



腾讯优图



C3P: Cross-domain Pose Prior Propagation for Weakly Supervised 3D Human Pose Estimation

Cunlin Wu¹, Yang Xiao^{1*}, Boshen Zhang², Mingyang Zhang¹, Zhiguo Cao¹, and Joey Tianyi Zhou³

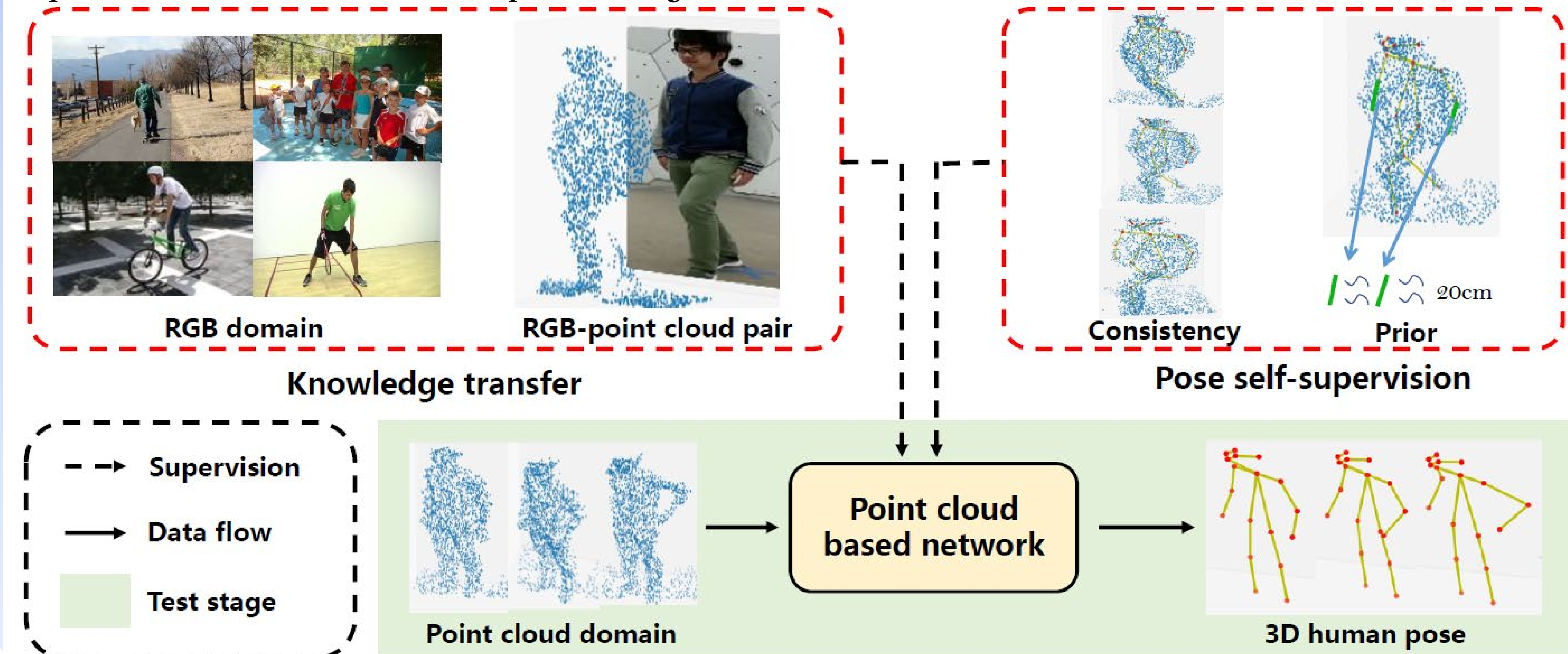
¹ School of AIA, Huazhong University of Science and Technology

