IJMRSET

e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 8, August 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

© 2024 IJMRSET | Volume 7, Issue 8, August 2024| DOI: 10.15680/IJMRSET.2024.0708054

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Suspicious Activity Recognize Implementing Deep Learning Approach

Dr D Kirubha, Deekshitha S, Netra, Spoorthi B S

Associate Professor, Department of Computer Science & Engineering, Rajarajeswari College of Engineering, Bengaluru, Karnataka, India

UG Student, Department of Computer Science & Engineering, Rajarajeswari College of Engineering, Bengaluru, Karnataka, India

ABSTRACT: The widget now has video and deep learning capabilities. Using the aforementioned combinations, the area has remarkable methods for distinguishing various suspicious behaviors from live photo surveillance. Human behavior is the most unpredictable, and it is quite difficult to tell whether it is suspicious or normal. In this article, we classified human activities into two categories: normal and suspicious. Common activities include sitting, walking, jogging, and waving. Running, boxing, and fighting are all examples of suspicious activities. We use convolutional neural networks to perform this classification. First, sophisticated picture features are extracted using a convolutional neural network. Convolutional network classification is assessed, and the aggregate layer's final result is extracted before making the final prediction.

KEYWORDS: suspicious activity, deep learning, and convolutional neural networks.

I. INTRODUCTION

In today's environment, we can see that crime has increased despite the presence of cameras everywhere. To detect suspicious activity, a modelmust be created that decreases.

In this situation, the surveillance camera is in the time it takes to notice it and take action in the form of film. The most efficient technique to handle the video is to split it into images and then edit them [1], [2]. There are numerous machine learning methodologies available today for image processing, but as the dataset grows larger, the accuracy drops, thus we turned to deep learning algorithms.

Automated video monitoring can assist avoid violence caused by overcrowding and fighting in public facilities such as parking lots, jails, military bases, mosques, borders, and public transit stops. each other, and those carrying weapons capable ofcausing harm to others, such as bombs, robbery, vandalism, and so on [3], [4].

Video surveillance is an essential component of improving the security of banks and ATMs [5], [6]. The deployment of autonomous security cameras in banks will help avoid armed robberies and heists. ATMs are a popular target for robbers, and autonomous surveillance cameras could help tostrengthen their security [7], [8].

Surveillance cameras may aid in the detection of disruptive behavior among students on campus, such as bullying and fighting [9, 10]. They can alsohelp improve the camera's anti-theft capabilities.

Safety cameras are increasingly being used in small businesses, factories, and shopping centers [12], [13]. They're used to apprehend shoplifters and robbers, as well as to keep armed robberies at bay [14], [15]. Security cameras are also used to track supplies and inventory held in warehouses and to detect employee bribery and theft [16], [17].

Platforms, routes, roads, tunnels, and parking lots are all monitored by security cameras in railways and bus stations. Terrorists may use these areas as a staging ground for explosive attacks by leaving a bag containing explosives [4], [18].

Automated security cameras can detect discarded bags and warn officials, who can then remove them to protect passengers and facilities [19].

Video monitoring can be used to keep an eye on patients in hospitals and elderly people in their homes [20]. It is capable of detecting abnormal behavior in patients, such as vomiting, fainting, or any other irregular behavior [14]. As a result, given the wide range of applications, we must devise a method for detecting suspicious activities in videos [21], [22]. The remaining paper is organized as follows: literature survey is discussed in Section 2, activity classification and the CNN models are presented in Section 3, the proposed framework is presented in Section 4, details of the dataset are presented in Section 5, and results are discussed inSection 6 and we conclude the paper in Section 7.

II. LITERATURE SURVEY

A. Security camera research for detecting violent activity

In this part, we'll go through some of the research that's been done in the field of detecting violent behavior in security cameras. Fighting, vandalism, punching, kicking, scratching, peeping,shooting, and other violent acts are examples.

A non-tracking, real-time algorithm that detects suspicious behavior, is very useful in crowded and public areas [23]. Instead of object tracking, the algorithm keeps track of low-level measurements in a series of fixed spatial locations. This algorithmhas the downside of not providing sequential tracking.

