Safe Deep Semi-Supervised Learning for Unseen-Class Unlabeled Data

School of Industrial and Management Engineering, Korea University

Jinsoo Bae

KOREA UNIVERSITY

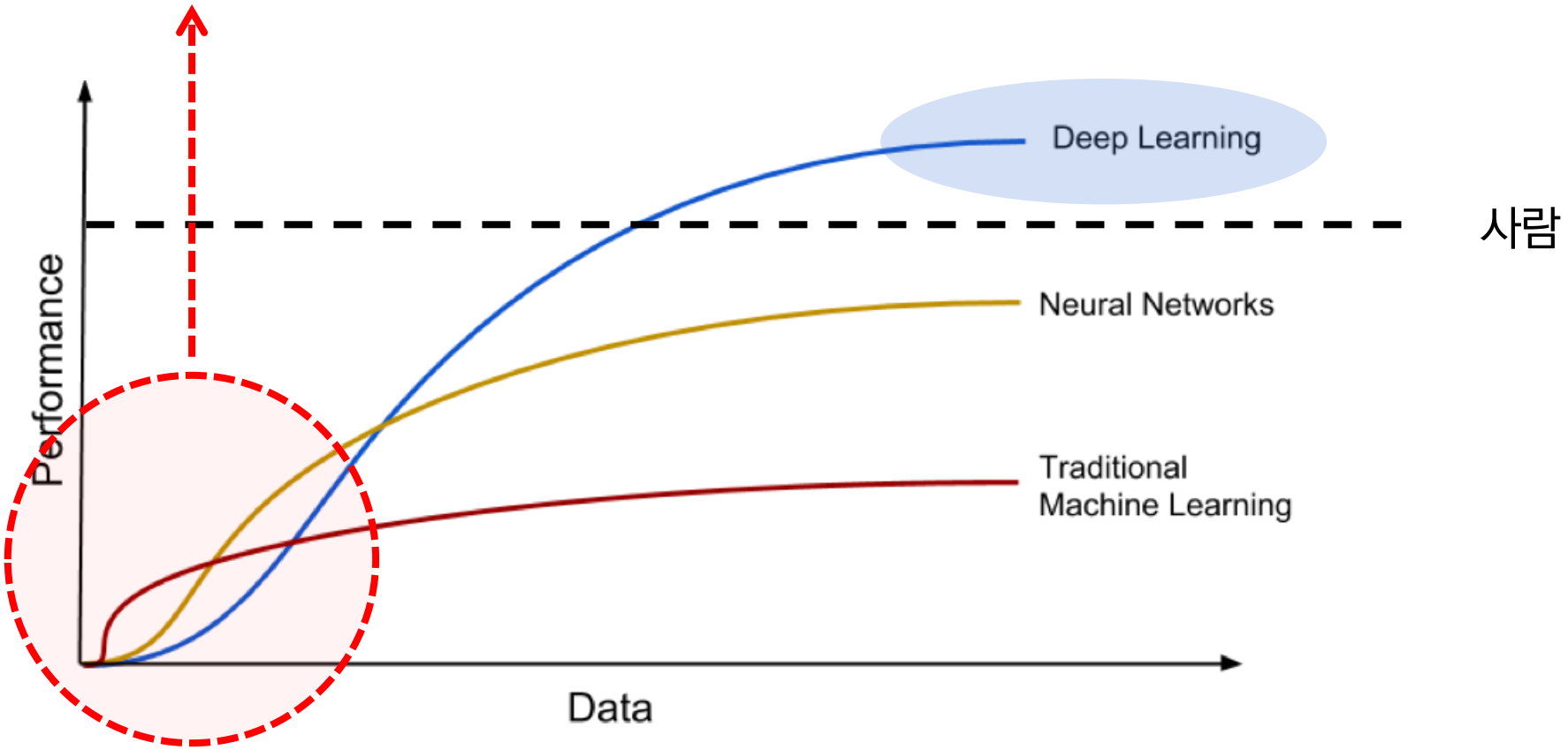
DMQA hcai Human-Centered Artificial Intelligence Center

