



2025 International Conference on Intelligent Computing

July 26-29, Ningbo, China

<https://www.ic-icc.cn/2025/index.php>

# A Robust Intelligent Framework for Long Jump Action Scoring: From Pose Estimation to Motion Blur-Resistant Recognition

Zhiliang Qiu<sup>1</sup> [0000-0002-7445-9397], Yanyan Su<sup>1</sup> [0009-0000-9548-384X], Min Lu<sup>1</sup>, Jun Xiang<sup>2</sup>(✉) and Shenglian Lu<sup>1,3</sup>(✉)

<sup>1</sup> Key Lab of Education Blockchain and Intelligent Technology, Ministry of Education, Guangxi Normal University, Guilin 541004, China  
lsl@gxnu.edu.cn

<sup>2</sup> College of Physical Education and Health, Guangxi Normal University, Guilin 541004, China  
xiangjunqk@163.com

<sup>3</sup> Guangxi Key Lab of Multisource Information Mining & Security, College of Computer Science & Engineering, Guangxi Normal University, Guilin 541004, China

**Abstract.** With the advancement of deep learning, sports motion analysis has become increasingly data-driven. However, techniques such as pose estimation, action recognition, and scoring often operate independently. To address this limitation, a unified framework is proposed for structured and objective long jump analysis. One major challenge in real-world scenarios is motion blur, which greatly reduces the accuracy of pose estimation. To mitigate this issue, a long jump dataset was collected from 30 athletes, annotated across four movement phases, multiple lighting conditions, and four levels of motion blur. Based on this dataset, a simple MetaFormer-based model named BaseFormerPose is developed, using uniformly stacked window self-attention. It achieves 91.0 AP on the long jump motion-blur dataset. An automatic scoring module is also introduced, and its outputs show strong agreement with pose-based scores from three expert coaches, suggesting improved consistency and reduced subjectivity in long jump evaluation.

**Keywords:** Human Pose Estimation, Deep Learning, Performance Evaluation, Motion Blur, Automatic Scoring.

## 1 INTRODUCTION

In recent years, motion analysis systems for track and field sports have progressed significantly, evolving from expensive, marker-based laboratory setups to more flexible, portable solutions enabled by computer vision [1, 2]. Traditional systems like Vicon [3] offered high accuracy but required extensive calibration and controlled environments, limiting practical deployment. Later systems such as Dartfish [4] increased accessibility by removing markers, but still relied heavily on manual annotation. With the widespread use of smartphones and consumer-grade cameras, there is a growing demand for automatic analysis without complex equipment or human involvement. This

shift highlights two essential requirements for modern motion analysis: full automation (no manual intervention) and non-invasiveness (no markers or sensors).

Parallel to this hardware evolution, advancements in deep learning have introduced powerful methods for automated and non-invasive sports analysis. Among these, human pose estimation [5-7] has become the cornerstone for extracting motion information from video. It supports higher-level tasks such as action recognition [8], trajectory analysis [9], and scoring by providing robust skeletal keypoints [10]. Numerous models have demonstrated strong performance in controlled environments. However, in real-world high-speed sports scenarios like long jump, motion blur [11, 12]—caused by rapid movement—and inconsistent lighting [13] continue to degrade pose estimation accuracy. Motion blur blurs object edges, making limbs indistinct and keypoints difficult to localize. While several image classification models have introduced adaptive attention mechanisms [14] to combat blur, their application to human pose estimation remains limited. Moreover, most existing datasets [15] are captured under ideal conditions, lacking the diversity in environmental variables needed to train models that generalize well to real-world sports footage.

To address these limitations, we propose a unified intelligent framework for long jump performance analysis, built on two key contributions: (1) a custom long jump dataset designed to simulate real-world blur scenarios, using hang-style long jump as a representative case, and (2) a simple pose estimation model tailored for robustness and scalability. Our dataset includes video sequences from 30 athletes, annotated across four distinct movement phases (approach, takeoff, flight, landing), multiple lighting conditions, and four levels of motion blur. This rich annotation allows for fine-grained training and evaluation of models under various real-world challenges. On this dataset, we develop BaseFormerPose, a simplified multi-stage pose-estimation model based on the MetaFormer. It uses uniformly stacked window-based self-attention to maintain modularity and implementation simplicity, while effectively enhancing robustness to blur without the need for complex attention variants.

