

BioPose: Biomechanically-accurate 3D Pose Estimation from Monocular Videos

Farnoosh Koleini*, Muhammad Usama Saleem*, Pu Wang, Hongfei Xue, Ahmed Helmy, Abbey Fenwick
University of North Carolina at Charlotte

{fkoleini, msaleem2, pu.wang, hongfei.xue, ahmed.helmy, afenwick}@charlotte.edu*

Abstract

Recent advancements in 3D human pose estimation from single-camera images and videos have relied on parametric models, like SMPL. However, these models oversimplify anatomical structures, limiting their accuracy in capturing true joint locations and movements, which reduces their applicability in biomechanics, healthcare, and robotics. Biomechanically accurate pose estimation, on the other hand, typically requires costly marker-based motion capture systems and optimization techniques in specialized labs. To bridge this gap, we propose BioPose, a novel learning-based framework for predicting biomechanically accurate 3D human pose directly from monocular videos. BioPose includes three key components: **a Multi-Query Human Mesh Recovery model (MQ-HMR)**, **a Neural Inverse Kinematics (NeurIK) model**, and **a 2D-informed pose refinement technique**. MQ-HMR leverages a multi-query deformable transformer to extract multi-scale fine-grained image features, enabling precise human mesh recovery. NeurIK treats the mesh vertices as virtual markers, applying a spatial-temporal network to regress biomechanically accurate 3D poses under anatomical constraints. To further improve 3D pose estimations, a 2D-informed refinement step optimizes the query tokens during inference by aligning the 3D structure with 2D pose observations. Experiments on benchmark datasets demonstrate that BioPose significantly outperforms state-of-the-art methods.

1. Introduction

Biomechanically-accurate 3D human pose estimation (BA-HPE) refers to the process of predicting a human's 3D body posture, including joint positions and movements, in a way that aligns with the actual anatomical and physical constraints of the human skeletal system. BA-HPE plays a crucial role in fields like physical therapy, sports science, and ergonomics [17, 31, 35], where precise modeling of human motion is necessary for injury prevention, rehabilitation, or performance analysis. It is also useful in robotics, anima-

*Equal contribution.

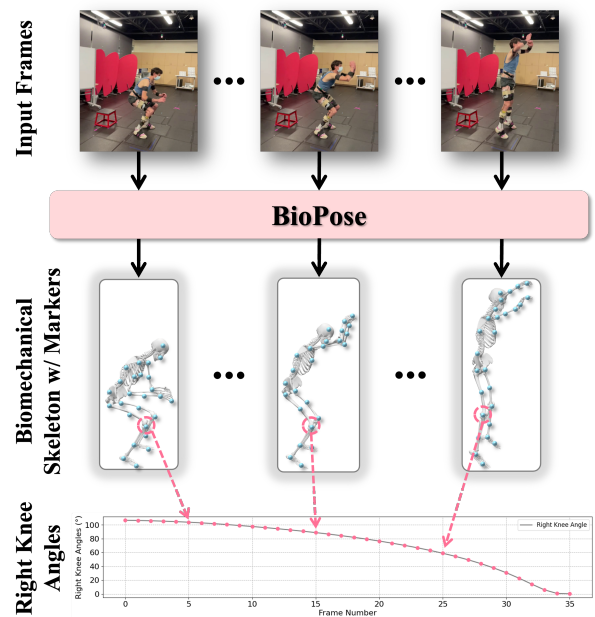


Figure 1. BioPose is a framework for biomechanically accurate 3D pose estimation from monocular videos. It first integrates Multi-Query Human Mesh Recovery (MQ-HMR), which leverages multi-scale image features for precise 3D mesh reconstruction, secondly with Neural Inverse Kinematics (NeurIK), ensuring biomechanical constraints for anatomically valid 3D poses.

tion, and human-computer interaction [1, 27, 38, 40, 41].

Traditionally, marker-based motion capture systems have been the gold standard for obtaining biomechanically accurate 3D pose data. These systems utilize multiple cameras to track reflective markers placed on the human body in controlled laboratory environments [33, 34]. The captured marker data is then processed using sophisticated biomechanical optimization systems, such as OpenSim [36], which often requires skilled professionals to configure and refine. Although these marker-based systems provide high accuracy, they are expensive, labor-intensive, and impractical for use outside of specialized labs, particularly in dynamic, real-world environments.

On the other hand, significant advancements have been

made in 3D pose estimation from monocular videos (single-camera footage) [11, 14], where deep neural networks are used to infer 3D poses, specifically, the rotation angles of body joints, from 2D image sequences, leveraging parametric human body models like SMPL [21, 28]. While monocular pose estimation offers greater accessibility compared to traditional marker-based systems, it faces notable limitations due to the anatomical simplifications inherent in body models like SMPL. These body models are designed to produce visually plausible poses, but fail to achieve biomechanical accuracy, particularly in joint positioning and skeletal movements.

To address these challenges, we propose *BioPose*, a novel learning-based framework for biomechanically accurate 3D human pose estimation from monocular videos. BioPose consists of three core components: a multi-query human mesh recovery model (MQ-HMR), a neural inverse kinematics (NeurIK) model, and a 2D-informed pose refinement technique. The *MQ-HMR* model utilizes a multi-query deformable transformer to extract fine-grained, multi-scale image features from monocular video frames, enabling precise recovery of a 3D human mesh. This model simultaneously estimates 3D body pose, shape parameters, and camera parameters, resulting in a accurate and detailed body mesh. In the next stage, the *NeurIK* model treats the recovered mesh vertices as virtual markers and uses a spatial-temporal network to regress biomechanically accurate 3D poses. This process is guided by the biomechanical skeleton model with anatomical realism, ensuring the predicted poses adhere to human biomechanics, such as anatomical locations and degrees of freedom of the joints. To further enhance accuracy, BioPose introduces a *2D-informed refinement step* during inference, optimizing pose queries in latent space to align the predicted 3D poses with 2D pose cues from the video. This refinement corrects 3D-to-2D projection discrepancies, ensuring both visual coherence and biomechanical precision. Our key contributions are summarized as follows:

- We introduce BioPose, a novel learning-based framework for biomechanically accurate 3D pose estimation from monocular videos, offering performance comparable to traditional marker-based optimization methods while retaining the convenience and accessibility of monocular learning-based approaches.
- We propose a MQ-HMR model for accurate 3D mesh reconstruction and virtual marker tracking along with the NeurIK module that incorporates a biomechanical skeleton model to ensure anatomically valid 3D poses from monocular videos.
- We develop a novel 2D-informed refinement technique that further enhances 3D pose estimation via inference-time optimization that aligns the predicted 3D structure with 2D pose cues.

