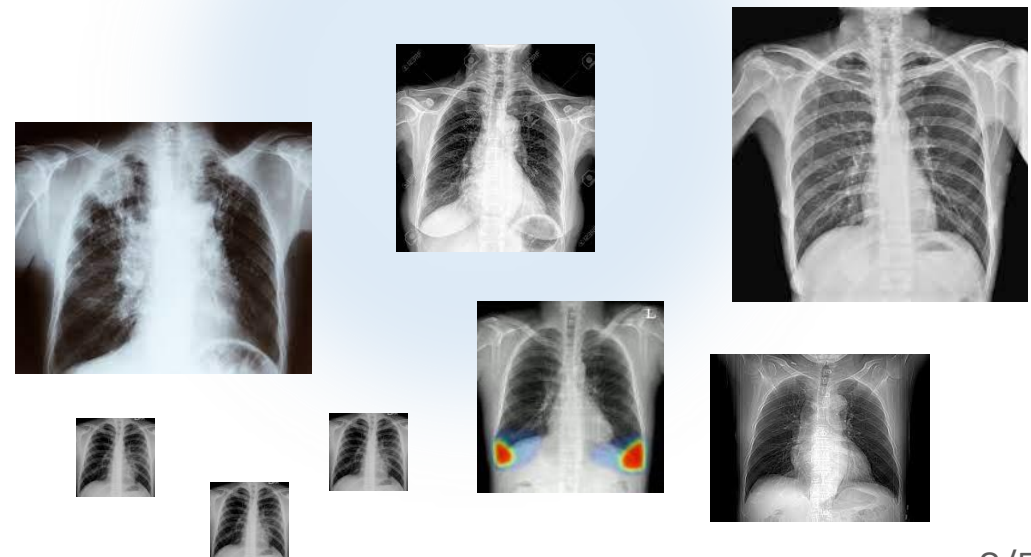
Labeled Data

Supervised Learning



Unlabeled Data

Unsupervised Learning



Deep Semi-Supervised Learning

- ❖ 상대적으로 수집 비용이 저렴한 데이터셋을 적극 활용해 성능을 높이는 방안 존재 → Semi-Supervised Learning [2]

Semi-Supervised Learning (SSL)

Self-Supervised Learning

Transfer Learning



Labeled Data

Unlabeled Data

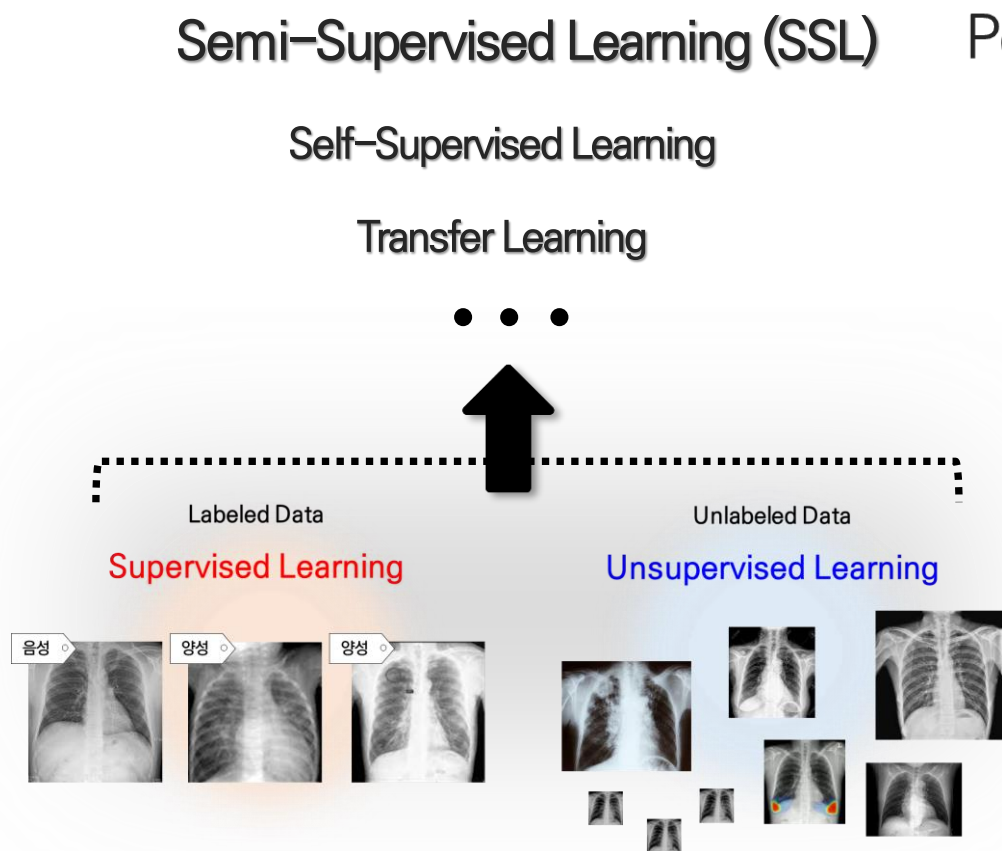
Supervised Learning

Unsupervised Learning

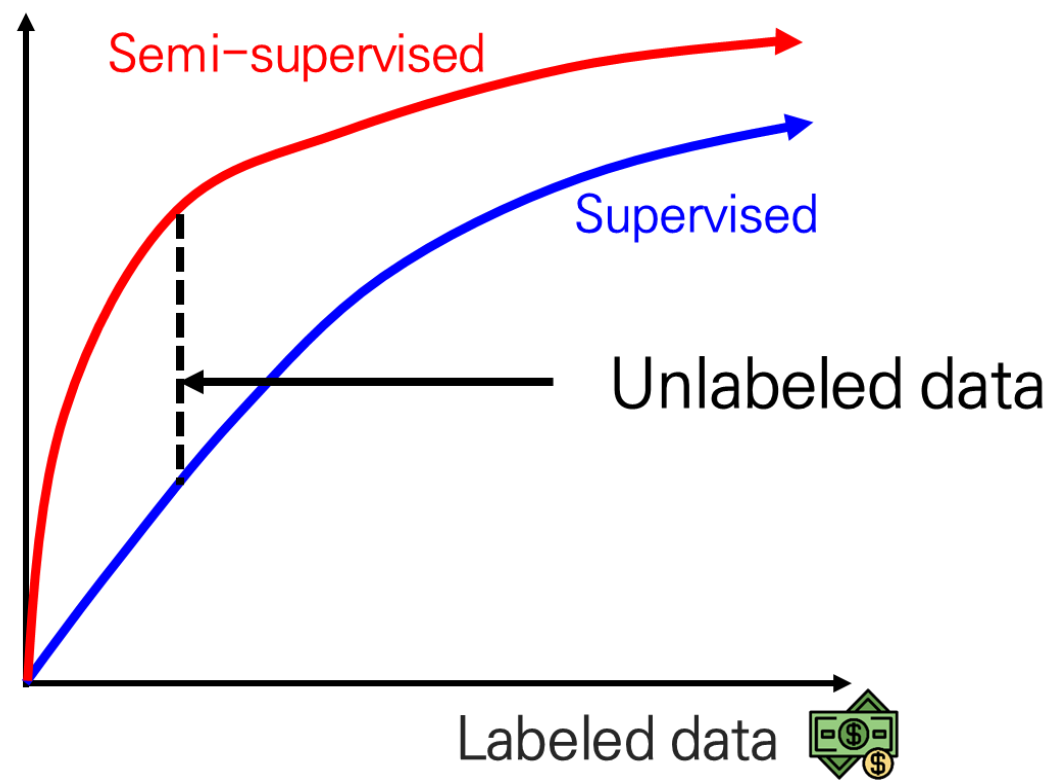


Deep Semi-Supervised Learning

- ❖ 상대적으로 수집 비용이 저렴한 데이터셋을 적극 활용해 성능을 높이는 방안 존재 → Semi-Supervised Learning [2]



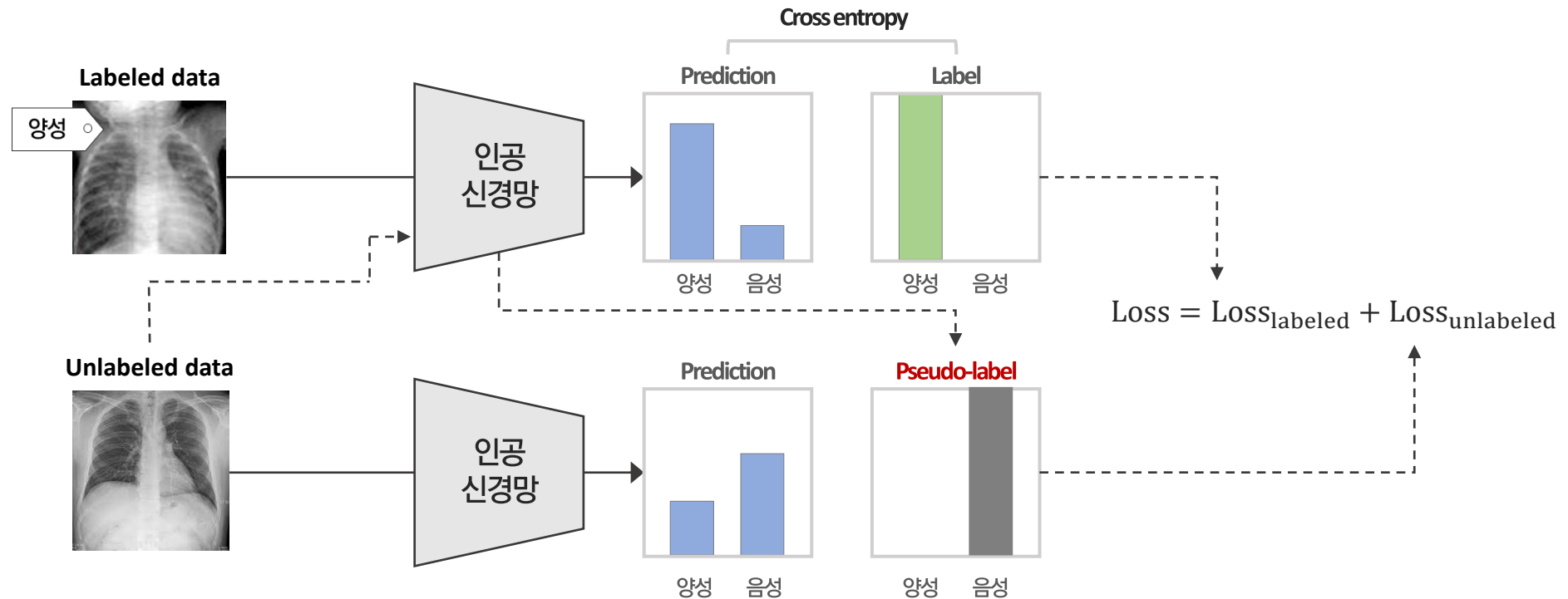
Performance



Deep Semi-Supervised Learning

Self-Training [3]

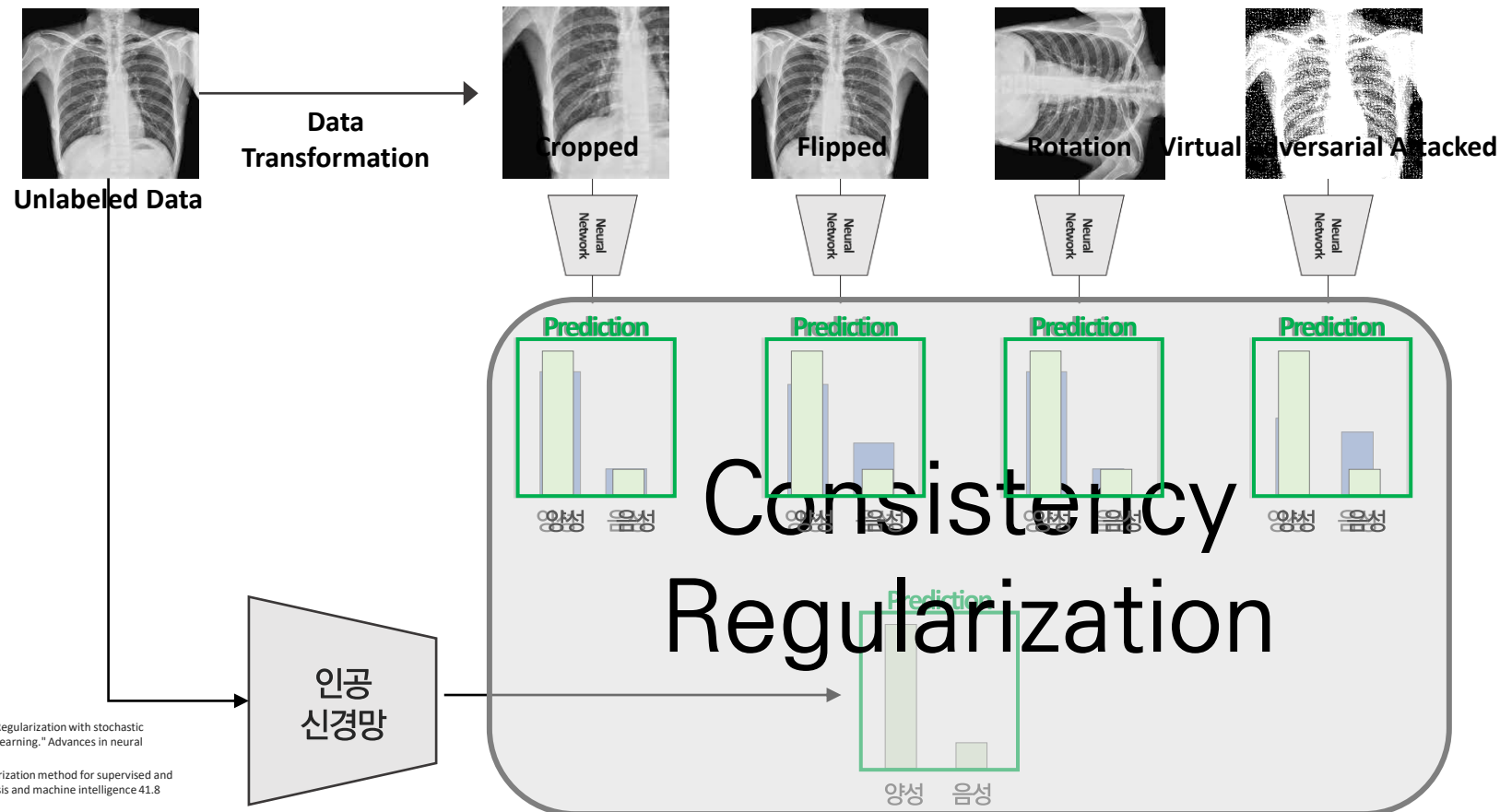
Semi-Supervised Learning → Unlabeled Data를 어떻게 활용하여 성능을 높였을까?



