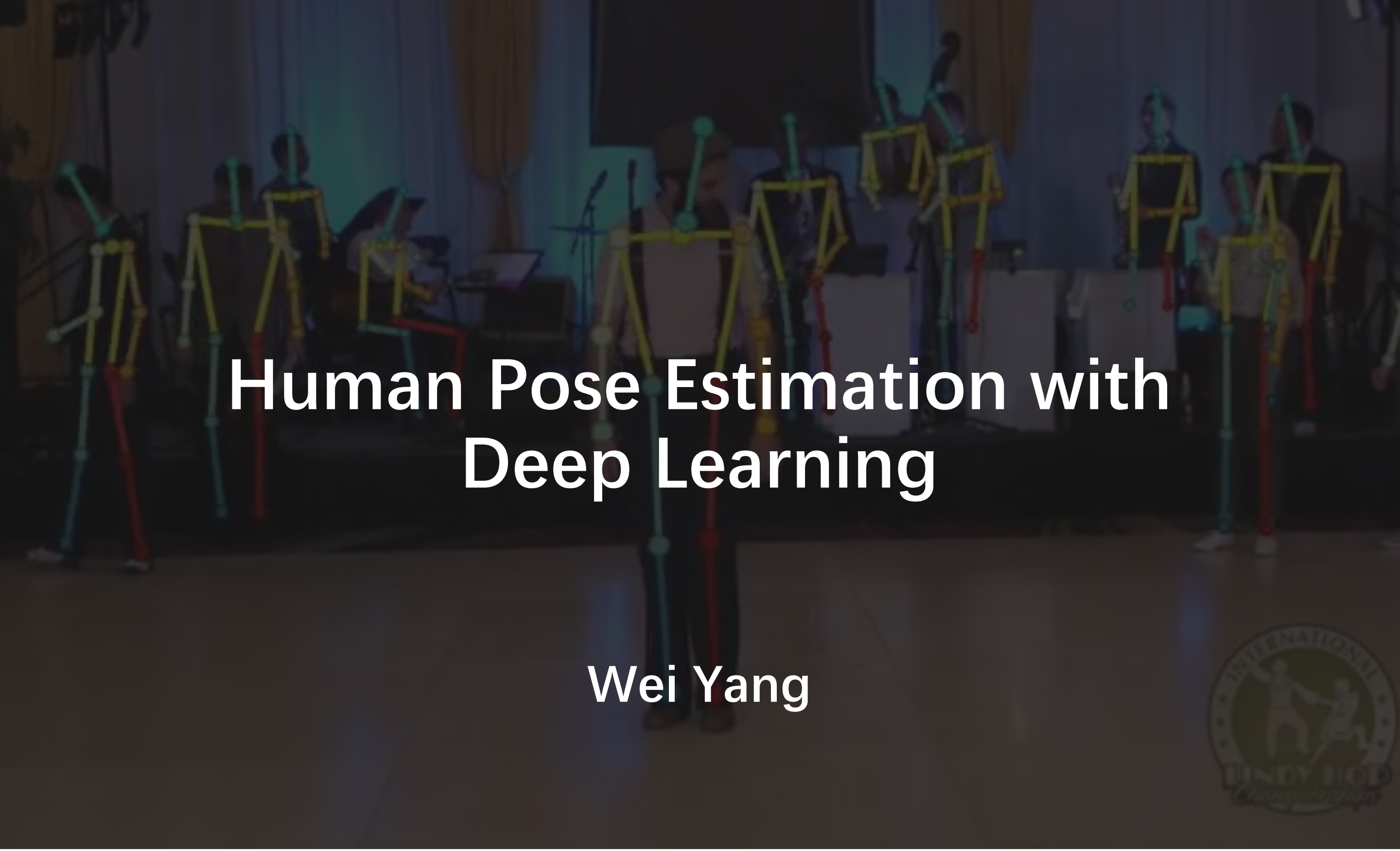
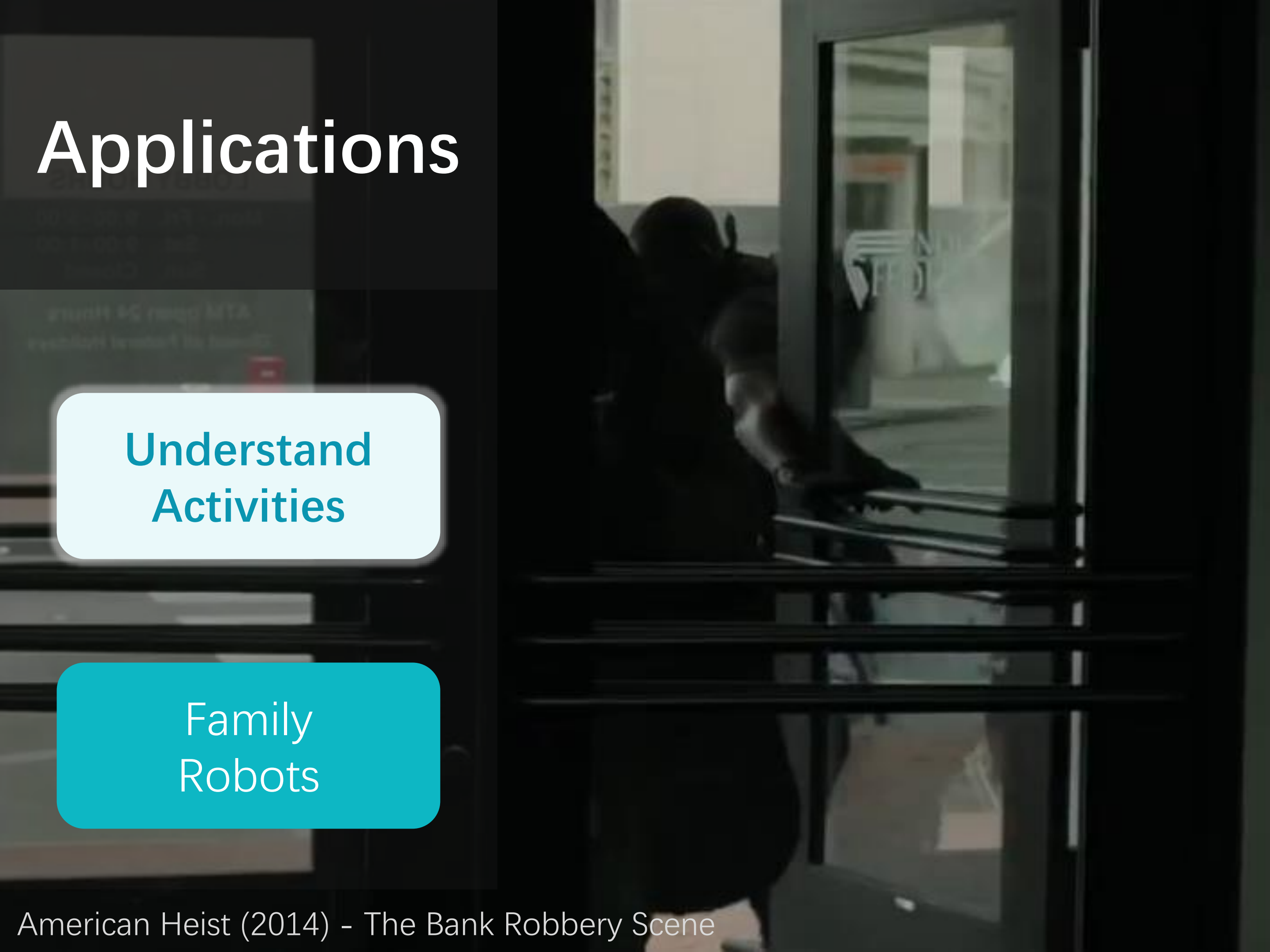


