WESTLAKE UNIVERSITY

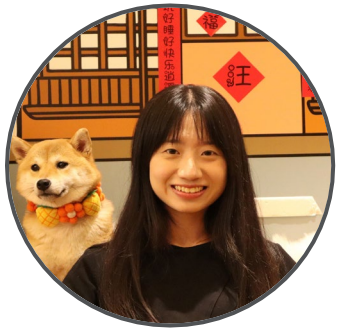
MAX PLANCK INSTITUTE FOR INFORMATICS

Tübingen AI Center
tuebingen.ai

EBERHARD KARLS
UNIVERSITÄT TÜBINGEN

Human3R

Everyone Everywhere All at Once


Yue Chen


Xingyu Chen


Yuxuan Xue


Anpei Chen


Yuliang Xiu


Gerard Pons-Moll

Human understanding requires not only human

Humans interact with the world, we need to reconstruct not only human, but also the environment.

3D human recon.
alone

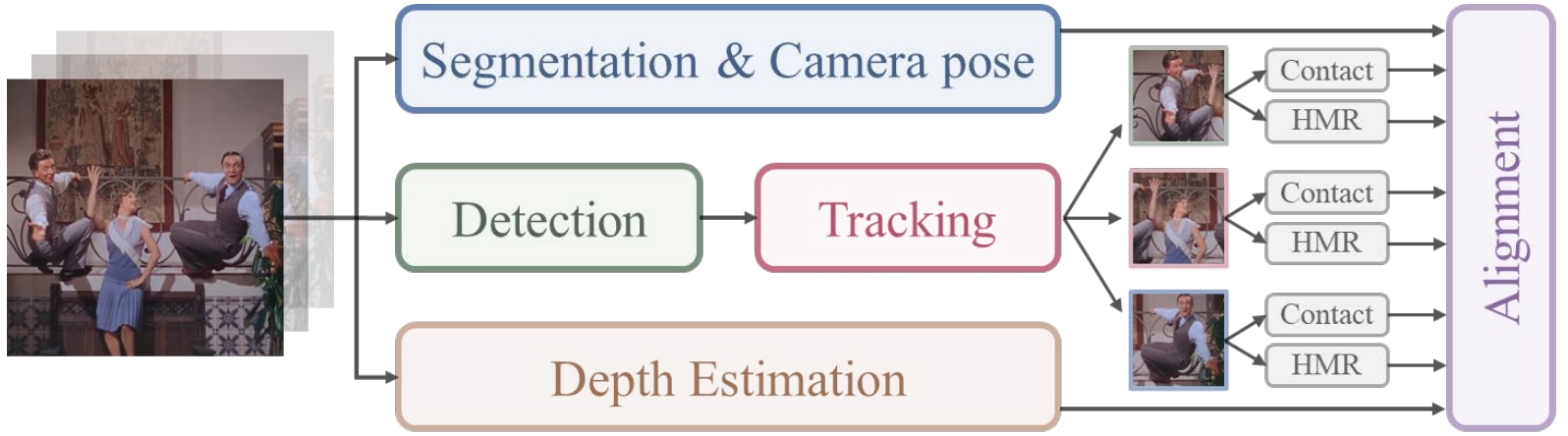
3D human recon.
w/ scene context

Raw capture

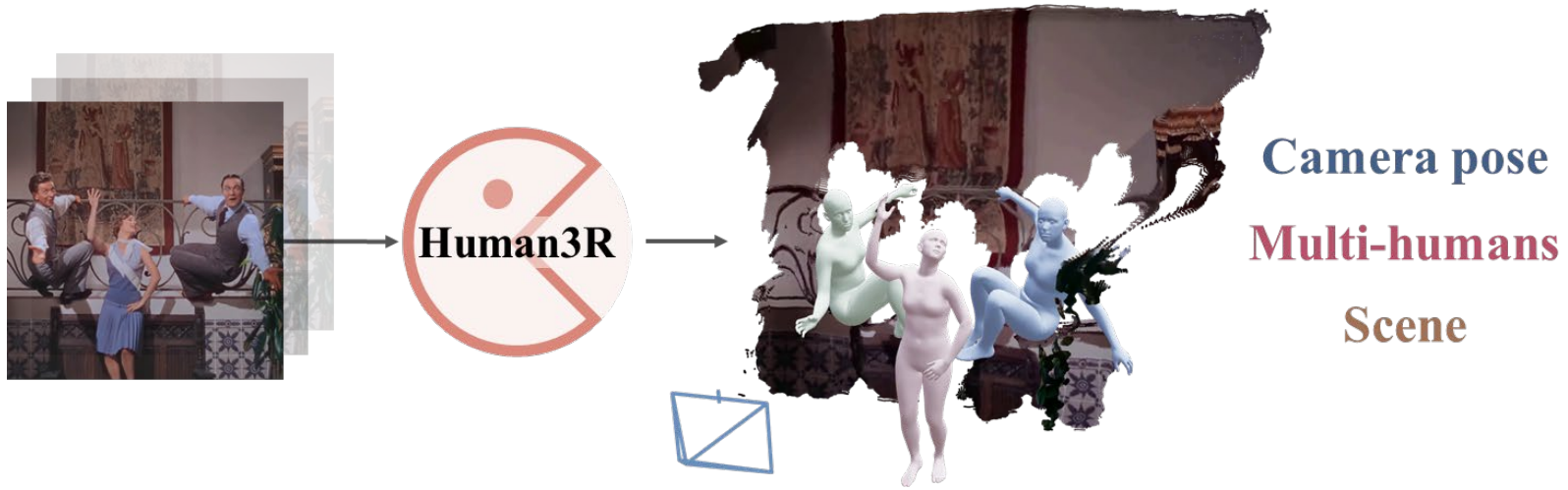
Just input a RGB video -> Global SMPL + 3D scene + Camera poses in **real time**