To extend the framework beyond pose estimation, a pose-based automatic scoring module is further introduced to evaluate technical execution from keypoint sequences. BaseFormerPose is designed to handle the challenges of motion blur more effectively than commonly used backbones such as ResNet. The scoring module operates on joint trajectories and produces objective scores for performance evaluation. To assess its reliability, pose sequences were manually rated by three professional track and field coaches. The automatic scores exhibited high consistency with the averaged expert ratings, suggesting strong potential for reducing subjectivity and improving fairness in long jump evaluation.

In summary, the main contributions of this work are:

- **Motion Blur-Oriented Dataset Design:** We construct a dedicated hang-style long jump dataset involving 30 athletes, annotated across four movement phases, multiple lighting conditions, and four levels of motion blur. This dataset enables models to learn robust representations under real-world degradation, filling the gap of motion blur scenarios in existing datasets.

- **Simple and Scalable Pose Estimation Model:** We propose BaseFormerPose, a simplified and extensible MetaFormer-based model that utilizes uniformly stacked window-based self-attention. It achieves competitive accuracy with lower complexity, demonstrating effectiveness under blur while maintaining implementation simplicity.
- **Expert-Validated Automatic Scoring Framework:** We develop an automatic scoring module based on predicted pose sequences. Its outputs show high consistency with scores given by three professional coaches, demonstrating the framework's potential to reduce subjectivity and improve fairness in long jump evaluation.

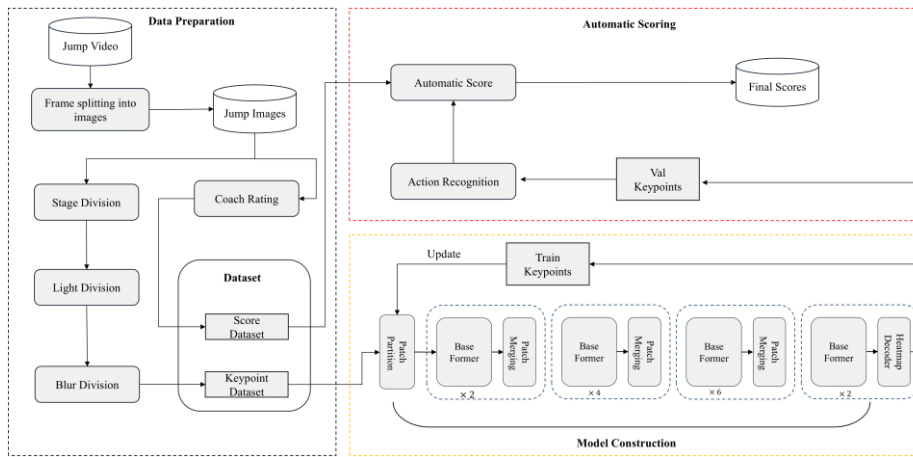


Fig. 1. The flowchart of the proposed 2D human pose estimation and scoring framework.

## 2 Method

The entire experiment process can be divided into two parts: 1) Data Preparation; 2) Model Construction. Initially, long jump videos are captured from athletes and subjected to preprocessing. Individual frames are extracted from the videos, followed by keypoint annotation across selected frames to establish ground truth for pose estimation. Subsequently, the BaseFormerPose model is constructed and trained using the annotated dataset. A detailed overview of the entire process is presented in Fig. 1.

### 2.1 Data preparation

The dataset was collected as part of a university-level sports science course focused on hang-style long jump, where students were trained in the technical execution of standard jumps with emphasis on the swinging posture characteristic of the hang style. Given the natural variability in athlete experience, biomechanics, and execution quality, the dataset includes a wide range of motion sequences—from technically proficient to suboptimal performances. Unlike traditional datasets that focus solely on ideal

