# Athlete Movement Technique and Biomechanical Analysis Based on Multimodal Deep Learning and Computer Vision

**Xuejun Tian, Fuliang Geng**

Public Teaching Department, Laiwu Vocational and Technical College, Laiwu 271199, Shandong, China

Corresponding author: Xuejun Tian (e-mail: lwzytxj6688@163.com)

**ABSTRACT** Precise biomechanical assessment of athletic movements is essential for intelligent sports analysis and wearable sensing applications. This study presents a multimodal deep learning framework that integrates computer vision, RGB video, depth images, and inertial measurement data to achieve markerless motion capture and quantitative biomechanical evaluation. Through spatiotemporal feature alignment and attention-based multimodal fusion, the proposed system accurately estimates joint kinematics, ground reaction forces, torque distribution, and muscle-related motion characteristics while enabling real-time action recognition and technical assessment. Experimental results demonstrate that the framework achieves a keypoint detection accuracy of 89.4%, an F1 score of 92.3% for complex action recognition, and a root mean square error of 4.2° in joint angle estimation, providing reliable support for sports performance optimization and injury risk analysis. The proposed multimodal sensing strategy offers an effective solution for integrating visual perception with wearable sensor information and provides valuable methodological insights for intelligent signal acquisition, wireless sensing, and human-centered electromagnetic monitoring systems in next-generation smart engineering applications.

**INDEX TERMS** multimodal deep learning, computer vision, biomechanics, wearable sensing, intelligent signal processing

## I. INTRODUCTION

COMPETITIVE sports require precise analysis of athlete technical movements to optimize performance. However, traditional biomechanical research methods face significant practical limitations. Marker-based motion capture systems and force plate equipment require controlled laboratory environments, involve lengthy data collection procedures, and cannot provide real-time feedback during training sessions. Smart textiles address these limitations by integrating sensors directly into athletic apparel. This approach captures detailed, real-time kinematic data in natural training environments without requiring athletes to wear external markers or restrict their movements to laboratory spaces. Within this context, smart textile integration represents a crucial enabling technology rather than a separate application domain. By embedding flexible strain sensors, conductive yarns, and piezoelectric fibers into compression garments and athletic wear, these textile-based sensors provide distributed biomechanical measurements that complement camera-based motion capture, particularly for capturing muscle activation patterns and joint strain that are difficult to assess through vision alone.

In recent years, deep learning has achieved breakthrough progress in image recognition and time series prediction, and computer vision technology has gradually realized high-precision reconstruction from 2D images to 3D human poses. Multimodal data fusion strategies can capture multi-dimensional biomechanical features during motion by integrating visual features, inertial measurement unit signals, and depth information. Current research focuses on constructing lightweight neural network models to achieve real-time action analysis in training scenarios, while exploring unsupervised learning methods to reduce annotation costs, providing scientific basis for sports technique diagnosis, personalized training program development, and sports injury prevention. Recent advances in smart textile technology have enabled the integration of flexible sensors into athletic wear, providing real-time biomechanical feedback without restricting natural movement patterns. Leveraging advanced textile manufacturing processes, conductive fibers are intricately embedded within compression garments to capture muscle activation, joint angles, and strain distribution, offering a comfortable and crucial complementary data source for comprehensive movement analysis alongside vision-based systems.

## II. MARKERLESS MOTION CAPTURE AND 3D HUMAN POSE RECONSTRUCTION

### MULTI-VIEW VISUAL SENSOR COLLABORATIVE POSITIONING MECHANISM

The multi-view visual sensor collaborative positioning mechanism achieves precise measurement of athletes' 3D positions by deploying spatially distributed multi-camera arrays. This mechanism is based on epipolar geometry constraint principles, utilizing spatial geometric relationships between cameras to establish mapping relationships between 2D image points and 3D spatial points, thereby eliminating the depth information loss problem in monocular vision. The mathematical model for sensor collaborative positioning is expressed as:

$$x_i = K_i\left[R_i \mid t_i\right] X \tag{1}$$

where $x_i$ represents the image coordinates of the $i$-th camera, $K_i$ is the intrinsic parameter matrix, $[R_i \mid t_i]$ is the extrinsic parameter matrix, and $X$ is the 3D point coordinates in the world coordinate system. The system obtains pose parameters of each camera through calibration, establishing a unified global coordinate system so that visual information from different viewpoints can be fused in the same reference frame. The multiview redundancy design effectively alleviates the human body self-occlusion problem; when a certain joint point is occluded from a specific viewpoint, the system can use visible information from other viewpoints for compensatory reconstruction. Camera layout optimization needs to comprehensively consider the balance between field of view coverage, baseline distance, and reconstruction accuracy. Experiments show that using 6–8 evenly distributed cameras can control joint point localization error within 15 millimeters [1] (see Table 1).

