ECE 6554:Advanced Computer Vision

Pose Estimation

Sujay Yadawadkar, Virginia Tech
,

Agenda:

- Pose Estimation:
- Part Based Models for Pose Estimation
- Pose Estimation with Convolutional Neural Networks (Deep pose)
- Pose Estimation with Sequential Prediction (Pose Machines)

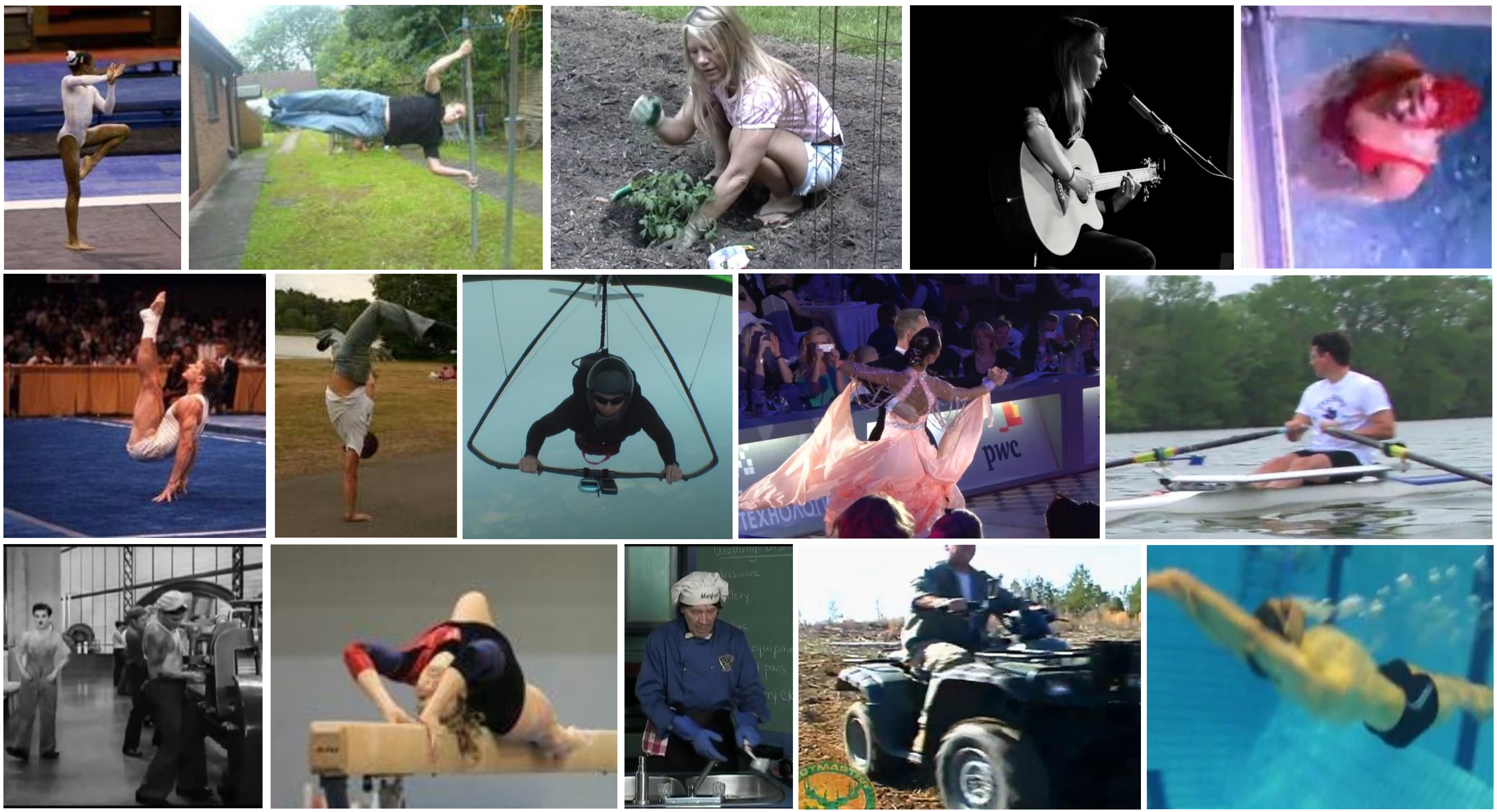
Estimating Articulated Poses

Localizing Body Joints from Monocular Images

Estimating Articulated Poses from Monocular Images

Why it is Hard?

large variance occlusion L/R ambiguity

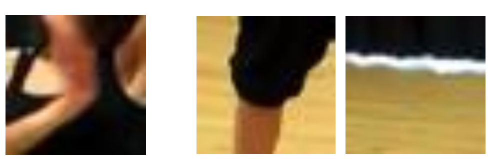


Estimating Articulated Poses from Monocular Images
Direct Mapping

$$\xrightarrow{\quad f \quad} \begin{pmatrix} x_1 \\ y_1 \\ \vdots \\ x_P \\ y_P \end{pmatrix}$$

Part-based Models

Recognizing Local Appearance



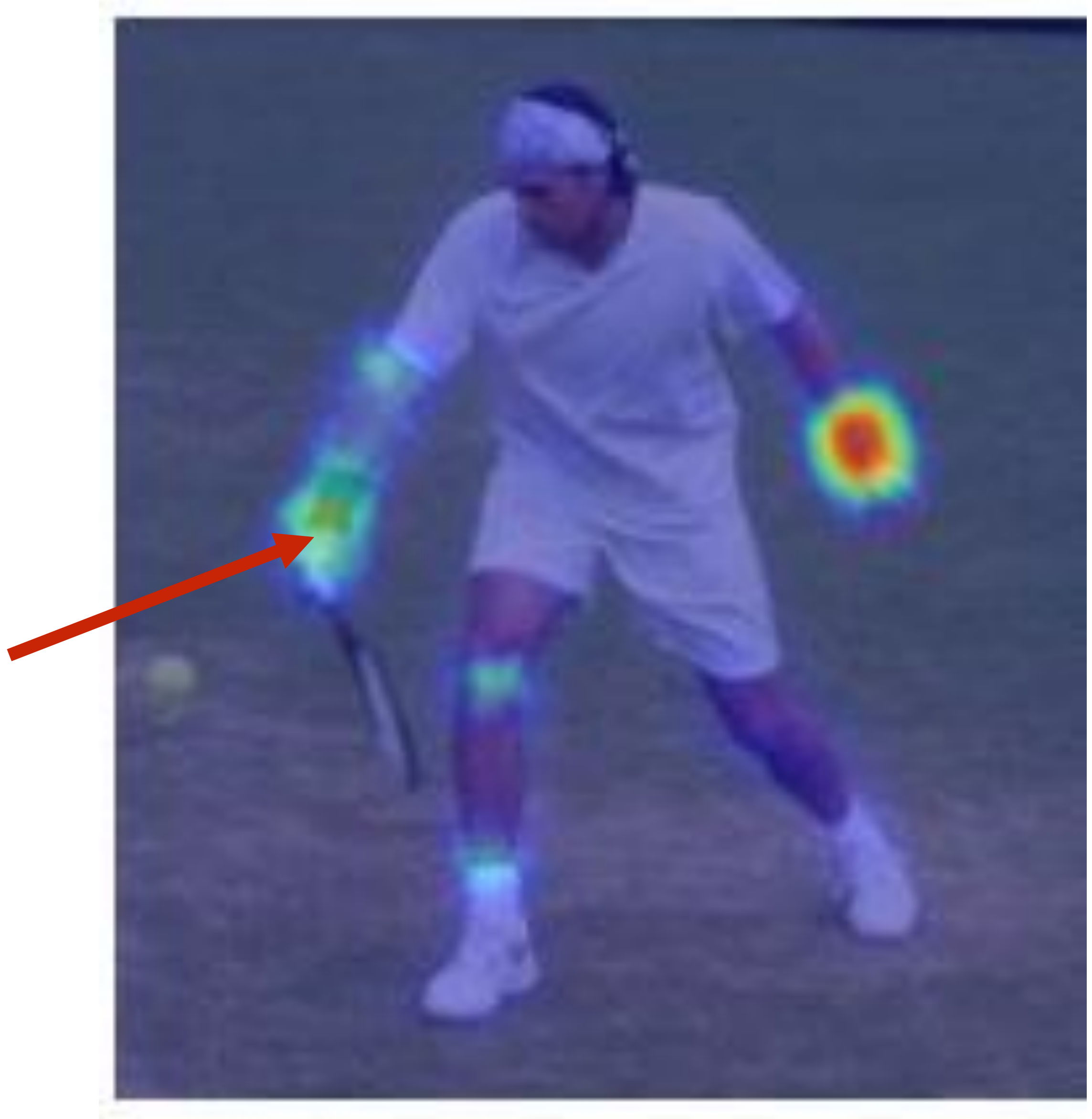
Part detector for wrist

0.999 0.001 0.001

Confidence maps

Non-parametric Uncertainty on Confidence Maps

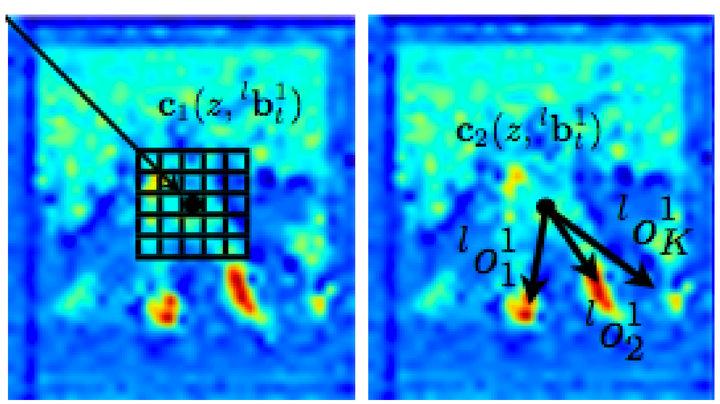
right wrist



Extracting Features from Confidence Maps Loses Uncertainty

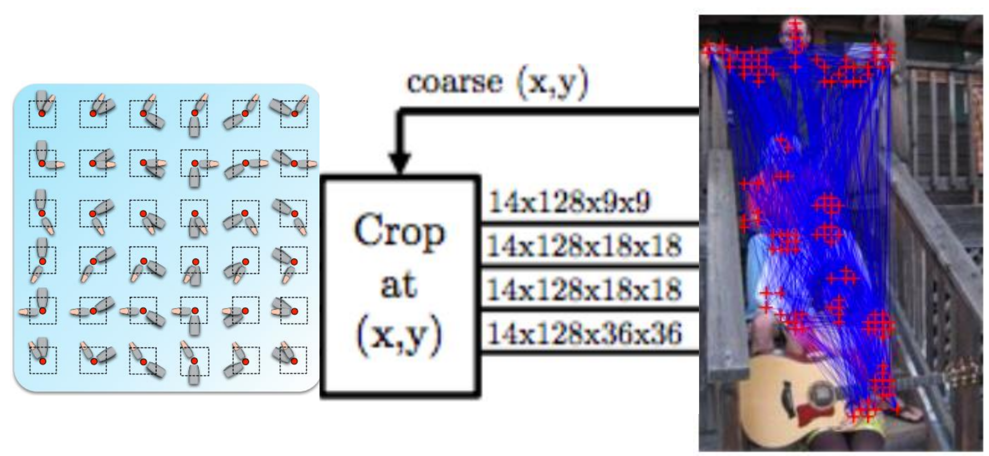
Hand-crafted
Context feature

[Ramakrishna14]



