Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 8, July, 2021:7465 – 7470

Research Article

Anomaly Detection And Classification Using Deep Learning Techniques

[1*]S. Jothishri, [2]S. Jothilakshmi, [3]G.Jawaherlalnehru and [4] S. Nagamani

Abstract

To protect and control the public and private crowd, the anomaly detection system is introduced. In addition, the security mechanism should be improved. When an abnormality is detected, an anomaly detection system must be used to warn the crowd. The government, particularly in private and public congested areas, requires a low-cost solution to offer safety today. Thus the Deep Learning based computer vision technique provides efficient methods for private and public safety. The anomaly detection system will be very cooperative if the event videos on the web can be routinely categorized into predefined classes. Video event holds visual information of anomaly which can be detected on a frame basis using Convolution Neural Network (CNN). The main goal of the proposed system is to recognize anomaly on various crowd videos. The proposed system has applied CNN baseline and VGG-16 for crowd video anomaly detection. The computation result of the anomaly detection system is analyzed and quantified as a good results.

Key words: Convolution Neural Network (CNN), Deep Learning, VGG-16, CNN baseline, Anomaly Detection.

1. INTRODUCTION

Now a days the developed countries are improving the security system to defend and manage the public and private crowd. People need security in mass assemblies, public and private events. Anomaly detection is a perilous issue in a crowded places. Since Anomaly has made injuries and damages in public area. Sometimes if any anomaly has occurred in a crowded area, the anomaly detection is essential to protect people and the environment without any severe impairment. When the anomaly is perceived, alerting crowd by an alerting system is very imperative. Particularly in private and public crowded area, the government needs a solution to provide safety now- a- days with low cost. Thus the Deep Learning based computer vision technique [1],[2],[3],[4] provides efficient methods for private and public safety. And also the technique affords real time video surveillance system for crowd management.

The proposed system of anomaly detection implements an essential and requisite phase in the process of assessing the video events (Musical Function, Public Meeting, Bazaar, and Protest). Video event [5,6] holds visual information of anomaly which can be detected on a frame basis using Convolution Neural Network (CNN). In the proposed system CNN model[7] has been initiated and implemented with high resolution video event frames. The Figure 1 shows the steps of anomaly detection system. The huge amount of trained data is used for working out the CNN model. The anomaly detection system[8] will be very cooperative if the event videos on the web can be routinely categorized into predefined classes.The alerting

[1]Associate Professor, [2]Associate Professor, [3] Associate Professor, [4] Associate Professor
[2]Swarna Bharathi Institute of Science & Technology, Khammam , *sri.sankar6@gmail.com
[3]Annamalai University, Tamilnadu,[3]St. Martin's Engineering College, Secunderabad, Telangana,
[4]Swarna Bharathi Institute of Science & Technology, Khammam

system is in different forms such as tones, voice and alert message. After detecting anomaly in the crowd, the alarm system should intimate through message or make sound automatically.

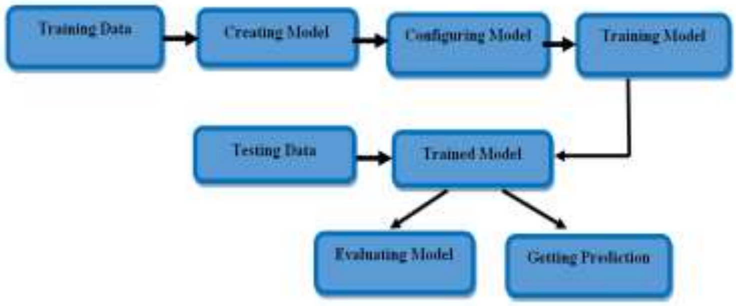

Figure1: Steps involved in the anomaly detection system.

II. RELATED WORK

Solmaz *et al.*, (2018) [9] proposed a framework to identify multiple crowd behaviors through stability analysis for dynamical systems, without the need for object detection, tracking, or training. The proposed method is deterministic and cannot capture the randomness inherent in the problem without a stochastic component.