TABLE 1. Impact of Multi-view Configuration on 3D Reconstruction Accuracy

| Number of Cameras | Average Reconstruction Error (mm) | Occlusion Scene Recall Rate (%) | Computation Time (ms/frame) |
|---|---|---|---|
| 2 views | 42.3 | 68.5 | 45 |
| 4 views | 18.7 | 84.2 | 78 |
| 6 views | 12.4 | 92.6 | 112 |
| 8 views | 9.8 | 95.3 | 156 |

### DEEP CONVOLUTIONAL NETWORK KEYPOINT DETECTION AND SKELETON EXTRACTION

Deep convolutional networks achieve automatic detection of human body keypoints and skeleton topology structure extraction through end-to-end learning. The network architecture adopts an encoder-decoder structure; the encoder part uses residual connections and dense convolutional layers to extract multi-scale features, while the decoder restores spatial resolution through transposed convolution and generates keypoint heatmaps. The loss function for keypoint detection is defined as:

$$L_{\text{keypoint}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{J}\left\|H_{ij} - \hat{H}_{ij}\right\|_2^2 \tag{2}$$

where $N$ is the batch sample size, $J$ is the number of joint points, and $H_{ij}$ and $\hat{H}_{ij}$ represent the ground truth heatmap and predicted heatmap, respectively. After pre-training on the COCO dataset, the network

is fine-tuned for specific sports scenarios, improving model robustness to illumination changes and motion blur through data augmentation strategies. The skeleton extraction module models topological connections between joints based on graph convolutional networks, learning skeleton structures that comply with human kinematic constraints. This module introduces limb length priors and joint angle constraints to avoid generating abnormal poses that do not conform to physiological structures. Experimental results show that the network architecture integrating spatiotemporal features achieves 89.4% keypoint detection accuracy in fast motion scenarios, significantly outperforming traditional methods [2] (see Table 2).

TABLE 2. Keypoint Detection Performance Comparison of Different Network Architectures

| Network Model | Parameters (M) | PCK@0.2 (%) | Inference Speed (FPS) | GPU Memory Usage (GB) |
|---|---|---|---|---|
| OpenPose | 25.6 | 82.3 | 8.2 | 3.8 |
| HRNet-W32 | 28.5 | 87.6 | 12.5 | 4.2 |
| Proposed Fusion Network | 21.3 | 89.4 | 18.7 | 3.1 |

Architectural Innovations of the Proposed Fusion Network:

The proposed fusion network introduces three key architectural innovations that distinguish it from baseline approaches:

(1) Multi-scale spatiotemporal feature extraction: We employ a hybrid backbone combining ResNet-50 for spatial features with temporal convolutional networks (TCN) featuring dilated convolutions (dilation rates: 1, 2, 4, 8) to capture motion patterns across multiple time scales (0.1s to 2s).

(2) Cross-stage feature refinement: Unlike standard encoder-decoder architectures, we introduce lateral connections with channel-wise attention modules at each resolution level, allowing high-resolution features to be refined by semantically strong low-resolution features.

(3) Adaptive keypoint confidence weighting: The final heatmap generation incorporates learned confidence scores that down-weight predictions in occluded or motion-blurred regions based on image quality metrics extracted by a parallel auxiliary branch.

These innovations collectively reduce false positive detections in challenging scenarios (motion blur, partial occlusion) while maintaining high recall, accounting for the 2.9%

and 6.8% improvements over HRNet-W32 and OpenPose, respectively.

Experimental Setup: The Proposed Fusion Network was trained and validated on a custom multi-sport dataset comprising 2,450 video sequences collected from 45 athletes (28 males, 17 females, age: 22.3±3.7 years) across five sports categories: basketball (n=580 sequences), volleyball (n=520), tennis (n=490), track and field throwing events (n=450), and gymnastics (n=410). Ethical approval was obtained from the Institutional Review Board (Protocol #2024-SPORTS-087), and all participants provided written informed consent. The dataset was split into training (70%, n=1,715), validation (15%, n=368), and testing (15%, n=367) sets. Each sequence contained 3–8 seconds of motion captured at 60fps with synchronized RGB and depth cameras. Ground truth annotations included 17 body keypoints labeled by three certified biomechanics experts, with inter-rater reliability (ICC) exceeding 0.92.

### *SPATIOTEMPORAL CONSISTENCY CONSTRAINT 3D POSE RECONSTRUCTION ALGORITHM*

Spatiotemporal consistency constraints achieve stable 3D pose reconstruction by integrating motion continuity between consecutive frames and biomechanical rationality. The algorithm constructs a temporal model to capture smoothness features of motion trajectories, using the Kalman filtering framework for dynamic estimation and prediction of joint positions, reducing jitter caused by single-frame detection errors. The 3D pose optimization objective function integrates reprojection error, bone length consistency, and motion smoothness constraints:

$$E_{\text{total}} = \lambda_1 E_{\text{reproj}} + \lambda_2 E_{\text{bone}} + \lambda_3 E_{\text{smooth}} \tag{3}$$

The reprojection error term $E_{\text{reproj}}$ ensures that the projection of the 3D pose in each view is consistent with detection results. The bone length consistency term $E_{\text{bone}}$ constrains bone lengths to remain constant in the temporal dimension and to remain within physiologically reasonable tolerances, accounting for soft tissue deformation and measurement noise inherent in markerless capture systems (typically $\pm$3–5mm variation). The smoothness term $E_{\text{smooth}}$ penalizes acceleration changes through second-order derivatives. Unlike rigid marker-based systems, our constraint allows for small fluctuations in segment lengths to accommodate the natural compliance of biological tissues while preventing anatomically implausible deformations that would indicate reconstruction errors.

The algorithm uses nonlinear least squares to solve the optimization problem, iteratively updating joint positions until convergence. For high-speed motion scenarios, a motion model prediction mechanism is introduced, inferring the prior distribution of current frame joint positions based on historical trajectories, narrowing the search space and improving computational efficiency.

