Multi-classifier Combined Anomaly Detection Algorithm Based On Feature Map In Underground Coal Mine

Yan Fu\textsuperscript{1a}, Zemin Cui\textsuperscript{2b}, Ou Ye\textsuperscript{3}

\textsuperscript{1}School of Computer Science and Technology Xi’an University of Science and Technology Xi’an, China
\textsuperscript{2}School of Computer Science and Technology Xi’an University of Science and Technology Xi’an, China
\textsuperscript{3}School of Computer Science and Technology Xi’an University of Science and Technology Xi’an, China
\textsuperscript{b}www.401960380@qq.com

Abstract. The detection of abnormal activities in deep learning is of great significance for preventing the occurrence of abnormal disasters in mine production. As the underground scenes of coal mines are characterized by much noise and uneven light, the traditional manual feature extraction method has little obvious effect in the underground and low accuracy of anomaly detection. To solve the above problems, a feature extraction method combining CNN+LSTM is proposed. Secondly, the obtained features are matched by graph structure. Finally, multiple classifiers are used to classify the features before and after matching. In this paper, experiments are carried out in coal mine dataset and UCSDped1 dataset respectively, and comparisons are made with some classical algorithms. Experimental show that the algorithm achieves high recognition accuracy in different abnormal event datasets.

1. Introduction
Abnormal activity detection technology is an application of image processing. It is a technology to extract feature information from the image and then find out the abnormal information based on feature classification modeling. Abnormal activity detection is one of the most active research topics in the field of computer vision. Its core is to detect, identify and track targets from images and understand and describe their behaviors by using image analysis, computer vision and other technologies, so as to find abnormal activities.

At present, there are two kinds of methods for abnormal activity detection, one is the traditional manual method, the other is the new deep learning method in recent years.

According to the characteristics used, abnormal behavior detection methods based on traditional characteristics can be divided into two categories: track-based abnormal detection\textsuperscript{[1]} and visual feature-based abnormal detection\textsuperscript{[2]}. These features are essentially hand-designed features. For different recognition problems, the extracted features have a direct impact on the performance of the system. Therefore, researchers need to conduct in-depth research on the problem areas to be solved in order to design more adaptive features to improve the performance of the system.

The anomaly detection method based on deep learning\textsuperscript{[3,4]} mainly uses the neural network to extract the features of video activities and then classify the abnormal activities, taking advantage of the high accuracy and strong anti-interference ability of deep learning. However, most of the anomaly detection
algorithms based on deep learning often adopt a single classifier, which is difficult to ensure the accuracy of classification. Especially for complex scenes, the interference of light source, occlusion and other factors in the video will obviously affect the accuracy of the algorithm.

Both the traditional manual method and the convolutional neural network method have great room for improvement in classification accuracy. Aiming at the problem of uneven light and much noise in coal mine. In this paper, a method is proposed to extract features by combining the convolutional neural network with the cyclic neural network, which makes use of the advantages of the convolutional neural network in spatial features and the advantages of the cyclic neural network in video sequence timing features. At the same time, the graph structure is used to match the effective features obtained by the neural network. Then the characteristics before and after matching with multiple classifiers are combined to complete the detection of abnormal activity in underground coal mine.

2. Multi-classifier Combined Anomaly Detection Based on Feature Graph

The implementation framework of this algorithm is shown in Fig. 1:

First, VGG16 network is used to extract the spatial features of the video sequence, and then it is passed into LSTM network to further extract the temporal features. The obtained features are matched with the graph structure, and the main features are screened. The main features and the spatial and temporal features obtained from the deep network are introduced into multiple classifiers for joint classification, and the results are obtained.

2.1. Depth Feature Extraction

Convolutional Neural Network (CNN)[5] is a typical example of deep learning and performs well in various algorithms. VGG16 convolutional neural network is used in the experiment. The VGG16 convolutional neural network model adopts one pooling after two convolution operations with 64 convolution kernels by inputting images (224*224*3). And then after two 128-convolution kernels, a pooling is used. After three convolution operations with 256 convolution kernels, a pooling is used. After three convolution operations with 512 convolution kernels, pooling is adopted. Repeat the convolution twice with three 512's and then pool it again. Finally, adjust the vector characteristics of the output (1*4096). Fig. 2 shows the network model of VGG16.

