Anomaly Detection In Surveillance Videos Using Deep Learning

03CS7914 Project (Phase II)

CHN20MT013 CHN20CSIP06 Shivangi M
chn20csip06@ceconline.edu
M. Tech. Computer Science & Engineering (Image Processing)

Department of Computer Engineering
College of Engineering Chengannur
Alappuzha 689121
Phone: +91.479.2165706
http://www.ceconline.edu
hod.cse@ceconline.edu

College of Engineering Chengannur
Department of Computer Engineering

C E R T I F I C A T E

This is to certify that, this report titled ***Anomaly Detection in Surveillance Videos using deep learning*** is a bonafide record of the work done by

CHN20MT013 CHN20CSIP06 Shivangi.M

Fourth Semester M. Tech. Computer Science & Engineering (Image Processing)

student, for the course work in **03CS7914 Project (Phase II)**, under our guidance and supervision, in partial fulfillment of the requirements for the award of the degree, M. Tech. Computer Science & Engineering (Image Processing) of **APJ Abdul Kalam Technological University**.

Guide

Shiny B
Asst. Professor
Computer Engineering

Coordinator

Ahammed Siraj K K
Associate Professor
Computer Engineering

Head of the Department

July 16, 2022

Dr. Manju S Nair
Associate Professor
Computer Engineering

Permission to Use

In presenting this project dissertation at College of Engineering Chengannur(CEC) in partial fulfillment of the requirements for a postgraduate degree from APJ Abdul Kalam Technological University, I agree that the libraries of CEC may make it freely available for inspection through any form of media. I further agree that permission for copying of this dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the Head of the Department of Computer Engineering. It is understood that any copying or publication or use of this dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to CEC in any scholarly use which may be made of any material in this project dissertation.

Shivangi M

Statement of Authenticity

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at College of Engineering Chengannur(CEC) or any other educational institution, except where due acknowledgement is made in the report. Any contribution made to my work by others, with whom I have worked at CEC or elsewhere, is explicitly acknowledged in the report. I also declare that the intellectual content of this report is the product of my own work done as per the **Problem Statement** and **Proposed Solution** sections of the project dissertation report. I have explicitly stated the major references of my work. I have also listed all the documents referred, to the best of my knowledge.

Shivangi M

Acknowledgements

The success and final outcome of any project requires a lot of guidance and assistance from many people and I am extremely fortunate to have all these along the completion of this project work. At the very offset of this project I am ineffable indebted to GOD ALMIGHTY for his showers of blessings on us for making this project a success.

With deep sense of reverence, I express my sincere thanks to Dr. Smitha Dharan, Principal, College of Engineering Chengannur for extending all the facilities required for doing my project, the Head of the Computer Science and Engineering Department Dr. Manju S Nair, Project Guide Mrs.Shiny B and Project Coordinator Mr. Ahammed Siraj K K for conscientious guidance and encouragement to accomplish this project.

I would like to express my gratitude to my parents for their constant moral and economic support. Last but not the least goes to all my friends who have directly or indirectly helped me for the completion of this project.

Shivangi M

Abstract

One of the most challenging and ongoing problems in computer vision is the detection of abnormal occurrences. Many surveillance cameras have been erected in public areas, including as airports, plazas, subway stations, and train stations, in response to the growing need for public safety. These cameras produce enormous volumes of visual data, making the process for processing that data laborious and wasteful for a human operator to look for unusual or suspicious activity.Many techniques have been tested to stop aggressive behaviour, which includes setting up surveillance equipment. It will be very good significance if monitoring technology can quickly identify violent behaviour and alert users or warning sign.The project proposal outlines a technique for identifying unusual occurrences in a any movie.All acts of violence are viewed as anomalous in this scenario.

Contents

List of Figures

Chapter 1

Introduction

While keeping an eye on public violence is crucial for safety and security, surveillance devices are now frequently used in public spaces and infrastructure. A vital function of video surveillance is detecting unusual events, such as accidents, robberies, or illegal activity. However, the majority of current monitoring systems also require manual examination and human operators (prone to disturbances and tiredness). Consequently, the demand for clever computer vision algorithms for automated video anomaly/violence identification is rising. Building algorithms to detect a specific anomalous occurrence, such as a violence detector, fight action detection, or traffic accident detector, is a tiny step toward resolving the detection of anomalies problem. Due to its auspicious performance in recent years, video action recognition has attracted more attention.