Deep Semi-Supervised Learning

Consistency Regularization [4,5]

Semi-Supervised Learning → Unlabeled Data를 어떻게 활용하여 성능을 높였을까?



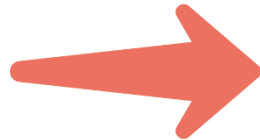
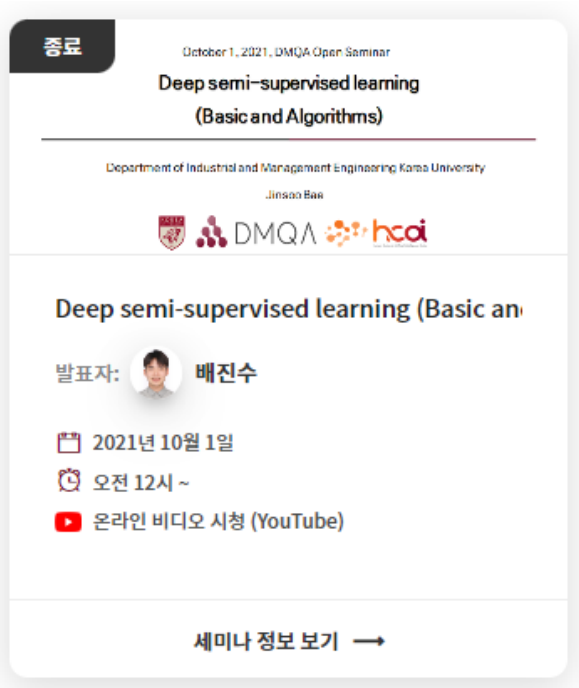
[4] Sajjadi, Mehdi, Mehran Javanmardi, and Tolga Tasdizen. "Regularization with stochastic transformations and perturbations for deep semi-supervised learning." Advances in neural information processing systems 29 (2016): 1163-1171.

[5] Miyato, Takeru, et al. "Virtual adversarial training: a regularization method for supervised and semi-supervised learning." IEEE transactions on pattern analysis and machine intelligence 41.8 (2018): 1979-1993.

Deep Semi-Supervised Learning

DMQA 연구실에서 진행한 준지도학습 세미나

- <http://dmqa.korea.ac.kr/activity/seminar>



준지도학습의 구체적인 원리 이해 가능

대표적인 준지도학습 알고리즘 5개 설명

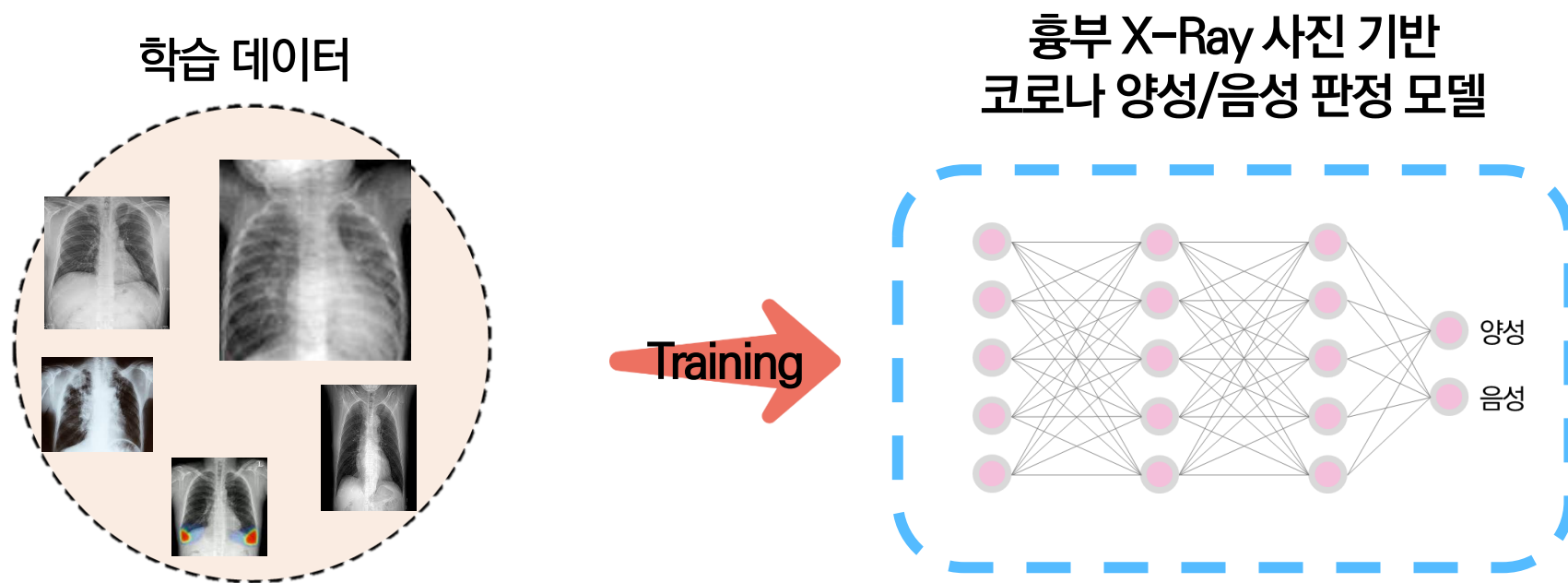
2. Out-of-distribution (O.O.D) Unlabeled Data

Out-of-distribution (O.O.D) Unlabeled Data

- ❖ Out-of-distribution Data : 인공지능 모델 학습에 사용되었던 데이터셋에 속하지 않은 데이터를 의미

Out-of-distribution (O.O.D) Unlabeled Data

- ❖ Out-of-distribution Data : 인공지능 모델 학습에 사용되었던 데이터셋에 속하지 않은 데이터를 의미

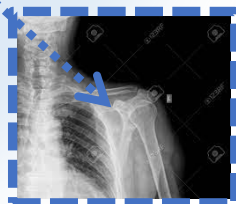
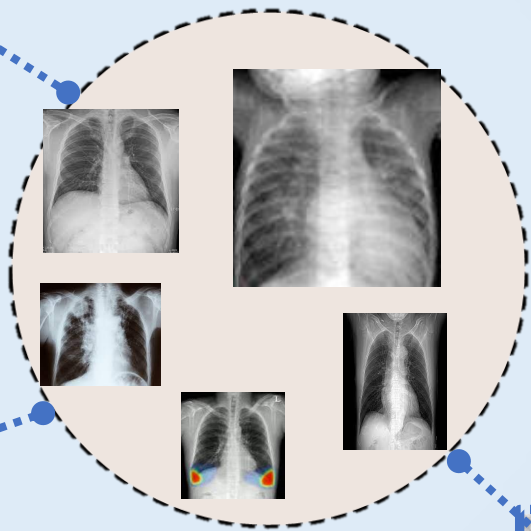
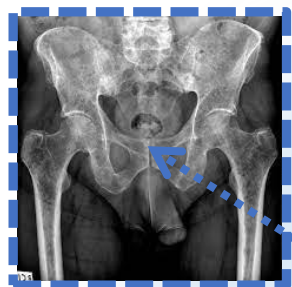


Out-of-distribution (O.O.D) Unlabeled Data

❖ Out-of-distribution Data : 인공지능 모델 학습에 사용되었던 데이터셋에 속하지 않은 데이터를 의미

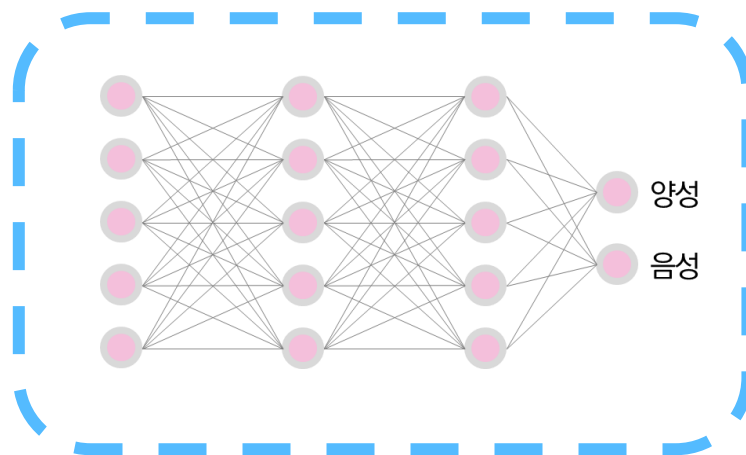
Out-of-distribution

학습 데이터

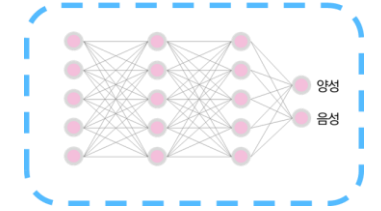


Training

흉부 X-Ray 사진 기반
코로나 양성/음성 판정 모델

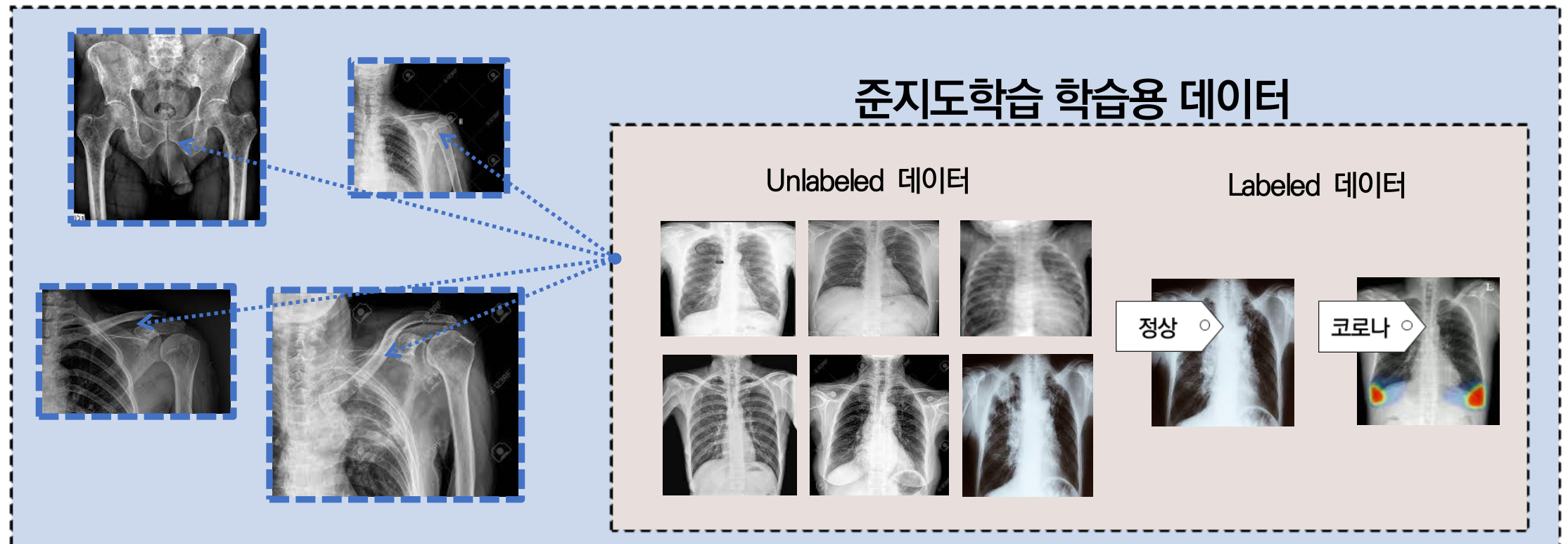


Out-of-distribution (O.O.D) Unlabeled Data



- ❖ Out-of-distribution Unlabeled Data : 인공지능 모델 학습에 사용된 Labeled 데이터셋에 속하지 않은 Unlabeled 데이터

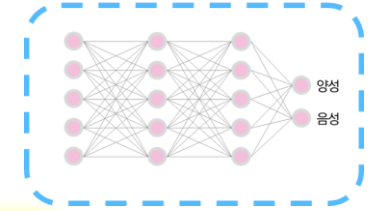
Out-of-distribution Unlabeled Data



Out-of-distribution (O.O.D) Unlabeled Data

O.O.D Unlabeled Data와 SSL이 무슨 관계를 갖고 있을까?

Out-of-distribution (O.O.D) Unlabeled Data



O.O.D Unlabeled Data와 SSL이 무슨 관계를 갖고 있을까?

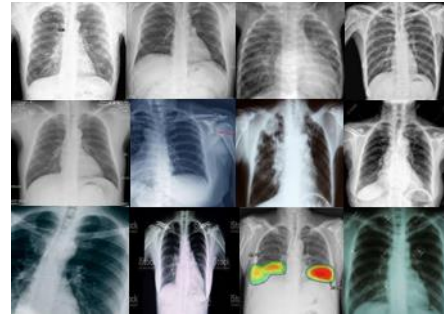
Case. 1

Labeled

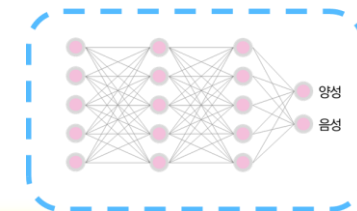
Unlabeled

정상 ○

코로나 ○



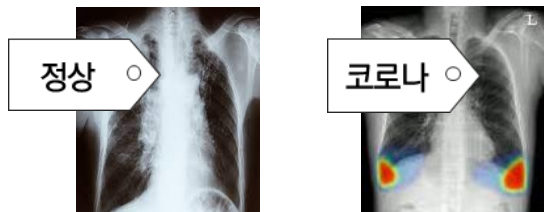
Out-of-distribution (O.O.D) Unlabeled Data



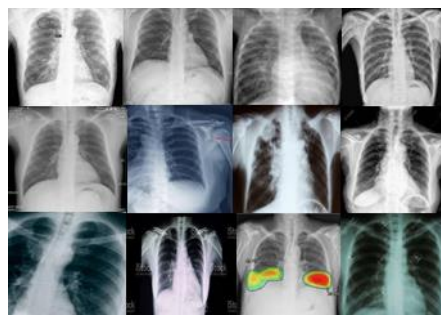
O.O.D Unlabeled Data와 SSL이 무슨 관계를 갖고 있을까?

