2021 ICCV OCTOBER 11-17 VIRTUAL

An Empirical Study of the Collapsing Problem in Semi-Supervised 2D Human Pose Estimation

Rongchang Xie, Chunyu Wang, Wenjun Zeng, Yizhou Wang

北京大学 PEKING UNIVERSITY

Microsoft Research 微软亚洲研究院

Task

Semi-Supervised Human Pose Estimation

- Leverage *large-scale unlabeled images* for training 2D human pose estimator

Why it is important?

- Labeled datasets are small (MPII, COCO)
- Poor generalization accuracy (Variations of poses, appearance, background)
- Abundant images and videos containing person

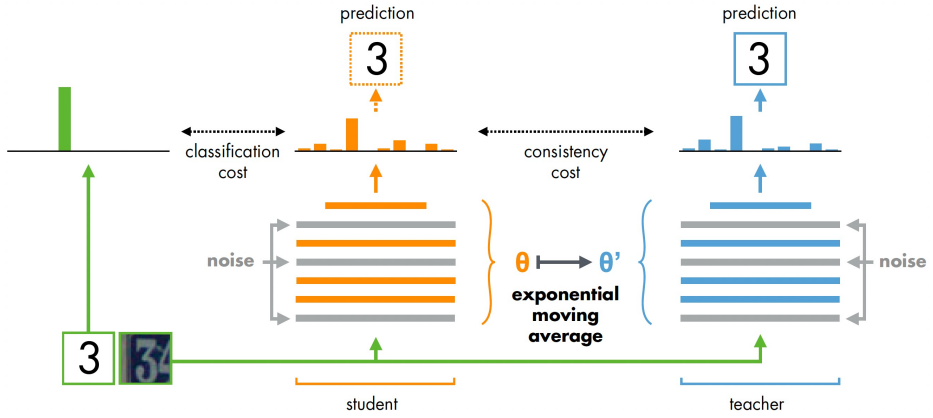



Related works

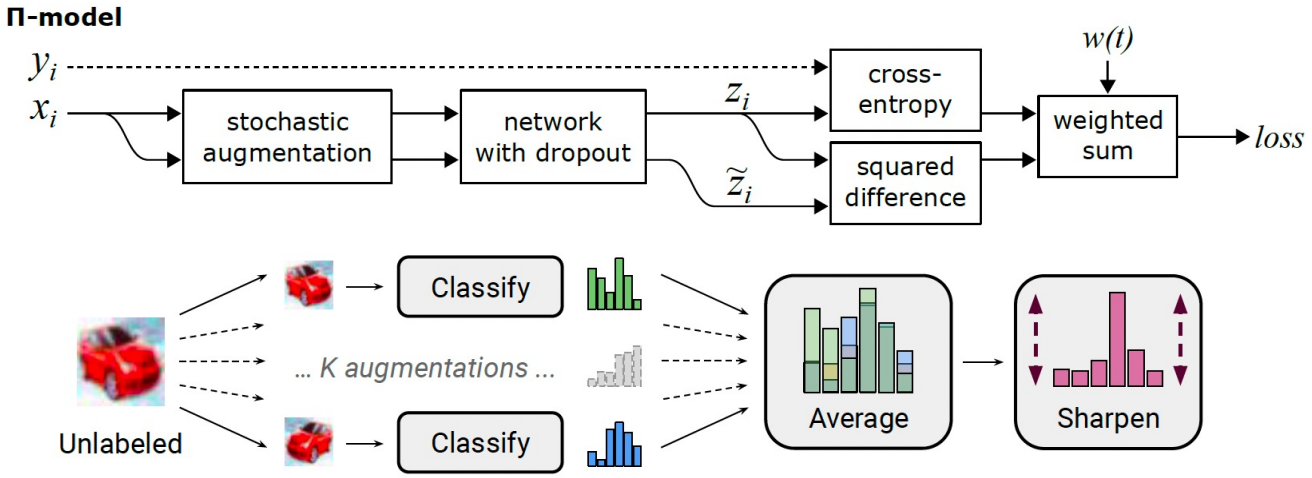
Consistency regularization

$$L_u = \mathbb{E}_{\mathbf{I}\in\mathcal{U}} \left|\left| f(\mathbf{I}_\eta, \theta) - f(\mathbf{I}_{\eta'}, \theta') \right|\right|^2 .$$

