Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 1140611, 11 pages
https://doi.org/10.1155/2021/1140611

Research Article

Big Data and Deep Learning-Based Video Classification Model for Sports

Lin Wang,[1] Haiyan Zhang,[2] and Guoliang Yuan[2]

[1]*Department of Physical Education, North China University of Science and Technology, 063210 Tangshan, Hebei, China*
[2]*College of Physical Education, Hengshui University, Hengshui, 053000 Hebei, China*

Correspondence should be addressed to Haiyan Zhang; zhy18903182110@126.com

Received 28 August 2021; Revised 8 September 2021; Accepted 14 September 2021; Published 7 October 2021

Academic Editor: Yuanpeng Zhang

Information technologies such as deep learning, big data, cloud computing, and the Internet of Things provide key technical tools to drive the rapid development of integrated manufacturing. In recent years, breakthroughs have been made in big data analysis using deep learning. The research on the sports video high-precision classification model in this paper, more specifically, is the automatic understanding of human movements in free gymnastics videos. This paper will combine knowledge related to big data-based computer vision and deep learning to achieve intelligent labeling and representation of specific human movements present in video sequences. This paper mainly implements an automatic narrative based on long- and short-term memory networks to achieve the classification of sports videos. In the classical video description model S2VT, long- and short-term memory networks are used to learn the mapping relationship between word sequences and video frame sequences. In this paper, we introduce an attention mechanism to highlight the importance of keyframes that determine freestyle gymnastic movements. In this paper, a dataset of freestyle gymnastics breakdown movements for professional events is built. Experiments are conducted on the data and the self-constructed dataset, and the planned sampling method is applied to eliminate the differences between the training decoder and the prediction decoder. The experimental results show that the improved method in this paper can improve the accuracy of sports video classification. The video classification model based on big data and deep learning is to provide users with a better user experience and improve the accuracy of video classification. Also, in the experiments of this paper, the effect of extracting features for the classification of different lifting sports models is compared, and the effect of feature extraction network on the automatic description of free gymnastic movements is analyzed.

1. Introduction

With the rapid development of computers, networks, multimedia, and other related technologies, multimedia data has shown an exponential growth trend. A video is a common form of multimedia data, and it is one of the important components of multimedia data [1], which is closely related to our daily life. Video contains the richest data information, with a more complex structure and a large amount of data. Faced with such a huge video data, automatic video description can better manage and utilize these rich video resources, which can help users to improve the indexing speed as well as the search quality of online videos, so that they can play a greater role. For people with impaired vision, the automatic description of videos and combined with text-to-speech conversion technology converts the text within the computer into continuous natural language for communication. It can help them to understand the content in the video better, thus making life easier for the visually impaired. In the field of automatic video description research, automatic human action-based video analysis and understanding have gradually become a popular research problem in computer vision and pattern recognition in recent years. It has a wide application prospect in the fields of intelligent life assistance, advanced human-computer interaction, and content-based video retrieval and is closely followed by researchers at home and abroad [2].

Existing research results for high-precision classification algorithms and their conceptual drift still rely mainly on data structures, and algorithm optimization on data mining

as well as detection of conceptual drift is still mainly done by standalone computers with limited computational resources.

The growing and escalating levels of data and data complexity make it insufficient to rely solely on the algorithm itself and single-computer computing resources [3]. The use of distributed computing platforms to cope with the huge consumption of time complexity and space complexity of algorithms in big data environments and to address the problem of conceptual drift in data streams has become a major concern.

Faced with the current problems in the research of sports video high-precision classification models [4], such as low-level video features cannot accurately reflect high-level human semantic concepts, high time complexity, and low recognition accuracy of action recognition algorithms in traditional RGB videos, and the use of single features cannot meet the massive growth of existing video data and its recognition of complex actions, the study of the automatic description of videos represented by competitive sports events has important theoretical research significance and extensive practical application value. In terms of theoretical research, the study of the automatic description of sports video is a cross-cutting topic that integrates multiple disciplines such as big data analysis, machine learning, pattern recognition, video analysis, computer vision, and cognitive science, which provides a good research object for these fields, and its in-depth research can promote the development of related disciplines. With the progress of deep neural network research and the emergence of large-scale datasets in the fields of image classification and object recognition, a number of approaches have attempted to use convolutional neural networks to learn the semantic representation of images and then use recurrent neural networks to achieve their correspondence with natural language. Traditional supervised learning is mainly single-label learning, while real-life target samples are often complex, have multiple semantics, and contain multiple labels [5].

In recent years, automatic video understanding has gradually become a popular research direction in the field of computer vision [6–8]. Compared with image content research, the content of video contains more information, and a single label is not able to completely characterize the content of the video, so most of the problems for automatic video content understanding are multilabel problems. A number of learning algorithms on multilabeling have been proposed in existing research, and based on the problem-solving perspective, these algorithms can be divided into two major categories: the first category is based on problem transformation approaches, where the main difficulty of multilabeling learning lies in the explosive growth of the output space, and to cope with the exponential complexity of the label space, the correlation between labels needs to be mined. Effective mining of correlations between labels is the key to the success of multilabel learning. According to the strength of mining correlations, multilabel algorithms can be divided into three order strategies. First-order strategy: ignore the correlations between labels, e.g., decompose multilabel into multiple independent binary classification problems. Second-order strategy: consider pairwise correlations between labels, such as ranking relevant and irrelevant labels. Higher-order strategies: consider correlations between multiple labels [9], such as considering the effects of all other labels for each label. The second category is based on algorithmically applicable methods. Problem transformation-based methods focus on transforming problem data to make it applicable to existing algorithms; algorithm-applicable methods are those that extend for a particular algorithm to be able to handle multilabel data, improve the algorithm, and apply the data. The video classification model uses a 3D convolution kernel to process space and time dimensions at the same time. However, the 3D convolution model is shallow and has a huge amount of parameters, which is very bloated. Finally, C3D is used. This model achieves the same as the 2014 dual-stream method, accuracy of close video behavior classification. It uses 3D convolution and 3D pooling and fully connected layers to form an 11-layer shallow network. Its biggest advantage lies in speed. However, the size of the C3D model reaches 321 MB, which is even larger than 152-layer ResNet 235 MB model. Such a model is difficult to train and cannot be pretrained on a large-scale image data set like ImageNet. The shallow network also limits the classification performance of the model.

