Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 2839244, 13 pages
https://doi.org/10.1155/2022/2839244

Research Article

Automatic Detection Algorithm of Football Events in Videos

Yunke Jia

School of Physical Education, Liaocheng University, Liaocheng 252000, China

Correspondence should be addressed to Yunke Jia; 1910050103@stu.lcu.edu.cn

Received 8 March 2022; Revised 6 April 2022; Accepted 20 April 2022; Published 14 May 2022

Academic Editor: Vijay Kumar

The purpose is to effectively solve the problems of high time cost, low detection accuracy, and difficult standard training samples in video processing. Based on previous investigations, football game videos are taken as research objects, and their shots are segmented to extract the keyframes. The football game videos are divided into different semantic shots using the semantic annotation method. The key events and data in the football videos are analyzed and processed using a combination of artificial rules and a genetic algorithm. Finally, the performance of the proposed model is evaluated and analyzed by using concrete example videos as data sets. Results demonstrate that adding simple artificial rules based on the classic semantic annotation algorithms can save a lot of time and costs while ensuring accuracy. The target events can be extracted and located initially using a unique lens. The model constructed by the genetic algorithm can provide higher accuracy when the training samples are insufficient. The recall and precision of events using the text detection method can reach 96.62% and 98.81%, respectively. Therefore, the proposed model has high video recognition accuracy, which can provide certain research ideas and practical experience for extracting and processing affective information in subsequent videos.

1. Introduction

With the rapid development of the economy, society, and the Internet, data appearing on the Internet keep increasing, and their types are various. People's demand for network data is also growing [1]. As a comprehensive expression form, including text, image, and audio, video, allows users to get comprehensive information, which has become the most significant information type in data processing today [2]. As the threshold for video shooting and uploading on major video websites has been lowered, the number of videos on the Internet has increased dramatically. Video information has become an indispensable part of people's lives. However, as the amount of information available to people increases, it is difficult for many users to extract useful information because there is too much information in the videos, which reduces the user experience [3]. Sports game videos provide vital video information and have a very large audience. Besides, the industry from which sports game videos can be extended also has substantial commercial values [4]. Football is one of the users' favorite sports videos, and extracting useful information from football game videos has attracted much attention. The analysis and retrieval of football game videos aim to analyze and research various football game videos, establish a bridge between low-level semantics and high-level semantics, and ultimately meet the needs of users [5]. However, the current detection of football videos is often limited by problems such as complicated background and low accuracy [6]. Therefore, how to obtain information that users are interested in from loads of video information data to meet the different needs of different users has become a scientific problem that needs to be solved urgently.

At present, most research on automatic detection algorithms for game videos focuses on using different algorithm combinations to improve the accuracy of the model. Chambers et al. (2019) used a random forest classification model to develop a specially designed algorithm to detect tackles and tackle events. They also used eight other international competition data sets and video footage to verify this algorithm. It turned out that the algorithm based on video detection could correctly identify all tackles and tackle events in the games, and the detection accuracy rate could remain at 79% [7]. Daudpota et al. (2019) built an automatic video detection model using the rule-based multimodal

classification method and the shots and scenes in the video as classifiers. By detecting 600 selected videos with a duration of more than 600 hours, they found that the accuracy and recall of this model were 98% and 100% [8].

Here, the research object is football videos because compared with other sports games, football games have a more considerable amount of data, which is conducive to data analysis. Second, football videos have a wider audience group and sparse content, which is conducive to data processing. The purpose here is to find, cut, and extract various events that the audiences are interested in from the lengthy football games. The research approaches here include reasonable segmentation of shots, research and analysis of shots and semantic annotation, and extraction of the shot sequences that may be the target events using artificial rules by analyzing the rules of video shooting. The machine learning algorithm is employed to build a model. The model identifies the shot sequences of the suspected target events, thereby accomplishing high-precision extraction of useful information in the videos.

The innovative points include (i) from the perspective of camera labeling, artificial rules are utilized to determine the position of key events in the videos. (ii) Utilizing the improved HMM model and adding the genetic algorithm to achieve high-precision extraction of key events with comparatively few training samples.

There are five sections in total. The first section is the introduction, which introduces the problems encountered in extracting useful information from videos, where the research objects and research foundation are determined. The second section is the Literature Review, which analyzes and summarizes the research on video analysis and detection algorithms of football semantic events. The third section introduces the research method, which clarifies the models that need to be built, parameter settings, sample data, and performance testing methods. The fourth section is the Results and Discussions, which analyzes the proposed model with specific examples and compares the model with different algorithm models. The fifth section is the Conclusion, which elaborates on the actual contributions, limitations, and prospects of the results obtained.

2. Literature Review

2.1. Research Status of Video Analysis. The video analysis methods are developed based on structured analysis methods. Events with distinctive features in sports games, such as scores, fouls, and breaks, are detected to better summarize the videos, which enables users to browse videos conveniently and quickly [9]. The events in video analysis are defined as a series of behaviors and actions that are sudden and interesting after a period of straightforward content, which is unexpected and random. Therefore, some unfocused or dull videos cannot be analyzed by the video analysis method based on events [10]. This type of method is applied to dialog detection in movies, sudden events detection in surveillance videos, and target event detection in sports videos. However, because each event needs to be analyzed in combination with the characteristics of its category, it is impossible to establish a general semantic analysis model. The lack of practicality hinders the popularization [11].

Most detection methods for sports video events are based on audio, video, and texture features extracted directly from video data. Combining the actual situation of sports competitions, Lu et al. (2020) proposed an endpoint detection algorithm based on variance features and comprehensively designed a speech recognition model based on the Markov model. The results proved that the model was accurate and had excellent performance, providing a reference for applying artificial intelligence to sports video detection [12]. Morra et al. (2020) proposed a comprehensive method to detect various complicated events in football videos starting from location data. The model could effectively extract key information from sports game videos [13]. Sauter et al. (2021) investigated mental health problems by means of video games. Through video analysis, the results show that the social environment of game players has a great influence on their mental health, which may be combined with game motivation. This became a strong predictor of a clinically relevant high-risk population in the game [14]. The fundamental idea of these methods is to extract low-level or middle-level audio and video features and then use rule-based or statistical learning methods to detect events. These methods can be further divided into single-modal methods and multimodal methods. The single-modal methods believe that only one-dimensional features in the video can be used for event detection. These methods have lower computational complexity and lower accuracy of event detection. The reason is that the live videos are fusions of multidimensional information. The inherent information in the sports videos cannot be adequately expressed by the single-modal features alone [15]. Hence, multimodal methods are introduced to analyze exciting events in sports videos to improve the reliability of event detection performance. Compared with the single-modal methods, the multimodal methods can provide a higher event detection rate, but it comes with higher computational complexity and longer calculation time.

2.2. Football Video Event Detection. Lots of machine learning algorithms are applied to automate the detection of events in football games, which can be divided into two categories. The first category is machine learning algorithms based on generative models. The Yess model is developed as a foundation, including algorithms based on the Bayesian belief network model, algorithms based on the dynamic Bayesian network model, and algorithms based on Hidden Markov Model (HMM). Ji et al. (2019) employed the Markov model combined with other algorithms to detect fouls, scoring, and shooting events. The average precision and recall reached 83.65% and 83.4%, respectively. This type of machine learning algorithm required training samples to train and generate the algorithm model. However, the training samples must be standard and sufficient to get a model with good performance [16]. Since the model generated by this type of algorithm detects target events by simulating the actual situation of events, it can only detect a

single event. The second category is machine learning algorithms based on discriminant models proposed by Zhang et al. (2020), including event detection algorithm based on support vector machine (SVM) and event detection algorithm based on neural networks and conditional random field. They are used for the classification of multiple events [17].