Contents

Background

❖ **Deep Semi-Supervised Learning (SSL)**

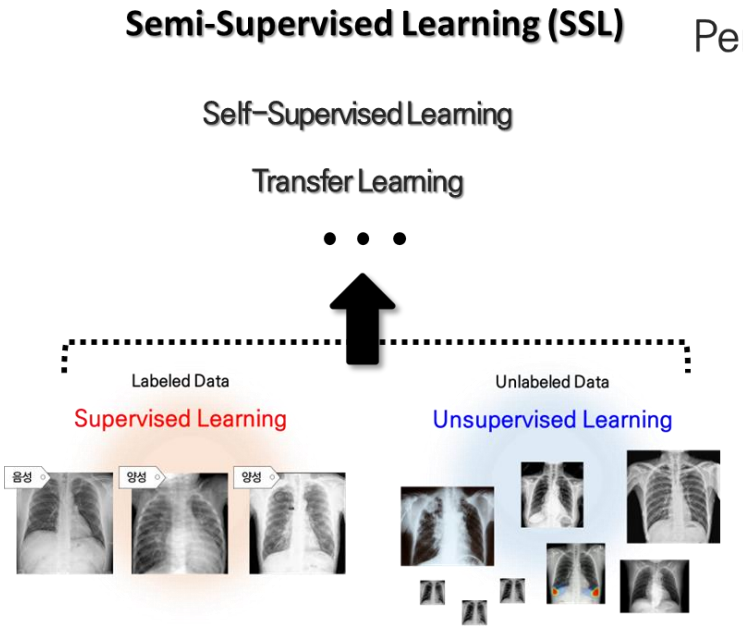
- Labeled 데이터의 개수가 부족할 때 성능이 떨어지는 딥러닝을 보완하고자 등장한 연구

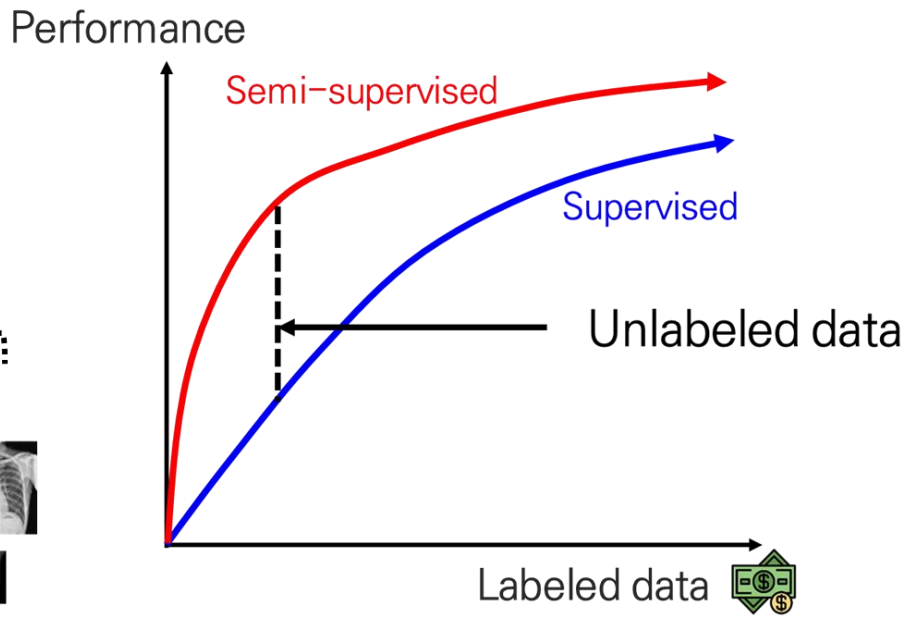


Background

❖ Deep Semi-Supervised Learning (SSL)

