LDBODY: BENCHMARKING HUMAN POSE AND MESH RECOVERY WITH RESIDUAL LIMB AND PROSTHESES

Anonymous authors
Paper under double-blind review

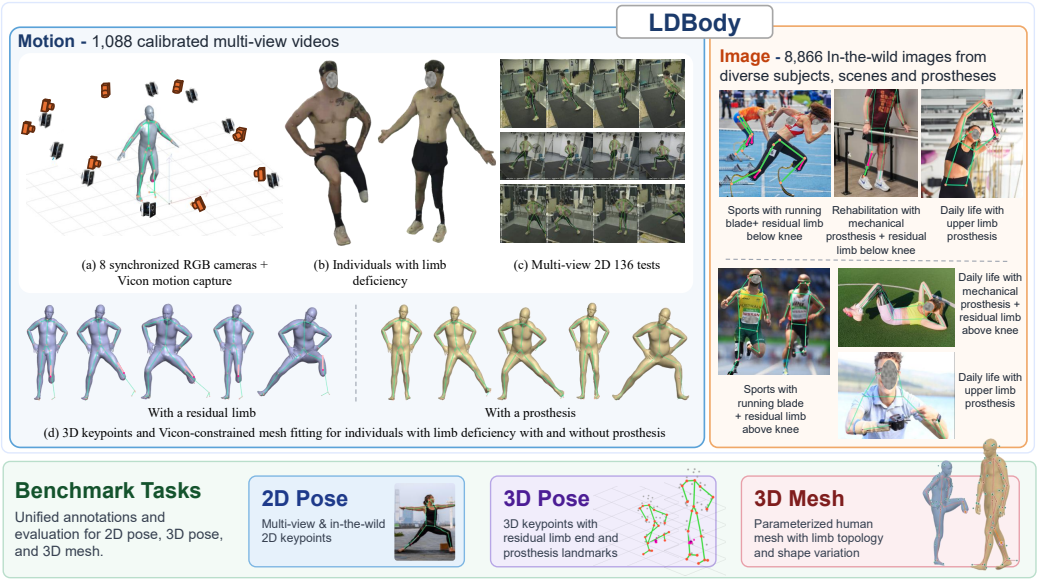

Figure 1: **Overview of LDBODY.** LDBODY combines calibrated multi-view motion capture with diverse in-the-wild images to support unified evaluation of 2D pose, 3D pose, and human mesh recovery for individuals with limb deficiency and prosthesis users.

ABSTRACT

Human pose and mesh benchmarks largely assume limb-complete bodies. This leaves residual anatomy, prosthesis-supported motion, and variation in body proportions insufficiently evaluated. We introduce LDBODY, a dataset for pose and 3D human mesh evaluation. To our knowledge, it is the first dataset to provide calibrated human body mesh evaluation with separate references for residual anatomy and prosthesis-supported function. Its annotation framework distinguishes body joints, residual-limb endpoints, and prosthesis-supported landmarks. Multiple internationally accredited Para Athletics classifiers guide the annotation protocol. The dataset contains an in-the-wild image component, LDBODY Images, and a calibrated motion component, LDBODY Motion, with shared landmark definitions. The image component covers upper- and lower-limb differences and varied body proportions, activities, scenes, clothing, and viewpoints. It adds prosthesis annotations to 8,866 LDPose images while retaining existing body-joint and residual-limb labels. The motion component captures real movements with eight synchronized RGB cameras and Vicon. It covers multiple prosthesis types and conditions with and without prostheses. The recordings provide 1,088 view-specific videos and 963,104 RGB frames. Vicon-derived keypoints and Vicon-constrained fitted body meshes support separate evaluation of residual anatomy and prosthesis-supported function. We benchmark 2D pose, multi-view 3D pose, residual-limb localization, and human mesh recovery. Additional analyses examine movement, viewpoint, and prosthesis conditions. Results on the evaluated limb-deficiency cases reveal persistent errors at residual endpoints and affected distal joints. Similar joint errors can also coexist with large differences in fitted-surface agreement. LDBODY supports research on accessible motion analysis, assistive technologies, and inclusive sport by making diverse bodies explicit evaluation targets.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

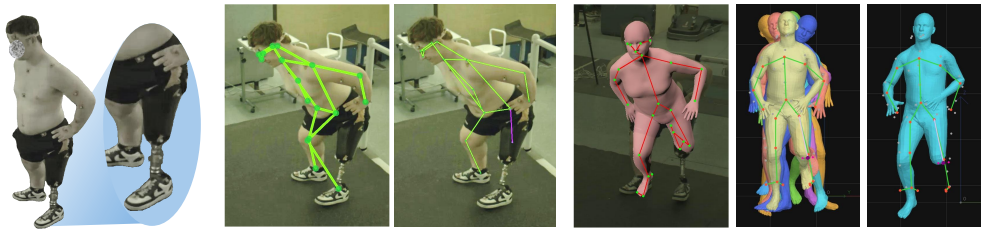

Figure 2: **Topology gap in pose and mesh estimation for prosthesis-assisted motion.** (a) A worn lower-limb prosthesis supports motion but does not follow the articulated topology of a biological limb. (b) Existing 2D pose estimators can mislocalize prosthesis-supported landmarks or residual-limb endpoints, as shown by ViTPose-l (Xu et al., 2022) and LDPose (Ying et al., 2025). (c) Current mesh and 3D reconstruction methods, including CameraHMR (Patel & Black, 2025), HSMR (Xia et al., 2025), and ResiHMR (Ying et al., 2026), can produce plausible outputs that remain inconsistent with prosthesis geometry, residual-limb structure, or Vicon-derived 3D references. These examples motivate LDBODY as a benchmark that reports standard pose, prosthesis-supported keypoints, residual-limb endpoints, and fitted mesh recovery separately.

1 INTRODUCTION

Human pose estimation and human body reconstruction have advanced with large-scale benchmarks. COCO (Lin et al., 2014) and MPII (Andriluka et al., 2014) have standardized 2D keypoint evaluation. Human3.6M (Ionescu et al., 2013), AMASS (Mahmood et al., 2019), MOYO (Tripathi et al., 2023), TotalCapture (Trumble et al., 2017), and RICH (Huang et al., 2022) have advanced 3D supervision, motion capture, and reconstruction evaluation. However, these benchmarks largely assume limb-complete bodies and use standard skeletal layouts and body models such as SMPL (Loper et al., 2023). Strong performance on these benchmarks does not necessarily imply reliable predictions across different body proportions, residual limbs, and prosthesis configurations.

These differences require more than treating missing limbs as occluded joints. As illustrated in Fig. 2, standard pose and mesh methods can hallucinate intact limbs, mislocalize residual-limb endpoints, confuse prosthetic components with anatomical limbs, or produce reconstructions that disagree with multi-view evidence. A residual-limb endpoint identifies existing body structure. A worn prosthesis can support standing, balance, or locomotion without sharing the anatomy of the missing limb. Evaluation must distinguish these targets while also accommodating variation in body proportions. Unlike existing inclusive resources centered on 2D pose, running-specific prosthesis geometry, or synthetic body meshes (Ying et al., 2025; Du et al., 2026; Wen et al., 2026; Cho et al., 2025), LDBODY supports pose and human body mesh evaluation on real calibrated recordings.

We introduce LDBODY, a dataset for pose and 3D human mesh evaluation. To our knowledge, it is the first dataset to provide calibrated human body mesh evaluation with separate references for residual anatomy and prosthesis-supported function. Its annotation framework distinguishes body joints, residual-limb endpoints, and prosthesis-supported landmarks. Multiple internationally accredited Para Athletics classifiers guide the annotation protocol. We preserve residual-limb endpoints and add standard-layout landmarks on worn prostheses when visible and functionally relevant. These landmarks are functional targets rather than anatomical joints. This design retains compatibility with common pose protocols without conflating residual anatomy and prosthesis function. For calibrated recordings, we provide Vicon-derived 2D/3D keypoints and Vicon-constrained fitted body meshes. In limb-deficiency cases, the anatomical mesh reference represents residual-limb structure. The kinematic reference retains a standard body layout for prosthesis-supported motion. They allow joint and endpoint accuracy to be evaluated separately from fitted-surface consistency.

As summarized in Fig. 1, LDBODY provides broad visual coverage and calibrated geometric references under shared landmark definitions. Its coverage includes upper- and lower-limb differences, dwarfism, and multiple prosthesis types. Diversity also spans body proportions, activities, scenes, clothing, and viewpoints rather than only disability categories. We add prosthesis-supported landmarks to 8,866 public LDPose images (Ying et al., 2025) while retaining existing body-joint and residual-limb labels. We also provide 1,088 view-specific videos containing 963,104 RGB frames.

Table 1: **Comparison of representative datasets for 3D human pose and mesh evaluation.** Capture fields describe recorded or rendered sequences. Marker MoCap denotes a marker-based capture system. People per sequence is not the total participant count. Body mesh denotes a human body reference, not a prosthesis surface or a baseline prediction. No indicates that the cited dataset does not explicitly provide the corresponding reference or focus. ^{p} denotes 2D points obtained by projection from calibrated 3D references. ^{a}The Panoptic entry lists its HD and depth cameras, not the additional VGA array. The LDBODY capture fields describe its calibrated recordings. Its in-the-wild images are included in Table 2. RSP3D provides prosthesis-surface geometry, which is distinct from the body mesh column.

Dataset	Image cameras	Marker MoCap	People per sequence	Limb-deficiency focus	Environment	2D points	3D points	Body mesh	Residual endpoints
Human3.6M (Ionescu et al., 2013)	4	Optical	1	No	Indoor	✓	✓	No	No
TotalCapture (Trumble et al., 2017)	8	Vicon	1	No	Indoor	✓	✓	No	No
MPI-INF-3DHP (Mehta et al., 2017)	14	No	1	No	Indoor and outdoor	✓	✓	No	No
Panoptic Studio (Joo et al., 2015)	31 HD + 10 RGB-D^{a}	No	Multiple	No	Indoor	✓	✓	No	No
MOYO (Tripathi et al., 2023)	8	Vicon	1	No	Indoor	✓	✓	✓	No
HuMMan (Cai et al., 2022)	10 RGB-D	No	1	No	Indoor	✓	✓	✓	No
RICH (Huang et al., 2022)	7 to 9	No	1	No	Indoor and outdoor	✓	✓	✓	No
EgoHumans (Khirodkar et al., 2023)	Variable	No	Multiple	No	Outdoor	✓	✓	✓	No
EgoBody (Zhang et al., 2022)	Variable	No	2	No	Indoor	✓	✓	✓	No
RSP3D (Wen et al., 2026)	16 RGB	OptiTrack	1	Yes	Indoor	✓^{p}	✓	No	No
BEDLAM2.0 (Tesch et al., 2025)	1 virtual	N/A	Multiple	No	Synthetic	✓	✓	✓	No
LDBODY (Ours)	**8 RGB**	**Vicon**	**1**	**Yes**	**Indoor**	✓	✓	✓	✓

Eight synchronized RGB cameras and Vicon record real movements, including trials with and without prostheses. The recordings cover static/postural, gait/locomotion, transitional/functional, and running/sport dynamic activities. This design lets the benchmark assess localization under varied visual conditions and reconstruction against calibrated geometry. For example, synchronized views can expose errors in a reconstruction that appears plausible from only one camera.

We benchmark 2D pose estimation, multi-view 3D pose estimation, residual-limb localization, and human mesh recovery. Additional analyses examine participant variation, movement categories, viewpoints, and prosthesis conditions. Among the evaluated limb-deficiency cases, residual endpoints remain difficult to localize, and affected distal joints show persistent errors. Similar joint errors can also coexist with substantially different fitted-surface agreement. These findings show why standard pose accuracy, residual-limb localization, and mesh consistency need separate evaluation rather than a single aggregate score.

LDBODY expands whose bodies human pose and mesh benchmarks are designed to evaluate. It provides a foundation for research on accessible motion analysis, assistive technologies, and inclusive sport. Its contribution is to make differences in body structure and prosthesis-supported function explicit evaluation targets.

Our contributions are as follows:

- **LDBODY.** We introduce a dataset for pose and 3D human mesh evaluation with diverse in-the-wild images and real calibrated recordings. It covers limb differences, dwarfism, and multiple prosthesis types.
- **Expert-guided annotations.** We distinguish body joints, residual-limb endpoints, and prosthesis-supported landmarks under a protocol guided by multiple internationally accredited Para Athletics classifiers. Vicon-constrained fitted body meshes extend this distinction to residual anatomy and prosthesis-supported function in calibrated recordings.
- **Benchmark protocols and diagnostic findings.** We evaluate 2D pose, multi-view 3D pose, residual-limb localization, and human mesh recovery. The analyses identify target-specific localization errors and distinguish joint accuracy from fitted-surface consistency.

Project page. The anonymous project page is available at Anonymous project page.

2 RELATED WORK

Human pose and mesh datasets. Human3.6M (Ionescu et al., 2013), Panoptic Studio (Joo et al., 2015), and TotalCapture (Trumble et al., 2017) provide multi-view observations and 3D pose references. MOYO (Tripathi et al., 2023), RICH (Huang et al., 2022), HuMMan (Cai et al., 2022), and M3GYM (Xu et al., 2025) extend human capture and reconstruction evaluation. AMASS (Mahmood et al., 2019) unifies motion capture through body-model fitting, while BEDLAM2.0 (Tesch et al.,

Table 2: **Comparison of datasets for limb differences and prosthesis use.** Image and video-frame counts remain separate. Prosthesis labels describe annotation content rather than the presence of prosthesis users. Body mesh denotes a human body reference rather than device geometry. No indicates that the corresponding annotation is not explicitly provided. ^{p} denotes 2D points obtained by projection from calibrated 3D references. RSP denotes a running-specific prosthesis. ProGait provides paired frontal and sagittal views but does not report calibrated 3D references. A3D provides synthetic body-model parameters and 2D/3D joints. The LDBODY image count refers to LDPose images with added prosthesis annotations. Its video count is view-specific. Calibrated 3D and fitted mesh references apply to its motion recordings.

Dataset	Data	Population	Images or video frames	Videos	Prosthesis labels	Views	2D	3D	Body mesh	Residual endpoints
LDPose (Ying et al., 2025)	Images	Limb differences	28k images	N/A	No	Single	✓	No	No	✓
InclusiveVidPose (Du et al., 2026)	Videos	Limb differences	327k frames	313	Status	Single	✓	No	No	✓
ProGait (Yin et al., 2025)	Videos	Transfemoral prosthesis users	Not reported	412	2D landmarks	2 paired views	✓	No	No	No
A3D (Cho et al., 2025)	Synthetic images	Simulated limb loss	1.0M images	N/A	No	N/A	✓	✓	✓	No
RSP3D (Wen et al., 2026)	RGB and MoCap	RSP users	5.6M frames	Not reported	3D device surface	16 calibrated RGB views	✓^{p}	✓	No	No
LDBODY (Ours)	Images, RGB and MoCap	Limb differences and dwarfism	8,866 images / 963,104 frames	**1,088**	Functional landmarks	Single images / 8 calibrated views	✓	✓	✓	✓

2025) provides synthetic human data. These resources mainly use standard skeletal layouts and body models. Residual endpoints and prosthesis-supported function are not their central evaluation targets. LDBODY extends pose and human mesh evaluation to these targets using real calibrated recordings and diverse in-the-wild images. Table 1 compares capture settings and reference annotations.

Limb-deficiency and prosthesis-user resources. LDPose (Ying et al., 2025) provides residual-limb endpoint annotations for image pose estimation. InclusiveVidPose (Du et al., 2026) adds video pose, tracking, and prosthesis-status annotations. ProGait (Yin et al., 2025) provides paired walking views with segmentation, pose, and gait annotations for transfemoral prosthesis users. Biomechanics resources provide measurements for gait analysis and prosthesis evaluation (Hunt et al., 2025; Aghababa & Andrysek, 2024). Concurrent work RSP3D (Wen et al., 2026) focuses on running-specific prosthesis users and provides multi-view recordings with 3D body-joint and prosthesis-surface references. Its surface reference describes the prosthesis rather than a fitted human body mesh. LDBODY covers multiple prosthesis types and broader body variation, including dwarfism. Its annotations distinguish body joints, residual endpoints, and prosthesis-supported landmarks, and provide calibrated fitted human body meshes. Table 2 summarizes these differences.

Pose and mesh recovery methods. Most pose and human mesh recovery methods use standard keypoint layouts or complete body templates. AJAHR (Cho et al., 2025) and ResiHMR (Ying et al., 2026) address limb loss in body reconstruction. The A3D data introduced with AJAHR provide synthetic images with 2D/3D joints and body-model parameters. Our focus is real calibrated evaluation with distinct references for residual anatomy and prosthesis-supported function. We evaluate representative pose and mesh methods in Sec. 4 and provide extended comparisons in Appendix B.

3 LDBODY DATASET

LDBODY uses shared landmark definitions across in-the-wild images and calibrated motion recordings.

