

DATA DISTILLATION

TOWARDS OMNI-SUPERVISED LEARNING

REFERENCE

- ▶ Radosavovic, Ilija, et al.
"Data distillation: Towards omni-supervised learning."
Proceedings of the IEEE conference on computer vision
and pattern recognition. 2018.

ABSTRACT

ABSTRACT

- ▶ Omni-supervised learning
 - ▶ Semi-supervised learning 의 특수한 형태
 - ▶ 학습자(신경망)가 가능한 레이블링 된 데이터들에 대해 모두 학습하고
 - ▶ 추가로 인터넷-스케일 출처의 레이블링 되지 않은 데이터들에 대해 학습

ABSTRACT

- ▶ Omni-supervised 학습
 - ▶ 하한: 가지고 있는 레이블링 된 데이터로부터 나올 수 있는 성능
 - ▶ 상한: 모든 데이터가 레이블링 된 상황
 - ▶ 상한에 가까운 SOTA 성능

ABSTRACT

- ▶ Data distillation
 - ▶ 데이터 증류
 - ▶ Omni-supervised 학습을 위한 방법
- ▶ 레이블링 되지 않은 데이터의 다양한 변형을
 - ▶ 하나의 모델로 예측해서
 - ▶ 그 결과를 앙상블
 - ▶ 레이블을 만듦

INTRODUCTION

DATA DISTILLATION

- ▶ 데이터로부터의 지식 증류
 - ▶ 저자들은 모델 지식 증류로부터 아이디어를 얻음

DATA DISTILLATION

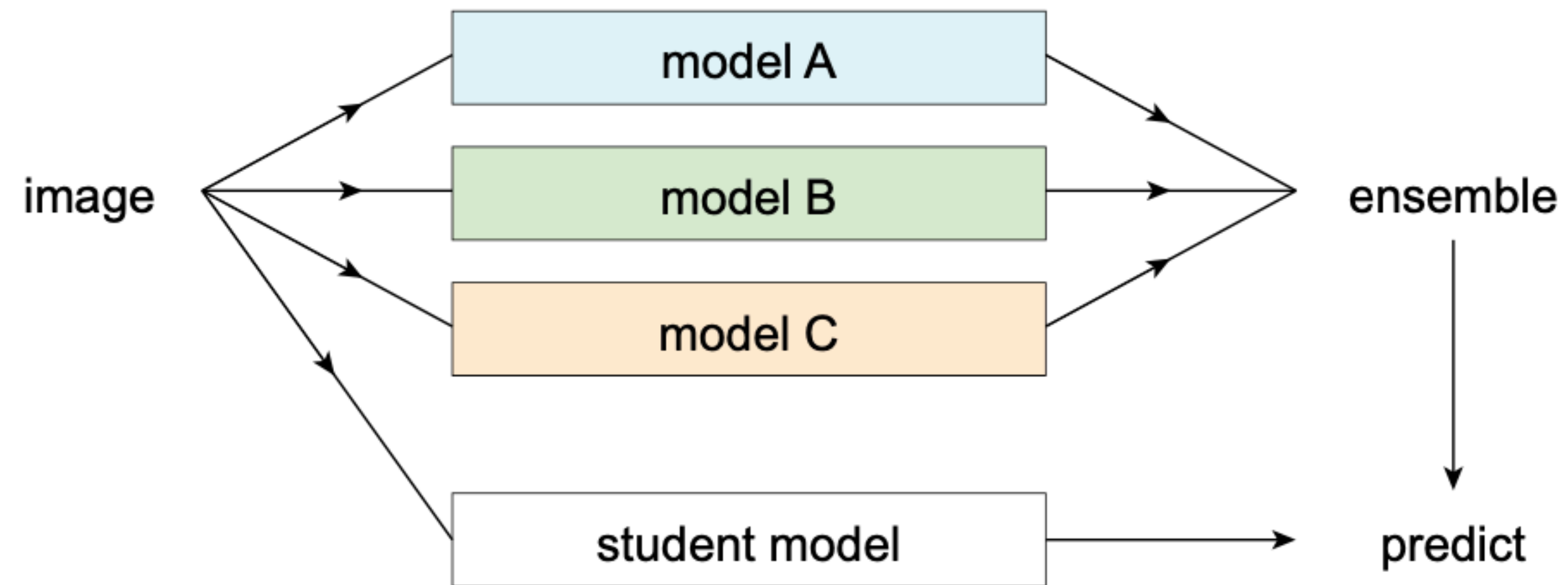
- ▶ Unlabeled 데이터에 주석(레이블)을 생성
 - ▶ 많은 양의 labeled 데이터로부터 학습한 모델을 이용해서
- ▶ 이렇게 추가로 생성된 주석으로부터
 - ▶ 데이터를 가지고 모델을 학습시킴
- ▶ 그러나 자기가 만들어낸 데이터를 가지고 학습하는 것이 과연 유의미한가?