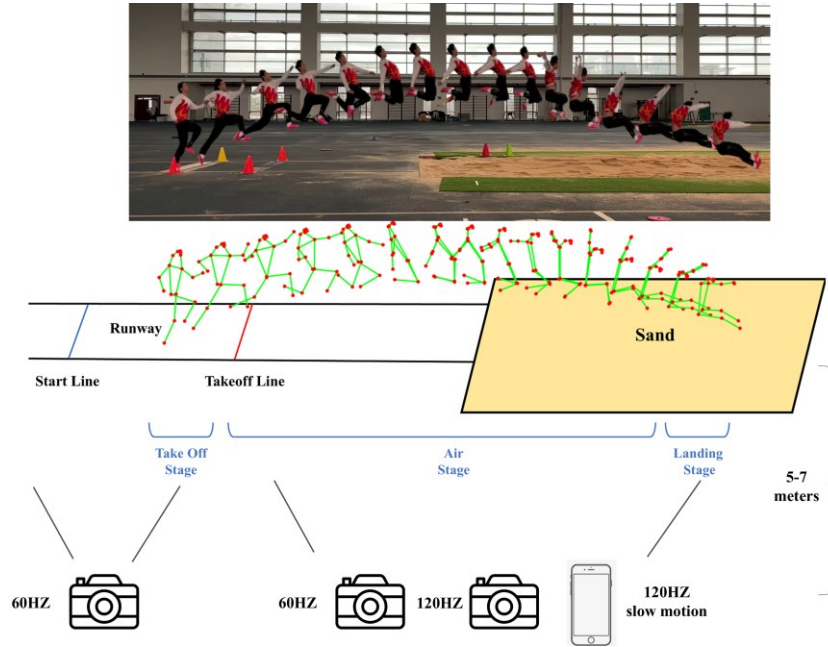
movements, our collection strategy captures all valid jumps, enabling the model to learn from diverse and realistic postural variations.

To simulate motion blur systematically, we introduced two key variables during video capture: lighting conditions and camera frame rate. These factors were adjusted to replicate varying levels of motion blur commonly observed in real-world sports scenarios. The dataset is annotated across four standard long jump phases—approach, takeoff, flight, and landing—with particular emphasis on the latter three stages, which involve intense body deformation and rapid limb motion, making them more susceptible to blur. By training on these challenging sequences, we aim to improve the robustness of pose estimation in dynamic sports contexts and ensure broader applicability in performance analysis.

**Setup and Parameters.** To capture the full complexity of long jump movements, we employed a multi-camera setup consisting of four cameras arranged in two functional groups. The first group included a 60 Hz camera (manual focus, 2.9mm focal length, 120° distortion-free lens, 1920×1080 resolution) placed on the left side of the runway, primarily focused on the approach phase where limb motion is relatively moderate. The second group, positioned on the right side, covered the takeoff, flight, and landing phases—stages characterized by rapid movement and high susceptibility to motion blur. This group comprised two cameras operating at 60 Hz and 120 Hz (with the same specifications as above), along with an iPhone 15 capturing at 120 Hz in slow motion. The horizontal distance from each camera to the runway was maintained between 5–7 meters, providing a wide, distortion-free field of view.

**Recording Conditions and Data Collection.** Data collection spanned four months and reflected three key stages in athlete training: beginning, middle, and end of the course. While all participants had prior experience in standard long jump, the focus was on improving the hang-style swinging posture. Each session included one to three jumps per athlete, with no attire restrictions and scheduled breaks to minimize fatigue. Lighting conditions were controlled across sessions, with the first two conducted in well-lit environments and the third simulating poorly-lit conditions to study lighting-induced motion blur variability. This design enabled the creation of a diverse dataset reflecting real-world motion dynamics.

**Scoring and Evaluation.** Posture scoring was based on a standardized rubric that assessed three specific postural features. During the final evaluation session, three expert coaches independently rated the athletes' movements. In parallel, our automatic scoring module generated scores using the same evaluation criteria. By comparing the predicted scores with the experts' ratings, we assessed the model's consistency and reliability. All raw data were thoroughly annotated prior to training to ensure compatibility with deep learning frameworks and enable structured analysis. An overview of the camera setup and keypoint annotation process is shown in Fig. 2.



**Fig. 2.** Visualization of athletes' keypoint postures and camera equipment arrangements throughout the hang-style long jump.

