International Engineering Journal For Research & Development

Vol.4
Issue 5

E-ISSN NO:-2349-0721

Impact factor : 6.03

A SYSTEM TO DETECT ANOMALY IN LIVE FEED OF AUTONOMOUS DRONE SURVEILLANCE USING COMPUTER VISION APPROACH

Sachin S. Deshmukh
Dept. of CSE,
PRMCEAM, Badnera
Amravati, Maharashtra
sachindeshmukh24@gmail.com

Dr. M. S. Ali
Dept. of CSE,
PRMCEAM, Badnera
Amravati, Maharashtra
softalis@gmail.com

Dr. D. G. Harkut
Dept. of CSE,
PRMCEAM, Badnera
Amravati, Maharashtra
Dg.harkut@gmail.com

Vikram M. Kakade
Dept. of EXTC,
PRMCEAM, Badnera
Amravati, Maharashtra
vikram.kakade160@gmail.com

Abstract-

Now a days, every city in the world has installed CCTV cameras for the monitoring the crowd and for public safety. But as these cameras are increasing, it has become difficult to manually monitor the feed which is received from these cameras. Also CCTV cameras are stationary, which is making them capturing the footage from particular angle while monitoring a place. These problems can be overcome by using the autonomous drone technology. By virtue of recent industrial developments, drones are increasingly employed in various domains. They can also be employed for surveillance as moving cameras. But again the problem of autonomous drone surveillance is that, it can only fly around without human interaction and transmit the live feed. It cannot interpret what is in that live feed. The solution to this problem is Computer Vision application through Deep Learning approach. Deep learning is a trend field of Computer Science which is helping to solve the problem like to autonomous car, detection of cancer cell and many other applications where it can possibly replace human eyes to see and interpret the data. In this paper, we are proposing a system design which will make use of the deep learning to detect the anomaly in the live feed received from an autonomous drone. The suggested anomaly detector system makes use of a deep neural network composed of a convolution neural network and a recurrent neural network, trained using supervised learning. Although we have not implemented a system but our main aim is to promote research efforts to resolve the impenetrability of anomaly detection in live videos.

Keywords— Anomaly detection, Deep Learning, Drone, Surveillance, Computer Vision

INTRODUCTION

The general meaning of anomaly is something that deviates from what is standard, normal, or expected. This makes anomaly detection a special area of interest for researchers around the world. For example, anomaly can be intrusion in the network, abnormal behavior of node in the network, in the surveillance case an abnormal event might be an intrusion or a person running on a pedestrian path.

In medical area, it might be the presence of a cell or tumor. Ignored anomalies can present costly ramifications. Therefore, their detection is an important task that arises in various domains. However, by definition, an anomaly is a rare event. For this reason, their detection is tedious. For example, in the case of surveillance, human operators must watch continuous video streams carefully, possibly from several sources, for hours. For this reason, an automatic system that performs this task is desired.

Current surveillance systems commonly employ fixed position cameras that present some difficulty in anomaly tracking. By virtue of recent industrial developments, drones are increasingly used in various applications related to agriculture, traffic monitoring, construction industry, and disaster area investigations. In the perspective of surveillance, drones are useful as mobile cameras. They overcome weaknesses of fixed position cameras. It is particularly interesting that no recent reports describe studies addressing the difficulty of finding anomalies in videos with a dynamic background such as those recorded by drones. Recent contributions in anomaly detection for the surveillance case rely upon the assumption of a static background using data representations such as optical flows [1]. However, such methods are difficult to apply for videos with dynamic backgrounds.

For many years, information extraction from videos was done by using the features which are considered and feel important to the researchers. However, this limited feature problem can be resolved by using the deep learning approach. Also we have huge amount of data available publicly which can be used to train the deep learning model which can learn to extract useful features automatically. Consequently, tools such as convolution neural networks (CNN) and recurrent neural networks (RNN) are promising for extraction of spatiotemporal information from videos. As described herein, we propose with a supervised learning approach, a model that combines both a CNN and an RNN for the task of anomaly detection in live feed of a autonomous drone.

