TEST Engineering & Management

May - June 2020
ISSN: 0193-4120 Page No. 5228-5233

Crime Scene Prediction by Detecting Threatening Objects using Deep Learning Techniques

T Rishitha[1], Sini Robin[2], Sailaja Thota[3], Tejaswini N[4], Lakshmi K[5]

[1,2,3,4,5]School of Computing & Information Technology
REVA University, Bangalore, India
[1]rishithachandran.124@gmail.com, [2]sinirobin584@gmail.com, [3]sailajathota@reva.edu.in,
[4]tejum089@gmail.com, [5]lakshmik@gmail.com

Article Info
Volume 83
Page Number: 5228-5233
Publication Issue:
May - June 2020

Article History
Article Received: *19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *16 May 2020*

Abstract

Last couple of decades security has been a pressing issue and to overcome this, surveillance cameras are installed in several public places which helps in segregating crimes. Surveillance cameras allow us to watch and helps in creating a secure society. The data captured using cameras play a vital role in monitoring, predicting events and goal-driven analysis applications including anomalies and intrusion detection. The process followed in providing the input on the crime is to analyze the frames captured by the surveillance cameras and detecting the anomalies and sending an alert message to concerned authorities. This paper aims to put forth a unique method for anomaly detection based on deep learning techniques, which is designed by studying various existing models. Max pooling and ReLU are used in Convolutional Neural Network (CNN). Max pooling is a pooling operation that calculates the maximum, or largest, value in the patch of each feature map. This method has been evaluated using UCSD dataset and showed an increase in accuracy. Incorporating such techniques can help in crime detection at an early stage.

Keywords: *crime prediction, deep learning techniques, goal-driven analysis, anomaly detection, CNN, max pooling*

1. Introduction

As innovation advances and criminal plans become more and more sophisticated, police forces are counting on AI to secure society and ANN has a significant role in it. Crime can be categorized into five main types namely personal crimes, property crimes, statutory crimes, inchoate crimes and financial crimes. Identifying anomalies by observing normal patterns can have significant and various applications [2]. Furthermore, a Crime detection process is reliant on nature, setting and Crime situation [4,5]. In various situations, anomalies will be unique [2]. Existing regulated strategies for Crime detection, for example, simple CNN based techniques require labels which are hard to accomplish because of the video high dimension information. High resolution of video influences portrayal and production of a model. Right now, detection happens based on recordings of surveillance cameras. Slight anomalies in moving pictures like videos are difficult to identify as it requires video processing for detection methods.

The handling of surveillance cameras data in jam-packed scenes presents genuine difficulties and challenges. On the off chance if this procedure is online the complexity increases even more. The method proposed for processing data and at the same time accomplishing goal-oriented pattern is the implementation or use of advanced machine learning approaches such as deep learning methods. The benefit of these sorts of techniques, which typically have a high dimensional information, can be followed back to the presence of an end to end framework. End to end frameworks computerize extraction [6]. One of the fundamental reasons for using deep learning is to retrieve meaningful information from multi-dimension data.

The recent works done regarding this are done by detecting the threatening objects only in images. Here rime detection cannot be done in videos [3]. Even if surveillance cameras are used, they are used only for monitoring [9]. And in some only anomalies are detected which are different based on the scenario [7].

This paper presents a Crime detection strategy dependent on profound learning techniques. This technique has two principle components such as train network and detection classifier. The train network focuses on feature extraction and the subsequent one focuses on detection. Each part in detection component delivers a distinguished class as well as score. Finally, through the values of class and score, the ensemble classifier plays out the last detection and declares it. Advantage of this method is that the deep learning techniques can be used in the train-phase as well as test-phase. Using deep learning techniques in all phases of Crime detection is the primary intention of this work.

2. Literature Survey

A crime can be detected in a scenario by using various techniques. Crime is an action that is unethical, offensive and punishable according to our existing laws and anomaly is something that deviates from the norms of the society [1]. Predicting these anomalies in advance by taking the real-time data from the surveillance cameras can help the police to safeguard the society better. Many dynamics come into play while predicting a crime scene. To predict a crime, we need to ascertain the factors like the attacker's behavioral pattern, objects in his/her possession and background accurately. Blood, knives, firearms, toxins etc., can be considered as anomalies to predict a crime scene.