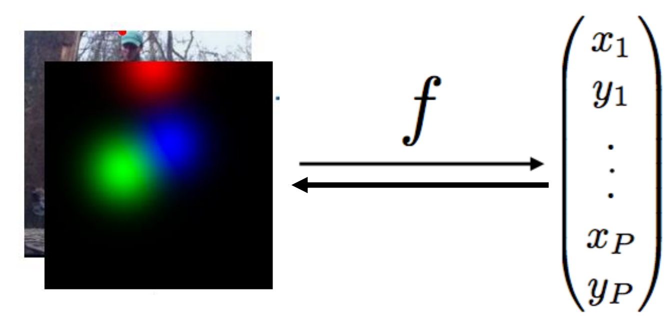
Peak Candidates for
Graphical Models

[Chen14] [Tompson15] [Pishchulin16]

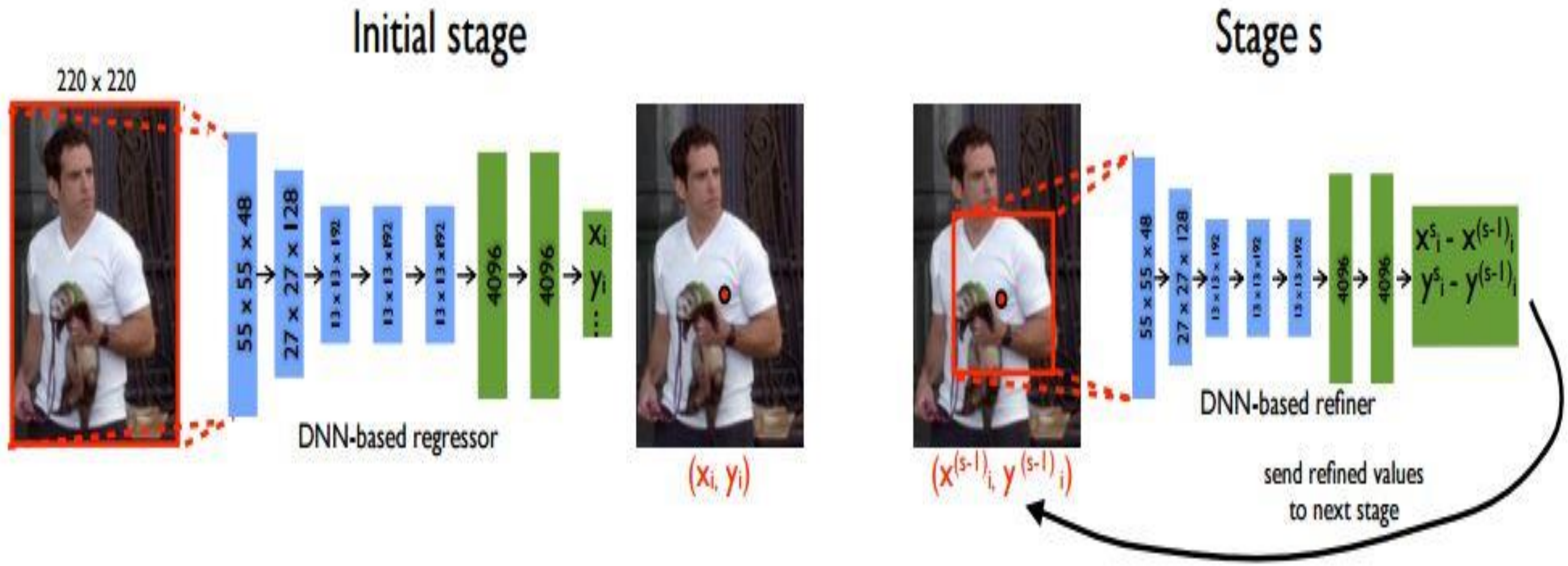


Regress to
Displacement

[Carriera16]



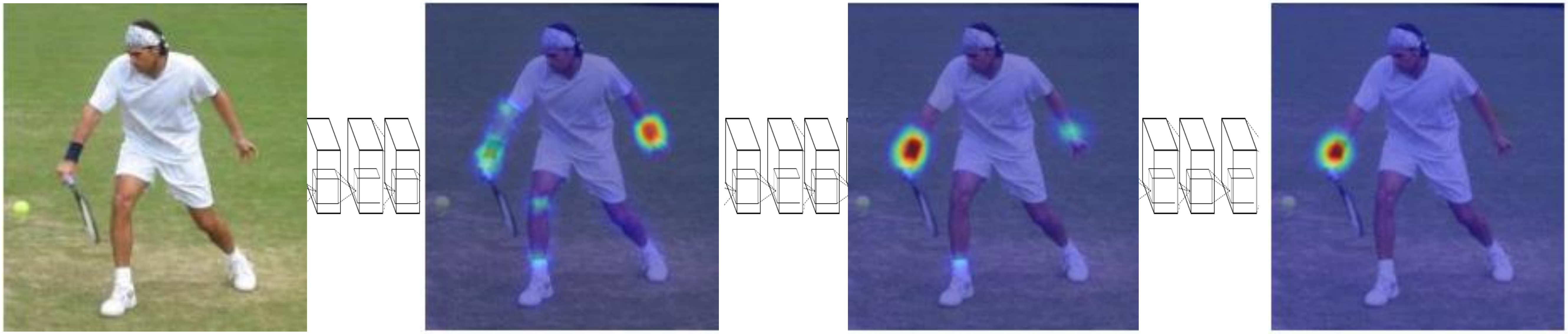
Pose Estimation with CNN



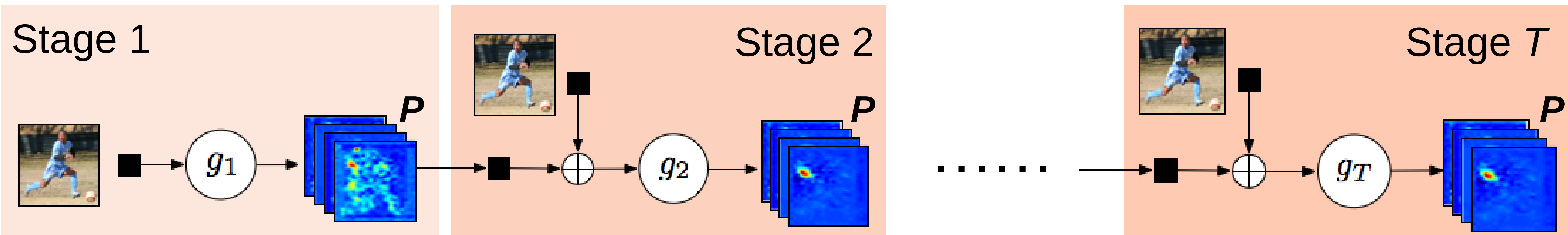
- Consider Pose Estimation as a regression problem.
- Loss function: L2 distance between ground truth of the pose vector and estimated pose vector.

Convolutional Pose Machines

1. Capture local appearance with CNNs
2. CNNs on confidence maps to capture long-range part dependencies (preserve uncertainty)
3. Iteratively refine confidence maps with global cues

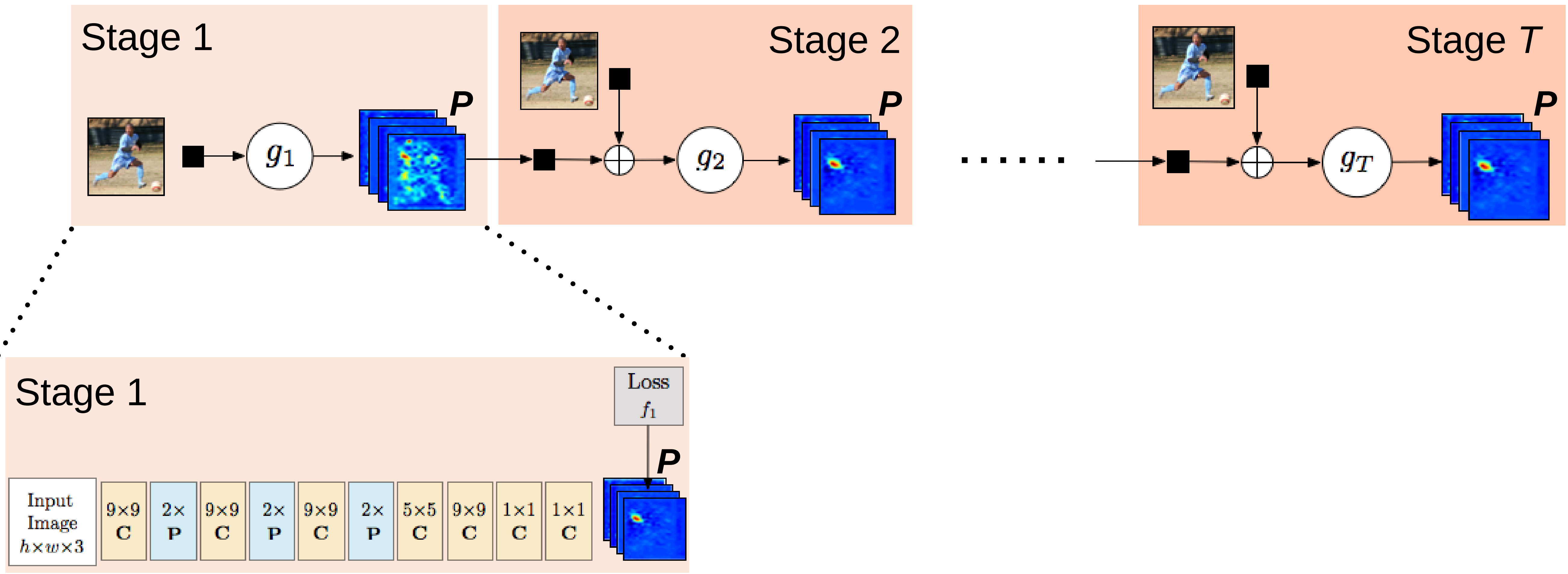


Convolutional Pose Machines (CPMs)