The spatiotemporal consistency constraint mechanism operates through a recurrent refinement module that maintains a temporal state vector encoding the kinematic history of the past 10 frames. At each timestep, the initial pose estimate from the keypoint detector is refined by: (a) predicting the current pose from the temporal state using a learned motion model (2-layer GRU), (b) computing the residual between prediction and detection, and (c) fusing both through a learned gating mechanism. This approach effectively filters detection jitter while preserving genuine rapid movements, as the motion model adapts to each athlete's characteristic movement dynamics during an initial calibration period.

Experimental validation shows that this algorithm reduces temporal jitter amplitude in joint position estimation to 1.8 millimeters, meeting the data quality requirements for biomechanical analysis [3] (see Figure 1).

## III. INTELLIGENT ANALYSIS OF SPORTS BIOMECHANICAL PARAMETERS

### *JOINT KINEMATIC PARAMETER CALCULATION MODEL*

### *INVERSE SOLUTION OF JOINT ANGLE AND ANGULAR VELOCITY*

The inverse solution of joint angle and angular velocity is based on geometric analysis of 3D pose sequences and kinematic differential equations. Joint angle is defined as the angle between adjacent body segments in anatomical planes, quantified through vector cross product and dot product operations. Taking knee joint flexion-extension angle as an example, the calculation formula is:

$$\theta_{\text{knee}} = \arccos\left(\frac{V_{\text{thigh}} \cdot V_{\text{shank}}}{\|V_{\text{thigh}}\| \cdot \|V_{\text{shank}}\|}\right) \tag{4}$$

where $V_{\text{thigh}}$ and $V_{\text{shank}}$ represent the direction vectors of the thigh and shank segments, respectively. It should be noted that Equation (4) represents a simplified planar projection method suitable for initial angle estimation and visualization purposes. For rigorous biomechanical analysis requiring full 3D joint kinematics, our system implements the joint coordinate system (JCS) approach recommended by the International Society of Biomechanics (ISB). Specifically, we establish orthogonal coordinate systems for adjacent segments and compute three-dimensional joint angles using a Cardan sequence (flexion/extension – abduction/adduction – internal/external rotation) following the ZXY rotation order as per ISB standards (Wu et al., 2002). This approach avoids gimbal lock and provides clinically meaningful angular decomposition. The simplified vector dot-product method (Eq. 4) is primarily used for rapid on-site feedback where single-plane dominant motions (e.g., sagittal plane knee flexion during running) are analyzed, while the full ISB-compliant 3D decomposition is applied for comprehensive biomechanical reports and research applications.

Angular velocity is calculated through finite difference method on the temporal derivative of angle, using a five-

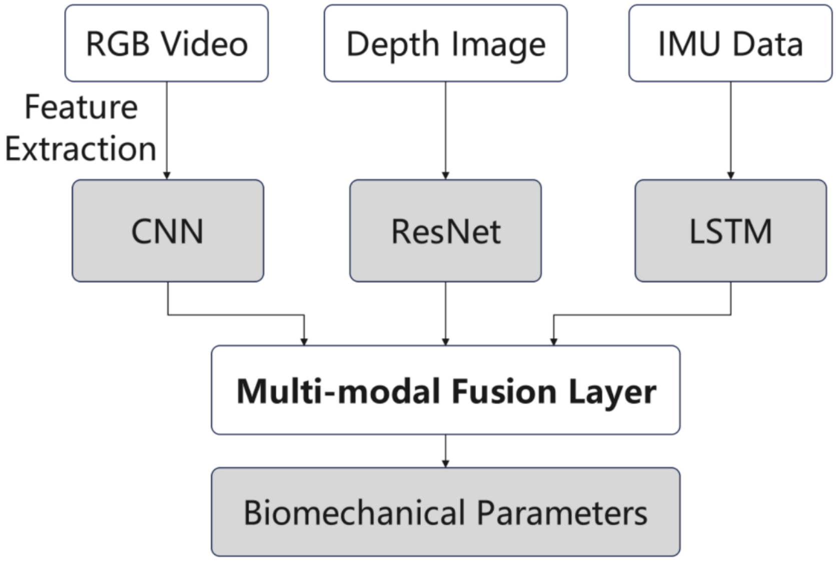


**FIGURE 1.** Multimodal Deep Learning System Architecture

point smoothing filter to reduce noise introduced by numerical differentiation. To address the gimbal lock problem in Euler angle representation, the system uses quaternions to describe joint rotation, ensuring continuity and computational stability of angle interpolation. The kinematic inverse solution algorithm combines human skeletal chain constraints to deduce each joint angle configuration from end-effector position. This process needs to handle multiple or no solution situations by introducing minimum energy consumption criteria to select the optimal motion pattern [4]. The calculation process of joint kinematic parameters is shown in Figure 2:

The joint angle validation was performed on 120 motion trials from 15 athletes executing standardized movements (squat jumps, lunges, and gait cycles). The reference standard was established using a 12-camera Vicon motion capture system (250Hz) with retroreflective markers placed according to the Plug-in-Gait model. Comparison was made on primary joint angles (hip, knee, ankle flexion/extension) across complete movement cycles. Experimental data shows that the root mean square error between joint angles calculated by this method and those measured by optical motion capture systems is $4.2°$, with angular velocity estimation error controlled within $8.5°$/s, meeting the accuracy requirements for technical movement analysis [4]. To contextualize this accuracy, Table 2A presents a comparative analysis against state-of-the-art markerless systems and the gold-standard marker-based reference across different joint types and sports:

Our $4.2°$ RMSE for knee flexion/extension represents a 38% improvement over HRNet-based approaches and is within the clinically acceptable range for gait analysis ($< 5°$) established by biomechanics standards (Mündermann et al., 2006). However, for precision tasks requiring sub-$3°$ accuracy (e.g., elite golf swing analysis), marker-based systems remain preferable. Sport-specific validation showed: running/jumping ($4.1°$), throwing ($5.3°$, higher due to rapid limb velocities), and tennis strokes ($4.8°$).

