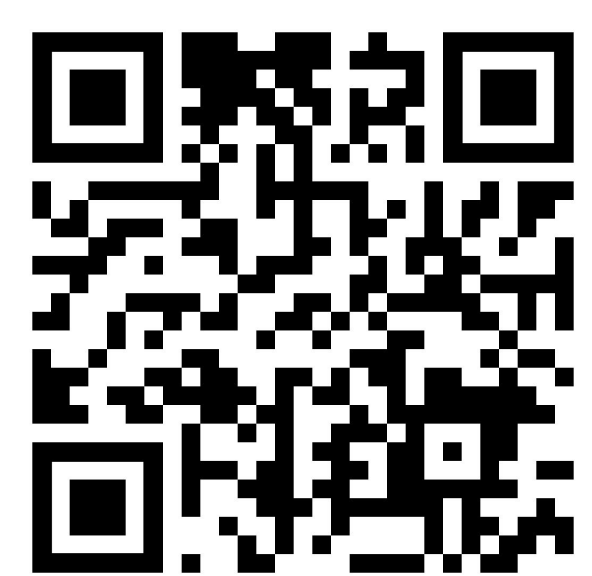




- Online demo
- Inference code
- Blogpost

Multi-HMR 2: Multi-Person Camera-Centric Human Detection, Recovery and Tracking

Guénolé Fiche Philippe Weinzaepfel Romain Brégier Fabien Baradel

