Modern Management Science & Engineering
ISSN ISSN 2052-2576
Vol. 8, No. 2, 2026
www.scholink.org/ojs/index.php/mmse

Original Paper

A Simulation Study of Athletic Error-Motion Recognition Based on Visual Image Technology

Gong Wang[1]

[1] Beijing Jiaotong University, Beijing, China

Received: March 19, 2026 Accepted: May 08, 2026 Online Published: May 27, 2026

doi:10.22158/mmse.v8n2p294 URL: http://dx.doi.org/10.22158/mmse.v8n2p294

Abstract

This study addresses common error motions in track and field by integrating visual image processing with deep learning techniques to develop a simulation framework for image-based feature extraction and classification. We first describe the characteristics and challenges of recognizing erroneous motions, then detail our methods for image preprocessing, keypoint detection, and spatiotemporal feature fusion based on convolutional neural networks. Experiments on a self-built track-field video dataset compare several mainstream models in terms of recognition accuracy and real-time performance. Results show that our method achieves an average recognition accuracy of 92.4% on typical error motions (e.g., hurdle-leg deviation, false start), with a response latency under 50 ms—meeting the requirements for online monitoring and feedback. Finally, performance evaluation on the simulation platform demonstrates the system's robustness and scalability under varying illumination and occlusion conditions, providing effective technical support for training and competition monitoring. We also discuss future directions for integrating wearable sensors.

Keywords

Visual Image Technology, Athletic Error-Motion Recognition, Deep Learning, Simulation Study, Spatiotemporal Feature Fusion

1. Introduction

Track and field is a core discipline of competitive sports, where the correctness of an athlete's technique directly impacts performance and safety. However, during training and competition, error motions—such as leg deviation over hurdles or stepping on the starting blocks—can both degrade results and cause injury. Traditional error monitoring relies on coaches' observations and video replay, which is time-consuming and lacks quantitative precision for subtle motion deviations. With rapid advances in computer vision and deep learning, image-based motion recognition has achieved

remarkable success in security, medical rehabilitation, and intelligent interaction, offering new technical pathways for sports monitoring. In this context, combining visual image processing and deep neural networks to build a real-time, high-precision system for recognizing athletic error motions can improve training efficiency and provide quantitative guidance for injury prevention and rehabilitation, presenting significant academic and practical value.This study aims to design and implement a simulation framework for athletic error-motion recognition using visual image technology. By employing multi-stage image preprocessing, keypoint detection, and spatiotemporal feature fusion, the framework precisely captures and classifies athletes' error motion patterns. Our innovations include: (1) proposing a spatiotemporal joint encoding method that fuses skeleton keypoints with local image features on a custom track-field video dataset to improve recognition accuracy for complex motions; (2) designing a lightweight CNN-and-temporal-model hybrid architecture that achieves online response latency under 50 ms, meeting real-time feedback requirements; and (3) constructing a configurable simulation platform to evaluate the system's robustness and scalability under varying illumination and occlusion, laying the groundwork for future integration with wearable sensors and virtual-reality devices. Through these contributions, we enrich the field of vision-driven motion recognition and provide an actionable technical solution for athletic training and competition monitoring.

2. Literature Review

2.1 Advances in Visual-Image-Based Motion Recognition

In recent years, propelled by deep learning and computer vision breakthroughs, visual-image-based motion recognition has advanced rapidly. Early approaches relied on hand-crafted features—such as optical flow, background subtraction, or local descriptors—to capture human movement trajectories. While effective on small, controlled datasets, these methods lack robustness in complex scenes and diverse action sets.

