Accident Detection System using Surveillance Camera's

Rohit Hooda[1], Rishabh Kawediya[2], Krish Sonigara[3], Snehal Andhare[4]

[1,2,3]Student, Dept. of Computer Engineering, Vidyalankar Institute of Technology, Mumbai, India
[4]Professor, Dept. of Computer Engineering, Vidyalankar Institute of Technology, Mumbai, India

---***---

Abstract- Accidents have been main reason for the road casualties from a long period of time. From past decade surveillance cameras have been installed for security purposes on several roads and are still being installed, but those surveillance cameras aren't being used to their fullest. Also, there aren't enough of man power to survey each and every road, each and every surveillance video. Also more of manual surveillance leads to more of errors, delays, miscommunications and ultimately can be prove to be fatal in serious scenarios. Hence, automation in this sector is very much important. By using Artificial Intelligence all the surveillance videos that are being recorded and can revert back directly to emergency services without any intervention. This could save the time wasted by the manual communication as every second is important so save lives during an accident. The paper proposes how human life on road can be simplified by using extraction techniques on the surveillance video. This study proposes ways and steps to achieve the objective by using different algorithms.

Keywords— *Accident detection, C3D feature extraction, multiple instance learning, surveillance videos.*

I. INTRODUCTION

Accident Detection has been one of the basic necessities for modern day safety as new and new cars are being added on the road every day. With increasing numbers of travelers, the number of accidents has gone up considerably. Hence, there is a need of modern solution to this modern-day problem. Adding the accident detection algorithm to the surveillance cameras would provide a great safety feature for the detection and to alert the emergency services without any waste of time made due to manual work.

As surveillance cameras are being used more and more on the roads, for safety purpose, anomaly detection is also a huge and important part of road safety as the road accidents can be said as byproduct of road usage. Although road accidents are rare events and might not be a everyday activity, but anytime one takes place, it could be a life and death situation. Thus video rendering and filtering and detecting anomalies from that video plays an important part in road safety.

Detecting accidents in real world is quite difficult task to perform, as there are many different factors involved such as human interactions, weather, camera quality, etc. These factors may affect the accuracy of detection of anomalies. Proper anomalous videos as well as normal videos should be treated for training the model. However, the main aim for anomaly detection is not detecting the accident accurately but to detect and differentiate the anomalous videos from the normal videos. However, since there are drastic changes into the environment due to climate changes and changing in the lightning, the approach creates false alarm rates for different kind of accidents.

II. BACKGROUND AND RELATED WORK

Here the work carried out and researched by other researches will be illustrated which are relevant and related to our research work. The sub-parts below are the most important key areas in our study.

A. Deep Spatio-Temporal Representation for Detection of Road Accidents Using Stacked Auto Encoder

In the proposed paper the approach was to automatically learn the feature representation of the videos treated to them from the spatio-temporal volume. In Spatio-temporal volume, the videos are stacked upon each other to find out spatial as well as temporal information of the object. In this paper, any accident I detected as unusual activity. Autoencoders which are trained on normal videos are used to extract deep representations. An autoencoder is an artificial neural network that learns to compress data from the input layer into a short code i.e. deep representation, and then uncompress that code into something that closely matches the original data. The accident is being detected by reconstructing error and likelihood of deep representation. The false alarm rate is being reduced by the intersection point of vehicles trajectories.

B. Real World Anomaly Detection in Real World

In this paper, to detect accidents in real world, both anomalous as well as non anomalous videos are being exploited. Annotating tags to anomalous or normal clips is quite time consuming therefore the paper proposes to learn anomaly through deep multiple instance ranking framework leveraging weakly labeled training videos i.e. training labels at video level instead of clips.

In this approach, they consider normal and anomalous videos as bags and video segments as instances and multiple instance learning (MIL) is used. It is a type of supervised learning in which instead of receiving a set of instances which are individually labeled, the learner receives a set of labeled *bags*, each containing many instances.

