

EVALUATING 2D POSE ESTIMATION MODELS FOR BIOMECHANICAL ANALYSIS USING A HIGH-FIDELITY DATASET

Chris-Mari Schreuder^{1,3}, Rynhard Fourie¹, Oloff Bergh², and Rensu P. Theart¹

¹Electrical and Electronic Engineering, Stellenbosch University, Stellenbosch, South Africa

²Neuromechanics Unit, Central Analytical Facility, University of Stellenbosch, Stellenbosch, South Africa

³ Email: chrismari.schreuder@gmail.com

Introduction

Human Pose Estimation (HPE) plays a vital role in sports science and rehabilitation, enabling accurate movement tracking for performance analysis and injury prevention. While modern 2D pose estimation models have achieved strong results on benchmark datasets, they typically use limited keypoints and general-purpose data, which limits their applicability in clinical and other biomechanical contexts. This research addresses that gap by evaluating the adaptability of state-of-the-art 2D HPE models to dense, medically relevant keypoint datasets. The objective is to identify models that are both accurate and efficient enough for practical use in high-stakes, real-time environments. Particular focus is placed on assessing each model's ability to support biomechanical analysis, including fine-grained joint tracking and robustness to movement complexity.

Methods

Three representative HPE architectures were selected: HRNet (high-resolution CNN), ViTPose (transformer-based), and YOLO-Pose (real-time bottom-up regression model). Each model was fine-tuned and evaluated on the SUMediPose dataset [1], which contains over 3 million multi-angle images of 26 subjects performing functional movements at different speeds. The dataset provides 63 anatomically accurate 3D keypoints derived from a Vicon motion capture system. To assess adaptability and scalability, three increasingly complex keypoint subsets (17, 34, and 59 points) were used.

Performance was quantified using the Normalized Mean Absolute Error (NMAE), a resolution-independent metric designed for consistent evaluation across varying image scales and keypoint complexities. The top-performing models were then used for multi-view 3D reconstruction, where triangulated keypoints were analysed to compute joint angles in the sagittal, frontal, and transverse planes*. These angles were compared against ground truth Vicon data to evaluate the models' suitability for biomechanical analysis.

Results and Discussion

Following initial evaluation, ViTPose-B-simple, ViTPose-L-classic, and HRNet-w48-768 were identified as the top performers and selected for full fine-tuning. HRNet-w48-768 achieved the highest accuracy across

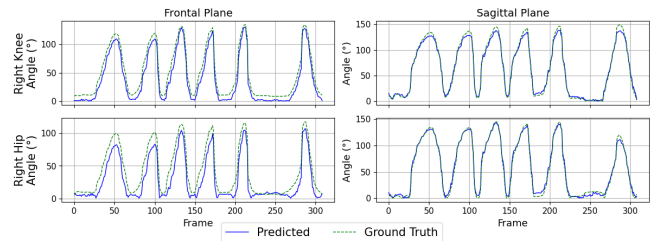


Figure 1: Example of a single subject's knee (top) and hip (bottom) joint angle trajectories during a squat movement, shown over time in the frontal (left) and sagittal (right) planes. Predicted angles from the model (blue) are compared with ground truth Vicon data (green).

all keypoint subsets but had the highest computational cost, with an inference time of 17.03 ms per frame. In contrast, ViTPose-B-simple and ViTPose-L-classic offered significantly faster speeds—3.31 ms and 5.72 ms respectively—while maintaining strong accuracy.

For 3D reconstruction, HRNet-w48-768 achieved the lowest mean error against ground truth Vicon data, with a reconstruction error of 25.87 mm. When joint angle trajectories over time were analysed in the sagittal, frontal, and transverse planes, as seen in Figure 1, HRNet consistently produced the most biomechanically accurate results. The angular error was 3.54° in the sagittal plane and 9.08° in the frontal plane, demonstrating strong alignment with clinical reference standards.

Conclusions

This research demonstrates the viability of adapting state-of-the-art 2D pose estimation models for high-resolution biomechanical analysis using medically grounded datasets. By extending these models to support dense keypoints and evaluating their biomechanical accuracy, the study bridges a critical gap between conventional computer vision methods and clinical-grade motion analysis. The findings support the use of markerless, video-based systems in contexts where traditional motion capture is impractical. Future work will focus on enhancing model robustness through temporal filtering, increasing anatomical detail, and optimizing for deployment in real-time sports and rehabilitation settings.

References

- [1] Chris-Mari Schreuder, Oloff Bergh, Lizé Steyn, and Rensu P. Theart. Sumedipose: A 2d-3d pose estimation dataset. *Data in Brief*, 60:111579, 2025. ISSN 2352-3409. doi: <https://doi.org/10.1016/j.dib.2025.111579>.

*https://youtu.be/p_7MBKxOPPc