Human Pose Estimation with Deep Learning

Wei Yang

香港中文大學
The Chinese University of Hong Kong

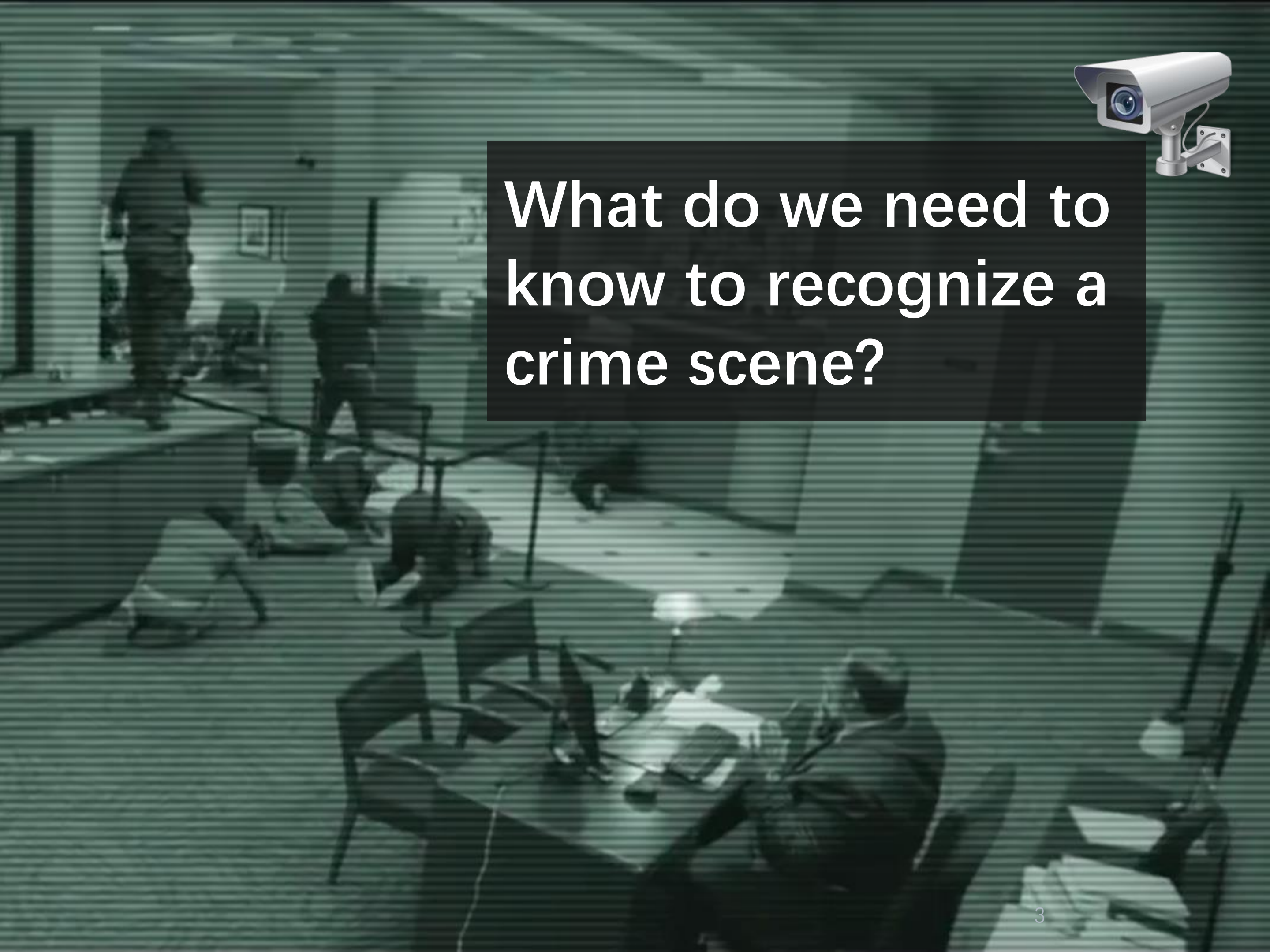
Applications



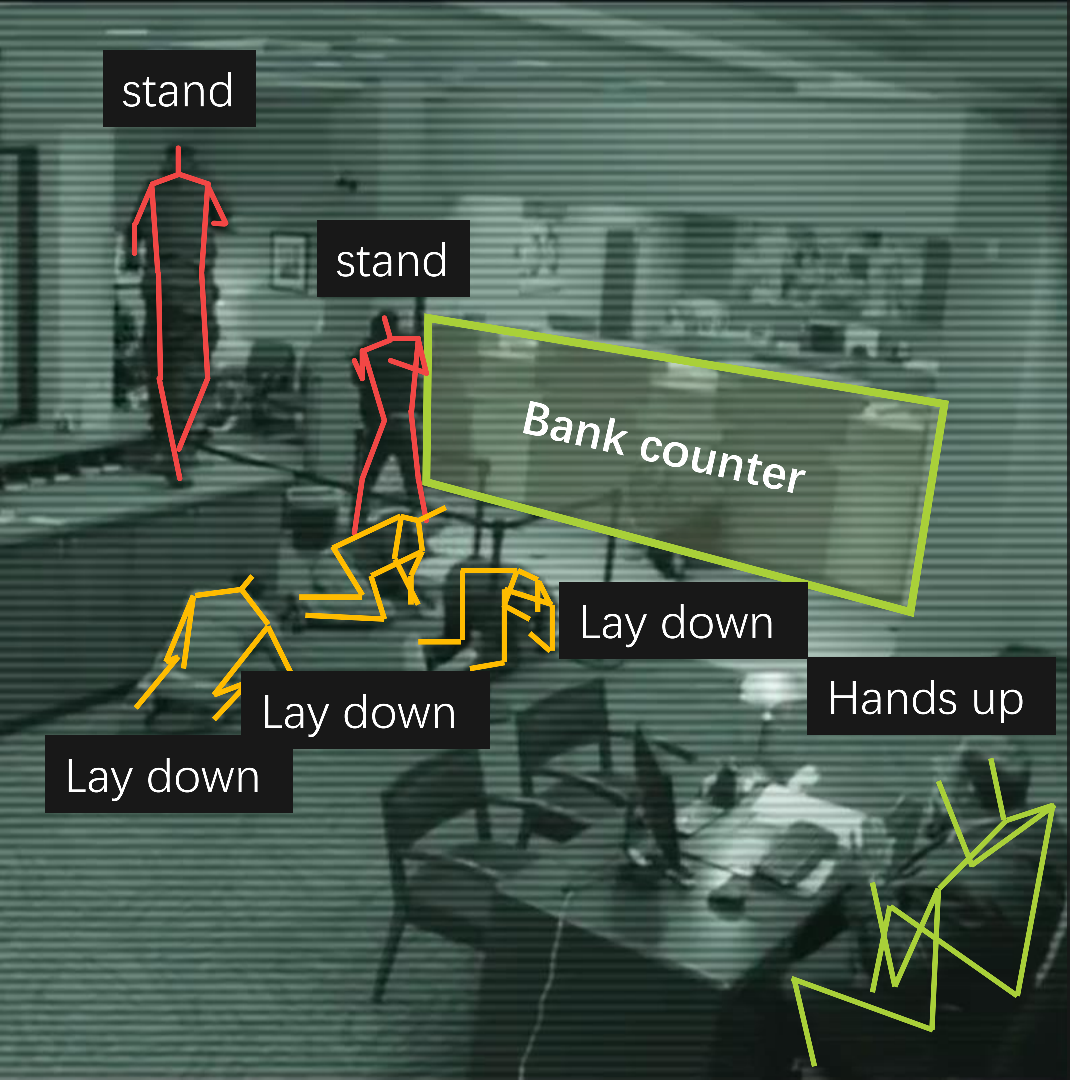
Understand Activities

Family Robots

American Heist (2014) - The Bank Robbery Scene



What do we need to know to recognize a crime scene?