CURRENT STATUS AND CHALLENGES

Numerous publicly available datasets have been designed for anomaly detection in the context of surveillance. All of these dataset are created from the stationary CCTV cameras; very few of those include videos recorded by drones. As described in this paper, we specifically consider the drone video for anomaly detection. During our literature survey we found only two dataset useful for our proposed system. The mini-drone video (MDV) dataset [2] can be used to train the proposed system's deep learning model. Another dataset is from University of Minnesota (UMN) [3] on Detection of Unusual Crowd Activity.

Our survey shows that difference between the results obtained on the MDV dataset and those obtained on the UMN dataset indicates the former as more complex, both in varieties of scenes and conditions under which videos were recorded. Furthermore, the model struggles to detect anomalies that are similar to normal patterns.

By definition, anomalies are rare events. Therefore, a natural idea is to estimate the data distribution and to detect outliers. We found many papers on the theme of anomaly detection using deep learning. But very few are focused on drone video dataset. The detail of finding in the literature survey is mentioned below. Sun et al. [4] proposed an online growing neural gas that adapts itself to changing environments to estimate the data distribution topology. During tests, when the neuron representing a given input video frame is far away the densest parts of the gas, the corresponding frame is regarded as abnormal. Another approach was suggested by Xu et al. [1], who proposed the use of three auto-encoders to extract appearance, motion and aggregation of both features from video frames. Then, three one-class Support Vector Machines (SVM) are trained for anomaly detection with features extracted using auto-encoders. During tests, the scores of the three SVMs are aggregated to compute the global abnormal score of a given video frame.

Another 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, a frame providing a large reconstruction error is likely to be abnormal. Ren et al.[5] proposed to learn dictionaries representing normal patterns. Chong et al. [6] reported the use of a spatiotemporal auto-encoder with an encoder consisting of convolutional layers and a decoder of deconvolutional layers. To extract temporal features, the auto-encoder's bottleneck is made with convolutional long short-term memory (conv-LSTM) units. Hasan et al. [7] proposed a similar approach, except that the model includes no conv-LSTM units. Rather, to extract temporal information, the method inputs a stack of several consecutive frames to the auto-encoder so that the input volume has a temporal dimension. The idea that large reconstruction errors are useful for anomaly detection has also been investigated in a medical context by Schlegl et al. [8] for retinal fluid or hyper-reflective foci detection. To do so, the authors suggest the use of a Generative Adversarial Network (GAN). Because GAN only learns to map latent representations to images, the authors also suggest learning reverse mapping, so that a given image can be reconstructed and evaluated during testing.

Recently, Munawar et al. [9] defined anomalies as ob- served events that differ from expectations. Consequently, they suggested comparison of the prediction of a model and the actual observation. A model is trained to predict a representation of the frame at time-step t + 1 given the frame at time-step t.

In this paper, we are specifically considering the mini-drone video dataset for which the training set includes both normal and abnormal patterns. Therefore, we decided to work first on training a model under a supervised learning approach to propose baseline challenges and current status. All methods cited above tackle the anomaly detection problem with semi-supervised or unsupervised learning approaches. They use only normal patterns during training and try to detect outliers during tests. Manuwar et al. [10] investigated supervised learning for anomaly detection based on reconstruction error. They suggested training a model by alternating positive and negative learning phases. A positive learning phase searches for the parameters that minimizes the reconstruction error of normal frames, while a negative phase maximizes the reconstruction error of abnormal frames.

Mini-drone video dataset's anomalies are essentially determined by actors' actions. Thus, one can consider the problem as action detection. Hong et al. suggested methods to extract the skeleton from the silhouette of an actor by using sparse coding in [11], and a multimodal deep auto- encoder in [12]. However, the mini-drone video dataset includes a few examples per action. Furthermore, some actions are present in the test set but not in the train set. Thus, action detection is hardly conceivable. Consequently, this work does not attempt to recognize each action, but to classify them as normal or not.

Although the research on anomaly detection is beneficiary and received a lot of attention from research community but it also faces some challenges are listed below:

1. Due to different motion patterns of different subjects at different time, the accuracy of activity recognition decreases.
2. It is difficult for any classification algorithm to recognize the motion during the transition period between two activities.
4. Smaller datasets
5. Smart human computer interaction
6. Detection of human activities in lesser resolution pictures

PROBLEM FORMULATION

Convolutional Neural Network

The first component of our model is designed to extract visual (spatial) information contained in video frames. To do so, we use VGG-16, a famous CNN architecture proposed by Simonyan and Zisserman [13]. This architecture includes thirteen convolutional layers, five max pooling operations and three fully connected layers as depicted in Figure 1.

