

이름 정민규

지도교수 유종빈

### 연구배경

deep learning을 활용한 기술들은 현재 많은 곳에서 사용되고 있다. 적용될 수 있는 분야는 무궁무진하며 이에 따라 다양해진 task들에 대해 새로운 연구들이 진행되고 있다. 따라서 본 연구는 전반적인 latest deep learning theory들을 이해하고 실습해 보는 데 초점을 둔다. 또한 이해한 이론들을 바탕으로 FixMatch method를 적용한 model의 성능을 개선시켜보며 해당 architecture의 동작 방식을 살펴본다.

### 연구 진행 과정

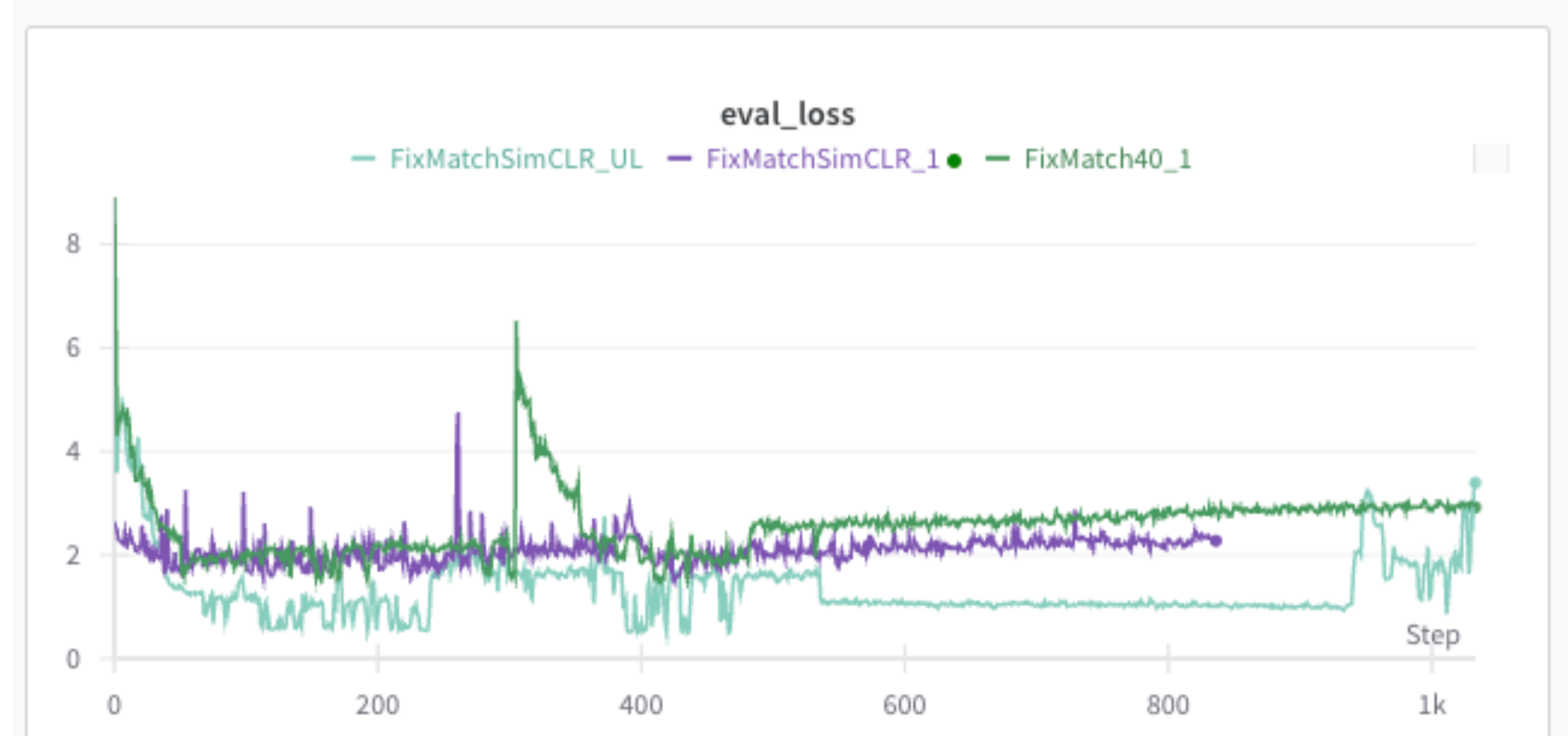