**Data Processing.** To systematically control and annotate motion blur, we categorized data based on lighting conditions—well-lit (approximately 500–700 Lux) and poorly-lit (100–200 Lux)—as illumination significantly influences blur intensity. Environmental light levels were estimated using a smartphone-based Lux Light Meter. Under each lighting category, motion blur levels were classified independently to reflect actual recording conditions and reduce misclassification caused by heterogeneous image characteristics.

**Table 1.** Blur Level Classification Definitions and Scoring Ranges (n/a indicates not applicable).

Blur Level	Scoring Range (BRISQUE)	Manual Annotation Standard
	Well-lit / Poor-lit	
Clear	0 - 25 / n/a	Sharp details, well-defined edges
Slightly Blurred	26 - 40 / 26 - 50	Slightly blurred details, discernible edges
Blurred	41 - 70 / 50 - 70	Noticeably blurred edges, loss of details
Highly Blurred	n/a / 70 - 100	Edges indistinguishable, merged with background

To ensure accurate blur grading, we applied a hybrid approach combining no-reference image quality assessment (IQA) and manual validation. Specifically, the BRISQUE algorithm was used to quantify spatial quality by analyzing natural scene

statistics, and score thresholds were defined to classify blur levels. Manual inspection by experts followed to correct borderline cases. Under well-lit conditions, three blur levels were defined: clear, slightly blurred, and blurred; under poorly-lit conditions, the categories were slightly blurred, blurred, and highly blurred. Table 1 summarizes the corresponding scoring ranges and definitions.

Following exclusion of unusable videos (e.g., incomplete jumps, occlusions, camera errors), a total of 614 sequences were retained. Frames were extracted at fixed intervals, and a cleaning process removed redundant or excessively degraded images. The final dataset comprises 17,113 images, each manually verified for quality. All images were labeled with their respective long jump phase: approach, takeoff, flight, or landing. As shown in Table 2 and Figure 3, the landing phase accounts for 45% of the dataset due to its duration and movement complexity, while the takeoff phase, though shorter, still provides 936 useful images. The dataset also reflects a balanced distribution across lighting and blur categories, ensuring robustness and diversity for downstream learning.

**Table 2.** The number of images in the dataset under different conditions.

Stage / Condition		Approach	Take-Off	Flight	Landing	All
Well-lit	clear	0	31	12	3496	3539
	slightly blurred	607	286	1431	1103	3427
	blurred	1570	221	1064	662	3517
Poorly-lit	slightly blurred	76	31	15	1715	1837
	blurred	1311	284	528	706	2829
	highly blurred	785	83	1074	22	1964
All		4349	936	4124	7704	17113

**Training Details.** After processing, the dataset was randomly divided into five groups based on lighting, blur level, and stage. One group was used for validation, and the remaining four for training. The model was trained on the training set and evaluated on the validation set. The exclusion of the validation set from training ensured a fair evaluation. The process followed standard human pose estimation training protocols. Standard training, evaluation settings, and data augmentation strategies from the MMPose [16] framework were followed. Data augmentation techniques, including affine transformations, random cropping, and random masking, were applied during training. These strategies ensured dataset diversity and enhanced model robustness.



Fig. 3. Proportion of stages and lighting conditions in the dataset.

## 2.2 Model Construction

**BaseFormerPose Architecture.** BaseFormerPose is a simple yet effective human pose estimation model designed to enhance robustness under motion-blurred conditions. It is built upon the generalized MetaFormer framework, which abstracts a neural network into two functional components: a token mixer for spatial interaction and a channel MLP for inter-channel information processing. This design paradigm provides architectural flexibility and allows for efficient adaptation to task-specific needs while maintaining low coupling between modules.

Specifically, BaseFormerPose employs window-based self-attention as the token mixer, following the MetaFormer layout (Fig. 4). This approach restricts attention to non-overlapping local windows, reducing computation while retaining spatial modeling capability. Each block consists of normalization, window attention with residual connection, followed by another normalization and a lightweight channel MLP. Compared to convolutional layers with fixed receptive fields, window attention enables adaptive, content-aware feature aggregation. This allows the model to more effectively capture local pose-related structures under motion blur without increasing architectural complexity.



