PBADet: A One-Stage Anchor-Free Approach for Part-Body Association

Zhongpai Gao[1], Huayi Zhou[2], Abhishek Sharma[1], Meng Zheng[1],
Benjamin Planche[1], Terrence Chen[1], Ziyan Wu[1]

[1]United Imaging Intelligence, Burlington, MA 01803, USA
[2]Shanghai Jiao Tong University, Shanghai 200240, China

2024-04-17

Motivation

- **Problem:** detecting human parts and associating them with the correct individual.
- **Application:** complex scenarios involving hand gestures from multiple people (*e.g.,* gesture from whom?).

Existing methods often involve two-stage processes

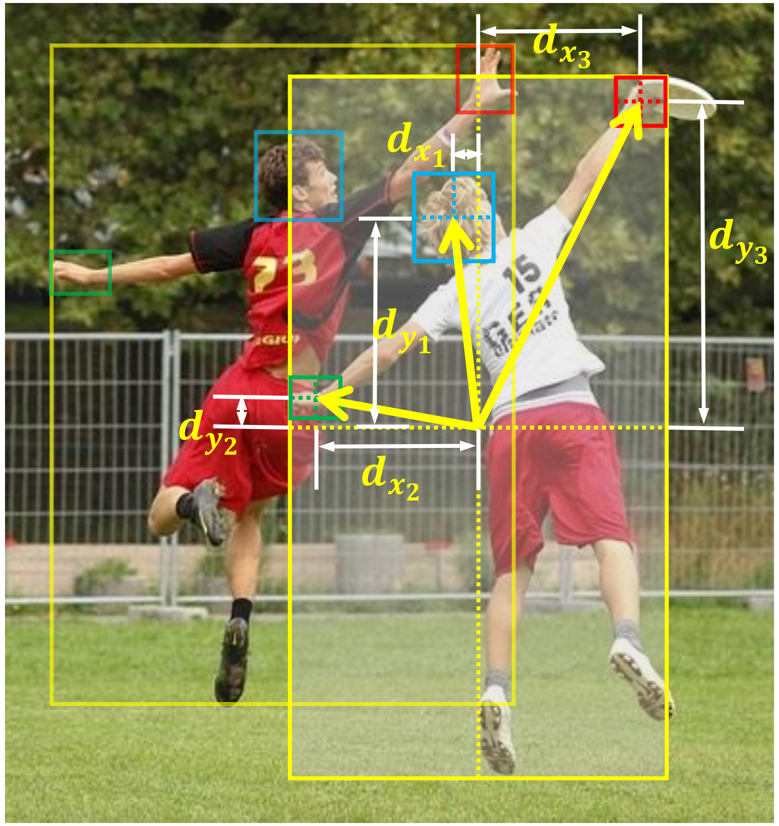
- Detect hands and human body pose, then using heuristic strategies to match [1]
- Detect hands and bodies, then utilizing association network to predict [2]

[1] Huayi Zhou, Fei Jiang, and Ruimin Shen. *Who are raising their hands? hand-raiser seeking based on object detection and pose estimation*. ACML 2018

[2] Supreeth Narasimhaswamy, Thanh Nguyen, Mingzhen Huang, and Minh Hoai. *Whose hands are these? hand detection and hand-body association in the wild*. CVPR 2022

Motivation – Comparison with BPJDet

- BPJDet: **anchor-based** and **body-to-part association** on body objects [3]
- PBADet (ours): **anchor-free** and **part-to-body association** on part objects



Human Body Object

$\mathcal{O}^b = (1, b^b_x, b^b_y, b^b_w, b^b_h, 1,0,0,0, d^b_{x_1}, d^b_{y_1}, d^b_{x_2}, d^b_{y_2}, d^b_{x_3}, d^b_{y_3})$