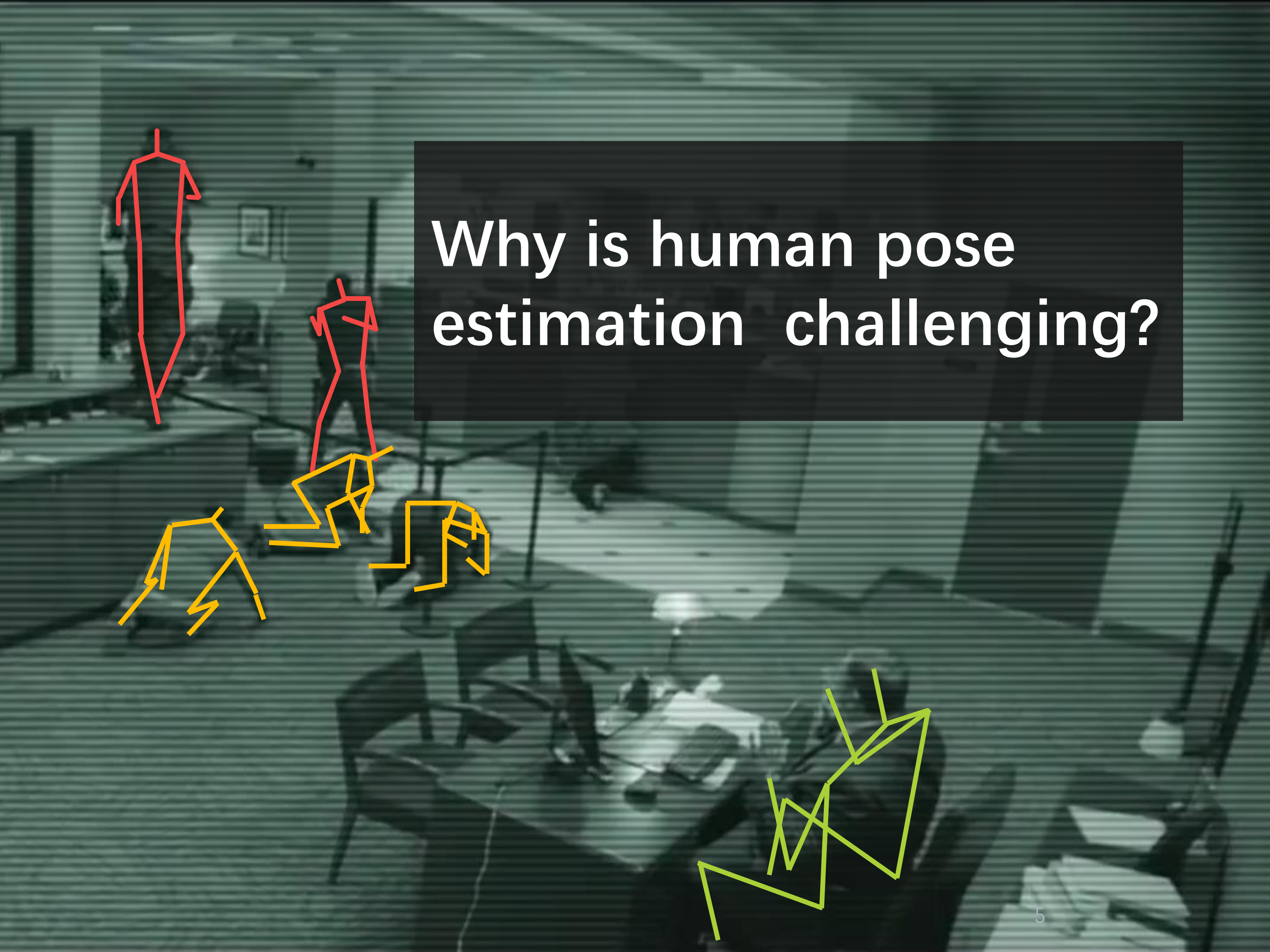
Cues

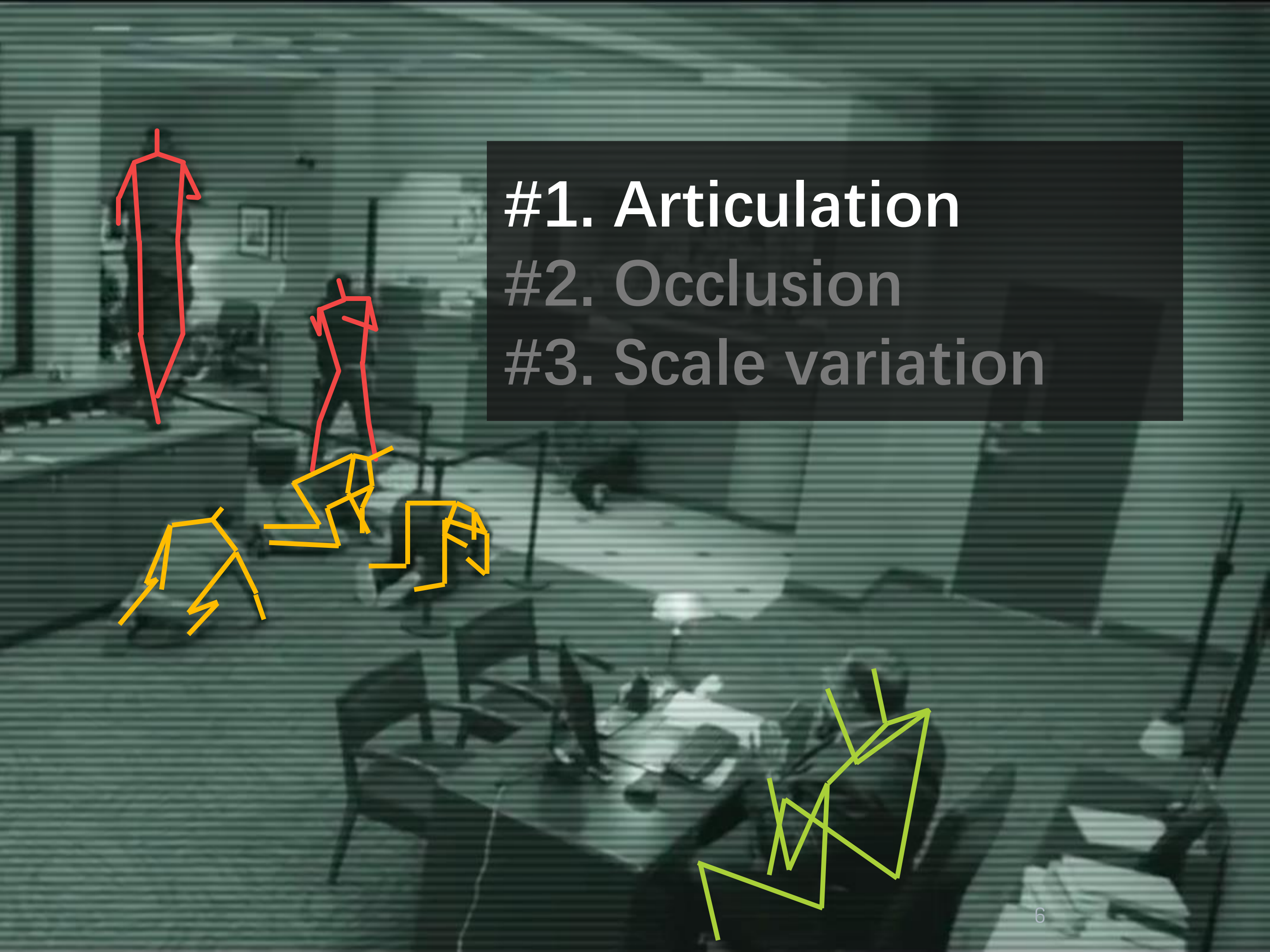
Scene: bank

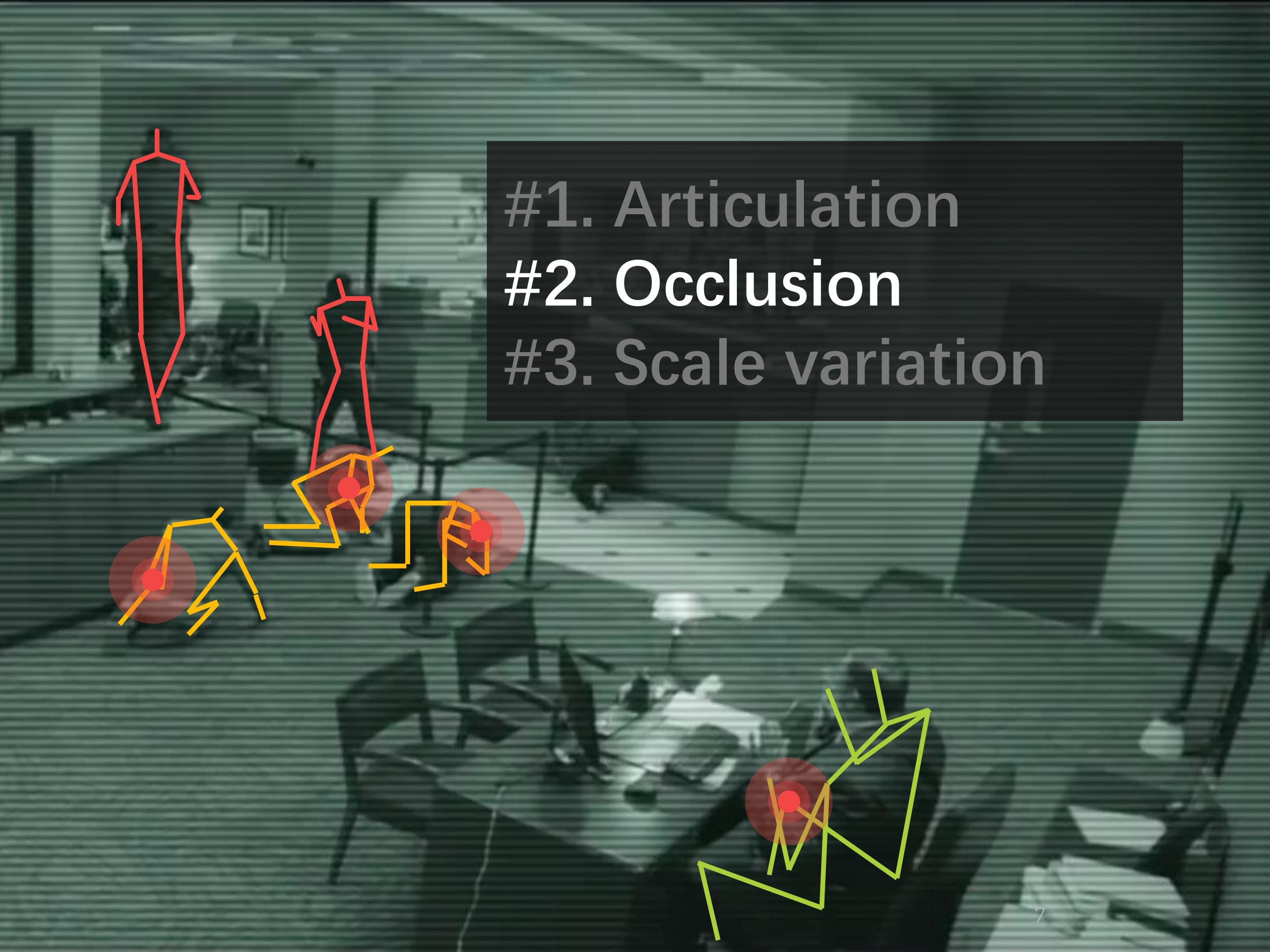
Abnormal **pose**

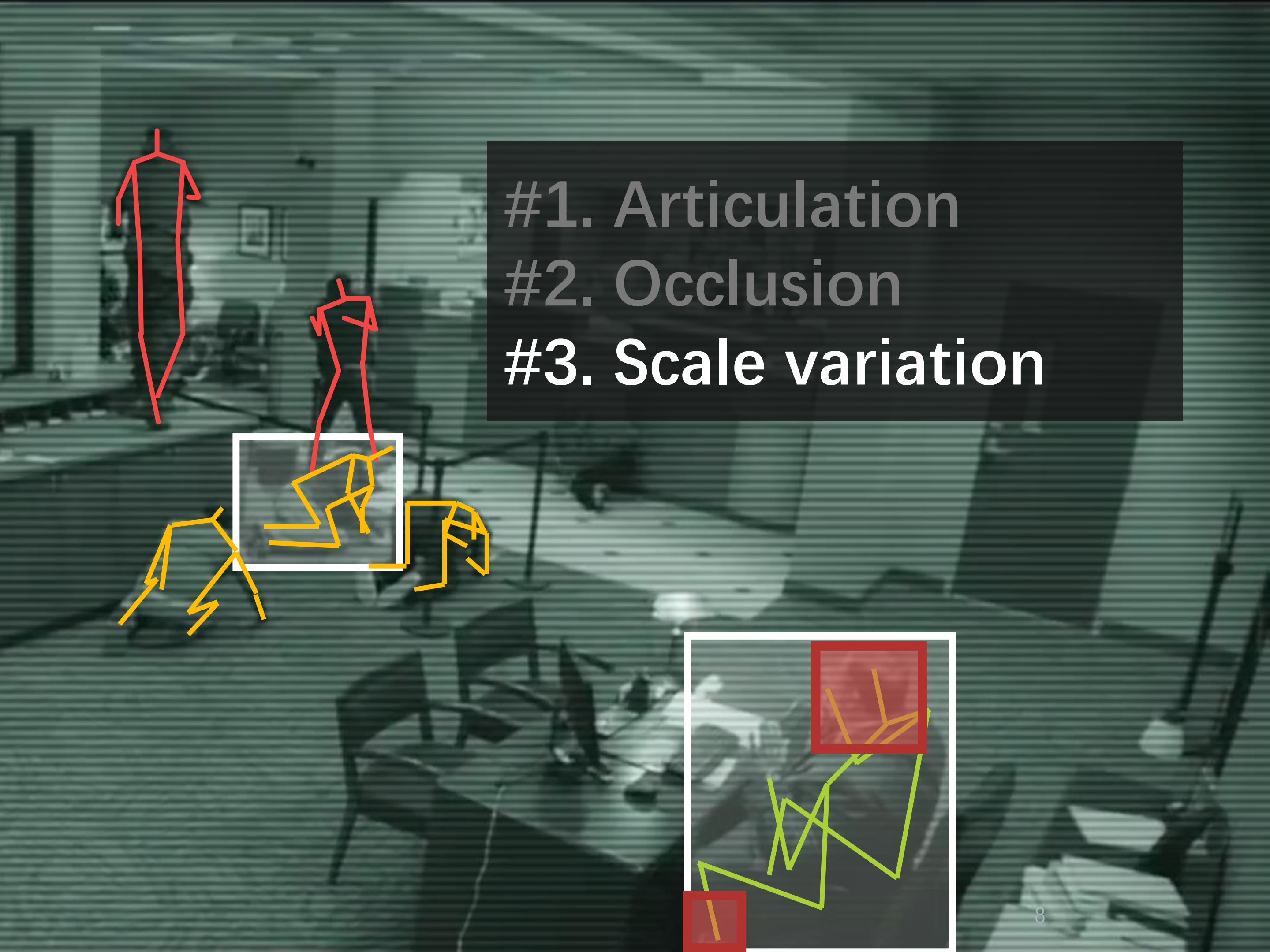
Activity: **robbery**

Why is human pose estimation challenging?