The event detection method based on artificial rules aims to artificially reduce the difference between low-level features and high-level semantics, formulate a set of useful rules based on previous experience and the rules summarized by oneself, and cut across from low-level features to high-level semantics [18]. Nowadays, video event detection has developed excellently; however, there are some problematic issues that need to be studied. (i) Currently, classic machine learning methods have different problems, resulting in low precision and recall of event detection. (ii) The artificial rule-based method is simple to implement and can effectively bridge the semantic gap between low-level features and high-level semantics, providing better event detection performance; however, it depends too much on people's subjective observations and consumes a lot of labors. (iii) Unlike static goals, the key semantic events in sports videos are all dynamic, and their patterns are more complicated.

3. Research Method

3.1. Automatic Detection Algorithms

3.1.1. Shot Segmentation. Simply speaking, shot segmentation detects the boundary frames of each shot in the video through the boundary detection algorithm, which can divide the complete video into a series of independent shots through these boundary frames. The general steps of shot segmentation are (i) calculating the changes in characteristics between frames through a particular algorithm; (ii) obtaining a value that can serve as a basis for judgment as a threshold using experience or algorithm calculation; (iii) once the changes between a frame and its following frame are more significant than the preset threshold, this frame is marked as the boundary frame of the shot for shot segmentation [19]. Because scenes in the football videos are not complicated, and there are comparatively many sudden shots, a simple shot segmentation method based on pixel comparison is selected, considering efficiency and accuracy. The equation for the difference between two frames is

$$D(k, k+1) = \frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} \left| I_k(x, y) - I_{k+1}(x, y) \right|. \quad (1)$$

In (1), $I_k(x, y)$ and $I_{k+1}(x, y)$, respectively, refer to the brightness value of the k-th frame and the k+1-th frame at (x, y), and M and N, respectively, stand for the frame height and width. If the value of $D(k, k+1)$ is small, the changes between the two frames are small; on the contrary, there are some considerable changes between the two frames. When $D(k, k+1)$ is greater than a given threshold, the two frames are considered to belong to two different shots. Specifically, the Twin Comparison algorithm is selected. The algorithm is a dual-threshold technique capable of identifying sudden and gradual changes. Figure 1 is a schematic diagram of the sudden change lens and the gradual change lens after the segmentation processing of the algorithm. They all come from live videos of FIFA World Cup matches. The algorithm can well balance computational complexity and precision.

3.1.2. Keyframe Extraction. A keyframe refers to a frame or several frames in the shot that can be representative. The conservative principle of making mistakes rather than missing one frame is generally adopted when extracting the keyframes to make the video content expressed by the extracted keyframes as comprehensive and complete as possible. The difference between the internal frames of the shots divided by the Twin Comparison algorithm is comparatively small. Therefore, the keyframe extraction method based on the camera boundary is selected by analyzing the advantages and disadvantages of different algorithms and integrating the time cost and effect. Finally, the intermediate frames are decided as the keyframes of the shots by analyzing the structure of football videos [20].

3.1.3. HMM. HMM is a double random process model; the first is the random function set of observable vectors, and the second random process is a hidden Markov chain with some states. After the sequence of the semantic shots is marked, the football game videos can be regarded as a sequence of semantic shots composed of a series of semantic shots. Thus, the sequence of semantic shots can be regarded as a sequence that can be observed by a computer, which is the observation vector in HMM. When people watch this video, different impressions in the human brain will be formed, creating a coherent semantic sequence, which cannot be observed by the computer. This semantic sequence is the hidden Markov chain in HMM. In previous studies, loads of experiments have shown that HMM can indeed describe the production process of video information very accurately [21].

Usually, two probability matrices are used to assist in describing the Hidden Markov Model (HMM). One is used to generate the state sequence (Markov chain), the other is used to constrain the observation sequence, and an initial distribution of the generated Markov chain is needed, which represents the distribution law of each hidden state when the Markov chain tends to be stable [22]. Let M be the size of the observation space, the observation element set is $V = V_1 V_2 V_3 \cdots V_M$, N is the size of the state space, and the state set element is $S = S_1 S_2 S_3 \cdots S_N$. The matrix that generates the state sequence is $A = \{a_{ij}\}_{N*N}$, where $i, j \in 1, 2, \ldots, N$, the matrix that generates the observations is $B = \{b_i(k)\}_{N*M}$, where $i \in 1, 2, \cdots, N; k \in 1, 2, \cdots, M$, the initial state matrix is $PI = \{\pi_1, \pi_2, \cdots, \pi_N\}$. For any t, there is a relationship shown in equation (2):

$$\begin{cases} a_{ij} = P(q_{t+1} = S_j | q_t = S_i), \\ b_i(k) = P(O_{t+1} = v_k | q_t = S_i), \\ \pi_i = P(q_t = S_i), \quad i \in 1, 2, \ldots, N. \end{cases} \quad (2)$$

Figure 1: A schematic diagram of shot segmentation effect.

Among them, A refers to an N^*N matrix. B refers to an N^*M matrix. Meanwhile, PI is a vector of length N. HMM is composed of five parameters: state transition matrix, emission probability (observation) matrix, and initial state matrix, namely HMM is denoted as $\lambda = \{M, N, A, B, PI\}$. Theoretically, when these five model parameters are known, computer simulation generates the corresponding HMM sequence. The basic generator algorithm has four steps:

First, select an initial state $q_t = S_i$ according to the initial state matrix. Currently, $t = 1$;

Second, according to the emission probability, the state is selected as $q_1 = S_i$, and the emission element (observable element) when $t = 1$ is $b_i(k)$;

Third, generate a new state $q_{t+1} = S_j$ according to the state transition matrix $a_{ij} = P(q_{t+1} = S_j | q_1 = S_i)$ and update $t = t + 1$ at the same time;

Fourth, repeat steps 2 and 3 until the target data amount is generated and terminate the program.

3.1.4. Genetic Algorithm. The genetic algorithm simulates the reproduction, mating, and mutation phenomena that occur in natural selection and genetic evolution. Started from any initial population, a group of new and better individuals can be generated through random selection, crossover, and mutation operations. The group evolves to a better area in the search space so that it continues to evolve from generation to generation, and finally, converges to a group of optimal individuals and then selects the optimal solution. The genetic algorithm does not require complicated calculations for optimization problems if the three genetic algorithm operators can obtain the optimal solution [23]. The precise calculation process is displayed in Figure 2:

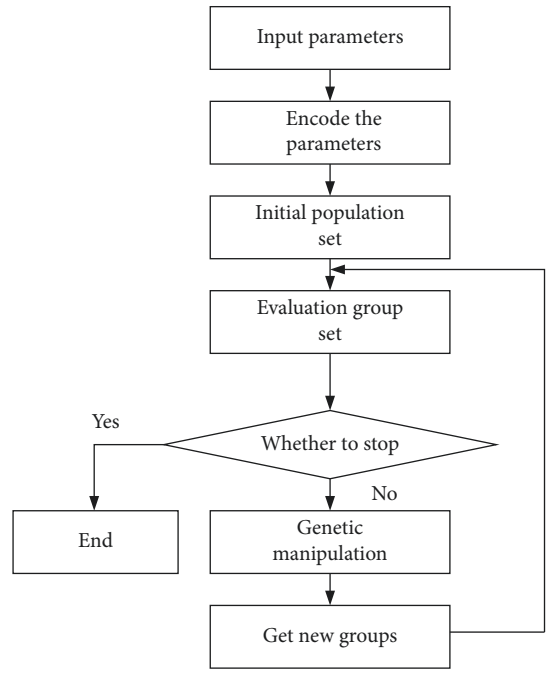

Figure 2: A schematic diagram of the genetic algorithm structure.

