



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 10, October 2021

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 7.569

9940 572 462

6381 907 438

[ijrset@gmail.com](mailto:ijrset@gmail.com)

[www.ijrset.com](http://www.ijrset.com)



# Real Time Anomaly Detection using Drone Surveillance

Raj Shah<sup>1</sup>, Aparna Burhade<sup>2</sup>, Parth Kulkarni<sup>3</sup>

U.G. Student, Department of Computer Engineering, Vishwakarma Institute of Information Technology, Pune, Maharashtra, India<sup>1</sup>

U.G. Student, Department of Electronics & Telecommunication Engineering, Modern Education Society's College of Engineering, Pune, Maharashtra, India<sup>2</sup>

U.G. Student, Department of Computer Engineering, P.E.S. Modern College of Engineering, Pune, Maharashtra, India<sup>3</sup>

**ABSTRACT:** Deep learning has shown significant performance in many domains including natural language processing, recommendation systems, and self-driving cars in current years. From all the available applications detecting anomalies is a key problem that has been studied within research domains. The purpose is to assist with recognizing individual actions and detecting whether it is an anomaly or normal activity. To address this challenge of a detection algorithm for action recognition the author has presented a 3-dimensional convolutional neural network model to detect real-time anomalies using drone surveillance. The initial stage of the model extracts features from each person in the video and represents the data. Analysis of each extracted sequence to detect the associated actions is also proposed. Access dataset and UCF- Rooftop dataset was used for training and testing purposes. The results of this work revealed that the proposed 3-dimensional convolutional neural network method provides more accurate anomaly detection achieving around 90%.

**KEYWORDS:** Anomaly Detection, 3-Dimensional Convolutional Network, Action Recognition, Suspicious activity, Video recognition.

## I. INTRODUCTION

In recent years, there is a reduction in the number of unusual activities happening around which has resulted in fewer security problems in public life. But when some unusual activity happens the authorities cannot reach the location at the time of the incident. So, developing a system that could potentially identify these incidents so that law authorities can reach the location rapidly is the hour of need. However, the problems we faced in the past were lack of coverage, low computational power, and models based on video and action recognition were scarce and did not achieve very good performance and results.

CCTV Surveillance is being used in public places to detect anomalous activity in the surroundings and for public safety. A crucial piece of work in video-based drone surveillance is the detection of anomalous activity such as robbery, accidents, explosions, or misleading activities. Generally, anomalous activity frequency is less than the usual activity. Hence, to ease the cost of time and labour, evolving intelligent deep learning and computer vision algorithms for real-time anomaly detection using drone surveillance is the current requirement. The purpose of real-time anomaly detection is to indicate action as either normal activity or anomaly activity. In short, activity recognition is required to result in whether it is anomalous or normal patterns.

The issue is addressed by developing an anomaly detection algorithm to identify particular anomalous incidents. Anomaly detection in the real world is quite interesting, varied, and multiplexed. Real-time anomaly detection using drone surveillance uses action recognition. Research done in Video Recognition models, especially those related to 3D convolutional networks, we see better accuracy and high-performance models being constructed for this task. Since we are working with videos to check whether it's normal or an anomaly the basic step is feature selection. The number of trainable and non-trainable parameters is high. The optimization techniques for hyper-parameters are recommended. Previous other approaches like 2D Convnets + LSTM model are being also developed and have achieved modest results. The environment in which videos are captured by surveillance camera changes drastically like for example the drone movement, human position, activity point of view, angle of drone, etc.



In this paper, the author tried to implement real-time anomaly detection using drone surveillance techniques on 3DCNN Model Architecture by using action recognition.

Action Recognition is a technique that does features extraction from each video from the collected dataset. Analysis of each extracted sequence to detect the associated actions is also proposed by training the 3DCNN model. Training is done with the training data followed by testing it on random video. The model is repeatedly performed until we get a satiate outcome. The developed dataset includes video clips that cover the area from multiple heights, angles, and situations. Our desired output for the model is to identify the scene that checks humans performing normal action or anomalous activity.

The main contributions of this paper are: -

- We introduce a large-scale video anomaly detection dataset consisting of 9500 real-world surveillance videos of 23 different anomalous events and normal events captured by surveillance cameras.
- We have merged the UCF Rooftop dataset and DCSASS dataset to get this huge dataset.
- Results concluded that the model achieves better performance as compared to state-of-the-art anomaly detection approaches.
- An attempt at creating a practical model that can identify suspicious activities and distinguish them from normal events.