Body Part Object

$\mathcal{O}^p = (1, b^p_x, b^p_y, b^p_w, b^p_h, 0,1,0,0, -,-,-,-,-,-)$ → *head*

$\mathcal{O}^p = (1, b^p_x, b^p_y, b^p_w, b^p_h, 0,0,1,0, -,-,-,-,-,-)$ → *left hand*

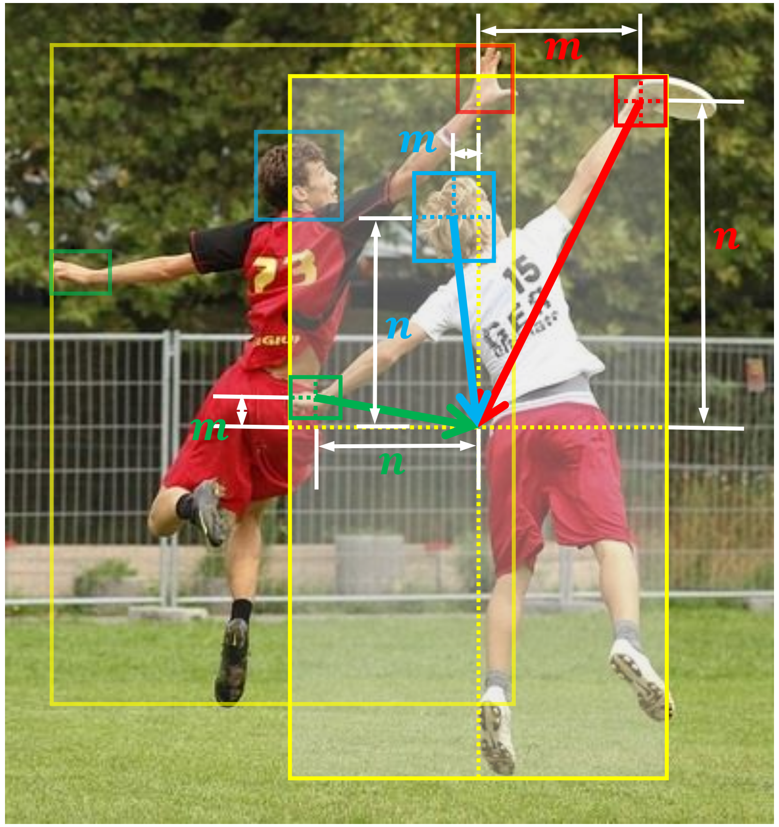
$\mathcal{O}^p = (1, b^p_x, b^p_y, b^p_w, b^p_h, 0,0,0,1, -,-,-,-,-,-)$ → *right hand*

No Object

$\mathcal{O}^0 = (0,-,-,-,-, 0,0,0,0, -,-,-,-,-,-)$

Representation Length:
6+K+2K (K= # of Parts)