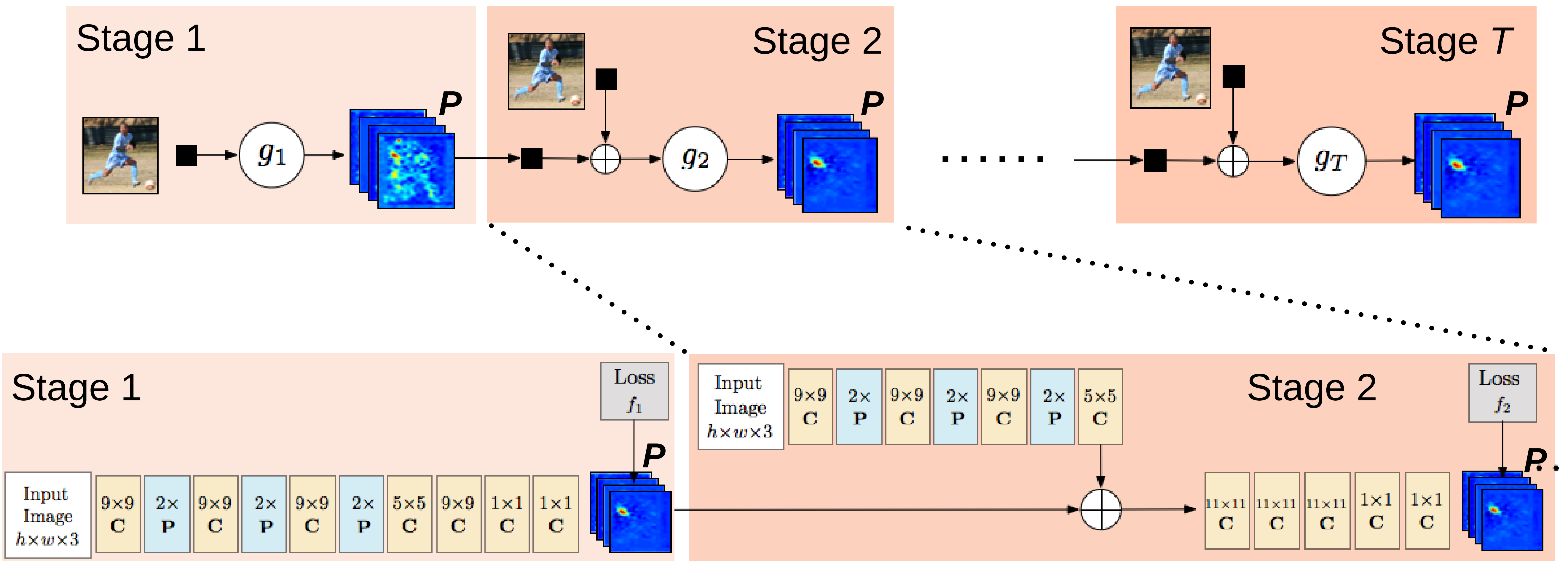
Convolutional Pose Machines

Capturing Local Appearance by FCNN



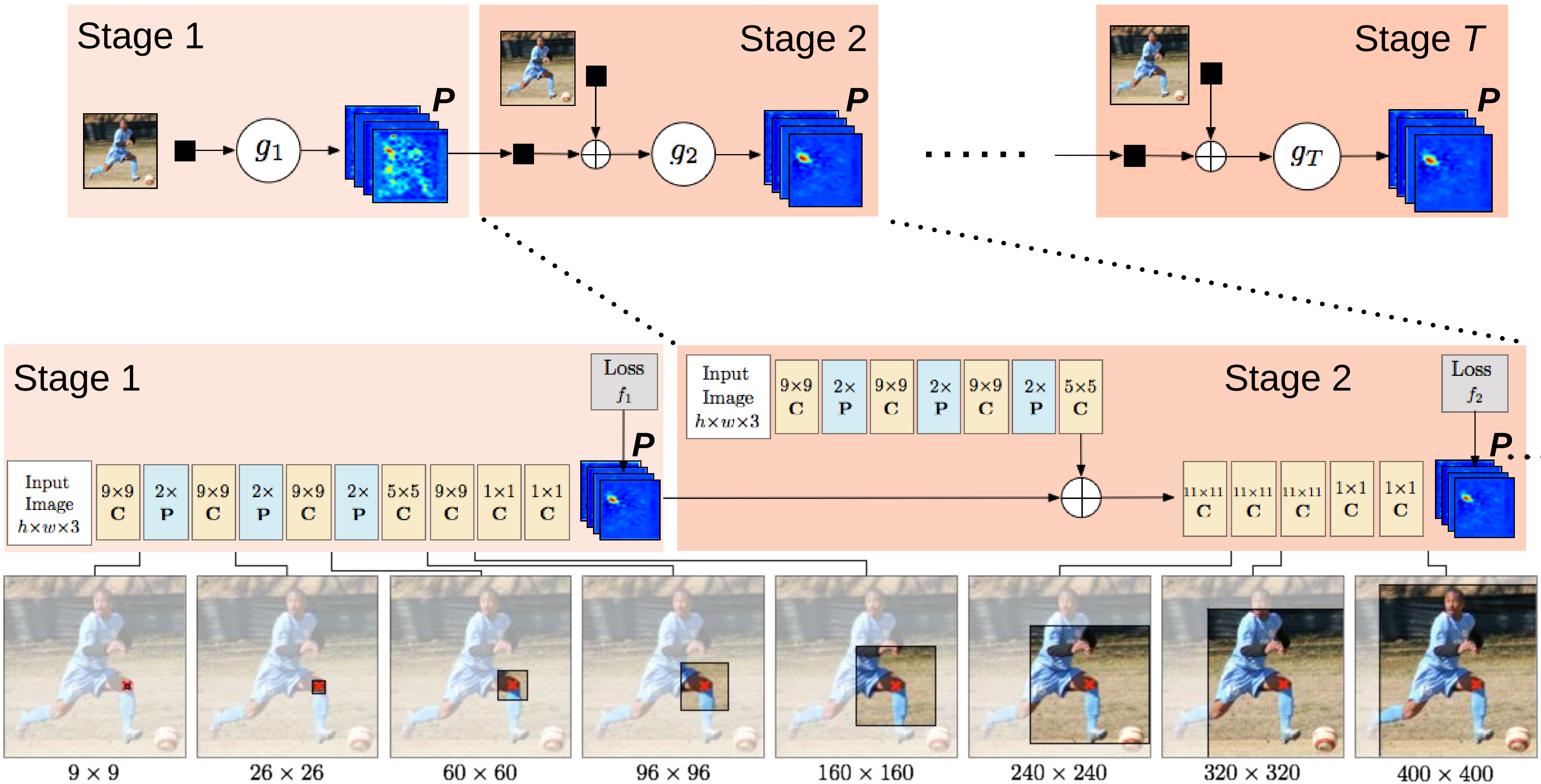
Convolutional Pose Machines

Learning Image-dependent Spatial Model



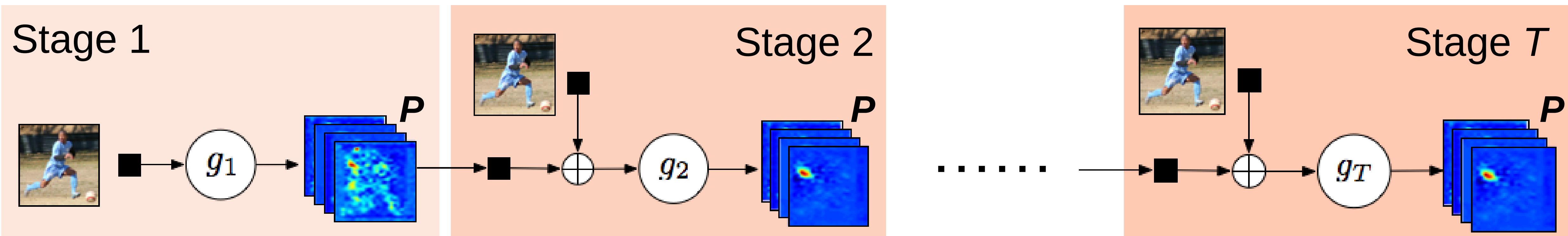
Convolutional Pose Machines

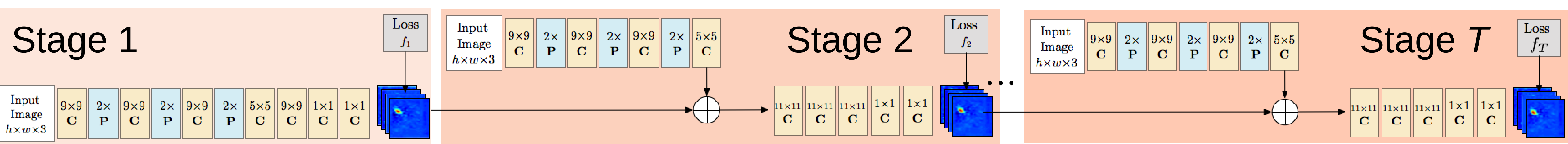
Large Receptive Field



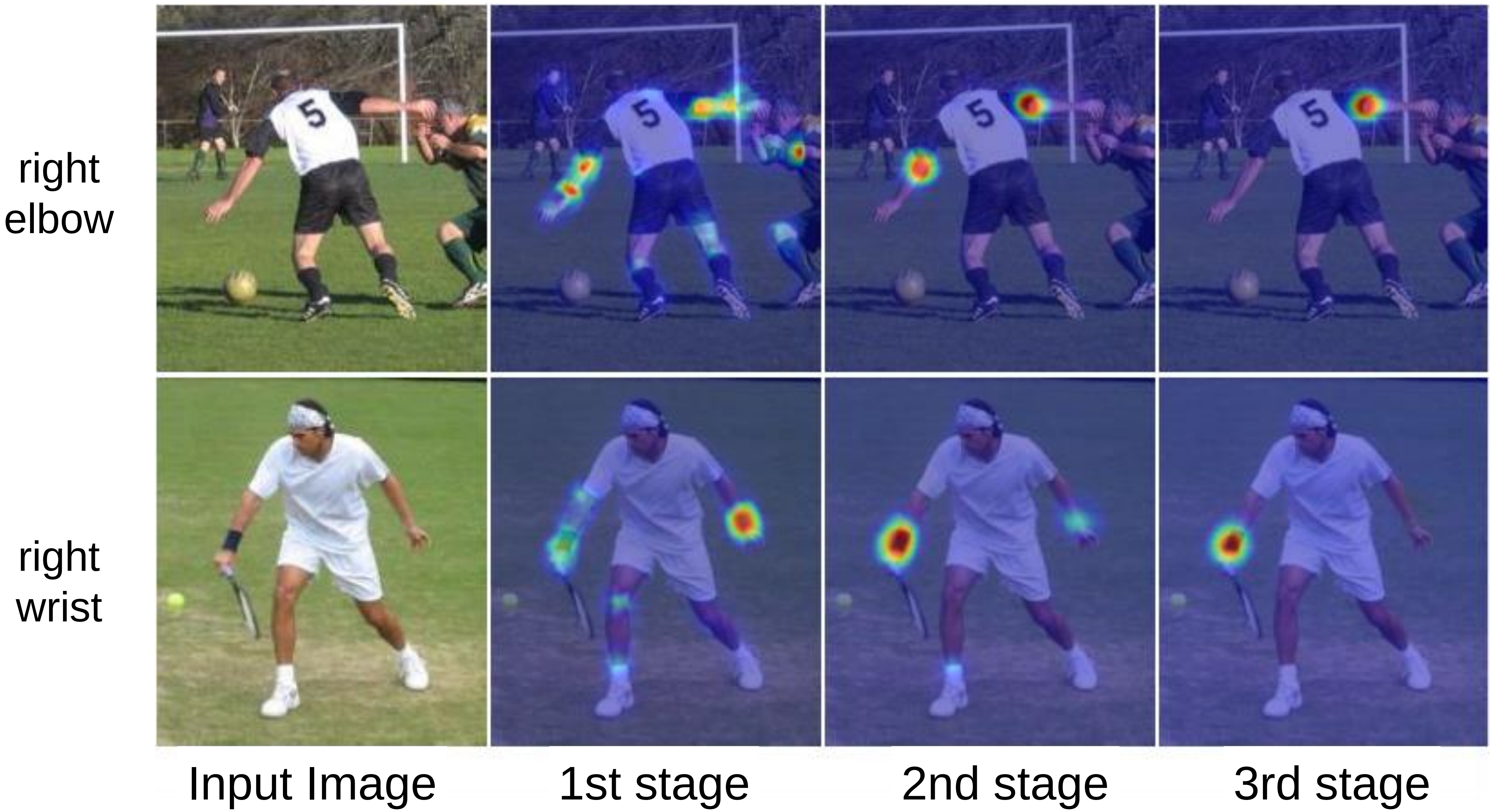
Convolutional Pose Machines

Overall Architecture





Iteratively Refined Confidence Maps



Iteratively Refined Confidence Maps

Recover from False Negative

1st stage

R. Elbow

2nd stage

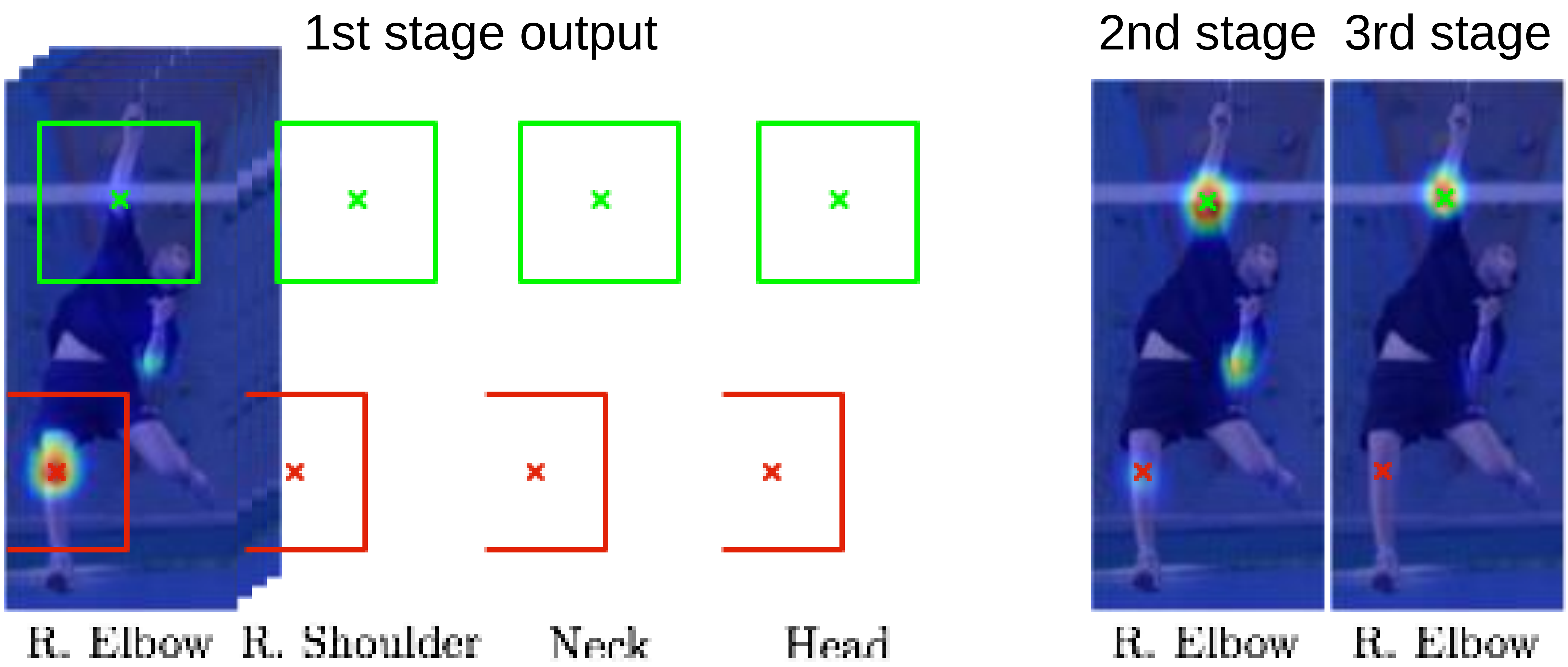
R. Elbow

3rd stage

R. Elbow