3.2. System Model Construction. The designed scoring event detection process of football game videos is based on real applications. As shown in Figure 3, the first step is shot segmentation, which not only makes the video description more convenient but also reduces the time cost of event detection through the extraction and application of keyframes. After the shot segmentation is finished, only some shots are obtained, while the computer does not know what meaning these shots represent, so that these shots need to be semantically annotated. After the semantic annotation of all the shots is completed, a unique shot-based event positioning method is proposed. The scoring events are taken as examples to verify the feasibility of this method. The suspected event positioning is to extract the sequence of the semantic shots of suspected scoring events by combining some simple artificial rules with some algorithms. Research results of the event detection algorithm in recent years suggest that the HMM model is the most commonly used and classic model. In this regard, HMM is selected as the final detection algorithm.

In the constructed key event monitoring model, the hidden Markov model of the scoring event includes the state set of the HMM model of the scoring event = {the game is on, the game is suspended}. The observation set of the HMM model for scoring events, i.e., the set of semantic shots = {far shot, medium shot, close-up shot, spectator shot, playback shot}. The observation sequence in the scoring event HMM model is defined as a semantic shot sequence obtained by a video segment marked by semantic shots.

K video sequences describing goal-scoring events are selected to perform artificial semantic annotation on the segmented physical shots. K semantic shot sequences are used as the training data set. The game state of each semantic shot is judged, and K state sequences are obtained. Let the initial state probability π_i:

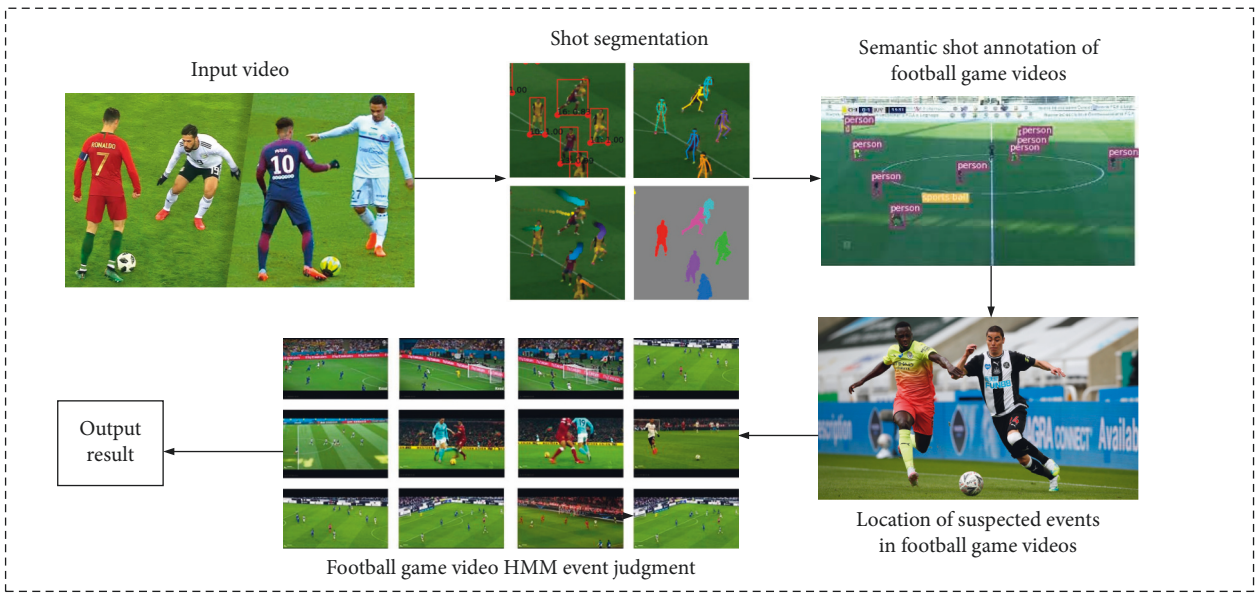

Figure 3: Key event detection1 model based on semantic analysis.

$$\pi_i = \frac{n_i}{n}, \quad 1 \le i \le N, \tag{3}$$

In (3), among the K training sequences, n_i refers to the number of shots in state θ_i, n refers to the number of all semantic shots, N refers to the number of HMM states, and $N = 2$ means that the game is on and the game is suspended.

Let the state transition matrix $A = \{a_{ij}\}_{N*N}$. Among them, a_{ij} is shown in (4):

$$a_{ij} = \frac{n_{(i,j)}}{n_{(i,*)}}, \quad 1 \le i, j \le N. \tag{4}$$

Among them, the K training sequences $n_{(i,j)}$ refer to the number of shots transferred from state θ_i to state θ_j. $n_{(i,*)}$ refers to the number of shots that transition from state θ_i to any state. Let the observation matrix $B = \{b_i(k)\}_{N*M}$. Among them, $b_i(k)$ is shown in (5):

$$b_i(k) = \frac{n_{i,k}}{n_i}, \quad i \in 1, 2, \ldots, N; k \in 1, 2, \ldots, M. \tag{5}$$

Among them, in the K training sequences, $n_{i,k}$ refers to the number of kth semantic shots in state θ_i, n_i refers to the number of lenses in state θ_i, and M refers to the number of semantic shot types in the observation set and $M = 5$.

The weighted semantic sums are normalized to remove the interference caused by detecting video length. Semantic information I_k is the amount of information contained in the semantic shot S for the goal event, as shown in (6) and (7):

$$I_k = -\log\left(\overline{P}(s_k|goal)\right) 1 \le k \le 5, \tag{6}$$

$$\overline{P}(s_k|\text{goal}) = \frac{1}{K}\sum_{x=1}^{K} P_x(s_k|\text{goal}). \tag{7}$$

Among them, K refers to the number of samples in the training set. s_k belongs to any shot in the semantic shot set. The goal is the scoring event, and $\overline{P}(s_k|\text{goal})$ refers to the average probability of the shot s_k appearing in the scoring event. $P_x(s_k|\text{goal})$ refers to the probability of occurrence of shot s_k in the xth goal segment. The semantic observation weight is W_k, as shown in (8):

$$W_k = \frac{1}{I_k}, \quad 1 \le k \le 5. \tag{8}$$

The semantically weighted sum S′ refers to the semantically weighted sum of video clips containing m shots, as shown in (9) and (10).

$$S' = \sum_{k=1}^{5} W_k \times n_{s_k}, \tag{9}$$

$$\sum_{k=1}^{5} n_{s_k} = m. \tag{10}$$

Among them, W_k refers to the semantic observation weight of the shot s_k. n_{s_k} refers to the number of semantic shots s_k in the video clip. The normalized semantic weighted sum S represents the normalized semantic weighted sum of the video clips containing m semantic shots, as shown in (11):

$$S = \frac{1}{m} \times S'. \tag{11}$$

3.3. Model Data Training. The videos selected for the experiment come from live videos of football games, including Fédération Internationale de Football Association (FIFA), Union of European Football Association (UEFA), and Liga De Primera Division Banco Bilbao Vizcaya Argentaria (LIGA BBVA). The video data are divided into a training set and a test set in the event detection experiment. The video

Table 1: Information about scoring events in experiment videos.