² YouTu Lab, Tencent ³ A*STAR Centre for Frontier AI Research (CFAR), Singapore



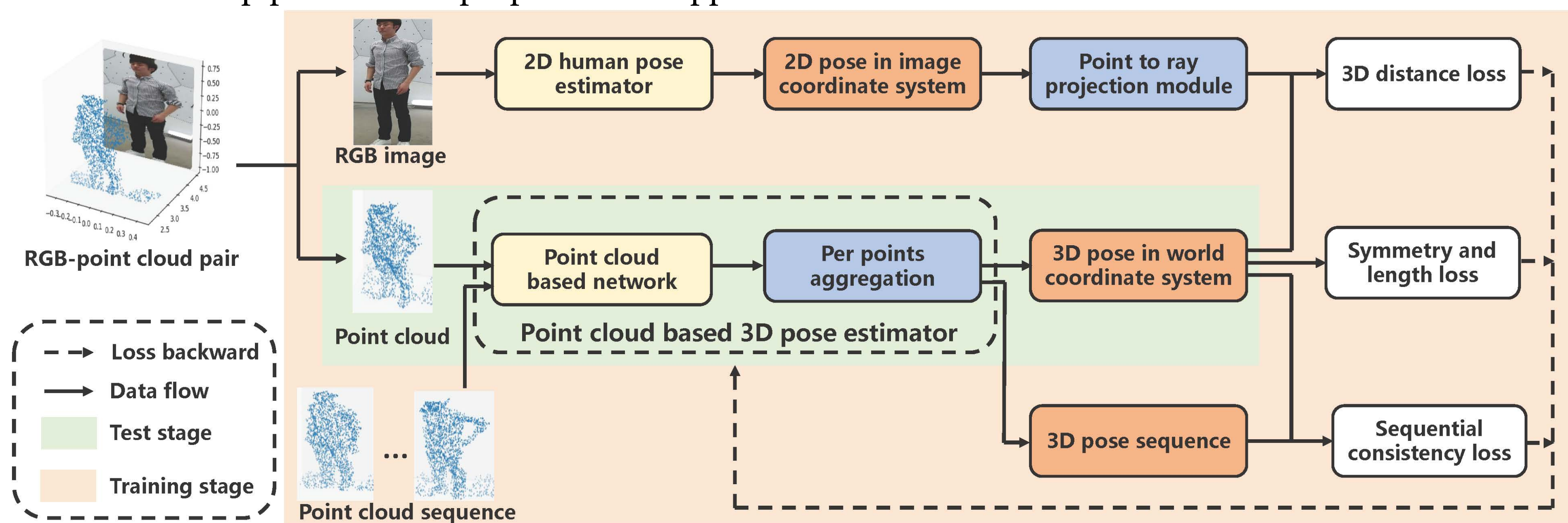
Introduction

We process C3P: a novel weakly supervised 3D human pose estimation approach that relies on unlabeled RGB-point cloud sequence. RGB-point cloud pairs bridge 2D and 3D domain. 3D priors of human pose in point cloud sequences are also utilized as self-supervision signals.