Recognizing anomalous events is challenging since it must be done on real-time videos captured by a large number of surveillance cameras at any time and in any location. It should be able to make reliable real-time detection and alert corresponding authorities as soon as violent activities occur. Public video surveillance systems are widespread around the world and can provide accurate and complete information in many security applications. However, having to watch videos for hours reduces your ability to make quick decisions. Video surveillance is essential to prevent crime and violence. In this regard, several studies have been published on the automatic detection of scenes of violence in video. This is so that authorities do not have to watch videos for hours to identify events that only lasts a few seconds. Recent studies have highlighted the accuracy of deep learning approaches to anomaly detection

Anomalies in the real world come in many different and complex forms. It is challenging to enumerate all potential abnormal occurrences. In order to avoid this, it is preferable that the anomaly detection algorithm not rely on any previous knowledge of the events. So, anomaly detection should be carried out with the least amount of monitoring possible. Approaches based on sparse-coding [2], [3] are regarded as exemplary techniques that produce cutting-edge anomaly detection outcomes. The first segment of a video is used to generate the normal event dictionary in these methods since it is assumed that just a modest initial piece of a movie contains normal events.

Several methods have been put forth for video action classification as a result of deep learning's successful image classification demonstration. [2] However, getting annotations for training is challenging and time-consuming, particularly for videos. To develop the model of typical behaviors, deep learning-based auto-encoders [4] used reconstruction loss to identify anomalies. The concept behind this study is to use the MobileNetv2 network to convert videos into image frames, classify them into anomalous occurrences like fighting, accidents, and fire, and then construct a real-time alert system. The model's output aberrant frames are saved, and then these frames, along with the time and place of the incident, are sent as an alarm to the necessary officials via the alert module.

1.1 Proposed Project

The proposed project describes a method for determining anomalous events in any video.In this scenario all violence activities is considered as an anomaly and raise an alarm message when anomaly is detected.

1.1.1 Problem Statement

- To develop an approach for detecting anomalous events from surveillance videos.

1.1.2 Proposed Solution

- To develop a deep learning based method to identify and classify anomalous events like traffic accidents, arson, fighting and normal event present in surveillance videos using MobileNetV2.
- Raise an alert message when an anomalous event occurs.

Chapter 2

Report of Preparatory Work

2.1 Literature Survey Report

- Localizing Anomalies From Weakly-Labeled Videos IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 30, 2021 [1]
  In this work it focuses on anomaly localization in surveillance videos and introduces a weakly supervised anomaly localization(WSAL) method focusing on temporally localizing anomalous segments within anomalous videos.A weak supervision enhancement strategy is further proposed to deal with noise interference and the absence of localization guidance in anomaly detection.Inspired by the appearance difference in anomalous videos, the evolution of adjacent temporal segments is evaluated for the localization of anomalous segments.implement and compare different SOTA anomaly detection approaches on the UCF-Crime and our TAD dataset. The experimental results showed that the proposed anomaly detector has performed significantly better than previous methods.

- MIST: Multiple Instance Self-Training Framework for Video Anomaly Detection [2]
  In this work, a multiple instance self-training framework (MIST) is developed.It is a self learning approach.Here, a multiple instance self-training framework that assigns clip-level pseudo labels to all clips in abnormal videos via a multiple instance pseudo label generator is introduced.The key of MIST is to design a two stage self training strategy to train a task-specific feature encoder for video anomaly detection.

- Real-world Anomaly Detection in Surveillance Videos [3]
  A deep learning approach to detect real world anomalies in surveillance videos is introduced.This approach begins with dividing surveillance videos into a fixed number of segments during training.In this approach, normal and anomalous videos is considered as bags and video segments as instances in multiple instance learning (MIL), and automatically learn a deep anomaly ranking model that predicts high anomaly scores for anomalous video segments. These segments make instances in a bag. Using both positive (anomalous) and negative (normal) bags, and trains the anomaly detection model using the deep MIL ranking loss

- Anomaly Detection With Particle Filtering for Online Video Surveillance.[4]
  With growing security threats, many online and offline frameworks have been proposed for

anomaly detection in video sequences. However, existing online anomaly detection techniques are either computationally very expensive or lack desirable accuracy. This research work proposes a novel particle filtering based framework for online anomaly detection which detects video frames with anomalous activities based upon the posterior probability of activities in a video sequence. The method also detects anomalous regions in anomalous video frames. Here a novel prediction and measurement models to accurately detect anomalous video frames and anomalous regions in video frames is introduced.