Research is being done for predicting human behaviour pattern in real-time. A Convolutional Long Short-Term Networks (CLSTN) technique is considered appropriate and implemented for dynamic human behaviour pattern detection and classification in videos [8]. Backdrop of a scene also plays a major role in accurate predictions. Foreground estimation is used in surveillance to eliminate background and there are many algorithms and techniques in computer vision to achieve the above [9]. One such method proposed for eliminating the background repetitive images is by using Gaussian Mixture Models (GMMs) and Kernel density estimation (KDE) and this is proven to be accurate while detecting or removing both static objects or moving objects in background [10].

Forensic team use methods like Ballistics, DNA Sequencer, luminol spray to accumulate evidence from a crime scene which can destroy a piece of crucial evidence. Hyperspectral imaging of the crime scene can facilitate in detecting and identifying bloodstains remotely which can help in preserving the crime scene [11]. CNN algorithms such as AlexNet, GoogLeNet, ResNet can be used for background elimination and object detection in the crime scene such as blood, knife, gun etc. accurately [3].

3. Proposed System

The data from the cameras are stored in the cloud storage and processed. Each module will perform its specified task. The frames are going to be tested and anomalies are identified and validated based on score. The concerned authorities are then notified to proceed further.

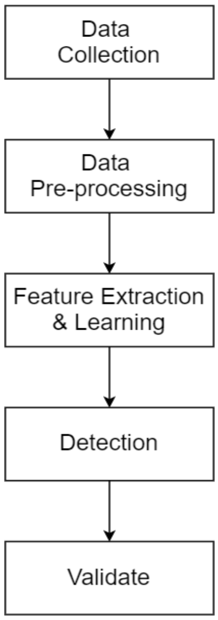

Figure 1: Crime Detection System Flow Diagram

4. Algorithms & Techniques Used

CV2 in simple terms is a library used for Image Processing. Its usability is mainly regarding operations associated with Images. Take a video as input and using CV2 break the video into frame by frame and save those frames. These images are then given as input to the CNN which is the algorithm, we will be using here to train our model. Consider any image, as we know in simple terms every image represents some pixels. We analyze the influence of nearby pixels in a picture by using a filter (can be called as weights, kernels or features). Filter is a tensor that keeps records of spatial information. It learns to extract features of objects in a convolutional layer. The key part is to detect edges within the images and these are detected by the filters. It helps to eliminate unwanted information to amplify images. There are agile filters where the changes in intensity occur very quickly like from white to black pixel and vice-versa.

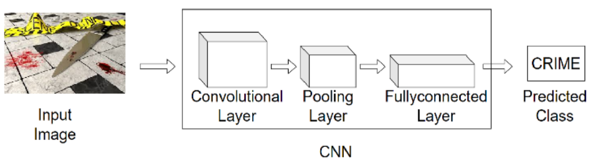

Figure 2: Working of CNN

Let's consider an image of size 4X4 size and 1 filter 2X2 size. Usually, these filters are continuous values. We took 0 and 1 for our convenience.

Here, Filter Size(X) = 2x2 → 2
Input size(Y) = 4x4 → 4
Stride(S) = 1x1→ 1 (1 No Cell moves)
Padding(P) = 0x0 → 0 (No padding)
Output Size = **(Y – X + 2P) * S + 1**
= (4 – 2 + 2*0) * 1 +1
= 3

So, the output matrix here will be 3X3.

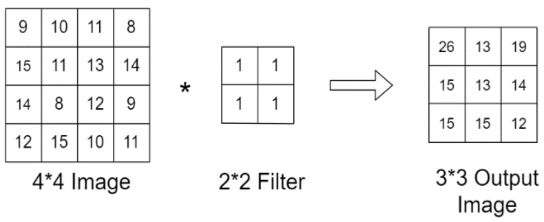

Figure 3: Figure showing how feature map is obtained

The filter convolutes with the image to detect features and patterns. Then the convolution of 4x4 image matrix multiplies with 2 x2 filter matrix called "Feature Map". We apply the dot product to the scaler value then move the filter by the stride over the whole image. Sometimes filter doesn't fit perfectly fit the input image then we need to pad the image with zeros so the filter perfectly fits the given image. It is called as padding. Next, if the images are too large, we need to scale down the size of it. Adding pooling layer decreases the scale of the image and hence decrease the complexity and computations. Next Step, is Normalization for which activation function ReLU (Rectified Linear Unit) is employed before max pooling.

