Deep Learning Based Hybrid Approach for Crowd Anomalous Behavior Detection

Aniruddha Prakash Kshirsagar and L. Shakkeera

1 Introduction

With the rapid development of information and technology, surveillance video system has been widely used in public like highways and stations, and a large amount of abnormal activities has been recognized and analyzed in video data. In actual application, it is a significant direction to recognize various actual scenes with high accuracy and missing report rate. It is necessary to study the recognition method in video based on deep learning, which is helpful to reduce the safety hidden trouble caused by abnormal activities [1].

Computer vision [2] and other methods have been used to recognize the abnormal activity. At present, existing researches mainly combine human and intelligent video surveillance to monitor and warn against abnormal activity. Manual recognition is still the main method and is supplemented by automation and information technology, thus the standard of abnormal activity recognition needs to be improved. Because the SVM network model can accurately describe the semantic characteristics of video time series changes, and is suitable for identifying abnormal activity with relatively long intervals and delays in videos. So, the SVM network can be used to perceive the semantic characteristics of abnormal activities in videos, which is conducive to the early recognition of hidden security problems and effectively alleviating the problems caused by manual recognition. Dubey et al. [2] proposed a method based on the combination of trajectory and pixel analysis to measure the velocity and direction of the moving target trajectory and realized the recognition of abnormal activity through a clustering algorithm. The accuracy of trajectory feature extraction has a great influence on the result and is not applicable to video data with many noises. AI and deep learning are ideas that are regularly covered. There can be a slight disarray between the terms, Machine learning utilizes a bunch of calculations

A. P. Kshirsagar (✉) · L. Shakkeera
School of Computing Science and Engineering, VIT Bhopal University, Madhya Pradesh, India
e-mail: anipk2007@gmail.com

R. J. Kannan et al. (eds.), *International Virtual Conference on Industry 4.0*, Lecture Notes in Electrical Engineering 1003, https://doi.org/10.1007/978-981-19-9989-5_2

to dissect and decipher the information, gain from it, and in light of the learnings, settle on the most ideal choices. Then again, deep learning structures the calculations into various layers to make a "fake neural organization". This neural organization can gain from the information and settle on shrewd choices all alone.

1.1 Deep Learning

Customary AI strategies will in general capitulate to ecological changes while profound learning adjusts to these progressions by steady criticism and work on the model. Profound learning is worked with by neural organizations which mirror the neurons in the human cerebrum and installs numerous layer design (few noticeable and few covered up). It is a high-level type of AI, which gathers information, gains from it, and enhances the model. Regularly a few issues are mind boggling to the point that it is essentially outlandish for the human cerebrum to understand it, and subsequently programming it is an unrealistic idea. Crude types of Siri and Google Aides are a fitting illustration of customized AI as they are found compelling in their modified range. However, Google's profound psyche is an extraordinary illustration of profound learning. Profound learning implies a machine, which learns without anyone else through numerous experimentation strategies. Frequently a couple hundred million times.

1.2 Existing Approaches

In article [3], Sultani et al. combined histogram, PHOG and HMOEOF features to recognize abnormal activity through SVM. However, their method requires an amount of calculation and the final classification accuracy needs to be improved. Kavikuil and Amudha [4] proposed an anomaly activity recognition model based on the AlexNet network, but the imbalance of recognized data is an important factor that affects the algorithm's training feature.

Compared with the mentioned methods, the extracted feature's quality was affected by data noise, the video sequence information utilization rate is low, and poor classification results a multiple feature fusion based on CNN and SVM abnormal activity recognition method was proposed and introduced the attention mechanism [5] to SVM, then analysis the correlation between the features, which can effectively extract features to reduce the long sequence information and the information shortage.

2 Proposed Convolutional SVM Approach

A. *Activity representation with deep learning models*

We transform the issue of abnormal activity recognition into an outlier recognition problem of space–time sequence, and the output is divided into two types: normal and abnormal activities. The spatial–temporal features were extracted by CNN and SVM. SVM can effectively avoid long-term dependence problems, and the gradient will not disappear after time back-propagation training [6]. In addition, attention mechanism was introduced to effectively analyze the correlation between model input and output, avoiding the influence of background noise and long sequence, to obtain more information (Fig. 1).

Each frame is the input of CNN model for convolution operation in the video, and finally a 2048-dimensional feature vector $\mathbf{C_r}$ will be chosen as spatial output through the fully connected layer for transmission to the SVM Attention layer.

