CVPR 2019, Long Beach, California

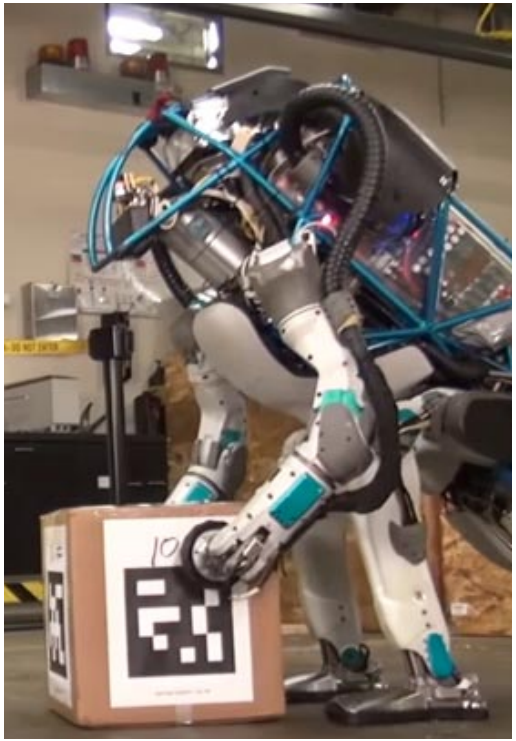
Estimating 3D Motion and Forces of Person-Object Interactions from Monocular Video

Zongmian Li [1,2], Jiri Sedlar [3], Justin Carpentier [1,2]

Ivan Laptev [1,2], Nicolas Mansard [4], Josef Sivic [1,2,3]

[1] Inria [2] DI ENS, PSL [3] CIIRC, CTU in Prague [4] LAAS-CNRS

Motivation: learn how to interact with objects



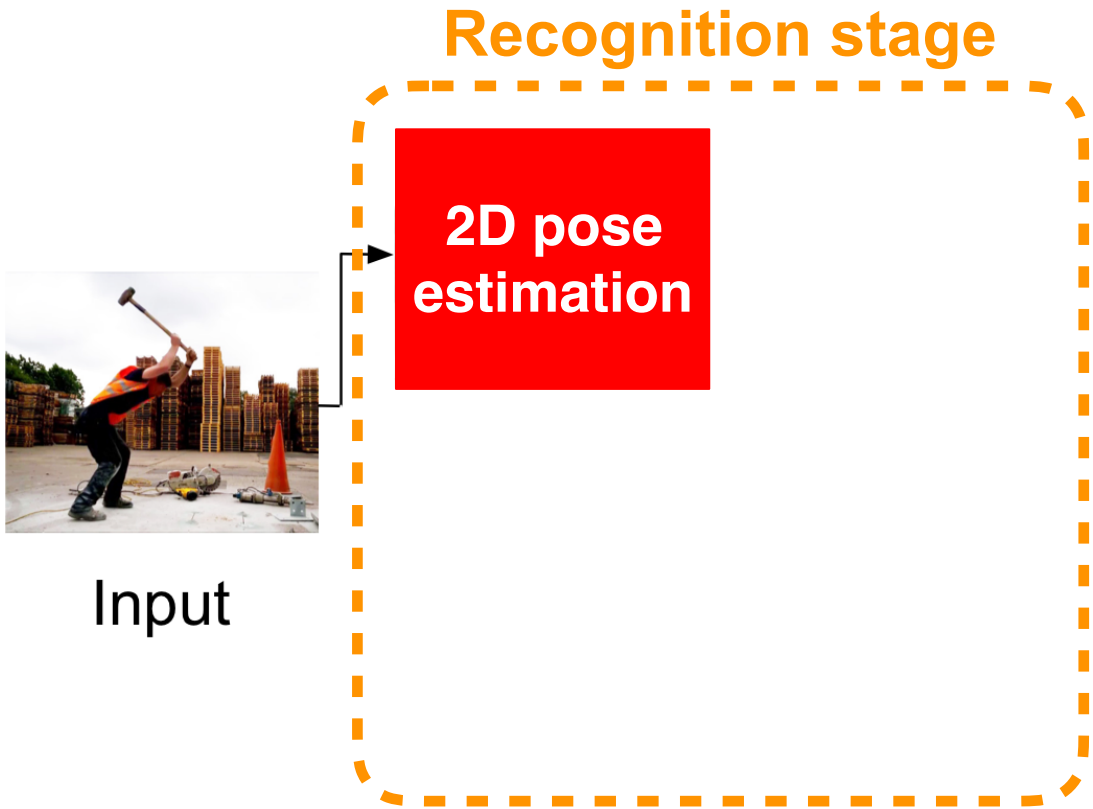
“Atlas” by Boston Dynamics

Goals

Input:

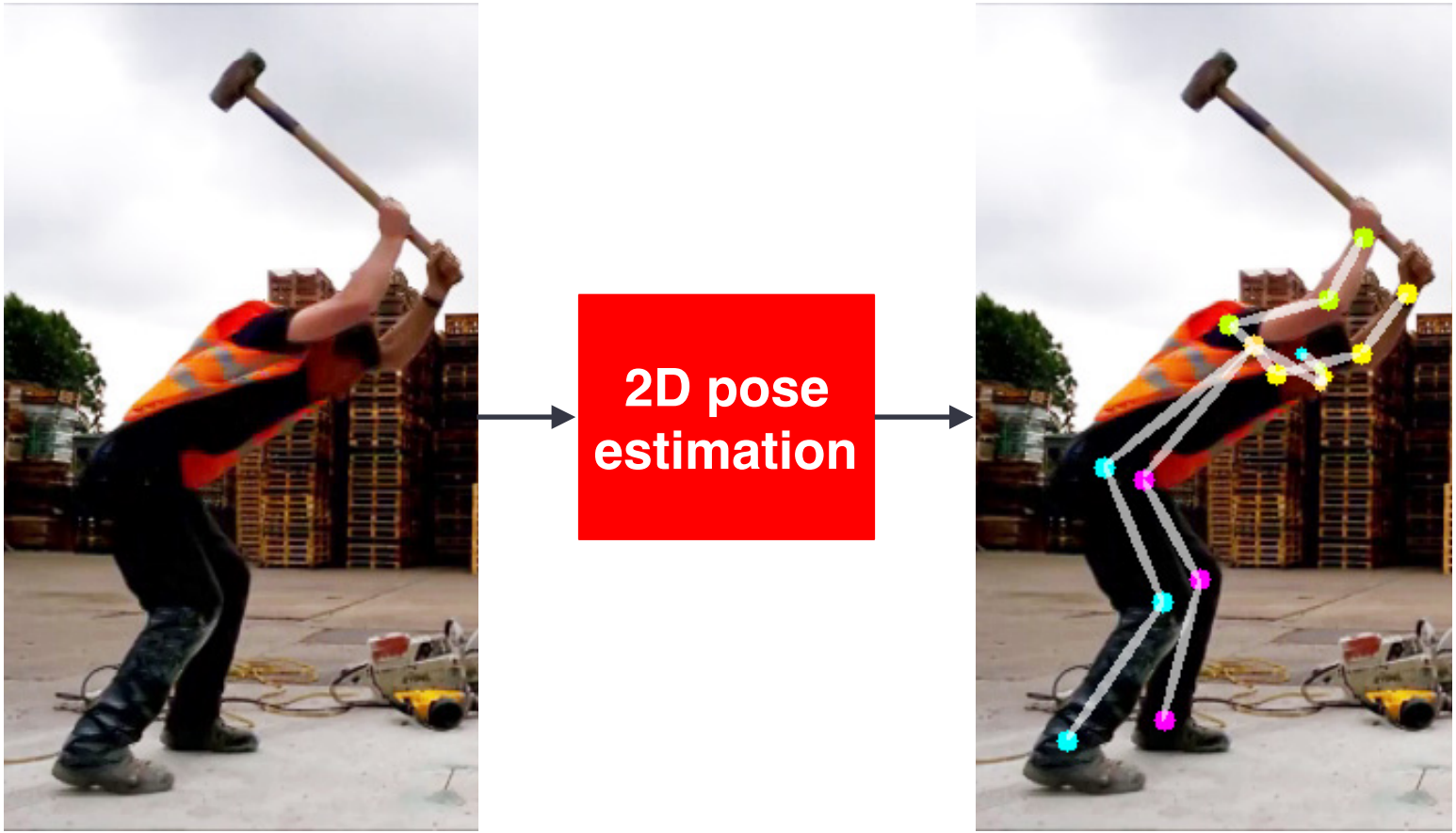
- A monocular RGB video

Method: a two-stage approach

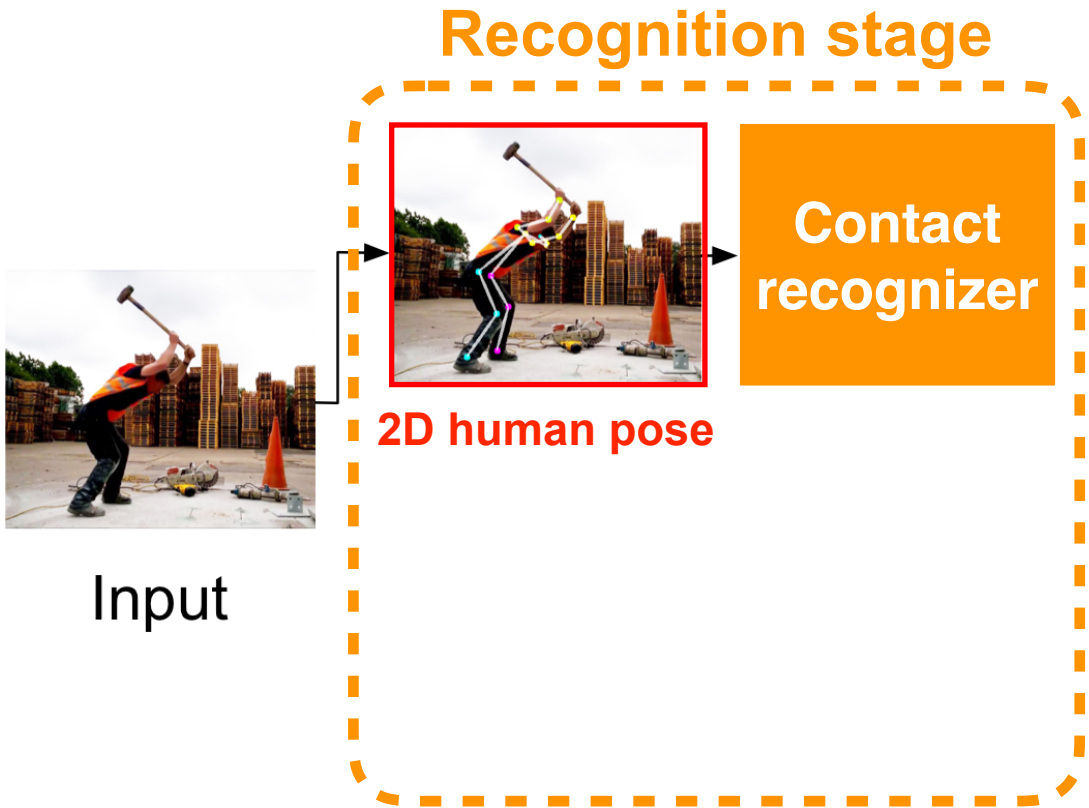


[Cao et al., CVPR 2017]

