

# HAKE: Human Activity Knowledge Engine

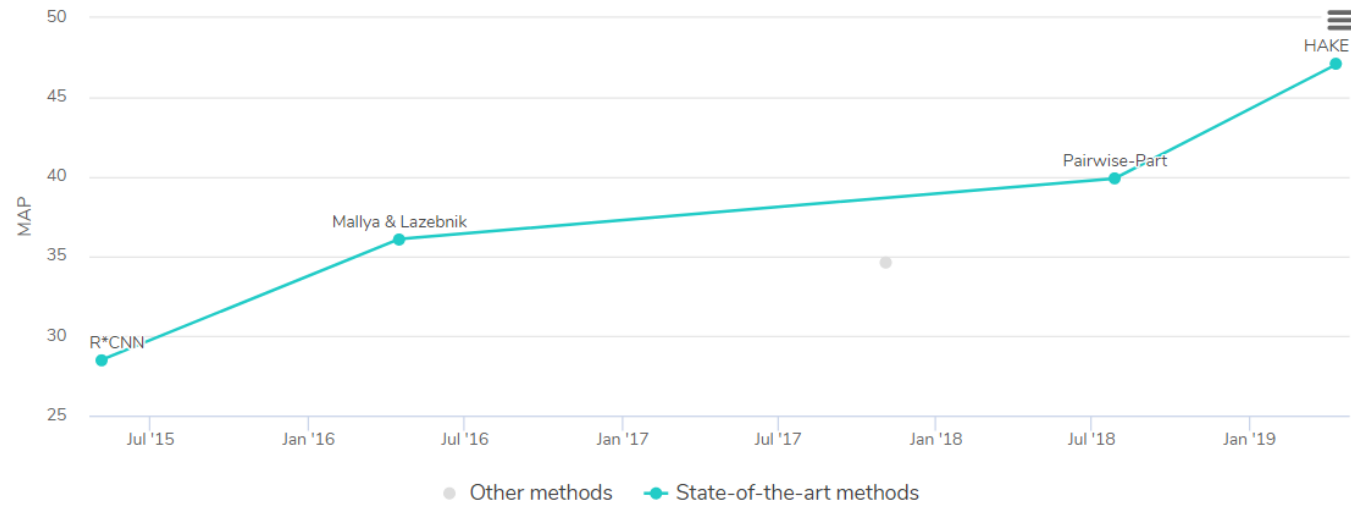
Yong-Lu Li, Liang Xu, Xinpeng Liu, Xijie Huang, Yue Xu, Mingyang Chen, Ze Ma,  
Shiyi Wang, Hao-Shu Fang, Cewu Lu\*

Shanghai Jiao Tong University

CVPR 2019

인공지능 연구실  
석사과정 구자봉

# Human-Object Interaction Detection on HICO



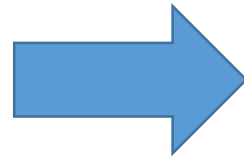
View: All methods

Edit

RANK	METHOD	MAP	EXTRA TRAINING DATA	PAPER TITLE	YEAR	PAPER	CODE
1	HAKE	47.1	✓	HAKE: Human Activity Knowledge Engine	2019		
2	Pairwise-Part	39.9	×	Pairwise Body-Part Attention for Recognizing Human-Object Interactions	2018		
3	Mallya & Lazebnik	36.1	×	Learning Models for Actions and Person-Object Interactions with Transfer to Question Answering	2016		
4	Girdhar & Ramanan	34.6	×	Attentional Pooling for Action Recognition	2017		
5	R*CNN	28.5	×	Contextual Action Recognition with R*CNN	2015		

# 문제 정의 :

## HOI(Human-object interaction)



<human, ride, bicycle>  
<human, sit\_on, bicycle>  
<human, straddle, bicycle>

# 기존 방법의 문제점 제시 :

불균형 한 데이터 분배, 동작 모호성, 복잡한 시각적 패턴과 같은 문제가 아직 남아있다.  
원 이미지에서 직접 인간의 행동을 추출하는 것은 성능이 떨어진다.

One-stage  
Paradigm:



Pixels

CNN

Direct, Brute force

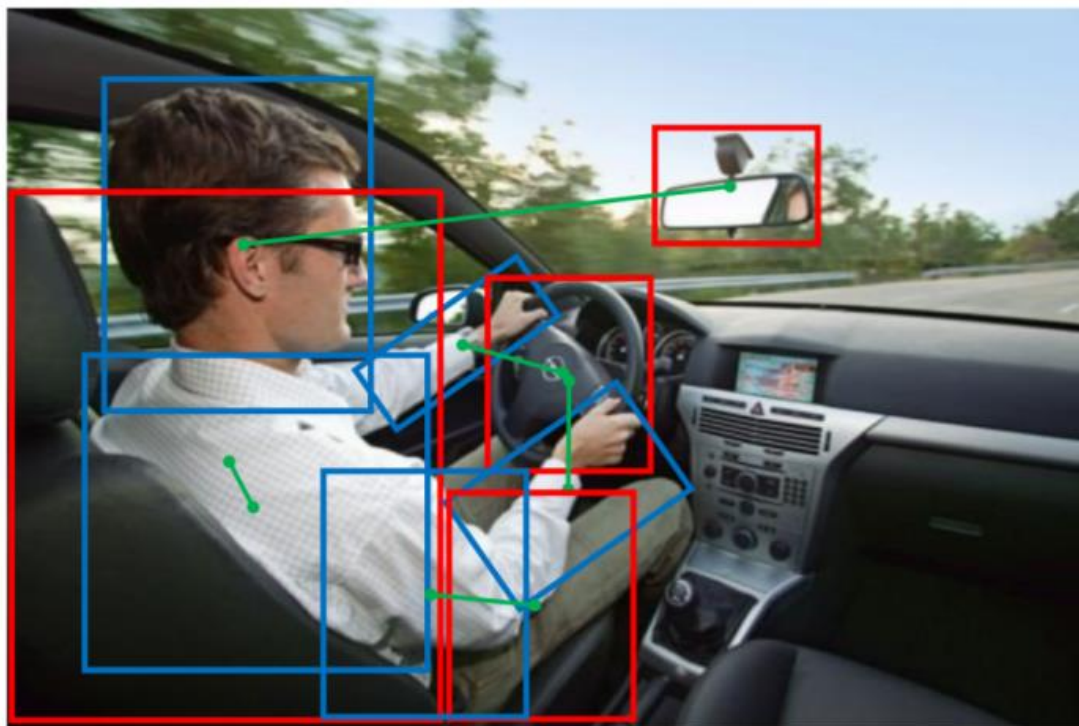
person-kick-football

Instance-level action label

# 동기 :

인간의 행동은 모든 신체 부위의 원자 행동 또는 상태로 구성된다.

- 1) 부분 상태는 동작의 기본 구성 요소이다.
- 2) 부분 상태 인식을 하면 이미지 공간과 시맨틱(의미론적)공간 사이의 간격을 크게 좁힌다.
- 3) 부분 상태 임베딩은 더 나은 해석을 제시한다.(모델의 최종 선택의 이유를 쉽게 알 수 있다.)



head-inspect-rearview

right\_hand-hold-wheel

left\_hand-hold-wheel

torso-lay\_on-chair\_back

hip-sit\_on-chair\_seat

# HAKE: Human Activity Knowledge Engine

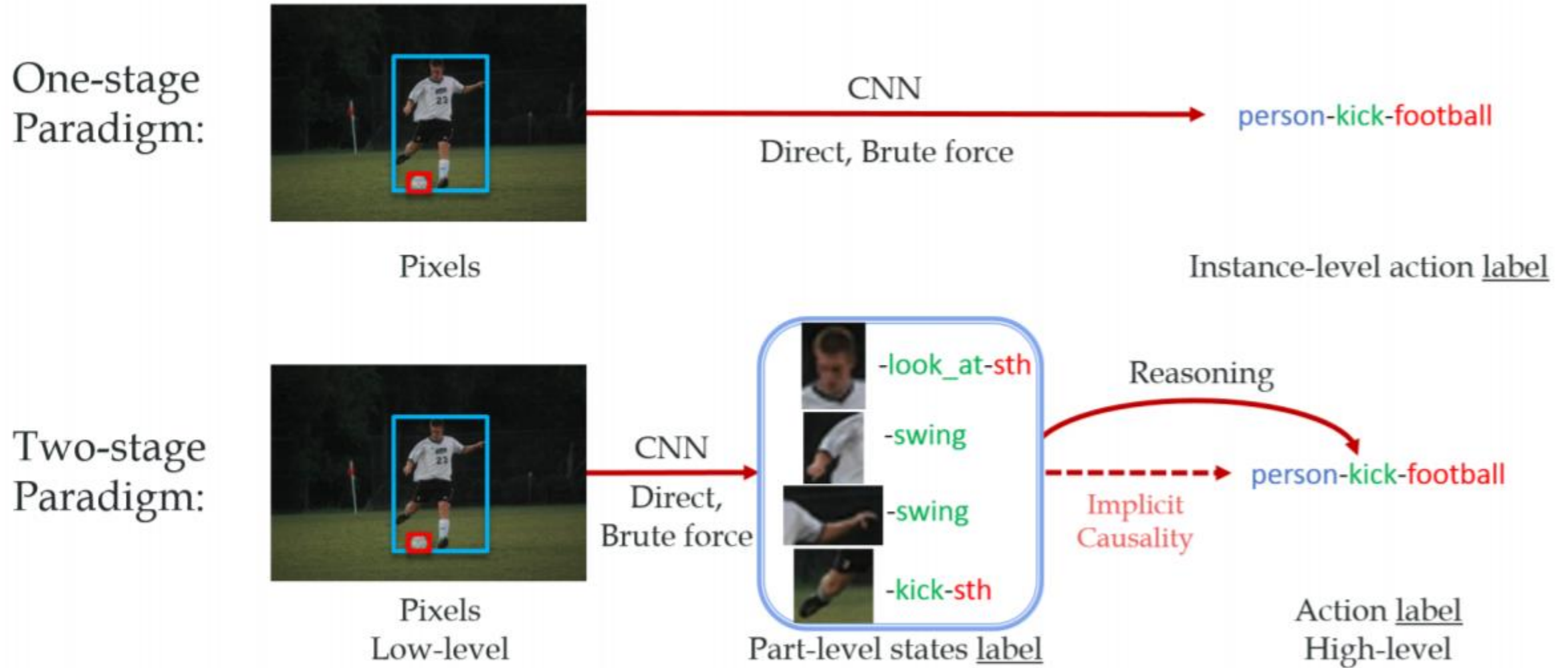
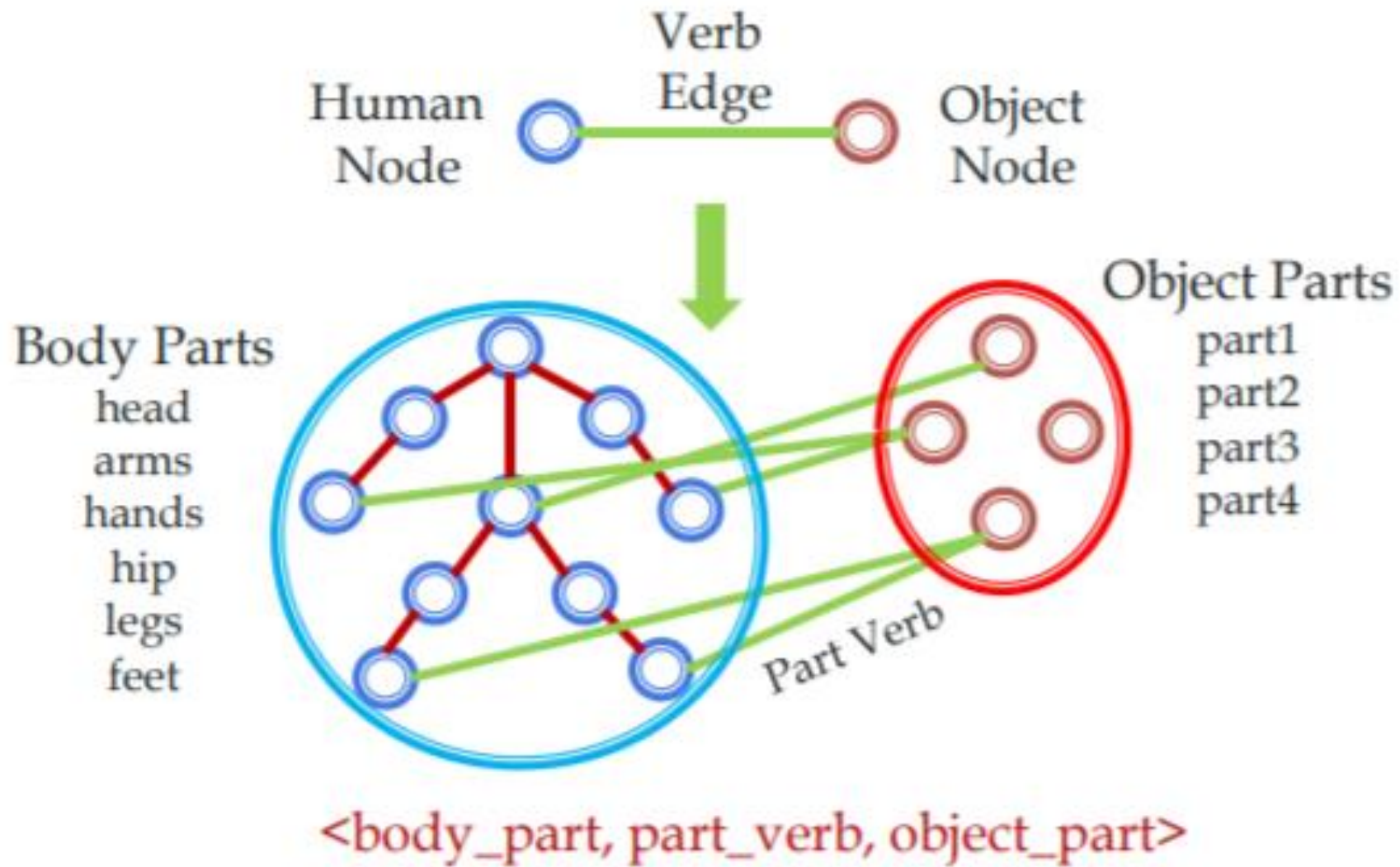