Body-to-Part Association
（BPJDet）



Human Body Object

$\mathcal{O}^b = (b^b_l, b^b_t, b^b_r, b^b_b, 1,0,0,0, -,-)$

Body Part Object

$\mathcal{O}^p = (b^p_l, b^p_t, b^p_r, b^p_b, 0,1,0,0, m, n)$ → *head*

$\mathcal{O}^p = (b^p_l, b^p_t, b^p_r, b^p_b, 0,0,1,0, m, n)$ → *left hand*

$\mathcal{O}^p = (b^p_l, b^p_t, b^p_r, b^p_b, 0,0,0,1, m, n)$ → *right hand*

No Object

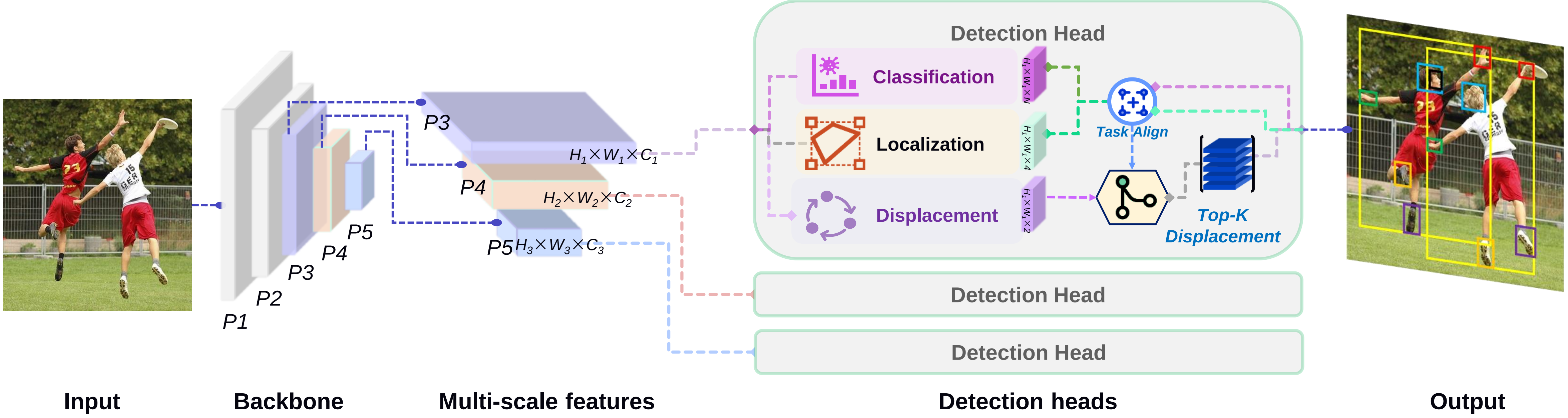
$\mathcal{O}^0 = (-,-,-,-, 0,0,0,0, -,-)$

Representation Length:
5+K+2 (k= # of Parts)

Part-to-Body Association
（Our proposed **PBADet**）

[3] Huayi Zhou, Fei Jiang, and Hongtao Lu. *Body-part joint detection and association via extended object representation*. ICME 2023

Method – Pipeline



Method – Definition

- **Part-to-body Association Definition**

We predict the part-to-body center offset (m_i, n_i) in multi-scale satisfying:

$$\left\lfloor \frac{c_x^b}{s_i} \right\rfloor = x_i + \lambda m_i, \qquad \left\lfloor \frac{c_y^b}{s_i} \right\rfloor = y_i + \lambda n_i,$$

with s_i stride at i^{th} feature layer, (x_i, y_i) feature point coordinates, c^b is the center of the body, and λ scaling factor of the center offset to control the NN output range.