**Fig. 4.** Overview of the BaseFormerPose (left) and structure of the BAFormer module (right).

Formally, for an input feature  $X \in \mathbb{R}^{N \times D}$ , where  $N = H \times W$  and  $D$  is the embedding dimension, the computations are as follows:

$$X' = \text{WindowAttention}(\text{Norm}(X)) + X \quad (1)$$

$$Y = \text{ChannelMLP}(\text{Norm}(X')) + X' \quad (2)$$

WindowAttention computes attention within each window  $N_w$ , aggregating local features as:

$$\text{WindowAttention}(X) = \sum_{j=1}^{N_w} \text{Attention}(Q_j, K_j, V_j) \quad (3)$$

Here,  $Q_j$ ,  $K_j$ ,  $V_j$  are the query, key, and value projections of the input tokens in each window. This localized attention mechanism helps retain structural consistency while reducing the effect of irrelevant background noise—especially crucial under motion blur conditions.

To predict final keypoint locations, we adopt a heatmap regression approach. The output feature is passed through a deconvolutional head to produce  $H \in \mathbb{R}^{K \times H' \times W'}$ , where  $K$  is the number of keypoints. Each channel in  $H$  represents the probability map of a specific keypoint, and the prediction is optimized using mean squared error between predicted and ground-truth heatmaps:

$$L_{\text{heatmap}} = \frac{1}{KH_h W_h} \sum_{k=1}^K \sum_{h=1}^{H_h} \sum_{w=1}^{W_h} (Y_{\text{pred}}^{k,h,w} - Y_{\text{gt}}^{k,h,w})^2 \quad (4)$$

Through this architecture, BaseFormerPose offers an elegant balance between computational efficiency and representation power. Its modular design ensures scalability across different deployment environments and its performance under motion blur is empirically validated in our experiments.



**Action Recognition.** Once BaseFormerPose is constructed and trained, keypoints are extracted from the input images and used for downstream action recognition. According to the findings of Jayaneththi and Chandana [17], the performance of long jump athletes is strongly influenced by air-phase posture, which encompasses a sequence of standardized movement patterns. In this study, we identify three representative postures characteristic of the hang-style long jump: takeoff, mid-air hip extension, and hip flexion with knee tuck near landing. These postures serve as critical indicators for evaluating technical execution. The definitions and associated joint angle ranges for each posture are detailed in Table 3.

**Table 3.** Define score-standard posture and evaluation indicators.

Evaluation Posture	Graphical	Posture description	Evaluation
Take-off		Explosively push off with the take-off leg to convert horizontal velocity into vertical motion. Maintain a controlled leg swing to stabilize posture and prepare for landing.	$170^\circ \leq \theta_1 \leq 180^\circ$ $60^\circ \leq \theta_3 \leq 90^\circ$
Hip Extension		Lift hips and extend body in the air during the jump. Raise arms while swinging legs downward.	$130^\circ \leq \theta_1 \leq 160^\circ$ $0^\circ \leq \theta_2 \leq 30^\circ$ $0^\circ \leq \theta_3 \leq 30^\circ$ $35^\circ \leq \theta_1 \leq 45^\circ$
Hip Flexion and Tuck		Lift legs up and lean torso forward as far as calves go	$160^\circ \leq \theta_2 \leq 180^\circ$ $160^\circ \leq \theta_3 \leq 180^\circ$

$\theta_1, \theta_2, \theta_3$  represents the characteristic angles of a posture, which determine a posture. For each athlete's motion data sequence, we calculate the similarity between the three feature angle vectors of the tested posture and the defined posture using a weighted Euclidean distance, denoted as  $d$ . A smaller  $d$  indicates a higher similarity between the tested and defined postures.  $\omega$  represents the weight of this feature angle vector in the entire  $d$ . The posture with the smallest  $d$  is considered the successfully matched posture, as shown in the formula:

$$d = \sqrt{\sum_{i=1}^n \omega_i (\theta_i - \theta'_i)^2} \quad (5)$$

**Automatic Scoring.** After calculating the Euclidean distances between the three tested poses and the defined pose, we apply a linear function to convert these distances into scores:

$$S = k \times d + c \quad (6)$$

$S$  represents the final score, while  $k$  and  $c$  are the quantitative values derived from the data. Given  $n$  training samples, we use the least squares method to find the optimal values for  $k$  and  $c$  by minimizing the sum of the squared differences between the fitted score  $S$  and the coach's actual score  $S'$ :