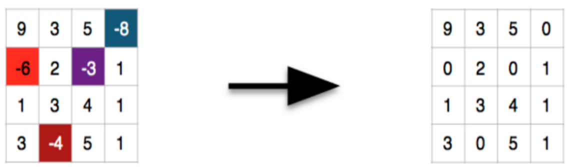

Figure 4: Figure showing how ReLU works

The purposefulness of ReLU is to add non-linearity to the convolutional network. In general, we want our network to learn non-linear values from the real-world data which may be negative sometimes. A rectified linear unit returns output 0 if the input is less than 0, or original input otherwise as shown in Fig 4. The output is function ***f(x) = max (0, x)***.

$$RELU(x) = \begin{cases} 0 \ if \ x < 0 \\ x \ if \ x >= 0 \end{cases}$$

Max pooling returns only maximum values and returns a maximum negative value if the input is less than 0. So, this is used after ReLU. It chooses the maximum element from the area of the feature map and gives as output which is a feature map that contains most important features of the earlier feature map. Fig 5 shows how max-pooling returns the max elements from the rectified filter feature map in two different cases.

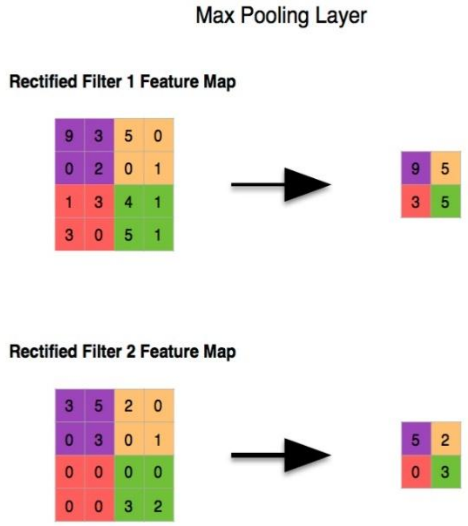

Figure 5: Figure showing how Max Pooling is employed

The final step is to convert our matrix into a column vector and feed the values to fully connected layer as shown in Fig 6. Then, we need to train the model to detect anomalies the same way that we train other neural networks.

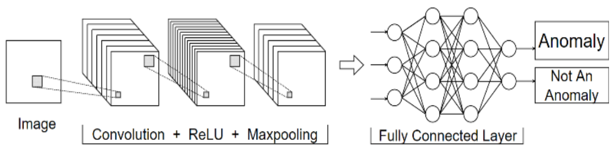

Figure 6: Step by step process of how CNN detects Anomalies

5. Methodology & Modules Identified

The proposed method of this work relies on deep learning techniques for identifying anomalies in video. Even this machine learning approach has two main phases; train phase and test phase. Three main modules in this method are i) Data Pre-processing step which is expounded to background elimination, ii) the feature selection and learning and iii) Identifying anomalies. Background elimination, Feature selection and Learning are part of train Phase while Detection module is parts of test-phase. Dataset that has normal frames is used for training the features in the train-phase. The dataset of frames that contain abnormalities will be given as input in the test phase.

Feature selection and Learning has four main features; object detection, the density of each color, direction of object flow and scene recreation; called learned features. In a few types, to reduce the cost and train time, feature extraction processes are performed on single frames and others are based on patch frames. Comparing each frame to the prior and next frames determines the detection score. The ultimate score is created based on the comparison of frames and average speed. The combination of these features is used to identify objects and generate scores which are used to identify anomalies.