C. Adaptive Video-Based Algorithm for Accident detection on Highways

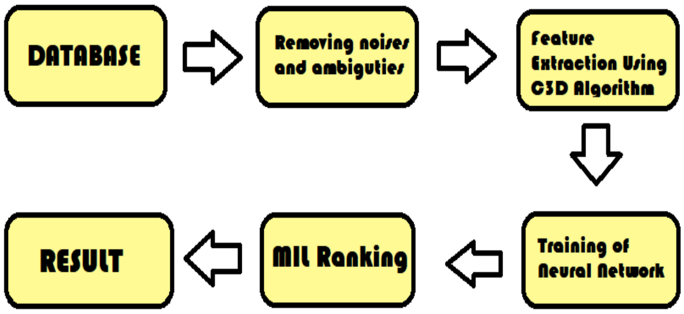
In this approach, first the traffic in a motion scene is considered as normal motion flow. An accident including one or more vehicles could be detected when there is a sudden or a sharp change their velocity.

It tracks the position of the pixels of the image and estimates the position of the next moving pixels between two frames. First the noise is being detected and filtered and then create a model to detect an accident.

III. Proposed System

As stated earlier, the proposed system would be focusing on creating the positive or negative bags of video instances for the training purpose, gathered from the dataset. Then extract features of that video and then give ranks to the bags to classify them as anomalous or normal incident.

So before getting into specifications of the proposed system, let us understand the overall flow of the system which is given below in the figure.



The overall flow of the system can be explained with the help of following steps:

1)Initially we will gather wide range of dataset which will contain anomalous as well as non anomalous videos.

2) Then we would remove all the videos which contains noise or other activities which would be non-suitable for the training set.

This can be done using a separate machine learning algorithm or manually $||V|| = ((x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2)^{-2}$ (1) θ = tan

3) Then we would run a supervised algorithm on these noise free videos along with machine instance learning (MIL)[7,8].

4)This algorithm will segregate these videos into negative bags and the positive bags.

The negative bags will the one which will contain normal videos whereas the positive bags will contain videos with anomalies.

5) After segregating the videos into positive and negative bags , these bags are given to the C3D feature extraction algorithm.

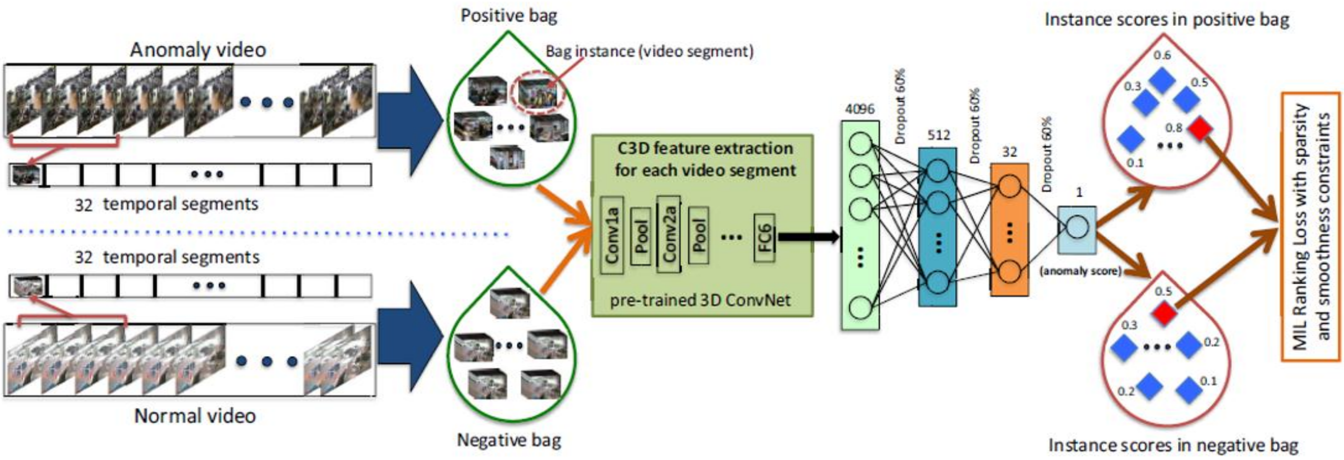


Figure 1. The flow diagram of the proposed anomaly detection approach. Given the positive (containing anomaly somewhere) and negative (containing no anomaly) videos, we divide each of them into multiple temporal video segments. Then, each video is represented as a bag and each temporal segment represents an instance in the bag. After extracting C3D features [6] for video segments, we train a fully connected neural network by utilizing a novel ranking loss function which computes the ranking loss

After running this algorithm on the video segments, we will get the video segments into binary format.