![VGG16 Network model frame diagram.](image-url)
The basic unit of Long Short Term Memory network (LSTM) consists of a memory unit, activation function and three gates (input \(i_t\), forget gate \(f_t\) and output gate \(o_t\)). The cell structure of the LSTM model that is most widely used is shown in Fig. 3. Its forward calculation method can be expressed as (1) (2) (3) (4) and (5):

\[
\begin{align*}
i_t &= \sigma(W_x i_t + W_h h_{t-1} + W_c c_{t-1} + b_i) \\
f_t &= \sigma(W_x f_t + W_h h_{t-1} + W_c c_{t-1} + b_f) \\
c_t &= f_t + i_t \tanh(W_x c_t + W_h h_{t-1} + b_c) \\
o_t &= \sigma(W_x o_t + W_h h_{t-1} + W_c c_{t-1} + b_o) \\
h_t &= o_t \tanh(c_t)
\end{align*}
\]

Where \(i, f, c\) and \(o\) are input gate, forget gate, cell state and output gate respectively. \(W\) and \(b\) are the corresponding weight coefficient matrix and the bias term respectively; \(\sigma\) and \(\tanh\) are the sigmoid and hyperbolic tangent activation functions respectively.

In the experiment, the \((1*4096)\) dimensional feature vectors extracted by VGG16 convolutive neural network were used as the input of LSTM neural network to obtain \((64*256)\) dimensional features with temporal characteristics.

2.2. Graph Structure Matching

The \((64*256)\) dimensional feature of VGG16+LSTM output needs further processing before it can be used for graph structure matching. The tensor needs to be stretched and stretched into a one-dimensional \((1*1024)\) eigenvector. The 256-dimensional feature of the vertical axis is partitioned into 16 parts, and the 256-dimensional feature of the vertical axis is adjusted to the 16-dimensional feature by using 2 norms. The \((64*16)\) is obtained, and the feature of \((1*1024)\) is finally stretched and stretched for subsequent matching calculation. The new feature vector \(F = \{f_i\}_{i=1}^{1024}\) is obtained by adjusting the feature. Now let's represent the video frame with a graph, where \(G = (f_i,E)\), \(f_i\) is a value in the eigenvector, and \(E\) is the edge distance between two values of \(f_i\) and \(f_j\) in the eigenvector.

The calculation process of fuzzy membership fraction of edge is shown in (6).

\[
d_{ij} = \sqrt{\lambda_1(x_i-x_j)^2 + \lambda_2(y_i-y_j)^2}
\]

Where, \((x_i,y_i)(x_j,y_j)\) are the positions of two values in the eigenvector in two-dimensional coordinates, and the existence of an edge can be judged by calculating the distance between the two positions. If the distance is far, the correlation between the two points is low. If the distance is close, the correlation between the two points is high. So the problem of edge existence turns into a problem of distance, where \(d_{ij}\) is less than a certain value. The weighted sum of the arguments \(\lambda_1\) and \(\lambda_2\) for \((x_i-x_j)^2\) and \((y_i-y_j)^2\) filters out the edges that form when the x-coordinate is far away and the y-coordinate is close. The adjacency matrix \(A\) of graph \(G\) in this eigenvector can be expressed by (7).

\[
A = \begin{cases} 
1, & d_{ij} < d_T \\
0, & d_{ij} \geq d_T 
\end{cases}
\]
Where, $d_T$ is the threshold value to judge the existence of edge. When the distance is within the threshold range, it indicates the existence of edge, which is connected by a straight line. The correlation between all the features in the feature (1*1024) was calculated by formula (6) and (7). Finally, the feature diagram containing several graph structures was obtained, and the graph structure $G$ was shown in (8).

$$G = \sum_{i=1}^{m} (f_i, 1), \quad i \in N^+, \ i \in [1,1024]$$

(8)

Where, $f_i$ is the size of the feature in the eigenvector, $I$ is the abscissa of the eigenvector in that coordinate, $n$ and $m$ are the subscripts of the continuous features constituting the graph structure, $N^+$ is the positive integer between 1 and 1024, and 1 is the correlation between adjacent features $f_i$.

