Anomaly Detection in Videos Using Deep Learning Techniques

[1]Rakesh C S, [2]K P KARTHIK, [3]ISHWARAPPA HAVIN, [4]Puneethraj B S, [5]MANU Y M

[1]Student, [2]Student, [3]Student, [4]Student, [5]Assistant Professor,

Computer science department

B G S Institute of technology B G Nagara, Mandya(di),Karnataka, India.

Abstract: People's safety in a public place is very important and this can be attained with the help of anomaly detection. This paper presents an approach to automatically detect abnormal activities in crowded scene. In light of this, we are developing a model in deep learning algorithms namely CNN and VGG16. We are collecting the anomaly and normal CCTV videos to build this project. Then we train with those videos with our algorithms to achieve best precision.

***IndexTerms* - CNN, Deep Learning, Anomaly detection, VGG16.**

I. INTRODUCTION

Public places are usually crowded, such as airports , train stations, theme parks and shopping centers. Due to urbanization, the number of people grows each year. Such rise is causing the risk of tragic incidents in crowded public places to increase. These places should be monitored constantly to avoid these dangerous situations and any unusual activity that is being spotted should be reported back immediately. Public health and security are primary considerations for an study of crowds. The visual screens are continuously observed by human operators to detect any event of interest that becomes challenging to tackle the whole time. This process seems to be more taxing than it sounds and therefore, researchers are developing an automatic system that allows the user to monitor the scene.

The computer vision and signal processing groups have done substantial work in the area of anomaly detection in crowds. Recently, several attempts have been made to manipulate deep learning models to escape any complicated methods for the extraction and processing of hand crafted apps. Given extensive research work and progress in this field, as discussed in, there are deficiencies such as the availability of ground reality, form of anomaly etc. The computer vision group also faces a variety of obstacles in designing successful solutions for detecting anomalies. This includes cameras being unavailable, adverse weather conditions, problems with night vision etc. The researchers are unable to detect the abnormal scenarios often because of these obstacles. Smart phones are suitable alternatives for other issues, as discussed in, in order to fix these issues. Increasing the number of smart phones has enabled the signal processing community to analyze and understand crowd dynamics. You may manipulate the data collected from these smart phones to expose knowledge about individual activities and group dynamics within a crowd.

Sensors such as accelerometer , gyroscope, bluetooth, proximity sensor, camera, ambient light sensor etc. within smart phones help to capture more accurate videos. These sensors acquire data which helps to understand people's behavior in crowds. Most research can be done relating to the identification of single human action structures. Such systems can catch events such as walking, sitting, biking, swimming etc. The main aim of the system developed and proposed by this paper is to detect unusual activities in a crowd by processing synchronously acquired non-visual data from a bunch of individual smart phones at the same time. This is the first work to our knowledge that exploits non-visual data acquired from mobile accelerometers and gyroscopes to detect group based anomaly in crowds.

One of this paper's key contributions is the use of smart phones embedded with rich sensors to collect data collectively and detect anomalous behavior in crowds. In addition, this paper provides a complete dataset along with video-based ground truth, acquired from accelerometer , gyroscope, and multiple fixed cameras, which could be used for research purposes to analyze individual behaviors in crowd.

The remainder of the paper is structured as follows: Section 2 analyzes the state of the art while Section 3 explains the new idea 's methodology. Section 4 reports the performance of the proposed system and finally , Section 5 concludes and discusses future directions for further improvements to the research.

II. ANOMALY DETECTION RECENT SURVEYS

Deep learning algorithms allow the systems to track behavior based on actual data. It is possible to develop algorithms that allow computers to view behaviors learned from the trained model. Deep learning algorithms are used on a given video to study abnormal instances. Could train multidata algorithms. This concept is used to identify and track real-time anomalies. Deep learning is important to the training and study of anomalous behaviour.