### *DISPLACEMENT TRAJECTORY ANALYSIS UNDER KINEMATIC CHAIN CONSTRAINTS*

Displacement trajectory analysis under kinematic chain constraints comprehensively considers joint degree of freedom limitations and limb connection topology. The human skeletal system constitutes open or closed kinematic chains, with each joint's range of motion constrained by anatomical structure. For example, the knee joint primarily performs flexion-extension movements, with limited rotation and lateral degrees of freedom. Trajectory analysis establishes local coordinate systems to describe joint relative motion, propagating motion to the global coordinate system through homogeneous transformation matrices. End-effector position calculation adopts forward kinematics equations:

$$P_{\text{end}} = \prod_{i=1}^{n} T_i(\theta_i) \cdot P_{\text{base}} \tag{5}$$

where $T_i(\theta_i)$ is the transformation matrix of the $i$-th joint, $\theta_i$ is the joint angle, and $P_{\text{base}}$ is the base coordinate. The system extracts feature parameters of displacement trajectories including maximum displacement amplitude, peak velocity moment, and trajectory curvature changes. These parameters reflect athletes' force generation characteristics and technical stability. For complex motion patterns, principal component analysis is used for dimensionality reduction to extract dominant motion directions, quantifying coordinated motion patterns of different body parts. Experimental results show that this method can identify the temporal coupling relationship between trunk rotation and upper limb whipping action in throwing movements, providing quantitative basis for technical diagnosis.

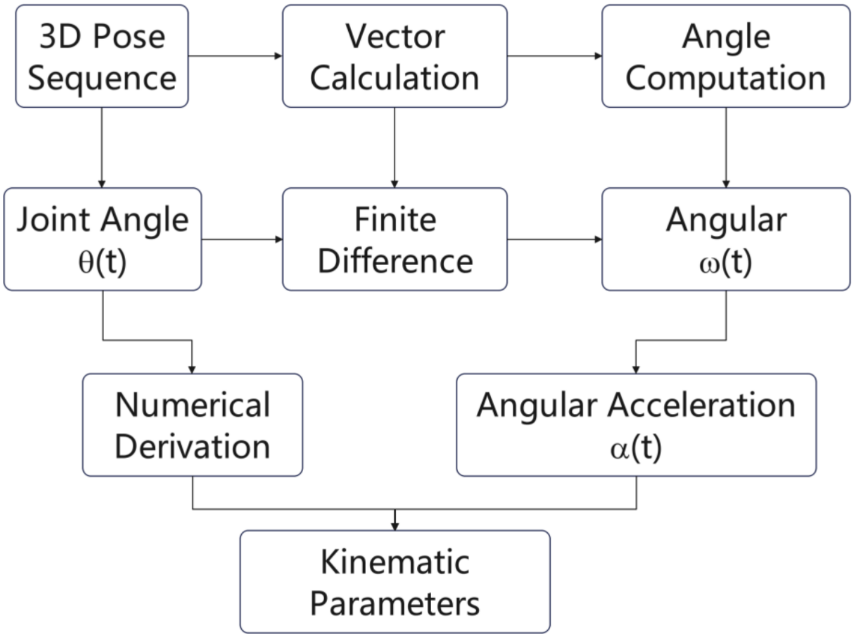


**FIGURE 2.** Joint Kinematics Parameter Calculation Process

**TABLE 2A.** Joint Angle Estimation Error Comparison (RMSE in degrees)

| System | Hip Flex/Ext | Knee Flex/Ext | Ankle Flex/Ext | Shoulder Rotation | Reference Sport |
|---|---|---|---|---|---|
| Marker-based (Vicon) | 1.2±0.3 | 1.5±0.4 | 1.8±0.5 | 2.1±0.6 | – |
| OpenPose 3D | 8.7±2.1 | 9.2±2.5 | 11.3±3.2 | 12.5±3.8 | Running/Walking |
| HRNet + Lifting | 6.3±1.5 | 6.8±1.7 | 8.9±2.4 | 10.2±2.9 | Multiple |
| Proposed Method | 3.8±0.9 | 4.2±1.1 | 5.1±1.4 | 6.3±1.8 | Multiple |