Figure 2. Previous one-stage paradigm and our hierarchical two-stage paradigm.



# HAKE 인간 신체 부위

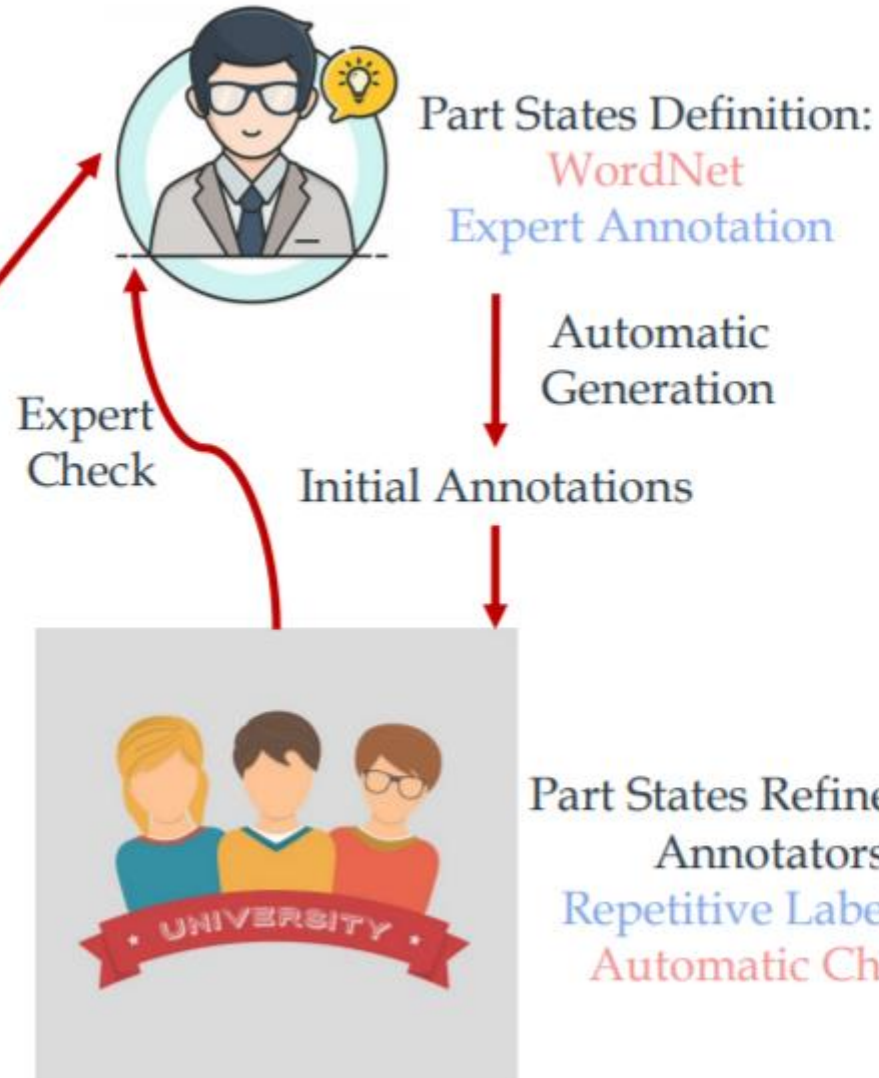


# HAKE 구성

## Existing Activity Datasets with Instance-level Annotations



방대한 양의 데이터셋  
HICO-DET [12], VCOCO [16],  
OpenImage [17], HCVRD  
[22], HICO [13], MPII [1],  
AVA [10]