- Weakly-supervised Video Anomaly Detection with Robust Temporal Feature Magnitude Learning. [5]
  Here it is supervised anomaly detection.A Robust temporal feature magnitude (RTFM) approach is introduced.RTFM learns a temporal feature magnitude mapping function that
  1) detects the rare abnormal snippets from abnormal videos containing many normal snippets.
  it guarantees a large margin between normal and abnormal snippets.RTFM-enabled model learns more discriminate features that improve its ability in distinguishing complex anomalies.

- MobileNetV2: Inverted Residuals and Linear Bottlenecks.[6]
  Here it describes a new mobile architecture, MobileNetV2, that improves the state of the art performance of mobile models on multiple tasks and benchmarks as well as across a spectrum of different model sizes and describe efficient ways of applying these mobile models to object detection in a novel framework and call it SSD Lite. Additionally, demonstrate how to build mobile semantic segmentation models through a reduced form of DeepLabv3 which we call Mobile DeepLabv3,is based on an inverted residual structure where the shortcut connections are between the thin bottleneck layers. The intermediate expansion layer uses lightweight depthwise convolutions to filter features as a source of non-linearity. Additionally, find that it is important to remove non-linearities in the narrow layers in order to maintain representational power and demonstrate that this improves performance and provide an intuition that led to this design.

- Deep Residual Learning for Image Recognition.[9]
  Deeper neural networks are more difficult to train and present a residual learning framework to ease the training of networks that are substantially deeper than those used previously. It explicitly reformulate the layers as learning residual functions with reference to the layer inputs, instead of learning unreferenced functions. It provide comprehensive empirical evidence showing that these residual networks are easier to optimize, and can gain accuracy from considerably increased depth. On the ImageNet dataset it evaluate residual nets with a depth of up to 152 layers and deeper than VGG nets but still having lower complexity. An ensemble of these residual nets achieves 3.57 error on the ImageNet test set. The depth of representations is of central importance for many visual recognition tasks. Solely due to our extremely deep representations, it obtain a 28 percent relative improvement on the COCO object detection dataset, ImageNet localization, COCO detection, and COCO segmentation.

2.2 System Study Report

Anomaly detection is one of the most challenging and long standing problems in computer vision.Real-world anomalous events are complicated and diverse. Most existing monitoring systems, however, also need human operators and manual inspection which is prone to disturbances and tiredness.Manual inspection is tiresome and time consuming,the data generated from the surveillance camera is large resulting in an inefficient and exhausting process for a human operator to find suspicious or anomalous occurrences.It is difficult to list all of the possible anomalous events. Therefore, it is desirable that the anomaly detection algorithm does not rely on any prior information about the events. In other words, anomaly detection should be done with minimum supervision.

Based on the experimental setting on the training data, video anomaly detection methods can be generally classified into three categories, i.e. unsupervised, weakly-supervised, and supervised.[12] Since real-world anomaly events happen with low probability, it is hard to capture all types of anomaly. However, normal videos are easy to access from social media and public surveillance, unsupervised methods are thus motivated to detect anomaly events with only normal videos in the training set. Although the unsupervised methods are not able to achieve satisfactory performance on complex real-world scenarios, they are believed to have better generalization ability on unseen anomaly patterns.[12].Early unsupervised methods mainly adopt classic machine learning techniques with hand-crafted features as well probability models.deep learning techniques were able to take advantage of large-scale data set and powerful computation resource. Following the setting of unsupervised anomaly detection, a number of works are proposed based on deep AE (auto encoder).For certain application scenarios where the anomaly activities are well defined, the performance can be significantly improved by introducing supervision information. Recent works follow the weakly supervised setting where only video-level annotation is available for training. That is the training videos are labeled with norma roads and cars for highway traffic accidents detection, recent works are usually based on the frame-level annotated training videos (i.e. the temporal annotations of the anomalies in the training videos are available – supervised setting). A popular solution is to leverage the geometric prior knowledge and object detection with additional supervision from other public data sets.[12]

Chapter 3

Project Design

3.1 Resource Requirements

Many resources were used for developing the project.Python language is used.The data is trained in colab to use google's gpu instance.The data is a total of 5GB.

3.1.1 Hardware & Software Requirements

Google Colab : Environment for running python and able to use Google's GPU.

Jupyter notebook: Environment used for running python code in system.

