Multi-Scale Structure Learning for Human Pose Estimation

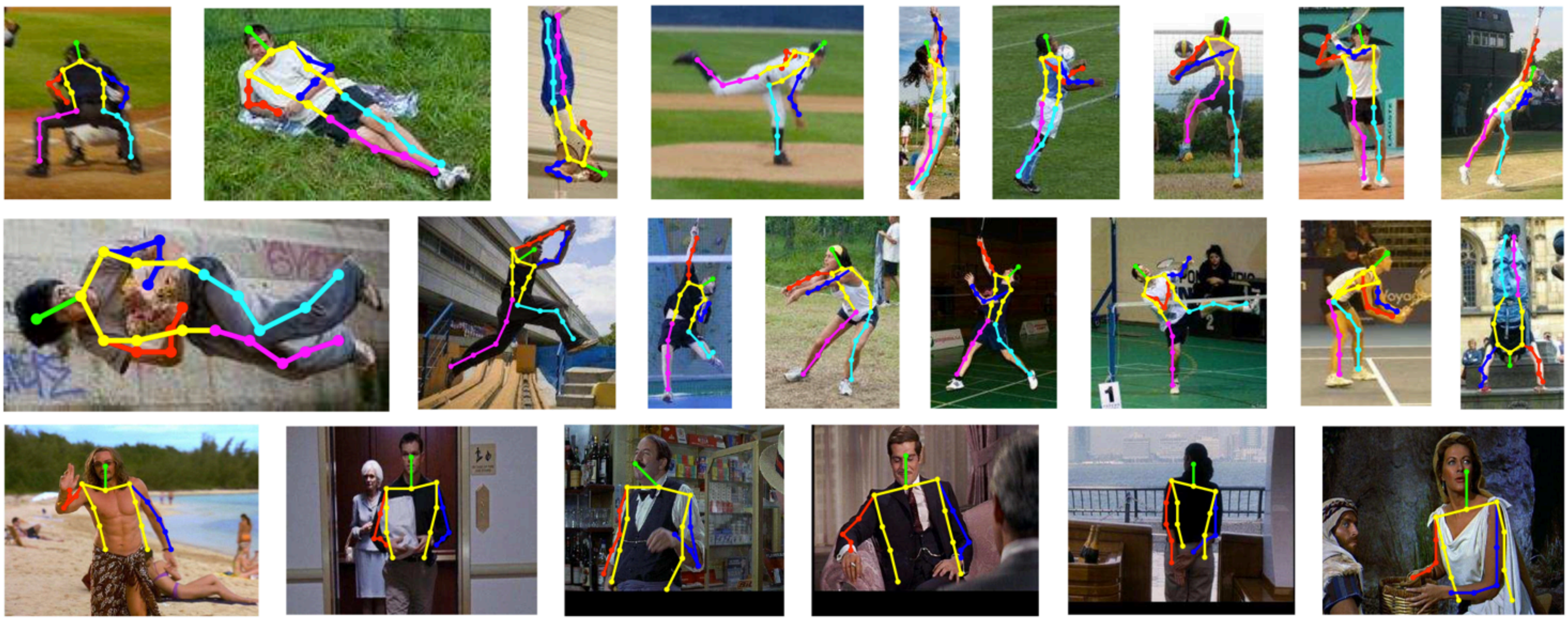
Lipeng Ke[1], Ming-Ching Chang[1], Honggang Qi[2] and Siwei Lyu[1]

[1]University at Albany, State University of New York, USA
[2]University of Chinese Academy of Sciences