#1. Articulation
#2. Occlusion
#3. Scale variation


#1. Articulation
#2. Occlusion
#3. Scale variation


#1. Articulation
#2. Occlusion
#3. Scale variation

Applications

Understand Activities

Family Robots

3D Human Poses

Real-Time Imitation of Human Whole-Body Motions by Humanoids.
J. Koenemann, F. Burget, and M. Bennewitz. ICRA, 2014.

Deep Learning Based Methods

P heatmaps H_p

Regression with Euclidean Loss: $L = \frac{1}{2}\sum_{p=1}^{P} \left\| \widehat{H}_p - H_p \right\|_2^2$

where $\widehat{H}_p \sim N(l_p, \Sigma)$, $\quad s.t., p = 1, \cdots, P$

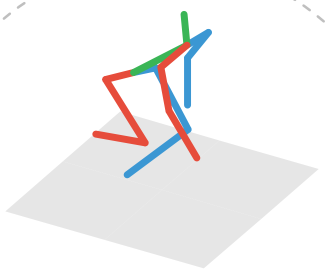
Outline

Scale

Feature pyramid learning

ICCV 2017

3D Pose



In-the-wild 3D pose estimation

CVPR 2018

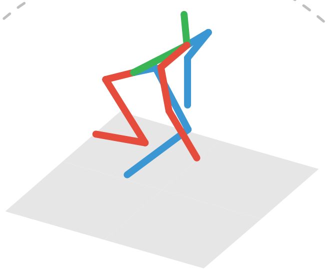
Outline

Scale

Feature pyramid learning

ICCV 2017

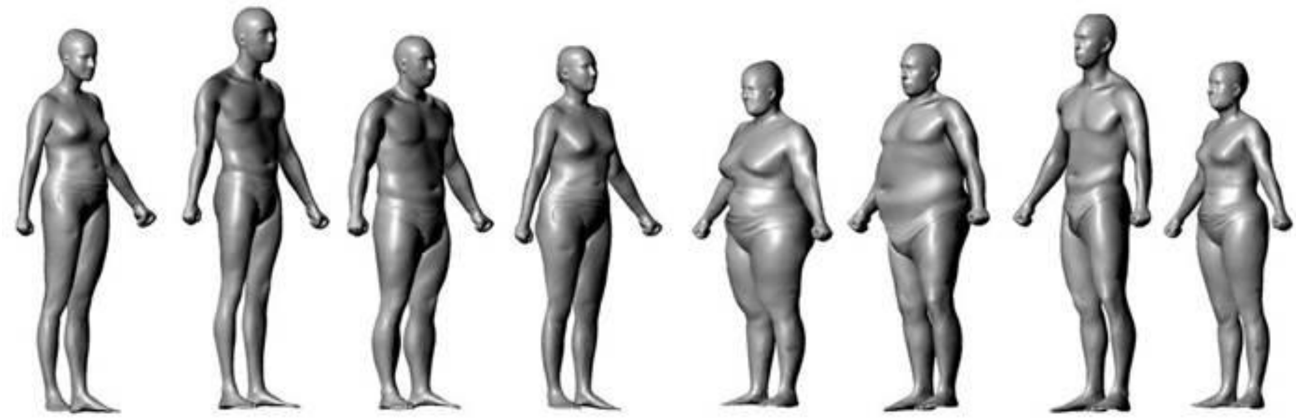
3D Pose



In-the-wild 3D pose estimation

CVPR 2018

Why the Scale Matters?



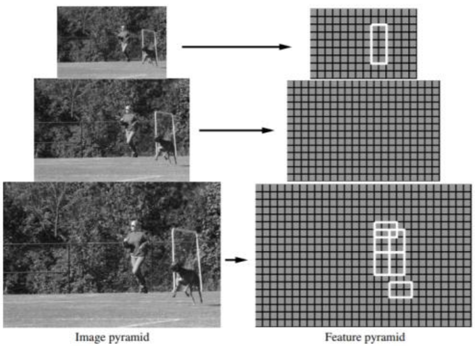
Yipin Yang, Yao Yu, Yu Zhou, Sidan Du, James Davis, Ruigang Yang. Semantic Parametric Reshaping of Human Body Models. In 3DV Workshop on Dynamic Shape Measurement and Analysis, 2014.

Why the Scale Matters?

Learning Feature Pyramids for Human Pose Estimation
Wei Yang , Shuang Li, Wanli Ouyang, Hongsheng Li, Xiaogang Wang
ICCV, 2017

Previous work

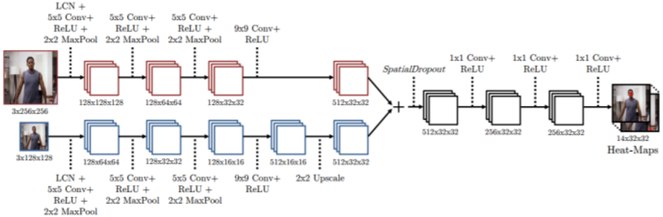
Multi-scale testing



😞 The model itself is not scale invariant

Felzenszwalb, Pedro F., et al. "Object detection with discriminatively trained part-based models." *TPAMI, 2010*.

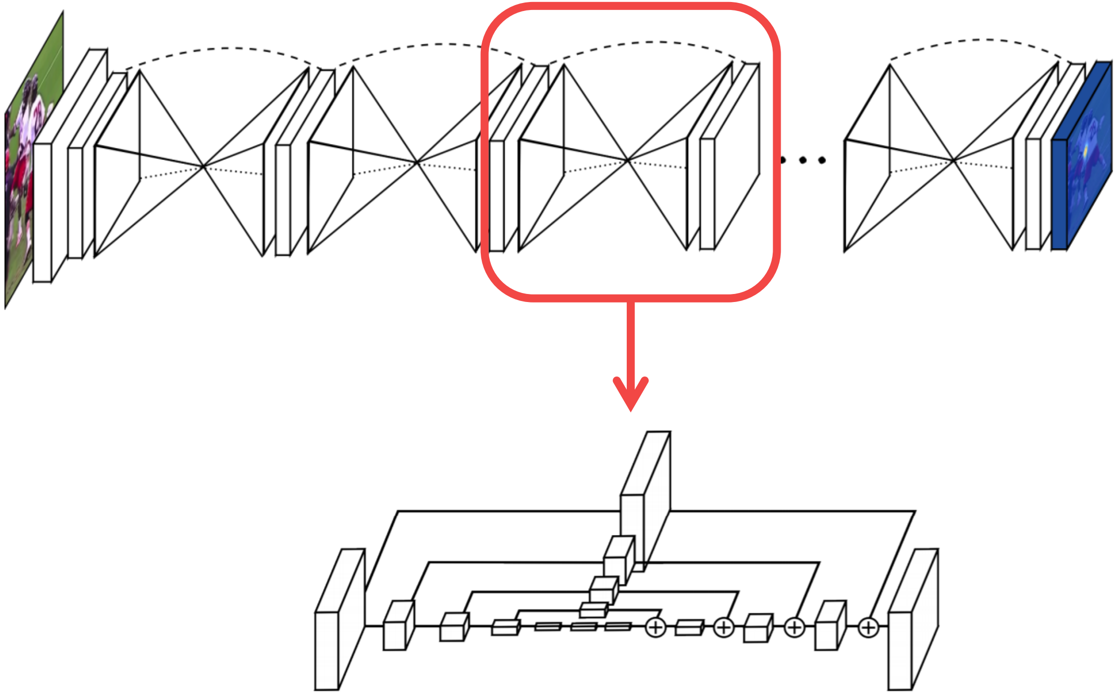
Multi-branch network



😞 Need much more memory and computation

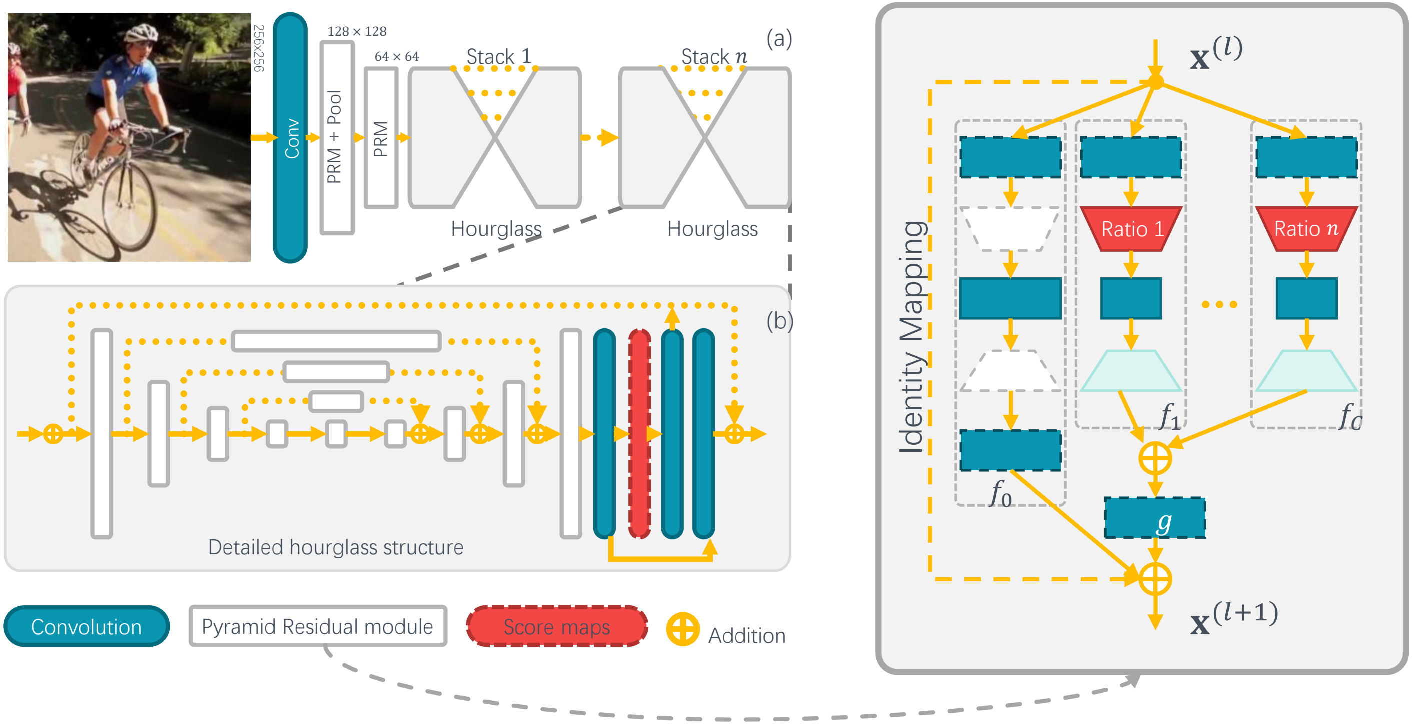
Tompson, Jonathan, et al. "Efficient object localization using convolutional networks." *CVPR*. 2015.

