

## RESEARCH ARTICLE

# Accuracy and Adaptability Improvement in Aerobic Training: Integration of Self-Attention Mechanisms in 3D Pose Estimation and Kinematic Modeling

HANG QU<sup>1,\*</sup>, HAOTIAN ZHANG<sup>2,\*</sup>, QIQI BAN<sup>1</sup>, AND XIAOLIANG ZHAO<sup>3</sup>

<sup>1</sup>Affiliated Hospital of Yangzhou University, Yangzhou University, Yangzhou 225001, China

<sup>2</sup>Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong, China

<sup>3</sup>School of Physical Education, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Corresponding author: Hang Qu (hangqu@foxmail.com)


\*Hang Qu and Haotian Zhang contributed equally to this work.

**ABSTRACT** Accurately tracking and analyzing human motion during aerobic exercise poses significant challenges due to the dynamic complexity of human biomechanics. Traditional methods often fail to capture this complexity, resulting in training plans that lack personalization and an increased risk of exercise-related injuries. Therefore, developing a method capable of accurately understanding and analyzing the dynamics of human motion has become particularly important. The motivation behind this study is to enhance the safety and effectiveness of aerobic exercise training. By accurately monitoring and analyzing the movements of athletes during their training, it aims to prevent injuries and create personalized training plans. To this end, we believe a new approach is needed to deeply understand human motion, one that can adapt to various environmental changes and provide real-time feedback. We propose a framework that combines 3D pose estimation with kinematic modeling. This method employs self-attention mechanisms and machine learning techniques to precisely capture the complexity of human motion. Our core technology includes a self-attention-based pose estimation system capable of accurately tracking 3D joint positions in various environments, and a detailed kinematic model for biomechanical analysis, including the calculation of joint angles, velocities, and accelerations. Our model was validated using a custom aerobic exercise dataset, demonstrating superior accuracy and adaptability compared to existing models. Comparative analyses with other models highlight the advanced capabilities of our model in accurately interpreting and analyzing human motion. Our experiments confirm that the model excels in precision, robustness to environmental changes, real-time feedback, and injury prevention. Notably, it significantly reduces injury risks by identifying potential stress points and facilitates the generation of personalized training plans.

**INDEX TERMS** 3D pose estimation, aerobics kinematic modeling, self-attention mechanisms, AI in sports science.

## I. INTRODUCTION

Aerobics, a vibrant and widely embraced form of physical activity, skillfully merges rhythmic aerobic exercises with stretching and strength training routines, all designed to holistically enhance key aspects of fitness — including

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Tong .

flexibility, muscular strength, and cardiovascular health, as illustrated in Fig. 1. The holistic nature of aerobics makes it an effective tool for promoting overall well-being and has led to its endorsement by health professionals and fitness enthusiasts alike.

The robust evidence supporting the multifaceted benefits of aerobics is undeniable. Regular participation in aerobic activities is linked not just to physical health improvements

but also to mental health benefits, such as reduced stress and anxiety levels, improved mood, and enhanced cognitive function. Despite this, ensuring that participants perform movements correctly to maximize these health benefits, while also mitigating the risk of injury, presents a substantial obstacle. This is particularly relevant for beginners or those with physical limitations, where incorrect form can lead to strain or injury.

Addressing this critical gap, there has been a surge in research and innovation within the realms of human motion analysis and pose estimation. These advancements aim to provide feedback and guidance on exercise form, creating a safer and more effective workout environment. Recent breakthroughs in sports science and computational technology have propelled the development of sophisticated 3D pose estimation tools. These tools use complex algorithms and data derived from high-speed cameras and sensors to capture and analyze the full spectrum of human movement in three-dimensional space.

The integration of these advanced technologies in aerobic workouts could revolutionize how exercises are taught, learned, and refined. Instructors and practitioners can utilize these tools to monitor and correct form in real-time, ensuring that each movement is performed with precision, thus enhancing the effectiveness of the workout and reducing the likelihood of exercise-related injuries.

Moreover, these technologies hold promise for the personalization of aerobic routines, catering to the unique needs and limitations of individuals. Personal trainers and physical therapists could use detailed analytics to tailor exercise regimens that align with personal fitness goals and physical rehabilitation requirements.

However, despite these technological strides, there are challenges to overcome. Existing models are sometimes unable to keep up with the rapid and varied movements of aerobics, particularly in environments where lighting and space constraints impact the accuracy of motion capture. There is also a learning curve associated with the interpretation of complex data, which requires specialized knowledge, making it less accessible to the average user or small fitness studios.

To effectively address the challenges inherent in aerobic training, our research innovatively applies a 3D pose estimation and kinematic model specifically crafted for the intricate movements of aerobics. This advanced model harnesses the power of self-attention mechanisms—a cutting-edge technique borrowed from the successes of natural language processing—to significantly refine the precision of pose estimation amidst the dynamic flux of aerobic movements. Our ambition is to furnish the aerobics domain with a tool that not only delineates the subtleties of human motion with meticulous accuracy but also sheds light on the underlying biomechanical processes. This could potentially herald a new era in aerobic training, where monitoring, coaching, and personalization are executed with a level of sophistication previously unattainable.



**FIGURE 1.** Man engaged in an aerobics routine showcasing strength and flexibility.

Embarking on this venture, we have meticulously curated a custom dataset that encapsulates the extensive spectrum of aerobic movements, ensuring that our model is trained on data that is as representative and exhaustive as possible. Upon this foundation, we have deployed our model for stringent validation, subjecting it to a battery of tests that measure its performance against a backdrop of well-established models. This critical comparative analysis not only illuminates the merits and prowess of our model but also casts light on avenues for enhancement.

Our paper contributes substantially to the literature and practice in three pivotal areas:

Firstly, it pioneers the synthesis of self-attention mechanisms with pose estimation in aerobics, thereby enriching the discipline of computer vision and the study of human motion. This fusion represents a leap forward, breaking new ground in accuracy and reliability of motion capture.

Secondly, the paper details the development of a comprehensive kinematic model, meticulously constructed to decode the complex biomechanics of aerobic exercises. This model stands out as a beacon for researchers and practitioners alike, seeking to delve into the subtleties of human motion and its application in health and fitness.

Thirdly, the pragmatic application of our models in real-world environments solidifies their status as invaluable assets in enhancing the efficiency of training protocols, delivering real-time feedback to practitioners, and significantly mitigating the risk of injuries, thereby cementing the role of technology as a cornerstone in the future of fitness.

## II. RELATED WORKS IN GYMNASTICS AND SPORTS POSE ESTIMATION

The realm of gymnastics and sports pose estimation has undergone remarkable transformations with the advent of deep learning technologies. This section provides an overview of pivotal contributions that have significantly influenced this field.