- Extensive experiments demonstrate that MQ-HMR model outperforms state-of-the-art methods in monocular human mesh recovery from single images and BioPose system achieves the very competitive performance, compared with golden-standard multi-camera marker-based techniques.

2. Related Work

2.1. Biomechanically-accurate 3D Pose Estimation

The gold standard for biomechanically accurate 3D pose estimation combines multi-camera, marker-based tracking systems with biomechanical optimization tools. This process involves three main steps: first, retro-reflective markers are attached to the subject's body and tracked using synchronized infrared cameras. Next, a calibration pose (T-pose) is captured, allowing tools like OpenSim [36] to scale a biomechanical skeleton model to the subject's anatomy. Finally, the subject performs the target motion, and the marker data, along with the scaled skeleton model, is used to compute 3D joint rotation angles through inverse kinematics optimization.

To reduce the cost of marker-based systems, marker-less multi-camera approaches like OpenCap [15, 35] and Pose2Sim [31] have emerged. These methods use 2D pose estimation algorithms like OpenPose [6] to detect keypoints from video frames, triangulating them from multiple camera views to reconstruct 3D positions. However, the resulting keypoints are sparse and anatomically imprecise, requiring post-processing techniques like keypoint augmentation [32] to improve anatomical accuracy. These systems also require tedious camera calibration and synchronization, along with optimization tools for final 3D pose estimation. To address these limitations, D3KE [4] was recently proposed, using deep neural networks to directly regress biomechanically accurate 3D poses from monocular videos. However, D3KE struggles with generalization due to limited paired training data. In contrast, BioPose leverages a novel MQ-HMR model to generate robust 3D human meshes from in-the-wild videos, serving as virtual markers. These markers, combined with our biomechanics-guided NeurIK model and 2D-informed inference-time optimization, yield highly generalizable 3D pose estimations with strong anatomical accuracy.

2.2. HMR from Monocular Images

Human Mesh Recovery (HMR) from monocular images has evolved significantly, focusing on estimating 3D meshes from single 2D images. Early approaches mainly leverage convolutional neural networks to directly regress the parameters of parametric human model SMPL [28], from images [10, 18, 22] and videos [7, 9, 19]. More recently, vision transformers [2] have been adopted for HMR tasks,

yielding state-of-the-art results. Models like HMR 2.0 [14] and TokenHMR [11] exploit the transformer’s ability to model long-range dependencies, allowing them to better capture relationships between different body parts and improve mesh reconstruction accuracy. Besides human 3D mesh, these HMR methods also can produce the estimated 3D pose (joint rotation angles) of its underlying skeleton structure. However, such skeleton is physically inaccurate, inherently leading to inaccurate pose estimations. To address these challenges, our BioPose framework utilizes the SMPL mesh vertices as the virtual tracking markers because SMPL mesh is designed to accurately capture the deformable 3D body surface. Using these virtual markers as inputs, NeurIK model is trained to learn the 3D pose parameters of a biomechanical skeleton model with anatomical details. Since the pose estimation accuracy depends on the mesh reconstruction quality, we propose MQ-HMR model leads to improved mesh recovery results compared with SOTA solutions by harvesting multi-scale image features via multi-query deformable transformer along 2D-informed query optimization at inference.

3. Proposed Method: BioPose

The goal of BioPose is to predict biomechanically accurate 3D human poses directly from monocular videos. As shown in Figure 3, BioPose has three core components. The MQ-HMR model uses a multi-query deformable transformer decoder to extract multi-scale image features from vision transformer encoder, enabling precise recovery of 3D human meshes (Section 3.2). These meshes are then used by the NeurIK model, which treats the mesh vertices as virtual markers, applying a spatial-temporal network to infer biomechanically accurate 3D poses while maintaining anatomical constraints (Section 3.3). To further improve accuracy, a 2D-informed pose refinement step aligns the 3D predictions with 2D observations, enhancing both visual coherence and biomechanical validity (Section 3.4).

3.1. Preliminaries

3.1.1 SMPL Human Mesh Model

We make use of the SMPL model, a differentiable parametric framework for representing human surface geometry [28]. This model encodes the human body using pose parameters $\theta \in \mathbb{R}^{24 \times 3}$ and shape parameters $\beta \in \mathbb{R}^{10}$. The pose parameters $\theta = [\theta_1, \dots, \theta_{24}]$ include both the global orientation $\theta_1 \in \mathbb{R}^3$ of the entire body and the local joint rotations $[\theta_2, \dots, \theta_{24}] \in \mathbb{R}^{23 \times 3}$, with each θ_k describing the axis-angle rotation of joint k relative to its parent joint in the kinematic tree. By combining these pose and shape parameters, the SMPL model produces a detailed 3D mesh $M(\theta, \beta) \in \mathbb{R}^{3 \times N}$, where $N = 6890$ vertices represent the surface of the body. The positions of the body joints

$J \in \mathbb{R}^{3 \times k}$ are then derived as a weighted sum of these vertices, formulated as $J = MW$, where $W \in \mathbb{R}^{N \times k}$ contains the predefined weights that map vertices to the corresponding joints.

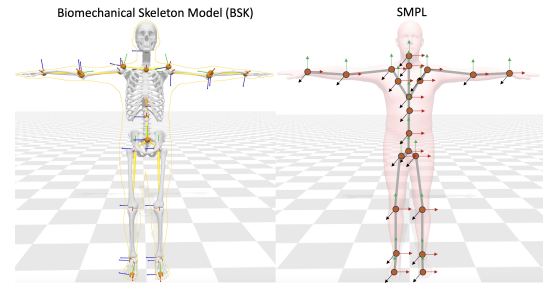


Figure 2. **Left:** Biomechanical skeleton model has anatomical details with accurate joint locations and bone orientations. **Right:** SMPL body model has deformable 3D body surface that includes an approximate skeleton geometry with inaccurate joint location and bone orientations.