- Requiring the model to have *similar* predictions for *different augmentations* of the same image
- Consistency-based training works well and achieves SOTA results for semi-supervised classification



Tarvainen et al. 2017



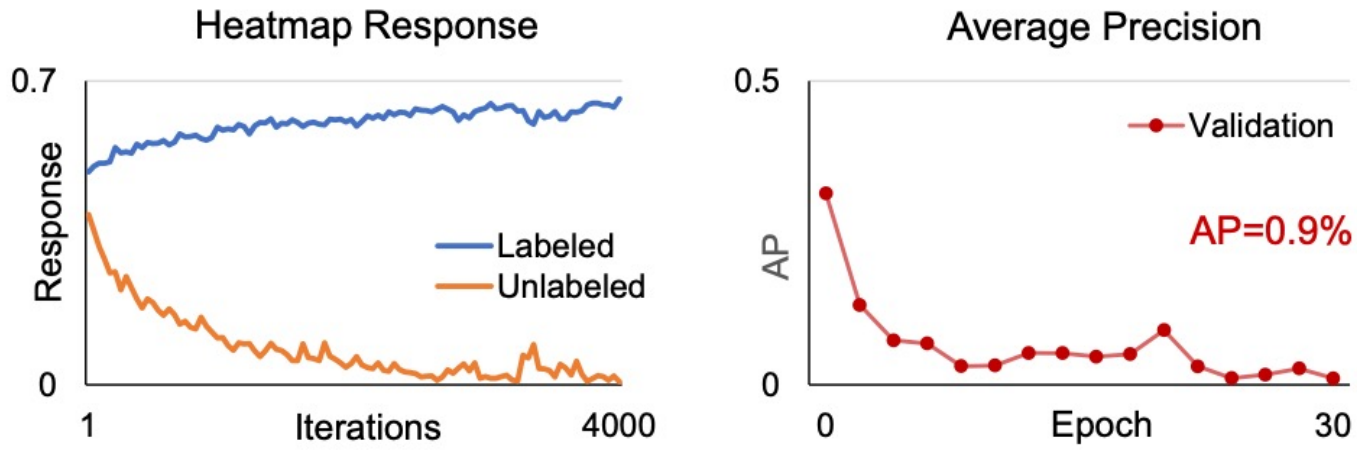
Laine et al. 2017 and Berthelot et al. 2019

Challenges

The collapsing problem

The consistency-based model *collapses* (predict all pixels as *background*) when applied to human pose estimation

Supervised Naïve consistency



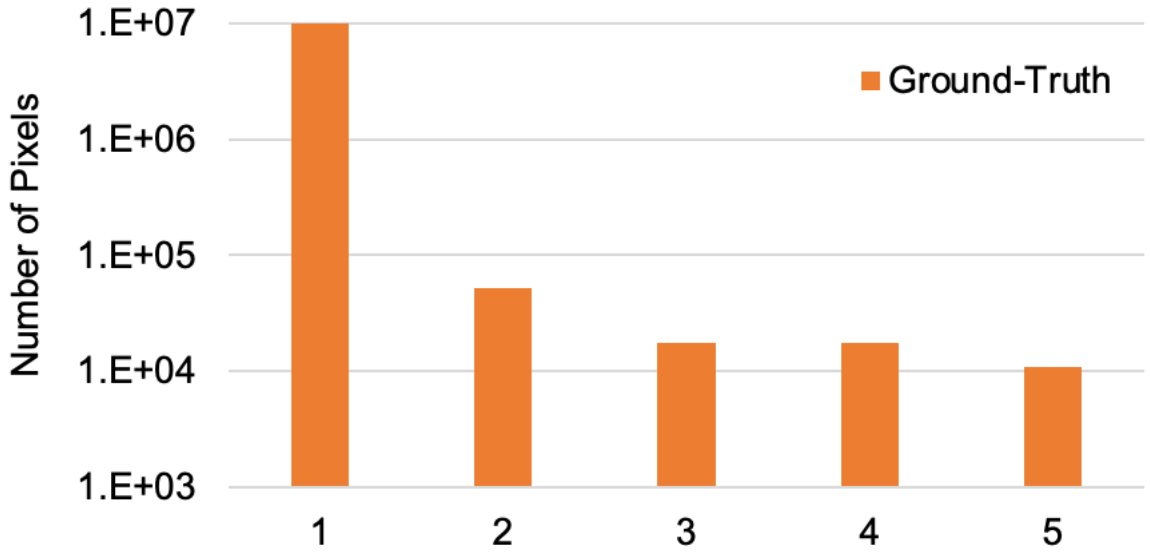
The results of naïve consistency learning