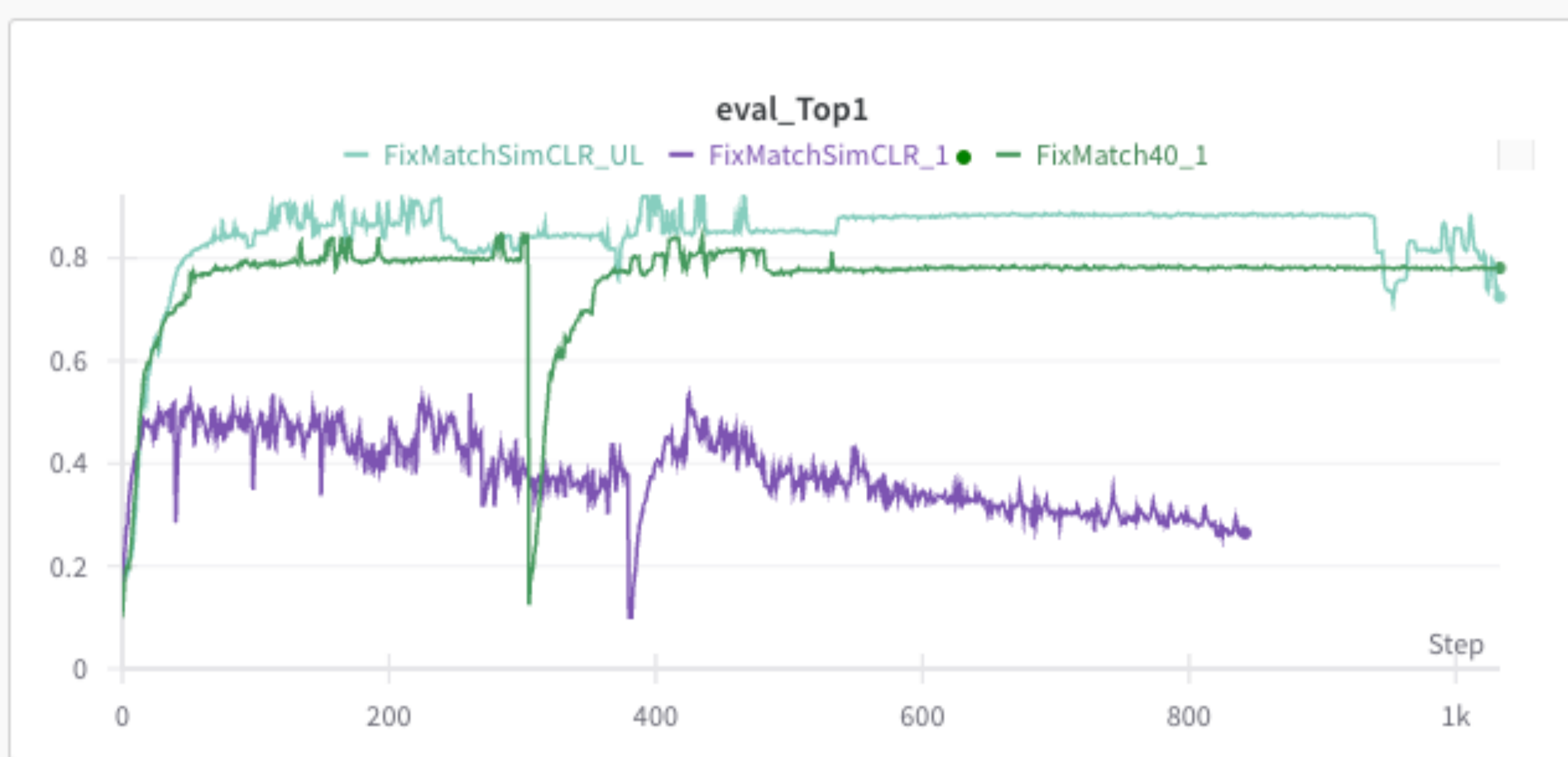
본 연구는 image classification, object detection, network analysis 등의 deep learning 분야에서 선두 주자의 역할을 하고 있는 논문들을 읽고, 각각 실험을 한 뒤 분석하고 이해하는 것이 목표이다. 실험을 해보며 어떻게 하면 해당 방법론의 성능을 개선시킬 수 있을지 고찰해본다. 본 SOFTCON에서는 FixMatch 방법론에 SimCLR 방법론을 적용하여 기존 FixMatch 논문 의 성능을 비교한 것을 중점으로 발표한다. 외의 연구 내용은 자기주도연구 결과보고서에서 정리한다.