A few important findings in this area of study have been published over the last ten years or so. The authors of [124] used visual trajectories to analyze the identification, tracking, scenario analysis and behavioural perception. The study addressed surveillance, knowing the behaviour and detecting incidents from the video. The study discussed is may be the first work covering strategies for anomaly detection. To find phenomena it includes individuals, including detection processes, learning processes and scenario simulation. From the point of view of object identification, monitoring and analysis of actions detailing the success of the last decade of works, this was portrayed with an object oriented approach .Multi-camera study on surveillance in multi-camera setups presented in the research. Authors addressing events that are regarded as a type of anomalous events needing urgent intervention, arising accidentally, rapidly and spontaneously. The work discussed in [144] addresses machine vision related implementations from the perspective of intelligence, intelligence and law enforcement. The analysis described in [181] addresses the characteristics of human behavior and mechanisms for conduct awareness. Authors of [25] describe the task of knowing human nature through human actions and experiences. Intelligent camera systems have been studied in[105], covering facets of analytics. Surveillance systems were introduced in [213], with different implementation areas. Datasets used for identification of phenomena are covered in [140]. In [115], Scientists presented anomalous human behaviour recognition analyzes with a emphasis on task representation and emulation, object extraction approaches, classification and behavior analysis models, performance assessment techniques and video surveillance system sample repositories. Summarizes major computer-based vision studies undertaken over the last 10 years. Summarizes significant computer-based vision research conducted in the past 10 years. In our sample we concentrate in particular on the anomaly detection studies. Anomalies are by their very definition subjective. The principles used in detections of phenomena cannot be uniformly generalized across various situations of operation. From the application viewpoint, we examine the capabilities of anomaly detecting approaches utilized in road camera surveillance.

III. METHODOLOGY

We are using two separate models in the method to identify the videos of anomaly. The model is convolutional neural networks (cnns Network(CNN), and architecture VGG16. We train the pattern in this project with videos of real time anomaly. We construct the model CNN and VGG16 independently but we use the same data for both. So that we can able to verify the performance of the models here.

3.1 CNN Model

A convolutional neural networks (cnns Network (CNN / ConvNet) is a Deep learning algorithm capable of capturing an input image, assigning context (learnable weights and biases) to various aspects / objects in the image, and distinguishing one from another. In a ConvNet the pre-processing requirement is much lower than other algorithms of classification. While filters are hand-engineered in primitive ways, ConvNets has the ability to know with appropriate planning those filters / characteristics.

In this project we are using conv2, max pooling, dropout, relu, softmax like all the layers and activation functions.

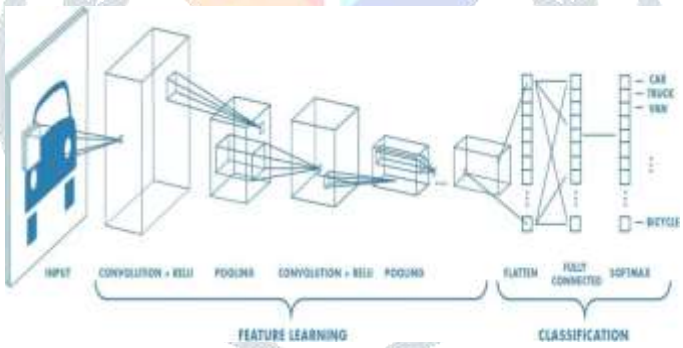


Fig 1: CNN Structure

3.2 VGG 16 Model

VGG16, introduced by K, is a blueprint for the convolutionary neural network. Simonyan, A. Zisserman, of Oxford University, in the article "Very Broad Convolutionary Object Recognition Networks." The algorithm reaches 92.7 percent of the top-5 measurement accuracy in ImageNet, which is a dataset of over 14 million images belonging to 1000 categories. It was one of the best-known model ILSVRC-2014 it had submitted. It builds on AlexNet by replacing large kernel filters (11 and 5 respectively in the first and second convolution layers) with miniature, one by one, 33 kernel-sized filters. For weeks, VGG16 was conditioned, and used NVIDIA Titan Black GPUs.

