VIDEOBEHAVIOR POSSIBLE IDENTIFICATION AND RECOGNITION OF ABNORMALITIES AND NORMAL BEHAVIOR PROFILING FOR ANOMALY DETECTION USING CNN MODEL

VENU MAJJI[1], DR. VANITHA KAKOLLU[2], MPJ SANTOSH KUMAR[3], K NAGA SOUJANYA[4], DR. B. KANTHAMMA[5], DR. G. BABU RAO[6]

venumajji@asg.com[1], vkakollu@gitam.edu[2], smadabat@gitam.edu[3], sketha@gitam.edu[4], kbokka@gitam.edu[5], bgumpina@gitam.edu[6]

UX Manager[1], Assistant Professor[2,3,4,5,6]

[1]ASG Technologies, [2,3,6]Department of CS, [4]Department of CSE, [5]Department of EECE, GITAM (Deemed to be University) Visakhapatnam

ABSTRACT

The aim of this Paper is to unravel the matter of modeling video behavior recorded in surveillance videos to be used in online normal behavior recognition and anomaly detection applications. With non-manual marking of the training data collection, a replacement architecture is made for automated behavior profiling and online anomaly sampling/detection. The subsequent are the core components of the framework supported discrete scene event detection, a compact and efficient behavior representation method is developed. Modeling each pattern employing a Dynamic Bayesian Network is employed to gauge the similarities between behavior patterns (DBN). A completely unique spectral clustering algorithm supported based on unsupervised model selection and have selection on the eigen vectors of a normalized affinity matrix is employed to get then actual grouping of behavior patterns. To detect abnormal behavior, a runtime accumulative anomaly measure is implemented, while normal behavior patterns are recognized when adequate visual evidence is out there supported a web survey. This enables the fastest possible identification and recognition of abnormalities and normal behavior. Experiments with noisy and broken data sets gathered from both indoor and outdoor monitoring scenarios show the efficacy and robustness of our approach. It's is demonstrated that in detecting anomaly from an unseen video, a behavior model trained with an unlabeled data set out performs those trained with an equivalent but labeled dataset.

Keywords*: Dynamic Bayesian Network, Anomaly, CNN, Adaptive Video Conversion*

1. INTRODUCTION

In order to protect and control the public and private crowds, developing countries are upgrading their security systems. In a crowded location, detecting anomalies can be risky. Since the Anomaly has caused injury and property damage in the public domain. When an anomaly occurs in a crowded area, anomaly detection is often needed to keep people and the environment healthy. When an anomaly is detected, an alerting device must be used to warn the crowd. The alerting device comes in a variety of formats, including tones, speech, and text. When an anomaly in the crowd is detected, the warning system can immediately send out a message or tone. The government, especially in private and public crowded areas, requires a low-cost solution to provide protection. In large crowds, public and private events, people need protection. As a consequence, the Deep Learning-based computer vision approach [2][3][4] offers a wide variety of capable methods for private and public safety. The method also allows for a real-time video monitoring system for crowd control. The proposed anomaly detection system serves as an important and necessary step in the process of

Journal of Theoretical and Applied Information Technology
31st July 2022. Vol.100. No 14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

evaluating video events (Musical Function, Public Meeting, Bazaar, and Protest). If the massive detection in event videos on the web can be systematically grouped into predefined groups, the anomaly detection system would be very cooperative. Convolution Neural Networks can be used to detect anomalies in video events on a frame-by-frame basis (CNN). The proposed framework includes a CNN model that has been initialized and implemented with high-resolution video event frames. The CNN model was built using a massive amount of trained data. Automated anomaly detection is extremely useful in minimizing the amount of data that must be manually analyzed by concentrating attention on a small portion of the data while ignoring large quantities of irrelevant data. The problem of anomaly detection, on the other hand, is highly interpretable, and research efforts are dispersed not only in terms of method, but also in terms of understanding of the problem, expectations, and goals. By reviewing the problem formulations and approach methods used in anomaly detection research as applied to automate surveillance, this analysis will attempt to add synergy to these diverse efforts.