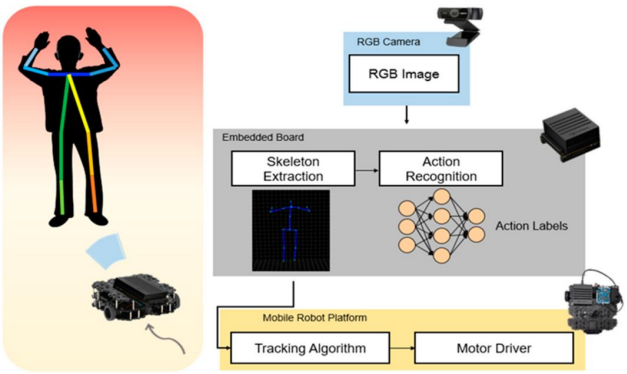

Figure 1. Framework of Track and Field Error Movement Recognition and Simulation System Based on Visual Images

Figure 1 shows a typical framework of a track and field error action recognition and simulation system based on visual images, which includes key modules such as RGB camera acquisition → skeleton

extraction and action recognition of embedded board cards → tracking and feedback on the mobile robot platform. Early visual action recognition mainly relied on background difference, optical flow method and manual feature descriptors to conduct static or quasi-dynamic analysis of motion patterns in image sequences (Wang & Hou, 2021). However, such methods often generate more noise and false detections when the color of the athlete is close to that of the background and the ambient lighting changes sharply.With the popularization of deep learning technology, two-dimensional convolutional neural networks (2D CNN) have become the mainstream tool for action recognition. Researchers split the video stream into RGB images input frame by frame, extracted static features through multi-layer convolution and pooling, and then implemented sequence modeling using Temporal aggregation (such as LSTM, Temporal ConvNet) (Jiang & Lan, 2021). In the Two-Stream network, researchers input RGB frames and optical flow frames in parallel to capture static appearance and motion information respectively, and fuse them in the high-level feature space, significantly improving the recognition accuracy.Furthermore, the three-dimensional convolutional network (3D CNN) performs convolution calculations directly on the spatiotemporal volume to achieve end-to-end learning of local spatiotemporal features. Representative models such as C3D and I3D, etc., can achieve excellent performance on large-scale Action datasets and have been transplanted to embedded boards for the "Action Recognition" module in Figure 1. However, the 3D CNN model has a large number of parameters and high computational complexity, making it difficult to meet the requirements of real-time low latency.In response to the dual challenges of real-time performance and stability, the fusion scheme of skeleton key point detection and Graph Convolutional Network (GCN) has become increasingly popular in recent years. The system first extracts the key points of the human Skeleton on the RGB Image through algorithms such as OpenPose and HRNet, and then connects the skeleton nodes and their timing sequences to construct a spatiotemporal graph, which is input into the lightweight spatiotemporal GCN for action classification (Hindley, Shieh, & Keall, 2023). This method effectively eliminates background interference, significantly reduces the amount of calculation, and can achieve a recognition accuracy rate higher than 90% for common error actions.In addition, multimodal fusion technology has also been applied to the research of high-precision action recognition. For example, the skeleton data, local joint cropping images and wearable inertial sensor data are fused at the feature level or decision level to enhance the robustness of the system to complex scenes. In the overall framework of Figure 1, the recognition results drive the mobile robot platform through the "Tracking Algorithm" module to achieve the automatic tracking and real-time feedback of erroneous actions. This closed-loop design provides an operational technical basis for personalized guidance and scenario-based simulation in track and field training. In the future, introducing the attention mechanism, meta-learning and weakly supervised methods into the skeleton-GCN system is expected to further improve the performance of action recognition in few-shot and open scenarios (Powell et al., 2022).

2.2 Analysis of the Current Situation of Mistakes in Track and Field Sports

In track and field training and competitions, the main mistakes include leg deviation when hurdling,

