Design of Juvenile Chain Boxing Scoring System Based on Deep Learning

Mingxuan Li[1](✉), Feng Tian[2], Tianfeng Lu[1], and Shuting Ni[1]

[1] Shanghai Film Academy, Shanghai University, Shanghai, China
halfspring116@163.com, {22723213,22723206}@shu.edu.cn
[2] Shanghai Film Special Effects Engineering Technology Research Center, Shanghai, China

Abstract. Computer vision technology has significant implications for advancing martial arts education. The Ministry of Education aims to integrate martial artsinto the campus curriculum evaluation system as part of promoting martial arts culture. In 2016, Shanghai added martial arts as a subject to the senior high school entrance examination for the first time. The system introduced in this paper is designed to provide real-time feedback and objective assessment of martial arts movements, particularly in the context of the senior high school entrance examination in Shanghai. Unlike traditional exam scoring methods, this system utilizes computer vision technology, including human body pose estimation and action recognition, and is based on the Transformer architecture. The system provides accurate posture matching scores, feedback, and a guided juvenile chain boxing teaching system, which can improve learning efficiency and reduce assessment costs. The aim is to promote fair and objective sports scoring in the Entrance Examination and aid the Shanghai Wushu Sports Entrance Examination. The system has also been deployed in a client software for testing purposes.

Keywords: Martial arts scoring · action quality evaluation

1 Introduction

Wushu is deeply rooted in Chinese culture, representing the spiritual pursuit and cultural characteristics of the Chinese people. Martial arts education plays a crucial role in inheriting and promoting this culture, further developing traditional Chinese culture. Juvenile Chain Boxing, which features fluid and graceful movements, is easy to learn, and can be performed at varying speeds according to the student's proficiency level [1]. It is particularly suitable for martial arts beginners, and its inclusion in the high school entrance examination project provides young people with more opportunities to learn and experience traditional martial arts culture.

Traditional scoring methods rely heavily on subjective judgment, leading to significant subjective influence on the final score due to referees' need to maintain strict adherence to rules. However, with the advent of artificial intelligence (AI) technology, AI + traditional scoring methods have become an integral part of modern sports, offering benefits such as improved fairness, elimination of subjective factors, and reduced scoring

C. F. Peng et al. (Eds.): EIMT 2023, AHSSEH 8, pp. 842–847, 2023.
https://doi.org/10.2991/978-94-6463-192-0_109

errors and deviations. This article aims to apply this model to the juvenile serial boxing scoring system by collecting motion data through a front camera and analyzing it using human pose estimation and regression algorithms. By utilizing this system, users can receive comprehensive scoring and real-time recognition feedback, promoting learning efficiency and enhancing the objectivity and fairness of raters. Additionally, testers can obtain effective practice through feedback, while raters can gather quantitative data for auxiliary scoring, improving the overall scoring process.

2 Related Work

2.1 Action Quality Assessment (AQA)

Sports video analysis has gained significant attention due to its practical applications in healthcare [2], sports [3], and video retrieval [4]. Although available action scoring datasets are limited, Wnuk K et al. [5] provides a new diving dataset that poses three key challenges: tracking, classifying, and judging action quality, which lays a foundation for future research. Pirsiavash H et al. [6] proposes a learning-based framework for evaluating diving and figure skating performance by training a regression model from spatio-temporal pose features to expert referee scores. Parmar P et al. [3] introduces three frameworks for evaluating Olympic sports using spatio-temporal features learned by 3D convolutional neural networks, and LSTM plus SVR combination modules for score regression. Xu C et al. [7] proposes a deep architecture with two complementary components, self-attention LSTM and Multi-scale Convolutional Skip LSTM, to effectively learn local and global sequence information in each video. Furthermore, the FisV dataset, a large-scale figure skating sports video dataset, consisting of 500 videos annotated by two scores from 9 different referees, provides inspiration for multi-score evaluation in future scoring research.

2.2 Pose Feature Extraction Method

In computer vision, pose feature extraction involves extracting human body pose information from various types of data, such as images and videos, and has a wide range of applications, including human action recognition [8], human pose tracking [9], and human-computer interaction [10]. There are several commonly used methods for extracting pose features, including joint position-based [11], image descriptor-based [5], and deep learning-based methods [7, 13]. Deep learning-based methods, especially those with spatiotemporal pose extraction modules [12] and view-invariant trajectory descriptors, have significantly improved the accuracy and stability of pose estimation. Moreover, combining different methods, such as the JCA and ADA blocks proposed by Nekoui M et al. [14], can lead to better pose estimation results through fusion.

3 System Design and Function Realization

3.1 Basics of System Development

This system has two main components: real-time recognition of action categories and intelligent scoring. Real-time recognition is achieved using the mmaction2 algorithm. Intelligent scoring relies on gesture feature extraction and a score data regression model.

