

Martial Arts Routine Action Recognition and Scoring Using YOLOv8 and OpenPose

Dan Zhao¹, Yuanqi Ma¹

¹School of Physical Education, Hechi University, Hechi 546300, Guangxi, China

Corresponding author: Yuanqi Ma (e-mail: hcxyuanqi@163.com)

ABSTRACT Martial arts routines are highly dynamic, multi-pose, and fast-paced, posing significant challenges to automated recognition and scoring. Such complex spatiotemporal characteristics are also representative of intelligent sensing tasks in advanced electromagnetic-aware environments, where robust human motion perception is essential for integrated visual and wireless monitoring systems. This paper proposes a joint framework combining YOLOv8 and time-optimized OpenPose to mitigate pose estimation jitter and detection inaccuracies caused by rapid motion and occlusion in human motion analysis. YOLOv8 first performs high-precision human target detection using bounding boxes to define regions of interest, after which the cropped images are processed by OpenPose to extract 17 keypoints, with IoU matching ensuring cross-frame identity association for temporal consistency. The keypoint sequence is modeled as a multi-channel time series and refined through a bidirectional LSTM network to predict smooth pose trajectories. The optimized keypoints are further used to calculate joint angles and movement velocities, which are integrated with dynamic thresholds for motion segmentation. DTW-based alignment and similarity matching with a standard motion library are subsequently employed to evaluate posture accuracy, motion amplitude, and rhythmic consistency, producing a comprehensive scoring result. Experimental results demonstrate average mAP@0.5 values of 0.863–0.942 (FPS 115.2–124.3), average PCK values of 88.7%–90.3% under occlusion (average MPJPE 48.2–51.6 pixels), and Pearson correlation coefficients of 0.925–0.941 for scoring consistency across diverse martial arts routines. The proposed framework provides an effective solution for intelligent motion analysis and offers technical reference for multimodal perception and dynamic scene understanding in advanced electromagnetic sensing applications.

INDEX TERMS YOLOv8, openpose, action recognition, intelligent motion analysis, electromagnetic sensing

I. INTRODUCTION

As a complex movement system that integrates combat logic and artistic expression, the martial arts routine contains highly sophisticated body coordination mechanisms in its movement sequences. This required coordination, where every limb and joint must move in precise harmony [1,2]. In competitive judging and skill transmission, the accuracy, strength, and rhythm of movements form the core of evaluation, and relying on subjective interpretation can easily apply bias [3-5]. In recent years, intelligent perception technology has provided an objective and quantitative path for sports quality assessment [6]. By analyzing the evolution of human posture through visual systems, it is not only possible to restore the spatial configuration of the movement but also to capture the dynamic characteristics in the temporal dimension [7,8], promoting the transformation of sports evaluation to data-driven mode. Current vision-based action recognition systems face multiple technical

obstacles when processing highly dynamic martial arts sequences. During rapid leaps, rotations, and strikes, key point detection is prone to coordinate jumps [9], especially when limbs are occluded or clothing is reflective, some joints are frequently lost or misplaced, resulting in distorted posture representation [10,11]. Existing methods rely on simple coordinate proximity matching in cross-frame tracking and lack the ability to model motion continuity, resulting in frequent identity switching or trajectory breakage [12]. Even if a region cropping strategy is used to narrow the detection range, it is still difficult to avoid background interference, misleading the posture estimation module [13]. Furthermore, the original key point sequence contains a large amount of high-frequency noise, and directly using it for action feature extraction will affect the reliability of subsequent analysis [14,15]. Although some studies have attempted to apply sliding window smoothing, this type of static filtering method cannot adapt to changes in movement

rhythm and has difficulty retaining key transient features [16]. In addition, movement segmentation relies on fixed time windows or preset thresholds, ignoring the individual differences in speed and amplitude when different athletes perform the same movement, resulting in deviations in segment boundary determination [17,18]. The scoring process often uses Euclidean distance comparison, ignoring the nonlinear scaling characteristics on the time axis and unable to accurately measure the similarity of movement morphology [19,20]. These problems jointly restrict the system's generalization ability and discrimination accuracy in real scenarios, and there is an urgent need to build an end-to-end analysis framework with strong robustness and adaptability.

II. YOLOV8 AND OPENPOSE FUSION FRAMEWORK CONSTRUCTION

Figure 1 illustrates the YOLOv8 and OpenPose fusion framework. After frame extraction from the raw video, YOLOv8 achieves high-precision human body detection and ensures cross-frame identity consistency through IoU matching. The detection bounding box guides OpenPose to extract 17 keypoints within the cropped region, combining interpolation and coordinate normalization to enhance pose integrity. The keypoint sequence is fed into a bidirectional LSTM network for temporal modeling, outputting smooth, motion-compliant pose trajectories. Based on this, kinematic features such as joint velocity, acceleration, and angular rate of change are extracted. Nonlinear temporal alignment with a standard action library is performed using the DTW algorithm to calculate morphological similarity. Rhythm consistency and a keypoint deviation penalty mechanism are also incorporated, and a weighted fusion process is used to generate a final quantitative score. The system integrates object detection, pose estimation, timing optimization, and intelligent scoring, creating a closed-loop analysis process from raw video to action quality assessment. Designed specifically for highly dynamic martial arts movements, it offers robustness and technical interpretability.

A. MULTI-FRAME HUMAN TARGET DETECTION AND LOCALIZATION BASED ON YOLOV8

1) Frame-by-Frame Human Detection and Multi-Scale Feature Extraction Based on YOLOv8

During the initial processing of video sequences, the system preprocesses the raw input frames, uniformly scaling them to a 640×640 resolution to accommodate the YOLOv8 network input requirements while maintaining the aspect ratio and padding the edges to avoid distortion. The system utilizes YOLOv8's anchor-free detection mechanism, eliminating the traditional reliance on pre-set anchor boxes and directly predicting the offset and size of bounding boxes on feature maps. This improves its adaptability to the diverse poses and scales of human subjects found in martial arts movements. Its backbone network, based on the CSPDarknet architecture, effectively mitigates the vanishing gradient

problem through cross-stage partial connections, enhancing the expressiveness of deep features. During forward propagation, the network concurrently extracts semantic and localization information at three feature layers of different scales (S3/8, S4/16, and S5/32). Low-level features preserve spatial details to accurately locate limb edges, while high-level features capture contextual semantics to distinguish moving subjects in complex backgrounds. A Path Aggregation Network (PANet) architecture enables bidirectional top-down and bottom-up feature fusion, enhancing response consistency for multi-scale objects. Ultimately, the model outputs the coordinate parameters (x, y, w, h) for each detection box, where x and y represent the normalized position of the box's center in the image coordinate system, and w and h represent its width and height. This is accompanied by the category confidence c and the detection confidence s, forming a five-tuple (x, y, w, h, c·s). This confidence score is used to subsequently filter high-reliability detection results, retaining only bounding boxes with a confidence score above a threshold $\tau_{\text{det}} = 0.7$ as valid human regions, thereby suppressing interference from false background detections on the pose estimation module.