Analysis of the Collapsing Problem

The Imbalanced Distribution

- The class distribution is extreme *imbalanced* in 2D pose estimation
- Background pixels dominate the number of samples in heatmap

Keypoint Heatmap

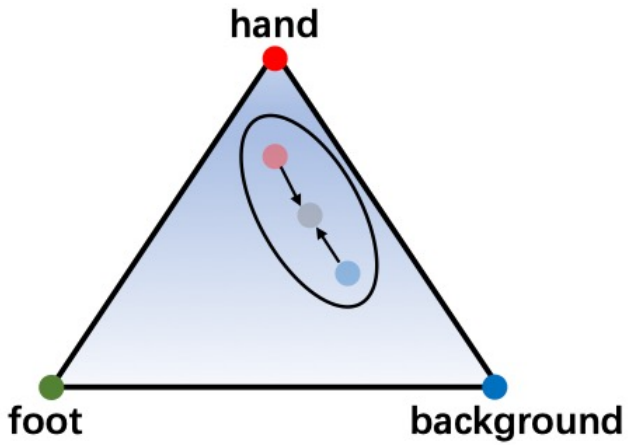


The histogram of GT heatmap

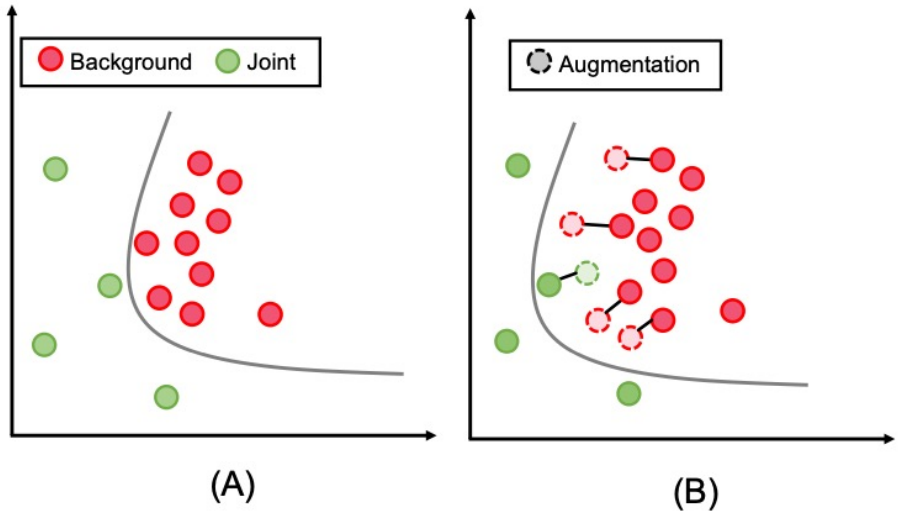
Analysis of the Collapsing Problem

The Imbalanced Distribution

- The naive consistency regularization moves data and their augmentations to their middle points
- More data will be close to the decision boundary which pushes the decision boundary to pass through the areas of minor class that is sparse globally



