Monitoring Method of Transmission Line Breaking Prevention Based on Deep Learning

JIANG Yan[1], LI Qiang[2], WANG Guanyao[1,*], WANG Ben[1], DENG Wei[1]

[1]Beijing Fibrlink Communications Co., Ltd, Beijing 100071, China;
[2]State Grid Information & Telecommunication Group Co., Ltd, Beijing 102211, China

Abstract. With the rapid development of the national economy, the national power consumption level continues to increase, which puts forward higher requirements on the power supply guarantee capacity of the power grid system. The distribution range of the transmission line is wide and densely, most lines are exposed to the unguarded field without any shielding or protective measures, which are vulnerable to man-made destruction or natural disasters. Therefore, it is very important for the early monitoring and prevention of the external force breaking of the transmission lines. The method for preventing external breakage of transmission lines based on deep learning proposed in this paper utilizes the video data collected by the cameras erected on the transmission line roads to perform feature extraction and learning through 3D CNN and LSTM networks, and obtains a monitoring model for external breakage prevention of transmission lines. The model was tested on public data sets and verified that it has a good performance in the field of transmission lines against external damage. The method in this paper makes full use of the existing video acquisition equipment, and the process does not require human intervention, which greatly reduces the cost of line monitoring and the hidden dangers of accidents.

1 Introduction

With the rapid development of the national economy and the continuous deepening of the industrialization process, people's demand for electricity continues to increase, which also brings huge challenges to the power supply system. Especially with the rise of new energy power generation in recent years, the application of many large power generation equipment has greatly increased the difficulty and risk of transmission line operation and maintenance. The emergence of new energy generation methods will inevitably require the support of new types of facilities. The construction of these facilities also poses a certain degree of threat to the safety of transmission lines. Therefore, as an important part of ensuring the smooth and efficient operation of State Grid system, and even as an important guarantee for country's economic development, it is very necessary to avoid damage to transmission lines by external forces to the greatest extent. The application scenarios of transmission lines are special, which the distribution range of the transmission line is wide and densely, most lines are exposed to the unguarded field without any shielding or protective measures, which are vulnerable to natural disasters such as thunderstorm. In addition to natural factors, the proportion of line damage, power outages, and even transmission accidents caused by human factors is also increasing year by year. As a result, the regulatory authorities are paying more and more attention to the destruction of lines by external forces. Compared with the uncertainty and force majeure of natural factors, man-made transmission line damage is easier to avoid. If the possible time of occurrence can be monitored in advance before man-made damage, early warning and early intervention, the impact of man-made transmission line external damage can be minimized.

At present, domestic and foreign researchers have proposed a series of monitoring devices for preventing external damage of transmission lines to solve the problem of frequent damage to transmission lines. To resolve the problem of external damage, foreign power companies use aircraft inspection methods to conduct safety inspections on transmission lines. Golightly [1] proposed a method based on helicopter video surveillance, which uses cameras to automatically align the targets of interest, so as to realize the monitoring of transmission lines. This method uses corner point monitoring and matching methods to make the camera's attention always stable during the helicopter flight. The disadvantages of the method of using manned helicopters for video monitoring of transmission lines are the high risk of accidents and the high cost of monitoring. Therefore, it is not suitable for the densely distributed and wide-covered power grids of our country. With the continuous maturity and upgrading of aircraft technology, the method of using unmanned aerial vehicles to conduct inspections on transmission lines has gradually attracted attention. Based on this method, Larrauri [2] proposed an automatic monitoring system for transmission lines. The system analyzes and