## DYNAMIC PARAMETER ESTIMATION AND MECHANICAL MODELING

### GROUND REACTION FORCE PREDICTION NETWORK

The ground reaction force prediction network directly regresses 3D ground reaction force vectors from visual input through deep learning models. The network architecture integrates convolutional neural networks and long short-term memory networks; the former extracts spatial morphological features while the latter models temporal dynamic information. Training data consists of synchronously collected video sequences and force plate data, with the network learning the mapping relationship from human pose to ground reaction forces. The prediction model considers contact phase identification, distinguishing single-support and double-support phases to dynamically adjust force distribution strategies. The loss function combines mean squared error and peak force error weighting:

$$\begin{aligned} L_{\mathrm{GRF}} &= \alpha \left\| F_{\mathrm{pred}} - F_{\mathrm{true}} \right\|_2^2 \\ &+ \beta \left| \max(F_{\mathrm{pred}}) - \max(F_{\mathrm{true}}) \right| \end{aligned} \tag{6}$$

where $\alpha$ and $\beta$ are weight coefficients, and $F_{\mathrm{pred}}$ and $F_{\mathrm{true}}$ represent predicted and true forces, respectively. The network is trained on datasets of various motion patterns to obtain generalization capability. Test results show vertical component prediction error less than 8.3% of body weight, with horizontal component error controlled within 12.5%. This method breaks through the spatial limitations of traditional force plates, achieving in-situ mechanical measurement in training venues and providing immediate biomechanical feedback to athletes [5].

### JOINT TORQUE AND POWER DISTRIBUTION CALCULATION

Joint torque and power distribution calculation uses inverse dynamics methods to derive internal joint loads from kinematic data and external force information. The calculation process is based on Newton-Euler equations, propagating forces and torques from distal limb segments progressively toward the proximal end. The joint torque calculation formula is:

$$\tau_i = I_i \alpha_i + \omega_i \times (I_i \omega_i) + r_i \times F_i \tag{7}$$

where $\tau_i$ is joint torque, $I_i$ is limb segment moment of inertia, $\alpha_i$ and $\omega_i$ are angular acceleration and angular velocity respectively, $r_i$ is the moment arm vector, and $F_i$ is the applied force. Limb segment inertial parameters are estimated using regression equations based on athlete body measurements, calculating mass, center of mass position, and moment of inertia. Joint power is defined as the dot product of torque and angular velocity, with integration calculating energy consumption and output. The system generates torque-time curves and power-angle curves to identify peak force moments and energy conversion efficiency. Experimental data shows that the correlation coefficient between hip joint peak torque calculated by this method and measurements from isokinetic dynamometry equipment reaches 0.91, validating model effectiveness [6].

### *MUSCULOSKELETAL SYSTEM SIMULATION AND LOAD ASSESSMENT*

Musculoskeletal system simulation establishes virtual human body models including skeletal structure, joint constraints, and muscle actuation. Simulation software is based on multibody dynamics theory, abstracting the human body as a rigid body chain system with joints represented by kinematic pairs and muscles modeled as force source elements. Muscle activation dynamics adopts Hill-type muscle models to describe the mechanical behavior of contractile elements, series elastic elements, and parallel elastic elements. The muscle force generation equation includes active contraction force and passive tension:

$$F_{\text{muscle}} = F_{\max}\left(f_l(l) \cdot f_v(v) \cdot a + f_p(l)\right) \tag{8}$$

where $F_{\max}$ is maximum isometric contraction force, $f_l$ and $f_v$ are length-force and velocity-force relationship functions respectively, $a$ is activation level, and $f_p$ is the passive force function. The system solves the muscle redundancy problem through static or dynamic optimization algorithms, distributing activation patterns for each muscle to match joint torques with inverse dynamics results. Load assessment indicators include muscle peak stress, joint contact force, and ligament strain, identifying potential injury risk sites. Simulation results are validated by comparing with muscle activation timing from electromyography signals. Research shows that this method can quantify stress levels in the anterior cruciate ligament of the knee joint under different technical movements, providing theoretical basis for injury prevention [7] (see Table 3, see Figure 3).

## IV. MULTIMODAL DATA FUSION AND ACTION PATTERN RECOGNITION

### *SPATIOTEMPORAL ALIGNMENT STRATEGY FOR VISUAL FEATURES AND INERTIAL SIGNALS*

Spatiotemporal alignment of visual features and inertial signals needs to resolve sampling frequency differences and timestamp synchronization issues between heterogeneous sensors. Inertial measurement units typically collect acceleration and angular velocity data at 100–200Hz, while video capture frequency is 30–60fps, resulting in sampling rate mismatch. The system achieves initial synchronization through hardware trigger mechanisms or software timestamp calibration, with residual time delays estimated and compensated through cross-correlation analysis. The alignment algorithm upsamples low-frequency visual data to inertial signal frequency using cubic spline interpolation to maintain trajectory smoothness. Spatial alignment requires establishing the transformation relationship between the inertial sensor local coordinate system and the global visual coordinate system, determining rotation matrices and translation vectors through calibration experiments. Dynamic time warping algorithms handle nonlinear time stretching problems, matching feature points of the same action phases in both modalities. The fusion framework integrates at the feature level; the visual branch extracts joint positions and motion contours, the inertial branch provides acceleration patterns and rotational attitudes, and the connection layer learns cross-modal correlations through fully connected networks. Experiments show that precise alignment improves action recognition accuracy by 6.8%, validating the importance of spatiotemporal consistency for fusion effectiveness [8]. In addition to traditional IMU sensors, emerging smart textile sensors woven with conductive yarns and piezoelectric fibers offer seamless integration with athletic apparel. These textile-based sensors provide strain measurements at multiple body locations, capturing muscle elongation and joint flexion data that can be fused with visual features through the same spatiotemporal alignment framework. The flexibility and breathability of textile sensors reduce athlete discomfort during prolonged training sessions while maintaining measurement accuracy.