Video source	ID	Games played	Date	Score	Video length
FIFA	F1	England vs. the United States	2010.6.13	1 : 1	108 : 31
	F2	Germany vs. Australia	2010.6.14	4 : 0	102 : 26
	F3	Spain vs. Switzerland	2010.6.16	0 : 1	107 : 53
	F4	Germany vs. Argentina	2010.7.3	4 : 1	108 : 34
UEFA	U1	Real Madrid vs. Dinamo Zagreb	2011.11.23	6 : 2	95 : 40
	U2	Bayern Munich vs. Villarreal	2011.11.23	3 : 1	106 : 24
	U3	Naples vs. Manchester city	2011.11.23	2 : 1	100 : 35
	U4	AC Milan vs. Barcelona	2011.11.24	2 : 3	109 : 58
La liga	L1	Real Madrid vs. Osasuna	2011.11.6	7 : 1	96 : 04
	L2	Bilbao vs. Barcelona	2011.11.7	2 : 2	99 : 57
	L3	Real Madrid vs. Atletico madrid	2011.11.27	4 : 1	98 : 28
	L4	Real Madrid vs. Barcelona	2011.12.11	1 : 3	104 : 50

Figure 4: Some video images in the datasets.

sample data of scoring events used for training include a total of 40 footage, with 20 scoring footage and 20 non-scoring footage. The test data include 23 scoring footage and 30 nonscoring footage. The numbers of training footage for corner kicks, penalty kicks, and red and yellow card events are 40, 40, and 62, respectively; the respective numbers of test footage are 70, 50, and 62. The video information of scoring events is shown in Table 1. Different video images at various angles in the video sample data are shown in Figure 4.

3.4. Model Performance Analysis. In order to evaluate the performance of the constructed model, it is analyzed from multiple perspectives, such as semantic clue extraction, different parameter changes, different model comparison, different data set testing, and key event detection results. Among them, when analyzing the changes of different parameters, the relationship between the hidden state n and the window length w is analyzed to detect multiple exciting events. Among them, the hidden state n is taken from 1 to 4, from small to large. This algorithm model is compared with literature A [24], literature B [25], and literature C [26] proposed by scholars in related fields. The advantages of model performance under different models are compared. In the analysis of the key event detection results, the algorithm model is compared with the pieces of literature E [27] and F [28] proposed by scholars in related fields.

The precision and recall are used as experimental evaluation criteria to verify the performance of the proposed model. The precision denotes the proportion of the correctly recognized positive categories to all the positive category samples. Its calculation is as follows:

$$\text{Precision} = \frac{M}{M+N}. \tag{12}$$

The recall is the proportion of all positive category samples that are correctly identified as positive categories. It is calculated as follows:

$$\text{Recall} = \frac{M}{M+P}. \tag{13}$$

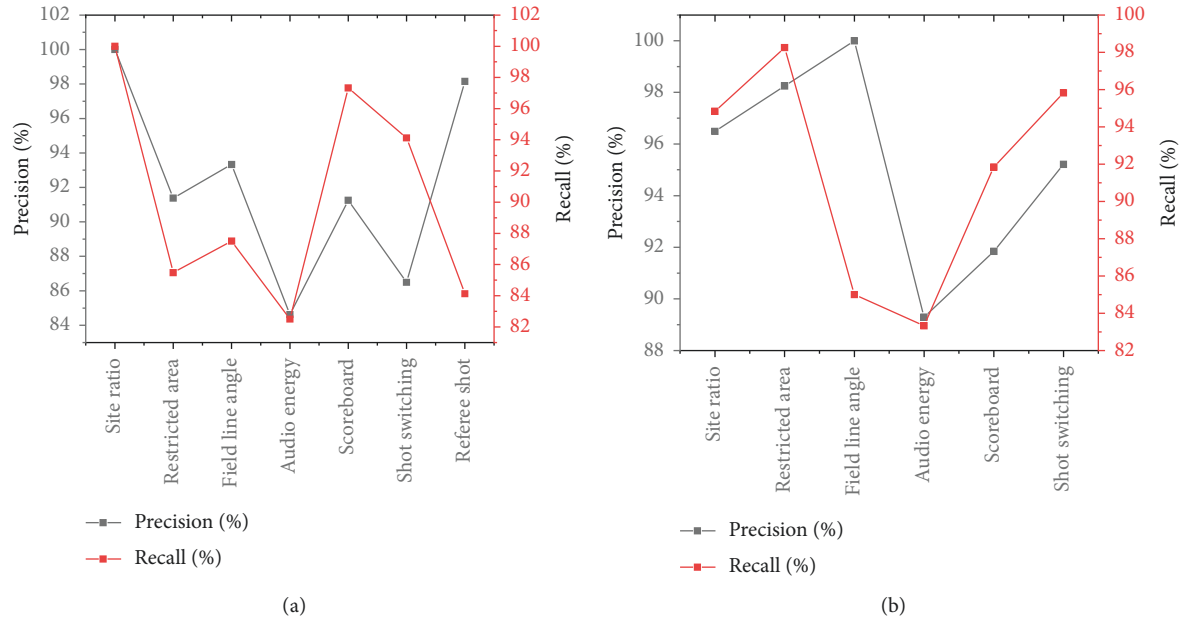

Figure 5: Extraction results of low-level features ((a) Training set; (b) Test set).

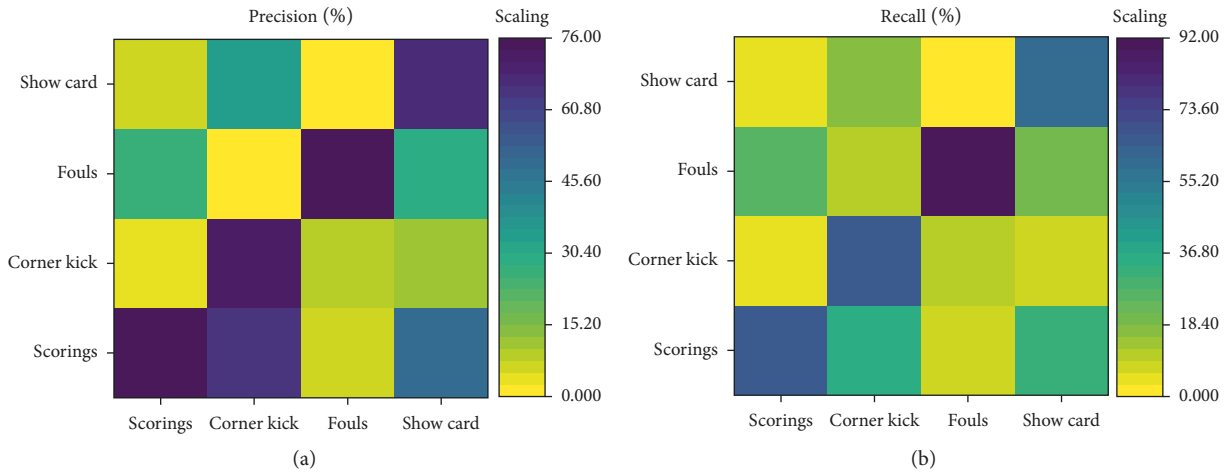

Figure 6: Football video clustering results ((a) precision; (b) recall).

