



Video Inference for Human Body Pose and Shape Estimation

Muhammed Kocabas^{1,2}, Nikos Athanasiou¹, Michael J. Black¹

¹Max Planck Institute for Intelligent Systems,

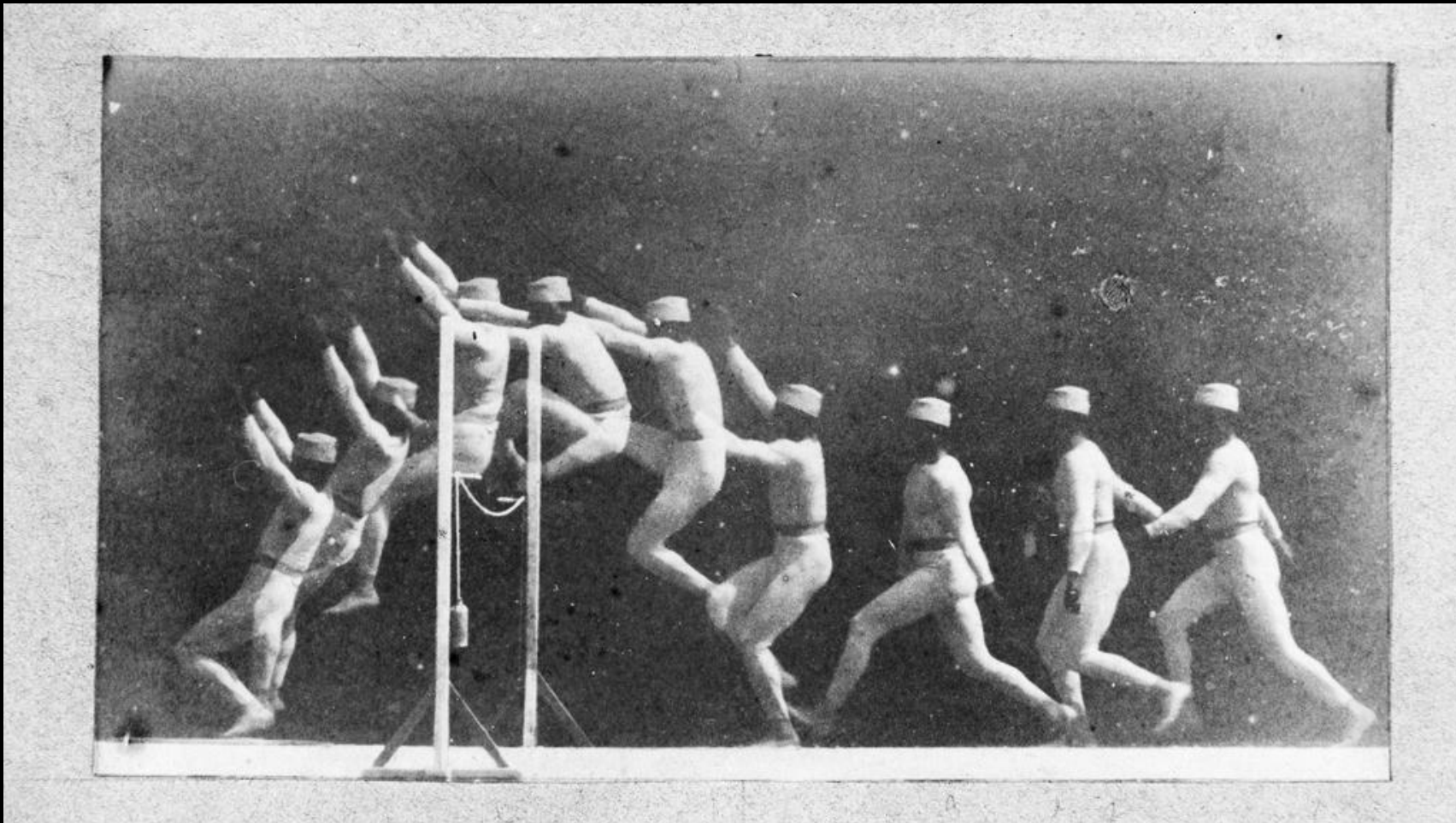
²Max Planck ETH Center for Learning Systems



MOVIECLIPS.COM

MOVIECLIPS.COM

Bodies in Motion



E. J. Marey, c.1892

Bodies in Motion



D Hogg, Image and Vision Computing, Vol 1 (1983)

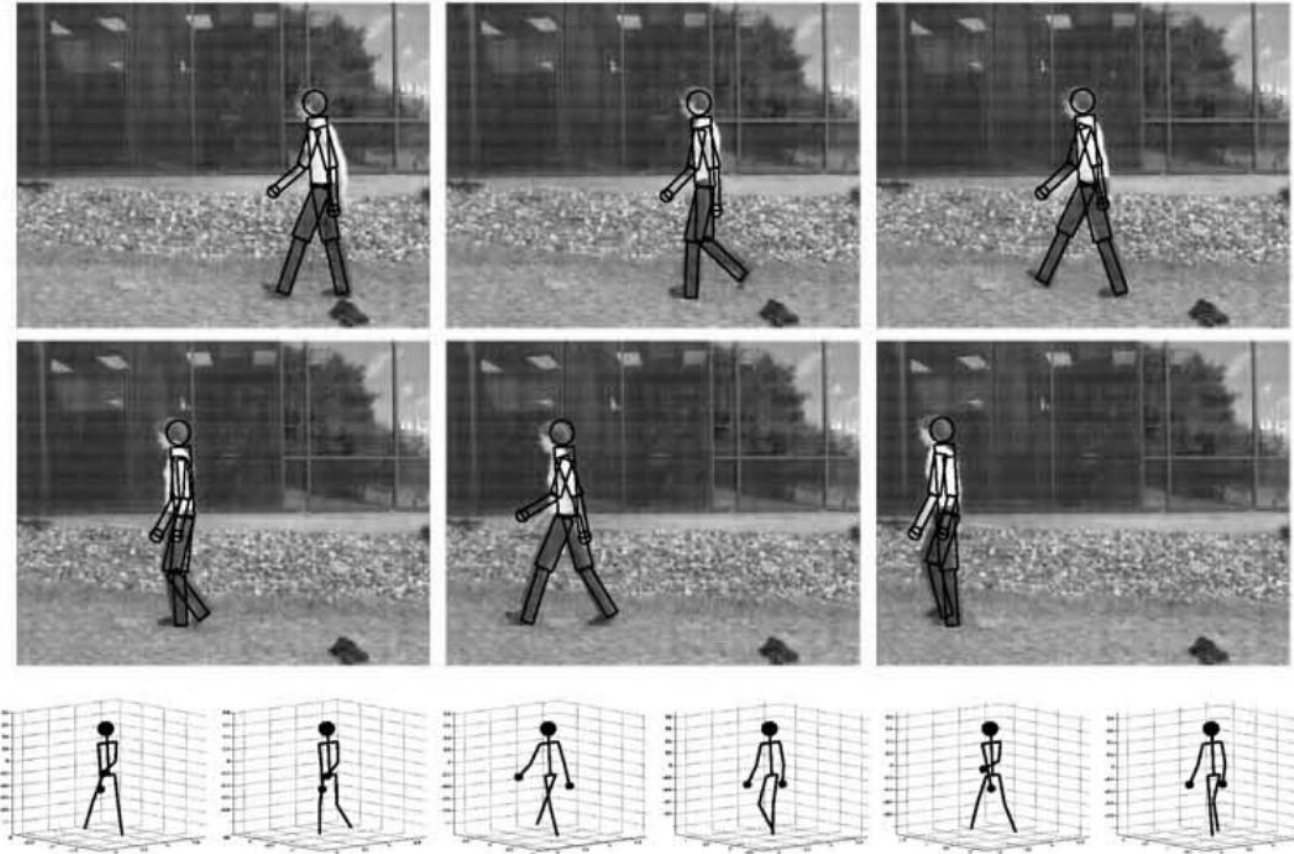
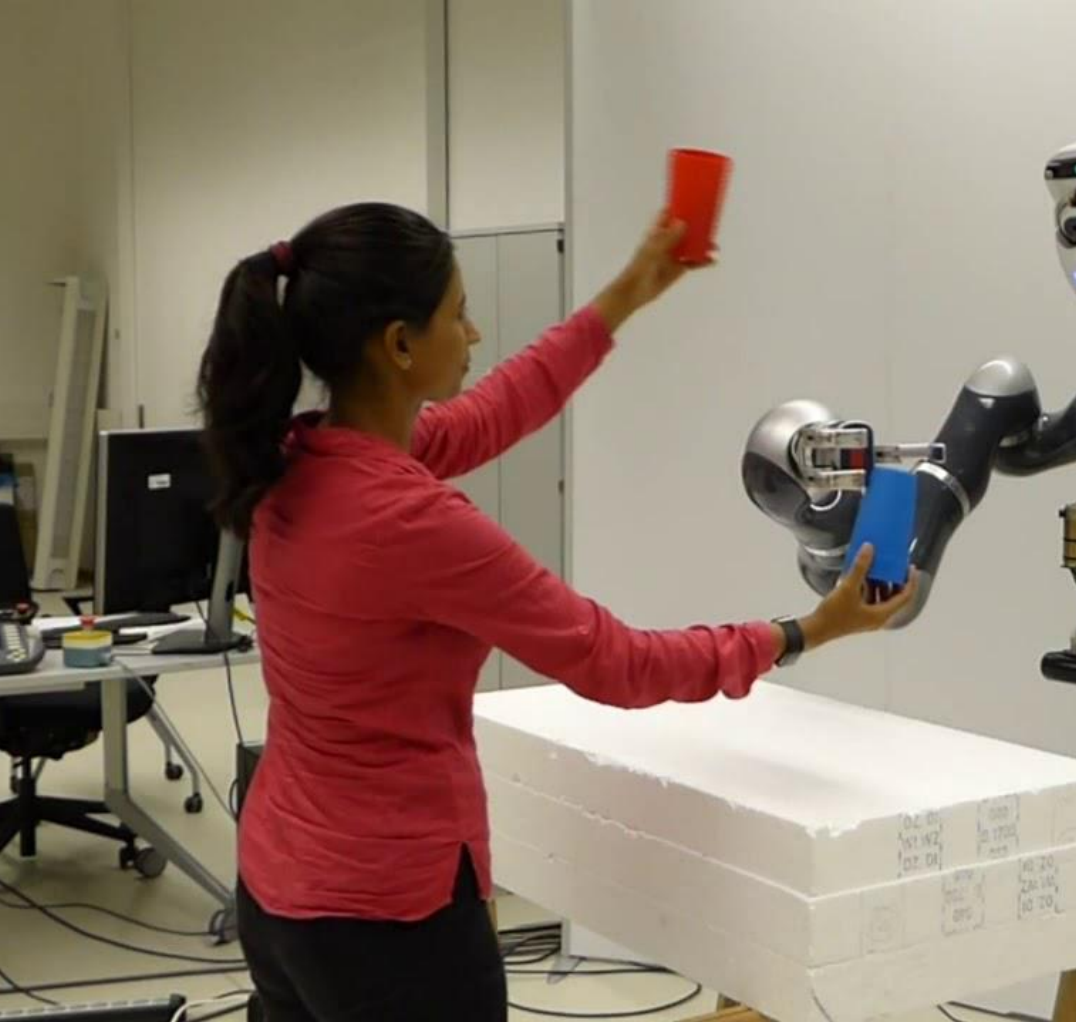