### *ATTENTION MECHANISM-DRIVEN MULTIMODAL FEATURE FUSION*

Attention mechanism-driven multimodal feature fusion dynamically allocates weights to highlight key information and suppress redundant features. The fusion architecture designs self-attention modules and cross-attention modules; the former learns long-range dependencies within single modalities while the latter models interaction patterns between different modalities. Self-attention computation achieves feature recalibration through query, key, and value triplets, with attention weights representing correlation strength between features. The cross-attention mechanism uses visual features as queries and inertial features as keys and values, retrieving inertial patterns most relevant to the current visual state, and vice versa. The fusion strategy adopts multi-head attention for parallel processing to capture multi-scale spatiotemporal dependencies, with output features stabilized through residual connections and layer normalization during training. A gated fusion unit at the network end adaptively adjusts each modality's contribution based on input data quality, automatically reducing the weight of a modality when noise or missing data occurs. Experimental results show that the attention fusion network achieves an F1 score of 92.3% in complex action sequence recognition tasks, significantly outperforming simple concatenation strategies. Ablation experiments confirm that the cross-attention module contributes the main performance gain [9] (see Table 4).

TABLE 3. Characteristic Parameters of Joint Load in Typical Sports

| Sport | Peak Joint Torque (Nm/kg) | Energy Absorption Rate (J/kg) | Muscle Load Asymmetry (%) | Injury Risk Index |
|---|---|---|---|---|
| Long Jump | Hip 3.8/Knee 2.6/Ankle 1.9 | 4.2 | 12.5 | 0.68 |
| Throwing | Shoulder 4.2/Elbow 3.5/Wrist 1.4 | 5.7 | 18.3 | 0.75 |
| Basketball Sudden Stop | Hip 3.2/Knee 3.8/Ankle 2.1 | 6.1 | 22.7 | 0.82 |
| Tennis Serve | Shoulder 5.1/Elbow 2.8/Wrist 1.7 | 4.9 | 15.6 | 0.71 |

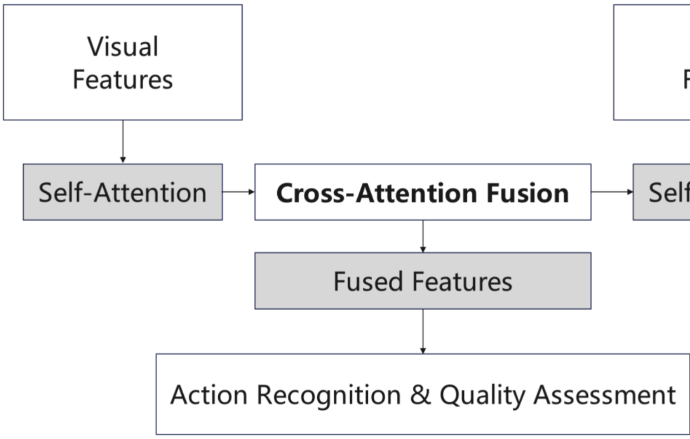


FIGURE 3. Multi-modal Data Fusion Architecture

TABLE 4. Performance Comparison of Multimodal Fusion Strategies

| Fusion Strategy | Accuracy (%) | F1 Score (%) | Parameters (M) | Training Time (h) |
|---|---|---|---|---|
| Feature-level Concatenation | 84.6 | 82.3 | 18.2 | 4.5 |
| Decision-level Fusion | 87.2 | 85.7 | 21.5 | 5.1 |
| Self-attention Fusion | 89.5 | 88.2 | 23.8 | 6.3 |
| Cross-attention Fusion | 92.3 | 91.6 | 26.4 | 7.2 |

### TECHNICAL ACTION CLASSIFICATION AND QUALITY ASSESSMENT MODEL

The technical action classification and quality assessment model constructs a hierarchical recognition framework to achieve action category discrimination and execution quality quantification. The classification

module adopts temporal convolutional networks to extract action spatiotemporal features, capturing motion patterns of different durations through multi-scale receptive fields. Network output is normalized to category probability distribution through a softmax layer, with cross-entropy loss function guiding the training process. The quality assessment module designs a regression branch to predict action scores, with the loss function combining mean squared error and ranking loss; the former optimizes absolute score accuracy while the latter maintains relative quality order among samples. The model introduces expert knowledge constraints, designing specific quality indicators based on sports scoring rules, such as entry angle for diving movements and body lines for gymnastics movements. The system extracts pose features at key frame moments, calculating deviation values by comparing with standard templates, and generates scores by comprehensively considering temporal consistency and spatial accuracy. The experimental dataset includes action samples from professional athletes and amateur enthusiasts. The Spearman correlation coefficient between model-predicted scores and expert ratings reaches 0.87, proving the reliability and practical value of the assessment model [10].