$$\min_{k,c} \sum_{i=1}^n (S_i - S'_i)^2 \quad (7)$$

By substituting the formula and taking the partial derivatives, the optimal values of  $k$  and  $c$  can be obtained. It is worth noting that as the sample size increases, the values of  $k$  and  $c$  converge toward the most accurate estimates:

$$k = \frac{n \sum_{i=1}^n d_i S'_i - \sum_{i=1}^n d_i \sum_{i=1}^n S'_i}{n \sum_{i=1}^n d_i^2 - (\sum_{i=1}^n d_i)^2} \quad (8)$$

$$c = \frac{n \sum_{i=1}^n S'_i - k \sum_{i=1}^n d_i}{n} \quad (9)$$

### 2.3 Evaluation Metrics

**Human Pose Estimation Evaluation.** We evaluate pose estimation performance using the standard mean Average Precision (mAP) and mean Average Recall (mAR) metrics on the COCO dataset [18]. These metrics are computed based on Object Keypoint Similarity (OKS):

$$OKS = \frac{\sum_i \exp(-\hat{d}_i^2 / 2s^2 k_i^2) \sigma(v_i > 0)}{\sum_i \sigma(v_i > 0)} \quad (10)$$

where  $\hat{d}_i$  is the Euclidean distance between the  $i$ -th predicted keypoint and its true position,  $V_i$  is the visibility flag of the keypoint,  $S$  is the object scale, and  $K_i$  is a keypoint-specific constant. Two widely used thresholds are mAP@0.5 and mAP@0.75. mAP@0.5 allows moderate distance between predicted and ground truth keypoints, while mAP@0.75 requires closer predictions for accuracy. mAP (M) and mAP (L) assess model precision on medium and large objects based on object area.

Similarly, mAR (mean Average Recall) is used to assess the model's recall capabilities. mAR@0.5 and mAR@0.75 measure recall at OKS thresholds of 0.5 and 0.75,

respectively. mAR (M) and mAR (L) focus on recall for medium and large objects. Recall is calculated as:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (11)$$

Where *True Positives* are correctly detected keypoints, and *False Negatives* are keypoints present in the ground truth but missed by the model.

**Posture Scoring Evaluation.** Three key long jump postures were extracted from the athletes, and three sports experts were invited to manually score each posture for every individual. The scoring criteria for both expert evaluation and automatic scoring followed a predefined quantitative rubric, which standardizes the assessment based on three key postures. However, while the automatic scoring system strictly calculates scores based on joint angles, human experts, though still using angle-based assessment, intentionally retain a degree of subjective judgment to account for real-world variations in execution quality.

A total of 614 video sequences, each representing a unique long jump trial, were collected. From each trial, we extracted three video sub-sequences corresponding to the three key postures and selected the most representative frame for each posture. Among the 614 sequences (1842 images), 600 were used for model training and refinement, while the remaining 14 sequences were reserved for testing and evaluating the automatic scoring system. The closer the automatic score is to the average expert score, the better the system's performance.

### 3 Experiments

#### 3.1 Results of Human Pose Estimation

**Table 4.** Comparison on our MotionBlur validation dataset. All data is trained through the mmpose framework [16].

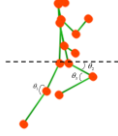



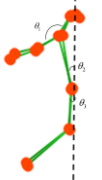







Method	MotionBlur Validation Dataset↑							
	AP	AP <sup>50</sup>	AP <sup>75</sup>	AP <sup>L</sup>	AR	AR <sup>50</sup>	AR <sup>75</sup>	AR <sup>L</sup>
ResNet-50 [19]	85.9	97.6	91.0	90.2	89.8	98.7	93.8	92.8
HRNet-W32 [5]	90.4	98.6	94.0	93.8	93.2	99.2	95.8	95.6
DARK-Res50 [7]	90.2	<b>98.7</b>	94.0	93.8	92.8	99.1	95.6	95.3
SimCC [6]	91.0	<b>98.7</b>	95.0	<b>94.4</b>	<b>94.0</b>	99.2	<b>96.5</b>	<b>96.2</b>
SwinTransformer [20]	90.6	98.6	93.9	94.1	93.2	99.2	95.8	95.7
PVT [21]	90.2	98.6	93.8	93.9	93.0	99.1	95.7	95.5
BaseFormerPose (Ours)	<b>91.0</b>	<b>98.7</b>	<b>95.6</b>	91.2	93.0	<b>99.5</b>	<b>96.5</b>	93.1