3.1 DATA COLLECTION

Coverage and image selection. LDBODY covers upper- and lower-limb differences, dwarfism, and multiple prosthesis types. The motion recordings include people with limb deficiency and people with dwarfism. The in-the-wild images provide broader variation in body proportions, activities, scenes, clothing, and viewpoints. For LDBODY Images, we add prosthesis-supported landmarks to 8,866 public LDPose images (Ying et al., 2025) with visible prostheses while retaining their existing body-joint and residual-limb annotations. The image records and motion recordings follow the same landmark definitions. Image selection, annotation examples, split files, and source licensing are documented in Appendix C.3.

Capture setup. We prospectively collected LDBODY Motion in an indoor motion-capture space using eight synchronized RGB cameras and a Vicon optical motion-capture system. The cameras surround the capture volume, as shown in Fig. 1(a). RGB videos have a resolution of 2048×1080

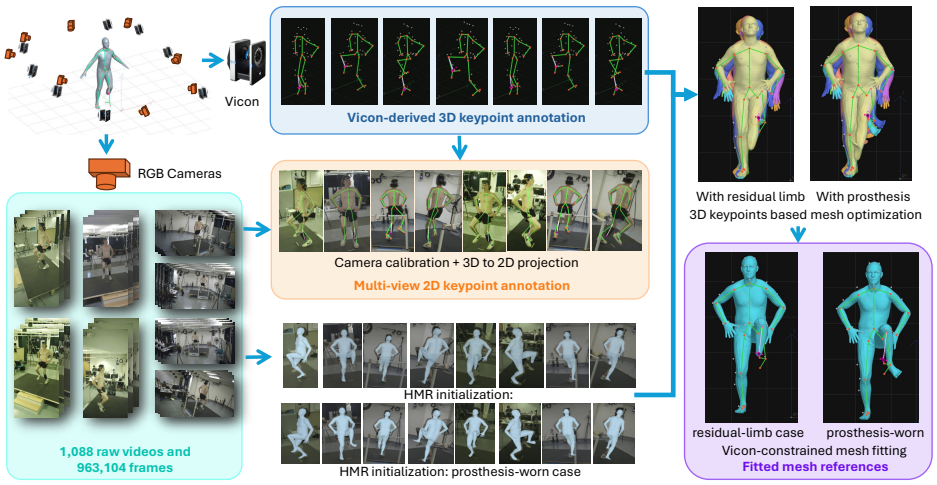

Figure 3: **Keypoint and fitted mesh reference generation for the calibrated recordings in LD-BODY.** Vicon observations provide 3D keypoints and calibrated 2D projections. HMR initialization supplies the common base reconstruction. ResiHMR (Ying et al., 2026) converts it to residual-limb topology for the anatomical branch. Camera-space hypotheses are aligned in Vicon coordinates and refined for their respective targets. Only the anatomical branch uses a residual-endpoint constraint. Grey, orange, and pink dots denote Vicon markers, standard-layout keypoints, and residual-endpoint references. Green lines denote skeletal connections.

and a frame rate of 50 fps. Vicon marker trajectories are recorded at 100 Hz. We use the full-body Plug-in Gait marker set for body motion. For participants with terminal limb deficiency, four additional markers around the residual-limb region provide the source for residual-endpoint references. Appendix C.2 describes camera layout, synchronization, and marker placement.

Capture conditions and motion protocol. The capture space contains visible equipment and ordinary indoor lighting rather than uniform background panels. A support handle is available during standing, balance, walking, and transition tasks. We retain frames with handle-related occlusion. Each synchronized sequence contains one participant performing structured motion and coordination tests. The protocol covers static postures, balance, locomotion, functional transitions, and sport-related movements. Prosthesis users perform tests with and without their devices where applicable. The recorded devices include blade-shaped prostheses, exposed-pylon prostheses, and prostheses with cosmetic covers.

3.2 Keypoint Definitions and Annotation

Shared landmark definitions. Multiple internationally accredited Para Athletics classifiers guided the annotation protocol. We distinguish body joints, residual-limb endpoints, and prosthesis-supported landmarks. For terminal limb deficiency without a prosthesis, we retain the residual endpoint and leave unavailable distal body keypoints unlabeled. When a prosthesis is worn, we preserve the residual endpoint and add applicable standard-layout landmarks on the device. These landmarks represent functional motion targets rather than anatomical joints. For a visible mechanical knee or ankle, placement follows the joint centre or apparent rotation axis. For a running blade without a discrete ankle joint, the protocol defines a virtual functional ankle using the device centreline and the corresponding anatomical level. It is not the blade tip or ground-contact point. Appendix C.3 gives device-specific placement and ambiguity rules.

Image and Vicon annotation. Three annotators add the image landmarks under the written protocol. For calibrated recordings, Plug-in Gait joint centres and marker combinations provide body keypoints. Additional residual markers provide a marker-derived endpoint reference. We compute this reference as the mean of available valid residual markers and assign its category from the participant's deficiency level using the LDPose convention (Ying et al., 2025). A frame without a valid endpoint reference

is excluded from residual-endpoint supervision. All 3D points are represented in the Vicon world coordinate system and projected into calibrated RGB views, as shown in Fig. 3. Appendices C.4 and C.5 specify marker validity, coordinate transforms, and projection rules.

Compatible keypoint formats. We provide an OpenPose-style format with 25 body keypoints (Cao et al., 2019) and a COCO-style format with 17 body keypoints (Lin et al., 2014). Each is augmented with the same eight residual-limb slots defined by LDPose (Ying et al., 2025). The resulting layouts contain 33 and 25 keypoint slots, respectively. The OpenPose-style layout is used for mesh recovery evaluation. The LDPose-style layout is used for 2D and 3D pose evaluation. Unavailable reference landmarks are masked during evaluation.

3.3 Vicon-constrained Fitted Mesh References

We construct Vicon-constrained fitted mesh references for human mesh recovery evaluation. These references convert synchronized motion-capture evidence into fitted parametric body representations. The construction is not restricted to a particular HMR initializer: any HMR method providing a compatible limb-complete reconstruction can supply the base estimate. We use HSMR (Xia et al., 2025) for the default references and evaluate PromptHMR (Wang et al., 2025) as an alternative initializer.

Anatomical and kinematic reference branches. For each base initializer, we derive two reference representations from the same initial reconstruction. The *kinematic branch* retains the standard full-body layout, using prosthesis-supported landmarks to represent the functional motion chain when a prosthesis is worn. This representation supports whole-body motion evaluation rather than reconstruction of the prosthesis surface. The *anatomical branch* applies ResiHMR (Ying et al., 2026) residual anchor–factor optimization and residual-limb reconstruction to the base estimate. It represents the remaining body and residual limb, excluding the missing distal segment and prosthetic components. The conversion retains the initial camera and body-shape parameters while modifying the affected limb's geometry and kinematic chain.

Vicon-constrained refinement. The branch-specific hypotheses are transformed into the Vicon world coordinate system and refined separately against frame-matched Vicon-derived constraints. Both objectives include joint alignment and regularization of pose, shape, and translation around their initial estimates. The anatomical branch additionally uses valid residual-endpoint constraints, whereas the kinematic branch uses standard-layout functional landmarks, including those on prosthetic components where available. Candidate selection, optimization objectives, and implementation details are provided in Appendix D.

Reference validation and evaluation scope. We assess the fitted references through Vicon fitting residuals, shared-region discrepancies between branches, and sensitivity to alternative base initializations. These diagnostics quantify constraint agreement and sensitivity to the reference construction. Vicon-derived joint and residual-endpoint metrics provide the primary geometric evaluation, while *reference-relative mean vertex error (MVE)* measures agreement with the corresponding fitted parametric surface. We report general full-body and residual-topology results under their corresponding evaluation protocols. Detailed reference diagnostics are provided in Appendix D.

3.4 Data Statistics, Quality Control, and Release

Scale and coverage. The motion recordings comprise 136 synchronized tests from five adult male participants aged 20–65 years with lower-limb terminal deficiency. The cohort includes three participants with transtibial deficiency and two with transfemoral deficiency. Four participants were active Para athletes at the time of recording. Each test contributes eight synchronized camera videos of the same motion, yielding 1,088 view-specific videos and 963,104 RGB frames across views. The recordings include 120 prosthesis-worn and 16 non-worn tests, covering four movement categories: static/postural, gait/locomotion, transitional/functional, and running/sport dynamic. The 8,866 in-the-wild images with additional prosthesis annotations are counted separately from the motion-video frames. Detailed participant characteristics, prosthesis configurations, and recording statistics are provided in Appendix C.1.

Table 3: **2D keypoint estimation within LDBODY.** Each paired entry reports LDBODY Images / LDBODY Motion. Fine-tuning uses the training images in LDBODY Images. AP and AR follow the COCO-style evaluation over the LDPose 25-keypoint schema and are reported on a 0 to 100 scale. The settings differ in visual conditions and reference availability. Their scores do not directly measure relative subset difficulty. Bold denotes the highest reported score for each metric and setting.

Method	Fine-tuning	AP ↑	AP^{50} ↑	AP^{75} ↑	AR ↑	AR^{50} ↑	AR^{75} ↑
YOLOPose-l (Maji et al., 2022)	No	58.4 / 45.7	84.4 / 73.5	67.7 / 50.8	64.2 / 51.8	87.9 / 82.9	72.5 / 56.8
	Yes	62.8 / 32.7	82.3 / 46.2	68.6 / 35.6	65.8 / 35.4	84.8 / 49.2	71.2 / 38.0
RTMPose-l (Jiang et al., 2023)	No	63.3 / 69.1	87.9 / 97.8	74.1 / 82.6	67.3 / 72.8	88.3 / 98.6	76.4 / 86.2
	Yes	77.4 / 72.6	89.9 / 96.9	84.4 / 81.3	79.7 / 77.6	90.3 / 97.6	85.0 / 84.2
ViTPose-l (Xu et al., 2022)	No	65.0 / 71.8	88.9 / 98.8	76.0 / 87.1	69.2 / 75.5	89.0 / 99.0	78.7 / 90.2
	Yes	**80.5 / 81.9**	**90.8 / 99.0**	**86.4 / 91.4**	**82.9 / 85.9**	**91.0 / 99.4**	**88.0 / 93.4**

Quality control. Image annotators complete reviewed practice annotations before production. We randomly sample 10% of completed image annotations for manual quality review. For motion records, we inspect synchronization, camera calibration, projected keypoints, and mesh-fitting diagnostics. Stored fitting metadata make the reference-generation process auditable. Detailed participant attributes, marker mappings, and annotation rules are provided in Appendix C.

Ethics and release. Prospective motion capture was conducted under institutional human research ethics approval, with informed consent obtained from all participants prior to data collection. LDBODY will be released as a versioned research resource. Potentially identifiable assets, including raw RGB videos, Vicon recordings, and fitted mesh references, will be provided to approved non-commercial researchers under a data-use agreement and access-review process. Derived annotations and metadata will be released under the Creative Commons BY-NC-SA 4.0 license where permitted by participant consent. Source LDPose images and inherited annotations remain subject to their original access and licensing conditions.

4 BENCHMARKS AND BASELINE EVALUATION

We evaluate 2D pose estimation, multi-view 3D pose estimation, residual-limb localization, and human mesh recovery on LDBODY. These evaluations distinguish standard-layout landmarks, residual endpoints, and fitted body surfaces. We use YOLOPose-l (Maji et al., 2022), RTMPose-l (Jiang et al., 2023), and ViTPose-l (Xu et al., 2022) for 2D prediction and calibrated multi-view triangulation. We evaluate TokenHMR (Dwivedi et al., 2024), CameraHMR (Patel & Black, 2025), PromptHMR (Wang et al., 2025), and ResiHMR (Ying et al., 2026) for human mesh recovery. The pose models are evaluated before and after fine-tuning on the training images in LDBODY Images. The calibrated motion recordings are not used for fine-tuning. Appendix E describes the splits, model configurations, keypoint mappings, valid-point handling, camera processing, and evaluation metrics.

2D keypoint estimation. Table 3 reports COCO-style average precision (AP) and average recall (AR) using the LDPose 25-keypoint schema. We evaluate in-the-wild images and calibrated video frames separately within LDBODY. Fine-tuning improves image AP for all three models. RTMPose-l increases from 63.3 to 77.4 AP, and ViTPose-l increases from 65.0 to 80.5 AP. On the calibrated frames, their AP increases from 69.1 to 72.6 and from 71.8 to 81.9, respectively. YOLOPose-l improves from 58.4 to 62.8 AP on the images but falls from 45.7 to 32.7 AP on the calibrated frames. Image supervision therefore yields model-dependent changes across visual conditions. The overall AP changes motivate the target-specific diagnostics in Appendix F.

Multi-view 3D keypoint estimation. At each synchronized time step, we predict keypoints independently in the calibrated RGB views and triangulate them using the camera geometry. Table 4 reports mean per-joint position error (MPJPE), its Procrustes-aligned form (PA-MPJPE), and residual-endpoint error in millimeters. Without fine-tuning, RTMPose-l and ViTPose-l obtain standard-keypoint MPJPE of 47.89 and 46.07 mm. Fine-tuned models additionally predict residual endpoints and are evaluated over an extended keypoint set. Their aggregate errors therefore cannot be interpreted as a matched comparison over standard joints alone. The separate residual-endpoint errors are 110.0 mm for RTMPose-l, 90.12 mm for ViTPose-l, and 375.94 mm for YOLOPose-l.

Table 4: **Multi-view 3D keypoint estimation on the evaluated LDBODY Motion recordings.** All errors are in millimeters. Fine-tuning uses LDBODY Images. Standard denotes standard-layout keypoints. Extended also includes residual endpoints. The aggregate MPJPE and PA-MPJPE columns therefore use different keypoint coverage before and after fine-tuning. N/A denotes unavailable endpoint predictions, not zero error. Endpoint results must be interpreted with reconstruction coverage.

Method	Fine-tuning	Keypoint coverage	MPJPE ↓	PA-MPJPE ↓	MPJPE-ResidualEnd ↓
YOLOPose-l (Maji et al., 2022)	No	Standard	209.28	96.96	N/A
	Yes	Extended	594.64	130.72	375.94
RTMPose-l (Jiang et al., 2023)	No	Standard	47.89	45.10	N/A
	Yes	Extended	58.83	53.35	110.0
ViTPose-l (Xu et al., 2022)	No	Standard	46.07	43.56	N/A
	Yes	Extended	50.25	47.61	90.12

Table 5: **Human mesh recovery on the evaluated LDBODY Motion recordings.** All errors are in millimeters. Kinematic rows use the standard full-body protocol. †The anatomical row uses present joints and residual-topology vertices. It is not ranked against the kinematic rows. MVE is reference-relative fitted-surface consistency, not scan-based surface accuracy. ResiHMR also participates in anatomical-reference construction. Bold compares only the three kinematic rows. N/A denotes no residual-endpoint prediction.

Method	Fitted reference	MPJPE ↓	PA-MPJPE ↓	MPJPE-ResidualEnd ↓	MVE ↓
TokenHMR (Dwivedi et al., 2024)	Kinematic	148.81	87.30	N/A	177.81
CameraHMR (Patel & Black, 2025)	Kinematic	229.97	120.71	N/A	198.53
PromptHMR (Wang et al., 2025)	Kinematic	**148.09**	**87.25**	N/A	**107.44**
ResiHMR† (Ying et al., 2026)	Anatomical	135.32	76.14	154.22	101.47

Models without endpoint predictions are marked N/A rather than assigned zero error. These results motivate reporting residual localization separately from standard pose, with reconstruction coverage considered alongside endpoint error.

3D human mesh recovery. Table 5 separates two evaluation protocols. TokenHMR, CameraHMR, and PromptHMR use the kinematic fitted reference with a standard full-body layout. ResiHMR uses the anatomical reference with residual-limb topology. Joint and endpoint errors are evaluated against Vicon-derived keypoints. Mean vertex error (MVE) measures agreement with the corresponding fitted surface. Within the full-body protocol, TokenHMR and PromptHMR have similar MPJPE of 148.81 and 148.09 mm and PA-MPJPE of 87.30 and 87.25 mm. Their MVE differs substantially, at 177.81 and 107.44 mm. Joint alignment therefore does not determine fitted-surface agreement. ResiHMR obtains 135.32 mm MPJPE and 101.47 mm MVE under its residual-topology protocol, while its endpoint error remains 154.22 mm. Appendix D details the reference construction and topology-specific interpretation.

Reference sensitivity. We also evaluate fixed predicted meshes against kinematic references initialized from HSMR and PromptHMR. Replacing the HSMR-based reference with the PromptHMR-based reference increases MVE by 25.067 mm for PromptHMR, 37.331 mm for TokenHMR, and 47.391 mm for CameraHMR. Their ordering remains unchanged under the two references. This separates stability of method ordering from stability of the absolute surface-error values. The complete comparison is provided in Appendix D.5. We retain Vicon-derived joint and endpoint errors as primary geometric measures and interpret MVE as reference-relative fitted-surface consistency.

Participant and side-specific errors. For the four mesh methods, participant-equal MPJPE differs from the pooled values by 1.62 to 4.53 mm. The MPJPE ordering within the full-body protocol is unchanged when each participant is excluded in turn. Appendix F.1 reports the participant-level standard deviations and leave-one-participant-out ranges.