2. Related Work

Early approaches to automatic video description were rule-based. A language model is used as the basis for predicting the subject, predicate, and object and then complementing the final description of the other constituent videos. For example, the literature describes human activities by introducing a behavioral concept hierarchy, and the literature uses a semantic hierarchy to learn semantic relations between different segments. The literature uses conditional random fields to model objects and activities and generates semantic features for description. In addition, the literature proposes a unified framework consisting of a semantic language model, a deep video model, and a joint embedding model to learn associations between videos and natural sentences. However, all of the above approaches rely excessively on well-defined rules and are limited by fixed syntactic structures, resulting in generated sentences that are too rigid for everyday descriptions. With the advances in deep neural network research and the availability of large-scale datasets in the fields of image classification and object recognition, a number of approaches have attempted to use convolutional neural networks to learn the semantic representation of images and then use recurrent neural networks to implement their correspondence with natural language.

The literature proposes a long short-term memory neural network, which effectively solves the problem of artificially prolonged time tasks that are difficult to solve by RNNs, and addresses the problem that RNNs are prone to gradient disappearance. The literature introduces the forgetting gate mechanism, which enables the LSTM (long short-term memory) to reset the state. The literature proposes a bidirectional long short-term memory neural network (BLSTM), which is one of the most widely used LSTM models. LSTM is a member of deep learning techniques,

and its basic structure is complex and has high computational complexity, which makes it more difficult to perform deeper learning, for example, Google Translate also only applies 7-8 layers of LSTM network structure. The literature is the first implementation of direct text generation from video, which uses convolutional neural networks to extract features from all frames in the video and then perform average pooling before sending them to the LSTM to decode and generate text. The S2VT proposed in the literature uses a long- and short-term memory network in both the encoder and decoder. The literature uses different models trained on different kinds of features thus to generate the description of the video and then uses an evaluation network to evaluate the correlation between the generated sentences and the video features and selects the best correlation as the final video description. The literature uses multiple types of features such as image features, video features, ambient sound features, speech features [10], and kind features to fuse them as a representation of the video. The literature proposes a transfer unit to model the high-level semantic properties of incoming images and videos that are presented as supplementary knowledge for video representations to facilitate sentence generation.

In the methods of sports video content analysis and recognition research, most of the existing sports video research focuses on the semantic analysis of sports and its recognition research. And in the study of sports classification breakdown, except for ball sports such as football, basketball, golf, and badminton, other sports research is rarely involved. The literature borrows a dynamic Bayesian network to analyze sports video, then extracts image features by Kalman filtering principle and uses EM algorithm to complete DBN parameter learning. The literature proposes a semantic content analysis model based on perceptual concepts and finite state machines to automatically describe and detect meaningful semantic content in sports videos. The literature proposes a pose constraint-free upper-body human target detection algorithm. The literature proposes two novel nonlinear feature fusion models and designs an automatic sports video classification algorithm based on the novel features and support vector machines. The literature proposes an automatic detection method based on convolutional neural networks for the detection problem of multiscale athletes in sports videos [11]. The literature uses feature filtering and support vector machines for sports video recognition, which improves the accuracy of sports video recognition and speeds up the speed of sports video recognition. There are offline computing, real-time stream computing, and two types of computing frameworks.

3. Video Classification Model Based on Big Data and Deep Learning

Big data and the deep market is entering a period of rapid development with scale applications in industries such as healthcare. In the market, there are video classification modes based on partial differential equations, and video classification modes based on big data and deep learning, education, the Internet, and commerce. The huge amount of data resources such as user's behavioral characteristics need to be further analyzed and mined to build commercial sports video classification models to provide to producers and sellers to improve the product experience. Big data and deep learning technologies will not only help to increase the added value of the product but also maximize the value of the customer experience. Countries now have a huge market for big data technology and also have the data resources and technology accumulation base for big data to enable more accurate sports video classification [12].

3.1. Streaming Data Model for Sports Video Big Data

3.1.1. Basic Introduction to Streaming Data for Sports Video Big Data. Streaming data is a set of sequential, large, fast, and continuous data sequences. In general, a data stream can be regarded as a dynamic data collection that grows indefinitely over time. It is used in the fields of network monitoring, sensor network, aerospace, meteorological measurement and control, and financial services. The study of streaming data is the study of a new type of data processing model and therefore requires a different approach to data query mining than the traditional method, adapted for streaming data application scenarios, to implement a streaming data query and mining algorithm based on a distributed streaming computing framework and form a prototype algorithm system to query and mine the latest data as fast as possible and give results [13], solving the streaming data loop practical problems in the environment. Whether it is in the field of data mining or database research, more and more experts and scholars at home and abroad are paying attention to the research to stream data will be a focus that cannot be ignored in the future, and top conference sessions such as SIGMOD, ICDE, VLDB, and ICDM will have related topics every year. The research on data analysis and processing for big data is mainly in two aspects, stream data query and stream data mining. Data flow management system is mainly for the research of data flow query and on this basis continue to deepen and build some data management solutions for specific application background items; stream data analysis and processing methods have significant differences with the traditional data processing; first of all, stream data is real-time dynamic data flow and is dynamic relative to the traditional static data form. Secondly, the data elements of streaming data are independent and random, and the dynamic changes of data flow are not easy to predict. More importantly, the real-time nature of streaming data and the large volume of data makes it more time and space complex for the system compared to traditional data. Streaming data requires real-time in-memory processing and outputting the results, rather than traditional data by storing it to disk first and then computing it. Streaming data processing differs significantly from traditional data processing in terms of data form, processing characteristics, data storage methods, update rates, and real-time requirements (as shown in Table 1).