B. *SVM attention models*

In sequential tasks, it is critical to learn the time dependence between the inputs. As a special time recurrent network, SVM obtains higher level information by stacking together [7]. The cell structure is shown in Fig. 2.

SVM network is controlled and updated by input gate i_t, forgot gate $f_{t,}$ and output gate o_t, where there is an input, if i_t is activated, its information will be stored in the cell. Also, if f_t is turned on, the unit state c_t is forgotten. The latest feature in the fully connected layer. Next, the 1 * n-dimensional feature vectors are feeding into the SVM unit output c_t is determined by whether $o_t$is propagated to the final state *h*.

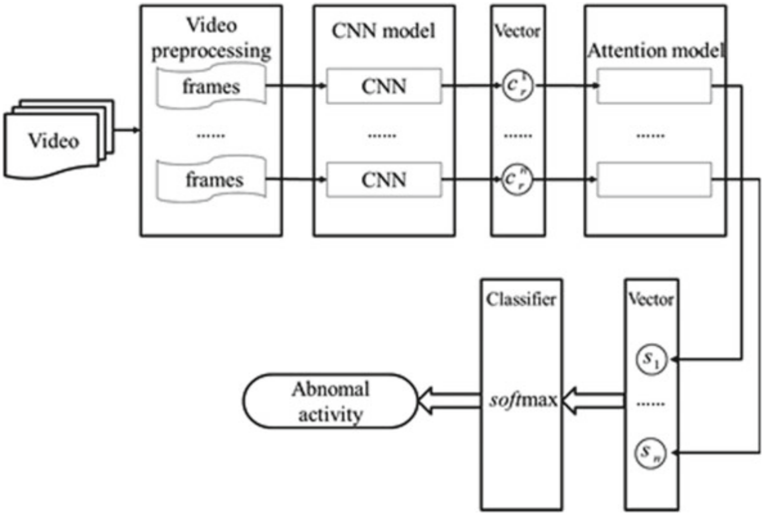

Fig. 1 Method of abnormal recognition-based CNN SVM attention models

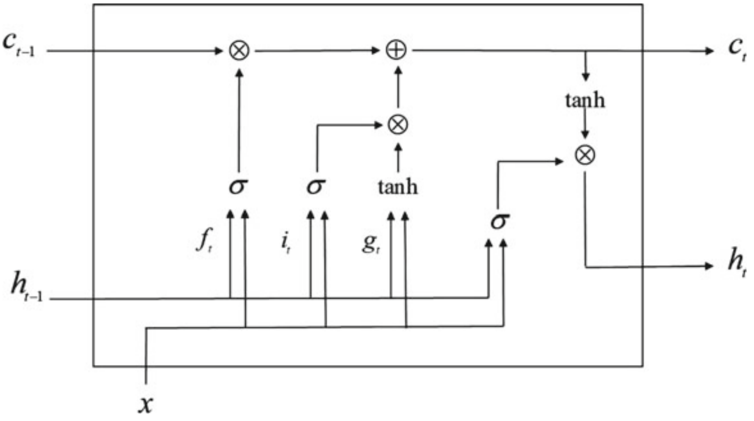

Fig. 2 SVM cell structure

The state of each cell can be expressed by the attention model to train the time series features. The attention mechanism can distinguish key features from the hidden state output of the SVM layer.

C. *CNN models*

The essence of CNN is to extract the visual features between data through convolution and pooling operations, and the extracted features will become more and more abstract with the increase of the number of layers, and finally converge at the full connection layer. Due to the good performance in the process of feature extraction, we chose inception-v3 model to extract features that are different from traditional CNN models, it convoluted images through different convolution verification operations, and then combined different convolution layers in parallel. The dataset used in the experiment is a publicly dataset UMN [8] with a resolution of 320 * 240, and it contains normal and abnormal activity in the crowd. Dataset contains 11 videos in 3 scenes where some people walking normally and suddenly running after some time, and all video scenes are taken in a permanent position with a static background. We trained on a normal section of 5 videos of all scenes and tested on all videos. The experimental parameters settings: experiment SVM super parameter of the model is obtained by cross-validation, and using the Singh and Mohan [9] optimization neural network model, it can weight vector update and set up according to the model, using the batch size of 64, every time training for the whole is represented by feature vector after pooling layer. CNN as input of SVM network, output vector used in this experiment single-layer SVM network and the attention of the input layer, in the attention layer, to compute the weight vector, and then the weight vector and the input vector to merge the current layer, a new vector **s** and as a weighted vector and all of the time step characteristics, its overall structure is shown in the Fig. 3.