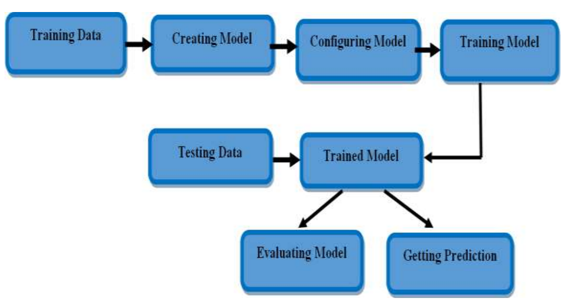

Fig1: Anomaly System Model

2. RELEATED WORK

Human activity identification has a wide range of uses. Recognizing a human activity may be useful in hospitals for monitoring patients and in public areas for providing protection and security. The installation of CCTV (Close Circuit Television) cameras has become mandatory due to the rise in various types of crime rates. However, manpower is needed to track the surveillance cameras. As a result, extensive research has been undertaken to simplify the monitoring process and detect suspicious human behavior. The Viola Jones algorithm is used to detect and recognize objects, which in this case is the human face, as stated in the research paper [1] The Viola Jones algorithm can be used to detect and classify items, such as a human face, in an examination hall to reduce invigilator workload. This is the human face in this case. The integral image is used to represent the features, which are then computed using pixel-based operations. The summation of the left, top, and affected pixels yields the integral image. The SURF (Speed up Robust Features) algorithm is used to find the interesting point. This approach also finds and matches features. AdaBoost is used to classify the extracted features. Cascade is a classifier that is used for even more complicated classification. As a result, this device detects any unusual human activity in a testing space.

The task is traced as a binary classification problem, normal and unusual [8], in one of the research papers done for anomaly detection, and it has been proven successful in terms of accuracy. Unlabeled videos are useful for training the system since this technique needs very little to no supervision. Introducing auto encoders [2] with the goal of replacing low-level features with learned hierarchical features to reduce the amount of labor needed for feature engineering, resulting in data representation that can support effective machine learning for sparse coding. Additionally, auto encoders have been used to replace it, eliminating the distinction between feature extractions and learning a model for certain features. Allowing the auto encoder to have several layers of hidden unit's results in hierarchical feature learning. The total number of input units is greater than the total sum of output units, which is used to minimize dimensionality. The back propagation method is used to train an auto encoder unsupervised. Recurrent Neural Networks (RNN) function similarly to feed forward networks, but unlike conventional feed forward neural networks, they have a temporal dependence between inputs since they are a series. The output here is influenced not only by the previous input, but also by the entire history of inputs. However, since the gradients disappear, looking back for long sequences is not realistic. As a result, the LSTM (Long Short-Term Memory) architecture was developed, which includes the forget gate, which is also a recurrent gate. This stops back propagated errors from disappearing and bursting. Working on longer sequences that are stacked together for higher level information capture becomes possible as a result. As a result, this anomaly detection approach employs the principle of integrating the temporal sequencer with the spatial feature extractor, resulting in the development of an end-to-end model for anomaly detection that can be trained using only semi-supervised images. Human activity recognition is useful in protection systems because it assists in the identification of suspicious activities. There has been extensive research into various strategies for anomaly detection, one of which is the use of PNN (Probabilistic Neural Networks) [3], which classifies behavior into regular and abnormal categories. Blob analysis is used to track the object (Local Binary Pattern). A broad variety of advances have been made in the fields of machine learning, activity recognition, and anomaly detection for smart surveillance systems. The study's main goal is to differentiate current protocols and draw attention to their major flaws. HMM (Hidden Markov Model) is a