2) Temporal Correlation of Cross-Frame Detection Boxes and Stable Region Generation

To ensure the identity consistency of human targets in consecutive video frames and avoid detection frame jumps or losses caused by rapid motion or short-term occlusion, a cross-frame matching strategy based on intersection-over-union (IOU) is applied. For each bounding box B_t^i detected in the current frame t , its spatial overlap with all detection boxes B_{t-1}^j in the previous frame $t-1$ is calculated, which is defined as:

$$\text{IOU}(B_t^i, B_{t-1}^j) = \frac{|B_t^i \cap B_{t-1}^j|}{|B_t^i \cup B_{t-1}^j|} \quad (1)$$

$B_t^i \cap B_{t-1}^j$ represents the intersection area of the two boxes, and $B_t^i \cup B_{t-1}^j$ represents the union area of the two boxes. Using a greedy matching algorithm, each detection box in the current frame is associated with the box in the previous frame with the maximum IOU value that exceeds a preset threshold of 0.5, establishing a cross-frame correspondence. If a detection box has no matching target in the previous frame, it is considered a newly appeared individual. Conversely, if a box in the previous frame has no matching target in the current frame, it is temporarily marked as temporarily lost, allowing backtracking matching within three consecutive frames to address transient occlusions. This strategy essentially builds a short-term memory buffer to maintain the target identity during the complete occlusion, and then restore the trajectory through spatial proximity and motion continuity when it reappears. This mechanism effectively maintains the identity continuity of athletes during complex movements, preventing misjudgments caused by sudden changes in posture or reflective clothing. After

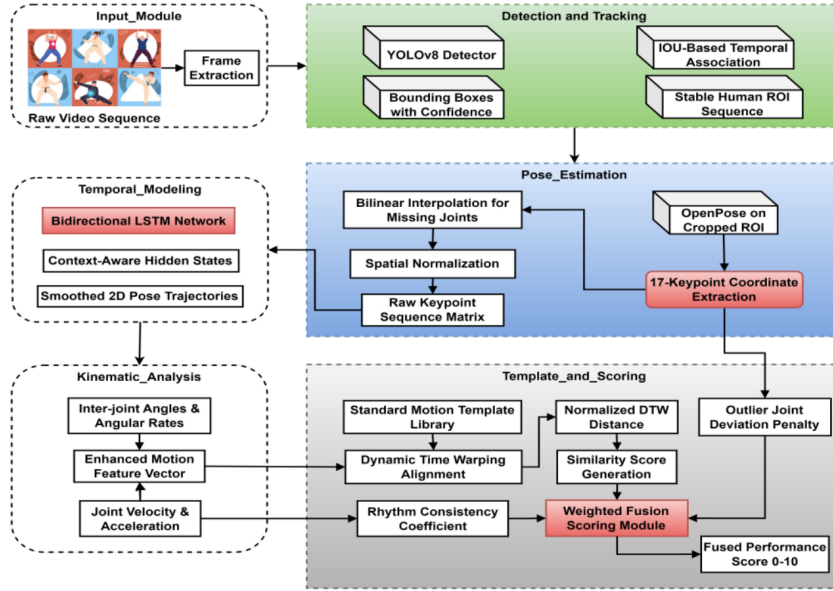


FIGURE 1. YOLOv8 and OpenPose fusion framework

matching is completed, the system maintains a queue of detection frame tracks for each tracked individual, and only successfully matched detection frames are used as valid input areas for subsequent OpenPose pose estimation. By cropping the bounding box from the original image and appropriately expanding it (by a factor of 1.2), it ensures that key limb parts are not cut off while suppression of background interference. The resulting cropped area serves as a stable and accurate Region of Interest (ROI), significantly reducing OpenPose's false detection rate and computational redundancy in complex backgrounds, and providing a reliable spatial constraint foundation for subsequent pose analysis of highly dynamic martial arts movements.

B. IMPROVING OPENPOSE KEYPOINT EXTRACTION AND TEMPORAL ALIGNMENT

1) OpenPose Keypoint Detection and Coordinate Compensation Based on Region Constraints

Based on the stable human bounding boxes generated by YOLOv8, it performs refined pose estimation on the cropped region of interest. The cropped image is scaled to 368×368 resolution to fit the OpenPose network input scale. It then uses a joint prediction mechanism using Part Affinity Fields (PAFs) and confidence heatmaps to simultaneously regress the spatial positions of 17 COCO standard joints (nose, left eye, right eye, left ear, right ear, left shoulder, right shoulder, left elbow, right elbow, left wrist, right wrist, left hip, right hip, left knee, right knee, left ankle, and right ankle). Through multi-stage iterative optimization, the network gradually refines the keypoint response map and limb joint field, enhancing its ability to model the topological structure of slender limbs. In the process of inference, the linear interpolation strategy based on time neighborhood is adopted to reconstruct the key points with low confidence

degree by using the coordinate pairs of the previous and the following frame. Under the condition of fixed frame rate, the method is equivalent to midpoint interpolation. For joint points with a confidence score lower than the threshold in a certain frame, they are determined to be potential missing points and are not discarded directly. Instead, they are reconstructed by combining the information of the previous and next frames in the temporal neighborhood. Specifically, let the coordinates of a joint point in the t th frame be p_t . If its confidence score is lower than the threshold, the coordinates p_{t-1} and p_{t+1} of the point in the previous frame $t-1$ and the next frame $t+1$ are used to reconstruct the point by linear interpolation:

$$\hat{p}_t = \frac{d_{t+1}}{d_{t-1} + d_{t+1}} p_{t-1} + \frac{d_{t-1}}{d_{t-1} + d_{t+1}} p_{t+1} \quad (2)$$

d_{t-1} and d_{t+1} are the temporal distances (in frames) between frame t and frames $t-1$ and $t+1$, respectively. Under uniform sampling, they are taken to be 1, so this equation degenerates into midpoint interpolation. This method effectively fills in temporarily lost keypoints while maintaining motion continuity, avoiding distortion in subsequent time series modeling due to data discontinuities. After interpolation, all keypoint coordinates are uniformly mapped back to the original image coordinate system to ensure physical consistency of spatial positions. Figure 2 shows the visualization of key point detection in Tai Chi movements. The four images, arranged in chronological order, show the typical evolution of the posture of the same athlete during martial arts movements. Red dots mark COCO's 17 joints, and red lines connect adjacent joints, forming the human skeleton structure. As the movement progresses, the upper limbs rotate in a spiral while the lower limbs maintain stable support. The key point trajectories

clearly reflect limb coordination and movement coherence. This sequence verifies the proposed framework's high-precision positioning capability in complex backgrounds and low-speed movements. It can also maintain stable key point output even in scenes with crossed arms and torso twists, providing a reliable foundation for subsequent temporal modeling and movement scoring.

2) Coordinate System Normalization and Key Point Sequence Organization

To eliminate spatial inconsistencies between frames due to body position offset, scale change, and perspective differences, the extracted keypoint coordinates are subjected to local normalization based on the detection frame. Taking the center point (c_x, c_y) of the current frame detection frame as the origin, a relative coordinate system is constructed, converting the absolute coordinates (x_i, y_i) of each joint point into an offset relative to the detection frame center:

$$\tilde{p}_i = \left(\frac{x_i - c_x}{w_b}, \frac{y_i - c_y}{h_b} \right) \quad (3)$$

w_b and h_b are the width and height of the detection box, respectively, and are used for scale normalization, ensuring that the distribution of key points is not affected by the distance of the athlete or the image size. This normalization strategy ensures that the posture representation of the same action at different spatial locations or imaging scales is consistent, improving the generalization capability of subsequent time series modeling. After completing the coordinate transformation, the system chronologically integrates the two-dimensional coordinates of the 17 joint points in each frame and their confidence levels into a three-dimensional tensor $K \in \mathbb{R}^{T \times 17 \times 3}$, where T is the total number of frames in the video sequence, and each entry contains the normalized coordinates and the original detection confidence level. This tensor forms the original keypoint sequence matrix, serving as the input for the subsequent timing optimization module. "Time optimization" specifically refers to the process of using a bidirectional LSTM network to perform temporal smoothing and trajectory reconstruction on keypoint sequences after OpenPose inference, thereby enhancing the temporal continuity and physical consistency of pose estimation. Throughout the sequence construction process, strict timestamp alignment is maintained to ensure that the pose data for each frame is precisely synchronized with the video timeline, avoiding timing misalignments caused by frame rate fluctuations or processing delays. Through the aforementioned spatial constraints, missing compensation, and coordinate normalization processes, the system achieves high-fidelity, structurally complete keypoint sequence extraction in complex martial arts scenes, providing a reliable data foundation for in-depth analysis of the dynamic characteristics of the movements. Figure 3 shows the distribution of localization errors for 17 standard keypoints in martial arts action videos. The horizontal axis shows the keypoint names, and the vertical axis shows the error