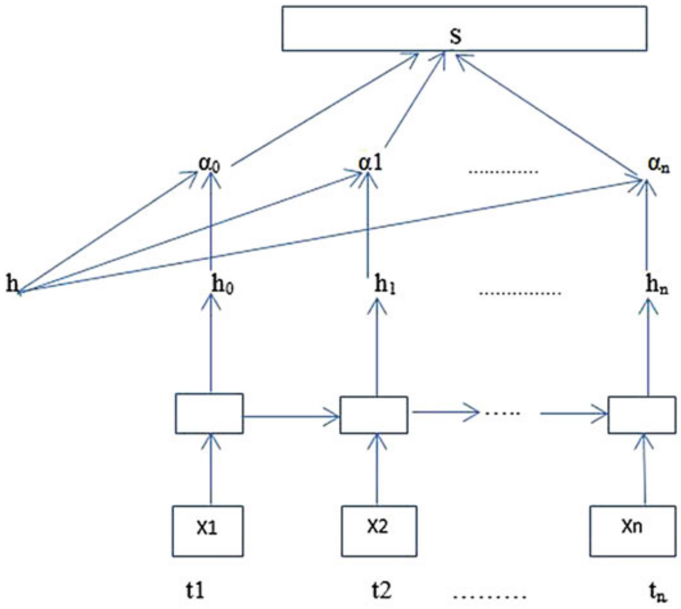

Fig. 3 The model of SVM attention

2.1 C-SVM Working

After the multiple steps like data selection and data preprocess, the proper sorted and clean data are used for extracting the unused feature in the proposed approach which can be done by the SVM approach by considering the dataset into the vector by removing the grid of images. In the proposed hybrid approach with have used the combination of CNN and SVM so can find exact abnormal activity through a proper approach with automatic indication. Following some steps gives the detail overview of the approach.

1. New Trainingset {xi, yi, i=1…l+1}
2. New Coefficients qi, i=1…l+1
3. New Bias b
4. New Training Set Partition
5. New R Matrix

The output contains all the values given in the input updated. In the above steps, we have created and verified the dataset by maintaining the dataset and properly patinating the data to check the abnormal entry. The output contains all values given in the input updated.

Forgetting Algorithm

```
IF(c € REMAINING SET)
REMOVE SAMPLE FROM REMAINING SET
REMOVE SAMPLE FROM TRAINING SET
EXIT
IF(c € SUPPORT SET)
REMOVE SAMPLE FROM SUPPORT SET
IF(c € ERROR SET)
REMOVE SAMPLE FROM ERROR SET
```

In the above step check whether the normal activity is available in movement or not by cross-checking the activity with the store dataset if it’s available and the activity is normal then it works as it is or is detected as abnormal activity in a particular area the following section gives the experimental analysis of real-time video.

3 Experiment and Analysis

A. *Experiment environment and dataset*

The experiment used Python to program and TensorFlow for the training model, and the Python version is 3.6. Anocanda3 is used to build the experimental model in the Linux operating system server version Ubuntu 14.

B. *Results and analysis*

The accuracy represents the proportion of samples correctly classified in all classifications. The precision rate indicates how many of the predicted samples (such as positive samples) actually samples of a certain type. Recall is how much of a sample is correctly predicted.

```
For Prediction and Accuracy
Algorithm 1 (Predict,Accur)=C-SVM(Train,Div,Test,Test-Final, ɵ)
ɵ=Termination condition Ensure: Predict->Predicted sentiment output
Accur->Accuracy
Net->Create Network
Network_initialize(Net)
for error>=do
error Network_Train(Net,Train,Div)
end for
/*Training completed*/
Featureopt->C-SVM(Train,Div)
HTrain->GetTop_HiddenLayer(Net,Train)
Train_combined<-HTrain + Featureopt
ModelSVM<-SVMLinear(Traincombined)
HTest<-GetTop_HiddenLayer(Net,Test)
```

Table 1 Comparison of experiments

Method	Accuracy (%)	Precision (%)	Recall (%)
CNN	83.13	76.63	99.32
SVM	89.41	83.23	96.77
C-SVM	94.30	89.24	96.61

```
12. Testcombined<-HTest + Featureopt
13. Predict<-SVMLinear(ModelSVM,Testcombined)
14. Accur<-Evaluation(Test-Final,Predict)
15. return(Predict,Accur)
```

The comparison of proposed approach with existing method in Table 1.