- 상대적으로 수집 비용이 저렴한 Unlabeled 데이터를 활용하는 방법론
- 결론적으로, Labeled 데이터만을 사용했을 때보다 높은 성능을 보이며 우수성이 입증 됨





Background

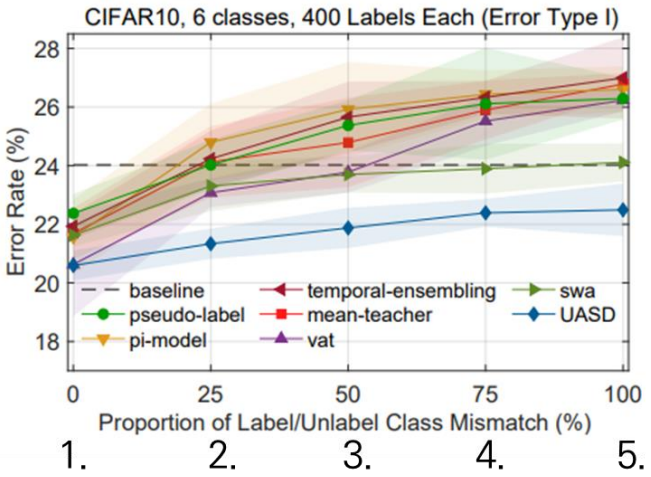
❖ Class-Mismatched Unlabeled와 SSL이 무슨 관계를 갖고 있을까?

- Google Brain 연구원들이 NeuraIPS에 게재한 최신 논문[1]에 따르면, Class-mismatched Unlabeled 데이터는 SSL 학습에 부정적인 영향을 끼치는 것이 확실함

  * Class-mismatched Unlabeled Data: Labeled 데이터셋에서 관측된 적이 없는 클래스를 가진 Unlabeled 데이터

	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%

Class-Mismatched Unlabeled Data



[1] Oliver, A., Odena, A., Raffel, C., Cubuk, E. D., & Goodfellow, I. J. (2018). Realistic evaluation of deep semi-supervised learning algorithms. 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada.

Background

❖ **"Safe Deep Semi-Supervised Learning for Unseen-Class Unlabeled Data"**

- 2020년 37회 PMLR 학회에 게재된 논문으로, 86회 인용 됨 (2022.09.30 기준)
- Labeled 데이터와 Unlabeled 데이터의 클래스 분포와 불일치할 때의 Semi-Supervised Learning (SSL) 방법론 연구

* 기존 대부분의 SSL 연구들은 Unlabeled 데이터와 Labeled 데이터의 클래스 분포가 정확히 일치하는 경우만 집중하였기에, 유의미한 연구

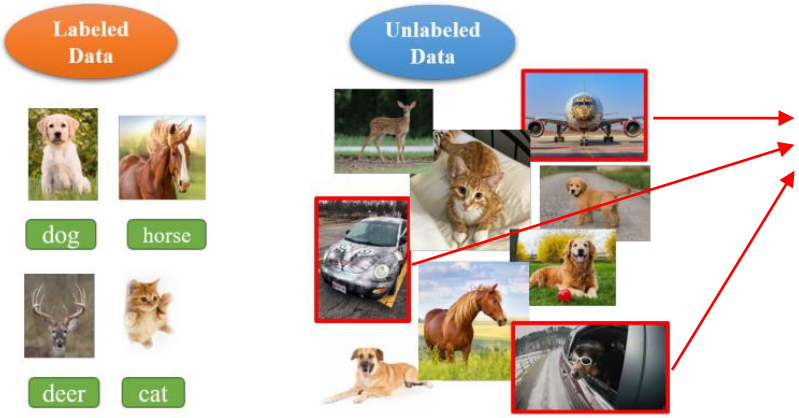


Figure 1. One example of class distribution mismatch. Unlabeled data contains classes that are not seen in the labeled data (indicated with red bounding boxes).