UNIVERSITY AT ALBANY
State University of New York

CVML@UA

Pose Estimation

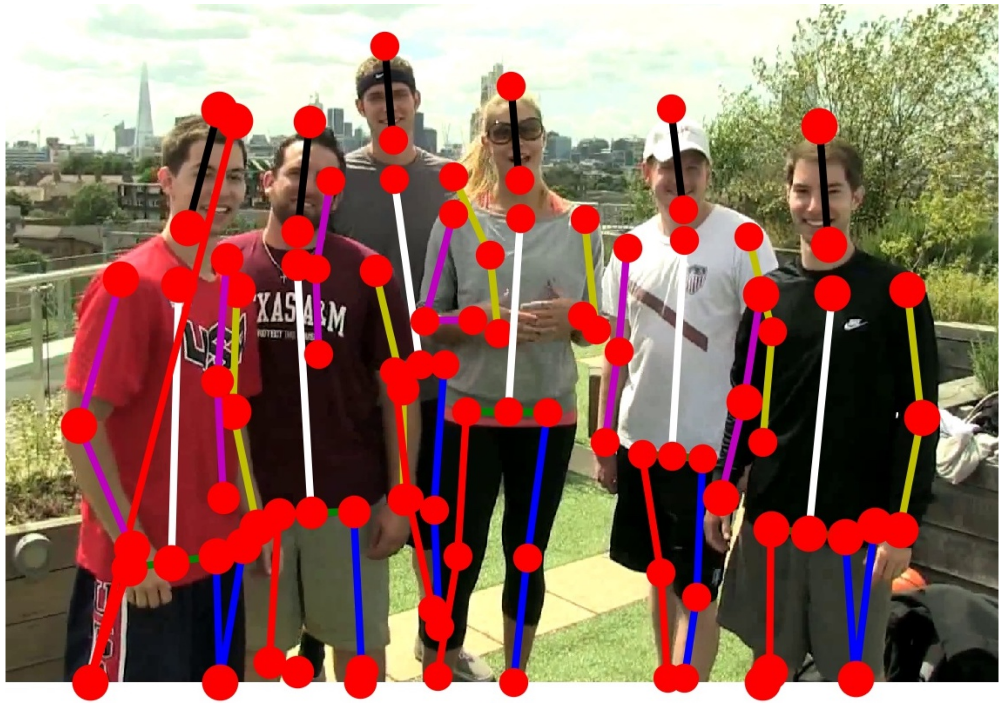


ICIP 2018 - ATHENS, GREECE

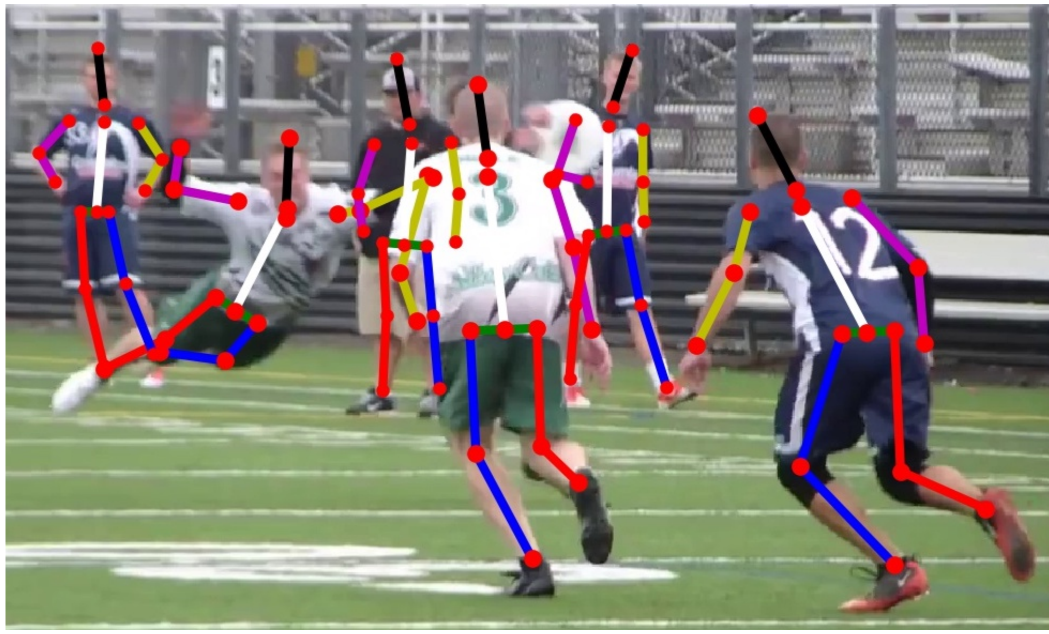
UNIVERSITY AT ALBANY
State University of New York