The first thing to consider with stream data mining algorithms is what data to process. The most accurate solution is, of course, the one that is most accurate for all.

Table 1: Characteristics of stream data processing and comparison of traditional requirements with traditional data.

Feature requirements	Traditional data	Streaming data
Data format	Continuous and stable data flow	High-speed data flow
Data storage method	Passive	Initiative
Efficient	Low	High

All the data that have arrived are mined without any omission, but the characteristics of stream data make the method impossible to achieve. Stream data processing model is the study of data stream summary model, that is, the analysis of which part of the data stream needs to be processed for analysis. For stream data, its large data volume characteristics is so often taken according to the algorithm needs to select a range of data within the accuracy requirements for processing, rather than all data as the processing object [14]. The current data stream processing is mainly based on the approximation of the data selection time range, so the selection of the range can be divided into snapshot model, boundary marker model, and sliding window model.

3.1.2. Stream Data Mining Algorithm Features

(1) Incrementalization: traditional data mining algorithms are mined for offline datasets, where data may be accessed by the algorithm multiple times to get the final converged model, whereas real-time mining algorithms emphasize making full use of the latest data for predictive model training or finding knowledge and patterns in the latest data. Therefore, real-time mining algorithms need to be able to continuously update the original training model using the latest incremental data

(2) Anti-interference: traditional data mining algorithms run data mining with a single background principle or pattern of the target data, i.e., the target data set is influenced by the same factors or contains knowledge of the same pattern, whereas the actual existence of real-life stream data contains a model that may change and may have multiple variations in a form called "conceptual drift" [15]. Therefore, real-time mining algorithms need to be able to effectively resist the interference of the concept drift phenomenon; otherwise, the extracted patterns or the trained models will be outdated and inaccurate

(3) Labeling problem: for data mining algorithms, classification, and regression in supervised learning, the dependent variables of their training sets need to be labeled. For traditional data mining algorithms, although the cost of training set labeling is high and time-consuming, it can still accomplish the set goal because it is offline mining; for real-time data mining, because the training data flows into the system incrementally and in real time, the timing of labeling is fleeting, and further research on semisupervised algorithms is needed to form a closed-loop process for training set labeling

3.2. Big Data Distributed Computing Framework. The video big data streaming computing architecture requires several requirements such as low latency, scalability, high reliability, and fast recovery. The computing platform needs to be able to track the execution of each message and deliver the message quickly, and the forwarding delay needs to be below milliseconds; the computing platform needs to support horizontal scaling so that when the computing needs require system expansion [16], it is easy to operate and does not affect the existing data processing operations; the computing platform needs to be highly stable, support the rapid detection of node failure, and be able to resist the avalanche phenomenon.

3.2.1. Storm. The storm is an open-source real-time distributed computing system developed by Twitter, which features scalability, high system fault tolerance, etc. Storm defines the original language for developers to use based on platform's real-time computing and development and has convenient APIs for developers to use, such as real-time online queries, machine learning, and data mining. Storm can be used by configuring ZooKeeper, thereby facilitating the scaling of its distributed clusters. The computational principles are as follows.

$$\mathrm{Ft} = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i \cdot \left(\frac{x-\mu}{\sigma}\right). \tag{1}$$

Storm can process millions of messages in a second, in terms of sports video classification. The development languages for Storm are Clojure and Java, and non-JVM languages can communicate with Storm via stdin/stdout using the JSON protocol. One of the advantages of Storm is that its topology writing and message processing components can be freely used with a variety of programming languages. The most important feature of Storm is that it guarantees that messages are processed, a feature that will benefit the reliability of the entire architecture. By ensuring that messages are processed, Storm ensures that the entire streaming processing layer of the architecture is guaranteed to process messages. The computational framework is shown in Figure 1.

3.2.2. YahooS4. YahooS4 with distributed, pluggable, scalable, and partitioned fault-tolerant features was the first system used to process ad click traffic data. S4 uses the Actor architecture model using a simple programming interface to enable large-scale development. S4 has no central control node, and each working node is of equal status, which makes the system highly usable. S4 has the Actor computing the processing unit PEs in the S4 system pass the completed data through the Actor architecture in the form of events. Each PE is independent of each other and interacts with each

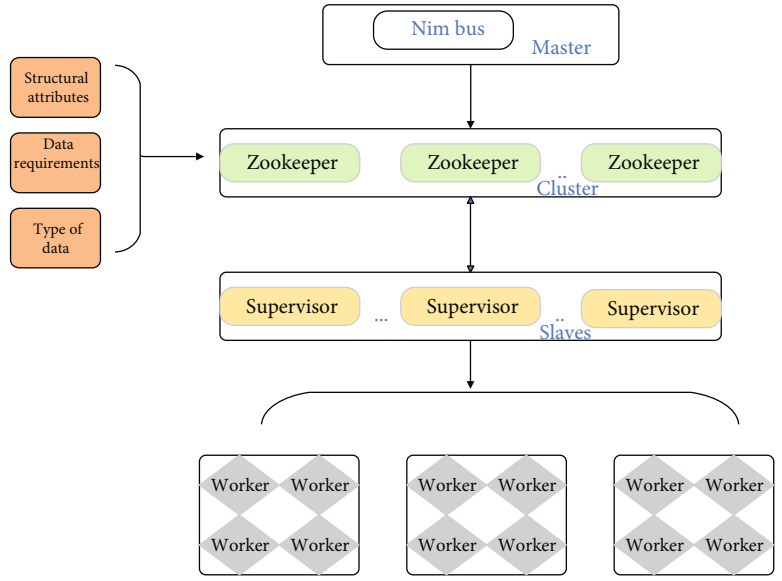

Figure 1: Storm stream data computation framework diagram.