인간을 10개의 부위로 분해

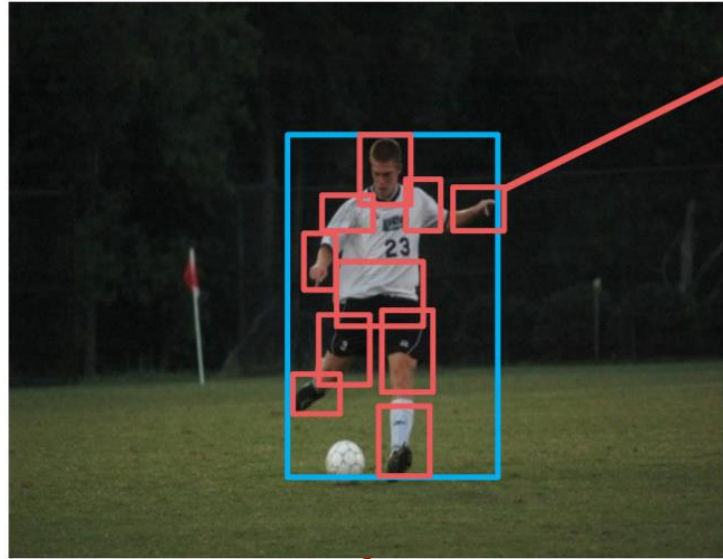
WordNet의 동사를 기반으로  
200개의 부품상태 풀을 구축,  
e.g. "hold", "push", "pick" for  
hands, "listen to", "eat", "talk  
to" for head

부품상태 개선을 위해  
NPMI를 사용하여 최  
종 부품 상태를 선택



# HAKE 구축을 위한 반 자동 라벨링을 통해 학습 데이터 생성

- 1) 기존의 데이터셋에는 인간 및 객체의 바운딩 박스와 그 사이의 인터랙션 라벨이 있음
- 2) 포즈 추정을 사용해 키 포인트와 부분 경계상자를 얻음
- 3) 9명의 전문가를 초대, 154개의 모든 인스턴스 작업을 기준으로 10,000개의 이미지에 주석을 담
- 4) 모든 이미지의 초기 부품 상태 생성
- 5) 여러 전문가 및 어노테이터들이 반복 확인 및 여러 주석을 담으로써 품질을 향상 시킴



# HAKE 계층적 Paradigm

Activity2Vec을 통한  
지식 추출을 통한  
부분 상태인식



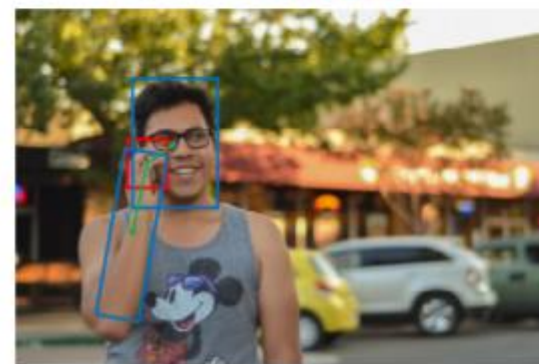
파트 상태에서 인스턴스 활동을 추론



head-drinks\_with-bottleneck  
right\_hand-hold-bottle\_body



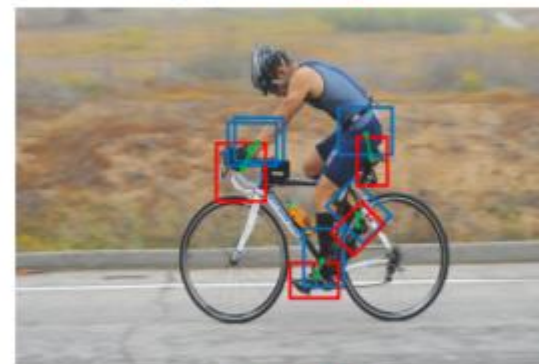
human-drink\_with-bottle



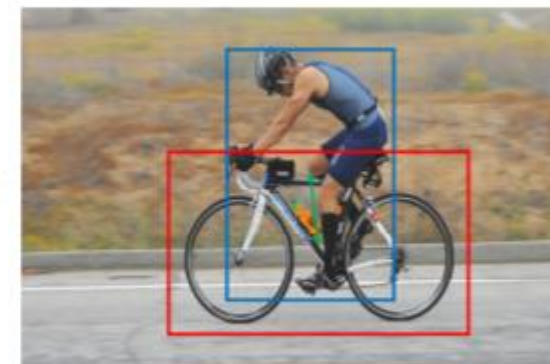
head-talk\_on-cellphone  
right\_hand-hold-cellphone



human-talk\_on-cellphone



hip-sit\_on-seat  
right\_hand-hold-handle  
left\_hand-hold-handle  
right\_foot-step\_on-pedal  
left\_foot-step\_on-pedal



human-ride-bike

# HAKE

## Part State Recognition & Activity2Vec

[http://openaccess.thecvf.com/content\\_ECCV\\_2018/papers/Haoshu\\_Fang\\_Pairwise\\_Body-Part\\_Attention\\_ECCV\\_2018\\_paper.pdf](http://openaccess.thecvf.com/content_ECCV_2018/papers/Haoshu_Fang_Pairwise_Body-Part_Attention_ECCV_2018_paper.pdf)

