

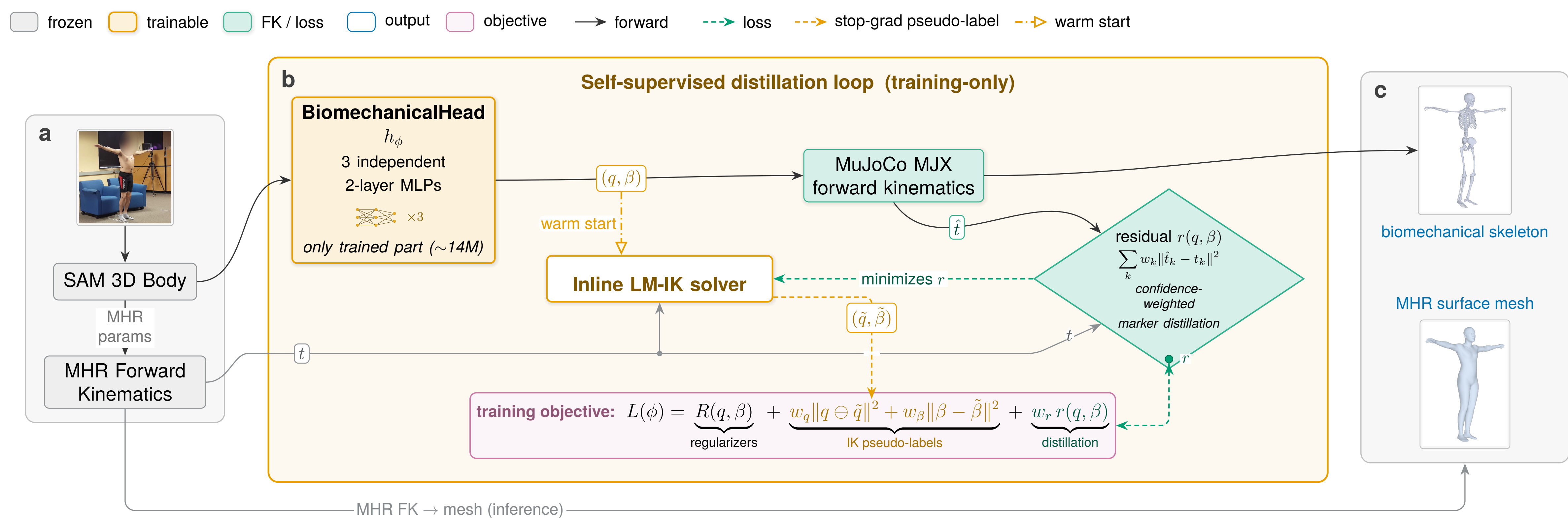
OVERVIEW

Problem Body-recovery foundation models predict on *their own* kinematic tree; clinics need ISB joint angles on an anatomical skeleton — and *no large paired image* → *biomechanics dataset* exists.

Approach A small **biomechanical head** on a *frozen* SAM-3D-Body backbone, **zero biomechanical labels**: differentiable MuJoCo FK plus a **Levenberg–Marquardt IK solver in the training loop** supply all the supervision. 1.49 M public images, **134 GPU-hours**.

Result **Competitive with the optimization-based state of the art** on three external datasets — including a 109-participant clinical cohort — at **two to three orders of magnitude lower cost**; clearly ahead of the other feed-forward method.