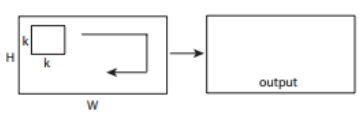
* Corresponding author: wangguanyao@sgitg.sgcc.com.cn

processes line monitoring images collected by infrared cameras, identifies poor-quality conductive locations and hot spots in lines, transformers and even substations, and can calculate the distance between the transmission line and nearby vegetation and buildings to ensure The line is not damaged. Compared with helicopters, UAV-based transmission line monitoring systems effectively avoid accidents such as casualties during inspections, and save costs to a large extent, but this type of method still has drawbacks. The UAV video surveillance system uses a round-robin monitoring method. The inspection cycle is long. Real-time long-term monitoring of moving targets in a fixed area cannot be achieved. It is not suitable for monitoring abnormal behaviors in transmission lines, such as illegal construction. Therefore, Fang [3] used laser scanning to obtain multi-point spatial position information on suspicious targets around the wire, and processed it into a complete gray-scale image, and proposed a suspicious target monitoring in the transmission line corridor based on image processing technology. Methods. Image monitoring methods are mostly based on cameras that have been installed on transmission lines. The advantages are low equipment costs and strong real-time performance. However, the current monitoring methods using cameras are mostly based on images and lack the measurement of the temporal characteristics of the video. It is often only possible to detect whether the construction equipment enters the monitoring area, or to identify the abnormal line after the accident occurs. It is impossible to make accurate judgments on external breaches, early warning and early protection. Therefore, this paper proposes a monitoring method based on 3DCNN+LSTM, which fully considers the timing correlation between images, extracts deep spatiotemporal features through 3D CNN network, and then learns the features through the LSTM network. Experimental results show that this method can effectively monitor abnormal behaviors to achieve accurate judgment and timely warning.

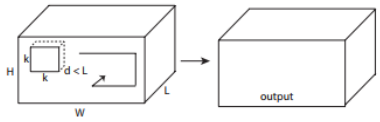
2 Deep Learning Network

Compared with 2D CNN commonly used for feature extraction, 3D CNN considers information in the time dimension in the process of convolution and pooling. 2D CNN only performs spatial convolution and pooling operations on a single image, while 3D CNN performs convolution and pooling operations on continuous frame sequence images, so it can more comprehensively characterize and describe video content efficiently [4-5].

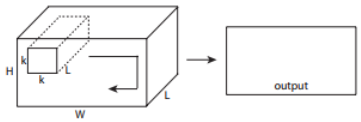
Fig.1 illustrates the structural differences between 3D CNN and 2D CNN. Figure 1(a) describes the 2D convolution process of a single image, Figure 1(b) describes the 2D convolution process of multiple frames of images, and Figure 1(c) is the 3D convolution process.



(a) 2D convolution of a single frame



(b) 2D convolution of multi-frame images



(c) 3D convolution

Fig.1. Comparison of 2D convolution and 3D convolution

The size of the image sequence is H×W, the sequence length is L, the size of the 2D convolution kernel is k×k, and the size of the 3D convolution kernel is k×k×d.

It can be seen from Fig.1 that whether it is a 2D convolution based on a single frame or a 2D convolution based on multiple frames, the output result is in the form of a feature map, which cannot characterize the temporal characteristics of the video. This also means that the 2D convolutional network extracts the spatial features of the image, and the 3D convolution can extract not only the spatial features of the image, but also the temporal correlation features between image frames. In other words, 3D convolution can extract the depth and space-time features of the video.

Simonyan [6] used a 2D convolutional network of multiple images to extract deep space features. After the first convolutional layer, all feature information in the sequence time dimension has been lost.

Tran [7] tried to determine the best 3D CNN network architecture based on experience. Because training deep networks on large video data sets is very time-consuming, the author first tried to use UCF101 (a medium-scale data set) to search for the best architecture, and strive to verify the results of large-scale data sets through a small number of network experiments. According to the research on 2D CNN, the author found that the 3×3 size convolution kernel works best in the deep network architecture. Therefore, in the study of the 3D CNN architecture, the author fixed the spatial size of the convolution kernel to 3×3, and only changed the time depth of the 3D convolution kernel to find the best convolution kernel size.

The article also proposes a 3D CNN network structure with 8 convolutional layers, five pooling layers and two fully connected layers, called C3D network, and its network architecture is shown in Fig.2.

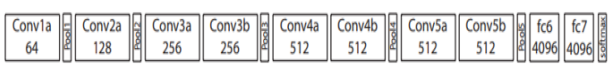


Fig.2. C3D network

The size of all 3D convolution filters is 3×3×3, and the step size is 1×1×1. The size of the first pooling layer is 1×2×2, and the step size is 1×2×2. The size of all other 3D pooling layers is 2×2×2, and the step size is 2×2×2.