3.1.2 Biomechanical Skeleton (BSK) Model

The BSK model, e.g., widely-adopted OpenSim models, is represented by a series of bone segments that are interconnected through movable anatomical joints, which possess anatomical movement constraints, such as Degrees of Freedom, to limit the range of motion of the respective body parts. In particular, the BSK model generally consists of 24 rigid bone segments, which are represented by three sets of parameters (q^o, q^r, s) . The anatomical joint orientation $q^o \in [q_1^o, \dots, q_{24}^o]$ with $q_i^o \in \mathbb{R}^3$ defines the relative orientation of each joint with respect to its parent joint along the kinematic skeleton tree. Therefore, q^o are determined by the anatomical structure of human skeleton. $q^r \in [q_1^r, \dots, q_{24}^r]$ represents the motion-induced joint rotation with $q_i^r \in \mathbb{R}^{D_i}$ and $D_i \leq 3$ represents the Euler’s angle rotation of joint i relative to its parent in the kinematic tree under the constraints imposed by the degree of freedom D_i of each joint i . The bone scale $s \in [s_1, \dots, s_{24}]$ with $s_i \in \mathbb{R}^3$ aims to tailor the generic anatomical skeleton model (in the rest pose) by scaling each bone length and shape along with the (x, y, z) axis. The scaled skeleton yields the body joints $p_J \in \mathbb{R}^{3 \times 24}$ at anatomical positions. The differences between SMPL and BSK models are shown in Fig. 2

3.1.3 Optimization-based Biomechanical Kinematics

With assistance of the BSK model, the kinematic analysis aims to find the optimal pose, i.e., joint rotation angles q^r , which can best fit the BSK model to the motion capture sequences. Towards this goal, a set of model markers is first attached to the bone segments in such a way that each bone segment is associated with at least D_i markers to ensure the unique solutions of the derived rotation angles at joint i with

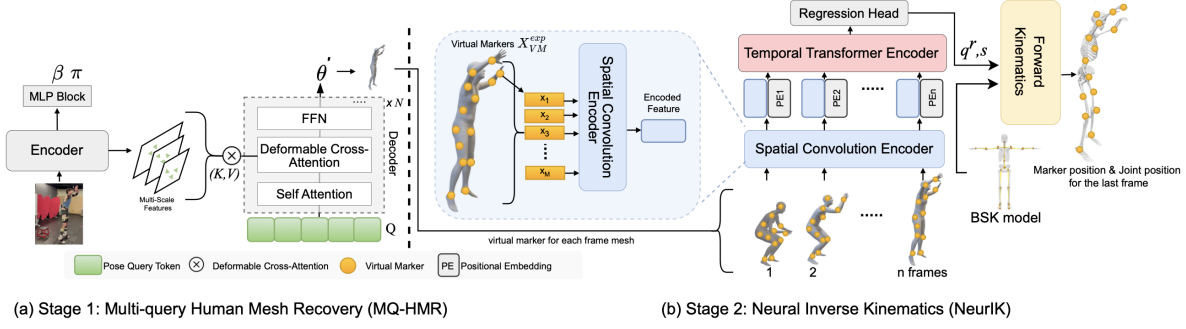


Figure 3. Overview of BioPose, comprising two key components: (1) the **MQ-HMR** model, which leverages a multi-query deformable transformer decoder to extract multi-scale image features from a vision transformer, enabling precise 3D mesh recovery, and (2) the **NeurIK** model, which uses the mesh vertices as virtual markers and applies a spatio-temporal network to infer biomechanically accurate 3D poses while maintaining anatomical constraints.

D_i degrees of freedom. Then, a set of corresponding experiential markers is placed on the human subject. Then, the pose q^r and bone scale s can be obtained by solving an optimization problem that minimizes the distance between each experimental marker and its corresponding model marker. i.e.,

$$q^{r*}, s^* = \arg \min_{q^r, s} \sum_{i=1}^M \|f_{FK}(q^r, s, q^o, p_i) - x_i^{exp}\|$$

where M is the number of markers. $p_i \in \mathbb{R}^3$ denotes the position of i -th model marker in the local coordination system of the body segment to which it is attached. $f_{FK}(q^r, s, q^o, p_i)$ is the forward kinematics transformation that converts the model marker i from its local coordination frame to the world coordination system under the scaled skeleton with the pose of q^r . $x_i^{exp} \in \mathbb{R}^3$ is the position of the experiential marker i in the world coordination system. We leverage OpenSim, the classic biomechanical optimizer, and BML-MoVi dataset [13] to obtain q^{r*}, s^* , which serve as the ground-truth data to train the NeurIK model.

3.2. MQ-HMR

The goal of MQ-HMR is predicting accurate virtual experiential markers, X_{VM}^{exp} , from the 3D mesh, which serves as the inputs for NeurIK model. Towards this goal, MQ-HMR model consists of two key components: the image encoder and the multi-query deformable transformer decoder.

Image Encoder. Our image encoder is based on the Vision Transformer (ViT), specifically the ViT-H/16 variant [2, 11]. The encoder begins by dividing the input image into 16x16 pixel patches, which are processed through multiple transformer layers to produce a set of feature tokens that encode the visual information. To enhance this process, we implement a multi-scale feature extraction approach [2]. After generating the initial feature map, the en-

coder upsamples it to produce feature maps at various resolutions. These higher-resolution maps capture fine-grained visual details, such as joint positions and orientations, while lower-resolution maps preserve broader, high-level semantic information, such as the overall human skeleton structure. This multi-scale strategy allows the encoder to capture both intricate local details and global contextual information simultaneously, which is crucial for accurate pose estimation.

Multi-Query Deformable Transformer Decoder. Building upon multi-scale feature maps, the multi-query deformable transformer decoder introduces a novel mechanism designed to recover precise 3D human poses by extracting fine-grained semantic information from diverse resolutions. The multi-query approach initializes multiple pose queries as zero pose tokens, which interact with the encoder’s multi-scale feature maps. These queries generate pose anchors crucial for accurately estimating complex human poses, especially in challenging scenarios like occlusions or ambiguous body positions. To process high-resolution feature maps efficiently, MQ-HMR incorporates a multi-scale deformable cross-attention mechanism [25, 39], focusing each query on a small set of sampling points near the pose anchors and dynamically adjusting attention to the most relevant regions. This optimizes both computational efficiency and accuracy, allowing the model to concentrate on critical spatial features with minimal overhead. The deformable attention mechanism (DAM) for multi-scale features is formulated as:

$$\text{DAM}(\mathbf{Q}, \hat{\mathbf{r}}_k, \{F^s\}_{s=1}^S) = \sum_{s=1}^S \sum_{m=1}^M \alpha_{ksm} \cdot \mathbf{W} F^s (\hat{\mathbf{r}}_k + \Delta r_{ksm})$$

Where \mathbf{Q} represents the pose token queries, $\hat{\mathbf{r}}_k$ denotes the learnable reference points, Δr_{ksm} refers to the learnable sampling offsets around the reference points, $\{F^s\}_{s=1}^S$

are the multi-scale image features, α_{ksm} are the attention weights, and \mathbf{W} is a learnable weight matrix. This deformable cross-attention strategy allows the model to capture fine-grained details while balancing computational efficiency. From this accurately reconstructed 3D mesh, virtual markers $\mathbf{X}_{\text{VM}}^{\text{exp}}$ are extracted, serving as input to the NeurIK module for further refinement and achieving precise biomechanical accuracy.