Python,Tensorflow,Keras,openCV ,Python Flask : supporting software and libraries.

Operating system : Windows

3.1.2 Data Requirements

The model is trained using data from the UCF-Crime dataset. Long uncut surveillance footage covering 13 real-world abnormalities, such as abuse, arrest, arson, assault, road accidents, burglaries, explosions, fighting, robberies, shooting, shoplifting, and vandalism, make up this collection. Only three groups of anomalies—out of a total of 13—are taken into consideration because they significantly affect public safety. 950 of them are regular films, while the remaining videos each have at least one anomalous incident. 800 regular videos and 810 abnormal videos make up the training set. Temporally annotated footage from the remaining 140 anomalous and 150 regular videos has been verified. All 13 anomalous events are included in both the training and testing sets. Every video is accurate for use in actual surveillance applications.

Another data set having real-world violent and nonviolent activities is used and this dataset is obtained from Kaggle. This dataset Contains 1000 Violence and 1000 non-violence videos collected from youtube videos, violence videos in the data set contain many real street fight situations in several environments and conditions also non-violence videos from this dataset are collected from

many different human actions like sports, eating, walking, etc.

3.2 Method

Here, the data is trained using MobileNetV2. Use of the transfer learning principle. A convolutional neural network design called MobileNetV2citeref6 aims to function well on mobile devices. It is built on an inverted residual structure where the bottleneck layers are connected by residual connections. Lightweight depth-wise convolutions are used in the intermediate expansion layer as a source of non-linearity to filter features. The architecture of MobileNetV2 includes a 32-filter initial fully convolution layer as well as 19 additional bottleneck layers.

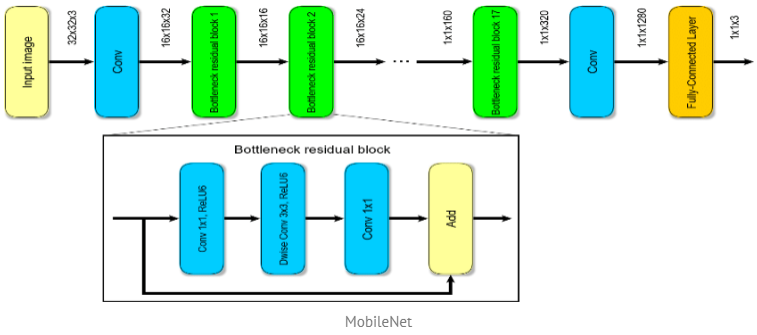

Figure 3.1: MobileNet

The MobileNet V2 model features 1 AvgPool with around 350 GFLOP and 53 convolution layers. Inverted Residual Block and Bottleneck Residual Block make up its two primary parts. In the MobileNet V2 design, there are two different types of convolution layers: 1x1 and 3x3 depth-wise.

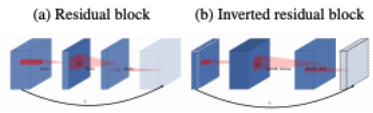

Figure 3.2: Blocks of MobilenetV2

There are two categories of blocks in MobileNetV2. With a stride of 1, one is the remaining block. Another is a block with a two-step stride for shrinking. Both sorts of blocks have three levels. One-to-one convolution with ReLu6 makes up the initial layer this time. Convolution based on depth is the second layer. If ReLu is employed, the deep networks only have the ability of a linear classifier on the non-zero volume portion of the output domain and there is an expansion factor t. The third layer is another 11 convolution but this time without any non-linearity. And for all major experiments, t=6. Input 64 channels, internal output 64t=646=384 channels, and so on.

The core network (width multiplier 1, 224224) typically utilizes 3.4 million parameters and costs

Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	k	-	

MobileNetV2 Overall Architecture

Figure 3.3: MobileNet v2 Architecture Summary

300 million multiply-adds to compute. (MobileNetV1 introduces the width multiplier.) Further investigation of the performance trade-offs is done for input resolutions ranging from 96 to 224 and width multipliers ranging from 0.35 to 1.4. While the model size varies between 1.7M and 6.9M parameters, the computational cost of the network can reach 585M Adds. With a batch size of 96, 16 GPU are used to train the network.