In (3), M stands for the number of correct negative semantics recognized by this model, N refers to the number of positive semantics of unrecognized errors, and P denotes the number of negative semantics of unrecognized errors.

4. Results and Discussions

4.1. Extraction and Analysis of Semantic Clues. Figure 5 presents the results of model semantic clue extraction. The definition extraction method of low-level features can express the characteristics of key events, with an accuracy of over 82%. Furthermore, the experimental results suggest that the defined semantic clue extraction method can express the potential laws of scorings, fouls, corner kicks, and red and yellow card events effectively. The method is efficient and straightforward and provides a theoretical basis for the subsequent event detection effectively.

Figure 6 shows the results of the football video feature extraction. The recall of various football events using feature clustering remains in the range of 68%~76%, and the precision is 60%~92%. This result shows that the preliminary screening of semantic clues by clustering can accurately reflect the underlying laws of football events, show the unique characteristics of various events, and distinguish various events automatically and effectively to get the emotional feature combination of various events.

4.2. Changes of Different Parameters. Figure 7 shows the model performance results under different parameter changes. When the parameter $n = 1$, because there is only one hidden state, the internal structure of the input observation cannot be successfully trained to simulate the internal structure of the input observation value. No matter

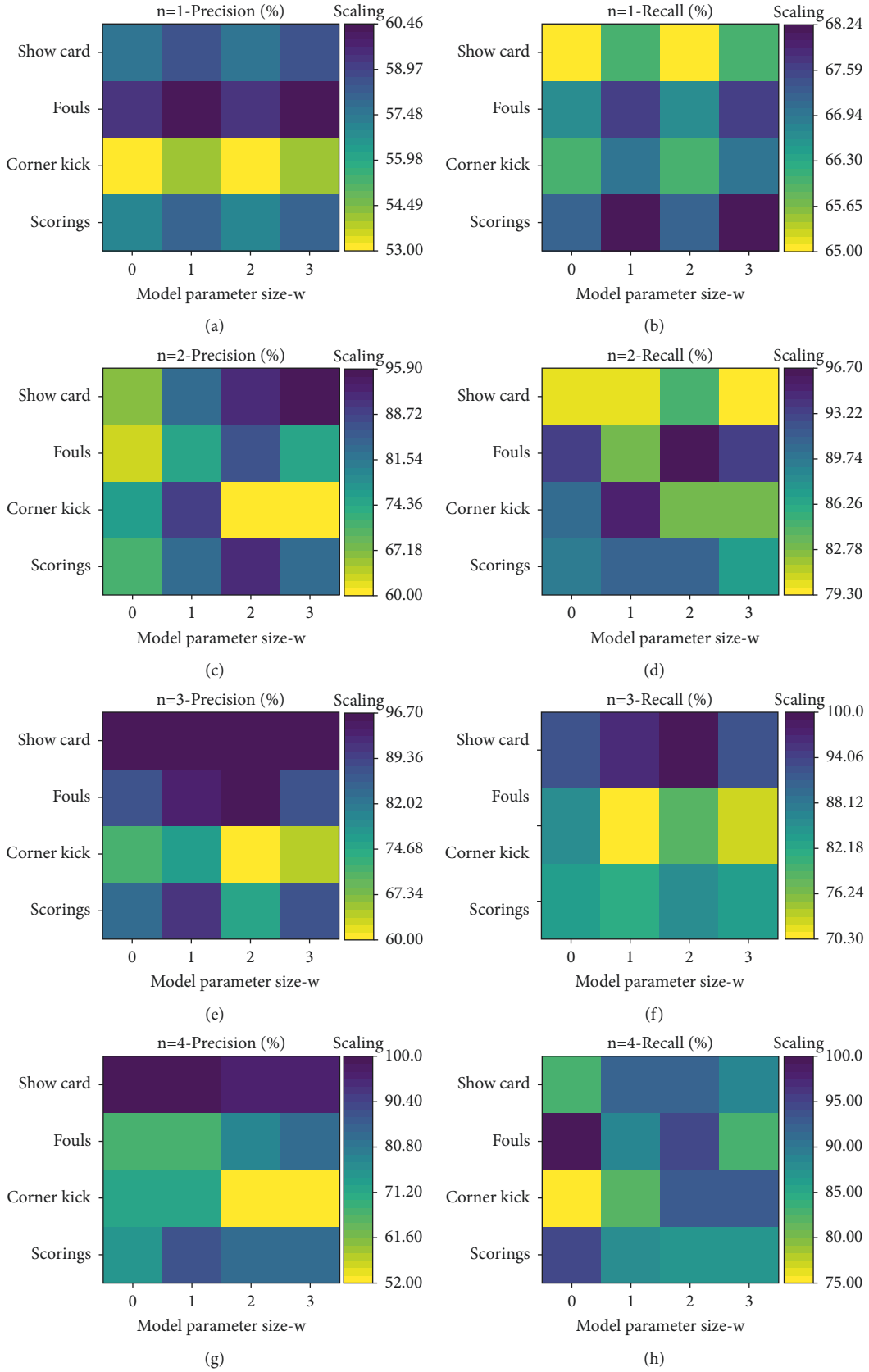

Figure 7: Model performance results under different parameter changes (a) $n = 1$-precision; (b) $n = 1$-recall; (c) $n = 2$-precision; (d) $n = 2$-recall; (e) $n = 3$-precision; (f) $n = 3$-recall; (g) $n = 4$-precision; (h) $n = 4$-recall).