CVML@UA

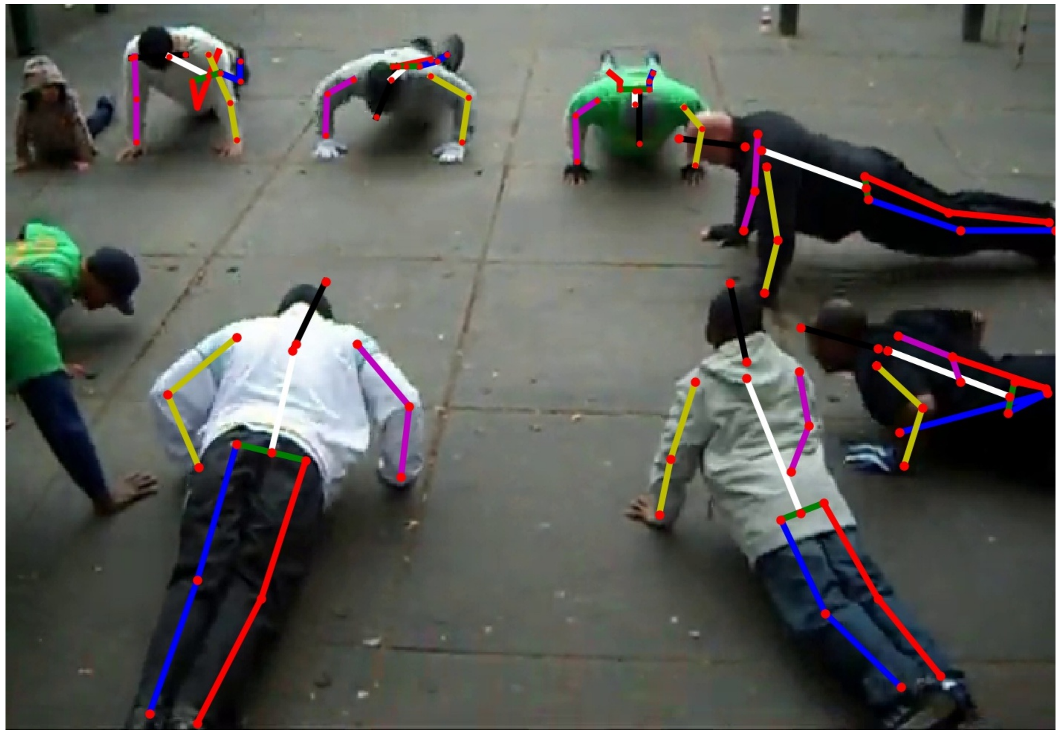
Challenges in Pose Estimation



Multi-Person



Occlusion



Scale Instability
Deformation

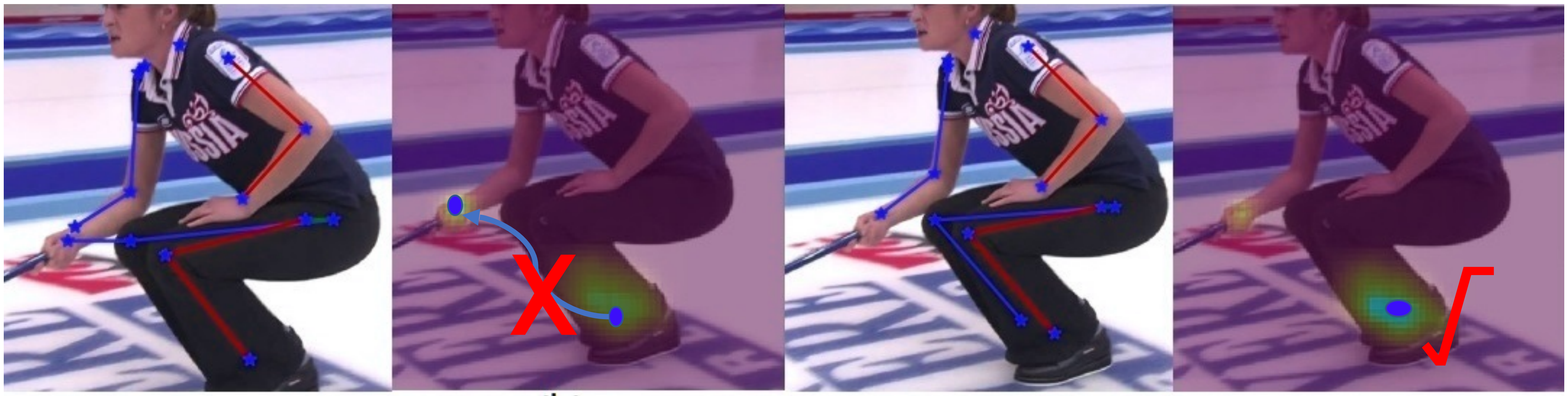
UNIVERSITY AT ALBANY
State University of New York

CVML@UA

Method Overview

256x256

Multi-scale regression