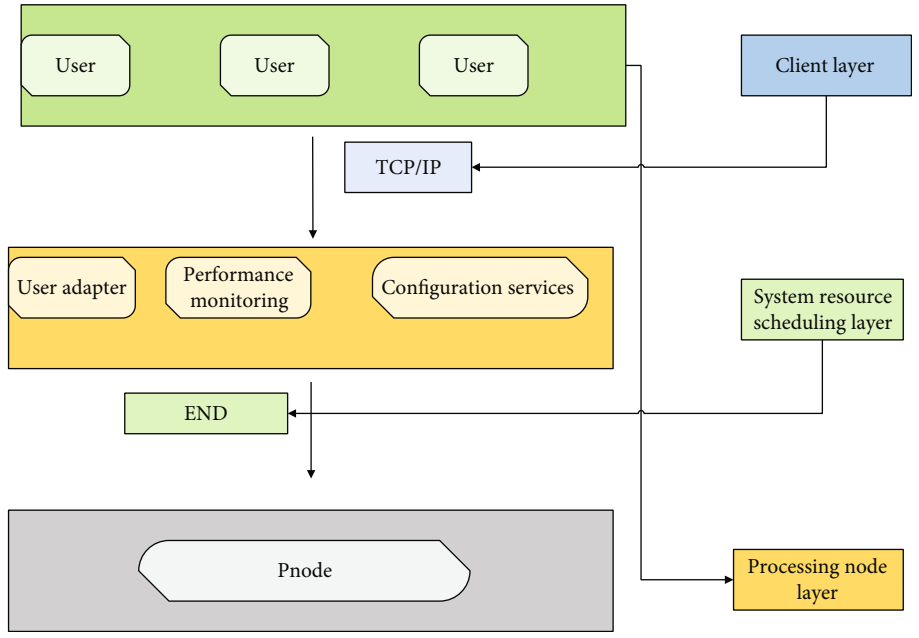

Figure 2: S4 system framework diagram.

other by issuing events and interaction events. The specific principle formula is

$$G(x,y) = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 + \sum_{i=1}^{n} X_i Y_i . \frac{1}{n}\sum_{i=1}^{n} X_i. \quad (2)$$

S4 draws on the MapReduce model and solves the single point of fault tolerance problem. The independent peer-to-peer architecture not only improves the scalability of the cluster but also ensures the efficiency of the system. S4 system supports Java language development and modular packaging. The functional components of S4 are divided into three major categories: Clients, Adapters, and PNodeCluster. Figure 2 shows the framework of the S4 system.

3.3. Deep Learning Based on Video Classification. Deep learning is a recent area of research that has received much attention and plays an important role in machine learning. The history of deep learning development, in terms of timeline, can be seen as three stages of neural network development. The first generation of neural networks [17], which can be called artificial neural networks, was studied starting from the proposal of the M-P model. Figure 3 illustrates the M-P structural model.

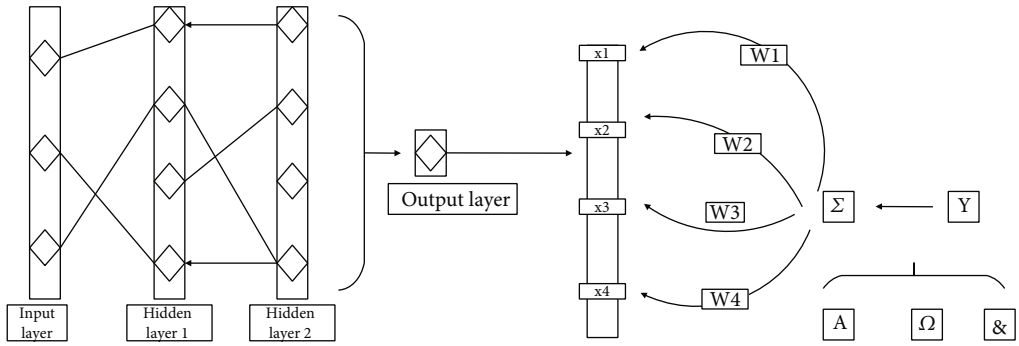

Figure 3: Basic model diagram of neural network.

The backpropagation algorithm for artificial neural networks was proposed, which not only gave hope to the development of machine learning but also opened the statistical model-based machine learning that is still of deep research significance today. The third stage is deep learning, and in the second stage, deep learning enters a period of rapid development by pretraining to quickly train deep belief nets. In the third phase, researchers reduced the top 5 error rate of the ImageNet image classification problem to 15% and deep learning entered an explosive period [18, 19].