We further compare affected and contralateral sides within the same participant and recorded test using participant-balanced macro-MPJPE. Five of the six multi-view 3D pose configurations have higher affected-side errors under both prosthesis-worn and non-worn conditions. The differences range from 11.71 to 45.35 mm and from 4.98 to 32.66 mm, respectively. Fine-tuned YOLOPose-l is the exception. All four mesh methods have higher affected-side ankle errors by 28.1 to 46.2 mm.

Hip differences are smaller at 3.9 to 10.2 mm. The ankle has the largest side-wise gap among the evaluated hip, knee, and ankle landmarks for every mesh method. These reconstructed-joint errors identify spatial asymmetry within the evaluated lower-limb cases. Appendix F.2 provides the full protocol and results.

Summary. The benchmark reveals three complementary findings. Image fine-tuning improves in-the-wild AP for all evaluated pose models, but its effect on calibrated frames depends on the model. Residual endpoints remain difficult to localize. The HMR side comparison shows the largest affected-side error increase at the ankle among the evaluated hip, knee, and ankle landmarks. Human mesh evaluation adds information beyond joint alignment. TokenHMR and PromptHMR have similar joint errors, yet their MVE differs by 70.37 mm under the same kinematic reference. Changing between the two tested references shifts absolute surface errors while preserving the ordering of the full-body methods. Together, these results distinguish adaptation to visual conditions, target-specific localization, and fitted-surface agreement. LDBODY makes these aspects of pose and mesh recovery separately measurable within one benchmark.

5 DISCUSSION

LDBODY treats body structure and prosthesis-supported function as complementary targets of pose and mesh evaluation. Residual endpoints localize the remaining limb, while prosthesis-supported landmarks identify functional locations on the worn device. Preserving both allows the person and the device to remain distinct within a common pose layout. Shared landmark definitions support evaluation across variations in body proportions, limb configuration, and prosthesis appearance. The in-the-wild images broaden the visual conditions, while calibrated recordings add geometric evidence for examining reconstruction.

The results suggest concrete directions for model development. Changes in image AP need to be examined alongside endpoint localization and multi-view reconstruction. The separate annotation targets support studying how learning standard body landmarks interacts with learning residual and prosthesis-supported landmarks. The anatomical and kinematic references further support studying representations of residual body structure and the functional motion chain. This extends evaluation from locating joints to examining the body surfaces and motion relationships represented by a model.

Real recordings also provide a basis for evaluating future augmentation and synthetic training methods against observed geometry. Target-specific evaluation can examine whether these methods improve residual localization and reconstruction alongside overall pose scores. Controlled research access supports this work while protecting participant recordings. Appendix G discusses coverage, annotation uncertainty, reference dependence, and responsible use.

6 CONCLUSION

We presented LDBODY, a dataset for pose and 3D human mesh evaluation across diverse in-the-wild images and real calibrated motion. A protocol guided by multiple internationally accredited Para Athletics classifiers distinguishes body joints, residual-limb endpoints, and prosthesis-supported landmarks. Vicon-derived keypoints and fitted body meshes support evaluation of 2D pose, multi-view 3D pose, residual-limb localization, and human mesh recovery. The evaluated models retain substantial errors at residual endpoints and affected distal joints. Similar joint errors can also coexist with markedly different fitted-surface agreement. By making body variation and prosthesis-supported function explicit evaluation targets, LDBODY provides a foundation for accessible motion analysis, assistive technologies, and inclusive sport.

AI USE STATEMENT

In this work, we used ChatGPT to assist with literature exploration and revising manuscript text, improving readability, and formatting LaTeX tables and other manuscript content. We have not used generative AI tools for dataset protocol design, collection, annotation, quality check, and experiment results.

The authors reviewed and revised all AI-assisted content included in the submission. Specifically, we checked methodological descriptions against the implemented procedures, checked interpretations against the experimental results, and verified cited references against the original sources. All final research decisions and conclusions were made by the authors. We take full responsibility for the final content of this work, including text, claims, and research artifacts produced with the assistance of generative AI tools.

ETHICS STATEMENT

The curation of LDBODY was conducted under institutional human research ethics approval, with the institution and approval identifier omitted for double-blind review. All recruited participants provided informed consent before data collection, including consent for motion-capture recording, multi-view RGB video capture, and controlled research use of the collected data.

LDBODY is intended to support more inclusive pose estimation and human mesh recovery for individuals with limb deficiency and prosthesis users. We use coded participant identifiers and face mosaicing in derived visualizations to reduce direct identifiability. However, these measures do not guarantee anonymity: body shape, movement patterns, prosthesis configuration, and other visual characteristics may still enable re-identification. Accordingly, potentially identifiable RGB recordings, motion-capture data, and fitted mesh references will be made available to approved non-commercial research users under a data-use agreement and access review. The agreement prohibits attempted re-identification, unauthorized redistribution, surveillance, unsupported clinical decision-making, and use outside the stated research scope.

The motion results characterize the recorded participants and protocol. They should not be used to make clinical or eligibility decisions about individual participants. Prosthesis-supported keypoints are functional motion landmarks rather than assertions of biological anatomy, and fitted mesh references are model-derived geometric estimates rather than clinical surface ground truth. For the image records in LDBODY, reuse and redistribution of source images remain subject to the original dataset's licensing and access conditions.

REPRODUCIBILITY STATEMENT

Section 3 and Appendix C describe dataset composition, capture procedures, landmark definitions, and keypoint conversion. Appendix D details fitted mesh construction and reports reference diagnostics. Section 4 and Appendix E describe the benchmark tasks, evaluation metrics, validity rules, aggregation, and computational resources. This includes Procrustes alignment for PA-MPJPE. Appendix F reports participant, condition, movement, side-specific, and viewpoint analyses. Appendix A describes review access and the assets required for metric recomputation and end-to-end evaluation. Experiments involving restricted participant recordings require approved access under the data-use agreement.

REFERENCES

Mohammad Pourmahmood Aghababa and Jan Andrysek. Exploration and demonstration of explainable machine learning models in prosthetic rehabilitation-based gait analysis. *Plos one*, 19(4): e0300447, 2024.

Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, and Bernt Schiele. 2d human pose estimation: New benchmark and state of the art analysis. In *Proceedings of the IEEE Conference on computer Vision and Pattern Recognition*, pp. 3686–3693, 2014.

Federica Bogo, Angjoo Kanazawa, Christoph Lassner, Peter Gehler, Javier Romero, and Michael J. Black. Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image.

In *Computer Vision – ECCV 2016*, Lecture Notes in Computer Science. Springer International Publishing, October 2016.

Zhongang Cai, Daxuan Ren, Ailing Zeng, Zhengyu Lin, Tao Yu, Wenjia Wang, Xiangyu Fan, Yang Gao, Yifan Yu, Liang Pan, Fangzhou Hong, Mingyuan Zhang, Chen Change Loy, Lei Yang, and Ziwei Liu. HuMMan: Multi-modal 4d human dataset for versatile sensing and modeling. In *17th European Conference on Computer Vision, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII*, pp. 557–577. Springer, 2022.

Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE transactions on pattern analysis and machine intelligence*, 43(1):172–186, 2019.

Hyunjin Cho, Giyun Choi, and Jongwon Choi. Ajahr: Amputated joint aware 3d human mesh recovery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7925–7935, 2025.

Heming Du, Jiaying Ying, Sen Wang, Xue Li, Kaihao Zhang, and Xin Yu. Inclusivevidpose: Bridging the pose estimation gap for individuals with limb deficiencies in video-based motion. In *The Fourteenth International Conference on Learning Representations*, 2026. URL `https://openreview.net/forum?id=SyQqXAdWUq`.

Sai Kumar Dwivedi, Yu Sun, Priyanka Patel, Yao Feng, and Michael J. Black. TokenHMR: Advancing human mesh recovery with a tokenized pose representation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

Hao-Shu Fang, Jiefeng Li, Hongyang Tang, Chao Xu, Haoyi Zhu, Yuliang Xiu, Yong-Lu Li, and Cewu Lu. Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time. *IEEE transactions on pattern analysis and machine intelligence*, 45(6):7157–7173, 2022.

Zigang Geng, Ke Sun, Bin Xiao, Zhaoxiang Zhang, and Jingdong Wang. Bottom-up human pose estimation via disentangled keypoint regression. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14676–14686, 2021.

Shubham Goel, Georgios Pavlakos, Jathushan Rajasegaran, Angjoo Kanazawa, and Jitendra Malik. Humans in 4D: Reconstructing and tracking humans with transformers. In *ICCV*, 2023.

Chun-Hao P. Huang, Hongwei Yi, Markus Höschle, Matvey Safroshkin, Tsvetelina Alexiadis, Senya Polikovsky, Daniel Scharstein, and Michael J. Black. Capturing and inferring dense full-body human-scene contact. In *Proceedings IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 13274–13285, June 2022.

Grace R Hunt, Lukas Gabert, Colby Hansen, K Bo Foreman, and Tommaso Lenzi. Open dataset of kinetics, kinematics, and electromyography of above-knee amputees during stand-up and sit-down. *Scientific Data*, 12(1):1–31, 2025.

Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE transactions on pattern analysis and machine intelligence*, 36(7):1325–1339, 2013.

Tao Jiang, Peng Lu, Li Zhang, Ningsheng Ma, Rui Han, Chengqi Lyu, Yining Li, and Kai Chen. Rtmpose: Real-time multi-person pose estimation based on mmpose. *arXiv preprint arXiv:2303.07399*, 2023.

Hanbyul Joo, Hao Liu, Lei Tan, Lin Gui, Bart Nabbe, Iain Matthews, Takeo Kanade, Shohei Nobuhara, and Yaser Sheikh. Panoptic studio: A massively multiview system for social motion capture. In *Proceedings of the IEEE international conference on computer vision*, pp. 3334–3342, 2015.

Marilyn Keller, Keenon Werling, Soyong Shin, Scott Delp, Sergi Pujades, C Karen Liu, and Michael J Black. From skin to skeleton: Towards biomechanically accurate 3d digital humans. *ACM Transactions on Graphics (TOG)*, 42(6):1–12, 2023.

Rawal Khirodkar, Aayush Bansal, Lingni Ma, Richard Newcombe, Minh Vo, and Kris Kitani. Egohumans: An egocentric 3d multi-human benchmark. *arXiv preprint arXiv:2305.16487*, 2023.

Rawal Khirodkar, Timur Bagautdinov, Julieta Martinez, Su Zhaoen, Austin James, Peter Selednik, Stuart Anderson, and Shunsuke Saito. Sapiens: Foundation for human vision models. In *European Conference on Computer Vision*, pp. 206–228. Springer, 2024.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer vision–ECCV 2014: 13th European conference, zurich, Switzerland, September 6-12, 2014, proceedings, part v 13*, pp. 740–755. Springer, 2014.

Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. Smpl: A skinned multi-person linear model. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pp. 851–866. 2023.

Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, et al. Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172*, 2019.

Naureen Mahmood, Nima Ghorbani, Nikolaus F. Troje, Gerard Pons-Moll, and Michael J. Black. AMASS: Archive of motion capture as surface shapes. In *International Conference on Computer Vision*, pp. 5442–5451, October 2019.

Debapriya Maji, Soyeb Nagori, Manu Mathew, and Deepak Poddar. Yolo-pose: Enhancing yolo for multi person pose estimation using object keypoint similarity loss. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2637–2646, 2022.

Dushyant Mehta, Helge Rhodin, Dan Casas, Pascal Fua, Oleksandr Sotnychenko, Weipeng Xu, and Christian Theobalt. Monocular 3d human pose estimation in the wild using improved cnn supervision. In *2017 international conference on 3D vision (3DV)*, pp. 506–516. IEEE, 2017.

Priyanka Patel and Michael J. Black. CameraHMR: Aligning people with perspective. In *International Conference on 3D Vision (3DV)*, 2025.

Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed AA Osman, Dimitrios Tzionas, and Michael J Black. Expressive body capture: 3d hands, face, and body from a single image. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10975–10985, 2019.

Tianye Qi, Tengyue Zhang, Jiaying Ying, Tianqing Zhu, and Xin Yu. Topology-unified 2d pose estimation across intact, residual and prosthetic limbs. *arXiv preprint arXiv:2608.13047*, 2026.

Joachim Tesch, Giorgio Becherini, Prerana Achar, Anastasios Yiannakidis, Muhammed Kocabas, Priyanka Patel, and Michael J. Black. BEDLAM2.0: Synthetic humans and cameras in motion. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2025.

Shashank Tripathi, Lea Müller, Chun-Hao P. Huang, Taheri Omid, Michael J. Black, and Dimitrios Tzionas. 3D human pose estimation via intuitive physics. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4713–4725, 2023. URL `https://ipman.is.tue.mpg.de`.

Matt Trumble, Andrew Gilbert, Charles Malleson, Adrian Hilton, and John Collomosse. Total capture: 3d human pose estimation fusing video and inertial sensors. In *2017 British Machine Vision Conference (BMVC)*, 2017.

Yufu Wang, Yu Sun, Priyanka Patel, Kostas Daniilidis, Michael J Black, and Muhammed Kocabas. Prompthmr: Promptable human mesh recovery. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025.

Yilin Wen, Kechuan Dong, Fumiya Suginaka, Ken Endo, and Yusuke Sugano. Prosthesis-aware 3d human pose estimation: A dataset and benchmark for rsp users. In *European Conference on Computer Vision*, pp. 430–447. Springer, 2026.

Yan Xia, Xiaowei Zhou, Etienne Vouga, Qixing Huang, and Georgios Pavlakos. Reconstructing humans with a biomechanically accurate skeleton. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 5355–5365, 2025.

Qingzheng Xu, Ru Cao, Xin Shen, Heming Du, Sen Wang, and Xin Yu. M3gym: A large-scale multimodal multi-view multi-person pose dataset for fitness activity understanding in real-world settings. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 12289–12300, 2025.

Yufei Xu, Jing Zhang, Qiming Zhang, and Dacheng Tao. Vitpose: Simple vision transformer baselines for human pose estimation. *Advances in neural information processing systems*, 35:38571–38584, 2022.

Zhendong Yang, Ailing Zeng, Chun Yuan, and Yu Li. Effective whole-body pose estimation with two-stages distillation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4210–4220, 2023.

Xiangyu Yin, Boyuan Yang, Weichen Liu, Qiyao Xue, Abrar Alamri, Goeran Fiedler, and Wei Gao. Progait: A multi-purpose video dataset and benchmark for transfemoral prosthesis users. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8984–8993, 2025.

Jiaying Ying, Heming Du, Kaihao Zhang, Lincheng Li, and Xin Yu. Ldpose: Towards inclusive human pose estimation for limb-deficient individuals in the wild. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9865–9875, 2025.

Jiaying Ying, Heming Du, Kaihao Zhang, Sean M Tweedy, and Xin Yu. Resihmr: Residual-limb aware single-image 3d human mesh recovery for individuals with limb loss. *arXiv preprint arXiv:2604.28025*, 2026.

Siwei Zhang, Qianli Ma, Yan Zhang, Zhiyin Qian, Taein Kwon, Marc Pollefeys, Federica Bogo, and Siyu Tang. Egobody: Human body shape and motion of interacting people from head-mounted devices. In *European conference on computer vision*, pp. 180–200. Springer, 2022.

A APPENDIX

A REVIEW ACCESS AND REPRODUCIBILITY ASSETS

Reading guide. Appendix B places the contribution relative to prior work. Appendix C specifies data provenance, coverage, annotation semantics, and coordinate conversion. Appendix D describes fitted mesh construction and reference sensitivity. Appendix E defines the evaluation protocols. Appendix F reports participant, condition, movement, deficiency-level, side-specific, and camera diagnostics. Appendix G discusses interpretation, remaining uncertainty, and responsible use. All sections use the same LDBODY terminology and landmark definitions.

A.1 REVIEW-ONLY ARCHIVE

The review package contains potentially identifiable images, videos, and motion-capture records. It uses a password-protected, header-encrypted split 7-Zip archive with nominal 10 GB volumes. The accompanying review-only access note contains the case-sensitive password. That note is intended for a review-restricted channel rather than the public manuscript. Access for review does not authorize redistribution or reuse outside the consented research scope.

The existing archive retains its original file names. These are download identifiers, not separate dataset names. All volumes must be placed in the same directory without renaming. Extraction begins at the first volume and uses subsequent volumes automatically:

```
LDBody_raw.7z.001
LDBody_raw.7z.002
LDBody_raw.7z.003
...
```

On Linux or another system with `7z`, test and extract with:

```
7z t LDBody_raw.7z.001
7z x LDBody_raw.7z.001 -oLDBody_extracted
```

On macOS, the corresponding executable may be `7zz`:

```
7zz t LDBody_raw.7z.001
7zz x LDBody_raw.7z.001 -oLDBody_extracted
```

On Windows, use a 7-Zip-compatible application to extract the first volume, or run:

```
& "C:\Program Files\7-Zip\7z.exe" x `
  .\LDBody_raw.7z.001 -o.\LDBody_extracted
```

Enter the password at the interactive prompt rather than placing it in a command line or shell history. If archive testing fails, first check that every volume is present and that downloads are complete. Testing or extraction of a later volume in isolation is not the supported procedure. A short destination path can avoid path-length problems. These commands describe extraction of the supplied format rather than verification that a particular download is complete.