Labeled 데이터에서 관측된 적 없었던
클래스를 가진 Unlabeled 데이터

↓

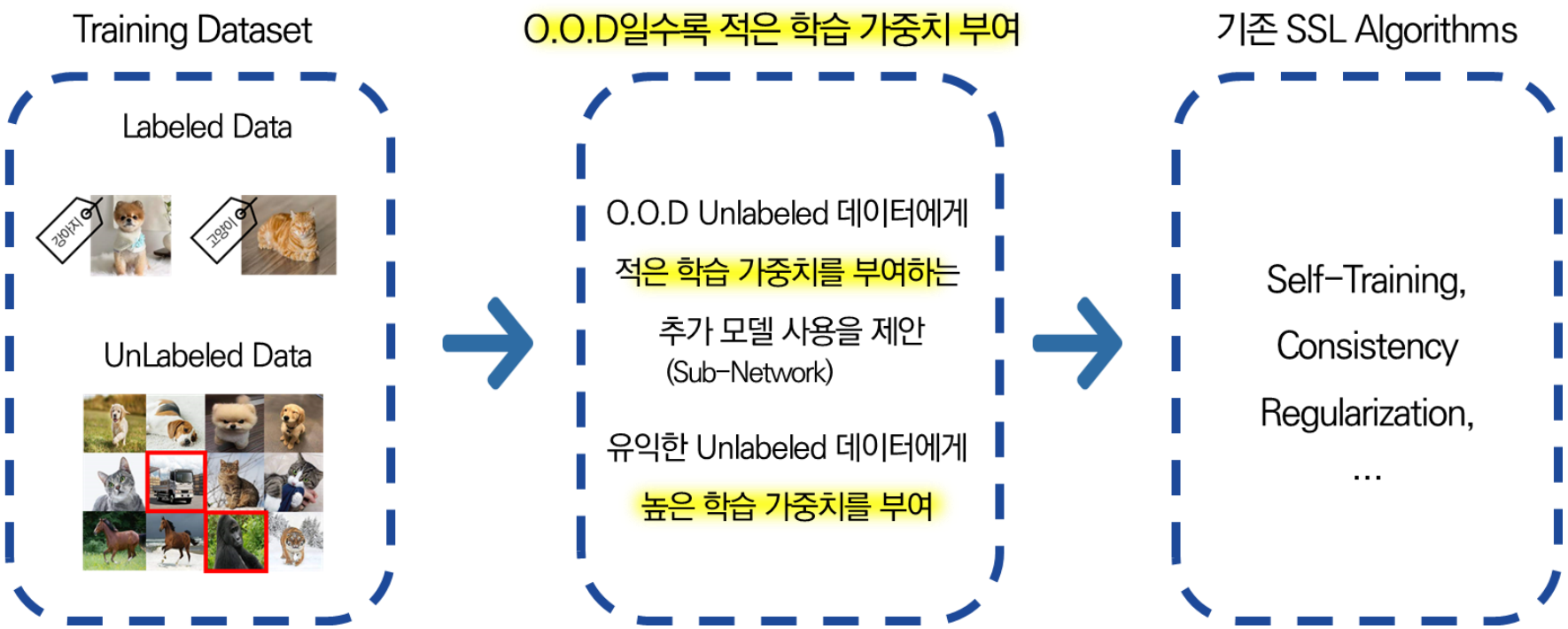
SSL의 학습 성능을 떨어트리기 때문에
유의하여 SSL을 학습시켜야 함

Proposed Method (DS3L)

Overall Concept

❖ 어떻게 SSL에서 Class-mismatched Unlabeled 데이터의 악영향을 줄일 수 있을까?