Deep learning is a part of machine learning, which can be divided into two research phases, shallow machine learning and deep learning, in terms of research history. Shallow machine learning models and deep learning models have commonalities and important differences. Shallow machine learning models do not use distributed representations and require human extraction of manual features in feature extraction, which is not only time-consuming, heavy, and difficult but also requires accurate domain knowledge of the relevant specialty when researching a particular domain. The model itself can only make further prediction or classification based on the extracted features, and the quality of human-extracted features directly determines the performance of the whole system to a large extent. Deep learning can learn to get the essential features of the whole data set from a smaller number of data samples when performing feature extraction. A nonlinear deep network structure is obtained by training the network to learn and achieve a distributed representation of the input data by representing complex functions with a small number of parameters. The essence of deep learning is to improve the accuracy of classification or prediction by building machine learning models with a number of hidden layers and huge amounts of training data to learn the desired features. The hidden layer in deep learning is equivalent to a linear combination of input features, and the weight between the hidden layer and the input layer is equivalent to the weight of the input features in the linear combination [20], and the capability of the deep learning model grows exponentially with the depth of the network, and its specific principle is formulated as

$$St = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i. \tag{3}$$

Recurrent neural networks (RNNs) are a special kind of neural network structure inspired by the fact that humans rely on experience and memory during cognition. The reason why RNNs are called recurrent neural networks is that RNNs assign not only a memory function to the input of the previous moment but also the input of the next moment is referred to the memory of the previous moment; the current output of a sequence is jointly determined by its input and the output of the previous sequence determines. The specific process is manifested in that the current output is computed by applying it to the output remembered from the previous moment. RNNs differ from CNNs in that in RNNs, the input data have a temporal order of precedence, thus forming a sequence. This is the key difference between RNNs and other neural networks and is the reason why the "loop" can be established [17]. The nodes between the hidden layers, which are unconnected in CNN, become connected in RNN, and the input of the hidden layer contains the output of the input layer and the output of the hidden layer at the previous moment. Its mechanism is illustrated in Figure 4. The hidden layer of the simplest structure of the RNN is expanded in time.

3.4. Algorithmic Framework for Deep Learning in Sports Video Classification. With the basic framework in place, the first and foremost problem to be addressed during the research is the lack of experimental data. There are some publicly available sports video datasets, such as the Sport-1M dataset, which is currently the largest video classification benchmark dataset consisting of 1.1 million sports videos. Each video belongs to one of 487 sports categories, and this dataset does not tag the decomposed actions of a particular category of videos. In the current lack of professional sports datasets, to achieve an automatic description of free gymnastics videos, in this work, a free gymnastics decomposition action dataset containing videos of professional events such as the Olympic Games and National Games is constructed [21], and the description of videos is labeled according to professional commentary. For the discriminative power calculation of video frames, a number of methods have been proposed to solve this problem. One of the commonly used strategies is to use an attention mechanism. Therefore, in the approach of this paper, the attention mechanism is fused into the existing video description network to calculate the

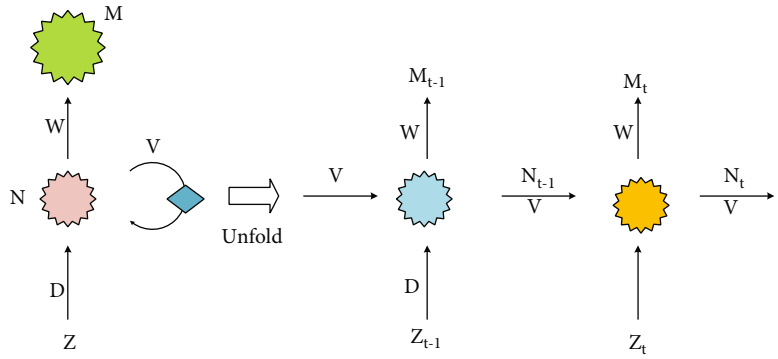

Figure 4: RNN hidden layer unfolding.

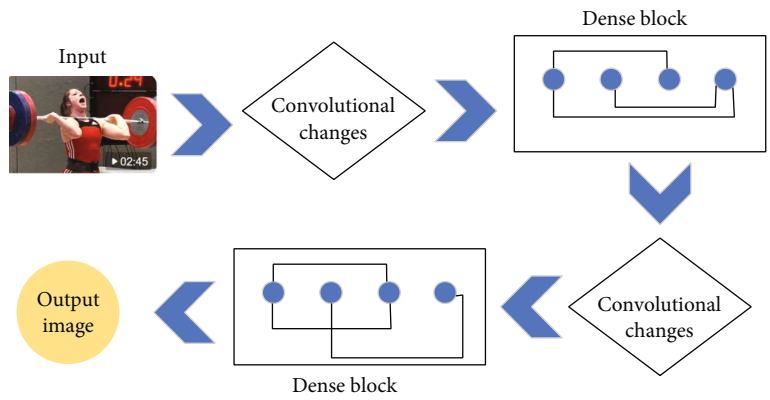

Figure 5: Basic flow chart.

weights between different video frames and improve the accuracy of the automatic video description. As shown in Figure 5, firstly, the free gymnastics decomposition action dataset is constructed; for the video data, feature extraction is performed by the convolutional neural network, and for the labeled text, a dictionary is constructed to extract the corresponding dictionary features; Figure 5 shows basic flow chart, then the training set is input to the video-to-text learning model for training, and the trained text that can accurately recognize the corresponding free gymnastics video is obtained; finally, the tested video features are input to the trained model to get the description of the free gymnastics video [22].

For the automatic description network of sports videos, the input data includes video sequences and text sequences. The video features are first extracted by a convolutional neural network, and then, the text features are extracted using natural language text processing.

3.4.1. Video feature processing. Convolutional neural networks have some degree of invariance to geometric transformations, deformations, and illumination. Trained convolutional neural networks can scan the entire image with a small computational cost and hence are widely used for image feature extraction. And while using its feature extraction, the specific morphology of the features is not considered at all.

3.4.2. Description of text processing. In this paper, we use one-hot vector coding to transform the descriptors of free gymnastics videos into features. The words in the annotated text of sports videos are first counted to construct a dictionary. One-hot computation process: the input is a sentence, and the output is a feature. The calculation is done by first calculating the total number N of all words described in the free gymnastics dataset and then representing each word as a $1 * N$ long vector; in that long vector, there are only two values taken, 0 and 1. There is only one value, 1, in that vector, and the position where this value, 1, is located in the position of the word in the word list at the moment, and the rest of the values are 0. Its schematic diagram is shown in Figure 6.