Technique for detecting suspicious activity that is resistant to brightness changes, changes induced by climatic or Weather changes in the outdoors, and background noise. This primarily focuses on object tracking and movement detection. The context subtraction approach is used to modify differences in the bitrates of video streams in the technique, which is based on the probabilistic neural network strategy. Since this is a PNN (Probabilistic Neural Network) with multiple layers, the probabilities for classification of targeted and non-targeted behavior are chosen by the previous layer output participating in the transfer function. This paper claims that the PNN approach for anomaly detection outperforms other techniques such as [1] HMM and Dynamic Bayesian network, whose output must be evaluated in an unstructured situation. [9] For movement detection, Bayesian and optical flow methods were used. The aspect ratio of the object is used to identify and track it. The latent Dirichlet allocation is another strategy that aids behavior recognition. [Twenty] The previously available two-stream CNNs [12] for action recognition achieves state-of-the-art efficiency, but they are too computationally costly for real-time deployment. It is made up of a spatial network and a temporal network that take the RGB image and the discretionary stream as inputs separately. Even if it is done effectively, the measurement of optical flow is the most expensive step. As a consequence, the classical two-stream model, which is built by combining the context subtraction method with the hidden Markov model, has a significant downside that prevents real-time detection. However, in a crowded situation, this approach did not work. It involves comprehensive Saliency detection testing. Even the KNN classifiers [7] were unable to effectively classify the anomalous behaviors, leading to the suggestion of using detectors that can handle a larger number of attributes. For the definition of local block movement with respect to the corresponding reference frame in this work, optional flow is replaced with motion vector [3]. In the context of a deep convolution neural network, it investigates motion vector. However, when specifically trained using those motion vectors, the low resolution and imprecise definition of fine movements causes a depletion of recognition execution efficiency. Several training methods for improving the motion vector have been implemented in the process of resolving this problem. Since they are intrinsically linked, the facts and features extracted in the optical flow CNN must be transferred and used to direct the motion vector CNN. The optical flow CNN's efficiency could be improved as a result of this. Initialization of teachers, supervision transition, and a mixture of the two are some of the methods used to accomplish this. The initialization of the network has a significant effect on the final results. However, there is a chance of losing the information that was initially transferred during the fine-tuning process, which leads to the implementation of supervision transfer, which keeps track of the additional supervision from the teachers' net while the Motion vector CNN is being trained. The final strategy incorporates initialization, which involves the teacher's parameters, with supervision transition to improve efficiency. The classification of the pose template results in the development of shape and motion models. The SVM (Support Vector Machine) algorithm, which is based on a semi-supervised structure, is emphasized. Despite the need for a 2D view, this model fits well for most applications. When comparing the different methodologies for human activity recognition and anomaly detection, it is clear that CNNs perform extremely well in visual recognition. CNNs extract local features from a video sequence of frames and then merge them to create more complex higher order features. The 3D convolution [5] method extracts features by stacking several continuous frames to form a cube into which a 3D kernel is convolved. The motion information is captured by linking feature maps in the convolution layer to several contiguous frames on the previous layer. Since kernel weights are replicated throughout the entire cube, the 3D convolution kernel from frame cube may extract a single form of function. Back propagation and a supervised algorithm with an accuracy of 80-90 percent are used to train this model. As a consequence, the 3D-CNN model shows promise in detecting irregular human behavior.

The spatial multi-scale area convolution 3D Network [6], which uses a temporal feature pyramid to reflect activities of corresponding temporal scales, is another new and powerful technique for activity detection. Using the activity classifier and activity proposal detector, each stage of the pyramid detects behaviors on the temporal scale.