Case. 1

Labeled

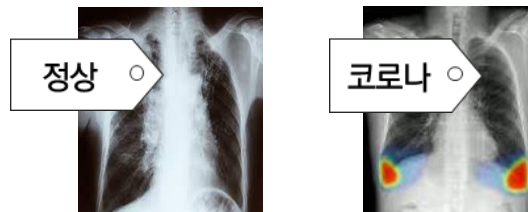


Unlabeled



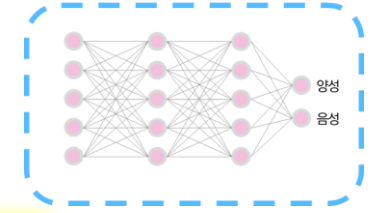
Case. 2

Labeled data



Unlabeled data



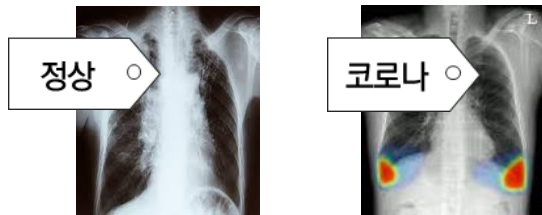


Out-of-distribution (O.O.D) Unlabeled Data

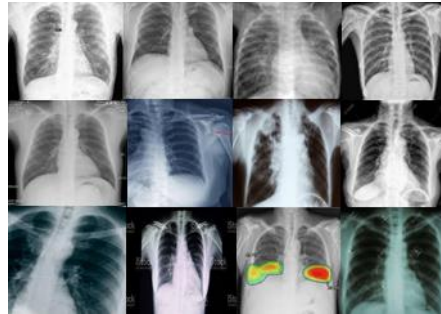
O.O.D Unlabeled Data와 SSL이 무슨 관계를 갖고 있을까?

Case. 1

Labeled



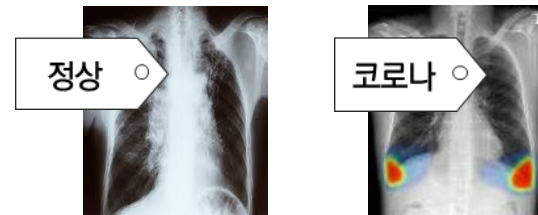
Unlabeled



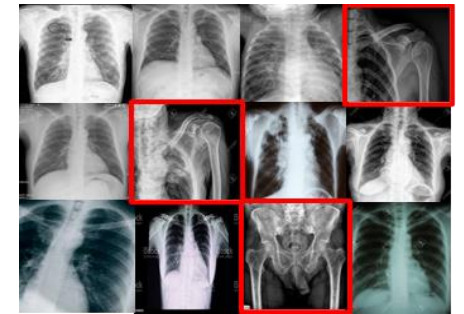
수집된 모든 Unlabeled 데이터가
Labeled 데이터셋과 유사한 것은 매우 이상적인 상황

Case. 2

Labeled data



Unlabeled data



Unlabeled 데이터는 사람의 검수를 받지 않고 수집되기 때문에
Labeled 데이터셋과 무관한 O.O.D 데이터일 확률이 높음

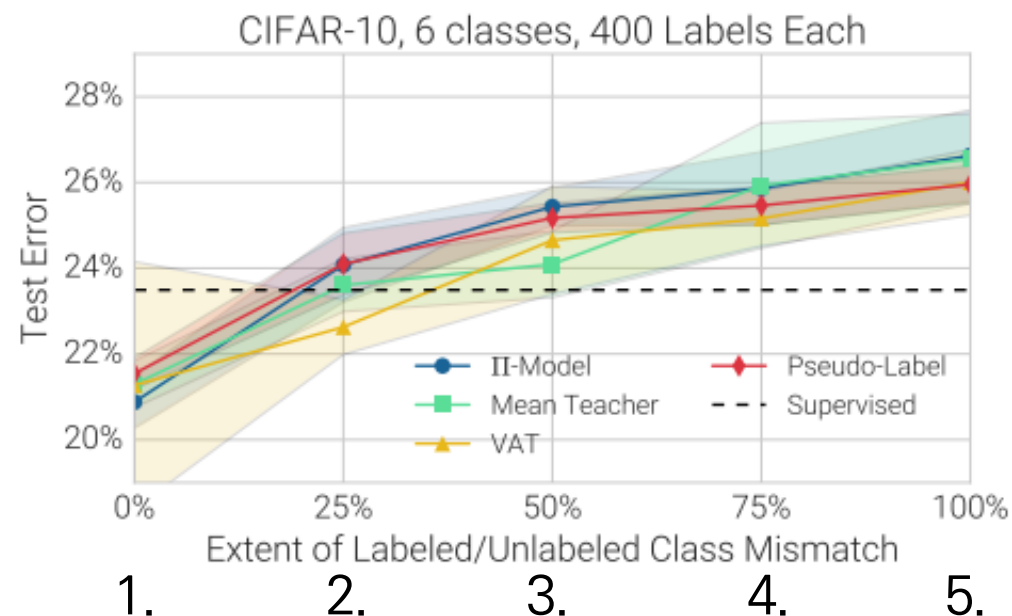
Out-of-distribution (O.O.D) Unlabeled Data

O.O.D Unlabeled Data와 SSL이 무슨 관계를 갖고 있을까?

“Google Brain 연구원들이 NeuralPS에 게재한 최신 논문 [2]에 따르면,
O.O.D Unlabeled 데이터는 SSL 학습에 부정적인 영향을 끼침”

CIFAR 10 class info		Mismatch ratio
Labeled data (400)	Unlabeled data (4100)	
1. Bird,cat,deer,dog,fog,horse	Deer, dog, fog, horse	0%
2. Bird,cat,deer,dog,fog,horse	Airplane, dog, fog, horse	25%
3. Bird,cat,deer,dog,fog,horse	Airplane, automobile, fog, horse	50%
4. Bird,cat,deer,dog,fog,horse	Airplane, automobile, ship, horse	75%
5. Bird,cat,deer,dog,fog,horse	Airplane, automobile, ship, truck	100%

→ O.O.D Unlabeled Data ($\cong$ Open-set Data)



Out-of-distribution (O.O.D) Unlabeled Data

O.O.D Unlabeled Data와 SSL이 무슨 관계를 갖고 있을까?

“Google Brain 연구원들이 NeuralPS에 게재한 최신 논문 [2]에 따르면,
O.O.D Unlabeled 데이터는 SSL 학습에 부정적인 영향을 끼침”



Safe SSL Against O.O.D Unlabeled Data

= Unlabeled 데이터 내의 O.O.D 데이터 존재와 상관 없이, 지도학습 대비 우수한 성능을 낼 수 있는

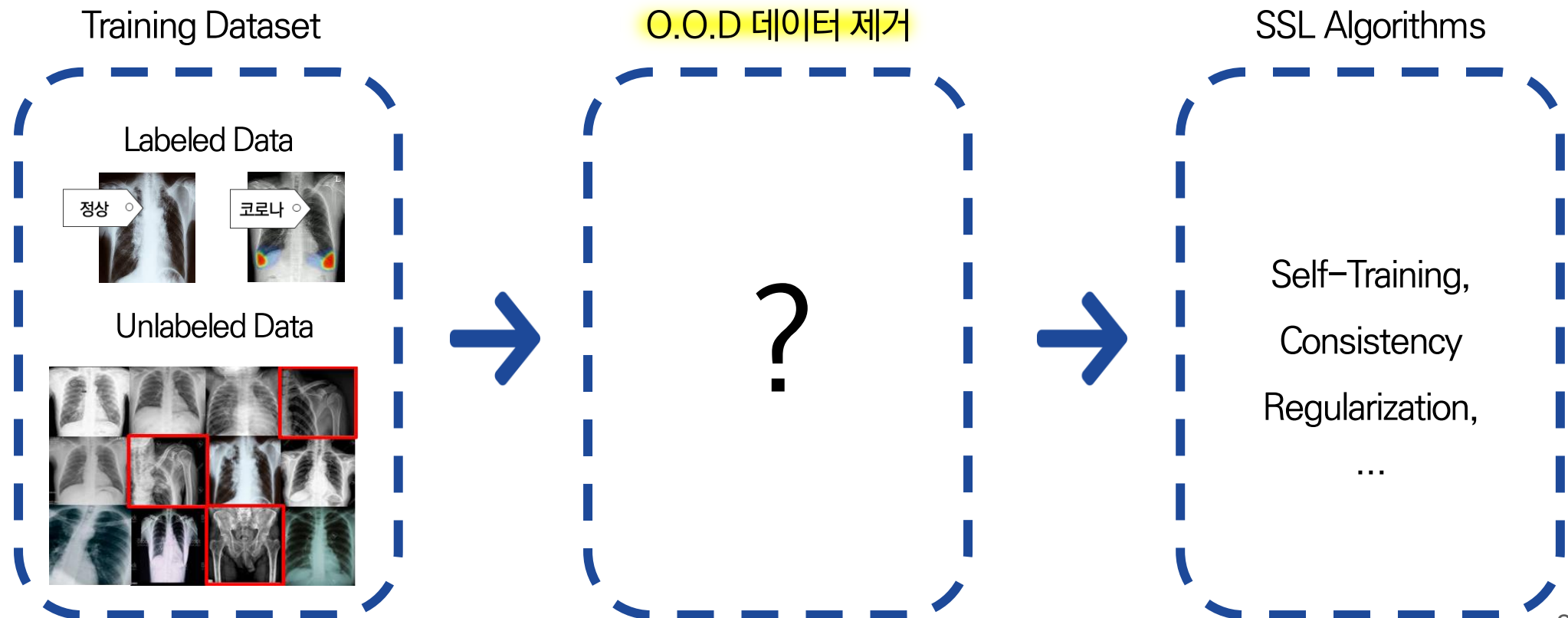
3. Safe SSL Against O.O.D Unlabeled Data

→ UASD, DS3L, SafeUC

Safe SSL Against O.O.D Unlabeled Data

1. Uncertainty-Aware Self-Distillation (UASD, 2020년 AAAI 게재)

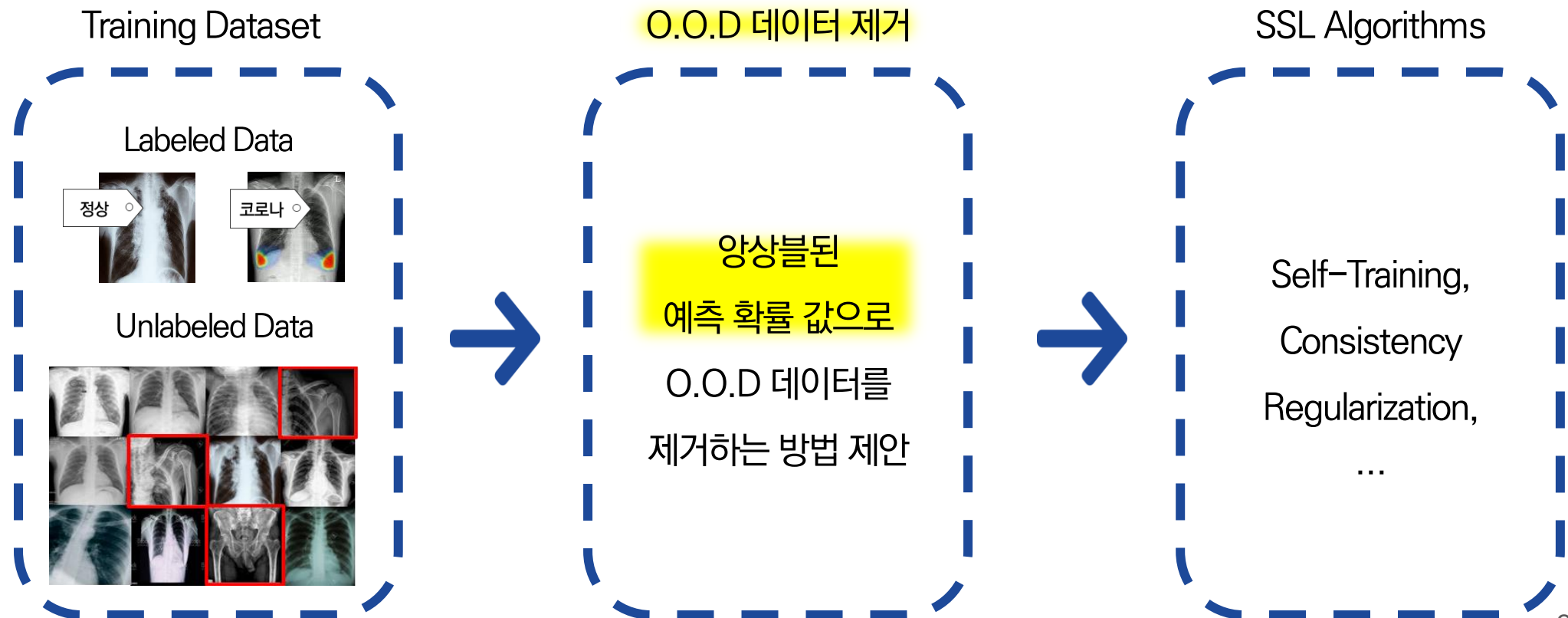
SSL에서 O.O.D Unlabeled Data의 악영향을 어떻게 줄일 수 있을까?