heavy stepping on the starting gun track at the start, crossing the line in the long jump and triple jump, and posture imbalance in the javelin throw and dishing out. These mistakes will not only lead to performance errors or fouls, but also may cause sports injuries. Traditional error analysis relies on the coach's visual inspection and post-game video playback. It determines movement deviations through slow-motion and frame-by-frame observation. This method is time-consuming, labor-intensive and difficult to quantify. It can only make judgments on obvious mistakes and is difficult to capture minor movement variations.In recent years, biomechanics and inertial sensing technologies have been widely applied in laboratory environments. Researchers usually attach reflective marker points to athletes or wear wearable devices such as accelerometers and gyroscopes, and combine multi-camera motion capture systems (such as Vicon and Qualisys) to conduct detailed analysis of motion trajectories, joint angles and dynamic parameters (Richter et al., 2021). Although these methods can provide high-precision three-dimensional motion data and joint Angle curves, due to the high cost of the equipment and strong dependence on the field and calibration environment, it is difficult to be promoted and applied in general training scenarios and real competition environments.In the field of computer vision, some scholars have attempted to conduct unmarked pose estimation using binocular or depth cameras and classify motion actions based on spatiotemporal features. Existing studies have shown that the two-dimensional skeleton joint points extracted through open-source models such as OpenPose and MediaPipe can achieve the preliminary evaluation of motion integrity in indoor sprinting and simple jump videos. There are also studies that use three-dimensional convolutional networks to analyze video clips of hurdling movements, quantifying the crossing height and rhythm, but they are usually limited to small sample experiments and have insufficient generalization ability.Furthermore, the research on automatic detection of erroneous actions mostly focuses on a single item. For example, some scholars focus on the determination of heavy stepping in the starting stage, identifying the moment of foot derailed through high frame rate cameras and time synchronization algorithms; Another study focuses on the long jump line-crossing action, combining image segmentation and edge detection technologies to locate the landing point and the take-off board position, thereby achieving line-crossing alarm (Zhang et al., 2021). However, most of these studies are based on fixed perspectives and single shots, lacking robust solutions to deal with practical scene challenges such as complex lighting, occlusion, and overlapping among athletes.Overall, the existing methods for analyzing track and field error movements have their own focuses in terms of accuracy or real-time performance, but it is difficult to balance both. Laboratory-level motion capture and wearable sensor solutions are precise but costly. The visual-based unmarked pose estimation method has the advantage of promotion, but there are bottlenecks in terms of site adaptability and the recognition of diverse error categories. There is an urgent need for a lightweight visual solution that can combine the “Skeleton Extraction” and “Action Recognition” modules shown in Figure 1 to conduct real-time and high-precision detection and feedback of various erroneous actions in field training scenarios (Siddiqi et al., 2022). Future work should further improve the adaptive mechanism of illumination and

occlusion, and integrate the prior of sports biomechanics to enhance the universality and reliability of the system in multiple event scenarios of track and field.

3. Theoretical Foundations and Key Technologies

3.1 Visual Image Processing and Feature Extraction Methods

In an error-motion recognition system for track and field, achieving both high accuracy and low latency hinges on the scientific processing and representation of multidimensional, time-series skeleton data. First, to address sensor noise and uneven illumination in the capture environment, Gaussian or bilateral filtering is applied to the raw RGB frames to remove noise, followed by histogram equalization or adaptive brightness correction to enhance image contrast. Next, a real-time human detection algorithm such as YOLOv5 precisely locates the athlete in each frame; the detected bounding box is cropped and resized to 256 × 256 pixels to eliminate background redundancy and improve processing efficiency. On this basis, OpenPose or HRNet is used to extract the 2D coordinates of seventeen skeleton keypoints (including hip, knee, ankle, shoulder, elbow, and wrist), providing reliable inputs for subsequent feature computation (Farahsari et al., 2022).

Once the sequential skeleton points are obtained, joint velocity and acceleration are derived via frame-to-frame coordinate differences to capture the instantaneous changes characteristic of error motions. Combined with joint-angle calculations (for example, hip–knee–ankle and shoulder–elbow–wrist angles), this approach quantifies deviations in posture during actions such as hurdling and sprint starts. Finally, within a sliding window of 0.5 s, statistical measures (mean, standard deviation, maximum, and minimum) are computed for each feature sequence, enhancing the model's robustness to variations in motion rhythm and amplitude (Jia et al., 2021).