Iterratively Refined Confidence Maps

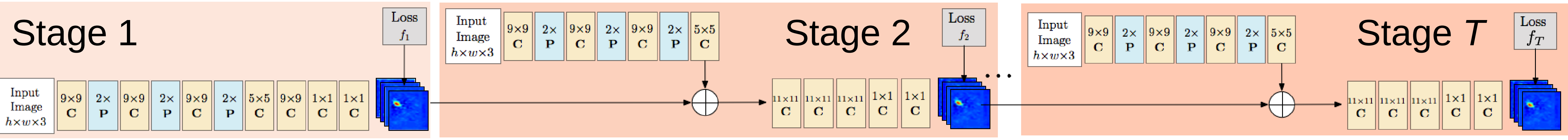
Recover from False Negative

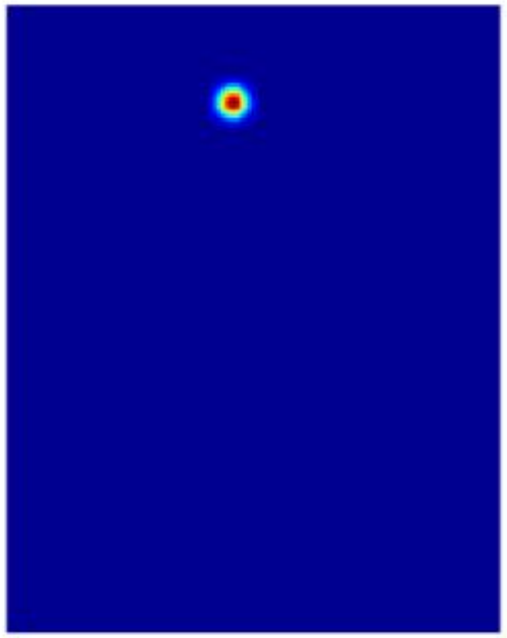
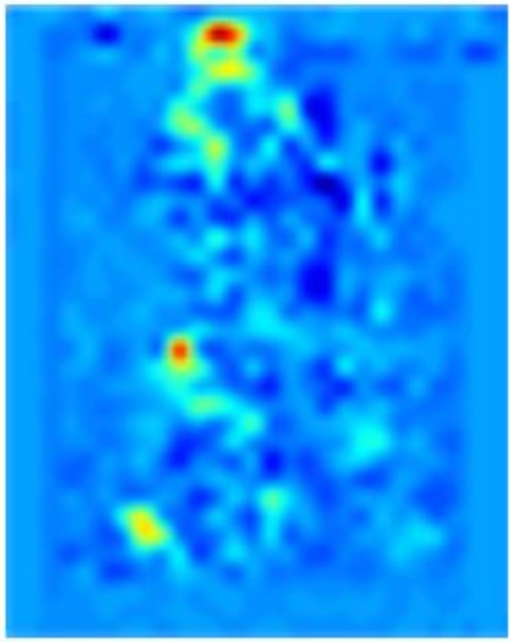


Iteratively Refined Confidence Maps

Training CPMs

Ideal Confidence Maps for Intermediate Supervisions

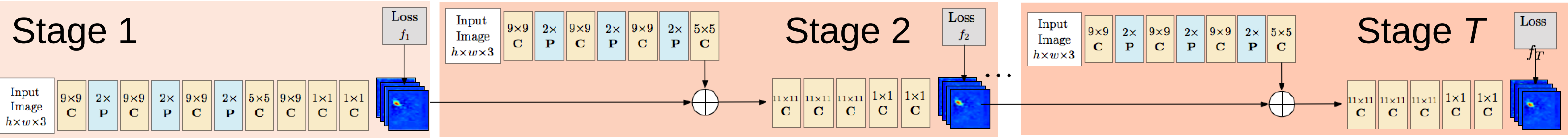


$f_t = \|$  $-$  $\|_F$

overall loss $\mathcal{F} = \sum_{t=1}^{T} f_t$

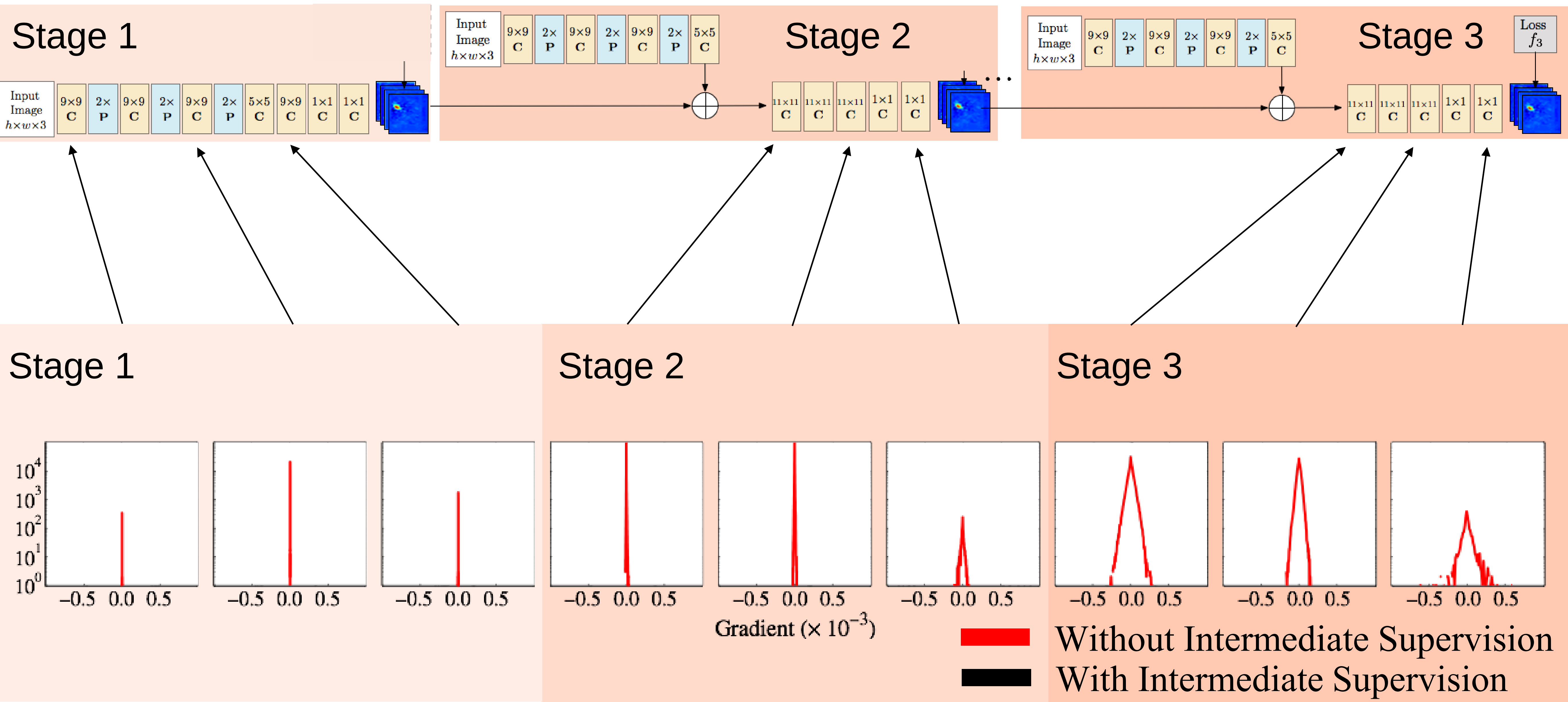
Training CPMs

Joint training with Intermediate Supervisions



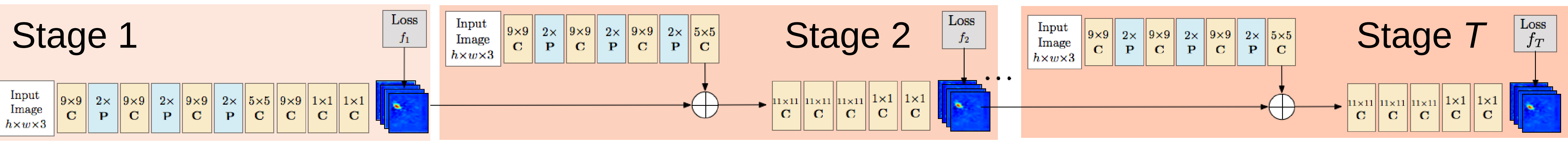
Training CPMs

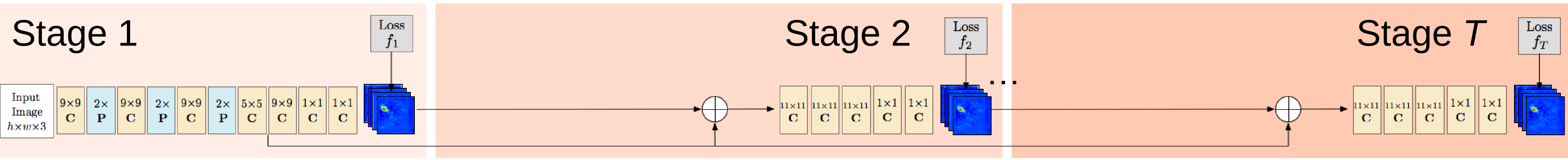
Intermediate Supervisions Resolves Gradient Vanishing