Safe SSL Against O.O.D Unlabeled Data

1. Uncertainty-Aware Self-Distillation (UASD, 2020년 AAAI 게재)

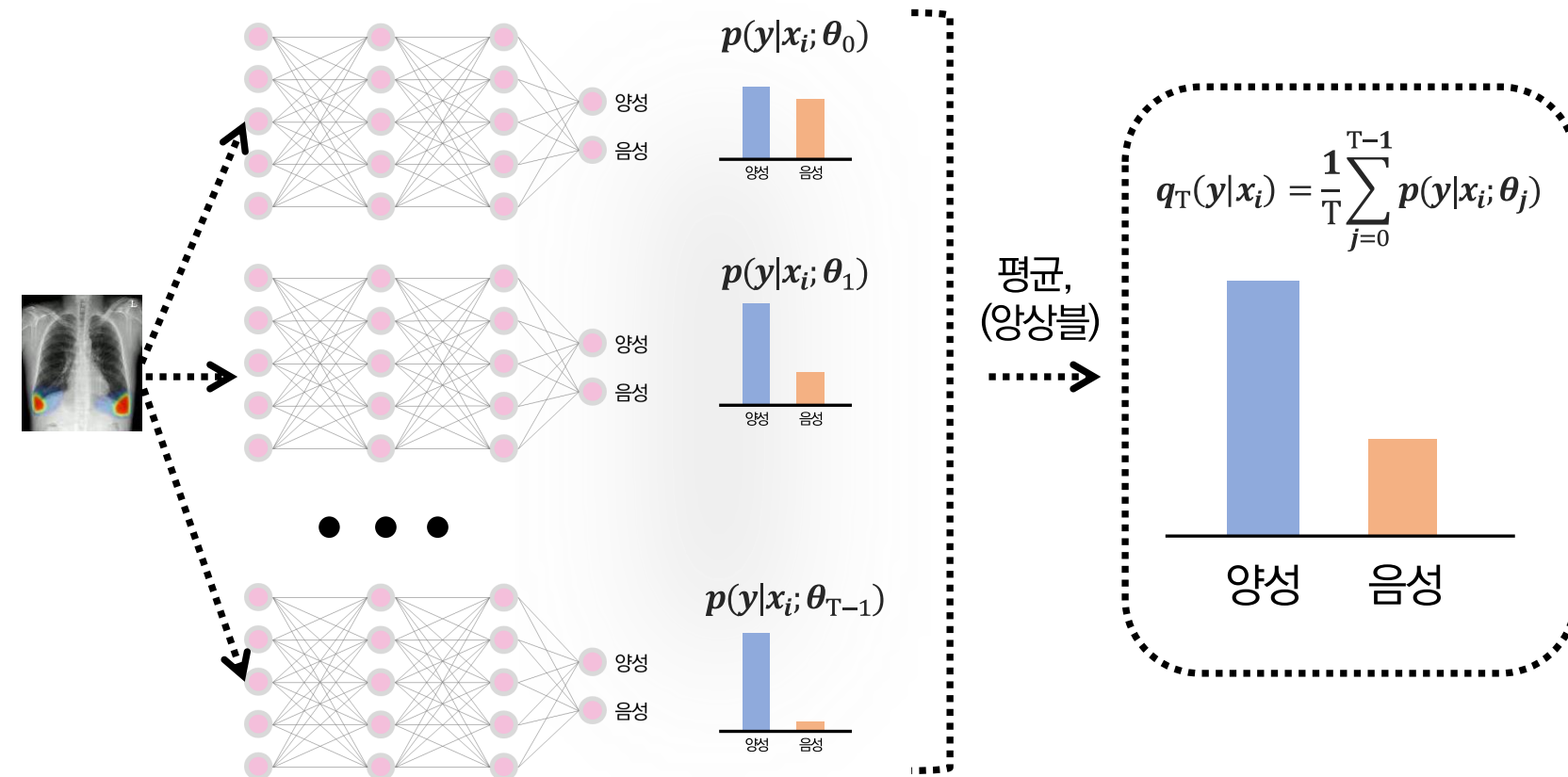
SSL에서 O.O.D Unlabeled Data의 악영향을 어떻게 줄일 수 있을까?



Safe SSL Against O.O.D Unlabeled Data

1. Uncertainty-Aware Self-Distillation (UASD, 2020년 AAAI 게재)

- ❖ 매 에폭 별 학습된 인공지능망 모델의 예측 확률 값들을 앙상블하여, O.O.D 데이터인지 아닌지 구분짓는 방법 제시
 - 앙상블 적용 이유 1. 단일 모델의 예측 확률 값의 경우, O.O.D 데이터에 대한 인식 능력이 현저히 떨어짐 (Over-Confidence Problems)

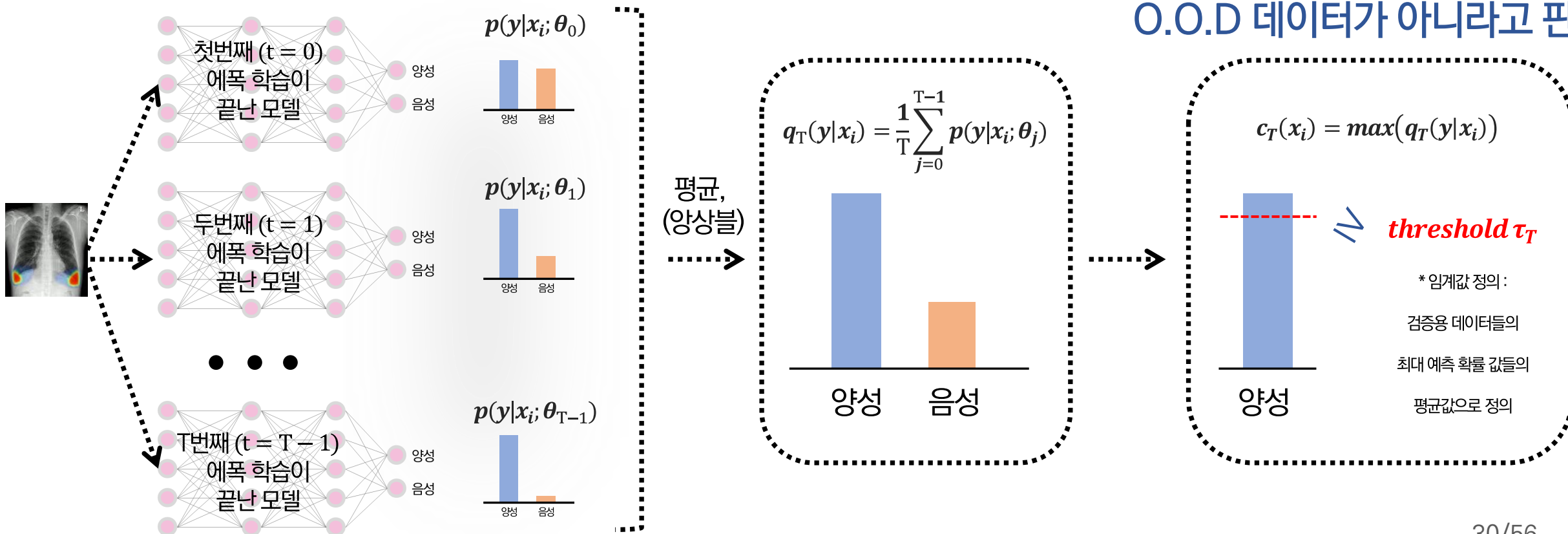


Safe SSL Against O.O.D Unlabeled Data

1. Uncertainty-Aware Self-Distillation (UASD, 2020년 AAAI 게재)

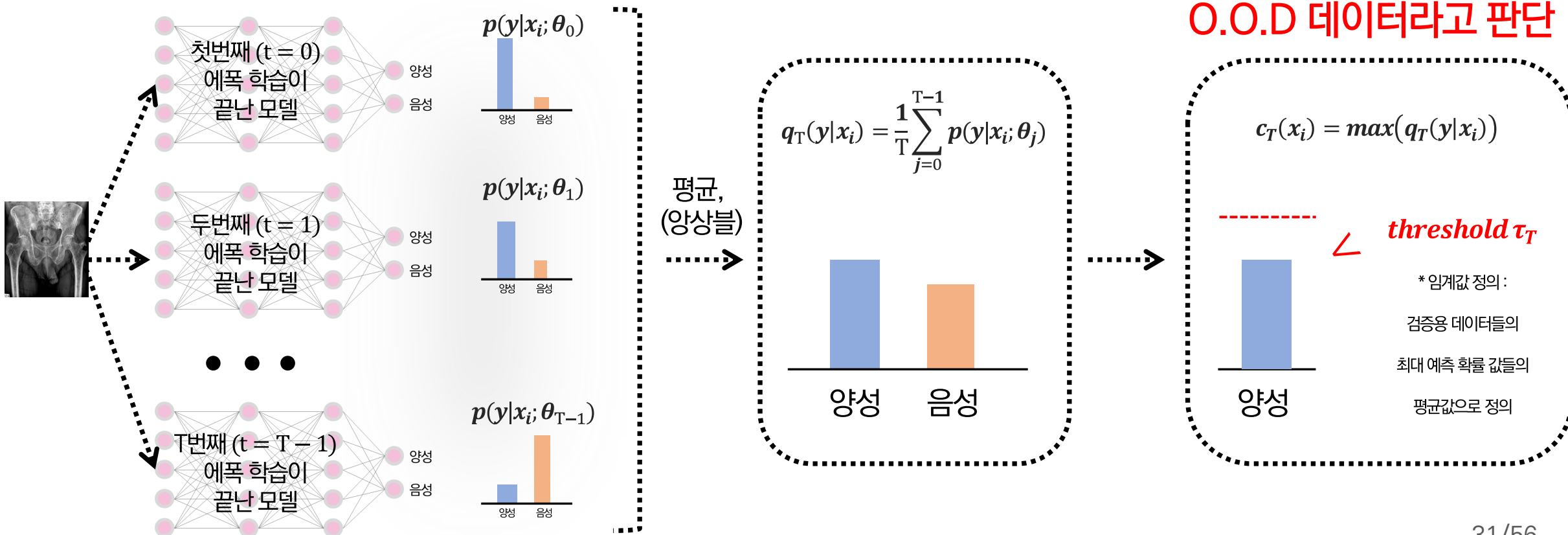
- ❖ 매 에폭 별 학습된 인공지능망 모델의 예측 확률 값들을 앙상블하여, O.O.D 데이터인지 아닌지 구분짓는 방법 제시
 - 앙상블 적용 이유 1. 단일 모델의 예측 확률 값의 경우, O.O.D 데이터에 대한 인식 능력이 현저히 떨어짐 (Over-Confidence Problems)

O.O.D 데이터가 아니라고 판단



1. Uncertainty-Aware Self-Distillation (UASD, 2020년 AAAI 게재)

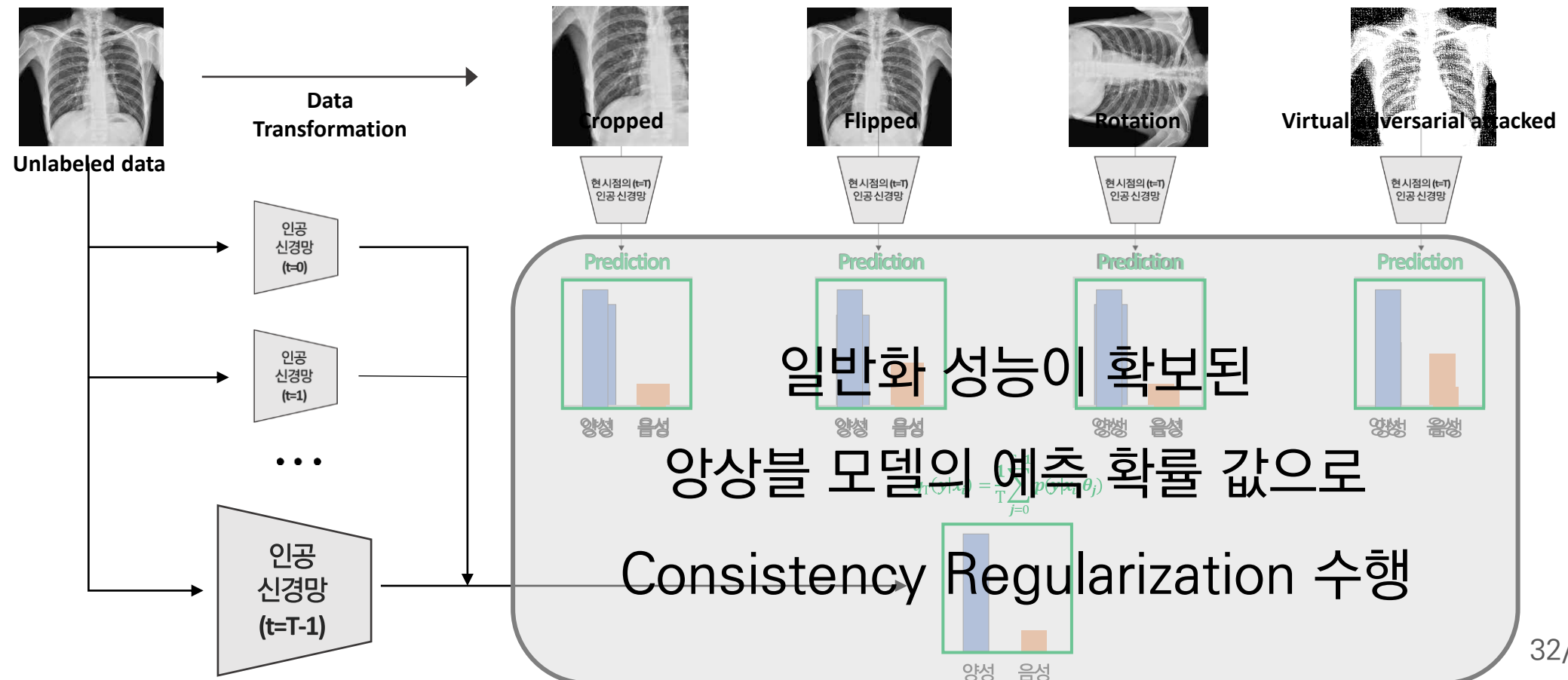
- ❖ 매 에폭 별 학습된 인공신경망 모델의 예측 확률 값들을 앙상블하여, O.O.D 데이터인지 아닌지 구분짓는 방법 제시
 - 앙상블 적용 이유 1. 단일 모델의 예측 확률 값의 경우, O.O.D 데이터에 대한 인식 능력이 현저히 떨어짐 (Over-Confidence Problems)



Safe SSL Against O.O.D Unlabeled Data

1. Uncertainty-Aware Self-Distillation (UASD, 2020년 AAAI 게재)

- ❖ 매 에폭 별 학습된 인공지능망 모델의 예측 확률 값들을 앙상블하여, O.O.D 데이터인지 아닌지 구분짓는 방법 제시
 - 앙상블 적용 이유 2. 일반화 능력이 좋은 앙상블의 예측 확률 값으로 Consistency Regularization을 수행해 현 시점의 예측 모델 성능 향상



