International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 06,2020
ISSN: 1475-7192

Crime Detection and Evidence Extraction Using Machine Learning on Cloud

[1]Dr. E. Karunakaran,[2]Dr. J. Kumaran @ Kumar,[3] R.Sudhakar

ABSTRACT--*Digital video plays a vital role in evidence identification, analysis, presentation, and report recently. With the convenience use of smartphones and the increasing popularity of surveillance camera, visual data are hugely being used in digital crime investigation. Extracting evidence becomes a time-consuming process when searched manually. To ensure reduction in possible human errors and for better accuracy, machine learning model is used which speeds up the investigation by extracting crime scene objects and masked faces from the crime scene footages. The main approach of this paper is to develop such machine learning model to assist forensic investigation for better evidence extraction with higher accuracy. In order to further reduce the processing time of the object extraction model, parallel processing is introduced which helps the model to perform faster in larger data sets. The evidence extracted from the video are stored in the cloud storage repository. Storing of these extracted evidences in the cloud helps to compare the extracted evidences with the past crimes and these extracted crime scene objects can be used to train the model further in order to improve the detection rate.*

Key words—*detection, evidence extraction, learning.*

I. INTRODUCTION

The Closed-circuit television (CCTV) camera are mainly used on most of the places like malls, traffic signals, bank, shops, including home. These cameras and many daily use devices such as mobiles, smartwatches help to record videos which can be used for crime scene investigation [1]. Searching videos and other resources for forensic investigation manually is a very time-consuming process. This process can be simplified with the help of machine learning. An image detection and classification model built using deep neural network is fine tuned to recognize crime scene objects which can be used to extract evidences from video footages. This model is trained on a large collection of dataset images in order to identify the crime scene places and objects automatically. Accordingly, due to the computational cost of training such models, it is a common practice to import such models from published literatures such as AlexNet, VGG, Inception, MobileNet and train further effectively. This process uses a simple CNN model, AlexNet and ResNet34 models to train our custom data. To further improvise the model, parallel computing is used which helps to reduce the time taken to process and analyse large data sets.

The CCTV systems used in surveillance allows continuous recording of videos and stores it as footages. Analysing huge volumes of these footages manually is a very time-consuming and challenging process. In order to solve this issue, this paper focuses on using these footages extracted manually by crime scene investigators to automatically analyse a huge volume of video files with the help of machine learning. These footages extracted

[1] *Associate Professor, Department of Computer Science and Engineering, Pondicherry Engineering CollegePillaichavady, Puducherr605014.*
[2] *Assistant Professor, Department of Computer Science and Engineering, Pondicherry Engineering CollegePillaichavady,Puducherry605014.*
[3] *M.Tech., DCS, Department of Computer Science and Engineering, Pondicherry Engineering CollegePillaichavady, Puducherry-605014*

Received: 22 Sep 2019 | Revised: 13 Oct 2019 | Accepted: 15 Jan 2020

can also be used in crime scene investigation when such suspicious activities take place. However, the strength of evidence extracted from these footages significantly depends on the contents recorded by CCTV cameras. Therefore, the quality of the footage is very important for the contents-based evidence recorded which needs to be re-formatted or converted to a suitable format for easier investigation.

This approach mainly uses the following techniques or streams to build a model for evidence analysis and extraction – 1) Machine Learning 2) Deep Learning 3) Parallel Computing 4) Cloud Computing. The main aim of forensic investigation is to identify strong evidence at different level. In this paper, we focus on the video footages to develop efficient video analysis technique with huge accuracy from the viewpoint of forensics using machine learning on cloud.

II. RELATED WORKS

Mostly images and videos obtained from the surveillance devices are widely used to retrieve major evidence for detecting crime scenes. But the evidence was always extracted manually by the crime investigation team. Gathering such evidences manually was not just a tedious process but, time-consuming as well. As a result of this manual process, many pre-trained model and frameworks were created to simplify the process in an automated way.

Various work done related to automated crime detection involves evidence analysis and extraction, hyper-parameter optimization, CCTV quality assessment, video stream analysis in cloud, classification and search based on spark.