2.3. Classifier union classification

Most traditional neural networks use softmax as a classifier for classification, but softmax as a classifier is not strong generalization ability, and is often used in multi-classification problems. To solve the dichotomy problem of anomaly detection, this paper proposes the joint classification using Support Vector Machines (SVM) [7] and Random Forest (RF) [8].

Support vector machines often encounter the problem that only two classes are needed in classification. SVM classification method is generally used when considering dichotomy. Its model is a plane equation. For a given training dataset, the optimal segmentation plane is shown in (9).

$$w^* \cdot x + b^* = 0$$

(9)

Where, $w$ is the weight matrix and $b$ is the bias. The classification decision function is shown in (10).

$$f(x) = \text{sign}(w^* \cdot x + b^*)$$

(10)

Random forest is a classifier that uses multiple Decision Trees to train and predict samples. Random forest trains multiple decision tree weak classifiers, then combines these weak classifiers to form a strong classifier, and gets the final classification result by voting. The random forest has a good classification effect on the multi-dimensional feature dataset, strong generalization ability, and can also select the importance of the feature, with high operational efficiency. The random forest algorithm is described as follows:

- Input training dataset $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$ the number of iterations $T$. Output the final strong classifier $f(x)$.
- For $t = 1, 2, \ldots, T$:
  - Perform the $t$th dataset generation in the original dataset. Through $m$ times of put back random sampling, the sampled dataset $D_m$ is obtained, which contains the same number of samples as the original dataset.
  - The $m$th decision tree $G_m(x)$ was constructed by using the sampling set $D_m$.
  - Learn $T$ decision trees in total through step 1. Each tree is a weak classifier. In the classification, each decision tree gives a classification result and votes according to the result. The final classification result adopts the result with the most votes.

3. Experiment and result analysis

This experiment was performed on a personal computer with i7-8750h processor and 16GB memory. The system environment was Windows10 and the programming environment was Python3.6. The TensorFlow1.13.1 deep learning framework was adopted.

3.1. Dataset and Evaluation Index

The dataset used in this paper is the video dataset of the same scene in coal mine [9]. There are 73 videos in the coal mine dataset for training and surveying, with each frame being a (224*224) three-channel image. In this experiment, 6 videos were selected from 3 scenes with a total of 6,879 frames of images. The normal data in the dataset has no coal accumulation, while the abnormal video has a lot of coal accumulation. At the same time, this paper also verifies the effectiveness and robustness of this method in the public exception event dataset UCSDped1[10]. There are 34 training sets and 36 test sets in
UCSDped1. There are about 200 frames of images in each clip, and each frame is a (238*158) three-channel image.

In order to verify the effectiveness of the algorithm, this paper evaluates the experimental results by three indexes: Accuracy, Precision, and Recall. The calculation process is shown in (11), (12) and (13).

\[
\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}
\]

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

\[
\text{Recall} = \frac{TP}{TP + FN}
\]

Where, TP: the sample is positive and the predicted result is positive; FP: The sample was negative and the predicted result was positive; TN: The sample is negative, and the predicted result is negative; FN: The sample is positive and the predicted result is negative.

3.2. Experimental scheme

In order to verify the effectiveness and robustness of the algorithm, the Accuracy, Precision and Recall of this algorithm and other deep learning algorithms (CNN+softmax, CNN+LSTM[11], CNN+LSTM+SVM[12]) were compared with the same parameters in the coal mine dataset and UCSDped1 dataset respectively.

3.3. Experimental analysis

The experimental results of different algorithms on the two datasets are shown in Table 1 and Table 2:

| Methods                  | A (%) | P (%) | R (%) |
|--------------------------|-------|-------|-------|
| This paper               | 99.07 | 98.96 | 99.97 |
| CNN+softmax              | 81.12 | 83.34 | 87.47 |
| CNN+LSTM                 | 92.45 | 91.33 | 91.27 |
| CNN+LSTM+SVM             | 98.75 | 98.68 | 99.87 |

| Methods                  | A (%) | P (%) | R (%) |
|--------------------------|-------|-------|-------|
| This paper               | 99.39 | 99.32 | 99.91 |
| CNN+softmax              | 82.56 | 79.23 | 83.44 |
| CNN+LSTM                 | 90.32 | 89.46 | 86.24 |
| CNN+LSTM+SVM             | 98.70 | 98.98 | 99.38 |