Hourglass



Newell A, Yang K, Deng J. Stacked hourglass networks for human pose estimation[C]//European Conference on Computer Vision. Springer, Cham, 2016: 483-499.

Pyramid Residual Modules

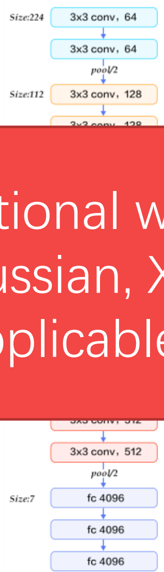


Newell et al. Stacked Hourglass Networks for Human Pose Estimation. ECCV, 2016

Initialization of Multi-Branch Networks

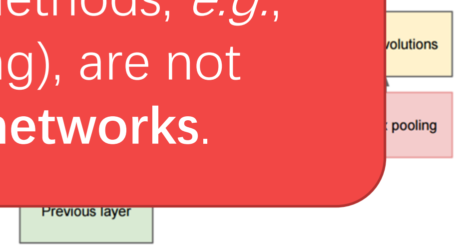
Single-branch networks

VGG



Multi-branch networks

Inceptions



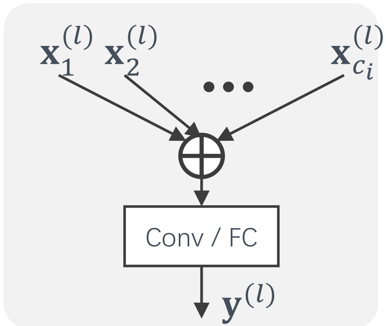
Full Inception module

Traditional weight initialization methods, *e.g.*, Gaussian, Xavier, MSRA (Kaiming), are not applicable for **multi-branch networks**.

Xavier Glorot, Yoshua Bengio ; Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, PMLR 9:249-256, 2010.

Initialization of Multi-Branch Networks

Forward

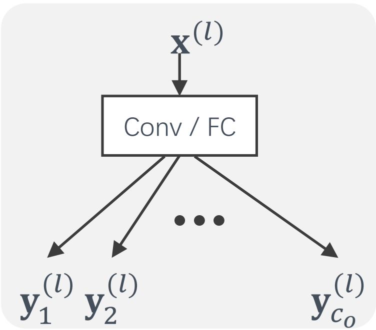


$$\mathbf{y}^{(l)} = \mathbf{W}^{(l)} \sum_{c=1}^{c_i^{(l)}} \mathbf{x}_c^{(l)} + \mathbf{b}^{(l)}$$

$$\mathbf{x}^{(l+1)} = f(\mathbf{y}^{(l)})$$

$$\alpha C_i^{(l)} n_i^{(l)} \mathrm{Var}\left(\omega^{(l)}\right) = 1$$

Backward



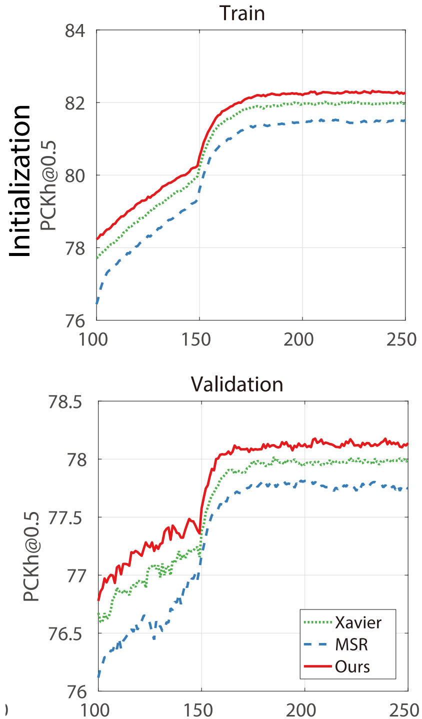
$$\Delta\mathbf{x}^{(l)} = \sum_{c=1}^{c_o^{(l)}} \mathbf{W}^{(l)T} \Delta\mathbf{y}^{(l)}$$

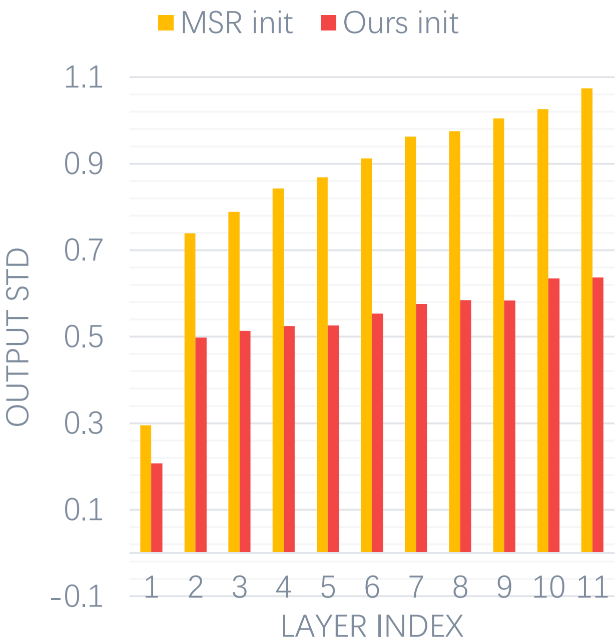
$$\Delta\mathbf{y}^{(l)} = f'(\mathbf{y}^{(l)}) \Delta\mathbf{x}^{(l+1)}$$

$$\alpha C_o^{(l)} n_o^{(l)} \mathrm{Var}\left(\omega^{(l)}\right) = 1$$

$*\ \alpha = 0.5$ *for ReLU and* 1 *for Tanh and Sigmoid.*

Initialization of Multi-Branch Networks





He, Kaiming, et al. "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification." ICCV. 2015.

Qualitative Results

MPII dataset

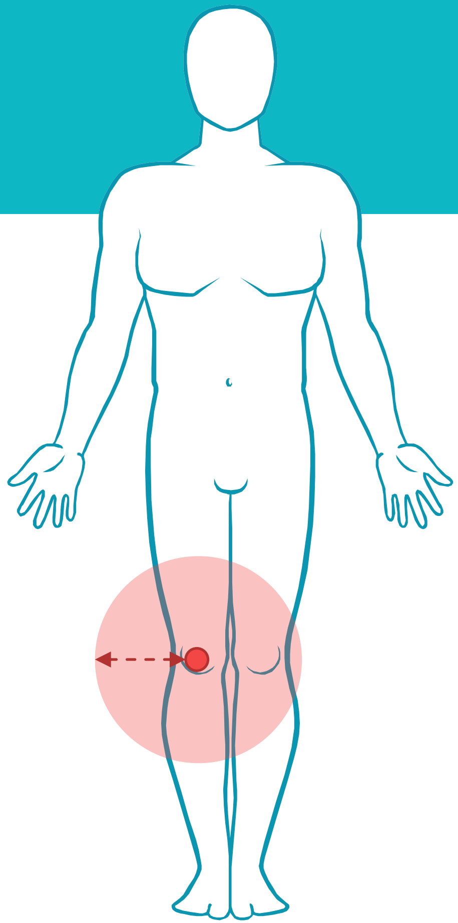
LSP dataset

Evaluation Metric

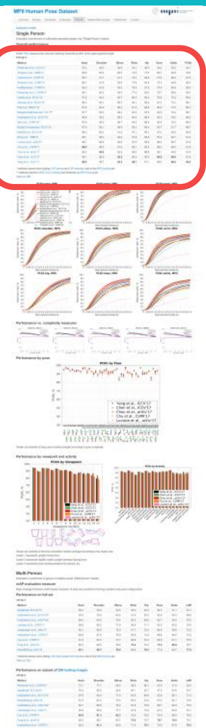
PCK:

Percentage of Correct Keypoints

$$\alpha \cdot \max(h, w)$$



Results on MPII Human Pose



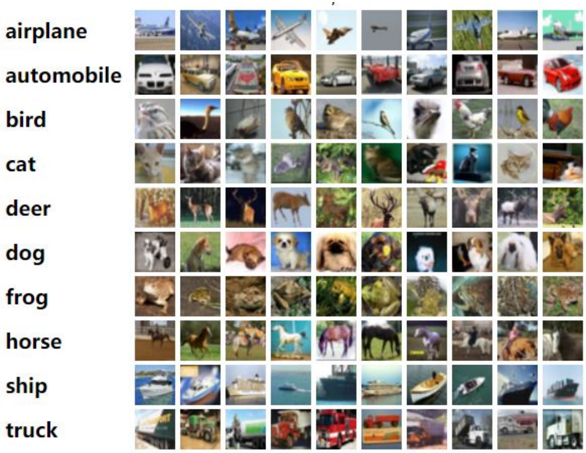
Method	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	PCKh
Pishchulin et al., ICCV'13	74.3	49.0	40.8	34.1	36.5	34.4	35.2	44.1
Tompson et al., NIPS'14	95.8	90.3	80.5	74.3	77.6	69.7	62.8	79.6
Carreira et al., CVPR'16	95.7	91.7	81.7	72.4	82.8	73.2	66.4	81.3
Tompson et al., CVPR'15	96.1	91.9	83.9	77.8	80.9	72.3	64.8	82.0
Hu&Ramanan., CVPR'16	95.0	91.6	83.0	76.6	81.9	74.5	69.5	82.4
Pishchulin et al., CVPR'16*	94.1	90.2	83.4	77.3	82.6	75.7	68.6	82.4
Lifshitz et al., ECCV'16	97.8	93.3	85.7	80.4	85.3	76.6	70.2	85.0
Gkioxary et al., ECCV'16	96.2	93.1	86.7	82.1	85.2	81.4	74.1	86.1
Rafi et al., BMVC'16	97.2	93.9	86.4	81.3	86.8	80.6	73.4	86.3
Belagiannis&Zisserman, FG'17**	97.7	95.0	88.2	83.0	87.9	82.6	78.4	88.1
Insafutdinov et al., ECCV'16	96.8	95.2	89.3	84.4	88.4	83.4	78.0	88.5
Wei et al., CVPR'16*	97.8	95.0	88.7	84.0	88.4	82.8	79.4	88.5
Bulat&Tzimiropoulos, ECCV'16	97.9	95.1	89.9	85.3	89.4	85.7	81.7	89.7
Newell et al., ECCV'16	98.2	96.3	91.2	87.1	90.1	87.4	83.6	90.9
Ning et al., TMM'17							82.7	91.2
Luvizon et al., arXiv'17							82.7	91.2
Chu et al., CVPR'17							85.0	91.5
Chou et al., arXiv'17		**96.8**	92.2	88.0	**91.3**	89.1	84.9	91.8
Chen et al., ICCV'17	[illegible]	96.5	**92.5**	88.5	90.2	**89.6**	**86.0**	91.9
Yang et al., ICCV'17	**98.5**	96.7	**92.5**	**88.7**	91.1	88.6	**86.0**	**92.0**