A.2 ACCESS TIERS AND REPRODUCTION SCOPE

Metric recomputation and end-to-end reproduction require different assets. Recomputing an existing score requires fixed predictions, the applicable references, record identifiers, validity masks, coordinate and topology mappings, and the evaluator configuration. Running a new RGB-based model additionally requires access to the corresponding RGB inputs and a reproducible model configuration. Reconstructing multi-view keypoints requires camera calibration and the frame mapping between views. Regenerating fitted meshes further requires the base predictions, body model, topology conversion, fitting parameters, and selection rule. Access to annotations alone therefore does not establish that RGB inference or mesh fitting can be rerun.

The prospective recordings and their derived identifiable assets are intended for approved non-commercial research under a data-use agreement. An application specifies the research purpose, verifiable institutional affiliation, responsible investigator, requested assets, intended outputs, and a data-security plan. The intended use must remain consistent with participant consent. Original LDPose images and inherited annotations retain their source access and licensing terms (Ying et al., 2025). Additional annotation access does not override restrictions on the underlying images.

The review archive is an access mechanism, not a public deployment or a hosted evaluation service. No fixed approval turnaround or operational server-side evaluation is assumed here. Public release documentation must identify the responsible access-review body and the procedures for renewal, delayed requests, revocation, and appeal once approved. The release should state which tasks can be recomputed from accessible derived assets and which require restricted RGB. The existence of a project page or a model name alone is not evidence of executable reproduction.

Version and audit information. A reproducible evaluation is identified by its dataset version, exact record selection, model and checkpoint version, preprocessing, keypoint layout, reference version, valid-prediction rule, alignment, and aggregation. For a split archive, the volume inventory and file sizes support completeness checking. When checksums are supplied, they should be verified before extraction. Appendix E.1 lists the settings needed to interpret or rerun each task.

B Extended Related Work

Comparison criteria. We distinguish the inclusion of prosthesis users from annotations of prosthesis status, functional landmarks, and device geometry. Paired video views are also different from calibrated 3D references. A human body mesh and a prosthesis surface describe different targets. Tables 1 and 2 compare reference annotations rather than any outputs a model could predict from a dataset. Access conditions are separate from these technical capabilities.

General human pose and reconstruction datasets. Human3.6M (Ionescu et al., 2013), TotalCapture (Trumble et al., 2017), and MPI-INF-3DHP (Mehta et al., 2017) provide images with 3D pose supervision. Panoptic Studio (Joo et al., 2015) emphasizes multi-person social interaction. MOYO (Tripathi et al., 2023), HuMMan (Cai et al., 2022), RICH (Huang et al., 2022), and M3GYM (Xu et al., 2025) extend body reconstruction, interaction, or multi-view evaluation. EgoHumans (Khirodkar et al., 2023) and EgoBody (Zhang et al., 2022) address egocentric and interacting settings. AMASS (Mahmood et al., 2019) unifies motion capture through parametric body fitting. BEDLAM2.0 (Tesch et al., 2025) provides synthetic human data. These resources already include substantial pose and body-model supervision. The distinction of LDBody is its explicit treatment of residual anatomy and prosthesis-supported function, not an assertion that earlier datasets lack all 3D or body-mesh references.

Inclusive image and video annotations. LDPose (Ying et al., 2025) adds residual-limb endpoints to standard body keypoints. InclusiveVidPose (Du et al., 2026) provides video pose, segmentation, tracking identities, and per-limb prosthesis status. Presence or status labels do not by themselves specify functional landmarks on a device. LDBody preserves the source LDPose body and residual labels while adding prosthesis-supported landmarks. It uses shared landmark meanings in its images and calibrated recordings. This does not assign a 3D body reference to every in-the-wild image.

Gait and biomechanics resources. ProGait (Yin et al., 2025) records transfemoral prosthesis users in paired frontal and sagittal views. It includes several knee configurations and supports segmentation, 2D pose, and gait analysis. Paired views should not be described as single-view capture, but do not by themselves establish calibrated 3D body-mesh references. Other biomechanics datasets supply kinematic, kinetic, electromyographic, or force measurements for gait and prosthesis analysis (Hunt et al., 2025; Aghababa & Andrysek, 2024). These are valuable targets in their own right. A pose-landmark error in LDBody is not a substitute for those measurements or a clinical outcome.

Concurrent prosthesis-aware 3D evaluation. RSP3D (Wen et al., 2026) focuses on running-specific prosthesis users and includes daily activities as well as exercise. It uses 16 RGB cameras and

OptiTrack. Body-joint references support joint errors, and markers along the device define a surface for geometric evaluation. That surface is not a fitted human body reference. LDBODY provides explicit residual endpoints, prosthesis-supported functional landmarks, and fitted human bodies with anatomical and kinematic targets. Conversely, a fitted kinematic body in LDBODY does not replace direct prosthesis-surface measurements. The two resources address related but distinct geometric questions.

Topology-aware 2D representation. ProPose (Qi et al., 2026) introduces Omni-Pose annotations with Biological, Prosthetic, or Absent keypoint types. It includes residual endpoints and distal landmarks, expands real and synthetic image training data, and uses ProLoss to model dependencies within limbs. This work addresses 2D localization and semantic classification. LDBODY shares the motivation to distinguish biological structure from prosthesis-supported function. Its additional focus is real calibrated 3D pose and human body mesh evaluation. The contribution is not the general idea of assigning keypoints to prostheses, but the dataset and reference design for separately evaluating the associated geometric targets.

2D pose estimation methods. OpenPose (Cao et al., 2019) uses a bottom-up formulation. AlphaPose (Fang et al., 2022) and ViTPose (Xu et al., 2022) provide representative person-centric pose estimators. YOLOPose (Maji et al., 2022), RTMPose (Jiang et al., 2023), and MediaPipe (Lugaresi et al., 2019) address practical pose estimation. DEKR (Geng et al., 2021), DWPose (Yang et al., 2023), and Sapiens (Khirodkar et al., 2024) examine regression, distillation, and large-scale pretraining. The three evaluated pose families offer distinct baseline configurations. Their results are not an exhaustive evaluation of all pose architectures.

Human mesh recovery and body representations. SMPL (Loper et al., 2023) and SMPL-X (Pavlakos et al., 2019) are common parametric representations. SMPLify (Bogo et al., 2016) and SMPLify-X (Pavlakos et al., 2019) fit body parameters to observations. HMR2.0 (Goel et al., 2023), TokenHMR (Dwivedi et al., 2024), CameraHMR (Patel & Black, 2025), and PromptHMR (Wang et al., 2025) develop learning-based reconstruction. SKEL (Keller et al., 2023) and HSMR (Xia et al., 2025) incorporate biomechanically motivated skeleton modeling. AJAHR (Cho et al., 2025) introduces the synthetic A3D data with approximately one million images, 2D/3D joints, SMPL parameters, and amputation-region labels. ResiHMR (Ying et al., 2026) also addresses limb-deficient body reconstruction. These works establish that both synthetic amputation-aware supervision and residual-topology reconstruction precede this benchmark. They motivate real calibrated evaluation rather than a claim that LDBODY is the first work to represent an amputated body mesh.

Scope of the contribution. The priority claim in the main text concerns calibrated human body mesh evaluation with separate references for residual anatomy and prosthesis-supported function. It does not claim the first prosthesis-user dataset, the first 2D residual labels, the first real 3D prosthesis geometry, or the first synthetic amputated body. The dataset also supports diverse image-based pose evaluation under the shared keypoint definitions. This task and reference distinction is more specific than counting the number of modalities or applying a single broad label such as "inclusive."

C Data Collection and Annotation Details

C.1 Data Provenance, Coverage, and Counting

LDBODY contains real in-the-wild images and prospectively captured motion recordings under shared landmark definitions. The image records retain the public LDPose source identifiers and inherited body-joint and residual-limb annotations (Ying et al., 2025). The additional annotation work adds prosthesis-supported landmarks to 8,866 selected images. This count refers to annotated images, not unique participants, newly collected images, or all images in the original LDPose collection. The motion records provide synchronized RGB and Vicon observations with point and fitted mesh references.

The dataset includes upper- and lower-limb differences, dwarfism, and multiple prosthesis types. Image diversity also covers body proportions, activities, scenes, clothing, and viewpoints. Dwarfism is present in the motion collection as well. Coverage categories can overlap. A prosthesis user can

also belong to a body-proportion category, and category presence does not imply an equal number of independently recorded people or an equal set of reference annotations. We distinguish data inclusion from inclusion in a particular benchmark analysis.

Table 6 lists the motion participants and recordings. The motion collection contains five adult male participants aged 20 to 65 years. The table records limb-deficiency and device attributes. Body-proportion categories can overlap these attributes. Image coverage is not restricted to the demographics of the motion cohort. The present tables do not provide a separate performance estimate for dwarfism.

Table 6: **Coded lower-limb evaluation records in LDBODY.** Videos and RGB frames are view-specific. Durations are in seconds. The sum describes the records listed here. Body-proportion attributes and subgroup performance for dwarfism are not tabulated separately.

Participant	Deficiency level	Side	Prosthesis category	Tests	Videos	Frames	Mean duration
AP001	Transtibial	Left	Blade-shaped	27	216	192,528	17.83
AP002	Transfemoral	Left	Exposed-pylon	22	176	135,928	15.45
AP003	Transtibial	Left	Blade-shaped	25	200	155,104	15.51
AP004	Transtibial	Left	Blade-shaped	30	240	183,296	15.27
AP005	Transfemoral	Right	Cosmetic cover	32	256	296,248	23.14
Total	N/A	N/A	N/A	136	1,088	963,104	17.70

Each synchronized test contributes eight view-specific RGB videos. For the tabulated block,

$$136 \times 8 = 1,088, \qquad \frac{963,104}{1,088 \times 50} = 17.70 \text{ seconds, approximately.}$$

The 963,104 frames count all views. A simultaneous observation from eight cameras is not eight independent motions. Counts of source frames also need not equal the number of frames or valid landmarks retained for a particular score. The image and video-frame counts are not added into a single interchangeable sample count.

Table 7: **Coverage of the reported motion diagnostics.** Movement groups partition the 136 tests. Deficiency groups partition the five coded participants. Frames count all RGB views. The prosthesis-condition split contains 120 worn and 16 non-worn tests.

Grouping	Category	Participants	Tests	Frames
Movement	Static/postural	5	31	177,816
Movement	Gait/locomotion	5	21	125,968
Movement	Transitional/functional	5	52	439,240
Movement	Running/sport dynamic	5	32	220,080
Deficiency level	Transtibial	3	82	530,928
Deficiency level	Transfemoral	2	54	432,176

The movement taxonomy used for diagnostics has four categories. It is distinct from the more detailed activity names used during collection. The worn and non-worn conditions contain 120 and 16 tests in the tabulated block. Deficiency level, side, device appearance, participant identity, and activity distribution are not fully crossed. Four coded participants have left-side deficiency. AP005 is the only right-side case. The exposed-pylon and cosmetic-cover categories each contain one coded participant. These counts support descriptive stratification but not independent side or device-type effect estimates.

C.2 CAPTURE SETUP AND PROTOCOL

Sensors and timing. Eight synchronized RGB cameras surround an indoor capture volume. RGB videos are recorded at 2048×1080 resolution and 50 fps. The Vicon optical system records marker trajectories at 100 Hz. Calibration associates the RGB cameras with the Vicon coordinate system. A 20 ms RGB frame interval corresponds nominally to two Vicon sampling intervals. The sampling-rate ratio alone does not determine the recording offset or identify the frames actually scored. The timestamp association and any frame-index fallback belong to the conversion configuration.

Environment and safety. The capture space contains equipment, stored objects, and ordinary indoor illumination rather than uniform visual background panels. A support handle is available for standing, balance, locomotion, and transition tasks. Frames with handle-related occlusion are retained. The handle supports safe data collection and is part of the observed scene, not a digitally added occluder. Indoor clutter does not make the recordings equivalent to unconstrained outdoor capture. It provides realistic visual constraints within a calibrated setting.

Recruitment and participation requirements. Recruitment extended beyond the final recorded cohort. Participation required the willingness to attend an in-person capture session and the ability to perform the prescribed movements safely. The protocol included standing, balance, locomotion, and functional transitions under researcher supervision. Prosthesis users performed activities with and without their devices where applicable. These requirements narrowed the pool of eligible and available volunteers. Five participants met the participation requirements and completed recording. The resulting cohort reflects the movement and attendance requirements of prospective capture. Participant-balanced summaries and leave-one-participant-out analyses describe variation within this recorded cohort.

Quality inspection. The processing includes synchronization inspection, calibration checks, Vicon-to-image projection checks, keypoint visualization, and fitting diagnostics. These checks serve different purposes. Projection consistency checks coordinate conversion, while fitting residuals describe agreement with constraints used in optimization. Neither is automatically an independent measurement of the true body surface. This appendix reports the available fitting diagnostics. It does not assign an unmeasured calibration, synchronization, or marker-placement error bound.

C.3 Shared Landmark Semantics and Image Annotation

Annotation design. Multiple internationally accredited Para Athletics classifiers guided the protocol. The framework distinguishes body joints, residual-limb endpoints, and prosthesis-supported landmarks. It preserves compatibility with standard pose layouts while assigning these locations different meanings. A standard slot such as "ankle" can describe a biological joint on an intact limb or a functional landmark on a prosthesis. The slot name alone is not a statement about biological tissue. The participant and limb configuration determine the applicable interpretation.

Without a worn prosthesis, the residual endpoint remains labeled and unavailable distal body points remain unlabeled. With a worn prosthesis, the residual annotation is retained and applicable functional landmarks are added on the device. Residual and distal functional points can therefore coexist. This differs from replacing the residual limb with a hypothetical complete biological limb. Masks, boxes, and prosthesis-presence labels can complement the framework, but do not specify the point locations used by pose and reconstruction protocols.

Image selection and annotation provenance. The 8,866-image annotation set is drawn from public LDPose images with visible prosthetic components. Original body and residual annotations are retained. New labels concern the applicable prosthesis-supported points rather than a complete relabeling of the inherited data. Source-image identifiers and the designated image splits preserve the relation to the original resource. Existing rights in source images and annotations are not changed by the LDBody name. Images that cannot be redistributed remain subject to the original access procedure.

Device-specific rules. For anatomically shaped prostheses, placement follows the corresponding intact-limb convention. For a visible mechanical knee or ankle, the point is placed at the projected joint centre or apparent rotation axis. For a running blade without a discrete ankle articulation, a virtual functional ankle is defined on the device centreline at the anatomical-equivalent level. The contralateral ankle, body proportions, and the transition into the curved blade section provide placement cues. The point is not the blade tip or the instantaneous ground-contact point. Its error measures agreement with this operational definition, not blade deformation or a measured physical ankle rotation axis.

Visibility and ambiguity. For manually annotated images, $v = 1$ denotes an occluded landmark whose location can be reliably inferred. An ambiguous location is not assigned speculatively. Out-of-

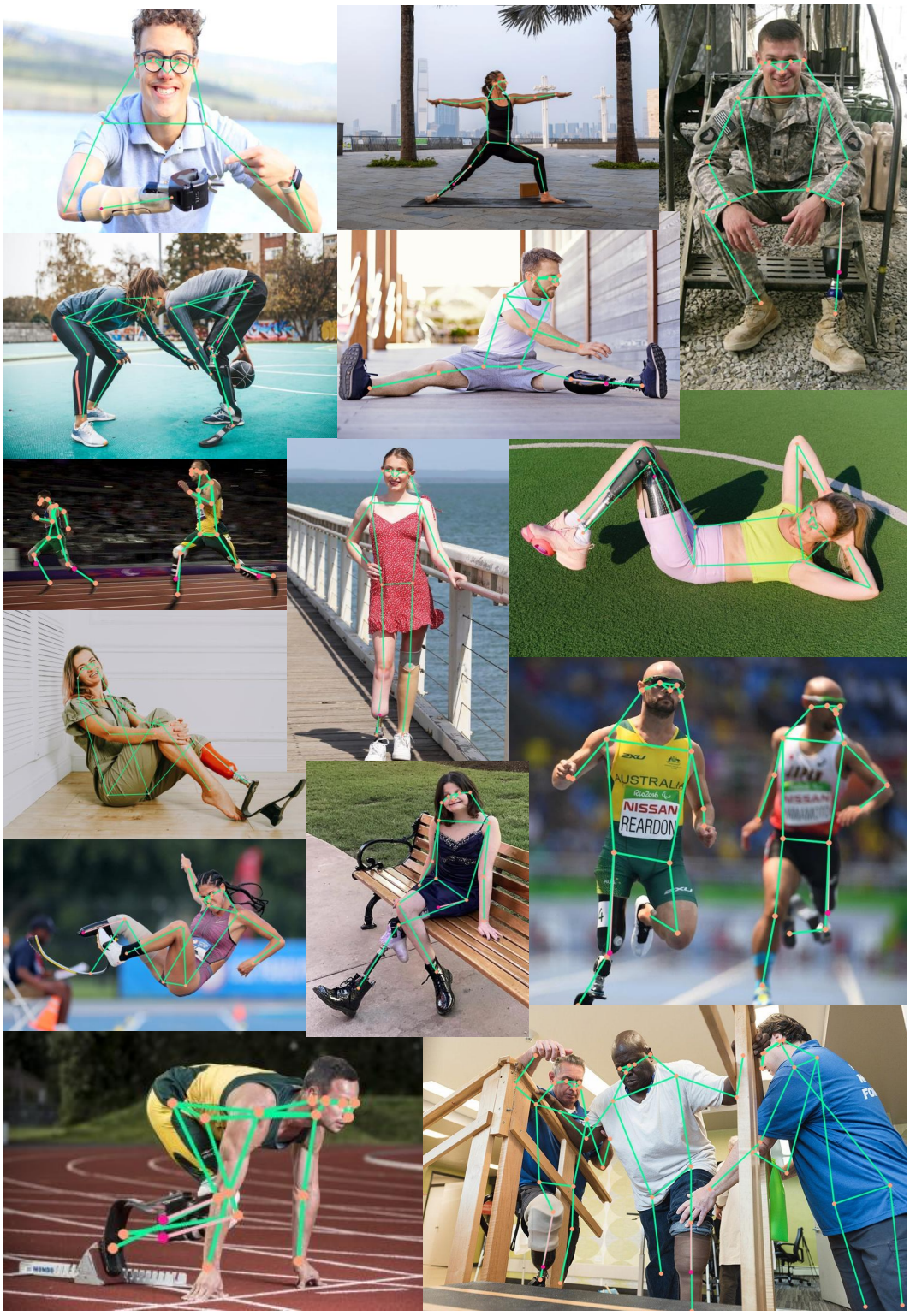

Figure 4: **In-the-wild keypoint annotations in LDBODY.** Examples illustrate variation in activity, clothing, viewpoint, body configuration, and prosthesis appearance. Original LDPose body-joint and residual-limb annotations are retained. Added device landmarks describe functional locations rather than biological joints. These examples do not enumerate every coverage category.

image handling follows the inherited image annotation convention. An existing but occluded point, an unavailable label, and an anatomically absent distal joint are distinct cases. The projection-derived availability flag described below is not an independent manual visibility annotation.