Figure 2: Tracking of person walking, 10000 samples. Upper rows: frames 0, 10, 20, 30, 40, 50 with the projection of the expected model configuration overlaid. Lower row: expected 3D configuration in the same frames.









chair

keyboard

desk

keyboard

desk



chair

keyboard

chair

desk

keyboard

desk



person 100%

chair 70%

tie 95%

tv 66%
laptop 53%

book 63%

book 95%

book 92%

book 64%

tv 77%

person 94%

chair 76%

tie 74%

book 51%

person 97%

laptop 55%

person 100%

person 98%

tie 61%

person 93%

laptop 67%

laptop 70%

tie 94%

tv 73%

mouse 7%

cup 99%

book 76%

mouse 98%

DWIGHT K. SCHRUTE

tv 68%

book 60%

book 80%

book 66%

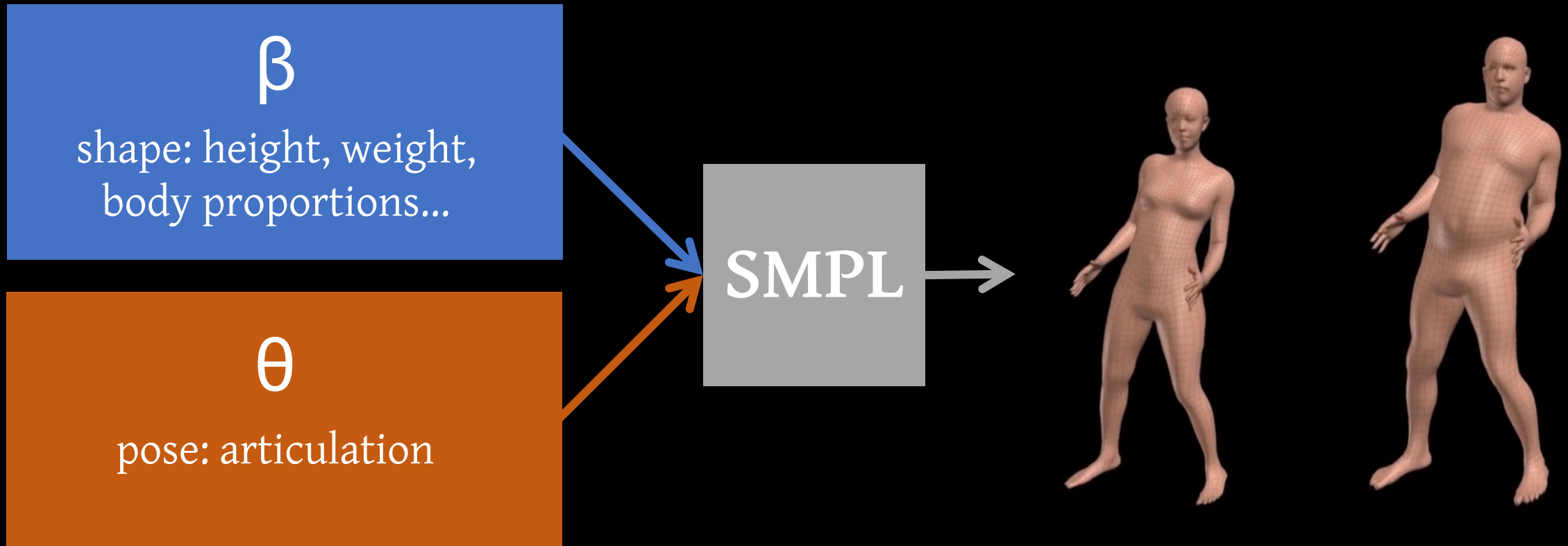
book 54%

book 65%

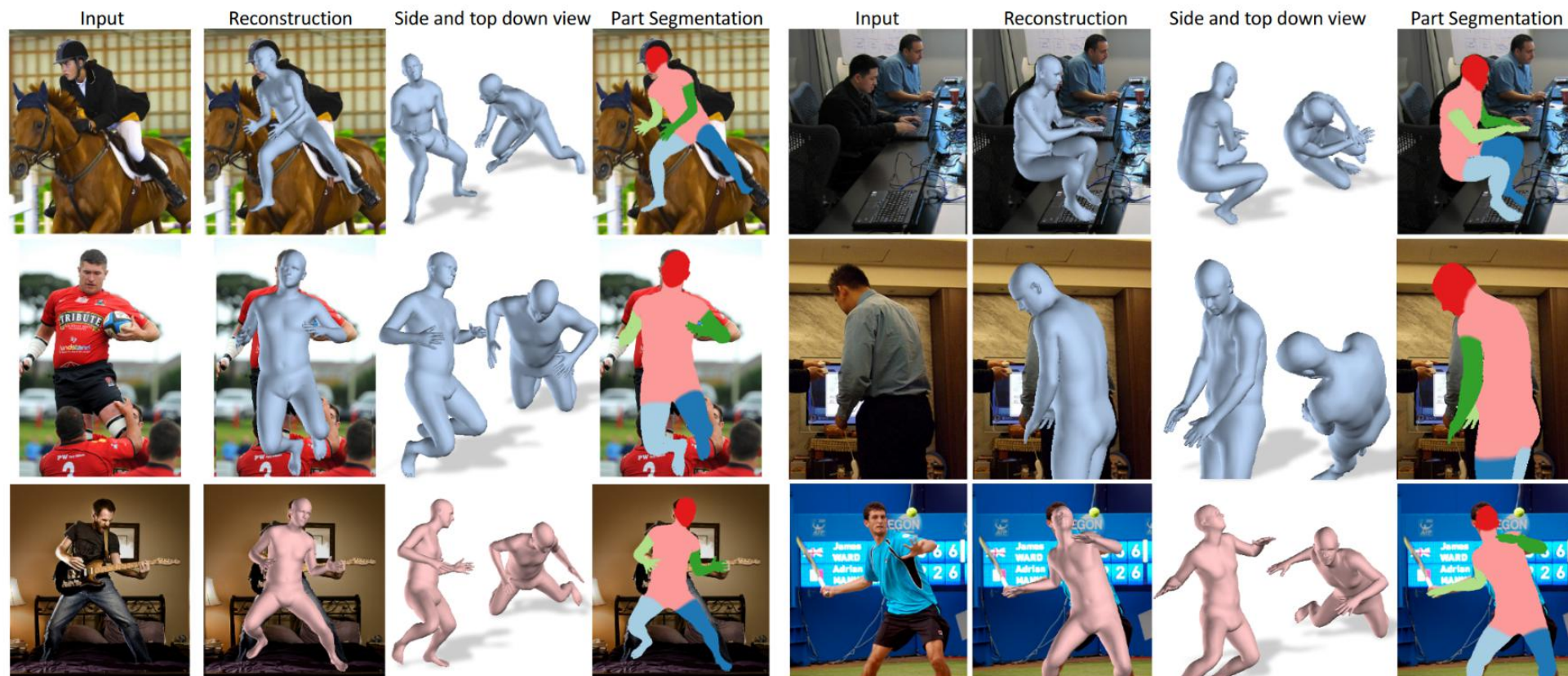
cup 62%



SMPL Body Model



Human Mesh Recovery (HMR)

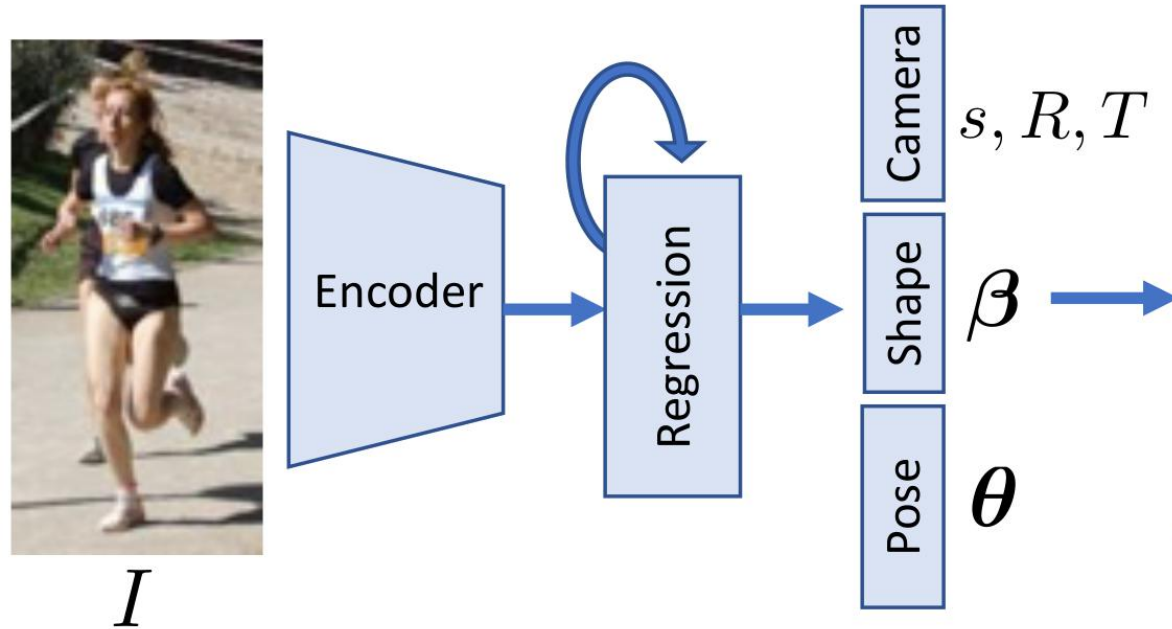


Human Mesh Recovery (HMR)