6) Then we input these features into a 3 layer neural network. The first layer has 512 units, the second layer has 32 units and the third layer has 1 unit of FC layer.

7)Then each video is divided into 32 non-overlapping segments and each segment is considered as an instance of the bag.

8) Now, we would calculate cosine similarity for all the sentences present in the document using the vector representation of the sentences and not the original sentences.

Each video segment in the bad is given scores for the training of the model ranging from 0 to 1.

9) After all the above steps, we use Machine Instance Learning algorithm for ranking loss with sparsity and smoothness constraints.

IV. Overview of the System

A. Dataset

The dataset used in the previous approaches were inconsistent and contained lots of noises which was defying our purpose and wasn't suitable for training our model.

The videos used for training in the previous videos were recorded and contained videos of only one locality. Also, the recorded videos in the dataset were short also some were abnormal and unrealistic.

The dataset we collected and used had many of the anomalies and other activities too such as human interactions and different lightning conditions. Hence, we segregated all the videos which contained such noises before we could treat it to our model for training. Doing so would result in increased accuracy of our model and give a high precision rate for the accident detection. All the videos collected and used are taken directly from the actual videos present and recorded from the surveillance cameras from the roads.

B. Implementation

After working on dataset and removing the noises and all the ambiguity from the dataset we work to extract the features from the datasets and work on it.

Once the training of the model is done, two bags of the video segments are being created. One is the positive bag; this bag contains the all the video segments that contains that contains anomalies or any of the accidents. The other bag; the negative bag, contains the videos of all the normal videos of car passing through the road.

After creating the bags, they are being extracted for the visual features. This is being done with use of the fully connected(FC) layer FC6 of the C3D network. Each of the video segment is being resized with each video frame to 240 x 320 pixels and the frame rate is being fixed to 30 fps. The C3D features are being computed for every 16 frames video clips which is then followed by l_2 normalization. The average of 16 frame video clip is taken within that segment to obtain features for that video segment. Then these features(4096) are then given as input to a 3-layer FC neural network. The first FC layer has 512 units , second FC layer has 32 units and the third has 1 unit FC layer. 60% dropout regularization [3] is used between FC layers. ReLU[4] activation and Sigmoid activation is used for the first and last layer respectively and employ Adagrad[5] optimizer with the initial learning rate od 0.001. The parameter of sparsity and smoothness constraints in the MIL ranking loss are set to $\lambda_1 = \lambda_2 = 8 \times 10^{-5}$ and $\lambda_3 = 0.01$ for the best performance.

Each video is then divided into 32 non-overlapping segments and consider each video segment as an instance of bag. The number of segments (32) is an empirically set. Also, multi-scale overlapping temporal segment was experimented but it does not affect the efficiency. Selection of 30 negative bags and 30 positive bags were made randomly. Gradient is computed by reverse mode automatic differentiation on computation graph by Theano[1].

Following previous work on anomaly detection , we use frame based receiver operating characteristics (ROC) curve and corresponding area under curve (AUC) to evaluate the performance of our method. As there is fault in anomaly detection using equal error rate (EER)[2] we do not use it.