A. EVIDENCE ANALYSIS AND EXTRACTON

The paper [2] uses YOLO (You Only Look Once) [3] network model which is a fast-real time multi object detection algorithm. The paper uses yolov3 pre-trained weights to detect objects and faces from video footages. Though algorithm works fast, the detection rates are low.

B. HYPER-PARAMETER OPTIMIZATION

Hyper-parameter tuning [4] chooses a set of optimal hyper-parameters for a learning algorithm. This technique helps in improving the performance of the algorithm. It uses Apache SPARK for parallel video analysis which has an accuracy of 97% and precision of 96%.

C. CCTV QUALITY ASSESSMENT

The CCTV video quality differs from one device to another. Hence, in order to get better results, the CCTV video quality is enhanced [1]. AwareABIS (Automated Biometric Identification System) supporting fingerprint, face and iris modalities are widely used for large-scale biometric identification and de-duplication. The camera position also impacts the results. The farther the camera, the lesser the detection rate.

D. CLASSIFICATION AND SEARCH BASED ON SPARK

PF-Face provides distributed data storage mechanism, which can transfer massive data from local file system to Hadoop Distributed File System (HDFS) [5]. Spark is used as the computing layer to achieve the process with iterative algorithms efficiently, in which Resilient Distributed Databases (RDDs) can be continuously stored in memory. In order to enable deep learning, training and testing on PF-Face, CaffeOnSpark is used to process the deep learning algorithms.

E. VIDEO STREAM ANALYSIS IN CLOUD

An Object Detection and Classification Framework for High Performance Video Analytics [6] focus on building a scalable and robust cloud computing platform for performing analysis of thousands of recorded video streams with high accuracy and detection automatically. This criterion defines parameters for detecting objects of interests such as face, car, van or truck and size/colour-based classification of the detected objects. The recorded video streams are then automatically obtained from the cloud storage which are decoded and analysed on cloud resources. Some of the limitations and disadvantages from all these works led to the development of this pre-trained model on cloud which helps to detect the crime scenes and identify the evidences and the objects at a much higher rate of accuracy.

III. CRIME DETECTION AND EVIDENCE EXTRACTION

Digital videos from mobile devices, CCTV and Internet (Social Network – Facebook, Instagram, etc.,) are being widely used as a key evident source in crime detection and analysis. Video and images extracted from these devices are helpful in identifying the crime place and many other suspicious activities. The video quality plays a key role in forensic analysis. High quality video plays a major role to conduct efficient crime investigation. So, the video extracted from the devices should be enhanced and resized accordingly in order to proceed further. This process can be simplified with the help of machine learning and parallel processing.

An image detection and classification model built with the help of deep neural network (Deep Learning) is fine tuned to recognize crime scene objects which can be used to extract evidences from these video footages. To further improvise the model, parallel computing is used which helps to reduce the time taken to process and analyse large data sets.