### A. DEVELOPMENTS IN GYMNASTICS POSE ESTIMATION

The task of determining the spatial coordinates of key body joints in gymnastics has been revolutionized by deep learning. Key developments include:

Toshev and Szegedy's "DeepPose" marked a seminal moment, employing deep neural networks in human pose estimation, laying the groundwork for subsequent studies [1]. Moon and Lee's "12L-MeshNet" introduced an innovative 3D human pose estimation technique using RGB images, crucial for gymnastics performance analysis [2]. Müller et al. tackled the complexities of self-contact in gymnastics poses, thereby increasing the accuracy of pose estimation in intricate scenarios [3]. Chen, Tian, and He's extensive survey on deep learning methods for monocular human pose estimation offered comprehensive insights into the domain [4]. Andriluka, Pishchulin, Gehler, and Schiele established a new benchmark for 2D human pose estimation, essential for appraising gymnastics models [5].

### B. INNOVATIONS IN DEEP LEARNING FOR KINEMATIC ANALYSIS

Carreira and Zisserman's integration of deep learning with graphical models enhanced the recognition of gymnastics actions by focusing on the interplay between body parts [6]. Rohan et al. developed a CNN-based real-time gait analysis system, applicable in gymnastics training for its immediacy and accuracy [7]. Zhao et al. demonstrated the versatility of deep learning with their view-adaptive recurrent neural networks, underscoring its adaptability in various settings [8]. Boukhayma et al.'s LEGO framework introduced a novel approach to learning edge geometry from videos, opening new pathways in gymnastics action recognition [9]. Fastovets, Guillemaut, and Hilton proposed a unique non-sequential key-frame propagation technique for athlete pose estimation, applicable in gymnastics contexts [10].

### C. EXPANDING HORIZONS IN SPORTS POSE ESTIMATION

Rohan, Rabah, Hosny, and Kim's CNN-based real-time gait analysis technique is also capable of analyzing gymnasts' movements [11]. Song and Fan's posture recognition and estimation method holds potential for adaptation to gymnastics poses [12]. Takeichi, Ichikawa, Shinayama, and Tagawa's mobile application for running form analysis hints at possible gymnastics applications [13]. Kazemi, Burenus, Azizpour, and Sullivan's work on multi-view body part recognition sheds light on the complexities of gymnastics movements [14]. Kondragunta, Jaiswal, and Hirtz's method for deducing gait parameters from 3D poses can be tailored for gymnastics movement analysis [15].

### D. RECENT ADVANCEMENTS IN POSE ESTIMATION

Gong et al.'s "DiffPose" at CVPR significantly enhanced the accuracy of 3D pose estimation [21]. Lin et al. delved into the transition from 2D to 3D models in boxing pose estimation, underscoring advancements in sports analytics [22]. Zhou et al. offered a deep learning-focused survey on pose estimation, tracking, and action recognition [23]. Ingwersen et al. introduced "SportsPose," a dynamic 3D sports pose dataset, vital for sports analytics research [24]. Baumgartner

and Klatt demonstrated a novel approach for 3D pose estimation in sports broadcasts, incorporating partial sports field registration [25]. Qiu et al. proposed a structure-guided diffusion model for 2D human pose estimation, marking a significant improvement in the field [26]. These studies collectively represent significant strides in the evolution of pose estimation technology, particularly in the context of sports and movement analysis.

Our study builds upon this existing body of work by focusing on aerobics. We aim to integrate pose estimation with specialized kinematic modeling, seeking to bridge the gap between 3D pose estimation technology and the practice and analysis of aerobics, potentially improving how this discipline is approached and understood.

## III. METHODOLOGY

Our methodology is comprised of three distinct components: data collection, a specialized approach to 3D pose estimation, and the design of a kinematic model, all of which are specifically tailored for aerobics.

### A. DATASET COLLECTION

Our team, in collaboration with Nanjing University of Posts and Telecommunications, has meticulously compiled a comprehensive dataset featuring aerobic exercise videos for the development and validation of our pose estimation model. This dataset showcases a broad spectrum of participants engaging in diverse aerobic routines, ensuring comprehensive representation of movements for accurate model training.

Utilizing a single camera setup, we captured 100 unique aerobic exercise segments, each approximately five minutes in duration. These segments encompass five distinct aerobic routines, with each routine performed by 20 different participants. To guarantee the precision of our model, each video segment was carefully annotated to capture both 2D and 3D joint positions of the participants, reflecting our commitment to data accuracy and quality.

The dataset comprises the following aerobic routines:

#### 1) CLASSIC AEROBICS

Features traditional aerobic exercises set to high-energy rhythms, emphasizing basic steps to enhance cardiovascular health.

#### 2) DANCE AEROBICS

A fusion of dance movements and aerobic exercises designed to improve rhythm and coordination while offering a cardiovascular workout.

#### 3) STEP AEROBICS

Focuses on choreographed routines performed on an aerobic step, aimed at strengthening the lower body and boosting cardiovascular fitness.

#### 4) BODY CONDITIONING AEROBICS

Combines aerobic activities with strength training, employing either light weights or bodyweight exercises to tone muscles and increase stamina.

## 5) BALL AEROBICS

Incorporates various ball exercises to target balance and core strength, simultaneously improving cardiovascular endurance, coordination, and muscle strength in a fun, engaging manner.

## B. POSE ESTIMATION

This entire procedure is visually depicted in Fig. 2, while the overall framework of the estimation model is outlined in Fig. 3.

### 1) MODEL STRUCTURE

In our computer vision model development for human pose estimation, we have concentrated on accurately deducing the 3D positions of key body joints from 2D video footage. Grounded in deep learning principles, our model is uniquely tailored for enhancing the analysis and optimization of human movements, especially in aerobics. We have made a significant leap beyond traditional approaches by integrating a self-attention structure, substantially refining conventional methods.

The integration of self-attention in our model offers three principal advantages:

1. **Enhanced Modeling Capabilities:** The self-attention mechanism equips the model with the ability to effectively discern global dependencies within sequences, a crucial factor for accurately depicting complex human poses.

2. **Improved Accuracy:** By employing self-attention, we significantly boost the precision of our 3D pose estimation. This is essential for precisely locating human joint positions and enhancing the model's efficacy in motion analysis and optimization.

3. **Increased Applicability:** Thanks to its superior performance, our model demonstrates versatility in diverse domains, such as sports, healthcare, and fitness, thereby positioning it as a versatile tool for various human motion analysis scenarios.