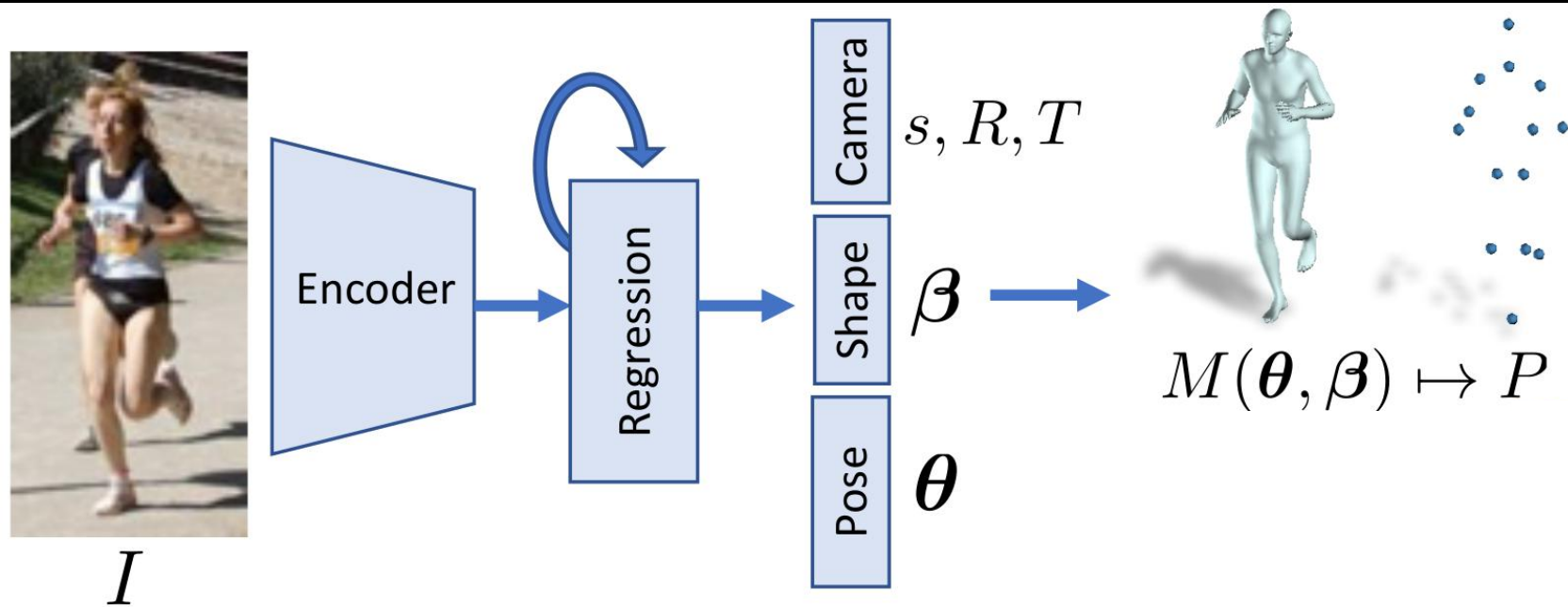


I

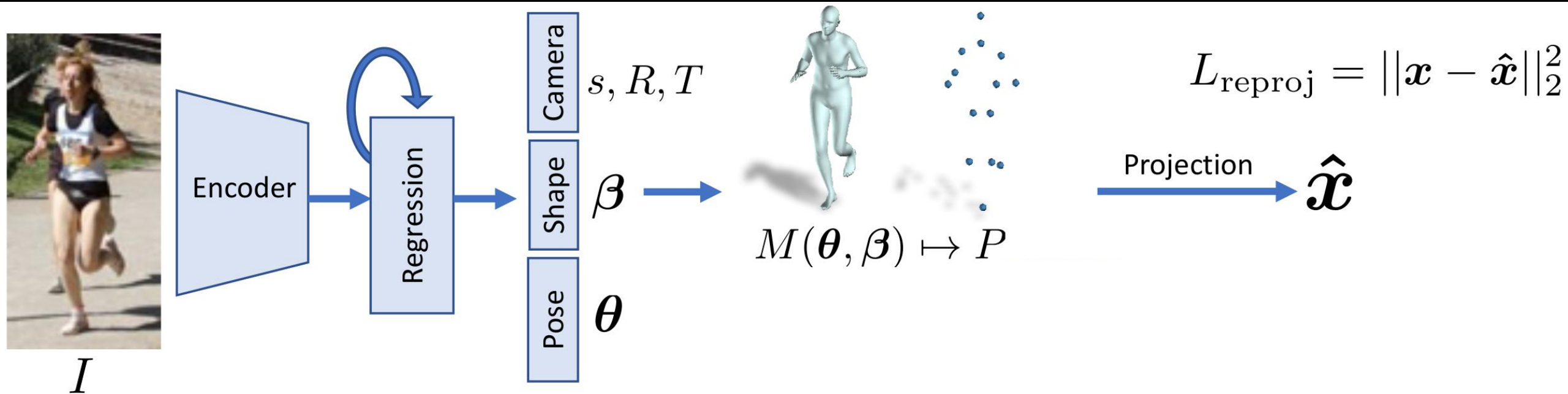
Human Mesh Recovery



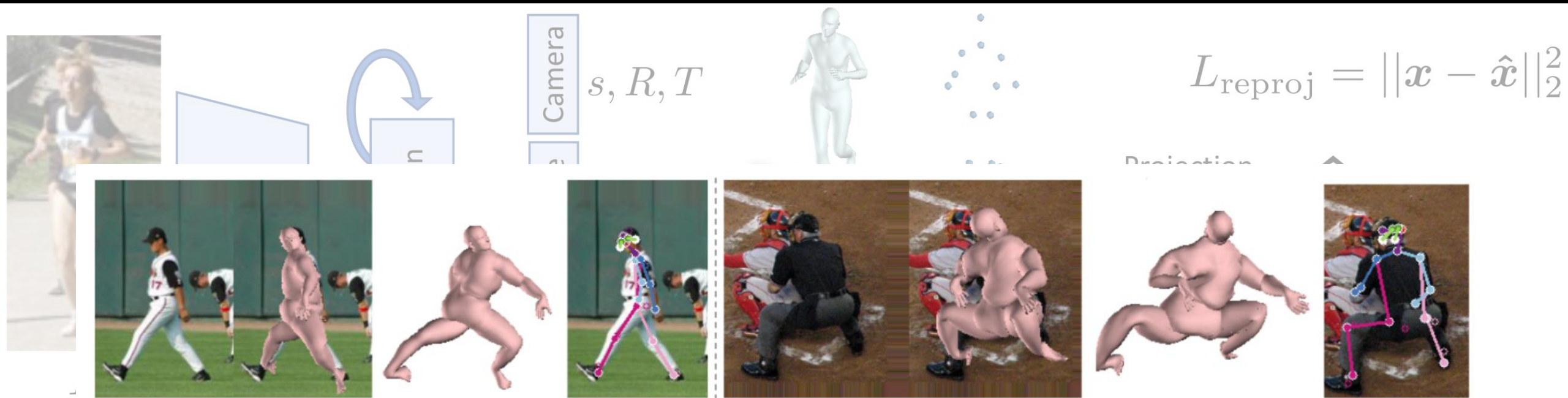
Human Mesh Recovery



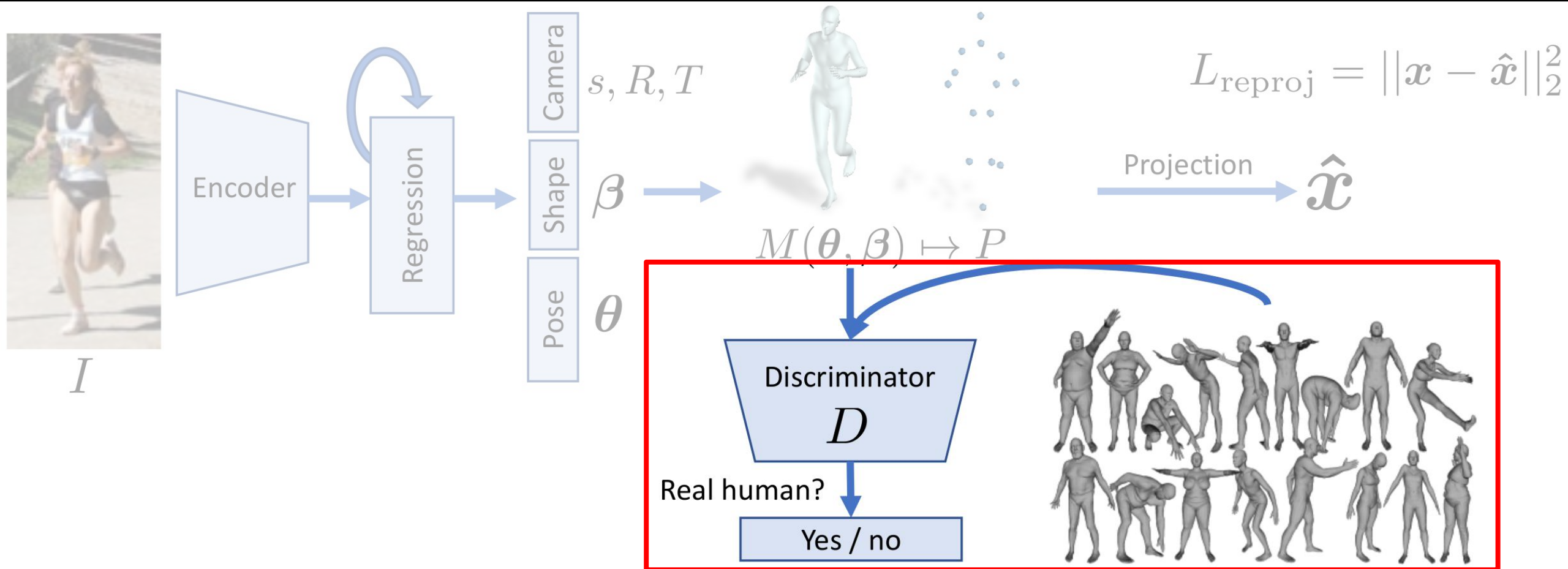
Human Mesh Recovery



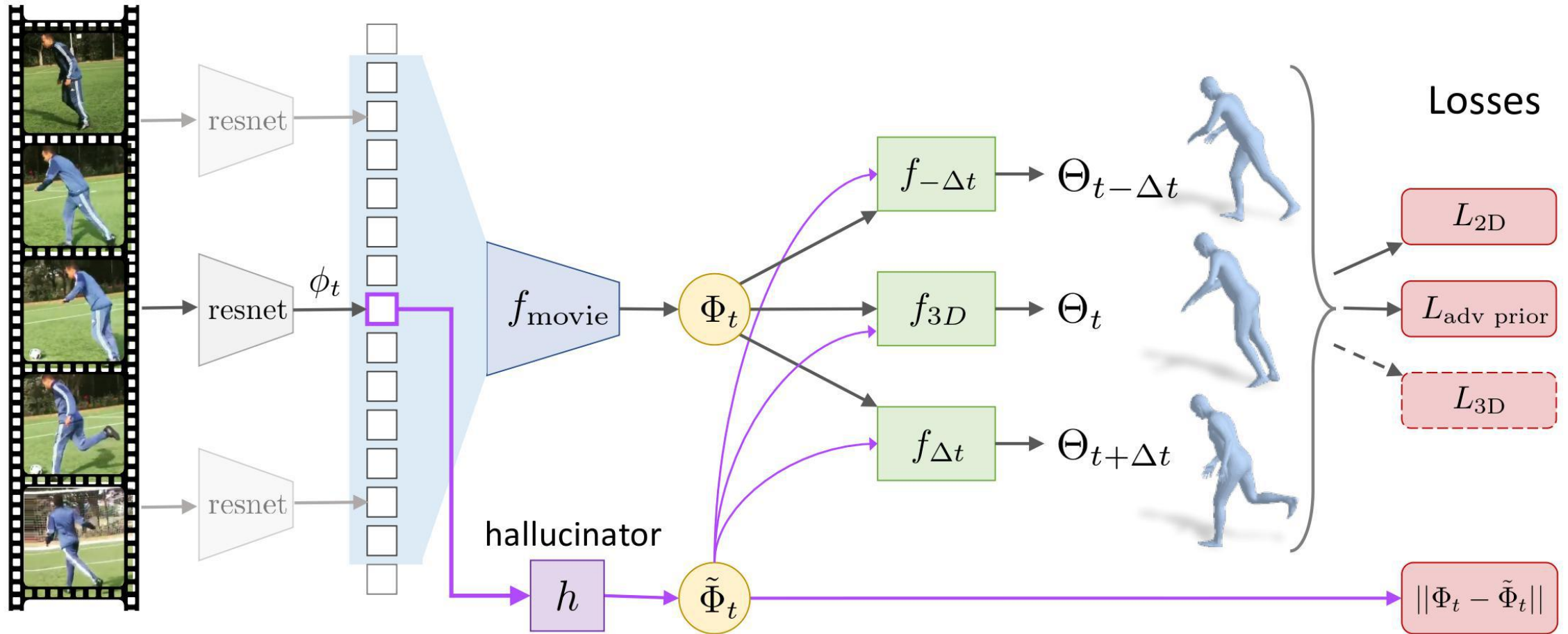
Human Mesh Recovery



Inspiration