Annotator training and quality control. Three annotators completed practice annotations that were reviewed before production. They retained the inherited labels and added the applicable prosthesis landmarks. Ten percent of completed image annotations were randomly sampled for manual quality review. Classifier guidance concerns the annotation protocol and does not imply that classifiers independently reannotated every image. The 10% review is a quality audit, not a formal inter-annotator agreement experiment. No inter-annotator precision, kappa, or millimeter error is inferred from the audit percentage.

C.4 Vicon Marker Placement and Keypoint Conversion

The full-body Plug-in Gait marker set provides body-motion references. For terminal limb deficiency, four additional markers provide the residual-endpoint configuration. Figure 5 shows the placement examples. All output 3D coordinates use the Vicon world frame and millimeters. The semantic landmark definitions are shared with the image annotations, while their measurement and annotation procedures differ.

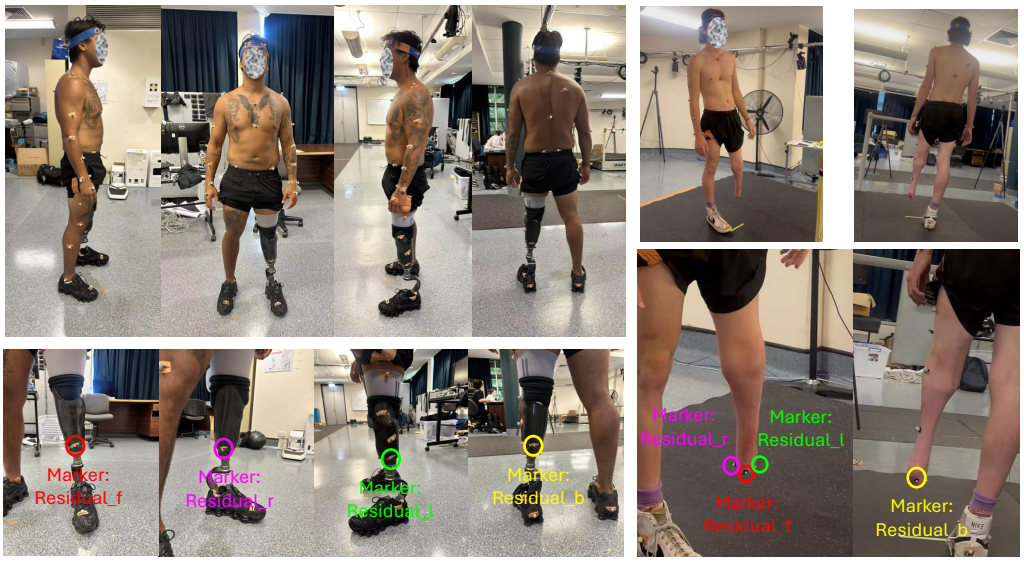

Figure 5: **Marker placement for the illustrated LDBody recordings.** The Plug-in Gait marker set provides body-motion references. The four additional markers `residual_f`, `residual_r`, `residual_l`, and `residual_b` define the configuration used for the residual-endpoint reference. Examples show prosthesis-worn and non-worn conditions. The illustration is a placement example rather than a direct scan of internal residual anatomy.

Marker validity and aggregation. Let $M_f(m) \in \mathbb{R}^3$ denote a marker or joint-centre source at Vicon frame f. The conversion accepts finite coordinates that pass its zero-value validity check and treats invalid or missing sources as unavailable. Write

$$\mathrm{fetch}(m, f) = \begin{cases} M_f(m), & \text{if the source is valid,} \\ \mathrm{NaN}, & \text{otherwise.} \end{cases}$$

For a keypoint with source set S_k, the conversion uses

$$J_f(k) = \mathrm{nanmean}_{m \in S_k} \mathrm{fetch}(m, f).$$

The mean uses available valid sources. If none is valid, the keypoint has no reference. The term "non-zero" describes the converter's missing-data convention, not a general geometric claim that any point with a zero coordinate is physically invalid. The exact zero-value predicate remains part of the conversion code.

Body-keypoint layouts. Table 8 gives the 25-body-slot OpenPose ordering (Cao et al., 2019). Table 9 gives the 17-body-slot COCO ordering (Lin et al., 2014) used by LDPose. Appending eight residual slots produces 33 and 25 slots, respectively. These two uses of "25" must not be confused. Face landmarks and small toes without a reliable source remain unavailable. Schema capacity is not evidence that every keypoint exists or has a valid reference in every record.

Table 8: **Vicon to OpenPose-style body-keypoint conversion.** Indices are zero-based. N/A denotes no reliable Vicon source, not an absent anatomical part. The eight residual slots are appended after these 25 body slots. The mean uses the available valid source points.

Index	Body keypoint	Vicon source
0	nose	N/A
1	neck	mean(C7, CLAV)
2	right shoulder	RSJC
3	right elbow	REJC
4	right wrist	mean(RWJC, RWRA, RWRB)
5	left shoulder	LSJC
6	left elbow	LEJC
7	left wrist	mean(LWJC, LWRA, LWRB)
8	mid hip	PelvisCOM
9	right hip	RHJC
10	right knee	RKJC
11	right ankle	RAJC
12	left hip	LHJC
13	left knee	LKJC
14	left ankle	LAJC
15	right eye	N/A
16	left eye	N/A
17	right ear	N/A
18	left ear	N/A
19	left big toe	LTOE
20	left small toe	N/A
21	left heel	LHEE
22	right big toe	RTOE
23	right small toe	N/A
24	right heel	RHEE

Table 9: **Conversion to the COCO-style body layout used by LDPose.** The 17 body slots precede the eight residual slots. Missing Vicon sources remain unavailable. This 25-slot layout is different from the 25-body-slot OpenPose ordering.

Index	Body keypoint	OpenPose index	Vicon source
0	nose	0	N/A
1	left eye	16	N/A
2	right eye	15	N/A
3	left ear	18	N/A
4	right ear	17	N/A
5	left shoulder	5	LSJC
6	right shoulder	2	RSJC
7	left elbow	6	LEJC
8	right elbow	3	REJC
9	left wrist	7	mean(LWJC, LWRA, LWRB)
10	right wrist	4	mean(RWJC, RWRA, RWRB)
11	left hip	12	LHJC
12	right hip	9	RHJC
13	left knee	13	LKJC
14	right knee	10	RKJC
15	left ankle	14	LAJC
16	right ankle	11	RAJC

Residual endpoints. We retain the LDPose residual types and their ordering (Ying et al., 2025). Let S_{res} contain `residual_b`, `residual_r`, `residual_f`, and `residual_l`. The exported endpoint is

$$r_f = \text{nanmean}_{m \in S_{\text{res}}} \text{fetch}(m, f).$$

A valid subset is sufficient under this rule. Losing one marker does not necessarily exclude the entire frame. If no residual marker is valid, the endpoint is unavailable for endpoint supervision. Table 10 lists the slots. Table 11 gives the assignments for the coded lower-limb records. Dwarfism or altered body proportions alone do not activate a residual slot.

Table 10: **Residual-limb slot ordering.** The same eight residual types are used in both formats. Indices are zero-based. A slot is active only when it applies to the record and has a valid reference.

Residual index	Residual keypoint	LDPose index	OpenPose-extended index
0	above left elbow	17	25
1	above right elbow	18	26
2	below left elbow	19	27
3	below right elbow	20	28
4	above left knee	21	29
5	above right knee	22	30
6	below left knee	23	31
7	below right knee	24	32

Table 11: **Residual-slot assignments for the coded lower-limb evaluation records.** The assignments follow the recorded deficiency level. They do not define a residual endpoint for people who do not have terminal limb deficiency.

Participant	Residual slot	Skeleton anchor
AP001	below left knee	left knee
AP002	above left knee	left hip
AP003	below left knee	left knee
AP004	below left knee	left knee
AP005	above right knee	right hip

The endpoint is a point derived from the documented marker configuration. It is not a complete residual-limb surface or a direct measurement of internal anatomy, tissue condition, or socket fit. In worn conditions, a device or socket surface and an internal biological endpoint must not be equated merely because their marker names share a prefix. The reported procedure establishes the marker-derived reference, not a separately validated internal-point reconstruction. If the available marker subset changes, its centroid can shift even without movement of the full configuration. This is a possible uncertainty source whose magnitude is not separately quantified by the present results.

C.5 Coordinate Alignment, Timing, and Projection

Vicon-to-RGB transform. We estimate a rigid transform using camera centres associated between the Vicon XCP and Theia/RGB XCP calibrations:

$$(R, t) = \arg \min_{R \in SO(3),\, t} \sum_i \left\| R C_i^{\text{Vicon}} + t - C_i^{\text{RGB}} \right\|_2^2 .$$

Kabsch alignment using singular-value decomposition provides the rigid transform. A Vicon point becomes

$$P^{\text{RGB}} = R P^{\text{Vicon}} + t.$$

This conversion requires the corresponding camera centres and consistent coordinate units. The transform estimates coordinate alignment rather than participant movement or temporal synchronization.

Frame association. RGB frames are associated with Vicon frames through timestamps when available, with a frame-index fallback. Association requires the actual recording offset and indexing convention. A 100 Hz marker source cannot simply be read at the same numerical index as a 50 fps

RGB stream. Reproduction must use the recorded association rather than infer it from the nominal rates. The actual evaluated frame selection can also differ from the full captured frame count. No claim of framewise evaluation at 100 Hz follows from the Vicon sampling rate.

Calibrated image projection. We project into the undistorted image using the calibrated intrinsics and world-to-camera extrinsics:

$$P_{\mathrm{cam}} = R_{\mathrm{wc}} P^{\mathrm{RGB}} + t_{\mathrm{wc}}, \qquad u = f_x \frac{X_{\mathrm{cam}}}{Z_{\mathrm{cam}}} + c_x, \quad v = f_y \frac{Y_{\mathrm{cam}}}{Z_{\mathrm{cam}}} + c_y.$$

The documented export accepts a valid source point whose projected coordinates are finite, pass the conversion's zero-value check, and lie inside the undistorted image. Let $a_{f,k,c}$ denote this projection-availability test for frame f, keypoint k, and camera c. The stored triplet is

$$(x, y, c_{\mathrm{ann}}) = \begin{cases} (u, v, 2), & a_{f,k,c} = 1, \\ (0, 0, 0), & a_{f,k,c} = 0. \end{cases}$$

Here $c_{\mathrm{ann}} = 2$ denotes an available projected reference under this export rule. It does not certify unobstructed visual evidence. Occlusion by the body, support handle, or equipment is not resolved by an image-bounds check. Conversely, a zero triplet is a missing-reference encoding and is not a target at the image origin. This interpretation preserves the recorded conversion rule without substituting a different visibility detector.

Left-right consistency. Conversion does not flip left and right. Right-side sources such as `RSJC`, `RHJC`, `RKJC`, and `RAJC` map to right-side slots. Left-side sources map to left-side slots. Image coordinates use the top-left origin of the undistorted image. Any later crop, resize, or model augmentation needs its corresponding inverse mapping before camera-space evaluation.

D Mesh Reference Construction and Diagnostics

D.1 Base Initialization and Reference Branches

The construction converts image-based body estimates into Vicon-constrained fitted references. It does not introduce a new image-to-mesh architecture. A compatible full-body reconstruction supplies the base estimate. HSMR (Xia et al., 2025) provides the default initialization using SKEL (Keller et al., 2023). PromptHMR (Wang et al., 2025) provides the alternative base used in the initialization analysis. ResiHMR (Ying et al., 2026) performs residual-topology conversion after the base reconstruction rather than acting as an unrelated stand-alone initializer.

Kinematic reference. The kinematic branch retains the standard full-body representation. Its joint targets include prosthesis-supported functional landmarks where applicable. A standard lower-limb chain provides a representation of prosthesis-supported whole-body motion, not a physical model of the device surface, articulation, or material deformation.

Anatomical reference. The anatomical branch applies ResiHMR residual anchor and factor optimization and residual-limb reconstruction to the base estimate. It represents the remaining body and residual limb while excluding the missing distal segment and prosthetic component. The default conversion retains the initial camera and body-shape parameters and modifies the affected limb's pose, geometry, and kinematic chain. The two branches consequently differ before their final optimization. An anatomical reference is not defined merely by whether a prosthesis was removed during capture. The same worn record can supply constraints relevant to both reference types.

D.2 Coordinate Alignment and Constrained Refinement

For each camera and frame, the construction reads the initial pose, shape, camera translation, predicted joints, and mesh hypothesis. It transforms camera-space predictions into Vicon world coordinates using calibrated RGB camera parameters and the rigid transform between the RGB and Vicon systems. Where multiple camera hypotheses exist for one frame, it rigidly aligns each candidate to the Vicon body keypoints and selects the smallest joint-alignment error for initialization. This Vicon-guided

selection belongs to reference construction. It is not an additional RGB-only inference capability of the evaluated baseline.

The default SKEL-based refinement optimizes pose, shape, and translation around the selected estimate. The branch objectives are

$$\mathcal{L}_{\text{anat}} = \mathcal{L}_{\text{joint}} + \lambda_{\text{res}}\mathcal{L}_{\text{res}} + \lambda_\theta\mathcal{L}_\theta + \lambda_\beta\mathcal{L}_\beta + \lambda_t\mathcal{L}_t, \tag{1}$$

$$\mathcal{L}_{\text{kin}} = \mathcal{L}_{\text{joint}} + \lambda_\theta\mathcal{L}_\theta + \lambda_\beta\mathcal{L}_\beta + \lambda_t\mathcal{L}_t. \tag{2}$$

The joint loss uses the reference landmarks appropriate to the branch. The anatomical branch additionally constrains a residual endpoint when a valid marker-derived reference exists. The kinematic branch uses applicable prosthesis-supported landmarks and no endpoint loss. The remaining terms regularize pose, shape, and translation relative to the branch initialization. The loss weights, optimizer, stopping rule, and model-specific parameter conversion belong to the fitting configuration. The objectives alone do not specify their numerical settings.

The branches use the same frame-matched Vicon evidence and are refined separately. There is no explicit final cross-branch consistency loss. Sharing the initialization does not require identical optimized shape, pose, or surface coordinates. In particular, direct correspondence is not imposed on affected distal regions that have different semantics and topology. Meshes are exported in Vicon world coordinates in millimeters. Associated metadata record the source hypothesis, transforms, optimized parameters, alignment errors, and refinement status. Qualitative fitted-reference examples are shown in Figure 6 and Figure 7

D.3 Fitting Residuals and Alternative Initializations

Table 12 reports residuals against the Vicon constraints for both base initializers and reference branches. Endpoint error is not applicable to a kinematic mesh that does not represent the endpoint. The reported global standard deviations describe error dispersion. They are not confidence intervals or standard deviations across independent training runs. These measurements assess agreement with constraints used in the fit. They do not independently validate unobserved body surfaces.

Table 12: **Fitting residuals against Vicon-derived constraints in millimeters.** Values are the reported global mean $\pm$ standard deviation. These standard deviations are not participant confidence intervals. Anatomical branches apply ResiHMR conversion after the indicated base initializer. Endpoint error is not applicable to the kinematic reference.