MQ-HMR Losses. In line with established practices in HMR research [11, 14], we train our MQ-HMR model using a combination of losses based on SMPL parameters, 3D keypoints, and 2D keypoints. The final overall loss is:

$$L_{\text{total}} = L_{\text{smpl}} + L_{3\text{D}} + L_{2\text{D}} \quad (1)$$

where L_{smpl} minimizes the error between the predicted and ground-truth SMPL pose (θ) and shape (β) parameters, $L_{3\text{D}}$ supervises the accuracy of the 3D keypoint predictions, and $L_{2\text{D}}$ enforces consistency between the projected 3D keypoints and their corresponding 2D annotations.

3.3. NeurIK

After extracting virtual markers $\mathbf{X}_{\text{VM}}^{\text{exp}}$ from MQ-HMR, the NeurIK module processes these markers to predict biomechanically accurate 3D poses q_r and bone scale s . To achieve this, it employs three key components: i) a Spatial Convolution Encoder to model spatial relationships among body parts, ii) a Temporal Transformer Encoder to capture dynamic motion patterns over time, and iii) multiple loss functions that incorporate biomechanical constraints from a musculoskeletal model.

Spatial Convolution Encoder. The Spatial Convolution Encoder is designed to extract high-dimensional spatial features from a single frame. Given the 3D human mesh generated by the pre-trained MQ-HMR model, we extract M virtual markers from the mesh, where each marker m has 3D coordinates (x_m, y_m, z_m) . These marker positions $\mathbf{X}_{\text{VM}}^{\text{exp}} \in \mathbb{R}^{M \times 3}$ are first projected into a higher-dimensional space using a trainable linear projection, resulting in the spatial embedding $\mathbf{Z}_{n_i} \in \mathbb{R}^{M \times c}$, where c is the spatial embedding dimension. To capture spatial relationships across the body, this spatial embedding is processed through a series of 1-D convolutional layers, which capture both local relationships between neighboring markers and global dependencies across the entire body structure. The output of this spatial convolution process for frame n_i , $\mathbf{Z}_{n_i} \in \mathbb{R}^{M \times c}$, represents a refined spatial feature embedding, which is passed to the Temporal Transformer Encoder for temporal modeling.

Temporal Transformer Encoder. After encoding high-dimensional spatial features for each individual frame, the Temporal Transformer Encoder models the dependencies

across the sequence of frames. For frame n_i , the spatially encoded feature matrix $\mathbf{Z}_{n_i} \in \mathbb{R}^{M \times c}$ is flattened into a vector $\mathbf{z}_{n_i} \in \mathbb{R}^{1 \times (M \cdot c)}$. We concatenate these vectors across all n frames to form the sequence matrix $\mathbf{Z}_{\text{seq}} \in \mathbb{R}^{n \times (M \cdot c)}$, which represents the spatial features for the entire motion sequence. To capture temporal relationships, we add a learnable temporal positional embedding $\mathbf{PE}_n \in \mathbb{R}^{n \times (M \cdot c)}$ to the sequence matrix. The temporal transformer then applies multi-head self-attention blocks and feed-forward layers to model both short-term and long-term dependencies across frames. This allows the model to understand the progression of motion and how body parts evolve over time. The final output of the temporal transformer, $\mathbf{Y}_n \in \mathbb{R}^{n \times (M \cdot c)}$, is used to predict key biomechanical parameters such as body scales s and joint angles q^r . These predictions are further refined through a Forward Kinematics (FK) module to ensure biomechanically accurate marker and joint positions.

NeurIK Losses. Spatial and temporal models are trained using multiple supervisions, including joint positions, marker positions, body scales, and joint angles. The joint positions correspond to anatomical landmarks in the musculoskeletal model, ensuring biomechanical accuracy. The overall loss function L is a weighted sum of four terms: L_j for joint positions, L_m for marker positions, L_s for body scales, and L_q for joint angles. These terms are weighted by coefficients λ_j , λ_m , λ_s , and λ_q , respectively, to control their contributions to the total loss. The **overall loss function** is defined as:

$$L_{\text{neurIK}} = \lambda_j L_j + \lambda_m L_m + \lambda_s L_s + \lambda_q L_q \quad (2)$$

In particular, both L_j and L_m incorporate biomechanical constraints during training through the forward kinematics (FK) layer. As shown in Fig. 3, the FK layer transforms the rest-pose BSK model markers and anatomical joints to the new positions according to the estimated rotation angles q^r in such a way that the model markers best match the virtual experimental markers. Since we employed a full-body skeletal model from OpenSim [33], the FK transformation is inherently constrained by the realistic degrees of freedom and range of motions of body joints. The details of loss functions are shown in the supplementary material.

3.4. 2D Pose-Informed Refinement at Inference

During inference time, to further reduce uncertainties in the 3D reconstruction process, we fine-tune the pose query tokens Q within the MQ-HMR model while keeping the rest of the network frozen as shown in Figure 4. This refinement ensures alignment between the predicted 3D poses and 2D pose data, leveraging robust 2D pose detectors such as OpenPose [6]. The process begins with the pose query tokens generated by MQ-HMR, and the initial pose parameters θ' are derived from the output of MQ-HMR for a given

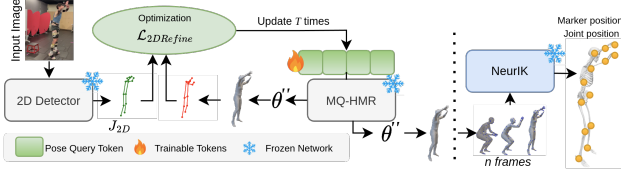


Figure 4. Our **2D Pose-Guided Refinement** process fine-tunes the pose query tokens to minimize 3D reconstruction errors by aligning the 3D body mesh with 2D pose estimates, improving visual coherence and providing more reliable virtual markers for the NeurIK module to further refine 3D biomechanical pose accuracy.

