Hindawi
Advances in Mathematical Physics
Volume 2022, Article ID 7592180, 10 pages
https://doi.org/10.1155/2022/7592180

Research Article

Construction and Application of Video Big Data Analysis Platform for Smart City Development

Xu Wu[1], **Guifeng Yan**[1], **Xintian Xie**[1], **Yan Bao**[2] and **Wei Zhang**[1]

[1]*Department of BIM RESEARCH, Nantong Institute of Technology, Nantong 226002, China*
[2]*The Key Laboratory of Urban Security and Disaster Engineering of China Ministry of Education, Beijing University of Technology, Beijing 100124, China*

Correspondence should be addressed to Yan Bao; baoy@bjut.edu.cn

Received 13 August 2022; Revised 7 September 2022; Accepted 12 September 2022; Published 21 September 2022

Academic Editor: Miaochao Chen

Copyright © 2022 Xu Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the progress of society and the rapid development of science and technology, daily data volume also shows an exponential upward trend. From the research report of the Internet data center, we can see that the growth rate of data will change from the original slow growth to a sharp rise within 10 years. This shows that the era of big data has arrived, and video data plays an important role in it. Video comes from all aspects of life. As a typical unstructured data, video has the characteristics of large memory, and with the leap of society, this characteristic is becoming increasingly obvious. Taking video data analysis as the starting point, this paper proposes a long-term and short-term memory neural network integrating attention mechanism and verifies it in the experimental data set. The experiment shows that this method has superior performance in model accuracy and work efficiency. Therefore, the application of this method to the construction and application of video big data analysis platform is an important step to promote the development of smart cities.

1. Introduction

In the field of urban public security, the development of cities is progressing day by day, and a large number of population collectives appear, which puts forward higher requirements for public security management level and urban governance ability. However, in the traditional management system, although a large number of cameras and other infrastructure are arranged in the urban area, due to the limitation of technical level, the method of manual real-time observation, playback, and viewing of video data is generally adopted, which only effectively controls the onsite situation in some key areas such as densely populated areas, checkpoints, and urban trunk roads, and it is difficult to find all public safety problems and emergencies at the first time [1]. At the same time, in the video monitoring system, the application of big data technology will replace manual processing of huge data streams, screen out useless data, extract high-value data for visual presentation, help managers quickly find emergencies and security incidents, and reserve sufficient time for subsequent work. During the operation of video monitoring big data system, currently, we are mainly faced with the problems of independent operation of monitoring systems at all levels, which form an information island. A single video monitoring system is difficult to extract enough high-value information from limited video data, and the powerful data processing and logical computing capabilities of big data system have not been brought into full play, resulting in performance redundancy [2]. In view of this, video surveillance systems at all levels and supporting databases need to be integrated. On the one hand, a unified data processing platform is established. Video monitoring systems at all levels submit tasks such as data processing and operation analysis to the data processing platform, as well as upload the captured image data to the data platform. Personnel of all departments directly access the data processing platform to view multidimensional information such as people, places, and objects within the scope of authority, so as to effectively meet the application needs of video monitoring big data. For example, the public security department inquires the image and video of a specific time period in the data processing platform to find

the details of the suspect's facial feature information and wearing feature information. The rail transit operation department grasps the real-time road conditions by consulting the image and video and data reports and checks whether there are problems such as line congestion [3]. On the other hand, considering that the data collected by the video monitoring system is composed of multisource heterogeneous data, taking intelligent transportation and intelligent behavior as examples, and collecting relational data such as the number of violations and individual driving age, as well as time series data such as individual geographical location; there are obvious differences in the characteristics, distribution, and production of different types of data. If a unified processing method is adopted, the processing capacity of the video monitoring big data system will be weakened and reduced the actual utilization of data [4]. Based on this kind of problem, it is necessary to classify videos, which can be classified through AI+ video monitoring technology.

Video analysis technology based on artificial intelligence has been deeply integrated into various industry fields. Video objects include people, vehicles, environment, and objects; relevant management departments need to make corresponding technology choices in combination with industry characteristics and video characteristics, so as to achieve efficient analysis and utilization of video. This behavior plays an important role in urban public safety, network security, emergency disposal, and other fields [5]. The smart city analysis platform needs to include the storage, analysis, classification, sharing, data mining, data early warning, and other functions of video data. The system platform needs to show the panorama of the city, reflect the key characteristics of the city, and have the functions of emergency early warning and intelligent scheduling [6]. On this basis, this paper studies the LSTM video analysis model based on the integrated attention mechanism, aiming to create a video big data analysis platform for smart cities and promote the construction of smart cities.

The innovation contribution of this research is to propose an LSTM neural network model combining attention mechanism. This model inherits the advantages of recurrent neural network and has good advantages in sequence task processing. The LSTM model based on the fusion attention mechanism is tested on the data set. The results show that this method has obvious advantages in model accuracy and work efficiency and has strong advantages in video feature extraction and video classification. Therefore, applying this method to intelligent city construction will greatly promote the development of cities. Video big data technology focuses on helping all kinds of customers to quickly find high-value information from the increasingly massive unstructured video data. Assist customers to improve the efficiency and accuracy of their decisions.