(a) (b) (c) (d)

Before Multi-Scale Regression
High Activation on both ankle &wrist
Left ankle ——> Right Wrist

After Multi-Scale Regression
Lower the activation on left wrist
Correctly Predict left ankle Pose

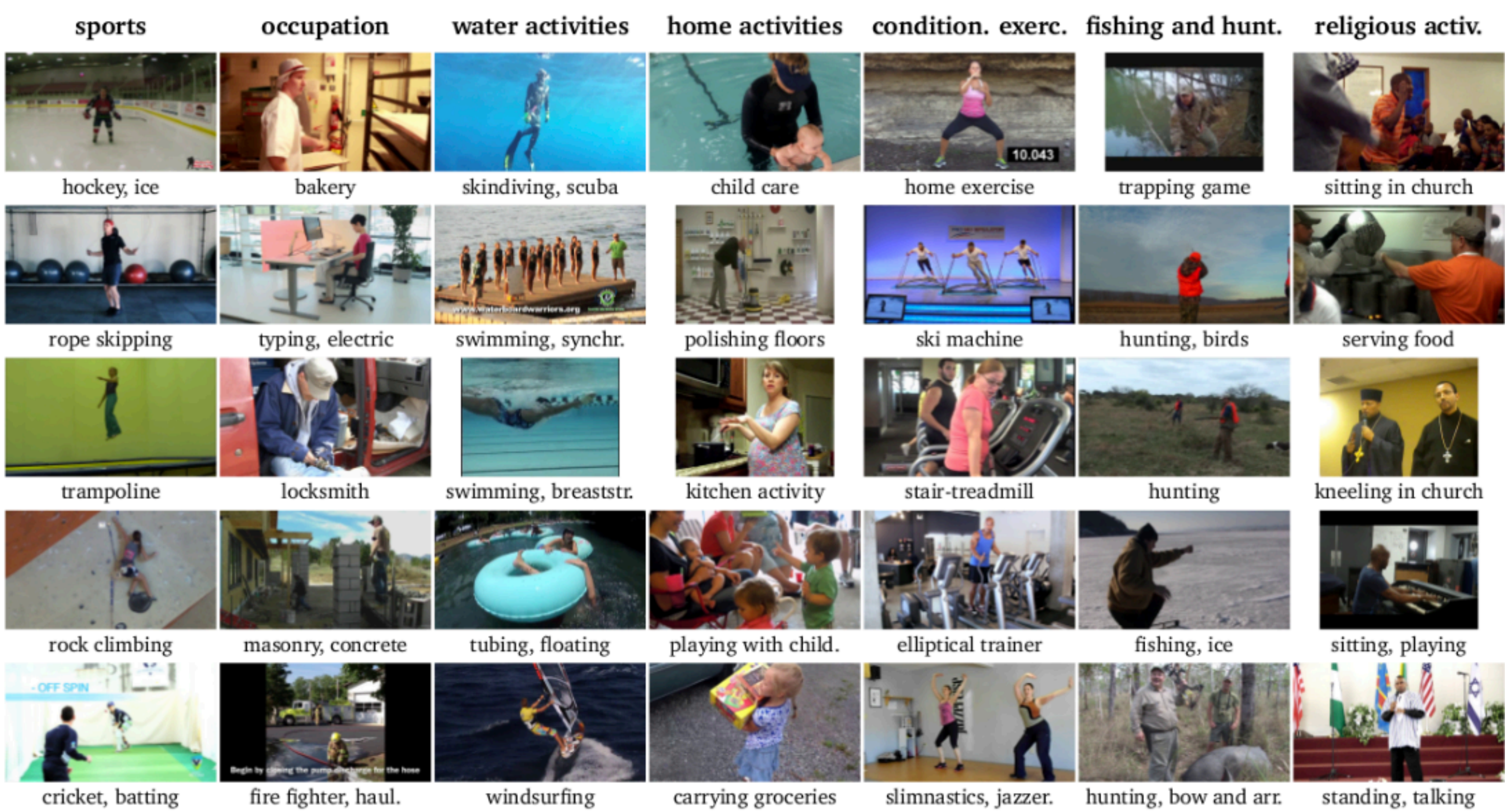
UNIVERSITY AT ALBANY
State University of New York