input image. These tokens and parameters are iteratively adjusted to minimize discrepancies between the inferred 3D and observed 2D poses by optimizing a guidance function $F(Q_t, J_{2D}, \theta')$. This function penalizes misalignments between the two pose domains while enforcing regularization through the following expression:

$$Q^+ = \arg \min_{Q_t} (\mathcal{L}_{2DRefine}(J'_{3D}) + \lambda_{\theta'} \mathcal{L}_{\theta'}(\theta')) \quad (3)$$

where the term $\mathcal{L}_{2DRefine}(J'_{3D})$ aims to align the reprojected 3D joints J'_{3D} with the detected 2D keypoints J_{2D} , using the following relation:

$$\mathcal{L}_{2DRefine}(J'_{3D}) = \|\Pi(K(J'_{3D})) - J_{2D}\|^2 \quad (4)$$

where $\Pi(K(\cdot))$ represents the perspective projection function governed by the camera intrinsics K . Simultaneously, the regularization term $\mathcal{L}_{\theta'}(\theta')$ constrains the pose parameters θ' , ensuring that they do not diverge excessively from their initial values, thereby avoiding anatomically implausible body configurations. The pose tokens Q_t are iteratively updated at each step t via gradient-based optimization as follows:

$$Q_{t+1} = Q_t - \eta \nabla_{Q_t} F(Q_t, J_{2D}, \theta') \quad (5)$$

where η determines the step size for the update, and $\nabla_{Q_t} F(Q_t, J_{2D}, \theta')$ denotes the gradient of the objective function with respect to the pose tokens at the current iteration t . This iterative refinement process, carried out over T total iterations, gradually fine-tunes the pose representation, reducing the gap between the 3D estimates and their 2D counterparts while keeping realistic pose configurations.

4. Experiments

4.1. Datasets

To train MQ-HMR, we employ a diverse collection of datasets. In line with prior research [14] and to ensure consistency in comparisons with other baselines, a diverse set

of standard datasets (SD) is used for training, including Human3.6M (H36M) [16], COCO [26], MPI-INF-3DHP [30], and MPII [3]. For NeurIK, we utilize BML-MoVi [13] for training, which provides rich biomechanical motion capture and video data from multiple actors performing everyday activities.

4.2. Evaluation Metrics

The performance of MQ-HMR and NeurIK is evaluated using a set of key metrics designed for evaluating 3D human pose accuracy and biomechanical estimation. For MQ-HMR, the metrics used include Mean Per Joint Position Error (MPJPE) and Mean Vertex Error (MVE) to evaluate the accuracy of 3D pose estimation. Additionally, we report Procrustes-Aligned MPJPE (PA-MPJPE), which measures the alignment between predicted and ground-truth poses after a rigid transformation, providing a more precise comparison of pose structure. MQ-HMR is evaluated on the Human3.6M testing split, consistent with prior studies [23], and for *generalization* testing, we assess the model on challenging real-world datasets like 3DPW [37] and EMDB [20]. Importantly, **no** training is performed on these datasets, ensuring a fair evaluation on unseen data.

For NeurIK, we employ metrics that focus on biomechanical accuracy. These include Mean Per Bony Landmarks Position Error (MPBLPE), which measures the accuracy of predicted bony landmarks against ground truth positions, and Mean Absolute Error for body scale (MAE_{body}), which evaluates the correctness of predicted body segment dimensions by comparing their longest axes in millimeters. Additionally, we report Mean Absolute Error for joint angles (MAE_{angle}), which assesses the precision of joint angle predictions in degrees, critical for biomechanical simulations such as joint force and muscle force analysis. To assess computational efficiency, we use the metric Average Inference Time per Image (AITI) in seconds, similar to the average inference or optimization time per image described by OpenPose [6]. Unlike that, AITI tracks the processing time per iteration, allowing for a more detailed evaluation of our 2D Pose-Informed Refinement technique. AITI is obtained on a single NVIDIA Quadro RTX A4000. Lower values in these metrics indicate better model performance. For NeurIK, we test the model on the BMLmovi [13], OpenCap [35], and BEDLAM [5] datasets. NeurIK is **not** trained on OpenCap and BEDLAM datasets to show its cross-dataset generalization performance.

4.3. Comparison to State-of-the-art Approaches

4.3.1 Quantitative Evaluation of MQ-HMR

As shown in Table 1, MQ-HMR demonstrates significant improvements over existing methods across multiple benchmark datasets, particularly on Human3.6M, 3DPW, and

Table 1. MQ-HMR Reconstruction Evaluation in 3D: Reconstruction errors (measured in mm) on the Human3.6M, 3DPW, and EMDB datasets. Smaller values (\downarrow) represent superior performance. Underlined values highlight the second-best performance in each column. **Blue** highlights the improvements of our MQ-HMR approach over the second-best method. - indicates missing results.

Methods	Venue	Human3.6M		3DPW			EMDB		
		PA-MPJPE	MPJPE	PA-MPJPE	MPJPE	MVE	PA-MPJPE	MPJPE	MVE
		(\downarrow)	(\downarrow)	(\downarrow)	(\downarrow)	(\downarrow)	(\downarrow)	(\downarrow)	(\downarrow)
FastMETRO [8]	ECCV 2022	33.7	52.2	65.7	109.0	121.6	72.7	108.1	119.2
PARE [22]	ICCV 2021	50.6	76.8	50.9	82.0	97.9	72.2	113.9	133.2
Virtual Marker [29]	CVPR 2023	-	-	48.9	80.5	93.8	-	-	-
CLIFF [24]	ECCV 2022	<u>32.7</u>	47.1	46.4	73.9	87.6	68.8	103.1	122.9
HMR2.0 [14]	ICCV 2023	33.6	<u>44.8</u>	<u>44.5</u>	<u>70.0</u>	<u>84.1</u>	<u>61.5</u>	<u>97.8</u>	120.1
VQ HPS [12]	ECCV 2024	-	-	45.2	71.1	84.8	65.2	99.9	<u>112.9</u>
TokenHMR [11]	CVPR 2024	36.3	48.4	47.5	75.8	86.5	66.1	98.1	116.2
MQ-HMR	Ours	28.5 (4.2\downarrow)	42.5 (2.3\downarrow)	39.5 (5.0\downarrow)	69.0 (1.0\downarrow)	79.8 (4.3\downarrow)	52.1 (9.4\downarrow)	92.5 (5.3\downarrow)	98.9 (14.0\downarrow)

Table 2. NeurIK Biomechanical Analysis in 3D: Evaluation of biomechanical accuracy on the BML-MoVi, BEDLAM, and OpenCap datasets. Metrics include body scale error (MAE_{body}), marker positions (MPBLPE), and joint angle error (MAE_{angle}). Smaller values (\downarrow) indicate better performance. Underlined values highlight the fourth-best performance in each column, achieved by existing HMR2.0 [14] + NeurIK (our NeurIK module combined with HMR2.0). **Blue** highlights the improvements of our MQ-HMR + NeurIK[†] over the fourth-best method. - indicates missing results.