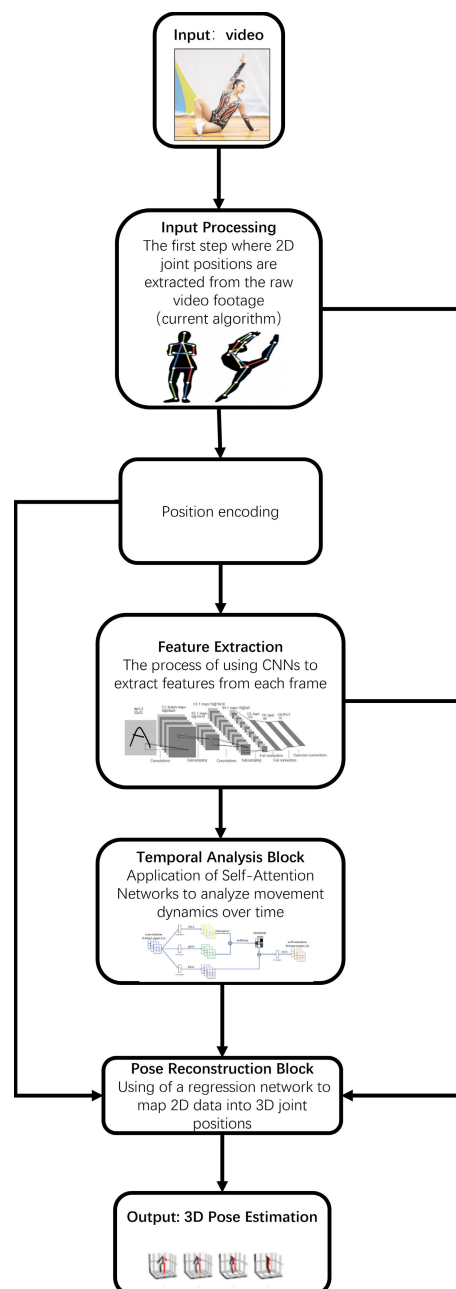
Integrating self-attention into the pose estimation architecture marks a major stride forward, offering a deeper and more nuanced understanding of human movement while greatly enhancing the model's precision and adaptability in multiple applications.

Our model comprises several integral modules:

1. **2D Pose Generation (Using AlphaPose):** We leverage the pre-trained AlphaPose model, known for its exceptional 2D pose estimation accuracy. It detects human figures in images and identifies key joints, forming the basis of our 2D pose matrix for further analysis.

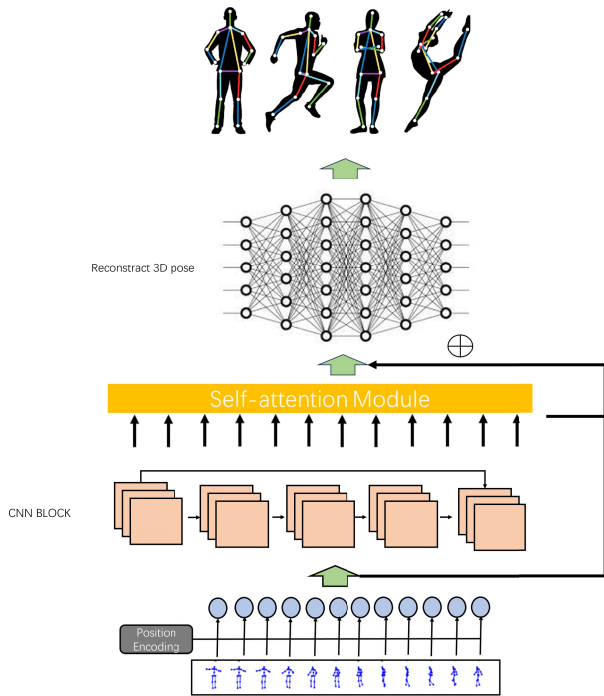
2. **Position Encoding:** In the initial phase, we utilize position encoding to preserve the sequential order of data. This is vital for correctly interpreting the temporal dynamics of human movement, offering advantages over traditional RNNs in terms of processing speed and stability.

3. **Feature Extraction with CNNs:** Our model features a five-layer CNN architecture inspired by ResNet principles,



**FIGURE 2.** Overview of a pose estimation pipeline: This schematic illustrates the complex stages of a pose estimation workflow, starting with the application of AlphaPose for 2D joint detection, followed by position encoding to retain the sequential order of data. Feature extraction is performed using a Convolutional Neural Network (CNN) that draws upon ResNet principles, merging the outputs of different layers to preserve original information. Temporal modeling with self-attention mechanisms captures the dynamics of human movement across frames, employing Query (Q), Key (K), and Value (V) components for an enhanced analysis. The final stage involves a regression network, translating 2D features and temporal data into accurate 3D joint positions, with a 10-layer fully connected architecture incorporating elements of residual learning for improved stability and performance.

extracting key features from 2D input data to build a comprehensive feature set detailing posture and spatial relations.



**FIGURE 3.** An illustration of a 3D pose estimation process for human motion during aerobic exercise, utilizing a self-attention neural network module. The diagram shows a sequence of pose detections leading to a neural network, indicating the flow from raw data input through position encoding to the self-attention mechanism that processes and interprets the complex patterns of movement.

4. Temporal Modeling with Self-Attention Mechanisms: This component captures the complex temporal dependencies in human movement across frames. The model dynamically adjusts its focus on the importance of joint positions and their movements at each time step.

5. 3D Pose Estimation: We employ a specially-designed regression network to convert the 2D features and temporal data into accurate 3D joint positions. The network includes a 10-layer fully connected architecture with residual learning elements for enhanced stability and performance.

## 2) GENERATE 2D IMAGE

In the process of generating 2D points for image representation, our system utilizes a pre-trained iteration of AlphaPose, a distinguished tool in the realm of computer vision, celebrated for its unparalleled 2D pose estimation capabilities. AlphaPose excels in identifying human figures within images and accurately mapping out their key body joints. This proficiency is crucial for creating intricate 2D representations of human poses. By leveraging a pre-trained AlphaPose model, our system efficiently transforms video data into precise 2D pose estimations. This adaptation harnesses the advanced functionalities of AlphaPose, ensuring elevated accuracy and performance in diverse contexts. Incorporating AlphaPose into our methodology markedly amplifies our ability to analyze and decode human movements within a two-dimensional perspective.

The output of this phase is encapsulated in a matrix of pixel points, formatted as [17, 2], which outlines the coordinates of crucial body joints on the 2D plane. Formally, this matrix can be represented as:

$$P = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_{17} & y_{17} \end{bmatrix} \quad (1)$$

where each row in  $P$  corresponds to a unique joint, and the two columns represent the  $x$  and  $y$  coordinates of that joint within the image domain. This structured output acts as a pivotal element for subsequent stages in our pose estimation workflow, facilitating advanced analysis and further applications.

## 3) POSITION ENCODING

In the initial stage of our model's operation, we integrate position encoding - a technique that enhances the model's ability to process data in parallel, presenting a marked improvement over the sequential data handling by traditional Recurrent Neural Networks (RNNs). Position encoding accelerates computation by facilitating simultaneous processing of multiple data points, which significantly reduces the time required for processing when compared to the step-by-step approach of RNNs.