Chen *et al.*, (2013)[10] proposed a real-time system, in this, each isolated region is considered as a vertex. A human crowd is thus modeled by a graph. Delaunay triangulation is used to systematically connect vertices. Therefore the problem of event detection in human crowds are formulated by measuring the topology variation of consecutive graphs in temporal order.

Nuria Pelechano *et al.*, (2006) [11] have shown a significant improvement in evacuation rates when using inter-agent communication. The grouping behavior that emerges can also be observed when there are a high percentage of dependent agents in the crowd. Only a relatively small percentage of trained leaders yield the best evacuation rates. These results can be realized in real-time with either simple 2D or 3D viewer. This approach could be improved by adding individualism into Helbing's model so that agents would have different local motions depending on their roles.

Crowd analysis **(Andrej Karpathy *et al.*, 2014)[12]** had incredible advancement from crowd scene understanding. The scene understanding is a must in computer vision based technique. The standard video classification method involves three main levels, first local visual features that describe a region of the video contents are extracted either densely or at a sparse set of interest points. Next, the features get combined into a set sized video stage description. Finally it quantizes all features using a learned dictionary and accumulates the visual phrases over the image into histograms of varying spatial-temporal positions. Using local features to learn body parts is the best approach to human detection. Part-based approaches that model an object as a rigid or deformable configuration of parts are shown to be very effective for occlusion handling.

Video clips with different characteristics mainly captured in different scenes to identify objects based on scenes in the crowd from various cultures **(Dehghan *et al.*2014)[13]**. Background modeling of each scene is essential thing in a video content analysis system. Thus the system determined to separate the video clips in order to identify scene types.

III. PROPOSED SYSTEM

The main goal of the anomaly detection system is to recognize anomalies (abnormal events) in various crowd videos as shown in Fig.2 In the proposed system, the CNN model detects four abnormal events namely fire, fighting, protest, and running with fear in the crowd video. In this work, two architectures namely, CNN baseline and VGG-16 models are proposed for crowd video anomaly detection.
The performance of both the architectures is evaluated and compared to identify the best model for anomaly detection.