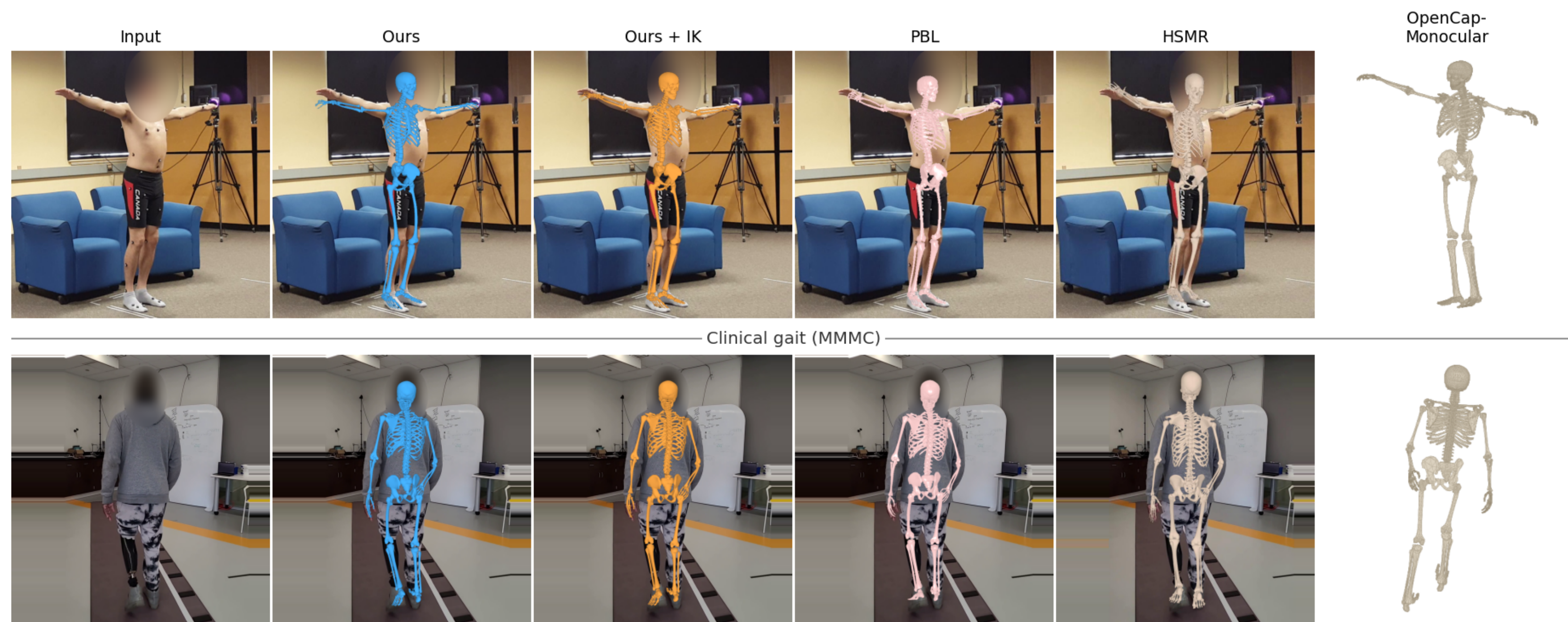
METHODS — ONE HEAD ON A FROZEN BACKBONE



$$\mathcal{L}(\phi) = \underbrace{w_r r(q, \beta)}_{\text{distillation}} + \underbrace{w_q \|q \ominus \hat{q}\|^2 + w_\beta \|\beta - \hat{\beta}\|^2}_{\text{IK pseudo-labels}} + \underbrace{\mathcal{R}(q, \beta)}_{\text{priors}} \quad r = \sum_k w_k \|\hat{t}_k - t_k\|^2$$

- Frozen SAM-3D-Body backbone** → pose token → three two-layer MLPs: **~13.6 M parameters**, the only weights that move.
- Full-body MuJoCo humanoid: **99** coordinates, **102** marker sites, **14** segment scales + offsets.
- MuJoCo MJX FK** (JAX/Equinox) poses it to sites $(J, x) = \mathcal{B}(q, \beta)$; gradients flow from image features through MuJoCo.
- No shared marker set**: each marker maps to one MHR joint center, keypoint, or vertex via an **offline EM**.
- Distillation**: backprop through MJX FK — no ground truth in the loop.
- Pseudo-labels**: staged Levenberg–Marquardt IK every step, stop-gradiented — cold-seeded, then **warm-started from the head's own prediction**.

QUALITATIVE



Held-out overlays: one BML-MoVi subject (top), one clinical MMMC subject (bottom); OpenCap ships no overlay code, so its skeleton is drawn alone.

TAKEAWAYS

- Direct inference**: distilling biomechanics onto a frozen SAM-3D-Body backbone is **state of the art for feed-forward** — with no biomechanical labels in training.
- By a wide margin**: ahead of HSMR and OpenCap-Monocular on every metric, and the gap widens on the clinical cohort — PA-MPJPE **41.9** vs 75.8 and 82.4 mm.
- A gap to optimization remains**: our Portable Biomechanics Lab still sets the bar for monocular biomechanics — at **two to three orders of magnitude** more compute per frame.

