Does Lower MPJPE Mean Better Biomechanics? Evaluating Joint Angle Fidelity of State-of-the-Art 3D Pose Estimation Models

Jason Wang
University of Virginia
School of Data Science
jyw5hw@virginia.edu

Natalie Kupperman
University of Virginia
School of Data Science
nk4qm@virginia.edu

Stephen Baek
University of Virginia
School of Data Science
baek@virginia.edu

Abstract

Three-dimensional human pose estimation models are typically evaluated using Mean Per Joint Position Error (MPJPE), a metric that measures positional accuracy of predicted joint locations. However, biomechanical applications require joint angles such as flexion, abduction, and rotation, which are the clinically meaningful representation of human movement. We evaluate three state-of-the-art 3D pose estimation models (MotionBERT, VideoPose3D, SimpleBaseline3D) on the AthletePose3D dataset, a real-world athletic motion capture dataset, and compute ISB-compliant joint angles from predicted poses via a learned pose-to-angle mapping. Our key finding reveals a paradox: models with lower *MPJPE produce* higher *joint angle error. MotionBERT, the best model by MPJPE (37.2 mm), yields the worst angle accuracy (25.19°), while SimpleBaseline3D, the worst by MPJPE (58.5 mm), achieves the best angle accuracy (19.21°). These results demonstrate that MPJPE alone is insufficient for evaluating biomechanical fidelity and argue for adopting angle-based evaluation metrics.*

1. Introduction

The computer vision community evaluates 3D human pose estimation models primarily through Mean Per Joint Position Error (MPJPE) [1], which measures the average Euclidean distance between predicted and ground-truth joint positions. This metric has driven substantial progress: recent models such as MotionBERT [12] achieve MPJPE below 40 mm on standard benchmarks, representing significant improvements over earlier methods [5, 6].

However, the downstream consumers of pose estimation which includes biomechanists, clinicians, and sports scientists, do not reason about raw joint positions. Instead, they care about joint angles: flexion, abduction, and rotation at anatomical joints. Joint angles are invariant to body size and camera perspective, making them the clinically meaningful representation of human movement [9, 10]. Whether assessing injury risk, evaluating surgical outcomes, or analyzing athletic performance, the angle between body segments, not the absolute position of skeletal landmarks, determines the biomechanical interpretation. In practice, practitioners select pose estimation models based on their MPJPE rankings on standard benchmarks such as Human3.6M, treating benchmark performance as a proxy for real-world deployment quality. Whether this proxy holds when the downstream task is joint angle estimation in unconstrained settings has not been examined. As computer vision-based pose estimation is increasingly adopted in clinical and sports settings as a scalable alternative to marker-based motion capture [3, 7], understanding whether metric improvements translate to meaningful biomechanical accuracy becomes urgent.

This disconnect raises a critical question: does lower MPJPE translate to better joint angle estimation? There are reasons to suspect it may not. Small positional errors can propagate nonlinearly into large angle errors, particularly at multi-axial joints such as the hip and shoulder, where angular motion depends on the relative orientation of multiple body segments. A model that distributes small errors uniformly across joints may achieve low MPJPE while distorting the angular relationships between segments. Crucially, answering this question requires evaluation under realistic deployment conditions, not the controlled laboratory environments of standard benchmarks like Human3.6M [1].

In this work, we investigate this question empirically. Our contributions are:

1. We develop a learned pose-to-angle mapping that computes 12 ISB-compliant joint angles from 3D poses, achieving 2.95° validation mean absolute error (MAE) on ground-truth poses.
2. We benchmark three state-of-the-art 3D pose estimation models on AthletePose3D [11], a real-world athletic dataset whose challenging conditions reflect practical deployment scenarios, evaluating both MPJPE and joint angle accuracy.
3. We demonstrate that model rankings derived from Hu-

man3.6M MPJPE do not predict joint angle accuracy on real-world athletic data, challenging the assumption that benchmark positional accuracy implies biomechanical fidelity.