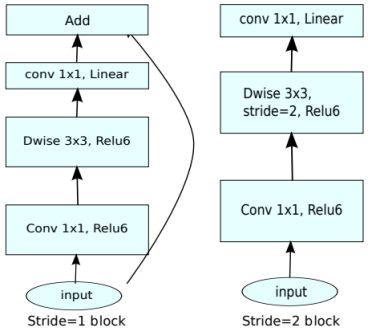

Figure 3.4: MobileNetv2

Depthwise Separable Convolutions

The fundamental building element for many effective neural network topologies is depth-wise separable convolutions. This is used by MobileNetv2 as well. The fundamental concept is to substitute a fully convolutional operator with a factorized variant that divides convolution into two distinct layers. A single convolution filter is applied to each input channel in the first layer, which is referred to as a depth-wise convolution and does light filtering. A 1 x 1 convolution, also known

as a point-wise convolution, makes up the second layer and is used to create new features by computing linear combinations of the input channels.

Linear Bottlenecks

MobileNetV2 is built on a depth-separable convolution with residuals that encounters bottlenecks. Because each bottleneck block contains an input, numerous bottlenecks, and then expansion, they resemble residual blocks. [9] The information needed is actually stored in the bottlenecks, whereas an expansion layer is only an implementation detail that goes along with a non-linear tensor transformation. Direct routes are chosen between bottlenecks in this area.that happens when the tensor undergoes a non-linear transformation.Here, shortcuts are taken directly between bottlenecks.

3.3 Classification Using MobileNetV2

The data is classified using the MobileNetV2 network. Tensorflow is used to train the model[10]. TensorFlow, an open-source library used to create a machine learning and deep learning models, was one of the libraries used. It develops input, hidden, and output layers in deep neural networks. and add unique layers to the MobilenetV2 that have already been trained using TensorFlow. The classification model will benefit from this and perform better. The pre-trained MobilenetV2 model is downloaded from the Tensorflow Hub, an open-source repository that has pre-trained models for image classification and natural language processing applications. To visualize the image dataset and create diagrams, we utilize the Matplotlib module. Results of the model's forecast will be displayed. The image dataset will be turned into an array by Numpy.
Based on the existing MobilenetV2's pre-trained

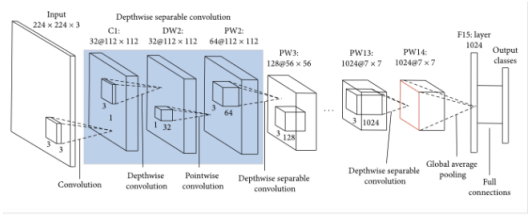

Figure 3.5: MobileNetv2 Architecture

Chapter 4

Implementation

The aim is to develop a real-time surveillance system that can recognize anomalous activities and give alerts to notify the concerned authorities. As the initial step, the videos from surveillance videos are broken down into frames. A four-class classification is done using the MobileNetv2 model. It makes use of a transfer learning strategy. Transfer learning is not a novel concept. We have been searching for strategies to aid machines in remembering what they have already learned since we first started teaching them how to learn, classify, and forecast data. Both humans and machines have a difficult time learning new things. It is a difficult, resource-intensive, and time-consuming activity, thus it was crucial to build a strategy that would prevent a model from forgetting the learning curve that is obtained from a specific dataset and also enable it to learn more from new and varied datasets. Transfer learning is the process of applying a model that has already been trained on one dataset to make predictions on a new dataset.
The frames are given as input to the MobileNet v2 classifier for detecting anomalous activities like Accident, Arson,Fighting, and to detect a normal event in the given sequence of input frames.

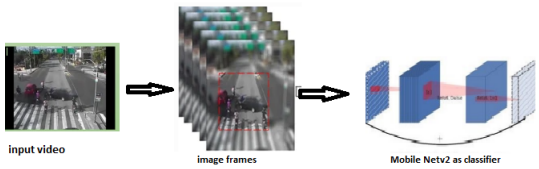

Figure 4.1: Video to Imageframes and Classification

The second step is to build the alert module, first, the normal frames are discarded that is if no anomalous activity is recognized then the respective frames are discarded. Image Enhancement is performed on the frames that are obtained as output. This is performed using the inbuilt functions provided by the Python Imaging Library. The anomaly detected frame is obtained and it is enhanced for better clarity. That frame, along with the location is sent to the nearest authorities using the Telegram bot.