Naïve consistency learning

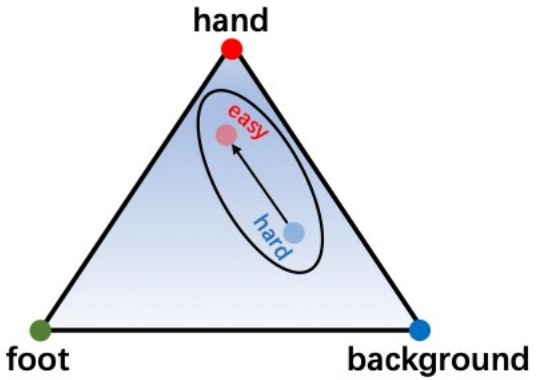


The decision boundary in naïve consistency

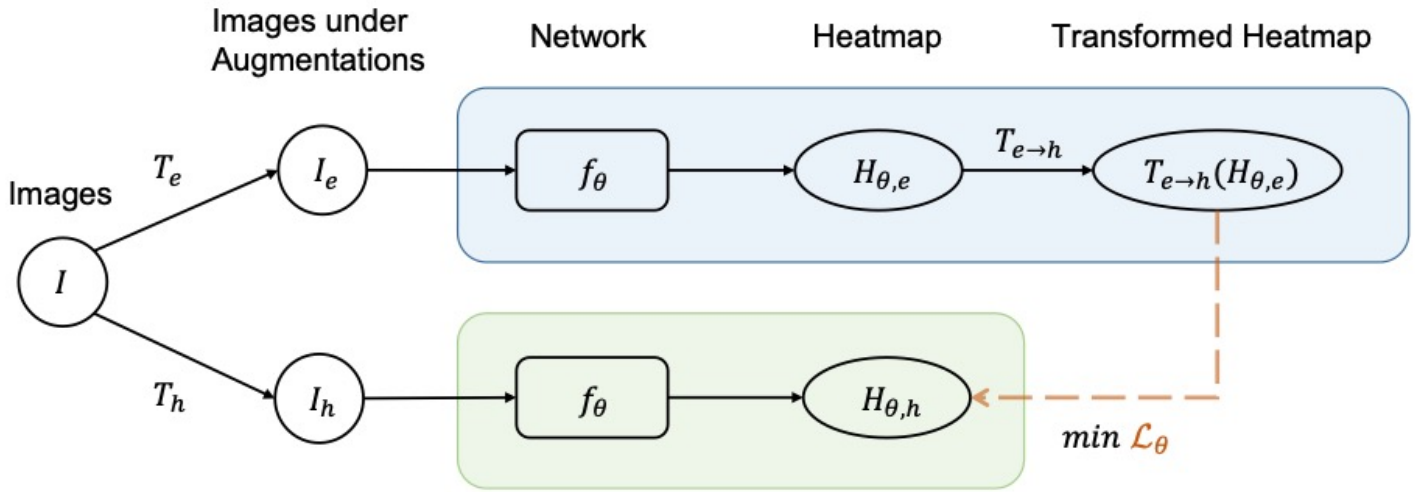
Method

Address Collapsing by Easy-Hard Augmentation

Drive the less accurate predictions which are close to the decision boundary to the direction of more accurate predictions