MARKER-BASED REFERENCES

Metric	Ours	Ours+IK	PBL [†]	HSMR	OpenCap [†]
<i>BML-MoVi (optical mocap, 85 subjects)</i>					
Lower-limb	3.25 (1.38)	3.15 (1.43)	2.73 (1.29)	4.76 (3.81)	3.84 (2.86)
Upper-limb	6.44 (5.57)	6.31 (5.52)	7.25 (6.41)	12.05 (9.54)	10.31 (5.97)
PA-MPJPE	37.1	36.0	33.9	49.8	47.2
<i>BioCV (marker-based MuJoCo IK, 14 subjects, 308 trials)</i>					
Lower-limb	4.04 (2.15)	3.96 (2.29)	4.10 (2.43)	6.11 (4.12)	5.39 (3.88)
Upper-limb	6.25 (6.38)	6.36 (6.05)	6.69 (6.91)	12.05 (12.57)	10.35 (9.98)
PA-MPJPE	62.2	62.7	52.5	67.4	64.9

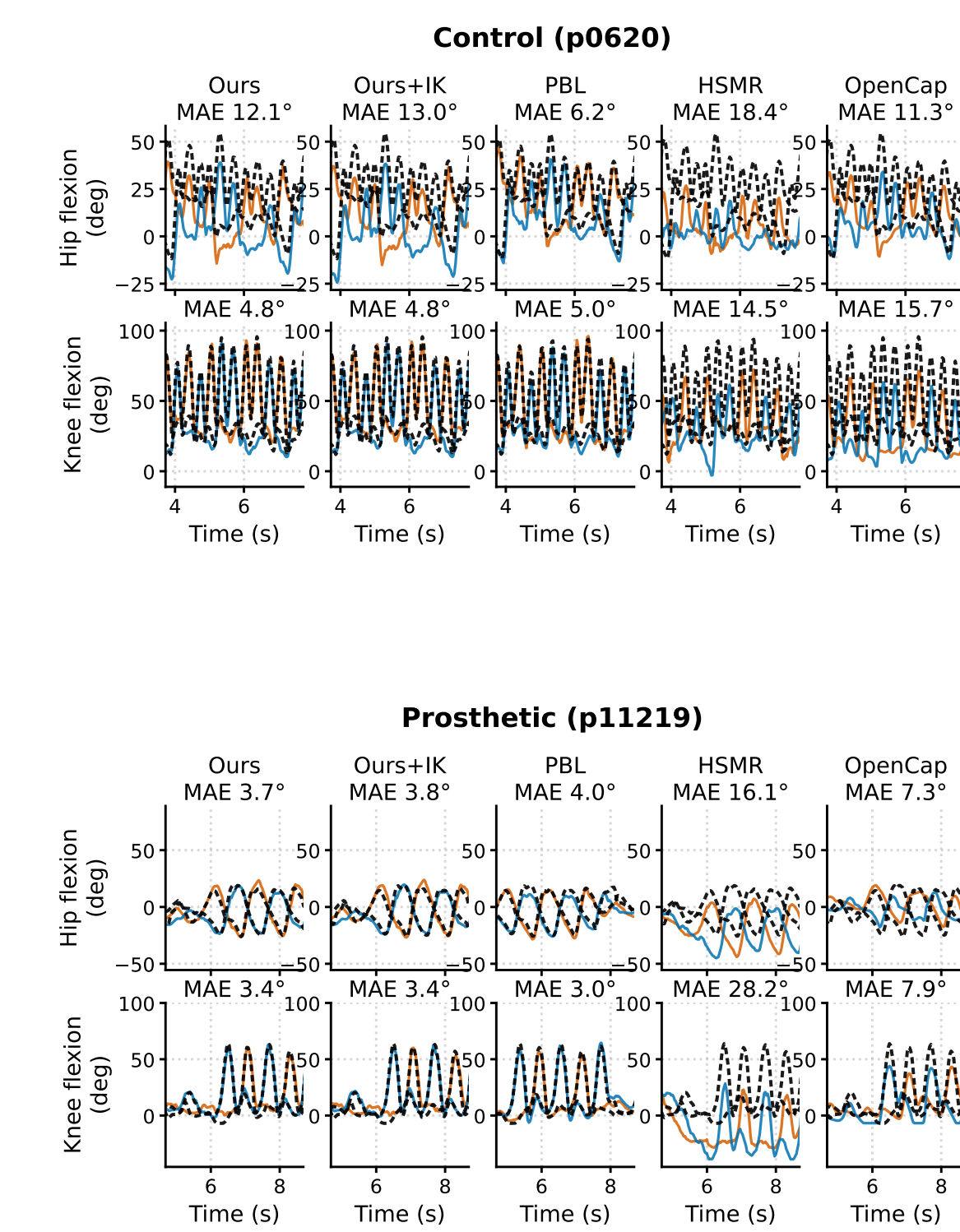
Median (nIQR) bias-corrected joint-angle error (deg) and PA-MPJPE (mm); best in **bold**, [†] = optimization. Held-out data, all methods reconciled to ISB.