William et al. [24] used contextual information to identify suspicious behavior in that study. A data stream clustering algorithm, a device inference algorithm, and a context space model were the three components he used. Continuous information upgradation from incoming videos was possible using a data stream type clustering algorithm. The Inference algorithm makes a decision based on a combination of contextual information and machine awareness. The framework used two datasets: two clips from the Queensland University of Technology's Z-Block dataset and 23 clips from the CAVIAR dataset. The AUC of this method is 0.787, with 0.135 errors. Ghazal et al. [25] discovered that videos could be used to detect vandalism such as graffiti and theft. The writer used a history model and a Gaussian model that is additive in nature for segmentation. A frame difference is applied between the current frame andthe historical model. To find the area's main features as well as the color histogram, LPF with adaptive thresholding is used, as well as contour tracing and morphological edge detection. He usedthe shape and motion features to monitor objects.

Goshala et al. [26] discovered fraudulent practices in exam halls. He used the student's head role to detect fraudulent activities such as theft, transferring sheets of paper between students, and conversing with other students, among other things. He did so by combining adaptive background subtraction with sequential and periodic modeling of the background. His machine, on the other hand,couldn't manage occlusion.

Tripathi et al. [16] provided a model that detects suspicious ATM behaviors such as (forcefully taking money, and customer fights), and an alarm isactivated if the activity is detected. The videos' main features were extracted using Hu and MHI moments. The features are classified using an SVM classifier, and the dimension of the features is reduced using PCA. A window-size study based onMHI has been carried out.

B. Research in theft detection in surveillancecameras

Centered on ontology, Akdemir et al. [19] proposed the identification of human behavior in banks and other places in this paper. The authors used design consistency, ontology consistency, minimal coding bias, extensibility, and minimal ontology binding as criteria. The model was put to the test on six videos, four of which depicted robbery and two of which depicted normal behavior. Many of the videos include footage from inside the bank. The color-based motion and appearance are used to keep track of the object in motion. The presented model reliably detects robbery using a single- threaded ontology, but the model's key flaw is that it is unable to detect robberies in which more than one person is involved. The algorithm of fuzzy k-means, which was based on histogram ratio, was used by Chuanget al. [20] to recognize suspicious behavior. Using a system known as GMM, the suspicious activity was correctly identified. The entity is detected in this model using a commonly used ratio histogram. The fuzzy color histogram was used to solve the problem of color similarity. By tracking the transferring state, abnormal behaviors have been discovered.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

C. Security camera research for detectingabandoned objects

Abandoned object detection can be difficult, particularly in densely populated areas where the object may be partially or fully obscured from viewby cameras. Many researchers have focused ondetecting abandoned objects using surveillance cameras to protect people and public facilities frompossible explosives in the bag.

Sacchi and Regazzoni [27] proposed a model that uses security camera footage to detect an object left behind at a train station. If the left-behind objectis detected in the model, an alarm is activated in thenearby station, and proper authorities are notified, allowing the danger to be avoided. This model uses multiple access with direct sequence code sharing to create a noise-tolerant device and ensure a secure connection between remotes and stations. This model is designed to work with monochrome cameras. By using colored images, the model shown can be enhanced in the event of a false alarm or an object that is identified by accident. But this comes with a major disadvantage that it increases the computational time of the system and hence itcannot be used as a real-time system.

Ellingsen [28] proposed a model that uses mean pixel intensity and pixel standard deviation to detect fall artifacts. A foreground image is formed by subtracting a frame from a background image containing multiple objects. This approach is used to find objects that are moving. The features extracted to locate the object dropped by the individual are region, minor axis, major axis, the center of mass, and so on because it contain more than enough information about it. It is essential to function on a learning mechanism and an automated feature vectors classifier.

In this article we used many videos from real- world surveillance cameras, as well as some videos from the caviar dataset, to train and test our system. Human behaviors are divided into three categories: common, suspicious, and unusual. Sitting, walking, jogging, and hand waving are all popular practices. Running, boxing, war, and other suspicious activities are examples. Convolutional Neural Networks are used to accomplish this grouping. To begin, high-level features from images are extracted using a convolutional neural network. In doing so, the convolutional network classification is taken into account, the final pooling layer result is extracted, and the final prediction is made.

III. PRELIMINARIES

A. Activity Classification

We classify 4 datasets here: normal activities and suspicious activities. In context first one is Arson and second one is Burglary and the third one is Fighting where some people fight each other and the fourth dataset is the normal which captures normal things.