The purpose of this design is to retain time information as much as possible in the early stage, and more effectively express the temporal and spatial characteristics of the video.

In view of the advantages of 3D CNN in spatiotemporal feature extraction, this paper chooses 3D CNN to extract local spatial features and short-term temporal features of video, and uses LSTM to extract long-term features of video.

LSTM is a variant of Recurrent Neural Network (RNN), which aims to learn the long-range dependence between input frames and output tags, and has achieved excellent performance in tasks such as speech recognition and action recognition [8]. Compared with 3D CNN which extracts video spatio-temporal feature information in the form of convolution, LSTM directly processes the video sequence and simulates the dynamic evolution of the video content state through the memory unit module. Since the classic RNN network retains the exponential decay of the video frame context information, it is limited in learning the long-term representation of video sequences [9]. In order to overcome this limitation, LSTM is designed as an architecture composed of a series of memory cells. Each memory cell contains internal states to store information from the sequence input to the current moment. In order to manage contextual information over a long period of time, three types of gate units are incorporated into the LSTM to control which information will enter or leave the memory unit over time. These gate units are composed of the non-linear function of the input/output sequence and the internal state to ensure that they have enough ability to simulate the dynamically changing context.

The gate unit of LSTM is divided into the following three forms:

(1) Input gate: Control the degree of influence of the input information entering the storage unit at a certain moment on the internal state of the storage.

(2) Forgetting gate: adjust the state before the current moment to control its contribution to the current state.

(3) Output gate: Controlling the information output from the memory unit will affect the future state of the LSTM unit.

The specific LSTM network structure is shown in Fig.3.

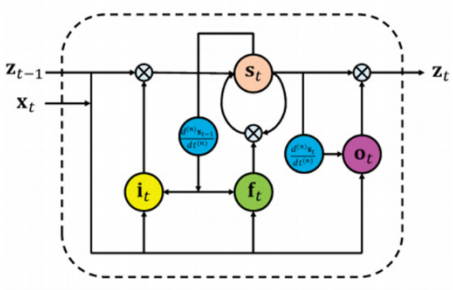


Fig.3. LSTM network

As can be seen from Fig.3, the implementation process of the basic unit of LSTM can be roughly summarized as follows: After the output state at the previous moment and the input state at the current moment are each judged on the degree of influence on the context information, the weighted sum is performed, and then output to the next unit. The overall LSTM network structure is an iteration of the above process.

Because it always retains all the information from the beginning to the end of the video, it constantly learns the relationship between existing information and new input information, and dynamically adjusts the weight of context information. Therefore, LSTM has shown great advantages in extracting long-term features of videos, and is widely used in various fields such as video classification and gesture recognition.

3 3D CNN+LSTM-based monitoring model for transmission line protection against external damage

The biggest difficulty of image technology-based transmission line protection methods is how to judge whether the current behavior will cause damage to the transmission line based on the information of a single image. Taking the construction situation as an example, when line poles and construction vehicles appear in a picture at the same time, the image-based monitoring method can easily judge the situation as external force damage. After watching the full video, you may find that this is just a construction vehicle passing near the tower. The above example illustrates the difference in principle between video-based monitoring technology and image-based monitoring technology in the monitoring scenarios of transmission line damage prevention. Video-based methods focus on abnormal behaviors. Compared with image methods based on target monitoring, they have higher prediction accuracy, and reduce the economic loss and personnel costs caused by false alarms to a certain extent.

A large number of studies have shown that 3D CNN and LSTM have good performance in characterizing the temporal and spatial characteristics of video content. In this paper, 3D CNN and LSTM are combined to fully extract deep spatiotemporal features, and then complete the efficient characterization and description of video content. Next, we will introduce how to use deep space-time features to model transmission line protection against external damage.

At present, most of the current transmission line monitoring methods for preventing external damage often ignore the characteristics of the video content, so the false alarm rate of the established model is very high. Due to the strong time correlation of video, considering the time and computational cost, this paper combines 3D CNN and LSTM to extract deep spatio-temporal features, and proposes a 3D CNN+LSTM-based transmission line monitoring model based on external damage. The 3DCNN+LSTM network structure proposed in this paper is shown in Fig.4.