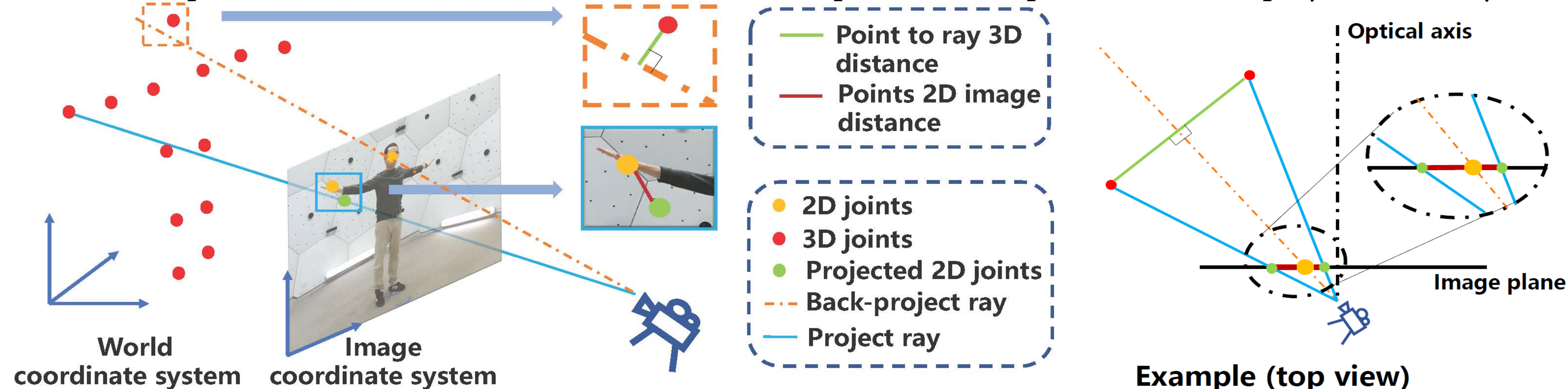
Method

The main technical pipeline of the proposed C3P approach.



Weak supervision signal from RGB image

Different supervision signals. We propose to back-project the pre-computed keypoints on image plane to ray forms in 3D space, and then calculate the distance between predicted 3D points and back-projected 3D rays



Learning procedure

The weakly supervised signal from 2D keypoints

$$\mathcal{D}_{p-ray,k} = \frac{\|K_{rgb}^{-1} \hat{z}_k [u_k \ v_k \ 1]^T - \hat{j}_k\|_2}{\|K_{rgb}^{-1} [u_k \ v_k \ 1]^T\|_2}, \quad \mathcal{L}_{2d} = \sum_{k=1}^K \frac{\mathcal{D}_{p-ray,k}}{\mu}$$

The self supervision signal from bone length and human body symmetry

$$\mathcal{L}_{len} = \sum_{n=1}^N \|B_n - \bar{B}_n\|_2^2, \quad \mathcal{L}_{sym} = \sum_{n=1}^N \left\| \frac{B_n}{B_n^{cor}} - 1 \right\|$$

The self supervision signal from motion consistency

$$\mathcal{L}_{con1} = \sum_{k=1}^K \sum_{i=2}^{I-1} \left\| j_k^i - \frac{j_k^{i-1} + j_k^{i+1}}{2} \right\|, \quad \mathcal{L}_{con2} = \sum_{n=1}^N \left\| \frac{B_n}{\bar{B}_n} - 1 \right\|$$