Methods	BML-MoVi Dataset			BEDLAM Dataset			OpenCap Dataset		
	MAE_{body}	MPBLPE	MAE_{angle}	MAE_{body}	MPBLPE	MAE_{angle}	MAE_{body}	MPBLPE	MAE_{angle}
	(\downarrow)	(\downarrow)	(\downarrow)	(\downarrow)	(\downarrow)	(\downarrow)	(\downarrow)	(\downarrow)	(\downarrow)
D3KE [4]	5.90	36.98	3.54	8.26	39.45	6.72	7.15	38.62	5.92
OpenCap Multi-Camera [35]	-	-	-	-	-	-	-	-	4.50
HMR2.0 [14] + NeurIK (Ours)	<u>5.86</u>	<u>35.90</u>	<u>3.31</u>	<u>6.01</u>	<u>36.62</u>	<u>3.85</u>	<u>5.89</u>	<u>36.07</u>	<u>3.41</u>
HMR2.0 [14] + NeurIK (Ours)	4.11	26.31	2.73	4.84	27.02	3.55	5.29	27.52	3.21
MQ-HMR + NeurIK (Ours)	5.26	28.39	3.87	5.62	30.12	4.16	5.42	32.17	4.02
MQ-HMR + NeurIK[†] (Ours)	3.97 (32%\downarrow)	25.76 (28%\downarrow)	2.84 (14%\downarrow)	4.28 (29%\downarrow)	26.54 (27%\downarrow)	3.14 (18%\downarrow)	4.87 (17%\downarrow)	26.34 (27%\downarrow)	3.19 (6%\downarrow)

[†] indicates MQ-HMR with 2D Pose-Informed Refinement during inference. OpenCap method is the multi-camera markerless motion capture system. D3KE is the single-camera solution. HMR2.0 + NeurIK indicates the virtual markers come from the mesh vertices of HMR 2.0, the SOTA solution for monocular HMR.

EMDB. The model’s performance is especially noteworthy on EMDB, which simulates complex, real-world scenarios, achieving substantial reductions in PA-MPJPE (by 9.4 mm) and MVE (by 14.0 mm), showcasing its ability to excel in intricate environments with enhanced precision. Similarly, on 3DPW, known for its in-the-wild settings, MQ-HMR further demonstrates its effectiveness by lowering MPJPE by 5.0 mm and MVE by 4.3 mm, confirming its robust performance in diverse 3D pose reconstruction tasks. The key innovation driving these improvements is MQ-HMR’s use of multi-scale deformable attention, enabling simultaneous capture of fine-grained details and broader contextual information. This balanced approach enhances both precision and computational efficiency. Consequently, these results highlight MQ-HMR’s ability to generalize effectively across both controlled and real-world conditions.

4.3.2 Quantitative Evaluation of NeurIK

As per Table 2, the proposed MQ-HMR + NeurIK method introduces significant advancements in biomechanically accurate 3D human pose estimation, as demonstrated across the BML-MoVi, BEDLAM, and OpenCap

datasets. By focusing on key metrics like joint angle and marker position accuracy—crucial for biomechanical applications—our model outperforms previous approaches. On the BML-MoVi dataset, the joint angle error (MAE_{angle}) is reduced by 14.2%, and marker position error (MPBLPE) improves by 28.3%, underscoring the model’s precision in estimating detailed biomechanical movements. The improvement is even more pronounced on the BEDLAM dataset, where joint angle error sees an 18.4% reduction, a testament to the model’s robustness in handling complex synthetic motions. Marker position accuracy also improves by 27.5%, highlighting the precision of our model in predicting anatomically correct body movements even in challenging environments. For OpenCap dataset, our method still delivers a 6.4% reduction in joint angle error, compared with HMR2.0 + NeurIK baseline, and an impressive 29% reduction compared with the OpenCap multi-camera tracking system, a clinically-proved motion analysis solution for biomechanical and healthcare applications.

Table 3. Impact of Number of Iterations on 2D Pose-Informed Refinement

# of Iteration	BML-MoVi			BEDLAM			OpenCap			AITI (s)
	MAE _{body}	MPBLPE	MAE _{angle}	MAE _{body}	MPBLPE	MAE _{angle}	MAE _{body}	MPBLPE	MAE _{angle}	
1	4.89	26.23	3.47	5.61	27.61	4.11	5.41	27.82	3.98	0.293
5	4.12	25.92	2.97	5.13	27.19	3.45	5.33	27.15	3.97	0.405
10	3.97	25.76	2.84	4.28	26.54	3.14	5.12	26.61	3.52	0.620
20	4.42	25.81	3.01	4.76	27.12	3.94	4.87	26.34	3.19	0.838

5. Ablation Study

5.1. Effectiveness of Pose Query Token

The Effectiveness of Pose Query Tokens in MQ-HMR highlights the significant influence of token quantity on model performance across datasets as shown in Table 4. Starting with 4 tokens, the model underperforms, showing an MPJPE of 98.8 mm on EMDB. As the pose token count increases, there is a clear improvement, with 96 tokens yielding the best results: an MPJPE of 69.0 mm and MVE of 79.8 mm on 3DPW, and 98.9 mm on EMDB. However, beyond 96 tokens, the improvements diminish. At 192 tokens, MPJPE increases to 74.7 mm, and performance plateaus or declines further with 384 tokens. This indicates that 96 tokens strike the optimal balance between maximizing accuracy and maintaining computational efficiency.