1) Arson
2) Burglary
3) Fighting
4) Normal

All these activities were shot fromcameras at different angles in the dataset.

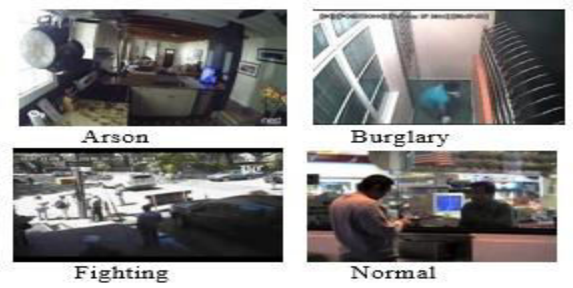

Fig. 1. Different shapes activities

Activities like boxing and fighting arecategorized as suspicious activities. Sample imagesfor various forms of activities are shown in Fig. 1.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

B. Convolutional Neural Network

Convolutional Neural Networks (CNN) are deep learning networks that are commonly used as image classifiers. This network takes an image as input and assigns importance to different objects in the image so that different groups of images can be distinguished. A sample CNN architecture is shownin the picture. 2.

In classification, Convolutional Neural Networks apply various filters to images to obtain their spatial and temporal dependencies. Since it uses fewer parameters than an ANN and allows re-usability of weights, a CNN is able to provide a better match to the image dataset. As a result, a CNN can be equipped to understand the image's complexities and sophistication far better than traditional ANNs.

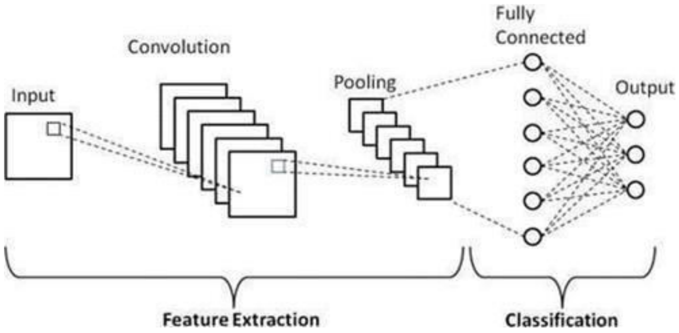

Figure 2: Convolutional Neural Network

IV. CNN MODEL FOR SUSPICIOUSACTIVITY DETECTION

Various CNN models have been proposed depending on the target outputs. LeNet is among the initial networks proposed for various image- processing applications. The architecture consists of seven layers, including two sets of convolutional layers, two sets of average pooling layers, and a flattening convolutional layer. After that, there are two thick com- completely connected layers and a SoftMax classifier. A graphical representation of the LeNet CNN architecture is presented in Fig. 3.

Layer 1: A convolutional layer with a kernel size of 3×3, a stride of 1×1, and a total of 6 kernels. As a result, a 28x28x1 input image yields a 26x26x6 image. Let's take a look at how many criteria are needed. The convolution kernel is 3x3 in size, witha total of 6×(3×3+1) =60 parameters, where +1 means that the kernel is biased.

Layer 2: A scale of 2×2 kernels, a step of 2×2 kernels, and a total of 6 kernels of pooling layers. This pooling layer works uniquely. The receptive input values are added, multiplied by the trainable parameters (1 per filter), and the result is added to the trainable biases (1 per filter). Finally, the output has undergone tanh activation. As a result, input previous sheet, which was 26x26x6, is sub-sampled to 13x13x6. (trainable parameters) + 1 (trainable biases)] * 6 = 12 total parameters in the plane.

Layer 3: This layer is a convolutional layer with the same configuration as layer 1, except that it has 16 filters instead of 6. result, previous layer's input of 13x13x6 yields an output of 11x11x16. Total layer parameters: (3x3x6x16 + 16) + 16 = 880.

Layer 4: This layer, like Layer 2, is a pooling layer, except this time it has 16 filters. The tanh activation function These facts for pass the outputs. The previous layer's input of 11x11x16 is sub-sampled to 5x5x16. (1 + 1) * 16 = 32 total parameters in layer.