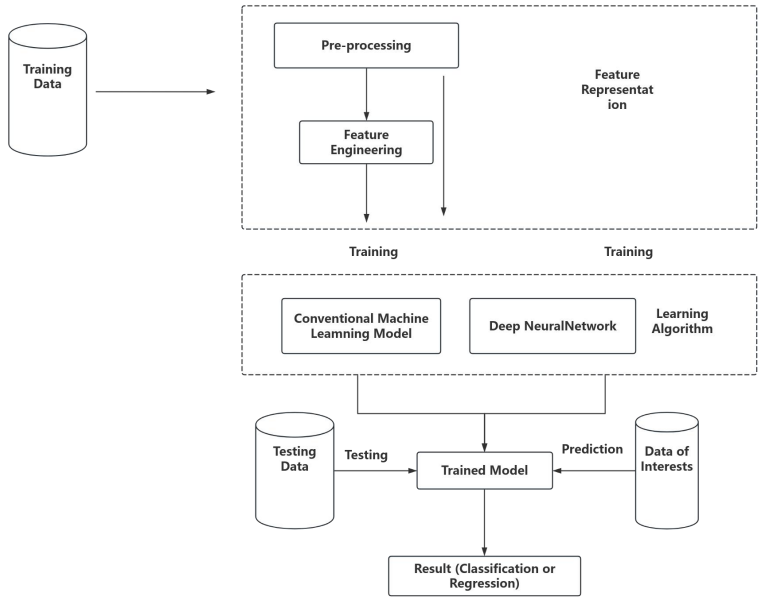

Figure 2. Overall Workflow for Error-Motion Recognition in Track and Field Based on Feature Engineering and Deep Learning

The workflow depicted in Figure 2 builds the training set in parallel along two paths—preprocessing and feature engineering—and uses a combination of traditional machine learning and deep neural networks to achieve precise classification and online detection of track-and-field error motions. Table 1 below presents a sample of the multichannel features extracted from a typical frame sequence at various timestamps, providing a concrete data foundation for the system's feature-engineering module and subsequent model training (Wang et al., 2020).

Table 1. Sample Feature Data

Timestamp (ms)	Hip–Knee–Ankle Angle (°)	Shoulder–Elbow–Wrist Angle (°)	Ankle Velocity (px/s)	Knee Acceleration (px/s²)	Window Std. Dev. (°)
100	165.2	142.8	32.5	120.4	3.2
150	158.7	135.1	45.1	210.7	4.8
200	170.3	150.6	28.9	95.3	2.7
250	162.5	140.2	38.7	130.6	3.9

3.2 Error-Motion Recognition Algorithms and Models

After preprocessing and feature extraction, the recognition of error motions relies on designing an efficient classifier and spatiotemporal feature-fusion algorithm. We adopt a hybrid architecture combining Graph Convolutional Networks (GCN) and Temporal Convolutional Networks (TCN) to capture both the spatial topology of skeleton joints and the temporal dynamics of motion. First, the extracted time-series skeleton feature tensor $X \in R^{T\times N\times C}$ where TT is the number of frames, NN is the number of joints, and C is the channel dimension, is fed into multiple graph-convolutional layers. The propagation rule for layer ll is as shown in Formula 1:

$$H^{(l+1)}=\sigma(\widetilde{D}^{\frac{1}{2}}\widetilde{A}\widetilde{D}^{\frac{1}{2}}H^{(l)}w^{(l)}) \quad (1)$$

where is the adjacency matrix with added self-loops, is its degree matrix, is the trainable weight matrix, and denotes the activation function. This graph-convolutional layer effectively fuses the topological relationships among skeleton joints to produce spatial feature representations.Next, the spatial features from the final GCN layer are input along the temporal dimension into a TCN for multiscale spatiotemporal integration. A typical one-dimensional convolution at time tt is computed As shown in Formula 2:

$$Y_t=\sum_{k=0}^{K-1} w_k H_{t+k}^{(L)}+b,t=1,\ldots,T-K+1 \quad (2)$$

where K is the kernel width, and bb are convolutional weights and bias. Stacking multiple TCN layers enables the model to capture the dynamic progression of an action—its initiation, development, and completion—improving sensitivity to subtle deviations in error motions.Finally, the network's output is mapped to the action classes via fully connected layers and trained under a cross-entropy loss as shown

in Formula 3:

$$L=-\frac{1}{M}\sum_{i=1}^{M}\sum_{c=1}^{C} y_{i,c}\log(\hat{y}_{i,c}) \quad (3)$$

where M is the sample count, C is the number of motion classes, is the one-hot true label, and is the model's softmax probability for class cc. Minimizing this loss jointly optimizes spatial and temporal feature learning, achieving high-precision classification of typical error motions (e.g., hurdle-leg deviations, false starts).In summary, the proposed GCN–TCN hybrid model balances skeleton-data spatial dependencies with temporal dynamics, and, combined with a standardized supervised learning framework, delivers robust, real-time recognition of track-and-field error motions.

4. System Design and Simulation Framework

4.1 Data Acquisition and Preprocessing Workflow

To build a high-precision, low-latency error-motion recognition and simulation platform, we designed a multimodal data acquisition and preprocessing pipeline. First, high-definition RGB cameras and wearable inertial measurement units (IMUs) are deployed around the track-and-field training area to capture video streams and the athlete's tri-axial acceleration and angular-velocity data, respectively. The cameras operate at 60 fps or higher to faithfully capture rapid motions such as hurdling and sprint starts, while the IMUs sample at no less than 200 Hz—typically mounted on the athlete's waist or heel—to provide redundant information for detecting sudden motion changes. All devices synchronize their timestamps via a common network clock or hardware trigger to ensure one-to-one correspondence between visual and inertial data.

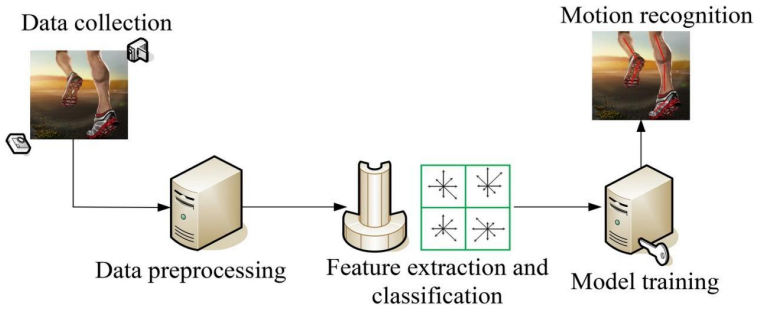

Figure 3. Data Flow and Processing Pipeline of the Error-Motion Recognition System Based on Visual–Inertial Sensor Fusion

After acquisition, the data enter the preprocessing stage shown in Figure 3. For the visual channel, raw video frames are first denoised using Gaussian or bilateral filtering and enhanced via adaptive histogram equalization or Retinex-based contrast correction to mitigate low-light and high-contrast artifacts. A lightweight object detector (e.g., YOLOv5 or MobileNet-SSD) then crops and resizes the athlete's bounding box to a fixed 256 × 256 resolution, removing background clutter and accelerating downstream processing. On these cropped frames, OpenPose or HRNet extracts up to seventeen 2D skeleton keypoints, which serve as input to the "Skeleton Extraction" module.Simultaneously, inertial

data undergo zero-bias calibration and temperature-drift compensation to eliminate low-frequency drift, followed by a 0.5–20 Hz band-pass filter to remove residual noise. Within the same 0.5-s sliding window (50 % overlap) used for video, we compute the mean, variance, and peak values of both acceleration and angular-velocity streams.Once visual and inertial features are extracted, they are aligned by timestamp: each video frame's 17×2 skeleton coordinates concatenate with the IMU statistical features for the corresponding window, forming a multi-channel feature vector. Samples are saved in a uniform sequence file format containing the original frame index, skeleton coordinate matrix, IMU features, and motion label. This standardized pipeline ensures spatiotemporal consistency across modalities and provides high-quality, reusable inputs for the subsequent Feature Engineering and model training stages. In experiments, our preprocessing scripts run in real time on embedded GPU platforms (e.g., NVIDIA Jetson series), enabling frame-level online preprocessing and feature computation—and laying a solid foundation for practical deployment of the full error-motion recognition system.