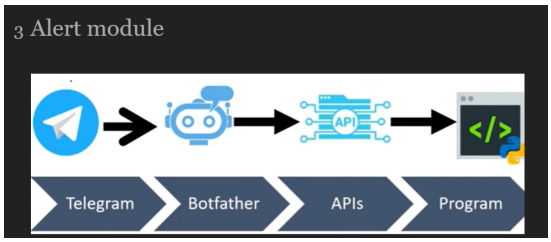

Figure 4.2: Alert Module

When an anomalous event is identified in a frame, the system sets a counter variable to one and checks subsequent frames to see if they have the same anomaly detected true. Each time an abnormality occurs in a consecutive frame, the counter is increased. If a frame is false, the counter variable is reset to zero, and the next frame is checked to see if an abnormal behavior has been detected. If the violence is recognized for 30 consecutive frames, the current time is calculated using a built-in python function, and an alert is issued to a Telegram group consisting of officials from higher authorities. The image of the identified anomalous action, as well as the current time, are included in the Alert message.

The proposed model is implemented using Keras and TensorFlow framework using the Google Colab platform. The model is then trained with ucf crime dataset and real-time violence dataset with a batch size of 96 and an image size of 224×224. Adam optimizer is used and the learning rate is 1e-4. And does not use any pre-processing strategies. For training, the input data which are videos are converted into image frames as a first step. The model is trained using these image frames. Four classifications of anomalous activities are done here. The model is tested using anomalous videos like Arson, Accident, Fighting, and normal videos. The output from the model will be image frames with the type of anomaly written on each frame using OpenCV. Finally, as output, the image frames are converted into videos.