- Unlabeled 데이터를 활용하는 방식은 기존 SSL과 동일하나,
  - ✓ Class-mismatched Unlabeled 데이터에게 적은 양의 학습 가중치를 부여하는 방식으로 접근
  - ✓ Class-matched Unlabeled 데이터에게는 높은 양의 학습 가중치를 부여하는 방식으로 접근

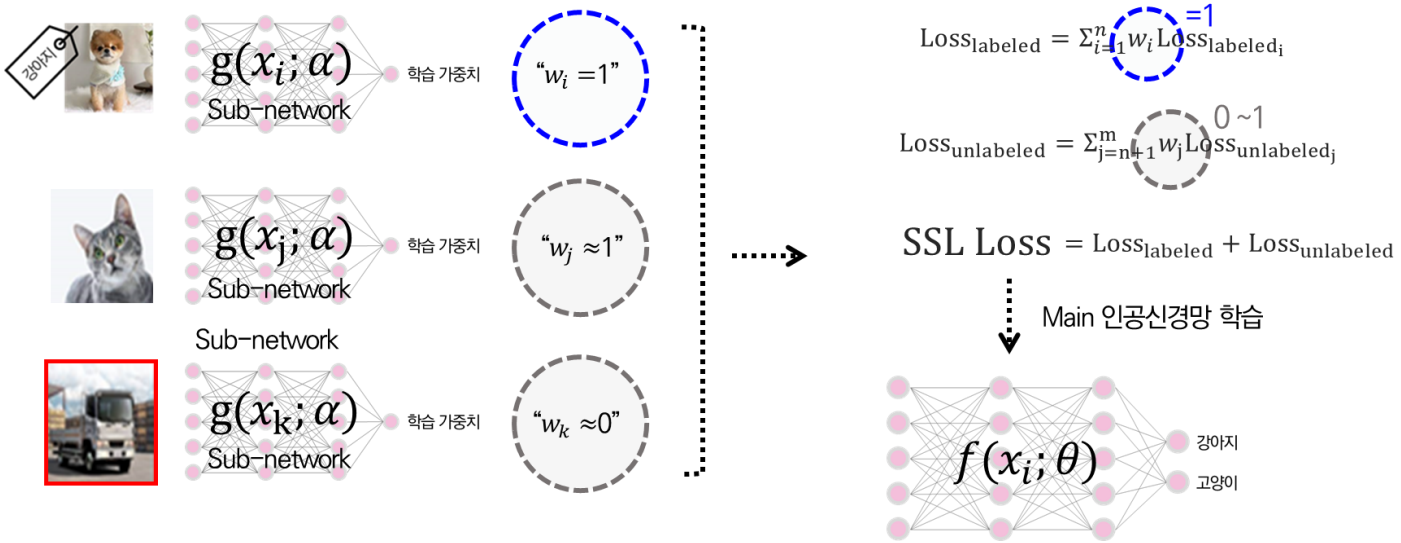


Proposed Method (DS3L)

Main Network(f(x;θ))와 Sub Network(g(x;α)) 두 가지 예측 모델을 사용함

❖ Class Mismatched Unlabeled 데이터에게 적은 학습 가중치를 제공할 수 있는 별도의 추가 모델(Sub-network, g(x; α))제안

- 0과 1 사이의 값 내에서, Labeled 데이터에게는 학습 가중치 1을 부여하고, Matched Unlabeled 데이터에게는 1에 가까운 학습 가중치를 부여하고, Mismatched Unlabeled 데이터에게는 0에 가까운 학습 가중치를 부여



Proposed Method (DS3L)