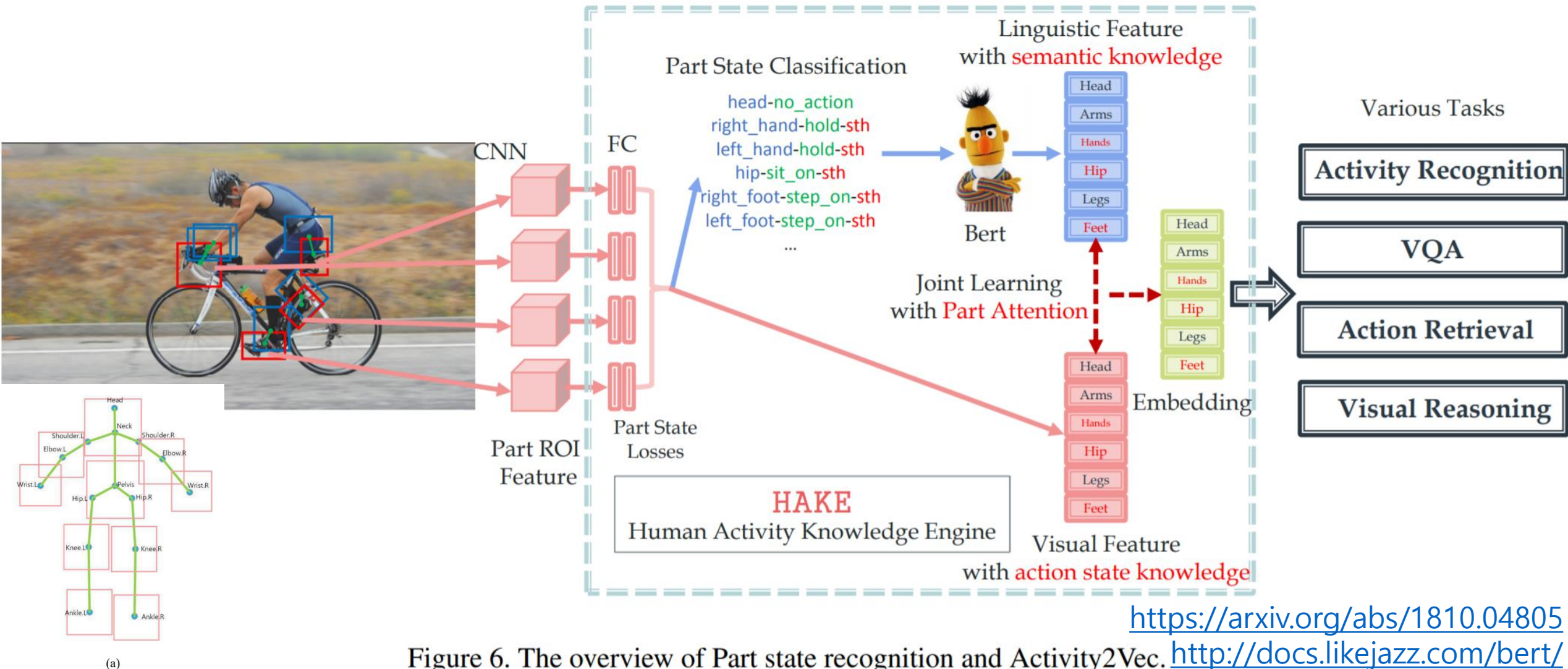


Figure 6. The overview of Part state recognition and Activity2Vec. <https://arxiv.org/abs/1810.04805> <http://docs.likejazz.com/bert/>



# HAKE

## Reasoning from Part States to Instance Activity

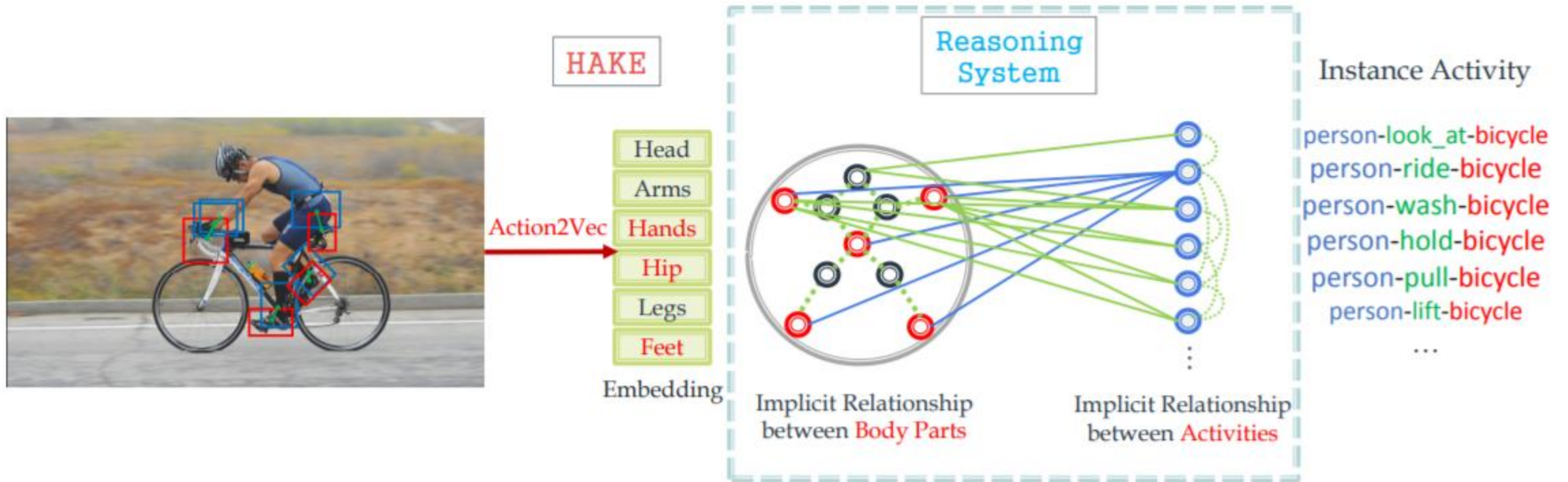


Figure 7. Reasoning from part states to instance activities.