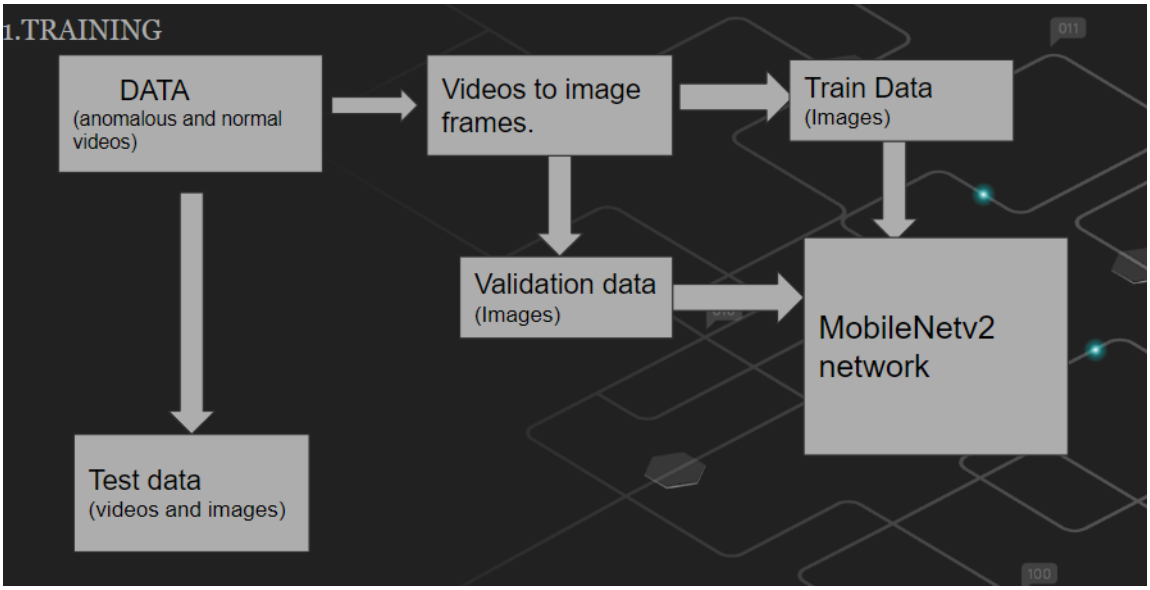

Figure 4.3: Training Block diagram

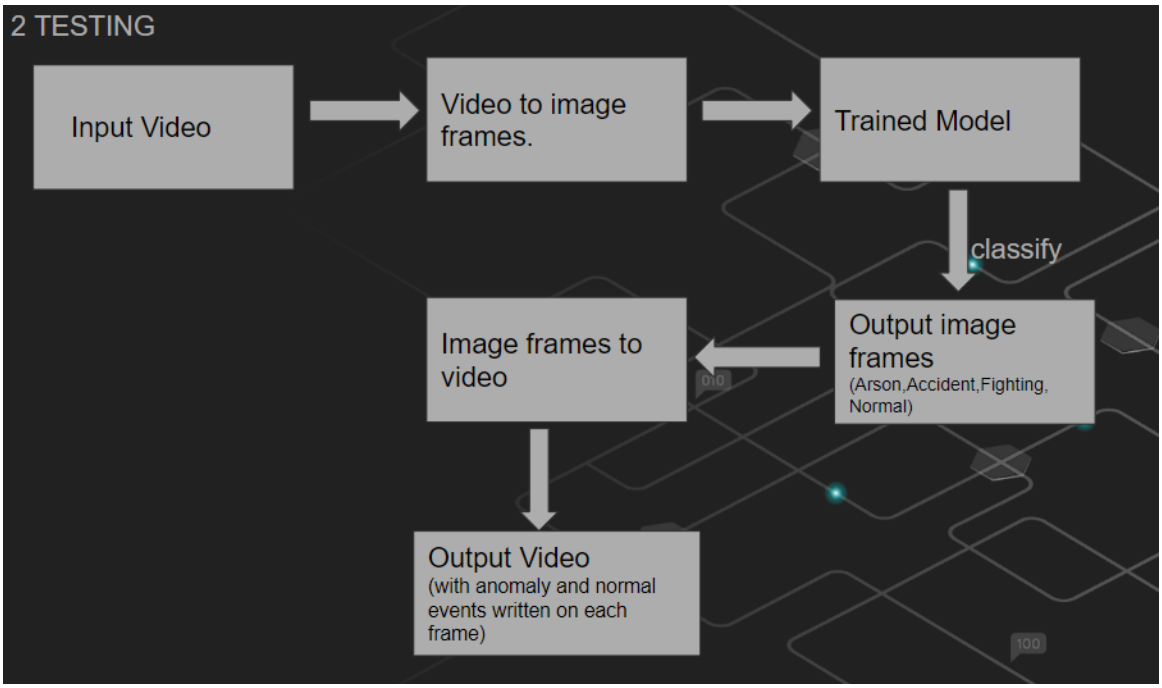

Figure 4.4: Testing Block diagram

Chapter 5

Results & Conclusions

5.1 Performance Analysis

The model is tested with multiple videos from different sources like youtube etc. When the model was tested with 12 videos among the 12 videos ,the image frames generated from 10 videos gave accurate result. Performance accuracy of the model is calculated using the equation, (number of correct predictions/Total predictions) $\times$ 100.

When an anomaly is detected the telegram bot sends the image of an anomalous event to the telegram group where the concerned authorities are notified.

5.2 Conclusion

The project is based on the detection of anomalies in surveillance videos, here in this scenario anomaly is considered as a violent or criminal activity like fighting, robbery, vandalism, etc. A MobileNetV2 network is used as it is a better option for real-time implementation. The dataset is preprocessed that is which is converted from video to image frames for testing purposes. It is a real-time violence detector using MobileNetV2 pretrained model, giving the output in the form of images with the result printed and written on each image using OpenCV, implemented using python. Due to the different materials and substantial fluctuations in quality, detecting anomalies in real-time is a difficult task. The MobileNet v2 model is used in this study. It's very quick to compute, making it excellent for application in time-sensitive situations, low-cost devices, and applications.

A MobileNetV2 network is used as it is a better option for realtime implementation.The dataset is preprocessed that is it is converted from video to image frames for testing purpose.It is a real-time violence detector using MobileNetV2 pretrained model, giving the output in the form of images with the result printed written on each image using OpenCV, implemented using python.Due to the different materials and substantial fluctuations in quality, detecting anomalies in real-time is a difficult task. The MobileNet v2 model is used in this study. It's very quick to compute, making it excellent for application in time-sensitive situations,low-cost devices and applications.