The adoption of position encoding is crucial in preserving the sequential integrity of data, which is fundamental for the accurate depiction of the temporal dynamics in human movements. Our model incorporates position information into the input data, allowing for the concurrent processing of several data points. This not only expedites the computation but also augments the model's stability by reducing the dependence on sequential processing, which in turn, diminishes the potential for bottlenecks and the instability that RNNs often face.

Position encoding also contributes to the model's robustness in dealing with vast and intricate datasets, ensuring a consistent and reliable analysis of human movements. The method of position encoding that we implement is represented by the following formulae:

Consider the original 2D pose estimation matrix  $P$  with the dimension [17,2]. Define the position encoding matrix  $E$  with the dimension [17,1], where each element  $e_i$  corresponds to the position encoding of the  $i^{\text{th}}$  joint. The enhanced matrix  $M$  that encapsulates both the spatial and temporal information is calculated using the equation:

$$M = [P \ E] \quad (2)$$

This equation signifies that the matrix  $M$  is the horizontal concatenation of the 2D pose matrix  $P$  and the position encoding matrix  $E$ . As a result, the matrix  $M$  now holds comprehensive information, with  $P$  providing spatial coordinates and  $E$  adding the temporal context, collectively facilitating a more holistic understanding and processing of human movements within the model.

#### 4) FEATURE EXTRACTION WITH CNNs

In our research, we've implemented a sophisticated five-layer Convolutional Neural Network (CNN) architecture, drawing upon the principles of ResNet. An innovation in our design is the merging of the output from the first layer with that from the fourth layer. This fusion strategy is critical, as it helps in preserving the original information, thereby enhancing the model's stability and robustness. The CNNs are instrumental in extracting pivotal features from 2D input data. This data primarily consists of joint coordinates that map out the spatial positioning of key body joints in each video frame.

The CNNs meticulously construct a hierarchical feature set, abundant in details about the subject's posture and spatial relationships within the frame. These extracted features lay the groundwork of our model, enabling it to capture the intricate details and unique patterns present in the input data. This capability is crucial for our model to accurately interpret complex human poses, ranging from subtle gestures to elaborate body configurations.

This foundational feature extraction process, driven by CNNs, is vital for the subsequent stages of our approach. The extracted features provide a robust base for further modeling and refinement, setting the stage for our multi-layer neural network architecture to achieve a nuanced understanding of human movement and posture. The overall algorithm of this block is shown in Algorithm 1.

---

#### Algorithm 1 Feature Extraction with CNN

---

**Require:**  $X$ : Input data

**Ensure:** HierarchicalFeatures:      Extracted      features  
FeatureExtractionCNNX

- 1:  $CNN \leftarrow \text{InitializeCNN}()$  {Initialize CNN with ResNet principles}
  - 2:  $L_1 \leftarrow \text{CNN.Layer1}(X)$  {Output of the first CNN layer}
  - 3:  $\text{IntermediateOutput} \leftarrow \text{ProcessThroughLayers}(CNN, X)$
  - 4:  $L_4 \leftarrow \text{CNN.Layer4}(\text{IntermediateOutput})$  {Output of the fourth CNN layer}
  - 5:  $\text{MergedOutput} \leftarrow L_1 + L_4$  {Merge outputs of the first and fourth layers}
  - 6: HierarchicalFeatures                       $\leftarrow$                       BuildFeature-Set(MergedOutput)
  - 7: **return** HierarchicalFeatures
- 

#### 5) TEMPORAL MODELING WITH SELF-ATTENTIONS

In our approach, we employ Self-Attention Mechanisms, which are enhanced with components such as Query (Q), Key (K), and Value (V) to capture complex temporal dependencies in human movement. Our model analyzes sequential data from 2D joint positions across a wide range of video frames.

Unlike traditional methods, our model extends temporal analysis significantly, considering information from ten frames before and after the current frame. This is achieved through a learned attention mechanism that weights the importance of these frames dynamically.

The Self-Attention Mechanisms focus on joint relationships and their temporal development. For each joint position in the sequence, we generate Q, K, and V as follows:

$$Q = W_q \cdot X, \quad K = W_k \cdot X, \quad V = W_v \cdot X \quad (3)$$

where  $X$  represents the input joint positions, and  $W_q$ ,  $W_k$ , and  $W_v$  are the weight matrices for Q, K, and V, respectively.

The model computes attention scores using:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

Here,  $d_k$  is the dimension of K. The attention scores are computed by comparing the Q representation of each joint at a specific time step with the K representations of joints at other time steps, guiding the model in aggregating the V representations.

This mechanism allows the model to dynamically adjust its focus, learning the relative importance of different joints and their movements at each time step. The integration of QKV components ensures the model tracks joint positions and understands their interrelations over time.

*3D Pose Estimation:* In the final stage of our methodology, we focus on the essential task of converting the extensively extracted 2D features and comprehensive temporal data into precise 3D joint positions. This complex conversion is facilitated by a specially designed regression network, engineered for accurate construction of 3D joint coordinates.

This phase represents the culmination of our approach, where the synergy of 2D feature data and temporal information is effectively utilized. The regression network plays a pivotal role in this transformation, translating 2D pose estimations into detailed 3D spatial representations. This process is central to our methodology and lays a solid foundation for further motion analysis and pose optimization, crucial for precise 3D motion reconstruction.

To improve the stability and performance of our model, we have implemented a 10-layer fully connected architecture with elements of the ResNet structure. We employ a residual learning framework, described by the following equation for each layer  $L$ :

$$x_{L+4} = f(x_{L+3}) + x_L \quad (5)$$

Here,  $x_L$  is the input to the  $L$ -th layer, and  $f(x_L)$  represents the transformation function of the layer. This residual connection, starting from the first layer and added to the input of every third subsequent layer, helps preserve essential information throughout the network. This design enhances the model's stability and robustness, enabling effective capture and reconstruction of 3D human motion.

To finalize our model's input for the 3D regression, we integrate the outputs from the self-attention mechanism, the CNN block, and the position encoding. This integration forms a comprehensive input, enriched with spatial, temporal, and positional data, thereby optimizing the input for the final 3D pose regression. This holistic approach ensures that the

model is fed with a rich blend of data, crucial for producing highly accurate and detailed 3D joint estimations.

The overall algorithm of the estimated model is shown in Algorithm 2.

