Smart Internz

PROJECT TITLE:

<u>VIOLENCE DETECTOR USING DEEP LEARNING</u>

Introduction:

<u>Overview:</u>

With the rapid growth of surveillance cameras to monitor the human activity demands such system which recognize the violence and suspicious events automatically. In smart buildings, this technology is mandatory. *Automatic Violence Detection will prevent any mis-happenings from taking place by raising alerts to the security guards.*

<u>Purpose:</u>

Our goal is to reduce public violence by having smart AI driven security cameras monitoring the areas 24x7.

Literature survey:

<u>Existing problem:</u>

Public places like shopping centers, avenues, banks, etc. are increasingly being equipped with CCTVs to guarantee the security of individuals. Subsequently, this inconvenience is making a need to computerize this system with high accuracy. Since constant observation of these surveillance cameras by humans is a near impossible task.

It requires workforces and their constant attention to judge if the captured activities are anomalous or suspicious.

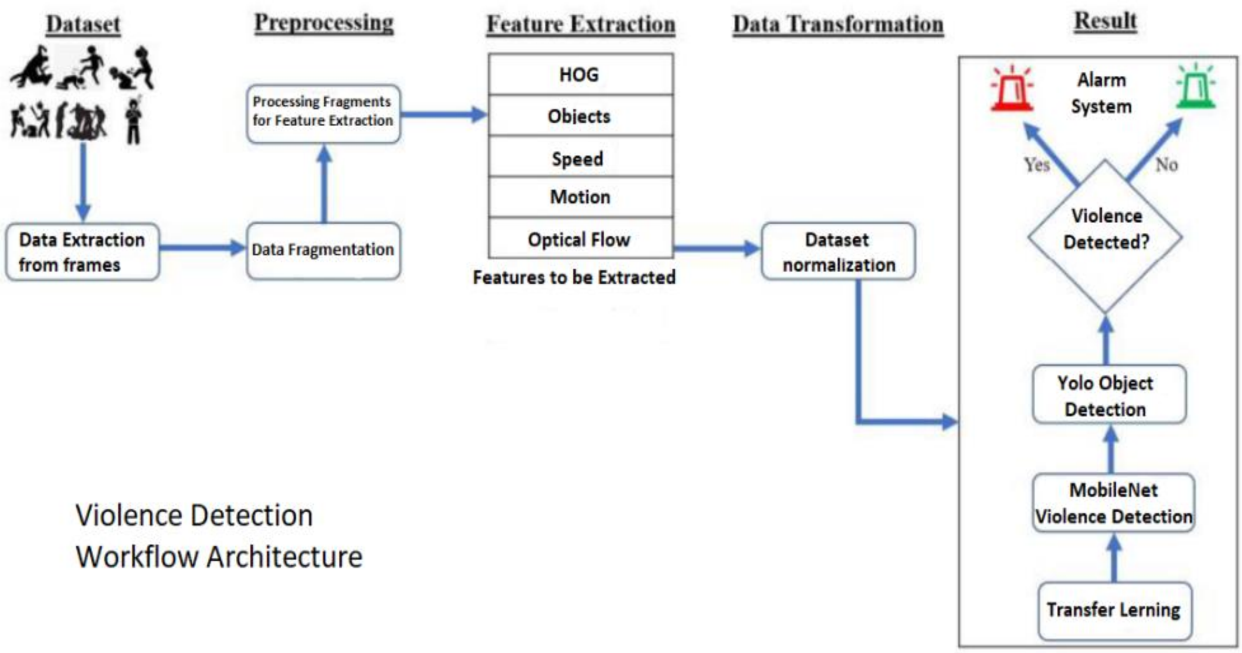
Proposed solution:

There is a need to display which frame and which parts of the recording contain the uncommon activity which helps the quicker judgment of that unordinary action being unusual or suspicious. Therefore, to reduce the wastage of time and labour, *we are utilizing deep learning algorithms for automated violence detection system.* Its goal is to automatically identify signs of aggression and violence from the video, which filters out irregularities from normal patterns. *We intend to utilize MOBILE NET , a pre built model to identify and classify levels of high movement in the frame.* From there, we can raise a detection alert for the situation of a threat, indicating the suspicious activities at an instance of time.