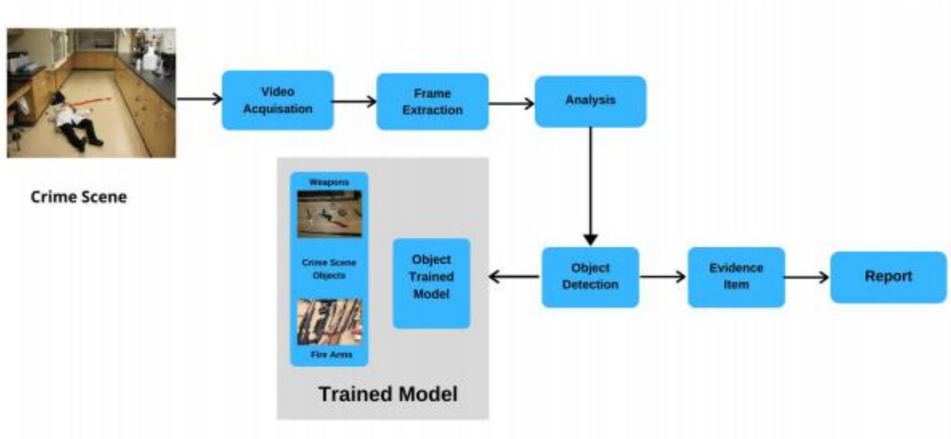

Figure 1: Architectural Design of Existing Systems

This process includes four main basic components – 1) Identifying the crime scene and gathering the video data from the crime place. 2) Extracting the frame from the video data to be accessed parallelly. 3) Enhancing the video quality for further investigation. 4) Gathering the evidence and the crime items in an automated way using a trained model based on machine learning.

A. IDENTIFYING THE CRIME SCENE AND GATHERING THE VIDEO DATA

A crime scene is a location where criminal or suspicious activity occurred. Crime scenes contain physical evidence which is needed for criminal investigation. This evidence from the crime place are collected by crime scene investigators (CSIs) and Law enforcement. All the possible videos collected from the crime place using devices such as CCTV, smartphones, Internet etc. are further used to collect evidences.

B. FRAME EXTRACTION FROM VIDEO

Since the original video retrieved from the crime place usually contains large volumes of duplicate frames, many frames are extracted from the collected video. Extracting case related videos and grouping them in an organized manner are two key tasks, which helps in identifying variety of multimedia evidences. To extract frames from video, OpenCV can be used. OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library which comes with many powerful video editing functions. In current scenario, techniques such as frame extraction and frame resizing can be accomplished using OpenCV.

C. ENHANCEMENT OF FRAME AND VIDEO QUALITY

Due to the different varieties of devices, many digital video evidences are not presented in standard formats. The video quality differs from one another when collected from various sources. It is mandatory to convert these videos to standard format without degrading the video quality. The organized video frames or clips need to be improved to provide additional features and uncover hidden information. This stage should be capable of reducing many different types of noises in videos such as noisy audio recordings as well as perform video editing, enhancement, filtering products, extraction *etc.*

D. OUTPUT AND EVIDENCE EXTRACTION

The Evidence items are extracted from the enhanced videos using a trained model. The objects extracted using these trained models can be further analysed to showcase as evidence. These objects can be used for future investigations and can be used as data set for training the model to improve its performance. Once the crime scene objects are extracted, they are sent to the result storage repository which consists of various crime scene objects extracted from several crime scenes. They can also be used in the future to analyse a different crime scene.

Video footages from crime scenes are collected. The videos are then pre-processed. The pre-processing stage involves extraction of frames and resizing the frames to best suit the model in which the frames are to be processed. The extracted frames are further enhanced to proceed with the investigation. This enhanced video helps to classify the object and the crime items using a pre-trained model. The proposed model helps to speed up the process and reduces possible human errors, which in turn helps to move forward with further investigation.

IV. CRIME DETECTION AND EVIDENCE EXTRACTION PROCESS USING MACHINE LEARNING ON CLOUD

The proposed system uses cloud for data storage and processing. Video footages from crime scenes are collected. The videos are then pre-processed. The pre-processing stage involves extraction of frames and resizing the frames to best suit the model in which the frames are to be processed. The models AlexNet [6] and ResNet-34 [7] are trained and fine-tuned with custom data consisting of crime scene objects to detect such suspicious items used for object classification. In addition, Parallelism is introduced in order to speed up the process of evidence extraction.