Temporal HMR



Temporal HMR Result



Limitations

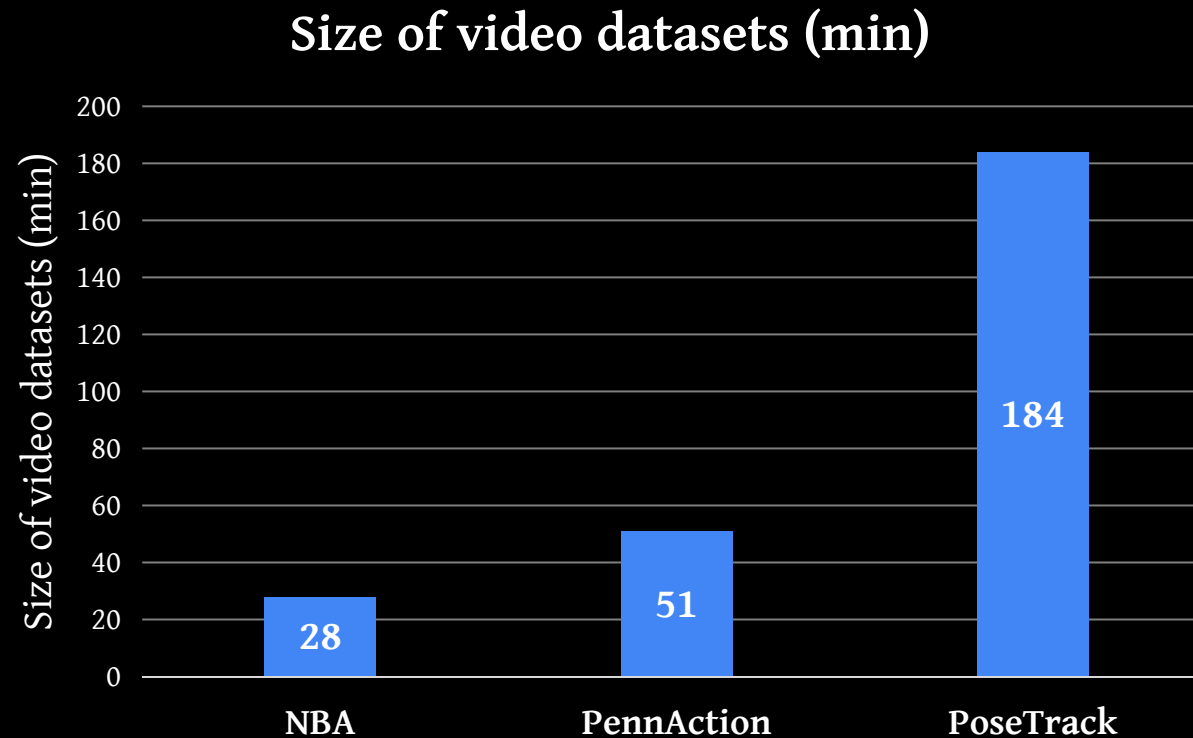
(1) Indoor 3D datasets are limited in:

- number of subjects,
- range of motions,
- and image complexity.



Limitations

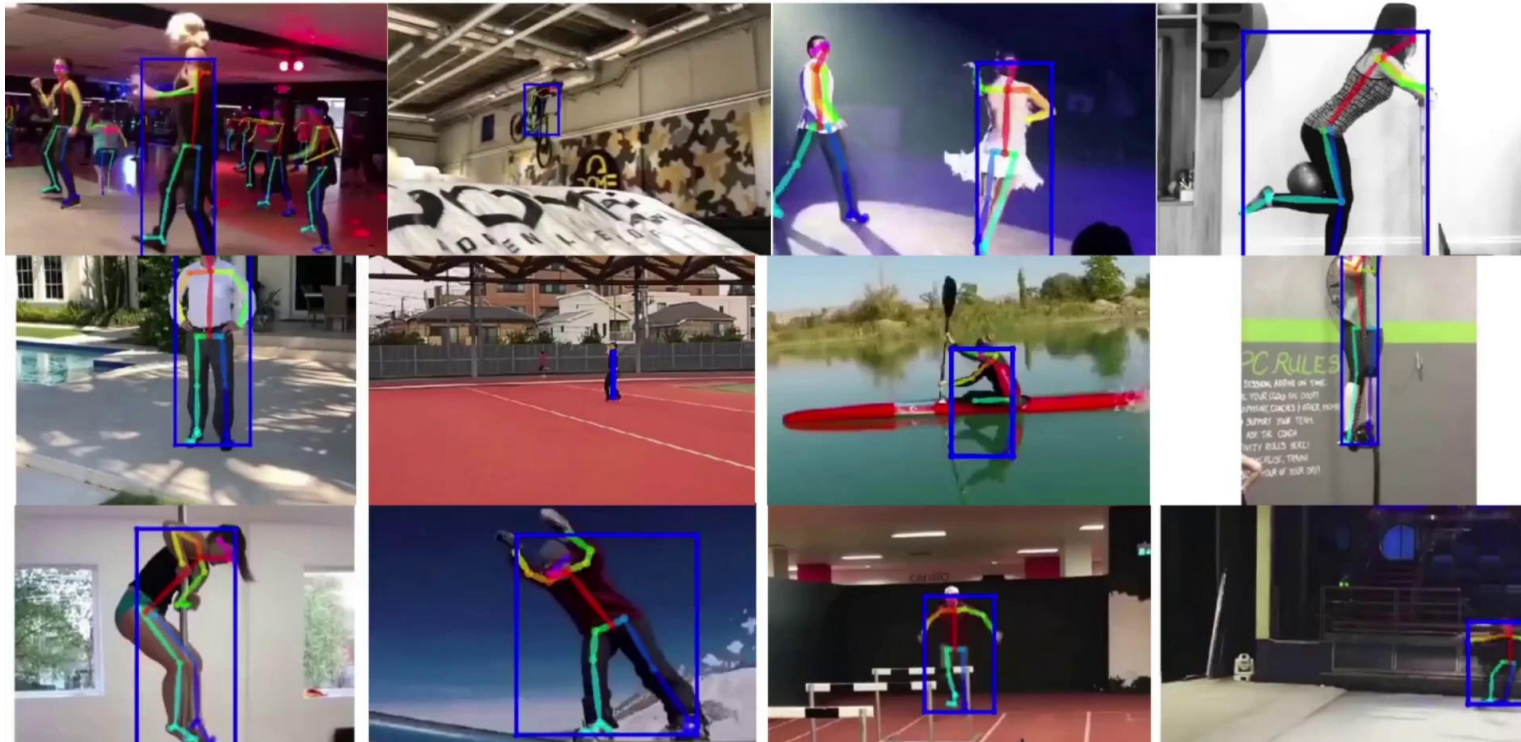
(2) Videos with ground-truth 2D pose labels are insufficient