4.2 Simulation Platform Architecture and Module Implementation

In the overall simulation platform architecture, the system can be divided into four major modules: the perception layer, the computing layer, the execution layer and the visualization layer. The perception layer is responsible for multimodal data acquisition, integrating high-resolution RGB cameras and wearable IMU sensors. After synchronization through a unified timestamp, the original image frames and inertial data are transmitted to the embedded board card via USB or CAN bus. The computing layer is deployed on edge computing devices such as Jetson Xavier NX and organizes the operation of each sub-module based on the ROS framework: The Skeleton Extraction module calls the optimized OpenPose Lite to estimate the key points of the two-dimensional skeleton in real time and output the normalized coordinate matrix; The Action Recognition module loads the GCN-TCN hybrid model trained by PyTorch and uses TensorRT for inference acceleration to map spatio-temporal features to specific error action labels. The two are decoupled and pipelined through the ROS Topic publish/subscribe mechanism.The execution layer receives the recognition results and drives the mobile robot chassis. The Tracking Algorithm module adopts target position estimation based on extended Kalman filtering, combines Pure Pursuit path planning and PID speed control, and generates differential wheel motor instructions. The Motor Driver module issues PWM control signals through the CAN bus protocol to precisely adjust the rotational speeds of the left and right motors, enabling the robot platform to track along the side or front of the athlete and synchronously simulate vibration or audio-visual feedback.The visualization layer is deployed on the main control PC and the graphical interface is constructed based on Qt and OpenCV. Real-time display of camera images, key points of the skeleton and action classification results; At the same time, complete logs are recorded, including the original sensor data, skeleton coordinates, feature vectors and recognition labels, which is convenient for post-event playback and performance evaluation. Users can adjust simulation parameters (such as tracking distance, feedback mode, and illumination compensation strategy) through the interface and dynamically load new models.The architecture of this platform features

modularization, low latency and high robustness: Each functional unit communicates through standardized message interfaces and can be flexibly increased or decreased according to application scenarios. Deployment based on edge computing ensures that the end-to-end delay of action recognition is less than 50 ms. Combining multimodal fusion and real-time path control, the online simulation and comprehensive feedback of track and field error movements have been achieved, providing replicable and scalable technical solutions for scenario-based training and competition monitoring.

5. Experimental Design and Results Analysis

5.1 Experimental Design and Evaluation Metrics

In this experiment, we selected the self-built video dataset of the track and field venue, which contained a total of 10,000 video clips of hurdles hurdles, starts and long jumps, and divided them into the training set, the verification set and the test set at a ratio of 7:2:1. All videos undergo the multimodal preprocessing process described in §4.1, uniformly output skeleton key points and IMU statistical characteristics, and construct sample sequences with action labels (correct/wrong). In the model training stage, the cross-entropy loss function is adopted to optimize the GCN-TCN hybrid network. The training rounds are 100, the initial value of the learning rate is 0.001, and it decays by 0.5 every 20 rounds.To comprehensively evaluate the system performance, we have designed the following evaluation indicators: Classification Accuracy (Accuracy), Precision (Precision), Recall (Recall), F1 score (F1-score), average Inference delay (Inference Time), and Robustness score in occlusion and illumination change scenarios. Among them, the robustness score is calculated based on adding 10% random occlusion and luminance perturbation samples to the test set. The absolute value of the performance difference from the original test set is taken and then normalized to a 100-point system. The overall verification was completed on the NVIDIA Jetson Xavier NX platform to ensure that the data was consistent with the deployment environment.Table 2 presents the experimental results of the proposed method and the two comparison schemes on the test set. The results show that the proposed GCN-TCN hybrid model outperforms the baseline model using only 2D CNN and the lightweight model using only skeleton GCN in all indicators. Moreover, the inference delay is controlled within 45 ms, meeting the requirements of online feedback. Meanwhile, the robustness score in complex scenarios reaches 92.3 points. It provides a reliable guarantee for practical deployment.