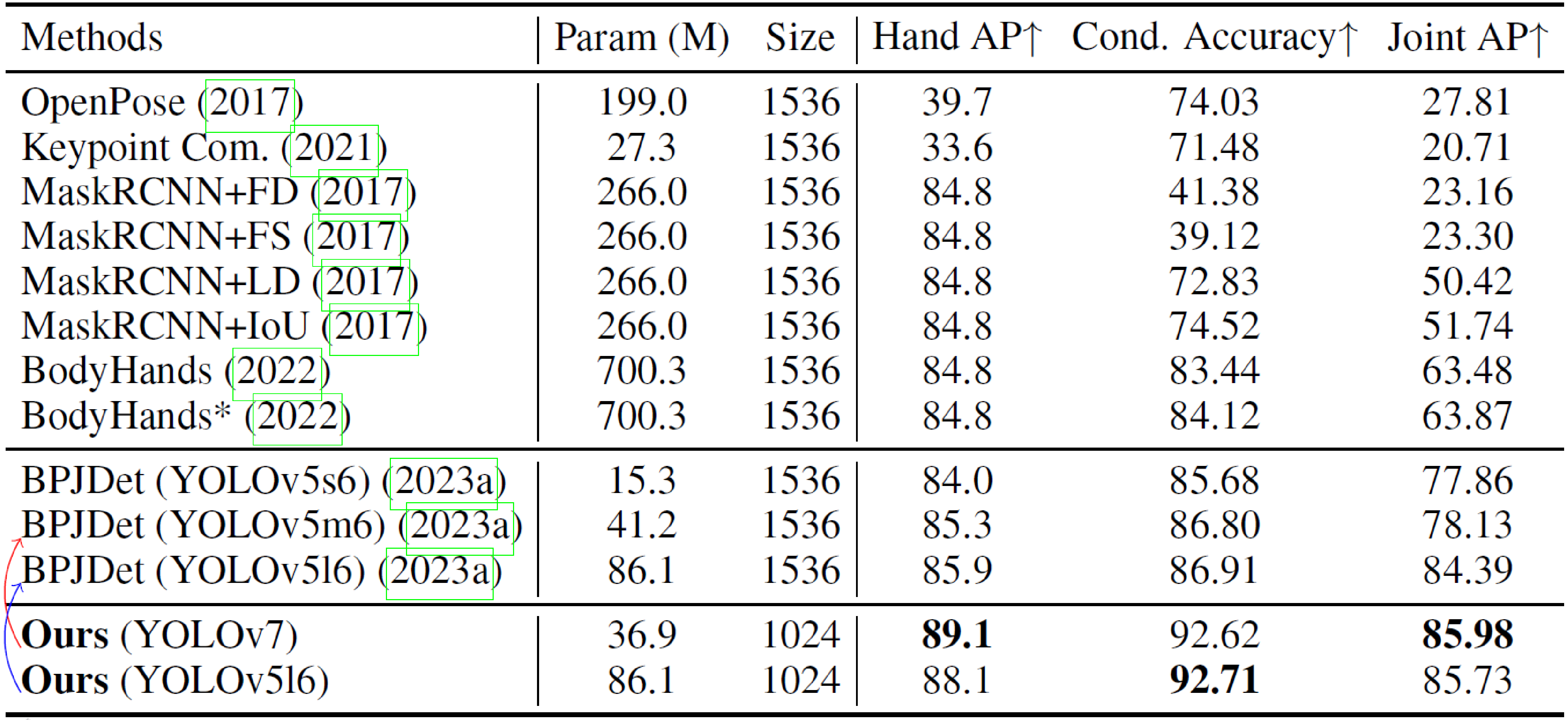
- **Decoding Part-to-body Associations**

The anticipated body center is computed as:

$$\hat{c}_x^b = s_i(x_i + \lambda m_i), \qquad \hat{c}_y^b = s_i(y_i + \lambda n_i).$$