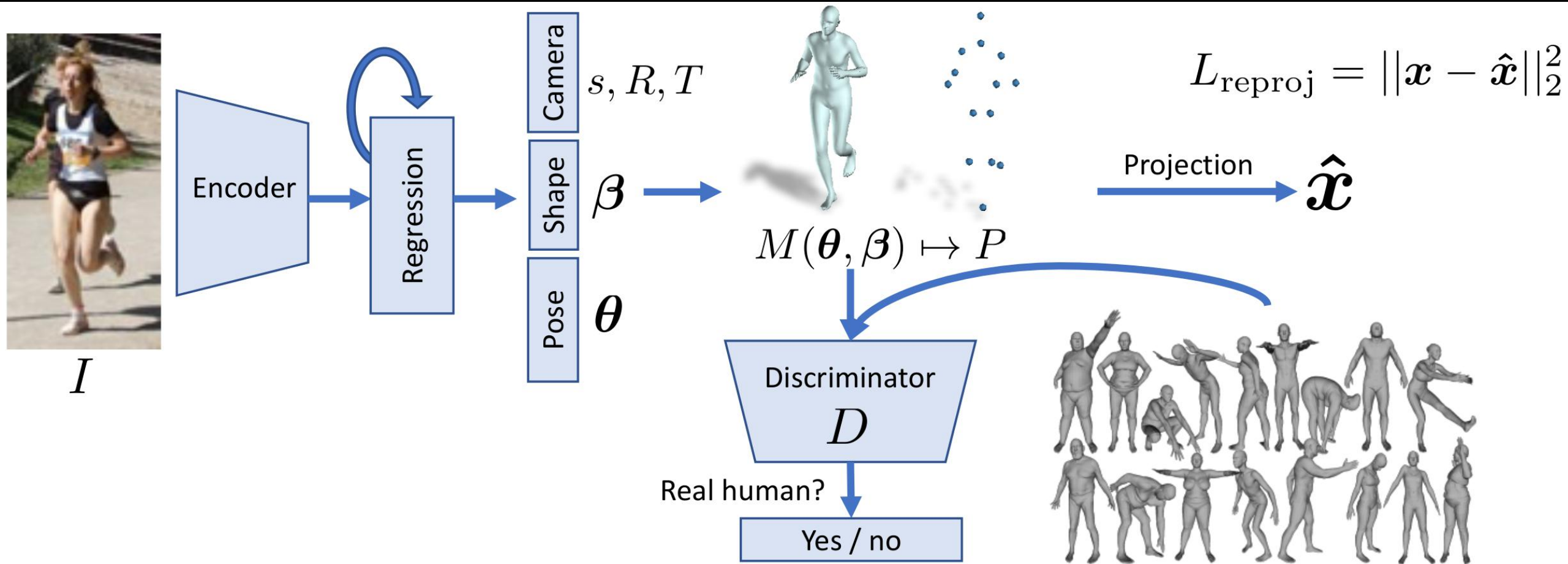
Limitations

(3) pseudo-ground-truth 2D labels are not reliable

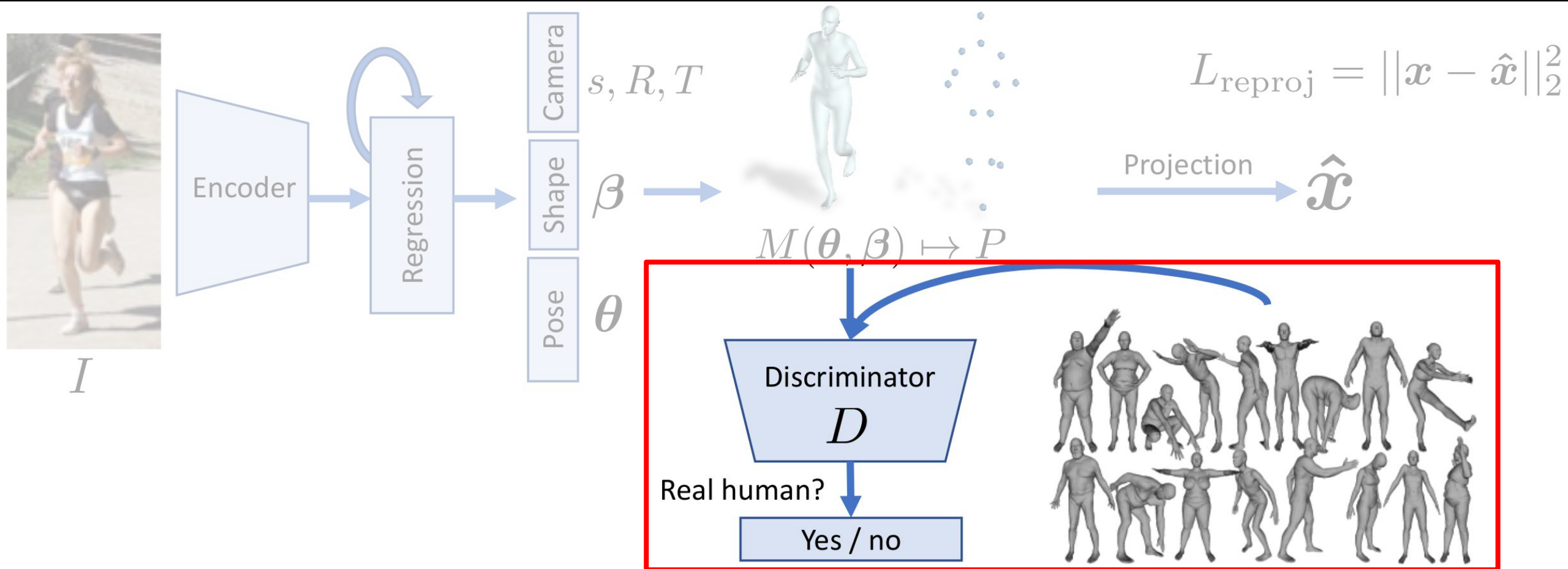
InstaVariety



Inspiration

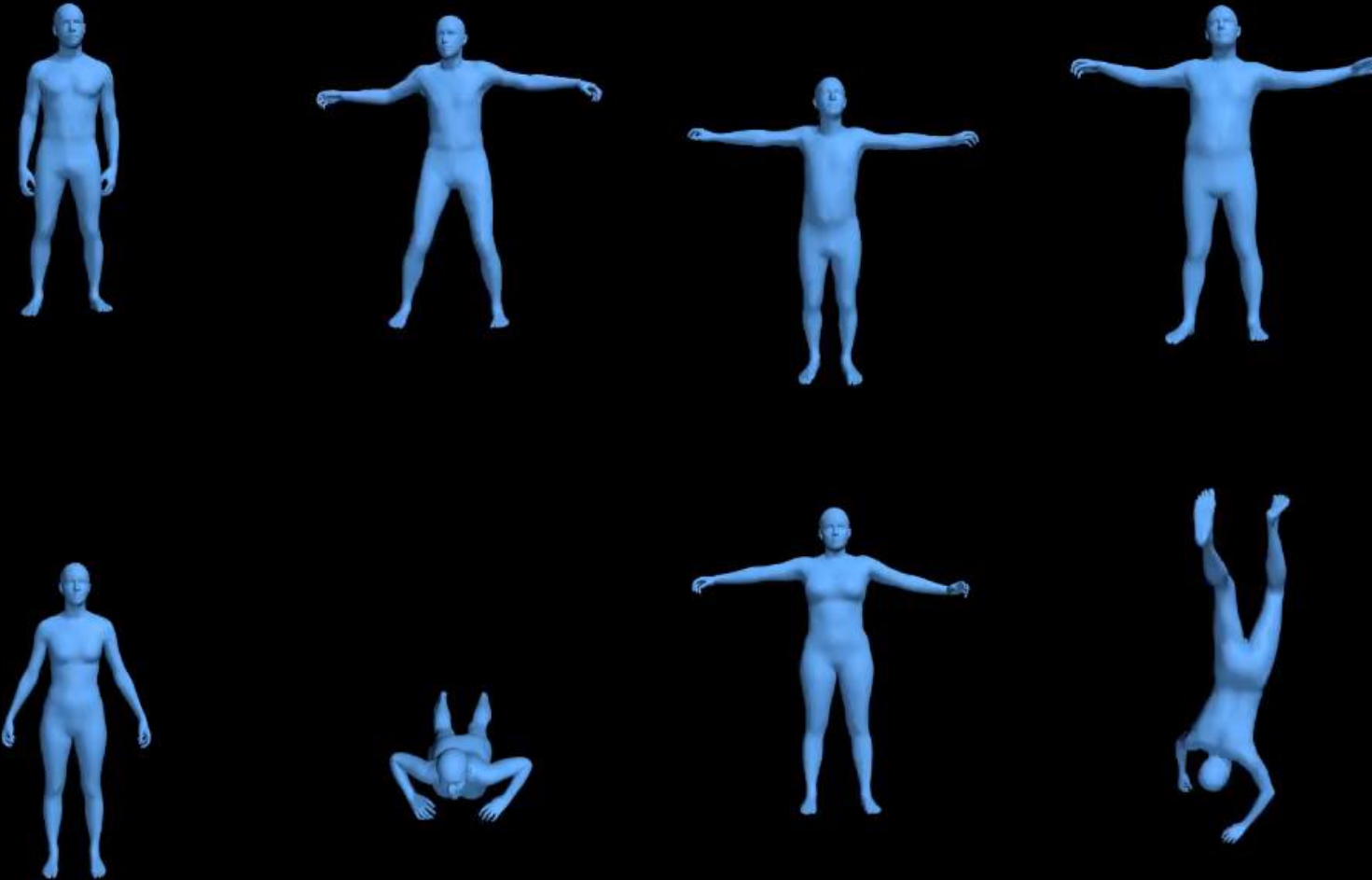


Inspiration



AMASS

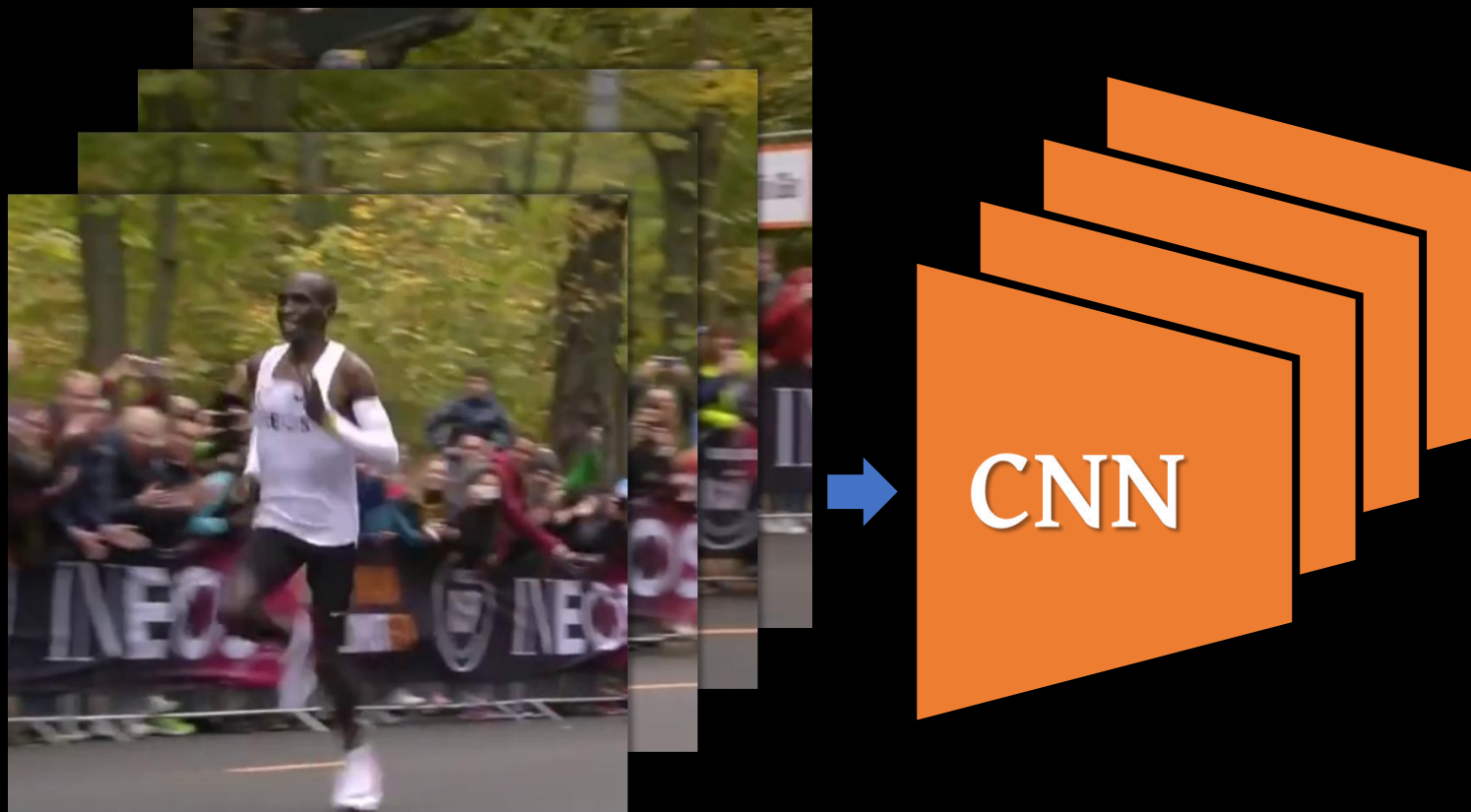
Archive of Motion Capture as Surface Shapes