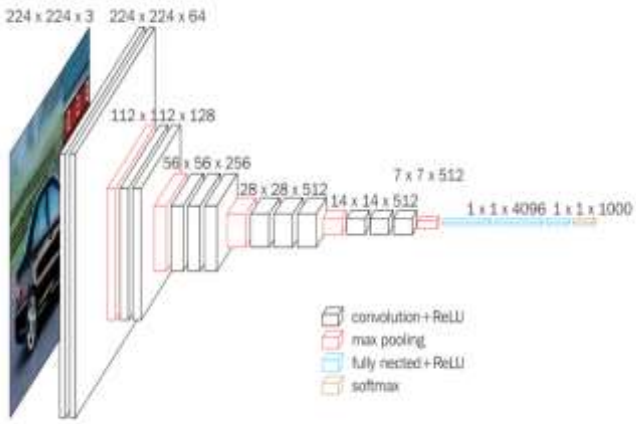


Fig 2: VGG 16 Structure

IV. IMPLEMENTATION

This system is developed using Python programming with Anaconda framework. Datasets are collected from UCF Center for Research in Computer Vision. Videos labeled according to their relevant classification then feed into the system. Videos are trained and tested with CNN Model and VGG 16.

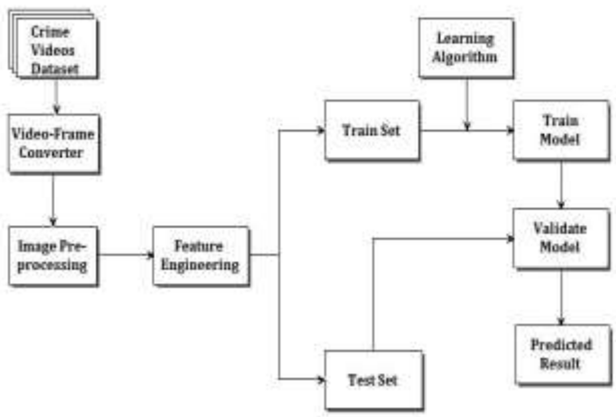


Fig 3: Anomaly Detection System Flow

4.1 Implementation Steps

- Collect the data set(videos)
- Using cv2 we extracted the frames from videos
- Once we get the frames we are going to use cv2 for resize all the images into (64,64,3)
- Now build the algorithms CNN and VGG16.
- Split the data into training and testing.
- Using training data we are going to train our CNN and VGG16
- Once we trained the model utilizing the test data we are going to evaluate our model

4. 2 DATASET

We gather the dataset from this hyperlink https://www.Crcv.Ucf.Edu/projects/actual-international/. We use five categories: ***Arson, Robbery, Fire, Protection, and Routine***. We have 25 videos in each category totally one hundred twenty-five videos that we used on this venture.

Real-world Anomaly Detection in Surveillance Videos

Surveillance recordings can capture a variety of plausible phenomena. Within this article, we recommend investigating phenomena with the aid of manipulating both anomalous and daily images. To stop annotating the anomalous fragments or excerpts in educational films that can be very time-consuming, we recommend learning anomaly by the more than one sample ranking system by exploiting weakly classified educational films, i.e. The training labels (anomalous or normal), instead of the clip-level, are at video level. Throughout our methodology, we bear in mind standard and anomalous motion pictures as bags and video segments as instances in a couple of examples acquiring awareness of (MIL), and regularly analyze a deep anomaly rating variant forecasting strong anomaly ranks for anomalous video segments. In addition, we incorporate sparsity and transient smoothness restrictions within the rating failure feature to find anomaly higher all through preparation.