Training a deep CNN such as VGG-16 requires a large dataset and requires much time. Consequently, the proposed method uses transfer learning to ease the training process by initializing the weights of VGG-16 to those resulting from training on the ImageNet [14] dataset. This dataset includes more than fourteen million images divided up into one thousand classes, making it a challenging natural image classification dataset.

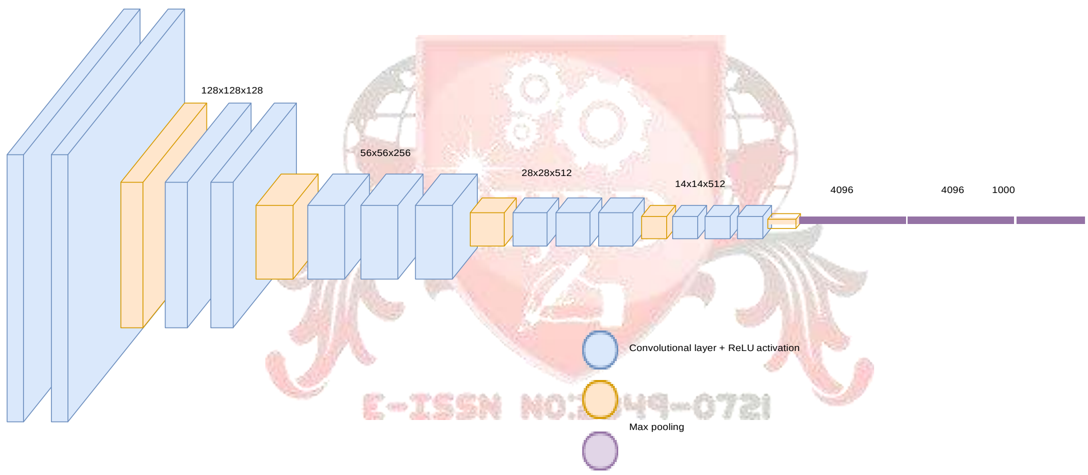


Figure 1 VGG16 architecture

Recurrent Neural Network

The second component of our system is aimed at extract- ing temporal information contained between video frames. Standard neural networks and CNN do not extract time dependencies between the current and previous input features. A common approach for the extraction of temporal information is to employ an RNN. This class of neural network extracts information from the current input while considering features from earlier inputs.

Specifically, the proposed method employs a Long Short-Term Memory (LSTM) [15]. This model is commonly used to cope with the vanishing/exploding gradient problem commonly encountered with standard RNN. LSTM tackles this problem, by using gates that maintain the recurrent state c by removing or adding information according to the current input vt and the previous output

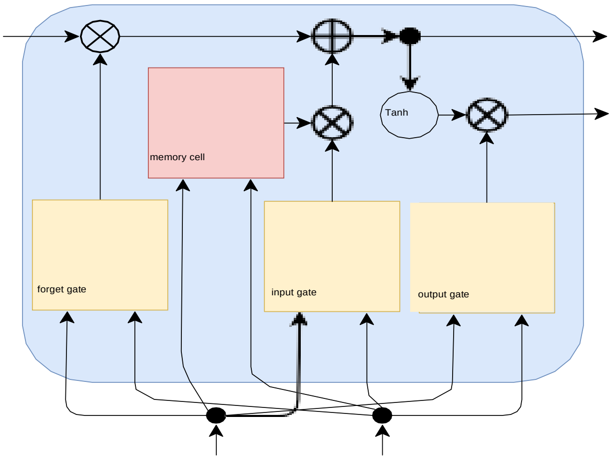


Figure 2 LSTM diagram

In this paper, we specifically examine the difficulty of anomaly detection as a binary classification task. The actual system proposed in this paper is the combination of the two components presented in earlier sections. Additionally, it ends up with a final fully connected layer that computes evidence for the classes (normal = 0 or abnormal = 1). Here, it consists of a single neuron with a sigmoid activation. Figure 3 depicts an abstract view of the proposed model.