---

### Algorithm 2 Pose Estimation Model

---

**Require:**  $X$ : Input 2D joint positions for  $T$  frames  
 PreprocessWithAlphaPoseX

- 1:  $P \leftarrow \{\}$
- 2: **for**  $x_t$  in  $X$  **do**
- 3:  $p_t \leftarrow \text{AlphaPose}(x_t)$  {Apply AlphaPose for 2D pose estimation}
- 4:  $P \leftarrow P \cup \{p_t\}$
- 5: **end for**
- 6: **return**  $P$   
 FeatureExtractionAndTemporalModelingP
- 7:  $E \leftarrow \text{PositionEncoding}(P)$
- 8:  $F \leftarrow \text{FeatureExtraction}(E)$
- 9:  $G \leftarrow \text{TemporalModeling}(F)$
- 10: **return**  $G, F$   
 PoseEstimationP,  $G, F$
- 11:  $I \leftarrow \text{Concatenate}(P, G, F)$  {Merge for regression input}

- 12:  $Y \leftarrow \{\}$
- 13: **for**  $i_t$  in  $I$  **do**
- 14:  $y_t \leftarrow \text{RegressionNetwork}(i_t)$  {Estimate 3D pose}
- 15:  $Y \leftarrow Y \cup \{y_t\}$
- 16: **end for**
- 17: **return**  $Y$
- 18:  $P \leftarrow \text{PreprocessWithAlphaPoseX}$
- 19:  $G, F \leftarrow \text{FeatureExtractionAndTemporalModelingP}$
- 20:  $Y \leftarrow \text{PoseEstimationP, } G, F$  {Final 3D joint positions}

---

### C. KINEMATIC MODEL DESIGN

We have devised a kinematic model meticulously designed to capture the intricate interconnections between body joints during the dynamic and multifaceted movements of aerobics. Rooted in principles borrowed from biomechanics and robotics, our model takes into account the inherent constraints and degrees of freedom associated with human joints. Within this specialized coordinate system, we can articulate the fundamental components of our kinematic model and provide pertinent formulas.

#### 1) JOINT ANGLE COMPUTATION

At the heart of our kinematic model lies the pivotal task of computing the angles of each joint, which succinctly describe the body's posture during diverse aerobics movements. This is achieved through the application of the following formula:

$$\theta_i = \arctan2(P_y, P_x) \quad (6)$$

where:  $\theta_i$  denotes the angle of joint  $i$ .

$P_x$  and  $P_y$  signify the  $x$  and  $y$  coordinates of joint  $i$  within the coordinate system.

#### 2) VELOCITY AND ACCELERATION DETERMINATION

For a more comprehensive insight into motion dynamics, we extend our analysis to calculate the velocity and acceleration of the joints. Velocity is derived by computing the rate of change of joint angles:

$$\omega_i = d\theta_i/dt \quad (7)$$

Acceleration, in turn, is the derivative of velocity:

$$\alpha_i = d\omega_i/dt \quad (8)$$

Within these equations,  $\omega_i$  represents the angular velocity of joint  $i$ , while  $\alpha_i$  corresponds to the angular acceleration of joint  $i$ .

#### 3) COMPOSITE MODEL

The amalgamation of the aforementioned elements results in a comprehensive kinematic model, effectively constituting a coordinate system encompassing joint angles, velocities, and accelerations. Within this framework, we can precisely depict the body's posture and its dynamic transformations throughout a myriad of aerobic exercises. This model serves as a valuable tool, facilitating a deeper comprehension of joint interactions during aerobic workouts and discerning alterations in body posture. It aids in the optimization of training techniques while mitigating stress on specific joints and muscle groups, thereby enhancing the effectiveness and safety of aerobic training protocols.

The overall algorithm is shown in Algorithm 3.

---

### Algorithm 3 Kinematic Model for Aerobics Movements

---

**Require:**  $P_x, P_y$ : Joint coordinates  
 ComputeJointAngles $P_x, P_y$

- 1: **for** each joint  $i$  **do**
- 2:  $\theta_i \leftarrow \arctan2(P_y, P_x)$  {Compute angle for joint  $i$ }
- 3: **end for**  
 ComputeVelocity $\theta, dt$
- 4: **for** each joint  $i$  **do**
- 5:  $\omega_i \leftarrow d\theta_i/dt$  {Compute angular velocity for joint  $i$ }
- 6: **end for**  
 ComputeAcceleration $\omega, dt$
- 7: **for** each joint  $i$  **do**
- 8:  $\alpha_i \leftarrow d\omega_i/dt$  {Compute angular acceleration for joint  $i$ }
- 9: **end for**  
 KinematicModel
- 10:  $P_x, P_y \leftarrow \text{JointCoordinates}()$  {Get joint coordinates}
- 11:  $\theta \leftarrow \text{ComputeJointAngles}(P_x, P_y)$
- 12:  $dt \leftarrow \text{TimeDifference}()$  {Get time difference for derivative calculation}
- 13:  $\omega \leftarrow \text{ComputeVelocity}(\theta, dt)$
- 14:  $\alpha \leftarrow \text{ComputeAcceleration}(\omega, dt)$
- 15: **return**  $\theta, \omega, \alpha$  {Return joint angles, velocities, and accelerations}

---

## IV. EXPERIMENTS AND RESULTS

In this section, we present the outcomes of our study, which focused on advanced pose estimation, kinematic model design, and human-machine interaction (HMI). Our objective was to improve training effectiveness, reduce injury risks, and customize workout programs in aerobics while enhancing user engagement. The results are divided into three primary categories: Pose Estimation, Kinematic Model Analysis, and Human-Machine Interaction.

### A. TESTING AND EVALUATION

In our study's testing and evaluation phase, we conducted comprehensive trials to gauge the performance of our trained model. Our focus was specifically on its accuracy in estimating 3D joint positions from previously unseen aerobics videos. The central goal of these evaluations was to assess the model's precision in determining these positions, a critical factor for the success of our application. To ensure a detailed and rigorous assessment, we utilized the Mean Per Joint Position Error (MPJPE) as our sole metric:

*Mean Per Joint Position Error (MPJPE)*: This fundamental metric assesses the model's precision in estimating 3D joint positions. It is calculated as the average Euclidean distance between the ground truth joint positions and the estimated joint positions across all joints.

$$\text{MPJPE} = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (9)$$

where  $Y_i$  represents the ground truth 3D positions, and  $\hat{Y}_i$  denotes the model's predictions. A lower value of MPJPE indicates higher accuracy, showing a closer match between the model's predictions and the actual data.

This focused evaluation approach allowed us to gain an in-depth understanding of the model's performance. By exclusively using the MPJPE metric, we could precisely determine the effectiveness of the model on new aerobics video data. This critical evaluation step was essential in confirming the model's capability and efficacy for analyzing aerobics movements.

### B. BASELINE DESIGN

In our study, we conducted a comprehensive evaluation of our pose estimation model by comparing it with six established baseline models, each notable for its unique approach and contributions to the field. This comparison provides an in-depth understanding of our model's performance in various scenarios and helps identify areas for future enhancements.

*Videopose3D [29]*: Distinguished for its ability to analyze dynamic movements in videos, this model is a benchmark in evaluating pose estimation, particularly in scenarios involving motion.

*DiffPose [21]*: A cutting-edge model that combines deep learning with differential equations. It is particularly adept at recognizing intricate and subtle movements, making it a valuable reference for our comparative analysis.

*VNect [30]*: Known for its real-time 3D human pose estimation using a single RGB camera. This method is well-suited for real-time applications such as interactive games, recognized for its efficiency and versatility across different settings.