## II. RELATED WORK

Anomaly is all about some abnormal activities that happen in day-to-day life. This is a reasonably commonly used type of machine learning. Anomaly is also referred to as outliers in the statistics and data mining literature. By anomaly detection definition, anomalies happen in rare events. Therefore, a natural idea to estimate the data distribution and detect outliers. The common approach presented in the literature is to learn the representation basis of normal patterns and then to try to reconstruct an input frame based on this representation. In this case, video is providing as input and this video convert into a-frames and these frames providing a large reconstruction error are likely to be abnormal. To improve the action recognition performance recent works employed various deep learning models. Since human action is extracted from multiple moments of the human body or some parts of it is necessary that the recognition process should involve video extracting over time to learn the patterns of visual appearance changes. To achieve these existing deep learning models based on the 2D the convolutional network can be extended into the 3D domain so that the 3D domain captures the temporal information. For example, use the motion in consecutive frames to extract information for action recognition using a two ConNet method that incorporates both spatial and temporal. 3DCNN based method using a stack frame cubes of actions to capture 3D spatial-temporal action signals. In order to recognize human action, many methods use several different input videos. Use a combination of 2DCNN and LSTM to recognize human action. Use 2DCNN-LSTM a model with two streams as input the first represents the human body skeleton and the second represent the video segment data.

### **Suspicious human activity recognition: a review-By Rajesh Kumar, TripathiAnand Singh Jalal, Subhash Chand Agrawal**

The research team working on this paper produces a very thorough and in-depth analysis of many of the major works done in this field. They also highlight and specify the various techniques used by the major research papers written over the years to provide a guiding light for new research to begin and be successful. The team compiles the pros as well as the cons of many of the models prepared by various research institutions and researchers to form a very detailed and structured review for the field of surveillance and intelligent video monitoring systems.

### **Real-world Anomaly Detection in Surveillance Videos-Waqas Sultani, Chen Chen, Mubarak Shah**

Department of Computer Science, Information Technology University, Centre for Research in Computer Vision (CRCV), University of Central Florida (UCF)

The research team proposes to learn anomalies by exploiting both normal and anomalous videos. To avoid annotating the anomalous segments or clips in training videos, which is very time-consuming, the team proposes to learn anomaly through the deep multiple instances ranking framework by leveraging weakly labelled training videos, i.e., the training labels (anomalous or normal) are at video level instead of clip-level. In their approach, they consider normal and anomalous videos as bags and video segments as instances in multiple instances learning (MIL), and automatically learn a deep anomaly ranking model that predicts high anomaly scores for anomalous video segments. Furthermore,

they introduce sparsity and temporal smoothness constraints in the ranking loss function to better localize anomalies during training.

Our approach consists of creating a separate Class based on the same videos of the given dataset for which the model will distinguish between Anomalous activities and Normal activities and extract features from the videos for better detection.

### III. METHODOLOGY

#### A. Dataset Selection

The "DCSASS" dataset [2] and the "UCF Rooftop" dataset [3] were merged to obtain larger amounts of data and train the models with that dataset. Surveillance videos cover over 10 to 13 real-world anomalies for example Burglary, Assault, etc. These unusual activities are taken into consideration because they have an impact on public safety. The dataset that we have created contains videos that are recorded at a height of about 10-15 meters with actions such as 'Running,' 'Walking,' 'Fighting,' 'Stealing', etc. These videos have been recorded at different times of the day and from different angles to train the model properly. Lastly, the distribution of this dataset is as follows: There is a total of 11758 videos, where 6150 videos are labelled as Normal and 5608 as abnormal. In that dataset, every class consists of several videos ranging from 50-150 to make the data balanced. Using python script 4 we have separated the Normally labelled video clips from the other classes and combined them all into two separate classes of training data and testing data. This would result in the model extracting features from videos.

#### B. Pre-processing

Due to a large number of video datasets, it is not feasible and efficient to load the entire dataset into RAM during training. So first we extracted the frames from each video separately and split them all into separate two folder training data and the testing data first we train the model with the training data and then test the model with the testing data. Extraction was done using OpenCV. This enabled us to resume model training and testing without having to extract frames every time. We also made use of Keras's util. The sequence class allowed us to only load the data which we required for a particular mini-batch. During training, we only load videos. Due to the uneven duration of the clips, we only select each frame for each video. We then resize the images to 32\*32\*32. The shape of an input video is (12,32,32, 3). and run the model with small changes in the parameters until we get a satisfying outcome.