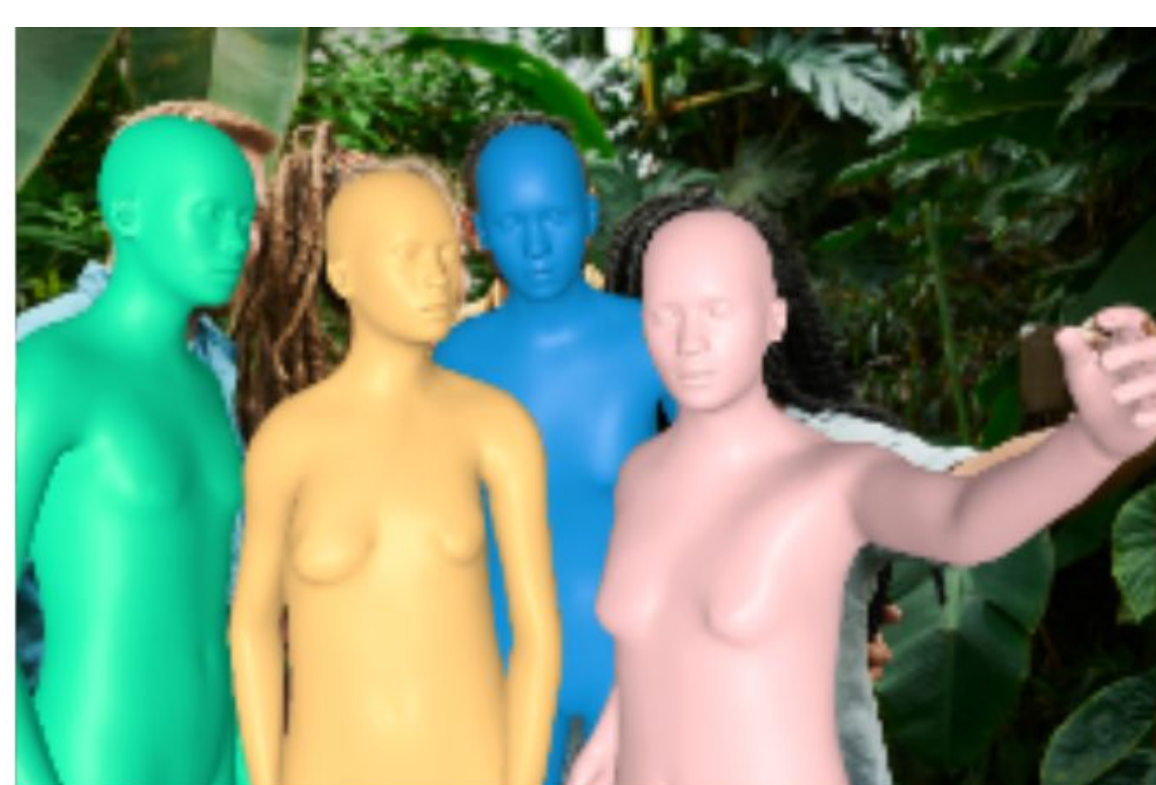
<https://github.com/naver/multi-hmr2>

Motivation

Task: Recover all human meshes given an image in a single shot



Input image



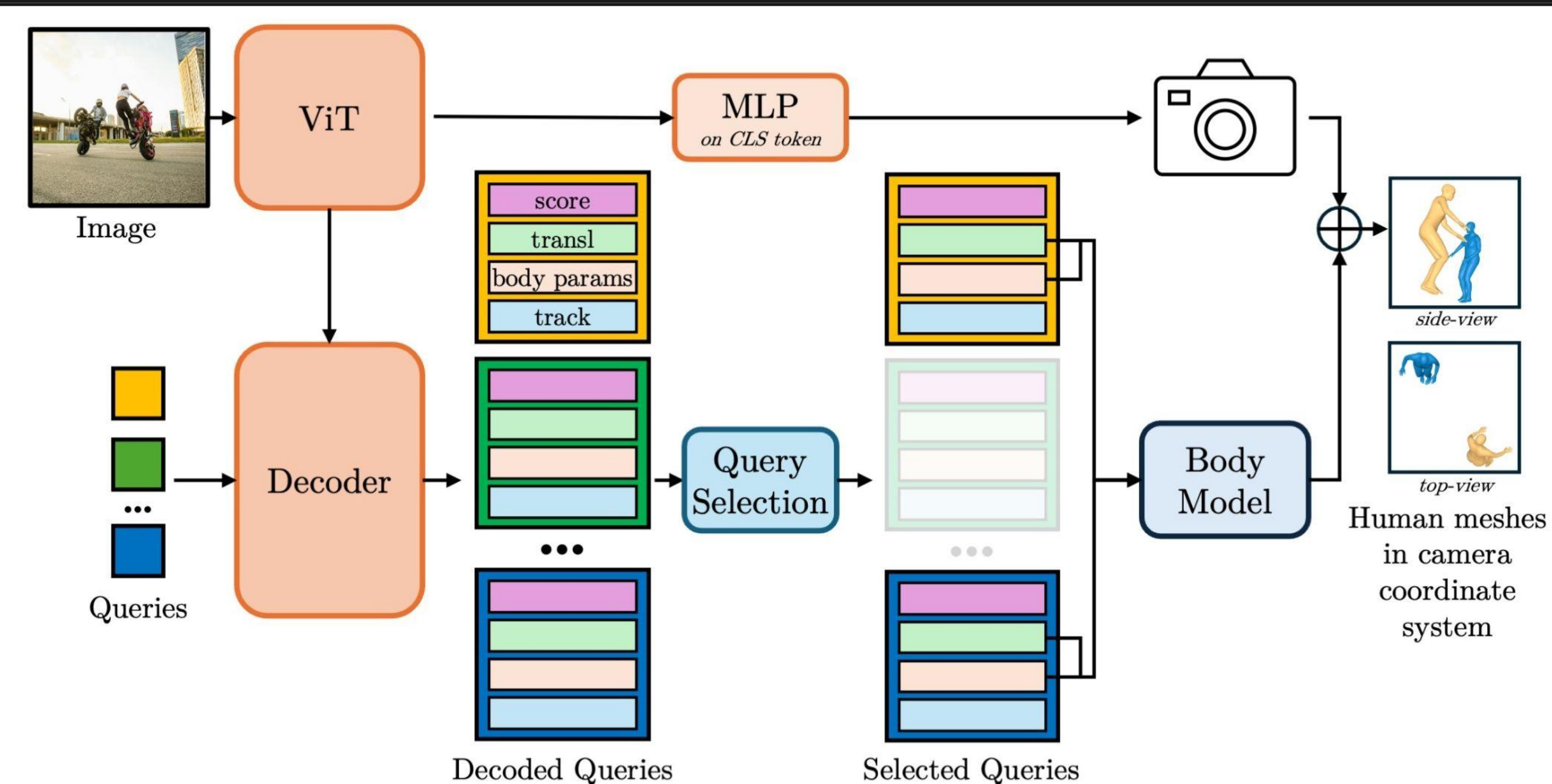
Prediction overlay