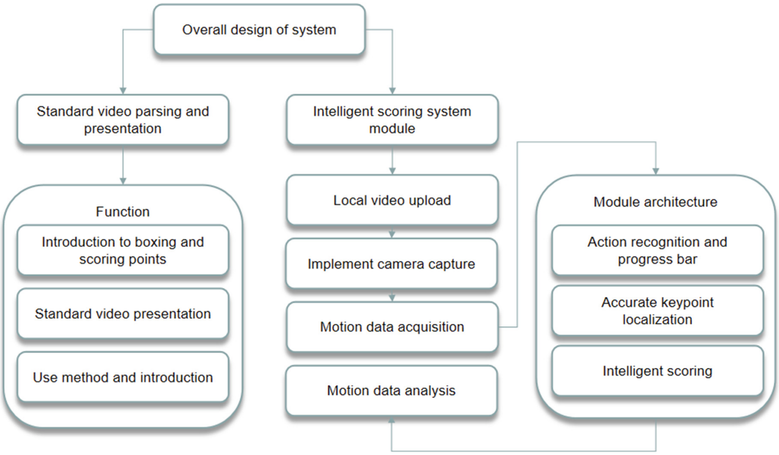
Overall design of system
Standard video parsing and presentation
Intelligent scoring system module
Function
Introduction to boxing and scoring points
Standard video presentation
Use method and introduction
Local video upload
Implement camera capture
Motion data acquisition
Motion data analysis
Module architecture
Action recognition and progress bar
Accurate keypoint localization
Intelligent scoring

Fig. 1. System overall function design drawing

To support this, we created a juvenile serial boxing scoring dataset, which includes video footage and four scores: Standard score (SS), Fluency score (FS), Attitude score (AS), and Integrity score (IS). These scores measure technical standard, movement coherence, unique mental state, and completeness of the serial boxing movements. The dataset was created with guidance from a founder of Juvenile Lianhuanquan and a professional martial arts teacher.

3.2 Overall System Design

Figure 1 shows the overall functional design of the system.

3.3 Design of Juvenile Serial Boxing Scoring System

The flow chart for the intelligent scoring module is presented in Fig. 2.

3.4 Design of Action Scoring System

The scoring system captures sequential frames from a camera or local video, extracts key point coordinates of the human body, and sends them back to the client for analysis of their matching degree with the scoring model, resulting in a score. This scoring method is based on the 3D pose of the human body and the features of the score model table trained in the preparation phase. The process of designing the scoring model is illustrated in Fig. 3.

The performance evaluation process for assessing martial arts performance is based on four aspects: movement standard, fluency, spirit, and integrity. The VideoPose3D algorithm, is first used to extract the pose sequence from the test video. The pose features are then encoded and sent to the space-time Transformer codec architecture.

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

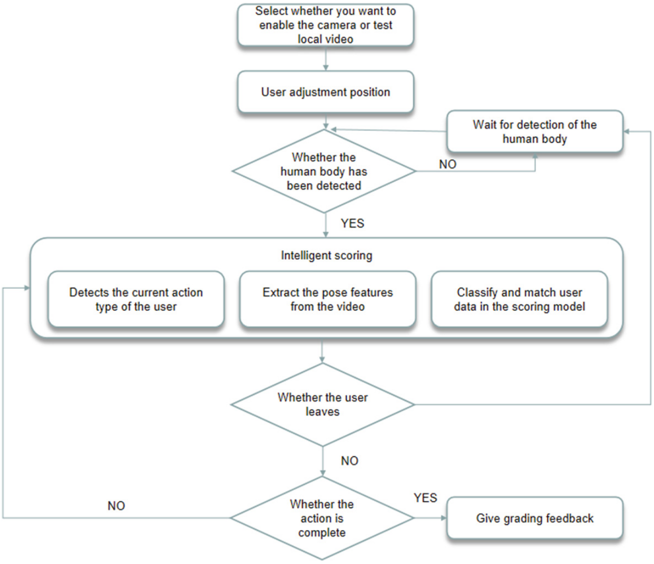
Select whether you want to enable the camera or test local video
User adjustment position
Wait for detection of the human body
Whether the human body has been detected
NO
YES
Intelligent scoring
Detects the current action type of the user
Extract the pose features from the video
Classify and match user data in the scoring model
Whether the user leaves
NO
NO
Whether the action is complete
YES
Give grading feedback

Fig. 2. The flow chart of Boat fist training part system design.

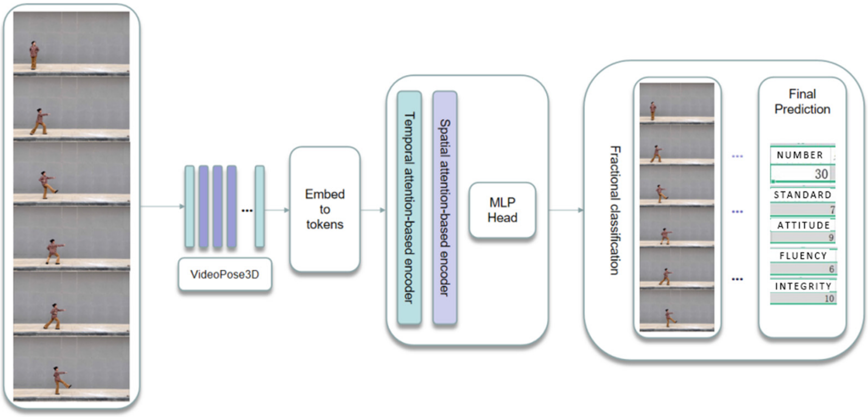

Fig. 3. System Process Architecture Design.