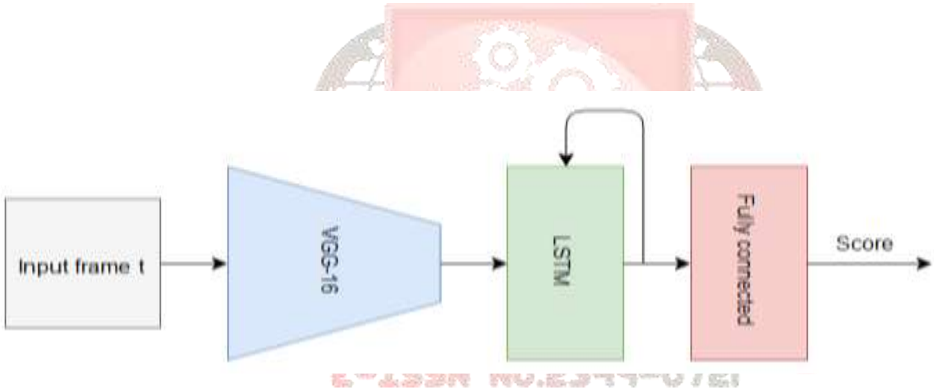
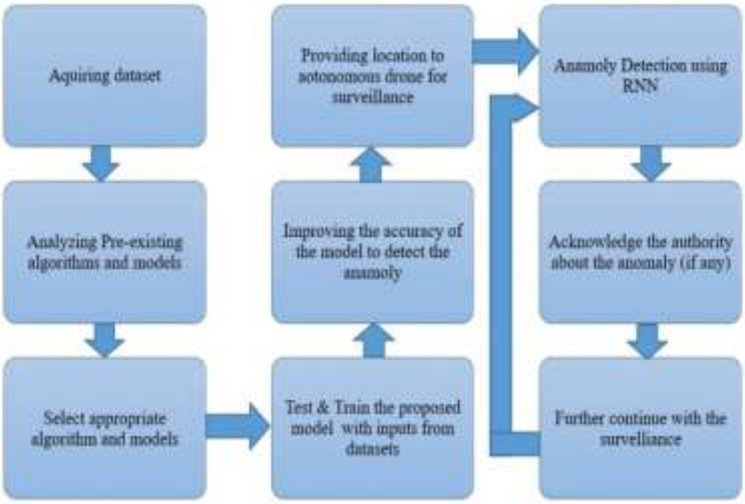


Figure 3 Diagram representing the suggested system.

IMPACT ON SOCIETY

Impact on society: due to our video surveillance systems it is very easy to detect suspicious behavior in our society. This system is very useful for detective agencies. Generally we can see that in our Indian society there are various illegal works openly done by terrorism group which can produce violence in our society but by using this human anomaly action detection surveillance system it is easy to track these activities and take action over it. By introducing this technology in our society the security level of our society will increase and be of a great help. Due to the self analyzing feature of our system, it will reduce the human error so as a result the trust on our security system of our society will increase. Currently our country have a much more population due to this our local police capability of analyzing anomaly actions is decreased but our system uses a neural network which can have a much better capability than our human brain, an eye to analyze and detect anomaly actions that's why due to introduction of our system the capability of our local police to analyzing anomaly actions will increase. This increasing capability of our police of analyzing anomaly action provides various benefits to our society. In recently we can see that terror group make trap for our soldiers and destroy their vehicle but if we use our drone in that area then we can easily catch there anomaly action and save our soldiers life .the use of our system in border site local areas will make a big impact on anomaly action of terrorists in border local areas. Our country have various of religious program in each year where the big crowed is gathered and terrorist take a advantage of this and make a illegal activities in crowed due to lack of crowed they can't be catch easily

DESIGN APPROACH AND NOVELTY



The proposed system will have autonomous drones spread across the city or the area. These drones will recording the area through their camera and will be sending it to the base station. The base station will no longer need to appoint a human to monitor the live fed which is being transmitted by the drones. The live feed which is being received will be given as input to our deep learning model, which will do the video analysis in real time.

The model will detect any anomaly in the live feed, if any. If the anomaly is detected then it will be flagged and the concerned policing authority in that area will be notified.