Base initializer	Branch	Global MPJPE	Endpoint error
HSMR	Kinematic	34.327 ± 34.418	N/A
HSMR	Anatomical	48.268 ± 61.373	16.922 ± 32.524
PromptHMR	Kinematic	31.263 ± 19.453	N/A
PromptHMR	Anatomical	31.205 ± 19.330	85.183 ± 42.726

The kinematic global MPJPE differs by 3.064 mm between the two initializations. For the anatomical branch, the PromptHMR-based setting reduces global MPJPE by 17.063 mm but increases endpoint error by 68.261 mm relative to the HSMR-based setting. A better fit to the ordinary joints therefore need not yield a better residual endpoint. This result supports keeping these constraint residuals separate and shows that Vicon refinement does not remove all dependence on the initial body estimate.

D.4 Consistency Between Reference Branches

We compare the default anatomical and kinematic references on deduplicated paired frames. Shared regions include the pelvis, torso, upper limbs, contralateral intact limb, and anatomically shared proximal affected-limb joints. The affected distal region is excluded from the direct shared-region comparison because the branches represent different structures there. Table 13 reports world-coordinate joint discrepancies, shared-surface Chamfer distances, and residuals against Vicon constraints.

The participant-macro discrepancies are 20.787 mm for shared joints and 14.882 mm for shared surfaces. These are measured differences between the fitted branches, not evidence that the branches are identical. Shared-surface Chamfer distance is a set-distance diagnostic rather than the corresponding-vertex MVE used for baseline evaluation. Its region selection, point sampling, and averaging

Table 13: **Consistency of the anatomical and kinematic references.** Results use deduplicated paired frames from the default HSMR-based construction. Distances are in millimeters. Participant-macro rows give equal weight to participants. The affected distal region with different topology is not forced into correspondence.

Diagnostic	Mean ± SD
Shared-joint discrepancy in world coordinates	21.085 ± 10.323
Shared-joint discrepancy, participant macro	20.787 ± 2.106
Shared-surface Chamfer distance	14.986 ± 5.164
Shared-surface Chamfer distance, participant macro	14.882 ± 1.201
Anatomical branch to Vicon shared-joint residual	37.201 ± 9.884
Kinematic branch to Vicon shared-joint residual	31.861 ± 9.897
Anatomical residual-endpoint error	16.922 ± 32.524

convention are part of the diagnostic configuration. The global and participant-macro rows have different aggregation units and should not be substituted for one another.

D.5 Reference Sensitivity and Model Overlap

Fixed predictions with kinematic references. Table 14 holds predicted meshes fixed within each row and changes the base initializer used to construct the kinematic reference. HSMR is included as an initializer-overlap diagnostic because it supplies the default reference initialization. It is not added to the main-table baseline set.

Table 14: **Kinematic-reference sensitivity of MVE in millimeters.** Predictions are fixed within each row. The columns identify the base initializer used to construct the reference. Δ is HSMR-reference MVE minus PromptHMR-reference MVE. HSMR† is an initializer-overlap diagnostic rather than an additional main-table baseline.

Predicted mesh	PromptHMR reference	HSMR reference	Δ
PromptHMR	132.507	107.440	-25.067
HSMR†	136.670	135.970	-0.700
TokenHMR	215.141	177.810	-37.331
CameraHMR	245.921	198.530	-47.391

Changing the reference shifts MVE by 0.700 to 47.391 mm while preserving the ordering of these full-body predictions under the two tested references. The overlap conditions do not show a consistent same-initializer benefit. PromptHMR has higher error against its own initialized reference, while the HSMR prediction changes by only 0.700 mm. This is evidence about the tested references and predictions. It does not establish independence from body-model priors or rule out all reference compatibility effects.

Anatomical configurations. Both anatomical constructions apply ResiHMR after the selected base reconstruction. Table 15 contains the available same-base configurations. The base pipeline and its fitted reference change together across the two rows. Their difference therefore does not isolate a reference-only effect. In particular, changing HSMR to PromptHMR does not remove the common ResiHMR conversion stage.

Table 15: **Same-base-initializer anatomical configurations.** Both prediction and reference change across rows. Each uses ResiHMR for residual-topology conversion. This is not a fixed-prediction reference swap. MVE is in millimeters.

Predicted pipeline	Reference construction	MVE
HSMR + ResiHMR	HSMR + ResiHMR	101.470
PromptHMR + ResiHMR	PromptHMR + ResiHMR	195.245

A fixed-prediction anatomical comparison would require evaluating each saved prediction against both anatomical references under a compatible target mapping. The two reported diagonal configurations do not supply those cross evaluations. We do not infer their unreported values or describe the available comparison as a complete crossed sensitivity experiment.

Interpretation of baseline mesh scores. TokenHMR, CameraHMR, and PromptHMR use the kinematic reference in the main experiment. ResiHMR uses anatomically present joints and residual-topology vertices. The latter protocol changes the evaluated targets and shares a topology-conversion method with its reference. Its score is therefore a topology-matched diagnostic rather than a cross-topology winner. Vicon-derived joint and endpoint metrics do not use fitted surface vertices as references. They still depend on marker validity, point definitions, and alignment. MVE provides complementary agreement with the corresponding fitted surface, not direct anatomical or prosthesis-surface accuracy.

E Benchmark Implementation Details

E.1 Task Configuration and Reproduction Requirements

Benchmark design. The evaluation uses four target categories within LDBody: 2D keypoints, multi-view 3D keypoints, residual endpoints, and fitted human body surfaces. Residual localization is reported separately even when the endpoint belongs to an extended pose output. The pose baselines are YOLOPose-l (Maji et al., 2022), RTMPose-l (Jiang et al., 2023), and ViTPose-l (Xu et al., 2022). Human mesh baselines are TokenHMR (Dwivedi et al., 2024), CameraHMR (Patel & Black, 2025), PromptHMR (Wang et al., 2025), and ResiHMR (Ying et al., 2026). Baseline names denote the evaluated configurations, not a claim that all methods have identical backbones, supervision, or computational budgets.

Image fine-tuning and evaluation selection. Fine-tuning uses the designated training images with added prosthesis annotations. The calibrated motion records are not used to fine-tune these pose models. Image identifiers define the image training and evaluation partitions. Motion records require participant, trial, camera, and frame identifiers. The two record types share an annotation framework but are not assumed to contain paired people or paired motions. The comparison examines behavior across visual conditions within one dataset. It is not an experiment that varies only annotation type while holding the subjects and scenes constant.

An image-ID split alone does not establish person-disjoint or source-disjoint evaluation. Likewise, unused motion training data do not prove that no image of a participant was seen during external pretraining. Claims about identity separation, near-duplicate removal, and source leakage require the corresponding records or audit. The reported protocol should not be interpreted as providing those stronger guarantees solely from the use of distinct image IDs. All synchronized views of one test are related observations. Randomly treating its frames or views as independent participants would overstate the evidence.

Input and coordinate handling. Model input crops, resize transforms, detection boxes, confidence outputs, and predicted keypoints must be mapped back to the undistorted RGB coordinates before calibrated triangulation. The geometric association uses the same physical time across cameras. Nominal capture rates do not specify the actual evaluation stride. For multi-view reconstruction, the selected views, confidence thresholds, minimum valid-view count, triangulation routine, and failure tests affect both conditional error and prediction coverage. The phrase "calibrated triangulation" does not uniquely specify these settings.

Model and reference configurations. An exact run is identified by the model implementation and checkpoint, inherited pretraining, input resolution, augmentation, output-keypoint mapping, fine-tuning schedule, optimizer, and random seed. A metric also requires the reference version, record list, confidence and validity rules, coordinate alignment, and aggregation order. Mesh evaluation additionally requires the body-model version, joint regressor, vertex correspondence, and reference branch. These fields should be read from the corresponding archived run configuration. The table

values do not determine missing hyperparameters, and a method citation does not replace a checkpoint identifier.

E.2 Evaluation Metrics

Reference validity and prediction validity. A reference can be geometrically available even when the landmark is visually occluded. In manually annotated images, a reliably inferred occluded landmark has $v = 1$ and remains eligible under the $v > 0$ OKS rule. An unavailable reference has $v = 0$. In calibrated projections, the value 2 follows the projection-availability convention in Appendix C.5. These encodings must not be described collectively as "only visible joints." An absent anatomical segment, an unobserved existing joint, and a failed model prediction are different cases.

A reference mask identifies eligible targets. A prediction-validity mask identifies targets for which the model and reconstruction pipeline returned a scoreable result. Discarding a missing reference avoids assigning an invented target. Discarding a failed prediction changes the population on which a conditional localization error is computed. Such failures cannot be treated as correct predictions or silently used to improve the average. Reported unavailable outputs are denoted N/A, not zero error.

2D keypoint metrics. We use the LDPose-style schema with 17 standard-layout body slots and eight residual slots. COCO-style average precision and average recall use object keypoint similarity (OKS):

$$\mathrm{OKS} = \frac{\sum_i \exp\left(-\frac{\|\hat{\mathbf{x}}_i - \mathbf{x}_i\|_2^2}{2s^2\kappa_i^2}\right)\mathbf{1}[v_i > 0]}{\sum_i \mathbf{1}[v_i > 0]}.$$

Here s is the object scale and κ_i is the keypoint-specific normalization factor. The evaluator configuration determines object scale, the complete normalization vector including residual slots, matching, detection limits, and threshold averaging. The usual body-keypoint vector does not by itself define normalization for eight additional endpoints. An instance without any eligible reference keypoint does not obtain a perfect OKS from an empty denominator. Its treatment follows the evaluator's instance-handling rule.

The main 2D table expresses AP and AR on a 0 to 100 scale. The diagnostic tables in Appendix F also use the 0 to 100 scale. Values originally reported on a 0 to 1 scale are multiplied by 100, including standard deviations, ranges, and AP differences. This changes the display scale rather than the underlying evaluation. AP is a detection and localization metric, not an average coordinate error. The supplemental 2D MPJPE is a pixel-distance summary and has different units and matching implications. Overall AP also does not isolate the accuracy of body joints, residual endpoints, or device landmarks individually.

3D joint error. For a declared target set $\mathcal{K}$ of valid evaluated keypoints, MPJPE in millimeters is

$$\mathrm{MPJPE} = \frac{1}{|\mathcal{K}|}\sum_{i\in\mathcal{K}} \|\hat{\mathbf{X}}_i - \mathbf{X}_i\|_2.$$

Predictions and references must use the same coordinate convention and physical units. A world-coordinate error, a root-relative error, and an error after a rigid alignment are not interchangeable merely because each is called MPJPE. The declared alignment and target set belong to the result configuration. The direct world-coordinate comparisons use Vicon coordinates, while any reported alignment operation must be identified separately.

PA-MPJPE applies a similarity Procrustes alignment to the selected target set before averaging distances. This alignment fits rotation, translation, and scale. It can reduce global position or scale errors while leaving other localization errors. A large MPJPE with a much smaller PA-MPJPE therefore does not establish that the original reconstruction was geometrically valid in the calibrated world. The same target mask must be used to define and score a matched alignment comparison.

Standard, extended, and shared targets. Non-fine-tuned pose models do not predict residual endpoints. Fine-tuned configurations use an extended target set. Their main-table aggregate MPJPE is not a standard-joint-only ablation. The movement and deficiency diagnostics separately report the original "shared" target summaries. Shared targets must be understood through their actual

keypoint mapping and masks, not assumed to equal every standard schema slot. The schema includes unavailable face or toe sources, and device-supported locations are functional targets rather than distal biological anatomy.

Residual-endpoint error and coverage. For scoreable endpoint observations $\mathcal{R}$,

$$\mathrm{MPJPE}_{\mathrm{ResidualEnd}} = \frac{1}{|\mathcal{R}|} \sum_{r \in \mathcal{R}} \|\hat{\mathbf{R}}_r - \mathbf{R}_r\|_2.$$

The endpoint reference is the marker-derived point or image annotation defined in Appendices C.3 and C.4, not complete residual shape. A model without endpoint outputs is not evaluated on this metric. To distinguish conditional accuracy from reconstruction success, let $\mathcal{E}$ be eligible observations with valid endpoint references, and let $\mathcal{Q}$ be those with valid predictions. Prediction coverage among eligible observations is

$$C_{\mathrm{endpoint}} = 100 \frac{|\mathcal{E} \cap \mathcal{Q}|}{|\mathcal{E}|} \,\%.$$

The denominator, minimum-view rule, and missing-value tests must be fixed for reproduction. Reference availability across all records is a separate quantity. Participant-macro coverage averages participant percentages rather than pooling every valid endpoint. The reported endpoint error is interpreted jointly with this coverage.

Human mesh metrics. Human mesh recovery uses Vicon-derived joint or endpoint references for point errors and the appropriate fitted reference for MVE:

$$\mathrm{MVE} = \frac{1}{|\mathcal{V}|} \sum_{j \in \mathcal{V}} \|\hat{\mathbf{V}}_j - \mathbf{V}_j\|_2.$$

The set $\mathcal{V}$ contains corresponding vertices under the declared topology and mapping. Equal vertex counts alone do not prove semantic correspondence. Nearest-surface distance or Chamfer distance is not interchangeable with this corresponding-vertex metric. A result configuration must identify any model conversion, vertex subset, units, and alignment applied before comparison. The anatomical and kinematic protocols have different target sets and are not combined into one surface ranking.

Vicon keypoints supply a reference that does not use fitted vertices. This independence from the fitted surface does not make the point reference error-free or prove an independent anatomical surface. Conversely, low fitting residuals against the same points used in optimization are not a second independent test. These distinctions separate target accuracy, reference agreement, and reference-generation quality.

E.3 Aggregation and Uncertainty

For participant p and test t, let $e_{p,t}$ be the declared test-level error after within-test aggregation. Participant-balanced summaries are

$$m_p = \frac{1}{|\mathcal{T}_p|} \sum_{t \in \mathcal{T}_p} e_{p,t}, \qquad \bar{m} = \frac{1}{P} \sum_{p=1}^{P} m_p.$$

The sample standard deviation and the mean after removing participant p are

$$s_m = \sqrt{\frac{1}{P-1} \sum_{p=1}^{P} (m_p - \bar{m})^2}, \qquad \bar{m}_{-p} = \frac{1}{P-1} \sum_{q \neq p} m_q.$$

The reported delete-one range is the minimum and maximum of these P means. Predictions remain fixed. This is evaluation sensitivity, not retraining-based cross-validation. The participants in Table 6 are the units for the corresponding analyses. Multiple views, frames, or trials do not increase P.

A pooled score can weight participants by the number of frames, valid observations, or detections. A participant-macro score weights their summaries equally. For AP, pooled detection evaluation and averaging participant AP are also nonlinear alternatives. They must be labeled separately. For

a paired condition comparison, the standard deviation of the within-participant differences is not obtained by subtracting the two marginal standard deviations. Global fitting residual dispersion, participant dispersion, and variability across training seeds likewise describe different quantities. No confidence interval, hypothesis test, or multi-seed result is inferred from a reported standard deviation or delete-one range.

E.4 Compute Resources

The reported experiments used a local workstation with two NVIDIA RTX 4090 GPUs, each with 24 GB memory, and 64 GB system memory. GPU workloads included inference, fine-tuning, and mesh recovery. CPU workloads included data loading, projection, metric computation, and post-processing. These hardware specifications do not specify GPU hours, peak utilization, or the number of repeated runs. Runtime and software-version claims require the corresponding execution records.

F Extended Diagnostic Evaluation

We examine participant variation, side-wise errors, prosthesis conditions, movement categories, deficiency levels and camera viewpoints. These are descriptive analyses of the motion cohort listed in Table 6. Metric conventions and the distinction between result families are given in Appendices C.1 and E.

F.1 Participant-Level Summaries and Leave-One-Out Sensitivity

Tables 16–19 report equally weighted participant summaries, sample SDs and leave-one-participant-out ranges. The summaries characterize variation within the evaluated cohort using fixed predictions.

Table 16: **Participant-balanced pose summaries.** AP uses a 0–100 scale; 3D errors are in mm. LOO denotes the range of leave-one-participant-out means. Standard/extended target coverage follows the reported configuration.

Model	2D AP mean ± SD	AP LOO range	3D MPJPE mean ± SD	MPJPE LOO range
RTMPose-l	71.89 ± 2.84	71.23–73.03	47.29 ± 3.51	46.18–48.17
RTMPose-l FT	75.89 ± 3.66	74.72–76.92	59.15 ± 6.21	56.55–60.29
ViTPose-l	74.52 ± 2.43	73.93–75.45	45.40 ± 2.61	44.52–46.23
ViTPose-l FT	85.48 ± 2.87	84.57–86.13	49.14 ± 2.66	48.33–49.98
YOLOPose-l	37.00 ± 15.58	32.93–41.50	226.68 ± 157.63	177.20–262.08
YOLOPose-l FT	24.98 ± 12.23	21.03–28.18	600.76 ± 323.70	475.23–682.69

ViTPose-l FT improves overall 2D AP for all five evaluated participants, while RTMPose-l FT improves it for four of five. YOLOPose-l configurations show larger participant-level dispersion. These observations concern the overall pose protocol; they do not establish target-specific benefits for prosthesis landmarks or matched shared-joint comparisons.