MNIST  
MNIST를 이용한 단순화 실험

HICO (Humans Interacting with Common Objects)  
HICO를 이용한 HOI 실험

Train/test images : 38,116 / 9,658

## Task 1: HOI Classification

The input is an image and the output is a set of binary labels, each representing the presence or absence of an HOI class.



Sample annotations in the HICO benchmark

## Task 2: HOI Detection

The input is an image and the output is a set of bounding box pairs, each localizes a human plus an object and predicts an HOI class label.

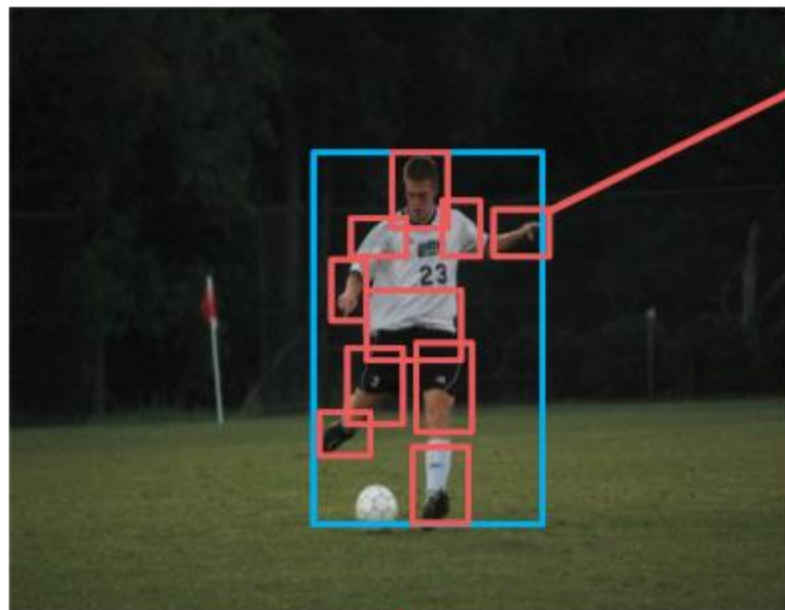


Sample annotations in the HICO-DET benchmark



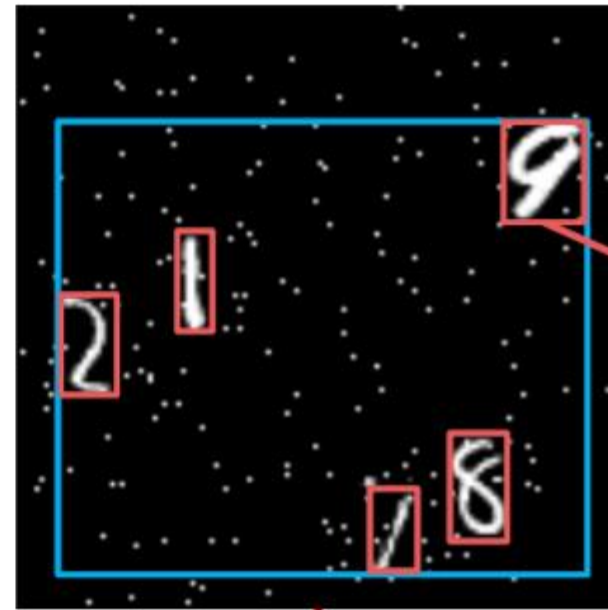
# 단순화 MNIST를 이용한 비교 실험

## Part-level Understanding



"Part Bounding Box"  
from pose keypoints,  
Annotators  $\rightarrow$  Part States

Analogy



"Part Bounding Box",  
Annotators  $\rightarrow$  numbers

$f(I) \rightarrow f'(p_1, p_2, \dots, p_{10}) ?$

Part State Labels:

head-look\_at-sth right\_arm-swing  
left\_arm-swing right\_foot-kick-sth

Labels:

person-kick-football

$g(x) \rightarrow g'(x_1, x_2, x_3, x_4, x_5) ?$

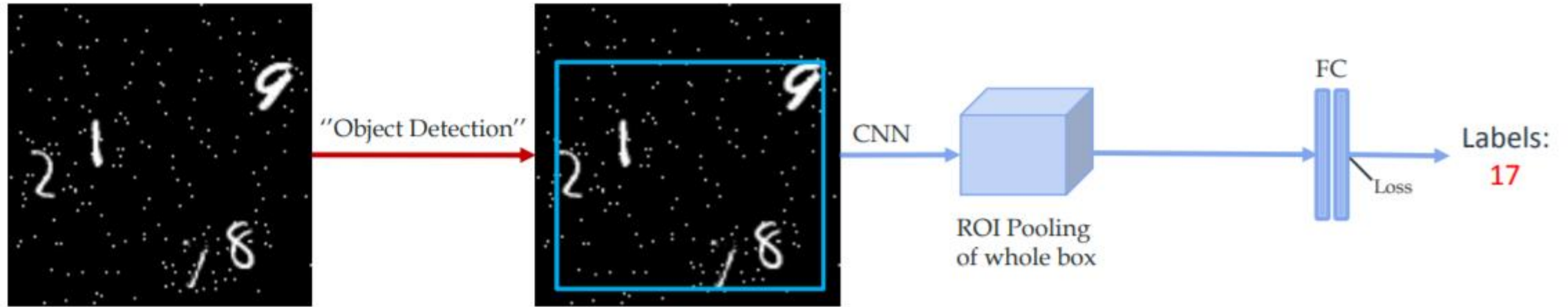
Number Labels:

9 8 1 1 2

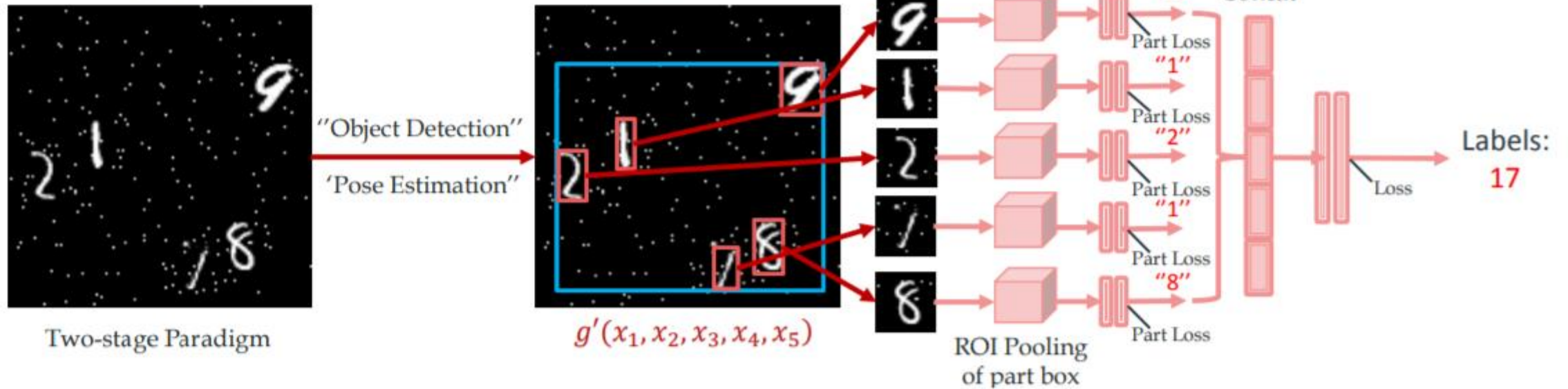
Labels:

17

# 단순화 MNIST를 이용한 비교 실험



One-stage Paradigm

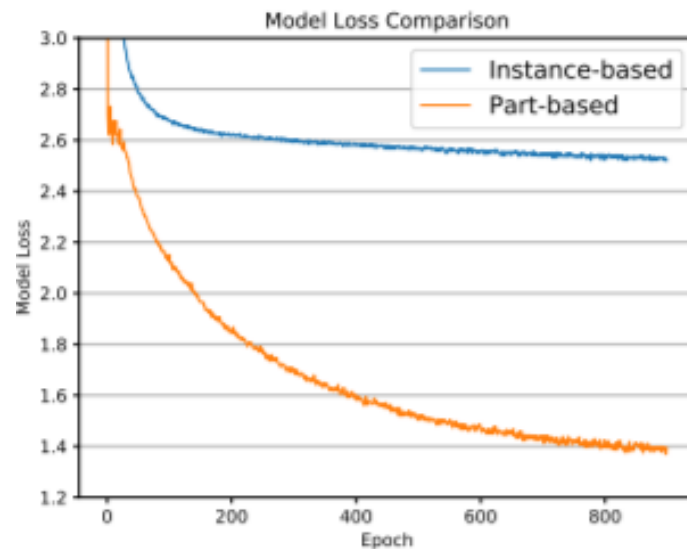


Two-stage Paradigm

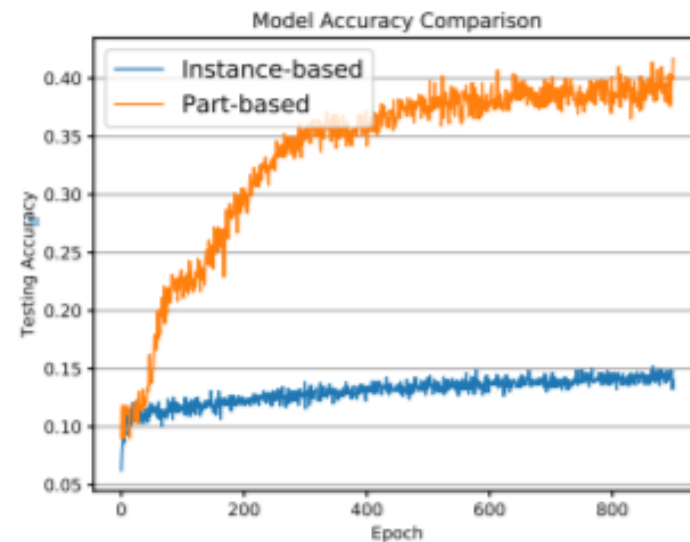
# 단순화 MNIST를 이용한 비교 실험

Method	Test Accuracy
Instance Based Paradigm	15.2
Part Based Paradigm	<b>41.7</b>

Table 1. Comparison of accuracy on our dataset



(a) Loss



(b) Test Accuracy

Figure 10. Comparison of loss and accuracy

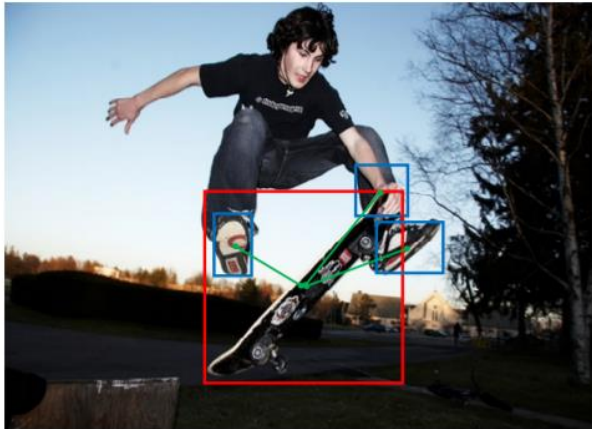
# HICO 실험

Method	mAP
AlexNet+SVM [13]	19.4
R*CNN [7]	28.5
Girdhar & Ramanan [6]	34.6
Mallya & Lazebnik [19]	36.1
Pairwise [14]	39.9
Pairwise [14]+HAKE-GT	<b>62.5</b>
Pairwise [14]+HAKE	<b>47.1</b>
Gain	<b>7.2</b>