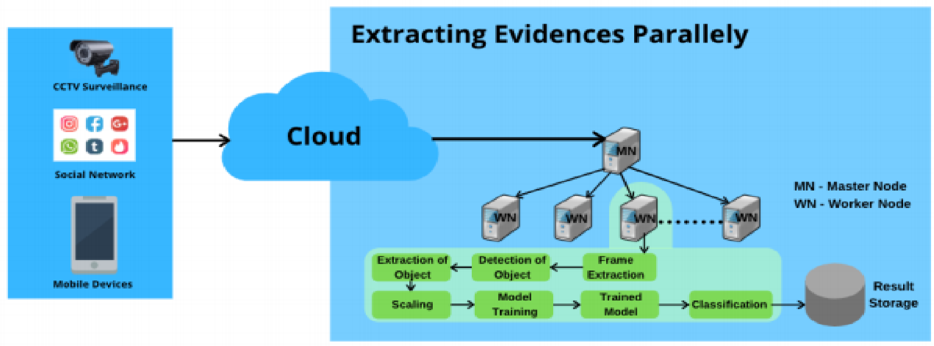

Figure 2: Architectural Diagram of Proposed Model

The various stages involved in developing this trained model are as follows-

1) CRIME SCENE DETECTION

Videos from different sources such as CCTV, Mobile Devices from crime scene and Videos found on Social Networks related to the crime are collected. The collected data are sent to the cloud for storage and evidence extraction.

2) CLOUD STORAGE

The videos collected related to the crime scene are sent to the cloud for storage and processing. Storing the crime scene footages on cloud has several advantages such as, it is easy to investigate evidences related to the crimes from the past footages and link them with the current crime for which the evidences are to be extracted.

Pre-processing and object classification are done in the cloud which has less maintenance and low cost when compared to other kind of processing. [8] The evidence extracted is stored in the result storage repository from which the evidences can be accessed any time.

3) PRE-PROESSING AND FRAME ANALYSIS

Since the original video extracted usually contains large volumes of duplicate frames, many frames are extracted from the collected video. However, the strength of evidence extracted from these footages significantly depends on the contents recorded. Therefore, the quality of the video footage needs to be re-formatted or converted to a suitable format for easier investigation. The re-formatted video should be further enhanced in order to increase the quality of the video. To extract and enhance the quality of frames from the video, OpenCV can be used which is the leading open source library for computer vision, image processing and machine learning. Additionally, it features GPU acceleration for real-time operation.

4) ROLE OF VARIOUS NODES ON FRAME EXTRACTION

Video footages collected from the crime scenes are large and are difficult to process. Hence, the frames are extracted and are sent as batches to multiple nodes where they are processed [9]. Frames received at each node are processed and crime scene objects from the frames are extracted. The extracted objects from each node are sent back to the result storage repository. The parallel computing used has a master node and several worker nodes.

a) MASTER NODE

The master node collects data from the cloud storage. The data are then pre-processed, and frames are extracted. The extracted frames are rescaled and are sent to the worker nodes as batches. Each batch consists of a set of frames extracted from the video based on the video length.

b) WORKER NODE

The worker node processes every frame received from the master node. A custom model built to extract crime scene objects is used to process the frames. Once the crime scene objects are extracted, they are sent to the result storage repository so that they can be used in the future to analyse a different crime scene.

5) OBJECT CLASSIFICATION

The pre-trained model uses the dataset consisting of crime scene images to classify the object and crime items using machine learning. There are many pre-trained models available. Each model has its own advantages and disadvantages. Models such as AlexNet, VGG, ResnNet, DenseNet and MobileNet are studied and are trained with custom datasets having a large collection of crime scene objects.

The performances of these models are analysed and the model with higher accuracy rate is chosen to classify the crime scene objects and images.

6) AUTOMATIC CRIME OBJECT DETECTION USING TRAINED MODEL

A custom model is built by fine tuning based on the results from the above-mentioned models. The task of fine-tuning a network is to tweak the parameters of an already trained network so that it adapts to the new task at hand. Fine-tuning a network is a procedure based on the concept of transfer learning. We start training a CNN model to learn features for a broad domain with a classification function targeted at reducing error in that domain. Then, we replace the classification function and optimize the network again to minimize error in another, more specific domain where we transfer the features. This trained model is developed using Python 3.6 with Ubuntu 16.04 as the operating system. Along with this, the packages TensorFlow and Keras are used to train the model more efficiently. The data set consist of images of different pixels. Hence the images are rescaled to a fixed resolution such that it matches the input size of the training model. AlexNet accepts a resolution of 224 x 224 and

ResNet accepts a resolution of 256 x 256 as input. The images are rescaled using the OpenCV library functions. The cv2.resize() function is used to resize the original image. Two separate dimensions are given for AlexNet and ResNet models. Three different models are trained with custom data having crime scene objects. The results from various models are compared with different epochs.