Experiments

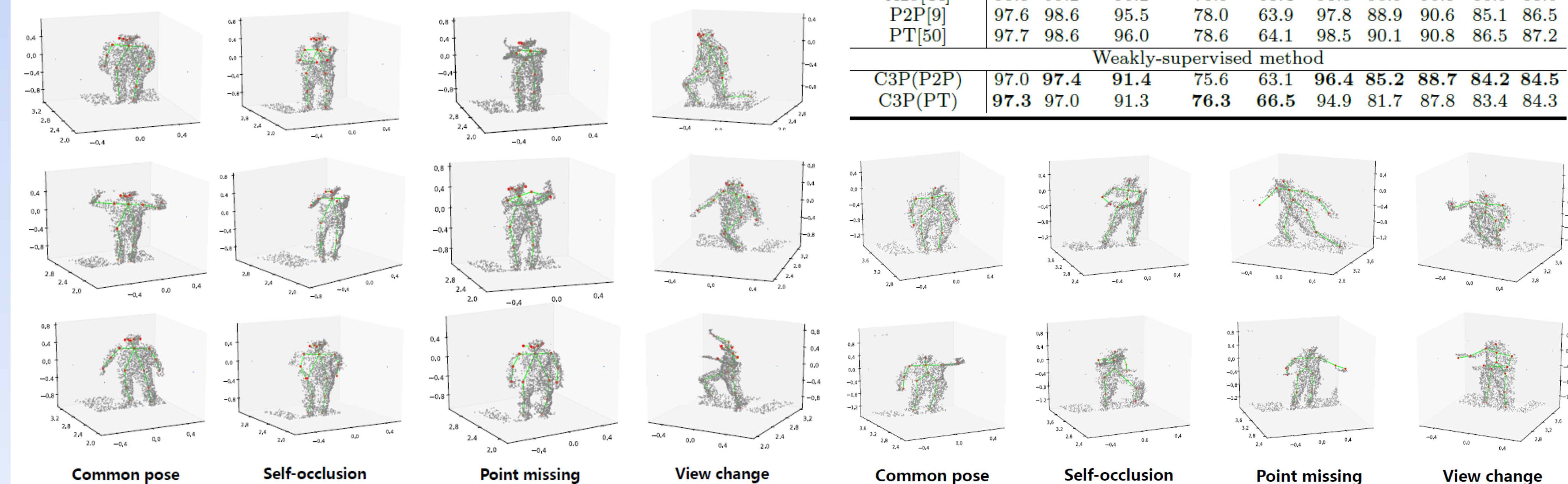
Comparison with fully-supervised method

Performance comparison on CMU Panoptic Dataset

	Nose	Eyes	Ears	Shoulders	Elbows	Wrists	Hips	Knees	Ankles	mAP	Error
Fully-supervised methods											
HandPoint[8]	79.8	78.9	79.6	80.6	5.3	0.2	89.6	84.5	75.1	62.8	12.0
P2P[9]	98.1	98.1	98.2	96.9	95.0	89.8	94.5	94.1	93.7	95.2	4.1
PT[50]	99.6	99.3	99.2	98.8	97.0	92.0	96.8	95.0	95.7	96.9	3.3
Weakly-supervised methods											
C3P(P2P)	96.3	95.8	95.0	93.9	91.4	81.5	90.9	78.4	85.2	89.4	6.1
C3P(PT)	99.1	98.6	95.9	95.4	94.4	85.4	93.1	91.1	94.0	93.8	5.3

Performance comparison on ITOP Dataset

	Head	Neck	Shoulders	Elbows	Hands	Torso	Hips	Knees	Feet	mean
Fully-supervised methods										
RF[34]	63.8	86.4	83.3	73.2	51.3	65.0	50.8	65.7	61.3	65.8
RTW[47]	97.8	95.8	94.1	77.9	70.5	93.8	90.3	68.8	68.4	80.5
IEF[6]	96.2	85.2	77.2	45.4	30.9	84.7	83.5	81.8	80.9	71.0
VI[11]	98.1	97.5	96.5	73.3	68.7	85.6	72.0	69.0	60.8	77.4
CMB[41]	97.7	98.5	75.9	62.7	84.4	96.0	87.9	84.4	83.8	83.3
REN-9*6*6[10]	98.7	99.4	96.1	74.7	55.2	98.7	91.8	89.0	81.1	84.9
V2V*[27]	98.3	99.1	99.2	80.4	67.3	98.7	93.2	91.8	87.6	88.7
A2J[44]	98.5	99.2	96.2	78.9	68.4	98.5	90.9	90.8	86.9	88.0
P2P[9]	97.6	98.6	95.5	78.0	63.9	97.8	88.9	90.6	85.1	86.5
PT[50]	97.7	98.6	96.0	78.6	64.1	98.5	90.1	90.8	86.5	87.2
Weakly-supervised method										
C3P(P2P)	97.0	97.4	91.4	75.6	63.1	96.4	85.2	88.7	84.2	84.5
C3P(PT)	97.3	97.0	91.3	76.3	66.5	94.9	81.7	87.8	83.4	84.3



Contact information

Projection: <https://github.com/wucunlin/C3P>

E-mail: {cunlin_wu, Yang_Xiao, izmy, zgcao}@hust.edu.cn, boshenzhang@tencent.com, zhouty@ihpc.a-star.edu.sg