4.3 UCF-Crime Dataset

To test our strategy, we compile a brand new, huge-scale dataset, dubbed UCF-crime. It includes long uncontrolled surveillance films covering 13 real-world anomalies, including abuse, arrest, arson, assault, road accidents, burglary, explosion, fights, robbery, shooting, stealing, shoplifting, and vandalism. Such findings are selected because they have a huge effect on public safety. Please seek advice from our report for more information on the UCF-Crime dataset. Here is a short summary of each anomalous event.

Video frames for Arson

Example of This event contains Arson video showing people deliberately setting fire to property.

Video Frames for Burglary

Example of This event consists of Burglary video that show humans (thieves) getting into a constructing or residence on the way to dedicate theft. It no longer consists of exerting pressure on humans.

Video Frames for Fighting

Example of this event contains Fighting video, which shows two or more people attacking on each another.

Video Frames for Explosion

Description of This event contains Explosion video, showing destructive event of something blowing apart. This case does not contain images in which a human intentionally causes a fire or starts an explosion.

Video Frames for Normal

Example of This event contains Normal Event video where no crime occurred. These videos include both indoor and outdoor scenes (such as a shopping mall), as well as day and night scenes.

Classification in deep learning refers to a problem of predictive modeling where a class mark is predicted for a given example of input data. Types of classification problems include: Classify whether it is Anomaly or not, provided an example.

4.1 VGG16 & CNN Model Implementation

We discussed about the dataset, once we collect dataset from that website, we have to arrange data based on the classification and make it ready for training process. The frames need to be extracted from videos using OpenCV library for training process. Once we extracted frames it doesn't mean all the frame size of a pixel is same, so it has to be resize all the frames into (64*64*3) using OpenCV.

After these process dataset will be ready, then we build model. In convolutional neural networks (CNN) it make use of conv2d, maxpooling2d, dropout, dense layer and softmax, relu activation functions, using this model will be build. The test result shows this model achieve 88% accuracy. Next we 're building the model VGG16, it's a pre-trained model with the same dataset that gives overall accuracy of 90 percent. We suggested a two model in this paper each achieving the highest accuracy.

V. RESULTS & ANALYSIS

Anomaly detection in videos is one of the emerging requirements in this jiffy. Using deep learning technique CNN and VGG 16 we build the model to classify the Anomaly in videos and the results shows 88% accuracy and 90% accuracy correspondingly with those two models.

5.1 Classification Report (CNN)

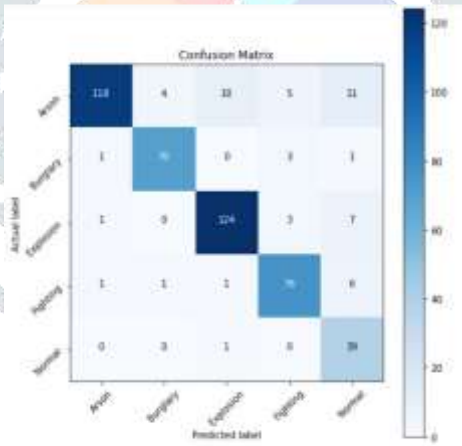
```
print(classification_report(y_true=y_true,y_pred=y_predict))

              precision    recall  f1-score   support

           0       0.98      0.80      0.88       148
           1       0.93      0.93      0.93        75
           2       0.91      0.92      0.92       135
           3       0.87      0.89      0.88        85
           4       0.61      0.97      0.75        40

    accuracy                           0.88       483
   macro avg       0.86      0.90      0.87       483
weighted avg       0.90      0.88      0.89       483
```

5.2 Confusion matrix (CNN)



5.3 Classification report (VGG 16)

```
print(classification_report(y_true=actual,y_pred=y_predict))

              precision    recall  f1-score   support

           0       0.95      0.91      0.93       148
           1       0.95      0.93      0.94        75
           2       0.83      0.99      0.90       135
           3       0.99      0.86      0.92        85
           4       0.78      0.62      0.69        40

    accuracy                           0.90       483
   macro avg       0.90      0.86      0.88       483
weighted avg       0.91      0.90      0.90       483
```

5.4 Confusion matrix (VGG 16)

5.5 Overall Accuracy for CNN and VGG16

- CNN gives us 88% accuracy
- VGG16 gives us 90% accuracy.