DATA DISTILLATION

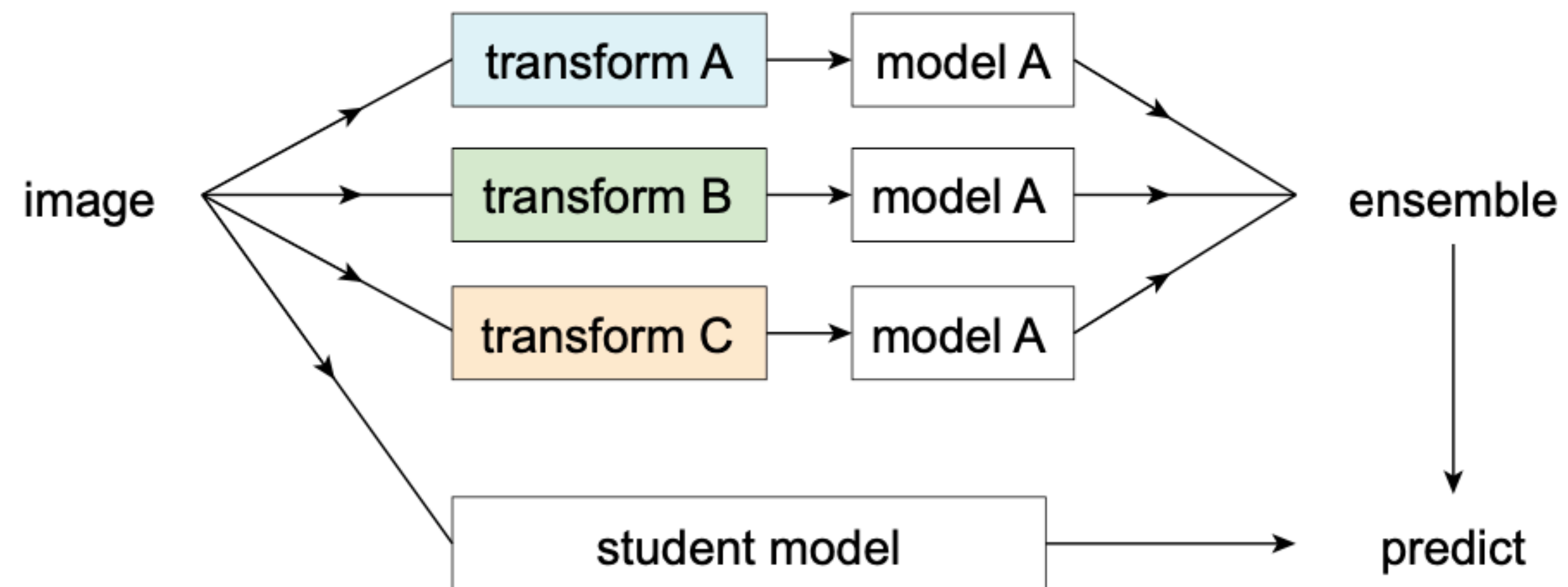
- ▶ 저자들은
 - ▶ 서로 다른 변형(뒤집기, 스케일링)을 거친
 - ▶ Unlabeled 이미지를
 - ▶ 하나의 모델을 가지고 결과들을 내고
 - ▶ 앙상블을 거쳐 레이블링
- ▶ 주로 test time에 한 모델의 정확도를 올리기 위해 사용하는 방법