Under the MotionBlur validation dataset, BaseFormerPose outperforms all compared methods in AP, AP<sup>75</sup>, and AR<sup>75</sup>, achieving 91.0, 95.6, and 96.5 respectively, indicating superior keypoint localization accuracy under stricter thresholds. It also achieves competitive results in AP<sup>50</sup>, AR, and AR<sup>50</sup>. However, on AP<sup>L</sup> and AR<sup>L</sup>, BaseFormerPose

slightly lags behind DARK-Res50, likely due to the deeper convolutional structure of DARK-Res50 providing stronger hierarchical feature extraction on large-scale key-point regions. Overall, BaseFormerPose achieves the best balance between performance and model simplicity.

### 3.2 Results of Action Recognition and Automatic Scoring

**Table 5.** Comparison of evaluation outcomes based on three randomly sampled datasets extracted from the complete results for analysis.

Key Pos	Standard Pos	Number	Target Pos	E1	E2	E3	Score	Score
Take-off		1		80	80	72	77.3	<b>78</b>
		2		85	82	75	80.7	<b>80</b>
		3		86	84	78	82.7	<b>81</b>
Hip Extension		1		81	78	70	76.3	<b>74</b>
		2		77	70	74	73.7	<b>75</b>
		3		78	76	79	77.7	<b>78</b>
Hip Flexion-and Tuck		1		75	74	70	73	<b>72</b>
		2		65	67	69	67	<b>71</b>
		3		80	76	74	76.7	<b>76</b>

To ensure the validity and consistency of expert scoring, Pearson Correlation Coefficients (PCCs) were used to assess score alignment among three experts across 614

trials, each covering three key postures: take-off, hip extension, and hip flexion with tuck. PCCs were preferred over Intraclass Correlation Coefficients (ICCs) since each trial represents an independent movement with unique biomechanical characteristics. Unlike ICCs, which assume repeated measurements of the same target, PCCs are better suited to evaluating trend agreement across non-overlapping samples. The average PCCs were 0.88 (E1–E2), 0.87 (E1–E3), and 0.86 (E2–E3), indicating strong inter-rater consistency and supporting the reliability of the manual evaluation process.

Despite the high overall correlation, noticeable score variations persist at the trial level, reflecting the subjectivity of human judgment. As shown in Table 5, scores assigned to the same posture in a given trial often differ across experts, even when standardized visual references are provided. For example, in the take-off posture of Trial 3, expert scores range from 78 to 86, suggesting different interpretations of movement quality. In contrast, the automatic scoring module applies a fixed evaluation rubric based on joint-angle deviations from reference poses, enabling consistent and objective assessments. While small differences may arise between predicted and expert-averaged scores, the model maintains stable performance across all postures. Trained on 600 annotated trials and capable of continuous improvement with more data, the system shows promise as a reliable component of a scalable, data-driven framework for standardized long jump evaluation.

## 4 Conclusion

This work presents a unified framework for long jump performance analysis under motion-blurred conditions, integrating pose estimation and automatic scoring within a structured pipeline. A domain-specific dataset was constructed to capture key phases of the long jump under diverse lighting and motion blur levels, supporting robust model training. The proposed BaseFormerPose, employing uniformly stacked window self-attention, demonstrates strong performance in keypoint detection with 91.0% AP on the motion blur subset. In addition, the automatic scoring module shows high alignment with expert evaluations, reinforcing its reliability for posture assessment. Together, these components offer a scalable, objective, and data-driven solution to assist coaches in evaluating athletic performance in real-world track and field environments.

**Acknowledgements.** This work was funded by the Key Lab of Education Blockchain and Intelligent Technology, Ministry of Education (Grant No. EBME24-04).