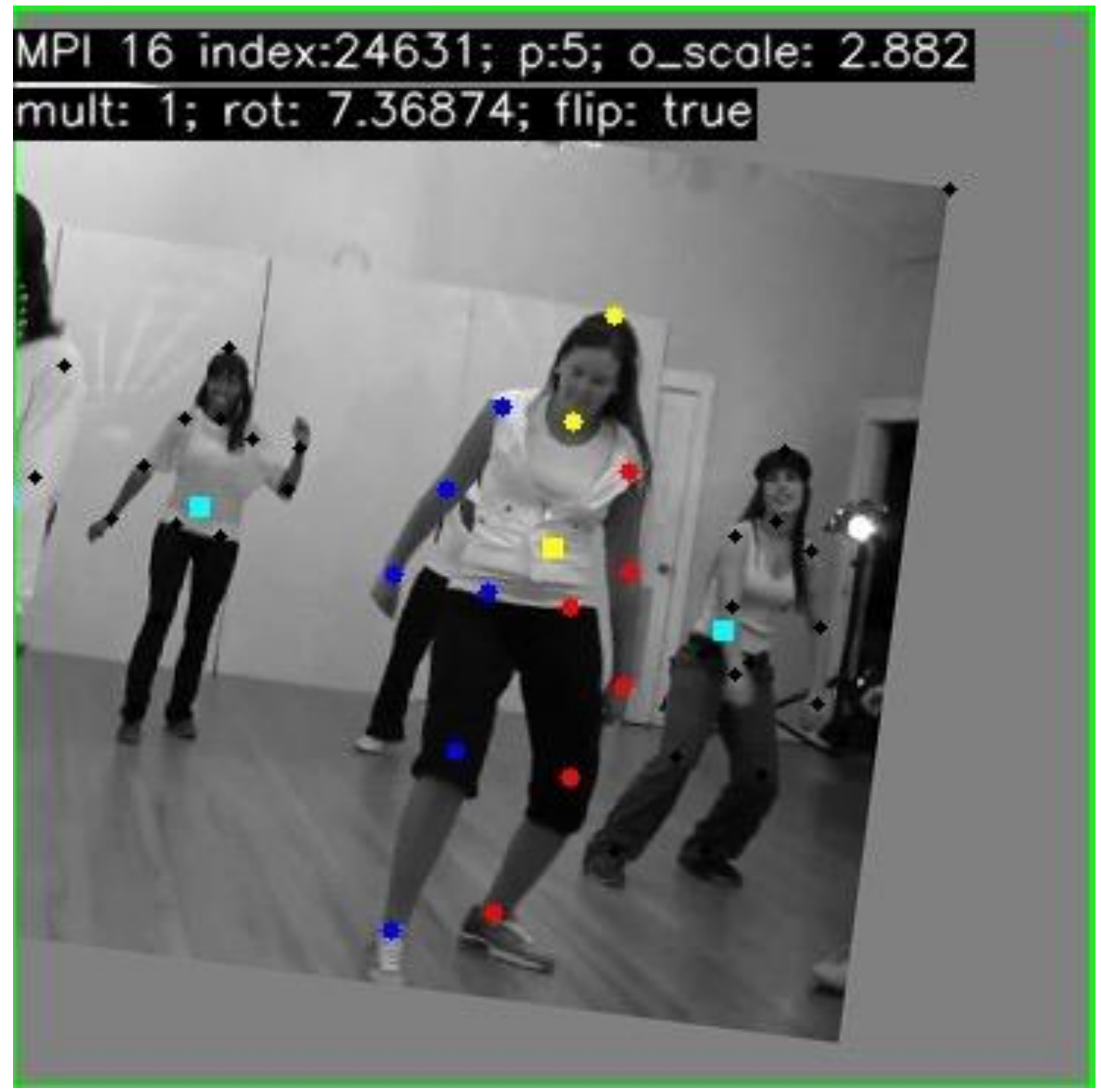
Convolutional Pose Machines

Overall Architecture with Shared Image Features





Training CPMs
Data Prepare and Augmentation



Analysis and Results

Benchmark Datasets

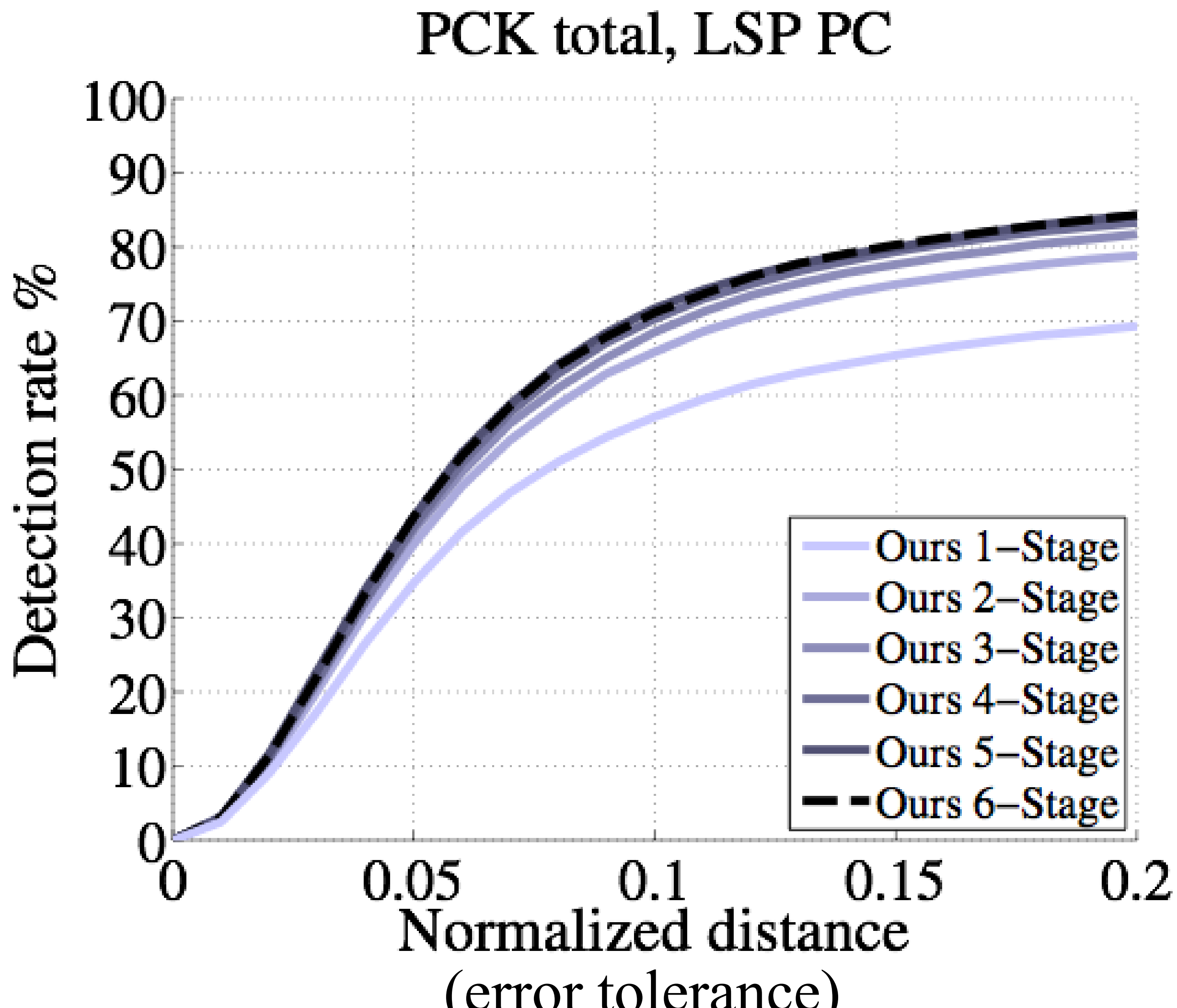
	FLIC	LSP	MPII
size	3987 training 1016 testing	11000 training 1000 testing	29116 training 11823 testing
type	movie scenes	sports	diverse
annotation	upper body	full body	full body w/ truncation