Method: a two-stage approach

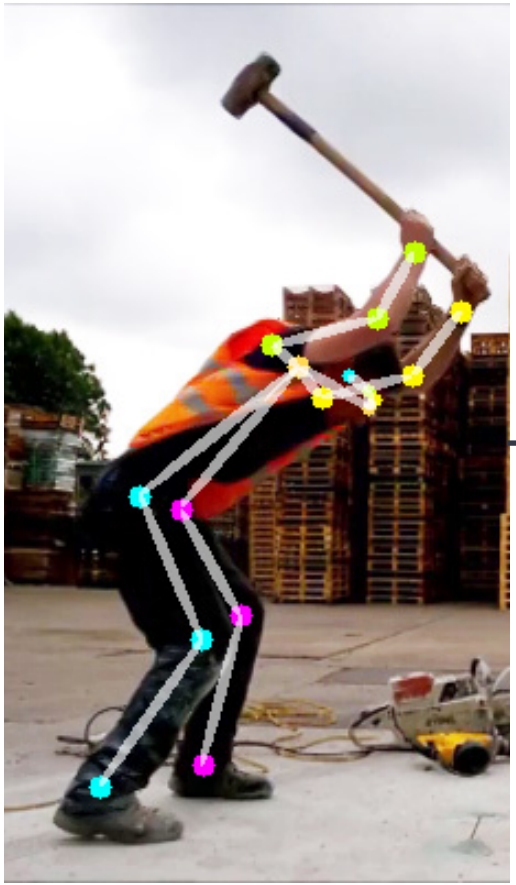


Method: a two-stage approach



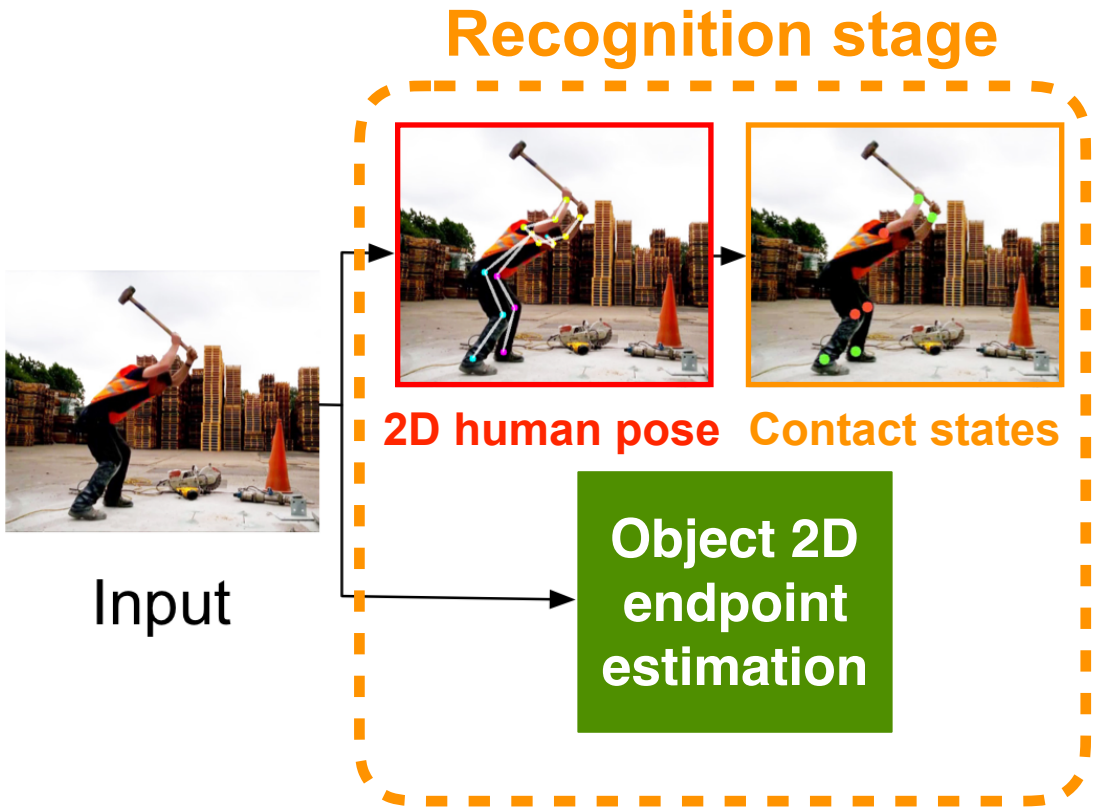
[He et al., CVPR 2016]