2. The Related Works

The video monitoring equipment that can be seen everywhere in China is the basic hardware equipment of the video big data analysis platform, but the monitoring equipment in most parts of China has the functions of video acquisition, storage and output, and cannot realize intelligent video analysis. At present, the function of monitoring system is too single, and it can only support viewing, which requires manual video classification, feature retrieval, and other tasks. Video monitoring equipment and stored management equipment lack intelligent video analysis function, or the function is very single, and only supports event classification and location classification; complex tasks such as finding and searching video features need to be carried out manually, which not only consumes a lot of human and material resources, but also easily leads to feature leakage; task completion is not up to standard. Therefore, there are great loopholes in video data mining. It is easy to waste data resources [7]. The use of video resources in various places only stays on tasks such as data collection, vehicle search, and person tracking, which are mainly used in public security management and personal security. At present, video analysis is still in the stage of low technology analysis. Only simple intelligent technology or no intelligent method is used for video analysis and classification. Therefore, it is easy to find videos that cannot be found or are too slow to find. It is also easy to find videos with low reliability and too much workload in the search process. More importantly, it is easy to ignore the features we need to find in videos in this work. These problems all point to the low-end of video analysis means and low intelligence [8].

At this stage, in the smart city, the application of video surveillance big data technology effectively solves the problem of low efficiency of data processing and can complete the analysis and processing of huge data streams in a short time. However, due to the complex environment, camera resolution and other factors, some video images taken are ambiguous, and it is difficult for the big data platform to extract sufficient and real data information. As a result, data processing results and decision-making suggestions to users lack practical reference value. For example, in simple and pure scenes, the big data platform can extract real feature information and obtain accurate detection results. In scenes with large traffic and a large number of facilities and obstacles, the detection accuracy of the algorithm will be affected by factors such as light and color, so it is difficult to obtain accurate detection results, and it is impossible to correctly distinguish the behavior of all people and effectively predict potential problems [9]. To solve these problems, we should start from the technical level and take three measures: image enhancement, image restoration, and image super-resolution reconstruction to provide high-quality, high-resolution, and complete detailed video image data for the big data platform. First of all, image enhancement is to use new algorithms such as image defogging, image denoising, and image dark detail enhancement to replace the original image filtering algorithms, so as to improve the image quality and clarity. Secondly, image restoration relies on image degradation knowledge to build a degradation model, and uses Wiener filtering algorithm, wavelet algorithm, and other methods to carry out inverse process processing in the model, gradually restore the image, eliminate the image blur caused by motion and other factors, and obtain a clear

image. Finally, the technical principle of image super-resolution reconstruction technology comes from the signal processing method, using high-frequency components to improve the resolution, and generating a large number of restored images on the basis of low-resolution images, and then screening [10].

There are extensive achievements in the analysis and research of video big data. Mohammadi et al. proposed an image analysis method based on Hadop method. In this method, there is an HDFS module, which can ensure the storage of images. In addition, he also used a distributed framework for image analysis. This method has the advantages of good analysis effect and fast speed, but it is not suitable for dynamic image data processing [11]. Some scholars also studied the storage and search of massive data. Nelson et al. developed a massive image retrieval system based on Hadop technology. He also applied HDFS module for storage, but he added Lucene module to the former to provide retrieval [12]. At present, the mainstream technology of video storage is to compress video frames and pictures, and this technology is also relatively mature. Therefore, the research focus of the above scholars is not on how to compress video, but on how to quickly store and retrieve video images. One solution is to clip the video and store it in the HDFS module in a complete and appropriate size. When video is needed, download it, and use third-party technology for processing [13]. Another method uses the segmentation attribute of HDFS to store the video distributed, and then uses the decoding technology of the module to decode the video, but the subsequent operations need to be considered to splice the cut video [14]. Hadop technology also has strong applications in other video processing and analysis fields. There is still no good way to solve the problem of obtaining the main information of video, but this application is the most needed function in the era of big data. Analyzing video according to video content is an important progress in the field of video analysis. This method can enable people to quickly read a large amount of video data and obtain useful information from it, but it still has the disadvantages of insufficient applicability and low efficiency. On the basis of previous studies, this paper proposes an LSTM video big data analysis method based on attention fusion mechanism. Experiments show that this method has good adaptability in video feature extraction and video classification tasks.

3. Video Big Data Processing Method Based on the Fusion of LSTM and Attention Mechanism

This paper investigates the recognition of video big data analysis platform in promoting the construction of smart city. The subjects of the survey are relevant participants in the construction of smart city, relevant government departments of smart city, citizens, and university research institutions. The questionnaire was distributed online, and the results showed that only 2% of the people said they did not agree. Most people believe that the construction of video big data analysis platform can promote the construction of smart cities. The results are shown in Figure 1.

Figure 2 shows the video big data analysis platform for smart city development. This paper mainly studies how to carry out tasks such as feature extraction and video classification for video big data, and carry out preoperations such as video information acquisition, storage, download, and acquisition based on this goal Figure 2. At the same time, it also solves the difficulties of improving video processing efficiency and storage efficiency. In distributed storage technology, the most important algorithm is load balancing algorithm. The principle of this method is actually a reasonable allocation algorithm of computer resources. Its work is to allocate resources between computer groups and internal hardware of computers and finally maximize the utilization of resources. The algorithm can ensure the reasonable allocation of tasks, improve work efficiency, and balance the load of each hardware of the computer, so as to protect equipment resources. In the task of video classification, the traditional time series model has some shortcomings, such as low efficiency and poor accuracy. Therefore, on the basis of video storage, this paper studies the time series video prediction task and video classification task based on long-term and short-term memory neural network (LSTM) and adds the attention mechanism as the core algorithm in the smart city data analysis platform. After experimental verification, the algorithm shows high correctness and wide applicability. Each time step of the test data set will be executed one at a time. A model will be used to predict the time step, and then, the actual expected value of the next month will be obtained from the test set and provided to the model for the prediction of the next time step. This simulates a real scenario in which new data can be obtained every month and used for the next prediction. This will be simulated by testing the structure of the data set. All predictions on the test data set will be collected and error scores calculated to summarize the model's skills for each prediction time step. The root mean square error (RMSE) is used to punish the larger error, and the score obtained is the same as the unit of the prediction data.