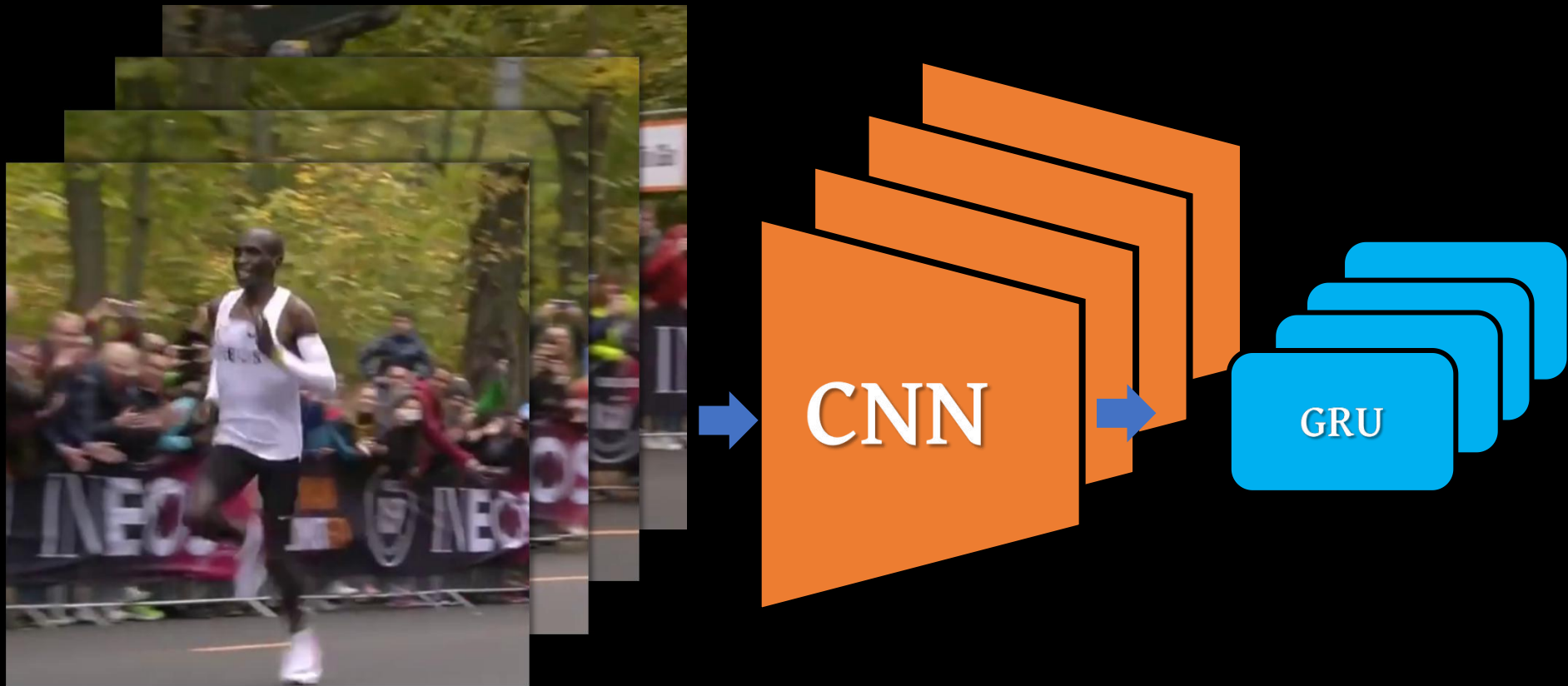


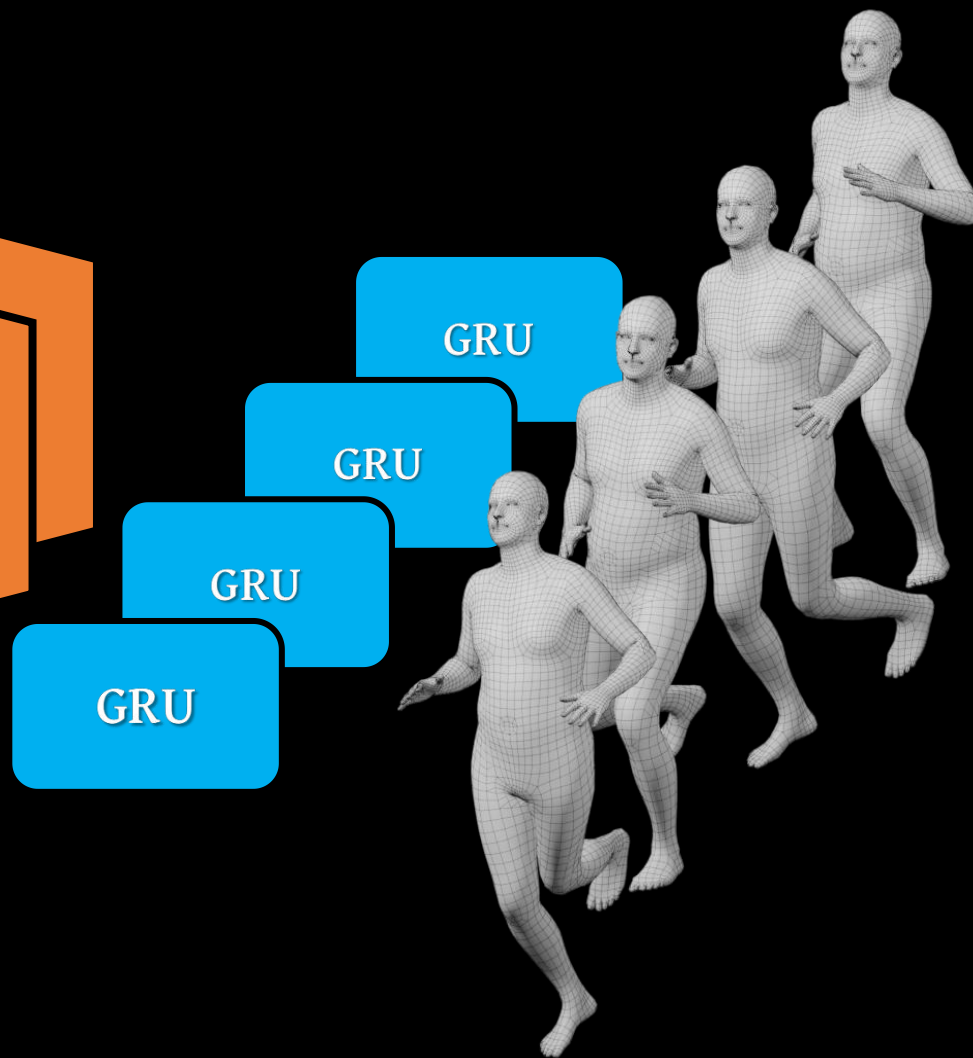
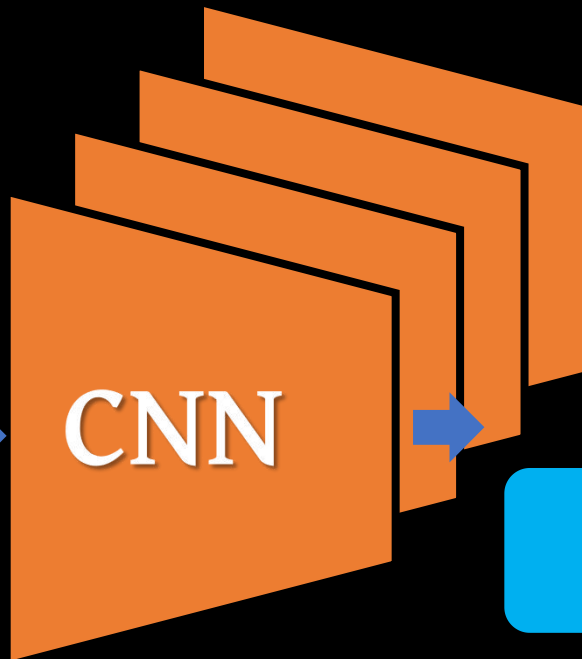
Approach



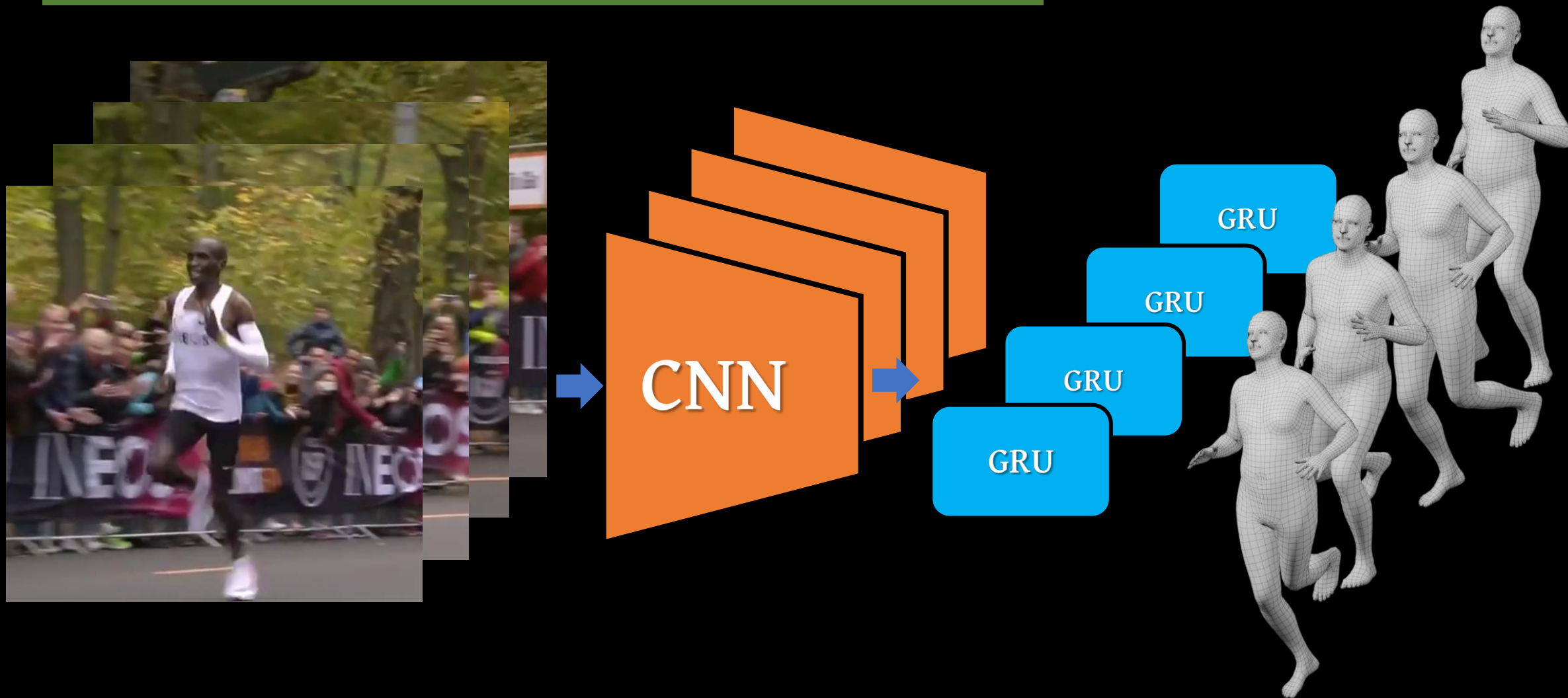


*CNN feature extractor is taken from SPIN (Kolotouros et al., ICCV19) and not updated during training