Table 17: **Residual-endpoint localization and reconstruction coverage.** Error and LOO range are in mm; coverage is percent. All summaries give participants equal weight.

Model	Endpoint MPJPE mean ± SD	LOO range	Coverage (%) mean ± SD
RTMPose-l FT	126.04 ± 72.86	103.68–143.35	100.00 ± 0.00
ViTPose-l FT	99.79 ± 57.76	82.01–112.89	99.22 ± 1.74
YOLOPose-l FT	394.89 ± 52.92	379.14–410.41	35.62 ± 17.02

Endpoint error varies substantially across participants. In the reported endpoint analysis, the method ordering remains unchanged in every leave-one-out summary. YOLOPose-l FT's low coverage requires its conditional MPJPE to be interpreted together with reconstruction availability, rather than as performance on all eligible endpoints.

The reported full-body MPJPE ordering remains unchanged after removing any one participant. ResiHMR is reported separately; its participant-balanced endpoint error is 155.62 ± 26.68 mm with a

Table 18: **Participant-balanced HMR MPJPE (mm).** Residual topology is not ranked against the full-body protocol.

Method	Protocol	Mean ± SD	LOO range
CameraHMR	Full-body	225.44 ± 29.27	212.73–230.70
PromptHMR	Full-body	143.90 ± 28.31	131.26–147.45
TokenHMR	Full-body	147.19 ± 21.59	137.64–150.46
ResiHMR	Residual	133.62 ± 26.11	122.27–139.00

Table 19: **Participant-balanced HMR PA-MPJPE (mm).**

Method	Protocol	Mean ± SD	LOO range
CameraHMR	Full-body	119.74 ± 5.19	118.76–121.91
PromptHMR	Full-body	84.33 ± 11.41	79.74–87.20
TokenHMR	Full-body	86.78 ± 5.92	84.62–88.26
ResiHMR	Residual	75.63 ± 6.72	72.93–77.13

leave-one-out range of 147.33–163.65 mm. Participant-balanced HMR MPJPE differs from the main pooled values by 4.53 mm for CameraHMR, 4.19 mm for PromptHMR, 1.62 mm for TokenHMR and 1.70 mm for ResiHMR. These comparisons do not equate the participant-summary result suite with the condition/view suites below.

The participant analysis also identifies AP002 as more challenging across the four mesh methods. It does not isolate whether participant geometry, capture conditions, movement composition, or model behavior accounts for this difference.

F.2 Affected-Side versus Contralateral-Side Errors

The affected side is the side with lower-limb deficiency; it includes prosthesis-supported targets when a device is worn. Its contralateral side is evaluated in the same participant and recorded test, sharing capture-session conditions. For error measures, results are aggregated within test, within participant, and then equally across participants. We report $\Delta = E_{\text{aff}} - E_{\text{contra}}$, so positive error differences indicate higher affected-side error. For AP, a negative difference indicates lower affected-side AP.

The reported 2D side comparison uses the hip, knee, and ankle labels. The mesh side table also describes reconstructed joints at these named locations, not local surface-vertex error. Its standard-layout functional indices should not be interpreted as asserting that every corresponding biological joint exists. For non-worn and residual-topology cases, the exact active-joint map and missing-reference mask are required to decide which distal locations contribute. The summary tables do not establish a biological ankle or knee reference where no such anatomical target exists.

Participant-balanced aggregation. Let $e_{p,t,s}$ be the declared test-level error for participant p, test t, and side s. The side summaries are

$$E_s = \frac{1}{P} \sum_{p=1}^{P} \left(\frac{1}{|\mathcal{T}_p|} \sum_{t \in \mathcal{T}_p} e_{p,t,s} \right), \qquad P = 5, \tag{3}$$

with difference

$$\Delta = E_{\text{aff}} - E_{\text{contra}}. \tag{4}$$

Positive error differences indicate higher affected-side error. AP differences have the opposite interpretation because larger AP is better. The side-specific target sets, condition selection, and validity masks are part of this calculation, not quantities that can be inferred from the final side means.

2D pose. The reported side-specific AP uses COCO-style OKS matching for the hip, knee and ankle landmarks on each side. Table 20 reports lower affected-side AP in all six configurations. This does not imply that the side difference holds for every individual participant separately.

Table 20: **Side-specific 2D AP.** AP and differences use a 0–100 scale. FT denotes image-based fine-tuning.

Model	Affected AP	Contralateral AP	Δ
RTMPose-l	79.99	89.48	-9.49
RTMPose-l FT	70.63	78.78	-8.15
ViTPose-l	84.41	92.94	-8.53
ViTPose-l FT	82.20	88.29	-6.09
YOLOPose-l	51.43	64.04	-12.61
YOLOPose-l FT	32.38	34.07	-1.69

Table 21: **Side-wise multi-view 3D MPJPE (mm).** W and N denote prosthesis-worn and non-worn tests; A and C denote affected and contralateral sides. Reported differences are preserved at source precision.

Model	W: A	W: C	W: Δ	N: A	N: C	N: Δ
RTMPose-l	59.35	38.77	$+20.58$	46.76	41.23	$+5.53$
RTMPose-l FT	79.54	52.66	$+26.88$	58.94	46.32	$+12.62$
ViTPose-l	54.23	36.39	$+17.83$	50.35	43.07	$+7.28$
ViTPose-l FT	56.18	44.47	$+11.71$	50.18	45.20	$+4.98$
YOLOPose-l	217.13	171.78	$+45.35$	211.18	178.53	$+32.66$
YOLOPose-l FT	650.62	903.02	-252.40	304.35	315.32	-10.97

Multi-view 3D pose. Five of six configurations have higher affected-side MPJPE under both prosthesis conditions. For these configurations, the differences range from 11.71 to 45.35 mm in worn tests and from 4.98 to 32.66 mm in non-worn tests. Fine-tuned YOLOPose-l has differences of -252.40 and -10.97 mm and is retained. The reported analysis associates its instability with participant-dependent outliers; the reversed differences are not interpreted as a protective effect of the affected side.

Table 22: **Side-wise reconstructed-joint errors for HMR outputs.** Entries are affected / contralateral participant-balanced MPJPE in mm, not regional vertex errors.

Method	Hip	Knee	Ankle
CameraHMR	86.0 / 81.3	260.2 / 236.4	371.9 / 325.7
PromptHMR	99.2 / 89.0	165.8 / 140.1	252.7 / 210.9
ResiHMR	60.9 / 55.6	143.7 / 148.3	220.3 / 184.8
TokenHMR	57.8 / 53.9	160.6 / 135.2	225.0 / 196.9

Lower-limb HMR landmarks. Every method has higher affected-side ankle error, by 28.1–46.2 mm, while hip differences are 3.9–10.2 mm. CameraHMR, PromptHMR and TokenHMR have 23.8–25.7 mm higher affected-side knee error; ResiHMR has 4.6 mm lower affected-side knee error. Among the evaluated hip, knee and ankle landmarks, the largest side-wise difference is at the ankle for each method. This is not a measurement of error concentration across every body region or mesh surface.

Scope. A shared recording does not make the two sides identical in visibility, motion or landmark meaning. Contralateral compensations can occur. The comparison is a descriptive within-participant diagnostic, not a causal estimate of limb deficiency or a substitute for separately recruited matched limb-complete controls.

F.3 Prosthesis-Worn and Non-Worn Conditions

The recorded subset contains 120 worn and 16 non-worn tests. Condition-specific results are participant-aware but are not interpreted as causal effects of wearing a prosthesis: activity composition, residual visibility and valid reconstruction coverage differ. In this subsection, Δ denotes non-worn minus worn, reversing the side-wise definition used above.

2D pose. Table 23 gives the reported AP and pixel-space MPJPE. All six configurations have higher AP on the worn subset, by 3.2–6.4 points, whereas pixel-error changes are method-dependent.

Table 23: **2D pose by prosthesis condition.** AP uses a 0–100 scale; positional errors are pixels. W/N denote worn/non-worn; differences retain the reported source precision.

Model	AP W	AP N	Δ AP	px W	px N	Δ px
RTMPose-l	72.8	67.5	−5.3	24.27	25.51	+1.24
RTMPose-l FT	76.6	71.1	−5.5	32.47	31.64	−0.82
ViTPose-l	75.5	69.1	−6.4	21.67	24.55	+2.89
ViTPose-l FT	86.2	80.9	−5.2	24.78	26.48	+1.70
YOLOPose-l	37.4	34.1	−3.2	59.64	62.26	+2.62
YOLOPose-l FT	25.4	21.1	−4.3	90.39	75.91	−14.48

Table 24: **3D pose by prosthesis condition (mm).** M denotes MPJPE, PA denotes PA-MPJPE. † flags the raw failed-trial-sensitive result, retained rather than replaced.

Model	M W	M N	Δ M	PA W	PA N	Δ PA
RTMPose-l	47.78	42.68	−5.10	43.84	38.71	−5.13
RTMPose-l FT	58.94	50.82	−8.12	53.28	46.50	−6.78
ViTPose-l	45.06	45.59	+0.53	41.56	41.55	−0.01
ViTPose-l FT	1,145.04†	44.32	−1,100.72†	46.28	41.62	−4.65
YOLOPose-l	221.18	227.77	+6.58	98.07	77.95	−20.12
YOLOPose-l FT	634.34	322.87	−311.46	132.76	106.08	−26.68

Failed trial. The raw ViTPose-l FT worn-condition MPJPE is dominated by the test `AP001/05_Bent_Over_Twist`. The reported test-level MPJPE is 104,106.16 mm, while PA-MPJPE is 86.47 mm. Table 25 retains the raw condition aggregate and reports the removal of that test only as a sensitivity analysis. It does not replace the primary raw condition result or establish a prospective failure-rejection rule. A low aligned error does not negate the severe world-coordinate failure.

Table 25: **ViTPose-l FT failed-trial sensitivity (mm).** Only the explicitly named trial is removed in the second row.

Condition summary	Worn	Non-worn	Δ
Raw reported MPJPE	1,145.04	44.32	−1,100.72
Excluding identified failed trial	49.72	44.32	−5.40

HMR condition comparisons. The HMR condition analysis weights trials equally within each participant/condition, then weights participants equally. The difference column reports paired within-participant non-worn minus worn differences and their sample SDs; its SD is not obtained by subtracting the two marginal SDs.

PromptHMR, ResiHMR and TokenHMR have lower MPJPE in the non-worn subset, whereas CameraHMR changes in the opposite direction. PA-MPJPE is lower in the non-worn subset for all four methods, with substantial participant-level variation. These table-specific patterns do not establish that either condition is intrinsically easier.

F.4 Movement-Category Diagnostics

The four categories and their coverage are given in Table 7. Every category includes all evaluated participants. The following tables preserve the reported category-specific summaries. They are not independent participant cohorts.

Pose. Table 27 separates overall 2D AP, shared-keypoint 2D AP, shared-keypoint 3D MPJPE and residual-endpoint error. S, G, T and R abbreviate static/postural, gait/locomotion, transitional/functional and running/sport dynamic.

Table 26: **HMR by prosthesis condition (mm).** Entries are participant means ± SD; Δ is the paired non-worn minus worn difference. ResiHMR remains a separate residual-topology protocol.

Method	Metric	Worn	Non-worn	Paired Δ
CameraHMR	MPJPE	176.24 ± 24.79	180.07 ± 64.93	$+3.83 \pm 49.41$
CameraHMR	PA-MPJPE	95.87 ± 5.96	91.96 ± 21.69	-3.91 ± 18.56
PromptHMR	MPJPE	143.53 ± 23.04	125.17 ± 32.59	-18.36 ± 11.35
PromptHMR	PA-MPJPE	81.24 ± 9.13	69.88 ± 17.68	-11.36 ± 11.05
ResiHMR	MPJPE	135.32 ± 26.40	122.04 ± 30.56	-13.28 ± 5.13
ResiHMR	PA-MPJPE	75.98 ± 6.39	71.92 ± 13.49	-4.06 ± 9.72
ResiHMR	Endpoint	156.70 ± 28.75	150.32 ± 18.59	-6.37 ± 23.54
TokenHMR	MPJPE	131.44 ± 26.01	121.70 ± 31.92	-9.74 ± 6.77
TokenHMR	PA-MPJPE	71.19 ± 5.08	68.49 ± 10.93	-2.70 ± 7.48

Table 27: **Participant-macro pose results by movement category.** AP is on a 0–100 scale; errors are mm. Shared denotes the diagnostic common-keypoint set. N/A denotes no endpoint prediction.

Model	Cat.	Overall AP	Shared AP	Shared 3D MPJPE	Endpoint MPJPE
RTMPose-l	S	65.1	87.3	34.8	N/A
RTMPose-l	G	74.8	90.2	35.6	N/A
RTMPose-l	T	73.0	88.8	35.3	N/A
RTMPose-l	R	74.8	90.7	35.3	N/A
RTMPose-l FT	S	72.3	73.9	46.4	126.3
RTMPose-l FT	G	76.9	81.1	51.3	124.8
RTMPose-l FT	T	75.7	79.9	47.5	128.6
RTMPose-l FT	R	78.3	82.7	45.2	126.5
ViTPose-l	S	68.4	91.7	35.2	N/A
ViTPose-l	G	77.1	93.0	33.4	N/A
ViTPose-l	T	75.6	91.4	34.3	N/A
ViTPose-l	R	76.9	93.3	33.6	N/A
ViTPose-l FT	S	83.9	85.3	39.1	91.9
ViTPose-l FT	G	86.8	89.3	42.5	103.2
ViTPose-l FT	T	84.9	87.5	41.0	105.5
ViTPose-l FT	R	86.6	89.9	39.8	94.8
YOLOPose-l	S	29.4	15.0	358.58	N/A
YOLOPose-l	G	38.9	49.1	189.47	N/A
YOLOPose-l	T	40.0	51.9	198.19	N/A
YOLOPose-l	R	39.4	50.8	146.11	N/A
YOLOPose-l FT	S	21.0	22.8	833.27	363.4
YOLOPose-l FT	G	28.7	30.7	200.79	368.3
YOLOPose-l FT	T	24.7	26.9	919.32	379.8
YOLOPose-l FT	R	25.9	27.6	346.48	348.0

The reported non-FT RTMPose-l shared-keypoint MPJPE ranges from 34.8 to 35.6 mm and non-FT ViTPose-l from 33.4 to 35.2 mm across categories. ViTPose-l FT endpoint error is 91.9–105.5 mm, compared with 124.8–128.6 mm for RTMPose-l FT. YOLOPose-l FT has greater category variation; endpoint coverage ranges from 21.6 to 38.4% across categories. This range does not specify the coverage for each individual category.

The displayed shared-keypoint AP and 3D error do not uniformly follow overall AP changes after fine-tuning. This motivates target-separated analysis but does not by itself identify an intrinsic training trade-off or isolate the effect of adding prosthesis annotations.

Table 28: **HMR movement-category results (mm).** Entries are participant means ± SD. Category coverage is in Table 7; ResiHMR has a separate topology protocol.

Category	Method	MPJPE	PA-MPJPE
S	CameraHMR	163.61 ± 29.21	89.97 ± 9.74
S	PromptHMR	135.81 ± 32.17	77.65 ± 13.08
S	ResiHMR	132.01 ± 24.09	75.44 ± 7.83
S	TokenHMR	124.98 ± 27.65	68.40 ± 7.27
G	CameraHMR	192.78 ± 22.35	99.23 ± 4.69
G	PromptHMR	144.04 ± 25.06	81.96 ± 9.03
G	ResiHMR	136.64 ± 27.58	74.89 ± 5.60
G	TokenHMR	131.99 ± 26.42	68.62 ± 3.50
T	CameraHMR	184.27 ± 33.48	100.27 ± 10.76
T	PromptHMR	150.79 ± 20.10	85.98 ± 9.14
T	ResiHMR	137.88 ± 24.02	79.67 ± 6.19
T	TokenHMR	133.68 ± 24.94	75.40 ± 6.12
R	CameraHMR	162.31 ± 45.09	90.29 ± 9.87
R	PromptHMR	131.19 ± 23.63	72.15 ± 8.81
R	ResiHMR	130.82 ± 35.89	71.51 ± 11.17
R	TokenHMR	131.26 ± 30.74	68.91 ± 8.03

Within this reported result suite, transitional/functional movements have the highest PA-MPJPE for each method. Global MPJPE is not uniformly largest in the same category. No direct cross-topology ordering is made between ResiHMR and the full-body methods.

F.5 Descriptive Deficiency-Level Results

The transtibial subset contains three participants, 82 tests, 656 videos and 530,928 RGB frames. The transfemoral subset contains two participants, 54 tests, 432 videos and 432,176 frames. Identity, side, device configuration and movement distribution are confounded with deficiency level in this small subset. The following are descriptive subgroup summaries, not general deficiency-level effects.

Table 29: **Pose results by deficiency level.** TT/TF denote transtibial/transfemoral; parentheses give participants. AP is 0–100 and errors are mm.