CLINICAL COHORT

Metric	Ours	Ours+IK	PBL [†]	HSMR	OpenCap [†]
Lower-limb	3.66 (1.61)	3.59 (1.66)	3.24 (1.96)	6.36 (4.35)	7.16 (5.77)
Upper-limb	7.40 (6.99)	7.44 (6.99)	6.18 (5.49)	10.45 (10.25)	11.43 (17.50)
PA-MPJPE	41.9	42.0	27.6	75.8	82.4
Rejected	0	0	0	0	400

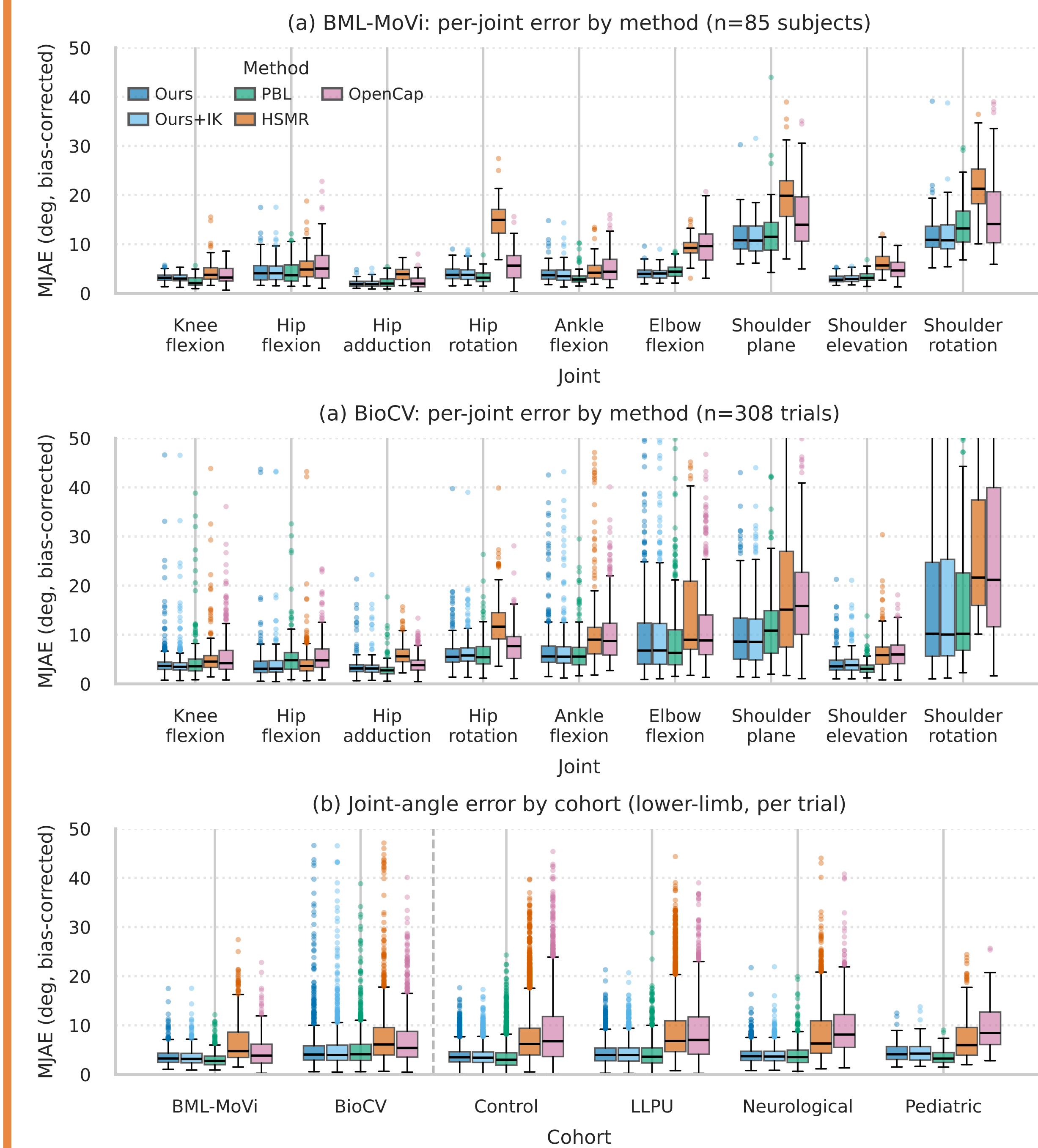
Multiview markerless reference: 109 participants, 1693 trials, four clinical populations. OpenCap's front-end gate **rejected 400 of 1693 inputs**.