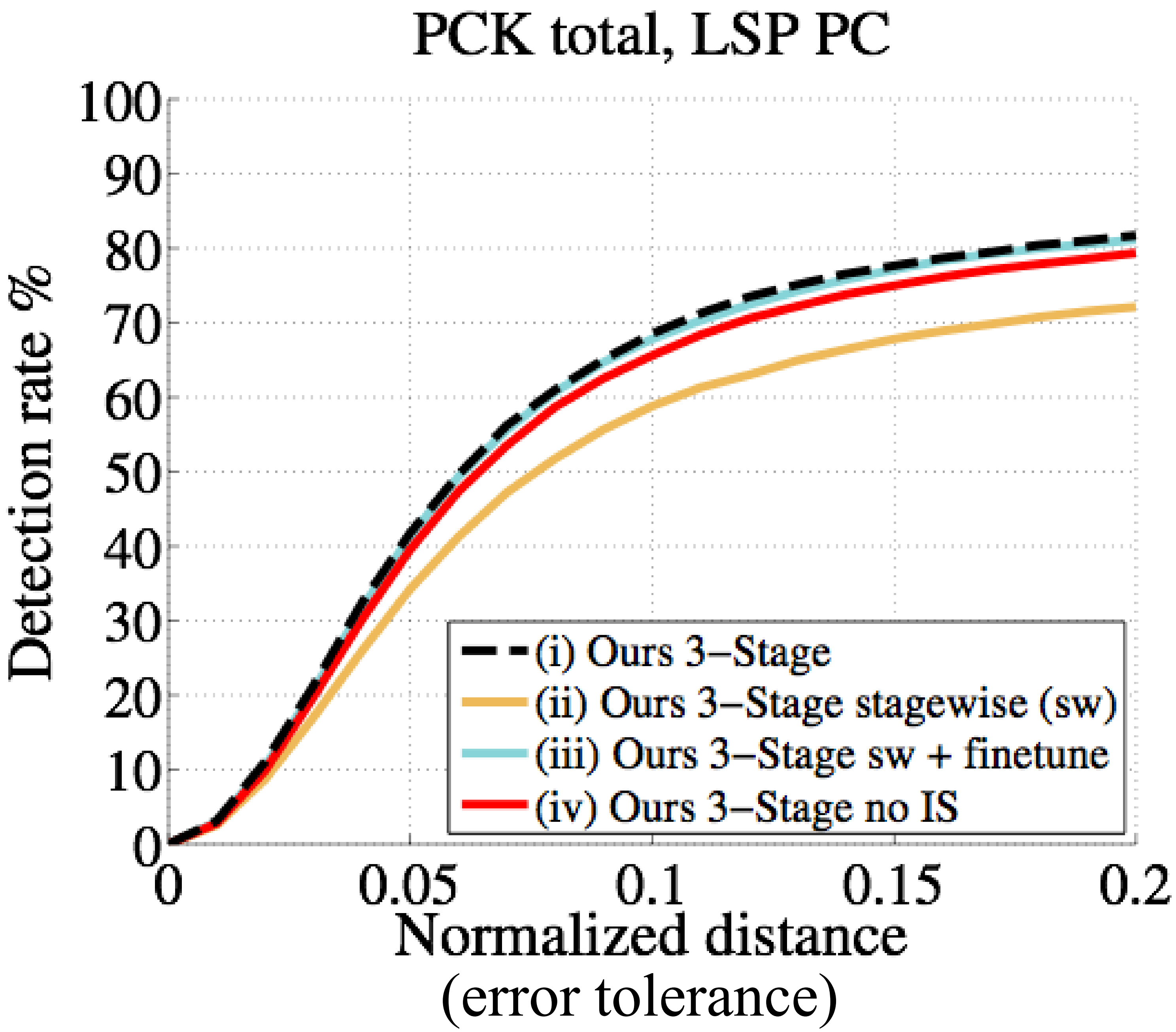

Number of Stages

PCK 0.2



Training Methods

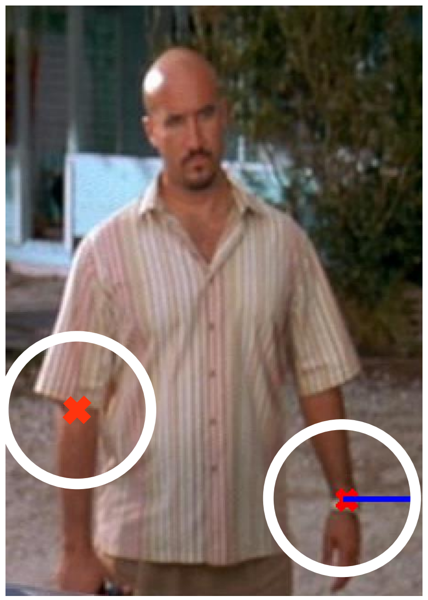
PCK 0.2



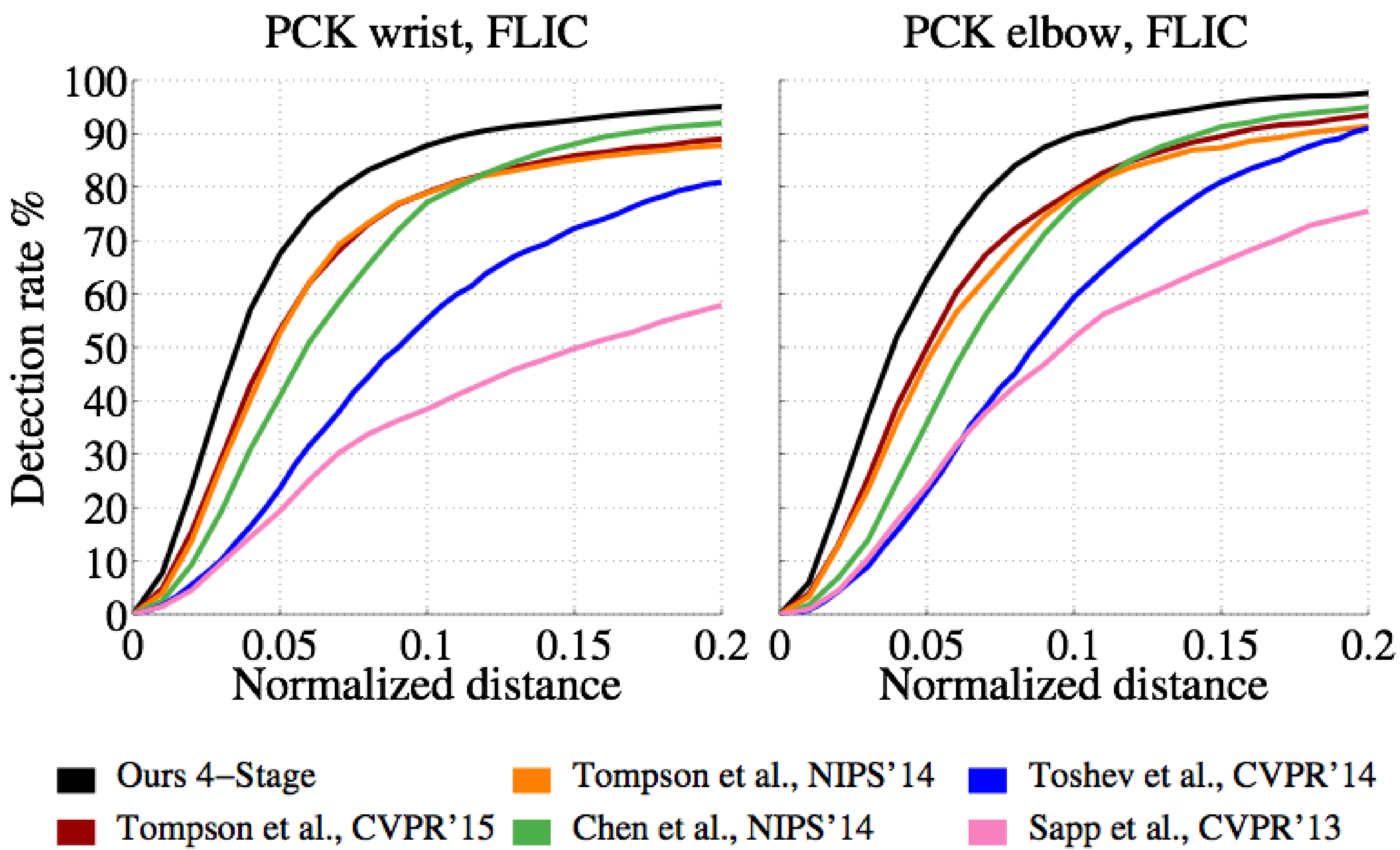
Quantitatively Results

FLIC Upper Body with Observer Centric (OC) Annotations

PCK 0.2



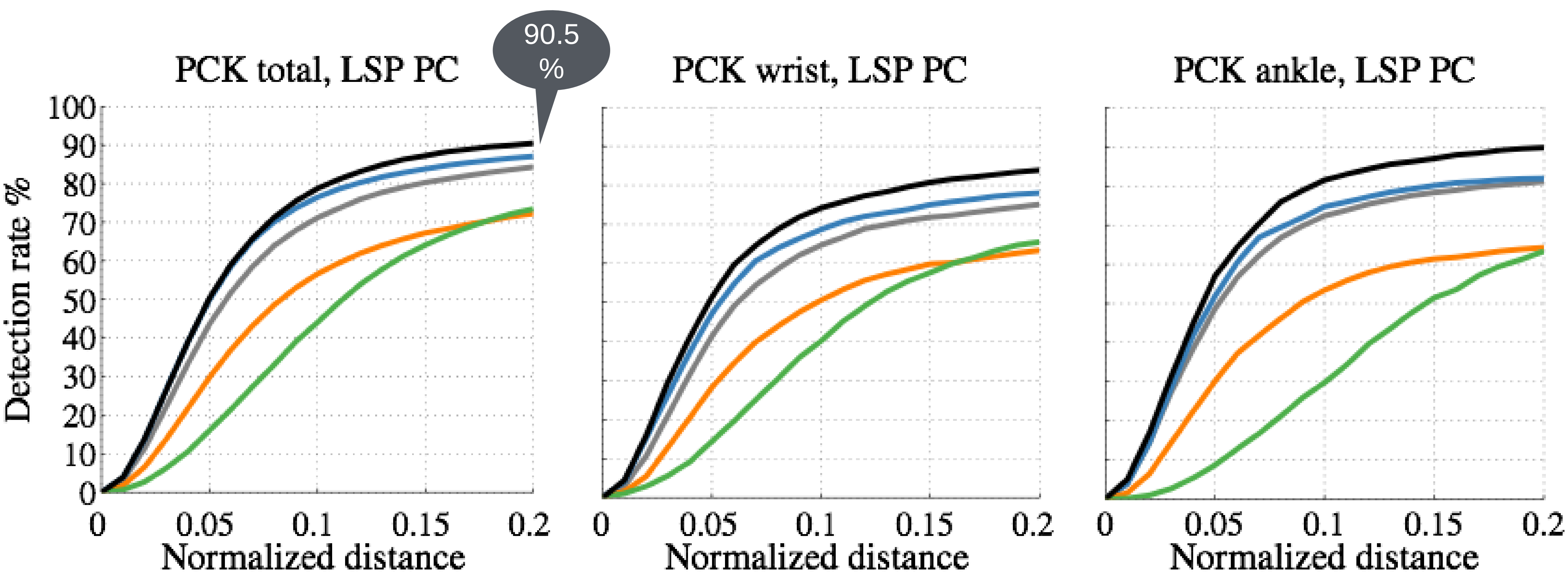
PCK 0.1



Quantitatively Results

LSP Dataset with Person Centric (PC) Annotations

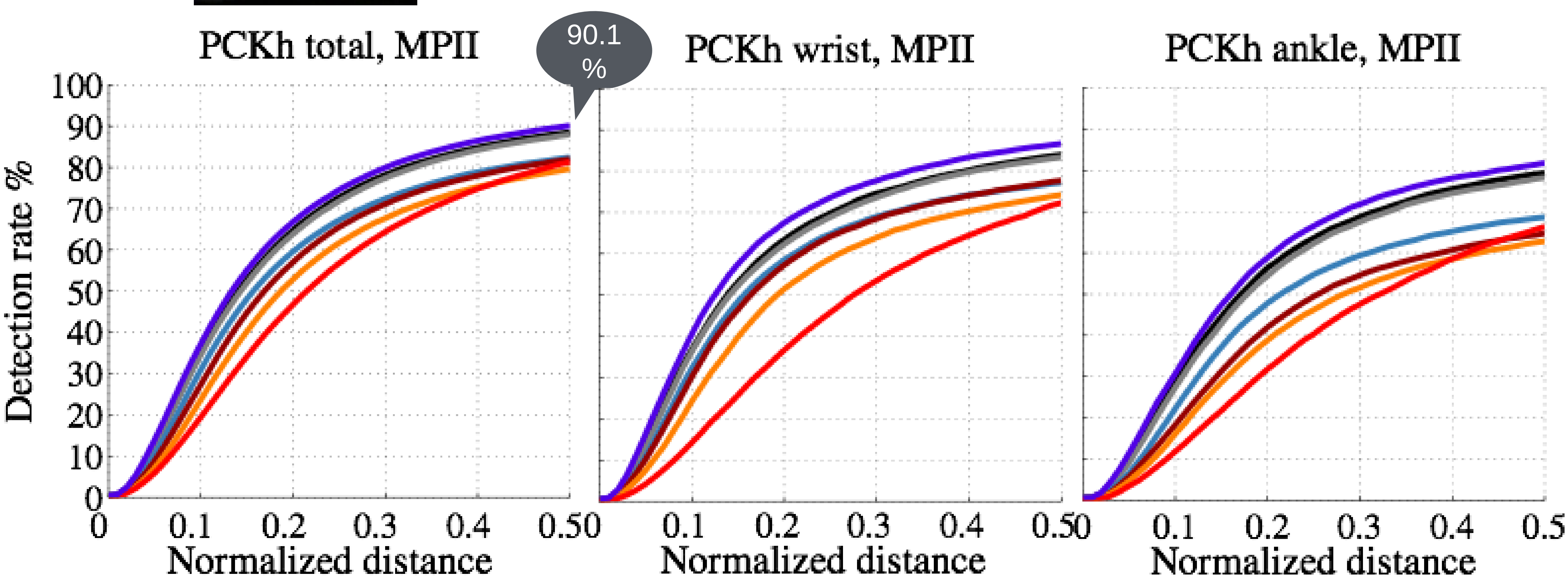
PCK 0.2



Quantitatively Results

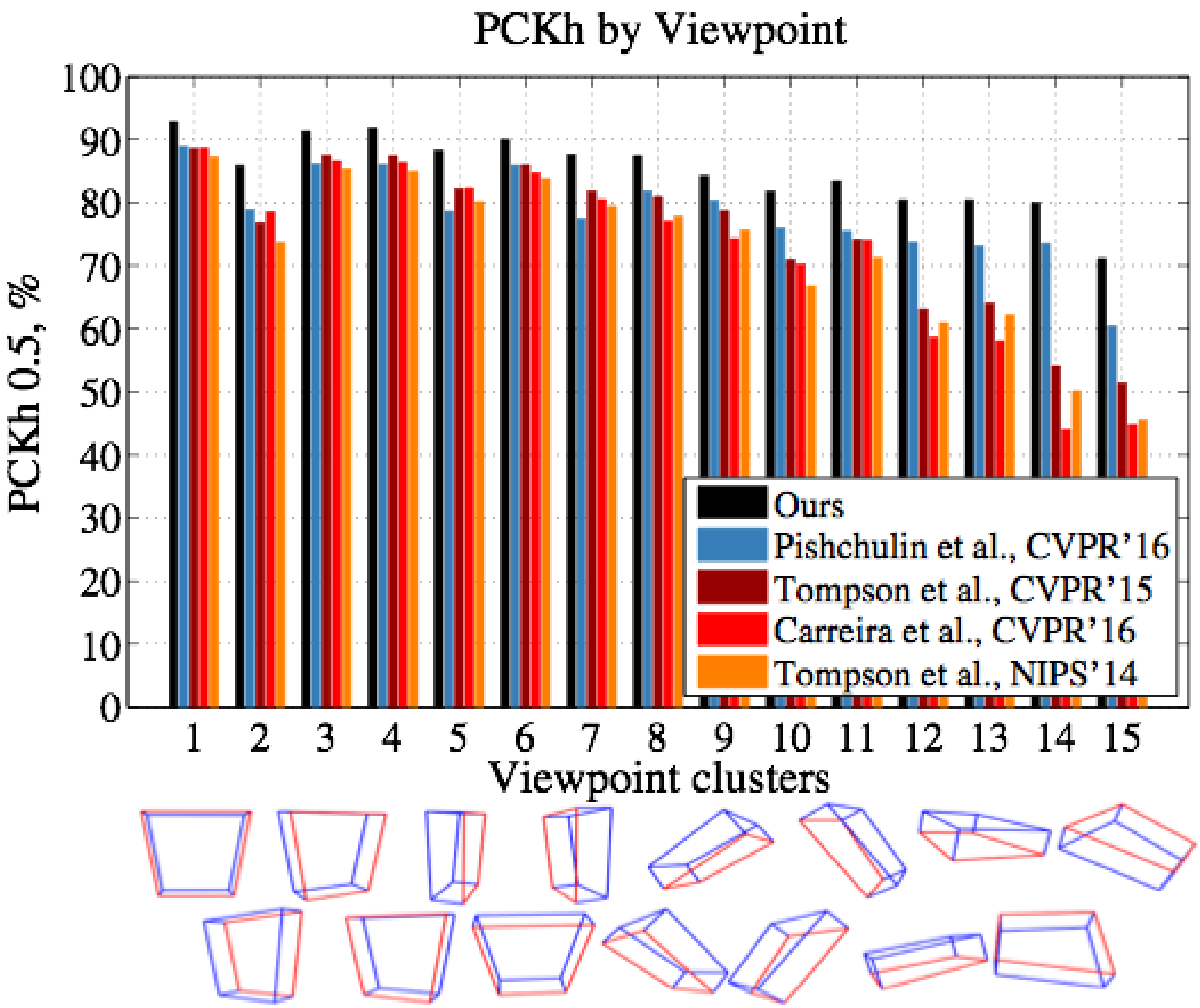
MPII Dataset with PC Annotations

PCKh 0.5



Quantitatively Results

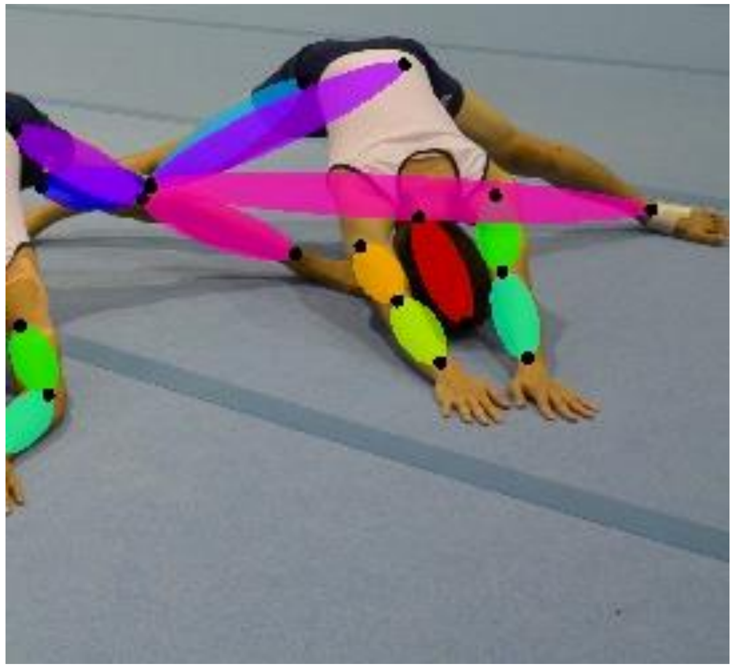
MPII Dataset: Viewpoints