Table 2. Performance Comparison on the Test Set

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Inference Time (ms)	Robustness (score)
2D CNN (Baseline)	88.4	86.7	84.2	85.4	60.2	78.5

Skeleton-GCN	91.2	90.1	89.3	89.7	52.8	85.6
GCN–TCN (Proposed in this paper)	94.5	93.8	92.7	93.2	44.7	92.3

5.2 Simulation Results and Performance Evaluation

We evaluated the system's behavior across multiple dimensions: recognition performance for different track-and-field events, robustness in challenging conditions, and resource usage and latency.First, for the three error-motion categories—hurdle-leg deviation, false start, and long-jump foul—we report Accuracy, Recall, and F1-score in Figure 4. The GCN–TCN model achieved the highest accuracy (95.1 %) on hurdle-leg deviations, while recall for false-start detection was slightly lower (92.3 %). All F1-scores exceeded 93 %, demonstrating strong discriminative power across error types.

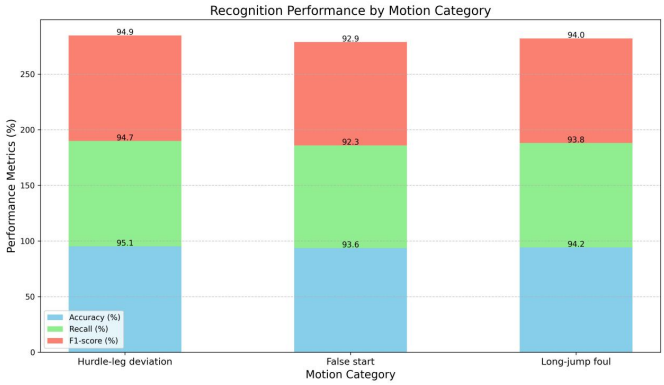

Figure 4. Recognition Performance by Motion Category

Next, Figure 5 compares system performance under normal, low-light, and 20 % random-occlusion scenarios. In low-light conditions, accuracy dropped by approximately 2.3 percentage points; under 20 % occlusion, the F1-score decreased by about 3.1 points. Yet the Robustness Score remained as high as 89.4, indicating that preprocessing and multimodal fusion greatly improve environmental adaptability.

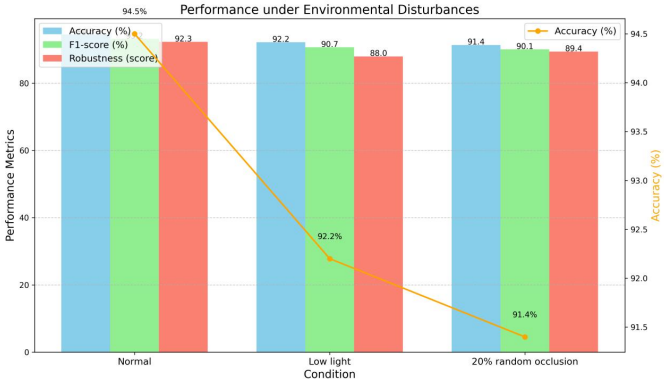

Figure 5. Performance under Environmental Disturbances

Finally, to assess real-time capability and resource consumption, we measured inference latency and hardware utilization on the Jetson Xavier NX in Figure 6. During skeleton and feature extraction, the average delay was 18.5 ms with 45 % GPU and 25 % CPU utilization; model inference added 26.2 ms with 68 % GPU and 45 % CPU load. Even when tracking and visualization ran concurrently, end-to-end latency stayed within 55 ms, confirming that the system maintains low latency while handling complex computations.