(in pixels) between the detected keypoints and the ground-truth annotations, the red plus sign indicates outliers. The errors for some keypoints on the head and face are relatively low, indicating that these joints are limitedly affected by occlusion and background interference during rapid motion, resulting in relatively stable detection. The slightly higher errors for the shoulders, elbows, and wrists reflect some jitter during rapid upper limb swings and crossovers. The higher median value for the elbows indicates that joint localization is prone to momentary deviation during large bends or rapid extensions. The overall error for the lower limb joints (hip, knee, and ankle) is relatively high, primarily due to the lower limbs being frequently occluded by the upper limbs or parts of the body during movement. Furthermore, the large displacements during flight or rotation increase OpenPose detection uncertainty. The red line shows the average error trend for each joint, showing a gradual increase from the upper body to the lower limbs. This demonstrates that the difficulty of joint detection in highly dynamic martial arts movements is closely related to spatial position and occlusion level, providing a clear reference for subsequent timing optimization and smoothing.

Figure 4, using the right wrist joint as an example, illustrates the longitudinal motion trajectory and its temporal alignment across consecutive video frames. The horizontal axis represents the frame index, and the vertical axis represents the Y-axis position of the joint in the image coordinate system. The black curve represents the ideal true trajectory, reflecting the dynamic pattern of the action under standard execution. The red dashed line represents the original unsmoothed detected trajectory. Significant jitter and localized offsets can be observed during rapid arm swings and complex limb crossings. This is due to the instability of OpenPose's keypoint localization under high-speed motion or partial occlusion. The blue solid line represents the trajectory after timing optimization and smoothing, significantly reducing high-frequency noise and transient outlier offsets, bringing the motion trend closer to the true trajectory. This improvement not only restores the continuity of the wrist trajectory but also more accurately reflects the biomechanical characteristics of the action, providing reliable basic data for subsequent joint angle calculation, motion velocity analysis, and action scoring. The trajectory differences clearly demonstrate the importance of temporal modeling in addressing keypoint jitter and transient false detections.

C. LSTM NETWORK-DRIVEN KEYPOINT SEQUENCE SMOOTHING

1) Bidirectional LSTM Network Structure Design and Multi-Channel Time Series Modeling

To address the issue of keypoint coordinate jitter caused by high-speed movement, limb occlusion, or sudden posture changes in martial arts movements, a temporal smoothing architecture based on a bidirectional long short-term memory (Bidirectional LSTM) network is constructed. The nor-

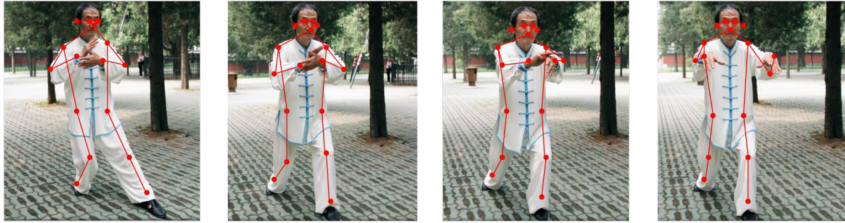


FIGURE 2. Key point detection

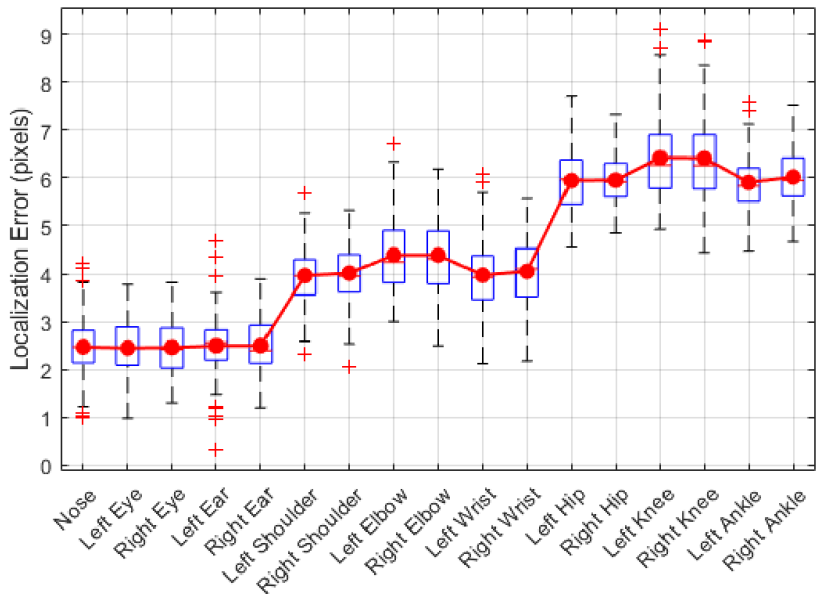


FIGURE 3. Distribution of positioning errors of key points

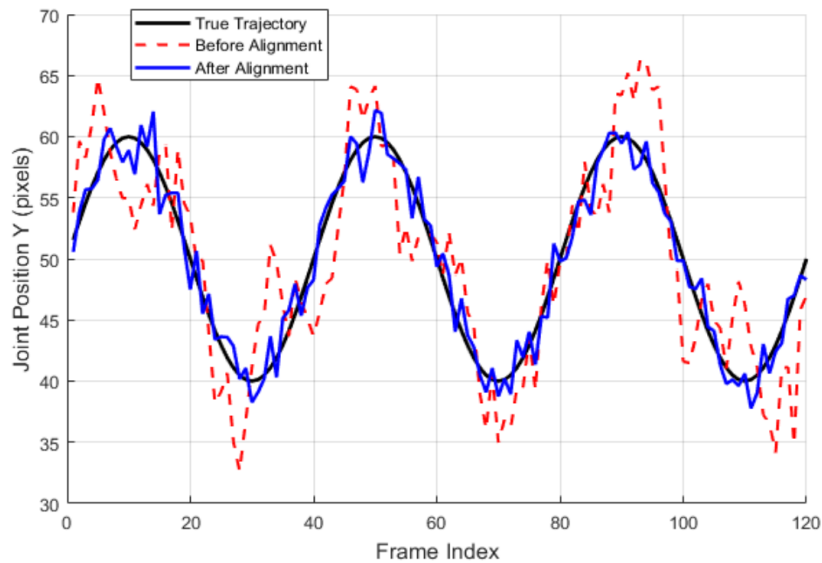


FIGURE 4. Time alignment of motion trajectories

malized keypoint sequence output from the previous stage is organized into a multi-channel time series input. Each time step contains the two-dimensional spatial coordinates of 17 joint points and their corresponding confidence scores, form-

ing a dimensional input matrix. This architecture processes the coordinates and detection confidence of each joint point as independent channels in parallel, preserving the integrity of the original observations. The network constructs chains

of LSTM units in both the forward and backward directions along the timeline. The forward chain captures historical motion trends from the starting frame to the current frame, while the backward chain models future contextual information from the last frame to the current frame. Therefore, the input of each time step constitutes a 51-dimensional feature vector, in which the first 34 dimensions represent the normalized (x, y) coordinates of 17 key points, and the last 17 dimensions correspond to the confidence score of each point. At each time step, the forward LSTM unit receives input and updates its hidden state, while the backward LSTM unit updates based on information from subsequent frames. Ultimately, the two are concatenated into a joint hidden representation:

$$h_t = [\vec{h}_t; \overleftarrow{h}_t] \quad (4)$$

$\vec{h}_t$ represents the forward hidden state, and $\overleftarrow{h}_t$ represents the backward hidden state. This mechanism enables the model to simultaneously reference the initial evolution path and subsequent convergence trend of the movement when predicting the pose in the current frame, enhancing its ability to discern transient outliers. Through the synergistic effect of the forget gate, input gate, and output gate within the LSTM, it selectively retains long-term motion patterns and suppresses high-frequency noise interference. This is particularly true during intense movements such as takeoffs and rotations, effectively preventing jumps in key points caused by brief periods of lag. The network stacks three layers of bidirectional LSTM, each containing 128 hidden units. A dropout mechanism (with a dropout rate set to 0.3) is applied between layers to prevent overfitting and improve the model's generalization performance across different athletes and execution rhythms.