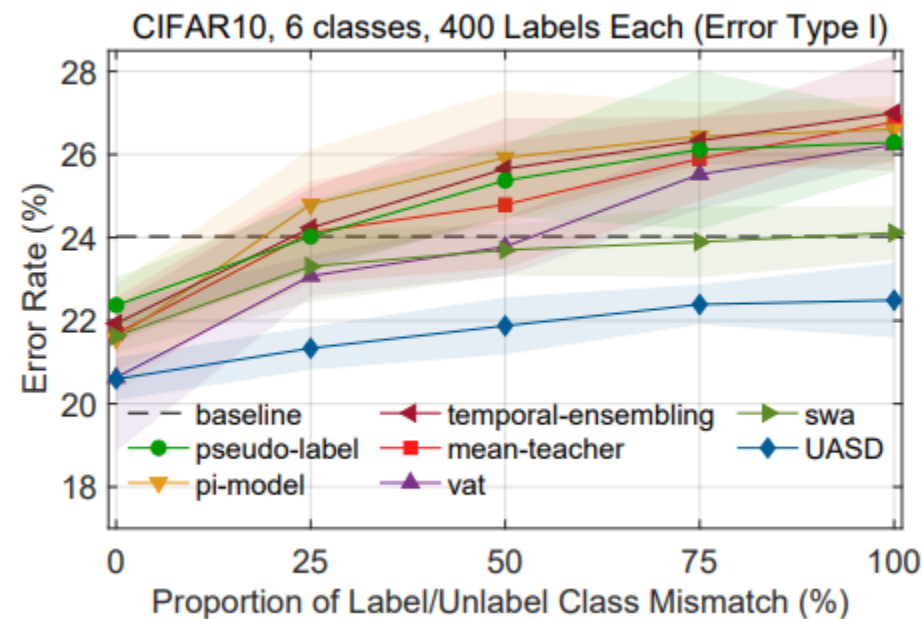
Safe SSL Against O.O.D Unlabeled Data

1. Uncertainty-Aware Self-Distillation (UASD, 2020년 AAAI 게재)

- ❖ UASD는 앙상블 아이디어를 활용해 O.O.D Unlabeled 데이터의 악영향을 줄이며, 안전한 준지도학습을 수행함
 - 기존 준지도학습 방법론들의 경우, O.O.D 데이터로 인해 성능이 떨어지는 경향을 보임
 - UASD 방법론은 O.O.D 데이터의 존재 여부에 상관없이 지도학습보다 좋은 성능을 보이며, 안전성을 입증함

CIFAR 10 class info		Mismatch ratio
Labeled data (400)	Unlabeled data (4100)	
1. Bird,cat,deer,dog,fog,horse	Deer, dog, fog, horse	0%
2. Bird,cat,deer,dog,fog,horse	Airplane, dog, fog, horse	25%
3. Bird,cat,deer,dog,fog,horse	Airplane, automobile, fog, horse	50%
4. Bird,cat,deer,dog,fog,horse	Airplane, automobile, ship, horse	75%
5. Bird,cat,deer,dog,fog,horse	Airplane, automobile, ship, truck	100%

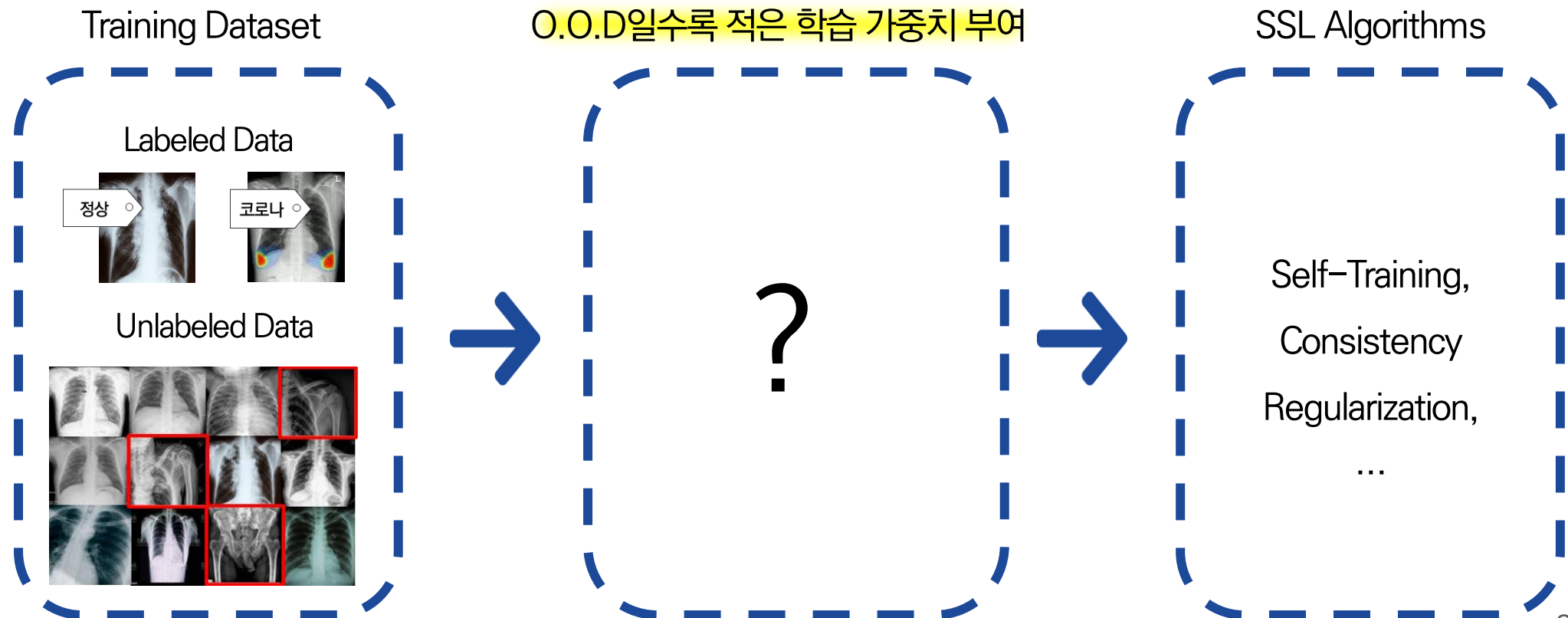
→ O.O.D Unlabeled Data ($\cong$ Open-set Data)



Safe SSL Against O.O.D Unlabeled Data

2. Deep Safe Semi-Supervised Learning (DS3L, 2020년 PLMR 게재)

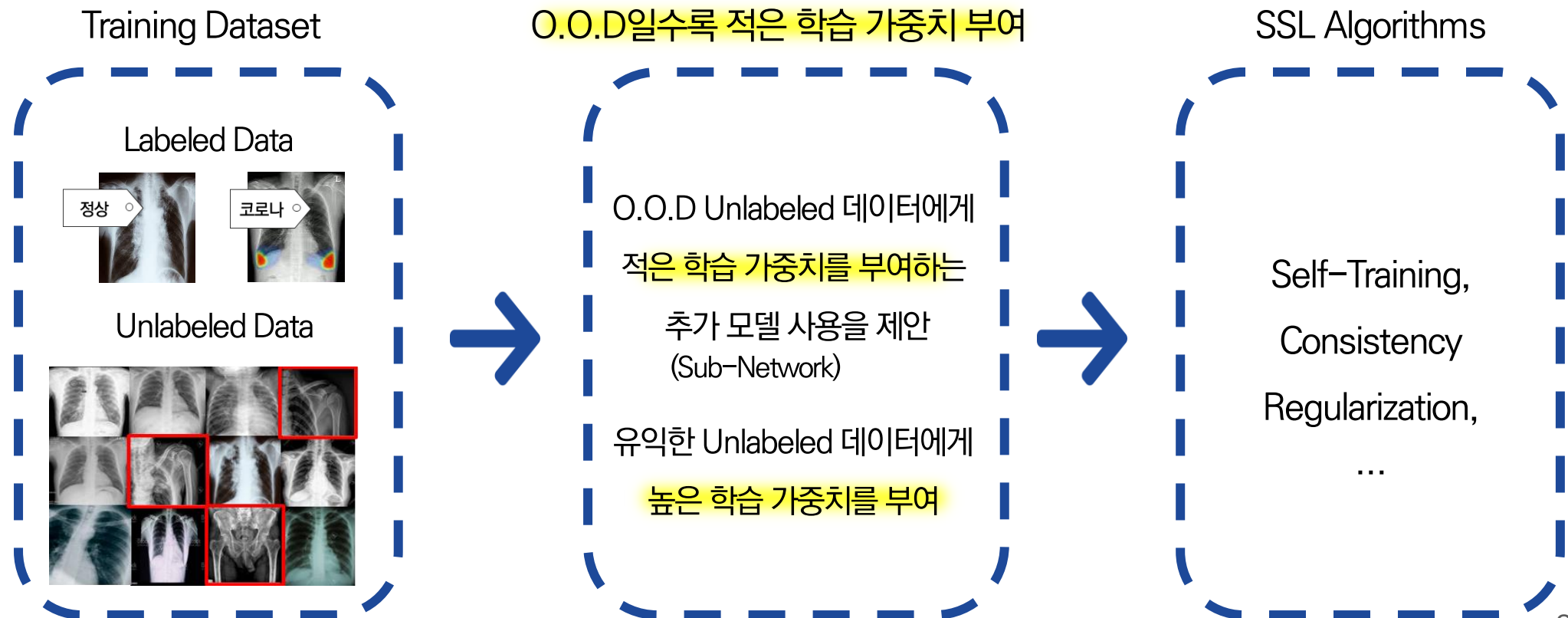
SSL에서 O.O.D Unlabeled Data의 악영향을 어떻게 줄일 수 있을까?



Safe SSL Against O.O.D Unlabeled Data

2. Deep Safe Semi-Supervised Learning (DS3L, 2020년 PLMR 게재)

SSL에서 O.O.D Unlabeled Data의 악영향을 어떻게 줄일 수 있을까?

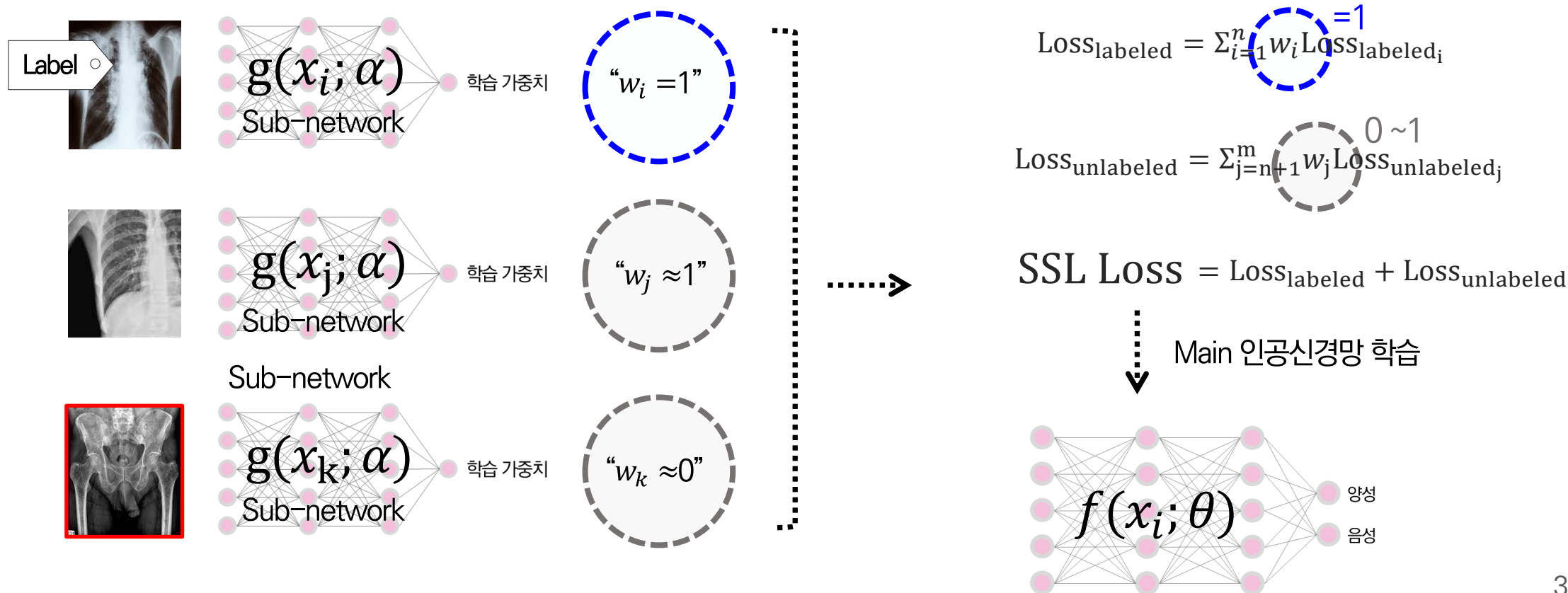


Safe SSL Against O.O.D Unlabeled Data

2. Deep Safe Semi-Supervised Learning (DS3L, 2020년 PLMR 게재)

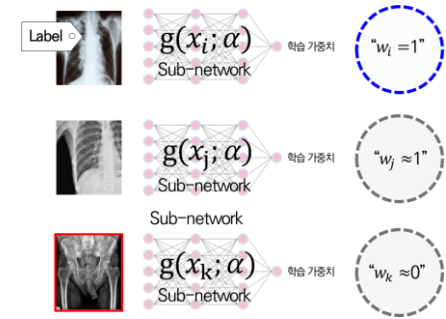
(Sub-network, $g(x; \alpha)$)

- ❖ Unlabeled 데이터가 유익할수록 높은 가중치를 부여하고, O.O.D 데이터는 낮은 가중치를 부여하는 별도의 추가 모델 제안
 - 0과 1 사이의 값 내에서, Labeled 데이터에게는 학습 가중치 1을 부여하고, O.O.D 데이터는 0에 가까운 학습 가중치를 부여



Safe SSL Against O.O.D Unlabeled Data

2. Deep Safe Semi-Supervised Learning (DS3L, 2020년 PLMR 게재)

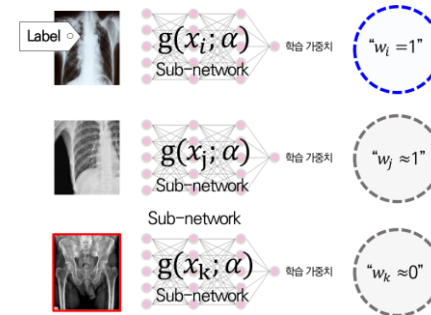


DS3L은 O.O.D 데이터의 가중치를 감소시키는 Subnetwork를 어떻게 학습할까?