Convolutional Pose Machines: Model trained from MPII Dataset

Failure Cases

L/R confusion



rare viewpoint

rare pose

severe occlusion

right wrist

Summary

- Monocular human pose estimation are becoming reliable.
- CPMs capture complex long-range part dependencies by iteratively refining confidence maps with preserved uncertainty.
- CPMs naturally avoid the problem of vanishing gradient by intermediate supervisions.

What’s Next?

From Single to Multi-Person

Challenge: Identifying number of people and part-person association

Multi-person Human Pose Estimations

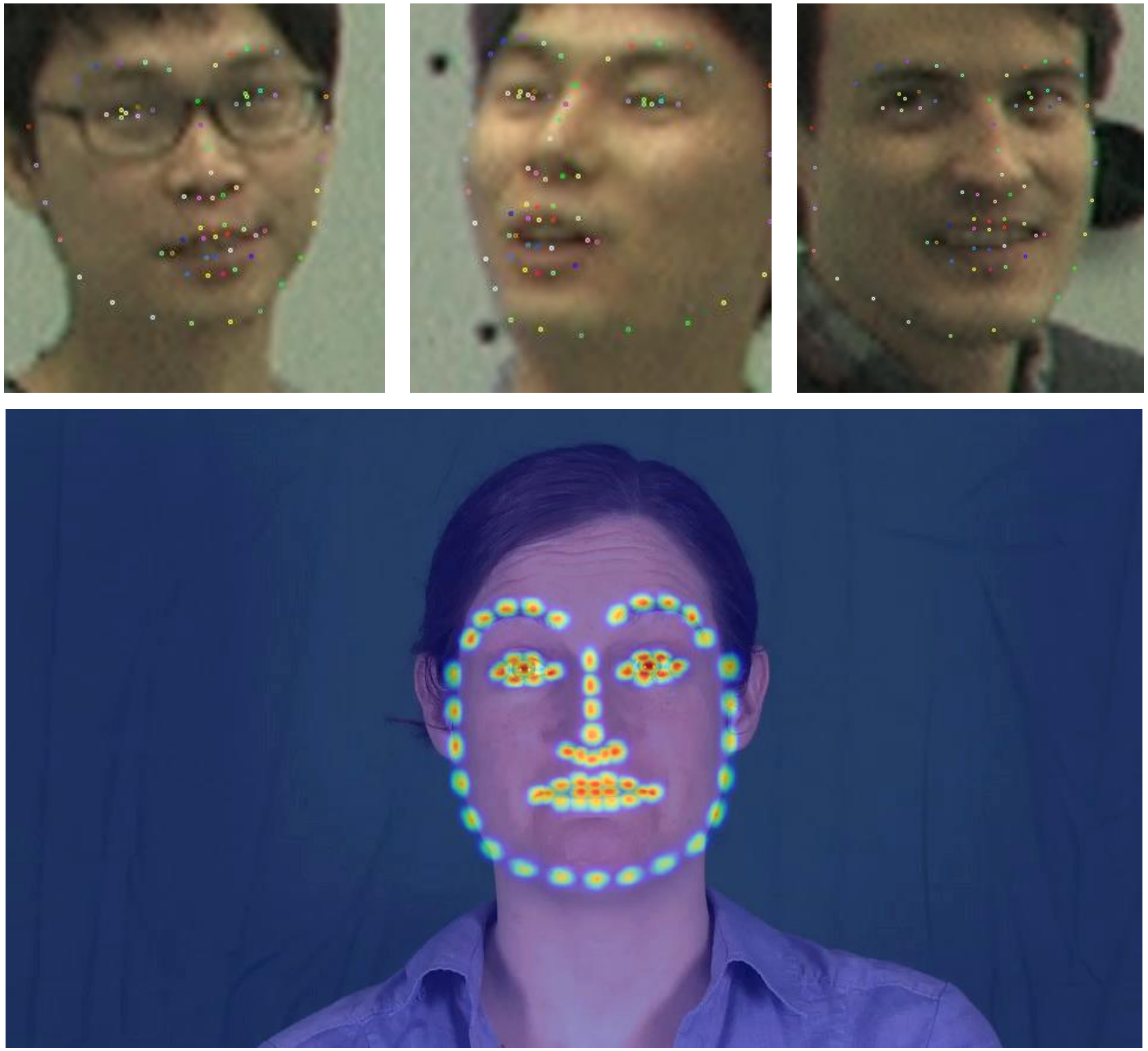
Naive Two-phase Method

CPM with P = 1 (person detector)