Proposed consistency learning

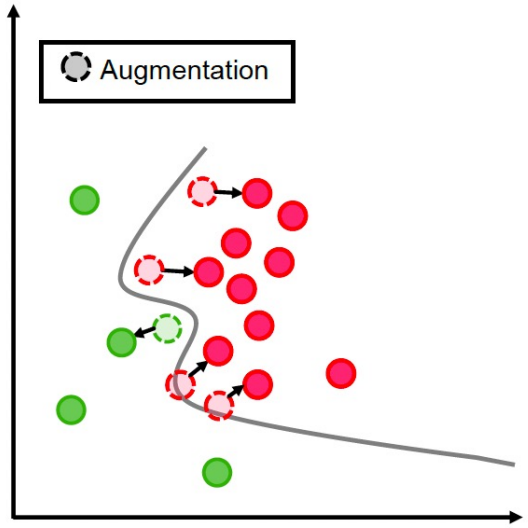


The optimization is unidirectional

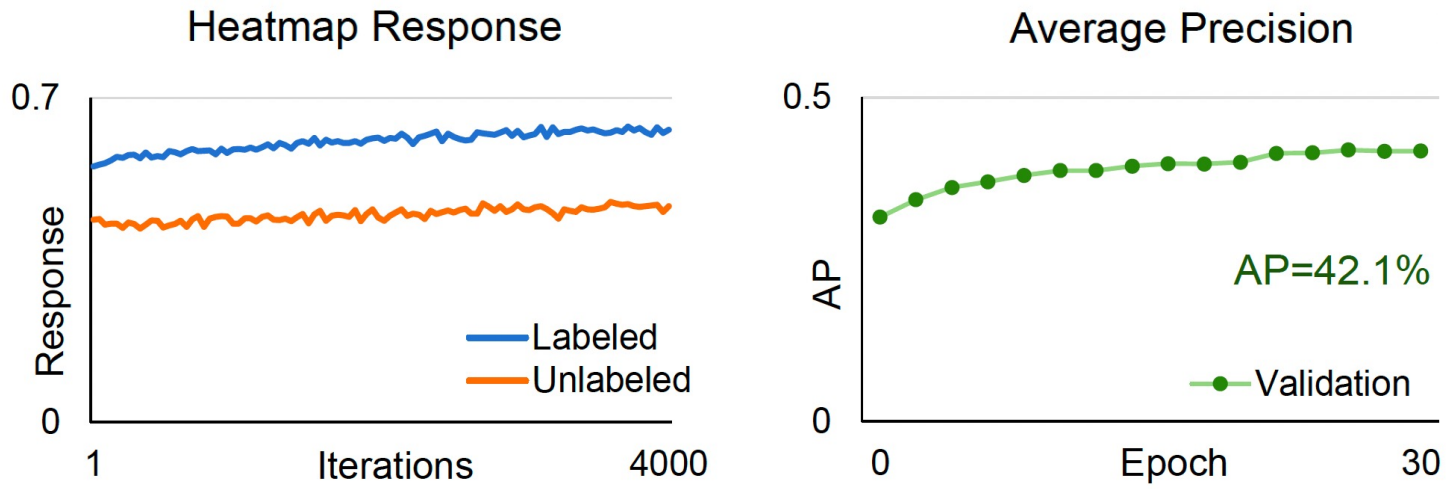
Method

Address Collapsing by Easy-Hard Augmentation

The proposed method stabilize the training and avoid the collapsing problem



Proposed consistency learning

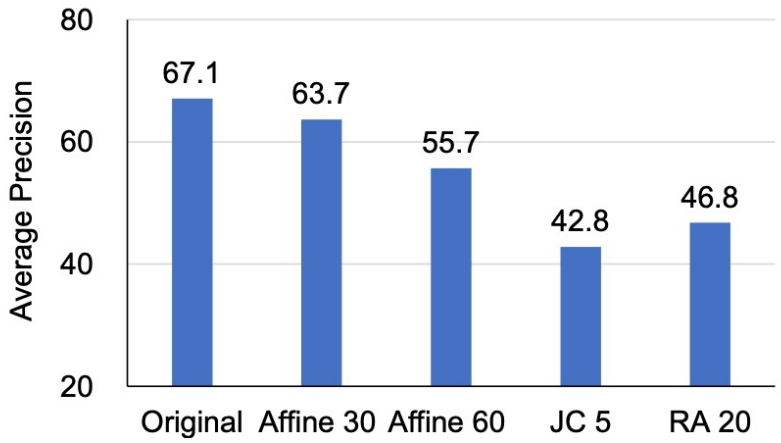


The results of proposed methods

Method

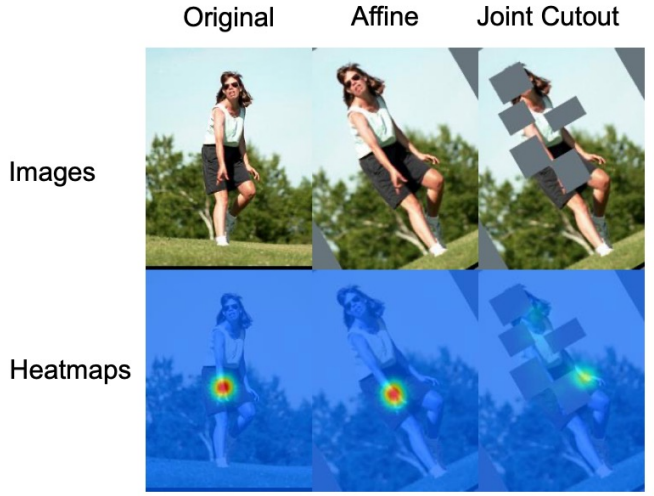
How to construct augmentation pairs