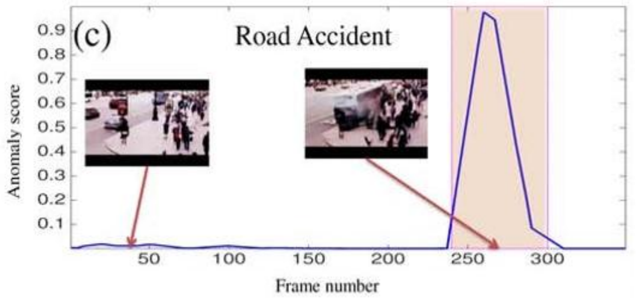


Figure 7. Qualitative results of our method on testing videos. Colored window shows ground truth anomalous region. It shows videos containing, *road accident*. The video frame segment showed a spiked up graph which contains accident.

V. ANALYSIS

- False Alarm Rate.

In real world, accidents are considered as a rare event and their frequencies are very much low compared to normal driving conditions. Along with this there are several factors involved which affects the accuracy of our model and leads to giving false alarm i.e. tagging or identifying normal, non anomalous video as anomalous video. A false alarm is usually detected when a vehicle is leading to or the gap between two vehicles decreases considerably such that the model detects its as accident, although the driver may prevent the accident. Also, it may give false alarm during the bad weather conditions as the visibility of the surveillance cameras goes down. During a rainy season, when the roads are wet and reflects image of the moving car, can be considered as a car and intersection of this reflection and actual car may be given as accident, leasing to false alarm. A sturdy and robust anomaly detection system should have its false alarm rate as low as possible on normal videos. Our system has comparatively low false alarm rate as compared to previous methods. Further false alarm rates can be further decreased by training the model more on the normal videos rather than anomalous videos and hence evaluating this model will give much lower false alarm rate.

- Model Training.

The assumption made in this approach is that to give lots of positive anomalous and negative non-anomalous videos with the video labels and the instance ranking to the model so that the neural network can automatically learn to predict the location of the anomaly in the videos. To achieve this, the model has to give highest rating to the video segment having anomalies during the training iterations. As we keep on training the model more and more on the anomalous and normal videos, the score of the anomalous videos goes on increasing and the score of the normal videos goes in decreasing. Although we don't use any segment level annotations, the network is able to predict the temporal location of an anomaly in terms of anomaly scores.

VI. Conclusion and Future Scope

In this paper, we have proposed and explained a deep learning approach to detect accidents taken place in real world using videos captured by surveillance cameras. As real-world anomalies are such a rare event we train our model on more on the normal videos and normal data. But training only on normal data may not be optimal for detection of anomalies, hence we attempt to exploit both anomalous and non-anomalous videos. We create positive and negative bags for anomalous and non-anomalous videos by training the model on the video segments using deep Machine Instance Learning framework with weakly labeled data. For validation of our approach, we introduce a new large scale anomaly dataset consisting of various accidents in real world. Resulting which we get a better accuracy of anomalies compared to the other baseline proposed systems. Further we demonstrate the usefulness of our dataset for the accident detection.

Our proposed systems can be installed and attached to the surveillance cameras to already installed.

VII. REFERENCES

[1] Theano Development Team. Theano: A Python framework for fast computation of mathematical expressions. *arXiv preprint arXiv:1605.02688*, 2016.

[2] W. Li, V. Mahadevan, and N. Vasconcelos. Anomaly detection and localization in crowded scenes. *TPAMI*, 2014.

[3] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 2014.

[4] G.E Hinton Rectified linear units improve restricted Boltzmann machines vinod nair. In ICML,2010.

[5] J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.*, 2011.

[6] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. Learning spatiotemporal features with 3d convolutional net- works. In *ICCV*, 2015.

[7] S. Andrews, I. Tsochantaridis, and T. Hofmann. Support vec- tor machines for multiple-instance learning. In *NIPS*, pages 577–584, Cambridge, MA, USA, 2002. MIT Press.

[8] T. G. Dietterich, R. H. Lathrop, and T. Lozano-P´erez. Solving the multiple instance problem with axis-parallel rectan- gles. *Artificial Intelligence*, 89(1):31–71, 1997.

[13] S. Ding, L. Lin, G. Wang, and H. Chao. Deep fea- ture learning with relative distance comparison for person re-identification. *Pattern Recognition*, 48(10):2993–3003, 2015. [14] J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.*, 2011.