→ O.O.D Unlabeled 데이터의 악영향은 Main 인공지능망 모델의 Labeled 데이터에 대한 학습에 부정적인 영향을 끼칠 것이다 가정

Safe SSL Against O.O.D Unlabeled Data

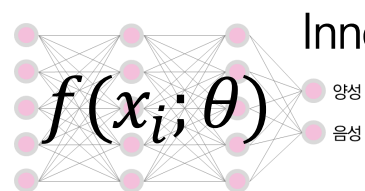
2. Deep Safe Semi-Supervised Learning (DS3L, 2020년 PLMR 게재)



DS3L은 O.O.D 데이터의 가중치를 감소시키는 Subnetwork를 어떻게 학습할까?

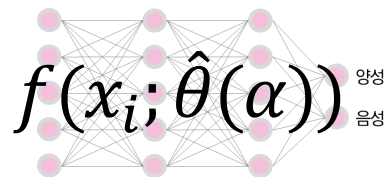
→ O.O.D Unlabeled 데이터의 악영향은 Main 인공지능망 모델의 Labeled 데이터에 대한 학습에 부정적인 영향을 끼칠 것이다 가정

$$\hat{\alpha} = \operatorname{argmin}_{\alpha} \{ \text{Supervised Loss}(\alpha) = \sum_{i=1}^n w_i \text{Loss}_{\text{labeled}_i} \}$$



Inner optimization : Update θ

$$\operatorname{argmin}_{\theta} \{ \text{SSL Loss}(\theta) = \sum_{i=1}^n w_i \text{Loss}_{\text{labeled}_i} + \sum_{j=n+1}^m w_j \text{Loss}_{\text{unlabeled}_j} \}$$



$$\hat{\theta}(\alpha) = \operatorname{argmin}_{\theta} \{ \sum_{i=1}^n w_i(\alpha) \text{Loss}_{\text{labeled}_i} + \sum_{j=n+1}^m w_j(\alpha) \text{Loss}_{\text{unlabeled}_j} \}$$

Outer optimization : Update α

Safe SSL Against O.O.D Unlabeled Data

2. Deep Safe Semi-Supervised Learning (DS3L, 2020년 PLMR 게재)

- ❖ DS3L은 별도의 추가 네트워크 학습을 통해 O.O.D Unlabeled 데이터의 악영향을 줄이며, 안전한 준지도학습을 수행함
 - 기존 준지도학습 방법론들의 경우, O.O.D 데이터로 인해 성능이 떨어지는 경향을 보임
 - DS3L은 기존 준지도학습들보다 우수한 성능을 보이며, 안전성을 입증함

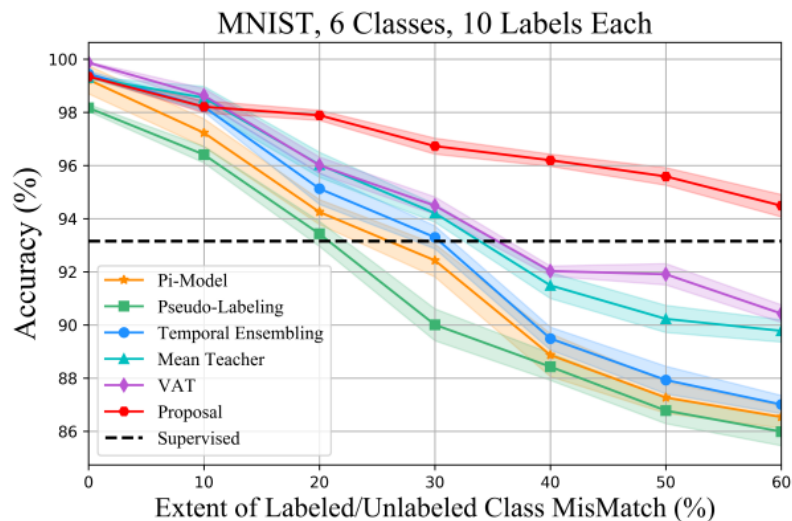


Figure 5. Classification accuracy of compared deep SSL techniques and DS³L on MNIST data set (class 1 – 6) with varying class mismatch ratio between labeled and unlabeled data. Shaded regions indicate standard deviation over five runs.

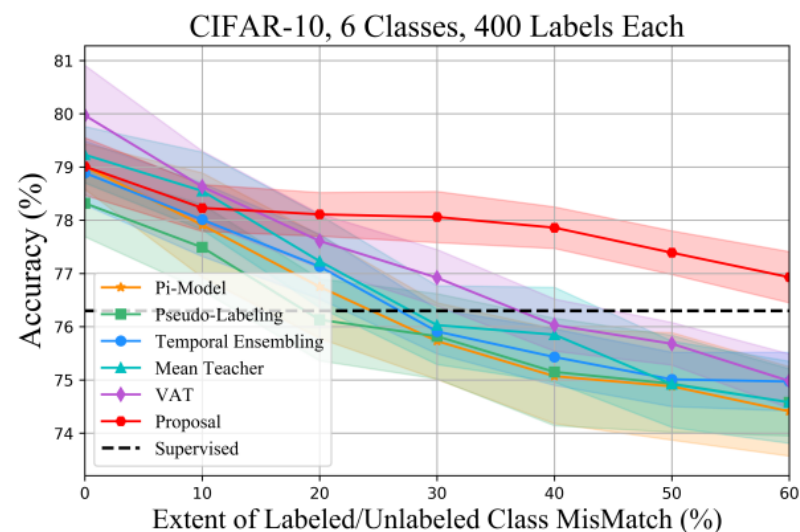
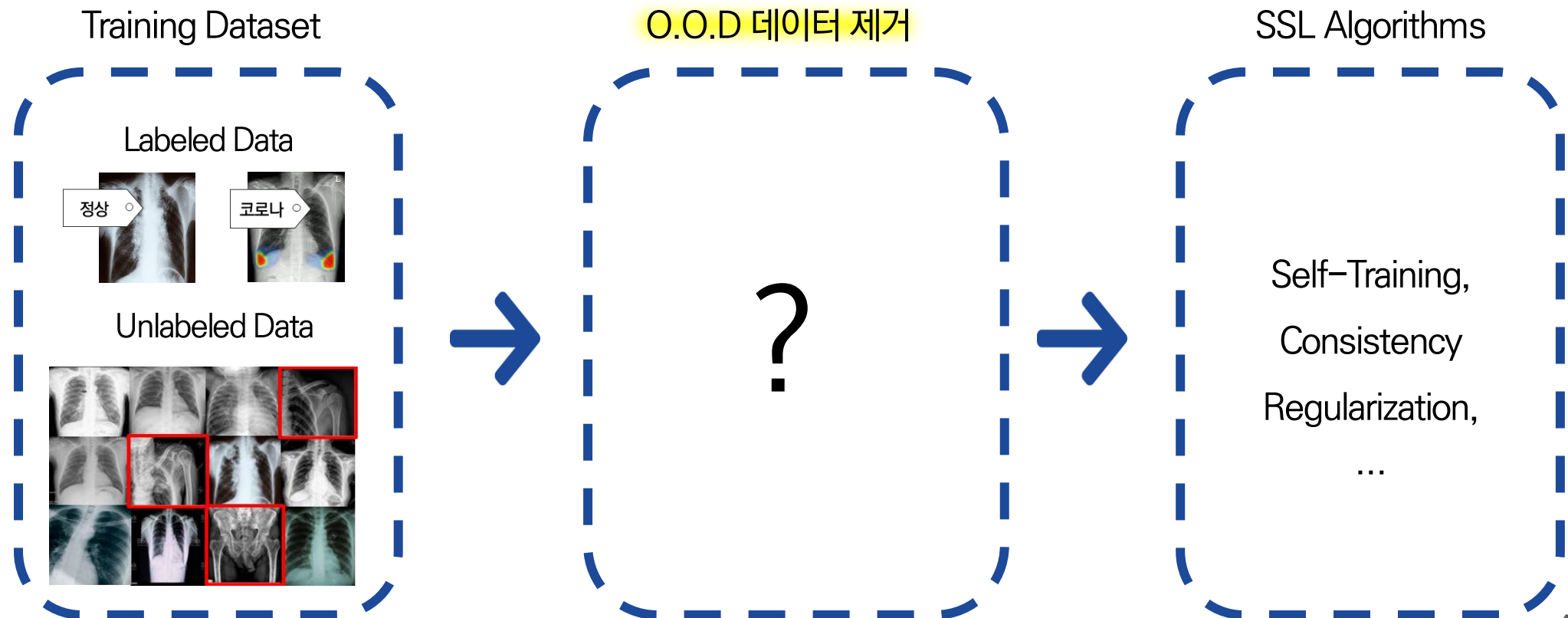


Figure 6. Classification accuracy of compared deep SSL techniques and DS³L on CIFAR-10 data set with varying class mismatch ratio between labeled and unlabeled data. Shaded regions indicate standard deviation over five runs.

Safe SSL Against O.O.D Unlabeled Data

3. Safe Uncertainty-based Consistency Training (SafeUC, 2022+년 Information Sciences Accept)

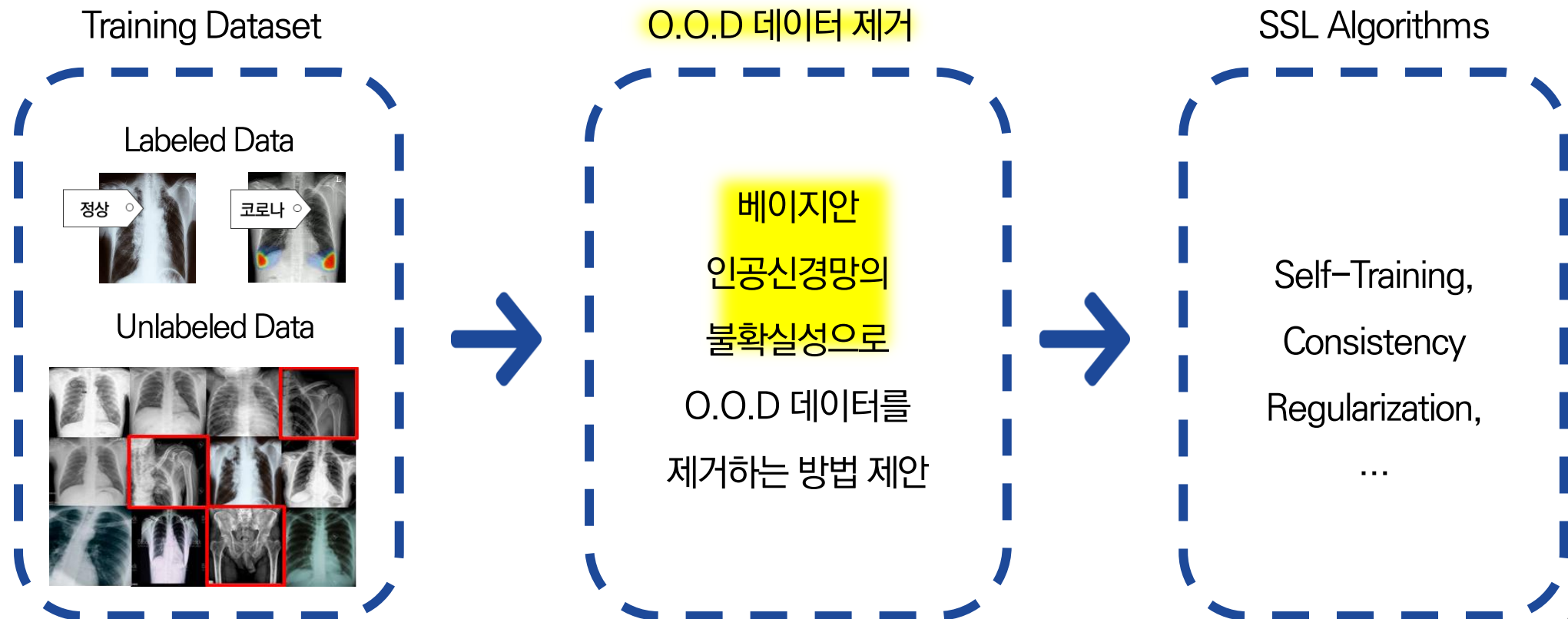
SSL에서 O.O.D Unlabeled Data의 악영향을 어떻게 줄일 수 있을까?



Safe SSL Against O.O.D Unlabeled Data

3. Safe Uncertainty-based Consistency Training (SafeUC, 2022+년 Information Sciences Accept)

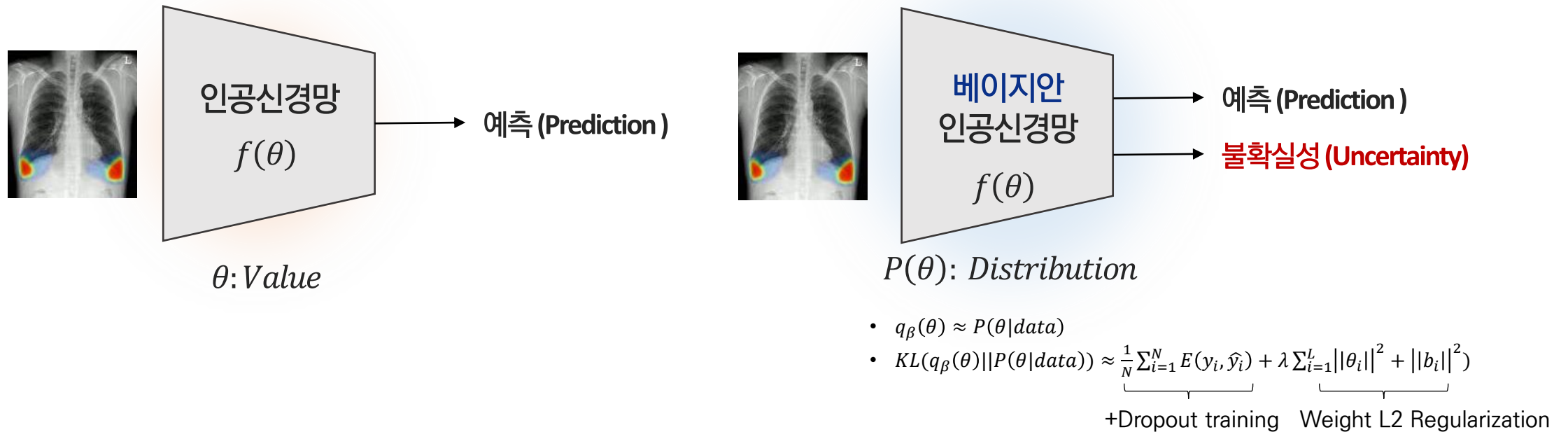
SSL에서 O.O.D Unlabeled Data의 악영향을 어떻게 줄일 수 있을까?