- Determine “easiness” and “hardness” based on its effect on estimation accuracy in testing
- The selection seems to generalize across different datasets



Easy Aug	Hard Aug	Avoid Collapsing
Affine 30	Affine 30	No
Affine 60	Affine 60	No
Original	Affine 60	**Yes**
Affine 30	Affine 60	**Yes**
Affine 60	JC 5	**Yes**

The effect of different easy-hard augmentation pairs



Joint Cutout

Failed attempts

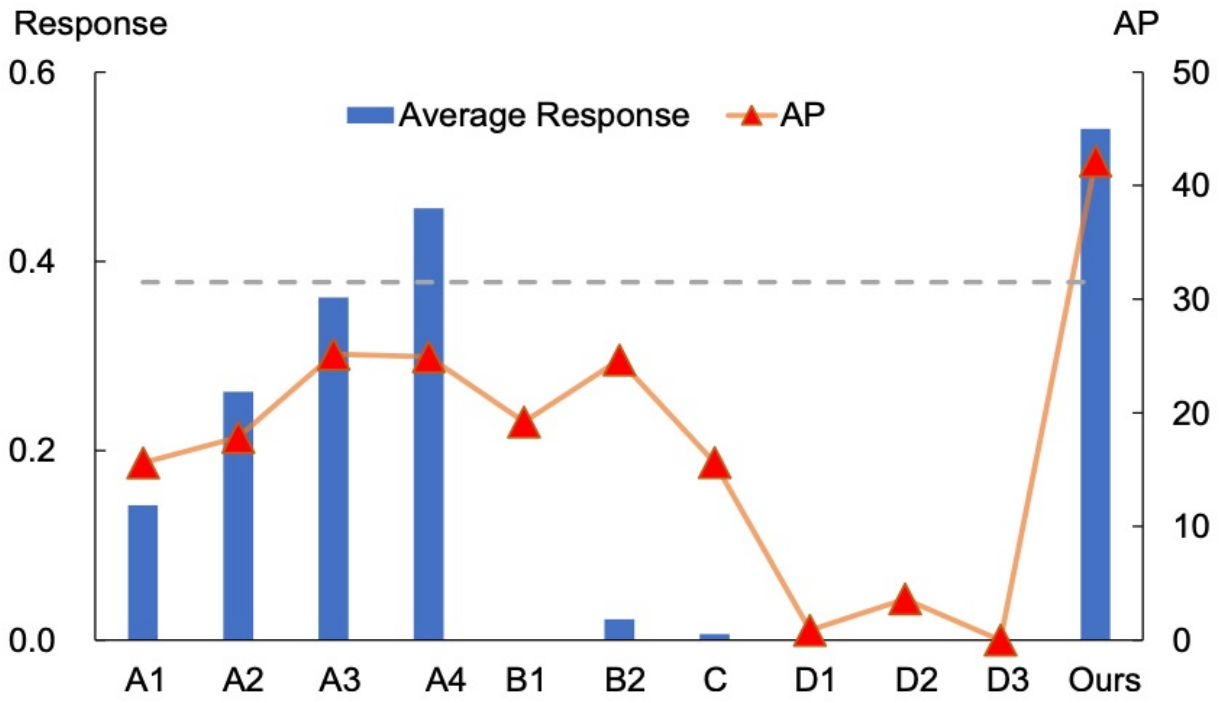
We tried other several methods to avoid collapsing but they all failed