Credit: Zhe Cao

Pose Estimations in Finer Scales

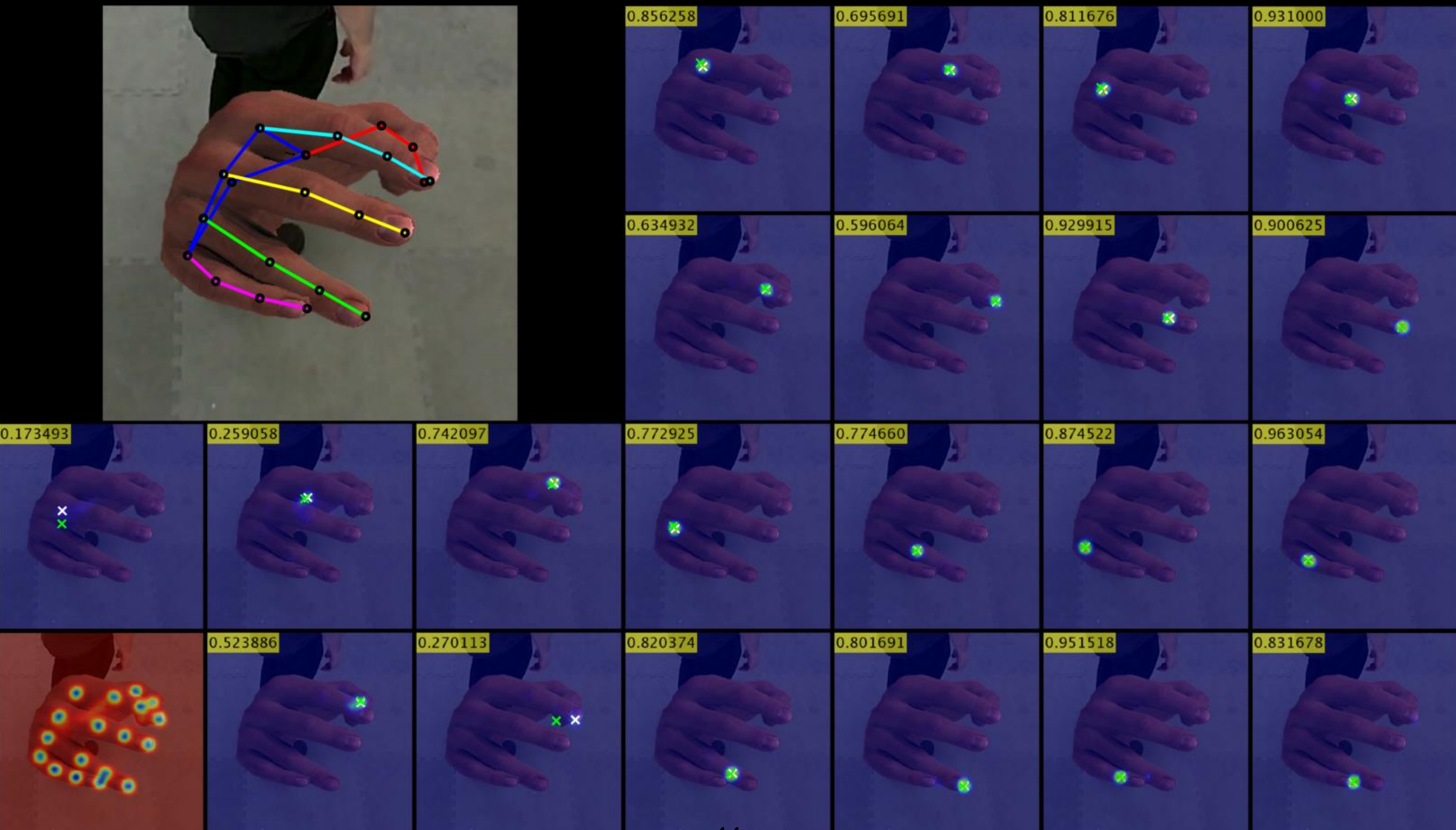
Faces



Source: Tomas Simon

Pose Estimations in Finer Scales

Hands



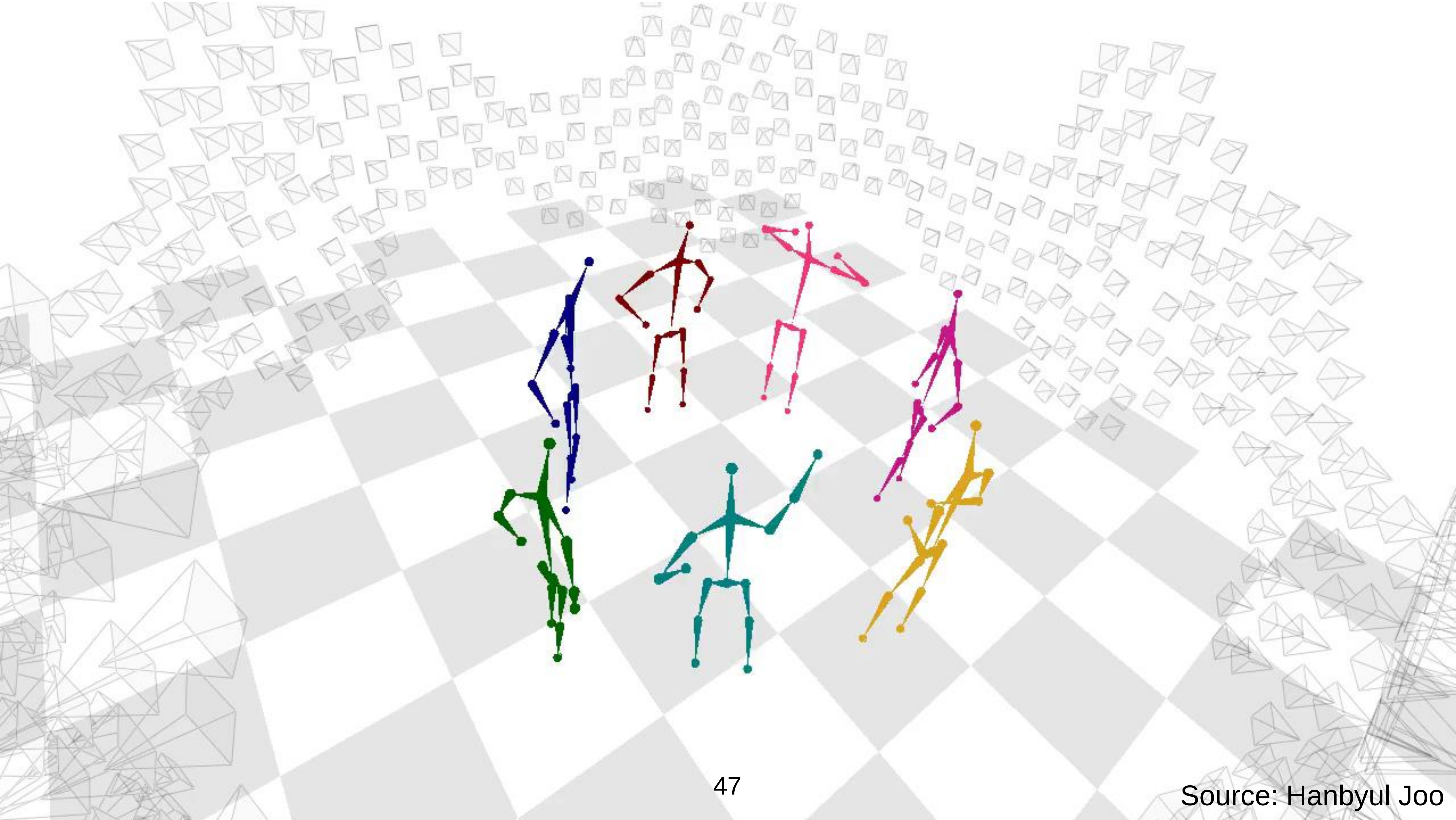
CMU Panoptic Studio

500 Synced Cameras

Multiple Views for 3D Recon

Right Wrist

Multiple Views for 3D Reconstruction



Source: Hanbyul Joo

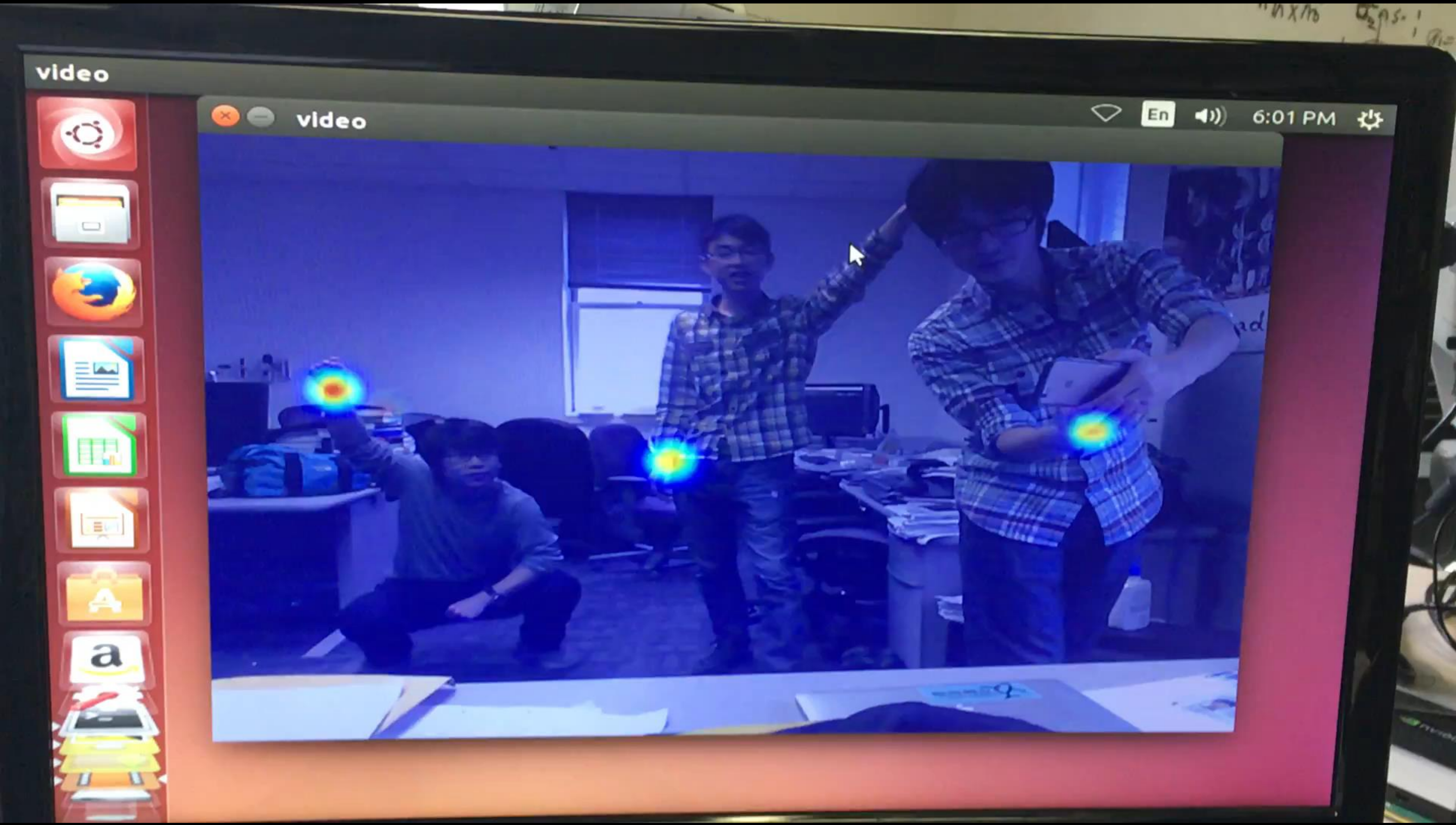
Multiple Views for 3D Recon

Projected Full Poses

Live Demo!

Real-time CPMs

Real-time CPMs: Confidence Map of Right Wrist



Future Directions

- Analysis on failure cases and data distribution
- One-shot multi-person pose estimation
- Direct 3D reasoning
- Temporal CPMs

Thank you
Questions?

Check our Paper, Github, and Youtube Channel!