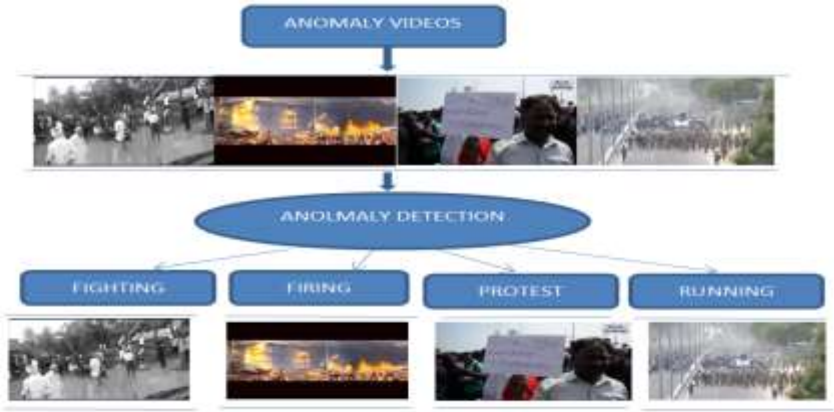

Fig. 2 Anomaly Detection in Crowd Videos

3.1 Proposed CNN Baseline Model for Anomaly Detection

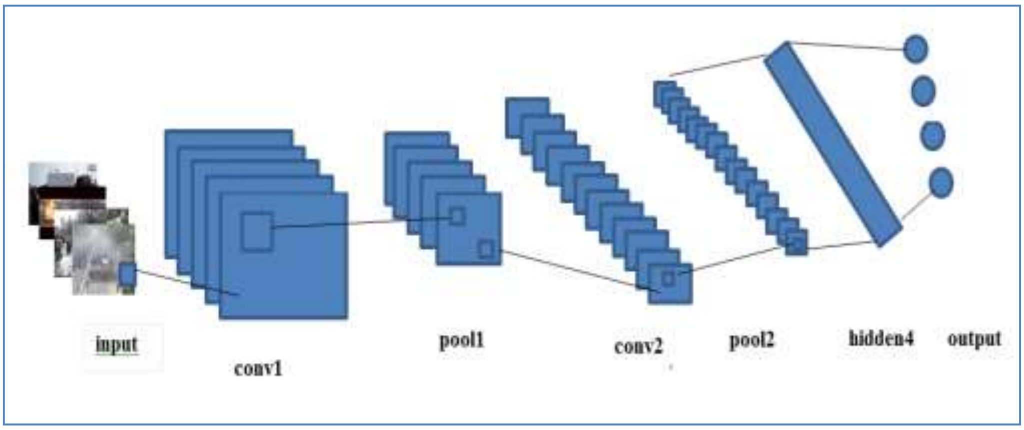

Fig. 3 Proposed CNN Architecture for Crowd Density Estimation

Fig. 3 shows the architecture of the CNN baseline model used in the anomaly event classification system. In the input layer, each pixel of the input image is represented by a neuron. The hidden layer consists of three convolution layers.

Table 1. Parameters of CNN Baseline Model

Parameter	Description
Input size	148 ×148

Colour channels	3 (RGB)
Filter size	3 × 3
Activation function	Rectification Linear Unit (ReLU)
Pooling layer	Max Pooling
Max pooling size	2 × 2
Stride	1
Final layer	Softmax
Dropout	0.5

The convolution output size must be the same as the size of the input image (150 × 150). The convolution operation is performed for RGB color channels of the input image. The output softmax layer classifies four anomaly events. The architecture used for the event detection system is used in this anomaly event detection task as well. Table 1 shows the parameters of CNN model for detecting four different anomalies in task.

3.2 VGG-16 Model for Anomaly Detection

Table 2 shows the parameters of VGG-16 modelfor detecting four different anomalies in this the task.

Table 2 Model Parameters of VGG-16

Parameter	**Description**
Input size	224 × 224
Colour channels	3 (RGB)
Filter size	3 × 3
Activation function	Rectification Linear Unit (ReLU)
Pooling layer	Max pooling
Max pooling size	2 × 2
Stride	2
Final layer	Softmax (multi class)
Total number of layers	16
Frozen layers	1to12

IV EXPERIMENTS AND RESULTS

A new data set has been created with 5000 images. The abnormal event frames are collected from different events namely Jallikattu, Marriage, Shopping, Sports, Temple, Bazzaar, and Protest. In the data set, randomly selected 1000 images for each anomaly category are a training set and 500 images from all four categories are a validation set. The test set has 500 images to test the performance of the anomaly detection system. From each anomaly category the test set has 100 frames, and 100 false images collected from other videos also stored in a test container.

Deep learning is established for anomaly detection by training the CNN models through 4000 frames of training samples. The validation phase used 500 frames from four differentevents. Table 3 shows the validation results of CNN models for a different number of epochs. The validation results show that baseline model performs well. Fig. 4 shows the training and validation accuracy of the CNN Baseline and VGG-16 model for different epochs.

Table 3 Effect of Epochs on Validation Accuracy

Epoch	Baseline Model Accuracy (%)	VGG-16 Accuracy (%)
21	89	82
22	89	84
23	90	85
24	92	86
25	92	86
26	96	86

The validation results of the CNN baseline and VGG-16 model are illustrated in Fig 4. The system shows significant variations in performance for the CNN baseline for each epoch. The CNN baseline gives 92% validation accuracy

.