4. Experimental Results and Analysis

4.1. Experimental Setup. The experiments in this section were done on the operating system application Ubuntu 16.04, and the code used to implement the experiments was based on the Tensorflow 1.6.0 framework, using the language Python 2.7. Its classification model diagram is shown in Figure 7.

The network model was trained on two NVIDIATitan 1080 graphic cards with 11 GB of memory. The input video data is sampled at every 5 frames, the input to the C3D

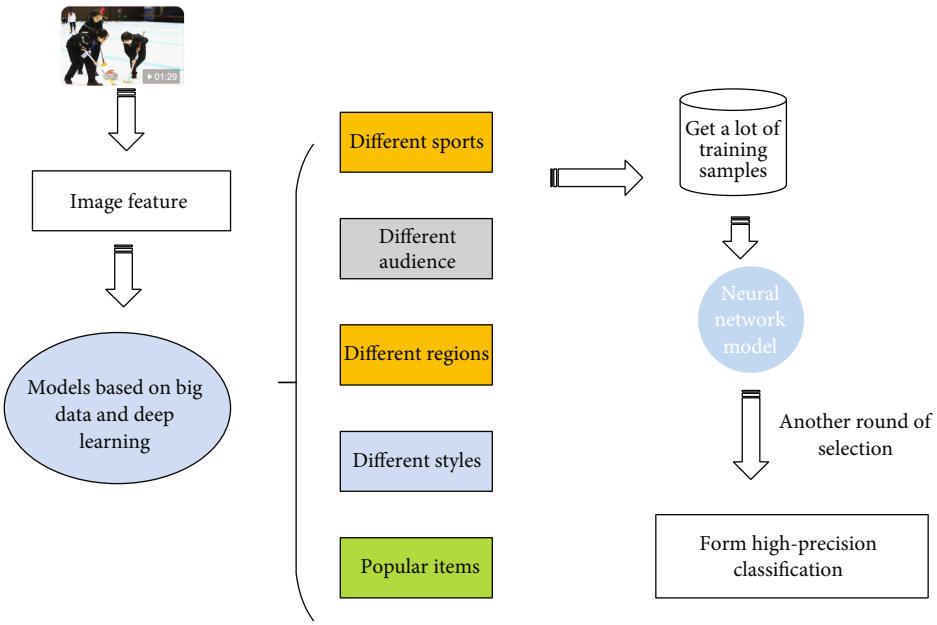

Figure 6: Codec structure diagram.

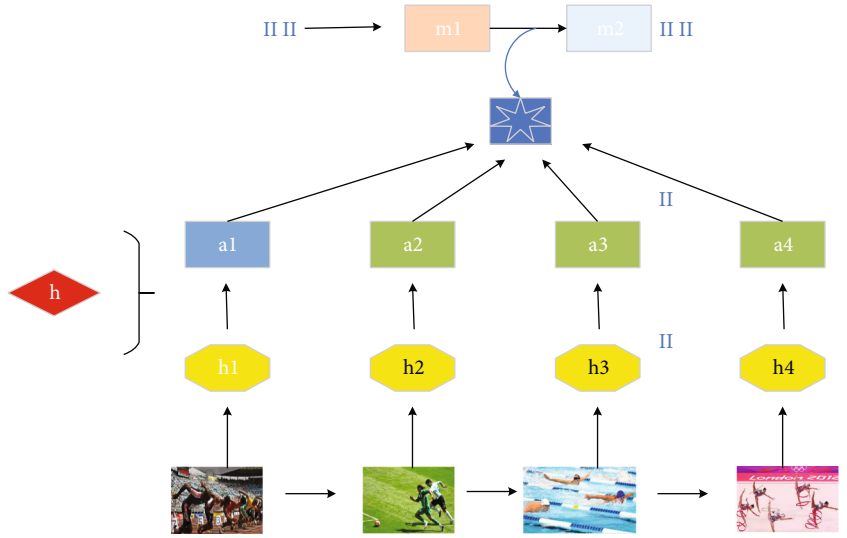

Figure 7: Sports-based video classification process.

feature extraction model is 16 frames long clips with 8 frames of overlap between two consecutive clips, and the fc6 activations of these clips are averaged to obtain 4096-dimensional video descriptors. This section uses LIBSVM classification tool, LIBSVM is a set of libraries for support vector machines, mainly used for classification, which needs to preprocess the fc6 learned video features in the format <label> <index 1>: <value 1> <index 2>: <value 2>.... (Table 2), for the databases that appear for each category.

This chapter is using a classification approach, so the data is described and categorized according to a sporting event professional actions; each video data consists of more than one decomposed action, so each video data is labeled as at least one category, and the categories are replaced with positive integers (starting from 1), for a total of 31 categories [23]. Each category may require more than one word for its description, and the number of occurrences of this one category is shown in Table 2. It can be seen that half of the categories occur very infrequently below 10 times, with a relatively small number of categories occurring particularly high. Figure 8 shows an analysis of the frequency of the 16 categories that occur more than 10 times, with only a few categories having a particularly high number of occurrences.

4.2. Analysis of Experimental Results

4.2.1. Analysis of Multicategorical Evaluation Indicators and Experimental Results. Accuracy is the most common performance metric in classification models and applies to binary classification models, but can also be used for

Table 2: Databases in which the video category appears.