CVML@UA

Dataset

MPII Human Pose dataset is a state of the art benchmark for evaluation of articulated human pose estimation

- ~ 25K image, > 40K individuals, 410 different activity

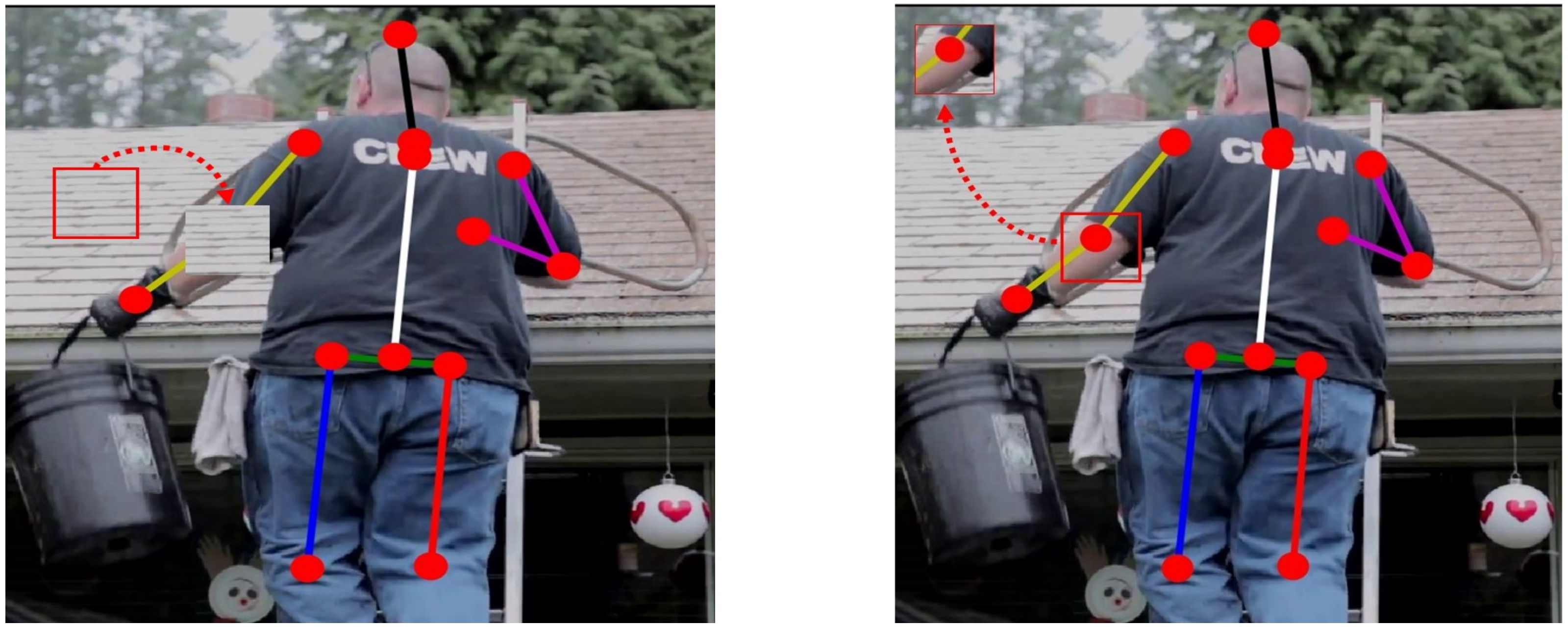


UNIVERSITY AT ALBANY
State University of New York

CVML@UA

Training data augmentation

- MPII Training data lack examples with occluded keypoints and multi-person scenario
  - We argument training data using key point masking by copy-move patches in the image



Model training

- Device: 4 GTX 1080 Ti
- Network: 8 x MSS-Net + MSR-Net layers
- Optimizer: SGD + ADAM
- Scheme: 150 epoch MSS-Net —> 75 epoch MSR-Net —> 75 epoch Jointly Training
- Data augmentation: key point masking in 75 epoch Jointly Training

UNIVERSITY AT ALBANY
State University of New York

CVML@UA

Quantitative Results

On the FLIC dataset

	Elbow	Wrist
Tompson *et al.* CVPR'15 [23]	93.1	92.4
Wei *et al.* CVPR'16 [26]	97.8	95.0
Newell*et al.* ECCV'16 [18]	99.0	97.0
Our model	**99.2**	**97.3**

On the MPII dataset

	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Total	AUC
Our method	**98.5**	**96.8**	**92.7**	88.4	90.6	89.3	**86.3**	**92.1**	63.8
Chen *et al.* ICCV'17 [7]	98.1	96.5	92.5	**88.5**	90.2	**89.6**	86.0	91.9	61.6