Deep learning techniques have been proven effective in extracting spatial-temporal features from videos, i.e., features that represent the motion information contained in a sequence of frames, in addition to the spatial information contained in a single frame.

Theoretical analysis:

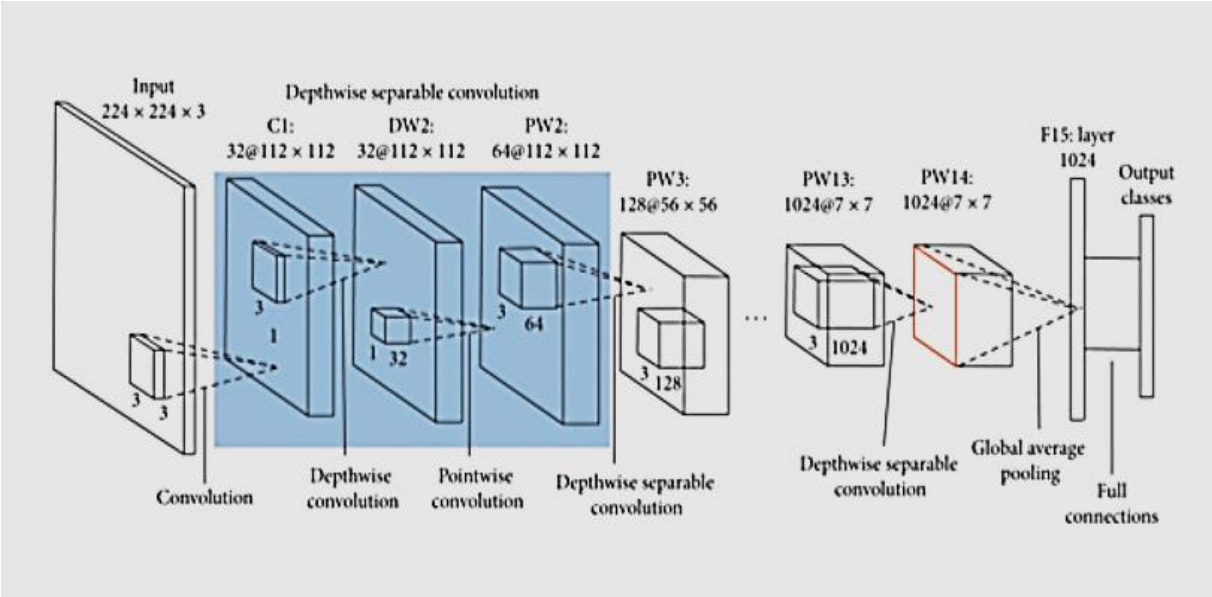
Block diagram:



Violence Detection
Workflow Architecture

MobileNet is a streamlined architecture that uses depth wise separable convolutions to construct lightweight deep convolutional neural networks and provides an efficient model for mobile and embedded vision applications. Transfer learning will be applied to this to train the model for violent activities detection as well as object detection.

Here is the architecture of mobileNET:



Hardware/Software components required:

HARDWARE COMPONENTS

1. *Laptop Camera* - For capturing realtime video, laptop camera was used.