Generator



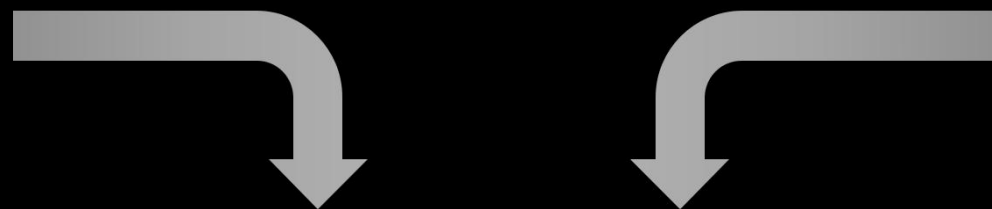
Predictions



AMASS
Archive of Motion Capture

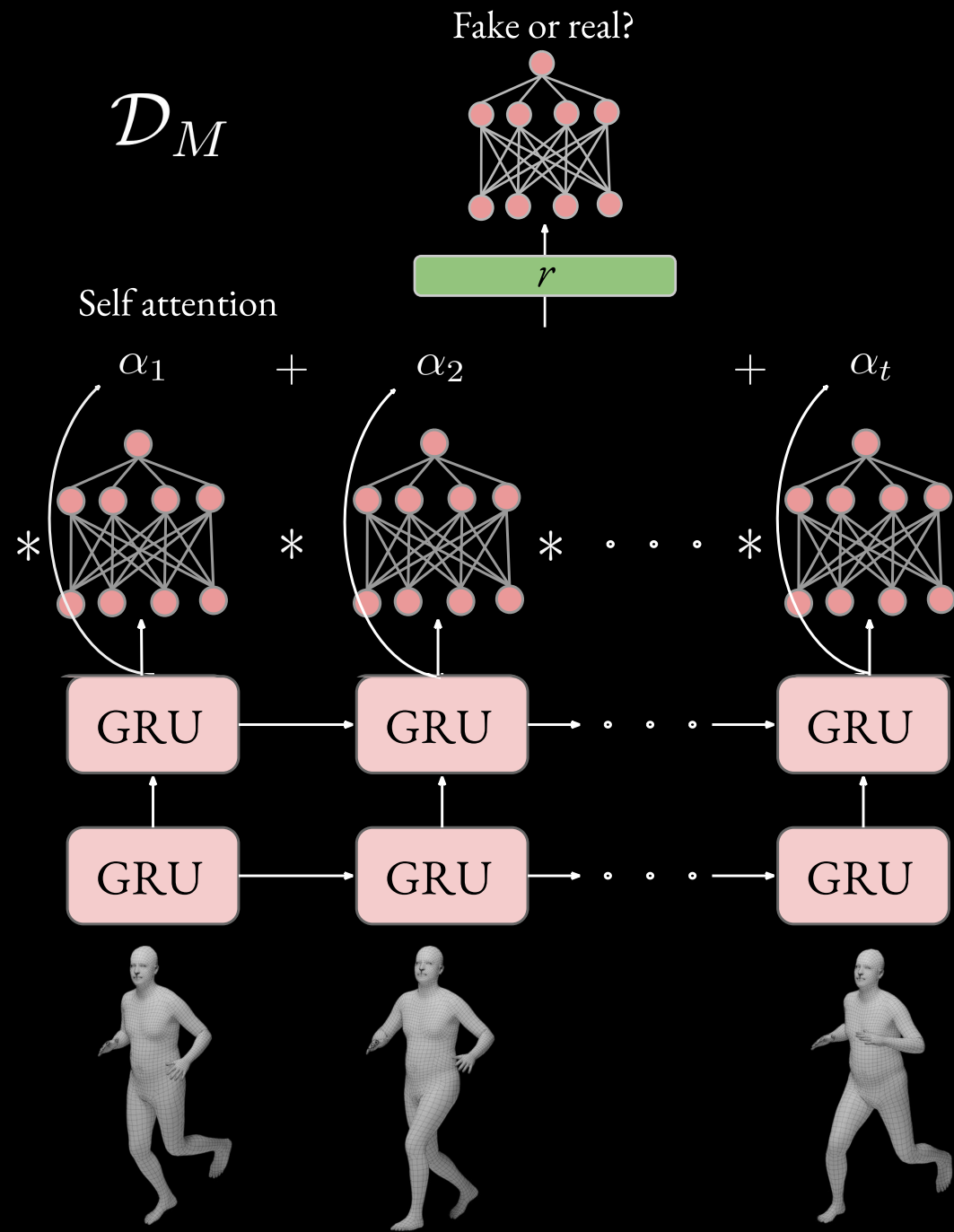


Realistic Motion?



Motion
Discriminator





VIBE Loss

$$L = L_{2D} + L_{3D} + L_{\theta} + L_{\beta} + L_{D_M}$$

2D keypoint projection loss

3D joint loss*

SMPL Pose and Shape loss*

Motion Discriminator loss

*when 3D annotations are available

Training Datasets

2-D datasets

Examples of 2-D Datasets

3-D datasets



Results

Evaluation Dataset



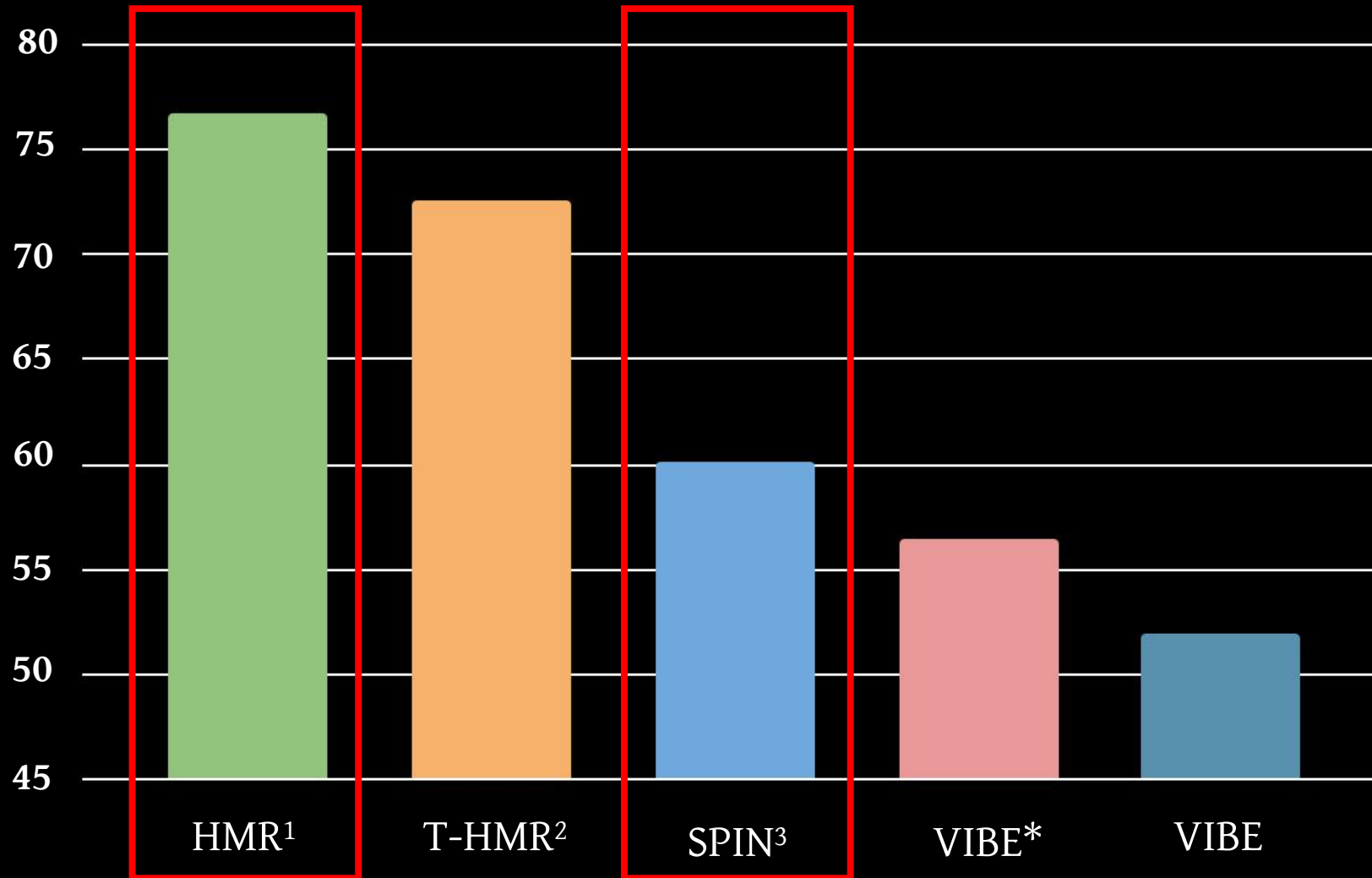
3D poses in-the-wild dataset

Eval Metric:

MPJPE

(Mean per joint position error)

Reconstruction Error (mm) in 3DPW Test Set



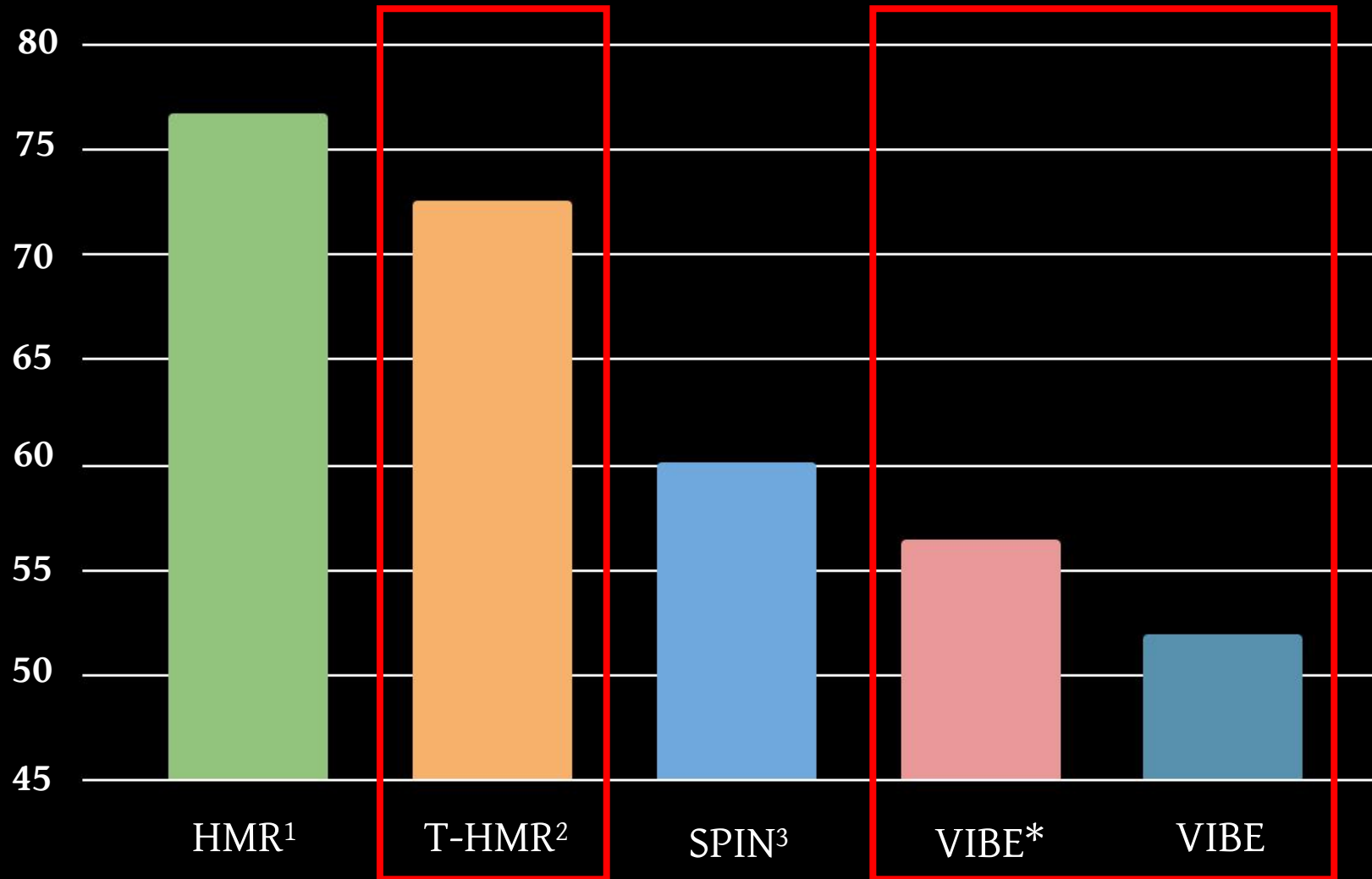
¹Kanazawa *et al.*, End-to-end Recovery of Human Shape and Pose, CVPR 2018.

²Kanazawa *et al.*, Learning 3D Human Dynamics from Video, CVPR 2019

³Kolotouros *et al.*, Learning to Reconstruct 3D Human Pose and Shape via Model-fitting in the Loop, ICCV 2019.

VIBE* denotes the model trained with datasets identical to T-HMR.