Side view

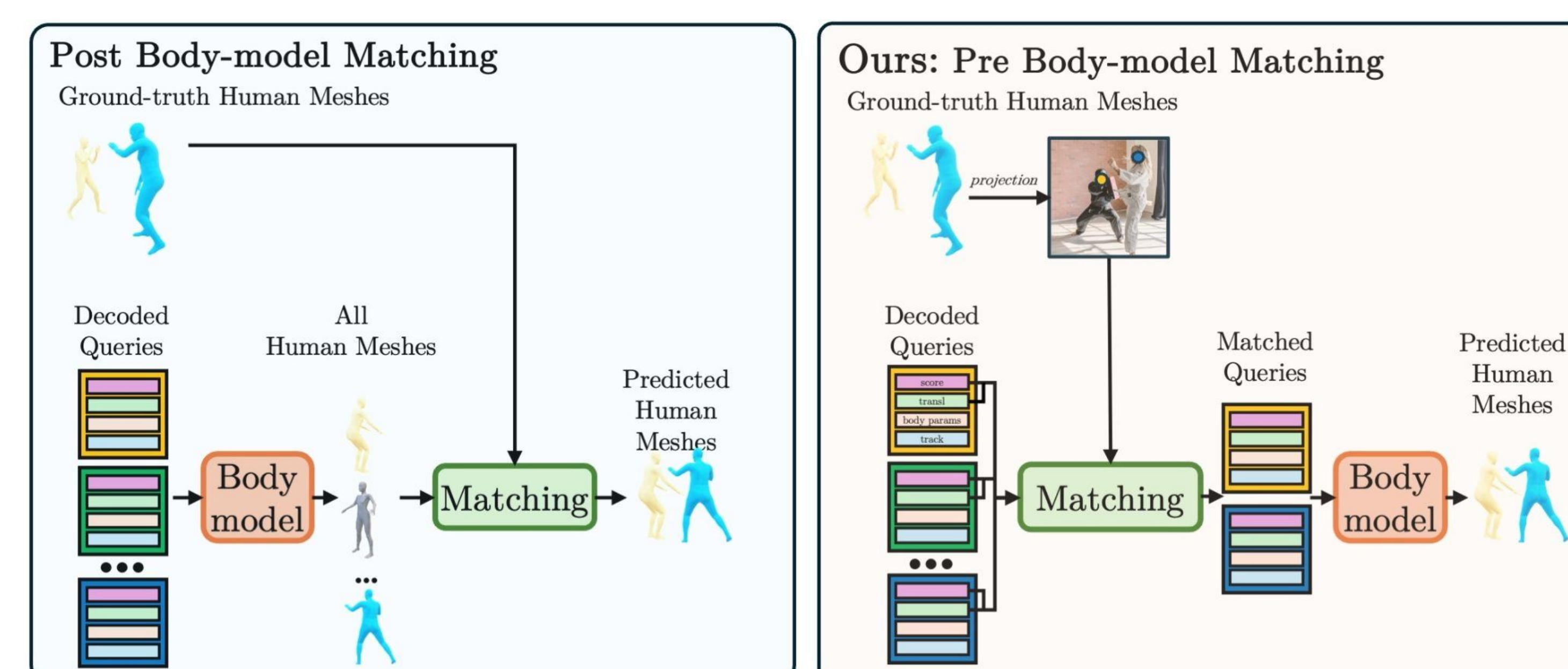
Feature	Multi-HMR (ECCV'24)	SAT-HMR (CVPR'25)	Multi-HMR 2 (ours)
Single shot	✓	✓	✓
Whole-body	●	✗	●
Child-aware	✗	✗	✓
Camera-centric	●	✗	✓
Interactions	✗	●	✓
Identity tracking	✗	✗	✓