Model	Level (N)	Overall AP	Shared AP	Shared 3D MPJPE	Endpoint MPJPE
RTMPose-l	TT (3)	72.9	90.7	35.2	N/A
RTMPose-l	TF (2)	70.5	86.9	35.3	N/A
RTMPose-l FT	TT (3)	78.0	81.6	47.5	107.1
RTMPose-l FT	TF (2)	72.6	76.8	47.2	156.7
ViTPose-l	TT (3)	75.2	93.8	33.2	N/A
ViTPose-l	TF (2)	73.5	89.7	36.0	N/A
ViTPose-l FT	TT (3)	86.9	89.8	40.6	92.3
ViTPose-l FT	TF (2)	83.4	85.2	40.5	109.6

All configurations listed in Table 29 have lower overall AP in the transfemoral subgroup. Shared-keypoint 3D errors of the top-down configurations are similar across the two subgroups, while their residual-endpoint errors are higher in the transfemoral subgroup. These are observations in the recorded participants, not a general causal effect of deficiency level.

Table 30: **HMR results by deficiency level (mm).** Entries are participant means ± SD. The three-versus-two comparison remains confounded by participant-specific factors.

Method	Level	MPJPE	PA-MPJPE	Endpoint error
CameraHMR	TT	165.19 ± 13.25	96.72 ± 7.94	N/A
CameraHMR	TF	187.50 ± 47.06	92.32 ± 5.79	N/A
PromptHMR	TT	131.16 ± 5.09	78.27 ± 3.53	N/A
PromptHMR	TF	156.80 ± 38.98	83.29 ± 17.59	N/A
ResiHMR	TT	120.48 ± 8.17	73.60 ± 3.37	139.24 ± 16.41
ResiHMR	TF	154.19 ± 36.75	79.17 ± 10.78	183.00 ± 9.84
TokenHMR	TT	116.68 ± 6.89	70.21 ± 4.27	N/A
TokenHMR	TF	150.94 ± 36.23	72.53 ± 8.27	N/A

MPJPE is descriptively higher in the transfemoral subgroup for each method, whereas PA-MPJPE differences are smaller and not uniform in direction. ResiHMR endpoint error is also higher in that subgroup. This does not establish that a given deficiency level is intrinsically harder independently of the associated participants and capture conditions.

F.6 Complete Eight-Camera HMR Diagnostics

Each of the eight views covers the same motion cohort, 136 tests and 120,388 view-specific RGB frames. Camera identifiers are retained in Table 31. This is a single-view HMR breakdown over synchronized views, not a per-camera 2D pose analysis or a leave-one-camera-out triangulation experiment.

Table 31: **Camera records in the HMR view analysis.** Each row covers the same motion cohort and 136 tests.

View	USERID	Camera serial	RGB frames
1	0	55788813	120,388
2	1	55788789	120,388
3	2	55788786	120,388
4	3	60416142	120,388
5	4	55788787	120,388
6	5	55788788	120,388
7	6	60416105	120,388
8	7	60416141	120,388

Table 32: **HMR MPJPE by camera (mm).** Entries are participant means ± SD. ResiHMR retains its separate residual-topology evaluation; no cross-topology best scores are highlighted.

View	CameraHMR	PromptHMR	ResiHMR	TokenHMR
1	116.66 ± 20.91	104.12 ± 27.63	117.57 ± 27.88	108.27 ± 28.84
2	125.66 ± 23.77	100.43 ± 28.45	118.76 ± 29.29	113.94 ± 28.75
3	118.32 ± 27.16	96.54 ± 27.69	127.04 ± 25.84	112.33 ± 29.53
4	241.52 ± 43.97	199.76 ± 24.69	149.72 ± 23.88	157.17 ± 23.60
5	257.58 ± 62.25	223.45 ± 23.54	147.90 ± 30.61	145.16 ± 23.28
6	197.42 ± 43.18	152.04 ± 23.32	130.64 ± 23.22	132.69 ± 24.26
7	230.95 ± 29.97	187.56 ± 13.41	150.68 ± 19.59	148.26 ± 24.51
8	126.52 ± 24.62	110.23 ± 30.71	131.00 ± 39.98	128.51 ± 37.21

View-dependent MPJPE ranges are 116.66–257.58 mm for CameraHMR, 96.54–223.45 mm for PromptHMR, 117.57–150.68 mm for ResiHMR and 108.27–157.17 mm for TokenHMR. ResiHMR endpoint error varies from 134.35 to 178.36 mm. These describe within-method view sensitivity under each method's protocol.

Among the full-body methods in this diagnostic suite, PromptHMR has lower MPJPE than TokenHMR on Views 1–3 and 8, while TokenHMR has lower MPJPE on Views 4–7. This is a

Table 33: **HMR PA-MPJPE by camera (mm).** Conventions match Table 32.

View	CameraHMR	PromptHMR	ResiHMR	TokenHMR
1	68.34 ± 3.63	59.26 ± 9.02	65.11 ± 6.76	61.80 ± 5.75
2	75.84 ± 4.24	57.32 ± 8.60	65.72 ± 9.93	61.69 ± 7.65
3	71.50 ± 6.09	56.48 ± 9.90	73.43 ± 10.38	60.47 ± 8.74
4	120.40 ± 15.15	117.76 ± 10.20	85.13 ± 5.43	85.30 ± 5.20
5	133.49 ± 11.43	122.99 ± 11.43	85.63 ± 7.45	82.59 ± 3.37
6	108.11 ± 16.55	91.40 ± 10.40	76.41 ± 7.34	75.46 ± 7.37
7	115.32 ± 9.00	101.51 ± 6.20	83.93 ± 4.44	83.48 ± 6.26
8	76.42 ± 9.48	59.94 ± 10.21	72.09 ± 10.00	59.29 ± 8.82

Table 34: **ResiHMR residual-endpoint error by view (mm).** Entries are participant means $\pm$ SD.

View	Endpoint MPJPE
1	144.78 ± 32.98
2	134.35 ± 31.10
3	141.32 ± 26.66
4	174.46 ± 25.69
5	175.95 ± 25.25
6	154.04 ± 34.21
7	178.36 ± 48.54
8	150.77 ± 36.59

within-protocol comparison; ResiHMR is not included in a cross-topology ranking. Camera views are repeated observations, not additional independent subjects. Neither this view dependence nor the qualitative failure examples alone establish a causal mechanism such as image rotation or topology mismatch.

G Additional Discussion

G.1 Contribution, Coverage, and Inferential Scope

The curation of LDBody was conducted under institutional human research ethics approval. Participants in prospective motion capture provided informed consent for the specified recording and research uses. The institution and approval identifier are omitted during double-blind review. Reuse of source LDPose images and inherited annotations also follows their original access and licensing terms.

Dataset inclusion, availability of a particular annotation, and inclusion in an experiment are different statements. The dataset covers upper- and lower-limb differences, dwarfism, and multiple device types. The calibrated records also include dwarfism. The motion cohort is enumerated in Table 6. Its limb-deficiency and device attributes do not enumerate every body-proportion attribute or image category in the dataset. Their conclusions cannot be extended automatically to every image category or body configuration. Equally, limitations of that evaluation block do not imply that the entire dataset consists only of those visual conditions.

Independent people remain the relevant unit for demographic inference. Additional synchronized views improve geometric observability, and additional tests reveal within-person variation. Neither turns one participant into several independent people. The participant and delete-one analyses in Appendix F show how some conclusions depend on the recorded individuals. They do not establish population representativeness or a distribution-free guarantee of generalization. Broader image coverage and accurately defined 3D targets address different gaps, and neither replaces the other.

G.2 Why These Annotation Targets Are Needed

An object mask can delineate a visible prosthesis, but it does not specify the functional locations expected by standard whole-body pose systems. Conversely, replacing a residual endpoint with an

ordinary distal joint can erase the body structure that is actually present. Retaining both endpoint and applicable device landmarks makes these questions separately testable. A virtual functional landmark is an operational annotation target. It is not a claim that a device has a biological joint or a particular mechanical articulation.

The value of a fitted body reference is also different from that of a point reference. Joint agreement constrains locations but does not uniquely determine the intervening surface. This motivates surface-level diagnostics when methods output meshes. The fitting residuals, branch-consistency results, and reference-swap analysis make the source of those references inspectable. They do not convert a model-derived surface into a scan. Keeping these layers explicit is more informative than either discarding all mesh evaluation or treating a fitted mesh as anatomical truth.

The remaining uncertainty is localized. Image ambiguity concerns operational landmark placement and occlusion. Marker uncertainty concerns source validity, marker-derived points, and calibration. Surface uncertainty concerns body-model assumptions, initialization, and topology conversion. The corresponding rules and measured diagnostics appear in Appendices C and D. Formal repeated-annotation agreement, independent surface measurements, and a complete crossed anatomical reference experiment are not reported as completed studies.

G.3 Measurement Gaps and Causal Questions

The need for residual-endpoint and prosthesis-function references is not conditional on all general-purpose pose metrics being worse for every person with limb deficiency. Even equal ordinary-joint error would leave a benchmark without these targets unable to evaluate them. This is the measurement motivation for the dataset. A different question asks how much error is caused by body configuration rather than activity, visibility, or capture conditions.

The affected-side comparison reduces several between-record differences by using the same participant and test. It does not eliminate side-specific visibility, compensation, or target-semantic differences. A separately recruited group without limb deficiency performing matched tasks with the same equipment would address additional attribution questions. Such a group is not supplied by the contralateral-side analysis. Condition and deficiency-level results are therefore descriptive. A larger aligned error, lower AP, or missing prediction is not itself evidence of a unique causal mechanism.

The baseline findings support three specific conclusions. Image fine-tuning produces model-dependent changes across visual conditions. Residual localization and affected distal landmarks retain errors that overall pose metrics can obscure. Joint errors and fitted-surface errors can differ substantially in their ordering or magnitude. These findings do not establish that all methods fail, that one data type is intrinsically harder, or that a particular model design has been causally validated.

G.4 Real Recordings and Synthetic Data

Real recordings provide observations of body structure, device appearance, and movement occurring together. Synthetic data can add controlled variation and more training examples, including rare configurations. Its usefulness depends on the target task and realism of the generated evidence. A plausible rendered body or prosthesis does not by itself validate socket fit, device articulation, residual geometry, or movement strategy.

The present work does not compare real-only and synthetic training under a matched budget. It therefore does not establish that synthetic data are unsuitable or that real-only training is superior. The calibrated records provide a basis for evaluating future simulation and augmentation methods on real observations. Such studies should distinguish visual improvements from endpoint, joint, and surface performance and should preserve the independence of their final evaluation records.

G.5 Privacy, Consent, and Responsible Access

Prospective motion capture was conducted under institutional human research ethics approval with informed consent for the specified recording and research uses. Institution and approval identifiers are omitted in the review manuscript to preserve author anonymity. The underlying approval and consent records remain part of responsible data governance. Image reuse follows the original LDPose access and licensing terms rather than being covered by the prospective motion-capture consent.

Face mosaicing and coded identifiers reduce direct identification but do not anonymize body shape, movement, prosthesis configuration, clothing, or trial context. Derived trajectories and fitted bodies can also retain identifying information. Controlled access, storage safeguards, and a data-use agreement are consequently part of the release design. Prohibited uses include re-identification, unauthorized linkage or redistribution, surveillance, publication of identifiable participant material without permission, and unsupported clinical inference. Access and reproduction requirements are consolidated in Appendix A.

Author anonymity and participant privacy are separate. Removing author names from a PDF does not make participant records anonymous. A review package must also avoid disclosing authors through repository metadata, personal file paths, account names, or document properties. A password-protected archive is an access control, not proof that its contents have no identifiers. The review-only credential note should remain separate from the public manuscript.

G.6 Intended Research Use and Future Work

The dataset supports research toward accessible motion analysis, assistive technologies, and inclusive sport. Expert-guided annotation helps define relevant visual targets. It does not establish clinical validity, sport classification decisions, device suitability, rehabilitation progress, or improved quality of life. Endpoint and fitted-surface errors do not measure pain, tissue condition, socket fit, or clinical mobility. Responsible downstream work requires validation appropriate to those decisions.

Future work should increase independent participant coverage and improve geometric validation across body proportions, limb configurations, devices, activities, and environments. Categories already included can still require more independent examples and complete reference coverage. Repeated annotation, additional surface measurements, matched control recordings, and controlled reference comparisons would address distinct uncertainties. These are future extensions rather than experiments claimed by the current paper.

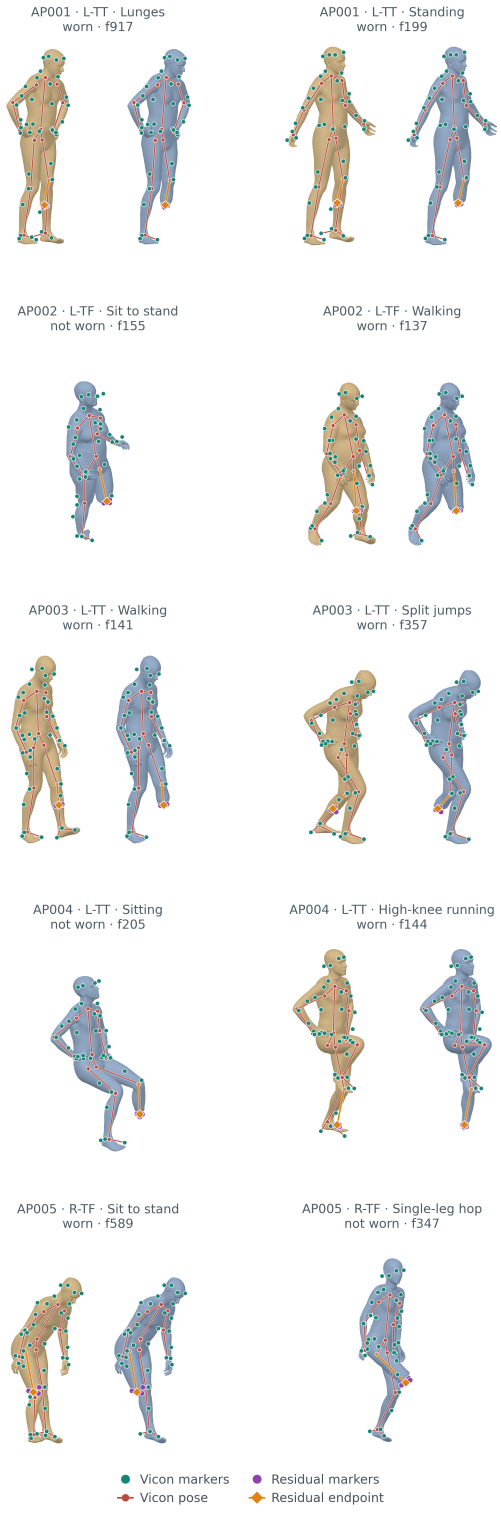

Figure 6: **Qualitative fitted-reference examples across all five participants.** Prosthesis-worn cases show both the kinematic full-body and anatomical residual-limb references, while non-worn cases show the anatomical reference. Examples span different deficiency levels, sides, and motion categories, with Vicon-derived constraints overlaid.

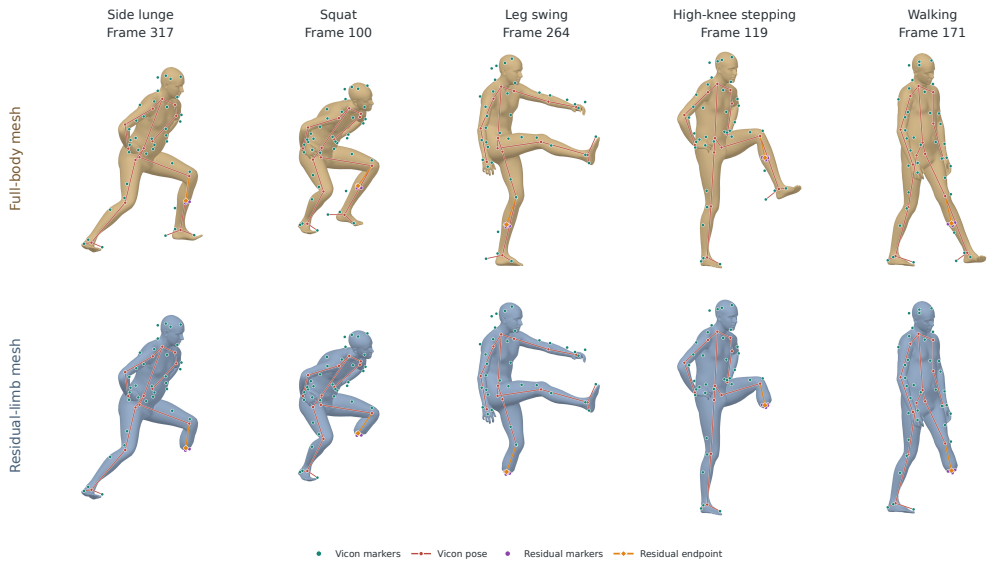

Figure 7: **Qualitative fitted-reference examples across motion categories.** Full-body kinematic references (top) and residual-limb anatomical references (bottom) are shown with Vicon-derived pose and residual markers across representative movements. The two representations share the measured body evidence but differ in the affected distal-limb semantics.