Method: a two-stage approach



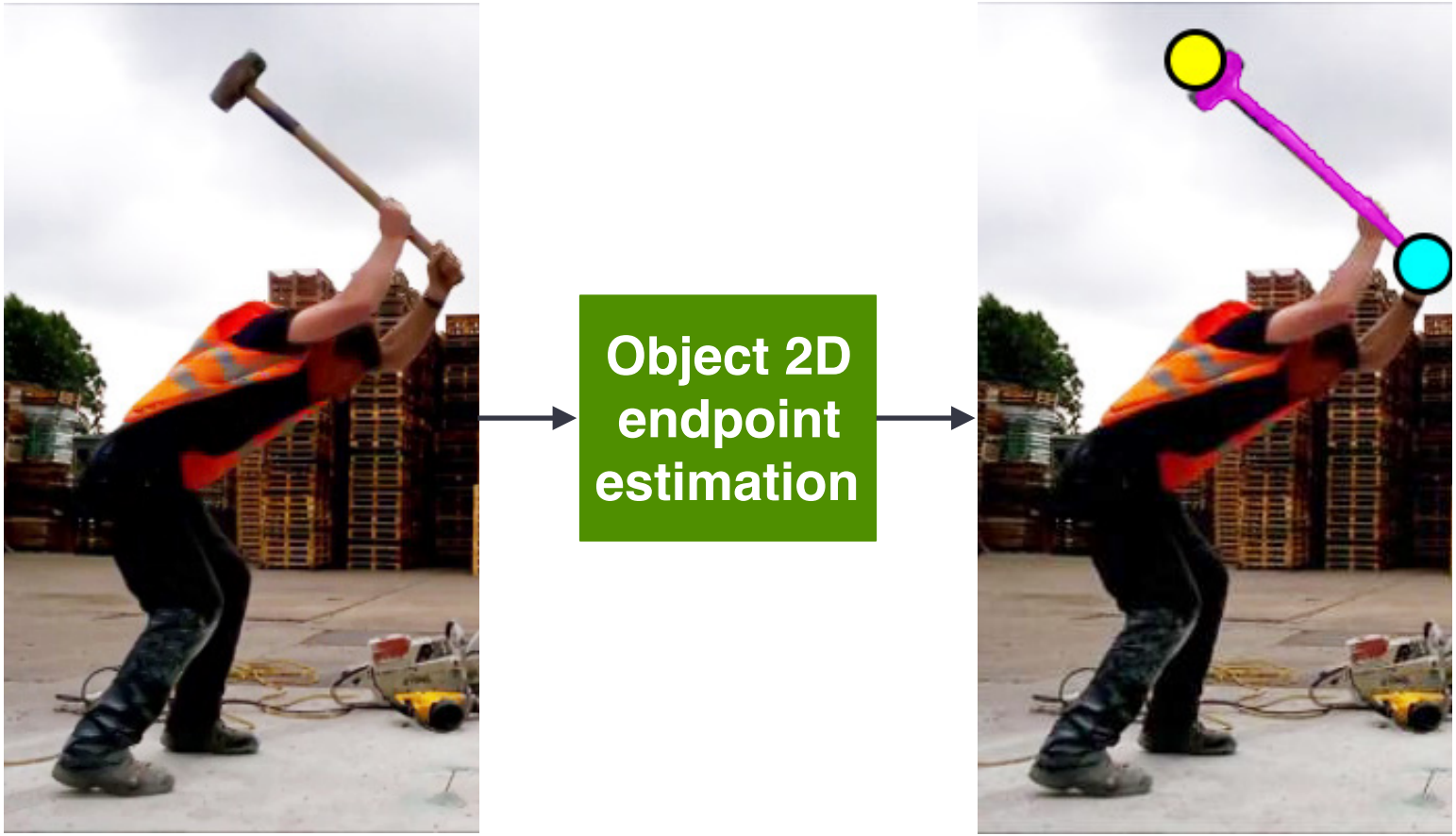
Contact recognizer