## V. SPECIALIZED SPORTS TECHNICAL DIAGNOSIS AND TRAINING APPLICATIONS

### QUANTITATIVE INDICATORS FOR TYPICAL COMPETITIVE SPORT MOVEMENT TECHNIQUES

### FORCE SEQUENCE ANALYSIS FOR TRACK AND FIELD THROWING EVENTS

Force sequence analysis for track and field throwing events focuses on the spatiotemporal coordination characteristics of the momentum transfer chain. Events such as javelin and shot put require athletes to maximize implement velocity through sequential force generation patterns of lower limb extension, trunk rotation, and upper limb whipping. The system quantifies the activation timing and peak torque moments of each segment, calculating time delays between adjacent joints to evaluate power chain transmission efficiency. Force sequence indicators include ground push moment, maximum hip joint rotation velocity moment, peak shoulder joint torque moment, and wrist release instant, with ideal technical patterns showing proximal-to-distal sequential activation. The analysis extracts joint power flow direction, quantifying the proportion of energy transfer from lower to upper limbs and identifying energy dissipation segments. Experimental comparison of force sequences between elite and ordinary athletes shows the former has a 0.08-second hip-shoulder activation time difference while the latter reaches 0.15 seconds; delay causes momentum transfer interruption, reducing release velocity. The system generates personalized technical reports, marking force rhythm deviations and providing improvement suggestions. After training intervention, athletes' throwing performance improved by an average of 4.3%, validating the guiding role of quantitative analysis for technical optimization [11].

### BIOMECHANICAL FEATURES OF BALL SPORT STRIKING MOVEMENTS

Biomechanical features of ball sport striking movements encompass multi-dimensional parameters in preparation, acceleration, and follow-through phases. Taking tennis forehand as an example, the system analyzes key indicators such as shoulder joint internal rotation velocity, elbow joint extension torque, and wrist joint flexion angle. The preparation phase assesses body center of mass transfer and lower limb ground reaction force; the acceleration phase quantifies explosive power output of upper limb joints; and the follow-through phase analyzes deceleration mechanisms and energy dissipation patterns. Biomechanical feature parameters establish regression models with ball velocity, spin, and landing accuracy, identifying dominant factors affecting stroke quality. Research shows shoulder joint internal rotation peak velocity is highly positively correlated with ball velocity ($r = 0.78$), while elbow joint extension torque determines stroke stability. The system compares feature distributions of athletes at different levels, finding that professional players have 18° greater trunk rotation angle at ball contact moment than amateur players, resulting in greater energy storage space. Technical diagnosis identifies non-optimal motion patterns such as premature upper limb activation leading to lower limb force waste. Targeted training programs strengthen power chain coordination, with athletes showing significant technical rating improvements [12].

### SPORTS INJURY RISK FACTOR IDENTIFICATION AND WARNING

Sports injury risk factor identification is based on long-term monitoring and statistical analysis of biomechanical loads. The system tracks cumulative load indicators during training, including joint impact force peaks, muscle fatigue index, and movement symmetry deviations. The risk assessment model integrates acute-to-chronic load ratio, triggering warnings when short-term load increases dramatically exceed long-term adaptation levels. Joint load thresholds are customized based on individual athlete characteristics and injury history, using machine learning algorithms to learn precursor patterns of injury occurrence from historical data. The system identifies high-risk movement patterns such as knee valgus angle exceeding safe ranges, significantly increasing anterior cruciate ligament stress. Such movements are more common in female athletes, resulting in higher injury rates. The warning mechanism provides immediate feedback to coaches and athletes upon detecting risk factors, recommending training intensity adjustments or technical movement corrections. Prospective research tracking sports teams using this system shows injury incidence reduced by 37% compared to control groups, proving the preventive efficacy of intelligent warning systems and providing safety assurance for sustainable development of competitive sports [13].

### PERSONALIZED TECHNICAL FEEDBACK SYSTEM CONSTRUCTION AND PRACTICAL VALIDATION

The personalized technical feedback system integrates multimodal analysis results to generate visualization reports and real-time guidance information. The system interface displays athletes' 3D pose replay, joint angle curves, mechanical parameter distribution, and comparison with standard templates. Coaches mark technical deficiencies through an interactive interface and formulate training plans. Feedback content includes quantitative indicator deviations, key frame pose comparisons, and improvement suggestion text. Information presentation follows pedagogical principles, highlighting core issues while avoiding information overload. The real-time feedback module deploys edge computing devices at training sites [14], generating analysis results within 3 seconds after movement completion. Athletes wear augmented reality glasses to receive immediate prompts, achieving closed-loop learning. The system underwent 6-month practical validation at the National Training Center from March to August 2024, involving 32 athletes (18 track and field, 14 tennis players) from the national team. Pre-post testing included technical performance metrics, biomechanical assessment scores, and coach satisfaction surveys (n=8 coaches, 5-point Likert scale). This study was approved by the Sports Ethics Committee, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form. Participating athletes showed 19% improvement in technical stability indicators and 82% improvement rate in specialized performance. Coach satisfaction surveys showed universal recognition of system effectiveness. User feedback indicated the system reduced time costs for technical analysis, allowing coaches to focus on training strategy development, promoting digital transformation of training processes [15].

## VI. CONCLUSION

The integrated application of multimodal deep learning and computer vision technology, combined with the real-time biomechanical and physiological data captured by smart textile-integrated sensors, marks the leap of sports biomechanics research from laboratory-controlled environments to authentic training scenarios. By constructing an end-to-end intelligent analysis framework, real-time quantitative assessment of athlete technical movements and multi-dimensional biomechanical parameter analysis are achieved. Compared to traditional marker-based systems, data collection efficiency and spatial adaptability are significantly improved. However, current models still have room for improvement in robustness under complex occlusion scenarios and generalization capability across sports. The impact of multimodal data temporal synchronization accuracy on biomechanical parameter estimation accuracy requires in-depth exploration. Future research should focus on few-shot learning methods to reduce annotation dependence, combining biomechanical prior knowledge to constrain network training processes, and exploring federated learning architectures for multi-center data collaborative modeling. The resulting maturity of this technical system, underpinned by the integration of smart textiles directly woven with sensing capabilities, will furnish universal, comfortable tools for scientific training in competitive sports, personalized youth athletic skill development, and widespread public fitness guidance, fundamentally promoting the evolution of sports science research paradigms toward a combination of data-driven and knowledge-guided approaches.