State-of-the-art performance

http://human-pose.mpi-inf.mpg.de/#results

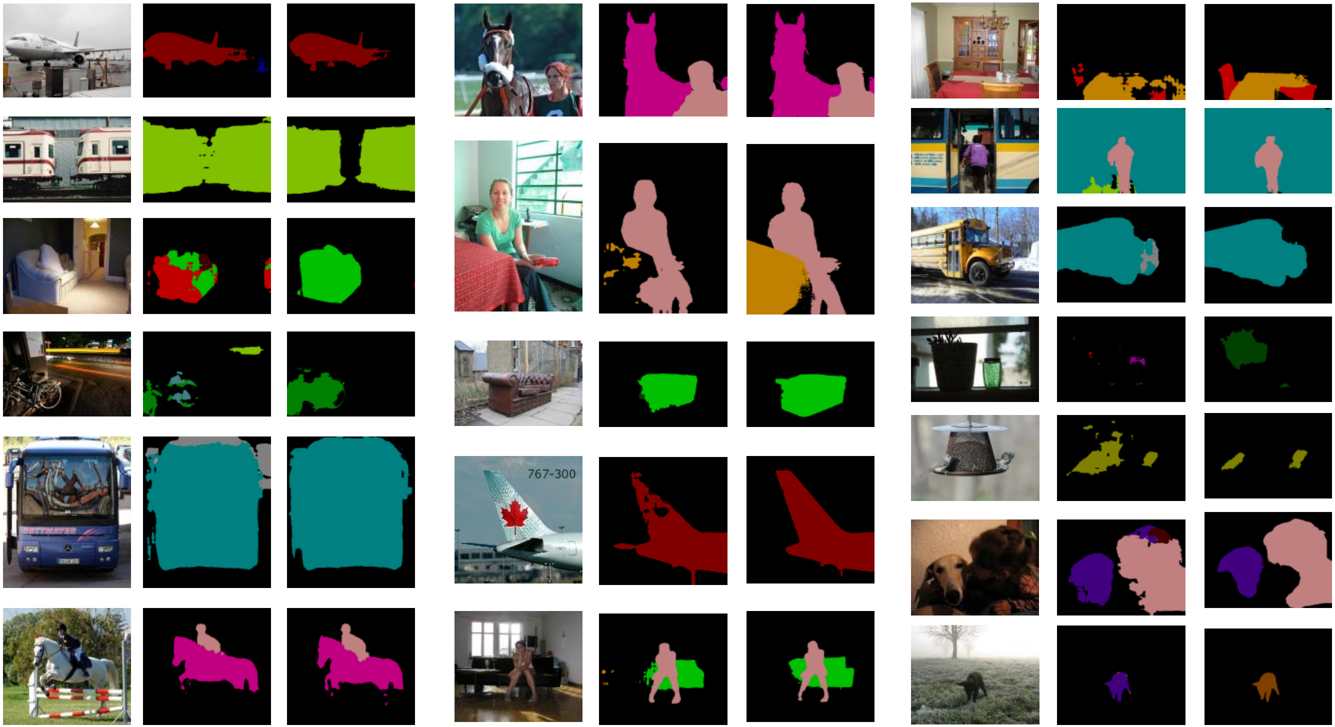
Image Classification

Top-1 Test Error on **CIFAR-10**



method	#params	GFLOPs	top-1
WRN-28-10 [64]	36.5	10.5	4.17
Ours-28-9	36.4	9.5	3.82
Ours-28-10	42.3	11.3	3.67
ResNeXt-29, 8 × 64d [56]	34.4	48.8	3.65
ResNeXt-29, 16 × 64d [56]	68.2	184.5	3.58
Ours-29, 8 × 64d	45.6	50.5	3.39
Ours-29, 16 × 64d	79.3	186.1	**3.30**

Semantic Segmentation:
PASCAL VOC 2012 dataset



(a) Image (b) DeepLab (c) DeepLap+PRM (a) Image (b) DeepLab (c) DeepLap+PRM (a) Image (b) DeepLab (c) DeepLap+PRM

Section Summary

- Feature pyramid module
- Generalizable for various networks and tasks
- Weight initialization for multi-branch networks