VI. CONCLUSION AND FUTURE SCOPE

In this paper, a deep learning based models CNN & VGG16 are proposed for detecting anomalies in surveillance videos. To train and test the models 125 videos are used. The result shows VGG 16 slightly performing better than CNN model. Video classification is much differing from Image classification and also it is complex, in video classification achieving 90% accuracy shows our model works perfectly.

These models are created and tested using Python on desktops, actual usage of video anomaly detection is required in CCTV camera systems, in future these models can be inserted in hardware chips and it can be added in hardware system to provide Artificial Intelligence (AI), so that when anomaly detected hardware itself sent a alert notification to the corresponding departments.

REFERENCES

[1] S. Zhou, W. Shen, D. Zeng, M. Fang, Y. Wei and Z. Zhang, "Spatial-temporal convolutional neural networks for anomaly detection and localization in crowded scenes", *Signal Processing: Image Communication*, vol. 47, no. Supplement C, pp. 358-368, 2016.

[2] C. Piciarelli, C. Micheloni and G. L. Foresti, "Trajectory-based anomalous event detection", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1544-1554, Nov 2008.

[3] H. Tan, Y. Zhai, Y. Liu and M. Zhang, "Fast anomaly detection in traffic surveillance video based on robust sparse optical flow", *2016 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp. 1976-1980, March 2016.

[4] F. Li, W. Yang and Q. Liao, "An efficient anomaly detection approach in surveillance video based on oriented GMM", *2016 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp. 1981-1985, March 2016.

[5] X. Tang, S. Zhang and H. Yao, "Sparse coding based motion attention for abnormal event detection", *2013 IEEE International Conference on Image Processing*, pp. 3602-3606, Sept 2013.

[6] J. Kim and K. Grauman, "Observe locally infer globally: A space-time mrf for detecting abnormal activities with incremental updates", *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2921-2928, June 2009.

[7] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks" in Advances in Neural Information Processing Systems 25, Curran Associates, Inc., pp. 1097-1105, 2012.

[8] L. Wang, Y. Qiao and X. Tang, "Action recognition with trajectory-pooled deep-convolutional descriptors", *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4305-4314, June 2015.

[9] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury and L. S. Davis, "Learning temporal regularity in video sequences", *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 733-742, June 2016.

[10] S. Hochreiter and J. Schmidhuber, "Long short-term memory", *Neural Comput.*, vol. 9, no. 8, pp. 1735-1780, Nov. 1997.

[11] I. Sutskever, O. Vinyals and Q. V. Le, "Sequence to sequence learning with neural networks" in Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2 ser. NIPS'14, Cambridge, MA, USA:MIT Press, pp. 3104-3112, 2014.

[12] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-k. Wong and W.-c. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting" in Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1 ser. NIPS'15, Cambridge, MA, USA:MIT Press, pp. 802-810, 2015.

[13] M. D. Zeiler, D. Krishnan, G. W. Taylor and R. Fergus, "Deconvolutional networks", *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2528-2535, June 2010.

[14] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", *CoRR vol. abs/1412.6980*, 2014.

[15] V. Saligrama and Z. Chen, "Video anomaly detection based on local statistical aggregates", *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2112-2119, June 2012.

[16] D. Xu, E. Ricci, Y. Yan, J. Song and N. Sebe, Learning deep representations of appearance and motion for anomalous event detection, BMVC. BMVA Press, pp. 8.1-8.12, 2015.

[17] V. Mahadevan, W. Li, V. Bhalodia and N. Vasconcelos, "Anomaly detection in crowded scenes", *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1975-1981, June 2010.