Table 4. Impact of Pose Query Token on MQ-HMR

# of Pose Tokens	H36M		3DPW		EMDB	
	MPJPE	MPJPE	MVE	MPJPE	MVE	MPJPE
4	44.4	75.3	85.7	98.8	103.7	102.3
24	43.9	73.6	83.9	95.5	102.3	100.4
48	45.1	73.1	83.2	95.5	100.4	98.9
96	42.5	69.0	79.8	92.5	98.9	101.6
192	44.5	74.7	84.6	95.9	101.6	99.7
384	44.1	74.1	84.5	95.7	99.7	

5.2. Impact of Feature Resolutions

The ablation study of MQ-HMR reveals clear trends in how multi-scale feature resolutions impact 3D human pose estimation performance as shown in Table 5. As feature resolutions increase from 1× to 16×, the MPJPE consistently improves, reducing errors by 3.5 mm on 3DPW and 2.9 mm on EMDB. Including intermediate scales such as 4× and 8× further reduces errors, showing that the model benefits from a balance of both global and local information. However, adding the highest resolution scale 32× brings only minimal improvement, suggesting diminishing returns. Conversely, when lower resolutions (1× or 4×) are excluded, errors increase sharply, particularly on 3DPW, where MPJPE rises by 6.1 mm, indicating the importance of low-resolution features for maintaining overall pose structure. These trends emphasize the effectiveness of multi-scale feature fusion in MQ-HMR, balancing fine detail and broader spatial context for optimal performance.

Table 5. Impact of feature resolutions on MPJPE error on 3DPW and EMDB datasets. ✓ indicates inclusion, ✗ indicates exclusion, with deltas (Δ) showing error differences with all feature maps.

Feature Scale						MPJPE (↓)	
1×	4×	8×	16×	32×		3DPW	EMDB
✓	✓	✓	✓	✓		69.0	91.5
✓	✗	✗	✗	✗		75.1 _{Δ6.1}	95.5 _{Δ4.0}
✓	✓	✗	✗	✗		73.5 _{Δ4.5}	93.4 _{Δ1.9}
✓	✓	✓	✗	✗		72.5 _{Δ3.5}	92.9 _{Δ1.4}
✓	✓	✓	✓	✗		69.0 _{Δ0.0}	92.5 _{Δ1.0}
✗	✓	✓	✓	✓		70.6 _{Δ1.6}	90.8 _{Δ0.7}
✓	✗	✓	✓	✓		71.5 _{Δ2.5}	91.7 _{Δ1.2}
✓	✓	✗	✓	✓		72.3 _{Δ3.3}	91.4 _{Δ0.9}
✓	✓	✓	✗	✓		70.1 _{Δ1.1}	90.5 _{Δ1.0}

5.3. Effectiveness of 2D Pose-Informed Refinement

Table 3 shows the effect of varying iteration numbers on the 2D pose-informed model across the BML-MoVi, BEDLAM, and OpenCap datasets. Performance generally improves with more iterations, peaking at 10 iterations for BML-MoVi and BEDLAM, where the model achieves the lowest (MAE_{body} and MAE_{angle}, 3.97 and 2.84 for BML-MoVi, and 4.28 and 3.14 for BEDLAM). However, beyond 10 iterations, performance slightly declines, showing diminishing returns. For the OpenCap dataset, performance remains optimal at 20 iterations, with MAE_{body} of 4.87 and MAE_{angle} of 3.19. This underscores the method’s ability to achieve significant improvements in just a few iterations while balancing computational time. For example, at 5 iterations, the model performs well (MAE_{body} of 4.12 on BML-MoVi) with an AITI of just 0.405 seconds (NVIDIA Quadro RTX A4000). This shows that effective refinement is achieved early, making the approach highly efficient for real-time applications.

6. Conclusion

We introduce BioPose, a framework for biomechanically accurate 3D human pose estimation from monocular video. BioPose integrates human mesh recovery with biomechanical analysis, combining Multi-Query Human Mesh Recovery (MQ-HMR) for mesh reconstruction, Neural Inverse Kinematics (NeurIK) for biomechanical refinement, and 2D-informed pose alignment. Experiments show its significant gains over state-of-the-art methods, making it a powerful tool for clinical assessments, sports analysis, and rehabilitation technology.