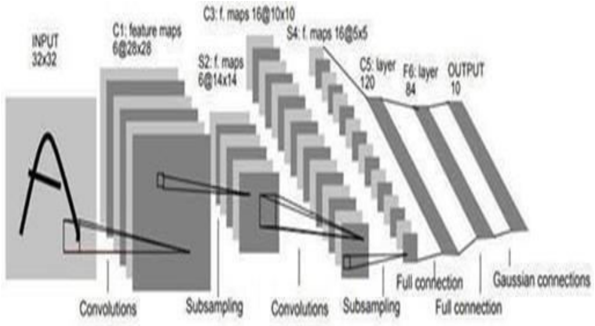

Figure 3: Seven layers of the LeNet CNN Architecture

Layer 5: After that, the data is flattened, yielding 400 neurons (5x5x16).

Layer 6: This is a 128-parameter dense sheet. Total parameters: 400 x 128 + 128 = 51328. tanh was the activation mechanism used in this case.

Layer 7: Finally, a dense layer with two units is used, which is a completely connected Softmax output layer.

V. DATASETS

We have taken several videos from real-life surveillance cameras, and some videos from the KTH dataset and Nanyang Technological University (CCTV-Fights dataset) for training and created our own dataset for testing. We have categorized people's navigation into two categories, they are normal and suspicious. The CCTV-Fights dataset includes 1,000 videos of real-world fights captured on security cameras or cell phones. Videos for the dataset were gathered from YouTube. The fights have a wide range of acts and characteristics, such as punching, kicking, pushing, grappling, fighting with two or more people, and so on. Common activities include sitting, walking, and normal. Suspicious activities include burglary, explosion, fighting, and arson. We achieve this classification using convolutional neural networks.

First, we use a convolutional neural network to extract high-level features from the image. Convolutional network classification is taken into account, the final result of the aggregate layer is extracted and the final prediction is made.

Videos were split into split images, which were then loaded in the form of a NumPy array with their labels. We know this by handling large datasets, the model must be loaded several times. So, to keep things easy, we'll create a single pickle file that contains all of the NumPy arrays and their labels, allowing us to load them quickly every time. The model architecture was done on TensorFlow and Keras environment in Python

The majority of papers using a deep learning method only detect suspicious behavior. As a result, an effective mechanism is needed to notify security in the event of any suspicious activity. When your device detects suspicious activity, it will send an SMS to the appropriate authorities. This frameworkwas built in Python on an open-source platform.

You can send SMS by creating an account with Twilio and installing the Twilio library in Python. Twilio allows you to programmatically make and receive phone calls, and send and receive text messages. Twilio offers a specific phone number aswell as an Account SID and Auth Token for sendingand receiving text messages.

TABLE I: PERFORMANCE ON TEST IMAGES

Total images	Performance parameters							
	True positive	**True negative**	**False positive**	**False negative**	**Accuracy (%)**	**ecall (%)**	**Specificity (%)**	**recision (%)**
20222	17842	2092	111	177	98.57	99.01	94.96	99.38

TABLE II: VARIOUS PERFORMANCE PARAMETERS

Total training images	Performance Parameters							
	True positive	True negative	False positive	False negative	Accuracy (%)	Recall (%)	Secificity (%)	recision(%)
80884	72343	8494	27	20	99.94	99.97	99.68	99.96

The KTH database contains six types of people's behavior: walking, jogging, running, boxing, waving, and clapping. This is an open dataset available online for the Recognition of human actions. This dataset was then split, since this dataset contains videos, we have to extract frames to get images in JPEG format.

The Nanyang Technological University (CCTV- Fights Dataset) includes 1,000 videos of real-world fights captured on security cameras or cell phones. Videos for the dataset were gathered from YouTube. The fights have a wide range of acts and characteristics, such as punching, kicking, pushing, grappling, fighting with two or more people, and so on. The dataset contains 280 CCTV videos of various forms of combat ranging in length from 5 seconds to 12 minutes with an average duration of 2 minutes. Also included are 720 videos of live action from various sources (hereafter referred to as Non CCTV).

VI. RESULTS & DISCUSSION

The proposed framework was implemented in Python 3.2 with the OpenCV library. The system hardware specifications are as follows: Intel(R)Core(TM) i5-8300H @ 2.30GHz, 8.00GB RAM,

Windows Operating System (64-bit).
The training dataset consisted of 10700 frames of non-suspicious (safe) activity and 96800 frames of suspicious activity. The output shown in Table III was calculated using a testing dataset of 20000 frames of non-suspicious (safe) and suspicious behavior. Table I represents the performance of the test set. The confusion matrix for training data and the corresponding performance are presented inTable IV and Table II respectively.