7) OBJECT DETECTION AND ANALYSIS

Once the model is trained, the trained model is used to classify a set of test images. The test image set consist of images from all the classes that are detected in the crime place. These images help to extract the evidence automatically in order to proceed further for the forensic investigation. In addition, this dataset can also be compared with the past crime scenes and can be used in future for further enquiries. The evidence and the objects collected using this model are finally sent to the result storage repository.

8) EVIDENCE AND OBJECT STORAGE REPOSITORY

The result storage repository consists of various crime scene objects extracted from several crime scenes. These objects can be used for future investigations and can be used as data set for training the model to improve its performance.

V. EXPERIMENTAL SETUP

There are many pre-trained models available such as AlexNet, VGG, ResNet, DenseNet and MobileNet.

Each model has its own advantages and disadvantages. In this paper, three models namely, AlexNet, ResNet and DenseNet are studied and are trained with custom datasets having a large collection of crime scene objects.

The three models are trained with custom datasets consisting of 600 images each. The dataset consists of collection of images related to crime scene. Five different classes of images are collected namely handguns, machine guns, masked faces, RPG (rocket-propelled grenade) and grenades. Once the model is trained, the trained model is used to classify a set of test images which consists of images from all classes. The trained models are tested using testing data consisting of 200 images each from which the accuracy rate is calculated at different epochs which is shown in Fig.3.

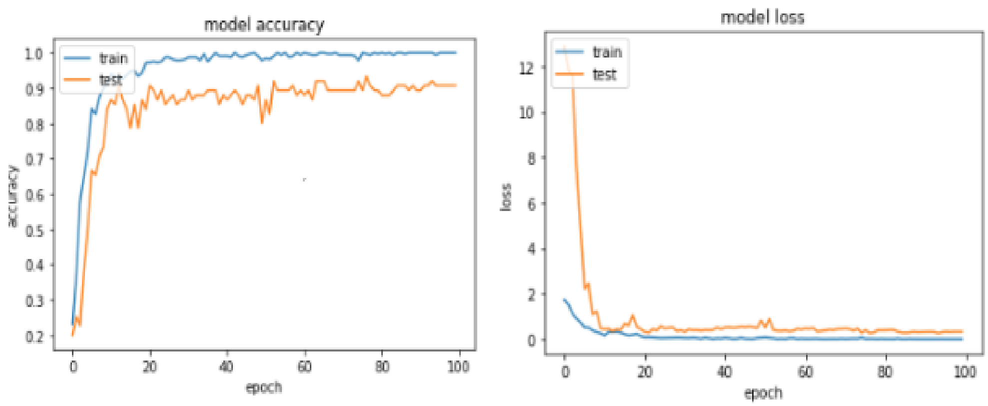

Figure 3: ResNet-34 – Epochs: 100 – Accuracy and Loss Graph

Received: 22 Sep 2019 | Revised: 13 Oct 2019 | Accepted: 15 Jan 2020

The results from various models are compared with different epochs and the performance of these models are analysed. The Performance Comparison of these trained models are shown in table 1.

Table 1: Accuracy Comparison of Trained Models

Model	Epochs-10	Epochs-25	Epochs-50	Epochs-100
AlexNet	61.33%	82.67%	86.67%	85.33%
ResNet - 34	65.33%	80.00%	84.00%	90.67%
DenseNet	63.42%	78.16%	79.36%	83.57%