## References

1. Suo, X., Tang, W., Li, Z.: Motion capture technology in sports scenarios: a survey. *Sensors* 24, 2947 (2024)
2. Hellsten, T., Karlsson, J., Shamsuzzaman, M., Pulkkinen, G.: The potential of computer vision-based marker-less human motion analysis for rehabilitation. *Rehabilitation Process and Outcome* 10, 11795727211022330 (2021)

3. Merriaux, P., Dupuis, Y., Bouteau, R., Vasseur, P., Savatier, X.: A study of vicon system positioning performance. *Sensors* 17, 1591 (2017)
4. Kassay, A.D., Daher, B., Lalone, E.: An analysis of wrist and forearm range of motion using the Dartfish motion analysis system. *Journal of Hand Therapy* 34, 604-611 (2021)
5. Sun, K., Xiao, B., Liu, D., Wang, J.: Deep high-resolution representation learning for human pose estimation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5693-5703. (2019)
6. Li, Y., Yang, S., Liu, P., Zhang, S., Wang, Y., Wang, Z., Yang, W., Xia, S.-T.: Simcc: A simple coordinate classification perspective for human pose estimation. In: *European Conference on Computer Vision*, pp. 89-106. Springer, (2022)
7. Zhang, F., Zhu, X., Dai, H., Ye, M., Zhu, C.: Distribution-aware coordinate representation for human pose estimation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7093-7102. (2020)
8. Duan, H., Zhao, Y., Chen, K., Lin, D., Dai, B.: Revisiting skeleton-based action recognition. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2969-2978. (2022)
9. Sepas-Moghaddam, A., Etemad, A.: Deep gait recognition: A survey. *IEEE transactions on pattern analysis and machine intelligence* 45, 264-284 (2022)
10. Yan, S., Xiong, Y., Lin, D.: Spatial temporal graph convolutional networks for skeleton-based action recognition. In: *Proceedings of the AAAI conference on artificial intelligence*. (2018)
11. Suo, X., Tang, W., Mao, L., Li, Z.: Digital human and embodied intelligence for sports science: advancements, opportunities and prospects. *The Visual Computer* 1-17 (2024)
12. Bright, J., Chen, Y., Zelek, J.: Mitigating motion blur for robust 3d baseball player pose modeling for pitch analysis. In: *Proceedings of the 6th International Workshop on Multimedia Content Analysis in Sports*, pp. 63-71. (2023)
13. Samkari, E., Arif, M., Alghamdi, M., Al Ghamdi, M.A.: Human pose estimation using deep learning: a systematic literature review. *Machine Learning and Knowledge Extraction* 5, 1612-1659 (2023)
14. Chen, Z., Zhu, Y., Zhao, C., Hu, G., Zeng, W., Wang, J., Tang, M.: Dpt: Deformable patch-based transformer for visual recognition. In: *Proceedings of the 29th ACM international conference on multimedia*, pp. 2899-2907. (2021)
15. Gan, Q., El-Yacoubi, M.A., Fenaux, E., Cl  men  on, S.: Human Pose Estimation Based Biomechanical Feature Extraction for Long Jumps. In: *2024 16th International Conference on Human System Interaction (HSI)*, pp. 1-6. IEEE, (2024)
16. MMPose Contributors: Openmmlab pose estimation toolbox and benchmark, <https://github.com/open-mmlab/mmpose>
17. Sachini Jayaneththi, J.P., Suraj Chandana, A.W.: The Effect of Movement Pattern in Flight Phase for Long Jump Performance. *European Journal of Science, Innovation and Technology* 2, 98-104 (2022)
18. Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Doll  r, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: *Computer Vision  ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V* 13, pp. 740-755. Springer, (2014)
19. Xiao, B., Wu, H., Wei, Y.: Simple baselines for human pose estimation and tracking. In: *Proceedings of the European conference on computer vision (ECCV)*, pp. 466-481. (2018)
20. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10012-10022. (2021)



*2025 International Conference on Intelligent Computing*

*July 26-29, Ningbo, China*

<https://www.ic-icc.cn/2025/index.php>

21. Wang, W., Xie, E., Li, X., Fan, D.-P., Song, K., Liang, D., Lu, T., Luo, P., Shao, L.: Pvt v2: Improved baselines with pyramid vision transformer. *Computational Visual Media* 8, 415-424 (2022)