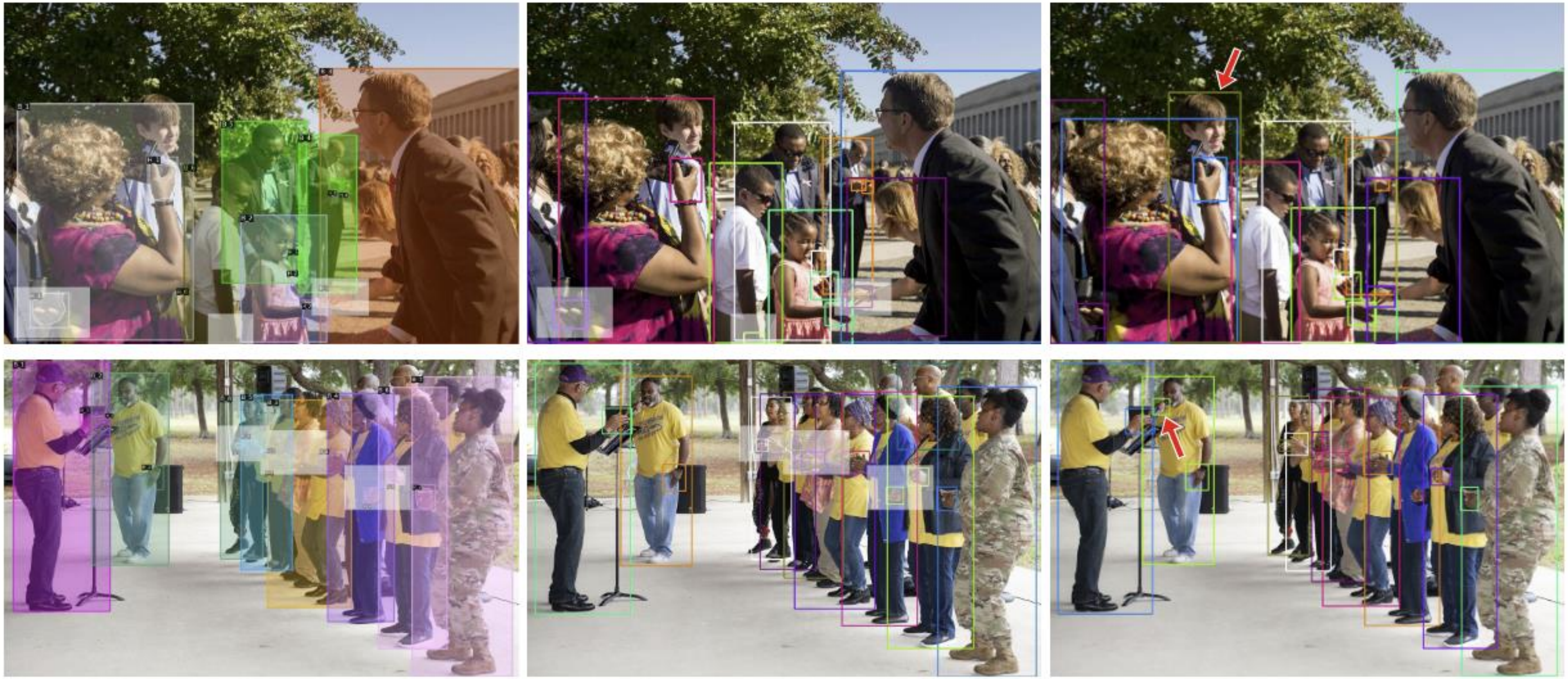
The Euclidean (ℓ_2) distance is used to determine the association.

Results – BodyHands



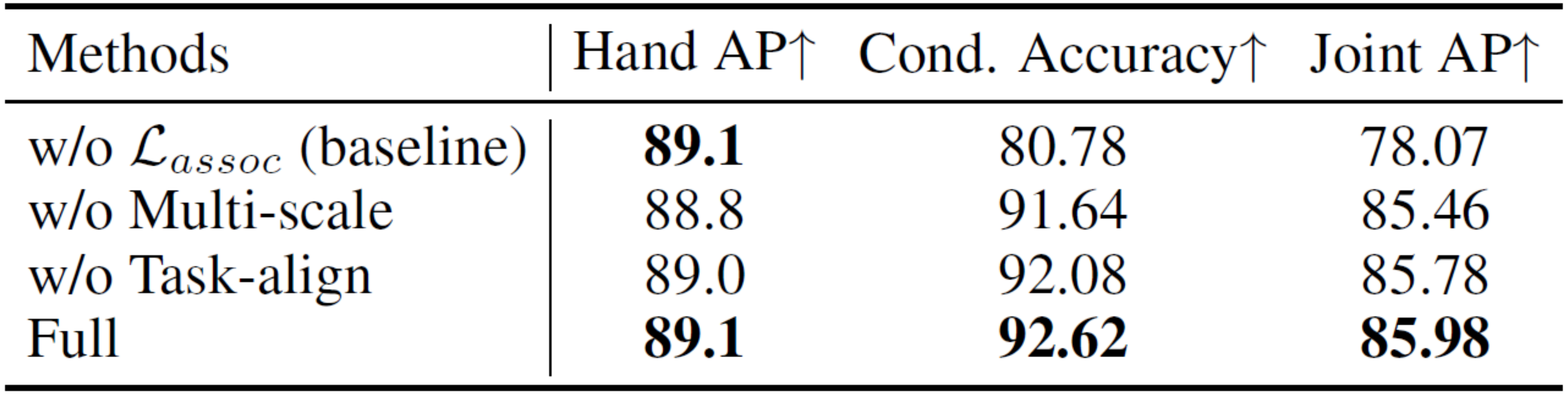
Methods	Param (M)	Size	Hand AP↑	Cond. Accuracy↑	Joint AP↑
OpenPose (2017)	199.0	1536	39.7	74.03	27.81
Keypoint Com. (2021)	27.3	1536	33.6	71.48	20.71
MaskRCNN+FD (2017)	266.0	1536	84.8	41.38	23.16
MaskRCNN+FS (2017)	266.0	1536	84.8	39.12	23.30
MaskRCNN+LD (2017)	266.0	1536	84.8	72.83	50.42
MaskRCNN+IoU (2017)	266.0	1536	84.8	74.52	51.74
BodyHands (2022)	700.3	1536	84.8	83.44	63.48
BodyHands* (2022)	700.3	1536	84.8	84.12	63.87
BPJDet (YOLOv5s6) (2023a)	15.3	1536	84.0	85.68	77.86
BPJDet (YOLOv5m6) (2023a)	41.2	1536	85.3	86.80	78.13
BPJDet (YOLOv5l6) (2023a)	86.1	1536	85.9	86.91	84.39
Ours (YOLOv7)	36.9	1024	**89.1**	92.62	**85.98**
Ours (YOLOv5l6)	86.1	1024	88.1	**92.71**	85.73