Overall Architecture



DETR-based framework with a matching strategy during training based on 2D root location only:

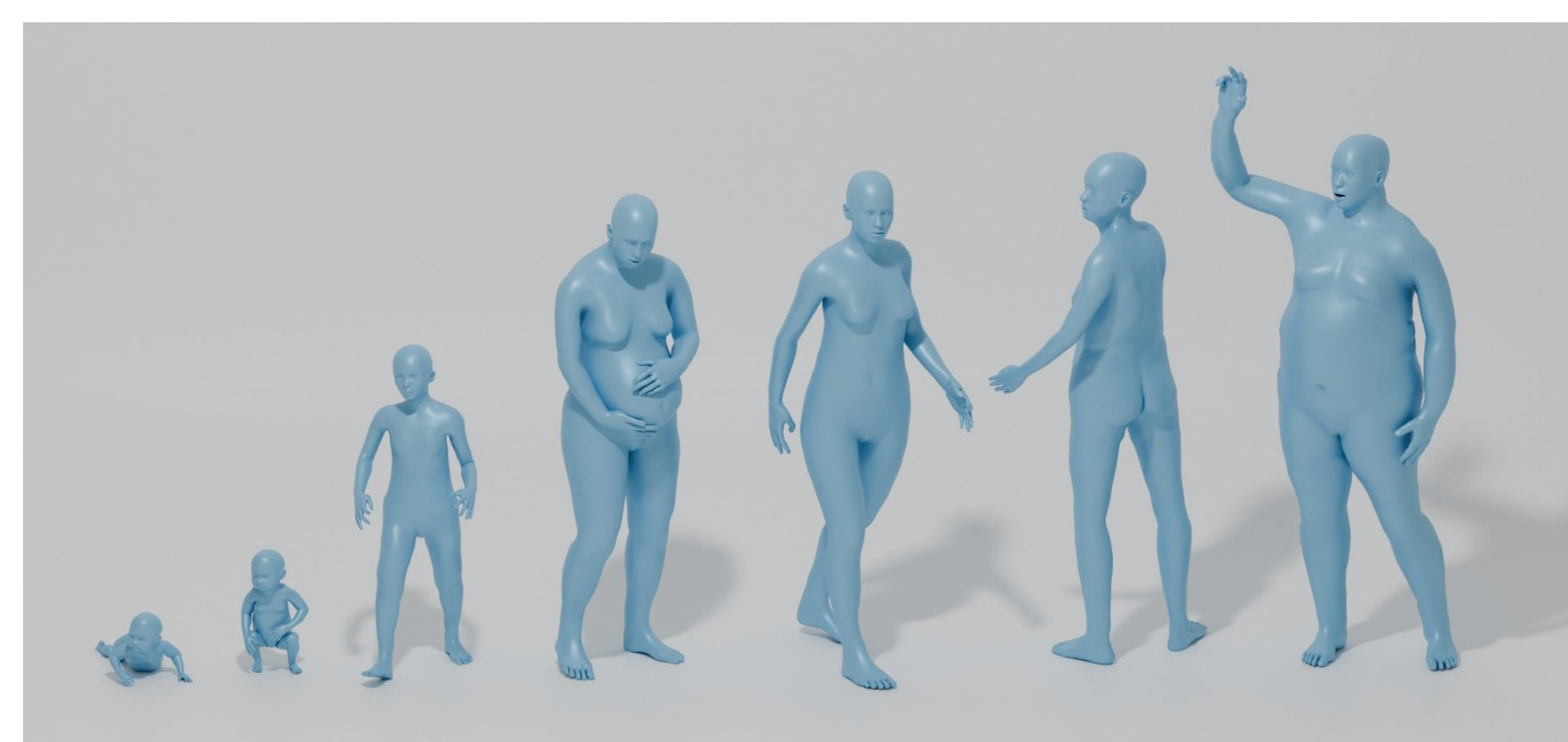
- No need to forward the body model on all queries (= fast)
- Can leverage data without 3D annotation