The earliest sequence task is to process text, and video is composed of frames. Therefore, the study of text processing methods has a great inspiration for video frame sequence processing. Language is not only a means to distinguish between others and animals, but also an important way to distinguish different ethnic groups. The first object used in text processing is Latin language, which has a high degree of independence and is easier to be encoded in matrix form. The encoded text can be processed simply by calculating the distance between different units. The research in this field has a long history. Compared with text processing, speech processing is more complex. Speech information not only contains the text information we need, but also contains a lot of noise information we do not need. Therefore, when processing voice text, we must first carry out noise reduction to filter out the impurities in the voice signal and then compare the voice data before and after processing to ensure the integrity of information features. In addition, the most

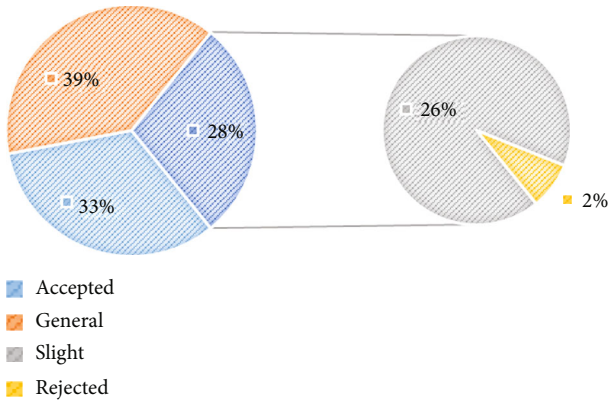

Figure 1: People's recognition of video big data analysis platform.

important link in the process of speech information processing is to distinguish language types, which is a priori condition to ensure the smooth progress of the follow-up work. The core idea of speech data processing lies in logical judgment. The correct logical connection is the key to speech analysis. The core of video data processing is also logical judgment. This paper uses the LSTM network model integrating attention mechanism to classify video events, so as to solve the practical problems encountered during the construction and management of smart cities and promote the healthy development of smart cities.

3.1. Basis of Recurrent Neural Network. The basic idea of recurrent neural network (RNN) is to process the data with logical relationship. Its structure has high repeatability. Its processing object is to analyze the logical relationship between adjacent units in the data. The weight of this model has the advantages of popularity and collinearity. RNN will have a multilayer network structure in the sequence tasks at multiple time points. The number of sequences is consistent with the number of layers of the network structure, with a high degree of correspondence. Its structure is shown in Figure 3. In Figure 3, the structure correspondence and sequence characteristics in RNN are introduced in detail. S represents the hidden layer, which has the function of data storage and memory; U represents the weight to be added during the transmission of input data to the hidden layer; O is the output value but not the final output; V is the weight matrix through which the data is transmitted from the hidden layer to the output layer; L is the loss function of the model; and Y is the final result of the model output [15]. Figure 3 shows the structure of the recurrent neural network.

By analyzing the above figure, the input at time t in the expanded structure diagram can be expressed as x_t, and the hidden layer is s_t at this time. It can be seen from the figure that the data of the hidden layer should not only be combined with the input at this time, but also consider the value of the hidden layer at the previous time. The above structure diagram clearly shows the forward propagation theory of RNN, according to which tasks such as prediction at a certain time can be carried out.

$$\widehat{y}_t = \sigma(o_t), \tag{1}$$

$$o_t = g(V \cdot s_t + c), \tag{2}$$

$$s_t = f(U \cdot X_t + W \cdot s_{t-1} + b), \tag{3}$$

where σ and f in the above formula are activation functions. The two common activation functions are $soft$ max activation function and tanh activation function, respectively. B in the formula means the offset of the function, and $\widehat{y}_t$ represents the final output of the model, that is, the predicted value. In addition, the RNN model parameters are mainly determined by back propagation. The gradient descent method is used to iterate the model, and finally, the parameters with the highest accuracy and the best model performance are calculated. The direction of gradient descent is controlled by the loss function, and its formula is as follows:

$$L = \sum_{t=1}^{T} L_t. \tag{4}$$

The determination of model performance is to determine the weight matrix of each stage in the model and other parameters in the formula. The gradient calculation formula is as follows:

$$\frac{\partial L}{\partial c} = \sum_{t=1}^{T} \frac{\partial L_t}{\partial c} = \sum_{t=1}^{T} \frac{\partial L_t}{\partial o_t} \cdot \frac{\partial o_t}{\partial c} = \sum_{t=1}^{T} \widehat{y}_t - y_t, \tag{5}$$

$$\frac{\partial L}{\partial V} = \sum_{t=1}^{T} \frac{\partial L_t}{\partial V} = \sum_{t=1}^{T} \frac{\partial L_t}{\partial o_t} \cdot \frac{\partial o_t}{\partial V} = \sum_{t=1}^{T} \left(\widehat{y}_t - y_t\right)^T. \tag{6}$$

The determination of the above parameters basically determines the network structure of RNN.