Results – BodyHands

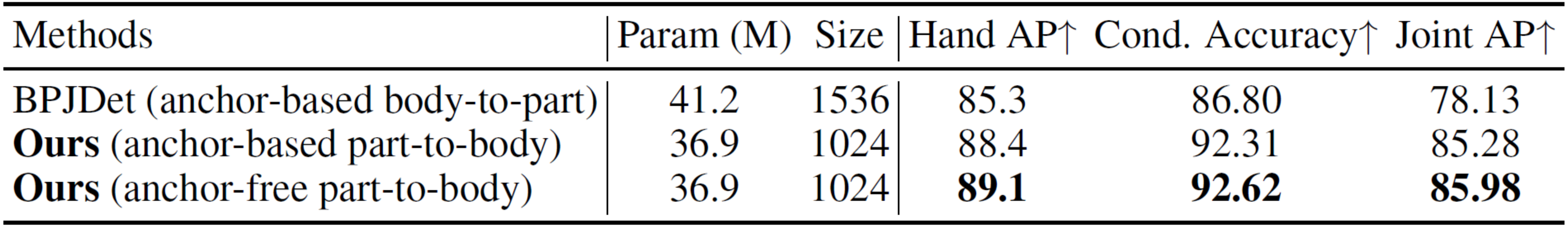


BodyHands BPJDet (YOLOv5l6) **Ours** (YOLOv5l6)