Anny Body Model

Multi-HMR 2 leverages Anny:

- Human diversity (age, body types and proportions)
- Interpretable body shape parameters
- Open source
- Compatible with SMPL, SOMA and others

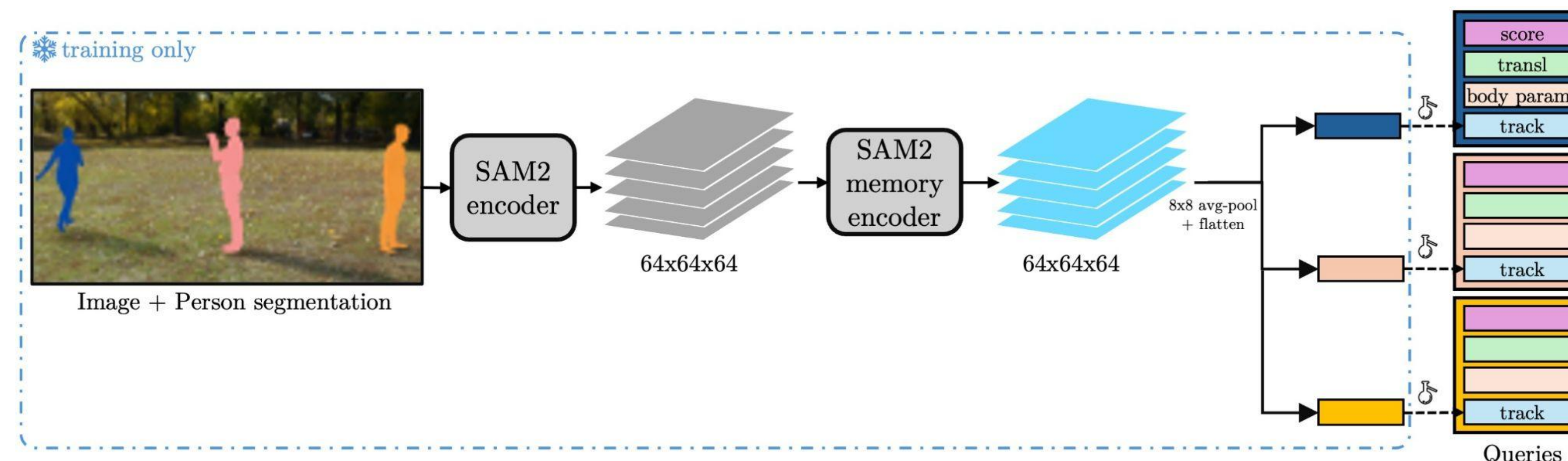


Demo and code under the Apache2.0 licence

Learning Tracking From Single Images

Prior works rely on video or mocap-based training to learn human tracking.

We propose to regress **features distilled from the SAM2 memory** encoder for each human, allowing Multi-HMR 2 to **track** humans without any training on video data.



Heterogeneous Training Data

- First training stage: **3D annotations** only (synthetic data + pseudo-gt meshes)
- Second training stage: finetune with **large-scale 2D keypoints** data (manually annotated or estimated with SOTA) + additional **consistency losses** on pose parameters and 3D joints compared to the first-stage model predictions



+20% AP and AR on MSCOCO
+15% MOTA and IDF1 on Posetrack



Preserved 3D
performance

Results



SOTA

- Detection
- Pelvis-centric 3D error
- Camera-space metrics
- Close interactions
- 2D reprojection
- Tracking