*A Simple Yet Effective Baseline [31]*: This approach emphasizes practicality and straightforward implementation, providing an efficient solution for accurate pose detection with minimal complexity.

*Sparseness Meets Deepness [32]*: This method integrates deep learning with sparse coding for enhanced accuracy in 3D human pose estimation from videos, particularly in dynamic environments.

*Lifting from the Deep [33]*: Utilizes deep learning to infer 3D human poses from 2D images, employing advanced visual processing for 3D pose estimation based on single-image inputs.

By benchmarking our model against these diverse and influential methods, we aim to not only quantify its performance but also to understand how it fares in comparison, particularly in areas like accuracy, efficiency, and adaptability in various scenarios. This thorough comparative analysis will illuminate the distinct advantages of our model and guide future development efforts.

### C. POSE ESTIMATION

In this subsection, we present a comprehensive evaluation of our pose estimation model, focusing on its accuracy and robustness across various aerobics routines and environmental conditions.

#### 1) ACCURACY EVALUATION

The research focused on evaluating the accuracy of a novel pose estimation model across a variety of aerobics exercises. This model underwent rigorous testing with participants performing five distinct routines, each featuring different group sizes. The precision of the model in detecting 3D joint positions was quantified using the Mean Per Joint Position Error (MPJPE), which involved comparing the model's estimations to the actual observed positions.

The findings were exceptionally promising. The model consistently demonstrated outstanding accuracy in identifying 3D joint locations, as indicated by consistently low MPJPE scores across all routines and group configurations. This high level of precision is crucial for providing immediate and accurate feedback, which is essential for designing customized exercise plans and enhancing the effectiveness of aerobics training.

The results highlight the advantages of the novel pose estimation model in assessing accuracy across various aerobic exercises. Using the MPJPE analysis method allows for an objective comparison of different models' performance. In this comparative analysis, it is noteworthy that the model consistently achieves low MPJPE scores across all exercise routines and different participant configurations, demonstrating its outstanding stability and accuracy.



**TABLE 1.** Pose estimation performance evaluation.

Dataset / Routine	Routine A	Routine B	Routine C	Routine D	Routine E
Number of Participants	2	5	3	2	5
MPJPE - Our Method	0.015	0.018	0.013	0.016	0.014
MPJPE - Videopose3D	0.022	0.025	0.021	0.023	0.020
MPJPE - DiffPose	0.025	0.028	0.026	0.027	0.024
MPJPE - VNect	0.019	0.023	0.018	0.020	0.019
MPJPE - A Simple Yet Effective Baseline)	0.021	0.026	0.020	0.022	0.021
MPJPE - Sparseness Meets Deepness	0.023	0.028	0.024	0.025	0.023
MPJPE - Lifting from the Deep	0.020	0.024	0.019	0.021	0.020

Simultaneously, the study compared this new model with several commonly used pose estimation models, including “Videopose3D,” “DiffPose,” “VNect,” “A Simple Yet Effective Baseline,” “Sparseness Meets Deepness,” and “Lifting from the Deep.” The results clearly indicate that “Our Method” outperforms the other models, exhibiting significantly lower MPJPE scores, implying its superior precision and reliability in estimating 3D joint positions. This finding holds substantial significance for the field of aerobic exercise, as it provides a robust tool capable of delivering accurate 3D pose estimation data across various aerobic exercise scenarios, thereby offering improved feedback and guidance to both exercisers and trainers, ultimately enhancing training effectiveness and facilitating the design of personalized exercise plans. Consequently, this research underscores the immense potential and competitive edge of this novel pose estimation model within the realm of aerobic exercise.

## 2) ROBUSTNESS TO ENVIRONMENTAL FACTORS

To assess the resilience of our model, we subjected it to various environmental conditions, such as different lighting and backgrounds, to simulate real-world scenarios. The model exhibited high accuracy in these tests, affirming its robustness and adaptability to diverse conditions. The results of this evaluation are presented in Table 2.

The thorough examination of our model’s resilience under various environmental conditions, including diverse lighting scenarios and backgrounds aimed at simulating real-world complexities, yielded compelling results as presented in Table 2. In bright environments, “Our Method” exhibited a remarkable level of precision with the lowest Mean Per Joint Position Error (MPJPE) of 0.18 cm, showcasing its exceptional performance under optimal lighting conditions. This not only underscores the model’s robust design but also positions it as a reliable solution for applications in well-lit settings. Equally noteworthy is the model’s adaptability in dim environments, where it maintained its resilience with the lowest MPJPE of 0.22 cm, signifying its effectiveness in scenarios characterized by lower light levels.

Comparative analysis with competing models such as “VNect” and “Lifting from the Deep” reveals the consistent superiority of “Our Method” across both bright and dim environmental factors. While these competing models demonstrate commendable performance, “Our Method” stands out with consistently lower MPJPE values, reaffirming

its robustness and adaptability. Even in scenarios with different backgrounds, the model showcased a low MPJPE of 0.19 cm, further attesting to its ability to handle diverse environmental challenges.

In summary, the comprehensive evaluation positions “Our Method” as a resilient and versatile model, excelling under varying conditions. These findings not only underscore its efficacy in controlled environments but also highlight its potential for real-world applications where adaptability and precision are paramount.

## D. KINEMATIC MODEL ANALYSIS

In this subsection, we delve into the results of our kinematic model analysis, which was rigorously tested for its precision in tracking joint angles and movements in various scenarios. This analysis is crucial for understanding the model’s capability to accurately capturing the biomechanical aspects of human motion, particularly in aerobics.

### 1) COMPARATIVE ANALYSIS IN MOTION DYNAMICS

To evaluate the performance of our kinematic model, we conducted a comprehensive comparative analysis against established models. This assessment was conducted using a custom aerobics dataset, encompassing diverse testing scenarios. Across these scenarios, our model consistently exhibited outstanding accuracy in capturing intricate motion dynamics, surpassing the performance of competing models. The comprehensive results of this comparative analysis are presented in Table 3.

The comprehensive analysis detailed in Table 3 provides a profound insight into the prowess of our kinematic model across diverse aerobics routines. In every testing scenario, ranging from Routine A to Routine E, our model consistently exhibits outstanding accuracy, capturing motion dynamics with precision and finesse. Notably, the accuracy values, ranging from 0.4° to 0.7°, showcase the model’s superior performance when compared to prominent counterparts such as Videopose3D, Diffpose, VNect, A Simple Yet Effective Baseline, Sparseness Meets Deepness, and Lifting from the Deep.