2. Related Work

3D Human Pose Estimation. Modern 3D pose estimation approaches commonly adopt a 2D-to-3D lifting paradigm. Martinez *et al*. [5] demonstrated that a simple fully-connected network can achieve competitive results by lifting detected 2D keypoints to 3D. VideoPose3D [6] introduced temporal convolutions to exploit motion coherence across frames. MotionBERT [12] leverages a transformer architecture with dual-stream spatiotemporal modeling for state-of-the-art performance. Alternative approaches include direct regression from images [8] and parametric body model fitting [2, 4].

Evaluation Metrics. MPJPE remains the dominant metric for 3D pose evaluation, typically reported on the Human3.6M benchmark [1]. Procrustes-aligned MPJPE (PA-MPJPE) factors out global rotation and translation, while scale-corrected variants additionally remove global scale differences. These position-based metrics have driven steady progress, with state-of-the-art models now achieving MPJPE below 40 mm. However, MPJPE treats all positional errors equally regardless of their direction relative to joint axes, and does not capture whether the geometric relationships between joints, or the angular structure of the pose, are preserved. Two poses with identical MPJPE can have very different joint angles if their errors are distributed differently across the kinematic chain.

Biomechanics and Pose Estimation. The International Society of Biomechanics (ISB) has established standards for reporting joint angles using anatomical coordinate frames and Euler/Cardan decomposition sequences [9, 10]. These standards define how to construct segment coordinate systems and decompose relative orientations into clinically interpretable angles. Several recent studies have validated markerless motion capture against marker-based gold standards. Kanko *et al*. [3] found that markerless systems can achieve clinically acceptable accuracy for gait analysis but with joint-specific limitations, particularly at the hip and shoulder. Scataglini *et al*. [7] conducted a meta-analysis across multiple markerless systems, reporting variable accuracy that depends heavily on the joint, the movement performed, and the system used. These validation studies compare specific systems against motion capture, but none have examined the fundamental question of whether the metric used to develop pose estimation models (MPJPE) aligns with the downstream biomechanical quantities of interest. Our work addresses this gap directly.

3. Methodology

3.1. Dataset

We evaluate on the AthletePose3D dataset [11], which contains approximately 1.2 million frames of multi-camera video with synchronized marker-based motion capture ground truth. The dataset provides both 17-joint H36M-style 3D skeleton poses and ISB-compliant joint angles derived from 142 optical markers. Our test set comprises 996 clips from a competitive athletic domain.

The choice of AthletePose3D is deliberate. Standard benchmarks such as Human3.6M [1] capture subjects performing scripted actions in controlled laboratory settings with ideal lighting and minimal occlusion, conditions far removed from how pose estimation is actually deployed. AthletePose3D represents a real-world deployment scenario, featuring fast dynamics, self-occlusion, and complex multiplanar motion including jumps with full-body rotation, spins at extreme joint ranges, and rapid transitions between extended and tucked postures. Evaluating on such data ensures our findings reflect the performance gap practitioners would encounter when applying these models in practice.

3.2. Ground-Truth Joint Angle Computation

Ground-truth angles are computed from 142 optical markers (e.g. at the pelvis, thorax, and limbs) provided by a Vicon system at 120Hz following ISB standards. We compute 12 clinical degrees of freedom (DOF): bilateral hip flexion/extension and abduction/adduction (ZXY sequence), bilateral shoulder plane of elevation and elevation angle (YXY sequence), and bilateral elbow and knee flexion. Internal/external rotations for the hip and shoulder, as well as axial rotation for the shoulder, are intentionally omitted. These components are notoriously difficult for sparse 2D-to-3D lifting models to resolve without explicit segment-orientation markers. Consequently, knee and elbow motions are simplified to 1D vector angles between adjacent segments, matching the constraints of current pose estimation paradigms.

3.3. Pose-to-Angle Mapping Model

We train a learned mapping from the 17-joint 3D skeleton to 12 ISB-compliant joint angles. Rather than using raw joint coordinates, we first extract bone-relative features: the 16 bone vectors connecting adjacent joints in the H36M skeleton are each decomposed into a unit direction vector (3 components) and a scalar bone length, yielding a 64-dimensional input representation. This bone-relative encoding is invariant to global translation and captures the inter-segment orientations that directly determine joint angles, providing a more natural parameterization than absolute coordinates.