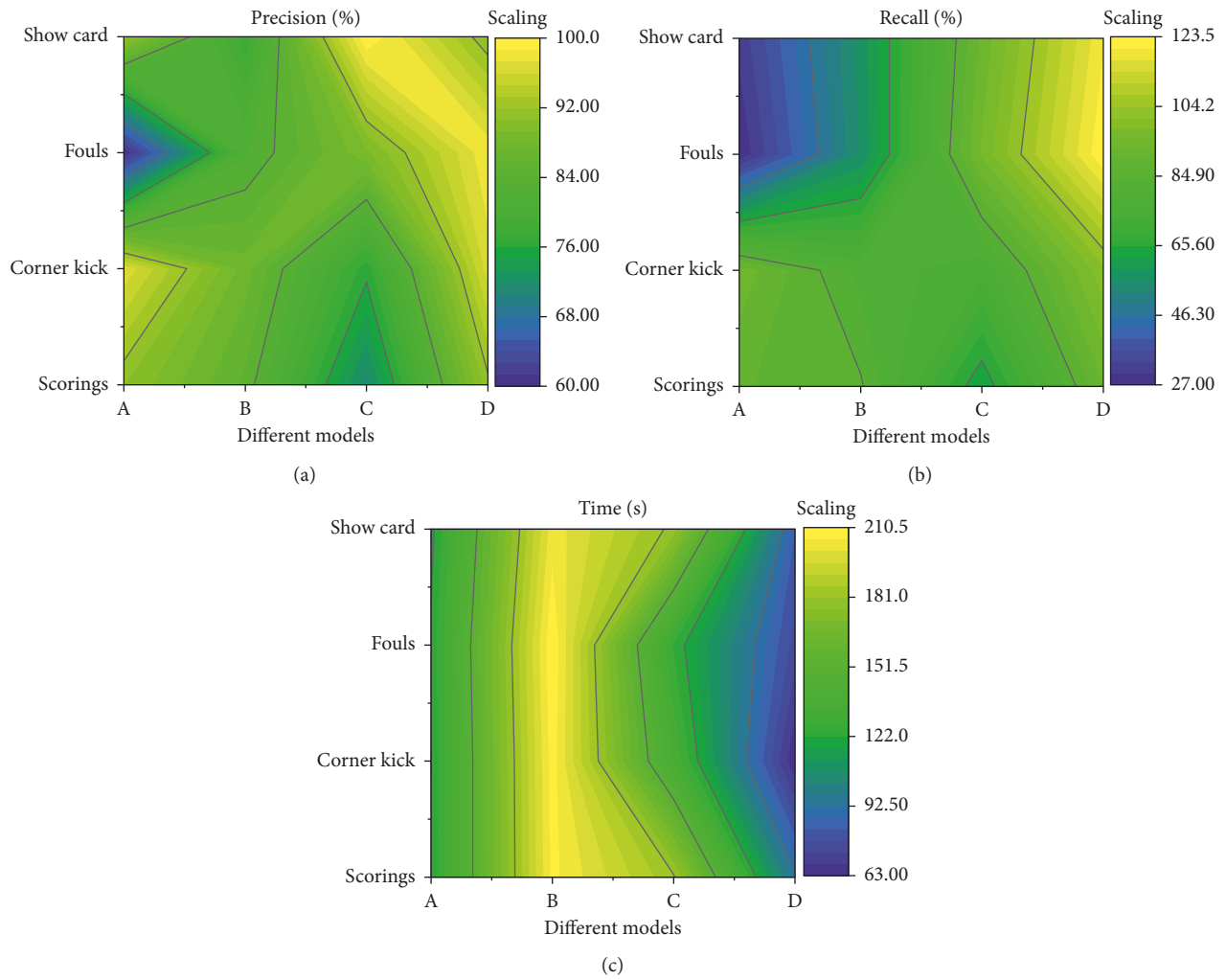

Figure 8: Comparison of the event detection results with algorithms of the references ((a) precision; (b) recall; (c) time required).

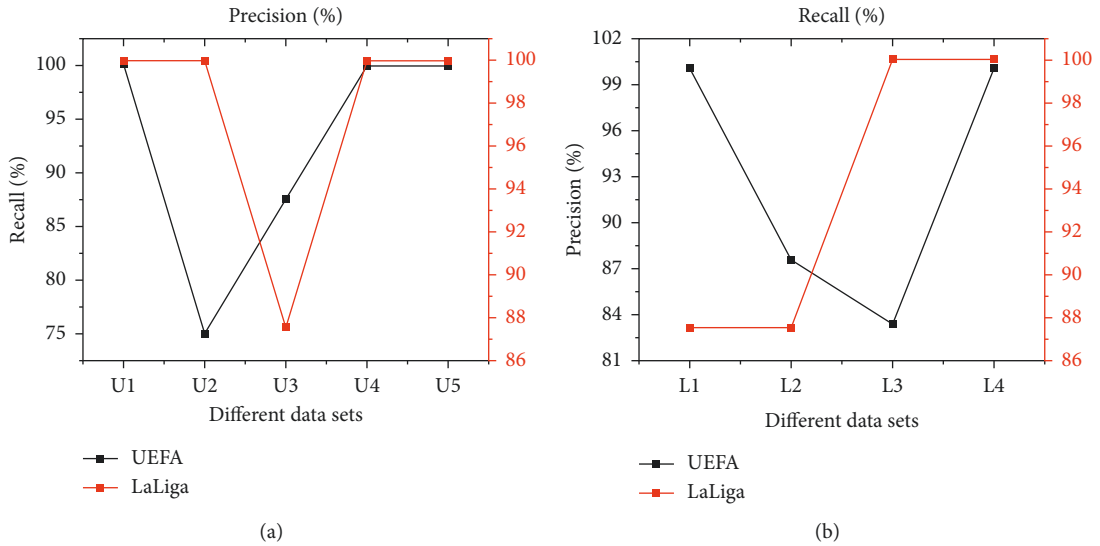

Figure 9: Test results on different data sets ((a) UEFA; (b) La Liga).