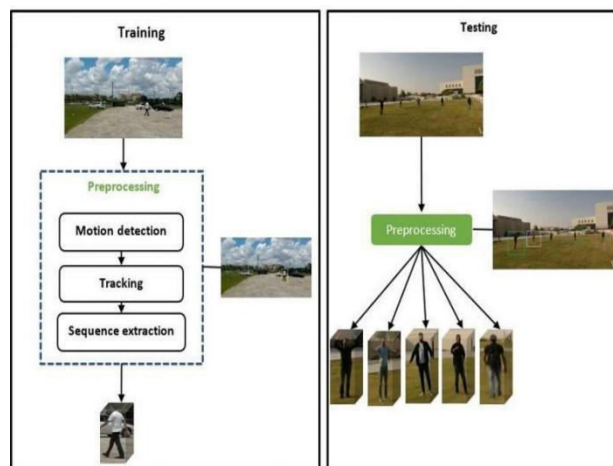


Fig 1: Flow extraction process

#### C. Architecture

3DCNN networks can be described as a single stream architecture that operates at two different frame rates of training and testing. We know that the 1D-CNN model is usually used for statistic knowledge and therefore the 2D-CNN model has the best accuracy in image classification and once it involves the 3D-CNN the dimension is for the statistic since the videos area unit temporal knowledge. The output for the 3D-CNN is four-dimensional. Achievement of 3D convolution is done by convolving a 3D kernel to cube-shaped. This process is done by stacking multiple contiguous frames along. This results in the feature map within the convolution layer are connected to multiple contiguous frames. We extract each frame from each video clip5 and resize the frames to 32\*32 with depth 12 before computing features and so we tend to pass the information to Conv3d layer, Maxpooling3d, Flatten, Dense, Batch- Normalization, dropout



with activation functions, and extracted the utmost options from the input file and born-again it into the dimensions of labels or categories. The dataset is taken consists of twenty-three categories.

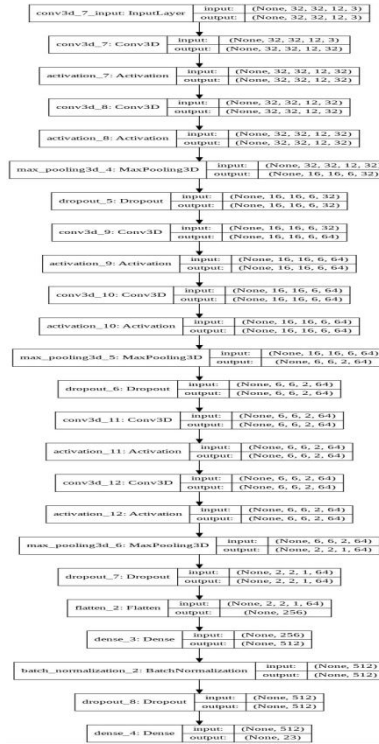


Fig 2: Layered Configuration of proposed 3DCNN architecture.

#### IV. EXPERIMENTAL RESULTS

The following section presents details of the experiment's implementation and its result. The experiment was set up and performed on the DCSASS ad UCF-Rooftop dataset as a whole that had around 9500 video clips. The frequency of videos in each class ranged from 50-100. For activity recognition, divide The dataset into an 80/20 ratio for training and testing. The summary of the experiments for real-time anomaly detection is as follows.

##### A. Training Result

3DCNN model is trained and results obtained on training and the testing dataset are shown in Figures.

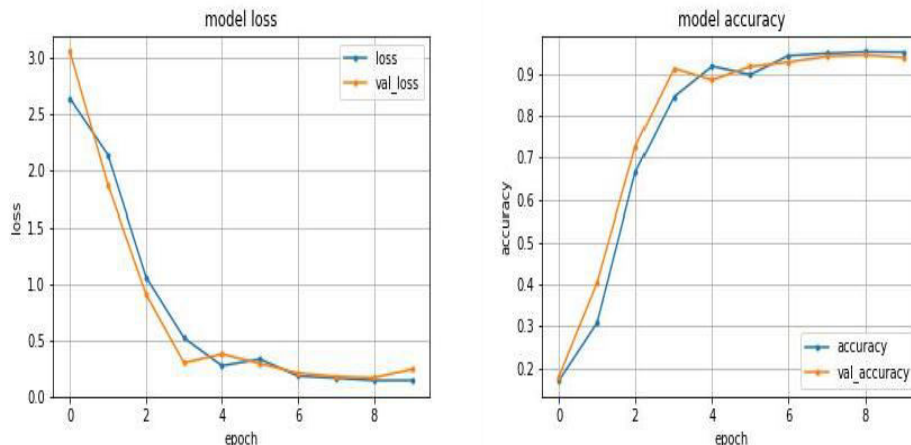


Fig 3 & 4. Accuracy and Loss comparison of training and testing(validation) of the 3DCNN model with a depth of 12 frames, Batch size 32, and Number of Epochs as 10.



The model was trained for 10 epochs with a mini-batch size of 32. Fig 3 and 4 give the logs for the training and testing(validation) accuracy and losses during the entire training process.