Safe SSL Against O.O.D Unlabeled Data

3. Safe Uncertainty-based Consistency Training (SafeUC, 2022+년 Information Sciences Accept)

- ❖ 베이지안 인공지능망은 모델 파라미터에 사전 분포를 가정하고, 예측 확률 분포를 학습하여 예측 불확실성 산출이 가능함 [3]
 - Dropout과 Weight L2 Regularization 기법을 통해 임의의 인공지능망을 베이지안 인공지능망으로 학습할 수 있음이 입증됨 [4,5]



[3] Anzai, Yuichiro. Pattern recognition and machine learning. Elsevier, 2012.

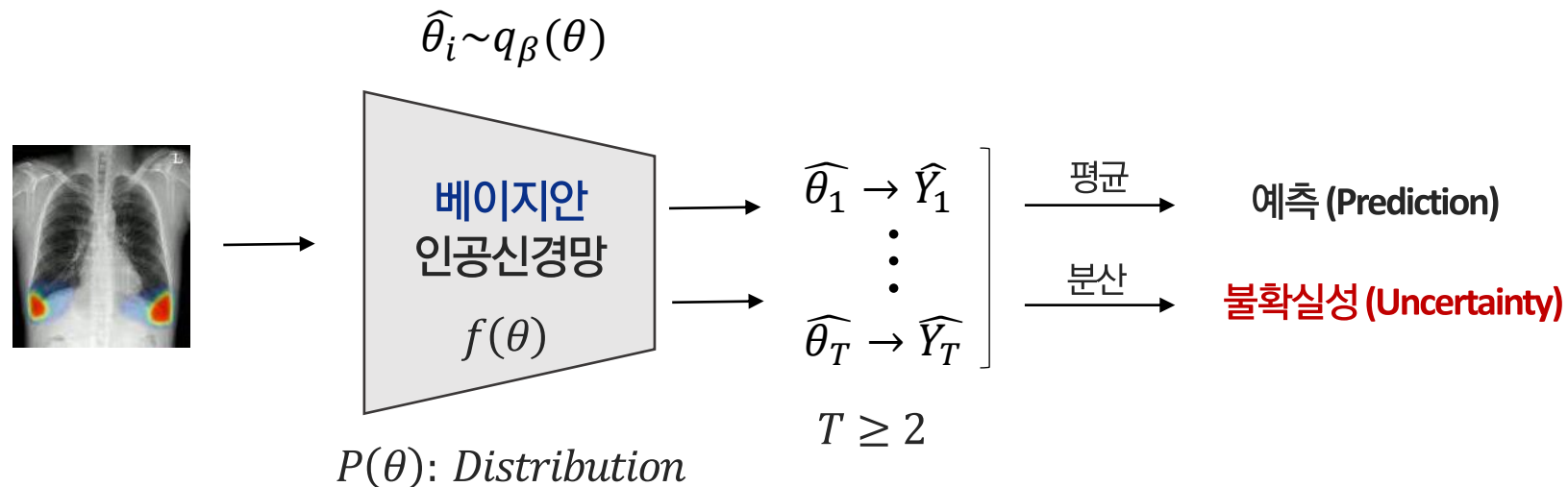
[4] Gal, Y., & Ghahramani, Z. (2016, June). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In international conference on machine learning (pp. 1050-1059). PMLR.

[5] Gal, Y., Hron, J., & Kendall, A. (2017). Concrete dropout. arXiv preprint arXiv:1705.07832.

Safe SSL Against O.O.D Unlabeled Data

3. Safe Uncertainty-based Consistency Training (SafeUC, 2022+년 Information Sciences Accept)

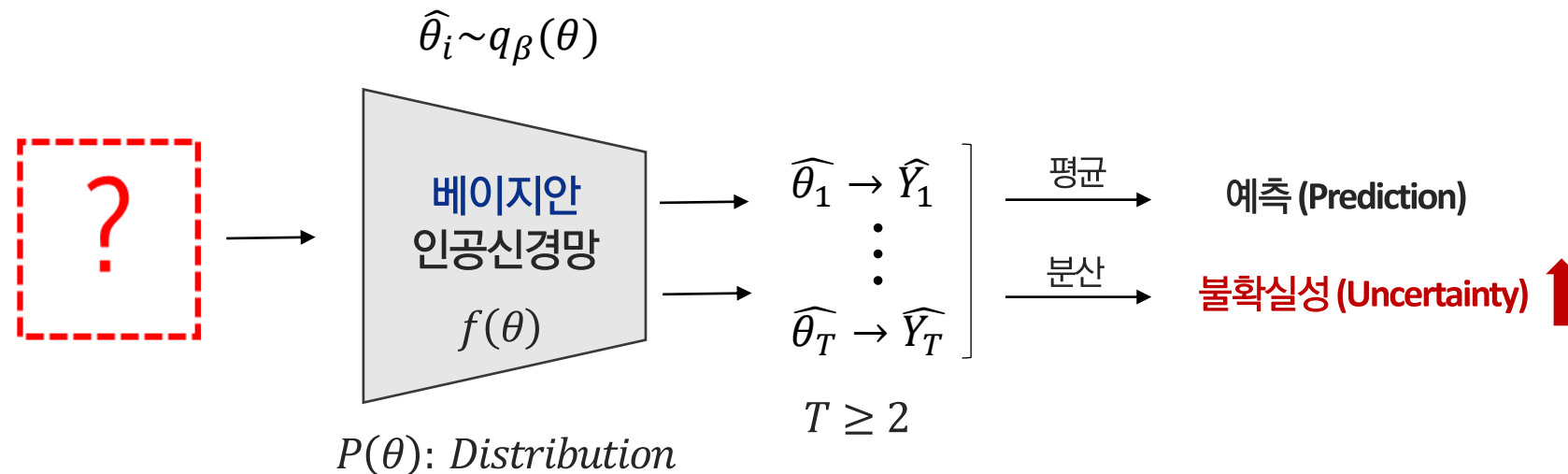
- ❖ 베이지안 인공지능망의 불확실성은 사후 분포에서 샘플링한 2개 이상의 모델 파라미터에 의해 얻어진 예측 확률 값들의 분산 [4]
 - 학습 데이터 분포에서 멀리 벗어난 데이터일수록 높은 값을 가진 불확실성이 산출되며, 모델이 일관되지 않은 예측을 진행했음을 의미함 [4,6]



Safe SSL Against O.O.D Unlabeled Data

3. Safe Uncertainty-based Consistency Training (SafeUC, 2022+년 Information Sciences Accept)

- ❖ 베이지안 인공지능의 불확실성은 사후 분포에서 샘플링한 2개 이상의 모델 파라미터에 의해 얻어진 예측 확률 값들의 분산 [4]
 - 학습 데이터 분포에서 멀리 벗어난 데이터일수록 높은 값을 가진 불확실성이 산출되며, 모델이 일관되지 않은 예측을 진행했음을 의미함 [4,6]



Safe SSL Against O.O.D Unlabeled Data

3. Safe Uncertainty-based Consistency Training (SafeUC, 2022+년 Information Sciences Accept)

DMQA 연구실에서 진행한 베이지안 인공지능 세미나

- <http://dmqa.korea.ac.kr/activity/seminar>

종료 Reluctance for Consistent Training
Robust Architecture
Robustness against adversarial attacks and OOD data is a key requirement for deep learning models in real-world applications. This seminar introduces a novel framework for robust training, which leverages the concept of "reluctance" to prevent the model from overfitting to the training data and improve its generalization ability.

Uncertainty Quantification in Deep Learn

발표자: 이지운

📅 2022년 1월 28일
🕒 오후 1시 ~
📺 온라인 비디오 시청 (YouTube)

세미나 정보 보기 →

종료 2019/02/28 DMQA Seminar
Understanding Uncertainty and Bayesian Convolutional Neural Networks
이민정

Understanding Uncertainty and Bayesian

발표자: 이민정

📅 2019년 2월 28일
🕒 오후 1시 30분 ~
📍 고려대학교 신공학관 218호

세미나 정보 보기 →

종료 Bayesian Deep Learning
for Safe AI
Jiyeon Lee
April 24, 2020

Bayesian Deep Learning for Safe AI

발표자: 이지운

📅 2020년 4월 24일
🕒 오후 1시 ~
📍 화상 프로그램 이용 (Zoom)

세미나 정보 보기 →

Safe SSL Against O.O.D Unlabeled Data

3. Safe Uncertainty-based Consistency Training (SafeUC, 2022+년 Information Sciences Accept)

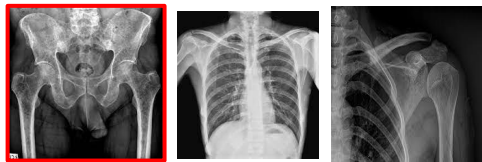
베이지안 인공지능망이 O.O.D Unlabeled Data의 악영향을 어떻게 줄일 수 있을까?
(불확실성)

Safe SSL Against O.O.D Unlabeled Data

3. Safe Uncertainty-based Consistency Training (SafeUC, 2022+년 Information Sciences Accept)

베이지안 인공지능망이 O.O.D Unlabeled Data의 악영향을 어떻게 줄일 수 있을까?
(불확실성)

〈 Unlabeled Data 〉

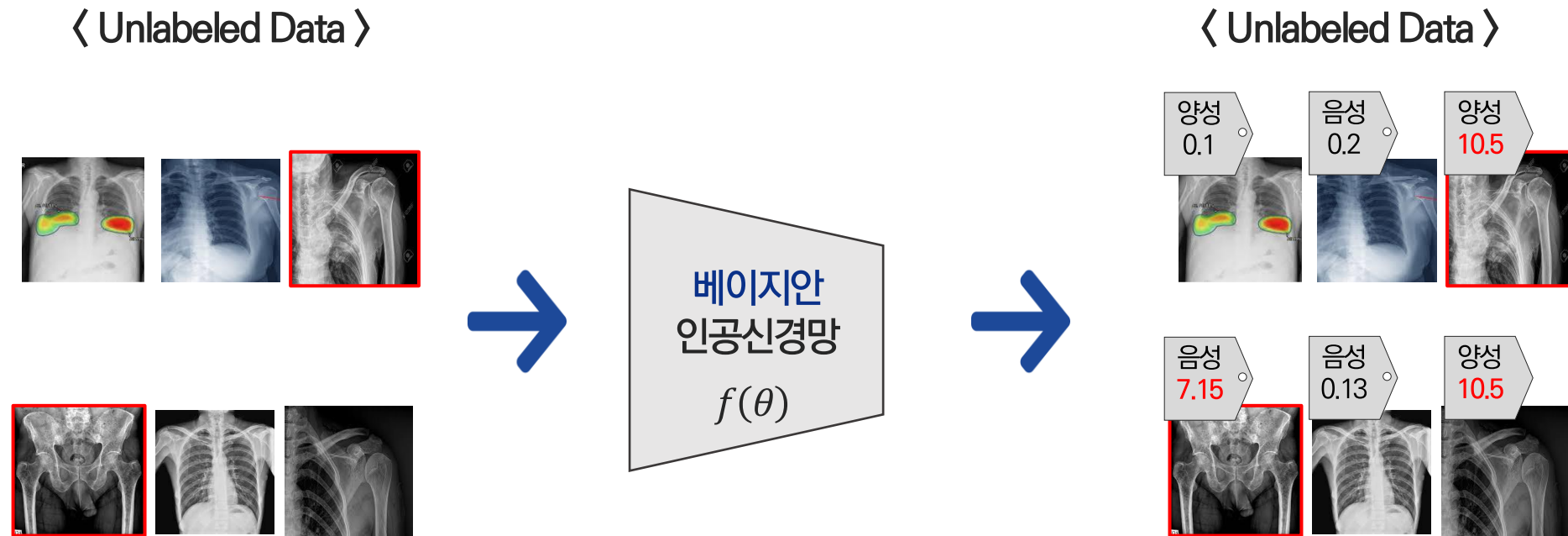


베이지안
인공지능망
 $f(\theta)$

Safe SSL Against O.O.D Unlabeled Data

3. Safe Uncertainty-based Consistency Training (SafeUC, 2022+년 Information Sciences Accept)

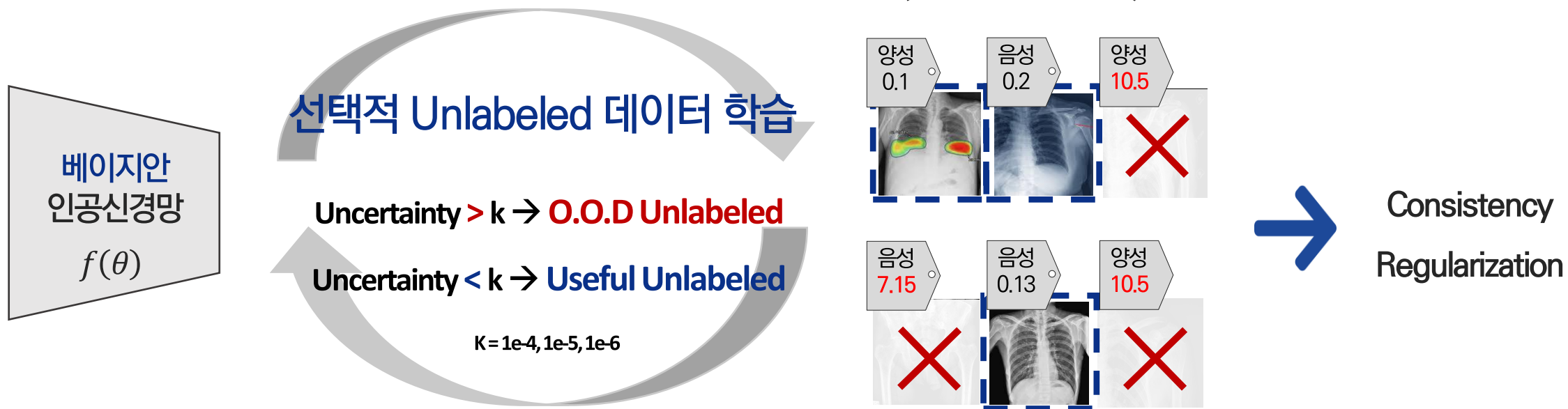
베이지안 인공지능망이 O.O.D Unlabeled Data의 악영향을 어떻게 줄일 수 있을까?
(불확실성)