Category	Quantity	Category	Quantity
1	23	1	34
2	33	2	35
3	21	3	32
4	34	4	12
5	32	5	43
6	34	6	47
7	21	7	55
8	21	8	42
9	32	9	33

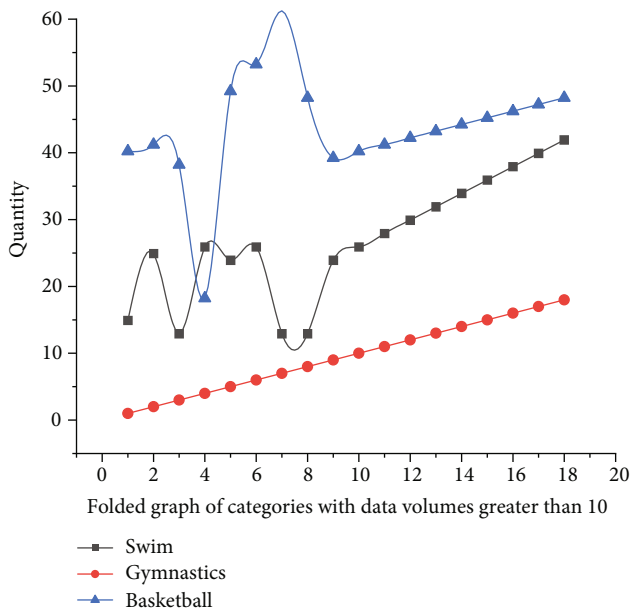

Figure 8: Folded graph of categories with data volumes greater than 10.

multiclassification models. The calculation of accuracy is also relatively simple. Assuming the classification model is g, its formula for calculating accuracy is as follows [24].

$$T(x) = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 + \sum_{i=1}^{n} X_i^2. \quad (4)$$

4.2.2. Example Analysis. The experiments are representative frames of two videos from the test data of the self-built dataset, and due to the page limit, only two examples are given for reference; this example is the same video as the example in Section 3. The experimental results of the automatic description of free gymnastics on the self-built dataset with different models and the multilabel classification method in this chapter are compared. Compared to the original model, S2VT, the model with the attention mechanism is similar in the direction of the "forward" test in blue, but in the video, the improved model is more specific. In the classification problem, this video contains two categories, and the classification results are significantly better. Figure 9 shows a comparison of the differences in the results after classification.

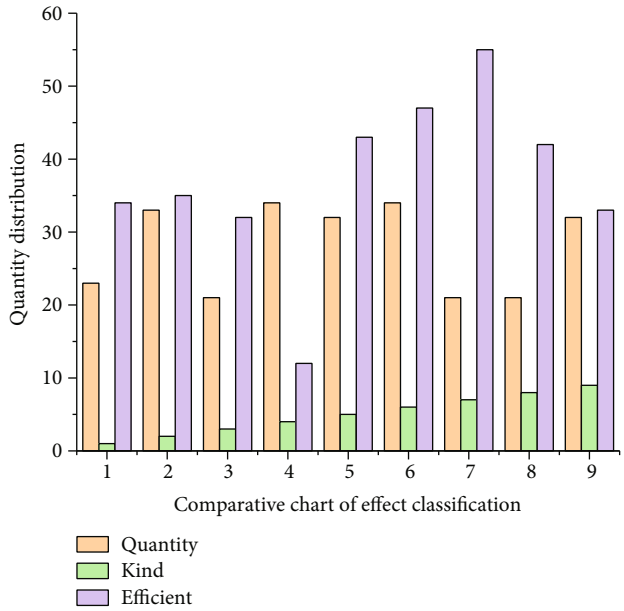

Figure 9: Comparative chart of effect classification.

From the experimental results, it can be seen that the end-to-end deep learning algorithm proposed in this chapter for time-series accurate classification of sports video classification big data has a large improvement in the accuracy performance F compared to other algorithms for classifying sports video classification time-series data [25]. The LSTM-based model improves 14.22% over the BPNN model used on sports video classification and 3.91% over EE, the superior integrated learning algorithm in time series classification. The GRU-based model improves by 14.03% over the BPNN model used for sports video classification and by 3.72% over EE, the superior integrated learning algorithm for time series classification. However, in terms of classification time consumption, both LSTM and GRU have a large speedup over the EE algorithm, but both of them take 6.15 times and 6.03 times more time than the BPNN algorithm used for sports video classification, respectively [26]. The accuracy of these two algorithms can be applied to sports video classification large data temporal classification, but the operational efficiency can only be applied to nonreal-time sports video classification application scenarios.

5. Conclusion

In this paper, the methods and related technical theories of automatic video description and video classification are introduced, and the main framework and implementation steps of the free gymnastics video automatic description method based on long- and short-term memory networks and the sports video automatic classification based on support vector machine multilabel classification are described in detail, and the feasibility of the improved methods in this paper is verified by comparison experiments. In this paper, we first take the automatic video description method as an

entry point and review the current state of research on automatic video description methods, sports video research, and video classification. Then, from deep learning, we introduce its related concepts and development history, focus on describing the derivation and specific architectures of three important types of neural networks, and make structural dissection of typical network models, such as special recurrent neural network LSTM, respectively. Finally, the current state of research on feature extraction of video is analyzed, and the commonly used convolutional neural network feature extraction methods are classified and compared.

In this paper, an automatic description method of free gymnastics based on big data and deep learning classification is proposed. The research object sports video data is not only characterized by the presence of a large number of keyframes, but the definition of its high-level semantic things also is more fixed, and sports videos have certain choreography rules, so the automatic video description problem is transformed into a video classification problem, and the important technical support to promote this problem transformation is the existing video classification model with high accuracy. To preserve the temporal signal of the video, this paper uses a C3D feature extractor and feeds the extracted feature vectors into multiple binary SVM classifiers to accomplish the task of multilabel classification. Video classification based on big data and deep learning can bring users a better experience in the future and can better perform high-precision classification for sports videos. To verify the feasibility of the problem transformation, the classification results are mapped into natural language descriptions. The experimental results show that the classification model of sports video based on big data and deep learning can effectively improve the accuracy of sports video classification rapidly.