Results – Ablation Study

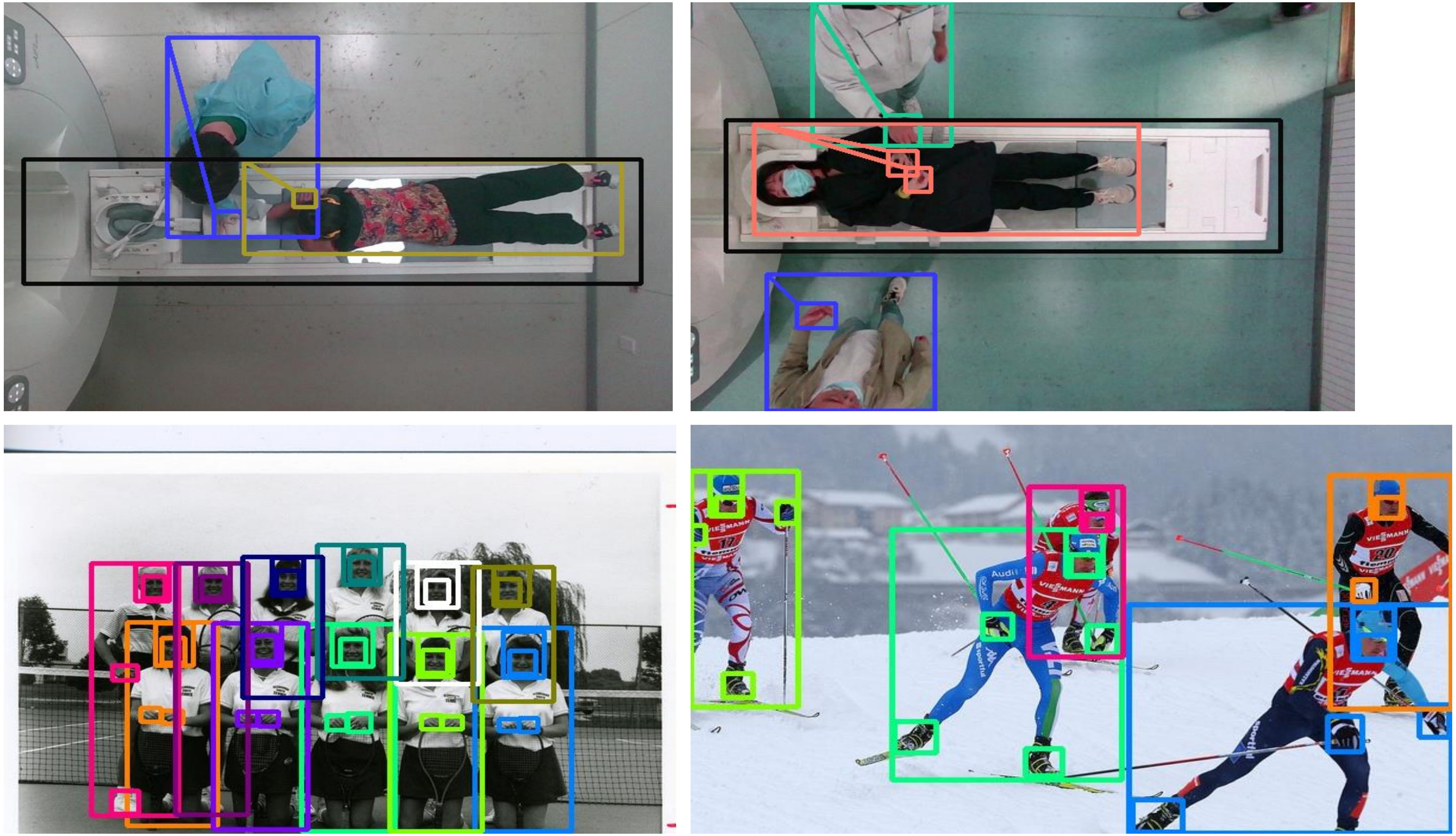


Methods	Hand AP↑	Cond. Accuracy↑	Joint AP↑
w/o $\mathcal{L}_{assoc}$ (baseline)	**89.1**	80.78	78.07
w/o Multi-scale	88.8	91.64	85.46
w/o Task-align	89.0	92.08	85.78
Full	**89.1**	**92.62**	**85.98**



Methods	Param (M)	Size	Hand AP↑	Cond. Accuracy↑	Joint AP↑
BPJDet (anchor-based body-to-part)	41.2	1536	85.3	86.80	78.13
Ours (anchor-based part-to-body)	36.9	1024	88.4	92.31	85.28
Ours (anchor-free part-to-body)	36.9	1024	**89.1**	**92.62**	**85.98**

Results – More Qualitative Examples



Leading CHANGE

引领改变