### 연구 가설 및 정리

FixMatch 방법론은 consistency regularization과 pseudo-labeling을 결합하여 semi-supervised learning에서 좋은 성능을 보인 방법론이다. 그리고 SimCLR 방법론은 contrastive learning을 위하여 제안된 framework이다. 이에 model이 SimCLR을 통하여 unlabeled data에 대한 representation을 더 학습할 수 있을 것임이 가설이다. 두 방법론을 적용하는 접근은 다음 두 가지로 나눌 수 있었다.

- 1) labeled data에 SimCLR 기법을 적용(L)하고 loss function을 수정 -> labeled data에서의 model이 추출한 feature는 일반화 성능이 좋을 것이고. 또한 labeled data의 정보량을 늘려 model이 더 나은 representation을 학습할 것이다. 하지만 상대적으로 적은 수의 labeled data를 올바르게 학습할 수 있을지에 대하여 살펴볼 필요가 있다.
- 2) unlabeled data에 SimCLR 기법을 적용(U)하고 loss function을 수정. -> 이는 FixMatch에서 생성한 augmented unlabeled data를 사용하려는 접근이다. 이에 따라 학습에서 unlabeled data의 활용성을 높여 model의 representation 능력을 향상시킬 수 있을 것이다. 즉, labeled data에 대해서 기존 FixMatch와 같이 학습을 하고 data의 다양성을 늘리는 방식이다. 그러나 다양해진 data는 같은 분포 내에서 크게 벗어나지 않기 때문에 task에 따라서도 보다 나은 학습을 할 수 있을 것을 기대한다.

### 결과 및 분석



좌측은 O와 L, U에 대한 evaluation top 1 accuracy를 그린 그래프이며 우측은 evaluation loss를 그린 그래프이다. 학습이 진행될 수록 O와 U는 점진적인 성능 향상을 이끌어내고 있지만 L의 경우 성능이 하락하고 있다. L의 경우 labeled data에 대한 불확신성을 loss를 통하여 받고 있고, model이 이전까지 학습했던 것들과 다르게 학습이 진행됨을 알 수 있다. 따라서 학습이 진행됨에 따라 같은 label을 갖는 비슷한 data에 대해서도 unlabeled data가 생성한 pseudo label이 부정확해지며 model의 학습을 방해한다. 반면 U의 경우, labeled data에 대해서 학습을 제대로 진행할 수 있고, SimCLR을 통하여 unlabeled data에 대해 다양한 representation을 학습할 수 있다. 따라서 일반화 능력 또한 함께 학습한 것이다. 특히 학습 시간에 대한 관점에서도 이점을 가진다. 아래의 결과를 보면 L의 경우 동일한 epoch에서 약 이틀에 걸친 학습을 진행하였는데 이는 매 iteration마다 상대적으로 cost가 더 큰 augmentation을 계속해야 하기 때문이다.

| Model      | FixMatch(baseline, O) | At supervised data(L) | At unsupervised data(U) |
|------------|-----------------------|-----------------------|-------------------------|
| Top-1      | 88.1                  | 53.2                  | 92.4                    |
| Train time | 18h 27m               | 2d 1h 9m              | 23h 7m                  |

### 오픈소스 URL

[https://github.com/Gnaroshi/univ\\_ajou-24\\_summer-self\\_directed\\_research](https://github.com/Gnaroshi/univ_ajou-24_summer-self_directed_research)