3. PROPOSED WORK

The definition of a simple deep neural network model for anomaly detection has been widely adopted in this work. DNN is a mathematical model that analyses a large number of classified video input frames in general terms. Neural network architecture is based on the basic nerve system found in animal brains. CNN's Network [7][8] is a multi-perception network that is entirely connected. Weights and biases are called network neurons in the critical network. Dot product and neuron weight are two of the network's key processes. As shown in Figure 2, the Convolution Neural Network has two important layers: pooling and completely linked layers. The convolution network's key function is performed by the CNN network layers. At the same time, each neuron in the network layer is linked to a small region of nearby frame data. The receptive field refers to each person small area. The CNN model detects anomalies in video events in the proposed scheme. Frames from images are used to derive anomaly case characteristics. YouTube videos have been added to the latest dataset. For anomaly detection, the features are averaged from video frames and fed into a regular CNN classifier. Anomaly features from the layers of the model structured with different kernels for anomaly detection are used to train the CNN model. Using a feed forward process, the feature representations [9][11] are generated sequentially in a fully connected layer. The proposed framework includes a new dataset of 4000 images from a number of anomaly events that fall into four categories: fire, war, running with fear, and accidents. Figure 4 describes the proposed system's anomaly cases.

The CNN model detects anomalies in video events in the proposed scheme. Frames from images are used to derive anomaly case characteristics. YouTube videos have been added to the latest dataset. For anomaly detection, the features are averaged from video frames and fed into a regular CNN classifier. Anomaly features

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

from the layers of the model structured with different kernels for anomaly detection are used to train the CNN model. Using a feed forward process, the feature representations [10] are generated sequentially in a fully connected layer. The proposed framework includes a new dataset of 4000 images from a number of anomaly events that fall into four categories: fire, war, running with fear, and accidents. Figure 4 describes the proposed system's anomaly cases.

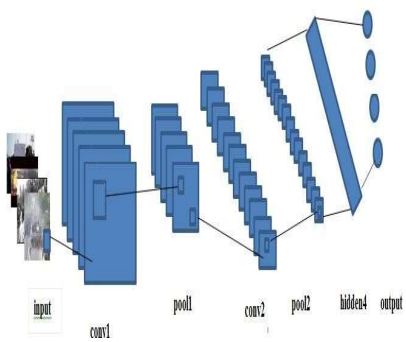

Fig2: Proposed Architecture

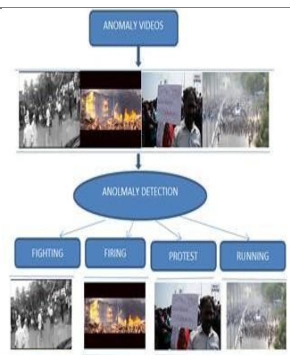

Fig 3: Anomaly Detection

4. METHODOLOGY

In This module the anomaly detection method uses a CNN Architecture to identify *Dataset Collection (SNM)* 2000 data sets and shows significant improvements in the experimental outcome. Based on features developed by the training dataset, the CNN Model provides a detection result. On the basis of an event classifier trained over 2000 frames of video, a Deep Learning model is developed. For each of the four case groups, 500 images were randomly chosen as a training set and 500 images were used as a validation set. After testing, the CNN model had 90% accuracy in detecting anomalies in the validation data collection.

5. IMPLEMENTATION FRAMEWORK

Python 3.5 and the Anaconda Library's required packages are used to implement the well-known recent Deep learning technique in the proposed framework. The OpenCV3.3.0 library can also be used to quickly identify anomalous video and image files. Jupiter Notebook's popular navigator comes in handy when compiling python code for anomaly detection.

(i) Features:

The CNN model is designed with some important parameters in the proposed framework. The activation feature of layers Rely and max pooling layer is used to implement the three convolution layers. There are filters in this layer. The size of the kernel is 22. The model has been educated for four different levels. The output layer has four neurons. Categorical-cross entropy is the network's special activation function for classifying the dataset.

(ii) Training dataset

This project's training process includes ten epochs and 2000 training samples to implement the model for extracting critical features and providing good training. The required video frame dataset is stored in a stack array with a changed size of 250 x250. The dataset's stacked array is converted to a batch file, which is used to feed data to the CNN model during preparation. The model uses 0 and 1 for anomalies and extracts the features over epochs to detect anomalies.