2) Context-aware Keypoint Trajectory Reconstruction and Smooth Output

After completing the bidirectional encoding of the entire sequence, the network maps the final hidden state back to the original output space through the fully connected layer to generate a sequence of key point coordinates that has been optimized for time series. This process is not a simple filtering, but a trajectory reconstruction based on the learned laws of motion dynamics to ensure that the output posture conforms to the biomechanical constraints of the human body and maintains the spatiotemporal coherence of the movement details. For each joint point, its coordinate sequence is regarded as a controlled dynamic system in the time dimension. LSTM uses internal memory units to memorize typical motion patterns, such as the sinusoidal characteristics of arm swings and the stability of the lower limb support phase, so that when a short-term deviation occurs during detection, the most likely true position can be inferred based on the context. In particular, the model uses the mean squared error (MSE) loss function to drive optimization during the training phase. The goal is to

minimize the spatial deviation between the smoothed output $\hat{P}_t$ and the true normalized key point P_t :

$$L = \frac{1}{T} \sum_{t=1}^T \|\hat{P}_t - P_t\|^2 \quad (5)$$

The norm $\|\cdot\|$ represents the square of the Euclidean distance. This loss function forces the network to minimize coordinate jitter while preserving the motion morphology. During the inference phase, a complete sequence of key points is input, and the model outputs a stream of corrected coordinates frame by frame, forming a smooth, continuous, and logically self-consistent posture trajectory. Notably, because the bidirectional architecture relies on complete sequence input, the system employs a sliding window strategy in real-time analysis, using fixed-length local segments as input and combining overlapping frames to ensure boundary continuity. The resulting keypoint sequence not only suppresses high-frequency oscillations caused by unstable detection but also corrects local distortions due to occlusion or false detection, significantly improving the physical plausibility and temporal consistency of the pose representation, providing a high-fidelity motion foundation for subsequent action segmentation and scoring. This entire process achieves data-driven, end-to-end trajectory optimization without requiring a priori motion models, enhancing the system's robustness and adaptability in complex martial arts scenarios. Table 1 lists the core training and structural parameters used by the bidirectional LSTM network during the key point sequence smoothing phase. These parameters cover optimization strategies, model capacity, training control, and regularization mechanisms, ensuring the network maintains numerical stability and generalization capabilities while learning complex motion patterns. The parameter settings comprehensively consider the requirements for action sequence continuity and computational efficiency, enabling accurate trajectory modeling of highly dynamic martial arts movements. All parameters are fine-tuned through multiple experiments to achieve optimal smoothing and convergence performance.

TABLE 1. Bidirectional LSTM network parameter configuration

Parameter Name	Value	Description
Initial Learning Rate	0.001	Step size for optimizer at the beginning
Batch Size	32	Number of video frames processed per batch
Training Epochs	150	Total number of training iterations over dataset
Sliding Window Length (frames)	60	Number of consecutive frames used as input
Gradient Clipping Threshold	5	Maximum norm to clip gradients and prevent explosion
Weight Decay	0.0001	L2 regularization coefficient to reduce overfitting
Early Stopping Patience	20	Number of epochs with no improvement before stopping
Learning Rate (LR) Patience	8	Epochs with no loss decrease before reducing LR

D. ACTION MATCHING AND SCORE GENERATION BASED ON DYNAMIC TIME WARPING

1) Kinematic Feature Extraction and Standard Action Template Library Construction

Based on the smooth and continuous keypoint trajectories obtained, their dynamic evolution patterns are further analyzed to extract discriminative kinematic features to support movement quality assessment. For each joint point, first- and second-order derivatives are calculated based on its normalized coordinate sequence in the time domain, representing the velocity and acceleration of the point, respectively. The derivatives are numerically approximated using the central difference method to avoid phase delays applied by forward or backward differencing, ensuring that the features are highly synchronized with the original movement rhythm. Simultaneously, joint angle models of key kinematic chains are constructed. Typical limb structures, such as the shoulder-elbow-wrist and hip-knee-ankle, are selected, and the evolution of joint flexion and extension angles is calculated using the vector angle formula. The time-varying nature of the angle sequence is further analyzed by first-order differences within a sliding window to extract angular velocities, reflecting the force rhythm and transition rate of the movement. All features—including joint linear velocities, accelerations, joint angles, and their rates of change—are organized into a high-dimensional feature vector sequence in a time-aligned manner, forming a dynamic representation space for the movement. This feature set not only preserves the spatial configuration of the posture but also encodes key evaluation dimensions of movement execution, such as force, fluidity, and coordination. On this basis, the system can adaptively identify the action transition point by detecting the local peak and significant fluctuation of the joint angle change rate, and realize the dynamic division of the starting and ending boundary, so as to overcome the inadaptability of the fixed time window to the individual rhythm difference.

2) DTW-Based Nonlinear Alignment and Comprehensive Scoring Mechanism

To achieve accurate comparison between the test action and the standard template, the Dynamic Time Warping (DTW) algorithm is used to address the common rhythm differences during action execution. Traditional Euclidean distance requires strict time synchronization between the two sequences, which makes it difficult to cope with the acceleration, deceleration, or pauses caused by individual differences in martial arts movements. DTW, on the other hand, constructs a nonlinear alignment path, allowing for flexible scaling on the time axis, and finding the optimal matching path while maintaining the action structure. Let the test action feature sequence be $X = \{x_1, x_2, \dots, x_m\}$ and the standard template sequence be $Y = \{y_1, y_2, \dots, y_n\}$. DTW recursively calculates the cumulative distance matrix D , whose elements are defined as:

$$D(i, j) = \|x_i - y_j\| + \min\{D(i-1, j), D(i, j-1), D(i-1, j-1)\} \quad (6)$$

$\|x_i - y_j\|$ is the Euclidean distance between the feature vectors of frame i and frame j , and $D(i, j)$ represents the minimum cumulative mismatch cost from starting point to position (i, j) . The final DTW distance reflects the overall deviation between the two sequences under optimal alignment. This distance value is length-normalized and converted into a preliminary similarity score $S_{\text{sim}} = \exp(-\alpha \cdot D_{\text{norm}})$, where α is the scale adjustment factor to ensure that the score is interpretable within the range $[0, 1]$. To further enhance the fairness and technical sensitivity of the scoring, a two-dimensional weighting mechanism is applied. On the one hand, a threshold for joint spatial deviation is set to penalize key posture errors that exceed a reasonable range; on the other hand, a rhythm consistency coefficient is calculated to measure the degree of consistency between the test movement and the template in terms of start and end timing, force nodes, and pause positions. The final comprehensive score is:

$$\text{Score} = w_1 \cdot S_{\text{sim}} + w_2 \cdot C_{\text{rhythm}} - \beta \cdot E_{\text{outlier}} \quad (7)$$

C_{rhythm} represents the rhythm matching degree; E_{outlier} represents the over-limit error term; w_1 , w_2 , and β are adjustable weight parameters, calibrated according to martial arts scoring rules. This scoring mechanism considers both morphological similarity and execution standardization, enabling a comprehensive evaluation from "formal resemblance" to "spiritual resemblance," providing a scientific and quantifiable basis for automated scoring of martial arts movements.