3.2. Basis of Long and Short-Term Memory. RNN network model is mainly born to understand the task of time series prediction, but this model has a well-known drawback. When the network structure of the model gradually increases, the gradient will disappear. The gradient vanishing problem is mainly due to the high learning ability of the hidden layer, which leads to excessive learning, resulting in the smooth function curve, and finally leads to the failure of the prediction and classification task. RNN model has many variants. LSTM (long- and short-term memory) is one of them. It can process the data of time series and effectively avoid the problem of gradient disappearance. This model mainly includes forgetting gate and input-output gate. In RNN, the hidden layer is the main structure that exists at any time. Its state depends on the input information at that time and the hidden information at the previous time, and the hidden information at this time affects the hidden information at the next moment [16]. Compared with the simple iterative problem of RNN, LSTM designs a more complex structure called forgetting gate, so it can avoid the gradient disappearance problem.

In Figure 4 above, the input at time t and the hidden information at the previous time enter the activation

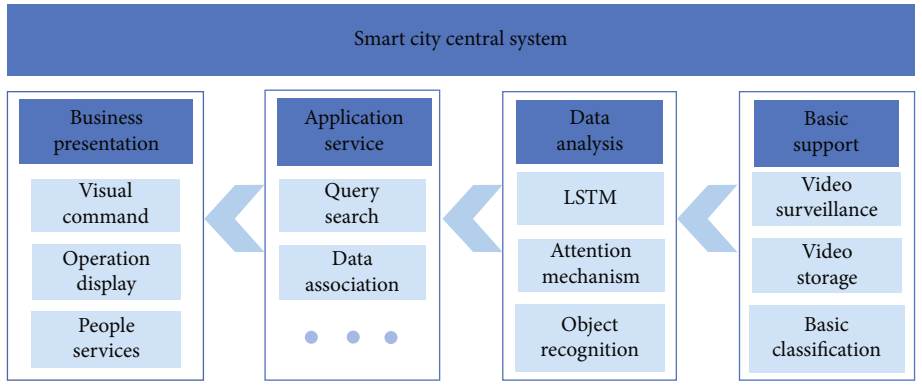

Figure 2: Video big data analysis platform for smart city development.

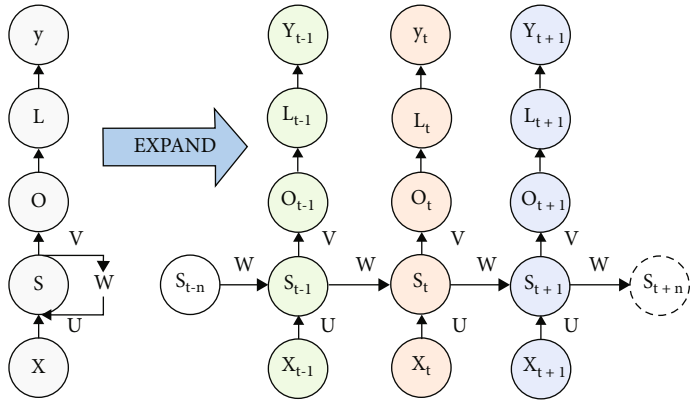

Figure 3: Structure of recurrent neural network.

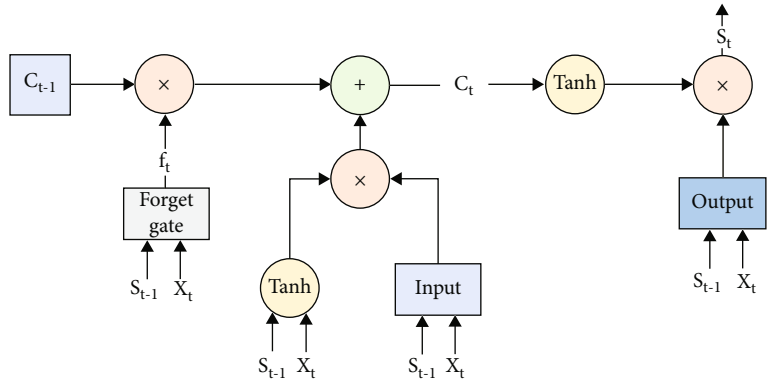

Figure 4: Structure diagram of LSTM model.

function at the same time and then get the output for the next step. This process is the work of the forgetting gate. The meaning of the output value f_t represents the probability that the information at the previous time is deleted, and its formula is as follows:

$$f_t = \sigma\left(W_f s_{t-1} + U_f x_t + b_f\right). \tag{7}$$

Input the T, time information, and the last time hidden state information into the tanh activation function, and then, multiply its output with the t, time information, and the last time hidden state information to obtain this part of the output at

$$a_t = \tanh\left(W_a s_{t-1} + U_a x_t + b_a\right). \tag{8}$$

Then, input the a_t and computed CT into the tanh activation function, and multiply it by the output to obtain the hidden information at that time:

$$o_t = \tanh\left(W_o s_{t-1} + U_o x_t + b_o\right), \tag{9}$$

$$h_t = o_t \odot \tanh c_t. \tag{10}$$

The forward calculation formula of LSTM model can be obtained by accumulating the above formulas:

$$\widehat{y}_t = \sigma(\mathrm{V}\mathrm{h}_t + c). \tag{11}$$

Compared with RNN, LSTM has the characteristics of complex structure, but the structure is clear and easy to understand, has strong adaptability, and can also solve the problem of RNN gradient disappearance. The load balancing algorithm introduced earlier in this paper also uses this model, which predicts the load of nodes, so as to make dynamic adjustment, form a closed-loop control system to automatically allocate tasks, and improve the efficiency and stability of tasks [17].