Global Human Motion Estimation

Previous
multi-stage methods



Our
one-stage method



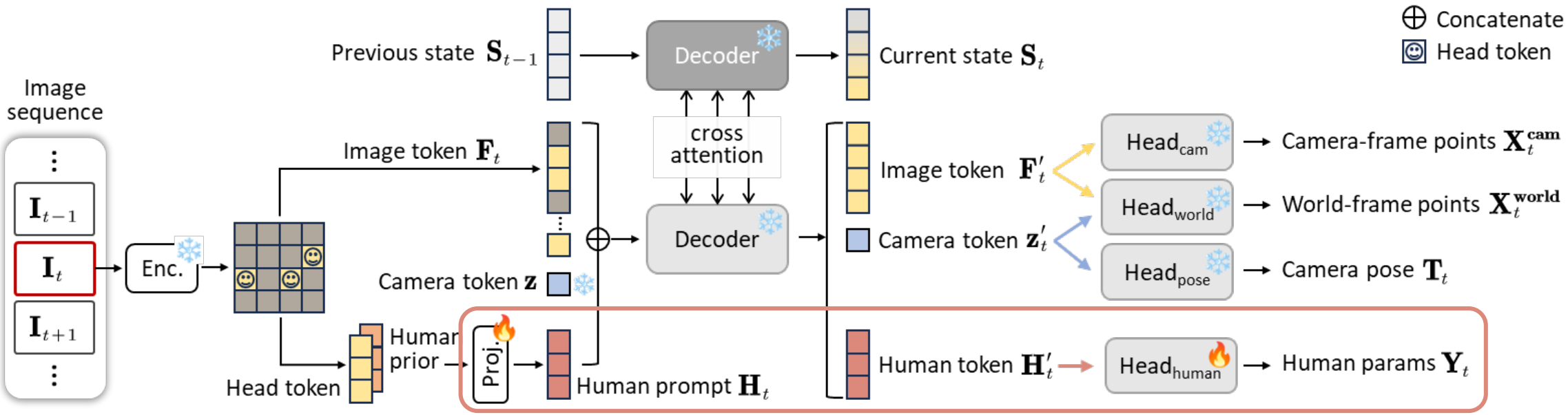
Challenges: **Data!**

Building such a unified model needs **large-scale video datasets**
with global human motion, 3D scene, and camera pose

Our solution

Reading out humans from **4D foundation model** - CUT3R with strong priors
via **minimal tuning** on small-scale human-scene dataset

Method Overview



Only human-related layers are fine-tuned, other parameters remain frozen and are initialized from CUT3R

Training this versatile model is easier than you think

One day, One GPU! (48GB) 🫠

● Prediction ● Ground-truth

● Prediction ● Ground-truth

Human3R: Everyone Everywhere All at Once

Yue Chen[1] Xingyu Chen[1*] Yuxuan Xue[2] Anpei Chen[1] Yuliang Xiu[1†] Gerard Pons-Moll[2,3]

[1]Westlake University [2]University of Tübingen, Tübingen AI Center [3]Max Planck Institute for Informatics

*Project Lead †Corresponding Author

arXiv Code Play NOW

Inference with One model, One stage; Training in One day using One GPU

https://fanegg.github.io/Human3R

Code, model and 4D interactive demo now available