DATA DISTILLATION

- ▶ Model Distillation
vs.
Data Distillation



Model Distillation



Data Distillation

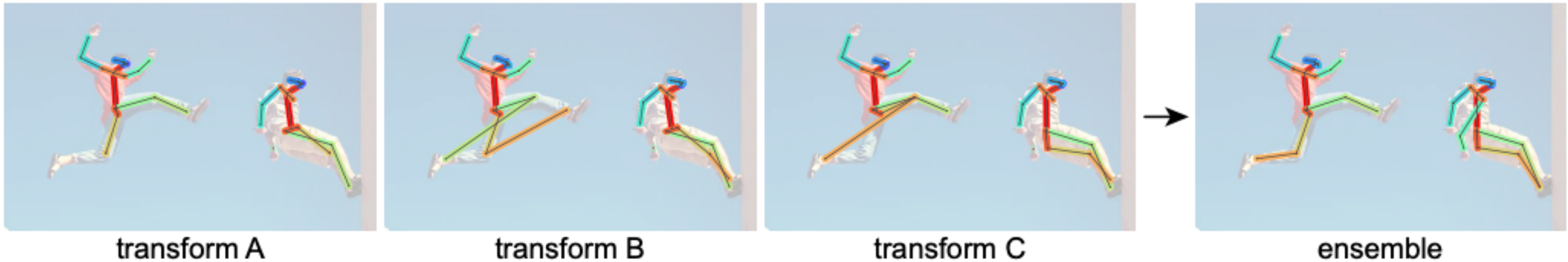
DATA DISTILLATION

DATA DISTILLATION

- ▶ 1. 모델을 (수동으로) 레이블링한 데이터에 대해 학습
- ▶ 2. 학습된 모델을 여러 변형이 가해진 레이블링 되지 않은 데이터에 적용
- ▶ 3. 예측 결과들을 앙상블을 이용해 레이블로 변경
- ▶ 4. 수동 레이블된 데이터와 자동 레이블된 데이터의 합집합으로 모델을 다시 학습시킴

DATA DISTILLATION

- ▶ 다양한 예측 결과를 통합하면
 - ▶ 하나만 사용할 때 보다 좋은 성능을 낼 가능성이 큼



DATA DISTILLATION

- ▶ 통합된 예측이 새 지식을 생성하고
 - ▶ 이 정보를 레이블 생성에 활용하고
 - ▶ 스스로 학습하는 데 쓸 수 있지 않을까?

DATA DISTILLATION

- ▶ 어떻게 통합할 것인가
- ▶ 분류 확률들의 평균
 - ▶ 문제 1: Soft label을 생성하므로 문제가 됨 (레이블이 아님)
 - ▶ 이 경우 Soft label을 다루도록 손실 함수를 수정해야 함
 - ▶ 문제 2: 출력 공간이 object detection이나 human pose 추정 등이면
 - ▶ 평균을 통한 결과가 자연스럽지 못함

DATA DISTILLATION

- ▶ Hard label을 생성하도록 하고
- ▶ 문제에 따른 형태/구조를 다르게 통합해야 함

DATA DISTILLATION

- ▶ 여러 번 **inference** 하기 때문에
 - ▶ **학습**을 여러 번 하는 model distillation 보다
 - ▶ 효율적임

DATA DISTILLATION

- ▶ 지식 증류
 - ▶ 새롭게 생성된 레이블링 되지 않은 데이터에 대한 지식(Hard label)으로
 - ▶ (같거나 다를 수 있는) 학생 모델이 학습
- ▶ 지도 학습에 사용한 데이터와
 - ▶ **자동으로 레이블을 생성한 데이터의 합집합**으로 학습

DATA DISTILLATION

- ▶ 합집합으로 학습하는 것은
 - ▶ 직관적으로 보임
 - ▶ 손실 함수의 수정도 필요 없음

DATA DISTILLATION

- ▶ 각 학습 미니배치는 수동/자동 레이블 데이터의 혼합으로 진행
 - ▶ 각 미니배치가 일부 진실된 레이블을 가지게 강제함으로써
 - ▶ 기울기 추정을 보다 좋게 만들어 줌

EXPERIMENTS

EXPERIMENTS

- ▶ 키폰트 검출
 - ▶ COCO 데이터셋

EXPERIMENTS

- ▶ 모델
 - ▶ Mask R-CNN 사용
 - ▶ ResNet & ResNeXt
 - ▶ Feature Pyramid Networks (FPN)
- ▶ 선생과 학습 모델에 동일한 구조 사용

EXPERIMENTS

- ▶ 데이터 변형
 - ▶ 스케일링
 - ▶ 수평 뒤집기
- ▶ 예측 결과를 통합 가능한 형태로 만들기 위한 변형들

EXPERIMENTS

- ▶ 재학습

- ▶ 원본:생성 = 6:4 비율로 미니배치

EXPERIMENTS

▶ 작은 규모의 데이터에 대한 결과

labeled	unlabeled	AP	AP ₅₀	AP ₇₅	AP _M	AP _L
co-35		54.9	80.5	59.0	50.1	62.8
co-35	co-80	60.2	83.8	65.4	55.2	68.4
co-115		65.1	86.6	70.9	59.9	73.6