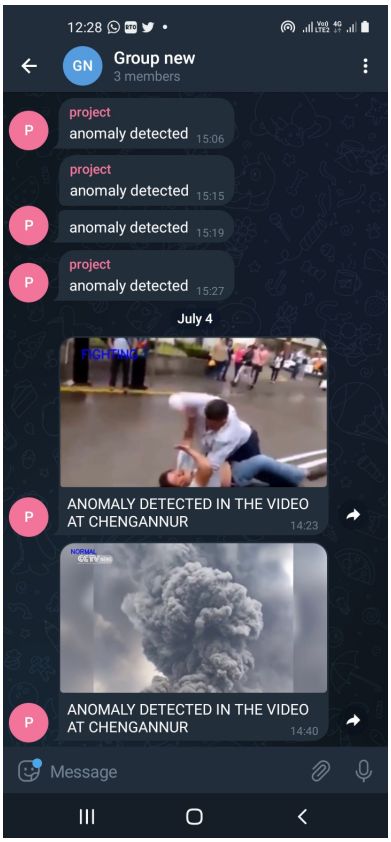

Figure 5.1: Screenshot of Alert message in Telegram group

5.3 Future Scopes

This model might be updated to simultaneously operate on many cameras connected by a single network.Together with the alert message, a short video of the aggressive conduct could be included. Due to the numerous variables, may identification in real time is a difficult task.Human recognition can also be included,so that the criminal can be identified from the resultant image frames.It can be considered as a simple and easy to implement method for detecting anomalous activities.

References

[1] Zhen Cui Member Chunyan Xu Yon Li Hui Lv, Chuanwei Zhuo and JianYang. Localizing anomalies from weakly-labeled videos. IEEE TRANSACTIONS ON IMAGE PROCESSING, .

[2] Jia-Chang Feng, Fa-Ting Hong, and Wei-Shi Zheng. Mist: Multiple instance self-training framework for video anomaly detection. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 2021.

[3] Deep anomaly detection through visual attention in surveillance videos Nasaruddin Nasaruddin, Kahlil Muchtar, Afdhal Afdhal Alvin Prayuda Juniarta Dwiyantoro Journal of Big Data

[4] HAROON FAROOQ ABDUL JALEEL ATA-URREHMAN, SAMEEMA TARIQ and SYED MUHAMMAD WASIF. Anomaly Detection With Particle filtering for Online Video Surveillance. IEEE ACESS, 7(3), September.

[5] Yuanhong Chen Johan W.Verjans Gustavo Carneiro. Yu Tian, Guansong Pang. Weakly-supervised video anomaly detection with robust temporal feature magnitude learning. IEEE Conference on Computer Vision and Pattern Recognition, 7(3), September 2021.

[6] MobileNetV2: Inverted Residuals and Linear Bottlenecks CVPR 2018, Mark Sanler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen

[7]] Real-world Anomaly Detection in Surveillance Videos CVPR 2018 · Waqas Sultani, Chen Chen, Mubarak Shah ·

[8] Real-Time Anomaly Detection and Feature Analysis Based on Time Series for Surveillance Video May 2021 · Ruoyu Xue, Jingyuan Chen, Yajun Fang

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. CoRR, abs/1512.03385, 2015

[10] Mart´ın Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mane,´ Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viegas, Oriol Vinyals, Pete Warden, Martin Wat- ´ tenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org. 5, 6.

[11] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam.Mobilenets: Efficient convolutional neural networks for mobile vision applications. CoRR, abs/1704.04861, 2017.

[12] Zhen Cui Member Chunyan Xu Yong Li Hui Lv, Chuanwei Zhou and JianYang. Localizing anomalies from weakly-labeled videos. IEEE TRANSACTIONS ON IMAGE PROCESSING, 30(1), AUGUST 2021.

[13] Chang Feng, Fa-Ting Hong, and Wei-Shi Zheng. Mist: Multiple instance self-training framework for video anomaly detection. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 2021.

[14] Mubarak Shah Waqas Sultani, Chen Chen. Real-world Anomaly Detection in Surveillance Videos. IEEE ACESS, 13(3), March 2020.

[15] HAROON FAROOQ ABDUL JALEEL ATA-UR-REHMAN, SAMEEMA TARIQ and SYED MUHAMMAD WASIF. Anomaly Detection With Particle filtering for Online Video Surveillance. IEEE ACESS, 7(3), September 2021.

[16] Yuanhong Chen Johan W.Verjans Gustavo Carneiro. Yu Tian, Guansong Pang. Weaklysupervised video anomaly detection with robust temporal feature magnitude learning. IEEE Conference on Computer Vision and Pattern Recognition, 7(3), September 2021.

[17] Afdhal Afdhal Nasaruddin Nasaruddin, Kahlil Muchtar and Alvin Prayuda Juniarta Dwiyantoro. Deep anomaly detection through visual attention in surveillance videos. JOURNAL OF BIG DATA, 7(8), SEPTEMBER 2020.

[18] Rob Fergus Lorenzo Torresani Manohar Paluri Du Tran, Lubomir Bourdev. Learning spatiotemporal features with 3d convolutional networks. In Proc. ICCV VS-PETS,, 5(2), october 2015