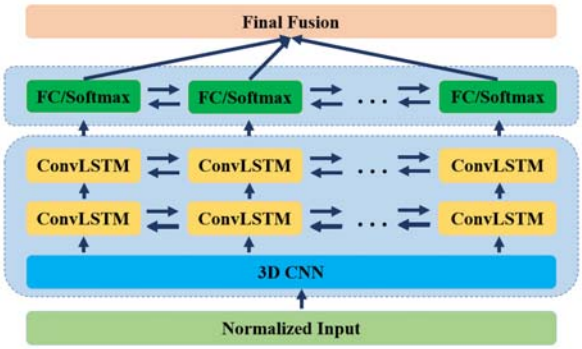


Fig.4. 3D CNN+LSTM network

As shown in Fig.4, this paper adopts a supervised learning method for modeling, the input is the normalized video frame, and the output is the behavior category. Considering that the length of the video sequence is too long, this paper uses a two-layer nested LSTM network to further learn the extracted 3D CNN features to speed up the learning rate, and finally output through the fully connected layer and the SoftMax layer.

The network structure proposed in this paper makes full use of the advantages of 3D CNN for spatio-temporal feature extraction and LSTM for long-term sequence learning. At the same time, it avoids the problem that 3D CNN is limited by computing power and cannot process too long sequences. , And the disadvantage of LSTM in feature extraction. Compared with the existing method, its advantage is that it considers the time characteristics of the external damage situation and fully analyzes the possibility of the external damage behavior. Therefore, the built model has extremely high accuracy.

4 Results and Analysis

The essence of the method of preventing external damage based on video processing is to effectively classify the video and accurately identify, which type of behavior in the video will cause damage to the transmission line. Taking into account the variety of external damage behaviors, there are certain difficulties in video capture, so this paper chooses the UCF101 video data set specifically for behavior classification to conduct experiments. In order to be more in line with the actual situation, this article uses five categories for experiments, 70% of the data is used for training, and the remaining 30% is used to test the accuracy of the model.

In the experiment, the 3D CNN network parameters are set as follows: contains a total of 5 convolutional layers and pooling layers, all convolution kernels are 3×3×3, except for the first pooling layer that has a size of 1×2×2, And the rest are 2×2×2. LSTM is divided into two layers, and the parameters are set as follows: the first layer network output dimension is 1024, and the second layer network output dimension is 512. The training batch of the overall network is 32, and the number of iterations is 100.

In order to verify the monitoring accuracy of the model proposed in this chapter, this paper uses different deep neural network structures for modeling, including C3D deep neural network structure, Conv_3d network structure and Long-term Recurrent Convolutional Networks (LRCN) as a comparative test.

The C3D network structure has been introduced in detail above, so I won't repeat it here. Conv_3d is an improved 3D CNN model based on the C3D model. The size of the first convolution kernel in Conv_3d is 7×7×3, and the size of the second convolution kernel is 7×6×3. Where the size of the pooling layer is 2×2×2, and finally it is output through a layer of two-dimensional convolution and a layer of fully connected layer. The network structure is shown in Fig.5.

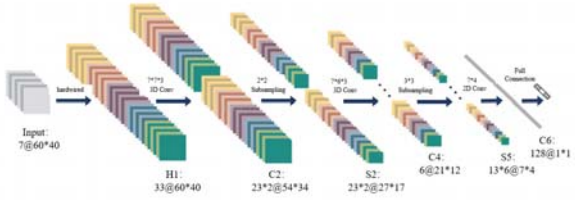


Fig.5. Conv_3d network

LRCN is similar to the network architecture proposed in this article. The difference is that the 2D CNN network is used to extract spatial features, while the method in this article uses the 3D CNN network. The schematic diagram of the LRCN network structure is no longer given here [10].