III. PERFORMANCE VERIFICATION OF YOLOV8 AND OPENPOSE FUSION FRAMEWORK

A. EXPERIMENTAL DATA

The experimental data is collected from a standardized video library at a national martial arts training base, covering the disciplines of Changquan, Nanquan, and Tai Chi. It contains 180 complete routines performed by 60 professional athletes, with 60 segments for each style, sampled at 30 frames per second. All videos are recorded under natural lighting and complex background conditions and include typical movements such as jabs, kicks, jumps, tumbles, balances, and rotations, fully capturing the high dynamics, multi-posture, and self-occlusion characteristics of martial arts. All videos are double-blindly scored by five national-level referees, with the average score serving as the expert scoring benchmark for system output consistency analysis. Key movement nodes are annotated by professional coaches for segmentation boundary verification. The entire dataset has been anonymized, and the research process follows the ethical principles of research involving human subjects in the Declaration of Helsinki, conforms to scientific research ethics, and ensures the authenticity and reproducibility of the experimental results.

B. YOLOV8 DETECTION CONFIDENCE AND CROSS-FRAME IOU CHANGES

YOLOv8 detection is performed frame by frame on the input video sequence to obtain the athlete's bounding box coordinates and confidence scores. Candidate boxes with confidence scores above a threshold are selected as valid detection results. Subsequently, a cross-frame matching method based on IOU is used to calculate the overlap between the detection boxes of the current frame and the previous frame. If the IOU value exceeds a set threshold, it is considered a continuous trajectory of the same target. Otherwise, it is marked as a potentially lost or newly appeared target. During this process, corresponding detection trajectory sequences are established for three targets performing different martial arts movements (Target 1: Standing Front, Target 2: Side Kick, and Target 3: Spinning). Their IOU trends across the entire video clip are then analyzed. By recording and archiving IOU values frame by frame, the stability of cross-frame matching is reflected. The number of frames with IOU values below the threshold is used to calculate the target's tracking reliability during complex movements. This establishes a quantitative evaluation framework, providing an accurate input foundation for subsequent pose estimation and time series modeling. Figure 5 shows Frame Index (frame number, variable frames, range 1–120) on the horizontal axis and Detection Confidence (human detection confidence of YOLOv8 in each frame) on the vertical axis. Three types of data columns are shown: 1) the original frame-by-frame confidence (shown as blue scattered dots), 2) the moving average based on a window size of 7 (blue solid line) to highlight the overall trend, and 3) the local standard deviation calculated based on the same window to construct the light blue band of "confidence $\pm$ std" (std is short for standard deviation). In addition, the detection

threshold of 0.7 is marked with a red dotted line, and all frames below the threshold (< 0.7) are highlighted with red hollow dots. The data also includes a small confidence histogram, showing the confidence distribution (frequency distribution). The data shows fluctuations in the overall confidence level, which fluctuates roughly between 0.6 and 1. There are significant drops in the data around frames 40–44 and 88–95, corresponding to widening of the light blue confidence intervals and clustering of raw points below the threshold. This change indicates that detection quality degrades during these short periods (typically corresponding to fast motion, blur, or occlusion), potentially compromising the input quality for subsequent pose estimation. The moving average effectively smooths out the noise, and the widening confidence intervals indicate increased uncertainty, which is valuable for deciding whether to discard or back-track frames.

Figure 6 shows the Frame Index on the horizontal axis and the IOU on the vertical axis. The three curves show the IOU values of the three targets across consecutive frames (target 1 shows good stability, target 2 shows moderate fluctuation, and target 3 shows the greatest fluctuation). A gray semi-transparent band highlights two simulated occlusion windows (frames 38–44 and 88–95), and an IOU threshold of 0.5 is drawn (black dashed line). When the IOU of a frame falls below the threshold, the script draws a prominent filled dot at that point as a warning. The inset bar graph in the lower right corner counts the number of frames in which the IOU of each target falls below 0.5, quantifying the frequency of tracking failures/instabilities. The data reveals that target 1 experiences a significant IOU drop in the occluded section but then recovers (indicating that the backtracking matching mechanism may be effective after brief occlusion). Target 3 frequently drops below 0.5 at multiple moments, indicating that this target is prone to identity confusion or severe bounding box jitter during tracking. The trough of IOU corresponds to the period when detection confidence decreases (that is, when the detection value is low, the subsequent IOU matching also weakens), which shows that there is a significant coupling relationship between detection quality and cross-frame tracking stability - when the detection confidence decreases, IOU matching is more likely to be inaccurate, and additional backtracking or appearance feature assistance is needed to avoid identity switching.

C. COMPARISON OF HUMAN BODY DETECTION STABILITY

To comprehensively evaluate the robustness and real-time performance of different detection models in martial arts action scenarios, six typical difficult movements—jabs, kicks, jumps, rolls, balances, and rotations—are selected as samples. Under standardized data pre- and post-processing thresholds, the proposed YOLOv8, along with traditional YOLOv8, Faster R-CNN (Faster Region-based Convolutional Neural Network), YOLOv5, and SSD (Single Shot