Method: a two-stage approach

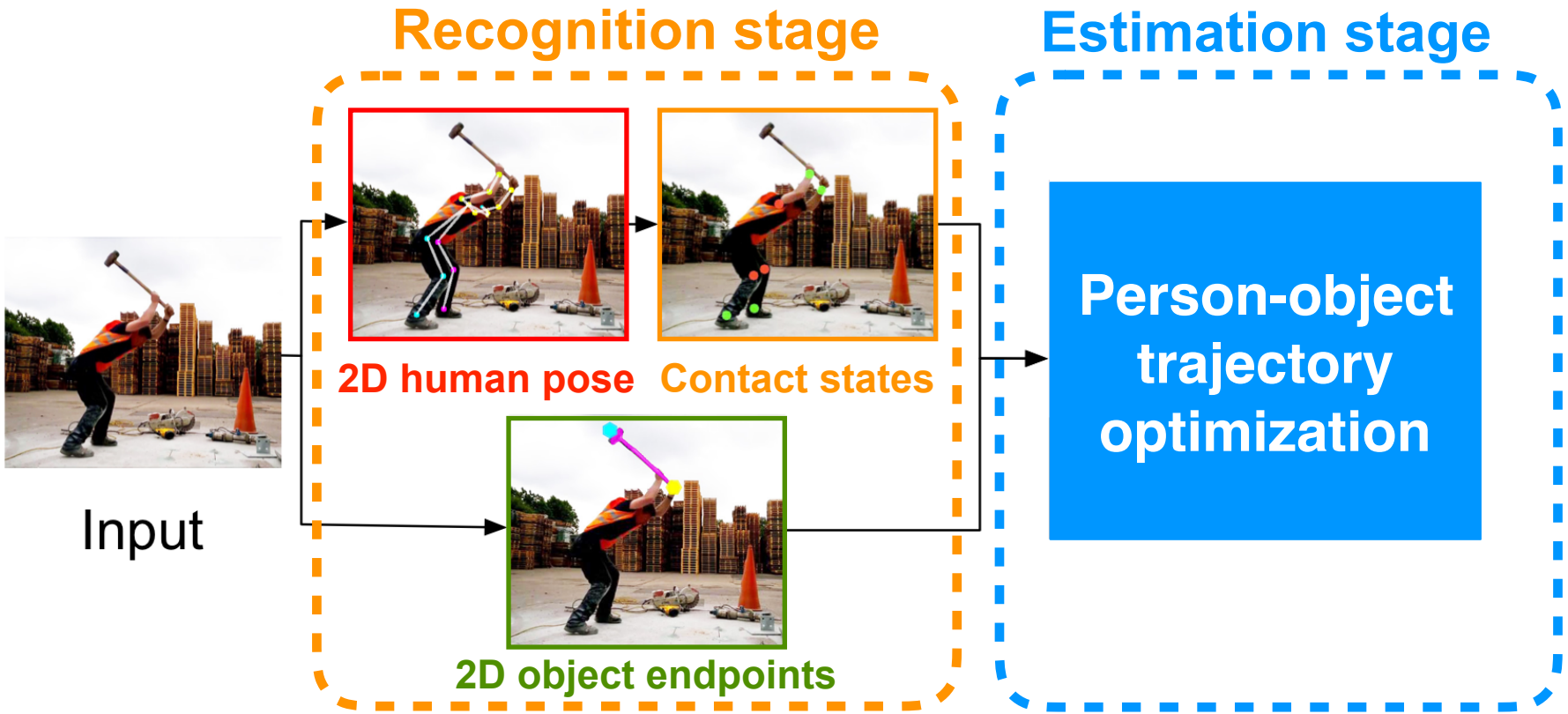


[He et al., ICCV 2017]