GAIT WAVEFORMS



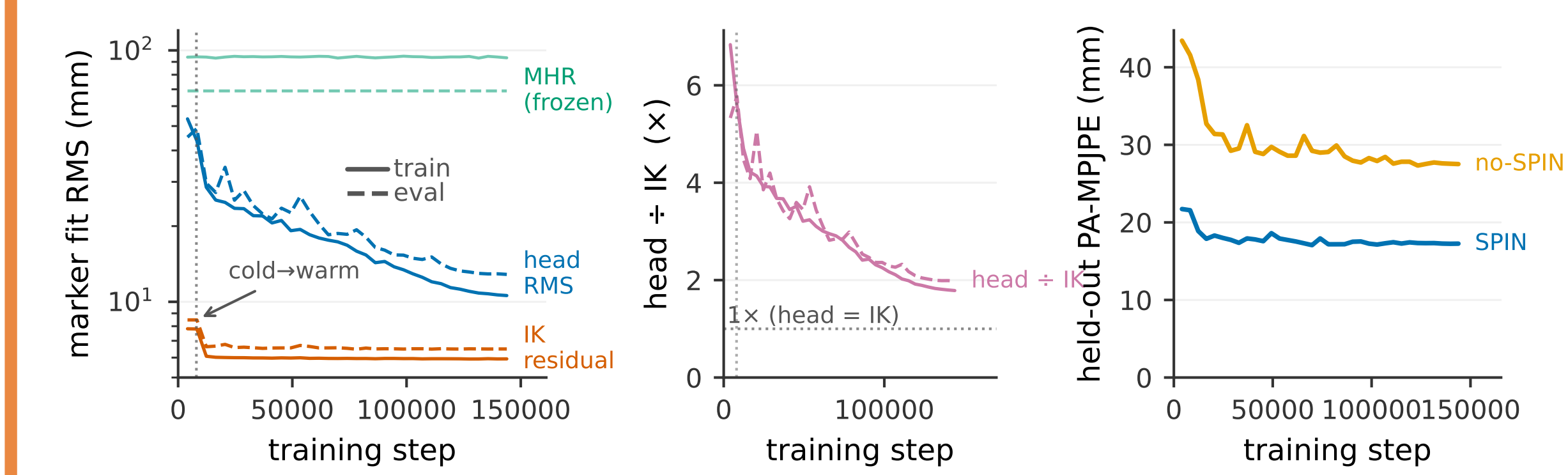
A control (top) and a prosthesis user (bottom), bias-corrected as in the tables; HSMR's knee decorrelates on the prosthetic trial.

ERROR BREAKDOWN



MJAE (a) per joint, BML-MoVi; (b) per-trial lower-limb by cohort. Ours stays **flat into the impairment cohorts**; HSMR inflates.

TRAINING DYNAMICS & ABLATION



(a) IK residual drops at the cold→warm switch; (b) head-to-1x ratio → 1x; (c) vs. SPIN-disabled.

- No in-loop pseudo-labels: marker fit **12.9** → **21.1 mm**, PA-MPJPE **17.3** → **27.5 mm**.
- Cold IK at inference: heavier tail, **180 vs 41 ms/frame**.