(iii) Testing Dataset

The final step of testing in the CNN model for detecting anomalies involves taking multiple video events and converting them into frames. In a batch file, 50 video events are stored in a stack array with a resolution of 250x250. The event video frames were gathered from various events such as sports, marches, temples, and so on. 25

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

frames are taken from each video and stored in a

Model	True positive	True negative	False positive	False negative
Base line	90	40	40	1
Vgg16	95	30	30	10

test container. Ten false datasets are obtained from other videos and stored in the test container in this process. Testing data is sent to the qualified model from the batch file. The CNN model recognizes type of anomalies and correctly names each of them.

6. RESULTS AND DISCUSSION

Table 1: Compares CNN Baseline And VGG16 Model Validation Data. In The CNN Baseline And VGG 16 Model Models, The System Discovered The Variations Of Each Epoch. Variations Of The CNN Model That Have Been Validated.

Epoch	Baseline Validation	VGG 16Validation
1	63	82
2	80	84
3	81	85
4	84	85
5	83	83
6	82	86
7	78	85
8	80	87
9	82	89
10	84	90

Model	precession	Recall	F1-score
baseline	82.7%	81.5%	80%
VGG16	87%	89%	90%

Table 2: Confusion matrix results for the CNN

Table 3: The computation result of the anomaly detection system is analyzed and quantified as precision, recall, and f1score. The results of CNN baseline and VGG-16 are given in Table

7. CONCLUSION

Using deep learning methods, the proposed method detects anomalies. The device generates voice alerts and alarms in three different ways after detecting the anomaly zone. The approach is effective in identifying the source of the fire. The proposed system can be used in crowded areas with a large number of people. Overall, the precision is greater than 90%. The method may be used to introduce more different forms of anomalies in the future.

REFERENCES

[1]. AbidMehmood et al.," Abnormal Behavior Detection in Uncrowded Videos with Two-Stream 3D Convolutional Neural Networks", April 2021.

[2]. Andrew D. et al., "Adaptive Video Compression ForVideo Surveillance Applications" IEEE conference June 2011

[3]. Oluwatoyin P. et al., "Video-Based Abnormal Human Behaviour Recognition—A Review" IEEE conferenceApril 2016

[4]. B Ravi Kiran, et al., "An overview of deep learning-based methods for unsupervised and semi-supervised anomaly detection in videos" arXiv:1801.03149v2 [cs.CV] 30 Jan 2018

[5]. Yan Zhang, et al., "Temporal Human Action Segmentation via Dynamic Clustering" arXiv:1803.05790v2 [cs.CV] 18 Mar 2018

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[6]. William Robson Schwartz et al., “Novel Anomalous Even tDetection based on Human-object Interactions” In Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2018) -Volume 5: VISAPP, pages 293-300 ISBN: 978-989-758-290-5

[7]. Dr. U. Srinivasulu Reddy et al., "Article Location and Following in Group Condition - A Survey" Procedures of the Global Gathering on Innovative Registering and Informatics (ICICI 2017) IEEE Xplore Agreeable - Part Number: CFP17L34-Craftsmanship, ISBN: 978-1-5386-4031-9Number: CFP17L34-ART, ISBN:978-1-5386-4031-9.

[8]. Yong Shean Chong, et al., “Abnormal Event Detection in Videos using Spatiotemporal Auto encoder” arXiv:1801.03149v2 [cs.CV] Jan 2017

[9]. Karuna B. et al., “Analysis of anomaly detection techniques in video surveillance,” IEEEconference June 2018

[10]. B. Zhang, et al., “Real-time ActionRecognition with Enhanced Motion Vector CNNs,” Apr. 2016

[11]. Sathyashrisharmilhaet al., “Using 3D Convolutional Neural Network in Surveillance Videos for Recognizing Human Actions” The International Arab Journal of Information Technology, Vol. 15, No. 4, July 2018

[12]. Tao Xiang et al., “Video behavior profiling for anomaly detection”, IEEE Trans Pattern Anal Mach Intell 2008 May;30(5):893-908.