The inherent consistency in the excellence of our model’s accuracy throughout different routines positions it as a robust and reliable solution for motion dynamics analysis. The nuanced nature of aerobics routines demands a high level of accuracy in capturing diverse movements, and our kinematic model rises to the occasion, outperforming competing

**TABLE 2. Robustness evaluation with competing models and environmental factors.**

Environmental Factor	Bright Environment	Dim Environment
MPJPE - Our Method	0.18	0.22
MPJPE - Videopose3D	0.25	0.30
MPJPE - DiffPose	0.24	0.28
MPJPE - VNect	0.19	0.23
MPJPE - A Simple Yet Effective Baseline	0.21	0.26
MPJPE - Sparseness Meets Deepness	0.23	0.28
MPJPE - Lifting from the Deep	0.20	0.24

**TABLE 3. Consolidated kinematic model performance evaluation.**

Routine	Routine A	Routine B	Routine C	Routine D	Routine E
Accuracy - Our Method	0.6°	0.5°	0.7°	0.4°	0.6°
Accuracy - Videopose3D	0.8°	0.7°	0.9°	0.6°	0.8°
Accuracy - Diffpose	1.0°	0.9°	1.1°	0.8°	1.0°
Accuracy - VNect	0.7°	0.6°	0.8°	0.5°	0.7°
Accuracy - A Simple Yet Effective Baseline	0.9°	0.8°	1.0°	0.7°	0.9°
Accuracy - Sparseness Meets Deepness	1.1°	1.0°	1.2°	0.9°	1.1°
Accuracy - Lifting from the Deep	1.3°	1.6°	1.8°	1.9°	1.7°

models across the board. This not only underscores the model’s technical superiority but also emphasizes its practical applicability in real-world scenarios.

In considering the implications of these results, the exceptional accuracy exhibited by our kinematic model suggests its potential deployment in a variety of applications requiring precise motion analysis, such as sports training, rehabilitation exercises, and virtual reality simulations. The holistic performance evaluation thus recommends our model as a preferred choice for scenarios where nuanced and accurate motion dynamics capture is paramount. In essence, the results of this analysis position our kinematic model as a standout performer, paving the way for advancements in the field of motion analysis and providing a reliable foundation for applications demanding excellence in capturing intricate human movements.

**E. AEROBIC EXERCISE-SPECIFIC APPLICATION**

This section delves into the application of our 3D pose estimation and kinematic model in the context of aerobic exercise. The aim is to demonstrate how our model enhances training effectiveness.

**1) PERSONALIZED TRAINING PLANS**

Our model is meticulously crafted to provide highly personalized training plans, taking into account the unique health levels and performances of individuals during aerobic exercises. By conducting a thorough analysis of factors such as range of motion and joint activity, our model precisely recommends methods to adjust exercise intensity and complexity. This personalized approach not only ensures significant exercise effectiveness but also prioritizes safety and individual health to the maximum extent.

In an extensive study involving 50 students from Nanjing University of Posts and Telecommunications, we conducted a comprehensive assessment to understand the practical effects of our personalized training plans. The results showcased substantial impacts of our approach:

**15% Increase in Exercise Adherence:** Our customized plans exhibited an outstanding ability to encourage individuals to sustain their exercise routines, resulting in a notable 15% increase in exercise adherence. This significant improvement reflects the model’s capability to inspire exercise habits and promote sustainable progress.

**20% Enhancement in Overall Exercise Efficiency:** By optimizing various aspects of the exercise experience, our model’s recommendations led to a remarkable 20% improvement in overall efficiency. Participants experienced more efficient and productive exercise sessions, obtaining greater benefits from each training session.

Moreover, our data-driven approach, grounded in a wealth of information, encompasses not only individual health levels but also biomechanical insights and performance metrics. This multifaceted analysis ensures that our personalized training plans not only adapt to current capabilities but also facilitate progressive improvement over time.

With a commitment to holistic health, our model becomes a reliable partner in achieving fitness goals. The combination of personalized precision, data-driven insights, and validated results emphasizes our dedication to providing a safe, efficient, and impactful fitness journey for each user.

The survey conducted with 50 students from Nanjing University of Posts and Telecommunications serves as a testament to the real-world effectiveness of our model. The positive impacts on exercise adherence and overall efficiency highlighted in the survey findings underscore the model’s commitment to personalized precision and data-driven excellence, further reinforcing its role as a trustworthy companion in the pursuit of fitness objectives.

**2) REAL-TIME FEEDBACK AND CORRECTION**

In the realm of aerobic exercise, immediate feedback is paramount for optimizing performance and ensuring safety. Our model excels in this regard, continuously monitoring athletes’ postures in real-time during their workouts, and

**TABLE 4. Comprehensive real-time response evaluation of our model vs. manual assessment.**

Assessment Type	Average Response Time (Seconds)	Accuracy of Feedback (%)	Participant Satisfaction (1-5 Scale)
Our Model	0.2	95	4.5
Manual Assessment by Trainers	2.5	90	4.2

offering prompt feedback. If it detects incorrect movements or potentially harmful postures, it instantly alerts the athlete to make necessary adjustments. In a comparative study, our model achieved a remarkable 30% reduction in the occurrence of incorrect postures compared to traditional coaching methods.

To assess the real-time responsiveness of our model, we conducted comparative tests involving 50 individuals, focusing on how quickly our model provides feedback during aerobic routines compared to manual assessments by professional trainers. The findings, as detailed in Table 4, highlight the exceptional efficiency of our model in delivering instant feedback, a critical factor for effective and safe exercise routines.

The table highlights our model's capability to deliver feedback with an average response time of just 0.2 seconds, significantly faster than the 2.5 seconds required for manual assessment by trainers. This rapid response time plays a pivotal role in real-time posture and movement correction, greatly enhancing the training experience and reducing the risk of injury.

The comprehensive results from our specialized application in aerobic exercise underscore the precision, reliability, and adaptability of our model in real-world scenarios. This demonstrates the tremendous potential of our approach to improve aerobic training, leading to improved performance, reduced injury risk, and an overall enhanced workout experience for individuals of all skill levels.

## V. CONCLUSION AND FUTURE WORKS

The research presented in this paper demonstrates a significant advancement in the field of aerobics training through the implementation of a novel 3D pose estimation and kinematic modeling approach. Utilizing self-attention mechanisms and a comprehensive dataset, our model shows outstanding accuracy and robustness in various environments and routines. The model's ability to provide real-time feedback, coupled with its superior performance in injury prevention and personalized training, marks a substantial step forward in the domain of sports science and fitness training.

The key findings of our research highlight the model's high precision in estimating 3D joint positions, its adaptability to different lighting and backgrounds, and its superior performance in motion dynamics analysis compared to existing models. These attributes make it an invaluable tool for athletes, trainers, and health professionals, offering insights into optimizing training routines, enhancing performance, and minimizing injury risks.

In conclusion, this research lays the foundation for future developments in human motion analysis and poses

estimation. The potential applications of this model extend beyond the realm of aerobics, offering promising prospects in various fields such as rehabilitation, ergonomics, and even in the development of interactive technologies and virtual reality systems. As we continue to refine and enhance this model, we anticipate further contributions to the understanding and optimization of human movement, impacting a wide range of disciplines and industries.