3.3. Attention Mechanism. In the actual monitoring system, there are generally 20 cameras working in a cluster. The frame rate is calculated according to 25 seconds per second, and the resolution is calculated according to a single 2 million. Based on this data, the monitoring system needs to process such huge data per second, which is obviously a task that a single device cannot complete. Similarly, the amount of video and picture data stored in a day is also massive, which is also a great test for the storage and analysis system. Therefore, a distributed system is needed for data storage and analysis. Based on the above distributed data storage design, the video classification steps can realize multidirectional parallel operation. On this basis, in order to improve the accuracy of video classification, this paper integrates attention mechanism on the basis of LSTM to improve the accuracy of video classification. The introduction of attention mechanism can reduce the computational burden of processing multidimensional data input, select the data with a high degree of coincidence with the target information through structured means for detailed processing, and only pay attention to the part of the target concerned. This method can enable the algorithm system to focus on processing data objects that overlap with the target features and can greatly improve work efficiency and task quality.

The attention mechanism is essentially an automatic weighting scheme. In the traditional model, the decoder can only obtain the fixed hidden vectors of a certain layer of the encoder (generally using the last layer) as input each time it predicts. From the perspective of weighting, it is actually a simple global average of the hidden vectors of all layers of the encoder [18]. With the introduction of attention mechanism, each time step model will be weighted sum all the hidden vectors of the encoder according to the automatically calculated weight probability and get a new context vector. Because the weight of the hidden layer of each time step is different, the input context received by each time step decoder is no longer fixed. So that each time step decoder can focus on processing the most relevant information in the original module and the current output [19].

After the introduction of attention mechanism, the original encoding and decoding work has become relatively complex, in which the interval has also changed from a single value to a group of vectors. The output of the encoder also becomes a multidimensional vector, from which the decoder obtains a vector with high reliability for calculation. The calculation formula is as follows:

The weight a_t^i of the attention mechanism is calculated by the hidden unit of the encoder and decoder. Note that the mechanism adopts a quantitative calculation of the improvement effect. Let us first define that in the example above, the query item is the hidden state of the decoder, and the key item and the value item are both the hidden state of the encoder. In the sense, note that the input of the mechanism includes the query item and the key item and value item corresponding to the query item, wherein the value item is a group item that needs to be weighted average. In the weighted average, the weight of the value item is used to calculate the query item and the key item corresponding to the value item.

$$a_t^i = \frac{\exp\left(\mathrm{score}(s_t, h_i)\right)}{\sum_{j=1}^{n} \exp\left(\mathrm{score}\left(s_t, h_j\right)\right)}. \tag{12}$$

In the above formula, the expression of the fractional function score is variable, and there are two common ones below.

$$\mathrm{score\_addition}\left(s_t, h_j\right) = V^T \tanh\left(W_a[s_t ; h_i]\right), \tag{13}$$

$$\mathrm{score\_multiplication}(s_t, h_i) = s_t^T W_a h_i. \tag{14}$$

Combined with the attention weight, the front and rear semantic vector c_t is calculated according to the front and rear sequence vectors.

$$c_t = \sum a_i^t * h_i. \tag{15}$$

The hidden value h_t and semantic vector of the decoder can get the final weight through tanh activation function.

$$a_t = f(c_t, h_t) = \tanh\left(W_a[c_t ; h_t]\right). \tag{16}$$

Input the attention weight to the next unit through the following formula.

$$y_t = f(h_t, y_t, c_t), \tag{17}$$

where W and V mean the weight matrix, a represents the attention weight value, f is the activation function, and c_t is the semantic vector. The essence of attention mechanism is to add the target elements to the network, so that the model will pay attention to the sequence related to the target elements in the operation process, so as to control the resource allocation and finally improve the work efficiency [20].

It is difficult to distinguish the correlation between input and target only relying on the encoding and decoding module of LSTM. Therefore, this model introduces a temporal attention mechanism between the encoder and decoder corresponding to each video feature, automatically learns the

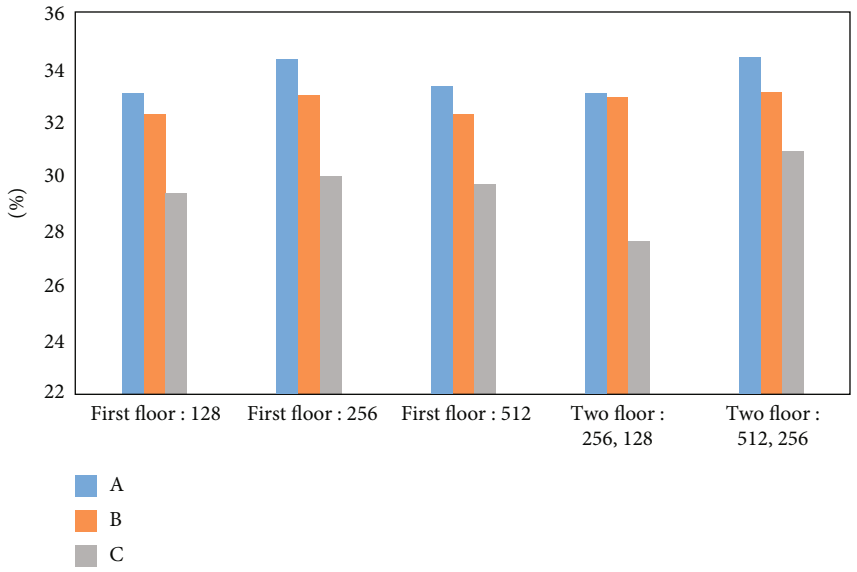

Figure 5: Experimental results of LSTM under different parameter states.