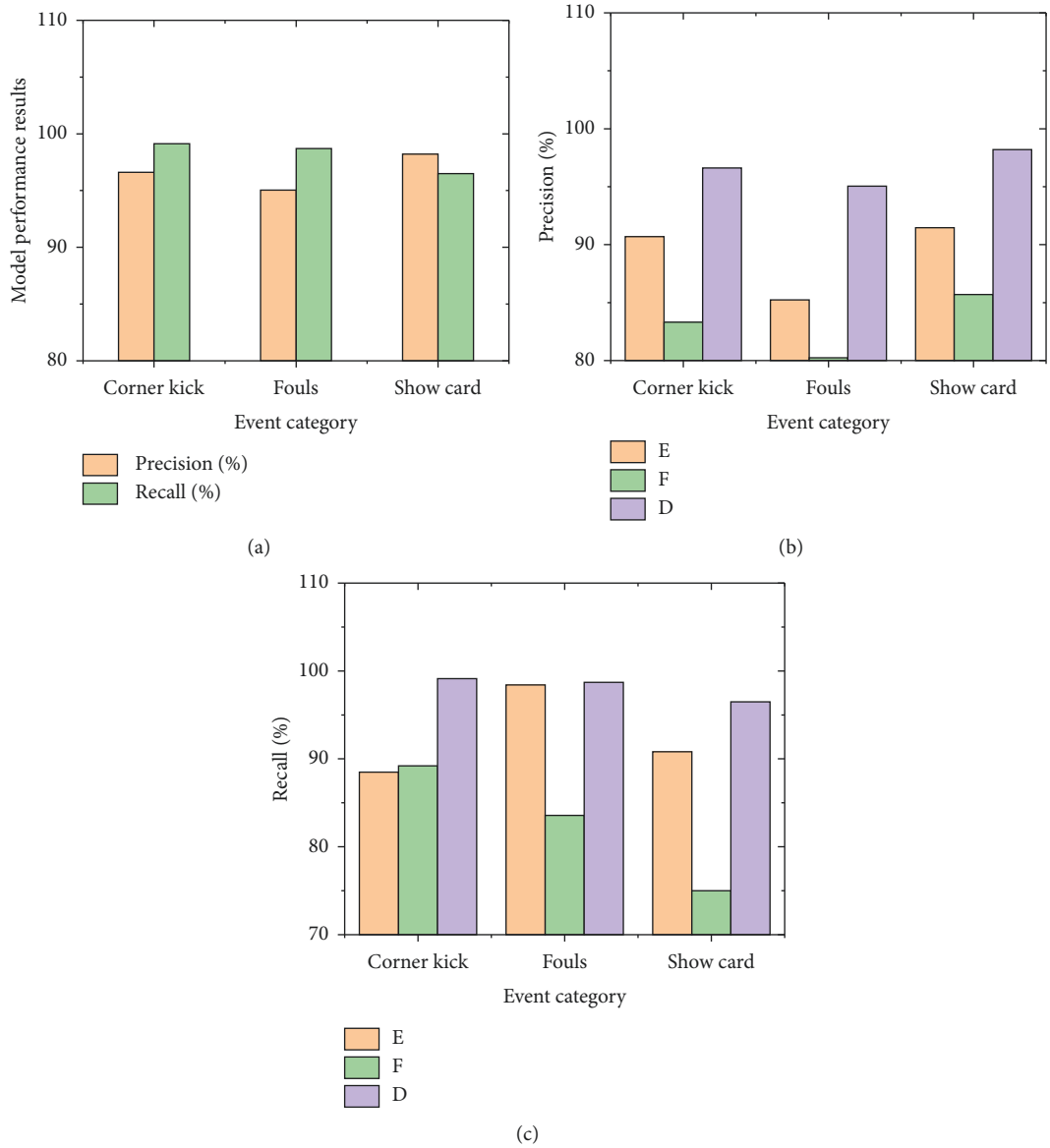

Figure 10: Detection results of key events (a) precision and recall results of the model constructed; (b) precision results compared with other methods; (c) recall results compared with other methods).

how the window length is changed, the detection result cannot be improved. When the model increases a hidden state number, and $n = 2$, the increase in the model state leads to an increase in the expressive power of events, which can better simulate the relationship between the input observations objectively. At $n = 1$, the recall and prescience reach 62.5% and 93.33%, respectively. In the meantime, when the number of hidden states is fixed and the parameters are increased, the correlation between the front and back observation vectors is taken into consideration when modeling, which is more in line with the actual occurrence of events. Hence, the detection performance of events has been further improved. When the model parameter n is = 3, the precision of the event detection increases from 62.5% to 87.5%. The reason is the increase in the number of hidden states, which strengthens the ability to describe events. The recall and prescience can reach 96.67% and 100%, respectively. When $n = 4$, there are too many hidden states, while the input observations do not need these many hidden states to simulate the probability prediction model. At this time, changing the window length parameter cannot improve the efficiency of event detection, and the detection performance of the target under this model parameter is not optimal. In summary, the above analysis suggests that the number of hidden states is the key to determining the expressive ability of the model. When the window length is too large and exceeds the objective reality, the complexity of the model will increase, bringing more computational complexity.

4.3. Performance Comparison of Different Models. Figure 8 shows the performance comparison results of different models. The effect of the proposed model is better than that of the reference methods. It measures the

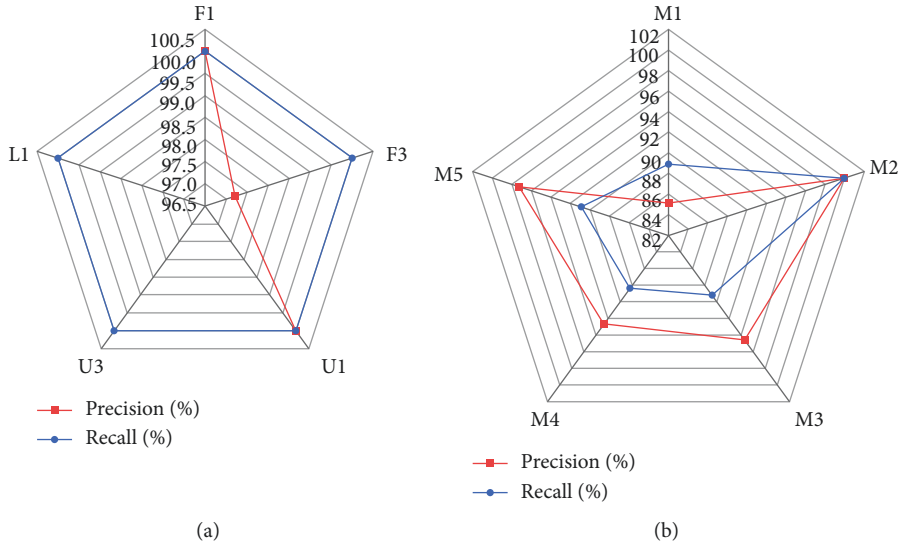

Figure 11: Comparison of results of key event detection on different data sets.

timeliness of the method by recording the time it takes to extract features. The event detection time is 64.79% of reference A [24], 48.89% of reference B [25], and 37.23% of reference C [26]. The reason is that the proposed method first filters 13 semantic features for each different event by clustering. Afterward, only two ~ four features of event detection in this method are used, reducing the number of feature types required for detection and effectively improving the timeliness of event detection. However, reference A requires nine features to detect each event; in contrast, reference B and reference C use 7 and 17 features, respectively. Therefore, they require more types of features and consume more computing time and resources to extract video features, thereby reducing the timeliness of event detection.

4.4. Test on Different Data Sets. As shown in Figure 9, the adaptability tests are performed to prove the effectiveness of this method. The videos of various leagues, such as the UEFA and LIGA BBVA, are selected as test data, and the key event detection is tested. Figure 9 shows the annotated results of the footage of red and yellow cards. The average recall and precision of this method for the adaptability test of the test video are 95.83% and 92.59%. Hence, this method has a broader application scope.

4.5. Detection Results of Key Events. Figure 10 presents the key event detection results. The recall and precision of events using the text detection method can reach 96.62% and 98.81%, respectively. Hence, the proposed keyword definition method is simple and effective. It can dig deeper into the structural semantics and potential laws of network text description and accurately find the location of key events in the text. Compared with the state-of-the-art references *E* [27] and F [28], the average precision and recall of the proposed method are 5.55% and 7.49% higher than that of the BN method in reference *E* and 15.71% and 12.90% higher than the method in reference F. The reason is that the text keywords are effectively defined by integrating the artificial rule into event detection. Moreover, the time labels of the key events are accurately found, which overcomes the common problems of unclear semantics in general event detection methods, thereby improving the precision and recall of event detection.

Figure 11 shows the comparison results of key event detection of methods proposed in different references. The recall of the model optimized by the genetic algorithm reaches 99.39%, and the precision reaches 100%. Therefore, the proposed video time extraction method has a strong resolving power and apparent advantages in different data sets. This model can accurately align the occurrence time of events with the time of the text, laying a foundation for the subsequent accurate segmentation of the start and end frames of video footage.

5. Conclusions

Based on the results of predecessors, this study aims at the problems of high time cost, low detection accuracy, and difficult standard training samples in the current detection of key events in football videos. Based on semantic analysis, it innovatively uses lens annotation. In this way, to ensure accuracy, the time cost of semantic annotation is reduced. By dividing the shots, the range of the resolution shots is greatly reduced, and the accuracy of adding artificial rule models to the shots is improved. The genetic algorithm is used to improve the HMM algorithm to make the data more stable during the training process and can generate a more optimized precision model with fewer training samples. The proposed video event detection model based on the combination of artificial rules and machine learning algorithms can effectively save event costs and improve the detection accuracy of the model. Although the constructed model is suitable for football event detection, it still has several shortcomings. First, establishing artificial rules will consume

time and cost, which will significantly affect the efficiency of video analysis. Therefore, further optimization processing is required for artificial rules. Second, the accuracy and learning ability of the model used for video key event detection may not be as good as the latest deep learning algorithms. Some state-of-the-art models put forward higher requirements on equipment and computer configuration; however, the performance will be improved. These two directions will be explored and analyzed in-depth in the following investigations to improve the proposed video key event detection model.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Z. Lv, R. Lou, J. Li, A. K. Singh, and H. Song, "Big data analytics for 6G-enabled massive internet of things," *IEEE Internet of Things Journal*, vol. 8, no. 7, pp. 5350–5359, 2021.