The model architecture consists of an input projection layer followed by 4 residual blocks totaling 8.4M parameters.

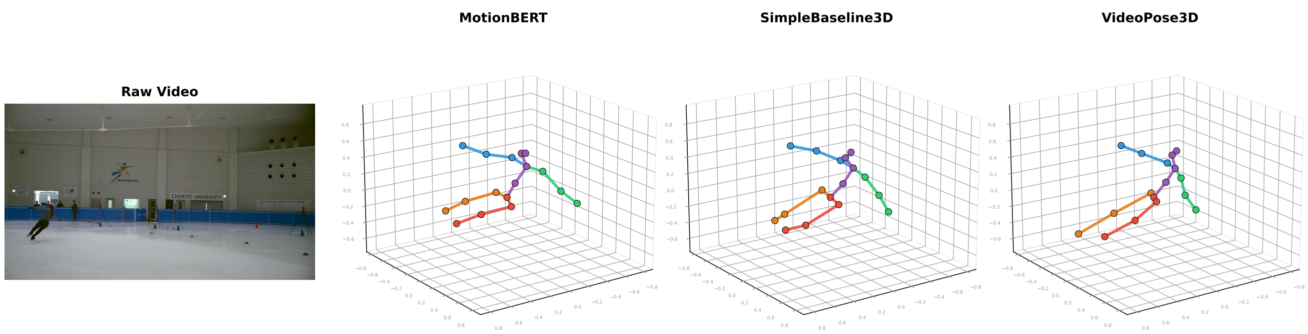

Figure 1. Raw video with three 3D pose estimation outputs from our three test models.

Training uses mean squared error(MSE) loss with AdamW optimization (lr=10^{-4}, weight decay $= 0.1$), cosine annealing with 10-epoch warmup, for 300 epochs. The best model achieves 2.95° MAE on validation ground-truth poses, establishing a strong baseline.

3.4. 3D Pose Estimation Models

We evaluate three representative 2D-to-3D lifting models spanning different architectural paradigms:

- **SimpleBaseline3D** [5]: A fully-connected network that lifts single-frame 2D detections to 3D, representing the simplest effective approach.
- **VideoPose3D** [6]: A temporal convolutional network that leverages motion context across multiple frames.
- **MotionBERT** [12]: A transformer-based model with dual-stream spatial and temporal attention, achieving state-of-the-art MPJPE on Human3.6M.

All models are applied to video clips from the AthletePose3D test set, producing per-frame 3D poses in H36M 17-joint format. These models span three architectural paradigms, single-frame feedforward, temporal convolution, and transformer, providing a representative cross-section of the current 2D-to-3D lifting landscape.

3.5. Evaluation Protocol

The evaluation pipeline proceeds as follows: for each video clip, each pose estimation model produces per-frame 3D poses, which are passed through the same trained model to obtain predicted joint angles. These are compared against ground-truth ISB angles via mean absolute error (MAE), computed per-angle per-frame and averaged across frames within each clip, then across clips.

MPJPE values are reported from published Human3.6M benchmarks (SimpleBaseline3D: 58.5 mm, VideoPose3D: 46.8 mm, MotionBERT: 37.2 mm) to contextualize which models are considered better performers. This mirrors how practitioners currently select models: by consulting published benchmark rankings rather than evaluating on their target domain. Our evaluation tests whether this selection strategy leads to better biomechanical outcomes. We emphasize that MPJPE is used here solely as a reference for model ranking, our primary evaluation metric is angle MAE on the AthletePose3D test set.

Table 1. MPJPE vs. joint angle MAE. Lower MPJPE does not yield lower angle error, the relationship is *inverted*.

Model	MPJPE (mm) ↓	Angle MAE (°) ↓
GT Poses + Our Model	—	2.95
SimpleBaseline3D(SB3D)	58.5	19.21
VideoPose3D(VP3D)	46.8	21.77
MotionBERT	37.2	25.19

4. Results

4.1. MPJPE vs. Angle MAE

Tab. 1 presents the central finding of this work. As MPJPE decreases (indicating better positional accuracy by CV standards), angle MAE increases. SimpleBaseline3D, the "worst" model by MPJPE (58.5 mm), achieves the best angle accuracy (19.21°), while MotionBERT, the "best" by MPJPE (37.2 mm), yields the worst angle accuracy (25.19°). Our pose-to-angle model achieves 2.95° MAE on ground-truth poses, establishing that the 16–22° gap between the baseline and the pose estimation results is caused entirely by pose estimation error, and this gap *widens* as MPJPE improves.