The final results are:

Training loss - 0.1401, Training accuracy - 0.9541

Testing loss - 0.2406, Test accuracy - 0.9392

**B. Testing**

For testing, we collected a new set of data which was collected from local and YouTube videos which are typically followed by pre-processing step, feature extraction of each frame in sequence from a video. Thus, the model would then really try to predict every second as the video frames.

Test Video	Predicted	Actual	Comments
	Arson	Arson	Bright flames is learned by the model to associate with arson.
	Normal	Road Accident	Model is not able to distinguish due to harsh street lights
	Road Accident	Road Accident	Model can easily predict car crashes on empty road

From the sample results in Table 1, we conclude that there are some predictions that the 3DCNN model fails to detect.

**V. CONCLUSION**

We have attempted an approach to detecting anomalous activities based on drone surveillance video clips. We have successfully designed the 3dcnn model which was able to predict and differentiate between the normal and anomalous activities successfully. We achieved 93% test accuracy which is greater than most past models and we have also used pre-processing techniques to increase the efficiency and performance of the model. The proposed anomaly detection algorithm can be used to detect anomalies by taking into account more datasets. More data related to more normal scenarios would help the model classify normal activities better.

**A. Learning and Reflection**

Through this project, we have learned a lot about how the process of designing, planning, and implementing a real-scale deep learning project works. We now have hands-on experience in designing models, obtaining data, and the importance of pre-processing. Working with video data also proved to be another challenge and it helped us understand just how new this field in deep learning is and the many difficulties that come with video data and we learned to overcome them.

**B. Limitations**

1. More data related to more generalized scenarios would help the model classify normal activities better.
2. Hardware limitations prevent us from attempting to extract more frames and large resolution videos for training
3. Restrictions in Google Colab prevent us from training the data continuously to maintain consistency.



## REFERENCES

- [1]. Waqas Sultani, Chen Chen, Mubarak Shah, 'Real-world Anomaly Detection in Surveillance Videos' arXiv(5), 1801.04264 (2018).
- [2]. DCSASS Dataset, Kaggle, <https://www.kaggle.com/mateohervas/dcsass-dataset>
- [3]. UCF Rooftop Dataset, <https://www.crcv.ucf.edu/data/UCF-ARG.php>
- [4]. Patel, Z., Senjaliya, N., & Tejani, A. (2019). AI-enhanced optimization of heat pump sizing and design for specific applications. *International Journal of Mechanical Engineering and Technology (IJMET)*, 10(11), 447-460.
- [5]. Raghavendra Chalapathy, Sanjay Chawla 'Deep Learning for Anomaly Detection: A Survey' <https://arxiv.org/abs/1901.03407>
- [6]. J.Sasidevi, R.Sugumar, Fuzzy C-means Clustering and Elliptic Curve Cryptography using privacy preserving in cloud, *International Journal of Business Intelligence and Data Mining*, Volume 15, Issue 3, pp.273-287, July 2019.
- [7]. Balmukund Mishra, Deepak Garg, Pratik Narang, Vipul Mishra 'A hybrid approach for search and rescue using 3DCNN and PSO', [https://doi.org/10.1007/s00521-020-050017\(0123456789\).vol V](https://doi.org/10.1007/s00521-020-050017(0123456789).vol V)
- [8]. Video Analysis to Detect Suspicious Activity Based on Deep Learning, <https://medium.com/@everisUS/video-analysis-to-detect-suspicious-activity-based-on-deep-learning-fee2032ea14a>
- [9]. G Jaikrishna, Sugumar Rajendran, Cost-effective privacy preserving of intermediate data using group search optimisation algorithm, *International Journal of Business Information Systems*, Volume 35, Issue 2, September 2020, pp.132-151.
- [10]. Noor Almaadeeda, Omar Elharroussa, Somaya Al-Maadeeda, Ahmed Bouridaneb, Azeddine Beghdadi 'A Novel Approach for Robust Multi Human Action Detection and Recognition based on 3-Dimensional Convolutional Neural Networks' journal homepage: [www.elsevier.com](http://www.elsevier.com)
- [11]. Joao Carreira, Andrew Zisserman, 'Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset' arXiv(5), 1705.07750(2017).
- [12]. Senjaliya, N., & Tejani, A. (2020). Artificial intelligence-powered autonomous energy management system for hybrid heat pump and solar thermal integration in residential buildings. *International Journal of Advanced Research in Engineering and Technology (IJARET)*, 11(7), 1025-1037.
- [13]. Tripathi, R.K., Jalal, A.S. & Agrawal, S.C. Suspicious human activity recognition: a review. *Artif Intell Rev* 50, 283-339



**Impact Factor:  
7.569**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details