Method: a two-stage approach

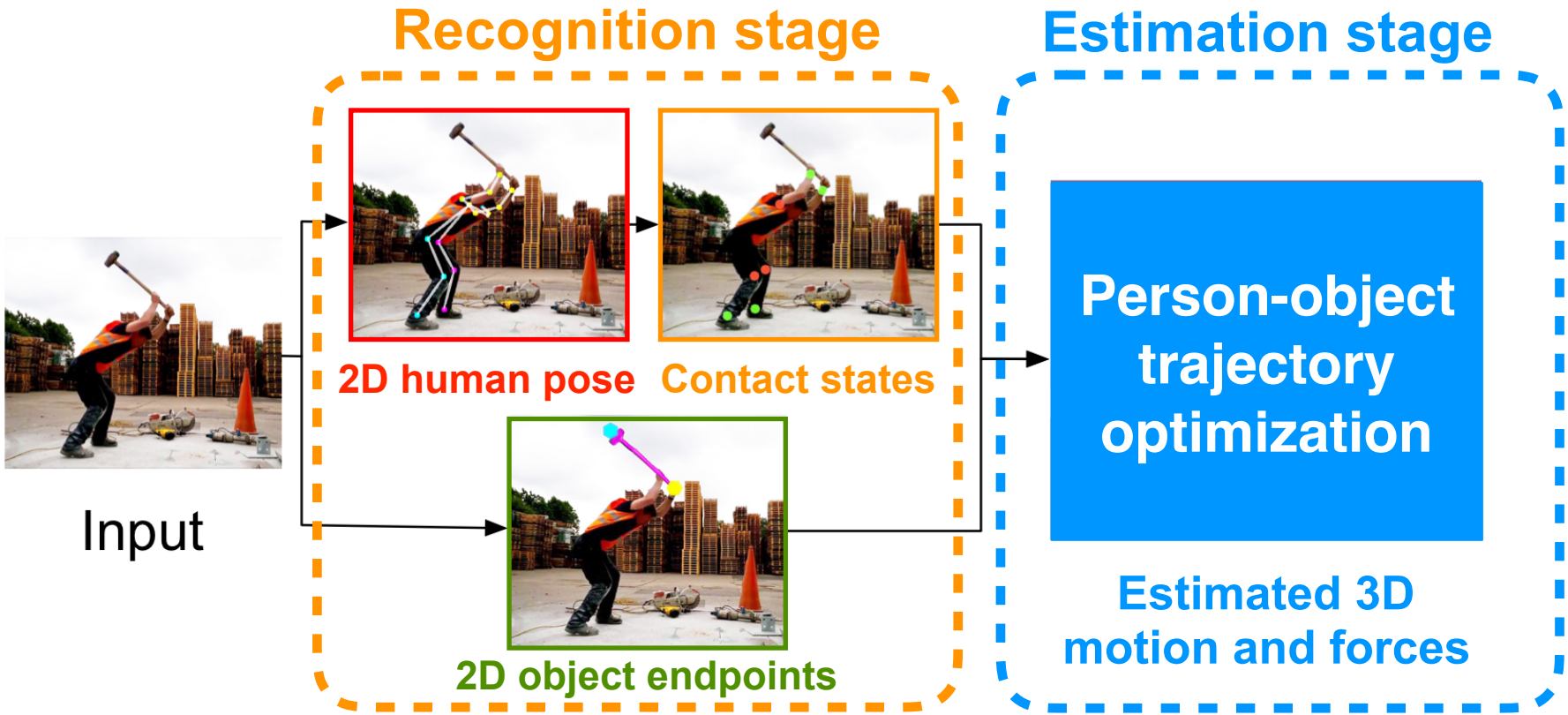


Method: a two-stage approach



[Tassa et al., IROS 2012]
[Carpentier et al., TRO, 2018]

Method: a two-stage approach

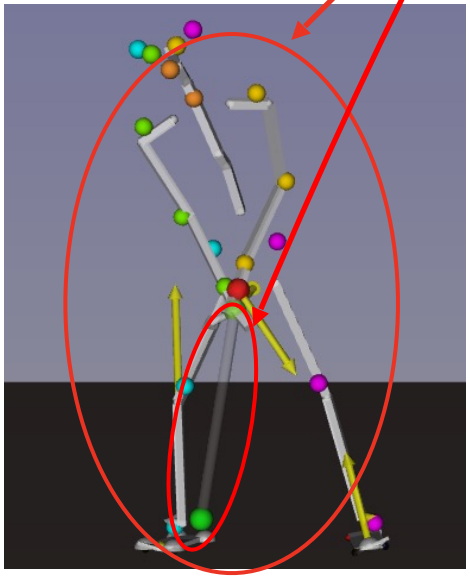


[Tassa et al., IROS 2012]
[Carpentier et al., TRO, 2018]

Estimation stage

$$\underset{\boldsymbol{x},\underline{u},\underline{c}}{\text{minimize}} \quad \sum_{e\in\{\mathrm{h,o}\}} \int_0^T l^e\left(\boldsymbol{x}, u, c\right) \mathrm{d}t$$