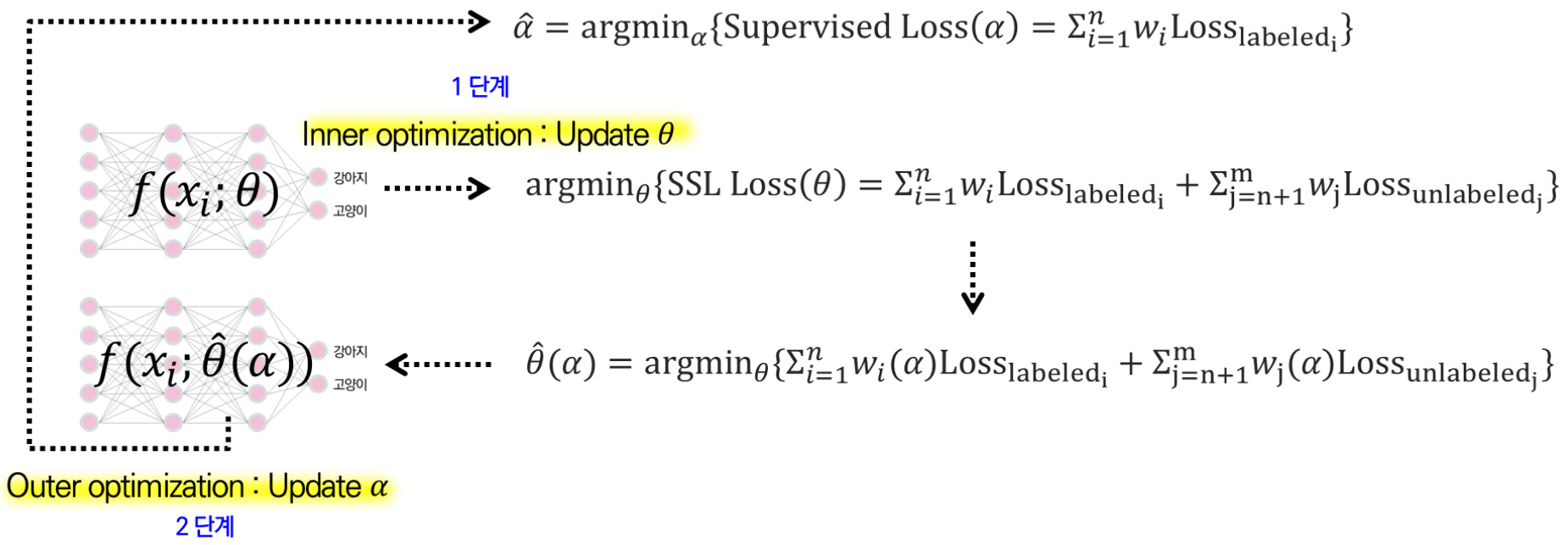
Main Network($f(x;\theta)$)와 Sub Network($g(x;\alpha)$) 두 가지 예측 모델을 사용함

❖ 특정 데이터에게 적은 학습 가중치를 제공할 수 있는 별도의 추가 모델을 어떻게 학습시킬 수 있을까?

- Mismatched Unlabeled 데이터의 악영향은 Main 인공신경망 모델의 Labeled 데이터에 대한 이해도에 부정적인 영향을 끼칠 것으로 가정

  Step 1: Labeled, Unlabeled 데이터에 대한 SSL Loss를 Main Network 모델의 파라미터 θ에 대해 메인 모델 업데이트

  Step 2: 가중치를 출력하는 Sub Network 모델의 파라미터 α에 대해, Labeled 데이터에 대한 Supervised Loss를 α에 대해 서브 모델 업데이트



Proposed Method (DS3L)

Main Network($f(x;\theta)$)와 Sub Network($g(x;\alpha)$) 두 가지 예측 모델을 사용함

❖ 특정 데이터에게 적은 학습 가중치를 제공할 수 있는 별도의 추가 모델을 어떻게 학습시킬 수 있을까?