A: Confidence threshold

B: EMA parameters

C: Re-weighting

D: Easy-easy, hard-hard and hard-easy augmentation



Experiments

Semi-supervised setting

The proposed method outperforms other semi-supervised methods

Methods	Aug.	1K	5K	10K	All
Supervised [41]	A	31.5	46.4	51.1	67.1
PseudoPose	A	37.2	50.9	56.0	—
DataDistill [26]	A	37.6	51.6	56.6	—
Ours (Single)	A	38.5	50.5	55.4	—
Ours (Dual)	**A**	**41.5**	**54.8**	**58.7**	—
Ours (Single)	A+JC	42.1	52.3	57.3	—
Ours (Dual)	A+RA	43.7	55.4	59.3	—
Ours (Dual)	A+JC	44.6	55.6	59.6	—

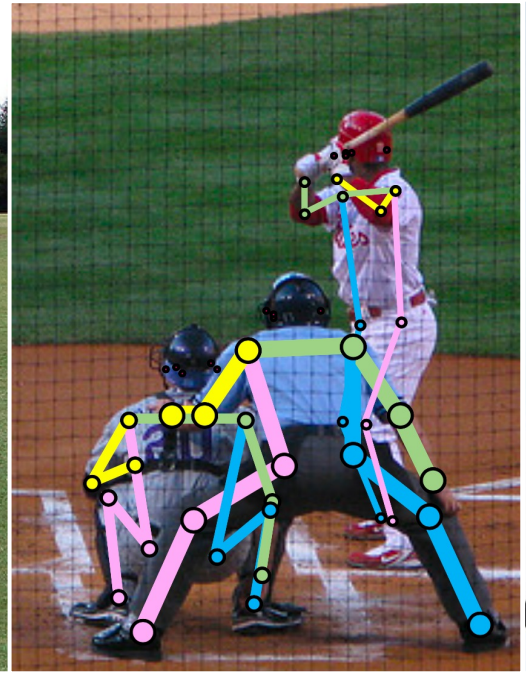
Experiments

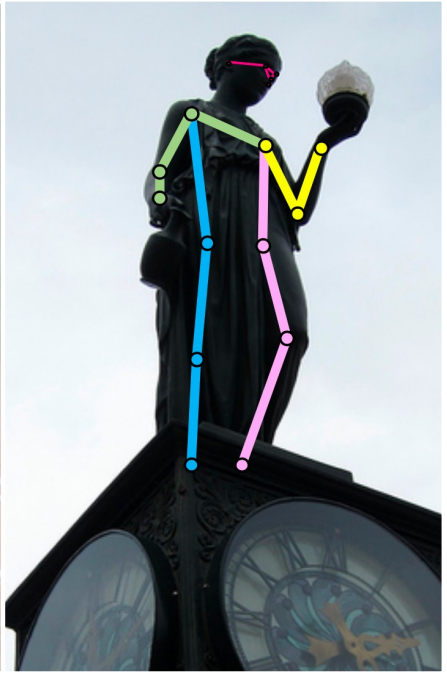
Full labels setting

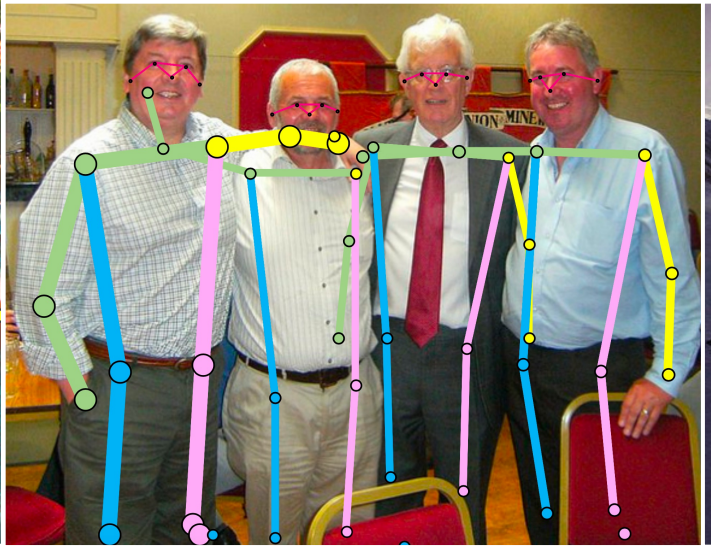
The proposed method can further improve performance when full labels are used in COCO dataset

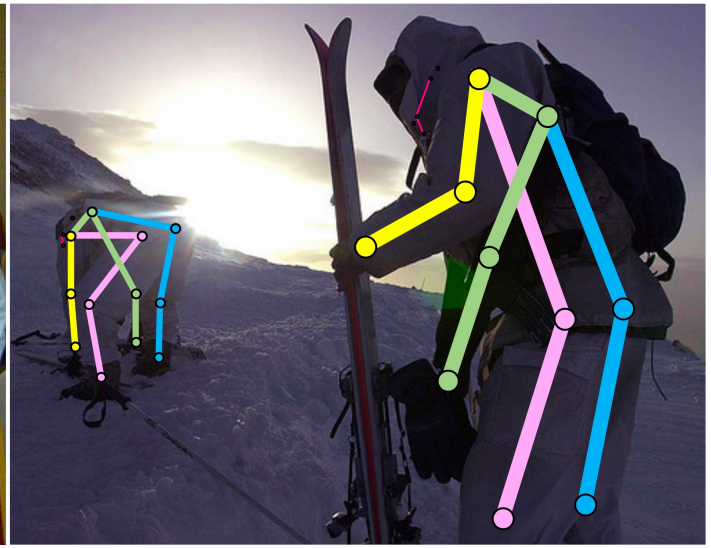
Method	Network	Input Size	GFLOPS	#Params	AP	AP0.50	AP0.75	APM	APL	AR
SB [41]	ResNet50	256 × 192	8.9	34.0	70.2	90.9	78.3	67.1	75.9	75.8
SB [41]	ResNet152	256 × 192	15.7	68.6	71.9	91.4	80.1	68.9	77.4	77.5