## AUTHOR CONTRIBUTIONS

Conceptualization – Xuejun Tian and Fuliang Geng; methodology – Xuejun Tian and Fuliang Geng; investigation – Xuejun Tian and Fuliang Geng; writing-original draft preparation – Xuejun Tian and Fuliang Geng. All authors have read and agreed to the published version of the manuscript.

## CONFLICTS OF INTEREST

The authors declare no conflict of interest.

## HUMAN RESEARCH SUBJECTS

This study was approved by the Sports Ethics Committee, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

## FUNDING

This research received no external funding.

## ACKNOWLEDGEMENTS

Not applicable.

## REFERENCES

[1] K. D. McNish, Z. Gao, R. Trama, et al., “Utilised traction and foot biomechanics of male and female rugby athletes during a cutting movement,” Footwear Science, vol. 17, no. sup1, pp. S266-S267, 2025, doi: 10.1080/19424280.2025.2499508.

[2] C. Yang, L. Wei, X. Huang, et al., “Comparison of lower limb kinematic and kinetic estimation during athlete jumping between markerless and marker-based motion capture systems,” Scientific Reports, vol. 15, no. 1, pp. 18552-18552, 2025, doi: 10.1038/S41598-025-02739-9.

[3] M. E. Nijmeijer, D. F. Brals, M. Kempe, et al., “How are athletes trained to move? A systematic review exploring the effects of implicit and explicit learning on biomechanics of sport-specific tasks,” Journal of Biomechanics, vol. 184, pp. 112671, 2025, doi: 10.1016/J.JBIOMECH.2025.112671.

[4] S. P. Glazier, “Markerless motion capture in sport: panacea or Pandora’s box? Sports biomechanics,” Sports Biomechanics, pp. 1-8, 2025, doi: 10.1080/14763141.2025.2458485.

[5] W. Dhahbi, “Editorial: Advancing biomechanics: enhancing sports performance, mitigating injury risks, and optimizing athlete rehabilitation,” Frontiers in Sports and Active Living, vol. 7, pp. 1556024-1556024, 2025, doi: 10.3389/FSPOR.2025.1556024.

[6] Z. Zhu, “Design and implementation of an intelligent sports management system (ISMS) using wireless sensor networks,” PeerJ Computer Science, vol. 11, pp. e2637, 2025, doi: 10.7717/PEERJ-CS.2637.

[7] Y. Wang, “A study on the optimization of athletes’ fitness assessment and training program based on sports biomechanics,” Applied Mathematics and Nonlinear Sciences, vol. 10, no. 1, pp. 20251035-20251035, 2025, doi: 10.2478/AMNS-2025-1035.

[8] X. Tang, B. Long, and L. Zhou, “Real-time monitoring and analysis of track and field athletes based on edge computing and deep reinforcement learning algorithm,” Alexandria Engineering Journal, vol. 114, pp. 136-146, 2025, doi: 10.1016/J.AEJ.2024.11.024.

[9] H. Juan, “Analysis of changes in lower limb joints of athletes during the movement of wild horses’ mane parting based on biomechanics,” Journal of Vibroengineering, vol. 26, no. 8, pp. 1783-1790, 2024, doi: 10.21595/JVE.2024.24240.

[10] C. Liu and Y. Xie, “Innovative Application of Computer Vision and Motion Tracking Technology in Sports Training,” EAI Endorsed Transactions on Pervasive Health and Technology, vol. 10, no. 1, 2024, doi: 10.4108/EETPHT.10.5763.

[11] P. Xing, L. Zhang, T. Wang, et al., “A deep learning algorithm that aids visualization of femoral neck fractures and improves physician training,” Injury, vol. 55, no. 12, pp. 111997, 2024, doi: 10.1016/J.INJURY.2024.111997.

[12] W. Peng, “Research on Sports Training Action Recognition Based on Deep Learning,” Scientific Programming, vol. 2021, 2021, doi: 10.1155/2021/3396878.

[13] H. S. Lim, S. Hong, J. H. Park, et al., “Development of Coastal Disaster Simulation System using Marine Digital Twin Technology,” Journal of Coastal Research, vol. 113, no. Sup1, pp. 6, 2024, doi: 10.2112/JCR-SI113-192.1.

[14] S. Ryu, K. Ko, and K. W. J. Hong, “Performance Analysis of Applying Deep Learning for Virtual Background of WebRTC-Based Video Conferencing System,” IEICE Proceeding Series, vol. 67, pp. 53-56, 2021, doi: 10.34385/PROC.67.PS1-14.

[15] Q. Xu, H. Yan, J. Yang, et al., “Biomechanical Differences in Bilateral Lower Limb Movement During the Back Kick Technique of Outstanding Taekwondo Athletes,” Life, vol. 15, no. 12, pp. 1822-1822, 2025, doi: 10.3390/LIFE15121822.