4.2. Per-Joint Angle Analysis

Fig. 2 presents the per-joint breakdown across all 996 test clips. Key observations:

- **Upper limb joints have highest errors**: Shoulder and elbow angles exhibit the highest errors (17–33°), likely due to greater degrees of freedom and susceptibility to occlusion during athletic motion.
- **Lower limb joints have lowest errors**: Hip abduction (8–12°) and knee flexion (13–18°) show the lowest errors

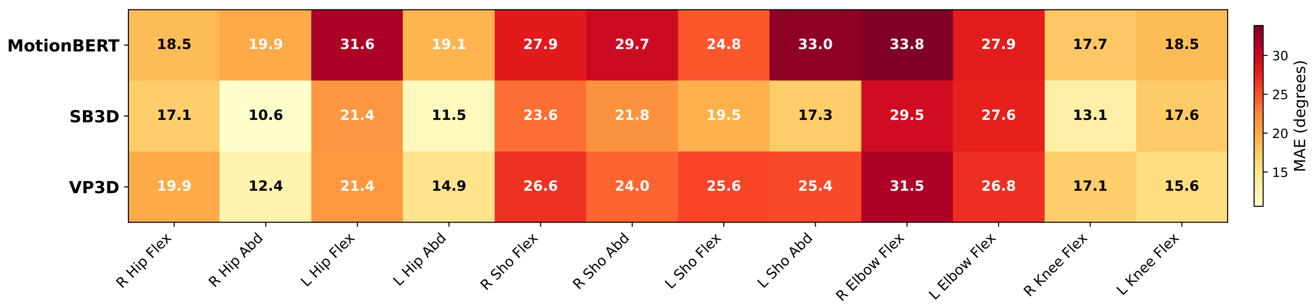

Figure 2. Heatmap of joint angle output errors for each pose estimation algorithm.

among the estimation models, corresponding to joints with simpler kinematics.

- **The inverse trend holds per-joint**: MotionBERT consistently shows higher angle MAE than SimpleBaseline3D for shoulder and elbow angles, despite its superior MPJPE.

4.3. Per-Subject Variation

Crucially, the inverse relationship between MPJPE and angle MAE is preserved within each subject: MotionBERT remains the worst by angle MAE despite its lowest MPJPE, and SimpleBaseline3D remains the best. The per-subject gap also suggests that angle accuracy is highly sensitive to factors beyond aggregate positional error, such as how well a model handles specific movement patterns or body proportions.

5. Discussion

The MPJPE–Angle Paradox. Our results reveal a counterintuitive finding: models that achieve lower MPJPE does not guarantee biomechanical fidelity. Our finding is that model rankings established on Human3.6M do not transfer to joint angle accuracy on real-world athletic data. We hypothesize two complementary mechanisms. First, models optimized for average positional error may learn to distribute errors across joints in ways that minimize MPJPE but distort the angular relationships between body segments. MPJPE weights all joints equally and is agnostic to error direction where a 5 mm error tangential to a joint's arc of motion has a much larger effect on the computed angle than a 5 mm error along the bone axis, yet both contribute identically to MPJPE. Second, more expressive models (transformers with temporal attention) may learn stronger spatial priors from their training data (Human3.6M) that regularize joint positions toward plausible-looking poses but do not generalize to the angular structure of out-of-distribution athletic movements. The simpler SimpleBaseline3D, lacking such priors, may produce noisier but less systematically biased poses.

Implications for the CV Community. These findings suggest that MPJPE, while useful for measuring positional accuracy, is insufficient as the sole evaluation metric when pose estimates are used for biomechanical analysis. We advocate for reporting joint angle accuracy alongside MPJPE, particularly for applications in sports science, clinical assessment, and ergonomics. The pose-to-angle mapping model developed here provides a practical and lightweight tool for computing such metrics from standard 17-joint pose outputs, requiring no additional hardware or marker data at inference time. We note that our finding does not diminish the value of MPJPE, as it remains a necessary measure of positional accuracy. Rather, we argue it is not sufficient as a model that achieves excellent MPJPE but poor angle accuracy may be unsuitable for downstream biomechanical applications.