It can be concluded from Table 1 that only USING CNN has a general effect. The LSTM neural network method introduced on the basis of CNN can improve the accuracy of abnormal activity detection in coal mines, and then the SVM classification method is adopted to further improve the accuracy. On the basis of the previous methods, this paper introduces the LSTM neural network, then combines the method of graph structure, and USES the joint classification of multiple classifiers, which further improves the accuracy. It can be seen from Table 2 that this experimental method can also perform well on the common data set. Therefore, on the basis of neural network CNN+LSTM, this experiment can achieve more than 99% detection accuracy by combining the method of graph structure feature matching.

Experimental results on coal mine data set and UCSDped1 data set are shown in Fig. 4 and Fig. 5 respectively.
4. Conclusion
This paper proposes an improved mine video abnormal activity detection algorithm based on graph structure matching. Compared with most abnormal activity detection algorithms, this algorithm combines CNN+LSTM, and introduces an improved graph structure feature matching algorithm based on this, and uses multiple classifiers for joint classification, which greatly improves the accuracy of classification results. After many experimental tests, the algorithm presented in this paper performs very well on the coal mine dataset and the UCSDped1 dataset, with the accuracy reaching 99.07% and 99.39% respectively.

Acknowledgment
This work was supported in part by Natural Science Basic Research Plan in Shaanxi Province of China under Grants Nos. 2018JQ5095, 2020JQ-758, 2020JM-533 and Chinese Postdoctoral Science Foundation under Grant No. 2020M673446.

References
[1] Imran N. Junejo. Using dynamic Bayesian network for scene modeling and anomaly detection[J]. Signal, Image and Video Processing,2010,4(1):pp. 1-10.
[2] Tran D, Yuan J. Optimal spatio-temporal path discovery for video event detection[C]// Proc. computer vision and pattern recognition, 2011, pp. 3321-3328.
[3] HU X, HU SQ, HUANG Y P, et al, Video anomaly detection using deep incremental slow feature analysis network[J]. IET Computer Vision, 2016, 10(4):pp. 258-267.
[4] Qiu Peng , Kim Sumi. Anomaly detection in a crowd using a cascade of deep learning networks[J]. Intelligent Systems and Computing.2018,53(5) : pp. 596-607.
[5] T. Gupta, V. Nunavath and S. Roy, “CrowdVAS-Net: A deep-CNN based framework to detect abnormal crowd-motion behavior in videos for predicting crowd disaster,”[C]// Proc. IEEE International Conference on Systems, Man and Cybernetics. 2019, pp. 2877–2882.
[6] Thomas Fischer,Christopher Krauss. Deep learning with long short-term memory networks for financial market predictions[J]. European Journal of Operational Research,2018,270(2). pp. 654-669.
[7] Li Guo,Samia Boukir. Fast data selection for SVM training using ensemble margin[J]. Pattern Recognition Letters,2015,51. pp. 112-119.
[8] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): pp. 5-32.
[9] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, “Anomaly detection in crowded
scenes,"[C]/ Proc. IEEE Conf. CVPR, San Francisco. 2010, pp. 1975–1981.

[10] O. Ye, Z. Li and Y. Zhang, “Near-duplicate video cleansing method based on locality sensitive hashing and the sorted neighborhood method,"[C]/ Proc. 2nd EAI International Conference on Robotic Sensor Networks, Kitakyushu, Japan, Aug. 2018, pp. 129–139.

[11] D. Arifoglu, A. Bouchachia, “Detection of abnormal behavior for dementia sufferers using convolutional neural networks,"[J] Artificial Intelligence in Medicine, vol. 94, pp. 88–95, March. 2019. DOI: 10.1016/j.artmed.2019.01.005

[12] K. Vignesh, G. Yadav, and A. Sethi, “Abnormal event detection on BMTT-PETS 2017 surveillance challenge,"[C]/ Proc. IEEE Conf. CVPR, Honolulu, USA, Aug. 2017, pp. 2161–2168.