







INTRODUCTION

contact forces from a single RGB video.





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

Zongmian Li^{1,2}, Jiri Sedlar³, Justin Carpentier^{1,2}, Ivan Laptev^{1,2}, Nicolas Mansard⁴, Josef Sivic^{1,2,3} ¹Ecole Normale Supérieure, PSL ³CIIRC, CTU in Prague ²INRIA ⁴LAAS-CNRS

- Mask R-CNN trained for each type of object
- Synthetic and labeled
- Infer endpoints from

- "In the wild" Internet instructional videos
- > 4 manipulation actions,
- Annotated human 3D & object 2D poses and

Input

[1] Z. Cao, et. al., Realtime multi-person 2d pose estimation using part affinity fields. In CVPR, 2017. [2] F. Bogo, et. al., Keep it SMPL: Automatic estimation of 3dhuman pose and shape from a single image. In ECCV, 2016. [3] A. Kanazawa, et. al., End-to-end recovery of human shape and pose. In CVPR 2018.

RESULTS: PARKOUR DATASET

Description > A MoCap dataset with ground truth motion and forces

	Jump	Move-up	Pull-up	Нор	Avg	
SMPLify [2]	121.75	147.41	<u>120.48</u>	169.36	139.69	
HMR [3]	<u>111.36</u>	<u>140.16</u>	132.44	<u>149.64</u>	135.65	
Ours	98.42	125.21	119.92	138.45	122.11	
	L. So	ole R.	Sole	L. Hand	R. Hand	
Force (N)	144.	23 13	8.21	107.91	113.42	
Moment (N.m)	23.7	71 22	2.32	131.13	134.21	
			A state of the sta		With the second se	
					1 •	

RESULTS: HANDTOOL DATASET

S		Barbell	Spade	Hammer	Scythe	Avg
	SMPLify [2]	130.69	135.03	93.43	112.93	118.02
	HMR [3]	<u>105.04</u>	<u>97.18</u>	96.34	115.42	<u>103.49</u>
sults	Ours	104.23	95.21	<u>95.87</u>	<u>114.22</u>	102.38

Rotated view

Output Rotated view Output lnput