2. *GPU*- GPU used to run the project was Nvidia GeForce GTX

1650 TI.

SOFTWARE COMPONENTS

1. Python

2. Jupyter Notebook

3. Google Collab

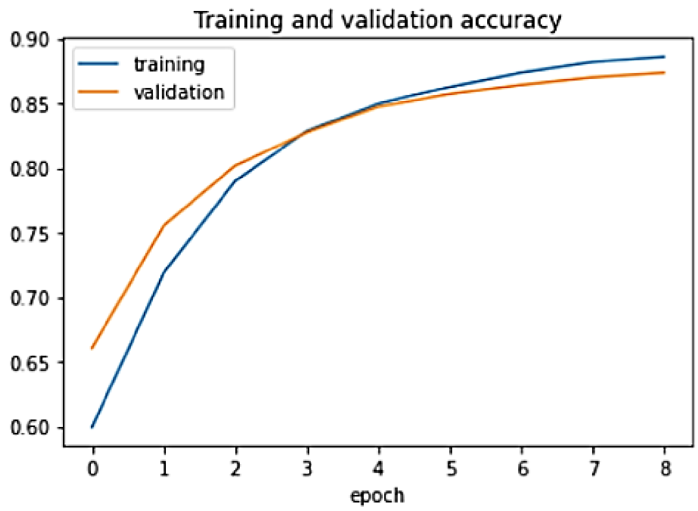
EXPERIMENTAL INVESTIGATIONS:

Because of limited space allotted on Google Collab, I was only able to train the model with 700 videos (350 violence, 350 non-violence). So, the accuracy of the model is 89% but it can be 95-99%. If we are able to train it with 2000 videos.

Best Epoch	9
Accuracy on train	0.888836801
Accuracy on test	0.87644875
Loss on train	0.26514703
Loss on test	0.284462363

Table 1- Evaluation Matrix

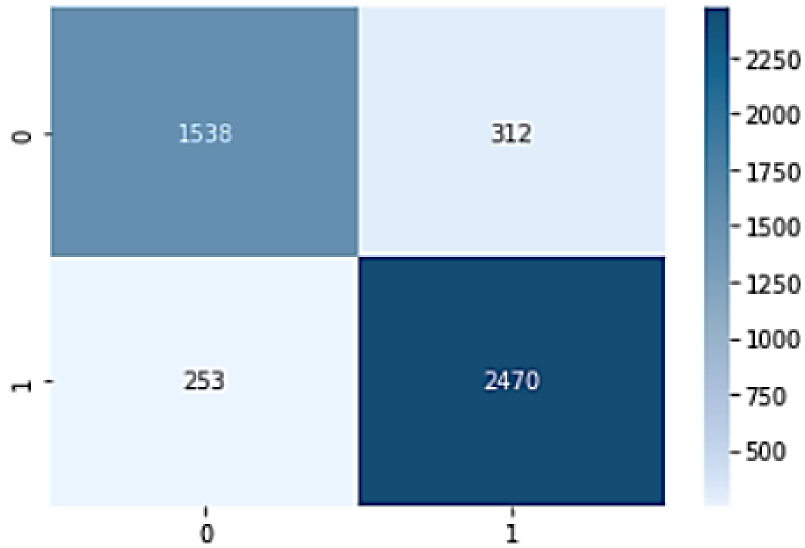
The training loss indicates how well the model is fitting the training data, while the validation loss indicates how well the model fits new data. The training and validation loss and training and validation accuracy curves are given below:



Confusion Matrix below depicts the different outcomes of the prediction and results of a classification problem and helps visualize its outcomes.

The confusion matrix of the violence detection model prediction is given below:

```
> Correct Predictions: 4008
> Wrong Predictions: 565
```