Chou *et al.* arXiv'17 [9]	98.2	**96.8**	92.2	88.0	**91.3**	89.1	84.9	91.8	63.9
Chu CVPR'17 [12]	**98.5**	96.3	91.9	88.1	90.6	88.0	85.0	91.5	63.8
Luvizon *et al.* arXiv'17 [17]	98.1	96.6	92.0	87.5	90.6	88.0	82.7	91.2	63.9
Ning *et al.* TMM'17 [19]	98.1	96.3	92.2	87.8	90.6	87.6	82.7	91.2	63.6
Newell ECCV'16 [18]	98.2	96.3	91.2	87.1	90.1	87.4	83.6	90.9	62.9
Bulat ECCV'16 [4]	97.9	95.1	89.9	85.3	89.4	85.7	81.7	89.7	59.6
Wei CVPR'16 [26]	97.8	95.0	88.7	84.0	88.4	82.8	79.4	88.5	61.4
Insafutdinov ECCV'16 [13]	96.8	95.2	89.3	84.4	88.4	83.4	78.0	88.5	60.8
Belagiannis FG'17 [2]	97.7	95.0	88.2	83.0	87.9	82.6	78.4	88.1	58.8

UNIVERSITY AT ALBANY
State University of New York

CVML@UA

Examples

UNIVERSITY AT ALBANY
State University of New York

CVML@UA

Summary

- A new method for DL based pose estimation
  - Multi-Scale Supervision helps multi-scale feature learning
  - Multi-Scale Regression optimize both scale and structure
  - Embedded structure loss learns human skeleton in DNNs
  - Hard sampling adds robustness in challenge case

- An improved model published in our ECCV 18 paper leads performance on the MPII human pose estimation challenge

Thank you

UNIVERSITY AT ALBANY
State University of New York

CVML@UA