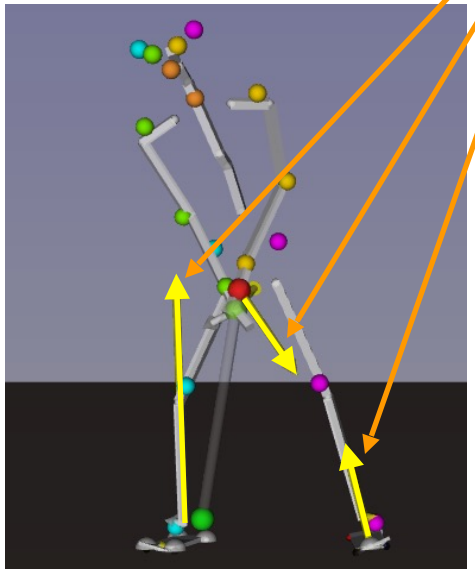
Person 3D poses



Object 3D poses

Estimation stage

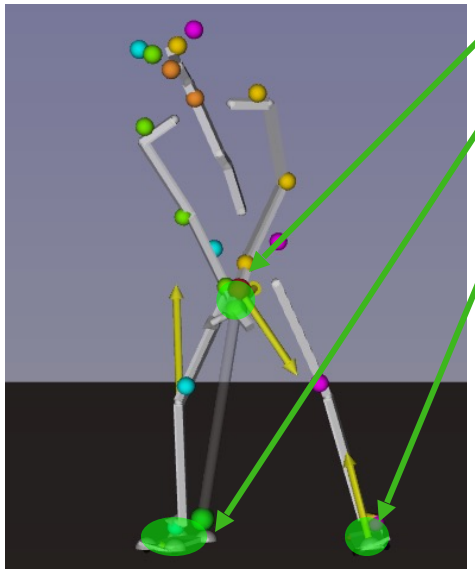
$$\underset{\underline{x},\boldsymbol{u},\underline{c}}{\text{minimize}} \quad \sum_{e\in\{\mathrm{h,o}\}} \int_0^T l^e\left(x, \boldsymbol{u}, c\right) \mathrm{d}t$$



Person-object
person-ground
contact forces

Estimation stage

$$\underset{\underline{x},\underline{u},\boldsymbol{c}}{\text{minimize}} \quad \sum_{e\in\{\mathrm{h,o}\}} \int_0^T l^e\left(x, u, \boldsymbol{c}\right) \mathrm{d}t$$



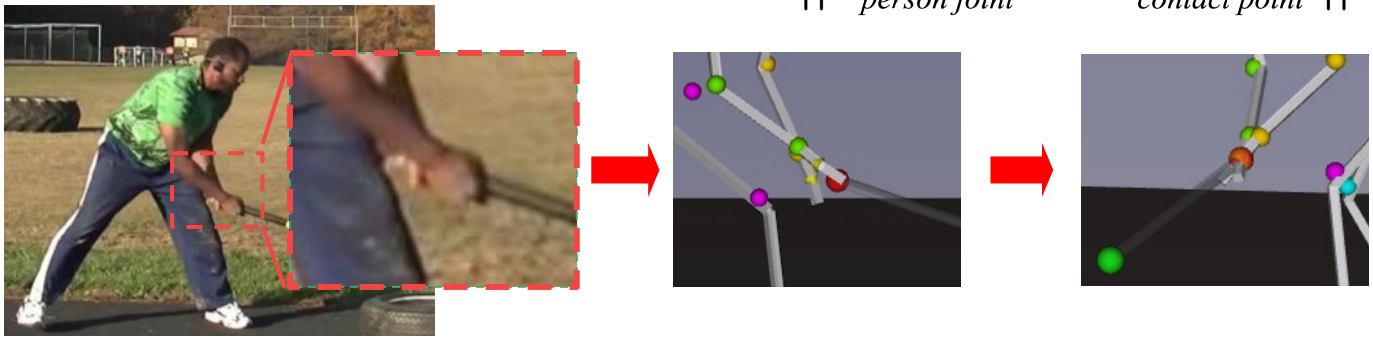
Contact positions

Estimation stage

$$\underset{\underline{x},\underline{u},\underline{c}}{\text{minimize}} \quad \sum_{e\in\{\mathrm{h},\mathrm{o}\}} \int_0^T l^e\left(x,u,c\right)\mathrm{d}t$$

Subject to

1. Contact motion model: $\left\| p_{\textit{person joint}} - p_{\textit{contact point}} \right\| = 0$



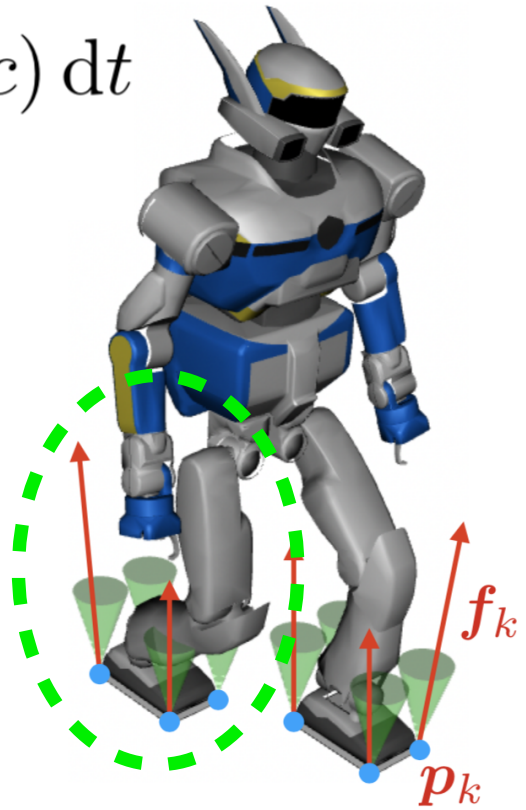
Estimation stage

$$\underset{\underline{x},\underline{u},\underline{c}}{\text{minimize}} \quad \sum_{e \in \{\text{h,o}\}} \int_0^T l^e\left(x, u, c\right) \mathrm{d}t$$

Subject to:

2. Contact force constraints:

- Prevent the feet from sliding



Estimation stage

$$\underset{\underline{x},\underline{u},\underline{c}}{\text{minimize}} \quad \sum_{e\in\{\text{h,o}\}} \int_0^T l^e\left(x,u,c\right)\mathrm{d}t$$

Subject to:

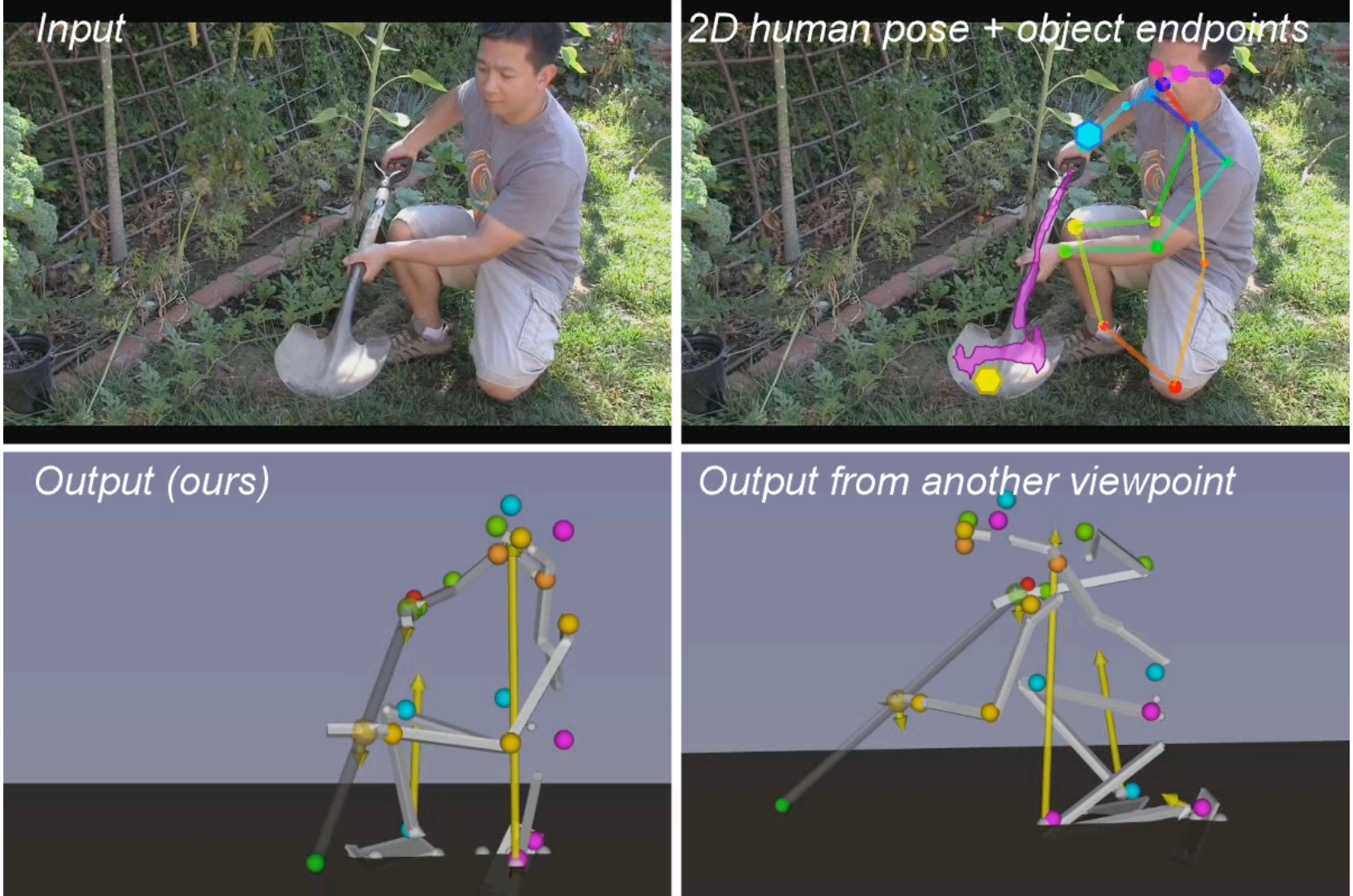
3. Lagrangian dynamics equation

$$M(x)\ddot{x} + b(x,\dot{x}) = \tau(u,c)$$

[Carpentier, et al. Pinocchio. https://stack-of-tasks.github.io/pinocchio. 2015-2019]

Experiments

A new dataset: Handtool videos


Input
2D human pose + object endpoints
Output (ours)
Output from another viewpoint

CVPR 2019, Long Beach, California

At the poster (#202):

- Additional details
- Quantitative results
- Discussion

Zongmian Li [1,2], Jiri Sedlar [3], Justin Carpentier [1,2]

Ivan Laptev [1,2], Nicolas Mansard [4], Josef Sivic [1,2,3]

[1] Inria [2] DI ENS, PSL [3] CIIRC, CTU in Prague [4] LAAS-CNRS