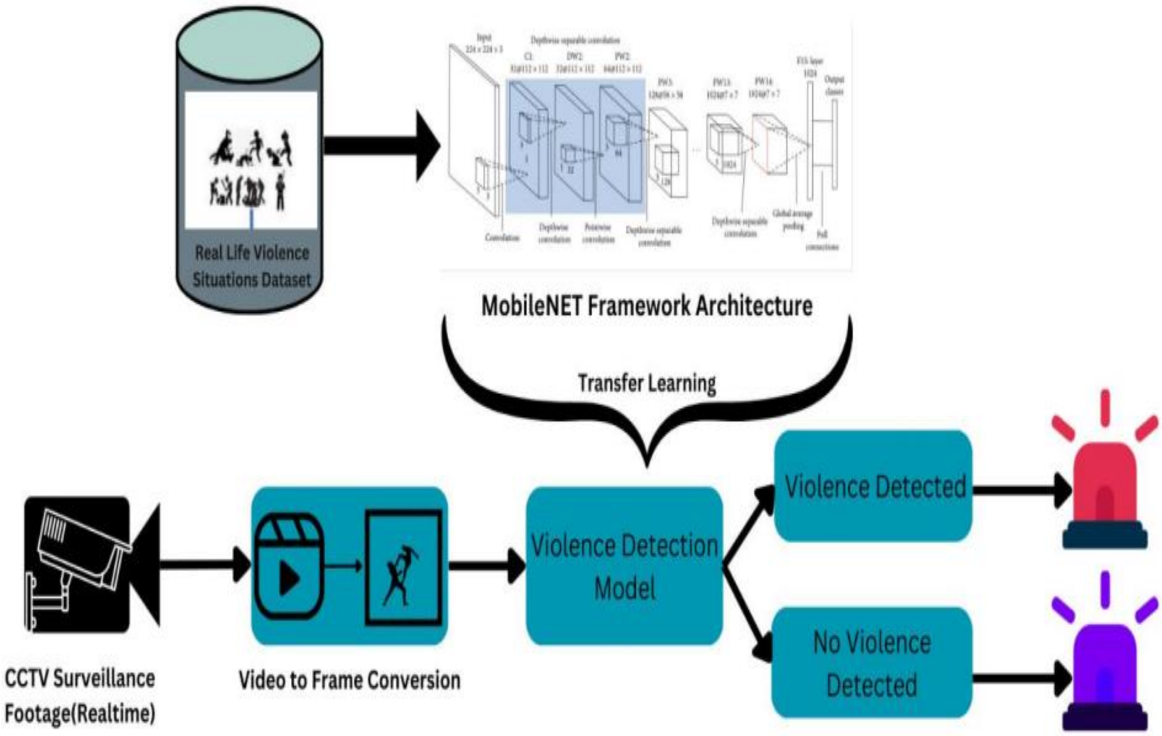
Confusion Matrix		
	Actually Positive	Actually Negative
Predicted Positive	True Positive = 1538	False Positive = 312
Predicted Negative	False Negative = 253	True Negative = 2470

Table 2- Confusion Matrix Values

Clasiffication Report				
	Precision	Recall	f1-score	Support
NonViolence	0.86	0.83	0.84	1850
Violence	0.89	0.91	0.9	2723
Accuracy			0.9	4573
Macro avg	0.87	0.87	0.87	4573
Weighted avg	0.88	0.88	0.88	4573

Table 3 – Classification Report

FLOWCHARTS:



Results:

<u>Screen shots:</u>

2 people standing-

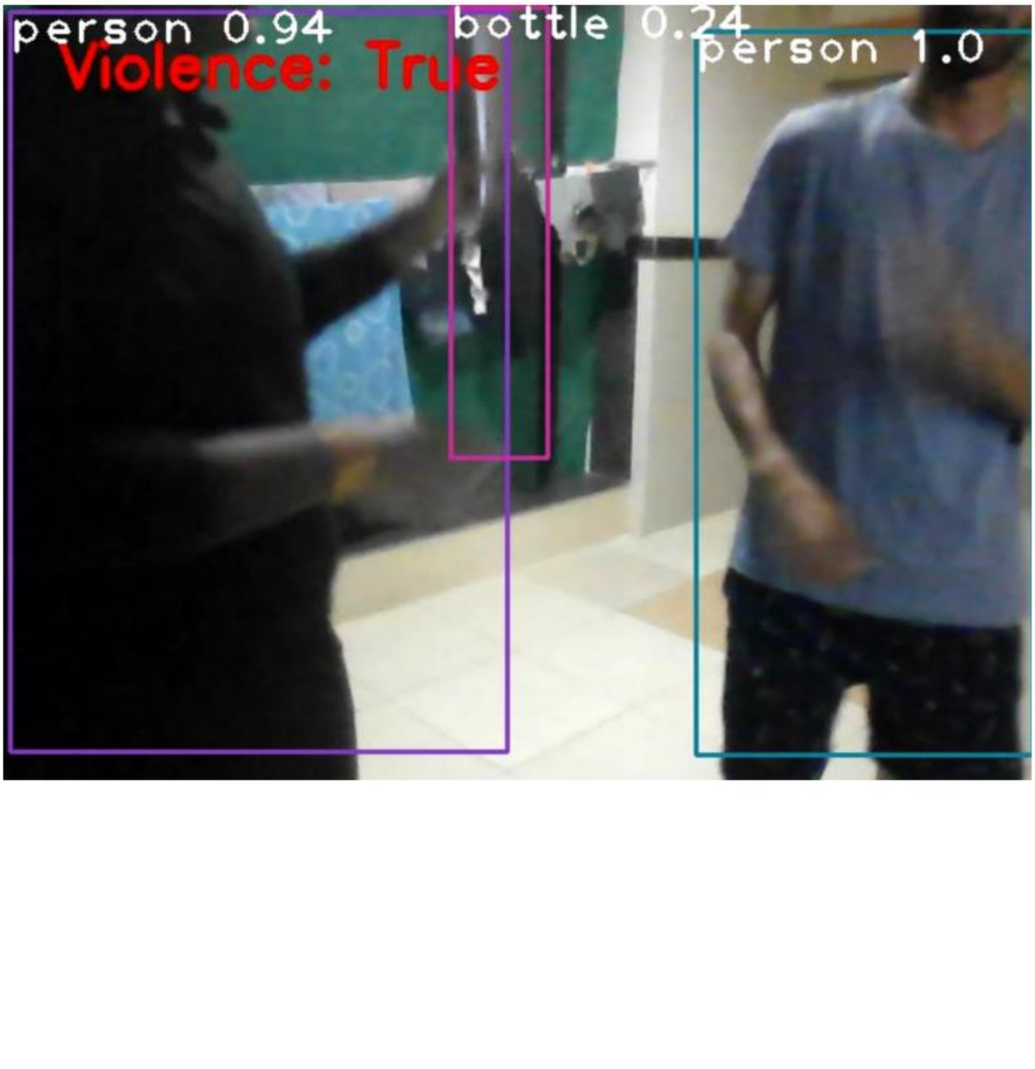
After slap-

Model does not detect hug as violence-

When the person is taken down, violence is detected-

Choke-

When the person is hit with an object, in this case a bottle, violence is detected-



Advantages/Disadvantages

ADVANTAGES-

1. Automatic and fast detection of violence.
2. No more tedious work for the security guards.
3. Less cost for labor.

DISADVANTAGES-

1. The model might not be able to pick up violence from far away.
2. Strong internet connectivity will be needed.
3. Won't be able to detect gun shots.
4. We were able to train the model only on 700 videos out of 2000 due to shortage of space in google collab.

Applications-

Public video surveillance systems are common all over the world, being capable of providing accurate and rich information in many security applications. However, the need of watching hours of video footages undermines the chance to take decisions in a short time, which *is essential in video surveillance for crime and violence prevention.*

Conclusion-

Artificial intelligence is crucial for spotting violence in video footage. Violence detection is still an issue even though it appears to be complicated at the moment. With this paper, we aimed to demonstrate how violence detection is possible and how it can be easily implemented using the most basic techniques currently available, all due to the ongoing developments in deep learning and AI. These models have learned how to extract dynamic and practical features from everyday images and use them as a jumping-off point for learning new tasks and also alerts when violence is detected and produce the beep sound, so that the nearest authority can take the action according to the situation. After being trained on more than a million images, these networks can categorize images into violent and non violent classes. Transfer learning was used with pre-trained networks because it is frequently simpler and faster than creating or training a network from inception.

Future scope-

1. Model performance and accuracy can be improved with more number of footage videos.
2. Videos should contain various kinds of action movements so that the model can generalize better.
3. Alerting system could also be made efficient and quick.