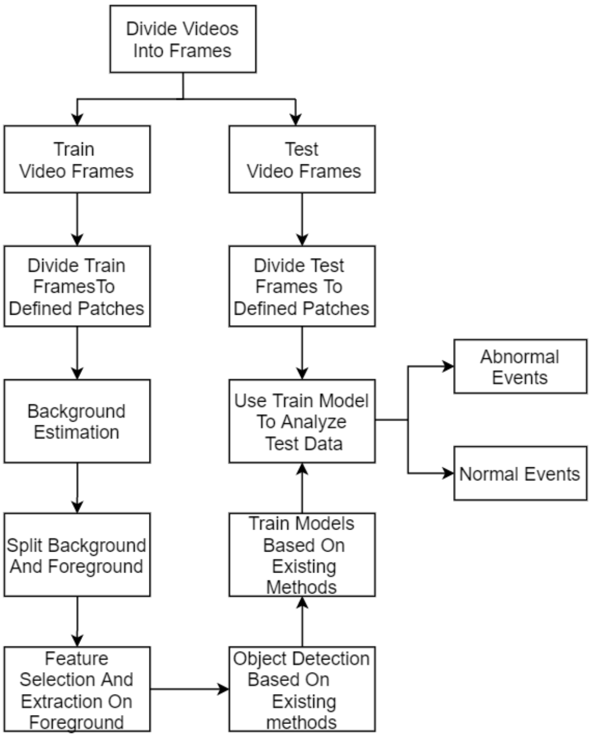

Figure 7: Video crime Detection Flow Diagram

A. Background Elimination

The primary action before feature selection and learning module is to evaluate and take out the background which will reduce the cost of computing and processing time. Techniques for expelling background will vary as it won't be same for different situations. On the off chance it can be comprised of void spaces or road outskirts, at that point, its elimination depends on the Most Occurrence of Frequency (MOF) in video patches. Contingent upon pixels and their area in the picture, at first, a histogram is produced for each frame in the video patch. Later the histogram of the frames in each patch is contrasted and one another. At that point, the most extreme qualities in each patch are recognized as background and are greyed out.

B. Feature selection and Learning

The feature selection and learning module has four components to suit each feature. The essential component for object selection incorporates a stacked denoising auto-encoder (SDAE) with 6 layers encode layer and unravel layer with a similar structure. Each image is convolved with 1*1 filter which fuses cushioning and walk. The yield of this part is a data consisting of objects which are to be distinguished which is utilized in the identifying anomalies and as input to generate feature map. This aides in increasing the accuracy of identifying anomalies.

The density of each color in the image is computed by CNN using 8*8 filter. The output of this component is a feature map. The loss function is determined utilizing the square error. In this part is the background isn't taken to consideration.

The third component is to remove the movement of the object flow. It performs feature selection in the same direction of moving objects within the frames of video patches. This part has a system like the segment used for object selection however it relies upon frames in the video patches. Optical flow of the frames determined based on the comparison of frames in a patch in after feeding them to the model. The output of this component is the direction of the object flow which will prove to be useful later on.

The last component is a reconstruction component to reconstruct the scene. For which convolutional Auto-Encoder is used which is similar to CNN but has a fully connected auto-encoder within the convolutional layer. It has encoder and decoder with a similar structure to regenerate the scene. The encoder regenerates the scene to reconstruct the frames dependent on the prior and the next frames. The decoder compares the created scene with the original one to calculate reconstruction error which is used in identifying anomalies. The reconstruction error generated during the train-phase; which is generally low; will be the measure for detecting the anomalies.

Toward the finish of the train-phase, learned feature components are generated. Which can be utilized separately or joined with other components to predict Crime.

C. Identifying anomalies

In this module, the output generated after the object detection in train-phase is given to the classifier in a fully connected layer as input with two labels; normal events and abnormal events. The output generated after object detection and after reconstruction of the scene is given combinedly to identify anomalies accurately. Anomalies are identified accurately because of lower reconstruction error in the respective frame.

Density of each color and direction of motion of objects are given combinedly cause these two are complementary and the direction of optical flow should be the same as the direction of the density of each color increases.

6. Result and Discussion

A video is uploaded which is divided into frames using CV2. Pretrained models helps in detecting anomalies. Finally, an output video is obtained with the alert message “Anomaly Detected” printed as shown in Fig 8 and also the concerned authorities are notified via message as shown in Fig 9. A Receiver Operating Characteristic (ROC) curve is plotted by measuring true positive rate and false positive rate using matplotlib as shown in fig 10. This helps us assessing the performance of the model.