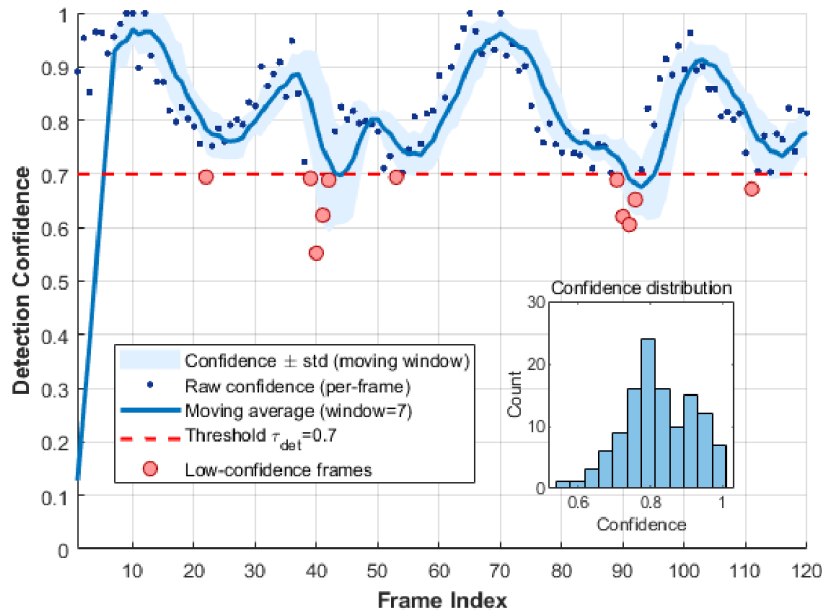


FIGURE 5. YOLOv8 detection confidence per frame

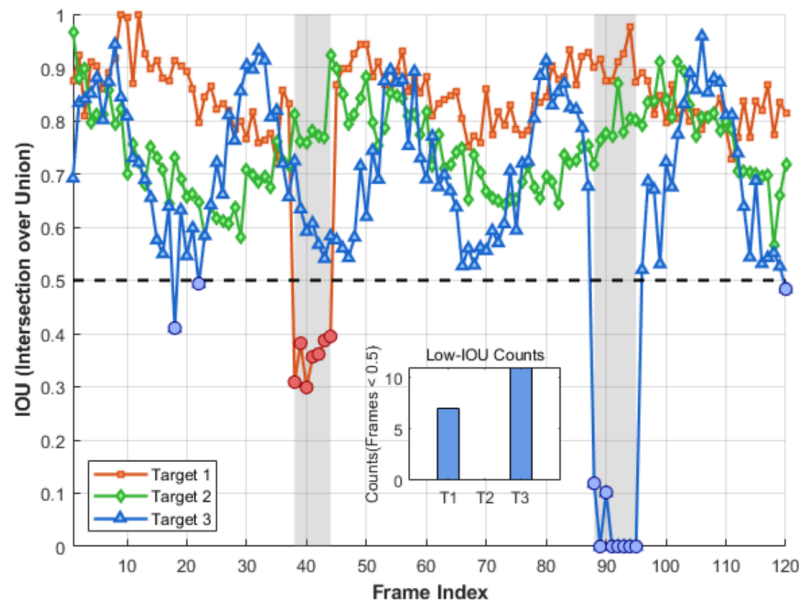


FIGURE 6. Multi-object cross-frame IOU trajectory

MultiBox Detector)—are deployed to perform inference on each action category. Detection accuracy is measured using mAP@0.5 (mean Average Precision (IoU = 0.5)) to reflect the consistency of the model's object localization under different movements. The system's end-to-end Frames Per Second (FPS) is also recorded to assess its feasibility for real-time analysis. Each experiment is repeated five times, and the mean and standard deviation of the performance metrics are calculated. Error bars are applied to analyze the volatility of the results to ensure statistical significance and reproducibility. Figure 7 compares the performance of five

detection models in terms of detection accuracy (mAP@0.5) and processing speed (FPS) for six typical martial arts movements. For fast movements such as jabs and kicks, the proposed YOLOv8 method maintains optimal detection stability, with average mAP@0.5 values of 0.863-0.942, thanks to its improved feature fusion mechanism and sensitive response to small-scale key gestures. For complex movements such as jumps and tumbles, which involve dramatic displacements and brief occlusions, traditional YOLOv5 and SSD experience significant accuracy degradation. The proposed method effectively mitigates false and missed

detections through enhanced context modeling. While Faster R-CNN possesses strong static recognition capabilities, its region proposal mechanism lags in response in high-speed motion scenes, limiting its overall performance. In terms of real-time performance, the proposed method maintains high accuracy while maintaining an advantageous frame rate, with average FPS values of 115.2–124.3. While SSD offers a higher frame rate, this comes at the expense of detection reliability. In this paper, YOLOv8 achieves a better balance between accuracy and speed, showing stronger robustness in different categories of actions while having a lower standard deviation. This verifies its adaptability to the highly dynamic and changeable posture scenes of martial arts, and provides a stable and reliable detection foundation for subsequent posture estimation and scoring.

D. POSITIONING ACCURACY OF POSTURE KEY POINTS

To evaluate the robustness of different pose estimation methods under occlusion, two typical occlusion scenarios, crossed arms and overlapping legs, are selected to capture the common limb overlap and spatial occlusion encountered in martial arts. The core evaluation metrics used are PCK (Percentage of Correct Keypoints) and MPJPE (Mean Per Joint Position Error), reflecting the relative accuracy and absolute positioning error of keypoints, respectively. Under the same input conditions, the performance of the proposed OpenPose+LSTM temporal optimization is compared with the original OpenPose, AlphaPose, and HRNet (High-Resolution Network) models. Each experiment is repeated five times, and the mean and standard deviation of each metric are calculated for both occlusion modes. Error bars are used to analyze the stability of the methods. A PCK threshold of 0.5 is set, and MPJPE directly reflects the accumulated pixel-level error, ensuring that the evaluation results are both biomechanically plausible and spatially accurate. Figure 8 shows the comparison of keypoint localization accuracy between the proposed OpenPose+LSTM framework and the original OpenPose, AlphaPose, and HRNet in two typical occlusion scenarios: crossed arms and overlapping legs. In terms of the PCK metric, the proposed method achieves the best joint recognition rates under both occlusion conditions, with average rates ranging from 88.7% to 90.3%. The advantage is particularly pronounced when arms are crossed, demonstrating that temporal modeling effectively mitigates joint confusion caused by overlapping limbs. By applying a bidirectional LSTM to dynamically smooth the keypoint sequence, the system leverages contextual information from frames before and after the action to reasonably infer the potential positions of occluded joints, thereby improving spatial consistency. In contrast, static pose estimation methods such as the original OpenPose lack temporal compensation mechanisms, making keypoints prone to misalignment or drift under severe occlusion. While HRNet maintains high-resolution features, it struggles to maintain a stable response in scenes subject to rapid

motion and deformation. MPJPE results further validate the proposed method's ability to control pixel-level positioning error, with an average range of 48.2 pixels to 51.6 pixels. This demonstrates that temporal optimization not only improves overall accuracy but also enhances the geometric precision of keypoints. This advantage is also evident in scenes with overlapping legs, demonstrating the model's robust adaptability to complex spatial occlusions. Overall, the results demonstrate that the posture correction mechanism integrated with temporal modeling plays a key role in the analysis of highly dynamic martial arts movements, providing a high-fidelity posture foundation for subsequent movement segmentation and scoring.

E. COMPARISON OF TIME SERIES SMOOTHING EFFECTS

To quantitatively evaluate the temporal stability of different pose estimation methods in dynamic martial arts routines, three distinct styles of boxing are selected for testing: Changquan, Nanquan, and Tai Chi. These styles represent high-frequency, fast-paced, aggressive, and compact, and low-speed, continuous movement patterns, respectively. Keypoint jitter variance (KJV) and trajectory continuity error (TCE) are used as evaluation metrics. KJV measures the mean variance of coordinate fluctuations at the same joint within a short time window, reflecting local temporal smoothness; TCE detects trajectory abrupt changes by using forward and backward velocity differences to assess the physical plausibility of the motion path. Both metrics are calculated for 17 joints in the complete action sequence of each routine, and the mean and standard deviation are calculated. All tests are conducted under the same pre- and post-processing conditions to ensure comparability and statistical consistency of the evaluation results. Table 2 compares the temporal stability of different pose estimation methods for three distinct martial arts styles: Changquan, Nanquan, and Tai Chi. The proposed OpenPose+LSTM framework significantly outperforms the original OpenPose, AlphaPose, and HRNet in both keypoint jitter variance and trajectory continuity error, with mean KJV values of $(1.52\text{--}2.03) \times 10^{-3}$ pixels² and mean TCE values of $(0.28\text{--}0.36)$ pixels/frame², demonstrating its superior modeling capabilities in suppressing keypoint jitter and maintaining motion coherence. This advantage stems from the bidirectional LSTM's explicit learning of long-range temporal dependencies, enabling it to leverage the motion context of previous and subsequent frames to effectively identify and correct transient anomalies caused by rapid motion, limb occlusion, or detection misalignment, thereby generating smooth trajectories that conform to biomechanical laws. In contrast, static frame-by-frame detection methods lack temporal consistency constraints and are prone to coordinate jumps in highly dynamic movements, especially during the leaps of Changquan and the short bursts of force in Nanquan. Although HRNet has high-resolution feature extraction capabilities, its spatial accuracy gain has not been