- Mismatched Unlabeled 데이터의 악영향은 Main 인공신경망 모델의 Labeled 데이터에 대한 이해도에 부정적인 영향을 끼칠 것으로 가정

  Step 1: Labeled, Unlabeled 데이터에 대한 SSL Loss를 Main Network 모델의 파라미터 θ에 대해 메인 모델 업데이트

  Step 2: 가중치를 출력하는 Sub Network 모델의 파라미터 α에 대해, Labeled 데이터에 대한 Supervised Loss를 α에 대해 서브 모델 업데이트

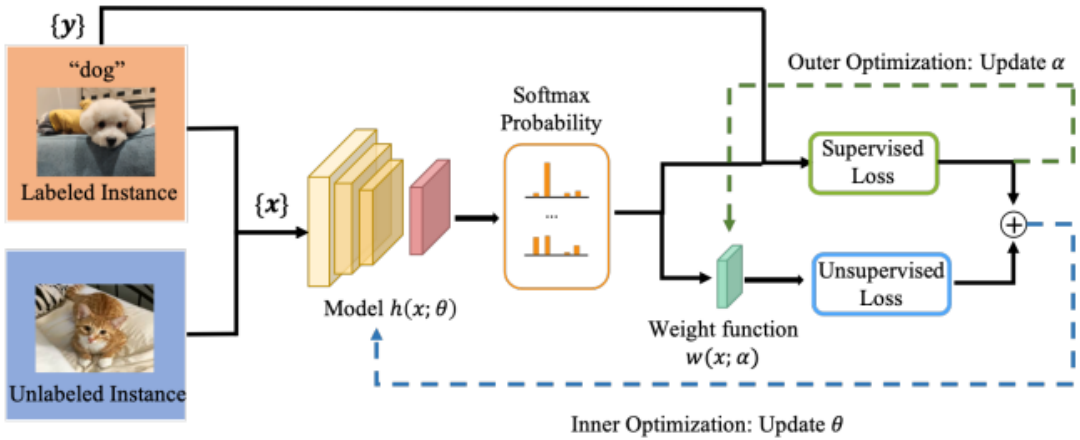


Figure 3. Illustration of the DS^{3}L framework.

Experiment

❖ DS3L은 별도의 서브 모델 학습을 통해 Mismatched 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
  - 학습 데이터 : MNIST, CIFAR-10, 예측 모델 : Wide-ResNet 28-10
  - Mismatched Unlabeled 데이터의 개수(비율)을 변화하며 제안 방법론의 Mismatch에 대한 강건성을 확인

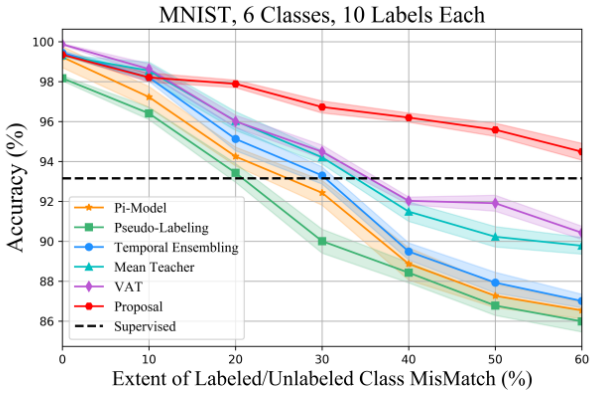


Figure 5. Classification accuracy of compared deep SSL techniques and DS^{3}L on MNIST data set (class $1-6$) with varying class mismatch ratio between labeled and unlabeled data. Shaded regions indicate standard deviation over five runs.

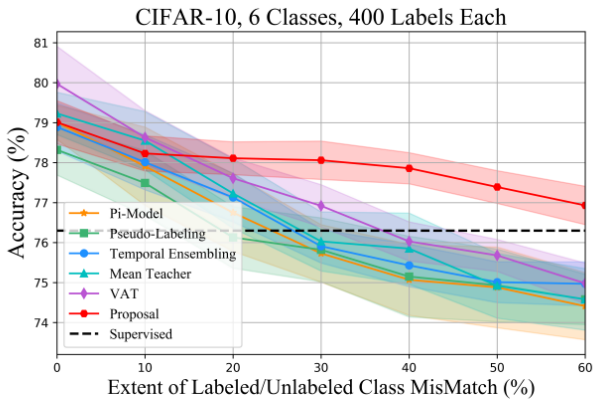


Figure 6. Classification accuracy of compared deep SSL techniques and DS^{3}L on CIFAR-10 data set with varying class mismatch ratio between labeled and unlabeled data. Shaded regions indicate standard deviation over five runs.