Reconstruction Error (mm) in 3DPW Test Set



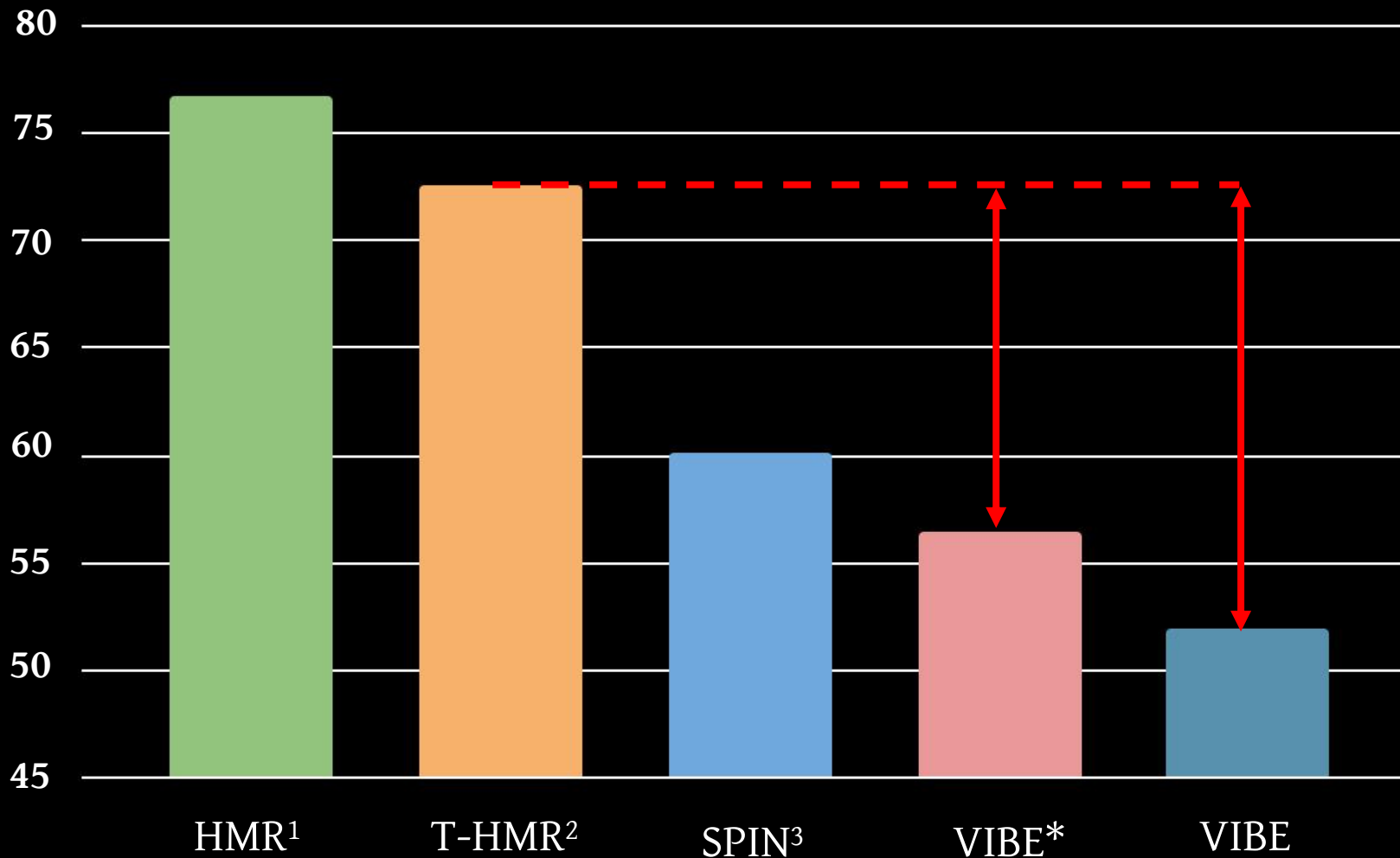
¹Kanazawa *et al.*, End-to-end Recovery of Human Shape and Pose, CVPR 2018.

²Kanazawa *et al.*, Learning 3D Human Dynamics from Video, CVPR 2019

³Kolotouros *et al.*, Learning to Reconstruct 3D Human Pose and Shape via Model-fitting in the Loop, ICCV 2019.

VIBE* denotes the model trained with datasets identical to T-HMR.

Reconstruction Error (mm) in 3DPW Test Set



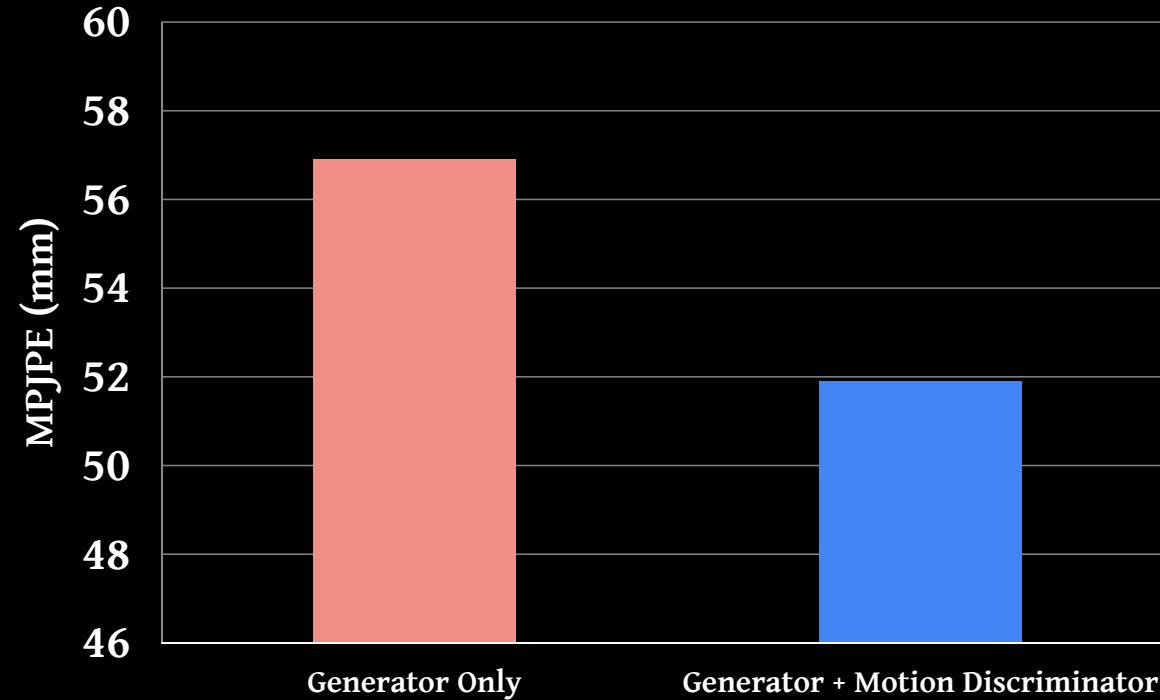
¹Kanazawa *et al.*, End-to-end Recovery of Human Shape and Pose, CVPR 2018.

²Kanazawa *et al.*, Learning 3D Human Dynamics from Video, CVPR 2019

³Kolotouros *et al.*, Learning to Reconstruct 3D Human Pose and Shape via Model-fitting in the Loop, ICCV 2019.

VIBE* denotes the model trained with datasets identical to T-HMR.

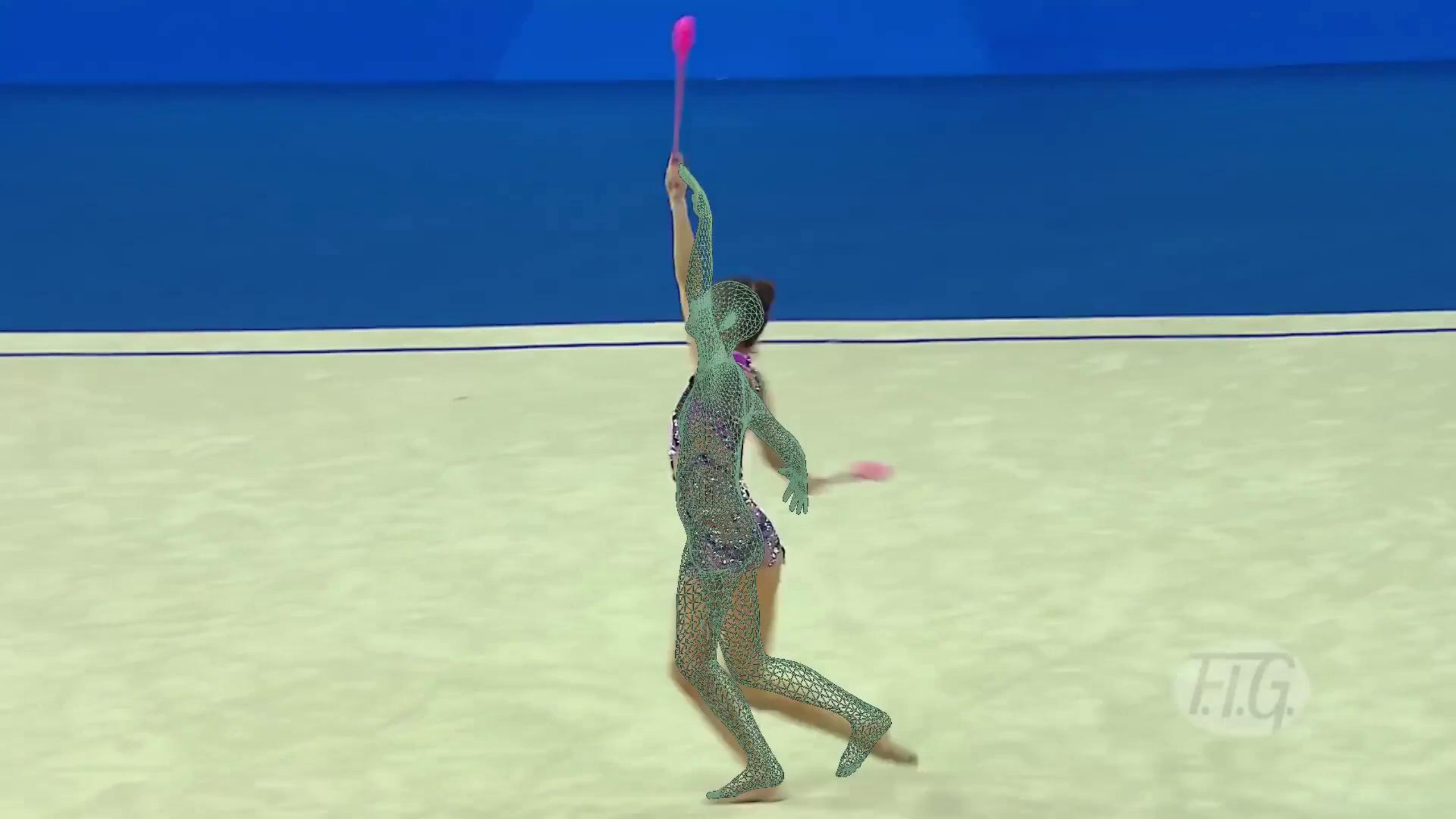
Ablation - Motion Discriminator



Results on 3DPW test set with and without motion discriminator

Qualitative Results





F.I.G.



DWIGHT K. SCHRUTE



Thanks for listening!

Resources

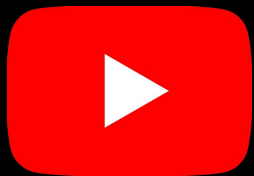
arXiv

<https://arxiv.org/abs/1912.05656>



colab

<https://github.com/mkocabas/VIBE>



<https://youtu.be/rIr-nX63dUA>