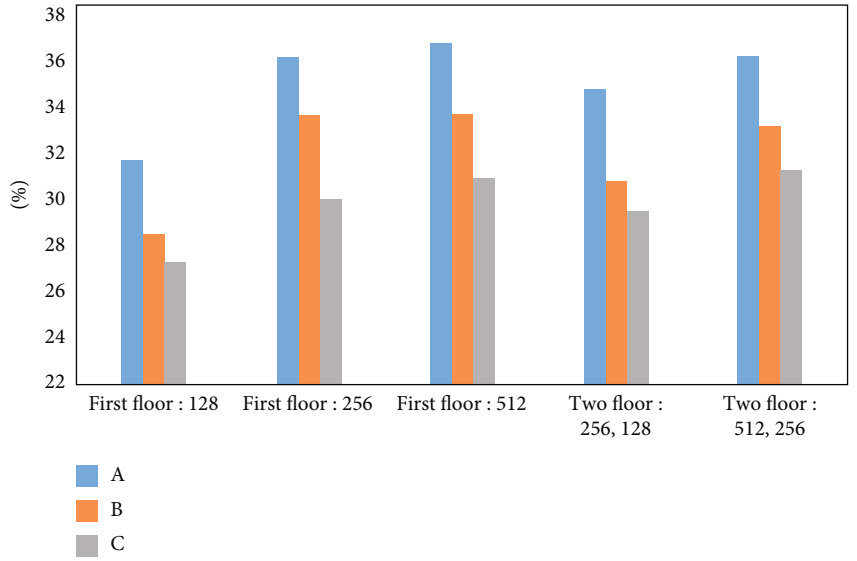

Figure 6: Experimental results of GRU under different parameter states.

correlation between the decoder's predicted output and the encoder's hidden vector, and is used to simulate the attention allocation of different video features.

4. Analysis of Simulation Results

The data processed by RNN series models are time series data containing time information, so increasing the width of the network has a better effect on improving the performance of the network model. Considering that different data have different characteristic dimensions, this paper uses comparative experiments to illustrate the specific situation. Both LSTM and Gru networks are variants of RNN, but LSTM has one more gate unit than Gru, which can control the direction of information flow, so it has structural and functional advantages. At the same time, in order to verify the difference in accuracy between the two variants of the network, this paper sets up a comparative experiment: In the experiment, each video is set to take 50 frames for calculation, and the time interval is automatically selected according to the time length of the video. In this paper, the LSTM and Gru networks in the cyclic neural network are compared, and three experiments are carried out with different structural parameters. The accuracy results are shown in Figure 5.

As can be seen from Figures 5 and 6, LSTM and Gru networks have high similarity in model accuracy, but LSTM has obvious advantages in structure and function. It can be seen from the figure that the accuracy of the LSTM model still needs to be improved. In order to improve the accuracy of the model, this paper adds an attention mechanism. In order to more objectively verify the performance of the LSTM

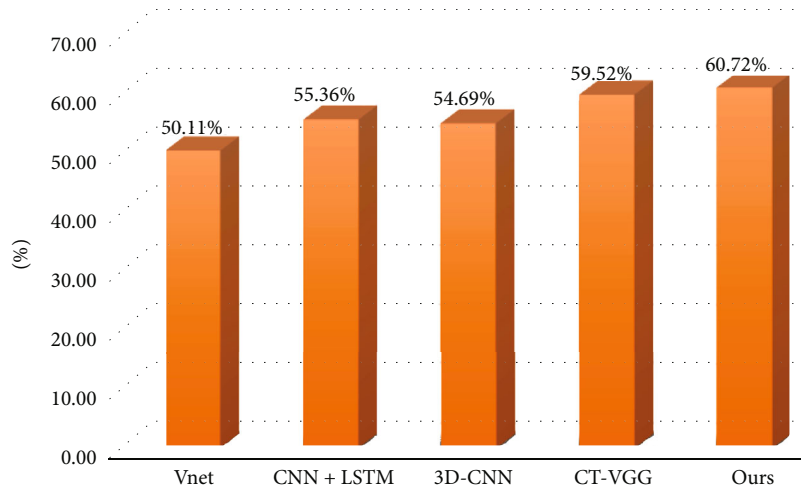

Figure 7: Comparison of recognition results of different network models applied to BAUM-1s dataset.

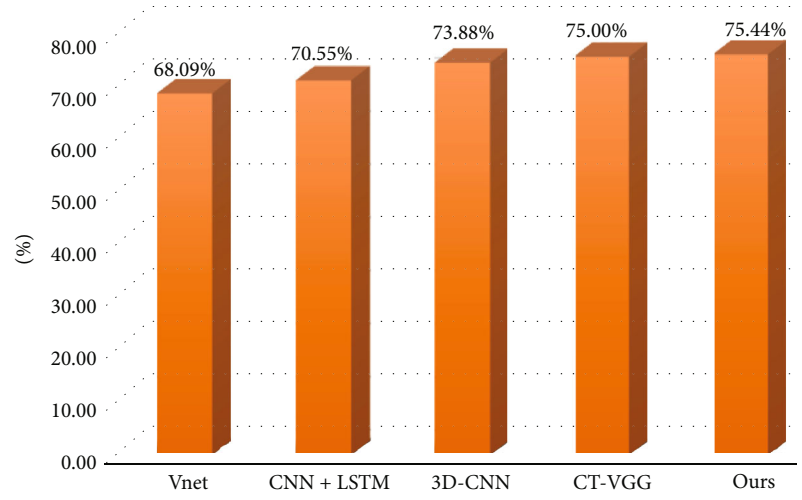

Figure 8: Comparison of recognition results of different network models applied to RML dataset.

network model after adding an attention mechanism, this paper uses baum-1s and RML data sets to verify the performance of the model.