HRNet [30]	HRNetW48	384 × 288	32.9	63.6	75.5	92.5	83.3	71.9	81.5	80.5
MSPN [21]	ResNet50	384 × 288	58.7	71.9	76.1	93.4	83.8	72.3	81.5	81.6
DARK [45]	HRNetW48	384 × 288	32.9	63.6	76.2	92.5	83.6	72.5	82.4	81.1
UDP [13]	HRNetW48	384 × 288	33.0	63.8	76.5	92.7	84.0	73.0	82.4	81.6
Ours (+SB)	ResNet50	256 × 192	8.9	34.0	**72.3** (↑ **2.1**)	91.8	80.5	69.3	77.8	77.7
Ours (+SB)	ResNet152	256 × 192	15.7	68.6	**73.7** (↑ **1.8**)	92.1	82.1	71.0	79.0	79.1
Ours (+HRNet)	HRNetW48	384 × 288	32.9	63.6	**76.7** (↑ **1.2**)	92.5	84.3	73.5	82.5	81.8
Ours (+DARK)	HRNetW48	384 × 288	32.9	63.6	**77.2** (↑ **1.0**)	92.6	84.5	73.9	82.9	82.2

Visualization









北京大学前沿计算研究中心
Center on Frontiers of Computing Studies, Peking University

An Empirical Study of the Collapsing Problem in Semi-Supervised 2D Human Pose Estimation

Rongchang Xie, Chunyu Wang, Wenjun Zeng, Yizhou Wang

The code and pretrained models are released at
https://github.com/xierc/Semi_Human_Pose

2021 ICCV OCTOBER 11-17 VIRTUAL