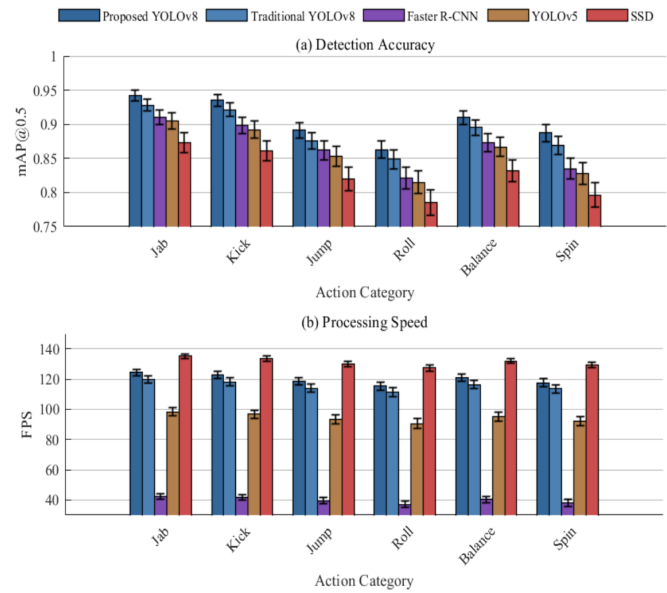


FIGURE 7. Human body detection stability

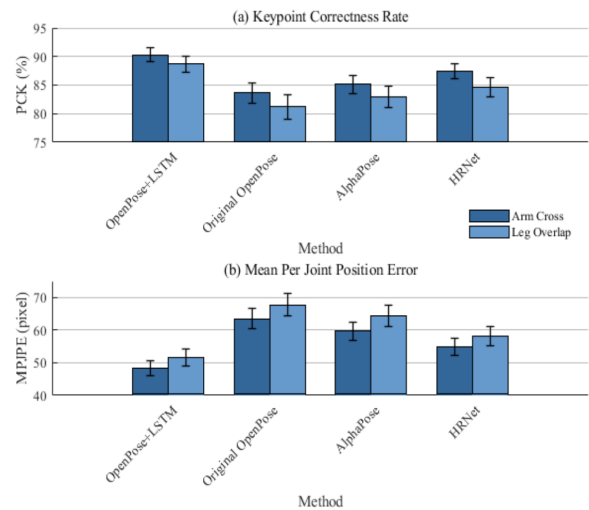


FIGURE 8. Key point positioning accuracy

fully converted into improved temporal stability, and local oscillations still exist. Notably, the proposed method demonstrates optimal performance in slow, continuous Tai Chi-like movements, demonstrating its suitability not only for high-frequency motion suppression but also for finely capturing subtle changes in low-speed movements. The overall results validate the necessity of temporal modeling in martial arts motion analysis. The proposed framework achieves robust posture optimization across rhythmic and structural motion patterns, providing a high-fidelity dynamic representation foundation for subsequent motion segmentation and scoring.

TABLE 2. Time series smoothing effect

Method	Style	KJV ($\times 10^{-3}$ pixel ²)	TCE (pixel/frame ²)
OpenPose+LSTM	Changquan	1.87 $\pm$ 0.13	0.32 $\pm$ 0.03
OpenPose+LSTM	Nanquan	2.03 $\pm$ 0.16	0.36 $\pm$ 0.04
OpenPose+LSTM	Taiji	1.52 $\pm$ 0.11	0.28 $\pm$ 0.02
Original OpenPose	Changquan	4.96 $\pm$ 0.31	0.89 $\pm$ 0.07
Original OpenPose	Nanquan	5.21 $\pm$ 0.35	0.93 $\pm$ 0.08
Original OpenPose	Taiji	3.84 $\pm$ 0.27	0.75 $\pm$ 0.06
AlphaPose	Changquan	4.12 $\pm$ 0.28	0.76 $\pm$ 0.06
AlphaPose	Nanquan	4.43 $\pm$ 0.30	0.81 $\pm$ 0.07
AlphaPose	Taiji	3.35 $\pm$ 0.24	0.68 $\pm$ 0.05
HRNet	Changquan	3.68 $\pm$ 0.25	0.69 $\pm$ 0.05
HRNet	Nanquan	3.94 $\pm$ 0.28	0.73 $\pm$ 0.06
HRNet	Taiji	2.97 $\pm$ 0.21	0.61 $\pm$ 0.05

F. ANALYSIS OF SCORING CONSISTENCY AND STANDARD FIT

To verify the alignment of the constructed scoring system with human expert judgment, the automatically generated quantitative scores are compared with the average of the independent scores of five national-level judges. The Pearson correlation coefficient (r) is used to measure the consistency of the scoring trends, reflecting the model's sensitivity to differences in movement quality. The root mean square error (RMSE) is also calculated to assess the absolute deviation between the system output and the expert mean, reflecting the accuracy of the scoring. The tests are conducted using the same standard movement template library and DTW matching parameters, comparing the scoring results driven by five detection frameworks: the improved YOLOv8 proposed in this paper, the traditional YOLOv8, Faster R-CNN, YOLOv5, and SSD. All methods integrated the same pose estimation and scoring logic to ensure that the differences are solely due to detection stability. Table 3 shows the consistency and fit between the scoring systems and expert judges driven by different object detection methods. The improved YOLOv8 proposed in this paper significantly outperforms traditional YOLOv8, Faster R-CNN, YOLOv5, and SSD in both Pearson correlation coefficient (mean 0.925-0.941) and root mean square error (mean 0.39-0.45), demonstrating that its generated scores are not only highly consistent with expert trends but also have lower bias. Its detection robustness in complex martial arts movements is enhanced—it maintains high-confidence localization despite rapid motion, limb occlusion, and small objects, ensuring the accuracy of subsequent action segmentation and template matching. In contrast, while traditional detectors offer some real-time performance, they are prone to inter-frame jitter in highly dynamic scenes, resulting in loss or distortion of key action segments. Faster R-CNN, limited by its region proposal mechanism, experiences response lags under high-speed motion, compromising temporal integrity. SSD, due to its insufficient fusion of multi-scale features, increases the misjudgment rate for fine movements, further widening the scoring bias. Notably, this method demonstrated optimal consistency across both the explosive movements of Changquan and the continuous transitions of Tai Chi, validating its broad adaptability to multi-rhythm and multi-structured martial arts routines. Overall results demonstrate that accurate and stable front-end detection is a key prerequisite for achieving high-confidence automated scoring. The proposed framework significantly improves the cognitive alignment between the system and human judges through the collaborative optimization of detection, posture, and scoring.

TABLE 3. Comparison of scoring consistency across different front-end detection and tracking strategies

Detection Method	Style	Pearson r	RMSE (Score)
Proposed YOLOv8	Changquan	0.932 ± 0.018	0.42 ± 0.05
Proposed YOLOv8	Nanquan	0.925 ± 0.021	0.45 ± 0.06
Proposed YOLOv8	Taiji	0.941 ± 0.016	0.39 ± 0.04
Traditional YOLOv8	Changquan	0.896 ± 0.025	0.53 ± 0.07
Traditional YOLOv8	Nanquan	0.883 ± 0.028	0.57 ± 0.08
Traditional YOLOv8	Taiji	0.902 ± 0.023	0.51 ± 0.06
Faster R-CNN	Changquan	0.847 ± 0.031	0.68 ± 0.09
Faster R-CNN	Nanquan	0.832 ± 0.034	0.72 ± 0.10
Faster R-CNN	Taiji	0.856 ± 0.029	0.65 ± 0.08
YOLOv5	Changquan	0.864 ± 0.029	0.62 ± 0.08
YOLOv5	Nanquan	0.851 ± 0.032	0.66 ± 0.09
YOLOv5	Taiji	0.873 ± 0.026	0.58 ± 0.07
SSD	Changquan	0.821 ± 0.036	0.75 ± 0.11
SSD	Nanquan	0.808 ± 0.039	0.79 ± 0.12
SSD	Taiji	0.834 ± 0.033	0.71 ± 0.10

IV. CONCLUSION

This paper constructs an automated recognition and scoring framework for martial arts routines, aiming to analyze the complex "fabric of motion" inherent in these forms. It integrates YOLOv8 and OpenPose for high-precision human body detection and keypoint extraction, effectively treating the body's kinetic chain as a series of connected threads to precisely capture the nuanced, woven sequences of movement. Cross-frame identity consistency is ensured through IoU matching. A bidirectional LSTM is applied to model the keypoint sequence temporally, significantly suppressing pose jitter caused by rapid motion and occlusion. Kinematic features are calculated based on smoothed trajectories, and dynamic thresholding is combined to achieve adaptive segmentation of action segments, overcoming the limitations of fixed time windows. The DTW algorithm is used in the scoring phase for nonlinear alignment with a standard movement library, effectively addressing individual differences in rhythm and amplitude among athletes and improving both fairness and accuracy in the comparison. Experiments demonstrate that the method demonstrates excellent detection stability, pose estimation accuracy, and temporal smoothing across a variety of martial arts styles. The scoring results are highly consistent with expert judgment, with mean mAP@0.5 values ranging from 0.863 to 0.942, mean FPS values ranging from 115.2 to 124.3, mean PCK values ranging from 88.7% to 90.3%, and mean MPJPE values ranging from 48.2 pixels to 51.6 pixels. This framework implements end-to-end closed-loop analysis from video input to quantitative scoring, much like the advanced quality control systems to ensure consistent output. This provides objective and repeatable technical support for martial arts competition judging while also verifying its generalization potential in analyzing the complex, high-dynamic sports movements found beyond the "woven fabric" of martial arts.