Table 1 shows the results of different depth neural networks. These models are based on the temporal and spatial characteristics of the video frame sequence to build models. Use classification accuracy and RSME to evaluate the performance of the model. It can be seen from the table that the accuracy of the model obtained by using the method proposed in this paper can reach 99.16%, and the RMSE is only 0.1104. The performance far exceeds other deep neural network structures, which proves that the model can accurately predict the external breaking behavior of transmission lines.

Table 1. Comparison results of models using different deep learning

MODEL	ACCURACY	RMSE
Conv_3d	0.8211	0.5083
C3D	0.8304	0.4314
LRCN	0.9557	0.1526
Proposed Method	0.9716	0.1322

From the comparison of the experimental results of different deep neural networks, the following conclusions can be drawn:

① The deep neural network based solely on 3D CNN only considers the spatiotemporal information of several adjacent video frames in the sequence, and the prediction results of long-term sequences with continuous meaning in video content are not good.

② Compared with the method of using 3D CNN to extract features, using 2D CNN to extract features will cause a certain decrease in the accuracy of the model. This is because when 3D CNN extracts features, it first focuses on the spatial characteristics in the first few frames and pays attention to the significant motion in the subsequent frames. This is the most essential difference compared to 2D CNN. In other words, compared with 2D CNN, 3D CNN has stronger spatiotemporal feature extraction and expression capabilities, so 3D CNN is very suitable for spatiotemporal feature learning.

③ LSTM learns the long-term temporal features of the video sequence, which makes up for the lack of 3D CNN that only extracts temporal feature information for a few adjacent frames. After joining the LSTM network, the isolated time features in the entire video sequence are connected in series and learned as a whole, so the model has more accurate prediction results.

5 Conclusion

In summary, the 3D CNN+LSTM-based transmission line monitoring method for external damage prevention proposed in this paper is more comprehensive in the extraction of deep temporal and spatial features and can accurately determine the occurrence of external damage. Therefore, it is shown in comparison with other networks It has good monitoring accuracy.

This paper proposes a video monitoring method based on deep learning for external damage in transmission lines. This method uses 3D CNN and LSTM network to conduct in-depth analysis of video content to determine whether the behavior in the video will cause damage to the transmission line. Making full use of the good video capture device set up on the existing line has a greater advantage in cost and safety compared to other methods. Validated by the public data set, this method has shown good performance in the accuracy of the classification of external damage behaviors.

In the future research, the research will focus on the collection of video data. The model is adjusted according to the data collected by the monitoring camera of the actual transmission line to meet the actual needs of monitoring the transmission line against external damage.

References

1. GOLIGHTLY I, JONES D. Corner detection and matching for visual tracking during power line inspection[J]. Image & Vision Computing, 2003, 21(9):827-840.
2. LARRAURI J I, SORROSAL G, GONZÁLEZ M. Automatic system for overhead power line inspection using an Unmanned Aerial Vehicle - RELIFO project[C]// International Conference on Unmanned Aircraft Systems. 2013.
3. FANG S, LI L, ZHENG Y, et al. Protection Technology for Transmission Lines Based on Laser Range Imaging and Image Processing [J]. Electrical Automation. 2017, 39(3):6-8.
4. LIU J, JING Z, HUI Z, et al. Extracting Deep Video Feature for Mobile Video Classification with ELU-3DCNN [J]. 2017.
5. HAQUE A, ALAHI A, FEIFEI L. Recurrent Attention Models for Depth-Based Person Identification [C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016.
6. SIMONYAN K, ZISSERMAN A. Two-Stream Convolutional Networks for Action Recognition in Videos [J]. 2014.
7. TRAN D, BOURDEV L, FERGUS R, et al. Learning Spatiotemporal Features with 3D Convolutional Networks [J]. 2014..
8. GERS F A, SCHMIDHUBER J, CUMMINS F. Learning to Forget: Continual Prediction with LSTM [J]. Neural Computation, 2000, 12(10):2451-2471.
9. XIAO A, LIU J, LI Y, et al. Two-phase rate adaptation strategy for improving real-time video QoE in mobile networks [J]. China Communications, 2018, 15(10):12-24.
10. BU S J, CHO S B. A Hybrid Deep Learning System of CNN and LRCN to Detect Cyberbullying from SNS Comments [M]. 2018. English version of this patent.