Learning Feature Pyramids for Human Pose Estimation

__Wei Yang__ , Shuang Li, Wanli Ouyang, Hongsheng Li, Xiaogang Wang

ICCV, 2017

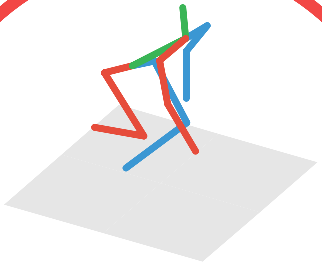
Outline

Scale

Feature pyramid learning

ICCV 2017

3D Pose

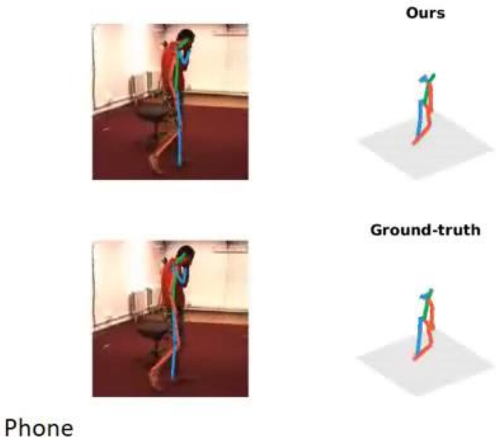


In-the-wild 3D pose estimation

CVPR 2018

Challenges: No Annotation

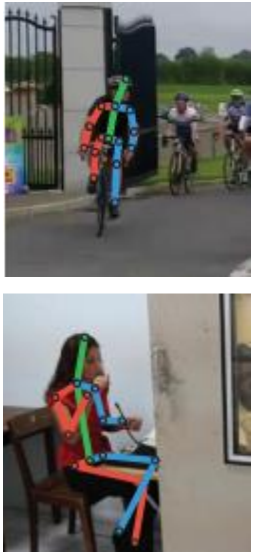
Constrained scenes



Domain

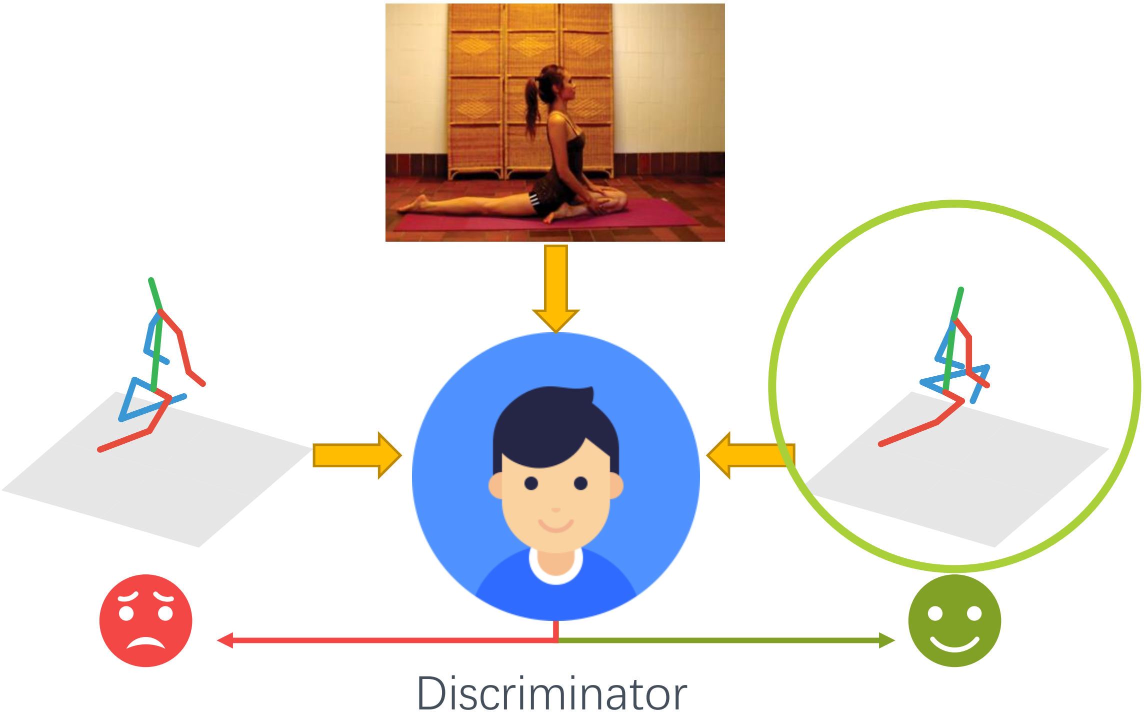
Discrepancy

In-the-wild scenes

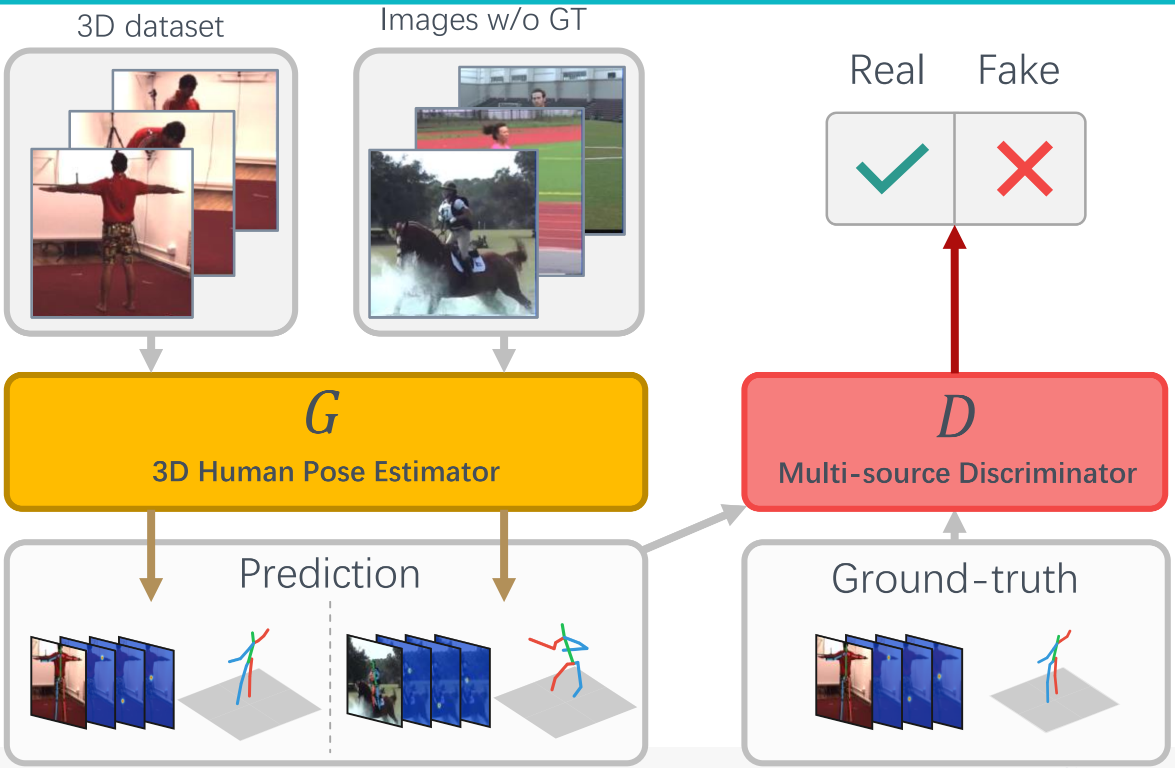


No annotation

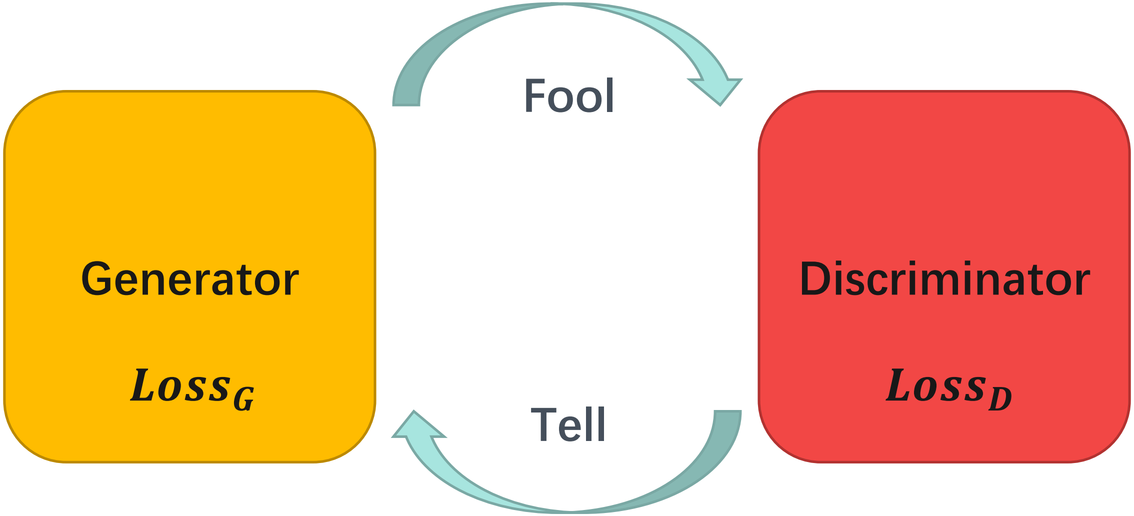
Which one is more plausible?



Weakly Supervised Adversarial Learning



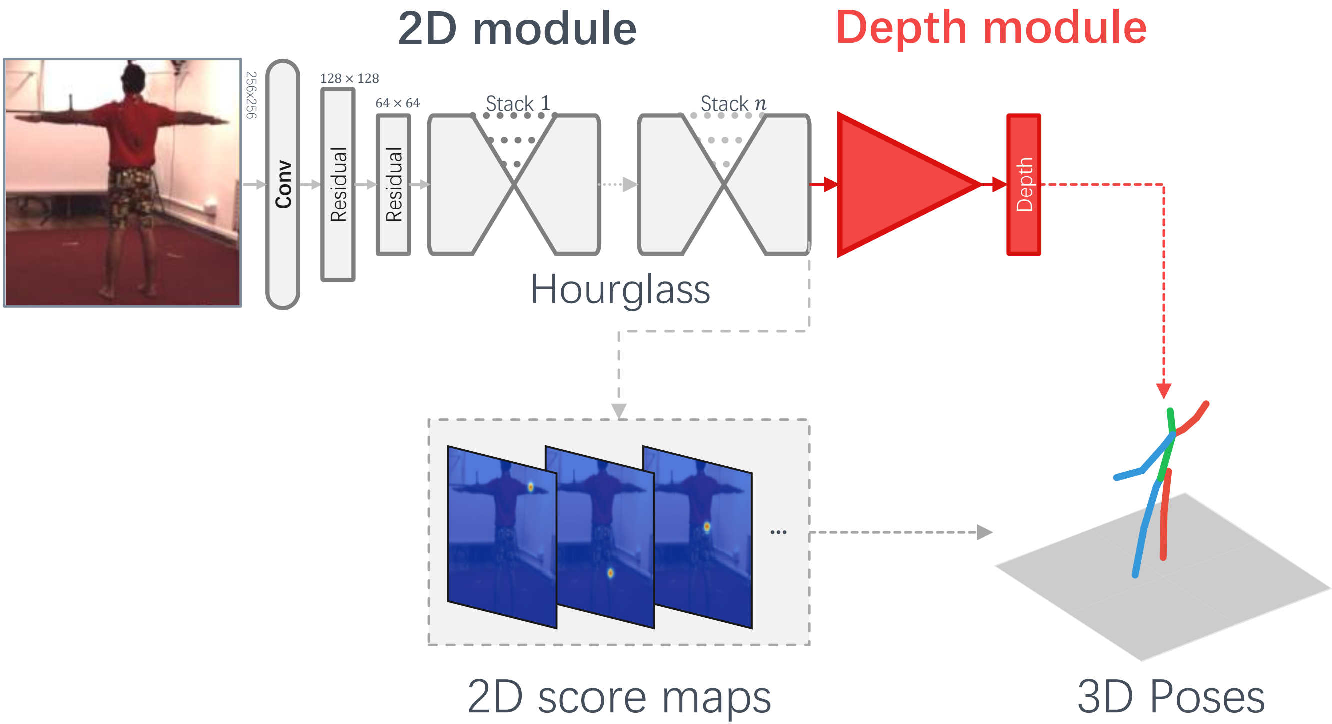
Adversarial Learning



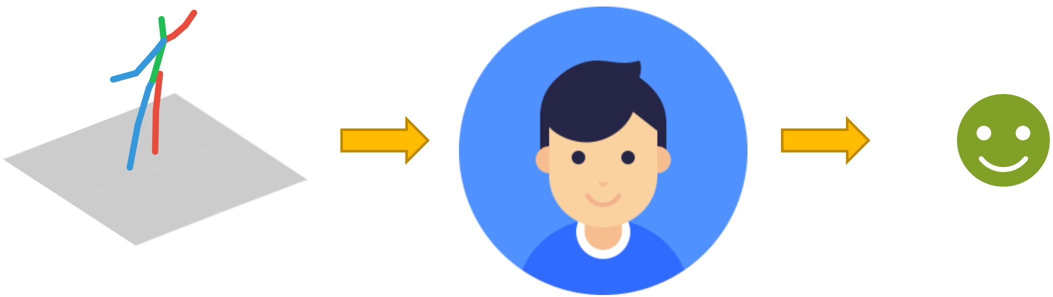
Euclidean Loss

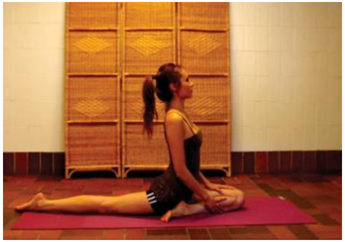
Classification Loss

Generator


2D module
Depth module
256x256
Conv
128 × 128
Residual
64 × 64
Residual
Stack 1
Stack n
Depth
Hourglass
...
2D score maps
3D Poses