VII. CONCLUSION & FUTURE SCOPE

We provided a detection tool based on frames extracted from videos and deep learning-based algorithms in this article. To detect the operation, this novel and special method necessarily require the use of minimal computational resources. It is flexible and cell-based because of the shortage of unique hardware components. As a result, this cost-effective tool may be easily

TABLE III: CONFUSION MATRIX FOR TEST DATA

Test	Predicted	Predicted
	0	1
Actual	2092	111
0	TN	FP
Actual	177	17842
1	FN	TP

TABLE IV: CONFUSION MATRIX FOR TRANING DATA

Test	Predicted	Predicted
	0	1
Actual	8494	27

0	TN	FP
Actual	45	72343
1	FN	TP

installed in surveillance cameras and can evaluate a video in real time, alerting the police and other appropriate authorities as soon as the model detects suspicious activity. Frames were extracted from the videos for this project. All of the frames that were extracted were properly categorized. We were able to get 98.57% reliable results by using this particular method.

Additional work will include thorough testing, increased computational performance, and the incorporation the model trained separately for each field to make it more robust and accurate. It also entails the addition of a new classification parameter that is potentially suspicious behavior. If the individual is caught doing this, he will be given extra attention until he demonstrates that he is not a threat. The proposed method was developed and tested using a dataset provided by NanyangTechnological University, KTH Dataset, and othervideos captured by security cameras.

ACKNOWLEDGEMENT

The work presented in this article was supported by the SERB (Science and Engineering Research Board) sponsored research project "Develop a simple and affordable unified intelligen video surveillance system. suspicious human activity recognition in wide-area surveillance" (EEQ/2018/001103)

REFERENCES

1) P. Gajbhiye, C. Naveen, and V. R. Satpute, "Virtue: Video surveillance for rail-road traffic safety at unmanned level crossings;(incorporating indian scenario)," in 2017 IEEE Region 10 Symposium (TENSYMP), pp. 1–4, IEEE, 2017.
2) V. Kamble and K. Bhurchandi, "Noise estimation and quality assessment of Gaussian noise corrupted images," in IOP Conference Series:Materials Science and Engineering, vol. 331, p. 012019, IOP Publishing, 2018.
3) A. A. Bhadke, S. Kannaiyan, and V. Kamble, "Symmetric chaos-based image encryption technique on image bit-planes using sha-256," in Twenty Fourth National Conference on Communications (NCC), pp. 1–6, IEEE, 2018.
4) A. L. Alappat and V. Kamble, "Image quality assessment using selective contourlet coefficients," in 11th International Conference on Compute- ing, Communication and Networking Technologies (ICCCNT), pp. 1–7, IEEE, 2020.
5) C. Sacchi and C. S. Regazzoni, "A distributed surveillance system for detection of abandoned objects in unmanned railway environments," IEEE Transactions on Vehicular Technology, vol. 49, no. 5, pp. 2013– 2026, 2000.
6) K. Ellingsen, "Salient event-detection in video surveillance scenarios," in Proceedings of the 1st ACM workshop on Analysis and retrieval of events/actions and workflows in video streams, pp.57–64, 2008.
7) M. Ghazal, C. Va´zquez, and A. Amer, "Real- time automatic detection of vandalism behavior in video sequences," in 2007 IEEE International Conference on Systems, Man and Cybernetics, pp.1056–1060, IEEE, 2007.
8) D. Gowsikhaa, S. Abirami, et al., "Suspicious human activity detection from surveillance videos.," International Journal on Internet & Dis- tributed Computing Systems, vol. 2, no. 2, 2012.
9) C. Sacchi and C. S. Regazzoni, "A distributed surveillance system for detection of abandoned objects in unmanned railway environments," IEEE Transactions on Vehicular Technology, vol. 49, no. 5, pp. 2013– 2026, 2000.
10) K. Ellingsen, "Salient event-detection in video surveillance scenarios," in Records of the First ACM workshop on Analysis and retrieval of events/actions and workflows in video streams, pp.57–64, 2008.

INNO SPACE
SJIF Scientific Journal Impact Factor

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com