Method	Few@1	Few@5	Few@10
Pairwise [14]	13.02	19.79	22.28
Pairwise [14]+HAKE	<b>25.40</b>	<b>32.48</b>	<b>33.71</b>
Gain	<b>12.38</b>	<b>12.69</b>	<b>11.43</b>

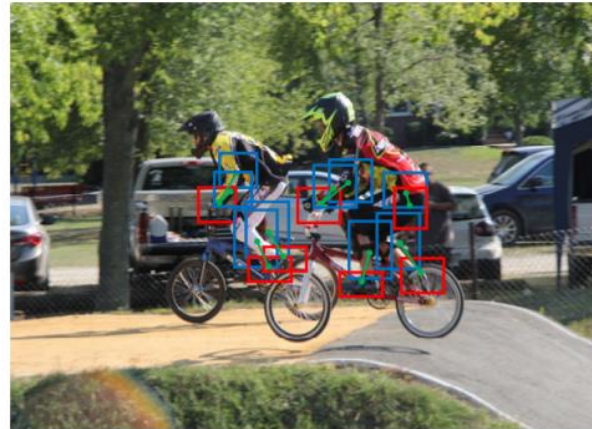


# HICO 실험



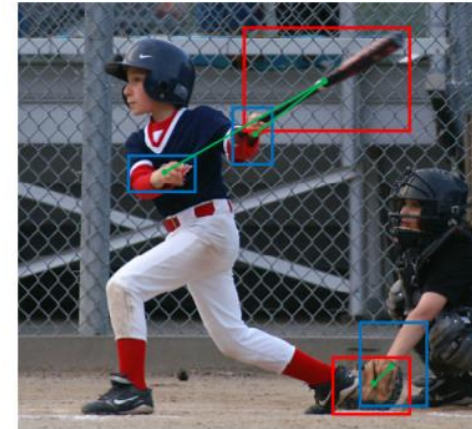
left\_hand-hold-skateboard  
right\_foot-jump\_from-skateboard  
left\_foot-jump\_from-skateboard

✓ jump\_from-skateboard ✓ hold-skateboard ✓ ride-skateboard



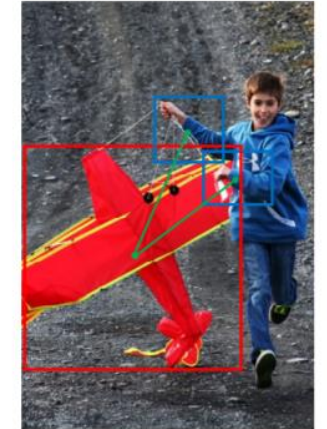
hip-sit\_on-seat  
right\_hand-hold-handle  
left\_hand-hold-handle  
right\_foot-step\_on-pedal  
left\_foot-step\_on-pedal

✓ ride-bicycle ✓ straddle-bicycle ✓ jump-bicycle



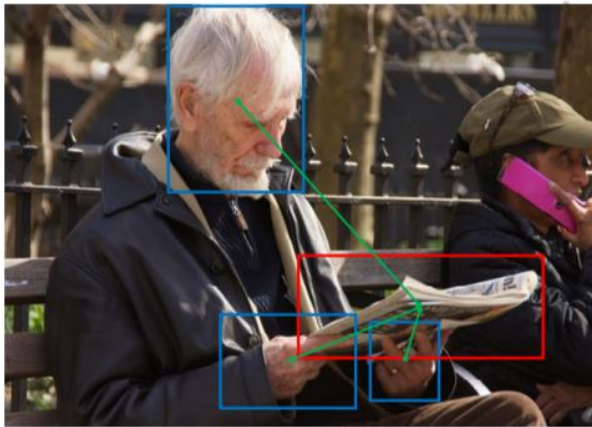
left\_hand-hold-baseball\_bat  
right\_hand-swing-baseball\_bat  
left\_hand-swing-baseball\_bat  
left\_hand-wear-baseball\_glove

✓ hold-baseball\_bat ✓ swing-baseball\_bat ✓ wear-baseball\_glove



left\_hand-hold-kite  
right\_hand-hold-kite  
left\_hand-launch-kite  
right\_hand-launch-kite

✓ hold-kite ✓ fly-kite ✓ launch-kite



left\_hand-hold-book  
right\_hand-hold-book  
head-read-book

✓ hold-book ✓ read-book



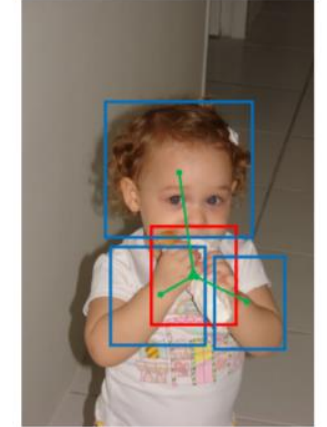
left\_knee-lean\_on-bench  
right\_knee-lean\_on-bench  
hip-sit\_on-bench

✓ sit\_on-bench ✗ lie\_on-bench



right\_foot-dribble-sports\_ball  
right\_foot-kick-sports\_ball  
left\_foot-kick-sports\_ball

✓ kick-sports\_ball ✓ dribble-sports\_ball ✓ block-sports\_ball



left\_hand-hold-hot\_dog  
right\_hand-hold-hot\_dog  
head-eat-hot\_dog

✓ hold-hot\_dog ✓ eat-hot\_dog



Q & A