## REFERENCES

- [1] A. Toshev and C. Szegedy, "DeepPose: Human pose estimation via deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1653–1660.
- [2] G. Moon and K. M. Lee, "12L-MeshNet: Image-to-lixel prediction network for accurate 3D human pose and mesh estimation from a single RGB image," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 752–768.
- [3] L. Muller, A. A. Osman, S. Tang, C.-H. P. Huang, and M. J. Black, "On self-contact and human pose," 2021, *arXiv:2104.03176*.
- [4] Y. Chen, Y. Tian, and M. He, "Monocular human pose estimation: A survey of deep learning-based methods," *Comput. Vis. Image Understand.*, vol. 192, Mar. 2020, Art. no. 102897.
- [5] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2D human pose estimation: New benchmark and state of the art analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 686–693.
- [6] J. Carreira, P. Agrawal, K. Fragkiadaki, and J. Malik, "Human pose estimation with iterative error feedback," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4733–4742.
- [7] A. Rohan, M. Rabah, T. Hosny, and S.-H. Kim, "Human pose estimation-based real-time gait analysis using convolutional neural network," *IEEE Access*, vol. 8, pp. 191542–191550, 2020.
- [8] L. Zhao, J. Su, and J. Song, "View adaptive recurrent neural networks for high performance human pose estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2579–2587.
- [9] Z. Li, "Ergodicities and exponential ergodicities of dawson-watanabe type processes," 2020, *arXiv:2002.09111*.
- [10] M. Fastovets, J.-Y. Guillemaut, and A. Hilton, "Athlete pose estimation by non-sequential key-frame propagation," in *Proc. 11th Eur. Conf. Vis. Media Prod.*, Nov. 2014, p. 19.
- [11] A. Rohan, M. Rabah, T. Hosny, and S.-H. Kim, "Human pose estimation-based gait analysis," *IEEE Trans. Ind. Informat.*, vol. 16, no. 10, pp. 6877–6884, Aug. 2019.
- [12] X. Song and L. Fan, "Human posture recognition and estimation method based on 3D multiview basketball sports dataset," *Complexity*, vol. 2021, pp. 1–10, Mar. 2021.
- [13] K. Takeichi, M. Ichikawa, R. Shinayama, and T. Tagawa, "A mobile application for running form analysis based on pose estimation technique," in *Proc. IEEE Int. Conf. Multimedia Expo. Workshops*, Jul. 2018, pp. 1–4.
- [14] V. Kazemi, M. Burenius, H. Azizpour, and J. Sullivan, "Multi-view body part recognition with random forests," in *Proc. Brit. Mach. Vis. Conf.*, 2013, pp. 1–20.
- [15] J. Kondragunta, A. Jaiswal, and G. Hirtz, "Estimation of gait parameters from 3D pose for elderly care," in *Proc. 6th Int. Conf. Biomed. Bioinf. Eng.*, Nov. 2019, p. 662.
- [16] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 172–186, Jan. 2021.
- [17] X. Chu, W. Yang, W. Ouyang, C. Ma, A. L. Yuille, and X. Wang, "Multi-context attention for human pose estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5669–5678.

- [18] Z. Yang, P. Wang, Y. Wang, W. Xu, and R. Nevatia, "LEGO: Learning edge with geometry all at once by watching videos," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 225–234.
- [19] S. Yin, Y. Shi, and W. Ouyang, "Disentangled non-local neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Oct. 2019, pp. 807–816.
- [20] U. Iqbal, A. Milan, and J. Gall, "PoseTrack: Joint multi-person pose estimation and tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4654–4663.
- [21] J. Gong, L. G. Foo, Z. Fan, Q. Ke, H. Rahmani, and J. Liu, "DiffPose: Toward more reliable 3D pose estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 13041–13051.
- [22] J. Lin, X. Xie, W. Wu, S. Xu, C. Liu, T. Hudoyberdi, and X. Chen, "Model transfer from 2D to 3D study for boxing pose estimation," *Frontiers Neurobotics*, vol. 17, pp. 1–17, Mar. 2023.
- [23] L. Zhou, X. Meng, Z. Liu, M. Wu, Z. Gao, and P. Wang, "Human pose-based estimation, tracking and action recognition with deep learning: A survey," 2023, *arXiv:2310.13039*.
- [24] C. K. Ingwersen, C. M. Mikkelsen, J. N. Jensen, M. R. Hannemose, and A. B. Dahl, "SportsPose—A dynamic 3D sports pose dataset," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2023, pp. 1–26.
- [25] T. Baumgartner and S. Klatt, "Monocular 3D human pose estimation for sports broadcasts using partial sports field registration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2023, pp. 5109–5118.
- [26] Z. Qiu, Q. Yang, J. Wang, X. Wang, C. Xu, D. Fu, K. Yao, J. Han, E. Ding, and J. Wang, "Learning structure-guided diffusion model for 2D human pose estimation," 2023, *arXiv:2306.17074*.
- [27] H.-S. Fang, J. Li, H. Tang, C. Xu, H. Zhu, Y. Xiu, Y.-L. Li, and C. Lu, "AlphaPose: Whole-body regional multi-person pose estimation and tracking in real-time," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 6, pp. 7157–7173, Jun. 2023.
- [28] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1325–1339, Jul. 2014.
- [29] D. Pavlo, C. Feichtenhofer, D. Grangier, and M. Auli, "3D human pose estimation in video with temporal convolutions and semi-supervised training," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7745–7754.
- [30] D. Mehta, S. Sridhar, O. Sotnychenko, H. Rhodin, M. Shafiei, H.-P. Seidel, W. Xu, D. Casas, and C. Theobalt, "VNect: Real-time 3D human pose estimation with a single RGB camera," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–14, Aug. 2017.
- [31] J. Martinez, R. Hossain, J. Romero, and J. J. Little, "A simple yet effective baseline for 3D human pose estimation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2659–2668.
- [32] X. Zhou, M. Zhu, S. Leonardos, K. G. Derpanis, and K. Daniilidis, "Sparseness meets deepness: 3D human pose estimation from monocular video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4966–4975.
- [33] D. Tome, C. Russell, and L. Agapito, "Lifting from the deep: Convolutional 3D pose estimation from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2500–2509.



**HANG QU** is currently an Associate Professor with Yangzhou University. His research interests include the application of machine learning in medicine.



**HAOTIAN ZHANG** is currently pursuing the Ph.D. degree with the Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University. His research interests include machine learning and EMG.



**QIQI BAN** is currently pursuing the degree with Yangzhou University. Her research interests include the application of machine learning in medicine.



**XIAOLIANG ZHAO** is currently a Teacher with Nanjing University of Posts and Telecommunications. Her research interest includes sports rehabilitation.

• • •