[2] M. Sadrishojaei, N. J. Navimipour, M. Reshadi, and M. Hosseinzadeh, "A new preventive routing method based on clustering and location prediction in the mobile internet of things," *IEEE Internet of Things Journal*, vol. 8, no. 13, pp. 10652–10664, 2021.

[3] G. Manogaran, M. Alazab, H. Song, and N. Kumar, "CDP-UA: cognitive data processing method wearable sensor data uncertainty analysis in the internet of things assisted smart medical healthcare systems," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 10, pp. 3691–3699, 2021.

[4] A. Tejero-de-Pablos, Y. Nakashima, T. Sato, N. Yokoya, M. Linna, and E. Rahtu, "Summarization of user-generated sports video by using deep action recognition features," *IEEE Transactions on Multimedia*, vol. 20, no. 8, pp. 2000–2011, 2018.

[5] H. Zhou, G. Sun, S. Fu, L. Wang, J. Hu, and Y. Gao, "Internet financial fraud detection based on a distributed big data approach with node2vec," *IEEE Access*, vol. 9, pp. 43378–43386, 2021.

[6] D. J. Samuel R.Samuel R, F. E, G. Manogaran et al., "Real time violence detection framework for football stadium comprising of big data analysis and deep learning through bidirectional LSTM," *Computer Networks*, vol. 151, pp. 191–200, 2019.

[7] R. M. Chambers, T. J. Gabbett, R. Gupta et al., "Automatic detection of one-on-one tackles and ruck events using microtechnology in rugby union," *Journal of Science and Medicine in Sport*, vol. 22, no. 7, pp. 827–832, 2019.

[8] S. M. Daudpota, A. Muhammad, and J. Baber, "Video genre identification using clustering-based shot detection algorithm," *Signal, Image and Video Processing*, vol. 13, no. 7, pp. 1413–1420, 2019.

[9] S. Wan, X. Xu, T. Wang, and Z. Gu, "An intelligent video analysis method for abnormal event detection in intelligent transportation systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4487–4495, 2020.

[10] C. Loukas, "Video content analysis of surgical procedures," *Surgical Endoscopy*, vol. 32, no. 2, pp. 553–568, 2018.

[11] X. Zhang, Q. Yu, and H. Yu, "Physics inspired methods for crowd video surveillance and analysis: a survey," *IEEE Access*, vol. 6, pp. 66816–66830, 2018.

[12] Y. Lu and S. An, "Research on sports video detection technology motion 3D reconstruction based on hidden Markov model," *Cluster Computing*, vol. 23, no. 3, pp. 1899–1909, 2020.

[13] L. Morra, F. Manigrasso, G. Canto, C. Gianfrate, E. Guarino, and F. Lamberti, "Slicing and dicing soccer: automatic detection of complex events from spatio-temporal data," 2020, https://arxiv.org/abs/2004.04147.

[14] M. Sauter, T. Braun, and W. Mack, "Social context and gaming motives predict mental health better than time played: an exploratory regression analysis with over 13,000 video game players," *Cyberpsychology, Behavior, and Social Networking*, vol. 24, no. 2, pp. 94–100, 2021.

[15] X. Xu, Q. Wu, L. Qi, W. Dou, S.-B. Tsai, and M. Z. A. Bhuiyan, "Trust-aware service offloading for video surveillance in edge computing enabled internet of vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1787–1796, 2021.

[16] S. Ji, Y. Shan, and X. Li, "WITHDRAWN: video structure syntax mining based on Hidden Markov Model in sports video," *Journal of Visual Communication and Image Representation*, pp. 102695–102706, 2019.

[17] R. Zhang, L. Wu, Y. Yang, W. Wu, Y. Chen, and M. Xu, "Multi-camera multi-player tracking with deep player identification in sports video," *Pattern Recognition*, vol. 102, pp. 107260–110773, 2020.

[18] Z. Li, L. Yao, X. Chang, K. Zhan, J. Sun, and H. Zhang, "Zero-shot event detection via event-adaptive concept relevance mining," *Pattern Recognition*, vol. 88, pp. 595–603, 2019.

[19] J. Xing and X. Li, "WITHDRAWN: feature extraction algorithm of audio and video based on clustering in sports video analysis," *Journal of Visual Communication and Image Representation*, pp. 102694–102703, 2019.

[20] K. Kumar, D. D. Shrimankar, and N. Singh, "Eratosthenes sieve based key-frame extraction technique for event summarization in videos," *Multimedia Tools and Applications*, vol. 77, no. 6, pp. 7383–7404, 2018.

[21] Y. Jin, Y. Long, C. Chen, Z. Zhao, Q. Dou, and P.-A. Heng, "Temporal memory relation network for workflow recognition from surgical video," *IEEE Transactions on Medical Imaging*, vol. 40, no. 7, pp. 1911–1923, 2021.

[22] M. Stoeve, D. Schuldhaus, A. Gamp, C. Zwick, and B. M. Eskofier, "From the laboratory to the field: IMU-based shot and pass detection in football training and game scenarios using deep learning," *Sensors*, vol. 21, no. 9, p. 3071, 2021.

[23] P. P. Roy, P. Kumar, and B.-G. Kim, "An efficient sign language recognition (SLR) system using Camshift tracker and hidden Markov model (hmm)," *SN Computer Science*, vol. 2, no. 2, pp. 79–15, 2021.

[24] P. Bauer and G. Anzer, "Data-driven detection of counterpressing in professional football," *Data Mining and Knowledge Discovery*, vol. 35, no. 5, pp. 2009–2049, 2021.

[25] L. F. Gabler, S. H. Huddleston, N. Z. Dau et al., "On-field performance of an instrumented mouthguard for detecting head impacts in American football," *Annals of Biomedical Engineering*, vol. 48, no. 11, pp. 2599–2612, 2020.

[26] R. Izzo, T. D'isanto, G. Raiola, A. Cejudo, N. Ponsano, and C. H. Varde'i, "The role of fatigue in football matches, performance model analysis and evaluation during quarters

using live global positioning system technology at 50hz," *Sport Science*, vol. 13, no. 1, pp. 30–35, 2020.

[27] G. J. Tierney, C. Kuo, L. Wu, D. Weaving, and D. Camarillo, "Analysis of head acceleration events in collegiate-level American football: a combination of qualitative video analysis and in-vivo head kinematic measurement," *Journal of Biomechanics*, vol. 110, pp. 109969–109974, 2020.

[28] L. Morra, F. Manigrasso, and F. Lamberti, "SoccER: computer graphics meets sports analytics for soccer event recognition," *Software*, vol. 12, pp. 100612–100623, 2020.