Data Availability

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

Conflicts of Interest

All the authors do not have any possible conflicts of interest.

References

[1] J. Wang, Y. Yang, T. Wang, R. S. Sherratt, and J. Zhang, "Big data service architecture: a survey," *Journal of Internet Technology*, vol. 21, no. 2, pp. 393–405, 2020.

[2] C. Qi, "Big data management in the mining industry," *International Journal of Minerals, Metallurgy and Materials*, vol. 27, no. 2, pp. 131–139, 2020.

[3] Z. Lv and L. Qiao, "Analysis of healthcare big data," *Future Generation Computer Systems*, vol. 109, pp. 103–110, 2020.

[4] M. Gao, W. Cai, and R. Liu, "AGTH-Net: attention-based graph convolution-guided third-order hourglass network for sports video classification," *Journal of Healthcare Engineering*, vol. 2021, Article ID 8517161, 10 pages, 2021.

[5] S. Shilo, H. Rossman, and E. Segal, "Axes of a revolution: challenges and promises of big data in healthcare," *Nature Medicine*, vol. 26, no. 1, pp. 29–38, 2020.

[6] M. Zhao, A. Jha, Q. Liu et al., "Faster mean-shift: GPU-accelerated clustering for cosine embedding-based cell segmentation and tracking," *Medical Image Analysis*, vol. 71, article 102048, 2021.

[7] W. Chu, P. S. Ho, and W. Li, "An adaptive machine learning method based on finite element analysis for ultra low-k chip package design," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 11, no. 9, pp. 1435–1441, 2021.

[8] M. Zhao, Q. Liu, A. Jha et al., "VoxelEmbed: 3D instance segmentation and tracking with voxel embedding based deep learning," 2021, https://arxiv.org/abs/2106.11480.

[9] M. A. Amanullah, R. A. A. Habeeb, F. H. Nasaruddin et al., "Deep learning and big data technologies for IoT security," *Computer Communications*, vol. 151, pp. 495–517, 2020.

[10] R. H. Hamilton and W. A. Sodeman, "The questions we ask: opportunities and challenges for using big data analytics to strategically manage human capital resources," *Business Horizons*, vol. 63, no. 1, pp. 85–95, 2020.

[11] S. Khanra, A. Dhir, and M. Mäntymäki, "Big data analytics and enterprises: a bibliometric synthesis of the literature," *Enterprise Information Systems*, vol. 14, no. 6, pp. 737–768, 2020.

[12] H. Tamiminia, B. Salehi, M. Mahdianpari, L. Quackenbush, S. Adeli, and B. Brisco, "Google Earth Engine for geo-big data applications: a meta-analysis and systematic review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 164, pp. 152–170, 2020.

[13] M. Amani, A. Ghorbanian, S. A. Ahmadi et al., "Google earth engine cloud computing platform for remote sensing big data applications: a comprehensive review," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5326–5350, 2020.

[14] J. A. Leopold, B. A. Maron, and J. Loscalzo, "The application of big data to cardiovascular disease: paths to precision medicine," *The Journal of Clinical Investigation*, vol. 130, no. 1, pp. 29–38, 2020.

[15] M. Holmlund, Y. van Vaerenbergh, R. Ciuchita et al., "Customer experience management in the age of big data analytics: a strategic framework," *Journal of Business Research*, vol. 116, pp. 356–365, 2020.

[16] M. N. I. Sarker, B. Yang, Y. Lv, and M. M. Md Enamul, "Climate change adaptation and resilience through big data," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 3, pp. 533–539, 2020.

[17] K. Yu, L. Tan, L. Lin, X. Cheng, Z. Yi, and T. Sato, "Deep-learning-empowered breast cancer auxiliary diagnosis for 5GB remote E-health," *IEEE Wireless Communications*, vol. 28, no. 3, pp. 54–61, 2021.

[18] S. Min, B. Lee, and S. Yoon, "Deep learning in bioinformatics," *Briefings in Bioinformatics*, vol. 18, no. 5, pp. 851–869, 2017.

[19] D. Ravì, C. Wong, F. Deligianni et al., "Deep learning for health informatics," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 4–21, 2017.

[20] X. Hao, G. Zhang, and S. Ma, "Deep learning," *International Journal of Semantic Computing*, vol. 10, no. 3, pp. 417–439, 2016.

[21] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," *IEEE Access*, vol. 9, pp. 44304–44321, 2021.

[22] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning (still) requires rethinking generalization," *Communications of the ACM*, vol. 64, no. 3, pp. 107–115, 2021.

[23] L. Lu, X. Meng, Z. Mao, and G. E. Karniadakis, "DeepXDE: a deep learning library for solving differential equations," *SIAM Review*, vol. 63, no. 1, pp. 208–228, 2021.

[24] A. Echle, N. T. Rindtorff, T. J. Brinker, T. Luedde, A. T. Pearson, and J. N. Kather, "Deep learning in cancer pathology: a new generation of clinical biomarkers," *British Journal of Cancer*, vol. 124, no. 4, pp. 686–696, 2021.

[25] C. P. Hensley, E. M. Lenihan, K. Pratt et al., "Patterns of video-based motion analysis use among sports physical therapists," *Physical Therapy in Sport*, vol. 50, pp. 159–165, 2021.

[26] L. I. U. Ziyu, "Application of college basketball training teaching based on sports video analysis under network multimedia," *Solid State Technology*, vol. 64, no. 1, pp. 170–181, 2021.