Novelty of our approach is that we are first to propose the use of drone surveillance in context of detecting anomaly using computer vision problem. This system will free up the human effort which is required to monitor or may not be able to detect anomaly in the video surveillance.

CONCLUSION

As described herein, we specifically studied anomaly detection in a surveillance context, especially for the mini- drone video dataset that consists of surveillance videos recorded by a drone.

The system proposed in this paper was not evaluated. This report is the first of an attempt to propose anomaly detection in live feed of the drone. We observed the system which might get to be sensitive to the dynamic background of the MDV dataset's videos. Consequently, our future work will specifically focus on the development of actual system to implement the proposed design.

REFERENCES

[1] Dan Xu, Yan Yan, Elisa Ricci, and Nicu Sebe. "Computer Vision and Image Understanding", pp 117–127, 2017.
[2] Margherita Bonetto, Pavel Korshunov, Giovanni Ramponi, and Touradj Ebrahimi. "Privacy in mini-drone based video surveillance", *IEEE International Conference on Image Processing (ICIP)*, pages 2464–2469, 2015.
[3] Department of Computer Science University of Minnesota and En- gineering. Detection of unusual crowd activity. Video dataset. http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi.
[4] Qianru Sun, Hong Liu, and Tatsuya Harada. "Online growing neural gas for anomaly detection in changing surveillance scenes. Pattern Recognition", pp 187–201, 2017.
[5] Yong Shean Chong and Yong Haur Tay. "*Abnormal Event Detection in Videos Using Spatiotemporal Autoencoder"*, pages 189–196. *Springer International Publishing*, Cham, 2017.
[6] Mahmudul Hasan, Jonghyun Choi, Jan Neumann, Amit K Roy- Chowdhury, and Larry S Davis "Learning temporal regularity invideo sequences" *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 733–742, 2016.
[7] Thomas Schlegl, Philipp Seeböck, Sebastian M. Waldstein, Ursula Schmidt-Erfurth, and Georg Langs "*Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery"*, pages 146–157. *Springer International Publishing*, Cham, 2017.
[8] Asim Munawar, Phongtharin Vinayavekhin, and Giovanni De Magistris "Spatiotemporal anomaly detection for industrial robots through prediction in unsupervised feature space" *In Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on*, pages 1017–1025, 2017.
[9] K. Simonyan and A. Zisserman "Very deep convolutional networks for large-scale image recognition *CoRR"*, abs/1409.1556, 2014.
[10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. "ImageNet: A Large-Scale Hierarchical Image Database" *In CVPR09*, 2009.
[11] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory, *Neural Computation*, 9(8):1735–1780, November 1997.
[12] Diederik P. Kingma and Jimmy Lei Ba. Adam "A method for stochastic optimization", *The Third International Conference for Learning Representations*, 2015.

[13] Ian Goodfellow, Yoshua Bengio, and Aaron Courville "*Deep Learning*", chapter 7, pages 241–249. MIT Press, 2016.
[14] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov "Dropout: A simple way to prevent neural networks from overfitting" *Journal of Machine Learning Research*, 15:1929–1958, 2014.
[15] D. Doermann and D. Mihalcik. "Tools and techniques for video performance evaluation", *15th International Conference on Pattern Recognition. ICPR-2000*, pages 167–170, 2000.
[16] Statistical Visual Computing Laboratory University of California San Diego. Ucsd anomaly detection dataset. Video dataset. http:
//www.svcl.ucsd.edu/projects/anomaly/dataset.html.
[17] Cewu Lu, Jianping Shi, and Jiaya Jia. "Abnormal event detection at 150 fps in matlab", *Proceedings of the IEEE International Conference on Computer Vision*, pages 2720–2727, 2013.
[18] Radu Tudor Ionescu, Sorina Smeureanu, Bogdan Alexe, and Marius Popescu "Unmasking the abnormal events in video" *IEEE International Conference on Computer Vision (ICCV)*, pages 2914– 2922, 2017.
[19] Allison Del Giorno, J Andrew Bagnell, and Martial Hebert "A discriminative framework for anomaly detection in large videos", *European Conference on Computer Vision*, pages 334–349. Springer, 2016.