Limitations.

Our evaluation uses a single athletic domain, which features extreme ranges of motion that may amplify the MPJPE–angle discrepancy. The learned angle mapping, while accurate on ground-truth poses (2.95° MAE), introduces a modeling layer between poses and angles. We evaluate 12 angles across 4 joint types; a more comprehensive assessment would include wrist, ankle, and spinal angles. Despite these limitations, the consistent inverse relationship between MPJPE and angle MAE across three architecturally diverse models suggests this finding generalizes beyond the specific domain studied here.

Conclusion. We have demonstrated that lower MPJPE does not imply better biomechanical fidelity. As the computer vision and biomechanics communities increasingly collaborate, evaluation metrics must reflect the end-use of pose estimates. We call for the adoption of angle-based evaluation metrics and the incorporation of biomechanical constraints into pose estimation training objectives.

References

[1] 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 TPAMI*, 36(7):1325–1339, 2014. 1, 2

[2] Angjoo Kanazawa, Michael J. Black, David W. Jacobs, and Jitendra Malik. End-to-end recovery of human shape and pose. In *CVPR*, pages 7122–7131, 2018. 2

[3] Robert M. Kanko, Elise K. Laende, Elyse M. Davis, W. Scott Selbie, and Kevin J. Deluzio. Concurrent assessment of gait kinematics using marker-based and markerless motion capture. *Journal of Biomechanics*, 127:110665, 2021. 1, 2

[4] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM TOG*, 34(6):248:1–248:16, 2015. 2

[5] Julieta Martinez, Rayat Hossain, Javier Romero, and James J. Little. A simple yet effective baseline for 3D human pose estimation. In *ICCV*, pages 2640–2649, 2017. 1, 2, 3

[6] Dario Pavllo, Christoph Feichtenhofer, David Grangier, and Michael Auli. 3D human pose estimation = 2D pose estimation + matching. In *CVPR*, pages 7035–7043, 2019. 1, 2, 3

[7] Sofia Scataglini, Elise Abts, Carl Van Bocxlaer, Maxim Van den Bussche, Silvio Meletani, and Steven Truijen. Accuracy, validity, and reliability of markerless camera-based 3D motion capture systems versus marker-based 3D motion capture systems in gait analysis: A systematic review and meta-analysis. *Sensors*, 24(11):3686, 2024. 1, 2

[8] Xiao Sun, Bin Xiao, Fangyin Wei, Shuang Liang, and Yichen Wei. Integral human pose regression. In *ECCV*, pages 529–545, 2018. 2

[9] Ge Wu, Sorin Siegler, Paul Allard, Chris Kirtley, Alberto Leardini, Dieter Rosenbaum, Mike Whittle, Darryl D. D'Lima, Luca Cristofolini, and Hartmut Witte. ISB recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion—Part I: ankle, hip, and spine. *Journal of Biomechanics*, 35(4):543–548, 2002. 1, 2

[10] Ge Wu, Frans C. T. van der Helm, H. E. J. Veeger, Mohsen Makhsous, Peter Van Roy, Carolyn Anglin, Jochem Nagels, Andrew R. Karduna, Kevin McQuade, and Xuguang Wang. ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion—Part II: shoulder, elbow, wrist and hand. *Journal of Biomechanics*, 38(5):981–992, 2005. 1, 2

[11] Ethan Yeung, Chris Bigelow, Meghana Vummidigiridhar, Anirudh Jamkhandi, Jason Wang, and Stephen Baek. AthletePose3D: A large-scale multi-view 3D human pose dataset for athletic motion analysis. In *CVPR*, 2025. 1, 2

[12] Wentao Zhu, Xiaoxuan Ma, Zhaoyang Liu, Libin Liu, Wayne Wu, and Yizhou Wang. MotionBERT: A unified perspective on learning human motion representations. In *ICCV*, pages 15085–15099, 2023. 1, 2, 3