The result in Table 1 shows that the accuracy obtained from UMN dataset is higher than the other three existing methods. Figure 4 shows that the initial loss value caused by the increase in complexity increases and the convergence speed is fast after the attention mechanism is introduced into the SVM network during model training.

Figures 5, 6, 7 and 8 indicate that the attention mechanism improves the prediction of final classification to some extent. By introducing the attention mechanism into the hidden layer of the SVM network, feature loss caused by a long sequence can be effectively solved and important features can be highlighted, thus improving the performance of the model.

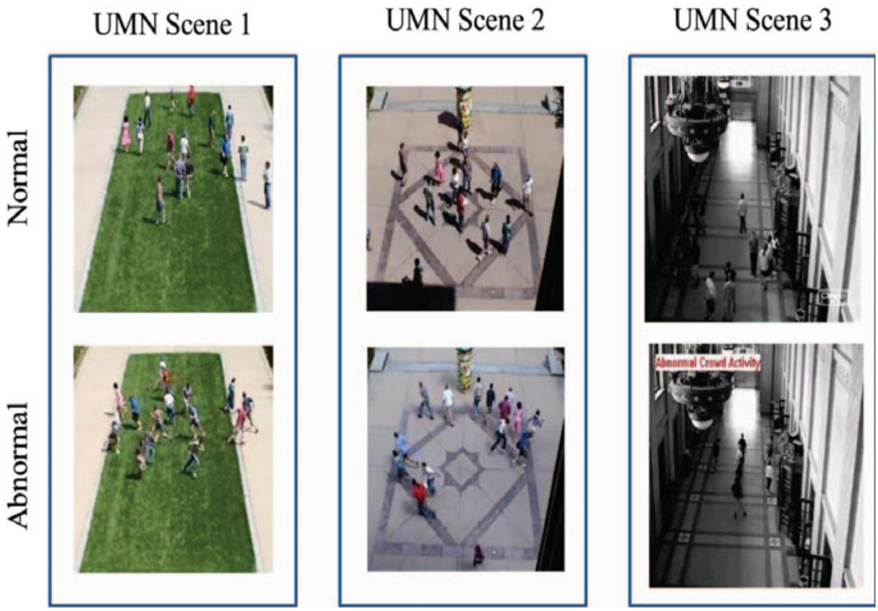

Fig. 4 Normal and abnormal activity in dataset

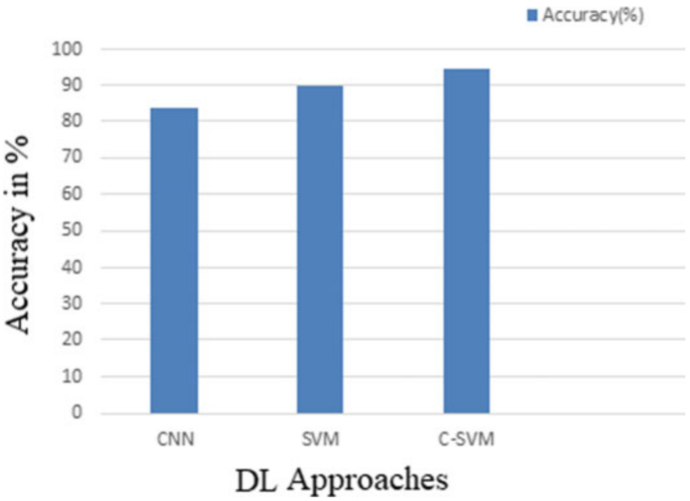

Fig. 5 Analysis on basis of accuracy

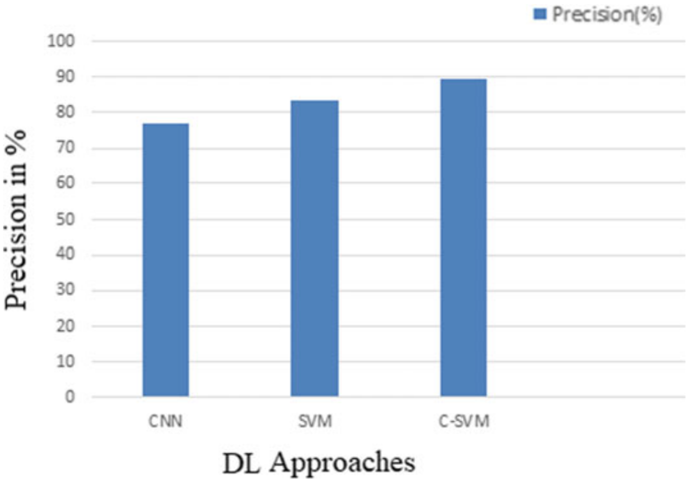

Fig. 6 Analysis on basis of precision

4 Conclusion

We proposed a new method of abnormal activity recognition that applies the deep learning and attention mechanism to the issues of recognition successfully. Experimental result shows that the proposed method has been tested on the **UMN dataset** and outperforms the existing used methods, which proves the efficacy of the proposed method. The proposed approach cannot only fully extract the deep features of video