Safe SSL Against O.O.D Unlabeled Data

3. Safe Uncertainty-based Consistency Training (SafeUC, 2022+년 Information Sciences Accept)

베이지안 인공지능망이 O.O.D Unlabeled Data의 악영향을 어떻게 줄일 수 있을까?
(불확실성)



Safe SSL Against O.O.D Unlabeled Data

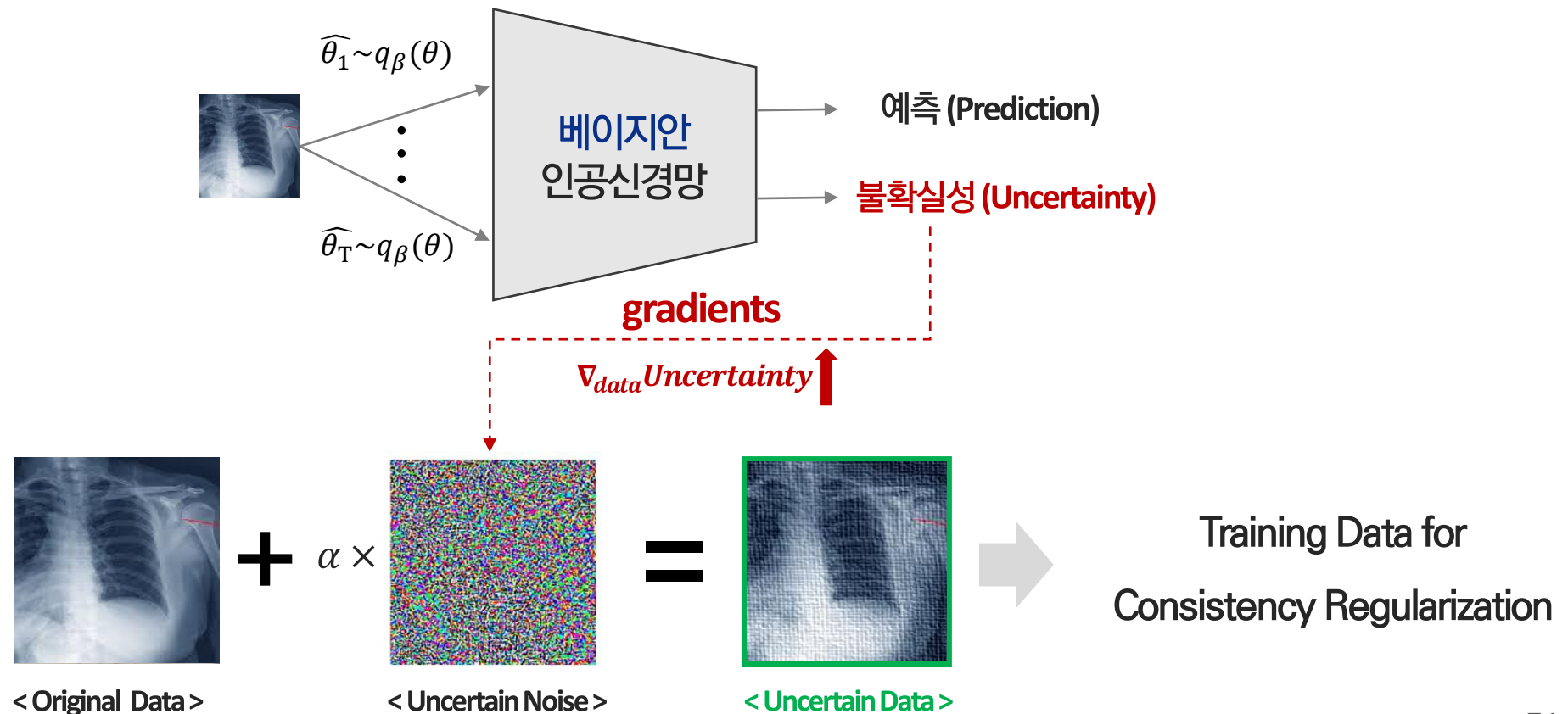
3. Safe Uncertainty-based Consistency Training (SafeUC, 2022+년 Information Sciences Accept)

베이지안 인공지능망이 선택된 Unlabeled Data와 Labeled Data를 어떻게 활용할까?

Safe SSL Against O.O.D Unlabeled Data

3. Safe Uncertainty-based Consistency Training (SafeUC, 2022+년 Information Sciences Accept)

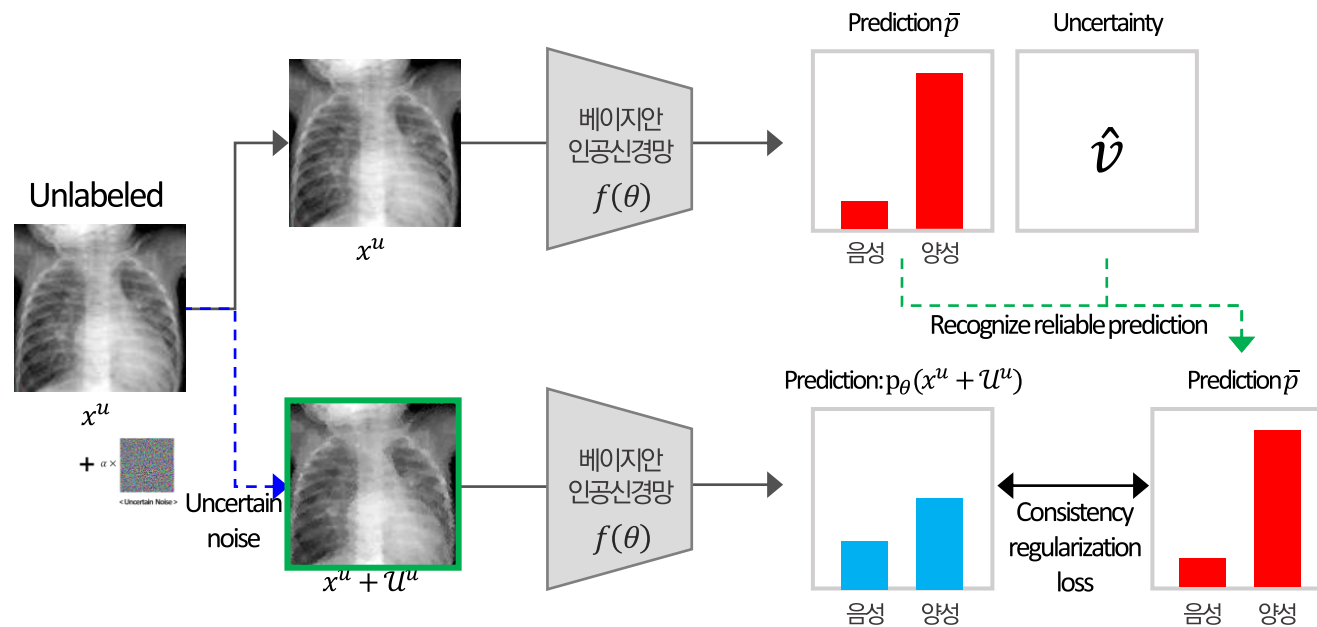
베이지안 인공지능망이 선택된 Unlabeled Data와 Labeled Data를 어떻게 활용할까?



Safe SSL Against O.O.D Unlabeled Data

3. Safe Uncertainty-based Consistency Training (SafeUC, 2022+년 Information Sciences Accept)

베이지안 인공지능망이 선택된 Unlabeled Data와 Labeled Data를 어떻게 활용할까?



Safe SSL Against O.O.D Unlabeled Data

3. Safe Uncertainty-based Consistency Training (SafeUC, 2022+년 Information Sciences Accept)

- ❖ SafeUC는 베이지안 불확실성 기반 선택적 Unlabeled Data 학습 및 데이터 변형 기법을 통해, 안전한 준지도학습을 수행함
 - SafeUC는 UASD 방법론을 포함한 모든 준지도학습보다 우수한 성능을 보이며, 제안 방법론의 안전성을 입증함
 - Unlabeled 데이터 내에 O.O.D 데이터의 존재 비율(%)을 다양하게 조절해가며 여러 실험을 진행함

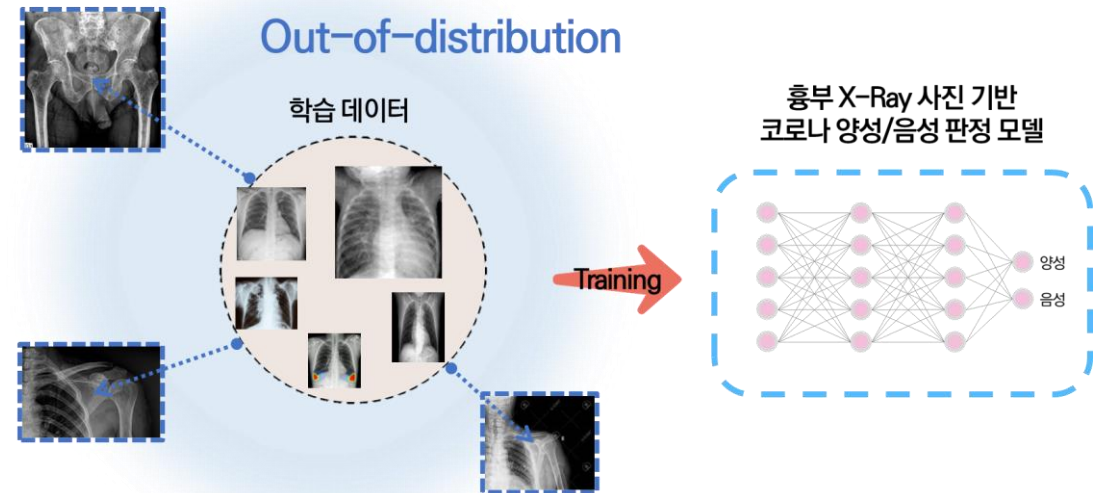
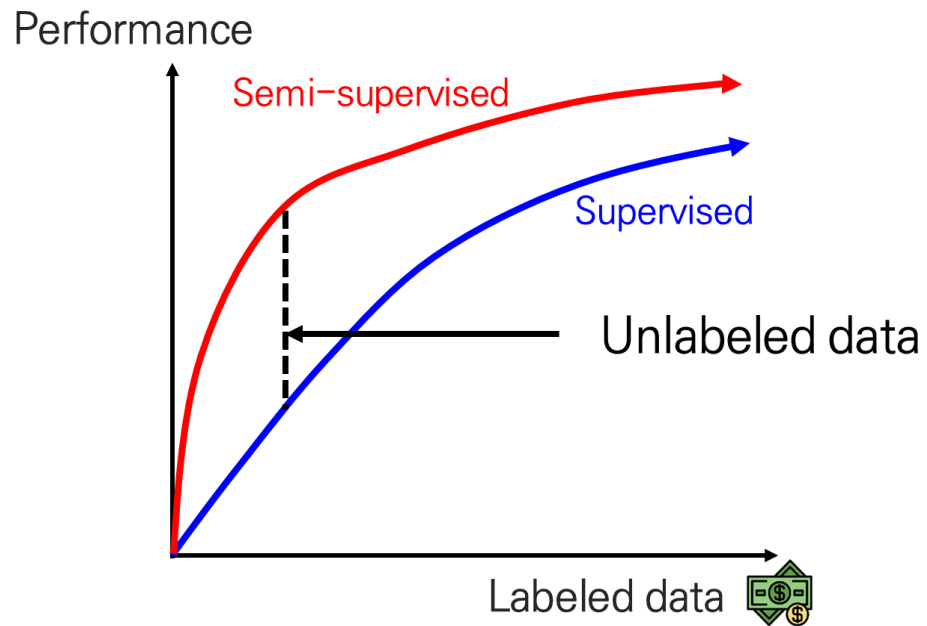
Method	CIFAR-10				
	0%	25%	50%	75%	100%
Baseline	76.04 (0.69)				
Pseudo-Labels	78.43 (0.86)	75.93 (0.75)	74.73 (0.96)	74.09 (1.61)	73.51 (0.81)
Mean-Teacher	78.28 (0.98)	75.83 (1.24)	75.19 (1.34)	74.08 (1.31)	73.44 (0.35)
VAT	79.33 (1.97)	77.12 (0.67)	76.43 (0.75)	74.25 (0.69)	73.97 (0.63)
SWA	79.17 (0.31)	76.79 (0.75)	76.63 (0.56)	76.1 (0.83)	75.59 (0.49)
UASD	79.61 (0.49)	78.53 (0.67)	78.14 (0.64)	77.57 (0.43)	77.55 (0.79)
SafeUC (proposed)	80.3 (0.56)	79.77 (1.03)	78.51 (0.30)	78.67 (0.59)	77.77 (0.62)

Method	SVHN		
	0%	50%	100%
Baseline	84.11 (3.16)		
Pseudo-Labels	84.54 (2.14)	80.39 (2.58)	76.72 (1.94)
Mean-Teacher	85.76 (0.81)	82.65 (3.00)	80.67 (2.46)
VAT	86.28 (1.75)	81.09 (1.14)	79.26 (3.22)
SWA	85.30 (1.67)	83.24 (1.43)	82.89 (1.38)
UASD	82.54 (2.39)	81.06 (2.50)	82.39 (2.38)
SafeUC (proposed)	87.39 (1.25)	85.09 (1.68)	83.26 (1.40)

4. Conclusions

Conclusions

- ❖ 금일 세미나에서는 Semi-Supervised Learning과 Out-of-distribution Unlabeled Data의 개념을 설명함.



Conclusions

- ❖ 하지만, 현실적으로, 대부분의 Unlabeled Dataset 내에는 Labeled Data의 학습 데이터셋 분포에서 벗어난 Out-of-distribution Data가 포함되어 있을 확률이 비일비재함. 이러한 상황에서도 지도학습 대비 우수한 성능을 낼 수 있는 알고리즘이 필요.
- ❖ UASD, DS3L, SafeUC가 어떻게 Out-of-distribution Unlabeled Data의 악영향을 줄였는지, 본 세미나에서 다룸.

Case. 2

Labeled data



Unlabeled data



Unlabeled 데이터는 사람의 검수를 받지 않고 수집되기 때문에
Labeled 데이터셋과 무관한 O.O.D 데이터일 확률이 높음

Safe SSL Against O.O.D Unlabeled Data

= Unlabeled 데이터 내의 O.O.D 데이터 존재와 상관 없이, 지도학습 대비 우수한 성능을 낼 수 있는

Thank you.