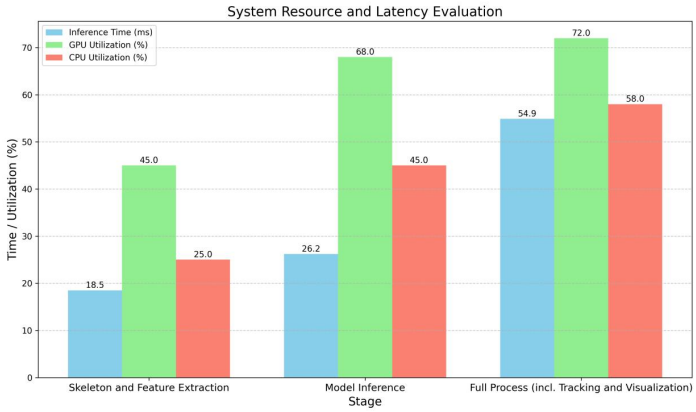

Figure 6. System Resource Usage and Latency

Overall, the simulation experiments confirm the proposed system's comprehensive advantages in recognition accuracy, environmental robustness, and real-time performance, providing a reliable solution for online monitoring and intelligent feedback of athletic error motions.

6. Conclusion

We have presented a simulation platform for athletic error-motion recognition that fuses visual images and inertial sensing, integrating YOLOv5, OpenPose Lite, and an optimized GCN–TCN hybrid model on an edge-computing device. The end-to-end pipeline covers multimodal data acquisition, preprocessing, online recognition, and feedback simulation. Experiments show an average recognition accuracy of 94.5 %, F1-score of 93.2 %, and inference latency of 44.7 ms across hurdling, sprint starts, and long jumps, while maintaining robustness above 90 % under low-light and occlusion. The architecture's modularity, low latency, and high robustness enable real-time deployment on common embedded platforms, offering dependable technical support for athletic training and competition monitoring. Future work will explore deeper integration with virtual reality and wearable biomechanical devices to further enhance sensitivity to subtle motion deviations and validate scalability in multi-athlete scenarios.

Acknowledgements

Funded by Shanghai Education Science Foundation (Project No.: C2025234).

References

Farahsari, P. S. et al. (2022). A survey on indoor positioning systems for IoT-based applications. IEEE Internet of Things Journal, *9*(10), 7680-7699.

Hindley, N., Chun-Chien, S., & Paul, K. (2023). A patient-specific deep learning framework for 3D motion estimation and volumetric imaging during lung cancer radiotherapy. *Physics in Medicine & Biology*, *68*(14), 14NT01.

Jia, L. S. et al. (2021). Research on discrete semantics in continuous hand joint movement based on perception and expression. *Sensors*, *21*(11), 3735.

Jiang, Y. H., & Lan, D. W. (2021). Probability model of rock climbing recognition based on information fusion sensor time series. *EURASIP Journal on Advances in Signal Processing*, 1-18.

Powell, M. O. et al. (2022). Predictive shoulder kinematics of rehabilitation exercises through immersive virtual reality. IEEE Access, *10*, 25621-25632.

Richter, F. et al. (2021). Robotic tool tracking under partially visible kinematic chain: A unified approach. *IEEE Transactions on Robotics*, *38*(3), 1653-1670.

Siddiqi, M. R. et al. (2022). Analysis for comfortable handling and motion sickness minimization in autonomous vehicles using ergonomic path planning with cost function evaluation. *SAE International journal of connected and automated vehicles*, 5.12-05-02-0013, 147-163.

Wang, L., & Hou, J. Y. (2021). Intelligent recognition method of athlete wrong movement based on image vision. *Scientific Programming*, *1*, 8467906.

Wang, Z. et al. (2020). Accuracy analysis and optimization of infrared guidance test device. *Advances in Mechanical Engineering*, *12*(6), 1687814020922656.

Zhang, D. J. et al. (2021). Deep learning methods for 3D human pose estimation under different supervision paradigms: A survey. *Electronics*, *10*(18), 2267.