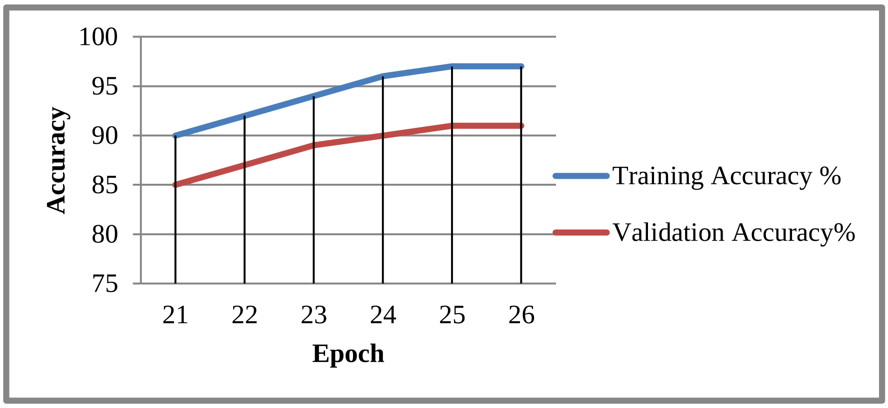

Fig. 5 Validation Accuracy of Anomaly Detection System

During the testing phase, the proposed model classifies and labels the frames in the test set. The performance of the anomaly detection system is analyzed and quantified as precision, recall, and F-Score. The results of CNN baseline and VGG-16 models are given in Table 4. The results show that the CNN baseline model performs well with 0.92 F-score when compared to the VGG-16 model which shows 0.87 F-Score.

Table 4 Performance of Anomaly Detection System

Model	Precision	Recall	F-Score
CNN Baseline	0.92	0.93	0.92
VGG-16	0.86	0.87	0.87

V. CONCLUSION

The two different tasks of crowd analysis are introduced two different CNN classifiers (CNN Baseline and VGG-16). The CNN Baseline is used with three layers with suitable modified parameters. The VGG-16 model is introduced with a transfer learning method. The CNN performs well for all tasks of crowd management system when compared to VGG-16. The performance of the CNN baseline and VGG-16 model are evaluated as true positive, true negative, false positive and false negative. The CNN baseline method gives 92 percent of accuracy.

References

1. Amin Seyyedi, S. and Ivanov, N. (2012), Statistical Image Classification for Image Stenographic Techniques, International Journal of Image, Graphics and Signal Processing (IJIGSP), Vol. 6(8), pp. 19-24.
2. Anagha V., Joshi, and Nikita Hattiwale,(2017), Optimal Fire Detection Using Image Processing, International Journal of Recent and Innovation Trends in Computing and Communication, Vol.5, pp.248-252.
3. Brahmbhatt and Samarth,(2013), Embedded computer vision: running OpenCV programs on the Raspberry Pi, Practical OpenCV. Apress, Berkeley CA, pp.201-218.
4. Carolina, Garate, François Bremond, and Piotr Bilinsky,(2009), Crowd event recognition using hog tracker, Proceedings of Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance. IEEE.
5. Chan, C.S., Lim, M.K., and Kok, V.J.,(2016), Crowd behavior analysis: A review where physics meets biology, Neurocomputing, Vol.177, pp.342-362.
6. Chen Sun and Ram Nevatia, (2013), Large Scale Web Video Event Classification by use of Fisher Vectors, University of Southern California, Institute for Robotics and Intelligent Systems Los Angeles, USA.
7. Ding, L., Fang, W., Luo, H., Love, P.E., Zhong, B., and Ouyang, X.,(2018), A deep hybrid learning model to detect unsafe behavior: integrating convolution neural networks and long short-term memory. Automation in Construction, Vol.86, pp.118-124.
8. Divya, B.and Dhanujalakshmi, R. (2017), Image Processing Based Fire Detection System using Raspberry Pi System, International Journal of Computer Science and Engineering, Vol.4, pp.18-20.
9. Solmaz, B.(2018), Video-based detection of abnormal activities in crowd using a combination of motion-based features, In Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies II International Society for Optics and Photonics.
10. Chen, Duan-Yu, and Po-Chung Huang, (2013), Visual-based human crowds behavior analysis based on graph modeling and matching, IEEE Sensors Journal, Vol. 13(6), pp.2129-2138.
11. Pelechano, Nuria, and Norman I. Badler, (2006), Modeling crowd and trained leader behavior during building evacuation. IEEE computer graphics and applications Vol. 26(6), pp. 80-86.
12. Karpathy., George Toderici, and Sanketh Shetty, (2014), Large-scale Video Classification with Convolutional Neural Networks, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.1725-1732.
13. Dehghan, A., Idrees, H., Shah, M., and Zamir, A.R., (2014), Automatic Detection and Tracking of Pedestrians in Videos with Various Crowd Densities, In Pedestrian and Evacuation Dynamics, pp. 3-19.