Discriminator



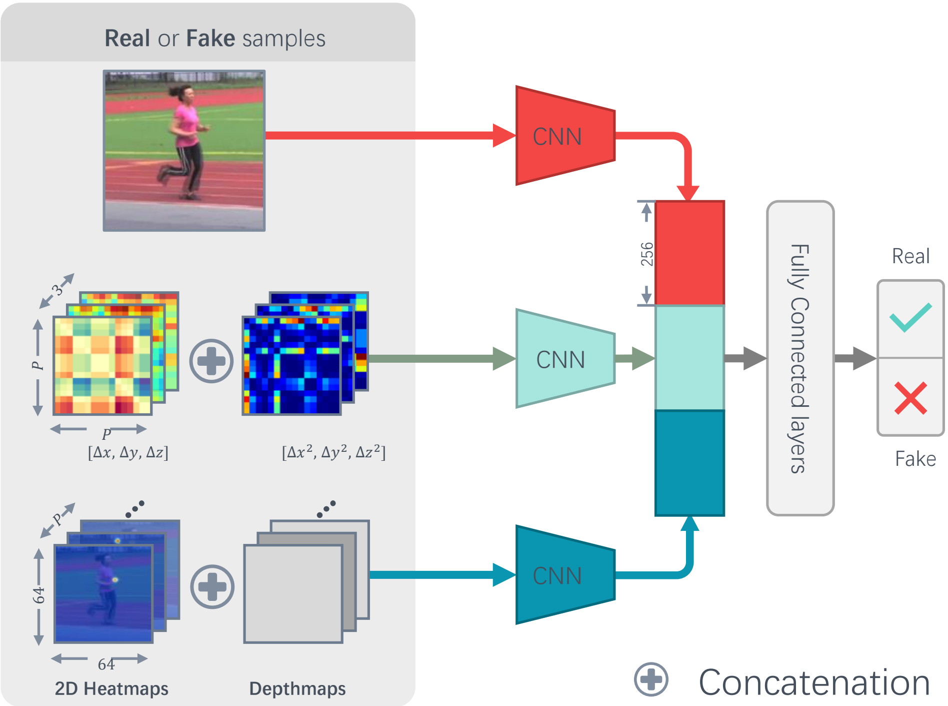


Multi-Source Discriminator

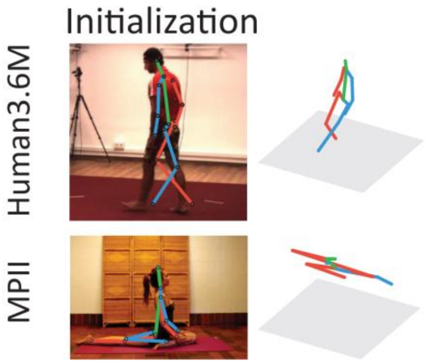
Image I

Geometric descriptor

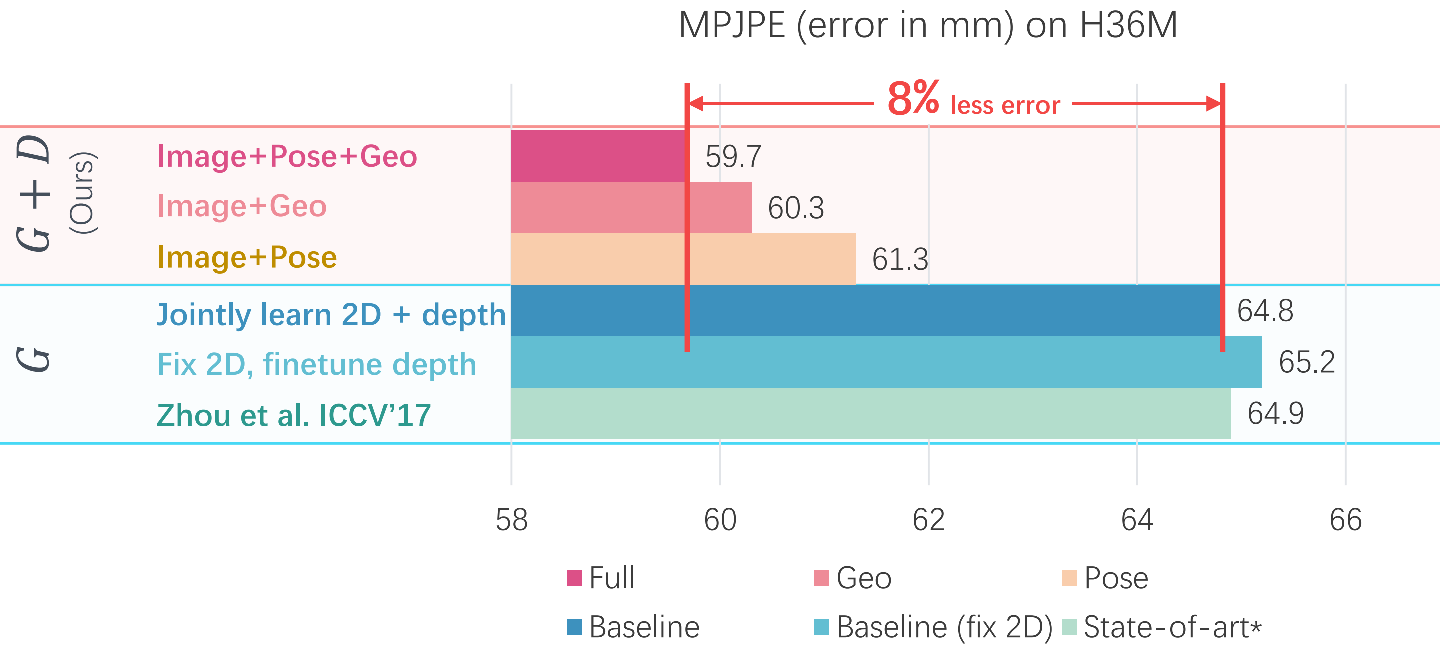
Raw poses



Effectiveness of Adversarial Learning

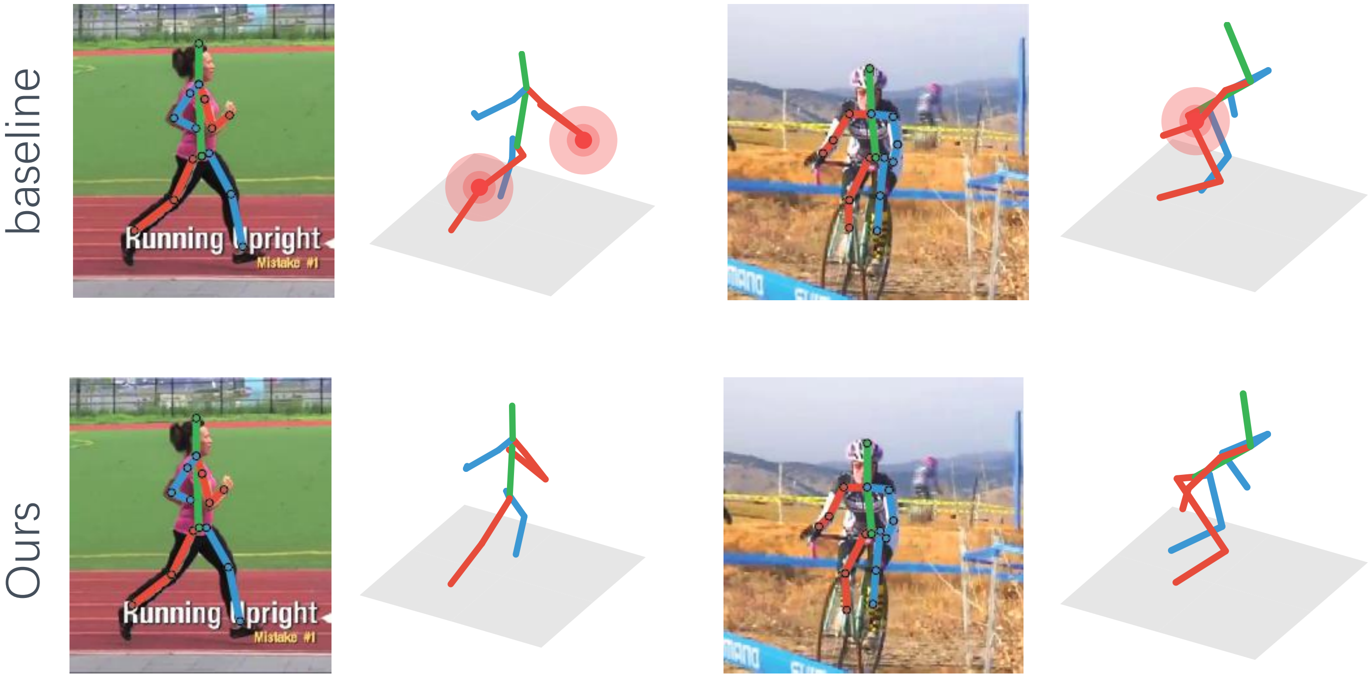


Ablation Study on H36M Dataset

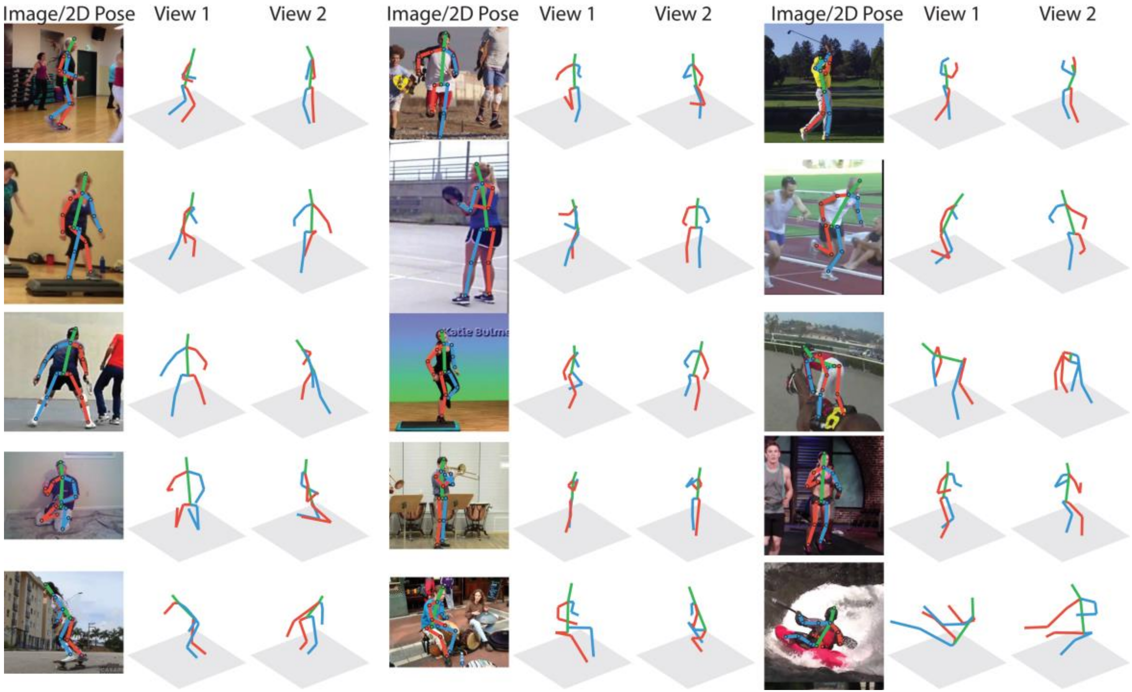


*Zhou et al. ICCV'17

Results on Images in the Wild



Multi-view Results



Section Summary

- Weakly supervised adversarial learning for 3D pose estimation in the wild
- Multi-source discriminator

3D Human Pose Estimation in the Wild by Adversarial Learning

Wei Yang , Wanli Ouyang, Xiaolong Wang, Hongsheng Li, Xiaogang Wang

CVPR, 2018

Code

- Open-source PyTorch code
  - https://github.com/bearpaw/pytorch-pose

- ICCV 17
  - https://github.com/bearpaw/PyraNet

Thanks!

wyang@ee.cuhk.edu.hk

http://www.ee.cuhk.edu.hk/~wyang/

@bearpaw