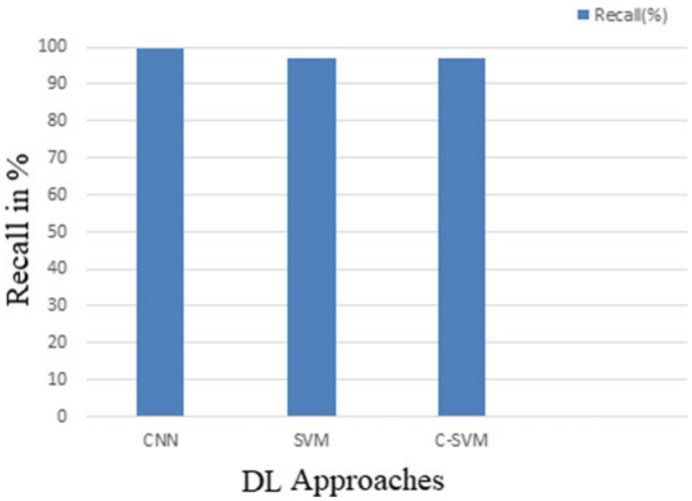

Fig. 7 Analysis on basis of recall

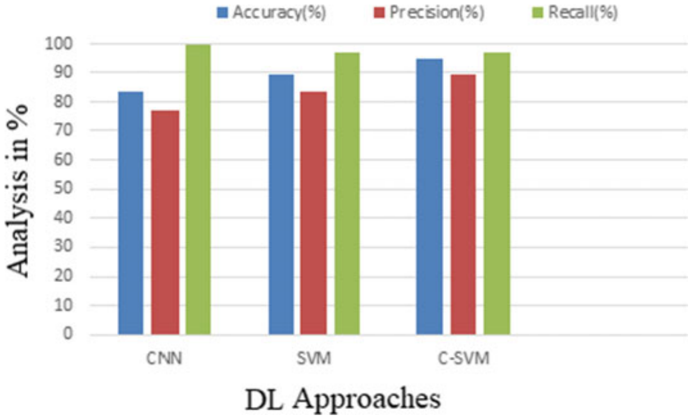

Fig. 8 Overall performance analysis of the proposed system

frames, but also focus on behavioral features that have a greater impact on results. So, it has a greater potential compared with common deep learning and traditional manual feature extraction methods. However, due to the large amount of calculation, real-time performance of this method is difficult to be applied to the multi-channel recognition system with high real-time requirements, this will be the focus of our future research.

References

1. Amrutha CV, Jyotsna C, Amudha J (2020) Deep learning approach for suspicious activity detection from surveillance video. 978-1-7281-4167-1/20/$31.00 ©2020 IEEE
2. Dubey S, Boragule A, Jeon M (2020) 3D ResNet with ranking loss function for abnormal activity detection in videos. IEEE
3. Sultani W, Chen C, Shah M (2019) Real-world anomaly detection in surveillance videos. In: Computer vision and pattern recognition (CVPR)
4. Kavikuil K, Amudha J (2019) Leveraging deep learning for anomaly detection in video surveillance. Advances in intelligent systems and computing
5. Pang H, Li H (2018) Intelligent detection simulation for crowded pedestrian abnormal behavior. Comput Simul 35:405–408
6. Cosar S, Donatiello G, Bogorny V, Garate C, Alvares LO, Bremond F (2017) Toward abnormal trajectory and event detection in video surveillance. IEEE Trans Circ Syst Video Technol 27:683–695
7. Mnih V, Heess N, Graves A. Recurrent models of visual attention. In: Advances in neural information processing systems, Montreal, pp 2204–2212
8. Ding L, Fang W, Luo H, Love PED, Zhong B, Ouyang X (2018) A deep hybrid learning model to detect unsafe behavior: integrating convolution neural networks and long short-term memory. Autom Constr 86:118–124
9. Singh D, Mohan CK (2017) Graph formulation of video activities for abnormal activity recognition. Pattern Recogn 65:265–272