In order to verify the effectiveness of the LSTM network integrating attention mechanism in video feature extraction, a comparative experiment is set up in this paper. Figures 7 and 8 show the comparison between this method and other methods. By analyzing the figure, the LSTM model with attention mechanism in this paper achieved an average accuracy of 60.72% and 75.44%, respectively, in the comparative experiment. This paper sets up four groups of comparative experiments. The first group uses deep CNN (VNET) to extract video dynamic features. The second group adopts CNN+LSTM method to extract video dynamic features. The third group uses 3d-cnn to extract video features. The fourth group used CT-VGG for dynamic video feature extraction. Through the experimental data, we can see that the LSTM model with attention mechanism has the highest accuracy, so it shows that this model can effectively carry out the task of video dynamic feature extraction.

In the process of determining the network model, the performance accuracy of the model will change with the length of the time step. It is verified by experiments that there is a positive correlation between the increase of the time step and the performance of the model. The application of multiple time steps can improve the generalization ability of the model, because in the operation of multiple time steps, the model can automatically eliminate the influence of contingency and maintain the reliable stability of the model. In this paper, several groups of multitime step comparative experiments are set up, and the false alarm rate is used as the model evaluation parameter. The lower the false alarm rate is, the more reliable the stability of the model is. The results are shown in Figure 9. With the increase of time step, the false alarm rate gradually decreases, and the LSTM model integrating attention mechanism proposed in this paper always has the lowest false alarm rate.

In order to verify the effect of the method proposed in this paper from multiple dimensions, this paper also makes a statistical comparison of the running time of the model. The results show that the running time of the

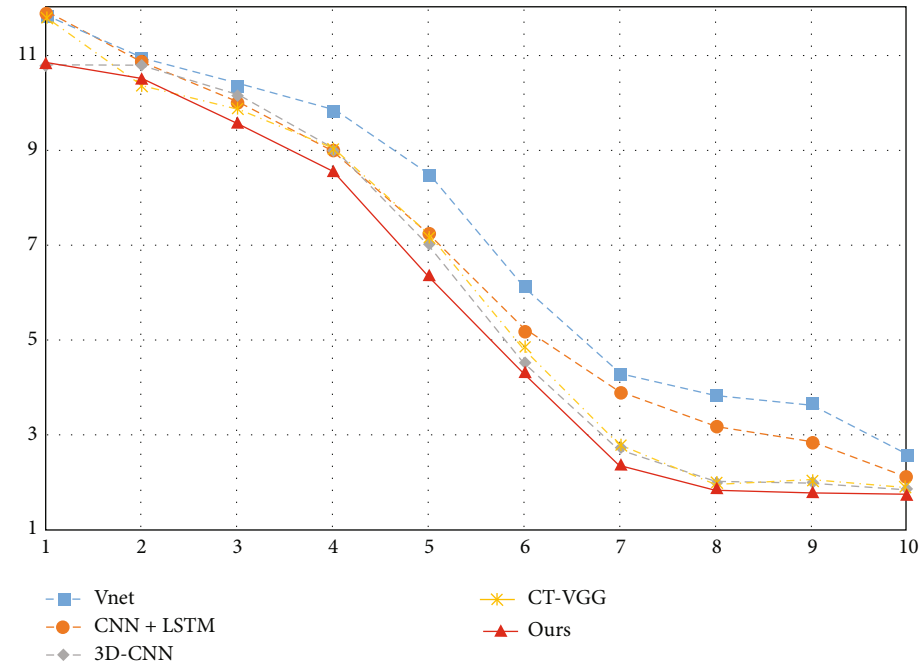

Figure 9: Comparison of false alarm rates of various models with different time step lengths.

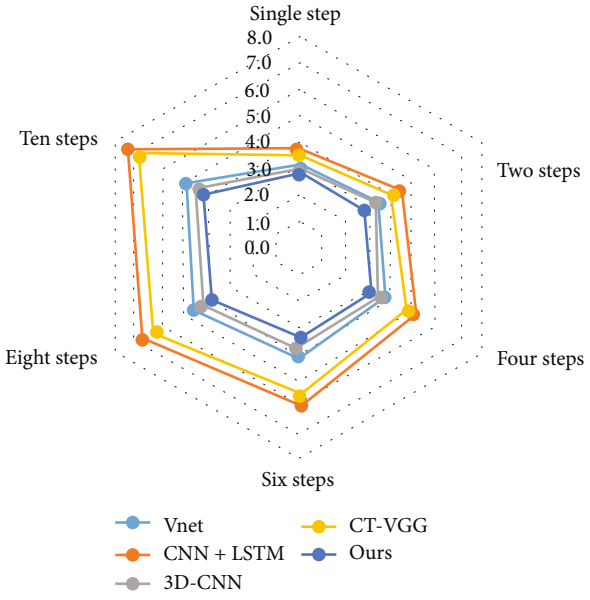

Figure 10: Comparison of running time of different models.

LSTM model integrating attention mechanism proposed in this paper is significantly lower than that of other models, because the LSTM model starts from two ends, and the computational efficiency is higher than that of other models. Moreover, it can be seen from Figure 10 that with the increase of time step, the time comparison between several models becomes more obvious, and the gap also gradually increases.

Through three groups of comparative experiments, this paper studies the performance of the model from three levels: model accuracy, false alarm rate, and running time. Finally, it shows that the LSTM model with attention mechanism has strong performance and is suitable for video big data analysis.