Figure 8: Screenshot of the video when an anomaly is detected

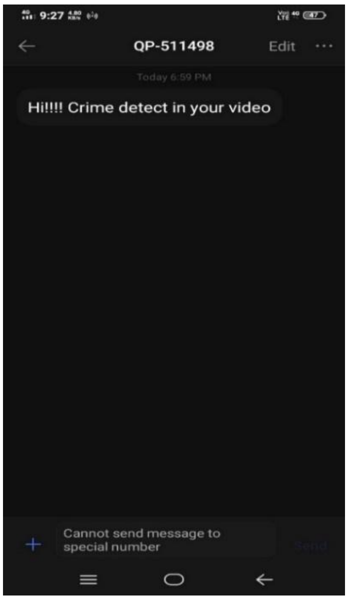

Figure 9: Screenshot of the message

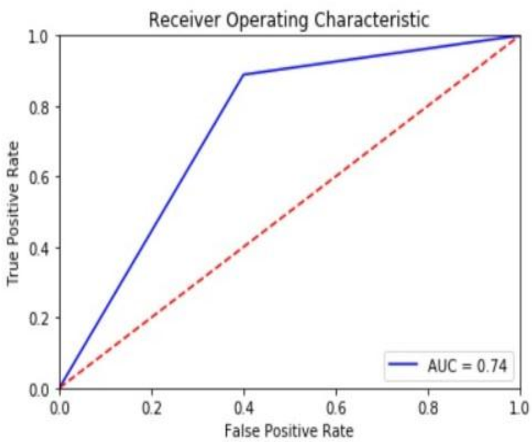

Figure 10: ROC Curve

7. Conclusion

In the current work, a unique method based on deep learning is used for Crime detection through footage in video surveillance cameras. ReLU (an activation function) and Max Pooling are used in CNN to obtain the desired results. Widely used crime detection dataset, UCSD dataset, was used to evaluate the modules of this method. Isolation of training phase is another added advantage in using this method. Hence, it can be used as a pre-train model in similar works.

In future, we can include a feature which might add descriptions to each of the detection classifiers or down to the last one which will enable to ascertain the crime accurately.

References

[1] Saini, Dinesh Kumar, Dikshika Ahir, and Amit Ganatra. "Techniques and challenges in building intelligent systems: anomaly detection in camera surveillance." *Proceedings of First International Conference on Information and Communication Technology for Intelligent Systems: Volume 2*. Springer, Cham, 2016.

[2] Kiran, B. Ravi, Dilip Mathew Thomas, and Ranjith Parakkal. "An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos." *Journal of Imaging* 4.2, 2018: 36.

[3] Nakib, Mohammad, et al. "Crime Scene Prediction by Detecting Threatening Objects Using Convolutional Neural Network." *2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*. IEEE, 2018.

[4] Hinami, Ryota, Tao Mei, and Shin'ichi Satoh. "Joint detection and recounting of abnormal events by learning deep generic knowledge." *Proceedings of the IEEE International Conference on Computer Vision*,2017.

[5] Ribeiro, Manassés, André Eugênio Lazzaretti, and Heitor Silvério Lopes. "A study of deep convolutional auto-encoders for anomaly detection in videos." *Pattern Recognition Letters* 105, 2018: 13-22.

[6] Sabokrou, Mohammad, et al. "Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes."*Computer Vision and Image Understanding* 172,2018: 88-97.

[7] Chong, Yong Shean, and Yong Haur Tay. "Abnormal event detection in videos using spatiotemporal autoencoder." *International Symposium on Neural Networks*. Springer, Cham, 2017.

[8] Wang, Shuqin, et al. "Dynamic Human Behavior Pattern Detection and Classification." *2019 IEEE Fifth International Conference on Big Data*

May - June 2020
ISSN: 0193-4120 Page No. 5228-5233

Computing Service and Applications (BigDataService). IEEE, 2019.

[9] Kaur, Shanpreet. "Background Subtraction in Video Surveillance.", 2017.

[10] Edelman, G. J., T. G. Van Leeuwen, and M. C. G. Aalders. "Hyperspectral imaging of the crime scene for detection and identification of blood stains." *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIX*. Vol. 8743. International Society for Optics and Photonics, 2013.

[11] Olmos, Roberto, Siham Tabik, and Francisco Herrera. "Automatic handgun detection alarm in videos using deep learning." *Neurocomputing* 275 ,2018,: 66-72.

[12] Hinami, Ryota, Tao Mei, and Shin'ichi Satoh. "Joint detection and recounting of abnormal events by learning deep generic knowledge." *Proceedings of the IEEE International Conference on Computer Vision*, 2017.

[13] Chackravarthy, Sharmila, Steven Schmitt, and Li Yang. "Intelligent Crime Anomaly Detection in Smart Cities Using Deep Learning." *2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC)*. IEEE, 2018.

[14] Ye, Muchao, et al. "Anopcn: Video anomaly detection via deep predictive coding network." *Proceedings of the 27th ACM International Conference on Multimedia*, 2019.