AUTHOR CONTRIBUTIONS

Conceptualization – Dan Zhao and Yuanqi Ma; methodology – Dan Zhao and Yuanqi Ma; investigation – Dan Zhao; writing-original draft preparation – Dan Zhao and Yuanqi Ma. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

The authors were supported by, 1. Construction of a Dynamic Inheritance Model for Wolf Warrior Martial Arts in Western Guangxi from the Perspective of "Integration of Sports and Tourism" Project Source: 2024 Guangxi University Young and Middle aged Teachers' Research Basic Ability Enhancement, Project Project Number: 2024KY0616 2. "Research on the Practical Teaching System of Physical Education Majors in Local Universities Based on Teacher Education Professional Certification Standards" Source: Hechi University 2024 School level Education and Teaching Reform Project, Project Number: 2024EB014

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] Y. Pang, Y. Wang, Q. Wang, F. Li, C. Zhang, and C. Ding, "Applications of AI in martial arts: A survey," *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, vol. 239, no. 2, pp. 301-331, 2025, doi: 10.1177/17543371241273827.
- [2] Y. Shang, "Advancing Martial Arts Training: Neural Network-Based Recognition and Assistance Systems in Biotechnological Applications," *Journal of Commercial Biotechnology*, vol. 29, no. 5, pp. 84-94, 2024, doi: 10.5912/jcb1913.
- [3] Z. Li, "A method for recognising wrong actions of martial arts athletes based on keyframe extraction," *International Journal of Biometrics*, vol. 16, no. 3-4, pp. 256-271, 2024, doi: 10.1504/IJBM.2024.138228.
- [4] L. Song, R. A. B. A. Rashid, Y. Li, and B. Wang, "Predictive Decision Analytics for Membership Retention and Expansion in Martial Arts Organisations," *Decision Making: Applications in Management and Engineering*, vol. 8, no. 2, pp. 21-35, 2025, doi: 10.31181/dmame8220251474.
- [5] Y. Hou, F. M. Seydou, and S. Kenderdine, "Unlocking a multimodal archive of Southern Chinese martial arts through embodied cues," *Journal of Documentation*, vol. 80, no. 5, pp. 1148-1166, 2024, doi: 10.1108/JD-01-2022-0027.
- [6] C. R. Chung, S. C. Yeh, E. H. K. Wu, and S. Y. Lin, "Qigong master: A Qigong-based attention training game using action recognition and balance analysis," *IEEE Transactions on Games*, vol. 16, no. 2, pp. 376-383, 2023, doi: 10.1109/TG.2023.3274148.
- [7] J. Zhuang, J. Sun, and G. Yuan, "Arrhythmia diagnosis of young martial arts athletes based on deep learning for smart medical care," *Neural Computing and Applications*, vol. 35, no. 20, pp. 14641-14652, 2023, doi: 10.1007/s00521-021-06159-4.
- [8] V. H. Le, "Deep learning-based for human segmentation and tracking, 3D human pose estimation and action recognition on monocular video of MADS dataset," *Multimedia Tools and Applications*, vol. 82, no. 14, pp. 20771-20818, 2023, doi: 10.1007/s11042-022-13921-w.
- [9] X. Cui and R. Hu, "Application of intelligent edge computing technology for video surveillance in human movement recognition and Taekwondo training," *Alexandria Engineering Journal*, vol. 61, no. 4, pp. 2899-2908, 2022, doi: 10.1016/j.aej.2021.08.020.
- [10] X. Sun, Y. Wang, and J. Khan, "Hybrid LSTM and GAN model for action recognition and prediction of lawn tennis sport activities," *Soft Computing*, vol. 27, no. 23, pp. 18093-18112, 2023, doi: 10.1007/s00500-023-09215-4.
- [11] V. Romanenko, S. Piatysotska, L. Podrigalo, M. Baibikov, N. Boychenko, and O. Volodchenko, "Methodology for evaluating the "Go/No-Go" reaction in martial arts," *Journal of Physical Education and Sport*, vol. 24, no. 12, pp. 2139-2146, 2024, doi: 10.7752/jpes.2024.12312.
- [12] M. Rodrigo, C. Cuevas, D. Berjón, and N. García, "Automatic highlight detection in videos of martial arts tricking," *Multimedia Tools and Applications*, vol. 83, no. 6, pp. 17109-17133, 2024, doi: 10.1007/s11042-023-16003-7.
- [13] S. Qu, "Cloud IoT-Oriented Neural Network-Based Taekwondo Teaching Scheme," *Journal of Multimedia Information System*, vol. 11, no. 1, pp. 83-96, 2024, doi: 10.33851/JMIS.2024.11.1.83.
- [14] H. Mohammadzade, S. Hosseini, M. R. Rezaei-Dastjerdehei, and M. Tabejamaat, "Dynamic time warping-based features with class-specific joint importance maps for action recognition using Kinect depth sensor," *IEEE Sensors Journal*, vol. 21, no. 7, pp. 9300-9313, 2021, doi: 10.1109/JSEN.2021.3051497.
- [15] L. Pei, S. Xia, L. Chu, F. Xiao, Q. Wu, W. Yu, et al., "MARS: Mixed virtual and real wearable sensors for human activity recognition with multidomain deep learning model," *IEEE Internet of Things Journal*, vol. 8, no. 11, pp. 9383-9396, 2021, doi: 10.1109/IIOT.2021.3055859.
- [16] Y. Qiao, "Application of Virtual Simulation Technology in Martial Arts Routine Movement Analysis," *Journal of Electronics and Information Science*, vol. 8, no. 4, pp. 7-13, 2023, doi: 10.23977/jeis.2023.080402.
- [17] F. E. Ait-Bennacer, A. Aaroud, K. Akodadi, and B. Cherradi, "Applying Deep Learning and Computer Vision Techniques for an e-Sport and Smart Coaching System Using a Multiview Dataset: Case of Shotokan Karate," *Int. J. Online Biomed. Eng.*, vol. 18, no. 12, pp. 35-53, 2022, doi: 10.3991/ijoe.v18i12.30893.
- [18] L. Li and K. Wang, "Research on automatic recognition method of basketball shooting action based on background subtraction method," *International Journal of Biometrics*, vol. 14, no. 3-4, pp. 318-335, 2022, doi: 10.1504/IJBM.2022.124676.
- [19] R. Dang, C. Liu, M. Liu, and Q. Chen, "Channel attention and multi-scale graph neural networks for skeleton-based action recognition," *AI Communications*, vol. 35, no. 3, pp. 187-205, 2022, doi: 10.3233/AIC-210250.
- [20] H. Liang, "Role of artificial intelligence algorithm for taekwondo teaching effect evaluation model," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 2, pp. 3239-3250, 2021, doi: 10.3233/JIFS-189364.