References

- [1] Danial Abshari, Chenglong Fu, and Meera Sridhar. Llm-assisted physical invariant extraction for cyber-physical systems anomaly detection. *arXiv preprint arXiv:2411.10918*, 2024. 1
- [2] Dosovitskiy Alexey. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 2, 4
- [3] 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*, pages 3686–3693, 2014. 6
- [4] Marian Bittner, Wei-Tse Yang, Xucong Zhang, Ajay Seth, Jan van Gemert, and Frans CT van der Helm. Towards single camera human 3d-kinematics. *Sensors*, 23(1):341, 2022. 2, 7
- [5] Michael J Black, Priyanka Patel, Joachim Tesch, and Jinlong Yang. Bedlam: A synthetic dataset of bodies exhibiting detailed lifelike animated motion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8726–8737, 2023. 6
- [6] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7291–7299, 2017. 2, 5, 6
- [7] Hanbyel Cho, Jaesung Ahn, Yooshin Cho, and Junmo Kim. Video inference for human mesh recovery with vision transformer. In *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*, pages 1–6. IEEE, 2023. 2
- [8] Junhyeong Cho, Kim Youwang, and Tae-Hyun Oh. Cross-attention of disentangled modalities for 3d human mesh recovery with transformers. In *European Conference on Computer Vision*, pages 342–359. Springer, 2022. 7
- [9] Hongsuk Choi, Gyeongsik Moon, Ju Yong Chang, and Kyoung Mu Lee. Beyond static features for temporally consistent 3d human pose and shape from a video. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1964–1973, 2021. 2
- [10] Hongsuk Choi, Gyeongsik Moon, JoonKyu Park, and Kyoung Mu Lee. Learning to estimate robust 3d human mesh from in-the-wild crowded scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1475–1484, 2022. 2
- [11] Sai Kumar Dwivedi, Yu Sun, Priyanka Patel, Yao Feng, and Michael J Black. Tokenhmr: Advancing human mesh recovery with a tokenized pose representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1323–1333, 2024. 2, 3, 4, 5, 7
- [12] Guénolé Fiche, Simon Leglaive, Xavier Alameda-Pineda, Antonio Agudo, and Francesc Moreno-Noguer. Vq-hps: Human pose and shape estimation in a vector-quantized latent space. *arXiv preprint arXiv:2312.08291*, 2023. 7
- [13] Saeed Ghorbani, Kimia Mahdavian, Anne Thaler, Konrad Kording, Douglas James Cook, Gunnar Blohm, and Nikolaus F Troje. Movi: A large multi-purpose human motion and video dataset. *Plos one*, 16(6):e0253157, 2021. 4, 6
- [14] Shubham Goel, Georgios Pavlakos, Jathushan Rajasegaran, Angjoo Kanazawa, and Jitendra Malik. Humans in 4d: Reconstructing and tracking humans with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14783–14794, 2023. 2, 3, 5, 6, 7
- [15] Yoni Gozlan, Antoine Falisse, Scott Uhlrich, Anthony Gatti, Michael Black, and Akshay Chaudhari. Opencapbench: A benchmark to bridge pose estimation and biomechanics. *arXiv preprint arXiv:2406.09788*, 2024. 2
- [16] 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. 6
- [17] Xiaoxu Ji, Ranuki O Hettiarachchige, Alexa LE Littman, and Davide Piovesan. Using digital human modelling to evaluate the risk of musculoskeletal injury for workers in the health-care industry. *Sensors*, 23(5):2781, 2023. 1
- [18] Angjoo Kanazawa, Michael J Black, David W Jacobs, and Jitendra Malik. End-to-end recovery of human shape and pose. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7122–7131, 2018. 2
- [19] Angjoo Kanazawa, Jason Y Zhang, Panna Felsen, and Jitendra Malik. Learning 3d human dynamics from video. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5614–5623, 2019. 2
- [20] Manuel Kaufmann, Jie Song, Chen Guo, Kaiyue Shen, Tianjian Jiang, Chengcheng Tang, Juan José Zárate, and Otmar Hilliges. Emdb: The electromagnetic database of global 3d human pose and shape in the wild. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14632–14643, 2023. 6
- [21] 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. 2
- [22] Muhammed Kocabas, Chun-Hao P Huang, Otmar Hilliges, and Michael J Black. Pare: Part attention regressor for 3d human body estimation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 11127–11137, 2021. 2, 7
- [23] Nikos Kolotouros, Georgios Pavlakos, Michael J Black, and Kostas Daniilidis. Learning to reconstruct 3d human pose and shape via model-fitting in the loop. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 2252–2261, 2019. 6
- [24] Zhihao Li, Jianzhuang Liu, Zhensong Zhang, Songcen Xu, and Youliang Yan. Cliff: Carrying location information in full frames into human pose and shape estimation. In *European Conference on Computer Vision*, pages 590–606. Springer, 2022. 7
- [25] Jing Lin, Ailing Zeng, Haoqian Wang, Lei Zhang, and Yu Li. One-stage 3d whole-body mesh recovery with component aware transformer. In *Proceedings of the IEEE/CVF Con-*

- ference on Computer Vision and Pattern Recognition, pages 21159–21168, 2023. 4
- [26] 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*, pages 740–755. Springer, 2014. 6
- [27] Wansong Liu, Xiao Liang, and Minghui Zheng. Task-constrained motion planning considering uncertainty-informed human motion prediction for human–robot collaborative disassembly. *IEEE/ASME Transactions on Mechatronics*, 28(4):2056–2063, 2023. 1
- [28] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 34(6):248:1–248:16, oct 2015. 2, 3
- [29] Xiaoxuan Ma, Jiajun Su, Chunyu Wang, Wentao Zhu, and Yizhou Wang. 3d human mesh estimation from virtual markers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 534–543, 2023. 7
- [30] 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)*, pages 506–516. IEEE, 2017. 6
- [31] David Pagnon, Mathieu Domalain, and Lionel Reveret. Pose2sim: An end-to-end workflow for 3d markerless sports kinematics—part 1: Robustness. *Sensors*, 21(19):6530, 2021. 1, 2
- [32] Chan-Il Park and Chae-Bong Sohn. Data augmentation for human keypoint estimation deep learning based sign language translation. *Electronics*, 9(8):1257, 2020. 2
- [33] Ajay Seth, Jennifer L Hicks, Thomas K Uchida, Ayman Habib, Christopher L Dembia, James J Dunne, Carmichael F Ong, Matthew S DeMers, Apoorva Rajagopal, Matthew Millard, et al. Opensim: Simulating musculoskeletal dynamics and neuromuscular control to study human and animal movement. *PLoS computational biology*, 14(7):e1006223, 2018. 1, 5
- [34] Victor Sholukha, Bruno Bonnechere, Patrick Salvia, Fédor Moiseev, Marcel Rooze, and S Van Sint Jan. Model-based approach for human kinematics reconstruction from markerless and marker-based motion analysis systems. *Journal of biomechanics*, 46(14):2363–2371, 2013. 1
- [35] Scott D Uhlich, Antoine Falisse, Łukasz Kidziński, Julie Muccini, Michael Ko, Akshay S Chaudhari, Jennifer L Hicks, and Scott L Delp. Opencap: Human movement dynamics from smartphone videos. *PLoS computational biology*, 19(10):e1011462, 2023. 1, 2, 6, 7
- [36] Mukul Verma, Muskan Dawar, Prashant Singh Rana, Neeru Jindal, and Harpreet Singh. An opensim guided tour in machine learning for e-health applications. In *Intelligent Data Security Solutions for e-Health Applications*, pages 57–75. Elsevier, 2020. 1, 2
- [37] Timo Von Marcard, Roberto Henschel, Michael J Black, Bodo Rosenhahn, and Gerard Pons-Moll. Recovering accurate 3d human pose in the wild using imus and a moving camera. In *Proceedings of the European conference on computer vision (ECCV)*, pages 601–617, 2018. 6
- [38] Wentao Zhu, Xiaoxuan Ma, Dongwoo Ro, Hai Ci, Jinlu Zhang, Jiaxin Shi, Feng Gao, Qi Tian, and Yizhou Wang. Human motion generation: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 1
- [39] Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable detr: Deformable transformers for end-to-end object detection. *ArXiv*, abs/2010.04159, 2020. 4
- [40] Arastoo Zibaeirad, Farnoosh Koleini, Shengping Bi, Tao Hou, and Tao Wang. A comprehensive survey on the security of smart grid: Challenges, mitigations, and future research opportunities. *arXiv preprint arXiv:2407.07966*, 2024. 1
- [41] Arastoo Zibaeirad and Marco Vieira. Vulnllmeval: A framework for evaluating large language models in software vulnerability detection and patching. *arXiv preprint arXiv:2409.10756*, 2024. 1