▶ 숫자는 데이터의 수(k)

▶ AP: Average Precision

EXPERIMENTS

- ▶ 큰 규모, 유사한 분포의 데이터에 대한 결과
- ▶ co-115 with labels
- ▶ un-120 without labels
- ▶ DD: Data Distillation

backbone	DD	AP	AP ₅₀	AP ₇₅	AP _M	AP _L
ResNet-50		65.1	86.6	70.9	59.9	73.6
ResNet-50	✓	67.1	87.9	73.4	62.2	75.1
ResNet-101		66.1	87.7	71.7	60.5	75.0
ResNet-101	✓	67.8	88.2	73.8	62.8	76.0
ResNeXt-101-32×4		66.8	87.5	73.0	61.6	75.2
ResNeXt-101-32×4	✓	68.7	88.9	75.1	63.9	76.7
ResNeXt-101-64×4		67.3	88.0	73.3	62.2	75.6
ResNeXt-101-64×4	✓	69.1	88.9	75.3	64.1	77.1

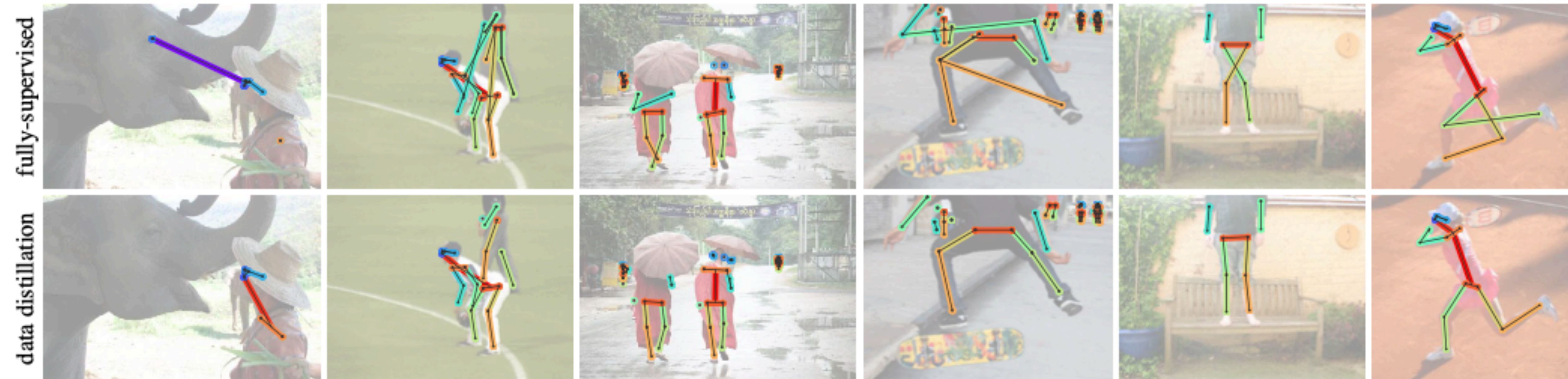
EXPERIMENTS

- ▶ 큰 규모, 다른 분포의 데이터에 대한 결과
- ▶ co-115 with labels
- ▶ slm-120 without labels

backbone	DD	AP	AP ₅₀	AP ₇₅	AP _M	AP _L
ResNet-50		65.1	86.6	70.9	59.9	73.6
ResNet-50	✓	66.6	87.3	72.6	61.6	75.0
ResNet-101		66.1	87.7	71.7	60.5	75.0
ResNet-101	✓	67.5	87.9	73.9	62.4	75.9
ResNeXt-101-32×4		66.8	87.5	73.0	61.6	75.2
ResNeXt-101-32×4	✓	68.0	88.1	74.2	63.1	76.2
ResNeXt-101-64×4		67.3	88.0	73.3	62.2	75.6
ResNeXt-101-64×4	✓	68.5	88.8	74.9	63.7	76.5

EXPERIMENTS

- ▶ 큰 규모, 다른 분포의 데이터에 대한 결과



CONCLUSION

CONCLUSION

- ▶ Omni-supervised 학습
 - ▶ 모든 가능한 지도된 데이터와
 - ▶ 많은 양의 unlabeled 데이터를 이용한 학습
- ▶ 데이터 종류를 통해 달성

CONCLUSION

- ▶ 추후 더 규모가 크고 실용적인 환경에서 실험할 것
 - ▶ Facebook AI Research (FAIR)

DATA DISTILLATION

TOWARDS OMNI-SUPERVISED LEARNING