5. Summary and Outlook

In this paper, an LSTM neural network model combined with attention mechanism is proposed. This model inherits the advantages of recurrent neural network and has good advantages in sequence task processing. At the same time, the model can well solve the gradient disappearance problem in the recurrent neural network. The LSTM model proposed in this paper is tested on the data set. Compared with other RNN variants, LSTM has a more flexible model structure. Finally, the attention mechanism is integrated into the LSTM network to form the core method of this paper. The model with attention mechanism can carry out adaptive attention classification according to different types of videos, which greatly improves the efficiency of the model. The results show that this method has obvious advantages in model accuracy and work efficiency and has strong advantages in video feature extraction and video classification. Applying this method to the construction of intelligent city will greatly promote the development of the city. However, the study still has some limitations. Video analysis and feature extraction models have room for improvement in both structure and performance and are difficult to meet the work requirements in the big data environment. Therefore, further analysis is needed in future research and development.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Nantong Nature Fund Project YEAR2020 Nantong Science and Technology Bureau (No. JC2020123) and the Provincial Key Platform Cultivation Project of Nantong Institute of Technology (No. XQPT202102).

References

[1] S. Mujeeb, N. Javaid, M. Ilahi, Z. Wadud, F. Ishmanov, and M. Afzal, "Deep long short-term memory: a new price and load forecasting scheme for big data in smart cities," *Sustainability*, vol. 11, no. 4, p. 987, 2019.

[2] Z. Lei, "Development and deep application of intelligent video analysis technology," *Smart Cities Computer Knowledge and Technology*, vol. 16, no. 35, pp. 251-252, 2020.

[3] H. Li, T. Xiezhang, C. Yang, L. Deng, and P. Yi, "Secure video surveillance framework in smart city," *Sensors*, vol. 21, no. 13, p. 4419, 2021.

[4] H. Hu, G. Zhang, W. Gao, and M. Wang, "Big data analytics for MOOC video watching behavior based on spark," *Neural Computing and Applications*, vol. 32, no. 11, pp. 6481–6489, 2020.

[5] D. S. Jat, L. C. Bishnoi, and S. Nambahu, "An intelligent wireless QoS technology for big data video delivery in WLAN," *International Journal of Ambient Computing and Intelligence (IJACI)*, vol. 9, no. 4, pp. 1–14, 2018.

[6] T. C. Phan, A. C. Phan, H. P. Cao, and T. N. Trieu, "Content-based video big data retrieval with extensive features and deep learning," *Applied Sciences*, vol. 12, no. 13, p. 6753, 2022.

[7] A. Yuloskov, M. R. Bahrami, M. Mazzara, and I. Kotorov, "Smart cities in russia: Current situation and insights for future development," *Future Internet*, vol. 13, no. 10, p. 252, 2021.

[8] G. Sreenu and S. Durai, "Intelligent video surveillance: a review through deep learning techniques for crowd analysis," *Journal of Big Data*, vol. 6, no. 1, pp. 1–27, 2019.

[9] D. Byler, "Producing "Enemy Intelligence": Information Infrastructure and the Smart City in Northwest China," *Information & Culture*, vol. 57, no. 2, pp. 197–216, 2022.

[10] A. Alam and Y. K. Lee, "TORNADO: intermediate results orchestration based service-oriented data curation framework for intelligent video big data analytics in the cloud," *Sensors*, vol. 20, no. 12, p. 3581, 2020.

[11] M. Mohammadi and A. Al-Fuqaha, "Enabling cognitive smart cities using big data and machine learning: approaches and challenges," *IEEE Communications Magazine*, vol. 56, no. 2, pp. 94–101, 2018.

[12] A. Nelson and O. Neguriță, "Big data-driven smart cities," *Geopolitics, History, and International Relations*, vol. 12, no. 2, pp. 37–43, 2020.

[13] K. Wade, J. Vrbka, N. A. Zhuravleva, and V. Machova, "Sustainable governance networks and urban internet of things systems in big data-driven smart cities," *Geopolitics, History, and International Relations*, vol. 13, no. 1, pp. 64–74, 2021.

[14] R. Dubman, "The digital governance of data-driven smart cities: sustainable urban development, big data management, and the cognitive internet of things," *Geopolitics, History, and International Relations*, vol. 11, no. 2, pp. 34–40, 2019.

[15] H. Sixi, "Research on video model construction and motion recognition strategy based on RNN," *Journal of Jiamusi University (NATURAL SCIENCE EDITION)*, vol. 37, no. 5, pp. 752–754, 2019.

[16] L. Meng and R. Li, "An attention-enhanced multi-scale and dual sign language recognition network based on a graph convolution network," *Sensors*, vol. 21, no. 4, p. 1120, 2021.

[17] M. Chen and S. Jie, "Motion prediction of multimodal LSTM based on self attention," *Computer engineering and design*, vol. 43, no. 4, pp. 1083–1088, 2022.

[18] Y. Mo, Q. Wu, X. Li, and B. Huang, "Remaining useful life estimation via transformer encoder enhanced by a gated convolutional unit," *Journal of Intelligent Manufacturing*, vol. 32, no. 7, pp. 1997–2006, 2021.

[19] D. Adams, A. Novak, T. Kliestik, and A. M. Potcovaru, "Sensor-based big data applications and environmentally sustainable urban development in internet of things-enabled smart cities," *Geopolitics, History, and International Relations*, vol. 12, no. 2, pp. 108–118, 2021.

[20] M. Montagnuolo, P. Platter, A. Bosca, N. Bidotti, and A. Messina, "Realtime semantic enrichment of video streams in the age of big data," *SMPTE Motion Imaging Journal*, vol. 128, no. 1, pp. 1–8, 2019.