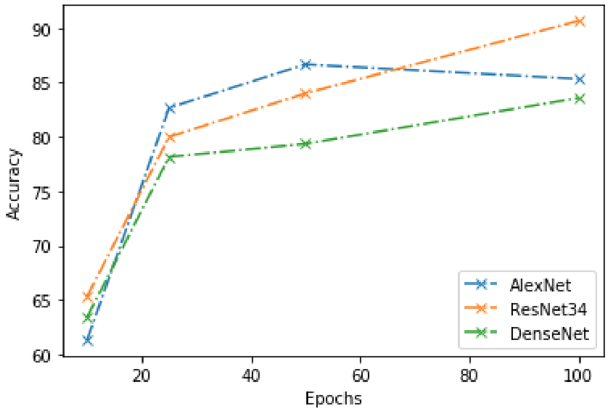

Figure 4: Accuracy Graph of Trained Models

Each model is trained at Epochs-10, Epochs-25, Epochs-50 and Epochs-100 respectively. A custom model is built by fine tuning based on the results from the above-mentioned models. Compared to AlexNet and ResNet, DenseNet provided lower accuracy rate at all epochs. AlexNet provided good accuracy rate of 82.67% at Epochs-25 and 86.67% at Epochs-50 respectively. Compared to AlexNet, ResNet-34 provided a higher accuracy rate of 65.33% at Epochs-10 and 90.67% at Epochs-100 respectively as shown in Fig. 4. In order to classify objects, the model with highest accuracy rate is chosen for better performance. Since, ResNet-34 has higher accuracy rate and performance comparatively, ***ResNet-34*** is chosen to classify the objects from the dataset.

VI. CONCLUSION

Evidence extraction in digital forensic investigation is a tedious process. The proposed model will be able to reduce the time for extracting evidences and minimize the human power needed. It can also be used for comparing extracted objects from past records to the current crime scene. In this paper, the performance of models such as AlexNet and ResNet are compared with custom data as input. The results will be used to fine tune a custom model to increase the performance. The model will be implemented in cloud and parallelism will be used to reduce the processing time. The future work will include building of a custom model to extract crime scene objects with better accuracy rate.

REFERENCES

1. Mohamad Firham Efendy Md Senan, Siti Norul Huda Sheikh Abdullah, Wafa Mohd Kharudin, Nur Afifah Mohd Saupi. "CCTV Quality Assessment for Forensics Facial Recognition Analysis." In 2017 7th International Conference on Cloud Computing, Data Science & Engineering, pp.649-655. IEEE, 2017.
2. Jianyu Xiao, Shancang Li, Qingliang Xu. "Video-Based Evidence Analysis and Extraction in Digital Forensic Investigation", Special Section on Deep Learning: Security and Forensics Research Advances and Challenges. IEEE, 2019.
3. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. "You Only Look Once: Unified, Real-Time Object Detection." arXiv preprint arXiv: 1506.02640v5, 2016
4. Muhammad Usman Yaseen, Ashiq Anjum, Omer Rana, and Nikolaos Antonopoulos. "Deep Learning Hyper-Parameter Optimization for Video Analytics in Cloud" In Systems, Man, And Cybernetics: Systems, pp.1-12. IEEE, 2018.
5. Jinna Lv, Bin Wu, Chang Liu, Xiaoyan Gu. "PF-Face: A Parallel Framework for Face Classification and Search from Massive Videos Based on Spark." In 2018 IEEE Fourth International Conference on Multimedia Big Data, IEEE, 2018.
6. Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks." papers.nips.cc. https://papers.nips.cc/paper/4824 -imagenet-classification-with-deep-convolutionalneural-networks.pdf
7. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun." Deep Residual Learning for Image Recognition" arXiv preprint arXiv: 1512.03385v1, 2015
8. Yuyong Cui, Zhiyuan Zeng, Liu Liu "An Object Detection Algorithm Based on the Cloud Model" In
9. Anjum, T. Abdullah, M. Tariq, Y. Baltaci, and N. Antonopoulos, "Video stream analysis in clouds: An object detection and classification framework for high performance video analytics." IEEE Trans. Cloud Computing. IEEE, 2016.
10. Seoyoung Choi, Eli Salter, Xuyun Zhang, Burkhard C. Wunsche. "Bird Eyes: A CloudBased Object Detection System for Customisable Surveillance" In 2018 International Conference on Image and Vision Computing New Zealand (IVCNZ). IEEE, 2018.