where, Q is the query vector, K is the key vector, V is the value vector, d_k is the dimension of the key vector, and the softmax function is used to calculate the attention weight. The attention function multiplys the query vector with the key vector, divides it by $\sqrt{d_k}$ to get the scaled similarity, and finally gets the attention weight through the softmax function. Then multiply the attention weight with the value vector to get the context feature vector. This architecture can simultaneously capture temporal continuity correlation.

We extract the pose feature sequence extracted from the video, expressed as $x \in \mathbb{R}^{17}$, and we have the corresponding score data $y \in \mathbb{R}$. We need to build a regression model that maps the input attitude characteristics to the output score, i.e.

$$f(x) = wx + b$$

In this equation, $w \in \mathbb{R}^{1\times 17}$ represents the weight matrix, and $b \in \mathbb{R}$ is the bias term. The model parameters w and b are determined the mean squared error (MSE) between the predicted values and the true values based on the provided training dataset (x_i, y_i).

$$\underset{w,b}{min}\frac{1}{n}\sum_{i=1}^{n}(f(x_i) - y_i)^2$$

where n is the size of the training dataset. The score test dataset comprises 323 data points, which are divided into 226 training sets, 65 validation sets, and 32 test sets. The video pose feature sequences and corresponding scores are trained to form a mapping relationship, which results in better prediction performance.

4 Conclusion

The Juvenile Lianhuanquan teaching system aims to provide a fair and quantifiable evaluation system for assessing the quality of martial arts movements. In its development, we constructed a dataset for juvenile Lianhuanquan scoring and boxing types, and designed an action quality evaluation model. The system has facilitated the advancement of martial arts quality scoring through its in-depth learning and utilization. However, further improvements are required in terms of functionality, interfaces, and data collection. Moreover, individual actions can still be improved. The action quality evaluation technology used in this system can also be applied in fields such as medical rehabilitation, fitness, human-computer interaction, work safety, and ergonomics. Thus, the research and development of the Juvenile Lianhuanquan scoring system has vast application prospects.

References

1. The Dilemma and Path of Including Wushu in the Middle School Physical Education Entrance Examination" by Zheng Haijuan, published in Shandong Sports Science and Technology in 2021, volume 43, issue 4, pages 62–66.
2. Panesar A. Machine learning and AI for healthcare[M]. Coventry, UK: Apress, 2019.Van Derlofske, J. F., “Computer modeling of LED light pipe systems for uniform display illumination,” Proc. SPIE 4445, 119–129 (2001).
3. Parmar P, Tran Morris B. Learning to score olympic events[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017: 20–28.
4. Dong Z, Wei J, Chen X, et al. Face detection in security monitoring based on artificial intelligence video retrieval technology[J]. IEEE Access, 2020, 8: 63421-63433.
5. Wnuk K, Soatto S. Analyzing Diving: A Dataset for Judging Action Quality[C]//ACCV Workshops (1). 2010: 266-276
6. Pirsiavash H, Vondrick C, Torralba A. Assessing the quality of actions[C]//Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VI 13. Springer International Publishing, 2014: 556-571

7. Xu C, Fu Y, Zhang B, et al. Learning to score figure skating sport videos[J]. IEEE transactions on circuits and systems for video technology, 2019, 30(12): 4578-4590
8. Yan S, Xiong Y, Lin D. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]//Proceedings of the AAAI conference on artificial intelligence. 2018, 32(1).
9. Xiu Y, Li J, Wang H, et al. Pose Flow: Efficient online pose tracking[J]. arXiv preprint arXiv: 1802.00977, 2018.
10. 11. Liu H, Liu T, Zhang Z, et al. ARHPE: Asymmetric relation-aware representation learning for head pose estimation in industrial human–computer interaction[J]. IEEE Transactions on Industrial Informatics, 2022, 18(10): 7107-7117.
11. 12. Pucci D, Becattini F, Del Bimbo A. Joint-Based Action Progress Prediction[J]. Sensors, 2023, 23(1): 520.
12. 13. Li H, Lei Q, Zhang H, et al. Skeleton-based deep pose feature learning for action quality assessment on figure skating videos[J]. Journal of Visual Communication and Image Representation, 2022, 89: 103625.
13. 14. Sardari F, Paiement A, Hannuna S, et al. Vi-net—view-invariant quality of human movement assessment[J]. Sensors, 2020, 20(18): 5258.
14. Nekoui M, Cruz F O T, Cheng L. EAGLE-Eye: Extreme-Pose Action Grader Using Detail Bird's-Eye View[C]//Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2021: 394–402.
15. Pavllo D, Feichtenhofer C, Grangier D, et al. 3d human pose estimation in video with temporal convolutions and semi-supervised training[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 7753–7762.