Experiment

❖ DS3L은 별도의 서브 모델만 활용하는 방법을 제안했기에, 기존 준지도학습 방법론들에 쉽게 추가 적용될 수 있음

- 기존 준지도학습 방법론들만 사용했을 경우에는, Mismatched Unlabeled 데이터로 인해 정확도의 성능이 쉽게 떨어짐
- 별도의 서브 모델을 활용했을 경우, 기존 방법론들 모두 예측 성능이 쉽게 떨어지지 않는 강건성을 갖게 됨

Sub-Network 모델 활용 O →

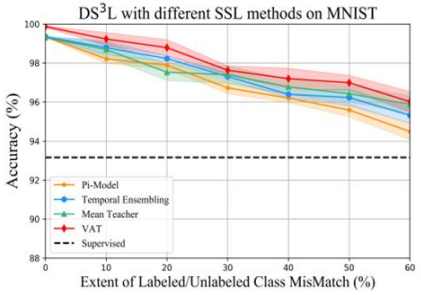


Figure 7. Classification accuracy of DS^{3}L incorporated with four deep SSL methods on MNIST data set.

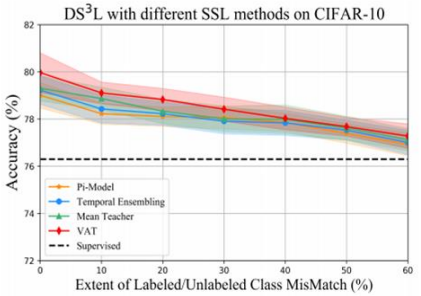


Figure 8. Classification accuracy of DS^{3}L incorporated with four deep SSL methods on CIFAR-10 data set.

Sub-Network 모델 활용 X →

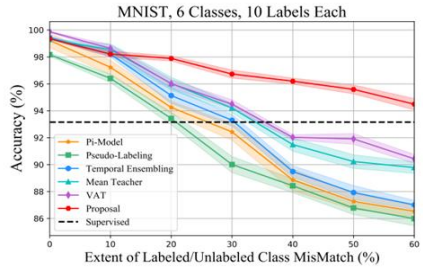


Figure 5. Classification accuracy of compared deep SSL techniques and DS^{3}L on MNIST data set (class $1-6$) with varying class mismatch ratio between labeled and unlabeled data. Shaded regions indicate standard deviation over five runs.

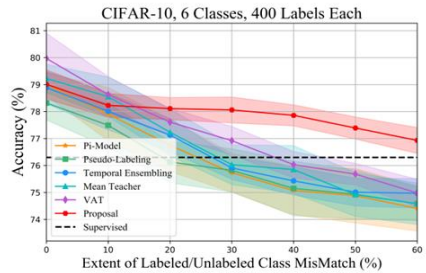


Figure 6. Classification accuracy of compared deep SSL techniques and DS^{3}L on CIFAR-10 data set with varying class mismatch ratio between labeled and unlabeled data. Shaded regions indicate standard deviation over five runs.

Conclusion

- 준지도학습은 Unlabeled 데이터를 활용하여 지도학습 보다 우수한 성능을 보이는 방법론
- 하지만, Unlabeled 데이터 내에 Class-Mismatched 데이터가 존재하는 경우 성능이 하락함
- 위 경우는, 현실적인 상황에서 자주 발생할 수 있는 문제이기에 해결 필요성이 높음
- 본 연구는 Class-Mismatched Unlabeled 데이터의 악영향에 강건한 준지도학습을 연구함
- 별도의 Sub-Network 모델과 2 단계 구조의 Optimization을 통한 준지도학습 방법론을 제시
- 실험적으로도 우수한 성능을 입증하였고, 논문 내에는 수식적으로도 우수성을 증명함


Thank You